UNIVERSITAT DE BARCELONA

# The Onset of Geometry in Complex Networks

## From Stuctural Properties to Dynamical Processes

Jasper Eibertus van der Kolk

*PhD Thesis*

# The Onset of Geometry in Complex Networks

## From Stuctural Properties to Dynamical Processes

**Jasper van der Kolk**


UNIVERSITAT DE BARCELONA

# THE ONSET OF GEOMETRY IN COMPLEX NETWORKS

## From structural properties to dynamical processes

Programa de doctorat en Física

Autor: JASPER EIBERTUS VAN DER KOLK

Directors: MARIÁN BOGUÑÁ ESPINAL, M. ÁNGELES SERRANO MORAL

Tutor: GIANCARLO FRANZESE


UNIVERSITAT DE BARCELONA

Jasper Eibertus van der Kolk: *The Onset of Geometry in Complex Networks,* From structural properties to dynamical processes

## SUMMARY

Over the past few decades, the use of complex networks to describe the properties of systems of many interacting parts has become widespread in many fields of science. Surprisingly, networks from disparate disciplines share a wide range of basic properties, such as small worldness, high levels of clustering and broad degree distributions. One of the most promising frameworks to explain this observation is that of network geometry, where nodes are assumed to live in some underlying metric space that conditions their connectivity. The fact that this approach can reproduce all the basic network properties and symmetries as well as produce strong results in practical tasks such as community detection and link prediction has led many to wonder if there is a way to determine if real networks are indeed geometric in nature. However, these studies do not contemplate the fact that the transition between non-geometric and geometric networks might not be sharp. In this thesis, we study the effect of the underlying metric space on the complex network for different geometric coupling strengths. We show that three different regimes can be identified: In the non-geometric regime, where the coupling is extremely weak, results are similar to those of the configuration model, which is explicitly non-geometric. Increasing the coupling slightly leads us to the quasi-geometric regime, where the decrease of the clustering coefficient with the system size is extremely slow, leading to significant levels of this quantity for finite systems. Additionally, we show that, here, geometric information can be extracted from the topology alone through network embedding, and that it is essential for obtaining self-similar network replicas through geometric renormalization. A large number of empirical networks are best described in the quasi-geometric regime. Increasing the coupling further leads to a phase transition which we show to be topological in nature. We enter the geometric region, which is the regime typically studied in past works where the effects of the underlying metric space are strong and where clustering remains finite in the thermodynamic limit. Motivated by these results for single-layer graphs we also study geometric multiplexes. We introduce the mutual network, which is made up of the edges that are shared by all layers. This network allows us to obtain rigorous results on edge and triangle overlap. We show that it is generally the more geometric links that lead to overlap, increasing the effective geometric coupling in the mutual network. Finally, having extensively investigated the structural properties at various coupling strengths, we lastly turn to dynamical processes running of top of the network. Specifically, we show that the underlying metric space reveals periodic Turing patterns, both in the quasi- and strongly geometric regimes as well as in empirical networks. All these results show that the underlying geometry is essential for understanding complex networks, both from a structural as well as dynamical point of view.

## RESUM

En les darreres dècades, l'ús de xarxes complexes per descriure les propietats de sistemes amb moltes parts que interactuen ha esdevingut habitual en molts camps de la ciència. Sorprenentment, xarxes de disciplines diverses comparteixen una àmplia gamma de propietats bàsiques, com ara la propietat "small world", alts nivells d'agrupament i distribucions de grau àmplies. Un dels marcs teòrics més prometedors per explicar aquesta observació és el de la geometria de xarxes, on s'assumeix que els nodes es troben en un espai mètric subjacent que condiciona la seva connectivitat. En aquesta tesi estudiem l'efecte de l'espai mètric subjacent en xarxes complexes per diferents intensitats de l'acoblament geomètric. Identifiquem tres règims diferents: en el règim no geomètric, on l'acoblament és extremadament feble, els resultats són similars als de models explícitament no geomètrics. Un augment lleuger de l'acoblament ens porta al règim quasi-geomètric, on l'escalat del coeficient d'agrupació amb la mida del sistema és molt lent, cosa que genera nivells significatius d'aquesta quantitat per a sistemes finits. A més, mostrem que, aquí, es pot extreure informació geomètrica només de la topologia i i que aquesta informació és essencial per obtenir rèpliques auto-similars de xarxes mitjançant la renormalització geomètrica. Finalment, estudiem un gran nombre de xarxes empíriques i mostrem que es descriuen millor en el règim quasi-geomètric. Augmentar l'acoblament encara més condueix a una transició de fase de naturalesa topològica. Entrem en la regió geomètrica, que és el règim típicament estudiat en treballs anteriors, on els efectes de l'espai mètric subjacent són forts i on l'agrupament es manté finit en el límit termodinàmic. Motivats per aquests resultats en grafs de capa única, també estudiem grafs geomètrics de múltiples capes, els multiplexos. Introduïm la xarxa mútua, que està formada per les arestes compartides per totes les capes. Aquesta xarxa ens permet obtenir resultats rigorosos sobre la superposició d'arestes i triangles. Mostrem que, en general, són els enllaços més geomètrics els que es conserven a la xarxa mútua, augmentant l'acoblament geomètric efectiu. Finalment, després d'haver investigat extensament les propietats estructurals a diferents intensitats d'acoblament, ens centrem en processos dinàmics sobre la xarxa. Concretament, mostrem que l'espai mètric subjacent revela patrons periòdics de Turing, tant en els règims quasi-geomètrics com en els fortament geomètrics, així com en xarxes empíriques. Tots aquests resultats mostren que la geometria subjacent és essencial per entendre les xarxes complexes, tant des d'un punt de vista estructural com dinàmic.

# SAMENVATTING

In de afgelopen decennia is het gebruik van complexe netwerken om de eigenschappen van systemen met veel interagerende onderdelen te beschrijven, gebruikelijk geworden in veel wetenschappelijke disciplines. Verrassend genoeg delen netwerken uit verschillende vakgebieden een breed scala aan basiskenmerken, zoals de "small world"-eigenschap, hoge niveaus van clustering en brede graaddistributies. Een van de meest veelbelovende kaders om deze waarneming te verklaren, is "network geometry", waarbij wordt aangenomen dat de knopen zich in een onderliggende metrische ruimte bevinden die hun connectiviteit bepaalt. In dit proefschrift bestuderen we het effect van de onderliggende metrische ruimte op complexe netwerken bij verschillende geometrische koppelingsterktes We identificeren drie verschillende regimes: in het niet-geometrische regime, waar de koppeling extreem zwak is, zijn de resultaten vergelijkbaar met die van expliciet niet-geometrische modellen. Een lichte toename van de koppeling leidt ons naar het quasi-geometrische regime, waar de clustercoëfficiënt zeer langzaam schaalt met de systeemgrootte, wat significante niveaus van deze grootheid genereert voor eindige systemen. Bovendien laten we zien dat in dit regime geometrische informatie kan worden afgeleid enkel uit de topologie door netwerkinbedding toe te passen, en dat deze informatie essentieel is voor het verkrijgen van zelfgelijke netwerkstructuren via geometrische renormalisatie. Ook identificeren we een groot aantal empirische netwerken die beter kunnen worden beschreven in het quasi-geometrische regime. Een verdere verhoging van de koppeling leidt tot een faseovergang, waarvan wij stellen dat deze van topologische aard is. We betreden de geometrische regio, het regime dat typisch is bestudeerd in eerdere werken. Hier zijn de effecten van de onderliggende metrische ruimte sterk en blijft clustering eindig in de thermodynamische limiet. Gemotiveerd door deze resultaten voor éénlaagse grafen bestuderen we ook geometrische multiplexen. We introduceren het wederzijds netwerk, dat bestaat uit de randen die door alle lagen worden gedeeld. Dit netwerk stelt ons in staat om rigoureuze resultaten te verkrijgen over rand- en driehoeksoverlap. We betogen dat het over het algemeen de meer geometrische verbindingen zijn die leiden tot overlap, waardoor de effectieve geometrische koppeling in het wederzijds netwerk toeneemt. Na uitgebreid te hebben stilgestaand bij de structurele eigenschappen bij verschillende koppelingssterkten, richten we ons op dynamische processen in het netwerk. We tonen aan dat de onderliggende metrische ruimte periodieke Turing-patronen onthult, zowel in de quasi- als sterk geometrische regimes, alsook in empirische netwerken. Deze resultaten tonen aan dat de onderliggende geometrie essentieel is voor het begrijpen van complexe netwerken, zowel structureel als dynamisch.

## PUBLICATIONS

This thesis is based on the following publications

- "An anomalous topological phase transition in spatial random graphs", *Communications Physics* **5**, 245 (2022), J. van der Kolk, M.Á. Serrano and M. Boguñá.
- "Emergence of Geometric Turing Patterns in Complex Networks", *Physical Review X* **13**, 021038 (2023), J. van der Kolk, G. García-Pérez, N.E. Kouvaris, M.Á. Serrano and M. Boguñá.
- "Random graphs and real networks with weak geometric coupling", *Phyiscal Review Research* **6**, 013337 (2024), J. van der Kolk, M.Á. Serrano and M. Boguñá.
- "Renormalization of networks with weak geometric coupling", *Physical Review E* **110**, L032302 (2024), J. van der Kolk, M. Boguñá and M.Á. Serrano.
- "Link overlap and mutual clustering in multiplex network", in preparation (2024), J. van der Kolk, D. Krioukov, M. Boguñá, M.Á Serrano.

It has also produced an updated version of the following open source software

- Mercator, a reliable embedding method to map real complex networks into their hyperbolic latent geometry, which can be found at `https://github.com/networkgeometry/mercator`.

# ACKNOWLEDGMENTS

These past four years working on this thesis have been a blessing, and that is in no small part thanks to my wonderful friends, family and colleagues. That is why I would like to dedicate a few words to those that made my PhD possible.

Antes que nada me gustaría darles las gracias a mis directores de tesis, Marián y Mariángeles. Tuve mucha suerte de poder venir a Barcelona para la entrevista del doctorado, justo dos días antes de que se cerraran las fronteras por la pandemia. Muchas gracias por confiar en mí, y darme la oportunidad de venir a trabajar con vosotros en un momento complicado. No podría haber tenido mejores supervisores. Obviamente a nivel de conocimiento —he aprendido más de lo que creía posible— pero también a nivel personal. Me he sentido siempre muy cómodo con vosotros, ya sea en las reuniones y comidas semanales o en las cenas grupales. Espero poder seguir trabajando con vosotros y volver a vernos mucho, ya sea en conferencias o cuando venga de visita a Barcelona.

To my delightful coworkers, you made coming to university an absolute joy. Robert, you can't imagine how happy I was to have someone to share the group with after a year of being the only PhD student. You are an extremely generous person, full of ideas, creativity and energy. Thank you for being my travel buddy in Mallorca, Japan and Quebec, those experiences would definitely not have been the same without you. Laia, com m'alegro que completessis el nostre equip de doctorands. Ets una persona meravellosa, tan forta i organitzada. Espero que vingueu amb el Robert a visitar-me a Viena a repetir la nostra experiència allà. To Lucía and Roya, thank you for the group lunches, and making la Festa de la Ciència a great experience.

My research stay in Boston was one of highlights of my PhD. Dima, thank you for hosting me and for all I learned from you. I especially enjoyed our meeting with Santo philosophizing about the existence of exponential densities of states in far flung fields of Physics. Narayan, your love for science is inspiring, and I enjoyed our discussions, huddled around the tiny white board next to your desk. Harrison, thank you for your insights and making me feel included in the nonrandom project. Brennan, I remember fondly our meeting talking about careers in science. Thank you for putting me in contact with Laszlo, I owe you a lot. Finally, thank you Jesseba, Eline, Erik, Rushali and Sharaj for making my time at the institute as well as our trip to Acadia unforgettable.

A los Inglorious Bravas —Sergio, Javi y Íñigo— gracias por recogerme cuando acababa de llegar a Barcelona, y cuando la uni todavía estaba completamente vacía. Los cafés por la mañana realmente me hicieron bien. Estar escribiendo la tesis más o menos a la vez que vosotros me ha ayudado mucho a acabarla.

Irina, gràcies per organitzar els seminaris d'UBICS amb mi i, juntament amb el Quim i el Sergi, per haver fet de Manso 52 la meva llar aquests últims mesos.

Carol, dar el lab de menánica contigo fue un placer, y te quiero dar las gracias por llevarme a escalar tantas veces para distraerme de la tesis. A todos los demás colegas de la UB, Marc R., Blanca, Carla B., Marc S., Irene, Sergio C., Mattia, Adrià, Josep M., Josep F., Andreu B., Andreu F., Carla R., Álex, gracias por acompañarme estos últimos años. Los voy a echar de menos.

Dan mijn Lekkere Dingen in Nederland: Jochem, Yorick, Luuk, Maik, Vicky, Romee en June. Ik ben heel erg blij dat wij elkaar 10 jaar na het afstuderen van de middelbare en ondanks de afstand die ons scheidt, nog steeds zo regelmatig zien en spreken. Zoveel vrienden hebben die tegelijkertijd met mij promoveren is een absoluut privilege, en ik put zeer veel motivatie uit het zien wat jullie allemaal wel niet bereiken. Jullie zijn super, bedankt dat ik jullie tot mijn vrienden mag rekeken.

Teresa, gracias por compartir esta experiencia en Barcelona conmigo, no habría sido lo mismo sin ti. Gabriel, gracias por las paellas en tu terraza, y por siempre ser tan generoso. Juan, Pablo, gracias por venir a visitarme tantas veces, y por todo lo demás.

Aan mijn familie, mama, papa, Floris, jullie staan altijd voor me klaar. Het klinkt als een cliché, maar het is echt zo: zonder jullie zou ik niet zo ver gekomen zijn.

Finalmente, Alej, gracias por ser la persona más maravillosa del mundo. Tu inteligencia, fuerza y compasión me inspiran cada día. Gracias por aguantarme cuando estaba estresado, gracias por las cenas asiáticas, gracias por las conversaciones infinitas. Te quiero mucho, somos un equipo imparable.

# CONTENTS

# PREFACE

# 1

# INTRODUCTION

## 1.1 MOTIVATION

In his book on the prestigious Santa Fe institute, David Krakauer described the study of complex systems as the study of "Worlds hidden in plain sight" [1]. Modern physics traditionally studies objects that are too small, too distant, too slow, or too fast to be observed directly, and provides us with a plethora of tools for investigating them. Complexity science, however, often focuses on phenomena that we experience, directly or indirectly, in our day to day. Examples include cities, ecosystems, the internet, economies and our own nervous system. Paradoxically, these systems are often the ones we understand the least. Their functioning is, as Krakauer puts it, "hidden [...] through non-linearity, randomness, collective dynamics, hierarchy and emergence".

One of the most powerful ways of studying these disparate phenomena is through the lens of *complex networks* [2, 3]. Here, the agents of a complex system are represented by nodes, with the links connecting them encoding their interactions. This field has increased our understanding of various domains, including ecology [4, 5], neurology [6, 7], the social sciences [8, 9] and technology [10, 11].

Among the most surprising results about complex networks is the apparent universality of their structural properties. Many networks representing completely unrelated systems are, for example, small worlds [12], meaning that they can be traversed in only a small amount of links, even if they are very large. This allows them to be highly navigable, a feature that seems essential for many complex systems [13, 14]. Another seemingly universal feature is the presence of a large amount of triangles in networks representing transitive relations; if a node has two neighbors, it is exceedingly likely those neighbors are also connected to each other [15]. Other properties include sparsity [16], degree heterogeneity [2], the presence of communities [17], etc. All these features have proven essential for understanding the complex systems these networks represent.

The ubiquity of the aforementioned structural properties has led to the development of many models that can reproduce and explain their origins. One framework that has proven especially fruitful is that of *network geometry* [18–20]. Here, nodes are assumed to live in some underlying metric space, where distances can be defined. Sometimes this space is explicit [20], as is the case for airport networks [21, 22], power grids [23–25], or urban networks [26, 27], but oftentimes

it is hidden, or latent, such that the coordinates of the nodes in this space encode their properties in some abstract way [28–31]. In both cases, the distances between the nodes contribute to the connectivity pattern of the network.

Of special importance here is the clustering coefficient, which quantifies the amount of triangles present in a network. Triangles are the shortest possible cycles, thus introducing redundancy in the network connectivity; it is now possible to take several different paths when traveling between two nodes. Furthermore, triangles encode important transitive relationships between the nodes of the system. In the geometric framework, the high value this quantity takes in many real networks is a straightforward consequence of the triangle inequality of the underlying metric space; if a node lies close to its neighbors, those neighbors must also lie close to one another. If metric closeness implies a high probability of being connected, as is assumed in this framework, this immediately leads to a triangle forming in the network. This fact makes clustering the quintessential geometric network property.

One particularly successful class of geometric models are those that explain the observed network structure through the interplay between the similarity and popularity dimensions, where it is assumed that more similar and more popular nodes are more likely to be connected. This framework provides a simple explanation to many of the observed topological properties of real networks, including heterogeneous degree distributions [29, 30, 32], clustering [29, 32–35], small-worldness [36–38], as well as percolation [39, 40], spectral [41] and self-similarity properties [30, 42, 43]. These models have been extended to growing networks [44], weighted networks [45], directed networks [46], multilayer networks [47, 48] and networks with community structure [49–51]. Finally, the framework has produced strong results in practical tasks such as community detection [52, 53], information routing [54–56] and link prediction [57, 58]

## 1.2 AIM

From the works mentioned above, it becomes clear that geometry plays an essential role in shaping the properties of complex networks. However, to what extend the latent space influences the topology differs from network to network. In the past, it has often been assumed that there are two classes of networks; those that are well described in the geometric framework and those that are not. Generally, the distinction between these two classes is made on the basis of some measure of the amount of triangles in the networks [59, 60]. However, these studies do not contemplate the fact that the transition between non-geometric and geometric networks might not be sharp. In this thesis, we investigate how different levels of coupling between the geometry and the topology of a networks affect its properties.

We investigate this question using the $\mathbb{S}^1/\mathbb{H}^2$-model [29, 30]. This similarity $\times$ popularity model is maximally random, in the sense that no unwarranted assumptions about the networks it produces are made. They also contain a model parameter $\beta$ that can be used to tune the coupling strength between the network

topology and its underlying metric space, making them the ideal lens with which to study our central question.

We approach our main research question from various perspectives; we investigate the effect of geometry on the structural properties of networks, specifically the level of clustering. We ask how different geometric couplings influence the the possibility of embedding and renormalizing both artificial as well as empirical networks. We extend our structural investigations to multiplex networks, asking what influence geometry has on edges that are simultaneously present in multiple layers. Even though the main focus of this thesis is on the structural properties of the network, we are also interested in how geometry shapes the dynamical processes running on top of the network.

## 1.3 THESIS STRUCTURE

This thesis consists of three parts, which are then further divided into chapters.

**Theoretical background:** The first part gives an introduction to the fields of complex networks and network geometry.

In Chapter 2 we introduce the complex network as a mathematical object and give an overview of some of its most important properties. In Chapter 3 we provide a selection of random graph models that explain these properties. We especially focus on maximal entropy models as these contain the geometric models we employ in this thesis. Network geometry is then introduced in Chapter 4. We give a short historic overview of this field after which we define the $\mathbb{S}^1$ and $\mathbb{H}^2$ models, which form the basis of this thesis. We summarize some of the other models that exist in this field and finally introduce the concept of network embedding, where the hidden metric space of a network can be revealed through various computational methods.

**Structural properties:** This part constitutes the bulk of the thesis and is based on various publications [35, 61, 62] as well as some yet unpublished work [63].

In Chapter 5 we investigate the phase transition that exists in the $\mathbb{S}^1/\mathbb{H}^2$-model between a region of finite clustering when the geometric coupling is strong to a region where this quantity vanishes in the thermodynamic limit when the coupling is weak [30]. We map the model to a gas of non-interacting fermions in order to show that the transition exhibits novel thermodynamic properties, such as a diverging entropy density. We argue that the transition is topological in nature, driven by a reorganization of the chordless cycles. Finally, we study the finite size scaling behavior of the clustering coefficient and show that it decreases extremely slowly with the network size in the critical and sub-critical regions, otherwise known as the weak coupling regime. We show that this region can be divided in a non-geometric regime, where clustering decays in the same way as when no geometry is present, and a quasi-geometric regime, where the clustering coefficient decays so slowly that it remains significant for most realistic system sizes. These results on the clustering coefficient lead us to wonder what role the latent space

plays in the weakly geometric regime. In Chapter 6 we approach this question from the perspective of network embeddings. We extend Mercator [64], a computational tool that combines machine learning and model based techniques to recover the hidden coordinates of a network from its topology alone, to the weakly geometric regime. By studying artificial networks generated with the $\mathbb{S}^1/\mathbb{H}^2$-model, we show that geometric information can oftentimes indeed be extracted, even though the coupling is weak. Interestingly, we find that the region for which this is possible coincides with the quasi-geometric regime identified above. Finally, we show that even when geometry is absent, the coordinates provided by Mercator can still be used for practical tasks such as greedy routing.

In Chapter 7 we show that geometric information is important when studying networks across scales. By extending the geometric renormalization framework [42] to the weakly geometric regime, we show that, in the quasi-geometric region, self-similarity in several important structural network properties can only be achieved when geometric information is taken into account.

The results in Chapters 5-7 are theoretical in nature, in the sense that they are based on artificial networks generated with the model. In Chapter 8 we show that the results carry over to emperical networks as well. We argue that many real networks are best described in the weakly-geometric regime, and show that, here, Mercator is able to provide faithful network embeddings. In this regime self-similar network replicas can also be obtained using the extended geometric renormalization technique.

In Chapter 9 we extend our investigation of weakly geometric graphs to multiplex networks. These are networks where different types of interactions exist between the nodes. These interactions are then modeled as existing in different layers of the multiplex. We define the mutual network, which is constructed from the edges that exist in all layers simultaneously. Studying its properties we are able to prove rigorous results on the effects of both heterogeneity as well as geometry on the scaling behavior of edge and triangle overlap. Analyzing several empirical multiplexes we find that they generally sport low levels of edge overlap and high levels of mutual clustering, results that are in line with our theoretical predictions.

**Dynamical processes:** In this part of the thesis, which is based on Ref. [65], we investigate how geometry influences dynamical processes on complex networks, focusing on diffusion-driven Turing patterns.

In Chapter 10 we study spontaneous pattern formation within the Turing framework, and show that the underlying metric space can reveal periodic patterns in real and artificial networks that remain hidden when approaching the network from a purely topologic point of view. We show that by approximating the network by its annealed counterpart, where links are resampled at a high rate, the frequencies of the observed patterns can be predicted from the structural and dynamical parameters alone. Finally, we show that several real networks are capable of sustaining patterns.

## 1.4 NOTATION

Unless otherwise stated, we use the following notational conventions:

- If $f(N) \simeq g(N)$ then $\lim_{N \to 0} \frac{f(N)}{g(N)} = 1$. In words, this means they are asymptotically equivalent.
- If $f(N) \sim g(N)$ then $\exists k_1 > 0$, $\exists k_2 > 0$ and $\exists N_0 > 0$ such that $\forall N > N_0$, $k_1 f(N) \leq g(N) \leq k_2 f(N)$. In words, this means the two functions asymptotically have the same scaling behavior with respect to $N$.
- If $f(N) \propto g(N)$ then $\exists k \neq 0$ such that $f(N) = kg(N)$. In words, this means they are proportional to one another.
- Matrices are denoted with capitalized, boldface roman letters, e.g., $\boldsymbol{A}$.
- Vectors are denoted with lower-case, boldface roman letters, e.g., $\boldsymbol{v}$.

# Part I

# THEORETICAL BACKGROUND

# 2

# NETWORKS

At the most basic level and in its simplest form, a network is a collection of nodes connected by links. These nodes represent the constituents of a system: Individuals in social networks, organisms in trophic networks, countries in trade networks, etc. The links in turn represent the interactions between these constituents. Once again, these interactions can take on a wide range of forms depending on the system in question. For example, in social networks they might represent friendships, whereas in trophic networks they may encode the predator-prey relations. In this Chapter we will introduce the concepts necessary to describe these systems and introduce some of the most important properties shared by many real world networks, irrespective of what systems they represent.

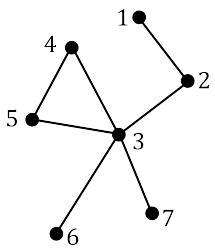
## 2.1 THE BASICS

Mathematically, networks are often denoted as $G(V, E)$, where $V$ is the set of *nodes* (or *vertices*) and $E$ the set of *links* (or *edges*). The nodes are generally labeled by integers $V = \{1, 2, ..., N\}$, where $N \equiv |V|$ is the amount of nodes. The links in turn can be labeled by their endpoints such that $l = \{i, j\} \in E$, with $i, j \in V$. An example of such a network can be seen in Fig. 2.1. We see that in this specific example we have seven nodes, $N = 7$, and seven links, $M \equiv |E| = 7$.

Of course, there are many ways to organize seven links among seven nodes, and we need a way to encode the specific structure of a network. Throughout this thesis, this structure will be referred to as the *topology* or *connectivity* of the network. An efficient way to encode the structure is through the *adjacency matrix* $\boldsymbol{A}$. This is an $N \times N$ matrix, where the row and column indices are both related to a specific node. The entries therefore represent pairs of nodes, or potential links. We then define the adjacency matrix as

$$A_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{else.} \end{cases} \tag{2.1}$$

Note that this matrix is symmetric, and that its diagonal elements are zero as we do not allow for self-loops, i.e. links that connect a node to itself. We also do not allow for multiple edges between the same pair of nodes. Such networks are called *simple*. In the case of our example, the adjacency matrix is given in Fig. 2.1.




$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \boldsymbol{k} = \begin{pmatrix} 1 \\ 2 \\ 5 \\ 2 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

Figure 2.1: An example of a simple graph with 7 nodes and 7 edges. The associated adjacency matrix and degree sequence are also shown.

An important property of a node $i$ is its *degree* $k_i$, which counts the amount of neighbors it has. This quantity can be obtained by summing the rows (or columns) of the adjacency matrix

$$k_i = \sum_{j=1}^{N} A_{ij}. \tag{2.2}$$

For our specific example, $\boldsymbol{k}$ is given in Fig. 2.1.

The above definitions are for the most definition of a network. However, many generalizations exist. One could for example study systems where the interactions between agents are not symmetric. In that case we obtain a *directed* network, where it could happen that a link $l_1 = \{i, j\} \in E$ but where $l_2 = \{j, i\} \notin E$. An example of such a network could be a citation network, where author $i$ cites author $j$ but not the other way around. This then leads to a non-symmetric adjacency matrix and two types of degrees: If one sums over the columns one obtains the *in-degree* $k_{in,i}$ of a node $i$, whereas summing over the rows leads to the *out-degree* $k_{out,i}$. In the example of the citation network, the in degree would encode how many people cite author $i$ and the out degree how many others are cited by this author.

In a *weighted* network, the interaction strength between two constituents is taken into account. For example, in a trade network one might also wish to encode the trade volume between two states. In this case the entries of adjacency cease to be binary, but are rather given by integer or real numbers, depending on whether the interaction strength is measured in terms of countable or uncountable units. In this case we also need to generalize the idea of the degree of a node to its *strength* $s_i$, which is the sum over the weights of the links attached to node $i$.

There might also be different types of interactions between nodes. In this case we can represent the system in terms of a *multiplex*. Here, the adjacency matrix is promoted to tensor with entries $A_{ij}^{(l)}$, where $l$ encodes the type of interaction between nodes $i$ and $j$. One can visualize this as a network with multiple layers, where each layer represents a type of interaction, but where nodes are shared between the layers. Note that here we can generalize even further to where the

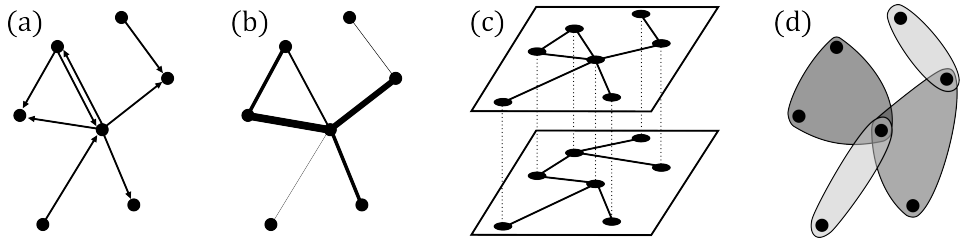



Figure 2.2: Examples of an (a) directed network, (b) weighted network, (c) multiplex network and (d) hypergraph.

constituents in the different layers are also different; this then leads to *interconnected* networks.

The final type of generalization we will mention here takes into account that interactions between agents might not be pairwise. When modeling a biomolecular network, three or more proteins can interact together to produce a certain effect in a cell. When one of these proteins is missing, the process will not occur. It would therefore not be realistic to model this biomolecular network as a set of pairwise interactions between the proteins. Two ways of modeling such higher-order interactions is through *simplicial complexes* and *hypergraphs*.

Of course, these generalizations can be combined: Some hypergraphs might be directed, or one could think of applications where a weighted multiplex might be the proper representation of the system. In this thesis we will mostly focus on simple networks, briefly expanding our scope to include multiplexes in Chapter 9. In Fig. 2.2 we display the directed, weighted, multiplex and hypergraph versions of the example studied in the previous section.

## 2.2 NETWORK PROPERTIES

As we have seen in the introduction, networks are ubiquitous and can be used to represents all kinds of real world systems. One might expect networks from completely distinct fields of study to be vastly different. However, one of the most fascinating results of network science tells us that this is not the case. There are several properties that most real world networks share, independent of which real world system they represent. These properties can also be used as a definition of a *complex* network[1]: A network is complex if it exhibits one or several of the properties described in the following.

1 This is of course not the only possible definition of this rather vague term.

### 2.2.1 *Graph density*

The *graph density* encodes how strongly a network is connected. For simple graphs it is defined as

$$\rho = \frac{M}{\binom{N}{2}} = \frac{2M}{N(N-1)}, \tag{2.3}$$

i.e., the amount of links present in the graph divided by the total possible amount of links. Networks can be divided into two types depending on how this quantity scales with the system size $\rho \sim N^{\alpha}$. When $\alpha = -1$, the network is called *sparse*. If $\alpha > -1$, the network is called *dense*[2].

Another way of encoding how strongly a networks is connected is through the *average degree*

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij} = \frac{2M}{N}. \tag{2.4}$$

Sparse networks, like the ones studied in this thesis, have average degrees that do not depend on the system size: $\langle k \rangle \sim N^0$. Of course, for many real networks the scaling with respect to the system size is not an accessible quantity. Therefore, when talking about sparsity in the context of real systems we mostly mean that $\langle k \rangle \ll N$. This is a feature many real networks exhibit.

An important question to answer when studying a network is how many *connected components* it has. When the bulk of the network is contained in a single connected entity, this component is called the *giant connected component*, or GCC. Whether or not a GCC exists depends on the amount of links in the network, and is studied in the field of *network percolation*. In this thesis we will only work with networks where a GCC is present, unless otherwise specified.

### 2.2.2 *Degree heterogeneity*

The *degree distribution* $P(k)$ gives the probability of a randomly chosen node $i$ having degree $k$, and can thus be estimated by the fraction of nodes that have degree $k$

$$P(k) = \frac{N_k}{N}, \tag{2.5}$$

where $N_k = |\{i \in V \,|k_i = k\}|$. Another, related, quantity is the *complementary cumulative degree distribution*, also knows as the *tail distribution*, given by

$$P_c(k) = \sum_{q=k+1}^{N-1} P(q). \tag{2.6}$$

How broad the degree distribution is captures the degree heterogeneity of a graph. At one extreme we have the *regular graphs*, where every node has the same degree. For a $k$-regular graph, the degree distribution is then $P(q) = \delta_{kq}$. Adding a bit

2 Note that the definition of sparse and dense is different in the math literature. There, $\alpha = -1$ is called *ultra-sparse*, $-1 < \alpha < 0$ are called sparse and only when $\alpha = 0$ is a graph dense.

more spread to the degrees might lead to a Poisson distribution $P(k) = \frac{1}{k!}\langle k\rangle^k e^{-\langle k\rangle}$. The variance of such a distribution is $\langle k\rangle$, which, if we assume a sparse graph, is small. This implies that such networks are still very *homogeneous*. It has been shown, though, that a common property of a wide range of real networks is that they are *heterogeneous*, i.e. that their degree distribution has a large variance. These networks are also characterized by the presence of *hubs*, nodes with a degree much larger than the average.

One way of modeling such heterogeneity is by assuming that the degree distribution can be described well by a power law[3]

$$P(k) = \zeta k^{-\gamma}, \tag{2.7}$$

where $\zeta$ is some constant [16, 66–68]. The parameter $\gamma$ governs how fast the distribution falls as one increases $k$. Note that this distribution is the only normalizable distribution that is invariant under rescaling. It does not matter at which scale we observe the distribution, its shape is always the same. In mathematical notation this can be summarized as $P(bk) = b^{-\gamma}P(k)$. This distribution, and networks whose degrees follow it, are therefore often called *scale-free* [69].

We can define three different regimes depending on the characteristic exponent $\gamma$. First, when $\gamma \leq 2$ the average degree $\langle k\rangle$ diverges as $N$ goes to infinity, which is not in line with our sparsity assumption. Second, when $\gamma > 3$ the distribution falls off very fast. In real data it is therefore very difficult to distinguish such degree distributions from more homogeneous distributions like the Poisson [68]. Finally, we have the region $2 < \gamma \leq 3$. Here, the average degree remains finite as $N \to \infty$, but the second moment, and thus the variance of the distribution, does not. We thus obtain a sparse graph where the fluctuations around the mean are large. This is another, more imprecise definition of scale-free: If one picks a node at random one has no way of knowing what the scale of its degree will be [68]. It might be that the node is connected to only a handful of neighbors or to all other nodes in the network.

It is good to take a moment and discuss some subtleties. Because, what does it mean for a "distribution to be described well by a power law"? If we take a real network and extract its degree distribution, how are we supposed to know if it is actually scale-free? One might argue that Eq. (2.7) only has to hold for the tail of the distribution $k > k_l$ for some $k_l > k_{\min}$ [70]. Or maybe we can expect finite size effects to introduce a cut-off in the distribution such that is should only hold in some region $k_l < k < k_u$ for some $k_l > k_{\min}$ and $k_u < k_{\max}$ [71, 72]. Even if we clear up these concerns, it is still not always clear if the power law fits the data well [73]. It might fit well in comparison to something like a Poisson, but maybe there are other distribution that would fit equally well or even better [74].

The lack of precision in the definition of scale-free networks has led to contradictory results on their ubiquity. In their highly cited work [74], Broido and Closet show that the log-normal distribution fit a wide range of real networks better than the power law, and therefore conclude that "scale-free networks are rare". However,

3 Such a distribution is also called a Pareto distribution.

Voitalov *et al.* [75] argue that trying to fit an exact power law to data is unrealistic due to the presence of noise and fluctuations. Even network models which are designed to produce pure power laws fail to do so [76]. The authors therefore argue that one should focus on the extreme behavior of the distribution, i.e. on the behavior of the largest degrees, as these are solely responsible for a wide range of structural and dynamical properties of the network. They extend the definition of scale-freeness to include all networks whose degree distribution is regularly varying, i.e. whose probability distribution function is given by $P(k) = l(k)k^{-\gamma}$, where $l(k)$ is slowly varying[4] such that the tail of the distribution is a power law [78, 79]. They show that many real networks fall into this class and conclude that scale-free networks are widespread.

They base this claim on extreme value theory [80]. Here, the idea is to draw $N$ random numbers i.i.d. from some distribution. The maximum value in this sample is then again a random variable, and extreme value theory tries to characterize its distribution, which is called the *extreme value distribution*. It can be shown that all regularly varying distributions lead to extreme value distribution that fall into the same family, characterized by the *extreme value index* $\xi > 0$ [81]. This index is related to the exponent of the power law tail as $\xi = 1/(\gamma - 1)$. The expected largest value in a sample of size $N$ is then given by

$$k_c \propto N^{\xi}, \tag{2.8}$$

which we will refer to as the *natural cut-off* [82]. Note that all this relies on several assumptions about the degree distribution, most notably that the degrees are drawn i.i.d. However, this is generally not the case as node degrees must also comply with constraints set by the network structure. For example, a node of course has a maximum degree of $k_{max} = N - 1$, i.e. it cannot be connected to more nodes than exist in the network. We will see examples of other degree correlations in the next section. All in all, the topic of scale-freeness in networks still has many unanswered questions.

To avoid entering in this discussion, in this thesis we will refer to real networks as either heterogeneous of homogeneous, depending on whether their degree distribution is relatively broad or not. We will not claim anything about their scale-freeness, nor try and fit their degree distribution.

### 2.2.3 *Degree correlations*

In the previous section we discussed the distribution of the degrees of individual nodes. Of course, these nodes are connected through links, and it is interesting to ask how the degrees of the nodes at the ends of these links are correlated. Are hubs generally connected to each other? Or do they avoid one another and mostly connect to low degree nodes?

4 Mathematically, a function $l$ is slowly varying if, $\forall a > 0 \ \frac{l(ak)}{l(k)} = 1$. Examples include functions that converge to a constant or to (an arbitrary power of) a logarithm. [77]

We can answer these questions by studying the *degree correlation function* [83, 84], defined as

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k), \tag{2.9}$$

where $P(k'|k)$ gives the probability that a neighbor of a node with degree $k$ has degree $k'$. The degree correlation function therefore represents the average degree of the neighbors of a node with degree $k$.

On the basis of the degree correlation function we can define three types of networks [15, 69]: 1) In *assortative* networks, $k_{nn}(k)$ is an increasing function of $k$; low degree nodes generally have neighbors which also have small degrees and hubs mostly have neighbors which are also hubs. 2) In *neutral* networks, the degree correlation function is given by a constant; no degree correlations are present and the average degree of the neighbors of a node is independent of its degree. In this case $P(k'|k)$ is just the probability that a randomly chosen node connects to a node with degree $k'$. The probability of connecting to a specific node $i$ with degree $k'$ is $k'/(N\langle k\rangle)$. To understand this, let us cut all edges in half so we end up with $2M = N\langle k\rangle$ ends of edges, which we call *stubs*. Then, $k'$ of these stubs belong to node $i$, leading to our result. Now, there are $N_{k'} = NP(k')$ such nodes with degree $k'$, so the final result is

$$P_{neutral}(k'|k) = \frac{k'P(k')}{\langle k\rangle}, \tag{2.10}$$

which leads to

$$\bar{k}_{nn,neutral}(k) = \frac{1}{\langle k\rangle}\sum_{k'} k'^2 P(k') = \frac{\langle k^2\rangle}{\langle k\rangle}, \tag{2.11}$$

which in independent of $k$ [68]. 3) In *disassortative* networks, the degree correlation function decreases with $k$. This implies that hubs prefer to connect to low degree nodes and vice versa.

In scale-free networks, something surprising occurs when attempting to create neutral networks [82, 85]. Let us study the expected number of connections between two nodes $i$ and $j$ with degrees $k_i$ and $k_j$ respectively. If no degree correlations are present, the expected amount of links between these two nodes will be

$$\epsilon_{ij} = \frac{k_i k_j}{N\langle k\rangle}. \tag{2.12}$$

This can be understood as follows. Draw two stubs at random; the probability that one belongs to $i$ and the other to $j$ is $2k_ik_j/(2M)^2$, where the factor two comes from the fact that we are working with an undirected network; we now repeat this $M$ times to connect all stubs, leading to the expectation shown in Eq. (2.12). Of course, we are working with simple graphs, so this quantity should never exceed 1. However, we saw in the previous section that for a scale-free network with exponent $\gamma$, the expected maximum degree scales is $N^{1/(\gamma-1)}$. If we assume the second largest degree to scale similarly, the expected amount of links between these two nodes scales as $\epsilon_{\max} \sim N^{(3-\gamma)/(\gamma-1)}$, which is larger than 1 for $\gamma \leq 3$. Enforcing that $\epsilon_{ij} \leq 1$ for all $i$ and $j$ means that hubs cannot be connected to

other hubs, at least not as often as one would expect for a neutral network. This introduces *structural correlations* into the network, particularly causing it to become disassortative. One can avoid these correlations by introducing a *structural cut-off* $k_{\max} \sim \sqrt{N}$ such that $\epsilon_{max} \sim N^0$.

### 2.2.4 *Clustering*

Real networks often contain a large amount of triangles. In a social networks, for example, this can be understood with the conventional wisdom "the friend of my friend is my friend". The friend of my friend is very likely to live in a similar social milieu as me and we are thus very likely to also be friends, thus creating a triangle in our social network. How strong this effect is is quantified by the *clustering coefficient*, which can be defined both locally and globally. The local clustering coefficient is given for each node separately. For node $i$, it can be defined as

$$c_i = \frac{(\text{number of pairs of neighbors of } i \text{ that are connected})}{(\text{number of pairs of neighbors of } i)} \tag{2.13}$$

Note that the numerator is equivalent to the amount of triangles attached to node $i$, and that the denominator can be replaced by the maximum possible amount of triangles attached to $i$, which is given by $k_i(k_i - 1)/2$. Of course, this latter quantity is zero for nodes with degree zero or one. For this reason we generally only take nodes with degree strictly larger than one into account when calculating this quantity.

In the example in Fig. 2.1, we see that there is one triangle connected to the central node, which has five neighbors in total. Thus, the local clustering coefficient of that node is $c_i = 2/(5(5-1)) = 1/10$.

An important (global) quantity that we will use throughout the thesis is the *average local clustering coefficient*, which is given by

$$\overline{c} = \frac{1}{N_{k>1}} \sum_{i=1}^{N_{k>1}} c_i. \tag{2.14}$$

Note that here we normalize by $N_{k>1} \equiv |i \in V | k_i > 1|$ as we can only define the local clustering coefficient for these nodes.

One could also define a global clustering coefficient directly using the concept of triplets. A triplet is a set of three nodes connected by either two or three links. This measure is also called the *transitivity* and can be defined as

$$C = \frac{3 \times (\text{number of triangles})}{(\text{number of connected triplets})}. \tag{2.15}$$

The fact that there are various measures describing the amount of triangles in a networks sometimes leads to confusion in the literature. To avoid such confusion in this thesis we will only work with Eq. (2.14), unless otherwise stated.

### 2.2.5 *Communities*

We have already seen several examples of how to the structure of complex networks. We now want to ask if structure can also be found in the relation between groups of nodes. One way of answering this questions is through the concept of *network communities* [86, 87]. A community can (loosely) be defined as follows:

> A community is a set of nodes with many internal links and few external ones.

This is of course a rather vague definition as it is not clear what exactly is meant by "many" or "few". This is a problem because when finding communities in networks, a process known as *community detection*, it helps to know what exactly one is looking for. One of the ways of addressing this ambiguity is by considering that the amount of internal and external links should be compared to the expected amount of links if connections were made at random. This is the idea behind community detection algorithms based on *modularity* [88].

Say we divide our network into $N_C$ communities $\{V_i\}_{i=1}^{N_C}$ such that $\cup_{i=1}^{N_C} V_i = V$ and $V_i \cap V_j = \emptyset$ for $i \neq j$. Put simply, all nodes should be contained in exactly one of $N_C$ communities. For a given node $i$ we define $\sigma(i) \in \{V_j\}_{j=1}^{N_C}$ as a classifier which tells us in which community it is located. We can then define the modularity as follows

$$Q = \frac{1}{M} \sum_{i<j} \left( A_{ij} - \frac{k_i k_j}{2M} \right) \delta(\sigma(i), \sigma(j)), \tag{2.16}$$

where $\delta(\sigma(i), \sigma(j)) = 1$ if $\sigma(i) = \sigma(j)$ and zero otherwise. How should we interpret this? Say we take two nodes $i$ and $j$ that lie in the same community. The first term in the summand checks whether or not a links is present between them. We then want to compare this with the expected amount of links between node $i$ and $j$ if edges are placed randomly conditioned on their degrees. We studied this quantity in the previous section and found that it was given by Eq. (2.12), exactly the second term in the summand in Eq. (2.16).

The modularity thus represents the difference between the actual and expected number of links within communities in the network, normalized w.r.t. the total number of links. If modularity is positive, there are more internal links than we would expect and if it is negative there are less. Of course, there being more (less) internal links than expected immediately implies that there are also less (more) external ones. A high modularity therefore implies that the communities $\{V_i\}_{i=1}^{N_C}$ comply with our definition for them.

For community detection, a good strategy might then be to find the partition of the network that maximizes the modularity. This is, however, not a trivial endeavor as 1) one does not usually know how many communities are contained in the network and 2) there are many different ways $N$ nodes can be distributed over $N_C$ communities [89]. Checking all possible distributions is, therefore, mostly not feasible. Many other optimization methods exist that try and find the optimal

partition without going through all the possibilities. Possibly the most widely used of these methods is the Louvain algorithm [90].

Even though modularity methods can be very useful, there are some settings in which they are less suitable [91]. It has been shown that these methods overfit, in the sense that they find communities even in random data [92]. They also have a resolution limit, meaning that they cannot observe communities below a certain scale [93]. Therefore, it is important to not directly give *explicative* meaning to the communities found when maximizing modularity. When looking at real data, the communities found might not really 'exist', in the sense that they are not representative of any meaningful group. In a friendship network, for example, these methods might not give you actual friend groups. Other methods, mostly based on statistical inference and generative models, are more suitable for problems where we want to explain real data [94]. Modularity maximization is *descriptive*, meaning that it finds a partition with communities that comply with our initial definition: They have many internal and few external links. For the purposes of this thesis, this is sufficient; it tells us something about the structure of the observed network. We do not claim that the communities found necessarily correspond to any real groups in the data.

### 2.2.6 *Small world*

The *small world property* is best illustrated intuitively using Milgram's small world experiment, which originated the now famous idea of six degrees of separation [95, 96]. Milgram asked several people in Nebraska and Kansas to send a package to a specific person who lived in Boston. The catch was, however, that they were not allowed to send it directly to the recipient. Instead, they were asked to send the package to someone in their direct social circle that they thought might have a better chance of knowing the recipient personally. This friend would then do the same, until finally the package arrived at someone who was on a first name basis with the target and could thus forward it to them. Surprisingly, even though most packages never reached their destination, those which did arrived in only a small amount of steps. In the specific case of Milgram's experiment, the average path length was six, hence "six degrees of separation".

Of course, this specific number needs to be taken with a grain of salt [97]. However, the general idea holds: Even though the system under study is vast (in Milgram's case it theoretically encompassed the entire US population), the typical distance between agents is small [98, 99].

Let us now formalize this idea. First, we define a *walk* as a sequence of alternating nodes and edges, starting and ending with a node, such that each node is connected to the subsequent edge and each edge to the subsequent node. The *length* of a walk is the amount of edges in the sequence and a *path* is a special type of walk where each edge can only be used once. Finally, the *geodesic topological distance* $d_{ij}$ is the length of the shortest path between nodes $i$ and $j$. A network is said to exhibit the small world property when the average geodesic distance between any pair of nodes $\langle d \rangle = \frac{2}{N(N-1)} \sum_{i<j} d_{ij}$ scales with the system size slower than any

power law. When the scaling is extremely slow and $\langle d \rangle \sim \ln \ln N$, the network is said to be ultra-small world. Note that, as in the case of sparsity, this definition is not generally applicable to real networks. Once again for these kind of networks small worldness generally implies that $\langle d \rangle \ll N$.

### 2.2.7 *Other properties*

The above list of network properties is in no way exhaustive. There are, for example, many different *centrality* measures [100, 101], which quantify the importance of the nodes within the network, in turn defining a *hierarchical* ordering. One can also define a nested hierarchy of nodes, where higher level elements consist of and contain lower level elements [102–104]. An example of such a hierarchy is achieved through the *k-core* decomposition [105, 106], where nodes are removed according to their degree until different nested sets of nodes are obtained. Another type of hierarchy is based on flows. This type is mostly relevant for directed networks, where lower level nodes are influenced by higher order ones [107, 108].

One can also study networks from the point of view of motifs [109, 110]. These are small subgraphs, such as fully connected graphs of a few nodes. One can then study their frequency within the larger graph, where it is often found they occur far more often than in randomized networks. Different classes of networks contain different types of motifs, indicating that they are important in defining the network function.

# 3

# RANDOM GRAPHS

In the previous Chapter, we saw that many real networks share important features. To understand why this happens, many generative network models have been proposed. One class of such models is given by *random graphs*. Here, the connections between the nodes are made at random in such a way that specific network properties are reproduced. Studying these network models has many advantages. Firstly, random graphs allow us to tune all kinds of network properties. This flexibility allows us to study the effect of these properties on the networks structure or on dynamical processes on top of the network. Secondly, where for real networks we generally only have a single instance, random graphs allow us to create large ensembles of graphs with similar properties. One can then contrast real data with these emsembles; how likely is it that our observed network was obtained with this random process. The ease with which artificial data can be generated also allows for better statistics than when working with finite, often incomplete, real data. Finally, random graphs allow us to create networks of different sized and therefore to study size dependent effects.

For these reasons, in this thesis we will mostly work with random graphs, although we will make a point of always comparing our results with real data. We will introduce the specific random graph model on which this thesis is based, the *geometric* $\mathbb{S}^1$ *model* in Chapter 4. In this section we will give an overview of some other models that we deem relevant, either because they are of historical importance or because they will be referenced in the thesis. However, it is important to note that the list of models described here is in no way exhaustive. For a more complete overview we refer the reader to Refs. [16, 111–114].

## 3.1 ERDÖS RENYI

Arguably the simplest, and best-known random graph model is the Erdös-Renyi (ER) model, named after the two mathematicians who published a celebrated series of papers on it [115]. It allows us to fix the graph density or, equivalently, the average degree. There are two versions of this model: If we want to fix the density exactly, we must turn to the $G(N, M)$ model, where we sample uniformly from the set of all graph with exactly $N$ nodes and $M$ links. The ensemble generated by this procedure is, then, *microcanonical* [116]. We can sample networks from this

ensemble by distributing $M$ links uniformly over the $\binom{N}{2}$ pairs of nodes in the network.

We can also relax the constraint somewhat and only fix the expected density. In this case we define the $G(N,p)$ model [117], where we go through all $\binom{N}{2}$ pairs of nodes and create a link between them with probability $p$. The expected density of such a network is then given by $\langle \rho \rangle_{ER} = p$. As we are fixing only the expected density, we are working in the *canonical*-ensemble[1]. Both models have very similar properties, but the latter is easier to work with and we therefore refer to it when discussing the ER model. Note that is was actually Gilbert who first introduced the $G(N,p)$ model [117]. However, for historical reasons, it is Erdös and Renyi whose names are mostly linked to both models.

One of the real world features present in the ER model is the small world property; it can be shown that $\langle d \rangle_{ER} \sim \log N$ [118, 119]. However, even though the ER model is the quintessential random graph model, it also lacks several key properties of real networks. First of all, is does in general not lead to a clustered network as $\bar{c}_{ER} \sim N^{-1}$ vanishes in the thermodynamic limit [16]. Furthermore, the degree distribution is given by a binomial distribution, which in the large $N$, sparse limit takes on a Poisson form. It is, thus, very homogeneous [112].

## 3.2 WATTS STROGATZ

Our next objective is a model which leads to a substantial amount of clustering, i.e. where the density of triangles is high. A very simple candidate is the regular ring lattice (RRL) shown in Fig. 3.1a. It can be shown that for this graph we have

$$\bar{c}_{RRL} = \frac{3}{4}\frac{(\langle k \rangle - 2)}{(\langle k \rangle - 1)}, \tag{3.1}$$

which is independent of the amount of nodes [120]. This fact can be understood by noting that connections here are local; you can make the circle as large as you like, the local neighborhood of a node stays the same. This locality is also a problem, however, as it causes the graph to be large world. When moving from node $i$ to $j$, you necessarily have to follow the circle, and the largest steps you can take are of size $\langle k \rangle / 2$. Thus, if $i$ and $j$ are $l$ lattice spacings apart, the shortest path will have $2l/\langle k \rangle$ hops. Averaging this over all pairs of nodes we obtain that

$$\langle d \rangle_{RRL} = \frac{N}{2\langle k \rangle}, \tag{3.2}$$

1 In this thesis, naming ensemble classes is not as straightforward as in standard statistical physics. There, canonical ensembles have a fixed particle number but variable energy, while grand canonical ensembles allow both to fluctuate. But in the case of the ER ensemble there is no energy, and only the number of links is free to fluctuate. Is this then a canonical ensemble because there is only one constraint that is fixed up to expectation? Or do we have a grand canonical ensemble because the amount of links is reminiscent of the particle number? To avoid addressing these questions for each encountered model, we refer to all ensembles where the constraints are fixed in expectation as "canonical", regardless of the type and number of these constraints. The only exception is the $S^1$-model introduced in Chapter 4.2, where both an energy and particle number can be defined, making the mapping to the grand canonical ensemble one-to-one.

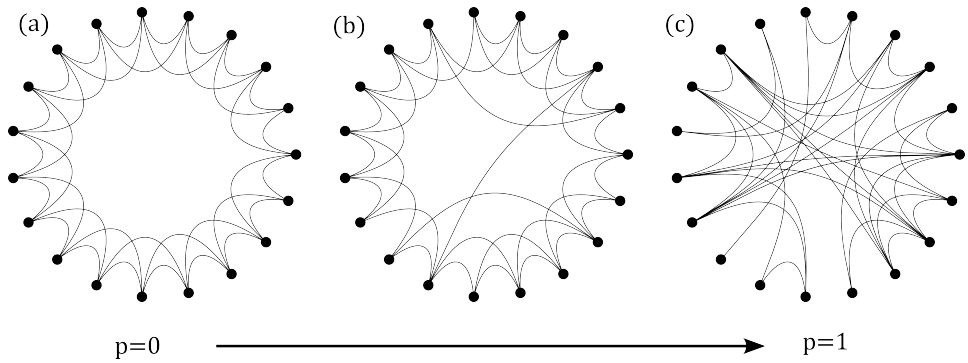



Figure 3.1: Realizations of the WS for three different values of the rewiring probability $p$.

which scales linearly with the system size, implying a large world network [16].

We now have two types of networks: The RRL with its high levels of clustering and the ER model with its small worldness. A third model, the *Watts-Strogatz* (WS) model, combines these two by interpolating between them [121]. We start with the RRL, and rewire each of the $M$ edges with probability $p$. The rewiring is done by removing the original link and placing a new one between two uniformly chosen nodes. Note that, when $p$ is small, an alternative way to implement the model is to add a set of random links to the RRL. This alternative is more analytically tractable and therefore often preferable.

It is clear that the model produces the original RRL when $p = 0$ and leads to a realization of the ER model when $p = 1$. The surprising result here is that for a relatively large range of small nonzero $p$, the model produces networks that both have high levels of clustering and short mean geodesic distances. This result is visualized in Fig. 3.1. Even though exact analytical results for this model are hard to obtain, it can be shown that, for small values of the rewiring probability, the model leads to networks with $\langle d \rangle_{WS} \sim \log N$ and $\bar{c} \sim N^0$, i.e., to networks that are both small world as clustered [121, 122].

We can understand why this happens by thinking again about locality. When $p$ is zero, all connections are local. However, when the rewiring probability gets tuned away from zero, a few short-cuts are created, linking distant parts of the network. This allows one to cover large distances in a relatively small amount of hops, leading to small world behavior. It can be shown that only a few shortcuts are necessary to obtain this effect. The amount of rewirings is then also small, such that locally the network still looks like the RRL, leading to high levels of clustering. This way of thinking will come in handy when studying geometric network models.

## 3.3 THE CONFIGURATION MODEL

We now introduce a set of related models which generalize the ER graph models to allow for heterogeneous degree distributions. These models will turn out to be a limiting case of the geometric $\mathbb{S}^1$-model on which this thesis is mainly based.

### 3.3.1 *The classic configuration model*

The first model we study is the *configuration model* (CM) [123, 124]. This model focuses on the degree distribution and is defined in a similar way as the $G(N, M)$. Here, we sample from the set $\mathcal{G}$ of all networks those that have a certain *degree sequence* $\boldsymbol{k} = \{k_i\}_{i \in \mathcal{V}}$. Thus, as was the case for the $G(N, M)$, the CM defines a microcanonical ensemble.

Where sampling uniformly from the $G(N, M)$ ensemble was trivial, the same cannot be said for the configuration model. The most common sampling method is based on stubs. The basic idea is that we visualize our degree sequence as $N$ nodes $i$ with each $k_i$ stubs attached to them. We then choose two stubs uniformly at random, and connect them[2]. The final network is obtained by continuing this process until all stubs are connected [125, 126].

There are, however, several problems with this method, all having to do with the creation of multiple links between pairs of nodes as well as self-loops between a node and itself. Both of these objects are at odds with the assumption that we are working with simple graphs. The creation of these types of links also violates the assumption that each network in the ensemble is created with equal likelihood [16]. It can be shown that the expected amounts of multiedges and self-loops are constants with respect to the system size, as long as $\langle k^2 \rangle$ is as well, i.e., if the variance remains finite in the thermodynamic limit. In this case the density of these types of edges vanishes in the thermodynamic limit, implying that there the proposed method does properly sample from the ensemble [112, 127, 128]. For power law degree sequences with exponent $\gamma \leq 3$, the second moment $\langle k^2 \rangle$ diverges with $N$. This results in a macroscopic amount of self-loops and multiedges, which implies that for these types of degree sequences the stub matching method fails even in the thermodynamic limit [127][3]. One might wonder why we cannot just reject the stub matchings that lead to non-desirable edge types; the problem is that this would lead to non-uniform sampling from the ensemble.

Other, uniform, sampling methods can be employed that do not lead to multiedges nor self-loops, the most notable being degree-preserving rewiring [15, 129, 130]. Here, we start with a single network realization with no self-loops nor multiedges[4]. Then, two edges, $l_1 = \{i, j\}$ and $l_2 = \{l, m\}$, are selected and rewired

2 The attentive reader recognizes in this process the implicit null-model in the definition of modularity.

3 We already observed this result in the section on degree correlations. There, the stub matching method was used to create neutral networks, and also there we saw that for $\gamma \leq 3$ the expected amount of multiedges was too high.

4 Not all degree sequences permit such a realization. Sequences for which it is possible to generate a simple graph are called *graphical* [131].

such that $l_1' = \{i, m\}$ and $l_2' = \{j, l\}$. The rewiring is rejected if it leads to a double edge or self-loop. An important open problem is the *mixing time* [132] of such a procedure, i.e., how many swaps should one perform before the new graph can be seen as independent from the old? One popular method is to take some network measure and compare its variance for a sequence of sampled graphs to its variance across multiple sequences. One might also study the the autocorrelation of such measures within a sequence to see when to stop rewiring.

### 3.3.2 *The soft configuration model*

Another way to circumvent the issues with sampling from the microcanonical ensemble defined by the CM is by relaxing the constraint of having an exact degree sequence and rather asking for an expected degree sequence[5]. This then leads us to define the *soft configuration model* (SCM), where, as in the case of the $G(N, p)$, we work with a canonical ensemble [116, 134–137].

Let $\boldsymbol{k}^*$ denote the desired degree sequence. We then want to find the probability distribution $P(G)$ over all simple graphs $G \in \mathcal{G}$ that 1) is normalized, 2) is maximally random, i.e. that maximizes the Shannon entropy

$$S = -\sum_{G\in\mathcal{G}} P(G)\log\left(P(G)\right), \tag{3.3}$$

and 3) leads to the expected degree sequence being the desired one:

$$\sum_{G\in\mathcal{G}} \boldsymbol{k}(G)P(G) = \boldsymbol{k}^*, \tag{3.4}$$

where $\boldsymbol{k}(G)$ denotes the degree sequence of the network $G$. We can obtain the distribution that accomplishes these demands through the method of Lagrange multipliers. We introduce the Lagrange multipliers $\alpha$ and $\{\nu_i\}_{i\in V}$ and maximize the entropy as

$$\frac{\delta}{\delta P(G)}\left(S + \alpha\left(1 - \sum_{G\in\mathcal{G}} P(G)\right) + \sum_{i\in V}\nu_i\left(k_i^* - \sum_{G\in\mathcal{G}} k_i(G)P(G)\right)\right) = 0. \tag{3.5}$$

This then leads to

$$\log P(G) + 1 + \alpha + \sum_{i\in V}\nu_i k_i(G) = 0, \tag{3.6}$$

which can be rewritten as

$$P(G) = \frac{1}{Z}e^{-H_{SCM}(G)}, \tag{3.7}$$

where $H_{SCM}(G)$ is the *graph Hamiltonian*

$$H_{SCM}(G) = \sum_{i\in V}\nu_i k_i(G) \tag{3.8}$$

5 This might also actually be a more realistic constraint as the degree sequence taken from a real network cannot necessarily be taken to be exact, as measurement errors and noise might have affected it [133].

and $Z$ is the *partition function*

$$Z = e^{1+\alpha} = \sum_{G \in \mathcal{G}} e^{-H_{SCM}(G)}. \tag{3.9}$$

In the last step we set $\alpha$ such that the probability is indeed normalized[6].

With this we have defined the ensemble, but how do we sample from it? First, we note that by using $k_i = \sum_{j \in V} A_{ij}$ we can rewrite the Hamiltonian as

$$H(G) = \sum_{i<j} (\nu_i + \nu_j) A_{ij}(G), \tag{3.10}$$

where the sum goes over all $\binom{N}{2}$ pairs of nodes in the network and $A_{ij}(G)$ denotes the adjacency matrix of network $G$. We then rewrite the partition function as

$$Z = \sum_{\{\boldsymbol{A}\}} \prod_{i<j} e^{-(\nu_i+\nu_j)A_{ij}} = \prod_{i<j} \sum_{A_{ij}=0}^{1} e^{-(\nu_i+\nu_j)A_{ij}} = \prod_{i<j} \left(1 + e^{-(\nu_i+\nu_j)}\right) \tag{3.11}$$

where we have used that $A_{ij} \in \{0, 1\}$ because we are working with simple graphs. Plugging this expression into Eq. (3.7) we obtain

$$P(G) = \prod_{i<j} \frac{e^{-(\nu_i+\nu_j)A_{ij}(G)}}{1 + e^{-(\nu_i+\nu_j)}} = \prod_{i<j} p_{ij}^{A_{ij}} (1 - p_{ij})^{1-A_{ij}}, \tag{3.12}$$

where we have defined the *connection probability*

$$p_{ij} = \frac{1}{1 + e^{\nu_i+\nu_j}}. \tag{3.13}$$

The reader might have noticed that this connection probability has the same form as the Fermi-Dirac distribution describing the occupation of energy states in a gas of free fermions. This was to be expected, as the two systems are surprisingly similar. In the case of the SCM, the available states are defined by the pairs of nodes, and each state can either be filled or not, reminiscent of the Pauli exclusion principle.

We have thus shown that the probability $P(G)$ can be decomposed into $\binom{N}{2}$ Bernoulli trials with success probability $p_{ij}$ conditioned on the Lagrange multipliers, which immediately gives us a way to sample from the ensemble: We just need assign to each node $i$ a value $\nu_i$ and then connect each pair of nodes $i$ and $j$ with probability $p_{ij}$. We often choose to introduce the suggestively named *hidden degrees* $\kappa_i = \eta e^{-\nu_i}$ with which we rewrite the connection probability as

$$p_{ij} = \frac{1}{1 + \frac{\eta^2}{\kappa_i \kappa_j}}. \tag{3.14}$$

6 This procedure can be generalized to include many different types of constraints on graph properties, leading to the large class of *exponential random graph models.*

The only remaining task is to figure out how to set the Lagrange multipliers $\{\nu_i\}_{i\in V}$, or equivalently, the hidden degrees $\{\kappa_i\}_{i\in V}$. We do this by noting that

$$k_i^* = \bar{k}_i = \sum_{j\neq i} p_{ij}, \tag{3.15}$$

where $\bar{k}_i$ is the expected degree of node $i$ and where $k_i^*$ again represents the desired degree. This defines a system of $N$ equations which fully fixes the values of the Lagrange multipliers and hidden parameters.

### 3.3.3 *The Chung-Lu model*

If we now assume for a moment that $\kappa_i \ll \eta$ for all $i \in V$, we can approximate Eq. (3.14) as $p_{ij} \simeq \frac{\kappa_i\kappa_j}{\eta^2} \equiv \tilde{p}_{ij}$. Plugging this into Eq. (3.15) we obtain

$$\bar{k}_i \simeq \sum_{j\neq i} \tilde{p}_{ij} = \kappa_i \frac{1}{\eta^2} \sum_{j\neq i} \kappa_j = \kappa_i \frac{N\langle\kappa\rangle}{\eta^2}, \tag{3.16}$$

which is proportional to $\kappa_i$. We can then choose $\eta$ such that the hidden degree $\kappa_i$ is equal to the expected degree of node $i$, justifying its name. The connection probability then becomes

$$\tilde{p}_{ij} = \frac{\kappa_i\kappa_j}{N\langle k\rangle}. \tag{3.17}$$

This connection probability belongs to the *Chung-Lu (CL) model* [118, 134] and is equivalent to the expected amount of edges in the configuration model (see Eq. (2.12)). There, we noted that for power law distributions with exponent $\gamma \leq 3$, the expected amount of edges between hubs is larger than one. Then, $\tilde{p}_{ij} > 1$ and, therefore, not a proper probability. In this case also the assumption $\kappa_i \ll \eta = \sqrt{N\langle k\rangle}$ fails for the largest hubs as $\kappa_c \sim N^{1/(\gamma-1)}$. In the CL, this is solved by redefining the connection probability as

$$p_{ij}^{CL} = \min\left(1, \frac{\kappa_i\kappa_j}{N\langle k\rangle}\right). \tag{3.18}$$

Note that neither Eq. (3.14) nor Eq. (3.18) can be factorized, i.e., they cannot be written as $p_{ij} \neq f(\kappa_i)g(\kappa_j)$ [137]. This implies that, *a priori*, degrees will be correlated. This is in line with our expectations for a power law network with $\gamma \leq 3$. Only when Eq. (3.17) is a good approximation of the connection probabilities do we expect to obtain neutral networks.

It has been shown that Eq. (3.18) and (3.14) lead to asymptotically equivalent graphs if $\langle\kappa\rangle$ and $\langle\kappa^2\rangle$ are finite [138]. If the second moment diverges, the resulting graphs might not be asymptotically equivalent, but many properties will still be the same. For example, all results derived in Ref. [139] using a generalization of Eq. (3.18) (resulting in the *inhomogeneous random graph model*) are shown to also hold when using Eq. (3.14).

### 3.3.4 *The hypersoft configuration model*

In the SCM, the degree sequence $\boldsymbol{k}^*$ sets the expected degree of each node in the network. This might, however, still be too strong of a constraint. Many networks are constantly evolving, leading to changing node degrees. What often not does change, however, is the degree distribution. The *hypersoft configuration model (HSCM)* is based on this idea [133, 140–143]. Here, instead of the hard constraint $\boldsymbol{k} = \boldsymbol{k}^*$ of the CM or the soft constraint $\bar{\boldsymbol{k}} = \boldsymbol{k}^*$ of the SCM, we have the hypersoft constraint that states that the degree distribution of the graphs in the sample should converge to the desired one. In order to achieve this we promote the hidden degrees $\kappa_i$ to random variables drawn from some distribution $\rho(\kappa)$. This model then defines a hypercanonical ensemble.

Let us assume the distribution of hidden degrees is power law:

$$\rho(\kappa) = \begin{cases} (\gamma-1)\kappa_0^{\gamma-1}\kappa^{-\gamma} & \text{if } \kappa_0 < \kappa \\ 0 & \text{else,} \end{cases} \tag{3.19}$$

i.e., let $\kappa$ have a Pareto distribution with *scale parameter* $\kappa_0$ and *shape parameter* $\gamma - 1$. Note that the mean of such a distribution is $\langle\kappa\rangle = \frac{\gamma-1}{\gamma-2}\kappa_0$. The HSCM equivalent of Eq. (3.15) is[7]

$$\bar{k}(\kappa) = N\int \rho(\kappa')p(\kappa,\kappa')\mathrm{d}\kappa', \tag{3.22}$$

and plugging in Eq. (3.19) we obtain

$$\bar{k}(\kappa) = N\int_{\kappa_0}^{\infty} \frac{(\kappa')^{-\gamma}}{1+\frac{\eta^2}{\kappa\kappa'}}\mathrm{d}\kappa' = N\,{}_2F_1\left(1,\gamma-1;\gamma;-\frac{\eta^2}{\kappa_0\kappa}\right), \tag{3.23}$$

where ${}_2F_1(a,b;c;z)$ is the ordinary hypergeometric function.

If we assume that $\eta^2 \gg \kappa\kappa_0$ this reduces to

$$\bar{k}(\kappa) \simeq \frac{N\langle\kappa\rangle}{\eta^2}\kappa. \tag{3.24}$$

We can then choose $\eta = \sqrt{N\langle\kappa\rangle}$ such that the hidden degrees and expected degrees coincide. We then immediately see that $\langle k\rangle = \langle\kappa\rangle$.

7 This can be seen as follows: Given a certain sequence of hidden degrees $\{\kappa_i\}_{i\in V}$, we can find the expected degree of node $i$ using Eq. (3.15). Now, we want to marginalize over the hidden degrees as they are random variables as well. In mathematical notation this can be written as

$$\bar{k}(\kappa_i) = \mathbb{E}\left(\bar{k}(\kappa_i)|\{\kappa_j\}_{j\neq i}\right) = \mathbb{E}\left(\sum_{j\neq i} p(\kappa_i,\kappa_j)|\{\kappa_j\}_{j\neq i}\right). \tag{3.20}$$

We now use the fact that in the SCM, edges are drawn independently, which allows us to switch the sum and the expectation value leading to

$$\bar{k}(\kappa_i) = \sum_{j\neq i}\int \rho(\kappa')p(\kappa_i,\kappa')\mathrm{d}\kappa' = (N-1)\int \rho(\kappa')p(\kappa_i,\kappa')\mathrm{d}\kappa', \tag{3.21}$$

where in the last step we have used that all $\kappa_j$ are drawn from the same distribution. If we now relabel $\kappa_i \to \kappa$ and assume $N \gg 1$, we obtain Eq. (3.22).

Let us return for a moment to the assumption that $N\langle\kappa\rangle = \eta^2 \ll \kappa\kappa_0$, which can be rewritten as $N \ll \kappa\,(\gamma-2)/(\gamma-1)$. Intuitively, this should hold for all $\kappa$, as the degree of a node can never exceed the system size. However, $\kappa$ is not the degree of the node, but rather a hidden parameter that we assign to each node. In principle there is no reason why it should be impossible for it to be larger than $N$. In practice, $\kappa$ is drawn from a Pareto distribution, and we know that $\overline{\kappa_{\max}} \sim N^{1/(\gamma-1)}$, which is much smaller than $N$ for $\gamma > 2$. Yet, $\overline{\kappa_{\max}}$ gives the expected maximal hidden degree. Fluctuations might cause some nodes to have hidden degrees larger than this value, and even larger than $N$. In fact, it can be shown[8] that, for finite systems, $P(\kappa_{\max} \geq N)$ is rather large when $\gamma \gtrsim 2$. In this region it might therefore be useful to introduce an artificial cut-off in the distribution, enforcing $\kappa_{\max} < N$. In this case, we redefine the distribution of the hidden degrees as

$$\tilde{\rho}(\kappa) = \begin{cases} \frac{(\gamma-1)\kappa_0^{\gamma-1}}{1-\left(\frac{\kappa_c}{\kappa_0}\right)}\kappa^{-\gamma} & \text{if } \kappa_0 < \kappa < \kappa_c \\ 0 & \text{else.} \end{cases} \tag{3.25}$$

Different cut-offs can of course be chosen, but in this thesis we will tend to set $\kappa_c = \kappa_0 N^{1/(\gamma-1)}$, the natural cut-off[9].

We have shown that for the HSCM with power law hidden degree distribution, $\overline{k}(\kappa) = \kappa$. However, this result does not imply that the hidden degree distribution $\rho(\kappa)$ and actual degree distribution $P(k)$ are equal. It can be shown that the actual degree distribution will approach a mixed Poisson distribution

$$P(k) \simeq \frac{1}{k!}\int \kappa^k e^{-\kappa}\rho(\kappa)\mathrm{d}\kappa, \tag{3.26}$$

which, in the case of the Pareto distribution Eq. (3.19), leads to

$$P(k) \simeq (\gamma-1)\kappa_0^{\gamma-1}\frac{\Gamma(k+1-\gamma,\kappa_0)}{k!}, \tag{3.27}$$

where $\Gamma(a,b)$ is the incomplete gamma function. This distribution has a power law tail with exponent $\gamma$, implying the resulting network is scale-free in the sense that its degree distribution is regularly varying.

### 3.3.5 *The relation between the SCM and HSCM*

So far, we have seen three related ways to generate networks with non-trivial degree distributions. First, the configuration model allows us to fix the degree

8 Given $\rho(\kappa) = (\gamma-1)\kappa_0^{\gamma-1}\kappa^{-\gamma}$, the probability of a single event being smaller than $N$ is $P(\kappa \leq N) = \int_{\kappa_0}^{N}\mathrm{d}\kappa\rho(\kappa) = 1-(\kappa_0/N)^{(\gamma-1)}$. The probability of all events being smaller than $N$, which is equivalent to the maximal value being smaller than $N$ is equal to $P(\kappa_{\max} \leq N) = P(\kappa \leq N)^N$. The probability of the maximal value being larger than $N$ is then $P(\kappa_{\max} > N) = 1 - P(\kappa \leq N)^N$. For a network with $N = 10^6$, $\gamma = 2.1$ and $\kappa_0 = 2$, this leads to $P(\kappa_{\max} > N) \approx 0.4$. Note, however, that this is a finite size effect as $\lim_{N\to\infty} P(\kappa_{max} > N) \to 0$ for $\gamma > 2$.

9 Note that one can also introduce a soft cut-off by setting $\rho(\kappa) \propto e^{-\kappa N^{1/(1-\gamma)}}\kappa^{-\gamma}$. In Ref. [75] it was shown that this distribution converges to the Pareto distribution in the thermodynamic limit.

sequence exactly. Second, relaxing this constraint we arrive at the soft configuration model, where we fix the expected degree sequence. Finally, in the hypersoft configuration model we are no longer interested in the degree sequence but rather in the degree distribution.

It turns out that in many cases, the SCM and HSCM are equivalent for all practical proposes. To justify this statement, let us draw a set of hidden degrees $\{\kappa_i\}_{i\in V}$ from some distribution $\rho(\kappa)$. As we have fixed the hidden degrees, the ensemble we need to use to draw graphs is the SCM. Therefore, to obtain the expected degree of a node we use

$$\bar{k}_i = \sum_{j\neq i} p(\kappa_i, \kappa_j). \tag{3.28}$$

However, let us now define the following measure,

$$\rho_N(\kappa) \equiv \frac{1}{N}\sum_{i\in V} \delta(\kappa - \kappa_i). \tag{3.29}$$

We can then rewrite Eq. (3.28) as

$$\bar{k}_i = N\int \mathrm{d}\kappa' \rho_N(\kappa') p(\kappa_i, \kappa'). \tag{3.30}$$

Now, if we assume that in the thermodynamic limit $\rho_N(\kappa) \to \rho(\kappa)$, we see that we obtain the definition of the expected degrees in the HSCM as defined in Eq. (3.22). In fact, it was shown in Ref. [139] that for many network properties it does not really matter if we fix the set of hidden degrees and work in the SCM or keep them as random variables and work in the HSCM[10], as long as for every set $A$ in the support of $\rho(\kappa)$, it can be shown that $\rho_N(A) \equiv |\{i \in V | \kappa_i \in A\}|/N \to \rho(A)$ as $N \to \infty$.

Even though many properties might be the same for both distributions, there are also differences. For example, let us study the Shannon entropy of the HSCM [133]. This entropy, which we denote as $S(G_{HSCM})$, can be lower bounded as

$$S(G_{HSCM}) \geq \mathbb{E}\left(S(G_{SCM})|\{\kappa_i\}_{i\in V}\right). \tag{3.31}$$

Let us now recall that for the SCM, $P(G_{SCM})$ could be decomposed into $\binom{N}{2}$ Bernouilli trials with different success probabilities $p_{ij}$, for which we know that the entropy is

$$S(\mathrm{Be}(p)) = -p\log(p) - (1-p)\log(1-p). \tag{3.32}$$

This then implies that

$$S(G_{SCM}) = \sum_{i<j} S(\mathrm{Be}(p_{ij})), \tag{3.33}$$

10 Borrowing from the field of disordered systems, we can call the SCM *quenched* in this setting and the HSCM *annealed*. In the former case we assume the expected degree sequence to be fixed but unknown, whereas in the latter they are continuously resampled [142].

and finally that

$$S(G_{HSCM}) \geq \binom{N}{2} \iint S(\mathrm{Be}(p(\kappa', \kappa'')))\rho(\kappa')\rho(\kappa'')\mathrm{d}\kappa'\mathrm{d}\kappa''. \tag{3.34}$$

The right hand side is once again the entropy of the SCM if we do the same large system approximation as before (see Eq. (3.29)). However, we see that the two entropies are, *a priori*, not the same. The fact that the hidden degrees are random variables in the HSCM adds some uncertainty, and therefore increases the entropy[11]. Calculating how much entropy is added is not a trivial task. In Ref. [133], the authors employ the theory of *graphons* to show that the missing entropy is subleading to that of the SCM[12].

Because of these similarities, we will not be very careful in distinguishing between these two models in this thesis. The same holds true when discussing their geometric counterparts, which we will introduce in the next Chapter. In general, results will be assumed to hold for both ensembles. If not, we will assume that it is clear from the context whether the hidden parameters are random variables or not.

### 3.3.6 *Clustering*

So far, the discussion of these models has been limited to the properties related to the node degrees. Let us now turn to another key property: the clustering coefficient. For the CM, it has been shown that the average local clustering coefficient, as defined in Sec. 2.2.4 scales as

$$\bar{c}_{CM} = \frac{1}{N}\frac{(\langle k^2\rangle - \langle k\rangle)^2}{\langle k\rangle^3}. \tag{3.35}$$

For the derivation of this result it needs to be assumed that the probability of an edge connecting nodes $i$ and $j$ is equal to $k_i k_j/(2M)$. However, as we have now seen several times in this introduction, for scale-free networks this is only the case when $\gamma > 3$, i.e. when the largest hubs have relatively few neighbors. In that case, $\langle k^2\rangle \sim N^0$, and we can conclude that $\bar{c} \sim N^{-1}$. This is the same fast decay as in the case of the ER model.

When $\gamma \leq 3$, $\langle k^2\rangle$ diverges. This can be fixed by including the natural cut-off $k_c \propto N^{1/(\gamma-1)}$ such that $\langle k^2\rangle \sim N^{(3-\gamma)/(\gamma-1)}$. Plugging this into Eq. (3.35) we obtain that $\bar{c} \sim N^{(7-3\gamma)/(\gamma-1)}$, which diverges in the thermodynamic limit when $\gamma < 7/3$. This is of course unphysical as the local clustering coefficient was defined as the amount of triangles attached to a node divided by the total possible amount

11 Another way of thinking about this is as follows: Say we know there is a link between node $i$ and node $j$, i.e., $A_{ij} = 1$. In the HSCM, we do not know the hidden degrees as they are sampled a new for each realization. Now, the fact that $A_{ij} = 1$ tells us something about the hidden degrees of node $i$ and $j$; they are likely to be high. This then implies that $\rho(\kappa_i = \kappa | A_{ij} = 1) \neq \rho(\kappa)$, and so we cannot decompose the entropy of the graph into the entropy of $\binom{N}{2}$ Bernoulli trials.

12 The leading order term of the entropy of the microcanonical ensemble in the CM is also equal to that of the the SCM and HSCM; all three scale as $S \simeq \frac{\langle k\rangle}{2} N \ln N$ [133, 142].

of triangles, and can thus never exceed 1. This result is related to the creation of large amount of multilinks between hubs in the configuration model when $\gamma \leq 3$.

In Ref. [144], the clustering of the HSCM with power law hidden degree distribution was studied. The authors found that the scaling behavior for general $\kappa_c$ is

$$\overline{c}_{HSCM} \simeq \frac{1}{\kappa_s^{2(\gamma-2)}} \begin{cases} \xi_1(\gamma) & \text{if } \kappa_{\max} = \kappa_s \gg 1 \\ \xi_2(\gamma) \ln\left(\frac{\kappa_{\max}}{\kappa_s}\right) & \text{if } \kappa_{\max} \gg \kappa_s \gg 1, \end{cases} \tag{3.36}$$

where $\xi_1(\gamma)$ and $\xi_2(\gamma)$ are both functions of $\gamma$. In the first case, when $\kappa_{\max} = \kappa_s \sim \sqrt{N}$, i.e. before structural correlation enter the picture, Eq. (3.35) holds. However, if the cut-off is set to $\kappa_{\max} = \kappa_c \gg \kappa_s$, Eq. (3.36) reduces to

$$\overline{c}_{HSCM} \sim N^{2-\gamma} \ln N. \tag{3.37}$$

These results were then generalized to other hidden parameter models such as the CL where the same results were found [145].

The scaling in Eq. (3.37) is clearly very slow, especially close to $\gamma = 2$. In Ref. [144] it was shown that $\gamma = 2.1$, clustering remains significant even for system sizes as large as $N = 10^8$.

# 4

# NETWORK GEOMETRY

In the previous Chapters we saw that real networks share several important properties. Many network models attempt to capture these features, but so far we have not encountered any that is able to reproduce them all (see Tab. 4.1). In this Chapter we will give an introduction to the field of *network geometry*, which is based on the idea that nodes live in some underlying metric space that conditions their connectivity. We will see that this can lead to networks that possess the necessary properties, i.e., to networks that are simultaneously small world, clustered, heterogeneous and sparse. This framework will form the basis of this thesis.

## 4.1 SIMILARITY SPACE

So far in this introduction, the only network model that has been able to produce clustered networks is the Watts-Strogatz model. In Sec. 3.2, we noted that this is due to the locality of the connections; nodes are distributed along the circle, and for small $p$ the connections are mainly between close neighbors. Let us now try and understand why this leads to many triangles. We will do so by looking at the very first geometric network model, the *random geometric graph* (RGG) introduced by Gilbert in 1961 [23].

Gilbert assumed nodes to be placed following a Poisson point process on the infinite real plane $\mathbb{R}^2$. The nodes are then connected if they lie closer together than some critical distance $r$. In other words, two nodes $i$ and $j$ with associated

Table 4.1: Summary of the different model classes introduced in Chapter 3. For each model we state whether or not its networks are small world, clustered and heterogeneous.

| | Small World | Clustered | Heterogeneous |
|---|---|---|---|
| *Erdös-Renyi* (ER) | Yes | No | No |
| *Watts-Strogatz* (WS) | Yes | Yes | No |
| *Configuration Model* (CM) | Yes | No | Yes |

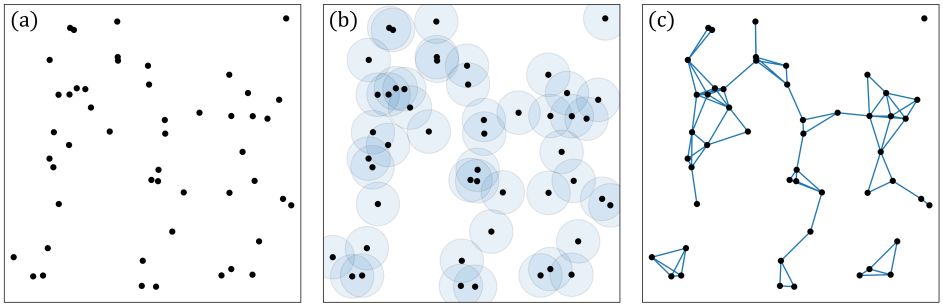



Figure 4.1: The construction of an RGG: (a) We start with a metric space (in this case $[0,1]^2$) and distribute nodes uniformly over it. (b) We then show the disks with radius $r/2$, where $r$ is the critical distance, centered around the nodes. (c) If two of these disks intersect, the nodes are connected, leading to the graph shown in this panel.

coordinates $\boldsymbol{x}$ and $\boldsymbol{y}$ are connected if $\|\boldsymbol{x}-\boldsymbol{y}\| \leq r$. We visualize this process in Fig. 4.1[1].

We immediately see that this model results in a large amount of triangles[2]. How can we understand this? First, we note that the underlying metric space must obey the triangle inequality, which states that if we take three points $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$, then

$$\|\boldsymbol{x}-\boldsymbol{y}\| + \|\boldsymbol{x}-\boldsymbol{z}\| \geq \|\boldsymbol{y}-\boldsymbol{z}\|. \tag{4.1}$$

This has immediate implications for the connectivity of the network, as is introduces correlations between the links. Say we have three points, $i$, $j$ and $k$, located at $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ respectively and assume we know that $i$ and $j$ and $i$ and $k$ are connected. By definition, this mean that $\|\boldsymbol{x}-\boldsymbol{y}\| \leq r$ and $\|\boldsymbol{x}-\boldsymbol{z}\| \leq r$. Plugging this into the triangle inequality we obtain that $\|\boldsymbol{y}-\boldsymbol{z}\| \leq 2r$. So, where originally $P(A_{jk}=1)$ was very small, as both points could be arbitrarily far away from each other, the fact that we know $A_{ij}=1$ and $A_{ik}=1$ increases this probability significantly. In other words, if we know a node is connected to two neighbors, the probability of these neighbors being close together and therefore connected is high. And, of course, if the neighbors are also connected we create a triangle in the network.

We can now ask why it might be natural that nodes are embedded in some metric space, and why it makes sense to connect close by points. Gilbert's motivation was straightforward; in many real world systems, nodes have an explicit spatial locations that condition their connectivity. For example, in communications networks, nodes might be radio towers that have a range of $r$. Another application would be the spread of infectious diseases. People occupy some position in space,

1 The system depicted here is not exactly the same as the one introduced by Gilbert. Here, instead of working with a Poisson point process on $\mathbb{R}^2$, we work on $[0,1]^2$ where $N$ nodes are distributed uniformly. It can, however, be shown that these two models are often interchangeable [146, 147].

2 It can be shown that the clustering coefficient of this model is given by $\bar{c} = 1 - \frac{3\sqrt{3}}{4\pi} \approx 0.6$, independent of $r$ and $N$ [148]

and if the disease is airborne it might be assumed to be able to infect others only within some range $r$.

Even though this interpretation is intuitive, it is also limiting. First of all, not all systems are explicitly embedded in real space. And second, even if they are, the connectivity is rarely only the result of the spatial distance between nodes. To overcome these limitations, let us introduce the concept of a *similarity space*[3]. Here, a node's location encodes some information about it[4]. Then, these properties are assumed to condition the connectivity. If nodes are close together, their properties are similar and they are more likely to be connected.

The idea that more similar nodes are more likely to be connected comes from the social sciences [149]. There, it is called *homophily* and a typical example is a friendship network: People are more likely to befriend those with similar ethnic, educational, socioeconomic, etc. profiles [150]. This same mechanism has been found to affect many other fields of study, from citation networks [151] to the internet [152] to international trade [53].

It is important to note that the similarity space is not equivalent to a feature space where the dimensions directly encode different node properties. We do not claim that in a social network, for example, one can assign one dimension to income, one to ethnicity, one to geographic location, one to age, etc., and then expect to create a realistic network based on the distances in this extremely high dimensional space. In general, many intrinsic node properties can affect the connectivity of a network, and at the same time, not all of these properties will be equally important. Furthermore, here we exclusively consider properties whose values can be compared to those of other nodes such that the similarity framework is applicable[5]. Taking this into account, we define the similarity space as an abstract low dimension space where the location of a node encodes its relevant properties.

We can understand this as follows: say we start with some extremely high dimensional feature space, where each dimension encodes some property of a node. The idea is then that, through some unspecified dimension reduction process, one can summarize all this information into a few dimensions. That such a dimensional reduction is feasible is sustained by several observations. First, it is a well known fact in data science that oftentimes one can extract meaningful low-dimension information from high-dimensional data sets [155–157]. Second, it has been shown that the clustering coefficient in the RGG goes to zero as the dimension goes to infinity [148], which is not in line with observations. Finally, in the context of the model we will introduce next, low dimensional similarity spaces are best able to explain network structures [158, 159].

3 We will also refer to this space as a *latent*, *hidden* and *underlying* metric space.

4 Of course, geographical location can be one of these properties, meaning that explicitly geometric systems can also be described in this framework.

5 Properties such as popularity, influence, rank, i.e., those that affect the connectivity on their own will be discussed later. Note that there are also other properties that can be compared but where similarity does not lead to higher connection probabilities. This is considered in the *complementarity* framework, where it is noted that sometimes it is beneficial for agents to be different when making connections [153, 154].

## 4.2 THE $\mathbb{S}^1$-MODEL

The RGG is intuitive, easy to implement and leads to clustered networks. However, it also has a couple of important problems. Just like in the case of regular ring lattice, it is large world. All interactions are local, and there are thus no "short-cuts" that might reduce the mean hop distance between nodes (see Fig. 4.1c). In addition, the RGG does not lead to heterogeneous degree-distributions. In fact, it can be shown that, as in the case of the ER graph, the degree distribution is given by a Poisson distribution. While several models have been introduced that address (some of) these issues [160–162], in this section we focus on the $\mathbb{S}^1$-model, first introduced in 2008 by Serrano *et al.* [30], because of its unique properties. As we will see, this model allows for the production of sparse graphs with arbitrary degree distributions. They can be clustered or tree-like, small world or large world. All this can be achieved with a single connectivity law that also makes its ensemble maximally random. For the derivation of this model we follow Ref. [163].

### 4.2.1 *The homogeneous case*

The first challenge we tackle is that of the small world property. We approach this in a similar way to the WS model; we want to introduce long range connections into the network. To do so, we use the methods introduced in Sec. 3.3.2. We want to find a $P(G)$ over all the simple graphs $G \in \mathcal{G}$ that maximizes the entropy and that satisfies some convenient constraints, in this case the total amount of edges as well as the total energy of the system, which fixes the amount of long range connections. The energy should therefore be related to the distance between nodes. Let us say that we assign some quantity

$$\epsilon_{ij} = f(d_{ij}) \tag{4.2}$$

to each pair of nodes, where $f(d_{ij})$ is some function of the distance between nodes. For reasons that will become clear later on, we will refer to this quantity as the energy of the potential link.

Now, in order for us to maintain the clustered nature of our network, we want connections to mostly be local, implying that they should be easier to create, and therefore be associated to lower energies. As was the case in the WS-model, we only want a few long-range links, which should thus be difficult to create. Hence, they will be assigned high energies. This implies that $f(d_{ij})$ should be an increasing function of the distance. In fact, it was shown in Ref. [163] that the only function that allows for both small-world and clustered networks is given by

$$f(d_{ij}) = \ln(ad_{ij}), \tag{4.3}$$

where $a$ is a constant. If the function grows faster, there are not enough long range connections and the network ends up being large world. If it grows slower, too many are created and we loose the clustered nature of the resulting graph.

We can now use this energy per node pair to define a new constraint for the entropy maximization procedure by fixing the expected total energy of the system

$$E^* = \langle E \rangle = \sum_{G \in \mathcal{G}} P(G) E(G) = \sum_{G \in \mathcal{G}} P(G) \sum_{i<j} \epsilon_{ij} A_{ij}(G). \tag{4.4}$$

At the same time, we also want to fix the expected average degree of the system, which can be achieved with the following constraint:

$$M^* = \langle M \rangle = \sum_{G \in \mathcal{G}} P(G) M(G) = \sum_{G \in \mathcal{G}} P(G) \sum_{i<j} A_{ij}(G). \tag{4.5}$$

Following the same steps as in Sec. 3.3.2, one then obtains

$$P(G) = \frac{1}{Z} e^{-\beta(E(G) - \mu M(G))}, \tag{4.6}$$

reminiscent of the *grand canonical* Gibbs measure with inverse temperature $\beta$ and chemical potential $\mu$. Using again the fact that we are working with simple graphs, we can decompose this as in Eq. (3.12) where the connection probability is now given by

$$p_{ij} = \frac{1}{1 + e^{\beta(\epsilon_{ij} - \mu)}}. \tag{4.7}$$

This is exactly the Fermi-Dirac occupation distribution for a state with energy $\epsilon_{ij}$. There is thus a one to one mapping between this $\mathbb{S}^1$-model and a gas of free fermions: Pairs of nodes $\{i, j\}$ define the energy states and links represent the fermionic particles that can occupy them. When the Lagrange multiplier $\beta$ is high, the temperature of the system is low and only low-energy fermions are formed. Interactions are therefore short ranged. Conversely, when $\beta$ is low and the temperature is high we have fermions at all energy levels, and there will be many long range links.

The parameter $\beta$ can also be interpreted as quantifying the *geometric coupling*. When $\beta$ is high and connections are short ranged, the geometry plays a very important role in the topology. Nodes connect to their spatial nearest neighbors. Hence, the connectivity is strongly coupled to the geometry. Inversely, when links are long ranged, the geometry does not influence the connectivity much. Nodes can connect to any other in the network. Therefore, the geometric coupling is weak.

As we are working in the (grand) canonical ensemble, we can fix the lagrange multipliers $\beta$ and $\mu$ by solving

$$E^* = \langle E \rangle = \sum_{i<j} p_{ij} \epsilon_{ij} \tag{4.8}$$

and

$$M^* = \langle M \rangle = \sum_{i<j} p_{ij}, \tag{4.9}$$

respectively.

Let us now assume our similarity space is given by the $\mathbb{S}^1$ circle with radius $R = N/(2\pi)$, such that each node $i$ gets assigned an angular coordinate $\theta_i$. Note that this radius implies that the node density along the circle is always one, irrespective of the amount of nodes in the system. The distance between two nodes $i$ and $j$ is then given by $d_{ij} = R\Delta\theta_{ij}$, where $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$ is the angular separation between the nodes. If we now sample the angular coordinates i.i.d. from a uniform distribution, i.e. $\theta_i \sim \mathcal{U}(0, 2\pi)$, we can rewrite[6] Eq. (4.9) as

$$\langle M \rangle = \frac{N^2}{2} \iint \rho(\theta')\rho(\theta'')p(\theta', \theta'')\mathrm{d}\theta'\mathrm{d}\theta'. \tag{4.10}$$

Plugging in the connection probability (4.7) together with $\epsilon_{ij} = \ln(R\Delta\theta_{ij})$ and using the rotational symmetry of the system leads to

$$\langle M \rangle = \frac{N^2}{2} {}_2F_1\left(1, 1/\beta; 1 + 1/\beta; -\left(\frac{aN}{2\hat{\mu}}\right)^{\beta}\right), \tag{4.11}$$

where we have introduced $\hat{\mu} = e^{\mu}$, reminiscent of the *fugacity* from statistical physics. Using $\langle M \rangle = N\langle k \rangle/2$ and assuming that $aN/(2\hat{\mu}) \gg 1$, this can be rewritten as

$$\langle k \rangle = \frac{1}{1-\beta}\left(\frac{2\hat{\mu}}{a}\right)^{\beta} N^{1-\beta} + \frac{2\pi\hat{\mu}}{a\beta\sin(\pi/\beta)} + \mathcal{O}\left(N^{1-2\beta}\right). \tag{4.12}$$

Depending on whether $\beta$ is smaller or larger than one, either the first or second term will dominate. When $\beta > 1$,

$$\hat{\mu} \simeq \frac{\beta\sin(\pi/\beta)\langle k \rangle a}{2\pi}. \tag{4.13}$$

Conversely, when $\beta < 1$, we obtain

$$\hat{\mu} \simeq \frac{a}{2}(1-\beta)^{1/\beta}\langle k \rangle^{1/\beta} N^{1-1/\beta}, \tag{4.14}$$

which decreases with the system size.

How can we understand this change in behavior at $\beta = 1$? When $\beta \to \infty$, the connection probability approaches a step-function; $p_{ij} = 1$ when $d_{ij} < \hat{\mu}/a \equiv d_c$, and zero otherwise. As the density of points on the $\mathbb{S}^1$ circle with radius $R$ is constant, $d_c$ and therefore $\hat{\mu}$ must also be constant in order for us obtain a sparse graph. On the other extreme, where $\beta = 0$, the connection probability does not depend on the spatial coordinates. To obtain a sparse graph, $p_{ij}$ needs to then be proportional to $1/N$, and this can only be true if $\hat{\mu}$ decreases with the system

6 This can be interpreted in two different ways, just like in the case of Eq. (3.22). First, we can assume to be working in a large system where the sum in Eq. (4.9) can simply be approximated well by an integral. In this case we are working in the quenched case as the angular coordinates are fixed but unknown; we are working in the (grand) canonical ensemble. Second, the expected amount of edges can be written as $\mathbb{E}(M^*|\{\theta_i\}_{i\in V})$, in which case the ensemble is hyper grand canonical and we are working with annealed disorder.

size. Indeed, plugging Eq. (4.14) into the connection probability we observe that $p_{ij} = 1/(1 + N/\langle k \rangle) \simeq \langle k \rangle / N$, which is exactly the sparse $G(N, p)$ connection probability. It turns out the transition point between these two behaviors lies at $\beta = 1$. This point also marks the transition between a regime where geometry induces many triangles ($\beta > 1$) to a regime where the density of these triangles vanishes in the thermodynamic limit ($\beta \leq 1$) [30]. We will investigate this transition in depth in Chapter 5. In the following we will refer to the regime $\beta > 1$ as *strongly* geometric and the regime $\beta \leq 1$ as *weakly* geometric.

### 4.2.2 *The heterogeneous case*

Now that we have found that assigning an energy to each pair of nodes lets us tune the amount of long range connection in our network, and therefore allows us to generate networks that are both clustered as well as small world, we can add the heterogeneous component to our model. This is done in the same way as for the SCM, by assigning to each node $i$ an expected degree $k_i^*$ and constraining on it during the entropy maximization process. Where similarity between nodes is responsible for the clustering in our graphs, individual *popularity*, encoded by this expected degree, leads to heterogeneity. Constraining on both the energy and the expected degree[7] gives us the following connection probability

$$p_{ij} = \frac{1}{1 + e^{\beta \epsilon_{ij} + \nu_i + \nu_j}}, \tag{4.15}$$

where we can use Eq. (3.15) to fix the Lagrange multipliers $\{\nu_i\}_{i \in V}$ and Eq. (4.8) to fix $\beta$.

Upgrading both $\{\theta_i\}_{i \in V}$ and $\{\nu_i\}_{i \in V}$ to (quenched or annealed) random variables, we find the expected degree by

$$\begin{aligned} \bar{k}(\nu) &= \frac{N}{\pi} \iint \mathrm{d}\nu' \mathrm{d}\theta' \rho(\nu') p(0, \theta', \nu, \nu') \\ &= N \int \mathrm{d}\nu' \rho(\nu') {}_2F_1 \left( 1, 1/\beta; 1 + 1/\beta; -\left( \frac{N}{2} e^{\frac{\nu + \nu'}{\beta}} \right)^{\beta} \right), \end{aligned} \tag{4.16}$$

where we have once again used the rotational invariance of the system. If we assume the argument of the hypergeometric function to be large[8], we can use the same expansion as in Eq. (4.12) to obtain

$$\bar{k}(\nu) \simeq \int \mathrm{d}\nu' \rho(\nu') \left( \frac{2^{\beta} N^{1-\beta}}{1-\beta} e^{-\nu - \nu'} + \frac{2\pi}{\beta \sin(\pi/\beta)} e^{-\frac{\nu + \nu'}{\beta}} \right), \tag{4.17}$$

where we once again notice the change of behavior at $\beta = 1$. For $\beta > 1$, the second term dominates, and we obtain

$$\bar{k}(\nu) \simeq \frac{2\pi}{\beta \sin(\pi/\beta)} \left\langle e^{-\nu/\beta} \right\rangle e^{-\nu/\beta}. \tag{4.18}$$

7 We do not need the constraint on the total amount of links as this follows directly from the expected degree sequence

8 We prove this more rigorously for the power law case in Appendix A.1.1

Taking the expectation of this expression we find the relation between $\langle e^{-\nu/\beta}\rangle$ and the average degree $\langle k\rangle$. In order for $\bar{k}(\nu) = \kappa$, some desired degree, one needs to fix

$$\nu = -\beta\left(\ln\kappa + \frac{\mu}{2}\right), \tag{4.19}$$

where we have defined $\mu$ as

$$\mu = \ln\left(\frac{\beta\sin(\pi/\beta)}{2\pi\langle k\rangle}\right). \tag{4.20}$$

When $\beta < 1$, the first term in Eq. (4.17) dominates. Following the same steps as in the strongly geometric case, one obtains

$$\nu = -\left(\ln\kappa + \frac{\mu}{2}\right) \tag{4.21}$$

where

$$\mu = \ln\left(\frac{1-\beta}{2^\beta\langle k\rangle}N^{\beta-1}\right). \tag{4.22}$$

When $\beta = 1$, the approximation in Eq. (4.17) breaks down as both terms are of equal size. Using the fact that ${}_2F_1(1,1,2,-x) = -\ln(1+x)/x$ we can rewrite Eq. (4.16) as

$$\bar{k}(\nu) \simeq 2\ln(N)\int \mathrm{d}\nu'\rho(\nu')e^{-\nu-\nu'}. \tag{4.23}$$

Following the same steps as before, we then define

$$\nu = -(\ln\kappa + \frac{\mu}{2}), \tag{4.24}$$

where

$$\mu = \ln\left(\frac{1}{2\langle k\rangle\ln N}\right). \tag{4.25}$$

All in all we can combine these results to obtain the final form of the connection probability

$$p_{ij} = \frac{1}{1+\chi_{ij}}, \tag{4.26}$$

where

$$\chi_{ij} = \frac{(R\Delta\theta_{ij})^\beta}{(\hat{\mu}\kappa_i\kappa_j)^{\max(1,\beta)}}. \tag{4.27}$$

Here, we have once again introduced the fugacity-like parameter $\hat{\mu} = \exp\mu$. Note that these results reduce to the homogeneous case when $\kappa_i = \langle k\rangle\ \forall i \in V$[9]. One needs to set $a = 1/\langle k\rangle^2$.

In the $\beta \to 0$ limit, this connection probability reduces to

$$p_{ij} = \frac{1}{1+\frac{\langle k\rangle N}{\kappa_i\kappa_j}}, \tag{4.28}$$

9 For $\beta < 1$ one needs to also redefine $\mu \to \mu/\beta$, but this is a matter of convention.

which is exactly the form derived for the (H)SCM. This also motivates the $\max(1,\beta)$ in Eq. (4.27): Without it, the limiting connection probability would be $p_{ij} = 1/(1+N/\langle k\rangle) \simeq \langle k\rangle/N$, i.e., the limiting ensemble would be that of the ER model. The inverse temperature $\beta$ would then not only encode the coupling to the similarity space, but also tune the importance of the popularity of the nodes.

**Generating an $\mathbb{S}^1$-network**

The following algorithm generates an instance of the $\mathbb{S}^1$-model:

1. Fix the number of nodes $N$, the inverse temperature $\beta$ and the target average degree $\langle k\rangle$.
2. Assign a hidden degree $\kappa_i$ and an angular coordinate $\theta_i$ to each node $i \in V$, drawn from $\rho(\kappa)$ and $\rho(\theta) = 1/(2\pi)$, respectively[a].
3. We now have two options for fixing the fugacity $\hat{\mu}$:
   - If $N$ is large enough, one can set
   $$\hat{\mu} = \begin{cases} \frac{\beta\sin(\pi/\beta)}{2\pi\langle k\rangle} \text{ if } \beta > 1 \\ \frac{1}{2\langle k\rangle \ln N} \text{ if } \beta = 1 \\ \frac{1-\beta}{2^\beta\langle k\rangle} N^{\beta-1} \text{ if } \beta < 1 \end{cases} \tag{4.29}$$
   - If finite size effects are relevant, one needs to solve $\kappa = \bar{k}(\kappa)$ numerically, where
   $$\bar{k}(\kappa) = N\int \mathrm{d}\kappa'\rho(\kappa')\,{}_2F_1\left(1,\frac{1}{\beta},1+\frac{1}{\beta},-\frac{(N/2)^\beta}{(\hat{\mu}\kappa\kappa')^{\max(1,\beta)}}\right). \tag{4.30}$$
4. Connect every pair of nodes with probability
$$p_{ij} = \left(1+\frac{(R\Delta\theta_{ij})^\beta}{(\hat{\mu}\kappa_i\kappa_j)^{\max(1,\beta)}}\right)^{-1}. \tag{4.31}$$

This algorithm leads to a network instance like the one shown in Fig.4.2a.

[a] If one resamples these variables for every network instance, we are drawing from the hyper grand canonical ensemble. If one keeps them fixed, the ensemble is grand canonical. We will comment on this in Sec. 4.2.3

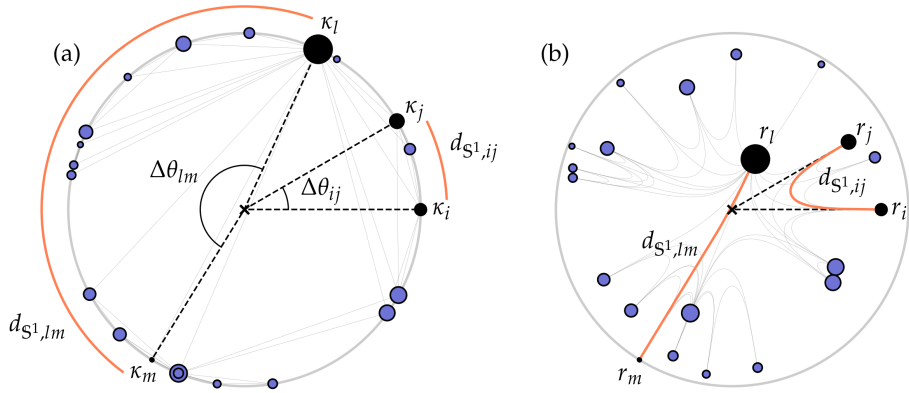



Figure 4.2: (a) An example of network generated by the $\mathbb{S}^1$-model. The angular separation $\Delta\theta$ and the distance in similarity space $d_{\mathbb{S}^1}$ are shown for two pairs of nodes. The size of a node reflects the magnitude of its hidden degree. (b) The same network mapped to the $\mathbb{H}^2$ model. Hidden degrees are now explicitly geometric and represented by a radial coordinate. The geodesics of the pairs of the same two pairs of nodes are shown, encoding their distance in hyperbolic space $d_{\mathbb{H}^2}$.

### 4.2.3 *Statistical properties*

The first question we would like to ask is what type of ensemble the $\mathbb{S}^1$-model defines. Let us start by analyzing similarity and popularity separately.

When it comes to the degree distribution, the situation is *a priori* equivalent to the configuration model. We can fix the degrees exactly, in expectation or by their distribution. In Sec. 3.3.5 we noticed that the HSCM is a probabilistic mixture of canonical ensembles, where the hyperparameters $\kappa$'s or $\nu$'s are random variables. Notice that the disorder here arises from the ensemble constraints $k_i^*$, which then translate to the Lagrance multipliers.

In the case of the energy, this would be equivalent to upgrading $\beta$ to a random variable, fixing the energy of the system only up to its distribution. In general, we will not go this route, but rather always assume that energy is fixed at the level of the expectation value, meaning that the constraint is canonical.

There are, however, other parameters related to the energy; namely, the angular coordinates $\{\theta_i\}_{i\in V}$. These parameters do not play the same role as the hidden degrees as they are not one-to-one related to any constraint. However, we can still choose to upgrade them to random variables. If they are considered fixed but drawn from some distribution, the disorder introduced by doing so is quenched. If they are continuously redrawn from this distribution, it is annealed. As was the case for the CM, we assume that for all practical purposes, these two situations are equivalent. For example, when fixing the average degree in Eq. (4.10), both lead to the same result. However, we do not claim these ensembles to be generally equivalent. In fact, at the level of the entropy they are probably not: For reasons similar

to the ones presented in Sec. 3.3.5, the random nature of the hyperparameters in the annealed case will most likely add uncertainty to the system, increasing the entropy. In general it can be assumed that we work with the quenched situation in this thesis, especially when it comes to to analytic results.

One final consideration arises from observing that by plugging the expression Eq. (4.19) for the Lagrange multiplier $\nu$ into Eq. (4.15) we can obtain the exact Fermi-Dirac form $p_{ij} = 1/(1 + e^{\beta(\epsilon_{ij}-\mu)})$ if we redefine

$$\epsilon_{ij} = \ln\left(\frac{d_{ij}}{\kappa_i \kappa_j}\right). \tag{4.32}$$

Thus, for $\beta \geq 1$, the heterogeneous case is exactly equivalent to a non interacting gas of fermions with energy levels given by Eq. (4.32). Note that in this case the $\{\kappa_i\}_{i\in V}$ are no longer Lagrange multipliers as we arrive at exactly the same ensemble by only constraining on the energy and the total amount of links. The hidden degrees are now parameters that fix the energy levels, in the same way as the angular coordinates.

When $\beta < 1$, things are not as straightforward. In this case, $\nu = -\ln(\kappa) - \mu/2$, implying that we can only achieve the Fermi-Dirac form when

$$\epsilon_{ij} = \ln\left(\frac{d_{ij}}{(\kappa_i \kappa_j)^{1/\beta}}\right). \tag{4.33}$$

We note that the energy levels now depend on the (inverse) temperature. This is non-standard, but also not unheard off. In fact, there is a large body of literature studying temperature dependent energy levels [164–170]. These appear in many physical systems [171, 172], and lead to generalized thermodynamic laws. This generalized thermodynamics can be shown to be non-extensive, in the sense that the extensive properties (energy, entropy, particle number) do not scale linearly with one another. Both non extensivity and temperature dependent energy states are related to systems with long range interactions [173, 174]. As we will see in Chapter 5, our system loses extensivity at $\beta < 1$, where links become long ranged. In this same region, the chemical potential depends on the system size $N$, as already observed in Eq. (4.22). This is also at odds with standard extensive thermodynamics as there the chemical potential is an intensive quantity. The interpretation of these facts are left as open questions; in Chapter 5 we will report on the non-standard thermodynamic properties of our system, but will not claim to place them in a broader physical context.

### 4.2.4 *Structural properties*

Just like in the SCM, $\overline{k}(\kappa) = \kappa$ does not imply that the degree distribution is the same as the hidden degree distribution. Indeed, it can once again be shown that the actual degree distribution approaches a mixed Poisson distribution given by Eq. (3.26). In the case of a Pareto hidden degree distribution, the degrees are

distributed according to Eq. (3.27), which has a power law tail. Thus, for all $\beta$, the $\mathbb{S}^1$-model is able to generate scale-free networks [29, 30, 32, 175, 176]. Of course, the means of the hidden and real degree distributions are identical due to our choice of $\mu$. Therefore, if $\langle\kappa\rangle \sim N^0$, so is $\langle k\rangle$, leading to sparse graphs.

One of the principle motivations for introducing the geometric framework was the promise of highly clustered graphs. This will be the topic of Chapter 5, but some results are already known [30, 32, 34, 163, 176, 177]. First of all, it can be shown that in the region $\beta > 1$, the average local clustering coefficient is a constant with respect to $N$. As the geometric coupling decreases, so does the clustering coefficient, and when $\beta \leq 1$, clustering vanishes entirely in the thermodynamic limit. This is in line with our expectations; when $\beta$ is large, geometry is important and the triangle inequality will induce many triangles. When $\beta$ is small, the coupling to the geometry is weak and clustering vanishes.

Another essential feature of the $\mathbb{S}^1$ is its ability to produce small-world graphs. In the homogeneous case, the mean shortest path scales as $\langle d\rangle \sim (\ln N)^{\Delta}$, for some $\Delta \leq 1$, when $1 < \beta < 2$ [178, 179]. When $\beta > 2$, there are not enough long-ranged links (short cuts) in the network to produce the small world effect; the mean shortest path scales linearly with the system size and our system is large world [163, 177, 180]. When $\beta < 1$, the network is similar to the ER, and $\langle d\rangle \sim \ln N$ [177]. The results for scale-free networks with $\gamma > 3$ are comparable [163, 175, 181]. For heterogeneous networks with $2 < \gamma \leq 3$, it can be shown that for any $\beta$, $\langle d\rangle \sim \ln\ln N$, indicating ultrasmall world networks [175].

Networks sampled from the $\mathbb{S}^1$ ensemble can be shown to contain no degree correlations, except for the structural ones inevitably generated in scale-free simple graphs with $\gamma \leq 3$. In fact, it can be shown that $\epsilon_{ij} \propto \ln d_{ij}$ is the only functional form that leads to this result [163].

Finally, we make some brief remarks on community structure. Here, the geometric nature of our model has very profound implications. Any set of nodes belonging to the same angular sector, regardless of how the sector has been defined, satisfies the classical definition of community, namely, a group of nodes with an internal average degree larger than the external. Any partition of the network in angular sectors has high modularity, even if the partition is totally arbitrary [182, 183][10]. Of course, these partitions are irrelevant from an explicative point of view. The communities found do not reflect any meaningful group structure.

However, these results do not imply that the $\mathbb{S}^1$-model is incompatible with community structure. Indeed, if one lets go of the homogeneous distribution of nodes on the circle, *soft communities* can be generated [49–51]. These are regions of high node density, which are separated by gaps that are larger than those expected from randomly placing nodes. Due to the geometric connection probability, such a non-uniform distribution of nodes will inevitably lead to topological communities [51, 184]. As we will see in Sec. 4.5, the angular coordinates of real network can be inferred from their adjacency matrix. Interestingly, these coordinates will gener-

10 These results all assume $\beta > 1$. They should fail as $\beta \to 1$, as there our model approaches the HSCM/ER. The behavior of modularity for $0 < \beta \leq 1$ remains an open problem, although some numerical results in Ref. [184] indicate it indeed decreases in this regime.

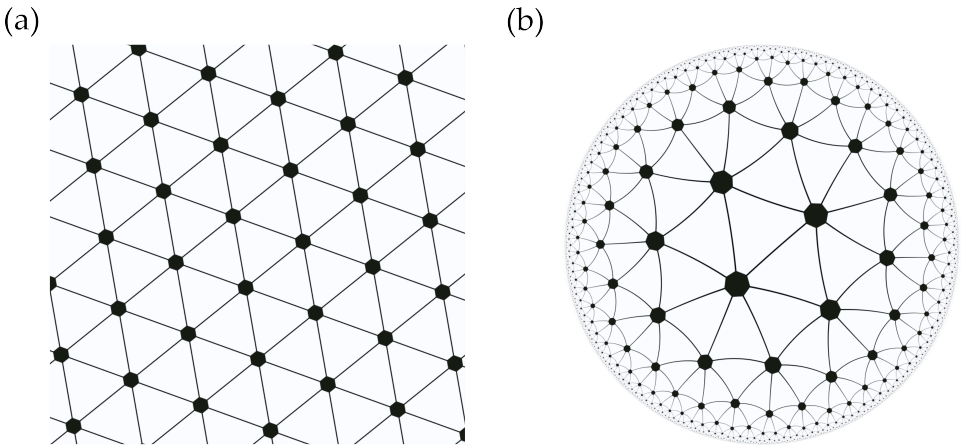



Figure 4.3: The regular tiling of (a) the Euclidean and (b) the hyperbolic plane. Here, "regular" refers to the fact that each triangle occupies the same area on the plane. The fact that the triangles appear to get smaller in (b) is due to the embedding into flat space: As one increases the radius of the disc, more and more triangles "fit" in the hyperbolic space. In the projection this is represented by smaller and smaller triangles.

ally not be homogeneously distributed, but rather show soft community structure, which are compatible with metadata about group membership. For example, the soft communities found for the World Trade Web [53] are highly congruent with geopolitical aspects, placing nodes in the same continent in the same community. In Ref. [54] it was found that the partitions identified in structural brain networks were highly congruent with both neuroanatomical regions as well as functional clusters. The partition based on the soft communities can therefore be considered explicative.

## 4.3 THE $\mathbb{H}^2$-MODEL

In the $\mathbb{S}^1$-model, similarity and popularity are treated differently. The similarity dimension is explicitly geometric, whereas popularity is encoded in hidden degrees linked to the nodes. In Ref. [29] it was shown that these hidden degrees can be mapped to radial coordinates on the hyperbolic plane, leading to a fully geometric model. Before we turn to the details of this mapping, let us shortly recap some properties of the hyperbolic geometry. There are three homogeneous and isotropic[11] spaces; Euclidean (flat), spherical (positively curved) and hyperbolic (negatively curved). Where the first two geometries are intuitive, hyperbolic spaces are difficult to visualize as they cannot be embedded perfectly into flat space. Several models exist, each focusing on different properties of hyperbolic space, but none capturing them all. We choose to work with the *native* representation of

11 Meaning that the geometry is the same in all directions.

the hyperbolic manifold. Here, points on the two dimensional hyperbolic plane $\mathbb{H}^2$ with constant negative curvature $K = -\zeta^2 < 0$ are mapped to a disk with radius $R$ in such a way that radial coordinates in the original space are equal to the radial coordinates on the disk. This allows one to observe an important feature of hyperbolic spaces: They expand faster than their Euclidean counterparts. Consider the expressions for the circumference $L(r)$ and area $A(r)$ of a circle of hyperbolic radius $r$

$$L(r) = 2\pi \sinh(\zeta r), \tag{4.34}$$

$$A(r) = 2\pi \left(\cos(\zeta r) - 1\right). \tag{4.35}$$

Both these quantities grow as $e^{\zeta r}$, much faster than the polynomial growth observed in Euclidean geometry. This property is visualized in Fig. 4.3 and is crucial for understanding why hyperbolic geometry is the natural choice for modeling networks. Let us take a heterogeneous network, and create a strictly nested set of nodes $\tilde{V}_k = \{i | k_i \geq k\}$. This division uncovers a hidden hierarchy in the network based on the node degree[12]. If we now wanted to embed such a network in some space, we could imagine the nodes lying in overlapping discs $D_k$ related to the sets $\tilde{V}_k$ they belong to, all centered at the origin. It might be reasonable to assume that the surface area of the discs are proportional to the amount of nodes contained in the related set, such that $A(D_k) \propto \left|\tilde{V}_k\right|$. One can imagine that each node is located at the center of one of the triangles in the regular tiling in Fig. 4.3b. The sizes of these discs then grow very fast; there are many nodes with low degree and very few with high degrees. In order for $A(D_k)$ to grow sufficiently fast, we need to assume that the space we are embedding our network into is hyperbolic.

In the hyperbolic plane, distances between points are defined by the hyperbolic law of cosines:

$$\cosh(\zeta d_{ij}) = \cosh(\zeta r_i)\cosh(\zeta r_j) - \sinh(\zeta r_i)\sinh(\zeta r_j)\cos(\Delta\theta_{ij}). \tag{4.36}$$

It can be shown that this equation reduces to the simpler form

$$d_{ij} \simeq r_i + r_j + \frac{2}{\zeta}\ln\left(\sin\left(\frac{\Delta\theta_{ij}}{2}\right)\right) \tag{4.37}$$

when $r_i, r_j \gg 1$ and

$$\sin\left(\frac{\Delta\theta_{ij}}{2}\right) \gg \sqrt{e^{-2\zeta r_i} + e^{-2\zeta r_j}}. \tag{4.38}$$

If one also assumes that $\Delta\theta_{ij} \ll 1$ this can further be approximated as

$$d_{ij} \simeq r_i + r_j + \frac{2}{\zeta}\ln\left(\frac{\Delta\theta_{ij}}{2}\right). \tag{4.39}$$

We will comment on the validity of these assumptions later on.

12 Note that this is not the only hidden hierarchy a network can have. One might divide a social network based on nested community membership [185], or a citation network based on the topics of papers [186].

With this new distance function we can again define the energy of a link as $\epsilon = \ln(d_{ij})$, following the same logic as in the $\mathbb{S}^1$-model. Maximizing the entropy under the constraints of a specific expected total energy and amount of links, we once again retrieve the Fermi-Dirac connection probability

$$p_{ij} = \frac{1}{1 + e^{\frac{\beta\zeta}{2}\left(d_{ij} - R_{\mathbb{H}}\right)}}, \tag{4.40}$$

where $R_{\mathbb{H}}$ is the radius of the hyperbolic disk, which in this case plays the role of the chemical potential. In order for both models to be equivalent, this probability must be equal to Eq. (4.26). This is achieved by setting

$$\kappa(r) = \kappa_0 \exp\left(\frac{\beta\zeta}{2\max(1,\beta)}\left(R_{\mathbb{H}^2} - r\right)\right), \tag{4.41}$$

where we have defined the radius such that $\kappa(R_{\mathbb{H}^2}) = \kappa_0$. In this case, the hyperbolic radius is given by

$$R_{\mathbb{H}^2} = \frac{2}{\zeta}\ln\left(\frac{N}{\pi}\right) - \frac{2\max(1,\beta)}{\beta\zeta}\ln(\hat{\mu}\kappa_0^2). \tag{4.42}$$

Eq. (4.41) can be inverted to obtain

$$r(\kappa) = R_{\mathbb{H}^2} - \frac{2\max(1,\beta)}{\beta\zeta}\ln\left(\frac{\kappa}{\kappa_0}\right). \tag{4.43}$$

This relation is in line with the intuition we gained before: Low degree nodes are numerous and must therefore lie at the periphery of the hyperbolic disc, where there is a lot of "space" for them. This relationship also uncovers an intimate relation between scale free networks and the hyperbolic latent geometry. Say $\rho(\kappa)$ is given by the Pareto distribution with exponent $\gamma$. By virtue of Eq. (4.41), the distribution of radial coordinates is then given by $\rho(r) = \alpha\exp(\alpha(r - R_{\mathbb{H}^2}))$, where $\alpha = \frac{\beta\zeta}{2\max(1,\beta)}$. Interestingly, this is approximately the distribution of radial coordinates of nodes when they are *quasiuniform* sprinkled on the hyperbolic disc. When $\alpha = \zeta$, the sprinkling becomes exactly uniform. This is the case when $\gamma = 2\max(1, 1/\beta) + 1$. The fact that this trivial, quasiuniform, node distribution naturally leads to a power law degree distribution further motivates the use of hyperbolic geometry to model heterogeneous networks.

So far we have not remarked on the role of the curvature $K = -\zeta^2$ on the model. Interestingly, in the Fermi-Dirac connection probability, $2/\zeta$ can be interpreted as the Boltzmann constant. Similarly to how setting this constant to one in statistical physics implies a change of units, a change in the curvature in hyperbolic space can always be reabsorbed into a change of length scale. We are thus free to choose $\zeta$ in any way that is convenient. For $\beta > 1$, we choose to set $\zeta = 1$, as is typically done when studying this regime [19]. In the case $\beta < 1$, we set $\zeta = \beta^{-1}$, leading to an infinite negative curvature at $\beta = 0$ [29].

The latter choice has several advantages. First, it is the only definition that leads to a finite hyperbolic radius at $\beta = 0$, which is important as the $\mathbb{H}^2$-model is mostly

used for visualization purposes. Second, it allows for an intuitive interpretation of the hyperbolic distance defined in Eq. 4.39. As $\beta \to 0$, the dependence on the angular distance $\Delta\theta_{ij}$ vanishes, which is in line with the fact that this limit corresponds to the hyper-soft configuration model, where only the popularity dimension plays a role. If the curvature were set to some other value, the typical length scale in the popularity dimension would diverge whereas it would remain constant in the similarity dimension, effectively leading to the same situation.

### 4.3.1 *The equivalence of the $\mathbb{S}^1$ and $\mathbb{H}^2$ models*

In order to derive Eq. (4.41), leading to the equivalence of the two models, one must first show that Eq. (4.39) is a good approximation of Eq. (4.36). Hence, let us now return to the assumptions that were necessary for this approximation to hold.

First, we know that the smallest radial coordinate is related to the node with the largest expected degree. For homogeneous networks, $\kappa_{\text{max}} = \langle k \rangle \sim N^0$, and so according to Eq. (4.41), $r_{\text{min}} \simeq 2 \ln N$. For scale-free networks, $\kappa_{\text{max}} \sim N^{1/(\gamma-1)}$, such that $r_{\text{min}} \simeq 2\frac{\gamma-2}{\gamma-1} \ln N$. In both cases $r_i \gg 1 \; \forall i \in V$ as $N \to \infty$. This motivates the first assumption needed for Eq. (4.37).

Turning now to Eq. (4.38), we note that if the assumption holds for $r_i = r_j = r_{\text{min}}$ it must necessarily hold for all $r_i, r_j$. Plugging in the scaling of the minimum radius, the assumption reduces to $\sin\left(\Delta\theta_{ij}/2\right) \gg \sqrt{2}N^{-\xi\zeta}$, where $\xi$ depends on the specific degree distribution. In the domain of $\Delta\theta_{ij}$, the sine is a monotonously growing function, and so we can study the smallest angular separation, where the inequality reduces to $\Delta\theta_{ij} \gg \frac{1}{2}\sqrt{2}N^{-\xi\zeta}$. The expected minimal distance between two consecutive nodes goes as $N^{-1}$, and so the inequality holds for all nodes when $\xi > 1/\zeta$.

Now, for the homogeneous case, $\xi = 2$, and the inequality thus holds for all $\zeta > 1/2$. In the scale-free case, $\xi = 2(\gamma-2)/(\gamma-1)$. This then leads to the inequality $\gamma > (4\zeta-1)/(2\zeta-1)$. For $\beta \geq 1$, $\zeta = 1$, and the inequality reduces to $\gamma > 3$. In this range, our assumption holds for all pairs of nodes. When $\gamma \leq 3$, the amount of nodes that fall into the range $\Delta\theta_{ij} < \sqrt{2}N^{-\xi\zeta}$ scales as $\sim N^{-\xi\zeta}/N^{-1} = N^{\frac{3-\gamma}{\gamma-1}}$, which is vanishing fraction of the total amount of nodes. Consequently, the number of pairs of nodes for which the approximation does not hold is a vanishing fraction in the thermodynamic limit. When $\beta < 1$, we argued that a convenient choice for the curvature was $\zeta = 1/\beta$. When $\beta = 0$, we then have $\gamma > 2$, i.e., our approximation is always valid. For intermediate $\beta$ we then expect the approximation to break down for a microscopic fraction of nodes for some $\gamma$ between two and three.

The final assumption $\Delta\theta_{ij} \ll 1$ clearly does not hold for all node pairs. However, the angular distance between two nodes at most scales as a constant, whereas $r_i$ and $r_j$ scale as $\ln N$. For the pairs of nodes where the assumption does not hold, the correction to $d_{ij}$ coming from the angular separation is therefore irrelevant in both Eq. (4.37) and Eq. (4.39). This implies that they can be used interchangeably also when $\Delta\theta_{ij}$ is not small.

## 4.4 OTHER GEOMETRIC MODELS

In this Chapter, we have focused on the $\mathbb{S}^1$-model and its fully geometric counterpart, the $\mathbb{H}^2$-model. However, many other geometric models exist [20, 113]. Here we will mention two other important models that, together with the $\mathbb{S}^1$-model, can be seen as specific examples of the subclass of geometric random graphs whose connection probabilities are a general function of $\chi_{ij} \propto d_{ij}/(\kappa_i\kappa_j)$, first introduced in Ref. [30].

The first model was introduced for lattices in Ref. [187] and generalized to continuous (Euclidean) similarity space in Ref. [188] and is often referred to as the Scale Free Percolation model. Here, the connection probability is modified to be

$$p_{ij} = 1 - \exp\left(-\left(\frac{(\hat{\mu}\kappa_i\kappa_j)}{d_{ij}}\right)^{\beta}\right). \tag{4.44}$$

This form is also a special case of the connection probability derived by Garuccio *et al.* in Ref. [189]. In this reference the authors show that this functional form is the only one that leads to networks that are exactly scale invariant under renormalization. This advantage comes at the cost of no longer generating maximally random ensembles. We will come back to these points in Chapter 7.

The second model is that of the Geometric Inhomogenous Random Graph (GIRG) [175, 176]. Here, the connection probability is once again modified and reads

$$p_{ij} = \Theta\left(\min\left(1, \frac{(\hat{\mu}\kappa_i\kappa_j)^{\max(1,\beta)}}{d_{ij}^{\beta}}\right)\right), \tag{4.45}$$

where the big $\Theta(g(N))$ notation implies that $p_{ij}$ scales with the system size in the same way as $g(N)$. Sometimes GIRGs are defined without this generalization, and are simply taken to be exactly $g(N)$ [113, 190].

Intuitively, all three connection probabilities (Eqs. (4.26),(4.44) and (4.45)) are related. For one, when $\hat{\mu}\kappa_i\kappa_j \ll d_{ij}$, all three are approximately equivalent, at least for $\beta \geq 1$[13]. Of course, this will not be the case for all node pairs. Other, more rigorous, relations between the different models can be established as well. For example, it has been shown that the $\mathbb{H}^2$ model is a special case of the GIRG model defined with Eq. (4.45) [176]. In Refs. [191, 192] the GIRG and SFP models are compared. It can be shown that they are comparable with respect to many important properties such as clustering and the small world property [181, 193].

All in all, which connection probability is chosen depends on what feature is deemed most important. For example, GIRGs are the most easy to work with analytically. However, if one wants explicit self-similarity, using the SFP connection probability is the way to go. Finally, if a maximally random ensemble is required, one needs to turn to the $\mathbb{S}^1/\mathbb{H}^2$-models.

13 For $\beta > 1$, all three connection probabilities are a function of $\chi_{ij} = (\hat{\mu}\kappa_i\kappa_j/d_{ij})^{\beta}$. When $\chi_{ij} \ll 1$, $p_{ij}^{(\mathbb{S}^1)} = \chi_{ij}/(1+\chi_{ij}) \approx \chi_{ij}$. The same goes for Eqs. (4.44) and (4.45).

## 4.5 NETWORK EMBEDDING

The hypothesis that the connectivity of a network arises from the location of the nodes in some underlying metric space has been able to explain many of the key properties observed in real networks. The natural next step is then to ask whether the node coordinates of real networks can be extracted from their connectivity. Finding such *network embeddings* can shed new light on the relations between the nodes and explain how the network was formed. They also allow for an appealing way of visualizing the network, and can help us analyze its properties.

The task of finding a meaningful mapping between the network topology and its latent space is not trivial. The most conceptually intuitive model-based methods are based on *maximum likelihood estimation* (MLE) [28, 194, 195]. Here, we ask the following question: Given a set of hidden coordinates $\{\kappa_i, \theta_i\}_{i \in V}$, how likely is it that we produce the observed network with the model? One can then find the set of hidden coordinates that maximizes this likelihood function. Unfortunately, this optimization problem is, in general, NP-hard and one must therefore rely on heuristics to obtain reasonable approximate solutions. Hence, MLE techniques are often slow, and their accuracy depends strongly on the chosen estimation method.

Another option to embed networks is using dimension reduction techniques, which are often much faster than MLE [196, 197]. Here, we assume nodes lie in some high dimensional space, and the idea is to find a proper mapping to a low dimensional space such that the distances between points are preserved. One way of doing so is given by Laplacian Eigenmaps, a technique designed for data embedded in high dimensional Euclidean space [198]. The problem here is that our network is *a priori* not embedded in any high dimensional real space, and that we therefore do not know anything about the distances between the nodes. Network embedding techniques based on Laplacian Eigenmaps therefore require the use of heuristics to estimate these distances. Another problem is that these methods embed the network into Euclidean space, implying that they can only be used for the similarity space. Other methods must therefore be employed if one also wants to take into account the popularity dimension, or directly embed into hyperbolic space.

Finally, there is a relatively large body of literature from the computer science community which uses neural networks [199–201]. These methods are generally fast, but difficult to interpret and their setup is more complicated in comparison to the other techniques mentioned in this section. They have the added benefit that node attributes[14] can be included in the embedding procedure.

In this thesis we generally make use of a specific embedding tool which makes use of both MLE and Laplacian Eigenmaps; *Mercator* is a ready to use C++ code to embed real networks into the $\mathbb{S}^1/\mathbb{H}^2$-model [61, 64]. In the following we will give a short overview of the most important steps. Note that this method was initially

14 This is additional information about the nodes that is not directly included in the network topology. Think for example of the age of a person, the continent a country lies in or the academic position of a researcher.

only designed for $\beta > 1$, i.e. for networks with a strong geometric coupling. The extension to the weakly geometric $\beta \leq 1$ regime is developed in Chapter 6.

First, Mercator chooses a random geometric coupling $\beta$ and sets $\hat{\mu}$ according to Eq. (4.20). The next step is to infer the hidden degrees $\{\kappa_i\}_{i \in V}$. *A priori*, all nodes with the same degree are assumed to be equivalent. This allows us to work only with degree classes, not individual nodes. The goal is then to set the hidden degrees such that, for all degree classes, $|\bar{k}(\kappa(k)) - k| < \epsilon$, where $\epsilon$ is some arbitrary small value. To be able to estimate $\bar{k}(\kappa(k))$ given the model, we assume that nodes are distributed homogeneously.

In the previous step, the hidden degrees sequence of the network was inferred for some arbitrary $\beta$. Next, this inverse temperature is refined. This is done by calculating the expected mean local clustering of the network, given the hidden degrees inferred in the previous step as well as the geometric coupling $\beta$. If this clustering is lower than the observed clustering, $\beta$ needs to be increased. If is higher, $\beta$ should decrease. After adjusting $\beta$ we return to the previous step and infer the hidden degrees. This process continues until the expected level of clustering matches the real value.

Now that the popularity dimension has been sorted and the proper geometric coupling has been inferred, we turn to the similarity dimension. This is where Laplacian Eigenmaps come in. As mentioned earlier, this method was designed for dimensional reduction of data. Here, each point $i$ has a coordinate $\boldsymbol{x}_i \in \mathbb{R}^l$, and the goal is to find new coordinates $\boldsymbol{y}_i \in \mathbb{R}^m$ such that $m < l$. Normally, the first step in this procedure is to constructs a graph, for example by connecting nodes that lie closer than some distance $r$ to one another. The links are then weighted through some weighting function $W_{ij} = A_{ij} f(||\boldsymbol{x}_i - \boldsymbol{x}_j||)$, where $A_{ij}$ is the adjacency matrix of the constructed graph and $f(||\boldsymbol{x}_i - \boldsymbol{x}_j||)$ is some decreasing function. The idea is then to find the coordinates $\{\boldsymbol{y}_i\}_{i=1}^N$ that minimize the weighted distance function

$$\epsilon = \sum_{i<j} ||\boldsymbol{y_i} - \boldsymbol{y_j}||^2 W_{ij}, \tag{4.46}$$

i.e., that ensure that connected points stay close together. We first solve the generalized eigenvalue equation $\boldsymbol{Lv} = \boldsymbol{Dv}$, where $D_{ij}$ is the weighted degree matrix $D_{ij} = \delta_{ij} \sum_k W_{ik}$ and $L$ is the weighted Laplacian matrix $L_{ij} = W_{ij} - D_{ij}$. It can then be shown that the desired coordinates of node $i$ are obtained by taking the $i$'th component of the first $m$ generalized eigenvector, ordered by their eigenvalues. Note that the zeroth eigenvector is trivial, and therefore ignored.

In the setting of network embeddings, some adjustments need to be made to this process. First, we do not need to construct a graph as we already have one. Second, this dimensional reduction leads to coordinates in $\mathbb{R}^m$, when our similarity space is given by $\mathbb{S}^1$. In our case we will take $m = 2$ and define

$$\theta_i = \text{atan2}(y_i^1, y_i^2) \tag{4.47}$$

and disregard the radial coordinate associated to the node. Finally, and most problematically, we do not actually have access to the original coordinates $\{\boldsymbol{x}_i\}_{i \in V}$,

which means $W_{ij}$ needs to be estimated. This is done by calculating the expected angular distance $\langle \Delta\theta_{ij} \rangle$ between two nodes conditioned on their inferred degrees and the fact that they are connected. We then estimate the distance between the points as

$$||\boldsymbol{x}_i - \boldsymbol{x}_j|| = 2 \sin\left(\frac{\langle \Delta\theta_{ij} \rangle}{2}\right), \tag{4.48}$$

the expected chord length, and define the weighting function as $f(||\boldsymbol{x}_i - \boldsymbol{x}_j||) = \exp\left(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\right)$. Next, we find the generalized eigenvalues of the weighted Laplacian as described above and obtain a set of hidden coordinates $\{\theta_i\}_{i \in V}$. We can further improve this result by making order-preserving adjustments, where we fix the ordering of the coordinates but make sure the gap between consecutive nodes is in line with expected gap given the model. This then concludes the first "fast" version of Mercator, which has time complexity $\mathcal{O}(N)$. This version can already accurately reproduce the structural properties of the original network.

Mercator also provides the option of further refining the results given by the fast mode. This refinement is based on maximizing the likelihood. The idea is to perturb all node coordinates in an ordered way, and choose the coordinates that maximize the likelihood that the observed network was generated by the model. Where starting with this step was impossible due to the vast configuration space of the problem, the coordinates in the previous step provide us with an initial guess that lies very close to the optimal configuration. This step is still rather slow, though, and increases the time complexity to $\mathcal{O}(N^2)$.

The final step refines the set of hidden degrees $\{\kappa_i\}_{i \in V}$ found in the first step. There, we did not use any information about the angular coordinates of the nodes. Now that we have obtained a set of angular coordinates we can repeat the first step, taking into account the inferred $\theta_i$ of each node.

The code of the Mercator embedding tool is publicly available at https://github.com/networkgeometry/mercator

# Part II

# STRUCTURAL PROPERTIES

# 5

# THE CLUSTERING PHASE TRANSITION

In Chapter 4, we saw that in geometric network models the triangle inequality induces topological triangles in the associated graph. We also saw that including long range connections weakens the coupling between the geometry and the topology. In the $\mathbb{S}^1$-model, this coupling strength is encoded by the inverse temperature $\beta$, and it has been shown that non-geometric and non-clustered SCM graphs are obtained when $\beta = 0$. In fact, in the thermodynamic limit, clustering vanishes in the entire region $\beta \leq 1$, where the geometric coupling is so weak that long-ranged connections dominate. In this Chapter, which is an adaptation of Ref. [35], we study the phase transition between the strong and weak coupling regimes.

This Chapter is organized as follows: First, by mapping the network to a gas of fermions, in Sec. 5.1 we are able to analytically derive the thermodynamic properties of the system, revealing a diverging entropy density at the critical point. In Sec. 5.2 we argue that the transition is topological in nature, driven by the reorganization of chordless cycles. Finally, in Sec. 5.3, we uncover interesting, atypical finite size scaling behavior as compared with standard continuous phase transitions, where one observes a power law decay at the critical point and a faster decay in the disordered phase. Instead, at the critical point, the average local clustering coefficient decays logarithmically to zero for very large systems and, in the weakly geometric phase, where the coefficient decays as a power law, we discover a *quasi-geometric* region where the exponent that characterizes this decay depends on the temperature.

## 5.1 STATISTICAL PROPERTIES

### 5.1.1 *The networks as a gas of fermions*

The first question we want to ask is whether the transition in the local properties (the presence of triangles attached to nodes) affects the global behavior of the system (codified by the thermodynamic properties, specifically the entropy). To this end, we dig deeper into the mapping between the $\mathbb{S}^1$-network ensemble and that of a gas of non-interacting fermions [29, 116].

In the homogeneous case, we saw in Sec. 4.2.1 that the connection probability could be written in the Fermi-Dirac form, where the energy of a state was given $\epsilon_{ij} = \ln d_{ij}$. The fact that our networks are simple, reflecting the Pauli exclusion

principle, and that edges are unlabeled, leading to indistinguishable particles, completes the mapping between the two systems.

For heterogeneous networks, the hidden degrees $\{\kappa_i\}_{i\in V}$ complicate the situation somewhat. While for $\beta \geq 1$ the energy of a state still takes a standard form, which is independent of the Lagrange multiplier $\beta$, for $\beta < 1$ the energy levels become temperature dependent. This fact means we need to tread with care when working in this regime.

As an example of how this picture can help us, let us study the amount of edges/particles in the system. For the $\mathbb{S}^1$-model, this is given by

$$\langle M \rangle = \sum_{i<j} p_{ij} = \sum_{i<j} \frac{1}{1+e^{\beta(\epsilon_{ij}-\mu)}}, \tag{5.1}$$

a fact we have used various times in this work. Similar to what was done in Sec. 3.3.5, we now introduce the following measure

$$\rho_N(\epsilon) = \frac{1}{\binom{N}{2}} \sum_{i<j} \delta(\epsilon - \epsilon_{ij}), \tag{5.2}$$

which is normalized as there are $\binom{N}{2}$ states in the system (all node pairs). This allows us to rewrite Eq. (5.1) as

$$\langle M \rangle = \binom{N}{2} \int \frac{\mathrm{d}\epsilon \rho_N(\epsilon)}{1+e^{\beta(\epsilon-\mu)}}. \tag{5.3}$$

The goal is now to find a closed form expression for $\rho_N(\epsilon) \to \rho(\epsilon)$ that holds when $N \gg 1$, which we we will call the *density of states*. We know that $\epsilon_{ij} = \epsilon(\theta_i, \theta_j, \kappa_i, \kappa_j)$, and that these hidden coordinates were all drawn from nice probability distributions[1]. That is to say, we know that $\rho_N(\theta) \to \frac{1}{2\pi}$ and that $\rho_N(\kappa) \to \rho(\kappa)$ when $N \gg 1$. Using this fact, we write the density of states as

$$\rho(\epsilon) = \iint \mathrm{d}\kappa' \mathrm{d}\kappa'' \frac{\rho(\kappa')\rho(\kappa'')}{(2\pi)^2} \iint \mathrm{d}\theta' \mathrm{d}\theta'' \delta\left(\epsilon - \epsilon(\theta', \theta'', \kappa', \kappa'')\right) \tag{5.4}$$

Without loss of generality, the spherical symmetry of the system can be employed to set $\theta' = 0$. Rewriting the Dirac delta function and integrating over $\theta''$, this leads to

$$\rho(\epsilon) = \frac{2e^{\epsilon}}{N} \iint \mathrm{d}\kappa' \mathrm{d}\kappa'' \rho(\kappa')\rho(\kappa'')(\kappa'\kappa'')^{\frac{1}{\min(1,\beta)}} \Theta\left(1 - \frac{2e^{\epsilon}}{N}(\kappa'\kappa'')^{\frac{1}{\min(1,\beta)}}\right), \tag{5.5}$$

where $\Theta(.)$ denotes the Heaviside step function.

Without knowing the exact form of the hidden degree distribution, this is as far as we can go. In the following, we will work with this form as it is the most general. However, obtaining analytic results requires the assumption that

1 We are working here with *quenched* disorder; the same result for the density of states would be obtained were we to assume that the hidden variables were constantly redrawn from the distribution. In that case we would simply be able to marginalize Eq. (5.1) over $\epsilon_{ij}$.

$\frac{N}{2(\kappa'\kappa'')^{\max(1,1/\beta)}} \gg 1$, which does not always hold. In particular, when the degree distribution is power law and the network contains structural correlations ($\gamma \leq 3$) [82], this approximation will break down for the hubs. However, for these degrees $\rho(\kappa'), \rho(\kappa'') \ll 1$, implying that we might assume the contribution of any error arising from this procedure to vanish. In fact, the standard derivation of the hidden variable $\hat{\mu}$ in the thermodynamic limit, leading to sparse networks, relies on this assumption.

In this thesis we check the validity of any such results for the case of Pareto distributed hidden degrees. Assuming this distribution allows us to work out Eq. (5.5) further, leading to

$$\rho(\epsilon) = \frac{2a}{N} e^{\epsilon} \Theta\left(\epsilon_{\max} - \epsilon\right) \left(1 + e^{b(\epsilon_{\max} - \epsilon)} \left(b(\epsilon_{\max} - \epsilon) - 1\right)\right), \tag{5.6}$$

where $a = \min(1,\beta)\kappa_0^{2\max(1,1/\beta)}(\gamma-1)^2/(1+\min(1,\beta)(1-\gamma))^2$ and $b = 1 + \min(1,\beta)(1-\gamma)$. Note that $\epsilon_{\max} = \ln\left(\frac{N}{2\kappa_0^{2(\max(1,1/\beta))}}\right)$ is given by the energy of a pair of nodes exactly opposite one another on the $\mathbb{S}^1$-circle such that $\Delta\theta_{ij} = \pi$.

This now allows us to solve Eq. (5.3) exactly for $N \gg 1$. In Appendix A.1.1 it is shown that this leads to the same results for $\mu$ derived in Sec. 4.2.2 when $\langle M \rangle = \frac{N\langle k \rangle}{2}$. In the same Appendix we also verify all the thermodynamic results obtained in the following section.

### 5.1.2 *The exponential density of states*

An important observation is that the density of states is exponentially distributed, i.e., $\rho(\epsilon) \propto e^{\epsilon}$. This is a rather unusual form, as, in general, physical systems exhibit a power law density of states. For example, in the case of a 3 dimensional Fermi gas, the density of states goes as $\rho(\epsilon) \propto \sqrt{\epsilon - \epsilon_0}$. However, the exponential density of states makes an appearance in various fields of physics such as string theory [202, 203], nuclear physics [204], organic semi-conductors [205–207] and finally high energy physics.

In this latter field, the exponential density of states arises in the theory of hadrons. These subatomic particles, of which protons and neutrons are examples, were once thought of as the fundamental building blocks of matter. Of course, now we know they are comprised of quarks. However, in the first half of the twentieth century, this was unknown. Hadrons were assumed to be point particles, and it was observed that in high energy collision many different hadron resonances, with ever higher masses, were formed [208]. The explanation of this phenomenon came from Rolf Hagedorn in 1965 [209, 210], who proposed a thermodynamic interpretation. He postulated that the fundamental feature of the dynamics was new resonance formation, which allowed him to treat the hadrons, or "fireball" as he called them, as a system of non-interacting particles. Hagedorn's idea was then that each hadron was actually composed of smaller ones, while also being a

constituent of a larger hadron. This idea leads to a self-consistency relation for the density of states

$$\rho(m) = \delta(m - m_0) + \sum_{N=2}^{\infty} \frac{1}{N!} \int \prod_{i=1}^{N} \mathrm{d}m_i \rho(m_i) \delta \left( \sum_i m_i - m \right). \tag{5.7}$$

This equation, here shown in in a simplified, toy-model form, is also known as the *bootstrap equation* [211–213] and gives Hagedorn's model for hadron formation, the *Statistical Bootstrap Model (SBM)*, its name. The idea here is that the amount of resonances of mass $m$ is a result of the amount of all masses that can be combined to form a particle of mass $m$ (of course, because we are working at very high energies, mass and energy can be treated on the same footing). Solving this leads to an exponential density of states $\rho(m) \propto e^{m/T_H}$, where the temperature $T_H$ is called the *Hagendorn temperature.* A very intuitive explanation for this result was given in Ref. [214], where a model for ordered partitions of integers was introduced. The idea here is that each integer $n$ can be written as the sum of smaller integers. For example, $n = 3$ can be written as $3$, $2 + 1$, $1 + 2$ and $1 + 1 + 1$. Of course, the amount of partitions gets larger with $n$, and it can be shown that this growth is exponential, just like in the SBM; if I give you an energy of $m$ and $m$ is large, it is clear that there are many ways one can construct a resonance with that energy, leading to a high degeneracy and a high $\rho(m)$.

We can use the result to give an expression for the partition function,

$$\ln \mathcal{Z}(T, V) \sim \int_0^{\infty} \mathrm{d}m g(m, T, V) e^{-m\left(\frac{1}{T} - \frac{1}{T_H}\right)}, \tag{5.8}$$

where $g(m, T, V)$ is some function arising from the specifics of the physical system that are not important for this discussion. We note that the exponential density of states renders this integral infinite when $T > T_H$. Thus, $T_H$ is the ultimate temperature for hadronic matter. This can be explained as follows: Normally, when more energy is added to a system, the momentum of its constituents rise, leading to a higher temperature. Here, there is a competing process where added energy is used for producing ever heavier resonances. Eq. (5.8) tells us that as one approaches $T_H$, more and more energy is used for this second process, until finally no more energy is converted into momentum and the system remains at a fixed temperature. As more and more different particles are being formed, the entropy of the system does keep rising. Therefore, the entropy diverges with the energy, in this case at finite temperature.

Does this mean there exists an absolute maximum temperature in the universe? No, because Hagedorn's model assumes that hadrons are pointlike elemental particles. Of course, we know that this is not the case, but that they are rather made up of quarks and gluons. A subsequent model, the *bag model* [215, 216], explored this further, leading to the interpretation that at $T_H$, the system becomes deconfined, i.e., it marks the onset of a free quark-gluon plasma phase. This temperature is roughly 150MeV, approximately the mass of the lightest hadron, the pion [210, 217].

### 5.1.3 *The divergence of the entropy*

Clearly, an exponential density of states can lead to very interesting physical behavior. In our case, despite the fact that links in the model are noninteracting particles, the system undergoes a continuous phase transition at a critical temperature $T_c = \beta_c^{-1} = 1$. We can analyze the nature of the transition by studying the entropy of the ensemble. Given the mapping of the $\mathbb{S}^1$-model to a system of non-interacting fermions in the grand canonical ensemble, we start from the grand canonical partition function

$$\ln \mathcal{Z} = \sum_{i<j} \ln\left(1 + e^{-\beta(\mu-\epsilon)}\right). \tag{5.9}$$

Assuming that $N$ is large, we can employ the density of states given in Eq. (5.5) to obtain

$$\begin{aligned} \ln \mathcal{Z} &= \binom{N}{2} \int_{-\infty}^{\infty} \mathrm{d}\epsilon \rho(\epsilon) \ln\left(1 + e^{-\beta(\mu-\epsilon)}\right) && (5.10)\\ &= \binom{N}{2} \iint \mathrm{d}\kappa' \mathrm{d}\kappa'' \, \rho(\kappa')\rho(\kappa'') \Bigg[\beta + \ln\left(1 + \frac{1}{\xi}\right) \\ &\qquad - \frac{\xi\beta}{1+\beta} {}_2F_1\left(\begin{matrix} 1, 1+\frac{1}{\beta} \\ 2+\frac{1}{\beta} \end{matrix}; -\xi\right)\Bigg], && (5.11) \end{aligned}$$

where $\xi = \left(\frac{N}{2\hat{\mu}}\right)^{\beta} (\kappa'\kappa'')^{-\max(1,\beta)}$ and $\hat{\mu} = e^{\mu}$. We now focus on the region of $\beta > \beta_c = 1$, and assume that $\xi \gg 1$[2]. The grand partition function then reduces to

$$\ln \mathcal{Z} = N \frac{\hat{\mu}\pi\langle k\rangle^2}{\sin\left(\frac{\pi}{\beta}\right)}. \tag{5.12}$$

We can then use the above expression to find the grand potential $\Xi = -\beta^{-1} \ln \mathcal{Z}$ and the entropy as $S = \beta^2 (\frac{\partial \Xi}{\partial \beta})_{\mu}$. From this, we can find the entropy per link of the system as

$$\frac{S}{\langle M\rangle} = \beta - \pi \cot\frac{\pi}{\beta} \overset{\beta\to\beta_c^+}{\sim} \frac{1}{\beta - 1}, \tag{5.13}$$

where in the last step $\hat{\mu}$ was plugged in. Note that $\langle M\rangle = N\langle k\rangle/2$ is the number of links –and so particles– in the network. Interestingly, the entropy density is only a function of $\beta$, and so independent of the degree distribution.

From Eq. (5.13), we see that the entropy per link diverges at the critical temperature $\beta \to \beta_c^+ = 1$. This implies that there is a sudden change in the behavior of the system at the critical point $\beta = \beta_c$, which could indicate the presence of a phase transition. This transition is, however, anomalous –at odds with the

2 As mentioned in the previous section, this assumption might break down for for heterogeneous networks. In Appendix A.1 we redo this calculation for the specific case of Pareto distributed hidden degrees and find the same results, implying that this assumption seems to hold quite generally.

continuous entropy density usually observed in continuous phase transitions– and thus cannot be described by Landau's symmetry-breaking theory of continuous phase transitions.

In the regime $\beta < \beta_c$, one needs to be careful in applying the standard thermodynamic laws because of the non-standard temperature-dependent energy levels in this regime. In particular, we need to check the relation between $S$ and $\beta^2 \frac{\partial \Xi}{\partial \beta}$. Starting from Eq. (5.10), the latter can be written as

$$\beta^2 \left( \frac{\partial \Xi}{\partial \beta} \right)_\mu = \int_{-\infty}^{\infty} \mathrm{d}\epsilon \rho(\epsilon) \left( \ln \left( 1 + e^{-\beta(\epsilon-\mu)} \right) + \frac{\beta(\epsilon-\mu)}{1+e^{\beta(\epsilon-\mu)}} \right) + \Delta, \tag{5.14}$$

where

$$\Delta = -\beta \int_{-\infty}^{\infty} \mathrm{d}\epsilon \frac{\partial \rho(\epsilon)}{\partial \beta} \ln \left( 1 + e^{-\beta(\epsilon-\mu)} \right). \tag{5.15}$$

The first line in Eq. (5.14) can be recognized as $-\beta(\Xi + \langle E \rangle - \mu \langle M \rangle)$, whereas $\Delta$ is a direct consequence of the temperature dependence of the energy levels. It can then be shown that we can rewrite Eq. (5.14) as

$$\beta^2 \left( \frac{\partial \Xi}{\partial \beta} \right)_\mu = \int_{-\infty}^{\infty} \mathrm{d}\epsilon \rho(\epsilon) \left( p(\epsilon) \ln p(\epsilon) + (1-p(\epsilon)) \ln(1-p(\epsilon)) \right) + \Delta, \tag{5.16}$$

where we recognize the standard functional form of the Shannon entropy for a graph[3], leading to the final relation

$$S = \beta^2 \left( \frac{\partial \Xi}{\partial \beta} \right)_\mu - \Delta. \tag{5.17}$$

This, *a posteriori*, justifies the use of the relation $S = \beta^2 \frac{\partial \Xi}{\partial \beta}$ when $\beta > 1$, as $\Delta$ is zero in this regime. However, for $\beta < 1$, the energy levels are temperature dependent and $\Delta$ is non-vanishing. Starting once again from Eq. (5.11) and using $\xi \gg 1$, we obtain

$$\frac{S}{\langle M \rangle} = \ln N + \frac{2\beta - 1}{1-\beta} - \ln(1-\beta) + \ln\langle k \rangle - 2 \frac{\langle \kappa \ln \kappa \rangle}{\langle k \rangle}. \tag{5.18}$$

Note that for $\rho(\kappa) = \delta(\kappa - \kappa_0)$ one finds $\langle \kappa \ln \kappa \rangle = \langle k \rangle \ln \langle k \rangle$, whereas for $\rho(\kappa) = (\gamma - 1)\kappa_0^{\gamma-1} \kappa^{-\gamma}$ we obtain $\langle \kappa \ln \kappa \rangle / \langle k \rangle = (\gamma - 2)^{-1} + \ln \kappa_0$ when $\gamma > 2$. In both cases, the leading term of the entropy density is of order $\sim \ln N$, which diverges as $N \to \infty$. Also note that when $\beta \to 0$ we retrieve the entropy density found for the SCM derived in Ref. [142]. For $\beta = 1$, similar steps can be performed, leading to

$$\frac{S}{\langle M \rangle} = \frac{1}{2} \ln N + \ln \ln N + 1 + \ln\langle k \rangle - 2 \frac{\langle \kappa \ln \kappa \rangle}{\langle k \rangle}, \tag{5.19}$$

where we once again see that the divergence is logarithmic in $N$.

3 This is the entropy we are interested in this thesis. Note, however, that one could argue that $\Delta$ should actually be included into the definition of the entropy, as the temperature dependence of the energy levels arises from an unknown interaction between the heat bath and the system. Including this interaction energy leads to Hill's generalized thermodynamic theory [166, 218]

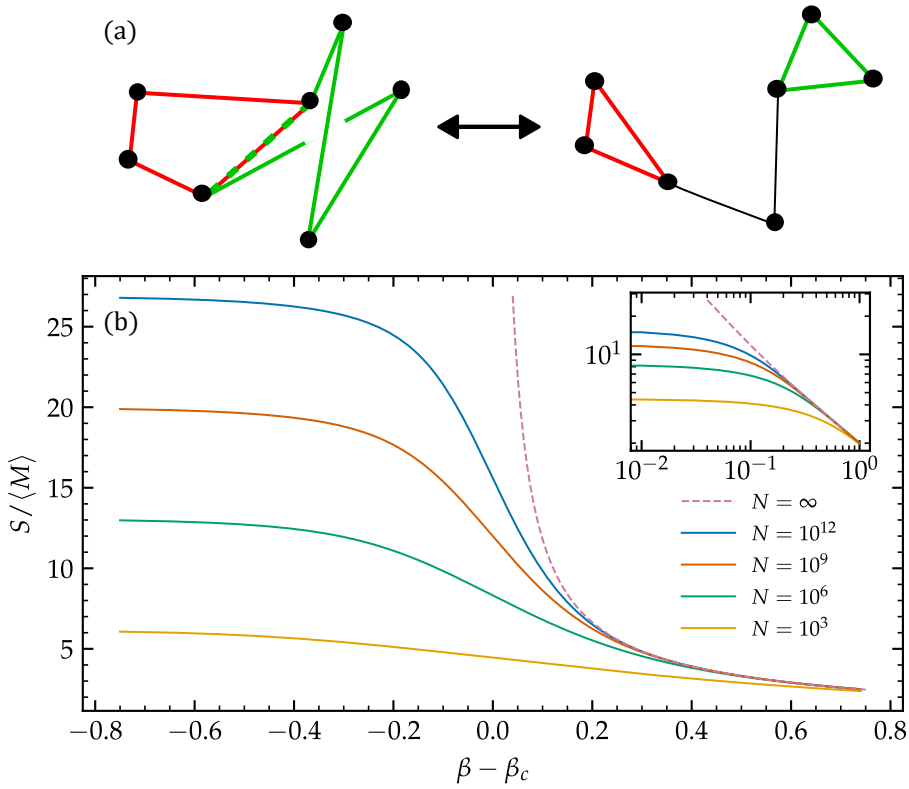



Figure 5.1: Behavior of the network model as one crosses the critical temperature, both from an entropic as topological perspective. (a) Sketch illustrating the different organization of cycles in the two phases, short-range at low temperatures and long-range –of the order of the network diameter– in the high temperature regime. (b) Entropy per link for $S^1$ geometric networks of different sizes with homogeneous degrees. Different curves are obtained by numerical integration. The inset shows the same curves in the region $\beta > \beta_c$ in logarithmic scale.

Figure 9.1 shows a numerical evaluation of the entropy for different system sizes in homogeneous networks confirming the divergence of the entropy per link at the critical temperature as predicted by our analysis. As expected, the entropy diverges logarithmically with the system size at $\beta = \beta_c$, implying that the divergence can only be detected for very large systems.

## 5.2 THE TOPOLOGICAL NATURE OF THE TRANSITION

For many years, Landau's theory of symmetry breaking was believed to be the ultimate explanation of continuous phase transitions [219]. In the liquid-crystal transition, for instance, the continuous translational and rotational symmetry at high temperatures break into a set of discrete symmetries in the low temperature phase. This paradigm was challenged by Berezinskii, Kosterlitz, and Thouless (BKT) in the two dimensional XY model [220–222]. For this model, the Mermin-Wanger theorem [223] states that there is no ordered phase even at zero temperature, so that a phase transition in Landau's sense cannot exist. Yet, BKT showed that, in fact, there is a finite temperature phase transition driven by topological defects: vortices and antivortices. At low temperature, vortex-antivortex pairs are bound together. Above the critical temperature, vortex-antivortex pairs unbind, moving freely on the surface. No symmetry is broken in the transition since both phases are rotationally invariant and so magnetization is zero in both phases. Topological order and topological phase transitions are nowadays fundamental to understand the properties of quantum matter [224].

In this section, we argue that similar observations can be made for the transition discussed in this Chapter, thus making it topological in nature. Notice first that the $\mathbb{S}^1$-model is also rotationally invariant both above and below the critical temperature, which implies that there is no symmetry breaking at the critical point. In fact, we argue that $\beta_c$ separates two distinct phases with different organization of the cycles, or topological defects, in the network. Indeed, the cycle space of an undirected network with $N$ nodes, $M$ links, and $N_{\text{com}}$ connected components is a vector space of dimension $M - N + N_{\text{com}}$ [225]. This dimension is also the number of independent chordless cycles in the network as they form a complete basis of the cycle space. We are typically interested in connected or quasi-connected networks, with a giant connected component extending almost to the entire network. In the $\mathbb{S}^1$-model this is achieved in the percolated phase when the average degree is sufficiently high, but still in the sparse regime, so that the vast majority of cycles are contained in the giant component. In this case, by changing temperature without changing the degree distribution, the number of nodes, links, and components remain almost invariant and so does the number of chordless cycles. Thus, the two different phases correspond to a different arrangement of the chordless cycles of the network, as illustrated in the sketch in Fig. 9.1a. This is again similar to the BKT transition since the number of vortices and antivortices is preserved in both phases. We notice, however, that the exact preservation of the number of cycles is not a necessary condition for the transition to take place.

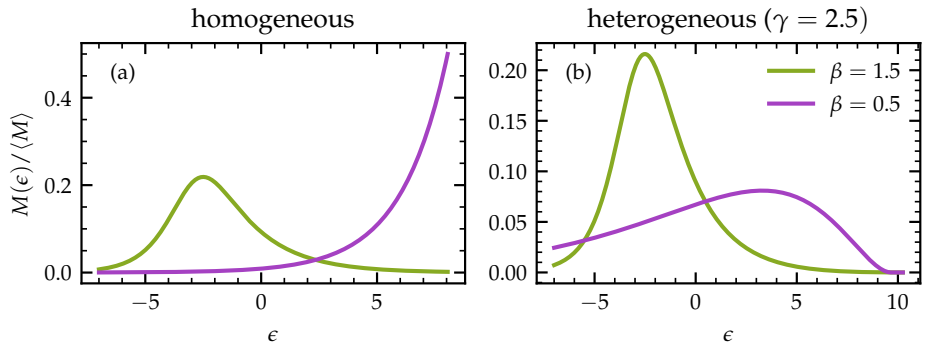



Figure 5.2: The number density per link $M(\epsilon)/\langle M \rangle$ as a function of the energy $\epsilon$ for (a) a homogeneous network and (b) a heterogeneous network with $\gamma = 2.5$. The purple line represents a weakly geometric network and the green a strongly geometric one.

This difference in arrangement of the cycles is caused by the following process. At low temperatures, the high energy associated to connecting spatially distant points causes the majority of links attached to a given node to be local. This defines the geometric phase at $\beta > \beta_c$ where the triangle inequality plays a critical role in the formation of cycles of finite size. As temperature increases, the number of energetically feasible links connecting very distant pairs of nodes grows, and at $\beta \leq \beta_c$ the number of available long range states becomes macroscopic due to the logarithmic dependence of the energy on distance, which causes the entropy per link to be infinite in this regime. This defines a non-geometric phase where links are mainly long ranged and the fraction of finite size cycles vanishes because the triangle inequality stops playing a role. This in turn implies that chordless cycles are necessarily of the order of the network diameter.

We can further clarify this point by looking at the number density of the system, which is given by the product of the density of states and the Fermi-Dirac occupation probability

$$M(\epsilon) = \binom{N}{2} \frac{\rho(\epsilon)}{1 + e^{\beta(\epsilon - \mu)}} \tag{5.20}$$

This quantity gives the expected amount of particles/links with associated energy $\epsilon$. As can be seen in Fig. 5.2, there is an abrupt change in the system when the critical point is crossed: While the vast majority of links has low energy when $\beta > 1$, the distribution is peaked at high energies for $\beta < 1$. This is a direct consequence of the exponential density of states of the system; in fact, we show in Appendix A.1.5 that for a simple toy model of non-interacting classical particles, an exponential density of states is sufficient to obtain the anomalous entropic behavior observed previously.

In the geometric phase, there are finite cycles of any order, although, as we show in Fig. 5.3, the density of triangles is much higher than the density of squares,

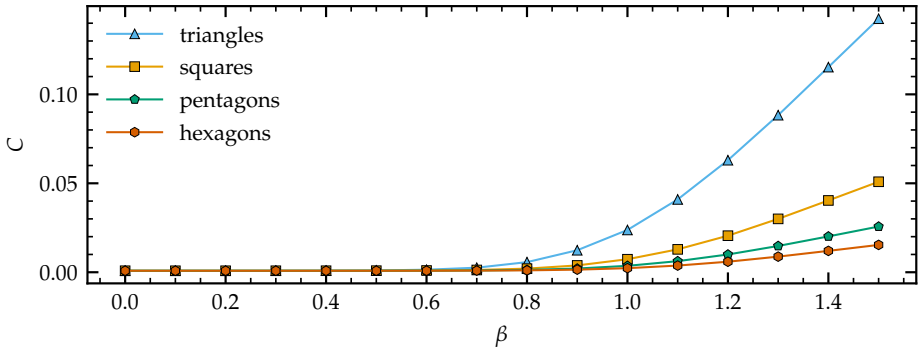



Figure 5.3: The global clustering coefficient for different sized chordless cycles as a function of the inverse temperature. The global clustering coefficient is defined as the ratio between the amount of closed $n$-lets and the total amount of $n$-lets, where $n$ goes from three (triangles) to six (hexagons). This coefficient is a measure for the amount of different sized chordless cycles, as a function of the inverse temperature $\beta$. The results shown are for networks of size $N = 5000$ and $\langle k \rangle = 6$. Errorbars representing the standard error are smaller than the data points and therefore not displayed.

pentagons, etc. In the non-geometric phase, the cycles are of the order of the network diameter. However, due to the (ultra) small-world property and finite size effects the diameter of the network can be quite small, so that the distinction between finite cycles of order higher than three and long range cycles can be difficult. Therefore, the average local clustering coefficient –measuring the density of the shortest possible cycles, which are also the most numerous– is the perfect order parameter to quantify this topological phase transition.

## 5.3 FINITE SIZE SCALING BEHAVIOR

To quantify the behavior of clustering in this transition, we compute the average local clustering coefficient, $\bar{c}$, as the local clustering coefficient averaged over all nodes in a network. The local clustering coefficient, as defined in Sec. 2.2.4, is given by the actual amount of triangles connected to a node as a fraction of the total amount of possible triangles connected to said node. In a hidden variable model, the local clustering coefficient for a given node $i$ with variables $\boldsymbol{h} = \{\theta_i, \kappa_i\}$ is defined as the probability that a pair of randomly chosen neighbors are neighbors themselves and, using results from [141], can be computed as

$$c_i = \frac{\sum_{j \neq i} \sum_{k \neq i} p_{ij} p_{jk} p_{ik}}{\left( \sum_{j \neq i} p_{ij} \right)^2}. \tag{5.21}$$

### 5.3.1 *Analytic results*

If we assume that $N \gg 1$, the sums in Eq. (5.21) can be replaced by integrals, leading to the expression

$$c(\boldsymbol{h}) = \frac{\int\int\int\int \mathrm{d}\theta' \mathrm{d}\theta'' \mathrm{d}\kappa' \mathrm{d}\kappa'' \rho(\kappa')\rho(\kappa'') p(\boldsymbol{h},\boldsymbol{h}') p(\boldsymbol{h},\boldsymbol{h}'') p(\boldsymbol{h}',\boldsymbol{h}'')}{\left(\int\int \mathrm{d}\theta' \mathrm{d}\kappa' p(\boldsymbol{h},\boldsymbol{h}')\right)^2}. \tag{5.22}$$

Note that, due to the rotational symmetry of the system, the node in question can be placed at the origin such that $\boldsymbol{h} = \{0, \kappa\}$. In principle, it is still necessary to marginalize this expression with respect to $\kappa$ to obtain the true average local clustering coefficient. However, as pointed out in Refs. [144, 145], $c(\kappa)$ is a monotonously decreasing function, which implies that the dominant scaling behavior can be obtained by plugging in any arbitrary, small, value for $\kappa$.

In Appendix A.2 we derive analytic results for the behavior of the average local clustering coefficient when hidden degrees follow a power law distribution $\rho(\kappa) \sim \kappa^{-\gamma}$ with $2 < \gamma < 3$ and a cutoff $\kappa < \kappa_c \sim N^{\alpha/2}$. Notice that the arguments above, presenting the average local clustering coefficient as an appropriate order parameter, should be valid for all choices of the distribution of the hidden degrees, as long as they lead to sparse graphs. Here, we choose this specific definition because it is the most common in the literature and allows for analytically tractable results. Notice also that it includes both the heterogeneous case with ($\alpha > 1$) and without ($0 < \alpha \leq 1$) degree-degree correlations [82], as well as the homogeneous case ($\alpha = 0$) where $\rho(\kappa) = \delta(\kappa - \langle k \rangle)$.

When $\beta > 1$, i.e. in the geometric region, the average local clustering coefficient behaves as [30]

$$\lim_{N\to\infty} \bar{c}(N,\beta) = Q(\beta), \tag{5.23}$$

for some constant $Q(\beta)$ that depends on $\beta$. Moreover, there exists a constant $Q'$ such that

$$\lim_{\beta\to 1^+} \frac{Q(\beta)}{(\beta-1)^2} = Q'. \tag{5.24}$$

The analytic results for $\beta \leq 1$ are derived by finding appropriate bounding functions $f(N,\beta) \leq \bar{c}(N,\beta) \leq g(N,\beta)$ that are both asymptotically proportional to $N^{-\sigma(\beta)}h(N,\beta)$, where $h(N,\beta)$ represents some non-power law function of $N$, implying that $\bar{c} \sim N^{-\sigma(\beta)}h(N,\beta)$ as well. When $\beta_c' < \beta \leq 1$,

$$\bar{c}(N,\beta) \sim \begin{cases} (\ln N)^{-2} & \text{if } \beta = 1 \\ N^{-2(\beta^{-1}-1)} & \text{if } \beta_c' < \beta < 1 \end{cases} \tag{5.25}$$

where the value of $\beta_c'$ depends on the parameter $\alpha$. If $\alpha > 1$, it is given by $\beta_c' = 2/\gamma$ and if $\kappa_c$ grows with $N$ slower than any power law ($\alpha = 0$) then $\beta_c' = \frac{2}{3}$. The fact that the decay in this regime is both temperature dependent and very slow, leads us to coin it the *quasi-geometric* region, as the geometry still induces triangles

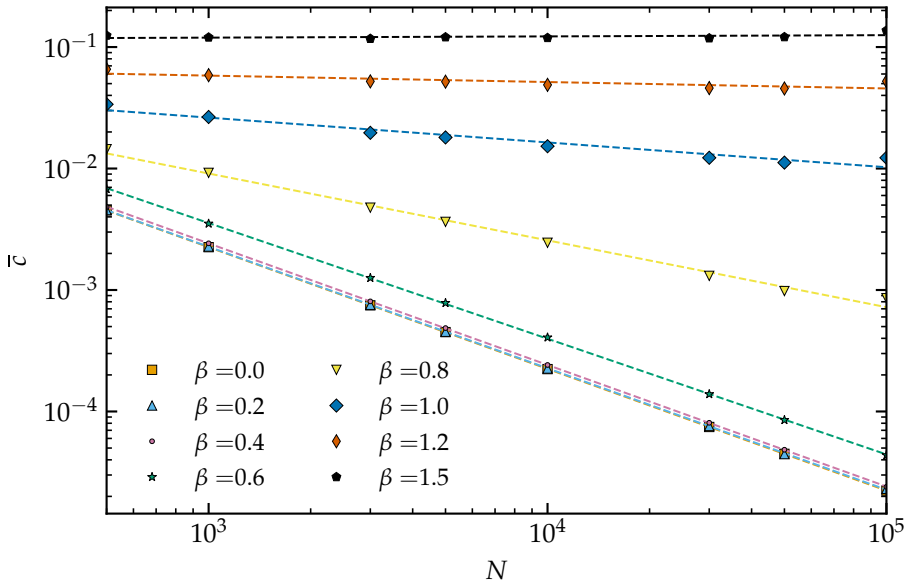



Figure 5.4: The average local clustering coefficient as a function of the network size. The networks were generated by applying the DPG technique to a configuration model network with a regular degree sequence $k_i = 4, \forall i \in V$. Dashed lines are power law fits used to estimate the exponent $\sigma(\beta)$ defined as $\bar{c} \sim N^{-\sigma(\beta)}$.

in the network for very large finite systems. Notice that the behavior in a close neighborhood of $\beta_c$ is independent of $\gamma$. The fact that the microscopic details of the model, in particular the hidden degree distribution, do not affect this scaling behavior points to the universality of our results.

Finally, when $\beta < \beta_c'$, the exact scaling behavior depends on $\alpha$ (see Appendix A.2 for the case $0 < \alpha \leq 1$):

$$\bar{c}(N, \beta) \sim \begin{cases} N^{-(\gamma-2)} \ln N & \text{if } \alpha > 1 \\ N^{-1} & \text{if } \alpha = 0. \end{cases} \tag{5.26}$$

The fact that these results coincide with those derived for the SCM [144], where there is no latent geometry, leads us to call this region *non-geometric*.

These results are remarkable in many respects. First, clustering undergoes a continuous transition at $\beta_c = 1$, attaining a finite value in the geometric phase $\beta > \beta_c$ and becoming zero in the non-geometric phase $\beta < \beta_c$ in the thermodynamic limit. The approach to zero when $\beta \to \beta_c^+$ is very smooth since both clustering and its first derivative are continuous at the critical point. Second, right at the critical point, clustering decays logarithmically with the system size, and it decays as a power of the system size when $\beta < \beta_c$. This is at odds with traditional continuous phase transitions, where one observes a power law decay at the critical point and an even faster decay in the disordered phase. Third, there is a quasi-geometric

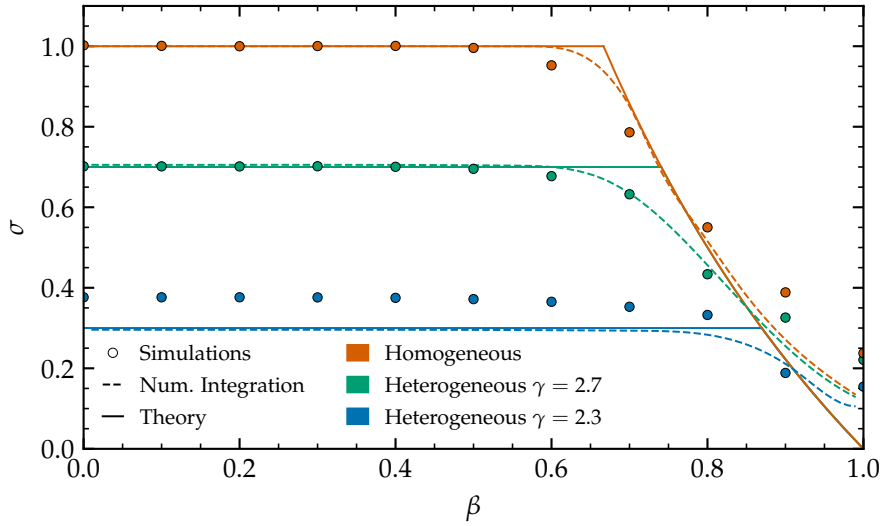



Figure 5.5: Exponent of the average local clustering coefficient as a function of the inverse temperature. The exponent $\sigma(\beta)$, defined by $\bar{c}(N,\beta) \sim N^{-\sigma(\beta)}h(N,\beta)$, with $h(N,\beta)$ a non-power law function of $N$, evaluated from numerical simulations (colored circles), numerical integration of Eq. (5.21) (dashed lines), and theoretical approach Eq. (5.25,5.26) (solid lines). Networks are generated with a homogeneous distribution of hidden degrees (red lines and circles) and a power law distribution with exponents $\gamma = 2.3$ and $\gamma = 2.7$, blue and green lines and circles, respectively.

region $\beta_c' < \beta < \beta_c$ where clustering decays very slowly, with an exponent that depends on the temperature. Finally, for $\beta < \beta_c'$, we recover the same result as that of the SCM for scale-free degree distributions [144].

The results in Eqs. (5.25,5.26) around the critical point suggest that $N_{\text{eff}} = \ln N$ plays the role of the system size instead of $N$. Indeed, in terms of this effective size, we observe a power law decay at the critical point and a faster decay in the weakly geometric phase, as expected for a continuous phase transition. Consequently, we expect the finite size scaling ansatz of standard continuous phase transitions to hold with this effective size. We then propose that, in the neighborhood of the critical point, clustering at finite size $N$ can be written as

$$\bar{c}(\beta, N) = (\ln N)^{-\frac{\eta}{\nu}} f\left((\beta - \beta_c)(\ln N)^{\frac{1}{\nu}}\right), \tag{5.27}$$

with $\eta = 2$, $\nu = 1$, and where $f(x)$ is a scaling function that behaves as $f(x) \sim x^{\eta}$ for $x \to \infty$.

### 5.3.2 *Numerical methods*

We test these results with numerical simulations and by direct numerical integration of Eq. (5.22). Simulations are performed with the degree-preserving geometric

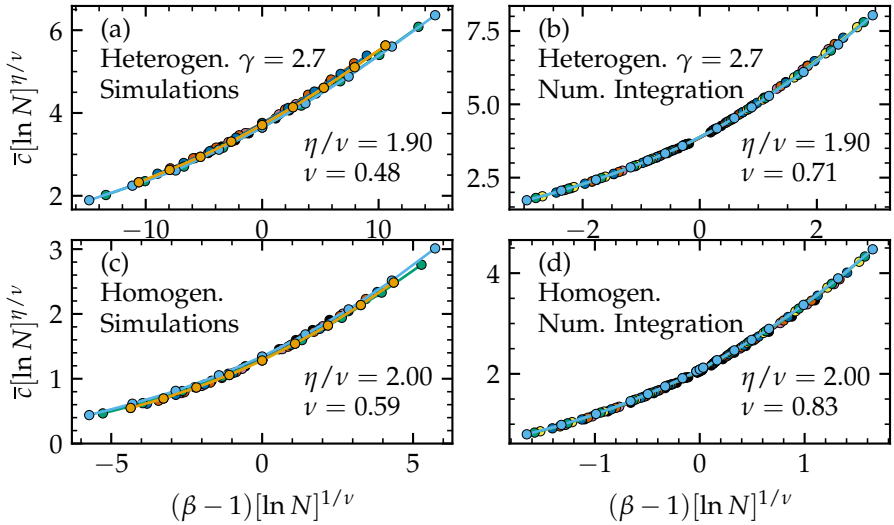



Figure 5.6: Finite size scaling analysis. Data collapse of the average local clustering coefficient at different sizes as defined in Eq. (5.27) for heterogeneous networks with $\gamma = 2.7$ (panels (a) and (b)) and homogeneous networks (panels (c), and (d)). The panels (a) and (c) correspond to numerical simulations with sizes in the range $N \in (5 \times 10^2, 10^5)$, whereas the panels (b) and (d) are obtained from numerical integration of Eq. (5.21) with sizes in the range $N \in (5 \times 10^5, 10^8)$. Different colors correspond to the different system sizes used.

(DPG) Metropolis-Hastings algorithm introduced in Ref. [226], that allows us to explore different values of $\beta$ while preserving exactly the degree sequence. Given a network, the algorithm selects at random a pair of links connecting nodes $i, j$ and $l, m$ and swaps them (avoiding multiple links and self-connections) with a probability given by

$$p_{swap} = \min\left[1, \left(\frac{\Delta\theta_{ij}\Delta\theta_{lm}}{\Delta\theta_{il}\Delta\theta_{jm}}\right)^{\beta}\right], \tag{5.28}$$

where $\Delta\theta$ is the angular separation between the corresponding pair of nodes. This algorithm maximizes the likelihood that the network is $\mathbb{S}^1$ geometric while preserving the degree sequence and the set of angular coordinates, and it does so independently of whether the system is above or below the critical temperature. Notice that the continuity of Eq. (5.28) as a function of $\beta$ makes it evident that, even if the connection probability takes a different functional form above and below the critical point, the model is the same.

Figure 5.4 shows the behavior of the average local clustering coefficient as a function of the number of nodes for homogeneous $\mathbb{S}^1$-networks with different values of $\beta$, showing a clear power law dependence $N^{-\sigma(\beta)}$ in the non-geometric phase $\beta < \beta_c$, with an exponent that varies with $\beta$ as predicted by our analysis. These results are used to measure the exponent $\sigma(\beta)$ as a function of the inverse

temperature $\beta$, which in Fig. 5.5 are compared with the theoretical value given by Eq. (5.25,5.26). The agreement is in general very good, although it gets worse for values of $\beta$ very close to $\beta_c$ and for very heterogeneous networks. This discrepancy is expected due to the slow approach to the thermodynamic limit in the non-geometric phase, which suggests that the range of our numerical simulations, $N \in [5 \times 10^2, 10^5]$, is too limited. To test for this possibility, we solve numerically Eq. (5.21) for sizes in the range $N \in [5 \times 10^5, 10^8]$ and measure numerically the exponent $\sigma(\beta)$. In this case, the agreement is also very good for heterogeneous networks. The remaining discrepancy when $\beta \approx \beta_c$ is again expected since, as shown in Eq. (5.25), right at the critical point clustering decays logarithmically rather than as a power law. Finally, Fig. 9.3 shows the finite size scaling Eq. (5.27) both for the numerical simulations and numerical integration of Eq (5.21). In both cases, we find a very good collapse with exponent $\eta/\nu \approx 2$ in all cases. The exponent $\nu$, however, departs from the theoretical value $\nu = 1$ in numerical simulations due to their small sizes but improves significantly with numerical integration for bigger sizes. We then expect Eq. (5.27) to hold, albeit for very large system sizes.

## 5.4 CONCLUSION

In this Chapter we have seen that the $\mathbb{S}^1$-model shows different behavior of the average local clustering coefficient on the left and right side of $\beta_c = 1$. Studying the system as a gas of non-interacting fermions reveals that the entropy diverges at the critical point, implying a change in the structural organization of the system as a whole. Because the model is rotational invariant in both regimes, one can conclude that this transition is not due to symmetry breaking. The behavior of clustering—non-zero on the right and vanishing on the left of the critical point—indicates that the transition is of topological nature related to the organization of chordless cycles. In particular, these cycles are short above $\beta_c$ and of the order $\ln N$ below it. Taking the clustering coefficient, which measures the triangle density in the network as an order parameter, we observe anomalous finite size behavior, with $\ln N$ and not $N$ playing the role of the system size. This slow approach to the thermodynamic limit is relevant for real networks in the quasi-geometric phase $\beta_c' < \beta \leq 1$, for which high levels of clustering can still be observed. In Chapter 6, we extend the network embedding tool Mercator, introduced in Sec. 4.5, and show in Chapter 8 that, indeed, many real networks live in the weakly geometric regime.

# 6

# NETWORK EMBEDDING

In the previous Chapter we saw that the decay of the average local clustering coefficient in the weakly geometric regime $\beta \leq 1$ is very slow, indicating that the transition between geometric and non-geometric graphs might not be sharp. This led us to divide the subcritical regime into two sub-regions: (1) the non-geometric regime, where clustering decayed in the same manner as for the SCM, and (2) the quasi-geometric regime, where clustering decayed extremely slowly and in a temperature dependent manner.

In order to better understand what this implies for the "geometricity" [59, 60] of a network, in the following two Chapters we attempt to discover how important the underlying geometry is for determining the structural properties of the network in these two regimes. In this Chapter, which is an adapted version of Ref. [61], we do so using network embeddings. First, we extend the embedding tool Mercator to the weakly geometric regime in Sec. 6.1. Then, in Sec. 6.2, we show that for a range of weak couplings the original geometry can be recovered based solely on the topology. We also show that this region coincides with the quasi-geometric regime. We note that this is not a finite size effect and that geometric information can be extracted for networks of any size in the quasi-geometric regime. Finally, in Sec. 6.3, it is found that Mercator can also provide useful embeddings when the network is explicitly non-geometric, even in the case of the configuration model [133]. Similarly to how fluctuations can induce spurious communities [92, 227], finite size effects can lead to an *effective geometry*, which can be used, for example, in greedy routing algorithms.

## 6.1 MERCATOR IN THE WEAKLY GEOMETRIC REGIME

Network geometry has important practical implications for real systems. For example, it can be used for routing information on the Internet [28], for community detection [52, 53] or for the prediction of missing links [57, 58], as well as for creating downscaled network replicas [42]. In order to do so, one needs to be able to faithfully embed real-world networks into the hidden metric space using only the information contained in their topology.

Even though there are many ways to obtain such an embedding [184, 194, 197, 228–233], here we focus on Mercator [64], which finds the hidden $\mathbb{S}^1/\mathbb{H}^2$-coordinates such that realizations based on these coordinates best reproduce the

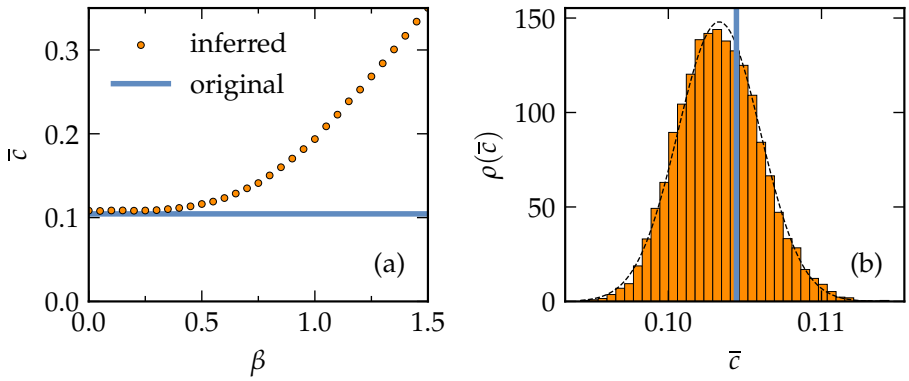



Figure 6.1: Inferring beta for the Ecological network (details in Tab. 8.1). **(a)** The orange points represent the inferred average local clustering coefficient given the fitted hidden degree and different betas, and the blue horizontal line indicates the original clustering coefficient of the network. **(b)** The histogram represents the probability density of the local clustering coefficient produced by degree preserving randomization of the connectivity of the original network. The black dotted line is the fitted normal distribution and the continuous blue line indicates the original clustering.

properties of the original network. As explained in Sec. 4.5, Mercator employs a combination of machine learning and maximum likelihood methods to achieve this goal, which allows it to be both precise as well as efficient.

In the original version, Mercator was only able to handle strongly geometric networks with $\beta > 1$. Extending it to the weakly geometric regime implies implementing the change in connection probability when crossing the critical temperature as found in Eq. (4.26). The same goes for the parameter $\hat{\mu}$. However, here we cannot just take the other asymptotic equation for $\beta < 1$ because we want Mercator to be applicable to small networks with $\beta \approx 1$ as well. Hence, we produced a new version of Mercator that is able to handle networks in the whole range of $\beta$ values and where $\hat{\mu}$ is determined numerically such that the observed average degree of the real network matches exactly the expected average degree of a $\mathbb{S}^1$-network with uniform node distribution and a hidden degree distribution that matches the observed one.

One of the challenges of embedding networks below $\beta = 1$ comes from the fact that the function $\overline{c}(\beta)$ flattens off as one approaches the infinite temperature limit $\beta \to 0$, as can be seen in Fig. 6.1a. Here, we take as an example an ecological network [234], where nodes represent taxa and edges trophic relationships. As per the initial steps of Mercator, we choose a certain $\beta$ and set the hidden degrees such that the degree distribution is reproduced and with this calculate the expected average local clustering coefficient $\langle \overline{c} \rangle$. Repeating this for a range of $\beta$'s, one observes that the function approaches a constant as $\beta$ approaches zero. The horizontal line in the figure represents the actual level of clustering in the ecological

network under study. We can say, with confidence, that $\beta_{real} \lesssim 0.5$, but cannot determine a lower bound. As the true value could be $\beta = 0$, i.e. the levels of clustering in the network could be described by the configuration model, the coupling between the geometry and topology of the network is extremely weak and it is, thus, effectively non-geometric. We conjecture that these networks either have no associated geometry to begin with, or are coupled so weakly to it that it cannot be reproduced. Thus, Mercator must be able to detect these types of networks, i.e., it must be able to filter out non-geometric networks. In order to do so we want Mercator to answer the following question: "Can the observed levels of clustering be plausibly explained by the configuration model?".

To answer this question, we need to add a step to the algorithm. Before the embedding of a network starts, a large amount of random copies are created using degree-preserving randomization [235]. This randomization step destroys all information contained in the network, except for the degrees of the nodes and structural correlations imposed by global constraints at finite sizes. Because the angular coordinate in the $\mathbb{S}^1/\mathbb{H}^2$-model functions as a proxy for all attributes of a node, except for its degree, it is clear that removing this information is equivalent to decoupling the network from its similarity dimension, exactly what happens at $\beta = 0$. Thus, these random copies are just realizations of the configuration model preserving the original degree distribution. We then calculate the average local clustering coefficient for all randomized copies, leading to the distribution shown in Fig. 6.1b. The observed level of clustering is given by the vertical line, and we can conclude that it is completely in line with the configuration model. Had the observed clustering been much larger, we might conclude that it is statistically unlikely that the network was generated with the configuration model and that $\beta > 0$. We stress, however, that networks not being congruent with the configuration model does not necessarily mean the $\mathbb{S}^1$-model is a good fit for explaining the observed network structure. It merely implies that Mercator has at least the potential to find a meaningful embedding. How well this embedding reproduces the network's structural properties must be studied separately. We do precisely this in Chapter 8.2.

## 6.2 RECOVERING GEOMETRIC INFORMATION

It is important to first study if it is possible to recover geometric information from the topology of a weakly geometric network when the ground truth about its geometry is known. We generate heterogeneous networks where the distribution of hidden degrees is given by the Pareto distribution with exponent $\gamma > 2$.

We focus on Mercator's ability to recover the angular coordinates of the original network, as it is only the coupling to the similarity dimension that becomes weaker as $\beta \to 0$. The performance with respect to the popularity dimension should not be much different than in the region $\beta > 1$, which has been already extensively studied in Ref. [64].

In Fig. 6.2 we show the performance of Mercator for various inverse temperatures $\beta$ and $\gamma = 2.5$. We see that, as expected, the embedding gets progressively

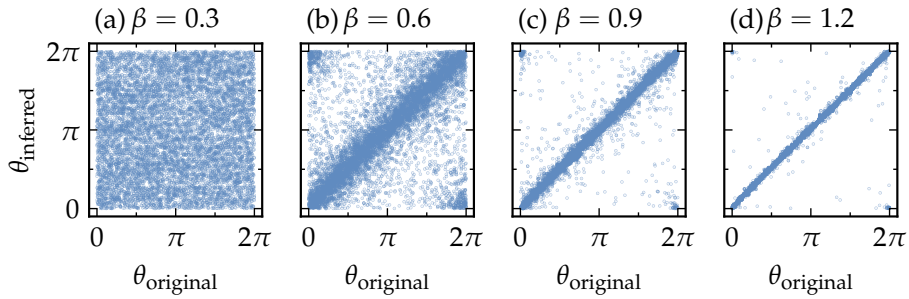



Figure 6.2: Examples of the inferred angular coordinate versus the original angular coordinate for artificial networks generated with the $S^1$-model at varying inverse temperatures (a) $\beta = 1.2$, (b) $\beta = 0.9$, (c) $\beta = 0.6$, and (d) $\beta = 0.3$. All networks were created with the following parameters: $N = 5000$, $\langle k \rangle = 12$, $\gamma = 2.5$.

worse as the temperature increases. As $\beta \to 0$, the coupling to the geometry becomes weaker and there is thus less geometric information contained in the topology. We do, however, observe that the embedding is still good even for $\beta$'s relatively far below the transition point. For small values of $\beta \lesssim 0.6$, Mercator is not able to infer the correct $\beta$ because the geometric coupling in these networks is extremely weak. Even so, by feeding Mercator with the correct value of $\beta$, we are able to obtain an embedding. However, as shown in Fig 6.2d, the obtained coordinates are completely different from the original ones so that, even if the nodes were originally placed on an underlying geometry, the resulting topology is not congruent with it.

Now that we have an intuitive idea of the performance of Mercator in the weakly geometric region $0 < \beta \leq 1$, the next step is to substantiate these results. To this end, we generate network realizations, embed these and test the quality of the embedding. As a quality measure we choose the *concordance* or *C-score*, which quantifies the similarity of two different orderings.

In our case, the first ordering is given by the set of vertices in a network, ordered by their original coordinates, and the second by ordering the indices according to the inferred coordinates. First introduced in Ref. [236], the C-score was adjusted to a system with periodic boundary conditions in Ref. [197], leading to the following definition:

$$\text{C-score} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \delta(i,j), \tag{6.1}$$

where $N$ is the total amount of nodes, $i$ and $j$ indicate two nodes and $\delta(i,j)$ is 1 if the shortest distance between $i$ and $j$ along the circle has the same direction (clockwise or counterclockwise) in both the original and inferred ordering, and 0 if the direction is different. Note that it is possible that Mercator returns an inverted ordering, which, for example, leads to an inverted diagonal in Fig. 6.2, as well as a C-score $< 0.5$. Of course, the orientation of the ring does not influence the quality of the embedding, as it is only the distance between points along the circle that

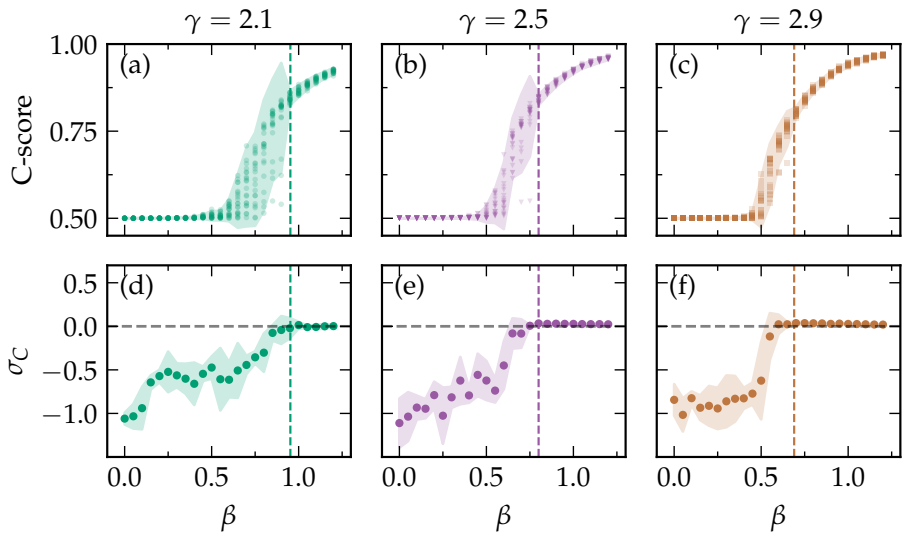



Figure 6.3: (a-c) Dots represent the C-score as a function of the inverse temperature $\beta$ for individual realizations of the $\mathbb{S}^1$-model for $N = 4000$. (d-f) The scaling exponent of the C-score, assuming $(\text{C-score} - 0.5) \propto N^{\sigma_C}$. For the fit, networks of size $N \in \{1000, 2000, 4000, 8000, 16000\}$ where used. For all panels $\langle k \rangle = 12$ and for (a,d) $\gamma = 2.1$ , (b,e) $\gamma = 2.5$ and (c,f) $\gamma = 2.9$. The vertical dashed lines indicate the critical inverse temperature $\beta_c' = 2/\gamma$ separating the quasi-geometric and non-geometric regimes. The shaded regions represent the $2\sigma$ confidence interval.

matters. Therefore, we are actually interested in using $\max(\text{C-score}, 1 - (\text{C-score}))$ as a measure, such that 1 implies perfect ordering and 0.5 means the inferred order is completely random.

In Fig. 6.3a-c we show the results for the C-score as a function of the geometric coupling $\beta$ for various $\gamma$. We observe a transition between almost perfect reproduction of the ordering by Mercator ($C = 1$) at high $\beta$, to a situation where the ordering is completely random ($C = 0.5$) at low $\beta$. This confirms the results we obtained in Fig. 6.2.

In Fig. 6.3d-f we investigate how the performance varies as a function of $N$. To this end, we fit the function $(\text{C-score} - 0.5)(N)$, assuming that it scales as $\propto N^{\sigma_C}$. We use networks of sizes $N \in [1000, 16000]$ to perform this fit. We then plot the exponent $\sigma_C$ as a function of the inverse temperature $\beta$. We see that $\sigma_C = 0$ for a range of $\beta$, and then quickly decreases. This implies that the performance remains constant for a wide range of weak couplings, meaning that here geometric information can always be extracted irrespective of the system size. This is true despite the fact that clustering vanishes in the thermodynamic limit in this region. While the presence of non zero triangle densities in this region is a finite size effect, the presence of geometric information is not.

We note that the region where the performance is constant is larger for higher $\gamma$, and that the transition point lies around $\beta = 2/\gamma$, indicated by the dotted lines in Fig. 6.3d-f. The fact that $\sigma_C < 0$ below this point and $\sigma_C = 0$ above it implies that, in the thermodynamic limit, we should observe a jump in the performance, reminiscent of a first order phase transition. However, the observation that at $\beta = 2/\gamma$ the variance in the C-score explodes, as can be seen in Fig. 6.3a-c, indicates a continuous phase transition. These somewhat contradictory results might indicate that the transition is actually explosive in nature [237, 238].

The location of this transition is in agreement with the theoretical results found in Chapter 5. There, $\beta = \beta'_c = 2/\gamma$ marked the transition between slow, temperature dependent decay of the average local clustering coefficient for $\beta > \beta'_c$ and a faster decay for $\beta < \beta'_c$, equivalent to the one observed in the soft configuration model [144]. The fact that we recover this transition here is a very profound result, as it confirms that the division of the region $\beta \leq 1$ into these two sub-regions is not just theoretical in nature but has very real, observable consequences.

## 6.3 GREEDY ROUTING

The second test we discuss here is more practical in nature and involves the performance of the greedy routing protocol [29]. In this protocol, a pair of nodes is selected at random, and the goal is to efficiently send a packet of information from one to the other. This is done by looking at the neighbors of the node that contains the packet, which is then forwarded to whichever neighbor is closer in hyperbolic space to the destination. This is repeated until one of two scenarios occurs. In scenario (1), the packet reaches the destination. In scenario (2), the neighbor closest to the goal is the node from which the parcel just arrived. In this latter case the packet is dropped as the destination cannot be reached using the greedy routing method.

One of the measures to define how well this algorithm performs is the success probability $p_s$, defined as the fraction of nodes pairs for which a greedy routing path exists. In Ref. [29] it was shown that information can be efficiently routed through the network if one uses the coordinates in the latent space. Of course, this works better when the connection to this underlying space is stronger, i.e. when $\beta$ is higher. This is confirmed in Fig. 6.4a: When using the original coordinates and the hyperbolic distance

$$d_{ij} = r_i + r_j + 2\min(1,\beta)\ln\frac{\Delta\theta}{2}, \tag{6.2}$$

as defined in Eq. (4.39), one observes that the success probability $p_s$ decays with $\beta$ until leveling out at $\beta = 0$. Here, the angular coordinates are no longer taken into account and the greedy routing is purely based on the degrees of the network. If we redefine the hyperbolic distance such that it reads

$$\hat{d}_{ij} = r_i + r_j + 2\ln\frac{\Delta\theta}{2}, \tag{6.3}$$

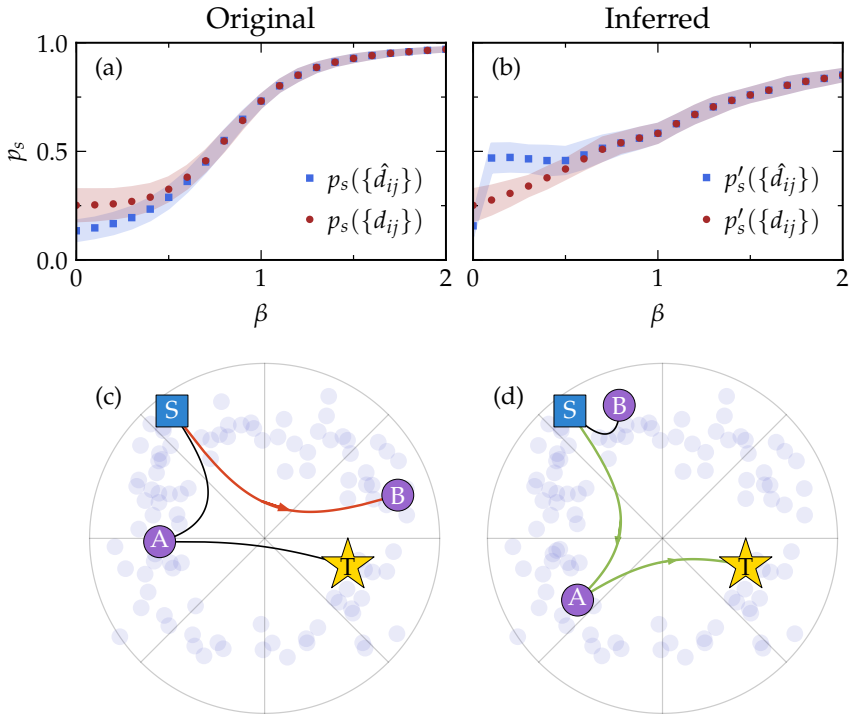



Figure 6.4: Results for the greedy routing routine. (a) Success probability $p_s$ based on the original coordinates, using both the definition of the hyperbolic distance where the contribution vanishes at $\beta = 0$ ($d_{ij}$) as well as where it remains constant ($\hat{d}_{ij}$). The shaded regions represent the $2\sigma$ confidence interval. Similar results are shown in panel (b) where now the inferred coordinates are used. Parameters used: $\{N, \langle k \rangle, \gamma\} = \{4000, 12, 2.5\}$. In panels (c) and (d) a schematic representation of greedy routing paths based on the original and inferred coordinates, respectively, are shown.

i.e. such that the effect of the angular coordinates does not diminish, we see that the results are worse. This is because, for extremely low $\beta$, the connection between the topology and the geometry is lost and the angular coordinates are thus meaningless, impeding proper routing. Let us now turn to the inferred coordinates, for which the results can be found in Fig. 6.4b. Returning to the original definition of the hyperbolic distance, we see that for $\beta \leq 1$, the results are better than in the case of the original coordinates. This can be understood as follows: As lowering $\beta$ can be interpreted as increasing the temperature, more of the connectivity is determined by noise (conditioned on the hidden degrees). However, Mercator will always try and find as much geometry as possible, and place nodes in such a way that the inputted network realization is most congruent with it. In practice, this means that two nodes that were originally far away from each other, but are connected due to the large fluctuations, will most likely be placed close together in the embedding. In other words, the fact that for finite systems even non-geometric

random graphs display clustering implies that an effective geometry can be found such that the effect of the triangle inequality on the topology is strongest. This is reminiscent of the fact that fluctuations in random graphs can lead to high modularity [92], which can lead to the detection of spurious communities [91]. In our case, Mercator is able to uncover an effective geometry, arising from the noise in the system (which makes this a finite size effect). However, where detecting spurious communities can be considered undesirable, the effective geometry can be useful.

For example, it is beneficial to the greedy routing routine, as nodes that are close together are now also connected. Of course, when using the original hyperbolic distance, this effect will eventually disappear as the angular coordinates are no longer taken into account, leading to a success probability that coincides with that of the original coordinates at $\beta = 0$. However, if we again use Eq. 6.3, keeping the influence of $\theta$ constant, we note that high success probabilities can be achieved, even for low betas. Note that exactly at $\beta = 0$ this is no longer the case as Mercator does not even try to find meaningful angular coordinates.

In Fig. 6.4c,d this effect is clarified graphically. Here, the goal is to send a packet from the source node labeled $S$ to the target node labeled $T$. There is only one correct path, passing through node $A$. In the original metric space, due to fluctuations, the source node is also connected to node $B$, even though it lies far away from it. As $B$ lies closer to the target, the packet will get forwarded there. There is, however, no connection between $B$ and $T$ and so the packet gets dropped. In the case of the inferred coordinates, node $B$ gets placed closer to the source, and further from the target, in accordance with its connectivity. Node $A$ now lies closer to the target than $B$ does, and so a successful routing is achieved.

## 6.4 CONCLUSIONS

The success of the framework of network geometry in describing real systems has led many to wonder if there is a way to determine if real networks are indeed geometric in nature [59, 60]. In general, (some function of) the amount of closed triangles in the network, expressed, for example, by the average local clustering coefficient, is taken to indicate the presence of geometry. In this Chapter we use the network embedding tool Mercator to shed light on this important question. After extending it to the weak coupling regime, we show that, in the quasi-geometric region, the tool is able to recover a significant amount of geometric information from the topology of the network alone. This implies that geometric information is indeed relevant in this regime. Only when the coupling strength is very close to zero does the geometricity completely vanish. Here, the properties of the network can be explained by the soft configuration model. We show that the presence of triangles in finite non-geometric random graphs allows for the definition of an effective geometry by Mercator. This effective geometry does not reflect an original underlying geometry, which is absent, but can still be useful for information routing problems.

# 7

# NETWORK RENORMALIZATION

In the last Chapter we observed that geometric information can be extracted from the topology alone in the quasi-geometric regime through network embedding. This can be taken to mean that this information is important for understanding the connectivity of the network, even though the geometric coupling is weak in this regime. In this Chapter, adapted from Ref. [62], we reach the same conclusion by studying network renormalization.

The Renormalization Group remains an essential tool in statistical physics to study systems at different length scales, and for revealing the scale invariance and universal properties of critical phenomena near continuous phase transitions, where fluctuations are strong [239]. The simplest technique for processes on regular lattices is that of the block spin method proposed by Kadanoff [240], where blocks of nearby nodes are grouped together into supernodes whose state is determined by some averaging rule. Extending this method to complex networks is complicated by their small world property, which makes the concept of closeness fuzzy and hinders the definition of supernodes [121].

Different methods have been proposed for circumventing this problem. The first renormalization scheme for complex networks was the box covering method, where closeness is defined in a chemical sense: nodes are grouped if they lie closer than a certain amount of hops away from one another in the topology of the network [241]. Other methods are based on the graph Laplacian, where closeness is based on diffusive distance [242, 243]. Finally, network geometry offers a natural framework for renormalizing networks, with closeness being related to the distance in similarity space.

In Ref. [43, 50], a renormalization procedure was defined in which adjacent nodes are coarse-grained into supernodes on the basis of their coordinates in their latent geometry. The geometric renormalization (GR) approach has revealed that scale invariance is a pervasive symmetry in real networks [50]. From a practical perspective, GR has also enabled the generation of scaled-down self-similar replicas—an essential tool for facilitating the computationally challenging analysis of large networks. Additionally, when combined with scaled-up replicas produced through a fine-graining reverse renormalization technique [43], it provides a means to explore size-dependent phenomena.

This Chapter is organized as follows: First, we develop and assess the performance of GR in the region of weak geometric coupling in Sec. 7.1. Then, in Sec. 7.2,

we show that in the quasi-geometric regime, geometric information is essential for obtaining self-similarity in important network measures across scales.

## 7.1 GEOMETRIC RENORMALIZATION IN THE WEAKLY GEOMETRIC REGIME

The first step in GR is to define non-overlapping sectors along the $\mathbb{S}^1$-circle containing each $r$ consecutive nodes. To determine the coordinates of the nodes in the geometric space and, hence, which nodes are consecutive in the similarity space, real networks must of course first be embedded using, for example, the Mercator tool extended in the previous Chapter. The second step is to coarse grain the nodes within a group to form a single supernode, whose angular coordinate and hidden degree are functions of the coordinates and hidden degrees of its constituents. It is essential that the supernode order along the circle preserves the order of nodes in the original layer. The connectivity of the new network is defined by connecting two supernodes if any pair of their respective constituents are connected. This procedure can be repeated iteratively starting from the original layer $l = 0$. Each layer $l$ is then $r^l$ times smaller than the original network. This defines the renormalization group flow.

In the original work, GR was only defined for $\beta > 1$, assuming that this is the region where real networks live. However, in the previous Chapters we have seen that weakly geometric networks, with $\beta \leq 1$, can also exhibit high levels of clustering. In the following, we extend this procedure to this regime. We use a compact notation that includes the results in [50] for $\beta > 1$.

### 7.1.1 *Self-similarity of the connection probability*

In the renormalization procedure described above, supernodes in layer $l+1$ are formed by combining $r$ adjacent nodes from layer $l$. If any constituent of supernode $\sigma$, denoted by the set $\mathcal{S}(\sigma)$, is connected to any of the constituents of supernode $\tau$ they are said to be connected. The probability of this being the case is given by

$$p_{\sigma\tau}^{(l+1)} = 1 - \prod_{\{i,j\}\in\mathcal{P}(\sigma,\tau)} (1 - p_{ij}^{(l)}), \tag{7.1}$$

i.e., one minus the probability that none of the constituents are connected. Here we have defined $\mathcal{P}(\sigma,\tau) = S(\sigma) \times S(\tau)$. Using $p_{ij}^{(l)} = 1/(1-\chi_{ij}^{(l)})$ we can rewrite this expression as

$$p_{\sigma\tau}^{(l+1)} = 1 - \frac{1}{\prod_{\{i,j\}\in\mathcal{P}(\sigma,\tau)}(1+(\chi_{ij}^{(l)})^{-1})}. \tag{7.2}$$

The denominator of the second term can be expanded as

$$1 + \sum_{\{i,j\}\in\mathcal{P}(\sigma,\tau)} (\chi_{ij}^{(l)})^{-1} + \sum_{\{i,j\}\in\mathcal{P}(\sigma,\tau)} (\chi_{ij}^{(l)})^{-1} \sum_{\{s,t\}\in\mathcal{P}(\sigma,\tau)\setminus\{i,j\}} (\chi_{st}^{(l)})^{-1} + \ldots \tag{7.3}$$

We now assume that $\chi_{ij}^{(l)} = (R^{(l)}\Delta\theta_{ij}^{(l)})^{\beta^{(l)}}/(\hat{\mu}^{(l)}\kappa_i^{(l)}\kappa_j^{(l)})^{\max(1,\beta^{(l)})} \gg 1$. This is a consequence of the sparsity of the graph and allows us to truncate the expansion at first order. We then realize that $\Delta\theta_{ij}^{(l)} \approx \Delta\theta_{\sigma\tau}^{(l+1)}$, the distance between the two supernodes. This is because the distances between the nodes within a single supernode are generally much smaller than the distances between nodes in different supernodes. This allows us to rewrite Eq. (7.2) as

$$p_{\sigma\tau}^{(l+1)} = \left(1 + \frac{\left(R^{(l)}\Delta\theta_{\sigma\tau}^{(l+1)}\right)^{\beta^{(l)}}}{\sum_{\{i,j\}\in\mathcal{P}(\sigma,\tau)}\left(\hat{\mu}^{(l)}\kappa_i^{(l)}\kappa_j^{(l)}\right)^{\max(1,\beta^{(l)})}}\right)^{-1}. \tag{7.4}$$

In order for this to be a proper connection probability in the renormalized layer, taking into account that $R^{(l+1)} = R^{(l)}/r$ and $\beta^{(l+1)} = \beta^{(l)} \equiv \beta$, we must demand $\hat{\mu}^{(l+1)} = \hat{\mu}^{(l)}/r^{\min(1,\beta)}$. Furthermore, the evolution of the hidden degrees is as follows

$$\kappa_\sigma^{(l+1)} = \left(\sum_{i\in\mathcal{S}(\sigma)}\left(\kappa_i^{(l)}\right)^{\max(1,\beta)}\right)^{1/\max(1,\beta)}. \tag{7.5}$$

Note that in the weak coupling regime $\max(1,\beta) = 1$, reducing the definition to a simple sum. This transformation respects the semi-group property of the renormalization as

$$\kappa_\sigma^{(l+2)} = \left(\sum_{i\in\mathcal{S}(\sigma)}\left(\kappa_i^{(l+1)}\right)^{\max(1,\beta)}\right)^{\frac{1}{\max(1,\beta)}} = \left(\sum_{i\in\mathcal{S}(\sigma)}\sum_{s\in\mathcal{S}(i)}\left(\kappa_s^{(l)}\right)^{\max(1,\beta)}\right)^{\frac{1}{\max(1,\beta)}}. \tag{7.6}$$

This final double sum is equivalent to a single sum over all $r^2$ nodes in the unrenormalized layer $l$ that make up the supernode in the layer $l+2$.

We have slightly more freedom for the similarity dimension, as we just need to find a definition of $\Delta\theta_{\sigma\beta}^{(l+1)}$ that (1) respects the semi-group property of the renormalization procedure, (2) respects the spherical symmetry of the system and (3) lies in the range defined by the angular coordinates of the constituent nodes and therefore respects the original node order. To this end we define

$$\theta_\sigma^{(l+1)} = \frac{\sum_{i\in S(\sigma)}(\kappa_i^{(l)})^{\max(1,\beta)}\theta_i^{(l)}}{\sum_{i\in S(\sigma)}(\kappa_i^{(l)})^{\max(1,\beta)}}, \tag{7.7}$$

which can been seen as a weighted average. Note that we do not choose the exact equation as given in Ref. [50] because that definition introduces a bias for the constituent node with the largest angular coordinate, which is not in line with the rotational symmetry of the system.

### 7.1.2 *The average degree*

The flow of the average hidden degree can be derived from Eq. (7.5). In the region $\beta \leq 1$ (the case $\beta > 1$ was already investigated in Ref. [50]), the hidden degree of a supernode is simply given by the sum of the hidden degrees of its constituents, which implies

$$\langle \kappa_\sigma^{(l+1)} \rangle = \Big\langle \sum_{i \in \mathcal{S}(\sigma)} \kappa_i^{(l)} \Big\rangle = \sum_{i \in \mathcal{S}(\sigma)} \langle \kappa_i^{(l)} \rangle = r \langle \kappa^{(l)} \rangle. \tag{7.8}$$

Thus, $\langle \kappa^{(l+1)} \rangle = r^\xi \langle \kappa^{(l)} \rangle$ where $\xi = 1$ for $\beta \leq 1$. Following similar steps as in Sec. 4.2.2, it can be shown that

$$\overline{k^{(l+1)}}(\kappa^{(l+1)}) = r^{\xi-1} \frac{\langle k^{(l)} \rangle}{\langle \kappa^{(l)} \rangle} \kappa^{(l+1)}. \tag{7.9}$$

We can then take the average over the hidden degree to get

$$\langle k^{(l+1)} \rangle = r^\nu \langle k^{(l)} \rangle, \tag{7.10}$$

where we define $\nu = 2\xi - 1$. The flow of the average degree in the weakly geometric regime is, thus, inversely proportional to the flow of the system size, which means that the amount of links $M = N\langle k \rangle / 2$ is a constant under renormalization, i.e., that no links are lost as one performs GR steps. This result implies that, on average, there is only one connection between the constituents of a pair of supernodes. This has to do with the fact that for $\beta \leq 1$ connections are long ranged in the thermodynamic limit. It is therefore exceedingly unlikely that a node is connected to two nodes so close together in the latent space that when we perform a GR step they end up in the same supernode.

Note that this result implies that the network will get denser along the renormalization flow. This can lead to problems when many renormalization steps are performed on relatively small networks. In this case, $\chi_{ij}^{(l)}$ as defined in Eq. (4.27), might cease to be large, resulting in Eq.(7.3) not being truncatable and the connection probability, therefore, loosing self-similarity. This is in line with the observations made in Ref. [189], where it was shown that the connection probability $p_{ij}^{(l)} = 1 - \exp(-1/\chi_{ij}^{(l)})$ is the only one that is explicitly self similar under renormalization. In the sparse regime, i.e., when $\chi_{ij}^{(l)}$ is large, this connection probability and the one in the $\mathbb{S}^1$-model are interchangeable, as discussed in Sec. 4.4. However, when the network becomes too dense and $\chi_{ij}^{(l)}$ ceases to be large, this equivalence breaks down, explaining the loss of self similarity in the $\mathbb{S}^1$-model. We still choose to work with the $\mathbb{S}^1$ connection probability for two reasons: First, the densification effect responsible for the loss of self-similarity can be mitigated through a geometric pruning procedure introduced in Ref. [42]. Second, in this Chapter we will see that even without pruning, self-similarity is obtained for a significant amount of renormalization steps. Only when the networks become truly small and dense does it break down, implying that for all practical purposes the $\mathbb{S}^1$ is sufficient, while having the added benefit of being explicitly maximally random.

### 7.1.3 *Quantifying the GR*

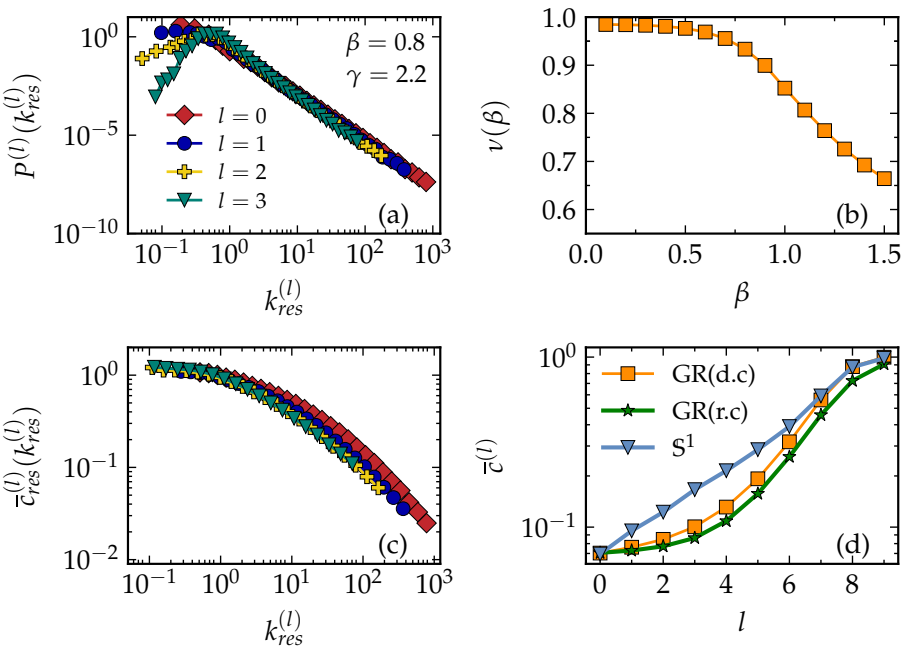



Figure 7.1: (a) The log-binned degree distribution $P^{(l)}(k^{(l)})$ as a function of the rescaled degrees $k^{(l)}_{res} = k/\langle k^{(l)}\rangle$. (b) The exponent $\nu$ in $\langle k^{(l+1)}\rangle = r^{\nu}\langle k^{(l)}\rangle$, as a function of $\beta$. (c) Rescaled average local clustering per degree class $\bar{c}^{(l)}_{res}(k^{(l)}) = (\bar{c}^{(l)}(k^{(l)}))/\bar{c}^{(l)}$ as a function of the rescaled degrees. (d) Average local clustering coefficient $\bar{c}^{(l)}$ as a function of the layer $(l)$. We display the flow under standard GR with deterministic links (orange squares), GR where links are made probabilistically (green stars), and new independent $\mathbb{S}^1$-realizations created in every layer (blue triangles). In this latter case, the networks size and the average degree match the GR in every layer. The original networks were generated with the $\mathbb{S}^1$-model for $N = 65536$ and $\langle k\rangle = 6$.

In Fig. 7.1, we show the behavior of several network properties in the flow of synthetic scale-free networks generated with the $\mathbb{S}^1$-model. In Fig. 7.1a, the tail of the degree distribution of rescaled degrees $k^{(l)}_{res} = k^{(l)}/\langle k^{(l)}\rangle$ for $\beta = 0.8$ in the quasi-geometric domain is self-similar under renormalization. This self-similarity is also proven analytically in Appendix B.1. For large enough $l$, this self-similarity will always be lost for finite systems like real networks. This is because the finite size induces a cut-off in the degree distribution, rendering its variance finite and therefore leading to the applicability of the central limit theorem, resulting in a Gaussian distribution.

In Fig. 7.1b, we plot the dependence of the exponent $\nu$ characterizing the flow of the average degree as a function of $\beta$. As discussed above, in the region $\beta \leq 1$ no

edges get destroyed in the renormalization flow as the long range nature of links in this regime makes it extremely unlikely that two or more edges connect nodes in the same two supernodes. However, such situations do arise for finite systems, leading to the loss of links along the flow and, thus, to $\nu < 1$, as can be observed in Fig. 7.1b. This finite size effect is stronger the closer to $\beta = 1$, and can therefore be seen as quasi-geometric behavior. When $\beta > 1$, the exponent $\nu$ decreases even further and we enter in the geometric regime described in Ref. [50].

In Fig. 7.1c we display the average local clustering coefficients per degree class, which is again self-similar when rescaled as $\bar{c}^{(l)}_{res}(k^{(l)}) = (\bar{c}^{(l)}(k^{(l)}))/\bar{c}^{(l)}$, where $\bar{c}^{(l)}$ is the average local clustering coefficient. Rescaling is necessary because $\bar{c}^{(l)}$ is not conserved under the RG flow for $\beta \leq 1$. This is confirmed by Fig. 7.1d, where the orange squares represent the evolution of $\bar{c}^{(l)}$ as a function of the renormalization step $l$ for networks at $\beta = 0.8$. This behavior is in contrast to the situation for $\beta > 1$, where $\bar{c}$ only depends on the inverse temperature $\beta$, which is unaffected by the renormalization procedure. For $\beta \leq 1$, clustering depends on the systems size and the average degree [35], which do change under the RG flow.

This result on its own is not in tension with the notion of self-similarity as, granted the network is well described by the $\mathbb{S}^1$-model, a smaller version of a certain network should indeed have a higher clustering coefficient. However, comparing networks obtained through GR (orange squares) and with the $\mathbb{S}^1$-model (blue triangles) in Fig. 7.1d, we see that the flows of $\bar{c}^{(l)}$ do not match. This discrepancy is caused by the fact that the largest contribution to the average local clustering coefficient comes from nodes with small degrees for which self-similarity is not fulfilled, as can be seen in Fig. 7.1a. To prove that the discrepancy does not stem from a lack of congruence with the $\mathbb{S}^1$ connection probability, we repeat the same analysis for networks where the hidden degrees of the supernodes were generated using Eqs. (7.5) and (7.7) but where the connections were made randomly following Eq. (4.26). In Fig. 7.1d, this case is represented by green stars and coincides with the GR flow. The discrepancy, thus, originates in the lack of self-similarity of the hidden degree distribution at small $\kappa$.

## 7.2 THE IMPORTANCE OF GEOMETRY

In this section we will argue that geometry is still important for renormalizing networks with weak geometric coupling. To this end we compare GR with a second scheme that is explicitly non-geometric – supernodes are created by choosing constituent nodes at random. In general, the random scheme will not lead to a conserved connection probability as the proof in Sec. 7.1.1 breaks down because $\Delta\theta^{(l)}_{ij}$ can no longer be assumed equivalent to $\Delta\theta^{(l+1)}_{\sigma\tau}$ as nodes within a single super-node might lie very far apart. However, one might argue that the angular coordinate is irrelevant as the regime $\beta \leq 1$ is, *a priori*, non-geometric in the thermodynamic limit and self-similar network copies could, thus, still be obtainable. As we show below, for finite networks this is only the case for extremely small values of $\beta \lesssim 0.5$, i.e., in the non-geometric regime.

### 7.2.1 *The circular mean*

In the random scheme, the angular coordinate of the supernodes is meaningless by construction. Nevertheless, for convenience, we redefine it such that it represents a proper average even for constituent nodes that lie far away from each other:

$$\theta_\sigma^{(l+1)} = \arg\left(\frac{\sum_{i\in S(\sigma)}(\kappa_i^{(l)})^{\max(1,\beta)} e^{\mathrm{i}\theta_i^{(l)}}}{\sum_{i\in S(\sigma)}(\kappa_i^{(l)})^{\max(1,\beta)}}\right). \tag{7.11}$$

Note that when the spread of the constituent angular coordinates is small, as is the case for GR, it does not matter which of the two definitions of $\theta_\sigma^{(l+1)}$ one takes: Let $\{\theta_1^{(l)},...,\theta_r^{(l)}\}$ be the set of constituent nodes of a supernode $\sigma$ in layer $l+1$, sorted in ascending order and where we assume that $\theta_r^{(l)} - \theta_1^{(l)} \ll 1$. Then we know that $\Delta\theta_{i1}^{(l)} \ll 1\ \forall i$, which allows us to approximate Eq. (7.11) as

$$\begin{aligned}\theta_\sigma^{(l+1)} &\approx \arg\left(\frac{e^{i\theta_1^{(l)}}\sum_{i\in\mathcal{S}(\sigma)}\kappa_i^{\max(1,\beta)}\left(1+\mathrm{i}\Delta\theta_{i1}^{(l)}\right)}{\sum_{i\in S(\sigma)}\kappa_i^{\max(1,\beta)}}\right)\\ &= \arg\left(e^{\mathrm{i}\theta_1^{(l)}}\left(1+\mathrm{i}\overline{\Delta\theta^{(l)}}\right)\right) \approx \arg\left(e^{\mathrm{i}\theta_1^{(l)}}e^{\mathrm{i}\overline{\Delta\theta^{(l)}}}\right) = \theta_1^{(l)} + \overline{\Delta\theta^{(l)}},\end{aligned} \tag{7.12}$$

where in the second step we have defined the weighted average of the angular differences

$$\overline{\Delta\theta^{(l)}} = \frac{\sum_{i\in\mathcal{S}(\sigma)}\kappa_i^{\max(1,\beta)}\left(\theta_i^{(l)}-\theta_1^{(l)}\right)}{\sum_{i\in\mathcal{S}(\sigma)}\kappa_i^{\max(1,\beta)}}, \tag{7.13}$$

which is assumed to be small. Eq. (7.12) can then be rewritten to obtain Eq. (7.7).

### 7.2.2 *The clustering spectrum*

We first study self-similarity of the clustering spectrum as clustering is the key property of geometric graphs due to its relation to the triangle inequality. In Fig. 7.1, it is shown that, for a scale-free synthetic network with $\beta = 0.8$, GR reveals self-similar behavior in the renormalization flow. In Fig. 7.2, we show the results for the randomized coarse-graining scheme. We see that self-similarity is obtained for the smallest $\beta$'s, implying that geometric information is not important here. However, the overlap between the different curves gets progressively worse as $\beta$ increases, reflecting the growing importance of the geometry. The self-similarity is lost at $\beta \approx 0.7$, very close to the theoretical transition point $\beta_c' = 2/\gamma$ between the non- and quasi-geometric regimes [35]. The curves flatten out with $l$, implying that more and more of the clustering in the network is due to high degree nodes. This is to be expected, as the random coarse-graining scheme destroys the coupling of the network to the geometry. This leads to networks that are similar to those generated

with the configuration model, where we know that most of the clustering is due to high degree nodes [144].

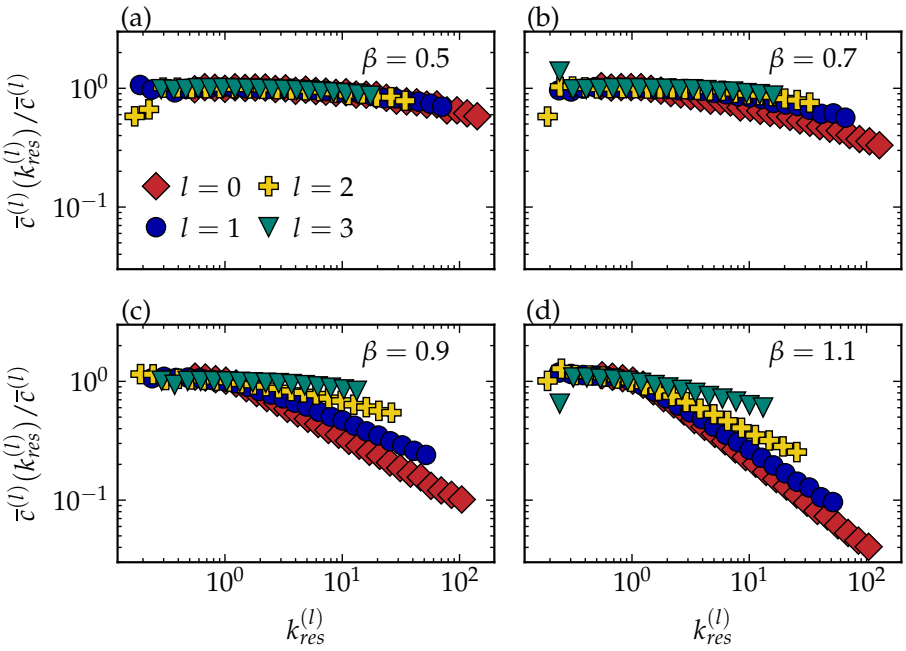



Figure 7.2: The flow of the rescaled average local clustering coefficient per rescaled degree class under the randomized coarse-graining scheme for different $\beta's$: (a) $\beta = 0.5$, (b) $\beta = 0.7$, (c) $\beta = 0.9$, (d) $\beta = 1.1$. Here, $l = 0$ represents the original network and we perform three consecutive renormalization steps with $r = 2$, leading to the cases $l = 1, 2$ and 3. The network parameters used to generate the original networks are $\{N, \gamma, \langle k \rangle\} = \{65536, 2.9, 6\}$.

### 7.2.3 *The empirical connection probability*

To quantify further how much poorer the results of the randomized coarse-graining scheme are in comparison to GR, we measure how well the empirical connection probability of the renormalized network fits the theoretical one in the $S^1$-model. After obtaining the hidden coordinates, the quantity $\chi_{ij} = (R\Delta\theta_{ij})^\beta / (\hat{\mu}\kappa_i\kappa_j)^{\max(1,\beta)}$ is determined for each pair of nodes. These values are binned logarithmically, and for each bin the proportion of links versus non-links is calculated to produce the inferred connection probability $p(\chi)$. The results of this analysis are shown in Fig. 7.3 where we have used networks in the quasi-geometric regime with $\beta = 0.8$. Fig. 7.3a shows the inferred connection probability of the different renormalized layers for the standard GR, where geometric information is used to define the supernodes. In Fig. 7.3b, we see the same results but for the case where the nodes are chosen at random. Clearly, while GR produces self-similar copies congruent

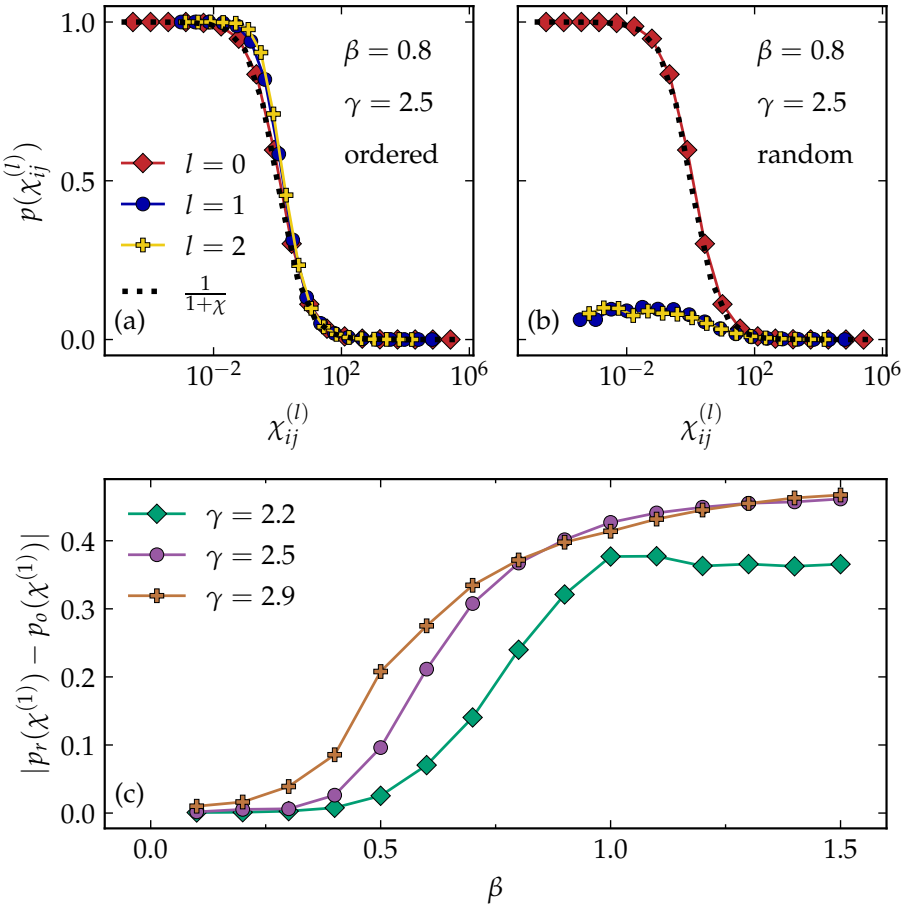



Figure 7.3: (a,b) The flow of the connection probability as a function of $\chi = (R\theta_{ij})^{\beta}/(\hat{\mu}\kappa_i\kappa_j)^{\max(1,\beta)}$ under the RGN where nodes are combined sequentially (a) or randomly (b) along the circle. The dotted lines give the theoretical curve. (c) The mean difference between the two previous cases for $l = 1$ and for three different $\gamma$'s. In all cases the networks were generated with the $S^1$-model with $N = 65536$ and $\langle k \rangle = 6$.

with the $\mathbb{S}^1$ connection probability, the random procedure does not. This confirms that in the quasi-geometric regime geometric information is important even though the geometric coupling is weak.

We plot the average difference between the connection probabilities of the two schemes at layer $l = 1$ as a function of the inverse temperature $\beta$ in Fig. 7.3c. To compute this difference, one first samples parameters $\chi_{ij}^{(l=1)}$ logarithmically. For each of these values, one finds the observed connection probability for the two schemes. One then takes the difference between these cases and averages it over the sampled distances. Once again, three different behaviors can be observed. In the geometric regime ($\beta > 1$), the difference between the two methods is large. For $\beta$'s in the quasi-geometric regime, the difference decreases, and it goes to zero in the non-geometric regime. The transition point between the non- and quasi-geometric regimes shifts to higher betas when the heterogeneity of the network is increased, in line with the theoretical prediction that this transition occurs at $\beta_c' = 2/\gamma$ [35]. The discrepancy between the curves at $\beta > 1$ comes from the fact that not only similarity but also popularity plays a role in the connection probability. As this second type of information is used equivalently in the renormalization procedure regardless of how the angular coordinates are chosen, the difference between these two methods can thus be expected to be smaller when popularity dimensions plays a more important role, which is the case when the degree distribution is more heterogeneous, i.e. when $\gamma$ is smaller.

## 7.3 CONCLUSIONS

In summary, we have extended the geometric renormalization scheme to networks in the weakly geometric regime. We have shown that also in this regime, self-similar scaled-down network replicas can be obtained, where self-similarity refers to important network properties such as the degree distribution and the clustering spectrum. In the quasi-geometric domain $0.5 \lesssim \beta \leq 1$, one must define supernodes by grouping consecutive nodes along the $\mathbb{S}^1$-circle in order to obtain self-similarity in the clustering spectrum and in the connection probability. This underlines the importance of geometric information for understanding the network topology even when the geometric coupling is weak. In constrast, for $\beta \lesssim 0.5$ it does not matter how nodes are grouped. This implies that here the connectivity is solely determined by the degree-distribution, making them effectively non-geometric.

# 8

# EMPIRICAL NETWORKS IN THE WEAKLY GEOMETRIC REGIME

In the previous Chapters, we have seen various confirmations that geometric information is relevant to understanding a network's topology, even when the geometric coupling is relatively weak. For one, clustering decays extremely slowly in the quasi-geometric $\beta_c' < \beta \leq 1$ regime, indicating that high levels of clustering might be a finite size effect. Secondly, the underlying coordinates of artificial $\mathbb{S}^1$-networks can be recovered faithfully by network embedding, even when $\beta$ is low. Finally, for a large range of $\beta \leq 1$, information about the location of nodes in their underlying similarity space is essential for obtaining self-similarity in important network measures when performing graph renormalization.

Even though these results are of clear theoretical interest, it still needs to be shown that they are of importance to real world networks. In this Chapter, which combines results from Refs. [61] and [62] we analyze several empirical data sets from various fields of study. In Sec. 8.1 we show that many of these networks are indeed best described as living in the weakly geometric regime. Then, in Sec. 8.2, we demonstrate that in this regime, faithful network embeddings can be obtained using Mercator. Finally, in Sec. 8.3 these networks are shown to be self-similar under the geometric renormalization group flow.

## 8.1 CLASSIFYING REAL NETWORKS

We first want to understand if the three regions identified in the previous chapters – strongly geometric, quasi-geometric and non-geometric– are also relevant for empirical data. In the case of real networks, the theoretical transition point $\beta_c' = 2/\gamma$ between the non- and quasi-geometric regions is less useful as it is mostly not possible to accurately extract the exponent $\gamma$. We therefore define a new classification of three distinct types of networks. Type I networks are classified by Mercator as being effectively non-geometric, i.e. their clustering can be explained by the configuration model. Type II networks live in the region $\beta \leq 1$, but Mercator is still able to determine their temperature as they have significantly more clustering than one would expect from a network generated by the configuration model. These networks can be considered quasi-geometric. Finally, type III networks are those network that have $\beta > 1$ and, thus, they are strongly geometric.

In Tab. 8.1 we show a selection of real networks and their properties, as well as their classification into the categories described above. There are several real networks for both categories I and II, where $\beta \leq 1$. Note also that the absolute value of the average local clustering is, on its own, not a good indicator for geometric coupling strength. For example, the value of $\bar{c}$ for the WordAdjacency–English network is relatively high, but it is still a type I network. This is because the triangles in the network can also be formed due to the presence of high degree nodes, meaning that this level of average local clustering can also be obtained in the configuration model. Another interesting observation is the presence of several gene regulatory and protein-protein interaction networks in the region $\beta \leq 1$, as well as the fact that several ecological networks are deemed to have extremely weak geometry.

Real networks of type III have been extensively studied in the literature [35, 42, 64, 244]. Living in the region $\beta > 1$, their coupling to the geometry is strong. In general, this leads to relatively high levels of clustering. However, once again exceptions arise. An example is the offline friendship network studied in Ref. [245]. It can be shown that for this network $\beta = 1.3$, squarely in the geometric regime. However, the average clustering $\bar{c} = 0.15$ is lower than that of many type I or II networks.

## 8.2 QUALIFYING THE EMBEDDING

Clearly, the levels of clustering in many real networks from a wide range of scientific field can be best explained by placing them in the weakly geometric regime. However, we still need to examine if the embeddings of these networks in this regime are meaningful; can the inferred hidden variables reproduce the structural properties of the network well? In order to answer this question we study four networks from Tab. 8.1. First, in Fig. 8.1 we study the Word Adjacency network of the English language, which is Type I. Here, nodes represent words which are connected if one directly follows the other in texts. In Fig. 8.2 we the turn to the Genetic Multiplex of S.cerevisiae, otherwise known as baker's yeast, where nodes are genes and the links represent different types of interaction, which are also treated equivalently. Then, in Fig. 8.3 we study the protein protein interaction network of H.sapiens –humans–, with nodes representing genes and the edges physical interactions between then. Finally, we study the MathOverflow network in Fig. 8.4. This is a network of users of the online Q&A site MathOverflow where an edge is created if the users interacted. The final three example networks are all of Type II.

For all four networks we reproduce the complimentary cumulative degree distribution (panel (d)), the clustering spectrum (panel (e)) and the degree-degree correlation spectrum (panel (f)) from the inferred coordinates (panel (a)). We also show the inferred connection probability (panel (c)), which is obtained as follows: After inferring the hidden coordinates, for each pair of nodes the parameter $\chi = x_{ij}^{\beta}/(\hat{\mu}\kappa_i\kappa_j)^{\max(1,\beta)}$ is calculated. These values are then binned logarithmically,

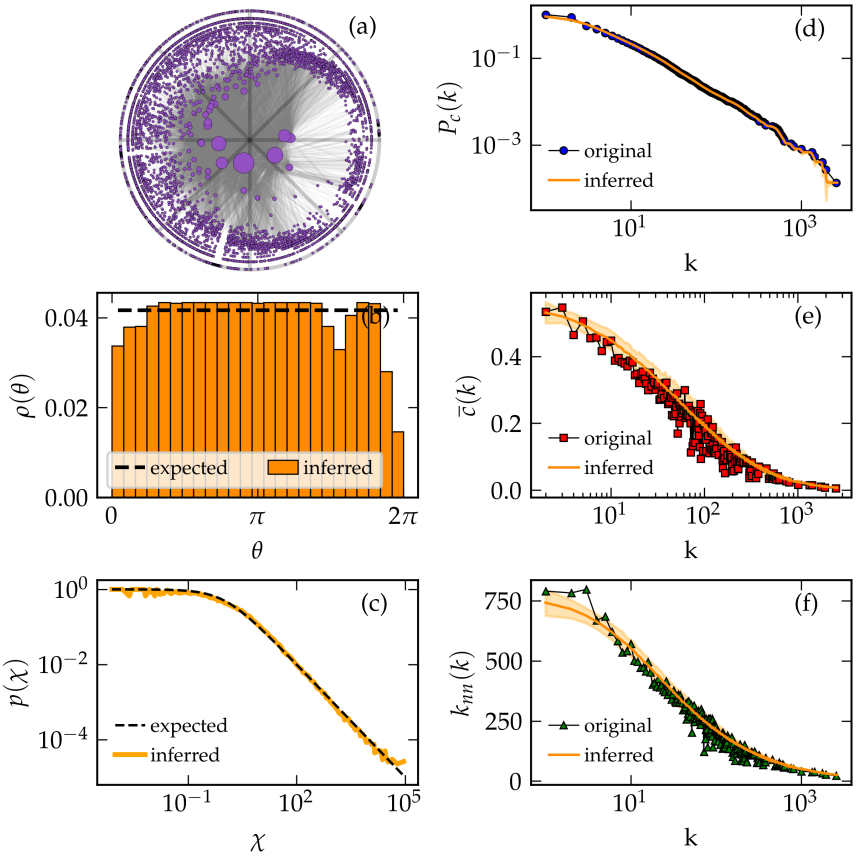



Figure 8.1: Summary of the results of Mercator for the WA–English network. (a) Representation of the embedding in the hyperbolic plane as defined by the $\mathbb{H}^2$-model. The top 20% most geometric edges are shown. (b) Comparison between the expected and inferred densities of nodes along the circle. (c) Comparison between the probability distribution as expected based on the model (expected) as well as the actual distribution based on the inferred coordinates (inferred). The reproduction of the topological properties is also given: (d) the complementary cumulative degree distribution, (e) the average local clustering coefficient per degree class and (f) the degree-degree correlations per degree class. The inferred results are obtained by generating 100 realizations of the $\mathbb{S}^1$-model based on the inferred coordinates. The orange shaded regions represent the $2\sigma$ confidence interval.

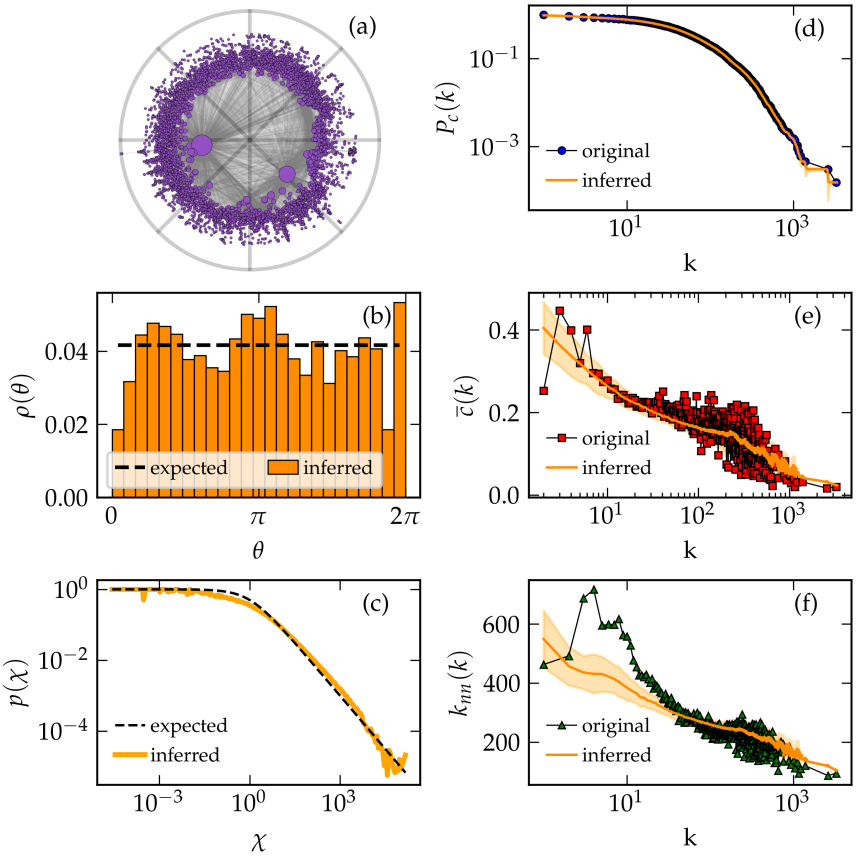



Figure 8.2: Summary of the results of Mercator for the GMP–S.cerevisiae network. (a) Representation of the embedding in the hyperbolic plane as defined by the $\mathbb{H}^2$-model. The top 10% most geometric edges are shown. (b) Comparison between the expected and inferred densities of nodes along the circle. (c) Comparison between the probability distribution as expected based on the model (expected) as well as the actual distribution based on the inferred coordinates (inferred). The reproduction of the topological properties is also given: (d) the complementary cumulative degree distribution, (e) the average local clustering coefficient per degree class and (f) the degree-degree correlations per degree class. The inferred results are obtained by generating 100 realizations of the $\mathrm{S}^1$-model based on the inferred coordinates. The orange shaded regions represent the $2\sigma$ confidence interval.

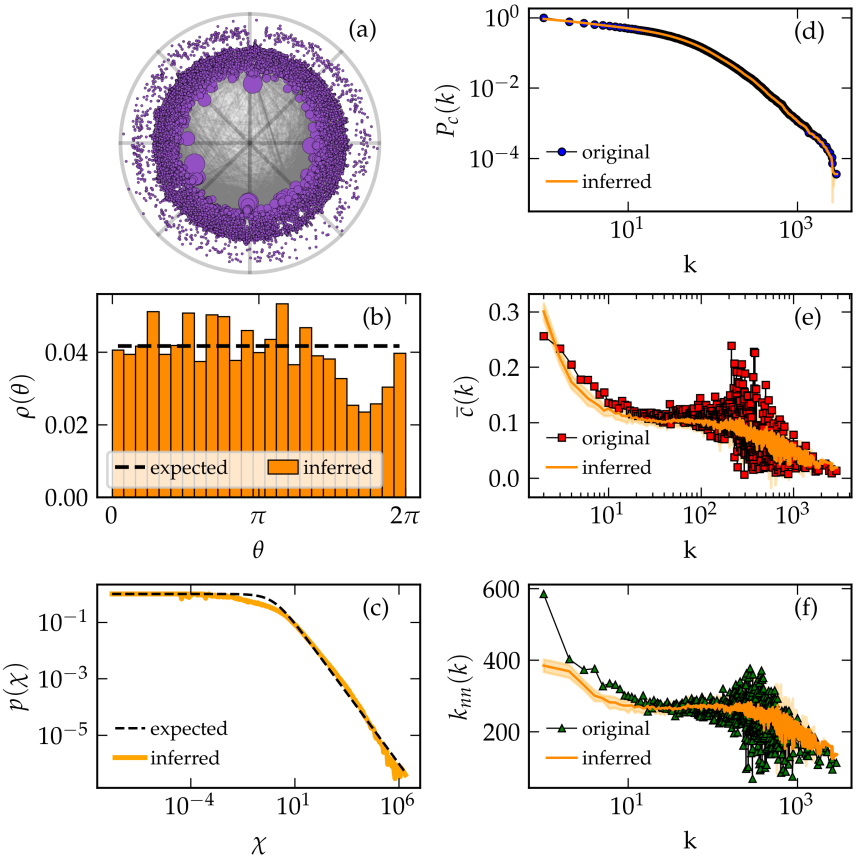



Figure 8.3: Summary of the results of Mercator for the PPI–H.sapiens network. (a) Representation of the embedding in the hyperbolic plane as defined by the $\mathbb{H}^2$-model. The top 4% most geometric edges are shown. (b) Comparison between the expected and inferred densities of nodes along the circle. (c) Comparison between the probability distribution as expected based on the model (expected) as well as the actual distribution based on the inferred coordinates (inferred). The reproduction of the topological properties is also given: (d) the complementary cumulative degree distribution, (e) the average local clustering coefficient per degree class and (f) the degree-degree correlations per degree class. The inferred results are obtained by generating 100 realizations of the $\mathbb{S}^1$-model based on the inferred coordinates. The orange shaded regions represent the $2\sigma$ confidence interval.

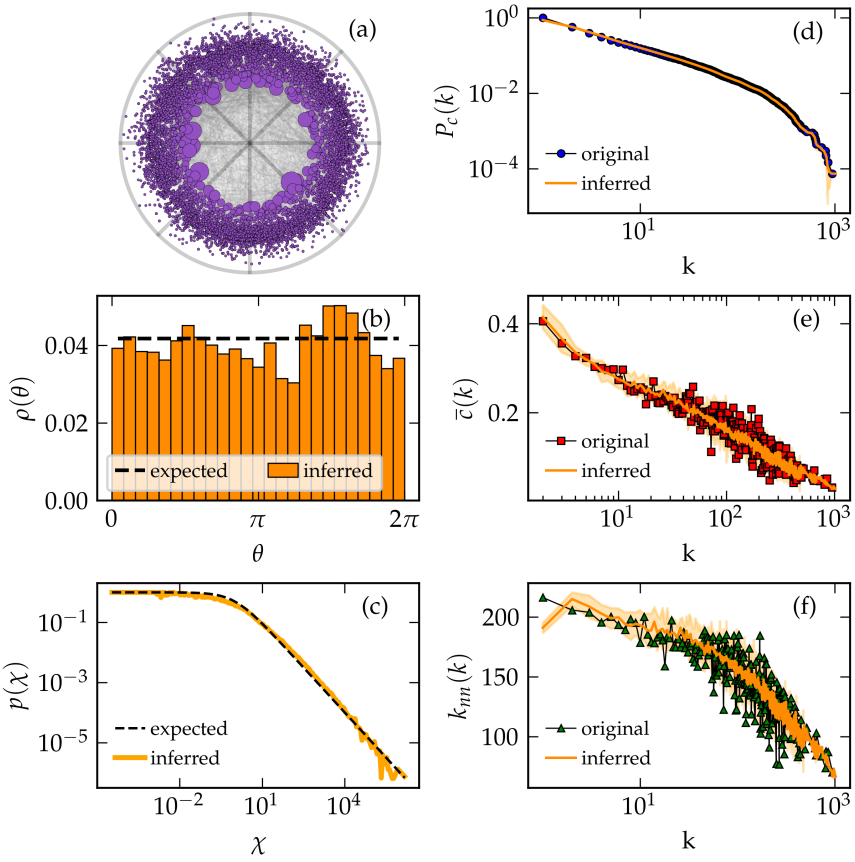



Figure 8.4: Summary of the results of Mercator for the MathOverflow network. (a) Representation of the embedding in the hyperbolic plane as defined by the $\mathbb{H}^2$-model. The top 5% most geometric edges are shown. (b) Comparison between the expected and inferred densities of nodes along the circle. (c) Comparison between the probability distribution as expected based on the model (expected) as well as the actual distribution based on the inferred coordinates (inferred). The reproduction of the topological properties is also given: (d) the complementary cumulative degree distribution, (e) the average local clustering coefficient per degree class and (f) the degree-degree correlations per degree class. The inferred results are obtained by generating 100 realizations of the $\mathbb{S}^1$-model based on the inferred coordinates. The orange shaded regions represent the $2\sigma$ confidence interval.

Table 8.1: Set of real networks classified by type: (I) non-geometric, (II) quasi-geometric. Network properties are also shown. The following abbreviations are used: (WA) WordAdjacency (MB) Metabolic, (GI) Genetic Interactions, (GMP) Genetic Multiplex, (PPI) Protein Protein Interactions, (PoP) Point of Presence. Detailed descriptions of the networks can be found in App. C.1.

| **Network** | **Area** | $N$ | $\langle k \rangle$ | $k_{\max}$ | $\bar{c}$ | $\beta$ | **Type** |
|---|---|---|---|---|---|---|---|
| Foodweb–Eocene | Ecological | 700 | 18.3 | 192 | 0.10 | $\approx 0$ | I |
| Foodweb–Wetland | Ecological | 128 | 32.4 | 110 | 0.33 | $\approx 0$ | I |
| WA–English | Language | 7377 | 12.0 | 2568 | 0.47 | $\approx 0$ | I |
| WA–Japanese | Language | 2698 | 5.9 | 725 | 0.30 | $\approx 0$ | I |
| MB–R.norvegicus | Cell | 1590 | 5.9 | 498 | 0.19 | $\approx 0$ | I |
| WikiTalk–Catalan | Social | 79209 | 4.6 | 53234 | 0.83 | $\approx 0$ | I |
| GI–S.cerevisiae | Cell | 5933 | 149 | 2244 | 0.17 | 0.63 | II |
| GMP–C.elegans | Cell | 3692 | 4.1 | 526 | 0.11 | 0.69 | II |
| Gnutella | Tech | 10876 | 7.4 | 103 | 0.01 | 0.73 | II |
| PPI–S.cerevisiae | Cell | 7271 | 45.0 | 3613 | 0.37 | 0.75 | II |
| PPI–D.melanogaster | Cell | 11319 | 23.7 | 889 | 0.10 | 0.84 | II |
| Transport–London | Transport | 369 | 2.3 | 7 | 0.03 | 0.86 | II |
| GMP–S.cerevisiae | Cell | 6567 | 68.1 | 3254 | 0.22 | 0.88 | II |
| Internet-PoP | Tech | 754 | 2.4 | 7 | 0.03 | 0.90 | II |
| PPI–H.sapiens | Cell | 27578 | 37.9 | 2883 | 0.15 | 0.91 | II |
| WikiVote | Social | 7066 | 28.5 | 1065 | 0.21 | 0.91 | II |
| MathOverflow | Social | 13599 | 10.5 | 949 | 0.32 | 0.99 | II |

and for each bin the proportion of links versus non-links is determined. This then gives the inferred connection probability $p(\chi)$, which we compare with the theoretical form $p(\chi) = 1/(1+\chi)$. Finally, we plot the distribution of nodes along the $\mathbb{S}^1$-circle in panel (panel (b)). This latter result allows us uncover the presence of soft-communities in the network, encoded in the non-uniformity of the distribution.

We observe in all four examples that the structural properties are reproduced well, and that the empirical connection probability matches the theoretical one. Interestingly, the fact that this is also the case for the Word Adjacency network, which is type I, indicates that Mercator can still be used, even though the networks is not geometric. This is a result of the fact that the $\beta \to 0$ limit of the $\mathbb{S}^1$-model is the SCM, where the topology of the networks is entirely determined by the hidden degrees. Mercator is still doing a good job at extracting these, even though the inferred hidden coordinates $\theta$ are meaningless. This latter fact is also the reason we observe a spiral structure in the node placement on the hyperbolic disk in Ref. 8.1a.

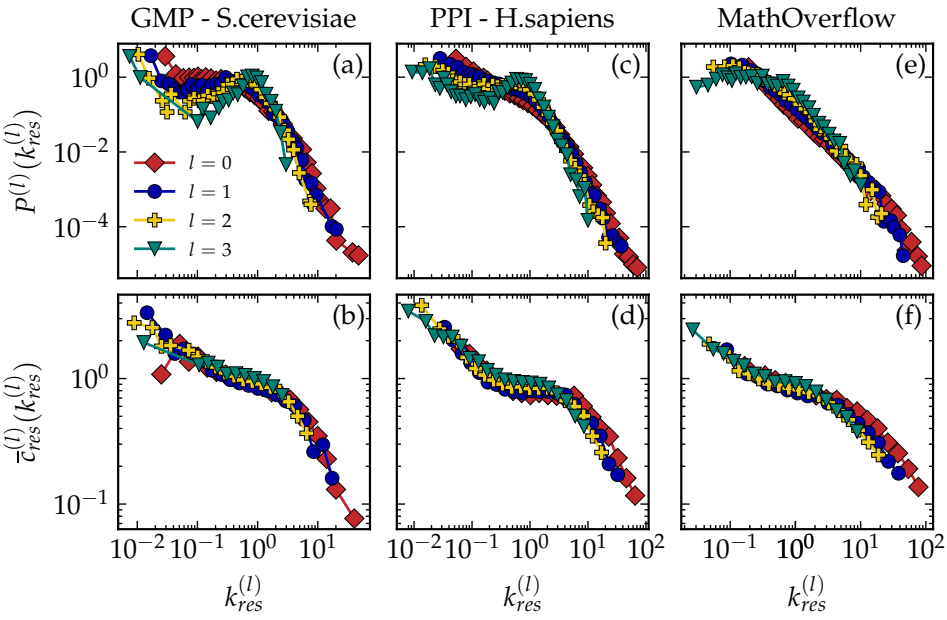



Figure 8.5: The degree distribution and rescaled average local clustering coefficient as functions of the rescaled degree for (a,b) the genetic multiplex of the yeast S.cerevisiae, (c,d) the Human protein-protein interaction network and (e,f) the interaction network of users on the online Q&A site MathOverflow. The details of these networks are given in Appendix C.1.

## 8.3 RENORMALIZING REAL NETWORKS

We now apply the geometric renormalization procedure introduced in Chapter 7 to the type II networks studied above. In Fig. 8.5 the flows of the degree distributions and clustering spectra are shown. The original networks correspond to layer $l = 0$, on which we perform three geometric renormalization steps with $r = 2$ leading to the layers $l = 1, 2$ and 3. In all cases, the curves remain invariant under repeated application of GR. Only for large $l$ does the degree distribution tend to a more homogeneous distribution. Indeed, in Appendix B.1 we showed that there is a maximum amount of renormalization steps one can perform after which the self-similarity of the degree distirbution breaks down due to finite size effects arising from the cutoff $\kappa_c$.

In Fig. 8.6 the networks embeddings on the hyperbolic disk are shown, both for the original network $l = 0$ and the thrice renormalized network $l = 3$. Note that these renormalized networks are $2^3 = 8$ times smaller than the original. We observe that the networks are self-similar, with important features being preserved along the flow. For example, the placement of hubs remains the same, as do the the presence and location of denser areas along the circle, indicating the preservation of soft communities [50]. One can also see that the networks become denser, with

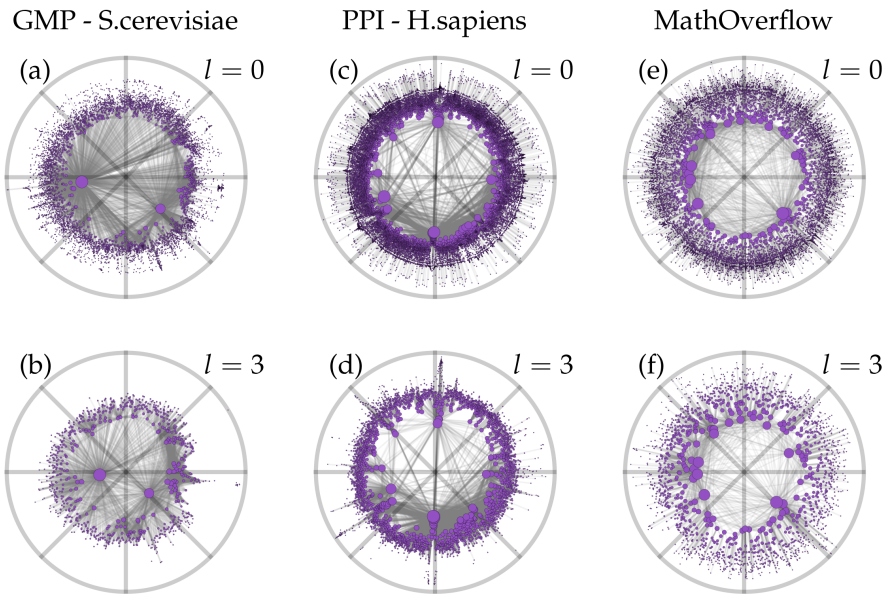



Figure 8.6: The hyperbolic embedding of the original ($l = 0$) and thrice renormalized ($l = 3$) versions of (a,b) the genetic multiplex of the yeast S.cerevisiae, (c,d) the Human protein-protein interaction network and (e,f) the interaction network of users on the online Q&A site MathOverflow. The details of these networks are given in Appendix C.1.

all nodes moving towards the center of the hyperbolic disk. This is an effect of evolution fo the hidden degrees derived in Sec. 7.1.2, and can be mitigated through geometric pruning [42].

## 8.4 CONCLUSIONS

All in all we have shown that the weakly geometric region is important from the point of view of real data, as well. The clustering coefficient of many real networks is best explained by this regime, where they can be faithfully embedded into the $\mathbb{S}^1$-model. This also allows us to renormalize them using the GR procedure, where self similarity is once again observed. The results for the other real networks mentioned in Tab. 8.1 can be found in the Supplementary Informations of Ref. [61] and Ref. [62].

# 9

# LINK OVERLAP AND MUTUAL CLUSTERING IN MULTIPLEX NETWORKS

Previously, we studied the transition from random graphs purely defined by their degree sequence to their geometric counterparts, where the underlying metric space induces clustering through the triangle inequality. We saw that this transition is not sharp, with clustering decaying very slowly in the weakly geometric regime. We also saw that one can define a quasi-geometric regime where geometric information is still relevant, even though the coupling is weak.

All these results were for simple, single layer networks. However, in the development of the science of complex networks it was observed fairly quickly that in many real systems nodes are connected through many different types of interactions [246, 247] – e.g., the same two neighborhoods might be connected by roads, the underground and tram connections. These different interactions lead to different network topologies, and it is often reductive to combine them into a single network. This observation led to the study of multiplex networks, multilayer objects where nodes are shared between layers, but where the connectivity profiles might be very different [248–256].

Even though the individual networks which make up the multiplex are, *a priory*, different, they will share many properties. For example, two important neighborhoods in a city will most likely be connected by all possible means of transportation. In the multiplex of online social networks, two users that are friends on one platform have a high probability of also being connected on another. These correlations lead to the relatively high levels of link overlap observed in real networks [257–264]. These observation has been explained in various different ways; generally the overlap can be thought of as resulting from correlations at the level of the links [265, 266] or at the level of the nodes [47]. Combinations of both these approaches are also possible [267].

In this Chapter we focus on the node-based approach: We assume that the correlations observed in real multiplexes can be understood through the lens of network geometry. In Ref. [47] the geometric framework was extended to multiplex networks, where the correlations between the different layers are encoded at the level of the hidden coordinates of the nodes; the hidden coordinates of one layer are correlated with that of another, leading to similar connectivity patterns. It was shown that these hidden correlations are indeed present in many real systems.

We employ the Geometric Multiplex Model (GMM), introduced in the aforementioned reference, to study the overlap of the different network layers by introducing the mutual network. This network is constructed by placing a link between two nodes $i$ and $j$ if and only if they are connected in all layers of the multiplex. Note that this subnetwork is but a tool for studying the overlap properties of the multiplex as a whole.

By studying the popularity and similarity dimensions separately, we show that only a strong coupling to the geometric similarity space in all layers leads to macroscopic levels of edge overlap; hubs do provide some additional common edges but their reduced number implies an ever sparser mutual network when the geometric coupling is weak. We supplement these results with an analysis of the triangle overlap in the multiplex, quantified by the clustering coefficient in the mutual layer. In contrast to the results on edge overlap, both degree heterogeneity and strong geometric couplings are sufficient to achieve high levels of mutual clustering. We note that in geometric multiplexes, it is generally the links most congruent with the underlying metric space that are preserved in the mutual network, increasing its effective geometric coupling. These results are corroborated in real networks where clustering coefficients generally remain high.

This Chapter is structured as follows: First, in Sec. 9.1 we introduce the geometric multiplex model. We continue by defining the mutual network in Sec. 9.2. Having introduced the tools required for the study of the edge and triangle overlap, in Sec. 9.3 we present analytic results on these quantities for multiplexes with homogeneous degree distributions and perfect correlations in the hidden variables. In Sec. 9.4 we generalize these results to heterogeneous multiplexes. Next, in Sec. 9.5 we study the effect of general coordinate correlations between the various layers. Finally, in Sec. 9.6 we investigate several empirical multiplexes, confirming the observations derived for artificial networks.

## 9.1 THE GEOMETRIC MULTIPLEX MODEL

In Ref. [47] it was noted that the geometric framework employed in this thesis can be extended to multiplexes. When embedding the layers of a real multiplex into the $\mathbb{S}^1$-model, i.e., when inferring the set of hidden variables $\{\theta_i^{(l)}, \kappa_i^{(l)}\}_{i\in V}$ for each layer $l$, correlations between the hidden variables in the different layers were found. Take a node $i$ and consider its angular coordinates $\theta_i^{(1)}$ in layer one and $\theta_i^{(2)}$ in layer two. Both these coordinates encode properties of the node that are important for establishing the connectivity of the respective network layers. Of course, as the layers represent distinct types of interactions, different properties might be relevant, and so the coordinates will most likely not be the same. However, the node is shared by both layers, and represents the same agent in the system. Therefore, it is intuitive for the two coordinates to be correlated. The same line of arguments might be followed to establish correlations between the hidden degrees of a node in the multiplex.

These correlations can be modeled by the GMM. This multiplex variant of the $\mathbb{S}^1$-model starts by drawing angular coordinates from the uniform distribution $\mathcal{U}(0, 2\pi)$ and hidden degrees from the Pareto distribution $\rho(\kappa) \propto \kappa^{-\gamma}$, where $\gamma$ encodes the heterogeneity of the distribution. These are then the assigned hidden variables of the first layer in the multiplex.

The similarity coordinates of the second layer are then drawn with respect to the first such that the correlation between them can be tuned. Take a node $i$ in layer $l = 1$ with angular coordinate $\theta_i$. We then want to draw a new coordinate from a Gaussian centered around this coordinate. Of course, we are working with a spherically symmetric system which implies that the support of this Gaussian cannot be infinite, but is rather $2\pi$. In order for this probability distribution to be normalized we must therefore use instead the truncated Gaussian. In formulas, this can be written as follows: The angular coordinate of node $i$ in layer $l = 2$ can be obtained using

$$\theta_i^{(2)} = \mod\left(\theta_i^{(l)} + \frac{2\pi l_i}{N}, 2\pi\right), \tag{9.1}$$

where $l_i \in [-N/2, N/2]$ is the arc-distance from the original coordinate along the circle with radius $R = N/(2\pi)$. It is drawn from the truncated Gaussian with probability density function

$$f_\sigma(l) = \frac{1}{\sigma} \frac{\phi\left(\frac{l}{\sigma}\right)}{\Phi\left(\frac{N}{2\sigma}\right) - \Phi\left(-\frac{N}{2\sigma}\right)}, \tag{9.2}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ is the probability density function of a standard normal distribution and $\Phi(x)$ is its cumulative distribution function $\Phi(x) = \frac{1}{2}(1 + \operatorname{erf}(x/\sqrt{2}))$. The scaling factor $\sigma$, which is related to the variance of the truncated Gaussian, is defined as

$$\sigma = \sigma_0 \left(\frac{1}{g} - 1\right), \tag{9.3}$$

where $\sigma_0 = \min(100, N/(4\pi))$. It is clear that $g \in [0, 1]$ is the parameter that tunes the correlation between the two layers. When $g \to 0$, $\sigma \to \infty$ and the Gaussian becomes flat. The coordinate of node $i$ in layer $l = 2$ is uniformly sampled, irrespective of its location in the first layer. The correlation between the two layers is therefore zero. On the other extreme, when $g = 1$, $\sigma$ vanishes, which implies that the Gaussian becomes a Dirac delta, leading to $\theta_i^{(2)} = \theta_i^{(1)}$ for all $i$, and therefore perfectly correlated coordinates between the layers.

In the popularity dimension things are slightly more complicated as we have the added constraint that the marginal distribution of hidden degrees in the second layer should still be Pareto with some average degree $\langle k^{(2)} \rangle$ and exponent $\gamma_2$, both of which might or might not be different to the ones in the first layer.

In the original publication Ref. [47], it was shown that this is achieved if the hidden degrees in layer $l = 2$ are drawn from the following cumulative distribution function:

$$F_\nu\left(\kappa_i^{(2)}|\kappa_i^{(1)}\right) = \exp\left(-\left(\varphi_1^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)}\right)^{1-\nu}\right)$$
$$\times\left(\varphi_1^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)}\right)^{-\nu} \frac{\varphi_1^{\nu/(1-\nu)}\kappa_0^{(1)}\left(\kappa_i^{(1)}\right)_1^{\gamma}}{\kappa_0^{(1)}\left(\kappa_i^{(1)}\right)_1^{\gamma} - \kappa_0^{(1)}\kappa_i^{(1)}}, \tag{9.4}$$

where

$$\varphi_l = -\ln\left(1 - \left(\frac{\kappa_0^{(l)}}{\kappa_i^{(l)}}\right)^{\gamma_l - 1}\right). \tag{9.5}$$

It is now the parameter $\nu \in [0, 1]$ that sets the strength of the correlation between the two layers. When $\nu = 0$ it can be shown that Eq. (9.4) reduces to $F_0(\kappa_i^{(2)}) = 1 - \kappa_i^{(2)}\left(\kappa_0^{(2)}\right)^{\gamma_2 - 1}$, which is just the Pareto cumulative density function, and does not depend on $\kappa_i^{(1)}$. The two layers are thus independent. When $\nu = 1$, we see that Eq. (9.4) becomes a step function, leading to $\kappa_i^{(2)} = \kappa_0^{(2)}\left(\kappa_i^{(1)}/\kappa_0^{(1)}\right)^{(1-\gamma_1)/(1-\gamma_2)}$. Here, the correlation between the two layers is perfect; the largest hidden degree in layer $l = 1$ is related to the largest hidden degree in layer $l = 2$ etc.

If more layers are present this process is continued, generating the hidden variables in layer $l$ on the basis of those in layer $l - 1$. In principle one could choose to vary $g$ and $\nu$ parameters, leading to some pairs of layers that are more strongly correlated than others. In this Chapter we choose them identical between all layers.

When all hidden variables are assigned, the nodes in the different layers are connected using connection probability (4.26), where each layer can have a distinct $\hat{\mu}_l$ and $\beta_l$. The process of connecting the nodes introduces no new correlation into the system; each edge is placed independently. This is in contrast to other multiplex models [266, 267] where correlations between the layers are a result of link persistence, where the placement of a link in layer $l$ is conditioned on its presence in layer $m$.

An example of a two layer multiplex generated using the GMM can be seen in Fig. 9.1, where $G_1$ and $G_2$ represent the two layers. We see that the locations of the nodes on the hyperbolic disk are highly correlated, only shifting slightly between layers. The third layer $\widetilde{G}$ represents the mutual network, which is constructed with the links that are present in both layers of the multiplex, and which will be the object of study in this Chapter.

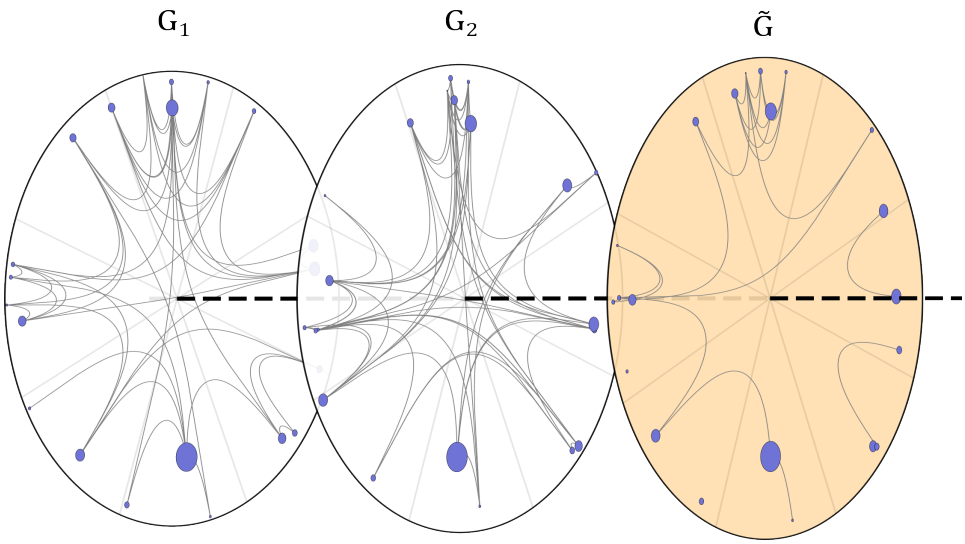



Figure 9.1: An example of a geometric multiplex with two layer $G_1$ and $G_2$. The mutual network $\widetilde{G}$ is also shown.

## 9.2 THE MUTUAL NETWORK

### 9.2.1 *The ensemble*

Say our multiplex is made up of $L$ layers, which we denote as $G_1, G_2, ..., G_L$. These networks $G_l$ are all single layer networks generated with the $\mathbb{S}^1$-model, each with an associated inverse temperature $\beta_l$, chemical potential $\hat{\mu}_l$ and adjacency matrix $\boldsymbol{A}^{(l)}$. We now define the mutual network $\widetilde{G}$ through the element-wise product of these adjacency matrices

$$\widetilde{A}_{ij} = \left[\bigodot_{l=1}^{L} \boldsymbol{A}^{(l)}\right]_{ij} = \prod_{l=1}^{L} A_{ij}^{(l)}, \tag{9.6}$$

where $\odot$ represents the element-wise Hadamard product. This means that a link is only placed between two nodes $i$ and $j$ in the mutual network $\widetilde{G}$ if these two nodes are connected in all $L$ single layer networks $\{G_l\}_{l=1}^{L}$ that make up the multiplex.

Note that different individual network realizations can lead to the same mutual network. This happens because it does not matter whether a link is absent in a single layer or in all layers, the effect is the same: No link will be present between these nodes in the mutual graph. The probability of a certain mutual adjacency matrix will then be given by

$$P(\widetilde{\boldsymbol{A}}) = \sum_{\widetilde{\boldsymbol{A}}=\bigodot_{l=1}^{L} \boldsymbol{A}^{(l)}} P(\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, ..., \boldsymbol{A}^{(l)}) \tag{9.7}$$

The single layer networks are independent at the level of the edges; all possible correlations are mitigated at the level of the hidden variables. Therefore, the

probability of a certain mutual network, given a certain distribution of hidden variables, can be factorized

$$P(\widetilde{\boldsymbol{A}}) = \sum_{\widetilde{\boldsymbol{A}}=\bigodot_{l=1}^{L}\boldsymbol{A}^{(l)}} \prod_{l=1}^{L} \left( \prod_{i<j} \left(p_{ij}^{(l)}\right)^{A_{ij}^{(l)}} \left(q_{ij}^{(l)}\right)^{1-A_{ij}^{(l)}} \right), \tag{9.8}$$

where we assume that the connection probability is pairwise, as in the case of the $\mathrm{S}^1$-model, and have introduced $q_{ij}^{(l)}$, the probability that a link is not present between nodes $i$ and $j$ in layer $l$. Finally, this equation can be rewritten as

$$P(\widetilde{\boldsymbol{A}}) = \prod_{i<j} \left(\widetilde{p}_{ij}\right)^{\widetilde{A}_{ij}} \left(\widetilde{q}_{ij}\right)^{1-\widetilde{A}_{ij}}, \tag{9.9}$$

where the mutual link probability is given by $\widetilde{p}_{ij} = \prod_{l=1}^{L}(p_{ij}^{(l)})$. The probability that a link is not present can trivially be given by $\widetilde{q}_{ij} = 1 - \widetilde{p}_{ij}$. This probability has many contributions: there are $2^L$ possible situations for a single node pair as in each layer a link can be either present or absent. Of these possibilities, $2^L - 1$ contribute to the probability of the link being absent in the resulting mutual graph.

### 9.2.2 *Multiplex hidden variable models*

Now that we have defined the mutual graph ensemble in very general terms, we turn to the structural properties of these graphs, in particular the amount of links and the average local clustering coefficient. To this end we view the system as a hidden variable model, such that the results from Ref. [141] can be adapted. Each node has a $L$-dimensional vector $\boldsymbol{h}$ associated to it, where the entries of the vector represent the different layers. For the two-layer, homogeneous GMM this means that $\boldsymbol{h} = \{\theta^{(1)}, \theta^{(2)}\}$. In the following, we assume the entries to be scalars for notational simplicity. However, the results are trivially extendable to higher dimensional entries, required for example in the case of the heterogeneous GMM, where also hidden degrees are present.

#### 9.2.2.1 *The average mutual degree*

Take two nodes with associated hidden variable vectors $\boldsymbol{h}$ and $\boldsymbol{h'}$. The probability that they are connected, here denoted by $\boldsymbol{h} \sim \boldsymbol{h'}$, is given by the mutual connection probability

$$\widetilde{p}(\boldsymbol{h}, \boldsymbol{h'}) \equiv P(\boldsymbol{h} \sim \boldsymbol{h'}) = \prod_{l=1}^{L} p_l(h^{(l)}, h'^{(l)}), \tag{9.10}$$

where we have used the fact that the edge placement is uncorrelated.

To obtain the total amount of edges, we just need to marginalize over the hidden variables $\boldsymbol{h}, \boldsymbol{h'}$ and multiply by $\binom{N}{2}$ as all nodes are, *a priori*, identical. To perform this marginalization we need the probability density function of the hidden variable vector $\rho(\boldsymbol{h})$. In line with the GMM, we assume that the assignment

of hidden variables in the different layers is Markovian; the hidden variables of layer $l+1$ only depend on layer $l$. This allows us to write the desired probability density function as

$$\rho(\boldsymbol{h}) = \rho(h^{(1)}) \prod_{l=2}^{L} \rho\left(h^{(l)}|h^{(l-1)}\right) \tag{9.11}$$

Finally, we obtain for the expected amount of links the following expression

$$\langle \widetilde{M} \rangle = \binom{N}{2} \iint d^L \boldsymbol{h} d^L \boldsymbol{h'} \rho(\boldsymbol{h}) \rho(\boldsymbol{h'}) \widetilde{p}(\boldsymbol{h}, \boldsymbol{h'}), \tag{9.12}$$

where we have introduced the $L$ dimension integral measure $d^L\boldsymbol{h} = \prod_{l=1}^{L} dh^{(l)}$. When we have perfect correlations and $\rho(h^{(l)}|h^{(l-1)}) = \delta(h^{(l)} - h^{(l-1)})$ this reduces to

$$\langle \widetilde{M} \rangle = \binom{N}{2} \iint dh dh' \rho(h) \rho(h') \prod_{l=1}^{L} p_l(h, h'). \tag{9.13}$$

On the other extreme, when no correlations are present and $\rho(h^{(l)}|h^{(l-1)}) = \rho(h^{(l)})$, Eq. (9.12) can be rewritten as

$$\langle \widetilde{M} \rangle = \binom{N}{2}^{1-L} \prod_{l=1}^{L} \langle M^{(l)} \rangle. \tag{9.14}$$

#### 9.2.2.2 *The average local clustering coefficient*

The average local clustering coefficient is defined as the probability that two neighbors of a node are also connected to each other. Say we have a node with hidden variable vector $\boldsymbol{h}$ connected to two nodes with $\boldsymbol{h'}$ and $\boldsymbol{h''}$, respectively. Given this situation, the probability of the triangle being closed is given by Eq. (9.10), the mutual connection probability.

The first step in deriving the average local clustering coefficient is finding the expected clustering coefficient of a node with hidden variable vector $\boldsymbol{h}$. To this end we need to marginalize the mutual connection probability over $\boldsymbol{h'}$ and $\boldsymbol{h''}$, which requires the probability that a node with hidden variable $\boldsymbol{h}$ is connected to a node with $\boldsymbol{h'}$ (and equivalently to a node with $\boldsymbol{h''}$). In Ref. [141] this was derived to be

$$P(\boldsymbol{h} \sim \boldsymbol{h'}|\boldsymbol{h}) = \frac{\rho(\boldsymbol{h'}) \widetilde{p}(\boldsymbol{h}, \boldsymbol{h'})}{\int d^M \boldsymbol{h''} \rho(\boldsymbol{h''}) \widetilde{p}(\boldsymbol{h}, \boldsymbol{h''})}, \tag{9.15}$$

the probability that a randomly chosen node has hidden variables $\boldsymbol{h'}$ times the probability that the node $\boldsymbol{h}$ is connected to it, where normalization has been taking into account. This can be rewritten as

$$P(\boldsymbol{h} \sim \boldsymbol{h'}|\boldsymbol{h}) = \frac{\rho(\boldsymbol{h'}) \widetilde{p}(\boldsymbol{h}, \boldsymbol{h'})}{N^{-1} \mathbb{E}(k|\boldsymbol{h})}, \tag{9.16}$$

where $\mathbb{E}(k|\boldsymbol{h})$ is the expected degree of a node with hidden variables $\boldsymbol{h}$. Marginalizing over the hidden variables of the neighbors we obtain

$$\mathbb{E}(c|\boldsymbol{h}) = \frac{\iint d^L \boldsymbol{h'} d^L \boldsymbol{h''} \rho(\boldsymbol{h'}) \rho(\boldsymbol{h''}) \widetilde{p}(\boldsymbol{h}, \boldsymbol{h'}) \widetilde{p}(\boldsymbol{h}, \boldsymbol{h''}) \widetilde{p}(\boldsymbol{h'}, \boldsymbol{h''})}{N^{-2} \mathbb{E}(k|\boldsymbol{h})^2}. \tag{9.17}$$

Finally, marginalizing over $\boldsymbol{h}$ we obtain the average local clustering coefficient

$$\bar{c} = \int d^L \boldsymbol{h}\, \mathbb{E}(c|\boldsymbol{h}). \tag{9.18}$$

## 9.3 HOMOGENEOUS LAYERS WITH CORRELATIONS

Let us first set the correlations between the hidden variables of the constituent networks to be perfect. All nodes lie in a common similarity space and have a single set of hidden degrees, which we assume to be constant for all nodes; the networks are homogeneous. Different networks are generated that represent the different layers of the multiplex. Note that these networks need not share the same $\beta$ and $\hat{\mu}$ parameters. Different layers can therefore have different levels of clustering and average degrees. Finally, as explained in Sec. 9.2, we remove all links that are not shared by all layers to obtain the mutual network.

We focus on this setting for several reasons. The first is practical; it is where analytical calculations are generally possible. However, studying the homogeneous case also allows us to isolate the effect of similarity, so that it is not distorted by effects from the popularity dimension. Finally, focusing on strong correlations gives the clearest view on its effect on the mutual network.

Following the structure of Chapter 5, we first investigate the statistical properties of mutual graph ensemble, focusing on the entropy density. We then turn to the link overlap, obtaining analytic results on its scaling behavior with the system size. Finally, we study the triangle overlap and find an extended region of constant clustering.

### 9.3.1 *Statistical properties of the ensemble*

First, we want to understand if the mutual network is maximally random, i.e. if the entropy of the ensemble is maximal under the proper constraints. We then need to study the properties of this entropy for different combinations of $\beta$. As in the case of the single layer network [35], one might expect it to diverge for low $\beta$ where the accessible phase space grows very fast, indicating a phase transition in the mutual clustering coefficient.

In the mutual network, a link is present if and only if it exists in all layers simultaneously. Using the Fermi-Dirac form of the single layer connection probabilities introduced in Sec. 4.2.1, this leads to

$$\widetilde{p}_{ij} = \prod_{l=1}^{L} \left( \frac{1}{1 + e^{\beta_l(\epsilon_{ij} - \mu_l)}} \right). \tag{9.19}$$

This connection probability does, *a priori*, not define an ensemble with Gibbs measure $P(\widetilde{G}) = \exp(-\boldsymbol{\lambda} \cdot \widetilde{\boldsymbol{H}})/\widetilde{Z}$ for some meaningful Hamiltonian $\widetilde{\boldsymbol{H}}$, Lagrange multipliers $\boldsymbol{\lambda}$ and partition function $\widetilde{Z}$. For example, we could have expected the total energy $\widetilde{E} = \sum_{i<j} \epsilon_{ij} \widetilde{A}_{ij}$ and total amount of links $\widetilde{M} = \sum_{i<j} \widetilde{A}_{ij}$ to be constraints of the system, such that $\widetilde{\boldsymbol{H}} = \{\widetilde{E}, \widetilde{M}\}$ and $\boldsymbol{\lambda} = \{\beta, -\beta\mu\}$. Eq. (9.9)

cannot be written in this form, implying that the ensemble is not maximally random with respect to these constraints; networks with the same total energy and amount of links might not be equally likely. This was to be expected, as the probability of a certain $\widetilde{G}$ is still conditioned on the non-mutual links, which do not affect the total energy of the mutual network.

Even though the entropy might not necessarily be maximized, we can still find its scaling behavior. To this end, let us note that for a single layer network we have the following relation

$$p_{ij} \sim p'_{ij} \equiv \min\left(1, e^{-\beta(\epsilon_{ij}-\mu)}\right), \tag{9.20}$$

i.e., that the $\mathbb{S}^1$ and GIRG (see Sec. 4.4) connection probabilities asymptotically scale in the same way, a fact that was proven in Ref. [176]. This means that when studying scaling relations we can always work with $p_{ij}$ and $p'_{ij}$ interchangeably. We assume that the same holds true for the "hole" probability: The probability of a connection not being present scales as $q_{ij} \sim q'_{ij} \equiv \min\left(1, e^{\beta(\epsilon_{ij}-\mu)}\right)$. For the mutual connection probability this implies

$$\widetilde{p}_{ij} \sim \widetilde{p}'_{ij} \equiv \prod_{l=1}^{L} \min\left(1, e^{-\beta_l(\epsilon_{ij}-\mu_l)}\right). \tag{9.21}$$

Interestingly, for the two layer network, when $\mu_1 \simeq \mu_2 \simeq \mu$, this reduces to $\widetilde{p}'_{ij} = \min\left(1, e^{-(\beta_1+\beta_2)(\epsilon_{ij}-\mu)}\right)$ for all node pairs, which is the form obtained when studying a single layer network with $\beta' \equiv \beta_1 + \beta_2$ and $\mu' \equiv \mu$. Thus, in this case, the mutual connection probability approximately defines a maximally random ensemble with effective temperature $\beta'$.

Returning to the general case, in order to find an expression for the scaling of the entropy, one needs to find an approximate expression for a link being absent in the mutual network. Previously we argued that this probability has $2^L - 1$ contributions; any non-zero amount of links being absent between two nodes in the multiplex leads these nodes to be disconnected in the mutual network. All these contributions can be written as products of $p'_{ij}$ and $q'_{ij}$, just like in Eq. (9.21). The probability $\widetilde{q}_{ij}$ will then scale as the sum of $2^L - 1$ distinct products. As we are only interested in the scaling behavior we only keep that term that has the largest scaling exponent.

In the (grand) canonical ensemble, the entropy of the network can be seen as that of $\binom{N}{2}$ independent Bernouilli trials with success probability $\widetilde{p}_{ij}$. When $N$ is large and energies are known to be drawn from some distribution $\rho(\epsilon)$, the total amount of entropy can then be written as

$$\widetilde{S} = \binom{N}{2} \int_{-\infty}^{\epsilon_{\max}} d\epsilon \rho(\epsilon) \left[\widetilde{p}(\epsilon) \ln\left(\widetilde{p}(\epsilon)\right) + \left(1 - \widetilde{p}(\epsilon)\right) \ln\left(1 - \widetilde{p}(\epsilon)\right)\right], \tag{9.22}$$

just like in Chapters 3.3.5 and 5.1. In general, we are interested in the entropy per link, or energy density of the system $\widetilde{s} \equiv \widetilde{S}/\langle \widetilde{M} \rangle$.

Let us now take $L = 2$, and assume, without loss of generality, that $\beta_1 \leq \beta_2$. If we then also take $\mu_1 \leq \mu_2$ (the results are qualitatively the same in the opposite case), we show in Appendix D.1.2 that

$$\widetilde{s} \simeq \begin{cases} \xi_1 & \text{if } 1 \leq \beta_1 \leq \beta_2 \\ \beta_1(\mu_2 - \mu_1) + \xi_2 & \text{if } 1 - \beta_2 \leq \beta_1 \leq 1 \\ \beta'\epsilon_{\max} - \beta_1\mu_1 - \beta_2\mu_2 + \xi_3 & \text{if } \beta' \leq 1, \end{cases} \tag{9.23}$$

where $\beta' \equiv \beta_1 + \beta_2$ and $\{\xi_i\}_{i=1}^3$ are constants that only depend on $\beta_1$ and $\beta_2$. We see that the entropy density is finite, independently of the chemical potentials, when both layers are strongly geometric. When $G_1$ becomes weakly geometric, the scaling behavior depends on the choice of $\mu_1$ and $\mu_2$. If both are constant, the entropy only diverges when $\beta' \leq 1$. However, when $\beta_l \leq 1$, a constant $\mu_l$ leads to a dense single layer graph $\langle M_l \rangle \sim N^2$, which is not realistic. If $\mu_l \simeq \alpha_l \ln N$ for some real constants $\alpha_l \leq 1$, necessary for a constant single layer average degree, the entropy density will in principle diverge in this regime. Only when $\alpha_1 = \alpha_2$ does the first term in the expression of $\widetilde{s}$ for $1 \leq \beta_1 + \beta_2 \equiv \beta'$ vanish, leading to a finite entropy density also there. This is in line with our observation that in this case the multiplex can be mapped to a single layer network with effective temperature $\beta'$ and chemical potential $\mu' \simeq \mu_1 \simeq \mu_2$. As $\epsilon_{\max} \sim \ln N$, the only way $\widetilde{s}$ can remain finite in the final regime is for $\beta_1\alpha_1 + \beta_2\alpha_2 = \beta_1 + \beta_2$. This is only possible when $\alpha_1 = \alpha_2 = 1$, leading to two dense single layer networks with $\langle M_l \rangle \sim N^2$.

In the rest of this paper we shall assume to be working with sparse single layer graphs, such that $\mu_l \simeq \min(0, 1 - 1/\beta_l) \ln N + N^0$. In this case, a finite entropy density can be achieved for $1 - \beta_2 \leq \beta_1 \leq 1$ only when $\beta_1 = \beta_2 = \beta$. Here, $\alpha_l = 1 - 1/\beta$ for $l = 1, 2$, and the transition point where $\widetilde{s}$ diverges will lie at $\beta' = 2\beta$.

Plugging this scaling behavior of $\mu_1$ and $\mu_2$ into Eq. (9.23), we note that for all choices of parameters the entropy density can be written as $\widetilde{s} \simeq \sigma \ln N + \tau$, where $\sigma$ and $\tau$ are constants. In Fig. 9.2a, the scaling behavior of the entropy encoded by the constant $\sigma$ is given. We also solve Eq. (9.22) numerically for $N \in [10^6, 10^9]$ using the full expression of $\widetilde{p}(\epsilon)$ as given by Eq. (9.19). There is a clear match between the two approaches, with only minor deviations due to finite size effects, justifying the use of the approximation.

In Fig. 9.2a, several observations can be made: First, we note that the entropy density is indeed only finite ($\sigma = 0$) when $\beta_1, \beta_2 > 1$ or when $\beta_1 = \beta_2 > 1/2$. Second, we see that there is an abrupt change in behavior at $\beta_1 + \beta_2 = 1$: Above this threshold, $0 \leq \sigma < 1$, whereas below it there is a sudden, discontinuous increase to $\sigma = 2$. Interestingly, in this region the entropy density of the mutual network becomes exactly the sum of that of the individual layers as found in Chapter 5.1. This is reminiscent of the fact that the joint entropy of two independent random variables is given by the sum of the individual entropies when the random variables are independent. In our case something similar happens: The

correlations in our system are due to the shared hidden variables, but these no longer matter when $\beta_1 + \beta_2 \leq 1$ as both networks are coupled extremely weakly to their respective underlying metric space.

### 9.3.2 *Link overlap and the average degree*

The average degree of the mutual network created from a multiplex with $L$ layers can be found using Eq. (9.12). In Appendix D.2.1 we find the scaling behavior for general $L$; here, we focus on the specific case $L = 2$, derived in Appendix D.1.1. As was the case for the entropy density, the scaling behavior can be divided into three cases:

$$\langle \widetilde{k} \rangle \simeq \begin{cases} \xi_1 e^{\mu_1} & \text{if } 1 \leq \beta_1 \leq \beta_2 \\ \xi_2 e^{(1-\beta_1)\mu_2 + \beta_1 \mu_1} & \text{if } 1 - \beta_2 \leq \beta_1 \leq 1 \\ \xi_3 e^{(1-\beta')\epsilon_{\max} + \beta_1 \mu_1 + \beta_2 \mu_2} & \text{if } \beta' \leq 1, \end{cases} \tag{9.24}$$

where we have, once again, assumed $\beta_1 \leq \beta_2$ and $\mu_1 \leq \mu_2$. Plugging in the expressions for $\mu_1$ and $\mu_2$ that lead to sparse single layer graphs, we obtain the scaling behavior in Fig. 9.2b, where $\sigma$ is defined through $\langle \widetilde{k} \rangle \sim N^{\sigma}$. It can be seen that a sparse mutual graph ($\sigma = 0$) is only possible when both constituent networks are strongly geometric. In this case, links in both layers are mostly connected to their neighbors in similarity space, leading to high link overlap and a high average mutual degree. When either of the constituent graphs becomes weakly geometric, the situation changes: $\sigma \leq 0$ and the mutual network sparsifies, eventually leading to a disconnected graph when the average mutual degree drops below the percolation threshold. Finally, when $\beta_1 + \beta_2 \leq 1$ and the geometric coupling of both networks is extremely weak, we reach the plateau $\sigma = -1$. Here, in the thermodynamic limit, the amount of links is constant. This is in line with the results found in Ref. [266], where the link overlap of two Erdös-Renyi graphs are studied; the effect of geometry is effectively lost in this region and the two networks are no longer correlated through their hidden coordinates.

An interesting question to ask is whether a constant average mutual degree is possible if we let go of the requirement that the constituent networks $G_1$ and $G_2$ are sparse, i.e. if we revert back to the general case of $\mu_i \simeq \alpha_l \ln N$ for $l \in \{1, 2\}$. For $\beta' \leq 1$, this would require $1 + \beta_1(\alpha_1 - 1) + \beta_2(\alpha_2 - 1) = 0$. Say we take $G_1$ to be sparse such that $\alpha_1 = 1 - 1/\beta_1$. This then requires us to set $\alpha_2 = 1$, leading to $\langle k_2 \rangle \sim N$, i.e. $G_2$ being dense. This can easily be understood by noting that combining any graph $G$ with the complete graph implies that $\widetilde{G} = G$. In the intermediate regime $1 - \beta_2 \leq \beta_1 \leq 1$, the two chemical potentials need to be set such that $\alpha_2 - \alpha_1 = 1/\beta_1$. Taking $G_1$ to be sparse once more leads to $\alpha_2 = 1$, which again implies a dense $G_2$. Note that the results for the intermediate region assume that $\mu_1 \leq \mu_2$; it can be shown that in the opposite case, and assuming that $G_2$ is sparse, intermediate results can be obtained where $0 < \alpha_1 \leq 1$. This then still leads to a dense graph $G_1$ when this is understood to mean a $\langle k_1 \rangle \sim N^{\sigma_{k_1}}$ with

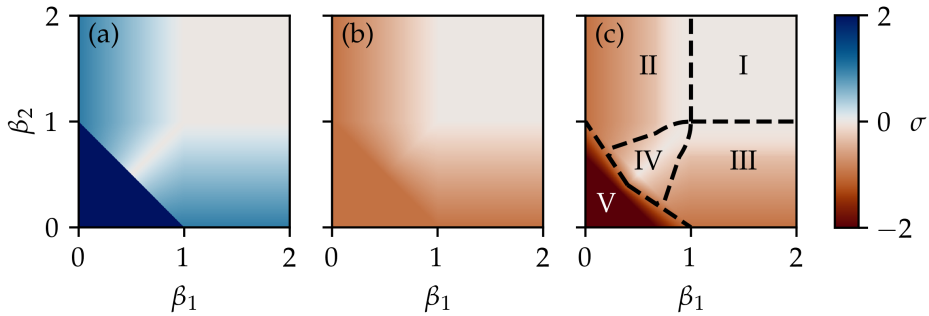



Figure 9.2: The scaling factors $\sigma$ as defined by (a) $\widetilde{s} = \sigma \ln N + \tau$, (b) $\langle \widetilde{k} \rangle \sim N^{\sigma}$ and (c) $\overline{\widetilde{c}} \sim N^{\sigma}$ as a function of the two single layer geometric couplings $\beta_1$ and $\beta_2$. In panel (c) we also define five different regions of parameter space based on how the mutual clustering coefficient relates to its single layer counterparts.

$0 < \sigma_{k_1} \leq 1$. Sparsity in one graph thus always needs to be compensated with density in the other in order to increase link overlap sufficiently to end up with a constant average mutual degree.

### 9.3.3 *The clustering coefficient*

In Chapter 5 we found that in the single layer case, the divergence of the entropy at $\beta_c = 1$ indicates a phase transition in the clustering behavior of the network. In order to find out if this is also the case in the multiplex setting, we investigate the finite size scaling behavior of the clustering coefficient.

For $L$ sparse, homogeneous, perfectly correlated layers with general $\{\beta_l\}_{l=1}^{L}$ and $\{\langle k \rangle_l\}_{l=1}^{L}$, Eq. (9.18) reduces to

$$\overline{\widetilde{c}} = \left( \frac{N}{\langle \widetilde{k} \rangle \pi} \right)^2 \int_0^{\pi} d\theta' \int_0^{\theta'} d\theta'' \prod_{l=1}^{L} f_l(\theta', \theta''), \tag{9.25}$$

where

$$f_l(\theta', \theta'') \equiv \frac{1}{1 + (\zeta_l \theta')^{\beta_l}} \frac{1}{1 + (\zeta_l \theta'')^{\beta_l}} \frac{1}{1 + (\zeta_l (\theta' - \theta''))^{\beta_l}}, \tag{9.26}$$

which implies that in all layers all three constituent links of a triangle should be present. In Appendix D.2.2 we find that the scaling exponents for $\overline{\widetilde{c}}$.

The results of $L = 2$ are shown in Fig. 9.2c. We can distinguish five regions: In region I, $\beta_1, \beta_2 > 1$ and so both layers are highly clustered. This behavior is carried over to the mutual network, where clustering remains finite. In region II, clustering is weaker than in layer $G_1$, where the geometry coupling is strong, but stronger than in layer $G_2$, where geometry is weak. This can be understood as follows: $G_1$ has many short range links, and few long ones, whereas in $G_2$, all possible length links are present. This means that in the mutual network only the short range links will be present, leading to higher levels of clustering than in $G_2$. However, clustering still decreases with $N$ as the mutual network is becoming

ever sparser. Hence, it is relatively unlikely that there exist groups of connected nodes for which at least three have a degree $k \geq 2$, a necessary requirement for a triangle to be formed. The same goes for region III. In region V, both networks are very weakly geometric, and there is very little link overlap. Therefore, the mutual clustering is very small.

Finally, in region IV something interesting happens: The mutual clustering is larger than in both the constituent networks $G_1$ and $G_2$. Here, both networks are in the weakly geometric regime with $\beta \lesssim 1$. Therefore, even though many long range links are present, the connection to the underlying geometry is still strong enough to cause a relatively large amount of short range ones. These are the ones that get translated to the mutual network: The relative phase space for short range links is much smaller than for long range ones, meaning that the probability of a short range link being present in both networks is much larger. This leads to a relatively "geometric" mutual network and high levels of clustering.

This effect is strongest when $\beta_1 = \beta_2 \geq 1/2$ where the average local clustering coefficient remains finite for all $N$. This is in line with our observation that for these values the network can be mapped to a single layer network with $\beta' = 2\beta_1 = 2\beta_2$. Therefore, this region can be seen to lie above the single layer critical point at $\beta' = 1$, implying a constant $\bar{\bar{c}}$. The fact that the entropy density also does not diverge in this region could also have predicted this result.

In the case of $L$ layers, we show in Appendix D.2.2 that the region of parameter space where clustering is constant is extended even further. Ordering the $\beta_l$'s in increasing order, it can be shown that several manifolds of constant $\bar{\bar{c}}$ can be found. These are described by the following set of equations

$$\begin{aligned} &1/2 \leq \beta_1 = \beta_2 < \beta_3 \leq ... \leq \beta_L, \\ &1/3 \leq \beta_1 = \beta_2 = \beta_3 < \beta_4 \leq ... \leq \beta_L, \\ &\vdots \\ &1/L \leq \beta_1 = \beta_2 = ... = \beta_L. \end{aligned} \tag{9.27}$$

We observe that when a larger amount of the smallest inverse temperatures are equal, the stable region grows. In the most extreme case, when all $\beta_l$'s are equal, this region extends to $\beta_c = 1/L$, implying that in the large $L$ limit one will always have a constant clustering coefficient and the phase transition will vanish.

One might wonder how it is possible for us to obtain finite levels of mutual clustering in the thermodynamic limit, even in the regime where the average mutual degree vanishes. In these cases, even though the network sparsifies, the overlapping edges that remain are the ones most congruent with the underlying metric space. The effect of the triangle inequality will therefore be strong, implying that these remaining mutual edges have a high probability of forming triangles. This effectively increases the geometric coupling in the mutual network, explaining for example the result that for a two layer multiplex with $\mu_1 \simeq \mu_2 \simeq \mu$ the mutual network can be mapped to a single layer network with $\beta' = \beta_1 + \beta_2$.

## 9.4 HETEROGENEOUS DEGREE DISTRIBUTIONS

In this section we generalize the above results on link overlap and mutual clustering to the case of heterogeneous networks. For the sake of simplicity, we focus on the two layer case with $\beta_1 = \beta_2 \equiv \beta$, which was shown previously to be the most interesting region.

### 9.4.1 *The soft configuration model*

Where in the previous section we completely ignored the popularity dimension and focused solely on similarity, here we begin with the opposite situation. We set $\beta = 0$ such that the connection to the similarity space is completely lost. The generative model of the individual layers is now the SCM, where the expected degrees can be fixed in expectation. We assume both layers share the same $\kappa$'s and that they are power law distributed $\rho(\kappa) \propto \kappa^{-\gamma}$ with the natural cut-off $\kappa_c \sim N^{1/(\gamma-1)}$.

In this setting, let us first investigate the scaling of the average mutual degree. To this end we apply Eq. (9.13) to the SCM, which gives a result that can be resolved analytically using Mathematica. This leads to many different terms, which can be shown to scale as

$$\langle \widetilde{k} \rangle \sim \begin{cases} N^{2-\gamma} \ln N & \text{if } 2 < \gamma < 3 \\ N^{-1} & \text{if } \gamma > 3 \end{cases} \tag{9.28}$$

Note that in the homogeneous $\gamma \to \infty$ limit, $\langle \widetilde{k} \rangle \sim N^{-1}$, which is in line with our results from Sec. 9.3.2. We observe that a more heterogeneous network leads to more edge overlap. This was to be expected, as the hidden degrees reintroduce correlations between the layers. A hub in one layer will also be one in the other, and there will be many shared connections. Note that even though these hubs increase the edge overlap, popularity alone is not able to provide us with a macroscopic amount of mutual edge. It is only similarity that can do so when both layers are sufficiently geometric.

When it comes to the average local mutual clustering coefficient, we use

$$c(\kappa) = \frac{\int d\kappa' d\kappa'' \rho(\kappa')\rho(\kappa'')p(\kappa',\kappa'')^2 p(\kappa,\kappa'')^2 p(\kappa',\kappa)^2}{N^{-2}\bar{k}(\kappa)^2}, \tag{9.29}$$

which is the SCM implementation of Eq. (9.17). In order to obtain the average local clustering coefficient, one would need to marginalize once more over $\kappa$. However, as argued in Ref. [144], $c(\kappa)$ is a monotonously decreasing function, implying that one only needs to study $c(\kappa)$ for some constant small $\kappa$ in order to obtain the

dominant scaling. Following similar steps to the one taken in the aforementioned reference, it can be proven that

$$\overline{c} \sim \begin{cases} N^0 & \text{if } 2 < \gamma < 3 \\ N^{-4\frac{\gamma-3}{\gamma-1}} & \text{if } 3 < \gamma < 5 \\ N^{-2} & \text{if } 5 < \gamma. \end{cases} \tag{9.30}$$

Surprisingly, the clustering coefficient is constant for scale-free networks where the second moment of the degree sequence diverges. In this region, it is the hubs that are responsible for link overlap. Indeed, it can be shown that the degree distribution of the mutual network becomes ever more heterogeneous, leading to a network with a few hubs and many degree 1 nodes. As these latter nodes do not contribute to the clustering coefficient, the effective network is just a set of hubs all connected to one another, leading to a maximal amount of triangles. In fact, numerical integration of Eq. (9.29) shows that the clustering coefficient stabilizes, after a long transient, on a value close to one.

After a transition region of fairly homogeneous power law degree distributions ($3 < \gamma < 5$) the clustering coefficient settles on its $\gamma \to \infty$ behavior already observed in Fig. 9.2.

### 9.4.2 *General $\beta$*

Moving away from the $\beta = 0$ limit we turn to numerical integration to study the mutual behavior. In Fig. 9.3a we plot the average degree scaling exponent $\sigma_k$ as a function of $\beta$ for $\gamma = 2.5$, $\gamma = 3.5$ and $\gamma = 6$. In the strongly geometric regime $\beta > 1$, a constant mutual degree is obtained for all $\gamma$. The networks become ever sparser as $\beta$ decreases, finally stabilizing around the non-geometric $\beta = 0$ results of Eq. (9.30). Note that the discrepancy for $\gamma = 2.5$ is due to the influence of the $\ln N$ prefactor in this equation. Dividing $\langle \widetilde{k} \rangle$ by this term leads to the black dotted line at $\sigma_k = -0.5$, perfectly in line with the analytic prediction.

We study the behavior of the mutual clustering in Fig. 9.3b, where $\sigma_c$ is plotted against $\beta$. Here, we observe that the constant clustering for $\beta_1 = \beta_2 = \beta > 1/2$ observed in the homogeneous case is also present for heterogeneous networks. In fact, for $2 < \gamma < 3$ this region is extended all the way to $\beta = 0$; the phase transition is completely absent. For less strongly heterogeneous networks $\gamma > 3$, the clustering does start to decrease for $\beta < 1/2$, finally reaching the non-geometric results from Eq. (9.30).

## 9.5 GENERAL COORDINATE CORRELATION

We now turn away from the assumption of perfect correlations between hidden variable in the different layers. As mentioned in the introduction, these correlations are parameterized by the variables $g$ and $\nu$ for the similarity and popularity dimensions, respectively.

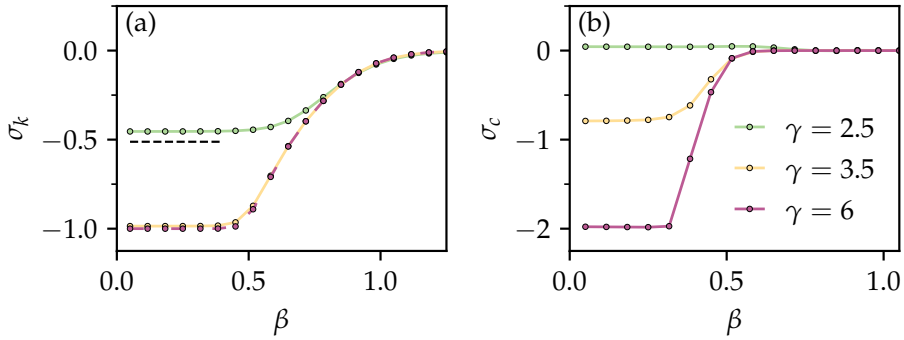



Figure 9.3: (a) The scaling exponents $\sigma_k$ assuming that $\langle \widetilde{k} \rangle \sim N^{\sigma_k}$ as a function of $\beta$ for various $\gamma$. The dashed black line was obtained by fitting $\langle \widetilde{k} \rangle \sim N^{\sigma_k} \ln N$ for $\gamma = 2.5$. All results were obtained by numerically integrating Eq. (9.13) for $N \in [10^6, 10^9]$. (b) Similar to the previous but now for the scaling exponent $\sigma_c$, obtained by numerically integrating Eq. (9.17). In all cases $\langle k \rangle = 15$.

In Fig. 9.4, we study the effect of the correlation strengths on the behavior of the average mutual degree and average mutual clustering coefficient in two extreme cases. For panels (a) and (b), we generate homogeneous networks at various $\beta$'s, varying the correlation strength from $g = 1$ to $g = 0$. For $g > 0$, we see relatively high values of $\langle \widetilde{k} \rangle$ and $\widetilde{\overline{c}}$ when the geometric coupling is strong and low values of these quantities when it is weak. When $g = 0$, the two sets of angular coordinates in the individual layers are sampled independently. This allows us to factorize the integrals in Eq. (9.12), leading to $\langle \widetilde{k} \rangle = \langle k \rangle^2 / N$ as shown in Eq. (9.14). Evidently, the average mutual degree no longer depends on $\beta$. The same line of reasoning can be applied to the mutual clustering coefficient, where it can be shown that $\widetilde{\overline{c}} = \langle c \rangle^2 / N^2$. This quantity does depend on $\beta$ as the average local clustering coefficient of the individual layers is temperature dependent. In particular, for $\beta < 2/3$, i.e. in the non-geometric phase, $\widetilde{\overline{c}} \sim N^{-3}$, decreasing rapidly with the system size.

For panels (c) and (d) similar results can be found, where we study the SCM ($\beta = 0$) for different values of $\gamma$ and varying $\nu \in [0, 1]$. Here we see high values of overlap and mutual clustering for heterogeneous network. Both of these values decrease as the network gets more homogeneous. The same effect is observed when $\nu$ is decreased, i.e. when the correlations are diminished. When $\nu = 0$, the same arguments described in the previous paragraph hold, leading to the same scaling behavior.

## 9.6 REAL NETWORKS

Testing the above findings on real data is no trivial task as most results presented in this paper relate to the scaling of the relevant quantities. Different sized multiplexes

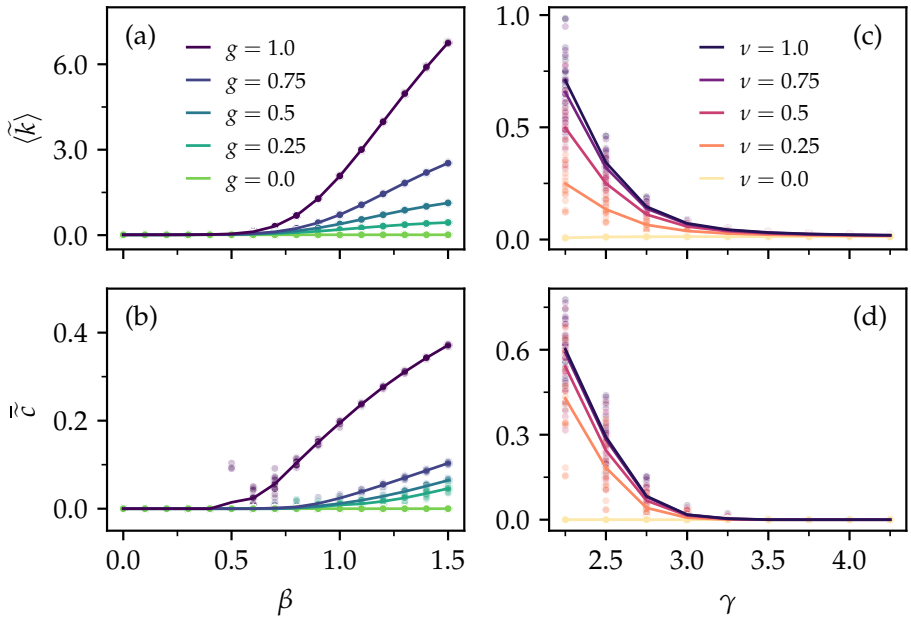



Figure 9.4: In (a) and (b) we study the average degree and the average local clustering coefficient of the mutual network, respectively, as a function of the geometric coupling $\beta$ for various correlation strengths $g \in [0, 1]$. In all cases the individual layers were generated with $N = 32000$, $\gamma = 50$, $\langle k \rangle = 20$ and $\nu = 1$. In (c) and (d) the same quantities are studied, now as a function of the degree exponent $\gamma$ used to generate the hidden degrees in the individual layers. Various correlations strengths $\nu \in [0, 1]$ are shown. The individual layers use $N = 32000$, $\beta = 0$, $\langle k \rangle = 20$ and $g = 1$.

representing the same real system are generally not available, and we are therefore restricted to studying the properties of a single realization. That being said, the scaling behavior should give us some intuition on the size of the node link overlap and mutual clustering, given the properties of the individual layers.

In contrast to the artificial networks studied above, real multiplexes, in general, do not overlap perfectly at the level of the nodes. Many nodes only participate in a couple of the layers. We therefore choose the two layers with the largest amount of shared nodes. We then select the giant mutually connected component, which has been shown to be the relevant object for many applications [268–270]. This pre-processing procedure leaves us with two single-component networks of equal size $N_l$. These two networks are then fed into Mercator [61, 64] in order to extract their geometric coupling strengths $\beta_l$. The networks layers are then used to construct the mutual network via Eq. (9.6). Finally, we calculate the average degree, maximum degree and average local clustering coefficient for each individual layers as well as for the mutual network. In Tab. 9.1 we show the results for the individual layers and in Tab. 9.2 the results for the corresponding mutual networks.

Table 9.1: The properties of two individual layers of various real multiplexes, where $N$ is the amount of nodes in both layers after extracting the mutually connected component. We also show $\beta_i$, the inverse temperatures of both layers as well as $\langle k^{(i)} \rangle$, $k^{(i)}_{\max}$ and $\overline{c^{(i)}}$, the single layer average degree, maximum degree and average local clustering coefficients, respectively. We also show which percentage of nodes the largest hub of the network is connected to. Detailed descriptions of the networks can be found in Appendix D.3. Note that GMP stands for genetic multiplex.

| Network | $N$ | $\beta_1$ | $\beta_2$ | $\langle k^{(1)} \rangle$ | $\langle k^{(2)} \rangle$ | $k^{(1)}_{\max}$ | $k^{(2)}_{\max}$ | $\overline{c^{(1)}}$ | $\overline{c^{(2)}}$ |
|---|---|---|---|---|---|---|---|---|---|
| GMP - Yeast | 2292 | $\approx 0$ | $\approx 0$ | 25.8 | 0.18 | 347 (15%) | 312 (14%) | 0.13 | 0.07 |
| arXiv | 2252 | $> 30$ | 9.73 | 6.47 | 7.07 | 60 (3%) | 80 (4%) | 0.79 | 0.76 |
| Trade | 193 | 1.31 | $\approx 0$ | 46.5 | 49.2 | 154 (80%) | 146 (76%) | 0.78 | 0.73 |
| Twitter | 24764 | 1.37 | 1.12 | 4.3 | 3.9 | 3698 (15%) | 2679 (11%) | 0.46 | 0.30 |
| GMP - Human | 9312 | 1.02 | 0.91 | 14.7 | 9.27 | 6717 (72%) | 1511 (16%) | 0.40 | 0.11 |

Table 9.2: The average degree, degree of the largest hub and clustering coefficient of the mutual networks constructed as described in Sec. D.3. We also show which percentage of nodes the largest hub of the network is connected to.

| Network | $\langle \widetilde{k} \rangle$ | $\widetilde{k}_{\max}$ | $\overline{\widetilde{c}}$ |
|---|---|---|---|
| GMP - Yeast | 0.18 | 10 (0.4%) | 0.06 |
| arXiv | 6.34 | 60 (3%) | 0.80 |
| Trade | 34.9 | 134 (69%) | 0.77 |
| Twitter | 0.67 | 683 (3%) | 0.26 |
| GMP - Human | 2.63 | 603 (5%) | 0.15 |

We notice that while the average mutual degree is oftentimes much lower than the average degrees of the individual layers, the average local clustering coefficient is more robust. This is in line with our theoretical observations: Where both strong geometry and network heterogeneity can produce high levels of clustering, only the former leads to constant average mutual degrees.

In the arXiv network, layers represent different subfields where two scientists are connected if they coauthered a paper in that field. In particular, layer 1 represents papers published in "Social and Information Networks" and layer 2 in "Data Analysis, Statistics and Probability". We see that both layers have a very large geometric coupling and that they are rather homogeneous. As predicted in Sec. 9.4, both edge overlap and mutual clustering are high for this network.

The opposite situation occurs for the genetic multiplex of Saccharomyces pombe, or fission yeast. Here, layers 1 and 2 represent additive and direct genetic interactions. We see that these network layers are rather heterogeneous but non geometric as $\beta_1 = \beta_2 \approx 0$. This latter results implies that the SCM is able to produce the observed levels of clustering, given the degree distribution. We see that node-overlap

is very small, with only a small fraction of the links being present in the mutual network. The mutual clustering, on the other hand, is of similar magnitude as the single layer clustering. These results are also in line with the theoretical predictions obtained in Sec. 9.4, where it was shown that heterogeneity alone leads to high mutual clustering but not necessarily to a macroscopic amount of overlapping edges.

Similar results are observed for the trade network, where nodes are countries and edges represent different traded products, the twitter network of the Moscow world cup in Athletics, where nodes are users and layers represent retweets and mentions, as well as for the human genetic multiplex, where edges represent direct interactions or physical associations. The individual layers of these multiplexes are rather heterogeneous, with the node with largest degree being connected to a large portion of the network. On the other hand, their couplings to the underlying metric space vary from extremely weak to relatively strong. This combination of heterogeneity and metricity leads to significant clustering and non-vanishing edge overlap.

## 9.7 CONCLUSIONS

In this Chapter we studied the overlap between different layers in multiplex networks, specifically the amount of edge and triangle overlap. To this end we introduced the mutual network, which is constructed from the edges of the network that are present in all layers. We studied this object in the context of the geometric multiplex model, where correlations between the layers are mediated through their underlying metric spaces. This framework allowed us to obtain analytic results for the scaling behavior of the mutual average degree and clustering coefficients in various limits. We showed that while both heterogeneous degree distributions alone as well as a strong geometric coupling can lead to large mutual clustering, only the latter can lead to non-vanishing edge overlap. Finally, we showed that many real networks seem to fit these predictions, although more research in this direction remains necessary.

# Part III

# DYNAMICAL PROCESSES

# 10

## THE EMERGENCE OF GEOMETRIC TURING PATTERNS

Throughout the earlier Chapters, we observed the importance of geometry for understanding the structure of complex networks, even when the coupling of these networks to their underlying metric space is weak. In this Chapter, which is an adaptation of Ref [65], we study the effects of geometry on the dynamics on top of the network. In particular, we look at spontaneous pattern formation. We find that the underlying geometry reveals order that cannot be observed from topology alone. We also find hints of the quasi- and non-geometric regimes when the dynamic process runs on weakly geometric networks, underlying the importance of these regions.

Spontaneous pattern formation was first described in 1952 by Alan Turing [271], who noted that competing species of diffusive particles can cause a wide variety of patterns, starting from a spatially homogeneous state. Ever since, reaction-diffusion processes have been shown to host a wide variety of self-organizational behavior [272–279]. Historically, these behaviors have mostly been studied on regular lattices or in continuous media, but there has long been evidence for their existence also in systems of interconnected elements with complex structures. For instance, in the framework of embryonic morphogenesis, which was Turing's original motivation, it has been argued that embryos should be thought of as a multicellular network rather than a continuous medium [280]. However, complex network architecture makes the investigation of dynamical processes difficult, and studies have often been restricted to small networks [281, 282].

In the context of network science, Nakao and Mikhailov took a step forward and demonstrated the existence of the Turing instability in systems of activator-inhibitor species diffusing in large random graphs [283]. In that work, the instability and the corresponding emerging patterns are intimately related to the degree heterogeneity usually present in real networks and are, thus, purely topological in nature.

Similar results have since been found for directed [284], non-normal [285] and temporal networks [286, 287], as well as for networks with higher order interactions such as hypergraphs [288] and simplicial complexes [289]. Another example of non-geometric pattern formation in networks can be found in the context of multilayered networks, where particles diffuse not only within network layers but also between them [290–293]. Here, the dynamical parameters can be chosen such that the particle concentration within a layer is homogeneous whereas

the distribution between layers is heterogeneous. Similar patterns can be found in networks displaying community structure [294], where concentrations can be chosen to be homogeneous within a community and heterogeneous between communities. It has been shown that these types of patterns can also be reproduced in empirical networks, provided their community structure is sufficiently strong.

While hints of explicitly geometric patterns in simple network models were found in several works [284, 295–297], no comprehensive study of this phenomenon has been undertaken. Most importantly, the network models described in these studies are not suitable to describe real systems. In fact, until the publication of th article on which this Chapter is based [65], geometric domains in a classical sense, typical of Turing patterns in lattices and continuous media, had not been observed in real complex networks. This is a consequence of the apparent lack of geometric structure in small-world networks, causing topological distances between nodes –measured by the shortest path lengths on the graph– to collapse around their average value, scaling logarithmically (or slower) with the system size.

However, as we have seen in previous Chapters, many networks do in fact have an associated underlying geometry with strong implications for the topology, even when the coupling is weak. In this Chapter, we show that the Turing instability triggers the emergence of purely geometric patterns that become evident in the latent space of real complex networks.

This Chapter is organized as follows: First, in Sec. 10.1 we introduce the Turing instability on complex networks, showing that a linear stability analysis can be performed when we expand the perturbations in terms of the eigenvectors of the graph Laplacian. Then, in Sec. 10.2, we use the annealed approximation to predict the periodicity of the resulting patterns based solely on the structural properties of the underlying graph. We then turn to weakly geometric graphs in Sec. 10.3, showing that pattering is also observed in the weakly geometric regime $0.5 \lesssim \beta \leq 1$. Finally, in Sec. 10.4 we analyze several real networks, showing that these can also support periodic patterns.

## 10.1 THE TURING INSTABILITY ON COMPLEX NETWORKS

The Turing instability arises as a consequence of the different diffusivity rates between activator and inhibitor species. Starting from a homogeneous concentration of particles, upon small perturbations, the system evolves towards a stable state with regions of high concentration of particles, coexisting with regions of low concentration. When the dynamics takes place on a metric space, these inhomogeneities give rise to geometric patterns.

### 10.1.1 *Diffusion on complex networks*

It is important to note that the first assumption of this framework is that diffusion alone, without the activation-inhibition mechanism, should lead to a homogeneous distribution of particles.

In continuous media, diffusion is modeled by Brownian particles and, in degree regular lattices by standard random walks. Diffusion in complex networks is, however, different. Indeed, a simple random walk process on a heterogeneous network leads to a steady state where the concentration of particles is proportional to the degree of each node and, so, highly heterogeneous. Thus, before adding the activation-inhibition dynamics, we have to carefully define diffusion on networks to ensure that a homogeneous concentration of species is a fixed point of the dynamics.

Let us focus on the diffusion of a single species on a simple undirected and globally connected network with adjacency matrix $\boldsymbol{A}$, and let $n_i(t)$ be the number of particles at node $i$ at time $t$. We assume particles do not interact with each other and that nodes can accumulate an arbitrary number of particles on them. Once at node $i$, a particle jumps away from it at constant rate $\zeta_i$. That is, each particle defines a continuous-time random walk (CTRW) [298, 299] with dwell time probability density $\psi_i(\tau) = \zeta_i e^{-\zeta_i \tau}$. When a particle jumps from node $i$, it chooses one of $i$'s neighbors at random. We can write the following equation for the stochastic evolution of $n_i(t)$ from $t$ to $t + dt$,

$$n_i(t+dt) = n_i(t) + \sum_{j=1}^{N} \frac{A_{ij}}{k_j} \sum_{l=1}^{n_j(t)} \eta_j^l(t, dt) - \sum_{l=1}^{n_i(t)} \eta_i^l(t, dt) \tag{10.1}$$

where $\eta_j^l(t, dt)$ is a random variable controlling the event that one of the $n_j(t)$ particles sitting at node $j$ jumps to a different node in the interval $(t, t + dt)$. That is, $\eta_j^l(t, dt) = 1$ with probability $\zeta_j dt$ and $\eta_j^l(t, dt) = 0$ otherwise. By first taking averages over the random variables $\eta$ and then over $n(t)$, we obtain the equation for the average number of particles at node $i$, $\langle n_i(t) \rangle \equiv u_i(t)$

$$\frac{du_i(t)}{dt} = \sum_{j=1}^{N} \zeta_j \left( \frac{A_{ij}}{k_j} - \delta_{ij} \right) u_j(t) = - \sum_{j=1}^{N} \frac{\zeta_j}{k_j} L_{ij} u_j(t), \tag{10.2}$$

where $L_{ij}$ is the Laplacian matrix [300]

$$L_{ij} = k_j \delta_{ij} - A_{ij}. \tag{10.3}$$

The steady state solution of Eq. (10.2) is given by the eigenvector of zero eigenvalue of the Laplacian matrix. This implies that

$$\frac{\zeta_i u_i^{st}}{k_i} = \text{cst}. \tag{10.4}$$

Therefore, if particles diffuse at the same rate in all nodes –so with $\zeta_i = \zeta$ , $\forall i$– then the average concentration at node $i$ is proportional to $k_i$ and the average concentration of particles is the same everywhere only when $\zeta_i \propto k_i$. We then make the choice $\zeta_i = \epsilon k_i$ such that the diffusion equation finally reads

$$\frac{du_i(t)}{dt} = -\epsilon \sum_{j=1}^{N} L_{ij} u_j(t). \tag{10.5}$$

Similarly to the case of diffusion in continuous media, we call $\epsilon$ the diffusion coefficient of the species in the network, even though, unlike diffusion in the continuum, it has dimensions of inverse of time.

### 10.1.2 *Local activation-inhibition dynamics*

In the diffusion process described by Eq. (10.5), particles do not interact when they meet in the same node. However, in realistic settings, particles may undergo all sorts of reaction processes. In particular, the Turing instability arises when two different species, $U$ and $V$, interact upon meeting at the same node, under an activation-inhibition process. Within each node, species $U$ undergoes an autocatalytic process and, simultaneously, favors the increase of the population of species $V$. At the same time, species $V$ inhibits the growth of species $U$ even though it cannot survive without it. When the average number of particles per node is large, we can neglect fluctuations in the number of particles and work under the mean field approximation.

Let us turn off diffusion for a moment, and study the local dynamics separately. Say $u(t)$ and $v(t)$ are the average numbers of particles of species $U$ and $V$, respectively. Their evolution equations are given by

$$\frac{du(t)}{dt} = f\left(u(t), v(t)\right) \tag{10.6}$$

$$\frac{dv(t)}{dt} = g\left(u(t), v(t)\right), \tag{10.7}$$

where functions $f(u, v)$ and $g(u, v)$ represent the local activation-inhibition dynamics.

In the absence of diffusion, we assume that each node has a stable stationary state, that is, that there exist $u^{st} = \bar{u}$ and $v^{st} = \bar{v}$ for which $f(\bar{u}, \bar{v}) = 0$ and $g(\bar{u}, \bar{v}) = 0$ such that $\{\bar{u}, \bar{v}\}$ is a fixed point of the dynamics.

We now introduce some constraints on the functions $f(u, v)$ and $g(u, v)$ and their first derivatives $f_u \equiv \partial_u f(\bar{u}, \bar{v})$, $f_v \equiv \partial_v f(\bar{u}, \bar{v})$, $g_u \equiv \partial_u g(\bar{u}, \bar{v})$, and $g_v \equiv \partial_v g(\bar{u}, \bar{v})$. The fact that $U$ catalyzes both its own growth as that of $V$ implies that $f_u, g_u > 0$. Conversely, the inhibitive nature of $V$ implies that $f_v < 0$. In order to ensure the stability of the fixed point, several extra constraints are necessary. Perturbing around the fixed point as $\{u(t), v(t)\} = \{\bar{u}, \bar{v}\} + \{\delta u(t), \delta v(t)\}$ and linearizing the equations we obtain

$$\frac{d}{dt}\begin{pmatrix}\delta u(t)\\ \delta v(t)\end{pmatrix} = \begin{pmatrix} f_u & f_v \\ g_u & g_v\end{pmatrix}\begin{pmatrix}\delta u(t)\\ \delta v(t)\end{pmatrix} \equiv \boldsymbol{J}\begin{pmatrix}\delta u(t)\\ \delta v(t)\end{pmatrix}, \tag{10.8}$$

which leads to an exponential solution whose decay or growth is determined by the sign of the eigenvalues of the Jacobian $\boldsymbol{J}$ shown above. The largest of these eigenvalues is given by $\lambda_+ = \frac{1}{2}(\mathrm{Tr}\boldsymbol{J} + \sqrt{(\mathrm{Tr}\boldsymbol{J})^2 - 4\Delta})$, where $\mathrm{Tr}\boldsymbol{J}$ is the trace of the matrix and $\Delta$ its determinant. In order for this eigenvalue to have a negative real part, and therefore lead to exponentially suppressed perturbations, we need

to impose $\mathrm{Tr}\boldsymbol{J} = f_u + g_v < 0$ (which directly implies $g_v < 0$ as we previously set $f_v > 0$) as well as $\Delta = f_u g_v - f_v g_u > 0$.

### 10.1.3 *The Laplacian expansion*

Now that we have fully understood both the local and the diffusive dynamics of the system we are ready to combine the two, leading to the evolution equations

$$\frac{du_i(t)}{dt} = f\left(u_i(t), v_i(t)\right) - \epsilon \sum_{j=1}^{N} L_{ij} u_j(t) \tag{10.9}$$

$$\frac{dv_i(t)}{dt} = g\left(u_i(t), v_i(t)\right) - \sigma\epsilon \sum_{j=1}^{N} L_{ij} v_j(t), \tag{10.10}$$

where $\epsilon$ and $\sigma\epsilon$ are the activator and inhibitor diffusion coefficients, respectively.

The Turing instability arises when, due to the differences in the diffusion coefficients of the two species, the fixed point becomes unstable so that any small perturbation drives the system away from it. Interestingly, the conditions for this to happen only depend on $f_u, f_v, g_u, g_v$ and the ratio between the diffusion coefficients of the two species $\sigma$. Similarly to the case of continuous media, where the perturbation around the fixed point can be expanded in Fourier modes –the eigenfunctions of the Laplacian $\nabla^2$– and where the eigenvalues are related to the wavenumbers, here the perturbation can be expanded in terms of the eigenvectors of the graph Laplacian $L_{ij}$ [280, 283]

$$\delta u_i(t) = u_i(t) - \bar{u} \approx \sum_{\alpha=1}^{N} c_\alpha e^{\lambda_\alpha t} \phi_i^\alpha, \tag{10.11}$$

where $c_\alpha$ and $\lambda_\alpha$ are some constants that depend on the parameters of the model Eqs. (10.9)-(10.10) and on the eigenvalue $\Lambda_\alpha$ (similarly for species $V$). It is this expansion that couples the dynamics ($\delta u_i(t)$) and the structure of the network ($\phi_i^\alpha$) through the Laplacian. This expansion is possible because the eigenvectors of the Laplacian matrix $\{\boldsymbol{\phi^\alpha} = \{\phi_1^\alpha, \phi_2^\alpha, \cdots, \phi_N^\alpha\}\}_{\alpha=1}^N$ form a complete orthonormal basis of $\mathbb{R}^N$. If $\lambda_\alpha < 0$ ; $\forall \Lambda_\alpha$, then perturbations around the fixed point are absorbed exponentially fast, so that the fixed point is stable. However, if there exists at least one $\Lambda_\alpha$ such that $\lambda_\alpha > 0$, then the fixed point becomes unstable. In this case, those nodes with $c_\alpha \phi_i^\alpha > 0$ will increase the concentration of species $U$ whereas those with negative value will decrease it. Note that for all eigenvectors associated to non-vanishing eigenvalues we have

$$\sum_{i=1}^{N} \phi_i^\alpha = 0, \tag{10.12}$$

which is a result of the fact that $\sum_{ij} L_{ij} = 0$. If we combine this with each component being associated to a given node of the network, we find that the dynamics will

necessarily evolve towards a state with some of the nodes containing a high concentration of species $U$ and others a low concentration.

Linearizing the system of equations given in Eq. (10.10) around the equilibrium values, and plugging in Eq. (10.11) it can be shown that there exist unstable eigenvalues (i.e $\lambda_\alpha > 0$) whenever

$$\sigma \geq \sigma_c = \frac{f_v g_u}{f_u^2}\left[\frac{f_u g_v}{f_v g_u} - 2 - 2\sqrt{1 - \frac{f_u g_v}{f_v g_u}}\right]. \tag{10.13}$$

In this regime, all eigenvalues that lay within the interval $\Lambda \in [\Lambda_-, \Lambda_+]$ with

$$\Lambda_\pm = \frac{1}{2\sigma\epsilon}[\sigma f_u + g_v \pm \sqrt{(\sigma f_u - g_v)^2 + 4\sigma f_v g_u}] \tag{10.14}$$

become unstable, with the most unstable one being

$$\Lambda_{\max} = \frac{1}{(1-\sigma)\epsilon}\left(f_u - g_v - (1+\sigma)\sqrt{\frac{-f_v g_u}{\sigma}}\right). \tag{10.15}$$

The functional dependence of Eqs. (10.13) and (10.14) implies that, for a given set of parameters $f_u, f_v, g_u, g_v$ and for one particular eigenvalue $\Lambda_\alpha$, it is always possible to select parameters $\epsilon$ and $\sigma$ such that $\Lambda_\alpha \in [\Lambda_-, \Lambda_+]$. In the case that $\Lambda^+ \approx \Lambda^- \approx \Lambda_\alpha$ one can assume that the differences in the concentration of species in the nodes are encoded in the eigenvector $\boldsymbol{\phi}^\alpha$.

### 10.1.4 *The Mimura-Murray dynamics*

Interestingly, the results described in Sec. 10.1.3 only depend on the first derivatives of functions $f(u,v)$ and $g(u,v)$ coupled to the diffusion coefficients. Therefore, there exists a whole class of systems where such instability may emerge, from chemical reactions [301–303] or biological morphogenesis [273, 304] to ecosystems [305–308] and game theory applied to ecological systems [309, 310]. In ecology, one such model is the Mimura-Murray model [306]. Here, $u$ and $v$ represent prey and predator densities, respectively, and the local dynamics are governed by the functions

$$f(u,v) = \left(\frac{a + bu - u^2}{c} - v\right)u, \tag{10.16}$$

$$g(u,v) = (u - (1+dv))\,v. \tag{10.17}$$

From Eqs. (10.9), (10.10), (10.16), and (10.17), we see that in the absence of preys, predators go extinct and in the absence of predators, preys attain a constant population. In general, there is a fixed point with positive densities of preys and predators.

In this Chapter we choose the parameters as $a = 35$, $b = 16$, $c = 9$ and $d = 2/5$, which gives the fixed point $\{\bar{u}, \bar{v}\} = \{5, 10\}$. In this case, the critical value of $\sigma$

above which the Turing instability arises is $\sigma_c = 15.5071$. We then set the value to $\sigma = 15.6$, slightly above the critical point. Using Eqs. (10.14) and (10.15), we find that unstable eigenvalues lie within the interval $[\Lambda_-, \Lambda_+]$ with

$$\Lambda_- = \frac{1.41026}{\epsilon} \text{ and } \Lambda_+ = \frac{1.66667}{\epsilon}, \tag{10.18}$$

and the most unstable one given by

$$\Lambda_{\max} = \frac{1.53325}{\epsilon}. \tag{10.19}$$

To select a given eigenvalue of interest $\Lambda^*$, we set it to $\Lambda^* = \Lambda_{\max}$ and use Eq. (10.19) to fix the value of $\epsilon$. The functional form of this relation implies that very large values of $\Lambda^*$ imply very small values of $\epsilon$. In these cases, the interval of unstable eigenvalues $|\Lambda_+ - \Lambda_-|$ might get very large, resulting in difficulties in selecting a single unstable eigenvalues. Here, it could be necessary to lower the value of $\sigma > \sigma_c$ in order to shrink the unstable range.

With all the parameters of the dynamics fixed, we set every node's initial condition to be the stationary one, that is, $\{u_i(0), v_i(0)\} = \{\bar{u}, \bar{v}\}$ for all $i$ except for a randomly chosen one, $j$, for which $\{u_j(0), v_j(0)\} = \{\bar{u} + \delta, \bar{v} + \delta\}$, where $\delta = 10^{-5}$ is a small perturbation. We then leave the system evolve towards its steady state.

### 10.1.5 *Spectral properties of the Laplacian matrix*

Before applying these findings on some actual networks, let us first note some important properties of the Laplacian matrix [300]. Most importantly, we want to understand which values it can take, i.e., which values for $\Lambda^*$ are reasonable.

The Laplacian is a real symmetric matrix with real eigenvalues $\Lambda_1 = 0 \leq \Lambda_2 \leq \cdots \leq \Lambda_N \leq 2k_c$, where $k_c$ is the maximum degree of the network. The average of all eigenvalues is equal to the network average degree, that is,

$$\langle \Lambda \rangle = \langle k \rangle \tag{10.20}$$

and the second moment is

$$\langle \Lambda^2 \rangle = \langle k^2 \rangle + \langle k \rangle. \tag{10.21}$$

These results suggest that eigenvalues are strongly correlated to the sequence of degrees in the network [311].

It is worth mentioning an interesting property of the $\mathbb{S}^1/\mathbb{H}^2$-model concerning its spectral gap, defined as the smallest non-null eigenvalue of the Laplacian $\Lambda_2$. Given a graph generated by the $\mathbb{S}^1$-model, $G_{\mathbb{S}^1}$, the Cheeger's inequalities state that

$$\frac{[h(G_{\mathbb{S}^1})]^2}{2k_c} \leq \Lambda_2 < 2h(G_{\mathbb{S}^1}), \tag{10.22}$$

where $h(G_{\mathbb{S}^1})$ is the isoperimetric (or Cheeger) constant of the graph [312, 313]. This constant defines the optimal cut of the graph in two disjoint sets of nodes.

For any partition $A$ and $B$ in two disjoint sets of sizes $N_A$ and $N_B$, the Chegger constant is defined as the ratio between the number of edges connecting the two sets $M_{AB}$ and the size of the smallest set, minimized over all possible bipartitions of the graph, that is,

$$h(\mathcal{G}_{\mathbb{S}^1}) = \min_A \left\{ \frac{M_{AB}}{\min(N_A, N_B)} \right\}. \tag{10.23}$$

While computing the Cheeger constant is, in general, not possible, in the $\mathbb{S}^1/\mathbb{H}^2$-model we can calculate its scaling behavior with the system size as follows: Assuming that the optimal cut minimizing the Cheeger constant is made by splitting the unit circle in two continuous regions, in the case of the homogeneous $\mathbb{S}^1$-model, Eq. 10.23 can be written as

$$h(\mathcal{G}_{\mathbb{S}^1}) = \min_x \left( \frac{2\pi}{xN} \int_0^x \int_x^{2\pi} \left( \frac{N}{2\pi} \right)^2 \frac{\mathrm{d}\theta_2 \mathrm{d}\theta_1}{1 + \left( \frac{N\Delta\theta_{12}}{2\pi\hat{\mu}\langle k \rangle^2} \right)^\beta} \right), \tag{10.24}$$

where $\Delta\theta_{12} = \pi - |\pi - |\theta_1 - \theta_2||$. Here we defined the two disjoint regions as $A = \{i \,|\, 0 < \theta_i < x\}$ and $B = \{i \,|\, x < \theta_i < 2\pi\} = V_{G_{\mathbb{S}^1}} \backslash A$. As we have chosen $0 < x < \pi$, in the model the least amount of nodes reside in $A$ and thus $\min(N_A, N_B) = N_A = xN/(2\pi)$, as seen in Eq. 10.24. The second part of this equation represents the amount of links between regions $A$ and $B$, i.e. $M_{AB}$. It can be shown that Eq. 10.24 can be written as

$$h(G_{\mathbb{S}^1}) = \min_x \left( \frac{1}{x} \frac{N}{\pi} \int_0^x \int_\theta^\pi \frac{\mathrm{d}\theta \mathrm{d}\Delta}{1 + \left( \frac{N\Delta}{2\pi\hat{\mu}\langle k \rangle^2} \right)^\beta} \right), \tag{10.25}$$

which in turn can be solved exactly, giving

$$h(G_{\mathbb{S}^1}) = \min_x \left( \frac{N}{2\pi} f(N, x) \right), \tag{10.26}$$

where

$$\begin{aligned} f(N,x) = 2\pi & \quad {}_2F_1 \left[ \begin{matrix} 1, \frac{1}{\beta} \\ 1 + \frac{1}{\beta} \end{matrix} ; -\left( \frac{N}{2\hat{\mu}\langle k \rangle} \right)^\beta \right] \\ + x \Bigg( & - {}_2F_1 \left[ \begin{matrix} 1, \frac{1}{\beta} \\ 1 + \frac{1}{\beta} \end{matrix} ; -\left( \frac{Nx}{2\pi\hat{\mu}\langle k \rangle} \right)^\beta \right] \\ & + {}_2F_1 \left[ \begin{matrix} 1, \frac{2}{\beta} \\ 1 + \frac{2}{\beta} \end{matrix} ; -\left( \frac{Nx}{2\pi\hat{\mu}\langle k \rangle} \right)^\beta \right] \Bigg). \end{aligned} \tag{10.27}$$

It can be shown that this scales as $h(G_{\mathbb{S}^1}) \simeq c_1 N^{\max(0,1-\beta)} + c_2 N^{-1}$, irrespective of the choice of $x$. For $\beta > 1$, the region on which we generally focus on in

this Chapter, this scaling relation implies that $h(G_{\mathbb{S}^1})$ approaches zero in the thermodynamic limit and that the spectral gap, therefore, does so too.

This implies that for very large networks we expect to find eigenvalues arbitrarily close to zero. This is important because small eigenvalues are the ones with more visible and clear patterns. This derivation is valid only in the case of a homogeneous hidden degree distribution. However, numerical analysis indicate that the Cheeger constant decays to zero in the case of a heterogeneous distribution as well.

These results imply that the smallest non-null eigenvalue of the Laplacian approaches zero in the thermodynamic limit. We assume that, in this limit, the spectrum is continuous on the positive real line as long as the node density along the circle and the connection probability are continuous functions and the degree distribution is unbounded. This implies that any $\Lambda^{\alpha}$ can be chosen in the procedure described above as any real positive number should lie arbitrarily close to an eigenvalue of the graph Laplacian for sufficiently large networks.

### 10.1.6 *Turing patterns on geometric networks*

To illustrate the role of geometry on the Turing instability, we run the Mimura-Murray dynamics on a real network, namely that of the connectome of a mouse described in Appendix E.1. Figure 10.1a shows the results on the concentration of species $U$ in the hyperbolic representation of the network. As can be seen, there is a clear pattern associated to the geometric organization of the network, with low concentrations of particles localized in the same angular position and high concentrations in the rest of the network. In panels 10.1b and 10.1c, we also represent the concentration as a function of the angular coordinate $\theta$ and the node index $i$, respectively. In the latter case the nodes are ordered in decreasing order of degree. It is important to notice that the Mimura-Murray dynamics does not use any information from the underlying metric space, so that the observed geometric pattern is a highly non-trivial result.

To shed light on this problem, we also run the dynamics on networks generated by the $\mathbb{S}^1/\mathbb{H}^2$-model with a heterogeneous degree distribution with exponent $\gamma$. Fig. 10.1d-i shows results of the dynamics on two different networks: one highly heterogeneous ($\gamma = 2.1$) and weakly clustered and a second one less heterogeneous ($\gamma = 2.84$) and highly clustered. In the strongly heterogeneous network shown in the figure, there are no patterns associated to the geometric character of the network. We do observe, however, topological patterns induced by degree, roughly resulting in high degree nodes holding a low concentration of species $U$. These are the type of patterns documented in [283]. Instead, in the case of the less heterogeneous network, we find a clear geometric pattern associated to the angular coordinates of nodes and very weak degree-related patterns, similar to the results found in the real network. These patterns emerge despite the fact that networks have the small-world property, so that diffusion induces flows of species between nodes that are far apart in the underlying metric space. These results suggest that some eigenvectors of the Laplacian of the $\mathbb{S}^1/\mathbb{H}^2$-model have a well

defined periodic structure in the similarity space. Next, we develop a theoretical framework allowing us to calculate an approximation for the frequency associated with such eigenvectors.

## 10.2 THE ANNEALED APPROXIMATION

The results in the previous section were obtained through trial and error: several diffusion constants $\epsilon$ were tested, leading to different eigenvectors becoming unstable, in turn resulting in different types of patterns. However, this method is highly inefficient, and it would be very useful to find a general relation between the eigenvalue and the periodicity of the resulting pattern. The goal of this section is therefore to determine the dispersion relation $\omega(\Lambda)$ between the eigenvalue $\Lambda$ and the frequency of its associated eigenvector. It needs to be stressed that this is a structural property of the network, and is not conditioned by the dynamics on top of the network. We focus here on the region $\beta > 1$, as this is where we expect our technique to work best due to the strong coupling with the underlying metric space. In Sec. 10.3 we also look into the weak coupling regime, albeit only from a numerical point of view.

Species whose dynamics are governed by the Mimura-Murray model diffuse within a network with a disordered but quenched structure; a single realization of the network ensemble defined by the $\mathbb{S}^1/\mathbb{H}^2$-model for a given sequence of hidden variables $\{\kappa_i, \theta_i\}_{i=1}^N$. Unfortunately, for quenched networks it is not possible to get any analytical insight on the structure of the eigenvectors of the Laplacian matrix. To overcome this problem, we use the annealed approximation, which has been extensively used in the literature to tackle a wide variety of problems, from neural networks and opinion dynamics to epidemic spreading [56, 314–321]. In this approach, the network structure is resampled from the ensemble at a rate faster than the diffusion dynamics. This allows us to replace the adjacency matrix $A_{ij}$ by its ensemble average, the connection probability $p_{ij}$, and consider the network not as a quenched system but as a dynamic one. With this approximation, the eigenvalue problem of the annealed Laplacian matrix can be written as

$$\sum_{j\neq i} \frac{\phi(\kappa_j,\theta_j)}{1+\left[\frac{x_{ij}}{\hat{\mu}\kappa_i\kappa_j}\right]^\beta} = \left[\sum_{j\neq i} \frac{1}{1+\left[\frac{x_{ij}}{\hat{\mu}\kappa_i\kappa_j}\right]^\beta} - \Lambda\right] \phi(\kappa_i,\theta_i), \tag{10.28}$$

where $\Lambda$ is the eigenvalue and $\phi(\kappa_i, \theta_i)$ the component of the corresponding eigenvector of a node $i$ with hidden variables $\{\kappa_i, \theta_i\}$. Notice that, as in the case of the quenched Laplacian matrix, $\Lambda = 0$ is an eigenvalue with constant eigenvector.

In the thermodynamic limit, the curvature of the circle goes to zero and, thus, nodes become distributed in $\mathbb{R}^1$ according to a Poisson point process of density one. In this limit, the distance between two nodes can be evaluated as $x_{ij} = |x_i - x_j|$, where $x_i$ and $x_j$ are the positions of nodes $i$ and $j$ in $\mathbb{R}^1$. Finally, we take the continuum limit in Eq. (10.28) by replacing the sum over index $j$ by a double

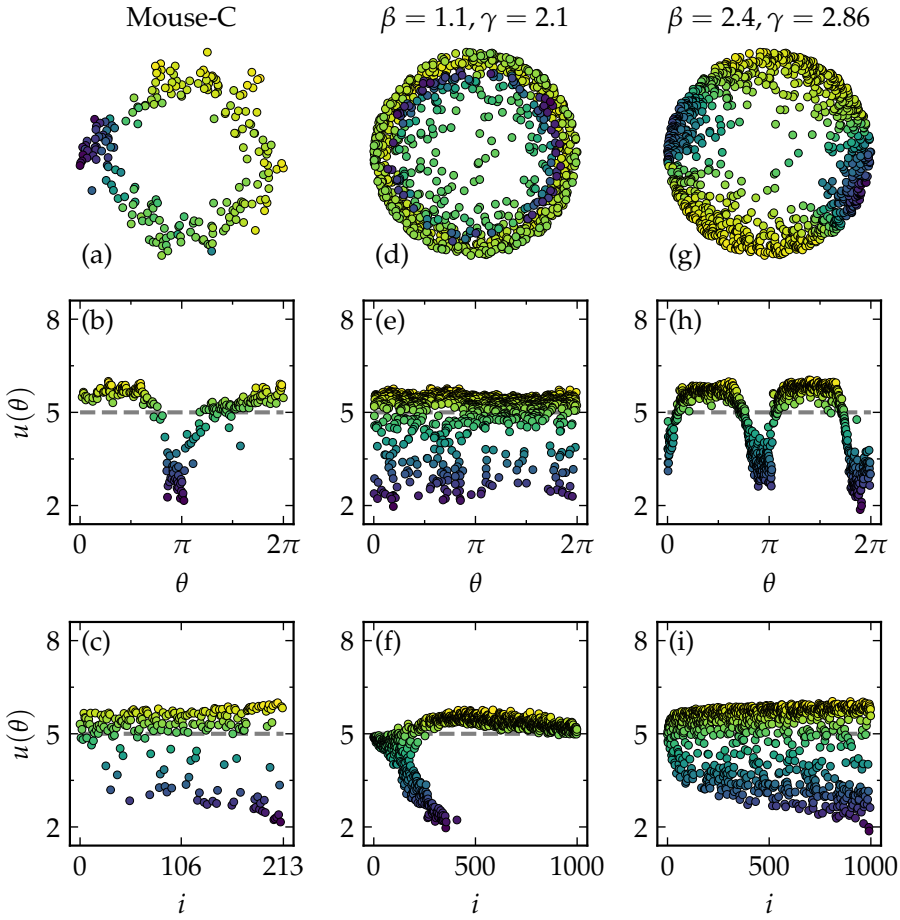



Figure 10.1: Examples of Turing patterns in networks. The first, panels (a-c), is a real network representing the connectome of a mouse, the second, panels (d-f), a strongly heterogeneous $\mathbb{S}^1/\mathbb{H}^2$ network with parameters $\{N, \langle k \rangle, \beta, \gamma\} = \{1000, 50, 1.1, 2.1\}$ and the third, panels (g-i) a weakly heterogeneous $\mathbb{S}^1/\mathbb{H}^2$ network with parameters $\{N, \langle k \rangle, \beta, \gamma\} = \{1000, 50, 2.4, 2.86\}$. In all panels the colors correspond to the density of the activators in the final stationary state. In panels (a,d,g), nodes are located according to their coordinates in hyperbolic space. Panels (b,e,h) show the density of activators as a function of nodes' angular coordinates and in panels (c,f,i) the activators densities are plotted against nodes' degree ranking.

integral over nodes coordinates $\{\kappa, x\}$. In this way, Eq. (10.28) can be written as the following integral equation

$$\int_{\kappa_0}^{\infty} \rho(\kappa')d\kappa' \int_{-\infty}^{\infty} dx' \frac{\phi(\kappa', x')}{1 + \left[\frac{|x-x'|}{\hat{\mu}\kappa\kappa'}\right]^{\beta}} = (\kappa - \Lambda)\phi(\kappa, x). \tag{10.29}$$

The spatial part of Eq. (10.29) has the form of a convolution integral. Thus, we can take advantage of the convolution theorem for Fourier transforms. By defining the Fourier transform of the eigenvector as

$$\hat{\phi}(\kappa, \omega) \equiv \int_{-\infty}^{\infty} e^{-i\omega x}\phi(\kappa, x)dx \tag{10.30}$$

Eq. (10.29) can be written in Fourier space as

$$\int_{\kappa_0}^{\infty} \frac{\kappa'\rho(\kappa')}{\langle k\rangle} \frac{\hat{\Psi}(\hat{\mu}\kappa\kappa'\omega)}{\hat{\Psi}(0)} \hat{\phi}(\kappa', \omega)d\kappa' = \frac{\kappa - \Lambda}{\kappa}\hat{\phi}(\kappa, \omega), \tag{10.31}$$

where we have defined $\hat{\Psi}$ as

$$\hat{\Psi}(z) \equiv \mathcal{F}\left[\frac{1}{1 + |x|^{\beta}}\right](z), \tag{10.32}$$

where $\mathcal{F}$ stands for Fourier transform. Notice that $\hat{\mu}^{-1} = \langle k\rangle\hat{\Psi}(0)$. From Eq. (10.31), it is easy to see that the integral over the spatial coordinate of non-trivial eigenvectors must vanish, that is, $\hat{\phi}(\kappa, 0) = 0$, similarly to the case of the quenched Laplacian matrix Eq. (10.12). Indeed, by setting $\omega = 0$ in Eq. (10.31), we conclude that $\hat{\phi}(\kappa, 0) = 0$ is the only solution when $\Lambda \neq 0$. Thus, the annealed approximation preserves the basic property of the Laplacian matrix given in Eq. (10.12).

### 10.2.1 *Homogeneous ensemble*

The analytic solution of Eq. (10.31) can be found in particular cases. One such case is a homogeneous ensemble where all nodes have the same hidden degree, that is $\rho(\kappa) = \delta(\kappa - \langle k\rangle)$, leading to a Poisson degree distribution of average $\langle k\rangle$. In this situation, the eigenvector is just a function of the frequency $\omega$ and Eq. (10.31) becomes

$$\left[\frac{\hat{\Psi}(\hat{\mu}\langle k\rangle^2\omega)}{\hat{\Psi}(0)} - \frac{\langle k\rangle - \Lambda}{\langle k\rangle}\right]\hat{\phi}(\omega) = 0. \tag{10.33}$$

Assuming that $\hat{\phi}(\omega) \neq 0$, this equation defines a dispersion relation between the eigenvalue $\Lambda$ and its characteristic frequency $\omega_c$ as a solution of the transcendent equation

$$\frac{\Lambda(\omega_c)}{\langle k\rangle} = \left(1 - \frac{\hat{\Psi}(\hat{\mu}\langle k\rangle^2\omega_c)}{\hat{\Psi}(0)}\right). \tag{10.34}$$

The characteristic frequency $\omega_c$ in Eq. (10.34) is constrained by the boundary conditions and the discretization of nodes in the space. Indeed, since nodes are

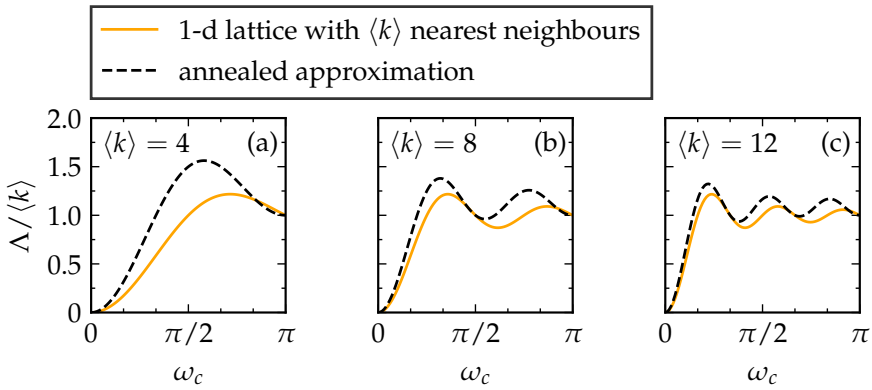



Figure 10.2: Dispersion relation for the annealed approximation and a degree-regular 1-dimensional lattice. Solid orange lines show exact results from Eq. (10.37) and black dashed lines for the annealed approximation in Eq. (10.36) for homogeneous networks with $\beta = \infty$. Results are shown for average degrees $k = \langle k \rangle = 4, 8$, and 12.

distributed on the line at density 1, frequencies above $\pi$ (and so wavelengths of the order 1) cannot be detected. Besides, if the system is finite of length $L = N$, frequencies below $2\pi/N$ will have associated wavelengths comparable to the system size and, so, will not be detected either. Therefore, we will look for solutions of Eq. (10.34) in the domain $\omega_c \in [2\pi/N, \pi]$.

It is illustrative to analyze in detail the homogeneous case with $\beta = \infty$, which corresponds to a connection probability being a step function. In this case, function $\hat{\Psi}(z)$ takes the form

$$\hat{\Psi}(z) = 2\frac{\sin z}{z}. \tag{10.35}$$

and $\hat{\mu} = 1/(2\langle k \rangle)$. Combining these results, the dispersion relation defining $\omega_c$ becomes

$$\frac{\Lambda(\omega_c)}{\langle k \rangle} = \left(1 - \frac{2\sin\frac{\langle k \rangle \omega_c}{2}}{\langle k \rangle \omega_c}\right). \tag{10.36}$$

In the case of $\beta = \infty$, nodes connect to nearest neighbors in the line that are below a certain critical distance, so that the average number of such neighbors is $\langle k \rangle$. Therefore, apart from the fluctuations in the number of neighbors, the dispersion relation in Eq. (10.36) should be equivalent to the dispersion relation of a one dimensional lattice with $\langle k \rangle$ symmetric nearest neighbors, introduced in Sec. 3.2 as the RRL, which reads

$$\frac{\Lambda}{\langle k \rangle} = 1 - \frac{2}{\langle k \rangle} \sum_{n=1}^{\langle k \rangle/2} \cos n\omega_c. \tag{10.37}$$

Figure 10.2 shows the comparison between Eqs. (10.36) and (10.37) for $\langle k \rangle = 4, 8$, and 12. Both expressions are very similar and become identical as the connectivity

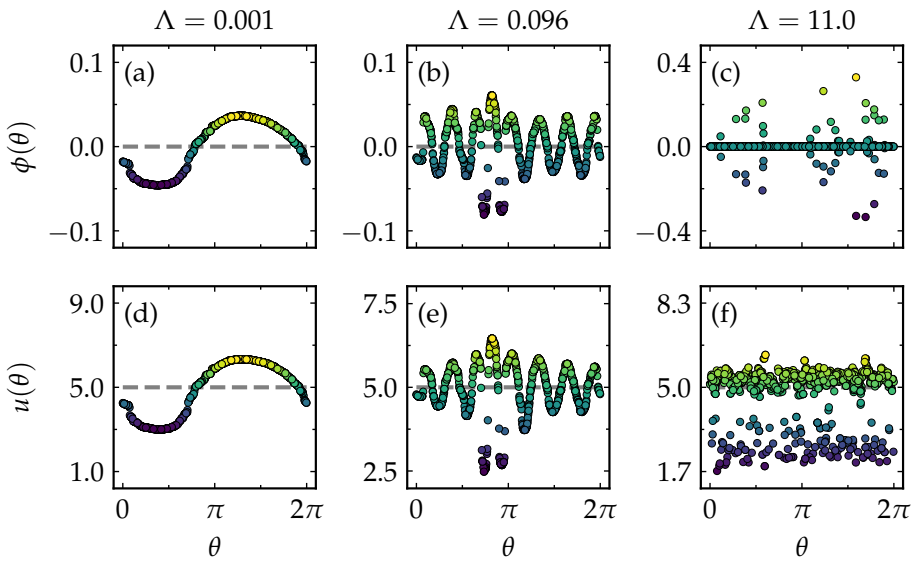



Figure 10.3: (a-c) Examples of three eigenvectors of the quenched Laplacian matrix with low, medium, and high associated frequencies for a single network generated by the homogeneous ensemble with $N = 1000$ and $\langle k \rangle = 20$ and $\beta = \infty$. (d-f) Concentration of species $U$ in the Mimura-Murray dynamics.

increases, in agreement with the fact that the continuum approximation becomes exact at infinite average degree.

Despite the good agreement shown in Fig. 10.2, the comparison with quenched networks generated by the $\mathbb{S}^1/\mathbb{H}^2$-model has to be made with care. Indeed, for homogeneous networks, the annealed approximation completely neglects fluctuations of nodes' degrees. In a quenched network this is far from true, as the degree distribution follows a Poisson distribution of average $\langle k \rangle$. However, the dispersion relations Eqs. (10.36) and (10.37) are strongly dependent on the network connectivity, especially in the high frequency domain. Therefore, due to the randomness of nodes' degrees, it is not possible to characterize each eigenvector of the quenched Laplacian by a unique frequency but by a collection of frequencies around a given average.

We cannot talk then of a deterministic dispersion relation but a fuzzy one where different frequencies may coexist in the same eigenvector. This will destroy any possible high frequency periodic pattern, leaving only low frequency eigenvectors (with $\Lambda \ll \langle k \rangle$) visible. The reason is that low frequency modes correspond to long wavelengths, so that local fluctuations of degrees become irrelevant. This is clearly visible in Fig. 10.3a-c, showing three different eigenvectors of the quenched Laplacian matrix corresponding to low, medium, and high frequencies of a graph generated by the homogeneous ensemble. The periodic pattern is very clear in the case of low eigenvalue with very low frequency. In the medium frequency case, we can still identify a periodic behavior, although distorted by noise. In

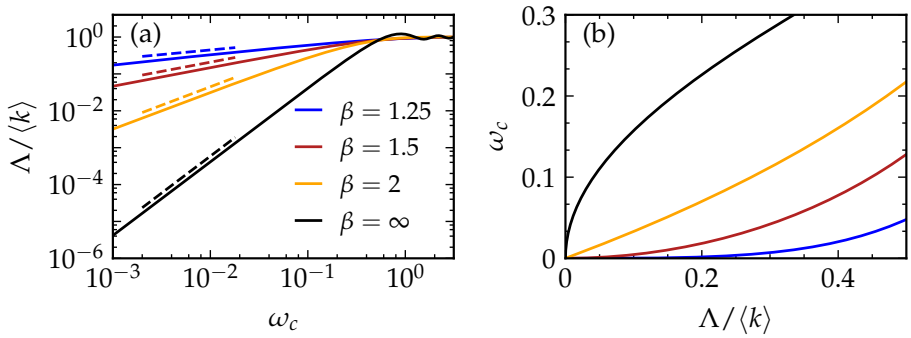



Figure 10.4: Characteristic frequency for the model for $\langle k\rangle = 10$ and $\beta = \infty, 2, 1.5$, and 1.25.

the case of high frequency, periodic behavior is totally absent. We also run the Mimura-Murray dynamics on the same network selecting parameters' values such that the same eigenvalues used in Fig. 10.3a-c become unstable. Fig. 10.3d-f show the concentration of species $U$ in the three cases. We observe that the patterns in the concentration of species $U$ closely follow the patterns of the corresponding eigenvectors, justifying the relevance of the structure of eigenvectors for dynamical processes on networks.

Based on these observations, we conclude that the annealed approximation provides good results in quenched networks for small eigenvalues such that $\Lambda \ll \langle k\rangle$, which corresponds to low frequencies, that is, $\hat{\mu}\langle k\rangle^2\omega_c \ll 1$. For arbitrary values of $\beta > 1$, we can then take the low frequency limit in Eq. (10.34) to derive the relation between $\omega_c$ and $\Lambda$. Using the definition of the Fourier transform, it is easy to see that for small $z$

$$1 - \frac{\hat{\Psi}(z)}{\hat{\Psi}(0)} \sim \begin{cases} z^{\beta-1} & 1 < \beta \leq 3 \\ z^2 & \beta > 3 \end{cases} . \tag{10.38}$$

Figure 10.4a shows the low frequency limit of the dispersion relation computed numerically for different values of $\beta$, which corroborates this scaling behavior.

Using this result, the characteristic frequency scales as

$$\omega_c \sim \begin{cases} \frac{1}{\langle k\rangle}\left[\frac{\Lambda}{\langle k\rangle}\right]^{\frac{1}{\beta-1}} & \beta \leq 3 \\ \\ \frac{1}{\langle k\rangle}\sqrt{\frac{\Lambda}{\langle k\rangle}} & \beta > 3. \end{cases} \tag{10.39}$$

The fact that $(\beta - 1)^{-1} < 1/2$ implies that, in this regime, a certain eigenvalue leads to a much lower frequency in the case of small $\beta$'s than it does for high $\beta$'s, an observation that is corroborated by Fig 10.4b.

Fig. 10.5 show the same analysis as in Fig. 10.3 but for the case $\beta = 1.5$, so that networks are deep into the small-world regime, with many long-range connections

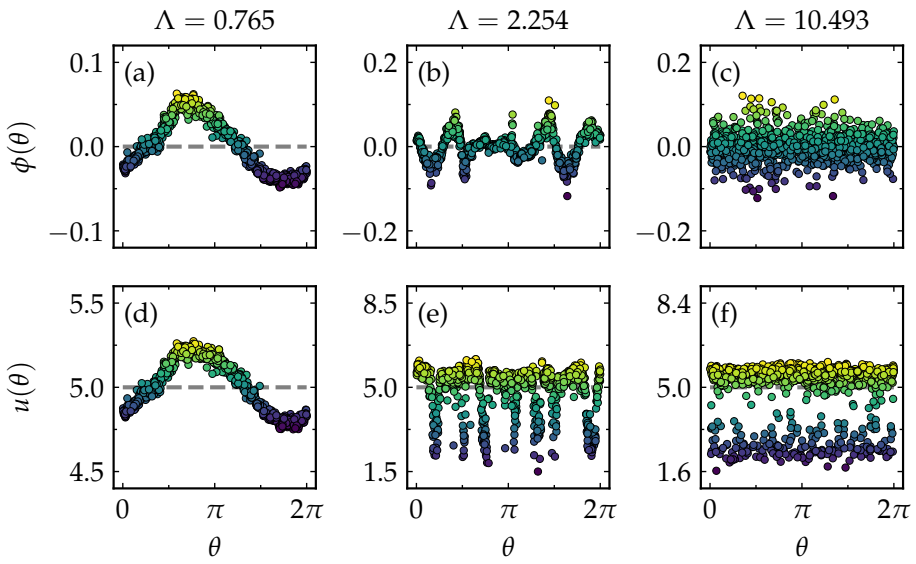



Figure 10.5: (a-c) Examples of three eigenvectors of the quenched Laplacian matrix with low, medium, and high associated frequencies for a single network generated by the homogeneous ensemble with $N = 1000$ and $\langle k \rangle = 20$ and $\beta = 1.5$. (d-f) Concentration of species $U$ in the Mimura-Murray dynamics.

among distant nodes in the metric space. We observe the same qualitative behavior as in the case $\beta = \infty$, except that, as expected, results have more noise. Yet, we can clearly identify geometric patterns both in the eigenvectors and in the steady state of the Mimura-Murray dynamics when the dynamic parameters are properly tuned. However, we notice that, in agreement with our theoretical prediction in Eq. (10.39), eigenvalues with similar associated eigenvector frequencies as those in Fig. 10.3 are significantly larger.

### 10.2.2 *Heterogeneous ensemble*

In real networks, the degree distribution is typically heterogeneous. Therefore, we have to solve the eigenvalue problem in Eq. (10.31) for a heterogeneous distribution of hidden degrees $\rho(\kappa)$. In particular, we choose a scale-free distribution $\rho(\kappa) \propto \kappa^{-\gamma}$. For this distribution of hidden degrees, Eq. (10.31) can be rewritten as

$$(\gamma - 2) \int_0^1 \frac{z'^{\gamma-3}}{1 - \tilde{\Lambda} z'} \hat{\Omega}\left(\frac{\tilde{\omega}}{z z'}\right) \hat{\Theta}(z', \tilde{\omega}) \mathrm{d}z' = \hat{\Theta}(z, \tilde{\omega}), \tag{10.40}$$

where we have redefined variables $(\kappa, \Lambda, \omega)$ as

$$z \equiv \frac{\kappa_0}{\kappa} \ , \ \tilde{\Lambda} \equiv \frac{\Lambda}{\kappa_0} \ , \ \tilde{\omega} \equiv \hat{\mu} \kappa_0^2 \omega \tag{10.41}$$

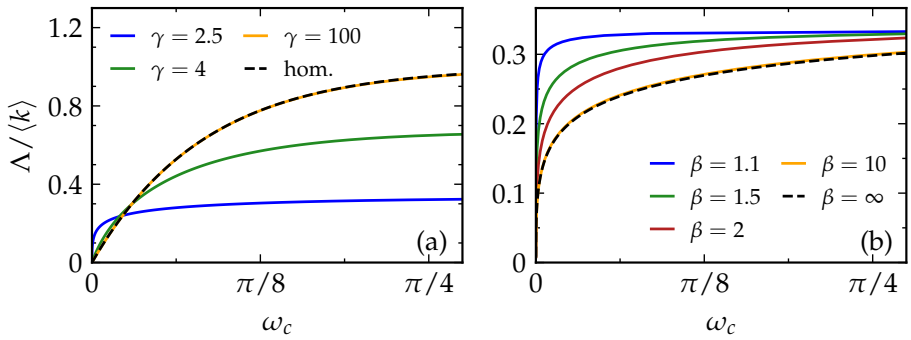



Figure 10.6: (a) Dispersion relation obtained with Gaussian quadrature for $\beta = 2$ for various $\gamma$'s and (b) the dispersion relation for $\gamma = 2.5$ and varying $\beta$. The average degree is $\langle k\rangle = 12$ in both panels.

and functions as

$$\hat{\Omega}(x) \equiv \frac{\hat{\Psi}(x)}{\hat{\Psi}(0)} \quad , \quad \hat{\Theta}(z,\tilde{\omega}) \equiv (1-\tilde{\Lambda}z)\hat{\phi}\left(\frac{\kappa_0}{z}, \frac{\tilde{\omega}}{\hat{\mu}\kappa_0^2}\right). \tag{10.42}$$

When $\tilde{\Lambda} < 1$, the singularity of the kernel in the integral of Eq. (10.40) falls outside the domain of integration. In this case, Eq. (10.40) can be solved numerically by Gaussian quadrature as the solution of the system of equations

$$\sum_{j=1}^{n}(\gamma-2)\frac{w_j z_j^{\gamma-3}}{1-\tilde{\Lambda}z_j}\hat{\Omega}\left(\frac{\tilde{\omega}}{z_i z_j}\right)\hat{\Theta}(z_j,\tilde{\omega}) = \hat{\Theta}(z_i,\tilde{\omega}), \tag{10.43}$$

where $z_i$, with $i = 1,\cdots,n$, are the zeros of the orthogonal polynomials used in the quadrature, $w_i$ their associated weights, and $n \gg 1$ the number of points within the domain of integration used to evaluate the integral.

Equation (10.43) defines a homogeneous system of linear equations with kernel matrix $\boldsymbol{K}(\Lambda,\omega)$ given by

$$K_{ij} \equiv (\gamma-2)\frac{w_j z_j^{\gamma-3}}{1-\tilde{\Lambda}z_j}\hat{\Omega}\left(\frac{\tilde{\omega}}{z_i z_j}\right) - \delta_{ij}. \tag{10.44}$$

A non-zero solution is found when $\det(\boldsymbol{K}(\Lambda,\omega)) = 0$, defining thus the dispersion relation $\Lambda(\omega_c)$. We use this condition to compute numerically the dispersion relation for different values of $\beta$ and $\gamma$. Figure 10.6a shows the results for $\beta = 2$ and different values of $\gamma$ and Fig. 10.6b shows the results for $\gamma = 2.5$ and different values of $\beta$. The dispersion relations are qualitatively similar to the homogeneous case with $\beta \leq 2$. The main difference appears in the asymptotic behavior of $\Lambda(\omega_c)$, which approaches $\kappa_0$ at high frequencies. Given the relation between the average degree and $\kappa_0$, this result implies that the ratio $\Lambda/\langle k\rangle$ approaches $(\gamma-2)/(\gamma-1)$,

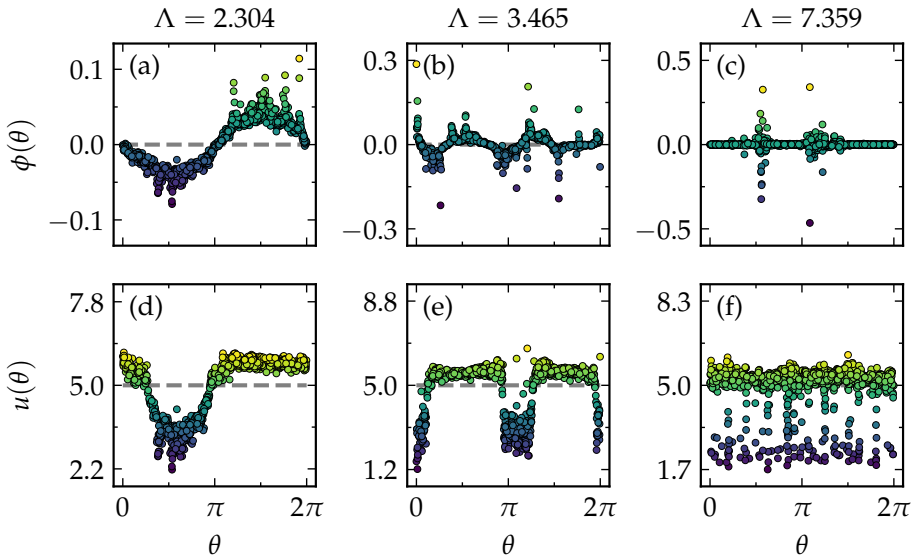



Figure 10.7: a-c) Examples of three eigenvectors with low, medium, and high associated frequencies for a single network generated by the heterogeneous ensemble with $N = 1000$, $\beta = 2.5$, $\gamma = 2.5$, and $\langle k \rangle = 20$. d-f) Concentration of species $U$ in the Mimura-Murray dynamics for the eigenvalues of the left column.

a value that is below 1. Therefore, the condition $\langle \Lambda \rangle = \langle k \rangle$ implies that the number of eigenvalues with periodic behavior can only account for a small fraction of all eigenvalues, and this fraction vanishes when $\gamma \approx 2$. We conjecture that the remaining eigenvalues, with $\Lambda > \kappa_0$, are those which cannot be characterized by a single frequency.

Figure 10.7 shows the same analysis performed in Figs. 10.3 and 10.5 but for heterogeneous networks with $\gamma = 2.5$ and $\beta = 2.5$. The low frequency eigenvector is clearly periodic, although we observe some deviations corresponding to low degree nodes. By increasing the eigenvalue, the periodicity is still preserved although the sinusoidal behavior is strongly modified. Finally, periodicity is destroyed at high eigenvalues. This behavior is, again, translated into the steady state of the Mimura-Murray dynamics so that, even in the presence of strong heterogeneity in the degree distribution, we are able to find geometric Turing patterns in the dynamics.

### 10.2.3 *Numerical results on synthetic $\mathbb{S}^1/\mathbb{H}^2$-networks*

To confirm numerically our theoretical predictions for the dispersion relations of quenched networks, we first generate a large number of them from the $\mathbb{S}^1$-model for a given set of parameters $\beta, \gamma$, and $\langle k \rangle$. For each such network, we compute its eigenvectors and eigenvalues sorted in increasing order. We then compute the

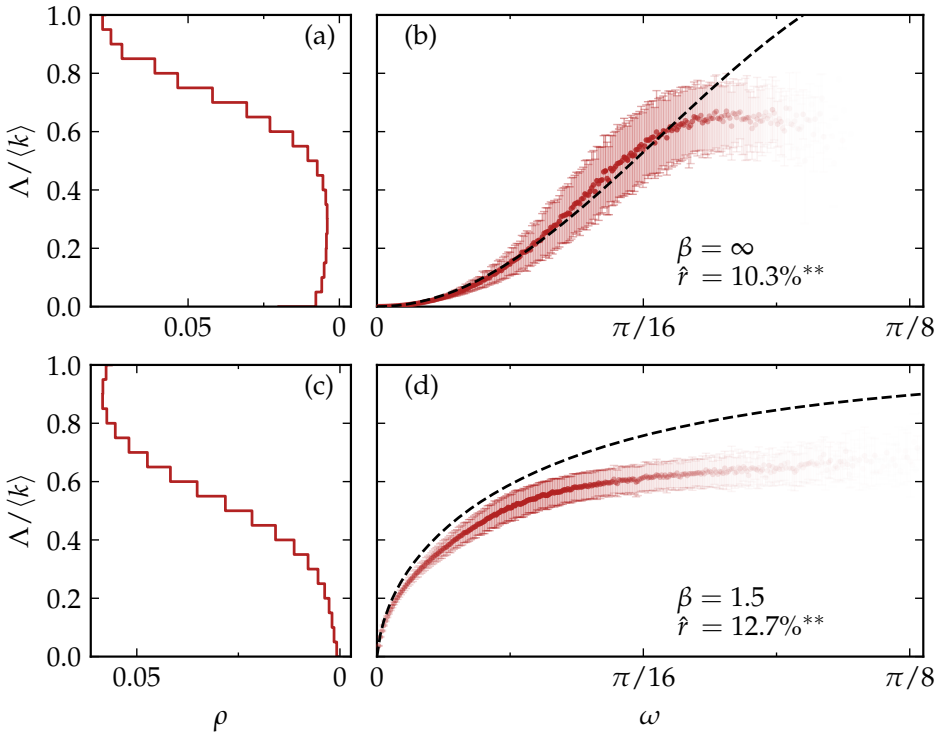



Figure 10.8: (a,c) Spectrum of an ensemble of networks generated by the homogeneous ensemble with $\beta = \infty$ (panel (a)) and $\beta = 1.5$ (panel (c)). In all cases the average degree is given by $\langle k\rangle = 12$. (b,d) Dispersion relation obtained by applying the Fast Fourier Transform to the eigenvectors and detecting the frequency with highest contribution. The opacity of the points is proportional to total amount of eigenvectors with the frequency corresponding to the point. The more transparent a point is the fewer eigenvectors have that frequency. The black dashed lines are the theoretical predictions given by the annealed approximation. The parameter $\hat{r}$ gives the percentage of eigenvectors classified as periodic. The notation $^{**}$ implies that the results are significant with $p = 0.01$.

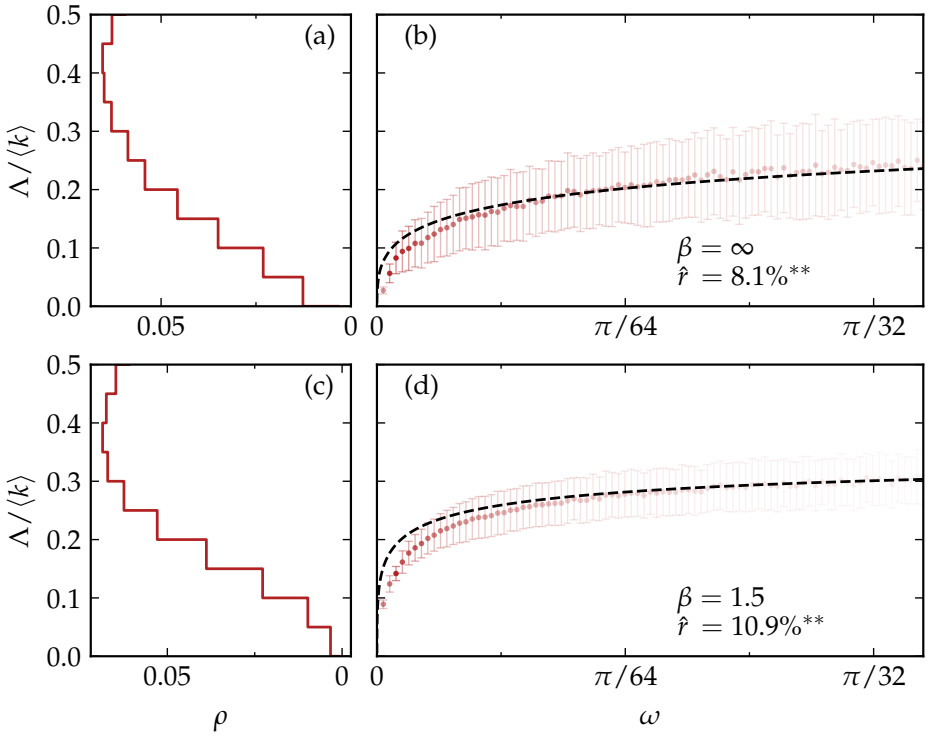



Figure 10.9: (a,c) Spectrum of an ensemble of networks generated by the heterogeneous ensemble with $\beta = \infty$ (panel (a)) and $\beta = 1.5$ (panel (c)). In all cases the average degree is given by $\langle k\rangle = 12$ and $\gamma = 2.5$. (b,d) Dispersion relation obtained by applying the Fast Fourier Transform to the eigenvectors and detecting the frequency with highest contribution. The opacity of the points is proportional to total amount of eigenvectors with the frequency corresponding to the point. The more transparent a point is the fewer eigenvectors have that frequency. The black dashed lines are the theoretical predictions given by the annealed approximation. The parameter $\hat{r}$ gives the percentage of eigenvectors classified as periodic. The notation $^{**}$ implies that the results are significant with $p = 0.01$.

discrete Fourier transform[1] of the eigenvectors and measure the characteristic frequency as the one corresponding to the highest peak in the Fourier spectrum. To asses whether a peak in the spectrum is statistically significant, we define a null model where the entries of the components of the eigenvector are independently drawn from a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma^2 = N^{-1}$. The first condition comes from Eq. (10.12) and the second from the fact that eigenvectors are normalized. Denoting $\hat{\phi}_i^{\text{rand}}$ as one of the $N$ entries of the DFT of a random eigenvector from our null model, it can be shown that it satisfies the following distribution function

$$\text{Prob}\left(|\hat{\phi}_i^{\text{rand}}|^2 < q\right) = 1 - e^{-q}. \tag{10.45}$$

With this in hand, we can find the value $q$ for which there is a probability $p$ that at least one peak in the white noise Fourier spectrum lies above $q$. We choose a probability $p = 10^{-2}$ and take the corresponding value of $q$ to be the minimal value a peak in the Fourier spectrum of one of the eigenvectors of a network generated by the model needs to cross to be considered periodic. Thus,

$$p = 1 - \prod_{i=1}^{N} \text{Prob}\left(|\hat{\phi}_i|^2 < q\right), \tag{10.46}$$

so that

$$q = -\ln\left[1 - (1-p)^{\frac{1}{N}}\right]. \tag{10.47}$$

For each network, we repeat this procedure for all its eigenvectors, leaving us with a set of pairs $(\Lambda, \omega_c)$ of statistically significant characteristic frequencies, from where we can derive the dispersion relation as an average over many network realizations.

The results of this program are shown in Fig. 10.8 for the homogeneous ensemble with $\beta = \infty$ and $\beta = 1.5$, and in Fig. 10.9 for the heterogeneous ensemble with $\gamma = 2.5$ and $\beta = \infty$ and $\beta = 1.5$. As can be seen, the agreement with the annealed approximation is very good in all cases, with small deviations in the case of $\beta = 1.5$. As mentioned in our previous discussion, it is not possible to find periodic behavior at high frequencies as the fluctuations of nodes' degrees induce incoherent signals in the eigenvectors' structures. In the examples shown in Figs. 10.8 and 10.9, the null model combined with the DFT –using a confidence level of two sigmas– is only able to detect between 8% and 12% of periodic eigenvalues.

## 10.3 WEAKLY GEOMETRIC GRAPHS

The results above have all been for the strongly geometric regime of the $\mathbb{S}^1/\mathbb{H}^2$-model. However, as argued in previous Chapters, geometry still plays a role in

1 We take a regular DFT, even though in principle the nodes are not regularly spaced along the circle. However, non-uniform discrete Fourier transforms (NUDFTs) are generally very slow to implement and therefore not suitable for large networks. Some preliminary testing has shown that the different between the approaches is negligible, most likely due to the fact that $\theta \sim \mathcal{U}(0, 2\pi)$

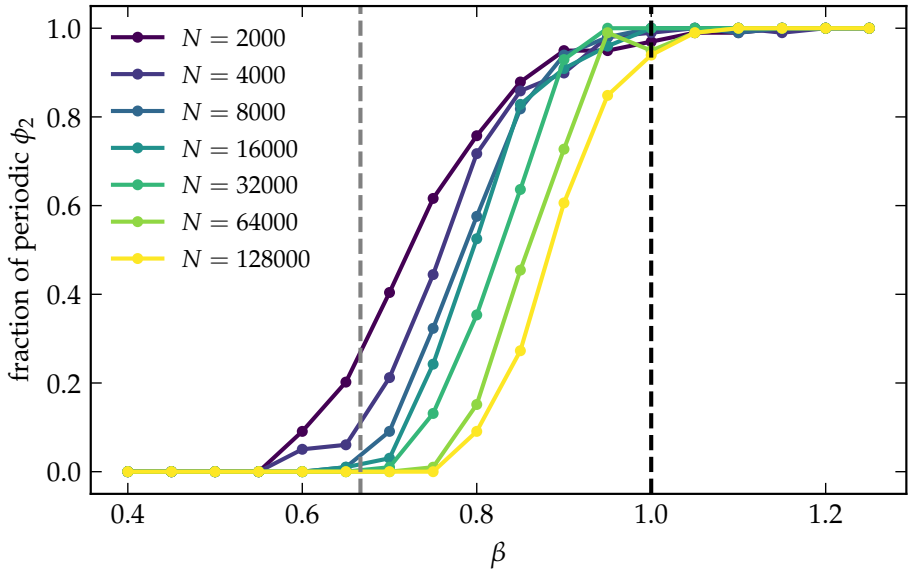



Figure 10.10: The fraction of network realizations that lead to a periodic $\phi^{(2)}$ as a function of $\beta$ for various $N \in [2000, 128000]$. For each $\beta$ and $N$ we generate 100 network realizations with $\langle k \rangle = 20$. The black dotted line marks $\beta_c = 1$ whereas the gray dotted line indicates $\beta_c' = 2/3$.

weakly geometric networks with $\beta \lesssim 1$. Therefore, in this section, we briefly explore this region through numerical means.

We focus on the periodic behavior of the eigenvector associated to the first non-zero eigenvalue, $\phi^{(2)}$, because its large wavelength leads to the least amount of noise. Therefore, if no periodicity can be detected for this eigenvector, it is very unlikely that periodicity is present in any of the other $N - 2$ non-trivial eigenvectors. We generate 100 homogeneous artificial networks with $\langle k \rangle = 20$ for $N \in [2000, 128000]$ and $\beta \in [0.4, 1.2]$. Finally we compute the fraction of realizations that lead to a periodic $\phi^{(2)}$ where we take the probability of the periodicity being a result of white noise to be $p = 0.001$.

The results of this procedure are shown in Fig. 10.10. We first notice that we can once again distinguish three regions. For $\beta > 1$, practically all realization lead to a periodic eigenvector whereas for $\beta \lesssim 0.5$, no periodicity is detected. In the intermediate region, a finite fraction of networks lead to a patterned eigenvector, even though this effect diminishes with the system size.

In Fig. 10.11 we plot $\phi^{(2)}$'s for two different realizations with $\langle k \rangle = 20$, $N = 8000$ and $\beta = 0.85$. For these parameters, about 80% of realization lead to a periodic eigenvector. Even though the parameters are the same, the results in panels (a) and (b) are very different. While (a) shows a very clean periodic signal, (b) is, at first glance, completely dominated by a single entry, which corresponds to the

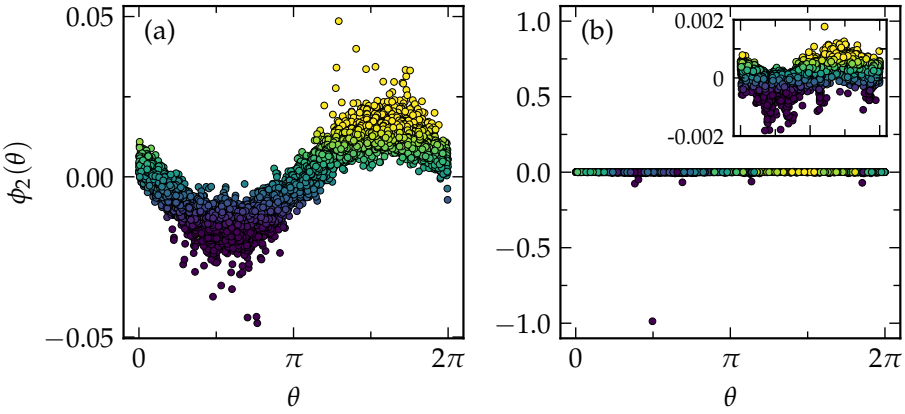



Figure 10.11: The eigenvector corresponding to the first non-trivial eigenvalue for two distinct homogeneous realizations with $\langle k \rangle = 20$, $N = 8000$ and $\beta = 0.85$. The realization in panel (a) shows a clean signal, whereas the result in panel (b) shows a signal dominated by an outlier. In both panels the color follows $\phi(\theta)$, rescaled such that the values lie between the 5th and 95th percentile of their original range. In panel (b), the inset shows an enhancement such that only the bulk nodes are shown.

node with the lowest degree. However, if one ignores this outlier and zooms into the bulk of the nodes as done in the inset of this panel, we see that a periodic signal is still present, even though its amplitude is very small. So small, in fact, that our DFT is not able to pick up the signal. Preliminary investigations indicate that it is the presence of these outliers, leading to very localized eigenvectors [311], which causes the finite size decay observed in Fig. 10.10.

These results can, surprisingly, be linked to those found in Sec. 6.2. There, we saw that for $\beta \lesssim 1$, the original nodes coordinates could be recovered with the embedding tool Mercator. Interestingly, this might be due to periodicity of the Laplacian eigenvectors in this range. As explained in the introduction of this thesis, the first step of Mercator uses Laplacian Eigenmaps (see Sec. 4.5 for details). Here, the angular coordinates are obtained by taking the points $\{\tilde{\phi}_i^{(2)}, \tilde{\phi}_i^{(3)}\}_{i=1}^N$, where $\tilde{\phi}_i^{(2)}, \tilde{\phi}_i^{(3)}$ are the second and third generalized eigenvectors of a Laplacian, weighted according to the expectation of links being present in the $\mathbb{S}^1$-model. Of course, if these eigenvectors form a perfect sine and cosine in the original underlying space, these points lie on a circle and the original coordinates are recovered. Noise like the one observed for example in Fig. 10.10a is taken care of by only keeping the angular coordinate of the aforementioned points. If the eigenvectors do not form a periodic pattern in the original space, it is impossible for the Laplacian Eigenmaps method to work. This is most likely why no geometric information can be extracted from the adjacency matrix in the non-geometric regime.

Of course, there are several differences between the results found by Mercator and the ones shown above. First of all, in Mercator one solves the generalized eigenvector equation, which diminishes the effect of the node degrees, leading to less outliers. Secondly, Mercator uses a weighted Laplacian matrix where information from the model is added to improve the embedding. And finally, Mercator performs many tuning steps after initializing using Laplacian Eigenmaps. All these differences might, for example, explain why Fig. 10.10 indicates that periodicity in the weakly geometric regime is a finite size effect, whereas the ability of Mercator to recover the original coordinates was not (see Fig. 6.3).

## 10.4 REAL NETWORKS

To analyze the emergence of Turing patterns in real networks, we first need to find an embedding of the network under study in the hyperbolic plane. This amounts to finding, for each node of the network, its hidden degree $\kappa$ and angular coordinate $\theta$. For this task, we use Mercator [64]. We run the Mimura-Murray dynamics on the network and look for geometric patterns in the angular similarity space inferred by Mercator.

Of course, in the case of real networks, we expect less clean results as compared to synthetic graphs generated by the $\mathbb{S}^1/\mathbb{H}^2$-model. Indeed, there are four main factors affecting the emergence of Turing patterns in real networks. First, the embedding method is noisy, so that periodic patterns will also be noisy. Second, even if the embedding was perfect, the degree distribution is generally heterogeneous so that, as in the $\mathbb{S}^1/\mathbb{H}^2$-model, the range of observable frequencies is small. Third, as shown in [158, 159], some real networks may be better represented in similarity spaces with dimension higher than $D = 1$. Thus, a geometric pattern in a high dimensional space can be distorted in a lower dimensional projection, making its detection difficult. Finally, the angular distribution in real networks is not perfectly uniform, with significant fluctuations defining geometric communities. As shown in Ref. [294], these communities have an impact on the eigenvector structure of the Laplacian matrix, so that in this case we should expect patterns with mixed effects from both the geometry of the network and its community structure. Despite all these limitations, we expect to find patterns in real networks associated to small non-vanishing eigenvalues.

Figure 10.12 shows results for four different real networks from different domains. Detailed definitions of these networks can be found in Appendix E.1. Other examples can be found in the Supplementary Information of Ref. [65]. Fig. 10.12a,d,g,j show the structure of the eigenvector corresponding to a small eigenvalue (and low frequency) for the four studied networks. The geometric patterns in the similarity space found by Mercator are very clear, with a wave-length of the order of the system size. The same structure is translated into the dynamics, as shown in Fig. 10.12b,e,h,k, where the same pattern can be recognized. We also plot a trendline, obtained by applying the Savitzky-Golay filter to the original signal. Finally, Fig. 10.12c,f,i,l show the concentration of species $U$ in the hyperbolic representation of the different networks inferred by Mercator. It is worth stress-

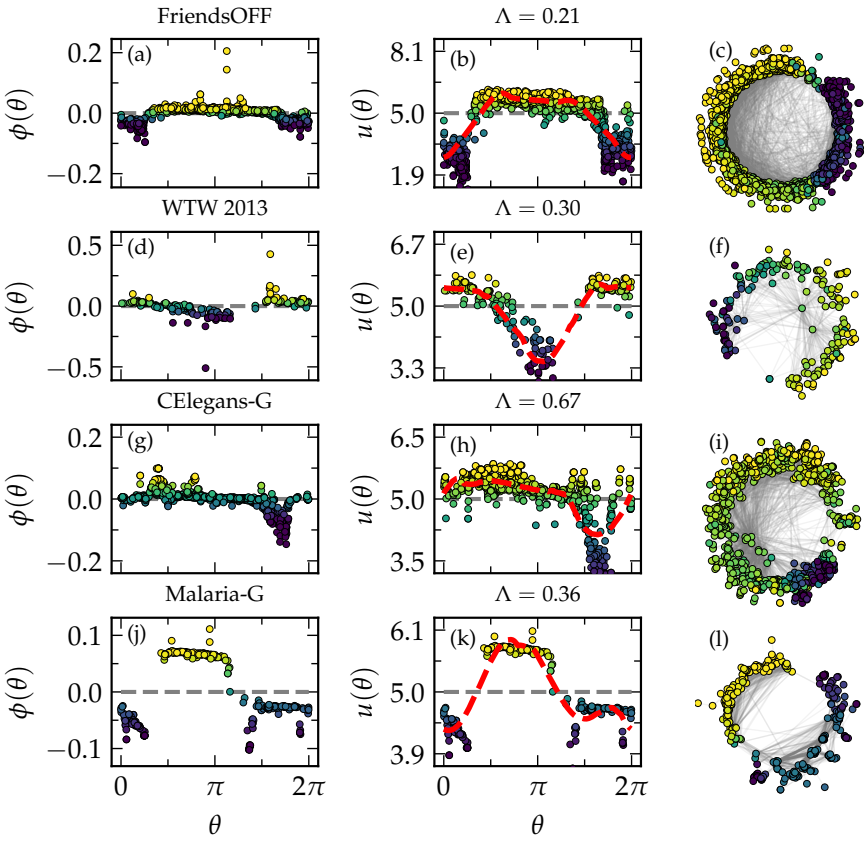



Figure 10.12: (a,d,g,j) Eigenvectors $\phi(\theta)$ corresponding to the eigenvalue $\Lambda$ as denoted in the figure for the real networks analyzed in this study, namely FriendsOFF, an off-line friendship network, WTW 2013, the World-Trade-Web, CElegans-G, a network of genetic interaction of the nematode C. Elegans and Malaria-G, a network of highly variable genes of the human Malaria parasite,. (b,e,h,k) Concentration of species $U$ for the Mimura-Murray dynamics setting the parameters to make these eigenvalues unstable. The red dotted line represents the trend obtained using the Savitzky-Golay filter with windowsize $\pi$. Panels (c,f,i,l) show the same as (b,e,h,k) but in the hyperbolic representation of the four networks.

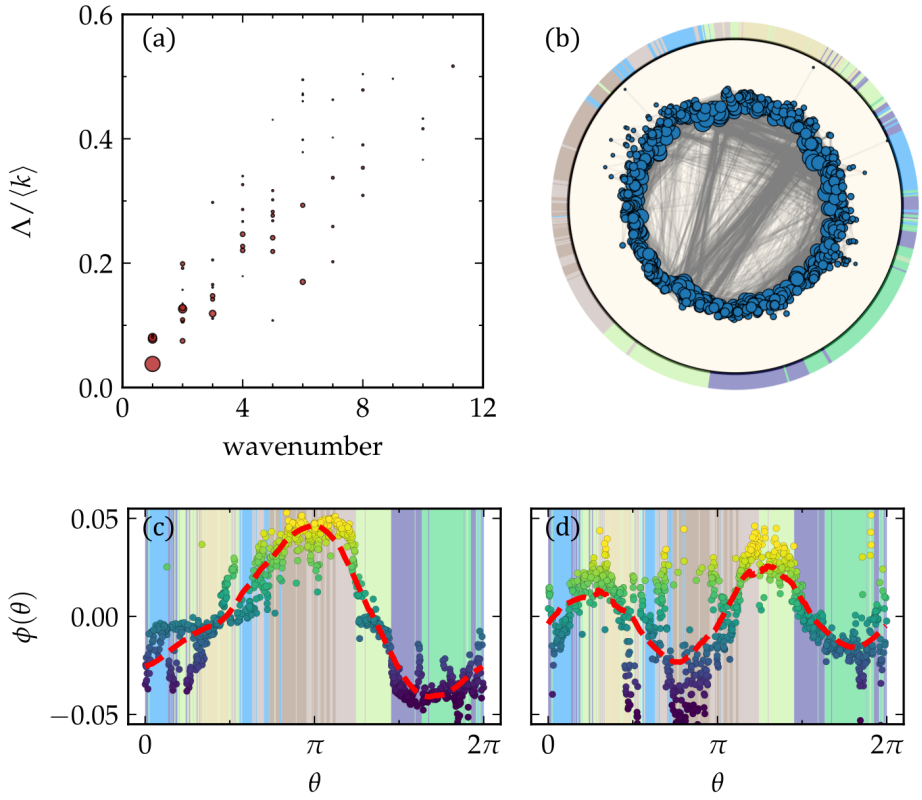



Figure 10.13: Analysis of a Human Connectome network [322]. a) Dispersion relation between the wavenumbers of the dominant mode of the Laplacian eigenvectors and their rescaled eigenvalues. The size of the points is proportional to the height of the largest peak in the Fourier spectrum. b) Hyperbolic representation of the network, where the node size is proportional to the hidden degree $\kappa$. The colored outer ring shows the partition in communities according to the Louvain algorithm [90]. c,d) The eigenvector with wavenumber one (c) and two (d) for which the pattern is most significant. The color follows $\phi(\theta)$ but rescaled such that the values lie between the 5th and 95th percentile of their original range. The background colors again represent the communities. The red dotted line represents the trend using the Savitzky-Golay filter with windowsize $\pi$.

ing how highly non-trivial this result is. Indeed, the Mimura-Murray dynamics generates steady states with correlation lengths of the order of the system size $N$ in small-world networks where any pair of nodes is separated by a very small number of intermediate steps, scaling as $\ln N$. Besides, this is achieved despite the fact that neither Mercator nor the Mimura-Murray dynamics use geometric information from the latent space as they only take the bare network topology as input.

Finally, we analyze in detail a human connectome network from [322] (see App. E.1 for details). This is an interesting example for which a significant number of patterns of different wavelengths can be detected. Figure 10.13a shows the dispersion relation between eigenvalues showing significant periodic signals versus their wavenumber. Despite the caveats discussed earlier, the dispersion relation agrees qualitatively well with the ones from the $\mathbb{S}^1$-model. In Fig. 10.13c-f we show two of these patterns, corresponding to wavenumbers one and two, where the patterns are clearly visible. Although the network displays a modular or community organization, as revealed by a Louvain algorithm analysis, this alone cannot account for the observed patterns. Rather, geometry is responsible for this phenomenon.

## 10.5 CONCLUSIONS

The small-world property present in complex networks, suggests, *a priori*, that networks in the real world are infinite dimensional structures, so that geometry plays a marginal role. Yet, network geometry stands today as the most parsimonious and comprehensive description of this class of systems. However, beyond providing an explanation to the observed topological structure of real networks, it is not clear how network geometry influences the dynamics taking place on them. In this Chapter, we have shown that geometric patterns may emerge in complex networks in a wide class of reaction-diffusion dynamics, as observed also in other types of dynamical processes such as opinion dynamics [323] and games [324]. Such patterns are already encoded in the structure of the Laplacian matrix as a result of the (hidden) metric structure of real networks. Interestingly, reaction-diffusion processes taking place on such networks are able to sense their geometric nature even when they show extreme disorder induced by heterogeneous degree distributions and a large fraction of long-range links making them small-worlds. This is especially surprising in the case of the quasi-geometric regime, where long-range links dominate.

We have provided a theoretical framework to understand this phenomenon and to quantify the role that the different topological properties of networks have on the emerging patterns. In addition to developing the theoretical tools for studying geometric patterns formation, significant insight is gained from the fact that Turing patterns can also be detected in real complex networks when the hidden similarity space is properly highlighted. This strongly suggests that there must exists a deep connection between networks' functions and the underlying geometry.

# CONCLUSIONS

# 11

# CONCLUSIONS

Many ubiquitous network properties can be explained by the presence of a latent geometric space. In this thesis we have explored how the coupling strength between the geometry and the topology of a network affects its structural properties as well as the dynamical processes that it supports. We have found that geometry plays an important role, even when the coupling is weak.

Using the $\mathbb{S}^1/\mathbb{H}^2$-model, we first studied the clustering phase transition in Chapter 5. By mapping the model to a gas of fermions, we showed that the entropy diverges at the critical point, driven by the sudden presence of a macroscopic amount of long-range links. This in turn leads to the reorganization of chordless cycles, causing the topological phase transition of the clustering coefficient. This sudden change in behavior at a finite temperature is due to the exponential density of energy states. If such a density of states could be engineered in a physical system, the predicted results might be confirmed experimentally.

The extremely slow decay of the clustering coefficient in the weakly geometric regime raises the question as to what the role of geometry is in this regime. In Chapters 6 and 7 we investigated this from various perspectives.

We started by studying the embeddibility of the network for $\beta \leq 1$. We show that geometric information can be extracted from the topology in this region only in the quasi-geometric regime $\beta_c' < \beta \leq \beta_c$, which coincides with the theoretical predictions based on the scaling of the clustering coefficient. Because Mercator, the embedding tool used in this thesis, uses the Laplacian Eigenmaps method as an initial step, we have argued that these results might have to do with the periodicity of the eigenvectors of the graph Laplacian, but further research in this direction is necessary.

We then turned to graph renormalization, and showed that geometric information is necessary to obtain self-similar network replicas in the quasi-geometric regime. Future work might compare these results to those found using other renormalization methods [189, 241–243].

In Chapter 8 we showed that many real networks are indeed best described in the weakly geometric regime, where we distinguish non- and quasi-geometric graphs. Even though a significant number of networks (metabolic, internet, foodwebs, etc.) were surveyed, more research is needed to be able to make statements about what types of networks tend to fall into which category. We also showed that these networks can be faithfully embedded and renormalized.

We extended our inquiry into the weakly geometric regime to the realm of multiplexes in Chapter 9. We showed that while degree heterogeneity is sufficient to obtain finite levels of mutual clustering, geometry is necessary in order to have a macroscopic amount of overlapping links. We presented some initial results showing that these tendencies can also be observed in empirical networks. However, more real networks need to be investigated. Future work might also focus on the interplay between geometry, degree heterogeneity and hidden coordinate correlations, as in this thesis their effects on the edge and triangle overlap were mostly studied separately.

Finally, in Chapter 10, we turned to dynamical processes on top of networks. Specifically, we studied the Turing instability and showed that the underlying geometry reveals patterns that are hidden when approaching the network from a purely topological perspective. We showed that patterns can also be observed in the weakly geometric regime, even though noise becomes ever more dominant as $\beta$ is decreased. We presented a method to predict the nature of the observed patterns based on the structural properties of the network and the parameters of the process alone. While we have shown that these predictions generally hold, future works needs to investigate their quality as a function of the structural properties. For example, it is as of yet unclear at what $\beta$ the approximations made break down. We also observed that while real networks also support geometric patterns, these are often distorted. It was hypothesized that this might be due to real networks living in higher dimension similarity spaces, meaning that their embedding onto the $\mathbb{S}^1$-circle might destroy geometric information, crucial for observing clean patterns. Studying Turing patterns on the higher dimensional $\mathbb{S}^D$-model [163, 325] is therefore a promising next step. This might also lead to new ways of inferring the proper dimension for real networks, complementing other works in this direction [158, 159].

Throughout the thesis, the $\mathbb{S}^1/\mathbb{H}^2$-model was used for both analytic calculations as well as numerical simulations. However, as we saw in the introduction, this model being maximum entropy implies that it can be defined in different ways. In particular, depending on whether quantities like energy, degrees, and amount of links are taken as hard, soft of hypersoft constraints, different ensembles emerge. As argued in the introduction, under some circumstances these might be interchangeable for some purposes, especially when the networks under study are large. We used this conjecture to choose the ensemble in which calculations/simulations where easiest. For example, in Chapter 5 numerical simulations were performed with the degree preserving geometric randomization procedure. Here, the degree sequence is fixed exactly, making it a hard constraint. At the same time, energy is preserved only in expectation. This thus leads to a hybrid between a microcanonical and canonical ensemble. At the same time, analytic calculations assumed the (hyper)canonical ensemble, where one could interpret the integration over the hidden coordinates as either being a result of their resampling or as a large $N$ approximations of the sums over the fixed quantities, drawn from some distribution. In other words, they could be interpreted as being the result of annealed or quenched disorder, respectively. In general we have always compared

our numerical and analytic results and have found that they coincide, motivating the interchangeable use of the different ensembles. However, more research needs to be done to prove these equivalences rigorously. While some results are known for the non-geometric $\beta \to 0$ limit of the model [133, 139, 142, 326–328], they are mostly lacking for the geometric case.

In general in this thesis we have worked with unweighted and undirected graphs. Extending these results to the weighted [45] and directed [46] versions of the $\mathbb{S}^1/\mathbb{H}^2$-models should, therefore, be a priority. Especially in the case of the Turing patterns, directed networks are relevant as they allow for more realistic instability conditions [284].

Zooming out, the field of network geometry provides us with many avenues of future work. The combinatorial nature of the topological picture of complex networks makes analytic treatment of many processes unfeasible. The fact that network geometry endows the study of complex networks with the ability to work with continuous spaces makes it possible to apply many tools from physics and mathematics to these intricate structures, once again making analytical results available. For example, once the mapping from a small-world network to its underlying metric space has been established, real space renormalization becomes available. This mapping might also allow for novel insights into the multi-scale behavior of dynamic processes on top of networks. Reversely, complex networks and their geometric treatment might lead to new insights into traditional subfields of physics. For example, recent progress [329] has been made in applying the network geometric framework to the quantum gravity, explaining the very fabric of space and time. In all these cases, it is essential to understand the effect of the geometric coupling strength on the associated network. In order to apply the geometric framework, one needs to know in which coupling regime a network is located. This is the topic that this thesis has studied.

All in all, this thesis shows that the geometric framework is able to explain many important properties of complex networks. We showed that only a weak geometric coupling is necessary for the underlying metric space to have an effect, extending the set of empirical networks for which geometric information needs to be taken into account.

# Part IV

# APPENDIX

# A

# APPENDIX TO CHAPTER 5

## A.1 STATISTICAL PROPERTIES OF THE $\mathbb{S}^1$

In Sec. 5.1 it was shown that the density of states of the $\mathbb{S}^1$ for a power law hidden degree distribution is given by

$$\rho(\epsilon) = \frac{2a}{N} e^{\epsilon} \Theta\left(\epsilon_{\max} - \epsilon\right) \left(1 + e^{b(\epsilon_{\max} - \epsilon)} \left(b(\epsilon_{\max} - \epsilon) - 1\right)\right), \tag{A.1}$$

where $a = \min(1,\beta)\kappa_0^{2\max(1,1/\beta)}(\gamma-1)^2/(1+\min(1,\beta)(1-\gamma))^2$ and $b = 1 + \min(1,\beta)(1-\gamma)$. Note that $\epsilon_{\max} = \ln\left(\frac{N}{2\kappa_0^{2(\max(1,1/\beta))}}\right)$ corresponds to the energy of a pair of nodes with minimal hidden degree $\kappa_0$, exactly opposite one another on the $\mathbb{S}^1$-circle such that $\Delta\theta_{ij} = \pi$.

We now use this functional form to confirm several results from the thesis rigorously for the specific hidden degree distribution $\rho(\kappa) = (\gamma-1)\kappa_0^{\gamma-1}\kappa^{-\gamma}$. In particular, we study the thermodynamic quantities of the system, including the average amount of particles or links $\langle M \rangle$.

### A.1.1 *Average amount of particles*

We first turn to the expected amount of particles in the system:

$$\begin{aligned}\langle M\rangle &= \binom{N}{2}\int_{-\infty}^{\epsilon_{\max}} \mathrm{d}\epsilon \frac{\rho(\epsilon)}{1+e^{\beta(\epsilon-\mu)}} \\ &= ae^{\epsilon_{\max}} N\Bigg(\frac{b^2}{(b-1)^2} + e^{\beta(\epsilon_{\max}-\mu)}\Bigg(-\frac{b\Phi\left[-e^{\beta(\epsilon_{\max}-\mu)}, 2, \frac{1-b+\beta}{\beta}\right]}{\beta^2} - \frac{{}_2F_1\left[\begin{matrix}1, 1+\frac{1}{\beta}\\ 2+\frac{1}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right]}{1+\beta} \\ &\qquad + \frac{{}_2F_1\left[\begin{matrix}1, 1+\frac{1-b}{\beta}\\ 2+\frac{1-b}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right]}{1+\beta-b}\Bigg)\Bigg).\end{aligned} \tag{A.2}$$

Here, $\Phi[z,a,b]$ is the Lerch zeta function. If we now assume $e^{\beta(\epsilon_{\max}-\mu)} \gg 1$, we can approximate this as

$$\langle M\rangle \simeq ae^{(1+\beta)\epsilon_{\max}-\beta\mu} N\left(\frac{1}{\beta}\pi\csc\left(\frac{\pi}{\beta}\right)e^{-(1+\beta)(\epsilon_{\max}-\mu)} + \frac{b^2}{(1-\beta)(b+\beta-1)^2}e^{-2\beta(\epsilon_{\max}-\mu)}\right). \tag{A.3}$$

We know that $\epsilon_{\max} \sim \ln N$. If we now also assume that $\mu \simeq c \ln N$ where $c < 1$, it can be shown that for all $c$ the dominant contributions are

$$\langle M \rangle \simeq \begin{cases} \frac{a\pi}{\beta} N e^{\mu} \csc\left(\frac{\pi}{\beta}\right) & \text{if} \quad \beta > 1 \\ a \epsilon_{\max} N e^{\mu} & \text{if} \quad \beta = 1 \\ \frac{ab^2 N}{(1-\beta)(b+\beta-1)^2} e^{(1-\beta)\epsilon_{\max} + \mu\beta} & \text{if} \quad \beta < 1 \end{cases} \tag{A.4}$$

If we take $\langle M \rangle = N\langle k \rangle / 2$ (sparse network) we obtain

$$\mu \simeq \begin{cases} \ln\left(\frac{\beta \sin\left(\frac{\pi}{\beta}\right)}{2\pi \langle k \rangle}\right) & \text{if } \beta > 1 \\ \frac{1}{2\langle k \rangle \ln N} & \text{if } \beta = 1 \\ \frac{1}{\beta} \ln\left(\frac{N^{\beta-1}(1-\beta)}{2^{\beta}\langle k \rangle}\right) & \text{if } \beta < 1 \end{cases} \tag{A.5}$$

Note that in all these cases $e^{\beta(\epsilon_{\max}-\mu)} \gg 1$ and that these are exactly the same results as found in Sec. 4.2.2.

### A.1.2 *Grand potential*

With this we can now study the grand potential as defined in Sec. 5.1.

$$\begin{aligned} \Xi &= -\frac{1}{\beta}\binom{N}{2} \int_{-\infty}^{\epsilon_{\max}} \mathrm{d}\epsilon \rho(\epsilon) \ln\left(1 + e^{-\beta(\epsilon-\mu)}\right) \\ &= -\frac{aN}{\beta} e^{\epsilon_{\max}} \left\{ \frac{b}{\beta(1-b)} \Phi\left[-e^{\beta(\epsilon_{\max}-\mu)}, 2, \frac{1-b}{\beta}\right] + (-1)^{-1/\beta} e^{-(\epsilon_{\max}-\mu)} B_{-e^{\beta(\epsilon_{\max}-\mu)}}\left[1 + 1/\beta, 0\right] \right. \\ &\frac{1 - 2b + (b-1)b\epsilon_{\max}}{1 - b + \beta} \frac{\beta}{(1-b)^2} {}_2F_1\left[\begin{matrix} 1, 1 + \frac{1-b}{\beta} \\ 2 + \frac{1-b}{\beta} \end{matrix} ; -e^{\beta(\epsilon_{\max}-\mu)}\right] e^{\beta(\epsilon_{\max}-\mu)} \\ &+ \frac{\beta b}{(b-1)^3}(1 - \epsilon_{\max} + b(-3 + b + \epsilon_{\max})) + \frac{b^2}{(1-b)^2} \ln\left(1 + e^{-\beta(\epsilon_{\max}-\mu)}\right) \\ &\left. - \frac{b\beta}{(1-b)^2} \epsilon_{\max} {}_2F_1\left[\begin{matrix} 1, \frac{1-b}{\beta} \\ 1 + \frac{1-b}{\beta} \end{matrix} ; -e^{\beta(\epsilon_{\max}-\mu)}\right] e^{\beta(\epsilon_{\max}-\mu)} \right\}, \end{aligned} \tag{A.6}$$

where $B_z[a, b]$ is the incomplete beta function. Again assuming that $e^{\beta(\epsilon_{\max}-\mu)} \gg 1$ and $b < 1$ we get the following dominant terms, after having divided out $\langle M \rangle$

$$\frac{\Xi}{\langle M \rangle} \simeq \begin{cases} -1 & \text{if} \quad \beta > 1 \\ -1 & \text{if} \quad \beta = 1 \\ -\frac{1}{\beta} & \text{if} \quad \beta < 1 \end{cases} \tag{A.7}$$

### A.1.3 *Energy*

Normally with this we have enough to calculate the entropy. However, in Chapter 5.1 we showed that $S = \beta^2(\frac{\partial \Xi}{\partial \beta})_\mu - \Delta$, where the correction $\Delta$ is due to the temperature dependence of the energy levels. We also showed that $S = -\beta(\Xi - \langle E\rangle + \mu\langle M\rangle)$ still holds, and here is it is easier to work with that formula. Thus, the final thing we need to do is find an expression for the average energy.

$$\begin{aligned}
\langle E\rangle &= \binom{N}{2}\int_{-\infty}^{\epsilon_{\max}} \mathrm{d}\epsilon \frac{\epsilon\rho(\epsilon)}{1+e^{\beta(\epsilon-\mu)}} \\
&= aNe^{\epsilon_{\max}}\Bigg\{ {}_2F_1\left[\begin{matrix}1, \frac{1-b}{\beta} \\ 1+\frac{1-b}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right]\epsilon_{\max} + \frac{1}{b-1}{}_2F_1\left[\begin{matrix}1, \frac{1-b}{\beta} \\ 1+\frac{1-b}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right]\epsilon_{\max} \\
&+ \frac{1+b\epsilon_{\max}}{(b-1)^2}{}_3F_2\left[\begin{matrix}1, \frac{1-b}{\beta}, \frac{1-b}{\beta} \\ 1+\frac{1-b}{\beta}, 1+\frac{1-b}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right] - {}_3F_2\left[\begin{matrix}1, \frac{1}{\beta}, \frac{1}{\beta} \\ 1+\frac{1}{\beta}, 1+\frac{1}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right] \\
&+ \frac{2b}{(b-1)^2}{}_4F_3\left[\begin{matrix}1, \frac{1-b}{\beta}, \frac{1-b}{\beta}, \frac{1-b}{\beta} \\ 1+\frac{1-b}{\beta}, 1+\frac{1-b}{\beta}, 1+\frac{1-b}{\beta}\end{matrix}; -e^{\beta(\epsilon_{\max}-\mu)}\right]\Bigg\}
\end{aligned} \tag{A.8}$$

We can again take the limit $e^{\beta(\epsilon_{\max}-\mu)} \gg 1$, dividing out $\langle M\rangle$, to obtain

$$\frac{\langle E\rangle}{\langle M\rangle} \simeq \begin{cases} \mu - \frac{\pi}{\beta}\cot\left(\frac{\pi}{\beta}\right) & \text{if} \quad \beta > 1 \\ \frac{1}{b} + \frac{1}{2}\epsilon_{\max} + \frac{1}{2}\mu & \text{if} \quad \beta = 1 \\ \epsilon_{\max} - \frac{b+3\beta-3}{(1-\beta)(b+\beta-1)} & \text{if} \quad \beta < 1. \end{cases} \tag{A.9}$$

### A.1.4 *Entropy*

Finally, this leads us to the entropy:

$$\frac{S}{\langle M\rangle} \simeq \begin{cases} \beta\left(1 - \frac{\pi}{\beta}\cot\left(\frac{\pi}{\beta}\right)\right) & \text{if} \quad \beta > 1 \\ \frac{1}{b} + \frac{1}{2}\epsilon_{\max} - \frac{1}{2}\mu + 1 & \text{if} \quad \beta = 1 \\ \beta\left(\epsilon_{\max} - \frac{b+3\beta-3}{(1-\beta)(b+\beta-1)} + \frac{1}{\beta} - \mu\right) & \text{if} \quad \beta < 1 \end{cases} \tag{A.10}$$

Now we plug in the remaining variables to obtain

$$\frac{S}{\langle M\rangle} \simeq \begin{cases} \beta - \pi\cot\left(\frac{\pi}{\beta}\right) & \text{if } \beta > 1 \\ \frac{1}{2}\ln N + \ln\ln N + 1 - \ln\langle k\rangle + 2\ln\left(\frac{\gamma-1}{\gamma-2}\right) + \frac{2}{2-\gamma} & \text{if } \beta = 1 \\ \ln N + \frac{2\beta}{1-\beta} - \ln(1-\beta) - \ln\langle k\rangle + 2\ln\left(\frac{\gamma-1}{\gamma-2}\right) + \frac{2}{2-\gamma} & \text{if } \beta < 1 \end{cases} \tag{A.11}$$

The final entropy is, as expected, equal to that of the an Erdös-Renyi graph with connection probability $\langle k \rangle / N$ when $\beta \to 0$ and $\gamma \to \infty$ and gives the entropy of the soft configuration model when only the first limit is taken [142].

### A.1.5 *Toy model*

Above we have seen some interesting bahavior, most notably the non-extensivity of the entropy above the critical temperature. We want to now investigate where this feature comes from, by looking at a simplified version of our model, where we assume classical indistinguishable particles with Maxwell-Boltzmann statistics. Suppose we have a system made of exactly $M$ non-interacting particles, each of which can attain states of energy $\epsilon \in (0, \epsilon_{\max})$. Suppose also that the density of states of energy $\epsilon$ grows as

$$\rho(\epsilon) = e^{\beta_c \epsilon} \tag{A.12}$$

with $\beta_c$ a fixed parameter. In the canonical ensemble, the probability that a certain particle $i$ is in state $\epsilon$ is then given by $p_i(\epsilon) = \frac{V\rho(\epsilon)}{Z_i} e^{-\beta\epsilon}$, where $Z_i = \int V g(\epsilon) e^{-\beta\epsilon} \mathrm{d}\epsilon$. Performing this integral we obtain

$$p_i(\epsilon) = \frac{\beta - \beta_c}{1 - e^{-(\beta-\beta_c)\epsilon_{\max}}} e^{-(\beta-\beta_c)\epsilon}. \tag{A.13}$$

We notice that here we find the same sudden change of behavior at the critical point $\beta = \beta_c$ as we found in the $\mathbb{S}^1$ model. Using Maxwell-Boltzmann statistics for identical particles in the canonical ensemble, the total partition function is then given by

$$Z = \frac{1}{M!} (Z_i)^M = \frac{V}{M!} \left( \frac{1 - e^{(\beta_c - \beta)\epsilon_{\max}}}{\beta - \beta_c} \right)^M \tag{A.14}$$

With this expression we calculate the entropy per particle using $S = \beta^2 \frac{\partial}{\partial \beta} (-\frac{\ln(Z)}{\beta})$ to obtain

$$\frac{S}{M} = \frac{\beta}{\beta - \beta_c} - \frac{\beta \epsilon_{\max}}{e^{(\beta-\beta_c)\epsilon_{\max}} - 1} - \ln \left[ \frac{M}{V} \frac{\beta - \beta_c}{1 - e^{-(\beta-\beta_c)\epsilon_{\max}}} \right] + 1, \tag{A.15}$$

where we have employed Stirling's approximation for $M \gg 1$.

The first two terms in this last equation are just the average energy per particle of the system. If the density of particles is kept fixed, so that $\lim_{M\to\infty} \frac{M}{V} = \mathrm{cst}$, then entropy is an extensive quantity as it is proportional to the number of particles. However, there is a clear change of behavior as one goes from $\beta > \beta_c$ to $\beta < \beta_c$

due to the change of behavior of the probability density Eq. (A.13). If besides $\beta\epsilon_{\max} \gg 1$, then the entropy behaves as

$$\frac{S}{M} \simeq \begin{cases} \frac{\beta}{\beta-\beta_c} - \ln\left[\frac{M}{V}(\beta-\beta_c)\right] + 1 & \beta > \beta_c \\ \frac{1}{2}\beta_c\epsilon_{\max} - \ln\left[\frac{M}{V\epsilon_{\max}}\right] + 1 & \beta = \beta_c \\ \beta_c\epsilon_{\max} + \frac{\beta}{\beta-\beta_c} - \ln\left[\frac{M}{V}(\beta_c-\beta)\right] + 1 & \beta < \beta_c \end{cases} \tag{A.16}$$

Thus, in the limit of $\epsilon_{\max} \to \infty$ the entropy per particle diverges at $\beta \to \beta_c^+$ and scales as $\epsilon_{\max}$ for $\beta \leq \beta_c$, just as in our model.

In the $\mathbb{S}^1$-model, the effective system size is given by the proportionality constant $V_{\text{eff}} = \binom{N}{2}\frac{2a}{N} \simeq aN$ and the amount of particles is given by $\langle M\rangle = N\langle k\rangle/2$ as we are working with a sparse graph. In this case we indeed satisfy $\lim_{\langle M\rangle\to\infty} \frac{\langle M\rangle}{V_{\text{eff}}} = \text{cst}$ and the entropy is thus in principle extensive. However, as in the full model there is the extra constraint $\langle M\rangle \leq N(N-1)/2$ and $\epsilon_{\max} = \ln(N/(2\kappa_0^2))$, we are obliged to also send $\epsilon_{\max}$ to infinity when going to the thermodynamic limit, thus resulting in a non-extensive entropy for $\beta < \beta_c$. We thus show that the essential feature of the $\mathbb{S}^1$ model that leads to a non-extensive entropy is the exponential dependence on the energy of the density of states.

## A.2 FINITE SIZE SCALING OF THE CLUSTERING COEFFICIENT IN THE $\mathbb{S}^1$

In this section we find the dominant finite size scaling of the clustering coefficient for $\beta \leq 1$. As was explained in Chapter 5, in this region clustering vanishes in the thermodynamic limit. We therefore study what happens when $N \gg 1$ but finite for any $\beta$ (we thus do not take any limit with respect to the temperature). As for $\beta \lesssim 1$ higher order finite size correction become important, we study the case $\beta = 1$ separately .

We start by manipulating the angular integrals of Eq. (5.21) as to simplify the task at hand later on. We then turn to the scaling when $\beta < 1$ and conclude with an analysis of the scaling when $\beta = 1$.

The basis of these calculations is the fact that we are looking for the scaling behaviour of the $\bar{c}$ with respect to the system size $N$. This allows us to always ignore terms that we know are smaller than than the main term, which simplifies the integrals that we study substantially. Once we have a term, say $A$, we want to know the scaling behaviour of, we use the fact that if the functions $f(N)$ and $g(N)$ in equation

$$f(N) \leq A \leq g(N) \tag{A.17}$$

have the same dominant scaling, one can immediately conclude that $A$ also has that exact dominant scaling. Therefore, by finding upper and lower bounds to the integrals in question we can extract there scaling behavior with respect to $A$. It is important to keep in mind that, even when the integrals representing the bounds

become very tedious, the strategy we employ remains the same throughout this section.

### A.2.1 *Angular manipulation*

We start by manipulating the angular integrals of Eq. (5.21) to make it easier to work with, i.e. get rid of the absolute values in the expressions for $\Delta\theta$. The equation has the following form:

$$\bar{c}(\kappa) = \frac{\iiint \mathrm{d}\kappa' \mathrm{d}\kappa'' \mathrm{d}\theta' \mathrm{d}\theta'' \rho(\kappa')\rho(\kappa'') p(\kappa,\kappa',\pi-|\pi-|\theta'||) p(\kappa,\kappa'',\pi-|\pi-|\theta''||) p(\kappa',\kappa'',\pi-|\pi-|\theta'-\theta''||))}{\iint \mathrm{d}\kappa' \mathrm{d}\theta' \rho(\kappa') p(\kappa,\kappa',\pi-|\pi-|\theta'||)}. \tag{A.18}$$

Here, without loss of generality, we have used $\theta = 0$. Let us first investigate the trivial case of the denominator, where we only focus on the angular integral

$$\begin{aligned}\int_0^{2\pi} \mathrm{d}\theta' p(\kappa,\kappa',\pi-|\pi-|\theta'||) &= \int_0^{\pi} \mathrm{d}\theta' p(\kappa,\kappa',\pi-|\pi-|\theta'||) + \int_{\pi}^{2\pi} \mathrm{d}\theta' p(\kappa,\kappa',\pi-|\pi-|\theta'||) \\ &= \int_0^{\pi} \mathrm{d}\theta' p(\kappa,\kappa',\theta') + \int_{\pi}^{2\pi} \mathrm{d}\theta' p(\kappa,\kappa',2\pi-\theta') = 2\int_0^{\pi} \mathrm{d}\theta' p(\kappa,\kappa',\theta'),\end{aligned} \tag{A.19}$$

where in the last step we have performed the transformation $t = 2\pi - \theta'$ and $t \to \theta'$ on the second integral.

The numerator can be rewritten in a similar way to obtain four terms

$$\begin{aligned}&\int_0^{2\pi} \mathrm{d}\theta' \int_0^{2\pi} \mathrm{d}\theta'' p(\kappa,\kappa',\pi-|\pi-|\theta'||) p(\kappa,\kappa'',\pi-|\pi-|\theta''||) p(\kappa',\kappa'',\pi-|\pi-|\theta'-\theta''||) \\ =&2\int_0^{\pi} \mathrm{d}\theta' \Bigg( \int_0^{\theta'} \mathrm{d}\theta'' p(\kappa,\kappa',\theta') p(\kappa,\kappa'',\theta'') p(\kappa',\kappa'',\theta'-\theta'') \\ &+ \int_0^{\theta'} \mathrm{d}\theta'' p(\kappa,\kappa',\theta'') p(\kappa,\kappa'',\theta') p(\kappa',\kappa'',\theta'-\theta'') \\ &+ \int_0^{\pi-\theta'} \mathrm{d}\theta'' p(\kappa,\kappa',\theta') p(\kappa,\kappa'',\theta'') p(\kappa',\kappa'',\theta'+\theta'') \\ &+ \int_{\pi-\theta'}^{\pi} \mathrm{d}\theta'' p(\kappa,\kappa',\theta') p(\kappa,\kappa'',\theta'') p(\kappa',\kappa'',2\pi-\theta'-\theta'') \Bigg).\end{aligned} \tag{A.20}$$

The first two terms are not exactly the same. However, as the full expression of the clustering coefficient also contains integrals over the hidden degrees, one can interchange $\kappa' \leftrightarrow \kappa''$. This shows that the first two terms contribute equally to the clustering coefficient. All in all, we will thus be working with the following three terms

$$\begin{aligned}&4\int_0^{\pi} \mathrm{d}\theta' \int_0^{\theta'} \mathrm{d}\theta'' p(\kappa,\kappa',\theta') p(\kappa,\kappa'',\theta'') p(\kappa',\kappa'',\theta'-\theta'') \\ &+2\int_0^{\pi} \mathrm{d}\theta' \int_0^{\pi-\theta'} \mathrm{d}\theta'' p(\kappa,\kappa',\theta') p(\kappa,\kappa'',\theta'') p(\kappa',\kappa'',\theta'+\theta'') \\ &+2\int_0^{\pi} \mathrm{d}\theta' \int_{\pi-\theta'}^{\pi} \mathrm{d}\theta'' p(\kappa,\kappa',\theta') p(\kappa,\kappa'',\theta'') p(\kappa',\kappa'',2\pi-\theta'-\theta'').\end{aligned} \tag{A.21}$$

Now, before we get started on finding the scaling with respect to the system size of each term individually, it might be that we can avoid doing so by some simple arguments. Indeed, we will show that the first term will always dominate the

others in the large $N$ limit, and so we only have to find its scaling. Let us start with the second term

$$\begin{aligned} &2\iint \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int_0^{\pi}\mathrm{d}\theta'\int_0^{\pi-\theta'}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',\theta'+\theta'') \\ \leq &2\iint \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int_0^{\pi}\mathrm{d}\theta'\int_0^{\pi}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',\theta'+\theta''). \end{aligned} \tag{A.22}$$

The above statement is true as the integrand is strictly positive and so extending the integration domain will only make the integral larger. Now, we can split the $\theta''$ integral and perform $\theta' \leftrightarrow \theta''$ and $\kappa' \leftrightarrow \kappa''$ on the second term to obtain

$$\begin{aligned} &2\iint \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int_0^{\pi}\mathrm{d}\theta'\int_0^{\pi}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',\theta'+\theta'') \\ &\quad = 4\iint \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int_0^{\pi}\mathrm{d}\theta'\int_0^{\theta'}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',\theta'+\theta'') \\ &\quad \leq 4\iint \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int_0^{\pi}\mathrm{d}\theta'\int_0^{\theta'}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',\theta'-\theta''). \end{aligned} \tag{A.23}$$

In the final step we use the functional form of $p$ with respect to the angular coordinate is

$$p(s) = \frac{1}{1+s^{\beta}}. \tag{A.24}$$

As $s^{\beta}$ is monotonously increasing, and $1/(1+s)$ is monotonously decreasing, $p(s)$ is monotonously decreasing. Thus, it is largest when $s$ is smallest. Obviously, $\theta'+\theta'' > \theta'-\theta''$ for all $(\theta',\theta'') \in [0,\pi]\times[0,\theta']$. We have thus proven that the first term in Eq. (A.21) dominates the second term. We can follow similar steps for the third term. We we will now only clarify steps if they are new.

$$\begin{aligned} &2\iint\limits_{\kappa',\kappa''} \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int\limits_0^{\pi}\mathrm{d}\theta'\int\limits_{\pi-\theta'}^{\pi}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',2\pi-\theta'-\theta'') \\ \leq &4\iint\limits_{\kappa',\kappa''} \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int\limits_0^{\pi}\mathrm{d}\theta'\int\limits_0^{\theta'}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',2\pi-\theta'-\theta''). \end{aligned} \tag{A.25}$$

Now, one knows that $2\pi-\theta'-\theta'' \geq \theta'-\theta''\ \forall(\theta',\theta'') \in [0,\pi]\times[0,\theta']$. For the same reasons as before, this then implies

$$\begin{aligned} &4\iint\limits_{\kappa',\kappa''} \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int\limits_0^{\pi}\mathrm{d}\theta'\int\limits_0^{\theta'}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',2\pi-\theta'-\theta'') \\ \leq &4\iint\limits_{\kappa',\kappa''} \mathrm{d}\kappa'\mathrm{d}\kappa''\rho(\kappa')\rho(\kappa'')\int\limits_0^{\pi}\mathrm{d}\theta'\int\limits_0^{\theta'}\mathrm{d}\theta'' p(\kappa,\kappa',\theta')p(\kappa,\kappa'',\theta'')p(\kappa',\kappa'',\theta'-\theta''), \end{aligned} \tag{A.26}$$

so this term is also dominated by the first term in Eq. (A.21).

### A.2.2 *Case* $0 < \beta < 1$

The first step is to perform the transformation $x = \frac{\kappa'}{\kappa_s}$ and $y = \frac{\kappa''}{\kappa_s}$, where we define $\kappa_s^2 \equiv N^\beta/((2\pi)^\beta \hat{\mu})$. Note that we use assume the functional form of $\hat{\mu}$ defined in Eq. (4.14), such that $\kappa_s \sim \sqrt{N}$. This leads to

$$\bar{c}(\kappa) \sim 2 \frac{\int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}x \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}y \int_0^{\pi} \mathrm{d}\theta' \int_0^{\theta'} \mathrm{d}\theta'' (xy)^{-\gamma} p(\kappa, \kappa_s x, \theta') p(\kappa, \kappa_s y, \theta'') p(\kappa_s x, \kappa_s y, \theta' - \theta'')}{\left( \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}x \int_0^{\pi} \mathrm{d}\theta' x^{-\gamma} p(\kappa, \kappa_s x, \theta') \right)^2}. \tag{A.27}$$

We investigate the numerator and denominator separately and define

$$A_- = \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}x \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}y \int_0^{\pi} \mathrm{d}\theta' \int_0^{\theta'} \mathrm{d}\theta'' (xy)^{-\gamma} p(\kappa, \kappa_s x, \theta') p(\kappa, \kappa_s y, \theta'') p(\kappa_s x, \kappa_s y, \theta' - \theta''). \tag{A.28}$$

$$B = \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}x \int_0^{\pi} \mathrm{d}\theta' x^{-\gamma} p(\kappa, \kappa_s x, \theta'). \tag{A.29}$$

It is also useful to define

$$A_+ = \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}x \int_{\kappa_0/\kappa_s}^{\kappa_c/\kappa_s} \mathrm{d}y \int_0^{\pi} \mathrm{d}\theta' \int_0^{\theta'} \mathrm{d}\theta'' (xy)^{-\gamma} p(\kappa, \kappa_s x, \theta') p(\kappa, \kappa_s y, \theta'') p(\kappa_s x, \kappa_s y, \theta' + \theta''). \tag{A.30}$$

Our investigation will focus on finding upper and lower bounds for these integrals. Note that from here on out we will drop the domains of the $x$ and $y$ integrals and assume them to be $[\kappa_0/\kappa_s, \kappa_c/\kappa_s]$ unless otherwise indicated. Using the fact that

$$\frac{1}{1 + \frac{(\theta' + \theta'')^\beta}{xy}} < \frac{1}{1 + \frac{(\theta' - \theta'')^\beta}{xy}}, \quad \forall \theta', \theta'', x, y, \tag{A.31}$$

we can conclude that $A_+ < A_-$. As numerical investigation leads us to expect that both have the same scaling, this implies that we do not need to worry about an upper bound for $A+$ nor the lower bound for $A_-$. If the functions $f(N)$ and $g(N)$ in equation

$$f(N) < A_+ < A_- < g(N) \tag{A.32}$$

have the same dominant scaling, one can immediately conclude that $A_-$ also has that exact dominant scaling. One might ask why we introduce $A_+$ in the first place, when in the end we are only interested in the scaling of $A_-$. The answer to this is that $A_+$ in general has nicer properties due to the lack of $(\theta' - \theta'')$, as it is thus easier to find a lower bound for it than for $A_-$.

We start with the simplest integral, the $B$-term, which can be solved exactly. To this end we first need to rewrite it a bit. By performing two substitutions

$$x' = \frac{\kappa_s}{\kappa_c} x \quad x' \to x, \qquad t = \frac{\theta'}{\pi} \quad t \to \theta', \tag{A.33}$$

one obtains

$$B = \pi \left(\frac{\kappa_c}{\kappa_s}\right)^{1-\gamma} \int_0^1 \mathrm{d}\theta' \int_{\kappa_0/\kappa_c}^1 \mathrm{d}x \frac{x^{-\gamma}}{1+\frac{(\pi\theta')^\beta \kappa_s^2}{\kappa_c \kappa x}}. \tag{A.34}$$

This then gives the following expression

$$\begin{aligned} B = \frac{\pi}{(\beta(\gamma-1)-1)(\gamma-1)} \Bigg\{ & (\gamma-1)\beta \left(\frac{\kappa_0}{\kappa_s}\right)^{1-\gamma} {}_2F_1\left[\begin{matrix}1, 1/\beta \\ 1+1/\beta\end{matrix}; -\frac{\pi^\beta \kappa_s^2}{\kappa\kappa_0}\right] \\ & -(\gamma-1)\beta \left(\frac{\kappa_c}{\kappa_s}\right)^{1-\gamma} {}_2F_1\left[\begin{matrix}1, 1/\beta \\ 1+1/\beta\end{matrix}; -\frac{\pi^\beta \kappa_s^2}{\kappa\kappa_0}\right] \\ & - \kappa_0 \left(\frac{\kappa_0}{\kappa_s}\right)^{1-\gamma} {}_2F_1\left[\begin{matrix}1, \gamma-1 \\ \gamma\end{matrix}; -\frac{\pi^\beta \kappa_s^2}{\kappa\kappa_0}\right] \\ & + \kappa_0 \left(\frac{\kappa_c}{\kappa_s}\right)^{1-\gamma} {}_2F_1\left[\begin{matrix}1, \gamma-1 \\ \gamma\end{matrix}; -\frac{\pi^\beta \kappa_s^2}{\kappa\kappa_0}\right] \Bigg\}. \end{aligned} \tag{A.35}$$

This expression can be expanded with respect to $N$, using that $\kappa_s \sim \sqrt{N}$. To lowest order, one finds that $B$ then scales as

$$B \sim N^{\frac{\gamma-3}{2}} \tag{A.36}$$

Next we turn to the $A_+$ term. Here we use the following fact to bound this integral. If $F = \int_{\mathcal{V}} f(\vec{x})$, where $\mathcal{V}$ is the volume over which to integrate the function, and $f(\vec{x}) \geq f_0$ for all $\vec{x}$, where $f_0$ some constant, then $F \geq f_0 \mathcal{V}$. From the form of the standard connection probability given in Eq. (A.24), we see that $A_+$ is smallest when the argument is largest, which is the case when $\theta', \theta''$ are largest, so when they are both $\pi$. Thus we can bound the angular integrals by replacing the integrand with its minimum, the same function where both angular coordinates are $\pi$. The integrand is then a constant so the bound is given by the value of that constant times the area of the integral. Plugging this in we obtain

$$\begin{aligned} A_+ &\geq \frac{1}{2}\pi^2 \int \mathrm{d}x\mathrm{d}y (xy)^{-\gamma} \frac{1}{1+\frac{\pi^\beta \kappa_s}{\kappa x}} \frac{1}{1+\frac{\pi^\beta \kappa_s}{\kappa y}} \frac{1}{1+\frac{(2\pi)^\beta}{xy}} \\ &= \frac{1}{2}\pi^{2-3\beta} \left(\frac{\kappa}{\kappa_s}\right)^2 \int \mathrm{d}x\mathrm{d}y (xy)^{2-\gamma} \frac{1}{1+\frac{\kappa x}{\pi^\beta \kappa_s}} \frac{1}{1+\frac{\kappa y}{\pi^\beta \kappa_s}} \frac{1}{1+\frac{xy}{(2\pi)^\beta}}. \end{aligned} \tag{A.37}$$

Now, this is exactly the same integral (with the exception of the $\pi$'s, but they will obviously not change scaling) as the one evaluated in Ref. [144]. As was found in the reference (Eq. (6)), the scaling depends on how we set $\kappa_c$ relative to $\kappa_s$. We distinguish two regimes. First, there is the regime where $\kappa_0 \ll \kappa_s \ll \kappa_c$. In this case, the scaling is

$$A_+ \geq c_{+,1} \kappa_s^{-2} \ln(\kappa_c/\kappa_s). \tag{A.38}$$

Then, there is the region where $\kappa_0 \leq \kappa_c \leq \kappa_s$ ($\kappa_0 \ll \kappa_s$ must be required to hold) where one obtains

$$A_+ \geq c_{+,2} \kappa_s^{2\gamma-8} \kappa_c^{6-2\gamma}. \tag{A.39}$$

This, however, does not give the full scaling behaviour, as numerical results show us that for large $\beta$ the scaling with respect to $N$ is different. To find where this different scaling comes from we take a step back and look at the full integral $A_+$ as given in Eq. (A.30). One might be tempted to, as in Ref. [144], expand the first two connection probabilities to first order. However, the presence of the angular coordinate makes this impossible. The argument of these connection probabilities has the form $s = \frac{\theta^\beta \kappa_s^2}{\kappa \kappa'}$. It becomes clear that for small enough $\theta$, $s$ is no longer large and the approximation thus breaks down. We thus expect different scaling behaviour to arise as a result of small angular coordinates. To investigate this further, we split the *angular* integration domain $[0,\pi]\times[0,t]$ in a convenient way and investigate the domain $\mathcal{D}_1 = [0,(xy)^{1/\beta}]\times[0,t]$. Note that we do not have to look at the other half of the original domain as we are only interested in the lower bound and our integrand is positive for all angles, which means that the integral over the full domain must be larger or equal to the integral over $D_1$. The domain $\mathcal{D}_1$ can only be defined in the case that $\kappa_c \leq \kappa_s$, as only then the angular coordinates remain smaller than the maximal possible value of $\pi$ for all $x$ and $y$. For the case that $\kappa_c \gg \kappa_s$ we define the more restrictive domain $\mathcal{D}_2 = [0,(\kappa_0/\kappa_s)^{2/\beta}]\times[0,t]$. Starting with the case $\kappa_c \leq \kappa_s$, bounding the integral as before (by replacing the integrand by its minimum), one finds

$$\begin{aligned} A_+ &\geq \frac{1}{1+2^\beta}\int \mathrm{d}x\mathrm{d}y (xy)^{2/\beta-\gamma} \frac{1}{1+\frac{\kappa_s y}{\kappa}}\frac{1}{1+\frac{\kappa_s x}{\kappa}} \\ &= \frac{(\kappa_s/\kappa)^{-4/\beta+2\gamma-2}}{1+2^\beta}\left(B_{\frac{\kappa}{\kappa_0+\kappa}}\left[\gamma-\frac{2}{\beta},1-\gamma+\frac{2}{\beta}\right]-B_{\frac{\kappa}{\kappa_c+\kappa}}\left[\gamma-\frac{2}{\beta},1-\gamma+\frac{2}{\beta}\right]\right)^2 \\ &\simeq c_{+,s,1}\kappa_s^{-4/\beta+2\gamma-2}+c_{+,s,2}\kappa_s^{-4/\beta+2\gamma-2}\kappa_c^{4/\beta-2\gamma}, \end{aligned} \tag{A.40}$$

For the case $\kappa_c \gg \kappa_s$ one obtains

$$\begin{aligned} A_+ &\geq \left(\frac{\kappa_0}{\kappa_s}\right)^{4/\beta}\int \mathrm{d}x\mathrm{d}y (xy)^{-\gamma}\frac{1}{1+\frac{\kappa_s}{\kappa x}\frac{\kappa_0^2}{\kappa_s^2}}\frac{1}{1+\frac{\kappa_s}{\kappa y}\frac{\kappa_0^2}{\kappa_s^2}}\frac{1}{1+\frac{2^\beta}{xy}\frac{\kappa_0^2}{\kappa_s^2}} \\ &\simeq \left(\frac{\kappa_0}{\kappa_s}\right)^{4/\beta}\int \mathrm{d}x\mathrm{d}y (xy)^{-\gamma} \simeq c_{+,s,3}\kappa_s^{-4/\beta+2\gamma-2}, \end{aligned} \tag{A.41}$$

where in the first step it was noted that irrespective of the value of $x$ and $y$, the argument of the connection probabilities is small.

We now have five different scaling behaviours. Which terms dominate will depend on the value of $\beta$ as well on $\kappa_c$. To quantify how the scaling varies with $\kappa_c$ we introduce the exponent $\alpha$ such that $\kappa_c \sim N^{\alpha/2}$. As $\kappa_s \sim N^{1/2}$, the different regimes of $\kappa_c$ described above correspond to $\alpha \in [0,1]$ for $\kappa_c \leq \kappa_s$ and $\alpha \in (1, \frac{2}{\gamma-1}]$ for $\kappa_c \gg \kappa_s$. Using these definitions and adding up the different scaling we found above, we conclude that

$$A_+ \geq \begin{cases} C_{+,1}N^{-2/\beta+\gamma-1}+C_{+,2}N^{-1}\ln N & \text{if} \quad \kappa_c \gg \kappa_s \\ N^{-1}\left(C_{+,3}N^{\gamma-2/\beta}+C_{+,4}N^{(1-\alpha)(\gamma-2/\beta)}+C_{+,5}N^{(1-\alpha)(\gamma-3)}\right) & \text{if} \quad \kappa_c \leq \kappa_s \end{cases}, \tag{A.42}$$

where $C_{+,i}$ are constants. Note that, for example, the scaling of Eqs. (A.38) and (A.41) can indeed be combined to the first of these two inequalities as both now

hold for all $\beta$. When $\beta > 2/\gamma$ the $C_{+,2}$-term vanished with respect to the $C_{+,1}$-term and we are left with inequality (A.41) and when $\beta < 2/\gamma$ the other term dominates and we are left with inequality (A.38).

Now obviously this is just a lower bound. To show that the clustering indeed scales like this we must also find an upper bound, which we do by turning to the $A_-$ term. We divide the integration domain in two: $\mathcal{D}_s = [0, (\kappa_0/\kappa_s)^{2/\beta}] \times [0, \theta']$ and $\mathcal{D}_l = [(\kappa_0/\kappa_s)^{2/\beta}, \pi] \times [0, \theta']$. We first turn to region $\mathcal{D}_l$.

$$
\begin{aligned}
A_{-,l} &= \iint_{\mathcal{D}_l} \mathrm{d}\theta' \mathrm{d}\theta'' \iint \mathrm{d}x\mathrm{d}y (xy)^{-\gamma} \quad \frac{1}{1+\frac{\theta'^{\beta}\kappa_s}{\kappa x}} \frac{1}{1+\frac{\theta''^{\beta}\kappa_s}{\kappa y}} \frac{1}{1+\frac{(\theta'-\theta'')^{\beta}}{xy}} \\
&\leq \left(\frac{\kappa}{\kappa_s}\right)^2 \iint_{\mathcal{D}_l} \mathrm{d}\theta' \mathrm{d}\theta'' \iint \mathrm{d}x\mathrm{d}y \quad (xy)^{2-\gamma} (\theta'\theta''(\theta'-\theta''))^{-\beta} \frac{1}{1+\frac{xy}{(\theta'-\theta'')^{\beta}}} \\
&= \left(\frac{\kappa}{\kappa_s}\right)^2 \iint_{\mathcal{D}_l} \frac{\mathrm{d}\theta' \mathrm{d}\theta''}{(\theta'\theta''(\theta'-\theta''))^{\beta}} \quad \left(\frac{\kappa_c}{\kappa_s}\right)^{2(3-\gamma)} \Phi\left[-(\theta'-\theta'')^{-\beta}\left(\frac{\kappa_c}{\kappa_s}\right)^2, 2, 3-\gamma\right] \\
&+ \left(\frac{\kappa}{\kappa_s}\right)^2 \iint_{\mathcal{D}_l} \frac{\mathrm{d}\theta' \mathrm{d}\theta''}{(\theta'\theta''(\theta'-\theta''))^{\beta}} \quad \left(\frac{\kappa_0}{\kappa_s}\right)^{2(3-\gamma)} \Phi\left[-(\theta'-\theta'')^{-\beta}\left(\frac{\kappa_0}{\kappa_s}\right)^2, 2, 3-\gamma\right] \\
&- 2\left(\frac{\kappa}{\kappa_s}\right)^2 \iint_{\mathcal{D}_l} \frac{\mathrm{d}\theta' \mathrm{d}\theta''}{(\theta'\theta''(\theta'-\theta''))^{\beta}} \left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{3-\gamma} \Phi\left[-(\theta'-\theta'')^{-\beta}\frac{\kappa_0\kappa_c}{\kappa_s^2}, 2, 3-\gamma\right]. \qquad \text{(A.43)}
\end{aligned}
$$

One sees that these three terms are similar, and so we treat the general integral

$$
I_\zeta = \iint_{\mathcal{D}_l} \mathrm{d}\theta' \mathrm{d}\theta'' \frac{\Phi\left[-(\theta'-\theta'')^{-\beta}\zeta, 2, 3-\gamma\right]}{(\theta'\theta''(\theta'-\theta''))^{\beta}} \zeta^{3-\gamma} = \iint_{\mathcal{D}_l} \mathrm{d}\theta' \mathrm{d}\theta'' \frac{\Phi\left[-\theta''^{-\beta}\zeta, 2, 3-\gamma\right]}{(\theta'\theta''(\theta'-\theta''))^{\beta}} \zeta^{3-\gamma}, \qquad \text{(A.44)}
$$

where the transformation $\theta''' = \theta' - \theta''$, $\theta''' \to \theta''$ was performed. Now, the argument of the Lerch zeta function can in principle be smaller and larger than one. If it is smaller, it can be shown that $\Phi[-(\theta'-\theta'')^{-\beta}\zeta, 2, 3-\gamma] < 2^{\gamma-3}$. If it is bigger than one can use the identity described in Ref. [144]

$$
\Phi[-z^2, 2, 3-\gamma] = z^{-2(3-\gamma)}\Big(2\psi(\gamma)\ln z + \vartheta(\gamma)\Big) + \frac{1}{z^2}\Phi\left[\frac{1}{z^2}, 2, \gamma-2\right], \qquad \text{(A.45)}
$$

where

$$
\psi(\gamma) = \Phi[-1, 1, 3-\gamma] + \Phi[-1, 1, \gamma-2] \qquad \text{(A.46)}
$$

$$
\vartheta(\gamma) = -\pi^2 \cot(\pi\gamma)\csc(\pi\gamma). \qquad \text{(A.47)}
$$

The argument of the Lerch zeta function is exactly one, which is the inflection point between the behaviours, when

$$
a = \zeta^{1/\beta} \qquad \text{(A.48)}
$$

We must thus split the integration domain $\mathcal{D}_l$ in three regions (where $b = (\kappa_0/\kappa_s)^{2/\beta}$): $\mathcal{D}_X = [a, \pi] \times [a, \theta']$, $\mathcal{D}_Y = [a, \pi] \times [0, a]$ and $\mathcal{D}_Z = [b, a] \times [0, \theta']$ as depicted in Supplementary Figure A.1. Now, the grey region is the one where the Lerch zeta function argument is bigger than one, in the hatched region we can bound the Lerch zeta function away and the black region is $\mathcal{D}_s$ and we thus do not care about it for the moment. Before going any further, let us note that

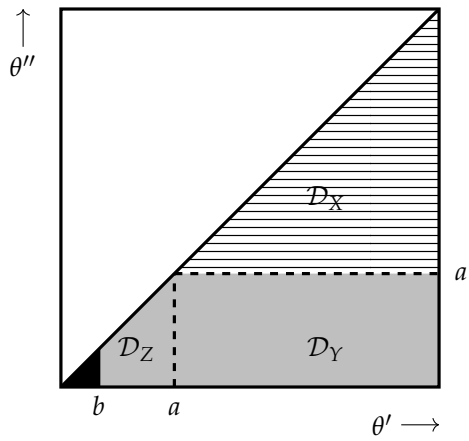



Figure A.1: Integration regions. In the grey region ($\mathcal{D}_Y + \mathcal{D}_Z$) the argument of the Lerch zeta function is bigger than one, in the hatched region ($\mathcal{D}_X$) it is not and the black region is $\mathcal{D}_s$.

Supplementary Figure A.1 looks slightly different for different $\kappa_c$ and $\zeta$. If $\kappa_c \gg \kappa_s$ and $\zeta = (\kappa_c/\kappa_s)^2$, then $\zeta \gg 1$ and thus so is $a$. However, $a$ as an integration limit must be smaller than $\pi$ and thus in this case the $\mathcal{D}_X$ and $\mathcal{D}_Y$ regions disappear. When $\kappa_c \leq \kappa_s$ this is not the case as for all $\zeta$, $a < \pi$. Finally, irrespective of the value of $\kappa_c$, for $\zeta = (\kappa_0/\kappa_s)^2$, $a = b$ and thus region $\mathcal{D}_Z$ vanishes. Implementing the transformation given by Eq. (A.45) in the grey region one obtains

$$I_\xi \leq \iint \mathrm{d}\theta'\mathrm{d}\theta''(\theta'(\theta'-\theta''))^{-\beta}\left\{\theta''^{\beta(2-\gamma)}\left[\psi(\gamma)\ln\left(\frac{\zeta}{\theta''^{\beta}}\right)+\vartheta(\gamma)\right]+\zeta^{2-\gamma}(\gamma-2)^{-2}\right\}. \tag{A.49}$$

As this leads to three different angular integrals, in the end we have seven different integrals to solve.

$$\begin{aligned}\iint_{\mathcal{D}_X}\mathrm{d}\theta'\mathrm{d}\theta''\left(\frac{1}{\theta'\theta''(\theta'-\theta'')}\right)^{\beta} &= \frac{a^{2-3\beta}}{3\beta-2}\Big\{B_1[2\beta-1,1-\beta]-B_1[1-\beta,1-\beta]\\ &+B_{\frac{a}{\pi}}[2\beta-1,1-\beta]+(a/\pi)^{3\beta-2}B_{\frac{a}{\pi}}[1-\beta,1-\beta]\Big\}\\ &+\frac{4^{\beta-1/2}\pi^{5/2-3\beta}\Gamma[1-\beta]}{\Gamma[3/2-\beta](3\beta-2)}\left((a/\pi)^{2-3\beta}-1\right) \qquad \text{(A.50)}\\ &=\boxed{c_{X_{11}}a^{2-3\beta}+c_{X_{12}}} \qquad \text{(A.51)}\end{aligned}$$

$$\begin{aligned}\iint_{\mathcal{D}_Y}\mathrm{d}\theta'\mathrm{d}\theta''\left(\frac{1}{\theta'(\theta'-\theta'')}\right)^{\beta} &= \frac{a^{2-2\beta}}{2(\beta-1)^2}\Big\{2(\beta-1)B_{\frac{a}{\pi}}[2\beta-1,1-\beta]\\ &-\pi^{-1/2}(\beta-1)\Gamma[1-\beta]\Gamma[\beta-1/2]-1\\ &+(1-{}_2F_1\left[\begin{matrix}2(\beta-1),\beta\\2\beta-1\end{matrix};a/\beta\right])(a/\pi)^{2\beta-2}\Big\} \qquad \text{(A.52)}\end{aligned}$$

$$\simeq \boxed{c_{Y_{11}} a^{2-2\beta} + c_{Y_{12}}} \tag{A.53}$$

$$\begin{aligned}\iint_{\mathcal{D}_Y} \mathrm{d}\theta' \mathrm{d}\theta'' \left(\frac{\theta''^{2-\gamma}}{\theta'(\theta'-\theta'')}\right)^{\beta} &= \frac{a^{2-\gamma\beta}}{\gamma\beta-2}\Big\{B_1[1+2\beta-\gamma\beta,1-\beta] \\ &- B_1[2\beta-1,1-\beta] + B_{\frac{a}{\pi}}[2\beta-1,1-\beta] \\ &- (a/\pi)^{\gamma\beta-2} B_{\frac{a}{\pi}}[1+2\beta-\gamma\beta,1-\beta]\Big\} \end{aligned} \tag{A.54}$$

$$\simeq \boxed{c_{Y_{21}} a^{2-\gamma\beta} + c_{Y_{22}} a^{1+2\beta-\gamma\beta}} \tag{A.55}$$

$$\begin{aligned}\iint_{\mathcal{D}_Y} \mathrm{d}\theta' \mathrm{d}\theta'' \left(\frac{\theta''^{2-\gamma}}{\theta'(\theta'-\theta'')}\right)^{\beta} \ln\left(\frac{\zeta}{\theta''^{\beta}}\right) &= \frac{\beta a^{2-\beta\gamma}\pi^{1-2\beta}}{(\beta(\gamma-2)-1)(\beta\gamma-2)^2} \\ &\times \Bigg\{\frac{4^{\beta-1}}{\pi^{\frac{3}{2}-2\beta}}(\beta(\gamma-2)-1)\Gamma[1-\beta]\Gamma\left[\beta-\frac{1}{2}\right] \\ &+ \frac{\pi^{2\beta-1}\Gamma[1-\beta]\Gamma[-\gamma\beta+2\beta+2]}{\Gamma[-\gamma\beta+\beta+2]} \\ &\times \left(\frac{1+(\gamma\beta-2)(H_{\beta(2-\gamma)}-H_{1+\beta-\gamma\beta})}{\beta(\gamma-2)-1}\right) \\ &- a^{2\beta-1}\Bigg(\frac{(\beta(\gamma-2)-1)}{2\beta-1}\,{}_2F_1\left[\begin{matrix}\beta, 2\beta-1 \\ 2\beta\end{matrix}; \frac{a}{\pi}\right] \\ &+ \frac{(\beta\gamma-2)}{\beta(\gamma-2)-1}\,{}_3F_2\left[\begin{matrix}\beta, -\gamma\beta+2\beta+1, -\gamma\beta+2 \\ \beta+1-\gamma\beta+2\beta+2, -\gamma\beta+2\beta+2\end{matrix}; \frac{a}{\pi}\right] \\ &+ {}_2F_1\left[\begin{matrix}\beta, \beta(-\gamma)+2\beta+1 \\ \beta(-\gamma)+2\beta+2\end{matrix}; \frac{a}{\pi}\right]\Bigg)\Bigg\} \end{aligned} \tag{A.56}$$

$$\simeq \boxed{c_{Y_{31}} a^{2-\gamma\beta} + c_{Y_{32}} a^{1+2\beta-\gamma\beta}} \tag{A.57}$$

$$\iint_{\mathcal{D}_Z} \mathrm{d}\theta' \mathrm{d}\theta'' \left(\frac{1}{\theta'(\theta'-\theta'')}\right)^{\beta} = \frac{a^{2-2\beta}-b^{2-2\beta}}{2(\beta-1)^2} = \boxed{c_{Z_{11}} a^{2-2\beta} + c_{Z_{12}} b^{2-2\beta}} \tag{A.58}$$

$$\iint_{\mathcal{D}_Z} \mathrm{d}\theta' \mathrm{d}\theta'' \left(\frac{\theta''^{2-\gamma}}{\theta'(\theta'-\theta'')}\right)^{\beta} = \frac{\Gamma[1-\beta]\Gamma[-\gamma\beta+2\beta+1]\left(b^{2-\beta\gamma}-a^{2-\beta\gamma}\right)}{(\beta\gamma-2)\Gamma[-\gamma\beta+\beta+2]} \tag{A.59}$$

$$= \boxed{c_{Z_{21}} a^{2-\gamma\beta} + c_{Z_{22}} b^{2-\gamma\beta}} \tag{A.60}$$

$$\iint_{\mathcal{D}_Z} \mathrm{d}\theta' \mathrm{d}\theta'' \left(\frac{\theta''^{2-\gamma}}{\theta'(\theta'-\theta'')}\right)^{\beta} \ln\left(\frac{\zeta}{\theta''^{\beta}}\right) = \frac{\Gamma[1+2\beta-\gamma\beta]\Gamma[1-\beta]\beta}{(\beta\gamma-2)\Gamma[2+\beta-\gamma\beta]}\left(a^{2-\gamma\beta}-b^{2-\gamma\beta}\right)$$

$$\times \left\{ H_{\beta(2-\gamma)} - H_{1+\beta-\gamma\beta} + \frac{1}{\gamma\beta-2} - \frac{1}{\beta}\ln(\zeta) \right.$$
$$\left. + \frac{a^{2-\gamma\beta}\ln a - b^{2-\gamma\beta}\ln b}{a^{2-\gamma\beta} - b^{2-\gamma\beta}} \right\} \tag{A.61}$$

$$= \boxed{\begin{array}{l} a^{2-\gamma\beta}\left(c_{Z_{31}} + c_{Z_{32}}\left(\ln(a) - \frac{1}{\beta}\ln(\zeta)\right)\right) \\ -b^{2-\gamma\beta}\left(c_{Z_{31}} + c_{Z_{32}}\left(\ln(b) - \frac{1}{\beta}\ln(\zeta)\right)\right) \end{array}} \tag{A.62}$$

The next step is to organise the different scalings (see Tab. (A.1), where we have defined $c_{Y_i} = c_{Y_{1i}} + c_{Y_{2i}} + c_{Y_{3i}}$ and similarly for $Z$) that were found and find which is dominant.

Let us note that the terms containing $\ln(\kappa_c/\kappa_0)$ cancel as the final results (Eq. (A.43)) contains $I_{\kappa_c^2/\kappa_s^2} - 2I_{\kappa_0\kappa_c/\kappa_s^2}$. We now have many different scaling behaviors, and the question of which one dominates again depends on the value of $\beta$ as well as $\kappa_c$. As a matter of fact, if one includes the $\kappa_s^{-2}$ prefactor in Eq. (A.43), one recovers the same behavior as was found for the lower bound

$$I_- \leq \begin{cases} C_{-,1}N^{-2/\beta+\gamma-1} + C_{-,2}N^{-1}\ln N & \text{if} \quad \kappa_c \gg \kappa_s \\ N^{-1}\left(C_{-,3}N^{\gamma-2/\beta} + C_{-,4}N^{(1-\alpha)(\gamma-2/\beta)} + C_{-,5}N^{(1-\alpha)(\gamma-3)}\right) & \text{if} \quad \kappa_c \leq \kappa_s \end{cases}, \tag{A.63}$$

where $C_{-,i}$ are constants.

This seems to go in the right direction. However, we have not explored the full integration domain yet. It turns out though that the integration domain $\mathcal{D}_s$ does not lead to any new scaling:

$$\begin{aligned} I_{-,s} &= \iint \mathrm{d}x\mathrm{d}y(xy)^{-\gamma} \iint_{\mathcal{D}_s} \frac{1}{1+\frac{\theta'^{\beta}\kappa_s}{\kappa x}} \frac{1}{1+\frac{\theta''^{\beta}\kappa_s}{\kappa y}} \frac{1}{1+\frac{(\theta'-\theta'')^{\beta}}{xy}} \\ &\leq \iint \mathrm{d}x\mathrm{d}y(xy)^{-\gamma} \iint_{\mathcal{D}_s} 1 \\ &= \left(\frac{\kappa_0}{\kappa_s}\right)^{4/\beta} \iint \mathrm{d}x\mathrm{d}y(xy)^{-\gamma} \\ &= \left(\frac{\kappa_0}{\kappa_s}\right)^{4/\beta} \frac{1}{(1-\gamma)^2}\left(\left(\frac{\kappa_c}{\kappa_s}\right)^{1-\gamma} - \left(\frac{\kappa_0}{\kappa_s}\right)^{1-\gamma}\right)^2 \\ &\simeq \frac{1}{(1-\gamma)^2}\left(\frac{\kappa_0}{\kappa_s}\right)^{2(1-\gamma+2/\beta)} \\ &\sim N^{-1+\gamma-2/\beta}. \end{aligned} \tag{A.64}$$

The contribution of $\mathcal{D}_s$ is thus subleading for small $\beta$ and equally dominant as the other contributions for large $\beta$. We have thus shown that for the the upper and lower bound the dominant scaling is the same. We now have the scaling of all distinct parts, $B, A_+, A-$, so we can now combine them all.

$$\bar{c} \simeq \begin{cases} C_1N^{2-2/\beta} + C_2N^{2-\gamma}\ln N & \text{if } \kappa_c \gg \kappa_s \\ C_3N^{2-2/\beta} + C_4N^{2-2/\beta-\alpha(\gamma-2/\beta)} + C_5N^{-1-\alpha(\gamma-3)} & \text{if } \kappa_c \leq \kappa_s \end{cases} \tag{A.65}$$

Table A.1: A summary of all terms contributing to (A.43). This table should be read as follows: We know that $A_{-,l} \leq (\kappa/\kappa_s)^2(I_{(\kappa_c/\kappa_s)^2} + I_{(\kappa_0/\kappa_s)^2} - 2I_{(\kappa_0\kappa_c/\kappa_s)^2})$. Each column gives the contributions of the different integration regions $X$, $Y$ and $Z$ to one of the three $I_\zeta$ in this equation. Note that for $\zeta = (\kappa_c/\kappa_s)^2$ the result depends on which of the two cut-offs is larger.

| $\zeta$ | $X$ | $Y$ | $Z$ |
|---|---|---|---|
| $\left(\frac{\kappa_0}{\kappa_s}\right)^2$ | $c_{X_{11}}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + c_{X_{12}}\left(\frac{\kappa_0}{\kappa_s}\right)^{2(3-\gamma)}$ | $c_{Y_1}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + c_{Y_2}\left(\frac{\kappa_0}{\kappa_s}\right)^{4+\frac{2}{\beta}-2\gamma}$ | |
| $\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)$ | $c_{X_{11}}\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{\frac{2}{\beta}-\gamma} + c_{X_{12}}\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{3-\gamma}$ | $c_{Y_1}\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{\frac{2}{\beta}-\gamma} + c_{Y_2}\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{2+1/\beta-\gamma}$ | $c_{Z1}\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{\frac{2}{\beta}-\gamma} + (c_{Z_{22}} - c_{Z_{31}})\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + \frac{c_{Z_{32}}}{\beta}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma}\ln\left(\frac{\kappa_c}{\kappa_0}\right) + c_{Z_{12}}\left(\frac{\kappa_0\kappa_c}{\kappa_s^2}\right)^{2-\gamma}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-4}$ |
| $\left(\frac{\kappa_c}{\kappa_s}\right)^2 \gg 1$ | | | $c_{Z_{11}}\pi^{2-2\beta}\left(\frac{\kappa_c}{\kappa_s}\right)^{2(2-\gamma)} + (c_{Z_{21}} + c_{Z_{31}})\,\pi^{2-\gamma\beta} + (c_{Z_{22}} - c_{Z_{32}})\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} - \frac{2}{\beta}c_{Z_{32}}\pi^{2-\gamma\beta}\ln\left(\frac{\kappa_c}{\kappa_s}\right) + \frac{2}{\beta}c_{Z_{32}}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma}\ln\left(\frac{\kappa_c}{\kappa_0}\right) + c_{Z_{23}}\pi^{2-\gamma\beta}\ln(\pi) + c_{Z_{12}}\left(\frac{\kappa_c}{\kappa_s}\right)^{2(2-\gamma)}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-4}$ |
| $\left(\frac{\kappa_c}{\kappa_s}\right)^2 \ll 1$ | $c_{X_{11}}\left(\frac{\kappa_c}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + c_{X_{12}}\left(\frac{\kappa_c}{\kappa_s}\right)^{2(3-\gamma)}$ | $c_{Y_1}\left(\frac{\kappa_c}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + c_{Y_2}\left(\frac{\kappa_c}{\kappa_s}\right)^{4+2/\beta-2\gamma}$ | $c_{Z1}\left(\frac{\kappa_c}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + (c_{Z_{22}} - c_{Z_{31}})\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma} + \frac{2c_{Z_{32}}}{\beta}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-2\gamma}\ln\left(\frac{\kappa_c}{\kappa_0}\right) + c_{Z_{12}}\left(\frac{\kappa_c}{\kappa_s}\right)^{4-2\gamma}\left(\frac{\kappa_0}{\kappa_s}\right)^{\frac{4}{\beta}-4}$ |

Let us discuss the limiting cases of $\alpha$. When $\alpha = 0$, $\kappa_0 \sim \kappa_c$, and the network thus has a homogeneous degree distribution. Then, $C \simeq (C_3 + C_4)N^{2-2/\beta} + C_5 N^{-1}$. If $\alpha = 1$, i.e. $\kappa_c \sim \kappa_s$, the scaling becomes $C \simeq C_3 N^{2-2/\beta} + (C_4 + C_5)N^{2-\gamma}$.

### A.2.3 *Case* $\beta = 1$

We now turn to the limit $\beta = 1$. The general practice of finding upper and lower bounds for the various relevant integrals will be again pursued here, and in many cases the integrals examined will be similar to the ones studied above. However, there are some important differences that force us to treat this case separately. For one, we know that in the case of $\beta = 1$, $\mu$ scales as $\hat{\mu} \sim (\ln N)^{-1}$ instead of $\hat{\mu} \sim N^{1-\beta}$, and thus $\kappa_s \sim \sqrt{N \ln N}$, which of course alters scaling. We will represent all integrals evaluated at $\beta = 1$ by a tilde ($\tilde{A}_-, \tilde{A}_+, \tilde{B}$). We start with $\tilde{B}$:

$$\begin{aligned}\tilde{B} &= \pi \left(\frac{\kappa_c}{\kappa_s}\right)^{1-\gamma} \int_0^1 \mathrm{d}\theta \int_{\kappa_0/\kappa_c}^1 \mathrm{d}x \frac{x^{-\gamma}}{1 + \dfrac{\pi\theta\kappa_s^2}{\kappa_c \kappa x}} \\ &= \pi \left(\frac{\kappa_c}{\kappa_s}\right)^{1-\gamma} \left\{ \frac{\kappa\kappa_c}{\pi\kappa_s^2} \frac{\ln\left(1 + \frac{\pi\kappa_s^2}{\kappa_c\kappa}\right)}{2-\gamma} + \frac{1}{\gamma - 2}\left(\frac{\kappa_0}{\kappa_c}\right)^{2-\gamma} \frac{\kappa\kappa_c}{\pi\kappa_s^2} \ln\left(1 + \frac{\pi\kappa_s^2}{\kappa\kappa_0}\right) \right. \\ &\left. + \frac{1}{(\gamma-2)(\gamma-1)} \left( {}_2F_1\left[\begin{matrix} 1, \gamma-1 \\ \gamma \end{matrix}; -\frac{\pi\kappa_s^2}{\kappa\kappa_c}\right] - \left(\frac{\kappa_0}{\kappa_c}\right)^{1-\gamma} {}_2F_1\left[\begin{matrix} 1, \gamma-1 \\ \gamma \end{matrix}; -\frac{\pi\kappa_s^2}{\kappa\kappa_0}\right] \right) \right\}. \end{aligned} \tag{A.66}$$

The second term is dominant and thus $\tilde{B}$ scales as

$$\tilde{B} \sim \kappa_s^{\gamma-3} \ln(\kappa_s) \sim N^{\frac{\gamma-3}{2}} (\ln N)^{\frac{\gamma-1}{2}}. \tag{A.67}$$

For the lower bound of the numerator of the clustering coefficient we can use the result found in Eq. (A.42) as nowhere was it assumed that $\beta < 1$. Irrespective of $\kappa_c$ this gives us

$$\tilde{A}_+ \leq \tilde{c}_+ N^{\gamma-3} (\ln N)^{\gamma-3}. \tag{A.68}$$

For the upper bound of $\tilde{A}_-$ we cannot follow the same path as was done in the case of general $\beta$. This is because the upper bound employed, given by Eq. (A.43), diverges in the $\beta = 1$ limit. Thus, we must find a stricter bound. This is done by once again dividing the angular integration domain, this time in four pieces: $\mathcal{D}_s = [0, (\kappa_0/\kappa_s)^2] \times [0, \theta']$, $\mathcal{D}_2 = [(\kappa_0/\kappa_s)^2, \pi] \times [0, (\kappa_0/\kappa_s)^2]$, $\mathcal{D}_3 = [(\kappa_0/\kappa_s)^2, \pi] \times [\theta' - (\kappa_0/\kappa_s)^2, \theta']$ and $\mathcal{D}_3 = [2(\kappa_0/\kappa_s)^2, \pi] \times [(\kappa_0/\kappa_s)^2, \theta' - (\kappa_0/\kappa_s)^2]$, as represented in Supplementary Figure A.2. Note that regions $\mathcal{D}_2$ and $\mathcal{D}_3$ overlap, but that is not a problem as our integrand is positive and counting a region double just increases the value of the integral, which in turn work for our purposes as we are only looking for an upper bound. For the region $\mathcal{D}_s$ we can use the result (A.64):

$$\tilde{A}_{-,s} \leq \tilde{c}_{-,s} N^{\gamma-3} (\ln N)^{\gamma-3}. \tag{A.69}$$

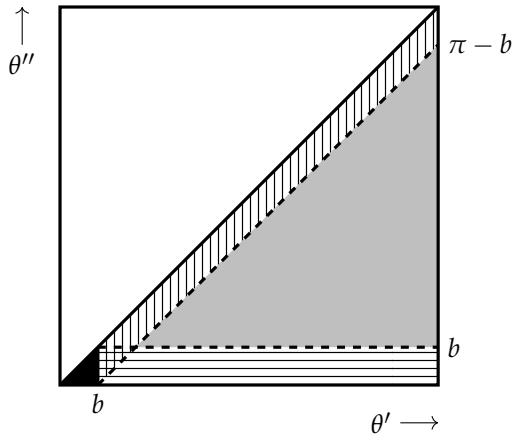



Figure A.2: Integration regions where $b = \frac{\kappa_0^2}{\kappa_s^2}$ The black region is region $\mathcal{D}_s$. The horizontally striped region is region $\mathcal{D}_2$. The vertically striped region is region $\mathcal{D}_3$. The grey region is region $\mathcal{D}_4$.

Turning to $\mathcal{D}_2$ we obtain

$$\begin{aligned}\tilde{A}_{-,2} &= \iint \mathrm{d}x\mathrm{d}y(xy)^{-\gamma} \iint_{\mathcal{D}_2} \frac{\mathrm{d}\theta'\mathrm{d}\theta''}{1+\frac{\theta'\kappa_s}{\kappa x}} \frac{1}{1+\frac{\theta''\kappa_s}{\kappa y}} \frac{1}{1+\frac{\theta'-\theta''}{xy}} \\ &\leq \frac{\kappa}{\kappa_s} \iint_{\mathcal{D}_2} \frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'} \int_{\kappa_s/\kappa_c}^{\kappa_s/\kappa_0} \mathrm{d}x \int_{\kappa_s/\kappa_c}^{\kappa_s/\kappa_0} \mathrm{d}y \frac{x^{\gamma-3}y^{\gamma-2}}{1+xy(\theta'-\theta'')} \end{aligned} \tag{A.70}$$

where we have bounded the integral by decreasing the size of the denominators of the first and second terms. We also performed a change of variables of $x$ and $y$. We now extend the lower bounds of the $x$ and $y$ integrals to zero, which can be done as our integral is positive, and so the resulting integral will be larger or equal to the original one.

$$\begin{aligned}\tilde{A}_{-,2} &\leq \frac{\kappa}{\kappa_s} \iint_{\mathcal{D}_2} \frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'} \int_0^{\kappa_s/\kappa_0} \mathrm{d}x \int_0^{\kappa_s/\kappa_0} \mathrm{d}y \frac{x^{\gamma-3}y^{\gamma-2}}{1+xy(\theta'-\theta'')} \\ &= \frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-3} \iint_{\mathcal{D}_2} \frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'} \left( \Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''),1,\gamma-2\right] - \Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''),1,\gamma-1\right] \right) \end{aligned} \tag{A.71}$$

We know again have the situation that depending on the values of the angular coordinates, the arguments of the $\Phi$'s diverge or go to zero. For the region $\mathcal{D}_{2s} = [b, 2b] \times [0, b]$, $\theta' - \theta'' \in [0, b]$, so the argument lies between zero and one. For the region $\mathcal{D}_{2l} = [2b, \pi] \times [0, b]$, $\theta' - \theta'' \in [b, \pi]$, so the argument is larger than one. We first turn to the second region. Here the argument can diverge and we should thus perform a similar transformation as Eq. (A.45). It is not exactly the

same as the second argument of the $\Phi$'s is now 1 and not two 2, but the derivation is equivalent. This leads us to

$$\begin{aligned}
&\frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-3}\iint_{\mathcal{D}_{2l}}\frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'}\quad\Bigg(\quad\Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''),1,\gamma-2\right]-\Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''),1,\gamma-1\right]\Bigg)\\
&=\frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-3}\iint_{\mathcal{D}_{2l}}\frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'}\Bigg(\quad\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{2-\gamma}\left(\Phi[-1,1,3-\gamma]+\Phi[-1,1,2-\gamma]\right)\\
&\qquad\qquad-\quad\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{-1}\Phi\left[-\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{-1},1,3-\gamma\right]\\
&\qquad\qquad+\quad\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{1-\gamma}\left(\Phi[-1,1,2-\gamma]+\Phi[-1,1,1-\gamma]\right)\\
&\qquad\qquad-\quad\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{-1}\Phi\left[-\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{-1},1,2-\gamma\right]\Bigg)\\
&\leq\frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-3}\iint_{\mathcal{D}_{2l}}\frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'}\Bigg(\quad\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{2-\gamma}\left(\Phi[-1,1,3-\gamma]+\Phi[-1,1,2-\gamma]\right)\\
&\qquad\qquad+\quad\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{1-\gamma}\left(\Phi[-1,1,2-\gamma]+\Phi[-1,1,1-\gamma]\right)\\
&\qquad\qquad-2\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)^{-1}\Bigg)\sim\kappa_s^{2(\gamma-3)}\sim N^{\gamma-3}\left(\ln N\right)^{\gamma-3}.
\end{aligned}\tag{A.72}$$

For $\mathcal{D}_{2s}$ we can immediately bound away the $\Phi$ to find

$$\begin{aligned}
&\frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-3}\iint_{\mathcal{D}_{2s}}\frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'}\left(\Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''),1,\gamma-2\right]-\Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''),1,\gamma-1\right]\right)\\
&\leq\frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-3}\iint_{\mathcal{D}_{2s}}\frac{\mathrm{d}\theta'\mathrm{d}\theta''}{\theta'}=\frac{\kappa}{\kappa_s}(\kappa_s/\kappa_0)^{2\gamma-5}\ln 2\sim N^{\gamma-3}\left(\ln N\right)^{\gamma-3}.
\end{aligned}\tag{A.73}$$

Combining the two results we find that $\tilde{A}_{-,2}\leq\tilde{c}_{-,2}N^{\gamma-3}\left(\ln N\right)^{\gamma-3}$ as expected.

Then we investigate to $\mathcal{D}_3$:

$$\begin{aligned}
\tilde{A}_{-,3}&=\iint\mathrm{d}x\mathrm{d}y(xy)^{-\gamma}\iint_{\mathcal{D}_3}\mathrm{d}\theta'\mathrm{d}\theta''\frac{1}{1+\dfrac{\theta'\kappa_s}{\kappa x}}\frac{1}{1+\dfrac{\theta''\kappa_s}{\kappa y}}\frac{1}{1+\dfrac{\theta'-\theta''}{xy}}\\
&=\iint\mathrm{d}x\mathrm{d}y(xy)^{-\gamma}\iint_{\mathcal{D}_2}\mathrm{d}\theta'\mathrm{d}\theta''\frac{1}{1+\dfrac{\theta'\kappa_s}{\kappa x}}\frac{1}{1+\dfrac{(\theta'-\theta'')\kappa_s}{\kappa y}}\frac{1}{1+\dfrac{\theta''}{xy}}\\
&\leq\left(\frac{\kappa}{\kappa_s}\right)^2\iint\mathrm{d}x\mathrm{d}yx^{1-\gamma}y^{1-\gamma}\iint_{\mathcal{D}_2}\mathrm{d}\theta'\mathrm{d}\theta''\frac{1}{\theta'}\frac{1}{\theta'-\theta''}\\
&=\frac{1}{(2-\gamma)^2}\left(\frac{\kappa}{\kappa_s}\right)^2\left(\frac{\kappa_0}{\kappa_s}\right)^{2(2-\gamma)}\left(\frac{\pi^2}{6}-\mathrm{Li}_2\left[\frac{\kappa_0^2}{\kappa_s^2\pi}\right]\right)\\
&\sim\kappa_s^{2(\gamma-3)}\sim N^{\gamma-3}\left(\ln N\right)^{\gamma-3}.
\end{aligned}\tag{A.74}$$

Here $\mathrm{Li}_2(z)$ is the dilogarithm. The final region to be studied is $\mathcal{D}_4$:

$$\begin{aligned}
\tilde{A}_{-,4} &= \iint \mathrm{d}x\mathrm{d}y(xy)^{-\gamma} \iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{1+\frac{\theta'\kappa_s}{\kappa x}} \frac{1}{1+\frac{\theta''\kappa_s}{\kappa y}} \frac{1}{1+\frac{\theta'-\theta''}{xy}} \\
&\leq \left(\frac{\kappa}{\kappa_s}\right)^2 \iint \mathrm{d}x\mathrm{d}y(xy)^{1-\gamma} \iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''} \frac{1}{1+\frac{\theta'-\theta''}{xy}} \\
&= \left(\frac{\kappa}{\kappa_s}\right)^2 \int_{\kappa_s/\kappa_c}^{\kappa_s/\kappa_0} \mathrm{d}x \int_{\kappa_s/\kappa_c}^{\kappa_s/\kappa_0} \mathrm{d}y(xy)^{\gamma-3} \iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''} \frac{1}{1+xy(\theta'-\theta'')} \\
&\leq \left(\frac{\kappa}{\kappa_s}\right)^2 \int_{0}^{\kappa_s/\kappa_0} \mathrm{d}x \int_{0}^{\kappa_s/\kappa_0} \mathrm{d}y(xy)^{\gamma-3} \iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''} \frac{1}{1+xy(\theta'-\theta'')} \\
&= \left(\frac{\kappa}{\kappa_s}\right)^2 \left(\frac{\kappa_s}{\kappa_0}\right)^{2(\gamma-2)} \iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''} \Phi\left[-\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta''), 2, \gamma-2\right] \\
&\leq \left(\frac{\kappa}{\kappa_s}\right)^2 \iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''} \left\{(\theta'-\theta'')^{2-\gamma}\left[\Psi(.)\ln\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right)+\vartheta(.)\right]\right. \\
&\left. + \left(\frac{\kappa_s}{\kappa_0}\right)^{2(\gamma-3)} (\theta'-\theta'')^{-1}(3-\gamma)^{-2}\right\}. \qquad \text{(A.75)}
\end{aligned}$$

Let us investigate the term with the logarithm first.

$$\begin{aligned}
&\left(\frac{\kappa}{\kappa_s}\right)^2 \int_{2\kappa_0^2/\kappa_s^2}^{\pi} \mathrm{d}\theta' \int_{\kappa_0^2/\kappa_s^2}^{\theta'-\kappa_0^2/\kappa_s^2} \mathrm{d}\theta'' \frac{(\theta'-\theta'')^{2-\gamma}}{\theta'\theta''} \ln\left(\frac{\kappa_s^2}{\kappa_0^2}(\theta'-\theta'')\right) \\
&= \left(\frac{\kappa}{\kappa_s}\right)^2 \left(\frac{\kappa_0}{\kappa_s}\right)^{2(2-\gamma)} \int_{2}^{\pi\kappa_s^2/\kappa_0^2} \mathrm{d}\theta' \int_{1}^{\theta'-1} \mathrm{d}\theta'' \frac{(\theta'-\theta'')^{2-\gamma}}{\theta'\theta''} \ln\left(\theta'-\theta''\right) \\
&= \left(\frac{\kappa}{\kappa_s}\right)^2 \left(\frac{\kappa_0}{\kappa_s}\right)^{2(2-\gamma)} \int_{2}^{\pi\kappa_s^2/\kappa_0^2} \mathrm{d}\theta' \int_{1}^{\theta'-1} \mathrm{d}\theta'' \frac{(\theta'')^{2-\gamma}}{\theta'(\theta'-\theta'')} \ln\left(\theta''\right). \qquad \text{(A.76)}
\end{aligned}$$

This can then be evaluated. The $\theta''$ integral leads to a variety of different terms, which need to be treated separately. Some variable transformations need to be performed, and some special functions need to be expanded to their series representation. It can be shown that the integral to leading order is constant in $N$, implying that the logarithm term of $\tilde{A}_{-,4}$ scales as $\kappa_s^{2(\gamma-3)}$. The other two terms in expression (A.75) are easier to evaluate:

$$\begin{aligned}
\iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''}(\theta'-\theta'')^{2-\gamma} &= \frac{\left(b^{2-\gamma}+\pi^{2-\gamma}\right)}{\gamma-2}\left\{B_{1-\frac{b}{\pi}}[3-\gamma,\gamma-2]-B_{\frac{1}{2}}[3-\gamma,\gamma-2]\right\} \\
&\quad + \frac{b^{2-\gamma}\ln\left(2-\frac{2b}{\pi}\right)}{\gamma-2} \sim \left(\frac{\kappa_0}{\kappa_s}\right)^{2(2-\gamma)} \qquad \text{(A.77)} \\
\iint_{\mathcal{D}_4} \mathrm{d}\theta'\mathrm{d}\theta'' \frac{1}{\theta'\theta''}(\theta'-\theta'')^{-1} &= \frac{2\ln\left(2-\frac{2b}{\pi}\right)}{b} - \frac{2\ln\left(\frac{\pi}{b}-1\right)}{\pi} \sim \frac{\kappa_s^2}{\kappa_0^2}. \qquad \text{(A.78)}
\end{aligned}$$

Plugging this back in we find that also the integral over the region $\mathcal{D}_4$ scales as $N^{\gamma-3}(\ln N)^{\gamma-3}$.

Thus, we can finally conclude that for $\beta = 1$, the clustering coefficient must scale as

$$\bar{c} \sim \frac{N^{\gamma-3}(\ln N)^{\gamma-3}}{N^{\gamma-3}(\ln N)^{\gamma-1}} = (\ln N)^{-2}. \tag{A.79}$$

With this we have found the critical exponent $\eta/\nu = 2$.

### A.2.4 *Exponent $\eta$*

In this section we show that the scaling exponent $\eta$ that encodes how the clustering approaches zero when $\beta \to \beta_c^+ = 1$. As this only requires working on the low temperature side of the transition, we can directly work in the thermodynamic limit (we thus take first the limit $N \to \infty$ and then $\beta \to 1$). To this end, we denote the general definition of the clustering coefficient with hidden degree $\kappa$ and (without loss of generality) spacial coordinate $r = 0$

$$\bar{c}(\kappa) = \frac{\int_{\kappa_0}^{\infty} \mathrm{d}\kappa' \int_{\kappa_0}^{\infty} \mathrm{d}\kappa'' \int_{-\infty}^{\infty} \mathrm{d}r' \int_{-\infty}^{\infty} \mathrm{d}r'' \rho(\kappa')\rho(\kappa'')p(\kappa,\kappa',|r'|)p(\kappa,\kappa'',|r''|)p(\kappa',\kappa'',|r'-r''|)}{\left(\int_{\kappa_0}^{\infty} \mathrm{d}\kappa' \int_{-\infty}^{\infty} \mathrm{d}r' \rho(\kappa')p(\kappa,\kappa',|r'|)\right)^2}. \tag{A.80}$$

where we can use connection probability (4.26) and $\hat{\mu}$ (4.20).

Let us first turn to the denominator:

$$\int \mathrm{d}\kappa' \rho(\kappa') \int_{-\infty}^{\infty} \frac{\mathrm{d}r'}{1 + \left(\frac{r'}{\kappa\kappa'\hat{\mu}}\right)^{\beta}} = \kappa, \tag{A.81}$$

where we have plugged in the definition of $\hat{\mu}$ and used that $\langle k \rangle = \frac{\gamma-1}{\gamma-2}\kappa_0$.

The next step is the numerator. We first perform the transformation $t = r'/(\kappa\kappa'\hat{\mu})$ and $\tau = r''/(\kappa\kappa''\hat{\mu})$ to obtain

$$\bar{c}(\kappa) = \frac{\hat{\mu}^2}{4}(\gamma-1)^2\kappa_0^{2\gamma-2} \iiint \frac{\mathrm{d}\kappa'\mathrm{d}\kappa''\mathrm{d}t\mathrm{d}\tau}{1+|t|^{\beta}} \frac{(\kappa'\kappa'')^{1-\gamma}}{1+|\tau|^{\beta}} \frac{1}{1+\left|\frac{\kappa t}{\kappa''} - \frac{\kappa\tau}{\kappa'}\right|^{\beta}}. \tag{A.82}$$

We know that $\hat{\mu}^2 \sim (\beta-1)^2$. This is exactly the scaling that we expect from numerical investigation for the clustering coefficient. Thus, all we need to prove is that at $\beta = 1$, the numerator is finite. If so, its $(\beta-1)$ dependence must be order $\mathcal{O}(1)$. If the full expression contained $(\beta-1)^{-n}$ terms with $n > 0$ it would diverge at the critical point and if the dominant term was $\mathcal{O}((\beta-1)^n)$ with $n > 0$ the numerator would go to zero at the critical point. And indeed, numerical

integration shows that at $\beta = 1$ the numerator is finite, leading to the conclusion that

$$\bar{c}(\kappa) \sim (\beta - 1)^2 \tag{A.83}$$

such that $\eta = 2$, which in turn implies that $\nu = 1$.

# B

# APPENDIX TO CHAPTER 7

## B.1 SELF SIMILARITY OF THE DEGREE DISTRIBUTION IN THE $\mathbb{S}^1$

### B.1.1 *The hidden degree distribution*

The goal of this section is to find the degree distribution at the $r$'th level of renormalization. We start by studying the hidden degree distribution, assuming that in the original network the distribution is given by

$$\rho(\kappa) = \mathcal{N}\kappa^{-\gamma}, \qquad \kappa_0 \leq \kappa \leq \kappa_c, \tag{B.1}$$

where $\mathcal{N}$ is the normalization constant. To obtain the distribution after renormalization we use Eq. (7.5). Note that we change our method slightly from this point onward. Instead of looking at the $l$'th layer of the iterative normalization procedure where each supernode is constructed with $r$ nodes, we now study only a single normalization step. Note however that, due to the semi-group property, these two approaches are equivalent as $l$ steps of size $r$ can always be replaced by a single step of size $r^l$. We first find the distribution $\tilde{\rho}(\tilde{\kappa})$, where $\tilde{\kappa} = \kappa^{\max(1,\beta)}$:

$$\tilde{\rho}(\tilde{\kappa}) = \tilde{\mathcal{N}}\tilde{\kappa}^{-\eta}, \qquad \tilde{\kappa}_0 \leq \tilde{\kappa} \leq \tilde{\kappa}_c, \tag{B.2}$$

where we have defined $\tilde{\mathcal{N}} = \mathcal{N}/\max(1,\beta)$, $\tilde{\kappa}_0 = \kappa_0^{\max(1,\beta)}$, $\tilde{\kappa}_c = \kappa_c^{\max(1,\beta)}$ and $\eta = 1 + (\gamma - 1)/\max(1,\beta)$. The next step is to find the distribution $\tilde{\rho}_r(\tilde{z})$ where $\tilde{z} = \sum_{i=1}^{r} \tilde{\kappa}_i$. We first state the result and follow with the proof:

$$\tilde{\rho}_r(\tilde{z}) = \sum_{n=1}^{r} \sum_{q=1}^{\infty} c_{n,q} \tilde{z}^{n(1-\eta)-q} \mathbf{1}_{[r\tilde{\kappa}_0, \tilde{\kappa}_c + (r-1)\tilde{\kappa}_0]}(\tilde{z}), \tag{B.3}$$

where $c_{n,l}$ are constants and $\mathbf{1}_{[a,b]}(z)$ denotes the indicator function, which evaluates to 1 when $z \in [a,b]$ and 0 otherwise. To obtain the distribution of the hidden degrees $z$ in the renormalized layer, we use the fact that $z = \tilde{z}^{1/\max(1,\beta)}$, which leads to

$$\rho_r(z) = \max(1,\beta)\tilde{\rho}_r(z^{\max(1,\beta)}) z^{\max(1,\beta)-1}. \tag{B.4}$$

Note that for $\tilde{z} \gg 1$, the dominant scaling in Eq.(B.3) is $\tilde{z}^{-\eta}$ ($n = 1, q = 1$). Plugging this into Eq.(B.4) proves that the distribution $\rho_r(z)$ scales as $z^{-\gamma}$, which in turn demonstrates the self-similarity of the scaling behavior of the hidden degree

distribution under renormalization. Note that the cut-off in the renormalized layer is given by $(\tilde{\kappa}_c + (r-1)\tilde{\kappa}_0)^{1/\max(1,\beta)}$, which is approximately $\kappa_c$ if $\tilde{\kappa}_c \gg (r-1)\tilde{\kappa}_0$.

We now prove Eq. (B.3) using induction. First, for $r = 2$, we know that the distribution $\tilde{\rho}_2(\tilde{z})$, where $\tilde{z} = \tilde{\kappa}_1 + \tilde{\kappa}_2$, is given by the convolution

$$\tilde{\rho}_2(\tilde{z}) = \int_{-\infty}^{\infty} \mathrm{d}\tilde{\kappa}\tilde{\rho}(\tilde{z}-\tilde{\kappa})\tilde{\rho}(\tilde{\kappa}). \tag{B.5}$$

Taking into account the support of $\tilde{\rho}(\tilde{\kappa})$ we can conclude that $\tilde{\kappa}_0 \leq \tilde{z} - \tilde{\kappa} \leq \tilde{\kappa}_c$ and $\tilde{\kappa}_0 \leq \tilde{\kappa} \leq \tilde{\kappa}_c$. We then rewrite Eq. B.5 as

$$\begin{aligned}\tilde{\rho}_2(\tilde{z}) = \frac{\tilde{\mathcal{N}}^2}{\tilde{z}^{2\gamma-1}}\Bigg[&\left(B_{1-\frac{\tilde{\kappa}_0}{\tilde{z}}}\begin{bmatrix}1-\eta\\1-\eta\end{bmatrix} - B_{\frac{\tilde{\kappa}_0}{\tilde{z}}}\begin{bmatrix}1-\eta\\1-\eta\end{bmatrix}\right)1_{[2\tilde{\kappa}_0,\tilde{\kappa}_c+\tilde{\kappa}_0]}(\tilde{z})\\ &+\left(B_{\frac{\tilde{\kappa}_c}{\tilde{z}}}\begin{bmatrix}1-\eta\\1-\eta\end{bmatrix} - B_{1-\frac{\tilde{\kappa}_c}{\tilde{z}}}\begin{bmatrix}1-\eta\\1-\eta\end{bmatrix}\right)1_{[\tilde{\kappa}_0+\tilde{\kappa}_c,2\tilde{\kappa}_c]}(\tilde{z})\Bigg].\end{aligned} \tag{B.6}$$

Here, $B_a\begin{bmatrix}b\\c\end{bmatrix}$ represents the incomplete beta function. We then note that this function can be expanded as

$$\begin{aligned}B_{1-x}\begin{bmatrix}a\\b\end{bmatrix} = \frac{\pi\csc(b\pi)}{a}&\left(\sum_{n=0}^{\infty}\frac{a_{(n)}}{n!}(-x)^n\right)\times\left(\frac{\Gamma(1+a)}{\Gamma(a+b)}\sum_{q=0}^{\infty}\left[\frac{(b-1)_{(q)}(-a)_{(q)}}{q!\Gamma(1-b+q)}x^l\right]\right.\\ &\left.-\frac{ax^b}{\Gamma(1-b)}\sum_{q=0}^{\infty}\left[\frac{(-a-b)_{(q)}}{q!\Gamma(1+b+q)}(-x)^q\right]\right)\end{aligned} \tag{B.7}$$

$$B_x\begin{bmatrix}a\\b\end{bmatrix} = x^a\sum_{n=0}^{\infty}\frac{(1-b)_{(n)}}{n!(a+n)}x^n \tag{B.8}$$

when $x \to 0$, where the $y_{(n)}$ represent the falling factorials: $y_{(n)} = y(y-1)(y-2)...(y-n+1)$. In the case that $\tilde{z} \in [2\tilde{\kappa}_0, \tilde{\kappa}_c + \tilde{\kappa}_0]$, $\tilde{z}/\tilde{\kappa}_0 \ll 1$ in the tail of the distribution. Thus, we can apply the expansions given above and show that the dominant scaling in this regime is $\tilde{\rho}_2(\tilde{z}) \sim \tilde{z}^{-\eta}$ and that the full behavior is given by Eq. (B.3). Crossing over to the regime $\tilde{z} \in [\tilde{\kappa}_0 + \tilde{\kappa}_c, 2\tilde{\kappa}_c]$, we get that $1 - \tilde{\kappa}_c/\tilde{z} \ll 1$, as least close to the transition. Using once again the series expansions of the beta functions we obtain that $\tilde{\rho}_2(z) \sim (1-\tilde{\kappa}_c/\tilde{z})^{1-\eta}$. This falls of hyperbolically and so we can take the probability density to be zero here. Therefore, we prove Eq. (B.3) for $r = 2$.

Now, assuming that Eq. (B.3) is true for some general $r$, let us investigate the case for $r+1$. In this case, we start with the convolution

$$\begin{aligned}\tilde{\rho}_{r+1}(\tilde{z}) = &\int_{r\tilde{\kappa}_0}^{\tilde{z}-\tilde{\kappa}_0}\mathrm{d}\tilde{\kappa}\rho_1(\tilde{z}-\tilde{\kappa})\rho_r(\tilde{\kappa})1_{[(r+1)\tilde{\kappa}_0,\tilde{\kappa}_c+r\tilde{\kappa}_0]}(\tilde{z})\\ &+\int_{\tilde{z}-\tilde{\kappa}_c}^{\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0}\mathrm{d}\tilde{\kappa}\tilde{\rho}_1(\tilde{z}-\tilde{\kappa})\tilde{\rho}_r(\tilde{\kappa})1_{[\tilde{\kappa}_c+r\tilde{\kappa}_0,2\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0]}(\tilde{z}),\end{aligned} \tag{B.9}$$

where have taken into account the respective domains of the two functions $\tilde{\rho}_1$ and $\tilde{\rho}_r$. The fact that $\tilde{\rho}_r(\tilde{z})$ can be expanded into a sum of terms $\tilde{g}_r(\tilde{z};\alpha) \sim \tilde{z}^{-\alpha}$, where $\alpha \geq \eta$, implies that $\tilde{\rho}_{r+1}(\tilde{z})$ can be expanded into a sum of integrals $\tilde{I}(\tilde{z};\alpha)$ evaluating to

$$\tilde{I}(\tilde{z};\alpha) = \tilde{\mathcal{N}}^r \tilde{z}^{1-\eta-\alpha} \left[ \left( B_{1-\frac{\tilde{\kappa}_0}{\tilde{z}}} \begin{bmatrix} 1-\alpha \\ 1-\eta \end{bmatrix} - B_{\frac{r\tilde{\kappa}_0}{\tilde{z}}} \begin{bmatrix} 1-\alpha \\ 1-\eta \end{bmatrix} \right) 1_{[(r+1)\tilde{\kappa}_0, \tilde{\kappa}_c + r\tilde{\kappa}_0]} \right.$$
$$\left. + \left( B_{\frac{\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0}{\tilde{z}}} \begin{bmatrix} 1-\alpha \\ 1-\eta \end{bmatrix} - B_{1-\frac{\tilde{\kappa}_c}{\tilde{z}}} \begin{bmatrix} 1-\alpha \\ 1-\eta \end{bmatrix} \right) 1_{[\tilde{\kappa}_c + r\tilde{\kappa}_0, 2\tilde{\kappa}_c + (r-1)\tilde{\kappa}_0]} \right]. \tag{B.10}$$

Using the same arguments as before, we can show that $\forall \alpha$ the integral falls off hyperbolically in the second region. When $\tilde{z} \in [(r+1)\tilde{\kappa}_0, \tilde{\kappa}_c + r\tilde{\kappa}_0]$, it can be shown that the expression can be rewritten in the form of Eq. (B.3), where the dominant scaling for large $\tilde{z}$ is once again $\sim \tilde{z}^{-\eta}$. With this we conclude the proof.

Note that this proof is contingent on some assumptions, most notably that $r\tilde{\kappa}_0 \ll \tilde{\kappa}_c$. Of course, for finite $\tilde{\kappa}_c$, there is always an $r$ for which this assumption breaks down. This has to do with the central limit theorem: For a finite cut-off $\tilde{\kappa}_c$, the variance of the distribution $\tilde{\rho}(\tilde{\kappa})$ is also finite, and thus the distribution $\tilde{\rho}_r(\tilde{z})$ necessarily approaches a Gaussian as $r \to \infty$. In the case of the model we in general assume that $\tilde{\kappa}_c = \tilde{\kappa}_0 N^{\max(1,\beta)/(\gamma-1)}$, which is very large for the network sizes we typically work with, and so one can perform several renormalization steps before one 'feels' the effect of the cut-off.

### B.1.2 *The degree distribution*

In Sec. 3.3.4 it was noted that the degree distribution is related to the distribution of hidden degrees by

$$P_r(k) = \frac{1}{k!} \int \mathrm{d}z \rho(z) \bar{k}(z)^k e^{-\bar{k}(z)}, \tag{B.11}$$

where $\bar{k}(\kappa)$ is the expected degree of a node with hidden degree $\kappa$ [141]. In the unrenormalized layer one can show that $\bar{k}(\kappa) = \kappa$ when $\hat{\mu}$ is chosen correctly. For this to be true for in the renormalized layer, however, one would need that $\langle \kappa_r \rangle = \langle k_r \rangle$, which is not generally the case as the scaling exponents determining the flow of these two quantities, $\xi$ and $\nu$, are not always equal. Using Eq. (7.9) and $\xi = (\nu+1)/2$, one obtains that

$$\overline{k_r}(\kappa_r) = r^{(\nu-1)/2} \kappa_r \tag{B.12}$$

We now note that we do not know the exact functional form of $\rho_r(\kappa)$, at least not for $\beta > 1$. To be able to plug in Eq. (B.3), we first need to transform (B.11). It can be shown that this integral is equivalent to

$$P_r(k) = \frac{1}{k!} \int_{r\tilde{\kappa}_0}^{\tilde{\kappa}_c + (r-1)\tilde{\kappa}_0} \mathrm{d}\tilde{z} \frac{r^{k(\nu-1)/2} \tilde{\rho}_r(\tilde{z}) \tilde{z}^{\frac{k}{\max(1,\beta)}}}{\exp\left(r^{(\nu-1)/2} \tilde{z}^{\frac{1}{\max(1,\beta)}}\right)}. \tag{B.13}$$

Then, combining the previous result with Eq. (B.3) and Eq. (B.11) one obtains

$$P_r(k) = \sum_{n=1}^{r} \sum_{q=1}^{\infty} \frac{c_{n,q} r^{k\frac{\nu-1}{2}}}{k!} \int_{r\tilde{\kappa}_0}^{\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0} \mathrm{d}\tilde{z} \frac{\tilde{z}^{n(1-\eta)-q+k/\max(1,\beta)}}{\exp\left(r^{\frac{\nu-1}{2}}\tilde{z}^{1/\max(1,\beta)}\right)}$$

$$= \sum_{n=1}^{r} \sum_{q=1}^{\infty} \frac{c_{n,l}\max(1,\beta) r^{\frac{\alpha(\nu-1)}{2}}}{k!} \Bigg[ \Gamma\left(\alpha+k, r^{\frac{\nu-1}{2}+1/\max(1,\beta)}\kappa_0\right)$$

$$+ \Gamma\left(\alpha+k, r^{\frac{\nu-1}{2}}\left(\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0\right)^{1/\max(1,\beta)}\right)\Bigg], \tag{B.14}$$

where $\alpha = \max(1,\beta)(1-q)+n(1-\gamma)$. When $k \gg (\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0)^{1/\max(1,\beta)}$, the two gamma functions cancel, meaning that the probability density vanishes. When $r^{1/\max(1,\beta)}\kappa_0 \ll k \leq (\tilde{\kappa}_c+(r-1)\tilde{\kappa}_0)^{1/\max(1,\beta)}$, the first term scales as $k^{-\gamma}$, whereas the second term falls off exponentially. This implies that the scaling behavior of the tail of the distribution is preserved under renormalization. Note that once again for large $\kappa_c$ the cut-off does not evolve under renormalization.

# C

# APPENDIX TO CHAPTER 8

## C.1 EMPIRICAL NETWORK DESCRIPTION

In this section we give an overview of the networks studied in Chapter 8 (Tab. 8.1).

- **Foodweb–Eocene** [234]: A reconstructed food web of an ecosystem from the early Eocene (48 million years ago). Nodes represent taxa and edges represent consumer-resource relations. The original network was directed.
- **Foodweb–Wetland** [330]: A network of carbon exchanges among species in the cypress wetlands of South Florida. Nodes represent taxa and edges represent consumer-resource relations. The original network was directed.
- **WordAdjacency–English** [331]: A network of word adjacency in English texts. Nodes represent words and two words are connected if one directly follows the other in texts. The original network was directed.
- **WordAdjacency–Japanese** [331]: A network of word adjacency in Japanese texts. Nodes represent words and two words are connected if one directly follows the other in texts. The original network was directed.
- **MB–R.norvegicus** [332]: A metabolic network of the rat (Ratus norvegicus), extracted from the Kyoto Encyclopedia of Genes and Genomes (KEGG). Nodes represent substances involved in enzymatic reactions and edges represent reactant-product pairs.
- **WikiTalk–Catalan** [333]: A network where nodes represents Wikipedia editors for a certain language (in this case Catalan), and where user $i$ and $j$ are connected if $i$ leaves a message on the talk page of $j$. The original network was directed.
- **GI–S.cerevisiae** [334]: A network based on the Molecular Interaction Search Tool (MIST) for baker's yeast (Saccharomyces cerevisiae). Here node represent genes and the edges indicate that the effects of mutations in one gene can be modified by mutations of another gene.
- **GMP–C.elegans** [335]: A multiplex network representing different types of genetic interactions for the nematode worm Caenorhabditis elegans. The

layers represent physical, association, co-localization, direct, suppressive and additive interactions. In this chapter we create a monolayer network by treating the different interaction types equally and removing repeated links. The original network was directed.

- **Gnutella** [336]: A snapshot of the Gnutella peer-to-peer file sharing network on August 4th 2002. Nodes are hosts and edges are connections between them. The original network was directed.

- **PPI–S.cerevisiae** [334]: A network based on the Molecular Interaction Search Tool (MIST) for baker's yeast (Saccharomyces cerevisiae). Here node represent genes and the edges indicate that there are physical interactions between their associated proteins.

- **PPI–D.melanogaster** [334]: A network based on the Molecular Interaction Search Tool (MIST) for the fruit fly (Drosophila melanogaster). Here node represent genes and the edges indicate that there are physical interactions between their associated proteins.

- **Transport–London** [337]: An multiplex network of the public transportation system in London. Nodes are London train stations and the links can represent either the underground, overground and DLR connections. There connections are treated equally as to create a mono-layer network.

- **GMP–S.cerevisiae** [335]: A multiplex network representing different types of genetic interactions for baker's yeast (Saccharomyces cerevisiae). The layers represent physical, association, co-localization, direct, suppressive and additive interactions. In this chapter we create a monolayer network by treating the different interaction types equally and removing repeated links. The original network was directed.

- **Internet-PoP** [338]: The Kentucky Datalink network, an internet graph at the Point of Presence (PoP) level. Nodes are physical network interface points and links physical connections between them.

- **PPI–H.sapiens** [334]: A network based on the Molecular Interaction Search Tool (MIST) for humans (Homo sapiens). Here node represent genes and the edges indicate that there are physical interactions between their associated proteins.

- **WikiVote** [339]: The network represents the voting process used to select Wikipedia administrators, which are contributors with access to additional technical features. Nodes represents Wikipedia users and an edge is created if user $i$ votes on the selections of user $j$. The original network was directed.

- **MathOverflow** [340]: An interaction network of users (nodes) on the online Q&A site MathOverflow. An edge from node $i$ to node $j$ indicates that $i$ responded to an answer by $j$. The original network was directed.

# D

# APPENDIX TO CHAPTER 9

## D.1 STATISTICAL PROPERTIES OF THE MUTUAL NETWORK ENSEMBLE

In this section we derive the scaling behavior of the results presented in Sec. 9.3. We start by noting that, because we are working with perfect correlations between the layers, we can work with the density of states derived in Sec. 5.1. Conveniently, as we are working in the homogeneous regime, this quantity takes the form

$$\rho(\epsilon) = \frac{2}{N} e^{\epsilon} \Theta(\epsilon_{\max} - \epsilon), \tag{D.1}$$

where $\epsilon_{\max} = \ln(N/2)$ and where we have used the definition $\epsilon = \ln(d_{ij})$ as presented in Sec. 4.2.1. Note that this also requires us to use the definitions Eq. 4.14 and 4.13 for $\mu$.

In the following, we assume that $\beta_1 \leq \beta_2$ and that $\mu_1 \leq \mu_2$. The opposite case, where $\beta_1 \leq \beta_2$ but $\mu_1 > \mu_2$ leads to qualitatively similar results.

### D.1.1 *The amount of particles*

Starting from Eq. (9.13) and plugging in $\widetilde{p}'_{ij}$ as defined in Eq. (9.21) for $L = 2$ we obtain

$$\begin{aligned}\langle \widetilde{M} \rangle &\simeq N \int_{-\infty}^{\mu_1} d\epsilon e^{\epsilon} + N \int_{\mu_1}^{\mu_2} d\epsilon e^{\epsilon} e^{-\beta_1(\epsilon-\mu_1)} + N \int_{\mu_2}^{\epsilon_{\max}} d\epsilon e^{\epsilon} e^{-\beta_1(\epsilon-\mu_1)} e^{-\beta_2(\epsilon-\mu_2)} \\ &= N\left( \frac{\beta_1}{\beta_1 - 1} e^{\mu_1} + \frac{\beta_2}{(1-\beta_1)(\tilde{\beta}-1)} e^{(1-\beta_2)\mu_2 + \beta_1\mu_1} + \frac{1}{1-\tilde{\beta}} e^{(1-\tilde{\beta})\epsilon_{\max} + \beta_1\mu_1 + \beta_2\mu_2} \right).\end{aligned} \tag{D.2}$$

### D.1.2 *Entropy*

We now start from Eq. (9.22) and plug in $\widetilde{p}'_{ij}$ and $\widetilde{q}'_{ij}$ for $L = 2$. This leads to

$$\begin{aligned}\widetilde{S} \simeq & -N\int_{-\infty}^{\mu_1} d\epsilon e^{\epsilon} e^{\beta_1(\epsilon-\mu_1)}\beta_1(\epsilon-\mu_1) + N\int_{\mu_1}^{\mu_2} d\epsilon e^{\epsilon} e^{-\beta_1(\epsilon-\mu_1)}\beta_1(\epsilon-\mu_1) \\ & + N\int_{\mu_2}^{\epsilon_{\max}} d\epsilon e^{\epsilon} e^{-\beta_1(\epsilon-\mu_1)} e^{-\beta_2(\epsilon-\mu_2)}\left(\beta_1(\epsilon-\mu_1)+\beta_2(\epsilon-\mu_2)\right) \\ & = N\left(\xi_1 e^{\mu_1} + \xi_2 e^{\mu_2(1-\beta_1)+\beta_1\mu_1} + \xi_3 e^{(1-\beta_1-\beta_2)\epsilon_{\max}+\beta_1\mu_1+\beta_2\mu_2}\right), \end{aligned} \tag{D.3}$$

where $\xi_1, \xi_2$ and $\xi_3$ depend at most linearly on $\mu_1$, $\mu_2$ and $\epsilon_{max}$, which in turn all scale at most logarithmically with $N$. Any power law scaling must therefore result from the exponential contributions to the three terms. However, we see that these are all the same as in Eq. (D.2), which means that they divide out when looking at the entropy per link $\widetilde{s} = \widetilde{S}/\langle\widetilde{M}\rangle$. Indeed, we obtain

$$\widetilde{s} \simeq \begin{cases} \frac{2(1+\beta_1^2)}{(1+\beta_1)^2}\frac{1}{\beta_1-1} & \text{if } 1 \le \beta_1 \le \beta_2 \\ \beta_1(\mu_2-\mu_1) + \frac{\beta_1\tilde{\beta}-1}{(1-\beta_1)(1-\tilde{\beta})} & \text{if } 1-\beta_2 \le \beta_1 \le 1 \\ \tilde{\beta}\epsilon_{\max} + \beta_1\mu_1 + \beta_2\mu_2 + \frac{\tilde{\beta}}{\tilde{\beta}-1} & \text{if } \beta_1+\beta_2 \le 1 \end{cases} \tag{D.4}$$

In $\beta_1 = \beta_2 \equiv \beta$ and $\mu_1 = \mu_2 \equiv \mu$ this reduces to

$$\widetilde{s} \simeq \begin{cases} \frac{3(1+2\beta^2)}{2(1+\beta)^2}\frac{1}{2\beta-1} & \text{if } 1/2 < \beta \\ 2\beta\left(\epsilon_{max} + \frac{1}{2\beta-1} + \mu\right) & \text{if } \beta < 1/2 \end{cases} \tag{D.5}$$

## D.2 HOMOGENEOUS MULTIPLEXES WITH PERFECT COORDINATE CORRELATIONS

In this section we study the scaling behaviors of the average mutual degree and clustering coefficients for $L$ sparse, homogeneous, perfectly correlated layers with general $\{\beta_l\}_{l=1}^{L}$ and $\{\langle k\rangle^{(l)}\}_{l=1}^{L}$.

### D.2.1 *The average amount of particles*

In the setting detailed above, Eq. (9.13) reduces to

$$\langle\widetilde{M}\rangle = \frac{N^2}{2\pi}\int_0^{\pi} d\theta \prod_{i=1}^{L}\frac{1}{1+(\zeta_i\theta)^{\beta_i}}. \tag{D.6}$$

Without loss of generality, we now assume that $\zeta_1 \ge \zeta_2 \ge \ldots \ge \zeta_L$. Note that we want to have sparse individual layers such that $\zeta_i = N/(2\pi\mu^{(i)}\langle k^{(i)}\rangle^2) \sim$

$N^{\max(1,1/\beta_i)}$. This implies that $\beta_1 \leq \beta_2 \leq ... \leq \beta_L$ when $N \gg 1$. This allows us to split the integral in $L+1$ intervals

$$\langle \widetilde{M} \rangle = \frac{N^2}{2\pi} \sum_{j=1}^{L+1} I_j, \tag{D.7}$$

where

$$I_j = \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta \prod_{i=1}^{L} \frac{1}{1+(\zeta_i\theta)^{\beta_i}} \tag{D.8}$$

and where we have defined $\zeta_0^{-1} = 0$ and $\zeta_{L+1}^{-1} = \pi$. We can now find the upper and lower bound of each $I_j$ separately. We start with the lower bound:

$$\begin{aligned} I_j \geq \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta \min_{\theta} \left( \prod_{i=1}^{L} \frac{1}{1+(\zeta_i\theta)^{\beta_i}} \right) &= \left(\zeta_j^{-1} - \zeta_{j-1}^{-1}\right) \prod_{i=1}^{L} \frac{1}{1+(\zeta_i/\zeta_j)^{\beta_i}} \\ &\simeq \zeta_j^{-1} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-\beta_i}. \end{aligned} \tag{D.9}$$

Here we make use of the fact that $\zeta_{j-1} \geq \zeta_j$ so $\zeta_{j-1}^{-1} \leq \zeta_j^{-1}$. Furthermore, the fraction $\zeta_i/\zeta_j$ is large only for $i < j$, and will thus only then show up in the Taylor expansion. Otherwise the fraction can just be approximated as 1. For the upper bound we will use the fact that $1 \geq 1/(1+x)$ and $1 \geq 1/x$. Now, we want to make as tight a bound as possible, so when $1 \geq 1/x$ we take $1/x$ and when $1 \leq 1/x$ we take 1. We thus only have a contribution to the integrand when $x \geq 1$. In our case this implies $\zeta_i\theta \geq 1$. This is the case for the entire integration range only when $i \leq j-1$. Thus, we can write the upper bound as

$$I_j \leq \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta \prod_{i=1}^{j-1} (\zeta_i\theta)^{-\beta_i} = \frac{\zeta_j^{-1} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-\beta_i} - \zeta_{j-1}^{-1} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_{j-1}}\right)^{-\beta_i}}{1 - \sum_{i=1}^{j-1} \beta_i}. \tag{D.10}$$

The product in the second term in the numerator can be taken instead to $j-2$ because the $j-1$ term just evaluates to one. Then, one sees that both terms are equivalent (ignoring prefactors) if for the second term one perform $j-1 \to j$. Thus, when summing over $I_j$, the second contribution will already come from the $I_{j-1}$ integral. We are left with

$$\sum_j \left( \zeta_j^{-1} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-\beta_i} \right) \leq \sum_j I_j \leq \sum_j \left( \zeta_j^{-1} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-\beta_i} \right). \tag{D.11}$$

We have then proven that

$$\langle \widetilde{M} \rangle \sim \frac{N^2}{2\pi} \sum_{j=1}^{L+1} \left( \zeta_j^{-1} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-\beta_i} \right) \sim N^{\max_j(\sigma_j)}. \tag{D.12}$$

Now, the question is to find out for which $\sigma_j$ dominates. Let us ask study $\sigma_j$ and $\sigma_{j-1}$. The difference between these two exponents is given by

$$\sigma_j - \sigma_{j-1} = \left(\max\left(1, \frac{1}{\beta_j}\right) - \max\left(1, \frac{1}{\beta_{j-1}}\right)\right)\left(\sum_{i=1}^{j-1}\beta_i - 1\right). \tag{D.13}$$

Now, we know that $\zeta_{j-1} \geq \zeta_j$, which implies that $\beta_{j-1} \leq \beta_j$. This means that $\left(\max\left(1, \frac{1}{\beta_j}\right) - \max\left(1, \frac{1}{\beta_{j-1}}\right)\right) \leq 0$ and so

$$\sigma_j \geq \sigma_{j-1} \quad \text{if} \quad \sum_{i=1}^{j-1}\beta_i \leq 1. \tag{D.14}$$

Thus, the largest exponent is given by the $j$ that satisfies

$$1 - \beta_j \leq \sum_{i=1}^{j-1}\beta_i \leq 1, \tag{D.15}$$

in which case

$$\sigma_j = 2 + \sum_{i=1}^{j-1}\beta_i\left(\max\left(1, \frac{1}{\beta_j}\right) - \max\left(1, \frac{1}{\beta_i}\right)\right) - \max\left(1, \frac{1}{\beta_j}\right). \tag{D.16}$$

Of course, this is the scaling exponent for the average amount of links. If we want to see how the average mutual degree scales we need to subtract this quantity by 1. When $L = 2$ this leads to

$$\sigma = \begin{cases} 0 & \text{if } 1 < \beta_1 < \beta_2, \\ \beta_1 - 1 & \text{if } 1 - \beta_2 < \beta_1 < 1,\ \beta_2 > 1, \\ (\beta_1 - 1)/\beta_2 & \text{if } 1 - \beta_2 < \beta_1 < 1,\ \beta_2 < 1, \\ -1 & \text{if } \beta_1 + \beta_2 < 1, \end{cases} \tag{D.17}$$

exactly what was found in Fig. 9.2b and equivalent to the results obtained in Sec. D.1.1.

### D.2.2 *Clustering coefficient*

We start from Eq. (9.25), where we define

$$\tilde{\bar{t}} \equiv \int_0^{\pi} d\theta' \int_0^{\theta'} d\theta'' \prod_{i=1}^{L} f(\theta', \theta''), \tag{D.18}$$

with

$$f(\theta', \theta'') = \frac{1}{1 + (\zeta_i\theta')^{\beta_i}} \frac{1}{1 + (\zeta_i\theta'')^{\beta_i}} \frac{1}{1 + (\zeta_i(\theta' - \theta''))^{\beta_i}} \tag{D.19}$$

such that $\bar{\tilde{c}} = \left(\frac{N}{\pi}\right)^2 \bar{\tilde{t}}/\langle\tilde{k}\rangle^2$. We then once again assume that $\zeta_1 \geq \zeta_2 \geq ... \geq \zeta_m$, and split the $\theta'$ prime integral accordingly:

$$\bar{t}^{(m)} = \sum_{j=1}^{L+1} I_j, \tag{D.20}$$

where

$$I_j = \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta' \int_0^{\theta'} d\theta'' \prod_{i=1}^{L} f(\theta', \theta''). \tag{D.21}$$

We can then find the upper and lower bound for each integral using the same logic as before. For the lower bound we have

$$\begin{aligned} I_j &\geq \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta' \int_0^{\theta'} d\theta'' \min_{\theta',\theta''} \left( \prod_{i=1}^{L} f(\theta', \theta'') \right) \\ &= \frac{1}{2}\left(\zeta_j^{-2} - \zeta_{j-1}^{-2}\right) \prod_{i=1}^{L} f(\zeta_j^{-1}, \zeta_j^{-1}/2) \sim \zeta_j^{-2} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-3\beta_i}, \end{aligned} \tag{D.22}$$

where we again use that $\zeta_i/\zeta_j > 1$ if $i < j-1$. For the upper bound we need to take some care. Naively, one starts as follows

$$\begin{aligned} I_j &\leq \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta' \int_0^{\theta'} d\theta'' \prod_{i=1}^{j-1} (\zeta_i^3 \theta' \theta'' (\theta' - \theta''))^{-\beta_i} \\ &= \frac{\Gamma(1-a_j)^2}{(2-3a_j)\Gamma(2-2a_j)} \left( \zeta_j^{-2} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-3\beta_i} - \zeta_{j-1}^{-2} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_{j-1}}\right)^{-3\beta_i} \right), \end{aligned} \tag{D.23}$$

where we have defined $a_j = \sum_{i=1}^{j-1} \beta_i$ and where we again see that the second term is just the first for $j \to j-1$. However, this integral is only defined for $a_j < 1$. We can thus only use this upper bound for $j$ such that $\sum_{i=1}^{j-1} \beta_i < 1$. For higher $j$ we make a looser upper bound

$$I_j \leq \int_{\zeta_{j-1}^{-1}}^{\zeta_j^{-1}} d\theta' \int_0^{\theta'} d\theta'' \prod_{i=1}^{\xi-1} (\zeta_i^3 \theta' \theta'' (\theta' - \theta''))^{-\beta_i}, \tag{D.24}$$

where $\xi$ is chosen such that $\sum_{i=1}^{\xi-1} \beta_i < 1$ and $\sum_{i=1}^{\xi} \beta_i \geq 1$. This integral can then be evaluated as before. However, the upper and lower bound do no longer match for $j > \xi$. This is not a problem as we will later show that other terms in the series will dominate even this upper bound, which means that the terms $j > \xi$ will always be sub-leading. Summarizing, we have found that

$$\sum_{j=1}^{\xi} I_j \sim \sum_{j=1}^{\xi} \zeta_j^{-2} \prod_{i=1}^{j-1} \left(\frac{\zeta_i}{\zeta_j}\right)^{-3\beta_i} = \sum_{j=1}^{\xi} N^{\sigma_j}, \tag{D.25}$$

and

$$I_j \leq \frac{\Gamma(1-a_\xi)^2}{(2-3a_\xi)\Gamma(2-2a_\xi)}\left(\zeta_j^{-2}\prod_{i=1}^{\xi-1}\left(\frac{\zeta_i}{\zeta_j}\right)^{-3\beta_i} - \zeta_{j-1}^{-2}\prod_{i=1}^{\xi-1}\left(\frac{\zeta_i}{\zeta_{j-1}}\right)^{-3\beta_i}\right)$$

$$\sim N^{\max(\tilde{\sigma}_j,\tilde{\sigma}_{j-1})} \quad \text{if } j > \xi. \tag{D.26}$$

First we find which of the $\sigma_j$ with $j \leq \xi$ dominates. We again look at two consecutive exponents:

$$\sigma_j - \sigma_{j-1} = \left(\max\left(1,\frac{1}{\beta_j}\right) - \max\left(1,\frac{1}{\beta_{j-1}}\right)\right)\left(3\sum_{i=1}^{j-1}\beta_i - 2\right). \tag{D.27}$$

Because $\zeta_{j-1} \geq \zeta_j$ we have that $\sigma_j \geq \sigma_{j-1}$ when $\sum_{i=1}^{j-1}\beta_i \leq \frac{2}{3}$. The dominant $\sigma_j$ is given by the $j$ that satisfies

$$\frac{2}{3} - \beta_j \leq \sum_{i=1}^{j-1}\beta_i \leq \frac{2}{3}, \tag{D.28}$$

in which case the scaling is given by

$$\sigma_j = 3\sum_{i=1}^{j-1}\beta_i\left(\max\left(1,\frac{1}{\beta_j}\right) - \max\left(1,\frac{1}{\beta_i}\right)\right) - 2\max\left(1,\frac{1}{\beta_j}\right). \tag{D.29}$$

For $j > \xi$ we have

$$\tilde{\sigma}_j - \tilde{\sigma}_{j-1} = \left(\max\left(1,\frac{1}{\beta_j}\right) - \max\left(1,\frac{1}{\beta_{j-1}}\right)\right)\left(3\sum_{i=1}^{\xi-1}\beta_i - 2\right). \tag{D.30}$$

Again, $\tilde{\sigma}_j \geq \tilde{\sigma}_{j-1}$ when $\sum_{i=1}^{\xi-1}\beta_i \leq \frac{2}{3}$. By definition of $\xi$ we also know that $\sum_{i=1}^{\xi}\beta_i \geq 1$. This can only be true if $j = \xi$, which does not satisfy $j > \xi$. Thus, we know that $\tilde{\sigma}_{j-1} > \tilde{\sigma}_j$ for all $j > \xi$. The largest exponent is then given by the term $I_{\xi+1} \sim N^{\tilde{\sigma}_\xi}$. However, $\tilde{\sigma}_\xi = \sigma_\xi$ which is the scaling of $I_\xi$ for which we know the exact scaling, not just an upper bound. The upper bounds on $I_j$ for $j > \xi$, and therefore also the integrals themselves, are therefore all sub-leading to $I_\xi$.

We can thus conclude that Eq. (D.29) is the scaling of $\bar{\tilde{t}}$. To obtain the scaling of the average local clustering coefficient $\bar{\tilde{c}} = (N/\pi)^2\bar{\tilde{t}}/\langle\tilde{k}\rangle^2$ we use that the scaling exponents are related as follows

$$\sigma_{i,j}^{(c)} = 2 + \sigma_i^{(t)} - 2\sigma_j^{(k)}. \tag{D.31}$$

Now, $i$ and $j$ are not necessarily the same. However, we can assume that $i \leq j$. This is because the condition given in Eq. (D.15) is weaker than that given in Eq. (D.28). This then allows us to write $\sigma_{i,j}^{(c)}$ as

$$\begin{aligned}\sigma_{i,j}^{(c)} = &\sum_{l=1}^{i-1} \beta_l \left( 3\max\left(1, \frac{1}{\beta_i}\right) - 2\max\left(1, \frac{1}{\beta_j}\right) - \max\left(1, \frac{1}{\beta_l}\right)\right) \\ &+ 2\sum_{l=i}^{j-1} \beta_l \left( \max\left(1, \frac{1}{\beta_l}\right) - \max\left(1, \frac{1}{\beta_j}\right)\right) \\ &- 2\left( \max\left(1, \frac{1}{\beta_i}\right) - \max\left(1, \frac{1}{\beta_j}\right)\right). \end{aligned} \tag{D.32}$$

Note that the second sum starts at $l = i$, not $l = 1$. A certain $\sigma_{i,j}^{(c)}$ dominates when

$$\frac{2}{3} - \beta_i \leq \sum_{l=1}^{i-1} \beta_l \leq \frac{2}{3} \quad \text{and} \quad 1 - \beta_j \leq \sum_{l=1}^{j-1} \beta_l \leq 1. \tag{D.33}$$

In words this is equivalent to saying: The dominant scaling $\sigma_{i,j}^{(c)}$ is given when the sum of the smallest $i-1$ $\beta$'s is smaller than $2/3$ but where adding one more $\beta$ makes the sum larger than $2/3$. Simultaneously, the sum of the smallest $j-1$ $\beta$'s is smaller than 1 but adding the $j$'th $\beta$ to the sum will make it larger than 1.

One notices that if $\beta_1 = \beta_2 = ... = \beta_i = ... = \beta_j$ all $\beta's$ that appear in Eq. (D.32) are equal, which leads to $\sigma_{i,j}^{(c)}$ being zero. Thus, for a multiplex with $L$ layers, we have several manifolds in parameter space where the clustering is constant. These are given by

$$\begin{aligned} &1/2 \leq \beta_1 = \beta_2 < \beta_3 \leq ... \leq \beta_L, \\ &1/3 \leq \beta_1 = \beta_2 = \beta_3 < \beta_4 \leq ... \leq \beta_L, \\ &\vdots \\ &1/L \leq \beta_1 = \beta_2 = ... = \beta_L. \end{aligned} \tag{D.34}$$

## D.3 EMPIRICAL MULTIPLEX DESCRIPTIONS

In this section we give a brief description of the real multiplexes used in Chapter 9.6

- **GMP–Yeast** [335]: A multiplex network representing different types of genetic interactions for Schizosaccharomyces pombe, also known as fission yeast. The layers represent (i) physical, (ii) association, (iii) co-localization, (iv) direct, (v) suppressive and (vi) additive interactions. Here the mutual network is created from the additive (layer 1) and direct (layer 2) interactions.

- **arXiv** [341]: A multiplex coauthership network of all researchers who published preprints on arXiv.org before May 2014. Differt layers represent different categories of publication. In this chapter the mutual network is created from the tags “Computer Science - Social and Information Networks (cs:SI)" for layer 1 and “Physics - Data Analysis, Statistics and Probability (physics.data-an)" for layer 2.
- **Trade** [342]: In this multiplex, nodes represent countries and edges different types of trade relations, where the data is obtained from the FAO (Food and Agriculture Organization of the United Nations). Here we focus on crude materials (layer 1) and food preparation necessities (layer 2).
- **Twitter** [343]: A twitter data set from the 2013 World Championships in Athletics held in Moscow. Here, nodes are users and edges represent retweets, mentions and replies. Here we focus on mentions (layer 1) and retweets (layer 2).
- **GMP–Human** [335]: A multiplex network representing different types of genetic interactions for humans. The layers represent (i) physical, (ii) association, (iii) co-localization, (iv) direct, (v) suppressive and (vi) additive interactions. Here the mutual network is created from the association (layer 1) and physical (layer 2) interactions.

# E

## APPENDIX TO CHAPTER 10

### E.1 EMPIRICAL NETWORK DESCRIPTIONS AND PROPERTIES

In the following section we elaborate on the properties of the real networks considered in Chapter 10.

- **CElegans-G** [344]: A genetic interaction network for the nematode Caenorhabditis elegans where the nodes represent genes and the links interactions on the genetic level.
- **FriendsOFF** [245]: An off-line friendship network among students, created from a survey where each student was asked to list five male and five female friends. The nodes are in this case the students and the edges represent these friendship relations. The original network was directed.
- **Human-C** [322]: A connectome of the Human brain where nodes represent small brain regions and the edges represent fibers connecting these regions based on white matter tractography, which is a method used to estimate axonal bundle trajectories. In this network both hemispheres are incorporated.
- **Malaria-G** [345]: The network of recombinant antigen genes from the human Malaria parasite Plasmodium falciparum, where nodes are var genes encoding for an antigen protein expressed on the surface of the infected red blood cell. These var genes are capable of interchanging bits of genetic information, thus creating a vast amount of slightly different genes that in turn create highly varying antigens that are hard for the human immune system to detect. Two nodes are connected if they share a substring of nucleotides of significant length, indicating that an interchange of genetic material between the two nodes has taken place.
- **Mouse-C** [346]: A connectome of the mouse brain where nodes represent anatomical subregions of the brain and the edges the physical connections between them. The original network was directed.
- **WTW-2013** [53, 347, 348]: The world trade web as of 2013, where nodes represent countries and edges trade relations.

Tab. E.1 shows the network properties of these networks.

Table E.1: Properties of the six real networks studied in Figs. 10.1, 10.12 and 10.13. The parameter $\hat{r}$ gives the percentage of eigenvectors classified as periodic by the procedure described in Sec.. The notation $^{**}$ implies that the results are significant with $p = 0.01$.

| Network | $N$ | $\langle k \rangle$ | $\beta$ | $k_{\max}$ | $\hat{r}$ |
|---|---|---|---|---|---|
| CElegans-G | 878 | 7.2 | 2.6 | 151 | 5.9%** |
| FriendsOFF | 2539 | 8.2 | 1.3 | 27 | 4.4%** |
| Human-C | 989 | 36.1 | 2.3 | 97 | 7.5%** |
| Malaria-G | 307 | 18.3 | 2.9 | 49 | 3.9%** |
| Mouse-C | 213 | 27.9 | 2.0 | 85 | 2.3%** |
| WTW 2013 | 189 | 5.8 | 1.9 | 110 | 5.8%** |

# BIBLIOGRAPHY


[1] D. C. Krakauer, ed., *Worlds hidden in plain sight* (SFI Press, 2019).

[2] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics **74**, 47–97 (2002).

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Complex networks: structure and dynamics, Physics Reports **424**, 175–308 (2006).

[4] J. Bascompte, Networks in ecology, Basic and Applied Ecology **8**, 485–490 (2007).

[5] J. Dunne, The network structure of food webs, in *Ecological networks: linking structure to dynamics in food webs*, edited by M. Pascual and J. Dunne (Oxford University Press, 2005), pp. 27–86.

[6] D. S. Bassett and O. Sporns, Network neuroscience, Nature Neuroscience **20**, 353–364 (2017).

[7] O. Sporns, Structure and function of complex brain networks, Dialogues in Clinical Neuroscience **15**, 247–262 (2013).

[8] D. Lazer et al., Computational social science, Science **323**, 721–723 (2009).

[9] L. C. Freeman, *The development of social network analysis* (Empirical Press, 2004).

[10] A.-L. Barabási, The physics of the web, Physics World **14**, 33–38 (2001).

[11] M. Nadini, L. Alessandretti, F. D. Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli, Mapping the nft revolution: market trends, trade networks, and visual features, Scientific Reports **11**, 20902 (2021).

[12] M. Buchanan, *Nexus: small worlds and the groundbreaking theory of networks* (W.W. Norton & Company, 2003).

[13] J. M. Kleinberg, Navigation in a small world, Nature **406**, 845–845 (2000).

[14] C. C. Cartozo and P. De Los Rios, Extended navigability of small world networks: exact results and new insights, Physical Review Letters **102**, 238703 (2009).

[15] M. E. J. Newman, Mixing patterns in networks, Physical Review E **67**, 026126 (2003).

[16] M. E. J. Newman, *Networks* (Oxford University Press, 2010).

[17] M. Girvan and M. E. J. Newman, Community structure in social and biological networks, Proc Natl Acad Sci USA **99**, 7821–7826 (2002).

[18] M. Boguñá, I. Bonamassa, M. D. Domenico, S. Havlin, D. Krioukov, and M. Á. Serrano, Network geometry, Nature Reviews Physics **3**, 114–135 (2021).

[19] M. Á. Serrano and M. Boguñá, *The shortest path to network geometry* (Cambridge University Press, 2022).

[20] M. Barthelemy, *Spatial networks* (Springer International Publishing, 2022).

[21] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, The architecture of complex weighted networks, Proceedings of the National Academy of Sciences **101**, 3747–3752 (2004).

[22] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles, Proceedings of the National Academy of Sciences **102**, 7794–7799 (2005).

[23] E. N. Gilbert, Random plane networks, Journal of the Society for Industrial and Applied Mathematics **9**, 533–543 (1961).

[24] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Classes of small-world networks, Proceedings of the National Academy of Sciences **97**, 11149–11152 (2000).

[25] R. Albert, I. Albert, and G. L. Nakarado, Structural vulnerability of the north american power grid, Physical Review E **69**, 025103 (2004).

[26] V. Latora and M. Marchiori, Is the boston subway a small-world network? Physica A: Statistical Mechanics and its Applications **314**, 109–113 (2002).

[27] A. Cardillo, S. Scellato, V. Latora, and S. Porta, Structural properties of planar graphs of urban street patterns, Physical Review E **73**, 066107 (2006).

[28] M. Boguñá, F. Papadopoulos, and D. Krioukov, Sustaining the internet with hyperbolic mapping, Nature Communications **1**, 62 (2010).

[29] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, Physical Review E **82**, 036106 (2010).

[30] M. A. Serrano, D. Krioukov, and M. Boguñá, Self-similarity of complex networks and hidden metric spaces, Phys. Rev. Lett. **100**, 78701 (2008).

[31] M. Boguñá, D. Krioukov, and K. C. Claffy, Navigability of complex networks, Nature Physics **5**, 74–80 (2009).

[32] K. Gugelmann Luca and Panagiotou and U. Peter, Random hyperbolic graphs: degree sequence and clustering, in *Automata, languages, and programming*, edited by A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer (2012), pp. 573–585.

[33] E. Candellero and N. Fountoulakis, Clustering and the hyperbolic geometry of complex networks, Internet Mathematics **12**, 2–53 (2016).

[34] N. Fountoulakis, P. van der Hoorn, T. Müller, and M. Schepers, Clustering in a hyperbolic model of complex networks, Electronic Journal of Probability **26** (2021).

[35] J. van der Kolk, M. Á. Serrano, and M. Boguñá, An anomalous topological phase transition in spatial random graphs, Communications Physics **5**, 245 (2022).

[36] M. A. Abdullah, N. Fountoulakis, and M. Bode, Typical distances in a geometric model for complex networks, Internet Mathematics (2017).

[37] T. Friedrich and A. Krohmer, On the diameter of hyperbolic random graphs, SIAM Journal on Discrete Mathematics **32**, 1314–1334 (2018).

[38] T. Müller and M. Staps, The diameter of KPKVB random graphs, Advances in Applied Probability **51**, 358–377 (2019).

[39] M. Á. Serrano, D. Krioukov, and M. Boguñá, Percolation in self-similar networks, Physical Review Letters **106**, 048701 (2011).

[40] N. Fountoulakis and T. Müller, Law of large numbers for the largest component in a hyperbolic model of complex networks, The Annals of Applied Probability **28** (2018).

[41] M. Kiwi and D. Mitsche, Spectral gap of random hyperbolic graphs and related parameters, The Annals of Applied Probability **28** (2018).

[42] G. García-Pérez, M. Boguñá, and M. Á. Serrano, Multiscale unfurling of real complex networks by geometric renormalisation, Nature Physics **14**, 583–589 (2018).

[43] M. Zheng, G. García-Pérez, M. Boguñá, and M. Á. Serrano, Scaling up real networks by geometric branching growth, Proceedings of the National Academy of Sciences **118** (2021).

[44] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, Popularity versus similarity in growing networks, Nature **489**, 537–540 (2012).

[45] A. Allard, M. Á. Serrano, G. García-Pérez, and M. Boguñá, The geometric nature of weights in real complex networks, Nature Communications **8**, 14103 (2017).

[46] A. Allard, M. Á. Serrano, and M. Boguñá, Geometric description of clustering in directed networks, Nature Physics **20**, 150–156 (2024).

[47] K.-K. Kleineberg, M. Boguñá, M. Á. Serrano, and F. Papadopoulos, Hidden geometric correlations in real multiplex networks, Nature Physics **12**, 1076–1081 (2016).

[48] K.-K. Kleineberg, L. Buzna, F. Papadopoulos, M. Boguñá, and M. Á. Serrano, Geometric correlations mitigate the extreme vulnerability of multiplex networks against targeted attacks, Physical Review Letters **118**, 218301 (2017).

[49] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov, Emergence of soft communities from geometric preferential attachment, Scientific Reports **5**, 9421 (2015).

[50] G. García-Pérez, M. Á. Serrano, and M. Boguñá, Soft communities in similarity space, Journal of Statistical Physics **173**, 775–782 (2018).

[51] A. Muscoloni and C. V. Cannistraci, A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities, New Journal of Physics **20**, 052002 (2018).

[52] M. A. Serrano, M. Boguna, and F. Sagues, Uncovering the hidden geometry behind metabolic networks, Mol. BioSyst. **8**, 843–850 (2012).

[53] G. García-Pérez, M. Boguñá, A. Allard, and M. Á. Serrano, The hidden hyperbolic geometry of international trade: world trade atlas 1870–2013, Scientific Reports **6**, 33441 (2016).

[54] A. Allard and M. Á. Serrano, Navigable maps of structural brain networks across species, PLOS Computational Biology **16**, e1007584 (2020).

[55] A. Gulyás, J. J. Bíró, A. Kőrösi, G. Rétvári, and D. Krioukov, Navigable networks as nash equilibria of navigation games, Nature Communications **6**, 7651 (2015).

[56] M. Boguñá, C. Castellano, and R. Pastor-Satorras, Langevin approach for the dynamics of the contact process on annealed scale-free networks, Physical Review E **79**, 036110 (2009).

[57] Z. Wang, Y. Wu, Q. Li, F. Jin, and W. Xiong, Link prediction based on hyperbolic mapping with community structure for complex networks, Physica A: Statistical Mechanics and its Applications **450**, 609–623 (2016).

[58] M. Kitsak, I. Voitalov, and D. Krioukov, Link prediction with hyperbolic geometry, Physical Review Research **2**, 043113 (2020).

[59] D. Krioukov, Clustering implies geometry in networks, Physical Review Letters **116**, 208302 (2016).

[60] R. Michielan, N. Litvak, and C. Stegehuis, Detecting hyperbolic geometry in networks: why triangles are not enough, Physical Review E **106**, 054303 (2022).

[61] J. van der Kolk, M. Á. Serrano, and M. Boguñá, Random graphs and real networks with weak geometric coupling, Physical Review Research **6**, 013337 (2024).

[62] J. van der Kolk, M. Boguñá, and M. Á. Serrano, Renormalization of networks with weak geometric coupling, Physical Review E **110**, L032302 (2024).

[63] J. van der Kolk, D. Krioukov, M. Boguñá, and M. Á. Serrano, Link overlap and mutual clustering in multiplex networks, In preparation, 2024.

[64] G. García-Pérez, A. Allard, M. Á. Serrano, and M. Boguñá, Mercator: uncovering faithful hyperbolic embeddings of complex networks, New Journal of Physics **21**, 123033 (2019).

[65] J. van der Kolk, G. García-Pérez, N. E. Kouvaris, M. Á. Serrano, and M. Boguñá, Emergence of geometric Turing patterns in complex networks, Physical Review X **13**, 021038 (2023).

[66] R. Albert, H. Jeong, and A.-L. Barabási, Diameter of the World-Wide Web, Nature **401**, 130–131 (1999).

[67] L. Li, D. Alderson, J. C. Doyle, and W. Willinger, Towards a theory of scale-free graphs: definition, properties, and implications, Internet Mathematics **2**, 431–523 (2005).

[68] A.-L. Barabási, *Network science* (Cambridge University Press, 2016).

[69] M. E. J. Newman, Assortative mixing in networks, Physical Review Letters **89**, 208701 (2002).

[70] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, SIAM Review **51**, 661–703 (2009).

[71] R. Pastor-Satorras and A. Vespignani, Epidemic dynamics in finite size scale-free networks, Physical Review E **65**, 035108 (2002).

[72] R. Khanin and E. Wit, How scale-free are biological networks, Journal of Computational Biology **13**, 810–818 (2006).

[73] Á. Corral and Á. González, Power law size distributions in geoscience revisited, Earth and Space Science **6**, 673–697 (2019).

[74] A. D. Broido and A. Clauset, Scale-free networks are rare, Nature Communications **10**, 1017 (2019).

[75] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, Scale-free networks well done, Physical Review Research **1**, 033034 (2019).

[76] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Structure of growing networks with preferential linking, Physical Review Letters **85**, 4633–4636 (2000).

[77] R. Bojanic and E. Seneta, Slowly varying functions and asymptotic relations, Journal of Mathematical Analysis and Applications **34**, 302–315 (1971).

[78] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation* (Cambridge University Press, 1987).

[79] S. I. Resnick, *Heavy-tail phenomena: probabilistic and statistical modeling* (Springer, 2007).

[80] R. A. Fisher and L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, Mathematical Proceedings of the Cambridge Philosophical Society **24**, 180–190 (1928).

[81] B. Gnedenko, Sur la distribution limite du terme maximum d'une serie aleatoire, The Annals of Mathematics **44**, 423 (1943).

[82] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Cut-offs and finite size effects in scale-free networks, The European Physical Journal B - Condensed Matter **38**, 205–209 (2004).

[83] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Dynamical and correlation properties of the internet, Physical Review Letters **87**, 258701 (2001).

[84] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Large-scale topological and dynamical properties of the internet, Physical Review E **65**, 066130 (2002).

[85] S. Maslov, K. Sneppen, and A. Zaliznyak, Detection of topological patterns in complex networks: correlation profile of the internet, Physica A: Statistical Mechanics and its Applications **333**, 529–540 (2004).

[86] S. Fortunato, Community detection in graphs, Physics Reports **486**, 75–174 (2010).

[87] S. Fortunato and D. Hric, Community detection in networks: a user guide, Physics Reports **659**, 1–44 (2016).

[88] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, Physical Review E **69**, 026113 (2004).

[89] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner, On modularity-np-completeness and beyond, ITI Wagner, Faculty of Informatics, Universität Karlsruhe (TH), Tech. Rep **19**, 2006 (2006).

[90] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment **2008**, P10008 (2008).

[91] T. P. Peixoto, *Descriptive vs. inferential community detection in networks* (Cambridge University Press, 2023).

[92] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, Modularity from fluctuations in random graphs and complex networks, Physical Review E **70**, 025101 (2004).

[93] S. Fortunato and M. Barthélemy, Resolution limit in community detection, Proceedings of the National Academy of Sciences **104**, 36–41 (2007).

[94] T. P. Peixoto, Bayesian stochastic blockmodeling, in (Wiley, Nov. 2019), pp. 289–332.

[95] S. Milgram, The small world problem, Psychology today **2**, 60–67 (1967).

[96] J. Travers and S. Milgram, An experimental study of the small world problem, in *Social networks* (Elsevier, 1977), pp. 179–197.

[97] S. Schnettler, A small world on feet of clay? A comparison of empirical small-world studies against best-practice criteria, Social Networks **31**, 179–189 (2009).

[98] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, Four degrees of separation, in *Proceedings of the 4th annual acm web science conference* (2012), pp. 33–42.

[99] P. S. Dodds, R. Muhamad, and D. J. Watts, An experimental study of search in global social networks, Science **301**, 827–829 (2003).

[100] F. Bloch, M. O. Jackson, and P. Tebaldi, Centrality measures in networks, Social Choice and Welfare **61**, 413–453 (2023).

[101] N. Perra and S. Fortunato, Spectral centrality measures in complex networks, Physical Review E **78**, 036107 (2008).

[102] S. Obernesser, Nested ecology: the place of humans in the ecological hierarchy, Interdisciplinary Studies in Literature and Environment **18**, 493–494 (2011).

[103] E. Mones, L. Vicsek, and T. Vicsek, Hierarchy measure for complex networks, PLoS ONE **7**, e33799 (2012).

[104] E. Ortiz, G. García-Pérez, and M. Á. Serrano, Geometric detection of hierarchical backbones in real networks, Physical Review Research **2**, 033519 (2020).

[105] Y.-X. Kong, G.-Y. Shi, R.-J. Wu, and Y.-C. Zhang, k-core: Theories and Applications, Physics Reports **832**, 1–32 (2019).

[106] S. B. Seidman, Network structure and minimum degree, Social Networks **5**, 269–287 (1983).

[107] L. da Fontoura Costa, The hierarchical backbone of complex networks, Physical Review Letters **93**, 098702 (2004).

[108] N. Tatti, Hierarchies in directed networks, in *IEEE international conference on data mining* (2015), pp. 991–996.

[109] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Network motifs: simple building blocks of complex networks, Science **298**, 824–827 (2002).

[110] U. Alon, Network motifs: theory and experimental approaches, Nature Reviews Genetics **8**, 450–461 (2007).

[111] A. Frieze and M. Karoński, *Introduction to random graphs* (Cambridge University Press, 2015).

[112] B. Bollobás, *Random graphs* (Cambridge University Press, 2001).

[113] R. van der Hofstad, *Random graphs and complex networks* (Cambridge University Press, 2024).

[114] B. Bollobás and O. Riordan, Mathematical results on scale-free random graphs, in *Handbook of graphs and networks* (Wiley-VCH, 2003).

[115] P. Erdös and A. Rényi, On random graphs I, Publicationes Mathematicae Debrecen **6**, 290–297 (1959).

[116] J. Park and M. E. J. Newman, Statistical mechanics of networks, Physical Review E **70**, 066117 (2004).

[117] E. N. Gilbert, Random graphs, The Annals of Mathematical Statistics **30**, 1141–1144 (1959).

[118] F. Chung and L. Lu, The diameter of sparse random graphs, Advances in Applied Mathematics **26**, 257–279 (2001).

[119] D. Fernholz and V. Ramachandran, The diameter of sparse random graphs, Random Structures & Algorithms **31**, 482–516 (2007).

[120] A. Barrat and M. Weigt, On the properties of small-world network models, The European Physical Journal B **13**, 547–560 (2000).

[121] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature **393**, 440–442 (1998).

[122] M. E. J. Newman, Models of the small world, Journal of Statistical Physics **101**, 819–841 (2000).

[123] E. A. Bender and E. Canfield, The asymptotic number of labeled graphs with given degree sequences, Journal of Combinatorial Theory, Series A **24**, 296–307 (1978).

[124] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Structures & Algorithms **6**, 161–180 (1995).

[125] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, European Journal of Combinatorics **1**, 311–316 (1980).

[126] J. Blitzstein and P. Diaconis, A sequential importance sampling algorithm for generating random graphs with prescribed degrees, Internet Mathematics **6**, 489–522 (2011).

[127] S. Janson, The probability that a random multigraph is simple, Combinatorics, Probability and Computing **18**, 205–225 (2009).

[128] O. Angel, R. van der Hofstad, and C. Holmgren, Limit laws for self-loops and multiple edges in the configuration model, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **55** (2019).

[129] J. Petersen, Die Theorie der regulären graphs, Acta Mathematica **15**, 193–220 (1891).

[130] B. K. Fosdick, D. B. Larremore, J. Nishimura, and J. Ugander, Configuring random graph models with fixed degree sequences, SIAM Review **60**, 315–355 (2018).

[131] P. Erdős and L. B. Richmond, On graphical partitions, Combinatorica **13**, 57–63 (1993).

[132] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, 3rd ed. (Chapman & Hall, 2013).

[133] P. van der Hoorn, G. Lippner, and D. Krioukov, Sparse maximum-entropy random graphs with a given power-law degree distribution, Journal of Statistical Physics **173**, 806–844 (2018).

[134] F. Chung and L. Lu, Connected components in random graphs with given expected degree sequences, Annals of Combinatorics **6**, 125–145 (2002).

[135] D. Garlaschelli and M. I. Loffredo, Maximum likelihood: extracting unbiased information from complex networks, Physical Review E **78**, 015101 (2008).

[136] T. Squartini and D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks, New Journal of Physics **13**, 083001 (2011).

[137] G. Bianconi, The entropy of randomized network ensembles, Europhysics Letters **81**, 28005 (2008).

[138] S. Janson, Asymptotic equivalence and contiguity of some random graphs, Random Structures & Algorithms **36**, 26–45 (2010).

[139] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, Random Structures & Algorithms **31**, 3–122 (2007).

[140] G. Caldarelli, A. Capocci, P. D. L. Rios, and M. A. Muñoz, Scale-free networks from varying vertex intrinsic fitness, Physical Review Letters **89**, 258702 (2002).

[141] M. Boguñá and R. Pastor-Satorras, Class of correlated random networks with hidden variables, Physical Review E **68**, 036112 (2003).

[142] K. Anand, D. Krioukov, and G. Bianconi, Entropy distribution and condensation in random networks with a given degree distribution, Physical Review E **89**, 062807 (2014).

[143] I. Voitalov, P. van der Hoorn, M. Kitsak, F. Papadopoulos, and D. Krioukov, Weighted hypersoft configuration model, Physical Review Research **2**, 043157 (2020).

[144] P. Colomer-de-Simon and M. Boguñá, Clustering of random scale-free networks, Physical Review E **86**, 026120 (2012).

[145] R. van der Hofstad, A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and C. Stegehuis, Local clustering in scale-free networks with hidden variables, Physical Review E **95**, 022307 (2017).

[146] M. Penrose, *Random geometric graphs* (Oxford University Press, 2003).

[147] Q. Duchemin and Y. D. Castro, Random geometric graph: Some recent developments and perspectives, in *High dimensional probability IX*, edited by R. Adamczak, N. Gozlan, K. Lounici, and M. Madiman (2023), pp. 347–392.

[148] J. Dall and M. Christensen, Random geometric graphs, Physical Review E **66**, 016121 (2002).

[149] P. Lazarsfeld and R. Merton, Friendship as a social process: a substantive and methodological analysis, in *Freedom and control in modern society*, edited by M. Berger (Van Nostrand, 1954), pp. 18–66.

[150] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: homophily in social networks, Annual Review of Sociology **27**, 415–444 (2001).

[151] K. Börner, J. T. Maru, and R. L. Goldstone, The simultaneous evolution of author and paper networks, Proceedings of the National Academy of Sciences **101**, 5266–5273 (2004).

[152] F. Menczer, Growing and navigating the small world web by local content, Proceedings of the National Academy of Sciences **99**, 14014–14019 (2002).

[153] S. Talaga and A. Nowak, Structural measures of similarity and complementarity in complex networks, Scientific Reports **12**, 16580 (2022).

[154] G. Budel and M. Kitsak, Complementarity in complex networks, arXiv:2003.06665, 2023.

[155] J. B. Tenenbaum, V. de Silva, and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science **290**, 2319–2323 (2000).

[156] A. Sarveniazi, An actual survey of dimensionality reduction, American Journal of Computational Mathematics **04**, 55–72 (2014).

[157] W. Jia, M. Sun, J. Lian, and S. Hou, Feature dimensionality reduction: a review, Complex & Intelligent Systems **8**, 2663–2693 (2022).

[158] R. Jankowski, A. Allard, M. Boguñá, and M. Á. Serrano, The D-mercator method for the multidimensional hyperbolic embedding of real networks, Nature Communications **14**, 7585 (2023).

[159] P. Almagro, M. Boguñá, and M. Á. Serrano, Detecting the ultra low dimensionality of real networks, Nature Communications **13**, 6096 (2022).

[160] B. Waxman, Routing of multipoint connections, IEEE Journal on Selected Areas in Communications **6**, 1617–1622 (1988).

[161] D. ben-Avraham, A. F. Rozenfeld, R. Cohen, and S. Havlin, Geographical embedding of scale-free networks, Physica A: Statistical Mechanics and its Applications **330**, 107–116 (2003).

[162] A. F. Rozenfeld, R. Cohen, D. ben-Avraham, and S. Havlin, Scale-free networks on lattices, Physical Review Letters **89**, 218701 (2002).

[163] M. Boguñá, D. Krioukov, P. Almagro, and M. Á. Serrano, Small worlds and clustering in spatial networks, Physical Review Research **2**, 023040 (2020).

[164] R. de Miguel and J. M. Rubí, Strong coupling and nonextensive thermodynamics, Entropy **22**, 975 (2020).

[165] R. de Miguel and J. M. Rubí, Statistical mechanics at strong coupling: a bridge between Landsberg's energy levels and Hill's nanothermodynamics, Nanomaterials **10**, 2471 (2020).

[166] R. de Miguel, Rayleigh–Schrödinger perturbation theory and nonadditive thermodynamics, The Journal of Physical Chemistry B **127**, 5089–5093 (2023).

[167] O. Shental and I. Kanter, Shannon meets Carnot: Generalized second thermodynamic law, Europhysics Letters **85**, 10006 (2009).

[168] E. W. Elcock and P. T. Landsberg, Temperature dependent energy levels in statistical mechanics, Proceedings of the Physical Society. Section B **70**, 161–168 (1957).

[169] C. B. Cuden, "Temperature dependence of the energy gaps in semi-conductors," PhD thesis (University of British Columbia, 1969).

[170] G. S. Rushbrooke, On the statistical mechanics of assemblies whose energy-levels depend on the temperature, Transactions of the Faraday Society **36**, 1055 (1940).

[171] Y. Varshni, Temperature dependence of the energy gap in semiconductors, Physica **34**, 149–154 (1967).

[172] C. E. P. Villegas, A. R. Rocha, and A. Marini, Anomalous temperature dependence of the band gap in black phosphorus, Nano Letters **16**, 5095–5101 (2016).

[173] I. Latella, A. Pérez-Madrid, A. Campa, L. Casetti, and S. Ruffo, Thermodynamics of nonadditive systems, Physical Review Letters **114**, 230601 (2015).

[174] A. Campa, T. Dauxois, D. Fanelli, and S. Ruffo, *Physics of long-range interacting systems* (Oxford University Press, 2014).

[175] K. Bringmann, R. Keusch, and J. Lengler, Average distance in a general class of scale-free networks with underlying geometry, (2016).

[176] K. Bringmann, R. Keusch, and J. Lengler, Geometric inhomogeneous random graphs, Theoretical Computer Science **760**, 35–54 (2019).

[177] K. Kosmidis, S. Havlin, and A. Bunde, Structural properties of spatially embedded networks, Europhysics Letters **82**, 48005 (2008).

[178] M. Biskup, On the scaling of the chemical distance in long-range percolation models, The Annals of Probability **32** (2004).

[179] M. Biskup, Graph diameter in long-range percolation, Random Structures & Algorithms **39**, 210–227 (2011).

[180] I. Benjamini and N. Berger, The diameter of long-range percolation clusters on finite cycles, Random Structures & Algorithms **19**, 102–111 (2001).

[181] P. Deprez, R. Hazra, and M. Wüthrich, Inhomogeneous long-range percolation for real-life network modeling, Risks **3**, 1–23 (2015).

[182] S. G. Balogh, B. Kovács, and G. Palla, Maximally modular structure of growing hyperbolic networks, Communications Physics **6**, 76 (2023).

[183] J. Chellig, N. Fountoulakis, and F. Skerman, The modularity of random graphs on the hyperbolic plane, Journal of Complex Networks **10** (2021).

[184] B. Kovács and G. Palla, The inherent community structure of hyperbolic networks, Scientific Reports **11**, 16050 (2021).

[185] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Identity and search in social networks, Science **296**, 1302–1305 (2002).

[186] S. Redner, How popular is your paper? An empirical study of the citation distribution, The European Physical Journal B - Condensed Matter and Complex Systems **4**, 131–134 (1998).

[187] M. Deijfen, R. van der Hofstad, and G. Hooghiemstra, Scale-free percolation, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **49** (2013).

[188] P. Deprez and M. V. Wüthrich, Scale-free percolation in continuum space, Communications in Mathematics and Statistics **7**, 269–308 (2019).

[189] E. Garuccio, M. Lalli, and D. Garlaschelli, Multiscale network renormalization: Scale-invariance without geometry, Physical Review Research **5**, 043101 (2023).

[190] R. Michielan and C. Stegehuis, Cliques in geometric inhomogeneous random graphs, Journal of Complex Networks **10** (2021).

[191] J. Komjáthy and B. Lodewijks, Explosion in weighted hyperbolic random graphs and geometric inhomogeneous random graphs, Stochastic Processes and their Applications **130**, 1309–1367 (2020).

[192] R. van der Hofstad and J. Komjathy, Explosion and distances in scale-free percolation, (2017).

[193] M. Heydenreich, T. Hulshof, and J. Jorritsma, Structures in supercritical scale-free percolation, The Annals of Applied Probability **27** (2017).

[194] F. Papadopoulos, R. Aldecoa, and D. Krioukov, Network geometry inference using common neighbors, Physical Review E **92**, 22807 (2015).

[195] Z. Wu, Z. Di, and Y. Fan, An asymmetric popularity-similarity optimization method for embedding directed networks into hyperbolic space, Complexity **2020**, 1–16 (2020).

[196] G. Alanis-Lobato, P. Mier, and M. A. Andrade-Navarro, Efficient embedding of complex networks to hyperbolic space via their laplacian, Scientific Reports **6**, 30108 (2016).

[197] A. Muscoloni, J. M. Thomas, S. Ciucci, G. Bianconi, and C. V. Cannistraci, Machine learning meets complex networks via coalescent embedding in the hyperbolic space, Nature Communications **8**, 1615 (2017).

[198] M. Belkin and P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in *Advances in neural information processing systems*, Vol. 14, edited by T. Dietterich, S. Becker, and Z. Ghahramani (2001).

[199] M. D. Ben Chamberlain and J. Clough, Neural embeddings of graphs in hyperbolic space, in *Proceedings of the 13th international workshop on mining and learning with graphs (mlg)* (2017).

[200] D. McDonald and S. He, Heat: Hyperbolic Embedding of Attributed Networks, in *Intelligent data engineering and automated learning – ideal 2020: 21st international conference* (2020), pp. 28–40.

[201] I. Chami, Z. Ying, C. Ré, and J. Leskovec, Hyperbolic graph convolutional neural networks, in *Advances in neural information processing systems*, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (2019).

[202] P. Salomonson and B.-S. Skagerstam, On superdense superstring gases: a heretic string model approach, Nuclear Physics **268**, 349–361 (1986).

[203] B. Sundborg, Thermodynamics of superstrings at high energy densities, Nuclear Physics B **254**, 583–592 (1985).

[204] H. A. Bethe, An attempt to calculate the number of energy levels of a heavy nucleus, Physical Review **50**, 332–341 (1936).

[205] M. C. J. M. Vissenberg and M. Matters, Theory of the field-effect mobility in amorphous organic transistors, Physical Review B **57**, 12964–12967 (1998).

[206] K. Hart, S. Hart, and J. P. Selvaggi, Modified charge carrier density for organic semiconductors modeled by an exponential density of states, Journal of Computational Electronics **20**, 259–266 (2021).

[207] F. R. Shapiro and D. Adler, Equilibrium transport in amorphous semiconductors, Journal of Non-Crystalline Solids **74**, 189–194 (1985).

[208] W. Broniowski, W. Florkowski, and L. Y. Glozman, Update of the Hagedorn mass spectrum, Physical Review D **70**, 117503 (2004).

[209] R. Hagedorn, Statistical thermodynamics of strong interactions at high energies, Nuovo Cimento, Suppl. **3**, 147–186 (1965).

[210] J. Rafelski, ed., *Melting hadrons, boiling quarks - from Hagedorn temperature to ultra-relativistic heavy-ion collisions at CERN* (Springer International Publishing, 2016).

[211] S. Frautschi, Statistical bootstrap model of hadrons, Physical Review D **3**, 2821–2834 (1971).

[212] J. Yellin, An explicit solution of the statistical bootstrap, Nuclear Physics B **52**, 583–594 (1973).

[213] R. Hagedorn, The long way to the Statistical Bootstrap Model: 1994, in *Melting hadrons, boiling quarks - from Hagedorn temperature to ultra-relativistic heavy-ion collisions at CERN* (Springer International Publishing, 2016), pp. 139–178.

[214] P. Blanchard, S. Fortunato, and H. Satz, The Hagedorn temperature and partition thermodynamics, The European Physical Journal C **34**, 361–366 (2004).

[215] R. Hagedorn and J. Rafelski, From hadron gas to quark matter 1, in *Proceedings of statistical mechanics of quarks and hadrons* (1980), pp. 237–251.

[216] M. Gorenstein, V. Petrov, and G. Zinovjev, Phase transition in the hadron gas model, Physics Letters B **106**, 327–330 (1981).

[217] R. Hagedorn, On the hadronic mass spectrum, in *Melting hadrons, boiling quarks - from Hagedorn temperature to ultra-relativistic heavy-ion collisions at CERN* (Springer International Publishing, 2016), pp. 223–228.

[218] T. L. Hill, *Thermodynamics of small systems, parts I & II* (Dover, 1994).

[219] P. Hohenberg and A. Krekhov, An introduction to the Ginzburg–Landau theory of phase transitions and nonequilibrium patterns, Physics Reports **572**, 1–42 (2015).

[220] V. L. Berezinskii, Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group I. classical systems. Soviet physics, JETP **32**, 493–500 (1971).

[221] V. L. Berezinskii, Destruction of long-range order in one-dimensional and two-dimensional systems possessing a continuous symmetry group. II. quantum systems. Soviet Physics, JETP **34**, 610–616 (1972).

[222] J. M. Kosterlitz, The critical properties of the two-dimensional XY model, Journal of Physics C: Solid State Physics **7**, 1046–1060 (1974).

[223] N. D. Mermin and H. Wagner, Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models, Physical Review Letters **17**, 1133–1136 (1966).

[224] C.-K. Chiu, J. C. Teo, A. P. Schnyder, and S. Ryu, Classification of topological quantum matter with symmetries, Reviews of Modern Physics **88**, 035005 (2016).

[225] J. L. Gross, J. Yellen, and M. Anderson, *Graph theory and its applications* (Chapman and Hall/CRC, 2018).

[226] M. Starnini, E. Ortiz, and M. Á. Serrano, Geometric randomization of real networks with prescribed degree sequence, New Journal of Physics **21**, 053039 (2019).

[227] T. P. Peixoto, Disentangling homophily, community structure, and triadic closure in networks, Physical Review X **12**, 011004 (2022).

[228] F. Papadopoulos, C. Psomas, and D. Krioukov, Network mapping by replaying hyperbolic growth, IEEE/ACM Transactions on Networking **23**, 198–211 (2015).

[229] T. Blasius, T. Friedrich, A. Krohmer, and S. Laue, Efficient embedding of scale-free graphs in the hyperbolic plane, IEEE/ACM Transactions on Networking **26**, 920–933 (2018).

[230] P. Goyal and E. Ferrara, Graph embedding techniques, applications, and performance: A survey, Knowledge-Based Systems **151**, 78–94 (2018).

[231] T. Bläsius, T. Friedrich, M. Katzmann, and A. Krohmer, Hyperbolic embeddings for near-optimal greedy routing, ACM Journal of Experimental Algorithmics **25**, 1–18 (2020).

[232] G. Alanis-Lobato, P. Mier, and M. A. Andrade-Navarro, Efficient embedding of complex networks to hyperbolic space via their Laplacian, Scientific Reports **6**, 30108 (2016).

[233] M. Keller-Ressel and S. Nargang, Hydra: a method for strain-minimizing hyperbolic embedding of network- and distance-based data, Journal of Complex Networks **8** (2020).

[234] J. A. Dunne, C. C. Labandeira, and R. J. Williams, Highly resolved early eocene food webs show development of modern trophic structure after the end-cretaceous extinction, Proceedings of the Royal Society B: Biological Sciences **281**, 20133280 (2014).

[235] G. W. Cobb and Y.-P. Chen, An application of markov chain monte carlo to community ecology, The American Mathematical Monthly **110**, 265 (2003).

[236] L. Zagar, F. Mulas, S. Garagna, M. Zuccotti, R. Bellazzi, and B. Zupan, Stage prediction of embryonic stem cell differentiation from genome-wide expression data, Bioinformatics **27**, 2546–2553 (2011).

[237] R. A. da Costa, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Explosive percolation transition is actually continuous, Physical Review Letters **105**, 255701 (2010).

[238] R. M. D'Souza and J. Nagler, Anomalous critical and supercritical phenomena in explosive percolation, Nature Physics **11**, 531–538 (2015).

[239] U. C. Täuber, Renormalization group: applications in statistical physics, Nuclear Physics B - Proceedings Supplements **228**, 7–34 (2012).

[240] L. P. Kadanoff, Scaling laws for ising models near $T_c$, Physics Physique Fizika **2**, 263–272 (1966).

[241] C. Song, S. Havlin, and H. A. Makse, Self-similarity of complex networks, Nature **433**, 392–395 (2005).

[242] P. Villegas, A. Gabrielli, F. Santucci, G. Caldarelli, and T. Gili, Laplacian paths in complex networks: information core emerges from entropic transitions, Physical Review Research **4**, 033196 (2022).

[243] P. Villegas, T. Gili, G. Caldarelli, and A. Gabrielli, Laplacian renormalization group for heterogeneous networks, Nature Physics **19**, 445–450 (2023).

[244] M. Zheng, A. Allard, P. Hagmann, Y. Alemán-Gómez, and M. Á. Serrano, Geometric renormalization unravels self-similarity of the multiscale human connectome, Proceedings of the National Academy of Sciences **117**, 20244–20253 (2020).

[245] J. Moody, Peer influence groups: identifying dense clusters in large networks, Social Networks **23**, 261–283 (2001).

[246] R. G. Little, Controlling cascading failure: understanding the vulnerabilities of interconnected infrastructures, Journal of Urban Technology **9**, 109–123 (2002).

[247] L. M. Verbrugge, Multiplexity in adult friendships, Social Forces **57**, 1286 (1979).

[248] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, Tastes, ties, and time: A new social network dataset using Facebook.com, Social Networks **30**, 330–342 (2008).

[249] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science **328**, 876–878 (2010).

[250] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, Diffusion dynamics on multiplex networks, Physical Review Letters **110**, 028701 (2013).

[251] M. Szell, R. Lambiotte, and S. Thurner, Multirelational organization of large-scale social networks in an online world, Proceedings of the National Academy of Sciences **107**, 13636–13641 (2010).

[252] R. G. Morris and M. Barthelemy, Transport on coupled spatial networks, Physical Review Letters **109**, 128703 (2012).

[253] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, and S. Boccaletti, Emergence of network features from multiplexity, Scientific Reports **3**, 1344 (2013).

[254] A. Solé-Ribalta, M. D. Domenico, N. E. Kouvaris, A. Díaz-Guilera, S. Gómez, and A. Arenas, Spectral properties of the Laplacian of multiplex networks, Physical Review E **88**, 032807 (2013).

[255] G. Bianconi, *Multilayer networks* (Oxford University PressOxford, 2018).

[256] O. Artime, B. Benigni, G. Bertagnolli, V. d'Andrea, R. Gallotti, A. Ghavasieh, S. Raimondo, and M. D. Domenico, *Multilayer network science* (Cambridge University Press, 2022).

[257] F. Battiston, V. Nicosia, and V. Latora, Structural measures for multiplex networks, Physical Review E **89**, 032804 (2014).

[258] F. Battiston, V. Nicosia, M. Chavez, and V. Latora, Multilayer motif analysis of brain networks, Chaos: An Interdisciplinary Journal of Nonlinear Science **27** (2017).

[259] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, The structure and dynamics of multilayer networks, Physics Reports **544**, 1–122 (2014).

[260] A. M. Abdolhosseini-Qomi, N. Yazdani, and M. Asadpour, Overlapping communities and the prediction of missing links in multiplex networks, Physica A: Statistical Mechanics and its Applications **554**, 124650 (2020).

[261] C. H. Kim, M. Jo, J. S. Lee, G. Bianconi, and B. Kahng, Link overlap influences opinion dynamics on multiplex networks of ashkin-teller spins, Physical Review E **104**, 064304 (2021).

[262] D. T. Luu and T. Lux, Multilayer overlaps and correlations in the bank-firm credit network of spain, Quantitative Finance **19**, 1953–1974 (2019).

[263] T. Dimitrova, K. Petrovski, and L. Kocarev, Graphlets in multiplex networks, Scientific Reports **10**, 1928 (2020).

[264] H. Wang, C. Ma, H. Chen, and H. Zhang, Effect of overlap on spreading dynamics on multiplex networks, Journal of Statistical Mechanics: Theory and Experiment **2020**, 043402 (2020).

[265] G. J. Baxter, G. Bianconi, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, Correlated edge overlaps in multiplex networks, Physical Review E **94**, 012303 (2016).

[266] G. Bianconi, Statistical mechanics of multiplex networks: Entropy and overlap, Physical Review E **87**, 062806 (2013).

[267] F. Papadopoulos and K.-K. Kleineberg, Link persistence and conditional distances in multiplex networks, Physical Review E **99**, 012322 (2019).

[268] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature **464**, 1025–1028 (2010).

[269] M. Á. Serrano, L. Buzna, and M. Boguñá, Escaping the avalanche collapse in self-similar multiplexes, New Journal of Physics **17**, 053033 (2015).

[270] S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, Percolation theory on interdependent networks based on epidemic spreading, Europhysics Letters **97**, 16006 (2012).

[271] A. M. Turing, The chemical basis of morphogenesis, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **237**, 37–72 (1952).

[272] A. Gierer and H. Meinhardt, A theory of biological pattern formation, Kybernetik **12**, 30–39 (1972).

[273] H. Meinhardt, *Models of biological pattern formation* (Academic Press, 1982).

[274] M. C. Cross and P. C. Hohenberg, Pattern formation outside of equilibrium, Reviews of Modern Physics **65**, 851–1112 (1993).

[275] A. J. Koch and H. Meinhardt, Biological pattern formation: from basic mechanisms to complex structures, Reviews of Modern Physics **66**, 1481–1507 (1994).

[276] A. M. Zhabotinsky, M. Dolnik, and I. R. Epstein, Pattern formation arising from wave instability in a simple reaction-diffusion system, The Journal of Chemical Physics **103**, 10306–10314 (1995).

[277] M. Cross and H. Greenside, *Pattern formation and dynamics in nonequilibrium systems* (Cambridge University Press, 2009).

[278] S. Kondo and T. Miura, Reaction-diffusion model as a framework for understanding biological pattern formation, Science **329**, 1616–1620 (2010).

[279] D. Walgraef, *Spatio-temporal pattern formation* (Springer New York, 1997).

[280] H. Othmer and L. Scriven, Instability and dynamic pattern in cellular networks, Journal of Theoretical Biology **32**, 507–537 (1971).

[281] W. Horsthemke, K. Lam, and P. K. Moore, Network topology and Turing instabilities in small arrays of diffusively coupled reactors, Physics Letters A **328**, 444–451 (2004).

[282] P. K. Moore and W. Horsthemke, Localized patterns in homogeneous networks of diffusively coupled reactors, Physica D: Nonlinear Phenomena **206**, 121–144 (2005).

[283] H. Nakao and A. S. Mikhailov, Turing patterns in network-organized activator–inhibitor systems, Nature Physics **6**, 544–550 (2010).

[284] M. Asllani, J. D. Challenger, F. S. Pavone, L. Sacconi, and D. Fanelli, The theory of pattern formation on directed networks, Nature Communications **5**, 4517 (2014).

[285] R. Muolo, M. Asllani, D. Fanelli, P. K. Maini, and T. Carletti, Patterns of non-normality in networked systems, Journal of Theoretical Biology **480**, 81–91 (2019).

[286] J. Petit, B. Lauwens, D. Fanelli, and T. Carletti, Theory of Turing patterns on time varying networks, Physical Review Letters **119**, 148301 (2017).

[287] R. A. V. Gorder, A theory of pattern formation for reaction–diffusion systems on temporal networks, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **477**, 20200753 (2021).

[288] R. Muolo, L. Gallo, V. Latora, M. Frasca, and T. Carletti, Turing patterns in systems with high-order interactions, Chaos, Solitons & Fractals **166**, 112912 (2023).

[289] S. Gao, L. Chang, M. Perc, and Z. Wang, Turing patterns in simplicial complexes, Physical Review E **107**, 014216 (2023).

[290] M. Asllani, D. M. Busiello, T. Carletti, D. Fanelli, and G. Planchon, Turing patterns in multiplex networks, Physical Review E **90**, 042814 (2014).

[291] M. Asllani, D. M. Busiello, T. Carletti, D. Fanelli, and G. Planchon, Turing instabilities on Cartesian product networks, Scientific Reports **5**, 12927 (2015).

[292] N. E. Kouvaris, S. Hata, and A. D. Guilera, Pattern formation in multiplex networks, Scientific Reports **5**, 10840 (2015).

[293] D. M. Busiello, T. Carletti, and D. Fanelli, Homogeneous-per-layer patterns in multiplex networks, Europhysics Letters **121**, 48006 (2018).

[294] B. A. Siebert, C. L. Hall, J. P. Gleeson, and M. Asllani, Role of modularity in self-organization dynamics in biological networks, Physical Review E **102**, 052306 (2020).

[295] M. Asllani, T. Carletti, and D. Fanelli, Tune the topology to create or destroy patterns, The European Physical Journal B **89**, 260 (2016).

[296] G. Cencetti, F. Battiston, T. Carletti, and D. Fanelli, Generalized patterns from local and non local reactions, Chaos, Solitons & Fractals **134**, 109707 (2020).

[297] M.-T. Hütt, D. Armbruster, and A. Lesne, Predictable topological sensitivity of Turing patterns on graphs, Physical Review E **105**, 014304 (2022).

[298] E. W. Montroll and G. H. Weiss, Random walks on lattices. II, Journal of Mathematical Physics **6**, 167–181 (1965).

[299] G. H. Weiss, *Aspects and applications of the random walk* (North-Holland, Amsterdam, 1994).

[300] P. van Mieghem, *Graph spectra for complex networks* (Cambridge University Press, 2010).

[301] I. Prigogine and R. Lefever, Symmetry breaking instabilities in dissipative systems. II, The Journal of Chemical Physics **48**, 1695–1700 (1968).

[302] V. Castets, E. Dulos, J. Boissonade, and P. D. Kepper, Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern, Physical Review Letters **64**, 2953–2956 (1990).

[303] Q. Ouyang and H. L. Swinney, Transition from a uniform state to hexagonal and striped Turing patterns, Nature **352**, 610–612 (1991).

[304] M. P. Harris, S. Williamson, J. F. Fallon, H. Meinhardt, and R. O. Prum, Molecular evidence for an activator–inhibitor mechanism in development of embryonic feather branching, Proceedings of the National Academy of Sciences **102**, 11734–11739 (2005).

[305] L. A. Segel and S. A. Levin, Application of nonlinear stability theory to the study of the effects of diffusion on predator-prey interactions, in *Aip conference proceedings* (1976), pp. 123–152.

[306] M. Mimura and J. Murray, On a diffusive prey-predator model which exhibits patchiness, Journal of Theoretical Biology **75**, 249–262 (1978).

[307] J. L. Maron and S. Harrison, Spatial pattern formation in an insect host-parasitoid system, Science **278**, 1619–1621 (1997).

[308] J. P. Gibert and J. D. Yeakel, Laplacian matrices and Turing bifurcations: revisiting Levin 1974 and the consequences of spatial structure and movement for ecological dynamics, Theoretical Ecology **12**, 265–281 (2019).

[309] A. K. Fahimipour, F. Zeng, M. Homer, A. Traulsen, S. A. Levin, and T. Gross, Sharp thresholds limit the benefit of defector avoidance in cooperation on networks, Proceedings of the National Academy of Sciences **119** (2022).

[310] A. Brechtel, P. Gramlich, D. Ritterskamp, B. Drossel, and T. Gross, Master stability functions reveal diffusion-driven pattern formation in networks, Physical Review E **97**, 032307 (2018).

[311] S. Hata and H. Nakao, Localization of laplacian eigenvectors on random networks, Scientific Reports **7**, 1121 (2017).

[312] G. Pólya and G. Szegö, *Isoperimetric inequalities in mathematical physics* (Princeton University Press, 1951).

[313] J. Cheeger, A lower bound for the smallest eigenvalue of the laplacian, in *Problems in analysis*, edited by R. C. Gunning (1969), pp. 195–199.

[314] B. Derrida and Y. Pomeau, Random networks of automata: a simple annealed approximation, Europhysics Letters **1**, 45–49 (1986).

[315] S.-i. Amari, N. Fujita, and S. Shinomoto, Four types of learning curves, Neural Computation **4**, 605–618 (1992).

[316] U. Bastolla and G. Parisi, Closing probabilities in the Kauffman model: An annealed computation, Physica D: Nonlinear Phenomena **98**, 1–25 (1996).

[317] B. Luque and R. V. Solé, Phase transitions in random networks: simple analytic determination of critical points, Physical Review E **55**, 257–260 (1997).

[318] T. Rohlf and S. Bornholdt, Criticality in random threshold networks: annealed approximation and beyond, Physica A: Statistical Mechanics and its Applications **310**, 245–259 (2002).

[319] D. Vilone and C. Castellano, Solution of voter model dynamics on annealed small-world networks, Physical Review E **69**, 016109 (2004).

[320] B. Guerra and J. Gómez-Gardeñes, Annealed and mean-field formulations of disease dynamics on static and adaptive networks, Physical Review E **82**, 035101 (2010).

[321] S. C. Ferreira, R. S. Ferreira, and R. Pastor-Satorras, Quasistationary analysis of the contact process on annealed scale-free networks, Physical Review E **83**, 066113 (2011).

[322] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, Mapping the structural core of human cerebral cortex, PLoS Biology **6**, e159 (2008).

[323] E. Ortiz and M. Á. Serrano, Multiscale voter model on real networks, Chaos, Solitons & Fractals **165**, 112847 (2022).

[324] K.-K. Kleineberg, Metric clusters in evolutionary games on scale-free networks, Nature Communications **8**, 1888 (2017).

[325] G. Budel, M. Kitsak, R. Aldecoa, K. Zuev, and D. Krioukov, Random hyperbolic graphs in $d + 1$ dimensions, Physical Review E **109**, 054131 (2024).

[326] T. Squartini, J. de Mol, F. den Hollander, and D. Garlaschelli, Breaking of ensemble equivalence in networks, Physical Review Letters **115**, 268701 (2015).

[327] K. Anand and G. Bianconi, Entropy measures for networks: Toward an information theory of complex topologies, Physical Review E **80**, 045102 (2009).

[328] Q. Zhang and D. Garlaschelli, Strong ensemble nonequivalence in systems with local constraints, New Journal of Physics **24**, 043011 (2022).

[329] M. Boguñá and D. Krioukov, Measuring spatial distances in causal sets via causal overlaps, Physical Review D **110**, 024008 (2024).

[330] R. E. Ulanowicz and D. L. DeAngelis, *Network analysis of trophic dynamics in south florida ecosystems*, 2005.

[331] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Superfamilies of evolved and designed networks. Science **303**, 1538–1542 (2004).

[332] M. Huss and P. Holme, Currency and commodity metabolites: their identification and relation to the modularity of metabolic networks, IET Systems Biology **1**, 280–285 (2007).

[333] J. Kunegis, Konect, in *Proceedings of the 22nd international conference on world wide web* (2013), pp. 1343–1350.

[334] Y. Hu, A. Vinayagam, A. Nand, A. Comjean, V. Chung, T. Hao, S. E. Mohr, and N. Perrimon, Molecular Interaction Search Tool (MIST): An integrated resource for mining gene and protein interaction data, Nucleic Acids Research **46**, D567–D574 (2018).

[335] M. D. Domenico, M. A. Porter, and A. Arenas, Muxviz: A tool for multilayer analysis and visualization of networks, Journal of Complex Networks **3**, 159–176 (2015).

[336] M. Ripeanu and I. Foster, Mapping the Gnutella network: Macroscopic properties of large-scale peer-to-peer systems, in (2002), pp. 85–93.

[337] M. D. Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, Navigability of interconnected networks under random failures, Proceedings of the National Academy of Sciences **111**, 8351–8356 (2014).

[338] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, The internet topology zoo, IEEE Journal on Selected Areas in Communications **29**, 1765–1775 (2011).

[339] J. Leskovec, D. Huttenlocher, and J. Kleinberg, Signed networks in social media, in *Proceedings of the sigchi conference on human factors in computing systems* (2010), pp. 1361–1370.

[340] A. Paranjape, A. R. Benson, and J. Leskovec, Motifs in temporal networks, in *Proceedings of the tenth acm international conference on web search and data mining* (2017), pp. 601–610.

[341] M. D. Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems, Physical Review X **5**, 011027 (2015).

[342] M. D. Domenico, V. Nicosia, A. Arenas, and V. Latora, Structural reducibility of multilayer networks, Nature Communications **6**, 6864 (2015).

[343] E. Omodei, M. D. Domenico, and A. Arenas, Characterizing interactions in online social networks during exceptional events, Frontiers in Physics **3** (2015).

[344] A. Cho, J. Shin, S. Hwang, C. Kim, H. Shim, H. Kim, H. Kim, and I. Lee, WormNet v3: a network-assisted hypothesis-generating server for Caenorhabditis elegans, Nucleic Acids Res. **42**, W76–W82 (2014).

[345] D. B. Larremore, A. Clauset, and C. O. Buckee, A network approach to analyzing highly recombinant malaria parasite genes, PLoS Computational Biology **9**, e1003268 (2013).

[346] S. W. Oh et al., A mesoscale connectome of the mouse brain, Nature **508**, 207–214 (2014).

[347] M. Á. Serrano and M. Boguñá, Topology of the world trade web, Physical Review E **68**, 015101 (2003).

[348] M. Á. Serrano, M. Boguñá, and A. Vespignani, Patterns of dominant flows in the world trade web, Journal of Economic Interaction and Coordination **2**, 111–124 (2007).