UNIVERSITAT DE BARCELONA

Evolution and ecology of the digital world

A complex systems perspective

Kaj Kolja Kleineberg

PhD Thesis

Evolution and ecology of the digital world

a complex systems perspective

Author: Kaj Kolja Kleineberg
Supervisor: Marián Boguñá
Tutor: Ignacio Pagonabarraga

March 4, 2016

UNIVERSITAT DE BARCELONA

Programa de Doctorat en Física

"We are all now connected by the Internet, like neurons in a giant brain."
STEPHEN HAWKING

Contents

Preface

"If our brains were simple enough for us to understand them, we'd be so simple that we couldn't."
Ian Stewart [1]

About this thesis

This thesis deals with challenges faced by an interconnected world that is strongly influenced by the recent digital revolution. Naturally, connectedness is represented in terms of networks: systems that consist of nodes and of edges that represent connections between the nodes. Nodes can be individuals, locations, technological entities such as Internet routers, or neurons in the brain, to mention only a few. Hence, edges represent physical connections or interactions, such as friendship relations or gene interactions.

Connectedness implies the coupling of many systems or components within a large entity. In reality, those large entities often exhibit behaviors that cannot be explained by the interaction of their components in isolation; a phenomenon called emergence. The study of emergence forms part of complexity theory, which is introduced in the following pages.

This thesis has two main parts titled *Evolution and ecology of the digital world* and *Geometry of multiplex networks*. In *Evolution and ecology of the digital world*, we develop a series of models based on empirical observations in order to describe the evolution of and competitive interactions between online social networks. The most natural description of these systems requires the treatment of interconnected networks, i.e. several networks that are coupled and hence interact in some way. We consider important properties of interacting networks in *Geometry of multiplex networks*, where we use a geometric approach to study the structure and function of so-called multiplex networks: systems that consist of multiple network layers and in which the same node can be present in different layers.

In the following pages, I provide a brief non-technical introduction to the digital revolution; present related opportunities and risks for society; and consider pre-digital social networks, complex systems, and complex networks. For a comprehensive introduction to the science of complexity, nonlinear physics, and complex networks, I recommend the following books:

- Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering by Steven Strogatz [2]
- Die Erforschung des Chaos: eine Einführung in die Theorie nichtlinearer Systeme by John Argyris, Gunter Faust, Maria Haase, and Rudolf Friedrich (in German) [3]

- Introduction to the Modeling and Analysis of Complex Systems by Hiroki Sayama [4]
- Networks: An Introduction by Mark Newman [5]
- Dynamical Processes on Complex Networks by Alain Barrat, Marc Barthlemy, and Alessandro Vespignani [6]
- The Structure and Dynamics of Networks by Mark Newman, Albert-László Barabási, and Duncan Watts [7]

For further reading on networks, the digital revolution, and its impact on society and the economy, I recommend the following non-technical books:

- The Rise of the Network Society by Manuel Castells [8]
- Everything Is Obvious: How Common Sense Fails Us by Duncan Watts [9]
- Why information grows: the evolution of order, from atoms to economies by César Hidalgo [10]
- Thinking Ahead - Essays on Big Data, Digital Revolution, and Participatory Market Society by Dirk Helbing [11]

Related publications

Parts of this thesis were published in the following papers/preprints:

- Evolution of the Digital Society reveals Balance between Mass Media and Viral Influence, Physical Review X (4) 031046, K. Kleineberg and M. Boguñá
- Digital Ecology: Coexistence and Domination among Interacting Networks, Sci. Rep. 5, 10268, K. Kleineberg and M. Boguñá
- Is bigger always better? How local online social networks can outperform global ones, arXiv:1504.01368, K. Kleineberg and M. Boguñá, *under review*
- Geometric correlations in real multiplex networks: multidimensional communities, trans-layer link prediction, and efficient navigation, arXiv:1601.04071, K. Kleineberg, M. Boguñá, M. A. Serrano, F. Papadopoulos, *under review*

Acknowledgements

I would like to thank my advisor Marián Boguñá for the support and many helpful discussions related to my research. He was always open to my ideas and also helped me to get back on the right path whenever necessary. It was a pleasure to have him as a travel companion to many project meetings and conferences as well as on a motorcycle tour in the Pyrenees. I could not have imagined having a better advisor and mentor for my PhD.

I thank all my collaborators and colleagues from the iSocial [12] project. I thank Šarūnas Girdzijauskas for great organization of the project in general and our project meetings in particular. I thank the fellows for the fun we had at our project meetings and all the questions and discussions during and after my talks. I thank Marios Dikaiakos and George Pallis for accepting my internship in Cyprus. I am especially grateful for the excellent collaboration and discussions with Fragkiskos Papadopoulos.

I also thank my colleagues from the group for discussions and for all the fun we had in the last years. Especially, I thank Pol for helping me with programming issues. I thank my other office mates Ula, Mario, Oriol, and Roberta. I thank Albert Díaz-Guilera for managing the research group and organizing all the group dinners, calçotadas, seminars, and so on. I thank M. Ángeles Serrano for helpful discussions on network geometry. I thank Pol and Ignacio for correcting the language of my Spanish summary.

I thank Alain Barrat for organizing the amazing Complex Networks Thematic School in Les Houches, where I learned many essential concepts in the field of complex networks and got into touch with the community. I thank all the organizers of the conferences I attended for the great organization and for giving me the opportunity to present my work.

Last but not least, I would like to thank my entire family. My parents always highly valuated education and their support brought me on the path that eventually led to this thesis.

Introduction

1 The digital revolution and the century of complexity

"The digital revolution is far more significant than the invention of writing or even of printing."
Douglas Engelbart

1.1 The digital revolution: an unprecedented opportunity

1.1.1 The rise of the Internet

During the Cold War, the Defense Advanced Research Projects Agency (DARPA) of the US Defense Department invented a system to ensure the preservation of communications in the event of a nuclear war. The system needed to be resilient against the removal or takeover of any of its parts. As a consequence, such a system had to be designed in a decentralized way, where every component could in general be replaced by every other component. Eventually, building on this architecture, the US Defense Department set up ARPANET, which later became the Internet that today links billions of devices worldwide and is used for communication in a post-Cold War, digitally connected world [8, 13].

The lack of any central control of the Internet means that it evolved and continues to evolve in a self-organized fashion. Indeed, the structure of the Internet exhibits a very interesting feature: the lack of scale. In particular, in [14], the authors showed that the degree distribution of the Internet at the Autonomous System[1] (AS) level is given by a power law with exponent $\gamma \approx 2.1$. This means that the probability that any randomly selected node has k connections is given by $P(k) \propto k^{-2.1}$.

The large size of the Internet presents a challenge: how to find efficient paths to route information from a given source to a given target. Nowadays, there are two versions of *Internet protocols*, namely IPv4 and IPv6, which are used to route Internet traffic. IPv4 is the version originally deployed in the ARPANET in 1983. IPv6 is the most recent version of the protocol. The networks corresponding to these two protocols can be represented as a two-layer system. In this thesis, in *Geometry of multiplex networks*, we discuss how routing paths can be found efficiently using both of these layers simultaneously and relying only on local knowledge. In particular, we offer an answer to the question "Under what conditions does routing using both layers of the system outperform the use of only a single layer?".

[1] In [15], an Autonomous System is defined as "a connected group of one or more IP prefixes run by one or more network operators which has a single and clearly defined routing policy".

1.1.2 Web 2.0: online social networks and computational social science

Whereas the part of this thesis titled *Geometry of multiplex networks* is devoted to the physical structure of the Internet and other networks, the part *Evolution and ecology of the digital world* provides a system-level perspective of platforms that enable interactions between users. The use of the Internet has experienced a paradigm shift from a collection of static pages, the original World Wide Web that we can term Web 1.0, to become an omnipresent user-centered interactive medium: Web 2.0. According to Cormode and Krishnamurthy, Web 1.0 is characterized by the following:

> *"Content creators were few in Web 1.0 with the vast majority of users simply acting as consumers of content."*
>
> Cormode, G. and Krishnamurthy, B., 2008 [16]

It was Cormode and Krishnamurthy who popularized the term Web 2.0, however, the expression dates back to Darcy DiNucci who in 1999 already predicted very precisely the influence the Internet would have and its use nowadays, more than 15 years later:

> *"The Web we know now, which loads into a browser window in essentially static screenfuls, is only an embryo of the Web to come. The first glimmerings of Web 2.0 are beginning to appear, and we are just starting to see how that embryo might develop. The Web will be understood not as screenfuls of text and graphics but as a transport mechanism, the ether through which interactivity happens. It will [...] appear on your computer screen, [...] on your TV set [...] your car dashboard [...] your cell phone [...] hand-held game machines [...] maybe even your microwave oven."*
>
> Darcy DiNucci, 1999 [16,17]

The nature of Web 2.0 is to promote collaboration on an unprecedented scale. As a consequence, new platforms referred to as "Wikis" have emerged, with Wikipedia the most popular example. The collaborative nature of Web 2.0 has also led to the rapid growth of large-scale open-source projects and more recently to crowdsourcing and crowdfunding. The key players in Web 2.0 are surely online social networks such as Facebook, Twitter, Google+, or LinkedIn, which are platforms designed to promote interaction between individuals in a digital environment at the global scale. Online social networks such as Twitter or Facebook have recently seen exceptional growth and they now[2] account for more than two billion active accounts [18] and cover 72% of online U.S. adults [19].

> *"Social science still has not found its Kepler. But three hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin..."*
>
> Duncan Watts [9]

[2] As of 2014.

Web 2.0 and especially online social networks provide researchers with data related to human interactions on previously unattainable scales. The availability of such data has led to the foundation of a new scientific discipline called "computational social science" [20]. Computational social science aims to provide data-driven, quantitative descriptions of human behavior. This new discipline is intrinsically multidisciplinary as it employs a collection of ideas, concepts and methods from physics, computer science, and social science.

The unprecedented success of online social networks is at the root of computational social science. However, the mechanisms that determine the fate of digital services at the system level are still poorly understood. In this thesis, we aim to provide insight into the evolution of and competitive interactions between online social networks using tools developed within the theory of complex systems. In the part titled *Evolution and ecology of the digital world*, we offer answers to questions such as: "How does the structure of online social networks emerge?", "Can several online social networks coexist or does competition always lead to one service dominating over others?" and "What mechanisms are responsible for Facebook overtaking its competitors and could things have turned out differently?".

1.1.3 Collective intelligence

"We are all now connected by the Internet, like neurons in a giant brain."
Stephen Hawking

Neurons perform quite simple tasks. In a nutshell, they react to a voltage; if that voltage is above a certain threshold, the neuron creates an electrochemical pulse that activates synaptic connections. Despite this simplicity, the combination of billions of interacting neurons in the human brain gives rise to an entity capable of developing a consciousness. Needless to say, the operations carried out by single neurons are remarkably simple compared to those executed by the whole brain. It is the connections, the synapses, that enable the emergence[3] of the intelligence of an individual.

Now, imagine a new kind of brain. A brain in which neurons are individuals, equipped with the capabilities of a whole human brain. It is impossible for us to imagine what such a system would be capable of – just as a single neuron cannot imagine consciousness. Let us call this new system "collective intelligence" [21]. We are to a neuron what the collective intelligence is to us. The connections in this new system, the synapses of collective intelligence, are digital services that connect billions of individuals, very nearly in real-time. We are still only witnessing at most the first glimpse of the potential emergent collective intelligence. Yet, it offers hope of solving problems in ways that we still cannot imagine.

1.1.4 The risk of totalitarian control and manipulation

Opportunity and risk always come together. It is no surprise that the same environment that provides the opportunity for the emergence of collective intelligence could also be

[3]Emergence or emergent behavior means that a system exhibits certain features or performs certain functions that cannot be explained by observing the constituents of the system in isolation. We introduce this concept in more detail in section 1.3 of the introduction.

the hatching ground of a dystopia. That disastrous society would be one in which a few entities had almost complete control over information [22] and where that power could be used to control and manipulate individuals on unprecedented scales [23, 24]. In that nightmare scenario, secret algorithms run by Google or Facebook would control the choices and emotions [25, 26] of individuals.

And then there is the opposite: a "digital democracy" where individuals are in control, and where digital diversity ensures the freedom of information and favors the emergence of collective intelligence [27]. In the following we discuss the fundamental principles that could sustain such a system.

1.1.5 Risk mitigation and digital prosperity: a manifesto for digital diversity and decentralization

The risks discussed in the previous section can be mitigated by two approaches, which can and should be combined. The first, which we refer to as digital diversity, can be seen by analogy with the situation involving traditional media. Every traditional medium (newspaper, TV station, or whatever) selects which news to cover and how. Hence, long before the digital age and the invention of the Internet, the so-called *framing effect* existed. However, the existence of a sufficient number of independent media providers (different TV stations, newspaper publishers, and so forth) mitigated the risk of public opinion being controlled or individuals being manipulated. In the digital world, only a few global players, such as Google and Facebook, control a huge share of the worldwide market. Whereas Google's search engine algorithms can be compared to framing in traditional media, the impact of Facebook is of a completely unprecedented type; for the first time, a centralized entity has the power to control social contagion and peer influence. The algorithms that determine what Facebook shows in the newsfeed of an individual have been shown to be capable of altering the emotional state of users and could be used to influence their decisions [28]. While such control offers almost unimaginable power to a few centralized entities, it would not arise in a diverse digital landscape where numerous services coexisted and competed for the attention of users. But is digital diversity possible from a system-level perspective or will the competition for the attention of users always lead to a single entity becoming dominant? In the part titled *Evolution and ecology of the digital world*, we provide a comprehensive answer to that question. In addition, it has been shown that diversity increases the performance of groups [27]; a finding that probably applies equally to the digital environment. Hence, a diverse digital landscape is a more fertile ground for collective intelligence to emerge and prosper, than one with a single dominant entity.

Diversity reduces the power of single centralized entities. Another approach to mitigate the risks of powerful, nontransparent monopolies of information is the decentralization of system architectures. Recently, Bitcoin [29], a decentralized system in the monetary sector, has attracted a lot of attention. The main difference between this currency and conventional ones is that by design it cannot be controlled by any single entity, such as a central bank. Analogously, decentralized information systems [12, 30] offer the advantage of transparency and impede abuse by single, centralized entities; in addition to offering better scalability.

So, such systems could pave the way for a true "digital democracy" [22]. A decentralized architecture, however, poses certain challenges for the operationality of the system; a key challenge being that of how to efficiently search and navigate with only local knowledge. In the part titled *Geometry of multiplex networks*, we provide an elegant solution to this challenge, especially for several interacting networks, as is the case in reality.

1.2 Pre-digital social networks

1.2.1 Milgram experiment

In 1967, Stanley Milgram published the results of an experiment he conducted to answer the following question:

> *"Given any two people in the world, person X and person Z, how many intermediate acquaintance links are needed before X and Z are connected?"*
>
> Stanley Milgram [31]

The experiment Milgram conducted was the following. Participants chosen as starting points of the study were provided with information about a target person to whom they had to send a letter. In particular, the information they were given was of the type that the target is the wife of a divinity school student or a stockbroker who works in Boston and lives in Sharon, Massachusetts. They were supposed to forward the letter only to people they knew on a first-name basis to avoid contacting the target directly if they did not know him or her personally. The results were quite surprising. On average, it took only five steps for the letter to reach the target for the 44 out of 160 that were successfully delivered [31]. Five steps means six degrees of separation; an expression that has become famous since Milgram's experiment.

For the messages that were successfully delivered, the number of steps it took is only an upper bound of the shortest possible path, since individuals do not possess knowledge of the whole social graph and hence their choice of the next step is not necessarily optimal. However, one can also argue that the more steps a message is forwarded through, the higher the probability that one of the individuals involved will decline to participate and will not forward the letter. Hence, some of the messages would actually have reached their target in a higher number of steps, which would increase the mean number of hops required to successfully deliver a letter. Given these considerations, how close is Milgram's result to the actual distance? As we will see in the following, it is in fact quite close.

1.2.2 From six to five degrees of separation

Nowadays, large-scale data related to social connections is available from online social networks such as Facebook. The availability of this data allows researchers to study questions such as that addressed by Milgram on unprecedented scales and allows them to measure the average number of hops between two individuals precisely.

In [32], the authors analyzed the entire graph of friendship relations of the online social network Facebook in May 2011. The network analyzed consists of 721 million active user accounts. They found that the average number of hops is around four. Hence, instead of Milgram's six degrees of separation, Facebook's social graph suggests only five degrees of separation.

More recently, in [33], researchers at Facebook measured the average number of hops between two individuals in Facebook's social graph again, which at that time[4] contained approximately twice the number of users as in the previous study. Remarkably, they found that the average number of hops has decreased. Hence, we are witnessing a shrinking average shortest pathlength. This type of behavior is predicted by our model presented in chapter 2 of the part titled *Evolution and ecology of the digital world*.

1.2.3 Why did the experiment work?

In the light of the findings from the analysis of Facebook, the result Milgram obtained is remarkable. Whereas knowledge of the social graph of Facebook allows us to compute the shortest number of hops, Milgram's experiment relied exclusively on the capacity of individuals to forward messages efficiently without knowledge of the entire social graph. Milgram's findings suggest that the paths found by the participants in his study were close to optimal. Why is that the case?

The answer to this question is due to the fact that individuals were given certain information about the target, including their geographical location and occupation. As Milgram states:

> *"... each intermediary moves the folder toward the target person. That is, a certain amount of information about the target person –his place of employment, place of residence, schooling, and so forth– is given to the starting subject, and it is on the basis of this information alone that he selects the next recipient of the folder."*
>
> Stanley Milgram [31]

What this means is that individuals can to some extend judge which of their contacts is closer to the target. The meaning of closer is obvious in a geographical sense; they can choose to forward the message to the contact who lives (or works) closest to the target. A chain of such forwarding hops presented in the original paper [31] confirms this trend: the source was located in Omaha, Nebraska, and forwarded the message to an individual in Council Bluffs, who then forwarded the message to someone in Belmont, Massachusetts, who then forwarded it to the town of the target, Sharon, Massachusetts. Through each of these steps, the geographical distance to the target is reduced. However, not only geographical information was used to forward the messages, and the notion of proximity is more abstract in other contexts or domains. Knowing that the target is a judge will have the effect that an individual would preferentially forward the message to someone who they deem

[4] As of 2016.

might know a judge, for example a lawyer. In an abstract sense, a lawyer is close to a judge in the occupation domain, or layer in the language of networks. Hence, Milgram's experiment was successful because individuals were able to select good candidates in a multilayer environment with only local knowledge.

We discuss the generalized notion of proximity in terms of multilayer networks embedded in a geometric space with application to the problem of routing messages relying only on local knowledge in part *Geometry of multiplex networks*.

1.3 Complex systems, complex networks, and beyond

1.3.1 More is different: what is a complex system?

There are in fact numerous definitions of complex systems. Whereas there is surely some redundancy among them, there is no single accepted definition; and some definitions or interpretations differ significantly. Here, I present some definitions that I consider to be especially useful, interesting, or historically relevant.

> *"Roughly, by a complex system I mean one made up of a large number of parts that interact in a nonsimple way. In such systems, the whole is more than the sum of the parts, not in an ultimate, metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole."*
>
> Herbert A. Simon [34]

Thus, according to Herbert A. Simon, complex systems are those in which the interaction of a large number of components yields new phenomena that are not trivial consequences of the parts and their interaction patterns. This process has been referred to as *emergence*; roughly the notion that "the whole is more than the sum of its parts". Phillip Warren Anderson pointed out the following in his famous paper entitled "More is different" [35]:

> *"The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. [...] The behavior of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other."*
>
> Philip Warren Anderson [35]

The interaction of many parts of a system can often –but not necessarily– be represented as a network, i.e. a collection of nodes and a set of edges which indicate connections or interactions between nodes. Along this line of reasoning, Hiroki Sayama [4] defines a complex system as follows:

> *"Complex systems are networks made of a number of components that interact with each other, typically in a nonlinear fashion. Complex systems may arise and evolve through self-organization, such that they are neither completely regular nor completely random, permitting the development of emergent behavior at macroscopic scales."*
>
> Hiroki Sayama [4]

The Encyclopedia Britannica [36] provides a concise definition of complexity as the scientific theory of complex systems. In particular, it states:

> *"Complexity, a scientific theory which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain."*
>
> Encyclopedia Britannica [36]

These different definitions agree on the following three necessary features: i) complex systems are made up of a large number of components; ii) the components interact in a nonlinear way (otherwise, the superposition principle would impede the following); and iii) the system as a whole exhibits behavior that is not observed in the interaction of the parts in isolation.

Complex systems and complexity science span across many disciplines, topics, and applications. In [4], Hiroki Sayama presents an organizational map of complex systems grouped by topics, as shown in Fig. 1.1. The topics spread over several disciplines, having implications for physics, social science, computer science, biology, and chemistry.

1.3.2 Complicated vs. complex: what is not a complex system?

From the definitions above, it is clear that a system that is not made up of many components or which is linear is not a complex system. Is a car a complex system? It is made up of many components (which are again usually made up of many subcomponents), as a whole it performs a function (transportation), which the individual parts cannot. It is, however, not a complex system. It is a system usually referred to as a complicated system. The function of a car as a whole *is exactly the sum of the functions of its parts.* The pistons within the cylinders contract and expand, and their movement is transmitted to the wheels. Each part has its own specialized function; and the whole is the sum of those functions. In a complex system, the organization of the whole is encoded in the interaction of the usually similar parts. A flock of birds is a complex system. It organizes in a way that is not observed when we only look at the interaction between two birds; and each part is similar to the others and interchangeable with them. If one bird dies, another can take its position in the flock. When it comes to a car, if a cylinder fails, it cannot be replaced by the wheel. The interchangeability of its parts can make complex systems highly robust and resilient to failures. This property played a central role in the design of the decentralized architecture of the Internet, as considered in section 1.1.1.

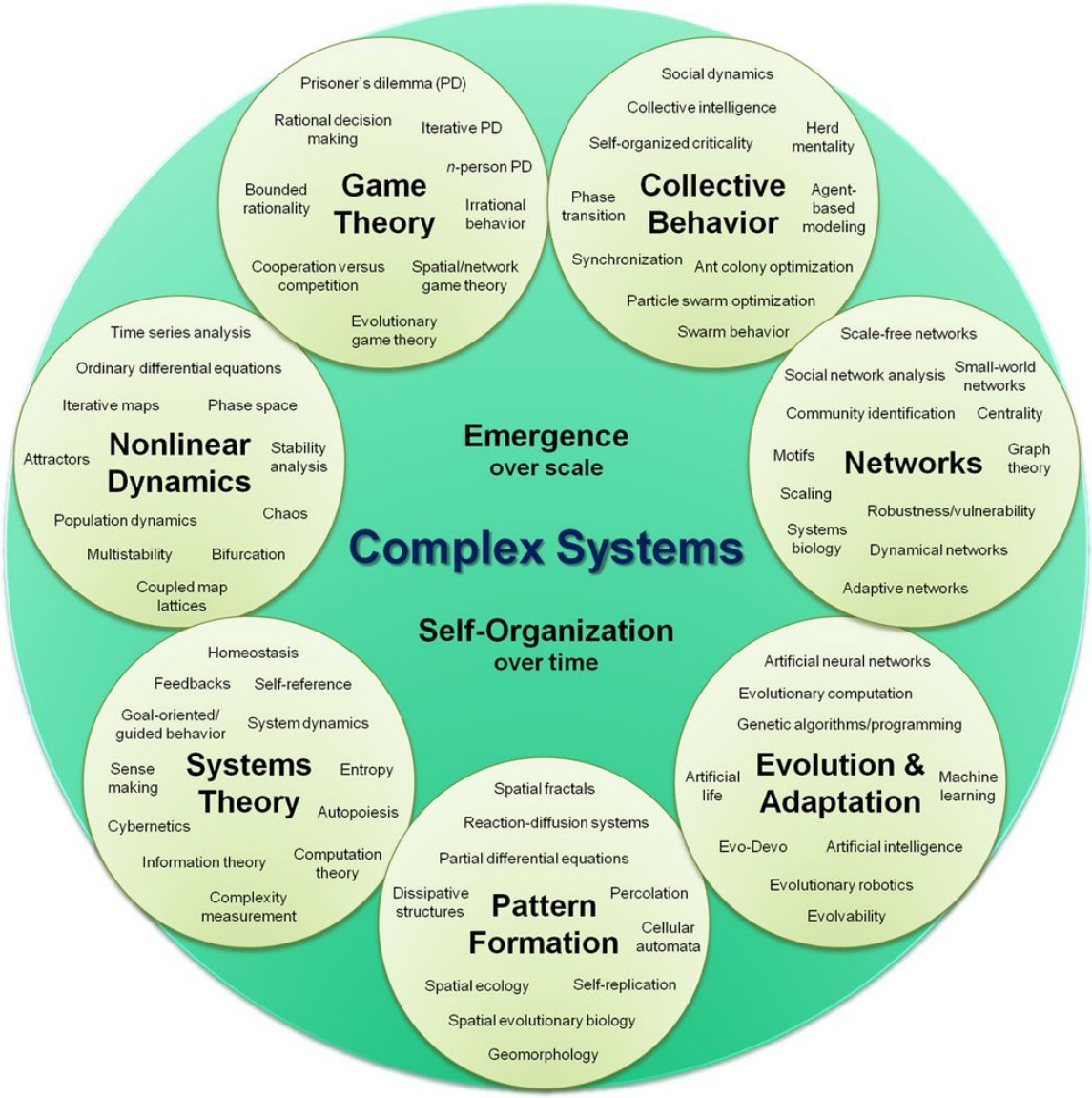

Figure 1.1: "Complex systems organizational map" by Hiroki Sayama, D.Sc. - Created by Hiroki Sayama, D.Sc., Collective Dynamics of Complex Systems (CoCo) Research Group at Binghamton University, State University of New York. Licensed under CC BY-SA 3.0 via Commons - https://commons.wikimedia.org/wiki/File:Complex_systems_organizational_map.jpg

1.3.3 Complex networks and beyond

Complex networks are networks with non-trivial topological properties[5]. These properties usually include a heterogeneous distribution of node degrees (the number of connections each node has; for example, friendship relations in a social network), the small world property (the average fewest hops from some source to a target scales with the logarithm of the network size) and a large clustering coefficient (the probability that two neighbors of a given node are themselves connected).

Examples for complex networks can be found in many different areas, disciplines, and applications: social networks (see Fig. 1.2), communication networks, neuronal networks, power grids, the Internet, and trade networks, to mention just a few.

Real networks are often not isolated entities, but instead form interacting parts of larger and more complex systems. An example for such a system is given in [37], where the authors consider the interaction between the Italian power grid and related Internet servers. The authors show that this interconnectedness leads to new phenomena; in particular, the emergence of a potential failure cascade also referred to as a catastrophic failure.

Interacting networks are found in many real-world systems and hence the study of their structure and functionality is of high importance. In this thesis, we focus on one common case of such systems called multiplex networks. Multiplex networks are systems that consist of several layers, and in which a node can be present simultaneously in different layers. Examples of such systems are social networks, where we can have individuals using both Facebook and Twitter, transportation networks, where two cities can be connected via rail and air travel, biological networks, where different layers encode different interactions, such as the structural and functional brain network, or scientific collaboration networks where different layers represent different scientific disciplines. In this thesis, in the part titled *Geometry of multiplex networks*, we discuss many of these examples in detail.

A further example of a type of interacting networks that has recently received a lot of attention are networks of networks [38]. These are systems where the interaction between different networks is provided by a network itself. In this thesis, we deal with a network of networks in chapter 4 in the part titled *Evolution and ecology of the digital world*. In particular, we deal with online social networks in different countries that interact globally. This interaction is provided by a network that encodes the cultural and social proximity between countries. This allows us to model the interaction between local online social networks and a globally operating one such as Facebook.

[5] Network topology refers to the structural properties of a network, i.e. the arrangement of edges between nodes.

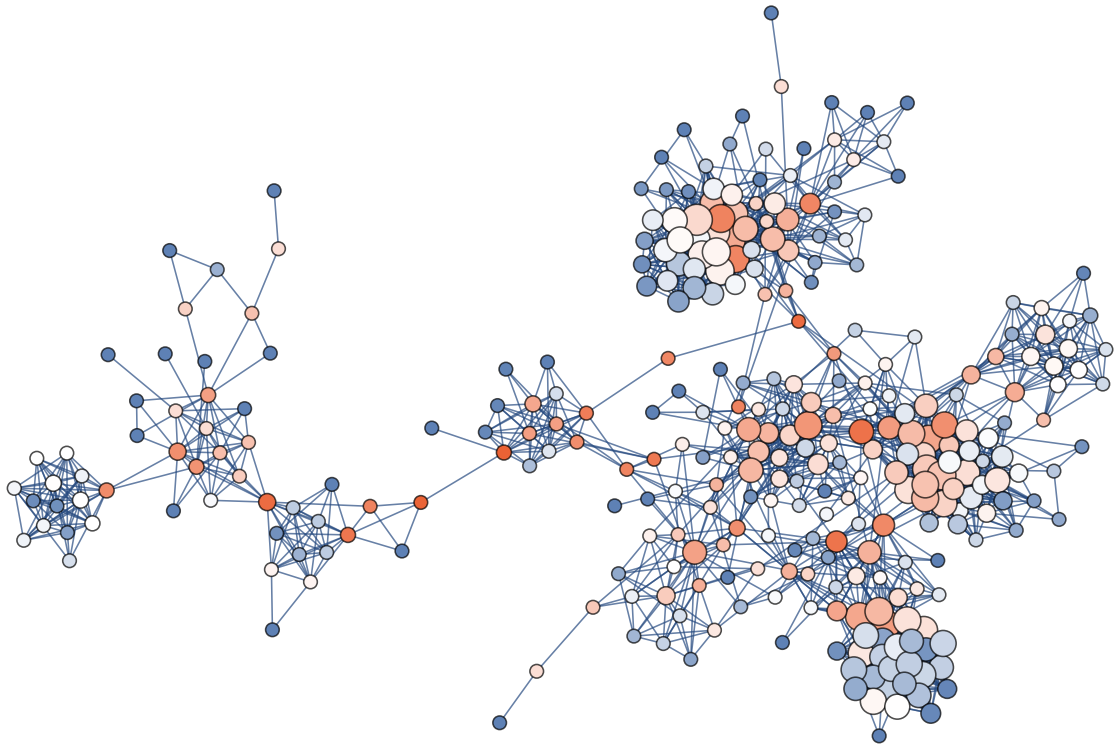

Figure 1.2: Relationships between my friends on Facebook. A link exists if two friends of mine are themselves friends and their profile settings allow me to access their friend list. Only the largest connected component is shown, i.e. the largest subgraph in which there exists a path between each pair of nodes that belongs to that subgraph. The color code represents the betweenness centrality [5], which is a measure of the number of shortest paths that pass through a certain node (blue is low, red is high). Node sizes represent degrees above some threshold.

Evolution and ecology of the digital world

2 Topological evolution of isolated online social networks

"...when it comes to finding out about new jobs - or, for that matter, new information, or new ideas - weak ties are always more important than strong ties."
Mark Granovetter [39]

This chapter was – with some small changes – published in "Physical Review X" by the American Physical Society under the title "Evolution of the Digital Society Reveals Balance between Viral and Mass Media Influence" [40]. A preprint version is available at [41]. Most figures are identical to the preprint version.

2.1 Introduction

The rapid growth of online social networks (OSNs), like Twitter or Facebook, is reshaping the social landscape and changing the way humans interact on a world-wide scale. Needless to say, social networks existed well before OSNs were even invented. However, OSNs offer us the unprecedented opportunity to map social interactions at a scale that was unattainable before the digital era. This has transformed OSNs into huge sociological laboratories, boosting social sciences up to the level of experimental sciences. There is, however, an important difference between traditional social networks and OSNs. Technology-mediated social interactions constitute accelerating phenomena already observed in conventional social networks. Nevertheless, in the case of OSNs these take place faster and on a world-wide scale. This is already changing the way companies try to promote or sell their products with viral marketing campaigns [42–45], the way influential people, e. g. politicians, interact with their followers in Twitter [46–48], or the way people self-organize and cooperate in protest movements [49, 50], crowdfunding [51], etc.

Nonetheless, this socio-technological revolution must come along with the development of new technologies able to sustain its growth. At this respect, one important issue concerns the design of the basic architecture of current OSNs. The best way to solve the scalability limitations of OSNs is to replace their centralized architecture by a fully decentralized one [12]. This can address privacy considerations and improve service scalability, performance, and fault-tolerance in the presence of an expanding base of users and applications. However, to accomplish this program successfully, it is necessary to take into account the

social structure of the underlying society and how it interacts with the system, a task that involves network, computer, and social sciences.

As a matter of fact, we already have a fairly good knowledge on the topological properties of the "social graph" among users of OSNs [32, 52, 53]. Indeed, large datasets of OSNs have allowed researchers to characterize their topology and to validate many principles from the social sciences, like the *"six degrees of separation"* [31, 54–57] by S. Milgram or the *"strength of weak ties"* by M. S. Granovetter [58–63]. However, these results concern static snapshots of the system and, thus, offer little insights into the fundamental mechanisms leading to the evolution of OSNs. Such insights can only be obtained from a detailed analysis of the temporal evolution of topologies of OSNs [64–70]. As we shall show, in the case of real OSNs, such temporal evolution follows an intricate path: an initial phase where the social graph is made of small clusters with increasing diameter and average degree, followed by a dynamical percolation transition and, finally, an epoch of increasing average degree and shrinking diameter, akin to the observations by J. Leskovec, J. Kleinberg, and C. Faloutsos in [71, 72]. Interestingly, this type of history cannot be explained by standard models of growing networks under preferential attachment-like mechanisms, thus calling for new fundamental principles.

In the following, we focus on a particularly important case study, the Slovakian OSN "Pokec" [73]. This network has a combination of unique properties that make it the perfect testbed for our purposes, namely, the following: i) It is the most popular friendship-oriented OSN in the country. ii) Its size represents 25% of the country's population. However, a simple demographic analysis of both the country and Pokec users suggests that, with its current size, it is covering a large fraction of the population susceptible to ever participate in OSNs. As shown in Fig. 2.1, among the age group of about 20 years over 90% of individuals are users of the network. iii) The slovak language is mostly spoken within the country and iv) we can reconstruct its temporal evolution since its birth. As a result, we have the full history of a quasi-isolated OSN whose final state is also a good proxy for the underlying offline friendship network. We hypothesize that such underlying social structure is essential for the emergence of OSNs. Under this premise, we introduce a simple model that incorporates viral dynamics and mass media influence [74] operating on a multiplex network [75] formed by the on- and offline social graphs. The model reproduces very well the topological evolution of the Pokec OSN. Nevertheless, the perfect match is only achieved by introducing into the model the "importance of weak ties" paradigm, yet another empirical evidence in support of Granovetter's theory.

2.2 Evolution of the OSN Pokec: An example of a dynamical percolation transition

2.2.1 The dataset

Pokec is a very large and popular friendship-oriented OSN in Slovakia [73]. By April of 2012, it gathered around 1.6 million users and 30 million directed friendship relations.

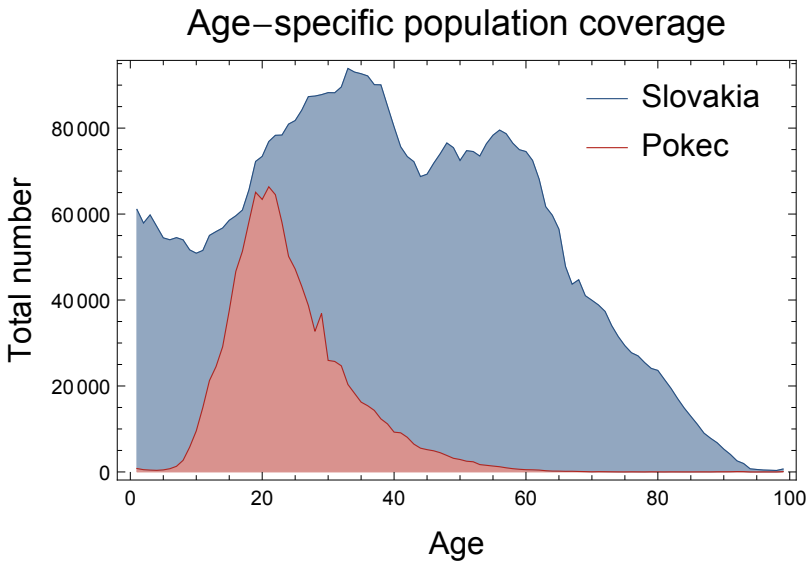

Figure 2.1: Coverage of Pokec users with respect to the whole population of Slovakia.

Nevertheless, not all directed links correspond to a real social tie: Alice might consider Bob as her friend while Bob may not have the same consideration for Alice. Thus, we discard all non-bidirectional links from the original graph and treat those left as undirected edges. The resulting filtered network is composed of 1.2 million users and 8.3 million bidirectional friendship connections. Interestingly, available users profile data contains the registration date of all users. Using this information, we can replay the history of the network topology by assuming that an edge between two users exists at a certain time if both users exist at that time. This approximation is reasonable due to observations from e.g. [65], which suggests that most edges are created in a short time period after the birth of its end nodes.

2.2.2 Dynamical percolation transition

In the inset of Fig. 2.2, we show the temporal evolution of the number of registered users. We clearly appreciate a sustained monotonous increase, suggesting that the popularity of Pokec has not diminished even after the onset of Facebook in the year 2004. This monotonous relation allows us to use the number of current users, $N(t)$, as a measure of time instead of the physical time t. While this is only a rescaling of the temporal axis, it makes the comparison with models easier. The main plot in Fig. 2.2 shows the evolution of the giant connected component (GCC) as a function of the network size. We observe a behavior that could be interpreted as a dynamical phase transition between a phase which consists of small disconnected clusters and a percolated phase where a macroscopic fraction of the network (99% at the end of the evolution) is connected. In percolation theory, the signature of such continuous transition is encoded in the divergence, at the critical point, of the susceptibility χ, defined as a measure of the ensemble fluctuations of the size of the GCC. Unfortunately,

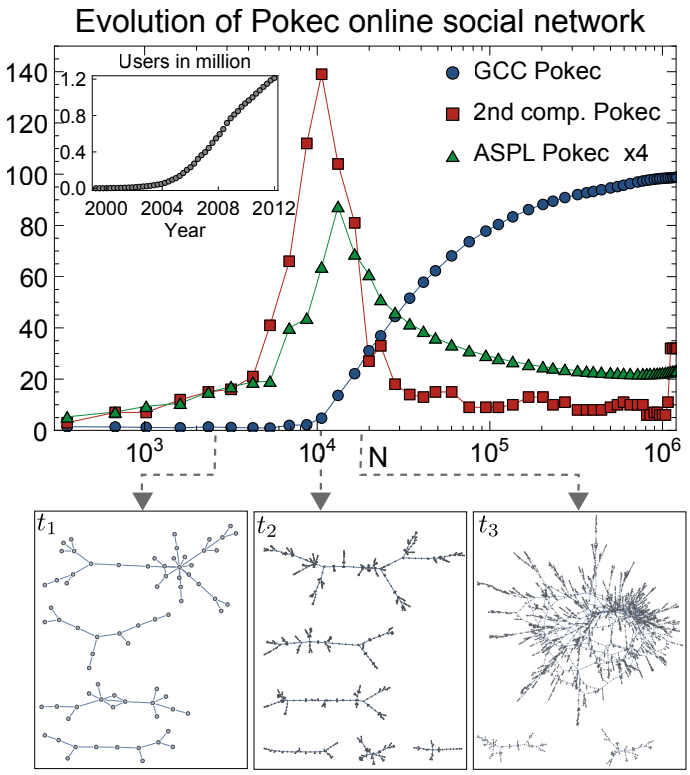

Figure 2.2: Topological evolution of the empiric network. **Top:** The inset shows the evolution of the network size from 1999 to 2012. The main plot shows the relative size of the GCC (blue circles), the size of the second largest component (red squares), and the average shortest path length (green triangles, multiplied by four for better readability). **Bottom:** The largest components of the network are visualized at three different times, before the critical point, t_1, at the critical point, t_2, and after it, t_3.

this technique cannot be applied in our case as the temporal evolution of the Pokec social graph is just one realization of the process. An alternative of the susceptibility is the size of the second largest connected component[1]. This measure is known to diverge at the critical point and, in a single realization of a finite system, it shows a maximum close to the critical percolation point.

Figure 2.2 shows a clear peak in the size of the second largest connected component, indicating that, indeed, we are observing a dynamical phase transition. The distribution of sizes of disconnected components at this point is a power law, another clear indication of the presence of a continuous phase transition, see Fig. 2.3. In Fig. 2.2, we also show the behavior of the average shortest path length (ASPL) within the largest connected component, which shows a quite interesting behavior. During the first stage of the evolution, the ASPL increases with the network size but its growth is not compatible with a logarithmic law,

[1]Alternatively, one can use the average size of components except the GCC, as we show in appendix B.4

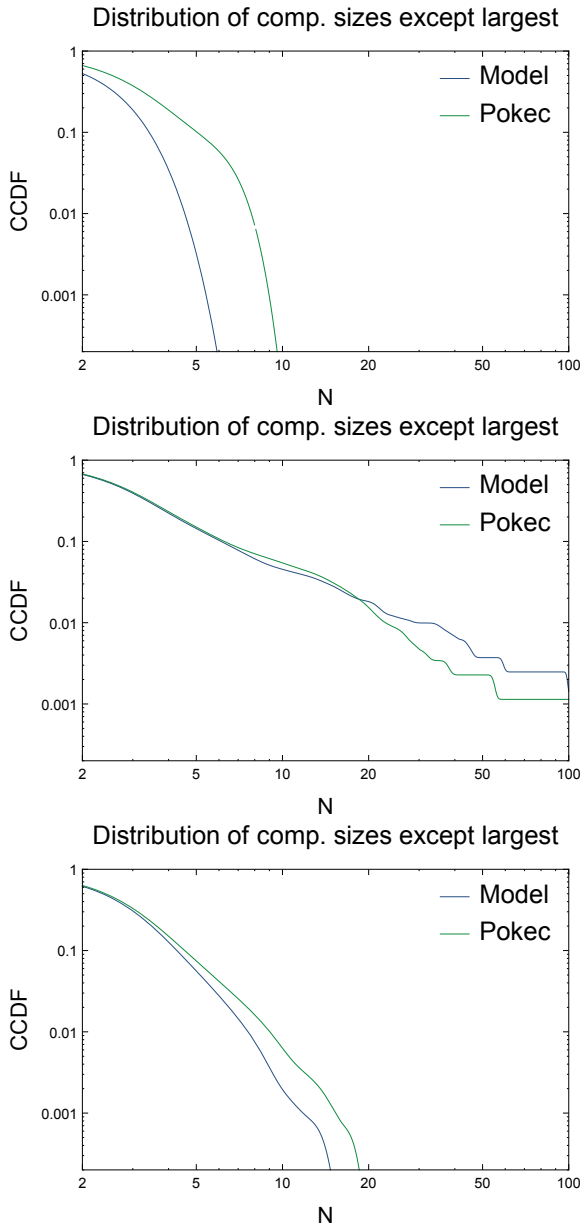

Figure 2.3: Distribution of component sizes except largest component for the basic model. **Top:** $N = 1000$, **middle:** $N = 10000$, **bottom:** $N = 29000$. The center row shows the distribution of sizes near the critical point. One sees that the distribution follows a power-law which is expected at the critical point of a phase transition.

as predicted by the small-world effect. Shortly after the critical point, the ASPL reaches its maximum and then decreases while the size and the average degree (see Fig. 2.10 B on page 29) of the network increase. At long times, the ASPL reaches a value of about five, which is compatible with the small world effect. This behavior was first observed in [71].

2.3 Basic model: balance between virality and mass media influence

2.3.1 Model mechanism

Growing network models based on preferential attachment [76] or similar mechanisms [77–84] were designed to describe systems whose functionality is essentially determined by their large-scale connectivity, for example the physical Internet, power grid networks, biological networks, road networks, etc. Consequently, such models do not show dynamical percolation transitions as they generate a giant connected component from the very beginning of the network evolution, a constant average degree, and an increasing average shortest path length as a function of the number of nodes. In this type of models, the pool of new nodes that are added to the system does not have any previous relation with existing nodes and the connections of newborn nodes to existing nodes are decided exclusively as a function of the current topological state of the network. However, in the case of friendship-orientated networks, there is a pre-existing underlying offline social network conditioning the growth of the OSN. Following this line of reasoning, we conjecture that the observed evolution is the result of a dynamical process that triggers potential users from the offline social network to subscribe to the OSN. Under this assumption, nearly all dynamics able to induce the recruitment of all potential users will yield a dynamical percolation transition. Yet, different dynamics induce different temporal orders in the evolution of OSNs and, therefore, different topological histories.

Following these ideas, we design a two-layer multiplex model for the evolution of OSNs. The upper layer represents the online social network whereas the bottom layer represents the offline social network. The latter can be considered the subgraph of all *a priori* susceptible individuals from the aggregation of all social interactions between individuals. Each individual can be in three different states depending on whether they are or are not enrolled in the OSN. Susceptible individuals are those not in the OSN but that might eventually become members of it. Active individuals belong to the OSN and are actively using it for their social interactions. Passive individuals also belong to the OSN but are not currently using it to interact with their social contacts, see Fig. 2.4. The populations of susceptible, active, and passive individuals are governed by a combination of an epidemic-like process between active users and susceptible or inactive ones and a mass media effect which equally affects the population of susceptible individuals. There are four possible events.

1. *Viral activation:* a susceptible node can be virally activated and added to the OSN by contact to an active neighbor in the traditional offline network. This event happens at rate λ per each active link.

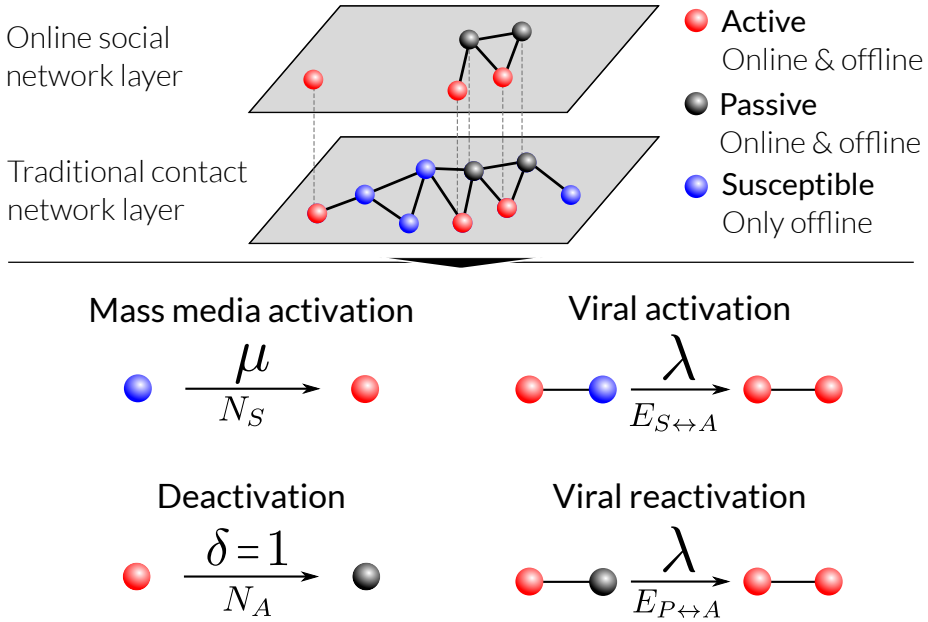

Figure 2.4: Illustration of the two-layer model. The upper layer represents the online social network and contains the active and passive nodes. The bottom layer corresponds to the underlying contact network which contains all nodes. The four dynamical processes are shown in topological illustrations below.

2. *Mass media effect:* each susceptible individual becomes active spontaneously at rate μ and is added to the OSN layer as a response to the visibility of the OSN.

3. *Deactivation:* active users become spontaneously passive at rate δ and no longer trigger viral activations nor reactivate other passive nodes.

4. *Viral reactivation:* at rate λ an active user can reactivate a passive neighbor. The neighbor then becomes active and can trigger both viral activations and viral reactivations.

We can arbitrarily set $\delta = 1$, which defines the timescale in units of the deactivation time. The model is then left with two independent parameters, the virality parameter λ and the mass media parameter μ. Finally, newborn users explore the OSN and connect to all their neighbors in the traditional offline social network that, at the time of the subscription, are either active or passive. In appendix B.6 we conduct an experiment to show that the impact of delayed edge formation can be neglected for reasonable time scales. The evolution of our model is illustrated in Fig. 2.5 for a small underlying network.

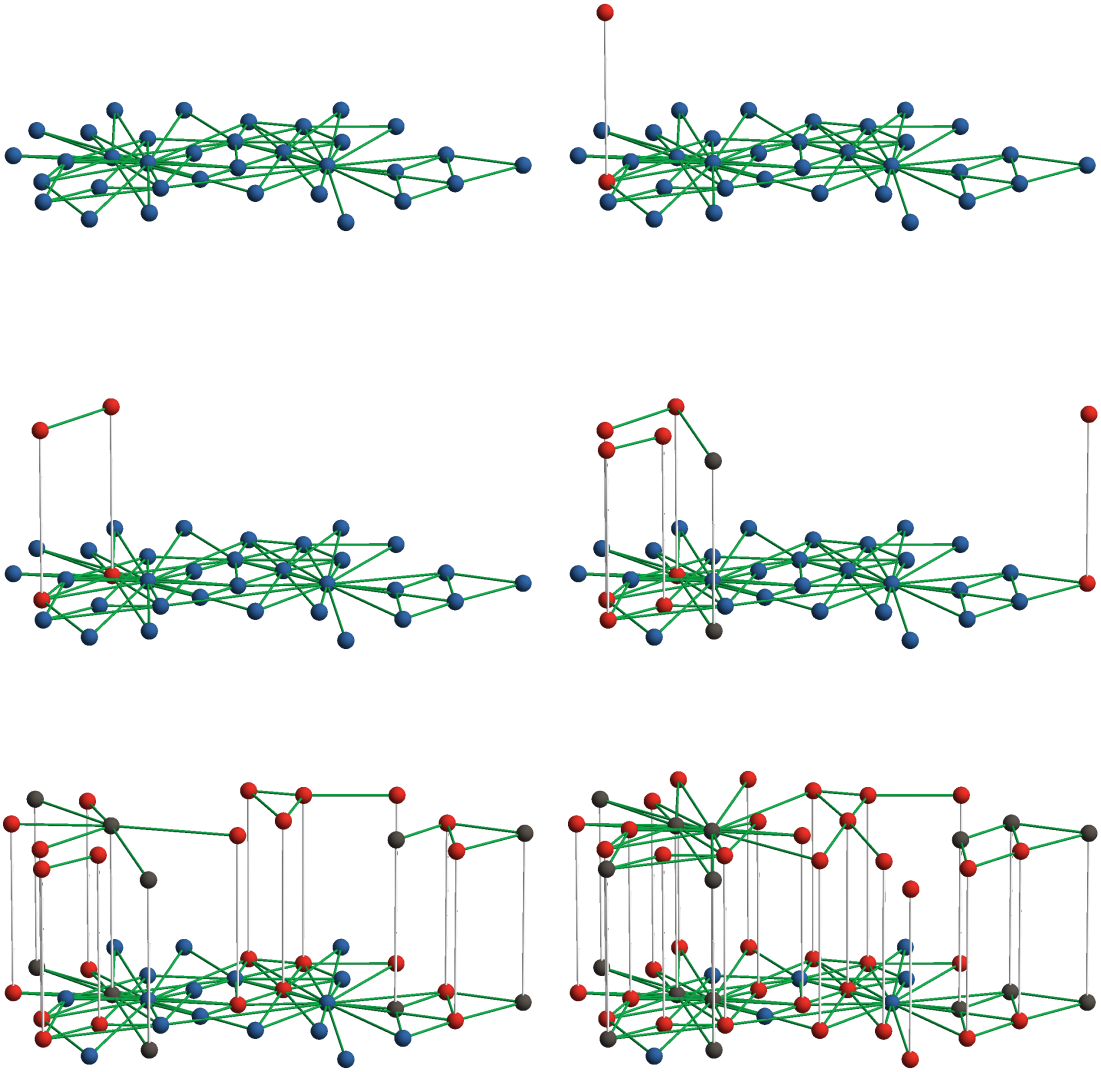

Figure 2.5: Illustration of the model dynamics. Time is increasing from top left to bottom right. Susceptible nodes are marked in blue, red denotes active nodes, and black are passive ones. At the beginning, the OSN layer (top) is empty. The dynamical processes described in the text then lead to the subscription of nodes to the OSN. A video of this sequence is available online [85].

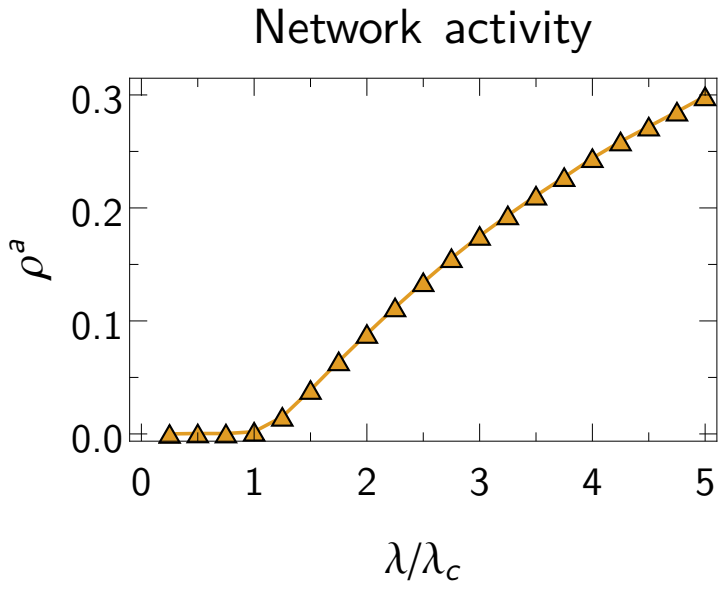

Figure 2.6: Activity as a function of the viral parameter λ for the limit $t \to \infty$ with the final snapshot of Pokec as underlying network. $\lambda_c = 0.02$ is the critical point below which activity is not sustained.

2.4 Relation to SIS model

It is worth to point out that the dynamics between active and passive users is equivalent to the susceptible-infected-susceptible (SIS) epidemic model [86]. As it happens in the SIS model, our model also has a critical rate λ_c below which the number of active users vanishes whereas above it the activity of the OSN is self-sustained, which is shown in Fig. 2.6. This makes the model extremely versatile as it can explain the different fates of OSNs. We also note that a mean field version of this dynamics has been recently and independently proposed to model users' activity of OSNs [87].

2.5 Model validation

The viral activation and the mass media effect play complementary roles in terms of their impact on the topological growth of the network. The mass media effect is very likely to create new components especially at the beginning of the network evolution whereas the viral activation leads to the growth of already existing components. The interplay between these complementary principles is the fingerprint of the evolution of the online social network and the balance between these mechanisms governs the appearance of the phase transition. Interestingly, this allows us to precisely quantify this balance, which we will show and discuss later.

Unfortunately, the rigorous validation of the model requires the precise knowledge of the

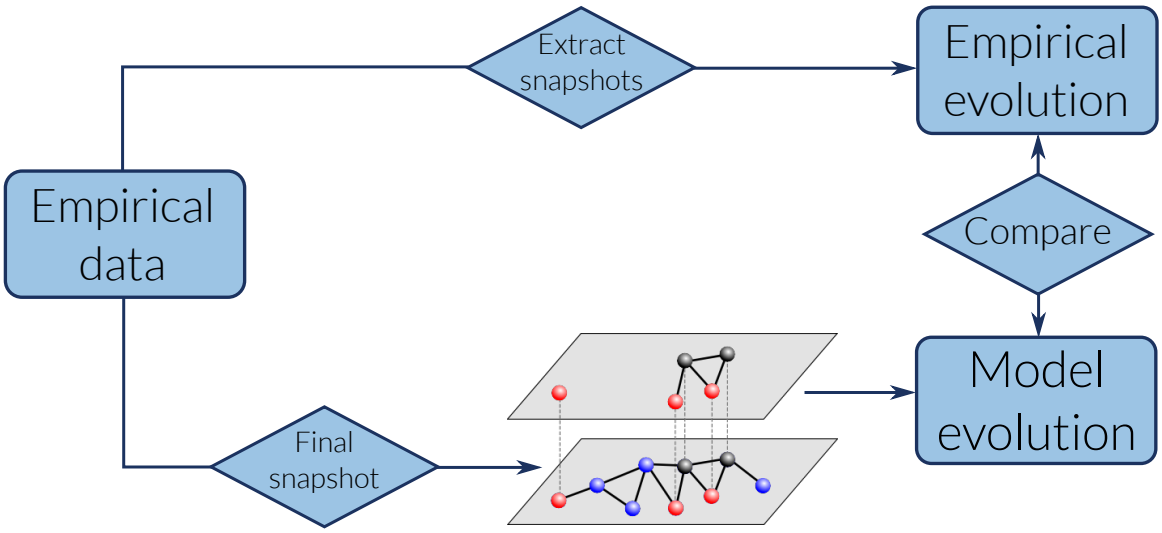

Figure 2.7: Model validation by reproduction of the empirical evolution with the final snapshot of Pokec as the underlying network in our model.

topology of the underlying social network. However, in the particular case of the Pokec network, its large coverage among the subgraph of potential users (see Fig. 2.1) suggests that we can consider the final snapshot of the Pokec OSN as a good proxy for the real underlying social network. As shown in Fig. 2.7, we take the final snapshot of the empirical network as the underlying layer in our model. Following this approach, we perform extensive numerical simulations of our model and compare the resulting evolution with the one we observe in the Pokec OSN. Of course, the real evolution of the Pokec network is still ongoing and, thus, we expect this approximation to fail as we approach the final size of the network. In particular, we do not expect the model to reproduce the network growth in physical time because as the model approaches the size of the empirical network a saturation process aparently slows down the dynamics. We deal with this problem by using the network size instead of physical time as the measure of the course of the evolution, which allows us to compare the topology of the model and the empirical network despite its ongoing evolution. This corresponds to dynamically rescaling λ, μ and δ in the same way at each timestep. Hence, the transformation preserves the fraction λ/μ which we discuss in terms of the balance between virality and mass media influence in the following.

2.5.1 Quantifying the balance between virality and mass media influence

The results of our model show the emergence of a dynamical phase transition from a disconnected to a connected state, in agreement with the empirical observations described earlier. The position of the critical point is related to the parameters λ and μ. For fixed λ, increasing μ leads to the creation of more new disconnected components while the rate at which they are merged by the viral mechanism is kept constant, hence the phase tran-

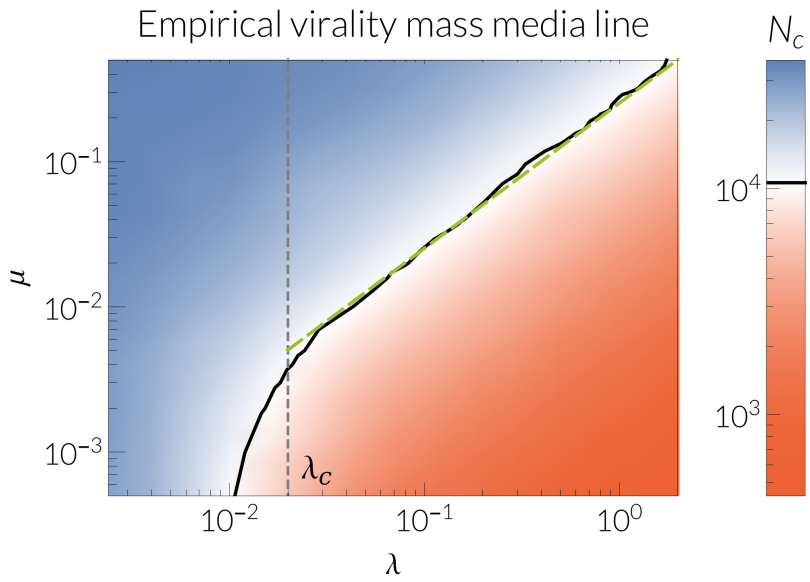

Figure 2.8: The color density plot represents the network size at the critical point for the respective parameters λ and μ. For constant λ, an increase of μ leads to a later transition (blue area) whereas for constant μ one observes an earlier transition (red area) for higher values of λ. The solid black line indicates the virality mass media line corresponding to the critical size of Pokec ($N_c^P = 10600$). The dashed green line shows a linear fit in the region above the sustained activity threshold $\lambda_c = 0.02$ (see Fig. 2.6) corresponding to $\mu(\lambda) = (0.25 \pm 0.01)\lambda$.

sition occurs at larger system sizes. Higher λ however increases the system's tendency to connect previously disconnected components, which leads to an earlier transition. We take advantage of the uniqueness of the critical point to adjust the parameters of the model by matching the network size of the model and the empirical network at the transition point. To this end, we compute the critical size for different values of the parameters λ and μ, as shown in Fig. 2.8. In the empirical network, the phase transition occurs at $N_c^P = 10600$, which is represented by the black line in the plot. The green dashed line shows a linear fit according to

$$\mu(\lambda) = (0.25 \pm 0.01)\lambda\,. \tag{2.1}$$

The virality mass media line given by Eq. (2.1) quantifies the balance between the importance of the viral effect and the mass media effect for the evolution of the network. At the light of this result, we conclude that the viral effect is about four times stronger than the mass media effect. In other words, in the particular case of the Pokec OSN, it is four times more likely to subscribe to the network as a result of the invitation of one active friend than as the result of the information about the network available through the mass media. However, Eq. (2.1) only holds above a critical value of the virality parameter $\lambda > \lambda_c$, which corresponds to the critical threshold for the self-sustained activity of the network. This

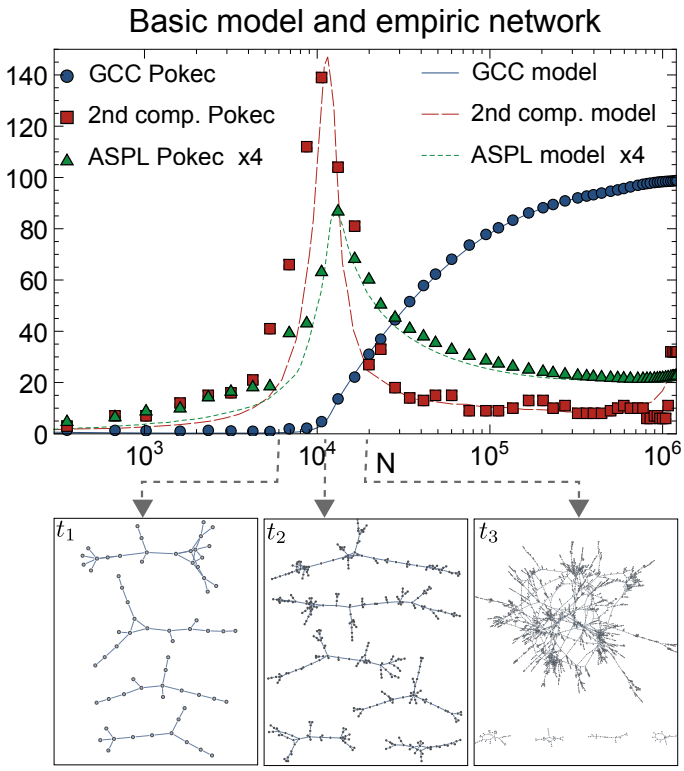

Figure 2.9: Comparison of model and Pokec network evolution. **Top:** The symbols represent the empirical data, whereas the solid lines correspond to the results from the model averaged over 100 realizations with $\lambda = 0.03$ and $\mu = 0.008$. Points correspond to the empiric network and lines represent the results from the model. **Bottom:** Snapshots of the topology of the model at different times similar to Fig. 2.2.

threshold is at $\lambda_c \approx 0.02$. Below this limit, the virality mass media line bends downwards and, in the limit of $\lambda \to 0$, it is not possible to match the position of the critical size (see appendix B.1). This implies that both virality and mass media influence are necessary and complementary mechanisms to explain the topological evolution of OSNs.

2.5.2 Evolution of the network topology: comparison with data

In the active phase, that is $\lambda > \lambda_c$ as shown in Fig. 2.6, the effect of changing the value of λ is very mild if the relation Eq. (2.1) is preserved. In our case, we choose the value of λ that best reproduces the evolution of the number of disconnected components (Fig. 2.10 A) and obtain the corresponding μ from Eq. (2.1). We then compare the results of the model with the empirical evolution of the Pokec OSN in Fig. 2.9. Interestingly, our two parameters model is able to reproduce the entire evolution of the network with an impressive precision for all measured global topological properties, such as the size of the giant component, the

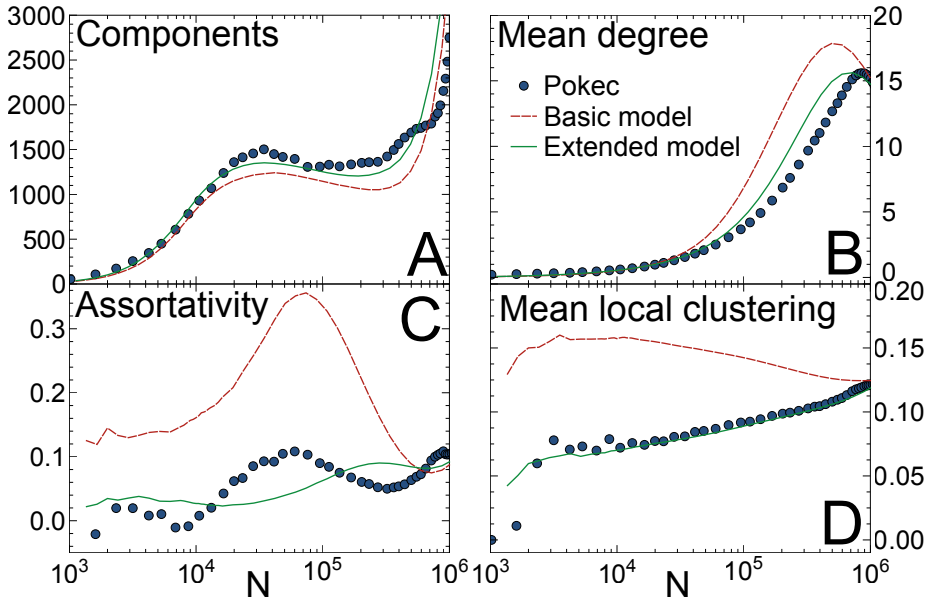

Figure 2.10: Topological evolution of the empiric network (blue circles), the basic model (red dashed lines), and the extended model (green lines). **A:** Evolution of the number of components of size $s > 1$. **B:** The evolution of the mean degree shows a densification of the network. **C:** The assortativity coefficient as defined in [5]. **D:** The evolution of the mean local clustering coefficient (of nodes with $k > 1$) exhibits an essential difference between the basic model (red dashed line) and the extended model with $\eta = -0.65$ (green line).

ASPL, and the size of the second largest connected component, see Fig. 2.9. However, the model is not able to reproduce the temporal trends of local quantities like the mean local clustering and assortativity coefficients, as shown in Figs. 2.10 C and D. The clustering coefficient of the Pokec OSN steadily increases since the beginning of the evolution whereas the model exhibits first a sudden increase followed by a decreasing clustering coefficient. The assortativity coefficient fluctuates both in the model and in the empiric network, although its value in the model is about three times higher. This disagreement suggests that a local mechanism must be incorporated to reproduce simultaneously the global and local evolution of the network topology. In the next section, we present an extended version of our model which takes into account the overlap of each node's neighborhood, with interesting implications concerning the *"strength of weak ties"* paradigm.

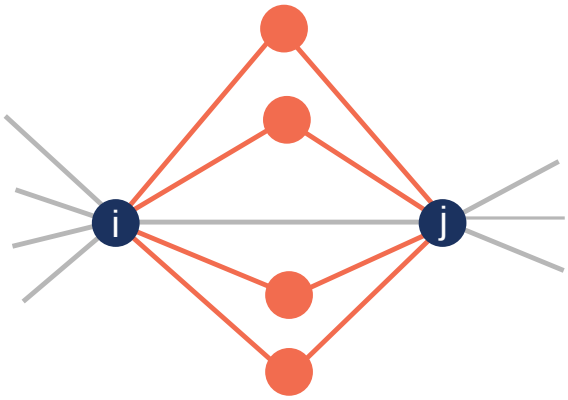

Figure 2.11: Illustration of the multiplicity m_{ij} in Eq. (2.2). The edge between i and j participates in four triangles, which are marked in orange. Hence, $m_{ij} = 4$.

2.6 Extended model: the strength of social ties

2.6.1 Overlapping neighborhood and strength of social ties

The viral activation mechanism of our model is completely blind to the network topology, that is, active users try to "infect" all their neighbors with the same probability. As a consequence, the model performs well at reproducing the evolution of the global topological quantities but it fails at reproducing trends in local quantities, like the clustering coefficient. However, according to Granovetter [58], the diffusion of information through a social tie is different depending on whether the tie is "strong" or "weak". Following Granovetter's idea, we use the overlap of two individuals' friendship network as a measure of the strength of their tie [58]. In particular, given an edge connecting users i and j, we define its social strength as

$$s_{ij} \equiv (m_{ij} + 1), \tag{2.2}$$

where m_{ij} counts the number of triangles going through the edge or, equivalently, the number of common neighbors of the two users. m_{ij} is called the multiplicity of the edge $i \leftrightarrow j$, which is illustrated in Fig. 2.11.

Our previous model can now be easily extended to account for the strength of social ties. We assume that viral activation and reactivation through the edge $i \leftrightarrow j$ is given by

$$\lambda_{ij} = \lambda \frac{s_{ij}^{\eta}}{\langle s^{\eta} \rangle}, \tag{2.3}$$

where $\langle \cdot \rangle$ denotes the global average over the whole network. In this way, the parameter λ has the same interpretation as in the basic model. The transmissibility-strength coefficient η represents the relationship between the viral transmissibility and the strength of the social tie. For $\eta > 0$, viral transmissibility is proportional to the strength of social ties, which puts special emphasis on the strong ties for viral spreading. Instead, for $\eta < 0$, high

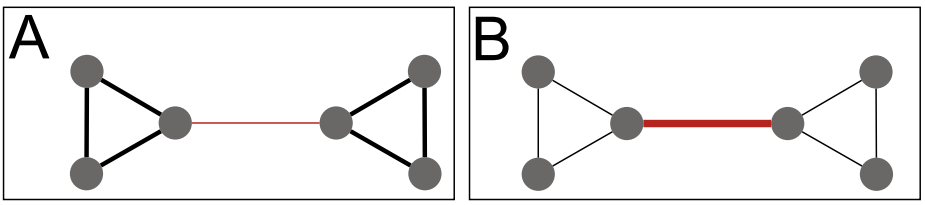

Figure 2.12: **A:** Illustration of the strength of social ties defined by Eq. (2.2). **B:** Illustration of viral transmissibilities for negative η defined by Eq. (2.3).

viral transmissibilities are assigned to edges with low multiplicities, which tend to act like connectors between different clustered groups (see Fig. 2.12). In the case of $\eta = 0$, we have $\lambda_{ij} = \lambda$ and we recover the basic model discussed in the previous section.

2.6.2 Quantifying the transmissibility-strength relationship

We quantify the transmissibility-strength coefficient η by comparing the evolution of the mean local clustering coefficient in the Pokec OSN with results from the extended model. Figure 2.10 D shows that the clustering coefficient of the Pokec OSN grows approximately linearly with the logarithm of the network size. Thus, we interpolate the evolution of the clustering coefficient of our model for different values of η and compare the obtained slopes with the empirical one, as shown in Fig. 2.13. The extended model exhibits an increasing clustering coefficient for $\eta < -0.2$ and the best match with the Pokec OSN is achieved at the value of $\eta = -0.65$ (see Fig. 2.13), which is, remarkably, a negative value. This is yet another empirical proof of Granovetter's theory on the importance of weak ties in processes of diffusion of information in social networks [58]. An alternative empirical validation of the same principle was provided in [63], where it was found that the probability of accepting an invitation to join an OSN is not proportional to the number of social contacts of the invited individual but to the number of different social contexts –the structural diversity– within the individual's life. Notice that a similar effect is achieved in our model when the exponent η is negative. Indeed, suppose that a user has a subset of k contacts forming a "strong" context, that is, these k contacts are all connected among them. This implies that the multiplicity of each link between our user and his/her k contacts is $m = k - 1$. Suppose now that $\eta = -1$, then the aggregated infectivity according to Eq. (2.3) is proportional to $k/(m + 1)$ =constant, which is precisely the main result in [63].

2.6.3 Topological evolution of the extended model

The introduction of weighted transmissibilities in our model does not affect significantly the evolution of the global topological properties. The evolution of the GCC, the size of the second largest connected component, and the average shortest pathlength exhibit a similar trend as in the basic model, as shown in Fig. 2.14.

For the empirical transmissibility, $\eta = -0.65$, the virality mass media line behaves like in

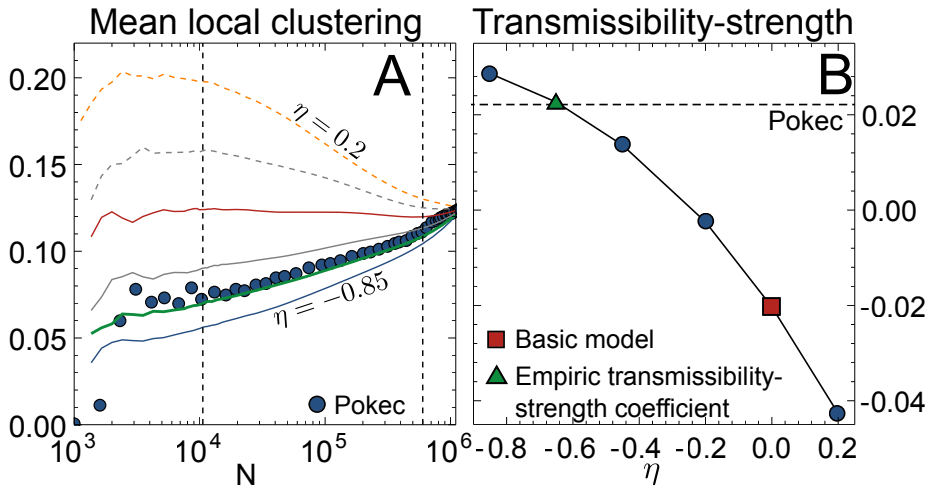

Figure 2.13: **A:** Evolution of the mean local clustering coefficient for different values of η from -0.85 to 0.2. The lines correspond to model results for different η and the symbols represent the evolution from the empiric network. We set $\lambda = 0.03$ and $\mu = 0.006$. **B:** Slopes of the evolution of the mean local clustering coefficient for different values of the transmissibility-strength coefficient. The dashed line represents the slope of the clustering evolution of the Pokec network. The empirical transmissibility-strength coefficient $\eta = -0.65$ is given by the intersection of this line with the curve representing the model results.

the basic model with the difference that now the relation between λ and μ is

$$\mu(\lambda) \approx (0.21 \pm 0.01)\lambda \tag{2.4}$$

as shown in Fig. 2.15.

In Fig. 2.10, we show results for the number of components, the mean degree, the assortativity coefficient, and the mean local clustering, which are all in very good agreement with their empirical counterparts.

To conclude, the strength of social ties has to be taken into account to correctly describe the evolution of the local topology. Interestingly, however, its impact on the global organization of the network is rather small.

2.7 Summary: Topological evolution of isolated online social networks

Comprehensive datasets on the evolution of OSNs offer us the opportunity to determine the principal mechanisms involved in social contagion and online activity of individuals. At this respect, the OSN Pokec, with its peculiar evolution and being almost isolated, is particularly appropriate. Interestingly, the evolution of Pokec's topology is characterized by a dynamical percolation transition, a rather peculiar behavior in real evolving networks. We have shown

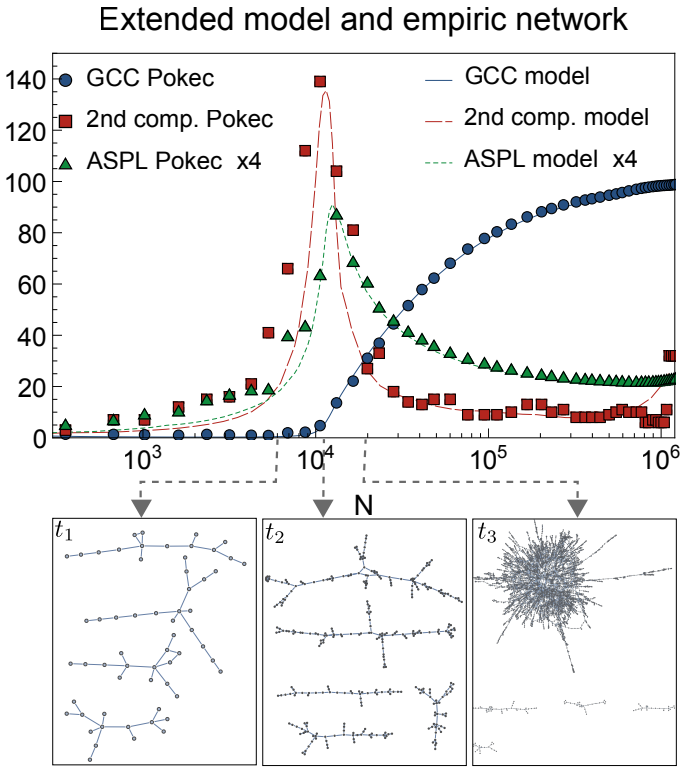

Figure 2.14: GCC (blue), size of second largest component (red), and average shortest path length (green) for the extended model for the parameters $\eta = -0.65$, $\lambda = 0.03$, and $\mu = 0.006$.

that this anomalous topological evolution can be explained very precisely on a quantitative level by a two-layer model, which accounts for the underlying real social structure, combined with two main mechanisms. First, a viral effect, responsible for the social contagion of new users and, second, a mass media effect, leading to random subscriptions of new users. Interestingly, the balance between these two mechanisms is what governs the topological growth of OSNs. In the particular case of the OSN Pokec, the quantification of this balance tells us that the viral effect is between four to five times stronger than the mass media effect. This can explain the proliferation of viral marketing campaigns, in detriment of traditional advertising [42]. To our knowledge for the first time a model with only very few parameters yields quantitatively precise insights about the topological formation of OSNs. This makes our model a necessary foundation for the development of next generation online social networking services.

Beyond the global behavior of our basic model, the social neighborhood of individuals has shown to be crucial to explain the evolution of local topological quantities in Pokec. We find that viral transmissibility is inversely proportional to the strength of social ties. This result is particularly interesting as it corroborates recent empirical findings concerning the

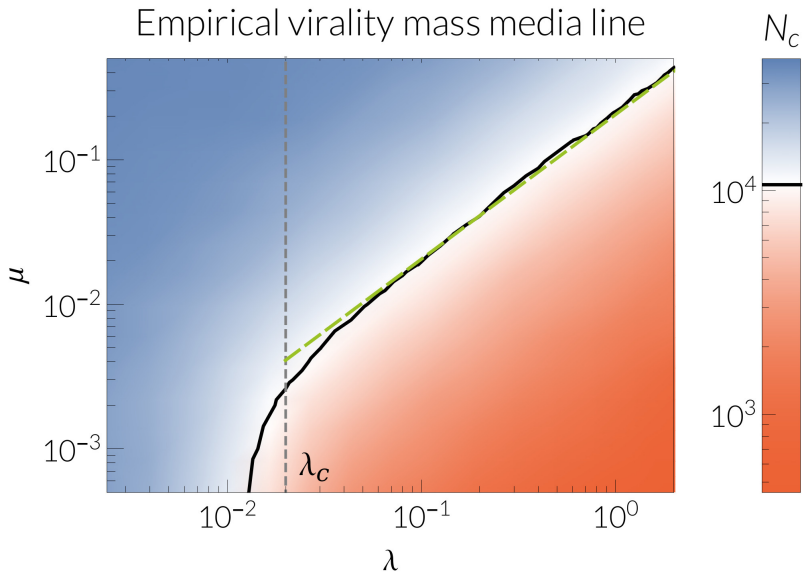

Figure 2.15: Matching of the critical point for the extended model for $\eta = -0.65$. The color density plot represents the network size at the critical point for the respective parameters λ and μ. For constant λ, an increase of μ leads to a later transition (blue area) whereas for constant μ one observes an earlier transition (red area) for higher values of λ. The solid black line indicates the virality mass media line corresponding to the critical size of Pokec ($N_c^P = 10600$). The dashed green line shows a linear fit corresponding to $\mu(\lambda) = (0.21 \pm 0.01)\lambda$.

role of "structural diversity" on social contagion processes by analyzing email invitations from Facebook users [63]. However, our model allows us to identify and quantify this effect exclusively from –and hence its impact on– the topological evolution of the OSN. Alongside with Granovetter's conclusion about the importance of weak ties for individual success, our results give rise to the interpretation that OSNs evolve in a way to improve the possibilities for individual success. This might constitute an important reason for the huge popularity of OSNs.

Our findings here suggest interesting future research lines. Indeed, the particular OSN analyzed here is a quasi-isolated system and, thus, allows us to gauge the fundamental mechanisms at play in the evolution of OSNs. However, in a general situation, an entire ecosystem of OSNs operate simultaneously, competing for the same users, which now become a scarce resource. In the following chapter, we will introduce competition among OSNs into our model which opens the possibility to develop an ecological theory of the digital world.

3 Digital ecology

"Just as a monopoly in economy is a threat to free markets, the lack of digital diversity poses a threat to the freedom of information."
ECOLOGY 2.0 [88]

This chapter was – with some small changes – published in "Scientific Reports" by Nature Publishing Group under the title "Digital Ecology: Coexistence and Domination among Interacting Networks" [89]. A preprint version is available at [90]. Most figures are identical to the preprint version.

3.1 Introduction

In the previous chapter we have studied the topological evolution of online social networks in an isolated environment and were able to identify the main mechanisms responsible for the evolution of quasi-isolated OSNs. However, most OSNs operate on a worldwide scale and are in constant competition for users' attention with numerous other services; a fact that makes it extremely challenging to model them. This competitive environment leads to the extinction of some networks, while others persist. This phenomenon suggests an ecological perspective on the interaction of multiple OSNs, from which networks are considered to form a complex digital ecosystem of interacting species that compete for the same resource: users' networking time.

In standard ecology theory, Gause's law of competitive exclusion [91] states that under constant environmental conditions, two species in competition for the same resource cannot coexist. This is because even the slightest advantage of one species over the others is amplified and eventually leads to the domination of this species. This mechanism is often referred to as rich-get-richer. Competitive exclusion is predicted by many theoretical models [92]. However, many observations of natural ecosystems seem to contradict Gause's law, as in the case of the famous plankton paradox [93]. Attempts to solve such paradoxes include the assumption of different roles (competition–colonization trade off [94, 95]), the increase of the dimension of the systems, the inclusion of further species properties, etc. (see [96] and references within). However, such models allow for an unlimited number of coexisting species, which thereby creates a new paradox. Indeed, real ecosystems usually consist of a moderate number of coexisting species. In this chapter, we show that the coexistence of networks that are in competition for the same resource, namely our society's networking time, is possible. Furthermore, our work predicts that the most probable outcome is the coexistence of a moderate number of networks.

Recent work [97] showed that the competition between Facebook and its competitors such as MySpace in the mid 2000s led to the extinction of Facebook's competitors and its own prevalence. However, the current existence of a large number of OSNs [98] suggests that the coexistence of multiple networks is indeed possible. This could be explained by analogy with the competition–colonization trade-off mentioned earlier, if we assume that different networks compete for different peer groups and hence one network can persist in each of these groups. Although the existence of different peer groups is certainly the case in reality[1], our aim in this chapter is to introduce a general and concise theory for competition between identical networks that are in competition for the same set of potential users that allows either the coexistence of any number of networks or the domination of a single network.

We show that the coexistence of competing networks can indeed be modeled by allowing for the interplay of two very common mechanisms: preferential attachment and diminishing returns. Preferential attachment [76–84] is a fundamental principle that can be applied to growing networks and which states that newborn nodes are most likely to connect to the more popular nodes; this leads to a rich-get-richer effect. The principle of diminishing returns—or diminishing marginal returns—is widely used in economic theories and refers to the negative curvature of production functions. For example, suppose that sowing 1 kilogram of seed in a certain place yields a crop of one ton. However, 2 kilograms of seed produces only 1.5 tons of crop; and 3 kilograms of seed produces 1.75 tons of crop. Thus, the marginal return per increment of seed diminishes with the increasing amount of seed used.

In this chapter, we demonstrate the following three points. First, multiple networks can coexist in a certain parameter region due to the interplay of a rich-get-richer mechanism and diminishing returns in the dynamics of the evolution of the networks. Second, we are most likely to observe only a moderate number of coexisting services. Finally, third, the influence of the mass media controls the observed diversity in the digital ecosystem.

3.2 From quasi-isolated online social networks to interacting networks

Suppose now that, instead of a single network as in the previous chapter, there are n_l networks competing for the same set of potential users. Each user can be active or passive in several networks simultaneously, as represented in Fig. 3.1, such that the long-term evolution of the fraction of active users in each layer determines the fate of the system: either several networks coexist or only a single network prevails. The first key point in the generalization of the model introduced in the previous chapter concerns the role of the viral parameter λ. This parameter is a proxy for users' engagement in online activities, such as inviting their friends to participate in the network, generating or forwarding content, etc. However, such activities require users to spend a given amount of their time on them and their time is, obviously, bounded. This implies that when users are simultaneously active

[1] We will discuss the influence of different sets of susceptible individuals in the following chapter.

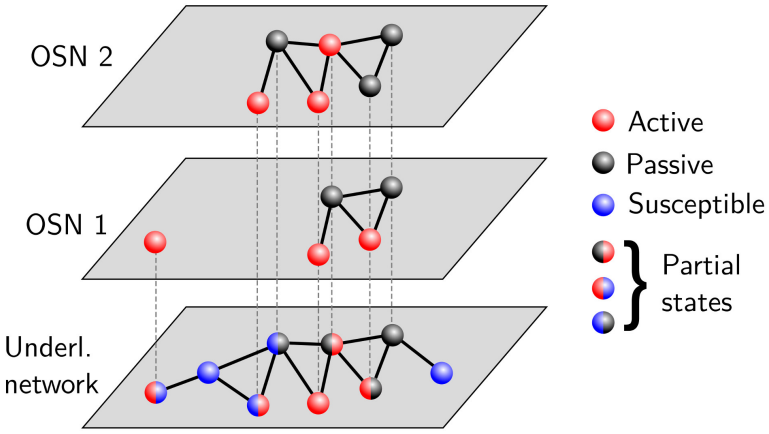

Figure 3.1: Multiplex layout of two online social network layers. The bottom layer represents the underlying social structure and the remaining layers represent each OSN.

in two or more different networks or services, they are forced to decide the amount of time they devote to each of them. We model this effect by assuming that the viral parameter for each layer is $\lambda_i = \lambda\omega_i$, where ω_i a set of normalized weights (that is, $\sum_{i=1}^{n_l}\omega_i = 1$) that quantify users' engagement with each OSN. In this way, $\sum_{i=1}^{n_l}\lambda_i = \lambda$ is a conserved quantity related to the physical and cognitive limitations of users. The second key point in our generalization concerns the dependence of the share, λ_i, of the total amount of virality for individual networks on the state of activity of the whole system, which is defined by the vector: $\boldsymbol{\rho}^a = (\rho_1^a, \rho_2^a, \cdots, \rho_{n_l}^a)^{\mathrm{T}}$. We assume that the weights ω_i are functions of $\boldsymbol{\rho}^a$ that obey the following two conditions:

1. *Symmetry:* All networks are considered intrinsically equal. Therefore, the weight functions must satisfy the symmetry conditions:

$$\omega_i(\cdot, \rho_i^a, \cdots, \rho_j^a, \cdot) = \omega_j(\cdot, \rho_j^a, \cdots, \rho_i^a, \cdot)$$
$$\omega_i(\cdot, \rho_j^a, \cdots, \rho_k^a, \cdot) = \omega_i(\cdot, \rho_k^a, \cdots, \rho_j^a, \cdot),$$

   for any i, j, and k. This implies that when the fraction of active users is the same in all of them, the viral parameters λ_i must also be equal in each network and, therefore, $\omega_i = 1/n_l\ \forall i$.

2. *Preferential attachment:* We assume that users are in general more likely to subscribe to and participate in more active networks. Hence, the weight of a given network i must be a monotonically increasing function of ρ_i^a. Following the same line of reasoning, we also assume that a network with zero activity is not functional, so that $\omega_i(\rho_i^a = 0) = 0$.

Finally, consistent with our findings in the previous chapter, we assume a linear relation between μ_i and λ_i,

$$\mu_i = \frac{\lambda_i}{\nu} = \frac{\lambda\omega_i(\boldsymbol{\rho}^a)}{\nu}, \tag{3.1}$$

where ν denotes the relative strength of the viral effect with respect to the influence of mass media, for which we found in the previous chapter that $\nu \approx 4 \sim 5$).

These conditions can be interpreted as coarse-grained preferential attachment in the bipartite graph consisting of users and networks. Users are in general more prone to connect to networks which exhibit higher activity and, once active in more than one network, they are also more inclined to engage with the most active one more often. Notice that we are introducing a feedback loop between the global dynamics of the system and the microscopic parameters λ_i. We are thus assuming that users are, somehow, able to sense the global activity of the system. This can be achieved in practice as a combination of the amounts of information that users receive from: the network itself [99–101], global media, the traditional offline social network, etc. Although preferential attachment induces a rich-get-richer mechanism, in what follows we show that the interplay of this mechanism with the dynamics of the networks leads to the emergence of stable coexistence of multiple networks across a certain parameter region.

3.3 Meanfield approximation

The effects of complex topologies on epidemic-like spreading processes are well understood nowadays and cannot be ignored. However, the dynamics of our model is rich and complex enough on its own to be analyzed in isolation. Therefore, in this section we perform a meanfield analysis which provides important insight into the emergence and stability of a state of coexistence of multiple networks. In particular, we replace the real social contact network by a fully mixed population with an average number of contacts per user $\langle k \rangle$. Section 3.4 contains numerical simulations of our dynamics using a real social network [40, 73]. We can confirm in advance that the general picture drawn in this section is also observed in the real system.

3.3.1 One-dimensional dynamics

For one network, the system is described by the following meanfield equations

$$\begin{aligned}
\dot{\rho}^a &= \underbrace{\lambda \langle k \rangle \rho^s \rho^a}_{\text{Viral activations}} + \underbrace{\lambda \langle k \rangle \rho^a \rho^p}_{\text{Reactivations}} + \underbrace{\mu \rho^s}_{\text{Mass media}} - \underbrace{\delta \rho^a}_{\text{Deactivations}} \\
\dot{\rho}^p &= - \underbrace{\lambda \langle k \rangle \rho^p \rho^a}_{\text{Reactivations}} + \underbrace{\delta \rho^a}_{\text{Deactivations}} \\
\dot{\rho}^s &= \underbrace{-\mu \rho^s}_{\text{Mass media}} - \underbrace{\lambda \langle k \rangle \rho^s \rho^a}_{\text{Viral activations}} .
\end{aligned} \tag{3.2}$$

The nontrivial steady-state solution is $\rho^s = 0$ and $\rho^a = 1 - \delta/\lambda\langle k \rangle$, which is stable only when $\lambda \geq \delta/\langle k \rangle \equiv \lambda_c^1$. This defines the critical value of λ below which activity is not possible, even in a single network. In the following, we assume that $\lambda > \lambda_c^1$ so that, even if coexistence is not possible, at least one network is always able to survive. Likewise, we

also fix the timescale of our model by setting $\delta = 1$ from now on as we did in the previous chapter.

3.3.2 Multiple competing networks

In the case of an arbitrary number of OSNs, the system is characterized by the fraction of active and passive users in each layer, ρ_i^a and ρ_i^p, and the fraction of individuals in the traditional offline social network that are susceptible to subscription in network i: ρ_i^s. We assume that the densities of active/passive/susceptible nodes are not correlated between different OSNs. Thus, the evolution equations in the meanfield approximation for the i-th layer are

$$\begin{aligned} \dot{\rho}_i^a &= \rho_i^a \Big\{ \lambda \langle k \rangle \, \omega_i(\boldsymbol{\rho}^a) \left[1 - \rho_i^a \right] - 1 \Big\} + \frac{\lambda}{\nu} \omega_i(\boldsymbol{\rho}^a) \rho_i^s \\ \dot{\rho}_i^s &= -\frac{\lambda}{\nu} \omega_i(\boldsymbol{\rho}^a) \rho_i^s \Big\{ 1 + \nu \langle k \rangle \, \rho_i^a \Big\} , \end{aligned} \tag{3.3}$$

where we have used $\rho_i^p = 1 - \rho_i^a - \rho_i^s$. Note that the coupling between different OSNs is encoded in the weights, $\omega_i(\boldsymbol{\rho}^a)$.

3.3.3 Stationary solution

The stationary solution of Eqs. (3.3) that corresponds to the complete coexistence of all the n_l networks is given by

$$\rho_i^{a*} = 1 - \frac{n_l}{\lambda \langle k \rangle} \text{ and } \rho_i^{s*} = 0, \ \forall i \tag{3.4}$$

for $\lambda > \lambda_c^{n_l} \equiv \frac{n_l}{\langle k \rangle}$. This again defines a critical threshold for λ below which complete coexistence is impossible. At the opposite extreme, the stationary solution for the prevalence of just one single network, j, is

$$\begin{aligned} \rho_j^a = 1 - \frac{1}{\lambda \langle k \rangle} \text{ and } \rho_j^s = 0, \\ \rho_i^a = 0 \text{ and } \rho_i^s = \text{const} \ \ \forall i \neq j, \end{aligned} \tag{3.5}$$

for $\lambda > \lambda_c^1$. It is easy to see that this last solution is always stable when $\lambda > \lambda_c^1$. However, the stability of the coexistence solution depends, in general, on the particular form of the weights $\omega_i(\boldsymbol{\rho}^a)$.

3.3.4 Stability

To investigate the stability of the stationary solution, we analyze the Jacobian matrix of the dynamical system defined in Eqs. (3.3) whose entries correspond to the following derivatives

$$\begin{aligned}
\left.\frac{\partial \dot{\rho}_i^a}{\partial \rho_j^a}\right|_{\boldsymbol{\rho}^*} &= \rho_i^{a*} n_l \left.\frac{\partial \omega_i(\boldsymbol{\rho}^a)}{\partial \rho_j^a}\right|_{\boldsymbol{\rho}^*} - \delta_{ij} \frac{\rho_i^{a*}}{1-\rho_i^{a*}} \\
\left.\frac{\partial \dot{\rho}_i^a}{\partial \rho_j^s}\right|_{\boldsymbol{\rho}^*} &= \delta_{ij} \frac{\lambda}{\nu n_l} \\
\left.\frac{\partial \dot{\rho}_i^s}{\partial \rho_j^a}\right|_{\boldsymbol{\rho}^*} &= 0 \\
\left.\frac{\partial \dot{\rho}_i^s}{\partial \rho_j^s}\right|_{\boldsymbol{\rho}^*} &= -\delta_{ij} \left[\frac{\lambda}{\nu n_l} + \frac{\rho_i^{a*}}{1-\rho_i^{a*}}\right] .
\end{aligned} \tag{3.6}$$

The Jacobian matrix (of dimension $2n_l \times 2n_l$) can be written as

$$\mathcal{J} = \begin{pmatrix} \mathcal{M}_{a,a} & \mathcal{M}_{a,s} \\ \mathcal{M}_{s,a} & \mathcal{M}_{s,s} \end{pmatrix} \tag{3.7}$$

where $\mathcal{M}_{\eta,\theta}$ represent $n_l \times n_l$ matrices with the following elements

$$\mathcal{M}_{\eta,\theta}(i,j) = \left.\frac{\partial \dot{\rho}_i^\eta}{\partial \rho_j^\theta}\right|_{\boldsymbol{\rho}^*} . \tag{3.8}$$

The matrix $\mathcal{M}_{s,s}$ is a diagonal matrix with all its diagonal equal to $d = -\left[\frac{\lambda}{\nu n_l} + \frac{\rho_i^{a*}}{1-\rho_i^{a*}}\right] < 0$. $\mathcal{M}_{a,s}$ is also a diagonal matrix with all its diagonal elements equal to some value c and finally $\mathcal{M}_{s,a}$ has all its elements equal to 0.

By using the Laplace expansion starting from the $2n_l, 2n_l$ entry and expanding row-wise, one finds after n_l iterations that

$$\det\left[\mathcal{J} - \Lambda \mathcal{I}\right] = (d-\Lambda)^{n_l} \det\left[\mathcal{M}_{a,a} - \Lambda \mathcal{I}\right] , \tag{3.9}$$

which means we have the eigenvalues $\Lambda_{n_l+1,\ldots,2n_l} = d$ with degeneracy n_l and the remaining eigenvalues are those of $\mathcal{M}_{a,a}$. Because $d < 0$, this means that the stability of the coexistence solution is exclusively determined by the dynamics in the limit $\nu \to \infty$, which reduces the dimensionality of the system from $2n_l$ to n_l and decouples the dynamics of ρ_i^a from ρ_i^s.

The matrix $\mathcal{M}_{a,a}$ has the form

$$\mathcal{M}_{a,a} = \begin{pmatrix} \alpha & \beta & \cdots & \beta \\ \beta & \alpha & \cdots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \cdots & \alpha \end{pmatrix} \tag{3.10}$$

and its eigenvalues are $\Lambda_1 = \alpha + (n_l - 1)\beta$ and $\Lambda_{2,\dots,n_l} = (\alpha - \beta)$, which has degeneracy $n_l - 1$. We have

$$\alpha = \rho_i^{a*} n_l \left.\frac{\partial \omega_i(\boldsymbol{\rho}^a)}{\partial \rho_i^a}\right|_{\boldsymbol{\rho}^*} - \frac{\rho_i^{a*}}{1 - \rho_i^{a*}} \tag{3.11}$$

and

$$\beta = \rho_i^{a*} n_l \left.\frac{\partial \omega_i(\boldsymbol{\rho}^a)}{\partial \rho_j^a}\right|_{\boldsymbol{\rho}^*} = -\rho_i^{a*} \frac{n_l}{n_l - 1} \left.\frac{\partial \omega_i(\boldsymbol{\rho}^a)}{\partial \rho_i^a}\right|_{\boldsymbol{\rho}^*} \tag{3.12}$$

(here we used $\sum_j \omega_j = 1$, hence $\frac{\partial \omega_i}{\partial \rho_j^a}|_{i \neq j} = -\frac{1}{n_l - 1}\frac{\partial \omega_i}{\partial \rho_i^a}$). From here, one obtains the result that Λ_1 is always negative and the stability is controlled by the eigenvalues $\Lambda_{2,\dots,n_l}$, which are

$$\Lambda_{2,\dots,n_l} = \alpha - \beta = \left.\frac{\partial \omega_i(\boldsymbol{\rho}^a)}{\partial \rho_i^a}\right|_{\boldsymbol{\rho}^*} \rho_i^{a*} \frac{n_l^2}{n_l - 1} - \frac{\rho_i^{a*}}{1 - \rho_i^{a*}}\,, \tag{3.13}$$

which have to be negative to satisfy stable coexistence. The coexistence state is stable if $\Lambda_{2,\dots,n_l} < 0$, which leads to the condition

$$\phi(\rho_i^{a*}) \equiv \frac{n_l^2}{n_l - 1}(1 - \rho_i^{a*}) \left.\frac{\partial \omega_i(\boldsymbol{\rho}^a)}{\partial \rho_i^a}\right|_{\boldsymbol{\rho}^{a*}} < 1\,. \tag{3.14}$$

3.3.5 Interplay between preferential attachment and diminishing returns

We have evaluated the stability of the system in the coexistence state by analyzing the Jacobian matrix earlier. Here, we show that the emergence of stable coexistence can be understood as the interplay between preferential attachment and diminishing returns. Preferential attachment affords an advantage in terms of respective weight, ω_i, for networks which already exhibit higher activity; inducing a rich-get-richer effect. However, this is damped by the intrinsic dynamics of the system, which exhibits diminishing returns in terms of activity with respect to an enhancement of the corresponding weight ω_i. As long as the preferential attachment mechanism is not strong enough to overcome this damping effect, any perturbation in the density of active nodes near the coexistence point will eventually decline. Hence, the coexistence is stable. From a mathematical point of view, this is equivalent to showing that, at the coexistence point, the function $\phi(\rho_i^{a*})$ is proportional to the dynamical return of the system when network i is perturbed. In other words, if the activity of network i is externally increased by a small amount $\Delta\rho_i^a$, after some relaxation time, the dynamics brings the perturbation to the new value $\Delta\tilde{\rho}_i^a = \phi(\rho_i^{a*})\Delta\rho_i^a$. Coexistence is stable whenever the dynamical perturbation $\Delta\tilde{\rho}_i^a$ is smaller than the external one $\Delta\rho_i^a$, which is explained in the following.

Here, we discuss the response of the system to a small perturbation, $\Delta\rho_i^a$, of the activity of one network. In the limit $n \gg 1$, we can neglect the effect of perturbing the i-th network on the remaining ones. The perturbation induces a shift in the corresponding weight according to

$$\Delta\omega_i \approx \left.\frac{\partial \omega_i}{\partial \rho_i^a}\right|_{\boldsymbol{\rho}^{a*}} \Delta\rho_i^a\,. \tag{3.15}$$

Our initial perturbation triggers the dynamical response $\Delta\tilde{\rho}$ from the system given by

$$\Delta\tilde{\rho}_i^a = \left.\frac{\partial\rho_i^{a*}(\omega_i)}{\partial\omega_i}\right|_{\omega_i=\frac{1}{n_l}} \Delta\omega_i \,, \tag{3.16}$$

where $\rho_i^{a*}(\omega_i) = 1 - 1/\lambda\langle k\rangle\omega_i$. With Eq. (3.15), we obtain

$$\Delta\tilde{\rho}_i^a = n_l(1-\rho_i^{a*}) \left.\frac{\partial\omega_i}{\partial\rho_i^a}\right|_{\boldsymbol{\rho}^{a*}} \Delta\rho_i^a. \tag{3.17}$$

The coexistence solution is stable if the perturbation decreases; this means that the dynamical response $\Delta\tilde{\rho}_i^a$ has to be smaller than the initial perturbation $\Delta\rho_i^a$. Mathematically, this leads to the condition

$$n_l(1-\rho_i^{a*}) \left.\frac{\partial\omega_i}{\partial\rho_i^a}\right|_{\boldsymbol{\rho}^{a*}} < 1 \,, \tag{3.18}$$

which is equivalent to Eq. (3.14) from the previous section in the limit $n_l \gg 1$. The left-hand side of Eq. (3.18) is proportional the ratio between the dynamical response of the system and the initial perturbation. If this ratio is smaller than one, the initial perturbation will decrease and the coexistence state is stable.

3.3.6 Existence of stable coexistence states

It is possible to see that $\phi(\rho_i^{a*})$ diverges at $\rho_i^{a*} = 0$ and is zero when $\rho_i^{a*} = 1$, and thus there is always a value of λ above which the inequality (3.14) is fulfilled. Our assumptions of symmetry and normalization allow us to write

$$\omega_i(\boldsymbol{\rho}^a) = \frac{\psi(\rho_i^a)}{\sum_{j=1}^{n_l}\psi(\rho_j^a)} \tag{3.19}$$

where $\psi(\rho_i^a)$ is an arbitrary monotonically increasing function with $\psi(0) = 0$, which is bounded on the interval $[0,1]$. We have

$$\left.\frac{\partial\omega_i}{\partial\rho_i}\right|_{\boldsymbol{\rho}^{a*}} = \frac{\psi'(\rho_i^{a*})}{\psi(\rho_i^{a*})}\frac{n_l-1}{n_l^2} \,, \tag{3.20}$$

which we can plug into Eq. (3.14) to obtain

$$\phi(\rho_i^{a*}) \equiv (1-\rho_i^{a*})\frac{\psi'(\rho_i^{a*})}{\psi(\rho_i^{a*})} < 1 \,. \tag{3.21}$$

Since $\psi(0) = 0$ and $\psi'(0) \neq \psi(0)$, the left-hand side of Eq. (3.21) diverges for $\rho_i^{a*} \to 0$. Because ψ is bounded, we have

$$\lim_{\rho_i^{a*}\to 1} \phi(\rho_i^{a*}) = 0 \,. \tag{3.22}$$

Therefore, there is always a ρ_i^{a*} (and so a value of λ) for which the inequality (3.21) is fulfilled.

Interestingly, a series of states of partial coexistence exist between the complete coexistence state and the prevalence of a single network, such that only a number $n_c < n_l$ of OSNs coexist simultaneously. The symmetries of the weights $\omega_i(\boldsymbol{\rho}^a)$ imply that any such case is exactly the same as the complete coexistence state if we replace n_l by n_c in Eqs. (3.4) and (3.14). Finally, we recall that the stability of the partial or complete coexistence solutions is independent of the value of ν (see Sec. 3.3.4). Therefore, we can discuss the stability in the limit $\nu \to \infty$, which reduces the dimensionality of the dynamical system.

As mentioned earlier, the symmetry and preferential attachment conditions of the weights $\omega_i(\boldsymbol{\rho}^a)$ combined with the normalization condition imply that, without loss of generality, $\omega_i(\boldsymbol{\rho}^a)$ can be written as

$$\omega_i(\boldsymbol{\rho}^a) = \frac{\psi(\rho_i^a)}{\sum_{j=1}^{n_l} \psi(\rho_j^a)}, \tag{3.23}$$

where ψ can be any monotonically increasing function bounded on $[0, 1]$ with $\psi(0) = 0$. To gain further insight, we consider the following form of function $\psi(\rho_i^a) = [\rho_i^a]^\sigma$. By adjusting a single parameter this form allows us to describe a system between a set of decoupled networks, when $\sigma = 0$, and very strongly coupled ones, when $\sigma \gg 0$. In this particular case, the stability condition of the coexistence state of n_c networks is given by

$$\sigma < \frac{\lambda - \lambda_c^{n_c}}{\lambda_c^{n_c}} \quad \text{with} \quad n_c = 2, \cdots, n_l \,. \tag{3.24}$$

This inequality defines a set of $n_l - 1$ critical lines $\sigma_c(\lambda; n_c)$ in the plane (λ, σ) that separate phases with n_c and $n_c - 1$ maximally coexisting networks. This is illustrated in Fig. 3.2 for the case of $n_l = 5$ competing networks.

However, the stability of the coexistence solution does not guarantee that it is reached from arbitrary initial conditions because, as we show above, there are several other stable fixed points, each with its own basin of attraction. This is illustrated in Fig. 3.3, where we show the vector field in the plane (ρ_1^a, ρ_2^a) for the case of two competing networks in the limit $\nu \to \infty$. For any fixed value of $\lambda > \lambda_c^2$ and $\sigma > \sigma_c(\lambda; 2)$, the coexistence solution is an unstable saddle point. This implies that one of the networks will eventually prevail, independently of the initial conditions (Fig. 3.3 top right). At the critical point $\sigma = \sigma_c(\lambda; 2)$, the system undergoes a subcritical pitchfork bifurcation with the appearance of two unstable saddle points moving away from the (now stable) coexistence solution as σ is decreased (Fig. 3.3 top left and bottom). The subcritical character of the bifurcation is akin to first-order phase transitions. Indeed, an infinitesimal increase in the value of σ near the critical point makes the system jump from stable coexistence to the domination of one of the networks. Decreasing the value of σ afterwards does not bring the system back into the coexistence state, as this type of bifurcation implies a hysteresis effect, as shown in the inset of Fig. 3.3. This behavior is particularly interesting as it implies that digital diversity – once lost – cannot be recovered without fluctuations in the order of the size of the system. The two saddle points that emerge below the critical line determine the basin of attraction

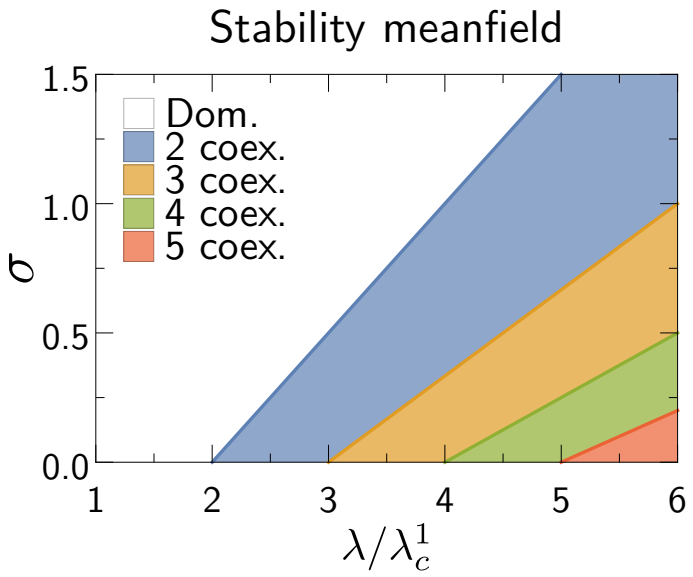

Figure 3.2: Regions of maximal possible coexistence in the meanfield approximation as a function of λ and σ for 5 networks evaluated from Eq. (3.24).

of the coexistence solution. This basin (depicted in blue in the top left plot of Fig. 3.3) is very narrow for low densities of active nodes, as found at the beginning of the evolution. This makes the system sensitive to stochastic fluctuations; a small perturbation of the initial conditions may push the system into a state of domination of one network. We finally note that, in contrast to other nonlinear models of population dynamics, our system does not exhibit limit cycles.

3.4 Numerical simulations

3.4.1 Real-world topology

The analysis presented in the previous section is based on two strong and unrealistic assumptions: the fully mixed hypothesis of the underlying offline social network and the absence of fluctuations in the densities of active users. The first assumption has a strong impact on the value of the critical threshold λ_c^1 and the fraction of active users in a single network when $\lambda > \lambda_c^1$. Fluctuations have an important impact mainly at the beginning of the evolution, when the number of active users is small, which is when the finite system size becomes especially relevant. Such fluctuations can induce the system to change stochastically from one basin of attraction to another, leading the system to different steady states—either coexistence or domination—even if it starts from the same initial configuration with identical parameters. Once the system is in the coexistence state and has approached its full size, the relative importance of fluctuations decreases as the expected time for the system to

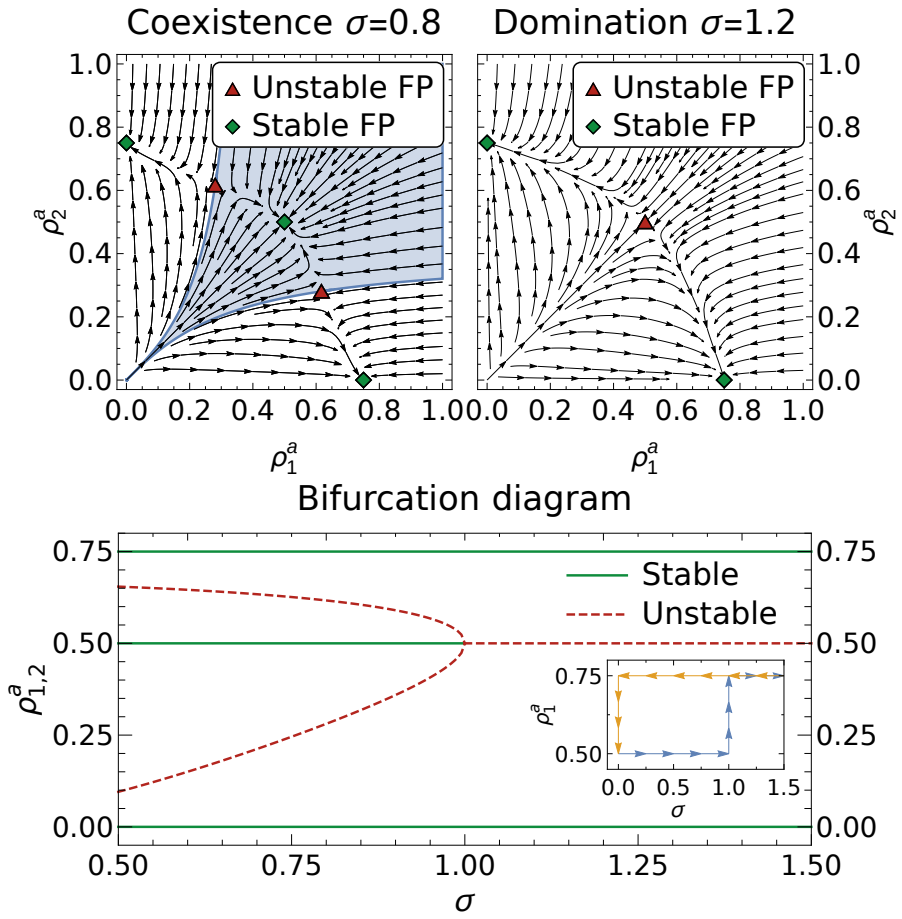

Figure 3.3: Meanfield approximation in the limit $\nu \to \infty$ (this reduces the system dimension from 4 to 2 and allows the diagram to be plotted, see Sec. 3.3.4). **Top:** Left: Stable coexistence solution ($\lambda/\lambda_c^1 = 4$, $\sigma = 0.8$). The basin of attraction for the coexistence solution is marked in blue. Right: Only the domination solution is stable ($\lambda/\lambda_c^1 = 4$, $\sigma = 1.2$). **Bottom:** Bifurcation diagram for two OSN layers showing subcritical pitchfork bifurcation at $\sigma = \sigma_c$ for $\lambda/\lambda_c^1 = 4$. The inset shows the hysteresis induced by this type of bifurcation.

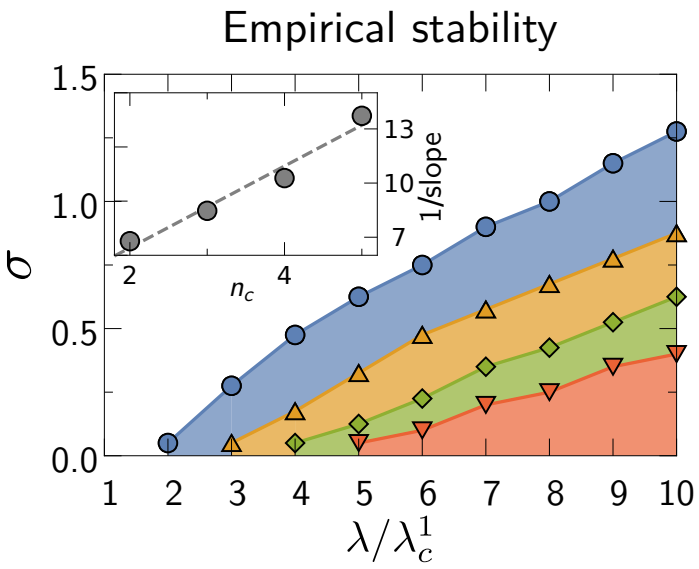

Figure 3.4: Stability regions for the full stochastic model with a real underlying topology. The inset shows n_c versus the inverse slope of linear fits to the respective lines.

jump out of the basin of attraction of the coexistence solution due to fluctuations diverges exponentially with the system size. To understand the effects of the above assumptions within a real scenario, we performed large-scale numerical simulations of our model on a real social network, the Slovakian friendship-oriented OSN Pokec [40, 73] in 2012, which we have analyzed in the previous chapter. We recall that the size of this network (1.2×10^6 users) represents 25% of the population of Slovakia but demographic analysis shows that it covers a much larger fraction of the population susceptible to ever participate in OSNs (see Fig. 2.1 on page 19). This makes Pokec a very good proxy of the underlying social structure.

3.4.2 Empirical stability

We first study the coexistence space in the plane (σ, λ) in the case $n_l = 5$. To do so, for each value of λ and σ we first set the system to the coexistence solution $\boldsymbol{\rho}^{a*}$. We then apply a small positive perturbation to one of the networks $\rho_1^{a*} \to \rho_1^{a*} + \delta\rho_1^a$. The evolution of the system after this perturbation can be used to determine the stability of the coexistence state (see appendix C.1 for details). The results are shown in Fig. 3.4. Even though the position of the critical point of a single network λ_c^1 of the real Pokec network is extremely different from the meanfield prediction, the critical lines as a function of the ratio λ/λ_c^1 follow a linear trend, as in the meanfield prediction. Interestingly, the slopes of these lines (although they are different from those in the meanfield case) scale with n_l in the same way as in the meanfield case (see the inset in Fig. 3.4).

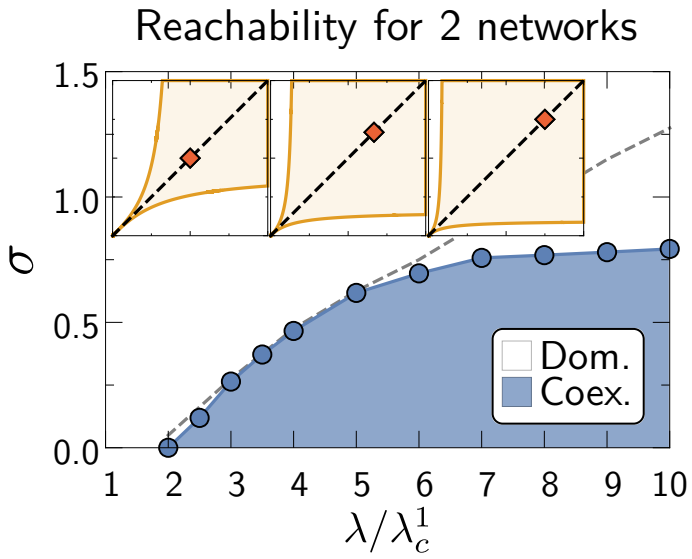

Figure 3.5: The most probable configuration reached from empty initial conditions for two networks. The dashed line corresponds to the empirical stability of the two networks. The insets (x and y axes each denote the activity from 0 to 1) show the basins of attraction in the meanfield approximation for $\sigma = 0.8$ and $\lambda/\lambda_c^1 = 4$ (left), $\lambda/\lambda_c^1 = 6$ (center), and $\lambda/\lambda_c^1 = 8$ (right).

3.4.3 Reachability

The stability of the coexistence solution *per se* does not guarantee that coexistence is reached from any initial configuration. This is particularly relevant when the evolution starts from empty networks, as fluctuations in the number of active users at the beginning of the evolution can induce the system to jump from one basin of attraction to another. Therefore, to determine the effective coexistence space in the plane (σ, λ), we evaluate the probability that a state of coexistence of a certain number of networks is reached when starting from empty networks. In the case of two competing networks, we define the effective critical line $\sigma_c^{eff}(\lambda; 2)$ as the line below which the probability of the two networks reaching coexistence is greater than $1/2$.

Figure 3.5 shows the results of this program for two competing networks and $\nu = 4$. The effective critical line follows the critical line in Fig. 3.4 for low values of λ and saturates at a constant value when $\lambda/\lambda_c^1 \gg 1$. This result can be understood in terms of the shape of the basin of attraction of the coexistence solution near the origin. Indeed, only in this region are fluctuations important enough to make the system change from one basin to the other. As an illustration, in the inset of Fig. 3.5 we show such a basin for $n_l = 2$ and different values of λ in the meanfield approximation. As can be observed, the shape of the basin in the neighborhood of $\rho_{1,2}^a \sim 0$ is almost independent of the value of λ, which explains why the probability of reaching the coexistence state saturates at a constant value.

This saturation effect is similarly observed for systems of more networks, where the ef-

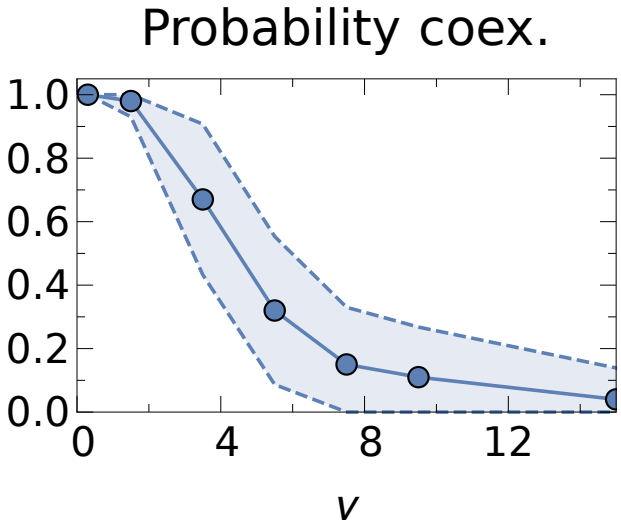

Figure 3.6: Probability of reaching the coexistence state for two networks for different values of ν, for $\lambda/\lambda_c^1 = 6$ and $\sigma = 0.70$. The blue area denotes one standard deviation (from top to bottom).

fective critical lines of higher coexistence states successively saturate at lower values; that is $\sigma_c^{eff}(\infty;2) > \sigma_c^{eff}(\infty;3) > \sigma_c^{eff}(\infty;4)\cdots$, which narrows the effective coexistence region in the plane (λ, σ) for large numbers of networks. This is particularly relevant because, although our theory allows for the coexistence of an arbitrarily large number of networks, the stochastic nature of the dynamics, coupled with the narrow form of the basin of attraction at low densities of active users, makes such coexistence highly improbable. Therefore, our model predicts—even without knowledge of the exact empirical parameters—a moderate number of coexisting networks in a large fraction of the parameter space.

3.4.4 Influence of mass media

The results shown in Fig. 3.5 are obtained for a fixed value of the parameter ν. While this parameter has no influence on the stability of the coexistence solution, and thus no effect on the results shown in Fig. 3.4, it has a strong influence on the probability of reaching coexistence. Indeed, when ν is finite, the last term in Eq. (3.3) acts, at the beginning of the evolution, as a temporal boost that increases the fraction of active users in each network. This mechanism drives the system closer to the coexistence state where its attractor is broader. Figure 3.6 shows the simulation results of the probability of reaching coexistence as a function of ν for two competing networks. For small values of ν, the initial boost is large and the system almost always ends up in the coexistence state. For larger values of ν, the probability decreases significantly. We conclude that a higher boost—hence a smaller value of ν—favors the effective reachability of the coexistence state; whereas a small boost reduces that probability dramatically. Since ν is related to the influence of mass media,

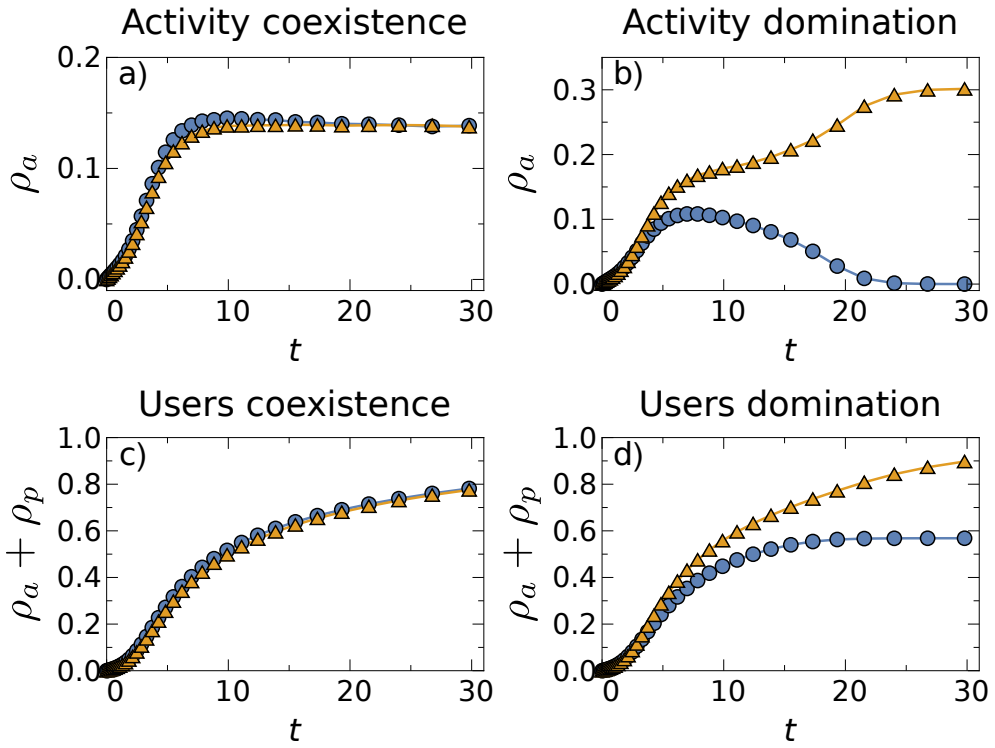

Figure 3.7: Evolution of the fraction of active users (top) and the fraction of total users (bottom) for two competing networks. The first column corresponds to the parameters $\lambda/\lambda_c^1 = 5$, $\sigma = 0.5$, and $\nu = 4$ which lies in the coexistence region. The second column represents the parameters $\lambda/\lambda_c^1 = 5$, $\sigma = 0.75$, and $\nu = 4$, which lies in the dominance region.

these results show that mass media influence plays a crucial role in the diversity of the digital ecosystem.

3.4.5 Temporal evolution

The temporal evolution of the process also shows interesting patterns. Figure 3.7 shows typical realizations of the process below and above the effective critical line in the case of two competing networks. It should be noted that in both cases, during the first stage of the evolution, the two networks acquire a very similar number of active users, making the forecasting of which network will eventually prevail very difficult. In a second stage, the symmetry is broken and one of the networks starts dominating, while the activity of the other declines. This pattern of "rise and fall" has been observed in many real OSNs [87]. In our model, however, such behavior is a consequence of the non-linear coupling between the networks, without the need to introduce an exogenous mechanism to explain it [97]. Meanwhile, the effective critical lines shown in Fig. 3.5 separate regions in a probabilistic way. This implies that in the vicinity of these lines, it is possible to find realizations that, with the same parameters and initial conditions, have opposite fates. This is illustrated in

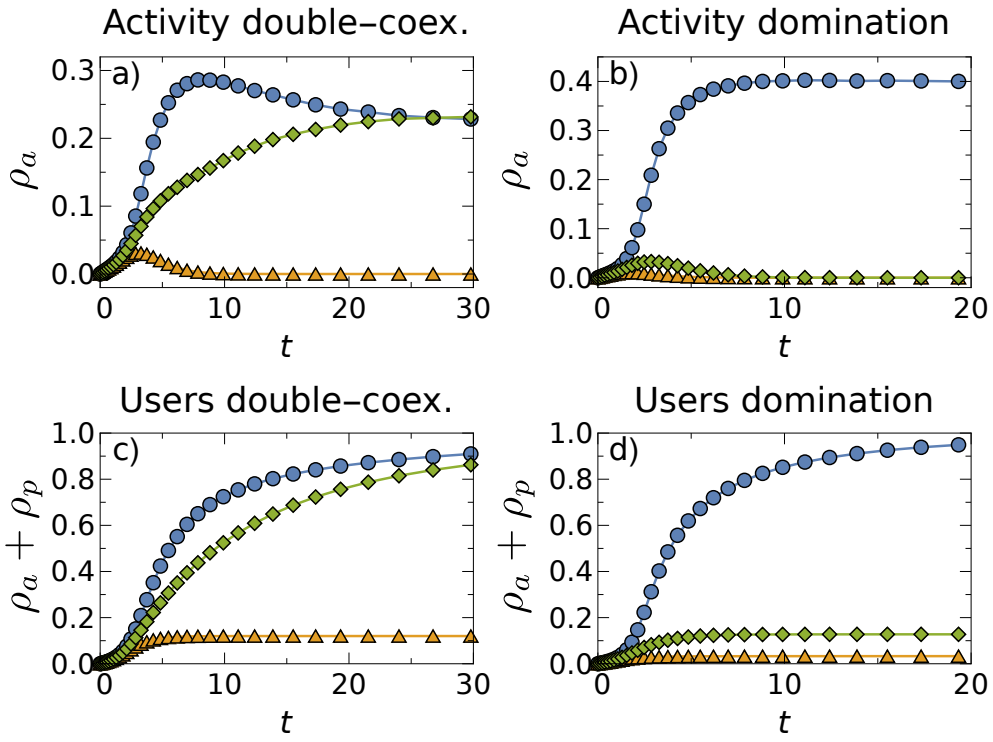

Figure 3.8: Evolution of the fraction of active users (top) and the fraction of total users (bottom) for three competing networks. Both columns correspond to the same parameters $\lambda/\lambda_c^1 = 7.5$, $\sigma = 0.8$, $\nu = 4$, but are different realizations.

Fig. 3.8 where we show two different realizations of three competing networks. In the first column of Fig. 3.8, we show one such realization where two out of three networks coexist and, in the second column, a realization where only one of the three networks prevails.

3.5 Summary: Digital ecology

OSNs constantly compete to attract and retain users' attention. From this point of view, OSNs and other digital services can be understood as forming a complex digital ecosystem of interacting species that compete for the same resource: our networking time. In this chapter, we have introduced a very general and concise theory of such an ecosystem. Akin to standard ecological theories of competing species, the fitness of OSNs increases with their performance following a preferential attachment (or rich-get-richer) mechanism. However, unlike the case of standard ecology, the total fitness of the system is a conserved quantity, which induces diminishing returns in the fitness of each network. Over a range of parameters, the combination of these two mechanisms leads to stable states of coexistence of many networks, in stark contrast to the competitive exclusion principle [91].

However, stable coexistence is only possible across a range of the parameter space, which

is delimited by a critical line. At that critical line the system undergoes a subcritical pitchfork bifurcation akin to a first-order phase transition. Our model thus predicts that a minimal change or perturbation in the interactions between the different networks can have a catastrophic effect on the fate of the system. In any case, due to the stochastic nature of the dynamics and the multitude of fixed points, a stable coexistence solution is not always reached. The probability of reaching such a solution is an indicator of the diversity observed in the digital ecosystem. Interestingly, we find that over a large proportion of the parameter space the most probable outcome is the coexistence of a moderate number of digital services; in agreement with empirical observations. This number is, in general, greatly affected by the magnitude of the mass media influence.

The flexibility of our theory allows us to reproduce, with only three parameters, a large number of possible outcomes that have been observed empirically. In the following chapter, we will account for a more complex situation in which networks are not *a priori* identical but can have different intrinsic fitnesses or are launched at different times. In particular, we will show how a globally operating network acquires a higher fitness compared to its local competitors because it provides users with the possibility to connect to individuals from different countries.

4 Competition between local and global networks

"The absence of 'counterfactual' versions of history tends to have the effect that we tend to perceive what actually happened as having been inevitable."
Duncan Watts [9]

This chapter is – with some small changes – available as a preprint at [102] and currently under review. Most figures are identical to the preprint.

4.1 Introduction

In the previous chapter we have introduced an ecological description of the digital world which deals with *a priori* identical networks. We have demonstrated that a moderate number of identical networks in competition for users' attention can coexist in the digital ecosystem, in contrast to the principle of competitive exclusion [91].

In contrast to the previous chapter, here we address the heterogeneity of networks. Networks can differ in functionality, features, and –most importantly– they can address different peer groups. Here, we show how the effect of different overlapping peer groups can be described in terms of different degrees of network fitness. We find that under certain conditions, the heterogeneity of degrees of fitness can impede coexistence which would indeed be possible for identical networks. This effect is particularly important for the competition between local networks and an international network. Unlike users of local networks, users of the international network have the possibility to interact with people in other countries, providing this network an advantage over local ones, similar to a higher fitness of a certain species. A proper modeling of this effect requires taking into account the network of interactions among countries in the world, which results in a highly complex and non-linear dynamical system made of interconnected multilayers – we are hence dealing with networks of multiplex networks [38, 103–105], in contrast to the previous chapter. Besides, we will show that inter-country interactions induce a different type of bifurcation as the symmetry of the system is broken, which constitutes a fundamentally new behavior not observed in the model presented in the previous chapter.

Empirical observations have shown that Facebook expanded massively in the middle of the first decade of this century, starting in the US, when local networks were the most popular services in most countries. Only a few years later, Facebook had become the most popular network in most countries. So, is the fate of the digital world to become dominated

by a single "big brother" as it takes over all our digital interactions? Alternatively, is digital diversity possible from a system-level perspective? In this chapter, we show that due to the nonlinear character of the underlying laws at work, the answer to both questions can be positive or negative depending on a range of parameters and, quite surprisingly, depending on chance.

4.2 Complex organization of the digital world

The digital world consists of highly connected and strongly coupled interacting subsystems. These basic building blocks are single networks, each of which obeys specific dynamics in the absence of coupling to the whole system. So the complexity of the digital world is a consequence of both the dynamics of networks in isolated environments, which we studied in chapter 2, and the interactions between many such networks, as studied in the previous chapter. Finally, not all of these building blocks are identical. Instead, different networks address different peer groups or have different functionalities. Hence, to reveal the fundamental mechanisms that determine the fate of the digital world, it is necessary to understand the interaction of heterogeneous networks, each driven by intrinsic dynamics.

4.2.1 Isolated dynamics of online social networks

The key actors in the digital world are OSNs; loosely defined as web-based platforms that enable digital social interactions over the Internet. However, societies were organized as networks long before OSNs were even thought of. From this point of view, the growth of OSNs can be described through the dynamical processes by which people in the traditional off-line social structure come to engage in OSNs. The topology of the OSN is now the digital counterpart of the underlying off-line social network [40, 106]. Recall from chapter 2 that, in isolation, this process of formation can be described by a set of simple dynamical mechanisms, namely a viral spreading mechanism and the influence of mass media acting on the pre-existing underlying social network. In chapter 2, we were able to rigorously validate the dynamics ruling OSNs in isolation; the fundamental building blocks of the digital world. These findings constitute the foundation for the development of a more comprehensive theory of interacting heterogeneous networks.

4.2.2 Competitive interaction between multiple networks

The simultaneous existence of multiple digital services in competition for the attention of users suggests an ecological perspective from which to explain the prevalence of a given network or the coexistence of multiple networks, as introduced in the previous chapter. Recall that the key principle that drives the competition between OSNs is the fact that, due to the physical and cognitive limitations of users, the time they devote to online activities is limited. As a consequence, the viral parameter, λ, constitutes a conserved quantity that is nevertheless distributed between the competing networks as $\lambda_i = \omega_i(\boldsymbol{\rho}^{\mathrm{a}})\lambda$, where $\omega_i(\boldsymbol{\rho}^{\mathrm{a}})$ represents a normalized set of weights, that is $\sum_i \omega_i(\boldsymbol{\rho}^{\mathrm{a}}) = 1$, and $\boldsymbol{\rho}^{\mathrm{a}} \equiv (\rho_1^{\mathrm{a}}, \rho_2^{\mathrm{a}}, \dots)$ is

a vector denoting the fraction of active nodes (activities) in the different networks. In general, users are more likely to subscribe to and engage in networks that are more active. Therefore, the viral activity of each network must be a function of the activity of the network itself. In particular, we model this by assuming that the weights function $\omega_i(\boldsymbol{\rho}^{\mathrm{a}})$ obeys $\partial\omega_i(\boldsymbol{\rho}^{\mathrm{a}})/\partial\rho_i^{\mathrm{a}} > 0$. From the previous chapter we recall that

$$\omega_i(\boldsymbol{\rho}^{\mathrm{a}}) = \frac{[\rho_i^{\mathrm{a}}]^{\sigma}}{\sum_{j=1}^{n_l}\left[\rho_j^{\mathrm{a}}\right]^{\sigma}}\,, \tag{4.1}$$

were n_l denotes the number of networks. This choice allows us to interpolate between a set of independent networks ($\sigma \ll 1$) and highly coupled ones ($\sigma \gg 1$). The activity affinity parameter, σ, then quantifies the tendency of users to subscribe to or engage in more active networks. Recall that, in contrast to the principle of competitive exclusion, multiple networks can coexist because the rich-get-richer mechanism is damped by the diminishing returns of the dynamics of network evolution.

4.2.3 Network heterogeneity leads to effective activity

As mentioned above, since its official launch in 2004, Facebook has become the most popular OSN in most countries; even in countries where there was already a popular OSN before Facebook was launched. To mimic the real evolution of the digital ecosystem at the worldwide scale, we assume that one local network exists in each country in addition to a globally operating, international network (see Fig. 4.1a). In the US, both networks are launched at the same time; whereas the international network is launched with a delay Δt in the remaining countries, to take into account the initial prevalence of local networks.

Once launched, the international network provides the user with the possibility to connect to individuals in different countries, in contrast to local networks, making it more attractive to users. For a given country, the advantage of the international network is directly related to the abundance of social ties between that country and the rest of the world. We use passenger air travel data as a proxy for the abundance of such ties. This choice is justified by the strong correlation between air travel flows and further measures of inter-country exchange, for instance email communication [107] or Twitter activity [108].

Users in country i experience the greater attractiveness of the international network as they perceive its activity with respect to the population of their own country and also with respect to their contacts in other countries. To account for this on a coarse grained level, in Eq. (4.1), we replace the activity of the international network by an effective activity as follows

$$\tilde{\rho}_{i,\mathrm{int}}^{\mathrm{a}} = \rho_{i,\mathrm{int}}^{\mathrm{a}} + \alpha\sum_j \Omega_{ij}\rho_{j,\mathrm{int}}^{\mathrm{a}}\,, \tag{4.2}$$

where

$$\Omega_{ij} = \frac{W_{ij}/N_i}{\max[W_{ij}/N_i]} \tag{4.3}$$

denotes the fraction of the number of air travel passengers between countries i and j, W_{ij}; and N_i, the population of country i. Notice that, in an ecological context, this corresponds

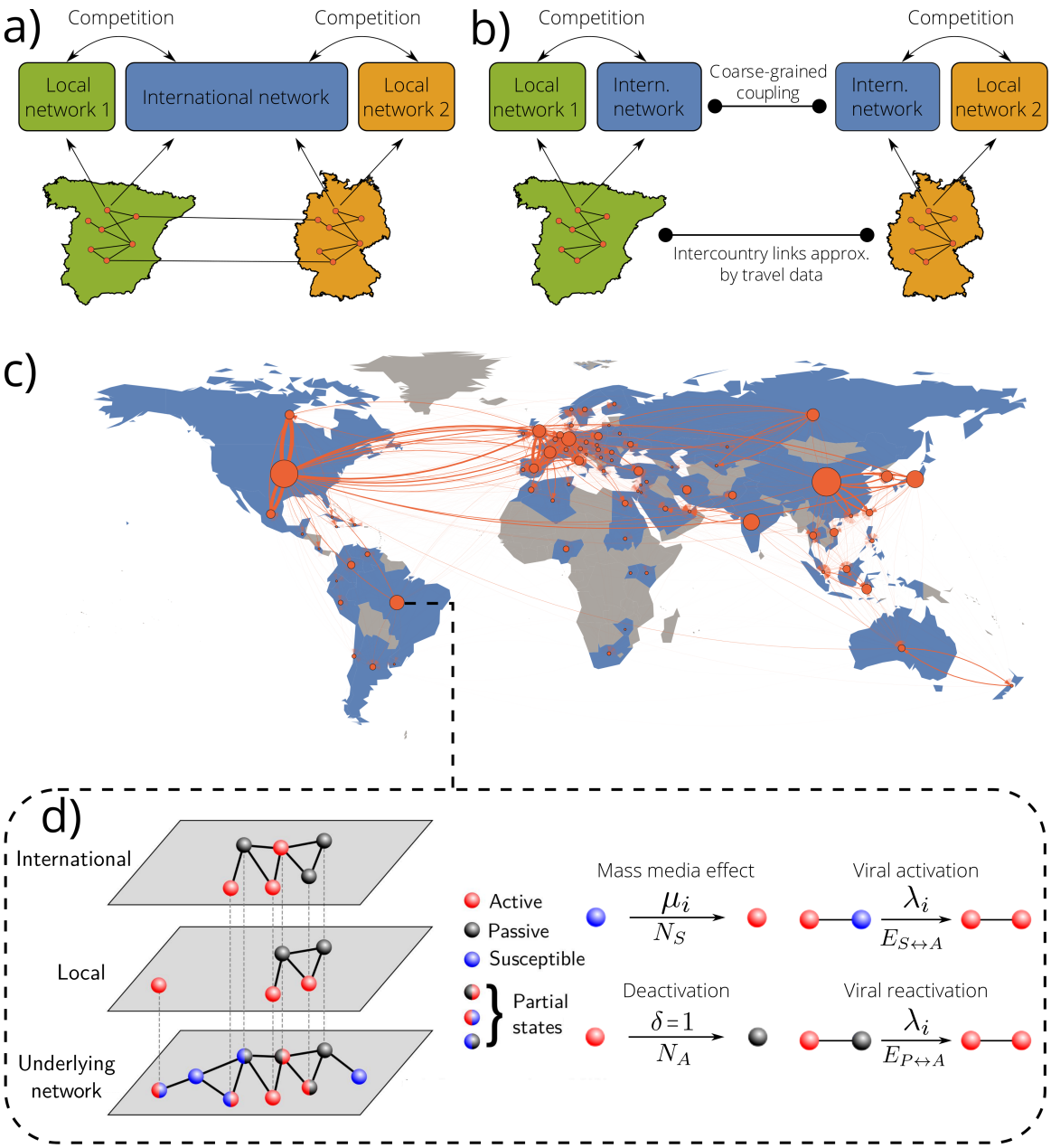

Figure 4.1: Constituents of our model. **a)** Design of the international network and local networks. **b)** Sketch of our model using coarse-grained coupling. **c)** Visualization of the flight network. The size of the nodes is proportional to the number of users in the respective countries with Internet access. The transparency and thickness of the links represents the density of passengers between the countries concerned. **d)** Illustration of the competition between the international network and the local network within one country.

to increased fitness of the international network. In Eq. (4.2), we have implicitly assumed proportionality between the number of passengers and the number of contacts in the respective countries, namely $N_{ij} \propto W_{ij}$. Finally, note that the arbitrary normalization in Eq. (4.3) serves the sole purpose of ensuring that reasonable values for the parameter α are of the order of unity.

Hereafter, we decompose the international network into a set of disjunct coupled subnetworks operating in each country and in competition with the respective local network (see Fig. 4.1b). These subnetworks are nevertheless not independent, as they are globally coupled via the effective activity defined in Eq. (4.2) and ultimately by the network representing the inter-country social ties. Hence, our model forms a network of networks [38, 103–105], where each node in Fig. 4.1c represents a three-layer multiplex network [75,109] in which the bottom layer corresponds to the underlying social structure and the two upper layers denote the local and international networks operating in the respective country (see Fig. 4.1d).

4.3 Double meanfield approximation reveals complex role of the activity affinity

To understand the qualitative behavior of the system, in this section we present a double meanfield approximation of the system. This reduces the system given by a network of networks to a set of evolution equations of the average activity in the international network and in local networks. As we show in section 4.4, the results of the full model with heterogeneous topologies exhibits similar behavior to that encountered by the double meanfield approximation.

The first meanfield approximation consists of assuming a fully mixed homogeneous population in each country. Let $\rho^{\mathrm{a}}_{i,l}$ denote the fraction of active users in network $l \in (\mathrm{loc}, \mathrm{int})$ in country i and $\rho^{\mathrm{s}}_{i,l}$ the fraction of nodes susceptible to joining this network. Then, the fraction of passive users is given by $1 - \rho^{\mathrm{s}}_{i,l} - \rho^{\mathrm{a}}_{i,l}$. As explained above, in each country the virality is distributed between the local and international network via the weight functions $\omega_{\mathrm{loc}}(\rho^{\mathrm{a}}_{i,\mathrm{loc}}, \tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}})$ and $\omega_{\mathrm{int}}(\rho^{\mathrm{a}}_{i,\mathrm{loc}}, \tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}}) = 1 - \omega_{\mathrm{loc}}(\rho^{\mathrm{a}}_{i,\mathrm{loc}}, \tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}})$, as introduced in Eq. (4.1). Here, $\tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}}$ denotes the effective activity of the international network as defined in Eq. (4.2). The evolution equations of the resulting system represent a generalization of the evolution equations for identical networks which we derived in the previous chapter, where one replaces the activity of the international network with the effective activity from Eq. (4.2) in the argument of the weights function from Eq. (4.1). This procedure yields

$$\begin{aligned} \dot{\rho}^{\mathrm{a}}_{i,l} &= \rho^{\mathrm{a}}_{i,l}\Big\{\lambda \langle k \rangle \, \omega_l(\rho^{\mathrm{a}}_{i,\mathrm{loc}}, \tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}}) \left[1 - \rho^{\mathrm{a}}_{i,l}\right] - 1\Big\} + \mu_{i,l}\rho^{\mathrm{s}}_{i,l} \\ \dot{\rho}^{\mathrm{s}}_{i,l} &= -\mu_{i,l}\rho^{\mathrm{s}}_{i,l}\Big\{1 + \nu \langle k \rangle \, \rho^{\mathrm{a}}_{i,l}\Big\}. \end{aligned} \tag{4.4}$$

We further assume the same linear relationship between virality and media influence in each

country as in the previous chapters, that is

$$\begin{aligned}\mu_{i,\mathrm{loc}} &= \frac{\lambda\omega_{\mathrm{loc}}(\rho^{\mathrm{a}}_{i,\mathrm{loc}}, \tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}})}{\nu} \\ \mu_{i,\mathrm{int}} &= \frac{\lambda\omega_{\mathrm{int}}(\rho^{\mathrm{a}}_{i,\mathrm{loc}}, \tilde{\rho}^{\mathrm{a}}_{i,\mathrm{int}})}{\nu}\,.\end{aligned} \tag{4.5}$$

Analogously to the previous chapter, the value of ν does not affect the stability of the system. In what follows, we perform the stability analysis in the limit $\nu \to \infty$. This decouples the evolution of $\rho^{\mathrm{a}}_{i,l}$ from $\rho^{\mathrm{s}}_{i,l}$, so that we only have to consider $\rho^{\mathrm{a}}_{i,l}$. Plugging in the weights function defined in Eq. (4.1) and the effective activity from Eq. (4.2) yields the evolution equations for the activities of the local and international networks in country i,

$$\begin{aligned}\dot{\rho}^{\mathrm{a}}_{i,\mathrm{loc}} &= \rho^{\mathrm{a}}_{i,\mathrm{loc}}\left[\frac{\lambda\,\langle k\rangle\,[\rho^{\mathrm{a}}_{i,\mathrm{loc}}]^{\sigma}}{[\rho^{\mathrm{a}}_{i,\mathrm{loc}}]^{\sigma} + (\rho^{\mathrm{a}}_{i,\mathrm{int}} + \delta_i)^{\sigma}}[1-\rho^{\mathrm{a}}_{i,\mathrm{loc}}] - 1\right] \\ \dot{\rho}^{\mathrm{a}}_{i,\mathrm{int}} &= \rho^{\mathrm{a}}_{i,\mathrm{int}}\left[\frac{\lambda\,\langle k\rangle\,(\rho^{\mathrm{a}}_{i,\mathrm{int}} + \delta_i)^{\sigma}}{[\rho^{\mathrm{a}}_{i,\mathrm{loc}}]^{\sigma} + (\rho^{\mathrm{a}}_{i,\mathrm{int}} + \delta_i)^{\sigma}}[1-\rho^{\mathrm{a}}_{i,\mathrm{int}}] - 1\right]\,,\end{aligned} \tag{4.6}$$

where $\delta_i = \alpha \sum_j \Omega_{ij}\rho^{\mathrm{a}}_{j,\mathrm{int}}$.

The second meanfield approximation consists of applying the hypothesis of a fully mixed homogeneous network for the inter-country social ties. We use $\bar{\Omega} = \alpha\,\langle\Omega_{ij}\rangle$ and define the mean activity of the local networks as $x \equiv \langle\rho^{\mathrm{a}}_{i,\mathrm{loc}}\rangle$ and the mean activity of the international network as $y \equiv \langle\rho^{\mathrm{a}}_{i,\mathrm{int}}\rangle$. Finally, our double meanfield approximation leads to the following system of coupled differential equations

$$\begin{aligned}\dot{x} &= x\left[\lambda\,\langle k\rangle\,\frac{x^{\sigma}}{x^{\sigma} + (y(1+\bar{\Omega}))^{\sigma}}[1-x] - 1\right] \\ \dot{y} &= y\left[\lambda\,\langle k\rangle\,\frac{(y(1+\bar{\Omega}))^{\sigma}}{x^{\sigma} + (y(1+\bar{\Omega}))^{\sigma}}[1-y] - 1\right]\,,\end{aligned} \tag{4.7}$$

which has three relevant parameters: $\lambda\,\langle k\rangle$, σ, and $\bar{\Omega}$. Note that by setting $\bar{\Omega}$ to zero, we recover the equations for identical networks presented in the previous chapter.

In what follows, we discuss the dynamical properties of the system given by Eq. (4.7). For constant σ, the system exhibits a saddle-node bifurcation at a critical value of the global connectivity $\bar{\Omega}_c(\sigma)$ (see Fig. 4.2). Above this point, coexistence is not possible and the only stable solutions correspond to the domination of either local networks or the international one. Both above and below the critical value $\bar{\Omega}_c(\sigma)$, the basin of attraction of the solution corresponding to the domination of local networks decreases with $\bar{\Omega}$, whereas that of the international network increases (see the rows of Fig. 4.2). Furthermore, at the critical point, the basin of attraction of the international network is amplified discontinuously as the region of coexistence in the subcritical regime is now merged with the basin of attraction of the domination of the international network.

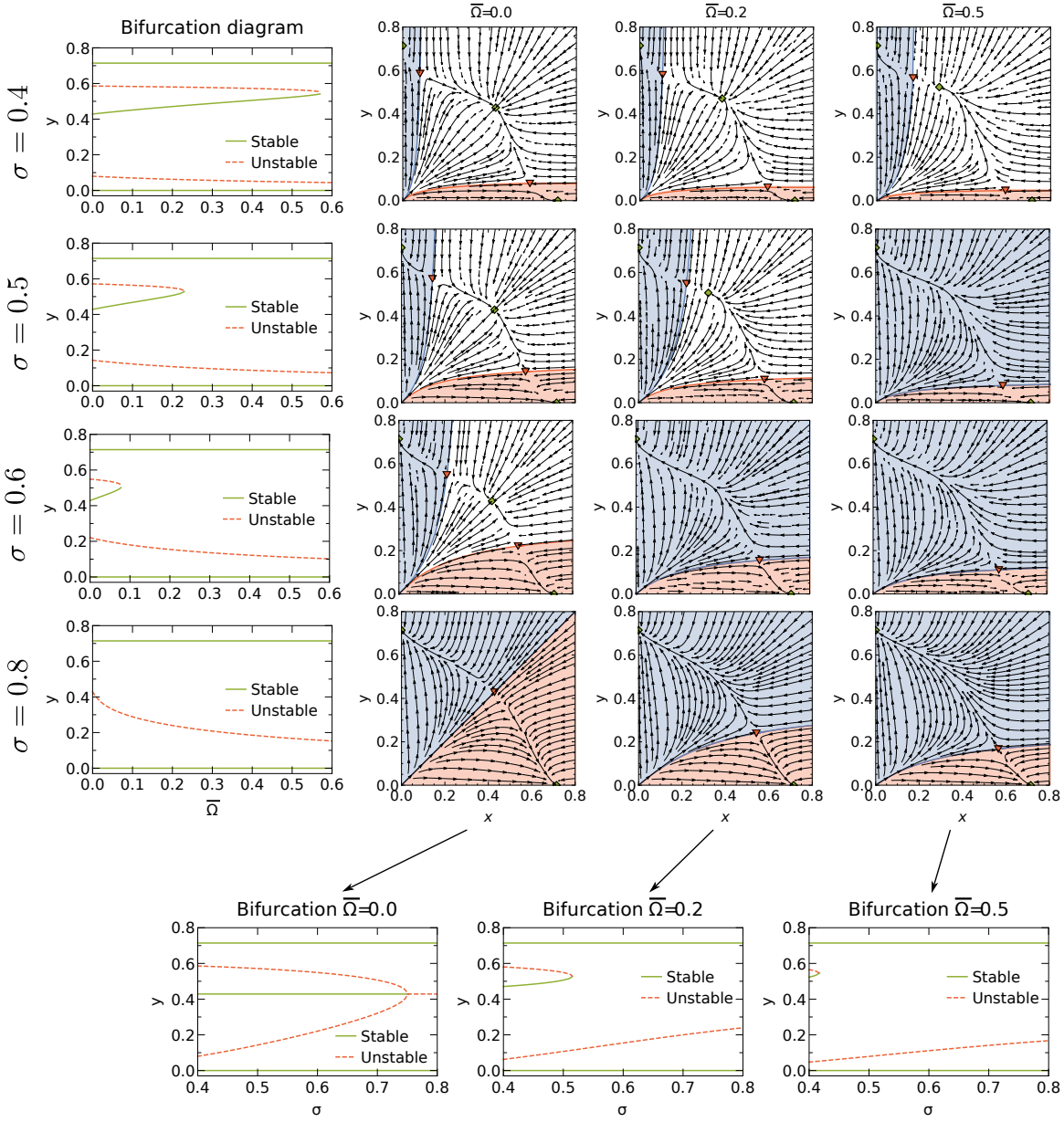

Figure 4.2: Bifurcation diagram and stream plots for the double meanfield approximation (4.7) for $\lambda \langle k \rangle = 3.5$ and $\nu \to \infty$. The basins of attraction for the domination of the international network are in blue, the basins of attraction for the domination of local networks are in red; the white areas correspond to the basins of attraction of the coexistence solution (if it exists).

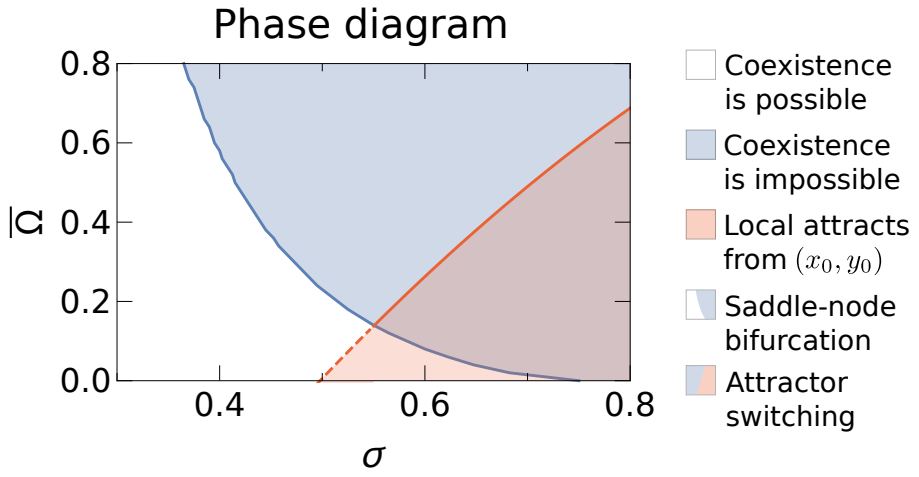

Figure 4.3: Phase diagram of the double meanfield approximation for $\lambda \langle k \rangle = 3.5$. The white area denotes the parameters for which coexistence is possible. The blue area denotes the parameters for which only domination of either local or the international network can occur. At the blue line, the system undergoes a saddle-node bifurcation in which the stable coexistence solution disappears. The red line denotes the combination of parameters for which the system switches attractors for the initial conditions given by Eq. (4.8), where we use $\beta = 0.2$ to reflect that the US contributes about 20% of the population. The red region shows the parameters for which the solution of domination of local networks is reached for these initial conditions. Above the red line, the system approaches the domination of the international network (solid red line) or the coexistence solution (dashed red line).

For constant $\bar{\Omega} > 0$, the system also exhibits a saddle-node bifurcation[1] at a critical value of the activity affinity $\sigma_c(\bar{\Omega})$. The evolution of the basins of attraction is more complex compared to the previous case. Below the critical point, both basins of attraction increase with σ. Above the critical point $\sigma_c(\bar{\Omega})$, the basin of attraction of the local networks increases whereas the basin of attraction of the international network decreases with σ (see the columns of Fig. 4.2). This is particularly interesting as it implies that an intermediate value of the activity affinity just slightly above the critical point $\sigma \gtrsim \sigma_c(\bar{\Omega})$ represents the worst scenario for the survival of local networks, since at this point the size of the basin of attraction of the domination of the international network is maximum.

In Fig. 4.3, the blue line indicates the critical line $\bar{\Omega}_c(\sigma)$ in the σ-$\bar{\Omega}$ plane, which separates a phase in the parameter space where coexistence is possible (white region) and one in

[1] In the previous chapter we showed that the system undergoes a subcritical pitchfork bifurcation with respect to the control parameter σ, above which no stable coexistence is possible. $\bar{\Omega} > 0$ breaks the symmetry of the pitchfork bifurcation and in this case the system undergoes a saddle-node bifurcation with respect to σ instead (see bottom panel of Fig. 4.2). This behavior is well known in bifurcation theory and results from adding a small error term to the normal form of the pitchfork bifurcation (see appendix D.5).

which only domination can occur (blue region). However, the increasing size of the basin of attraction of the domination of local networks above the critical point with respect to $\sigma > \sigma_c(\bar{\Omega})$ can dramatically alter the fate of the system for a given set of initial conditions. Assume, for instance, that the international network dominates in the US and starts with a significant delay in each other country, which causes the local networks to dominate in those countries. At the time when the international network is launched globally, the state of the system can be approximated as follows

$$\begin{aligned} x_0 &= (1-\beta)\left[1 - \frac{1}{\lambda \langle k \rangle}\right] \\ y_0 &= \beta \left[1 - \frac{1}{\lambda \langle k \rangle}\right], \end{aligned} \tag{4.8}$$

which we now use as initial conditions to study the further evolution. Notice that if one network dominates in country i, its activity is given by $\rho^{\mathrm{a}}_{i,l} = 1 - \frac{1}{\lambda \langle k \rangle}$. Hence, the initial conditions given by Eq. (4.8) reflect the fact that local networks dominate in the fraction $(1-\beta)$ of the system and the international one dominates in the remainder. The evolution of the basins of attraction makes the system approach different stationary solutions from these initial conditions for different parameters. Below the red line in Fig. 4.3, the system approaches the domination of local networks starting from the initial conditions given in Eq. (4.8). Above this line, the system either approaches coexistence (white area; crossing dashed red line) or domination of the international network (blue area; crossing solid red line). This means that in the red region, when the international network is launched globally, it is not able to overcome the initial advantage of the local networks due to its earlier launch.

To conclude, the double meanfield approximation predicts that intermediate values of the activity affinity most favor the international network; whereas the local networks can dominate for a high activity affinity and low global connectivity. We confirm these findings by numerical simulations in the following section.

4.4 Numerical simulations and synthetic networks

In this section, we go beyond the meanfield approximation and study, by means of numerical simulations, the effects of the real topology of inter-country social ties and of underlying social structures. To this end, we use the air travel network (see Fig. 4.1c and appendix D.2) as a proxy for inter-country social ties and construct 1:1000 scaled synthetic networks to model the structure of the 80 countries with most Internet users (see Tab. 4.1). To generate these networks, we make use of a model introduced in [110–112], which produces realistic topologies of the traditional off-line social networks, including heterogeneous node degrees and a high level of clustering (see appendix D.1).

Fig. 4.4 shows results from our model for the set of parameters that best matches empirical observations, which we will explain in section 4.5. The international network starts with a delay in all countries except the US; so that initially in these countries the respective local network dominates. After some time, the international network obtains a significant advantage and quickly takes over in most countries.

China	253.	UnitedStates	231.
Japan	90.9	India	81.
Brazil	64.9	Germany	62.
UnitedKingdom	48.8	Russia	45.2
France	42.9	SouthKorea	37.5
Indonesia	30.	Spain	25.2
Canada	25.1	Italy	25.
Turkey	24.5	Mexico	23.3
Iran	23.	Vietnam	20.8
Poland	18.7	Pakistan	18.5
Colombia	17.1	Malaysia	16.9
Thailand	16.1	Australia	15.2
Taiwan	15.1	Netherlands	14.3
Egypt	11.4	Argentina	11.2
Nigeria	11.	Ukraine	10.4
Morocco	10.3	Sweden	8.1
SaudiArabia	7.7	Belgium	7.3
Venezuela	7.2	Peru	7.1
Romania	6.1	CzechRepublic	6.
Austria	5.9	Hungary	5.9
Switzerland	5.7	Philippines	5.6
Chile	5.5	Denmark	4.6
Portugal	4.5	Finland	4.4
Greece	4.3	Sudan	4.2
SouthAfrica	4.2	HongKong	4.1
Algeria	4.1	Norway	3.9
Slovakia	3.6	Syria	3.6
Singapore	3.4	Kenya	3.4
Belarus	3.1	NewZealand	3.
Serbia	2.9	UnitedArabEmirates	2.9
Ireland	2.8	Tunisia	2.8
Bulgaria	2.6	Uganda	2.5
Uzbekistan	2.5	Kazakhstan	2.3
Lebanon	2.2	DominicanRepublic	2.1
Israel	2.1	Guatemala	2.
Croatia	1.9	Lithuania	1.8
Jamaica	1.5	Jordan	1.5
Azerbaijan	1.5	CostaRica	1.5
Cuba	1.4	Zimbabwe	1.4
Uruguay	1.3	Ecuador	1.3

Table 4.1: List of countries and estimated number of Internet users ($\times 10^6$) according to the Wolfram Alpha database.

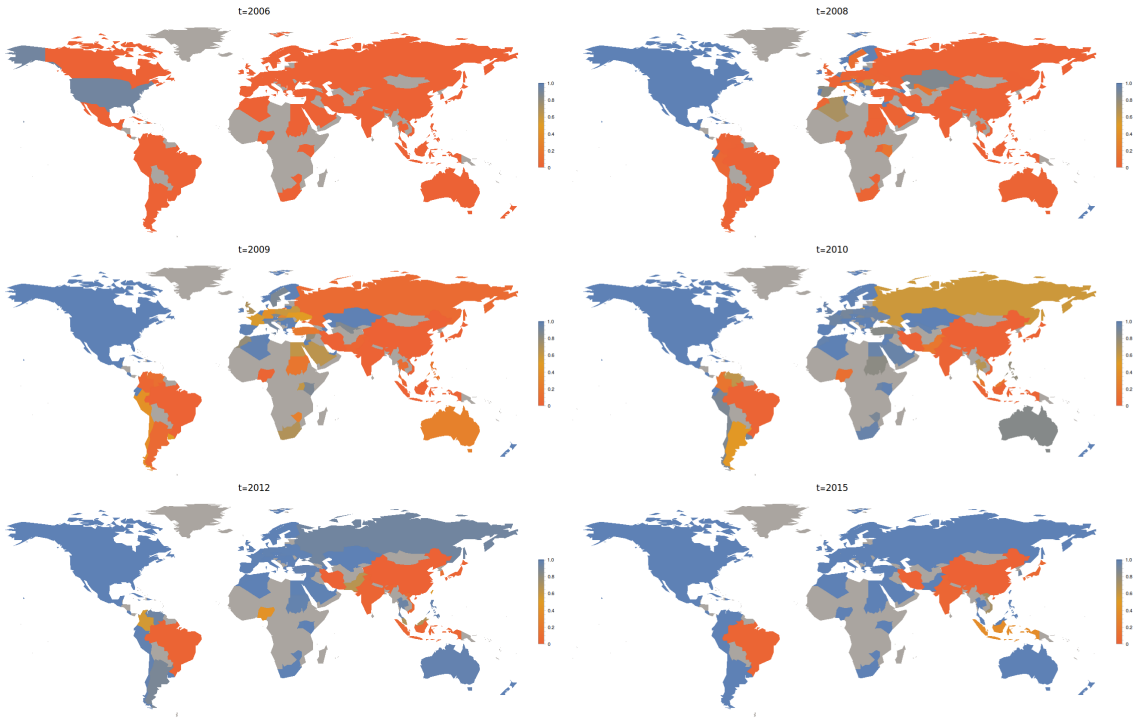

Figure 4.4: Evolution of our model averaged over several realizations for the parameters $\sigma = 1.25$, $\Delta t = 3$, and $\alpha = 2$ (here, we excluded realizations in which local networks dominate, which occurs with approximately 30% probability for these parameters). The mapping from model time to real time is explained in section 4.5. The relative importance of the influence of mass media compared to the viral spreading mechanism is governed by the parameter ν introduced in Eq. (4.5). In all numerical simulations, we set $\nu = 4$: the value found empirically in chapter 2; and $\lambda = 0.2$ (this corresponds to $\lambda/\lambda_c^1 \approx 4.3$ in 3). The relative prevalence of the international network, given by $\rho_{i,\mathrm{int}}^{\mathrm{a}}/(\rho_{i,\mathrm{int}}^{\mathrm{a}} + \rho_{i,\mathrm{loc}}^{\mathrm{a}})$, is color coded. We consider the international network to be banned in China and Iran. To model this, we set the values of $\Omega_{ij} = 0$ for each entry which involves one of these countries. This is equivalent to assuming that in these countries two local networks compete without any coupling to the rest of the world.

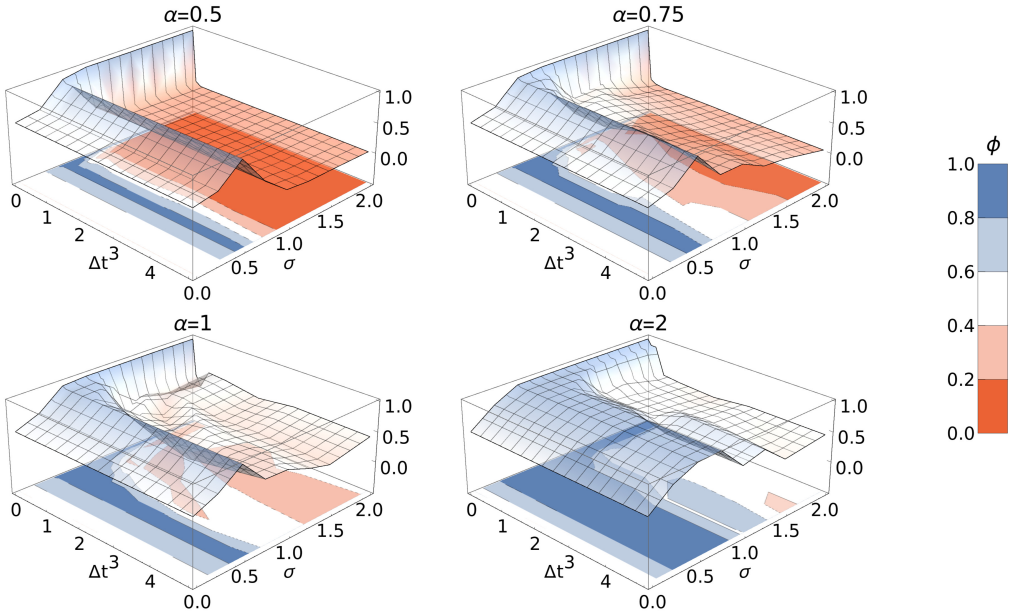

Figure 4.5: Relative prevalence Φ of the international network is plotted on the z axis as a function of the launch time delay Δt and the coupling strength σ. Averaged over 30 realizations.

To further study the properties of the model presented here, we define the relative prevalence of the international network compared to local networks as

$$\Phi = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{\rho^{\mathrm{a}}_{i,\mathrm{int}}\Big|_{\mathrm{st}}}{\rho^{\mathrm{a}}_{i,\mathrm{int}}\Big|_{\mathrm{st}} + \rho^{\mathrm{a}}_{i,\mathrm{loc}}\Big|_{\mathrm{st}}} , \tag{4.9}$$

where n_c denotes the number of countries, and $\rho^{\mathrm{a}}_{i,\mathrm{int}}\Big|_{\mathrm{st}}$ and $\rho^{\mathrm{a}}_{i,\mathrm{loc}}\Big|_{\mathrm{st}}$ are the activities of the international and local networks in country i in the stationary state. With this definition, a value of $\Phi \approx 0$ implies that local networks dominate in most countries, whereas $\Phi \approx 1$ corresponds to the domination of the international network. The relative prevalence of the international network averaged over many realizations is shown in Fig. 4.5 for different values of α as a function of the activity affinity, σ, and the launch time delay, Δt. For small values of Δt, we observe that when σ is small, the international and local networks coexist and we observe values around $\Phi \approx 0.5$ for the relative prevalence; then, increasing σ favors the international network, which dominates for values of $\sigma \gtrsim 0.5$ (see Fig. 4.6 top and center). For larger values of Δt, this behavior smoothly translates into a more complex case, which we discuss below.

We observe in Fig. 4.5 that for launch time delays $\Delta t \geq 2$, the actual length of the delay becomes irrelevant. This behavior corresponds to the limit of saturation of the evolution

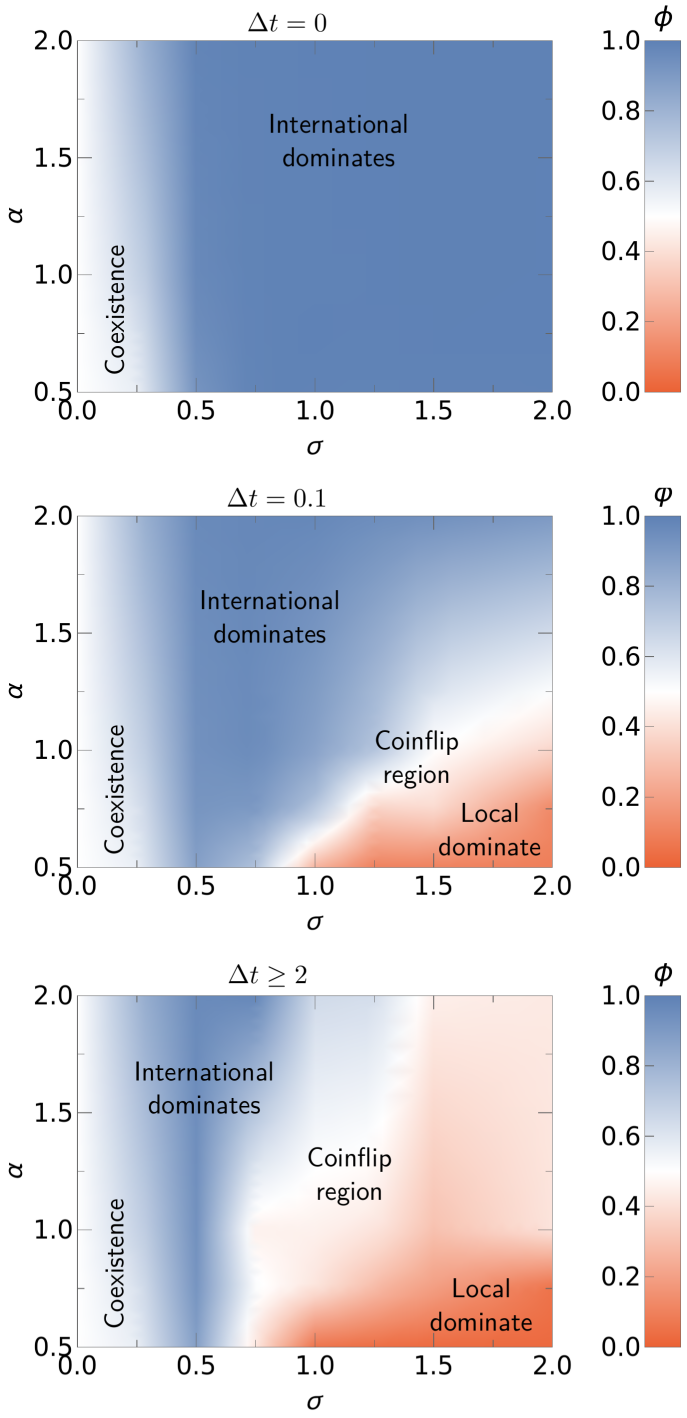

Figure 4.6: The prevalence of the international network for **top:** $\Delta t = 0$, **center:** $\Delta t = 0.1$, and **bottom:** averaged over time delays $\Delta t \geq 2$ as a function of the activity affinity (σ) and the global connectivity (α). Averaged over 30 realizations.

of local networks before the international network is launched; as discussed in section 4.3. We consider this limit by averaging over regions with $\Delta t \geq 2$ in Fig. 4.6 (bottom), which yields a two dimensional parameter space σ-α. Indeed, numerical simulations of the full model confirm the results from the meanfield analysis; in particular the complex role of the activity affinity σ. For small α and σ, local networks and the international network can coexist. Increasing σ or α favors the domination of the international network, which gives rise to the blue "V"-shaped region around $\sigma = 0.5$. This corroborates the saddle-node bifurcation predicted by the double meanfield approximation. See supplementary video [113] for an explicit realization. For high values of σ and small values of α (red region in the bottom right-hand corner of Fig. 4.6), local networks dominate. Note that partial states are also possible, in which the international network dominates in some countries and local networks dominate in the remaining countries. See supplementary video [114] for an explicit realization of this case.

Between the regions of domination of the international network and of local networks, there is a region in which the final fate of the system varies significantly between different realizations of the model ("coinflip region"). In this region, if the international network wins initially in the US, it will become dominant globally; otherwise, local networks maintain their initial prevalence. Although in this region the prevalence of the international network averaged over many realizations is about 0.5, as in the coexistence region in the bottom left-hand corner of Fig. 4.6, the behavior of the system differs dramatically from one to another. In the coexistence region, each realization of the model leads to the same final state: coexistence of local networks and the international one. In contrast, in the coinflip region, coexistence is not possible, as this region of the parameter space corresponds to the supercritical regime (the blue area in Fig. 4.3). In the coinflip region, about 50% of the realizations end up with domination of the international network, whereas the remaining 50% lead to the domination of local networks. As a consequence, even if we know the exact parameters, it is impossible to predict the fate of the system beforehand.

We can summarize these findings as follows. A higher value of α, which is a measure of the global connectivity of society, favors the prevalence of the international network and hinders the survival of the local ones. The role of the tendency of individuals to participate in more active networks (activity affinity), σ, is particularly interesting. Low values allow the networks to coexist, whereas intermediate values always lead to the prevalence of the international network and the extinction of local networks. A high activity affinity, however, enables the prevalence of local networks and thus can even lead to the extinction of the international network.

4.5 Comparison with empirical data

In this section, we compare the results of our model with empirical data on the recent expansion of Facebook at the cost of many local networks. In particular, we consider the evolution of the number of countries in which local networks (i.e. networks that are not Facebook) are the most popular ones, as measured in [115] using Alexa traffic data (see

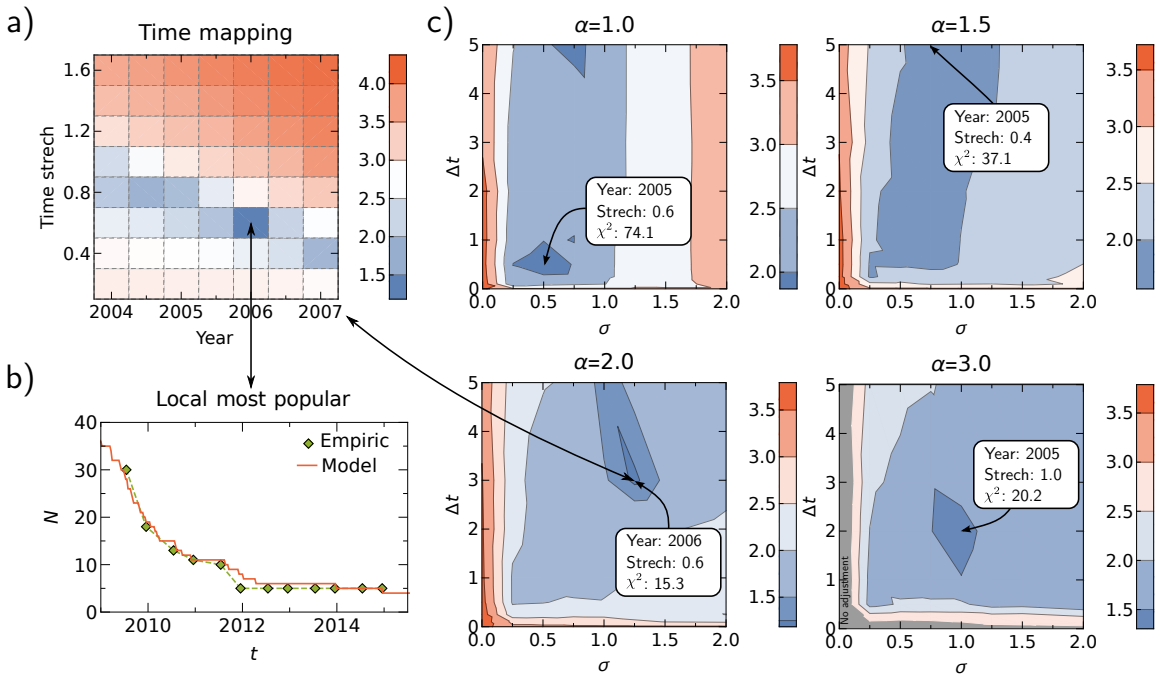

Figure 4.7: Comparison between model results and empirical data. **a)** χ^2 for different values of a_1 (year) and a_2 (time stretch) for $\sigma = 1.25$, $\alpha = 2$, and $\Delta t = 3$. **b)** Model results and empirical data. Here, the parameters are $\sigma = 1.25$, $\alpha = 2$, and $\Delta t = 3$ and the optimal mapping is given by $a_1 = 2006$ and $a_2 = 0.6$ (see **a)**. **c)** χ^2 for the respective best time mapping as a function of σ and Δt for different values of α. In each plot, the minimal value of χ^2 and the respective time mapping are shown in the boxes. The color coding in all plots represents the logarithm of χ^2.

Fig. 4.7b). We observe a significant decline of this number, which rules out the possibility that the empiric case corresponds to the domination of local networks. Because the past can be considered a single realization of a stochastic process [9], the empiric case can still be within the coinflip region of our model where –by chance– the international network was more successful. Hence, we will perform the following comparison only for realizations of our model in which local networks do not dominate.

The intrinsic timescale of the model is arbitrary and hence has to be mapped to real time. The optimal mapping is given such that it produces the best agreement with the empirical data. We quantify the agreement between model results and empirical data using the sum of the squared distances between the data points and model results. In particular, we use the χ^2 statistic defined as

$$\chi^2 = \frac{1}{\sigma_N^2} \sum_i \left[N_i - N_i^{\text{model}} \right]^2 , \tag{4.10}$$

where N_i denotes the number of countries where the local network is more popular and N_i^{model} is the corresponding result from the model. The index i denotes the individual

datapoints and $\sigma_N^2 = 1.5$ is the estimated variance of the data (see appendix D.3). The real time, t_R, is a linear function of the model time t_M given by $t_R = a_1 + a_2 t_M$, where a_1 is the starting year and a_2 represents the time stretch: how many years of real time correspond one model time step. For a given set of parameters, σ, α, and Δt, the optimal values for a_1 and a_2 are those that minimize χ^2, as shown in Fig. 4.7a.

We can also use the χ^2 statistic to estimate the parameters α, σ, and Δt which best reproduce the empirical observations. In Fig. 4.7c, we plot the values of χ^2 as a function of α, σ, and Δt, where –at each point– we applied the respective best time mapping, as described above. These results are averaged over several realizations of the model; however, in the coinflip region we exclude realizations where the local networks dominate, to mimic the empirical case. Interestingly, the overall best fit is achieved for $\alpha = 2$ at $\sigma = 1.25$ and $\Delta t = 3$, which lies in the coinflip region[2] with a probability for domination of the international network of 70%. This scenario corresponds to the time mapping $a_1 = 2006$ and $a_2 = 0.6$, meaning the system started at the beginning of 2006; while the launch time delay of $\Delta t = 3$ in the model translates to 1.8 years in real time. In Fig. 4.7b, we show the evolution of the number of countries where local networks are more popular for the optimal fit from the model.

4.6 Summary: Competition between local and global networks

Understanding the complex dynamics of the digital world constitutes an important challenge for interdisciplinary science. To meet this challenge, here we describe the worldwide web as a complex, digital ecosystem in which interacting networks play the role of species in competition for survival. In this chapter, we studied the competition between local networks operating in single countries and an international network that operates in all countries. Therefore, a proper description of this system must necessarily involve the network of worldwide social interactions between different countries.

We showed that the effect of inter-country social ties can be mapped to the increased fitness of the international network by means of an effective activity. Interestingly, there is a critical global coupling strength below which networks can coexist. However, above that threshold, only domination is possible: in general, local networks become extinct with a high probability. Yet, we find that if local networks are launched earlier they can persist and dominate the international network, which happens only if local networks have accumulated a sufficiently large active userbase when the global launch of the international network takes place. The accumulation of a sufficient base depends on the parameters; and for certain parameters on chance. For these parameters the final state of the system –whether local networks dominate or become extinct– can be completely unpredictable, as it varies randomly between different realizations of the model.

Quite remarkably, a thorough comparison of our model with empirical data from the recent takeover of Facebook indicates that the most probable launch date of Facebook was

[2]The optimal value of χ^2 is statistically consistent with the model, given the number of degrees of freedom in the data.

at the beginning of 2006 and its global launch was in late 2007. Facebook was in fact started in 2004, but opened to the public in 2006; in good agreement with the estimate from our model. Moreover, according to Google trend data (see appendix D.4), 2007 was the year when the global search volume for Facebook started to increase rapidly. Last but not least, our best estimation of the model parameters corresponds to the "coinflip" region, which means that the observed takeover of Facebook only had a probability of around 70%. With a 30% probability, we would have been living in a world where each country had its own successful local network and a network like Facebook would not exist [9].

Our findings suggest interesting future lines of research. On the one hand, even without adjusting the parameters on a country-by-country level, our model reproduces the main features empirically observed in the takeover of Facebook and the extinction of local networks in most countries for a certain parameter region. It remains an interesting task for future research to further increase the precision of the model. This could be done by improving the proxy for the similarity between countries or by adjusting parameters on a country-by-country basis. On the other hand, the model could be extended to account for several international networks and to study their global competition. For a second international network to overcome the first, a certain minimal difference of fitness is needed; which could be the result of different properties of the networks, such as features or functionalities. Finally, random fluctuations of fitness could be incorporated to describe Darwinian selection in the digital ecosystem.

Geometry of multiplex networks

5 Hidden geometric correlations: community detection and link prediction

"Birds of a feather flock together."
UNKNOWN

This chapter is – with some small changes – available as a preprint at [116] and currently under review. Most figures are identical to the preprint.

5.1 Introduction

Real networks are often not isolated entities but instead can be considered constituents of larger systems, called multiplexes or multilayer networks [62, 75, 109, 117–121]. Examples can be found everywhere. The most classical one is the multiplex consisting of the different social networks that a person may belong to, similar to the systems we studied in the previous chapters. Other examples include the Internet's IPv4 and IPv6 topologies, or the structural and functional networks in the brain. Understanding the relations among the networks comprising a larger multiplex is crucial for understanding the behavior of a wide range of real world systems [40, 89, 102, 122, 123]. However, despite the burst of recent research in studying the properties of multiplex networks, e.g., [75, 109, 124], a universal framework describing the relations among the single networks comprising a multiplex, and what implications these relations may have when it comes to applications, remains elusive.

In this chapter, we show that real multiplexes are not random combinations of single network layers. Instead, we find that their constituent networks exhibit strong *hidden geometric correlations*. These correlations are called "hidden" as they are not directly observable by looking at each individual network's topology. Specifically, each single network can be mapped (embedded) into a separate hyperbolic space [125–128], where node coordinates abstract the *popularity* and *similarity* of nodes [84, 110]. We find that node coordinates are strongly correlated across layers of real multiplexes, meaning that distances between nodes in the underlying hyperbolic spaces of the constituent networks are also strongly correlated.

The discovered geometric correlations yield a very powerful framework for answering important questions related to real multiplexes. Specifically, we first show that these correlations imply the existence of multidimensional communities, which are sets of nodes that are similar (close in the underlying space) in multiple layers, and which we can detect. Further, we show that strong geometric correlations imply accurate trans-layer link prediction, where connections in one layer can be predicted by knowing the hyperbolic distances among nodes in another layer. This is important for applications where we only know the

connections among nodes in one context, e.g., structural connections between brain regions, and we want to predict connections between the same nodes in some other context, e.g., likelihood of functional connections between the same brain regions.

It has been shown that many real (single layer) complex networks have an effective or hidden geometry underneath their observed topologies, which is hyperbolic rather than Euclidean [84, 125, 126, 129]. In this chapter, we extend the hidden geometry paradigm to real multiplexes and show that the coordinates of nodes in the different underlying spaces of layers are correlated. In the following chapter, we discuss how these correlations affect the navigability of multiplex networks, in particular the IPv4 and IPv6 topologies of the Internet.

5.2 Hyperbolic geometry: from single networks to multiplexes

5.2.1 Geometry of single layer networks

Nodes of real single-layered networks can be mapped to points in the hyperbolic plane, such that each node i has the polar coordinates, or hidden variables, r_i, θ_i. The radial coordinate r_i abstracts the node popularity. The smaller the radial coordinate of a node, the more popular the node is, and the more likely it attracts connections. The angular distance between two nodes, $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$, abstracts their similarity. The smaller this distance, the more similar two nodes are, and the more likely they are connected. The hyperbolic distance between two nodes, very well approximated by $x_{ij} \approx r_i + r_j + 2\ln\sin(\Delta\theta_{ij}/2)$ [125], is then a single-metric representation of a combination of the two attractiveness attributes, radial popularity and angular similarity. The smaller the hyperbolic distance between two nodes, the higher is the probability that the nodes are connected, meaning that connections take place by optimizing trade-offs between popularity and similarity [84].

Techniques based on Maximum Likelihood Estimation[1] for inferring the popularity and similarity node coordinates in a real network have been derived in [126] and recently optimized in [127, 128]. It has been shown that through the constructed hyperbolic maps one can identify soft communities of nodes, which are groups of nodes located close to each other in the angular similarity space [84, 126, 130]; predict missing links with high precision [127, 128, 130]; and facilitate efficient greedy routing in the Internet, which can reach destinations with more than 90% success rate, following almost shortest network paths [126–128]. Here, we extend this approach to multiplex networks.

5.2.2 Datasets of real multiplex networks

Within this chapter, we consider different real-world multiplex networks from diverse domains. Specifically, we consider the IPv4 and IPv6 topologies of the Internet's Autonomous Systems [131], the Indian airport and train networks [121], genetic interaction networks from the Drosophila Melanogaster (common fruit fly) [132, 133], synaptic junction networks from the C. Elegans Connectomme [134, 135], structural and functional networks from the

[1] We will explain this process in the following.

Name	Type	Nodes	Layer 1/Layer 2	Source
Internet	Technological	Autonomous Systems	IPv4 AS topology/IPv6 AS topology	[131]
Air/Train	Technological	Airports, Train stations	Indian airport network/Indian train network	[121]
Drosophila	Biological	Proteins	Suppressive genetic interaction/Additive genetic interaction	[132,133]
C. Elegans	Biological	Synapses	Electric synaptic junctions/Chemical monadic synaptic junctions	[134,135]
Brain	Biological	Brain regions	Structural network/Functional network	[123]
arXiv	Collaboration	Authors	physics.bio-ph category/cond-mat.dis-nn category	[136]

Table 5.1: Overview of the considered real-world multiplex network data.

human brain [123], and collaboration networks from two different categories of arXiv papers that have the word "networks" in the title or abstract [136]. An overview of the considered datasets is given in Tab. 5.1. In appendix E.1 we provide a detailed description of these datasets.

5.2.3 Hyperbolic mapping of multiplex networks

To map each layer of each real multiplex from Tab. 5.1 to its hyperbolic space we use the *HyperMap* method [127, 128], whose implementation is available at [137]. On its input the method takes the network adjacency matrix α_{ij} ($\alpha_{ij} = \alpha_{ji} = 1$ if there is a link between nodes i and j, and $\alpha_{ij} = \alpha_{ji} = 0$ otherwise), and the network parameters m, γ, T. It then computes radial and angular coordinates r_i, θ_i, for all nodes $i \leq N$ in the network. Parameter m is the expected minimum node degree, γ is the power law degree distribution exponent, and T is the temperature. The values of m, γ, T used to embed each layer are shown in Table 5.2.

To estimate the values of m, γ, T for each layer, we use the Extended Popularity×Similarity Optimization (E-PSO) model described in [127]. The E-PSO model grows synthetic complex networks and is equivalent to the hyperbolic $\mathbb{H}^2$ model [125]. It takes as input the final network size N, the average node degree $\bar{k}$, and the network parameters m, γ, T. We use the E-PSO model to construct synthetic networks with the same size N and average degree $\bar{k}$ as in each real layer, using different parameter values for m, γ, T. The estimated m, γ, T values for each layer are then the values that best match the degree distribution and average clustering between the layer and the corresponding synthetic network. We observe from Tab. 5.2 that in several cases the layers of the same multiplex have the same or similar estimated parameter values.

Layer	N	$\bar{k}$	$\bar{c}$	γ	m	T
Internet Layer 1	37563	5.06	0.63	2.1	1.0	0.5
Internet Layer 2	5163	5.21	0.55	2.1	1.0	0.5
Air/Train Layer 1	69	5.22	0.79	2.6	1.0	0.005
Air/Train Layer 2	69	9.33	0.48	2.9	1.0	0.4
Drosophila Layer 1	838	4.43	0.28	2.6	0.5	0.68
Drosophila Layer 2	755	3.77	0.29	2.8	0.5	0.65
C. Elegans Layer 1	253	4.06	0.24	2.9	2.0	0.65
C. Elegans Layer 2	260	6.83	0.21	2.9	3.4	0.7
Brain Layer 1	85	5.41	0.49	6.0	2.7	0.4
Brain Layer 2	78	5.48	0.40	6.0	1.0	0.5
arXiv Layer 1	2956	4.13	0.83	2.6	2.0	0.05
arXiv Layer 2	3506	4.19	0.81	2.6	2.0	0.05

Table 5.2: Topological properties and HyperMap parameter values for the considered empirical multiplex networks.

HyperMap is based on Maximum Likelihood Estimation. It finds the radial and angular coordinates r_i, θ_i for all nodes $i \leq N$, which maximize the likelihood

$$\mathcal{L} = \prod_{1 \leq j < i \leq N} p(x_{ij})^{\alpha_{ij}} \left[1 - p(x_{ij})\right]^{1-\alpha_{ij}}, \tag{5.1}$$

where the product goes over all node pairs i, j in the network, x_{ij} is the hyperbolic distance between pair i, j,

$$\begin{aligned} x_{ij} \quad &= \quad \operatorname{arccosh}\left(\cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta\theta_{ij}\right) \\ &\approx \quad r_i + r_j + 2 \ln \sin\left(\Delta\theta_{ij}/2\right), \\ &\approx \quad r_i + r_j + 2 \ln\left(\Delta\theta_{ij}/2\right), \\ \text{where} \quad &\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||, \end{aligned} \tag{5.2}$$

and $p(x_{ij})$ is the Fermi-Dirac connection probability,

$$p(x_{ij}) = \frac{1}{1 + e^{\frac{1}{2T}(x_{ij} - R)}}, \tag{5.3}$$

where $R \sim \ln N$. To efficiently and accurately maximize the likelihood in Eq. (5.1) the method follows the techniques described in [127, 128]. Here, we have used the most recent version of the method described in [128].

5.3 Geometric correlations in real multiplex networks

For each real multiplex we consider, we map each network layer independently to an underlying hyperbolic space using the *HyperMap* method described in the previous section, thus

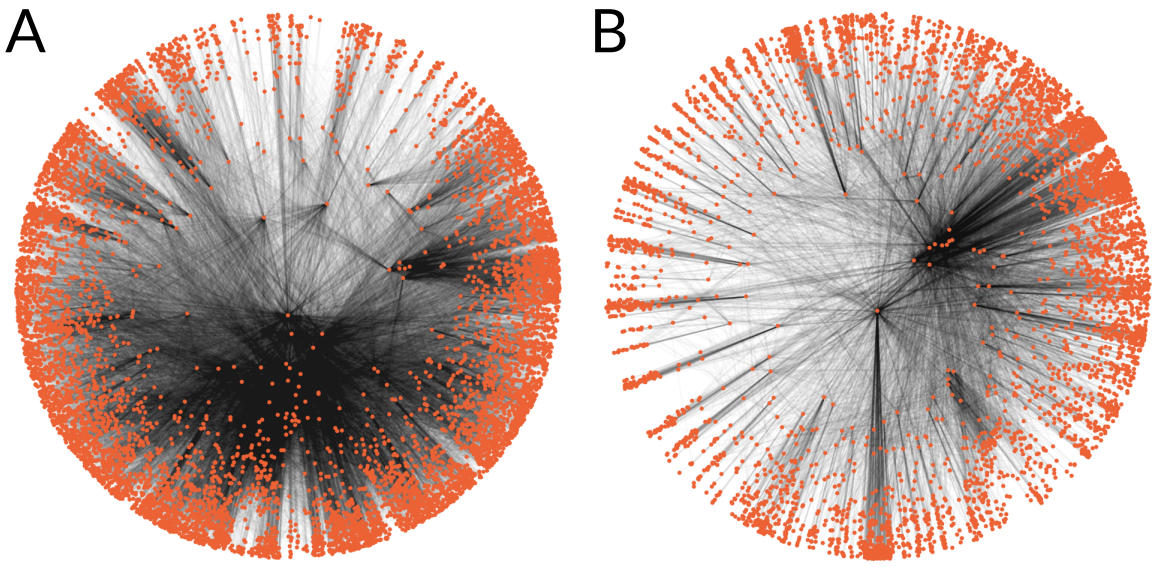

Figure 5.1: Hyperbolic mapping of the IPv4/IPv6 Internet. **A:** IPv4 topology—for clarity only nodes with degrees greater than 3 are shown. **B:** IPv6 topology.

inferring the popularity and similarity coordinates r, θ of all of its nodes. A visualization of the mapped IPv4 and IPv6 Internet layers is shown in Fig. 5.1. In all systems considered we find that node coordinates across layers are not independent. Specifically, we find that both the radial and the angular coordinates of nodes that exist in different layers are correlated, which we discuss in detail in the following.

5.3.1 Radial correlations

The radial popularity coordinate of a node i depends on its observed degree in the network k_i via $r_i \sim \ln N - \ln k_i$, where N is the total number of nodes [126–128]. Therefore, radial correlations are equivalent to correlations between node degrees, which have been recently found and studied [138–142]. Consistent with these findings, radial correlations are present in our real multiplexes and are encoded in the conditional probability $P(r_2|r_1)$, which is the probability that a node has radial coordinate r_2 in layer 2 given its radial coordinate r_1 in layer 1. $P(r_2|r_1)$ for the Internet is shown in Fig. 5.2, where we observe strong correlations between the radial coordinates of nodes in the IPv4 and IPv6 topologies.

5.3.2 Angular correlations and multidimensional communities

We find that in addition to the radial correlations the angular coordinates of nodes in different layers are also correlated. This is a fundamentally new result that has important practical implications. Fig. 5.3 shows the distribution of nodes that have angular coordinates θ_1, θ_2 in layers $1, 2$ of the real multiplexes. The figure also shows the corresponding

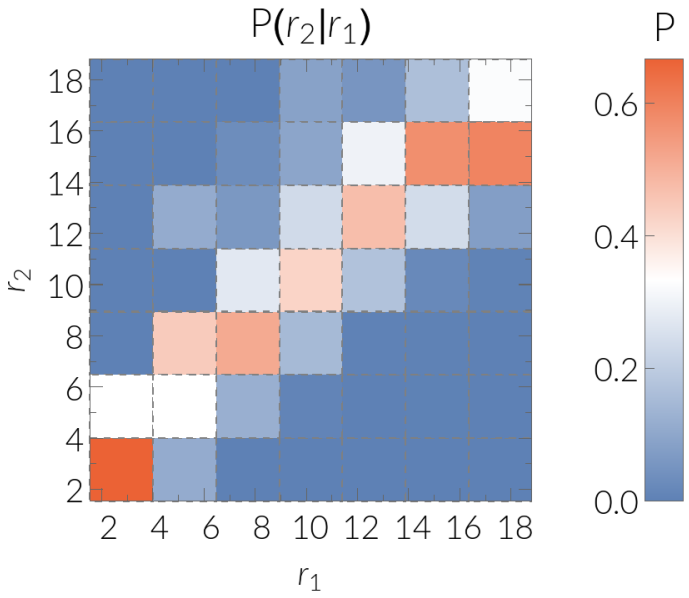

Figure 5.2: Conditional probability $P(r_2|r_1)$ that an AS has radial coordinate r_2 in the IPv6 topology given its radial coordinate r_1 in the IPv4 topology.

distributions in the reshuffled counterparts of the real systems, where we have destroyed the trans-layer coordinate correlations by randomly reshuffling node ids. Specifically, to destroy the geometric correlations in our real multiplexes, we randomly reshuffle the trans-layer node-to-node mappings. For each real multiplex we select one of its layers and we interchange the id of each node of the layer with the id of a randomly selected node from the same layer. The idea behind this process is that if a node with id i is node n_1 in layer 1 and node n_2 in layer 2 with correlated coordinates (r_{n_1}, θ_{n_1}), (r_{n_2}, θ_{n_2}), then, after reshuffling layer 2, the node will become some other node n_2' in this layer, with coordinates $(r_{n_2'}, \theta_{n_2'})$ that will not be correlated with (r_{n_1}, θ_{n_1}). We note that this reshuffling process is just a random id interchange among the nodes of a layer and does not alter the layer's topology. We used this process to create the reshuffled counterparts of the real multiplexes considered here. The reshuffled counterparts serve as a null model for what one would expect if there were no geometric correlations among the layers. From Fig. 5.3, we observe an overabundance of two-dimensional similarity clusters in the real multiplexes. These clusters consist of nodes that are similar, i.e., are located at small angular distances, in both layers of the multiplex. These similarity clusters do not exist in the reshuffled counterparts of the real systems, and are evidence of angular correlations.

The generalization of community definition and detection techniques from single layer networks to multiplex systems has recently gained attention [136, 143, 144]. Our approach here allows one to naturally define and detect multidimensional communities of nodes, which are sets of nodes that are similar, that is, close in the angular similarity space, in multiple layers simultaneously. Furthermore, it also provides a measure of distance between different

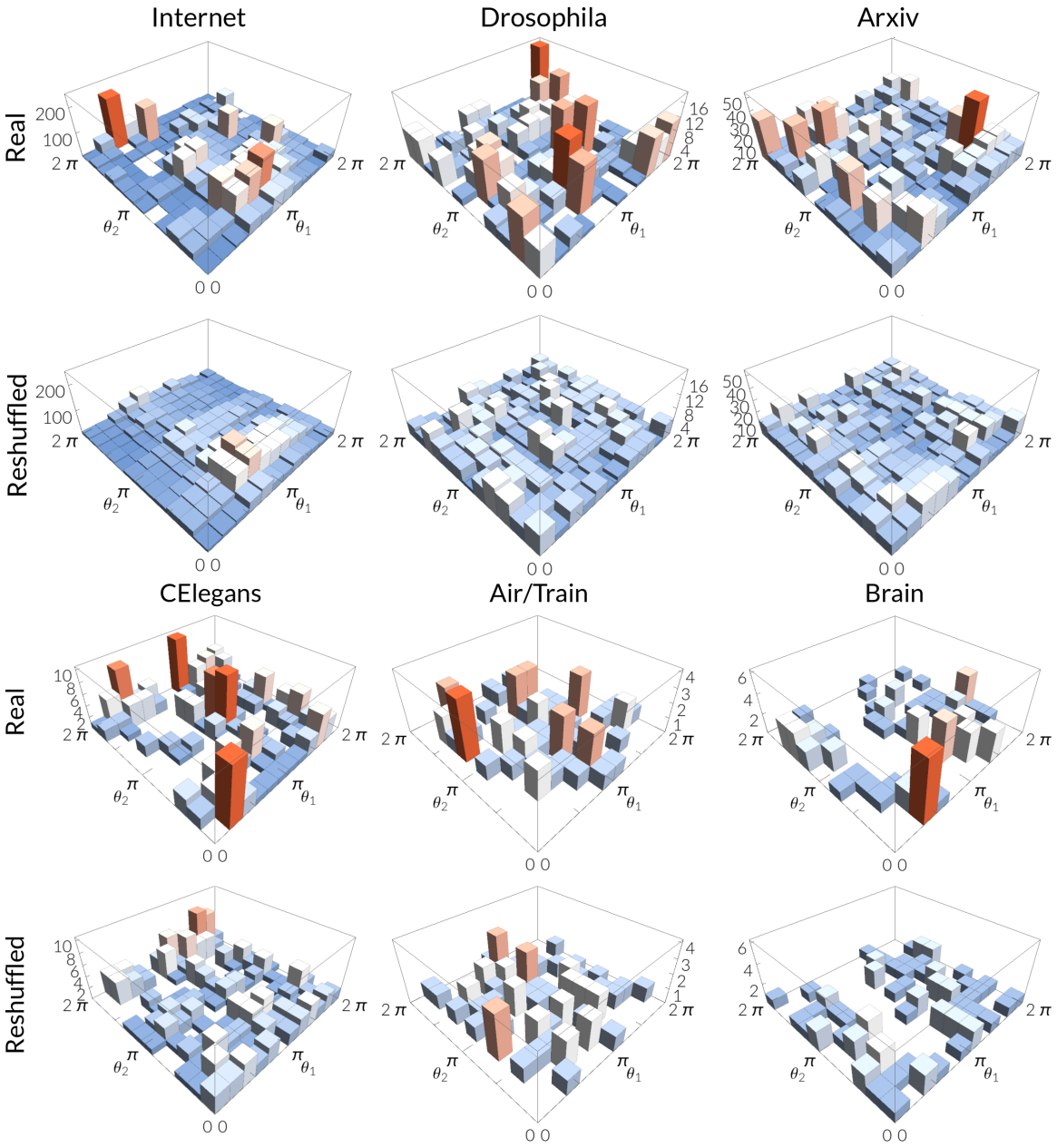

Figure 5.3: Distribution of nodes in the two-dimensional similarity space of the Internet, Drosophila, and arXiv multiplexes **(top row)** and C. Elegans, Air/Train, and human brain multiplexes **(third row)**. The plots correspond to nodes that exist in both layers of each system. The angular similarity coordinate of a node in layer 1 is denoted by θ_1 and in layer 2 by θ_2. The histogram heights are equal to the number of nodes falling within each two-dimensional similarity bin, and the colors in each case denote the relative magnitude of the heights. **Second and fourth row:** The same distributions as above but for the reshuffled counterparts of the real systems.

communities, cf. Fig. 5.3.

In [126–128], the authors have considered the IPv4 Internet topology. They have shown that the mapping of the topology to its underlying hyperbolic space yields meaningful results, since ASs belonging to the same country are mapped close to each other. Specifically, for the majority of the countries, they have shown that their ASs are localized in narrow angular (similarity) regions. The reason for this effect is that ASs belonging to the same country are usually connected more densely to each other than to the rest of the world, and the mapping method (HyperMap) correctly places all such ASs in narrow regions close to each other. We note that other reasons besides geographic proximity may affect the connectivity between ASs, such as economical, political, and performance-related ones. The mapping method does not favor any specific reason, but relies only on the connectivity between nodes (ASs in this case) in order to place the nodes at the right angular (and hyperbolic) distances.

In Fig. 5.4, we observe a similar effect in the two-dimensional similarity space of the IPv4/IPv6 Internet. The figure shows the distribution of ASs belonging to different regions and countries. The AS-to-country mapping is taken from the CAIDA AS Organizations Dataset [145]. In Fig. 5.4, we can see ASs from regions/countries that are narrowly distributed in the two-dimensional similarity space, as well as ASs from regions/countries that are more widely spread. The former group of ASs are the ASs that form strong communities, i.e., that are densely connected to each other, in both the IPv4 and IPv6 topologies. In the figure, these are the ASs belonging to the Post-Soviet and South America regions (Fig. 5.4A), as well as the ASs belonging to some distinct counties such as Austria, Japan, China, Taiwan (Figs. 5.4B,C). By contrast, ASs belonging to the US and Europe are more widely spread in both the IPv4 and IPv6 similarity spaces (Figs. 5.4A,C). We note that Europe in Fig. 5.4A represents not one country but a collection of 9 different countries. Finally, we also observe that there can be ASs from countries that are narrowly distributed in the one similarity space, but not in the other. This is the case for example with the ASs belonging to Poland, which are narrowly distributed in the IPv4 space but not in the IPv6 (Fig. 5.4D). This suggests that these ASs do not form a strong community in IPv6, while they do in IPv4.

5.4 Trans-layer link prediction

The radial and angular correlations across different layers suggest that the hyperbolic distances among nodes are also correlated. Since nodes that are closer hyperbolically have higher chances of being connected, we expect that by knowing the hyperbolic distances between nodes in one layer we can predict the likelihood that the nodes are connected in the other layer. Fig. 5.5 validates that this is indeed the case. The figure shows the empirical trans-layer connection probability, $P(1|2)$ ($P(2|1)$), that two nodes are connected in one of the layers of the multiplex, given their hyperbolic distance in the other layer.

To compute the trans-layer connection probability, we consider all nodes that exist in both layers. In each of the layers, we bin the range of hyperbolic distances between these

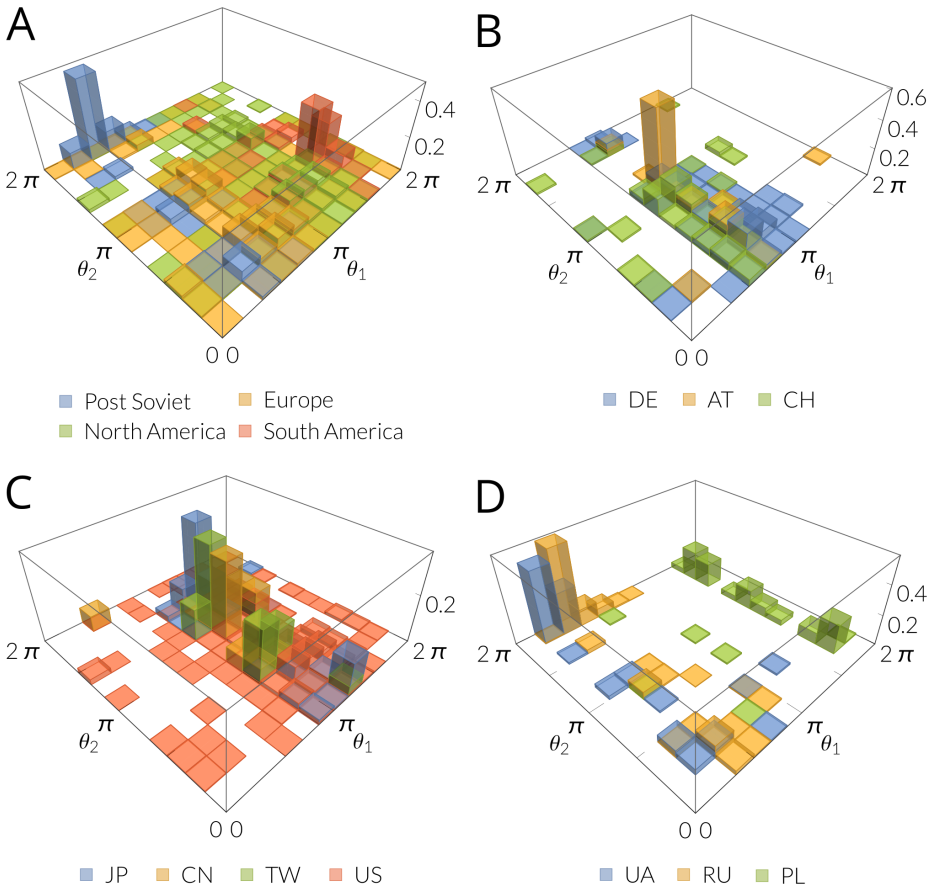

Figure 5.4: Distribution of ASs of the same region/country in the two-dimensional similarity space of the IPv4/IPv6 Internet. The plots correspond to ASs belonging to different regions/countries, which exist both in IPv4 and IPv6. The angular similarity coordinate of an AS in IPv4 is denoted by θ_1 and in IPv6 by θ_2. For each region/country, the histogram heights are normalized by the total number of ASs that belong to the region/country. In **A**, "Post Soviet" corresponds to the ASs belonging to Russia, Ukraine, Estonia, and Latvia; "Europe" corresponds to the ASs belonging to Germany, France, Spain, Finland, Austria, Netherlands, Sweden, Italy, and Greece; "North America" corresponds to the ASs belonging to the USA and Canada; and "South America" corresponds to the ASs belonging to Brazil, Uruguay, Argentina, and Colombia. The histograms in **B-D** correspond to the ASs belonging to 10 distinct countries. In **B**, the countries are Germany (DE), Austria (AT), and Switzerland (CH); in **C**, the countries are Japan (JP), China (CN), Taiwan (TW), and USA (US); and in **D**, the countries are Ukraine (UA), Russia (RU), and Poland (PL).

nodes from zero to the maximum distance into small bins. For each bin we then find all the node pairs located at the hyperbolic distances falling within the bin. The percentage of pairs in this set of pairs that are connected by a link in the other layer is the value of the empirical trans-layer connection probability at the bin. We observe (Fig. 5.5) that this probability decreases with the hyperbolic distance between nodes in all the real multiplexes. By contrast, in their reshuffled counterparts, which do not exhibit geometric correlations, this probability is almost a straight line. Fig. 5.5 shows that the trans-layer connection probability decreases with the angular distance between nodes, which provides an alternative empirical validation of the existence of strong similarity correlations across the layers.

The problem of link prediction has been studied extensively in the context of predicting missing and future links in single layer networks [146, 147]. Its generalization to real-world multilayer systems is recently gaining attention [148]. The trans-layer link prediction approach we described here is quite general, i.e., applicable to any real multiplex with geometric structure, and allows one to estimate the most probable connections among nodes in one layer of the multiplex, by knowing the hyperbolic distances among the same nodes in another layer.

5.5 Summary: Hidden geometric correlations, community detection, and link prediction

Numerous real-world systems are multiplex networks where nodes in one network layer can simultaneously exist in other network layers. Each single network layer can be mapped into its own hyperbolic space, where node coordinates abstract the popularity and similarity of nodes. We have found that in different real multiplexes the coordinates of nodes in different layers are correlated, meaning that the underlying hyperbolic distances of the single layers are also correlated.

Our findings yield a very powerful and general framework for understanding and analyzing real multiplexes. Specifically, we have shown that one can define and detect multidimensional communities, which are sets of nodes that are simultaneously similar in multiple layers. We have found that such communities are overabundant in different real multiplexes compared to their reshuffled counterparts that do not exhibit geometric correlations. Furthermore, we have also shown that one can facilitate trans-layer link prediction, where the most probable connections in one layer can be predicted by knowing the hyperbolic distances among the nodes in some other layer.

Our findings can have important applications ranging from understanding functional and structural brain networks and deciphering their precise relationship(s) [123] to predicting links among nodes (e.g., terrorists) in a specific network by knowing their connectivity in some other network. Finally, the discovered correlations are crucial for improving navigation using multiple layers simultaneously, as we show in the following chapter.

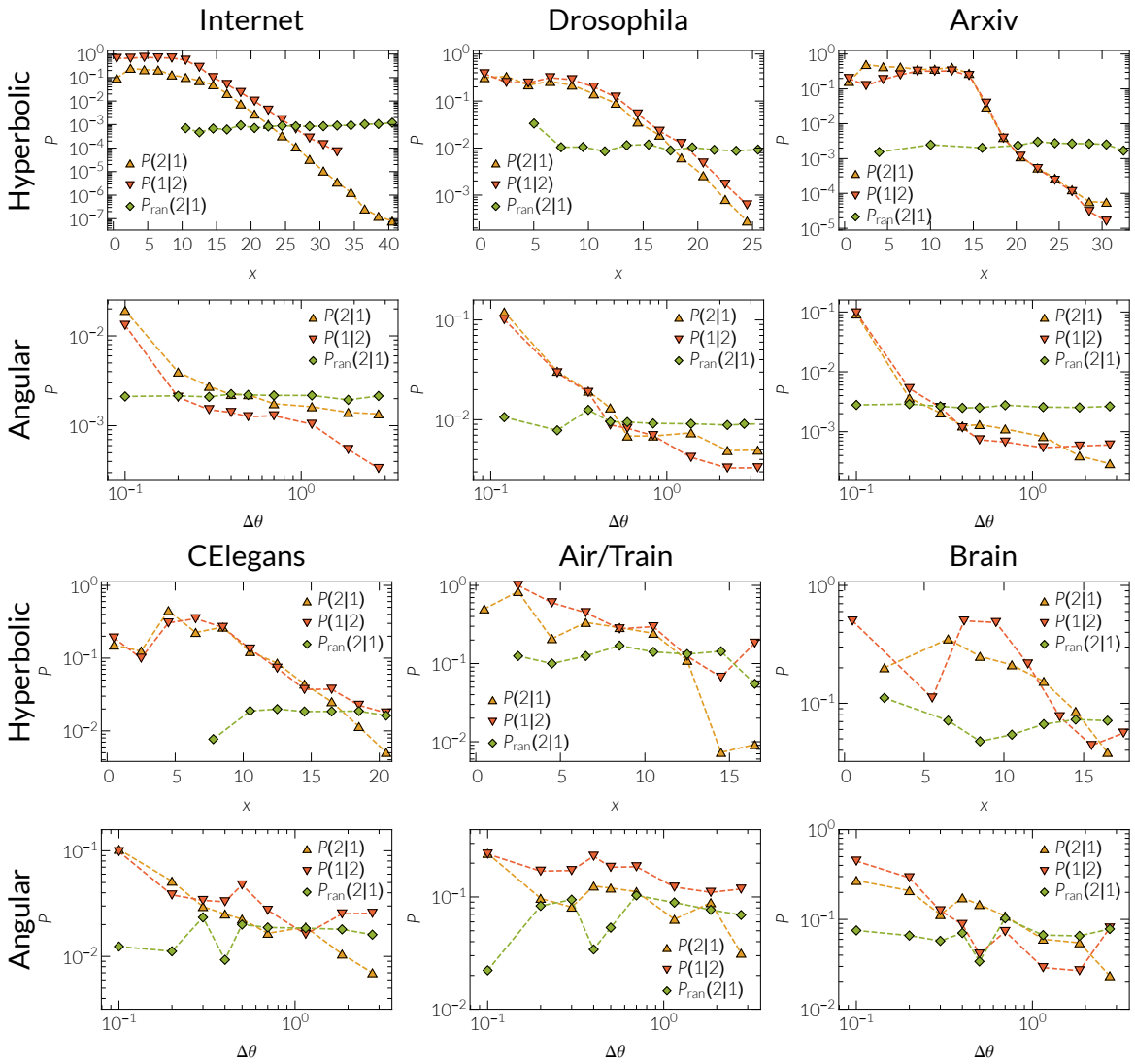

Figure 5.5: Trans-layer connection probability in the considered multiplexes. **Top row:** Trans-layer connection probability as a function of hyperbolic distance for the Internet, Drosophila, and arXiv multiplexes. $P(j|i)$ denotes the probability that a pair of nodes is connected in layer j given its hyperbolic distance x in layer i. $P_{\text{ran}}(j|i)$ denotes the same probability for the reshuffled counterpart of each real system. **Second row:** Corresponding trans-layer connection probabilities when considering only the angular (similarity) distance between nodes, $\Delta\theta$. **Third row:** Trans-layer connection probability as a function of hyperbolic distance for the C. Elegans, Air/Train, and human brain multiplexes. **Fourth row:** Corresponding trans-layer connection probabilities when considering only the angular (similarity) distance between nodes, $\Delta\theta$.

6 Hidden geometric correlations facilitate navigation

"My – it's a small world"
UNKNOWN

This chapter is – with some small changes – available as a preprint at [116] and currently under review. Most figures are identical to the preprint.

6.1 Introduction

In this chapter, we consider targeted navigation that uses only local knowledge. Targeted navigation is a key function of many real networks, where either goods, people, or information is transferred from a source to a destination using the connections of the network. It has been shown that single complex networks, like the IPv4 Internet or the network of airport connections, can be navigated efficiently by performing *greedy routing* in their underlying geometric spaces [111,125,126,149]. In greedy routing, nodes forward a message to their neighbor that is closest to the destination in the geometric space. The message either reaches its target, or it enters a loop, i.e. the message is given back to a node it already visited, and the delivery fails. To study navigation in multiplex systems, we extend the notion of greedy routing so that a node forwards a message to its neighbor that is closest to the destination in any of the layers comprising the system. We call this process *mutual greedy routing*.

Mutual greedy routing follows the same line of reasoning as greedy routing in Milgram's experiment [150], explained in detail in the introduction in section 1.2.1. For example, in the case of a single network, to reach a lawyer in Boston one might want to forward a message to a judge in Los Angeles (greedy routing). However, in the case of two network layers, it might be known that the lawyer in Boston is also a passionate vintage model train collector. An individual who knows a judge in Los Angeles and the owner of a vintage model train shop in New York, who might be attending all the vintage train meetings, would probably choose to forward the message to the latter (mutual greedy routing). Similarly, air travel networks can be supported by train networks to enhance the possibilities to navigate the physical world, individuals can use different online social networks to increase their outreach, and so on. In this chapter, we consider the real Internet, which is used to navigate the digital world, and show that mutual greedy routing in the multiplex consisting of the IPv4 and IPv6 Internet topologies [131] outperforms greedy routing in the single IPv4 and IPv6 networks. We also use synthetic model networks to show that geometric correlations improve the navigation

of multiplex systems, which outperforms navigation in the single layers if these correlations are sufficiently strong. In this context, we also investigate under which conditions adding more layers improves the navigability of a multiplex system.

6.2 Modeling geometric correlations and implications to mutual greedy routing

6.2.1 Mutual greedy routing in the Internet multiplex

The IPv4 Internet has been found to be navigable [126–128]. Specifically, it has been shown that greedy routing (GR) could reach destinations with more than 90% success rate in the constructed hyperbolic maps of the IPv4 topology in 2009. We find a similar efficiency of GR in both the IPv4 and IPv6 topologies considered here, which correspond to January 2015. Specifically, we perform GR in the hyperbolic map of each topology among 10^5 randomly selected source-destination pairs that exist in both topologies. We find that GR reaches destinations with 90% success rate in IPv4, and with 92% success rate in IPv6. Furthermore, we also perform angular GR, which is the same as GR but uses only the angular distances. We find that the success rate in this case is almost 60% in both the IPv4 and IPv6 topologies. We now consider mutual greedy routing. For any number of layers, (layer 1, ..., layer n), a node with a message first computes the distance between its neighbors and the destination of the message in layer 1, then it does the same for its neighbors and the destination in layer 2, and so on. The node then forwards the message to the neighbor that has the smallest distance to the destination across all computed distances. If a message is given back to a node it already visited, the delivery fails. The success rate is the percentage of messages that reach their destinations. We distinguish between hyperbolic mutual greedy routing (that hereafter we refer to as mutual GR) and angular mutual greedy routing, which uses only the angular distances. Hyperbolic mutual greedy routing between the same source-destination pairs as in the single layer case increases the success rate to 95%, while angular mutual GR increases the success rate along the angular direction to 66%. We are interested in angular mutual GR because its performance depends only on the angular similarity coordinates of nodes in the different layers[1].

The observations above raise the following fundamental questions. (i) How do the radial and angular correlations affect the performance of mutual GR? (ii) Under which conditions does mutual GR perform better than single-layer GR? (iii) How does the performance of mutual GR depend on the number of layers in a multiplex system? And (iv), how close to the optimal—in terms of mutual GR's performance—are the geometric correlations in the IPv4/IPv6 Internet? Answering these questions requires a framework to construct realistic synthetic topologies (layers) where correlations—both radial and angular—can be tuned without altering the topological characteristics of each individual layer. We develop such a framework in the following.

[1] In some situations nodes may not have knowledge of the popularity of their neighbors and hence can only rely on the angular component.

6.2.2 Modeling multiplex networks with geometric correlations

Our framework builds on the (single-layer) network construction procedure prescribed by the Newtonian $\mathbb{S}^1$ [110] and hyperbolic $\mathbb{H}^2$ models [125]. The two models are isomorphic and here we present the results for the $\mathbb{H}^2$ version even if for calculations it is more convenient to make use of the $\mathbb{S}^1$. We recall that to construct a network of size N, the $\mathbb{H}^2$ model first assigns to each node $i = 1, \ldots, N$ its popularity and similarity coordinates r_i, θ_i. Subsequently, it connects each pair of nodes i, j with probability $p(x_{ij}) = 1/(1+e^{\frac{1}{2T}(x_{ij}-R)})$, where x_{ij} is the hyperbolic distance between the nodes and $R \sim \ln N$. The connection probability $p(x_{ij})$ is nothing but the Fermi-Dirac distribution. Parameter T is the *temperature* and controls clustering in the network [151], which is the probability that two neighbors of a node are connected. The average clustering $\bar{c}$ is maximized at $T = 0$, nearly linearly decreases to zero with $T \in [0, 1)$, and is asymptotically zero if $T > 1$. It has been shown that the $\mathbb{S}^1$ and $\mathbb{H}^2$ models can construct synthetic networks that resemble real networks across a wide range of structural characteristics, including power law degree distributions and strong clustering [110, 125]. Our framework constructs single-layer topologies using these models, and allows for radial and angular coordinate correlations across the different layers. The strength of these correlations can be tuned via model parameters $\nu \in [0, 1]$ and $g \in [0, 1]$, without affecting the topological characteristics of the individual layers, which can have different properties and different sizes. The radial correlations increase with parameter ν—at $\nu = 0$ there are no radial correlations, while at $\nu = 1$ radial correlations are maximized. Similarly, the angular correlations increase with parameter g—at $g = 0$ there are no angular correlations, while at $g = 1$ angular correlations are maximized. In the following we present the details of our modeling framework.

6.2.3 $\mathbb{S}^1/\mathbb{H}^2$ model of single-layer networks

Instead of working directly with the $\mathbb{H}^2$ model, we make use of the $\mathbb{S}^1$ model [110] that is more convenient to work with, and which is isomorphic to the $\mathbb{H}^2$ model through a simple change of variables [125]. We first review the $\mathbb{S}^1$ model and its relation to the $\mathbb{H}^2$ model. Instead of radial and angular coordinates r_i, θ_i, each node i in the $\mathbb{S}^1$ model has hidden variables κ_i, θ_i. The hidden variable κ_i is the node's expected degree in the resulting network, while θ_i is the angular (similarity) coordinate of the node on a circle of radius $N/2\pi$, where N is the total number of nodes. To construct a network with the $\mathbb{S}^1$ model that has size N, average node degree $\bar{k}$, power law degree distribution with exponent $\gamma > 2$, and temperature $T \in [0, 1)$, we perform the following steps:

i. Sample the angular coordinates of nodes θ_i, $i = 1, 2, \ldots, N$, uniformly at random from $[0, 2\pi]$, and their hidden variables κ_i, $i = 1, 2, \ldots, N$, from the probability density function (PDF)

$$\begin{aligned} \rho(\kappa) &= (\gamma - 1)\kappa^{\min^{\gamma-1}}\kappa^{-\gamma}, \\ \kappa^{\min} &= \bar{k}\frac{\gamma - 2}{\gamma - 1}, \end{aligned} \tag{6.1}$$

where $\kappa^{\min}$ is the expected minimum node degree, which is a function of the average degree $\bar{k}$; [2]

ii. Connect every pair of nodes i, j with probability

$$\begin{aligned} r(\kappa_i, \theta_i; \kappa_j, \theta_j) &= \frac{1}{1 + \left[\frac{d(\theta_i,\theta_j)}{\mu\kappa_i\kappa_j}\right]^{\frac{1}{T}}}, \\ d(\theta_i, \theta_j) &= \frac{N}{2\pi}\Delta\theta_{ij},\ \Delta\theta_{ij} = |\pi - |\pi - |\theta_i - \theta_j|||, \\ \mu &= \frac{\sin T\pi}{2\bar{k}T\pi}, \end{aligned} \tag{6.2}$$

where $d(\theta_i, \theta_j)$ is the angular distance between nodes i, j on the circle.

The $\mathbb{S}^1$ model is equivalent to the $\mathbb{H}^2$ model after transforming the expected node degrees κ_i to radial coordinates r_i via

$$r_i = R - 2\ln\frac{\kappa_i}{\kappa^{\min}}, \tag{6.3}$$

where R is the radius of the hyperbolic disc in the $\mathbb{H}^2$ model within which all nodes reside,

$$\begin{aligned} R &= 2\ln\frac{N}{c}, \\ c &= \bar{k}\frac{\sin T\pi}{2T}\left(\frac{\gamma-2}{\gamma-1}\right)^2. \end{aligned} \tag{6.4}$$

It is easy to see that after the above change of variables the connection probability in Eq. (6.2) becomes the Fermi-Dirac connection probability in the $\mathbb{H}^2$ model,

$$p(x_{ij}) = \frac{1}{1 + e^{\frac{1}{2T}(x_{ij}-R)}}, \tag{6.5}$$

where $x_{ij} \approx r_i + r_j + 2\ln\frac{\Delta\theta_{ij}}{2}$ is the hyperbolic distance between nodes i, j [125]. We note that without loss of generality, we use here a hyperbolic plane of curvature $K = -1$. See [125] for further details.

6.2.4 Two-layer multiplex model

We now describe our framework for constructing a two-layer multiplex system with geometric correlations. Each single-layer (layer 1, layer 2) is constructed according to the $\mathbb{S}^1$ model, and we account for correlations among the hidden variables of nodes in the two layers, whose strength can be tuned. The extension of the framework to more than two layers is straightforward and described in 6.2.5. In a nutshell, our framework consists of the following steps:

[2] By sampling from a PDF $f(x)$ we mean that we first compute the CDF $F(x) = \int_{x_{\min}}^{x} \mathrm{d}x'\rho(x')$, where $x_{\min}$ is the minimum value of x, then generate a random number u_i uniformly at random from $[0, 1]$, and finally compute the value x_i such that $F(x_i) = u_i$. The value x_i is a sample from the PDF $\rho(x)$ (or the CDF $F(x)$).

i. Assignment of hidden variables $\kappa_{1,i}, \theta_{1,i}$ to each node i in layer 1 like in the $\mathbb{S}^1$ model (Eqs. (6.6), (6.8));

ii. Assignment of hidden variables $\kappa_{2,i}, \theta_{2,i}$ to each node i in layer 2, depending on the node's hidden variables in layer 1 (Eqs. (6.11), (6.22))—the assignment here is done such that the marginal (unconditional) distribution of $\kappa_{2,i}, \theta_{2,i}$ is still the one prescribed by the $\mathbb{S}^1$ model (Eqs. (6.9), (6.8));

iii. Creation of edges, by connecting node pairs in each layer with the corresponding $\mathbb{S}^1$ connection probability, which depends exclusively on the assigned hidden variables of nodes in each layer (Eqs. (6.24), (6.25));

iv. $\mathbb{S}^1$-to-$\mathbb{H}^2$ transformation, by mapping the hidden variables $\kappa_{1,i}, \kappa_{2,i}$ to radial coordinates $r_{1,i}, r_{2,i}$ (Eqs. (6.26), (6.27)).

Below, we describe these steps in detail. We assume that the two layers have the same number of nodes $N_1 = N_2 = N$. The extension of the framework to multiplexes with different layer sizes is described in 6.5.1.

i. Assignment of hidden variables in layer 1. For each node $i = 1, 2, \ldots, N$ in layer 1 we sample its hidden variable $\kappa_{1,i}$ from the PDF

$$\rho_1(\kappa_1) = (\gamma_1 - 1)\kappa_1^{\min \gamma_1 - 1}\kappa_1^{-\gamma_1}, \tag{6.6}$$

$$\kappa_1^{\min} = \bar{k}_1 \frac{\gamma_1 - 2}{\gamma_1 - 1}, \tag{6.7}$$

where $\bar{k}_1$ and $\gamma_1 > 2$ are respectively the target average degree and power law degree distribution exponent in layer 1. The angular coordinate $\theta_{1,i}$ of each node $i = 1, 2, \ldots, N$, is sampled from the uniform PDF

$$f(\theta) = \frac{1}{2\pi},\ \theta \in [0, 2\pi). \tag{6.8}$$

ii. Assignment of hidden variables in layer 2. We now want to assign to each node $i = 1, 2, \ldots, N$ its hidden variable $\kappa_{2,i}$ in layer 2, conditioned on the value of its hidden variable $\kappa_{1,i}$ in layer 1. At the same time, we want the $\kappa_{2,i}$'s to satisfy the marginal (unconditional) PDF

$$\rho_2(\kappa_2) = (\gamma_2 - 1)\kappa_2^{\min \gamma_2 - 1}\kappa_2^{-\gamma_2}, \tag{6.9}$$

$$\kappa_2^{\min} = \bar{k}_2 \frac{\gamma_2 - 2}{\gamma_2 - 1}, \tag{6.10}$$

where $\bar{k}_2$ and $\gamma_2 > 2$ are respectively the target average degree and power law degree distribution exponent in layer 2. Eq. (6.9) should be satisfied irrespectively of the correlation strength between the $\kappa_{2,i}$ and $\kappa_{1,i}$. To accomplish this, we sample the hidden variable $\kappa_{2,i}$

of each node $i = 1, 2, \ldots, N$, from the conditional cumulative distribution function (CDF)

$$F_\nu(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = e^{-(\varphi_1^{1/(1-\nu)}+\varphi_2^{1/(1-\nu)})^{1-\nu}} \times \\ \times \left[\varphi_1^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)}\right]^{-\nu} \frac{\varphi_1^{\nu/(1-\nu)} \kappa_1^{\min} \kappa_1^{\gamma_1}}{\kappa_1^{\min} \kappa_1^{\gamma_1} - \kappa_1^{\min^{\gamma_1}} \kappa_1}, \tag{6.11}$$

$$\varphi_i = -\ln\left[1 - (\kappa_i^{\min}/\kappa_i)^{\gamma_i - 1}\right], \text{ for } i = 1, 2, \tag{6.12}$$

where κ_1 is the value of the hidden variable of the node in layer 1, $\{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}$ are the network parameters defined earlier, and $\nu \in [0, 1]$ is the correlation strength parameter. The higher the value of ν the stronger is the correlation between $\kappa_{2,i}$ and $\kappa_{1,i}$. It is easy to see that when $\nu = 0$ (no correlation between $\kappa_{2,i}$ and $\kappa_{1,i}$), Eq. (6.11) becomes the marginal CDF of $\kappa_{2,i}$ given in Eq. (6.16) below. On the other hand, when $\nu \to 1$ (maximally correlated $\kappa_{2,i}$ and $\kappa_{1,i}$), Eq. (6.11) becomes

$$F_\nu(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \Theta\left[\kappa_2 - \kappa_2^{\min}\left(\frac{\kappa_1}{\kappa_1^{\min}}\right)^{(1-\gamma_1)/(1-\gamma_2)}\right], \tag{6.13}$$

where $\Theta[x]$ denotes the Heaviside step function. That is, when $\nu \to 1$

$$\kappa_{2,i} = \kappa_2^{\min}\left(\frac{\kappa_{1,i}}{\kappa_1^{\min}}\right)^{(1-\gamma_1)/(1-\gamma_2)}, \tag{6.14}$$

which yields $\kappa_{2,i} = \kappa_{1,i}$ if $\kappa_2^{\min} = \kappa_1^{\min}$ and $\gamma_1 = \gamma_2$.

To derive Eq. (6.11) we use copulas [152]. Copulas are multivariate probability distributions used to describe the dependence between random variables. In particular, any multivariate CDF $F(\kappa_1, ..., \kappa_n)$ of n random variables $\kappa_1, ..., \kappa_n$, can be written in the form $F(\kappa_1, ..., \kappa_n) = C(F_1(\kappa_1), ..., F_n(\kappa_n))$ where $F_1(\kappa_1), ..., F_n(\kappa_n)$ are the marginal CDFs of $F(\kappa_1, ..., \kappa_n)$, and C is called a copula. Each of the marginals of C is uniform in $[0, 1]$, and there are many parametric copula families available, which have parameters that control the strength of the dependence between the random variables [152].

In our case, the random variables are the node hidden variables κ_1, κ_2 in layers 1 and 2, whose marginal CDFs can be computed from Eqs. (6.6), (6.9),

$$F_1(\kappa_1) = 1 - \kappa_1^{(1-\gamma_1)} \kappa_1^{\min^{(\gamma_1 - 1)}}, \tag{6.15}$$

$$F_2(\kappa_2) = 1 - \kappa_2^{(1-\gamma_2)} \kappa_2^{\min^{(\gamma_2 - 1)}}. \tag{6.16}$$

For the copula function C, we use the bivariate Gumbel-Hougaard copula [152], defined as

$$C_\eta(u, v) = e^{-[(-\ln u)^\eta + (-\ln v)^\eta]^{1/\eta}}, \tag{6.17}$$

$$\eta \equiv \frac{1}{1-\nu} \in [1, \infty).$$

Hence, our copula reads

$$C_\eta(F_1(\kappa_1), F_2(\kappa_2)) = e^{-[(-\ln F_1(\kappa_1))^\eta + (-\ln F_2(\kappa_2))^\eta]^{1/\eta}}. \tag{6.18}$$

The joint PDF of κ_1 and κ_2, $\rho_\eta(\kappa_1, \kappa_2, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$, can be obtained by differentiating the copula with respect to κ_1, κ_2,

$$\rho_\eta(\kappa_1, \kappa_2, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \frac{\partial^2 C_\eta(F_1(\kappa_1), F_2(\kappa_2))}{\partial\kappa_1 \partial\kappa_2}, \tag{6.19}$$

while the conditional PDF $\rho_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$ can be written as

$$\rho_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \rho_\eta(\kappa_1, \kappa_2, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) \frac{1}{\rho_1(\kappa_1)}. \tag{6.20}$$

The conditional CDF $F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$ can be therefore computed as

$$\begin{aligned} F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) &= \int_{\kappa_2^{\min}}^{\kappa_2} \mathrm{d}\kappa' \, \rho_\eta(\kappa'|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) \\ &= \frac{\partial C_\eta(F_1(\kappa_1), F_2(\kappa_2))}{\partial\kappa_1} \frac{1}{\rho_1(\kappa_1)}, \end{aligned} \tag{6.21}$$

which yields Eq. (6.11).

The angular coordinate $\theta_{2,i}$ of each node $i = 1, 2, \ldots, N$ in layer 2 is obtained by[3]

$$\theta_{2,i} = \mod \left[\theta_{1,i} + \frac{2\pi l_i}{N}, 2\pi\right], \tag{6.22}$$

where $\theta_{1,i}$ is the angular coordinate of the node in layer 1, and l_i is a directed arc length on the $\mathbb{S}^1$ circle of radius $R = N/2\pi$, which is sampled from the zero-mean truncated Gaussian PDF

$$\begin{aligned} f_\sigma(l) &= \frac{\frac{1}{\sigma}\phi\left(\frac{l}{\sigma}\right)}{\Phi\left(\frac{N}{2\sigma}\right) - \Phi\left(-\frac{N}{2\sigma}\right)}, \quad -\frac{N}{2} \le l \le \frac{N}{2}, \\ \sigma &\equiv 100\left(\frac{1}{g} - 1\right), \end{aligned} \tag{6.23}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $\Phi(x) = \int \mathrm{d}x \, \phi(x)$, $\sigma \in (0, \infty)$ is the variance of the PDF, and $g \in [0, 1]$ is the angular correlation strength parameter. The higher the value of g the stronger is the correlation between $\theta_{2,i}$ and $\theta_{1,i}$. When $g \to 0$, $\sigma \to \infty$, $f_\sigma(l)$ becomes the uniform PDF, and $\theta_{2,i}$, $\theta_{1,i}$ are not correlated. When $g = 1$, $\sigma = 0$, and $l_i = 0$, meaning that the angles of each node i are identical in the two layers, $\theta_{2,i} = \theta_{1,i}$.

iii. Creation of edges. Once all node hidden variables are assigned, we connect each node pair i, j in layers 1 and 2 with the corresponding $\mathbb{S}^1$ connection probabilities given in

[3]Note that a rotation of the angular coordinates has no effect on the resulting system.

Eqs. (6.24), (6.25) below,

$$\begin{aligned} r_1(\kappa_{1,i},\theta_{1,i};\kappa_{1,j},\theta_{1,j}) &= \frac{1}{1+\left[\frac{d_1(\theta_{1,i},\theta_{1,j})}{\mu_1\kappa_{1,i}\kappa_{1,j}}\right]^{\frac{1}{T_1}}}, \qquad (6.24)\\ d_1(\theta_{1,i},\theta_{1,j}) &= \frac{N}{2\pi}\Delta\theta_{1,ij},\\ \Delta\theta_{1,ij} &= |\pi-|\pi-|\theta_{1,i}-\theta_{1,j}|||,\\ \mu_1 &= \frac{\sin T_1\pi}{2\bar{k}_1 T\pi},\\ r_2(\kappa_{2,i},\theta_{2,i};\kappa_{2,j},\theta_{2,j}) &= \frac{1}{1+\left[\frac{d_2(\theta_{2,i},\theta_{2,j})}{\mu_2\kappa_{2,i}\kappa_{2,j}}\right]^{\frac{1}{T_2}}}, \qquad (6.25)\\ d_2(\theta_{2,i},\theta_{2,j}) &= \frac{N}{2\pi}\Delta\theta_{2,ij},\\ \Delta\theta_{2,ij} &= |\pi-|\pi-|\theta_{2,i}-\theta_{2,j}|||,\\ \mu_2 &= \frac{\sin T_2\pi}{2\bar{k}_2 T_2\pi}, \end{aligned}$$

where $T_1 \in [0,1), T_2 \in [0,1)$ are the temperatures, which control clustering in each layer. We recall that the average node clustering is maximized at temperature $T = 0$, and nearly linearly decreases to zero with $T \in [0,1)$.

iv. $\mathbb{S}^1$*-to-*$\mathbb{H}^2$ *transformation.* Finally, we map the node hidden variables $\kappa_{1,i}, \kappa_{2,i}$ in layers 1, 2, to radial coordinates $r_{1,i}, r_{2,i}$ using the relations below,

$$\begin{aligned} r_{1,i} &= R_1 - 2\ln\frac{\kappa_{1,i}}{\kappa_1^{\min}},\ R_1 = 2\ln\frac{N}{c_1}, \qquad (6.26)\\ c_1 &= \bar{k}_1\frac{\sin T_1\pi}{2T_1}\left(\frac{\gamma_1-2}{\gamma_1-1}\right)^2,\\ r_{2,i} &= R_2 - 2\ln\frac{\kappa_{2,i}}{\kappa_2^{\min}},\ R_2 = 2\ln\frac{N}{c_2}, \qquad (6.27)\\ c_2 &= \bar{k}_2\frac{\sin T_2\pi}{2T_2}\left(\frac{\gamma_2-2}{\gamma_2-1}\right)^2, \end{aligned}$$

where $\kappa_1^{\min}, \kappa_2^{\min}$ are given in Eqs. (6.7), (6.10).

6.2.5 Modeling more than two layers

To construct a multiplex network consisting of n layers (layer 1, layer 2, ..., layer n), we work in the same way as with the two-layer system described before. Specifically, for each two consecutive layers $j-1, j$, for $2 \le j \le n$, we first fix their radial and angular correlation strength parameters $\nu_{j,j-1} \in [0,1]$, $g_{j,j-1} \in [0,1]$ to some desired values. Subsequently, we assign hidden variables $\kappa_{1,i}, \theta_{1,i}$ to nodes in layer 1 as described earlier (Eqs. (6.6), (6.8)), as well as hidden variables $\kappa_{2,i}, \theta_{2,i}$ to nodes in layer 2, conditioned on $\kappa_{1,i}, \theta_{1,i}$ (Eqs. (6.11),

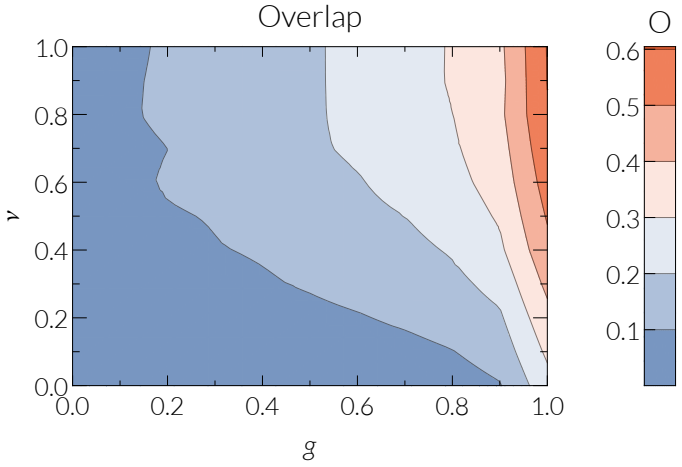

Figure 6.1: Edge overlap O in a two-layer synthetic multiplex as a function of the radial (ν) and angular (g) correlation strengths. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}, \bar{k} = 10$, and temperature $T = 0.4$.

(6.22)). Then, we continue by assigning hidden variables $\kappa_{j,i}, \theta_{j,i}$ to nodes in layer $3 \leq j \leq n$, conditioned on the values of the hidden variables $\kappa_{j-1,i}, \theta_{j-1,i}$ of the nodes in layer $j-1$. This conditional assignment is done in exactly the same manner as the assignment of $\kappa_{2,i}, \theta_{2,i}$, which is conditioned on the values of $\kappa_{1,i}, \theta_{1,i}$. Once all node hidden variables in all layers are assigned, we create edges in each layer by connecting each node pair with the corresponding $\mathbb{S}^1$ connection probability (cf. Eqs. (6.24), (6.25)). Finally, for each layer $1 \leq j \leq n$, we map the node hidden variables $\kappa_{j,i}$ to radial coordinates $r_{j,i}$ as described earlier (cf. Eqs. (6.26), (6.27)). We use this procedure to construct three- and four-layer multiplexes in the following, where we always set the correlation strengths between subsequent layers to the same value, $\nu_{j,j-1} = \nu \in [0, 1]$, $g_{j,j-1} = g \in [0, 1]$, $\forall j \geq 2$.

6.3 Geometric correlations lead to significant edge overlap

Radial and angular correlations across different layers naturally give rise to a significant amount of edge overlap between the layers, as observed in many real multiplexes [62, 121]. The edge overlap O between two layers (layer 1, layer 2) is formally defined as the ratio of the number of overlapping (i.e., common) edges between the layers, to the maximum possible number of common edges [153, 154],

$$O = \frac{\#(\text{overlapping edges})}{\min[\#(\text{edges in layer 1}), \#(\text{edges in layer 2})]} \,. \tag{6.28}$$

Fig. 6.1 shows the edge overlap in a synthetic two-layer multiplex as a function of the radial and angular correlation strength parameters ν, g. We observe that the overlap increases as

we increase the correlation strengths ν, g, and it is maximized at fully correlated coordinates, $\nu = 1, g = 1$. For uncorrelated coordinates, $\nu = 0, g = 0$, the overlap is minimized—it can be shown that when $\nu = 0, g = 0$, the overlap vanishes in the thermodynamic limit ($N \to \infty$), i.e., as the layer sizes increase.

The edge overlap also depends on the temperature of the layers. For fixed values of ν, g, a higher overlap is achieved when the temperature of the layers is lower. Specifically, if two layers have the same parameters $N, \gamma, \bar{k}, T$, and $\nu = g = 1$, i.e., the node coordinates in the two layers are identical, then at $T = 0$ the edge overlap is 100%, i.e., the topologies of the two layers are identical.

6.4 Geometric correlations increase routing performance

To investigate how radial and angular correlations affect the performance of mutual navigation, we consider two-, three- and four-layer multiplexes constructed using the modeling framework introduced earlier with different values of the correlation strength parameters ν and g. Each layer consists of $N = 30000$ nodes, has a power law degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 2.5$, average node degree $\bar{k} = 10$, and the same temperature T that we vary in $(0, 1)$. From Figs. 6.2–6.4 we observe that in general both mutual GR and angular mutual GR perform better as we increase the correlation strengths ν and g.[4]

When both radial and angular correlations are weak we do not observe any significant benefits from mutual navigation. Indeed, in Fig. 6.5 we observe that in the uncorrelated case ($\nu \to 0$, $g \to 0$) mutual GR performs almost identical to the single-layer GR, irrespectively of the number of layers. This is because when a message reaches a node in one layer after the first iteration of the mutual GR process, the probability that this node will have a neighbor in another layer that can get the message closer to the destination is small. That is, even though a node may have more options (neighbors in other layers) for forwarding a message, these options are basically useless.

Increasing the strength of correlations makes the different forwarding options that a node has more useful, as the probability to have a neighbor that can get the message closer to the destination in another layer increases. However, increasing the strength of correlations also increases the edge overlap between the layers (as shown in the previous section), which reduces the options that a node has for forwarding a message. We observe that very strong radial and angular correlations may not be optimal at low temperatures (cf. Figs. 6.2–6.4 for $T = 0.1$). This is because if the layers have the same nodes and the same parameters $\bar{k}, \gamma$, then as $\nu \to 1$, $g \to 1$, the coordinates of the nodes in the layers become identical. If at the same time the temperature of the layers is $T \to 0$, the connection probability $p(x_{ij})$ in each layer becomes the step function, where two nodes i, j are deterministically connected if their hyperbolic distance is $x_{ij} \leq R$. That is, as $T \to 0$, $\nu \to 1$, $g \to 1$, all layers

[4] Here we use the success rate as an indicator of the performance of mutual GR and angular mutual GR. In appendix E.2 we show that the stretch (i.e. the ratio between the paths found by greedy routing and the topologically shortest paths) is reduced if correlations are stronger. Hence, correlations do not only increase the probability to reach a target but also make the successful deliveries more efficient.

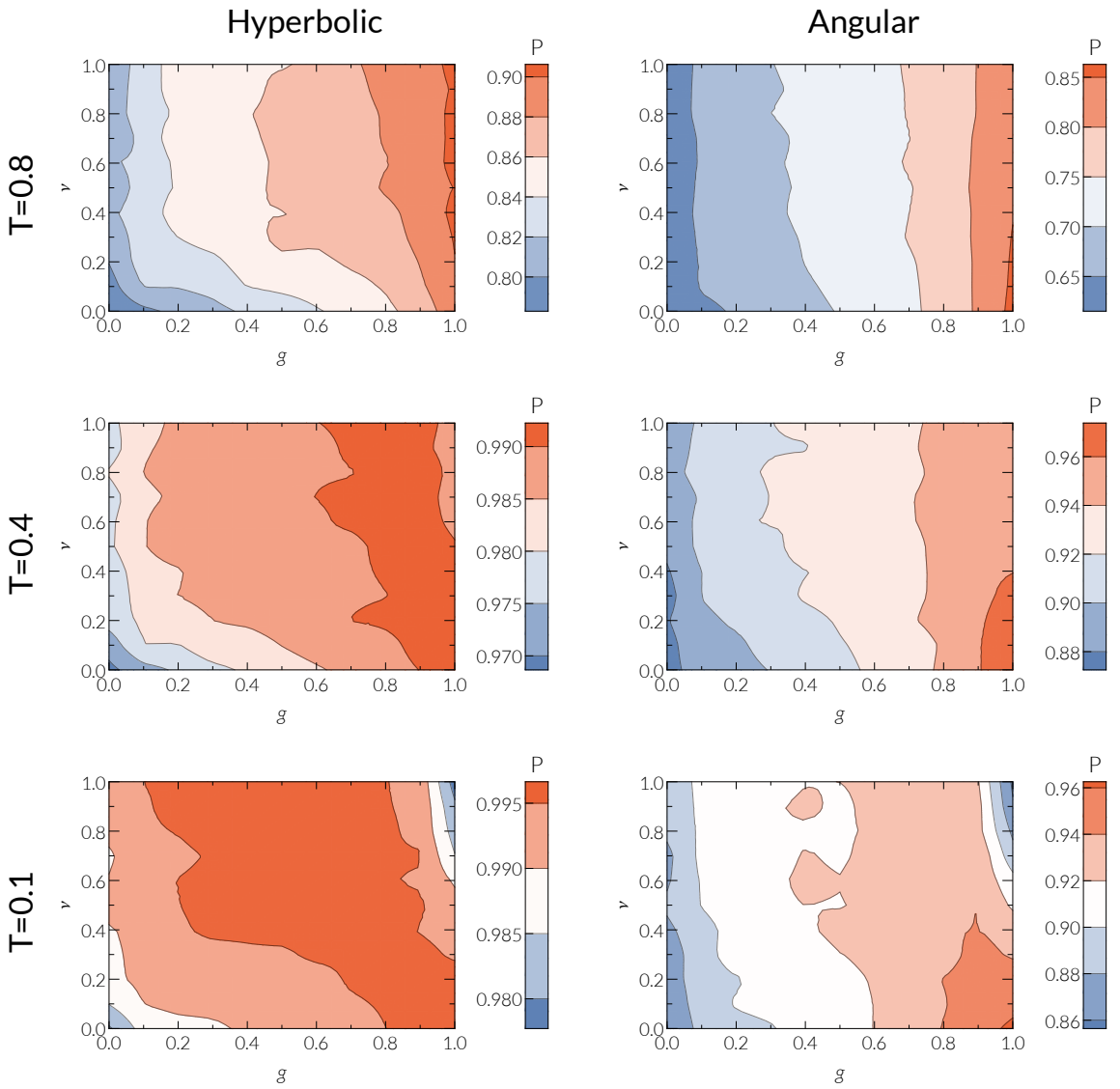

Figure 6.2: Success rate of mutual GR (hyperbolic routing, left column) and of angular mutual GR (angular routing, right column) for a two-layer multiplex system as a function of the radial (ν) and angular (g) correlation strengths. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From left to right, $T = 0.8, 0.4, 0.1$.

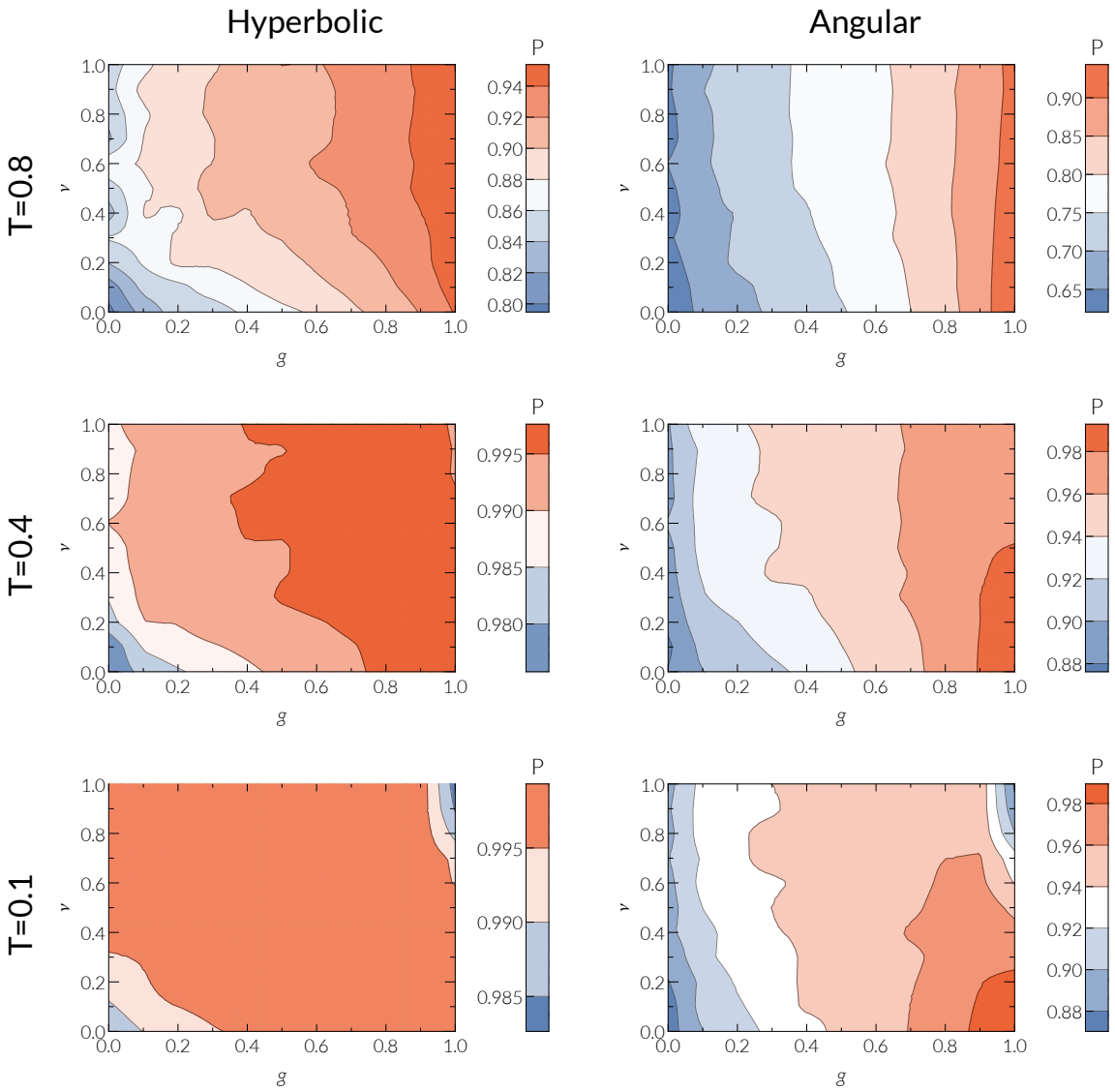

Figure 6.3: Success rate of mutual GR (hyperbolic routing, left column) and of angular mutual GR (angular routing, right column) for a three-layer multiplex system as a function of the radial (ν) and angular (g) correlation strengths. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From left to right, $T = 0.8, 0.4, 0.1$.

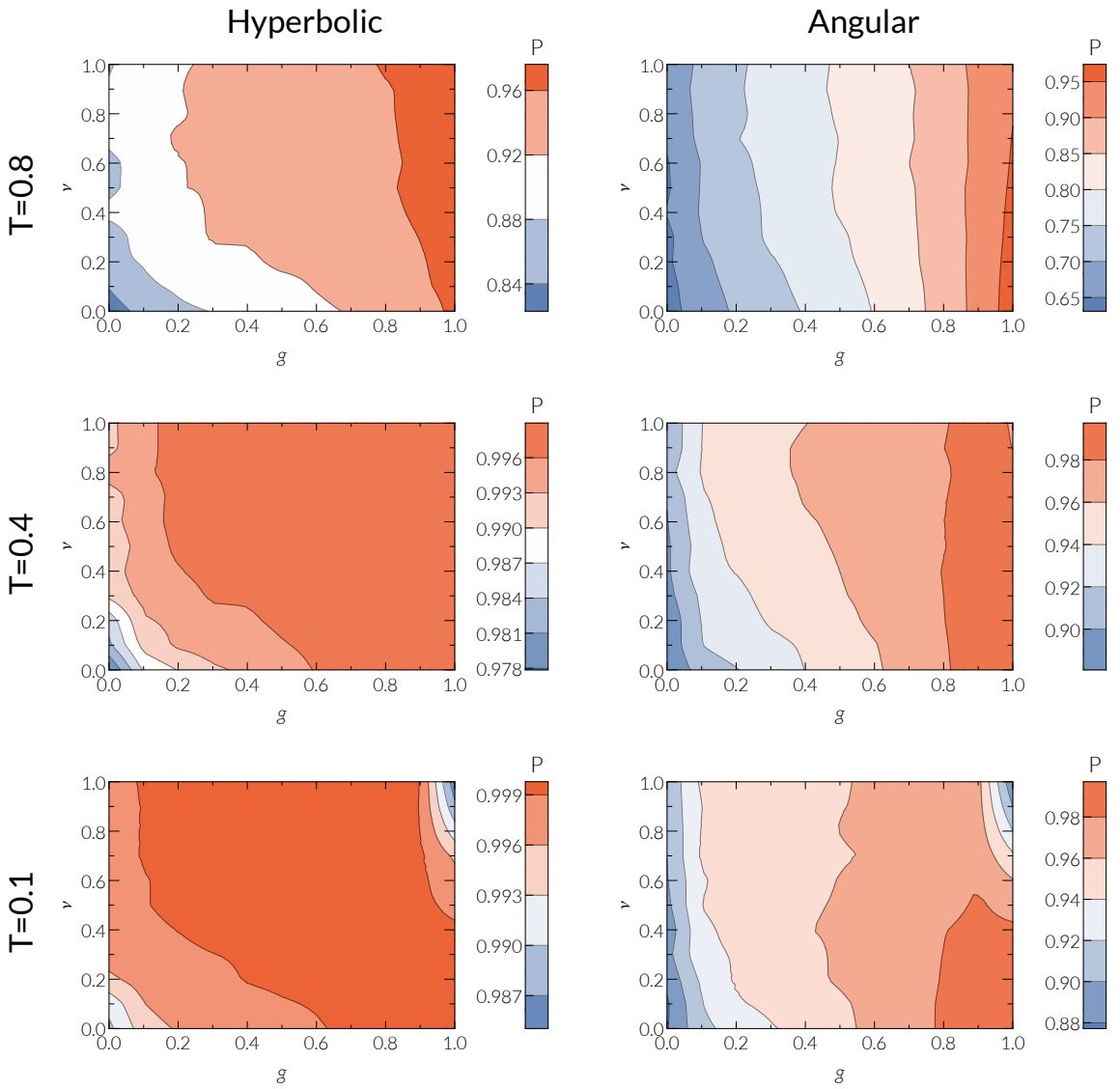

Figure 6.4: Success rate of mutual GR (hyperbolic routing, left column) and of angular mutual GR (angular routing, right column) for a four-layer multiplex system as a function of the radial (ν) and angular (g) correlation strengths. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From left to right, $T = 0.8, 0.4, 0.1$.

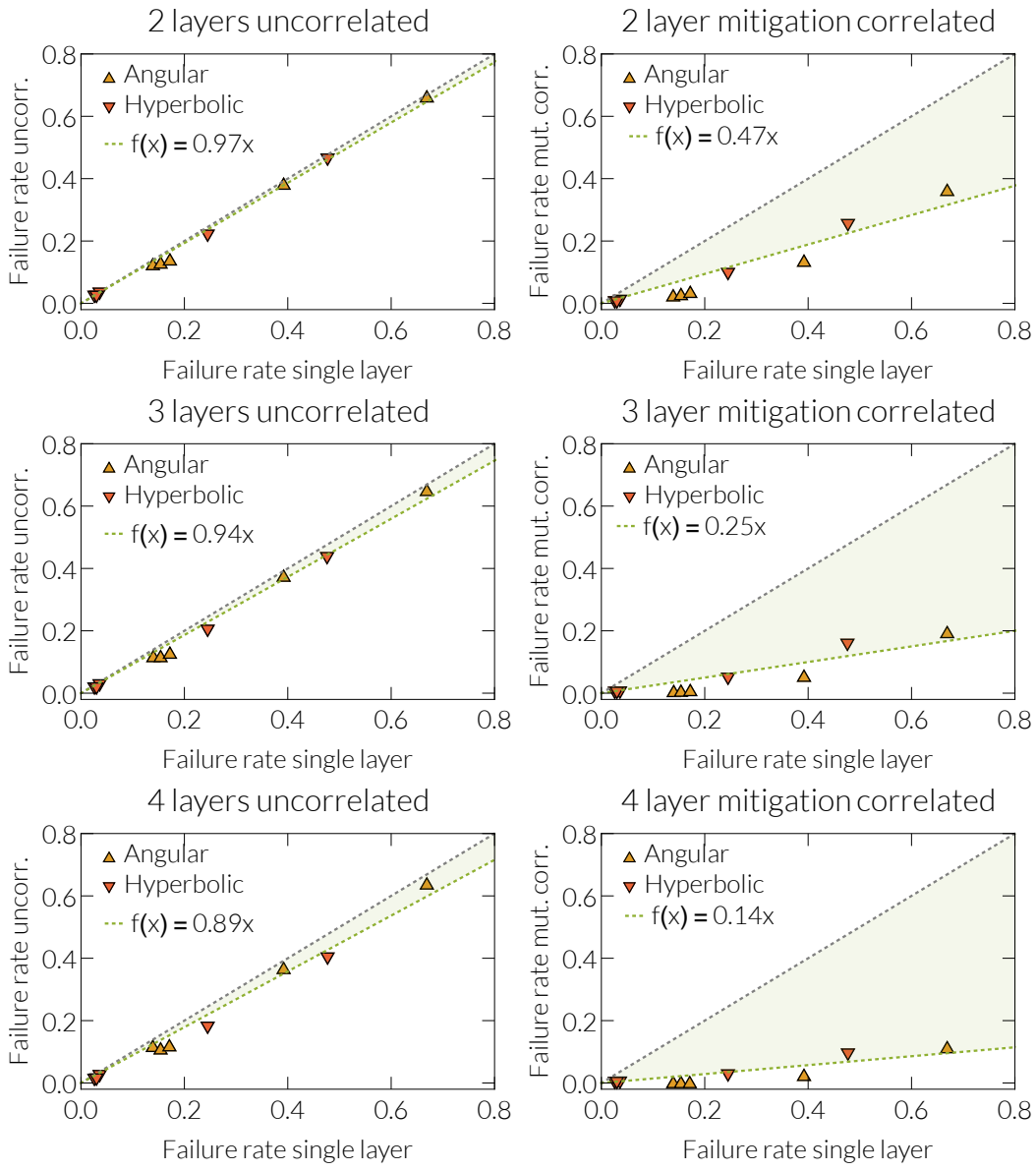

Figure 6.5: Failure rate (1−success rate) of mutual GR (red triangles pointing downwards) and angular mutual GR (yellow triangles pointing upwards). Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and the same temperature T that takes different values, $T = (0.1, 0.2, 0.4, 0.8, 0.9)$, corresponding for each navigation type respectively, from the leftmost triangle to the rightmost triangle. The left column corresponds to the case where there are no coordinate correlations among the layers, while the right column correspond to the case where there are optimal correlations, i.e., radial and angular correlation strengths that maximize the corresponding performance of mutual GR or angular mutual GR. From top to bottom we show the case of two, three, and four layers.

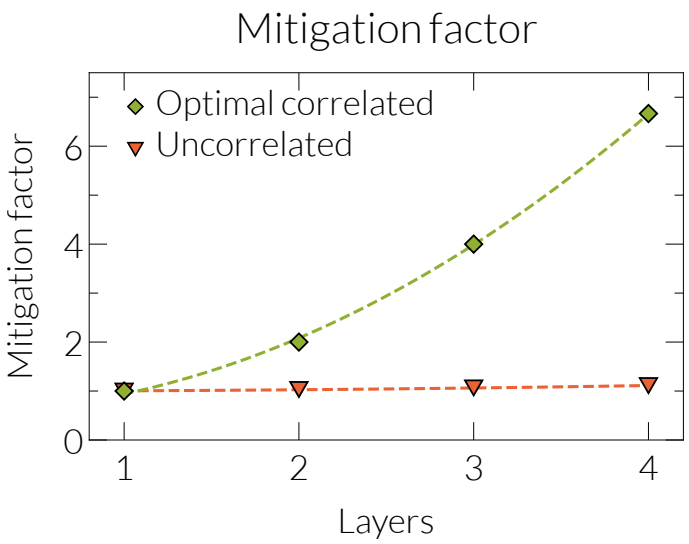

Figure 6.6: Mitigation factor as a function of the number of layers for optimal coordinate correlations and for uncorrelated coordinates.

become identical, and mutual GR degenerates to single-layer GR. We observe (Figs. 6.2–6.4) that the best mutual GR performance is always achieved at high angular correlations, and either high radial correlations if the temperature of the individual layers is high, or low radial correlations if the temperature of the layers is low. The best angular mutual GR performance is always achieved at high angular and low radial correlations.

From Fig. 6.5 we observe that for a fixed number of layers the failure rate (1−success rate) is reduced for optimal correlations[5] by a constant factor, which is independent of the navigation type (mutual GR or angular mutual GR) and the layer temperature. This factor, which we call *failure mitigation factor*, is the inverse of the slope of the best-fit lines in Fig. 6.5. Fig. 6.6 shows the failure mitigation factor for our two-, three-, and four-layer multiplexes for both uncorrelated and optimally correlated coordinates. Remarkably, if optimal correlations are present, the failure mitigation factor grows superlinerarly with the number of layers, suggesting that more layers with the right correlations can quickly make multiplex systems almost perfectly navigable. On the contrary, more layers without correlations do not have a significant effect on mutual navigation, which performs virtually identical to single-layer navigation.

6.5 Geometric correlations and routing in the Internet multiplex

Finally, we investigate how close to the optimal—in terms of mutual navigation performance—are the radial and angular correlations in the IPv4/IPv6 Internet. To this end, we use our

[5]With optimal correlations we refer to the values of ν, g where the respective routing success is maximized.

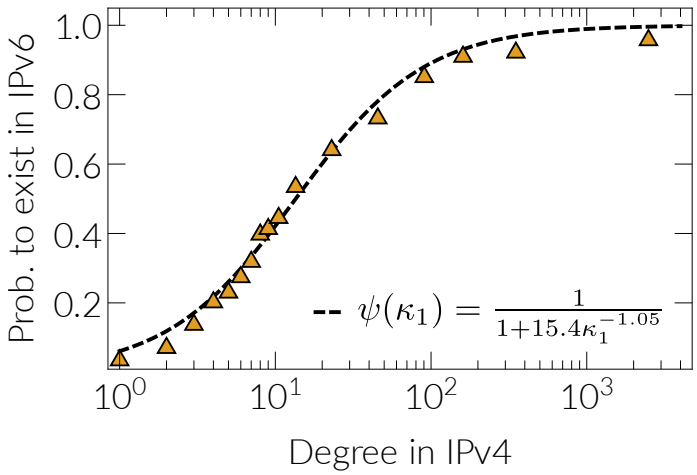

Figure 6.7: Probability that a node (AS) exists in the IPv6 Internet given its degree in the IPv4 Internet.

framework to construct a two-layer synthetic multiplex, where layer 1 has approximately the same number of nodes as in the IPv4 topology, $N_1 = 37563$ nodes, as well as the same power law degree distribution exponent $\gamma_1 = 2.1$, average node degree $\bar{k}_1 \approx 5$, and average clustering $\bar{c}_1 \approx 0.63$. Layer 2 has approximately the same number of nodes as in the IPv6 topology, $N_2 = 5163$ nodes, and the same power law exponent $\gamma_2 = 2.1$, average node degree $\bar{k}_2 \approx 5.2$, and average clustering $\bar{c}_2 \approx 0.55$.

The IPv4 topology is significantly larger than the IPv6 topology, and there are 4819 common nodes (Autonomous Systems) in the two topologies. We find that nodes with a higher degree in IPv4 are more likely to also exist in IPv6. Specifically, we find that the empirical probability $\psi(k)$ that a node of degree k in IPv4 also exists in IPv6 can be approximated by $\psi(k) = 1/(1 + 15.4k^{-1.05})$ (see Fig. 6.7). We capture this effect in our synthetic multiplex by first constructing layer 1, and then sampling with the empirical probability $\psi(k)$ nodes from layer 1 that will also be present in layer 2. A visualization illustrating the common nodes in the real Internet and in our synthetic multiplex is given in Fig. 6.8. We note that the fact that nodes with higher degrees in the larger layer have higher probability to also exist in the smaller layer has also been observed in several other real multiplexes [138]. However, our model for constructing synthetic multiplexes with different layer sizes is quite general, and allows for any sampling function $\psi(k)$ to be applied. We provide details of the extension of our model to multiplexes with different layer sizes in the following.

6.5.1 Model extension to multiplexes with different layer sizes

Here, we extend our framework to multiplexes with different layer sizes. Specifically, we consider a two-layer multiplex with layers 1, 2, which have number of nodes N_1, N_2. We assume that $N_1 > N_2$ and that there is a subset of N_{common} nodes in layer 1 that also exist

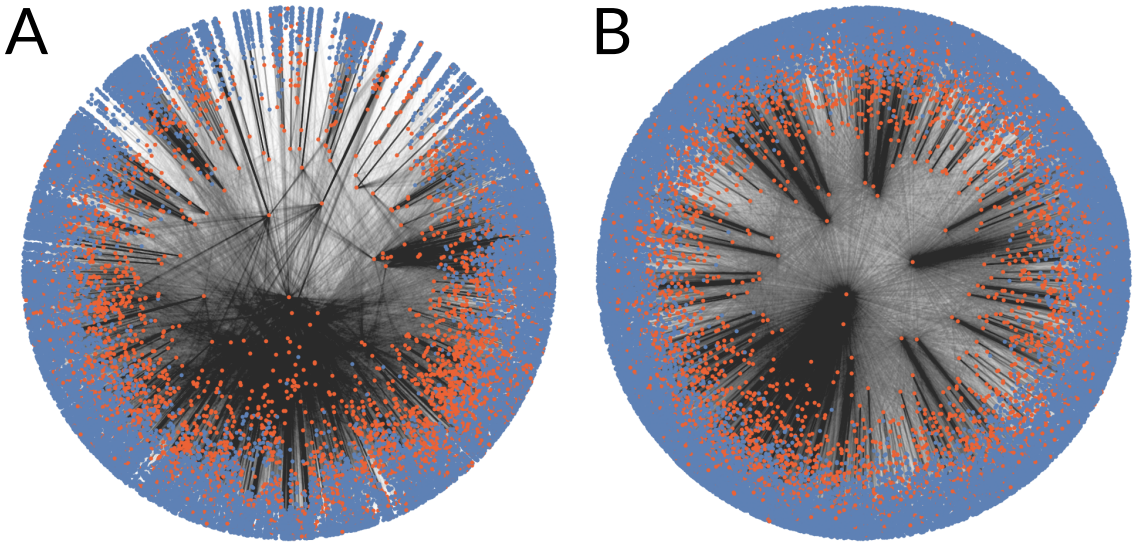

Figure 6.8: **A:** Hyperbolic mapping of the real IPv4 topology where nodes marked by red also exist in the IPv6 topology. **B:** Hyperbolic mapping of layer 1 of our Internet-like synthetic multiplex, where nodes marked by red also exist in layer 2.

in layer 2, $N_{\text{common}} \leq N_2$. To construct the two-layer multiplex we follow the steps below.

(i) Assignment of hidden variables in layer 1. For each node $i = 1, 2, \ldots, N_1$ in layer 1, we sample its hidden variable $\kappa_{1,i}$ as before, i.e., from the PDF $\rho_1(\kappa_1)$ in Eq. (6.6), and its angular coordinate $\theta_{1,i}$ from the uniform PDF $f(\theta)$ in Eq. (6.8).

(ii) Determining the common nodes. We now need to decide the N_{common} nodes from layer 1 that will also be present in layer 2. The simplest approach is to randomly select (approximately) N_{common} nodes from layer 1, by sampling each node from layer 1 with the same probability ψ,

$$\psi = \frac{N_{\text{common}}}{N_1}, \tag{6.29}$$

and declaring each sampled node as a common node that will also exist in layer 2. However, this random sampling approach may not be realistic. Indeed, as mentioned before, nodes with a higher degree in the IPv4 Internet (larger layer) have a higher probability to also exist in the IPv6 Internet (smaller layer). Fig. 6.7 shows the empirical probability for a node (AS) to exist in the IPv6 Internet given its degree in the IPv4 Internet. Recall that this probability can be approximated by

$$\psi(\kappa_1) = \frac{1}{1 + 15.4\kappa_1^{-1.05}}\,. \tag{6.30}$$

A dependence of the probability that a node exists in different layers on the degree of the node has also been observed in several other real multiplexes [138]. Therefore, a more

realistic and general approach is to sample each node from layer 1 with a probability $\psi(\kappa_1)$ that is a function of its expected degree $\kappa_{1,i}$, such that

$$\int_{\kappa_1^{\min}}^{\infty} \mathrm{d}\kappa_1 \psi(\kappa_1)\rho_1(\kappa_1) = \frac{N_{\text{common}}}{N_1}, \tag{6.31}$$

where $\kappa_1^{\min}$ is given in Eq. (6.7). Here, we use the $\psi(\kappa_1)$ from Eq. (6.30) to sample nodes from layer 1 that also exist in layer 2 of the synthetic multiplex that best mimics the real IPv4/IPv6 Internet (Fig. 6.8). The sampling yields $N_{\text{common}} \approx 4800$, which is approximately equal to the number of common ASs (4819) in the real IPv4/IPv6 Internet.

iii. Assignment of hidden variables in layer 2. For the nodes i in layer 2 that do not exist in layer 1 (non-common nodes), we sample their $\kappa_{2,i}$'s from the unconditional PDF $\rho_2(\kappa_2)$ in Eq. (6.9), and their $\theta_{2,i}$'s from the uniform PDF $f(\theta)$ in Eq. (6.8). For the common nodes, we assign hidden variables $\kappa_{2,i}, \theta_{2,i}$, as described below.

We first compute the PDF $\tilde{\rho}_1(\kappa_1)$ of the hidden variables κ_1 of the common nodes,

$$\tilde{\rho}_1(\kappa_1) = \frac{\psi(\kappa_1)\rho_1(\kappa_1)}{\int_{\kappa_1^{\min}}^{\infty} \mathrm{d}\kappa_1 \psi(\kappa_1)\rho_1(\kappa_1)}, \tag{6.32}$$

and the CDF $\tilde{F}_1(\kappa_1)$,

$$\tilde{F}_1(\kappa_1) = \int_{\kappa_1^{\min}}^{\kappa_1} \mathrm{d}\kappa' \tilde{\rho}_1(\kappa'). \tag{6.33}$$

Then, we compute the conditional CDF $F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\})$ in exactly the same manner as in section 6.2.2, with the only difference that in place of $F_1(\kappa_1)$ in Eq. (6.15), we use the $\tilde{F}_1(\kappa_1)$ that we compute in Eq. (6.33), and instead of $\rho_1(\kappa_1)$ we use $\tilde{\rho}_1(\kappa_1)$. That is,

$$F_\eta(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \frac{\partial C_\eta(\tilde{F}_1(\kappa_1), F_2(\kappa_2))}{\partial \kappa_1} \frac{1}{\tilde{\rho}_1(\kappa_1)}, \tag{6.34}$$

where $C_\eta(u, v)$ is the Gumbel-Hougaard copula (Eq. (6.17)) and $F_2(\kappa_2)$ is given in Eq. (6.16). The hidden variable $\kappa_{2,i}$ of each common node i is then sampled from the conditional CDF in Eq. (6.34). The angular coordinate of each common node $\theta_{2,i}$ is assigned using Eqs. (6.22), (6.23) with $N = N_2$.

iv. Creation of edges. The creation of edges in each layer is performed as before using the connection probabilities of the two layers in Eqs. (6.24), (6.25), with the only difference that now $d_1(\theta_{1,i}, \theta_{1,j}) = \frac{N_1}{2\pi}\Delta\theta_{ij}^1$ and $d_2(\theta_{2,i}, \theta_{2,j}) = \frac{N_2}{2\pi}\Delta\theta_{ij}^2$.

v. $\mathbb{S}^1$-to-$\mathbb{H}^2$ transformation. Finally, we again map the node hidden variables $\kappa_{1,i}, \kappa_{2,i}$ in layers 1, 2, to radial coordinates r_i^1, r_i^2 using Eqs. (6.26), (6.27), with the difference that now in these equations we have $R_1 = 2\ln\frac{N_1}{c_1}$ and $R_2 = 2\ln\frac{N_2}{c_2}$.

The above framework can be extended to more than two layers in the same manner as described in section 6.2.2. We note that when $\psi(\kappa_1)$ is the Internet's $\psi(\kappa_1)$ in Eq. (6.30),

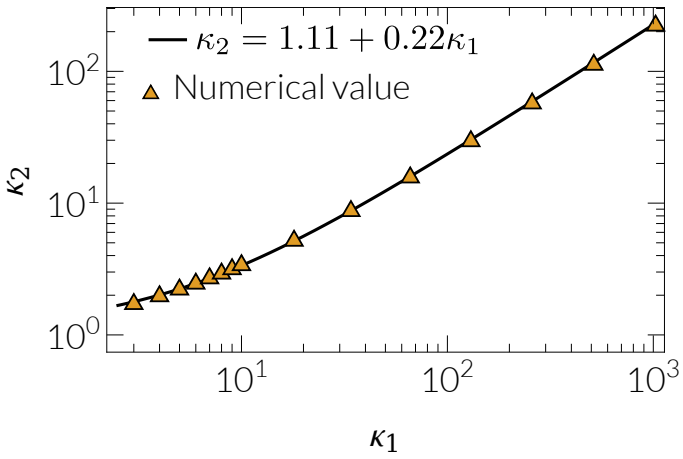

Figure 6.9: κ_2 as a function of κ_1 at the limit $\nu \to 1$. The triangles denote the values of κ_2 by numerically evaluating Eq. (6.34) using the $\psi(\kappa_1)$ in Eq. (6.30) and the IPv4/IPv6 Internet parameters $\gamma_1 = \gamma_2 = 2.1, \kappa_1^{\min} = 0.84, \kappa_2^{\min} = 1.30$. The best-fit line corresponds to $\kappa_2 = 1.11 + 0.22\kappa_1$.

the conditional CDF in Eq. (6.34) can be approximated by

$$F_\nu(\kappa_2|\tilde{\kappa}, \{\gamma_2, \kappa_2^{\min}\}) = e^{-(\tilde{\varphi}^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)})^{1-\nu}} \times \left[\tilde{\varphi}^{1/(1-\nu)} + \varphi_2^{1/(1-\nu)}\right]^{-\nu} \frac{\tilde{\varphi}^{\nu/(1-\nu)} \kappa_2^{\min} \tilde{\kappa}^{\gamma_2}}{\kappa_2^{\min} \tilde{\kappa}^{\gamma_2} - \kappa_2^{\min^{\gamma_2}} \tilde{\kappa}}, \quad (6.35)$$

where

$$\begin{aligned} \tilde{\varphi} &= -\ln\left[1 - \left(\kappa_2^{\min}/\tilde{\kappa}\right)^{\gamma_2 - 1}\right], \\ \varphi_2 &= -\ln\left[1 - \left(\kappa_2^{\min}/\kappa_2\right)^{\gamma_2 - 1}\right], \\ \tilde{\kappa} &= 1.11 + 0.22\kappa_1, \end{aligned} \quad (6.36)$$

and $\kappa_2^{\min}$ is given in Eq. (6.10).

Eq. (6.36) is obtained by considering the maximally correlated case $\nu \to 1$, where Eq. (6.34) converges to a Heaviside step function $F_\nu(\kappa_2|\kappa_1, \{\gamma_1, \gamma_2, \kappa_1^{\min}, \kappa_2^{\min}\}) = \Theta\left[\kappa_2 - \tilde{\kappa}\right]$. That is, for $\nu \to 1$, $\kappa_2 \approx 1.11 + 0.22\kappa_1 \equiv \tilde{\kappa}$. This relation is the analogue of Eq. (6.14), and is obtained by numerically evaluating Eq. (6.34) at $\nu \to 1$, see Fig. 6.9. Once this result is known, then for $\nu \neq 1$, we can approximate Eq. (6.34) with Eq. (6.35), which results from Eq. (6.11) if in place of $\kappa_1, \gamma_1, \kappa_1^{\min}$, we use $\tilde{\kappa}, \gamma_2, \kappa_2^{\min}$. The idea behind this approximation is that instead of directly correlating the $\kappa_{2,i}$ with the $\kappa_{1,i}$ via Eq. (6.34), we correlate them via Eq. (6.35) with the corresponding values of $\kappa_{2,i}$ at the maximal correlations ($\nu \to 1$), which in our case are given by $\kappa_{2,i} = 1.11 + 0.22\kappa_{1,i}$.

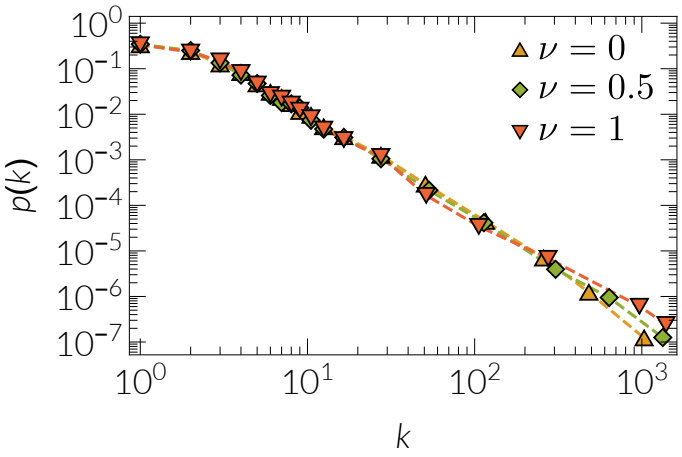

Figure 6.10: Degree distribution (PDF) in layer 2 of our synthetic Internet-like multiplex (Fig. 6.8) at correlation strengths $\nu = 0, 0.5, 1$.

In our synthetic Internet-like multiplex (Fig. 6.8), we sample the hidden variables $\kappa_{2,i}$ of the common nodes in layer 2 from the conditional CDF in Eq. (6.35). In Fig. 6.10, we show the marginal PDFs of the $\kappa_{2,i}$ of the common nodes at correlation strengths $\nu = 0$ (no correlations, where the $\kappa_{2,i}$ are sampled from their marginal CDF), $\nu = 0.5$ (partial correlations, where the $\kappa_{2,i}$ are sampled from the conditional CDF in Eq. (6.35)), and $\nu = 1$ (full correlations, where the $\kappa_{2,i}$ are directly obtained by Eq. (6.36)). In all cases, the marginal PDFs are nearly identical, validating the approximation described above.

6.5.2 Estimation of the radial and angular correlation strengths ν_E, g_E in the IPv4/IPv6 Internet

We now estimate which values of the model parameters ν and g represent the strength of radial and angular correlations in the real Internet. To estimate the empirical ν_E, we first compute the Pearson correlation coefficient between the inferred radial coordinates of common nodes (ASs) in the IPv4/IPv6 Internet. Then, we compute the same coefficient between the radial coordinates of common nodes in our synthetic Internet-like multiplex, at various radial correlation strengths $\nu \in [0, 1]$. The value of ν where the two coefficients are equal is the estimated value of ν_E. Fig. 6.11A shows the results, where we obtain $\nu_E \approx 0.4$.

To estimate g_E, we reconstruct IPv4-like and IPv6-like topologies as follows. We first consider all nodes in the real IPv4 topology with their inferred radial and angular coordinates, and connect each pair of nodes with the Fermi-Dirac connection probability $p(x_{ij})$ in Eq. (6.5), using the estimated temperature of the IPv4 topology, $T_1 = 0.5$, and such that the resulting network has the same average degree and power law degree distribution exponent as the real IPv4 topology, $\bar{k}_1 \approx 5$, $\gamma_1 = 2.1$. Subsequently, we consider all nodes in the real IPv6 topology. We assign to these nodes their inferred radial coordinates. For the nodes that also exist in IPv4, their angular coordinates are assigned using Eqs. (6.22),

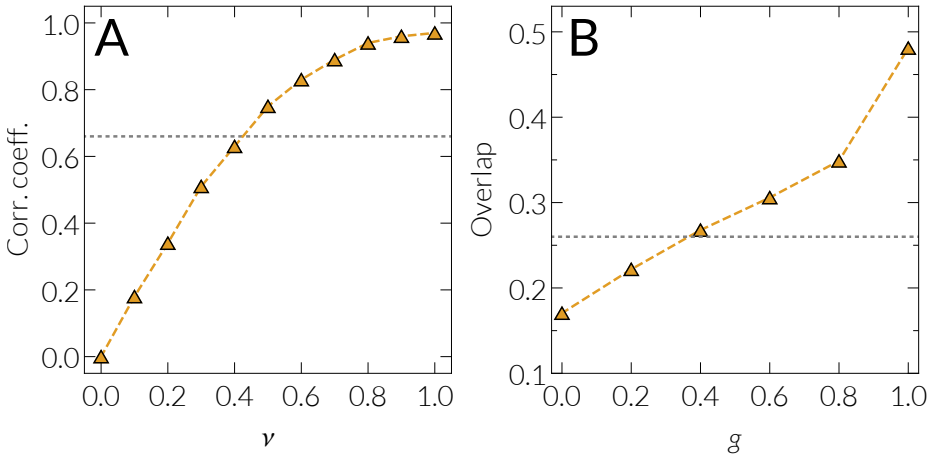

Figure 6.11: Estimation of ν_E, g_E in the IPv4/IPv6 Internet. **A.** Pearson correlation coefficient between the inferred radial coordinates of common ASs in the IPv4/IPv6 Internet (straight dashed line), and in our synthetic Internet-like multiplex at various radial correlation strengths ν. The two coefficients are equal at $\nu \approx 0.4 \equiv \nu_E$. **B.** Edge overlap O between reconstructed IPv4 and IPv6 topologies with inferred radial and angular coordinates (straight dashed line), and with synthetic angular coordinates for the common nodes in IPv6, at various correlation strengths g with their inferred IPv4 angles. The two overlaps are equal at $g \approx 0.4 \equiv g_E$.

(6.23), and are correlated to their real angular coordinates in IPv4, using different correlation strengths $g \in [0, 1]$. The angular coordinates of the non-common nodes are set equal to their inferred angular coordinates. Then, we connect each pair of nodes in IPv6 with the Fermi-Dirac connection probability $p(x_{ij})$ in Eq. (6.5), using the estimated temperature of the IPv6 topology, $T_2 = 0.5$, and such that the resulting network has the same average degree and power law degree distribution exponent as the real IPv6 topology, $\bar{k}_2 \approx 5.2$, $\gamma_2 = 2.1$. For the different values of g, we evaluate the edge overlap O (Eq. (6.28)) between the reconstructed IPv4 and IPv6 topologies. The value of g that matches the edge overlap obtained when we reconstruct the IPv6 topology with all nodes having their inferred angular coordinates is the estimated value of g_E. Fig. 6.11B shows the results, where we obtain $g_E \approx 0.4$.

6.5.3 Correlations present in the Internet multiplex improve navigability

For the nodes that exist in both layers of our multiplex, we tune the correlations among their coordinates as before, by varying the parameters ν and g. For each ν, g pair, we perform mutual navigation among 10^5 randomly selected source-destination pairs that exist in both layers. Fig. 6.12 shows respectively the performance of angular mutual GR and of mutual GR. We observe again that increasing the correlation strengths improves performance. In angular mutual GR, the success rate is 63% with uncorrelated coordinates, while with optimal correlations it becomes 75%. In mutual GR, the success rate with uncorrelated

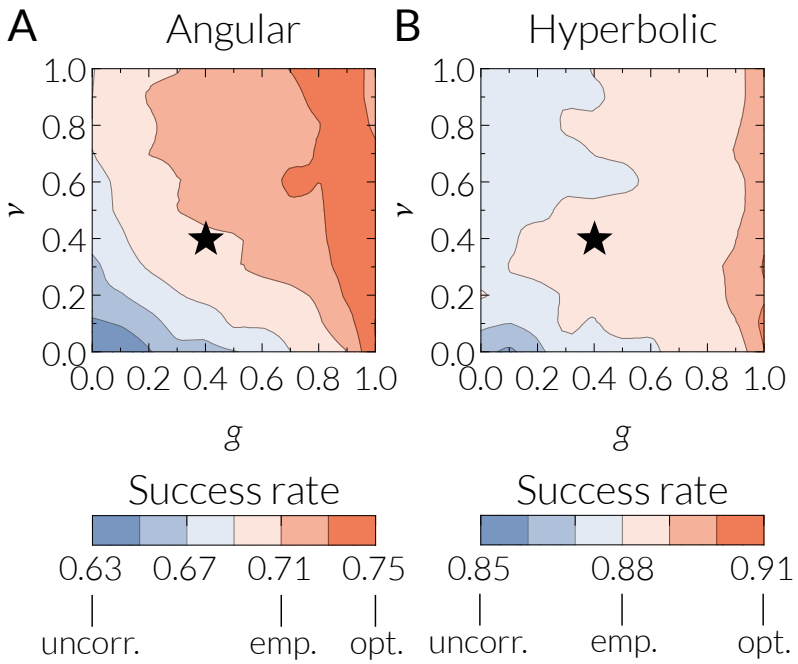

Figure 6.12: Performance of mutual navigation as a function of radial and angular correlation strengths (ν, g) in a two-layer synthetic multiplex that best mimics the real IPv4/IPv6 Internet. The black star indicates the achieved performance with the estimated correlation strengths in the real IPv4/IPv6 Internet, $\nu_E \approx 0.4, g_E \approx 0.4$. **A**: Performance of angular mutual GR. **B**: Performance of mutual GR.

coordinates is 85% and with optimal correlations is 91%. The star in Fig. 6.12 indicates the achieved performance with the empirical correlation strengths in the IPv4/IPv6 Internet, $\nu_E \approx 0.4, g_E \approx 0.4$, which are the empirical values estimated earlier. At $\nu = \nu_E, g = g_E$, the success rate of angular mutual GR is 71%, which is closer to the rate obtained with optimal correlations than to the uncorrelated case. For mutual GR, the success rate is 88%, which lies in the middle between the uncorrelated and optimally correlated case.

6.6 Summary: Hidden geometric correlations facilitate navigation

In the previous chapter we found that node coordinates in different layers of real multiplexes embedded into hyperbolic spaces are correlated, hence the underlying hyperbolic distances in the constituent layers are also correlated.

In this chapter, we have focused on mutual navigation, which uses the coordinates and connectivity of nodes in different layers to reach intended communication targets. We have shown that trans-layer geometric correlations improve the performance of mutual navigation, which outperforms navigation in the single layers only if these correlations are sufficiently strong. Our results also reveal that having more layers with the right correlations can

quickly make multiplex systems almost perfectly navigable. On the contrary, more layers without correlations do not improve mutual navigation.

Our findings can have important applications for improving information transport and navigation or search in multilayer communication systems and decentralized data architectures [22, 30].

Discussion

Summary and outlook

"We all decided to walk through the same door
on the Internet so we could think together."
Alec Couros

This chapter contains parts adapted from my publications [40, 89, 102, 116].

The emergence of Web 2.0 reshaped the digital landscape. Instead of there being only a few producers of content and many consumers, now nearly every Internet user also produces content. This is especially the case with online social networks, which have grown rapidly and nowadays include more than 2 billion users worldwide. These systems are key players in the Web 2.0 cosmos, where they evolve and compete in the absence of central control. Hence, tools, ideas, and techniques from the analysis of complex systems are suitable to describe the evolution of and competitive interactions between online social networks.

The digital world consists of highly connected and strongly coupled interacting subsystems. Those basic building blocks are single networks, each of which obeys specific dynamics in the absence of coupling to the whole system. So the complexity of the digital world is a consequence of both the dynamics of networks in isolated environments and the interactions between many such networks. Finally, not all of the building blocks are identical. Instead, different networks address different peer groups or have different functionalities. Hence, to reveal the fundamental mechanisms that determine the fate of the digital world, it is necessary to understand the interaction of heterogeneous networks, each driven by its own intrinsic dynamics. To meet this challenge, we started analyzing the evolution of online social networks in isolation. Building on that knowledge, we then studied competition between a priori identical networks, and the dichotomy consisting of coexistence and domination, in analogy to ecology theory. Finally, we describe competition between local networks and an international network, for which we had to take into account the network of interactions between different countries.

We found that the evolution of isolated online social networks follows an intricate path. The system is not connected globally from the beginning, but instead undergoes a dynamical percolation transition. This remarkable behavior rules out the adequateness of models based on preferential attachment, as such networks are either connected or not – they do not exhibit dynamical transitions. We have shown that the evolution of online social networks is governed by social networks that existed long before the invention of the Internet. These offline social networks underlie the formation of online social networks in a two-layer model. The dynamics of this system is governed by two main mechanisms: a dynamics of viral spreading and the influence of mass media. The former is similar to the SIS model

known from epidemic spreading; the latter is a homogeneous influence that triggers random subscriptions among the population susceptible to participate in online social networks. The model exhibits a dynamical percolation transition as observed in the "Pokec"[6] network. We chose the final snapshot of the empirical network as a proxy for the pre-existing underlying social structure in our model and were able to reproduce the entire topological evolution of the empirical network with astonishing precision. The viral spreading and media influence mechanisms play complementary roles in the connectivity structure of the network. The viral dynamics tend to connect components, whereas the media effect tends to create new, disconnected components. Hence, the balance between the relative strengths of these two mechanisms governs the position of the dynamical percolation transition in our model. We found that, to match the empirical evolution, the influence of one active friend in terms of the viral spreading dynamics is four to five times stronger than the influence of mass media.

Although the model precisely reproduces many topological features of the empirical network, such as the number of components, mean degree, the dynamical percolation transition and its critical point, and the evolution of the average shortest path length as well as network diameter, there is a strong disagreement when it comes to local topological quantities such as mean local clustering. This shortcoming indicates that at local scales our model lacks an important feature. We have shown that this feature is the strength of social ties, for which we follow the definition of Granovetter, the author of the "strength of weak ties" hypothesis. By comparing our model with the empirical evolution of the "Pokec" network, we find that indeed weaker social ties exhibit higher transmissibility for the viral spreading process. So, the extended version of our model takes into account the strength of social ties and reproduces well the global and local topological properties of the empirical network.

In reality, most online social networks do not evolve in isolation. Instead, they compete with other digital services for the attention of users, which is bounded due to the cognitive and physical limitations of individuals. We consider this attention to be a scarce resource and networks to be competing species in a digital ecosystem. Following this line of reasoning, we considered whether multiple online social networks can coexist or if the competitive interaction always leads to the domination of a single network. We considered a priori identical networks and assumed that users are more likely to engage in more active networks. This rich-get-richer effect creates a positive feedback loop, similar to preferential attachment. Interestingly, we found that despite the rich-get-richer effect, multiple networks can coexist depending on the parameters that describe the total virality (which represents the total amount of attention or time of individuals) and the activity affinity (which describes how much more likely individuals are to engage in more active networks). We found that at a critical point of the activity affinity, the system undergoes a subcritical pitchfork bifurcation such that coexistence is only possible in the subcritical regime. This type of bifurcation induces a hysteresis effect which means that once coexistence or digital diversity is lost, it cannot be recovered. Multistability in the subcritical regime in combination with noise present in the full stochastic model have interesting implications. Starting from empty networks, a stable coexistence solution is not always reached. In particular, we found that

[6]Pokec is a very large and popular OSN in Slovakia. See section 2.2 on page 18 for details.

for large values of the virality, the effective critical line that determines the region in which coexistence is reached with a probability of at least 50% saturates at constant values of the activity affinity, compared to the linear relationship of the stability region. This is of great importance when considering many networks, as the effective critical lines corresponding to the coexistence of a larger number of networks saturate to lower values of the activity affinity respectively. This means that without precise knowledge of the empirical parameters, our model explains the observation that there is a moderate number of coexisting digital services in agreement with empirical observations. Finally, the probability of reaching the coexistence solution depends on the relative strength of the influence of mass media. We found that a larger influence of mass media increases the chance of coexistence and hence augments the observed digital diversity.

We have seen how networks can coexist despite a rich-get-richer mechanism that provides more active networks with the benefit of more attention. However, we have considered all networks as a priori identical, which is not the case in reality. Especially the rise to supremacy of Facebook and the extinction of many local networks can only be explained by taking into account the intrinsic heterogeneity of networks. In the competition between local networks and an international network, the latter provides users with the possibility of connecting to individuals from other countries. This fact effectively gives the international network an advantage over local ones that is similar to an increased intrinsic fitness. Hence, one has to take into account the interaction network between different countries, which leads to a network of multilayer networks. In addition, compared to the case of identical networks, the symmetry of the bifurcation is broken, which gives rise to a saddle-node bifurcation instead of a subcritical pitchfork bifurcation. We showed that depending on the abundance of inter-country social ties, which we refer to as global connectivity, and the activity affinity, there are regions of qualitatively distinct behavior. Local networks can coexist alongside the international network for low activity affinity and global connectivity. For intermediate activity affinity and especially for high global connectivity, the international network always dominates. For high values of the activity affinity and low values of the global connectivity, local networks dominate if they are launched before the international network, which we assume to be the case in each country except the US. Interestingly, between the regions of the domination of local networks and the domination of the international network a region exists in which the final state of the system varies randomly between different realizations. In this region, which we call the "coinflip region", only the domination of either local networks or the international network is possible. Which of these solutions is approached is up to chance. We compared the evolution of the number of countries where local networks are the most popular services predicted by our model with empirical data and found that the most probable parameters lie in the coinflip region with a slight advantage towards the international network. In particular, for these parameters, local networks dominate in 30% of the cases whereas in the remaining 70% the international network ends up dominating. Hence, the global overtake of Facebook could have not taken place with a significant probability. Results of this type should warn us against the comparison of single realizations of stochastic processes, in particular those with strong non-linearities, such as the one we study here.

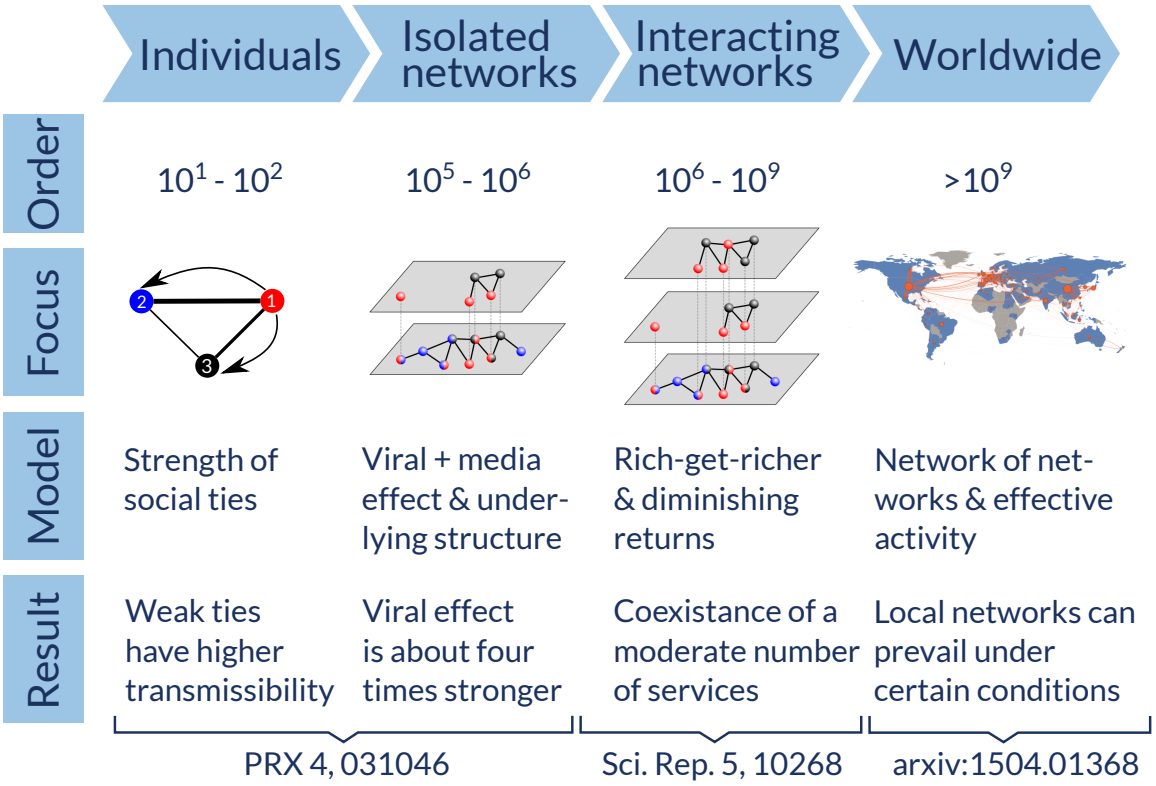

Figure 6.13: Overview of the evolution and ecology of the digital world organized by size.

We have described the evolution and competitive interaction of online social networks ranging from individual ties to globally interacting networks (see Fig. 6.13). Our model can be extended to include several internationally operating networks that are constantly evolving in analogy to natural systems. This can result in random fluctuations of the intrinsic fitness of networks and lead to a theory of Darwinian selection in the digital ecosystem. A deep understanding of the digital ecosystem is crucial to sustain diversity in the digital world, which is a requisite for the freedom of information and a "digital democracy".

Recently, decentralized architectures have attracted a lot of attention; for example, the virtual currency Bitcoin [29]. Such systems provide better scalability, more transparency, and by design cannot be controlled by central entities. Hence, decentralized systems will play an important role in future digital architectures [12,30]. Such systems have to function without central control and hence require solutions to particular challenges, such as search and navigation relying only on local information (greedy routing). We studied this case explicitly in terms of geometric spaces and geometric correlations in multiplex networks.

It has been shown that many complex networks have an effective hyperbolic geometry underlying their observed topologies, where node coordinates are abstractions of the popularity and similarity of nodes. Furthermore, real networks are often not isolated entities but

instead form interacting parts of larger and more complex systems, called multiplex or multilayer networks. Examples of such systems can be found in drastically different domains. The most classic example is provided by the different social networks that a person may belong to. Other examples are the structural and functional connections in the brain, or the IPv4 and IPv6 Internet topologies.

We found that different real multiplex systems, including brain networks, protein interaction networks, the Internet, and scientific collaboration networks, are not random combinations of single networks. Instead, we found that their constituent network layers exhibit strong geometric correlations. Specifically, each constituent network layer can be mapped into its own hyperbolic space; node coordinates in this space are abstractions of the popularity and similarity of nodes. The probability of the existence of an edge between two nodes in then a monotonically decreasing function (the Fermi-Dirac distribution) of their hyperbolic distance. Real networks can be embedded within hyperbolic space by techniques based on Maximum Likelihood Estimation to infer the popularity and similarity node coordinates. From the hyperbolic maps constructed for a single network, one can: identify soft communities of nodes, which are groups of nodes located close to each other in the angular similarity space [84, 126, 130]; predict missing links with high precision [127, 128, 130]; and facilitate efficient greedy routing. We found that node coordinates are strongly correlated across layers of real multiplexes, meaning that distances between nodes in the underlying hyperbolic spaces of the constituent network layers are also strongly correlated.

The geometric correlations we discovered yield a very powerful and general framework for understanding and analyzing real multiplex systems. One can use this framework to define and detect multidimensional communities, which are sets of nodes that are simultaneously similar in multiple layers. Nodes are considered similar in a certain layer if they are located at a small angular distance within that layer. Clusters of nodes that are simultaneously similar in different layers are overabundant compared to a random superposition of the constituent layer topologies. These clusters form multidimensional communities. In addition, the geometric framework allows us to quantify the degree of similarity between different communities, which is related to the multidimensional angular distance between them. The existence of geometric correlations furthermore enables accurate trans-layer link prediction, where connections in one layer can be predicted by observing the geometric space of another layer. In particular, nodes that are located at a small hyperbolic distance in one layer have a high probability of being close in a different layer and hence the probability that they are connected in that layer is high as well. As a consequence, knowing the hyperbolic distance in one layer allows us to predict the connection probability in a different layer.

We developed a model to create multiplex networks with geometric correlations. In our model, angular and radial correlations can be tuned individually and independently from the topological properties of the constituent layers. We created correlations of the hidden variables of nodes using copula techniques. We specifically accounted for different layer sizes, which occurs in the case of the IPv4/IPv6 Internet multiplex. In particular, high-degree nodes in the larger layer have a high probability of also existing in the smaller layer. The model allows us to explicitly study the effect of correlations on the performance of mutual greedy routing.

In greedy routing, nodes forward a message to their neighbor that is closest to the final destination in geometric space. The message either reaches its target, or it enters a loop, i.e. the message is given back to a node it already visited, and the delivery fails. To study navigation in multiplex systems, we extend the notion of greedy routing so that a node forwards a message to its neighbor that is closest to the destination in any of the layers comprising the system. We call this process mutual greedy routing. Interestingly, mutual greedy routing outperforms navigation in the constituent networks alone only if the geometric correlations are sufficiently strong. We show this explicitly for the example of the Internet, for which we create synthetic multiplexes whose individual layer topologies mimic those from the empirical data and where we can tune the geometric correlations. By comparing our model with the coordinates inferred from the empirical data, we are able to quantify the correlations present in the real Internet. These correlations indeed improve the mutual navigability of the system significantly for navigation using both hyperbolic and angular distances. We also studied the case of more than two network layers. Having more layers with the right correlations can quickly make multiplex systems almost perfectly navigable. In contrast, more layers without correlations do not improve mutual navigation.

Our findings can lead to a variety of applications in many disciplines, ranging from improving information transport and navigation in multilayer communication systems and decentralized data architectures, to understanding functional and structural networks in the brain and deciphering their precise relationship(s). They may also allow us to predict links among nodes, such as terrorists, in a specific network, by only knowing their connectivity in some other network.

In summary, in this thesis we dealt with challenges facing an interconnected world strongly influenced by the recent digital revolution. The rise of the Internet has connected individuals on unprecedented scales and Web 2.0 promotes worldwide collaboration and the nearly instantaneous exchange of ideas. However, the dominance of a few powerful information monopolies poses a threat to the freedom of both ideas and the decisions of individuals. Two factors are therefore essential for a prosperous future in the digital age: digital diversity and decentralization. Concerning the former, we introduced a set of models based on empirical observations, which improve understanding of the dynamics of and competitive interactions between online social networks: the key players in the Web 2.0 cosmos. In particular, our findings shed light on the conditions under which digital diversity can be sustained. Concerning the latter, the design of decentralized architectures poses certain challenges, among which we specifically addressed search and navigation with only local knowledge. We revealed under which conditions the existence of many interacting networks facilitates these tasks. Interestingly, many real systems fulfill these conditions. To conclude, from a system-level perspective, a prosperous future in the digital age comprised of a diverse digital landscape with interacting, decentralized architectures is possible; but so is the opposite. It remains a task for society to create sufficient awareness and the correct incentives to create the future we desire. I hope that our findings will help to accomplish this ambitious and important endeavor.

Resumen en castellano

La invención de la Web 2.0 ha cambiado totalmente el mundo digital. Antes, unos pocos eran creadores de contenido y los demás simplemente lo consumían. En la Web 2.0 casi todos los usuarios de Internet producen contenido. Esto es especialmente cierto para las redes sociales online que han crecido rápidamente y actualmente cubren más que dos mil millones de usuarios a nivel mundial. Estos sistemas son los actores clave en el cosmos de la Web 2.0 en el que se desarrollan y compiten en ausencia de un control centralizado. Por tanto, herramientas, ideas y técnicas de análisis de sistemas complejos son útiles y especialmente adecuadas para describir la evolución y competición entre redes sociales online.

El mundo digital se compone de subsistemas que interaccionan entre ellos y que están fuertemente acoplados. Estos componentes básicos son redes individuales, y cada una obedece una dinámica intrínseca en ausencia de acoplamiento con el resto del sistema. Por lo tanto, la complejidad del mundo digital es una consecuencia tanto de la dinámica de las redes en aislamiento como de la interacción entre muchas redes. Además, no todos los componentes básicos son idénticos. Diferentes redes atraen diferentes grupos de personas o tienen diferentes funcionalidades. Por lo tanto, para identificar los mecanismos fundamentales que determinan el destino del mundo digital, es necesario entender la interacción entre redes heterogéneas, de las que cada una sigue una dinámica intrínseca. Para afrontar este reto, comenzamos con el análisis de la evolución de las redes sociales online en aislamiento. Basándonos en este conocimiento, investigamos la competición entre redes a priori idénticas y estudiamos la dicotomía entre coexistencia y dominación en analogía con la ecología. Por último, describimos la competición entre redes locales y una red internacional, para lo cual tuvimos que tener en cuenta la red de interacciones entre diferentes países.

Hemos encontrado que la evolución de las redes sociales online aisladas sigue un camino particular. El sistema no está conectado a nivel global desde el principio, sino que exhibe una transición de percolación dinámica. Este comportamiento descarta modelos como *preferential attachment* ya que este tipo de redes no exhiben una transición dinámica – están conectadas o no lo están. Hemos demostrado que la evolución de redes sociales online se rige por las redes sociales que existían mucho antes de la invención del Internet. Estas redes sociales offline subyacen bajo la formación de redes sociales online en un modelo de dos capas. La dinámica de este sistema tiene dos mecanismos principales: una dinámica de propagación viral y la influencia de los medios centrales de comunicación. La primera es similar al modelo de SIS conocido a partir de propagación de epidemias y la última representa la influencia homogénea que causa suscripciones al azar entre la población susceptible de participar en redes sociales online. El modelo muestra una transición de percolación dinámica igual que en la evolución empírica de la red "Pokec". Elegimos la configuración final de la red empírica como un representante de la estructura social subyacente preexistente en nuestro

modelo que nos permitió reproducir con mucha precisión toda la evolución topológica de la red empírica. El mecanismo viral y la influencia de los medios de comunicación tienen un efecto complementario para la conectividad de la red. La dinámica viral tiende a conectar componentes mientras que el efecto mediático tiende a crear nuevos componentes desconectados. Por lo tanto, el equilibrio entre las fuerzas relativas de estos dos mecanismos regula la posición de la transición de percolación dinámica en nuestro modelo. Encontramos que para que coincida con la evolución empírica, la influencia de un amigo activo, en términos de la dinámica de propagación viral, es entre cuatro y cinco veces más fuerte que la influencia de los medios de comunicación.

A pesar de que el modelo reproduce muchas características topológicas de la red empírica, incluyendo el número de componentes, el grado medio, la transición de percolación dinámica y su punto crítico, la evolución de la longitud media del camino más corto y el diámetro de la red, existe un desacuerdo respecto a las medidas topológicas locales como el *mean local clustering*. Esta limitación indica que a escala local nuestro modelo carece de una característica importante. Hemos demostrado que esta característica es la fuerza de los enlaces sociales que cuantificamos seguiendo la definición de Granovetter, el autor de la hipótesis de la "fuerza de los enlaces debiles". Al comparar el resultado de nuestro modelo con la evolución empírica de la red "Pokec", encontramos que los enlaces sociales más débiles muestran una transmisibilidad mayor para el proceso de propagación viral. La versión extendida de nuestro modelo tiene en cuenta la fuerza de los enlaces sociales y reproduce bien las propiedades topológicas globales y locales de la red "Pokec".

En realidad, la mayoría de las redes sociales online no evoluciona de forma aislada. Al contrario, las redes compiten con otros servicios digitales por la atención de los usuarios, que está limitada debido a razones cognitivas y físicas de las personas. Esta atención la consideramos como un recurso escaso por el cual las redes, que representan diferentes especies, compiten en un ecosistema digital. Con este razonamiento, respondemos a la pregunta de si múltiples redes sociales online pueden coexistir o si la interacción competitiva siempre lleva a la dominación de una sola red. Hemos considerado redes a priori idénticas y asumimos que los usuarios son más propensos a involucrarse en las redes más activas. Este efecto *rich-get-richer* crea una retroalimentación positiva, similar a *preferential attachment*. Sorprendentemente, encontramos que a pesar del efecto *rich-get-richer* múltiples redes pueden coexistir en función de los parámetros que describen la viralidad total (representa la atención total o el tiempo de las personas) y la afinidad de actividad (describe cuanto más las personas prefieren interactuar con redes más activas). Hemos encontrado que en un punto crítico de la afinidad de actividad, el sistema exhibe una bifurcación *pitchfork* subcrítica de tal manera que la coexistencia sólo es posible en el régimen subcrítico. El tipo de bifurcación induce un efecto de histéresis, que significa que una vez que se pierde la coexistencia o diversidad digital éste no se puede recuperar. La multiestabilidad en el régimen subcrítico en combinación con el ruido presente en el modelo estocástico tiene implicaciones interesantes. A partir de las redes inicialmente vacías, el sistema no siempre llega a una solución de coexistencia aunque aquella sea estable. En particular, encontramos que para grandes valores de la viralidad la línea crítica efectiva que determina la región en la que el sistema alcanza la coexistencia con una probabilidad de al menos 50 % satura a valores constantes

de la afinidad de actividad; al contrario de la relación lineal de la región de estabilidad de la solución de coexistencia. Esto es muy importante cuando se consideran varias redes ya que las líneas críticas efectivas de soluciones de coexistencia de números más altos de redes saturan a valores más bajos de la afinidad de actividad respectivamente. Esto significa que, sin conocimiento preciso de los parámetros empíricos, nuestro modelo explica la coexistencia de un número moderado de servicios digitales, de acuerdo con observaciones empíricas. Finalmente, la probabilidad de llegar a la solución de la coexistencia depende de la fuerza relativa de la influencia de los medios de comunicación. Hemos encontrado que un efecto mediático más fuerte aumenta la probabilidad de coexistencia y entonces aumenta la diversidad digital observada en el sistema.

Hemos mostrado que varias redes pueden coexistir a pesar de un efecto *rich-get-richer* que beneficia a las redes más activas. Sin embargo, hasta ahora, hemos considerado que todas las redes son a priori idénticas, algo que claramente no se corresponde con la realidad. Especialmente, la dominación de Facebook y la extinción de muchas redes locales sólo se puede explicar teniendo en cuenta la heterogeneidad intrínseca de las redes. En la competición entre redes locales y una red internacional, esta última les da a sus usuarios la posibilidad de conectar con personas de otros países. Efectivamente este hecho da una ventaja a la red internacional sobre las locales similar a un aumento de la *fitness* intrínseca. Por lo tanto, uno tiene que tener en cuenta la red de interacciones entre diferentes países que conduce a una red de redes multi-capa. Además, en comparación con el caso de redes idénticas, la heterogeneidad de *fitnesses* intrínsecas rompe la simetría de la bifurcación y da lugar a una bifurcación *saddlenode* en lugar de la bifurcación *pitchfork* subcrítica. Hemos demostrado que, dependiendo de la abundancia de los enlaces sociales entre países, que llamamos conectividad global, y la afinidad de actividad hay regiones de comportamientos cualitativamente distintos. Redes locales pueden coexistir con la red internacional si la afinidad de actividad y la conectividad global son bajas. Para una afinidad de actividad intermedia y especialmente para valores altos de la conectividad global, la red internacional siempre domina. Para valores altos de la afinidad de actividad y bajos de la conectividad global las redes locales dominan si nacen antes de la red internacional, que suponemos que es el caso en cada país, excepto los Estados Unidos. Curiosamente, entre las regiones de la dominación de las redes locales y la dominación de la red internacional existe una región en la que el estado final del sistema cambia al azar entre las diferentes realizaciones. En esta región, que llamamos la región *coinflip*, sólo la dominación de las redes locales o de la red internacional es posible. Cual de estas soluciones es alcanzada por el sistema depende del azar. Comparamos la evolución del número de países en los que las redes locales son los servicios más populares según la predicción del modelo con los datos empíricos y encontramos que los parámetros más probables se encuentran en la región *coinflip* con una ligera ventaja hacia la red internacional. En particular, para estos parámetros las redes locales dominan en un 30 % de los casos, mientras que en los restantes 70 % la red internacional acaba dominando. Por lo tanto, la dominación mundial de Facebook podría no haber tenido lugar con una probabilidad significativa. Este tipo de resultados muestra que hay que tener cuidado al comparar realizaciones individuales de procesos estocásticos, en particular los altamente no lineales, como los que investigamos aquí.

Hemos desarrollado una descripción de la evolución y la interacción competitiva de las redes sociales online desde la escala de enlaces individuales hasta un sistema de redes que interactuan globalmente. Es posible extender nuestro modelo para incluir varias redes que operan a nivel internacional y que están en constante evolución en analogía con los sistemas naturales. Esto puede dar lugar a fluctuaciones aleatorias de la aptitud intrínseca de las redes y dar lugar a una teoría de selección darwiniana en un ecosistema digital. Un profundo conocimiento del ecosistema digital es crucial para preservar la diversidad en el mundo digital; un requisito para la libertad de información y una democracia digital.

Recientemente, las arquitecturas descentralizadas han atraído mucha atención; por ejemplo, la moneda virtual Bitcoin [29]. Tales sistemas proporcionan una mejor escalabilidad, mayor transparencia y por diseño no pueden ser controladas por entidades centralizadas. Por lo tanto, los sistemas descentralizados jugarán un papel importante en las futuras arquitecturas digitales [12, 30]. Estos sistemas tienen que funcionar sin control central y por lo tanto requieren soluciones a problemas particulares, como la búsqueda y la navegación con únicamente información local (*greedy routing*). Estudiamos este caso explícitamente en términos de espacios geométricos y correlaciones geométricas en redes multicapa.

Se ha demostrado que muchas redes complejas tienen una geometría hiperbólica efectiva debajo de sus topologías observadas, donde coordenadas de nodos representan la popularidad y la similaridad de nodos. Además, las redes reales normalmente no son entidades aisladas sino que forman parte de sistemas más grande y complejos, llamados redes múltiplex o multicapa. Ejemplos de tales sistemas se encuentran en dominios drásticamente diferentes. El sistema más clásico está compuesto por diferentes redes sociales a las que una persona pertenece. Otros ejemplos son las conexiones estructurales y funcionales en el cerebro, o las topologías IPv4 e IPv6 de Internet.

Hemos encontrado que diferentes sistemas reales múltiplex, incluidas las redes cerebrales, las redes de interacción de proteínas, Internet y redes de colaboración científica, no son combinaciones aleatorias de sus redes constituyentes. Por el contrario, encontramos que su redes constituyentes exhiben fuertes correlaciones geométricas. Específicamente, cada red constituyente puede ser mapeada en su propio espacio hiperbólico, en el que las coordenadas de los nodos representan su popularidad y su similaridad. La probabilidad de la existencia de un enlace entre dos nodos es una función continua monótonamente decreciente (la distribución de Fermi-Dirac) de su distancia hiperbólica. A través de los mapas hiperbólicos construidos para una sola red se puede identificar comunidades de nodos, que son grupos de nodos situados cerca uno del otro en el espacio de similaridad angular [84, 126, 130]; predecir las enlaces que faltan con alta precisión [127, 128, 130]; y facilitar el *greedy routing* de forma eficiente. Encontramos que las coordenadas de nodos en diferentes capas están fuertemente correlacionadas, lo que significa que las distancias entre los nodos en los espacios hiperbólicos subyacentes a las redes constituyentes están también fuertemente correlacionadas.

Las correlaciones geométricas descubiertas representan una herramienta potente para entender y analizar sistemas multiplex reales. Se puede utilizar para definir y detectar comunidades multidimensionales, que son conjuntos de nodos simultáneamente similares en múltiples capas. Dos nodos se consideran similares en una cierta capa si están a una distancia angular pequeña en esta capa. Las agrupaciones de nodos que son simultáneamente

similares en diferentes capas son sobreabundantes en comparación con una superposición aleatoria de las topologías de las capas constituyentes. Estos grupos forman comunidades multidimensionales. La estructura geométrica permite cuantificar el grado de similaridad entre las diferentes comunidades, que está relacionada con la distancia angular multidimensional entre ellos. Además, la existencia de correlaciones geométricas permite predecir enlaces de forma trans-capa, de modo que se puede predecir conexiones en una capa mediante la observación del espacio geométrico oculto en otra capa. En particular, los nodos que se encuentran a una pequeña distancia hiperbólica en una capa tienen una probabilidad alta de estar cerca en otra capa y, por lo tanto, la probabilidad de que estén conectados en esta capa es también alta. Como consecuencia de ello, conocer la distancia hiperbólica en una capa permite predecir la probabilidad de conexión en otra capa.

Hemos desarrollado un modelo para crear redes multiplex con correlaciones geométricas. En nuestro modelo, las correlaciones angulares y radiales se pueden ajustar de forma individual e independientemente de las propiedades topológicas de las capas constituyentes. Creamos las correlaciones de las variables ocultas de nodos utilizando técnicas cópula. En particular, consideramos tamaños diferentes de capas, lo cual es el caso en el múltiplex Internet IPv4 / IPv6. En particular, los nodos de grado alto en la capa más grande tienen una probabilidad alta de existir también es la capa más pequeña. El modelo nos permite estudiar de forma aislada el efecto de las correlaciones en el rendimiento de "mutual greedy routing".

En el proceso de *greedy routing*, los nodos pasan un mensaje a su vecino que esté más cerca del destino en el espacio geométrico. Los mensajes llegan a su destino o entran en una trayectoria cíclica y la entrega falla. Para estudiar navegabilidad en sistemas multiplex, extendemos el mecanismo de *greedy routing* de manera que un nodo pasa un mensaje a su vecino más cercano al del destino en cualquiera de las capas que comprenden el sistema. Llamamos a este proceso *mutual greedy routing*. Curiosamente, *mutual greedy routing* supera a la navegación que tiene lugar exclusivamente en las redes constituyentes sólo si las correlaciones geométricas son suficientemente fuertes. Para mostrar este efecto con el ejemplo de Internet creamos multiplexes sintéticas cuyas topologías en las capas individuales imitan a las del sistema real y en las que podemos ajustar las correlaciones geométricas. Comparando nuestro modelo con las coordenadas inferidas del sistema empírico podemos cuantificar las correlaciones presentes en el Internet real. De hecho, el nivel empírico de correlaciones mejora la navegabilidad mutua del sistema de manera significativa usando tanto las distancias hiperbólicas como las angulares. Además, hemos estudiado el caso de más de dos capas de redes. Aumentar el número de capas hace que el sistema sea casi perfectamente navegable siempre que las correlaciones sean suficientemente fuertes. Por el contrario, añadir capas sin correlaciones no mejora la navegabilidad mutua del sistema.

Nuestros descubrimientos pueden dar lugar a una variedad de aplicaciones en muchas disciplinas, desde mejorar el transporte de información y la navegación en los sistemas de comunicación de múltiples capas y arquitecturas de datos descentralizadas, a entender redes estructurales y funcionales del cerebro y descifrar sus relaciones precisas, a predecir conexiones entre nodos, cuando uno conoce sus patrones de conexión en la otra red.

En resumen, esta tesis está dedicada a los retos de un mundo interconectado que ha

emergido a partir de la reciente revolución digital. La penetración del Internet en las sociedades modernas juntamente con la Web 2.0 promueven hoy en día la colaboración global y el intercambio de ideas casi instantáneo entre usuarios de todo el mundo. Sin embargo, la dominación de unos pocos poderosos monopolios de información representa un peligro para la libertad de ideas y decisiones de individuos. Por tanto, dos factores son esenciales para un futuro próspero en la era digital: diversidad digital y decentralización. En cuanto al primero, hemos introducido modelos basados en observaciones empíricas que permiten entender mejor la dinámica y las interacciones competitivas de las redes sociales online, los sistemas claves en el cosmos de la Web 2.0. En particular, nuestros descubrimientos revelan las condiciones en las cuales la diversidad digital se puede sostener. Con respecto al segundo, el diseño de arquitecturas descentralizadas plantea retos específicos. De estos, nos hemos centrado el en problema de búsqueda y navegación basada exclusivamente en conocimientos locales. Hemos revelado en qué condiciones la existencia de muchas redes interaccionando facilita estas tareas. Afortunadamente, muchos sistemas reales cumplen estas condiciones. Para concluir, desde una perspectiva a nivel de sistema, un futuro próspero en el mundo digital compuesto por un paisaje digital diverso con arquitecturas descentralizadas en constante interacción es posible, pero no seguro. En esta situación, la conciencia, así como la creación de los incentivos adecuados, son retos importantes que nuestra sociedad debe afrontar. Crear conciencia suficiente e incentivos correctos para crear ese futuro sigue siendo un reto para la sociedad. Personalmente, espero que nuestros descubrimientos sean útiles para conseguir este ambicioso e importante reto.

Appendix

A Numerical simulations

This chapter was – with some small changes – published in the Supplementary Material of my paper "Digital Ecology: Coexistence and Domination among Interacting Networks" [89]. A preprint version is available at [90].

A.1 Gillespie algorithm

A.1.1 Single network

In the part *Evolution and ecology of the digital world* we simulate the dynamics of the evolution of online social networks. We take advantage of the fact that the temporal events in our model are independent Poisson point processes, which allows us to use the Gillespie [155, 156] algorithm (also known as the Doob-Gillespie [157] algorithm). For a single network, the algorithm works as follows:

1. Initialize the system and fix the rates corresponding to the respective events (here λ, μ, $\delta = 1$)

2. a) Evaluate the number of susceptible nodes (N_S), the number of active nodes (N_A) as well as the number of edges connecting susceptible and active nodes (E_{SA}) and the number of edges connecting active and passive nodes (E_{PA}). Evaluate the sum $\mathcal{S} = \mu N_S + \lambda \left[E_{SA} + E_{PA}\right] + \delta N_A$.

   b) Generate random numbers to choose the next event. The probabilities for the events are the following:

      - Mass media activation: $\mu N_S/\mathcal{S}$
      - Viral activation: $\lambda E_{SA}/\mathcal{S}$
      - Viral reactivation: $\lambda E_{PA}/\mathcal{S}$
      - Deactivation: $\delta N_A/\mathcal{S}$

   c) Evaluate the corresponding time step τ. The corresponding time step is given by $\tau = \mathcal{S}^{-1}$.

3. Update the status of the system. So, if in step 2 a mass media activation was chosen, we randomly choose a susceptible node and change its status to active. For a deactivation, we randomly choose an active node that then becomes passive. In the case of viral activation, we randomly choose an edge connecting a susceptible and an active node and activate the susceptible node at the end of the link. Similarly, for viral reactivation

we randomly choose an edge connecting a passive and an active node and the passive node at the end of this edge becomes active. We increase the time: $t \to t + \tau$. We iterate by returning to step 2 until the end of the simulation is reached.

A.1.2 Multiple network layers

The generalization of the algorithm to multiple layers is straightforward. One evaluates the probabilities of having a certain dynamical process in a certain layer; for example, mass media activation in layer i occurs with probability $\mu_i N_{S,i}$, where $N_{S,i}$ denotes the number of susceptible nodes with respect to layer i (all the nodes which are in the underlying network but not in the i-th layer) and μ_i is the corresponding rate in layer i. Accordingly, the probability of viral activation in layer i is given by $\lambda_i E_{SA,i}$, where $E_{SA,i}$ is the number of edges connecting active and susceptible nodes in layer i. One then chooses a dynamical process in a certain layer in accordance with these probabilities. Finally, τ is given by the inverse of the sum over all these probabilities in all the layers.

B Topological evolution of isolated online social networks

This chapter was – with some small changes – published in the Supplementary Material of my paper "Evolution of the Digital Society Reveals Balance between Viral and Mass Media Influence" [40]. A preprint version is available at [41]. Most figures are identical to the preprint version.

B.1 Null model

In Fig. B.1 we present the results from a null model for the role of the underlying empirical network. We again use the Pokec network as underlying network and add nodes completely randomly to the network. Note that this corresponds to our model for $\lambda = 0$ and arbitrary $\mu > 0$. However, the choice of μ then just fixes the model timescale, which we adjust implicitly by transforming physical time to the intrinsic network timescale given by the number of nodes. We observe that the phase transition takes place at a larger network size. Note that there is no more parameter to adjust. The number of components (see inset in Fig. B.1) also varies strongly between the empirical network and the presented null model.

We conclude that the occurrence of the phase transition is included in the structure of the underlying network. Nevertheless, a null model with exclusively random subscriptions fails to reproduce the critical point of the phase transition as well as the evolution of the number of components.

B.2 Pathlength and diameter

We observe the same behavior in the evolution of the average shortest path length and the network diameter (see Fig. B.2). In the connected regime the pathlength and the diameter within the GCC decrease. In the disconnected regime, the average shortest path length and the diameter increase and reach their maximum at the critical point. The extended model exhibits the same behavior.

B.3 Sustained activity threshold for extended model

In Fig. B.3 we compare the critical parameter λ_c for the basic and extended model for the PAP dynamics within the online social network layer, which is equivalent to the SIS model.

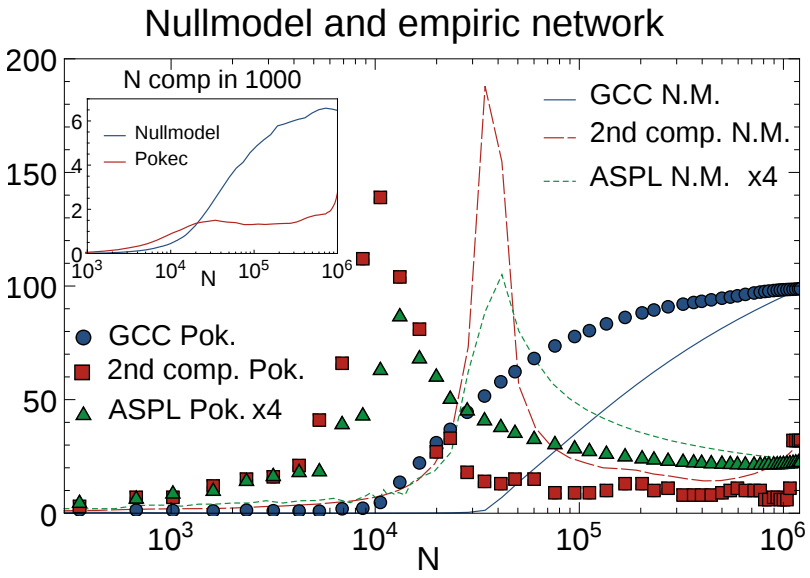

Figure B.1: A null model with underlying Pokec network that consists of randomly adding nodes to the online social network. Points correspond to the empirical network and solid lines correspond to the null model.

We find that

$$\lambda_c \approx 0.020 \sim 0.025\,, \tag{B.1}$$

which appears to be quite robust to the assignment of weights. Below this threshold, the whole network become passive. We suggest that this corresponds to the practical disappearance of the network as observed in many real online social networking services.

B.4 Average finite cluster size

Alternatively to the size of the second largest component one can consider the average finite cluster size, which is the average size of disconnected components without the largest one. The average finite cluster size also exhibits a peak at the critical point. In Fig. B.4 we show the average finite cluster size for all components with size $N > 1$.

B.5 Degree distribution of Pokec OSN

In Fig. B.5 we show the degree distribution of Pokec at different times.

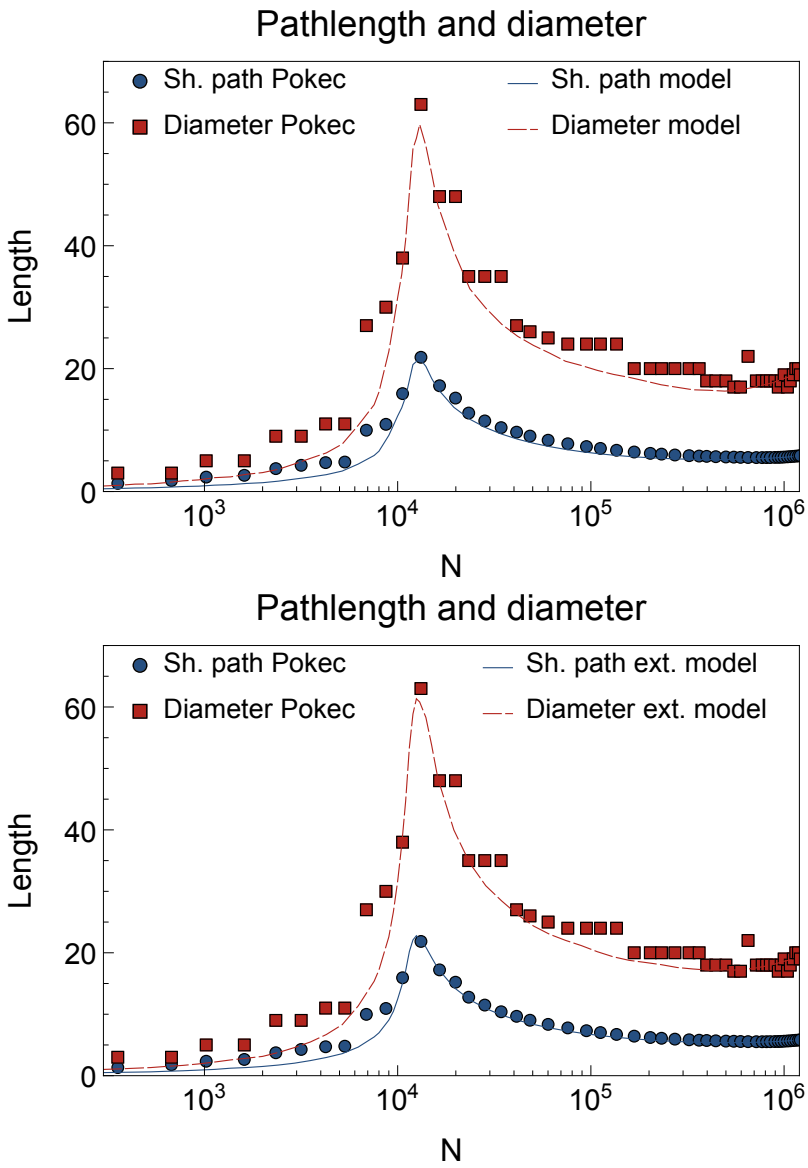

Figure B.2: Pathlength and diameter for basic model (top) and extended model (bottom).

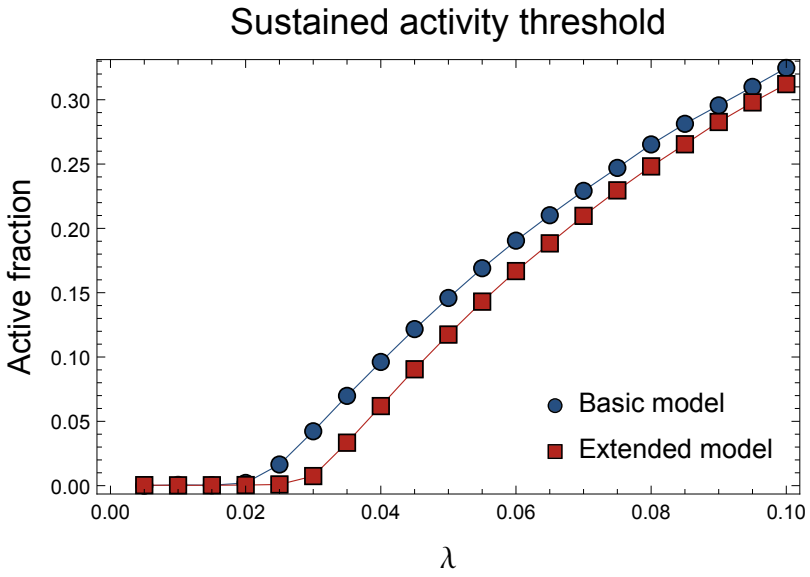

Figure B.3: Sustained activity threshold $\lambda_c \approx 0.020 \sim 0.025$. Below this threshold, the activity of the network is not sustained and eventually the whole network will become passive.

B.6 Delayed edge formation in OSN layer

To test our assumption of instantaneous link formation, we performed the following experiment. Instead of assuming the instantaneous existence of a link when both of its end nodes exist in the OSN layer, we delay its formation in the sense that we create them with a rate ξ from this point on. The rate ξ clearly has to be larger than δ, which is the rate of becoming inactive. In Fig. B.6 we present the results for $\xi = 10\delta$ and $\xi = 50\delta$, which we compare with our model ($\xi \to \infty$). We observe that the position of the critical point is barely affected by the edge creation delay (see Fig. B.6A and C). The initial increase in the clustering coefficient is shifted to slightly larger network times, however, for $\eta = 0$ it reaches values similar to the case of instantaneous link formation (see Fig. B.6B). The same tendency is observed for $\eta = -0.65$ (see Fig. B.6C). It is important to note that the assumption of instantaneous link creation is used for the Pokec network and the model consistently. To sum up, if we assume that links are created at a timescale which is significantly smaller than the timescale at which users stop to use the network, our approximation works fine.

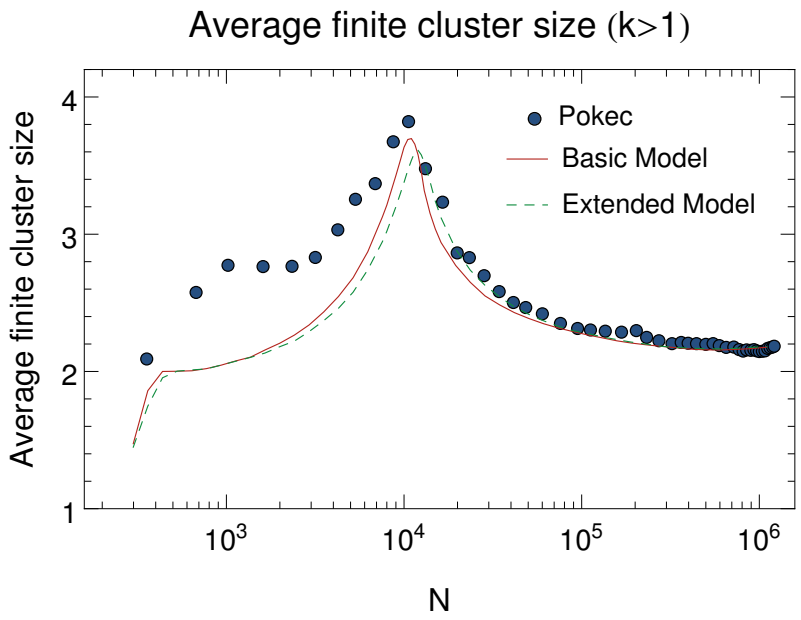

Figure B.4: The evolution of the average finite cluster size for the Pokec network, the basic model, and the extended model for the same parameters as used Fig. 2.2 on page 20.

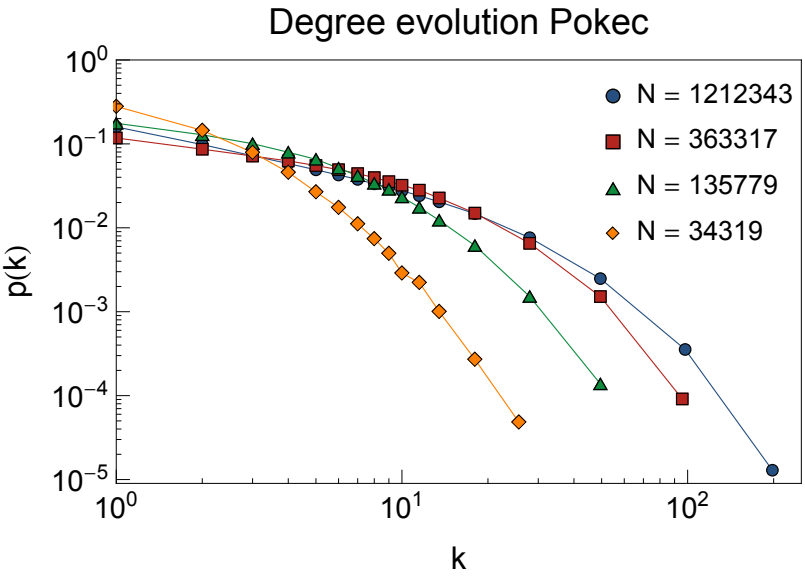

Figure B.5: Degree distribution of Pokec OSN at different times.

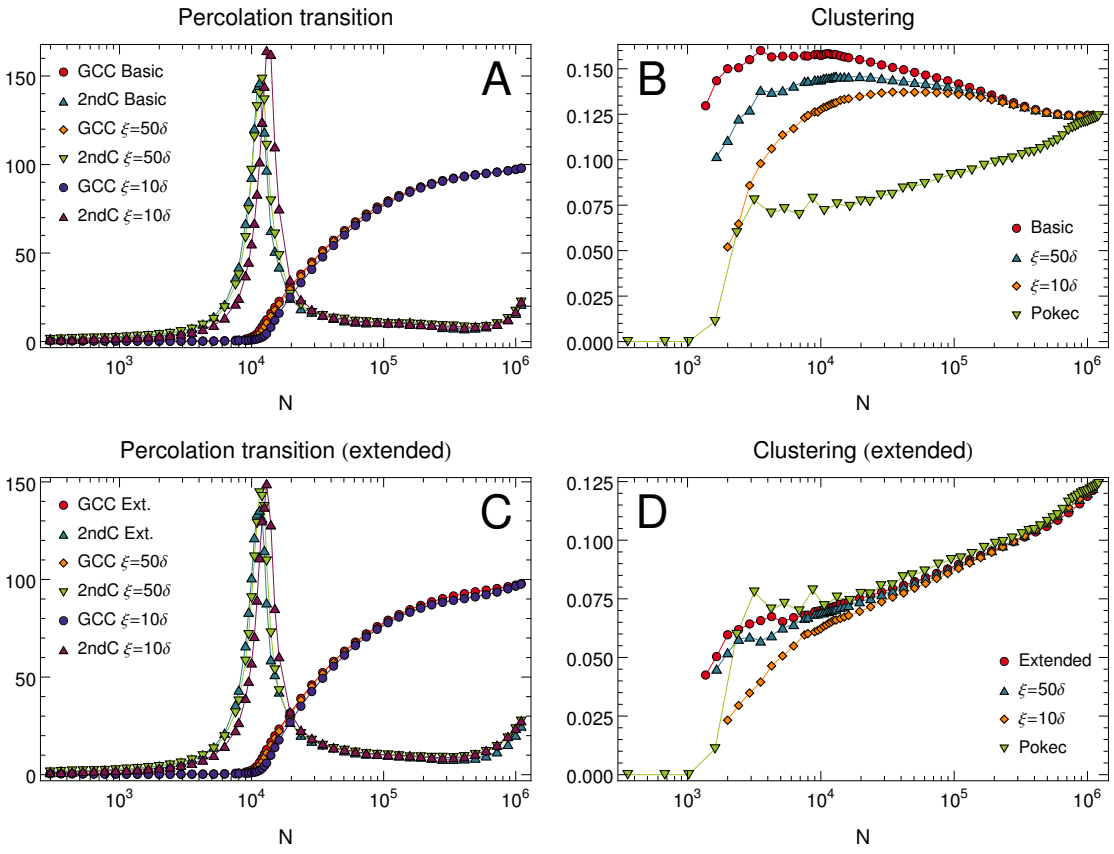

Figure B.6: Results for delayed edge creation for rate $\xi = 10\delta$ and $\xi = 50\delta$. **A:** Relative size of the GCC and absolute size of the second largest component for the basic model. **B:** Mean local clustering coefficient for the basic model. **C:** Relative size of the GCC and absolute size of the second largest component for the extended model. **D:** Mean local clustering coefficient for the extended model.

C Digital ecology

This chapter was – with some small changes – published in the Supplementary Material of my paper "Digital Ecology: Coexistence and Domination among Interacting Networks" [89]. A preprint version is available at [90].

C.1 Empirical stability

Independent of the topological properties of the network, the activities for the steady state solution for an arbitrary number of layers is encoded in the activity curve of a single layer, which we show in Fig. 2.6 on page 25. At the steady state of n_c coexisting networks, each prevailing layer has the same share of the total virality $\lambda_i = \omega_i \lambda = \frac{\lambda}{n_c}$, whereas the remaining ones have $\lambda_j = 0$. The steady state activity of the i-th network is then given by the activity value of a single layer shown in Fig. 2.6 at $\lambda = \lambda_i$.

D World model

This chapter is – with some small changes – available as a preprint at [102]. Most figures are identical to the preprint.

D.1 S1 model

We use the $\mathbb{S}1$ model [110–112] to generate the synthetic networks for the underlying societies in each country. The model allows us to specify the degree distribution and the level of clustering. The model is based on a circle as a hidden metric space and works as follows:

1. All nodes are placed on the circle with a randomly assigned variable, θ, which represents the polar coordinate. θ is uniformly distributed in $[0, 2\pi)$. To keep the average node density on the circle constant, its radius grows linearly with the number of nodes, to satisfy $N = 2\pi R$.

2. We assign each node a second hidden variable, κ, which represents its expected degree. κ is drawn from an arbitrary distribution $\rho(\kappa)$.

3. A pair of nodes is connected with a probability, r, that depends on their hidden variables (θ, κ) and (θ', κ')[1]

$$r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu\kappa\kappa'}\right)^{-\alpha} , \tag{D.1}$$

   with $\mu = \frac{\alpha-1}{2\langle k\rangle}$. Here, $d(\theta, \theta')$ denotes the geodesic distance between the two nodes on the circle and $\langle k\rangle$ the mean degree. Then, the expected degree, $\bar{k}(\kappa)$, of a node with hidden variable κ can be shown to be proportional to κ [112]. As a consequence, the degree distribution, $p(k)$, of the network follows the shape of the distribution $\rho(\kappa)$.

Here, we use an exponential distribution $\rho_\xi(\kappa) = \xi e^{-\xi\kappa}$ with $\xi = 10$. We set the parameters $\alpha = 1.5$ and $\mu = 0.02$. After generating the networks, we remove nodes with zero degree. Fig. D.1 shows the degree distribution and the clustering spectrum for the synthetic network created for the US.

[1] In part *Geometry of multiplex networks* we use a slightly different version of this model.

D.2 Air travel data

Air travel data aggregated on a country basis was taken from http://visualising.org/datasets/global-flights-network (date of access July 2014). The original data can be accessed at http://openflights.org/data.html. The network on a country basis incorporates 230 nodes and 4600 weighted edges which correspond to the number of routes between countries, i.e. the number of total flights. The dataset contains around 60.000 of such flights. We extract the subnetwork (see Fig. 4.1c on page 56) by constraining to the countries listed in Tab. 4.1 on page 62.

D.3 Estimation of data variance

We estimate the variance σ_N^2 of the data of the number of countries where local networks prevail by performing a fit and evaluating the deviation of the datapoints from this fit (see Fig. D.2). We find

$$\sigma_N^2 \approx 1.5\,. \tag{D.2}$$

D.4 Google trends data

In Fig. D.3 we show the evolution of the Google search volume for Facebook. The dashed line corresponds to the time of global launch in our model for the best parameter estimate.

D.5 Double meanfield approximation: $\bar{\Omega} > 0$ breaks symmetry of pitchfork bifurcation

The evolution equations for the double meanfield approximation contain an additional control parameter $\bar{\Omega}$. For $\bar{\Omega} = 0$ we recover the case of two competing identical networks as discussed in chapter 3. In this case, the system undergoes a subcritical pitchfork bifurcation. Such bifurcation is symmetric locally near the critical point. However, the additional control parameter $\bar{\Omega} > 0$ breaks this symmetry. As a consequence, the system undergoes a saddle-node bifurcation instead of the former pitchfork. See Fig. D.4.

D.6 Fate of single realizations in the coinflip region

In Fig. D.5 we show the evolution of the number of countries where local networks prevail for different realizations with the same parameters in the coinflip region. Clearly, there are two different classes of trajectories: first, those that are constant (red, international network becomes extinct), and secondly, those that decrease with time (blue, international network persists). The empirical case corresponds to the latter, and hence we only average over trajectories that belong to this class in section 4.5.

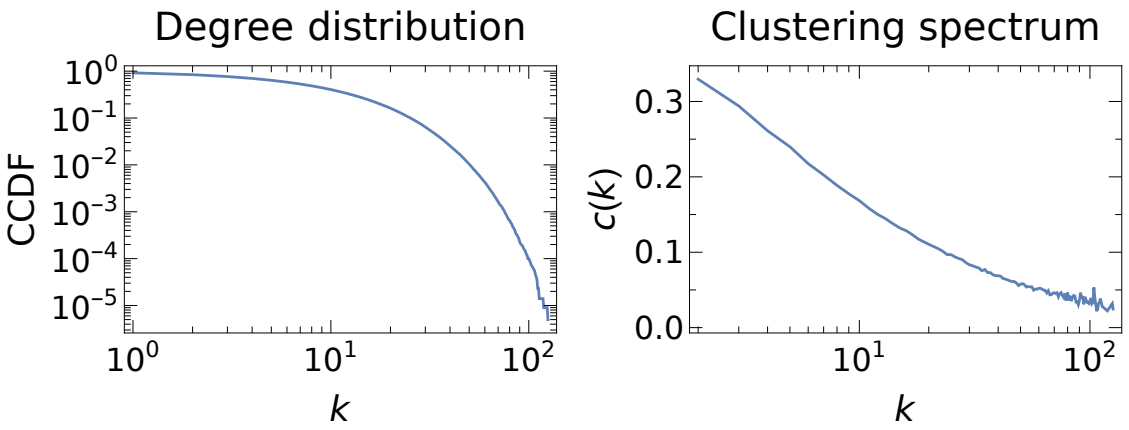

Figure D.1: Degree distribution and clustering spectrum for networks generated for the example of the US ($\approx$ 230.000 nodes).

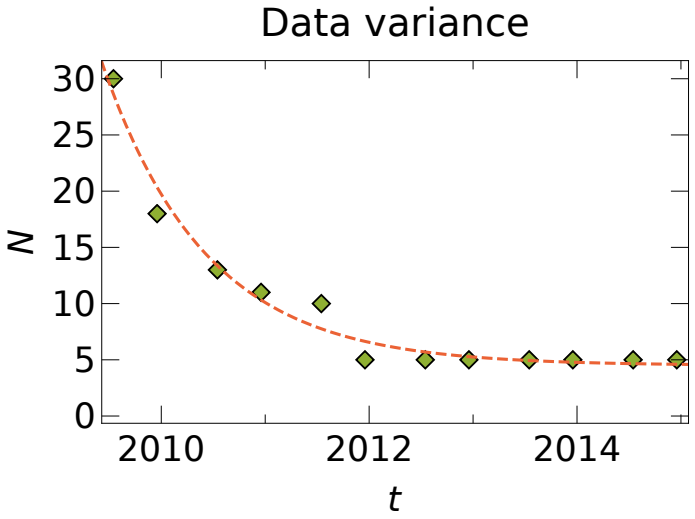

Figure D.2: Estimation of data variance. Symbols denote data and the red line a fit (exponential decay).

D.7 Guideline to develop fine grained description

In this section, we provide a guideline to generate a fine grained description for the evolution of networks in a certain country or region of interest. Similar to large eddy simulations in fluid dynamics [158, 159], we preserve a high accuracy in the region of interest for the prediction and rely on coarse grained approximative dynamics beyond.

The process to create a customized predictive model is as follows.

1. Choose a country of interest.
2. Define the region of influence given by the connections with the highest weight connected to the country of interest.
3. Gather empirical historical data of the evolution of the local networks and the international network in the region of influence.
4. Adjust parameters within the region of interest locally and use generic global parameters for the remaining countries. A suggestion of parameters to adjust locally in the influence region can be found in Tab. D.1.

Let us consider the example of Brasil (see Fig. D.6). A possible choice of the influence region would be the USA, Argentina, Uruguay, Spain, and Portugal. We then propose to adjust the following three parameters locally in the region of influence: the virality λ, the media influence μ, and the launch time delay Δt. Adjusting these parameters in each country would lead to 240 parameters that have to be adjusted simultaneously, which is not feasible and would require much data. However, by performing the above procedure, we reduce significantly the number of parameters (to $\approx 12 - 18$) and maintain an acceptable level of precision in the region of interest. Hence, this hybrid approach of adjusting parameters locally within the region of influence and globally beyond enables the development of precise customized predictions with a feasible effort.

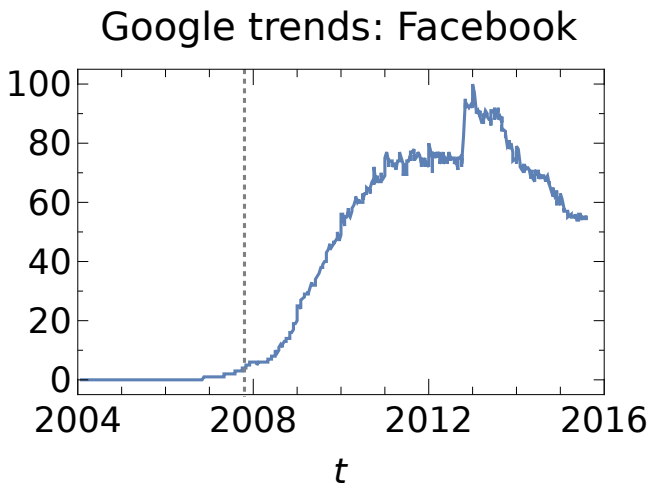

Figure D.3: Google trends results for Facebook. The gray dashed line shows the time of global launch predicted by our model.

Quantity	Parameter	Suggestion
Activity affinity	σ	Global
Global connectivity	α	Global
Virality	λ	Local in region of influence
Media influence	μ	Local in region of influence or educated guess [40]
Launch time delay	Δt	Local in region of influence

Table D.1: Suggestion of parameter adjustment.

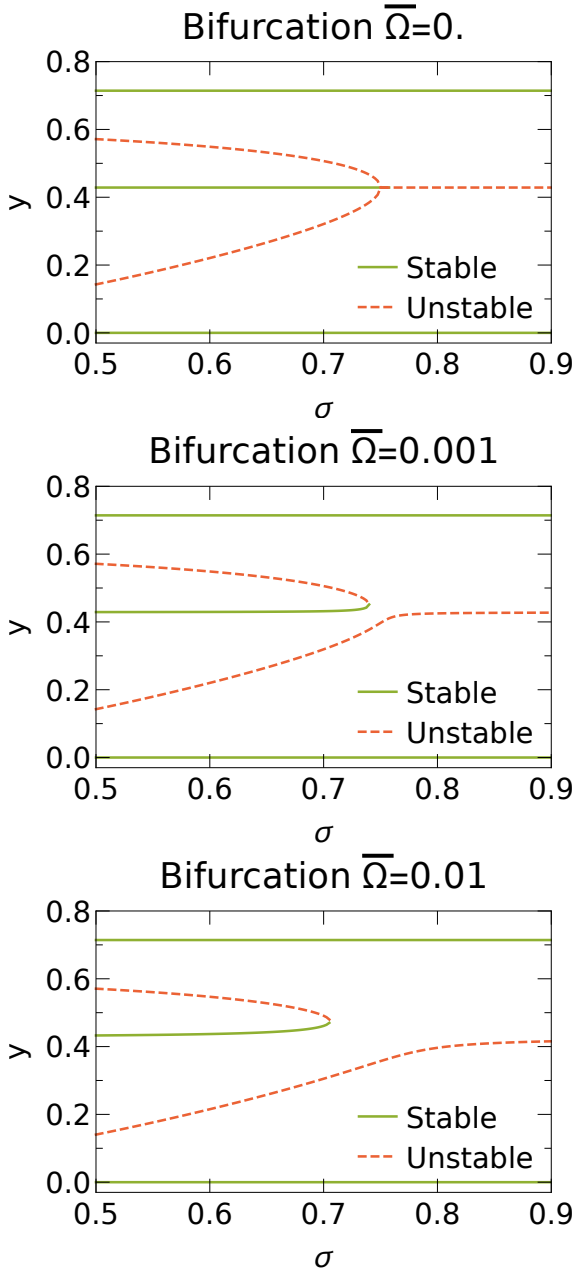

Figure D.4: Bifurcation diagram as a function of the control parameter σ for different values of $\bar{\Omega}$. Here, $\lambda \langle k \rangle = 3.5$.

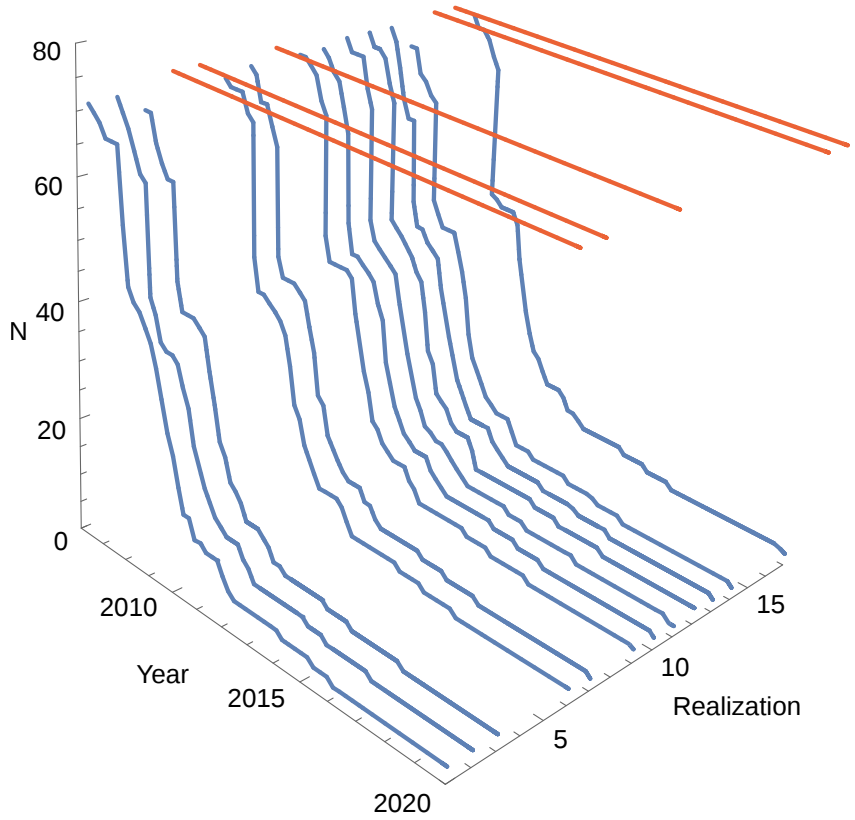

Figure D.5: Number of countries where local networks prevail as in Fig. 4.7b on page 67 for 17 realizations of the system. Parameters are as in Fig. 4.7b.

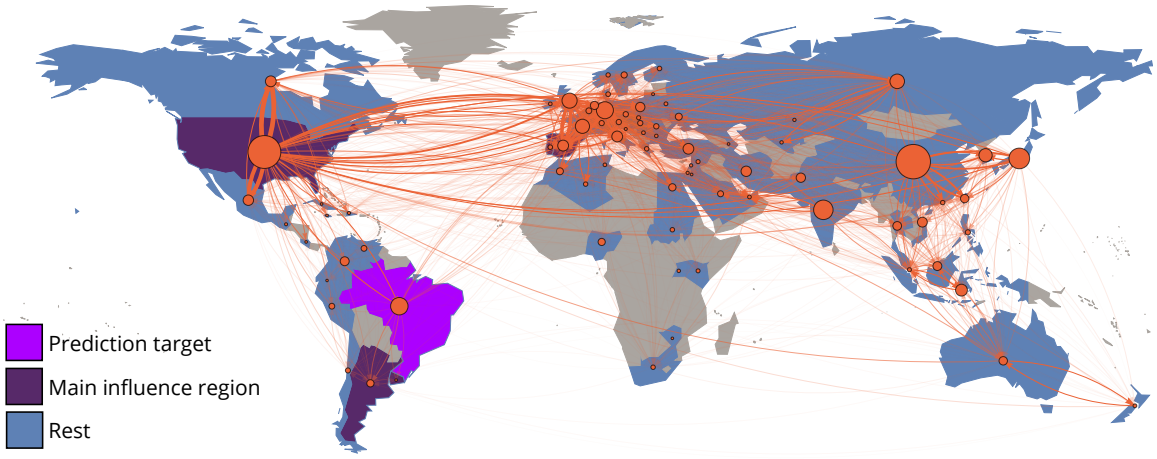

Figure D.6: Possible choice of region of influence for Brasil.

D.8 Explicit time series

In the following, we present explicit time series of single realizations of our model for different parameters.

Videos are available online, see [160, 161].

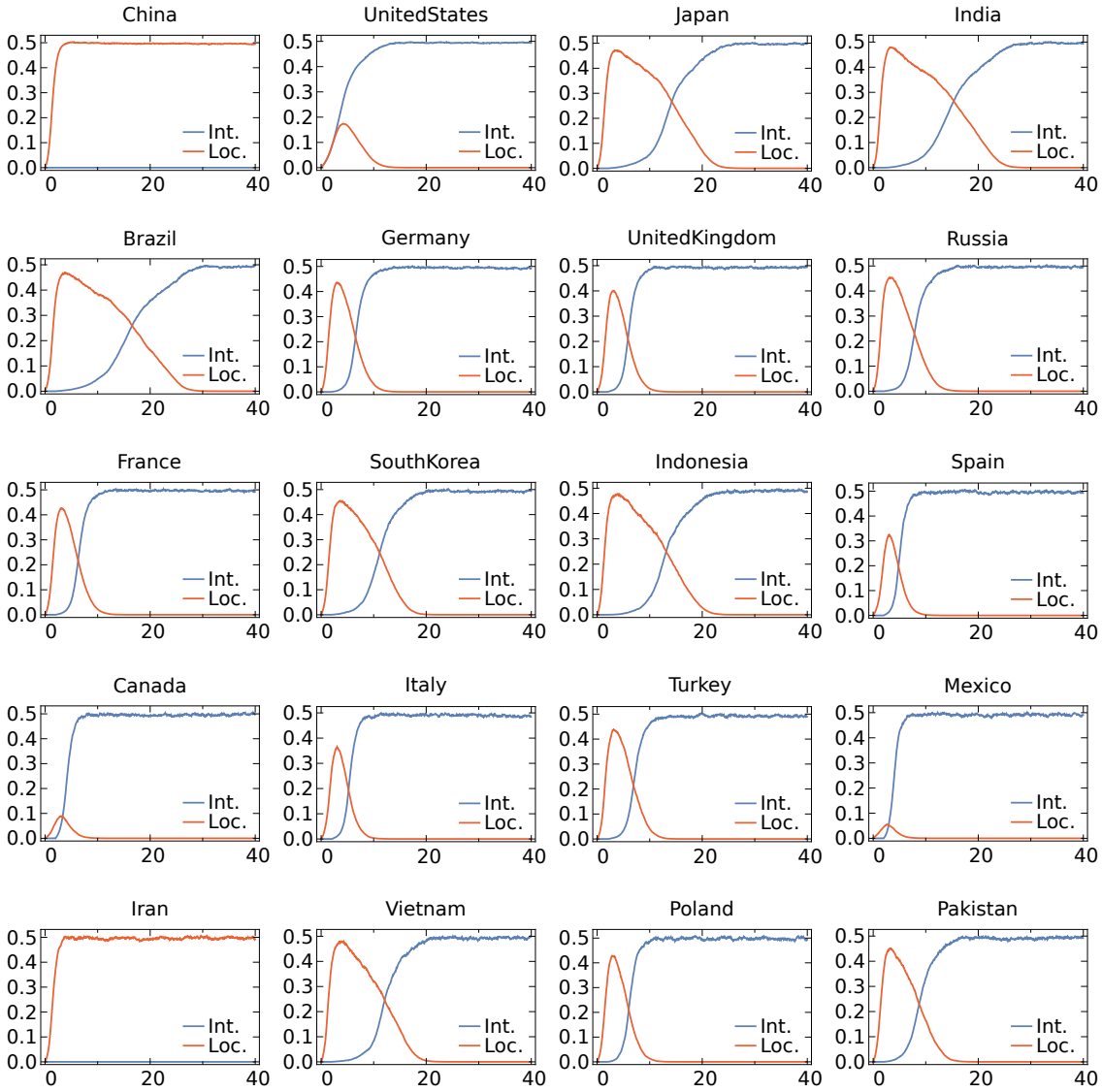

Figure D.7: Evolution of network activity for the first 20 countries. Here, $\lambda = 0.2$ per country, $\sigma = 0.75$, $\Delta t = 2$, and $\alpha = 2$.

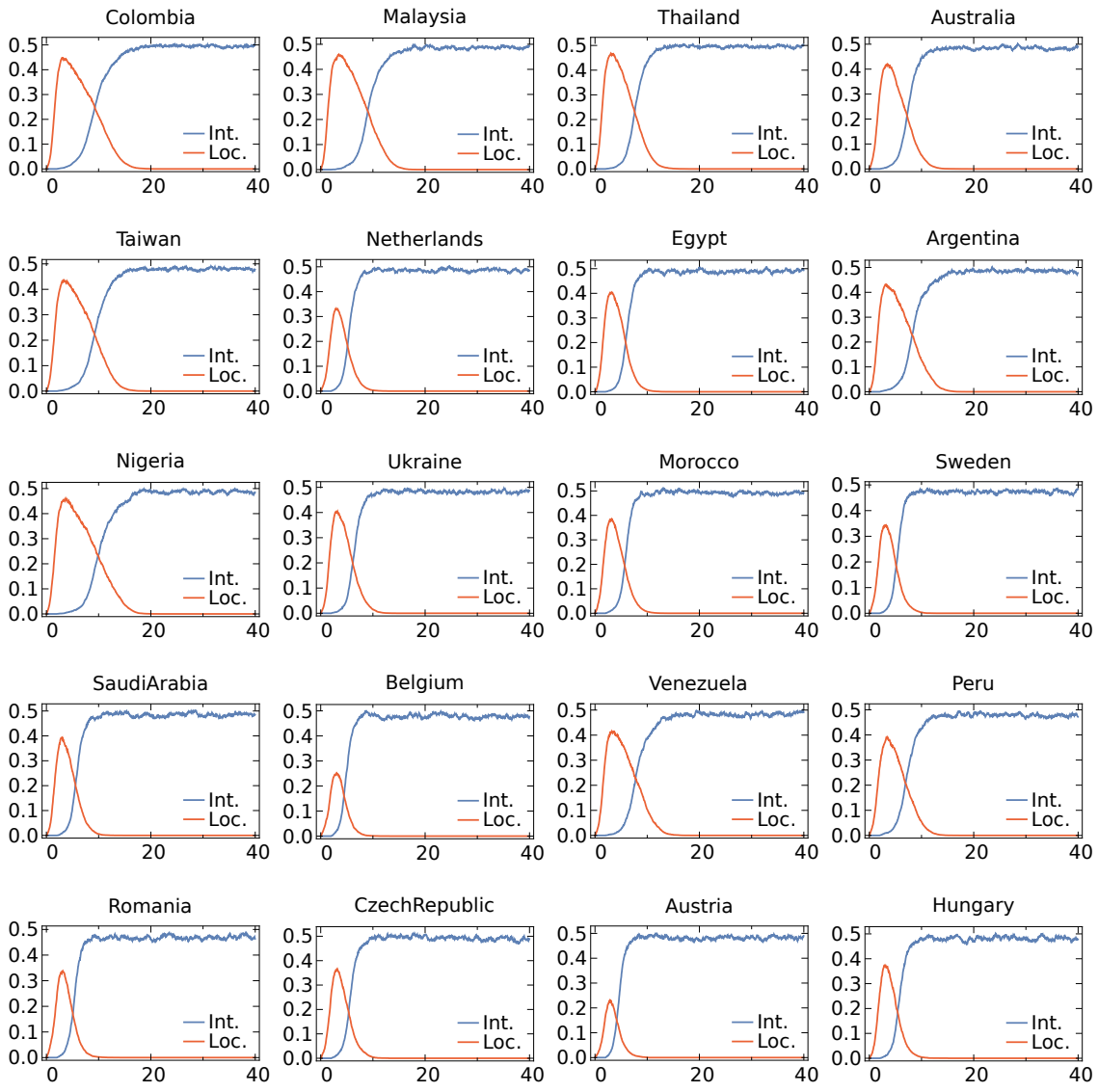

Figure D.8: Evolution of network activity for the second 20 countries (continuation from Fig. D.7).

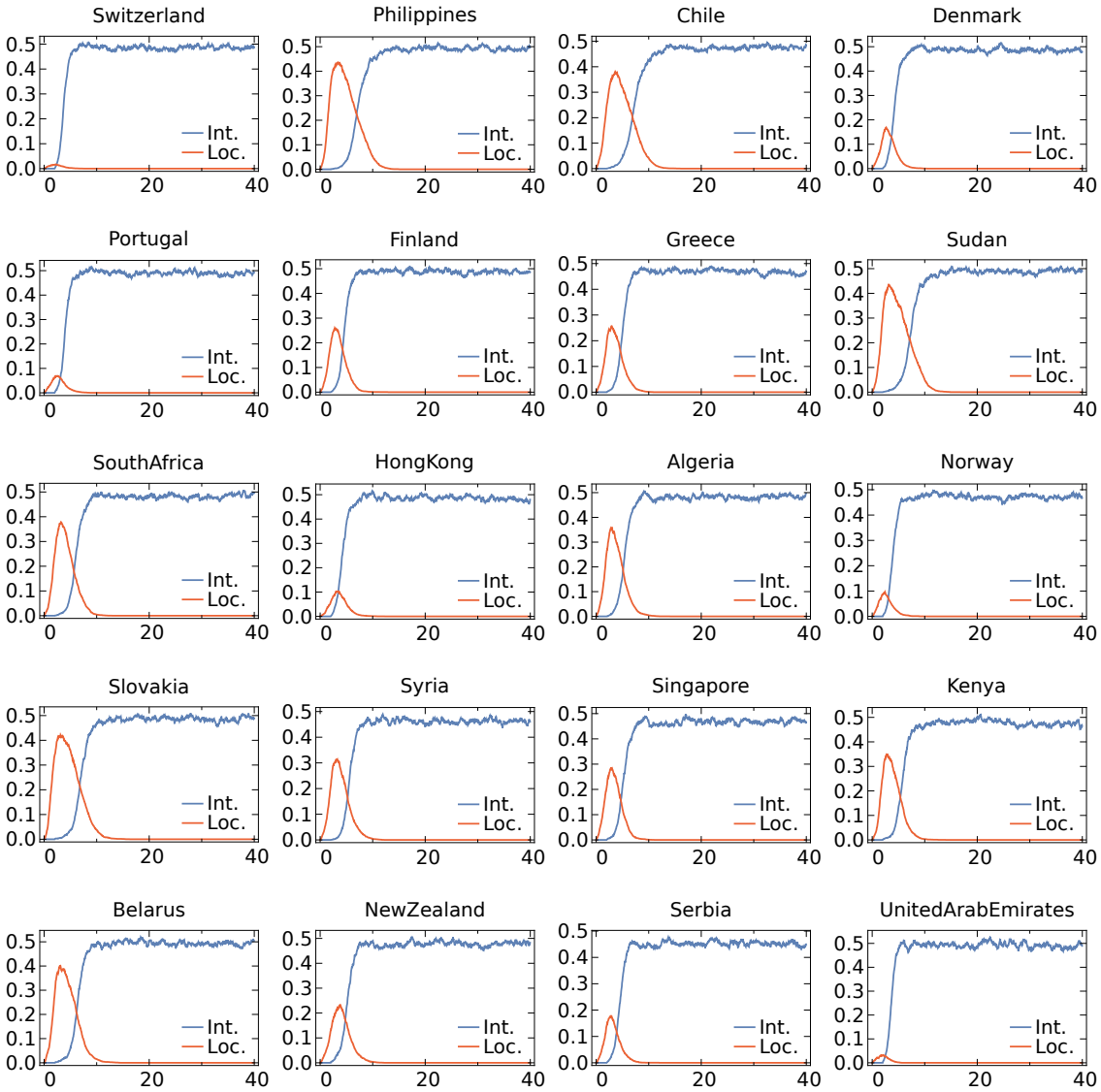

Figure D.9: Evolution of network activity for the last 20 countries (continuation from Fig. D.8).

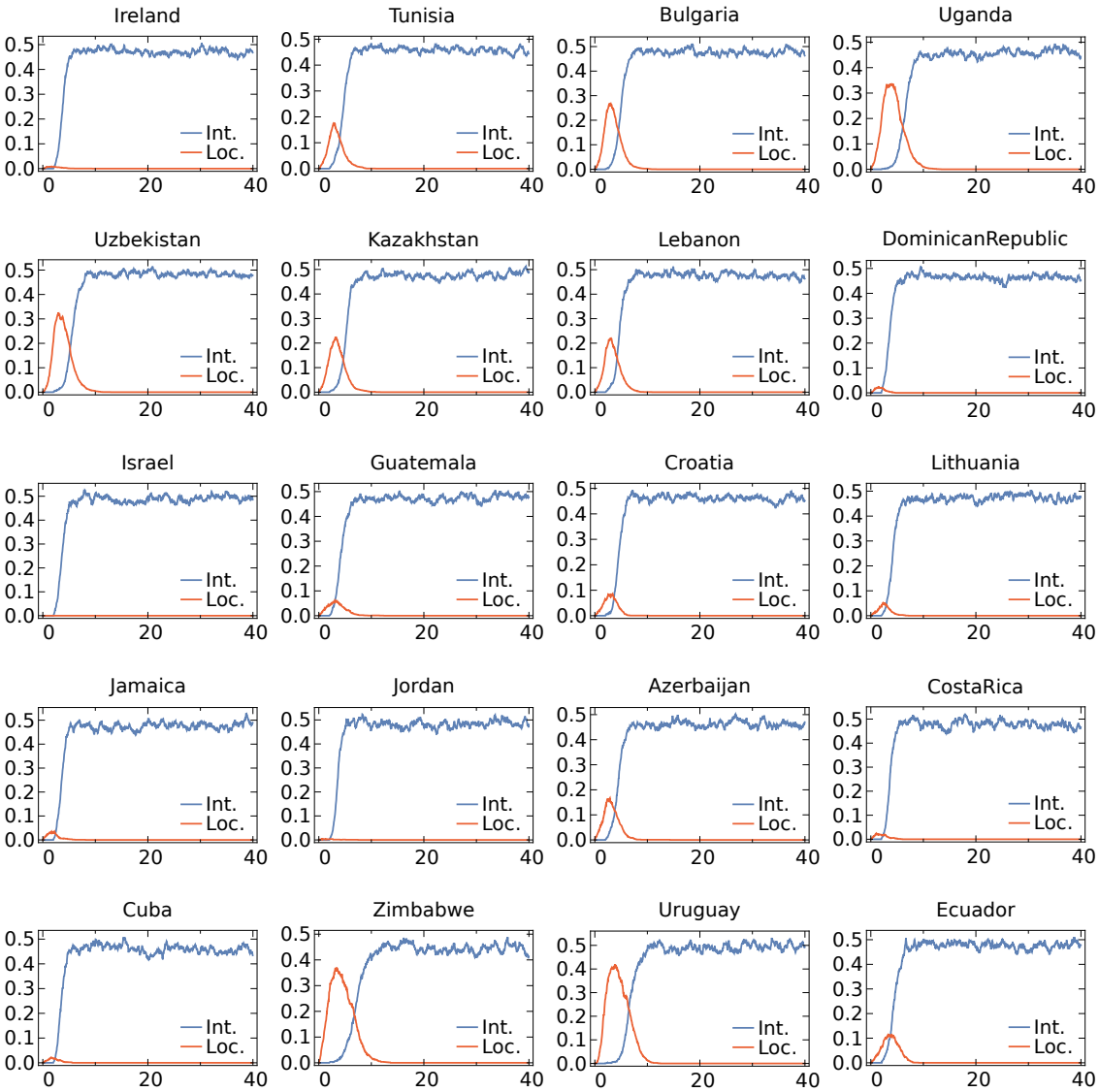

Figure D.10: Evolution of network activity for the last 20 countries (continuation from Fig. D.9).

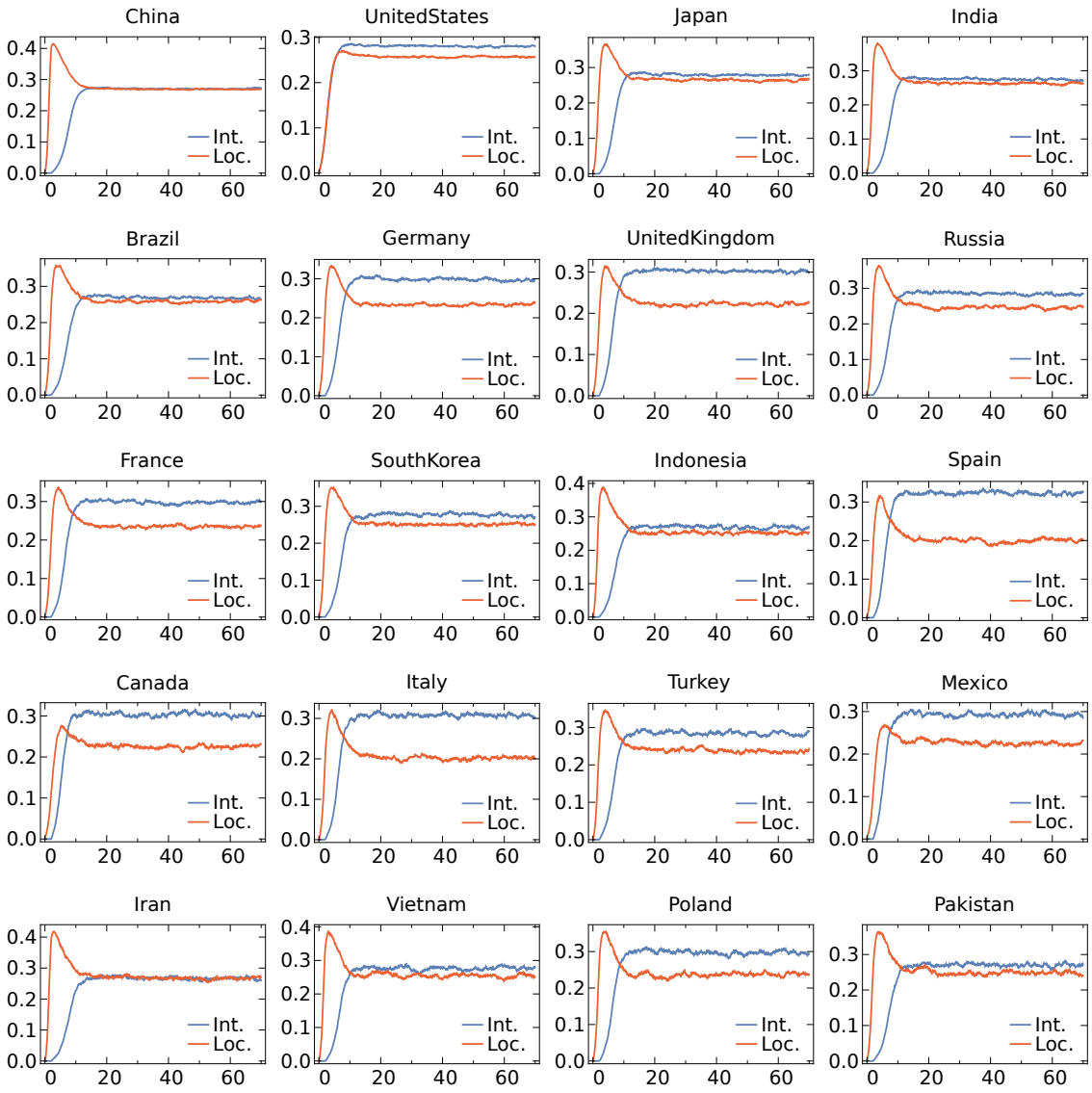

Figure D.11: Evolution of network activity for the first 20 countries. Here, $\sigma = 0.25$, $\Delta t = 2$, $\alpha = 0.75$, and $\lambda = 0.2$ per country.

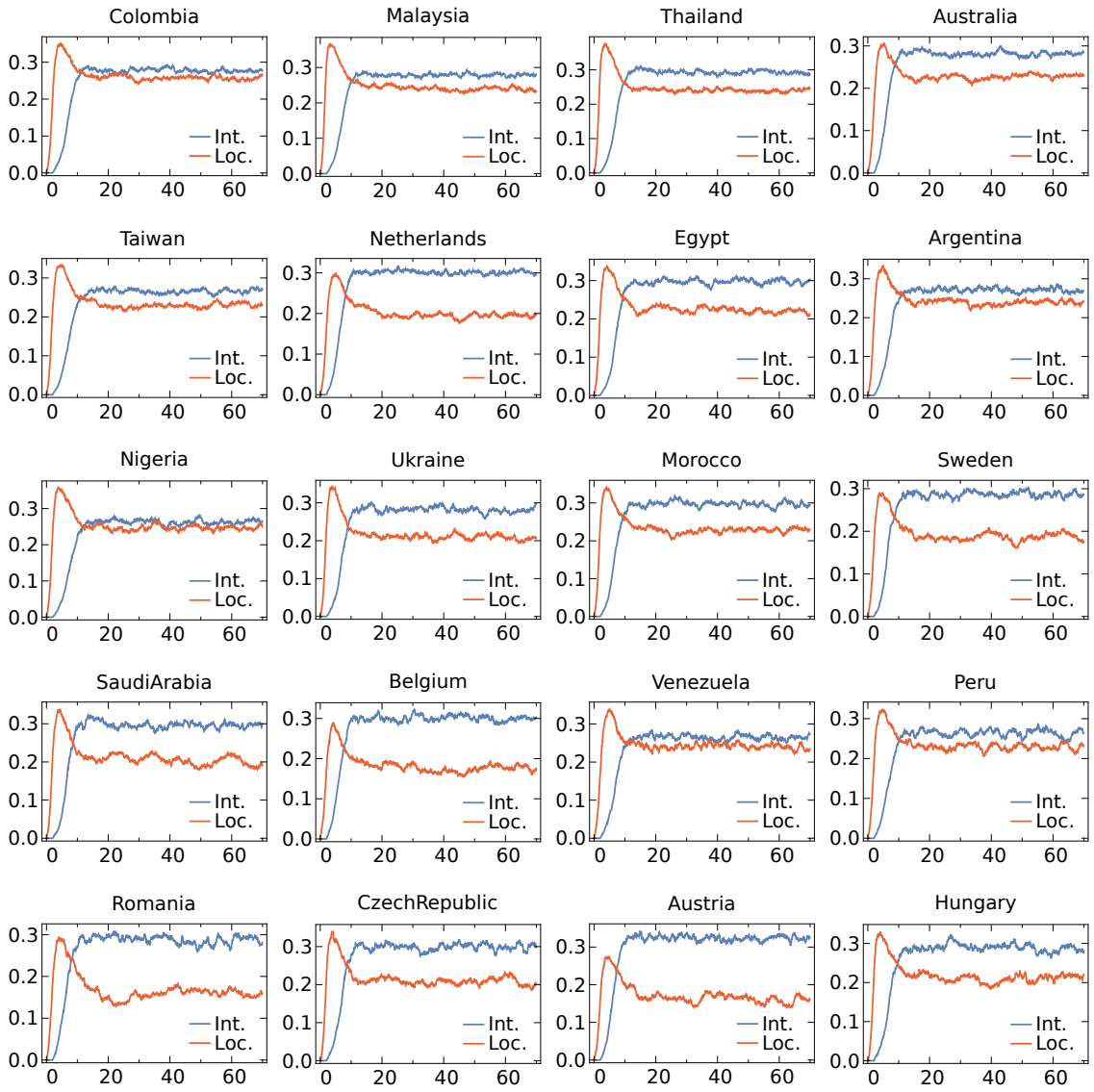

Figure D.12: Evolution of network activity for the second 20 countries (continuation from Fig. D.11).

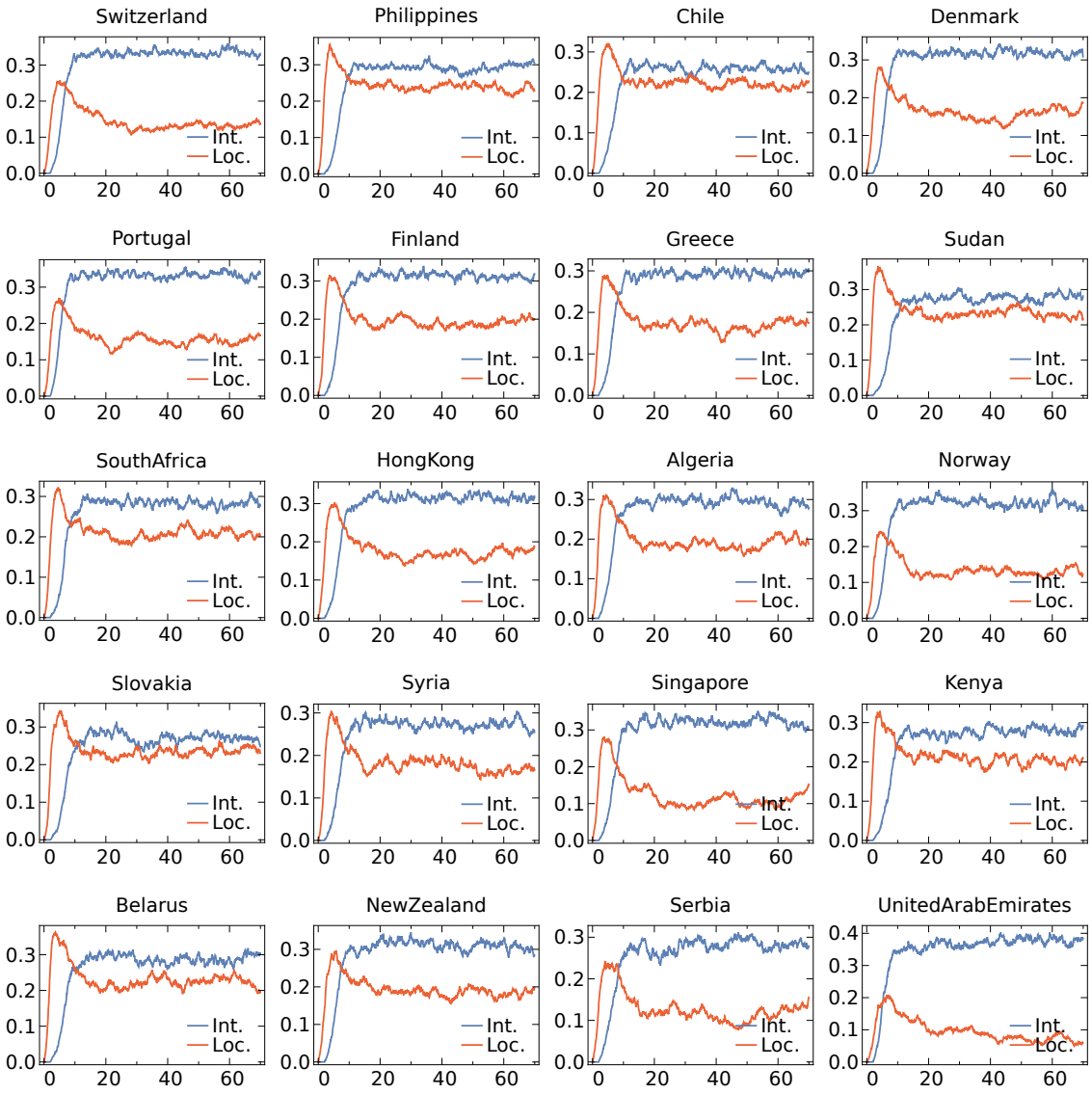

Figure D.13: Evolution of network activity for the last 20 countries (continuation from Fig. D.12).

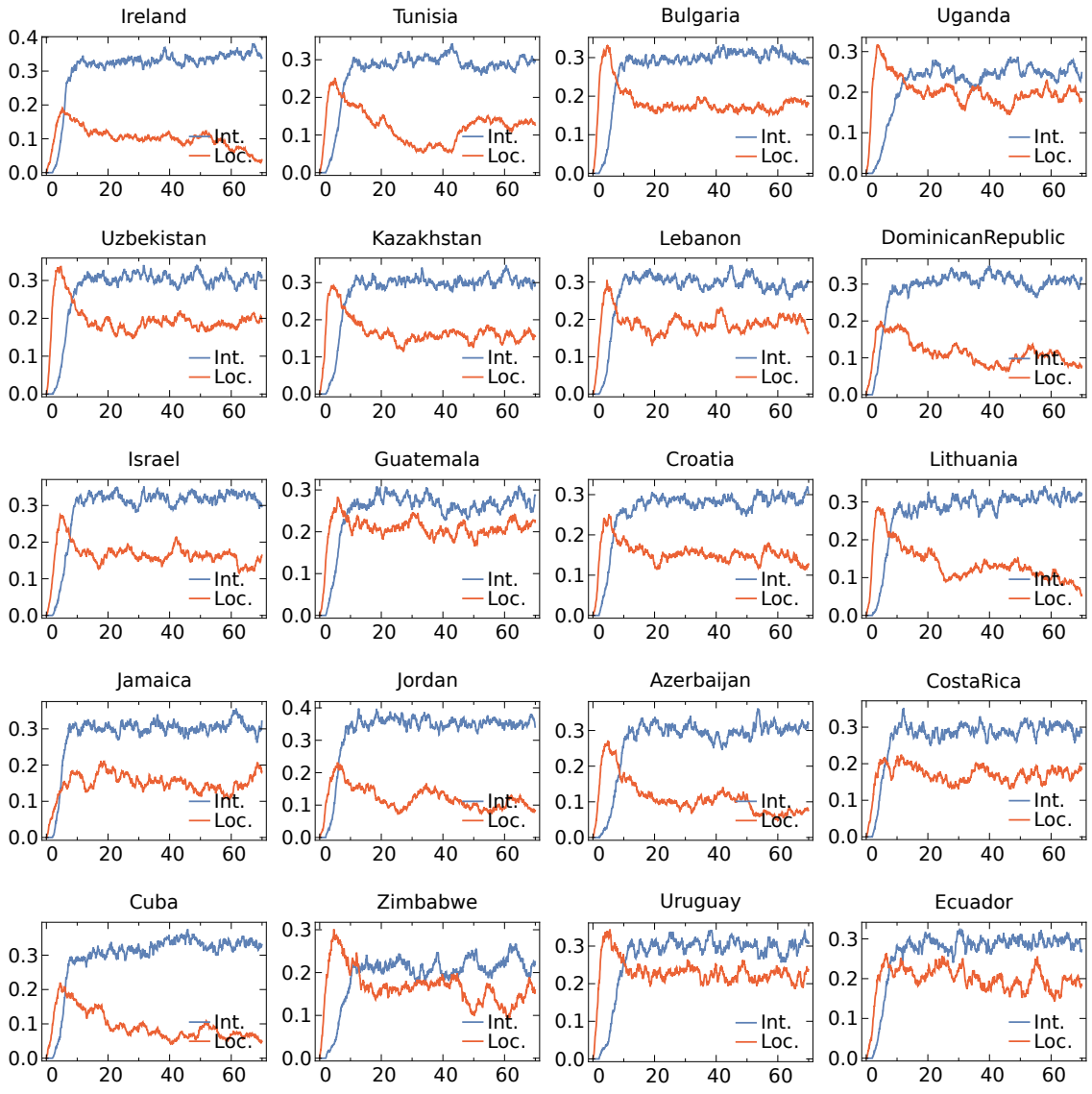

Figure D.14: Evolution of network activity for the last 20 countries (continuation from Fig. D.13).

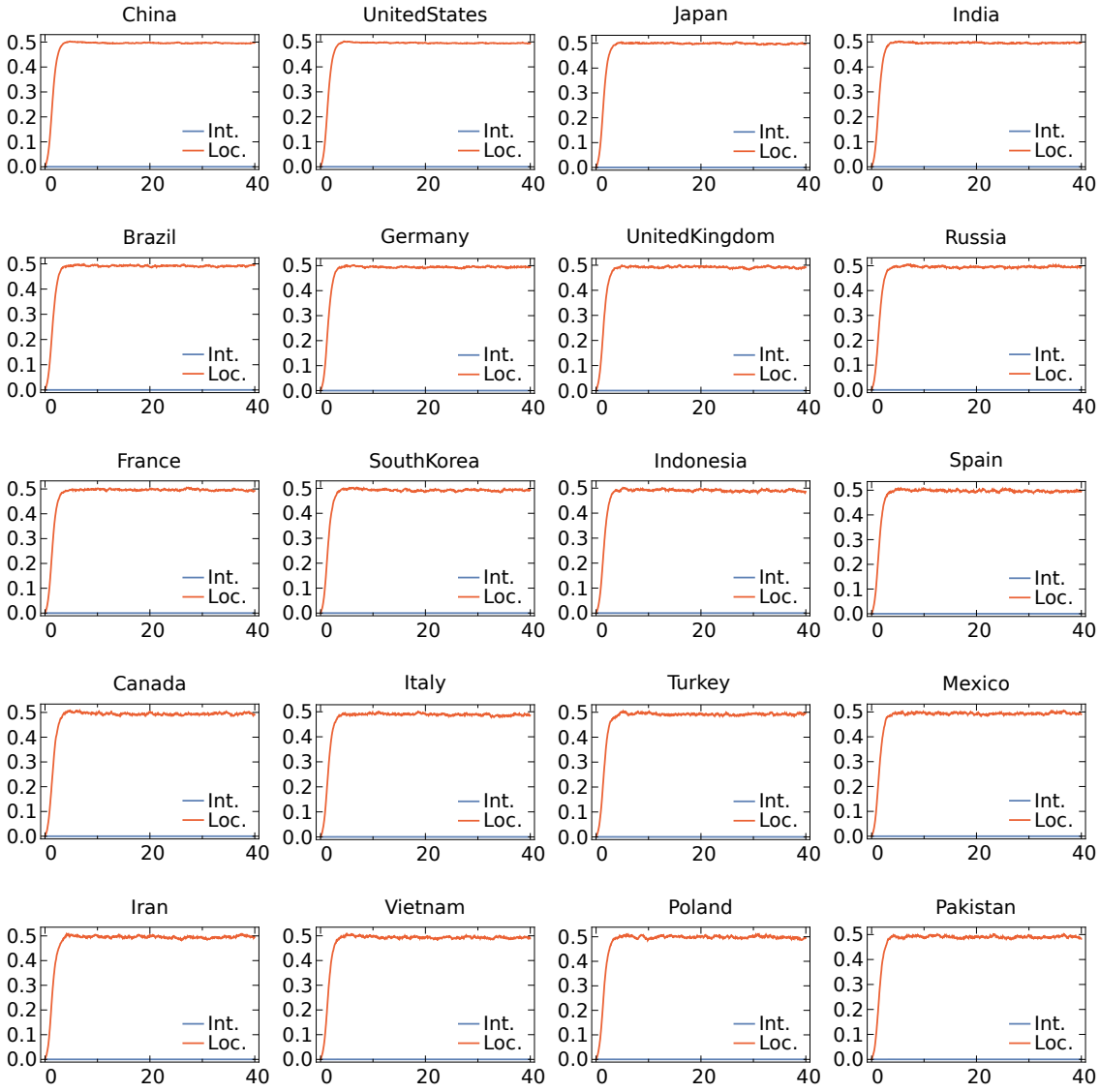

Figure D.15: Evolution of network activity for the first 20 countries. Here, $\sigma = 1.5$, $\Delta t = 3$, $\alpha = 2$, and $\lambda = 0.2$ per country.

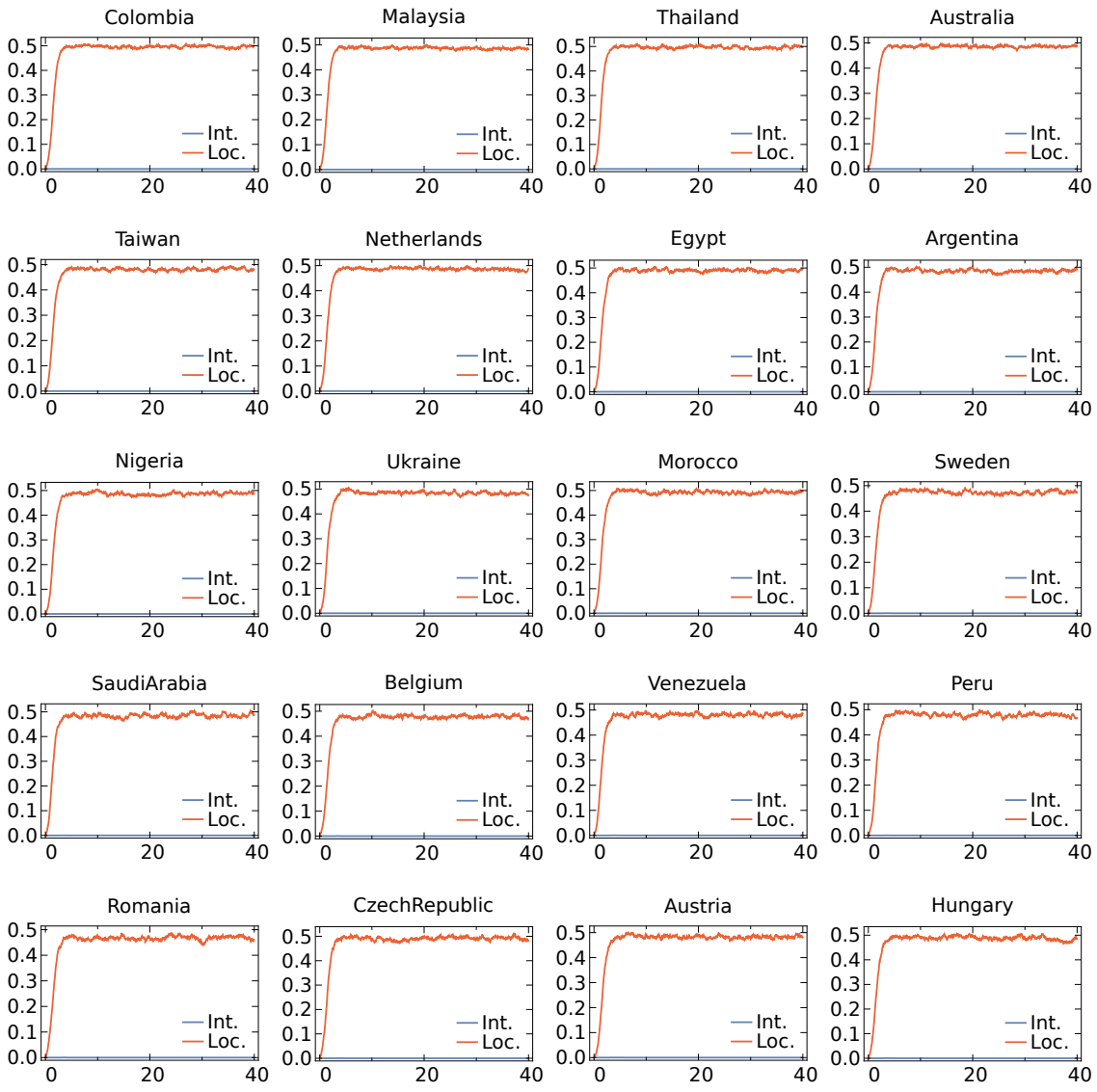

Figure D.16: Evolution of network activity for the second 20 countries (continuation from Fig. D.15).

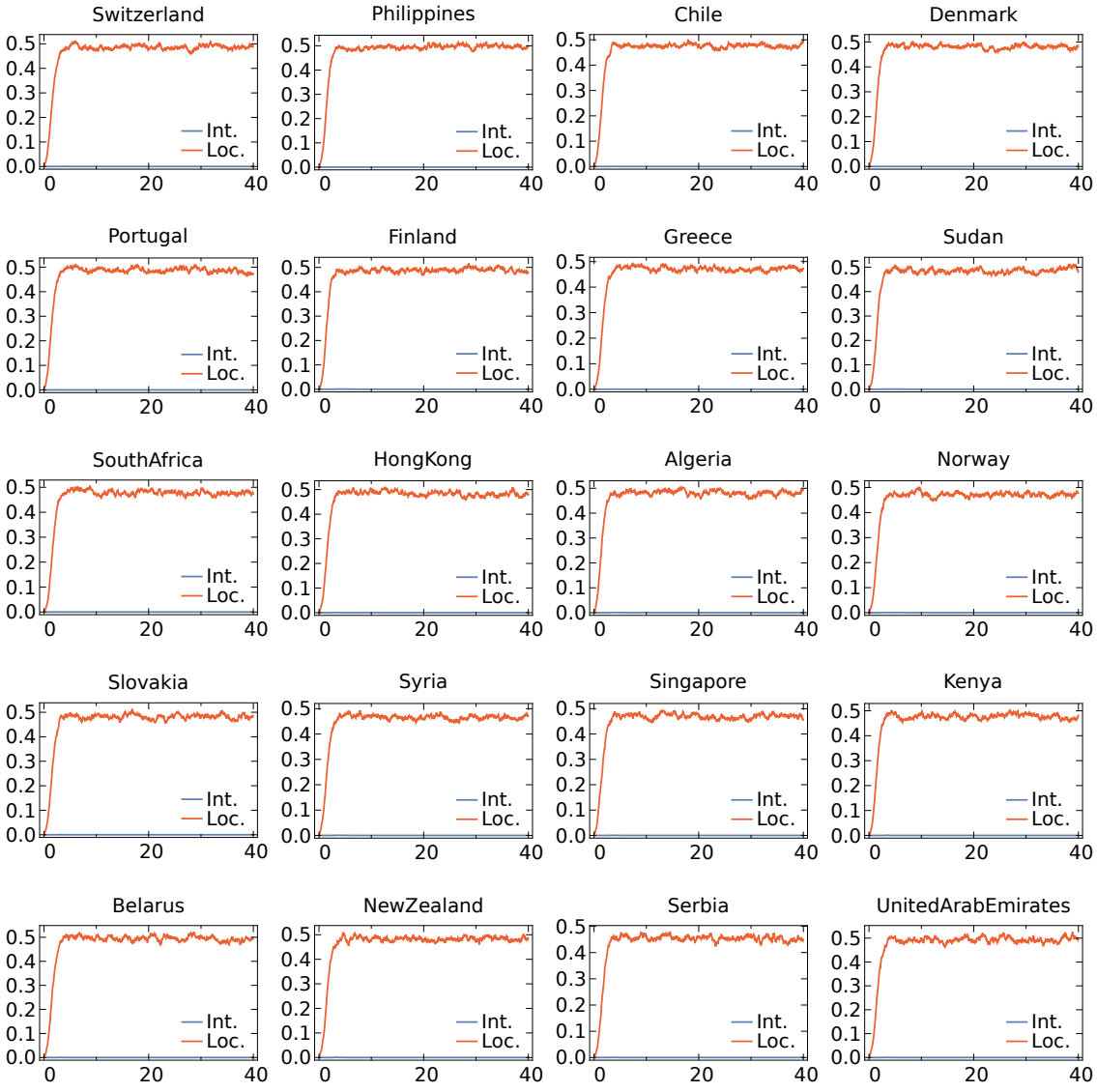

Figure D.17: Evolution of network activity for the last 20 countries (continuation from Fig. D.16).

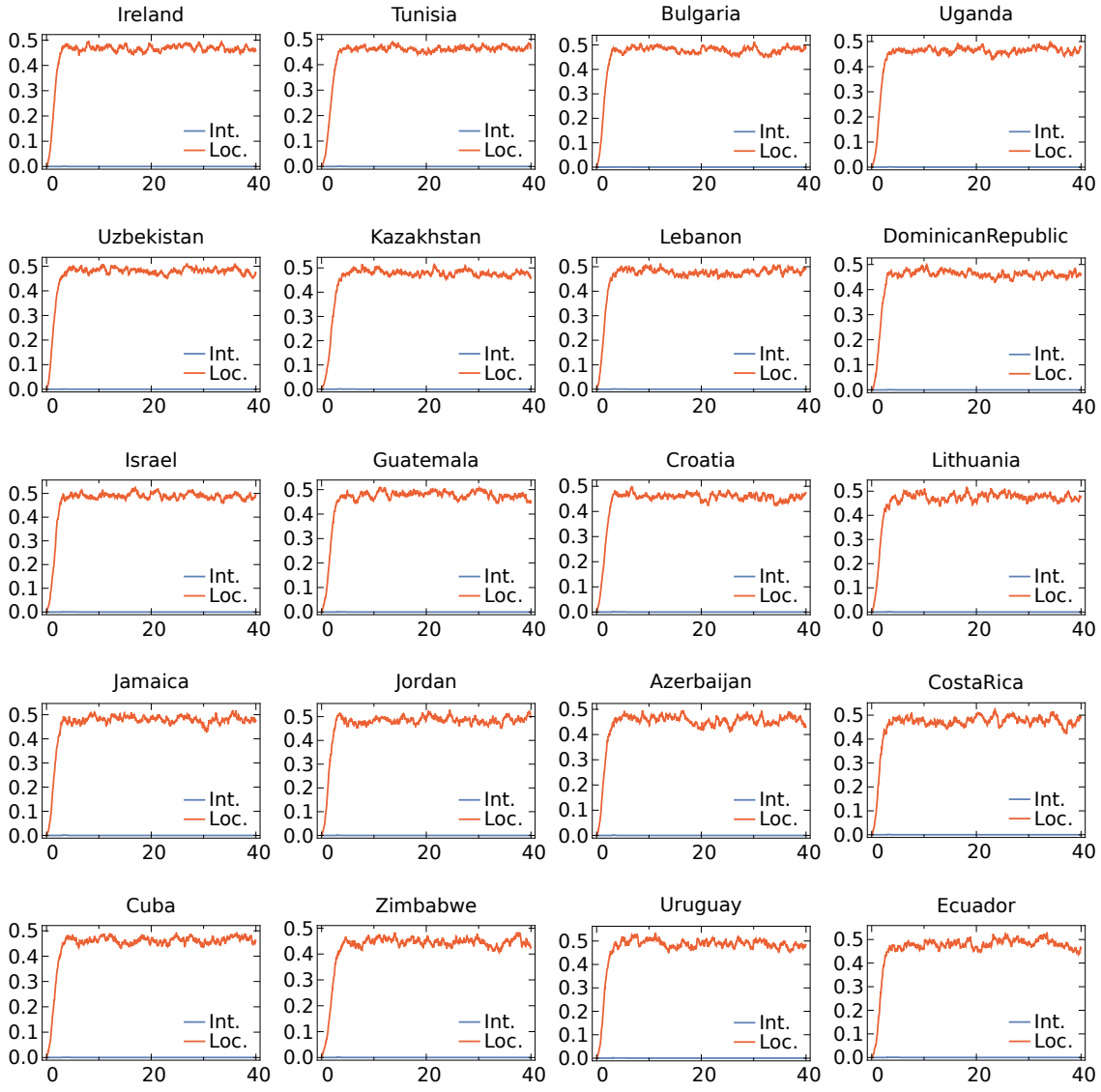

Figure D.18: Evolution of network activity for the last 20 countries (continuation from Fig. D.17).

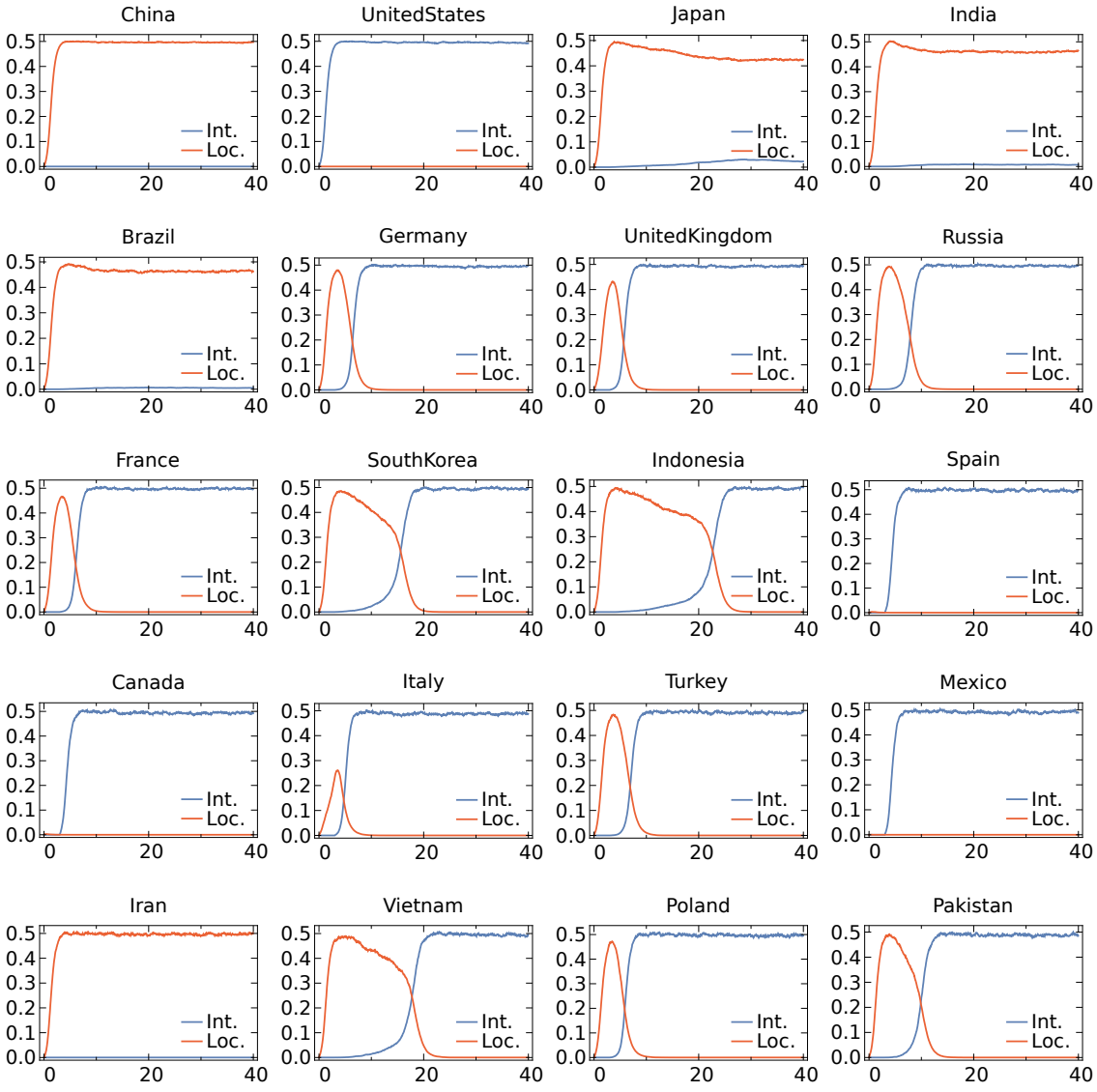

Figure D.19: Evolution of network activity for the first 20 countries. Here, $\sigma = 1.5$, $\Delta t = 3$, $\alpha = 2$, and $\lambda = 0.2$ per country.

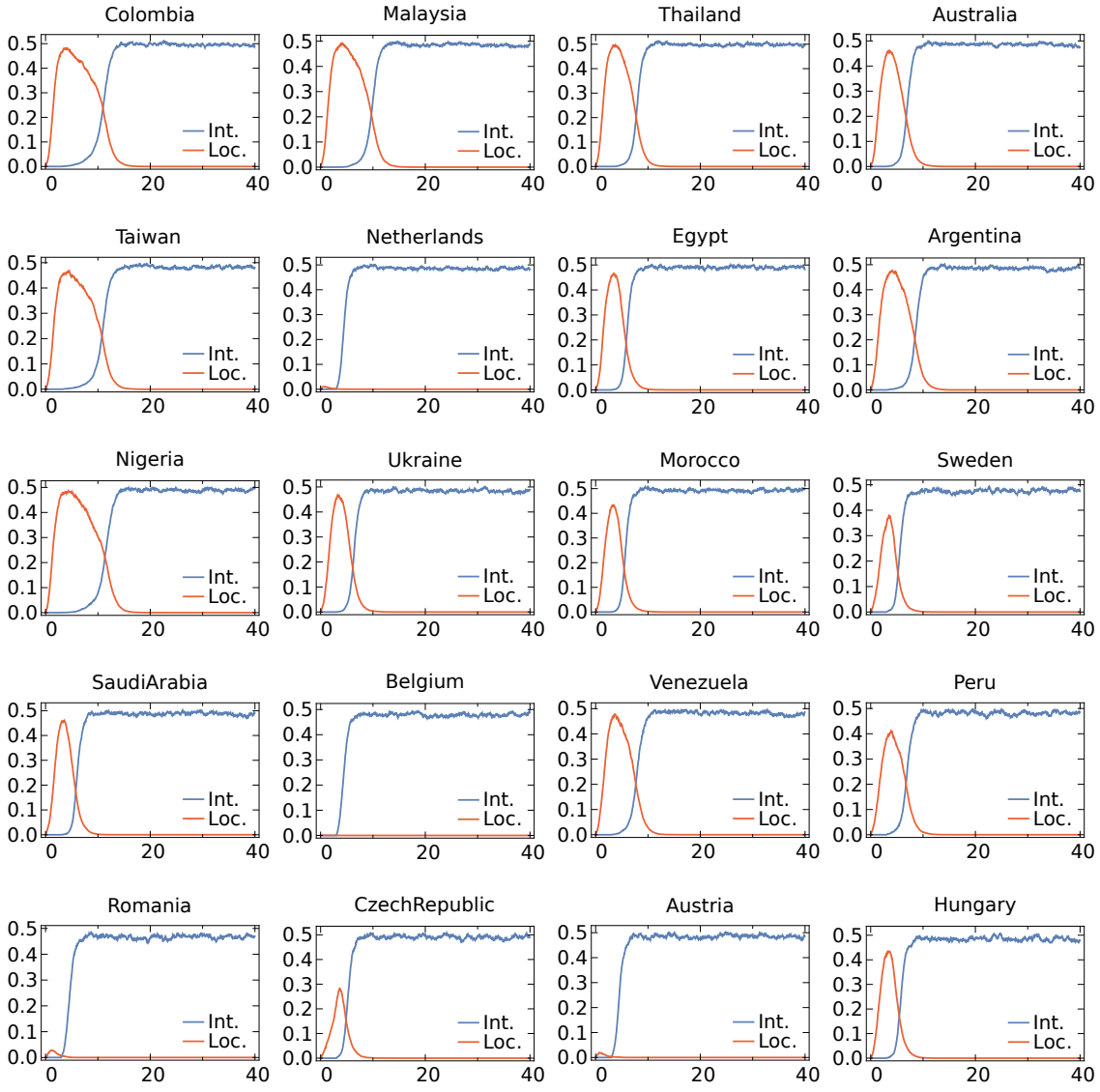

Figure D.20: Evolution of network activity for the second 20 countries (continuation from Fig. D.19).

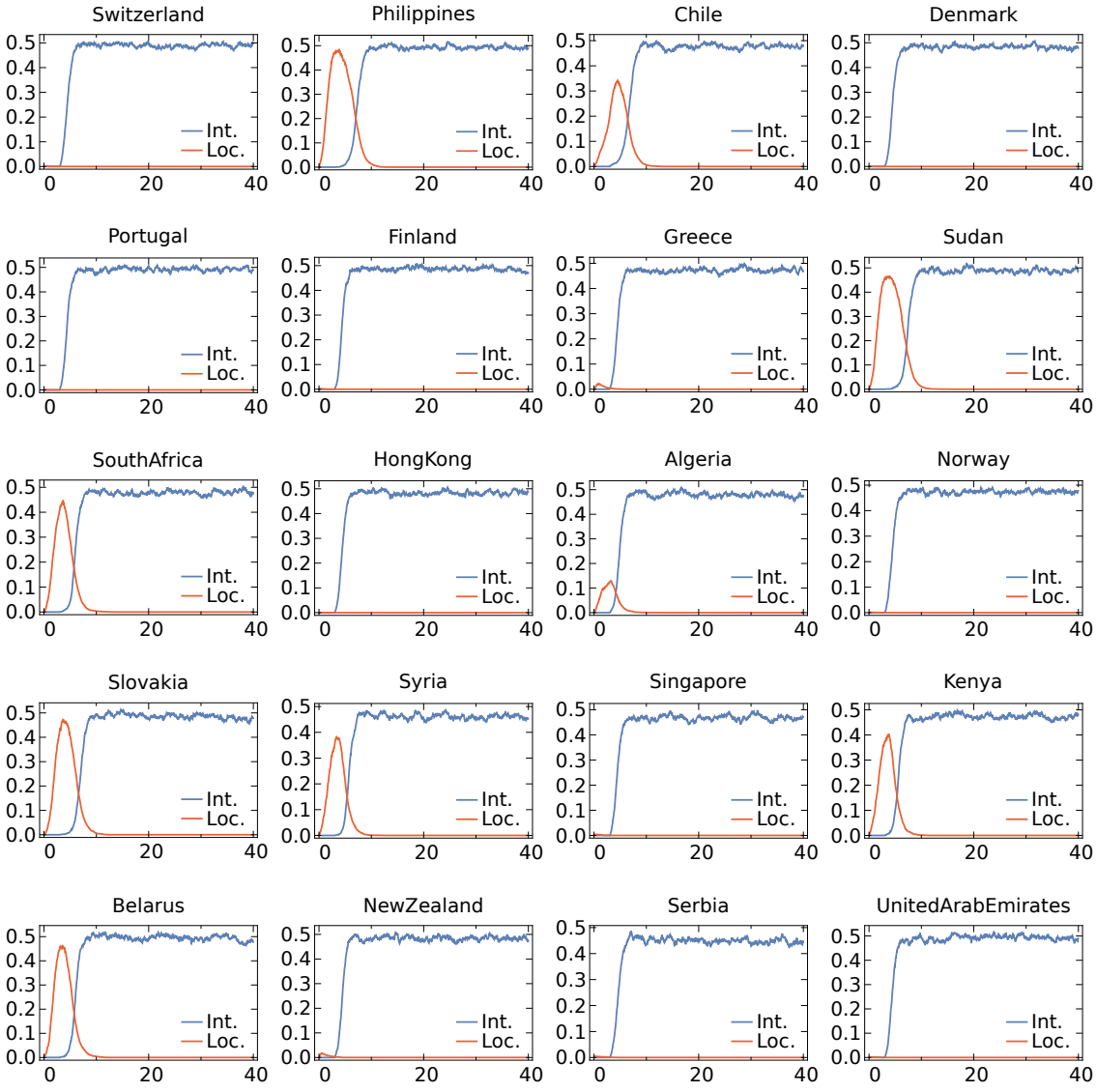

Figure D.21: Evolution of network activity for the last 20 countries (continuation from Fig. D.20).

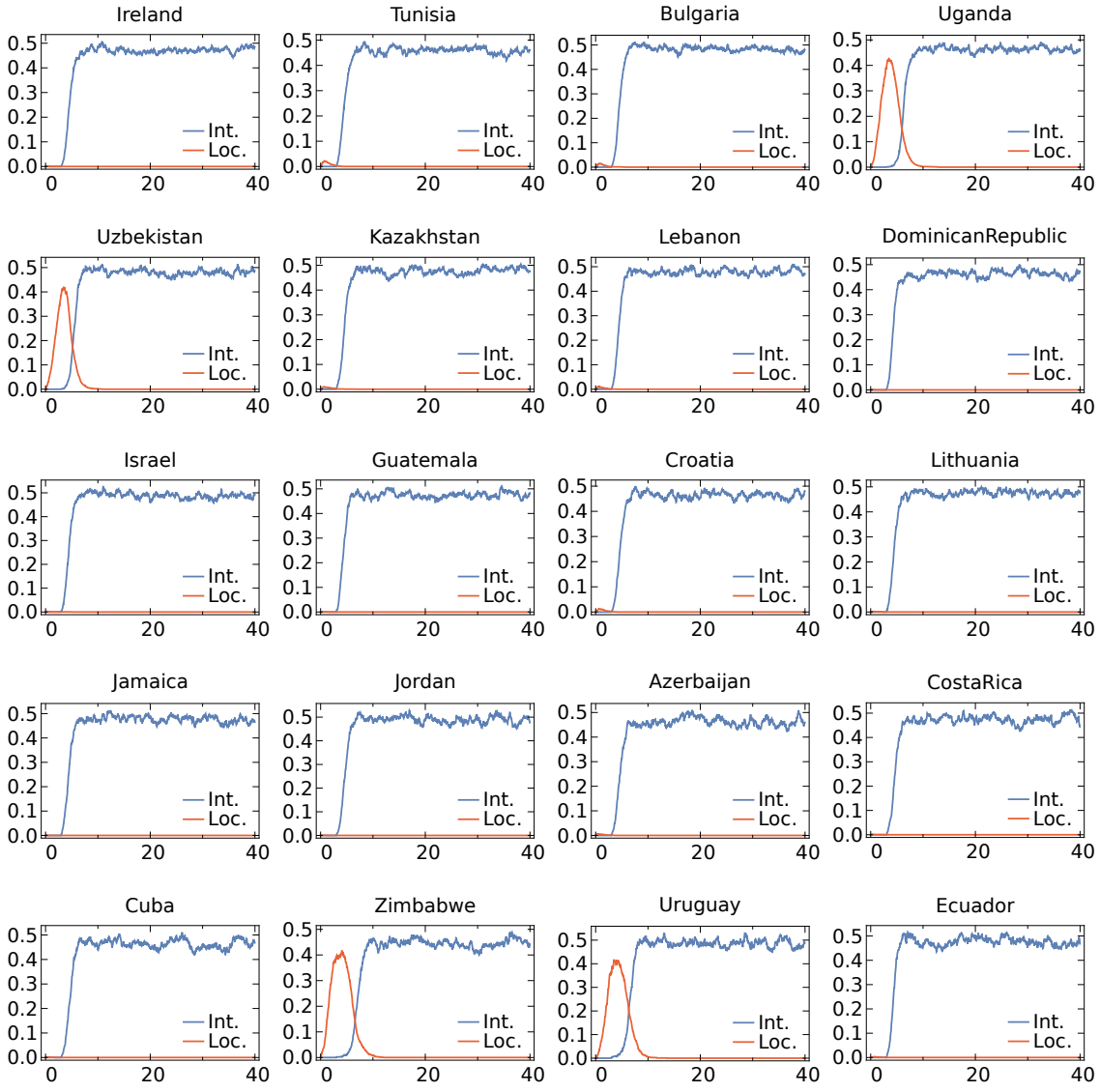

Figure D.22: Evolution of network activity for the last 20 countries (continuation from Fig. D.21).

E Geometry of multiplex networks

This chapter is – with some small changes – available as a preprint at [116] and currently under review.

E.1 Details of datasets

IPv4/IPv6 Internet. The IPv4 and IPv6 Autonomous Systems (AS) Internet topologies were extracted from the data collected by the Archipelago active measurement infrastructure (ARK) developed by CAIDA [162]. The connections in each topology are not physical but logical, representing AS relationships. An AS is a part of the Internet infrastructure administrated by a single company or organization. Pairs of ASs peer to exchange traffic. These peering relationships in the AS topology are represented as links between AS nodes. CAIDA's IPv4 and IPv6 [131] datasets provide regular snapshots of AS links derived from ongoing traceroute-based IP-level topology measurements. The IPv4 dataset consists of ASs that can route Internet packets with IPv4 destination addresses, while the IPv6 dataset consists of ASs that can route packets with IPv6 destination addresses. The considered IPv4 and IPv6 topologies were constructed by merging the AS link snapshots during the first 15 days of January 2015, which are provided at [163]. The IPv4 topology (Layer 1) consists of $N_1 = 37563$ nodes (ASs), and has a power law degree distribution with exponent $\gamma_1 = 2.1$, average node degree $\bar{k}_1 = 5.06$, and average clustering $\bar{c}_1 = 0.63$. The IPv6 topology (Layer 2) consists of $N_2 = 5163$ nodes, has a power law degree distribution with exponent $\gamma_2 = 2.1$, average node degree $\bar{k}_2 = 5.21$, and average clustering $\bar{c}_2 = 0.55$. There are 4819 common nodes in the two topologies, i.e., ASs that can route both IPv4 and IPv6 packets.

Air/Train. The Air/Train data is taken from [121]. The data contains the network of airports and the network of train stations in India, as well as the geographic distances between the airports and the train stations. For each airport, we aggregate all train stations that are within 50km from the airport into a supernode. Subsequently, we declare two supernodes connected if they have at least one train station in common, or if at least one train station from the one supernode is directly connected to a train station from the other supernode. If there are no train stations within 50km from an airport, we consider the nearest train station to the airport, which is considered a supernode on its own. Each supernode has the same id as its corresponding airport, i.e., it is considered to be the same node in the multiplex system. The idea behind this aggregation procedure is to relate train stations to the airports to which they are geographically close. The considered multiplex consists of the network of airports (Air) and the network of aggregated supernodes of train stations (Train). The two networks consist of $N_1 = N_2 = 69$ common nodes. The Air

network (Layer 1) has average degree $\bar{k}_1 = 5.22$, maximum degree $k_1^{\max} = 42$, and average clustering $\bar{c}_1 = 0.79$. The Train network (Layer 2) has average degree $\bar{k}_2 = 9.33$, maximum degree $k_2^{\max} = 41$, and average clustering $\bar{c}_2 = 0.48$.

Drosophila Melanogaster. The Drosophila Melanogaster dataset is taken from [132, 133]. In this dataset, the networks represent protein–protein interactions and the layers correspond to interactions of different nature. Layer 1 in our multiplex corresponds to suppressive genetic interaction, while layer 2 corresponds to additive genetic interaction. More details on the data can be found in [132, 133]. Layer 1 has $N_1 = 838$ nodes, average degree $\bar{k}_1 = 4.43$, and average clustering $\bar{c}_1 = 0.28$. Its degree distribution can be approximated by a power law with exponent $\gamma_1 = 2.6$. Layer 2 has $N_2 = 755$ nodes, average degree $\bar{k}_2 = 3.77$, and average clustering $\bar{c}_2 = 0.29$. Its degree distribution can be approximated by a power law with exponent $\gamma_2 = 2.8$. There are 557 common nodes in the two layers.

C. Elegans Connectomme. The C. Elegans dataset is taken from [134, 135]. It corresponds to the neuronal network of the nematode Caenorhabditis Elegans. The nodes are neurons and each layer corresponds to a different type of synaptic connection: Electric (Layer 1) and Chemical Monadic (Layer 2). Layer 1 has $N_1 = 253$ nodes, average degree $\bar{k}_1 = 4.06$, and average clustering $\bar{c}_1 = 0.24$. Layer 2 has $N_2 = 260$ nodes, average degree $\bar{k}_2 = 6.83$, and average clustering $\bar{c}_2 = 0.21$. The degree distribution in both layers can be approximated by a power law with exponent $\gamma_1 = \gamma_2 = 2.9$, and the two layers have 238 common nodes.

Human Brain. The human brain data is taken from [123]. The data consists of a structural (anatomical) network, as well as a functional network obtained by an algebraic aggregation procedure. In both networks, nodes are brain regions—there are 90 different brain regions in the data. The structural network is obtained by Diffusion Magnetic Resonance Imaging (dMRI). For each pair of brain regions, the data gives the probability that these regions are connected. The connection probability is proportional to the density of the axonal fibers between the regions. In our multiplex, we declare two regions of the structural network connected if their connection probability is larger than a threshold $th_s = 0.92$. The functional network is obtained by BOLD fMRI resting state recordings for the same brain regions. The probability that two regions are connected here is proportional to a correlation coefficient between the fMRI time series of the region voxels [123]. In our multiplex, we declare two regions of the functional network connected if their correlation coefficient is larger than the threshold $th_f = 0.67$. The resulting structural network (Layer 1) consists of a giant connected component of 85 nodes, with average degree $\bar{k}_1 = 5.41$, maximum degree $k_1^{\max} = 12$, and average clustering $\bar{c}_1 = 0.49$. The resulting functional network (Layer 2) has a giant connected component of 78 nodes, average degree $\bar{k}_2 = 5.48$, maximum degree $\bar{k}_2^{\max} = 14$, and average clustering $\bar{c}_2 = 0.40$. The two layers have 77 nodes in common.

arXiv. The arXiv data is taken from [136] and contains co-authorship networks from the free scientific repository arXiv. The nodes are authors that are connected if they have co-authored a paper. In arXiv, each paper is assigned to one or more relevant categories. The data considers only papers with the word "networks" in the title or abstract from 13 different categories up to May 2014. Layer 1 (physics.bio-ph) in our multiplex corresponds to the co-authorship network formed by the authors of papers in the "Biological Physics"

category. Layer 2 (cond-mat.dis-nn) corresponds to the co-authorship network formed by the authors of papers in the "Disordered Systems and Neural Networks" category. Layer 1 has $N_1 = 2956$ nodes, average degree $\bar{k}_1 = 4.13$, and average clustering $\bar{c}_1 = 0.83$. Layer 2 has $N_2 = 3506$ nodes, average degree $\bar{k}_2 = 4.19$, and average clustering $\bar{c}_2 = 0.81$. The degree distribution in both layers can be approximated by a power law with exponent $\gamma_1 = \gamma_2 = 2.6$, and the two layers have 1514 common nodes.

E.2 Stretch

Here, we consider the stretch which is the ratio of the shortest topological path compared to those found by greedy routing or mutual greedy routing. In particular, for each pair of nodes where greedy routing is successful, we evaluate this ratio. In Figs. E.1–E.6 we show both the mean of the stretch and its standard deviation as a function of radial and angular correlations for different temperatures as well as different routing types (hyperbolic/angular). We find that in most cases we have low values of the stretch, i.e. values close to one, and that in general correlations (both angular and radial) reduce the stretch as well as its standard deviation significantly. This means that –in addition to increasing the success rate[1]– geometric correlations increase the efficiency of successful deliveries.

[1]For the success rates see Figs. 6.2–6.4 on pages 95–97.

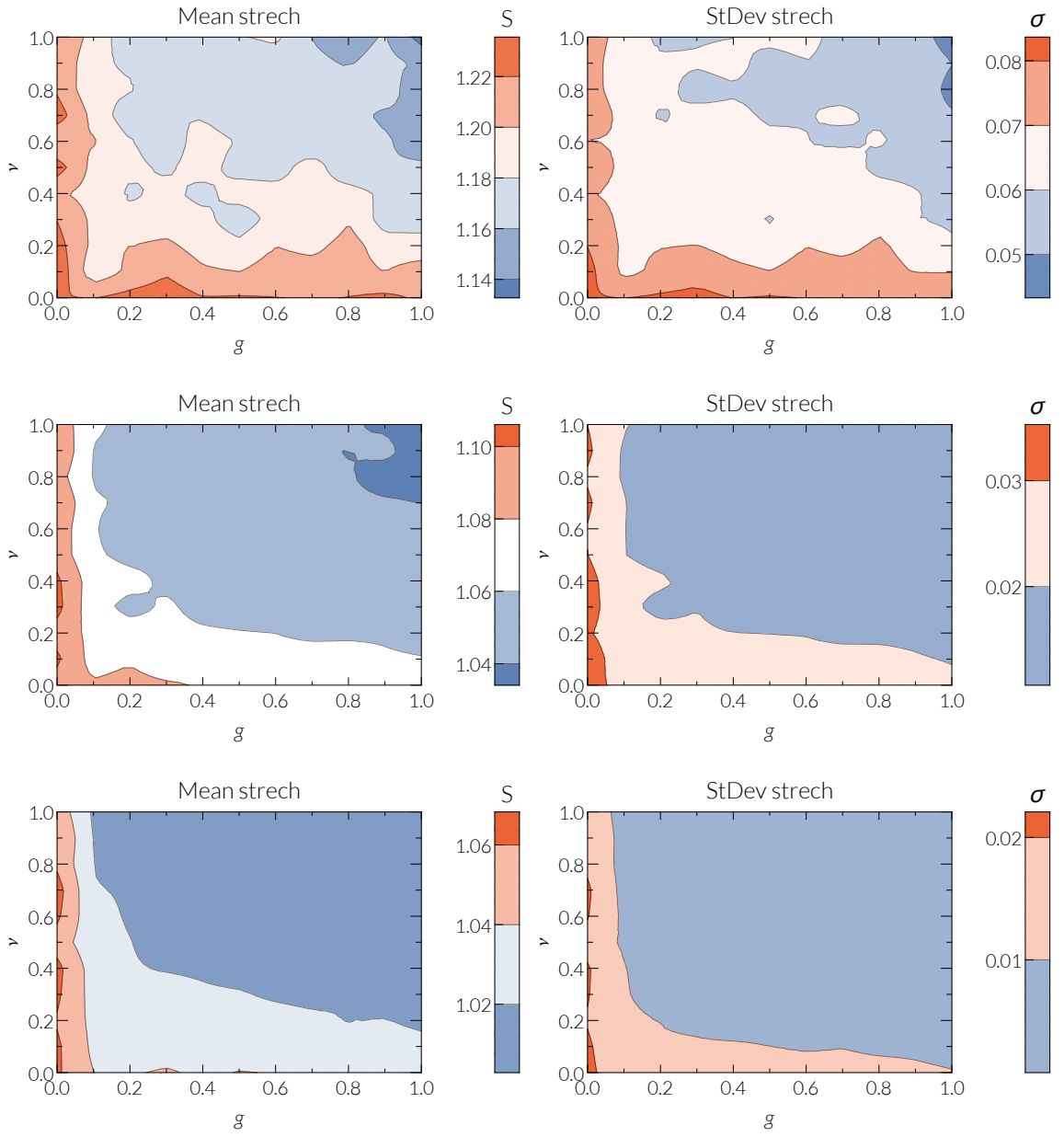

Figure E.1: Mean stretch and standard deviation for hyperbolic routing in two layers. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From top row to bottom row, $T = 0.8, 0.4, 0.1$.

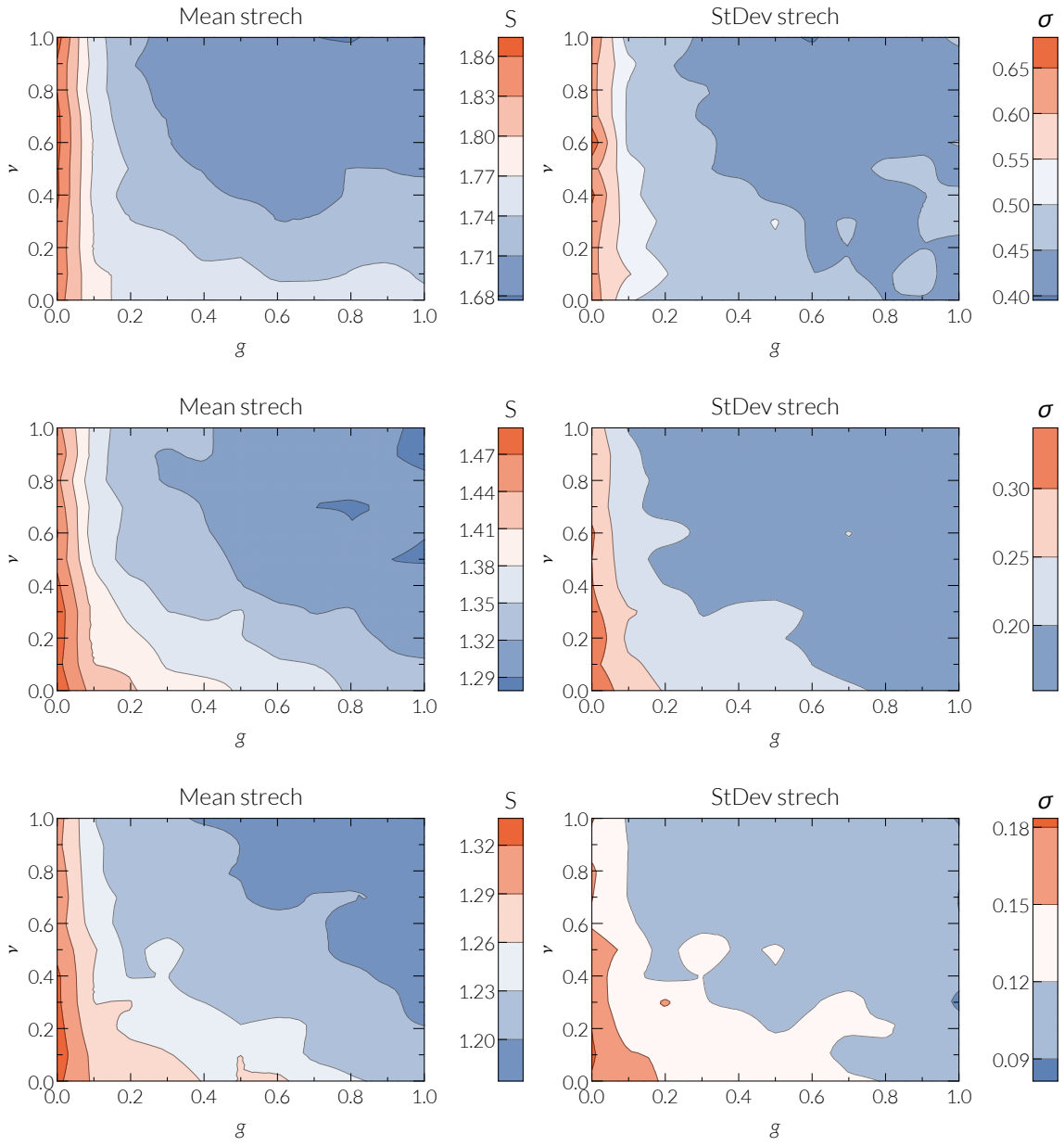

Figure E.2: Mean stretch and standard deviation for angular routing in two layers. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From top row to bottom row, $T = 0.8, 0.4, 0.1$.

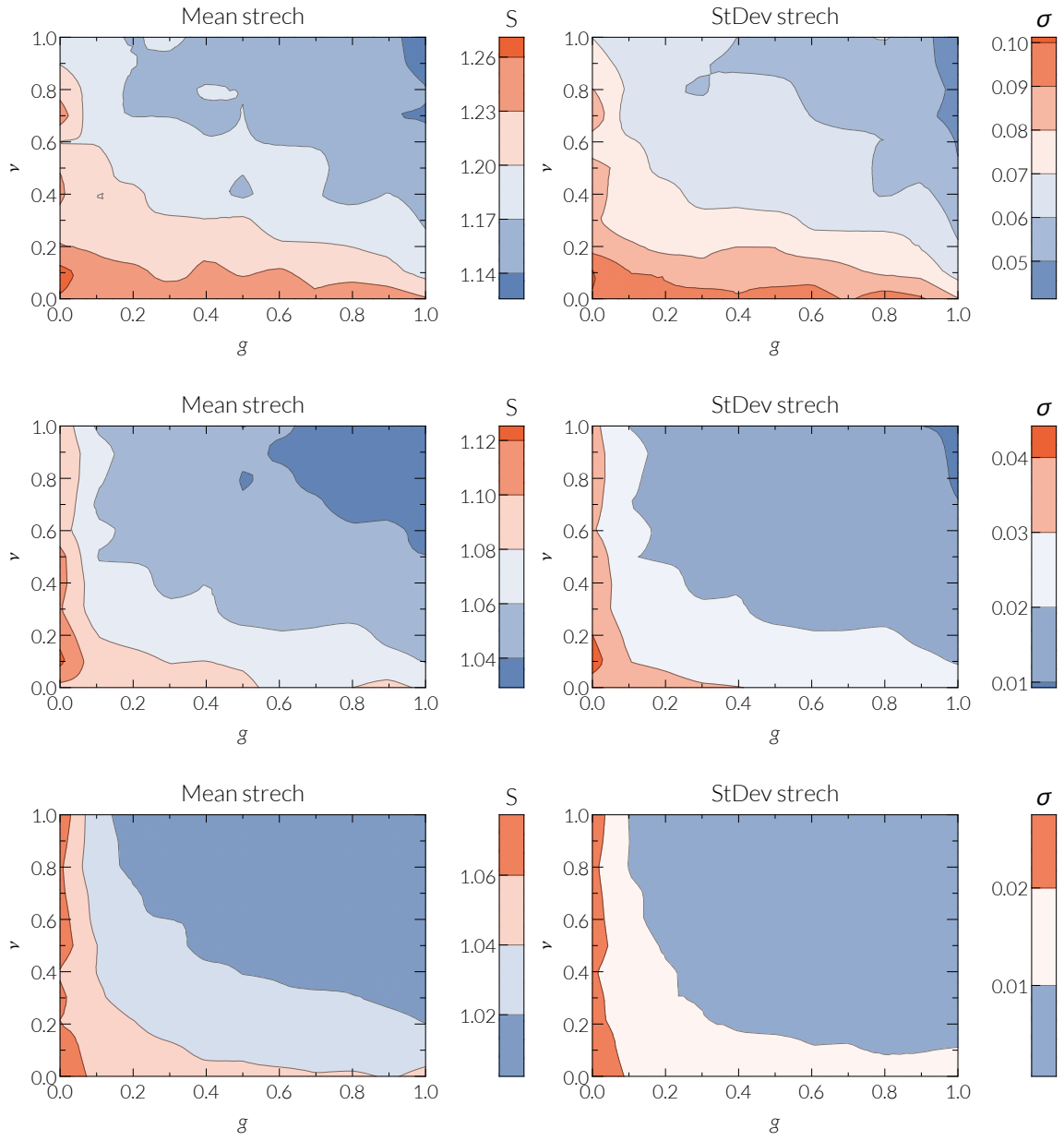

Figure E.3: Mean stretch and standard deviation for hyperbolic routing in three layers. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From top row to bottom row, $T = 0.8, 0.4, 0.1$.

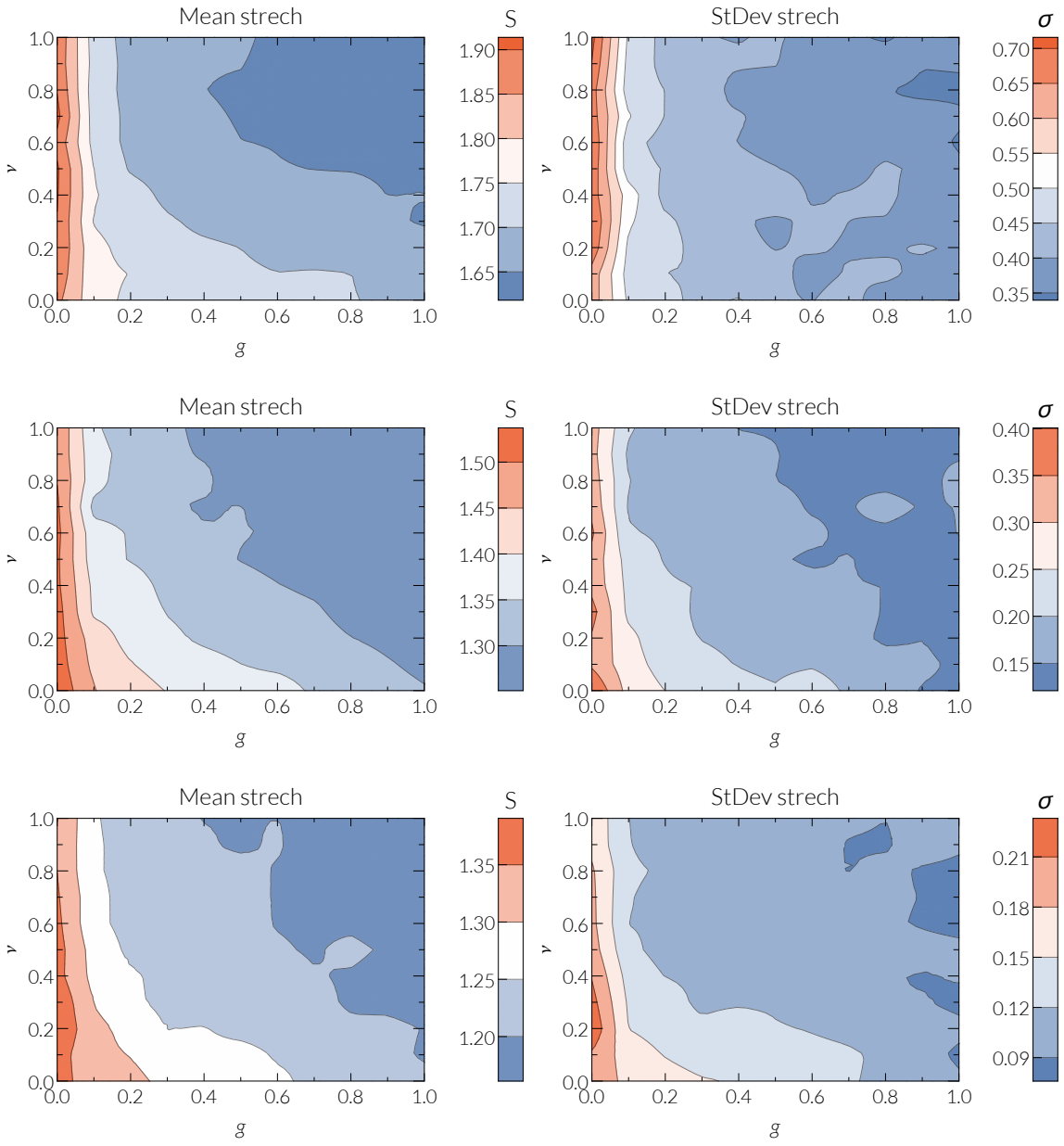

Figure E.4: Mean stretch and standard deviation for angular routing in three layers. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From top row to bottom row, $T = 0.8, 0.4, 0.1$.

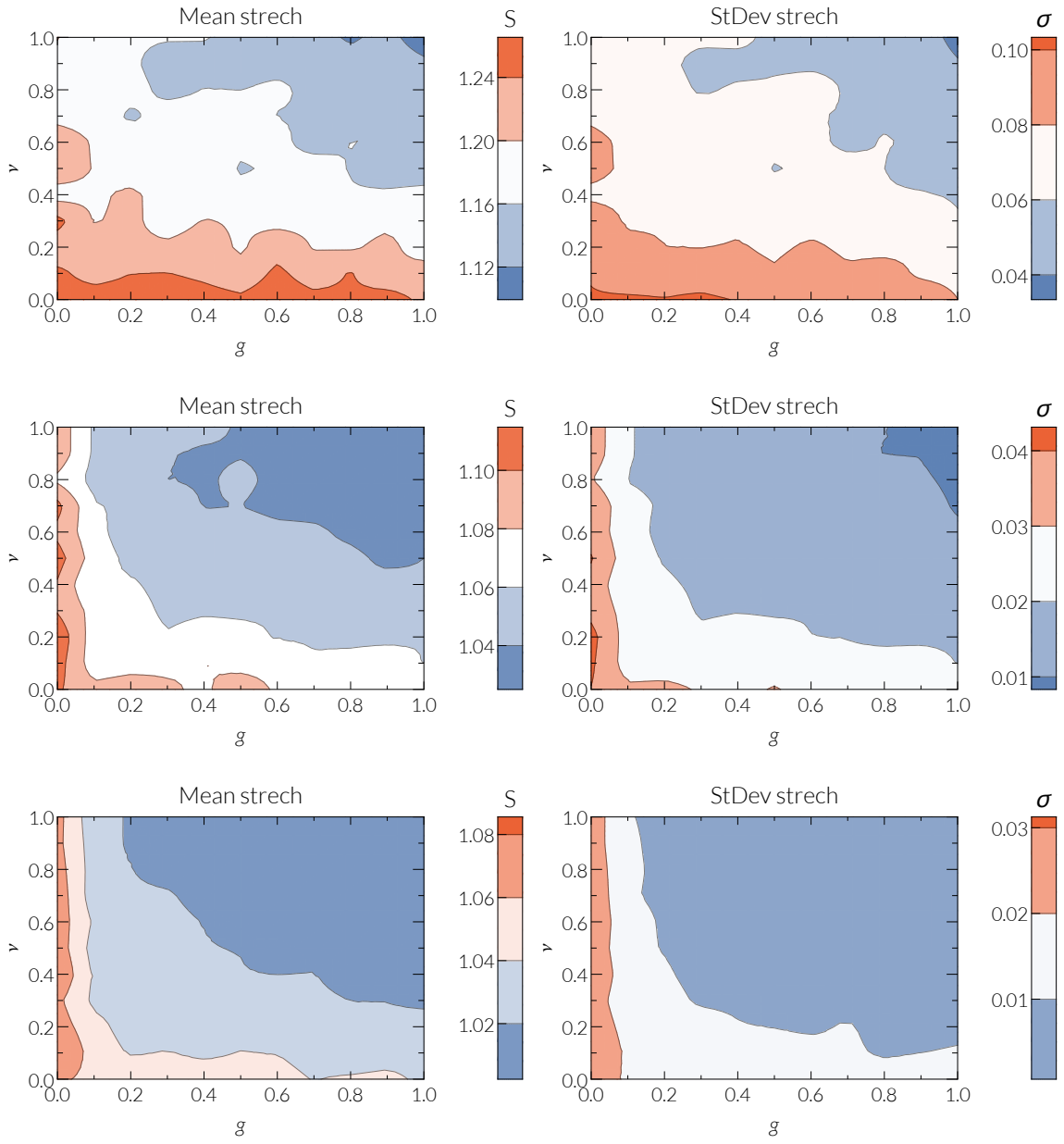

Figure E.5: Mean stretch and standard deviation for hyperbolic routing in four layers. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From top row to bottom row, $T = 0.8, 0.4, 0.1$.

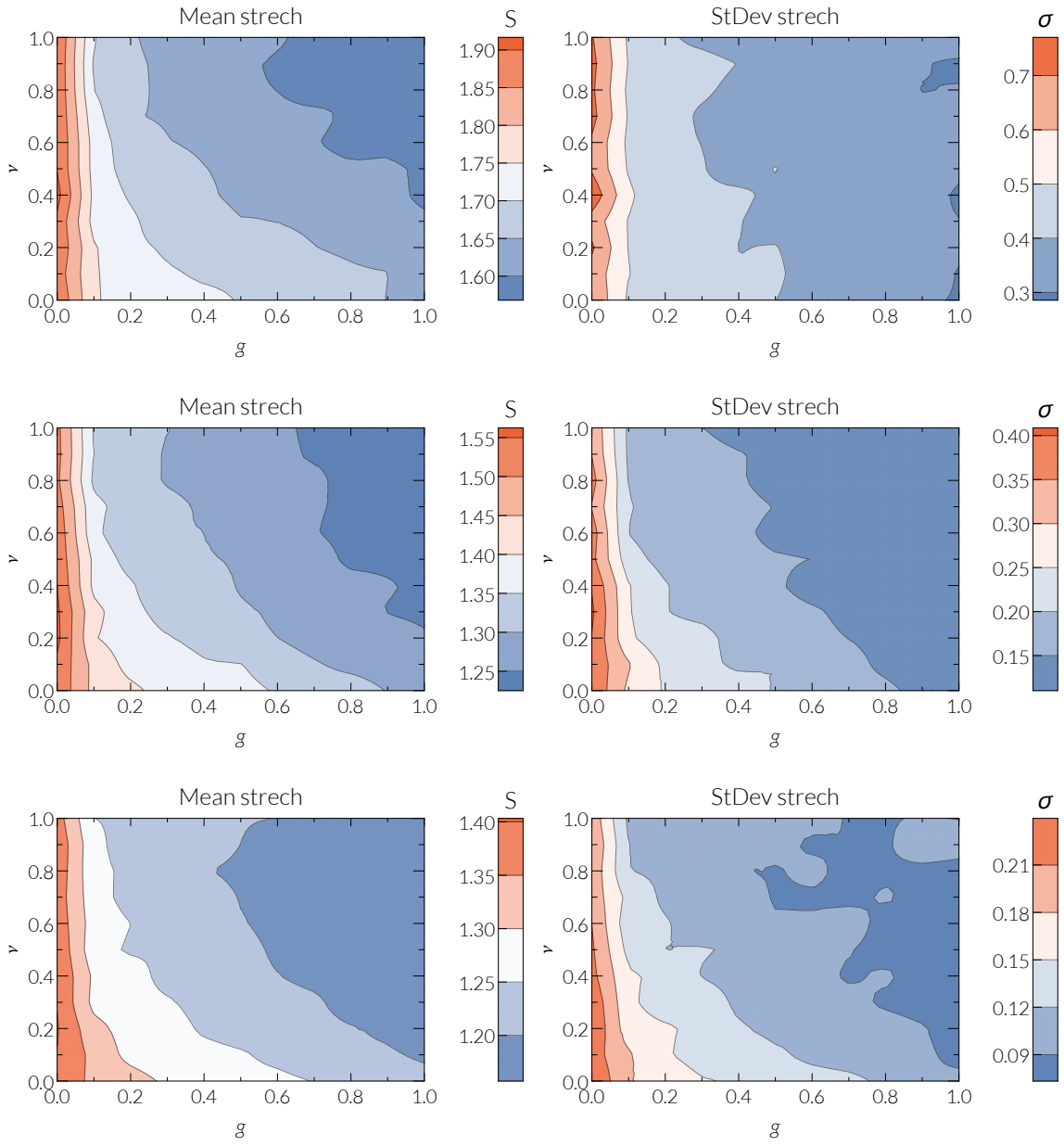

Figure E.6: Mean stretch and standard deviation for angular routing in four layers. Each layer has $N = 30000$ nodes, power law degree distribution $P(k) \sim k^{-2.5}$, $\bar{k} = 10$, and temperature parameter T. From top row to bottom row, $T = 0.8, 0.4, 0.1$.

Bibliography

[1] Ian Stewart and Jack Cohen. *The Collapse of Chaos: Discovering Simplicity in a Complex World*. Penguin Press Science, 1995.

[2] Steven Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Studies in nonlinearity. Westview Press, Cambridge (Mass.), 1994.

[3] John H. Argyris, Gunter Faust, Maria Haase, and Rudolf Friedrich. *Die Erforschung des Chaos: eine Einführung in die Theorie nichtlinearer Systeme*. Springer, Heidelberg, zweite edition, 2010.

[4] Hiroki Sayama. *Introduction to the Modeling and Analysis of Complex Systems*. Open Suny Textbooks, 2015.

[5] Mark Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[6] Alain Barrat, Marc Barthlemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[7] Mark Newman, Albert-Laszlo Barabasi, and Duncan J. Watts. *The Structure and Dynamics of Networks: (Princeton Studies in Complexity)*. Princeton University Press, Princeton, NJ, USA, 2006.

[8] Manuel Castells. *The Rise of the Network Society*. Blackwell Publishers, Inc., Cambridge, MA, USA, 2000.

[9] Duncan J. Watts. *Everything Is Obvious: How Common Sense Fails Us*. Crown Business, 2012.

[10] César Hidalgo. *Why information grows: the evolution of order, from atoms to economies*. Basic Books, 2015.

[11] Dirk Helbing. *Thinking Ahead - Essays on Big Data, Digital Revolution, and Participatory Market Society*. Springer, 2015.

[12] iSocial: Decentralized Online Social Networks project. http://isocial-itn.eu/. 2013-2016.

[13] https://en.wikipedia.org/wiki/internet. 2015.

[14] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, 29(4):251–262, 1999.

[15] Guidelines for creation, selection, and registration of an autonomous system (as). *https://tools.ietf.org/html/rfc1930*, Date of access: 10.12.2015.

[16] Wikipedia web 2.0. *https://en.wikipedia.org/wiki/Web_2.0*, Date of access: 10.12.2015.

[17] Darcy DiNucci. Fragmented future. *www.darcyd.com/fragmented_future.pdf*, 1999.

[18] WeAreSocial. http://wearesocial.net/tag/sdmw/, Date of access: 31/03/2015.

[19] Joanna Brenner and Aaron Smith. 72% of online adults are social networking site users. *Pew Research Internet Project*, 2013, Date of access: 31/03/2015.

[20] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009.

[21] Pierre Levy. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Helix Books, 1999.

[22] Society: Build digital democracy. *Nature*, 527:33–34, 2015.

[23] Robert Epstein and Ronald E. Robertson. The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33):E4512–E4521, 2015.

[24] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. Viking/Penguin, 2011.

[25] James H Fowler and Nicholas A Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *BMJ*, 337, 2008.

[26] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[27] Scott E. Page. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton Univ. Press, 2008.

[28] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam D I Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–8, 2012.

[29] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. provides a portrait of what bitcoin is and how it would be implemented. *http://www.bitcoin.org/bitcoin.pdf*, 2009.

[30] Jorge L. Contreras and Jerome H. Reichman. Sharing by design: Data and decentralized commons. *Science*, 350(6266):1312–1314, 2015.

[31] Stanley Milgram. The small world problem. *Psychology Today*, 1(1):61–67, 1967.

[32] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. The anatomy of the facebook social graph. *arXiv:1111.4503*, 2011.

[33] Sergey Edunov, Carlos Diuk, Ismail Onur Filiz, Smriti Bhagat, and Moira Burke. Three and a half degrees of separation. *https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/*, 2015.

[34] Herbert A. Simon. The architecture of complexity. In *Proceedings of the American Philosophical Society*, pages 467–482, 1962.

[35] Philip Warren Anderson. More is different. *Science*, 177, 1972.

[36] http://global.britannica.com/science/complexity-scientific-theory. 2015.

[37] Sergey V. Buldyrev, Roni Parshani, Gerald Paul, H. Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

[38] *Networks of Networks: The Last Frontier of Complexity*, volume XII of *Understanding Complex Systems*. Springer, 2014.

[39] Mark S. Granovetter. *Getting a job : a study of contacts and careers*. University of Chicago Press, 1995.

[40] Kaj-Kolja Kleineberg and Marián Boguñá. Evolution of the digital society reveals balance between viral and mass media influence. *Phys Rev X*, 4:031046, 2014.

[41] Kaj-Kolja Kleineberg and Marián Boguñá. Evolution of the digital society reveals balance between viral and mass media influence. *arXiv*, 1403.1437, 2014.

[42] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1):5:1, 2007.

[43] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 61–70, New York, NY, USA, 2002. ACM.

[44] José Luis Iribarren and Esteban Moro. Branching dynamics of viral information spreading. *Phys Rev E*, 84:046116, 2011.

[45] José Luis Iribarren and Esteban Moro. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.*, 103:038702, 2009.

[46] Tuan-Anh Hoang and Ee-Peng Lim. Virality and susceptibility in information diffusions. The AAAI Press, 2012.

[47] LarsKai Hansen, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni, and Michael Etter. *Good Friends, Bad News - Affect and Virality in Twitter*, volume 185 of *Communications in Computer and Information Science*. Springer Berlin Heidelberg, 2011.

[48] Przemyslaw A. Grabowicz, José J. Ramasco, Esteban Moro, Josep M. Pujol, and Victor M. Eguiluz. Social features of online networks: The strength of intermediary ties in online social media. *PLoS ONE*, 7(1):e29358, 2012.

[49] David G. Rand, Samuel Arbesman, and Nicholas A. Christakis. Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, 108(48):19193–19198, 2011.

[50] Alec Ross and Ben Scott. Social media: Cause, effect and response. *NATO Review Magazine*, 2011.

[51] Andrea Ordanini. Crowd-funding: transforming customers into investors through innovative service platforms. *Journal of service management.*, 2011.

[52] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, pages 29–42, New York, NY, USA, 2007. ACM.

[53] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 835–844, New York, NY, USA, 2007. ACM.

[54] Jeffrey Travers, Stanley Milgram, Jeffrey Travers, and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.

[55] Mark Newman, Albert-Laszlo Barabasi, and Duncan J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[56] R. Albert, H. Jeong, and A.L. Barabasi. The Diameter of the World Wide Web. *Nature*, 401:130–131, 1999.

[57] Jon M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems 14*, pages 431–438. MIT Press, 2002.

[58] M.S. Granovetter. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[59] N. Friedkin. A test of structural features of granovetter's strength of weak ties theory. *Social Networks*, 2(4):411–422, 1980.

[60] Linyuan Lü and Tao Zhou. Role of weak ties in link prediction of complex networks. In *Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information*, CNIKM '09, pages 55–58, New York, NY, USA, 2009. ACM.

[61] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[62] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.

[63] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

[64] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and evolution of online social networks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 611–617. ACM, 2006.

[65] Haibo Hu and Xiaofan Wang. Evolution of a large online social network. *Physics Letters A*, 373(12-13):1105–1110, 2009.

[66] Haibo Hu and Xiao-Fan Wang. Disassortative mixing in online social networks. *EPL (Europhysics Letters)*, 86(1):18003, 2009.

[67] Menghui Li, Shuguang Guan, Xiaofeng Gong, Kun Li, Jinshan Wu, Zengru Di, and Choy-heng Lai. From sparse to dense and from assortative to disassortative in online social networks. *arXiv:1309.7455*, 2013.

[68] Luca Maria Aiello, Alain Barrat, Ciro Cattuto, Giancarlo Ruffo, and Rossano Schifanella. Link Creation and Profile Alignment in the aNobii Social Network. *2010 IEEE Second International Conference on Social Computing*, pages 249–256, 2010.

[69] Gerald F. Frasco, Jie Sun, Hernán D. Rozenfeld, and Daniel ben Avraham. Spatially distributed social complex networks. *Phys. Rev. X*, 4:011008, 2014.

[70] Jameson L. Toole, Meeyoung Cha, and Marta C. González. Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE*, 7(1):e29528, 2012.

[71] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), 2007.

[72] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, pages 177–187, New York, NY, USA, 2005. ACM.

[73] L. Takac and M. Zabovsky. Data analysis in public social networks. *International Scientific Conference and International Workshop Present Day Trends of Innovations*, 2012.

[74] Juan Carlos González-Avella, Mario G. Cosenza, and Maxi San Miguel. A model for cross-cultural reciprocal interactions through mass media. *PLoS ONE*, 7(12):e51035, 2012.

[75] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A. Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Phys Rev X*, 3:041022, 2013.

[76] Albert-László Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[77] S N Dorogovtsev, J. Mendes, and A. Samukhin. Size-dependent degree distribution of a scale-free growing network. *Phys Rev E*, 63(6):062101, 2001.

[78] Ginestra Bianconi and A.-L. Barabási. Bose-Einstein Condensation in Complex Networks. *Phys Rev Lett*, 86(24):5632–5635, 2001.

[79] G Caldarelli, A Capocci, P De Los Rios, and and M. A. Muñoz. Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Phys Rev Lett*, 89:258702, 2002.

[80] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys Rev E*, 67(5):056104, 2003.

[81] R Pastor-Satorras, E Smith, and R V Sole. Evolving protein interaction networks through gene duplication. *J Theor Biol*, 222(2):199–210, 2003.

[82] Santo Fortunato, Alessandro Flammini, and Filippo Menczer. Scale-Free Network Growth by Ranking. *Phys Rev Lett*, 96(21):218701, 2006.

[83] Raissa M D'Souza, Christian Borgs, Jennifer T Chayes, Noam Berger, and Robert D Kleinberg. Emergence of tempered preferential attachment from optimization. *Proc Natl Acad Sci USA*, 104(15):6112–7, 2007.

[84] F. Papadopoulos, M. Kitsak, M. Serrano, M. Boguñá, and D. Krioukov. Popularity versus similarity in growing networks. *Nature*, 489:537–540, 2012.

[85] Kaj-Kolja Kleineberg. https://www.youtube.com/watch?v=58bet5htf-c.

[86] R. M. Anderson and R. M. May. *Infectious diseases in humans*. Oxford University Press, Oxford, 1992.

[87] Bruno Ribeiro. Modeling and predicting the growth and death of membership-based websites. *International World Wide Web Conference*, 2014.

[88] Kaj-Kolja Kleineberg. http://www.slideshare.net/kolja111/ecology-20-coexistence-and-domination-among-interacting-networks. 2015.

[89] Kaj-Kolja Kleineberg and Marián Boguñá. Digital ecology: Coexistence and domination among interacting networks. *Sci. Rep.*, 5:10268, 2015.

[90] Kaj-Kolja Kleineberg and Marián Boguñá. Digital ecology: Coexistence and domination among interacting networks. *arXiv*, 1410.8865, 2015.

[91] G. Hardin. The competitive exclusion principle. *Science*, 131:1292–1297, 1960.

[92] Hermann Haken. *Synergetics, An introduction*. Springer Verlag, 1977.

[93] G. E. Hutchinson. The paradox of the plankton. *The American Naturalist*, 95(882):137–145, 1961.

[94] Alan Hastings. Disturbance, coexistence, history, and competition for space. *Theoretical population biology*, 18:363–373, 1980.

[95] Marc William Cadotte. Competition-colonization trade-offs and disturbance effects at multiple scales. *Ecology*, 84(4):823–829, 2007.

[96] M. W. Palmer. Variation in species richness: towards a unification of hypotheses. *Folia Geobot Phytotaxon*, 29:511–530, 1994.

[97] B. Ribeiro and C. Faloutsos. Modeling Website Popularity Competition in the Attention-Activity Marketplace. *arXiv:1403.0600*, 2014.

[98] Wikipedia. List of social networking websites. 2014.

[99] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2, 2012.

[100] James P. Gleeson, Jonathan A. Ward, Kevin P. O'Sullivan, and William T. Lee. Competition-induced criticality in a model of meme popularity. *Phys. Rev. Lett.*, 112:048701, 2014.

[101] James P. Gleeson, Davide Cellai, Jukka-Pekka Onnela, Mason A. Porter, and Felix Reed-Tsochas. A simple generative model of collective online behavior. *Proceedings of the National Academy of Sciences*, 111(29):10411–10415, 2014.

[102] Kaj-Kolja Kleineberg and Marián Boguñá. Is bigger always better? how local online social networks can outperform global ones. *ArXiv:1504.01368*, 2015.

[103] Saulo D. S. Reis, Yanqing Hu, Andrés Babino, José S. Andrade Jr, Santiago Canals, Mariano Sigman, and Hernán a. Makse. Avoiding catastrophic failure in correlated networks of networks. *Nature Physics*, 2014.

[104] Filippo Radicchi. Driving Interconnected Networks to Supercriticality. *Phys Rev X*, 4(2):021014, 2014.

[105] Ginestra Bianconi and Sergey N. Dorogovtsev. Multiple percolation transitions in a configuration model of a network of networks. *Phys Rev E*, 89(6):062814, 2014.

[106] A. Z. Jacobs, S. F. Way, J. Ugander, and A. Clauset. Assembling thefacebook: Using heterogeneity to understand online social network assembly. *ACM Web Science Conference*, 2015.

[107] Bogdan State, Patrick Park, Ingmar Weber, Yelena Mejova, and Michael W. Macy. The mesh of civilizations and international email flows. *WebSci 2013 Proceedings*, 2013.

[108] Yuri Takhteyev, Anatoliy Gruzd, and Barry Wellman. Geography of twitter networks. *Social Networks*, 34(1):73–81, 2012.

[109] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2014.

[110] M. Serrano, Dmitri Krioukov, and Marián Boguñá. Self-Similarity of Complex Networks and Hidden Metric Spaces. *Phys Rev Lett*, 100(7):078701, 2008.

[111] Marián Boguñá, Dmitri Krioukov, and K. C. Claffy. Navigability of complex networks. *Nature Physics*, 5(1):74–80, 2008.

[112] Marián Boguñá and Romualdo Pastor-Satorras. Class of correlated random networks with hidden variables. *Phys Rev E*, 68(3):036112, 2003.

[113] Kaj-Kolja Kleineberg. 1:000 scale model of the digital world: Extinction of local networks. www.youtube.com/watch?v=z3dP3PD7ueA, Date of access: 31/03/2015.

[114] Kaj-Kolja Kleineberg. 1:000 scale model of the digital world: Partial persistence of local networks. www.youtube.com/watch?v=XkZTxnJd-eI, Date of access: 31/03/2015.

[115] Vincenzo Cosenza. World map of social networks. http://vincos.it/world-map-of-social-networks, Date of access: 31/03/2015.

[116] Kaj-Kolja Kleineberg, Marián Boguñá, M. Ángeles Serrano, and Fragkiskos Papadopoulos. Geometric correlations in real multiplex networks: multidimensional communities, trans-layer link prediction, and efficient navigation. *ArXiv: 1601.04071*.

[117] Filippo Radicchi and Alex Arenas. Abrupt transition in the structural formation of interconnected networks. *Nature Physics*, 9(11):717–720, 2013.

[118] Ginestra Bianconi. Multilayer networks: Dangerous liaisons? *Nature Physics*, 10:712–714, 2014.

[119] Filippo Radicchi. Percolation in real interdependent networks. *Nature Physics*, 11:597–602, 2015.

[120] Giulia Menichetti, Daniel Remondini, and Ginestra Bianconi. Correlations between weights and overlap in ensembles of weighted multiplex networks. *Phys. Rev. E*, 90:062817, 2014.

[121] Arda Halu, Satyam Mukherjee, and Ginestra Bianconi. Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles. *Physical Review E*, 89, 2014.

[122] Manlio De Domenico, Albert Solé-Ribalta, Sergio Gómez, and Alex Arenas. Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences*, 111(23):8351–8356, 2014.

[123] Tiago Simas, Mario Chavez, Pablo R. Rodriguez, and Albert Diaz-Guilera. An algebraic topological method for multimodal brain networks comparison. *Frontiers in Psychology*, 6:904, 2015.

[124] S. Boccaletti, G. Bianconi, R. Criado, C.I. 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):1 – 122, 2014.

[125] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

[126] Marián Boguñá, Fragkiskos Papadopoulos, and Dmitri Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature communications*, 1:62, 2010.

[127] F. Papadopoulos, C. Psomas, and D. Krioukov. Network mapping by replaying hyperbolic growth. *Networking, IEEE/ACM Transactions on*, 23(1):198–211, 2015.

[128] Fragkiskos Papadopoulos, Rodrigo Aldecoa, and Dmitri Krioukov. Network geometry inference using common neighbors. *Phys. Rev. E*, 92:022807, 2015.

[129] Dmitri Krioukov, Fragkiskos Papadopoulos, Amin Vahdat, and Marián Boguñá. Curvature and temperature of complex networks. *Physical Review E*, 80(3):035101, 2009.

[130] M. Ángeles Serrano, Marián Boguñá, and Francesc Sagues. Uncovering the hidden geometry behind metabolic networks. *Mol. BioSyst.*, 8:843–850, 2012.

[131] The IPv4 and IPv6 Topology Datasets, 2015. `http://www.caida.org/data/active/ipv4_routed_topology_aslinks_dataset.xml` and `https://www.caida.org/data/active/ipv6_allpref_topology_dataset.xml`.

[132] Biogrid: a general repository for interaction datasets. *Nucleic Acids Research*, 34:D535–D539, 2006.

[133] Manlio De Domenico, Vincenzo Nicosia, Alex Arenas, and Vito Latora. Structural reducibility of multilayer networks. *Nature Communications*, 6:6864, 2015.

[134] Beth L. Chen, David H. Hall, and Dmitri B. Chklovskii. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4723–4728, 2006.

[135] Manlio De Domenico, Mason A. Porter, and Alex Arenas. Muxviz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*, 3(2):159–176, 2015.

[136] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X*, 5:011027, 2015.

[137] HyperMap-CN Software Package. `https://bitbucket.org/dk-lab/2015_code_hypermap`.

[138] Vincenzo Nicosia and Vito Latora. Measuring and modeling correlations in multiplex networks. *Phys. Rev. E*, 92:032805, 2015.

[139] Byungjoon Min, Su Do Yi, Kyu-Min Lee, and K.-I. Goh. Network robustness of multiplex networks with interlayer degree correlations. *Phys. Rev. E*, 89:042811, 2014.

[140] Jung Yeol Kim and K.-I. Goh. Coevolution and correlated multiplexity in multiplex networks. *Phys. Rev. Lett.*, 111:058702, 2013.

[141] V. Gemmetto and D. Garlaschelli. Multiplexity versus correlation: the role of local constraints in real multiplexes. *Scientific Reports*, 5:9120, 2015.

[142] M Ángeles Serrano, Ľuboš Buzna, and Marián Boguñá. Escaping the avalanche collapse in self-similar multiplexes. *New Journal of Physics*, 17(5):053033, 2015.

[143] Chuan Wen Loe and Henrik Jeldtoft Jensen. Comparison of communities detection algorithms for multiplex. *Physica A: Statistical Mechanics and its Applications*, 431:29 – 45, 2015.

[144] Guangyao Zhu and Kan Li. A unified model for community detection of multiplex networks. In *Web Information Systems Engineering – WISE 2014*, volume 8786 of *Lecture Notes in Computer Science*, pages 31–46. Springer International Publishing, 2014.

[145] The CAIDA AS Organizations Dataset, 09-Jan-2015, 2015. `http://www.caida.org/data/as-organizations`.

[146] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150 – 1170, 2011.

[147] Aaron Clauset, Cristopher Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453:98–101, 2008.

[148] Desislava Hristova, Anastasios Noulas, Chloë Brown, Mirco Musolesi, and Cecilia Mascolo. A multilayer approach to multiplexity and link prediction in online geo-social networks. *arXiv:1508.07876*, 2015.

[149] Fragkiskos Papadopoulos, Dmitri Krioukov, Marian Boguna, and Amin Vahdat. Greedy Forwarding in Dynamic Scale-Free Networks Embedded in Hyperbolic Metric Spaces. *Proceedings of IEEE INFOCOM*, pages 1–9, 2010.

[150] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[151] S N Dorogovtsev. *Lectures on Complex Networks*. Oxford University Press, Oxford, 2010.

[152] Roger B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer-Verlag New York, 2006.

[153] Federico Battiston, Vincenzo Nicosia, and Vito Latora. Structural measures for multiplex networks. *Phys. Rev. E*, 89:032804, 2014.

[154] Elisa Omodei, Manlio De De Domenico, and Alex Arenas. Characterizing interactions in online social networks during exceptional events. *Frontiers in Physics*, 3(59), 2015.

[155] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403 – 434, 1976.

[156] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[157] J. L. Doob. Markoff chains–denumerable case. *Transactions of the American Mathematical Society*, 58(3):pp. 455–473, 1945.

[158] P Sagaut. *Large Eddy Simulation for Incompressible Flows*. Springer, 2006.

[159] S. Pope. *Turbulent Flows*. Cambridge University Press, 2000.

[160] Kaj-Kolja Kleineberg. https://www.youtube.com/watch?v=z3dp3pd7uea.

[161] Kaj-Kolja Kleineberg. https://www.youtube.com/watch?v=xkztxnjd-ei.

[162] K. Claffy, Young Hyun, K. Keys, M. Fomenkov, and D. Krioukov. Internet mapping: From art to science. In *Conference For Homeland Security, 2009. CATCH '09. Cybersecurity Applications Technology*, pages 205–211, 2009.

[163] IPv4 and IPv6 Topology Data, January 2015. `http://data.caida.org/datasets/topology/ark/ipv6/as-links/2015/01/` and `http://data.caida.org/datasets/topology/ark/ipv4/as-links/`.