Network geometry and the SIS epidemic model

Author: Àlex Garcia Herranz
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: Marián Boguñá

Abstract: We give a short introduction to network geometry and its relation to the SIS epidemic model. Approximations and simulations are provided for four synthetic networks and two real-world networks. We compare the quality of the different mean-field approximations to the simulations. Finally, we study the effect of the nodes's angular distribution in the hidden metric space on the prevalence of the epidemic.

I. INTRODUCTION

Historically, complex networks have been studied as graphs, even though they are not as random or regular. Some of the most interesting properties are small-world effect, connecting every pair of nodes in a few steps; scale-free of the distribution of node degrees, typically a power law of the form $P(k) \sim k^{-\gamma}$ with $\gamma \in [2, 3]$, heterogeneity and the presence of a lot of triangles, that is, they are clustered.

One may think that, because every node is connected to all nodes by a small number of intermediate edges, there is no metric structure on the system. Nevertheless, some networks are embedded in metric spaces, like airport networks, neural networks, trading routes... Moreover, clustering, in other words, the number of triangles, can be thought as a consequence of the triangular inequality. With this in mind, in [1] a model was developed which assumed that networks reside in a *hidden metric space*, meaning that the probability of connection between two nodes depends on their distance in this space, which, in turn, depends on the popularity and similarity of the nodes.

Networks are a key ingredient of epidemic modelling. Epidemic models assume that the population can be split in different categories depending on the disease's phase. One of the simplest models is the SIS or susceptible-infectious-susceptible model [2]. In this work, we will focus on how this model works and how is related to the network's hidden metric space geometry.

This thesis is organized as follows. Firstly, in section II we discuss the geometric model of networks. Secondly, in section III, we present the SIS epidemic model and some stationary state mean-field approximations. Then, in section IV we run some simulations on some synthetic and real-world networks and compare the results with the approximations presented before, and in section V we try to determine and explain if the angular distribution of the nodes affects the disease prevalence. Finally, in section VI we present some conclusions and we discuss possible future work.

*Electronic address: agarcihe34@alumnes.ub.edu

II. GEOMETRIC MODEL

As said before, complex networks are heterogeneous, nodes can be split in groups of smaller groups and so on. It can be approximated as a treelike structure. Thus, the hidden metric space is a *hyperbolic* space [3].

There are two ways to present this hyperbolic space: as the hyperbolic plane $\mathbb{H}^2$, typically the Poincaré disk representation, or as its quasi-isomorphic version, the $\mathbb{S}^1$ model.

A. The $\mathbb{S}^1$ model

In the $\mathbb{S}^1$ model, the hidden metric space is a circle of radius $R = N/2\pi$ where N is the number of nodes. Every node i is defined by two variables: a *hidden degree*, related to the popularity, κ_i and an *angular position* θ_i. The probability of connection between nodes has to decrease with the angular distance and increase with the product of hidden degrees. That is, similar nodes are angularly closer and, in consequence, probably connected. Two not-so-similar nodes have a low probability of being connected unless they are popular. In [4] it is proved that the only choice for the connection probability between nodes i and j that creates maximally random, clustered, small-world and heterogeneous networks is the Fermi-Dirac form

$$p_{ij} = \frac{1}{1 + \left(\frac{d_{ij}}{\mu\kappa_i\kappa_j}\right)^{\beta}} \tag{1}$$

where β controls the level of clustering of the network, d_{ij} is the angular distance between nodes i and j and μ controls the average degree. For $\beta < 1$, networks are unclustered and for $\beta > 1$, we obtain networks with finite clustering. Thus, there is a structural phase transition for $\beta = 1$.

It is not complicated to generate graphs from this model, with angular positions homogeneously distributed and hidden degrees uncorrelated. We fix the number of nodes N, the target average degree $\langle k \rangle$ and $\beta > 1$. The

constant μ is defined by

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

Then, we assign to each node i an angular position θ_i from a $[0,1]-$uniform distribution and a hidden degree κ_i from a distribution $\rho(\kappa)$ so that $\langle\kappa\rangle = \langle k\rangle$. Finally, we connect every pair of nodes with probability like Eq. (1). The values of hidden degrees κ_i can be sampled from any distribution, however, in many real-world networks, the observed distribution is a Pareto distribution $\rho(\kappa) = (\gamma-1)\kappa_0^{\gamma-1}\kappa^{-\gamma}, \gamma > 2$ which results in a power-law degree distribution $P(k) \sim k^{-\gamma}$ for $N >> 1$.

To simulate networks with $\gamma < 2$ and to compensate finite-size effects in networks with $\gamma \gtrsim 2$ we need to introduce a cutoff in $\rho(\kappa)$. All in all, the distribution is

$$\rho(\kappa) = \frac{(\gamma-1)\kappa_0^{\gamma-1}}{1-\left(\frac{\kappa_c}{\kappa_0}\right)^{1-\gamma}}\kappa^{-\gamma} \text{ with } \kappa_0 < \kappa < \kappa_c \tag{3}$$

where $\kappa_0 = \dfrac{1-N^{-1}}{1-N^{\frac{2-\gamma}{\gamma-1}}}\dfrac{\gamma-2}{\gamma-1}\langle k\rangle$ and $\kappa_c = \kappa_0 N^{\frac{1}{\gamma-1}}$.

In the thermodynamic limit, the expected degree of a node with hidden degree κ is $\bar{k}(\kappa) = \kappa$ for a network generated with this algorithm, which explains the name.

B. The $\mathbb{H}^2$ model

In the $\mathbb{H}^2$ model, the hidden metric space is the hyperbolic plane, represented by a disk of radius $R_{\mathbb{H}^2} = 2\ln\frac{N}{\pi\mu\kappa_0^2}$ and metric tensor

$$ds^2 = dr^2 + \sinh^2 r d\theta^2. \tag{4}$$

Nodes in $\mathbb{H}^2$ are characterized by two variables, a radius r and an angle θ. Analogously to II A, nodes that are closer are more likely to be connected, but this time the distance x_{ij} between nodes i and j can be approximated by

$$\tilde{x}_{ij} \approx r_i + r_j + 2\ln\frac{\Delta\theta_{ij}}{2} \tag{5}$$

when $\sin^2\dfrac{\Delta}{2} >> \dfrac{\cosh(r_i-r_j)}{\cosh(r_i-r_j)+\cosh(r_i+r_j)}$ and $r_i, r_j >> 1$. It can be proved [5] that almost for all nodes of the networks we are working with, this approximation holds.

The only choice of connection probability between node i and j that creates heterogeneous, clustered and small-world networks with maximum entropy is the Fermi-Dirac form, so we can interpret the network as a set of fermions (edges) that can be in different states (pairs of nodes) with the hyperbolic distance as the energies and β the inverse of temperature.

$$p_{ij} = \frac{1}{1+e^{\frac{\beta}{2}(x_{ij}-R_{\mathbb{H}^2})}}. \tag{6}$$

The model generates networks with a power-law degree distribution if the angular distribution of nodes is uniform and the radii of nodes are distributed with the probability density

$$\rho(r) = \alpha\frac{\sinh\alpha r}{\cosh\alpha R_{\mathbb{H}^2}-1},\ r\in[0,R_{\mathbb{H}^2}] \text{ and } \alpha \geq 1/2. \tag{7}$$

In this case, the degree distribution's exponential is $\gamma = 2\alpha + 1$.

The map that connects the two models is

$$(\kappa_i,\theta_i) \mapsto (r_i,\theta_i) = (R_{\mathbb{H}^2} - 2\ln\frac{\kappa_i}{\kappa_0}, \theta_i). \tag{8}$$

Thereby, popular nodes are closer to the center of the hyperbolic disk, whereas not-so-popular nodes are near the boundary. It is easy to check that, for large $R_{\mathbb{H}^2}$, that is, $N >> 1$, if hidden degrees follow Eq. (3), then via the mapping, radial coordinates are distributed as in Eq. (7).

Analogously, we obtain the connection probability

$$p_{ij} = \frac{1}{1+e^{\frac{\beta}{2}(\tilde{x}_{ij}-R_{\mathbb{H}^2})}}. \tag{9}$$

As it was pointed out, almost for all pair of nodes $\tilde{x}_{ij} \approx x_{ij}$, so we can conclude that, in the thermodynamic limit, both models are equivalent.

III. EPIDEMIC MODELLING: THE SIS MODEL

We assume that the total population is fixed in N individuals and we divide them in two compartments: susceptible (S), who can contract the infection and infectious (I), who are contagious. We assume that the disease does not confer immunity, so, under some conditions, the cycle $S \to I \to S$ can be sustained forever. Infected individuals recover ($I \to S$) at rate δ, and a susceptible individual, in contact with an infected one, contracts the disease at rate λ.

Consider a network with N nodes. Let $\{a_{ij}\}$ be the coefficients of the adjacency matrix $\mathbb{A}$. We define $n_i(t)$ for all $i \in \{1,2,\ldots,N\}$ as follows:

$$n_i(t) = \begin{cases} 1 \text{ if node } i \text{ is infected at time } t \\ 0 \text{ otherwise} \end{cases} \tag{10}$$

So our system's state is $\vec{n}(t) = (n_1(t),\ldots,n_N(t))$. In the SIS epidemic model the evolution of the system is a Markovian process. Thus,

$$n_i(t+dt) = n_i(t)\eta_i(dt) + (1-n_i(t))\xi_i(dt) \tag{11}$$

where η_i and χ_i are random variables with Bernoulli distributions. η_i will be 0 with probability δdt and 1 with probability $1-\delta dt$, ξ will be zero with probability $p = \lambda \sum_j a_{ij} n_j(t) dt$ and 1 with probability $1-p$.

We define the *prevalence* of the disease as $\rho(t) = \frac{1}{N}\sum_i \langle n_i(t)\rangle$, so $\langle n_i(t)\rangle = \rho_i(t)$ is the probability that node i has the disease at time t.

Then,

$$\langle n_i(t+dt)|\vec{n}(t)\rangle = \langle n_i(t)\eta_i(dt) + (1-n_i(t))\xi_i(dt)|\vec{n}(t)\rangle = n_i(t)\langle\eta_i(dt)\rangle + (1-n_i(t))\langle\xi_i(dt)\rangle =$$
$$n_i(t)(1-\delta dt) + (1-n_i(t))(\lambda dt \sum_j a_{ij} n_j(t)).$$

In consequence, $\langle n_i(t+dt)\rangle = (1-\delta dt)\langle n_i(t)\rangle + \lambda dt \sum_j a_{ij}\langle n_j(t)(1-n_i(t))\rangle$. And, finally,

$$\frac{d\rho_i}{dt} = -\delta\rho_i(t) + \lambda\sum_j a_{ij}(\rho_j(t) - \langle n_i(t)n_j(t)\rangle). \quad (12)$$

Resolving Eq. (12) is quite complex, the state of a node i depends on the state of its neighbours. Furthermore, we are often interested in stationary solutions of the equations. For these reasons, we will apply some mean-field approximations.

A. IBMF

The idea of the individual-based mean-field theory (IBMF) developed in [6] is that the state of node i is independent of the state of its neighbours. In other words, we replace the dynamic correlations $\langle n_i(t)n_j(t)\rangle$ with $\langle n_i(t)\rangle\langle n_j(t)\rangle = \rho_i(t)\rho_j(t)$, so we obtain

$$\frac{d\rho_i(t)}{dt} = -\delta\rho_i(t) + \lambda(1-\rho_i)\sum_j a_{ij}\rho_j \quad (13)$$

Of course, $\vec{\rho} = \vec{0}$ is a stationary solution and a fixed point of the equation. In order to study the stability of the solution we consider the limit $\rho \approx 0$, a reasonable situation at the beginning of the epidemic and linearize Eq. (13) obtaining $\frac{d\rho_i}{dt} \approx -\delta\rho_i + \lambda\sum_j a_{ij}\rho_j$. Thus, the stability equation is

$$\frac{d\vec{\rho}}{dt} = \mathbb{M}\vec{\rho} \quad (14)$$

with $\mathbb{M} = -\delta\mathrm{Id} + \lambda\mathbb{A}$. The zero solution is unstable when the largest eigenvalue of $\mathbb{M}$ is greater than 0, then, $\vec{\rho} = \vec{0}$ is a dynamic repeller and another nonzero solution is the actual stationary state called endemic state. Otherwise $\vec{\rho} = \vec{0}$ is the only solution. In terms of λ and δ, the threshold condition is

$$\left(\frac{\lambda}{\delta}\right)_c^{\mathrm{IBMF}} = \frac{1}{\Lambda_{\mathbb{A}}} \quad (15)$$

where $\Lambda_{\mathbb{A}}$ is the largest eigenvalue of the adjacency matrix. If $\frac{\lambda}{\delta} < \frac{1}{\Lambda_{\mathbb{A}}}$, we obtain a healthy state, that is, $\vec{\rho} = \vec{0}$.

B. DBMF

Degree-based mean-field approximation (DBMF), developed in [7] is a course-grained version of IBMF. The idea is that all nodes of degree k are equivalent, so that for any node of degree k the probability of being connected to a node of degree k' is the same $P(k'|k)$. As already said, we only consider the two-nodes correlations. For this reason, DBMF is a mean-field theory.

The SIS prevalence equation for a node of degree k is

$$\frac{d\rho_k(t)}{dt} = -\delta\rho_k(t) + \lambda k(1-\rho_k(t))\sum_{k'} P(k'|k)\rho_{k'}(t) \quad (16)$$

As in Eq. (12), the first term represents the recovery of nodes of degree k, the second one, the infection of new nodes. It is proportional to the probability that a node of degree k is susceptible $(1-\rho_k(t))$ times the probability that this node is connected to an infected one of degree k', $P(k'|k)\rho_{k'}(t)$, times all the possible edges through which the node can be infected, k.

If the network is not correlated, $P(k'|k) = k'P(k')/\langle k\rangle$. Then,

$$\frac{d\rho_k(t)}{dt} = -\delta\rho_k(t) + \lambda k(1-\rho_k(t))\Theta, \quad (17)$$

with $\Theta = \sum_{k'} \frac{k'P(k')}{\langle k\rangle}\rho_{k'}(t)$.

Introducing the expression of Θ in Eq. (17) and imposing stationary solution condition $\frac{d\rho_k}{dt} = 0$, we obtain

$$\Theta = \frac{1}{\langle k\rangle}\sum_{k'} k'P(k')\frac{\frac{\lambda}{\delta}k'\Theta}{1+\frac{\lambda}{\delta}k'\Theta} \quad (18)$$

We are looking for nonzero solutions, so $\Theta > 0$. Then,

$$1 = \frac{\frac{\lambda}{\delta}}{\langle k\rangle}\sum_{k'}\frac{k'^2P(k')}{1+\frac{\lambda}{\delta}k'\Theta} \leq \frac{\frac{\lambda}{\delta}}{\langle k\rangle}\sum_{k'} k'^2P(k')$$

In consequence, the threshold value is bounded by

$$\left(\frac{\lambda}{\delta}\right)_c^{\mathrm{DBMF}} \geq \frac{\langle k\rangle}{\langle k^2\rangle} \quad (19)$$

It is easy to check that, in fact, $\left(\frac{\lambda}{\delta}\right)_c^{\mathrm{DBMF}} = \frac{\langle k\rangle}{\langle k^2\rangle}$. Indeed, suppose that $\lambda/\delta < \langle k\rangle/\langle k^2\rangle$, from Eq. (18) we obtain

$$1 < \frac{1}{\langle k^2\rangle}\sum_{k'}\frac{k'^2P(k')}{1+\frac{\lambda}{\delta}k'\Theta} \leq 1 \quad (20)$$

So, the condition for nonzero stationary solutions is

$$\frac{\lambda}{\delta} \geq \left(\frac{\lambda}{\delta}\right)_c^{\mathrm{DBMF}} = \frac{\langle k\rangle}{\langle k^2\rangle} \quad (21)$$

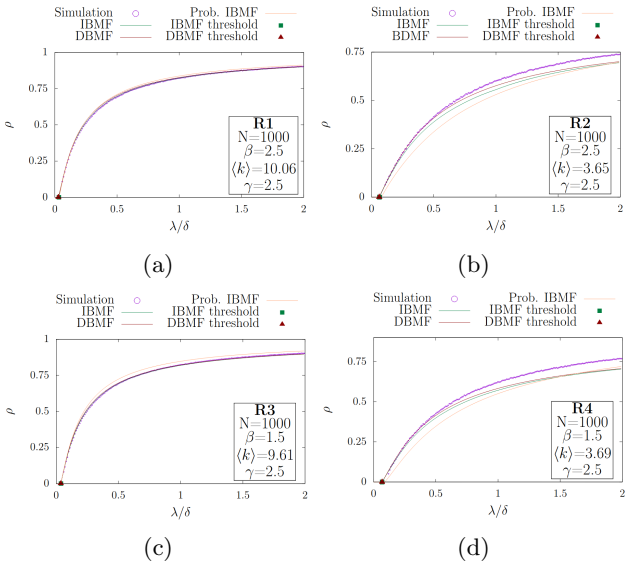


FIG. 1: Prevalence (ρ) of the disease in R1 (a), R2 (b), R3 (c) and R4 (d) networks as a function of λ/δ.

For scale-free networks, $P(k) \sim k^{-\gamma}$ with $\gamma \in [2,3]$, so $\langle k^2 \rangle \to \infty$ in the thermodynamic limit ($N >> 1$). Thus, $\left(\frac{\lambda}{\delta}\right)_c^{\text{DBMF}} \to 0$.

C. Probabilistic IBMF

The last mean-field approximation is a probabilistic one. Imagine a network whose edges appear and disappear at a higher rate than the disease ones. In this case, we cannot use the adjacency matrix $\mathbb{A}$ to compute the IBMF equation (13), but we can substitute each a_{ij} with the connection probability from Eq. (1).

Not only in this case it is interesting to use this approach, we can use it for all networks in order to obtain a reliable approximation of the epidemic prevalence.

IV. NUMERICAL SIMULATIONS AND APPROXIMATIONS

From now on, we will work with 6 different networks, four of them were generated with the algorithm on II A (named R1, R2, R3 and R4), the airports and preferred routes of USA [8] and email communication between members of Universitat Rovira i Virgili [9]. The details of each network are in Figure 1 and Figure 2

We compute the threshold values and perform the three mean-field approximation in the six networks, using Mercator [10], a tool to embed real complex networks in their hidden hyperbolic space, to obtain the hidden degree and angular position in the airports and emails networks.

We also run a SIS numerical simulation on each network on stationary solutions and compare it with the

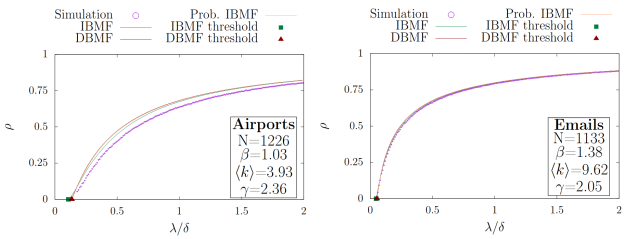


FIG. 2: Prevalence (ρ) of the disease in the airports (a) and emails (b) networks as a function of λ/δ.

approximations.

The results, in Figure 1 and Figure 2, show that, even though in the approximations we are considering the networks as tree-like structures, they work quite well, specially in R1, R3, airport and emails networks. Surprisingly, network R1 is very clustered, we are suppressing a considerable amount of triangles, however, being very connected (high degree), nodes are potentially in contact with infected neighbour nodes compensating the effect of not considering three-node connections. As expected, approximations are less reliable for networks with lower average degree since we are eliminating some of the few connections that can transmit the epidemic.

Although the results of the probabilistic IBMF approximation are similar to the simulations and the other approximations, it requires more information and is more resource consuming for not a huge improvement.

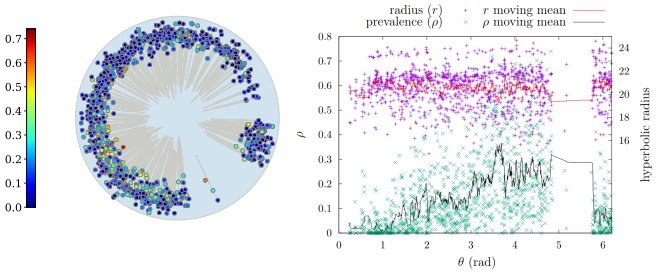
We can observe that threshold values computed following IBMF are very similar to the ones following DBMF, $\left(\frac{\lambda}{\delta}\right)_c^{\text{IBMF}} \approx \left(\frac{\lambda}{\delta}\right)_c^{\text{DBMF}}$ and both of them to the simulations. Another interesting point is that in all cases, the computed threshold values are very close to 0, as predicted.

V. ANGULAR EFFECT IN EPIDEMIC PREVALENCE

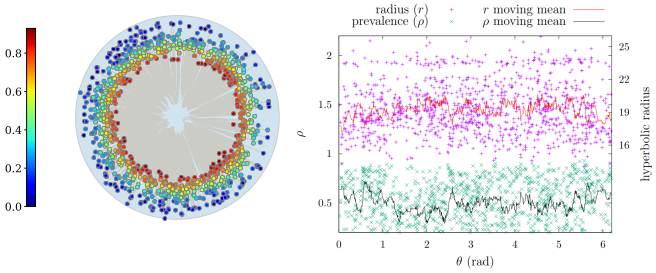
We now want to study if the angular situation of nodes affects the prevalence of the disease at the beginning of the epidemic, that is, λ/δ similar to the threshold value.

We run a simulation on the airports and emails networks on prevalence of stationary solutions of each node and we present the results embedded in the Poincaré disk, showing the geodesic connections between nodes too. We also represent these prevalences and the hyperbolic radii as functions of the angular position. The value of $\frac{\lambda}{\delta}$ on each network is different in order to get similar prevalences.

As it can be seen in Figure 3, the angular distribution has no effect whatsoever on the hyperbolic radius of the nodes, there is no pattern observed in all networks. Prevalence, unlike radius, in some cases can be affected by the angular distribution of the nodes. In some net-



(a)



(b)

FIG. 3: (a) Node prevalence (left) and hyperbolic radius and prevalence as a function of the angular position (right) in the airports network for $\lambda/\delta = 0.2$. (b) Node prevalence (left) and hyperbolic radius and prevalence as a function of the angular position (right) in the emails network for $\lambda/\delta = 0.075$.

works, under some circumstances, a more interconnected region can appear, maintaining the general degree distribution in it, which makes the disease more present in this type of regions, as can be seen in Figure 3a. Further study is needed to specify the exact behaviour of prevalence as a function of the angular distribution and the emergence of the more interconnected communities.

VI. CONCLUSIONS AND FUTURE WORK

In this work we studied a geometric network model and its relation with epidemic modelling. We first examined the hidden metric space network model and its possible embeddings, the circle $\mathbb{S}^1$ and the hyperbolic plane $\mathbb{H}^2$. Then we studied and justified different mean-field approximations to the susceptible-infected-susceptible (SIS) epidemic model. We compared these approximations to simulations in synthetic and real-world networks and proved that they are quite rigorous. We also showed that the angular distribution of the nodes in the network can affect the prevalence of the disease.

There are several ways in which this work could be expanded. Firstly, an interesting thing would be studying the effect of the angular distribution in the dynamics of the epidemic before arriving to a stationary state, how the disease is transmitted in the network dynamically. Secondly, how these mean-field approximations would hold in dynamic networks, graphs that their nodes and connections change with time, e.g. the Internet. Finally, we have not explained the emergence of more interconnected communities in networks and how their node angular distribution affects the prevalence.

Acknowledgments

I would like to thank my family and friends for their support. Last but not least, I would like to mention my advisor Dr. Marián Boguñá for his advice throughout the elaboration of this work.

[1] M. Ángeles Serrano, D. Krioukov, and M. Boguñá, "Self-similarity of complex networks and hidden metric spaces," *Physical Review Letters*, vol. 100, 2 2008.

[2] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, 8 2015.

[3] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, "Hyperbolic geometry of complex networks," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 82, 9 2010.

[4] M. Boguna, I. Bonamassa, M. D. Domenico, S. Havlin, D. Krioukov, and M. A. Serrano, "Network geometry," 1 2020. [Online]. Available: http://arxiv.org/abs/2001.03241http://dx.doi.org/10.1038/s42254-020-00264-4

[5] M. Ángeles Serrano and M. Boguñá, *The Shortest Path to Network Geometry*. Cambridge University Press, 1 2022.

[6] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking*, vol. 17, pp. 1–14, 2009.

[7] M. Boguñá and R. Pastor-Satorras, "Epidemic spreading in correlated complex networks," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 66, p. 4, 10 2002.

[8] "Air traffic control network dataset – KONECT," Oct. 2017. [Online]. Available: http://konect.cc/networks/maayan-faa

[9] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Phys. Rev. E*, vol. 68, no. 6, p. 065103, 2003.

[10] G. García-Pérez, A. Allard, Ángeles Serrano, and M. Boguñá, "Mercator: Uncovering faithful hyperbolic embeddings of complex networks," *New Journal of Physics*, vol. 21, 2019.