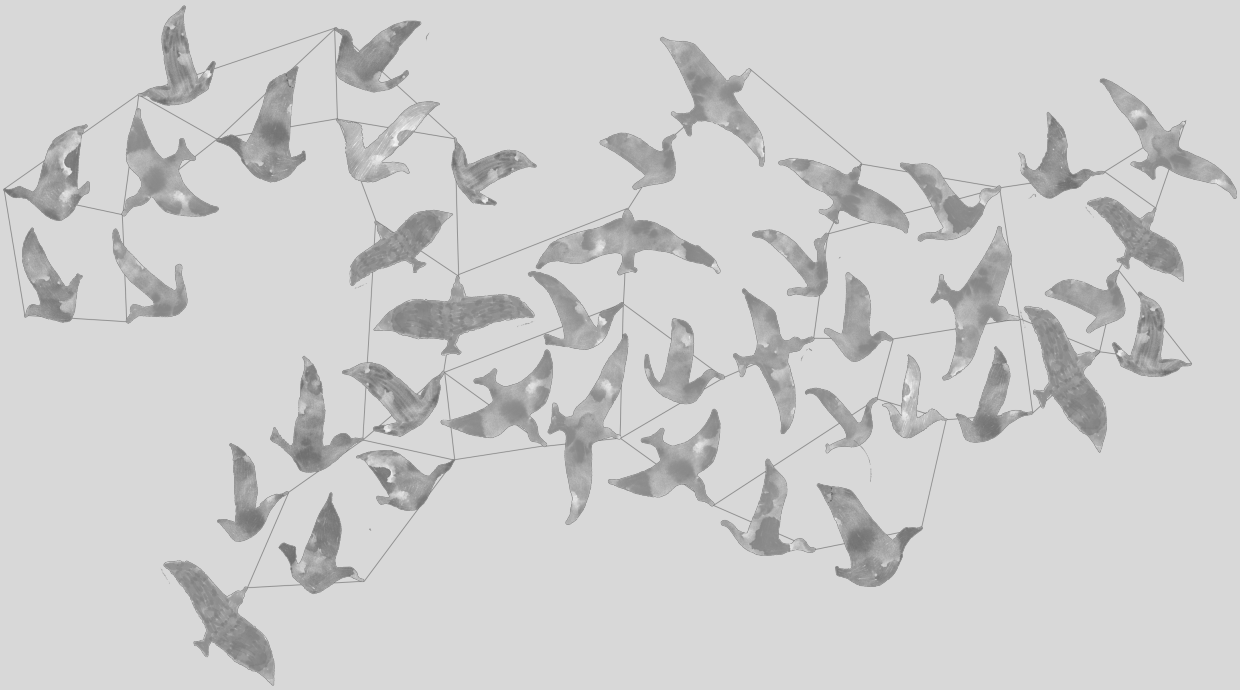
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS

Joan Tomàs Matamalas Llodrà

UNIVERSITAT
ROVIRA i VIRGILI

HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS

Joan T. Matamalas Llodrà



DOCTORAL THESIS
2019

JOAN T. MATAMALAS LLODRÀ

HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS

Ph.D. Thesis

Supervised by Dr. Alex Arenas and Dr. Sergio Gómez

DEPARTAMENT D'ENGINYERIA INFORMÀTICA I MATEMÀTIQUES

UNIVERSITAT
ROVIRA i VIRGILI

January 2019

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

UNIVERSITAT
ROVIRA i VIRGILI

WE STATE that the present study, entitled HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS, presented by JOAN T. MATAMALAS LLODRÀ for the award of the degree of Doctor, has been carried out under our supervision at the Department D'ENGINYERIA INFORMÀTICA I MATEMÀTIQUES of this university.

Tarragona, 7th January 2019

Doctoral Thesis Supervisors,

Dr. Alex Arenas Moreno

Dr. Sergio Gómez Jiménez

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

Als meus pares

A sa padrina

A la Marina

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

AGRAÏMENTS

No puc començar aquesta secció sense mostrar el meu agraïment als meus directors de tesi, el Dr. Arenas i el Dr. Gómez. Alex, Sergi res d'això no hauria estat possible sense la vostra guia i suport. Gràcies de tot cor per haver-me deixat formar part d'aquesta família que és l'Alephsys.

No puc oblidar-me tampoc de tots els membres d'aquesta família, els que són aquí i els que són fora, i que d'una manera o una altra han contribuït a que això haguí estat possible: la Julia, l'Elisa, l'Eugenio, la Serafina, el Manlio, la Clara, el Jordi.

Als meus dos germans petits acadèmics, el Benja i el Lluís, que són molt grans. Ànims, tot arriba!

Al tàndem de la UOC, Albert i Javi, pels matins arreglant el món.

A tots aquells amics que en moments de dubte, sempre m'han sabut aconsellar.

A la meva família, especialment als meus pares, pel seu suport i comprensió.

Finalment, a la Marina, sense tu segur que aquest viatge no hagués estat possible. Gràcies per cuidar-me i ajudar-me. I gràcies també per aquesta portada tan maca que m'has regalat.

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

SUMMARY

Dynamical processes running on complex networks have been largely studied due to their suitability to model very different scenarios. Understanding critical aspects of brain synapses, modeling how traffic flows in a city or modeling relations between species in ecosystem are a few examples of their potential applications. Nevertheless, there are some inherent limitations imposed by the way how these traditionally models are built, that could diminish their analytical power. For instance, consider the case of modeling mobility on a city. There, users can move between different places using different transportation systems, bus, subway, taxi, etc. Embedding this multimodality of systems inside our mobility model is critical to understand how it will behave in situations like congestion or failures since, different transportation modes have different capacities, costs or speeds. However, traditional network representations cannot embed this heterogeneity since the relations that they represent have to share a common definition. To overcome this issue, we should increase the structural complexity order of our model, enabling the use of different kinds of relations.

On this document, we focus on the understanding of how increasing the order of dynamical models, i.e., the amount of information represented, concerning structure, relational correlations, and temporal information and their effects on three topics that profoundly affect our lives: the emergence of cooperation, the spreading of epidemics and, finally, human mobility. We show that this increase on the complexity has a significant contribution to the way how we understand the way these processes work.

This document is organized as follows:

Chapter 1 provides a brief introduction to complex networks, explaining its origins and significant breakthroughs. Then, we move to provide a gentle introduction to the three subjects of study: cooperation, epidemic

spreading and human mobility. For each of them, we motivate the importance of the problem, a summary of the principal techniques used to analyze them and, finally, an analysis of their limitations.

In Chapter 2 we concentrate on the modeling of social dilemmas as a medium to study the emergence of cooperative behaviors. Using multilayer networks, hence, increasing the structural order of the model, and numerical simulations, we systematically study the evolution of cooperation in all four games in the $T - S$ plane, considering the density of cooperators, the convergence to equilibrium or the fluctuations of the system. More importantly, we show some remarkable and previously unknown features in the microscopic organization of the strategies, that are responsible for the important differences between cooperative dynamics in monoplex and multiplex networks. Specifically, we analyze the fact that, in the stationary state, there are individuals that play the same strategy in all layers (coherent), and others that do not (incoherent), proving their existence analytically in the harmony game, where only cooperation was expected.

We devote our efforts in Chapter 3 to explain a new way to model how epidemic diseases spread. We provide a new approach based on Markov chains which captures microscopic information about the epidemic properties of the links. The model is validated in synthetic and real networks, yielding an accurate determination of epidemic incidence and critical thresholds. Finally, thanks to the epidemic descriptors at the level of links, we propose a new containment strategy based on the removal of the links that mostly contribute to the spreading of the disease. We show how our method allows a faster containment of epidemic than other used methods, preserving the connectivity of the network, and hence, its functionality.

In Chapter 4, we expose the limitations that traditional mobility models have to capture recurrent patterns while, at the same time, retaining temporal information about the time spent on each location before moving somewhere else, an important issue, since these properties are inherent to human mobility. Then, we introduce an adaptive memory-driven approach to overcome such issues. At variance with the other compared models, it can realistically model conditional waiting times, i.e., the probability to stay in a specific area depending on individu-

als' historical movements. Our results demonstrate that, in standard mobility models, the individuals tend to diffuse faster than observed in reality, whereas the predictions of the adaptive memory approach significantly agree with observations. We show that, as a consequence, the incidence and the geographical spread of a disease can be inadequately estimated when standard approaches are used, with crucial implications on resources deployment and policy-making during an epidemic outbreak.

Finally, in Chapter 5, we provide a detailed analysis of the principal conclusions derived from the realization of this work. Analyzing our contributions, but also giving a critical view of our models, and exposing their limitations and possible ways to enhance them. At the end of this chapter, a brief detail of future work is depicted.

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

PUBLICATIONS

Publications covered in this Ph.D. Thesis:

[1] J. T. Matamalas, A. Arenas, and S. Gómez. "Effective approach to epidemic containment using link equations in complex networks." In: *Science Advances* 4.12 (Dec. 2018), eaau4212.

[2] J. T. Matamalas, M. De Domenico, and A. Arenas. "Assessing reliable human mobility patterns from higher order memory in mobile communications." In: *Journal of The Royal Society Interface* 13.121 (Aug. 2016), p. 20160203.

[3] J. T. Matamalas, J. Poncela-Casasnovas, S. Gómez, and A. Arenas. "Strategical incoherence regulates cooperation in social dilemmas on multiplex networks." In: *Scientific Reports* 5.1 (Apr. 2015).

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

People moving inside buildings, proteins interacting inside cells, cars traveling on cities or neurons interacting inside our brains have a thing in common: the behavior of an entity inside those systems cannot be understood without considering the interactions with the other bodies of the system. Straightforward interaction can produce complex phenomena like nonlinearities or emergence of structured behaviors like spontaneous order. For instance, consider bird flocks which can be composed of thousands of animals. Each bird tries merely to stay on curse and close to other birds by adjusting its direction and speed. Those animals also interplay with the environment following insects or avoiding predators. The sum of all these interactions between many components gives rise to the complex and beautiful movements that we can see on our skies. The study of these kinds of scenarios has given rise to complexity science; an interdisciplinary field that joins the knowledge of different areas like social sciences, mathematics, biology or physics to understand the underlying nature of many complex systems. Over the last decades, thanks to the availability of large volumes of data and the power of modern computers, the study of these systems has entered a golden age. It has enabled scientist to ask questions about crucial aspects of complex systems like how the relations between entities emerge and influence an interrelated system, giving rise to a new field of study: Complex Networks.

1.1 INTRODUCTION TO COMPLEX NETWORKS

One of the key aspects on the analysis of complex systems is understanding the influence and emergence of relations between the elements of the system. Graphs are the mathematical tools that enable the abstract representation of entities as a set of vertices, V, and the relations between them described as a set of edges, E. The bird of graph theory, the mathematical branch that study these structures, can be attributed to the Swiss physicist Leonhard Euler and his formalization and solution of the famous Königsberg bridges problem almost two centuries ago. Regular graphs have been deeply used since then. It was not until the beginning of the 20^{th} century when scientists focused on a different kind of networks.

The emergence of social science was the perfect substratum for the development of analysis techniques for networks that go beyond the properties of regular structured graphs since many human interactions can be modeled as a social network. One prominent example of the relationship between networks and social sciences can be found on an experiment designed to study the average path length on a social network, *small-world* experiment, proposed by the American psychologist Stanley Milgram [99, 155]. The empirical results of this experiment show that this distance is almost six, hence, the famous phrase "six degrees of separation"[1].

Almost thirty years afterwards, this empirical observation was theorized by Watts and Strogatz [162] who found that many real networks have properties like being highly clustered, common in regular graphs, and yet have small average path lengths, like the ones on random networks. They define the networks with this mixed properties *small-world* networks. It turns out that the definition of networks that have topological features beyond the ones that can be observed on regular graphs and random graphs is precisely the definition of complex networks.

A year later, another result boosts, even more, the popularity of networks science. Barabási and Albert discover that the distribution of

1 Although Milgram did not use this term himself.

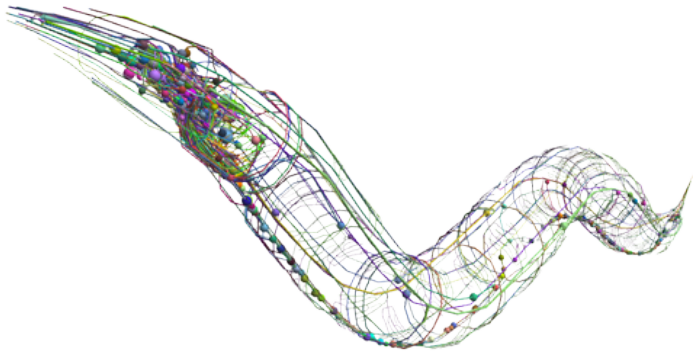


Figure 1.1: Neural network representation of the *Caenorhabditis Elegans*. *Credits: The OpenWorm Project*.

degrees on many real networks is highly heterogeneous [12]. They propose a model to understand how this phenomenon could be generated and propose the *preferentialattachment* model where nodes are sequentially added to the network, attaching them to the existing nodes with a probability proportional their degrees. This methodology generates networks that display a Power-law degree distribution, which is an essential result since many real networks like transportation systems [11], or human brain [148] have this property [116], hence, giving a plausible explanation of their formation.

Since these two breakthroughs, many other contributions have increased our understanding of these systems from different perspectives. From the way nodes and links between them are placed, information like the degree distribution, the clustering coefficient or the centrality measures can be gathered and interpreted, providing lots of useful insights. One of the aspects related with the structure that attracted the attention of many researchers was finding communities of densely connected nodes since they can be applied, for instance, to understand political polarization in social networks [33] or how proteins are grouped inside the cell [110].

Understanding networks as structures representing relations between entities is just the first step to unravel the dynamics that run on it. Take for instance the neural network of the *C. Elegans* nematode, Figure 1.1,

where nodes are represented by neurons and the links between them are represented by the synapses that connect them. Structural properties give powerful information, for instance about which node has a most central role on the network; however, the network has a functional purpose, there is a synaptic process running on top of it, and that is critical to understand this system [5]. In this case, there is an interplay between the structure of the network and the dynamical process that run on it, since the structure conditions the flow of information (signals) through the synapses, and the structure must be adapted to allow a correct functioning of the neural system.

From modeling the congestion on cities, [146] to understanding ecological food webs [75], the study of the dynamical process on complex networks has been intensely studied over the last years. On this document, we will focus on the understanding of how dynamical process and complex networks can help to understand three topics that profoundly affect our lives: the emergence of cooperation, the spreading of epidemics and, finally, human mobility. Although many scientists have already studied these topics, as we will see, we will offer a new way to analyze them, increasing the order of dynamical models, i.e., the amount of information represented, concerning structure, relational correlations, and temporal information. We will show that this increase on the complexity has a significant contribution to the way how we understand the way these processes work.

1.2 UNDERSTANDING COOPERATION

Cooperation is a ubiquitous and yet not-fully-understood phenomenon in Nature: from humans that cooperate to build complex societies to animals like wolves that hunt in packs to catch preys larger than they are, or meerkats that watch out for predators in turn while the rest of the colony feeds. Even small microorganism cooperate to survive in hostile environments. For instance, the *Dictyostelium discoideumu*, usually a solitary amoeba, when starves it associates with others to form a multicellular slug for the sake of survival. Explaining how cooperation has emerged and has resisted against more selfish behaviors is one of the biggest challenges in natural and social sciences.

		Opponent	
		Remain Silent	Betray
Player	Remain Silent	5 years	15 years
	Betray	0 years	**10 years**

Table 1.1: Tabular representation of the prisoner's dilemma game that specifies all possible outcomes. In bold we can see the equilibrium of the game.

A suitable mathematical way to model this kind of scenarios, where individuals can choose between cooperative and non-cooperative strategies, is game theory. Game theory is a field that studies the decision-making process when the final output does not just depend on the choice of one decision maker but also depends on the choices made by the other people. The study of this kind of situations is based on the perception of the outcome that an individual will receive depending on his choice an the other choices. The encoding of these perceptions on a matrix, the payoff matrix, is the usual way to represent a game. The study of this matrix leads to determine the best possible rational outcome. If all players are rational agents, and there is no information shared among the players, the best outcome is the one provided by Nash equilibrium [102], which describes the set of strategies that no player has the interest to change unilaterally.

To illustrate these concepts, let's consider the following example: imagine that the police arrests two persons accused of the same crime. The police confine them into two separate cells and make them an offer. They can either betray his partner (defect) or remain silent (cooperate). If both remain silent, they get a five-year sentence. If one betrays and the other remains silent, the one who betrays goes free and the other gets a fifteen-year sentence. Finally, if they betray each other, they get a ten-year sentence. See Table 1.1 for a full representation of the corresponding payoff matrix.

That is, maybe, the most famous dilemma in game theory, the prisoner's dilemma [7, 8, 125]. It explains a scenario where the best rational choice given by the Nash equilibrium is always defecting, no matter that a better possible outcome could be obtained if both subjects cooperate

(the Pareto optimal of the game). However, it turns out that in many real situations that exhibit the same distribution of payoffs and hence the same Nash equilibrium, a fraction of the players chooses to cooperate. The irrationality of some players could explain that but, as we will see, it can also be accounted for by the social structure of the individuals. All these situations could correctly hold in a real environment. Take for instance the case of meerkats that we mentioned at the beginning of the chapter. Clearly, those are not *rational* agents, at least from the human point of view, they are social animals and live in groups of up to 30 specimens where, when the colony feeds, some of the members watch out for predators. One possible way to model the emergence of cooperation in this scenario is to understand it as a repeated game played each time that the individuals change their roles. The subjects change their strategies observing not only how their strategies fit the environment but also observing how the strategies of their peers perform. Using this assumption, meerkats that choose to cooperate and remain vigilant for predators could have a longer survival rate and, as a consequence, the strategy of cooperation could be dominant in such a context. It turns out that modeling cooperation dilemmas as evolutionary environments where strategies compete and reproduce according to their fitness, are the perfect framework to study many cooperation scenarios; this theory is known as Evolutionary Game theory.

1.2.1 *Evolutionary Game Theory*

In the '70s, John Maynard Smith and George R. Price introduce [141] Evolutionary Game Theory as a version of classical game theory that uses concepts of Darwinian natural selection to model an evolving population of competing strategies. As we will see, this theory focuses not just in the quality of the strategy: it also depends in its frequency in the population.

There are two ways of modeling evolutionary scenarios. The first one is taking the agent approach and look at how each agent in the system evolve through time. This approach is good when the fluctuations, microscopic dynamics and the topology of the social neighborhood can be neglected. However, in many situations those requirements do not

need to be fulfilled and a mean-field approach that considers the entire population could be used. We will a focus on this last approach in this first part of the chapter.

Using evolutionary terms, the mean-field approach can be viewed as the evolution of a N individuals population of M species which evolves over generations according to the principles of Darwinian selection. In terms of game theory, it can be viewed as an iterative game where players change their strategies according to average payoffs of each strategy.

According to Szabó etal. [150], the following simplifying assumptions have to be made to apply the mean-field approach:

- The number of boundedly agents is very large.
- All agents are equivalent and have identical payoff matrices (they are playing a symmetric game), or they from two different groups for the two roles (asymmetric game).
- In each generation agents are randomly matched with equal probability.
- An update can be based on the average success rate of a strategy.
- All agents use the same strategy update.

Now, imagine that we have a population of two strategies: A with a frequency x_A and B with a frequency x_B. Let us define the vector $\vec{x} = (x_A, x_B)$ as the representation of the current population. Now we could define $f_A(\vec{x})$ and $f_B(\vec{x})$ as the fitness of both strategies in the current environment. Taking that into account, the following differential equations defines the change in the frequency of each strategy in the next generation:

$$\dot{x}_A = x_A \cdot (f_A(\vec{x}) - \phi) \tag{1.1}$$

$$\dot{x}_B = x_B \cdot (f_B(\vec{x}) - \phi) \tag{1.2}$$

where ϕ defines the average fitness of the game:

$$\phi = x_A \cdot f_A(\vec{x}) + x_B \cdot f_B(\vec{x}) \tag{1.3}$$

The population is constant, so $x_A + x_B = 1$. That means that we can express one variable in terms of the other, in this case $x_B = 1 - x_A$. Knowing that, the dynamics of the network is defined by the following differential equation:

$$\dot{x}_A = x_A \cdot (1 - x_A) \cdot (f_A(x_A) - f_B(x_A)) \tag{1.4}$$

That makes sense for a simple two strategy game, however a generalization of the dynamic is needed to cover a game with n strategies. The equation that specifies this behavior is known as the Replicator equation introduced by Peter Taylor and Leo Jonker [153]:

$$\dot{x}_i = x_i \cdot (f_i(\vec{x}) - \phi(\vec{x})) \tag{1.5}$$

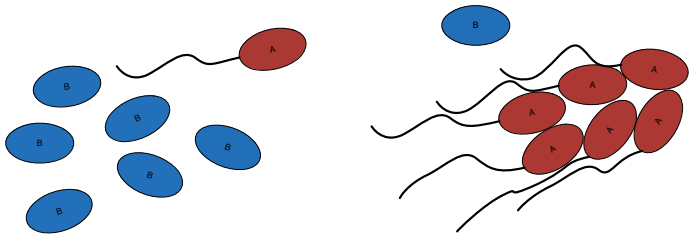
where the fitness of i could be computed as:

$$f_i(\vec{x}) = \sum_{j=1}^{n} a_{ij} \cdot x_j \tag{1.6}$$

a_{ij} is the respective entry in column i and row j of the payoff matrix $\mathbf{A}$. The average fitness $\phi(\vec{x})$ is calculated as:

$$\phi(\vec{x}) = \sum_{i} f_i(\vec{x}) \cdot x_i \tag{1.7}$$

To see how frequency based fitness works, we may use the example described by Nowak [106]. Imagine two phenotypes A and B, Figure 1.2. A can move while B cannot. A pays a certain cost for the ability to move, but also gains the associated advantage. Suppose the cost-benefit analysis leads to a fitness of 1.1 for A compared to a fitness of 1 for B. In this setting, fitness is constant, and A will certainly outcompete B, Figure 1.2a. But imagine that the advantage of being able to move is larger when few others are on the road, but diminishes as the highways get blocked up, Figure 1.2b. In this case, the fitness of A is not constant, but is a declining function of the frequency of A. A has a higher fitness than B when A is rare, but has a lower fitness than B when A is common. Then, frequency is crucial in evolutionary scenarios.



(a) Low proportion of A (b) High proportion of A

Figure 1.2: Example of frequency based dynamics in a population. We have two phenotypes A and B. A has the ability to move and B does not. If few other cells are moving, then A has a fitness larger than B. But if many other cells are moving, this fitness advantage is reversed. *Custom-made picture inspired from the original in [106].*

Evolutionary Stable Strategy

John Maynard Smith used evolutionary game theory to define a stable equilibrium in the population when he was unaware of the existence of Nash Equilibria in game theory [106]. However, a relation between the two definitions could be established if we consider that the fitness in classical game theory is the accumulated payoff of playing against all the players in the population in an iterative way.

We will start analyzing a symmetric game with just two available strategies defined in Table 1.8.

	A	B
A	a	b
B	c	d

(1.8)

We have the following situations:

- A is a strict Nash equilibrium if $a > c$
- A is a Nash equilibrium if $a \geq c$
- B is strict Nash equilibrium if $d > b$
- B is a Nash equilibrium if $d \geq b$

John M. Smith follows a different approach to arrive to a similar situation [141]. Supposing that we have a population of A. Could the introduction of an infinitesimally small quantity of B, ϵ, invade the entire population? To find a situation where the fitness of A is greater than the fitness of B, the following requirement has to be fulfilled:

$$a(1-\epsilon)+b\epsilon > c(1-\epsilon)+d\epsilon \tag{1.9}$$

Since ϵ is a really small value, we can cancel the terms where it appears. Therefore, the previous equation could be reduced to:

$$a > c \tag{1.10}$$

Nevertheless, if we found that $a = c$, then the Equation (1.9) is reduced to:

$$b > d \tag{1.11}$$

Consequently, to be a evolutionary stable strategy, the two conditions have to be fulfilled: $(a > c) \vee (a = c \wedge b > d)$

When that happens it can be guaranteed that the introduction of B in a population of A will not invade the entire population.

To see which is the relation between Nash Equilibrium and Evolutionary Stable Strategies, we will explore games with more than two strategies. The payoff for a strategy S_i versus S_j is given by $E(S_i, S_j)$. Thus, we have that a strategy S_k is a strict Nash Equilibria iff:

$$E(S_k, S_k) > E(S_i, S_k)\ \forall i \neq k \tag{1.12}$$

S_k is a Nash equilibrium if:

$$E(S_k, S_k) \geq E(S_i, S_k)\ \forall i \tag{1.13}$$

Finally, S_k is a evolutionary stable strategy if one of the following conditions is fulfilled:

$$E(S_k, S_k) > E(S_i, S_k)\ \forall i \neq k \tag{1.14}$$

$$E(S_k, S_k) = E(S_i, S_k) \ \wedge \ E(S_k, S_i) > E(S_i, S_i) \ \forall i \neq k \tag{1.15}$$

Observing the last equations, it can be verify how all the strategies that are strict Nash equilibriums are an evolutionary stable strategy too. However, that is not true for a Nash equilibrium. If $E(S_k, S_k) = E(S_j, S_k)$ and $E(S_k, S_j) < E(S_j, S_j)$, S_k is a Nash equilibrium but S_k cannot resist against an invasion by S_j. That means that strict Nash Equilibrium, implies evolutionary stable strategy that implies Nash equilibrium but not vice-versa.

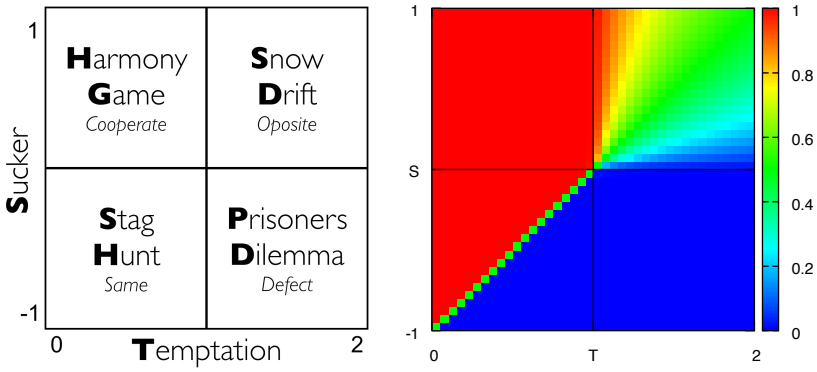
1.2.2 *Dyadic games*

For the sake of simplicity, in this chapter we will focus on dyadic games, strategy games between pairs of players. If we assume that each player in the system can either cooperate (C) or defect (D), a game can be defined according to its payoff matrix:

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} R & S \\ T & P \end{pmatrix}, \end{array} \tag{1.16}$$

where R represents the reward obtained by a cooperator playing against another cooperator, S is the sucker payoff obtained by a cooperator when it plays against a defector, the temptation payoff, T, is the payoff received by a defector when his opponent is a cooperator, and finally, P represents the penalty payoff obtained by a defectors which engages with another defector.

Traditionally the values of R and P are fixed to $R = 1$ and $P = 0$ to provide a fixed scale for the game payoffs [67, 105]. Applying this constraint, it turns out that the selection of the remaining parameters T and S enables the definition of several games according to their evolutionary stability on a T-S plane, as shown in Figure 1.3. Thus, if $R > S > P$ and $R > T > P$ the game is the harmony game [85]. The final state of a population playing this game will be total cooperation, regardless of the initial fraction of cooperators. Prisoner's dilemma [7, 8, 125], $T > R > P > S$, represents the opposite situation, and the population evolves towards total defection regardless of the initial



(a) $T - S$ Plane (b) Density of cooperators

Figure 1.3: (a) Representation of the Temptation-Sucker ($T - S$) plane. Over this plane the four possible equilibriums on dyadic games are represented. (b) Fraction of cooperator individuals on a population at equilibrium starting from a population composed by a 50% of cooperators and a 50% of defectors. Red represents a population composed entirely by cooperators and blue a population composed only by defectors.

conditions (although all players would be better off cooperating, hence the dilemma). A classical example of a coordination game, the Stag Hunt game [91, 134], is represented when the payoff values respect the order $R > T > P > S$, the output of this game will be either total defection or total cooperation, depending on the initial conditions. Finally, an anti-coordination game, the Snow Drift [142, 143], takes place if the payoff values satisfy $T > R > S > P$, where the final state will be a population made of both cooperators and defectors. The density of cooperators at the evolutionary stationary point at different regions of the $T - S$ plane can be seen in figure 1.3b.

1.2.3 *Evolutionary Game Theory on networks*

We have seen how we can define the evolutionary dynamics at population level using a mean-field approach, i.e, assuming that all players play against all the other players and the number of players is very large. However, the social structure of interactions could have a significant impact on dynamics and is one of the explanations behind the emergence

of cooperation [107]. If we have a network of interactions and a finite size of the system, we cannot use a population approach for studying how the process works, we have to use agent-level dynamics. These dynamics are usually described using strategy update rules, which describe how the agents perceive their environments, what information is necessary to evaluate their performances and the performances of their acquaintances, which are their expectations they form from former experience and, finally, how they use all that information to update their strategies. This strategy can mimic rational human learning or maybe be inspired by Darwinian natural selection.

There are an enormous variety of microscopic update rules in the literature, and the choice of which one to use depends on the particular problem that we are trying to solve. However, almost all the related work takes into account as a measure of the fitness of the individuals their own payoff. It could be the current payoff, the accumulated payoff over rounds, or maybe the prediction about the future payoff. For the sake of simplicity, we will assume that agents evaluate their fitness and the fitness of their neighbours looking just at the payoff obtained in the current round.

The round steps are always the same: first we calculate the fitness of the agents as an accumulated payoff received for playing against all their neighbours and then, the agents have the opportunity to update their strategy. However, there are two ways of doing that:

- **Synchronous:** All the agents compute their accumulated payoffs and then update their strategies at the same time step. This kind of update is applied, for instance, in biological models, when generations are clearly discernible in time [150].
- **Random sequential update:** players update their strategy independently of each other. One possible way of doing that is at each time step choose an agent at random and compute his current payoff and perform a strategy update. Thus, the chances that a given agent is selected for update are given by $\lambda = \frac{1}{N}$. We can also implement a Poisson clock inside each agent that makes the agent update his strategy at rate λ following a Poisson process.

Games on networks can be understood as a partial information game where players only can see how their peers perform but cannot evaluate where their payoff is coming from since they do not have complete information on the structure of the network. Thus, the selection of the strategy update rule becomes crucial.

As we have commented, the number of update rules to consider is large and worth to mention. We will start with imitation rules, a broad class of microscopic update rules. The essence of the imitation is that the agent who can revise his strategy takes into account the strategy of one of his acquaintances with some probability. It has to be said, that these rules are non-innovative, that means that they cannot introduce a new strategy that is not played in the population. The methods that belong to this class of update rule differ in two aspects: whom to imitate and with what probability.

First, we will start explaining a simpler deterministic rule, Unconditional Imitation. This rule works as follows: the player that has the option to update his payoff looks at his neighbors and chooses the one with higher payoff. If this neighbour has higher payoff than him then, the agent will change his strategy by the one that his neighbour is using. This very simplistic rule could work in many situations; however, it has the risk to converge too fast to a non-optimal solution [150].

The most important rule to our purpose is the proportional imitation rule, also known as the Replicator Rule. This rule is inspired in the replicator dynamics explained before because for large size of population, the evolution of the dynamics is equal, up to a time scale factor, to the one produced by replicator dynamics [128]. This rule has an explicit stochastic component and works as follows. Imagine that we have N agents, let be s_i the strategy of player i, W_i his payoff and N_i her neighborhood with k_i neighbors. With uniform probability, one agent j of his k_i neighbours is selected. The probability that player i adopts the strategy of player j is given by the following equation.

$$P(s_j^t \to s_i^{t+1}) = \begin{cases} \frac{W_j^t - W_i^t}{\Phi} & \text{if } W_j^t > W_i^t \\ 0 & \text{if } W_j^t \leq W_i^t \end{cases} \tag{1.17}$$

where Φ is a normalizing factor that has to ensure that probability goes between 0 and 1.

The next rule that is worth to mention is Moran update rule, inspired in the dynamic update rule for asexual replication [108]. At each time step, one individual is chosen for reproduction with a probability proportional to its fitness. An identical offspring is produced, which will replace another individual. That is for fully-connected populations, however when the individuals are connected in a network the process is a bit different. From the set of a node and its neighbours, one is selected proportionally to its fitness. The node then will change its strategy for the one that uses the selected one. Notice that with this rule a player can adopt, with low probability, the strategy of a neighbor that has done worse than himself.

Finally, we will comment one rule that is widely used in the literature, because of being analytically tractable. It is based on the Fermi distribution functions [128] and is known as the Fermi rule. A neighbour j of player i is selected at random with uniform probability. The probability that player i acquires the strategy of player j is given by the following equation:

$$P(s_j^t \to s_i^{t+1}) = \frac{1}{1 + e^{-\beta \cdot (W_j^t - W_i^t)}}, \tag{1.18}$$

where W_j^t and W_i^t are the payoffs of node j and i. The parameter β controls the intensity of selection, and can be understood as the inverse of temperature or noise in the update rule [128]. If parameter β is low, weak selection pressure is applied, and if is high, the selection pressure will be more discriminant.

Almost all the previous works in the field of games in structured populations has been done in the last two decades. In general, authors are focused in two main topics: how the structure of the network will impact on the evolution of cooperation and, on the other side, how the game dynamics will influence the structure of the network. We will follow this scheme to explain the most relevant works in the field. We will start looking at how the structure of the network conditions the behavior of players.

One of the first articles that make reference to the consequences of spatial structure was published by Nowak and May in 1992 [105]. In this work, they used a lattice, a bidimensional array, where agents play against the players in the adjacent cells. They discovered that, if they set all players to use cooperation, and strategically change to defection a few of them in particular positions, that will lead to the generation of kaleidoscopic patterns in the resultant population. However, lattices are not very useful to represent real-world scenarios, we need to look at how the dynamics work in complex structures.

Santos and Pacheco published one of the first works [137] about game theory in complex networks in 2005. They arrived at the conclusion that, in Prisoner's Dilemma and Snow Drift games, the use of networks generated via growth and preferential attachment leads to a dynamics where cooperation becomes the dominating trait. They asserted that the reason of such emergence is in the correlation between individuals. Those results are extended in [136], where the authors evaluated the resilience of cooperation for each one of the four equilibriums of games of two strategies.

The results obtained for scale-free networks are extended to more heterogeneous networks in papers like [57] and specially [128], where the authors checked the dynamics for several kinds of network structure, update rules and initial conditions. In [119] Poncela et al. discussed how cooperation is affected by the heterogeneity of the network. They used a mechanism described by Gómez-Gardeñes et al. [58] to generate network with a mixed degree of heterogeneity, between a random network and a network with power-law degree distribution, and they arrived at the same conclusion: heterogeneity enhances cooperation.

All these results are for spatial macroscopic descriptors of the network as the degree distribution. Nevertheless, the microscopic structure of the network could also lead to significant changes in the way how cooperation evolves. The study of features like how agents change strategies through time are quite interesting. Poncela and Gómez-Gardeñes [119] checked those features arriving at the conclusion that there are mainly three types of agents, mixed players (players that change their strategy as the evolutionary process runs), pure defectors and pure cooperators (players that never change their strategy). The organization of these

agents is worthy because always there exists a boundary of mixed players between pure cooperators and pure defectors. At the mesoscopic level, the community structure of the network present in many real social networks, has demonstrated to be important in the preservation of cooperation [89] under heavy temptation to defect conditions.

As we have said, structure modifies the game dynamics but dynamics can also modify structure. There are several papers related to this subject. Two of the most cited ones are from Poncela et al. [120, 121]. They proposed an evolutionary version of the preferential attachment model to grow a network with its main feature being that the capacity of a node to attract new links depends on a dynamical variable governed, in turn, by the node interactions. It turns out that the resultant network shows many features of real systems: scale-free degree distribution, hierarchical clustering and cooperative behavior.

1.2.4 *Structural higher-order*

Despite all these results, there are some important aspects regarding the cooperative behaviors not fully understood due to the limitations imposed by how we represent relational information. One of the most challenging is dealing with individuals that are interacting at the same time in multiple social contexts. Consider the following situation, where individuals interact in three different social scenarios: their workplace, their gym, and their home. On each of them, they could have social ties with different subjects, and they can behave differently on each situation. For instance, an individual could be gentle with the members of his family, but it could be a tyrant with his coworkers. However, the payoff that an individual obtains is an aggregation of the payoff obtained on all the social environments where he has ties and, since the peers of an individual have partial information about how he behaves in the context where they are peers, they could assume that behaves equally in all the contexts.

The implications of these situations are enormous and, being capable of modeling this scenario is the key to unveil complex cooperative patterns. These scenarios cannot be treated with classical network representations; thus, we need to resort to higher-order structures

that enable us to describe them. As will see in Chapter 2, the use of multilayer networks provides a perfect framework to tackle this problem and provides a new way to study the evolution of cooperation on structured populations.

Multilayer Networks

The emergence of the paradigm of multilayer networks has supposed a new way to understand and model complex interactions between entities. Nowadays, using the mathematical tools provided by this new framework, it is possible to describe simultanious relations among agents in different context and scenarios that can be structured in many levels of abstraction. Take for instance the example from [82] represented in figure 1.4; through the use of multilayer networks, they can model the behavior of scientists that may have attended to three major conferences, each one represented as a different level in the multilayer network, and also, as a consequence of the use of a second level of abstraction, they can also model different kinds of interactions, e.g., like if two scientists has talked to each other, or if they have attend to a talk by the other.

Notice that in this structure a scientist can be present or not in a conference layer depending on his attendence; however, if a scientist has attended to a particular conference, he will be present in both interaction layers. These kind of constraints is used to classify multilayer networks into different categories. For instance, if an entity is present along all the layers, we can classify this multilayer network as an interdependent network. If we also impose that the only allowed links between layers are those that connect an entity with its own representation in the other layers, we would talk about a multiplex network. It turns out that this last category of networks is quite useful to model many physical systems, for instance many transportation networks can be decomposed in layers either to represent their multimodal nature or to represent different attributes, e.g., the company that delivers the service, see Figure 1.5. For a complete description of the different multilayer categories see [82].

A formal way to define a single-layer graph G is in the form of a tuple $G = (V, E)$, where V is the set of nodes and where the set of

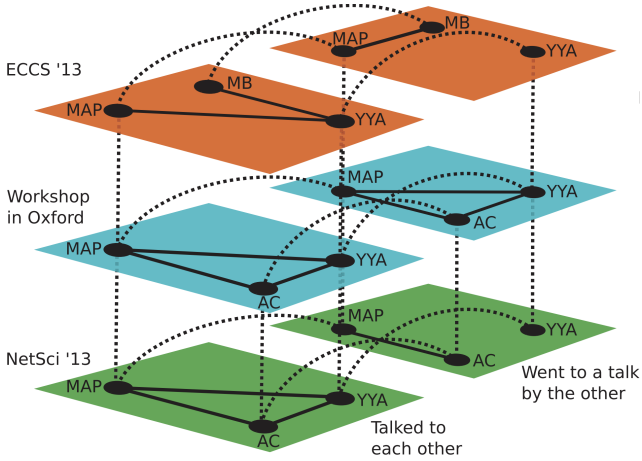


Figure 1.4: Representation of the Zachary Karate Club Club (ZKCC) multilayer network. The entities on this network are the winners of the karate club trophy: Aaron Clauset (AC), Cris Moore (CM), Mason A. Porter (MAP), Marián Boguñá (MB) and Yong-Yeol Ahn (YYA). This multilayer network encodes two aspects, the type of relationship between the scientists (talked to each other, went to a talk by the other) and a second aspect that stands for the conference in which the trophy was awarded. *Reprinted under Creative Commons CC BY license from [82].*

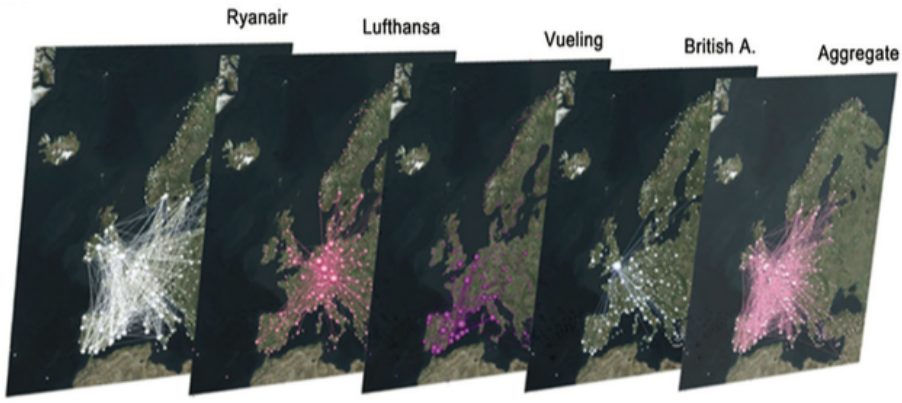


Figure 1.5: An example of a transportation network where each node is an airport and each layer represents the routes of an airline company. *Reprinted by permission from [36].*

edges that connect pairs of nodes is $E \subseteq V \times V$. Using this definition, one can extend this formalism to define any multilayer network M, where a node i in layer α is connected to any node j in layers β. As

single-layer graphs, multilayer networks contain a set of nodes V and edges E, but also an additional set of layers that represent different relational aspects, as we showed in the previous example. Thus, the organization of layers can be seen as a multidimensional representation of the different aspects, where each layer could be composed by a set of elementary layers[2]. Then, to represent a multilayer network with d aspects, a sequence $L = \{L_a\}_{a=1}^{d}$ of sets of elementary layers has to be defined such that there is one set of elementary layers L_a for each aspect a [82]. Using this definition, the set of layers that represents the multilayer network can be defined by the assembling of the set of all the combinations of elementary layers using the Cartesian product $L_1 \times L_2 \times \cdots \times L_d$.

In a general scenario, a node could be absent in some of the layers. Then, on the definition of a multilayer network, we should indicate whether a node is present on a layer or not. This can be done considering the case where a node is present in all the layers, defined as the Cartesian product $V \times L_1 \times L_2 \times \cdots \times L_d$, and selecting a subset that represents just that layers where the entity is present $V_M \subseteq V \times L_1 \times L_2 \times \cdots \times L_d$. It is also useful to define a proper way to represent in which layers a node i is present; this can be achieved using the tuple $(i, \alpha_1, \ldots, \alpha_d)$ that means that node i is present on layers $(\alpha_1, \ldots, \alpha_d)$.

In the same way, we have to find a proper way to define connections between pairs of node-layer tuples, considering that this connection could connect entities on the same layer, entities in different layers or even connect an entity with its representation in other layers. In single-layer networks, the set of edges is defined as a subset of the Cartesian product between all the vertices of the network $E \subseteq V \times V$. Making use of the definition of node-layer tuples V_M introduced before, we can extend this formalism to define the set of links in a multilayer network as $E_M \subseteq V_M \times V_M$.

Being all these components defined, now we can formalize a multilayer network as a quadruplet $M = (V_M, E_M, V, \mathbf{L})$. Note that this formalization is a generalization of the single-layer formalism that can be recovered setting the number of network aspects to $d = 1$.

2 Elementary layers are layers that define a particular aspect of a multilayer network.

The usual way to encode a single-layer network is using an adjacency matrix, $\mathbf{A}$, a square matrix with as many columns as vertices on the network, N, and where the value A_{ij} encodes the value of the relation between nodes i and j. If the network is undirected and unweighted then, $\mathbf{A}$ will be symmetric, and A_{ij} will be set to $\mathbf{1}$ if there is a link between nodes i and node j and set to $\mathbf{0}$ otherwise. This kind of network encoding can be extended for multilayer networks using a tensorial representation:

$$M_{\beta\tilde{\delta}}^{\alpha\tilde{\gamma}} = \sum_{\tilde{h},\tilde{k}=1}^{L} \sum_{i,j=1}^{N} w_{ij}(\tilde{h}\tilde{k}) \varepsilon_{\beta\tilde{\delta}}^{\alpha\tilde{\gamma}}(ij\tilde{h}\tilde{k}), \tag{1.19}$$

where $w_{ij}(\tilde{h}\tilde{k})$ stands for the intensity of the connection between nodes i on layer $\tilde{h}$ and node j in layer $\tilde{k}$, and $\varepsilon_{\beta\tilde{\delta}}^{\alpha\tilde{\gamma}}(ij\tilde{h}\tilde{k}) \equiv e^{\alpha}(i)e_{\beta}(j)e^{\tilde{\gamma}}(\tilde{h})e_{\tilde{\delta}}(\tilde{k})$ indicates the fourth order tensor of the canonical basis in the space $\mathbb{R}^{N\times L\times N\times L}$ [38].

Tensors represent a powerful way to represent multilayer networks, however for computational purposes sometimes it is more convenient to perform a flattening of the tensor into a matrix that enables the use of methods, tools and theoretical results for these kind of structures. Flattening a tensor into a $NL \times NL$ matrix generates an object called supra-adjacency matrix:

$$\bar{\mathbf{A}} = \begin{pmatrix} A_1 & C_{1,2} & \cdots & C_{1,j} \\ C_{2,1} & A_2 & \cdots & C_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ C_{i,1} & C_{i,2} & \cdots & A_i \end{pmatrix}, \tag{1.20}$$

where the diagonal block A_i is the adjacency matrix of the ith layer and the off-diagonal block C_{ij} represent the interlinks between layers i and j. Figure 1.6 shows an example from [82] where a multilayer network that represents the cognitive perception of social interactions of different subjects has been flattened into a supra-adjacency matrix.

The formal mathematical definition of multilayer networks has led to the extension of some structural properties present in single-layer networks [15, 34, 38] or to the definition of new ways to detect mesoscale

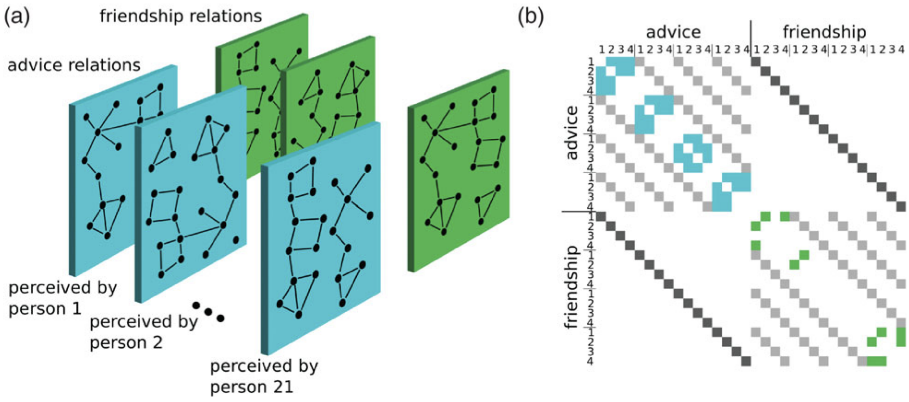


Figure 1.6: (a) Schematic of how different subjects perceive social structure represented as multilayer network. This network encodes two aspects, one for encoding the kind of relationship either advice relations or friendship relations (represented in different colors) and the second to represent the subject how is perceiving the interactions. (b) Representation of the encoding of the previous multilayer networks as a supra-adjacency matrix. Note that in this case the interlayer are just the connections between the same node in different layers, thus the interlayer blocks of the matrix are diagonal blocks. *Reprinted under Creative Commons CC BY license from [82].*

structures [37, 52, 101, 118]. One of these properties is centrality, the determination of the *central* nodes in a network, one of the most studied metrics on single-layer networks that in multilayer networks acquires a new dimension. Besides the redefinition of standard centrality measures like PageRank [66] or eigenvector centrality [144], the realization that multilayer centrality becomes a measure of node versatility [40] shows the potential of multilayer structures. Versatile nodes are those that play the most central role taking into account the whole multilayer structure; these nodes are those that bring cohesion to the multilayer structure since, through them, different aspects of the networks are bridged together. It can happen that nodes that are the most central on a particular aspect are not especially central when the whole structure is considered, see Figure 1.7 for an illustrative example.

One of the most relevant aspects derived from the use of higher-order structures on networks using multilayer networks, is their effect on the dynamics that run on top of them. Take diffusion, for instance,

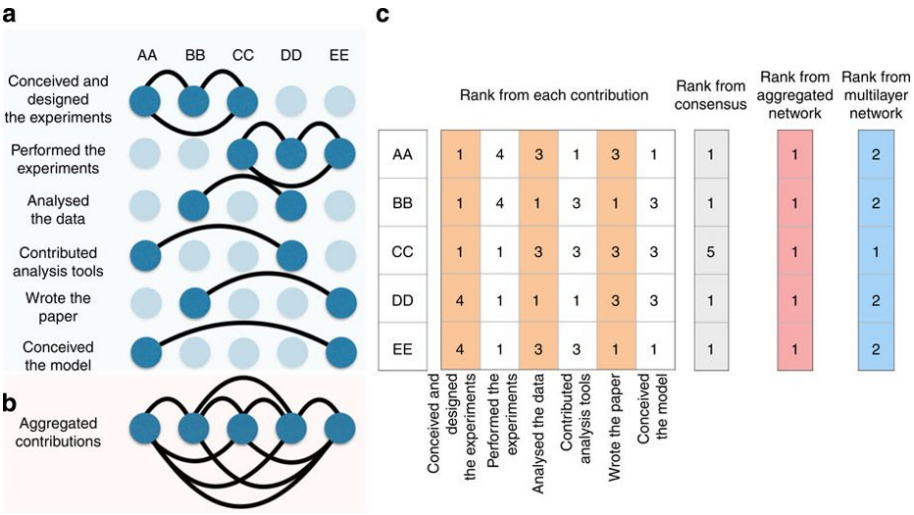


Figure 1.7: Consider five authors (AA, BB, CC, DD and EE) and their corresponding contributions. Each contribution represents one layer of a multilayer network (shown in a), where nodes are the above authors. In each layer, a clique is built with all authors participating in the corresponding activity. (b) shows the network resulting of the aggregation of the multilayer information. A table where columns report the rank of each author in each layer separately is shown in (c), obtained using eigenvector centrality. Interestingly, versatility and a weighted consensus, will rank author CC first. Therefore, four authors out of five are central in this network and it is not possible to remove this ambiguity by calculating centrality in the aggregated network. Conversely, the multilayer analysis provides a unique versatile (most central) author, CC, solving the ambiguity raising with classical approaches. However, after a more careful inspection, CC is the only author bridging the two layers with the largest number of contributions and for this reason her role must be more central than other authors, even if they contributed in more layers. *Reprinted by permission from [40].*

one of the most well-studied dynamics in single-layer networks; it has been proved [60, 145] that, when a multilayer structure is considered, diffusion can run faster than on individual layers. This is one of the aspects that can be studied analyzing the smallest positive eigenvalue Λ_2 of the supra-Laplacian matrix, an extension of the Laplacian matrix constructed upon the supra-adjacency matrix.

Random walks, another well-studied dynamic on single-layer networks, has also been reviewed for layered networks [39, 101, 124], a structure

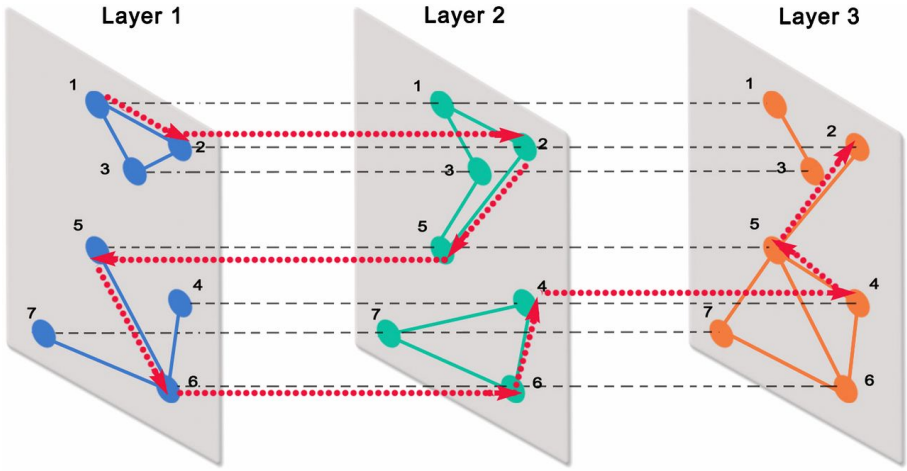


Figure 1.8: Example of the navigation on an interconnected network. Path (dotted red line) of a random walker exploiting the multiplex topology to find a way around to jump between disconnected components. In this example the walker is not allowed to switch between layer 1 and layer 3 in one time step. *Reprinted by permission from [39].*

that has significant implications on the way how walkers move. On multilayer networks, walkers can make use of interlayer edges to jump from one layer to another layer, see Figure 1.8. This new ability enables the definition of new metrics like network navigability [39], the fraction of nodes that are visited by a random walker in a finite time. This new property can be used to show that multilayer networks can be more resilient to random failures since the navigability can be more significant on those networks than on their aggregate projections.

Finally, another dynamic that deserves to mention is congestion since many transportation systems can be understood as a multiplex network due to their multimodal nature. It has been shown that multiplexity can induce congestion even when each independent layer remain decongested [146].

1.3 UNDERSTANDING EPIDEMIC SPREADING

No matter if they have been transmitted by air like influenza or tuberculosis, by sexual intercourse like AIDS or by contact like many

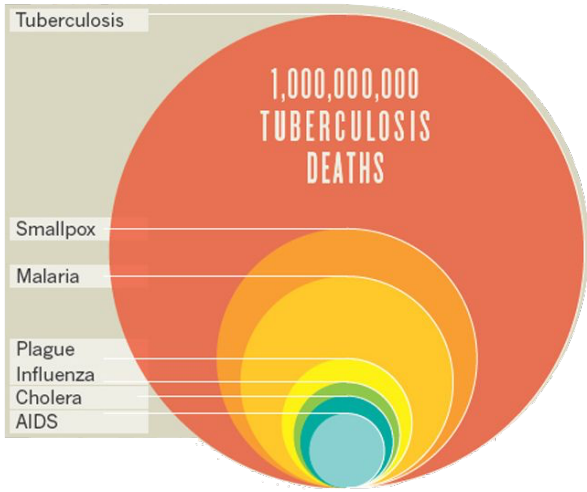


Figure 1.9: Number of deaths by infectious diseases over the last two centuries. *Reprinted by permission from [117]*

deadly parasites, the infection by an infectious pathogen has a significant impact on the way how we live our lives. Their effect could go from a cough and sore throat, like in common cold infection, to death like on infections produced by Ebola or tuberculosis, the most deadly disease responsible of the death of nearly 1 billion people over the last two centuries, figure 1.9. Their effects are not just constrained to individuals; since humans live on an interconnected complex society, many cultural traits have been affected and will keep being affected by the spreading of these microorganisms. It has changed the shape of country borders: in 1880 Tunisia was devastated by disease epidemics of cholera and typhoid, and the impact on the economy was so high that it was unable to pay its debts, leading to the occupation by France. It has changed how we eat: it is believed that the prohibition of eating pork in Judaism and Islam is an ancient way to avoid the infection by pig tapeworms or by the *Trichinella* [98]. The economy is also critically conditioned by infectious diseases: the embargo imposed by the European Union against British beef as a consequence of the spreading of Bovine spongiform encephalopathy had an impact over the British economy of $5.75 billions [51].

One prominent example of how an infectious disease changed at large scale the live of millions of people life is the European bubonic plague, also known as *the Black Death*, that had his deadly peak between 1347 and 1351. The cause of the pandemic plague was the bacterium *Yersinia pestis*, a pathogen commonly present in fleas carried by ground rodents. Transmitted by flea bites, contact with contaminated fluid or tissues or by infectious droplets [3] induces dark blotches on the skin an swollen lymph nodes to people who get the disease. At least 20 million people died in the peak years, a value that accounts for two-thirds of the European population at the time. Due to this death toll, European society was profoundly changed, since it slowed the industrial development and the economic growth, reversing the flow from people that had begun to move to cities from rural areas [68]. It would take centuries before Europeans recovered from the plague.

Pathogens are so critical to human life and societies that the control over the ways to treat and contain them has also shaped our culture. A paradigmatic example is the *quinine*. Over centuries Malaria was one of the most dangerous epidemics, accounting for a large number of deaths. The highly populated Rome was not an exception, where the disease was endemic; in fact, the word *malaria* came from the Italian *mal aria*, "bad air", since it was believed that the disease came from the marshes. The people in Rome was so concerned by this situation that, in 1623, after the death of ten cardinals, Pope Urban VII declares that a cure for malaria must be found [6]. Although Spanish Jesuits sent to America already knew the medicinal properties of the bark of chichona tree on the 16th century, it was not until 1631 that the Jesuit brother Agostino Salumbrino noticed that Quechua were using the bark to treat malaria successfully [129]. News of this discovery spread across the globe, unleashing a race to control the production of this plant, as a consequence, it became a commodity more valuable than gold [6]. Quinine played an essential role on the colonization of Africa since it enabled Europeans to deploy large armies without the burden of malaria. It also played an essential role on wars: during World War II, allied forces lost control over the production of quinine since Axis powers had occupied the countries were the main plantations were

3 https://www.cdc.gov/plague/transmission/index.html

installed. Concerned by this situation, US managed to obtain millions of seeds from plantations in the Philippines, under US control before the Japanese invasion, and moved them to plantations in Costa Rica. The production from these plantations was used to feed the need of the US Army forces fighting on Africa and the South Pacific [32].

Given the importance that epidemic spreading and their containment on society, the scientific community has put many efforts to understand their underlying mechanisms. Specialists from many fields like biologists, epidemiologists, social scientists or urban scientists try to unveil some fundamental properties of epidemics processes from different angles. Here, we are interested in the mathematical modeling of the underlying physical system that sustains the endimicity of diseases.

1.3.1 *Mathematical Modeling*

The modeling of physical processes supposes a new way to understand how epidemics works, from the analysis of the onset of an epidemic outbreak [22–24, 27, 61, 62, 74, 103, 112, 114, 159] to predict the effect of a vaccine [47, 55, 69, 92, 115]. The complexity of these models vary depending on the subject of study. Computational models built using machine learning techniques fitted by huge volumes of data could be useful when we are trying to predict with high accuracy which is the exact evolution of the incidence of infection or, also, to determine which are the phenotypical attributes that are related on getting a disease. However, due to their complexity, extract understanding from these models is almost an impossible task. Another approach is to use simpler models, assuming ideal situations, that enable us to make hypothesis about the general traits of the underlying dynamical process [3, 4, 35, 42, 70, 79, 113]. Although simpler, these models have been proven robust and their results can be extrapolated to real-life situations. We will focus on this last kind of models introducing three different complexity orders of the well-known compartmental models

Compartmental models are based on the segmentation of individuals regarding their pertinence to different epidemic stages. Individuals that are healthy and can get infected by a given disease are classified as *susceptible* (S). As a consequence of the epidemic process these

individuals can get infected by contact with the pathogen; when this happens, they do not become infectious immediately, the pathogen needs time to reproduce inside the host's body. This period is known as *incubation period*, and each disease has its own characteristic incubation period, e.g., chickenpox has a period of 14-16 days. When individuals are incubating a disease, they are classified as *exposed* (E), they are infected but cannot transmit the disease yet. At a given point the volume of pathogens inside the host's body arrives a critical point, when that happens the individuals become *infectious* (I), that means that they can spread the disease to other members of the population. At this point, two different things can happen: or the individuals recover from the disease or they die. If they recover, and through the antibodies produced fighting against the pathogen, they become immune to the disease, then they are classified as *recovered* (R).

Epidemic models built using compartments describe disease by their epidemic stages, see Figure 1.10. *SEIR* diseases are those where individuals pass for all the stages previously described, measles, rubella or chickenpox are some of them. In some situations, the use of an *exposed* compartment can be omitted since it adds complexity to the model, and in many scenarios, the incubation period of diseases is small enough. Another type of diseases are those defining situations where, once the host is infected, he will remain infected until he dies. In this situation, the only required compartments are *SI*; a classic example of a disease with this behavior is HIV/AIDS. *SIS* is one of the most studied compartmental models that, considering its construction, leads to the appearance of cyclic dynamics, since infected individuals become immediately susceptible to a new infection when they recover. This model is applied to diseases that display a repeated infection, like the common cold or sexually transmitted diseases like gonorrhea or chlamydia. If the immunization provided by antibodies last for a certain amount of time, *SIRS* models can be used to capture this peculiarity.

The rest of this section will be devoted to detail three different ways to analyze one of these models, *SIS*, since, as we have said, dynamics following these stages are interesting because they account for a large number of diseases and exhibit interesting behaviors like cycles. We will see how these dynamics operate under homogeneous populations,

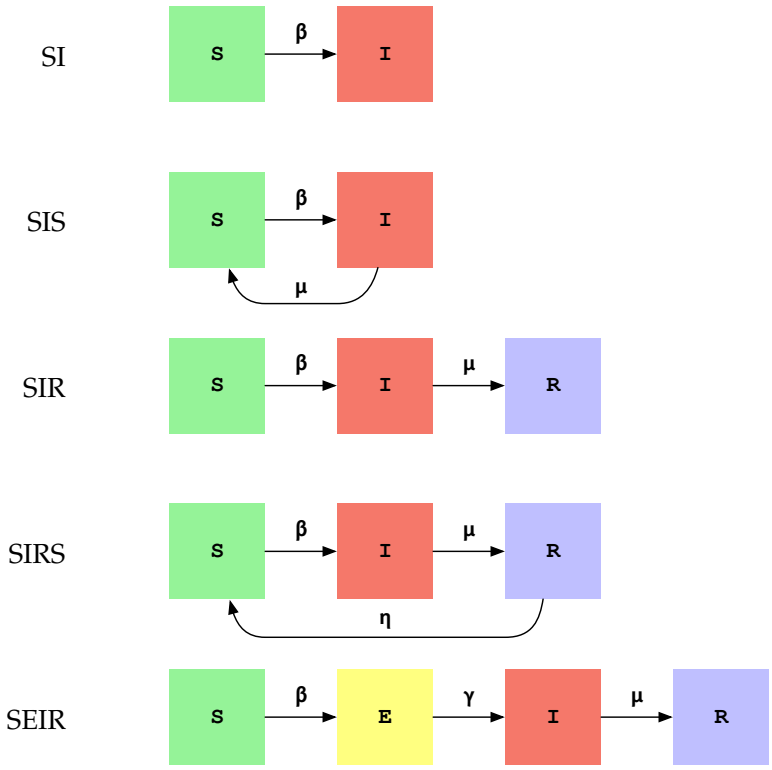


Figure 1.10: Illustration of different epidemic models build as a reaction-diffusion processes. Each box represent a different stage of an epidemic process, the variables near the arrows represent the transfer rates between each compartment.

heterogeneous populations and, finally, how they work at individuals microscopic level.

1.3.1.1 *Deterministic SIS under homogeneous population*

A classical way to analyze compartmental models is as reaction-diffusion systems. Those systems are traditionally used to model changes in space and/or time on the concentration of several chemical substances but, due to their generalization power, are often used also in biology, geology or ecology. In epidemiology, we can understand elements in the compartments as substances that can react and transform to other

substances. In the *SIS* compartmental model we have two possible reactions:

$$S + I \xrightarrow{\beta} I + I \tag{1.21}$$

$$I \xrightarrow{\mu} S \tag{1.22}$$

The first one accounts for the infection of healthy individuals (transitions $S \to I$). This reaction is proportional to the number of infected and susceptible individuals on the population, since infected individuals are needed to spread the disease at rate β. The second term stands for the espontaneous recovery of an infected individual at rate μ. Notice that assuming a Poisson process, $1/\mu$ determines the average infectious period.

For the sake of simplicity, assume that we have a constant population that does not change due to demographic effects like births, deaths or migrations. Also assume that we have a large population with homogeneous mixing, i.e., individuals have the same probability to interact with any other member of the population. Then, these epidemiological constants enable us to write the following set of differential equations.

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N} + \mu I \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \mu I\,, \end{aligned} \tag{1.23}$$

where S and I are the number of susceptible and infected individuals and, then, the total number of individuals on the population is $N = I + S$. Since the population is constant, we can express the previous equations in terms of fraction of individuals:

$$\begin{aligned} \frac{ds}{dt} &= -\beta si + \mu i \\ \frac{di}{dt} &= \beta si - \mu i\,, \end{aligned} \tag{1.24}$$

where s and i represent the fraction of susceptible and infected individuals on the population and then, it holds that:

$$\frac{ds}{dt} + \frac{di}{dt} = 1 \Rightarrow s(t) + i(t) = 1 \tag{1.25}$$

As the system is closed and the susceptible fraction of individuals can be determined as $s(t) = 1 - i(t)$, we can rewrite the Equation (1.24) as:

$$\frac{di}{dt} = (\beta - \mu)i - \beta i^2 \tag{1.26}$$

which is the logistic equation that implies that for all value of i grater than 0. This differential equation has two different asymptotic solutions:

$$\begin{aligned} \frac{\beta}{\mu} \leq 1 \quad &\Rightarrow \quad \lim_{t\to\infty} i(t) = 0 \\ \frac{\beta}{\mu} > 1 \quad &\Rightarrow \quad \lim_{t\to\infty} i(t) = \frac{\beta - \mu}{\beta} \end{aligned} \tag{1.27}$$

The ration β/μ is crucial on epidemiology because it determines the epidemic threshold, the point where an infectious disease can be established as endemic on a population. The value is also referred to as the *basic reproductive number*, or simply R_o, since it determines the average number of secondary cases arising from a primary case in a population composed entirely by susceptible individuals [42]. The rationale behind the epidemic threshold in the case of homogeneous populations is quite simple, since the only condition that has to be fulfilled to become an endemic is that the rate of infections β has to be larger than the recovery rate μ, ($\beta > \mu$), otherwise the outbreak will eventually die. On Figure 1.11, we can see the fraction of infected individuals on a population as a function of the *basic reproductive number* R_0, which is larger than 0 when $R_0 > 1$.

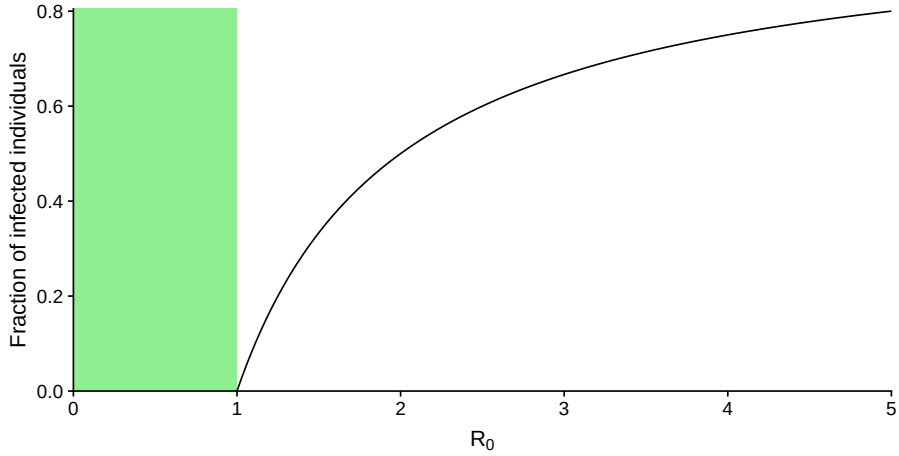


Figure 1.11: Fraction of infected individuals using homogeneous SIS mean-field dynamics as function of R_0. The green area represents a subcritical region where the epidemics does not survive.

Equation (1.27) can be solved for $R_0 > 1$ to analytically get the evolution of the fraction of infected individuals over time by, transforming the variable i to y^{-1} on the mean-field equations [71]:

$$i(t) = \frac{i_\infty}{1 + Ve^{-t(\mu-\beta)}}, \tag{1.28}$$

where the endemic infected population is defined by $i_\infty = (\beta - \mu)/\beta$ and $V = i_\infty/i_0 - 1$.

1.3.1.2 *Heterogeneous populations*

Despite its simplicity, the previous model enables us to understand some interesting properties about epidemic spreading like the basic reproductive number. However, it is built under some hard assumptions that have to be relaxed in order to understand the contribution of different aspects of population dynamics. Assuming that individuals interact homogeneously between them is a strong constraint because the population usually has some structure, and this structure is critical to understand spreading patterns. For instance, children are usually considered *super-spreaders*, and the elderly are more prone to be infected because they usually do not have a strong immunity system. One way to consider these new scenarios, it is increasing the order of the model

to consider that compartments now are also segmented in different types of individuals. The epidemic rates have to consider all possible different interactions between each different type of individuals. For example, if elderly and young people compose the population, we need a rate β_{ey} that stands for the rate at which the elderly will infect by their contact with young individuals.

Another way to consider heterogeneity on the population is assuming that the contacts between individuals are modeled using a complex network. Two individuals will be related if there is a contact between them. A way to approach this scenario is considering that there is an epidemiological equivalence between individuals that have the same number of neighbors, since they will be exposed to a similar number of contacts. Following the same idea that we explained before, we can increase the order of the homogeneous mean-field model by the segmentation of each compartment. Following this approach, each compartment α will be composed by different $\rho_k^\alpha(t)$ segments, each one representing the fraction of individuals of degree k inside it. This fraction can be seen as the probability that an individual with degree k can be found in the α compartment, since $\sum_\alpha \rho_k^\alpha(t) = 1$. The total fraction of individuals at α compartment can be obtained as $\rho^\alpha = \sum_k P(k)\rho_k^\alpha(t)$, where $P(k)$ is the probability of finding a node with degree k. Notice that on this approach, we do not use the complete information about the connectivity of the networks, this is summarized on the $P(k)$ and on the probabilities that a node of degree k is connected to a node of degree k', $P(k'|k)$.

This approach has successfully ben applied to SIS models [114]. In this situations the temporal evolution of an epidemic process can be described using the fraction of infected individuals $\rho_k^I(t)$:

$$\frac{d\rho_k^I(t)}{dt} = -\rho_k^I(t)\mu + \beta k[1-\rho_k^I(t)]\sum_k P(k'|k)\rho_{k'}^I(t)\,, \tag{1.29}$$

where β and μ are the infections and probability rates, respectively. The first term of the equation stands for the spontaneous recovery of infected individuals. The second one accounts for the infection of individuals. This value is proportional to the probability that a node of degree k is on the susceptible compartment, $1-\rho_k^I(t)$, to the number

of neighbors k, to the probability that a node is related to a node of degree k', $P(k'|k)$, to the probability that the node with degree k' is on the infected compartment, $\rho^I_{k'}(t)$, and to the infection rate β.

The fact that the system of Equations (1.29) depends on the probability of degree correlations makes the system unsolvable in a closed form. A way to overcome this issue is using linear stability analysis [48] that can be used to determine the epidemic threshold analytically:

$$\lambda_c = \frac{1}{\Lambda_{\max}}, \tag{1.30}$$

where $\Lambda_{\max}$ is the largest eigenvalue of a matrix whose elements are defined by:

$$C_{kk'} = kP(k'|k) \tag{1.31}$$

In [114] proved that, if the network does not show correlations between degrees, $P(k'|k) = \frac{k'P(k')}{\langle k \rangle}$, then for any $\lambda = \beta/\mu > \lambda_c$ the epidemic enters in the endemic state and the epidemic threshold can be computed as:

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{1.32}$$

As a consequence, if the network follows a power-law degree distribution with exponent between 2 and 3 (the ones most common in real networks), the epidemic threshold vanishes and the disease becomes endemic regardles of the epidemic parameters.

1.3.1.3 *Microscopic Approach*

The previous approach has the advantage of increasing the specificity of the model, allowing different infectivities based on an underlying network of contacts between subjects. However, the particular role of each of these individuals on the system is lost when we aggregate all the individuals with the same degree in the same category. This can not be underestimated since the topological properties of the network, like clustering coefficient or mesoscopic scale, affect genuinely how the epidemic process diffuses on the network. If we want to consider these properties on the model, we have to increase even more its order and

consider the weight of the state of each particular node on the network for the epidemic process.

A traditional way to overcome this issue is to rely on *Monte Carlo* simulations to capture the interaction between all nodes. However, due to its stochastic nature, this system can not be easily analyzed. A possible way to overcome this limitation is with the use of the probabilistic *microscopic Markov chain approach* (MMCA) [27, 61]. Besides capturing the global dynamics, this method also captures the microscopic state of the nodes, i.e., the probability that a given node belongs to S or I. In comparison to the other exposed methods, MMCA does not make aggregation based on any property; it uses all the information on the network of contacts encoded in the adjacency matrix $\mathbf{A}$. The only assumption that we made when we use this method is that the probability of a node i of being on any state is independent of the state of any of its neighbors.

Two types of processes can express the spreading on a network at the microscopic level: using a *contact process* where the infection is expanded to one neighbor at a time with a given rate. The other option is using a *reactive process* where the pathogens diffuse over all the node's neighbors also at a given rate. MMCA is powerful enough to model a full range of scenarios between these two process. Moreover, since the underlying mathematical grounds of MMCA are Markov chains, it can be applied independently if the network is weighted or unweighted.

To understand how this model works, assume that we have a network social contacts encoded in an adjacency matrix $\mathbf{A}$, where a node is connected to any other node for which there exists a contact that enables the transmission of the disease. These relationships can be weighted, encoding in w_{ij} the strength of the contact. To encode the process used to transmit the disease we consider λ as the number of trials that makes an infected individual to transmit the disease, the success of this trials is determined by the infection probability β. The fraction of endemic individuals ρ can be found when the process enters on a stable regime.

The probability that a node i is in contact with node j can be expressed in a matrix $\mathbf{R}$ where each entry is defined as

$$r_{ij} = 1 - \left(1 - \frac{w_{ij}}{w_i}\right)^{\lambda_i}, \tag{1.33}$$

where w_i is the total strength of node i. Notice that the tuning of λ_i is critical to determine how the epidemic process will spread. If $\lambda_i = 1$ for all nodes then, we recover the *contact process* since just one contact will be made by unit of time. Otherwise, if we set $\lambda_i = \infty$, we recover the *reactive process* since all the values of the matrix $\mathbf{R}$ different than 0 will be set to 1 regardless of the weights of the network; thus, the disease can spread to all the neighbors of node i with the same probability. All the values of λ_i between these two cases enable the generation of models that are interpolations between *reactive* and *contact* process.

We can obtain a system of equation that captures the SIS dynamics considering the probabilities that each node i is on an infected state p_i as

$$p_i(t+1) = (1 - q_i(t))(1 - p_i(t)) + (1 - \mu)p_i(t). \tag{1.34}$$

where q_i stands for the probability that a node i is not infected by any of its neighbors, which can be expressed as

$$q_i(t) = \prod_{j=1}^{N}(1 - \beta r_{ij} p_j(t)). \tag{1.35}$$

Equation (1.34) is very convenient since it enables to express the transition between the two stages straightforwardly. The first part stands for the probability that node i is susceptible and it gets infected by some of its neighbors. The second term encodes the probability that node i is infected at present and does not get espontaneously recovered. This system of equations can be solved at the stationary state:

$$p_i = (1 - q_i)(1 - p_i) + (1 - \mu)p_i, \tag{1.36}$$

where the trivial solution can be always found for $p_i = 0$ for all the nodes on the system. Non-trivial solutions can be computed by iteration.

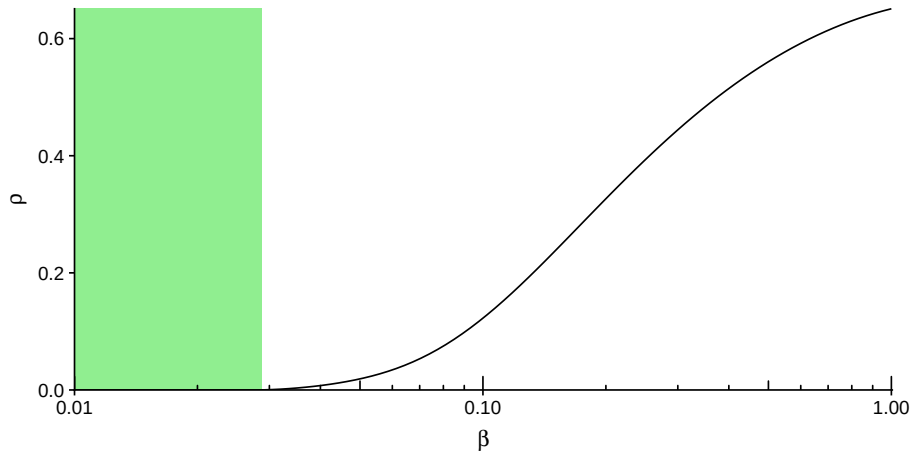


Figure 1.12: Fraction of endemic infected individuals ρ as function of infection probability *beta* using MMCA. The underlying network is a Barabasi-Albert of 10000 nodes with $\langle k \rangle = 4$. We have used a reaction process with a fixed value of $\mu = 0.5$. The green area represents a subcritical region where the epidemics survive computed using Equation (1.39).

The macroscopic order parameter ρ can be found as an average of p_i, $\rho = \frac{1}{N}\sum_i^N p_i$. We can see an illustration of how this parameter depends on the infection probability β in Figure 1.12.

As we commented on the other methods, the analytical computation of the epidemic threshold is important to understand the behavior of the system. We will start assuming that there exists a β_c such that, given a fixed μ and λ_i, $\rho > 0 \iff \beta > \beta_c$; otherwise, the system will have the trivial solution $\rho = 0$. To compute β_c we assume that when β is close to it, the probabilities are small, $p_i \approx \epsilon_i$, where $0 < \epsilon_i \ll 1$, then we can approximate Equation (1.35) to:

$$q_i \approx 1 - \beta \sum_{j=1}^{N} r_{ji}\epsilon_j \tag{1.37}$$

Using this equation on Equation (1.36) and neglecting second order terms in ϵ we obtain

$$\sum_{j=1}^{N} \left(r_{ji} - \frac{\mu}{\beta} \right) \epsilon_j = 0 \qquad \forall i = 1, \ldots, N \tag{1.38}$$

where δ_{ij} represents the Kronecker delta. The only non-trivial solutions of this equation can be found if and only if μ/β is an eigenvalue of $\mathbf{R}$. Since we are interested in the lowest value of β that satisfy the previous equation (we are looking for the onset of the epidemic) the epidemic threshold can be expressed as:

$$\beta_c = \frac{\mu}{\Lambda_{\max}(\mathbf{R})}, \tag{1.39}$$

where $\Lambda_{\max}(\mathbf{R})$ is the largest eigenvalue of the matrix $\mathbf{R}$.

1.3.2 *Relational Higher-Order*

From homogeneous interacting populations to the microscopical description of nodes behavior on complex networks, we have seen how increasing the order of complexity of the models can enhance our understanding of key aspects of epidemic processes. Regarding the last method, MMCA, having information about the probability of a specific node of being infected could be useful to understand which individuals are more prone to be affected by an infectious disease or to design and evaluate the impact of a vaccination campaign. However, we have made a critical assumption, we have considered that the probability of a node i of being infected is independent of the state of their neighbors. Maybe that is enough to model how the epidemic process affects individual nodes. Nevertheless, those entities are connected through relational edges; thus, those properties are also tied. Enabling the system to capture correlations between entities states could lead to a better understanding of how epidemic processes work. For instance, to spread through the network, an infectious disease uses connections between infected and non-infected individuals. Designing a relational higher-order model that accounts for the probability that this configuration happens at microscopic level will help to understand the possible contact paths that the infection will follow through its spreading.

In Chapter 3, we will expose a new method that enables us to better understand the paths of epidemics through the use of a higher-order approach. Using this new methodology, we can capture and study the microscopic epidemic properties of relations between entities. We will

apply this method to the problem of epidemic containment, showing the utility of having such epidemic descriptors.

1.4 UNDERSTANDING HUMAN MOBILITY

From home to supermarket to get food, from one city to the next one to go to work or, from one country to another to seek a better life, human mobility at any scale has shaped our world. When humans move from one place to another, we carry our language, our costumes, our technology and our vision of the world, spreading those things from one place to another. Understand human mobility, is to understand the history of the world: rise and fall of civilizations, the language spoken in different regions, or the economic development can be thought as the consequence of people movements and their eventual collision. Take for instance the case of America colonization, Europeans moving across the ocean determined the fate of millions of native Americans, expanding their dominion and religion in all the continent.

When people move from one place to another, other companions travel with us, pathogens. Human mobility plays a critical role in the way how epidemic outbreaks diffuse [9, 74, 127] and, as we said in the previous section, they are critical in our lives. Continuing with the example of the colonization of America, it is believed that the germs carried by the Spanish explorers, triggered a series of outbreaks of smallpox, influenza, measles, and typhus that virtually extinct the Taino [64, 88], indigenous people of the Caribbean, since they did not have antibodies for those diseases. Another paradigmatic example is the spreading of the *plage*. In the previous section, we have commented on the vast impact that this disease had on the peoples who lived in Europe on the 14^{th} century. It was believed that this disease travelled along the Silk Road from Asia to Europe, following the commercial route [51] and then it became endemic on the continent using European rats as a disease reservoir, causing a series of deathly outbreaks until the 18^{th} century. However, new studies [138] offer another explanation, proposing that the rodents infected on Europe at some point died out, and new outbreaks, triggered by a series of climatic fluctuations, arrived from Asia to Europe following the recurrent mobility patterns derived of the commercial trade, see Figure 1.13. Those recurrent patterns are

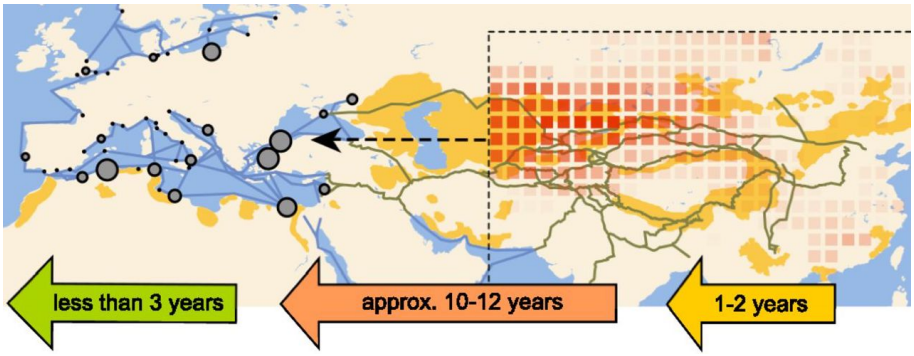


Figure 1.13: Schematic of the process of plague reintroductions into Europe. A new reintroduction of plague consists of three stages, namely, a rodent population crash (1–2 y) after the onset of the decline in climatic conditions, transport of the disease to Europe (10–12 y), and the spread of the disease through the European maritime trade network (<3 y). The average spatial extents of the nine climate fluctuations that were found to correlate with plague reintroductions in Europe are shown as red squares, showing the difference in the Monsoon Asia Drought Atlas (26) between the onset of the decline in the climate fluctuation and 2 y later. Also shown are the relevant trade routes (brown and blue lines), known wildlife plague reservoirs (orange), and the European harbors associated with new maritime introductions of plague (gray circles). *Reprinted by permission from [138].*

critical for the understanding of human mobility, as we will see on Chapter 4.

The scope of this section is to expose how scientists had tackled the problem of understanding human mobility, building models that are capable of capturing some essential mobility patterns.

1.4.1 *Mobility Models*

Mobility models are widely studied due to their essential implications on society: studying spatial economics [14], urban planning [16], location-based services [1] or, as we stated, understanding the spreading of diseases. However, obtaining data to build and validate models that enable us to make a hypothesis about human mobility, was not easy due to the lack of reliable data. Over the last years, thanks to the

development of modern technologies, researchers have turned their attention to investigate new ways to unveil human mobility using proxies like geolocalized phone calls, GPS traces or geolocalized tweets [9, 13, 63, 87, 140, 147]. Using this vast amount of data, scientists can build structures that store crucial components of human behavior.

One of these structures is the origin-destination (OD) matrix. If we tessellate the geographic area of study into N parts, we can store on the OD matrix T, the number of individuals that move from location i to location j into T_{ij}. The values encoded on this matrix are at the core of many transportation models [14]. We are especially interested in OD matrices since they can be understood as directed and weighted networks, and many of the tools used to analyze these structures can be used to study human mobility. This section aims to introduce some of the most popular models that can make use of the information stored on the OD matrices to build and validate mobility hypothesis.

1.4.1.1 *Gravity Model*

Humans usually move from one place to another for economic needs or social ties. Those travels have a short length since commonly, by constraints, we are more tied to nearby places: we work near home, we socialize with people of our surroundings. The farther we move from home, the more difficult is to find social connections and the more expensive to send goods. This observation was the seed behind using a *gravity law* inspired model to explain mobility [168].

The idea is to generate a theoretical OD matrix using

$$T_{ij} = K \frac{P_i P_j}{d_{ij}^{\sigma}}, \tag{1.40}$$

where P_i and P_j refers to the population of locations i and j, d_{ij} encodes the Euclidean distance, K is a constant that models the force of attraction, and σ is an exponent whose value depends on the system. The use of population as the *mass* of these model can be understood as an indirect indicator of socio-economical attributes that enhance the attractive power of a given place; the larger the population, the more substantial will be their attractiveness. These *masses* could be adapted

to represent different attributes, for instance, in an international trade scenario considering economic activity indicators makes the gravity approach suitable to model it [18].

A convenient way to represent the gravity model was proposed in [14] as

$$T_{ij} \sim P_i P_j f(d_{ij}) \,, \tag{1.41}$$

where the f encodes a function that describes the effect of space, which can vary depending on the scale, transportation mode or physical barriers in the data used to build it.

Using this approach, scientists have been able to find an agreement between empirical observations and gravity models on cases as diverse as cargo ship movements [78], highways trips [76], railway express journeys [168], global commuting [9, 158] or telecommunications [41].

Although the theoretical background is straightforward, the fitting and analysis of these models is not always easy since they are highly susceptible to:

- **Transportation modes:** different modes have associated different costs and travel times; they could be more constrained by natural barriers or, they could have different volume limitations. For instance, cars are suitable to travel short distances at a relatively small speed compared with high-speed trains that can travel faster, but they are only available for relatively distant destinations. Those transportation modes coexist, and their differences could make the modeling of local commuting a difficult task since they would affect our capacity to understand their specific contribution.

- **Discretization of the space:** choosing a proper scale when space is divided into different regions is a common problem on all the mobility models. If the scale is too large we will lose information about a large volume of local trips; if it is too small, we will introduce noise on the system since small perturbations would bias our model. Gravity models are not an exception, and relatively small changes in scale can have a huge effect on the parameters of the system [9].

- **Heterogenity of individuals:** the nature of the trip differs depending on the needs of the individuals. They can be commuting from home to work, business travels, migrations, etc. These different natures will have an impact on the duration of the trip, the distance or the associated costs. This heterogeneity will have important implications on the interpretation of the model parameters.

1.4.1.2 *Radiation Model*

If we talk about modeling the flow of commuters, there are some additional limitations to consider regarding the use of gravity model [140]:

- It is difficult to obtain the functional form of f (Equation (1.41)) since usually the parameters of gravity model are obtained using entropy maximization.
- Requires previous data to fit the model; thus, it is not able to generalize mobility behaviors to regions with a lack of empirical observations.
- There is an analytical inconsistency on Equation (1.41) since it predicts that if we increase the population on the destination j, P_j, the number of commuters will increase without limits, yet this number is bounded by the number of individuals in i, P_i.
- It is a deterministic method, meaning that it can not account for fluctuations in the number of travelers.
- Predictive discrepancies have been found when real commuting data is used to build the model, Figure 1.14a.

To overcome these issues, a model based on the daily commuting process is proposed in [140]. The authors assert that source and destination on commuting travels are determined by job selection. If we consider that a country is divided in counties, like in the US, then the process of job selection could be divided into two phases, Figure 1.14a, b.

On the first stage, individuals collect job offers for all counties. Each county has a number of jobs offers proportional to their resident population n, taking into account that there is one job offer for every n_{jobs}

individuals. The attractiveness of a job is determined by a combination of income, conditions, working hours, etc. To capture these attributes the authors propose to use values z randomly chosen from a distribution $p(z)$ for each n/n_{jobs} in each county.

On the second phase, individuals choose their jobs based on a straightforward rule: they choose the job whose benefits z are higher than the best offer available in their home county. This rule asses that commuting has priority over benefits since individuals will accept lesser jobs if they are close to their homes.

This process is applied proportionally to each individual in each country, assigning jobs in different locations to all of them. As a result of this process, daily commuting fluxes are established across the country.

This model has three unknown parameters: the job density n_{jobs}, the total number of commuters N_c, and the benefit distribution $p(z)$. However, the fluxes are independent to $p(z)$ and m_{jobs}, and N_c does not affect the flux distribution; thus the model is parameter-free.

Based on these premises, the authors propose a model that resembles the radiation of absorption process, that is why they call it the radiation model. The average flux from location i to location j, $\langle T_{ij} \rangle$, is computed as

$$\langle T_{ij} \rangle = T_i \frac{m_i, m_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}, \tag{1.42}$$

where m_i and m_j are the populations of i and j respectively, r_{ij} is the distance between them, and s_{ij} stands for the population that lives in a radius r_{ij} centered in i (excluding the populations on the source and destinations counties). Notice that this equation also takes into account the total number of commuters that start their journey from location i, $T_i = \sum_j T_{ij}$, which is proportional to the individuals living in i, that is the reason way N_c does not affect the flux distribution.

This model has some benefits compared with the gravitational model: it has a straightforward mathematical derivation, it is a parameter-free model, information about the surrounding populations encoded in s_{ij} makes the model robust to different population densities (see

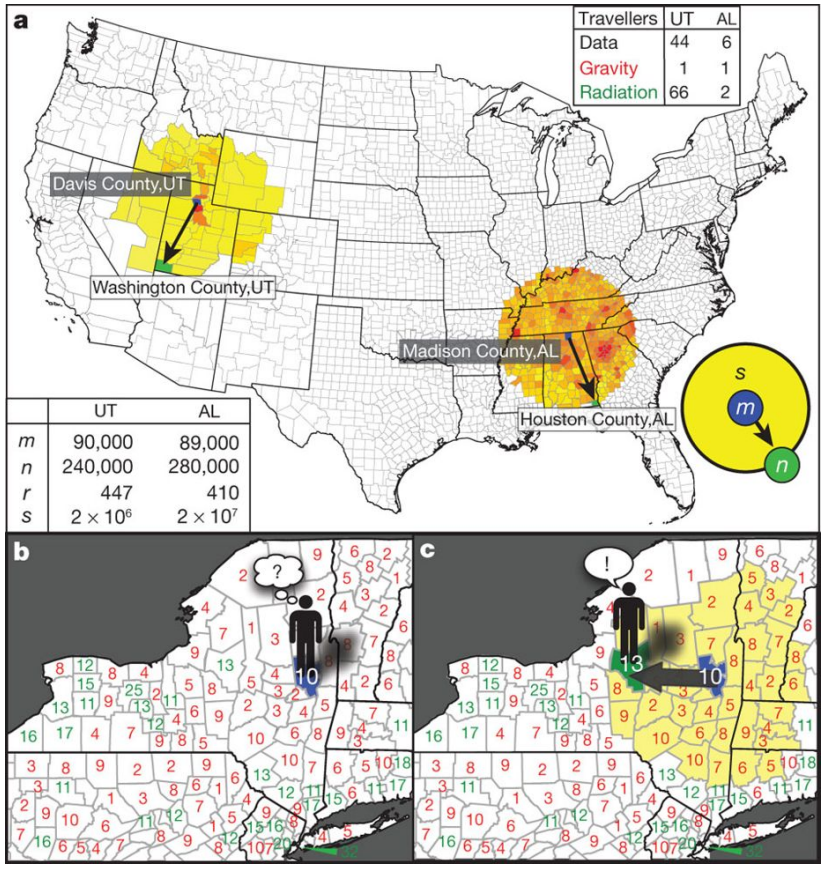


Figure 1.14: **(a)** logarithmicmitations of the gravity law we highlight two pairs of counties, one in Utah (UT) and the other in Alabama (AL), with similar origin (m, blue) and destination (n, green) populations and comparable distance r between them (see bottom left table). The gravity law predictions were obtained by fitting Equation (1.41) to the full commuting data set. The fluxes predicted by this equation are the same because the two counties pairs have similar m, n and r (top right table). Yet the US census 2000 reports a flux that is an order of magnitude greater between Utah counties. That difference is correctly captured by the radiation model. **(b)** The definition of the radiation model: an individual applies for jobs in all counties and collects potential employment offers. Each county is marked in green if its best offer is better (lower) than the best offer in the home county. **(c)** An individual accepts the closest job that offers better benefits than his home county. This process is repeated for each potential commuter, choosing new benefit variables in each case. *Reprinted by permission from [140].*

Figure 1.14) and, finally, it is a stochastic method and, since the variance can also be computed for T_{ij}, the method is robust to fluctuations.

However, this approach has an important issue: it just accounts for daily commuting routes between two locations, information about travels not related with commuting are not captured by this method. Another significant limitation is the heterogeneity on time spent at each location, the models capture fluxes between two locations, but not the amount of time that individuals will spend on those places.

1.4.1.3 *Markovian Models*

A standard approach to deal with mobility models of dynamics [17, 56, 86, 90] is to consider each node as a state of a Markov process. This process has been wildly studied: in chemistry, to study kinetic transitions; in statistics, for generating sequences of random numbers; in genetics, to describe the change in population frequencies affected by genetic drift; in computer science it is also used, PageRank is based on it; in physics, where they are used, among other things, to define the path integral formulation of quantum mechanics.

The power of Markov processes relies on their simplicity to capture probabilities of transition between different states from data. For instance, suppose that we have a large set of geolocalized phone calls, and we want to use them as a proxy for human mobility. To obtaining the flux between any pair of nodes, we consider the sequence of consecutive calls made by each individual. Then, we count the number of times the individual ℓ makes at least one call in location j after making at least one call in location i as $f_{ij}^{(\ell)}$. Then, the flux value is

$$F_{ij}^{(\ell)} = \frac{f_{ij}^{(\ell)}}{\sum_{k=1}^{n} f_{ik}^{(\ell)}}, \tag{1.43}$$

where n is the total number of locations.

It turns out that the matrix defined by $\mathbf{F}^{\ell}$ encodes a directed an weighted network that represents all the mobility patterns exhibited by the individual ℓ. This is very useful since we could use this matrix to feed

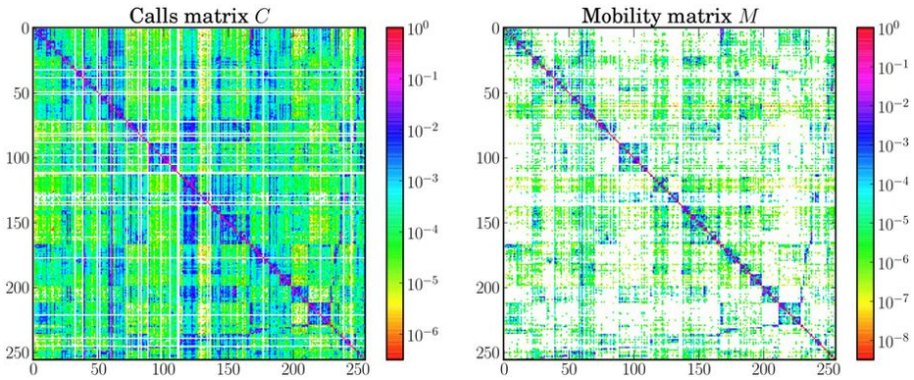


Figure 1.15: Logarithmic representation of the mobility matrix (left) and the call matrix (right). Null values are indicated using the white color. For both matrices, highest values are mostly concentrated along the diagonal, showing that communication and movement between sub-prefectures is highly uncommon. However, the calls matrix is visibly denser than the mobility matrix, confirming that phone contacts between different sub-prefectures are more usual than movement. *Reprinted under Creative Commons CC BY license from [86].*

a random walker and study some important characteristics regarding the coverage of the paths, i.e., which fraction of locations have been visited on a given time, which is the mean first passage time for each location, which nodes are the most central to this individual, etc. All this information gives us an idea of how diffusive is this individual.

If we want to generalize mobility for a given population, we use a similar approach to build the mobility matrix for each individual $\ell = 1, 2, ..., \mathcal{L}$ separately, and then we average over the whole set of mobility matrices, to obtain the transition probability of an average individual:

$$F_{ij} = \frac{1}{\mathcal{L}} \sum_{h=1}^{\mathcal{L}} F_{ij}^{(h)} \tag{1.44}$$

This matrix encodes the average mobility patterns found on the population. Notice that we did not impose a specific time window to calculate transitions, to avoid introducing biases and undesired effects due to the choice of the temporal range, and it is worth remarking that other normalizations can be considered depending on data and meta-

data availability [154]. Where not otherwise specified, we consider the mobility matrix obtained from the whole period of observation. This model is known as "first-order" (or 1-memory) because the present state is the only information required to choose the next state. Despite its simplicity and constrains, it has proved that it is capable of capturing mobility patterns in real data. Take for instance the example in Figure 1.15 where a mobility model has been built [86] from call detailed records (CDR) from people in Côte d'Ivoire. Data contains the sequence of calls made by individuals moving among different political regions, called sub-prefectures. It can be seen that both, the call matrix and the mobility matrix, show highest mostly concentrated along the diagonal, revealing that movements between sub-prefectures are highly uncommon. Nevertheless, the calls matrix is denser than the mobility matrix, showing that phone contacts between sub-prefectures are more usual than displacements.

1.4.2 *Temporal higher-order*

Markovian models are powerful tools but have a significant limitation: they can not account complex (more than one step) patterns without introducing mixing behaviors. This limitation comes from the way how they are built based on the Markov property [93], which asseses that the probability of future steps only depends upon the present state. However, those complex patterns are inherent to human mobility since they account for recurrent patterns: it is very likely that someone moves to a surrounding area to buy food on the supermarket and a few hours later he comes back to his home. This memory is an intrinsic property of human mobility and must be taken into account for realistic modeling of people movements between different geographic areas.

Temporal information, time that people spent on one place before moving to the next, is also important to understand the societies, if we realize that mobility has a profound effect on trade, technology, cultural spreading and also, as we commented at the beginning of this section, with the spreading of diseases. In fact, time is crucial on epidemics because as much time an infected individual spends on a particular location, more probable is that he transmit the infection to other individuals on the same spot. Being able of building models that

are capable of dealing with this complexity, both on the sequence of locations and time spent at each location, is mandatory to model this kind of scenarios.

"First-Order" Markovian models can capture time with an important constraint: the amount of time that an individual spends on a given location is independent regarding they previous location. This assumption cannot always be applied to human mobility since we can find many examples with heterogeneous waiting times before a move to another location is registered. For instance, consider holiday trips. Individuals making expensive intercontinental trips tend to spend more time visiting the destination than individuals making cheaper trips, achieving a good trade-off between the travel cost and the time spent. To tackle this issue we need to include more temporal information on our problem, thus increasing its order of complexity.

Chapter 4 will provide a detailed explanation of some techniques to overcome these issues. We will focus on a new method to deal with temporal higher-order patterns, studying the effect of these patterns on epidemic spreading processes. We will validate our approach using mobile call detail records (CDR) as a proxy to build our models.

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

2

STRUCTURAL HIGHER-ORDER: COOPERATION IN MULTIPLEX NETWORKS

2.1 INTRODUCTION

Networks are relational structures used to relate pairs of entities. The relation between these entities can be of any kind: biological, social, structural, spatial, etc. They can relate a few hundred entities, like on the neural network of the *C. elegans* worm [77], or millions of entities, like in many social networks. They can be sparse or dense, random or structured, connected or disconnected, etc. Nevertheless, regardless of their differences, there is something in common among all the networks, they are composed by nodes and links relating them, the way how we mathematically represent them does not change.

This framework holds while the definition of *relation* between two entities is common among all the other relations, thus, the meaning of the link has to be shared. For instance, if we want to build a social network, we can consider that a link exists between two persons if they know each other. The meaning of the link is general enough to consider many social networks; however, we have aggregated a large number of social ties under the same category. For example, a link will exist between two individuals disregarding of the way how these persons knew each other. It could be because they are co-workers, family members, clients of the same company or roommates.

In some scenarios, the aggregation of all these different relations under a general relational concept could be good enough, but in others, it is

crucial to use a high-order structure that enables us to represent all these different relational aspects under the same mathematical structure. If we want to represent a city transportation system it is essential that the structure that we use is capable of capturing the multi-modal nature of the connection between two locations: it is a bus line, a metro line or a train line. Using this structure, we can better understand how people move on a city or how the failure of one transportation system will affect the others.

As we have depicted in the previous chapter, the use of this structural higher-order using multilayer networks has a significant impact on how we model cooperative scenarios. Multiplex networks are interesting structures to try to understand the evolution of cooperative behaviors, because many social interactions entail a combination of interactions at different, independent levels, each one representing a different social scenario such as family, friends, coworkers, etc. An individual's behavior can be different at each level, but all of them ultimately condition it. This chapter aims to explain the different phenomenology derived from the use of multiplex networks as the higher-order structure of social interactions.

2.2 EVOLUTIONARY GAME THEORY ON MULTILAYER NETWORKS

Some previous work was devoted to understand the evolution of the Prisoner's Dilemma game on multiplex networks [59], or exploring different social dilemmas using a 2-layer network, where one layer was used for the accumulation of payoffs and the other for strategy updating, with emphasis on the degree correlations among layers [161]. There were also works that explored the problem of cooperation on coupled networks [160]. However, an in-deep analysis of how evolutionary dynamics works at each region of the $T - S$ plane is required to fully understand the implications of multilayer structures on game-theoretical frameworks, see Section 1.2.2 for further details.

To achieve this task, we propose a model [97] were the players sit on the nodes of a multiplex network of L layers. Each node is present in all layers, but in general, they have different connectivity in each layer. Every layer, $\ell \in L$, in the multiplex network is a connected and

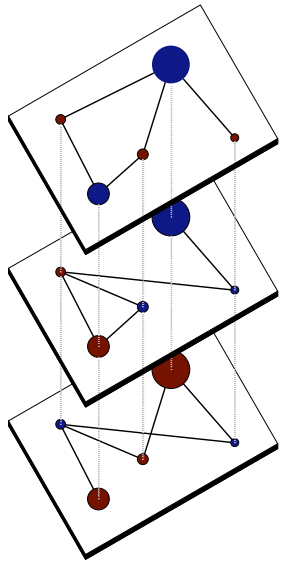


Figure 2.1: Example of a multiplex network with 3 layers, 5 nodes per layer and 5 links in each layer. The color of the nodes represents the strategy played in that layer (red for cooperators, blue for defectors. Their size is proportional to their global payoff.

homogeneous Erdős-Rényi (ER) network, with the same number of edges E and nodes N, and equal degree distribution, and the multiplex network is generated avoiding degree correlations between layers. Each layer is represented by an adjacency matrix A^ℓ, where $A^\ell_{ij} = 1$ if nodes i and j are connected in that layer, and $A^\ell_{ij} = 0$ otherwise. That representation enables the definition of the degree of node i in layer ℓ as $k^\ell_i = \sum_{j=1}^{N} A^\ell_{ij}$ and its global degree in the multiplex as $K_i = \sum_{\ell=1}^{L} k^\ell_i$.

Each round of the game is divided in two phases: payoff recollection and strategy update. Each node i can choose to play one of the two strategies, cooperation or defection, independently in each layer of the network and at every time step; we denote these states by $s^\ell_i(t)$, this value is set to 1 if the player i is cooperating in layer ℓ at time t and 0 if it is defecting. Within a specific payoff matrix, the node i's strategy determines the payoff, p^ℓ_i, that it obtains in a layer ℓ when it plays against all its k^ℓ_i neighbors. The total payoff of node i can be easily calculated as $P_i = \sum_{\ell=0}^{L} p^\ell_i$. At the end of each round, each player can

change the strategy in one of its layers, s_i^ℓ, using the Replicator-like rule: A node chooses a layer of the multiplex, ℓ', with uniform probability. Then it chooses with uniform probability one of its $k_i^{\ell'}$ neighbors, j', in that layer. If $P_i < P_{j'}$ and $s_i^\ell \neq s_{j'}^\ell$ the probability that node i changes its strategy in layer ℓ' is given by:

$$\prod_{i \to j'}^{\ell'}(t) = \frac{P_{j'}(t) - P_i(t)}{max(K_i, K_{j'}) \cdot (max(1, T) - min(0, S))} \tag{2.1}$$

It is important to notice that the update rule uses global information about the players: global degree and global payoff (that is, added up over all layers), to update the strategy of any particular layer. That is the way our model shares information between layers and relies in the social nature of layers' interdependency [59]: each player only has information about the strategy of its neighbour in their same layer (but not in those layers where they are not connected). However, it knows its neighbor's total benefits, and it makes the simplifying assumption that it is using the same strategy in every layer. As we will see later on, this fact has a profound impact on the outcomes of the dynamics, compare to the monoplex[1] scenario.

At the end of each time step the density of cooperators can be computed for each layer and for the entire multiplex using:

$$c(t) = \frac{1}{L}\sum_{\ell=1}^{L} c^\ell(t) = \frac{1}{L \cdot N}\sum_{\ell=1}^{L}\sum_{i=1}^{N} s_i^\ell(t)\,, \tag{2.2}$$

where $s_i^\ell(t)$ represents the state of the node i in layer ℓ and

To ascertain the outcome of the cooperative dynamics for the different games on multiplex networks, we will start by studying the stationary level of cooperation in the system, then we will study the effect of the initial fraction of cooperators, and finally, we will move to analyzing in detail the microscopic organization of cooperation for individuals across different layers.

1 We will use this term to refer to single-layer networks since we will use multiplex networks as multi-layer network configurations.

The results are obtained for a range of values of $T \in [0,2]$ and $S \in [-1,1]$ that defines the $T-S$ plane. The simulation runs on a multiplex network that has $N = 1000$ nodes and $E = 3000$ edges per layer distributed according an Erdős-Rényi degree distribution with $\langle k \rangle = 3$. For each possible pair of values of the game parameters the simulation runs 1×10^5 time steps, that is the transient time t_0 needed by the algorithm to generally reach a stationary state (we further discuss the matter of convergence time in Sections 2.4 and 2.5). After this time, the algorithm runs for another $t_\gamma = 2 \times 10^4$ time steps. All the quantities of interest are averaged over this second period of time. The experiments are repeated and averaged over $R = 64$ different networks and initializations to gain statistical confidence. The initial fraction of cooperators, c_0, is distributed randomly in each layer.

2.3 AVERAGE DENSITY OF COOPERATORS

One important aspect to consider on the analysis of evolutionary games is the stationary average value of cooperation that is defined according to the following equation:

$$\langle c \rangle = \frac{1}{t_\gamma \cdot R} \cdot \sum_{r=1}^{R} \sum_{t=t_0}^{t_0+t_\gamma} c(t)_r \tag{2.3}$$

where $c(t)_r$ is the is the value of $c(t)$ on the r simulation repetition.

In Figure 2.2 we present the average, stationary value of cooperation considering three different initial fractions of cooperators c_0 displayed at each row of the figure. We observe that our results for the monoplex case (left column) are consistent with those obtained by Roca et al. [128] for this kind of networks. The previously unreported results for multiplex networks show a large increase of the areas where both strategies coexist (that is, the areas in the plane that separate total cooperation from total defection), effect that is moderately enhanced increasing c_0. However, this coexistence is of a different nature depending on the evolutionary stability of the particular game (or quadrant).

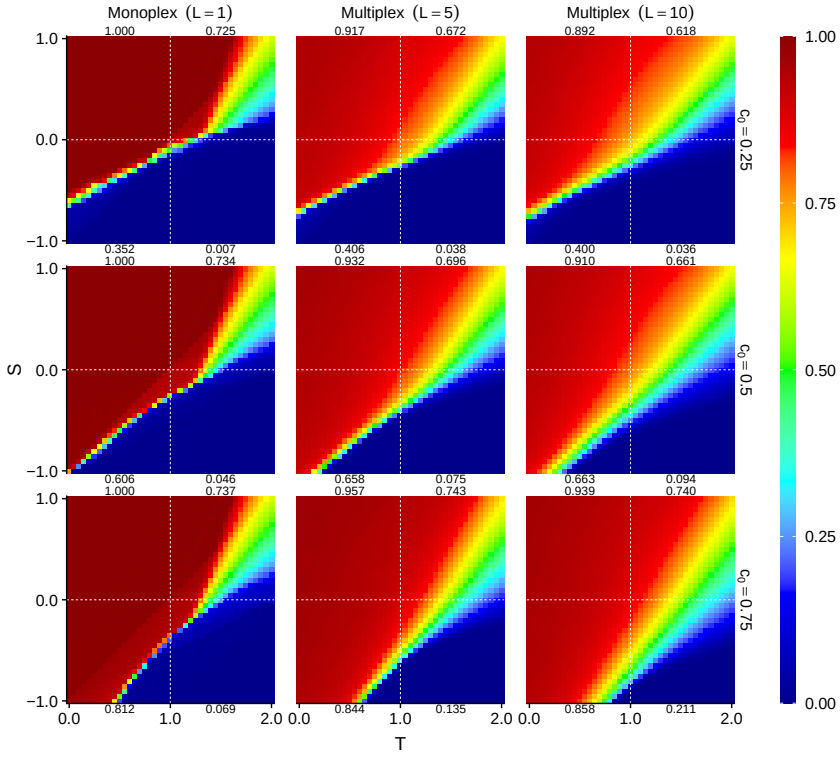


Figure 2.2: Asymptotic density of cooperators $\langle c \rangle$ for networks with different number of layers ($L = 1$ in the left column, $L = 5$ in the central column, $L = 10$ in the right column), and different initial fraction of cooperation ($c_0 = 0.25$ in the top row, $c_0 = 0.5$ in the central row, $c_0 = 0.75$ in the bottom row). The $T - S$ plane is divided into four major regions that correspond to the four games under study: the upper-left area is the Harmony Game, the upper-right is the Snow Drift, Stag-Hunt is in the lower-left, and the Prisoner's Dilemma is in the lower-right. The average asymptotic density of cooperators for each one of the games is also indicated, as a numerical value, next to the corresponding quadrant.

The Stag Hunt game has an unstable evolutionary equilibrium with mixed populations. This means that, when there is a structure, the population will evolve towards total cooperation or total defection depending on the initial population and type of structure of the network. For the monoplex we have a very narrow transition area between total cooperation or defection populations (left panel in Figure 2.2). This transition region widens with the number of layers, enabling the co-

existent of both strategies in a larger portion of the game parameter space. The explanation of such behavior can be found in the inter-layer dynamics: it is more likely that a cooperator or a defector node resists in hostile environments in a particular layer, because its fitness is not evaluated in just that layer, but also in the other layers where, due to its strategy or its topological configuration, the node might have better performance. The Stag Hunt game, where the maximum payoff possible is obtained when a cooperator plays against another cooperator, favors specially the resilience of cooperators nodes when the temptation value is low: a cooperator node i in layer ℓ' that has a big payoff P_i has higher probability of spreading its strategy to its defector neighbours in ℓ', thus increasing its payoff. This increase will propagate to the other layers, making the strategies of the player more robust against invasion. Playing defection in layer ℓ' when temptation value is small, does not have a big effect in the global payoff of the node. As a consequence, in this particular game the multiplex structure increases specially the resilience of cooperators, thus the first of all we have to define of cooperators in this game quadrant shows an statistically significant increase as we keep adding layers to the structure (Mann-Whitnet U test, $\alpha = 0.05$). Since this game has a heavy dependency on the initial conditions, an increase on c_0 has a direct impact in the outcome regardless of the number of layers.

In the Prisoner's dilemma game, defection dominates cooperation. Related papers [128] show that for ER networks using Replicator rule when temptation and sucker payoffs are not too large, cooperation can survive forming groups of cooperative clusters, thus resisting against the initial attempt of invasion by defectors, and then spread through the population. Our results for the monoplex are consistent with those. For the multiplex, we observe how the transition region between all-cooperator and all-defector situations is larger than for the monoplex, as in the case of Stag Hunt game. It is worth noticing that regions where we have all-cooperator populations in the monoplex, are not necessarily all-cooperator areas in its multiplex counterpart. This happens because the formation of cooperative clusters in one layer will also increase the fitness of these nodes in the other layers regardless of the strategy used in these other layers. And this can lead to a reinforcement of defector strategies due to the inter-layer dynamics, increasing their survival

rate. This inter-layer dynamics will led to a widening of the transition area that enables survival of cooperators in areas where they are not present in the monoplex scenario. If we take into account the whole Prisoner's Dilemma quadrant, the conclusions are the same that in the Stag Hunt game: a statistically significant increase in the average density of cooperators occurs as we increase the number of layers.

The Snow Drift game has a stable equilibrium in mixed populations: it is an anti-coordination game. Previous works [128] show that for ER networks there are some regions in the $T - S$ plane for which this game converges to single-strategy populations. For lower values of the temptation these regions are prone to cooperation. In multiplex networks however, single strategy regions are less common and mixed populations are the rule. That happens by the same inter-layer dynamics that we have explained earlier: the impact of a cooperator's benefits on the other layers of the multiplex structure. This entails a significant reduction on the average fraction of cooperators from 0.734 in the monoplex to 0.661 in the 10-layer multiplex for this quadrant.

Finally, the Harmony game has cooperation as its dominant strategy. For single-layer ER networks with Replicator update rule, Roca et al. [128] reported that the whole quadrant ends up in an all-cooperator configuration. However, in the case of multiplex scenarios, the average fraction of cooperators decreases significantly as we keep adding layers to the system: 0.932 for $L = 5$ and 0.910 for $L = 10$. This increasing resilience of defection can be explained as a consequence of the multiplex topology and the lack of degree correlations between layers: due to the payoff accumulated by an individual acting as cooperator in some layers, defector nodes can resist against cooperators in other layers.

2.3.1 *Defectors survival on Harmony Game*

We can mathematically prove that defectors can survive and be stable in the Harmony game on ER multiplex networks by analyzing the simplest situation: let's assume a multiplex structure with L layers. In one single layer (for simplicity we assume it will be the first one) we have one single node playing as defector, but it plays as cooperator

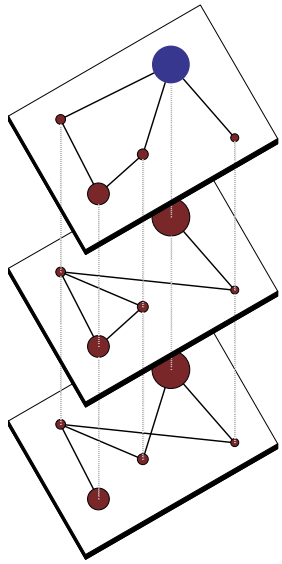


Figure 2.3: An example of a defector that is able to survive in a hostile environment due to the multiplex interlayer dynamics.

in all the other $L-1$ layers. There are no more defectors anywhere in the system. This node's connectivity in layer α is k_α, and, recalling that $R=1$ and $P=0$ (see Section 1.2.2), the total payoff of that node that is defecting in one single layer is given by:

$$P_d = Tk_1 + \sum_{\alpha=2}^{L} k_\alpha \tag{2.4}$$

The payoff of any of the node's neighbors (note that all of them play as cooperators), with a degree k'_α in layer α, is:

$$P_c = (k'_1 - 1) + S + \sum_{\alpha=2}^{L} k'_\alpha = \sum_{j=1}^{L} k'_\alpha + S - 1 \tag{2.5}$$

Thus, to survive as a defector in layer α, the following inequality must be fulfilled for each of the node's neighbours:

$$P_d \geq P_c \tag{2.6}$$

$$Tk_1 + \sum_{i=2}^{L} k_i \geq \sum_{j=1}^{L} k'_j + S - 1 \tag{2.7}$$

We can estimate both a soft and a hard limit for the previous inequality. As a soft limit, and assuming we have independent, uncorrelated Erdős-Rényi layers in our multiplex network, we can approximate every k_α by $\langle k \rangle$ and get:

$$(T + L - 1)\langle k \rangle \geq S - 1 + L\langle k \rangle \tag{2.8}$$

$$\langle k \rangle \geq \frac{S-1}{T-1} \tag{2.9}$$

On the one hand, a hard limit for the condition can be calculated by approximating k_i by k_{max} for the cooperator neighbours:

$$(T + L - 1)\langle k \rangle \geq S - 1 + L\langle k_{max} \rangle \tag{2.10}$$

$$\langle k \rangle \geq \frac{S - 1 + L\langle k_{max} \rangle}{T + L - 1} \tag{2.11}$$

On the other hand, we can calculate the probability of this topological situation happening. First of all, we have to define what is the probability of a node i to have degree k, $P(X = k)$. In our model, and to avoid the non-negligible effect of unconnected nodes, we impose a minimum connectivity, k_{min}. To get a more accurate approximation of our degree distribution we take into account this minimum:

$$P_{k_{min}}(X = k) = \frac{P(X = k)}{1 - P(X \leq k_{min})} \tag{2.12}$$

As it has been stated previously, the payoff of cooperators against cooperators is proportional to their degree, since we set $R = 1$: in this example we use $L - 1$ full cooperative layers, so the payoff obtained in this layers is proportional to the degree distribution of the aggregate network of this $L - 1$ layers. Moreover, the payoff distribution of the nodes

that play cooperation in all layers is proportional to the aggregation of all layers, L. Imposing that we do not have inter-layer degree correlation, the degree distribution of the aggregated networks can be modeled using the convolution of the single layer degree distributions.

$$P_L \sim P_{k_{min}} * \overbrace{\cdots}^{L} * P_{k_{min}} \tag{2.13}$$

The probability that a topological configuration that enables the fulfilment of the payoff conditions specified by Equation (2.7) exists, is given by:

$$\begin{aligned} P_{survival} \quad = \quad & \sum_{k_1=k_{min}}^{\infty} P_{k_{min}}(X = k_1) \cdot \\ \cdot \quad & \sum_{q=k_{min}\cdot L}^{\infty} P_{L-1}(X = q) \cdot P_L(X \leq \lfloor q + k_1 \cdot T - S + 1 \rfloor)^{k_1} \end{aligned} \tag{2.14}$$

where q is the payoff obtained by the defector node playing as a cooperator in $L-1$ layers. With that information, an upper bound for the aggregated degree of the defector's neighbours can be defined as $\lfloor q + k_1 \cdot T - S + 1 \rfloor$, and if all the neighbours have an aggregated degree below this upper bound, the defector can survive. It is worth noticing that the upper bound for the degree of a cooperator is a discretization of payoff values that involve S and T. This means that the survival probability of a defector only changes when the relation between S and T changes by an amount large enough.

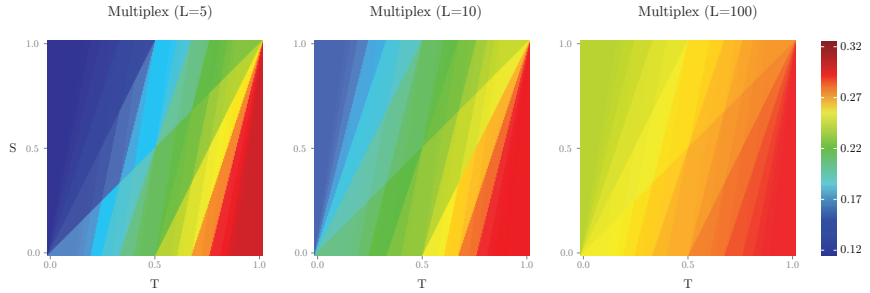
The expression for the degree distribution probability function is for an Erdős-Rényi network, assuming that we have a restriction for the minimum degree, so the degree distribution follows a Poisson distribution given by:

$$P_{k_{min}}(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k! \cdot (1 - e^{-\lambda} \cdot \sum_{i=0}^{\lfloor k_{min} \rfloor} \frac{\lambda^i}{i!})} \tag{2.15}$$

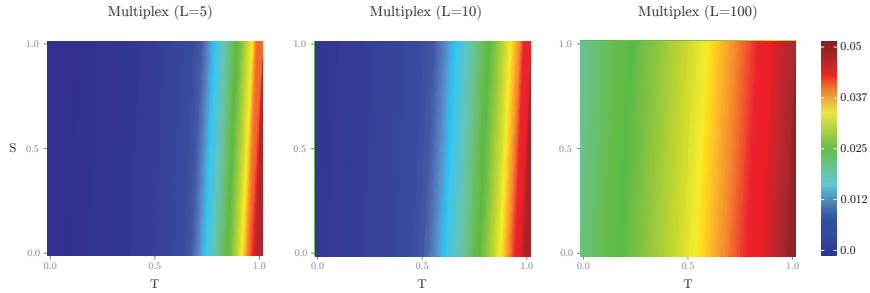
In Figure 2.4, we show the probability of a defector surviving in a full-cooperative population, calculated numerically using Equations (2.15) and (2.15). We observe that this probability increases naturally with T, because this is the payoff that a defector obtains against a cooperator, but it is only slightly dependent of the payoff of a cooperator against a defector, S. The number of layers has a huge impact on this probability: as the number of layers increases, the probability becomes more uniform in the $T - S$ plane, increasing in general. This can be explained by the relative contribution to the accumulated payoff that comes from layer 1 (the layer where the defector survives): the more layers are added to the system, the smaller this relative contribution. For a large number of layers, this implies that the values of S and T (that determine the payoff) are less important in the probability of a defector persisting in the system. For networks with a higher mean degree (Figure 2.4b), however, the chances of a defector surviving are lower: if the number of neighbours of the defector node is higher, then the probability that one of them has more payoff than him is also higher, thus the defector will tend to imitate the neighbour's behavior (or in other words, his chances of survival will decrease).

2.3.2 *Coherent and Incoherent players*

Prompted by the topological configurations described earlier, we can now define a "coherent cooperator" as a node that, at a given instant of time, plays as cooperator in all L layers of the system. Similarly, we can define a "coherent defector" as a node that, at a given instant of time, plays as defector in all L layers of the system. Finally, those individuals that are neither coherent cooperators nor coherent defectors will be called "incoherent" individuals. This new terms introduced here should not be mistaken for the concepts "pure cooperators", "pure defectors" and "fluctuating individuals" introduced in [57], which implied a *temporal* consistency of the agents' strategies. Also, we want to stress that a incoherent individual as defined here, is clearly different from the concept of a mixed population, that refers simply to a set of both strategies, coexisting together in a population. Moreover, we have to take into account that a coherent behavior is not trivial nor easily reachable, due to the fact that our simulations start with all mixed



(a) $\langle k \rangle = 3$



(b) $\langle k \rangle = 20$

Figure 2.4: Probability of a defector surviving in the Harmony Game for 5 layers (left), 10 layers (middle) and 100 layers (right), calculated according to Equations (2.15) and (2.15). The individual layers are ER with $k_{\min} = 1$, and $\langle k \rangle = 3$ (a) and $\langle k \rangle = 20$ (b).

populations (randomly distributed and uncorrelated strategies in all layers), so the dynamics that leads to coherence is specially interesting to study.

In Figure 2.5 we show the fraction of coherent cooperators (left column), coherent defectors (middle column) and incoherent individuals (right column) for 5 layers (top row) and 10 layers (bottom row). The formation of coherent cooperators is particularly complicated, and it is interesting to notice that even in the Harmony game there is a low fraction of them (except for a small area around the extreme case of $T = 0$ and $S = 1$). In the other quadrants, the fraction is very small (in particular, the Prisoner's Dilemma presents basically no coherent cooperation). This implies that most of the cooperation shown by the system comes from incoherent individuals. We also observe that the fraction of coherent cooperators decreases quickly with the number of

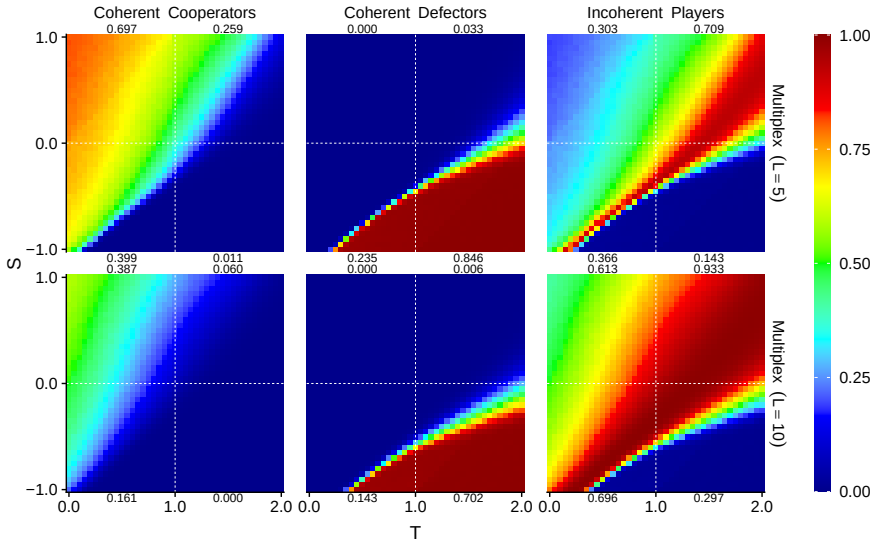


Figure 2.5: Average density of coherent cooperators (left column), coherent defectors (middle column) and incoherent individuals (right column) for networks with 5 layers (top row) and 10 layers (bottom row). The average density of the corresponding type of individuals is also provided for each one of the quadrants (upper-left is the Harmony Game, upper-right is the Snow Drift, Stag-Hunt is the lower-left, and the Prisoner's Dilemma in the lower-right). The initial fraction of cooperators, c_0 was set to 0.5.

layers for any game. As we have said, the origin of such results resides in the fact that a defector takes advantage of its own cooperative behavior in other layers, specially in regions of the $T - S$ plane prone to cooperation.

Conversely, regarding the fraction of coherent defectors, we observe that their presence is very strong in most of the Prisoner's Dilemma region and part of the Stag-Hunt area, and they decrease only slightly when increasing the number of layers from 5 to 10. This fact is easy to understand: the resilience of a cooperator in a hostile environment is based basically in how he performs as cooperator, the advantage of playing as defector in other layers is practically zero because in a large defector population the contribution to the payoff of a defector that plays against a defector is zero, $P = 0$. Thus, in these regions, the

survival rate of cooperation does not improve by playing as defector strategy in other layers.

Regarding incoherent individuals, we observe that they are very prevalent for all games (except for the extreme area of Harmony around $(T = 0, S = 1)$, where cooperation is very profitable, and the bottom-half area of the hard Prisoner's Dilemma where cooperation is extremely expensive). Incoherent individuals contribute significantly to the average density of cooperation in a large central area of the $T - S$ plane, particularly in the areas that separate full-cooperation from full-defection (see also Figure 2.6 for a detailed description of the fraction of incoherent individuals playing as cooperators). This area of prevalent incoherent individuals increases with the number of layers or, in other words, it gets harder and harder to be a coherent strategist as the number of layers increases.

Figure 2.5 also confirms what we showed analytically earlier: defection can survive in the Harmony game, as long as the individual defecting in a particular layer is a incoherent individual; it plays as cooperator in other layers and obtains enough payoff from them to avoid having to switch strategies (see also Figure 2.6 for further detail on the payoff of cooperators and defectors).

Interestingly enough, in Figure 2.5 we can observe how coherent players of opposite types do not coexist in the same population. Another important point is where coherent players can coexist with incoherent players. The area where coherent cooperators interact with incoherent players is wide and gets wider as we keep adding layers to the multiplex. However, the area of coexistence of coherent defectors and incoherent players is very narrowed and is only slightly affected when layers are added to the structure. This means that the coherent defection is a very dominant strategy that almost forbids the existence of any other kind of players.

2.4 CONVERGENCE TO EQUILIBRIUM

Since we are studying complex dynamics that could have a wild range of behaviors before reaching an equilibrium, a proper study of convergence times to stationarity seems appropiate. It is well know that the

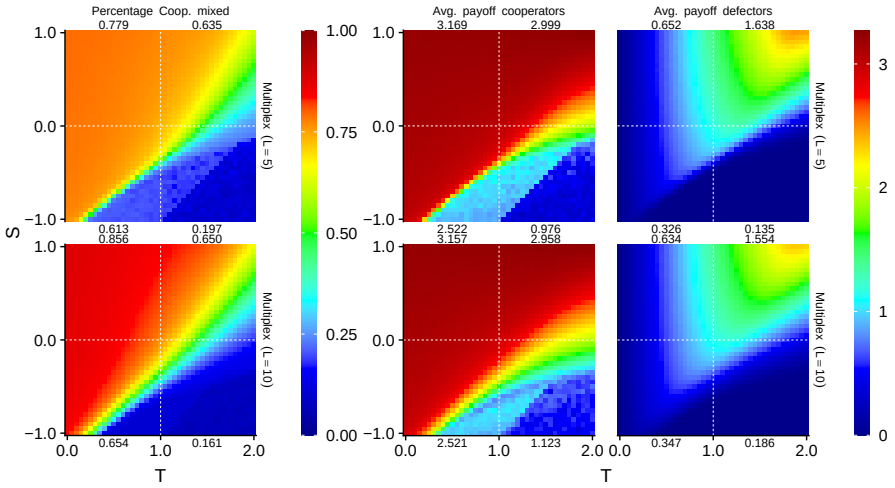


Figure 2.6: Percentage of cooperation in mixed individuals (left column), average direct payoff obtained playing as cooperator (middle column) and average direct payoff obtained playing as defector (left column) for 5 layers (top row) and 10 layers (bottom row) multiplex.

time evolution of cooperation on a monoplex, for a value of the parameters that allows the survival of at least some cooperation, usually follows a curve that initially decreases a bit, while cooperation rearranges itself from the random initial conditions into a more favorable setting, and then there is a new increase, followed by the achievement of the stationary state. In general, this whole transient time is relatively short (typically of $1 - 2 \times 10^4$). However, it hasn't been explored in detail until now the convergence process for the four games in the $T - S$ plane on multiplex networks.

To study such convergence we fit the last t_γ time steps to a linear model, $\widehat{c(t)} = \alpha + \beta t$ using QR decomposition method. Then we use the slope of the fitted model to compute the variation of the density of cooperators every 1000 time steps, $\Xi = 1000 \cdot \beta$. Thus, a near-zero value of this metrics indicates that the system has reached the stationary state, while a positive value would indicate that the average level of cooperation is still increasing in the system at that time, and vice versa. Figure 2.7 shows how every point of the $T - S$ plane performs on our measure of convergence during the last t_γ time steps of the simulation.

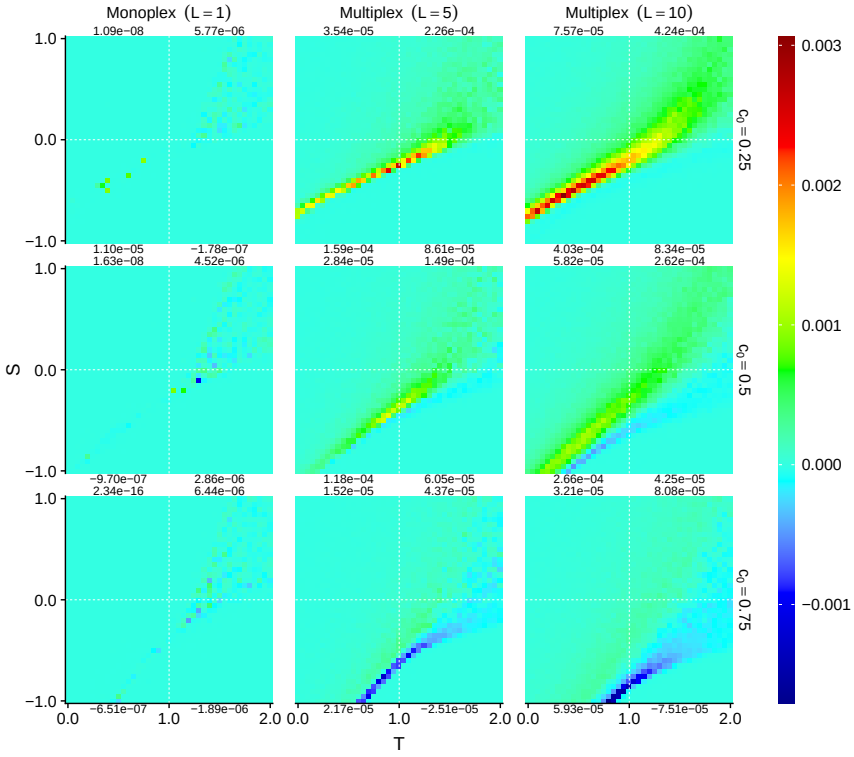


Figure 2.7: Convergence to the stationary point measured as the variation in the fraction of cooperators during 1000 time steps, measured using the slope of a linear model fitted at the end of t_γ steps of the simulation. The numbers in each quadrant represent the mean convergence value for each one of the four games. In rows we have the information for several values of initial fraction of cooperators ($c_0 = 0.25$, $c_0 = 0.50$, $c_0 = 0.75$).

Monoplex networks seem to reach the stationary state according to our convergence criteria for every point of the plane $T - S$ and independently of the initial fraction of cooperators. We observe, however, a small amount of stochastic noise for some regions of Snow Drift and Prisoner's Dilemma games, where our measure indicates that the stationary is not fully reached. Nonetheless, as we will discuss later, this noise is just an effect of the large fluctuations in the number of cooperators when the stationary is reached (See Figure 2.9).

In multiplex networks, on the other hand, there is a large area where convergence is not reached (red areas in the central and right panels

of Figure 2.7). In the most extreme cases, where the slope of the linear model β is largest, our measurements indicates an increment of the cooperators of even a 0.1% every 1000 time steps. That could seem a little increase in the fraction of cooperators, however if the evolutionary process runs over a large period of time the increase could be very significant.

To better illustrate this difference in the path to stability for monoplex vs. multiplex, we show in Figure 2.8 the time evolution of the level of cooperation, $\langle c \rangle$ as the simulation progresses (monoplex plotted in red and two multiplex networks with different number of layers represented in green and blue), for one point in the $T - S$ plane. This particular point has been picked as an extreme case, because it presents the maximum fluctuation values found in the entire plane (see Figure 2.9). We clearly observe that, while the time require for the monoplex system to achieve the stationary state is around $1 - 2 \times 10^4$, for the multiplex networks it can be at least one order of magnitude larger, and it increases with the number of layers, too. However, it is important to remember that this example shown here is a very extreme case, while the convergence process in multiplex is in general faster for regions of the plane that are far away from the transition area (that is, for areas where the final state is close to an all-cooperation or all-defection).

To understand the reason for such an increase in the convergence time for multiplex with respect to monoplex, one has to pay attention to which areas are more reluctant to reach stability. Such regions correspond again to the transition areas between those of total cooperation and total defection. In the Stag Hunt quadrant, the game has an unstable equilibrium with mixed population, that means that the game will tend to converge to total cooperation or total defection as happens in the monoplex network. However the multiplex structure of the network changes that behavior, as we commented before. In these structures, the transition region is larger than in the monoplex, and is in this transition region where the convergence is hard. The analysis of how the fraction of cooperators has an effect on the convergence gives us an insight about what is happening. We have already told that the interlayer dynamics has an important role in the survival rate of defectors and cooperators. If we look at the multiplex columns of the

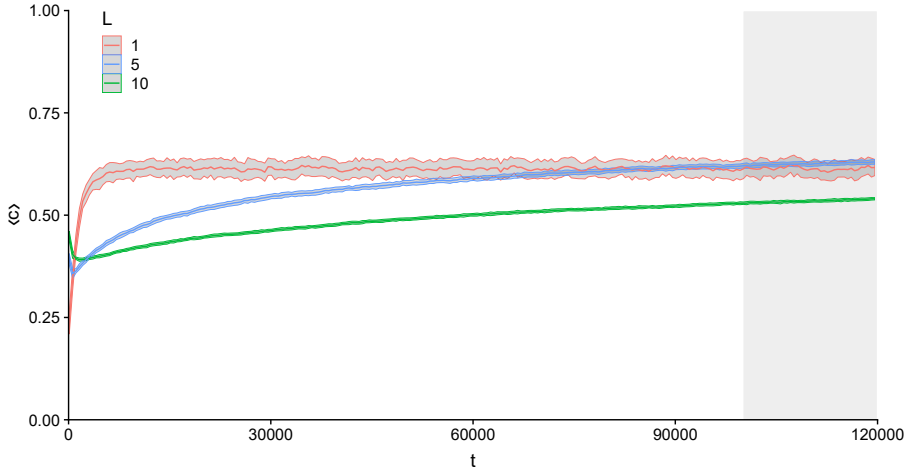


Figure 2.8: Example of the evolution of cooperation for the point in the $T - S$ plane with maximum fluctuations ($S = 0$ and $T = 1.35$). The shadowed area for each plot represents two standard deviation over the residuals the R iterations at each time step. It will be used to compare the size of the fluctuations between monoplex and multiplex networks. The grey vertical area corresponds to the interval $[t_0, t_0 + t_\gamma]$ where all the measures shown in the panel figures are computed.

Figure 2.7, we can observe how the convergence is strongly affected by the initial fraction of cooperators. On one hand, if the fraction of cooperators is small, they will need more time to reach their equilibrium because the have to fight against a bigger fraction of defectors that benefits of the interlayer dynamics. On the other hand, a bigger initial number of cooperators implies that the defectors will need more time to reach a stable point. However, the presence of a large number of initial cooperators has less impact on the convergence, a perfectly normal conduct giving that defectors get more profit of cooperating that the opposite case.

Similar conclusions could be obtained for the other games, taking into account that the transition regions between full-cooperation and full-defection are different in nature. For instance, in the Snow Drift game, this region is open so we can see how the effect of non-convergence is diluted across the SD quadrant.

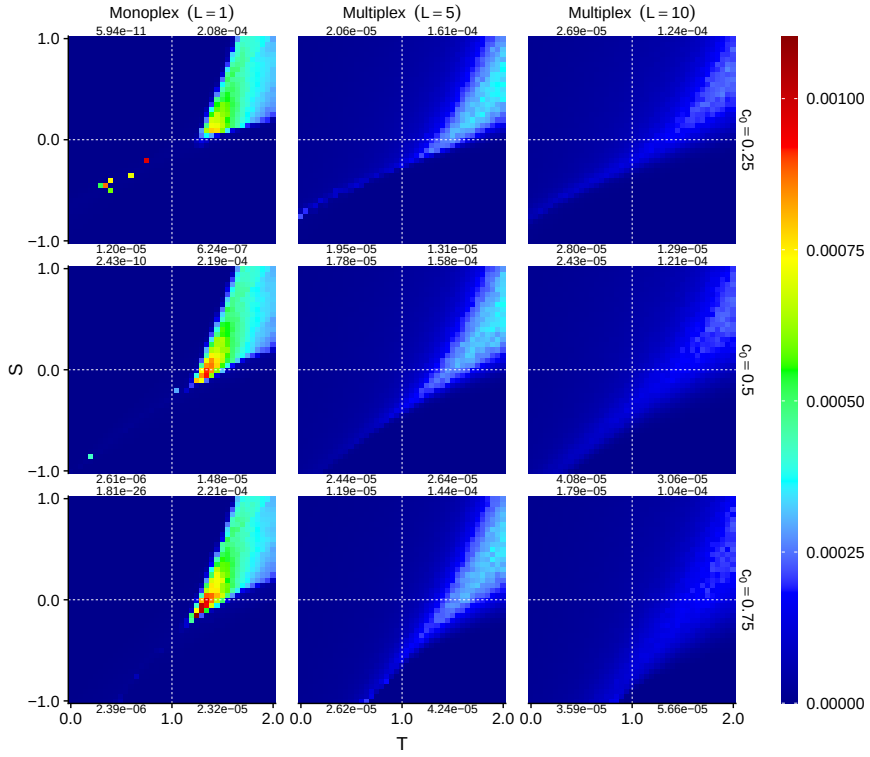


Figure 2.9: Fluctuation of the fraction of cooperator around the fitted trend at the final time steps of the simulation computed as described in the Equation (2.16), for each pair of values S and T. The number in each quadrant represents the average value for each one of the four games. The results are provided for 3 different of initial fraction of cooperators: top row $c_0 = 0.25$, middle row $c_0 = 0.5$ and bottom row $c_0 = 0.75$.

2.5 TEMPORAL FLUCTUATIONS

Last, we turn our attention to the fluctuations of the system in the stationary state. In the case of these four games on a monoplex network, it is well known that the level of cooperation in the stationary state fluctuates around a well-defined average value due to the effect of both the topological structure of the network and the Replicator updating rule. We propose a measure to quantify such effect and later compare it with the cases of multiplex networks. For each one of the R repetitions of the experiment we fit a linear model to the final t_γ time steps of the

simulation, $\widehat{c(t)}_r = \alpha_r + \beta_r t$. The effect of the trend, that is, a possible non-zero slope in the measure of fluctuations (see in Section 2.7), has to be considered, so we average the Mean Square Error (MSE) between the data from the simulations and the predictions of the linear model for the I iterations, calculated as:

$$\zeta = \frac{1}{R \cdot L} \sum_{r=1}^{R} MSE_r = \frac{1}{R \cdot L \cdot t_\gamma} \sum_{r=1}^{R} \sum_{t=t_0}^{t_\gamma + t_0} (c(t)_r - \widehat{c(t)}_r)^2 \tag{2.16}$$

The results for monoplex and multiplex networks are displayed in Figure 2.9. The simulations show small fluctuations in the quadrants of Stag-Hunt and Harmony Game. However in the Snow Drift monoplex results display a zone where fluctuations are larger than in the rest of the $T - S$ plane. That can be attributed to the nature of the game: it has an evolutionary stable equilibrium with mixed populations, so a consensus where both strategies coexist has to be reached. To achieve this objective some nodes have to alternate their strategies. These changes, due to the topological features of the network, can lead to a cascade effect of changes in a large portion of the network; the equilibrium gets disturbed, and the equilibrium has to be reached again. This causes the fluctuations that we have measured. It is worth noticing that, in the area of mutual coexistence of strategies, the fluctuations are larger where the temptation and sucker payoffs are not far from the payoffs of mutual cooperation and mutual defection. It is also noteworthy that, even when the Prisoner's Dilemma quadrant presents very small fluctuations in general, it does show a small but very significant spot near the line of weak Prisoner's Dilemma, where they are large. Again, this corresponds to the area of competition between Cooperation and Defection, where each of the strategies accounts roughly for half of the population.

The introduction of multiplex networks has an enormous effect on the measured fluctuations. The fluctuations are again in the region of coexistence of strategies, however, in the case of 5-layer multiplex the fluctuations are smaller than in the monoplex case. The results in the 10-layer multiplex displays even a larger reduction in the measure of fluctuations (compare also the three curves shown in Figure 2.8).

The nature of such reduction from monoplex to multiplex has to be found in the interlayer dynamics. Each layer is pushed to reach a stable equilibrium where both strategies can coexist, nonetheless the shared information between the layers establishes a way to constrict the range of the fluctuations. The change of strategy of a node in one layer is not conditioned by its performance in that layer, but by its global performance in the entire multiplex structure. That makes the system more robust to fluctuating nodes, at the expense of convergence time to the stationary equilibrium. It has to be said that the initial fraction of cooperators, c_0, does not have any observable influence on the size of the fluctuations.

Finally, it is worth mentioning that the fluctuations shown in Figure 2.8 are a modified version of equation (2.16), calculated as follows: we have to divide the time range in different slices in the interest of realizing local accurate measures of the fluctuations in each slice, so we fix the size of the time window to $t_w = 1000$. For each time slice we fit a linear model to each of the R runs of our simulation, then we compute the residuals as the difference between the linear model and the data from the simulations. We have R residuals that measure the size of the fluctuations at each time step, and we plot a range corresponding to twice their standard deviation to provide information about the size of the fluctuations at each time step.

2.6 SUMMARY

The introduction of multiplex networks is not only a more realistic representation of social systems, allowing for more sophisticated individual behaviors, but as it has been shown in other contexts too (see Section 1.2.4), it has a profound effect on the dynamics developing on top of them. Thus, as we commented at the beginning of the chapter, higher-order structures provide an enhancement on the understanding of network dynamics that cannot be neglected.

In particular, we have found that the stationary distribution of cooperation in the plane $T - S$ becomes less sharp as more layers are added. In the monoplex case there is a very narrow area that separates all-cooperator from all-defector areas for the Stag Hunt and Prisoner's

Dilemma games, but in the multiplex scenario we find that it becomes a wider region, with intermediate values of cooperation. We also find that the region of all-defectors shrinks as the number of layers increases. As a counter-effect though, we find a slight decrease in the value of cooperation (even in the quadrant of the Harmony game), from total cooperation to values around 90%. These results are consistent with and generalize those found by [59]: the introduction of a multiplex structure in the population helps promote cooperation in regions of the parameter space in which it can not survive in the monoplex scenario, at the expense of a moderate decrease of cooperation in those where traditionally it was very high. We explored the microscopic underpinnings for these phenomena, previously observed but unexplained in the aforementioned article.

Regarding the microscopic behavior of the nodes, we have found that in general, there are three types of individuals: those coherently acting as cooperators in all layers, those acting as coherent defectors, and a group of incoherent individuals, that play as cooperators in some layers and as defectors in others. The existence of this third incoherent group is at the root of the explanation of the survival of defection in the Harmony Game for a multilayered network, and it is also responsible for a large part of the cooperation in the central areas of the $T - S$ plane, where cooperation is lower in a monoplex. Also, we have analyzed how this three types of players interact among them, concluding that there are plenty of interaction between incoherent and coherent cooperators, fewer interactions between incoherent and coherent defectors, and practically no interaction between both types of coherent players. Moreover, this is a very plausible social scenario: some people may behave consistently in all their types of interactions (for example at work, at home, with friends,etc.) either cooperating or defecting, and some other may choose different strategies for different layers (for example, cooperate with family and defect at work). We have found that the fraction of incoherent players increases with the number of layers, which means that as the number of contexts where the players interact increases, it gets harder to maintain a coherent behavior in all of them. Regarding the dependence with the initial fraction of cooperation, we found that our system behaves consistently with what was found for the monoplex

network, and the effect of adding more layers is preserved or even enhanced with increasing initial fraction of cooperators.

We have shown how convergence to a stationary equilibrium is one of the descriptors that gets modified by the introduction of multiplex networks. More precisely, we have shown that, in the transition regions, the convergence is more difficult to reach. In this case, we have observed that the initial fraction of cooperators has a vast influence in the achievement of the stationary equilibrium. As an explanation, we have proposed that the interlayer dynamics supports more defectors playing cooperation in other layers than cooperators playing defective strategies in a distinct layer.

Finally, we have examined the influence of the multilayered network in the fluctuations of the number of cooperators around the stationary state. We have defined a new detrended ratio to measure such change. We have shown that the conclusions are significant. In the Snow Drift, the region where the fluctuations are observed, we have shown a decrease in their size as we keep adding layers to the system.

2.7 PERSONAL CRITICAL VIEW

Although the proposed method allows to understand how cooperation evolves in some situations where it is not supposed to be present, these results have to be considered cautiously when we try to extrapolate them to human behavior. Our model assumes that both players of a game have the same perception about the outcome of the game and they quantify in the same way the payoffs; however, that is not true in many real situations and, for instance, two individuals could evaluate the outcomes of a game differently due to some moral constrains. It also considers that the structure of the population does not change over time, a hard assumption considering that people could stop playing with peers that constantly defect on them, changing completely the observed results. Additionally, it is also important to highlight that the evolution of the game relies harshly on the rule used to update the strategies, an sometimes this dependency could hide the contribution of the structure on the outcome of the game.

3

RELATIONAL HIGHER-ORDER: EPIDEMIC LINK EQUATIONS, LINK IMPORTANCE, AND EPIDEMIC CONTAINMENT

3.1 INTRODUCTION

The modeling of dynamical processes to study complex scenarios is one of the critical aspects of networks science. They can be used to understand how traffic on a city flows, how neurons interact, how rumors spread, how new technology is adopted or, as we saw on the previous chapter, how cooperation evolves. In some situations, a macroscopic analysis is enough to understand many emerging properties of the system like the appearance of epidemic outbreaks or the appearance of congestion. However, in some other situations, a microscopic analysis of the evolution of some properties at entity level could lead to more comprehensive analysis, for instance, in epidemic processes. In these processes, information about the probability of a specific node of being infected could be useful to understand which individuals are more prone to be affected by an infectious disease or to study the impact of a vaccination campaign.

Although dynamical information at node level is effective to understand how complex dynamics affect individual entities, those entities are connected by relational links to other entities; thus those properties are also tied. Modeling those ties can be seen as a higher-order microscopic analysis at a relational level and can play a decisive role in our understanding of how links shape how dynamical processes work on networks. Going back to our example of an epidemic process, to spread

through the network, an infectious disease uses connections between infected and non-infected individuals. Designing a model that accounts for the probability that this configuration happens at microscopic level will help to understand the possible contact paths that the infection will follow through its spreading.

In this chapter, we confront this challenge and propose a set of discrete-time governing equations which can be closed and analyzed, assessing the contribution of links to spreading processes in complex networks. Our approach allows a scheme for the containment of epidemics, based on deactivating the most important links in transmitting the disease. The model is validated in synthetic and real networks, obtaining an accurate determination of the epidemic incidence and critical thresholds. Epidemic containment based on links' deactivation promises to be an effective tool to maintain functionality on networks while controlling the spread of diseases, as in air transportation networks.

3.2 EPIDEMIC LINK EQUATIONS

The contributed model [95] is built upon the relationships between the states of nodes connected by links, thus being related to pairwise approximations [26, 53, 80, 94, 149, 152, 156]. However, our model is microscopic, at the level of individual links as in [94], thus allowing for a clear identification of maximally infectious links, but with the additional advantage of being able to easily calculate the incidence of the epidemics and the importance of the links, not just the epidemic threshold.

3.2.1 *Description*

Let us consider a discrete time SIS dynamics which runs on top of a complex network of N nodes and L edges, with adjacency matrix A, and where each node i can be in one of two different states σ_i, either susceptible (S) or infected (I), i.e. $\sigma_i \in \{S, I\}$. We can say that a link (i, j) between nodes i and j is in state SI if $\sigma_i = S$ and $\sigma_j = I$. The parameters of the SIS dynamics are the infection and recovery probabilities, β and μ, respectively.

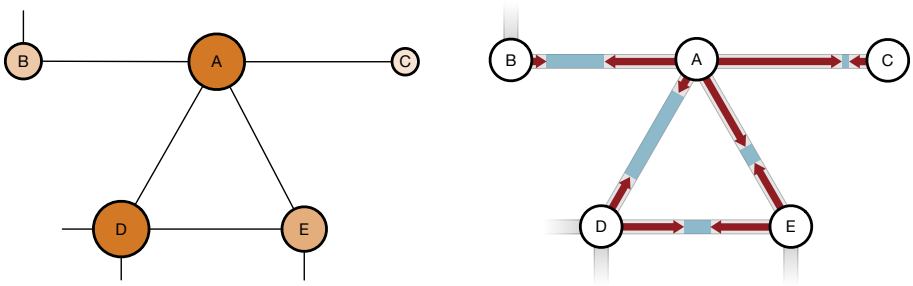


Figure 3.1: Left, state of the epidemic spreading at node level, where the size and the color of the nodes is proportional to the probability of being in an infected state. Right, quantification of the role of the links in the dynamics: the length of the arrow represents probability Φ, i.e. the probability that the link is in a state that enables the spreading of the disease, and the blue space between the arrows is proportional to the probability that a given link stays in a non-transmittable state ($\Theta^I + \Theta^S$). Both representations are complementary and help to understand the spreading processes in complex networks.

We are interested in building a probabilistic model that takes into account all the possible combination of states between pairs of nodes linked by a link because they provide information about how the spreading flows on the network. For instance, if we know the probability that a node i is on a susceptible state while node j, is infected we could determine the chances that infection goes from j to i. Mathematically, that probability is given by the joint probability $P(\sigma_i = S, \sigma_j = I)$, and knowing the value for of this probability for each link of the system is determinant, as we will see, to design containment strategies.

To simplify the notation, we first denote the previous joint probability as $\Phi_{ij} = P(\sigma_i = S, \sigma_j = I)$; the higher the Φ_{ij}, the larger the likelihood that the disease propagates from node j to node i. It is worth mentioning that this feature is in general asymmetrical, meaning that the propagation of the illness can be more probable from j to i than the other way around. In the same way, the epidemic is restrained by edges where the nodes are in the same state, thus it is convenient to define the probabilities $\Theta^S_{ij} = P(\sigma_i = \sigma_j = S)$ and $\Theta^I_{ij} = P(\sigma_i = \sigma_j = I)$ for all pairs of neighboring nodes. See Figure 3.1 for a graphical representation of these probabilities.

The evolution of the joint probability Φ_{ij} of one link depends on the probabilities Φ, Θ^I and Θ^S to the rest of the neighboring links, and the infection rules of the SIS dynamics. Thus, we can write the following equation for each link:

$$\begin{aligned}\Phi_{ij}(t+1) &= \Theta^S_{ij}(t)\, q_{ij}(t)\,(1-q_{ji}(t)) \\ &\quad + \Phi_{ij}(t)\,((1-\beta)q_{ij}(t))\,(1-\mu) \\ &\quad + \Phi_{ji}(t)\,\mu\,(1-(1-\beta)q_{ji}(t)) \\ &\quad + \Theta^I_{ij}(t)\,\mu(1-\mu)\end{aligned} \tag{3.1}$$

where $q_{ij}(t)$ stands for the probability that a susceptible node i is not infected by any of its neighbors (excluding node j). We have taken into account all the possible changes of state of the nodes i and j. The first term considers the probability that both nodes are in a susceptible state, and then node i remaining susceptible while node j is infected by any of its other neighbors. The second term accounts for both nodes remaining in the same state, node i is not infected by any of its neighbors and node j is not recovered from the infection. Then, the third term represents the transition in which node i is infected and recovers while node j is susceptible and it is infected by any of its other neighbors. Finally, in the fourth term both nodes are infected but node i recovers while node j does not. The asymmetry of probability Φ_{ij} multiplies the number of equations by two, since for each link between nodes i and j we need an equation for $\Phi_{ij}(t+1)$ and another for $\Phi_{ji}(t+1)$.

Similarly we can obtain an expression for probability Θ^I_{ij}:

$$\begin{aligned}\Theta^I_{ij}(t+1) &= \Theta^S_{ij}(t)\,(1-q_{ij}(t))\,(1-q_{ji}(t)) \\ &\quad + \Phi_{ij}(t)\,(1-(1-\beta)q_{ij}(t))\,(1-\mu) \\ &\quad + \Phi_{ji}(t)\,(1-\mu)\,(1-(1-\beta)q_{ji}(t)) \\ &\quad + \Theta^I_{ij}(t)\,(1-\mu)^2\end{aligned} \tag{3.2}$$

In this case we have only L equations, one per link, due to its symmetry. There is no need of extra equations for probability Θ^S_{ij} since the normalization leads to $\Theta^S_{ij} = 1 - \Phi_{ij} - \Phi_{ji} - \Theta^I_{ij}$.

The $q_{ij}(t)$ in equations (3.1) and (3.2) can be expressed as:

$$q_{ij}(t) = \prod_{\substack{r=1 \\ r \neq j}}^{N} (1 - \beta A_{ri} h_{ir}) \tag{3.3}$$

where h_{ij} defines the hostility of j against i, i.e. the probability that node j is infected when node i is susceptible, $h_{ij} = P(\sigma_j = I | \sigma_i = S)$. Note that we are assuming independence between the probabilities of not being infected by each of the neighbors. The hostility can be obtained in terms of Θ_{ij}^S and Φ_{ij} as:

$$h_{ij} = \frac{\Phi_{ij}}{\Phi_{ij} + \Theta_{ij}^S} \tag{3.4}$$

Note that the denominator in equation (3.6) is a property of node i given that $\Phi_{ij} + \Theta_{ij}^S = P(\sigma_i = S)$ for all neighboring nodes j of vertex i.

We call this system of $3L$ equations and unknowns our *Epidemic Link Equations* (ELE) model. It can be solved by iteration, starting from any meaningful initial condition, e.g. $\Theta_{ij}^I(0) = \rho_0^2$ and $\Phi_{ij}(0) = \Phi_{ji}(0) = \rho_0(1 - \rho_0)$ (for any $0 < \rho_0 \leq 1$), until a fixed point is found. At this point, the incidence of the epidemic process, the average number of infected nodes in the whole system, can be computed as:

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{k_i} \sum_{j=1}^{N} A_{ji} (\Phi_{ji} + \Theta_{ij}^I) \tag{3.5}$$

where k_i is the degree of node i.

Apart from the solution where all nodes are susceptible, $\Theta_{ij}^S = 1$ for all the links, a non-trivial one appears when the system is above the critical value of the epidemic spreading.

3.2.2 *Epidemic Threshold*

The determination of the epidemic threshold requires the consideration of states in which the probabilities of having infected nodes are very small, i.e. $\Phi_{ij}, \Phi_{ji}, \Theta_{ij}^I \ll 1$. Therefore, we may suppose that

$\Phi_{ij}, \Phi_{ji}, \Theta^I_{ij} \sim O(\varepsilon)$, with $\varepsilon \ll 1$, and in consequence $\Theta^S_{ij} \sim 1 - O(\varepsilon)$. Using these approximations the epidemic link equations become linear in these $O(\varepsilon)$ probabilities, since $O(\varepsilon^2)$ terms should be neglected.

We start with the linearization of the hostility:

$$\begin{aligned} h_{ij} &= \frac{\Phi_{ij}}{\Phi_{ij} + \Theta^S_{ij}} \\ &= \frac{\Phi_{ij}}{1 - (\Phi_{ji} + \Theta^I_{ij})} \\ &= \Phi_{ij}\left(1 + \Phi_{ji} + \Theta^I_{ij} + O(\varepsilon^2)\right) \\ &= \Phi_{ij} + O(\varepsilon^2)\,, \end{aligned} \tag{3.6}$$

where we have used the normalization $\Phi_{ij} + \Phi_{ji} + \Theta^I_{ij} + \Theta^S_{ij} = 1$, and we realize that terms $\Phi_{ij}\Phi_{ji}$ and $\Phi_{ij}\Theta^I_{ij}$ are both $O(\varepsilon^2)$. Note that here we have removed the dependence on time, to emphasize we are considering the steady state. Substituting hostility in the expression for q_{ij} we get:

$$\begin{aligned} q_{ij} &= \prod_{\substack{r=1 \\ r\neq j}}^{N} (1 - \beta A_{ri} h_{ir}) \\ &= \prod_{\substack{r=1 \\ r\neq j}}^{N} \left(1 - \beta A_{ri}\Phi_{ir} + O(\varepsilon^2)\right) \\ &= 1 - \beta \sum_{\substack{r=1 \\ r\neq j}}^{N} A_{ri}\Phi_{ir} + O(\varepsilon^2) \\ &= 1 - \beta \sum_{r=1}^{N} A_{ri}(1 - \delta_{rj})\Phi_{ir} + O(\varepsilon^2)\,, \end{aligned} \tag{3.7}$$

where the Kronecker δ_{rj}, which is 1 if $r = j$ and 0 otherwise, has been introduced to make zero the jth term of the sum. Now we are in condition to find the linear approximations of the main ELE model equations. First, the equation for Θ^I_{ij} becomes:

$$\Theta^I_{ij} = \Theta^S_{ij}\,(1 - q_{ij})\,(1 - q_{ji})$$

$$
\begin{aligned}
&+ \Phi_{ij}\,(1-(1-\beta)q_{ij})\,(1-\mu)\\
&+ \Phi_{ji}\,(1-\mu)\,(1-(1-\beta)q_{ji})\\
&+ \Theta^{I}_{ij}\,(1-\mu)^2\\
=\;& O(\varepsilon^2)\\
&+ \Phi_{ij}\,\beta(1-\mu) + O(\varepsilon^2)\\
&+ \Phi_{ji}\,(1-\mu)\beta + O(\varepsilon^2)\\
&+ \Theta^{I}_{ij}\,(1-\mu)^2\\
=\;& \beta(1-\mu)\,\Phi_{ij} + \beta(1-\mu)\,\Phi_{ji} + (1-\mu)^2\,\Theta^{I}_{ij} + O(\varepsilon^2)\,. \qquad (3.8)
\end{aligned}
$$

Note that, for the terms with a factor Φ_{ij}, Φ_{ji} or Θ^{I}_{ij}, which are $O(\varepsilon)$, we just need to keep $O(1)$ contributions in the rest of the term, thus we may use the approximations $q_{ij} = q_{ji} = 1 + O(\varepsilon)$.

The equation for Φ_{ij} reads:

$$
\begin{aligned}
\Phi_{ij} =\;& \Theta^{S}_{ij}\,q_{ij}(1-q_{ji})\\
&+ \Phi_{ij}\,((1-\beta)q_{ij})\,(1-\mu)\\
&+ \Phi_{ji}\,\mu\,(1-(1-\beta)q_{ji})\\
&+ \Theta^{I}_{ij}\,\mu(1-\mu)\\
=\;& \beta\sum_{r=1}^{N} A_{rj}(1-\delta_{ri})\,\Phi_{jr} + O(\varepsilon^2)\\
&+ \Phi_{ij}\,(1-\beta)(1-\mu) + O(\varepsilon^2)\\
&+ \Phi_{ji}\,\mu\beta + O(\varepsilon^2)\\
&+ \Theta^{I}_{ij}\,\mu(1-\mu)\\
=\;& \beta\sum_{r=1}^{N} \left(A_{rj}(1-\delta_{ri}) + \mu\delta_{ri}\right)\,\Phi_{jr}\\
&+ \Phi_{ij}\,(1-\beta)(1-\mu) + \Theta^{I}_{ij}\,\mu(1-\mu) + O(\varepsilon^2)\\
=\;& \beta\sum_{r=1}^{N}(A_{rj} - (1-\mu)\delta_{ri})\,\Phi_{jr} + (1-\beta)(1-\mu)\,\Phi_{ij}\\
&+\mu(1-\mu)\,\Theta^{I}_{ij} + O(\varepsilon^2)\,. \qquad (3.9)
\end{aligned}
$$

For the last step we have made use of $A_{rj}\delta_{ri} = A_{ij}\delta_{ri} = \delta_{ri}$, since these equations correspond to a link between nodes i and j, thus we are implicitly assuming that $A_{ij} = 1$.

Summarizing, the linearized equations of the ELE model can be expressed as:

$$\Theta^I_{ij} = \beta(1-\mu)\Phi_{ij} + \beta(1-\mu)\Phi_{ji} + (1-\mu)^2\Theta^I_{ij} \tag{3.10}$$

$$\Phi_{ij} = \beta\sum_r (A_{rj} - (1-\mu)\delta_{ri})\Phi_{jr} + (1-\beta)(1-\mu)\Phi_{ij} + \mu(1-\mu)\Theta^I_{ij} \tag{3.11}$$

From equation (3.10) we can write

$$\Theta^I_{ij} = \frac{\beta(1-\mu)}{\mu(2-\mu)}(\Phi_{ij} + \Phi_{ji}) \tag{3.12}$$

Now, calling $\varepsilon_i = \Phi_{ji} + \Theta^I_{ij} \ll 1$, which does not depend on node j since $P(\sigma_i = I, \sigma_j = S) + P(\sigma_i = I, \sigma_j = I) = P(\sigma_i = I)$, we make the following ansatz:

$$\Theta^I_{ij} = \mathrm{Y}(\varepsilon_i + \varepsilon_j) \tag{3.13}$$

$$\Phi_{ij} = X\varepsilon_i + Z\varepsilon_j \tag{3.14}$$

where Y, X and Z are constants independent of the link. These ansatz include the assumption of symmetry of Θ^I_{ij} and asymmetry of Φ_{ij}, respectively. We can determine the constants by substitution in equation (3.12) and using the definition of ε_i, which leads to

$$\mathrm{Y} = \frac{\beta(1-\mu)}{\mu(2-\mu) + 2\beta(1-\mu)} \tag{3.15}$$

$$X = -\mathrm{Y} \tag{3.16}$$

$$Z = 1 - \mathrm{Y} \tag{3.17}$$

Finally, we build equations for the ε_i by substituting equations (3.10) and (3.11) in $\varepsilon_i = \Phi_{ji} + \Theta^I_{ij}$, and using the ansatz. The result is

$$\frac{\mu}{\beta}\varepsilon_i = \sum_j B_{ji}\varepsilon_j \tag{3.18}$$

where B is a matrix whose elements depend on the adjacency matrix of the network, on Υ, and on the degrees k_i of the nodes:

$$B_{ij} = (1 - \Upsilon)A_{ij} - \Upsilon k_i \delta_{ij} \tag{3.19}$$

If μ/β is an eigenvalue of matrix B, equation (3.18) has a non-trivial solution. Hence, the onset of the epidemics β_c, the lowest value of β which yields non-trivial solutions of equation (3.18), is given by

$$\beta_c = \frac{\mu}{\Lambda_{\max}(B)} \tag{3.20}$$

where $\Lambda_{\max}(B)$ is the largest eigenvalue of matrix B. Note that matrix B depends on β and μ, thus equation (3.20) is implicit for β_c, which can be solved by iteration.

3.2.3 *Validation*

To test the agreement between our Epidemic Link Equations and empirical simulations we have analyzed the incidence of the epidemics, ρ, in different synthetic and real network structures, covering the full range of infection probabilities, β, see Figure 3.2.

The numerical incidence of the epidemics, ρ, is calculated using discrete-time and synchronous Monte Carlo simulations. We make use of the quasistationary approach (QS) [19, 50] to avoid the effect that large number of realizations end up in the absorbing state with no infected individuals in the system. Basically, the QS method focus the simulation on active configurations, i.e., with one or more infected individuals. Every time the system reaches the absorbing state, this state is replaced by one of the previously stored active states of the system. We keep 50 active configurations with an update probability of 0.20. We give the systems a transient time of 10^5 time steps and then, we calculate the ρ as an average over a relaxation time of 2×10^4 time steps.

We evaluate our methodology on synthetic and empirical networks. We have built a network with power-law degree distribution $P(k) \sim k^{-\gamma}$ with exponent $\gamma = 3$ and $\langle k \rangle = 6$ using the configuration model. To evaluate the impact of transitivity, we have also built another network

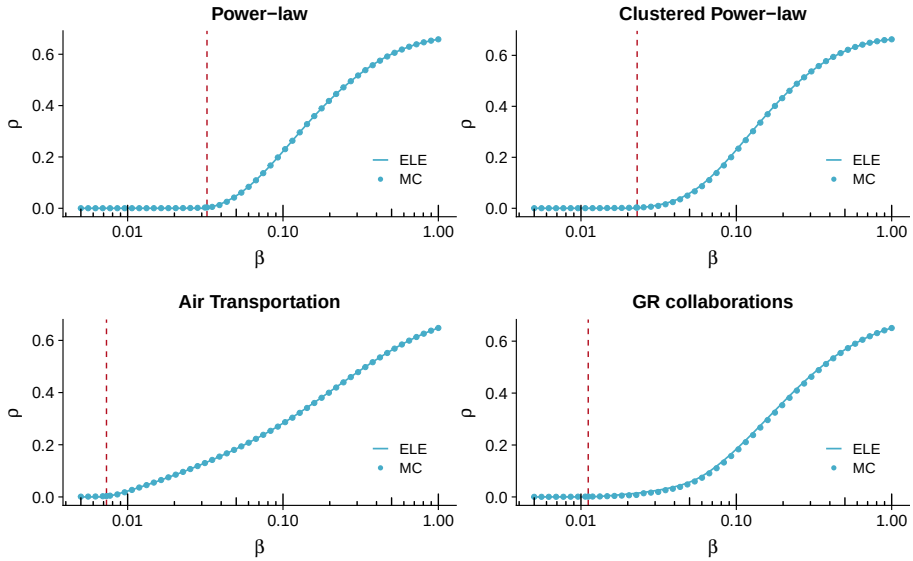


Figure 3.2: Incidence of the epidemic process ρ as a function of the infection probability β. We show the incidence level for the ELE model (solid lines) and for Monte Carlo simulations (circles). The theoretical epidemic threshold calculated using equation (3.20) is marked with a vertical line. We have made use of two synthetic and two real networks: two scale-free networks (top) with exponent 3, one of them with high clustering coefficient; the world air transportation network; and the network of scientific collaborations in the field of general relativity. We have set the recovery rate for all the networks to $\mu = 0.5$.

with the same characteristics of the previous one but with a clustering coefficient of 0.6, using the algorithm by Holme et al. [72] with a parameter $p = 0.8$.

We consider also two empirical networks: the air transportation network, and the network of scientific collaborations in the field of general relativity. The air transportation network has been constructed using data from the specialized website[1], with information about the traffic between airports updated to 2012. This network accounts for the largest connected component, with 3154 nodes and 18592 edges. The network of scientific collaborations has been obtained from [84]; it is composed by 5242 nodes linked by 14496 edges.

1 http://openflights.org

The results show a remarkable agreement between our ELE model and the Monte Carlo simulations, and a good prediction of the epidemic threshold for all synthetic and real networks, pointing out the validity of our model to describe the global impact of the epidemics. Note that all networks, except the first one, have a large clustering coefficient, making the determination of the incidence difficult for standard mean field methods, due to the effect of dynamical correlations.

3.3 EFFECTIVE EPIDEMIC CONTAINMENT

The design of effective containment strategies constitutes a major challenge. Measures like vaccination, improved hygiene, bio-security, cattle culling, or education to prevent contagions, operate on the biological aspects of the disease. On the other hand, isolation or mobility restrictions act on the physical spreading channels, which may transform a local event into a pandemic. Here, we concentrate on the role of the links of the spreading network. For example, if we identify the edges which are more involved in the propagation of a disease, it is possible to design targeted countermeasures which affect just specific links instead of whole nodes, while being more effective. This can be illustrated by a hypothetical pandemic disease propagated using the air transportation network: the isolation of one airport is a dramatic measure which is socially and politically difficult to accept and put into practice, but the suppression of just a few connections between selected airports could be more easily assumed, and at the same time achieving a better containment of the disease.

Previous works have directed their attention mostly towards schemes based on the actuation on single nodes, either randomly or according to node properties such as their degree, betweenness, PageRank or eigenvector centrality [28, 29, 73, 122]. Following the same idea, some authors have introduced link removal using properties of the adjacent nodes (degrees or centralities) or of the link itself (edge betweenness) [73, 139, 157]. A model of coevolution of epidemics with permanent and temporal link removals was proposed in [166], and methods from optimization and control have been applied to minimize the impact of the epidemics [109, 123, 167]. Currently, it is considered that the optimal approach consists in finding the minimum set of edges

whose removal leads to a maximum decrease of the spectral radius of the network, i.e., the largest eigenvalue of the adjacency matrix [20, 135, 157]. Since the epidemic threshold is, at first order approximation of a susceptible-infected-susceptible (SIS) epidemic dynamics, inverse to the spectral radius, it seems the best and more mathematically grounded option. Unfortunately, it turns out to be an NP-complete problem, thus only heuristics are available for large networks [157].

It must be emphasized that all the previous approaches make use only of the structural characteristics of the network to decide which nodes or edges have to be removed; the characteristics or parameters of the epidemic process are ignored. Even the spectral radius, which is closely related to the epidemic threshold, does not depend on the infection or recovery rates, the expected number of infected neighbors around a certain node, or any other local or global information of the spreading process.

Our proposal [95] concentrates on the role of the links in the spreading of the epidemics, quantifying the importance of each link [126], thus enabling containment strategies based on their removal. To this end, we define the epidemic importance of a link as its capacity to infect other individuals once this link has been used to propagate the disease. The calculation of this link epidemic importance requires the determination of the joint and conditional probabilities of the states of all the links, which can be done using our Epidemic Link Equations, see Section 3.2. Moreover, the quantification of the epidemic importance of the edges leads to a link removal strategy which in many cases outperforms the previous approaches, even those based on minimizing the spectral radius, and preserving at the same time most of the connectivity of the network.

3.3.1 *Link Epidemic Importance*

Our objective is to find an effective strategy to contain the SIS epidemic process through bond percolation. To determine which link should be removed first, we need a measure of the importance of each link in the spreading of the epidemics. A possible option would be to use the edge importance defined in [126], which accounts for the relative change of

the spectral radius when the edge is removed. However, this constitutes an indirect way of containment, since we aim at lowering the incidence of the epidemics as much as possible, whereas the actuation on the spectral radius is directed to increase the epidemic threshold; both are different targets. Additionally, the spectral radius just depends on the structure of the network, but not on the epidemic parameters of the process nor the participation of the link in the spreading of the disease.

We assume the system has reached the stationary state, which does not mean that the nodes remain in a certain fixed state, just that the average incidence of the epidemics is basically constant; thus, there is still margin for applying a containment strategy to minimize this incidence. In this regime, we can measure which are the probabilities of nodes and links to be in each of the epidemic states, e.g., the probability $P(\sigma_i = I)$ of node i being infected, or the joint probability Φ_{ji} of link (i, j) being in state IS.

Consider we have a link in state SS or II. In both cases, the next step of the epidemic dynamics is not going to use this link, since in the former there is no infected node to propagate the disease, and in the latter both nodes are already infected. Thus, to propagate the epidemics, a link must be either in an SI or IS state. Let us suppose we have a link (i, j) in state IS. First, with probability β, node i can infect node j through this link, changing to state II. Next, infected node j may transmit the disease to some of its neighbors. Thus, if we had removed link (i, j), we would have cut this path of infections initiated at node i. This means that, the larger the expected number of infected neighbors of node j, the largest the impact of removing link (i, j) for the spreading of the epidemics. Note that not only the degree of j is relevant, but also the probability of its neighbors being susceptible when j is infected, since you cannot infect nodes which are already infected. For example, if j is surrounded by many infected nodes, cutting link (i, j) is not going to have too much effect on the overall incidence of the epidemics. The expected number of infected nodes produced in this way can be expressed as

$$\begin{aligned}\bar{n}_{ij} &= \beta P(\sigma_j = S, \sigma_i = I) \sum_{r=1}^{N} A_{jr} \beta P(\sigma_r = S | \sigma_j = I) \\ &= \beta^2 \Phi_{ji} \sum_{r=1}^{N} A_{jr} \frac{\Phi_{rj}}{\Phi_{rj} + \Theta^I_{rj}}\,, \end{aligned} \tag{3.21}$$

where $P(\sigma_r = S|\sigma_j = I)$ is the conditional probability that node r is susceptible when its neighbor j is infected. Since this measure is asymmetric, and removing an edge affects the propagation of the disease in both directions, we define the *link epidemic importance* of a link, I_{ij}, as

$$I_{ij} = \bar{n}_{ij} + \bar{n}_{ji}\,. \tag{3.22}$$

Now, the problem reduces to finding the joint and conditional probabilities for each link, and this is accomplished using our *Epidemic Link Equations* (ELE).

3.3.2 *Connectivity*

An important property of our proposed definition of *link epidemic importance* is that it tends to maintain the connectivity of the network when the selected link is removed, unlike other options such as edge betweenness, which quickly tend to produce a large number of disconnected components, thus hindering the functionality of the network.

To give an intuition about why that happens, consider a network formed by two subnetworks A and B, which are connected by just one link (i_A, j_B); the link epidemic importance of that link is lower than that of another link internal to A or B. This means our containment strategy of removing the edge with largest link epidemic importance will not break the network in two disconnected components A and B, unlike the betweenness approach, for which the link (i_A, j_B) plays the role of a bridge and thus it has maximal edge betweenness.

Let $\langle k \rangle_A$ and $\langle k \rangle_B$ be the average degrees of subnetworks A and B, respectively, with $\langle k \rangle_A > \langle k \rangle_B$. Let us also call ρ^A and ρ^B their respective incidence of the epidemics.

Supposing independence of the states of the nodes, $\bar{n}_{ij}$ can be approximated as

$$\bar{n}_{ij} \approx \beta P(\sigma_j = S) P(\sigma_i = I) \sum_{r=1}^{N} A_{jr} \beta P(\sigma_r = S) \,. \tag{3.23}$$

In a homogeneous mean field approximation we may substitute $k_j = \langle k \rangle$, $P(\sigma_i = I) \approx \rho$, and $P(\sigma_j = S) \approx 1 - \rho$, which lead to the following expressions for the importance of a link:

$$I^A \approx 2\beta^2 \rho^A (1-\rho^A)^2 \langle k \rangle_A \,, \tag{3.24}$$
$$I^B \approx 2\beta^2 \rho^B (1-\rho^B)^2 \langle k \rangle_B \,, \tag{3.25}$$
$$I^{AB} \approx \beta^2 \left[\rho^A (1-\rho^B)^2 \langle k \rangle_B + \rho^B (1-\rho^A)^2 \langle k \rangle_A \right] . \tag{3.26}$$

Here, I^A and I^B denote the link epidemic importance of links inside A and B, respectively, and I^{AB} the link epidemic importance of the link connecting subnetworks A and B.

We need an expression relating the average degree $\langle k \rangle$ and the incidence of the epidemics ρ to be able to calculate the approximate values of the link epidemic importances in Equations (3.24) to (3.26). It can be obtained using the nonperturbative heterogeneous mean field (npHMF) equations in [62]. In particular, the npHMF equations for the SIS model without one-step reinfections (WOR) read as

$$0 = -\mu \rho_k + (1-\rho_k)(1-q_k) \,, \tag{3.27}$$
$$q_k = \prod_{k'} (1-\beta \rho_{k'})^{C_{kk'}} \,, \tag{3.28}$$

where ρ_k represents the fraction of infected nodes of degree k, q_k the probability that nodes of degree k are not infected by nodes of any other degree k', and $C_{kk'} = kP(k'|k)$ the expected number of links from a node of degree k to nodes of degree k'. In the nonperturbative homogeneous mean field (npHoMF) approximation, this reduces to

$$0 = -\mu \rho + (1-\rho)(1-q) \,, \tag{3.29}$$
$$q = (1-\beta \rho)^{\langle k \rangle} \,. \tag{3.30}$$

Thus, after some algebra we get

$$\langle k \rangle = \frac{\log\left(1 - \mu \dfrac{\rho}{1-\rho}\right)}{\log(1-\beta\rho)} . \tag{3.31}$$

An immediate consequence of Equation (3.31) is that $0 \leqslant \rho \leqslant 1/(1+\mu)$. We can see a plot of the npHoMF relationship between ρ and $\langle k \rangle$ in the inset of Figure 3.3. Note that ρ is an increasing function of $\langle k \rangle$, thus the larger the average degree, the greater the incidence of the epidemics.

Now, we can substitute Equation (3.31) for $\langle k \rangle_A$ and $\langle k \rangle_B$ into Equations (3.24) to (3.26) to obtain approximations of the three different link epidemic importances. The results are presented in Figure 3.3. Fixing a certain average degree $\langle k \rangle_B$ for subnetwork B, we consider subnetworks A with $\langle k \rangle_A \geqslant \langle k \rangle_B$. Since ρ increases with $\langle k \rangle$, this is equivalent to fixing ρ^B and consider subnetworks A with $\rho^A \geqslant \rho^B$. We observe that, in all cases, the link epidemic importance I^A of links in subnetwork A is larger than I^{AB} of the bridge link between subnetworks A and B ($I^A \geqslant I^{AB}$), thus confirming that our epidemic containment strategy driven by link epidemic importance does not disconnect the network. This result has been obtained under independence and homogeneous mean field approximations, and for a specific structural configuration of the network.

3.3.3 *Performance*

Our approach for effective epidemic containment consists in removing the links with largest *link epidemic importance*. This is possible once we have solved the ELE model, computing the I_{ij} for all the links in the network using equation (3.22), which can be expressed as

$$I_{ij} = \beta^2 \left(\Phi_{ji} \sum_{r=1}^{N} A_{jr} \frac{\Phi_{rj}}{\Phi_{rj} + \Theta^I_{rj}} + \Phi_{ij} \sum_{r=1}^{N} A_{ir} \frac{\Phi_{ri}}{\Phi_{ri} + \Theta^I_{ri}} \right) . \tag{3.32}$$

Note that the value of β does not affect the ranking of the links, but we do not remove it from equation (3.32) to preserve the semantics of I_{ij}. Since the structure of the network changes after each link re-

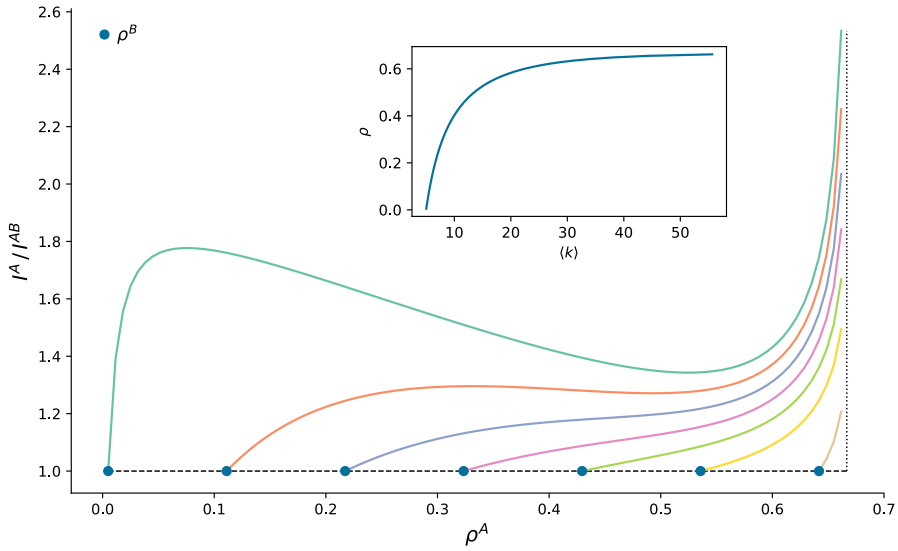


Figure 3.3: Ratio between the link epidemic importance I^A of a link in a subnetwork A and the link epidemic importance I^{AB} of a link which acts as the only bridge between subnetworks A and B. First, we fix the average degree $\langle k \rangle_B$ of subnetwork B (or equivalently, we fix its incidence ρ^B, the red circles), and then we consider subnetworks A with average degree (and epidemic incidence) larger than that of B, i.e. $\langle k \rangle_A \geqslant \langle k \rangle_B$ (thus, $\rho^A \geqslant \rho^B$). We can see that, in all cases, $I^A \geqslant I^{AB}$, meaning that the ranking by link epidemic importance will not be leaded by the bridges. The vertical dotted line highlights the asymptote at $\rho = 1/(1+\mu)$. The inset shows the relationship between the incidence and the average degree. We have set the epidemic parameters to $\mu = 0.5$ and $\beta = 0.1$, and the calculations rely on a nonperturbative homogeneous mean field approximation (npHoMF).

moval, it is convenient to recalculate the solution of the ELE model to ensure that we really remove the current link with largest link epidemic importance.

To measure how efficient is our approach in terms of epidemic containment, we compare it with four additional containment strategies. First, two strategies which just consider the structure of the network: removal based on maximal edge betweenness [73], and targeting the link with highest eigenscore, i.e., the product of the eigenvector centralities of the nodes connected by the link [157]. Then, we consider a measure

based on the epidemic process at the level of nodes, the removal of all the links of the node which has maximal probability of being infected. Finally, a simple random edge removal.

To perform the deactivation of links, we impose an adiabatic process: after each removal step, we let the system converge to the meta-stable equilibrium before removing any other link. For a fair comparison between different containment strategies, we remove on each deactivation step as many edges as we have removed using the node infectivity strategy. In the case of the real networks in Figure 3.14, we have set the maximum number of adiabatic processes to 1000, due to its large computational cost on the largest networks. This means that, if the network has 20000 links, we remove 20 links at each deactivation step. We consider we have reached total containment when the incidence of the epidemics becomes lower than $1/N$.

We have tested the performance of these methods in a complete set of networks with different topological properties (Figures 3.4 to 3.10), computing for each of them the evolution of the incidence (ρ), the incidence on the giant connected component (ρ_{GCC}), the number of components (C_r), and the size of the giant connected component (N_{GCC}) and the second largest connected components (N_{SLCC}), closely related to percolation processes [54, 65], as function of the fraction of removed links during the containment process (L_r/L). In all the cases, as we will show, we observe that link epidemic importance leads to the fastest extinction of the epidemics for the four considered networks, and it is the only method which preserves their connectivity (thus, functionality). Note that the strategy based on node infectivity performs poorly for all the networks despite having information about the epidemic process, although better than the random removal. This means that the use of information at the level of links is crucial to contain the epidemics.

For the power-law network, figure 3.4, our approach using link epidemic importance yields the best performance, but the results are very similar to the ones obtained using eigenscore and edge betweenness strategies (equivalent results hold for Erdős-Rényi networks, see Figure 3.5). However, when the transitivity of the network is increased, Figure 3.6, we can clearly see the benefits of using link epidemic im-

portance, both in epidemic containment and on preservation of the connectivity of the network.

The effect of the clustering coefficient is also present when we look at the epidemic containment results for the empirical networks. Moreover, as in most real networks, the air transportation, Figure 3.7, and the scientific collaborations networks, Figure 3.8, have a significant modular structure. This plays an important role on the epidemic containment process. Here, we can see how the strategy based on edge betweenness apparently performs better when few links are removed, due to the fact that links with higher edge betweenness are those connecting different modules [104]. When the bond percolation process isolates modules, each module may sustain its own epidemic process, and thus it may happen that some of the modules are subcritical for the given infection probability β. That will lead to a decrease of the global prevalence of the epidemics at expenses of loosing the connectivity of the network. Furthermore, if we look at the prevalence on the giant connected component, an important increase above the initial average number of infected individuals is revealed. A consequence of this fragmentation process is the appearance of multiple isolated supercritical components, for which the removal of a link in one of them does not affect the incidence on the other components. As a result, the edge betweenness procedure needs to remove more links to arrive to the total epidemic extinction than any of the other methods, even the random one.

To validate our hypothesis about the effect of modular structures, we have tested the containment procedures over two benchmark networks with community structure, one with an homogeneous degree distribution generated using Stochastic Block Models, Figure 3.9 and another one generated using the LFR algorithm [83] with an heterogeneous degree distribution, Figure 3.10. We can observe how the results match findings made on real networks, observing how the edge betweenness process leads to a fragmentation of the network, requiring the removal of more links to arrive to the total epidemic extinction.

To give a graphical illustration of how the containment process shapes the network, we will use the air transportation network, Figure 3.11 (top). We illustrate the survival links in the air transportation network after 33.3% of the edges have been removed according to our epidemic

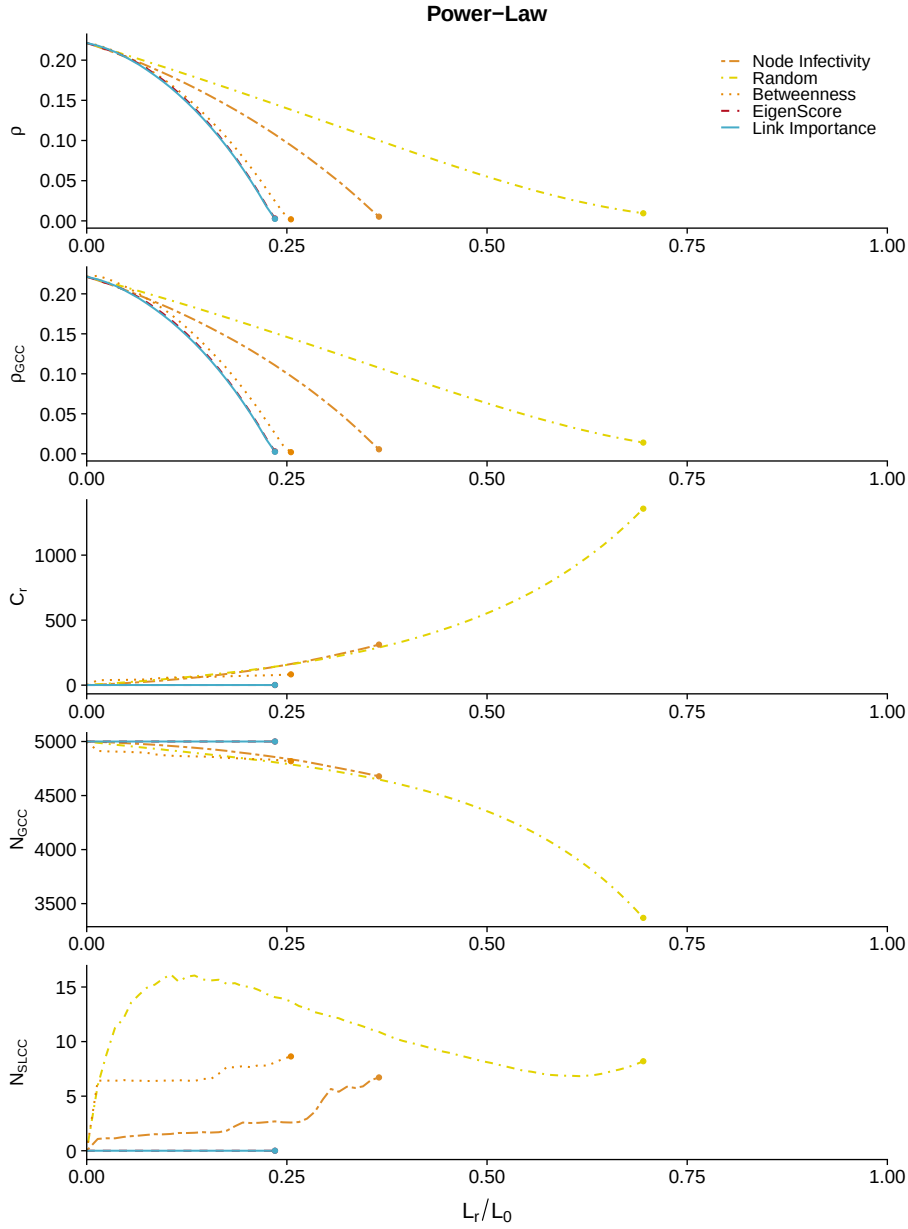


Figure 3.4: Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, and average degree $\langle k \rangle = 6$. The dots mark the achievement of total containment.

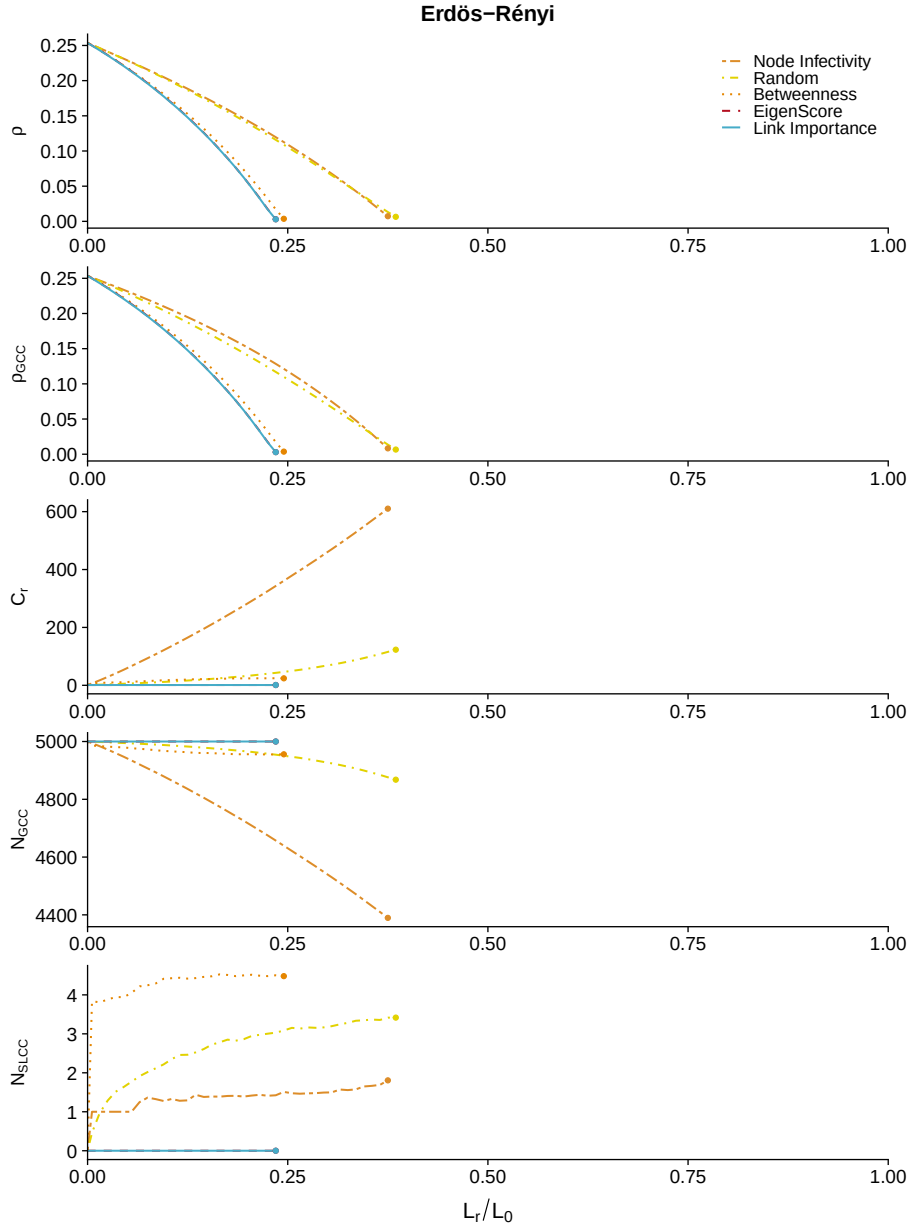


Figure 3.5: Epidemic containment for an Erdős-Rényi network with 5000 nodes and average degree $\langle k \rangle = 6$. The dots mark the achievement of total containment.

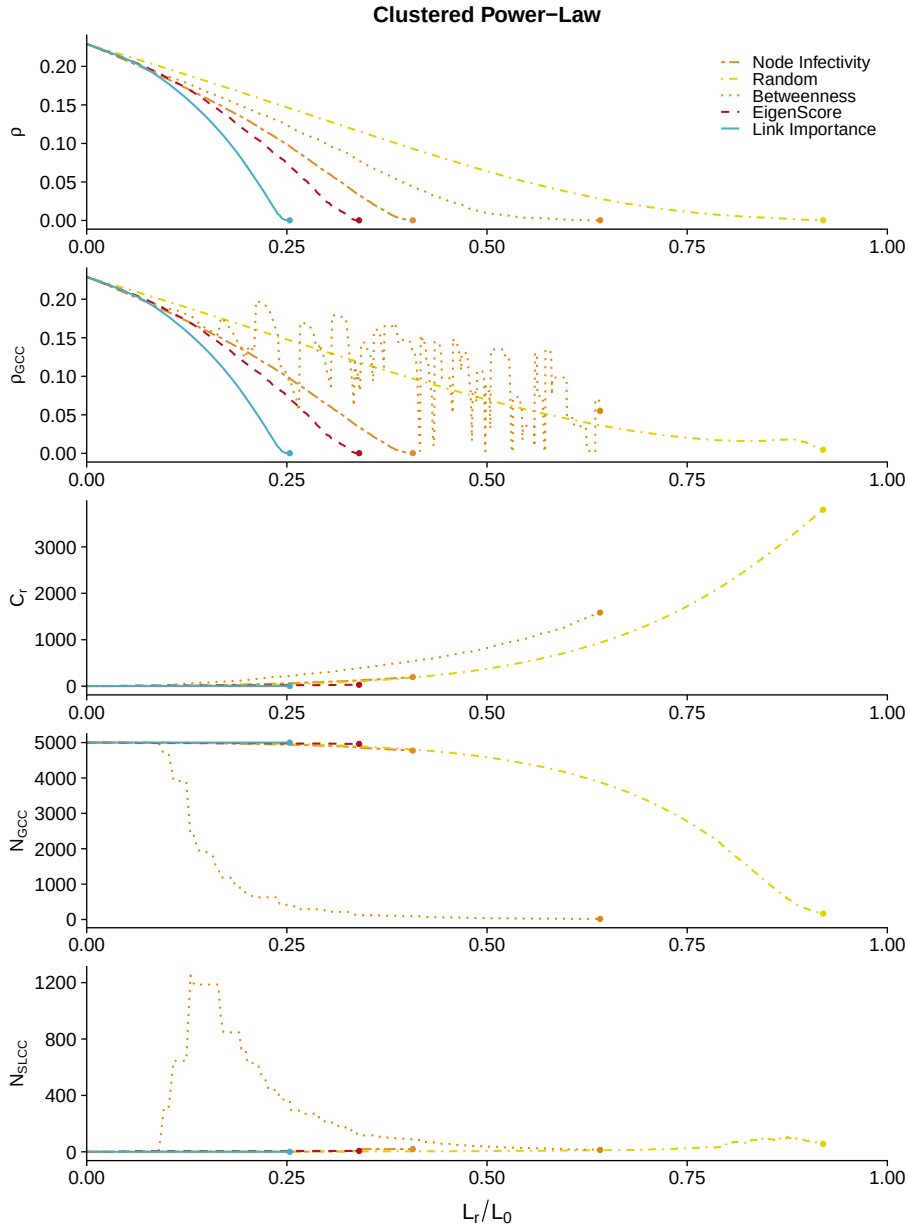


Figure 3.6: Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, high clustering coefficient, and average degree $\langle k \rangle = 6$. The dots mark the achievement of total containment.

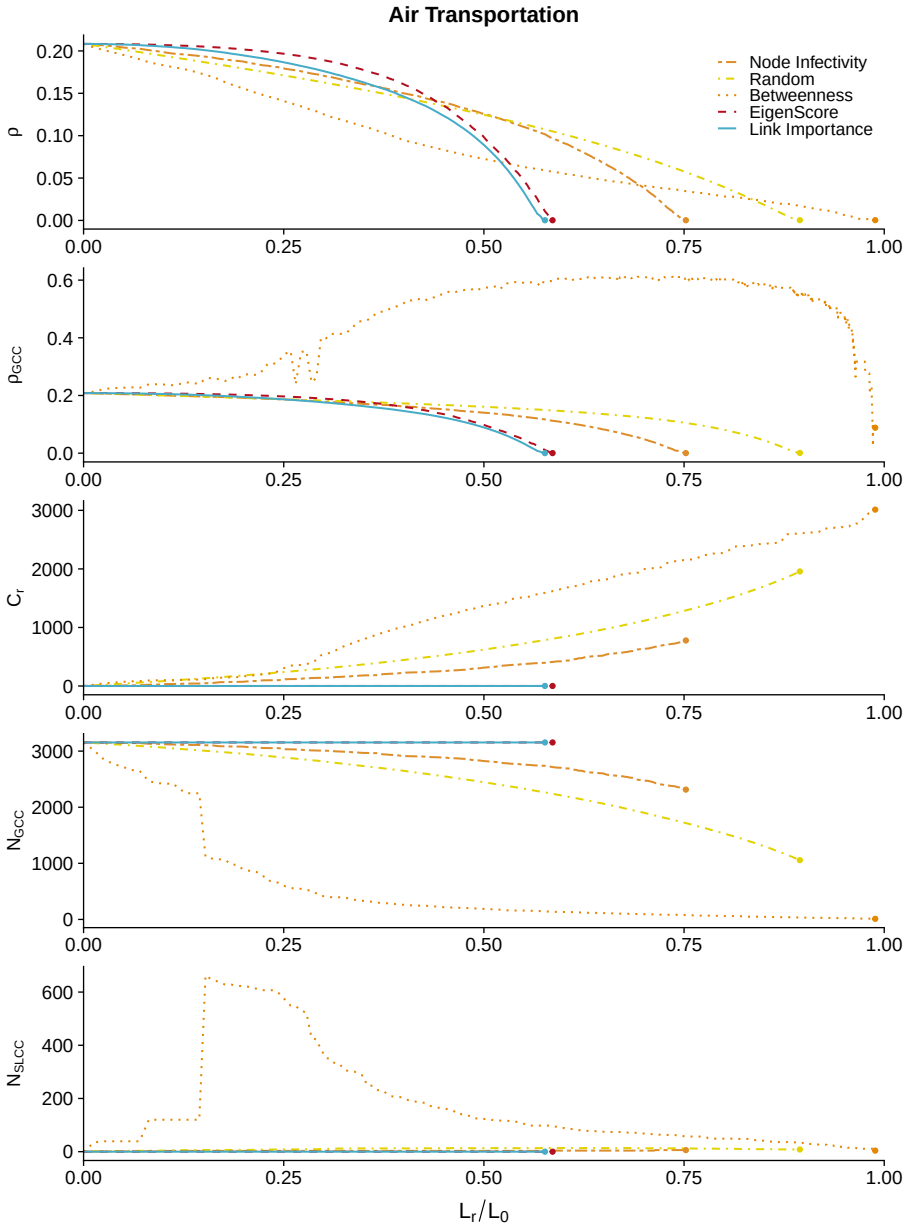


Figure 3.7: Epidemic containment for the air transportation network. The dots mark the achievement of total containment.

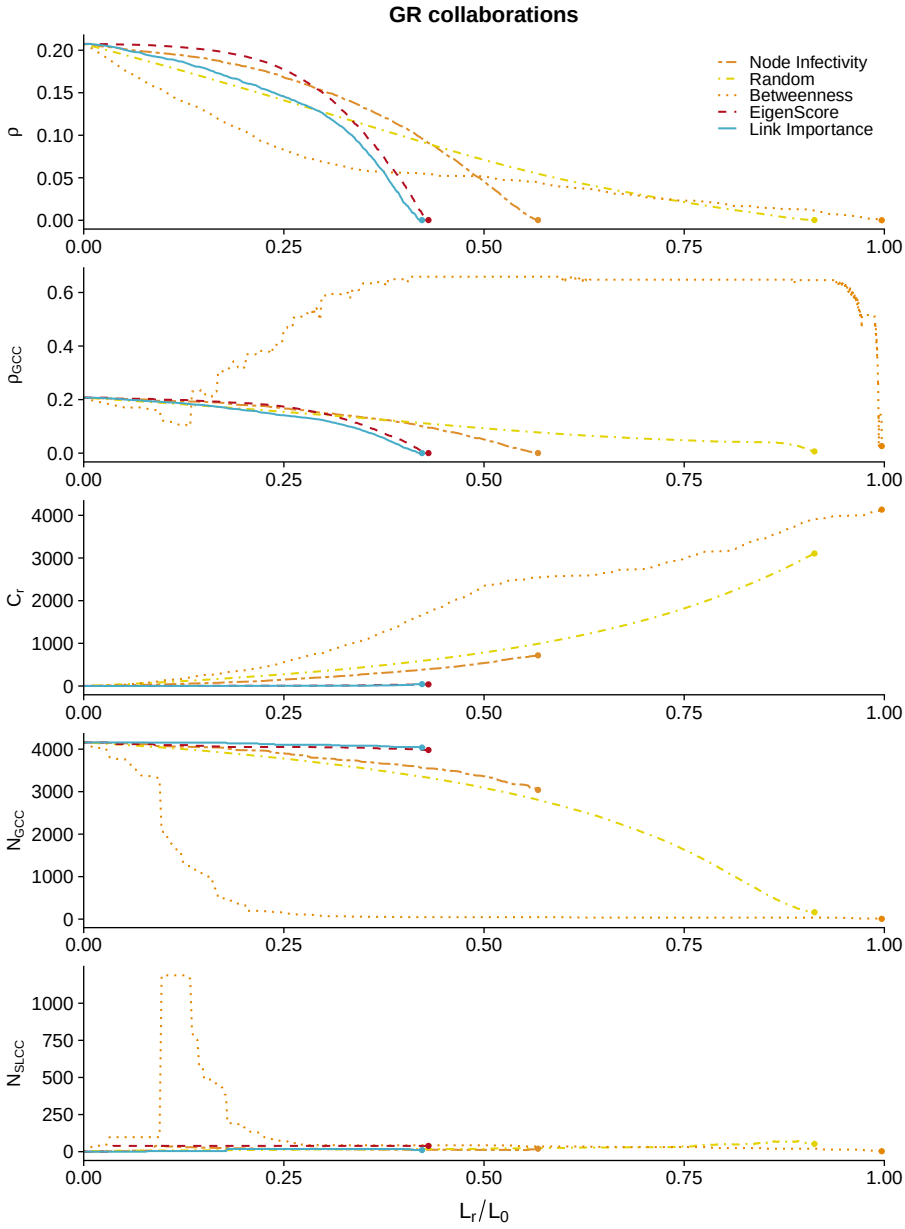


Figure 3.8: Epidemic containment for the general relativity collaborations network. The dots mark the achievement of total containment.

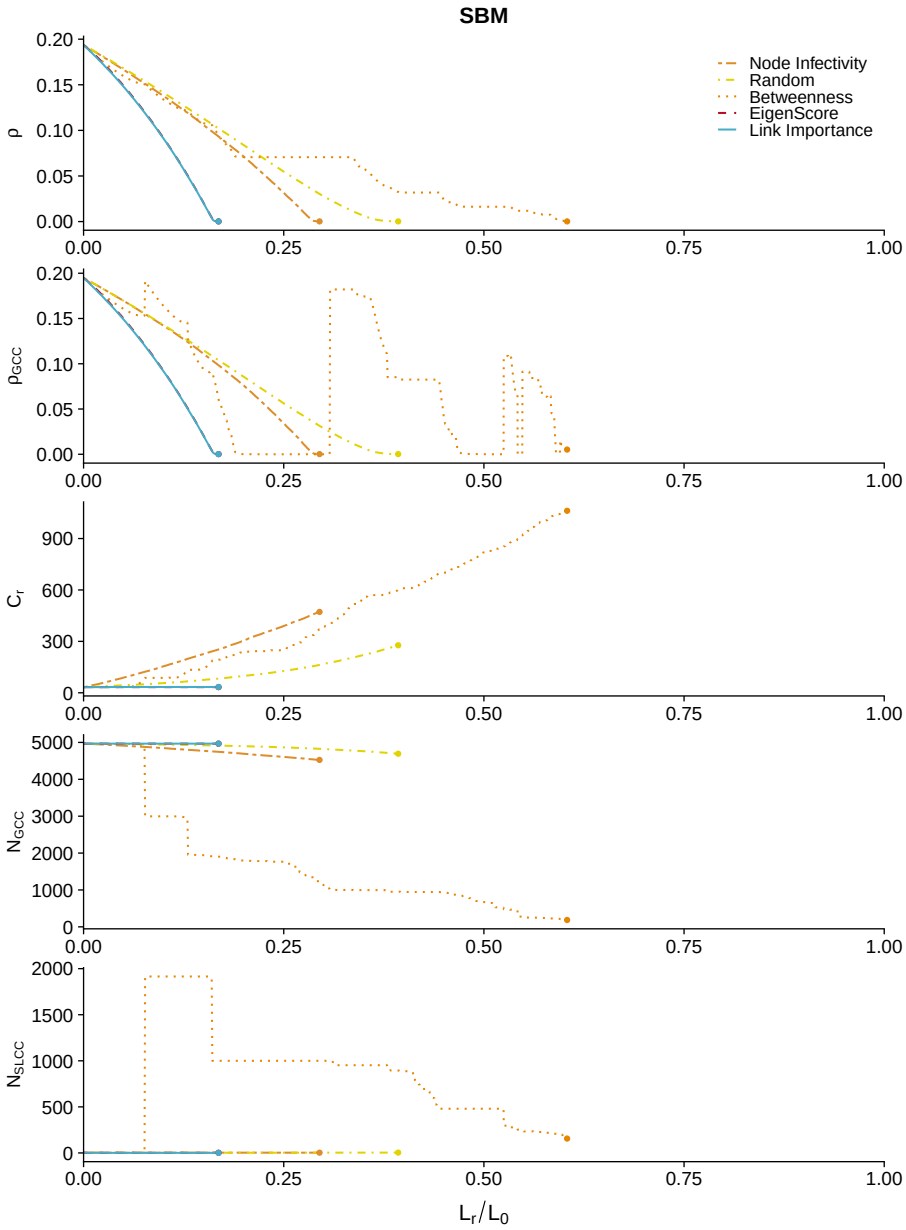


Figure 3.9: Epidemic containment for a network with 5000 nodes generated with a Stochastic Block Model, with 4 blocks of 250 nodes, 2 blocks of 1000 nodes, and 1 block of 2000 nodes, average degree 5 and mixing probability 0.3. The dots mark the achievement of total containment.

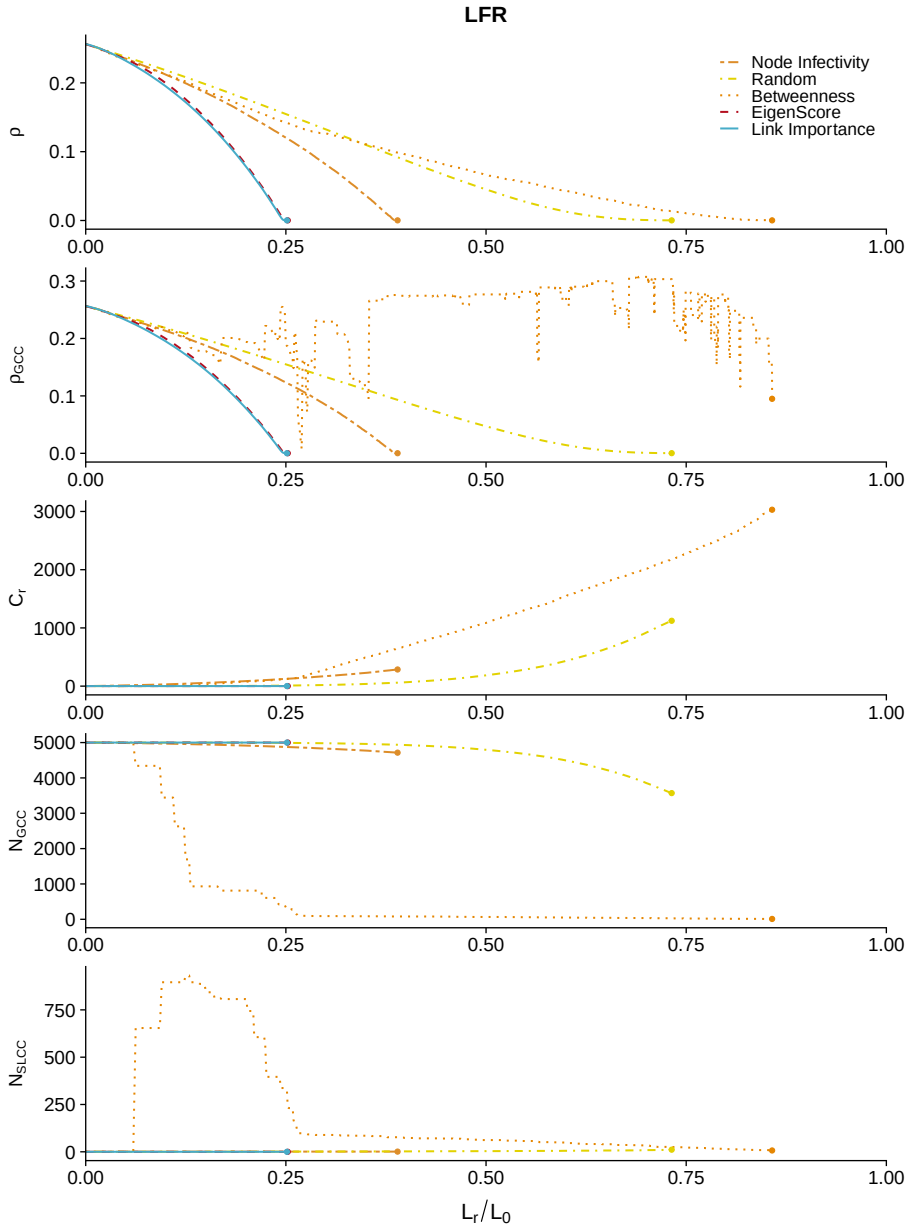


Figure 3.10: Epidemic containment for a network with 5000 nodes generated using the LFR algorithm [83], with average degree 6, exponent 3, and mixing probability 0.1. The dots mark the achievement of total containment.

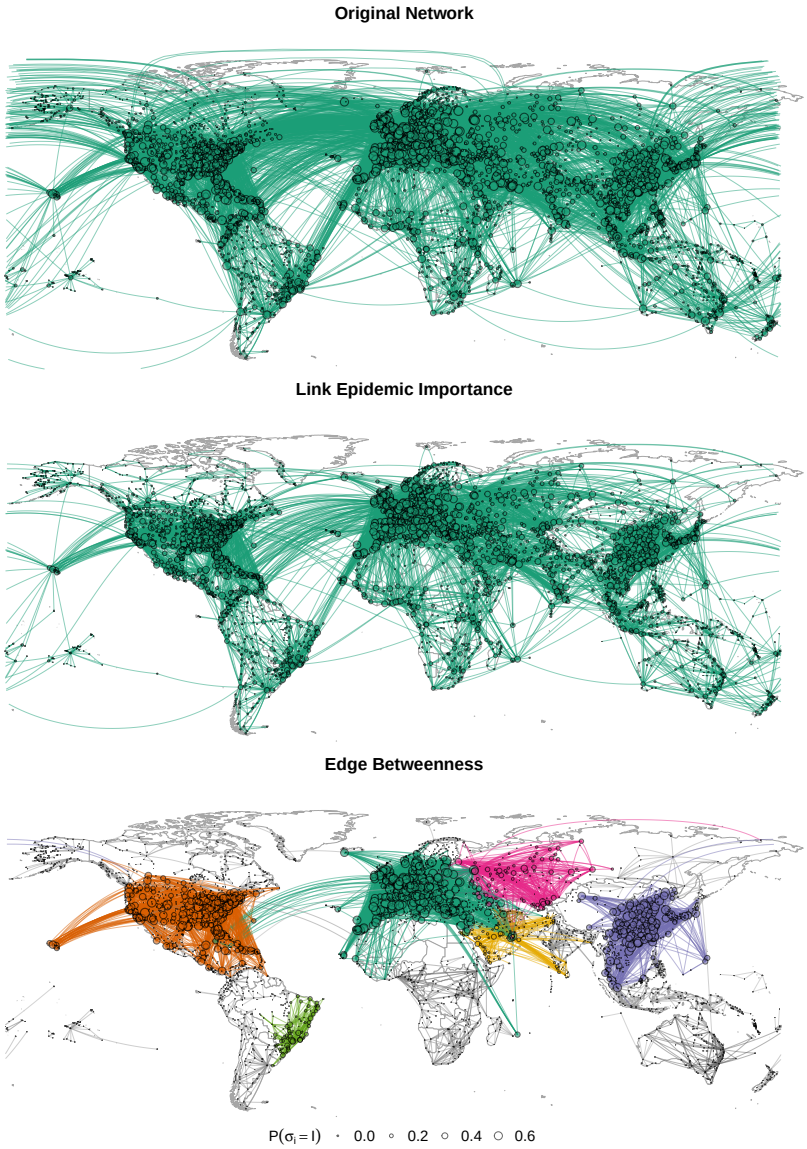


Figure 3.11: Original air transportation network (top) and the results after a removal of 33.3% of the links using link epidemic importance (middle) and edge betweenness (bottom). Nodes and edges with the same color belong to the same connected component, with subcritical components in gray scale and using darker gray for larger components. The area of the nodes is proportional to their probability of being infected. We have set the epidemic parameters to $\mu = 0.5$ and $\beta = 0.06$.

containment strategy proposal, Figure 3.11 (middle). As it is observed, the global connectivity, and thus functionality, of the worldwide connections is preserved (links of the same color are part of the same connected component). In Figure 3.11 (bottom) we plot the network after deactivating the same fraction of links (33.3%) using the recursive deactivation of links according to edge betweenness. The edge betweenness containment method, in contrast with our proposal, generates two main kinds of components: small or sparsely connected subcritical modules like the ones in Australia, Africa or South America, where the epidemics vanishes; and large supercritical communities in Europe, North America and Easter Asia, with a large prevalence of the epidemics. This means that, for instance, there is no path to go from London to New York, or from Tokyo to Los Angeles, thus disconnecting the world by air transportation.

For a better assessment of the performance of the different containment strategies, we show in Figure 3.12 their comparison in terms of the required fraction of removed links to attain total containment, L_{TC}/L_0, when applied to a large set of epidemic parameters and synthetic networks. Those networks have been generated with the model in [58], which interpolates between Erdős-Rényi (ER) and Barabási-Albert (BA) networks. In this way, we are able to evaluate the performance of the containment strategies on networks with degree distributions which range from Poisson (ER) to power-law (BA). Notice that these networks have no community structure and presents low transitivity. The observed results, 3.12, point to a clear advantage of the link epidemic importance method over the node infectivity and random approaches, and better or equal results with respect to edge betweenness and eigenscore. In fact, only eigenscore achieves results comparable to link epidemic importance, with a slight advantage for our method. Another interesting metric to evaluate the performance of containment methods is the number of connected components on the network when the total epidemic containment is achieved. In Figure 3.13, we can see how the use of link epidemic importance outperforms in many configurations the other methods, meaning that our method preserves better the functionality of the network.

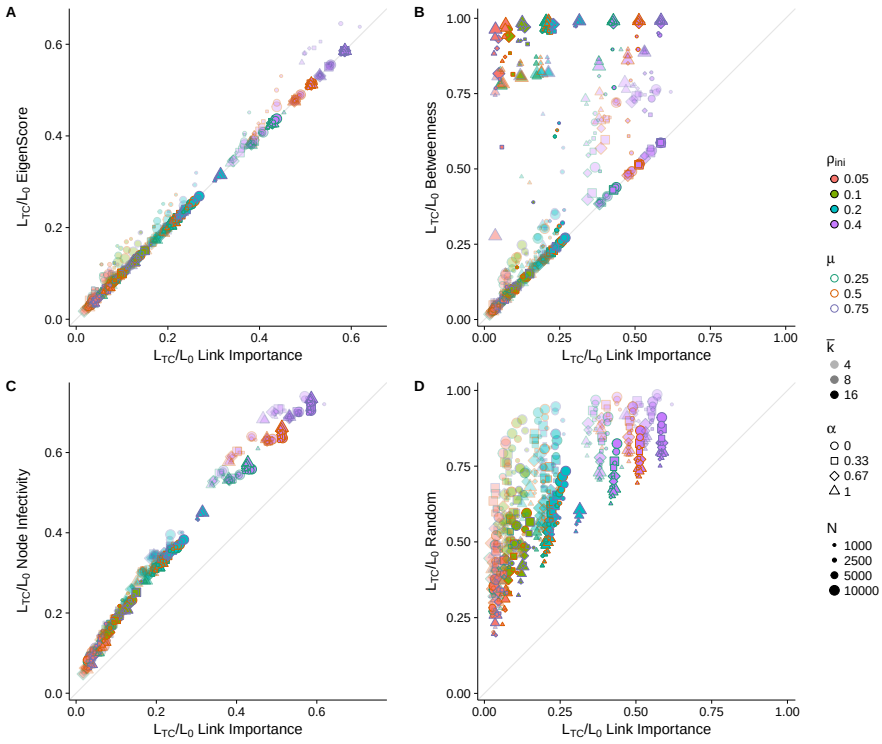


Figure 3.12: Fraction of links removed for total epidemic containment on synthetic networks. We show the fraction of links which have to be removed to obtain total epidemic containment using link epidemic importance, compared with the fractions for the other four strategies: (A) eigenscore; (B) edge betweenness; (C) node infectivity; (D) random removal. Each point represents a configuration consisting of a network and a set of epidemic parameters. The networks have been generated with the model in [58], which interpolates between Erdős-Rényi ($\alpha = 1$) and Barabási-Albert networks ($\alpha = 0$). We use four values of the interpolating parameter: $\alpha = 0.0, 0.33, 0.67, 1.0$. These networks are generated in four sizes ($N = 1000, 2500, 5000, 10000$) and three average degrees ($\langle k \rangle = 4, 8, 16$), thus amounting 48 different networks. For each network, we apply the five containment strategies for three different values of the recovery probability ($\mu = 0.25, 0.50, 0.75$), and four values of the infection probability β selected such that, before removing links, the incidence of the epidemics at the stationary state is equal to $\rho_{\text{ini}} = 0.05, 0.1, 0.2, 0.4$. Therefore, each plot contains 576 different configurations.

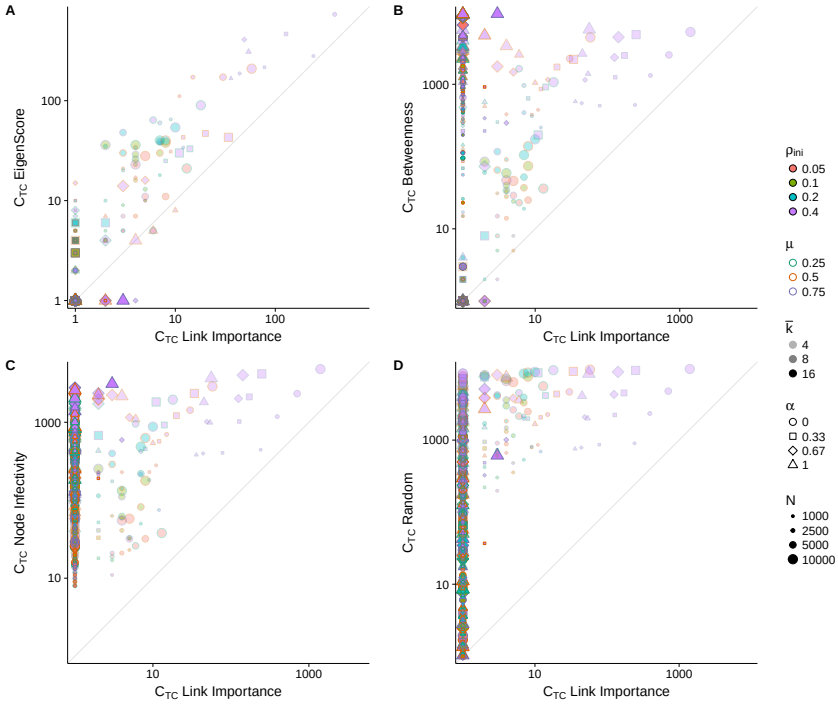


Figure 3.13: Comparison of the number of connected components after total containment between the link epidemic importance strategy and the other four methods. Each point represents a configuration consisting of a network and a set of epidemic parameters. The networks have been generated with the model in [58], which interpolates between Erdős-Rényi ($\alpha = 1$) and Barabási-Albert networks ($\alpha = 0$). We use four values of the interpolating parameter: $\alpha = 0.0, 0.33, 0.67, 1.0$. These networks are generated in four sizes ($N = 1000, 2500, 5000, 10000$) and three average degrees ($\langle k \rangle = 4, 8, 16$), thus amounting 48 different networks. For each network, we apply the five containment strategies for three different values of the recovery probability ($\mu = 0.25, 0.50, 0.75$), and four values of the infection probability β selected such that, before removing links, the incidence of the epidemics at the stationary state is equal to $\rho_{\text{ini}} = 0.05, 0.1, 0.2, 0.4$. Therefore, each plot contains 576 different configurations.

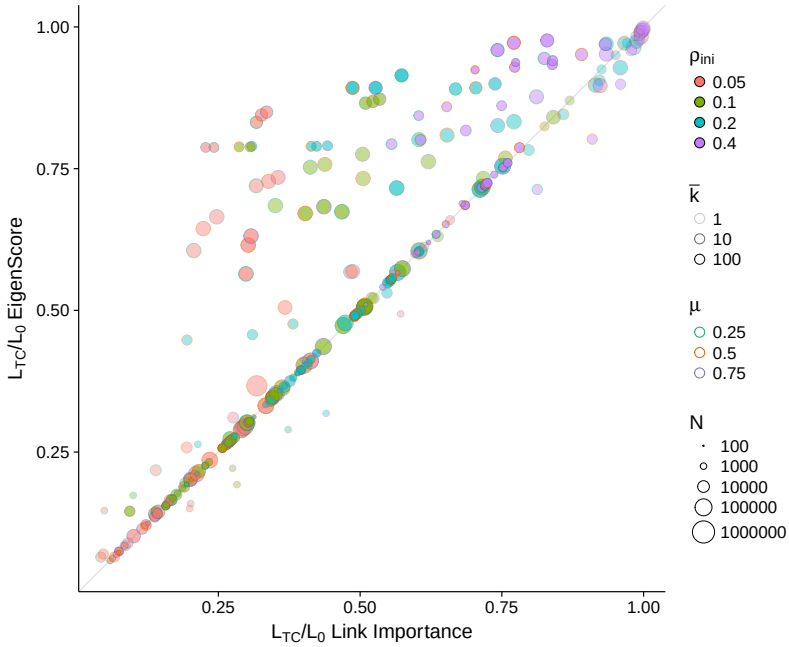


Figure 3.14: Fraction of links removed for total epidemic containment on real networks. We compare the link epidemic importance and eigenscore methods on a set of 27 real networks selected from the Network Repository [131], with sizes ranging from 410 to 404719 nodes. The epidemic parameters are the same as in Figure 3.12, thus amounting 324 different configurations.

We have repeated the same experiments over 27 real networks in Figure 3.14 obtained from the Network Repository [131], selecting only the largest connected component. They cover wide ranges of number of nodes (from 410 to 404719), number of links (from 1043 to 713319), average degree (between 2.04 and 84.82), average clustering coefficient (from 0.0023 to 0.1105) and assortativity (between -0.88 and 0.64), see Appendix A.1 for further details. The differences between link epidemic importance and eigenscore become more evident, showing again the effectiveness of our approach, but with some exceptions in which eigenscore performs better, see Figure 3.14. As in the case of synthetic networks, we have also analyzed the number of connected components when the total containment is achieved, Figure 3.15, which show similar results, a better performance of link epidemic importance to keep low the number of components, with just a few exceptions.

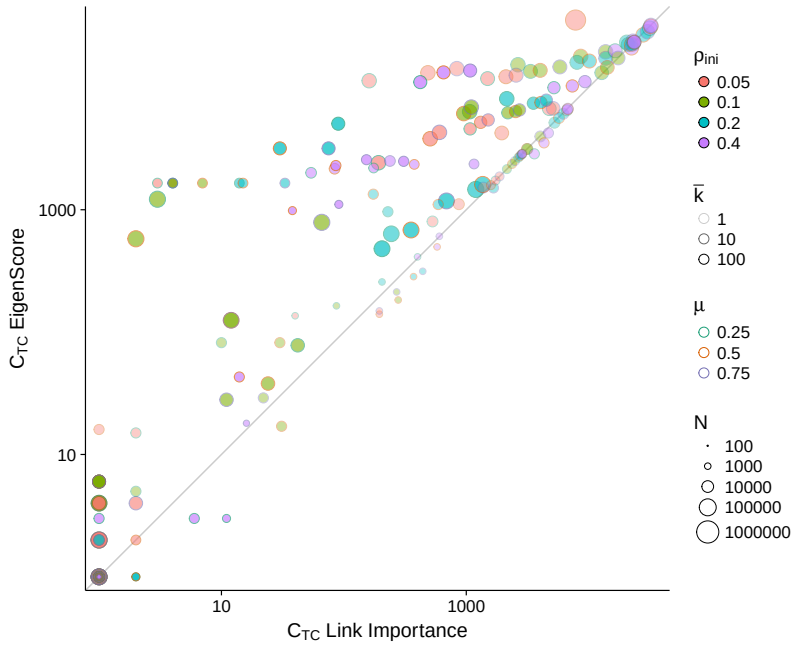


Figure 3.15: Comparison of the number of connected components after total containment between the link epidemic importance and eigenscore strategies. We compare the link epidemic importance and eigenscore methods on a set of 27 real networks selected from the Network Repository [131], with sizes ranging from 410 to 404719 nodes. The epidemic parameters are the same as in Figure 3.14, thus amounting 324 different configurations.

3.4 SUMMARY

We have presented a methodology for assessing epidemic spreading based on links instead of nodes. The model, named Epidemic Link Equations (ELE), allows the determination of the epidemic importance of each link in transmitting the disease. The method accounts for the first order correlations between links, although it could be extended to higher orders assuming a larger analytical and computational cost. The results are used to develop an epidemic containment strategy consisting in deactivating recursively the links with largest link epidemic importance while preserving the connectivity of the full network, i.e., avoiding fragmentation. We have validated our proposal in synthetic and empirical networks, comparing with other alternative containment

strategies, which show its better performance, with few exceptions. In the empirical case of the worldwide air transportation network, we identify the most important connections between airports for the spreading of epidemics and evaluate the epidemic incidence after its deactivation, considering an SIS epidemic spreading dynamics. Our results open the door to new approaches in the analysis of dynamical diffusive-like models on complex networks at the level of links instead of nodes.

3.5 PERSONAL CRITICAL VIEW

Despite the results obtained with our approach, some limitations need to be mentioned. Regarding the Link Epidemic Importance, it has to be noticed that the measure takes into account the effect over second neighbors. To ensure that the links that show higher link epidemic importance are the ones accountable for the major impact on the spreading of the epidemics, we should consider the full cascade of infections triggered by an initial spread through each link. Another point that has to be mentioned is the strong relation between EigenScore and link epidemic importance in some of our experiments. That could indicate the existence of some spectral characteristics of the networks that are intrinsically correlated with our measure. It is also worth to mention that our approach is not applicable to networks with a regular structure, like the grids. There, the dynamical correlations are too strong for our probabilistic assumptions to hold. Finally, we have to comment that the temporal scale used on the analysis of the containment process requires the arrival to the stationary state. However, on real-world scenarios, the actuation on the epidemics has to be done as fast as possible to stop the spreading before it becomes endemic; consequently, a revision of the containment procedure should consider the effects of the transitory state.

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

4

TEMPORAL HIGHER-ORDER: ADAPTATIVE MEMORY MODEL

4.1 INTRODUCTION

When we talk about dynamics running on top of complex networks, we usually focus on the state of the system after the dynamics reach an equilibrium point, regardless of the amount of time needed to achieve this stability, or the different transient states previously visited. Nevertheless, these transient states play a central role in the understanding of how many dynamical systems work. From biological pathways to human-mobility, the chain of states that precede the actual state of the system it is crucial to understand the future steps.

In this chapter, we will focus on human mobility to understand the critical role of previous locations on the mobility patterns to understand where individuals are going to move. They are decisive because, as we will see, humans mobility tends to display routine paths composed by a series of locations, e.g., from home to school, from school to work, from work to supermarket, from the supermarket to home again. To capture these n location patterns, we need to introduce models capable of dealing with this temporal higher-order complexity.

Regarding human displacements, not just the sequence of locations is important, the amount of time spent at each location it is also relevant when we consider the effect of other coupled dynamics that can run at the same time. For example, in epidemic processes, the longer the time an infected individual spends on a particular location, the more

probable is that he transmits the infection to other individuals at the same spot. Being capable of building models that may deal with this complexity, both the sequence of locations and time spent at each location is mandatory to analyze this kind of scenarios.

This chapter will provide a detailed explanation of how to proceed to model these situations. We propose a new method to deal with temporal higher-order mobility patterns, studying the effect of these patterns on epidemic spreading processes. We will apply our approach to mobile Call Detail Records (CDR) provided in the context of the D4D challenge.

4.2 HIGHER-ORDER MARKOVIAN MODEL

As we have said in the introductory chapter, first-order Markovian models have proved very useful to model human mobility. However, they have the fundamental disadvantage of not accounting for mobility memory because, according to the Markov property [93], the probability of future steps only depends upon the present state. Nevertheless, as we commented, recurrent patterns are inherent in human mobility: it is very likely that an individual moves to a neighboring area (by means of a car or public transportation) to work, and after a few hours he or she will come back to the original position. This memory is an intrinsic property of human mobility and must be taken into account for a realistic modeling of people movements between different geographic areas.

One method to overcome the issue of memory representation is using n-order Markovian models. On these models, the next position of an individual depends on the n previous visited locations:

$$P(X_{t+1}|X_t, X_{t-1}, \ldots, X_{t-(n-1)}) \tag{4.1}$$

When memory is taken into account, each physical node (e.g., $i \in \mathcal{A}$) is replaced by the corresponding state-nodes (e.g., $\mathrm{i} \triangleleft \mathrm{j} \in \tilde{\mathcal{A}}$ if memory is of order 2) encoding the information that an individual is in node i when he or she comes from j, see Figure 4.1e. While the first-order mobility matrix, $\mathbf{F}$, explained in Section 1.4.1.3, encodes information about

the network of n physical nodes, we need to introduce a new matrix $\mathbf{H}$ to represent information about the network of n^2 state-nodes, accounting for the allowed binary combinations (e.g. k $\triangleleft$ j, $j, k = 1, 2, ..., n$) between physical nodes. Similarly, higher-order memory can be taken into account by building appropriate matrices.

We use different mobility matrices to build different mobility models. Let $N_i(t)$ indicate the population of the physical node $i \in \mathcal{A}$ at time t, then the n mobility equations describing how the flux of people diffuses through the network are given by

$$N_i(t+1) = \sum_{j=1}^{n} F_{ji} N_j(t), \tag{4.2}$$

where $\mathbf{F}$ is the first-order Markovian model matrix.

In the case of $\tau-$memory, we indicate by $\tilde{N}_\alpha(t)$ the population of the state-node $\alpha \in \tilde{A}$ at time t and the n^τ mobility equations required to describe the same process are given by

$$\tilde{N}_\alpha(t+1) = \sum_{\rho=1}^{n^\tau} H_{\rho\alpha} \tilde{N}_\rho(t). \tag{4.3}$$

The population in each physical node at time t is given by the sum of the population in the corresponding state-nodes. It is worth remarking that, in general, the matrix $\mathbf{H}$ can be a function of time as well, and the equations would keep their structural form.

The way how matrices $\mathbf{F}$ and $\mathbf{H}$ are build differs. While to build a first-order matrix we count bigrams from data that stand for *from-to* patterns, to build a second-order matrix we count all the possible trigrams on the map that capture three locations patterns and we encode them on matrix H. Once this counting in completed, a proper normalization of the matrix produces the transition probabilities between different state nodes, keeping valuable temporal information about previous locations intact. An example of how these models are built can be seen in Figure 4.1.

The effect of memory has been shown to be relevant, for instance, to model mobility at country level, where individuals fly from one city to

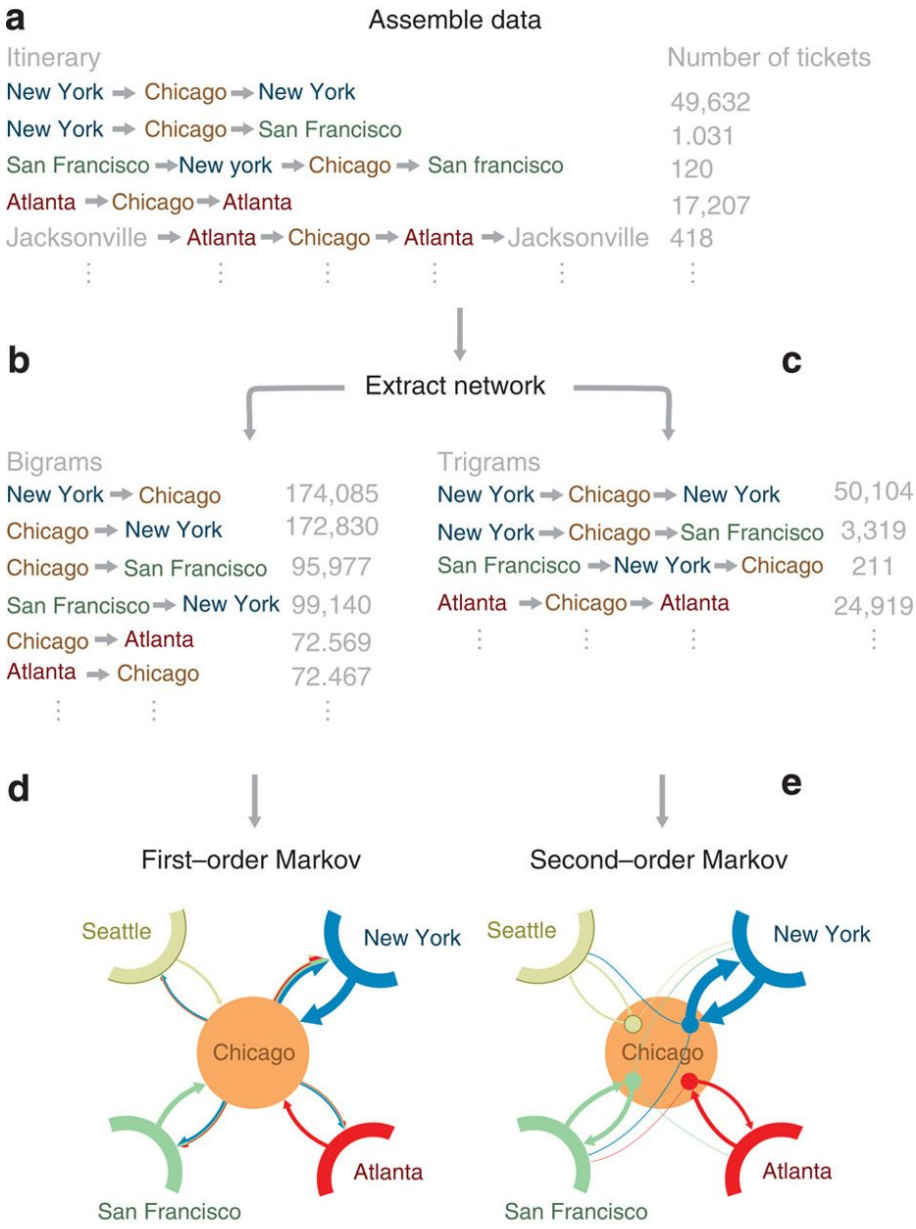


Figure 4.1: (a) Itineraries weighted by passenger number. (b) Aggregated bigrams for links between physical nodes. (c) Aggregated trigrams for links between memory nodes. (d) Network without memory. (e) Network with memory. *Reprinted by permission from [133].*

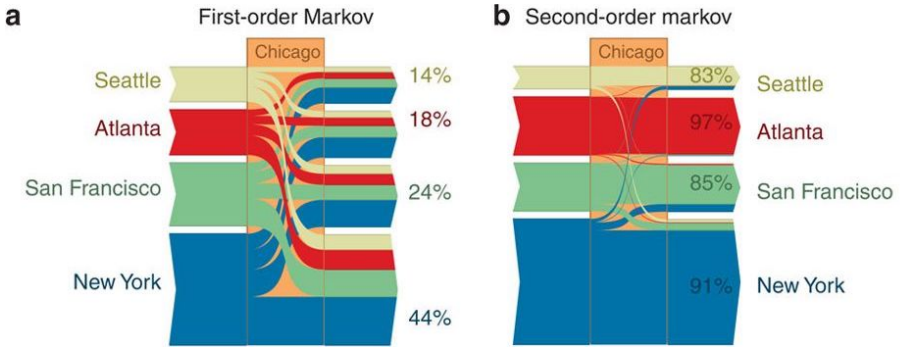


Figure 4.2: (a) In a first-order Markov approach, we model passengers' travel to a city to be proportional to the observed volume of traffic to that city, and irrespective of where the passengers come from. (b) In a second-order Markov model, passengers' travel to a city is still proportional to the traffic volume, but also dependent on where the passengers come from. In this example, out-and-back traffic to Chicago only dominates overtransfer traffic when second-order Markov dynamics are taken into account. The percentages represent the relative return flow. *Reprinted by permission from [133].*

another and often go back to their origin instead of moving towards a different city [133]. Looking at Figure 4.2, we can see a clear example of the gain produced by the higher-order model on these situations. The figure shows the counting of different trigrams that has Chicago as a middle location. We can clearly see how using first-order model, Figure 4.2a, we have a large set of different trigrams and also a low percentage of passengers that return to their original destinations. If we use a second order model to capture the mobility patterns, Figure 4.2b, it is evident that a large percentage of passengers come back to their previous locations; thus, the patterns shown by the first-order approach turn out to be spurious.

Additionally, the usage of higher-order models has a deep impact on the way how we understand human mobility when we combine them with other networks techniques. As an example, if we combine these models with community detection algorithms that make use of flow information [37, 132], like the one provided by mobility models, we can observe how the use of second-order dynamics contributes to find smaller and more informative modules [133], helping to unveil the true nature of some mobility patterns.

4.3 ADAPTIVE MEMORY MODEL OF HUMAN MOBILITY

Spatial human mobility is quite complex and, depending on the data used, (higher-order) Markovian dynamics might not be suitable to model peculiar patterns such as returning visits and conditional waiting times, i.e. the probability to stay in a location depending on the origin of the travel. This effect is clear if we consider data that has temporal information and not just a sequence of changes in location, like the one obtained by GPS tracking, geolocalized tweets or geolocalized phone calls. This last one is specially interesting since the collaboration between researchers and mobile operators recently opened new promising directions to gather information about human movements, country demographics and health, faster and cheaper than before [2, 21, 25, 41, 45, 46, 63, 154, 163, 165]. In fact, mobile phones heterogeneously penetrated both rural and urban communities, regardless of richness, age or gender, providing evidences that mobile technologies can be used to obtain real-time information about individual's location and social activity, in order to build realistic demographics and socio-economics maps of a whole country [164]. Mobile data have been successfully used in a wide variety of applications, e.g., to estimate population densities and their evolution at national scales [41], to confirm social theories of behavioral adaptation [44] and to capture anomalous behavioral patterns associated to religious, catastrophic or massive social events [43]. Even more recently, the public availability of mobile phone data sets further revolutionized the field, e.g., by allowing ubiquitous sensing to map poverty, to monitor social segregation and to optimize information campaigns to reduce epidemics spreading [2, 86], to cite just some of them [21].

Despite some limitations, mobile phone data still constitutes one of the most powerful tools for sensing complex social systems, and represent a valuable proxy for studies where human mobility plays a crucial role [9, 10, 30, 31, 49, 63, 86, 140, 147, 151, 154, 165]. Milestone works in this direction have shown that human trajectories exhibit more temporal and spatial regularity than previously thought. Individuals tend to return to a few highly frequented locations and to follow simple reproducible patterns [63, 87], allowing a higher accuracy in predicting their movements [147] and significantly affecting the spreading of

transmittable diseases [10]. However, the increasing interest for using mobile phone data in applications should be accompanied by a wise usage of the information they carry on. In fact, an inadequate model accompanied by incomplete data and scarce knowledge of other fundamental factors influencing the model itself, might lead, for instance, to a wrong estimation of the incidence of an epidemics and its evolution [111].

Let us consider, for instance, that we want to use as proxy of the mobility the call sequence *BBBBCCCSSS* made by an individual traveling between three American cities: Chicago, Boston and San Antonio. The main drawbacks of Markovian models – of order lower than three – become evident in a scenario like this one, because the number of consecutive calls in the same city exceeds the memory of the model and, thus, the spatial information about previously visited locations is lost. Clearly, in presence of more complicated patterns, increasing the order of the model will not solve the issue and some information will be inevitably lost. Alternatively, we could aggregate consecutive calls in the same place to a single identifier, e.g. the previous sequence would be reduced to *BCS*. In this case, a Markovian model would preserve the spatial information and correctly identify the transitions between the three cities, at the price of losing information about how many calls have been made in each place.

In absence of detailed temporal information about calling activity, the number of consecutive calls in a specific location can be used as a proxy: the higher the number of calls, the larger the waiting time. The temporal information about the amount of time spent in each location is critical for many dynamical processes like spreading or congestion. We assert that this time, like the next visited location, is conditioned by previous movements of the individuals. To illustrate this, we use the example shown in Figure 4.3, where people from three different places (nodes blue, green and orange) go to the same destination (node red), stay some time in there, and come back to the origin of their trip. The self-loops in the central (red) node represent the time spent there, the color encoding individuals coming from different origins and the size encoding the amount of time spent. For instance, individuals coming from the blue node wait more than individuals coming from the green

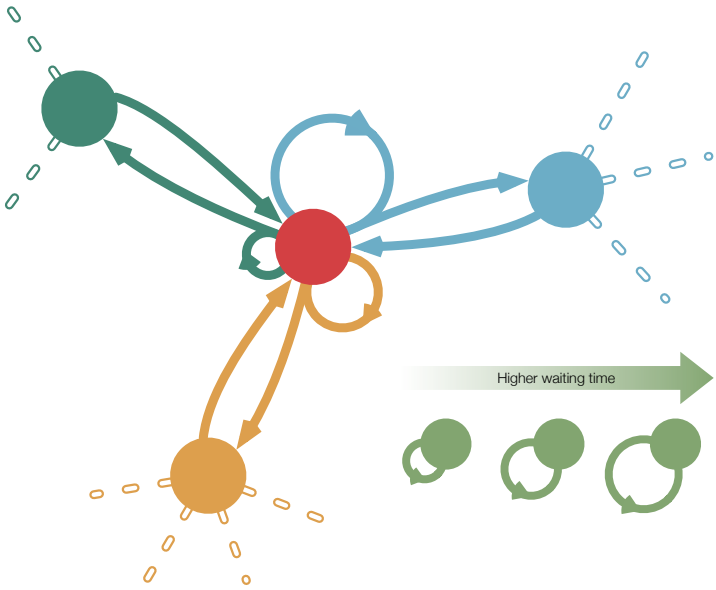


Figure 4.3: An example of human mobility between four different places. Individuals from green, blue and orange nodes move to the red central node and, after some time, go back to their previous location. The amount of time spent in the red node by individuals coming from the other nodes depends on their previous location, and it is represented by self-loops of different size.

node. This type of dependence is what we call conditional waiting time.

To better appreciate this fact, let us consider holiday trips. Individuals making expensive intercontinental trips tend to spend more time visiting the destination than individuals making cheaper trips, achieving a good trade-off between the travel cost and the time spent. Another emblematic case is urban mobility. For instance, the red node might be an expensive commercial area, the green node a wealthy neighborhood and the blue node a less wealthy area. In this scenario, that should be considered only for illustrative purposes, individuals coming from the less wealthy area are more likely to be qualified workers in the commercial one, with long and frequent visits. Conversely, individuals from the wealthy are more likely to make unfrequent and shorter visits for shopping, for instance.

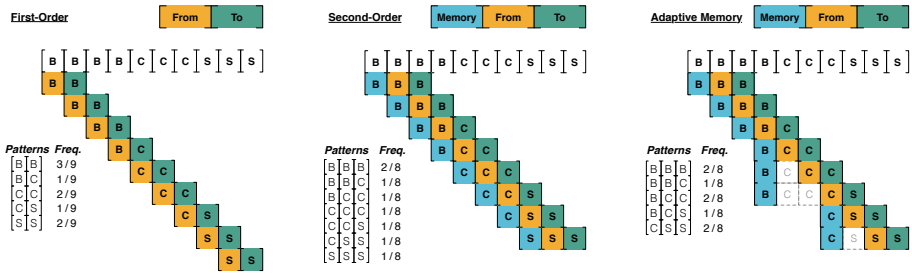


Figure 4.4: Mobility models built from a representative sequence of mobile phone calls (BBBBCCCSSS) made, for instance, by an individual during travels between three American cities, namely Chicago (C), San Antonio (S) and Boston (B). Let us focus on the pattern $S \leftarrow C \leftarrow B$, that is the real sequence of movements in the geographical space. The first-order model predicts a probability of $\frac{1}{81}$, the second-order model a probability of $\frac{1}{64}$, whereas the adaptive 2-memory estimates a probability of $\frac{1}{8}$, closer to observation.

The importance of accounting for conditional waiting times will be evident later, when we will consider the spreading of an epidemics in a country.

We propose a mobility model [96], that we call *adaptive memory*, able to account for conditional waiting times. At first order, the method is equivalent to a classical first-order Markovian model, whereas significant differences emerge for increasing memory with respect to standard approaches. For instance, at second order, the 2-memory mobility matrix is built between all possible pairs of nodes (2-states), as in a standard second-order Markovian model. However, instead of considering transitions between areas in the sequence of calls, as a second-order Markovian model does, transitions in the sequence of distinct geographical areas are considered. This point is crucial, and we better clarify it with the example shown in Figure 4.4, where the differences between adaptive memory and Markovian models, in terms of probability assigned to different mobility patterns, are reported.

The importance of such differences is reflected in the ability of each model to predict successive individual movements. In fact, the presence of spurious or under-represented patterns might significantly affect the results, as shown in Figure 4.5. In this example, two sequences of phone calls generated by two different users moving between three cities – B,

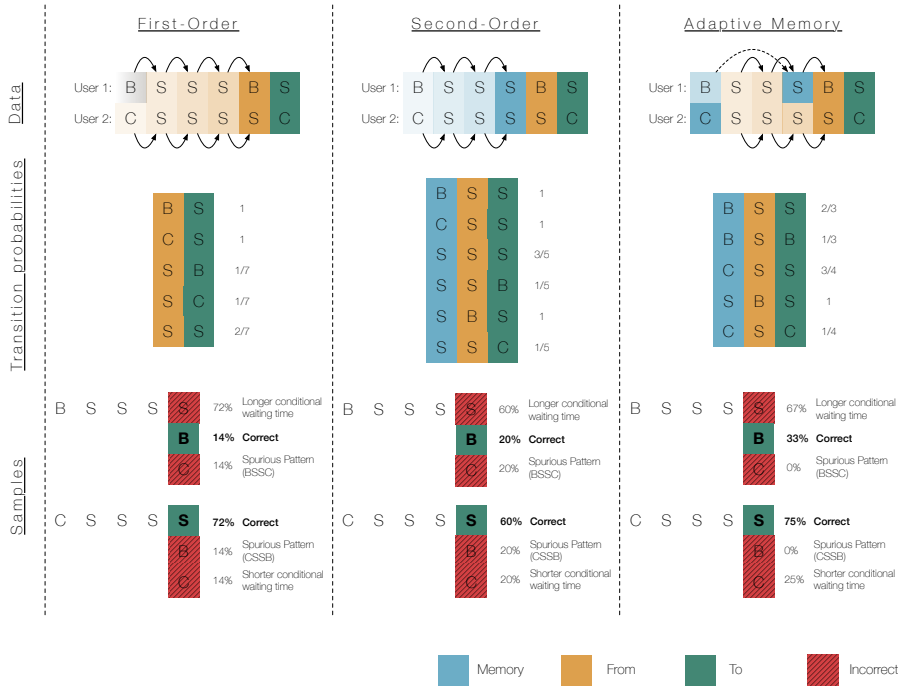


Figure 4.5: Using the sequence of calls made by two different users (1 and 2) – starting from two different locations (B and C) and visiting a new location S – we build first-order and second-order Markov models, as well as the adaptive memory one. We use each mobility model to generate the possible mobility sequences. Given that there are two empirical starting points, we originated the sampled sequences in B and C, respectively. In the figure, for each sample, we report the fraction of times it is reproducing observation ("Correct"), it is a non-observed mobility pattern ("Spurious Pattern") and it is underestimating or overestimating waiting times ("Longer/Shorter Conditional Waiting Time").

C and S – are considered. Markovian models generate spurious patterns that are never observed in the data, issue not affecting the adaptive memory model by construction. Morover, our approach predicts the next movement with more accuracy than Markovian ones, because it correctly takes into account conditional waiting times.

The difference between the adaptive memory and Markovian models becomes more evident when the corresponding transition matrices are compared. There is no difference at the first order, thus we will focus on

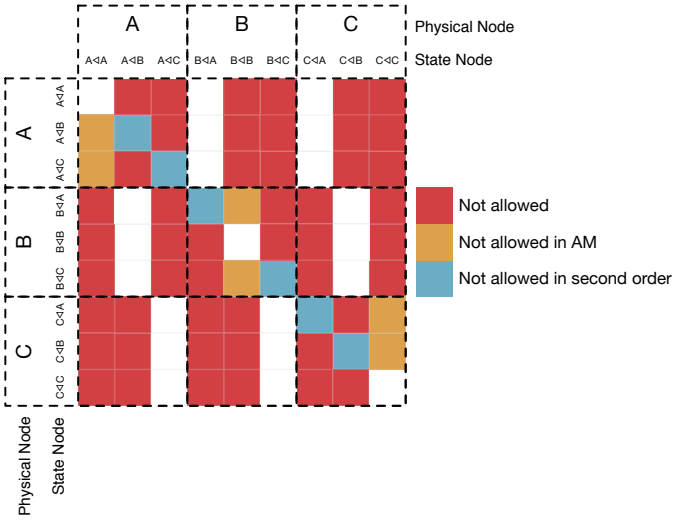


Figure 4.6: Second-order transition matrix for three physical nodes A, B and C. The state nodes are represented with the notation x $\triangleleft$ y meaning that walkers in this node have traveled from node y to node x. The cells in red are not used by either second-order markovian model or adaptive memory model. The cells in bluish are used only in adaptive memory model, while, the cells in orange are used only in the second-order Markovian model. The cells in white are used by both models.

the comparison between τ-order Markovian and adaptive $\tau-$memory models, in the following.

In both models, the number of possible transitions between state nodes is the same and equals $n^{2\cdot(\tau-1)}$, where n is the number of physical nodes. For instance, in second-order models, there are $n^2 \times n^2$ transition matrices with n^3 possible transitions between state nodes, as shown in Figure 4.6. However, the way how each model stores repeating calls in the same physical node is very different. While adaptive memory stores this information into the n^τ diagonal elements of the matrix, encoding the conditional waiting times discussed in the previous section, Markovian models redistribute this information among off-diagonal entries, because they do not allow this type of self-loops by construction. See Figure 4.7 for a comparison of how are represented internally the conditional waiting times.

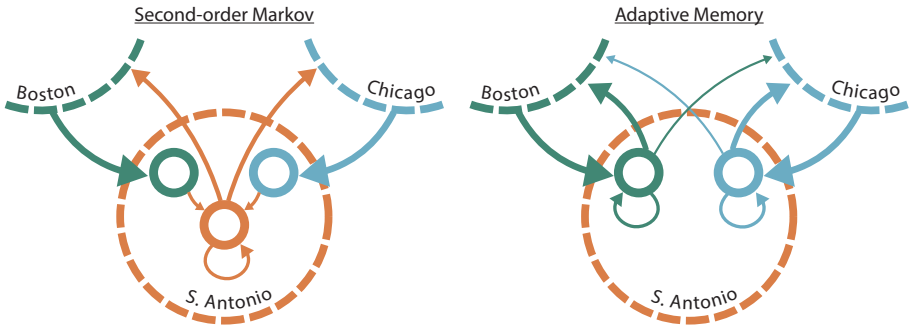


Figure 4.7: Comparison between the Second-order Markovian and Adaptive memory model in terms of allowed transitions between the internal states of a particular physical node. Physical nodes are delimited by a dashed line while solid lines define states nodes, the color of which represent the physical node from where the individuals came from. People that moves from Boston (green) and Chicago (blue) to San Antonio (orange) are assigned to their representative internal states nodes on the San Antonio physical node. On the Second-order model (left), if individuals perform two or more consecutive calls on San Antonio, they will transition to another internal state where the track of previous visited locations is lost, and any conditional information on conditional waiting times will be vanished. Using adaptive memory model (right), the transitions between state-nodes of the same physical node are forbidden, then individuals can keep track of previous visited locations while, at the same time, the conditional waiting time is stored on the self-loop.

More specifically, the information is redistributed among transitions between state nodes of the same physical node. The entries of off-diagonal blocks – corresponding to transitions between state nodes of different physical nodes – are the same in both models. Therefore, while the stationary probability of finding a random walker in a physical node is not different in the two models, it is different at the level of state nodes and, as we will see later, this significantly affects diffusion processes such as epidemics spreading.

4.3.1 *From Adaptive memory to Markovian models*

The previously discussed transition matrix features allow us to see τ-order Markov models as degenerated cases of τ-order adaptive memory

models, where the temporal information is aggregated according to a well-defined procedure. Through this procedure, we will derive the Markovian transition matrix $\mathbf{S}_{m\times m}$ from the transition matrix $\mathbf{A}_{m\times m}$ of the adaptive memory model. Here, we define $m = n^\tau$ as the number of state nodes in the system. For the sake of simplicity, we assume that $\tau = 2$. Nevertheless, the procedure can be easily extended to any arbitrary order.

If the state node network is fully-connected and does not have any absorbing state, the probability of finding a random walk in a given physical node will be the same for both models. In other words, if $\mathbf{u}^{[a]}$ is the normalized leading eigenvector of the transition matrix $\mathbf{A}$, i.e. the stationary distribution, and $\mathbf{u}^{[s]}$ is the normalized leading eigenvector of $\mathbf{S}$, then $\mathbf{M}\mathbf{u}^{[a]} = \mathbf{M}\mathbf{u}^{[s]} = \mathbf{v}$, where $\mathbf{v}$ is the stationary distribution of the first-order transition matrix $\mathbf{F}_{n\times n}$, and $\mathbf{M}_{n\times m}$ is an aggregation matrix defined as:

$$m_{i,j} = \begin{cases} 1 & i = \left\lceil \frac{j}{n} \right\rceil \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

where $\lceil \cdot \rceil$ is the ceiling function.

The idea is that the probability of finding a random walker in a given physical node is the sum of the probabilities of finding that walker in any of the state nodes which compose it.

Now, we define $\tilde{\mathbf{u}}^{[a]}$ as the state node stationary distribution normalized at physical level. That will give us the probability of finding a walker in a given state node of the physical node i knowing that the walker currently is in that physical node:

$$\tilde{u}_i^{[a]} = \frac{u_i}{v_{\lceil i/n \rceil}} \tag{4.5}$$

It has to be recalled that not all the transitions between state nodes are allowed in second-order and adaptive memory models, i.e., of n^4 possible state-node transitions, just n^3 are allowed. The restrictions about these transitions are different for both models, as we have mentioned

before. This fact enables us to reduce the $m \times m$ adjacency matrix $\mathbf{A}$ to a reduced form $\widehat{\mathbf{A}}_{m\times n}$.

$$\widehat{a}_{i,j} = \begin{cases} a_{i,i} & j = \left\lceil \frac{i}{n} \right\rceil \\ a_{i,\left\lceil \frac{i}{n} \right\rceil + (j+1)n} & \text{otherwise} \end{cases} \tag{4.6}$$

As has been said, the main difference between transition matrices $\mathbf{A}$ and $\mathbf{S}$ is the way how they encode transition times. To reduce adaptive memory to a Markov model, we have to create a transition matrix $\mathbf{T}^{[a\to s]}_{m\times m}$ which aggregates the independent transitions times of adaptive memory according to the normalized stationary probabilities $\tilde{\mathbf{u}}^{[a]}$. The intuition is that the aggregation in each physical node has to be weighted according to the volume of visitants of each of their sources. Note that, if the transition does not have an associated waiting time, the transformation leaves it as it is:

$$t^{[a\to s]}_{i,j} = \begin{cases} \tilde{u}^{[a]}_i & \left\lceil \frac{i}{n} \right\rceil = \left\lceil \frac{j}{n} \right\rceil \text{ and } i = (n+1)\left(\left\lceil \frac{j}{n} \right\rceil - 1\right) + 1 \\ 1 & i = j \text{ and } i \neq (n+1)\left(\left\lceil \frac{j}{n} \right\rceil - 1\right) + 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

To obtain the reduction form $\widehat{\mathbf{S}}_{m\times n}$ of the second-order transition matrix, we have to multiply the transformation matrix $\mathbf{T}^{[a\to s]}$ by the reduced transition matrix of the adaptive memory model $\widehat{\mathbf{A}}$:

$$\widehat{\mathbf{S}} = \mathbf{T}^{[a\to s]} \cdot \widehat{\mathbf{A}} \tag{4.8}$$

Finally, we have to expand the reduced from of the transition matrix to a complete form $\mathbf{S}$:

$$s_{i,j} = \begin{cases} \widehat{s}_{i,\lceil j/n \rceil} & j = \left\lceil \frac{i}{n} \right\rceil + \left(\left\lceil \frac{j}{n} \right\rceil - 1\right)n \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

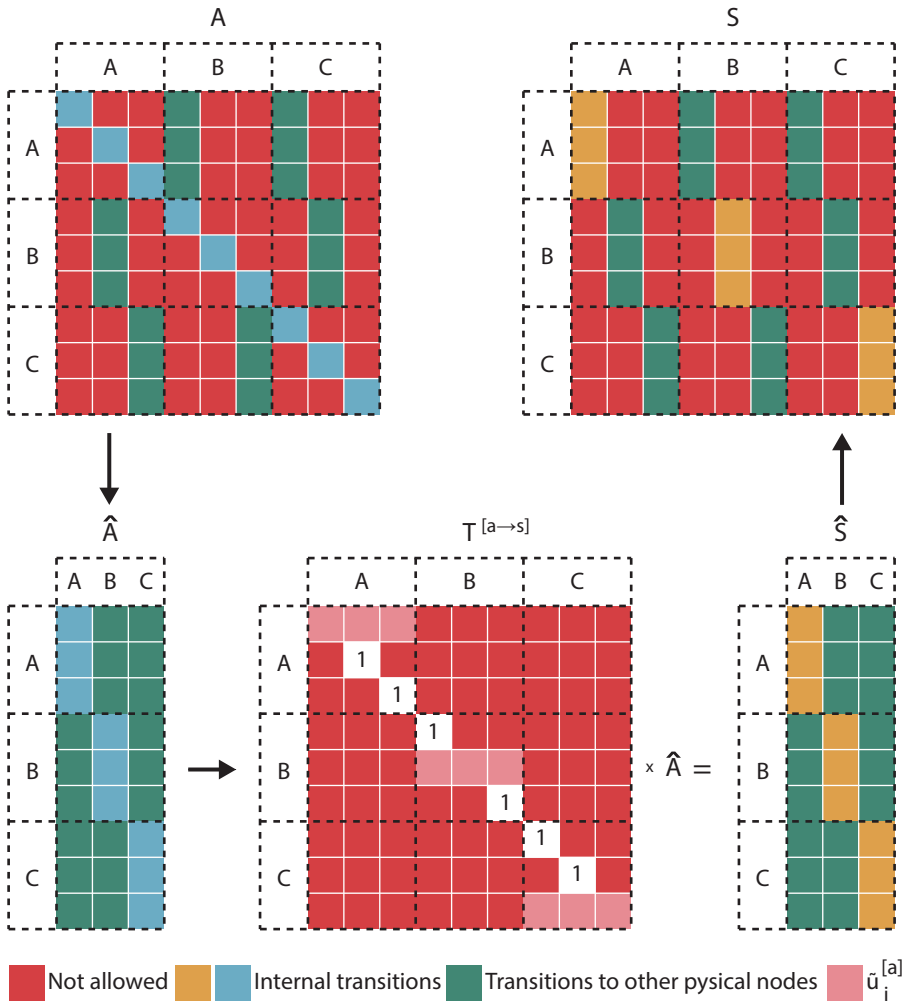


Figure 4.8: Graphical representation of the degradation procedure of an adaptive memory (matrix **A**) to a second-order Markovian model (matrix **S**). Red cells represent not allowed transitions in both models, those in orange and blue represent transitions between state nodes of the same physical node. Cells in green call for transitions between different physical nodes. Finally, cells in pink stand for the state node stationary distribution normalized at physical level, $\tilde{\mathbf{u}}^{[a]}$.

4.4 IMPORTANCE OF TEMPORAL HIGHER-ORDER DYNAMICS

4.4.1 *Dataset*

In this section, we will quantify the impact of adaptive memory on human mobility modeling by using data sets provided by the Data for Development Challenge 2014 [100] and some supplementary data sets provided by partners of the challenge. Mobile phone data consists of communications among 1666 towers distributed across Senegal. We exploit this information to map communication patterns between different areas of the country (i.e., the arrondissements). Another subset consists of 560 millions call records of about 150,000 users along one year at the spatial resolution of arrondissements. We use this information to map individuals' movements among different arrondissements. Demographics information has been obtained from the Senegal data portal [1], an official resource. It is worth noting that information has been manually checked against inconsistencies and data about population for the arrondissements of Bambilor, Thies Sud, Thies Nord, Ndiob and Ngothie were not available. We reconstructed the missing information by combining mobile phone activity and available demographics data (Figure 4.9). Such arrondissements did not exist at the time when the population census was obtained, because they were part of larger administrative areas. Information is available for older arrondissements, therefore we devise a procedure to infer the population in the new areas by using phone calls as a proxy to population density.

We have used the data to also infer more realistic contact rates to be used in viral spreading simulations. The contacts among individuals are generally quite difficult to track at country level. Their rate varies depending on several social and demographical factors such as age, gender, location, urban development, *etc.* [81, 130]. Nevertheless, there are evidences from European and African countries that, on average, the number of daily physical contacts among individuals range from 11 to 22 [81, 130]. There is no available data about contact rate in each arrondissement of Senegal, therefore we need to infer this information from available sources. We first estimate the population density for

1 http://donnees.ansd.sn/en/

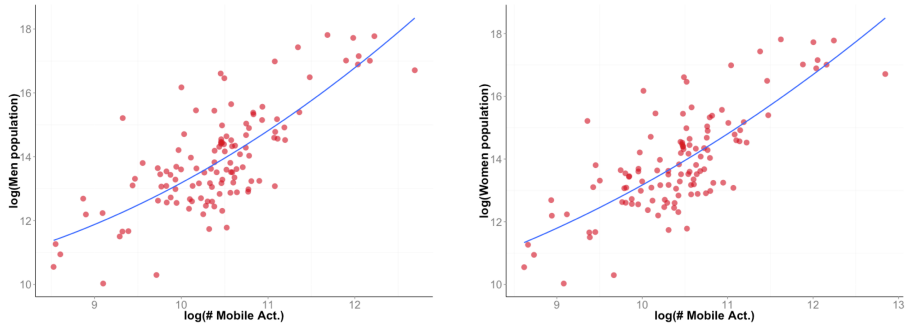


Figure 4.9: Second-order polynomial model (solid line) fitting the log-log relationships between the observed mobile phone data and demographics data (points). Men (left) and women (right) population were fitted separately, thanks to data availability, and have been used to infer the populations in the arrondissements of Bambilor, Thies Sud, Thies Nord, Ndiob and Ngothie.

each region, an administrative level coarser than arrondissement, using available data about number of inhabitants and area. As a plausible range of contact rates, we consider 10 and 25. Under the assumption that the contact rate is proportional to the population density, we assign a value to each region that ranges between 10 and 25, with extremal values assigned to the regions with lowest and highest population density, respectively. Finally, we assign the same contact rate to all arrondissements pertaining to the same region. We obtain a contact rate between 10 and 11 for all regions, except Dakar, which has the highest population density.

4.4.2 *Impact on human mobility flows*

We show in Figure 4.10 the significant differences in modeling the mobility flow using first-order (FO), second-order (SO), gravity, radiation and adaptive memory (AM) models. Gravity, radiation and Markovian models provide very similar transition patterns, whereas adaptive memory provides very different results. The adaptive memory model exhibits significantly less returning transitions than Markovian models, but – on average – with much higher probability of observing them. In fact, 47.4% of patterns captured by the first-order approach and 43.4% captured using second-order are spurious because they are not ob-

served in reality. Remarkably, the probability that an individual comes back to her origin is on average six times higher using adaptive memory models than using first-order, and five times higher using second-order. Radiation and gravity models performs similarly to First-order.

To compare the accuracy of both models against the mobility behavior observed in data, we use the coverage, defined as the fraction of nodes visited by an individual within a given amount of time. We calculate the coverage for each individual in the data, over a period of one month, and then we average over all arrondissements to obtain a measure at country level. For the same period of time, we generate three transition matrices **F**, **S** and **A** encoding the mobility dynamics for first-order, second-order and adaptive memory models, respectively. Using the same data we also generate gravitational and radiational models.

To better replicate the calling behavior of the individuals in the data set, we extract information about the distribution of time between calls and we use this information in our simulations. We found that human communications are bursty, with bunches of intense activities concentrated in a short period that are separated by longer periods of inactivity, Figure 4.11. This pattern is evident when two consecutive calls are made in the same arrondissement. However, in the case of two consecutive calls made from different arrondissements, the probability is lower because of the time required to move from one place to another. The arrondissement where the individual lives influences these activities, e.g., the Dakar region consists of several small arrondissements that are close together and the time required for movements is smaller when compared with the time observed for people living in Louga, a more sparse region. We considered these empirical facts in our simulations, by extracting two call distributions for each arrondissement, $P(T_{intra})$ and $P(T_{inter})$, to model the dynamics within the same arrondissement and among different arrondissements, respectively. An independent simulation is performed using each arrondissement as the starting point of a traveller's journey. At each time step, the traveller moves according to the transition probabilities of a specific mobility model. If the traveler decides to stay in the same arrondissement, the elapsed time within the simulation is increased by a random number drawn from the distribution $P(T_{in})$. Otherwise, if the traveller decides to change

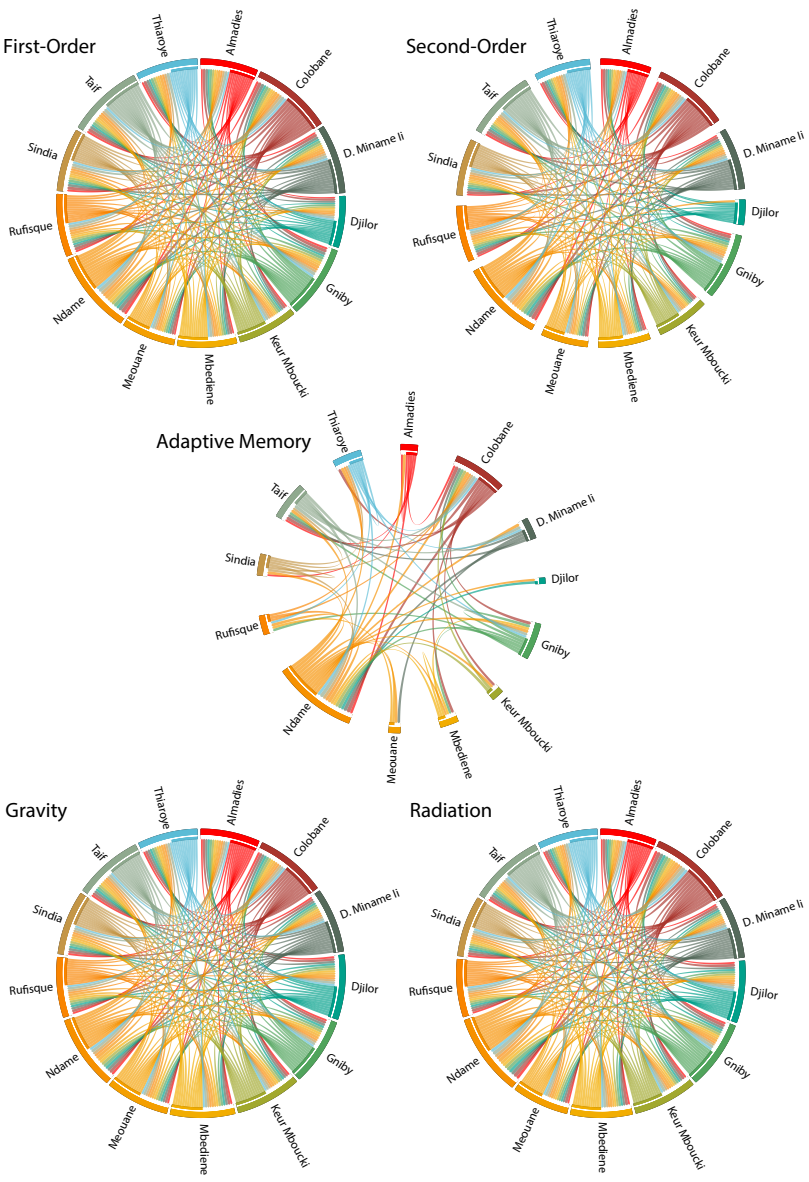


Figure 4.10: Mobility flow among a sub-set of Senegal's arrondissements. For simplicity, we illustrate the effects of each model by considering a subset of 13 arrondissements and patterns that goes through one specific arrondissement (Kael, in this example) after departing from their origin and before reaching their destination.

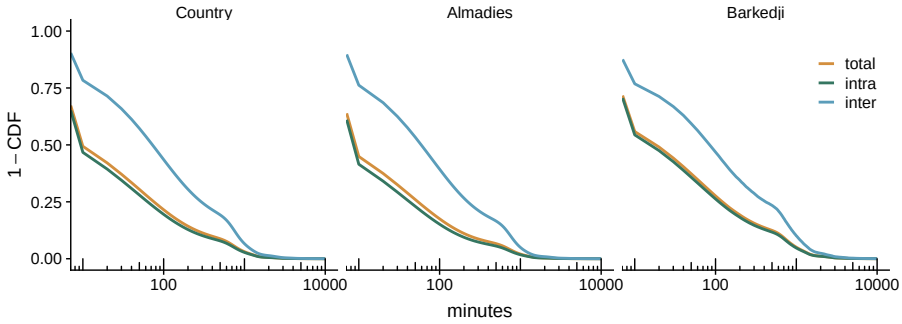


Figure 4.11: Complementary cumulative density function of time between calls at country level and in Almadies and Barkedji arrondissements. For each of them we consider three different actions: time between sequential calls in the same arrondissement (P_{intra}), time between consecutive calls performed in different arrondissements (P_{inter}), and time between any calls (P_{total}). We observe how using information at country level is not a good proxy for modeling the dynamics because the behaviour is significantly different depending on the arrondissement. The time between calls performed in Almadies is lower than in Barkedji due the high density of arrondissements in the region of Almadies, reducing the traveling time between different locations. Note that scale is logarithmic on the $x-$axis.

her location, the elapsed time is increased by a random number drawn from distribution $P(T_{out})$. The traveller stops her journey when the elapsed time reaches 30 days. The coverage for each arrondissement is computed as the average of five thousand independent realizations of this process.

To compare radiation and gravity models with the others we need to make additional assumptions since, by default, the two models do not take into account the waiting time at each site and, in fact, a comparison between their predictions and the ones obtained from Markovian models would be not fair. Therefore, we included waiting times through the calculation of the average waiting time at each location and accounting for it in simulations, discarding individuals' trips for which it is not possible to infer the waiting time.

In Figure 4.12 we show that people diffuse in the country too fast using Markovian models, whereas significantly slower diffusion is

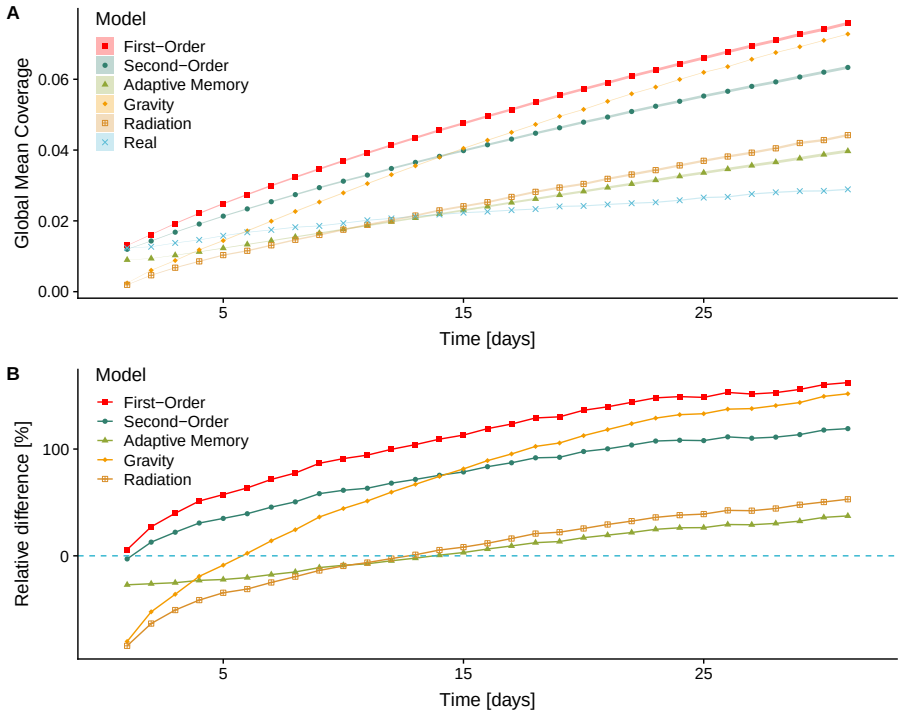


Figure 4.12: (A) Temporal evolution of the global mean coverage calculated from real data and from simulations using first-order, second-order, adaptive memory, gravity and radiation models. (B) Relative difference between the coverage observed in real human mobility and the one obtained from simulations. Here, waiting time equal to zero is assigned to individual's trips where we have incomplete information.

found with adaptive memory, in agreement with empirical observation. We show the result of the same analysis for both the gravity and the radiation models. We observe that the radiation model is not suitable to reproduce the observation, whereas the gravity model provides results comparable with the adaptive memory model proposed in this study.

These results have deep implications, for instance, in short-term or long-term predictions of epidemic spreading or national infrastructure planning.

4.4.3 *Impact on the spreading of epidemics*

How infectious individuals move among different locations has a strong influence in how diseases diffuse in a population. We considered each arrondissement as a meta-population where any individual can interact with a limited number of other individuals. We use a SEIR compartmental model [79] to characterize the epidemics evolution within each arrondissement and mobility models to simulate people traveling in the country.

The discrete time step of the following models is $\Delta t \approx 1$ hour, approximately the observed median between two successive calls from the same individual. The parameters are demographical and epidemiological. Demographics parameters include the birth $B = \tilde{B}\Delta t$ and death $\delta = \tilde{\delta}\Delta t$ probability, whereas epidemiological parameters correspond to the latent period τ_E of the infection, from which the probability $\epsilon = \Delta t/\tau_E$ to pass to the infectious state is calculated, and the infectious period τ_I, from which the probability $\gamma = \Delta t/\tau_I$ to recover from or die because of the infection is calculated. The last parameter is the effective transmission probability

$$\beta_i(t) = 1 - \left(1 - \tilde{\beta}\Delta t \frac{I_i(t)}{N_i(t)}\right)^{c_i \Delta t}, \tag{4.10}$$

an arrondissement-dependent parameter that depends on the average number of contacts per unit of time c_i experienced by an individual in node i, the fraction of infected individuals in that node, and the transmission risk $\tilde{\beta}\Delta t$ in case of contact with an infectious individual. In fact, the definition of $\beta_i(t)$ induces a type-II reaction-diffusion dynamics [30] accounting for the fact that each individual does not interact with *all* the other individuals in the meta-population, but only with a limited sample. If the number of infected agents is small (i.e. $I_i(t) \approx 0$) the Taylor expansion of $\beta_i(t)$ truncated at the first order gives the classical factor $\tilde{\beta}\Delta t c_i \Delta t \frac{I_i(t)}{N_i(t)}$ [79]. It follows that the equations describing the average spreading of a disease according to a SEIR model coupled to first-order mobility are given by

$$S_i(t+1) = \sum_{j=1}^{n} F_{ji} \left[(1 - \delta - \beta_j(t))S_j(t) + BN_j(t)\right]$$

$$
\begin{aligned}
E_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\epsilon-\delta)E_j(t)+\beta_j(t)S_j(t)\right] \\
I_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\gamma-\delta)I_j(t)+\epsilon E_j(t)\right] \\
R_i(t+1) &= \sum_{j=1}^{n} F_{ji}\left[(1-\delta)R_j(t)+\gamma I_j(t)\right]
\end{aligned} \tag{4.11}
$$

whereas the coupling to the second-order model is given by

$$
\begin{aligned}
\tilde{S}_\alpha(t+1) &= \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\delta-\tilde{\beta}_\psi^{(\alpha)}(t))\tilde{S}_\psi(t)+B\tilde{N}_\psi(t)\right] \\
\tilde{E}_\alpha(t+1) &= \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\epsilon-\delta)\tilde{E}_\psi(t)+\tilde{\beta}_\psi^{(\alpha)}(t)\tilde{S}_\psi(t)\right] \\
\tilde{I}_\alpha(t+1) &= \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\gamma-\delta)\tilde{I}_\psi(t)+\epsilon\tilde{E}_\psi(t)\right] \\
\tilde{R}_\alpha(t+1) &= \sum_{\psi=1}^{n^2} H_{\psi\alpha}\left[(1-\delta)\tilde{R}_\psi(t)+\gamma\tilde{I}_\psi(t)\right] \\
\tilde{\beta}_\psi^{(\alpha)}(t) &= 1-\left(1-\tilde{\beta}\Delta t\frac{\sum_{\rho=\left\lfloor\frac{\alpha}{n}\right\rfloor n+1}^{\left\lfloor\frac{\alpha}{n}\right\rfloor n+n}\tilde{I}_\rho(t)}{\sum_{\rho=\left\lfloor\frac{\alpha}{n}\right\rfloor n+1}^{\left\lfloor\frac{\alpha}{n}\right\rfloor n+n}\tilde{N}_\rho(t)}\right)^{c_i\Delta t}
\end{aligned} \tag{4.12}
$$

where $N(t)=\sum_{\psi=1}^{n^\tau}\tilde{N}_\psi(t)$ is the total population in the country at time t, $\lfloor\cdot\rfloor$ indicates the floor function and is used to identify the sub-set of state-nodes corresponding to the same physical node the population $\tilde{S}_\alpha$ belongs to. The equations for the adaptive memory model are the same, except that the transition matrix $\mathbf{A}$ is used instead of $\mathbf{H}$.

We initiate the simulation by infecting five individuals in Barkedji, at the center of Senegal. The differences between the diffusion of the infective process using each mobility model are quite visible in Figure 4.13. The spreading is faster for Markovian models, with some arrondissement populated by more infected individuals than adaptive memory. The

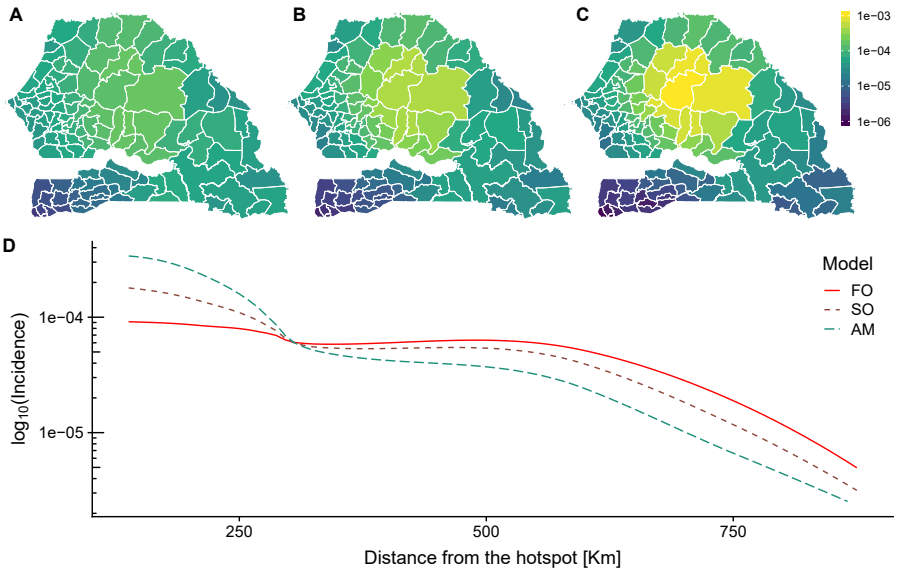


Figure 4.13: We show the incidence of an influenza-like virus over Senegal arrondissements a week after the infection onset, using first-order (A), second-order (B) and adaptive 2-memory (C) mobility models. The infection started in Barkedji (center of Senegal), where three individuals are initially infected. A SEIR compartmental dynamics with parameters $\beta = 0.05$, $\epsilon = 0.2$, $\gamma = 0.5$ is used to simulate the spreading of the disease within each arrondissement. We found that the number of arrondissements with infected individuals is higher using Markovian dynamics. Conversely, the adaptive memory favors a higher concentration of infected individuals in the arrondissements around the initial location of the infection. In fact, the location of the onset of the epidemic can be better identified using adaptive memory rather than Markovian models. (D) Relation between the incidence in a region and the distance from the hotspot of the infection using the three models. Adaptive memory models spread the incidence on regions closer to the hotspot and this effect is even more evident when higher memory is used.

incidence, i.e. the fraction of infected individuals in an arrondissement, follows different spatial patterns in the three models (see Figure 4.13A–C), with a higher incidence observed in the origin of the infection, that decreases as we move far from there. This effect is significantly stronger using adaptive memory because it tends to concentrate more infectious individuals close to the origin (see Figure 4.13D).

4.5 SUMMARY

Modeling how people move among different locations is crucial for several applications. Given the scarcity of information about individuals' movements, often human mobility proxies such as call detail records, GPS, *etc.*, are used instead. Here, we have shown that dynamical models built from human mobility proxies can be significantly wrong, underestimating (or overestimating) real mobility patterns or predicting spurious movements that are not observed in reality. We have proposed a general solution to this issue, by introducing an adaptive memory modeling of human mobility that better captures observed human dynamics and dramatically reduces spurious patterns with respect to memoryless or higher-order Markovian models. However, it is worth remarking that this approach, as all other methods in the literature, is based on the assumption that an individual makes a call in each place he or she visits. In fact, this is not always true and care must be taken when interpreting the results. Fortunately, an appropriate choice of the spatial granularity, for instance at administrative levels corresponding to cities or larger areas, reduces this unavoidable effect. We have validated our model on a data set consisting of 560 millions of call detail records from Senegal. We have found that individuals tend to diffuse faster with standard mobility models than what is observed in reality, whereas the adaptive memory approach reconciles empirical observations and theoretical expectations. Our findings have, for instance, a deep impact on predicting how diseases spread in a country. While standard approaches tend to overestimate the geographical incidence of the infection, the more realistic modeling obtained by means of adaptive memory can improve the inference of the hotspot of the infection, helping to design better countermeasures, e.g., more effective quarantine zones, improved resources deployment or targeted information campaigns.

4.6 PERSONAL CRITICAL VIEW

Despite their capacity to retain conditional waiting times and returning patterns, some important assumptions have to be made before building an adaptive memory model. Most of them are needed due to the

intrinsic nature of the data used to build the model, phone calls in the case of this work. The usage of this kind of data as a proxy to human mobility is always tricky due to the high heterogeneity of user behavior: a different number of calls per day, different inter-call distributions, different daily usage patterns, etc. Those differences could represent an important source of bias for any model; thus a proper data preparation is mandatory before using them.

Another significant limitation of the model refers to the spatial resolution. If small geographical areas compose the physical nodes, the model will need more memory to store returning patterns, since a large number of intermediate locations could be present between origin and destination. Otherwise, if the size of the geographical area is too large, the model will lose many patterns because they will appear inside the same geographical locations. Then, a proper selection of the spatial resolution is required to obtain significant results but, in many circumstances, this selection will depend on the available data.

Finally, notice that when we couple two different dynamics, there is a critical aspect to consider: the relative timescale between the two processes. For the sake of simplicity, here, we have considered that both processes share the same timescale; however, for more realistic results the scales should be accurately determined before any conclusion could be derived from the results.

5

CONCLUSIONS AND FUTURE WORK

This thesis dissertation seeks to answer the fundamental question of how increasing the order of dynamical models, i.e., the amount of information represented, helps to understand the emergence of complex behaviors on processes running on networks. During this document, we have provided evidence of how increasing this order of models concerning structure, relational correlations and temporal information has a deep impact on the understanding of critical questions like the emergence of cooperation, the spreading of epidemics and the understanding of human mobility.

In Chapter 2, we have focused on providing evidence of the impact of higher-order structures, modeled as multilayer networks, on our understanding of the emergence of cooperative behavior on social dilemmas. The introduction of multiplex networks has proved a more realistic representation of social systems. That has allowed us to use them to model more sophisticated individual behaviors that had a profound effect on how they can interact between them. As a consequence, we have documented the appearance of new phenomena not observed before. The dynamical interplay between the different layers of the network has proved essential to understanding the resilience of cooperation and defection on social dilemmas where classical approaches condition their resilience. As a paradigmatic example, we have proved that on Harmony Game, where previous results showed that the only valid strategy was cooperation, defective players can coexist using the interplay dynamics to cooperate in other social contexts and, hence, compensating the price of defection. Thus, higher-order structures pro-

vide an enhancement on the understanding of network dynamics that cannot be neglected.

As we commented in Section 2.7, there are a series of inherent limitations regarding the extrapolation of these results to the understanding of human behavior. We, humans, show a different behavior based on our culture, religion, social background, etc. Assuming that all individuals behave in the same way confronting similar dilemmas is a hard assumption. A possible way to relax this limitation is introducing this heterogeneity on the model, enabling parameters that account for memory about previous actions and their results or, also, correlations with the strategies of the neighbors, thus, increasing the temporal and relational order of the model, at expenses of computational costs. Additionally, the use of evolving networks could provide a plausible framework to model people change their social structure to accommodate their ties to a less defective environment. It is also important to mention, that over the recent years many efforts have been devoted to obtaining empirical data of social dilemmas. Designing social experiments to test our theoretical observations will help to refine the underlying model.

In Chapter 3, we concentrated on increasing the order of microscopic SIS epidemic models to capture effects that go beyond the description of the state of the nodes. We proposed a new approach to model epidemics at a level of links, called Epidemic Link Equations (ELE), which we proved that it provides critical descriptors to understand the role that each link plays on epidemic outbreaks. One of them is link epidemic importance, which we introduced as a measure of the impact of the transmission through a given link in its surrounding neighborhood. We used this metric to develop an epidemic containment strategy consisting in deactivating recursively the links with largest link epidemic importance, leading to effective results when compared with other network-based approaches. These results can be attributed to the dynamical information encoded inside this metric, which drives the containment procedure, unraveling the importance of accounting for relational higher-order dependencies when we model the spreading of epidemics on a complex network.

Some critical considerations have to be made about this approach, as we explained in Section 3.5. We tested our method with regular graphs, like lattices and cliques and, despite improving the results obtained with lesser-order methods, we found a significant discrepancy between numerical simulations and results provided by our model. We believe that these discrepancies can be attributed to the importance of dynamical correlations on regular structures that go beyond our assumption of second-order probabilistic independence. A higher-order method that probabilistically accounts for correlations for different cycle lengths could help to tackle the problem, at expenses of increasing the complexity of the functional form of the model.

Another point to consider is the way how we defined Link Epidemic Importance as a measure of the impact of the transmission of the disease. To guarantee that the edges that show higher link epidemic importance are the ones accounting for the major impact on the spreading of the epidemics, we should consider the full sequence of infections triggered by an initial spread through each link; however, our method just accounts for the impact on first neighbors. Our approach has some useful properties that could help on this effort; however, they need to be studied in further detail

It is worth mentioning that we found some performance similarities between the results by our method and EigenScore regarding the containment procedure, despite the EigenScore lack of information about dynamics. Further research is needed to understand the spectral properties that relate these two methods. We also have to comment, that the temporal scale used on the analysis of the containment process enforces the arrival to the stationary state. Nevertheless, the actuation on the epidemics has to be done as fast as possible to stop the spreading before it becomes endemic. Thus, a revision of the containment strategy should consider the effects of the transitory state.

Finally, in Chapter 4 we focused our attention on one of the most challenging topics of behavioral modeling, human mobility. We assessed that the heterogeneity about the time spent at a given place before moving somewhere else is critical to understand human mobility. We proposed a Markovian model, called Adaptive memory model, that captures these times while preserving information about recurrent

patterns, another essential attribute of human mobility. We tested our method against four well-grounded mobility models, fitting them with mobile phone records as a proxy for human mobility. We found that our model dramatically reduces the number of temporal and spatial spurious patterns compared with the other testing methods. We coupled these models with an epidemic dynamics, where temporal information is intrinsically related to the spreading process. Our findings showed that, since the recurrent patterns and temporal information are maintained on the Adaptive memory model, the diffusion imposed by our method is less diffusive than the one observed for the other methods, leading to more realistic spreading patterns.

As we commented on Section 4.6, our major concerns are related to the selection of proper space and time scales. As we said, space scales are critical when we study mobility models. If we impose small location areas, our model will need more memory to store returning patterns, since a large number of intermediate locations could be present between origin and destination. Otherwise, if the size of the geographical area is too large, the model will lose many patterns because they will appear inside the same geographical locations. The choice of a proper resolution scale depends on data and, to properly test our model under different scenarios, we would need to access a broad spectrum of datasets.

We observe the same problem regarding the selection of a proper temporal scale. Since we used mobile phone records as a proxy to human mobility, our model absorbes the temporal heterogeneity present in mobile communications, that could be a possible source of bias. To better asses our method capabilities, we should build it using a proxy that maintains homogeneous inter-event times between each sample, like the ones provided by GPS traces. However, open access to this kind of data is limited to datasets that encode the movements of few individuals.

Note that when we explained our model, we combined the information of all the individuals into a single mobility matrix. However, since we have individual information for every single user, we could build models that better retain the heterogeneity present in the dataset if we aggregate individuals by similar behaviors, akin to the degree classes

in heterogeneous mean field methods for epidemic spreading. Building a mobility matrix for each group can be seen as an increase in the structural order of the model, and it could lead to the emergence of refined mobility patterns.

In the near future, I plan to apply all these techniques to a particular field, medicine. Over the last years, thanks to the digitalization of their records, medical institutions have built massive datasets that contain information about medical treatments, patients demographics, results of medical tests, etc. The analysis of these data with modern data science techniques has already shown their potential, providing powerful diagnostic tools. However, as many computational techniques, the use of machine learning approaches cames at expenses of the interpretability of the models. The need to uncover understandable patterns is an opportunity for experts in mathematical modeling and analysis. I think that some of the tools that we have exposed on this document and all the know-how, background and experience learned during my Ph.D. studies should provide me a perfect set of tools to tackle any future problems in this field.

UNIVERSITAT ROVIRA I VIRGILI
HIGHER-ORDER DYNAMICS ON COMPLEX NETWORKS
Joan Tomàs Matamalas Llodrà

A

APPENDIX

A.1 DATA USED TO TEST LINK EPIDEMIC IMPORTANCE APPROACH

Description of the 27 real networks used in Figure 3.14 of the main text, and in Figure 3.15, sorted by increasing number of nodes. They have been obtained from the Network Repository (http://networkrepository.com). Table A.1 provides their main structural characteristics.

IA-INFECT-DUBLIN Human contact network where nodes represent humans and edges represent proximity (i.e., contacts in the physical world), during the Infectious SocioPatterns event that took place at the Science Gallery in Dublin, Ireland.

SOC-WIKI-VOTE Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes represent Wikipedia users and a directed edge from node i to node j represents that user i voted on user j.

CA-CSPHD Genealogy network of PhD's in computer science.

IA-FB-MESSAGES The Facebook-like Social Network originate from an online community for students at University of California, Irvine. The dataset includes the users that sent or received at least one message.

SOC-HAMSTERSTER Network of the friendship and family links between users of Hamsterster social network.

SOCFB-USFCA72 A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties.

SOCFB-NIPS-EGO A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties.

SOCFB-SANTA74 A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties.

CA-GRQC Collaboration network of arXiv General Relativity. Nodes represent scientists, and links coauthorship.

WEB-SPAM Web Spam Challenge 2008 network.

POWER-US-GRID US Power grid graph.

CA-ERDOS992 Erdős collaboration network. Nodes represent scientists, and links coauthorship.

SOC-ADVOGATO Advogato is a social community platform where users can explicitly express weighted trust relationships among themselves. The dataset contains a list of all of the user-to-user links.

P2P-GNUTELLA08 Gnutella peer to peer network from August 8 2002.

IA-REALITY Reality mining network data consists of human mobile phone call events between a small set of core users at the MIT whom actually were assigned mobile phones for which all calls were collected. A node represents a person; an edge indicates a phone call or voicemail between two users.

CA-HEPTH Collaboration network of arXiv High Energy Physics Theory. Nodes represent scientists, and links coauthorship.

SOC-ANYBEAT Anybeat is an online community, a public gathering place where you can interact with people from around your neighborhood or across the world.

CA-ASTROPH Collaboration network of arXiv Astrophysics. Nodes represent scientists, and links coauthorship.

CA-CONDMAT Collaboration network of arXiv Condensed Matter. Nodes represent scientists, and links coauthorship.

SOC-GPLUS Google+ social network.

TECH-AS-CAIDA2007 Internet network at the level of autonomous systems as of 2017. Nodes represent autonomous systems, and there exists a link between them if they have a business agreement for the routing of packets.

IA-EMAIL-EU The network was generated using email data from a large European research institution. Nodes are users and edges represent email exchanges between two users in both directions.

IA-ENRON-LARGE Network of emails exchanged between senor managers of Enron Corporation, during the period which lead to its bankruptcy.

SOC-BRIGHTKITE Brightkite is a location-based social networking service provider where users shared their locations by checking-in. The dataset contains all links among users.

SOC-EPINIONS Who-trust-whom online social network of the general consumer review site Epinions.com.

SOC-SLASHDOT A technology-related news website known for its specific user community. The dataset contains friend/foe tags between the users of slashdot.

SOC-TWITTER-FOLLOWS Twitter follower network.

The prefix in the name of each network indicates the cathegory it belongs to, namely: (ia) interaction networks; (soc) social networks; (ca) collaboration networks; (socfb) Facebook networks; (web) we graphs; (power) power networks; (p2p) peer to peer networks; (tech) technological networks.

Name	N	L	$\langle k \rangle$	c	r
ia-infect-dublin	410	2765	13.4878	0.1452	0.0000
soc-wiki-Vote	889	2914	6.5557	0.0901	-0.0556
ca-CSphd	1025	1043	2.0351	0.0023	-0.2532
ia-fb-messages	1266	6451	10.1912	0.6289	0.6392
soc-hamsterster	2000	16097	16.0970	0.0268	-0.3816
socfb-USFCA72	2672	65244	48.8353	0.0037	-0.3886
socfb-nips-ego	2888	2981	2.0644	0.3178	0.2013
socfb-Santa74	3578	151747	84.8222	0.2618	0.1253
ca-GrQc	4158	13422	6.4560	0.2294	0.0227
web-spam	4767	37375	15.6807	0.0006	-0.8764
power-US-Grid	4941	6594	2.6691	0.0925	-0.0952
ca-Erdos992	4991	7428	2.9766	0.0439	-0.0844
soc-advogato	5054	39374	15.5813	0.0207	0.0355
p2p-Gnutella08	6299	20776	6.5966	0.0420	-0.4531
ia-reality	6809	7680	2.2558	0.1032	0.0035
ca-HepTh	8638	24806	5.7435	0.1907	0.0917
soc-anybeat	12645	49132	7.7710	0.0024	-0.6753
ca-AstroPh	17903	196972	22.0044	0.4357	0.2258
ca-CondMat	21363	91286	8.5462	0.1273	-0.0288
soc-gplus	23576	39145	3.3207	0.0073	-0.1946
tech-as-caida2007	26475	53381	4.0326	0.0260	-0.0651
ia-email-EU	32430	54397	3.3547	0.0004	-0.6682
ia-enron-large	33696	180811	10.7319	0.2022	0.0706
soc-brightkite	56739	212945	7.5061	0.2811	0.2389
soc-epinions	61355	494372	16.1151	0.0217	-0.1234
soc-slashdot	70068	358647	10.2371	0.0851	-0.1165
soc-twitter-follows	404719	713319	3.5250	0.1105	0.0096

Table A.1: Structural characteristics of the 27 real networks obtained from the Network Repository (http://networkrepository.com) and used in Figure 3.14 of the main text, and in Figure 3.15. They correspond to the largest connected component of the networks with the same name in the repository. The structural descriptors shown are the number of nodes N, the number of links L, the average degree $\langle k \rangle$, the clustering coefficient c, and the assortativity r.

BIBLIOGRAPHY

[1] B. Alharbi and X. Zhang. "Exploring the significance of human mobility patterns in social link prediction." In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing - SAC '14* (2014).

[2] A. Amini, K. Kung, C. Kang, S. Sobolevsky, and C. Ratti. "The impact of social segregation on human mobility in developing and industrialized regions." In: *EPJ Data Science* 3.1 (2014), p. 6.

[3] R. M. Anderson, R. M. May, and B. Anderson. *Infectious diseases of humans: dynamics and control*. Vol. 28. Wiley Online Library, 1992.

[4] H. Andersson and T. Britton. "Models for endemic diseases." In: *Lecture Notes in Statistics* (2000), pp. 73–83.

[5] A. Arenas, A. Fernández, and S. Gómez. "A complex network approach to the determination of functional groups in the neural system of C. elegans." In: *Bio-inspired computing and communication*. Springer, 2008, pp. 9–18.

[6] D. F. Austin. "The Miraculus Fever-Tree. Malaria and the Quest for a Cure that Changed the World." In: *Economic Botany* 58.1 (Jan. 2004), pp. 132–132.

[7] R. Axelrod. "Effective Choice in the Prisoner's Dilemma." In: *J. Confl. Resolut.* 24.1 (Mar. 1980), pp. 3–25.

[8] R. Axelrod. "The Emergence of Cooperation among Egoists." In: *Am. Polit. Sci. Rev.* 75.2 (June 1981), p. 306.

[9] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani. "Multiscale mobility networks and the spatial spreading of infectious diseases." In: *PNAS* 106.51 (2009), pp. 21484–21489.

[10] D. Balcan and A. Vespignani. "Phase transitions in contagion processes mediated by recurrent mobility patterns." In: *Nature Phys.* 7.7 (2011), pp. 581–586.

[11] J. R. Banavar, A. Maritan, and A. Rinaldo. "Size and form in efficient transportation networks." In: *Nature* 399.6732 (May 1999), pp. 130–132.

[12] A.-L. Barabási and R. Albert. "Emergence of Scaling in Random Networks." In: *Science* 286.5439 (Oct. 1999), pp. 509–512.

[13] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini. "Human mobility: Models and applications." In: *Physics Reports* 734 (Mar. 2018), pp. 1–74.

[14] M. Barthélemy. "Spatial networks." In: *Physics Reports* 499.1 (2011), pp. 1–101.

[15] F. Battiston, V. Nicosia, and V. Latora. "Structural measures for multiplex networks." In: *Physical Review E* 89.3 (Mar. 2014).

[16] M. Batty. "The Size, Scale, and Shape of Cities." In: *Science* 319.5864 (Feb. 2008), pp. 769–771.

[17] V. Belik, T. Geisel, and D. Brockmann. "Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases." In: *Physical Review X* 1.1 (Aug. 2011).

[18] J. H. Bergstrand. "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade." In: *The Review of Economics and Statistics* 71.1 (Feb. 1989), p. 143.

[19] K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, S. McKay, et al. "Monte Carlo simulation in statistical physics." In: *Computers in Physics* 7.2 (1993), pp. 156–157.

[20] A. N. Bishop and I. Shames. "Link operations for slowing the spread of disease in complex networks." In: *Europhys. Lett.* 95.1 (July 2011), p. 18005.

[21] V. D. Blondel, A. Decuyper, and G. Krings. "A survey of results on mobile phone datasets analysis." In: *To appear in EPJ Data Science* (2015).

[22] D. Brockmann and D. Helbing. "The Hidden Geometry of Complex, Network-Driven Contagion Phenomena." In: *Science* 342.6164 (2013), pp. 1337–1342.

[23] C.-R. Cai, Z.-X. Wu, M. Z. Chen, P. Holme, and J.-Y. Guan. "Solving the dynamic correlation problem of the susceptible-infected-susceptible model on networks." In: *Physical Review Letters* 116.25 (2016), p. 258301.

[24] C.-R. Cai, Z.-X. Wu, and J.-Y. Guan. "Effective degree Markov-chain approach for discrete-time epidemic processes on uncorrelated networks." In: *Physical Review E* 90.5 (2014), p. 052803.

[25] F. Calabrese, L. Ferrari, and V. D. Blondel. "Urban Sensing Using Mobile Phone Network Data: A Survey of Research." In: *ACM Computing Surveys (CSUR)* 47.2 (2014), p. 25.

[26] E. Cator and P. Van Mieghem. "Second-order mean-field susceptible-infected-susceptible epidemic threshold." In: *Physical Review E* 85.5 (2012), p. 056111.

[27] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. "Epidemic thresholds in real networks." In: *ACM Transactions on Information and System Security (TISSEC)* 10.4 (2008), p. 1.

[28] F. Chung, P. Horn, and A. Tsiatas. "Distributing antidote using pagerank vectors." In: *Internet Mathematics* 6.2 (2009), pp. 237–254.

[29] R. Cohen, S. Havlin, and D. Ben-Avraham. "Efficient immunization strategies for computer networks and populations." In: *Physical Review Letters* 91.24 (2003), p. 247901.

[30] V. Colizza, R. Pastor-Satorras, and A. Vespignani. "Reaction–diffusion processes and metapopulation models in heterogeneous networks." In: *Nature Phys.* 3.4 (2007), pp. 276–282.

[31] V. Colizza and A. Vespignani. "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations." In: *J. Theor. Bio.* 251.3 (2008), pp. 450–467.

[32] C. D. Conner. *A People's History of Science: Miners, Midwives, and Low Mechanicks*. Hachette UK, 2009.

[33] M. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonçalves, F. Menczer, and A. Flammini. "Political polarization on twitter." In: *Icwsm* 133 (2011), pp. 89–96.

[34] E. Cozzo, M. Kivelä, M. De Domenico, A. Solé-Ribalta, A. Arenas, S. Gómez, M. A. Porter, and Y. Moreno. "Structure of triadic relations in multiplex networks." In: *New Journal of Physics* 17.7 (2015), p. 073029.

[35] D. J. Daley, J. Gani, and J. M. Gani. *Epidemic modelling: an introduction*. Vol. 15. Cambridge University Press, 2001.

[36] M. De Domenico, M. A. Porter, and A. Arenas. "MuxViz: a tool for multilayer analysis and visualization of networks." In: *Journal of Complex Networks* 3.2 (Oct. 2014), pp. 159–176.

[37] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall. "Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems." In: *Physical Review X* 5.1 (Mar. 2015).

[38] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. Porter, S. Gomez, and A. Arenas. "Mathematical Formulation of Multilayer Networks." In: *Phys. Rev. X* 3.4 (Dec. 2013), p. 041022.

[39] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas. "Navigability of interconnected networks under random failures." In: *PNAS* (2014), pp. 8351–8356.

[40] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas. "Ranking in interconnected multilayer networks reveals versatile nodes." In: *Nature Communications* 6.1 (Apr. 2015).

[41] P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem. "Dynamic population mapping using mobile phone data." In: *PNAS* 111.45 (2014), pp. 15888–15893.

[42] O. Diekmann and J. Heesterbeek. *Mathematical epidemiology of infectious diseases*. Vol. 146. Wiley, Chichester, 2000.

[43] A. Dobra, N. E. Williams, and N. Eagle. "Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data." In: *PLoS ONE* (2014), DOI:10.1371/journal.pone.0120449.

[44] N. Eagle, Y. de Montjoye, and L. M. Bettencourt. "Community computing: Comparisons between rural and urban societies using mobile phone data." In: *Computational Science and Engineering, 2009. CSE'09. International Conference on*. Vol. 4. IEEE. 2009, pp. 144–150.

[45] N. Eagle and A. Pentland. "Reality mining: sensing complex social systems." In: *Personal and ubiquitous computing* 10.4 (2006), pp. 255–268.

[46] N. Eagle, A. S. Pentland, and D. Lazer. "Inferring friendship network structure by using mobile phone data." In: *PNAS* 106.36 (2009), pp. 15274–15278.

[47] D. J. D. Earn, P. Rohani, B. M. Bolker, and B. T. Grenfell. "A Simple Model for Complex Dynamical Transitions in Epidemics." In: *Science* 287.5453 (2000), pp. 667–670.

[48] P. Erdos and A. Rényi. "On the evolution of random graphs." In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[49] N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. "Strategies for containing an emerging influenza pandemic in Southeast Asia." In: *Nature* 437.7056 (2005), pp. 209–214.

[50] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras. "Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results." In: *Physical Review E* 86.4 (2012), p. 041125.

[51] P. N. Fonkwo. "Pricing infectious disease. The economic and health implications of infectious diseases." In: *EMBO reports* 9 (July 2008), S13–S17.

[52] L. Gauvin, A. Panisson, and C. Cattuto. "Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach." In: *PLoS ONE* 9.1 (Jan. 2014). Ed. by Y. Moreno, e86028.

[53] J. P. Gleeson. "Binary-state dynamics on complex networks: Pair approximation and beyond." In: *Physical Review X* 3.2 (2013), p. 021004.

[54] J. P. Gleeson. "Bond percolation on a class of clustered random networks." In: *Physical Review E* 80.3 (Sept. 2009).

[55] J. Gómez-Gardeñes, P. Echenique, and Y. Moreno. "Immunization of real complex communication networks." In: *The European Physical Journal B - Condensed Matter and Complex Systems* 49.2 (Jan. 2006), pp. 259–264.

[56] J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas. "Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks." In: *Nature Physics* 14.4 (Dec. 2017), pp. 391–395.

[57] J. Gomez-Gardenes, M. Campillo, L. Floría, and Y. Moreno. "Dynamical Organization of Cooperation in Complex Topologies." In: *Phys. Rev. Lett.* 98.10 (Mar. 2007), p. 108103.

[58] J. Gómez-Gardeñes and Y. Moreno. "From scale-free to Erdos-Rényi networks." In: *Physical Review E* 73.5 (2006), p. 056124.

[59] J. Gómez-Gardeñes, I. Reinares, A. Arenas, and L. M. Floría. "Evolution of Cooperation in Multiplex Networks." In: *Sci. Rep.* 2 (Aug. 2012).

[60] S. Gómez, A. Díaz-Guilera, J. Gomez-Gardenes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas. "Diffusion Dynamics on Multiplex Networks." In: *Phys. Rev. Lett.* 110.2 (Jan. 2013), p. 028701.

[61] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno. "Discrete-time Markov chain approach to contact-based disease spreading in complex networks." In: *EPL (Europhysics Letters)* 89.3 (2010), p. 38009.

[62] S. Gómez, J. Gómez-Gardenes, Y. Moreno, and A. Arenas. "Nonperturbative heterogeneous mean-field approach to epidemic spreading in complex networks." In: *Physical Review E* 84.3 (2011), p. 036105.

[63] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. "Understanding individual human mobility patterns." In: *Nature* 453.7196 (2008), pp. 779–782.

[64] B. C. Hacker and G. Raudzens. "Technology, Disease, and Colonial Conquests, Sixteenth to Eighteenth Centuries: Essays Reappraising the Guns and Germs Theories." In: *The Journal of Military History* 66.4 (Oct. 2002), p. 1194.

[65] A. Hackett, D. Cellai, S. Gómez, A. Arenas, and J. P. Gleeson. "Bond Percolation on Multiplex Networks." In: *Physical Review X* 6.2 (Apr. 2016).

[66] A. Halu, R. J. Mondragón, P. Panzarasa, and G. Bianconi. "Multiplex PageRank." In: *PLoS ONE* 8.10 (Oct. 2013). Ed. by Y. Moreno, e78293.

[67] C. Hauert and M. Doebeli. "Spatial structure often inhibits the evolution of cooperation in the snowdrift game." In: *Nature* 428.6983 (Apr. 2004), pp. 643–646.

[68] J. N. Hays. *Epidemics and pandemics: their impacts on human history*. Abc-clio, 2005.

[69] L. Hébert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé. "Global efficiency of local immunization on complex networks." In: *Scientific Reports* 3 (2013), p. 2171.

[70] H. W. Hethcote. "The mathematics of infectious diseases." In: *SIAM Review* 42.4 (2000), pp. 599–653.

[71] H. W. Hethcote. "Three Basic Epidemiological Models." In: *Biomathematics* (1989), pp. 119–144.

[72] P. Holme and B. J. Kim. "Growing scale-free networks with tunable clustering." In: *Physical Review E* 65.2 (2002), p. 026107.

[73] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. "Attack vulnerability of complex networks." In: *Phys. Rev. E* 65.5 (May 2002), p. 056109.

[74] L. Hufnagel, D. Brockmann, and T. Geisel. "Forecast and control of epidemics in a globalized world." In: *Proceedings of the National Academy of Sciences USA* 101.42 (2004), pp. 15124–15129.

[75] T. C. Ings, J. M. Montoya, J. Bascompte, N. Blüthgen, L. Brown, C. F. Dormann, F. Edwards, D. Figueroa, U. Jacob, J. I. Jones, et al. "Ecological networks–beyond food webs." In: *Journal of Animal Ecology* 78.1 (2009), pp. 253–269.

[76] W.-S. Jung, F. Wang, and H. E. Stanley. "Gravity model in the Korean highway." In: *EPL (Europhysics Letters)* 81.4 (Jan. 2008), p. 48005.

[77] M. Kaiser and C. C. Hilgetag. "Nonoptimal Component Placement, but Short Processing Paths, due to Long-Distance Projections in Neural Systems." In: *PLoS Computational Biology* 2.7 (2006), e95.

[78] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. "The complex network of global cargo ship movements." In: *Journal of The Royal Society Interface* 7.48 (July 2010), pp. 1093–1103.

[79] M. J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

[80] I. Z. Kiss, L. Berthouze, T. J. Taylor, and P. L. Simon. "Modelling approaches for simple dynamic networks and applications to disease transmission models." In: *Proc. R. Soc. A* 468.2141 (2012), pp. 1332–1355.

[81] M. C. Kiti, T. M. Kinyanjui, D. C. Koech, P. K. Munywoki, G. F. Medley, and D. J. Nokes. "Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya." In: *PloS one* 9.8 (2014), e104786.

[82] M. Kivelä, A. Arenas, M. Barthelemy, G. J. P, Y. Moreno, and Porter. "Multilayer Networks." In: *J. Complex Netw.* 2.3 (2014), pp. 203–271.

[83] A. Lancichinetti, S. Fortunato, and F. Radicchi. "Benchmark graphs for testing community detection algorithms." In: *Physical Review E* 78.4 (Oct. 2008).

[84] J. Leskovec, J. Kleinberg, and C. Faloutsos. "Graph evolution: Densification and shrinking diameters." In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007), p. 2.

[85] A. N. Licht. "Games commissions play: 2x2 games of international securities regulation." In: *Yale J. Int. Law* (1999).

[86] A. Lima, M. De Domenico, V. Pejovic, and M. Musolesi. "Disease Containment Strategies based on Mobility and Information Dissemination." In: *Scientific Reports* 5.1 (June 2015).

[87] A. Lima, R. Stanojevic, D. Papagiannaki, P. Rodriguez, and M. C. González. "Understanding individual routing behaviour." In: *Journal of The Royal Society Interface* 13.116 (2016).

[88] J. Livingston and S. Watts. "Disease and Medicine in World History." In: *The International Journal of African Historical Studies* 36.3 (2003), p. 659.

[89] S. Lozano, A. Arenas, and A. Sánchez. "Mesoscopic Structure Conditions the Emergence of Cooperation on Social Networks." In: *PLoS ONE* 3.4 (Apr. 2008). Ed. by A. P. Galvani, e1892.

[90] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson. "Approaching the Limit of Predictability in Human Mobility." In: *Scientific Reports* 3.1 (Oct. 2013).

[91] R. D. Luce and H. Raiffa. *Games and Decisions*. Introduction and Critical Survey. New York, NY: Courier Dover Publications, Sept. 2012.

[92] N. Madar, T. Kalisky, R. Cohen, D. ben-Avraham, and S. Havlin. "Immunization and epidemic dynamics in complex networks." In: *The European Physical Journal B* 38.2 (Mar. 2004), pp. 269–276.

[93] A. A. Markov and N. M. Nagorny. "The Theory of Algorithms." In: (1988).

[94] A. S. Mata and S. C. Ferreira. "Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks." In: *EPL (Europhysics Letters)* 103.4 (2013), p. 48003.

[95] J. T. Matamalas, A. Arenas, and S. Gómez. "Effective approach to epidemic containment using link equations in complex networks." In: *Science Advances* 4.12 (Dec. 2018), eaau4212.

[96] J. T. Matamalas, M. De Domenico, and A. Arenas. "Assessing reliable human mobility patterns from higher order memory in mobile communications." In: *Journal of The Royal Society Interface* 13.121 (Aug. 2016), p. 20160203.

[97] J. T. Matamalas, J. Poncela-Casasnovas, S. Gómez, and A. Arenas. "Strategical incoherence regulates cooperation in social dilemmas on multiplex networks." In: *Scientific Reports* 5.1 (Apr. 2015).

[98] V. B. Meyer-Rochow. "Food Taboos." In: *Encyclopedia of Food Security and Sustainability* (2019), pp. 332–335.

[99] S. Milgram. "The small-world problem." In: *PsycEXTRA Dataset* (1967).

[100] Y.-A. de Montjoye, Z. Smoreda, R. Trinquart, C. Ziemlicki, and V. D. Blondel. "D4D-Senegal: The Second Mobile Phone Data for Development Challenge." In: *arXiv:1407.4885* (2014).

[101] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. "Community Structure in Time-Dependent, Multiscale, and Multiplex Networks." In: *Science* 328.5980 (May 2010), pp. 876–878.

[102] J. Nash. "Non-Cooperative Games." In: *The Annals of Mathematics* 54.2 (Sept. 1951), p. 286.

[103] M. E. Newman. "Spread of epidemic disease on networks." In: *Physical Review E* 66.1 (2002), p. 016128.

[104] M. E. Newman and M. Girvan. "Finding and evaluating community structure in networks." In: *Physical Review E* 69.2 (2004), p. 026113.

[105] M. A. Nowak and R. M. May. "Evolutionary games and spatial chaos." In: *Nature* (1992).

[106] M. A. Nowak. *Evolutionary Dynamics*. Exploring the equations of life. Cambridge, MA: Harvard University Press, Sept. 2006.

[107] M. A. Nowak. "Five Rules for the Evolution of Cooperation." In: *Science* 314.5805 (Dec. 2006), pp. 1560–1563.

[108] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. "Emergence of cooperation and evolutionary stability in finite populations." In: *Nature* 428.6983 (Apr. 2004), pp. 646–650.

[109] C. Nowzari, V. M. Preciado, and G. J. Pappas. "Analysis and control of epidemics: A survey of spreading processes on complex networks." In: *IEEE Control Systems* 36.1 (2016), pp. 26–46.

[110] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. "Uncovering the overlapping community structure of complex networks in nature and society." In: *Nature* 435.7043 (June 2005), pp. 814–818.

[111] A. Pandey, K. E. Atkins, J. Medlock, N. Wenzel, J. P. Townsend, J. E. Childs, T. G. Nyenswah, M. L. Ndeffo-Mbah, and A. P. Galvani. "Strategies for containing Ebola in west Africa." In: *Science* 346.6212 (2014), pp. 991–995.

[112] P. E. Paré, B. E. Kirwan, J. Liu, T. Başar, and C. L. Beck. "Discrete-time spread processes: Analysis, identification, and validation." In: *2018 Annual American Control Conference (ACC)*. IEEE. 2018, pp. 404–409.

[113] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. "Epidemic processes in complex networks." In: *Reviews of Modern Physics* 87.3 (2015), p. 925.

[114] R. Pastor-Satorras and A. Vespignani. "Epidemic spreading in scale-free networks." In: *Physical Review Letters* 86.14 (2001), p. 3200.

[115] R. Pastor-Satorras and A. Vespignani. "Immunization of complex networks." In: *Phys. Rev. E* 65 (3 Feb. 2002), p. 036104.

[116] R. Pastor-Satorras and A. Vespignani. "Measuring the global Internet." In: *Evolution and Structure of the Internet* (), pp. 19–35.

[117] T. Paulson. "Epidemiology: A mortal foe." In: *Nature* 502.7470 (Oct. 2013), S2–S3.

[118] T. P. Peixoto. "Inferring the mesoscale structure of layered, edge-valued, and time-varying networks." In: *Physical Review E* 92.4 (Oct. 2015).

[119] J. Poncela, J. Gómez-Gardeñes, L. M. Floría, and Y. Moreno. "Robustness of cooperation in the evolutionary prisoner's dilemma on complex networks." In: *New Journal of Physics* 9.6 (June 2007), pp. 184–184.

[120] J. Poncela, J. Gómez-Gardeñes, L. M. Floría, and Y. Moreno. "Growing Networks Driven by the Evolutionary Prisoner's Dilemma Game." In: *Handbook of Optimization in Complex Networks*. Gainesville, FL: Springer, 2012, pp. 115–136.

[121] J. Poncela, J. Gómez-Gardeñes, L. M. Floría, A. Sánchez, and Y. Moreno. "Complex Cooperative Networks from Evolutionary Preferential Attachment." In: *PLOS ONE* 3.6 (June 2008), e2449.

[122] B. A. Prakash, L. Adamic, T. Iwashyna, H. Tong, and C. Faloutsos. "Fractional immunization in networks." In: *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM. 2013, pp. 659–667.

[123] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas. "Optimal resource allocation for network protection against spreading processes." In: *IEEE Transactions on Control of Network Systems* 1.1 (2014), pp. 99–108.

[124] F. Radicchi. "Driving Interconnected Networks to Supercriticality." In: *Physical Review X* 4.2 (Apr. 2014).

[125] A. Rapoport. "Prisoner's Dilemma — Recollections and Observations." In: *Game Theory as a Theory of a Conflict Resolution*. Dordrecht, NL: Springer Netherlands, Jan. 1974, pp. 17–34.

[126] J. G. Restrepo, E. Ott, and B. R. Hunt. "Characterizing the dynamical importance of network nodes and links." In: *Physical Review Letters* 97.9 (2006), p. 094102.

[127] C. M. Rivers and S. V. Scarpino. "Modelling the trajectory of disease outbreaks works." In: *Nature* 559.7715 (July 2018), pp. 477–477·

[128] C. P. Roca, J. A. Cuesta, and A. Sánchez. "Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics." In: *Phys. Life. Rev.* 6.4 (Dec. 2009), pp. 208–249.

[129] F. M. Rodríguez. "Precisions on the History of Quinine." In: *Reumatología Clínica (English Edition)* 3.4 (Jan. 2007), pp. 194–196.

[130] P. Rohani, X. Zhong, and A. A. King. "Contact network structure explains the changing epidemiology of pertussis." In: *Science* 330.6006 (2010), pp. 982–985.

[131] R. A. Rossi and N. K. Ahmed. "The Network Data Repository with Interactive Graph Analytics and Visualization." In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. Vol. 15. 2015, pp. 4292–4293.

[132] M. Rosvall, D. Axelsson, and C. T. Bergstrom. "The map equation." In: *The European Physical Journal Special Topics* 178.1 (Nov. 2009), pp. 13–23.

[133] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte. "Memory in network flows and its effects on spreading dynamics and community detection." In: *Nature Communications* 5 (2014), p. 4630.

[134] J.-J. Rousseau. *A Discourse on Inequality*. London, UK: Penguin Books, Sept. 2003.

[135] S. Saha, A. Adiga, B. A. Prakash, and A. K. S. Vullikanti. "Approximation Algorithms for Reducing the Spectral Radius to Control Epidemic Spread." In: *Proceedings of the 2015 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics, July 2015, pp. 568–576.

[136] F. C. Santos, J. M. Pacheco, and T. Lenaerts. "Evolutionary dynamics of social dilemmas in structured heterogeneous populations." In: *Proceedings of the National Academy of Sciences* 103.9 (Feb. 2006), pp. 3490–3494.

[137] F. Santos and J. Pacheco. "Scale-Free Networks Provide a Unifying Framework for the Emergence of Cooperation." In: *Phys. Rev. Lett.* 95.9 (Aug. 2005), p. 098104.

[138] B. V. Schmid, U. Büntgen, W. R. Easterday, C. Ginzler, L. Walløe, B. Bramanti, and N. C. Stenseth. "Climate-driven introduction of the Black Death and successive plague reintroductions into Europe." In: *Proceedings of the National Academy of Sciences* 112.10 (Feb. 2015), pp. 3020–3025.

[139] C. M. Schneider, T. Mihaljev, S. Havlin, and H. J. Herrmann. "Suppressing epidemics with a limited amount of immunization units." In: *Phys. Rev. E* 84.6 (Dec. 2011), p. 061911.

[140] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási. "A universal model for mobility and migration patterns." In: *Nature* 484.7392 (2012), pp. 96–100.

[141] J. M. Smith. "Evolution and the Theory of Games: In situations characterized by conflict of interest, the best strategy to adopt depends on what others are doing." In: *American Scientist* (1976).

[142] J. M. Smith. *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press, Oct. 1982.

[143] J. M. Smith and G. A. Parker. "The logic of asymmetric contests." In: *Anim. Behav.* 24.1 (Feb. 1976), pp. 159–175.

[144] L. Solá, M. Romance, R. Criado, J. Flores, A. García del Amo, and S. Boccaletti. "Eigenvector centrality of nodes in multiplex networks." In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23.3 (Sept. 2013), p. 033131.

[145] A. Solé-Ribalta, M. De Domenico, N. E. Kouvaris, A. Díaz-Guilera, S. Gómez, and A. Arenas. "Spectral properties of the Laplacian of multiplex networks." In: *Physical Review E* 88.3 (Sept. 2013).

[146] A. Solé-Ribalta, S. Gómez, and A. Arenas. "Congestion Induced by the Structure of Multiplex Networks." In: *Physical Review Letters* 116.10 (Mar. 2016).

[147] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. "Limits of predictability in human mobility." In: *Science* 327.5968 (2010), pp. 1018–1021.

[148] O. Sporns. "Networks of the Brain." In: (2010).

[149] A. Szabó-Solticzky, L. Berthouze, I. Z. Kiss, and P. L. Simon. "Oscillating epidemics in a dynamic network model: stochastic and mean-field analysis." In: *Journal of Mathematical Biology* 72.5 (2016), pp. 1153–1176.

[150] G. Szabó and G. Fáth. "Evolutionary games on graphs." In: *Phys. Rep.* 446.4-6 (July 2007), pp. 97–216.

[151] A. J. Tatem, S. Adamo, N. Bharti, C. R. Burgert, M. Castro, A. Dorelien, G. Fink, C. Linard, J. Mendelsohn, L. Montana, et al. "Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation." In: *Popul. Health. Metr.* 10.8 (2012).

[152] M. Taylor, P. L. Simon, D. M. Green, T. A. House, and I. Z. Kiss. "From Markovian to pairwise epidemic models and the performance of moment closure approximations." In: *Journal of Mathematical Biology* Vol.64.No.6 (May 2011), pp. 1021–1042.

[153] P. D. Taylor and L. B. Jonker. "Evolutionarily Stable Strategies and Game Dynamics." In: *Mathematical biosciences* (Dec. 2001), pp. 1–12.

[154] M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam King, C. Schneider, and et al. "On the Use of Human Mobility Proxies for Modeling Epidemics." In: *PLoS Comput. Biol.* 10.7 (2014), e1003716.

[155] J. Travers and S. Milgram. "An experimental study of the small world problem." In: *The Structure and Dynamics of Networks* (Dec. 2011).

[156] P. Van Mieghem, J. Omic, and R. Kooij. "Virus spread in networks." In: *IEEE/ACM Transactions on Networking* 17.1 (2009), pp. 1–14.

[157] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. Van De Bovenkamp, D. Liu, and H. Wang. "Decreasing the spectral radius of a graph by link removals." In: *Physical Review E* 84.1 (2011), p. 016101.

[158] C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. "Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza." In: *Science* 312.5772 (Mar. 2006), pp. 447–451.

[159] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. "Epidemic spreading in real networks: an eigenvalue viewpoint." In: *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.* Oct. 2003, pp. 25–34.

[160] Z. Wang, A. Szolnoki, and M. Perc. "Evolution of public cooperation on interdependent networks: The impact of biased utility functions." In: *Europhys. Lett.* 97.4 (Feb. 2012), p. 48001.

[161] Z. Wang, L. Wang, and M. Perc. "Degree mixing in multilayer networks impedes the evolution of cooperation." In: *Phys. Rev. E* 89.5 (May 2014), p. 052813.

[162] D. J. Watts and S. H. Strogatz. "Collective dynamics of "small-world" networks." In: *Nature* 393.6684 (June 1998), pp. 440–442.

[163] A. Wesolowski, C. Buckee, L. Bengtsson, E. Wetter, X. Lu, and A. Tatem. "Commentary: Containing the Ebola outbreak–the potential and challenge of mobile network data." In: *PLoS Curr. Outbr.* (2014).

[164] A. Wesolowski, N. Eagle, A. M. Noor, R. W. Snow, and C. O. Buckee. "Heterogeneous mobile phone ownership and usage patterns in Kenya." In: *PLoS ONE* 7.4 (2012), e35319.

[165] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee. "Quantifying the impact of human mobility on malaria." In: *Science* 338.6104 (2012), pp. 267–270.

[166] D. H. Zanette and S. Risau-Gusmán. "Infection Spreading in a Population with Evolving Contacts." In: *J Biol Phys* 34.1-2 (Mar. 2008), pp. 135–148.

[167] M. Zargham and V. Preciado. "Worst-case scenarios for greedy, centrality-based network protection strategies." In: *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*. IEEE. 2014, pp. 1–6.

[168] G. K. Zipf. "The P 1 P 2/D hypothesis: on the intercity movement of persons." In: *American sociological review* 11.6 (1946), pp. 677–686.

UNIVERSITAT
ROVIRA i VIRGILI