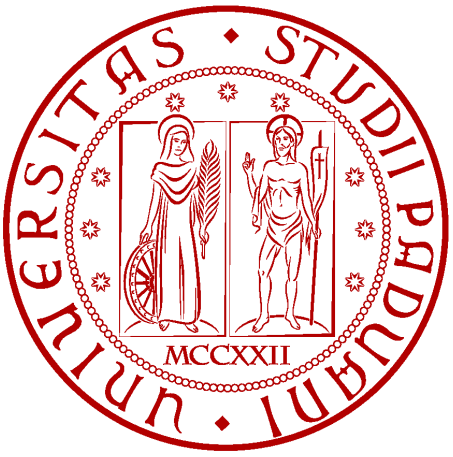

UNIVERSITÀDEGLI STUDI DI PADOVA

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"

Corso di Laurea Magistrale in Fisica

Study of phase transition on a temporal network, and applications to epidemiology

Relatore:
Prof. AMOS MARITAN
Dip. di Fisica e Astronomia "Galileo Galilei", UniPd
Correlatore:
D.ssa VITTORIA COLIZZA
Institut Pierre Louis d'Épidemiologie et de Santé Publique, INSERM e UPMC, Paris

Laureando:
ELENA SAGGIORO
Matricola:
1106832

Anno Accademico 2016 - 2017

Dedico questo lavoro alla mia famiglia.
Il vostro amore è per me fonte
di forza ed ispirazione.

Acknowledgements

I would like to express my gratitude to the Parisian research group of the EPIcx lab-Epidemics in complex environments, that welcomed me and introduced me to a completely new field of applied physics as epidemiology. My deep gratitude goes to my supervisor Dr. Vittoria Colizza, for her guidance, help and sharing of knowledge and enthusiasm. The most sincere thanks are also for Dr. Eugenio Valdano, Post Doc student at the laboratory. Without his patience and precious lessons I would have not made it through the task of learning so much in so little time. I would like to say thank you to all other member of the research group: senior researcher Dr. Chiara Poletto and PhD students Francesco Pinotti and Alexandre Darbon for having me as part of the family and making those months in Paris so enjoyable and stimulating.

To conclude, I would like express my deepest gratitude to Professor Amos Maritan, who has been my supervisor in Padua. His guidance and outstanding example in the physical research have been not only inspiring, but crucial in the last months of the work.

Contents

List of Figures

Chapter 1

Introduction

Infectious diseases represent a major burden on welfare and society. They directly threaten human health, and impact economy and development. In developing countries, they represent the top cause of death [1], and pose a particularly heavy burden on child health.Even in the developed World, where most deaths are due to non-infectious diseases, the situation may worsen in the near future. Bacteria strains are developing antibiotic resistance at a quicker pace than we can come up with new drugs [3], and vector-borne diseases (for instance, Dengue Fever [4]) are now reaching areas in which they were previously absent, as climate change impacts vector ecology. Finally, the ever more globalized World we live in is prone to breakouts of pathogens with pandemic potential, like SARS (2003), H1N1 flu (2009)[6], or, more recently MERS CoV and Ebola [7].

In the fight against infectious diseases, mathematical models have become crucial, as they provide tools to react promptly to emergencies, reduce the number of infections, optimally allocate limited resources and also design targeted containment strategies. Ever since their introduction in 1927 by Kermack & McHendrick [8], these models are based on the assumption that disease transmission and progression can be translated into a relative simple set of mechanistic equations that can be adapted to pathogens with very diverse pathophysiology and causative agents (bacteria, viruses, etc.). The epidemic is seen as an emerging collective behavior of the "microscopic" interactions among hosts [9]–[12]. This framework has allowed to tailor for epidemiology several tools and techniques borrowed from mathematics and statistical physics.

Up until two decades ago, however, the effective use of mathematical models in public health was limited by the lack of data concerning human interactions. Simplified and coarse-grained assumptions restricted the applicability to real scenarios. The picture has dramatically changed in the last years, with the outbreak of data science. The development of both new hardware and software technologies has made it possible to track real contacts and transports relevant for the spread of diseases. We now have detailed and wide records of how people interact in different settings (from schools to metro stations), and at different scales (from face-to-face proximity encounters, to mobility patterns). Detailed data do not concern only human activities: we can now keep track of livestock displacements between farms, which are spreading routes for many diseases threatening animal health, economy, but also human health.

This "data deluge" [1] has radically transformed infectious disease modelling, proving to be both a huge resource and a great challenge. High resolution data have made it possible to model entire populations down to single individuals [13]-[15], providing tailored real-time predictions of epidemic outbreaks. However, these schemes perform

[1] The data deluge, *The Economist*, 25th February 2010.

well only in specific settings, but cannot provide a general understanding of the unfolding of epidemic processes. The actual need to generalize epidemiology to realistic complex population structures has pushed towards the developing of new theoretical tools.

A very suitable way to represent disease unfolding of complex and highly resolved contact patterns is treating the disease as a dynamic process on networks [24],[21]. Networks have indeed become a common and successful tool to model populations in terms of nodes (hosts) and links (interactions among hosts) of a graph

Above all, the most challenging feature emerging from the new available data is that contacts among hosts are not fixed during the spread of the disease, but evolve in time. When the time scale of the network evolution is comparable to the one of the disease, the resulting interplay between the two has been proven to impact the outcome of the epidemic process in many non trivial ways. This is due to time correlations between contacts, that determine the length, shape and amount of routes along which the disease can spread. Traditional risk proxies, developed for static contact patterns, have shown to be insufficient to characterize this new phenomenology.Therefore, a methodology that aims at assessing the real threat a specific pathogen poses to a population must account for the temporal evolution of contacts within that population. Temporal networks represent an effective framework to model time-evolving contact patterns [22], but their interplay with spreading dynamics has been investigated mostly through numerical approaches, or in controlled settings. Recently a novel theoretical methodology has been developed to answer the need to use real data of contacts to find the epidemic threshold [2] of a system, while by-passing the microscopic simulation process [28]. In the present work we choose to exploit this methodology.

We study the diffusion of a disease on a data-based temporal network. Epidemiologically, this network is relevant for the spreading of sexually transmitted disease (STD). The main aim of our work is to compute the epidemic threshold of a supposed disease spreading in the community; that is, the critical value of a control parameter that tunes the phase transition between the over-all infectious/susceptible state of the population in the long time limit. We will firstly address the problem numerically, then we will correlate the results with the dynamical features of the underlying network and consequently tailor a theoretical model able to reproduce the above mentioned phase-transition diagram in the space of the two parameters characterizing the disease. We will also extend the space of the parameters to account for the possibility that men and women pass the disease with different probabilities, as it actually is the case of ST diseases.

1.1 Spreading of informations on a targeted population

It is no news that informations can spread among the individuals of a population. We experience it in our daily life, and we take part in the dynamical process: from the word-of-mouth going on in the neighbourhood, to the national channels of newspapers, up to the world-wide network of internet. What governs the diffusion of informations are actually two interrelated dynamics: how individuals get in touch with each other (we name it the *contact dynamics*) and how information is allowed to be transmitted (the *information dynamics*). In order to understand the meaning of these dynamics we can exploit the following simple example. Let us consider the diffusion of a rumour in a

[2]The epidemic threshold is the value of the disease contagion parameter above which a non zero fraction of the population eventually gets infected.

neighbourhood; for example we can address the following question: supposing that at instant t_0 a single person knows about the rumour, will that information reach a certain fraction of neighbours in the long time? To answer this question we firstly need to know the *contact dynamics*, that is how do individuals get in touch: are all people connected? do they change contacts over time? etc. Rules can be extracted from averaged observation of the people's behaviour, or we can be not interested in generalizing rules and just use the raw observational data, that is a sample of the detailed real contact pattern. Now, as we are interested in a rumour spreading among people, we need to know about the *information dynamics*. For instance, does a person tells the ruomur immediately when he/she knows, or waits a bit? Do men tell to women and vice-versa? Can a single eventually forget about the rumour and break the chain? Every study of information spreading needs to take into account both the contact and information dynamics. Two simple dynamics, once entangled, can give rise to the most various scenarios of information diffusion. In this same frame-work, it is possible to consider that the information spreading among people is a disease. In that case we are entering the field of epidemiology.

Chapter 2

Epidemiology

2.1 What is mathematical epidemiology?

Since the first mathematical approach to the spread of a disease by Daniel Bernoulli (1760), epidemic models lie at the core of our understanding about infectious diseases. As experimenting in vivo epidemics is not a viable option, modelling approaches have been the main resort to compare and test theories, as well as to gauge uncertainties in intervention strategies. The acclaimed work of Kermack and McKendrick (1927), defining the modern mathematical modelling of infectious diseases, has evolved through the years in an impressive body of work. At the same time, the epidemic modelling metaphor has been introduced to describe a wide array of different phenomena. The spread of information, cultural norms, and social behaviour can be conceptually modelled as a contagion process. How blackouts spread on a nationwide scale or how efficiently memes can spread on social networks are all phenomena whose mathematical description relies on models akin to classic epidemic models. Although the basic mechanisms of each phenomenon are different, their effective mathematical description often defines similar constitutive equations and dynamical behaviours framed in the general theory of reaction-diffusion processes (van Kampen, 1981). In recent years we witnessed a second golden age in epidemic modelling. Indeed, the real-world accuracy of the models used in epidemiology has been considerably improved by the integration of large-scale data sets and the explicit simulation of entire populations down to the scale of single individuals. Mathematical models have evolved into microsimulation models that can be computationally implemented by keeping track of billions of individuals. These models have gained importance in the public-health domain, especially in infectious disease epidemiology, by providing quantitative analyses in support of policy-making processes [23].

2.2 Compartmental models

Infectious diseases vary widely in their patho-physiology, clinical symptoms and etiology, resulting in diverse progression and transmission patterns. Viral diseases, like influenza, measles, chicken pox, usually confer permanent immunity after recovery, while bacterial diseases, like tuberculosis or syphilis, allow multiple re-infections of the same host. Many diseases are transmitted by direct contact between hosts, while others require vectors, such as malaria or blue-tongue. Some, like cholera, require the ingestion of contaminated water and food. Such diversity calls for modelling approaches that are

general and versatile enough to be adapted to each specific ailment, and still be a realistic description of its epidemiological features. This is commonly achieved through *compartmental* models.

Grouping individuals in health-compartments

Epidemic models generally assume that the population can be divided into different classes or compartments depending on the stage of the disease. The Susceptible individuals, denoted by S, are those who can contract the infection; the Infectious, I, are those who contracted the infection and are contagious, and recovered, R, are those who are no more infectious and cannot spread the disease any more, either because they have recovered from the disease or because they have died. Additional compartments can be used to signal other possible states of individuals with respect to the disease, for instance Exposed individuals, E, which have been infected by the disease but cannot yet transmit it. This framework can be extended to take into account vectors, such as mosquitoes for malaria, for diseases propagating through contact with an external carrier. Also, the total population in the system can be considered as fixed, or demographic process, such as migrations, births, etc., can be included.

Laws of transition between compartments

Epidemic modelling aims to describe the dynamical evolution of the contagion process within a population; namely, we are interested in the evolution of the number of infected individuals in the population as a function of time. In order to do so, we have to define the individual-level processes that govern the transition of individuals from one compartment to another. One of the simplest compartmentalizations is the SIS model: it has two states and only two possible transition rules. The first one, denoted $S \to I$, occurs when a susceptible individual interacts with an infectious individual and becomes infected. The second transition, denoted $I \to S$, occurs when the infectious individual recovers from the disease and returns to the pool of susceptible individuals. The SIS model assumes that the disease does not confer immunity and individuals can be infected over and over again, undergoing a cycle $S \to I \to S$, which, under some conditions, can be sustained forever. Another basic model is the classic three-state SIR model. In the SIR model, the transition $I \to S$ of the SIS process is replaced by $I \to R$, which occurs when an infectious individual recovers from the disease and is assumed to have acquired a permanent immunity, or is removed (e.g., has died). A diagrammatic representation for the major compartmental models is shown in Figure (2.1). The SIR and SIS models exemplify a basic classification of epidemic models given in terms of their long time behaviour. In the long time regime, the SIS model can exhibit a stationary state, the endemic state, characterized by a constant (on average) fraction of infected individuals. In the SIR model, instead, the number of infected individuals always falls to zero, but what can vary is the average number of people that contracts the infection in the active phase of the disease. In order to have a first insight in the typical SIR and SIS evolution dynamics, we present in Figure(2.2) a sketched behaviour for the density $i(t)$ of infected individuals in time. In Section (2.4) we will mathematically prove the long time limit of these results.

Spontaneous transitions and contact-dependent transitions

In the SIS and SIR models, the infection and recovery processes completely determine the epidemic evolution. The $I \to R$ and $I \to S$ transitions occur spontaneously after a certain time the individuals spend fighting the disease or taking medical treatments; the transition does not depend on any interactions with other individuals in the population. On the other hand, the $S \to I$ transition can occur only if the susceptible individual interacts with an infectious one; this is the point in which the interaction pattern between

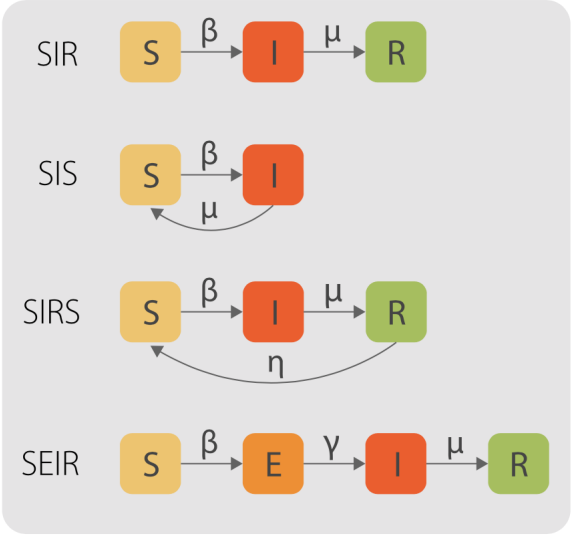

Figure 2.1: **Diagrammatic representation of epidemic models**. Boxes stand for different compartments, while the arrows represent transitions between compartments, happening stochastically according to their respective rates.

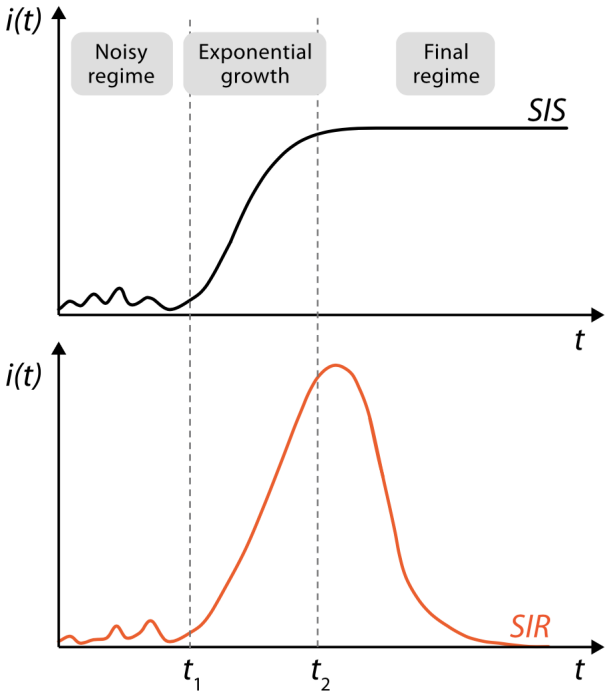

Figure 2.2: **Typical profile of the density** $i(t)$ of infected individuals vs time in a given epidemic outbreak, microscopically simulated with a stochastic infection process. In the first regime $t < t_1$, the outbreak is subject to strong statistical fluctuations. In the second regime $t_1 < t < t_2$, there is an exponential growth characterized by the details of the epidemic process. In the final regime ($t > t_2$), the density of infected individuals either converges to zero, for SIR-like models, or to a constant, possibly zero, for SIS-like models. We notice that at the early stage $t < t_2$ the two dynamics are not distinguishable.

hosts enters in the disease spreading study.

The transition probabilities

We have said that each transition between compartments happens probabilistically with a certain rate (in continuous time) or probability (in discrete time formulation). The two most relevant rates are the recovery rate μ (that is the inverse of the so called "infectious period") and the transmission rate β. We explain now how do the values of these quantities are estimated to represent real pathogens.

The distribution of the infectious period, defined as the average time in which an infected person can transmit the disease without recovering, can generally be estimated from clinical data. In a simplistic modelling scheme, the probability of recovery is often assumed constant, i.e. not dependent of the stage of the disease for the individual. In this way, for a discrete-time formulation one defines the recovery probability μ that an individual will recover at any time step, and therefore the mean infectious period is equal to μ^{-1} time steps. In a continuous-time formulation, and assuming a Poisson process, μ has to be defined as a rate (probability per unit time) and the probability that an individual remains infected for a time τ follows an exponential distribution $P_{inf}(\tau) = \mu e^{-\mu\tau}$. Thus the average infection period again is $\langle\tau\rangle = \mu^{-1}$. We notice that the Poisson assumption for the processes naturally couples to a Markovian description of epidemic models.

For what concerns the probability of the $S \to I$ transitions, the estimation from data is more complicated, because it is dependent on two factors: the characteristic infectiousness of the disease and the contact pattern. The disease-dependent part of the probability is called λ: the rate (or probability) of infection per contact in the continuous (or discrete) time formulation and is known for many diseases thanks to medical studies.The contact dependent part is the most challenging feature for epidemiologists in order to characterize the disease evolution. In the most simplistic case, if all people are in contact with an average number k of others, than the complete transmission rate $S \to I$ can be written as $\beta = \lambda k$.

The stoichiometric formalism

To scheme of transitions clear and in order to prepare a framework for a quantitative treatment of disease propagation, the epidemics is rephrased as a reaction process. Individuals belonging to the different compartments can be represented as different kinds of “particles” or “species” that evolve according to given rules for transitions among compartments. In total analogy with chemical reaction, this dynamic can be specified by means of appropriate so-called *stoichiometric equations*. In the continuum-time limit each transition is defined by an appropriate reaction rate.

The SIS model is thus governed by the stoichiometric equations:

$$\begin{aligned} S + I &\xrightarrow{\lambda} 2I \\ I &\xrightarrow{\mu} S \end{aligned} \tag{2.1}$$

where λ and μ are transition rates for infection per contact and recovery, respectively. In this model infection can be sustained forever for sufficiently large λ or small μ .

The SIR model is instead characterized by the three compartments S, I, and R, coupled by the reactions

$$\begin{aligned} S + I &\xrightarrow{\lambda} 2I \\ I &\xrightarrow{\mu} R \end{aligned} \tag{2.2}$$

For any values of λ and μ, the SIR process will always asymptotically die, after having affected a certain fraction of the population.

A useful variant of SIS is the SI model, which allows only for the first transition in reactions (2.1), i.e. individuals become infected and never leave this state. While the SI model is a somewhat strong simplification (valid only in cases where the time scale of recovery is much larger than the time scale of infection), it approximates the initial time evolution of both SIS and SIR dynamics.

More realistic models are defined in order to better accommodate the biological properties of real diseases. For instance, the susceptible-infected-recovered-susceptible (SIRS) model is an epidemic model incorporating a temporary R immunity. It can be defined from the SIR model by adding a last microscopic transition event:

$$\begin{aligned} S + I &\xrightarrow{\lambda} 2I \\ I &\xrightarrow{\mu} R \\ R &\xrightarrow{\eta} S \end{aligned} \tag{2.3}$$

where η is the rate at which the immunity of a recovered individual is lost, rendering him or her susceptible again. Another realistic model is the SEIR model. It includes the effects of exposed individuals, E, which have been infected by the disease but cannot yet transmit it, theus slowing down the dynamics with respect to SIR. The SEIR model is one of the paradigmatic models for the spreading of influenza-like illnesses and in the compact reaction-diffusion notation reads as

$$\begin{aligned} S + I &\xrightarrow{\lambda} E + I \\ E &\xrightarrow{\gamma} I \\ I &\xrightarrow{\mu} R \end{aligned} \tag{2.4}$$

All the above models can be generalized to include demographic effects (birth and death processes in the population), the age structure of the population other relevant compartments (such as asymptomatic infected individuals, gender category for sexually transmitted diseases), etc. The choice of the specific compartmental model, and the values of the parameters, are informed by the medical epidemiology of the specific disease under study.

2.3 Epidemic threshold

When a pathogen is introduced into a susceptible population, either it will cause an epidemic outbreak, or it will quickly go extinct. As we have already mentioned, the outcome will depend both on disease features, and on the structure of contacts between hosts. Researchers in computational epidemiology aim at finding a way of discriminating these two conditions, in terms of the intrinsic transmissibility λ of the pathogen. Systems show the existence of a critical value λ_c of transmissibility, called epidemic threshold, above which the disease is likely to turn epidemic. Conversely, when $\lambda \leqslant \lambda_c$, the outbreak will likely die out after a reasonably long time. Finding λ_c for a given system is crucial, as it allows both to predict the outcome of a potential pathogen introduction, and to assess the performance of prevention strategies. We notice that the choice of λ as the parameter for the definition of the critical state is arbitrary, given a disease characterized by more than one intrinsic parameter. For example, take SIS with its two tunable parameters: λ and μ. Finding $\lambda_c(\mu)$ is totally equivalent to finding $\mu_c(\lambda)$, once the endemic regime is understood for $\lambda > \lambda_c$ or $\mu < \mu_c$.

The definition of epidemic threshold is intuitively clear and simple; its mathematical formulation, on the contrary, is not always straightforward, and require caution. We need do distinguish between disease models that allow for an endemic state, and models that do not. In the former category falls the SIS model and in general all models that do not confer permanent immunity: after an initial transient, the number of infectious individuals will saturate around a steady value. In the latter category we have the SIR model, and all models giving permanent immunity. Every infectious agent will eventually recover, until no more infections occur. To address the problem generally, we will talk about SIS-like and SIR-like models. For SIS-like models the epidemic threshold discriminates between the existence of an endemic state (above λ_c threshold), and the condition where the only stable steady state is the disease-free state, i.e. there are no infectious agents (below λ_c threshold). For SIR-like models, instead, a threshold can be defined as related to the final attack rate, i.e. the total fraction of agents that will have been infected during the whole course of the epidemic. SIR-like models above threshold will lead to a large final attack rate, while below threshold the attack rate will be negligible. Despite the epidemic threshold bearing different meanings in these two different families of models, conceptually it always discriminate between the epidemic and the extinction scenarios, and for this reason it represents a valid tool for assessing the vulnerability of a population to pathogen introduction. From the mathematical point of view, too, the epidemic threshold has a unified interpretation in terms of phase transition. The two phases are clearly disease extinction and epidemic outbreak in the long time limit. By tuning model transmissibility λ (the control parameter, in the terminology of statistical mechanics) we pass from one phase to the other each time we cross the epidemic threshold. As we have seen, the measure that tells us in which phase we are (the order parameter) is different for the two families of models. For SIS-like models it is the average fraction of infected agents once the disease reaches the endemic state (i_∞), while for SIR-like models it is the final attack rate (i_{max}). The generic phase diagram is depicted in Figure (2.3), where the order parameter is called generically ρ. We observe the typical behaviour of a second-order phase transition: when one reduces transmissibility, the fraction of infected goes continuously down, until a tipping point where it becomes zero. Below this point it continues to be zero.

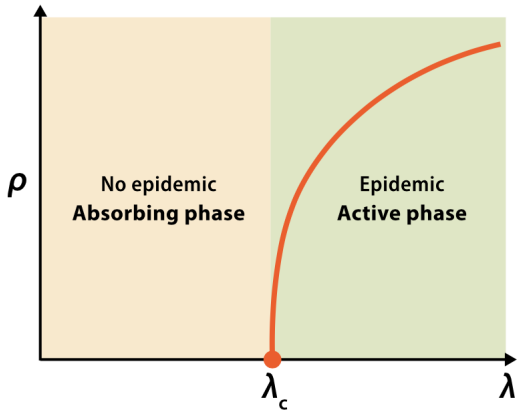

Figure 2.3: **Phase diagram of a typical absorbing state phase transition**. Below the critical point λ_c, the order parameter is zero (healthy phase in an epidemics interpretation). Above the critical point, the order parameter attains a non-zero average value in the long time regime.

As above noticed, the phase diagram $\rho(\lambda;\mu)$ for a SIS disease dynamics implicitly assumes that the recovery rate μ is fixed. We could however consider λ fixed, and find

the phase diagram $\rho(\mu;\lambda)$: this time we will find that the order parameter $\rho > 0$ for $\mu < \mu_c$, while for $\mu \geqslant \mu_c$ it goes to zero. Real systems rarely show phase diagrams as "clean" as the one shown in Fig.(2.3), as finite-size effects affects the system, making the threshold more difficult to define and compute.

2.4 Basic results from classical epidemiology

The goal of infectious disease modelling is to uncover the macroscopic behaviour resulting from the application of an epidemiological-compartmental model to a population interacting in a certain way. what we still need to address is the structure of contacts along which the disease spreads. The simplest framework, and the first to have been developed, is the so-called *homogeneous mixing*. Despite its simplicity, homogeneous mixing still represents a successful approximation in many contexts, especially within patches of meta-population models. The homogeneous mixing states that *every host in the population has the same probability of being in contact with any other host*. Therefore, each susceptible individual has the same probability of meeting an infectious individual and contract the infection in a single time step, and this probability is clearly proportional to the number of the presently infected individuals. This readily translates into a simple form of the *force of infection* α on an individual, that expresses the probability (also called the risk) at which one susceptible individual may contract the infection in a single time step:

$$\alpha = \beta \frac{I}{N} \tag{2.5}$$

where β is the rate of infection, I is the number of infected individuals and N the total size of the population. Thus, α is proportional to the fraction $i = \frac{I}{N}$ of infected individuals in the population. This form of the force of infection corresponds to the *mass-action law*, a widely used tool in the mean-field description of many dynamical processes in chemistry and physics. In some cases β explicitly split in two terms as $\beta = \lambda k$, where λ is the rate of infection per effective contact and k is the number of contacts every host establishes at each time. Knowing α, we can finally write down the time evolution of the epidemics via a set of deterministic differential equations. Let $s(t)$, $i(t)$, $r(t)$ be the densities of susceptible, infectious, recovered individuals at time t. The equations of the SIS model are :

$$\begin{cases} \frac{ds}{dt} = -\beta\, i\, s + \mu\, i \\ \frac{di}{dt} = +\beta\, i\, s - \mu\, i. \end{cases} \tag{2.6}$$

Analogously the SIR is described by:

$$\begin{cases} \frac{ds}{dt} = -\beta\, i\, s \\ \frac{di}{dt} = +\beta\, i\, s - \mu\, i \\ \frac{dr}{dt} = +\mu\, i. \end{cases} \tag{2.7}$$

Last line in both systems is redundant if we assume a fixed population ($dN/dt = 0$). Under this common assumption, correspondent to a closed system, we can express all remaining equations via the normalization conditions $r = (1 - s - i)$ and $s = (1 - i)$ (for the SIR and SIS models, respectively), thus getting rid of one variable (r or s).

If we consider the limit $i \simeq 0$, valid for example at the early stage of the epidemics, we can linearize the equations for $i(t)$ obtaining a simple linear expression for both the SIS and SIR models:

$$\frac{di}{dt} = (\beta - \mu)\, i + o(i^2). \tag{2.8}$$

Its solution is the exponential function

$$i(t) = i(0)e^{(\beta-\mu)t}, \tag{2.9}$$

that represents the evolution of small i. The early stage of infection is thus driven by a typical time scale:

$$\tau_{SIS/SIR} = \frac{1}{|\beta - \mu|}. \tag{2.10}$$

The number of infectious individuals grows exponentially if $\beta - \mu > 0$, otherwise it exponentially decrease to zero. If we define the quantity:

$$R_0 = \frac{\beta}{\mu} \tag{2.11}$$

that is generally known as the *basic reproduction number*, we see that, in the homogeneous mixing approximation, the value $(R_0 - 1)$ multiplied by μ accounts for the rate at which the disease grows exponentially the initial stage. There is therefore a condition on R_0 that discriminate the grows or the death of the disease at initial stage. This result allows one to quantitatively define the concept of epidemic threshold introduced in Section (2.3). If $R_0 > 1$ an infective agent can cause an outbreak of a finite relative size (in SIR-like models) or lead to a steady state with a finite average density of infected individuals, corresponding to an endemic state (in SIS-like models). If $R_0 \leqslant 1$, the relative size of the epidemics is negligibly small, vanishing in the thermodynamic limit of an infinite population (in SIR-like models) or leading to a unique steady state with all individuals susceptible (in SIS-like models). Besides the definition of R_0 ,that has been sometimes controversial and depends on the model, we stress again that the related concept of threshold behaviour is very general and present in all epidemic models .
For completeness, we solve Eqs.(2.6) in general case (i.e. no restriction of small i) and for fixed total population. The dynamics of infectious density is ruled by a logistic equation:

$$\frac{di}{dt} = (\beta - \mu)\, i - \beta i^2, \tag{2.12}$$

that allows for analytical solution [1]:

$$i(t) = \frac{i_{stat.}}{1 + V \exp[-(\beta - \mu)(t - t_0)]} \tag{2.13}$$

where $i_{stat} = (\beta - \mu)$, and $V = \frac{i_{stat}}{i(t_0)} - 1$. This solution gives a final infectious state $i(\infty)$ that is:

$$i(\infty) = \begin{cases} 0 & , R_0 \leqslant 1 \\ i_{stat} & , R_0 > 1 \end{cases} \tag{2.14}$$

We make a little remark here, in order to make clear that the deterministic set of equations we have used are an approximation of the real probabilistic process that governs an epidemic spreading. A more realistic analysis of epidemic models should in fact consider explicitly its stochastic nature. Accounting for this stochasticity is particularly important when dealing with small populations, in which the number of individuals in each compartment is reduced. For instance, while the epidemic threshold condition $R_0 > 1$ is a necessary and sufficient condition for the occurrence of an epidemic outbreak in deterministic systems, in stochastic systems this is just a necessary condition. Indeed, even

[1] by making a transformation of variables: $i = 1/y$.

for $R_0 > 1$, stochastic fluctuations can lead to the epidemic extinction when the number of infectious individuals is small (see Appendix A). Analogously, all the general results derived from deterministic mean-field equations can be considered representative of real systems only when the population size is very large (ideally in the thermodynamic limit) and the fluctuations in the number of individuals can be considered small. Indeed, most of the classical results of mathematical epidemiology have been obtained under these assumptions.

2.5 Connections with statistical physics models

From what we have just seen, it is natural that the statistical physics community has developed an interest in models for epidemic spreading, due to the close connection between these models and more standard non-equilibrium problems of statistical physics. In particular, the epidemic threshold concept is analogous to the concept of phase transition in non-equilibrium systems. A phase transition is defined as a change in the state (phase) of a system, characterized by qualitatively different properties, and that is experienced varying a given control parameter T. The transition is characterized by an order parameter ρ, which takes (in a system of infinite size) a non-zero value in one phase, and a zero value in another. The phase transition takes place at a particular value of the control parameter, the so-called transition point T_c, in such a way that for $T > T_c$ we have $\rho > 0$, while for $T < T_c$, $\rho = 0$. The SIS dynamics thus belongs to the wide class of non-equilibrium statistical models possessing absorbing states, i.e. states in which the dynamics becomes trapped with no possibility to escape. An example of a system with an absorbing state is the contact process (CP) (Harris, 1974), where all nodes of a lattice or network can be either occupied or empty. Occupied nodes annihilate at rate 1; on the other hand, they can reproduce at rate λ , generating one offspring that can occupy an empty nearest neighbour. Notice that the contact pattern here is fixed and regular, with every node contacting one of its neighbours at each time step. The contact process experiences an absorbing state phase transition at a critical point λ_c between an active phase, in which activity lasts forever in the thermodynamic limit, implying a finite average density of occupied nodes, and an absorbing phase, in which activity eventually vanishes, corresponding to an empty system. Drawing the connection with the SIS model, the active phase is given by the infected state, and the absorbing phase by the state where no individual is infected. The order parameter is therefore the density of infected individuals $\rho \to i_{stationary}$, and the control parameter is given by the basic reproductive number $T \to R_0$. The critical point is the epidemic threshold $R_{0,c}$, that separates the infected from the healthy phase. It is interesting to note that the dynamics of the SIS process on lattice is essentially identical to that of the contact process; indeed, the difference between the SIS and the contact process lies exclusively in the number of off-springs that an active individual can generate. While in the contact process one particle generates always on average one offspring per unit time, an infected individual in the SIS model can infect all his or her nearest neighbours in the same time interval.

Chapter 3

Epidemics on Static Networks

3.1 The convenient choice of networks

It is easy to understand that the homogeneous mixing assumption, used in the previous chapter to write the deterministic equations of an epidemic processes, may be inadequate in several real-world situations. As one can guess, in the real world individuals have large heterogeneity in the contact rate (i.e. β_i depends on the individual i and takes wide range of values), or present a specific frozen pattern of interaction (i.e. individuals do not mix), or are in contact only with a small part of the population (i.e. they do mix, but inside small subgroups). These features may have different relevance, depending on the disease or contagion process considered, but usually have an overall non trivial impact on the behaviour of the epidemics.
To actually capture real features of contacts among humans, epidemiologists have been reconstructing the net of "who had been in contact with whom" during a particular time frame through surveys and questionnaires, asking, for instance, to list list all people you had met that particular day. These data allowed researchers to uncover features of human interactions, which prompted the need of going beyond the homogeneous mixing assumption. The data showed clearly that most real-world systems have very complex connectivity patterns, dominated by large-scale heterogeneities that cannot be captured by average values and need to be described via heavy-tailed statistical distributions. Analogous findings concerned also the ecological and biological world. For all these systems, an averaged or homogeneous approximation of the contact pattern between individuals dramatically fails in predicting the spreading of epidemics. Therefore it becomes necessary to include individuals' contact pattern structure into the mathematical modelling approaches. Network theory turned out to be a suitable general framework, complete with powerful tools, to account for interactions among individuals in detail.

3.2 Mathematical representation

A network is a representation of an interacting population, in terms of a mathematical entity called the graph. A graph is an object composed of a set of nodes (vertices), and links (edges) that connect pairs of nodes. In our context, nodes represent the hosts of our population, and links represent the interactions relevant to the spread of the disease.
Edges can represent a bidirectional interaction between vertices, or indicate a precise directionality in the interaction. In the first case we talk about undirected networks, and in the second case about directed networks. From an epidemiological point of view, a network being directed is indeed relevant, since it imposes restrictions on the possible

paths of propagation of the contagion. Furthermore, links can be binary relationships (0,1), or weighted, encoding the amount of times the link has activated.
One of the advantages of the network approach is that a graph has a natural algebraic representation in terms of adjacency matrix. Given a graph of size N (i.e., with N vertices) we can always associate to it its adjacency matrix A. It is the $N \times N$ matrix with elements $A_{ij} = 1$ if an edge is connecting nodes i and j and zero otherwise. If the network is undirected, then A is symmetric ($A = A^{\dagger}$); if the network is weighted, A_{ij} can assume values other than 0, 1, encoding link weights. The study of properties of a graph in relationship to the characteristic polynomial, eigenvalues, and eigenvectors of its adjacency matrix is called spectral theory.The spectrum if a graph is the multiset of its adjacency matrix eigenvalues. While the adjacency matrix depends on the vertex labelling, its spectrum is a graph invariant. An undirected graph, that has a symmetric adjacency matrix, has real eigenvalues and a complete set of orthonormal eigenvectors.
Degree and degree distribution The first elementary concept that arises when dealing with networks is the one neighbourhood. Two nodes are neighbours if there is a link among them, and the neighbourhood of a node is the set of its neighbour nodes. This results in a theoretical measure called the *degree of a node*: the number of connections this node establishes with other nodes (the size of its neighbourhood). In formulas, the degree k_i of node i is therefore

$$k_i = \sum_{j|A_{ij}\neq 0} 1. \tag{3.1}$$

For directed networks, we discriminate between incoming and outgoing degree. For weighted networks, we introduce the strength of a node, i.e. the sum of the weights of its links:

$$s_i = \sum_j A_{ij} \tag{3.2}$$

Networks are often characterized in terms of their degree distribution: the statistical distribution of node degrees. When nodes establish links randomly, the resulting degree distribution is Poisson-like, with small dispersion around a mean value. On the other hand, networks with deeply non random connection patterns exhibit heterogeneous degree distributions, with large, sometimes diverging, variance. The most popular heterogeneous distribution in this context is the power-law $P(k) \sim k^{-\gamma}$, as many real networks are found to have such a degree distribution. It is often informing to consider the moments of the degree distribution $\langle k^n \rangle = \sum_k P(k)k^n$. The first moment, the average degree $\langle k \rangle = 2E/N$ is twice the ratio between the number E of edges and the number N of nodes and provides information about the density of the network. A network is called sparse if its number of edges E grows at most linearly with the network size N and the mean degree is less than one; otherwise, it is called dense. The second moment informs on the dispersion of the degree distribution around the mean value, therefore on the heterogeneity of the number of contacts each node establish.
Degree correlation Two-vertex degree correlations can be conveniently measured by means of the conditional probability $P(k'|k)$ that an edge departing from a vertex of degree k is connected to a vertex of degree k'. A network is called degree-uncorrelated if this conditional probability is independent on the degree of originating vertex k. In this case, $P(k'|k)$ can be simply estimated as the ratio between the number of edges pointing to vertices of degree k' ($k'P(k')N/2$) and the total number of edges ($\langle k \rangle N/2$), to yield

$$P(k'|k) = k'P(k')/\langle k \rangle. \tag{3.3}$$

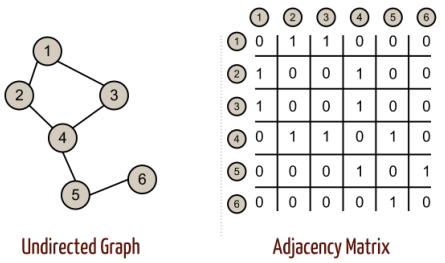

Figure 3.1: A graph and its adjacency matrix.

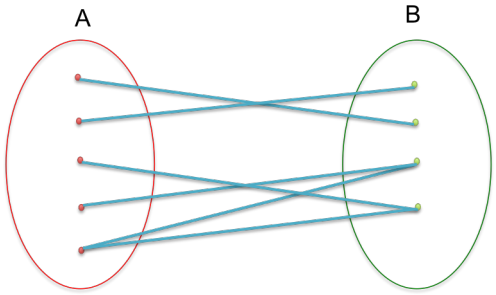

Figure 3.2: A bipartite graph.

The empirical evaluation of $P(k'|k)$ turns out to be quite noisy in real networks, due to finite-size effects. A related, simpler, measure of correlations is the average degree of the nearest neighbours of vertices of degree k, $k_{nn}(k)$ which is formally defined as

$$k_{nn}(k) = \sum_{k'} k' P(k'|k) \tag{3.4}$$

For degree-uncorrelated networks $k_{nn}(k) = \langle k^2 \rangle / \langle k \rangle$ does not depend on k, as $P(k'|k) = k'P(k')/\langle k \rangle$. Therefore, a varying $k_{nn}(k)$ is the signature of degree correlations. The analysis of empirical networks has suggested a broad classification of networks in two main classes, according to the nature of their degree correlations. *Assortative* networks exhibit an increasing $k_{nn}(k)$ with increasing values of k: indicative that high degree nodes tend to connect to high degree nodes, while low degree nodes are preferentially attached to low degree nodes. *Disassortative* networks, on the other hand, show a decreasing $k_{nn}(k)$ function, suggesting that high degree nodes connect to low degree nodes.
Bipartite graphs A type of graph that will be relevant in this work is the *bipartite graph*. In this graph nodes are dived into two groups (A and B for example) and edges can join only two nodes belonging to different groups.

3.3 Epidemic threshold on static networks

Different approaches have been developed for computing the epidemic threshold on static networks. Conceptually, they can be classified according to their way of taking into account the structure of the network:

- *quenched* network:
  The structure of the network is fixed, and expressed in terms of its adjacency matrix A. The resulting epidemic threshold is characteristic of that particular matrix;

- *annealed* network:
  Only one (or more) statistical property of the network is relevant.
  For example, it can be the degree distribution $P(k)$. Therefore an ensemble of adjacency matrices $\{A\}$ corresponding to the given $P(k)$ are equally probable and the threshold is computed over that ensemble.

We emphasize that static network means that its edges are fixed in time. Therefore while the epidemics temporally evolves on the networks by changing the "health" state of the nodes, the underling network structure does not change. Thus only one dynamics,

the disease's, has to be studied.

Here we review two approaches aimed at finding the epidemic threshold in static networks, because they are instructive for the work we will develop in the following chapters. The first approach is based on annealed networks, and uses deterministic differential equations; the second is based on quenched networks, and uses eigenvalue analysis of the adjacency matrix .

3.3.1 Annealed network: The degree-block approximation

This approach computes the threshold as a function of a given degree distribution $P(k)$ that defines the network structure [30]. Its original formulation requires the absence of degree-degree correlations, but can be generalized to include them. Its core idea is to use the same differential equation approach that we have seen for homogeneous mixing (Eq. (2.6)), by splitting each compartment by degree. An infectious node of degree k, and one of degree k' are no longer in the same compartment, but in I_k and $I_{k'}$, respectively. This approach is called *degree-block approximation*, or degree-based mean field as it assumes that all nodes with the same degree are statistically equivalent. Let S_k,I_k be the number of susceptible and infectious nodes in degree class k. Densities are obtained by dividing S_k,I_k by the number of nodes with degree k N_k.

$$\begin{cases} s_k = \frac{S_k}{N_k} \\ i_k = \frac{I_k}{N_k} \end{cases} \tag{3.5}$$

The average densities are obtained through the degree distribution :

$$s = \sum_k s_k P(k) \quad i = \sum_k i_k P(k)$$

The system of equations governing SIS dynamics, for small i, is:

$$\frac{di_k}{dt} = \sum_{k'} [\lambda k P(k'|k) - \mu \delta_{kk'}] i_{k'} + \mathcal{O}(i_k^2) = \sum_{k'} C_{kk'} i_{k'} + \mathcal{O}(i_k^2) \tag{3.6}$$

in which we have defined matrix $C_{kk'} = \lambda k P(k'|k) - \mu \delta_{kk'}$.
We now assume degree uncorrelation: $P(k'|k) = k'P(k')/\langle k \rangle$. If we define two column matrices: $(C_1)_k = k$, and $(C_2)_k = kP(k)$, we can rewrite the uncorrelated C matrix as $C = \lambda C_1 C_2 \dagger / \langle k \rangle - \mu I$. Therefore, C has rank equal to 1, and its principal eigenvector is C_1, with eigenvalue $\lambda \langle k^2 \rangle / \langle k \rangle - \mu$. By setting to zero the maximum eigenvalue , we recover the epidemic threshold:

$$\left(\frac{\lambda}{\mu}\right)^{SIS}_{critical} = \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{3.7}$$

For completeness, we write the threshold for SIR as well:

$$\left(\frac{\lambda}{\mu}\right)^{SIR}_{critical} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{3.8}$$

These results are especially interesting in the case of heterogeneous degree distributions. For common power law $P(k) \sim k^{-\gamma}$, we recall that a well-defined mean over $k \in [1, +\infty]$ exists inly when $\gamma > 2$, and has a finite second momentum only if $\gamma > 3$. Therefore, in

the thermodynamic limit $N \to \infty$, the second moment diverges $\langle k^2 \rangle \to \infty$ for $2 < \gamma < 3$. In this case the epidemic threshold vanishes in the thermodynamic limit: $(\lambda/\mu)_c \to 0$, meaning that every value of λ can give rise to an endemic state. For real systems that are not infinite, the result of Eq.(3.7) means that heterogeneous static networks (built via the given $P(k)$) get indefinitely more vulnerable as they get bigger, because hubs get more and more connected, and can easily sustain an epidemic.

3.3.2 Quenched network: Spectral theory

We treat this case more in detail, because it will be the starting point for the theory on temporal networks we will develop in the following Chapter.
A quenched network is a specific fixed network, represented by its adjacency matrix A. In this case we consider time as discrete, and at each time step a susceptible node can be infected by an infectious neighbour with probability λ and it can recover with probability μ. Each node is assigned a binary variable X_i, corresponding to the infectious status ($X_i = 1$) or susceptible status ($X_i = 0$).
Infection dynamic is translated into a discrete-time Markov process with 2N possible states, corresponding to the possible infection configurations. The transition matrix among all these states cannot be written in general. For this reason, it is customary to neglect correlation among infectious statuses at a given time[1], as in general the state of a node does not influence the state of another, at the same time step. Under this assumption, it is possible to decompose the joint probabilities in terms of single node probabilities:

$$\text{Prob}\left(X_i(t) = 1 \wedge X_j(t) = 1 \wedge X_k(t) = 1 \dots\right) = p_i(t)p_j(t)p_k(t)\dots \tag{3.9}$$

where we define $p_i(t) = \text{Prob}\left(X_i(t) = 1\right)$ as the probability of node i to be in state 1 (infectious) at time t. In this way, the probability of a particular N-nodes configuration can be written in terms of single-node probabilities $p_i(t)$. The advantage is that for these $p_i(t)$ an evolution equation can be written [26] :

$$p_i(t+1) = [1 - q_i(t)]\,[1 - p_i(t)] + (1-\mu)p_i(t) + \mu\,[1 - q_i(t)]\,p_i(t) \tag{3.10}$$

where $q_i(t)$ is the probability that node i does not get infected by *any* of its infectious neighbours:

$$q_i(t) = \prod_{j=1}^{N}\left(1 - \lambda r_{ji} p_j(t)\right) \tag{3.11}$$

being r_{ij} the contact probability and λ the probability of transmit the disease.

Equation(3.10) is composed by three terms, each accounting for a transition probability that leads to node i being infected at time $t+1$. The first term is the probability that node i is susceptible at time t, $(1 - p_i(t))$, and contracts the disease because of contagion with *at least one* infectious neighbour, $(1 - q_i(t))$; the second term is the probability that node i is infectious at previous time step and do not spontaneously recovers $(1-\mu)$; the third and last is the probability that node i is infectious at time t, recovers and gets re-infected. What remains to be defined is the contact probability r_{ij}: we can picture it as the transition probability of a random walker between nodes i and j. If a number n_i of r.w. leaves node i at each time step, then:

$$r_{ij} = 1 - \left(1 - \frac{w_{ij}}{s_i}\right)^{n_i} \tag{3.12}$$

[1] For a treatment of the two nodes correlation, see [25].

where w_{ij} , s_i are the weight of link i-j and the strength of node i, respectively. If $n_i = 1$, that is there is one try for infection per time-step: $r_{ij} = \frac{w_{ij}}{s_i}$; if $n_i \to \infty$, all reachable nodes are contacted at each time step, $r_{ij} = a_{ij}$, whether the network is weighted or not. By choosing the second limit $r_{ij} = A_{ij}$, and performing simple algebraic re-arrangement of Eq.(3.10), we eventually find :

$$\boxed{p_i(t+1) = 1 - [1 - (1-\mu)p_i(t)] \prod_j \Big(1 - \lambda A_{ji} p_j(t)\Big) \quad .} \tag{3.13}$$

This equation is the key to find the threshold.
We are interested in studying the behaviour of p_i around the the disease free state $p_i = 0$, that is clearly an equilibrium solution of $p_i(t+1) = p_i(t)$. Analogously to what we have done in the annelaed case, we wish to see if there is a λ_c for which the zero state passes from being a stable to an unstable equilibrium, and thus leading to an epidemic outbreak. In order to do that, we linearize the equation around $p_i \simeq 0$:

$$p_i(t+1) = (1-\mu)p_i(t) + \lambda \sum_j A_{ji} p_j(t) + \mathcal{O}(p_i^2). \tag{3.14}$$

This can be rewritten in matrix form, by interpreting $p_i(t)$ as the i-th element of a size N vector $\vec{p}(t)$:

$$\vec{p}(t+1) = \Big[(1-\mu)I + \lambda A^\dagger\Big] \vec{p}(t). \tag{3.15}$$

The above equation is a linear autonomous difference equation, as time is discrete [53]. General theory demonstrates that the zero solutions of Eq.(3.15) is asymptotically stable if and only if the spectral radius of matrix $(1 - \mu + \lambda A^\dagger)$ is larger than 1. The spectral radius of a matrix is the largest of the absolute values of the eigenvalues : $\rho(M) = \max\{|\lambda|, \lambda \text{ is eigenvalue of } M\}$. This is a fundamental result, that we will use also in the following chapters [2]. The stability condition coincides with the requirement for the threshold. We have:

$$\rho\Big[(1-\mu)I + \lambda A^\dagger\Big] = 1 \implies \left(\frac{\lambda}{\mu}\right)_{critical} = \frac{1}{\rho(A)}. \tag{3.16}$$

To derive the above relation we exploit the fact that $\rho(aI + B) = a + \rho(B)$, when $a = 1 - \mu > 0$ and $B = A$ has all non negative entries, therefore satisfies the Theorem of Perron-Frobenius [3]. Moreover, we use $A = A^\dagger$ of undirected graphs.

We recall that Eq.(3.16) for the epidemic threshold $(\lambda/\mu)_c$ is true under the independence of nodes approximation of Eq.(3.9) and it is known to be a lower bound estimate of the real epidemic threshold. However this result approaches the real value, computed via microscopic simulations of infection processes, with surprisingly high accuracy given the simplicity of the expression. This result, derived for the SIS model, holds for the SIR model too, as their equations are the same close to $p_i \simeq 0$.

In conclusion, in the quenched mean field approach, the epidemic threshold on a generic network of both the SIS and SIR model is completely determined by the maximum and absolute value of the adjacency matrix's eigenvalues.

[2] See Appendix D for details.
[3] See Appendix E for the detailed calculus.

Chapter 4

Epidemics on temporal networks

So far we have assumed that the topology defining the network is static: the set of nodes and links do not change over time. However, many real networks are far from static, their links being created, destroyed, and rewired at some intrinsic time scales. When contact patterns evolve in time, we need to go from a static network representations to a temporal version, where links activate and deactivate in time. In terms of adjacency matrix, we go from a single matrix A for a static network, to a matrix which is function of time A(t). Hence, while static networks are purely topological objects, temporal networks have an embedded dynamic process driving link evolution in time. We stress that while the definition of temporal network is a straightforward extension of the definition of static network, the emerging properties of this new object are conceptually and phenomenologically different, and cannot be in general recovered as a simple extension of what we know about static networks.

The first conceptual problem arising from the temporal dimension is the definition of path. A path on a static network is a set of edges such that an edge ends where the next edge in the path begins. Paths matter a lot, as far as the spread of disease is concerned, because if there is a path going from node i to node j, it means that i can affect the state of j. If the static graph is undirected, then every path connecting i to j will also connect j to i ; if it is directed, this symmetry is broken and a connection $i \to j$ does not in general imply $j \to i$. The reachability property "$\to$" is transitive also for directed networks: if $i \to j$, and $j \to k$, then $i \to k$. Things get more complicated when we deal with temporal networks. This happens because paths exist only if they are time-respecting. Two links, connecting i with j and j with k, represent a path from i to k if and only if link i-j activates *before* link j-k , otherwise, no information starting in i will be able to reach k. As a result, a temporal network is always directed, at least in the temporal dimension that encodes causality, and reachability is in general not transitive, because the chain $i \to j \to k$ implies $i \to k$ only if paths are arranged in a time-respecting way. These complex features of temporal networks matter a lot to us, because the spread of a disease on a network structure can be examined in terms of infection paths, which themselves must be time-respecting. Time can be treated as a continuous or a discrete variable. If time is discrete, the temporal network is a sequence of snapshots which themselves can be seen as static networks. Assuming discrete time can be seen as an approximation, but it is both practically sound, and theoretically convenient. Indeed, most part of empirical datasets regarding temporal networks are intrinsically discrete and as a result, discrete time becomes a natural choice in many empirical settings.

4.1 Mathematical representations of temporal networks

4.1.1 Lossy representation

The representations of temporal networks are divided into lossless and lossy. Lossless representations carry all the information about the temporal network, and in practice we can identify them with the network itself. In lossy representation, instead, some information about the original temporal network is lost. Here we describe only the representations that are relevant for our work.

The most common lossy representation consists in projecting all the temporal dynamics onto a static, aggregated, network. Here we consider two aggregation schemes, which we call HOM and HET. HOM aggregation consists of building a single static network where two nodes are linked if they are in contact at least once in the temporal network. The weight of links is the same for all and is equal to the average:

$$A_{ij}{}^{HOM} = w = \frac{\sum_t \sum_{ij} A_{ij}(t)}{\sum_{ij_{|A_{ij}\neq 0}} 1}. \tag{4.1}$$

HET's topology is the same as HOM's, but HET links are weighted by the number of times they are in contact in the temporal network, as it is clear from the definition in Eq.(4.2).

$$A_{ij}{}^{HET} = \sum_{t=1}^{T} A_{ij}(t) \tag{4.2}$$

HET aggregation, despite losing all temporal correlations, can account for the fact that some ties are stronger than others, i.e. they occur more frequently.
Another static, lossy representation is the accessibility graph: a directed graph in which a link going from i to j exists if there is a time-respecting path from i to j. Accessibility graph can be interpreted as a static entity, associated to the accessibility matrix:

$$\begin{aligned} \mathcal{P}_T = \prod_{t=1}^{T} [I + A_t] = I + A_1 + A_2 + ... + A_T \\ + A_1A_2 + A_1A_3 + ... \\ + ... \\ + A_1A_2A_3..A_T \end{aligned} \tag{4.3}$$

where time is assumed discrete, and A_t is the adjacency matrix of t-th snapshot of the temporal network. The name of accessibility matrix arises from the fact that $[A(t)A(t+1)]_{ij}$ counts the number of paths that start at node i at time t and end in node j at time t+1. The presence of the identity matrix in Equation (4.3) accounts for the fact that paths can contain waiting times. In other words, if we imagine a path as a route of a walker, without the identity term such walker would have to move from node to node at each time step. The identity terms allows it to stay still in one place indefinitely. Interesting studies on real dataset [46], have shown that the number of time respecting paths are far less than the number that exist in the HET aggregation scheme, thus proving the importance of causality in data based temporal network.

4.1.2 Lossless representation

The first, most natural lossless representation is the *sequence of snapshots*. Given a discrete-time temporal network of T time steps, we represent the links active at each time step

as a static graph. The temporal network thus becomes a ordered sequence of graphs, or, algebraically, an ordered sequence of adjacency matrices:

$$\mathcal{A} = \{A_1, A_2, \dots A_T\}. \tag{4.4}$$

The second representation is called *multilayer*. This retains a special importance for the understanding of the rest of the work. In general, a multilayer network is just a network whose nodes can be grouped in different subsets (called layers), with intra-layer and inter-layer connections. A temporal network can be seen as multilayer when we consider time snapshots as layers, with a copy of each node appearing in every layer. In Figure (4.1) we show the three possible "types" of links appearing in a multi-layer network:

- intra-layer: a link whose nodes belong to the same layer (black solid line);
- inter-layers diagonal: a link connecting the copy of the same node in two different layers (black dashed line);
- inter-layers non-diagonal: a link connecting two different nodes in two different layers (red dashed line).

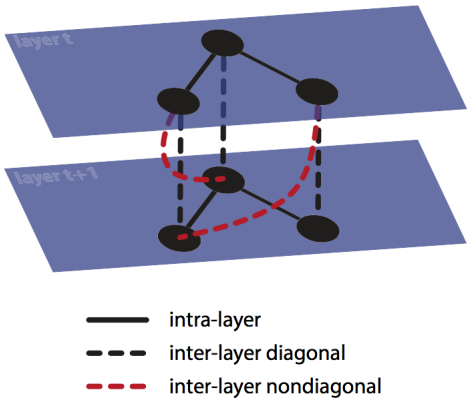

Figure 4.1: **Temporal network as a multilayer object**. We represent a temporal network of 3 nodes and 2 time steps. We picture the three possible types of links. Intra-layer links correspond to terms A_{ij}^{tt} of the adjacency tensor. Inter-layer diagonal links correspond to A_{ii}^{ts} terms. Inter-layer non-diagonal correspond to terms A_{ij}^{ts}, with both $t \neq s$ and $i \neq j$.

Now the natural question that arises is how to translate the links of the temporal network into the multi-layer object and its possible connection. We will use the choice made by Valdano et al in [28], and call it the *inter-layer representation*. Let us consider the sequence of snapshots $\mathcal{A}$ in Eq. (4.4). We want to build a corresponding multi-layer structure, that retains all contacts informations. The multi-layer object is composed of T layers, one for each snapshot and each containing a copy of all the N nodes. Hence, each node of the multilayer is identified by a pair (i, t), with $i = 1, \dots, N$ and $t = 1, \dots, T$. We now define links among layers with the following "rules":

- each node in layer t is connected to its own image in layer $t + 1$ via a directed link;
- a link between i and j at time t, $A_{ij}(t) = 1$, translates into a pair of directed links, one going from i at in layer t to j in layer $t + 1$, and one going from j at in layer t to i in layer $t + 1$;

- we enforce periodic boundary conditions: layer T is connected to layer 1 with the above rules, assuming the equivalence $T + 1 \equiv 1$

The multilayer network can therefore be expressed by means of the *adjacency tensor* $\mathbf{A}$, with four indices: two temporal and two referring to nodes.[1] The generic term is $\mathbf{A}_{ij}^{ts}$ and it encodes the "value" of the link between the multi-layer nodes (i, t) and (j, s). Following the above listed rules, the only non-zero terms are $\mathbf{A}_{ii}^{t\,t+1} = 1$ and $\mathbf{A}_{ij}^{t\,t+1} = A_{ij}(t)$. Summarizing, a component of the adjacency tensor is:

$$\mathbf{A}_{ij}^{tt'} = \delta^{t+1,t'} \left[\delta_{ij} + A_{ij}^{(t)} \right]. \tag{4.5}$$

We stress that in this particular multilayer there are no intra-layer links, i.e. no links among nodes belonging to the same layer, resulting in a multi-partite structure. In Figure (4.2) we schematically show the structure of our multilayer structure.

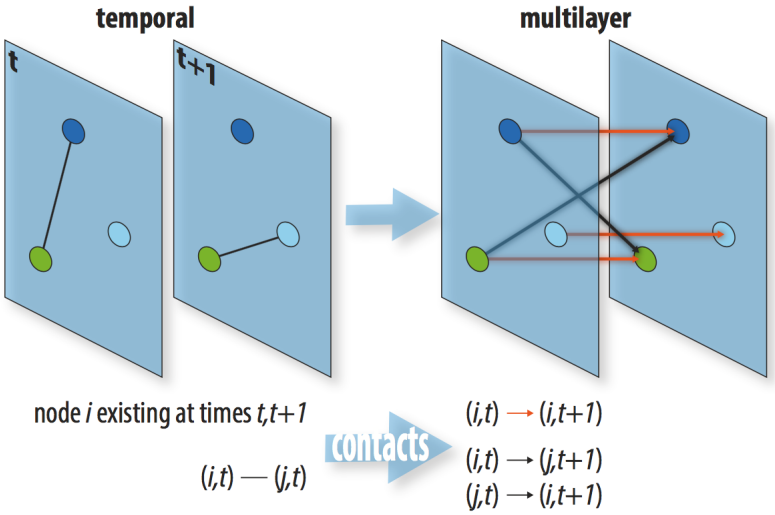

Figure 4.2: **Multilayer representation of the temporal network.** We show the rules to build the multilayer representation starting from the temporal network described as a time ordered sequence of snapshots . We focus on 3 nodes and 2 time steps.

4.2 New statistical properties for temporal networks

In many contexts we do not have, or we are not interested in, the exact structure of a network, but just in the statistical distribution of some microscopic properties, and want to study how they influence diffusion. In the context of static networks, researchers have characterized the role of different centrality measures, like degree, betweenness, clustering [24]. They have shown how different distributions of these quantities alter the epidemic outcome. Doing the same with temporal networks is not trivial. A simple generalization of tools used in static network does not always lead to meaningful result. Let us consider degree, for instance. Defined as the number of contacts a node performs, now its degree changes in time. Then one can define the aggregated degree, which is the number of nodes met at least once. These measures, however, have proven to be ineffective in some contexts, as they say nothing about the epidemic risk. As a result,

[1] For a detailed description see Appendix B.

one needs new centrality measures to account for the timely behaviour, and can be used as simple risk factors. Here we review two of the newly developed statistical measures for temporal networks.

In the context of information diffusion, we find the *social strategy* property. It uses both instantaneous degree and aggregated degree. Let us consider a discrete-time temporal network of T snapshots. We can fix a time window of δ, and a time t. We aggregate the network in the interval $[t, \delta + t]$, following the HET scheme. For every node i we consider k_i and s_i, its degree and strength, respectively. Social strategy γ_i of node i at time t, and with time window δ, is then defined as

$$\gamma_i = \frac{k_i}{\delta s_i}. \tag{4.6}$$

When $\gamma \to 0$ it means that the node makes almost all its contacts always with the same set of nodes in the time window $[t, \delta + t]$, while the opposite regime, $\gamma \to 1$ characterizes a node which changes its neighbours often. Social strategy is thus able to discriminate between a memory-driven behaviour, and an exploratory one.

Another temporal measure is called the *loyalty* [33]. Loyalty, like social strategy, aims at assessing node memory. However, while social strategy is averaged over a time window, loyalty quantifies the behaviour of a node from one snapshot to one that immediately follows. Specifically, loyalty measures the fraction of preserved neighbours of a node for a pair of two consecutive network configurations in time. Let us consider snapshots t and t+1, and let ν_i^t be the set of neighbours of node i in snapshot t. Then loyalty $\Theta_i^{t,t+1}$ of node i from time t to time (t+1) is defined as the Jaccard index between ν_i^t and ν_i^{t+1} :

$$\Theta_i^{t,t+1} = \frac{|\nu_i^t \cap \nu_i^{t+1}|}{|\nu_i^t \cup \nu_i^{t+1}|} \tag{4.7}$$

Loyalty takes values in interval [0, 1]. $\Theta = 0$ means no neighbours are retained, while $\Theta = 1$ means that the node keeps exactly the same neighbours from t to t+1. In the case of directed networks, loyalty can be defined for in-neighbour or out. In addition to Jaccard index, loyalty can also be expressed in the adjacency matrix formalism, as follows:

$$\Theta_i^{t,t+1} = \frac{\sum_j \left(A(t)A(t+1)^\dagger\right)_{ji}}{\sum_j \left[A_{ji}(t) + A_{ji}(t+1) - (A(t)A(t+1)^\dagger)_{ji}\right]}. \tag{4.8}$$

The informations that social strategy and loyalty carry regard the time changing dynamics of contacts, and they have been proved to be good estimators of the centrality of a node in a temporal network; for example they capture node risk with respect to disease dynamics [48] and can be exploited to design effective vaccination strategies.

4.3 Null models

Given a data-based temporal network, it is important to find out which is the dynamical feature that mostly impacts on the way an information flows onto it. One can achieve this knowledge by means of "randomization" of some structure of the original network. We start with the real temporal network, and design a numerical protocol that destroys some desired features, obtaining what is commonly called a *null model* of the original network. Conceptually, while designing synthetic models allows you to control exactly

which features you put in, a null model lets you choose what features you take out, leaving all the residual properties untouched. The objective is thus to isolate one specific temporal feature and to study its impact on, for instance, the disease spreading dynamics. It is often very hard to isolate a single property, as a null model often destroys more than one feature, being many of them tightly entangled. Therefore, null models have to be carefully used and understood. We describe here some null models that we have extensively used in order to characterize our real data-driven network.

- The fist null model is called RESHUFFLE and consists in a simple randomization of the snapshots' order.
- The second null model called RECONFIGURE acts instead on single links. Let us define a timestamped contact as a triplet (i,j;t), i.e., a link between i and j occurring at time step t. RECONFIGURE consists in a random reassignment of contact timestamps: two contacts (i,j;t), (k,l;s) are randomly selected, and their timestamp switched: (i,j;s), (k,l;t).

At the macroscopic level, both null models preserve the aggregated network. RESHUFFLE clearly destroys the activity timeline, i.e. the time-ordered number of active links, breaking all possible seasonal patterns. RECONFIGURE , on the other hand, preserves it. Microscopically, RESHUFFLE preserve link correlations inside each snapshot, while RECONFIGURE does not. They both destroy all time correlations and self-correlations in link activation: memory is lost, bursty inter-activation time is lost, time-respecting paths and temporal motifs, i.e. correlation patterns in link activations, are broken.

- Thirdly, in ANONYMIZE we randomize the identity of the nodes, independently inside each snapshot.
- Finally, SWAP is a cross-link operation. It picks two random edges inside a snapshot and, if they have no common nodes and if the new links do not already exist, switches the nodes.

SWAP and ANONYMIZE preserve the activity timeline and the "identity time line", i.e. the stronger request on the "identity" of the active nodes at each time step, while in turn destroys all correlations in link-activation and the aggregated network. SWAP moreover preserves the degree of each node at each time step.

- One last null model is RANDOM TIMES (R.T.), that takes the list of all contacts occurred in the total period and re-assigns each of them to a randomly chosen time step, therefore not respecting the original daily activity of links. The side effect is that it actually uniforms the activity time line of edges.

R.T. basically preserves the aggregated network and destroys all temporal correlations. RECONFIGURE and RESHUFFLE are therefore sub-cases of RANDOM TIMES: in RECONFIGURE we enforce the preservation of the activity time line while in RESHUFFLE we preserve snapshot topology.

Table 4.1: **Null Models for Temporal Networks**

Name	activity timeline	Aggregated network	Snapshot topology	microscopic time correlation	$k_{node}(t)$
RESHUFFLE	×	✓	✓	×	×
RECONFIGURE	✓	✓	×	×	×
ANONYMIZE	✓	×	×	×	×
SWAP	✓	×	×	×	✓
RANDOM TIMES	×	✓	×	×	×

Tab.(4.1) summarizes the features of all these null models, with × meaning "the null model destroys the feature" and ✓ meaning "the null model preserves the feature"
The design of appropriate null models allows to "switch off "selected types of temporal and topological correlations. To find out the role of those features on the (epidemic) process spreading on the temporal network, it is necessary to evaluate a spread parameter (such as the epidemic threshold) on several null model networks. Then, by contrast with the original, we deduce whether the "switched off feature" influences or not the criticality of the information dynamics. We highlight that we cannot tell the role of the preserved feature, but rather see how strong is the impact of the destroyed one.
It is worth saying that it can be hard to isolate one specific characteristic of the original network, because many topological/temporal features are reciprocally "entangled", meaning that removing one often modifies another. This implies that evaluating the importance of a single property can be sometimes impossible; other times we can do it indirectly, by comparing two null models that differ by a specific feature. In this sense, we see that comparing R.TIMES with RESHUFFLE enlightens on the influence of snapshot topology; while comparing R.TIMES with RECONFIGURE informs on the activity time line.

4.4 The quenched and annealed limits in temporal networks

They key feature that marks the signature between temporal and static networks is the separation of time scales characterizing the information dynamics and the intrinsic network dynamics. This statement in practice means that the temporality of the network has a non trivial impact on the spreading of information on top of it only if the time scale of the changing of network's configurations τ_{NET} is comparable with the time scale at which the information pervades a fraction of the system τ_{INFO}. When this happens, it is clear that the information, in our case is the disease, can "jump" and propagates from a node i to another j via a certain link i-j of the network only when the link is active (i,j;t) , therefore the disease can be slowed down by the necessary waiting time in order to see i-j active. This means that the two dynamics are strongly bounded and their interplay cannot be discarded. There are two limiting cases for which an approximation of the temporal network as a static entity can be still very good and informative. With analogue nomenclature as the one seen in Section 3.3, these limits are called Quenched and Annealed.

The first limit, the Quenched, takes place when the time scale of the network evolution is quite slow with respect to τ_{INFO}. The information spreading process thus do not sense the changing of the network topology, and therefore a single static network represented by a single adjacency matrix A_{Quen} provides a good approximation. A_{Quen} is a

fixed network that can be one of the original temporal sequence $\{A(t)\}$:

$$\tau_{INFO} \ll \tau_{NET} \qquad \underset{\forall t}{A(t)} \to A_{Quen} \tag{4.9}$$

where and $\to$ means "can be substitute with". The opposite limit defines the so-called Annealed network, and describes the case in which the evolution of the network is much faster than the dynamical processes. In this limit, the dynamical process unfolds on a network that is rapidly rewiring so that the dynamics effectively occurs on a time averaged network. An annealed network is thus described by a temporal average of the sequence of adjacency matrices:

$$\tau_{INFO} \gg \tau_{NET} \qquad \underset{\forall t}{A(t)} \to A_{Ann} = \sum_{t=1}^{T} \frac{A(t)}{T} \tag{4.10}$$

The above two cases are relevant to mark the limits of applicability of the most commonly used theoretical approaches. For instance, the ones described in Sec. (3.3). For the quenched limit the spectral theory can be used to asses for the epidemic threshold. For the annealed one can either use spectral theory on A_{Ann} or interpret each snapshot as belonging to a certain statistical ensemble ,e.g. the one defined by a degree distribution $P(k)$, if such ensemble can be meaningfully defined.
There are, however, several other instances of networks, notably in communication and in social systems, where the connectivity pattern varies over time scales comparable to those of the information processes. For those it is crucial to explicitly take into account the concurrent dynamics of the spreading process and the connectivity pattern. It is worth noticing that here we have not explicitly defined τ_{INFO} and τ_{NET}. Actually, τ_{NET} can be taken as the time scale intrinsically defined by the time-step of the sampling of the data describing the temporal network. For a daily varying network the time scale is one day, and so on. What is more challenging to define is τ_{INFO}. In fact, the intrinsic time-scale of the information, such as the infectious period $\tau_u = \mu^{-1}$ for a SIS disease, a priori do not say much on the time needed by the disease to spread, as that is highly dependent on the network dynamics. The time scale of a SIS in an homogeneous mixing approximation for the contact pattern is $\tau = |\lambda - \mu|^{-1}$, and we can think of this as the lower bound (faster evolution) of the disease. However it is clear that this is far from accurate. At a node level, we can think that a fast disease should at least transmit before the recovery, that is $\lambda/(1-\mu) \gg 1$. However, again this does not necessary mean the disease spreads fast all over the network. Therefore, given the rates λ and μ of a SIS and given a temporal network, a priori there is no safe way to asses if the dynamics can be brought back to one of the above mentioned limits. The only way to study a network that varies in time is to treat it in the most general way, i.e. considering all its temporal features. Only the result of microscopic infection simulations will reasonably inform on whether the system has fallen into one static limit or not.

4.5 Epidemic threshold on temporal networks: the method of the Infection Propagator [28]

Many contact structures relevant for the spreading of diseases do evolve in time. Human face-to-face interactions, responsible for the spread of airborne diseases, or sexual interactions, relevant for the diffusion of STDs, all result in highly dynamic contact networks.

Livestock transport networks, which are responsible for the spread of many animal diseases, exhibit a highly dynamic nature, too. The interaction between network evolution and disease diffusion is known to influence the dynamics and outcome of the epidemic process. Specifically, such interaction alters the epidemic threshold, affecting the conditions that lead to the wide-spreading regime. For this reason, being able to compute the epidemic threshold accounting for network dynamics is a crucial step towards assessing the vulnerability of the system. In the past years, researchers have analytically computed the epidemic threshold in the two limiting scenarios described in Section 4.4 : the quenched and annealed limits. These are accurate in the regime of *time scale separation*, i.e. when network evolution and disease diffusion occur at very different paces, with the network evolving much slower than disease (quenched) or much faster (annealed). When the two time scales are comparable, and thus fully coupled, researchers have so far resorted to numerical simulations, or provided analytical calculations only for specific cases. We list here few of the large number of studies in this new born field. One of the first model has been defined for epidemic spreading on a rewiring (links undergo a sort of swapping in their tips) network, highlighting that the rewiring process of links tends to suppress the infection. Similarly, another model has quantified the impact of the contact duration on the pathogen's diffusion. More recently, an activity driven model, in which the instantaneous interaction of agents is defined by a so called activity potential, has provided an analytical description of the epidemic threshold in such context [45]. Another model has shown the slowing down effect of a time network with weights of links assigned heterogeneously; this going in the opposite direction of the known acceleration of spreading in a static strength- heterogeneous network [47]. All these and other models have shown a non linear interplay of the disease and network dynamics.

In this section we exploit the *intra-layer representation* of the multilayer mapping , described in Section 4.1, in order to analytically compute the epidemic threshold of a generic temporal network. We will use this interesting and novel result, found in [28], in order to rapidly and accurately assess for the vulnerability of data-driven temporal network. This model in fact allows to numerically evaluate the epidemic threshold of a temporal network by-passing the usual time-demanding microscopic numerical simulation, while obtaining an exact result as nearly no hypothesis are made on the relevant features of the network dynamics and all contacts coming from data are included.

4.5.1 The SIS supra-adjacency matrix of a temporal network [28]

Let us recall for a moment the multi-layer defined in Section 4.1. We started from a time ordered sequence on adjacency matrices and defined a corresponding tensor of rank 4, thus rephrasing all contact informations into a one static object:

$$\{A^{(1)}, \dots, A^{(T)}\} \to \mathbf{A}_{ij}^{tt'} = \delta^{t+1,t'} \left[\delta_{ij} + A_{ij}^{(t)}\right]. \tag{4.11}$$

The ratio behind this choice of layer representation lies in the spreading process that we wish to couple to it. Let us consider a SIS spreading dynamics, with transmission probability λ and recovery probability μ. Suppose that node i is infectious at time t. If nodes i and j are connected at time t, there is a chance (λ) that i will infect j, resulting in an infectious j at time $t+1$. This explains why link i-j at time t is represented through non-diagonal links going from layer t to $t+1$. Moreover, going from t to $t+1$, node i may either recover or remain infectious. We interpret the probability that i does not recover, as the probability $1-\mu$ it transmits the disease to its future image at $t+1$. We implement that through the diagonal inter-layer links. In addition we stress that the

absence of undirected links and intra-layer links in the multilayer structure guarantees that causality in temporal evolution is preserved. The mapping we have defined thus naturally couples network dynamics to disease diffusion, which appears by tuning the intensity of the different types of links. Non-diagonal links will have a weight λ as they transmit the pathogen among different nodes, while diagonal links will be weighted by $1-\mu$, encoding "missed" recovery dynamics.

The *SIS tensor* coupled to the network dynamics $\mathbf{A}_{ij}^{tt'}$ therefore results:

$$\mathbf{M}_{ij}^{tt'} = \delta^{t+1,t'}\left[\delta_{ij}(1-\mu)+\lambda A_{ij}^{(t)}\right]. \tag{4.12}$$

Dealing with tensors is not very convenient, while matrices would be preferable. Instead of identifying each node of the multilayer with (i,t), we can consider the whole network as composed of NT distinct nodes, flattening out the tensor structure. Nodes are now identified by index $\alpha = 1,\dots,NT$, with the one-to-one mapping:

$$(i,t) \to \alpha = N(t-1)+i \tag{4.13}$$

where $i = 1,2,\dots,N$ and $t = 1,\dots T$. Algebraically, we are exploiting the fact that the tensor space $\mathbb{R}^N \otimes \mathbb{R}^T$ in which the multilayer node lives, is isomorphic to $\mathbb{R}^{NT}$. Thanks to this, instead of dealing with a tensor, we can now switch to the supra-adjacency matrix formalism, by writing the adjacency matrix of the corresponding NT graph. Such matrix will be of size $NT \times NT$, and composed of T^2 blocks of size $N \times N$, representing the original layers. In addition, given that we already have added disease dynamics with Eq.(4.12), we weigh links accordingly. The resulting matrix is the *SIS supra-adjacency matrix*:

$$M = \begin{bmatrix} 0 & 1-\mu+\lambda A^{(1)} & 0 & .. & 0 \\ 0 & 0 & 1-\mu+\lambda A^{(2)} & .. & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & .. & 1-\mu+\lambda A^{(T-1)} \\ 1-\mu+\lambda A^{(T)} & 0 & 0 & .. & 0 \end{bmatrix} \tag{4.14}$$

where each entrance is a $N \times N$ block matrix and $A^{(t)}$ is the adjacency matrix of time t. This matrix M will allow to compute the epidemic threshold of the dynamic network.

4.5.2 The Infection-Propagator matrix and the threshold

From Section (3.3.2) we recall that the SIS propagation on a generic network with N nodes and adjacency matrix A is given by

$$p_i(t+1) = 1-[1-(1-\mu)p_i(t)]\prod_j\left[1-\lambda A_{ji}p_j(t)\right]. \tag{4.15}$$

where $p_i(t)$ is the probability for the node i to be in the infectious state at time t. This model is Markovian and moreover is based on the mean-field assumption of the absence of dynamical correlations among the states of neighbouring nodes. In a temporal network, contacts are changing in time therefore one has $A(t)$ and it is easy to see that the evolution of the vector state of the infection at time t reads:

$$p_i(t+1) = 1-[1-(1-\mu)p_i(t)]\prod_j\left[1-\lambda A_{ji}(t)p_j(t)\right]. \tag{4.16}$$

where we have simply let the adjacency matrix depend on time. In order to ensure the asymptotic solution of the SIS process in a generic temporal network, we assume periodic boundary conditions for the network dynamics. With T being the total number of network time snapshots, we impose $A(T+1) = A(1)$. This does not imply any loss of generality given that T may be completely arbitrary. We notice that, as a consequence of the assumed periodic temporal dynamics of $A(t)$, the asymptotic solution of Eq. (4.16) is, in principle, periodic of period T.

We are now ready to use the supra adjacency matrix of Eq.(4.12) to re-write the Markov Equation (4.16). We can describe the Markov process of a period T via a trajectory in $\mathbb{R}^{NT}$ where the state vector $\hat{p}_\alpha(\tau)$, with $\alpha = N(t-1)+i$, represents the probability of each node i to be infected at each time step t included in the interval $[\tau, \tau+T]$. In this way Eq. (4.16) can be redrafted as:

$$\hat{p}_\alpha(\tau) = 1 - \prod_\beta \left[1 - M_{\beta\alpha}\hat{p}_\beta(\tau-1)\right]. \tag{4.17}$$

From this result is clear that M provides a network representation of the topological and temporal dimensions underlying the dynamics of Eq. (4.16), here interrelated and flattened. Its directed nature preserves the causality of the process, while its weights account for the SIS transition probabilities.

This equation encodes a one-period configuration, so we can find the T-periodic asymptotic state of the SIS process by solving the steady state equation

$$\hat{p}_\alpha = 1 - \prod_\beta \left[1 - M_{\beta\alpha}\hat{p}_\beta\right]. \tag{4.18}$$

As we are interested in the stability of the disease-free state, we can linearize Eq. (4.18) for small values of $\hat{p}$ and we find:

$$\hat{p}_\alpha = \sum_\beta M^\dagger_{\alpha\beta}\hat{p}_\beta. \tag{4.19}$$

The necessary and sufficient condition for the asymptotically stable zero solution is therefore $\rho(M^\dagger) < 1$, where ρ stands for the maximum absolute value of the M-eigenvalues. Considering that the uniform zero solution in the $\mathbb{R}^{NT}$ representation is mapped to a uniform zero solution in the original $\mathbb{R}^N$ representation, the threshold condition on the temporal network is:

$$\boxed{\rho\left[M^\dagger(\lambda, \mu, \{A_t\})\right] = 1} \tag{4.20}$$

from which the critical values of λ or μ are found. Eq.(4.20) is defined the *epidemic threshold condition* for the temporal sequence of snapshots $\{A_1, \ldots, A_T\}$ on which a SIS disease dynamics unfolds.

We notice that the spectral radius of M can be simplified with the following relation[2]:

$$\rho(M^\dagger) = \rho(P)^{1/T} \tag{4.21}$$

where P is a matrix called the *infection propagator* [28]:

$$\boxed{P = \prod_{t=1}^{T} \left[1 - \mu + \lambda A(T-t)\right].} \tag{4.22}$$

[2]See Appendix C.

It is worth noticing that P is nothing more than a weighted version of the accessibility matrix in Eq.(4.3), with the same meaning. P_{ij} is the number of paths that a disease can take to go from infectious node j at time 1 to node i at time T, each path weighted with its $(1-\mu)^n\lambda^{T-n}$, for some $n \in [0, T]$ depending on the path. Last Eq.(4.22) is particularly notable as it ensures a simplification of the numerical computation of the epidemic threshold, allowing an execution time scaling as $\sim TN^{5/2}$. The full analysis of the numerical performance of this approach can be found in [29]. In conclusion, given a recovery probability μ, the threshold is the value of the transmission parameter λ for which the spectral radius of P is equal to 1 :

$$\rho\left[P(\lambda_c, \mu)\right]^{1/T} = 1. \tag{4.23}$$

Conversely, if we fix λ we can find the critical recovery probability μ_c evaluating the same spectral radius.

The quenched and annealed regimes can be recovered within this general framework as particular limiting solutions. In the first case, it is to be noted that the sequence of temporal snapshots naturally defines the minimum time scale of the process. In order to consider contagion dynamics that are much faster than the time-varying process of the network, we thus rely on the commonly adopted assumption regarding the temporal network as static, so $A(t) \equiv A$. In this particular case, $P = (1-\mu+\lambda A)^T$. Therefore, $\rho(M) = \rho\,(1-\mu+\lambda A) = 1-\mu+\lambda\rho(A)$. The requirement $\rho(M) = 1$ thus recovers the expression known for the quenched case:

$$\left(\frac{\lambda}{\mu}\right)_c = \frac{1}{\rho(A)} \tag{4.24}$$

The study of the annealed regime is less trivial. In the assumption that λ and μ are very small, corresponding to a very slow disease dynamics with respect to the time scale of the network evolution (that here is a time step 1), it is possible to replace P with its linear expansion in $\lambda/(1-\mu)$, yielding:

$$P_{slow} = (1-\mu)^T\left[1 + \frac{\lambda}{1-\mu}A^{agg}\right] \tag{4.25}$$

where $A^{agg} = \sum_{t=1}^{T} A(t)$ is the weighted HET aggregation of the temporal network, a static network given by the sum of all snapshots. Temporal correlations are lost, and edges count for the number of times they are active during the whole period T. Equation (4.21) for the epidemic threshold thus simplifies to

$$\left(\frac{\lambda}{\mu}\right)_c = \frac{T}{\rho(A^{agg})} \tag{4.26}$$

and the aggregated matrix contains all the relevant information for spreading dynamics.

4.5.3 Weighted matrices

If the contacts are weighted $A_{ij}(t) \geqslant 1$, then the infection propagator matrix has to be defined as

$$P = \prod_{t=1}^{T}\left[1-\mu+\Lambda(T-t)\right]. \tag{4.27}$$

with

$$\Lambda_{ij}(t) = 1-(1-\lambda)^{A_{ij}(t)} \tag{4.28}$$

that gives the probability that at least one of the infected contacts transmits the disease.

Chapter 5

Study of phase transition on a data-driven temporal network

After an introduction to the concepts and methods involved in the study of an epidemics unfolding on a temporal network, we are ready to apply this knowledge to a real scenario. We dispose of a data set regarding a human contacts evolving in time. We would like to assess for the epidemic threshold of this data-driven temporal network, exploiting both numerical and theoretical analysis. The aim is to understand which are the driving features of this temporal network with respect to a disease unfolding on it, and to possibly implement them in a theoretical model that is able to reproduce the real trend of the epidemic threshold in the space of the SIS parameters (λ, μ).

The chapter is divided into three sections. In the first section, we introduce the data-set and describe the epidemiological relevance of investigating this particular community. In the second section, we characterize the topological and temporal features of the data-driven time-evolving network, exploiting the statistical and mathematical tools described in Chapter 4. In the third and last section, we perform a numerical computation of the epidemic threshold, for a SIS disease dynamics. We use the Infection Propagator approach of Section 4.5, that allows to explore a wide range of SIS parameters' values with low computational cost. Moreover, we exploit the bipartite nature of the population described by the data-set to enlarge the space of parameter from two (λ, μ) to three $(\lambda_A, \lambda_B, \mu)$, in order to assess for the different spreading potential of each group A and B of nodes.

5.1 The empirical network and its relevance in the spreading of ST diseases

The data set we have available is a daily contact list of encounters between female sex-sellers and male sex-buyers that are part of the escort business in Brazil. This real community has a Web counterpart based on a forum-like internet page that collects the rates that sex-buyers attributes to sex-sellers. The forum posts, grouped by day, can be translated into a daily temporal network of contacts. The analysis of the forum and the extraction of the daily resolved temporal network has been carried out in 2010 by Rocha, Liljeros and Holme and published in *Information dynamics shape the sexual networks of Internet-mediated prostitution*[34]. In the following years, the same data have been exploited in many others epidemiological and network science works, being this network relevant both for its time resolution and for the relevance of such community in the epi-

demiological studies on sexually transmitted diseases.

From the epidemiological point of view, this data set is of major interest because it allows for an insight in a community that is hard to detect via the usual survey channels. This is a type of business that can have a strong impact on public health, concerning particularly the containment of sexually transmitted diseases (STD) inside and outside the community. In fact, prostitution have long been considered to be reservoir, if not 'vector', for the transmission of STD. Survey records have confirmed this guess, and revealed a majority of sex sellers had active STD infections, with Gonorrhea and Chlamydia being the most common bacteria. For others disease, like HIV, the situation is different, but nevertheless sex workers often share common factors that make them vulnerable to transmission. Notably, these are: multiple partners, inconsistent condom use, social and legal factors that make sex workers more exposed to rapes and abuses; in addition in some communities they are more prone to injecting drug use and lastly they are often under educated and not aware of the risk connected to unsafe sex. As a result, in low and middle-income countries, HIV prevalence among sex workers is estimated at 12 %, even if there are significant variations between regions and countries [36].The clients of sex workers also act as a 'bridge population', transmitting HIV between sex workers and the general population. High HIV prevalence among the male clients of sex workers has been detected in studies globally [37] [38] [39]. All this considered, being able to study the dynamics of a real network of contacts in the prostitution business is strategic to public health policies that aims to prevent a potential spread of STD in the prostitution community, and consequently in the rest of the civil population outside the business.

Our interest in exploiting this temporal network is twofold. On one hand, it allows to explore a factual non trivial time-evolving contact dynamics, belonging to a human social context. On the other hand, it challenges to assess the vulnerability of such network to an hypothetical ST disease targeting the system.

5.2 Temporal and topological features of the network

The community we study is a Brazilian public on-line forum with free registration, that is financed by advertisements. In this community, male members grade and categorize their sexual encounters with female escorts, both using anonymous nicknames. The visible information of the forum contains anonymous users' nicknames of sex sellers and buyers, in which city the activity occurred, and the time of posting (that is taken as a rough estimate of the actual day-time of the real encounter). From these information, Rocha *et al* extracted a daily time evolving web of contacts. For the interest our analysis, we will focus only on the time evolution of the contacts:

$$(i, j; t) \tag{5.1}$$

with $i = 1, .., N_A$ and $j = 1, .., N_B$ and $t = 1, ...T$. Index i labels the sex buyer (always man in our system) and index j labels the sex seller partner (woman) in day t. Given the list of time-stamped contacts (5.1), we are able to built the networks of contacts for each day t. All the informations relative to the contact dynamics are so encoded in a sequence of time ordered networks. For matter of brevity, in the following, we will refer to this data-driven system as the data-set, while sex-sellers will be addressed as women and sex-buyers as men. We will rephrase this community an its dynamics in the mathematical formalism of a time evolving network of contacts, where people are nodes and their contacts are links.

Here we report some over all statistics of the network, once considered the total period of time covered by the data-set.

- Number of members $N = 16\,730$
- Number of men $N_A = 10\,106$
- Number of women $N_B = 6\,624$
- Period of observation $T = 2\,232$ (a bit more than six years)
- Number of contacts in the total period $N_c = 50\,185$
- Number of distinct couples $N_E = 40\,895$

5.2.1 Growth of the Community

The first characteristic to emphasise about the network is the fact it is **bipartite**. Women do not have contacts with other women, and the same is true for men. This results in a bipartite structure of contacts, where links exist only between members of different gender groups. Our data-set covers the time evolution of a community from its beginning (at $t_{in} = 0$ days) up to a stable and vast active state nearly 6 years after (at $t_{final} = 2\,232$ days).

In Figure (5.1) we can see the growth of the number of nodes in time. The plot shows the number of nodes active up to a certain day, separating female nodes from male[1].The x-axis covers from $t = 0$ days, that corresponds also to the beginning of the Website activity, up to the end of the available data, at $t_{final} = 2\,232$ days. The y-value of each point in the line is the number of vertices that have been active at least once up to the day correspondent to the x-value. In other words, the plot shows the cumulated number of active vertices *up to* day t. The magenta line corresponds to women and the light-blue to men. From the figure we notice that the first months see very few nodes active; then, after around day 500, a marked growth begins and it never stops up to the final day of the recorded data. At the end, the over all number of nodes that have been active at least once is $N_{nodes} \sim 10^4$. The slopes of the lines in Fig.(5.1) encode the amount of new people entering the system per day. After an initial period of about 1000 d, the slopes stabilize, meaning that membership grows at a fairly constant rate of about 6 new men and 5 new women per day. Therefore, the community constantly enlarges with new members entering the system around 11 each day.

The cumulated number of active nodes does not inform on the daily activity of the system. In Figure (5.2) we therefore show the per-gender activity time line. This quantity is defined as the number of nodes that are involved in at least one link in the day. It encodes for the daily load of business i.e. not cumulated but instantaneous amount of active people. The nearly twin-curves for men and women in Fig.(5.2) reproduce the growth of the system. In addition, they remarkably show a tendency towards a stabilization of the number of active people for $t > 1600$d. In last 300 days, we can see that the curves show a little fall followed by a new increase, probably due to "fashion" cycles in the web activity, or to seasonal behaviour of clients, like holidays. The final stable situation tells us that, despite of the constant enlarging of the network, the system is finally approaching a situation of saturation in the daily activity: the stable and sustainable business involves approximately 35 to 40 vertices of each kind, with a little higher

[1]The plot is a running mean of the data. This operation makes the otherwise noisy graphics readable.

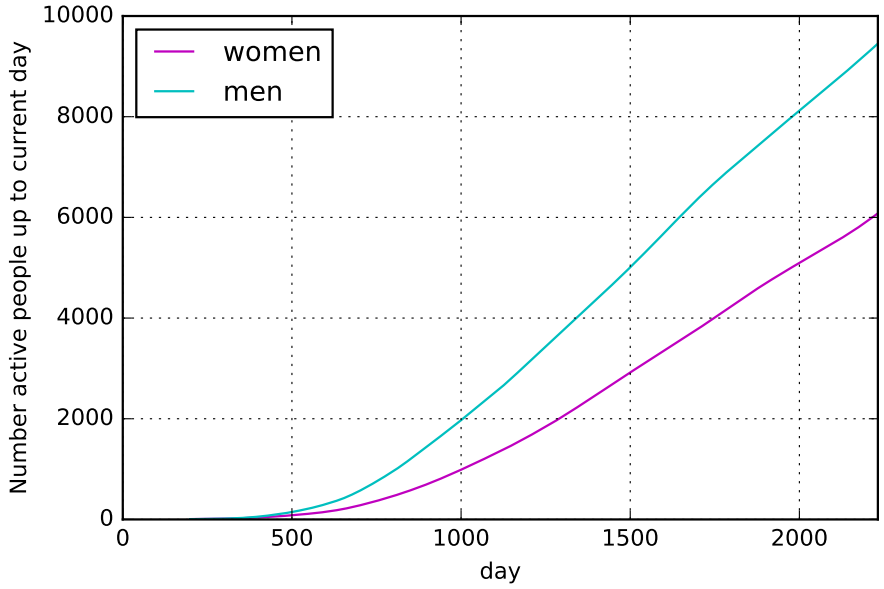

Figure 5.1: **Growth of the size of the network**

value for women. Remarkably, the average number of women being higher than men's shows an unexpected feature: clients performs more contacts than sellers when we look at them on a daily time window. In more details, comparing the active people with the cumulated in the same day we see that the ratio $N_{active}(t)/N_{cumul.}(t)$ is around 0.4% for men and 0.6% for women. In conclusion, the two plots tell us that the nodes do not remain active continuously after their first appearing in the system, but they have long periods of inactivity. Otherwise, we could not witness the relatively stable final state of the daily active people, that is moreover well below the cumulated activity. This behavioural feature is usually named turnover; and as one could expect, the turnover is higher for men while women tend to stay longer in the system.

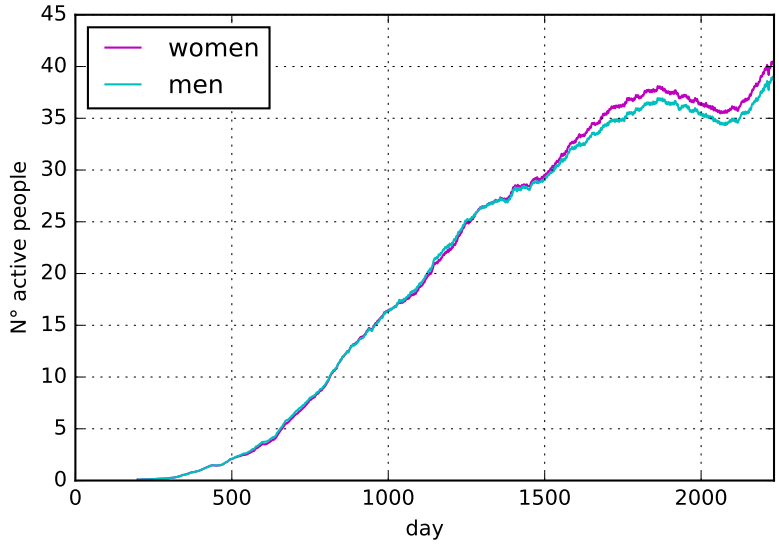

Figure 5.2: **Activity time line**.

In order to better quantify the turnover of individuals, we evaluate how many repeat their presence in the system, for instance in two consecutive weeks. This gives an idea of the variability of the population with weekly resolution; and in order to have an idea of the magnitude of the phenomenon it must be contrasted with the activity time line (aggregated in the same interval). Figure (5.3) shows that the permanence of people in

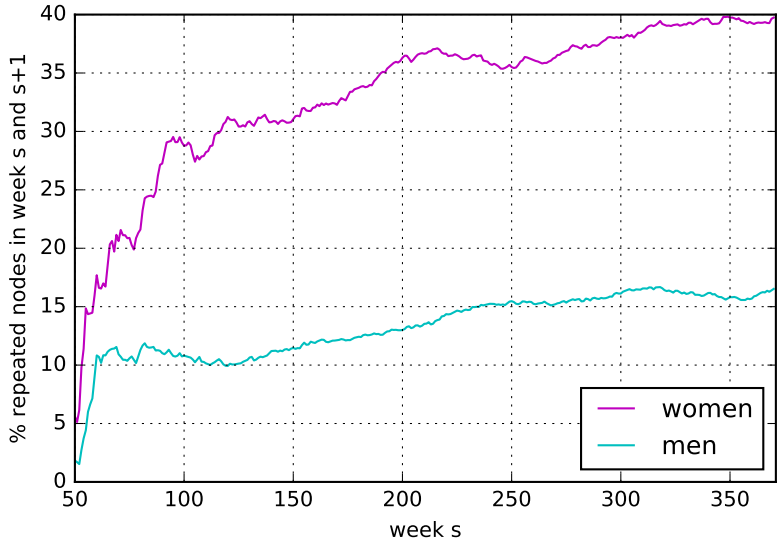

Figure 5.3: **Individuals that repeat their presence in consecutive weeks.** The y axis is the relative number of repeated people, normalized on the weekly activity time-line.

the system in very stable across the whole time length of the dataset (after discarding the first 100 weeks i.e. 600 days). About 40 % of women stay in the system at least two weeks, while for men the average fraction is 15 %. This difference is expected as women are the sellers in the business while men are the buyers, therefore a stronger permanence in the business is expected for sellers, at least in the short time. Complementary, we can say that the turnover is of 60% for women and 85 % for me.

Considering a smaller time interval (e.g. 2 days) leads to the trivial result than nearly no node repeat its presence consecutively: this is easily explainable once we recall this is an escort business so it is unlikely that individuals (especially men) repeat the activity *within* one-week time or less (figure not shown).

Finally, we can conclude that the turnover of active nodes is high and stable, confirming what we could already guess by comparing Figs (5.1-5.2). What we find in addition is that the daily turnover is nearly 90 % (figure not shown), while on a weekly scale it is still high but shows a marked difference for men and women. We can conclude that, despite the over-all continuous growth of the system, the average short term behaviour is quite stable, especially in the last 700 days.

5.2.2 Edges preservation and Loyalty

Up to now, we have studied the permanence of a group of people in the system for a period of time. It is clear that this may not mean that the same contacts are iterated over that period. Actually, besides the individuals' activity, what can be more epidemiologically relevant is the temporal dynamics of the links. Iteration of contacts may be a key feature for the spreading of a disease with a little viral load (like HIV), that needs the same contacts to happen many times in order for the disease to have a chance to spread.

We therefore inspect the preservation of edges i.e. when a couple is in contact for subsequent intervals of time, for intervals of width of one month $\Delta = 28$ days and semester $\Delta = 168$ days. Namely, we evaluate the fraction of preserved couples:

$$\Phi_{\Delta}^{t,t+1} = \frac{|E^t \cap E^{t+1}|}{|E^t \cup E^{t+1}|} \tag{5.2}$$

with E^t is the set of edges active in the t-th time window of chosen width Δ, and where $||$ is the cardinality of the set. Φ can take values between 0 and 1. Low values of Φ mean that nearly all couples are different with respect to the past interval, high values means many couples are retained. In Figures (5.4) and (5.5) we plot the fraction of common edges Φ expressed in %, for pairs of consecutive months and semesters respectively. Observing the monthly trend, we can see this is quite variable, but after 40th month it ranges always between 2.5 and 4.5 %. A similar trend emerges also in the semester plot, with Φ increasing and eventually becoming very stable in last 3 semesters, around the 4%. This mean that the fraction of people that tend to choose the same partner is fairly constant, and is of course low, as we can understand recalling the inherent high turnover of nodes. Concluding, around 95 % of the edges are newly formed with respect to the previous time interval.

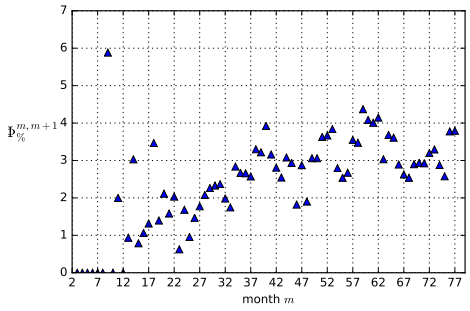

Figure 5.4: **Monthly common edges**

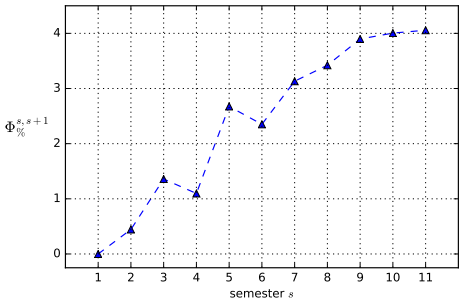

Figure 5.5: **Semester common edges.**

We can now return to an "egocentric" point of view and consider the behaviour of a node with respect to the choice of its contacts. Again, we can distinguish between the behaviour of men and women. We inspect the distribution probability of the the so-called loyalty parameter [33]. This variable measures the fraction of preserved neighbours of a node, for a pair of two consecutive network configurations in time. As before, the network can come from a time-aggregation of width Δ. Let us consider the t-th and t+1 time-windows, and let $\succeq_i^t$ be the set of neighbours of node i in window t. Then loyalty $\Theta_i^{t,t+1}$ of node i from time t to time (t+1) is defined as the Jaccard index between ν_i^t and ν_i^{t+1}:

$$\Theta_i^{t,t+1} = \frac{|\nu_i^t \cap \nu_i^{t+1}|}{|\nu_i^t \cup \nu_i^{t+1}|} \tag{5.3}$$

As Φ, loyalty takes values in $[0, 1]$. $\Theta = 0$ means no neighbours are retained, while $\Theta = 1$ means that the node keeps exactly the same neighbours from t to t+1, with no addition. Once a couple of time steps is fixed, loyalty has to be computed for each node, therefore it is convenient to show the results using the probability distribution $P(\Theta)$. This time, we consider a width of half semester $\Delta = 84$ days, so that we can inspect the loyalty 'inside' a semester. The results are shown in Figure (5.6), for three semesters in the last part of the dataset. We see that the three plots are quite similar, pointing out a certain

stability in the loyalty behaviour across time. Focusing, for instance, on the right-most plot we see that women tend to have low loyalty: they change often partner. Men, on the other hand, have a more heterogeneous behaviour. Most of them have an exploratory behaviour (Θ small) but there is a non negligible fraction of very loyal men that repeat more than 50 % (also 100 %) of their contacts. Notably, we have checked that all people repeat their presence at least once in the the chosen windows.

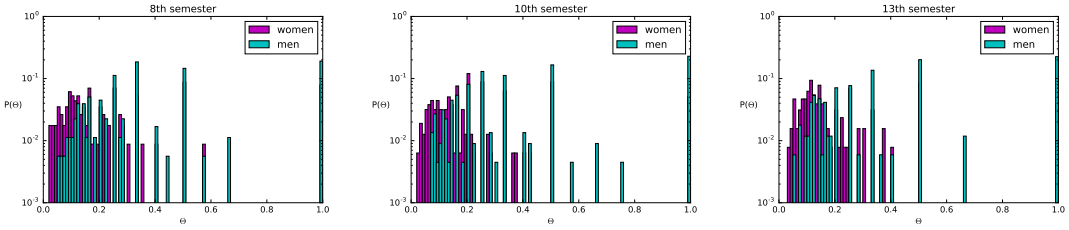

Figure 5.6: **Loyalty probability distribution** for 8-th,10-th and 13-th semester.

5.2.3 Degree distribution

We can now focus on the most common indicator for the connectedness of a static network: the degree. For a static network, each node i has a fixed degree k_i that counts the number of different links that have a tip in i. The mean degree of the network is therefore $\langle k \rangle = \sum_i k_i / \sum_i 1$. For a temporal network, the degree of node i can be defined on each snapshot t, $k_i(t)$; or for the aggregated network of Δ time-steps starting from snapshot t_0, $k_i{}^{(t_0,\Delta)}$. The mean degree is the average on the different nodes $\langle k \rangle^{(t_0,\Delta)}$.

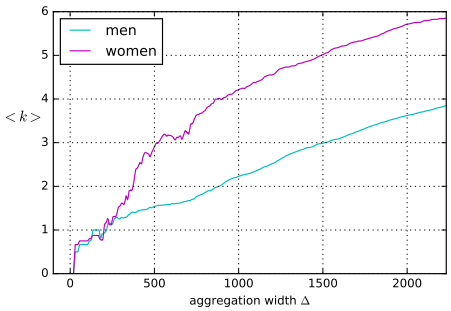

Figure 5.7

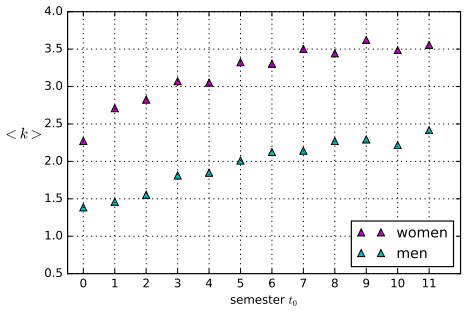

Figure 5.8

5.7: **Mean aggregated degree for growing aggregation time-window**. Normalization is on the cumulated active population. 5.8: **Mean degree per semester**, normalized on the current active population.

In Figure(5.7) we show the average degree $\langle k \rangle^{(0,\Delta)}$ for increasing time windows $\Delta = 1, 2, .., 2232$ days of aggregation. The average degree grows for both men and women. This means that our sampling time is shorter than the typical time that a man or woman has an active presence in the system. Or that the new entries are more and more active. The curve for the women, however, shows an incipient tendency toward saturation. This incipient saturation suggests that the timescale of a woman's activity life-time is not much longer than the sampling period, and also that men stay in the system longer than women. For $t < 250$ d, the average degree is approximately the same for both categories of nodes.

We can now fix the time width of aggregation $\Delta = 168$ days, i.e. a semester, and change the initial day t_0 shifting of +168 days. The semester-wide mean degree $\langle k \rangle^{(t_0, 168d)}$ is shown in Figure (5.8). After an initial transient, mostly caused by the enlarging of the system, the mean degree in a semester becomes very stable; the values are around 2.5 for men and 3.5 for women. These values are quite low, compared to the number of active men/women. This confirms that the network is very sparse, i.e. among the possible links only few are active. Moreover, we notice there is no big difference between the average degree of men and women, pointing out this escort business is actually of a different kind with respect to what one would expect from prostitution business.

The mean degree cannot inform us on the actual distribution of the degree for the single nodes, that could be a priori very widely distributed. We focus on one semester (the last, starting at $t_0 = 2064$d), and we consider the probability distribution $P(k)$ of the aggregated degrees $k_i(t_0 = 2064, \Delta = 168)$. We plot the distribution in Figure (5.9) in double log10 scale. Firstly, we see that men and women have different individual behaviour. In particular, the maximum cumulated degree for women is around 85, while for men is much lower, around 20. For both categories, the distribution is peaked around k=1 and decays for higher degrees. It is usual to compare the degree distribution to the power-law shape $P(k) \sim k^{-\gamma}$, typical if the free-scale network. In this case, we have too small values of k to go further in this comparison. To our interest, it is enough to notice the wide and heterogeneous distribution of the degrees. This allows to see that there are some nodes that can act potentially as major spreaders of disease, and that the contact pattern is far from homogeneous.

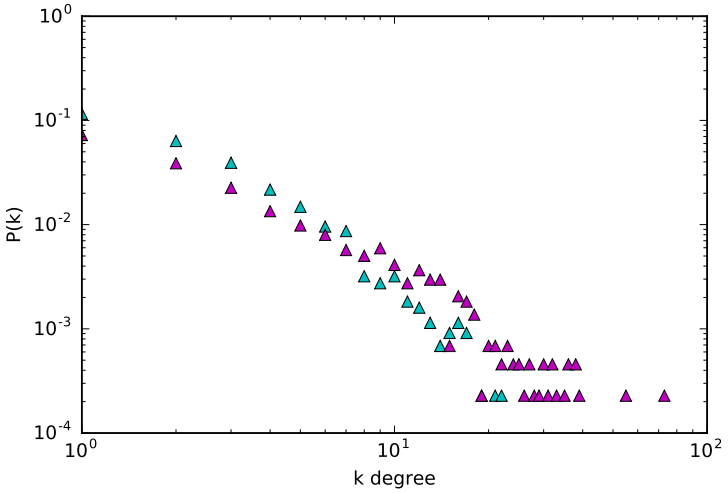

Figure 5.9: **Degree distribution of last semester.**

5.2.4 Assortativity

We finally would like to see if the aggregated network of last semester, for instance, shows the property of assortativity/disassortativity. These properties refer to the fact that nodes with large degree may preferentially choose to make links with nodes of large degree (assortative) or of small degree (disassortative). It is interesting to reveal this network feature because of its renowned non negligible effect on the spreading of information on top of the network. We thus compute function $k_{nn}(k)$ on the network

corresponding to the aggregation of last 168 days. $k_{nn}(k)$ is defined as the mean degree of the nearest neighbours of a node having itself degree k. If the network is neither assortative nor disassortative, $k_{nn}(k)$ would be constant in k. In particular, if we write $P(k|k') = kP(k)/\langle k\rangle$ we find $k_{nn} = \langle k^2\rangle/\langle k\rangle$. On the other hand, assortativity would mean k_{nn} decreases with increasing k; disassortative would show an opposite tendency. We plot our results in Figure (5.10) , where we have grouped together values of $k_{nn}(k)$ for k in a range of width 5 and separated men from women. The boxes show with a dark blue/magenta line the value the median of each group, the top and bottom sides of the box are the 25 and 75 percentiles and the vertical dashed bars cover from the 5 to 95 percentiles. This type of plot gives us an idea of the distribution of k_{nn} inside each k-group. We see that men make contacts with women of average neighbourhood ranging from 1 to 35. The median is quite stable, thus not showing preferential attachments to higher or lower degrees. Women behave in a random way as well, with a very stable value of media across their degree. The green and orange lines mark the uncorrelated esteem $k_{nn} = \langle k^2\rangle/\langle k\rangle$, and we see that this approximation is quite with our data. We do not see any marked trend, so no assortativity nor disassortativity is present in here.

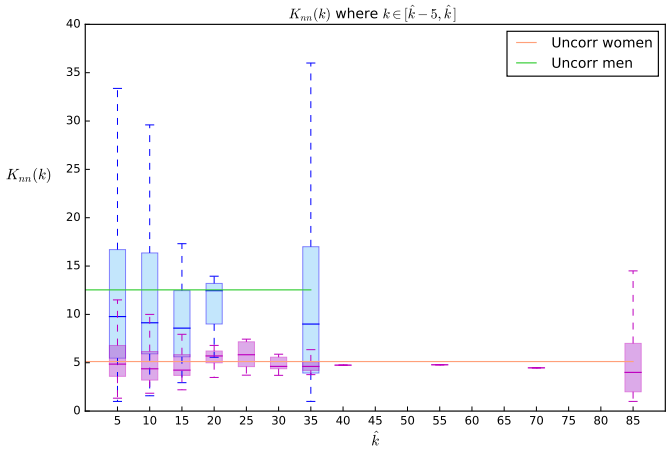

Figure 5.10: **Mean degree of nearest neighbours**. To make the graph more readable, we group $K_{nn}(k)$ by the value of k in a range of width 5. The boxes show in dark line the value the median of each group, the top and bottom sides of the box are the 25 and 75 percentiles and the vertical dashed bars cover from the 5 to 95 percentiles.

5.3 Epidemic threshold : a numerical analysis

It is now time to asses for the threshold of a SIS disease dynamics unfolding on our temporal network. The usual strategy in order to evaluate $\lambda_c(\mu)$, or equivalently $\mu_c(\lambda)$, consists in performing many microscopic simulations of the SIS dynamics on the network. This is done by simulating an infection process starting form a randomly infected node (the seed) in a completely susceptible population and iterate at each time step the possibility that the infectious nodes transmit to their neighbours and then possibly recover. At each time step, the number of infected nodes is evaluated and the simulation is arrested when that value becomes stable, i.e. the "stationary" state is reached. The final fraction of infected nodes i_{stat} for a given couple of parameters (λ, μ) comes from an average over the many different-seeds simulations. For a temporal network, the only difference is that at each time step the set of available links must be updated as they change in time.

In order to evaluate the epidemic threshold, for example with respect to rate λ with fixed μ , one needs to repeat the above described protocol for various values of λ. For each λ, a final i_∞ is evaluated and the threshold is defined as the value λ_c such that for $\lambda \geqslant \lambda_c$ the final infection becomes non zero: $i_\infty > 0$. Given the stochasticity of the simulation, when the transmission rate approaches the critical value there is an high chance that all infection processes "die out" even if above threshold. Therefore, the number of simulations to effectively evaluate i_∞ has to be very high; otherwise some side-way strategy, like the Quasi Stationary method [54], has to be used.

It is clear that the threshold computation via microscopic simulation is very laborious and time demanding. For this reason, we will exploit the method of the Infection Propagator (IP) for temporal networks. We have extensively proven and described it in Section 4.5 and here we exploit its rapidity in the estimation of the threshold in order to explore a wide range of the SIS parameters, overcoming the otherwise burdening computational cost. We will not prove herein the accuracy of the threshold esteemed via the IP, but we refer to [28] for an extensive application of the method to different temporal network data-set (among which notably we find ours) , contrasted with microscopic simulation. The method has been proved to be very robust and therefore we will use its output results as the best esteem of the "true" value of the threshold.

5.3.1 Infection Propagator matrix for bipartite network

In order to evaluate the threshold for a SIS dynamics unfolding on our temporal data-set, we need to implement the Infection Propagator method. We have already notices that our data-based network is bipartite. We would like to translate this feature in terms of a bipartite Infection Propagator matrix (Eq. 4.22), that we have generally defined in the previous chapter. If a network is bipartite then each node belongs either to Group A or Group B of nodes, with the condition that nodes from Group A make contacts (links) only with nodes from Group B, i.e. no contacts are seen between nodes belonging to the same group. When the underlying network of contacts is bipartite, the SIS becomes a 4-compartments model

$$\begin{cases} S_A \to I_A \to S_A \\ S_B \to I_B \to S_B \end{cases}$$

The first arrow stands for the infection process $S_A \xrightarrow[I_B]{\lambda_B} I_A$ meaning that a susceptible node of Group A needs to meet an infectious node from Group B in order to become

itself susceptible; the contagion happens with probability λ_B per contact. Viceversa: $S_B \xrightarrow[I_A]{\lambda_A} I_B$. In general the heterogeneity transmission ratio:

$$\epsilon \equiv \frac{\lambda_B}{\lambda_A} \neq 1 \tag{5.4}$$

The second arrow tells that each node can spontaneously heal with probability $\mu_{A,B}$: $I_A \xrightarrow{\mu_A} S_A$ and $I_B \xrightarrow{\mu_B} S_B$.
Implementing the bipartition of the network, we can write the infection propagator matrix P as:

$$\mathbf{P} = \prod_{t=1}^{T} \begin{bmatrix} 1-\mu_A & \lambda_B \mathbf{Z}(T-t) \\ \lambda_A \mathbf{Z}^\top (T-t) & 1-\mu_B \end{bmatrix}. \tag{5.5}$$

as the instantaneous adjacency matrix can be written in a block form:

$$\mathbf{A}(t) = \begin{bmatrix} \mathbf{0} & \mathbf{Z}(t) \\ \mathbf{Z}^\top (t) & \mathbf{0} \end{bmatrix} \tag{5.6}$$

with the rectangular adjacency matrix $\mathbf{Z}_{i,\alpha}$ that has indices $i = 1, .., N_A$ and $\alpha = 1, .., N_B$, with $N = N_A + N_B$. Notably, the instantaneous contact information are all found in a rectangular "quarter" of the usual adjacency matrix. *In the following analysis we will always consider $\mu_A = \mu_B \equiv \mu$, and possibly variate the value of ϵ above and under 1.*

It is interesting to notice what happens to the I.P. when one of the two transmission rates is zero. Let us consider for example $\lambda_2 = 0$, the expression for the infection propagator becomes:

$$\mathbf{P} = \prod_{t=1}^{T} \begin{bmatrix} 1-\mu & \mathbf{0} \\ \lambda_A \mathbf{Z}^\top (T-t) & 1-\mu \end{bmatrix} = \begin{bmatrix} (1-\mu)^T & \mathbf{0} \\ (1-\mu)^{T-1} \lambda_A \mathbf{Z}^\top_{aggr} & (1-\mu)^T \end{bmatrix}, \tag{5.7}$$

where $\mathbf{Z}^\top_{aggr} = \sum_{t=1}^{T} \mathbf{Z}^\top (t)$. The x eigenvalues of the infection propagator solve the equation $\det(\mathbf{P} - x\mathbf{I}) = 0$, that can be conveniently re-written:

$$\begin{aligned} 0 &= \det \begin{bmatrix} (1-\mu)^T - x & \mathbf{0} \\ (1-\mu)^{T-1} \lambda_A \mathbf{Z}^\top{}_{aggr} & (1-\mu)^T - x \end{bmatrix} \\ &= \det \left(\left((1-\mu)^T - x \right) \mathbf{I}_{N1xN1} \right) \det \left(\left((1-\mu)^T - x \right) \mathbf{I}_{N2xN2} \right) \\ \Leftrightarrow x &= (1-\mu)^T \end{aligned} \tag{5.8}$$

The x eigenvalue is $0 < x < 1$ because $\mu < 1$. Therefore, the threshold value for μ extracted by $\rho(P)^{1/T} = 1$ gives:

$$\rho[\mathbf{P}]^{1/T} = 1 - \mu_c = 1 \Leftrightarrow \mu_c = 0, \tag{5.9}$$

i.e. the system becomes endemic only if $\mu = 0$, meaning that the dynamics is not more a SIS but a SI. So, when one group of the bipartite network cannot pass the infection, the disease always eventually disappears for every transmission rate λ and every recovery rate $\mu > 0$.

5.3.2 Exploring the stable state respect to SIS threshold

We finally exploit the formula for the epidemic threshold given by Eq.(4.21) : $\rho(P)^{1/T} = 1$, to compute the epidemic threshold for our data-set. First of all, we need to notice that the infection propagator matrix depends on two parameters: the initial time t_0 of the considered time-sequence of networks, and the finite period-length T considered:

$$P = \prod_{t=t_0}^{T+t_0} [1 - \mu + \lambda A(T + t_0 - t)] = P(t_0, T). \tag{5.10}$$

Given a data-set corresponding to a sequence of networks $\{A(t)\}$ with t running from 0 to t_{fin}, there are a variety of possible choices of $(t_0; T)$, corresponding to as many IP matrices $P(t_0, T)$. One therefore needs to make a reasoned choice in the definition of the period $[t_0 : t_0 + T]$. The first choice one can make is the trivial $(0, t_{fin})$. However, one can be interested in isolating some particularly interesting stages of the dynamics that start at different t_0 and that last for specific period T. Moreover, it has been shown for many real contact networks[28] that, for a fixed t_0, the threshold evaluated on increasing T portion of the time-sequence converges to a final stable value well before the $T_{max} = t_{fin} - t_{in}$ available time window.

All this considered, we address now the problem of defining a reasonable portion $[t_0 : t_0 + T]$ of the complete time sequence of our data-set. For what concerns the choice of T, we want a time-window that is able to capture the dynamics of the network relevant to the spreading of a disease, but the shortest possible in order to speed up the computations. For what concerns t_0, we want to start considering the system after the initial phase of growth. In fact, what we aim to define is a temporal portion of the system in which the contact dynamics is captured with all its peculiar variability and that is representative of an over-all stable and "mature" stage in the network history. Epidemiologically, this stage is relevant to assess for the vulnerability of a typical and well affirmed contacts dynamics in a target population. Let us consider three period-lengths T to test. We choose them according to the network dynamics studied in the previous section, which pointed out that a reasonable time window can be of one semester or longer. We take therefore three periods of width $T = 168, 336, 672$ days, that are respectively one semester, one year, two years (where moths have 28 days). These intervals encode a long time dynamics, for a human social system based on short term relationships as ours. We numerically compute the threshold for different t_0, ranging from the beginning to the end of the data series. We aim to see how the threshold varies and if it approaches stability when the undergoing network dynamic does the same. For a matter of simplicity, for the moment we limit our self to the case of one transmission probability for both women and men: $\lambda_A = \lambda_B \equiv \lambda$. A priori, we could evaluate the threshold for μ or λ. At this stage, we evaluate λ_c because this allows to control the recovery time scale $\tau_\mu = 1\text{day}/\mu$ and choose it so that $\tau_\mu < T$, i.e. the recovery happens within the time window considered. Being $1/T \sim 0.005$ for the smaller T=168 days, we can choose as a good test the value $\mu = 0.02$, corresponding to $\tau_\mu \sim 50$ days.

Figure (5.11) shows the critical values of the transmission probability, it is computed for three time-length of the time sequence encoded in the IP matrix. For each choice of T, i.e. the color of the curve, and each starting time t_0, corresponding to the x-value of the points plotted minus T/2, the relative y-value is given by:

$$\lambda_c \qquad \text{so that} \qquad \rho\left[P(\lambda; t_0, T, \mu)\right] = 1. \tag{5.11}$$

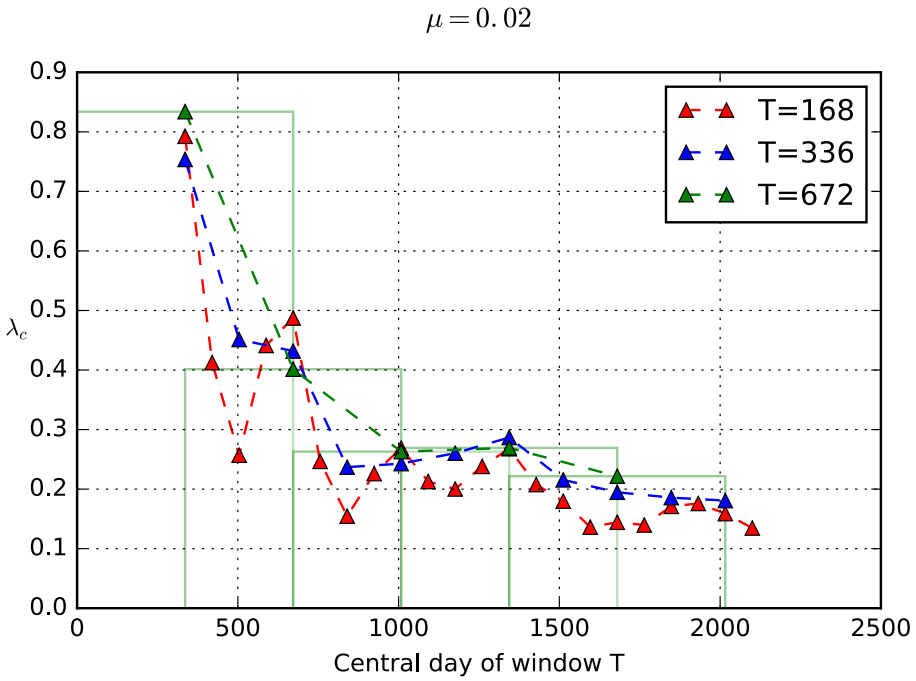

Figure 5.11: **Epidemic threshold for various T and** t_0.

The light green boxes in the fore ground are intended to make more clear the time window $t \in [t_0 : t_0 + T]$ (readable from the x-axis) whose time-snapshot $A(t)$ have been used to built the P matrix relative to the central green triangle point. We recall that the time resolution is always 1 day. Comparing the curves in different colors in Fig.(5.11), we notice a common trend and nearly the same values for λ_c inside the same green boxes. The common trend shows initial high values for λ_c (safe system), that rapidly lower (vulnerable system) and then stabilize, roughly after day 1000. What happens is that the system becomes more vulnerable with the passing of time. The stabilization of the threshold matches with the dynamics of the topological features: those as well stabilize in time, after a transient growth for days $t \in [0, \sim 1000]$ days.
The fact that the y-values are quite the same across the different size of T shows that the smaller window T=168 days is actually wide enough to include all the network dynamics that has an impact on the SIS spread. Lastly, we notice that the red curve allow to see a stable epidemiological situation for $t > 1500$ days; there are still small fluctuations that are explained by the small variation in the contact dynamics.

The results shown in Fig.(5.11) justify the choice of a representative period of the data-set to be taken of width 168 days and starting the first day of last semester:

$$\Delta T^* = [2063 : 2231]\, days. \tag{5.12}$$

5.3.3 Transition diagram in parameters' space (λ, μ)

Let us consider the portion of time ΔT^* above defined. The temporal network associated to this interval is represented as a sequence $\mathcal{A}$ of 168 time-ordered bipartite adjacency matrices $A(t)$, each corresponding to the configuration that the network takes at time $t \in \Delta T^*$. The time resolution is of the sequence is 1 day.
The total number of nodes active in this time window is $N = 4403$, formed by $N_A = 2629$ men and $N_B = 1774$ women. Each $A(t)$ is therefore a square matrix of dimension

N, formed by blocks $Z(t)$ of size $N_A \times N_B$, as described in Eq.(5.6). The network is not weighted because we find $A_{ij}(t) = 0, 1$ for every link and day. Thus we can perform the numerical computation of the threshold using the formula for the un-weighted Infection Propagator matrix of Eq.(5.10).
Unlike the previous computation of the threshold, this time we assess for the critical recovery probability:

$$\mu_c \qquad \text{so that} \qquad \rho\left[P(\mu; t_0, T, \lambda)\right] = 1. \tag{5.13}$$

Above μ_c the system is not endemic $i_{stat} = 0$, while for $0 \leqslant \mu \leqslant \mu_c$ the system reaches an endemic state $i_{stat} > 0$. At this stage, where we are considering $\lambda = \lambda_A = \lambda_B$, this choice seems arbitrary. However it will become necessary in the following section where we inspect the impact of $\lambda_A \neq \lambda_B$. Thus, to present a more linear dissertation, from here after we will always compute the threshold with respect to μ.

Varying the value of λ, the correspondent μ_c draws a so-called **phase transition diagram** in the space of parameters (λ, μ) of the SIS disease. The phase transition regards the order parameter i_{stat}: the Infectious state of the over all system in the long time limit. μ_c marks the transition from a situation in which the disease eventually disappears to the opposite, where the disease reaches a final non zero persistence. For $\mu \leqslant \mu_c$, the system is said to be in the endemic state; endemic in fact means that an infection is maintained in the population without the need for external inputs. The numerical results are shown in Figure(5.12) where we see the phase transition diagram $\mu_c(\lambda)$

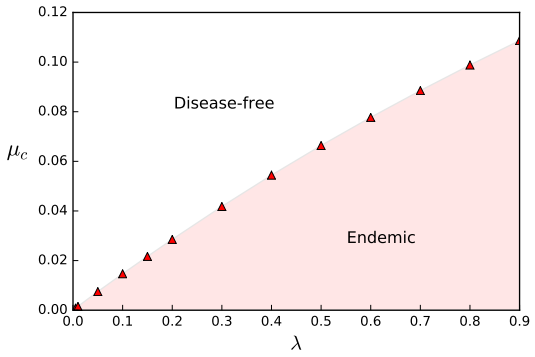

Figure 5.12: **Epidemic threshold for time window** ΔT^*, computed using the method of the Infection Propagator.

For every value of $\lambda \in [0, 1]$ there exists a non zero value of the threshold: this means that there is always a range of recovery probabilities that leads to an endemic state. The values of the threshold μ_c are all quite small: $\mu_c < 0.12$, corresponding to quite long infectious periods $\tau_c > 8.33$ days. Given, for example, a small transmission probability $\lambda = 0.2$, only for high mean infectious period $\tau_\mu > \tau_c \simeq 33$ days the disease becomes endemic in the population. This means that the system is quite resistant respect to low viral-load diseases, as normally recoveries happen faster than one per month. If we consider the total available space of parameters, the coloured endemic region is a small portion.

5.3.4 Transition diagram in parameters' space (λ, μ, ϵ)

We turn now to focus on the impact of the bipartite nature of the network on the epidemics. A priori, it is possible that the two categories (men or women) have different

transmission probabilities. This is actually the case for nearly all sexually transmitted diseases. Medical studies agree with the finding that men have naturally more chance to transmit a STD to a partner because of anatomic conformation. Instead, the value of the real-life recovery probability μ depends mostly on the promptness of the diagnosis of disease, and on the time of recovering proper of the disease. Therefore, μ depends on the disease and is also tunable with sanitary policies, while it does not intrinsically depend on the gender. We will therefore focus on the impact of different $\epsilon = \lambda_2/\lambda_1$.

Impact of ϵ

If we allow for $\epsilon \neq 1$, the space of the SIS parameters enlarges from two to three dimensions: (λ, μ, ϵ), having re-named $\lambda_1 = \lambda$ and $\lambda_2 = \epsilon\lambda$. The former phase diagram of Fig.(5.12) thus projected onto the plane defined by $\epsilon = 1$. Now we want to explore the ϵ dimension while keeping λ fixed, thus projecting this time on the $\lambda =$ cost plane. The threshold has to be understood as a two-dimensional object defined by equation $f_c(\lambda, \mu, \epsilon) = \rho\left(P(\lambda, \mu, \epsilon)\right)^{1/T} - 1 = 0$ in the three-dimension space of parameters. We observe that the domain of f_c is a region of the 3D space (λ, μ, ϵ) delimited by:

$$\begin{cases} 0 \leqslant \mu \leqslant 1 \\ 0 \leqslant \lambda \leqslant 1 \\ 0 \leqslant \epsilon \leqslant \frac{1}{\lambda} \end{cases} \tag{5.14}$$

because the transmission probability $\lambda_2 = \epsilon\lambda \in [0, 1]$ needs to have these bounds to remain meaningful. Fixing a λ , we show the critical function $\mu_c(\epsilon)$ with $\epsilon \in [10^{-2}, 1/\lambda]$ in Figure(5.13).

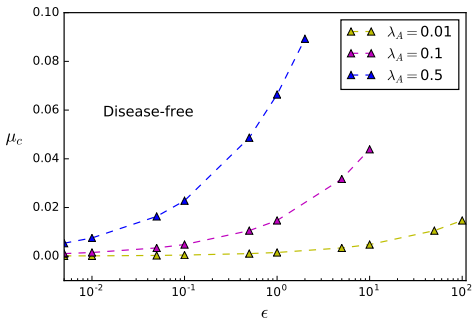

Figure 5.13: Threshold μ_c in function of transmission ratio $\epsilon = \lambda_B/\lambda_A$. The right most point for all three curves corresponds to $\epsilon_{max} = 1/\lambda_A$.

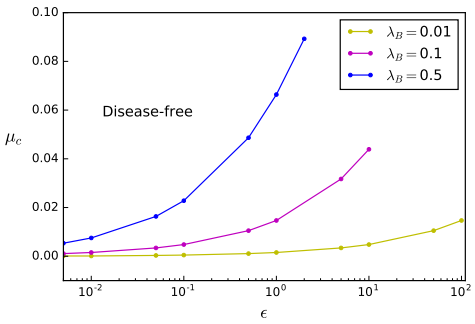

Figure 5.14: Threshold μ_c in function of transmission ratio $\epsilon = \lambda_A/\lambda_B$, the inverse of the one defined for the figure on the left. The values of μ_c are the same as in the figure on the left, up to the 6th decimal position.

Observing the plots for three fixed values of $\lambda \in \{0.01, 0.1, 0.5\}$, we see that obviously μ_c is an increasing function of ϵ. In fact, given λ fixed, ϵ tells how much the other $\lambda_B = \epsilon\lambda$ is smaller/bigger than λ_A. Thus an increase of ϵ means an over-all increasing of the transmission probability of nodes, and this makes the system always more vulnerable (high values of μ). For $\epsilon \lessapprox 10^{-2}$, the system approaches a situation where it is always disease-free $\mu_c \to 0$. This confirms the analytical finding that when one of the two group cannot transmit the disease, it always eventually dies out.

We have performed the same calculation inverting the λs in the definition of the ratio ϵ. The results for μ_c are reported in Fig.(5.14) and are nearly the same as the ones reported in the plot on its left (differences arises at order 10^{-4}). From this, we can conclude that the two categories of nodes do not have an a different impact on the threshold. Thus, although having different connectivity patterns, they have an over all symmetric impact on the spreading of the disease. This points out that an analytical expression of the threshold should be symmetric for the switching of A and B.

Fixing a mean transmissibility

In the previous paragraph, the effect of one group respect to the other in the spreading of the disease has been shown to be the symmetric. Now we use another strategy to explore the impact of ϵ on the system, and we will see its application to the prediction of an analytical form of the threshold in the following chapter.
We decide to fix a quantity that accounts for a sort of mean transmission probability of the population, and inspect in this context the variation of ϵ. We define as the mean transmissibility of the network the quantity

$$\langle\lambda\rangle = \frac{N_A\lambda_A + N_B\lambda_B}{N_A + N_B}. \tag{5.15}$$

Enforcing $\langle\lambda\rangle$ constant, we have only one degree of freedom in the transmissibility, that can be either λ or ϵ. If we choose it to be ϵ, than λ is determined by

$$\lambda(\epsilon;\langle\lambda\rangle) = \langle\lambda\rangle\frac{N_A + N_B}{\epsilon N_B + N_A}. \tag{5.16}$$

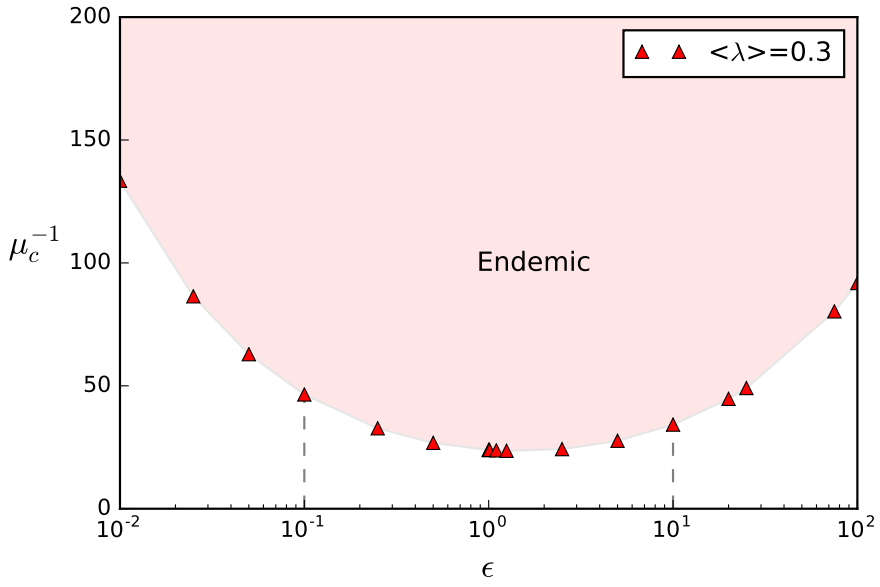

Figure 5.15: **Threshold $1/mu_c$ for fixed mean trasmissibility**.

What can be computed as a threshold is therefore the recovery probability μ_c. We fix now $\langle\lambda\rangle = 0.3$ and ranging with $\epsilon \in [0.01, 100]$ we compute the critical recovery probability [2]. We present a plot for the associated infectious period $\tau_\mu(\epsilon) = 1/\mu_c$, because plotting the infectious period is more effective in order to interpret the result. The critical curve plotted in Fig.(5.15) discriminates between a system in the endemic state for $\tau > \tau_c$, and a safe state that lies under the curve. We observe that there exists a τ minimum: the disease here leads the system to a condition of maximal vulnerability. By numerical inspection we find that $\epsilon_{min} \simeq 1.482$: this value is bigger than 1, thus the most vulnerable case happens when women have an higher transmission probability respect to men, provided that a mean transmissibility is fixed. This result from the necessity of compensate the lower number of women respect to men. In the following chapter we will be able to connect this minimum value to the analytical form of the threshold as function of λ_A , λ_B , μ.

5.3.5 Risk assessment for real STD

The critical values λ_c (reversing $\mu_c(\lambda)$) of the escort business network can now to be contrasted with the typical ones of the common sexually transmitted disease that could actually hit the network under study. We want to see if the system is a feasible root for the spreading of the most common STD. In Table (5.1) we summarize the per-contact rate of transmission for some viral and bacterial common diseases: HIV, Gonorrhea, and Chlamydia. The data are taken by both recent and classical epidemiological studies, and here we report just the average value of the transmission probability, in order to make the table more readable.

STD	Type of contact unprotected	λ_A man $\to$ woman	λ_B woman $\to$ man	λ_B/λ_A
HIV-1	V.	$0.08\,10^{-2}$	$0.04\,10^{-2}$	0.5
	A.	$1.4\,10^{-2}$	$0.42\,10^{-2}$	0.5
Gonorrhea		0.60	0.25	0.4
Chlamydia		0.05	0.05	1
Syphilis		0.01	0.01	1

Table 5.1: **Per contact transmission probability for common STD.**

Recalling Fig.(5.12) and focusing on λ values, we conclude that for HIV-1 the network is in a region where it is nearly always safe. For the portion of parameter-space around those λ, the endemic (coloured) region is very small. Our type of contact dynamics is not effective for low viral-load disease, for which the spreading needs repeated contacts between the same individuals in order to diffuse. This result founds support in the analysis carried out by Rocha et al. [40] on the same network. They found that this network is not a major reservoir for HVI-1 and that pathways other than his type of commercial sex are needed to explain the endemic state of HIV-1 epidemics in Brazil.

[2] We notice that this time the values that we can take, keeping $\lambda_{1,2} < 1$ are in the interval $[(\langle\lambda\rangle N - N_2)/N_1 : N_2/(\langle\lambda\rangle N - N_1)]$.

The same can be stated for Chlamydia and Syphilis: this network is not vulnerable for these diseases.

For Gonorrhea instead, the high per-contact probability of transmission makes the disease endemic for our system, when $\mu < 0.07$ i.e. when the infectious period is $\tau_\mu > 15$ days. This means that if people do not recover within 15 days, the disease becomes endemic.

Chapter 6

Epidemic threshold: a theoretical analysis

After having computed the threshold for our network using the Infection Propagator approach, we would like to find a theoretical model that reproduces to some degree of approximation the phase transition diagram, for both the case $\epsilon = 1$ and $\epsilon \neq 1$. We firstly inspect if any of the main theoretical models present in literature of mathematical modelling for epidemiology suits our case. The test of all the models is infeasible, and goes way beyond our scope. What is more interesting is to test some simple yet powerful analytical approaches that are usually claimed to give good results in real-world network. It is important to recall that most of those accepted and robust models apply to static networks (both quenched and annealed), therefore their straight forward application to our case-study is likely not to give the wished results. It is nevertheless interesting to see to what extent the model-based prediction differs for the true threshold μ_c, and if its value is smaller or bigger.

In the following chapter we present three analytical models that we found most interesting to test on our temporal network.

- The first one is the already mentioned *degree-block approximation*: it is based on the guess that the main feature characterizing the contacts of nodes is their degree-degree contact distribution $P(k|k')$. Therefore each node in placed in its degree class and makes contacts with the other on the basis of $P(k|k')$. The system of deterministic equations for the evolution of the infected nodes in each k-class is easy to write. This model comes from the field of static annealed networks, but we can tailor it to the temporal case with some suitable hypothesis. This has been proved to be a good model for real static network with heterogeneous degree, for which often $P(k)$ is a power-law, i.e. the network is called free-scale (Section 3.3.1). Here the task will be to generalize the equations to a bipartite population and to interpret the $P(k)$ for a temporal sequence of networks.

- The second approach we investigate comes from the field of epidemics on temporal networks. It is actually one of the few works that gives an explicit formula for the epidemic threshold of a SIS dynamics. Is is called the *Activity Driven model*, and describes each node of the system in terms of its class of temporal activity a. The system of deterministic equations for this models are written in total analogy with the degree-block, but this time the k-classes are substituted by a-classes.

- The third and last approach is what we call the *annealed* approximation of the tem-

poral network. It consist in the approximation of each snapshot with the time average of the sequence. As we pointed out in Section 4.5.2 , this approximation is very good in the case of slow disease respect to the network dynamics; however nothing ensures that for other ranges of the disease parameters this approximation still holds.

6.1 Bipartite Degree-Block model

To apply this model to a temporal network, we can consider each of our snapshots as belonging to the same statistical ensemble of graphs defined by a common degree distribution. In this sense, the degree-block is an ensemble view of the temporal network, as we guess that the disease sees an average network structure described by a $P(k)$.

In detail, we guess that an annealed approximation of our network can be described under the following assumptions:

- the degree-degree contact distribution $P(k|k')$ of the network is the only relevant statistical property needed to describe the contacts
- there is homogeneous mixing inside each sub-group of individuals with the same degree (called the *degree-block approximation*). It means that all nodes with the same degree are assumed to be statistically equivalent.
- there is no degree-degree correlation. This last hypothesis is not necessary, but is justified by the analysis on the degree-degree uncorrelation of our data-set, performed in Section 5.2.4 .

The calculation for a bipartite network is here developed extending the one found in the classical text *Dynamical Processes on Complex networks*[24]. Our result is confirmed by an analogous derivation found in an article about the spreading of sexual disease among an heterosexual population [31].

Let S_k^G,I_k^G (with G=A,B that identify each group of the bipartite net) be the number of susceptible and infectious individuals in degree class k. Densities s_k^G, $i_k^G$are obtained by dividing S_k^G, I_k^G by the number of nodes with degree k: $N_k^G = S_k^G + I_k^G$, that is constant being the system closed.
Thanks to the degree-block approximation, we can use a system of differential equations to describe the $S_A I_A S_A$,$S_B I_B S_B$ dynamics[1]:

$$\begin{cases} \dfrac{di^A{}_k}{dt} &= -\mu i^A{}_k(t) + \lambda_B k \left(1 - i^A{}_k(t)\right) \Theta_k^B(t) \\ \dfrac{di^B{}_k}{dt} &= -\mu i^B{}_k(t) + \lambda_A k \left(1 - i^B{}_k(t)\right) \Theta_k^A(t) \end{cases} \tag{6.1}$$

with

$$\begin{cases} \Theta_k^A(t) = \sum_l P(l|k) i_l^A(t) \\ \Theta_k^B(t) = \sum_l Q(l|k) i_l^B(t) \end{cases} \tag{6.2}$$

where Θ_k^A is the fraction of infected nodes of group A in contact with a node of degree k belonging to group B. As we assume degree-uncorrelated networks, $P(l|k) = \frac{lP^A(l)}{\langle l\rangle^A}$ and $Q(l|k) = \frac{lP^B(l)}{\langle l\rangle^B}$, so the $\Theta_k^G(t)$ becomes k-independent.

[1]Time is considered continuous.

We impose now the stationary condition for each equation in the system Eq.(6.1) and find

$$\begin{cases} (i_k^A)^{stat} & \equiv i_k^{A*} = f(\Theta^B) = \dfrac{\lambda_B k \Theta^{B*}}{\mu + \lambda_B k \Theta^{B*}} \\ (i_k^B)^{stat} & \equiv i_k^{B*} = f(\Theta^A) = \dfrac{\lambda_A k \Theta^{A*}}{\mu + \lambda_A k \Theta^{A*}} \end{cases} \tag{6.3}$$

Now, by substitution of (6.3) in Eq.(6.2), we find the system of equations :

$$\begin{cases} \Theta^{A*} & = g(\Theta^{B*}) = \dfrac{1}{\langle k \rangle_A} \displaystyle\sum_m m P^A(m) \dfrac{\lambda_B m \Theta^{B*}}{\mu + \lambda_B m \Theta^{B*}} \\ \Theta^{B*} & = g'(\Theta^{A*}) = \dfrac{1}{\langle k \rangle_B} \displaystyle\sum_l l P^B(l) \dfrac{\lambda_A l \Theta^{A*}}{\mu + \lambda_A l \Theta^{A*}} \end{cases} \tag{6.4}$$

The system can be solved substituting Θ^{B*} into $\Theta^{A*} = g(\Theta^{B*})$, that gives finally

$$\Theta^{A*} = \psi(\Theta^{A*}). \tag{6.5}$$

We are interested in finding the condition for which a solution $\Theta^{A*} > 0$ of Eq.(6.5) exists, because Θ^{A*} is the fraction of infectious nodes of type A that are in contact with at least one node of type B and a non zero stationary value implies an endemic stationary state for the whole system. We notice that the zero solution always holds, while a non zero can be found, by geometrical consideration on the slopes and convexity of curves $y_1 = \Theta^{A*}$ and $y_2 = \psi(\Theta^{A*})$, by imposing :

$$\left(\frac{d}{d\Theta_*^A} \psi \right)_{\Theta_*^A = 0} = 1. \tag{6.6}$$

Omitting long calculations, the left side of Equation (6.6) is found:

$$\left(\frac{d}{d\Theta_*^A} \psi \right)_{\Theta_*^A = 0} = \frac{\langle k^2 \rangle_A \langle k^2 \rangle_B}{\langle k \rangle_A \langle k \rangle_B} \frac{\lambda_A \lambda_B}{\mu^2}, \tag{6.7}$$

that finally gives the threshold condition:

$$\boxed{\left(\frac{\mu^2}{\lambda_A \lambda_B} \right)_c = \frac{\langle k^2 \rangle_A \langle k^2 \rangle_B}{\langle k \rangle_A \langle k \rangle_B}.} \tag{6.8}$$

We notice that this expression is symmetric in the exchange of A and B: this is consistent with the findings of Section 5.3.4.
We now want to contrast the degree-block theoretical prediction for the threshold given by Eq.(6.8) with the real values of the temporal-network threshold. The right hand coefficients of Eq.(6.8) are extracted from the data using the formulas:

$$\langle k_G \rangle = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N_G} \frac{k_i(t)}{N_G(t)} \qquad \langle k^2{}_G \rangle = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N_G} \frac{k_i^2(t)}{N_G(t)}. \tag{6.9}$$

where $N_G(t)$ is the number of people in group G active at time t. This mean we consider each snapshot as coming from a given ensemble of $P^A(k)$ and $P^B(k)$. The moments of the

	A women	B men
$\langle k \rangle$	1.059	1.106
$\langle k^2 \rangle$	1.231	1.461

degree distribution extracted by data in the time window of last semester $\Delta t* = [2\,063 : 2\,231]$ are:

It is interesting to notice that the daily behaviour of men and women with respect to k is very similar. It is quite surprising that the daily mean degree of men is higher that women's. This means that a man usually has more than one encounter when he decides to be active in the business. However, in the long time behaviour (e.g. one semester) women have higher cumulated degree ($k^A_{max} = 85$) then men ($k^B_{max} = 20$), thus restoring the expected sellers-buyers asymmetry that was showed in Section 5.2.3

Same transmission probability

Firstly, we consider the homogeneous transmission $\lambda_A = \lambda_B$. We check the theoretical prediction

$$\mu_c = m_k \lambda \qquad \text{with} \qquad m_k \equiv \sqrt{\frac{\langle k^2 \rangle_A \langle k^2 \rangle_B}{\langle k \rangle_A \langle k \rangle_B}} \simeq 1.239 \tag{6.10}$$

contrasting it with μ_c values obtained with the Infection Propagator. From plots (6.1)

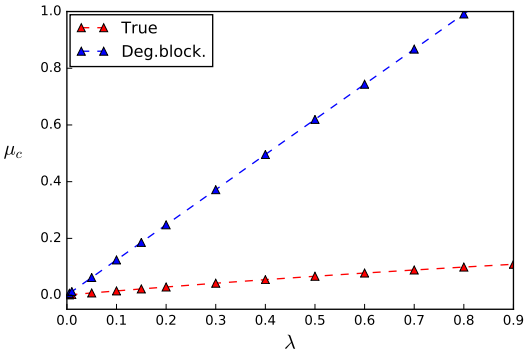

Figure 6.1: **Degree block threshold** vs true threshold.

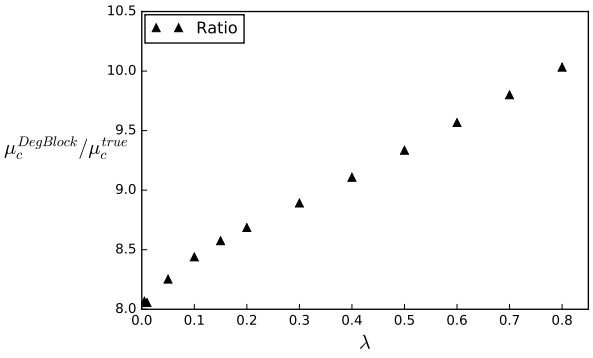

Figure 6.2: Ratio of the degree-block threshold over the true threshold.

and (6.2) we see that the degree-block approximation gives an about ten times higher recovery probability threshold. Thus it largely over-estimates the vulnerability of the network, as the endemic area (between the curve μ_c and the x-axis) is much wider than for true dynamics. The true temporally-changing network is much safer than the degree-approximation. We can understand this overestimation of vulnerability by the degree-block because it approximates *each* node of the network as having an average daily degree around 1. In the temporal network this is true, but only for the nodes that are active at that time step, which are a minority with respect to the total population (recalling that the daily network is very sparse). This result shows the temporal features of the network, or some topological properties other than $P(k)$, are the responsible for the behaviour of the empirical network.

Different transmission probability

Even if we have just proved that the degree-block approximation fails in this context, it is still interesting to see how it behaves with respect to an allowed heterogeneity in the transmission probability: $\lambda_A \neq \lambda_B$. Following the same analysis we have performed via numerical computation in Sec. 5.3.4, we can impose the constraint $\langle\lambda\rangle = \text{cost}$ on Eq.(6.8) and express the threshold τ_{μ_c} in function of ϵ :

$$\tau_{\mu,c}(\epsilon;\langle\lambda\rangle) = \frac{1}{m_k \langle\lambda\rangle} \frac{n_A + \epsilon n_B}{\sqrt{\epsilon}} \tag{6.11}$$

where $n_A = N_A/(N_A + N_B)$ and $n_B = N_B/(N_A + N_B)$. The function (6.11) has a minimum value, that we find by setting the derivative in ϵ equal to zero [2], and we get:

$$\epsilon_{min}^{Deg.Block} = \frac{N_B}{N_A} \simeq 1.4819 \tag{6.12}$$

This result shows that the critical $\tau_c(\epsilon;\langle\lambda\rangle)$ is asymmetric for the switch of A and B, with the point of maximal vulnerability ϵ_{min} that depends only on the number of nodes of each group. If we contrast this result with the numerical true value $\epsilon_{min}^{true} \simeq 1.4825$ we find a very good agreement, as we can see graphically see from Fig.(6.3). What can we learn from this?

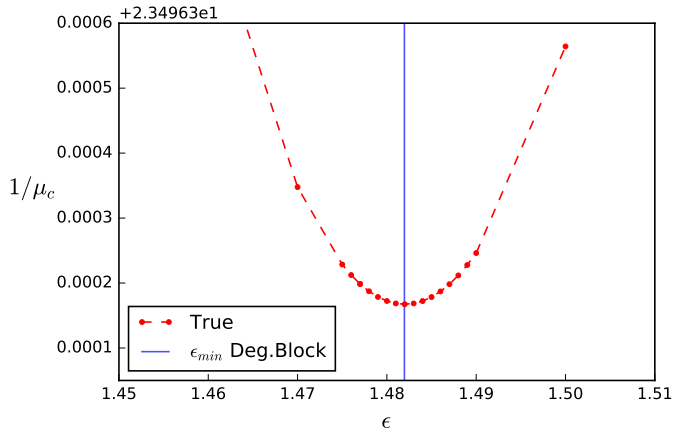

Figure 6.3: **Minimum value of τ_μ threshold, when we fix the mean transmissibility**.

It is important to notice that the value of ϵ_{min} equal to N_B/N_A depends on the functional form of τ_c we see in Eq.(6.11) regardless for the value of the coefficient m_k. Equivalently, every other theoretical model that gives a threshold for a bipartite network with the same functional form as Eq.(6.8) will have the same minimum ϵ:

$$\left(\frac{\mu^2}{\lambda_A \lambda_B}\right)_c = \text{cost} \implies \epsilon_{min} = \frac{N_B}{N_A} \tag{6.13}$$

Therefore, the agreement we find with ϵ_{min}^{true} tells us that every such model is approximately a good representation of our phase-transition diagram. Then, if λ is equal for both groups, then this means a linear approximation of the diagram $\mu_c(\lambda)$ is good, as it is confirmed by Figure (5.12).

[2] The second derivative positive is checked as well.

6.2 Bipartite Activity Driven model

The second model we inspect is the so called *activity-driven*. The activity-driven network class of models (Perra, Gonçalves et al., 2012; Starnini and Pastor-Satorras, 2013) is based on the concept of activity rate, defined as the probability per unit time that an individual engages in a social activity. All dynamics of the networks is encoded in the function $F(a)$ that characterizes the probability for a node to have an activity rate a. To better understand it, we describe how a synthetic activity-driven network can be built. The model considers N nodes whose activity a_i is assigned randomly according to the distribution $F(a)$. During each time step the node i is considered active with probability a_i. Active nodes generate u links that are connected to as many individuals chosen uniformly at random. Finally, time is updated $t \to t+1$ and the model is iterated with no memory of past configuration. The model output is a sequence of graphs, depending on the distribution $F(a)$, which is updated at every time step t. $F(a)$ can be very broad as empirical evidence shows that the activity rate varies considerably from individual to individual. An important result given by this model is that the aggregated network after time T has a degree distribution which depends on the activity distribution as $P_T(k) = (1/T)F(k/T - \langle a \rangle)$. The empirically observed power-law activity distributions $F(a)$ can thus explain the long tails in the real-life degree distribution of social networks. We will focus here of the work of Perra, Gonçalves et al. [45]. They considered the behaviour of the SIS model in activity-driven networks, writing dynamical mean-field equations for the infected individuals in the class of activity rate a, at time t, namely, $I^t{}_a$. The discrete-time dynamical evolution considers concurrently the dynamics of the network and the epidemic model, yielding

$$\begin{aligned} {I_a}^{t+\Delta t} &= -\mu \Delta t {I_a}^t + {I_a}^t \\ &+ \lambda u ({N_a}^t - {I_a}^t) a \Delta t \int da' \frac{{I_{a'}}^t}{N} \\ &+ \lambda u ({N_a}^t - {I_a}^t) \int da' \frac{a' \Delta t {I_{a'}}^t}{N} \end{aligned} \tag{6.14}$$

where ${N_a}^t$ is the total number of individuals with activity a at time t, λ is the probability with which infected individuals can propagate the disease to healthy neighbours per contact, μ is the rate of recovery of individuals. The first two terms on the right side of Eq.(6.14) accounts for the probability of a node in activity class a being infectious at time t and not to recover in the time step Δt, the third term takes into account the probability that a susceptible of class a is active and acquires the infection getting a connection from any other infected individual (summing over all different classes), while the last term takes into account the probability that a susceptible, independently of his activity, gets a connection from any infected active individual. We remember that a node i in class a cannot change its class of activity potential, meaning at every time step it will have the same probability a of activate/not activate. A linear stability analysis of Eq. (6.14) leads to the epidemic threshold for the activity driven model:

$$\left(\frac{\mu}{\lambda}\right)_c = u \left(\langle a \rangle + \sqrt{\langle a^2 \rangle} \right) \tag{6.15}$$

We generalize this approach to a bipartite network and find the relative threshold. The evolution equations for ${I_a}^A$ and ${I_a}^B$ are:

$$\begin{cases} I_a{}^A(t+\Delta t) &= (-\mu\Delta t+1)I_a{}^A(t)+\lambda_B u\left(N_a{}^A(t)-I_a{}^A(t)\right)a\Delta t\int da'\frac{I_{a'}{}^B(t)}{N_B}+ \\ &\quad +\lambda_B u\left(N_a{}^A(t)-I_a{}^A(t)\right)\int da'\frac{a'\Delta t I_{a'}{}^B(t)}{N_B} \\ I_a{}^B(t+\Delta t) &= (-\mu\Delta t+1)I_a{}^B(t)+\lambda_A u\left(N_a{}^B(t)-I_a{}^B(t)\right)a\Delta t\int da'\frac{I_{a'}{}^A(t)}{N_A}+ \\ &\quad +\lambda_A u\left(N_a{}^B(t)-I_a{}^B(t)\right)\int da'\frac{a'\Delta t I_{a'}{}^A(t)}{N_A} \end{cases} \tag{6.16}$$

These equations can be integrated in a and then coupled with another pair of equations coming from the above multiplied by a and integrated as well. By defining $i(t) = \int da I_a(t)/N$ and $\theta(t) = \int da\, a I_a(t)/N$, and neglecting terms $o(i^2)$ one finds:

$$\begin{cases} \frac{d}{dt}i^A(t) &= -\mu i^A(t)+\lambda_B u\langle a\rangle_A\, i^B(t)+\lambda_B u\theta^B(t) \\ \frac{d}{dt}i^B(t) &= -\mu i^B(t)+\lambda_A u\langle a\rangle_B\, i^A(t)+\lambda_A u\theta^A(t) \\ \frac{d}{dt}\theta^A(t) &= -\mu\theta^A(t)+\lambda_B u\left\langle a^2\right\rangle_A i^B(t)+\lambda_B u\langle a\rangle_A\,\theta^B(t) \\ \frac{d}{dt}\theta^B(t) &= -\mu\theta^B(t)+\lambda_A u\left\langle a^2\right\rangle_B i^A(t)+\lambda_A u\langle a\rangle_B\,\theta^A(t) \end{cases} \tag{6.17}$$

where we have also performed a limit to the continuous time. The max eigenvalue of the Jacobian matrix associated to the system gives the threshold condition, once we set it equal to zero. Thus we find:

$$\left(\frac{\mu^2}{\lambda_B\lambda_B}\right)_c = \frac{u^2}{2}\left(\left\langle a^2\right\rangle_A+2\langle a\rangle_A\langle a\rangle_B+\left\langle a^2\right\rangle_B\right)+ \\ \frac{u^2}{2}\sqrt{\langle a^2\rangle_A^2+\langle a^2\rangle_B^2+4\langle a\rangle_A\langle a\rangle_B\left(\langle a^2\rangle_A+\langle a^2\rangle_B\right)+4\langle a^2\rangle_A\langle a\rangle_B^2+4\langle a^2\rangle_B\langle a\rangle_A^2-2\langle a^2\rangle_A\langle a^2\rangle_B} \tag{6.18}$$

Given a data-set, the activity a_i^A of node i from group A from a data set can be evaluated by counting the number of contacts that involve node i in the typical time-step window Δt, and divided by the number of all contacts $N_e(\Delta t)$ involving all nodes of group A, in the same interval. What remains undefined is u. As described in the article, if each active nodes creates u links, then the total average edges per unit time are $E_t = \frac{u}{2}\left(\langle a\rangle_A N_A+\langle a\rangle_B N_B\right)$ yielding $u = \frac{2E_t}{\left(\langle a\rangle_A N_A+\langle a\rangle_B N_B\right)}$. Applying this model to our temporal network, we have the intrinsic $\Delta t = 1$ day. Therefore the activity rate of node i at time t is $a_i(t) = \sum_j A(t)_{ij}/N_e(t) = s_i(t)/N_e(t)$ it is actually the strength of node i at that time. Being the networks un-weighted,in our case is moreover true that $a_i(t) = k_i(t)/N_e(t)$ is the instantaneous degree of node i over the number of edges active at time t. Thus, the average activity associated to a node of group A is evaluated via:

$$\langle a\rangle_A = \sum_{i=1}^{N_A}\sum_{t=1}^{T}\frac{a_i(t)}{TN_A(t)}, \tag{6.19}$$

where $N_A(t)$ is the number of active nodes of group A at time t. Moreover $E_t = \sum_t N_{contacts}(t)/T = N_{contacts}/T$. The moments of the activity distribution extracted by data in the time window of last semester $\Delta t* = [2\,063 : 2\,231]$ are: Plugging the data

	A women	B men
$\langle a \rangle$	0.0277	0.0289
$\langle a^2 \rangle$	0.0011627	0.00124488
u	0.7243	0.7243

of Table in Eq.(6.18) we find:

$$\left(\frac{\mu}{\sqrt{\lambda_A \lambda_B}}\right)_c^{data} \simeq 0.0456 \tag{6.20}$$

Considering the simple case $\lambda_A = \lambda_B$, we see that this value is way too small respect to the true value of the threshold, as we can see from Fig.(6.4). Notably, this model pushes the threshold in the opposite direction with respect to the degree-block: the system for this activity pattern is more safe than the true temporal one. This is probably due to the fact that the heterogeneity of the connection pattern here is not considered. We also notice that this threshold gives again $\epsilon_{min} = N_B/N_A$.

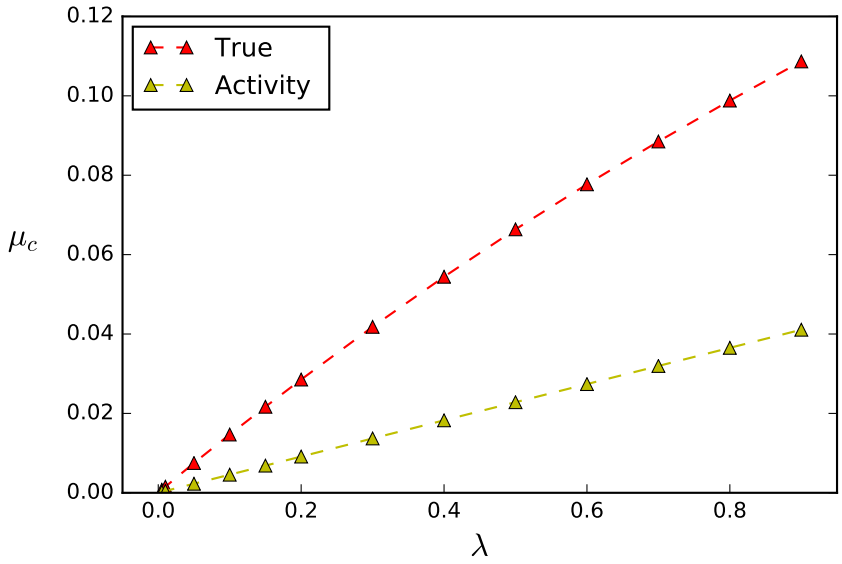

Figure 6.4: **Activity driven threshold vs true threshold.**

6.3 Average Temporal network

For the third and last analytical model, we consider what we call the *annealed approximation of the temporal network*. It consists in the approximation of each snapshot with the time average of the sequence:

$$A(t) \sim \bar{A} \qquad \forall t \qquad \text{where} \qquad \bar{A} \equiv \frac{\sum_{t=1}^{T} A(t)}{T} \tag{6.21}$$

and each $A(t)$ is the $N \times N$ adjacency bipartite matrix of the network at time step t.

Let us consider $\lambda_A = \lambda_B$. By means of this approximation, the SIS equation for the N-vector $p(t)$ describing probability of each node to be infectious at time t reads:

$$p(t+1) = (1-\mu)p(t) + \lambda\bar{A}p(t), \tag{6.22}$$

where we have already performed the linearization for small p. The threshold condition, given by the instability of the zero solution, is :

$$\left(\frac{\mu}{\lambda}\right)_c = \rho\left(\bar{A}\right) \tag{6.23}$$

where ρ has the usual meaning of spectral radius of the matrix. As we found in Sec. 4.5.2, this threshold is the same we recover starting from the Infection Propagator matrix $P = \prod_{t=1}^{T}(1-\mu+\lambda A(T-t))$ and considering the threshold condition $\rho(P)^{1/T} = 1$ in the limit case $\lambda/(1-\mu) \ll 1$. We therefore expect to find a good accordance for the time-average threshold μ^{Ann} and the temporal μ^{true} when $\lambda_c/(1-\mu) \ll 1$.
Figure (6.5) shows that the the time-average threshold μ^{Ann} is a good approximation of the true value of the threshold, throughout all the values of λ. As λ goes from 0 to 1, the value of $\lambda/(1-\mu_c{}^{true})$ rises from $o(10^{-2})$ to $o(1)$, thus explaining the increasing distance between the annealed approximation and the true threshold. Remarkably, the fact $\lambda/(1-\mu_c{}^{true}) < 1$ demonstrates that our temporal network is close to an annealed regime.

Lastly, if we consider $\lambda_A \neq \lambda_B$, equation (6.22) reads:

$$p(t+1) = (1-\mu)\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} p(t) + \lambda \begin{bmatrix} 0 & \epsilon\bar{Z} \\ \bar{Z}^\top & 0 \end{bmatrix} p(t), \tag{6.24}$$

thus yielding a threshold

$$\left(\frac{\mu}{\lambda}\right)_c = \rho \begin{bmatrix} 0 & \epsilon\bar{Z} \\ \bar{Z}^\top & 0 \end{bmatrix}. \tag{6.25}$$

We can conclude that the time-average network is a good approximation of the time evolving sequence. We will delve into the discovery of the reason behind this evidence in the following section.

6.4 Null models and the importance of the aggregated network

In the previous chapter we tested some well accounted models to reproduce the epidemic threshold of our temporal network. Only the time average approximation was able to get close to the true phase transition diagram, however we do not know the reason why this happens, besides the a posteriori finding that the threshold values of the SIS disease actually are in the region of the parameters $(\lambda, \mu, \epsilon = 1)$ in which the annealed approximation $(\lambda/(1-\mu) \ll 1)$ holds. We therefore want to inspect which is the

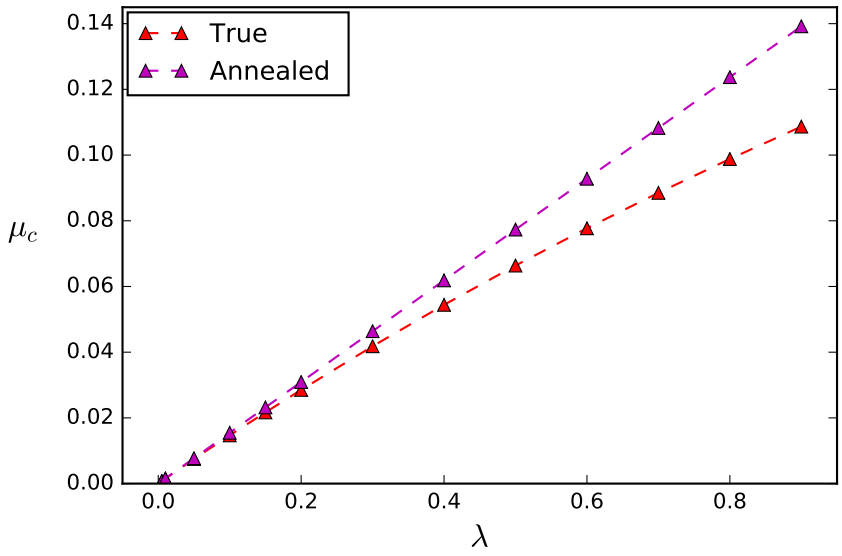

Figure 6.5: **Annealed threshold vs true threshold.**

main feature driving the epidemics on our system, and to do so we exploit the powerful numerical method of the *null models*.

We would like to analyse the impact of each specific temporal property of our network, disentangling it from the others. We recall briefly the description of the null models of Section 4.3 and summarize it in Table (6.1). Firstly, we have RESHUFFLE, that is a randomization of snapshots order. Secondly, RECONFIGURE consists in a random reassignment of contact timestamps: two contacts $(i, j; t_1)$, $(k, l; t_2)$ are randomly selected, and their time-stamp switched: $(i, j; t_2)$, $(k, l; t_1)$. Thirdly, in the ANONYMIZE null model we randomize the identity of the nodes, independently inside each snapshot. Finally, null model SWAP picks two random edges inside a snapshot and, if they have no common nodes and if the new links do not already exist, switches the nodes from one group of the bipartite network (cross-links) $(a1, b1; t)(a2, b2; t) \rightarrow (a1, b2; t)(a2, b1; t)$. One last null model is RANDOM TIMES (R.T.), that takes the complete list of all links occurred in the total period (with eventual repetitions) and re-assigns each of them to a randomly chosen time step, therefore not respecting the original daily activity of links. It basically preserves the aggregated network and destroys all temporal correlations. The random assignment actually results in a more uniform activity time-line.

Table 6.1: **Null Models for Temporal Networks**

Name	activity timeline	Aggregated network	Snapshot topology	microscopic time correlation	$k_{node}(t)$
RESHUFFLE	×	✓	✓	×	×
RECONFIGURE	✓	✓	×	×	×
RANDOM TIMES	×	✓	×	×	×
ANONYMIZE	✓	×	×	×	×
SWAP	✓	×	×	×	✓

In table, × means "the null model destroys the feature" and ✓ means "the null model preserves the feature". The markers with same colour emphasize which feature can be studied by contrasting the relative null models. We test the effect of each null model on

the epidemic spreading, by evaluating the threshold on each null model. The results are presented in Fig.(6.6). The shaded areas represents the values that lies between the 25% and the 75% percentiles of the 300 realizations we performed for each null model.

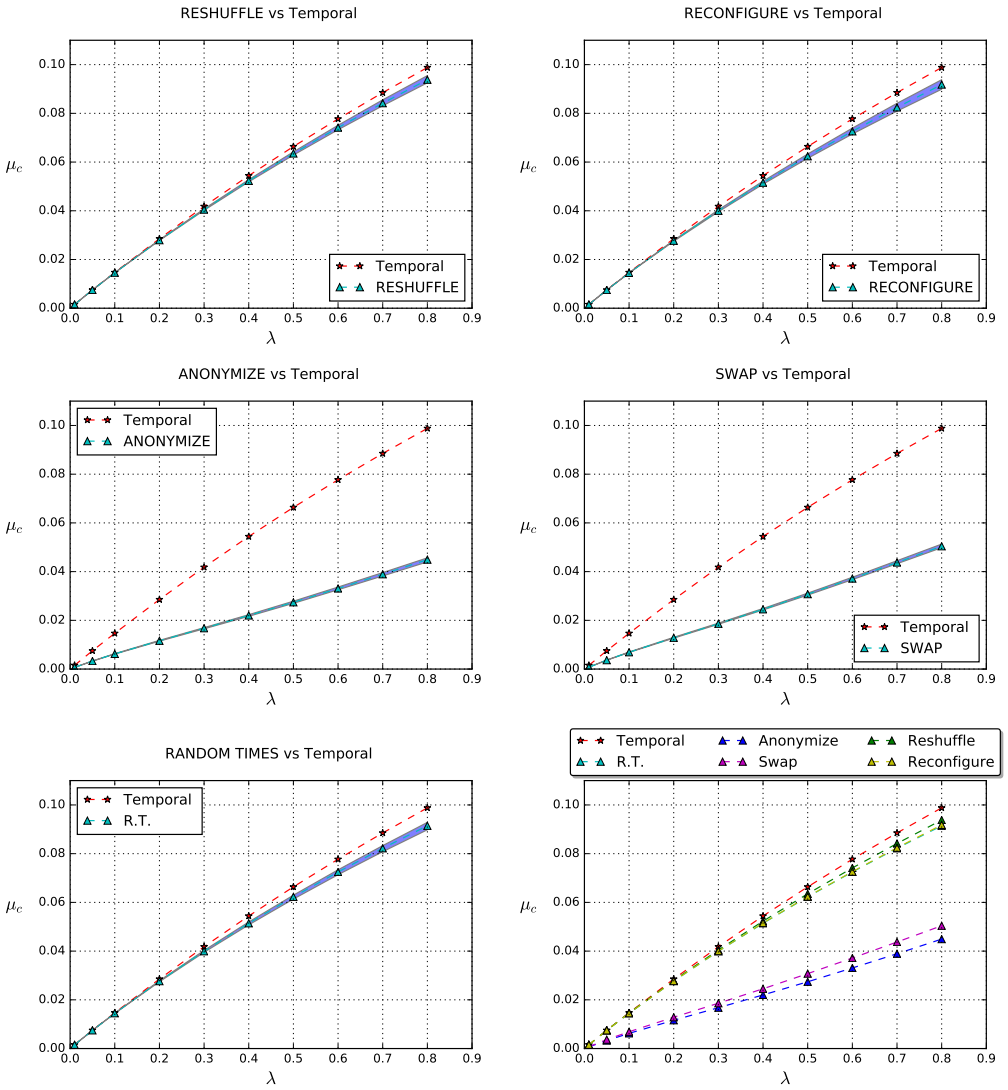

Figure 6.6: **Threshold computed on Null Models**

Before drawing conclusions from the plots presented in Fig.(6.6), we recall that the comparison between two models that differ for a single broken feature informs on the role of that feature on the threshold. We highlight that we cannot tell the role of the preserved feature, but rather see how strong is the impact of the destroyed one. It is worth saying that it can be hard to isolate one specific characteristic of the original temporal network, because many topological/temporal features are reciprocally "entangled", meaning that removing one often modifies another. This implies that evaluating the importance of a single property can be sometimes impossible; other times we can do it indirectly, by comparing two null models that differ by a that specific feature. In this sense: comparing R.TIMES with RESHUFFLE enlightens on the influence of snapshot topology (red markers in the Table); comparing R.TIMES with RECONFIGURE informs on the activity time line (orange markers) .
We can now observe Fig.(6.6): the first thing we notice is that all null models give a lower

value of the threshold μ_c, thus making the system safer. This shows that the original network is the most vulnerable. Going further in the comparisons, we find the following main results:

- The similarity of RESHUFFLE , RECONFIGURE and RANDOM TIMES with the original threshold tells that *destroying the activity time line and the single snapshot topology do not impact much on the spreading dynamics.* Notably, these features are related to temporal correlations. In particular, let us contrast R.T. vs RESHUFFLE and R.T. vs RECONFIGURE to test respectively the snapshot topology and the activity time line.

  R.T. breaks snapshot topology while RESHUFFLE preserves it: the results are almost equal, therefore the snapshot topology does not impact the spreading dynamics.

  R.T. breaks the activity time line while RECONFIGURE preserves it: the plots are very close, thus the original activity time line makes no difference respect to the spreading of the disease.

- ANONYMIZE and SWAP are the worst performing null models respect to the original network, meaning that breaking the aggregated network has a strong impact on the disease dynamics. ANONY-
  MIZE is like SWAP, with the additional breaking of the degree distribution: this further lowers the threshold value.

To be more precise in our conclusions, we compare RECONFIGURE and ANONYMIZE and their marked difference states that the relevant feature of the network with respect to a SIS dynamics is the aggregated graph. This finds a sustain in the goodness of the time-average $\bar{A}$ approximation for the temporal network. In fact $\bar{A}$ is the aggregated network with the weight of each link divided by the period T.
Notably, our finding is confirmed by a different analysis carried out by Rocha and Delvenne in *Diffusion on networked systems is a question of time or structure* [49]. In this work the time-dependence and the topology-dependence of the diffusion of a random walker on an empirical network is evaluated both numerically and analytically. The authors test their model on some empirical temporal networks, among which there is ours. The authors find that for our network is topology to be the responsible of the time scale of the stationary state of the R.W. . [3]

[3] In the work, topology is encoded by the spectral radius of the aggregated network, while the time-dependence in evaluated via the inter-contact waiting time distribution.

Chapter 7

Stochastic correction to the average temporal network

In the previous chapter, via Fig.(6.5), it has been numerically shown that the epidemic threshold for the sequence $\{A(t)\}$ does not differ much from the so-called annealed approximation, where each snapshot $A(t)$ is approximated with the temporal mean $\bar{A} = \sum_t A(t)/T$. Therefore we try to find a better esteem of the threshold by exploiting $\bar{A}$, while also including the fluctuations of real $A(t)$ with respect to $\bar{A}$. We will encode them into a stochastic noise matrix.

Let us start from recalling that the SIS spreading on a temporal network is described by a system of equations:

$$p_i(t+1) = 1 - [1 - (1-\mu)p_i(t)] \prod_j \left[1 - \lambda A_{ji}(t) p_j(t)\right]. \tag{7.1}$$

for $i = 1, .., N$ with N the number of nodes in the network. When $p_i \sim 0$, equation (7.1) can be linearized in p_i as:

$$p_i(t+1) = (1-\mu)p_i(t) + \lambda \sum_j A_{ji}(t) p_j(t) + o(p^2). \tag{7.2}$$

Considering vector formalism, Eq.(7.2) reads:

$$p(t+1) = [(1-\mu)I + \lambda A(t)]\, p(t). \tag{7.3}$$

with $p(t) \in \mathbb{R}^N$ and each A(t) is a $N \times N$ real valued non negative and symmetric matrix ($A^\dagger(t) = A(t)$), encoding the contacts between individuals. Being our population formed by men (N_1) and women (N_2, with $N = N_1 + N_2$), and provided that interactions can happen only between a man and a woman, we recall that we have arranged the contact matrix to take a convenient block form. If the vector p is defined to encode for men in the first N_1 components and for women in the last N_2, each matrix $A(t)$ takes the form:

$$A(t) = \begin{bmatrix} 0 & Z(t) \\ Z^\dagger(t) & 0 \end{bmatrix} \tag{7.4}$$

with the rectangular $N_1 \times N_2$ adjacency matrix Z of components $(Z)_{ij}$ where indices span respectively: $i \in \mathcal{I}_1 = \{1, .., N_1\}$ and $i \in \mathcal{I}_2 = \{N_1 + 1, .., N\}$. In the following we will address equivalently to this as *block-form* or *bipartite*. Moreover, the function A(t) is considered periodic $A(t+T) = A(t)$ considering that the period T is representative of a stable contact dynamics.

With an algebraic manipulation, we include the dependence on matrix $\bar{A}$, by re-writing Eq.(7.3) as follows:

$$\begin{aligned} p(t+1) &= (1-\mu)p(t) + \lambda\left(A(t) + \bar{A} - \bar{A}\right)p(t) \\ &= (1-\mu)p(t) + \lambda\bar{A}p(t) + \lambda\left(A(t) - \bar{A}\right)p(t) \\ &= \left[(1-\mu)I + \lambda\bar{A}\right]p(t) + \lambda\,\delta A(t)p(t) \\ p(t+1) &= [I+M]\,p(t) + \lambda\,\delta A(t)p(t) \end{aligned} \tag{7.5}$$

with the definition of the deterministic constant matrix :

$$M = -\mu I + \lambda\bar{A}, \tag{7.6}$$

In square brackets we have isolated a time-independent and deterministic evolution matrix, while in the right-most side there is a time dependent matrix $\delta A(t)$ that we call fluctuation matrix, encoding the time heterogeneous behaviour of contacts. We recall that $\delta A(t) = \delta A^{\dagger}(t)$, being the network undirected; moreover $\delta A_{ii}(t) = 0$, because for every t is clearly $A_{ii}(t) = 0$; lastly $\delta A(t)$ has the same block form as $A(t)$.
We make now the strong hypothesis that the fluctuation matrix can be considered as a *noise matrix*, with each entry being an independent Wiener process of mean 0 and variance σ^2_{ij}. Of course, being the fluctuation matrix non zero only in those entries where $\bar{A}$ is non zero, the independent Wiener processes are a total of N_E, that is the number of occupied links in matrix $\bar{Z}$. Therefore, the noise matrix is defined as having entries $W_{ij}(t) \sim \mathcal{N}(0, \sigma^2_{ij})$ with $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_2$ when $\bar{A}_{ij} > 0$, otherwise entries are 0. We require W to be symmetric as $\delta A(t)$ is: a "noise" contact happening on link (i,j) must be the same on link (j,i). Thus, Eq.(7.5) can be written as a *stochastic linear vector difference equation*:

$$p(t+1) = [(1-\mu)I + \lambda\bar{A}]p(t) + \lambda W(t)p(t) \tag{7.7}$$

7.1 Geometric Brownian motion, one dimension

We notice that Equation (7.7) is the vectorial and discrete time version of the well known geometric brownian motion. The Geometric Brownian motion $\mathbf{X} = \{X_t : t \in [0,\infty), X_t \in \mathbb{R}\}$ satisfies the stochastic differential equation

$$dX_t = \gamma X_t + \sigma X_t\, d\xi_t\,, \tag{7.8}$$

in the Ito prescription, where $\blacksquare = \{\xi_t : t \in [0,\infty)\}$ is a standard Brownian motion and $\gamma \in \mathbf{R}$ and $\sigma \in (0,\infty)$. The solution of Eq.(7.8), with initial condition $X_0 = x_0$, is given by:

$$X_t = x_0 \exp\left[\left(\gamma - \frac{\sigma^2}{2}\right)t + \sigma\xi_t\right], \qquad t \in [0,+\infty). \tag{7.9}$$

The stochastic process $\mathbf{X}$ is called Geometric Brownian motion with drift parameter γ and volatility parameter σ. The stochastic process at the exponent is itself a Brownian motion. We notice that the GBM is never negative, thus it will be in accord with our interpretation of X as a probability of infection. The distribution of X_t is lognormal, with parameters $\left(\gamma - \frac{\sigma^2}{2}\right)t$ and $\sigma\sqrt{t}$. The probability density function f_t is given by:

$$f_t(x) = \frac{1}{\sigma x\sqrt{2\pi t}} \exp\left(-\frac{\left[\ln(x) - (\gamma - \sigma^2/2)t\right]^2}{2\sigma^2 t}\right), \qquad x \in (0,\infty) \tag{7.10}$$

for each $t \in (0,\infty)$. The expectation values of moments are $\mathbf{E}\left(X_t{}^n\right) = e^{nt\left[\gamma-\frac{\sigma^2}{2}(n-1)\right]}$, so :

$$\begin{aligned}\mathbf{E}\left(X_t\right) &= e^{\gamma t}\\ \mathrm{Var}\left(X_t\right) &= e^{2\gamma t}\left(e^{\sigma^2 t}-1\right)\end{aligned} \tag{7.11}$$

and we note that the mean function $m(t) = \mathbf{E}\left(X_t\right) = e^{\gamma t}$ for $t \in (0,\infty)$ satisfies the deterministic part of the stochastic differential equation above (7.8).
An important property, that is especially relevant to our study, is the fact that the parameter $(\gamma - \sigma^2/2)$ establishes the asymptotic behaviour of the geometric Brownian motion. In fact:

$$\begin{aligned}&\text{If}\,\gamma > \sigma^2/2 \quad \text{then}\, X_t \to \infty \,\text{as}\, t \to \infty;\\ &\text{If}\,\gamma < \sigma^2/2 \quad \text{then}\, X_t \to 0 \,\text{as}\, t \to \infty;\\ &\text{If}\,\gamma = \sigma^2/2 \quad \text{then}\, X_t \,\text{has no limit as}\, t \to \infty.\end{aligned} \tag{7.12}$$

This follows from the law of the iterative logarithm: asymptotically, the term $(\gamma - \sigma^2/2)t$ dominates the term $\sigma\xi_t$ as $t \to \infty$. In fact the precise growth rate of a Brownian motion ξ_t is given by

$$\limsup_{t\to\infty} \frac{\xi_t}{\sqrt{2t\ln\ln t}} = 1,$$

called law of iterated logarithm [50]. This result, applied to Eq.(7.9), gives the asymptotic behaviour. We can understand the long time limit by looking at the probability density function on Eq. 7.10): a log-normal distribution can be strongly skewed, with the mode is quite smaller than the mean. More precisely, $\mathrm{mode}(X_t) = \exp\left(\gamma - \frac{3\sigma^2}{2}\right) t$ and it is therefore clear that the mode approaches 0 when $(\gamma - \sigma^2/2) < 0$, while the mean follows a very different course $\exp(\gamma t)$. Thus the mean value is not very informative on the long time limit of the single stochastic process.
Therefore, it is the sign of $(\gamma - \sigma^2/2)$ to define the threshold condition for a GBM: discriminating between the zero final state or the infinite final state. This parameter is the one we find at the exponent of the "deterministic part" of the solution in Eq.(7.9).

We can recover the same parameter $(\gamma - \sigma^2/2)$ if we consider the associated Fokker Plank equation:

$$\frac{dP(x,t)}{dt} = \frac{d}{dx}\left[-\gamma x P(x,t) + \frac{\sigma^2}{2}\frac{d}{dx}\left(x^2 P(x,t)\right)\right] = \frac{d}{dx}J(x,t). \tag{7.13}$$

The stationary probability distribution $P^*(x)$ that sets to zero the current J takes the form $P^*(x) \sim x^{2\left(\frac{\gamma}{\sigma^2}-1\right)}$. To account for its normalization, we firstly recall that in our interpretation of the random variable. To us X_t is a probability, therefore $X_t \in [0,1]$. Thus, we take Eq.(7.8) to be valid for small X and therefore we see that the previous expression for $P^*(x)$ may have problem of normalization only in the neighbourhood of 0. To ensure its normalizability, the stationary probability distribution finally reads:

$$P^*(x) = \begin{cases}\delta(x) & \text{if}\,\gamma \leqslant \sigma^2/2\\ \sim x^{2\left(\frac{\gamma}{\sigma^2}-1\right)} & \text{if}\,\gamma > \sigma^2/2\end{cases} \tag{7.14}$$

for $x \in [0,1]$. This results confirms the threshold role of $(\gamma - \sigma^2/2) = 0$.

7.2 Multi-dimensional geometric Brownian motion

We would like to rephrase the above result for the asymptotic behaviour of the one dimensional GBM to our N dimensional problem. First of all, we need to write the Ito Eq.(7.7) as a vector stochastic differential equation with independent gaussian noises. Then we can write it in the Stratonovich prescription, so that the usual rules of differential calculus are restored and a formal solution in the sense on Eq.(7.9) can be written. By inspecting the "deterministic part" of the exponent, we can guess we find the threshold coefficient for the asymptotic stability of the zero vector solution.
Let us consider the continuous time analogue of Eq.(7.7):

$$dp = Mp(t)\,dt + \lambda W(t)p(t), \tag{7.15}$$

with the symmetric noise $W_{ij}(t) \sim \mathcal{N}(0, \sigma_{ij}^2\,dt)$ for the (i,j) so that $\bar{A}_{ij} > 0$, and where the derivative is understood in the Ito prescription. We need do write it in the classical form

$$dx_i = A_{ij}x_j\,dt + B_{ij}^{\nu}x_j\xi^{\nu}(t) \qquad \textit{Ito} \tag{7.16}$$

where $\xi^{\nu}(t)$ are independent Wiener processes with $\nu = 1, .., M$ and $i = 1, .., N$ [51]. Repeated indices are understood in Einstein notation, i.e. as summed. With some algebraic manipulation, we can rewrite Eq.(7.15) as:

$$\boxed{dp_i = M_{ij}p_j\,dt + D_{ij}^{\nu}p_j\xi^{\nu}(t)} \tag{7.17}$$

with

$$\boxed{\begin{array}{l}\nu = (k,l) \qquad k \in \mathcal{I}_1\,,\ l \in \mathcal{I}_2 \,|\, \delta A_{kl}(t) \neq 0 \\ D^{(k,l)}{}_{ij} = \lambda\sigma_{ij}\left(\delta_{k,i}\delta_{l,j} + \delta_{k,j}\delta_{l,i}\right) \\ \xi^{(k,l)}(t) \sim \mathcal{N}(0, dt)\end{array}} \tag{7.18}$$

Now, we want to translate this into the the Stratonovich prescription, because we aim to integrate, at least formally, the vector Equation (7.17) using the usual rules of calculus. The Stratonovich formalism for differential equation

$$dx_i = \hat{A}_{ij}x_j\,dt + \hat{B}_{ij}^{\nu}x_j \circ \xi^{\nu}(t) \qquad \textit{Stratonovich} \tag{7.19}$$

can be derived from Ito's Eq.(7.16) using the following relations:

$$\begin{cases}\hat{A} = A - \frac{1}{2}\sum_{\nu}(B^{\nu})^2 \\ \hat{B^{\nu}} = B^{\nu}\end{cases} \tag{7.20}$$

We apply these equations to our case, and explicitly compute $\hat{M} = M - \frac{1}{2}\sum_{\nu}(D^{\nu})^2$. We consider for simplicity that $\sigma_{ij}^2 = \sigma^2$ is unique : we will see this is a necessary approximation due to the low statistics of the data from which we extract the σ value. Thus we compute

$$\frac{1}{2}\sum_{\nu}(D^{\nu})^2 = \frac{1}{2}\lambda^2\sigma^2 D_k, \tag{7.21}$$

where matrix D_k is the diagonal square matrix with entries :

$$d_{ii} = \sum_{j\,|\,\bar{A}_{ij}\neq 0} 1 = k_i(\bar{A}), \tag{7.22}$$

where $k_i(\bar{A})$ is the degree of node i in the network associated to $\bar{A}$. Therefore the Stratonovich matrix $\hat{M}$ reads:

$$\boxed{\hat{M} = -\mu I + \lambda\bar{A} - \frac{1}{2}\lambda^2\sigma^2 D_k.} \tag{7.23}$$

The evolution equation of vector p in the Stratonovich sense is :

$$dp_i = \hat{M}_{ij}p_j\,dt + D^{\nu}_{ij}p_j \circ \xi^{\nu}(t). \tag{7.24}$$

We make now a guess: that the analogy with the one dimensional geometric Brownian motion holds, so that the condition for which $p(t) \to 0$ as $t \to \infty$ is simply that $Re\left[\lambda(\hat{M})\right] < 0$, $\lambda(\hat{M})$ being an eigenvalue of $\hat{M}$. Rephrasing it in the discrete time formalism, we guess that the threshold for the recovery probability μ_c is the one such that :

$$\boxed{\rho\left(1 - \mu_c + \lambda\bar{A} - \frac{1}{2}\lambda^2\sigma^2 D_k\right) = 1.} \tag{7.25}$$

where ρ is the usual spectral radius of the matrix.
To have an idea of what this correction roughly means, we can imagine to have a symmetric matrix $\bar{A}$, with the only zero entries being the diagonals'. In this simple case, we have $D_k = (N-1)I$, therefore the threshold is simply:

$$\mu_c^{stoch} = \mu_c^{Annealed} - \frac{1}{2}\lambda^2\sigma^2(N-1), \tag{7.26}$$

where $\mu_c^{Annealed} = \lambda\rho(\bar{A})$ [1]. This means that the annealed approximation is corrected with a term $(\lambda^2\sigma^2/2)$ for every of the (N-1) contacts a single node performs. So here we have shown that the stochastic correction tends to lower the value of the annealed threshold $\mu_c^{Annealed}$, thus moving in the wished direction towards the temporal phase-transition line. Epidemiologically, the stochastic system is safer (i.e. lower μ_c) than the annealed approximation.

7.2.1 Numerical estimation of the stochastic correction

Now we want to test the new threshold found in Eq.(7.25) and see if it ameliorates the annealed approximation, with respect to the true temporal threshold.

What we need to extract from the data set is a value for σ^2, the variance of the noise matrix W(t). We recall that by means of W(t) we want to approximate the behaviour of $\delta A(t) = A(t) - \bar{A}$. The first main guess on this matrix has been that all entries are uncorrelated at every time. We can test the reliability of this hypothesis from the computation of $\langle \delta A_{ij}(t)\delta A_{lm}(t')\rangle$ and looking for correlations between different links at the same time (uncorrelation of noises) and between same link at different time. As usually defined, the average should be computed on many realizations $\delta A^{(r)}_{ij}(t)$ of the process

$$\langle \delta A_{ij}(t)\delta A_{lm}(t')\rangle = \frac{\sum_{r=1}^{R} A^{(r)}_{ij}(t)\delta A^{(r)}_{lm}(t')}{R}. \tag{7.27}$$

[1] To be precise, we can extract $1-\mu$ from the spectral radius only if $\lambda\bar{A} - \frac{1}{2}\lambda^2\sigma^2 D_k$ is a non-negative matrix.

However, we do not dispose of such data, having only one realization of the contacts dynamics $\{A^{(1)}(t)\}$. We can overcome this problem by guessing that $\langle \delta A_{ij}(t)\delta A_{lm}(t')\rangle \equiv G(|t-t'|, ij, lm)$, therefore ensemble average Eq.(7.27) can be rewritten as:

$$G(|t-t'|, ij, lm) = \langle \delta A_{ij}(t)\delta A_{lm}(t' = t \pm \tau)\rangle = \frac{\sum_{r=1}^{R} A_{ij}^{(r)}(t)\delta A_{lm}^{(r)}(t+\tau) + A_{ij}^{(r)}(t)\delta A_{lm}^{(r)}(t-\tau)}{2R}. \tag{7.28}$$

This function should be the same (i.e independent) for every t:

$$\begin{aligned} G(\tau, ij, lm) &= \langle \delta A_{ij}(t_1)\delta A_{lm}(t_1 \pm \tau)\rangle = \langle \delta A_{ij}(t_2)\delta A_{lm}(t_2 \pm \tau)\rangle \\ &= \frac{1}{(T-\tau)} \sum_{t=1}^{T-\tau} \frac{\langle \delta A_{ij}(t)\delta A_{lm}(t+\tau)\rangle + \langle \delta A_{lm}(t)\delta A_{ij}(t+\tau)\rangle}{2}. \end{aligned} \tag{7.29}$$

In conclusion, we can exploit the above equation to evaluate $G(\tau, ij, lm)$ by using one realization $\{A^{(1)}(t)\}$, discarding the ensemble average in favour of the temporal average. When all ensemble averages equal the corresponding time averages the signal is called ergodic.

$$G_{ij,lm}(\tau)_{data} = \sum_{t=1}^{T-\tau} \frac{\delta A_{ij}(t)\delta A_{lm}(t+\tau) + \delta A_{lm}(t)\delta A_{ij}(t+\tau)}{2(T-\tau)} \tag{7.30}$$

With formula in Eq.(7.30) we have an approximation of $\langle \delta A_{ij}(t)\delta A_{lm}(t \pm \tau)\rangle$ that we can extract from our data. Numerical calculations show that the correlation function can reasonably be approximated with the following form:

$$G_{ij,lm}(\tau)_{data} = \sigma_{ij}^2\, \delta(\tau)\left(\delta_{(ij),(lm)} + \delta_{(ij),(ml)}\right), \tag{7.31}$$

for ij and lm so that $\bar{A}_{ij} > 0$ and $\bar{A}_{lm} > 0$. We can understand it by looking at Figures (7.2) and (7.1).

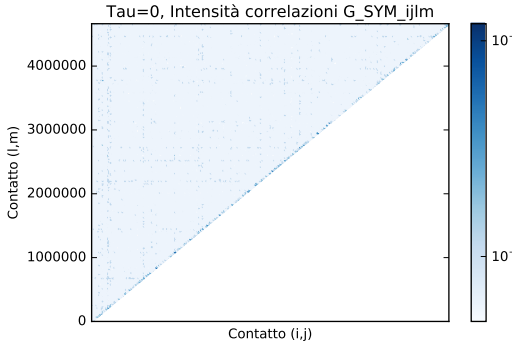

Figure 7.1: $G_{ij-lm}(\tau = 0day)$

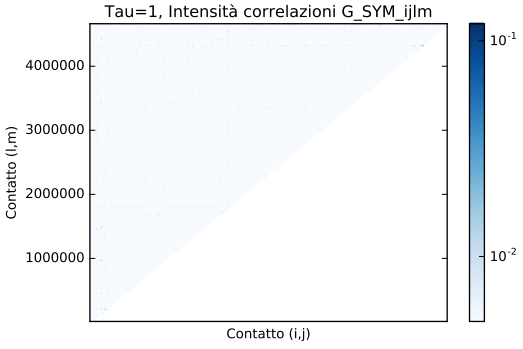

Figure 7.2: $G_{ij-lm}(\tau = 1)$

The x and y axis represent the links (i,j) and (l,m) for which the value $G_{ij,lm}(\tau)$ is computed; the figure on the left is for $\tau = 0$ and the other on the right is for $\tau = 1$. The intensity of the colour is log-proportional to the value of the correlation. We see that a strong colouring is found only along the diagonal of the plot with $\tau = 0$. This explains the form of function in Eq.(7.31), and its meaning: at each "noise process" $\delta A_{ij}(t)$ is uncorrelated with any other happening at the same time in another link, or in the same link but at different time (here we see for $\tau = 1$ day). Therefore, comparing Eq.(7.30) and

Eq.(7.31), it is easy to see that σ^2_{ij} is found by computing:

$$\sigma^2_{ij\,data} = G_{ij,ij}(0)_{data} = \frac{\sum_{t=1}^{T} \delta A^2_{ij}(t)}{T}. \tag{7.32}$$

More explicitly :

$$\sigma^2_{ij\,data} = \frac{\sum_t A^2_{ij}(t)}{T} - \left(\frac{\sum_t A_{ij}(t)}{T}\right)^2. \tag{7.33}$$

The specific data set has an extremely poor statistics for a single link (i,j): many of the links activate only one time in the whole period T, thus resulting in the impossibility of assessing for the variance σ^2_{ij} of such a sparse noise. Therefore, to find a meaningful value we need make the hypothesis that each noise is statistically equivalent, so that we compute an average noise variance as the mean of all σ^2_{ij}:

$$\sigma^2{}_{data} = \frac{\sum_{i\in\mathcal{I}_1}\sum_{j\in\mathcal{I}_2}\sigma^2_{ij}}{N_E} \equiv \frac{\Sigma^2}{N_E}. \tag{7.34}$$

where $\Sigma^2 \equiv \sum_{i\in\mathcal{I}_1}\sum_{j\in\mathcal{I}_2}\sigma^2_{ij}$ and $N_E = 6401$ is the number of edges that are found active in the period. The value obtained from the data is $\Sigma^2_{data} \simeq 44.837266$, giving the mean variance equal to:

$$\sigma^2{}_{data} \simeq 0.0070\,. \tag{7.35}$$

In conclusion, what we assume is that the fluctuation matrix $\delta A(t)$ is a block matrix with N_E independent Wiener processes of mean 0 (by definition) and variance $\sigma^2{}_{data}$, and $W_{ij}(t) = W_{ji}(t)$ is symmetric.

Now we need to compute numerically if the stochastic correction is sensible and to what extent. In Fig.(7.3) we report the stochastic threshold computed via formula Eq.(7.25) , i.e. with the diagonal degree matrix D_k. We notice that the stochastic correction goes much closer to the temporal that the annealed approximation, although being a little lower.

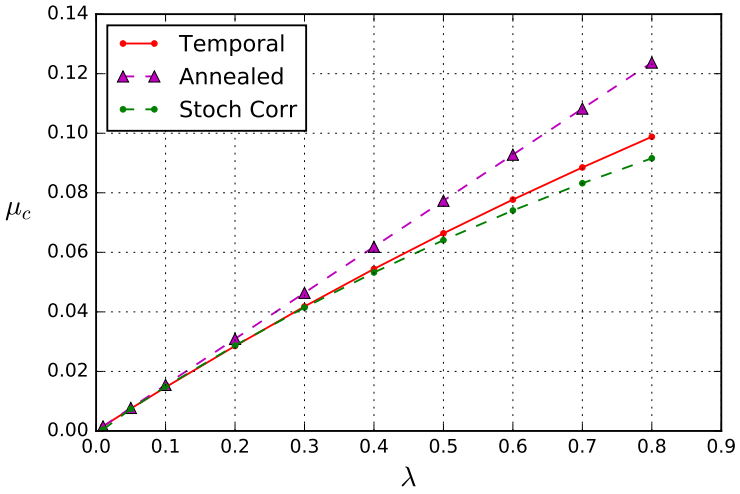

Figure 7.3: **Stochastic correction of the annealed approximation**.

We can contrast the result with the *Reshuffle null-model*. This null-model consists in reordering in a completely random fashion the sequence of snapshots $\{A(t)\}$, therefore

maintaining the average matrix $\bar{A}$, the average number of active links per day but loosing all time-correlation among link activation at different times. We show the relation between stochastic correction and Reshuffle in Figure (7.4).[2]

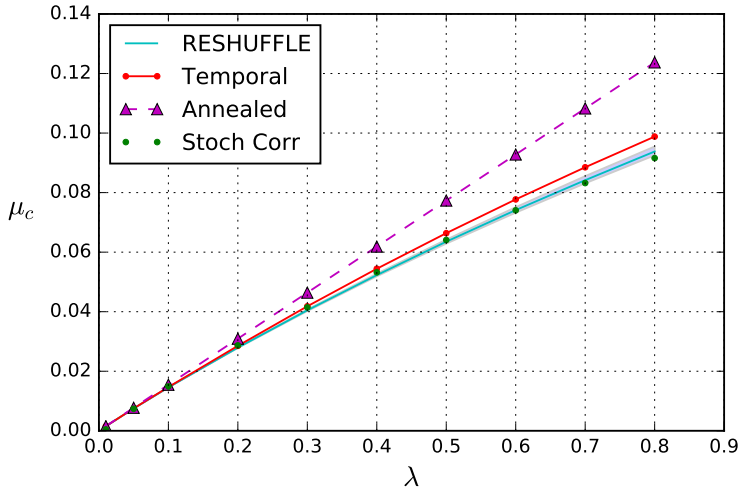

Figure 7.4: **Stochastic correction contrasted with the Reshuffle null model.**

What we would expect is a close relation of the stochastic critical line with the reshuffled, being all consequential information lost in the stochastic approximation. In fact, this is what we find in Figure (7.4).
The result we find really seems to capture the main features of the temporal network. However, we do not rely too much on the closeness of the stochastic curve with the true and reshuffled: that could be a on some extent a lucky coincidence between the evaluated σ^2_{data} and the needed distance to cover between the annealed and the true value. Moreover, we know that the perturbations of links are far from Gaussian: they may resemble more to shot noise or Poisson like activation but in any case we are very limited by the low statistic of the data. Nevertheless, what we can learn from this approach is the qualitative impact of randomness in the activation of links on the epidemic threshold. We can interpret the variance of the noise as a degree of heterogeneity in the activation of links in time: the bigger σ^2, the more oscillating is the activity time lime of a link with respect to the annealed case, in which the link is always active at each time step with same "probability" $\bar{A}_{ij}$. The heterogeneity in the activation of links is known [47] to be a factor that slows down the diffusion of a disease in a temporal network ($\mu_c \downarrow$). In fact, this implies that the number of available paths are less with respect to the annealed case, because at each time step some links decrease their activation probability. Our analytic result points in the same direction, giving a quantitative esteem of this lowering of the threshold proportional to the noise variance.

[2] We have performed 300 reshuffles and the plot is the median of those iterations, with the error evaluated via the 25th and the 75th percentiles values of μ_c.

Chapter 8

Conclusions

In this work we have addressed the problem of assessing the vulnerability for a system of interacting individuals targeted with a disease. The disease has been modelled using the SIS compartmental model, for which people can be either susceptible S or infectious I. The transition $S \rightarrow I$ takes place with probability λ, if a S. meets an I. Each I. can recover spontaneously with probability μ, giving the transition $I \rightarrow S$. For what concerns the population, we dispose of a data-set [34] regarding the interactions of a bipartite community, whose contacts change in time. We therefore exploited the formalism of temporal networks to encode this dynamics. We have focused on the so called *epidemic threshold*, e.t., in order to give a quantification of the vulnerability of the system. The epidemic threshold is the value of disease's parameters $\mu_c(\lambda)$ so that: for $\mu \leqslant \mu_c$ the number of infectious people in the asymptotic state is non zero, while for $\mu > \mu_c$ it is zero.

We have carried out the analysis of the threshold both numerically and analytically. In the numerical analysis, in order to fasten the computation of the e.t., we have exploited the findings of a recent work on temporal networks coupled with disease dynamics ([28],2015). This article shows that, when it comes to the determination of the epidemic threshold on a temporal network, the microscopic stochastic simulation of a disease-spreading can be replaced with a simpler and much faster computation of the spectral radius of a suitable matrix. This matrix encodes the coupled dynamics of disease and network: $P(t_0, T) = \prod_{t=t_0}^{T} [1 - \mu + \lambda A(T - t)]$, where the only additional guess is that the dynamics of the network is periodic of arbitrary period T. Using this approach, we have explored the threshold μ_c in various portion of the temporal data set and for varying λ. As this system regards a woman-man interaction relative to sexual contacts, we have allowed for men and women to have different λs probability of transmission of disease, as it is actually the case of most STDs. We have found that the impact of the two categories on the threshold is symmetric, with very good approximation, and that the present system is largely non vulnerable with respect to the most common STDs.

We have then addressed the question of which is the main temporal or topological feature that drives the disease on the network. By means of null models [4.3] we have broken various properties of the network in order to see their impact on the e.t. . We have concluded that the aggregated network is the structure one needs to preserve in order to capture the dynamic of the disease, even regardless of the time sequence of links activation.

To tackle the problem analytically, we have tested on our network three already developed analytical models that allow to write an explicit expression for the e.t. . We have generalized them to a bipartite network structure. Only one of these worked fine, namely the annealed approximation on the temporal network as an average network.

Lastly, we have tried to ameliorate the annealed approximation introducing a stochastic correction. We have written a N-dimensional Langevin equation for the vector state of nodes' infection probability $\vec{p}(t)$, when it is close to the disease-free state , i.e. $\vec{p}(t) \simeq \vec{0}$. The deterministic part of the equation is the same that describes the annealed regime. The stochastic part includes a Wiener noise-matrix, that aims to reproduce the fluctuations of the original temporal contact matrices respect to the annealed matrix. The resulting equation is a N-dimensional geometric Brownian motion. By extending the findings for the asymptotic state of a one dimensional GBM, we have found a correction to the epidemic threshold μ_c. The result shows good agreement with the "true" threshold.

For the future, it would be interesting to delve more into this stochastic approach we have addressed only in the last part of the work. For example, we would like to test its validity on other networks and on synthetic models. Moreover, some of the analysis we carried out aside suggest that it would be a good idea to test other types of noises, such as the Poissonian or shot noise. These are more resembling of the real behaviour of human contact dynamics with respect to gaussians'.

Appendices

Appendix A

The effect of stochasticity on the meaning of R_0

We have shown that the framework of deterministic differential equations allows a formal definition of epidemic threshold, in terms of basic reproductive number. We learn that for $R_0 > 1$ the disease will certainly turn epidemic, while when $R_0 < 1$ the growth is exponentially suppressed. We have already seen, however, that this kind of approach does not account for stochastic effects, and this qualitatively changes the definition itself of epidemic threshold. Let us examine this point through a Galton-Watson branching process. Let I_n be the number of (infectious) agents in generation n. Also let $X = 0, 1, 2..$ be the number of secondary cases each case generates. $P(X)$ be the probability distribution of such variable, with probability generating function $g(s)$. By definition of basic reproductive number, $E[X] = R_0$. Let $g_n(s)$ be the probability generating function of I_n. Clearly, the following relation holds

$$g_{n+1}(s) = g_n(g(s)). \tag{A.1}$$

From this relation we can compute the expectation values:

$$E[I_n] = g_n'(1) = g_{n-1}'(g(1))g'(1) = g_{n-1}'(1)g'(1) = E[I_{n-1}]R_0 = I_0 R_0^n. \tag{A.2}$$

Where I_0 is the initial number of infected. Analogously to differential equations, $R_0 = 1$ is a threshold value in the sense that discriminates the case when the average number of infected grows or decays exponentially. These are however only expectation values, and do not account for stochastic fluctuations. These fluctuations influence the extinction probability, i.e. the probability that the disease will at some point stop due to lack of infectious hosts.

Let us call $q_n = Prob[I_n = 0]$, i.e. the probability of extinction by the n-th generation. Clearly $q_n = g_n(0)$. For simplicity, we assume just one initial infected: $I_0 = 1$. Then the following relation holds $g_n(g(s)) = g(g_n(s))$, and we can use it to get a recurrence relation for q_n:

$$q_{n+1} = g(q_n). \tag{A.3}$$

We can find the limiting value q_∞ as the smallest positive root of

$$q = g(q) \tag{A.4}$$

$q = 1$ is always a solution of this equation. One can show that for $R_0 < 1$ this is also the smallest solution $q_\infty(R_0 < 1) = 1$. When the disease is below threshold, the disease

always goes extinct, as we already know from expectation values. In order to see what happens above threshold, we assume $X \sim \mathtt{Poisson}(R_0)$. Then the above equation reads

$$q = e^{R_0(q-1)} \tag{A.5}$$

When $R_0 > 1$, this equation has a solution in (0,1): this means that the extinction probability is non zero even above threshold. As a result, if you run your epidemic long enough, it will always goes extinct. Therefore the epidemic threshold in stochastic models formally does not discriminate between extinction and no extinction, but between a phase where the disease goes extinct exponentially fast ($R_0 < 1$), and a regime where the time to extinction is long, and increases with R_0 (above threshold). In practice the extinction time often becomes so long, that can be considered as infinite, recovering the meaning of the threshold found with differential equations. Formally, however, the active endemic state is always metastable, as fluctuations can always bring it to the absorbing state of no infected, which the system cannot then leave. These fluctuations become extremely suppressed only when R_0 is high, and/or the population is large.

Appendix B

Tensor representation of a temporal network

The tensor formulation of multilayer networks has been put forward in [27]. Here we restrict it to temporal networks. Let us assume a network of N nodes. We start by assuming it does not evolve in time. Let V be the vector space spanned by the basis $\{e_1, e_2, \ldots, e_N\}$ which is in one-to-one correspondence with the nodes in the network. In particular, $V \simeq \mathbb{R}^N$. We consider the dual space V^* as the vector space spanned by the linear maps $\{e_1^*, e_2^*, \ldots, e*_N\}$. They are defined as follows:

$$e_1^* : V \to \mathbb{R} \quad \text{and} \quad e_i^*(e_j) = \delta_{ij}. \tag{B.1}$$

This is a traditional definition of dual space of a vector space. Then, we can define the adjacency matrix A as a multi-linear map (i.e., a tensor) $A : V^* \otimes V \to \mathbb{R}$,defined as:

$$A = \sum_i \sum_j A_{ij} e_i^* \otimes e_j, \tag{B.2}$$

with component $A_{ij} = A(e_i^*, e_j)$ encoding the value of link i-j. A thus becomes a rank 2 tensor.
The generalization to the temporal case is now straightforward. Let us assume we have T snapshots. We then consider the space $W \simeq \mathbb{R}^T$ spanned by the basis $\{f_1, f_2, \ldots, f_T\}$ which is in one-to-one correspondence with the snapshots. Then we define the adjacency tensor $\mathbf{A}$ as a rank 4 tensor

$$\mathbf{A} = \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ts,ij} f_t^* \otimes f_s \otimes e_i^* \otimes e_j \tag{B.3}$$

with component $\mathbf{A}_{ts,ij} = \mathbf{A}(f_t^*, f_s, e_i^*, e_j)$ encoding the value of link from node i at time t, to node j at time s. This allows for both intra-layer (same time) and inter-layer (different times) links.

For a matter of convenience of representation, we will use high indices for time and low for nodes, therefore $\mathbf{A}_{ij}^{ts}$.

Appendix C

Maximum eigenvalue of M block matrix

Computing the eigenvalues of $M^{\dagger}$ in Eq.(4.12) means solving the equation

$$\det\left(xI - M^{\dagger}\right) = 0, \tag{C.1}$$

where the determinant is computed on the $\mathbb{R}^{NT}$ space ($\det_{NT}$). Given that $xI - M^{\dagger}$ is composed of T^2 blocks of size $N \times N$, we can use the findings in Ref. [52] to reduce the dimensionality of the problem, i.e., $\det_{NT} \rightarrow \det_{N}$. Moreover, given that several blocks of $xI - M^{\dagger}$ are zero, the general result simplifies to

$$\det_{NT}\left(xI - M^{\dagger}\right) = (-1)^{NT}\det_{N}\left(xI - P\right), \tag{C.2}$$

where P is the Infection Propagator matrix of Eq.(4.22).

Appendix D

Theory of Linear Difference Equation

In the present work we make an extensive use of equations describing the evolution of a certain quantity over time, with time considered *discrete*. This is the best choice when we deal with data-driven equations, with data collected in discrete time (e.g. each minute, hour, day, etc.). The mathematical representation of such dynamics is called a *difference equation*, For example, if a certain population has discrete generations n, the size of the $(n+1)$st generation $x(n+1)$ is a function of the n-th generation $x(n)$. This relation expresses itself in the difference equation

$$x(n+1) = f(x(n)). \tag{D.1}$$

If the function f in Eq.(D.1) is replaced by a function g of both x and time n, that is $g : \mathbb{Z}_+ \times \mathbb{R} \to \mathbb{R}$, where $\mathbb{Z}_+$ is the set of non negative integers and $\mathbb{R}$ is the set of real numbers, then we have

$$x(n+1) = g(n, x(n)) \tag{D.2}$$

Equation (D.2) is called non autonomous or time-variant, whereas (D.1) is called autonomous or time-invariant. If an initial condition $x(n_0) = x_0$ is given, then for $n \geqslant n_0$ there is a unique solution $x(n) = x(n, n_0, x_0)$ of D.2) such that $x(n_0, n_0, x_0) = x_0$, shown easily by iteration.

Linear First-Order Difference Equations The simplest special cases of (D.1) and (D.2) are linear equations. A typical linear homogeneous first-order equation is given by:

$$x(n+1) = a(n)x(n)\,, \qquad x(n_0) = x_0\,,\ n \geqslant n_0 \geqslant 0, \tag{D.3}$$

and the associated non homogeneous equation is given by

$$y(n+1) = a(n)y(n) + g(n)\,, \quad y(n_0) = y_0\,,\ n \geqslant n_0 \geqslant 0, \tag{D.4}$$

where in both equations it is assumed that $a(n) \neq 0$, and $a(n)$ and $g(n)$ are real-valued functions defined for $n \geqslant n_0 \geqslant 0$.

One may obtain the solution of (D.3) by a simple iteration, that leads to:

$$x(n) = \prod_{i=n_0}^{n-1} a(i)x_0 \tag{D.5}$$

Analogously, the unique solution of the non homogeneous (D.4) is:

$$y(n) = \prod_{i=n_0}^{n-1} a(i) y_0 + \sum_{r=n_0}^{n-1} \left[\prod_{i=r+1}^{n-1} a(i) \right] g(r) \tag{D.6}$$

An important special case of (D.4) is for a time independent:

$$y(n+1) = ay(n) + g(n), \quad y(0) = y_0 \tag{D.7}$$

Using formula (D.6) one may establish that

$$y(n) = a^n y_0 + \sum_{k=0}^{n-1} a^{n-k-1} g(k) \tag{D.8}$$

Notice that the solution of the non homogeneous *differential* equation

$$\frac{dy}{dt} = ay(t) + g(t), \quad y(0) = y_0 \tag{D.9}$$

is given by:

$$y(t) = e^{at} y_0 + \int_0^t e^{a(t-s)} g(s) ds \tag{D.10}$$

Thus the exponential e^{at} in differential equation corresponds to the exponential a^n and the integral $\int_0^t e^{a(t-s)} g(s) ds$ corresponds to the summation $\sum_{k=0}^{n-1} a^{n-k-1} g(k)$.

D.1 Linear system of difference equation

The above equations take values in $\mathbb{R}$. We can consider a generalization to $\mathbb{R}^k$, that is a *system of difference equations* where there are k dependent variables that evolve together. Systems of difference equations arise in various fields of scientific endeavour, like biology (the study of competitive species in population dynamics), physics (the study of the motions of interacting bodies), the study of control systems, neurology, and electricity, and in our thesis as well. More precisely, for our work we can restrict the study to the case of a *linear system of difference equation*.

Autonomous system

An *autonomous and homogeneous system* of difference of k linear equations is the following:

$$\begin{aligned} x_1(n+1) &= a_{11}x_1(n) + a_{12}x_2(n) + \cdots + a_{1k}x_k(n), \\ x_2(n+1) &= a_{21}x_1(n) + a_{22}x_2(n) + \cdots + a_{2k}x_k(n), \\ &\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\ x_k(n+1) &= a_{k1}x_1(n) + a_{k2}x_2(n) + \cdots + a_{kk}x_k(n), \end{aligned} \tag{D.11}$$

This system may be written in the vector form

$$\boxed{x(n+1) = Ax(n)} \tag{D.12}$$

where $x(n) = (x_1(n), x_2(n), ..., x_k(n))^\top \in \mathbb{R}^k$, and $A = (a_{ij})$ is a $k \times k$ real non singular matrix.

Here $\top$ indicates the transpose of a vector. System (D.11) is considered autonomous, or time-invariant, since the values of A are all constants. Non-autonomous, or time-variant, systems have $A(n)$.
If for some $n_0 \geqslant 0$, $x(n_0) = x_0$ is specified, then system (D.12) is called an initial value problem. With no loss of generality we take $n_0 = 0$. Then, by simple iteration (or by direct substitution into the equation), one may show that the solution is given by

$$\boxed{x(n, x_0) = A^n x_0} \tag{D.13}$$

where $A_0 = I$, the $k \times k$ identity matrix.
A fundamental result is that it is possible to express the solution of (D.12) in terms of eigenvalues and eigenvectors of A ([53], pg:135-137):

$$\boxed{x(n) = \sum_{i=1}^{k} c_i \lambda_i^n \xi_i} \tag{D.14}$$

where $c_i \in \mathbb{R}$ and

$$A\xi_i = \lambda_i \xi_i \qquad i = 1, \ldots, k. \tag{D.15}$$

This holds for A diagonalizable. A sufficient condition is to have k distinct eigenvalues. When the matrix A has repeated roots, then it is diagonalizable if it is normal, that is to say, if $A^\top A = AA^\top$. Examples of normal matrices are symmetric matrices, anti-symmetric and unitary.

We recall that a parallel theory exist for systems of linear differential equations. The solution of the initial value problem

$$\frac{dx}{dt} = Ax(t), \quad x(t_0) = x_0 \tag{D.16}$$

where $x \in \mathbb{R}^k$ and A is a $k \times k$ matrix, is given by:

$$x(t) = e^{A(t-t_0)} x_0 \tag{D.17}$$

with the usual definition of exponential matrix.

Non autonomous system

The *non autonomous* case of an homogeneous linear system is:

$$\boxed{x(n+1) = A(n)x(n)} \tag{D.18}$$

where $A(n) = (a_{ij}(n))$ is a $k \times k$ non singular [1] matrix function. The solution of the initial value problem $x(n_0) = x_0$ is proved to be unique for every choice of x_0 and takes the form:

$$\boxed{x(n) = \left[\prod_{i=n_0}^{n-1} A(i)\right] x_0,} \tag{D.19}$$

where the product is ordered as $\prod_{i=n_0}^{n-1} A(i) = A(n-1)A(n-2)\ldots A(n_0), \quad n > n_0$.

An important case is for $A(n)$ *periodic*, i.e. $A(n+N) = A(n)$ for a certain positive integer N, called the period. For long time behaviour, the periodic system behaves like an autonomous with matrix $A = \prod_{k=1}^{N} A(k)$. This will be exploited in the determination of the stability conditions for the zero solution.

[1] Matrix M is non singular when it in invertible i.e. $\det M \leqslant 0$.

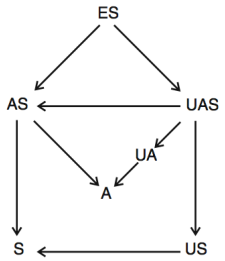

Figure D.1: Hierarchy of stability

D.2 Fixed point and stability

It is trivial to see that the vector 0 is a fixed point (or *equilibrium*) of system (D.12). Now we would like to see if the equilibrium is stable or not.

To do so, we first revise the definitions of equilibrium and stability.
A point $x^* \in \mathbb{R}^k$ is called an equilibrium point of (D.2) if $g(n, x^*) = x^*$ for all $n \geqslant n_0$.
We are now ready to introduce the various stability notions of the equilibrium point x^*:
Definition : The equilibrium point x^* of (D.2) is said to be:

- *Stable S* if given $\epsilon > 0$ and $n_0 \geqslant 0$ there exists $\delta = \delta(\epsilon, n_0)$ such that $\|x_0 - x\| < \delta$ implies $\|x(n, n_0, x_0) - x\| < \epsilon$ for all $n \geqslant n_0$, uniformly stable (US) if δ may be chosen independent of n_0, unstable if it is not stable.
- *Attractive A* if there exists $\mu(n_0)$ such that $\|x_0 - x\| < \mu$ implies $\lim_{n\to\infty} x(n, n_0, x_0) = x^*$, uniformly attracting (UA) if the choice of μ is independent of n_0.
- *Asymptotically Stable AS* if it is stable and attracting.
- *Exponentially Stable ES* if there exist $\delta > 0$,$M > 0$, and $\eta \in (0, 1)$ such that $\|x(n, n_0, x_0) - x\| \leqslant M\|x0 - x\|\eta^{n-n_0}$, whenever $\|x0 - x\| < \delta$.

The various definitions are placed in a hierarchic structure, exemplified if figure D.1:

When the system in autonomous and linear, Asymptotically stable and Exponentially are equivalent.

Now , we need to revise some linear algebra results that will help us in the stability analysis.

Theorem D.2.1 (The Jordan Canonical Form) *Any* $k \times k$ *matrix A is similar to a Jordan form given by the formula (D.20), where each* J_i *is an* $s_i \times s_i$ *matrix of the form (D.21), and* $\sum_{i=1}^{r} s_i = k$.

with

$$J = \text{diag}\,(J_1, J_2, \dots J_r)\,, \qquad 1 \leqslant r \leqslant k \tag{D.20}$$

and each J_i

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & 1 \\ 0 & 0 & \dots & \dots & \lambda_i \end{bmatrix} \tag{D.21}$$

is called a Jordan block, and λ_i are the eigenvalues of A.

We can now substantiate that the general solution of system (D.12) is

$$x(n) = A^n c = P J^n P^{-1} c, \tag{D.22}$$

where $P = [\xi_1, \dots \xi_k]$ the matrix with columns given by the eigenvectors of A.

The following corollary arises directly from an immediate consequence of the formula for J_i^n:

Corollary D.2.1.1 *Assuming that A is any* $k \times k$ *matrix, then* $\lim_{n\to\infty} A^n = 0$ *if and only if* $|\lambda| < 1$ *, for all eigenvalues* λ *of A.*

The importance of the preceding corollary lies in the fact that if $\lim_{n\to\infty} A^n = 0$, then

$$\lim_{n\to\infty} x(n) = \lim_{n\to\infty} A^n x(0) = 0 \tag{D.23}$$

that is if $|\lambda| < 1$ then all solutions $x(n)$ of (D.12) tend toward the zero vector as $n \to \infty$.

In the next theorem we summarize the main stability results for the linear autonomous systems (D.12).

Theorem D.2.2 *The following statements hold:*
(i) The zero solution of (D.12) is stable if and only if $\rho(A) \leqslant 1$ *and the eigenvalues of unit modulus are semi-simple.*
(ii) The zero solution of (D.12) is asymptotically stable if and only if $\rho(A) < 1$.

where ρ is the spectral radius of the matrix: $\rho(A) = \max\{|\lambda|, \lambda \text{ is eigenvalue of } A\}$. A similar stability condition holds also for non autonomous linear systems (D.18) that are also periodic. The asymptotic stability is recovered when

$$\rho(C) < 1\,, \qquad C = \prod_{i=1}^{N} A(N-i) \tag{D.24}$$

with N the period so that $A(n) = A(n+N)$.

Appendix E

Perron-Frobenius theorem and its applications

E.1 The theorem

The Perron-Frobenius theorem is an important result about eigenvalues of real positive, or more general, non negative, matrices. These matrices are the ones corresponding to linear functions $\phi : \mathbb{C}^n \to \mathbb{C}^n$ so that $\phi(\Pi) \subseteq \Pi$, where the subset Π is $\Pi = \{x \in \mathbb{R}^n | x_i \geqslant 0\}$. We define also the subset of strictly positive vectors $\Pi_0 = \{x \in \mathbb{R}^n | x_i > 0\}$, and the absolute value of a vector $|| : \mathbb{C}^n \to \Pi$ so that $v = x_1 e_1 + \cdots + x_n e_n \to |v| = |x_1| e_1 + \cdots + |x_n| e_n$. The theorem says:

Theorem E.1.1 ***(Perron's Theorem)*** *Be $\phi : \mathbb{C}^n \to \mathbb{C}^n$ a linear function so that $\phi(\Pi \setminus \{0\}) \subseteq \Pi_0$ and we denote by $r = \rho(\phi)$ the spectral radius of ϕ. The following statements hold true:*
(i) r is en eigenvalue of ϕ and if v is an eigenvector relative to an eigenvalue of modulus r, then $|v| \subseteq \Pi_0$ and it is eigenvector relative to the eigenvalue r.
(ii) r is the only eigenvalue with modulus equal to $\rho(\phi)$.
(iii)The exponent of $(x - r)$ in the minimal polynomial of ϕ is equal to 1.
(iv) The algebraic multiplicity of r is 1.
(v) No other eigenvalue of ϕ has eigenvectors belonging to Π_0.
(vi) Be $f : \Pi \setminus \{0\} \to \mathbb{R}$ defined by $f(v) = \min_{i=1,..n} \frac{e_i^* \circ \phi(v)}{e_i^* \circ v}$, with $e_i^* \circ v \neq 0$. Then $r = \max_{v \in \Pi \setminus \{0\}} f(v)$. f is called the Collatz-Wielandt function.

The request of $\phi(\Pi \setminus \{0\}) \subseteq \Pi_0$, i.e. that the associated matrix A has $a_{ij} > 0$, can be weakened by letting $a_{ij} \geqslant 0$ (non negative matrix). In this case, that is also the most common in statistical physics and stochastic processes, some of the above properties are lost, while some others resist.

Corollary E.1.1.1 *Be $\phi : \mathbb{C}^n \to \mathbb{C}^n$ a linear function so that $\phi(\Pi) \subseteq \Pi$. The following conditions hold true:*
(i) $r = \rho(\phi)$ is eigenvalue of ϕ.
(ii) An eigenvector $v \in \Pi \setminus \{0\}$ relative to r.
(iii) Be $f : \Pi \setminus \{0\} \to \mathbb{R}$ defined by $f(v) = \min_{i=1,..n} \frac{e_i^* \circ \phi(v)}{e_i^* \circ v}$, with $e_i^* \circ v \neq 0$. Then $r = \max_{v \in \Pi \setminus \{0\}} f(v)$.

The differences with respect to the positive matrix are that the spectral radius can be 0 ; the algebraic and geometric multiplicity of r can be different from 1 ; the eigenvalue

of maximal absolute value can be not unique. Frobenius found the property that ϕ non negative must satisfy in order to extend to it the uniqueness results of Perron's theorem: the function needs to be irreducible. We will not delve into this as it is out of our scope.

E.2 Applications

We make extensive use of this theorem when it is time to both analytically and numerically compute the epidemic threshold.

- In Eq.(3.16) we use Lemma (i) of the Perron's Theorem to prove that:

$$\rho\left((1-\mu)I+\lambda A^{\top}\right)=1-\mu+\lambda\rho(A^{\top}). \tag{E.1}$$

In fact, in general $\rho(A+B)\neq\rho(A)+\rho(B)$. We start by noticing that the eigenvalues of
$P=(1-\mu)I+\lambda A^{\top}$ are $x_P=1-\mu+\lambda x_A$, because every eigenvector of A is also eigenvector of the identity; moreover being A symmetric (by definition of undirected network) its eigenvalues are all real $x_A\in\mathbb{R}$. To compute the spectral radius, we have to evaluate $|x_P|$ and find its maximum value. Of course x_P belongs to a certain real interval $I_P=[x_{min},x_{Max}]$, therefore:

$$\max_{x_P\in I}|x_P|=\begin{cases}|x_{min}| & \text{if} \quad x_{min}<x_{Max}<0\\ |x_{min}| & \text{if} \quad x_{min}<0, x_{Max}>0, |x_{min}|>x_{Max}\\ x_{Max} & \text{if} \quad x_{Max}>x_{min}>0\\ x_{Max} & \text{if or} \quad x_{min}<0, x_{max}>0, x_{Max}>|x_{min}|\end{cases} \tag{E.2}$$

We now substitute $x_P=1-\mu+\lambda x_A$, with $x_A\in[x^A_{min},x^A_{max}]$. As A is non-negative , we can apply Lemma (i) in Perron's Corollary and therefore we know that $\rho(A)=x^A_{max}>0$ and $|x_A|\leqslant x^A_{max}$. Moreover $(1-\mu)>0$ by definition of μ as a probability. Therefore, we have restricted our possible intervals I_P to be $[1-\mu+\lambda x^A_{min},1-\mu+\lambda x^A_{max}]$. This is the case corresponding to $x^P_{max}>0$ and $x^P_{max}>|x^P_{min}|$, that we find in the last two cases of Eq.(E.2). In conclusion, the spectral radius of P is:

$$\rho(P)=\max_{x_P\in I}|x_P|=x^P_{Max}=1-\mu+\lambda x^A_{max}=1-\mu+\lambda\rho(A). \tag{E.3}$$

- Another important application of Perron's theorem is given by the direct use of Lemma (v) for computing the spectral radius of a non-negative matrix. The numerical method that implements it is usually called *power method*. It is sufficient to take a non zero (can be random) starting vector v_0 and iterate the recurrence relation

$$v_k=\frac{Av_{k-1}}{\|Av_{k-1}\|} \tag{E.4}$$

So, at every iteration, the vector v_k is multiplied by the matrix A and normalized. If we assume A has an eigenvalue that is strictly greater in magnitude than its other eigenvalues and the starting vector v_0 has a non-zero component in the direction of an eigenvector associated with the dominant eigenvalue, then the sequence $\{v_k\}$ converges to an eigenvector associated with the dominant eigenvalue. In the same way, the sequence of associated values $\{\lambda_k\}$ with:

$$\lambda_k=\frac{v_k^{\top}Av_k}{v_k^{\top}v_k}. \tag{E.5}$$

converges to the spectral radius of A.

Appendix F

Aggregating time windows and impact on the threshold

In many cases, information on network dynamics can be coarse. Data often report the temporal evolution at a lower resolution time scale (i.e. the time step is wider) than real one: $\Delta t_{data} > \Delta t_{real}$. The temporal-network we can reconstruct from data thus corresponds to an aggregated representation of the real and finer temporal dynamics. This means that all causal structures and temporal correlations that occur at time scales that smaller than the coarse data-resolution are lost. Since these structures can impact disease dynamics, it is crucial to assess how such coarser representation influences the description of epidemic processes ([48]). Here, we study the influence of two aggregation schemes on the epidemic threshold. HET scheme is a weighted aggregation of the snapshots, obtained by summing link weights: $W(t), W(t+1) \rightarrow W(t) + W(t+1)$ where W(t) is the weighted adjacency matrix of time t. HOM is topologically equivalent to HET, having the exact same set of links. Each link is given an equal weight corresponding to the average link weight of the weight distribution of the HET network aggregated over the same period. As a result, both schemes share the same average weight at every aggregation interval, but HET accounts for weight heterogeneity.
We apply these aggregation schemes to our network of encounters. We use the intrinsic transmissibility λ for comparison across different aggregation schemes and intervals, as it does not depend on weight. Starting from the finest resolution (Δt=1), we aggregate snapshots recursively, increasing each time the aggregation interval Δt_k. We consider the recovery rate m as an intrinsic property of the disease, thus not changing with aggregation. The probability of recovery *after* a time Δt is $ke^{-m\Delta t}$ Hence, we compute the recovery probability $\mu(\Delta t_k)$ at the aggregation interval $\Delta t_k = kdt = k$ as the probability of recovering *within* that interval $\mu(\Delta t_k) = 1 - e^{-mk}$. In this way we account for the changing of the time-scale of network dynamics via a rescaling of the recovery probability μ. We recall that high recovery rates mean short average infectious periods, and thus fast disease progression at node level. Conversely, low recovery rates induce long infectious periods, resulting in slower microscopical disease dynamics.

The curves of Figure (F.1) show that the prediction made on the aggregated sex-workers network deteriorates with the increase of the time aggregation window Δt. Because, as expected, the aggregation induces a loss of the temporal information with the result that the aggregated network performs poorly with respect to reproducing the behaviour obtained in the original network. By focusing on the HETEROGENEOUS time-aggregation (blue lines) we find $(\lambda_{\Delta t}/\lambda_1) < 1$: the HET network becomes more vulnerable with respect to the original time-resolution. The reason is that aggregation

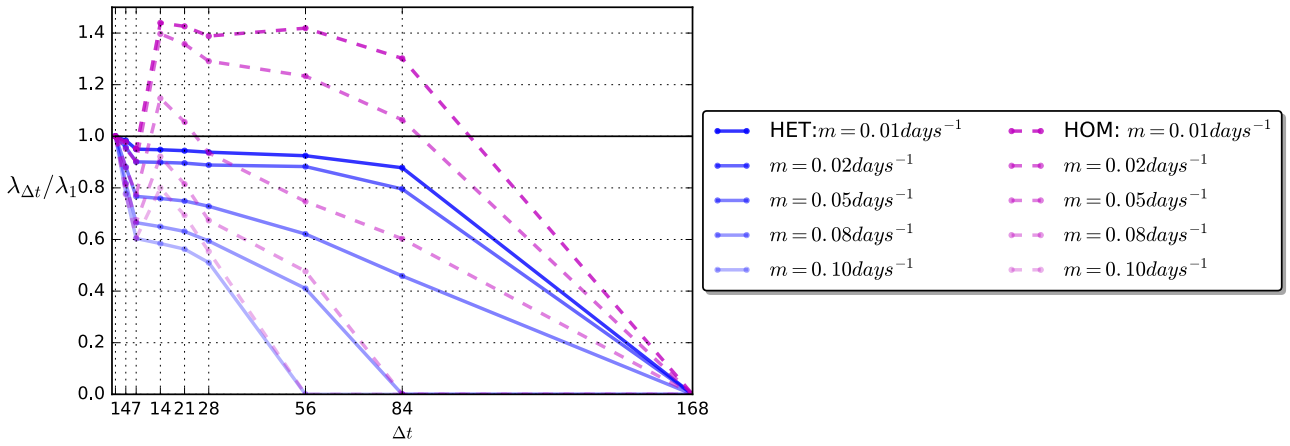

Figure F.1: **HET and HOM aggregation schemes compared.**

creates more paths of diffusion for the epidemic and also increases the link density ([48]). All these effects tend to facilitate the spread of a disease, so that the resulting epidemic threshold is lower than the one computed on the original temporal network. The effect is more rapid and stronger for the fast disease (e.g. m = 0.1 days^{-1} in shaded blue), because in that case the disease has the possibility to experience the entire landscape of dynamical changes the network undergoes through, thus differentiating a lot between the pattern obtained at the highest resolution and the aggregated one.

If we focus on the HOM aggregation scheme, we observe that the epidemic threshold predicted for a given Δt is systematically higher than the one obtained in the HET scheme for the same Δt value (same intensity of color, dashed lines vs. continuous lines in Fig.(F.1)). The reason lies in the way weights are distributed over the links of the aggregated networks. While the HET scheme preserves the heterogeneity of the duration of the contacts, cumulating the duration of the interaction established by each pair of individuals, this information is lost in the HOM scheme as the total contact duration is homogeneously distributed among all contacts. Heterogeneity of the weights has a strong effect on the evolution of epidemics and in many cases it has the effect of favouring the spread of diseases. This results in a lower epidemic threshold for HET than its homogeneous counterpart, for a given Δt. The faster the disease is, the smaller is the difference observed in the epidemic threshold obtained from the two aggregation schemes.

Bibliography

[1] WHO. Report, *www.who.int/mediacentre/factsheets/fs310/en/index1* (2012)

[2] L Liu, H Johnson, S Cousens, J Perin, S Scott, J E Lawn, I Rudan, H Campbell, R Cibulskis, M Li, C Mathers, R. E Black, *Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000.*, The Lancet, 379(9832):2151–2161 (2015)

[3] CDC Report, *Antibiotic Resistance Threats in the United States.*, www.cdc.gov/drugresistance/threat-report-2013/ (2013)

[4] C W Morin, A C Comrie, K Ernst, *Climate and dengue transmission: Evidence and implications.*, Environmental Health Perspectives, 121(11- 12):1264–1272 (2013)

[5] A McLean, *SARS a case study in emerging infections.*, Oxford University Press, Oxford New York (2005)

[6] M Tizzoni, P Bajardi, C Poletto, J J Ramasco, D Balcan, B Gonçalves, N Perra, V Colizza, A Vespignani, *Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm.*, BMC medicine, 10(1):165 (2012)

[7] C Poletto, MF Gomes, A Pastore y Piontti, L Rossi, L Bioglio, D L Chao, I M Longini, M E Halloran, V Colizza, A Vespignani, *Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic.*, Eurosurveillance, 19(42):20936 (2014)

[8] W O Kermack, A G McKendrick, *Contributions to the mathematical theory of epidemics I-II-III.*, Bulletin of Mathematical Biology, 53(1-2) (1991)

[9] N Bailey, *The mathematical theory of infectious diseases and its applications.*, Griffin, London (1975)

[10] R M Anderson,R M May, *Infectious Diseases of Humans: Dynamics and Control.*, Oxford University Press (1992)

[11] D Daley, *Epidemic modelling : an introduction*, Cambridge University Press, Cambridge New York (1999)

[12] M J Keeling, P Rohani, *Modeling Infectious Diseases in Humans and Animals.* , Princeton University Press (2007)

[13] S Eubank, H Guclu, V S Anil Kumar, M V Marathe, A Srinivasan, Z Toroczkai, N Wang, *Modelling disease outbreaks in realistic urban social networks.*, Nature, 429(6988):180–184 (2004).

[14] N M Ferguson, D 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.*, Nature, 437(7056):209–214 (2005)

[15] D Balcan, V Colizza, B Gonçalves, H Hu, José J Ramasco, A Vespignani, *Multiscale mobility networks and the spatial spreading of infectious diseases*, Proceedings of the National Academy of Sciences, 106(51):21484–21489 (2009).

[16] C T Butts, *Revisiting the Foundations of Network Analysis.*, Science, 325(5939):414–416 (2009)

[17] A Vespignani, *Predicting the Behavior of Techno-Social Systems.*, Science, 325(5939):425–428 (2009)

[18] M E J Newman, *Networks : an introduction.*, Oxford University Press, Oxford New York (2010)

[19] *Chapter 12 - An Overview of Social Networks and Economic Applications.* , Handbook of Social Economics, volume 1 (2011)

[20] A Vespignani, *Modelling dynamical processes in complex socio-technical systems.*, Nat Phys, 8(1):32–39 (2012)

[21] A L Barabàsi, *Network science.*, Cambdridge University Press, Cambridge UK, (2015)

[22] P Holme, J Saramaki,*Temporal networks.*, Physics Reports, 519(3):97–125 (2012)

[23] R Pastor-Satorras, C Castellano, P Van Mieghem, A Vespignani, *Epidemic processes in complex networks*, Reviews of Modern Physics, V.87 (2015)

[24] A Barrat, M Barthelemy, A Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, England (2008)

[25] AS Mata, SC Ferreira, *Pair quenched mean-field theory for the susceptible-infected- susceptible model on complex networks*, Europhys. Lett. 103, 48003 (2013)

[26] S Gómez, A Arenas, J Borge-Holthoefer, S Meloni, Y Moreno, *Discrete-time Markov chain approach to contact-based disease spreading in complex networks*, Europhys. Lett. 89(3), 38009 (2010)

[27] M De Domenico, A Solé-Ribalta, E Cozzo, M Kivelá, Y Moreno, M.A Porter, S. Gómez, A.Arenas,*Mathematical Formulation of Multilayer Networks*, Phys. Rev. X, 3(4):41022 (2013)

[28] E Valdano, L Ferreri, C Poletto, V Colizza, *Analytical Computation of the Epidemic Threshold on Temporal Networks.*, Phys. Rev. X 5, 021005 (2015)

[29] See Supplemental Material at *http://link.aps.org/ supplemental/10.1103/PhysRevX.5.021005* for performance of the algorithm used to compute the threshold in Valdano et al. *Analytical Computation of the Epidemic Threshold on Temporal Networks.*, Phys. Rev. X 5, 021005 (2015)

[30] R Pastor-Satorras, A Vespignani, *Epidemic spreading in scale-free networks.*, Phys. Rev. Lett., 86(14):3200–3203 (2001)

[31] J Gómez-Gardeñes, V Latora, Y Moreno, E Profumo, *Spreading of sexually transmitted diseases in heterosexual populations.*, PNAS , vol. 105 no. 5 1399–1404 (2008)

[32] E Valdano, *Computing the vulnerability of time-evolving networks to infections.*, Santé publique et épidémiologie, Université Pierre et Marie Curie-Paris VI (2015)

[33] E Valdanoa,C Poletto,A Giovanninic, D Palmac, L Savinic, V Colizzaa, *Predicting epidemic risk from past temporal contact data.*, PLoS Comput Biol, 1004152 (2015)

[34] L Rocha, F Liljerosb, P Holme, *Information dynamics shape the sexual networks of Internet-mediated prostitution.*, PNAS , 5706–5711, vol. 107, no. 13 (2010)

[35] C Harcourt, B Donovan, *The many faces of sex work.*, Sex Transm. Infect., 81:201–206 (2005)

[36] S Baral et al, *Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis*, The Lancet, 12(7):538-549 (2012)

[37] X Jin et al, *HIV prevalence and risk behaviors among male clients of female sex workers in Yunnan, China*, JAIDS ,53(1):131-135 (2010)

[38] T L Patterson et al, *Correlates of HIV, STIs and Associated High Risk Behaviors among Male Clients of Female Sex Workers in Tijuana, Mexico*, AIDS ,23(13):1765-1771 (2009)

[39] G do Espirito Santo, E Etheredge, *Male clients of brothel prostitutes as a bridge for HIV infection between high risk and low risk groups of women in Senegal*, Sexually Transmitted Infections, 81(4):342-344(2005)

[40] L Rocha ,F Liljeros, P Holme, *Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts.*, PLoS Comput Biol 7(3): e1001109 (2011)

[41] MC Boily ,RF Baggaley ,L Wang et al. *Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies.*, Lancet Infectious Diseases (2009)

[42] P Holmes, K King , D W Johnson,H J Trostle, *An estimate of the risk of men acquiring gonorrhea by sexual contact with infected females.*American Journal of Epidemiology 91.2 (1970)

[43] Kent ME, F Romanelli, *Reexamining syphilis: an update on epidemiology, clinical manifestations, and management*, Annals of Pharmacotherapy 42 (2) (2008)

[44] Teng, Yu, Nan Kong, Wanzhu Tu, *Estimating age-dependent per-encounter chlamydia trachomatis acquisition risk via a Markov-based state-transition model.*, Journal of clinical bioinformatics 4.1 (2014)

[45] N Perra, B Gonçalves, R Pastor-Satorras, A Vespignani, *Activity driven modeling of time varying networks.*, Sci. Rep. 2, 469 (2012)

[46] H Lentz, T Selhorst, I Sokolov, *Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks*, Phys. Rev. Lett. 110, 118701 (2013)

[47] L Ferreri,P Bajardi,M Giacobini, S Perazzo, E Venturino, *Interplay of network dynamics and heterogeneity of ties on spreading dynamics.*, Phys. Rev. E 90, 012812 (2014)

[48] E Valdano, C Poletto, V Colizza, *Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution.*, Eur. Phys. J. B , 88341 (2015)

[49] JC Delvenne, R Lambiotte, L Rocha, *Diffusion on networked systems is a question of time or structure.*, Nature Comm., 6:7366 (2015)

[50] B Øksendal, *Stochastic Differential Equations, An Introduction with Applications*, 5th Edition, Springer-Verlag Heidelberg (1998)

[51] CW Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 2nd Ed, Springer-Verlang New York (1985)

[52] P D Powell, *Calculating Determinants of Block Matrices*, arXiv:1112.4379.

[53] S Elaydi, *An Introduction to Difference Equations, 3rd ed.*, Springer, New York, NY (2005)

[54] K Binder, D W Heermann, *Monte Carlo Simulation in Statistical Physics, An Introduction*, Springer, Berlin (2010)