

# COMPLESSITÀ NEI SISTEMI SOCIALI

LORENZO DALL'AMICO  
*ISI Foundation*

Note per il corso in Fisica dei sistemi complessi

Università di Torino

April 2, 2024

Lorenzo Dall'Amico: *Complessità  
nei sistemi sociali*

# CONTENTS

<b>1</b>	<b>TEMPORAL GRAPHS</b>	<b>1</b>
1.1	Why temporal graphs? . . . . .	1
1.2	Representing temporal graphs . . . . .	2
1.3	Measuring proximity graphs . . . . .	5
1.4	Properties of temporal graphs . . . . .	8
1.5	Conclusion . . . . .	12
1.6	References . . . . .	12
<b>2</b>	<b>EPIDEMICS ON NETWORKS</b>	<b>13</b>
2.1	Epidemics . . . . .	13
2.2	Epidemic modeling on networks . . . . .	16
2.3	Extensions . . . . .	22
2.4	References . . . . .	23
<b>3</b>	<b>CAVITY METHOD</b>	<b>25</b>
3.1	Sparse and tree-like graphs . . . . .	25
3.2	Factorizing probability distributions on trees . . . . .	27
3.3	The non-backtracking matrix . . . . .	30
3.4	An application to epidemics . . . . .	31
3.5	Efficiently computing the spectral radius of $B$ . . . . .	33
3.6	Conclusion . . . . .	34
3.7	References . . . . .	34
<b>4</b>	<b>GRAPH FOURIER TRANSFORM</b>	<b>37</b>
4.1	The Fourier transform . . . . .	37
4.2	The graph Fourier transform . . . . .	38
4.3	Graph signal processing . . . . .	44
4.4	Conclusion . . . . .	48
4.5	References . . . . .	49
<b>5</b>	<b>COMMUNITY DETECTION</b>	<b>51</b>
5.1	Community detection . . . . .	51
5.2	Optimization approaches . . . . .	55
5.3	Inference in the DCSBM . . . . .	59
5.4	Spectral clustering . . . . .	64
5.5	Conclusion . . . . .	75
5.6	References . . . . .	76
<b>6</b>	<b>GRAPH EMBEDDINGS</b>	<b>77</b>
6.1	Graph embeddings . . . . .	77
6.2	Word2Vec . . . . .	78
6.3	Node2Vec . . . . .	81
6.4	Conclusion . . . . .	82
6.5	References . . . . .	83

# ACRONYMS

---

F2F	<i>Face-to-face</i>
ER	<i>Erdős-Rényi</i>
NMF	<i>naïve mean field</i>
BP	<i>belief propagation</i>
SIR	<i>Susceptible-Infected-Recovered model</i>
GW	<i>Galton Watson tree</i>
GFT	<i>Graph Fourier transform</i>
CD	<i>Community detection</i>
AMI	<i>Adjusted mutual information</i>
DCSBM	<i>Degree corrected stochastic block model</i>
SC	<i>Spectral clustering</i>

# SYMBOLS

---

- $\delta_{a,b}$  is the Kroeneker delta equal to 1 if  $a = b$  and equal to 0 otherwise.
- $\Lambda(M)$  is the set of eigenvalues of a matrix  $M$ .  $\lambda_i(M)$  is the  $i$ -th (smallest or largest, according to the context) eigenvalue of  $M$
- The spectral radius of  $M$  (largest eigenvalue) is denoted with  $\rho(M)$ .
- With the notation  $\mathbf{1}_n$  we denote the all-ones vector of size  $n$ .
- The entry-wise Hadamard product is denoted with  $\circ$ .
- The set of the first  $n$  integers is denoted with  $[n]$ .
- We adopt the Landau notation for the asymptotic behavior of variables. In particular  $x = O_n(y)$  is equivalent to  $\lim_{n \rightarrow \infty} \frac{x}{y} = c$  for some finite  $c$ . The notation  $x = o_n(y)$  instead means  $\lim_{n \rightarrow \infty} \frac{x}{y} = 0$ .
- The set of neighbors of a node  $i$  on a graph is  $\partial i = \{j \in \mathcal{V} : A_{ij} = 1\}$



# TEMPORAL GRAPHS

---

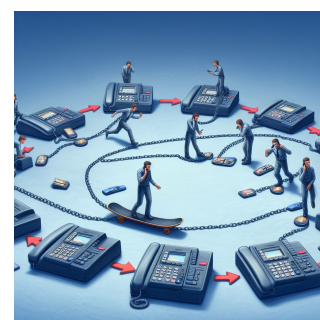
1.1	Why temporal graphs? . . . . .	1
1.2	Representing temporal graphs . . . . .	2
1.3	Measuring proximity graphs . . . . .	5
1.4	Properties of temporal graphs . . . . .	8
1.5	Conclusion . . . . .	12
1.6	References . . . . .	12

---

## 1.1 WHY TEMPORAL GRAPHS?

Graphs are an essential mathematical tool to represent interacting systems. When using static graphs, the interactions between two nodes is often represented as a single Boolean variable determining whether or not those two nodes interacted with one another. Yet, we can think of many examples in which this variable should depend on time. Think of a graph in which an interaction between two people is a phone call. In most instants, for most people, no interactions are recorded at all, and, in all cases at most one interaction per time may occur. So, how would we determine the Boolean interaction variable? One approach would be to aggregate time as determine that two people interacted if they had enough phone calls during a specific time window. In this way, connections are created between people that frequently call each other, but we loose an important piece of information: the order of events. Imagine an event like the fire of Notre Dame de Paris: in few moments people witnessing the fire spread the news to their contacts who, themselves reported to others in chain. If we had no idea of what happened and what is the content of the calls or messages, we could actually retrieve the geographical location from which the burst of information was initiated by retracing backwards the chain of events. If instead we use a static representation of the graph as we did earlier, all this information would be lost. A temporal graph is then an object capable of representing the interactions between the elements of a system together with a time stamp. Besides communication graphs, other notable examples are face-to-face proximity graphs, biological graphs, ecological graphs and many others.

As we will show in the remainder, in some cases it is necessary to keep the temporal dimension into account if one wants to understand the process happening on top of a graph. This is due to the fact that links may have



*A pictorial representation of a chain of calls*

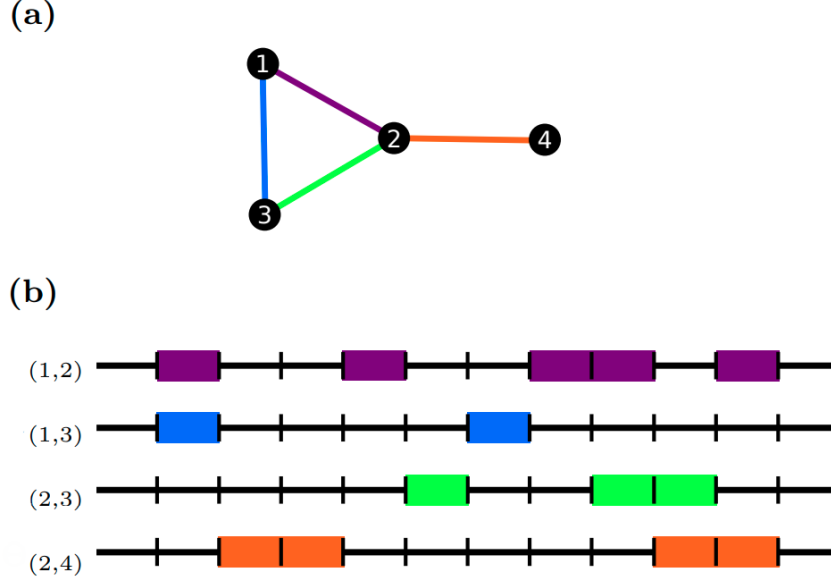


Figure 1.1: **A pictorial representation of a temporal graph.** (a) A graph with 4 nodes with  $\mathcal{V} = \{1, 2, 3, 4\}$  and  $\mathcal{E} = \{(1, 2), (1, 3), (2, 3), (2, 4)\}$ . (b) the temporal activation patterns of each edge (with color code). The  $x$  axis represents time. Picture taken from *Gauvin et al., Randomized reference models for temporal graphs*.

a causal relation (like the phone calls in the Notre Dame example) or simply because the aggregated static graph may be non representative of the interactions at any time stamp. We will then discuss how to mathematically define and represent temporal graphs, a relevant example of how to measure them and show some peculiar properties of temporal graphs.

## 1.2 REPRESENTING TEMPORAL GRAPHS

Let us consider a set of nodes  $\mathcal{V}$  and a set of edges  $\mathcal{E}$  connecting the nodes. Given a suited definition of interaction for our problem and a time window,  $\mathcal{E}$  is the set of all pairs of nodes  $(i, j)$  that interacted at least once in the considered time window. We now want to add the notion of *when* these interactions occurred. Each edge  $(i, j)$  can appear multiple times and for each interaction we can consider a time  $t$  at which the interaction begun and a time duration  $\tau$ .<sup>1</sup> We can then represent a temporal graph as a sequence of *temporal edges* in the form  $(i, j, t, \tau)$ . Figure 1.1 gives a pictorial representation of the temporal edges of a graph with 4 nodes. We now provide a more formal definition of a temporal graph.

<sup>1</sup> We could equivalently replace  $\tau$  with the time  $t_e$  at which the interaction ended.



### Temporal graphs

A temporal graph is a tuple  $\mathcal{G}(\mathcal{V}, \mathcal{E}_t)$ , where  $\mathcal{V}$  denotes the set of  $n$  nodes and  $\mathcal{E}_t$  of temporal edges. Each  $e \in \mathcal{E}_t$  can be written as  $(i, j, t, \tau)$  where  $i, j \in \mathcal{V}$ ,  $t$  is a time-stamp and  $\tau \in \mathbb{R}^+$  is a positive interaction duration, implying that the link between  $i$  and  $j$  was active from  $t$  to  $t + \tau$ . If a node has at least one connection at time  $t$  we say it is *active* at time  $t$  and it is inactive otherwise.

Temporal graphs

One can generalize the concept of adjacency matrix to the temporal setting by letting

$$\tilde{A}_{ij}^{(t)} = \begin{cases} 1 & \text{if } \exists e = (i, j, t_0, \tau) \in \mathcal{E}_t \text{ s.t. } t \in [t_0, t_0 + \tau] \\ 0 & \text{else.} \end{cases}$$

Temporal adjacency matrix

In this representation, however, time is kept as a continuous variable and, even for a finite observation time, we obtain an infinite number of adjacency matrices. For this reason, the snapshot representation – that uses a discrete notion of time – may be more suited. In particular, we assume that the interaction duration  $\tau$  is a multiple of a unit  $\Delta t$  that sets the temporal resolution of the graph. Let us define the concept of snapshot graphs, pictorially visualized in Figure 1.2.

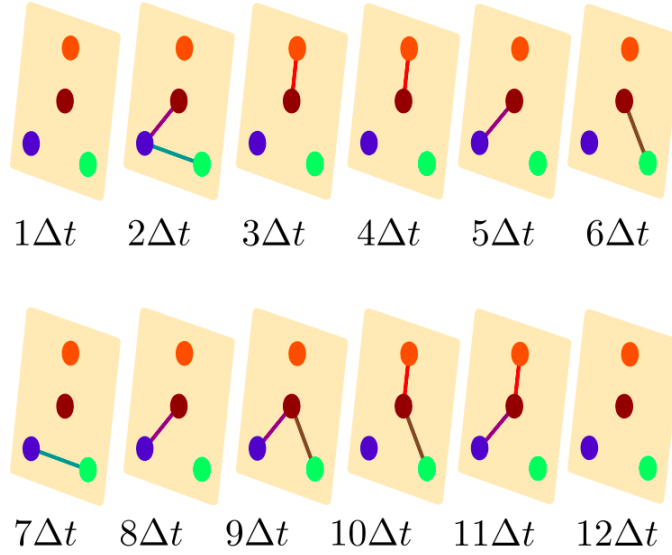


Figure 1.2: **A pictorial representation of a snapshot graph.** This plot represents the same graph of Figure 1.1 in which time was already discretized for convenience. Each *slice* corresponds to a different time step in which the edges progressively are activated. Picture taken from *Gauvin et al., Randomized reference models for temporal graphs*.

### Snapshot graphs

Snapshot graphs

A snapshot graph is a tuple  $\mathcal{G}(\mathcal{V}, \mathcal{E}_t)$ , where  $\mathcal{V}$  denotes the set of  $n$  nodes and  $\mathcal{E}_t$  of temporal edges. Each  $e \in \mathcal{E}_t$  can be written as  $(i, j, t)$  where  $i, j \in \mathcal{V}$ ,  $t$  is a *discrete* time-stamp and models an instantaneous interaction between  $i, j$  at time  $t$ .

With this representation at hand, one can obtain the adjacency matrix representation of a temporal graph as follows

The snapshot  
adjacency matrix

$$A_{ij}^{(t)} = \begin{cases} 1 & \text{if } (i, j, t) \in \mathcal{E}_t \\ 0 & \text{else.} \end{cases}$$

In this way we obtain a sequence of  $T$  adjacency matrices  $\{A^{(t)}\}_{t \in 1, \dots, T}$ , where  $T$  is the total number of snapshots. A natural question that poses is how which of the two representations is more appropriate and how much information is lost when choosing to discretize time. We address this point in the following remark.

### Discretizing time

Discretizing time

If we want to choose a discrete representation of time, the natural questions that arise are how to choose  $\Delta t$ , the minimal interaction duration and how much do we loose in performing this simplification. The first point we want to raise is that any measured quantity (including time) is not truly continuous and it is bounded by a resolution that makes time discrete by design. So, in all cases, one can say  $\Delta t$  is the measurement instrument resolution and the snapshot representation is always appropriate. Yet, if  $\Delta t$  is much smaller than the total observation time, the number of time frames  $T$  – even if finite – will tend to be very large, hence untractable. So, setting a longer  $\Delta t$  might be more appropriate in some cases and the choice of a good  $\Delta t$  is necessarily problem-dependent because  $\Delta t$  determines the scale at which we consider interactions to be *simultaneous*.

The coupling  
between the process  
and graph dynamics

Let us make two examples to make this point clearer. Suppose we have two spreading phenomena: in one case the propagation of a piece of information (such as the fire of Notre Dame) and in the other a flu-like illness transmission. In the former case, the propagation of the information from one person to the other moves very fast and so  $\Delta t$  must be small, in the order of seconds/minutes to capture the rapid dynamics of the news propagation. If we consider the flu-like propagation, instead, we know that a person, after being infected, is not typically able to infect someone for a couple of days, hence we can set  $\Delta t$  of the order of one day, assuming that one cannot change its own infectious state in the course of a day. So, summarizing, the proper time aggregation depends on the time-scale of the dynamic

process happening on top of the graph. If this process is much slower than the temporal evolution then we can simply aggregate the graph. If instead the two time scales (of the process and of the graph evolution) are similar, then we have an interesting coupling that must be taken into account.

To conclude this remark, there is still a quantity we want to preserve when we aggregate time, that is the cumulative interaction duration: what do we do with all interactions so that  $\tau < \Delta t$ ? Also in this case the answer depends on the problem under consideration. Take for instance the flu propagation with a time aggregation of 24 hours. If an infectious individual has a interaction with a susceptible one, the interaction duration is key to determine the probability of infection: an hour-long interaction is much more likely to propagate the disease than a minute-long interaction and all interactions will be shorter than  $\Delta t$  in this case. To preserve this piece of information, we might want to associate a weight to each edge, representing the cumulative interaction duration. We relate the continuous time and snapshot adjacency matrices as follows

$$W_{ij}^{(t)} = \int_t^{t+\tau_0} dt' \tilde{A}_{ij}^{(t')}.$$

If this representation may seem very reasonable, it must be noted that is not the only admissible one: in the case of the propagation of sexual diseases, for instance, the interaction duration is not relevant and may simply want to keep a Boolean representation of the edges, without attributing any weight.

*The weighted  
aggregated graph*

Now that we have introduced some of the main concepts related to temporal graphs, let us consider the specific set of proximity graphs as a case study, first describing a method to measure these graphs and then using the open source, real data to describe some relevant properties.

## 1.3 MEASURING PROXIMITY GRAPHS

In proximity graphs the edges represent a *Face-to-face* (F2F) contact between two persons. The interest of this type of graphs resides in the fact that F2F interactions are the vehicle of human communication and of infectious diseases and, more generally, they quantify how humans interact with one another. Measuring *Face-to-face* (F2F) proximity graphs is, however, a very challenging task. Among the most used approaches to quantify them we have the use of questionnaires in which the interactions one has are self-reported. It was shown that this method is biased towards long interactions – in the sense that short interactions tend to be forgotten – and can achieve a low temporal resolution. A very important contribution to the field of measur-

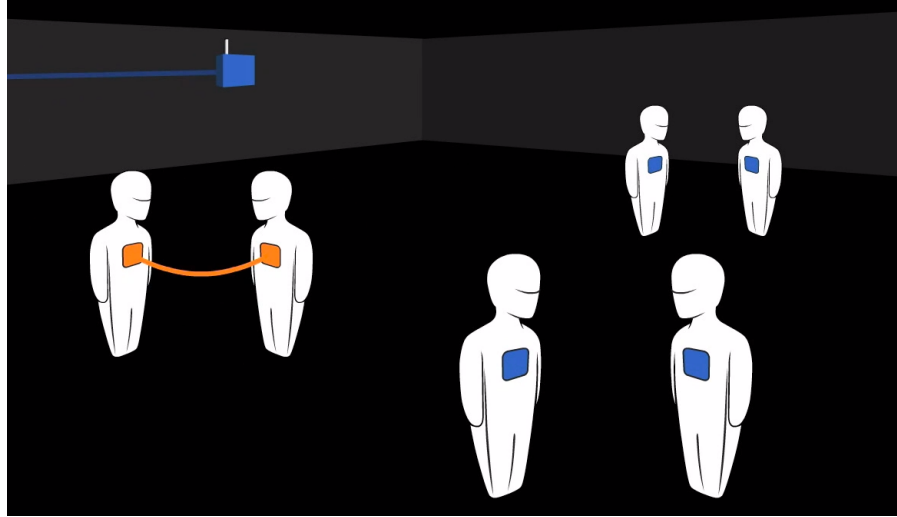
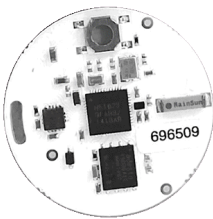


Figure 1.3: **Pictorial example of the use of SocioPatterns proximity sensors.** Six people in a room wearing a proximity sensor. The orange ones (with the line), indicate a recorded **F2F** interactions between two individuals. Picture taken from <http://www.sociopatterns.org/>.

ing **F2F** proximity graphs was made with the creation of the **SocioPatterns** collaboration by the **ISI Foundation**.

SocioPatterns was formed in 2008 and developed wearable proximity sensors that are capable to *measure* temporal proximity graphs. Their functioning is based on the transmission and exchange of information packets using radio-frequency electromagnetic waves. Briefly speaking, each device is associated with a code and continuously switches from a *listener* to a *speaker* mode. When it is in the *speaker* mode it emits an information packet containing its own code and the power at which the signal was emitted. When it is in listener mode, instead, the device intercepts the packets emitted by the “speakers” and records on its memory the code and the power declared, the time stamp at which this interaction occurs as well as the power of the received signal. The devices have to be worn on the chest of the participants and, by design, record **F2F** proximity. Figure 1.3 shows a demo of the functioning of the SocioPatterns proximity sensors.

The SocioPatterns  
collaboration



A proximity sensor

Inside the memory of each sensor we then have a list of entries of the type  $(j, \text{pow}_{\text{tr}}, \text{pow}_{\text{rec}}, t)$ , where  $j$  is the code of the sensor that emitted the signal,  $\text{pow}_{\text{tr}}$  is the transmission power,  $\text{pow}_{\text{rec}}$  is the received power and  $t$  is the time-stamp that has a temporal resolution of 20 seconds. Looking at the difference  $\text{pow}_{\text{tr}} - \text{pow}_{\text{rec}}$  one can measure the attenuation of the signal and filter out the interactions that are too attenuated, thus keeping only those that happened at a distance within approximately 2 meters. From this we can create a snapshot graph with  $\Delta t = 20 \text{ s}$  as described above.

The SocioPatterns sensors have been used in several contexts, including schools, hospitals, offices and rural African villages among others. They constitute a well known benchmark of temporal graph measurement that has

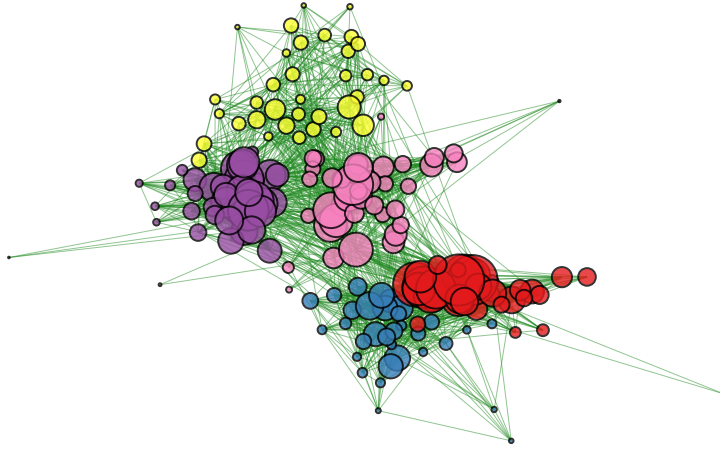


Figure 1.4: **The School dataset.** Pictorial representation of the *School* dataset (see Table 1.1) aggregated over all the observation time. The size of each node is determined by the total interaction time of that node, while the color is determined by the class the student belongs to.

been used in many applications and many of the collected datasets are publicly available at <http://www.sociopatterns.org/datasets/>. We will make use of some of these data to study some relevant properties of temporal graphs. Table 1.1 summarizes some descriptive properties of the considered graphs, while Figure 1.4 shows the aggregated graph collected in a high school.

Name	$n$	Observation time	Description
School	180	from a Monday to the Tuesday of the following week in November 2012.	interactions between students in a high school in Marseilles, France belonging to 5 classes.
Office	92	June 24 to July 3, 2013	interactions between individuals measured in an office building in France
Village	86	between 16th December 2019 and 10th January 2020	interactions between the people of Mdoliro village in Dowa district in the Central Region of Malawi.
Conference	405	June 4-5, 2009	interactions at the SFHH conference in Nice

Table 1.1: **Summary statistics of the SocioPatterns temporal networks.** The first column indicates the name used in these notes. The column indexed by  $n$  indicates the number of nodes appearing in the graph. The column *Observation time* describes the experiment duration, while *Description* provides a few details on the context of the data collection. For more information, refer to the [SocioPatterns](http://www.sociopatterns.org/) website.

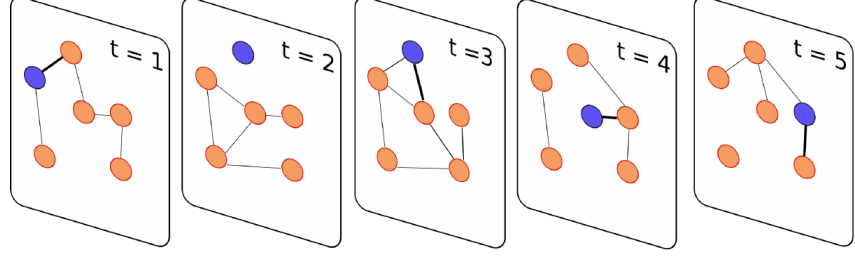


Figure 1.5: **Time respecting paths.** Five snapshots of a temporal graph in which unoccupied nodes are depicted in orange at each time, while the currently occupied node is in blue. A larger width is used to highlight the edge that causes the transition.

## 1.4 PROPERTIES OF TEMPORAL GRAPHS

We now proceed to describe some important concepts that characterize temporal graphs and use the four aforementioned datasets to show them on empirical data.

### TIME-RESPECTING PATHS

When we consider a graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$ , we define a path on it as an ordered sequence of nodes  $\{i_1, i_2, \dots, i_T\}$  for that, for all  $p \in [T]$ ,  $i_p \in \mathcal{V}$  and for all  $p \in [T - 1]$ ,  $(i_p, i_{p+1}) \in \mathcal{E}$ . In words, every step of a path allows one to only move from a node to one of its neighbors. When we consider a temporal graph, instead, we must generalize the concept of path, to encode the role played by time, introducing the *time-respecting paths*.

#### Time respecting paths

Given a temporal graph  $\mathcal{G}(\mathcal{V}, \mathcal{E}_t)$ , we denote a time respecting path as  $\{i_1(t_1), i_2(t_2), \dots, i_T(t_T)\}$  if it satisfies the following conditions

- For all  $p \in [T]$ ,  $i_p \in \mathcal{V}$ : the path is defined on the graph nodes.
- For all  $p \in [T - 1]$ ,  $t_p < t_{p+1}$ : these two times indicate the beginning of the residency on the respective nodes and time must be increasing.
- For all  $p \in [T - 1]$ ,  $\exists t_0, \tau$  s.t.  $(i_p, i_{p+1}, t_0, \tau) \in \mathcal{E}_t$  and  $t_{p+1} \in [t_0, t_0 + \tau]$ : the transition between one node and the other can only take place at a time at which the two nodes are connected.

*Time respecting  
paths*

This definition is given for a continuous time representation but it can simply be adapted to the discrete one. For this case, we give a simple representation of a time respecting path in Figure 1.5.

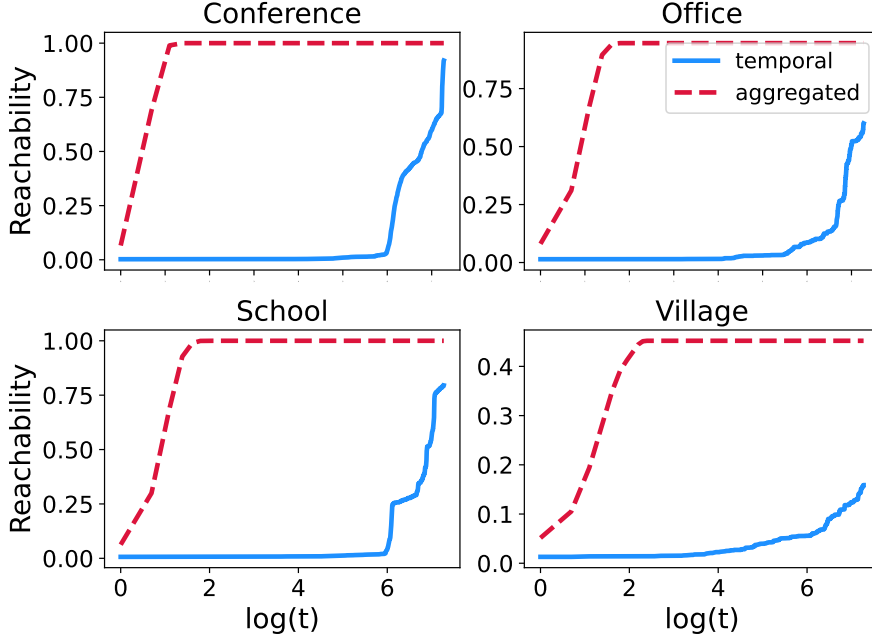


Figure 1.6: **Time-respecting vs aggregate reachability.** Each plot refers to one of the 4 *SocioPatterns* datasets described in Table 1.1, considering the first 8 hours of measurements. The red dashed line is the average of the reachability matrix  $R_t$  defined in Equation (1.1) as function of time, using for all  $t$   $A_t$  the weighted aggregated matrix over all the observation period. The blue continuous line, instead, is obtained from the snapshot adjacency matrices and encodes time-respecting paths.

Given a sequence of adjacency matrices, we now define the *reachability matrix*  $R_t$  as follows

$$R_t = \text{sign} \left[ \prod_{t'=1}^t (A_{t'} + I_n) \right], \quad (1.1) \quad \text{Reachability matrix}$$

where  $I_n$  is the identity matrix, the sign function has to be considered entry-wise, while the product has to be taken from right to left, i.e.  $\prod_{t=1}^3 A_t = A_3 A_2 A_1$ . The entry  $R_{t,ij}$  equals 1 if there exists a time-respecting path of length smaller or equal to  $t$  that allows one to go from  $i$  to  $j$ .

An important fact related to this matrix is that it is not necessarily symmetric. This comes from the fact that the product of matrices (such as the  $A_t$ 's) is symmetric only if the matrices commute. This is not the case in general and it implies that if there is a time-respecting path from  $i$  to  $j$ , that does not imply that there exists also a time-respecting path from  $j$  to  $i$ .

*Time-respecting  
paths are not  
symmetric*

By taking the average of the reachability matrix, we also have a measure of how well its nodes are connected. Figure 1.6 compares the reachability on the 4 real temporal graphs described above with the one obtained on their aggregated version and clearly shows that temporal graphs have a lower reachability. This is because the valid time-respecting paths are constrained

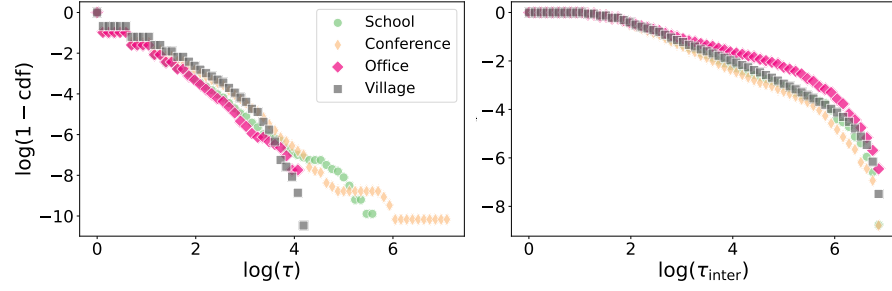


Figure 1.7: **Event and inter-event duration distributions.** The figures show the scatter plot in log-log scale of the interaction duration (left) and inter-event duration (right) distributions vs the  $1 - cdf$ , i.e. the complementary of the cumulative density function. Each line refers to one of the datasets described in Table 1.1 and is color and marker coded.

and are only a subset of all the possible paths that one can perform on the aggregate version of the graph. This is an important ingredient to consider when coupling a dynamic process with the graph, because only a fraction of all possible paths can actually take place.

## DURATION DISTRIBUTION AND BURSTINESS

We now focus on a very peculiar aspect of contact graphs, that is the contact duration distribution. It has been observed in many instances (with no apparent exception) that this distribution is very broad and follows approximately a power law decay that appears to be a universal behavior. Figure 1.7 (left plot) shows in log-log scale  $1 - cdf$  vs the interaction duration and confirms this trend, since the relation is approximately linear in the logarithmic scale. This is an important observation, because it tells us that very long interactions are much more common than what we would expect for a thin tail distribution, such as the Poisson. The consequence is that, if we have a process that needs a minimal time of interaction to consider the interaction to be valid, then, in practice, even if the threshold is very large, there will be valid interaction edges with high probability. On the other hand, we also know that most of the distribution is concentrated around small values.

A similar behavior is observed for the inter-event duration distribution. We define the inter-event duration as the time elapsed between two successive interactions of the same pair of nodes  $(ij)$ . Since this distribution is broad, we say that the interaction dynamics is *bursty*, i.e. that typically we have an alternation of time intervals in which the activity is very low and some in which it is very high. To best understand the effect that a bursty dynamics may have on a process, let us consider the following example.

Suppose we have a quantity  $\mathcal{Q}$  that is increased by one unit every time there is an interaction and it is decreased by a factor  $\alpha$  for each time step in which no interaction occurs. If  $\mathcal{Q}$  exceeds a threshold value  $\mathcal{Q}_{th}$ , then some

$$1 - cdf(x) = \mathbb{P}(\tau \geq x)$$



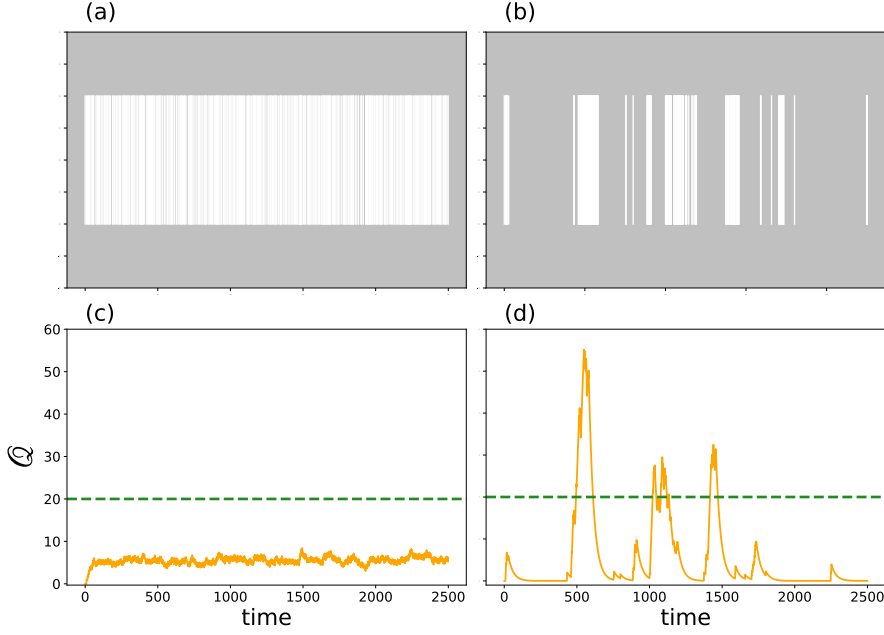


Figure 1.8: **The effect of bursty dynamics.** The first row represents a temporal time series for (a) a Poisson process and (b) the interaction times of a node of the *Conference* graph (see Table 1.1). Each horizontal line indicates an active time. The second line displays the dynamics of a quantity  $\mathcal{Q}$  evolving according to the process described in the main text for (c) the Poisson dynamics of (a), and for (d) the bursty dynamics of (b). The horizontal dashed line indicates an arbitrary selected threshold value that triggers some process when  $\mathcal{Q} > \mathcal{Q}_{\text{th}}$ . Image adapted from Holme, Saramaki, *Temporal networks*.

process is triggered otherwise it is not. In Figure 1.8 we compare this dynamical process on a Poisson temporal series made of 362 interaction events with one extracted from an individual activity pattern of the *Conference* graph. The bottom plots clearly evidence that the bursty dynamics, being highly concentrated in some time regions, allows one to go beyond the threshold several times, while this does not happen to the Poisson distribution.

A method to measure the burstiness level of a time series is as follows

$$B = \frac{s - m}{s + m},$$

where  $s$  and  $m$  are the standard deviation and mean of the inter-event duration distribution, respectively. For a periodic process,  $s = 0$  and  $B = -1$ , while for the burstiest of processes,  $s \rightarrow \infty$  and  $B \rightarrow 1$ . In our case, the Poisson dynamics of Figure 1.8(a) has  $B = -0.45$ , while the one of Figure 1.8(b) has  $B = 0.70$ .

## 1.5 CONCLUSION

Temporal networks are a powerful tool to model complex dynamical systems. Empirical networks often show very broad distribution of the interaction duration as well as bursty dynamics. These features are of great importance to some dynamical processes that may unfold on networks and the temporal framework is a relevant generalization of static graphs. However, the dynamic component of graph evolution must always be compared with the typical time scale of the process unfolding over the graph in order to understand whether it is necessary to have an additional layer of complexity given by time or, more in general, to choose an appropriate time discretization to perform the analysis.

## 1.6 REFERENCES

- P. Holme, J. Saramäki, *Temporal networks*. Physics reports, 519(3), 97-125 2012.  
This is *the* reference for an introduction to temporal graphs.

# EPIDEMICS ON NETWORKS

---

2.1	Epidemics . . . . .	13
2.1.1	Epidemic modeling . . . . .	13
2.1.2	The epidemic threshold . . . . .	14
2.2	Epidemic modeling on networks . . . . .	16
2.2.1	The state evolution equation . . . . .	16
2.2.2	Naïve mean field . . . . .	17
2.2.3	The reproductive number with naïve mean field . . . . .	18
2.2.4	Graph structure and reproductive number . . . . .	19
2.2.5	From theory to practice: epidemic mitigation . . . . .	21
2.3	Extensions . . . . .	22
2.4	References . . . . .	23

---

## 2.1 EPIDEMICS

Epidemiology is a branch of science that studies the determinants, distribution and dynamics of a propagation process in a population. Given its relation to a whole population, epidemiology is, by design complex and of great interest when shaping public health policies. We will focus on infectious disease epidemiology, *i.e.* on illnesses that can be transmitted from one person to the other. Notable examples include the bubonic plague, smallpox, the Spanish flu, HIV, influenza and, of course, Covid. Even though we will mainly have in mind “illness” propagation, a similar if not identical mathematical framework can be adopted to anythings that propagates through interactions, such as opinions, computer viruses or information.

Let us now describe in deeper detail some fundamental elements of epidemics and their modeling.

### 2.1.1 EPIDEMIC MODELING

The basic medical observation that we want to model and capture is that when a person is “sick” – for instance it is positive to influenza – and it is in contact with a person who is not – and never was – then it can pass the infection over to the healthy person. The meaning of “contact” depends on the disease we are considering; think to the extreme difference there

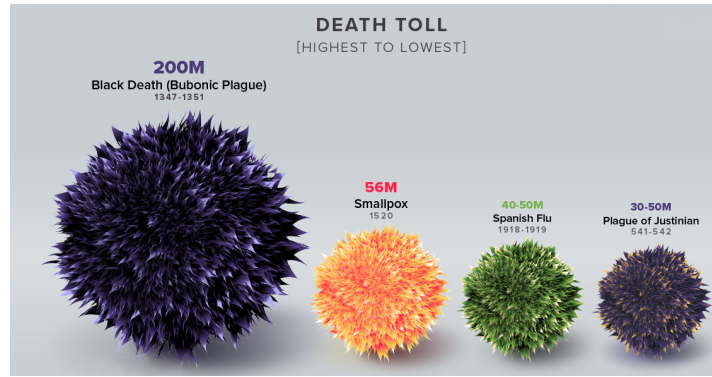
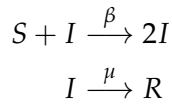
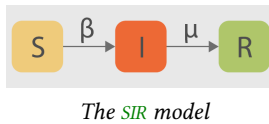


Figure 2.1: **The deadliest pandemics in history.**

Source: [visualcapitalist.com/history-of-pandemics-deadliest](https://visualcapitalist.com/history-of-pandemics-deadliest)

is between the HIV and the flu transmissions. Nonetheless, the dynamic processes we have in mind are very similar. We introduce the *Susceptible-Infected-Recovered model* (**SIR**), a cornerstone of epidemic modeling.

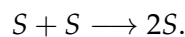
We define three possible states every individual can be in:  $S$ , susceptible;  $I$ , infected;  $R$  recovered. Susceptible people are healthy individuals that can contract the disease. Infectious ones are those who currently carry the disease and can spread it out when they interact with a susceptible individual. The recovered people, instead, are those who used to be infectious and now can no longer be infected. There are two important parameters that take part to this model:  $\beta$  that is the probability *per unit time* to infect a susceptible person and  $\mu$ , the probability *per unit time* to recover. We can summarize the **SIR** model with the following equation



The parameters  $\beta, \mu$  are disease-dependent and tell us how easily the infection runs across the population. Intuitively, if  $\beta \gg \mu$  we are in a situation in which people get infected at a much faster pace than they recover. As a consequence, the epidemic will swiftly spread across the population. On the opposite, if  $\mu \ll \beta$  people simply recover very quickly and the epidemic dies out. This concept, that we delineated in intuitive terms, is the so-called epidemic threshold that determines the necessary and sufficient condition for an epidemic spreading to occur and that we now more formally define.

### 2.1.2 THE EPIDEMIC THRESHOLD

We consider a population in which everybody is susceptible. This condition is a stationary state, because no infection can occur among susceptible people, or, in other words



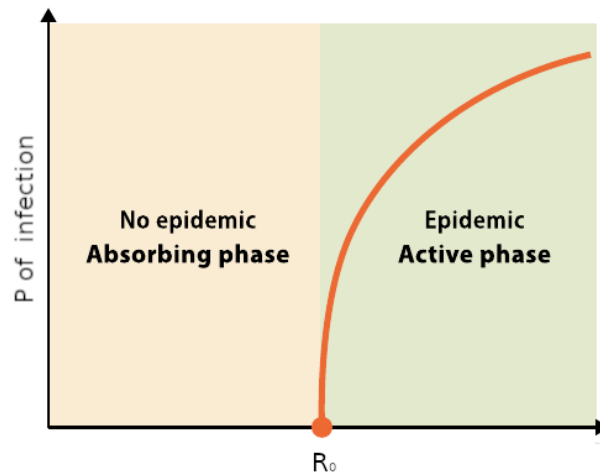


Figure 2.2: **Epidemic threshold phase diagram.** Below the critical value of  $R_0$  we are in an absorbing phase in which there is no epidemic, while for  $R_0$  larger than the critical value, the probability that each node has of being infected is non-null. Source: Satorras, Castellano, Van Mieghem, Vespignani. *Epidemic processes in complex networks*.

We now perturb this state introducing few infectious individuals and ask ourselves whether the system will fall back to an equilibrium having most people being unaffected by the disease, or, if it will spread hitting a large portion of the population. The simple approach to understand this problem lies in the following question

*At the beginning of the spreading,  
how many people do I infect before recovering?*

If the answer is “*more than one*”, then we have a cascade in which the propagation grows exponentially fast. If, on the opposite it is “*less than one*” than it dies out because, in most cases, every individual recovers before infecting someone else. We call this the *reproductive number*. As shown in Figure 2.2 it is the control parameter of a phase transition between a disease-free and an epidemic state.

Let us stress two important facts. The first one is that  $R_0$  is defined *at the beginning* of the epidemic. Notation-wise, this is why we call it  $R_0$ , while  $R_t$  is the reproductive number at a given time  $t$ . This is important to say because the exponential spreading can occur only on a short time scale and then saturate due to the finite population size. The second fact is that the question refers to a non-better specified “ $I$ ”, implying that the answer is the same for everybody. Of course, if we think of our everyday life, the probability of infection vary a lot across individuals, according to their sociability, the interaction with children, the spaces they occupy *etc.* So, of course, we want an *average* answer and we will study it in the next section.



*The effect of  $R_0$ ,  
some numbers*

Source: [visualcapitalist.com/history-of-pandemics-deadliest/](https://visualcapitalist.com/history-of-pandemics-deadliest/)

## 2.2 EPIDEMIC MODELING ON NETWORKS

As we mentioned, the epidemic spreading propagates through contacts that are well modeled by the edges of a graph. We now want to formally define the epidemic threshold transition for an arbitrary graph. This is of key importance to understand if, given the structure of the network and the disease parameters, the spread will touch a large fraction of individuals or not.

### 2.2.1 THE STATE EVOLUTION EQUATION

We consider an agent-based model in which every individual  $i \in \mathcal{V}$  is associated to a discrete variable  $x_i(t) \in \{S, I, R\}$  determining the state  $i$  is in. When a susceptible person  $i$  is in contact with an infectious one  $j$  ( $A_{ij} = 1$ ) for a time  $dt$ , then it gets infected with a probability  $\beta dt$ . An infected person recovers with a probability  $\mu dt$ . We can write the probability of being infected at the time-step  $t + dt$  as a function of  $t$  as follows:

*The infected state  
equation of the SIR on  
a graph*

$$\mathbb{P}(x_i(t + dt) = I) = \mathbb{E} \left[ \underbrace{\delta[x_i(t) = I](1 - \mu)}_{i \text{ was infected and did not recover}} + \underbrace{\delta[x_i(t) = S] \left( 1 - \prod_{j \in \mathcal{V}} (1 - \beta dt \cdot \delta[x_j(t) = I])^{A_{ij}} \right)}_{i \text{ was susceptible and got infected}} \right]. \quad (2.1)$$

This equation features two terms: the case in which  $i$  was susceptible and got infected and the case in which it was infected and did not recover. Note that the probability of being infected is written as 1 minus the probability of not being infected. We can simplify this equation in the limit for  $dt \rightarrow 0$ , focusing on the term describing the probability of not being infected.

$$\begin{aligned} & \lim_{dt \rightarrow 0} \prod_{j \in \mathcal{V}} (1 - \beta dt \cdot \delta[x_j(t) = I])^{A_{ij}} \\ & \stackrel{(a)}{=} \lim_{dt \rightarrow 0} \exp \left\{ \sum_{j \in \mathcal{V}} A_{ij} \log (1 - \beta dt \cdot \delta[x_j(t) = I]) \right\} \\ & \stackrel{(b)}{=} \lim_{dt \rightarrow 0} \exp \left\{ -\beta dt \sum_{j \in \mathcal{V}} A_{ij} \delta[x_j(t) = I] \right\} \\ & \stackrel{(c)}{=} \lim_{dt \rightarrow 0} 1 - \beta dt \sum_{j \in \mathcal{V}} A_{ij} \delta[x_j(t) = I], \end{aligned}$$

where in (a) we used the identity  $x = e^{\log x}$ ; in (b) we performed the expansion  $\log(1+x) = x + o(x)$  and in (c) the expansion  $e^x = 1 + x + o(x)$ . Substituting this expression in Equation (2.1) we obtain

$$\begin{aligned}\mathbb{P}(x_i(t+dt) = I) &= \mathbb{E}[\delta(x_i(t) = I)](1-\mu) + \beta dt \sum_{j \in \mathcal{V}} A_{ij} \mathbb{E}[\delta[x_i(t) = S] \delta[x_j(t) = I]] \\ &= \mathbb{P}(x_i(t) = I)(1-\mu) + \beta dt \sum_{j \in \mathcal{V}} A_{ij} \mathbb{P}(x_i(t) = S, x_j(t) = I).\end{aligned}$$

Taking the first term on the right hand-side to the left and dividing by  $dt$ , we obtain the derivative of the probability that reads

$$\partial_t \mathbb{P}(x_i(t) = I) = \beta \sum_{j \in \mathcal{V}} A_{ij} \mathbb{P}(x_i(t) = S, x_j(t) = I) - \mu \mathbb{P}(x_i(t) = I). \quad (2.2)$$

Following the same passages, we get the evolution equations for all three states. Note that,  $\partial_t \mathbb{P}(x_i(t) = S) + \partial_t \mathbb{P}(x_i(t) = I) + \partial_t \mathbb{P}(x_i(t) = R) = 0$ , because the probability of being in one of the three states sums up to one.

#### SIR model on a graph

$$\begin{aligned}\partial_t \mathbb{P}(x_i(t) = S) &= -\beta \sum_{j \in \mathcal{V}} A_{ij} \mathbb{P}(x_i(t) = S, x_j(t) = I) \\ \partial_t \mathbb{P}(x_i(t) = I) &= \beta \sum_{j \in \mathcal{V}} A_{ij} \mathbb{P}(x_i(t) = S, x_j(t) = I) - \mu \mathbb{P}(x_i(t) = I) \\ \partial_t \mathbb{P}(x_i(t) = R) &= \mu \mathbb{P}(x_i(t) = I).\end{aligned} \quad (2.3)$$

In order to obtain the reproductive number we must study the stability of this system of equations that, however, is still non-linear and hard to study because it involves the marginal distributions  $\mathbb{P}(x_i(t) = S, x_j(t) = I)$  for which we do not have an explicit expression. To cope with this problem, we adopt the simple naïve mean field approximation.

### 2.2.2 NAÏVE MEAN FIELD

The *naïve mean field* (NMF) approximation consists in considering all variables as independent *i.e.* in factorizing the marginals as follows

$$\mathbb{P}(x_i(t) = S, x_j(t) = I) = \mathbb{P}(x_i(t) = S) \mathbb{P}(x_j(t) = I).$$

*The NMF approximation*

This approximation greatly simplifies the problem. In fact, in (2.3) we have defined the evolution of  $3n$  equations concerning the node marginal probabilities, but there are  $2|\mathcal{E}|$  equations (where  $\mathcal{E}$  is the set of edges) that are unspecified. Factorizing the probabilities with naïve mean field, we simply get rid of these terms. Let us first make some comments about this approximation, its limits and when we expect to be a good method to proceed.

### Some notes on the naïve mean field approximation

Given its simplicity NMF is a commonly adopted strategy to first tackle a problem, but is it accurate? In other words, we are asking to what extent we can assume that the event that  $i$  is susceptible is independent from the event that its neighbor  $j$  is infected. Given the context of the model, the answer seems necessarily to be negative since the contagion is transmitted through the contacts.

The most relevant setting in which NMF should be considered is that of *dense* networks, *i.e.* those in which every node has a large degree. Suppose we have a fully connected network: the fact that the edge  $A_{ij}$  exists is simply irrelevant because all edges exist. One can show that indeed, in this setting the NMF approximation becomes asymptotically exact and, in general, the denser the network is the more the NMF approximation is accurate. An example of how to go beyond this approximation is discussed in chapter 3.

With NMF approximation at hand, Equation (2.2) turns into

$$\partial_t \mathbb{P}(x_i(t) = I) = \beta \mathbb{P}(x_i(t) = S) \sum_{j \in \mathcal{V}} A_{ij} \mathbb{P}(x_j(t) = I) - \mu \mathbb{P}(x_i(t) = I).$$

We now linearize this equation around the stationary state  $\mathbb{P}(x_i(t) = S) = 1$  to get the reproductive number.

### 2.2.3 THE REPRODUCTIVE NUMBER WITH NAÏVE MEAN FIELD

As we explained before, we want to see the effect of perturbing the stationary state in which everybody is susceptible by adding a small probability of being infected. For simplicity, we denote  $\mathbb{P}(x_i(t) = I) := p_i(t)$  and move to a vector form of the equations. We let  $P_i(x_i(t) = S) = 1$  and obtain

$$\partial_t \mathbf{p}(t) = (\beta A - \mu I_n) \mathbf{p}(t).$$

*The linearization  
around the  
diseases-free point  
under the NMF  
approximation*

If we want that  $\partial_t p_i(t) < 0$  for all  $i$  and all  $t$ , we must impose that  $\beta \rho(A) - \mu < 0$ , where  $\rho(A)$  denotes the spectral radius of  $A$ . If this condition is satisfied, then we end up in the disease-free region. If on the opposite  $\beta \rho(A) - \mu > 0$ , the probability of being infected grows at each time step and the virus has a broad diffusion on the network. We thus obtain the following value for the reproductive number

#### Reproductive number with the NMF approximation

$$R_0 = \frac{\beta \rho(A)}{\mu} \quad (2.4)$$



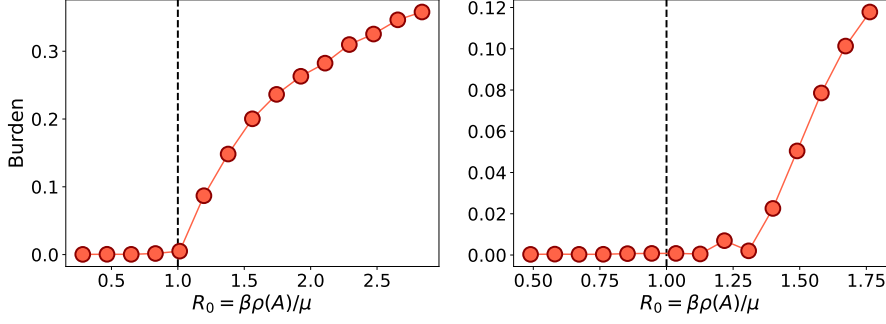


Figure 2.3: **Epidemic threshold: theoretical prediction versus simulated data.**

We run a **SIR** model for different  $\beta$  values on a dense graph (left panel) and on a sparse one (right panel) generated from the random configuration model. We plot the burden (*i.e.* the fraction of non-susceptible individuals) as a function of  $R_0$  as predicted by the **NMF** approximation.

In Figure 2.3 we compare this prediction with an empirical simulation on a dense (left panel) and on a sparse one (right panel), evidencing the goodness of the approximation only in the former case. Now, as a last step, we give some simple results relating the spectral radius of  $A$  with its structure.

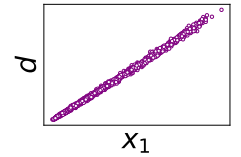
#### 2.2.4 GRAPH STRUCTURE AND REPRODUCTIVE NUMBER

Studying the spectral properties of the adjacency matrix for different generative models is a problem of great interest, that however goes beyond the scope of this course. Here we provide two simple examples with intuitive and non-rigorous arguments to characterize the value of  $\rho(A)$  and understand the role of density and degree heterogeneity in determining the threshold.

##### *Erdős Renyi random graph*

The spectral behavior of the *Erdős-Rényi* (**ER**) random graph changes dramatically according to whether its expected average degree grows with its size or not. Letting  $p$  be the probability of being connected, then we can define two different regimes: the dense one in which  $\log(n)/pn = o_n(1)$  and the sparse one in which the opposite is true, *i.e.*  $pn/\log(n) = o_n(1)$ . In words, if the average degree grows faster than  $\log(n)$  we say the network to be *dense*. This is a game-changer because, under this hypothesis, the degree distribution is concentrated, *i.e.* , for all large  $n$ , with probability one

$$\max |d_i - np| = o_n(np).$$



The leading eigenvalue of  $A$ ,  $x_1$  against the degree vector on a random dense graph

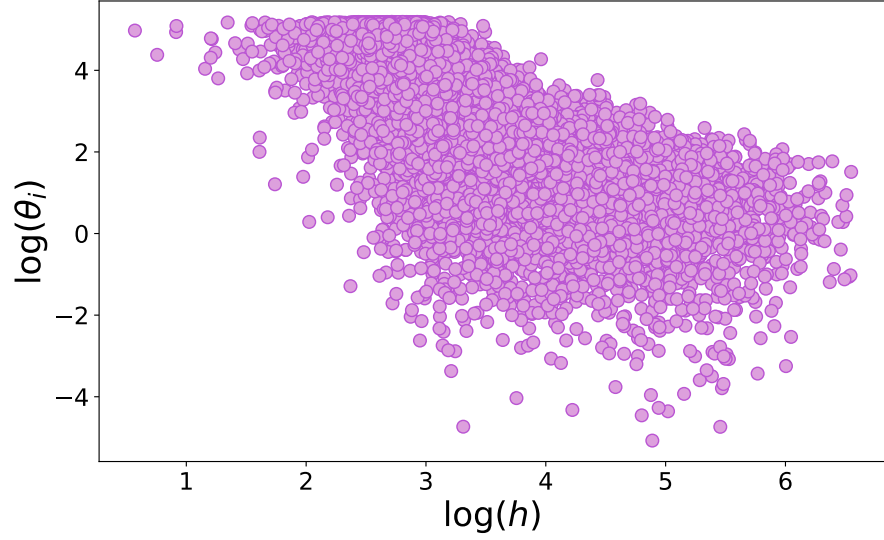


Figure 2.4: **Hitting time as a function of the expected degree.** We consider a random graph generated from the configuration model and denote with  $\theta_i$  the expected degree of node  $i$ . In the plot we show the scatter plot of  $\theta_i$  against the hitting time  $h$ , defined as the number of iterations after which the node got infected in a **SR** simulation.

Again, in words, this means that a dense **ER** graph is quasi regular. In this case, we can heuristically<sup>1</sup> write the following equation

$$\rho(A) = \langle d \rangle + o_n(d) \quad \text{on dense ER graphs}$$

$$(A\mathbf{1}_n)_i = \sum_{j \in \mathcal{V}} A_{ij} = d_i \approx np \approx \langle d \rangle (\mathbf{1}_n)_i.$$

This implies that  $\mathbf{1}_n$  is a close approximation of the leading eigenvector<sup>2</sup> and the average degree is an approximation of  $\rho(A)$ . From this result, we obtain that in denser networks an epidemic spreading runs faster, as one could reasonably expect.

Let us now consider the case in which the graph is generated from a configuration model with an arbitrary degree distribution.

### Configuration model

Once again we operate under the assumption of being in a sufficiently dense regime and derive a heuristic expression of the leading eigenvector of  $A$ , which we suppose in this case to be  $\mathbf{d}$ , the degree vector.

$$\rho(A) = \frac{\langle d^2 \rangle}{\langle d \rangle} + o_n(d) \quad \text{for dense random graphs with an arbitrary degree distribution}$$

$$(A\mathbf{d})_i = \sum_{j \in \mathcal{V}} A_{ij}d_j \approx \sum_{j \in \mathcal{V}} \frac{d_i d_j^2}{2|\mathcal{E}|} = d_i \frac{\langle d^2 \rangle}{\langle d \rangle}.$$

We thus get that  $\rho(A) = \langle d^2 \rangle / \langle d \rangle$  implying that a broad degree distribution makes the spreading run even faster on the network. This is because of the

<sup>1</sup> For simplicity we derive this result heuristically, but it can be formally proved.

<sup>2</sup> Due to Perron-Frobenius theorem.

role played by hubs, *i.e.* nodes with a high degree. Since they have a lot of connections, they are very likely to get infected in the earlier stages of the epidemic and then become super-spreaders. Notably, the value of  $\rho(A)$  that we just may diverge for scale-free networks in which the second moment of the degree distribution goes to infinity, making the transition go to zero. Figure 2.4 shows for each node the infection hitting time (*i.e.* the time it takes to get infected) as a function of their degree. The plot evidences a strong negative correlation, confirming the intuition that nodes with a large degree are the first to get infected and then they are responsible of the spreading. Note that in practice the second moment of the degree may only diverge in the asymptotic theoretical limit. All real world networks are finite and so is  $\langle d^2 \rangle$ . Nonetheless, in real-world settings, it can be overwhelmingly large and drive the epidemic to unfold very fast on the network.

### 2.2.5 FROM THEORY TO PRACTICE: EPIDEMIC MITIGATION

Let us now discuss some basic facts about epidemic mitigation based on our results. Suppose we have a vaccine and we add a fourth compartment to our model, that of vaccinated people that behaves exactly like the recovered one. From our analytical view-point, we can still deploy the results we obtained, because vaccinated people simply do not take part to the process, since they cannot change their compartment. For this reason it is as if they were not part of the network.

We ask ourselves how vaccination impacts the epidemic spread. To answer this question we add a Boolean variable  $s_i$  that equals 0 if  $i$  is vaccinated and cannot transmit the disease and is 1 otherwise. The matrix that determines the epidemic threshold is then now

$$W = A \circ (\mathbf{s}\mathbf{s}^T).$$

Using a non-rigorous argument, we will see how the vaccination determines the epidemic threshold.

#### Methodological remark

The approach we used to study  $\rho(A)$  was non rigorous but leads to the correct result because we assume  $A$  to be dense. By vaccinating a large portion of the population, instead  $W$  becomes sparse by design and not only the method we adopt but also the result is incorrect. Formally, we can only see what is the effect of vaccination on the spreading, assuming that a small fraction of the population has been vaccinated and  $W$  is still dense. This heuristic result, however, gives us some important intuition that we can verify numerically and that can be rigorously proved with other more rigorous methods.

Let  $\tilde{d}_i = d_i s_i$ . Then

$$(W\tilde{\mathbf{d}})_i = \sum_{j \in \mathcal{V}} A_{ij} s_i s_j \tilde{d}_j \approx \sum_{j \in \mathcal{V}} \frac{\tilde{d}_i d_j^2 s_j}{2|\mathcal{E}|} = \tilde{d}_i \frac{\mathbf{s}^T \mathbf{d}^2}{\mathbf{1}_n^T \mathbf{d}},$$

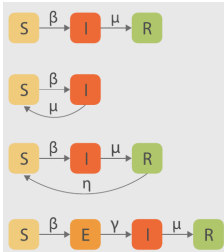
so  $\frac{\mathbf{s}^T \mathbf{d}^2}{\mathbf{1}_n^T \mathbf{d}}$  is a close approximation of  $\rho(W)$ . Now, if  $s_i$  is Bernoulli random variable with probability  $p$ , we get

$$\rho(W) \approx p \frac{\langle d^2 \rangle}{\langle d \rangle}.$$

So, the vaccination decreases the  $R_0$  and allows one to stay below the epidemic threshold. From a simple observation, however, one sees that this is not the optimal strategy. In fact, in we fix  $\mathbf{s}^T \mathbf{1}_n$ , *i.e.* the number of vaccines, and attempt to minimize  $\mathbf{s}^T \mathbf{d}^2$  one immediately sees that the solution lies in vaccinating the nodes with the highest degree, *i.e.* the *hubs*. This target immunization significantly helps in improving the mitigation effectiveness.

## 2.3 EXTENSIONS

In the previous sections we only considered the **SIR** to model an epidemic spreading. While this is one the most relevant models in epidemiology, it must be mentioned that several alternatives exist. The simplest one is the **SI** in which individuals cannot recover and is equivalent to the **SIR** for  $\mu = 0$ . In this case one can see that no epidemic threshold exists and, for how little is the transmission parameter, the epidemic will certainly involve all the population sooner or later. Different is the case of the **SIS** model in which an infected individual recovers but is once again susceptible. This model accounts for the fact that having experienced an infection does not imply one is immune, in some cases. From a mathematical perspective this slightly changes things: as we commented already, in the **SIR**, an infected individual cannot have been infected by a susceptible neighbor. On the opposite in the **SIS** this can happen: a infected individual can have been infected by someone who *now* is susceptible but that recovered. Actually, the difference between these two regimes cannot be understood from the **NMF** approximation because it is indeed not able to capture this dynamic. If one adopts a more refined approximation strategy, however, it is indeed possible to see that in the two cases the epidemic threshold varies.



Other models realistically add more compartments to the equations. Some of the most common are the *exposed*, *vaccinated* and *dead* compartments. Exposed people are those who already contracted the virus but that are still not contagious. After some time, they turn infected. The addition of these or other compartments may take into account of more complex medical and behavioral factors. On top of this, the model parameters may add further

depth. We assumed  $\beta, \mu$  to be constant and equal for all individuals, while one may assume that they depend, for instance, on age or mask-wearing.

It is worth mentioning that we only talked about simple contagion, *i.e.* the process in which one infected individual passes the disease over to a susceptible one. Thinking however of epidemiology in a broader sense, this is not the only possible alternative. Contagion may occur, for instance, only if one is exposed several times to an infected individual, each one passing a “piece”. Only when all “pieces” are passed one becomes infected. Alternatively one can imagine contagion as a process in which it is necessary to have several infectious people interacting *at once* for the disease to be transmitted. We talk in these cases of *complex contagion*.

## 2.4 REFERENCES

- A.L. Barabasi, *Network Science, Chapter 10*, [networksciencebook.com/chapter/10](http://networksciencebook.com/chapter/10)  
This chapter gives a wide and detailed view of epidemics on networks and can be used as a reading to get a bigger picture of the problem that we only treated in a very schematic way. From the analytical viewpoint, the book gives more results on the dynamics and considers also other models but does not use the NMF as it was presented here, but rather presents the degree-based mean field approach that is closely related to NMF.
- Pastor-Satorras, Castellano, Van Mieghem, Vespignani, *Epidemic processes in complex networks*, *Reviews of modern physics* 87.3 (2015): 925, [arxiv.org/pdf/1408.2701.pdf](http://arxiv.org/pdf/1408.2701.pdf)  
This is a long but very important review. The intent of this review is less pedagogical than the Barabasi’s book, but it takes a broader look at the problem from a technical perspective and summarizes different results.
- P. Van Mieghem, *Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold*, [arxiv.org/pdf/1402.1731.pdf](http://arxiv.org/pdf/1402.1731.pdf)  
This article details some rigorous results based on the NMF approximation with a notation very similar to the one adopted in this chapter.