

Lecture notes: Epidemics on Networks

Dr. Alejandro Lage Castellanos, and ChatGPT4

September 27, 2023

Contents

1	Classical results in mathematical epidemiology	5
1.1	A supersonic intro to Master equation	5
1.2	Daniel Bernoulli and smallpox	8
1.3	SIR compartmental models	9
1.3.1	Deterministic epidemics from master equation	12
1.3.2	Short time limit	14
1.3.3	Infinite time limit	15
1.3.4	Epidemics terminology	16
1.4	SIS model and epidemic thresholds	18
1.5	And beyond	19
1.6	Homework	19
2	Introduction to networks	21
2.1	Directed graphs	23
2.2	Weighted graphs	24
2.3	Graphs properties	24
2.4	Graph metrics	26
2.4.1	Density of a graph and mean degree	26
2.4.2	Diameter, radius, eccentricity	27
2.4.3	Jordan centrality measure	27
2.4.4	Degree centrality	27
2.4.5	Betweenness	28
2.4.6	Modularity	28
2.4.7	Clustering	28
2.5	Assortativity	29
2.6	Laplacian of a graph	29
2.7	Networks for benchmarking	30
2.7.1	Random graphs	30
2.7.2	Random regular graphs	31
2.7.3	Erdos-Renyi model	32
2.7.4	Albert Barabasi graphs	32
2.8	Practical Python Class: networks	33
2.9	Practical Homework	33
2.10	Percolation and epidemics	33

3	Monte Carlo simulation of epidemics	35
3.1	Rates and exponential distributions	35
3.2	Gillespie's Stochastic Simulation Algorithm	36
3.2.1	Gillespie's algorithm	37
3.3	Practical Python class	38
3.4	Practical Homework	38
4	Mean field solution of epidemics on Networks	41
4.1	Master equation	42
4.2	Mean field approximations to the SIS model	43
4.2.1	Individual-based mean field (IBMF) approximation	43
4.2.2	Pair-based mean field (PBMF)	44
4.3	Practical Python Class: simulating epidemics on networks	44
5	Reconstruction of epidemics	49
5.1	Inference in graphical models	50
5.1.1	Supersonic intro to statistical mechanics	51
5.1.2	Free energy as Kullback-Leibler distance	52
5.1.3	Mean field approximations	54
5.2	Beyond Mean Field: Belief Propagation	55
5.2.1	A simple example: coloring trees	57
5.2.2	A clever way to count	58
5.3	The zero patient problem	58
5.A	Functionals and Lagrange Multipliers	66
6	Cascade epidemics and contact tracing	69
6.1	Effects of contact tracing and quarantine	69

Chapter 1

Classical results in mathematical epidemiology

Epidemics are typical examples of complex phenomena. A mathematical description of an epidemic needs to take into account its non deterministic nature, and, therefore, rely on probabilities. An epidemic process is, essentially, a stochastic process.

In the dawn of mathematical epidemiology, researchers studied the dynamics in time of probabilities regarding the states of individuals. This is the case of the study of life expectancy and smallpox by Daniel Bernoulli, and so is the famous contribution of Kermack y McEndrick with the SIR models. In both cases, the resulting dynamic equations correspond to the Master equation of the given stochastic processes.

1.1 A supersonic intro to Master equation

Introduction to Stochastic Processes and Markov Chains

A **stochastic process** is a mathematical model that describes the evolution of a random system over time. It's a collection of random variables indexed by time, and can be used to model a wide range of phenomena, from the stock market to the weather.

A **Markov chain** is a specific type of stochastic process that has the **Markov property**: the future state of the system only depends on the current state, and not on any of the previous states. This makes Markov chains particularly useful for modeling systems where the future is dependent on the present, but not on the past.

Formally, a Markov chain is a sequence of random variables X_0, X_1, X_2, \dots that take values in a finite or countable set S , where the probability of moving from one state to another depends only on the current state. That is, for all $n \geq 0$ and all states $i, j \in S$,

$$\mathbb{P}(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{t+1} = j \mid X_t = i).$$

Markov chains have many interesting properties and applications, including their use in modeling queueing systems, random walks, and machine learning algorithms.

Master equation

The Master Equation is a fundamental tool in the study of stochastic processes, particularly in probability theory and statistical physics. It describes the time evolution of the probability distribution of a system in terms of transition rates between different states. In its simplest form, the Master Equation takes the form of a first-order differential equation:

$$\frac{d}{dt}P_i(t) = \sum_j W_{ji}P_j(t) - \sum_j W_{ij}P_i(t), \quad (1.1)$$

where $P_i(t)$ is the probability of the system being in state i at time t , and W_{ij} is the transition rate from state j to state i . The first term on the right-hand side of the equation represents the rate of transitions from all other states j to state i , weighted by the probability of being in state j . The second term represents the rate of transitions from state i to all other states j , weighted by the probability of being in state i .

Master equation can be seen as a continuity equation, or a mass conservation equation. In fluid dynamics, for instance, the rate of change of mass in a control volume must be equal to the net rate of mass flow into or out of the control volume. Mathematically, the continuity equation can be written as:

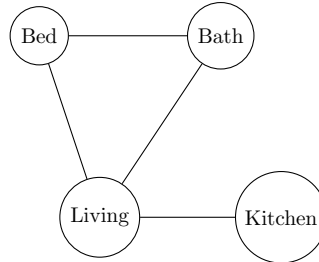
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = \frac{d}{dt} \int_V \rho dV + \oint_S \rho \mathbf{v} \cdot d\mathbf{S} = 0$$

where ρ is the fluid density, \mathbf{v} is the velocity vector, $\frac{\partial \rho}{\partial t}$ is the rate of change of density with respect to time, and $\nabla \cdot (\rho \mathbf{v})$ is the divergence of the mass flux density vector.

In other words, the continuity equation states that the change in the amount of fluid within a control volume is equal to the net flow of fluid across the boundaries of that control volume. This principle is based on the law of conservation of mass, which states that mass cannot be created or destroyed, only transferred or transformed.

Drunken man

A drunken man is moving around his own house, which is made of a Bedroom, a Bathroom, a Livingroom and a Kitchen, connected through doors as shown in the diagram:



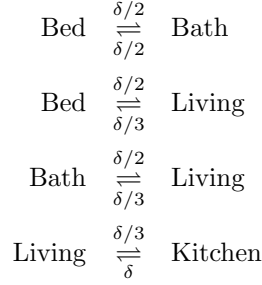
Everytime he enters a room, he chooses one of the doors in that room at random, and moves through it to the next room. The man starts in his bedroom.

1. What is the shortest path from the bedroom to the kitchen?
2. What is the largest path between bed and kitchen (without cycles)?
3. If the man keeps walking for a very long time, what amount of time he spends in every room?

4. What is the expected time that takes the man to visit the kitchen for the first time?

Solution in discrete time:

Let $\vec{p}(t) = (p_1, p_2, p_3, p_4)$ be the vector state defining the probabilities, at time t of finding the drunken man at (bed, bath, living, kitchen), respectively. This can be seen as the following set of stochastic decaying chemical reactions:



Consider the transition matrix

$$M = \begin{bmatrix} 1 - \delta & \delta/2 & \delta/3 & 0 \\ \delta/2 & 1 - \delta & \delta/3 & 0 \\ \delta/2 & \delta/2 & 1 - \delta & \delta \\ 0 & 0 & \delta/3 & 1 - \delta \end{bmatrix}$$

such that the evolution in discrete time of the Markov process is given by

$$\vec{p}(k+1) = M \cdot \vec{p}(k)$$

So the full evolution from $t = 0$ to t is given by

$$\vec{p}(k) = M^k \vec{p}(0)$$

Matrices as M , postive in their values, and such that every column add to 1, are known as stochastic matrices. It can be shown that the largest eigevalnue, in any stochastic matrix, is $\lambda = 1$, and its eigenvector is a probability vector.

In the current case, it is easy to prove, by substitution, that $\vec{p}_\infty = (2, 2, 3, 1)/8$ is such eigenvecto. This implies that

$$\vec{p}_\infty = M \cdot \vec{p}_\infty$$

meaning that \vec{p}_∞ is the stationary solution of the process.

A simple generalization proves that an umbiased random walk on a graph has a steaty state probability that only depends on the degree of each node, and its proportional to it.

Take home message: every discrete time homogenous in time Markov process corresponds to a transition (stochastic) matrix, whose $\lambda = 1$ eigenvector is the long time limit of the process.

Continuous time limit:

We can consider the scaling of this process as we devide the time scale in T units, and send $\delta \rightarrow \delta/T$. This process is known as continuous time limit, when $T \rightarrow \infty$. In this case

$$\vec{p}(k+1) - \vec{p}(k) = (M(\delta/T) - \mathbb{I}) \cdot \vec{p}(k)$$

resulting in

$$\frac{d\vec{p}}{dt} = (M(\delta) - \mathbb{I}) \cdot \vec{p}$$

Which also shows that the stationary solution is given by

$$\frac{d\vec{p}}{dt} = 0 \quad \Rightarrow (M(\delta) - \mathbb{I}) \cdot \vec{p} = 0$$

which is the definition of the eigenvalue $\lambda = 1$ for matrix M .

Attention, this eigenvector might not be unique. If it is unique, it is the solution in the long time limit. If it is not unique, then the solution will depend on the initial conditions.

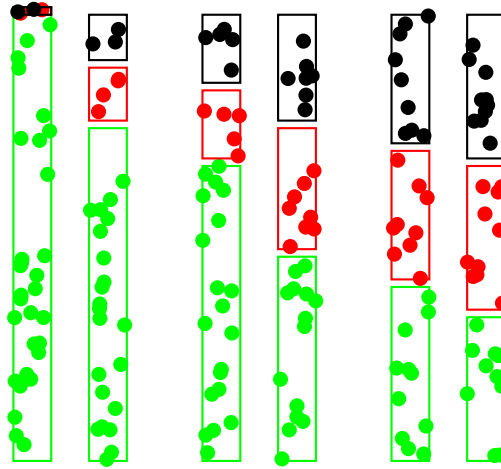
1.2 Daniel Bernoulli and smallpox

Daniel Bernoulli's study of the effect of smallpox in life expectancy and the possibility of inoculation as a vaccination strategy, a debated argument in the Royal Academy of Science in Paris, in the second half of the XVIII century.

Effect of inoculation in Smallpox, Royal Academy of Science, Paris, 1760, published in 1766

Scope: calculate the gain in life expectancy at birth if smallpox were to be eliminated as a cause of death. The model assumes a fixed amount of infections every year, so, it is not a model for the infection. Still, it deals with relevant questions in epidemiology.

The model of Bernoulli studied the age dynamics of the entire population in a country, as classified in three groups: susceptibles, immunes and dead¹. Susceptible was everyone that had not been in contact with the disease. Immune was everyone that had passed the disease. Dead were those that died, either from the disease or from other causes. Smallpox could either cause death, or cause life-long immunity. So, it is considered one of the SIR family of models, but rather, in this case, SR model.



Bernoulli imagined that people born in year $a = 0$, could go through the states (un-infected, immune, dead) with certain probabilities² The dynamics relied on the following dynamic parameters:

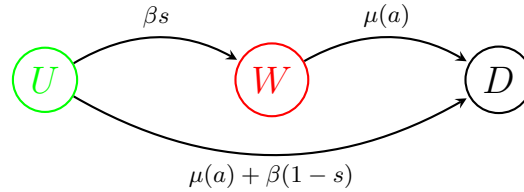
- $\mu(a)$ the rate of death of all causes except smallpox, for an individual of age a

¹As you can see, there is no characterization of those that are currently infected. The reason for this is that the study considered large time scales, such that the actual infectious period was negligible in that scale, and people were considered to transition immediately between susceptible to immune.

²The transitions between these states were described as a Markov chain, although Markov chains came centuries later.

- β is the **force of infection**: the rate at which susceptible individual get infected
- $c = 1 - s$ the fraction of individuals infected that dies of smallpox.

The Markov chain is described graphically as this:



The relevant quantities (dynamic variables) of the model depend on the age of the persons a . They are:

- $u(a)$ the probability for an individual of being alive and susceptible (uninfected) at age a
- $w(a)$ the probability for an individual of being alive and immune at age a
- $d(a)$ is the probability of being dead at age a , but since probabilities add up to 1, and there are only three possible states, $d(a) = 1 - u(a) - w(a)$.

Considering the rates at which individuals can move from one state to the other, we can compute the contributions of each state in the Master equation. The dynamical system derived by Bernoulli is given by:

$$\begin{aligned} \frac{du}{da} &= -(\beta + \mu(a))u(a) \\ \frac{dw}{da} &= \beta(1 - c)\mu(a)u(a) - \mu(a)w(a) \end{aligned}$$

The interest of Bernoulli was comparing the total amount of persons alive at a given age $u(a) + w(a)$ in the case where $c = 0$ (people do not die of smallpox), and $c \neq 0$.

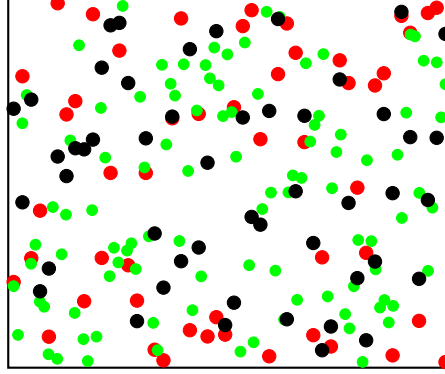
These equations are recognized as one of the first modern mathematical approaches to mathematical epidemiology. Their solution and Bernoulli conclusions is out of the scope of the present course (if interested see [?]). However, its interpretation and derivation are useful to introduce some probabilistic methods and epidemic concepts.

The Master Equation provides a powerful framework for analyzing a wide range of phenomena, including chemical reactions, population dynamics, and the behavior of complex systems. By solving the Master Equation, one can obtain the time-dependent probability distribution of the system, which can be used to calculate various statistical quantities of interest, such as the mean and variance of observables. The Master Equation can also be used to study the long-time behavior of the system, such as the emergence of steady-state solutions and the behavior of fluctuations around these solutions.

1.3 SIR compartmental models

The SIR (Susceptible-Infectious-Recovered) model is a widely used mathematical framework for studying the spread of infectious diseases in populations. It was first introduced by Kermack and

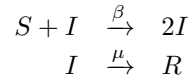
McKendrick in a seminal paper published in 1927. They used a system of differential equations to describe the time evolution of the number of individuals in each compartment, and derived an expression for the critical threshold of the disease, above which an epidemic occurs.



A first idealization of what actually happens in an epidemic, is to think of society as a gas of individuals. Each individual can be in one of the following states:

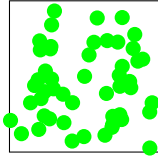
- **Susceptible** individuals, who are at risk of becoming infected;
- **Infectious** individuals, who are capable of transmitting the disease to susceptible individuals;
- **Recovered** individuals, who have either recovered from the disease and are immune, or have died from the disease, and are removed.

Under well mixed assumptions, these individuals-particles move around and interact with each other, occasionally. The process of contagion can be seen as a catalytic chemical reaction, in which the interaction of an Infectious individual with a Susceptible, can induce the latter to turn into infectious with a certain probability:

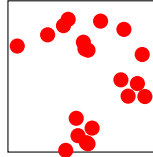


Simulation of this process in a computer, relies on Monte Carlo methods, more precisely in action diffusion equations and Gillespie algorithm, that we will study later.

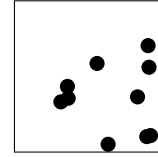
A mean field probabilistic description can be made, by considering the population to be classified into three different compartments, according to their state:



Compartment S

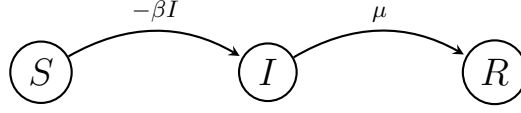


Compartment I



Compartment R

The evolution of the epidemic correspond to the process of changing individuals between compartments, with certain probabilities, as represented in the following diagram:



A key difference with the work of Bernoulli is that the *force of infection*: $\beta I(t)$ (the rate at which susceptible individuals contract the disease) is not fixed in time, but proportional to the amount of people infected in the population. This is typical for human-human transmitted diseases in a well mixed population, where the probability of getting infected depends on the chances of meeting an infectious individual. This probability is proportional to the current amount of individuals in that state.

The dynamic variables for this model are:

- $S(t)$, the fraction of the population susceptible to contracting the disease
- $I(t)$ the fraction that is currently infected and infectious
- $R(t) = 1 - S(t) - I(t)$ the fraction that is recovered or deceased.

We can write down the master equation of the stochastic process, which defines the evolution in time of $P_{S,I}(t)$, as

$$\frac{dP_{S,I}(t)}{dt} = \beta P_{S+1,I-1}(t) + \mu P_{S,I+1}(t) - \beta P_{S,I}(t) - \mu P_{S,I}(t)$$

By taking the average $S(t) = \frac{1}{N} \sum_{S,I} S P_{S,I}(t)$ and $I(t) = \frac{1}{N} \sum_{S,I} I P_{S,I}(t)$, we can obtain the deterministic version of this stochastic process, which characterize the evolution in time of the mean values:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \mu I \\ \frac{dR}{dt} &= \mu I \end{aligned} \tag{1.2}$$

where, the parameters

- β is the transmission rate or the rate at which susceptible individuals become infected.
- μ is the recovery rate or the rate at which infected individuals recover and become immune.

The first equation describes the rate at which susceptible individuals become infected. It assumes that the rate of transmission is proportional to the product of the number of susceptible individuals and the number of infectious individuals, with the proportionality constant β .

The second equation describes the rate at which infected individuals recover. It assumes that the rate of recovery is proportional to the number of infectious individuals, with the proportionality constant μ . The term βSI represents the rate at which susceptible individuals become infected, and the term μI represents the rate at which infected individuals recover.

The third equation describes the rate at which individuals recover from the disease and become immune. It assumes that the rate of recovery is proportional to the number of infectious individuals, with the proportionality constant μ .

Sometimes, the equations are written in terms of populations instead of probabilities. You can obtain one from the other by multiplying (dividing) by the population size N :

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{I}{N} S \\ \frac{dI}{dt} &= \beta \frac{S}{N} I - \mu I \\ \frac{dR}{dt} &= \mu I\end{aligned}\tag{1.3}$$

In this case, $\frac{I}{N}$ and $\frac{S}{N}$ can be interpreted as the density of infectious/susceptibles individuals, and β is still the rate of transmission contacts per unit time. The *force of infection* of a disease is the rate at which susceptible individuals become infected. Said differently, is the probability per unit time of a healthy individual of contracting the disease. In transmissible diseases will be $\beta \frac{I}{N}$ and it depends on the size of the infectious population.

These equations are a set of coupled first-order ordinary NON-LINEAR differential equations, which means that the values of S , I , and R change over time according to the values of β , μ , and the initial conditions. The nonlinearity, of course, causes some complications.

First epidemics models, however, were linear. For instance, if a model studies a disease that appears in the population at a fixed rate, let's say, skin cancer (non transmissible), then a deterministic model looks like:

$$\begin{aligned}\frac{dS}{dt} &= bS - \beta S - dS \\ \frac{dI}{dt} &= \beta S - \mu I - d_I I \\ \frac{dR}{dt} &= \mu I - d_R R\end{aligned}\tag{1.4}$$

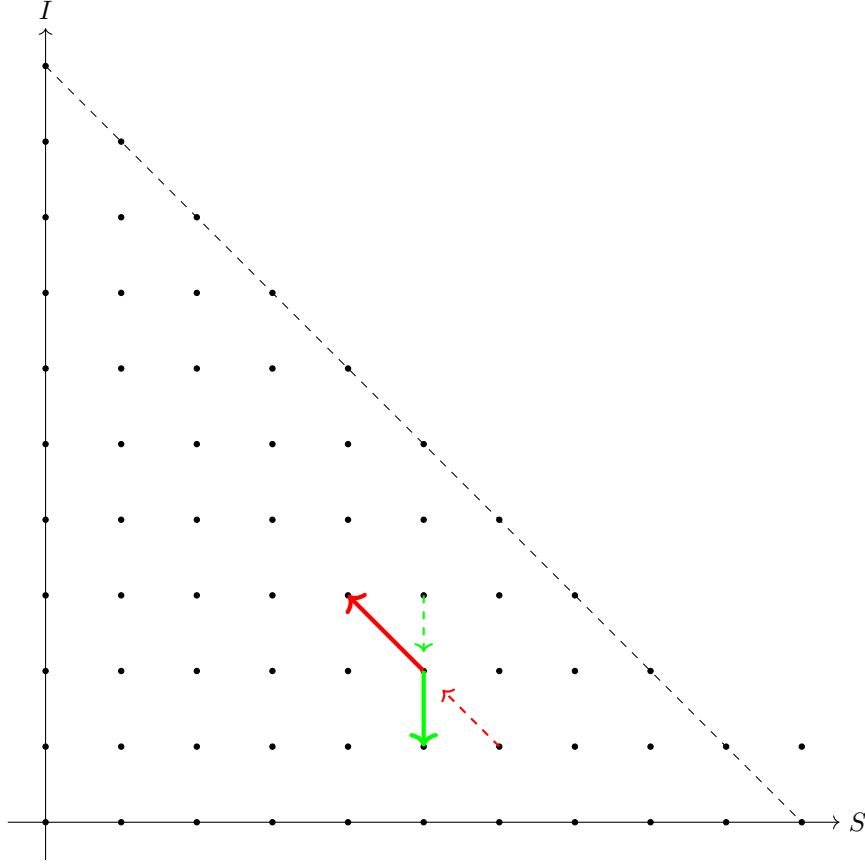
Notice that these equations are linear, therefore an scaling by N or any other factors is irrelevant.

1.3.1 Deterministic epidemics from master equation

Previous deterministic equations are similar in spirit to the master equation, but actually simpler and quite different in nature. While the stochastic process needs a full description of the probabilities of every possible state of the system, the SIR equations describe deterministic mean values of this process. Intuitively, following the law of large numbers, one expects that such large epidemics will be better described by the deterministic case.

Let us show how deterministic equations are obtained by averaging the master equation. Consider the SIR epidemic process. The full space of the stochastic process is given by the dots in the figure,

laying in the area $S + I \leq N$.



The red arrow exemplifies the contagion process, and the green arrow the recovery process. So, from any given state (S, I) there are two possible transitions

$$(S, I) \xrightarrow{W_I} (S-1, I+1) \quad (S, I) \xrightarrow{W_R} (S, I-1)$$

with transition rates given by

$$W_I = \beta \frac{I}{N} S \quad W_R = \mu I$$

The master equations reads

$$\frac{dP_t(S, I)}{dt} = \beta \frac{I-1}{N} (S+1) P_t(S+1, I-1) + \mu (I+1) P_t(S, I+1) - \beta \frac{I}{N} S P_t(S, I) - \mu (I+1) P_t(S, I)$$

Multiplying by S and tracing over S and I

$$\begin{aligned} \frac{d\langle S \rangle}{dt} &= \beta \sum_{S, I} \frac{I-1}{N} S (S+1) P_t(S+1, I-1) + \mu \sum_{S, I} S (I+1) P_t(S, I+1) - \\ &\quad - \beta \sum_{S, I} \frac{I}{N} S^2 P_t(S, I) - \mu \sum_{S, I} S I P_t(S, I) \end{aligned}$$

It suffices to assume that $P(S, I) \equiv 0$ whenever either of the arguments is negative or $S + I > N$, to change variables and obtain:

$$\begin{aligned} \frac{d\langle S \rangle}{dt} &= \beta \sum_{S, I} \frac{I}{N} (S-1) S P_t(S, I) + \mu \sum_{S, I} S I P_t(S, I) - \\ &\quad - \beta \sum_{S, I} \frac{I}{N} S^2 P_t(S, I) - \mu \sum_{S, I} S I P_t(S, I) \\ &= -\beta \frac{1}{N} \langle SI \rangle \end{aligned}$$

Repeating for the average of I , we get the equations

$$\begin{aligned} \frac{d\langle S \rangle}{dt} &= -\beta \frac{1}{N} \langle SI \rangle \\ \frac{d\langle I \rangle}{dt} &= +\beta \frac{1}{N} \langle SI \rangle - \mu \langle I \rangle \end{aligned}$$

which are similar but yet not the same as the SIR model equations. A further assumption $\langle SI \rangle \simeq \langle S \rangle \langle I \rangle$ is required to reproduce SIR model. This assumption shall be good if the variables S and I are actually independent, which is not mathematically the case. But it shall be a good approximation if the distribution $P_t(S, I)$ is sharply peaked around its mean values. This is the case for large systems, where we expect a sort of law of large numbers to concentrate the stochastic process on its typical values.

1.3.2 Short time limit

In the early stages of an epidemic, when the number of infected individuals is small compared to the total population size, the SIR model can be approximated by a linear differential equation. Suppose that the initial number of infected individuals is small compared to the total population size, i.e., $I(0) \sim 1/N$, where N is the total population size. Then, the number of susceptible individuals can be approximated as $S(t) \simeq 1$, and the number of recovered individuals can be approximated as $R(t) \sim 1/N$.

Under these assumptions, the dynamics of the infected compartment can be approximated by the following linear differential equation:

$$\frac{dI}{dt} = \beta S(t) I(t) - \mu I(t), \quad (1.5)$$

where β is the transmission rate, μ is the recovery rate, and $I(t)$ is the fraction of infected individuals at time t . During the initial epidemic phase, when few cases still exist compared to the population size, we can approximate $S(t) = S(0) = 1$, and write the linear equation:

$$\frac{dI}{dt} = \lambda I(t), \quad \text{where } \lambda = \beta - \mu. \quad (1.6)$$

The solution to this linear differential equation is an exponential function of the form:

$$I(t) = I(0) \exp(\lambda t). \quad (1.7)$$

The exponential solution of the SIR model in the low prevalence linear phase shows that the number of infected individuals changes exponentially, growing when $\beta > \mu$, and decreasing when $\beta < \mu$.

This exponential growth phase is characteristic of the beginning of an epidemic, when the number of infected individuals is small compared to the total population size and the epidemic has not yet reached a significant proportion of the population.

The basic reproduction number R_0 , is the expected number of new infections created by every single infected individual, in the early exponential phase of an epidemic. Since μ is the recovery rate, $\tau = 1/\mu$ is the expected duration of the infection on an individual. Since β is the rate of new cases, $R_0 = \beta\tau = \beta/\mu$ is the average number of new cases created by an infected individual during the period in which it is actively infectious. Notice that the criteria $\lambda \leq 0$ is equivalent to $R_0 \leq 1$.

1.3.3 Infinite time limit

Properties of SIR equations. Since the total amount of studied subjects remain invariant, only their categories changes, $S(t) + I(t) + R(t) = 1$. This can also be proven by adding (1.3). Furthermore, since

$$S'(t) + I'(t) = -\mu I(t)$$

is a decreasing non negative function, it converges to a limit $S_\infty + I_\infty$ in $t \rightarrow \infty$. Moreover, it is easy to see that I_∞ can only be zero, so $S + I \rightarrow S_\infty$.

Integration of the previous equation results in

$$1 - S_\infty = \mu \int_0^\infty I(t) dt$$

The first number $S_0 + I_0 = 1$ comes from assuming that all the population is originally in the states S or I.

Dividing the equation for $S'(t)$ by $S(t)$, we get

$$\frac{d \log S(t)}{dt} = -\beta I(t) \quad \text{that integrating} \quad \log(S_\infty/S_0) = -\beta \int_0^\infty I(t) dt$$

This results in a relation for the final amount of susceptible individuals

$$\log(S_0/S_\infty) = \frac{\beta}{\mu}(1 - S_\infty) = R_0(1 - S_\infty)$$

that is called the *final size relation* for the epidemic. If the variables are kept extensive $\tilde{S}(t) = NS(t)$, $\tilde{I}(t)$..., the final relation reads

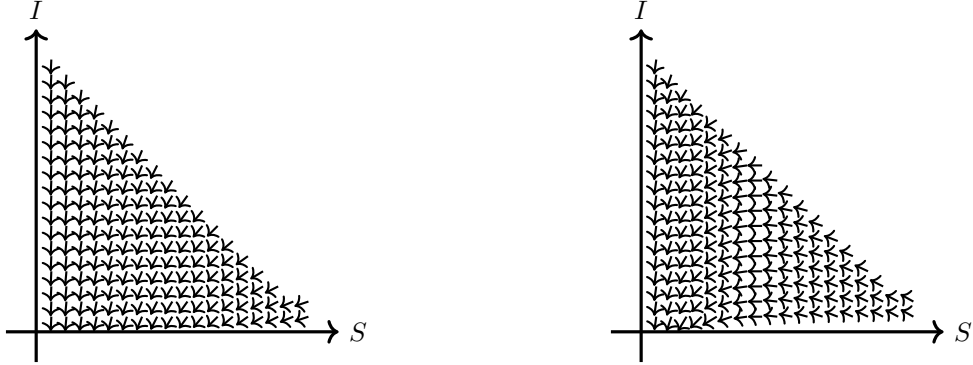
$$\log(\tilde{S}_0/\tilde{S}_\infty) = R_0(1 - \frac{\tilde{S}_\infty}{N}) \quad (1.8)$$

which is a useful relation between the basic reproduction number and the final size of the epidemic.

The total number of infected people in the population is given by $N - \tilde{S}_\infty$, and the fraction $1 - S_\infty$ or $1 - \frac{\tilde{S}_\infty}{N}$ is called the *attack rate*, although it is not a rate, but a fraction.

Another conclusion from (1.8) is that $S_\infty > 0$, and so, in every epidemic there is always a fraction of the population that will not be infected.

While the recovery rate μ is generally simple to estimate from studies on infected individuals. However the contact rate $\tilde{\beta} = \beta/N$ is much harder. Direct calculation of β , or of $R_0 = \beta/\mu$ can be difficult, while a post facto analysis of the epidemic through serological studies, can provide estimation for S_∞ , and through (1.8), to R_0 and β .



A similar procedure as the one we carried, but taken up to time t instead of ∞ , results in a parametric relation between $S(t)$ and $I(t)$:

$$I(t) + S(t) - \frac{\mu}{\beta} \log S(t) = 1 - \frac{\mu}{\beta} \log S_0 \quad (1.9)$$

This trajectories correspond to the flux lines in the diagram of I vs S , using the vector defined by the right hand side of (1.3), normalized.

The maximum number of infectives at any time is the number of infectives when the derivative of I is zero, that is, when $S = \mu/\beta$. This maximum is obtained by replacing this value into (1.9) as

$$I_{max} = I_0 + S_0 - \frac{\mu}{\beta} \log S_0 - \frac{\mu}{\beta} + \frac{\mu}{\beta} \log \frac{\mu}{\beta}$$

1.3.4 Epidemics terminology

A clear definition of terms is essential for the correct handling of epidemics. At the onset of any epidemic and in particular for emerging diseases, a clear statement of what is going to be consider a “case”, a “death caused by the disease”, a primary or secondary case, among others, is essential for a meaningful collection of statistical data, which is a key aspect of epidemiology.

Following WHO summary of terms for COVID (COVID-19, GLOSSARY: OUTBREAKS AND EPIDEMICS, A resource for journalists and communicators), here comes a selected list of useful terms related to epidemics:

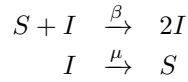
- Emerging disease: An unknown or newly appeared disease, usually of the infectious or communicable type.
- Reemerging disease: Resurgence or increase in the incidence of infectious or communicable diseases that were considered to be already under control.
- Seasonality of disease: Regular pattern of variation in the occurrence of some diseases (for example, the flu) during different seasons of the year.
- Incidence: The number of new cases of a disease in a population in a given period. Incidence measures the speed at which new cases occur during a given period in a specific population—for example, the number of new cases of HIV infection in gay men in a country in 2019.

- **Prevalence:** The total number of people who have a disease (new and existing cases) in a population or in a given place at a given time. Prevalence is an indicator of the magnitude of a disease—for example, the total number of people with tuberculosis in a country in 2019. Incidence and prevalence are essentially different ways of measuring the frequency of disease. The relationship between them varies from disease to disease.
- **Attack rate:** Number of people who contract a disease out of the entire group of people exposed to the disease. Expressed as a percentage.
- **Agent:** A microorganism, chemical substance, or form of radiation whose presence, excessive presence, or relative absence is essential for the occurrence of a disease. Agents may be biological (living organisms, such as viruses and bacteria) and non-biological (chemicals, such as pesticides, and physical phenomena, such as radiation).
- **Case-fatality rate:** The percentage of people affected by a disease or a particular event who die in a given period. Frequently used to describe the severity of an epidemic.
- **Contact:** A person who has been in contact with an infected person (case) such that he/she is considered to have had significant exposure and is therefore at risk of infection.
- **Contagious or infectious:** Terms that are often used interchangeably but that have subtle differences in meaning. “Contagious” is related to direct or indirect spread from person to person. The flu, for example, is very contagious, but Ebola is not. “Infectious” means that contact with a small amount of virus can cause disease. Ebola, for example, is very infectious.
- **Host:** A person or other living animal, including birds and arthropods, that affords subsistence or lodgment to an infectious agent under natural conditions.
- **Exposure:** Contact with an infectious agent or a risk factor that can cause a disease. Exposure has two dimensions: degree or level, and duration.
- **Incidence rate:** Rate of new cases in a population. The numerator is the number of new events that they occur in a given period. The denominator is the population at risk of experiencing the event of interest during this period.
- **Infectious period:** Period of time during which a person can transmit a disease. This period may precede the onset of symptoms and may last longer than the symptoms.
- **Latent period:** Time that elapses from exposure to an agent until the date that a person can transmit a disease (this is the period that immediately precedes the infectious period).
- **Mortality rate:** Percentage of people in a population who die out of the total population. May be expressed per 100, 1,000, or another number.
- **Virulence:** The ability of an infectious agent to produce severe and fatal cases. The measure of virulence is the ratio of the number of severe and fatal cases to total overt cases.
- **Basic reproduction number (R_0):** The spread of a disease in a population (especially in epidemics) depends on the basic reproduction number (or R_0) and the generation time (usually established on the basis of the serial interval). The R_0 is the number of people to whom an infected person can transmit the disease or the number of secondary cases that each primary case generates on average (during the contagious period).

- Generation time: the time between the onset of infection in the primary case and the onset of infection in a secondary case (i.e., a case infected by the primary case)
- Serial interval: In practical terms, this is the time between the onset of the disease in the primary case and the onset of the disease in the secondary case: the longer the serial interval, the more time there is to act on the problem, implement prevention and control measures, and, therefore, the better the chances of containing the epidemic. The shorter the generation time, the more difficult it is to control the outbreak.
- Isolation means separating sick or infected people from others to prevent infection from spreading.
- Quarantine means restricting the movement of healthy people who may have been exposed to a virus but are not sick.
- Herd immunity: Indirect protection against disease that results from a sufficient number of individuals in a community having immunity to that disease. With enough immune individuals, the transmission of a disease can be reduced, thus limiting the potential for any one individual to be exposed to it. Generally, PAHO recommends vaccination coverage of 95% or more to ensure herd immunity. Herd immunity does not apply to diseases such as tetanus that are not spread via person-to-person contact.

1.4 SIS model and epidemic thresholds

SIR epidemics mimic diseases that generate immunity, that is, that you can not take twice. This is the case of many (most) diseases, but it is not the case of many others, as, typically, sexually transmitted diseases, as chlamydia and gonorrhea. These diseases are actually best represented by SIS types of models. These are characterized by the transitions



and the corresponding deterministic equations

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \mu I \\ \frac{dI}{dt} &= \beta SI - \mu I \end{aligned} \tag{1.10}$$

While the epidemic phase $I > 0$ in an SIR model last a finite period of time, recurrent diseases can sustain epidemics for infinite time, normally referred as endemic phase. While SIS model is also non linear, the steady state can be easily deduced by setting to zero the time derivatives:

$$\begin{aligned} \frac{dS}{dt} \equiv 0 &= -\beta SI + \mu I \\ \frac{dI}{dt} \equiv 0 &= \beta SI - \mu I \end{aligned} \tag{1.11}$$

resulting in the equilibrium solution

$$S_{eq} = \frac{\mu}{\beta} \quad I_{eq} = 1 - S_{eq} = \frac{\beta - \mu}{\beta}$$

Of course this values will be meaningful if they are in the $[0, 1]$ range, meaning that only the case $\beta > \mu$ will generate an endemic phase. The value $\beta/\mu = 1$, corresponding also with $R_0 = 1$, defines the threshold above/below which endemic states do/do not exist.

Sometimes, SIS is a too simplistic approach, since almost every disease generates some level of immunity, although it can be lost in time.

1.5 And beyond

Reasons why deterministic description of epidemics is a very narrow approach:

- Well mixed systems are non realistic. Human behavior is a mixture of repetitive patterns (go to school, visit your parents, etc) and random interactions (in the public transport system), also influenced by distance constraints. The network of human contacts is best modelled by a dynamic network with rather stable clusterized subnetworks and random links whose probability decays with distance.
- Fluctuations: as the whole approach relies implicitly in the law of large numbers, fluctuations are completely ignored. Epidemic processes like SIS have an absorbing state, i.e. $I = 0$. Even epidemics above the epidemic threshold can be suppressed by fluctuations.
- Implies memoryless processes. The assumption that recovery times follow an exponential distribution is motivated by its mathematical convenience rather than its epidemiological plausibility.

1.6 Homework

1. Find out the incidence in a year and the prevalence of HIV in Cuba, China, India, USA, Haiti, Nigeria, Spain.
2. Consider the linear model in (1.4) as a model for skin cancer.
 - (a) What are reasonable values for the constants b and d ?
 - (b) Assume that half of people with skin cancer get cure, and half die of this condition. What can we say about μ and d_I ?
 - (c) Force of infection (not actually infection, but developing the disease), lets say is $\beta = 10^{-4}$. And for those who die of this disease, the average time from contraction to death is 1 year. Compute how many individuals are currently living with skin cancer.
3. Show that the number of infectious individuals always go to zero in the long time limit of the SIR model.
4. Consider $\theta = \beta/\mu$ and using

$$I = N - R - S_0 e^{-\theta R(t)}$$

show that the maximum number of infectious people occurs exactly at $S(t) = 1/\theta$. Show also that in the long time limit not everybody was infected. **Comment:** The first result can be derived in many different ways, some easier than the one proposed.

Chapter 2

Introduction to networks

Definition: A *graph* G is a pair (V, E) , where V is a set of vertices (or nodes), and E is a set of edges, where each edge is a pair of vertices $(u, v) \in V \times V$.

- The vertices are said to be adjacent if they are endpoints of an edge.
- The set of all edges incident to a vertex v is called the *neighborhood* of v , denoted by ∂v .

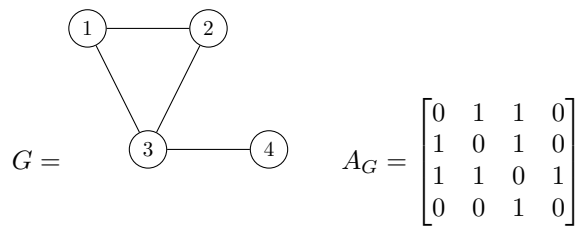
The *adjacency matrix* A_G of a graph G is:

- a square matrix of size $n \times n$, where $n = |V|$,
- if $(u, v) \in E$, then $A_G(u, v) = 1$
- else $A_G(u, v) = 0$

Consider the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$ and

$$E = \{(1, 2), (1, 3), (2, 3), (3, 4)\}$$

. The graph G and its adjacency matrix A_G are shown below:



Degree of a Node: The *degree* of a node v in a graph $G = (V, E)$ is the number of edges incident to v , i.e., the number of neighbors of v . The degree of v is denoted by $\deg(v)$. In a directed graph, the degree of a node is the sum of its in-degree (number of incoming edges) and out-degree (number of outgoing edges).

Neighborhood of a Node: The *neighborhood* of a node v in a graph $G = (V, E)$ is the set of all nodes that are adjacent to v , i.e., the set of all neighbors of v . The neighborhood of v is denoted by ∂v .

For example, in the previous graph, the degree of node 1 is 2, since it has 2 neighbors (nodes 2 and 3), i.e., $\deg(1) = 2$. Similarly, the degree of node 2 is 2 and the degree of node 3 is 3. The degree of node 4 is 1, since it has only 1 neighbor (node 3).

The neighborhood of node 1 is $\partial_1 = \{2, 3\}$, since nodes 2 and 3 are adjacent to node 1. Similarly, the neighborhood of node 2 is $\partial_2 = \{1, 3\}$, the neighborhood of node 3 is $\partial_3 = \{1, 2, 4\}$, and the neighborhood of node 4 is $\partial_4 = \{3\}$.

The adjacency matrix can be used to compute various properties of the graph, such as the number of paths of a certain length between two vertices, the number of cycles of a certain length in the graph, and many others.

Paths: A *path* in a graph $G = (V, E)$ is a sequence of distinct vertices v_1, v_2, \dots, v_k such that $(v_i, v_{i+1}) \in E$ for $i = 1, 2, \dots, k-1$. The length of the path is $k-1$, and it is denoted by $|P|$. A path is said to be *simple* if all its vertices are distinct.

Cycles: A *cycle* in a graph $G = (V, E)$ is a simple path v_1, v_2, \dots, v_k such that $(v_k, v_1) \in E$. The length of the cycle is k , and it is denoted by $|C|$.

Adjacency Matrix and Paths: The adjacency matrix A_G of a graph G can be used to compute the number of paths between two nodes u and v in the graph. The entry $(A_G)^n(u, v)$ gives the number of paths of length n between nodes u and v . Specifically, the (u, v) entry of A_G^n gives the number of paths of length n from node u to node v .

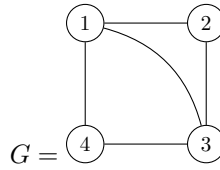
To see why this is true, consider the (u, v) entry of A_G^2 . This entry is given by $(A_G^2)_{u,v} = \sum_{w \in V} A_G(u, w)A_G(w, v)$, which counts the number of paths of length 2 from u to v by summing over all possible intermediate nodes w . By induction, we can show that $(A_G)^n(u, v)$ counts the number of paths of length n between nodes u and v .

In the previous graph, to compute the number of paths of length 2 between nodes 1 and 4, we can compute $(A_G)^2(1, 4)$, which is equal to $[(A_G)^2]_{1,4} = A_G(1, 2)A_G(2, 3) + A_G(1, 3)A_G(3, 4) = 1 \cdot 1 + 1 \cdot 1 = 2$. Therefore, there are two paths of length 2 between nodes 1 and 4 in the graph G .

The adjacency matrix of a graph G can also be used to compute the number of cycles of a certain length k in the graph. Specifically, the (i, j) entry of the matrix A_G^k gives the number of cycles of length k that contain the edge (i, j) .

To see why this is true, consider the (i, j) entry of A_G^k . This entry is given by $(A_G^k)_{i,j} = \sum_{w_1, w_2, \dots, w_{k-1}} A_G(i, w_1)A_G(w_1, w_2) \cdots A_G(w_{k-1}, j)$, which counts the number of paths of length k from i to j by summing over all possible intermediate nodes w_1, w_2, \dots, w_{k-1} . If the edge (i, j) is part of a cycle of length k , then we have $w_1, w_2, \dots, w_{k-1} \neq i, j$ and the last term $A_G(w_{k-1}, j)$ in the sum will be equal to 1 (since (w_{k-1}, j) is an edge in the cycle), while all other terms in the sum will be nonzero. Therefore, the (i, j) entry of A_G^k will count the number of cycles of length k that contain the edge (i, j) .

For example, let's consider the following graph G :



The adjacency matrix of G is:

$$A_G = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

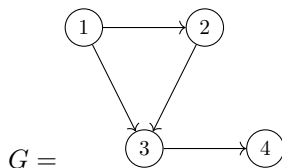
To compute the number of cycles of length 3 in the graph G , we can compute the matrix A_G^3 and count the number of entries that are nonzero. Since we are interested in cycles of length 3, we only need to consider the entries of A_G^3 that correspond to edges in the graph G . For example, the entry $(1, 2)$ of A_G^3 gives the number of cycles of length 3 that contain the edge $(1, 2)$. We have:

$$\begin{aligned}(A_G^3)_{1,2} &= \sum_{w_1, w_2} A_G(1, w_1)A_G(w_1, w_2)A_G(w_2, 2) \\ &= A_G(1, 4)A_G(4, 3)A_G(3, 2) \\ &= 1 \cdot 1 \cdot 1 \\ &= 1\end{aligned}$$

Therefore, there is exactly one cycle of length 3 in the graph G that contains the edge $(1, 2)$. Similarly, we can compute the other entries of A_G^3 to count the number of cycles of length 3 that contain the other edges in the graph. In this case, we find that there are three cycles of length 3 in the graph G .

2.1 Directed graphs

No problem, here's a directed version of the graph you requested:



The adjacency matrix of G is:

$$A_G = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

To compute the in-degree and out-degree of each vertex in the graph G , we can sum the entries of the corresponding rows and columns in the adjacency matrix. We have:

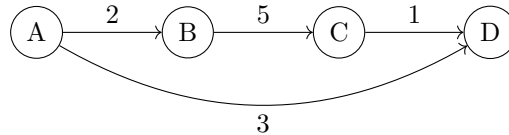
- Vertex 1: in-degree = 0, out-degree = 2
- Vertex 2: in-degree = 1, out-degree = 1
- Vertex 3: in-degree = 2, out-degree = 1
- Vertex 4: in-degree = 1, out-degree = 0

In this case, vertex 1 and vertex 4 are sources and sink, respectively.

2.2 Wheighted graphs

A weighted graph is a graph where each edge has a numerical value associated with it, called a weight. The weight often represents some sort of cost, distance, or importance associated with traversing the edge.

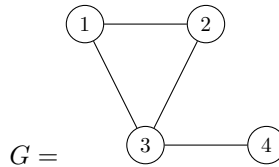
In mathematical terms, a weighted graph is represented by a graph $G = (V, E, w)$, where V is the set of vertices (nodes), E is the set of edges, and $w : E \rightarrow \mathbb{R}$ is a function that assigns a weight to each edge.



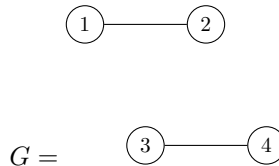
In this example, we define a weighted graph with four nodes (A, B, C, and D) and four edges with weights (2, 5, 1, and 3). In this graph, the weights on the edges represent the cost or distance associated with traversing the edge. For example, to get from node A to node D, we could either take the edge with weight 2 and the edge with weight 1, or we could take the edge with weight 3. The choice of which path to take would depend on the specific application of the graph.

2.3 Graphs properties

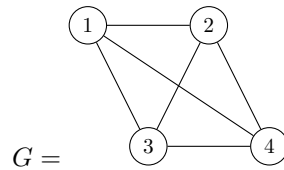
1. **Connected vs Disconnected Graphs:** A connected graph is a graph where there is a path between any two vertices in the graph. A disconnected graph is a graph that is not connected, meaning that there are at least two vertices that are not connected by a path. For example, the following graph is connected:



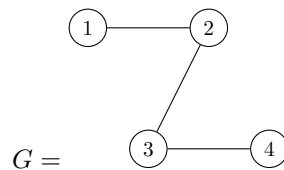
while the following graph is disconnected:



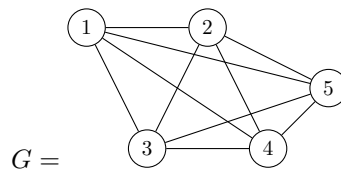
2. **Regular vs Non-Regular Graphs:** A regular graph is a graph where all vertices have the same degree, meaning that each vertex has the same number of neighbors. A non-regular graph is a graph where the degrees of the vertices are not all the same. For example, the following graph is regular:



while the following graph is non-regular:

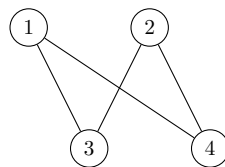


3. Planar vs Non-Planar Graphs: A planar graph is a graph that can be drawn on a plane without any edges crossing. A non-planar graph is a graph that cannot be drawn on a plane without any edges crossing. For example, the following previous non regular graph is planar, while the following graph is non-planar:



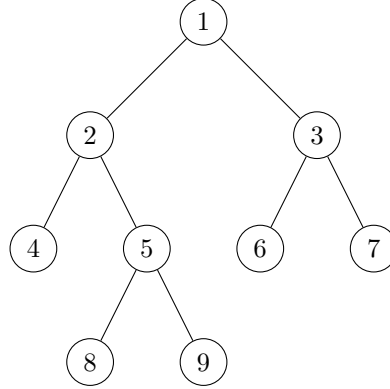
Less trivially, the previous example of a regular graph is also planar, although it doesn't seem so. Check that moving node 4 to the inside of the triangle created by the other three nodes, you can actually paint it without crossing lines.

4. Bipartite Graphs: A bipartite graph is a graph where the vertices can be divided into two sets such that there is no edge between vertices in the same set. The previous example of a non-planar graph, is also a non-bipartite graph. On the contrary, the following graph is bipartite with sets $\{1, 2\}$ and $\{3, 4\}$:



5. Trees: A tree is a connected graph with no cycles, meaning that there is exactly one path

between any two vertices in the graph. For example, the following graph is a tree:



Trees are important in many applications, such as in computer science where they are used to represent hierarchical structures like file systems or organization charts.

2.4 Graph metrics

Centrality measures are used to quantify the importance of individual nodes in a network. There are several types of centrality measures, including degree centrality, betweenness centrality, and eigenvector centrality. Degree centrality is based on the number of edges incident to a node, while betweenness centrality measures the fraction of shortest paths in the network that pass through a node. Eigenvector centrality takes into account the centrality of a node's neighbors, and assigns higher centrality to nodes that are connected to other highly central nodes.

2.4.1 Density of a graph and mean degree

For a graph $G(V, E)$, density is defined as

$$D = \frac{|E|}{|V|}$$

Graphs with $D \ll 1$ are named sparse.

The average degree of a graph is given by

$$\langle k \rangle = \sum_k kP(k) = \frac{2|E|}{|V|} = 2D$$

Second moment (variance):

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2 = \sum_k (k - \langle k \rangle)^2 P(k)$$

A graph is called heterogeneous if

$$\frac{\sigma_k}{\langle k \rangle} \gg 1$$

2.4.2 Diameter, radius, eccentricity

The eccentricity $\epsilon(v)$ of a vertex v is the greatest distance between v and any other vertex; in symbols,

$$\epsilon(v) = \max_{u \in V} d(v, u).$$

It can be thought of as how far a node is from the node most distant from it in the graph.

The radius r of a graph is the minimum eccentricity of any vertex or, in symbols,

$$r = \min_{v \in V} \epsilon(v) = \min_{v \in V} \max_{u \in V} d(v, u)$$

The diameter d of a graph is the maximum eccentricity of any vertex in the graph. That is, d is the greatest distance between any pair of vertices or, alternatively,

$$d = \max_{(v \in V)} \epsilon(v) = \max_{(v \in V)} \max_{(u \in V)} d(v, u)$$

To find the diameter of a graph, first find the shortest path between each pair of vertices. The greatest length of any of these paths is the diameter of the graph.

A central vertex in a graph of radius r is one whose eccentricity is r —that is, a vertex whose distance from its furthest vertex is equal to the radius, equivalently, a vertex v such that $\epsilon(v) = r$.

A peripheral vertex in a graph of diameter d is one whose eccentricity is d —that is, a vertex whose distance from its furthest vertex is equal to the diameter. Formally, v is peripheral if $\epsilon(v) = d$.

2.4.3 Jordan centrality measure

The center (or Jordan center) of a graph is the set of all vertices of minimum eccentricity, that is, the set of all vertices u where the greatest distance $d(u, v)$ to other vertices v is minimal. Equivalently, it is the set of vertices with eccentricity equal to the graph's radius:

$$\{v_1, \dots, v_n\} | \forall_{i \in 1 \dots n} \epsilon(v_i) = r$$

Thus vertices in the center (central points) minimize the maximal distance from other points in the graph.

2.4.4 Degree centrality

Degree centrality is a measure of the importance of a node in a network based on the number of edges incident to the node. Mathematically, the degree centrality C_d of a node i in a network G is defined as:

$$C_d(i) = \frac{k_i}{n-1}, \quad (2.1)$$

where k_i is the number of edges incident to node i , and n is the total number of nodes in the network. The degree centrality ranges from 0 to 1, with higher values indicating greater importance.

Degree centrality is a simple and intuitive measure of node importance, and is commonly used in social network analysis, biological networks, and other types of networks. However, it does not take into account the importance of the nodes to which a node is connected, and may not be the most appropriate measure for all types of networks.

2.4.5 Betweenness

Betweenness centrality is a measure of the importance of a node in a network based on the number of shortest paths that pass through the node. Mathematically, the betweenness centrality C_b of a node i in a network G is defined as:

$$C_b(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}, \quad (2.2)$$

where σ_{st} is the total number of shortest paths between nodes s and t , and $\sigma_{st}(i)$ is the number of shortest paths that pass through node i . The sum is taken over all pairs of nodes s and t that are not equal to node i .

The betweenness centrality ranges from 0 to 1, with higher values indicating greater importance. Nodes with high betweenness centrality are important for maintaining communication and facilitating the flow of information in the network.

The calculation of betweenness centrality can be computationally intensive for large networks, but efficient algorithms have been developed to approximate the betweenness centrality of nodes in large networks.

2.4.6 Modularity

Modularity is a measure of the division of a network into dense clusters or communities. So the value of the modularity is not associated to a graph by it self, but to a graph and a partition of it's nodes into different communities. Communities are specified by numbers $s_i \in \{1 \dots, k\}$, such that s_i is the number of the community that node i belongs to.

Mathematically, the modularity Q of a network G with a partition s of nodes into k communities is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(s_i, s_j), \quad (2.3)$$

where A_{ij} is the weight of the edge between nodes i and j , k_i and k_j are the degrees of nodes i and j , respectively, m is the total weight of edges in the network, and $\delta(s_i, s_j)$ is the Kronecker delta function that equals 1 if nodes i and j belong to the same community, and 0 otherwise.

The modularity ranges from -1 to 1, with higher values indicating a more modular network structure. A modularity of 0 indicates that the partition of nodes into communities is no better than chance. A modularity near 1 (homophily) indicate that nodes tend to connect to others of the same community, rather than a different one. Modularity near -1 indicates the opposite (paraphily).

The modularity can be used to evaluate the goodness of fit of different community detection algorithms and to compare the modular structure of different networks.

The calculation of modularity can be computationally intensive for large networks, but efficient algorithms have been developed to approximate the modularity of networks with millions of nodes and edges.

2.4.7 Clustering

The clustering coefficient is a measure of the degree to which nodes in a network tend to cluster together. It is defined as the fraction of the node's neighbors that are also neighbors of each other. Mathematically, the clustering coefficient C_i of a node i in a network G is defined as:

$$C_i = \frac{2e_i}{k_i(k_i - 1)}, \quad (2.4)$$

where e_i is the number of edges between the neighbors of node i , and k_i is the degree of node i , i.e., the number of edges incident to node i . The clustering coefficient ranges from 0 to 1, with higher values indicating that nodes in the network tend to form tightly interconnected clusters or communities. A low clustering coefficient indicates a more random and dispersed network structure.

The average clustering coefficient C of a network G is defined as the average of the clustering coefficients of all nodes in the network:

$$C = \frac{1}{n} \sum_{i=1}^n C_i, \quad (2.5)$$

where n is the total number of nodes in the network.

The clustering coefficient is a useful tool for characterizing the local structure of a network, and can be used to identify nodes that are part of tightly interconnected clusters or communities. However, it does not provide information about the global structure of the network or the connections between clusters.

These concepts are important tools for analyzing the structure and function of complex networks, and have applications in a wide range of fields, including social network analysis, biological networks, and transportation networks.

2.5 Assortativity

2.6 Laplacian of a graph

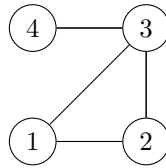
The Laplacian of a graph is a matrix that encodes important information about the structure of the graph. The Laplacian is defined as the difference between the degree matrix and the adjacency matrix of the graph. The degree matrix is a diagonal matrix whose entries are the degrees of each vertex, while the adjacency matrix is a binary matrix that indicates which pairs of vertices are connected by an edge.

Given a graph G with n vertices, the Laplacian matrix L is defined as:

$$L = D - A$$

where D is the degree matrix and A is the adjacency matrix of G . The Laplacian matrix is an $n \times n$ symmetric positive semidefinite matrix, and it has several important properties that make it useful in the study of graph theory and related fields.

Here's an example of how to compute the Laplacian matrix for a small graph with 4 vertices:



The adjacency matrix A and degree matrix D for this graph is:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The Laplacian matrix L is:

$$L = D - A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

The Laplacian of the graph encodes important information about the connectivity and structure of the graph. For example, the number of connected components of the graph is equal to the number of eigenvectors of the Laplacian with eigenvalue 0. The Laplacian also plays an important role in spectral graph theory, where it is used to analyze the spectral properties of graphs and to design graph clustering algorithms.

2.7 Networks for benchmarking

1. Zachary's karate club network: This is a social network of a karate club that was studied by Zachary in 1977. The network consists of 34 nodes representing members of the club and 78 edges representing friendships between members. It is often used as a benchmark for community detection algorithms.

2. Erdős-Rényi random network: This is a random graph model proposed by Erdős and Rényi in 1959. In this model, a graph is generated by connecting nodes with a certain probability p . Erdős-Rényi random networks are often used as a benchmark for studying properties of random graphs.

3. Barabási-Albert preferential attachment network: This is a random graph model proposed by Barabási and Albert in 1999. In this model, a graph is generated by starting with a small number of nodes and adding new nodes one at a time, with each new node preferentially attaching to existing nodes with higher degree. Barabási-Albert networks are often used as a benchmark for studying properties of scale-free networks.

4. Les Misérables co-occurrence network: This is a network of co-occurrences of characters in the novel Les Misérables by Victor Hugo. The network consists of 77 nodes representing characters and 254 edges representing co-occurrences in the same chapter. It is often used as a benchmark for community detection algorithms.

5. Power grid network: This is a network of the power grid in the western United States. The network consists of 4,941 nodes representing power stations and 6,594 edges representing power lines. It is often used as a benchmark for studying properties of real-world networks.

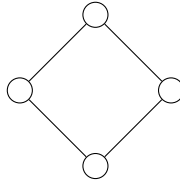
6. The Southern Women graph is a social network graph that represents the social interactions between women in a Southern US social club in the 1930s. The graph is an undirected graph with 32 nodes (women) and 89 edges (social interactions).

Each node in the graph represents a woman, and an edge between two nodes indicates that the corresponding women interacted socially at the club. The weight of each edge represents the number of social events at which the two women interacted.

2.7.1 Random graphs

Random graphs are mathematical models that are used to study the properties of graphs that arise from random processes. In a random graph, edges are created between nodes according to some probabilistic model, rather than being determined by some underlying structure or pattern.

Example. There are two possible regular graphs with 4 nodes and degree 2: the cycle graph C_4



and the disjoint union of two edges $K_2 \cup K_2$:



These are the only two possible regular graphs with 4 nodes and degree 2, as any other arrangement would either have nodes of degree greater than 2 or have disconnected nodes. An algorithm that randomly produces one of these two graphs, produces random regular graphs.

There are many different models of random graphs, each with their own set of assumptions and properties. Some common models include:

- Erdős-Rényi model: A model where each edge is present with a fixed probability, independently of all other edges.
- Watts-Strogatz model: A model where edges are rewired with a small probability, resulting in a graph with a high clustering coefficient but a short average path length.
- Barabási-Albert model: A model where new nodes are added to the graph one at a time, and each new node is connected to existing nodes with a probability proportional to their degree.

2.7.2 Random regular graphs

Random regular graphs are graphs where each vertex has the same degree. The algorithm to generate a random regular graph with n vertices and degree k is as follows:

1. If n is odd or k is odd and n is even, then no random regular graph exists. Otherwise, set $d = \frac{k}{2}$. 2. Create an empty graph G with n vertices. 3. For each vertex v , add an edge between v and the d vertices that appear next to it in a cyclic ordering of the vertices. 4. Shuffle the vertices randomly. 5. For each vertex v , rewire each of its d edges to a randomly chosen vertex, subject to the constraints that the resulting graph remains k -regular and does not contain self-loops or multiple edges.

The resulting graph is a random k -regular graph on n vertices.

Require: n : number of vertices, k : degree of each vertex

- 1: If n is odd or k is odd and n is even, then no random regular graph exists.
- 2: Otherwise, set $d = \frac{k}{2}$.
- 3: Create an empty graph G with n vertices.
- 4: **for** $i = 1$ to n **do**
- 5: **for** $j = 1$ to d **do**
- 6: Add an edge between vertex i and vertex $(i + j) \bmod n$.
- 7: Shuffle the vertices randomly.
- 8: **for** $i = 1$ to n **do**
- 9: **for** $j = 1$ to d **do**
- 10: Choose a random vertex v that is not adjacent to vertex i .
- 11: Remove the edge between vertex i and some neighbor of i that appears j positions after i in the cyclic ordering of vertices.

- 12: Add an edge between vertex i and vertex v .
- 13: Output the resulting graph G .

2.7.3 Erdos-Renyi model

The Erdős-Rényi (ER) algorithm is a popular method for generating random graphs. Given a set of n nodes and a probability p of any two nodes being connected by an edge, the ER algorithm generates a random graph by considering each possible edge and including it with probability p .

The ER algorithm can be implemented as follows:

1. Start with an empty graph G with n nodes.
2. For each pair of nodes i and j (excluding self-loops), add an edge between them with probability p . This can be done by generating a random number r between 0 and 1 for each pair of nodes, and adding an edge if $r \leq p$.
3. Output the resulting graph G .

Require: n : number of nodes, p : probability of an edge between any pair of nodes

- 1: Create an empty graph G with n nodes.
- 2: **for** $i = 1$ to $n - 1$ **do**
- 3: **for** $j = i + 1$ to n **do**
- 4: Generate a random number r between 0 and 1.
- 5: **if** $r \leq p$ **then**
- 6: Add an edge between nodes i and j .
- 7: Output the resulting graph G .

The ER algorithm generates a random graph with an expected number of edges equal to $p \cdot \binom{n}{2}$, which is the number of possible edges in the graph multiplied by the probability of each edge being included. The resulting graph is typically sparse for small values of p and becomes increasingly dense as p approaches 1.

2.7.4 Albert Barabasi graphs

The Barabási-Albert (BA) algorithm is a popular method for generating scale-free graphs, which are graphs with a degree distribution that follows a power law. The BA algorithm generates a graph by starting with a small number of nodes and adding new nodes one at a time, each with a fixed number of edges attached to existing nodes.

The BA algorithm can be implemented as follows:

1. Start with a small graph G with m_0 nodes, where m_0 is a parameter of the algorithm.
2. Add a new node i to the graph, and connect it to m existing nodes in the graph, where m is a parameter of the algorithm.
3. The probability that node i connects to node j is proportional to the degree of node j in the graph, i.e., nodes with higher degree are more likely to be connected to.
4. Repeat steps 2-3 until the desired number of nodes is reached.

Require: m_0 : number of initial nodes, m : number of edges to attach to each new node, n : final number of nodes in the graph

- 1: Create an empty graph G with m_0 nodes.
- 2: Create a list L of nodes, where each node appears m_0 times in the list.
- 3: **for** $i = m_0$ to $n - 1$ **do**
- 4: Create a new node v_i .

- 5: Choose m distinct nodes to connect v_i to, sampled from L with probability proportional to their degree.
- 6: Add edges between v_i and the chosen nodes.
- 7: Add m copies of v_i to L .
- 8: Output the resulting graph G .

The resulting graph generated by the BA algorithm has a degree distribution that follows a power law, which means that a few nodes have a very high degree while most nodes have a low degree. This property is characteristic of many real-world networks, and it has important implications for the behavior of these networks.

The BA algorithm is a simple and elegant way to generate scale-free networks, and it has been widely used in the study of network science and related fields.

2.8 Practical Python Class: networks

Graph metrics in Les Miserables

2.9 Practical Homework

Graph structure in written Chinese

Take the list of Chinese words. Define a graph G where each node is one Chinese character. Define a link between two nodes whenever there is at least one word with both characters.

1. What is the degree distribution in G ?
2. Show examples of the characters with highest and smallest degree.
3. Make a histogram of degree distribution? Is it heavytailed?
4. Define the same graph, but now directed, where direction is defined by the ordering of the characters in the words: $AB \Rightarrow (A \rightarrow B)$. Make the histogram of out-degree and in-degree for the directed graph.
5. Show examples of the characters with highest out and in degrees.
6. Make a histogram of the out and in degrees.
7. Define the unbalance, as the difference between out and in-degrees. Make a histogram of the unbalance.
8. Show examples of characters with the high positive and negative unbalances. Is there a linguistic reason for this?

2.10 Percolation and epidemics

Chapter 3

Monte Carlo simulation of epidemics

So, epidemics are stochastic processes. Some analytical results exist, for the simplest of cases, as the SIR or SIS dynamics. In more complex cases, simulations are the only tool at hand to study their dynamics.

As we mentioned in last chapter, epidemics can be seen as a set of stochastic reactions, turning healthy people into infectious, and turning them into recovered, at certain rates. Next we discuss how to simulate such types of processes in a computer.

3.1 Rates and exponential distributions

Exponential distribution

As you shall remember (see appendix) from your probabilities class, the exponential distribution is the continuous time limit of a process in which you flip a coin as many times as needed until you find the first “head” coming out ¹. It is clear from the nature of this process, that if a head has not shown up after we toss the coin 5 times, then the probability that it will appear in the next x shots, is the same probability that it had in any x shots, regardless of the fact that we already shot it 5 times. This results in a very important property of exponential distributions: they do not have memory.

The probability density of the exponential distribution is

$$f(x) = \lambda e^{-\lambda x}$$

The cumulative distribution function is therefore

$$F(x) = \int_{-\infty}^x f(x) = 1 - e^{-\lambda x}$$

(defined only for $x > 0$, with $F(x) = 0$ otherwise), and the mean value $\langle X \rangle = \frac{1}{\lambda}$.

memorylessness of exponential random variables

¹In appendix of this chapter, you will find a brief reminder of what conditional probability means, and the deduction of the continuous exponential distribution as a limit of a Bernoulli process, if you are interested to check.

Let's look at an interesting case. Let's prove that the exponential distribution has no memory. This means the following:

$$P(X > s + t | X > s) = P(X > t)$$

This property is called memorylessness because the probability of $X > s + t$ given that $X > s$ is independent of the value of s . Using the definition of conditional probability,

$$P(X > s + t | X > s) = \frac{P(X > s + t, X > s)}{P(X > s)}$$

But $P(X > s + t, X > s) = P(X > s + t)$, since the set $\{x : x > s + t\} \subset \{x : x > s\}$. Note that here the notation $P(A, B)$ corresponds to the probability of both A and B occurring simultaneously, that is, $P(A \cap B)$. Therefore,

$$P(X > s + t | X > s) = \frac{P(X > s + t)}{P(X > s)}$$

The probabilities $P(X > a) = 1 - F(a)$, so $P(X > s + t) = e^{-\lambda(s+t)}$ and $P(X > s) = e^{-\lambda s}$, from which

$$P(X > s + t | X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P(X > t)$$

thus concluding the proof.

This property of the exponential density has the following physical implication. The decay of a radioactive atom is a process that occurs with exponential probability. The lack of memory of this probability means that regardless of the past lifetime of an atom, its probability of decaying in the future is always a decreasing exponential.

Law of the minimum of independent exponential variables:

Another fundamental result for the following discussion is the following property of exponential r.v.:

$$\begin{cases} X_i \sim \text{Exp}(\lambda_i), \\ Y_i = \min_i X_i, \\ Z_i = \text{argmin}_i X_i \end{cases} \Rightarrow \begin{cases} Y \sim \text{Exp}(\sum_i \lambda_i) \\ \mathbb{P}(Z = i) = \frac{\lambda_i}{\sum_j \lambda_j} \end{cases}$$

The proof of which is left to the interested student.

3.2 Gillespie's Stochastic Simulation Algorithm

The Gillespie algorithm is a stochastic simulation algorithm that is often used to simulate chemical reaction systems. It simulates the time evolution of a reaction system by randomly selecting which reaction will occur next and when it will occur.

Consider the following chemical reaction system:



where A , B , C , and D are chemical species, and k_1 , k_2 , and k_3 are reaction rate constants.

The interpretation of this list of reactions is the following. We have a system with a certain amount of elements $X = (n_A, n_B, n_C, n_D)$ at time t . At any moment in time, any of the three reactions can take place. Let us call ΔX_r the vector of the effect that reaction r has on the system, in this case:

$$\Delta X_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \Delta X_2 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 2 \end{pmatrix} \quad \Delta X_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ -2 \end{pmatrix}$$

So, the stochastic process is actually defined by a sequence of growing times $t_{k-1} < t_k$ at which reactions occur, and the index of the reaction occurring:

$$\mathcal{P} = \{(t_1, r_1), (t_2, r_2)..\}$$

such that the stochastic variable is given by

$$X(t) = X(t_{k-1}) + \Delta X_{r_k}$$

where k is such that $t_{k-1} < t < t_k$. In other words, the t_k are the moments in which reactions occur, and at those points, the state of the system changes from $X(t)$ to $X(t+dt) = X(t^+) = X(t) + \Delta X_{r_k}$.

So, simulating such stochastic process amounts to defining a way to generate the stochastic sequence \mathcal{P} . Furthermore, this process is a Markov process, meaning that the values of (t_k, r_k) depend only on the current state of the system $X(t_{k-1}^+)$.

3.2.1 Gillespie's algorithm

To simulate this system using Gillespie's algorithm, we first define the propensity functions for each of the three reactions:

$$a_1 = k_1[A][B] \tag{3.4}$$

$$a_2 = k_2[C] \tag{3.5}$$

$$a_3 = k_3[D]^2[B] \tag{3.6}$$

where $[A]$, $[B]$, $[C]$, $[D]$, and $[E]$ are the concentrations of the corresponding chemical species.

We then randomly select the time τ until the next event occurs and the reaction that will occur next. The probability that reaction i will occur next is given by:

$$P_i = \frac{a_i}{\sum_j a_j} \tag{3.7}$$

where the sum is taken over all reactions.

Once we have determined which reaction will occur next, we update the concentrations of the chemical species according to the stoichiometry of the reaction. This is summarized in the algorithm 3.2.1.

This example illustrates how Gillespie's algorithm can be used to simulate the time evolution of a simple chemical reaction system. By defining the propensity functions and randomly selecting the next event, we can simulate the stochastic behavior of the system and gain insight into its dynamics.

Algorithm 1

Gillespie SSA

```

1:  $t = 0$ 
2:  $X(0) = X_0$  ▷ Set initial state
3: while  $t < t_{end}$  do
4:    $\text{Update}(a_i)$  ▷ Calculate each reactions propensity
5:    $a_0 = \sum_i a_i$  ▷ Total propensity
6:   for  $i = 1 \dots K$  do ▷ Compute probabilities for every reaction
7:      $P_i = \frac{a_i}{a_0}$ 
8:    $r_1 \sim U(0, 1)$  ▷ Generate random numbers,  $r_1$  and  $r_2$ 
9:    $r_2 \sim U(0, 1)$ 
10:   $\tau = \frac{1}{a_0} \ln \frac{1}{r_1}$  ▷ Time until the next reaction
11:   $k = i \mid \sum_k^{i-1} P_k < r_2 < \sum_k^i P_i$  ▷ Next reaction with probability  $P_i$ 
12:   $X(t + \tau) = X(t) + \Delta X_k$ 
13:  Set  $t = t + \tau$ 

```

3.3 Practical Python class

In this practical class we will program the Gillespie algorithm for stochastic simulations in Python. I will share with you a Jupyter Notebook.

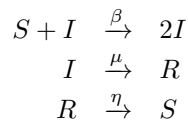
Once the algorithm is running, we will test it in three different models:

1. Drunken man model
2. SIR model
3. SIS model

In the first two cases, we will compare the results of the simulation with the known analytical solution.

3.4 Practical Homework

Consider the model SIRS, as follows:



1. Make a simulation of a system of $N = 1000$ individuals with parameters $\beta = 0.02$, and $\mu = \eta = 0.01$. Use as initial condition $(n_S, n_I, n_R) = (N/3, N/3, N/3)$.
2. By visual inspection in your previous simulation, define a value of T where the system is at equilibrium (approximately not changing).
3. Study the long time behavior of this system for a range of values of $\beta \in (0.001, 0.002, 0.003, \dots, 0.009, 0.01, 0.011, 0.012 \dots, 0.02)$

4. Write down the differential equations for this system.
5. Study the steady state analytically, by setting all time derivatives to zero, and solving for the other equilibrium values of S, I and R .
6. Compare to the long time study as a function of β .

Appendix: Conditional probability

The continuous conditional distribution has the usual expressions $P(A|B) = P(AB)/P(B)$. In the case of continuous random variables, the probability of any event is given by integrating the probability density over the set of points that constitute the event $P(A) = \int_A dx f(x)$, so that

$$P(A|B) = \frac{\int_{AB} dx f(x)}{\int_B dx f(x)}$$

The independence of events A and B is equivalent to either of the equalities

$$P(AB) = P(A)P(B) \quad P(A|B) = P(A)$$

Appendix: Exponential distribution

As you shall remember from your probabilities class, the exponential distribution is the continuous time limit of a process in which you flip a coin as many times as needed until you find the first “head” coming out. What follows is a deduction of such property.

Let’s consider the case of waiting time in a Bernoulli process with success probability q . Let $K \in \{1, 2, 3, \dots\}$ be the corresponding discrete random variable for the number of experiments that must be performed until the first successful result occurs. The distribution of K is given by:

$$P(K = k) = (1 - q)^{k-1} q$$

To understand the following steps, let’s think of K as the number of time intervals we must wait until we observe success. Now, suppose we want to discretize time into smaller intervals, and let’s focus on the variable $X = K/N$ where N is the number of intervals in which one second is divided. Let’s also consider, as is natural, that the divisions of the intervals do not affect the real probability of the phenomenon, so $q = q'/N$. In this way, the probability density of the variable X is obtained as the limit $N \rightarrow \infty$ from the variable K , known as the continuous limit:

$$\begin{aligned} P(X = k/N) &= P(K = k) = (1 - q)^{k-1} q \\ &= \left(1 - \frac{q'}{N}\right)^{Nx-1} \frac{q'}{N} \\ f(x) &= \lim_{N \rightarrow \infty} \frac{P(X = k/N)}{1/N} = q' e^{-xq'} \end{aligned}$$

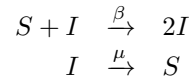
The resulting distribution is called the exponential density, and is usually expressed in terms of the parameter $\lambda = 1/q'$

$$f(x) = \frac{e^{-x/\lambda}}{\lambda}$$

Chapter 4

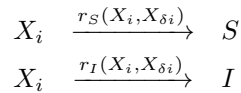
Mean field solution of epidemics on Networks

The well-mixed SIS model is represented by these “reactions”



and the dynamics variables corresponded to the ammount of elements of each specie: $\mathbf{n} = (n_S, n_I)$. In such model there is no consideration for the spatial structure of the epidemic, a factor that is fundamental in reality. A way to introduce space into epidemics is to study contact networks, *i.e.* a graph $G(V, E)$ of individuals $i \in \{1, \dots, N\}$, whose epidemic state is represented by $X_i(t) \in \{S, I, R\}$ or $X_i \in \{S, I\}$ in the SIS model, and whose possible interactions with other individuals is defined by the edges of the graph $(i, j) \in V$.

In what follows we focus on continuous time compartment epidemic models on networks. Now, instead of a global reaction with a global rate, we have a set of reactions for every given node of the graph:



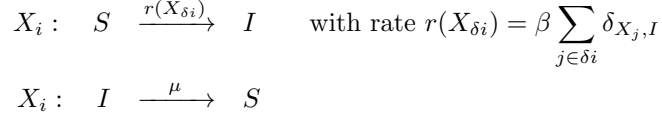
The standard susceptible-infectious-susceptible model (SIS) considers the nodes to be in either of two compartments (states):

- $X_i = 0 \equiv S \Rightarrow$ susceptible
- $X_i = 1 \equiv I \Rightarrow$ infectious

and is the simplest standard for recurrent transmissible diseases. The epidemic is thus a continuous time stochastic process with only two admitted transitions occurring at

- rate β , at which a link (i, j) can transmit the state 1 from node i to j ;
- rate μ at which state 1 decays to state 0 on any infectious node.

Explicitly the set of reactions on a single node are



assuming the simplest case of homegeous links in the network (all with the same β). The fact that

- transition rates are independent of time
- and infection rates are additive

are simplifications valid in the exponential case, but could be not valid (and are not) in general. For instance, the time a person is infecious from a given disease, has more a bell shape than an exponential.

These extensive set of reactions, define a transition rate over the entire stochastic system $\mathbf{X} = (X_i, \dots, X_N)$, susceptible of being treated with the same tools we studied before:

- Numerically: Gillespie's Stochastic Simulation Algorithm (SSA)
- Analytically: master equation

In this class we are going to study both approaches. An analytical treatment of the master equation will be in general impossible, and some approximations, under the general cathegory of mean field approximations, can be used instead.

4.1 Master equation

Any continuous time dynamic graphical model with binary variables $\sigma \in \{-1, 1\}$ (in our case $X \in \{S, I\}$), can be formally described by a master equation. The dynamics are fully defined by the rate $r_i(\sigma)$, which is the probability per unit of time that the variables change their states from $s_i \rightarrow -s_i$. The distribution $P(\sigma, t)$ is the probability that the system is in the state σ at time t , and the configuration space of this stochastic process is governed by the joint master equation:

$$\frac{dP(\sigma)}{dt} = - \sum_{i=1}^N \left[r_i(\sigma) P(\sigma) - r_i(F_i(\sigma)) P(F_i(\sigma)) \right], \quad (4.1)$$

where F_i represents the inversion operator on variable i , i.e. $F_i(\sigma) = \sigma_1, \dots, -\sigma_i, \dots, \sigma_N$ and we have omitted the time dependence of $P(\sigma) = P(\sigma, t)$ for simplicity of notation. This is a balance equation. The first term inside the summation accounts for transitions from the state σ of each of the elements of the system, and the second term accounts for transitions from any configuration of the system to the state σ .

This equation is exact, but it is impractical for treating medium-sized systems since it represents a set of 2^N (for an N -variable system) coupled differential equations. However, in a system of local interactions, where the connections are given by a graph and the transition rates $r_i(\sigma, t)$ depend only on the spin configuration i and its neighborhood ∂_i , the system can be described by differential equations of the individual probability $P(\sigma_i)$.

Taking into account the topology of the system, a master equation for the probability $P(\sigma_i)$ of each node in the system can be written as:

$$\frac{dP(\sigma_i)}{dt} = - \sum_{\sigma_{\partial i}} \left[r_i(\sigma_i, \sigma_{\partial i}) P(\sigma_i, \sigma_{\partial i}) - r_i(-\sigma_i, \sigma_{\partial i}) P(-\sigma_i, \sigma_{\partial i}) \right] \quad (4.2)$$

where ∂_i is the neighborhood of node i and $P(\sigma_i, \sigma_{\partial i})$ is the joint probability of a configuration of node i and its neighborhood. This new system of equations is still exact, and the probability $P(\sigma)$ can be calculated as the product of the individual probabilities $\left[\prod_{i=0}^N P(\sigma_i) \right]$ of all nodes in the system. But unlike equation (4.1), this set is not a closed set of differential equations, because the time evolution of the joint probabilities is not known.

4.2 Mean field approximations to the SIS model

Mean field approximations for deriving continuous-time differential equations for epidemic processes on contact networks are fundamentally based on considering correlations between the expected values of random variables up to a certain level of factorization, usually the first two.

Attempts to reduce the complexity start from factorizing the single master equation into many equations for each node marginals $P_i(X_i, t)$. Given that the X 's are two-state variables, $P_i(X_i)$ consists of two values $P_i(X_i = 0)$ and $P_i(X_i = 1)$, where we have identified the susceptible state as $X = 0$, and the infectious as $X = 1$. Furthermore, since $P_i(X_i = 0) = 1 - P(X_i = 1)$, only one value is enough to represent $P(X_i)$. In the particular choice of $S \rightarrow 0$ and $X \rightarrow 1$, it is the case that $P_i(X_i = 1) = E[X_i]$, such that the master equation can be written as an evolution of the expected values of the infectious state:

$$\frac{dE[X_i(t)]}{dt} = E \left[-\mu X_i(t) + \beta(1 - X_i(t)) \sum_{j=1}^N a_{ij} X_j(t) \right] \quad (4.3)$$

where a_{ij} are the elements of the adjacency matrix, meaning that $a_{ij} = 1$ if nodes i and j are neighbors ($(i, j) \in E$) and is zero otherwise.

However, the expectation value on the right hand side acts over products of variables $X_j(t)(1 - X_i(t))$, which requires a differential equation for the evolution of the correlations. Not surprisingly, the two point correlations functions depend on three point correlations and so on and so forth.

4.2.1 Individual-based mean field (IBMF) approximation

The simplest closure of equation eq. (4.3) is the individual-based mean field (IBMF) in which independence is assumed $E[X_i(t)X_j(t)] \approx E[X_i(t)] E[X_j(t)] \equiv \rho_i(t)\rho_j(t)$ (ρ_i is the probability that node i is infectious) and therefore eqs. (4.3) are now a closed set of non linear differential equations:

$$\frac{d\rho_i(t)}{dt} = -\mu \rho_i(t) + \beta[1 - \rho_i(t)] \sum_{j \in \partial i} \rho_j(t) \quad (4.4)$$

where we used the notation ∂i to represent the set of neighbors of node i .

4.2.2 Pair-based mean field (PBMF)

The second simplest way to close the equations is to consider a differential equation for the correlations between neighboring variables $X_i X_j$. We, therefore, need a master equation for the evolution of such products. Such equation is given by

$$\frac{dE X_i X_j}{dt} = E \left[-2\mu X_i X_j + \beta X_j (1 - X_i) \sum_{k=1}^N a_{ik} X_k + \beta X_i (1 - X_j) \sum_{k=1}^N a_{jk} X_k \right]$$

A reordering of the terms results in

$$\begin{aligned} \frac{dE [X_i X_j]}{dt} = & -2\mu E[X_i X_j] + \beta \sum_{k=1}^N a_{ik} E[X_j X_k] + \beta \sum_{k=1}^N a_{jk} E[X_i X_k] \\ & - \beta \sum_{k=1}^N (a_{jk} + a_{ik}) E[X_i X_j X_k] \end{aligned} \quad (4.5)$$

but a factorization is assumed for higher correlations. Different approaches have been used to approximate $E[X_i X_j X_k]$ in terms of smaller correlations. In this paper we will compare with the approximations proposed in [?] $E[X_i X_j X_k] \approx E[X_i X_j] E[X_k] \equiv \phi_{ij}(t) \rho_i(t)$:

$$\frac{d\rho_i(t)}{dt} = -\mu \rho_i(t) + \beta \sum_{j \in \partial i} \phi_{ij}(t) \quad (4.6)$$

$$\begin{aligned} \frac{d\phi_{ij}(t)}{dt} = & -(2\mu + \beta) \phi_{ij}(t) + \mu \rho_i(t) - \beta \phi_{ij}(t) \sum_{k \in \partial j \setminus i} \rho_k(t) \\ & + \beta [1 - \rho_i(t) - \phi_{ij}(t)] \sum_{k \in \partial i \setminus j} \rho_k(t) \end{aligned} \quad (4.7)$$

and [?] $E[X_i X_j X_k] \approx \frac{E[X_i X_j] E[X_j, X_k]}{E[X_j]} \equiv \frac{\phi_{ij}(t) \phi_{kj}(t)}{\rho_j(t)}$.

$$\begin{aligned} \frac{d\rho_i(t)}{dt} = & -\mu \rho_i(t) + \beta \sum_{j \in \partial i} \phi_{ji}(t) \\ \frac{d\phi_{ij}(t)}{dt} = & -(2\mu + \beta) \phi_{ij}(t) + \mu \rho_i(t) - \beta \phi_{ij}(t) / (1 - \rho_j(t)) \sum_{k \in \partial j \setminus i} \phi_{kj}(t) \\ & + \beta [1 - \rho_j(t) - \phi_{ij}(t)] / (1 - \rho_i(t)) \sum_{k \in \partial i \setminus j} \phi_{ki}(t) \end{aligned} \quad (4.8)$$

We will refer to the two different approximations shown in (4.7) and (4.8) as **PBMF-1** and **PBMF-2**, respectively.

In both approaches, IBMF and PBMF, the expected values evolving in time are intended to be expectations over different stochastic stories of the whole epidemic process. Therefore they are to be compared with averages over many Monte Carlo simulations of such process.

4.3 Practical Python Class: simulating epidemics on networks

Preclass: Install Epidemic on Networks (EoN) library on Python

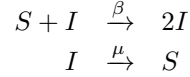
1. Simulate an SIR model on the Zachary club graph. Plot the results of $S(t)$, $I(t)$ and $R(t)$.
2. Make an animation of the process.

Homework

For this homework, do not code the epidemic process nor the numerical integration of IBMF and PBMF, yourself. Please base your code on the EoN library in python.

Task 1: Topologies and epidemic threshold for SIS.

As said in the conference, the SIS model corresponds to these reactions



We will study SIS on networks, for different values of the infectivity parameter β , and for fixed $\mu = 1.0$. It is known that for high values of β , the epidemic is sustained in time (call endemic), while for low values, the epidemic disappear. The critical value β_c is the smallest value that allows for a sustained epidemic, also known as endemic threshold.

1. Study an SIS epidemic for different values of R_0 in two topologies: Erdos-Renyi and scale free networks.

- Create two large networks ($N = 1000$) with the same average degree $\langle k \rangle = 6$, one with an Erdos-Renyi topology, and one with a scale free topology. For the latter, use an arbitrary exponent γ for the distribution, but adjust the parameters to have the same expected degree $\langle k \rangle = 6$.
- Show a histogram of the degree distribution in both networks. Compute the expected values $\langle k \rangle$ and $\langle k^2 \rangle$ for both networks.
- Use the EoN package in python to simulate epidemics on each of these networks, with starting conditions $I(0) = 200$ and $S(0) = 800$, randomly selected among the nodes of the network.
- Use the EoN package in python to compute the IBMF and PBMF numerical integration of the SIS epidemic on these networks.
- As an example, show the evolution in time of infected individuals for some β , in both networks, using all three approaches, the simulation, the IBMF and PBMF approximations.
- Program a function that, for a given value of β , computes the average number of infected individuals over many simulations (say 100 simulations). DO NOT USE IBMF/ PBMF, only simulations.
- Plot the average number of infected individuals at the end of simulations, for different values of β , for both types of networks.
- According to computations discussed in the lecture, the critical value should be related to the degree distribution of the network as

$$\lambda_c \equiv \frac{\beta_c}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle}.$$

Compare this prediction to the numerical results obtained.

- Do both topologies have the same epidemic threshold. Why?

1. Write down the Individual Based Mean Field equations for the SIR model on a network.
2. Consider SIS model on the Zachary Club graph, starting the epidemic from node 0.
 - Compute the average over many simulations of an epidemic.
 - Compare this average with the results of integrating IBMF and PBMF.
 - What is the average time of infection of node 19 in the simulations?
 - How does it compares with the moment in which probability of infection is 0.5, according to both approximations?

Research homework

Study an SIS epidemic model on the Les Misérables network.

1. Run many simulations for different R_0 values. Find approximately the epidemic threshold.
2. Study the epidemic slightly above the epidemic threshold.
3. Compute the fraction of time that each character is infected, in a long run.
4. What is the correlation of this time with: clustering coefficient, eccentricity, betweenness of each node?
5. What is the correlation of this time with the amount of times that each character is mentioned in the novel?

Chapter 5

Reconstruction of epidemics

Inference is something you already do. You already know, you are an expert. Except that you only do it intuitively.

Example 1 If I ask you the age of a person you have zero information about, well you might say it shall be anything between 0 and 120 years old. If I tell you is a friend of mine, well, you probably think that is more probable that his/her age is closer to mine. If I show you a picture, and the person has some gray hair, you move your probability distribution to the higher end of the range, say 50 – 70.

Example 2 If I ask you about the monthly income of a person, you might say that is something between 0 and 100 millions USD a month. If I then tell you that this person graduated from engineering, well, most probably he earns more than 500 USD a month. If I tell you that he lives in a third world country, chances are higher for a salary below 2000 USD. If I tell you that lives in India, and that all his friends are rich, that probably raises your salary expectation for this person.

Example 3 If a little kid (still not talking) comes with fever and crying, there is a lot bunch of diseases that might be the cause. If the mother tells you that in his kindergarten some kids have been suffering from stomach diseases, well, you can not be sure, but chances are higher for this to be the case.

In all these cases, there is a prior probability distribution, the one you have in advance of any new information coming to you (that humans live between 0 and 120 years old, most of the time). In all of these cases, that probability distribution changes after you receive new information. You are doing inference.

Mathematically, inference is the procedure by which you update probabilities of certain stochastic variables when you get incomplete information about them. To treat this mathematically, all you need to do is Bayes theorem

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \quad \boxed{\text{Posterior}} = \frac{\boxed{\text{Likelihood}} \cdot \boxed{\text{Prior}}}{\boxed{\text{Evidence}}} \quad (5.1)$$

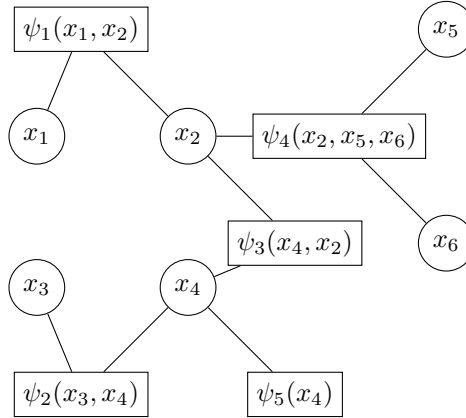
Terms in the Bayes formula have their own names. $P(A)$ is the *a priori* distribution, that is, what you know prior to any new information, while $P(A|B)$ –also a probability of event A – is the posterior distribution, what you know *a posteriori* that event B –named the evidence– took place. The term $P(B|A)$ is called likelihood, and $P(B)$ is the probability of the evidence, or just evidence for short.

5.1 Inference in graphical models

The algorithm we care about today has been derived in many different context, many times. It goes by the name of Sum-Product algorithm, or by the name of Belief-Propagation. It is an algorithm for inference in graphical models.

A graphical model is a type of probabilistic model that represents the dependencies between variables using a graph structure. The variables are represented as nodes in the graph, and the relationships between them are represented as edges. For instance a probabilistic model among 6 variables $x_1 \dots, x_6$ given by the interaction functions $\psi_1, \psi_2, \psi_3, \psi_4$ as

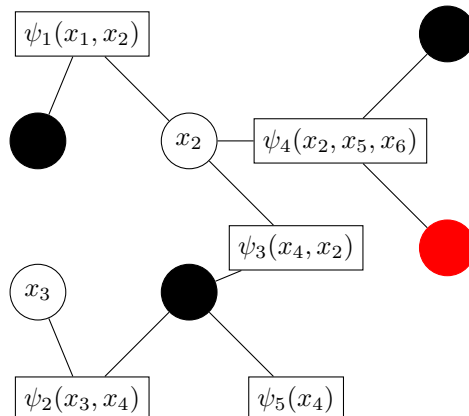
$$P(x_1 \dots, x_6) \propto \psi_1(x_1, x_2) \psi_2(x_3, x_4) \psi_3(x_4, x_2) \psi_4(x_2, x_5, x_6) \psi_5(x_4)$$



In this example, we have six variable nodes, x_1, x_2, x_3, x_4, x_5 and x_6 , and four five nodes, given by the functions ψ_a . The variable nodes represent the random variables in the model, and the factor nodes represent the factors that depend on these variables. The edges indicate the dependencies between the nodes. This is necessarily a bipartiate graph/.

Example: Graph coloring

Say that every variable takes on three different colors $x_i \in \{white, red, black\} \equiv \{0, 1, 2\}$. A valid coloring of a graph is such that interacting variables do not have the same color, as in this case



How do you define a probabilistic measure that concentrates on top of valid colorings? Well, just define

$$\begin{aligned}\psi_{1,2,3}(x, y) &= (1 - \delta_{x,y}) \\ \psi_4(x, y, z) &= (1 - (1 - (1 - \delta_{x,y})(1 - \delta_{y,z})(1 - \delta_{x,z}))) \\ \psi_5(x) &= 1\end{aligned}$$

and $P(x_1 \dots, x_6)$ will be zero in all non valid colorings, and will be $1/C$ in all C valid colorings.

Notice, *en passant*, that if you could compute $P(x_1, \dots, x_6)$ you could immediately know how many colorings there are, by just inverting any non zero value of $P(\dots) = 1/C$.

The relevance of graphical model is that often they are easy to write down, meaning, it is easy to propose a meaningful set of local interactions that define your particular problem. However, the full probability distribution is of little use, for the following reasons:

1. Most of the time, you want general properties (macro states), not a microscopic description of the system
2. Evaluation of $P(\underline{x})$ requires a normalization constant, and that needs a trace over all configurations of the system, which typically is exponentially many
3. Sometimes you care about specific variables $P_i(x_i) = \sum_{\underline{x} \setminus x_i} P(\underline{x})$, that also requires an unfeasible sum.

The first issue relates to the definition of observables in statistical mechanics, while the second relates to the partition function. The third issue is less typical of statistical mechanics, but yet, connected to the approximations we typically use to solve the first two.

5.1.1 Supersonic intro to statistical mechanics

In statistical mechanics, the partition function and the free energy are important quantities that describe the thermodynamic properties of a system. How does that have anything to do with inference?

Well, maybe it will be clearer later, but for the time being let us say that thermodynamic properties equals general macroscopic properties of a system, and statistical mechanics make the connection between those general properties and the details of the interactions in the system.

Consider, for instance, a graphical model as the previous one. Since the overall probability of the system is given by a product of functions, it is convenient (more on this later) to write each function as an exponential:

$$\psi_a(\underline{x}^a) = e^{-\beta E_a(\underline{x}^a)}$$

where $\underline{x}^a = \{x_{a_1}, x_{a_2} \dots\}$ is the set of variables that appear in the given function. In this expression β is just a scale, and you can think of the energy $E_a(x_{\partial a})$ as a cost function. The bigger the energy of a given configuration, the less probable that configuration is.

This means that these types of models can be defined in terms of an additive energy function $E(\underline{x}) = \sum_a E_a(\underline{x}^a)$ as

$$P_B(\underline{x}) = \frac{1}{Z} e^{-\beta E(\underline{x})} \quad (5.2)$$

The partition function Z is a function of the scale parameter β (equivalent to an inverse temperature), and is defined as the sum or integral over all possible states of the system, weighted by their Boltzmann factors:

$$Z = \sum_{\text{states}} e^{-\beta E}, \quad (5.3)$$

For continuous systems, the sum is replaced by an integral over the phase space of the system.

In general, the free energy can be split into two parts: the internal energy U , which is the average energy of the system, and the entropy S , which is a measure of the disorder or randomness of the system:

$$F = U - \frac{1}{\beta} S. \quad (5.4)$$

and the energy U and entropy S of a system are given as macroscopic averages

$$U = \sum_{\text{states}} P_B(\underline{x}) E(\underline{x}) \quad S = - \sum_{\{\underline{x}\}} P_B(\underline{x}) \ln P_B(\underline{x}), \quad (5.5)$$

where the sum runs over all configurations of the microscopic variables \underline{x} , *i.e* all the states.

By direct substitution of eq. (5.2) into the definition of F , we can relate F and Z as

$$F = -\frac{1}{\beta} \ln Z$$

5.1.2 Free energy as Kullback-Leibler distance

The Kullback-Leibler (KL) divergence, also known as the relative entropy, is a way to measure the difference between two probability distributions $p(x)$ and $q(x)$. The KL divergence is defined as:

$$D_{\text{KL}}(q||p) = \int q(x) \ln \frac{q(x)}{p(x)} dx, \quad (5.6)$$

where $q(x)$ and $p(x)$ are the two probability distributions over some variable x . The KL divergence measures the amount of information lost when approximating the true distribution $p(x)$ with the approximate distribution $q(x)$. It has the following properties:

1. $D(q||p)$ is convex in $q(x)$;

2. $D(q||p) \geq 0$;
3. $D(q||p) > 0$ unless $q(x) \equiv p(x)$

Since it is a convex non-negative quantity that is zero if and only if $q(x) = p(x)$ ¹, it is a good starting point to find approximations, in the sense that, the lower the KL distance, the better the approximation.

Through the glasses of statistical mechanics, the true distribution is the Boltzmann distribution (5.2). Simple calculations relate the KL divergence to a fundamental statistical mechanics function: the free energy. Just replace $p(\underline{x})$ by its Boltzmann definition

$$D_{\text{KL}}(q||P_B) = \sum_{\text{states}} q(x) (\beta E(\underline{x}) + \ln Z) \ln q(x) dx \quad (5.7)$$

$$= \beta \sum_{\text{states}} E(\underline{x}) q(x) + \sum_{\text{states}} q(x) \ln q(x) + \ln Z \quad (5.8)$$

$$= \beta \left(G[q(x)] + \frac{1}{\beta} \ln Z \right) \quad (5.9)$$

The functional

$$\begin{aligned} G[q(x)] &= \sum_{\text{states}} E(\underline{x}) q(x) - \frac{1}{\beta} \sum_{\text{states}} q(x) \ln q(x) \\ &= \langle E(\underline{x}) \rangle - \frac{1}{\beta} S[q(x)] \end{aligned}$$

its a weighted balance between the average energy $\langle E \rangle$ and the entropy

$$S[q] \equiv - \sum_{\text{states}} q(x) \ln q(x)$$

of the distribution $q(\underline{x})$, called Gibbs free energy. Since when $q(\underline{x}) \equiv P_B(\underline{x})$ the KL distance is zero, we get

$$G[P_B(\underline{x})] = -\frac{1}{\beta} \ln Z \equiv F \quad (5.10)$$

where the latter F is called free energy.

Going back to equation (5.9), it is evident that the minimum of the KL distance is found at the minimum of the Gibbs functional $G[q(\underline{x})]$, since Z is not a function of q .

Key message: When looking for probabilities distributions of a model:

- if you are a physicist, you minimize the Gibbs functional,
- if you are an information theoretician, you minimize KL divergence.

This relationship between the KL divergence and the free energy is useful in the context of computational physics and machine learning, where the KL divergence can be used as a measure of the quality of an approximate distribution compared to the true distribution. By minimizing the KL divergence, one can find the optimal approximate distribution that is closest to the true distribution, and therefore obtain a more accurate representation of the system.

¹almost everywhere

5.1.3 Mean field approximations

For instance, let us say we are interested in the previous graphical model. We do not know/want to compute the exact distribution. We could search for an approximated factorized distribution

$$q_{MF}(\underline{x}) = \prod_{i=1}^6 p_i(x_i)$$

that resembles the most to the full unknown distribution

$$p(\underline{x}) = \frac{1}{Z} \psi_1(x_1, x_2) \psi_2(x_3, x_4) \psi_3(x_4, x_2) \psi_4(x_2, x_5, x_6) = \frac{1}{Z} e^{-\beta \sum_{a=1}^6 \sum_{\underline{x}_a} E_a(\underline{x}_a)}$$

This results in the following Gibbs functional

$$G[q_{MF}(\underline{x})] = \sum_a \langle E_a(\underline{x}_a) \rangle + \frac{1}{\beta} \sum_i \sum_{x_i} q_i(x_i) \ln q_i(x_i)$$

Extremization of such functional results in

$$q_i(x_i) = \frac{1}{Z_i} e^{-\beta \sum_{a \in \partial i} \sum_{x_j | j \neq i, j \in \partial a} E_a(\underline{x}_a) \prod_j q_j(x_j)}$$

Why exponentials?

When you first study molecular physics, or statistical mechanics, you get that weird feeling that Boltzmann distribution and Entropy are connected some how. But is not clear which comes first. Why the equilibrium distribution has to be an exponential of the energy? Is this causing the shape of the entropy function or viceversa?

I haven't set my mind on this completely, but the way I understand it is that the functional form of the entropy is the single form that has a set of properties expected from the entropy, as extensiveness, positive definite, and additive over factorized distributions. So, most probably, entropy come first, causing the Boltzmann distribution.

Besides this chicken/egg controversy, there is some hand waving argument for the need of a Boltzmann factor with an extensive energy in the exponent. Since extremes of the Gibbs functional define equilibrium, and since it is a balance between energy and entropy, well, for both magnitudes to be comparable the expected value of $E(\underline{x})$ has to be of the order of the expected value of $\ln q(\underline{x})$, meaning that $q(\underline{x}) \sim e^{E(\underline{x})}$. Let me put an even more hand waving argument. Entropy is given by combinatorial terms, that grow very fast with the number of particles. Therefore, if disorder is going to have a non trivial interplay with temperature, you need to counter balance the disorder with something that grows as fast.

Only exponential of the energy makes sense. If it is not exponentiated, combinatorics always win!!

Mean field, details

Actually, we neede to introduce a Lagrange multiplier to force normalizaiton of the distributions when we did the Mean Field approach. So, instead of extremizing the Gibbs free energy

$$G[q_{MF}(\underline{x})] = \sum_a \langle E_a(\underline{x}_a) \rangle + \frac{1}{\beta} \sum_i \sum_{x_i} q_i(x_i) \ln q_i(x_i)$$

We actually need to extremize a Lagrange Function:

$$L[q_{MF}, \lambda_1, \dots, \lambda_N] = G[q_{MF}(\underline{x})] - \sum_i^N \lambda_i \left(\sum_{x_i} q_i(x_i) - 1 \right)$$

Extremization of such functional results in

$$q_i(x_i) = e^{1-\lambda_i/\beta} e^{-\beta \sum_{a \in \partial i} \sum_{x_j | j \neq i, j \in \partial a} E_a(\underline{x}_a) \prod_j q_j(x_j)}$$

The extremization over the Lagrange multipliers enforces the condition, so λ_i will take the required value for $q_i(x_i)$ to be normalized.

5.2 Beyond Mean Field: Belief Propagation

Mean field is quite poor. Essentially it loses all correlations in the system. We can do better, specially in trees. Lets see how.

By the way in which our graphical models are constructed, each factor node corresponds to a probability distribution of a group of variables in the absence of others, so independently, we have:

$$\forall_a p_a(\underline{x}_a) = \frac{1}{z_a} e^{-\beta E_a(\underline{x}_a)}$$

Can we build our full probability out of these factor probabilities? Yes indeed!

Bayes rule

As you shall remember from your probability course, the conditional probability is defined as

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

and is understood as the probability of having A given that B occurred.

So, if you have two sets of variables, that have a non null intersection, you can always write

$$p(\underline{x}_{a \cup b}) = p(\underline{x}_a | \underline{x}_b) p(\underline{x}_b)$$

for the joint probability. Furthermore, you can write

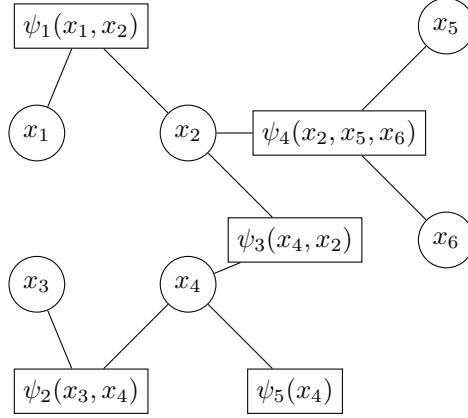
$$p(\underline{x}_{a \cup b}) = \frac{p(\underline{x}_a) p(\underline{x}_b)}{p(\underline{x}_{a \cap b})}$$

So you can iteratively merge your entire graph (as long as it is a tree) in order to write

$$P(\underline{x}) = \prod_a p_a(x_a) \prod_{a \cap b \neq \emptyset} \frac{1}{p(\underline{x}_{a \cap b})} \quad (5.11)$$

This is cool, since we have a new family of possible distributions to use as a test in Gibbs-KL minimization!!

Let us assume, for notation simplicity, that intersections between factors nodes happen always at a single variable, as is the case in the graphical model



where, for instance $\underline{x}_{1 \cap 3} = x_2 = \underline{x}_{1 \cap 4}$. Since the same set of variables $\underline{x}_{a \cap b}$ can arise from different a 's and b 's, it is more commonly written

$$P(\underline{x}) = \prod_a p_a(\underline{x}_a) \prod_i \frac{1}{p_i(x_i)^{d_i-1}} \quad (5.12)$$

where d_i is the degree of variable i .

Bethe free energy

Notice that any graphical model, not necessarily a tree, is susceptible of being approximated this way. In the case of trees, it is an exact representation of the probability, while in the rest of graphs, it is an approximation. To mark the idea that it not necessarily is exact, we call the local probability distributions as beliefs and replace $p_a(\underline{x}_a) \rightarrow b_a(\underline{x}_a)$.

If we replace the family (5.12) into the Gibbs free energy, we get the so called Bethe free energy is

$$G_{BP}[\{b_R(\mathbf{s}_R)\}] = \sum_a G_a[b_a(\underline{x}_a)] + \sum_i (1 - d_i) G_i[b_i(x_i)] \quad (5.13)$$

where d_i is the number of factors (functions ψ_a) that the i -th variable belongs to.

The distributions $b_a(\underline{x}_a)$ and $b_i(x_i)$ are the variables over which we minimize the free energy functional, and should approximate the marginals of the Boltzmann. For this reason, the functional minimization is not done considering the independent functions $b(\underline{x})$, but they are linked by the marginalization condition

$$\forall_a \forall_{i \in a} b_i(x_i) = \sum_{\underline{x}_{a \setminus i}} b_a(\underline{x}_a). \quad (5.14)$$

Constrained minimization

This relations can easily be enforced with a set of Lagrange multipliers, as we reminded you at the beginning. The Lagrange functional is

$$L[\{b_R(\mathbf{s}_R), \{m_{a,i}(x_i)\}\}] = G_{BP}[\{b_R(\mathbf{s}_R)\}] + \sum_a \sum_{i \in a} \sum_{x_i} m_{a,i}(x_i) \left(b_i(x_i) - \sum_{\underline{x}_{a \setminus i}} b_a(\underline{x}_a) \right)$$

Extremization of the Lagrange function produces the following relations

$$b_a(\underline{x}_a) = \frac{1}{z_a} e^{-\beta E_a(\underline{x}_a)} \prod_{j \in a} \prod_{b \in \partial j \setminus a} m_{b,j}(x_j) \quad (5.15)$$

$$b_i(\underline{x}_i) = \frac{1}{z_i} e^{-\beta E_i(\underline{x}_i)} \prod_{b \in \partial i} m_{b,i}(x_i) \quad (5.16)$$

relating the local probability distributions to the values of the local energy and the lagrange multipliers $m_{a,i}(x_i)$.

Message Passing equations

As usual, extremization with respect to the Lagrange multipliers reproduces the constraints they are enforcing, namely eq. (5.14). After writing this conditions in terms of eqs. (??), we get the following relation among Lagrange multipliers:

$$m_{a \rightarrow i}(x_i) = \sum_{\underline{x}_a \setminus i} \frac{1}{z_a} e^{-\beta(E_a(\underline{x}_a))} \prod_{j \in a \setminus i} \prod_{b \in \partial j \setminus a} m_{b,j}(x_j) \quad (5.17)$$

usually understood as **message passing** equations.

While the approximation is exact in a tree, and is very good in a graphs with large loops, in systems with short cycles the results are usually much worse.

5.2.1 A simple example: coloring trees

Let us go back to the **Graph coloring** example. We already with this functions

$$\begin{aligned} \psi_{1,2,3}(x, y) &= (1 - \delta_{x,y}) \\ \psi_4(x, y, z) &= (1 - (1 - (1 - \delta_{x,y})(1 - \delta_{y,z})(1 - \delta_{x,z}))) \\ \psi_5(x) &= 1 \end{aligned}$$

the probability measure

$$P(x_1 \dots, x_6) \propto \psi_1(x_1, x_2) \psi_2(x_3, x_4) \psi_3(x_4, x_2) \psi_4(x_2, x_5, x_6) \psi_5(x_4)$$

concentrates only on the valid colorings.

But, this measure doesn't look as we the exponential of an energy, since the $\psi_a(\cdot)$ functions aren't equivalent to exponentials of local interactions. First, this is, in itself not a problem. Although we wrote the BP equations in the physical way, we can directly replace the energy by the functions ψ_a , as in

$$m_{a \rightarrow i}(x_i) = \sum_{\underline{x}_a \setminus i} \psi_a(\underline{x}_a) \prod_{j \in a \setminus i} \prod_{b \in \partial j \setminus a} m_{b,j}(x_j) \quad (5.18)$$

But still, you might want not to do that. In that case, you can define

$$\begin{aligned} E_{1,2,3}(x, y) &= \delta_{x,y} \\ E_4(x, y, z) &= \delta_{x,y} + \delta_{y,z} + \delta_{x,z} \\ E_5(x) &= 0 \end{aligned}$$

and use a high value of β in your algorithm. The energy terms now are zero for valid local configurations of colors, and 1 for invalid ones. Since β is big, the exponentials

$$e^{-\beta\psi_a(\cdot)} = \begin{cases} 1 & \text{coloring is valid} \\ e^{-\beta} \ll 1 & \text{coloring is invalid} \end{cases}$$

which for practical purposes is equivalent to the hard constraint definition.

5.2.2 A clever way to count

In particular in the tree coloring problem above, message passing equations are completely trivial, since there is no bias towards any particular color, all messages turn out to be flat

$$m_{a \rightarrow i}(x_i) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$$

for the case of three-colors.

Yet, there are two possible ways in which we could use the BP solution. First, we could compute the entropy

$$S = \sum_a S[b_a(x_a)] + \sum_i (1 - d_i) \sum_{x_i} S[b_i(x_i)] \quad \text{where } S[b] = - \sum_x b(x) \ln b(x)$$

If you evaluate in this case, you get $S = \ln 48.0$, which coincide with the 48 valid configurations, and the fact that the entropy of an uniform distribution is the log of the number of states.

You could also repeat the message passing and the entropy calculation, but forcing one node, let's say x_4 to have a particular color. You can do that by simply changing ψ_5 :

$$\psi_5(x) = (1, 0, 0)$$

In that case, you get $S = \ln 16$, again consistent with the number of valid configurations.

5.3 The zero patient problem

Note: this is an extraction of the paper :

F. Altarelli. et. al., "Bayesian Inference of Epidemics on Networks via Belief Propagation, ", Phys. Rev. Lett. 112, 118701 –Published 17 March 2014

The SIR model on graphs The susceptible-infected-recovered (SIR) model of spreading is a stochastic dynamical model in discrete time defined over a graph $G = (V, E)$. At time t a node i can be in one of three states (*infected*, *susceptible* or *recovered*) represented by a variable $x_i^t \in \{S, I, R\}$. The dynamical process is defined as follows: at each time step t , each node i in the infected state attempts to infect each one of its susceptible neighbors $\{j \in \partial i : x_j^t = S\}$, succeeding with independent probabilities $p_{ij} \in]0, 1]$ (which we assume known); then, node i attempts to recover, succeeding with probability $r_i \in [0, 1]$ (also assumed known). The dynamics is clearly irreversible, as a given node can only undergo the transitions $S \rightarrow I \rightarrow R$, and thus a stationary condition will be reached in finite time (with all nodes either in the S or R state provided $r_i > 0$). Two important special cases of SIR are: **(a)** the Independent Cascade (IC) model, obtained when $r_i \equiv 1$, which has been extensively studied in the related context of optimization of the spread of information [?] and a particular sub-case of which is *bootstrap percolation* with threshold $\theta = 1$

(obtained when $p_{ij} \equiv 1$); **(b)** the susceptible-infected (SI) model, obtained when $r_i \equiv 0$. It should be noted that the IC model describes completely the SIR model for arbitrary r_i in the infinite time limit, while the infinite time limit of the SI model is rather trivial, as all nodes in the connected component of an initially infected node will be infected at time $t \rightarrow \infty$. Nevertheless, the evolution in time can show rich and interesting characteristics.

Self-consistent equations for spread computation. Assume that a certain set of nodes starts the infection at time $t = 0$, i.e. with $x_i^0 = I$. The SIR model can be re-stated as following: recovery times g_i and delay parameters s_{ij} are drawn respectively from $\mathcal{G}_i(g_i) = r_i(1 - r_i)^{g_i}$ and $\omega_{ij}(s_{ij}|g_i) = p_{ij}(1 - p_{ij})^{s_{ij}}$ for $s_{ij} \leq g_i$ and $\omega_{ij}(\infty|g_i) = \sum_{s > g_i} p_{ij}(1 - p_{ij})^s$; i.e. the recovery time follows a geometric distribution and the delay in transmission follows a geometric distribution truncated by g_i . Afterwards, the infection times t_i can be computed deterministically as the ones satisfying $1 = \phi_i = \delta(t_i, \mathbb{I}[x_i^0 \neq I] (\min_{j \in \partial i} \{t_j + s_{ji}\} + 1))$ for every i . This representation makes it possible to write the distribution of infection and recovery times \mathbf{t}, \mathbf{g} given the initial state $\mathcal{P}(\mathbf{t}, \mathbf{g}|\mathbf{x}^0) = \sum_{\mathbf{s}} \mathcal{P}(\mathbf{s}|\mathbf{g}) \mathcal{P}(\mathbf{t}|\mathbf{x}^0, \mathbf{g}, \mathbf{s}) \mathcal{P}(\mathbf{g})$ giving

$$\mathcal{P}(\mathbf{t}, \mathbf{g}|\mathbf{x}^0) = \sum_{\mathbf{s}} \prod_{i,j} \omega_{ij} \prod_i \phi_i \mathcal{G}_i \quad (5.19)$$

Bayesian inference in the SIR model Assume to know the state of an epidemics at time $t = T$, i.e. the state of every node $x_i^T \in \{S, I, R\}$ but to possess only probabilistic prior information about the set of initially infected nodes; specifically, assume that nodes are initially infected with i.i.d. *a priori* probability $\mathcal{P}(\mathbf{x}^0) = \prod_i \gamma_i(x_i^0)$. Starting with $\mathcal{P}(\mathbf{x}^T|\mathbf{t}, \mathbf{g}) = \prod_i \xi_i(t_i, g_i, x_i^T)$ where $\xi_i = \mathbb{I}[x_i^T = I, t_i \leq T < t_i + g_i] + \mathbb{I}[x_i^T = S, t_i < T] + \mathbb{I}[x_i^T = R, t_i + g_i \leq T]$, the posterior distribution $\mathcal{P}(\mathbf{x}^0|\mathbf{x}^T) \propto \sum_{\mathbf{t}, \mathbf{g}} \mathcal{P}(\mathbf{x}^T|\mathbf{t}, \mathbf{g}) \mathcal{P}(\mathbf{t}, \mathbf{g}|\mathbf{x}^0) \mathcal{P}(\mathbf{x}^0)$ can be expressed thanks to (5.19) as

$$\mathcal{P}(\mathbf{x}^0|\mathbf{x}^T) \propto \sum_{\mathbf{t}, \mathbf{g}, \mathbf{s}} \prod_{i,j} \omega_{ij} \prod_i \phi_i \mathcal{G}_i \gamma_i \xi_i \quad (5.20)$$

Belief Propagation equations for the likelihood function For the derivation of the Belief Propagation (BP) equations we will rewrite (5.20) by introducing $t'_j = t_j + s_{ji}$. Define $\phi_{ij} = \omega_{ij}(t'_j - t_i|g_i) \omega_{ji}(t'_j - t_j|g_j)$ and $\phi_i = \delta(t_i; \mathbb{I}[x_i^0 \neq I] (\min_{j \in \partial i} \{t'_j\} + 1))$; then

$$\mathcal{P}(\mathbf{x}^0|\mathbf{x}^T) \propto \sum_{\mathbf{t}, \mathbf{t}', \mathbf{g}} \prod_{i < j} \phi_{ij} \prod_i \phi_i \mathcal{G}_i \gamma_i \xi_i.$$

In particular, one can recover a factor graph topology equivalent to the one of the original graph of contacts (as in [?, ?]) by adding dynamical variables defined by triplets $(g_i^{(j)}, t_i^{(j)}, t'_j)$ and regrouping terms $\psi_i = \phi_i(t_i, \mathbf{t}'_{\partial i}) \prod_{j \in \partial i} \delta(t_i^{(j)}, t_i) \delta(g_i^{(j)}, g_i)$ (See Figs. 5.1c, 5.1d). We obtain the effective model

$$\mathcal{Q} \propto \prod_{i < j} \phi_{ij} \prod_i \psi_i \mathcal{G}_i \gamma_i \xi_i \quad (5.21)$$

in terms of which $\mathcal{P}(\mathbf{x}^0|\mathbf{x}^T) \propto \sum_{\mathbf{t}, \mathbf{t}', \mathbf{g}} \mathcal{Q}(\mathbf{g}, \mathbf{t}, \mathbf{t}', \mathbf{x}^0)$.

Belief propagation equations can be written for the efficient approximate computation of marginals of (5.21) (See SM). The BP update of ψ_i can be computed in time $O(G \cdot T \cdot |\partial i|)$, where G is the maximum allowed recovery delay, and the one of ϕ_{ij} in time $O(G^2 \cdot T)$. In practice, G can be

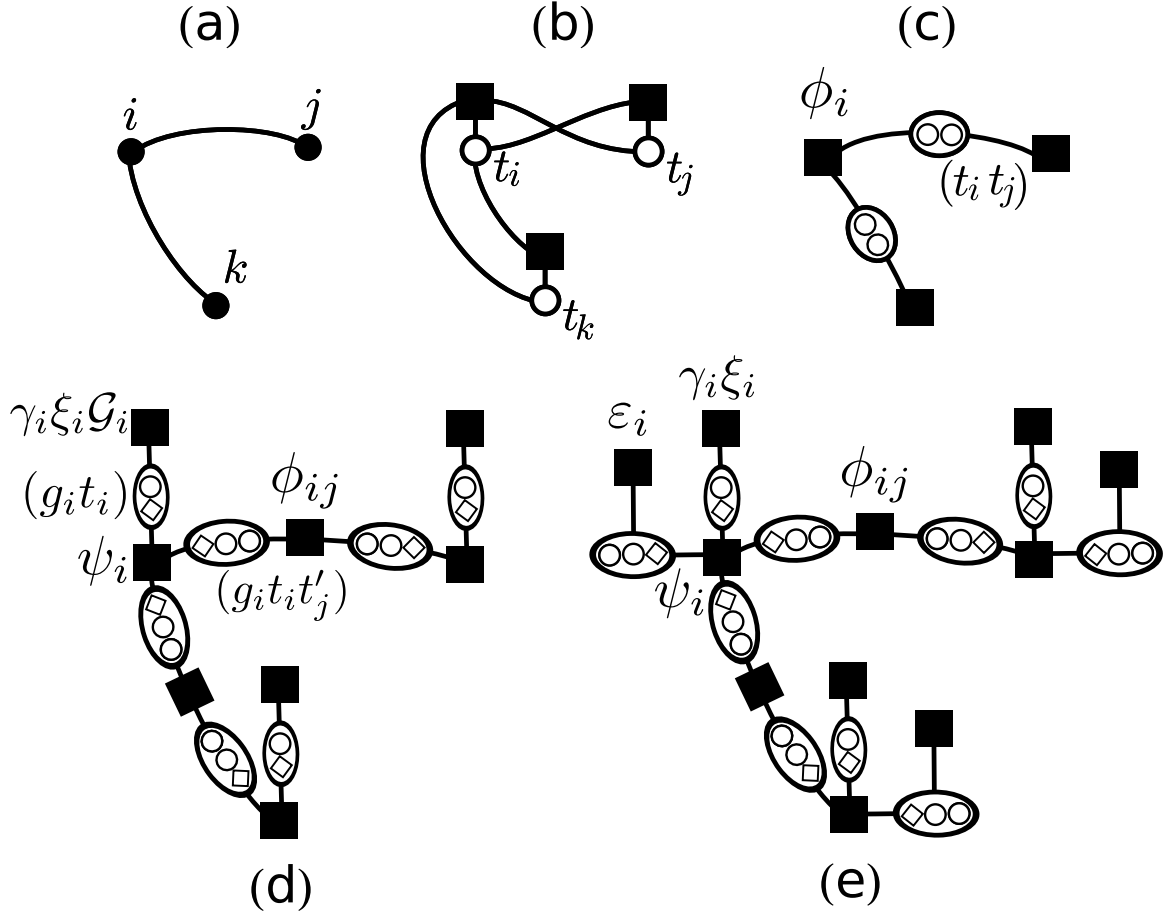


Figure 5.1: Factor graph representations for irreversible dynamics. (a) Original graph (b) Loopy, naive factor graph for an irreversible deterministic dynamics (c) Dual tree factor graph (d) Factor graph for the SIR model given in (5.21) with known epidemic age (e) Factor graph representation with unknown epidemic age.

taken constant for a geometric distribution \mathcal{G} . A single BP iteration can be thus computed in time $O(T \cdot G^2 \cdot |E|)$. We remark that the BP equations for the posterior distribution are exact (and have a unique solution) on tree factor graphs [?]. As the topology of the factor graph mirrors the one of the original graph, (5.21) allows the exact computation of posterior marginals for the SIR model on tree graphs (at difference with the DMP method).

Unknown epidemic age The start time of the epidemic was assumed up to now to be known, but this may be unrealistic. Coping with the case of unknown initial time is easy; starting with an upper bound on the epidemic duration, it suffices to consider the dynamical process to start from the all-susceptible state but to allow nodes to be spontaneously infected at an arbitrary time with (small) uniform probability. This is equivalent to the addition of a fictitious virtual neighbor to every node with no constraint in its activation time besides the prior probability $\varepsilon_i(g_i'', t_i'', t_i) = \delta(t_i'', \infty)(1 - 2\gamma) + \gamma$ of spontaneous infection (See Fig. 5.1e).

Identification of a single source Through out the rest of the paper we restrict ourselves to random regular graphs of degree $k = 4$ and $N = 1000$ nodes with homogeneous propagation probability $p_{ij} \equiv p$ and recovery probability $r_i \equiv r$ in order to make comparison with other inference methods simpler. An example of inference for an epidemic with recovery and transmission probabilities $\lambda = \mu = 0.5$ is depicted in Fig. 5.2. For each node, an infection time distribution is computed as a BP marginal and the mean and standard deviation of this distribution are plotted against the true time. We compared the inference performance of BP with the one of DMP in the case of a single epidemic source. The DMP method iterates the forward probabilistic transmission equations of [?] to obtain single-site trajectory probability distributions. Afterwards, using a single-site factorization hypothesis for joint distribution at the observation time, it gives a simple factorized expression of the likelihood of the observed epidemic. DMP is shown [?] to outperform other more classical estimations of the epidemic origin [?, ?, ?].

Two methods are used as benchmarks: Jordan centrality [?, ?, ?] and DMP. In figure 5.3 we show the average normalized rank given to the true zero patient by BP, DMP and Jordan centrality over 10^3 instances with observation time $T = 10$ and recovery probability $\mu = 0.5$. Since DMP requires a defined value of the observation time, we here assume that $T = 10$ is known for both DMP and BP. It is seen how BP outperforms DMP by a large margin in the probability of finding the epidemic origin. The inset shows also a remarkable difference in the normalized ranking of the true origin given by all three algorithms, again BP outperforming the other two.

Incomplete information In a more realistic setup, the state of many nodes in the population is usually not known. Calling ξ the fraction of unobserved sites (denoted in state X), we test the average ranking given to the real zero patient by BP, and compared it with other three methods in Fig. 5.4. We found that DMP can be improved over its original presentation if the dynamics MP equations are not iterated over all the nodes of the graph, but only over the connected component of I, R and X nodes. We refer to this version as DMP-restricted or DMP_r (see SM for details). Jordan centrality is also computed over this connected component as in [?]. BP is quite efficient in this regime, finding the true zero patient in more than 75% of instances up to 60% of unobserved sites (inner plot), and outperforming all other methods for almost all ξ .

Inference of multiple sources If the origin of the epidemics is not a single source, but a set of sources, methods based on the exhaustive exploration of initial states like DMP suffer a combinatorial explosion. This problem does not affect BP, as the trace over initial conditions is performed directly

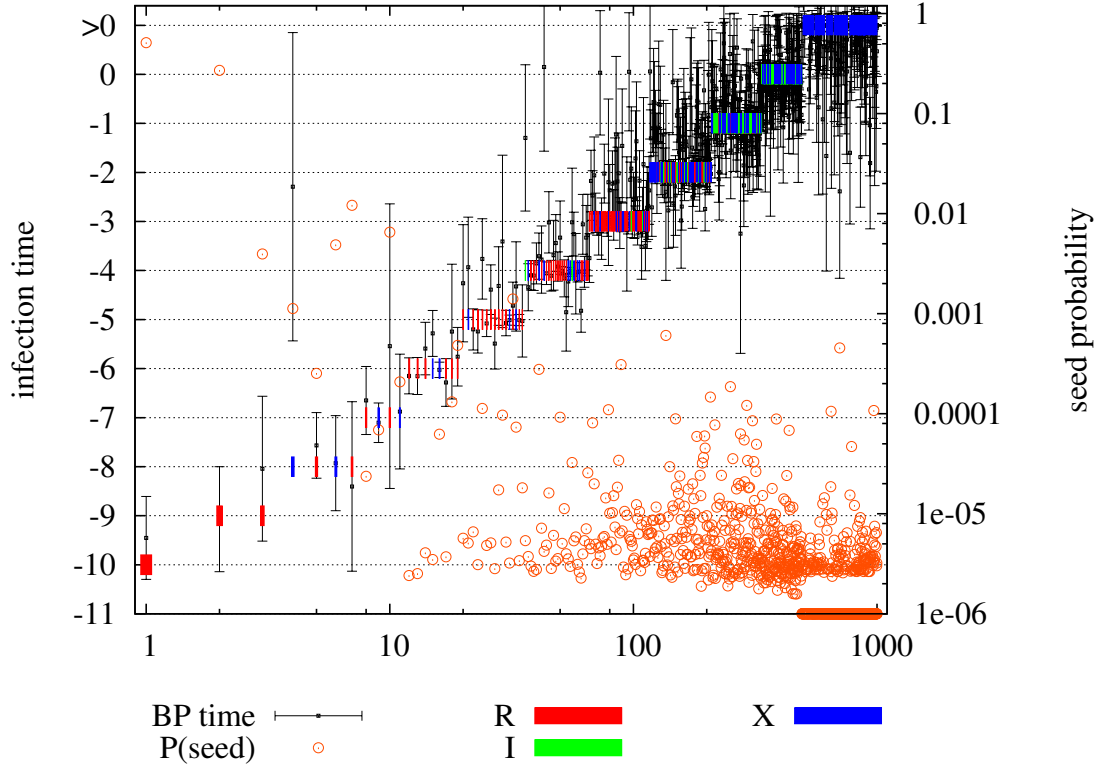


Figure 5.2: (Color online) An example of inference on a RRG with $N = 1000$, $k = 4$, with $\lambda = \mu = 0.5$ and prior seed probability 10^{-6} , and observations at time $T = 0$ on $\xi = 0.7$ of the nodes on an epidemic with $t_0 = -10$. Each point in the x axis corresponds to a vertex on the graph. Nodes are ordered by their real infection time (blocks, R=recovered, I=infected and X=unknown). The mean and standard deviation of their BP posterior marginal distribution of infection time is plotted (black dots and error bars) along with the marginal posterior probability of self-infection (orange, circles, right axis).

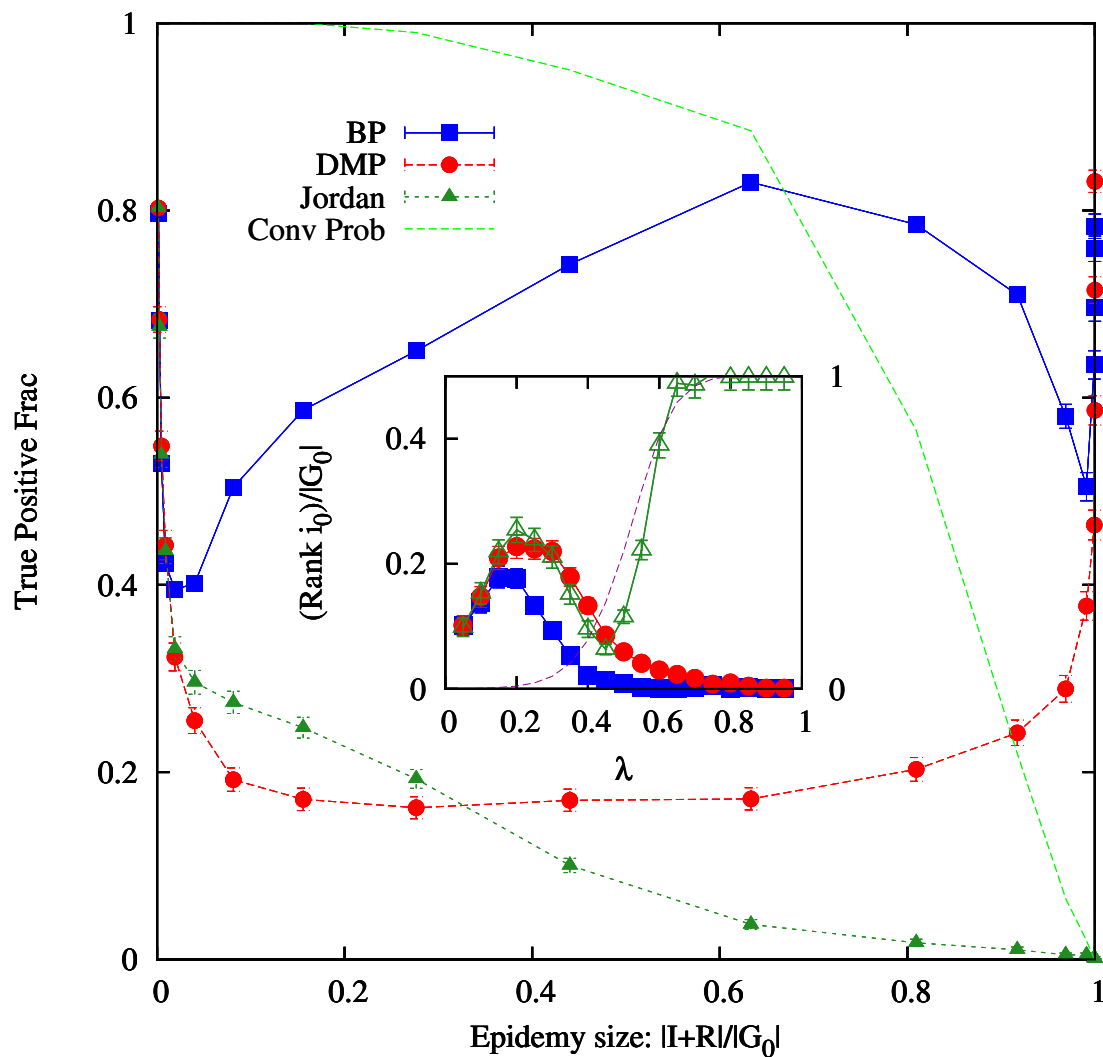


Figure 5.3: Normalized rank given to the true zero patient by Jordan, DMP and BP methods, as a function of the transmission probability λ , averaged over 10^3 instances of $N = 1000, k = 4$ RRG with observation time $T = 10$. Most relevant part is $0.2 < \lambda < 0.8$ where the epidemic size (purple, right axis) spans most of its range. For too large epidemics BP has convergence problems (green, right axis), nevertheless some relevant information is still present in the (unconverged) marginals. The inner plot shows the probability of finding the true origin of the epidemic using DMP and BP as a function of the average epidemic size for each value of λ .

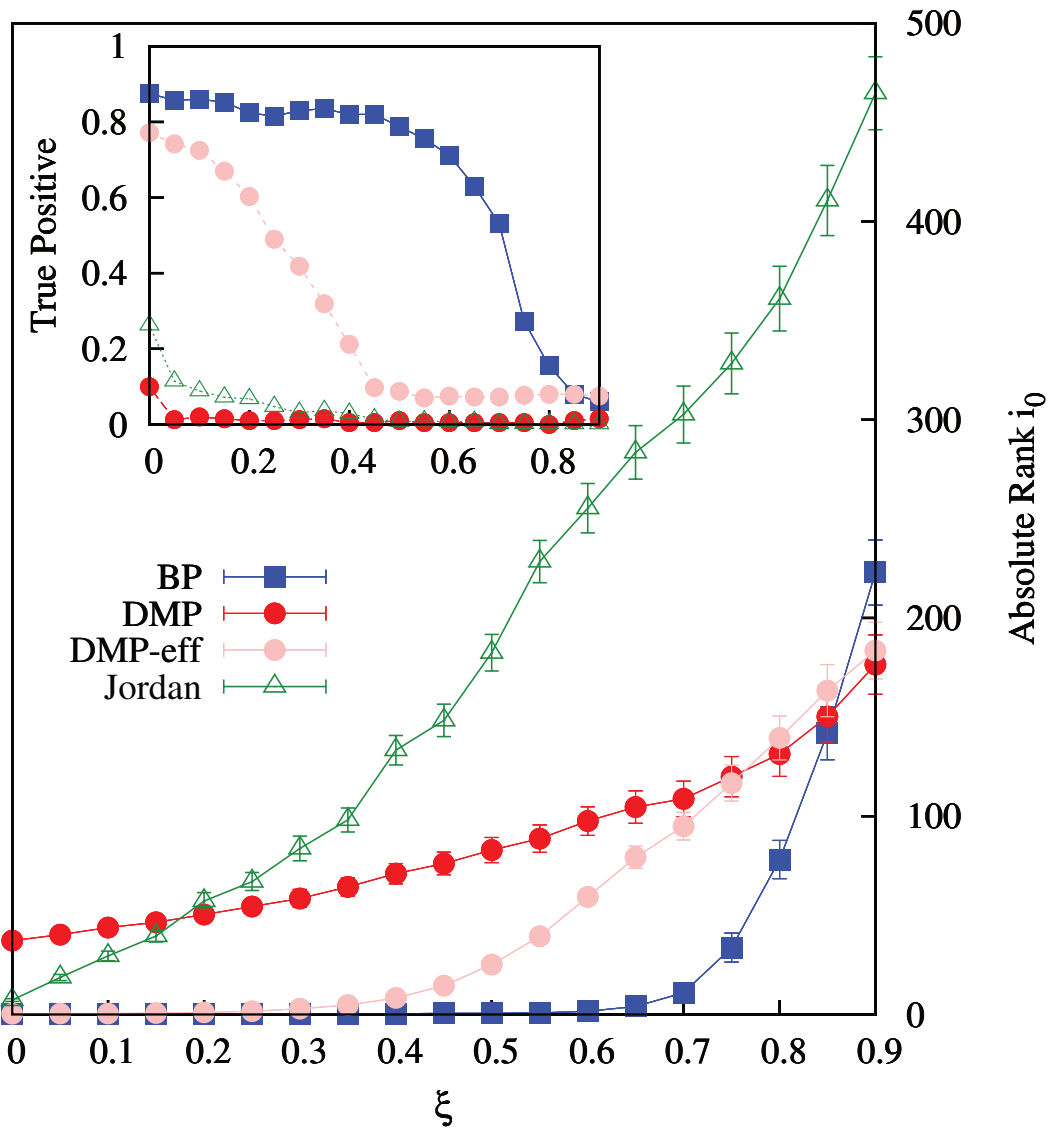


Figure 5.4:

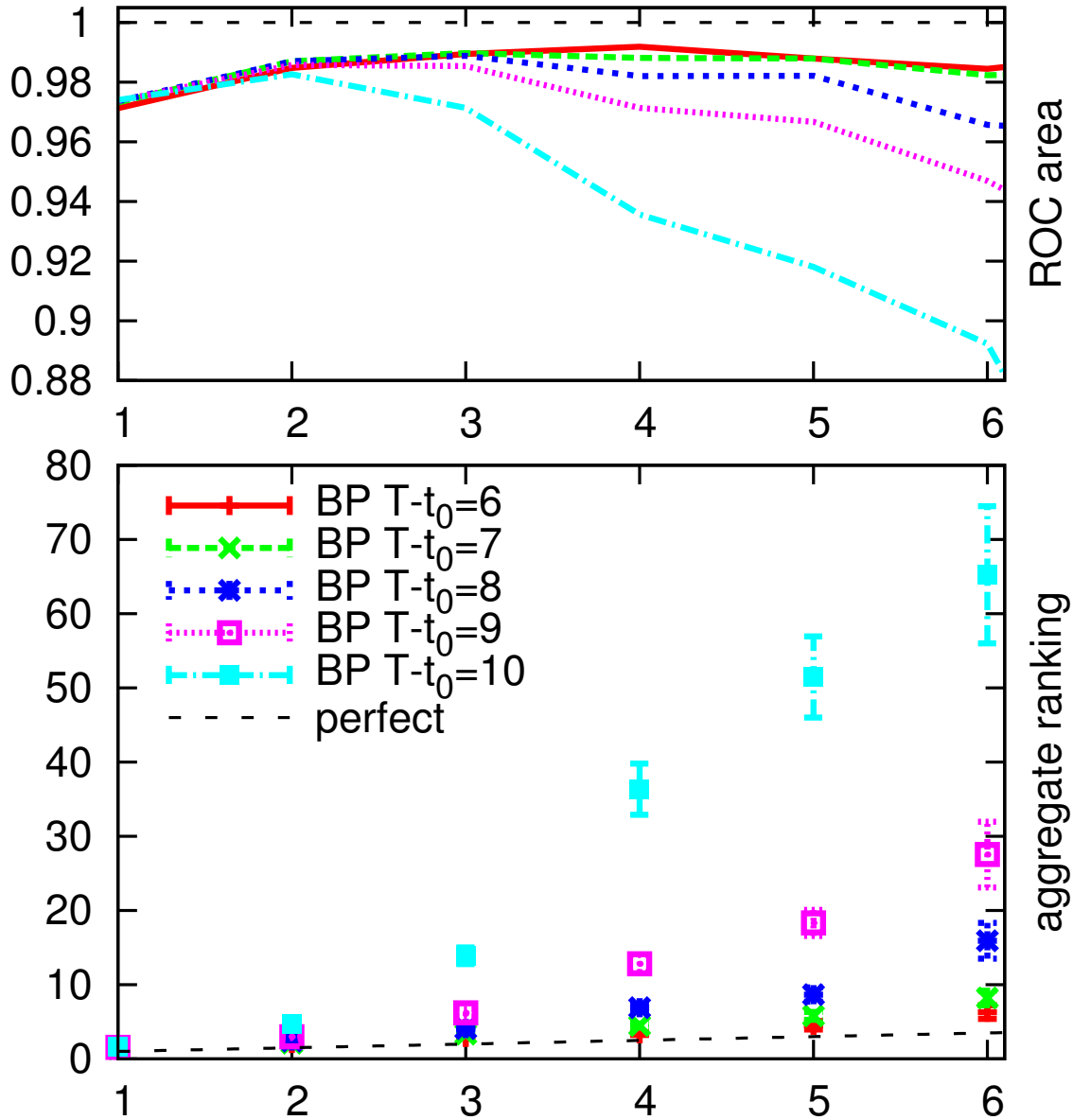


Figure 5.5: Inference of multiple seeds with $r = 1, p_{ij} = 0.5$, seed probability 0.002 and variable $T - t_0$ (unknown to the inference algorithm). Inference was attempted for each of 1 000 instances for every $T - t_0$, and the the sum of the rankings of the true seeds in the outcome order is plotted vs. the true number of seeds. Inset: the average area of the ROC curve. Each ROC curve was computed restricted to the epidemic subgraph.

within the framework. We show in Fig. 5.5 experiments with multiple seeds, showing that effective inference can also be achieved in this regime.

Dynamically evolving networks We studied the case of dynamically evolving networks, in which the transmission probability p_{ij} depends on time (representing the time-evolution of interactions between agents). For this purpose, we employed the interesting dataset [?], corresponding to a large list of time-stamped face-to-face contacts between couples of visitors in a cultural event. We aggregate the 20 second resolution of the data into T effective time steps, and simulate the progression of a virtual epidemic initiated by a random visitor. We employed the following set of parameters: probability $p_{ij}^{20s} = 0.1$ of contagion in a 20 second interval, recovery probability $r = 0.1$, and $T - t_0 = 30$. We employed the last four days of the event, in which the volume of visitors was larger, and simulated 2000 random propagations. We kept the 50 epidemics with larger number of infected individuals (98 ± 14). The correct origin was identified perfectly in 13 cases (26%), and in general its normalized ranking was in average 0.07 ± 0.11 . Finally, also infection times were recovered with good accuracy: the probability assigned to the correct infection time (for infected individuals) was in average 0.53 ± 0.06 , and the average absolute error in the inferred infection time was 0.76 ± 0.23 .

5.A Functionals and Lagrange Multipliers

Functionals

A functional is a function that takes a function as its input and produces a scalar value as its output. More formally, let X be a set of functions, and let F be a function that maps \mathcal{C} to the set of real numbers \mathbb{R} . Then, we say that F is a functional, and we write:

$$F : \mathcal{C} \rightarrow \mathbb{R}$$

In this context, the elements of \mathcal{C} are often called the "arguments" or "inputs" of the functional F . Sure, I'd be happy to explain how derivatives with respect to functional arguments are done.

When we take the derivative of a functional with respect to one of its arguments, we're asking how much the value of the functional changes as we change the function at that argument.

Formally, let $F : \mathcal{C} \rightarrow \mathbb{R}$ be a functional that maps a set of functions \mathcal{C} to the real numbers, and let $f \in \mathcal{C}$ be a function in the domain of F . Then, the derivative of F with respect to f is still a functional denoted by $D F[f]$ or $\frac{\delta F}{\delta f}[f, h]$, and is defined as follows:

$$D F[f, h] = \lim_{\epsilon \rightarrow 0} \frac{F[f + \epsilon h] - F[f]}{\epsilon}$$

where h is a small perturbation to f . In other words, we're looking at how much the value of F changes as we perturb f in the direction of $h(x)$, and then divide by the size of the perturbation ϵ .

Note that the derivative of a functional with respect to one of its arguments is itself a functional. The perturbation $h(x)$ is taken as a unit norm function. If we take $h(x) = \delta(x - x_0)$, the derivative is a function of x_0 , and a functional of $f(x)$. In such case, the derivation is carried as if $f(x)$ was itself a number. Lets see an example:

One of the simplest functionals, is the integral of a function, since it produces a real value:

$$I[f] = \int_a^b f(x) dx$$

where $f(x)$ is a function defined on the interval $[a, b]$.

The derivative of $I[f]$ with respect to f is given by:

$$\frac{\delta I[f]}{\delta f(x')} = \lim_{\epsilon \rightarrow 0} \frac{I[f + \epsilon \delta(x - x')] - I[f]}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_a^b \delta(x - x') \epsilon dx = \delta(x - x')$$

where $\delta(x - x')$ is the Dirac delta function. This result tells us that the derivative of $I[f]$ with respect to f at the point x' is equal to the Dirac delta function evaluated at x' .

Consider the functional $L[f]$ defined by:

$$L[f] = \int_a^b \sqrt{1 + (f'(x))^2} dx$$

where f is a function defined on the interval $[a, b]$. This functional takes a function f as its argument, and returns the length of the curve defined by the graph of f over the interval $[a, b]$.

To find the derivative of $L[f]$ with respect to f , we can use the calculus of variations. Specifically, we seek a function h that minimizes the value of the functional:

$$J[h] = L[f + \epsilon h]$$

subject to the constraint that $h(a) = h(b) = 0$. Intuitively, we're looking for the function h that produces the smallest change in the length of the curve as we perturb f by a small amount h .

Derivatives with respect to functional arguments are an important tool in many areas of mathematics, including calculus of variations, optimization, and partial differential equations.

Lagrange Multipliers

The method of Lagrange multipliers is a technique used in optimization to find the maximum or minimum of a function subject to one or more constraints. Suppose we want to find the maximum or minimum of a function $f(x_1, x_2, \dots, x_n)$ subject to a set of constraints of the form:

$$g_i(x_1, x_2, \dots, x_n) = 0, \quad i = 1, 2, \dots, m, \quad (5.22)$$

where m is the number of constraints. To solve this problem using the method of Lagrange multipliers, we introduce a set of Lagrange multipliers $\lambda_1, \lambda_2, \dots, \lambda_m$, and construct the Lagrangian function:

$$L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) - \sum_{i=1}^m \lambda_i g_i(x_1, x_2, \dots, x_n). \quad (5.23)$$

We then find the critical points of the Lagrangian function by setting its partial derivatives with respect to each variable to zero:

$$\frac{\partial L}{\partial x_i} = 0, \quad i = 1, 2, \dots, n, \quad (5.24)$$

$$\frac{\partial L}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, m. \quad (5.25)$$

The solutions to these equations give us the values of x_1, x_2, \dots, x_n and $\lambda_1, \lambda_2, \dots, \lambda_m$ that satisfy both the constraints and the maximum or minimum conditions.

The Lagrange multipliers have the interpretation of being the rate of change of the function f with respect to the constraints g_i . That is, if we move along the constraint surface in the direction of the i th constraint, then the rate of change of the function f is given by the Lagrange multiplier λ_i .

Chapter 6

Cascade epidemics and contact tracing

6.1 Effects of contact tracing and quarantine