Progetto dei Corsi di

Complex Systems and Network Science

Corso di Laurea Magistrale in Informatica

Analyzing Complexity

Corso di Laurea Magistrale in Relazioni Internazionali

A.A. 2020-2021

Dipartimento di Informatica — Scienza e Ingegneria
Università di Bologna

Ozalp Babaoglu Angelo Trotta

November 17, 2020

1 General Information about the Project

As part of your course requirement, you are to complete the project described below, which must be carried out individually. Submission of the project for evaluation must be done via email to the address: `angelo.trotta5@unibo.it`.

The deadline for submission is **23:59:59 hours on 04 January, 2021**. The email must have the subject field as *CSNS Project 2021* and must be sent from your University address (`name.surname@studio.unibo.it`).
You will receive a confirmation message within a few days of your submission. The email should contain an archive (in .zip or .tar.gz format) containing the following:

1. The source code that was developed (either in NetLogo or PeerSim);
2. A short paper, in PDF format, describing the model that was implemented, the experiments that were carried out using it, and a discussion explaining the results that were obtained.

Your full name, email address and student ID number (matricola) must be included in all of the source files, in the paper, and in the submission email that you send. The source code should be well documented and formatted, following good programming practices. The paper can be written in Italian or in English, and should be structured like a technical paper, thus containing a title, abstract and bibliography. It is strongly suggested that you limit the length to 16 pages and that you follow the Springer format for *Lecture Notes in Computer Science* (LNCS). Templates are available for both Word[1] and LaTeX[2]. You can use any text processing system you prefer (even though LaTeX is suggested) to write the paper as long as you submit the result as a PDF file.

The project must be done *individually*: no sharing of papers or source code is permitted. You are, of course, encouraged to discuss issues and solutions with fellow students or with the instructors.

2 Grading Policy

For your project to be satisfactory, it must satisfy the following requirements:

- The project must implement the specifications that follow. You are allowed (and encouraged) to apply modifications and extensions to the project, but they must be proposed to the instructors beforehand and approved by them.
- The model's implementation, and all of the related simulations and experiments, must be carried out using either the NetLogo or PeerSim software systems. If PeerSim is chosen, the cycle-driven simulation engine should

[1] Link to .doc template
[2] Link to .tex template

be used, and the simulator must be configurable by means of the standard PeerSim configuration file.

- Your paper must thoroughly describe the model that was implemented and justify all significant design decisions and extensions that were applied to it. You should also discuss the expected behavior of the model, by making previsions. Most importantly, you have to explain the experiments that you performed in terms of methodology and the results that you obtained. Significant implementation details can be inserted, if important in the context of the model, but should otherwise be kept as comments in the code itself.

You are encouraged to focus on a simple model and to apply extensions to it only if you completely understand the behavior of the base model. This can be achieved by working in modular fashion, thus incrementally (and carefully) adding new features, enriching your model. Ending up with a complex, unpredictable and difficult to understand model is very easy. On the contrary, you should prove through your experiments that you fully understand the behavior of your model and that you can interpret the results you obtained, and are able to relate them with real-world phenomena. Finally, you should try to find tipping points or interesting equilibrium states in your model.

If you are interested in these topics (e.g. you want to build better models or study other systems of this kind), do not hesitate to contact us when looking for a thesis topic.

3 Description of the Project

The purpose of this year's project is to study a contagion process through an *epidemic model* [1].
Communicable diseases that are endemic (always present in a population) cause many deaths. For example, in 2011 tuberculosis caused an estimated 1,400,000 deaths and HIV/AIDS caused an estimated 1,200,000 deaths worldwide. According to the World Health Organization, malaria still results in more than a million deaths. Measles, which cam be easily treated in the developed world, caused 160,000 deaths in 2011, but in 1980 the number of measles deaths was 2,600,000. The striking reduction in measles deaths is due to the availability of a measles vaccine. Other diseases such as typhus, cholera, schistosomiasis, and sleeping sickness are endemic in many parts of the world. The effects of high disease mortality on mean life span and on the economy in afflicted countries are considerable. Most of these disease deaths are in less developed countries, especially in Africa, where endemic diseases are a huge barrier to development.

For diseases that are endemic in some region, public health officials would like to be able to estimate the number of infected individuals at a given time as well as the rate at which new infections arise. The effects of quarantine or vaccine in reducing the number of victims are of importance. In addition, the possibility of defeating the endemic nature of the disease and thus controlling

or even eradicating the disease in a population is worthy of study. An epidemic, which acts on a short temporal scale, may be described as a sudden outbreak of a disease that infects a substantial portion of the population in a region before disappearing. Epidemics invariably leave part of the population untouched. Often, epidemic outbreaks recur with intervals of several years between outbreaks, possibly decreasing in severity as populations develop some immunity.

The 'Spanish" flu epidemic of 1919–1920 caused more than 50,000,000 deaths worldwide. The AIDS epidemic, the SARS epidemic of 2002–2003, recurring influenza pandemics such as the H1N1 influenza pandemic of 2009–2010, and outbreaks of diseases such as the Ebola virus are events of concern and interest to many people.

The essential difference between an endemic disease and an epidemic is that in an endemic disease there is a flow of new susceptibles into the population being studied through births, immigration, recovery from infection without immunity against reinfection, or loss of immunity in recovered individuals. This may result in a level of infection that remains in the population. In an epidemic, there is no flow of new susceptibles into the population and the number of infected individuals decreases to zero due to the scarcity of susceptibles.

There are many questions of interest to public health officials when confronted with a possible epidemic: How severe will the epidemic be? How many individuals will be affected and require treatment? What will the maximum number of people needing care at any particular time be? How long will the epidemic last? Can the epidemic be averted by vaccinating enough members of the population in advance of the epidemic? Will quarantining of victims reduce the severity of the epidemic?

3.1 The Compartmental Model

We formulate our descriptions of disease transmission as compartmental models, with the population under study being divided into compartments and with assumptions about the nature and the rate of transfers from one compartment to another. We assume a given population of size N divided into three classes labeled S, I, and R. Let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t. Let $I(t)$ denote the number of infected individuals, assumed to be infective and able to spread the disease by contact with susceptibles. Let $R(t)$ denote the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection. Removal can be carried out either through isolation from the rest of the population, through immunization against the infection, through recovery from the disease with full immunity against reinfection, or through death caused by the disease. These characterizations of removed members are different from an epidemiological perspective but are often equivalent from a modeling point of view which takes into account only the state of an individual with respect to the disease.

In many diseases, infectives return to the susceptible class on recovery because the disease confers no immunity against reinfection. Such models are

appropriate for most diseases transmitted through bacterial or helminth agents, and most sexually-transmitted diseases (including gonorrhea, but not diseases such as AIDS from which there is no recovery). We use **SIS** to denote the class of diseases with no immunity against reinfection, to indicate that the passage of individuals is from the susceptible class to the infective class and then back to the susceptible class.

We use **SIR** to denote the class of diseases that confer immunity against reinfection, where the passage of individuals is from the susceptible class S to the infective class I to the removed class R. Usually, diseases caused by viruses belong to the **SIR** class.

In addition to the basic distinction between diseases for which recovery confers immunity against reinfection and diseases for which recovered members are susceptible to reinfection, the **SIRS** class admits the possibility of temporary immunity. More complicated compartmental structures such as the **SEIR** and **SEIS** classes are also possible where an *exposed* (E) period is added between being infected and becoming infective.

In order to formulate a compartmental disease transmission model, we will need to make assumptions about the rates of flow between the compartments. Simplest models make assumptions about the rate of being infected and the rate of recovery from infection. One such assumption is *mass action incidence* where an individual causes new transmissions to another individual randomly chosen from the population with an *infection rate* a. Hence, in a population of size N, an individual on average makes $\mathsf{a}N$ contacts with other individuals per unit time. We assume that contacts are *effective*, in the sense that if there is a contact between an infective individual and a susceptible individual, infection is always passed to the susceptible. It is also possible that a contact leads to infection not with certainty but only with a certain probability. With mass action incidence (contact rate $\mathsf{a}N$), since the probability that a random contact by an infective with a susceptible is S/N, the number of new infections per unit time per infective is $(\mathsf{a}N)(S/N)$, resulting in $(\mathsf{a}N)(S/N)I = \mathsf{a}SI$ new infections. Alternately, given that the probability of contact with an infective is I/N, the rate of new infections per susceptible becomes $(\mathsf{a}N)(I/N)$, giving a rate of new infections $(\mathsf{a}N)(I/N)S = \mathsf{a}SI$.
We assume that individulas leave the infective class at a constant *recovery rate* $\mathsf{b}I$ per unit time, resulting in a mean infective period of $1/\mathsf{b}$.

3.2 The SIS Model

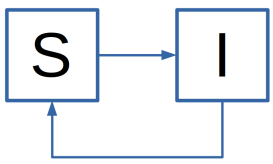

Figure 1: SIS model.

The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933.
The simplest SIS model, due to Kermack and McKendrick, is:

$$\begin{aligned} \frac{dS}{dt} &= -\mathsf{a}SI + \mathsf{b}I \\ \frac{dI}{dt} &= \mathsf{a}SI - \mathsf{b}I \end{aligned} \tag{1}$$

This model is based on the following assumptions:

1. The rate of new infections is given by mass action incidence.
2. Infectives leave the infective class at rate $\mathsf{b}I$ per unit time and return to the susceptible class.
3. There is no entry into or departure from the population.
4. There are no disease deaths, and the total population size is a constant N.

In an SIS model, the total population size N is equal to $S + I$. The hypothesis 3 really says that the time scale of the disease is much faster than the time scale of births and deaths so that demographic effects on the population may be ignored.

Because eq.1 implies $\frac{d}{dt}(S + I) = 0$, the total population $N = S + I$ is a constant. We can analyse the evolution of the system by looking at the variation of the infected set I. Substituting $S = N - I$ in eq.1 we have:

$$\begin{aligned} \frac{dI}{dt} &= \mathsf{a}(N - I)I - \mathsf{b}I = (\mathsf{a}N - \mathsf{b})I - \mathsf{a}I^2 \\ &= (\mathsf{a}N - \mathsf{b})I \left(1 - \frac{I}{N - \frac{\mathsf{b}}{\mathsf{a}}}\right) \end{aligned} \tag{2}$$

Here, we obtain a logistic equation in continuous time where we can sobstitute $(\mathsf{a}N - \mathsf{b}) = r$ and $N - \frac{\mathsf{b}}{\mathsf{a}} = K$:

$$\frac{dI}{dt} = rI \left(1 - \frac{I}{K}\right) \tag{3}$$

These two parameters can be interpretated as:

- r is the exponential growth rate
- K is the carrying capacity

The solution of the logistic differential equation is given by:

$$I(t) = K \frac{I(0)}{I(0) + (K - I(0))e^{-rt}} \tag{4}$$

Eq.4 shows different behaviors depending on the values of $I(0)$ and r (we assume $K > 0$ because otherwise there is no biological meaning in itself). If $I(0) = 0$ we are in a state called *disease-free equilibrium*, which corresponds to having $S = N$. Otherwise, analyzing the r value we have that:

- $r = (\mathsf{a}N - \mathsf{b}) < 0$ (or $\mathsf{a}N/\mathsf{b} < 1$), then all the solutions approach the limit zero as $t \to \infty$.
- $r = (\mathsf{a}N - \mathsf{b}) > 0$ (or $\mathsf{a}N/\mathsf{b} > 1$), then all the solutions approach the limit $K = N - \mathsf{b}/\mathsf{a}$ as $t \to \infty$. Here, the solution $I = N - \mathsf{b}/\mathsf{a}$ is called an *endemic equilibrium.*

The value of 1 for the quantity $\mathsf{a}N/\mathsf{b}$ is a tipping point in the sense that the behavior of the solution changes if this quantity passes through 1 because of some change in the parameters of the model.

Moreover, since $\mathsf{a}N$ is the number of contacts made by an average infective per unit time and $1/\mathsf{b}$ is the mean infective period, the quantity $\mathsf{a}N/\mathsf{b}$ is the number of secondary infections caused by introducing a single infective into a wholly susceptible population. The basic reproduction number is defined as the number of secondary infections caused by an average infective introduced into a wholly susceptible population over the course of the disease. The basic reproduction number is usually denoted by $\mathcal{R}_0$. Here, the basic reproduction number or contact number for the disease is:

$$\mathcal{R}_0 = \frac{\mathsf{a}N}{\mathsf{b}} \tag{5}$$

In studying an infectious disease model, the basic reproduction number is a central concept and its determination is invariably an essential first step. The value one for the basic reproduction number defines a threshold at which the course of the infection changes between disappearance and persistence. It is intuitively clear that if $\mathcal{R}_0 < 1$ the infection should die out, while if $\mathcal{R}_0 > 1$ the infection should establish itself. In more highly structured models than the simple one we have developed here, the calculation of the basic reproduction number may be much more complicated, but the essential concept obtains–that of the basic reproduction number as the number of secondary infections caused by an average infective over the course of the disease.

3.3 The SIR Model

One of the early triumphs of mathematical epidemiology was the formulation of a simple model by Kermack and McKendrick in 1927 whose predictions are very similar to the behavior, observed in countless epidemics, of disease that invade a population suddenly, grow in intensity, and then disappear leaving part of the population untouched. The epidemic models is:

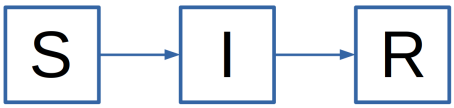

Figure 2: SIR model.

$$\begin{aligned} \frac{dS}{dt} &= -\mathsf{a}SI \\ \frac{dI}{dt} &= \mathsf{a}SI - \mathsf{b}I \\ \frac{dR}{dt} &= \mathsf{b}I \end{aligned} \tag{6}$$

It is based on the same assumptions that were made for the SIS model, except that recovered infectives go to a removed class rather than returning to the susceptible class. Moreover: $S(0) = S_0$, $I(0) = I_0$, $R(0) = 0$, $S_0 + I_0 = N$. We see that $N = S + I + R$ is constant.

We think of introducing a small number of infectious individuals into a population of susceptibles and ask whether there will be an epidemic. We remark that the model makes sense only so long as $S(t)$ and $I(t)$ remain non-negative. Thus if either $S(t)$ or $I(t)$ reaches zero we consider the system to have terminated. We observe that $\frac{dS}{dt} < 0$ for all t and $\frac{dI}{dt} > 0$ if and only if $\mathsf{a}S/\mathsf{b} > 1$. Thus I increases so long as $\mathsf{a}S/\mathsf{b} > 1$ but since S decreases for all t, I ultimately decreases and approaches zero.
The quantity $\mathcal{R}_0 = \mathsf{a}S_0/\mathsf{b}$ determines whether there is an epidemic. If $\mathcal{R}_0 < 1$, the infection dies out because $\frac{dI}{dt}|_{t=t'} < 0$ for all t', and there is no epidemic. Ordinarily, $S_0 \approx N$. If the epidemic is started by a member of the population being studied, for example by returning from travel with an infection acquired away from home, we would have $I_0 > 0$, $S_0 + I_0 = N$. A second way would be for an epidemic to be started by a visitor from outside the population. In this case, we would have $S_0 = N$. If $\mathcal{R}_0 > 1$, I increases initially and this is interpreted as saying that there is an epidemic.

3.3.1 Exposed Periods

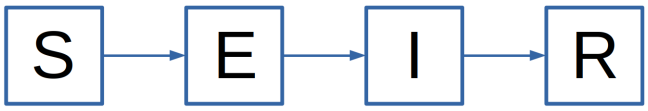

Figure 3: SEIR model.

In many infectious diseases there is an exposed period after the transmission of infection from susceptibles to potentially infective members but before these potential infectives develop symptoms and can transmit infection. To incorporate an exposed compartment with mean exposed period $1/\mathsf{k}$ we add

an exposed class E and use compartments S, E, I, R and total population size $N = S + E + I + R$ to give a generalization of the epidemic model:

$$\begin{aligned} \frac{dS}{dt} &= -\mathsf{a}SI \\ \frac{dE}{dt} &= \mathsf{a}SI - \mathsf{k}E \\ \frac{dI}{dt} &= \mathsf{k}E - \mathsf{b}I \\ \frac{dR}{dt} &= \mathsf{b}I \end{aligned} \tag{7}$$

The analysis of this model is similar to the previous analysis, but with I replaced by $E + I$. That is, instead of using the number of infectives as one of the variables we use the total number of infected members, whether or not they are capable of transmitting infection.

3.3.2 A Treatment Model

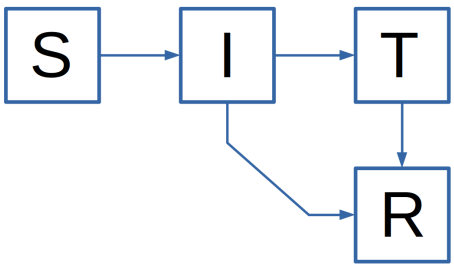

Figure 4: SITR model.

One form of treatment that is possible for some diseases is vaccination to protect against infection before the beginning of an epidemic. For example, this approach is commonly used for protection against annual influenza outbreaks. A simple way to model this would be to reduce the total population size by the fraction of the population protected against infection. In reality, such inoculations are only partly effective, decreasing the rate of infection and also decreasing infectivity if a vaccinated person does become infected. This may be modeled by dividing the population into two groups with different model parameters which would require some assumptions about the mixing between the two groups. In another case, if there is a treatment for infection once a person has been infected, this may be modeled by supposing that there is a rate c proportional to the number of infectives at which infectives are selected for treatment, and that treatment reduces infectivity by a fraction g. Suppose that the rate of removal from the treated class is n. This leads to the SITR model, where T is the treatment class, given by:

$$\begin{aligned}
\frac{dS}{dt} &= -\mathsf{a}S(I + \mathsf{g}T) \\
\frac{dI}{dt} &= \mathsf{a}S(I + \mathsf{g}T) - (\mathsf{b} + \mathsf{c})I \\
\frac{dT}{dt} &= \mathsf{c}I - \mathsf{n}T \\
\frac{dR}{dt} &= \mathsf{n}T + \mathsf{b}I
\end{aligned} \tag{8}$$

It is reasonable to assume that treatment does not slow recovery, so that $\mathsf{n} \geq \mathsf{b}$. Also, since the total time in the infectious and treatment stages should not be greater than the mean infective period, we should expect that:

$$\frac{1}{\mathsf{b}} \geq \frac{1}{\mathsf{c}} + \frac{1}{\mathsf{n}} \tag{9}$$

3.3.3 An Influenza Model

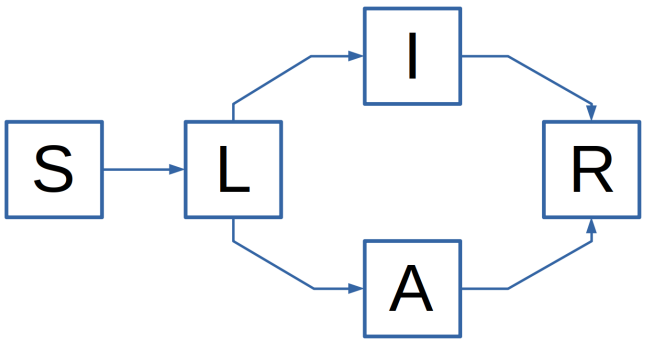

Figure 5: Influenza model.

In some diseases, such as influenza, at the end of a stage individuals may proceed to one of two stages. There is a latent period after which a fraction p of latent individuals L proceeds to an infective stage I, while the remaining fraction $(1 - p)$ proceeds to an asymptomatic stage A, with infectivity reduced by a factor g and a different period $1/\mathsf{n}$:

$$\begin{aligned}
\frac{dS}{dt} &= -\mathsf{a}S(I + \mathsf{g}A) \\
L' &= \mathsf{a}S(I + \mathsf{g}A) - \mathsf{k}L \\
\frac{dI}{dt} &= p\mathsf{k}L - \mathsf{b}I \\
A' &= (1 - p)\mathsf{k}L - \mathsf{n}A \\
\frac{dR}{dt} &= \mathsf{n}A + \mathsf{b}I
\end{aligned} \tag{10}$$

3.3.4 A Quarantine Isolation Model

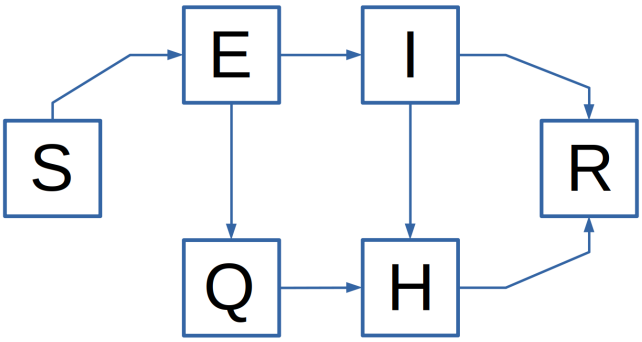

Figure 6: SEIQJR model.

For an outbreak of a new disease, where no vaccine is available, isolation of diagnosed infectives and quarantine of people who are suspected of having been infected (usually by tracing of contacts of diagnosed infectives) are the only control measures available. This model describes the course of an epidemic, when control measures are begun under the assumptions:

- Exposed members may be infective with infectivity reduced by a factor ε_E, $0 \leq \varepsilon_E < 1$.
- Exposed members who are not isolated become infective at rate k_E.
- We introduce a class Q of quarantined members and a class J of isolated (hospitalized) members and exposed members are quarantined at a proportional rate c_Q in unit time (in practice, a quarantine will also be applied to many susceptibles, but we ignore this in the model). Quarantine is not perfect, but reduces the contact rate by a factor ε_Q. The effect of this assumption is that some susceptibles make fewer contacts than the model assumes.
- Infectives are diagnosed at a proportional rate c_J per unit time and isolated. Isolation is imperfect, and there may be transmission of disease by isolated members, with an infectivity factor of ε_J.
- Quarantined members are monitored and when they develop symptoms at rate k_Q they are isolated immediately.
- Infectives leave the infective class at rate b_I and isolated members leave the isolated class at rate b_J.

These assumptions lead to the SEQIJR model:

$$
\begin{aligned}
\frac{dS}{dt} &= -\mathsf{a}S\left(\varepsilon_E E + \varepsilon_E\varepsilon_Q Q + \varepsilon_J J\right)\\
\frac{dE}{dt} &= \mathsf{a}S\left(\varepsilon_E E + \varepsilon_E\varepsilon_Q Q + \varepsilon_J J\right) - (\mathsf{k}_E + \mathsf{c}_Q)E\\
\frac{dQ}{dt} &= \mathsf{c}_Q E - \mathsf{k}_Q Q\\
\frac{dI}{dt} &= \mathsf{k}_E E - (\mathsf{b}_I + \mathsf{c}_J)I\\
\frac{dJ}{dt} &= \mathsf{k}_Q Q + \mathsf{c}_J I - \mathsf{b}_J J\\
\frac{dR}{dt} &= \mathsf{b}_I I + \mathsf{b}_J J
\end{aligned}
\tag{11}
$$

In this model the parameters c_Q and c_J are control parameters which may be chosen in the attempt to manage the epidemic. The parameters ε_Q and ε_J depend on the strictness of the quarantine and isolation processes and are thus also control measures in a sense. The other parameters of the model are specific to the disease being studied. While they are not variable, their measurements are subject to experimental error.

4 Networks and Epidemic Models

The historical study of networks has its grounding in two disparate fields: social sciences and graph theory [2]. Whereas in epidemiology, we speak of *hosts* and *contacts*, the social literature is based upon *actors* and *relations*, while graph theory uses the terms *nodes* and *edges*. In each case, however, it is the presence of a relationship between individuals in a population that is the issue of concern. The range of available vocabularies can hinder the transfer of ideas between these fields. We shall refer to *individuals* and their *contacts*; the set of contacts of an individual is their *neighbourhood* and the size of this neighborhood is the individual's *degree*.

Research in the social sciences is often concerned with the reason behind the network connections rather than the properties of the network structure itself. However, it provides a wealth of quantitative and qualitative information about social network connections, which are related to the mixing networks for airborne diseases. Network analysis has been used as an explanatory tool to describe the evolution and spread of ideas and innovations in societies, and observed social dynamics can often be understood through analysis of the social networks that underlie them. Attention has been given to the nature of connections, particularly properties such as symmetry (whether a relationship between A and B implies a relationship between B and A) and transitivity (whether the friend of a friend is a friend), which together provide measures of social cohesion. In addition, measures of the importance of individuals have also been derived; these range from the simple (such as the number of connections) to the highly

complex (number of paths between other actors in which an individual features. Because the social importance of an individual (i.e. the extent to which they dominate the network) is probably closely linked to their role in disease spread, such ideas are immediately relevant to epidemiology.

4.1 Network Analysis Using Graph Theory

Research in graph theory has provided a wealth of quantitative tools and mechanisms for describing networks, many of which have epidemiological applications. We can use an adjacency *matrix* or *sociomatrix*, $\mathcal{A}$, to describe the connections within a population, where $\mathcal{A}_{ij} = 1$ if there is a connection such that infection could pass from individual i to individual j; otherwise, $\mathcal{A}_{ij} = 0$. The matrix $\mathcal{A}$ summarizes all connections within the network. For example, a network (or graph) is said to be connected if any individual (or node) can be reached from any other by following network links; epidemiologically, this is equivalent to infection being able to reach the entire population from any starting point, which is the case if $\sum_{m=1}^{\infty} \mathcal{A}^m$ has no zero terms (here, $\mathcal{A}^m$ contains information about paths within the network of length m). Zeros in this matrices demonstrates that the network is divided into two or more separated components, none of which has any links to any of the others.

Determining a complete network requires knowledge of every individual in a population and every relationship between individuals. For all but the smallest populations, this is an impractically time-consuming task. Mainly, problems in building the graph concern data collection, but more fundamental is the question of how a network link is defined. If networks are to be used for epidemiological purposes, then connections should only be included if they describe relationships capable of permitting the transfer of infection. However, in many cases, it is not clear how to define such a relationship; how much contact is it necessary to have with someone with influenza, say, before there is a measurable risk? The issue is likely to be most acute for infections spread by casual contact, where some degree of arbitrariness is inevitable. Nevertheless the difficulties in build a complete graph, those networks that have been explored provide important insights into interaction patterns and their implication for disease transmission.

Three main techniques can be employed to gather network information: *infection tracing*, *complete contact tracing*, and *diary-based* studies (Figure 7). These have their own advantages and disadvantages, and the method applied depends on the resources available and the purpose for which the data is being gathered.

4.1.1 Infection Tracing

After an epidemic, field-based epidemiologists place considerable emphasis on determining the source of infection for each case. In this way, each infected individual is linked to one other from whom they caught the infection, and additionally, to a variable number of others to whom they transmitted the disease, thus providing a *transmission network* consisting of all the links through

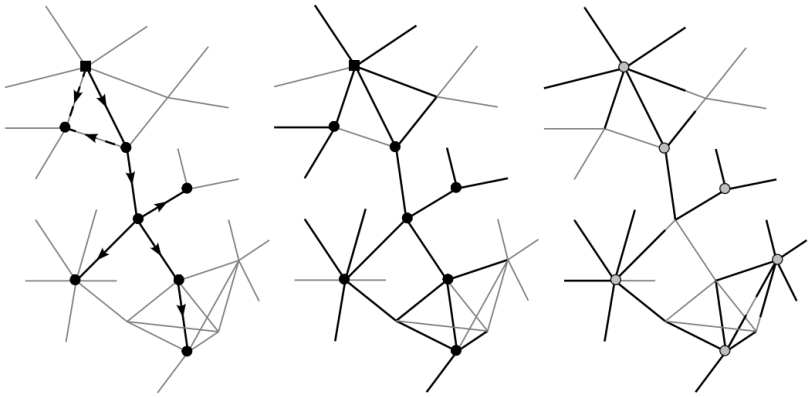

Figure 7: For the same simple network (thin grey lines), the type of network information that is achieved using infection tracing (left), contact tracing (middle) and diary-based studies (right). For infection and contact tracing, circles represent infected individuals, while the square shows the primary infectious case; for the diary-based study, those taking part are shown with open circles. For infection tracing, only sources of infection are traced and some individuals (e.g. top left) have multiple potential sources of infection. For contact tracing, a proportion of all contacts from infectious individuals are traced. Finally, with a diary-based study, although almost all links can be traced, the lack of a unique identifier means that often links from different individuals cannot be connected. [2]

which infection spread in a single outbreak. Because all connections represent actual transmission events, this method does not suffer from problems with the definition of links; however, interactions that did not happen to lead to the transmission of infection in this particular case will be omitted from the network. The networks observed are therefore liable to be tree-like, containing no loops, and this precludes any worthwhile evaluation of the more complex network measures. However, since such tracing is often an integral component of disease control policies, these partial samples of the network can be generated a little extra cost, and provide useful information about the individuals most involved in disease transmission.

4.1.2 Contact tracing

Contact tracing aims to identify all potential transmission contacts from a source individual (known as an *index case*). This reveals a new set of individuals who might be infected and who can be the subject of further tracing effort. Because it aims to identify potential transmission routes, contact tracing suffers from network definition issues; in addition, it is time consuming and relies on individuals providing complete and accurate data about personal relationships. Contact tracing has been commonly applied not as a network evaluation device but as a control tool. In such case, the purpose of contact tracing is to identify asymptomatic infected individuals who can then be treated or quarantined. This means that the contacts of uninfected individuals are not sought, and thus

only a subset of the full mixing network will be uncovered; we might expect a partial network with many dead-ends consisting of uninfected individuals. Although a network uncovered via contact tracing is not complete and has biases, it is generally most detailed in regions of the network with the highest disease burden; thus the network data obtained is of immediate epidemiological relevance.

4.1.3 Diary-based studies

The determination of networks through tracing is highly labour intensive and relies on the subject individuals being able to recall and willing to recount their contacts. In contrast, in diary-based studies subjects record contacts as (or shortly after) they occur, shifting the work-load from the researcher to the subject and allowing a larger number of individuals to be sampled in detail. The change of focus from the population approach of other tracing methods to the individual-level scale of diary-based studies has some associated problems. Firstly, the data collection is at the discretion of the subjects, thus the definition of a close contact may not be the same for all individuals. Secondly, while this method gathers detailed individual-level data, it may be difficult to link this information into a comprehensive network, as the names or identifiers of contacts may not be accurately or uniquely recorded. Indeed, unless the subject individuals come from a coherent group, such as work colleagues or residents in a single small community, it is probable that the study will result in a large number of unconnected sub-networks, each one representing the personal network of a few individuals.

4.2 Simulated Networks

While collecting network data is beset with difficulties, the simulation of disease transmission on networks is relatively easier. Here the infection transmission is based on the individual's neighborhood, i.e. the infection rate become $\mathsf{a}\hat{I}$, where $\hat{I} \subseteq I$ are the infectious in the individual's neighborhood. All network-based simulations are limited by fact that there is no simple way to ascertain the sensitivity of the epidemiological results to the details of the network structure. Such simulations are therefore always vulnerable to questions of 'what if?'; for example, we may ask whether the network is representative of an average realistic community, whether variation between communities will bias the results if large population sizes are considered and whether rare but epidemiologically important contact structures are missing from the network. It is difficult to answer such questions or gain an intuitive understanding of network structure if our experience is limited to simulations of sampled networks.

5 The Project

Purpose of the project is to study the evolution of an epidemic contagion using one of the compartmental models described in Section 3 (or any extension of

them). Moreover, you are to study possible actions to limit the epidemic spread and its consequences. Based on the specific case study, these actions can be (but not limited to):

- *Social distance* to avoid contacts among infected and non-infected agents
- *Quarantine* to isolate the infected agents
- *Focused restrictions* (train stations, events, school, etc.) to avoid agent clusters
- *Selected lockdown* to avoid agent mobility among different areas (eg. regions)
- *Total lockdown* to prevent completely agent mobility

However, in some other cases, we want the contagion to spread as fast as possible. In these cases, we need to study techniques to improve the "virus" diffusion. The containment measures should be supported by the network analysis for constant monitoring of the epidemic evolution. The study must include one of the techniques introduced in Section 4 (or a variant of them) to trace the diffusion of the contagion. In this case, the simulation needs to keep a network structure of the infection/contacts and analyse the feasability of this approach (how many resources I do need to build this network in real life? What if this network is not perfect? etc.).

In your report you are supposed to give the reader the necessary background knowledge on the problem/phenomenon you have chosen, as well as motivations for that choice (e.g. why this is an interesting problem, which applications it has etc). The simulation/modeling of the phenomenon has to be justified as clearly as possible. For instance, you should give the full details of your encoding/simulation of the phenomenon as well as argue why this encoding/simulation is a good one. During the study of the system, analyze its evolution (studyng all the used parameters, like the thresholds and the learning factors, and their impact on the system) and the behaviour during "stressing" events.

5.1 Some Ideas for Project Proposal

Here we list some possible topics for your final project. These examples can be also used as a cue for different types of problems.

5.1.1 Health

Consider an epidemic of a virus in a human population. Some example can be:

- *SARS-CoV-2* pandemic [3]
- *Influenza* [4] [5]
- *H1N1* [6]

5.1.2 Computer Science

Analysis of a computer network. Some example can be:

- *Peer-to-peer* (P2P) networks to understand the spread of information on computer networks [7]
- *Spread of computer virus* [8]
- *Delay Tolerant Networks* (DTN) where the moving agents need to share the information among the non-connected network [9]

5.1.3 Others

- *Zombies attack* to analyze the hypotesis of a zombie attack [10]

If you have doubts please do not hesitate to contact us. Please keep in mind that this is not a research project. You are not supposed to come up with new evolution of the epidemic models. Your task is to study in detail a (possibly already existing) application.

References

[1] Brauer, Fred, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*. Springer New York, 2019. https://www.springer.com/gp/book/9781493998265

[2] Keeling, Matt J., and Ken TD Eames. *Networks and epidemic models*. Journal of the Royal Society Interface 2, no. 4 (2005): 295-307. https://doi.org/10.1098/rsif.2005.0051

[3] Giordano, G., Blanchini, F., Bruno, R. et al. *Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy*. Nat Med 26, 855–860 (2020). https://doi.org/10.1038/s41591-020-0883-7

[4] Mummert, A., Otunuga, O.M. *Parameter identification for a stochastic SEIRS epidemic model: case study influenza*. J. Math. Biol. 79, 705–729 (2019). https://doi.org/10.1007/s00285-019-01374-z

[5] Hooten, Mevin B., Jessica Anderson, and Lance A. Waller. *Assessing North American influenza dynamics with a statistical SIRS model*. Spatial and spatio-temporal epidemiology 1.2-3 (2010): 177-185. https://doi.org/10.1016/j.sste.2010.03.003

[6] Mutalik, Anirudh. *Models to predict H1N1 outbreaks: a literature review*. International Journal Of Community Medicine And Public Health [Online], 4.9 (2017): 3068-3075. Web. 6 Nov. 2020. http://dx.doi.org/10.18203/2394-6040.ijcmph20173814

[7] D. F. Bernardes, M. Latapy and F. Tarissan, *Relevance of SIR model for Real-World spreading phenomena: experiments on large-scale P2P system.* IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 327–334, 2012.

[8] Xie Han, Qiulin Tan, *Dynamical behavior of computer virus on Internet*, Applied Mathematics and Computation, Volume 217, Issue 6, 2010, Pages 2520-2526, https://doi.org/10.1016/j.amc.2010.07.064.

[9] Luca Sciullo, Angelo Trotta, Marco Di Felice, *Design and performance evaluation of a LoRa-based mobile emergency management system (LOCATE)*, Ad Hoc Networks, Volume 96, 2020, https://doi.org/10.1016/j.adhoc.2019.101993

[10] Munz, Philip, Ioan Hudea, Joe Imad, and Robert J. Smith. *When zombies attack!: mathematical modelling of an outbreak of zombie infection.* Infectious disease modelling research progress 4 (2009): 133-150.