Social network analysis in epidemiology: Current trends and perspectives

2 authors:

Erick Stattner
University of the French Antilles
69 PUBLICATIONS   **350** CITATIONS

SEE PROFILE

Nicolas Vidot
University of the French Antilles
27 PUBLICATIONS   **436** CITATIONS

SEE PROFILE

Social Network Analysis in Epidemiology: Current Trends and Perspectives

Erick Stattner, Nicolas Vidot
LAMIA Laboratory,
University of the French West Indies and Guiana,
FRANCE
estattne@univ-ag.fr, nvidot@martinique.univ-ag.fr

Abstract—The increasing development of world trade has promoted the spread of diseases. Recent events, like the H1N1 outbreak and spread, have confirmed this tendency. Everyone could observe that this kind of infectious disease often has the ability to cross countries borders and spread rapidly. Preventing, monitoring and controlling the outbreak of these diseases have appeared to be a public health problem of primary importance for years. However, although numerous works have been conducted in order to understand and study epidemics, this subject remains a topical issue because of the complexity of the phenomena. In this paper, we present new opportunities offered by the field of social networks for understanding the spread of infectious diseases. This work focuses both on techniques already used in epidemiology, and very recent works in social network analysis that open new perspectives. We give an overview of methods currently used to model and study the spread of diseases and future directions that the social network approach allows to glimpse.

Index Terms—Spread of disease, Social Network, Epidemics, Infectious disease, Link Mining

Fig. 1. Distribution of main causes of death worldwide according to the WHO (2007)

I. INTRODUCTION

According to the World Health Organization (WHO), diseases are becoming a current major international issue, especially infectious diseases that threaten health, economy and security. Numerous WHO reports support this observation and clearly indicate the importance of the phenomenon [109], [110]. Figure 1 represents the main causes of death worldwide. We can notice that infectious diseases are at the second rank, after cardiovascular diseases. However, we should be careful since it is important to highlight the fact that these figures are below reality. For instance, cardiovascular or digestive diseases should be counted since they are often caused indirectly by an infectious disease. The rate of people infected by an infectious disease is increasing dramatically in countries of Southeast Asia and Africa. Today, it is estimated at 45% in developing countries, making infectious diseases the leading cause of mortality among children and young adults. Thus, infectious diseases are responsible for more than 13 million deaths each year, with one in two in developing countries. It is important to know that currently only six diseases are responsible for 90% of deaths in the world due to infectious spread.

However, the spread of infectious diseases is not new since the world experienced large waves of pandemics (Black Death, Spanish flu, Asian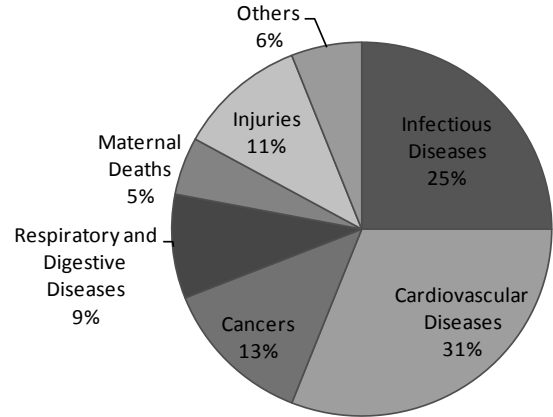 flu, etc.) More recently, an emerging disease of the 21st century, the Severe Acute Respiratory Syndrome (SARS), occurred in China in late 2002 and infected more than 8000 people worldwide in just a few months. Similarly, some diseases are still worrying, like the AIDS virus discovered in 1970 and which currently affects 33 million people or the H1N1 influenza epidemic in 2009.

WHO sounds the alarm today because the phenomenon is becoming of much concern for two reasons. First, the mutations, increasing resistance to antimicrobial medicines and failures of health systems continue to thwart the intervention plans implemented. The second reason concerns the high mobility of populations. A WHO recent report [111], explains that one major factor of how a disease is spreading is the high mobility of individuals. In the last decades, the considerable increase in mass movements of individuals, improved by the growth of the air traffic, has allowed infectious diseases to spread, both within countries and between continents in a few hours. Airlines now carry more than 2 billion passengers per year, and thus naturally participate in the spread of diseases.

In this context, we understand the importance of addressing this problem. Epidemiology is the science that is focused on infectious diseases. De and Das [33] defined epidemic theory as being "*the study of the dynamics of how contagious diseases spread in a population, resulting in an epidemic*". More formally, epidemiology is the study of patterns of health and

illness and associated factors at the population or individual level. It refers to all methods of modeling [113], analysis [96] or monitoring [25] the spread of infectious diseases in a given system, for identifying risk factors and determining optimal intervention approaches to clinical practice and for preventative medicine.

Although the biological interest for this phenomenon is undeniable, many other scientific communities have also contributed to its understanding: anthropology [104], mathematics [58], [59], computer science [40], [13]. In modeling, the mathematics was the first to be interested in this problem. Traditional methods were based on the biological and demographic aspects of disease [78] (state of disease, time of infection, duration of incubation, the social factor, etc.). However, such modeling is not suitable for diseases that spread by contacts between individuals.

In the context of anthropology J. A. Barnes [11] introduced for the first time in 1954 the concept of "social network" to represent social contacts between individuals. The craze for this kind of representation has been immediate, since it enables to take into account in the analysis both attributes and relationships between individuals. This type of representation has received strong interest from the scientific community, and was used in 1985 in epidemiology to study Sexual Transmitted Diseases (STDs). Indeed, STDs constitute a good example of diseases that depend on personal contacts for dissemination. There have been various studies analyzing the social network and demonstrating its relevance for STD research. The investigation of social networks is fundamental both to a better understanding of the process and the dynamics of transmission of an infectious that depends on social contact for spread and to determine the influence that the connections between people have in the transmission of a given disease.

In this context, new epidemiological techniques that aid public health officials in the prevention could be of great help in the attempt to control these diseases. In this study, we present an overview of works on the study of disease spread with a particular emphasis on the integration of social interaction into designed models. This work is focused both on traditional techniques used in epidemiology, and very recent works on social network analysis that may offer new perspectives.

This paper is organized in 8 sections as follows. In Section II we describe main issues in the modeling of disease spread among populations and distinguishes four types of models that are detailed in the following sections from the simplest to the most complex and according to an incremental point of view. Section III presents the *compartment models* that represent the simplest forms based on state changes only. Section IV is devoted to *metapopulations models* that in addition take into account the spatial structure of populations. In Section V, we present the *network models* that help to understand the impact of social interactions between people on the spread process. Section VI presents the very new approaches that are the most advanced and represent the complexity of the real world as a *multi-agent system modeling* the whole social network. In Section VII, we discuss perspectives of future works combining social network analysis techniques and epidemics modeling. We conclude in Section VIII.

II. Issues in Epidemiology

A major feature of the immune system is its ability to recognize and destroy infected cells [85]. An infection must be recognized at an early state so that the immune mechanisms are established efficiently. When an individual is infected, the pathogen responsible of the infection is trying to multiply in the body, while the immune system attempts to repulse it, involving complex biological processes occurring at different levels. In theory, to fully understand the spreading process of disease, all relevant biological factors should be considered, while in practice, such a complex study is not feasible on a human scale [17].

Fortunately, there exist other ways to approach the problem with simplified models of disease spread [7], [8], [90], [45]. Such approaches are not interested in the biology of the phenomenon, but rather try to understand the spread of disease over the network at different level communities and individuals. Modeling and automated simulations provide powerful tools that allow studying the spatio-temporal evolution of diseases [47], [46], [88], [50]. They can be used to understand the individual roles and the factors that influence the spread [26], [25] or measure the effects of various interventions such as pharmaceutical strategies or prevention campaign targeting people or the environment [84], [14]. We propose to classify these different according to four categories as follows: *Compartment Models* that are based on the assumption of a uniform mixing of individuals into different states and focus on changes among from one state to another, *Metapopulation Models* that introduce in addition the spatial structure of populations, *Network Models*, that exploit the idea that only the interactions between individuals or the environment allow the transmission of disease through social contacts and *Multi-Agent-Based Models*, that are the most advanced solution for modeling the complexity of social interactions in the real world. In the following sections, we detail these approaches and then we present potential future works.

III. Compartment Models

Epidemic modeling was initiated prior to the paradigm of social network. Indeed, first mathematical models for epidemiology seems to date back to Bernoulli, who published on complex analysis of smallpox vaccination, as described by Dietz and Heesterbeek [36]. Today, mathematical epidemiological models have been marked by a set of *coupled rate equations* generally referred as *compartment models*, since the early 1900s [92], [70], [15]. An overview of the main mathematical models can be found in [60], [77] or in [35] where Diekmann and Heesterbeek give a very complete description of these models.
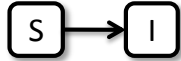
Compartment models assume that a population can be divided into a set of compartments (or states) according to the level of the disease development on individuals. In

such a model, the transmission occurs when individuals are in contact with infected individuals. These models assume that individuals have a regular relational structure between them and within compartments. Indeed, the mixing inside and between compartments is assumed to be random, or directed by transition rules that allow specifying how individuals switch to another compartment as proposed by Anderson et al. [5]. Thus, the number of new infections is proportional to the product of the number of infected individuals with the susceptible individuals. In other words, interactions between individuals leading to disease transmission are homogeneous, thus compartment models naturally are deterministic for a population in a given compartment. Subsequently, differential equations are commonly used to describe the evolution of the disease in the system, by mixing with certain probabilities individuals into the different compartments.

In addition, the structure of populations may be integrated within these models [51], [57], [54], [4] by adding new compartments corresponding to various individual features of the population such as the age, the risk behaviors, the social status, etc. The following subsections present the main compartment models used in epidemiology.

A. SI Model

We begin with perhaps the simplest model of contagion, that is called the *Susceptible-Infected Model*. In this kind of model, the disease within the host is only reduced to two states: *susceptible* (S) and *infected* (I). An individual in the susceptible state has not contracted the disease, but could catch it if he comes into contact with someone who is infected, whereas an individual in the infected state corresponds to someone who has contracted the disease.

This two-state model assumes that an infected individual retains its state forever. The approach here consists on modeling the spread of disease by a certain probability to pass in the infected state.

Let us denote S the number of susceptible individuals, X the number of infected individuals and $n = S + X$ the population size. Suppose β be the probability (per unit time) for a susceptible individual to contract the disease. Each infected individual has $\beta S/n$ chances to come into contact with a susceptible person, and hence new infected individuals can be estimated at $\beta SX/n$. Thus, the rate of change for X is expressed as a differential equation :

$$\frac{dX}{dt} = \beta \frac{SX}{n} \tag{1}$$

This equation may be solved by standard methods. Afterwards, as shown in Figure 2, it can be used to study the evolution of the number of infected individuals by varying the different parameters.
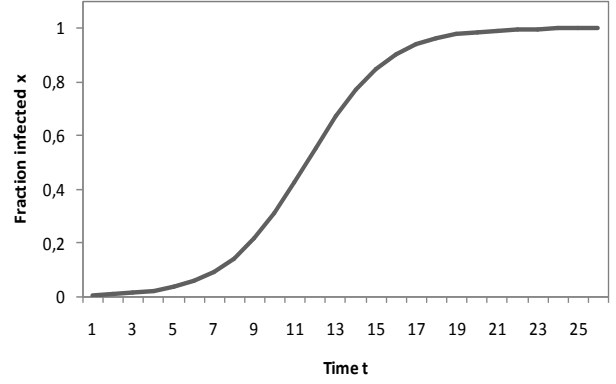

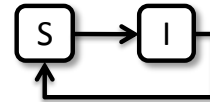Fig. 2.   Fraction of Infected Individual in the SI Model

B. SIR Model

On the same principle, the SIR model is an evolution of the SI model, that conserves the two states of the SI model and adds a new state: the *recovered* state. This model is called the *Suceptible-Infected-Recovered* model (SIR). It can be viewed as a more realistic approach in the case of diseases where people recover from infection after a certain time because their immune system rejects the pathogen responsible from the disease.

This kind of model assumes that people retain their immunity to the disease after such a recovery so that they cannot catch it again. This model can also be used to study other disease where people do not recover, but die after some interval. Indeed, in modeling term, both recovery and death can be represented by the R state of this model.

C. SIS Model

The SIS model is an another evolution of the SI model that allows *reinfection*. In the SIS model, individuals, once infected, can return to the S state, and can be infected again and again. To represent this behavior from the SI model, we must allow an infected individual (i.e. in the I state) to return in the S state.
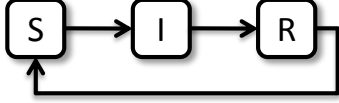

The SIS model is interesting for diseases that can infect people more than once. This may be due to an immune system that does not confer immunity to victims after recovery, or just confer a limited immunity. However, this kind of model seems limited, if one want to study the effect of various interventions on individuals, since it does not permit an individual to heal completely.

D. SIRS Model

The SIRS model takes advantage of SIR and SIS models by combining these two models. Indeed, it integrates the

recovery and the possibility of reinfection. This model takes into account the fact that immunity may be temporary. When individuals are in the R state, they gain immunity, like in the SIR model, but this immunity is temporary. After a certain time period, they lose their immunity and become susceptible again.

```
S → I → R
↑_____|
```

The SIRS model allows going beyond the limits set in the SIR and SIS models. It provides a model with oscillations more realistic since it includes the temporary immunity of the recovered individuals, a feature of many diseases.

A full description of these models can be found in [38], [83]. Other compartment models can be found in the literature [37], [53], [64]: SEIS, SEIR, MSIR, MSEIR and MSEIRS models. Similarly, other approaches try to integrate into existing models a *carrier* state [113], [74] for people who had an infectious disease that never completely recover and continue to carry it.

Compartment models are based on the assumption that the population structure may be represented only in terms of individual characteristics such as disease development status, age, behavior, etc. They rely on a strong and non realistic assumption that is the equal probability of individuals to come into contact with another. This approach proves to be interesting when we only want to follow the evolution of certain parameters inherent in the spread of disease as the number of susceptible, infected, recovered, etc. These approaches are very simple, since they do not reflect the real complexity of human interactions. Indeed, in the real world, people are actually connected to a small portion of individuals [102], and this portion is obviously not chosen randomly.

IV. METAPOPULATION MODELS

There is now a consensus on the idea that the role of population structure cannot be ignored in the understanding of the spread of human diseases. While compartment models assume a random distribution within populations and are based on individual attributes, it has become obvious that models have to integrate different types of social interactions. Among these models, so-called metamodels [98] represent a first stage that reduces the complexity of realistic network models. Since they introduce population structures, they can be considered as *compromise between compartment models and networks* [106]. Metamodels assume a random distribution among subpopulations like compartment models and introduce new types of interactions based on the spatial structure of populations. The metamodel approach is applied to understand the epidemic dynamics of populations structured in spatial subpopulations like families, villages, city locations, towns, cities, regions. The epidemic evolution is associated to the mobility of individuals among subpopulations [95], [98], [106], [103]. Metapopulation epidemic models are based on the spatial structure of the environment, and the detailed knowledge of

subpopulations interactions. Individuals within each subpopulation are classified into states like infected, susceptible, recovered and compartment approaches like SIR are applied to consider that individuals in the same location may get into contact and change their state according to the infection dynamics. The interaction among subpopulations is the result of the movement of individuals from one subpopulation to the other. A central issue is thus to describe the commuting patterns of people with precision. While reducing too much the complexity of real networks, metamodels help to get insight on the role of population structure on disease spread as shown by several contributions on this topic. For instance Watts et al. [106] assumes a hierarchical structure of subpopulations where interactions intra and inter subpopulations are distinguished. He shows that the final size of an epidemic is dependent on the structure of the population. With a similar idea, Vazquez [103] evaluates the ability of an epidemic outbreak to progress on the basis of the local outbreak size in subpopulations and social bridges rates between them. Colizza and Vespignani [27] show the importance of a local and global epidemic threshold that determine the epidemic outbreak at local and global scales. While metamodels have proved to be useful to understand the potential progression of epidemic outbreaks, all authors underlines the need to integrate more accurate data on population behaviours.

V. NETWORK MODELS

Although the concept of network has traditionally been used to represent a wide variety of objects and phenomena that have relations between them [107], the concept of social network was presented in 1954 for the first time by J. A. Barnes [11], in an article on anthropology, that emphasizes the social aspect of relationships between nodes in a network. In epidemiology, it is easy to understand that social networks have quickly found applications, since a significant factor of the outbreak and the evolution of diseases is the structure and the nature of the network of the interactions between which it spreads. Indeed, the transmission of disease often depends on the nature of the interactions that an individual maintains with his environment. Let us consider a simple example to highlight the importance of social networks. If we consider the context of STDs, the probability that an individual catches a disease, obviously depends on the characteristics of his social network, in other words, the nature of the social contacts with his entourage. Thus, network models are particularly useful in epidemiology, since the assumption of a uniform mixing between the compartments is proved to be clearly not realistic, for instance in the case of HIV/AIDS, influenza or Severe Acute Respiratory Syndrome (SARS) that only spread when social contacts are occurring. Figure 3 shows an example of sexual network. Each individual is represented by a circle and his sexual relations with other individuals are represented by links.

In 1985, Klovdahl [61] presented his study on STDs, that was the first application of social network analysis (SNA) in epidemiology. Recent studies [9], [79], [83] show that the
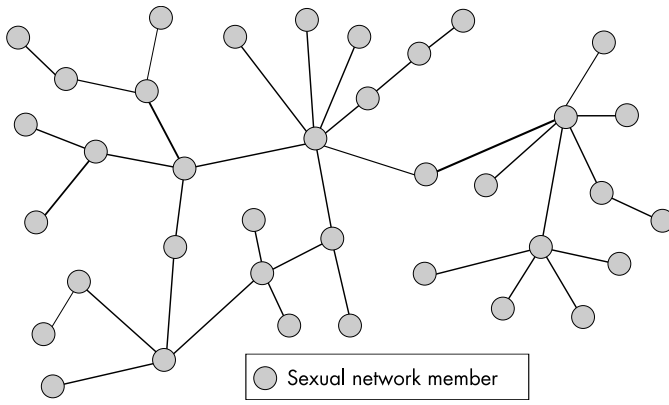
Fig. 3. Example of Sexual Network

analysis of such networks are of great interest to numerous scientific communities, because it finds applications in many fields. In epidemiology, social networks have received the attention of several communities: medicine [84], [86], biology [75], [69] or computer science [40], [42], [33]. The network approach has two main advantages:

1) Since, the world is made around social contacts between individuals, the network representation appears to be more suitable; it allows representing the dynamics of social contacts, that are known to greatly influence the way how a disease spreads [82], [91].

2) The network structure itself, plays an important role in finding efficient plans of intervention or awareness [25]. It provides powerful tools for understanding the underlying context of the "dynamic disease" and monitoring the evolution and the impact of various strategies.

So we can define a *social network model* for infectious diseases as a model that focuses on diseases spreading into the network of interactions between some entities. Usually, entities are individuals, but can also be groups or even geographic locations. Each entity has an associated vector of features that represent:

1) The individual state. It represents the level of the disease development into an individual as defined in existing compartment models (SI, SIR, etc.).

2) The individual characteristics. They correspond to all the features of a node that can influence the spread of the disease, represented as a attributes vector.

Unlike compartment models and metapopulation models, the transmission of the disease depends on the network structure and can also be combined with the distribution of the individual attributes (age, weight, ethnic, etc.). In this sense the appearance of social networks has played a major role in studying spread of diseases. The network is generally valued and oriented. Moreover, this kind of network is time dependent, since the interactions are evolving at each time unit.

There are two ways to obtain a social network. Either recorded statistical data are available and allow to generate and make evolve the network or the network is simulated with the

spread of disease. In both cases, the choice of interactions is fundamental. Chen et al. [24] explain that *"when SNA is applied to investigate disease transmission, the types of contact used to construct social networks should resemble the paths of the transmission"*. We can distinguish two types of contacts: personal contacts *(A)* or geographical contacts *(B)*.

A. Personal Contacts

Personal contacts are the first kind of contacts that have been exploited in epidemiology. Nevertheless, it is quite difficult to give a precise definition since they are used in so various and multidisciplinary ways. However, with the exception of applications in other areas [33], [42], we can state that, in epidemiology, personal contacts of an individual refer to kinds of human contacts with other people and are identifiable by the observation. They include for instance kinship, friendship, leisure activity, intimate contacts, etc that have been very efficiently considered in the study of STDs spread.

Numerous network models have used personal contacts to model the spread of diseases, because they are usually the most natural and identifiable paths, particularly in the field of sexually transmitted diseases, that seems to be more suited to this kind of model. For example, we can cite the large contribution of Klovdahl in the study of STDs [61], [63] who is considered as being a pioneer in the field. Inspired by Klovdahl, other works have been conducted on HIV/AIDS [93], [71]. Social networks have also been successfully applied to gonorrhea [49], syphilis [94], [91], chlamydia [56] or even tuberculosis [3] and Severe Acute Respiratory Syndrome [75]. Many applications of personal contact networks can be found in [86].

Numerous tools, are presented in the literature like an interesting simulation environment on HIV/AIDS, called *HIVSim* [2], proposed by Andrew et al. It simulates populations and allows evaluating the impact of vaccination. In 2005, Christley et al. [26], proposed a simple methodology to generate a small-word network. Due to frequent difficulties to obtain data sets, this work is interesting, since small-word networks prove to be contact networks close enough to reality [102]. In the same vein, Corley et al. presented *DynSNIC* [29] a computational simulator created to embody the intimate dynamic and evolving social networks related to the transmission of STDs and infections. We can also cite, an original work proposed by Barth-Jones et al. [14] that use Monte-Carlo simulation to analyze vaccine effects.

B. Geographical Contacts

Traditionally, social networks for disease investigations are constructed only with personal contacts. For sexually transmitted diseases (STDs), which require intimate contact for transmission, personal contacts are adequate to form a well-connected network for investigation. However, this kind of representation find its limits when we are focused in the case of disease like tuberculosis (TB) or SARS, that are transmitted not only through personal contacts. Indeed, a disease can spread by **(a)** direct contact between individuals, but can

also be transmitted by **(b)** the environment or **(c)** fomites. For cases (b) or (c), personal contacts do not deliver enough information, since they do not reveal the underlying context of the transmission. In their study on SARS, Chen et al. [24], show that the inclusion of geographical contacts in the social networks are a valuable contribution. They demonstrate indeed that incorporating geographical contacts into SNA *"provides a good way to find potential connections among patients and to see the role that those geographical locations play in disease outbreaks"*. Figure 4 presents an example of geographical contacts networks generating by spatial projection.
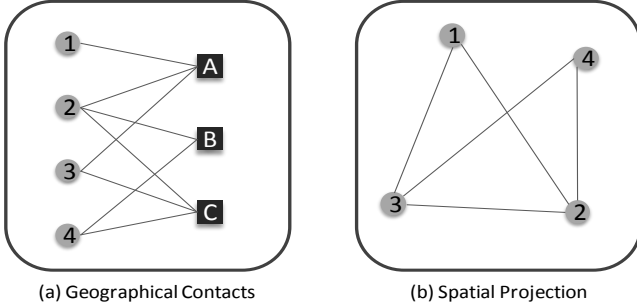

Fig. 4.   Example of Geographical Contacts

Numerous studies have recognized the importance of geographic location. For example, Klovdahl et al. [62] and McElroy et al. [73] included places as a type of node into the network. They both found that geographical contacts were critical for understanding the outbreak and the discovery of some potential contacts among patients. Another way is to connect two individuals when they have been in the same places [13]. This type of representation greatly increases the network connectivity, allowing to highlight nice properties of the network affecting the disease transmission [24]. Others applications of geographical contacts networks can be found in [32].

On the side of tools, we can cite *GERMS* [1], presented by Adams et al. that propose an epidemiological simulation tool for studying geographic and social effects on infection transmission. More recently, Stephen Eubank [41] proposed a tool for simulating the spread of disease among individuals in a large urban population over the course of several weeks. We can also cite the important contribution of Barrett et al. through *EpiSimdemic* [13], a scalable parallel algorithm to simulate the spread of contagion in large and realistic social contact networks. These simulations allow analyzing the effects of various interventions, such as vaccination or quarantine, targeted at individual, collective or geographic levels.

Thus, the type of contact used depends on the disease you study. Figure 5, presents examples of network construction [24]. We can see that the real interactions between individuals can have different semantic. Obviously, it is possible to generate a network composed of several types of interactions [63], [28].

Disease	Contact	
	Type	Semantic
Sexually Transmitted Disease (STD)	Personal contact	Sexual Contact
		Drug Use
		Needle Sharing
Tuberculosis (TB)	Personal contact	Social Contact
	Geographical contact	Frequenting the same place
Severe Acute Respiratory Syndrome (SARS)	Geographical contact	Frequenting the same place

Fig. 5.   Semantic of Networks

Both kind of networks on personal or geographical contacts are already subject of very active research, particularly in the search for relationships between network structure and the spread of the disease, and the study of the impact of various intervention plans on the transmission [25], [96], [97], [26], [49]. Finally, network models turn out to be a more realistic approach than simple models like compartment or metapopulation models, since they are more suited to the complexity of real relationships. However, mining such models is still difficult for several reasons. Firstly, the underlying social contact network is extremely wide, irregular and dynamic. Next, because real data sets are not available, network models often are based on generated data sets, that may not reflect fully the real world. Experimental studies on the impact of pharmaceutical, political, or awareness interventions require a very large number of executions, with results that can sometimes be skewed compared to reality. Third, to obtain complete and realistic models, models have to consider a large number of parameters. Indeed, the diversity among factors is crucial for understanding the spatial and temporal spread of a disease. However, focusing on all parameters is not always possible. Consequently, many models are restricted to a number of parameters considered as relevant a priori .

VI. Towards Multi-Agent-Based Models

Individual Based Models (IBM)[55] that were devised for the simulation of population in the context of ecology. The idea was to take profit of the increased available computing power to refine the modelling of population. Instead of modelling a population as a set of states and describing them by indices which hold aggregated values of state variables, each individual is represented as an actual entity. Thus a population is not reduced to a centroid of its component entities. It is a much complex representation that integrates the whole information on individuals, not just a resume. The population evolution model is then no longer just the result of differential equations but is the result of all complex interactions that take place between individuals. While rather new in epidemiology, these approaches follow diffrent directions among which we can observe a progression from cellular automata to multi-agent systems.

One of the most famous examples of such Individual Based Model making use of a *cellular automata* to simulate the behavior of a population is the game of life of Conway[48], which is based on original Von Neumann's works proposed a few decades earlier. The game is defined as an infinite

two-dimensionnal grid of square cells. Each cell can have two possible states "live" or "dead". At each step in time a transition occurs and is uniquely defined by the current state of the cell itself and the states of its neighbours. The global behavior of the population is a result of the local interactions between the entities for which only a simple behavior has to be specified [21]. Extensions to this basic principle have been be devised to modelize the propagation of an epidemic combined with a compartment model with more complex rules [89], [108].

As opposed to the original Conway game where cells represent some real world entities, numerous proposals use cells to represent areas in space [81], [89], [108]. Some cells act as interaction locations for individuals (transportation, public places, etc), as a prelude to the emergence of multi-agent systems [44]. *Multi-agent systems* (MAS) are the current most sophisticated individual based models. They represents individuals as agents that are independent entities driven by beliefs, desires and intentions [23]. Beliefs are what the agents know about the world (himself, others and the environment) and desires are roughly speaking what motivates the agents to act, intentions being those that the agent undertakes to achieve.

MAS are widely used for transportation systems simulation [12], [101]. The model is well-adapted to describe such systems in which numerous entities interact and where these interactions can have a great impact on the population behavior. In fact, these models are particularly suited to take into account the *spatial impacts* of the various risk events [31]. Based on the observation that the spread of an epidemic depends very closely on the individual contact network, Eubank proposed to devise a MAS simulating an epidemic propagation on a large urban population, from one simulating a transportation system [41] . The contact network was generated on the assumption that the transportation infrastructure determines people behavior and indirectly infers contacts. This way the MAS not only provides a finer granularity of the simulation, but also offers a more realistic approach thanks to a dynamic contact network. The flexibility of MASs enabled a variety of refinements to enhance the simulation [43], [39], [19]. Another interesting feature is the ability of such systems to scale as the model is enriched either by exploiting the environmental characteritics, the disease propagation scheme or other characteristics of the social network [13], [18]. All these aspects seems to designate multi-agent systems as a promising candidate for future developments.

VII. Perspectives

Since their introduction in 1954 [11], social network analysis has been a very active research field. Although first works have mainly concerned the social sciences [76], [66], particularly through the study of human behavior [22], [112], their natural ability to model a very large number of phenomena, has attracted other scientific communities, in biology [99], [68], mathematics [105], [10] or computer science [20] for instance. Indeed, it is now obvious that a very large number of real phenomena involve complex interactions between entities,

that is why networks appear to be a suitable representation. Moreover, social network allows gathering, in a single data description (the network), a large quantity of information that can be measured or analyzed according to techniques defined in several disciplines such as sociology, economics or mathematics. To sum up, the graph theory provides a formal framework for analysis relational data, through three levels of analysis:

 (i) Visualization analysis, to visually and identify some features of the network. It can be used to observe groups, subgroups or even the role of some individuals.
 (ii) Traditional analysis, to measure structural properties of networks. It allows studying a phenomenon by focusing on the network topology.
 (iii) Link mining, including contemporary analytical methods that apply or design data mining techniques to networks.

In epidemiology, network models have naturally allowed the application of these techniques to analyze, understand and control the spread of infectious diseases. However today, we can observe that only (i) and (ii) levels of the analysis seem to have really been exploited. If we focus on the epidemiology context:

At (i) level, we observe that SNA is applied to visualize the spread of diseases [94] by highlighting subgroups within groups and individuals that are high-risk such as those who act as bridges to transmit the disease from one subgroup to another. It can also be used to characterize a group or an individual [49] for understanding the underlying context of the transmission. However, such an approach requires to provide adequate graphical representation of networks, and this requirement constitutes a research problem in itself [30], and often limits the exploration to small networks.

At traditional level (ii), SNA is applied to measure structural properties of the constructed networks [26] to understand relationships between its structure and the transmission process [96], [25]. A frequent application is to use node properties, through centrality measures (degree, betweenness, shortest-path, etc.), to infer on the importance of individuals and identify high-risk nodes. An interesting study have been conducted by Christley et al. [26], who compare these measures to identify such individuals. Other works try to reveal potential bridges between subgroups [96]. Today, this kind of analysis is the most common because it allows identifying groups of people who are central in the disease transmission in order to design efficient strategies for breaking the chains of transmission [61], [97].

To the best of our knowledge, current trends in epidemiology do not exploit the possibilities recently offered by the SNA **at (iii) level**. Our goal here is not to detail these methods, but to present the perspectives they offer in epidemiology. In the following subsections, we present two fields of link mining that could be new tracks to study and control the spread of infectious diseases.

A. Graph Labeling

Faced to the increasing amount of social data, mainly due to the emergence of Web 2.0 and the improved performance of storage devices, a key issue in the field of social networks involves the automatic classification of nodes, also known as the "annotation in networks" or "graph labeling". Angelova and Weikum [6] present this problem as being "*a supervised learning technique for assigning thematic categories to data items.*". These last years, automatic labeling methods have been the stake of intense research [34], [52], [72], [16].

Two families of methods have been developed for learning labels in graphs. The first family of method exploit only the network structure [52], [16]. This kind of method propagate the labels iteratively through the network, without considering any node feature. The other family takes into account both the structure of the network and node features [72]. For example, the model proposed by Denoyer et Gallinari [34] is able to exploit simultaneously visual and textual content, metadata, and relational information (implicit like word or image similarities, or explicit, like friendship links) for image labeling in a social media.
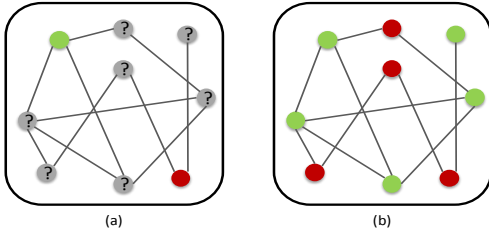

Fig. 6.   Graph Labeling in Epidemiology

In epidemiology, such methods could have a dual interest. First, numerous data sets are incomplete. Indeed, since many attributes are often missing or inaccurate, the learning methods are not enough accurate. Thus, the annotation in networks could be used to improve the quality of data sets, by completing or reassessing missing or erroneous values. On the other hand, we know that many individuals are sometimes carrying a disease but do not know it, either because the disease is not necessarily expressed, or because they have not consulted a specialist. These individuals, for which we have irrelevant data only, are particularly dangerous since they may transmit the disease without an appropriate treatment. Thus, by annotating nodes in a network, we could identify these individuals and allow early interventions to limit the spread as shown in Figure 6. In addition, a direct consequence of the annotation would result in improving the quality of data sets and the performances of learning algorithms.

B. Link Prediction

A considerable amount of attention has been devoted to the mechanisms through which social networks evolve. This question, referred as the *Link Prediction Problem*, is a fundamental issue that is still not well understood, but is also the subject of very active researchs [80], [65], [87], [100]. Liben-Nowell

et Kleinberg [67] present this problem as follows: "*Given a snapshot of a social network at time t, we seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time $t + \alpha$*".

We can distinguish two strategies to predict the formation of links. Indeed, there are methods that are based on the network structure only (a nice overview of these methods can be found in [67]), and other methods that include the nodes features too [87], [100]. Moreover, a fundamental aspect of some methods is that they work with a static snapshot of the network, rather than considering the network evolution as proposed by Liben-Nowell et Kleinberg.
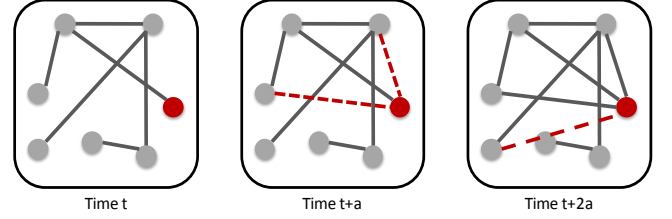

Fig. 7.   Link Prediction in Epidemiology

Such methods could be used to predict the occurrence of new links between susceptible individuals and infected individuals as shown in Figure 7, where the dotted lines represent predicted links. Thus, this solution could be applied to control the spread of infectious diseases, by attempting to contain the disease around an infected individual through interventions at two stages as depicted in Figure 8.
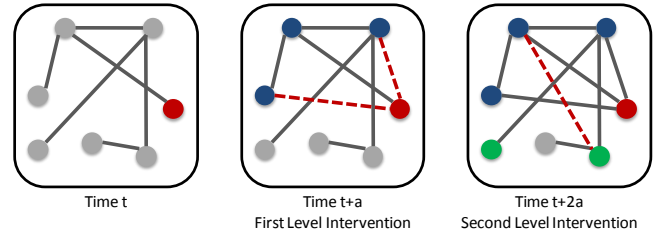

Fig. 8.   Example of interventions at two levels

1) At the first level: Interventions are targeted on individuals in contact with infected individuals and on individuals who have a high probability of being in contact with an infected individual.

2) At the second level: Interventions are targeted on individuals in contact with neighbors of an infected individual and on individuals who have a high probability of being in contact with the neighbors of an infected individual.

VIII.   CONCLUSION AND FUTURE DIRECTIONS

This article gives a quick overview of applications of social network analysis (SNA) in epidemiology and a sketch of possible future investigation. This work is motivated by the very recent advances in the field of SNAs. We have presented the variety of contributions that attempted to understand the spread processes of infectious diseases, with a particular focus

on those which integrate social interactions. We have traced the evolution of proposed models and presented the interest of social network analysis in this field.

Among solutions that integrate social contacts yet, we have distinguished network models and multi-agent-based models that tend to match the complexity of real social networks according to different degrees but are still limited. In particular, we have shown that current techniques would take big advantages from graph labeling and link mining. We have identified precise situations in which these techniques combined with a network model seem promising.

Thus, in the short term, we plan to implement these solutions and study how they compare to other techniques based only on some centrality measures. In the long term, it seems interesting to exploit the modeling power of multi-agent systems in order to obtain richer data from simulations for supply to analytical methods.

REFERENCES

[1] A.L. Adams, J.S. Koopman, S.E. Chick, and P.J. Yu. Germs: an epidemiologic simulation tool for studying geographic and social effects on infection transmission. *Winter Simulation Conference*, 2:1549–1556, 1999.

[2] Andrew L. Adams, Stephen E. Chick, Daniel C. Barth-Jones, and James S. Koopman. Simulations to evaluate hiv vaccine trial designs. *Simulation*, 71(4):228–241, 1998.

[3] AA Al-Azem, A Jolly, E Hershfield, and A Kabani. Social network analysis in tuberculosis. *Interscience Conference on Antimicrobial Agents and Chemotherapy*, 38(55), 1998.

[4] R. M. Anderson and R. M. May. Population biology of infectious diseases: Part i. *Nature*, 280:361–367, 1979.

[5] Roy M. Anderson, Robert M. May, and B. Anderson. *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications, 1992.

[6] Ralitsa Angelova and Gerhard Weikum. Graph-based text classification: learn from your neighbors. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 485–492, New York, NY, USA, 2006. ACM.

[7] N. T. J. Bailey. *The Mathematical Theory of Epidemics*. Griffin, 1957.

[8] N. T. J. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, 1975.

[9] Albert-Laszlo Barabasi. *Linked: The New Science of Networks*. Perseus Publishing, 2002.

[10] Albert-Lszl Barabsi and Rka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509 – 512, 1999.

[11] J. A. Barnes. Class and committees in a norwegian island parish. *Human Relations*, 7:39–58, 1954.

[12] C. Barrett, R. Beckman, K. Berkbigler, K. Bisset, B. Bush, K. Campbell, S. Eubank, K. Henson, J. Hurford, D. Kubicek, M. Marathe, P. Romero, J. Smith, L. Smith, P. Speckman, P. Stretz, G. Thayer, E. Eeckhout, and M.D. Williams. Technical report.

[13] Christopher L. Barrett, Keith R. Bisset, Stephen G. Eubank, Xizhou Feng, and Madhav V. Marathe. Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, SC '08, pages 37:1–37:12, Piscataway, NJ, USA, 2008. IEEE Press.

[14] Daniel C. Barth-Jones, Andrew L. Adams, and James S. Koopman. Monte carlo simulation experiments for analysis of hiv vaccine effects and vaccine trial design. In *Proceedings of the 32nd conference on Winter simulation*, WSC '00, pages 1985–1994, San Diego, CA, USA, 2000. Society for Computer Simulation International.

[15] N. Becker. The uses of epidemic models. *Biometrics*, 35(1):295–305, 1979.

[16] Mikhail Belkin and Partha Niyogi. Using Manifold Structure for Partially Labelled Classification. In *NIPS*, pages 929+, 2002.

[17] P.-M. Binder. Theories of almost everything. *Nature*, 455:884–885, 2008.

[18] Keith R. Bisset, Jiangzhuo Chen, Xizhou Feng, V.S. Anil Kumar, and Madhav V. Marathe. Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In *Proceedings of the 23rd international conference on Supercomputing*, ICS '09, pages 430–439, New York, NY, USA, 2009. ACM.

[19] Keith R. Bisset, Xizhou Feng, Madhav V. Marathe, and Shrirang M. Yardi. Modeling interaction between individuals, social networks and public policy to support public health epidemiology. In Ann Dunkin, Ricki G. Ingalls, Enver Yücesan, Manuel D. Rossetti, Ray Hill, and Björn Johansson, editors, *Winter Simulation Conference*, pages 2020–2031. WSC, 2009.

[20] Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, 2008.

[21] N. Boccara, O. Roblin, and M. Roger. Automata network predator-prey model with pursuit and evasion. *Phys. Rev. E*, 50(6):4531–4541, Dec 1994.

[22] E. Bott. *Family and social network*. New-York, 1957.

[23] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, November 1987.

[24] Yi-Da Chen, Chunju Tseng, Chwan-Chuen King, Tsung-Shu Joseph Wu, and Hsinchun Chen. Incorporating geographical contacts into social network analysis for contact tracing in epidemiology: a study on taiwan sars data. In *Proceedings of the 2nd NSF conference on Intelligence and security informatics: BioSurveillance*, BioSurveillance'07, pages 23–36, Berlin, Heidelberg, 2007. Springer-Verlag.

[25] Nicholas A. Christakis and James H. Fowler. Social network sensors for early detection of contagious outbreaks. *PloS one*, 5(9)(9), September 2010.

[26] R. M. Christley, G. L. Pinchbeck, R. G. Bowers, D. Clancy, N. P. French, R. Bennett 2, and J. Turner. Infection in social networks: Using network analysis to identify high-risk individuals. *American Journal of Epidemiology*, 162(10):1024–1031, 2005.

[27] Vittoria Colizza and Alessandro Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251:450–467, 2008.

[28] VJ Cook, SJ Sun, J Tapia, SQ Muth, DF Arguello, BL Lewis, RB Rothenberg, and PD McElroy. Transmission network analysis in tuberculosis contact investigations. *The Journal of infectious diseases*, 196(10):1517–1527, 2007.

[29] Courtney D. Corley, Armin R. Mikler, Diane J. Cook, and Karan Singh. Dynamic intimate contact social networks and epidemic interventions. *International Journal of Functional Informatics and Personalised Medicine*, 1(2):171 – 188, 2008.

[30] Darren P. Croft, Richard James, and Jens Krause. *Exploring Animals Social Networks*, chapter 3. Visual Exploration. Princeton, 2008.

[31] Eric Daude, Pierrick Tranouez, and Patrice Langlois. A multiagent urban traffic simulation. part ii: dealing with the extraordinary. *CoRR*, abs/0910.1026, 2009.

[32] P De, A E Singh, T Wong, W Yacoub, and A M Jolly. Sexual network analysis of a gonorrhoea outbreak. *Sex Transm Infect*, 80:280285, 2004.

[33] Pradip De and Sajal K. Da. *Algorithms and Protocols for Wireless Sensor Networks*, chapter 3. Epidemic Models, Algorithms, and Protocols in Wireless Sensor and Ad Hoc Networks. published by John Wiley and Sons, 2008.

[34] Ludovic Denoyer and Patrick Gallinari. A ranking based model for automatic image annotation in a social network. In *ICWSM*, 2010.

[35] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley Series in Mathematical & Computational Biology, 2000.

[36] K. Dietz and J. Heesterbeek. Daniel bernoullis epidemiological model revisited. *J. Math. Biosci.*, 180:121, 2002.

[37] d'Onofrio A. Stability properties of pulse vaccination strategy in seir epidemic model. *Math Biosci*, 179(1):5772, 2002.

[38] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, chapter 21. Epidemics. Cambridge University Press, 2010.

[39] Emmanuel Eliot and Eric Daude. Spatial diffusion of epidemics : Approaches of complexities in geography. *Espace populations socits*, 2(3):403–416, 2006.

[40] S. Eubank. Network based models of infectious disease spread. *Jpn J Infect Dis*, 58(6):S9–13, 2005.

[41] Stephen Eubank. Scalable, efficient epidemiological simulation. In *Proceedings of the 2002 ACM symposium on Applied computing*, SAC '02, pages 139–145, New York, NY, USA, 2002. ACM.

[42] Stephen Eubank, V. Anil Kumar, and Madhav Marathe. Epidemiology and wireless communication: Tight analogy or loose metaphor? In Pietro Lio, Eiko Yoneki, Jon Crowcroft, and Dinesh Verma, editors, *Bio-Inspired Computing and Communication*, volume 5151 of *Lecture Notes in Computer Science*, pages 91–104. Springer Berlin, Heidelberg, 2008.

[43] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, May 2004.

[44] Jacques Ferber. *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley Professional, February 1999.

[45] N. M. Ferguson and M. J. Keeling et al. Planning for smallpox outbreaks. *Nature*, 425(6959):681–685, 2003.

[46] J. Filipe and G. Gibson. Comparing approximations to spatio-temporal models for epidemics with local spread. *Bulletin of Mathematical Biology*, 63:603–624, 2001. 10.1006/bulm.2001.0234.

[47] J. A. N. Filipe. Studying and approximating spatio-temporal models for epidemic spread and control. *Philos Trans R Soc Lond B Biol Sci.*, 353(1378):21532162, 1998.

[48] M. Gardner. The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223:120–123, October 1970.

[49] A.C. Ghani, J. Swinton, and G.P. Garnett. The role of sexual partnership networks in the epidemiology of gonorrhea. *Sexually Transmitted Diseases*, 24(1):4556, 1997.

[50] T. R. Gottwald, K. M. Reynolds, C. L. Campbemm, and L. W. Timmer. Spatial and spatiotemporal autocorrelation analysis of citrus canker epidemics in citrus nurseries and grove in argentina. *Phytopathology*, 82(8):843–851, 1992.

[51] M. Elizabeth Halloran, Ira M. Longini Jr., Azhar Nizam, and Yang Yang. Containing bioterrorist smallpox. *Science*, 298(5597):1428–1432, 2002.

[52] Raymond Heatherly, Murat Kantarcioglu, and Xiaohu Li. Social Network Classification Incorporating Link Type. In *IEEE Intelligence and Security Informatics*, , Dallas, Texas,, June 2009.

[53] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42:599–653, 2000.

[54] Herbert W. Hethcote and James A. Yorke. *Gonorrhea Transmission Dynamics and Control*. Springer, 1984.

[55] Michael Huston, Donald Deangelis, and Wilfred Post. New Computer Models Unify Ecological Theory. *BioScience*, 38(10):682–691, 1988.

[56] A. Jolly, S. Muth, J. Wylie, and J. Potterat. Sexual networks and sexually transmitted infections: A tale of two cities. *Journal of Urban Health*, 78:433–445, 2001. 10.1093/jurban/78.3.433.

[57] M.J. Keeling and B.T. Grenfell. Individual-based perspectives on r0. *J. Theor. Biol.*, 203:51–61, 2000.

[58] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115:700–721, 1927.

[59] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. ii. the problem of endemicity. *Proceedings of the Royal Society of London. Series A*, 138:55–83, 1932.

[60] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. iii. further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A*, 141:94–122, 1933.

[61] A. S. Klovdahl. Social networks and the spread of infectious diseases: the aids example. *Soc Sci Med*, 21(11):1203–16, 1985.

[62] A. S. Klovdahl, E. A. Graviss, A. Yaganehdoost, M. W. Ross, A. Wanger, G. J. Adams, and J. M. Musser. Networks and tuberculosis: an undetected community outbreak involving public places. *Social Science & Medicine*, 52(5):681 – 694, 2001.

[63] A.S. Klovdahl, J.J. Potterat, D.E. Woodhouse, J.B. Muth, S.Q. Muth, and W.W. Darrow. Social networks and infectious disease: The colorado springs study. *Social Science & Medicine*, 38(1):79 – 88, 1994.

[64] Y. A. Kuznetsov and C. Piccardi. Bifurcation analysis pf periodic seir and sir epidemic models. *J Math Biol*, 32(2):109–21, 1994.

[65] V. Levchenkov and L. Levchenkova. The multirelation method for link prediction in social networks. *Computational Mathematics and Modeling*, 17:88–96, 2006. 10.1007/s10598-006-0015-3.

[66] K. Lewin. *Field theory in the social sciences*. New York, 1951.

[67] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.*, 58:1019–1031, May 2007.

[68] S. L. Lima and P. A. Zollner. Towards a behavioral ecology of ecological landscapes. *Trends in Ecology & Evolution*, 11(3):131–135, 1996.

[69] I. Longini and A. Nizam et al. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.

[70] G. Macdonald. The analysis of equilibrium in malaria. *Trop Dis Bull*, 49(9):813–829, 1952.

[71] KM Macqueen. The epidemiology of hiv transmission: trends, structure and dynamics. *Annu Rev Anthropol*, 23:509–26, 1994.

[72] Francis Maes, Stphane Peters, Ludovic Denoyer, and Patrick Gallinari. Simulated iterative classification a new learning procedure for graph labeling. In Wray Buntine, Marko Grobelnik, Dunja Mladenic, and John Shawe-Taylor, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 5782 of *Lecture Notes in Computer Science*, pages 47–62. Springer Berlin / Heidelberg, 2009.

[73] P. D. McElroy, R. B. Rothenberg, R. Varghese, R. Woodruff, G. O. Minns, S. Q. Muth, L. A. Lambert, and R. Ridzon. A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. *International journal of tuberculosis and lung disease*, 7(12):S486–S493, 2003.

[74] BJ McMahon, WL Alward, Hall DB, Heyward WL, Bender TR, Francis DP, and Maynard JE. Acute hepatitis b virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. *J Infect Dis*, 151(4):599–603, 1985.

[75] L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham. Network theory and sars: predicting outbreak diversity. *Journal of theoretical biology*, 232(1):71–81, January 2005.

[76] S. Milgram. The small world problem. *Psychology Today*, 1:61–67, 1967.

[77] Denis Mollison. *Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press, 1995.

[78] M. Morris. Epidemiology and social networks: Modeling structured diffusion. *Sociological methods & research*, 22(1):99–126, 1993.

[79] Martina Morris. *Network Epidemiology A Handbook for Survey Design and Data Collection*. Oxford University Press, 2004.

[80] Tsuyoshi Murata and Sakiko Moriyasu. Link prediction of social networks based on weighted proximity measures. *IEEE/WIC/ACM International Conference on Web Intelligence*, pages 85–88, 2007.

[81] Boccara N. A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behaviour. *Journal of Physics A: Mathematical and General*, 27:1585–1597(13), 1994.

[82] M. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167256, 2003.

[83] M. E. J. Newman. *Networks: An Introduction*, volume 1, chapter 17. Epidemics on Networks, pages 627–676(50). Oxford University Press, 25 March 2010.

[84] Institute of Medicine of the National Academies. C. on modeling community containment for pandemic influenza and i. of medicine, modeling community containment for pandemic influenza: A letter report. *Washington D.C.: The National Academies Press*, 2006.

[85] Peter Parham. *The Immune System*. Garland Science; 3 edition, 2009.

[86] André Reynaldo Santos R. Périssé and José Augusto da A. Costa Nery. The relevance of social network analysis on the epidemiology and prevention of sexually transmitted diseases. *Cadernos de saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública*, 23 Suppl 3, 2007.

[87] Alexandrin Popescul, Rin Popescul, and Lyle H. Ungar. Statistical relational learning for link prediction. 2003.

[88] K. M. Reynolds and L. V. Madden. Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology*, 78(2):240–246, 1998.

[89] C. J. Rhodes and R. M. Anderson. Dynamics in a lattice epidemic model. *Physics Letters A*, 210:183–188, 1996.

[90] S. Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301, 2007.

[91] Kahn R Kissinger P Rice J Kendall C Coughlin S Farley TA. Rosenberg D, Moseley K. Networks of persons with syphilis and at risk for syphilis in louisiana: Evidence of core transmitters. *Sexually Transmitted Diseases*, 26(2):108–114, 1999.

[92] Ronald Ross. *The prevention of malaria*. Princeton University Press, 1910.

[93] RB Rothenberg, JJ Potterat, DE Woodhouse, SQ Muth, WW Darrow, and AS Klovdahl. Social network dynamics and hiv transmission. *AIDS*, 12(12):1529–1536, 1998.

[94] RB Rothenberg, C Sterk, KE Toomey, JJ Potterat, D Johnson, M Schrader, and S. Hatch. Using social network and ethnographic tools to evaluate syphilis transmission. *Sexually Transmitted Diseases*, 25(3):154160, 1998.

[95] L.A. Rvachev and Ira M. Longini. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75:3?22, 1985.

[96] Marcel Salathe and James H. Jones. Dynamics and control of diseases in networks with community structure. *PLoS Comput Biol*, 6(4), 04 2010.

[97] Marcel Salathé, Maria Kazandjieva, Jung W. Lee, Philip Levis, Marcus W. Feldman, and James H. Jones. A high-resolution human contact network for infectious disease transmission. Aug 2010.

[98] Lisa Sattenspiel and Klaus Dietz. A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences*, 128:71–91, 1995.

[99] W. Sutherland. *From individual behavior to population ecology*. Oxford, 1996.

[100] Ben Taskar, Ming fai Wong, Pieter Abbeel, and Daphne Koller. Link prediction in relational data. In *in Neural Information Processing Systems*, 2003.

[101] Pierrick Tranouez, Patrice Langlois, and Eric Daude. A multiagent urban traffic simulation part i: dealing with the ordinary. *CoRR*, abs/0909.1021, 2009.

[102] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[103] Alexei Vazquez. Epidemic outbreaks on structured populations. *Journal of Theoretical Biology*, 245:125–129, 2007.

[104] Rodrick Wallace. Social disintegration and the spread of aids: Thresholds for propagation along [']sociogeographic' networks. *Social Science & Medicine*, 33(10):1155 – 1162, 1991.

[105] D.J Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature 393*, 6684:409–10, 1998.

[106] Duncan J. Watts, Roby Muhamad, Daniel C. Medina, and Peter S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11157–11162, 2005.

[107] Douglas B. West. *Introduction to Graph Theory (2nd Edition)*. Prentice Hall, 2000.

[108] S. Hoya White, A. Martn del Rey, and G. Rodrguez Snchez. Modeling epidemics using cellular automata. *Applied Mathematics and Computation*, 186(1):193 – 202, 2007.

[109] World Health Organization (WHO). World health organization: Report on infectious diseases. *http://www.who.int/infectious-disease-report/*, 1999.

[110] World Health Organization (WHO). Scaling up the response to infectious diseases a way out of poverty. pages 1–125, 2002.

[111] World Health Organization (WHO). International spread of disease threatens public health security. 2007.

[112] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.

[113] S Zhao, Z Xu, and Y Lu. A mathematical model of hepatitis b virus transmission and its application for vaccination strategy in china. *Int J Epidemiol*, 29(4):744–52, 2000.