Diffusion in Dynamic Social Networks: Application in Epidemiology

Erick Stattner, Martine Collard, and Nicolas Vidot

LAMIA Laboratory
University of the French West Indies and Guiana
France
estattne,mcollard,nvidot@univ-ag.fr
http://lamia.univ-ag.fr

Abstract. Structure and evolution of networks have been areas of growing interest in recent years, especially with the emergence of Social Network Analysis (SNA) and its application in numerous fields. Researches on diffusion are focusing on network modeling for studying spreading phenomena. While the impact of network properties on spreading is now widely studied, involvement of network dynamicity is very little known. In this paper, we address the epidemiology context and study the consequences of network evolutions on spread of diseases. Experiments are conducted by comparing incidence curves obtained by evolution strategies applied on two generated and two real networks. Results are then analyzed by investigating network properties and discussed in order to explain how network evolution influences the spread. We present the *MIDEN* framework, an approach to measure impact of basic changes in network structure, and *DynSpread*, a 2D simulation tool designed to replay infections scenarios on evolving networks.

Keywords: Information Spreading, Dynamic network, Evolution, Framework, Simulation.

1 Introduction

Network modeling involves a set of items, represented by nodes (also called vertexes), that are linked by connections. While seminal works were first conducted in mathematics through the graph theory, this domain has known a significant growth in recent years.

In last decades, network analysis has been the subject of an active research domain, so-called "*Science of Networks*" [2,5], an emerging scientific discipline that focuses on relationships maintained between entities, and not on entities themselves. Social science, with the occurrence of WEB 2.0 and Social Network Analysis (*SNA*), has provided the most popular works, with Milgram [19] on small world phenomenon, or Bott [4] on families. Other domains are also related to network analysis, such as biology [16], ethology [11] or computer science [3],

A. Hameurlain et al. (Eds.): DEXA 2011, Part II, LNCS 6861, pp. 559–573, 2011.

since this kind of representation is particularly fitted to understand information spreading phenomena. For instance spreading of viruses, diseases, rumors, knowledge or fashion behave very similarly [5,12].

This work is focused on diffusion phenomenon and addresses the particular issue of spread of diseases. Indeed, in numerous cases of disease spreading, social contacts seem to be main factors of transmission. Sexual Transmitted Diseases (STDs) are a good example of diseases that depend solely on personal contacts for dissemination. Thus, several studies have focused on social networks and have demonstrated their relevance in disease spreading [18,9,7]. However, although the effect of network properties on spreading is now widely studied, the impact of network changes is an emerging field. One obvious pitfall is the lack of real data and most works commonly handle snapshots of a network at a given time, that do not fully reflect real world networks.

In this paper, we address the issue of dissemination in dynamic networks. Unlike methods that study spreading on static networks, we propose a framework to assess impact of evolving mechanisms, by selecting and comparing some typical evolution strategies. Experiments are conducted by comparing incidence curves obtained with several evolution strategies and are then analyzed by investigating network properties. In this first stage, we have restricted the analysis to strategies thats only create new links. Our results provide an original insight about how and why network evolution impacts the spread. We show indeed that new links appearing in the social network generally emphasis the epidemic spread and speed up the process. Of course this assertion has to be modulated according to the evolution strategy applied. In epidemiology, understanding such implications is essential to help public health officials in the prevention and the development of appropriate strategies taking into account changes occurring on real world networks.

This paper is organized in 7 sections as follows. In Section 2 we present main previous works on networks, with particular emphasis on their application in epidemiology. Section 3 presents our motivations and objectives, and details the framework we propose to measure the involvement of network dynamics. Experiments and results are presented in Section 4. In Section 5, we discuss these results by investigating both effects of evolution strategies and changes on network properties. Section 6 is devoted to *DynSpread*, the 2D simulation tool that implements our framework. We conclude and present future directions in Section 7.

2 Previous Works

Epidemiology is the science that focuses on infectious diseases. De and Das [12] 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 population or individual level. It refers to all methods of modeling [24], analysis [22] or monitoring [7] of the spread in a given system, for identifying risk factors and determining optimal intervention approaches to clinical practice and preventative medicine.

2.1 Compartment Models

Although the biological interest for this phenomenon is undeniable, other scientific communities have contributed to its understanding: anthropology [4], mathematics [17] or computer science [10,3].

Mathematicians were the first to address modeling issues, through compartment models [24]. This kind of models assume that (1) a population can be divided into a set of compartments, according to the level of the disease development, and (2) individuals have equal probability to change compartment. The two main compartment models defined in epidemic literature are $Susceptible - Infected$ Model (SI Model), that assumes individuals may become infected with probability β and *SIR Model*, which adds a *Recover* state reached by infected individuals with a certain probability γ. On the same paradigm, many other models [24] can be found in the literature: SIS, $SIRS$, etc.

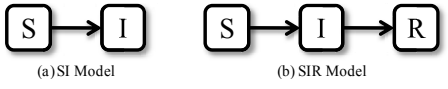


Fig. 1. Two Examples of Compartments Models

However, such approaches remain very simple and do not reflect the real complexity of human interactions. Indeed, in real world people are actually connected to a small portion of individuals, and this portion is obviously not chosen randomly. Introduced by pioneering works of Klovdahl [18] on AIDS, network modeling have found various applications in epidemiology, since a significant factor of the outbreak and the behavior of diseases is the structure and the nature of human interactions through which it spreads.

2.2 Networks and Epidemics

Traditionally, a network is described by a graph G, defined $G = (V, E)$, where V is the set of vertexes and E the set of edges in the graph $E \subseteq V \times V$. The neighbors $N(i)$ of vertex i is defined as $N(i) = \{j \mid ij \in E\}$. Other individual measures prove to be interesting for networks of very different types [20]. The *Degree* is the number of neighbors of a vertex. More formally, the degree d_i of node i is the cardinality of the set of its neighbors, *i.e.* $d_i = |N(i)|$. The *Clustering Coefficient* C_i of a node i indicates how close the neighbors of node i are to being a clique, $C_i = \frac{2t_i}{d_i(d_i-1)}$, where t_i is the number of triangles for which node i is a part.

Initially, networks were studied with the objective to understand various real systems in disciplines ranging from communication networks to ecological webs. Newman [20] classifies works according to three categories: Node-Based Measure, Statistical Properties of Networks and Dynamics of Networks.

At node level, networks are characterized by some individual properties of their nodes. Typical works try to classify or identify the role of nodes by

understanding which individuals are more connected to others or have most influence, or whether and how individuals are connected to one another through the network, etc. In epidemiology, such measures have been used to identify high-risk individuals. For example, Christley et al. [9] compare various node-based measures to identify this kind of individuals. Chen et al. [6] show the changes in the degree of nodes according to several kinds of interactions and their role for the transmission.

At global level, scientists try to classify networks focusing on the distribution of given properties. A typical example is the network classification we can found in the literature [1,20]. In epidemiology, works have been conducted to understand the propagation phenomena [14] and propose intervention strategies taking into account network topology [22,7].

According to the dynamic point of view, many recent studies try to understand and reproduce the manner in which a network evolves [25]. Thus, several models have been proposed to reflect growth processes inducing to particular structural features observed on real world networks [20,13]. However, it is interesting to note that the issues of *dynamics of networks* and the *dynamics on networks* are still independent fields. Indeed, the impact of network evolution on spread of infectious diseases is a very new research axis. In this area, the mathematical approach of Gross et al. [15] studies the impact of links deletion on spreading. Read et al. [21] show how changes in the frequency of encounters between individuals may impact the dissemination. More recently, Christensen et al. [8] have measured the effect of changes in demographic attributes within population on the disease transmission.

In this paper, we focus on this last issue of dynamicity in networks. The next section is devoted to the method we propose to understand how different changes in a network have consequences on the transmission of disease.

3 Objectives and Method

Networks are alive and animated objects, in which nodes can appear and disappear, links can be created, removed, or can even evolve. Thus, focusing on the dynamic issue in networks inevitably raises multiple questions, both on networks and on information diffusion: How does a network evolve? What do the changes operate on its properties? How is the network topology influenced by the way the network evolves and how does it affect the spread behavior?

3.1 Motivations and Objectives

Among social networks, we can identify on one hand, static structures that do not meet much evolution. For instance, co-author networks have few new links created. But on the other hand, networks based on geographic contacts are likely to meet much evolution with frequent deletion and creation of links, according to individual mobility. We focus on this kind of networks.

In epidemiology, a concrete motivation to study this question is demonstrated by intervention strategies that are currently proposed and are generally focused

on node-based measures. For example, the intervention strategy that gives best results is to vaccinating individuals with the highest degree. However, it is realistic to think that individuals with the highest degree at time t will probably not be in the same state at time $(t + 1)$, due to changes that occur in the network. Therefore, dynamic appears to have a strong and real impact on the spread, and may be an essential factor for the behavior of a disease.

Nevertheless, to the best our knowledge, no empirical or even comparative studies to assess the effect of different network evolutions are available. Indeed, we may assume that dynamic plays an important role in the diffusion of information through the network, and therefore should be taken into account to understand diseases behavior in evolving networks and propose suitable strategies to prevent and control epidemics. Thus, our objectives are **define a framework** to assess impacts of network evolution **measure and compare** the impact of network dynamic, by comparing the incidence curves induced by several evolution strategies, and **understand causes and effects** at global and individual levels.

3.2 Evolution Strategies

In order to carry out such a work, we compare the effects of four well known evolution models, accepted in the literature as reproducing changes observed in real world networks. These strategies are highlighted by schemes on Figure 2.

Random (R) is a process that creates links randomly between nodes. This evolution mechanism is particularly used to model evolution of networks for which we have no knowledge on the development.

Triadic Closure (TD) is one of the first social networking concepts. It can be expressed as "*friends of my friends become my friends*". More formally, it is the process through which a node is likely to create a link with the neighbors of its neighbors.

Global Connection (GC) corresponds to an abstracted mechanism of social link formation, through which a node creates links solely outside of its circle of close friends *i.e.* beyond friends of its friends.

Preferential Attachment (PA) is a kind of evolution in which a node is more likely to connect to one with high degree. Although it is fairly recent, it is now widely studied, particularly in social science where it makes sense. Indeed it is more likely for a person to connect to someone with a large number of connections, as those people tend to be more social and popular.

3.3 Studied Networks

In this work, we focus on four networks: two generated networks *GEN1*, *GEN2* and two real networks *HS* and *PL*. We have chosen these networks because they are representative of different types of networks currently referenced in the literature.

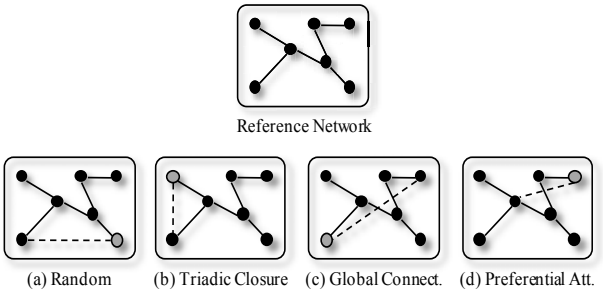


Fig. 2. Starting from a reference network, examples of links generated with (a)Random, (b)Triadic Closure, (c)Global Connection, (d)Preferential Attachment

- ***GEN1* and *GEN2*** are respectively obtained by Erdos-Renyi [13] and Barabasi-Albert [20] models. *GEN1* is a classic random network that has been the subject of intensives researches. *GEN2* represents the kind of network most commonly observed in real world networks, such as the Internet, telephone calls network, sexual network or friendship network, known as scale-free network.
- ***HS* and *PL*** are respectively obtained by real and synthetic situations. *HS* correspond directly to the high school interactions graph proposed by Salathe [23] and *PL* represents a synthetic population of the city of Portland extracted from *EpiSims*, an epidemiological simulation system prior to *EpiSimdemics* [3].

Generated networks are used because the work we are conducting is not confined to epidemiology, since mechanisms of transmission of information, or rumors are very similar. Therefore, it is important to understand how the disease behaves on evolving generic networks, to transpose it in other paradigms. Figure 3 details the main characteristics of the networks described above ($\#comp$ is the number of connected components and cc is the clustering coefficient of the network).

3.4 MIDEN Framework

For each network, we have no a priori knowledge about their evolutions. This allows us to make several assumptions about its development by fitting the network with evolution mechanisms. Thus, to assess the impact of network changes, we apply each evolution strategy (R, TD, GC and PA) to these networks, and compare their effect. In this preliminary work, we study the dynamics in an empirical manner, by setting the number of nodes and by considering the dynamic through the addition of new links only.

Let us give $T = \langle t_0, t_1, ..., t_m \rangle$, as the time sequence over which the disease transmission is studied with $\forall j \in [0..m]$, $t_j < t_{j+1}$ and $G = \langle G_{t_0}, G_{t_1}, ..., G_{t_m} \rangle$ as the sequence of networks, where each $G_{tj} = (V, E_{tj})$ represents the state of

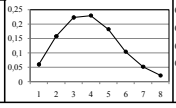
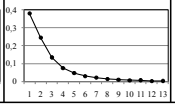
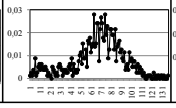
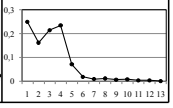
		Networks			
		GEN1	**GEN2**	**HS**	**PL**
general	**Origine**	Generated	Generated	High School	Portland
	#nodes	4771	3233	788	4829
	#links	7481	5154	26801	7455
	Density	0.000657	0.000986	0.086873	0.0006395
	#comp	17	1	1	1
degree	**avg**	3.136	3.188	68.195	3.087
	max	11	118	174	17
	Distribution				
cc	**avg**	0.00056	0.00427	0.49877	0.60880

Fig. 3. Detailed Networks Features

the network at time t_j, with $t_j \in T$. V is the set of individuals and E_{t_j} is the set of edges present in the network at time t_j, with $E_{t_{(j-1)}} \subseteq E_{t_j} \subseteq V \times V$.

Transmission of an infectious agent is simulated according to the $Suceptible - Infectious - Recover$ model.

Let us denote F_{t_j} the function that returns the state of a node i at time t_j, $F_{t_j} : V \rightarrow \{S, I, R\}$. Thus, at each time t_j of T, we have the set of infected nodes $I_{t_j} = \{v \in V | F_{t_j}(v) = I\}$. Let us also define $N^i_{t_j}$ as the set of infected neighbors of the node i, $N^i_{t_j} = \{v \in V | iv \in E_{tj} \ \ and \ \ F_{t_j}(v) = I\}$ at t_j.

Let β be the probability of transmission per contact and γ the probability of recover. We denote W the applied evolution strategy and Q the speed of evolution, *i.e.* the number of links created at each iteration.

1. Outbreak: Before introduction of the infectious agent, at t_0, the entire population is assumed to be susceptible and we assume that no link is created yet in the network G_{t_0}. We begin by selecting the patient zero, the first individual that will be contaminated by the pathogen. This step simply consists in randomly selecting an individual, z, among the population and change his state S to I, *i.e.* $\exists!\ z \in V, F_{t_0}(z) = I \ \ and \ \ \forall v \in V - \{z\}, F_{t_0}(v) = S$.

2. Wave Transmission ($S \rightarrow I$): Once the patient zero is infected, disease can spread, with a certain probability, from an infected individual to a susceptible individual if there exists an edge between them. More formally, if a susceptible node n_i has k_i infected neighbors at time t_j, it can become infected with the probability $1 - (1 - \beta)^{k_i}$ with $k_i = |N^i_{t_j}|$. In this way, the probability of being infected increases with the number of infected neighbors.

3. Network evolution : After the wave transmission, the proportion of infected individuals is stored and the network evolves to the $G_{(t_j+1)}$ state, where $(t_j + 1) \ \in \ T$. The evolution is carried out according to the strategy W chosen and the evolution speed Q. In practical terms, Q nodes are randomly selected from the network and link building is applied from these nodes.

4. Recovery ($I \rightarrow R$): Each infected individual has a probability γ to recover. Once the node is in R state, it cannot transmit the infectious agent again. Its immunity is supposed to be permanent and it cannot return in the S state.

Processes 2, 3 and 4 are repeated until susceptible individuals become fully extinct, *i.e.* $I_{t_j} = \emptyset$. Once the epidemic is over, duration of infection is stored. The algorithm below sums up the proposed approach.

Data: Probability of transmission β, and probability of recover γ
Input: Network G, Strategy W, and Evolution Speed Q
Result : List L, of infected individuals during T

Function MIDEN(G : **Network**, W : **Strategy**, Q : **Speed**) : **List**
    L : **List** $\leftarrow \emptyset$
    t : **Time** $\leftarrow 0$
    Infect Patient Zero
    While $(I_t \neq \emptyset)$ **do**
        Infect S-nodes i with probability $1-(1-\beta)^{k_i}$, with $k_i = |N_t^i|$
        add $\frac{|I_t|}{|V|}$ to L
        Network evolves to $G_{(t+1)}$ according to W and Q
        Recover I-nodes i with probability γ
        $t \leftarrow t+1$
    done
    return L
End

Algorithm 1: MIDEN(framework for Measure Impacts of Dynamic on Epidemic Networks)

4 Experiments and Results

The framework described above was experimented to analyze effects of these different evolution strategies. As a first step, this section shows and analyzes the direct effects of the different strategies on the strength of the epidemic and its appearance in time. Afterwards, in the next section, we deepen our analytical work by investigating explanations on the side of changes in network properties.

4.1 Test Bed

Disease behavior depends on many parameters such as the number of initial infected individuals, the probability of transmission or the probability of recover. However, changes in these parameters often influence only the virulence of the epidemic and are quite well known. In the issue we address, the most relevant parameter seems to be the evolution speed of the network. Thus to study the implication of this parameter on the behavior of a disease, we vary *the network*, *the evolution strategy*, and *the evolution speed*.

Epidemics may have varying durations, so we set $T = 120$. Thus the data collection was restricted over a period of 120 iterations. The probability of

transmission was set at 0.1 and the probability of recover was set at 0.2, *i.e.* $\beta = 0.1$ and $\gamma = 0.2$. Each test was performed upon 100 runs. Then, the average of the result obtained was calculated. For each test, a single evolution strategy was applied and the evolution speed of the network remained constant. All tests were performed by *DynSpread*, our simulation tool presented in Section 6 and were conducted with the following simulation environment: Intel Core 2 Duo P8600 2.4Ghz, 3Go Ram, Microsoft Windows Vista 32Bits, Java JDK 1.6.

4.2 Results

Figure 4 plots resulting an incidence curve for each strategy, according to the kind of network and the speed of evolution (x axis). For a given network and a given speed, results obtained by the four strategies are presented on the same scheme, through curves representing the percentage of infected nodes at each iteration of the *MIDEN* algorithm, also called incidence curves. The incidence curves obtained by epidemic *without evolution* strategy are always plotted as a reference to compare to the others.

If we first underline common characteristics observed on *GEN1*, *GEN2* and *PL*, we obviously observe the direct impact of evolution speed. Creating new links obviously emphasizes the epidemic spread. More precisely, two main observations can be made on: **(1)** The virulence of the epidemic, since for these three networks, peaks values increase with the speed. **(2)** The timing of epidemics, since we can observe that when the evolution speed increases, epidemic peaks appear earlier.

We can observe specific behavior, depending on speed evolution for each kind of network. *GEN1* and *PL* behave rather similarly. For example, we can observe that a speed of 10 links per iteration is not sufficient to generate an epidemic in networks *GEN1* and *PL*. However, when the *speed* is set at 50, *PA* is the only strategy able to generate an epidemic on these two networks. We notice that after a speed threshold approximately equals to 100, all strategies generate an epidemic peak that is increasing according to the speed.

Although all strategies generate an epidemic on network *GEN2*, at *speed* 10, the difference between strategies is insignificant and results remain very close to those obtained without evolution. From *speed* 50, the same trends, as for *GEN1* and *PL*, are observed: epidemic peak increases according to the speed.

The case of *HS* network is unusual. To understand why different strategies have no effect, we must address the data collection process. Data were collected by wireless sensors grafted on individuals, and interactions were recorded when two sensors were close enough. Although only interactions above 1 min were considered, the resulting network is very dense and thus allows a good spreading of the disease (in our test bed). We can conclude that for this kind of network, impact of dynamic is negligible. Moreover, it was impossible to generate the results for strategy *GC*, since all nodes rapidly become directly connected.

Finally, when the dynamic is high enough, common trends can be observed: *PA* strategy generally gives an epidemic curve with a peak that systematically is higher than in the others three strategies. It induces the earliest occurrence of the epidemic. *R* and *GC* strategies are always very close, since they have

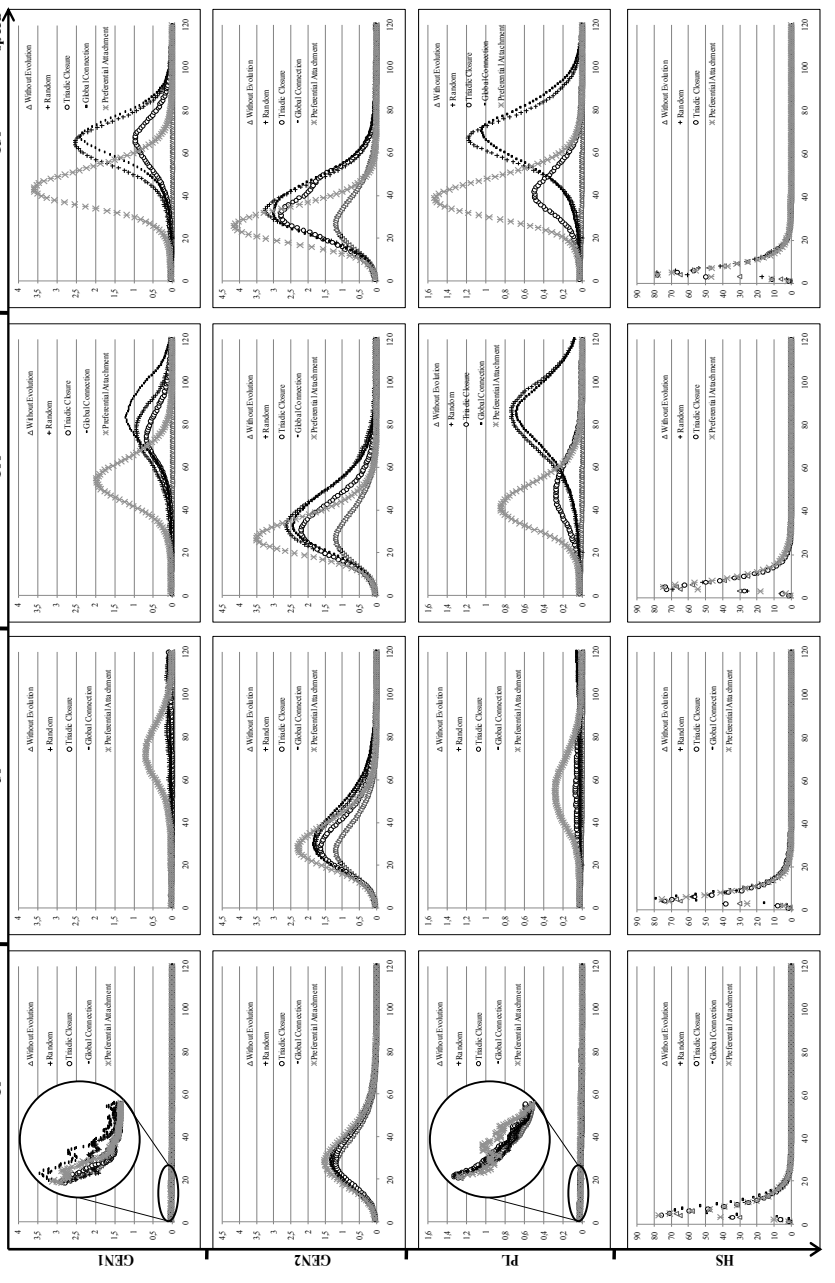
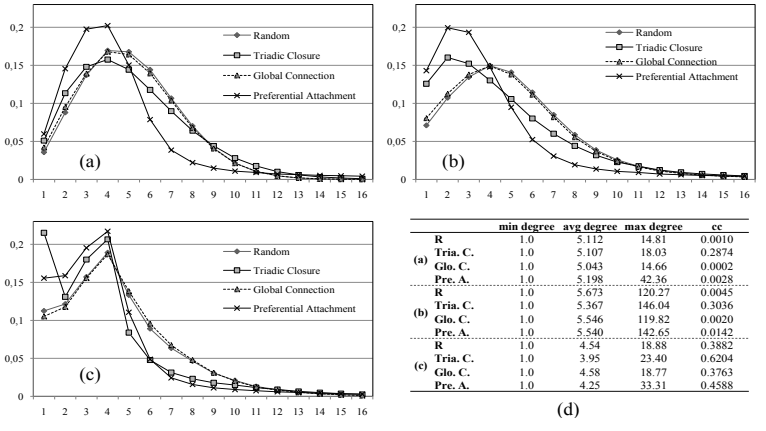


Fig. 4. Incidence curves according to the Network and the Evolution Speed

epidemic curves that follow the same variations in time. The peak obtained by *TD* strategy is always the lowest.

5 Discussion

Since the impact of the dynamic appears quite obviously in the results presented above, we investigate now the question and tend to explain what happens at the network level. For this, we focus on changes that occur on network features. As shown in Figure 4, there is a strong difference between strategies beginning from *speed* 100. Thus we have compared changes occurred on network features according to each strategy *R*, *TC*, *GC* and *PA*, after an epidemic diffusion with *speed* 100 and at time t_{120} (since we consider 120 iterations on *MIDEN* algorithm). Results are shown in Figure 5, and were obtained by averaging 100 runs. Degree distribution, for each evolution strategy, is plotted for *GEN1* 5(a), *GEN2* 5(b) and *PL* 5(c) and features are depicted on 5(d).



		min degree	avg degree	max degree	cc
(a)	**R**	1.0	5.112	14.81	0.0010
	Tria. C.	1.0	5.107	18.03	0.2874
	Glo. C.	1.0	5.043	14.66	0.0002
	Pre. A.	1.0	5.198	42.36	0.0028
(b)	**R**	1.0	5.673	120.27	0.0045
	Tria. C.	1.0	5.367	146.04	0.3036
	Glo. C.	1.0	5.546	119.82	0.0020
	Pre. A.	1.0	5.540	142.65	0.0142
(c)	**R**	1.0	4.54	18.88	0.3882
	Tria. C.	1.0	3.95	23.40	0.6204
	Glo. C.	1.0	4.58	18.77	0.3763
	Pre. A.	1.0	4.25	33.31	0.4588

(d)

Fig. 5. Network properties for each evolution strategy. Degree distribution is plotted for Network (a) *GEN1*, (b) *GEN2* and (c) *PL*, after epidemics (at time t_{120}) with *speed* 100. x axis is the degree and y axis is the fraction of node with degree k. For each strategy, main network features are shown in (d), where *cc* is the clustering coefficient

In order to explain the results, we discuss first the effect of strategies on the network properties **(1)** and then consequences on spread **(2)**.

(1) Direct effects on network properties. First of all, by comparing features of original networks presented on Figure 3 and those obtained on Figure 5, it is easy to observe direct effects of evolution strategies. While it is expected that the average degree increases, as our approach only considers the addition of new links, significant differences can be observed on degree distribution and clustering coefficient resulting from each strategy.

"Random" seems to tend towards a normal degree distribution, as we can particularly observe on Figure 5(b). For all networks, a high proportion of individuals moderately connected and a lower proportion of individuals weakly

and strongly connected can be observed. Indeed Figures 5(a), (b), (c) present an higher proportion of nodes with degree between 5 and 10 with this strategy. The most significant results is obtained for *GEN2*, where a notable change from the original distribution is highlighted.

"Triadic Closure" allows neighbors of a same node to become neighbor themselves. It strengthens links within groups of nodes that result in a significant increase in the overall clustering coefficient: from 0.00056 to 0.2874 for *GEN1* and from 0.00427 to 0.3036 for *GEN2*. Unlike *TC* other strategies may even reduce *cc*. For example, it is the case of network PL: at t_0 *cc* is 0.60880, with *R*, *GC* and *PA*, it respectively takes values 0.3882, 0.3763 and 0.4588 at t_{120}.

"Global Connection" allows a node to connect with any node outside its immediate community (*friends of friends*). This explains that, except for *cc*, observed properties with *R* and *GC* strategies are very close. Obviously, *cc* is low for this kind of evolution, since as shown on *cc* values of Figure 5, it does not allow creating "*triangles*" as the *R* strategy is likely to do. For example, at t_{120} in *GEN1*, *cc* is 0.0002 for *GC*, and is 0.0010 for *R*.

"Preferential Attachment" reinforces links of most connected nodes, since nodes prefer to make connections with most popular nodes, as observed on the growth of max degree. For example, at t_0, max degree of *PL* is 17 (Figure 3), against 18.88 for *R*, 23.40 for *TC*, 18.77 for *GC* and 33.31 for *PA* at t_{120}.

(2) Consequences on spread. As expected, strategies *R* and *GC* provide very similar results on spreading, since their effects on network properties prove to be very similar, as shown on Figure 5.

For networks *GEN1* and *PL*, and *speed* 50, as seen in Section 4.2, *PA* is the only strategy able to generate an epidemic (see Figure 4), because it enables emergence of individuals sufficiently connected to allow the transmission of disease within the network. Indeed, strategies *R*, *GC* and *TD* maintain, at same speed, low connected nodes that do not allow spreading in the network.

Beginning from *speed* 100, the trend is accentuated for *PA* that shows the earliest occurrence of the epidemic peak (see Figure 4, column 100 and 150). This peak is higher than with other strategies because the *max degree* is always very high, as shown on Figure 5: 42.36 for *GEN1*, 142.65 for *GEN2* and 33.31 for *PL*. While the strategy *TC* allows the emergence of highly connected nodes, it also generates a network with a high clustering coefficient. So the epidemic is less virulent, since the transmission occurs mainly within a same community.

Network *GEN2* contains highly connected nodes at t_0 (see Fig. 3), as a scale free network. Its *max degree* is 118 against 11 for *GEN1* and 17 for *PL*. For such a network, even without evolution, epidemic spreads and strategies *R*, *GC*, *TD*, and *PA* show different impacts although this is not noticeable before *speed* 50.

An issue not yet discussed remains the case of the occurrence of the epidemic peak when the network evolves according to *TC* strategy. As shown on Figure 4, the timings of epidemic peaks obtained with *TC*, for networks *GEN2* and *PL*, are close to *PA* peak. However in network *GEN1*, this trend is not confirmed, as we can observe on Figure 4: *TC* epidemic peak appears approximately at the same time than *R* and *GC* peaks. This phenomenon is due to the connected

components contained in *GEN1* (17 for $GEN1$ as shown on Figure 3). Indeed, the *TC* strategy does not allow creating links between components. For such an evolution, an epidemic can spread only through the component of patient zero.

Results presented on this section were confirmed even when we vary values of the two parameters transmission or recover probability between 0.1 and 1.

6 DynSpread Tool

Common tools designed for simulating spread of diseases in networks do not integrate the ability to manage the evolution of networks. For this reason, we designed `DynSpread`[1], a tool for studying the "Dynamic of Spreading", with the purpose of providing an experimental environment for simulating spread of diseases in evolving networks.

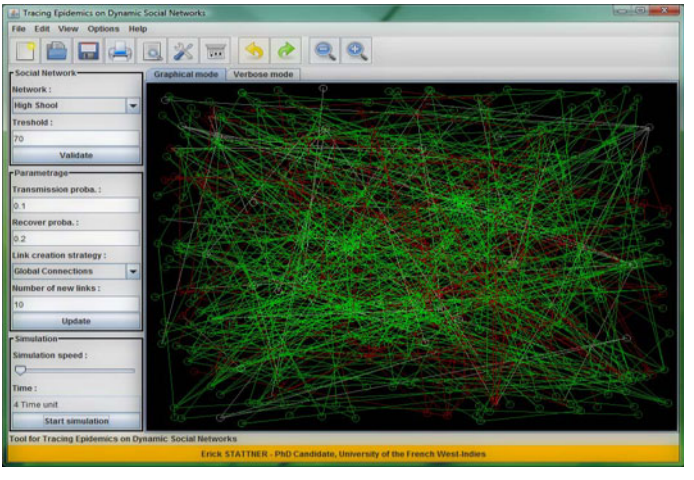


Fig. 6. Screenshot of *DynSpread* interface showing an epidemic in High School Network

DynSpread is a graphical application that shows two panels: one is used to load the network and calibrate the simulation, the other allows user to monitor the simulation either through a 2D view or a verbose mode. Green nodes correspond to susceptible individuals, red nodes to infected and white nodes to recovered. The *MIDEN* framework described in Section 3 was integrated in *DynSpread* and all tests presented in previous sections were performed with the tool. An example of the *DynSpread* interface, used for simulate the introduction of a disease in High School Network, is shown in Figure 6.

The tool is fully customizable and flexible. First, it is possible to load a network. Afterwards, user can define the probability of transmission, the probability of recover, the evolution strategy, the evolution speed and the simulation speed.

[1] DynSpread: `http://erickstattner.com/DynSpread/`

7 Conclusion and Future Works

In this work, we have addressed the question of information spreading and we have focused on the particular issue of disease spreading. Unlike works that handle static networks, we have tackled the emerging and fundamental issue of spreading in evolving networks. Our objectives were to understand how and why the evolution of networks could affect the spread. Our results provide an original contribution on three subjects:

About dynamic networks, we have provided original insight on evolution impact on network properties, by showing and comparing changes at features level, according to several network evolution strategies that have been restricted to link creation. We have shown that nodes degree and clustering coefficient can be differently modified from one strategy to another.

About epidemiology, this work gives an interesting view about the way a disease spreads through an evolving network. We have highlighted information on the impact of the network dynamics upon epidemics characteristics that should be useful for prevention campaign into communities.

About information diffusion, the *MIDEN* framework we have defined to measure impact of basic changes in network structure and the *DynSpread* tool that implements *MIDEN* to simulate diffusion in dynamic networks are flexible enough to fit similar spreading cases such as spreading of rumors, knowledge or fashion. Let us indeed refer to Borner et al.[5] who explained "If we are interested in the spreading of computer viruses, then epidemiological models can be readily applied even though the virus host is now a computer instead of a living being".

Moreover, *DynSpread* should help scientists and health professionals to have a better understanding of spread mechanisms in real world networks. Our future works in a very short term will be devoted to extending this study in order to capture full evolution strategies with link and node creation/deletion, inspired by real world networks.

References

1. Albert, R., Barabasi, A.L.: Statistical mechanics of complex networks. Reviews of Modern Physics 74, 51 (2002)
2. Barabasi, A.L.: Linked: The New Science of Networks. Perseus Books, Cambridge (2002)
3. Barrett, C.L., Bisset, K.R., Eubank, S.G., Feng, X., Marathe, M.V.: Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In: ACM/IEEE Conference on Supercomputing (2008)
4. Bott, E.: Family and social network, New-York (1957)
5. Brner, K., Sanyal, S., Vespignani, A.: Network science. In: Cronin, B. (ed.) Annual Review of Information Science & Technology, vol. 41, pp. 537–607 (2007)
6. Chen, Y.-D., Tseng, C., King, C.-C., Wu, T.-S.J., Chen, H.: Incorporating geographical contacts into social network analysis for contact tracing in epidemiology: A study on taiwan SARS data. In: Zeng, D., Gotham, I.J., Komatsu, K., Lynch, C., Thurmond, M., Madigan, D., Lober, B., Kvach, J., Chen, H. (eds.) Intelligence and Security Informatics 2007. LNCS, vol. 4506, pp. 23–36. Springer, Heidelberg (2007)

7. Christakis, N.A., Fowler, J.H.: Social network sensors for early detection of contagious outbreaks. PloS one 5(9)(9) (September 2010)
8. Christensen, C., Albert, I., Grenfell, B., Albert, R.: Disease dynamics in a dynamic social network. Physica A: Statistical Mechanics and its Applications 389(13), 2663–2674 (2010)
9. Christley, R.M., Pinchbeck, G.L., Bowers, R.G., Clancy, D., French, N.P., Bennett, R., Turner, J.: Infection in social networks: Using network analysis to identify high-risk individuals. American Journal of Epidemiology 162(10), 1024–1031 (2005)
10. Corley, C.D., Mikler, A.R., Cook, D.J., Singh, K.: Dynamic intimate contact social networks and epidemic interventions. International Journal of Functional Informatics and Personalised Medicine 1(2), 171–188 (2008)
11. Croft, D.P., James, R., Krause, J.: Exploring Animals Social Networks. Princeton University Press, Princeton (2008)
12. De, P., Das, S.K.: Epidemic Models, Algorithms, and Protocols in Wireless Sensor and Ad Hoc Networks, pp. 51–75. John Wiley & Sons, Chichester (2008)
13. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of networks. Adv. Phys. (2002)
14. Gallos, L.K., Liljeros, F., Argyrakis, P., Bunde, A., Havlin, S.: Improving immunization strategies. Phys. Rev. E 75(4) (April 2007)
15. Gross, T., D'Lima, C.J., Blasius, B.: Epidemic dynamics on an adaptive network. Physical Review Letters 96(20) (2006)
16. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabsi, A.-L.: The large-scale organization of metabolic networks. Nature 407, 651–654 (2000)
17. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London 115, 700–721 (1927)
18. Klovdahl, A.S.: Social networks and the spread of infectious diseases: the aids example. Soc. Sci. Med. 21(11), 1203–1216 (1985)
19. Milgram, S.: The small world problem. Psychology Today 1, 61–67 (1967)
20. Newman, M.E.J.: The structure and function of complex networks. Siam Review 45, 167–256 (2003)
21. Read, J.M., Eames, K.T.D., Edmunds, W.J.: Dynamic social networks and the implications for the spread of infectious disease. J. R. Soc. Interface 5(26) (2008)
22. Salathe, M., Jones, J.H.: Dynamics and control of diseases in networks with community structure. PLoS Comput. Biol. 6(4), 04 (2010)
23. Salathe, M., Kazandjieva, M., Lee, J.W., Levis, P., Feldman, M.W., Jones, J.H.: A high-resolution human contact network for infectious disease transmission (2010)
24. Stattner, E., Vidot, N., Collard, M.: Social network analysis in epidemiology: Current trends and perspectives. In: 5th IEEE Internatinal RCIS (2011)
25. Toivonen, R., Kovanen, L., Kivela, M., Onnela, J.P., Saramaki, J., Kaski, K.: A comparative study of social network models: network evolution models and nodal attribute models. Social Networks 31(4), 240–254 (2009)