

Diffusive Phenomena in Dynamic Networks: a data-driven study

Letizia Milli^{1,2}, Giulio Rossetti², Dino Pedreschi¹, and Fosca Giannotti²

¹ University of Pisa, Largo Bruno Pontecorvo, 2 Pisa, Italy
`name.surname@di.unipi.it`,

² KDD Lab. ISTI-CNR, via G. Moruzzi, 1 Pisa, Italy
`name.surname@isti.cnr.it`

Abstract. Everyday, ideas, information as well as viruses spread over complex social tissues described by our interpersonal relations. So far, the network contexts upon which diffusive phenomena unfold have usually considered static, composed by a fixed set of nodes and edges. Recent studies describe social networks as rapidly changing topologies. In this work – following a data-driven approach – we compare the behaviors of classical spreading models when used to analyze a given social network whose topological dynamics are observed at different temporal-granularities. Our goal is to shed some light on the impacts that the adoption of a static topology has on spreading simulations as well as to provide an alternative formulation of two classical diffusion models.

Keywords: Diffusion processes, Information Spreading, Dynamic Networks

1 Introduction

Since the last decade we are living two lives at the same time: one offline and one online. One of the facilities the WWW has granted us is the dismantling of physical distances, thus impacting the way diffusive phenomena evolve.

In the real world, we are used discussing the spread of viruses such as *passive* contagion processes that do not require active agents to unfold. The diffusion of ideas, conversely, is an example of *active* process: each can choose to adopt/advertise a new idea or not. When we move to the online world, we can experience both *passive* and *active* diffusion. These processes occur on top of social structures that have often been considered static. However, both passive and active processes require a direct contact with a content to spread from an already *infected* person to a susceptible one. Social interactions have a limited duration so that they dynamically shape the topology of our social graph.

In this work, we tackle the problem of understanding if, and how, dynamic network topology affects the diffusion of information. Is a static social network representation enough to simulate information spreading? Must topology dynamics be taken into account to understand the real diffusive phenomena better?

2 Related Works

Two different, yet related topics need to be reviewed and discussed: information spreading and dynamic social networks analysis.

Information Spreading. When we use the word “*spreading*” we think to contagious diseases caused by biological pathogens. However, a plethora of phenomena can be linked to the concept of the epidemic: such as the spread of computer viruses [1], mobile phone virus [2], or the diffusion of knowledge in an online social network [3]. Here, we focus on the diffusion of innovations/idea. Rogers developed the diffusion of innovation theory in 1962 [4]: it aims to explain how an idea or product diffuses through a specific population or social system.

Dynamic Social Networks. With the explosion of human-generated data, the time has started representing a non-negligible entity. During the last decade, several works have provided novel interpretations of known problems, porting them from static to temporal networks: motifs mining [5], Link prediction [6], community discovery [7] are only a few examples. Indeed, [8] showed that it is mandatory to consider different granularity of temporal abstraction. Once understood the importance of ties dynamics for the overall network topology it becomes natural to study how they affect spreading phenomena.

Spreading on Dynamic Networks. Recently, the analysis of diffusive processes in dynamic networks has started to capture the attention of the research community, such as in [9] or [10] where the authors used the SI and SIR model respectively in dynamic contests. [11] and [12] are some of the few investigations of how dynamic networks affects the spread of information. Finally, in [13] a data-driven study similar to ours was performed. However, the authors were forced to synthesize network topology evolution thus making impossible to observe the impact of characteristic phenomenon events on the diffusive process.

3 Problem definition

Our analysis will be focused on answering the following questions:

- Q1:** can analyzing spreading phenomena on a static social graph lead to an overestimate of the real volume of its diffusion?
- Q2:** do the choices made to keep track of topology dynamics impact the speed of diffusive processes?
- Q3:** is it safe to assume that spreading phenomena on a dynamic network topology unfold at a constant rate? Do the variations, as the diffusion progresses, of the number of nodes/edges impact the overall diffusion process?

To address such questions, we define three different scenarios. We model a network as an undirected graph denoted as $G = (V, E)$, where V is the set of the nodes and E is a set of interactions (edges), i.e., a triplet (u, v, t) where $u, v \in V$ and $t \in \mathbb{N}$ identify the time at which an *interaction* occurs between nodes u to v . We allow the presence of multiple interactions among the same pair of nodes. In the following, we will denote with E_{t_j} the set of interactions that appears in the graph at time t_j . We can formalize the problem in the following way:

Network	Nodes	Interactions	Edges	#Observation
WEIBO	1 656 615	6 759 012	3 394 566	90 days
FB07	19 561	304 392	67 077	365 days

Table 1. Base statistics of the analyzed interaction graphs.

Definition 1 (Spreading problem). Given a network $G = (V, E)$ observed for k consecutive snapshots, a diffusion model \mathcal{D} , and a set $I_{t_0} = \{n_1, n_2, \dots, n_j\} \subseteq V$ identifying the initial infected nodes we define the result of $\mathcal{D}(G, I_{t_0})$ as the ordered sequence $\mathcal{I} = \{I_{t_1}, \dots, I_{t_k}\}$ of the nodes infected during each snapshot.

The scenarios we will analyze in our data-driven investigation are:

-*S1 - Static topology.* For each time t_i with $i = 1, \dots, k$, we applied \mathcal{D} to the full network $G = (V, E)$ using as infected node set at time t_i the result of $\mathcal{D}(G, I_{t_{i-1}})$. The set of edges will be $E = E_{t_1} \cup E_{t_2} \cup \dots \cup E_{t_k}$.

-*S2 - Snapshot Evolution.* For each time t_i with $i = 1, \dots, k$ we compute $\mathcal{D}(G_{t_i}, I_{t_{i-1}})$ where $G_{t_i} = (V, E_{t_i})$.

-*S3 - Interaction Dynamics.* For each time t_i with $i = 1, \dots, k$ we apply \mathcal{D} incrementally to the ordered stream of interaction in E_{t_i} .

In **S1** a network will be built flattening all the interactions occurred in a single one, thus describing dynamic phenomena with a static structure. In **S2** a network will be built for each snapshot and the spreading process computed on each one of them starting, incrementally, from the previous infection status. Finally, in **S3** all the interactions among nodes that occur during each snapshot will be analyzed in their temporal ordering: no network will be explicitly built.

4 Data Driven Study

To address our research questions, we used the following datasets:

WEIBO³: it is based on data from the popular Chinese micro-blog service WEIBO⁴. An interaction represents a direct message from two users. We selected the first 90 days of the year 2011.

Facebook: the FB07 network is a sample of the WOSN2009 [14] dataset and describes online interactions between Facebook users during 2007.

In Table 1 are reported the main statistics of the networks.

On such datasets, we simulated two classical compartmental models SI and SIR detailed in 4.1. For each scenario, in 4.2 we compared the diffusion trends obtained while varying network dynamic and the model's parameters; in 5 we discuss our results and underline their relations with the topology dynamic.

4.1 Diffusion models

We chose SI and SIR to describe two different information diffusion scenarios:

D1 - Continuous advertising: after having adopted an idea/innovation an

³ <http://www.wise2012.cs.ucy.ac.cy/challenge.html>

⁴ <http://weibo.com>

Algorithm 1 Interaction-based SI

Require: I_{t_0} : set of initial infected node

```
1: for each  $t_i$  in  $\{1, \dots, k\}$  do
2:    $I_{t_i} = I_{t_{i-1}}$ 
3:   for each interaction  $(u, v, t_i)$  in  $E_{t_i}$  do
4:     if  $v$  in  $I_{t_{i-1}}$  then
5:        $p = \text{rand}(0, 1)$  ▷ Random value in  $[0, 1]$ 
6:       if  $\beta > p$  then
7:         add  $u$  to  $I_{t_i}$ 
8:       end if
9:     end if
10:  end for
11:  yield  $I_{t_i}$  ▷ Return daily status
12: end for
```

agent continue to advertise it to its neighbors during each interaction;

D2 – Diminishing advertising: after having adopted an idea/innovation an agent can decide to stop advertising it to its neighbors.

Since both models have been described for complete networks and static graphs, we will describe the modifications to apply them to the S2 and S3 scenarios.

SI: this model was introduced in 1927 by Kermack [15]. During the epidemics an individual can belong to two states, *infected* (I) and *susceptible* (S); we adopt SI to simulate diffusion scenario D1. SI assumes that if a susceptible node comes into contact with an infected one, it becomes infected with probability β .

S1: Static network. For every day t_i each node $u \in V$ having at least an infected neighbor is evaluated to decide if it will become infected or not. SI sets the probability of infection for a node having n infected neighbors as $n\beta$: the more the infected neighbors a node has the higher its chance to join the I set.

S2: Snapshot-based evolution. The model applied at day t_i will use $I_{t_{i-1}}$ and $E_{t_{i-1}}$. Therefore, the node sets and the interactions of consecutive snapshot could vary. Naturally, the nodes not present during t_i not take part in the diffusion process at time t_i . The probability of infection for a node u is $n_{t_i}\beta$ with $n_{t_i} \leq n$ restricting the set of infected neighbors to the ones that are present at time t_i .

S3: Interaction-based evolution. We can imagine such scenario as word of mouth spreading phenomena in which an idea or behavior can be shared/adopted only through a direct contact. We implement streaming SI as shown in Algorithm 1. In this model an actor u involved into m interactions with infected nodes during the day t_i has a probability of infection equal to $\sum_{i=1}^m \beta$.

SIR: this model represents a variation of the previous one. Each node belongs to three states during the epidemics: the state *infected* I , *susceptible* S and *removed* R . We adopt SIR to simulate diffusion scenario D2.

S1: Static network. We applied the classical formulation of the model on the flattened static graph. In SIR the idea/innovation is adopted with a $n\beta$ probability. Moreover, during each iteration, the probability that an infected node decides to stop advertising to its neighbors – thus joining the R set – is γ .

S2-S3: Dynamic networks. To comply with the topology dynamics described by S2 and S3, we adopted the SIR model with the same rationales used for SI.

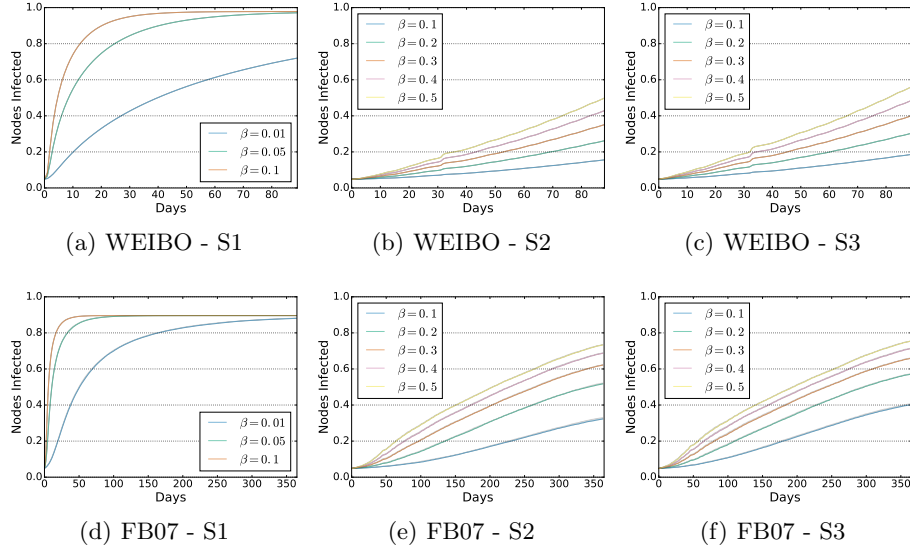


Fig. 1. Simulation of SI models on both WEIBO and FB07: the curves represent the average percentage of infected nodes over time while varying the model parameter.

We omit the pseudocode for the interaction-based version of SIR since it differs from the one reported in Algorithm 1 solely for the evaluation of the removal probability γ .

4.2 Diffusion Analysis

We organized our simulations as follow:

- i: for each dataset we randomly selected 10 sets of nodes each one covering 5% of the V : such sets identify I_{t_0} ;
- ii: for each dataset, scenario and initially infected status we executed the SI and SIR models while setting their parameters;
- iii: we build the infection trend as the iteration wise average of the runs over the 10 executions performed varying the initially infected nodes.

D1 - Continuous advertising. Figure 1 shows the results obtained by the simulation of the SI model on the two datasets.

Scenario S1. In WEIBO, Figure 1(a), setting $\beta = 0.01$ leads to an epidemic state covering almost 70% of the nodes. Increasing the values a significant speed up in the diffusion process allows reaching almost the 80% of the nodes, after only 15-20 iterations. In FB07, Figure 1(d), the impact of β is more evident: a slight increase double the number of nodes infected after the first 50 iterations.

Scenario S2. This scenario leads to a significant reduction of the diffusion speed; in both WEIBO and FB07 the infection trends do not reach saturation. Observing the FB07 trend, Figure 1(e), only for $\beta = 0.5$ we can reach a final percentage of infected nodes “comparable” to the lowest one obtained by the same model on S1. In WEIBO, Figure 1(b), the pattern is similar.

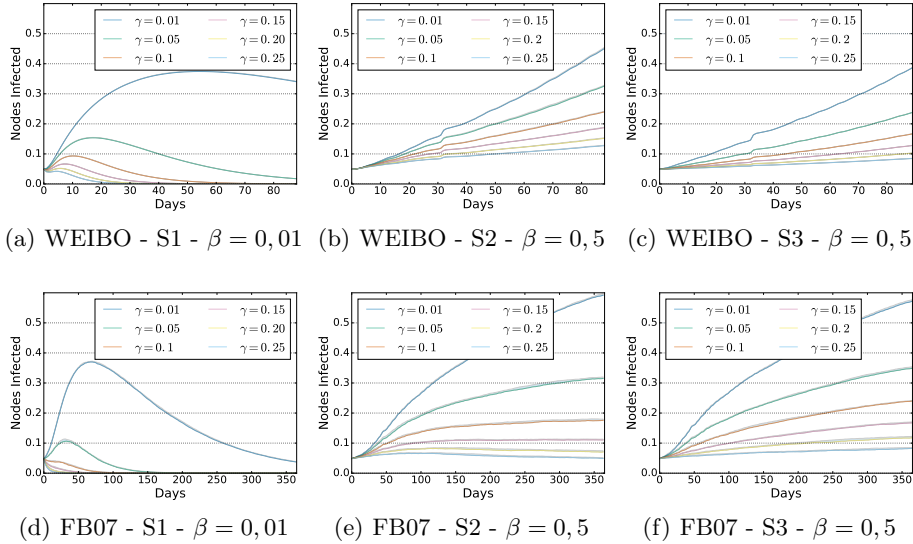


Fig. 2. Simulation of SIR models on both WEIBO and FB07: the curves represent the average percentage of infected nodes over time while varying the model parameters.

Scenario S3. We observe a behavior similar to the one identified in S2; however, in this scenario the infection trends grow always faster than the ones in S2. Such speedup is due to the different way the probability of infection is calculated: in S2 a node having n infected neighbors is subject to a $n\beta$ probability of being infected, in S3 the probability equals to $\sum_{i=1}^m \beta$ (where $m \geq n$ since during the same day multiple interactions can occur among the same pair of nodes).

D2 - Diminishing advertising. Figure 2 shows the results obtained by the simulation of the SIR model.

Scenario S1. In the simulation with SI, the diffusion reached in S1 with $\beta = 0.01$ is reachable in S2 and S3 when $\beta = 0.5$; so, we instantiate SIR fixing $\beta = 0.01$. In both datasets we observe, Figure 2(a,d), the classic decay experienced by the infection trend in a SIR model. With lower values of γ ($\gamma = 0.01$) we found a rapid growth in the first period followed that a rapidly decreased. For $\gamma \gg \beta$ the growing phase is not present since all the initial infected nodes are more likely to being removed than to spread the infection.

Scenarios S2-S3. In Figure 2(b,c,e,f) we report for S2 and S3 the infection trends for $\beta = 0.5$. Similar to what happened in S1, for values of γ comparable to the β ones the trend curves steadily die out. However, the velocity of both infection and recovery diffusions are extremely lower w.r.t. the ones in S1.

5 Discussion

Our results suggest that the particular characteristics possessed by a dynamic system deeply affect the way a word of mouth diffusion of an idea/innovation will

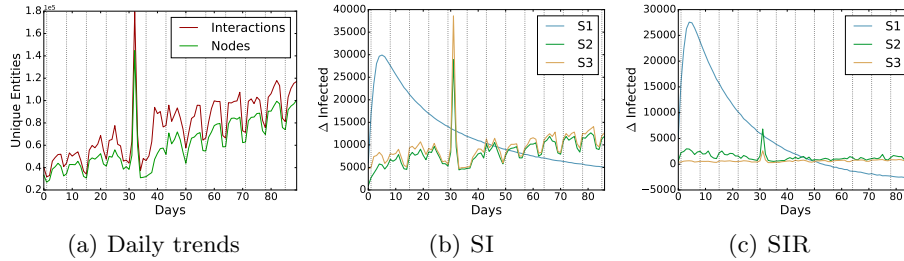


Fig. 3. (a) Daily trends in the WEIBO. Vertical lines identify Sundays. (b,c) Delta infection trend in SI (a) and SIR (b). The trends compare models having the following parameter settings - SI: S1 $\beta = 0.01$, S2-S3 $\beta = 0.5$; SIR: S1 $\beta = 0.01$ $\gamma = 0.01$, S2-S3: $\beta = 0.5$ $\gamma = 0.01$

spread. Now we concentrate our analysis on the WEIBO dataset. In Figure 3(a) are shown the patterns of daily interactions and node presences of the WEIBO interaction network. Such trends show an overall increase in the number of interactions and nodes. We identify the Sundays with vertical lines; the WEIBO users tend to diminish their presence during the weekends. We can also observe a sharp peak in the number of interactions and nodes on the 34th day: such day, 3 February 2011, identified the Chinese New Year. If we examine Figures 1(a,b,c) and 2(a,b,c), we can notice that in both S2 and S3, for all the tested parameters, a “small” jump highlight a sudden increase in the infected nodes while in S1 such behavior is not present. Therefore, adopting a flattened graph as in S1, not only we get an overestimate of the percentage of infected but also we do not capture the presence of special events. Such observations are confirmed by the *prevalence* plots shown in Figure 3 (b,c) where are reported for each day the number of novel infected nodes for SI and SIR respectively.

Once compared the diffusion trends in the three identified scenarios we can now provide answers to the research questions raised in Section 3:

- A1:** Yes, using an aggregate, static graph lead to an overestimate of the real network connectivity and, as a consequence, of all the diffusion processes.
- A2:** Yes, different temporal granularities for topology dynamics aggregation (e.g., snapshots and interactions) cause different spreading velocity.
- A3:** No, peculiar topology evolution patterns or the chosen diffusion model affect the rate of infection. In particular, cyclic patterns (weekend/weekdays) or special events (the Chinese New Year) characterize the rate at which diffusion occurs in SI, while the former loses their relevance with a SIR model.

6 Conclusions

In this work, we analyzed diffusive phenomena on dynamic social interaction graphs. We performed a data-driven study aimed to underline the real impact of network dynamics. After having modeled three different scenarios, we studied their impact on the outcome produced by classical compartmental models that

we redefined to handle topology dynamics⁵. Our results show that analyzing diffusive phenomena not considering topology dynamic lead to relevant over estimate of the real speed and not capture the presence of special events.

As future work, we plan to study the other side of the problem, namely the impact diffusive processes have on network topology.

Acknowledgments

This work is funded by the EU's H2020 Program under the funding scheme "FETPROACT-1-2014: Global Systems Science (GSS)", grant agreement # 641191 CIMPLEX and under the scheme "INFRAIA-1-2014-2015: Research Infrastructures", grant agreement # 654024 "SoBigData".

References

1. Szor, P.: Fighting computer virus attacks. USENIX (2004)
2. Havlin, S.: Phone infections. Science (2009)
3. Burt, R.S.: Social Contagion and Innovation: Cohesion Versus Structural Equivalence. AJS (1987)
4. Rogers, E.M.: Diffusion of innovations. (1962)
5. Kovanen, L., Karsai, M., Kaski, K., Kertész, J., Saramäki, J.: Temporal motifs in time-dependent networks. J. Stat. Mech. Theor. Exp. **2011**(11) (2011)
6. Tabourier, L., Libert, A.S., Lambiotte, R.: Predicting links in ego-networks using temporal information. EPJ Data Science **5**(1) (2016)
7. Rossetti, G., Pappalardo, L., Pedreschi, D., Giannotti, F.: Tiles: an online algorithm for community discovery in dynamic social networks. JMLR (2016)
8. Holme, P., Saramäki, J.: Temporal networks. Phys. Rep **519**(3) (2012)
9. Gulyás, L., Kampis, G.: Spreading processes on dynamically changing contact networks. EPJ ST **222**(6) (2013)
10. Liu, C., Zhang, Z.K.: Information spreading on dynamic social networks. Commun Nonlinear Sci Numer Simul **19**(4) (2014)
11. Miritello, G., Moro, E., Lara, R.: Dynamical strength of social ties in information spreading. Physical Review E **83**(4) (2011)
12. Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., Schifanella, R., Menczer, F., Flammini, A.: The role of information diffusion in the evolution of social networks. In: SIGKDD KDD. (2013)
13. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Régis, C., Pinton, J.F., et al.: Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. BMC medicine **9**(1) (2011) 87
14. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in facebook. In: WOSN. (2009)
15. Kermack, W.O., McKendrick, A.: A Contribution to the Mathematical Theory of Epidemics. Proc. Royal Soc. A **115**(772) (August 1927)

⁵ All methods were made available within the NDLlib library: <https://goo.gl/1tstvG>.