Information Diffusion in Complex Networks: The Active/Passive Conundrum

Letizia Milli[1,2(✉)], Giulio Rossetti[2], Dino Pedreschi[1,2], and Fosca Giannotti[2]

[1] University of Pisa, Largo B. Pontecorvo 2, Pisa, Italy
{letiziz.milli,dino.Pedreschi}@di.unipi.it
[2] KDD Lab., ISTI-CNR, via G. Moruzzi 1, Pisa, Italy
{giulio.rossetti,fosca.giannotti}@isti.cnr.it

Abstract. Ideas, information, viruses: all of them, with their mechanisms, can spread over the complex social tissues described by our interpersonal relations. Classical spreading models can agnostically from the object of which they simulate the diffusion, thus considering spreading of virus, ideas and innovations alike. Indeed, such simplification makes easier to define a standard set of tools that can be applied to heterogeneous contexts; however, it can also lead to biased, partial, simulation outcomes. In this work we discuss the concepts of *active* and *passive* diffusion: moving from analysis of a well-known *passive* model, the Threshold one, we introduce two novel approaches whose aim is to provide *active* and *mixed* schemas applicable in the context of innovations/ideas diffusion simulation. Our data-driven analysis shows how, in such context, the adoption of exclusively *passive*/*active* models leads to conflicting results, thus highlighting the need of *mixed* approaches.

1 Introduction

Information, ideas, viruses all of them have something in common: they describe different kinds of "contents" that need to be vehiculated by interacting agents to diffuse. Agents can be either individuals or animals as well as computers or other technological devices connected by a complex network describing their relations. Even if similar at a high abstraction level, diffusion process have their own characteristics that profoundly affect the way they evolve. One of such characteristics is undoubtedly tied to the degree of *activeness* of the agents they aimed to reach. Agents can be *passive* and doomed to suffer a diffusion process (e.g. during an outbreak of influenza) or *active* and voluntarily adopt a given behavior or idea just because they feel it right. Moreover, agents can also show both of such behaviors: in some circumstances a content can need both a certain degree of exposure of actors as well as their interest to be adopted. Indeed, such ambivalence is strictly tied to specific contents and contexts (e.g. individuals related in a social network). The *activeness* distinction regards prevalently phenomena of *social contagion*, where the diffusing object is either an idea or a piece of information. *Social contagions* are often modeled using a classical approach, the

C. Cherifi et al. (eds.), *Complex Networks & Their Applications VI*,
Studies in Computational Intelligence 689, https://doi.org/10.1007/978-3-319-72150-7_25

Threshold model [8] introduced by Granovetter in 1978, an approach able to capture only its *passive* component.

So far only a few empirical studies aimed to measure individual decision-making thresholds moving from a *passive* analysis (i.e. threshold describing peer pressure) to an *active* one (i.e. personal choice to adopt). Indeed, the *active-passive* dichotomy have not yet been adequately addressed nor formal models considering active users in network diffusion proposed: for this reason in this study we describe variants of the threshold model aimed to start filling such gap.

The paper is organized as follows: in Sect. 2 are introduced and discussed related works on spreading, diffusion of innovation and threshold modelling; in Sect. 3 we formalize our problem definition; in Sect. 4 are describe three scenarios of diffusion of innovations/ideas, namely *active*, *passive* and *mixed* diffusion; in Sect. 5 an *active* and a *mixed* diffusion models are introduced and the three scenarios previously described are evaluated on both synthetic and real data. Finally Sect. 6 concludes the paper.

2 Related Works

Commonly, when we use the word *"spreading"* the first images that come to our mind are related to contagious diseases caused by biological pathogens, like influenza, measles, chickenpox as well as sexually transmitted diseases. However, a plethora of phenomena can be linked to the concept of epidemic: think about the spread of computer viruses [17] where the agent is a malware or a digital virus that can transmit a copy of itself from computer to computer, or the spread of mobile phone virus [9,19], or the diffusion of knowledge, innovations, products in an on-line social network [4] – the so-called *"social contagion"*, where people are making the decision to adopt a new idea or innovation. Indeed, several elements determine the pattern by which epidemics spread through groups of people: the properties of the pathogen carries (its contagiousness, the length of its infectious period and its severity), the structure of the network and the mobility patterns of people are only a few examples. There are some analogies between the various types of spreading phenomena, the most important being the fact that all of them take place on top of complex networks whose nodes are characterized by a variable that represents the infectious state and links that represent interaction between nodes.

In this paper, we concentrate our analysis on a specific content of diffusion: innovations/ideas. The diffusion of innovation theory, developed by Rogers in 1962 [11], is one of the oldest social science theories: it aims to explain how an idea or product gains force and diffuses through a specific population or social system. The adoption of a new idea, behavior or product does not happen simultaneously in a social system; it is a process whereby some people are more suitable to adopt the innovation than others. When promoting innovation to a purpose population, it is important to understand the characteristics of the purpose population that will help or hinder adoption of the innovation. To address the diffusion of innovation problem are often adopted variants of the *Threshold Model* [8]: in such model an individual has two distinct and mutually exclusive behavioral alternatives, the decision to do or not do something, i.e., participate or not in a riot. A further requirement is that the decision depends in part on how many other people have made the same choice. Such behavior is modeled by employing individual *thresholds* to account for social pressure, e.g., a person's threshold for joining a riot can be defined as the proportion of the group he would have to see joined before he would do so. In [20], for instance, was shown that while applying such model in a network a global diffusion cascade can occur due to the interactions between nodes and individual thresholds. However, such model presents some limitations: (i) diffusion process is ignited by a single node status perturbation, while there are many situations where multiple sources of perturbation concur to the spreading (e.g. external impulses can arrive from mass media, advertising, friends), (ii) it does not consider the presence of individuals reluctant to adopt. When complex perturbations lie behind diffusion processes, we talk about *Complex contagion*, in which multiple sources of exposure to innovation are required before an individual adopts the change of behavior [2,3,5–7,12,15,20]. In such contexts, beyond the conventional threshold mechanism, recently were also investigated the effect of homophily [1,3,16] and the role of external media influence [18]. Conversely, the presence of reluctant individuals was addressed in [14] where was introduced a threshold-based model that includes blocked nodes as well as spontaneous adopters.

3 Problem Definition

In this work, we tackle a particular typology of network spreading: the diffusion of innovation. Innovation diffusion is a term often used to describe an active process: conversely from disease spreading each agent autonomously decide to adopt/publicize a given behavior/idea.

Although often treated as similar processes, diffusion of information and epidemic spreading can be easily distinguished by a single feature: the degree of *activeness* of the subjects they affect. Indeed, the spreading process of a virus does not require an *active* participation of the people that catch it (i.e., even though some behavior acts as contagion facilitators – scarce hygiene, moist and crowded environment – we can assume that no one chooses to get the flu on purpose); conversely, we can argue that the diffusion of an idea, an innovation, or a trend strictly depends not only on the social pressure but also to individual choices. Such dichotomy leads to our problem definition:

Definition 1 (Active-Passive Conundrum). Given a social context, modeled as a graph $G = (V, E)$, where a node $v \in V$ is an individual and an edge $(u, v) \in E$ identifies a social tie among $u, v \in V$, a new **information** ψ and a set of adopters $I_{t_0} \subset V$ of ψ: how can be modeled, and what characterize, *passive* and *active* diffusion processes of ψ over G?

To address passive diffusion processes are often adopted variants of the *Threshold Model* in which the adoption of a behavior by an individual is subject to a personal threshold (identified as the peer pressure exercised by his friends or, theoretically drawn by a given distribution). In the following section, we will propose and compare alternative models that take care of modeling the interest of individuals toward ψ so to understand the difference between the passive, active and mixed processes.

4 Information Diffusion Scenarios

Our aim is to compare alternative modeling choices that simulate both *passive* and/or *active* diffusion processes. We thus describe three different scenarios:

S1: Diffusion Model based on Peer Pressure. This scenario assumes that diffusion processes take place independently on the willingness of the individuals. For this model social contagion acts like virus spreading: once reached a sufficient peer pressure a novel idea will affect a target user.

S2: Diffusion Model based on Adopter Preference. This scenario assumes that the diffusion process is only apparent; each node decides to adopt or not a given behavior – once known its existence - only by its interests, completely ruling out peer pressure.

S3: Diffusion Model based on Adopter Preference and Peer Pressure. This scenario tries to relate *passive* and *active* processes so to shape information diffusion as a mix of the two. In such model, we assume interests as a preferential schema for individual tastes but relax the constraints imposed by the previous scenario by letting individually influenceable through the re-introduction of peer pressure mechanisms.

In the following section, we will detail the algorithmic choices made to provide simulations of the three scenarios as well as the results they produce on both synthetic generated networks and real world ones.

5 Passive/Active Diffusion: A Data-Driven Analysis

To compare the impacts of active and passive scenarios, we carried out a data-driven investigation modeling the social graph either with random networks and employing a graph extracted from an online social platform.

Datasets. The Erdós-Renyi random network was built with 10 k nodes and a probability for edge creation equal to 0.01. The real world network is a sample of 70 k users of the Last.fm on-line platform (network statistics are provided in [10]).

Diffusion models. In Sect. 4 we introduced three different scenarios describing different degree of *activeness* of the subjects that information diffusion affects. To understand the differences between such scenarios we model them with following diffusion models:

Algorithm 1 Node Profile

Require: I_{t_0}: infection seeds, k: iterations number, Γ: nodes' profiles, p: immunization probability

```
B = {}
for each t_i in {1,...,k} do
    I_{t_i} = I_{t_{i-1}}
    for each node v in G(V,E) do
        if v ∉ I_{t_{i-1}} and v ∉ B and {neighbors(v)} ∩ I_{t_{i-1}} ≠ ∅ then
            r = rand(0,1)                          ▷ Random value in [0,1]
            if r ≥ Γ(v) then
                add v to I_{t_i}
            else
                q = rand(0,1)                      ▷ Random value in [0,1]
                if q ≤ p then
                    add v to B
                end if
            end if
        end if
    end for
    yield I_{t_i}, B                               ▷ Return iteration status
end for
```

S1: We employ the classic *Threshold Model* to simulate a *passive adoption* process by using a theoretical distribution for the adoption threshold as done in [20]. Node's individual decision depends on the percentage of its neighbors that have already adopted. During the iteration t, every node is observed: iff the percentage of its neighbors that were infected at time $t-1$ is greater than its threshold it becomes infected as well.

S2: We design a novel *Node Profile* approach to simulate *active* adoptions. In such model, each adopter chooses to adopt the given information based only on his personal preferences. Each node carries its own "profile" describing degree by which it is likely to accept a behavior similar to the one that is currently spreading. The diffusion process starts from a set of nodes that have already adopted a given information/idea ψ. For each of the susceptible nodes in the neighborhood of a node u that has already adopted ψ, a balanced coin is flipped; if a positive result is obtained the susceptible node adopts the behavior: susceptible nodes that refuse to adopt can change their opinions during every iteration. We also implemented a variant of such model which contemplates blocked nodes, e.g., nodes that after having refused the adoption, with probability p decide to permanently stick with their choices. The pseudo-code for the introduced approach is described in Algorithm 1[1].

[1] All the methods have been implemented by the authors and are made available within the *"NDlib: Network Diffusion library"*, [13] https://github.com/GiulioRossetti/ndlib.

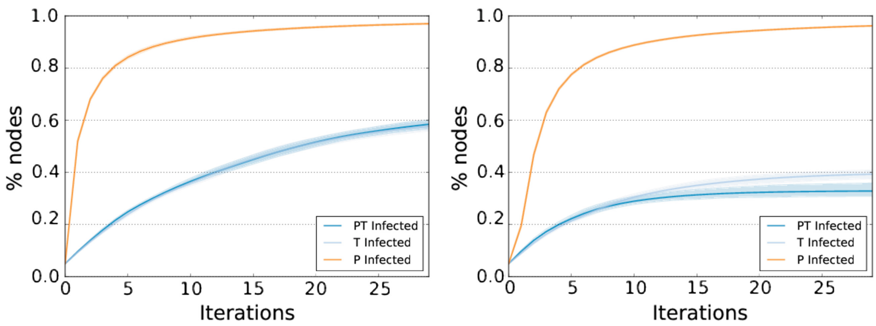

Fig. 1. Diffusion trends for *Passive* (T), *Active* (P) and *Mixed* (PT) models.

S3: To support mixed behaviors we implement a *Node Profile-Threshold* model that combines the previously described Node Profile model with theoretical peer pressure (i.e., classic threshold model). This model firstly evaluates if the peer pressure a node receives is enough to overcome its threshold, then if such constraint is satisfied, it evaluates the node profile. As for the Node Profile model, we implemented a variant that contemplates blocked nodes[2].

Experimental Settings. To compare the diffusion scenarios described, we organized our simulations as follow:

i. For each dataset we randomly selected 100 disjoint sets of nodes each one covering 5% of V: such sets identify, for each scenario and model, 100 different initial seeds of infections, e.g. I_{t_0};
ii. For each dataset, scenario and seed set we executed the active, passive and mixed diffusion models previously introduced;
iii. We compare the models by analyzing their infection trends as obtained by our simulation for each scenario, model, parameter configuration and dataset: in particular we consider for each configuration as infection trend the iteration wise average of the runs over the 100 executions performed over the identified seed sets.

In the following experiments both individual node thresholds as well as profiles were extracted from a normal distribution.

Results. In order to better characterize the obtained results we analyze separately models that contemplate blocked nodes from the ones that do not.

To the latter scenario belongs the classical Threshold model, and the simple implementations of both Profile and Profile-Threshold models. The diffusion trends of such diffusion models in both ER and Last.fm are shown in Fig. 1, where the neat lines identify the average trend overall the executions and the

[2] Pseudocode for the Profile-Threshold is not shown since it represents a combination of the two previously described approaches.

areas that surround them their interquartile ranges. As we can observe in both the synthetic graph and the real one the *active* diffusion trend shows the fastest growth: conversely, *passive* diffusion seems to be tied to a slower start and able to reach a reduced percentage of the nodes at the end of the observation period. Such results are somehow expected: the former model assumes that a susceptible node can decide to adopt when it discovers the existence of a given information (e.g. when at least a single of its neighbors has already adopted it) while the latter fix an exposure threshold below which the node does not come in contact with the information. Particular attention should be reserved to the *mixed* approach, described by the Profile-Threshold model: in the ER graph, the mixed and passive models behave alike while in the Last.fm network the Profile-Threshold trend stands below the Threshold one. The former result is due to the peculiar topology of the ER graph where the degree distribution follows a Poisson and on average a node that refuse to adopt the information is able to affect only a small fraction of the network nodes: in the latter case, conversely, the reduced diffusion speed is given by the temporarily denial of some hubs and bridge nodes to adopt the information (heavy-tail degree distribution and high clustering coefficient are topological features that also explain the slightly width increase of the observed interquartile range).

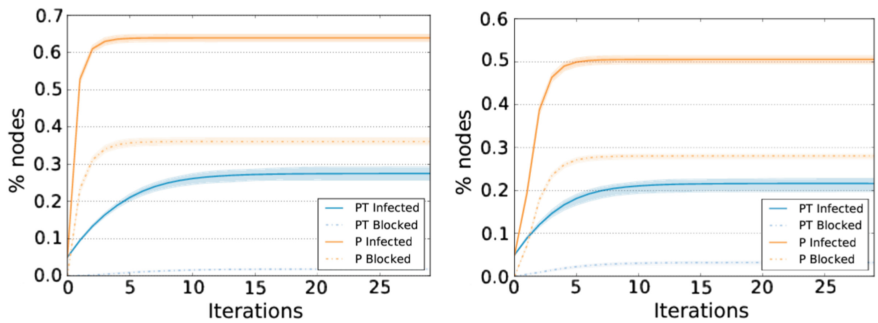

Fig. 2. Diffusion trends for *Active* (P) and *Mixed* (PT) models with blocked nodes ($p = 0.5$).

When we introduce the concept of "blocked nodes" the diffusion patterns change as shown in Fig. 2. Although we tested several values of p, the probability of a node to decide to permanently stick with his decision to not adopt, here we report only the trends for $p = 0.5$. As expected in both ER and Last.fm the infection trends of the *active* and *mixed* models experience a dumping w.r.t. what observed in the previous analysis. In particular, especially in the Last.fm context, we can observe that after the same period the percentage of infected nodes halves w.r.t. the previous simulation. Indeed, our experiments underline a linear correlation among the value chosen for p and the observed reduction of infected population.

6 Conclusion

In this work we studied a relevant difference between epidemic spreading and diffusion of information: the activeness of the involved entities. So far, both epidemic spreading and information diffusion have been studied using common modeling frameworks: among the plethora of models defined to simulate diffusive processes we focused our attention on the Threshold one. Differently from compartmental models (e.g., SI, SIR, SIS) the Threshold model once given an initial infection status produces a deterministic evolution of the diffusion: the lack of a stochastic component, along with the model rationale, make the diffusion produced by the observed model *passive*, i.e., a process during which the nodes involved do not play any active role. Such scenario is only able to capture one of the components that regulate the diffusion of an information/idea in a social context (e.g., *peer-pressure*), giving no credit to another important component: *individual preferences*. If such limitation is acceptable when dealing with disease spreading it deeply simplify the processes behind diffusion of information phenomena.

To cope with such limitation, we designed two stochastic models that reintroduce an *active* role for the nodes: namely, Profile and Profile-Threshold. In our experimentation we showed how *passive* and *active* strategies impact both the speed and overall width of the diffusion process. Moreover, we underline the need for a *mixed* approach that better simulate the real mechanics of information spreading.

As future works, we plan to extend the proposed models so to capture another relevant component: spontaneous adoptions. Finally, since social networks are constantly evolving realities where individuals, as well as interactions among them, rise and fall, we plan to reformulate our approaches to cope with dynamic topologies.

Acknowledgments. This work is funded by the European Community's H2020 Program under the funding scheme. "FETPROACT-1-2014: Global Systems Science (GSS)", grant agreement # 641191 CIMPLEX "Bringing CItizens, Models and Data together in Participatory, Interactive SociaL EXploratories"(CIMPLEX: https://www.cimplex-project.eu). This work is supported by the European Community's H2020 Program under the scheme "INFRAIA-1-2014-2015: Research Infrastructures", grant agreement #654024 *"SoBigData: Social Mining & Big Data Ecosystem"*(SoBigData: http://www.sobigdata.eu)

References

1. Aral, S., Muchnik, L., Sundararajan, A.: Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Natl. Acad. Sci. **106**(51), 21544–21549 (2009)
2. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. In: 2th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 44–54. ACM (2006)

3. Bakshy, E., Rosenn, I., Marlow, C., Adamic, L.: The role of social networks in information diffusion. In: 21st International Conference on World Wide Web, pp. 519–528. ACM (2012)
4. Burt, R.S.: Social contagion and innovation: cohesion versus structural equivalence. Am. J. Sociology (1987)
5. Centola, D.: The spread of behavior in an online social network experiment. Science **329**(5996), 1194–1197 (2010)
6. Centola, D.: An experimental study of homophily in the adoption of health behavior. Science **334**(6060), 1269–1272 (2011)
7. Gleeson, J.P., Cahalane, D.J.: Seed size strongly affects cascades on random networks. Phys. Rev. E **75**(5), 056103 (2007)
8. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociology (1978)
9. Havlin, S.: Phone infections. Science (2009)
10. Pennacchioli, D., Rossetti, G., Pappalardo, L., Pedreschi, D., Giannotti, F., Coscia, M.: The three dimensions of social prominence. In: International Conference on Social Informatics, pp. 319–332. Springer, Cham (2013)
11. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster, New York (2010)
12. Romero, D.M., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: 20th International Conference on World Wide Web, pp. 695–704. ACM (2011)
13. Rossetti, G., Milli, L., Rinzivillo, S., Sirbu, A., Pedreschi, D., Giannotti, F.: NDlib: Studying network diffusion dynamics. Accepted at DSAA, Tokyo (2017)
14. Ruan, Z., Iniguez, G., Karsai, M., Kertész, J.: Kinetics of social contagion. Phys. Rev. Lett. **115**(21), 218702 (2015)
15. Singh, P., Sreenivasan, S., Szymanski, B.K., Korniss, G.: Threshold-limited spreading in social networks with multiple initiators. arXiv:1304.7034 (2013)
16. Suri, S., Watts, D.J.: Cooperation and contagion in web-based, networked public goods experiments. PloS one **6**(3), e16836 (2011)
17. Szor, P.: Fighting computer virus attacks. USENIX (2004)
18. Toole, J.L., Cha, M., González, M.C.: Modeling the adoption of innovations in the presence of geographic and media influences. PloS one **7**(1), e29528 (2012)
19. Wang, P., González, M.C., Menezes, R., Barabási, A.L.: Understanding the spread of malicious mobile-phone programs and their damage potential. Int. J. Inf. Secur. (2013)
20. Watts, D.J.: A simple model of global cascades on random networks. Natl. Acad. Sci. **99**(9), 5766–5771 (2002)