Institute of Advanced Engineering and Science

International Journal of Information & Network Security (IJINS)
Vol.3, No.6, December 2014, pp. 336~342
ISSN: 2089-3299

Probabilistic Modeling of Information Diffusion in Online Social Networks (OSN): An Empirical Study

Andrea da Silva, Patrick Mukala
Department of Computer Science, University of Pisa

Article Info

Article history:

Received Jun 12th, 2014
Revised Aug 20th, 2014
Accepted Aug 26th, 2014

Keyword:

Model Checking with PRISM
Formal Verification
Information diffusion
Diffusion/epidemics models
Formal probabilistic model

ABSTRACT

Epidemics or diffusion models are popular as a potentially effective solution to capture information diffusion in online social networks (OSNs). Moreover, these models have been widely studied in the context of mobile ad-hoc networks, wireless sensor networks and peer-to-peer technologies for security and information diffusion purposes. In this paper, we describe a formal probabilistic model for information diffusion in online social networks. We consider specifically how new behaviours spread from user to user through an OSN. We use PRISM for the formal analysis of the diffusion process on a fixed OSN topology. We show some experiemental results pertaining to the speed and the probability of infection cascade as global properties considered in the formal analysis of the diffusion model.

Corresponding Author:

Andrea de Salve,Patrick Mukala,
Departement of Computer Science,
University of Pisa,
Largo Pontecorvo 5, 56127 Pisa PI
Email: andreadesalve@gmail.com ,patrick.mukala@gmail.com

1. INTRODUCTION

Epidemics or diffusion models/protocols have recently gained popularity as a potentially effective solution to capture information diffusion in online social networks (OSNs). Moreover, these models have been widely studied in the context of mobile ad-hoc networks, wireless sensor networks and peer-to-peer technologies for security and information diffusion purposes [4][5].

This work describes a formal probabilistic model for information diffusion in online social networks. We consider specifically how new behaviours spread from user to user through an OSN. We use PRISM for the formal analysis of the diffusion process on a fixed OSN topology [1][2]. Probabilistic model checking allows calculating the likelihood of the occurrence of certain events during the execution of the model and can be useful to establish properties of the model. The speed and the probability of infection cascade are the main local/global properties that we want to take into account in the formal analysis of the diffusion model [5].

Conventional probabilistic model checkers input a description of the system to be analysed (or model), represented as a state transition system, and a formal specification of the quantitative and qualitative properties of the system to be analysed. The description of a model may encode the probability of making a transition between states.

The model checker returns a result, indicating whether the model satisfies the specified properties, typically formulas in some temporal logic. The PRISM's property specification language includes several temporal logics such as Linear Temporal Logic (LTL) and Probabilistic Computational Tree Logic (PCTL) [1][2][3].

The rest of the paper is structured as follows. In section 2, we discuss the OSN with related grpahs, we lay the foundation for epidemics models, given details about our model and properties and present some results and finally in section 3, we conclude this experiment.

2. DIFFUSION MODELING ON OSNs: Problem Statement and Methodology

An OSN is described by an undirected graph $G=(V,E)$, where vertices in V correspond to users and edges in E to the social relationships between them. Let $\delta(u)$ the set of all the friends of a user u, defined as $\delta(u)=\{v\in V|(u,v)\in E\}$. Moreover, each user $u\in V$ has a fixed unique id in the system. The diffusion we want to model takes into account the spreading of a new behaviour A on G. At any time, a user may or not may adopt the new behaviour A. We consider a Susceptible-Infected epidemic model (SI) where each user is infected (I) if it adopts the new behaviour A, otherwise the user is in a susceptible (S) state. A susceptible user can remain susceptible or can be infected with the new behaviour, while an infected user can not leave his behaviour and come back in a susceptible state. The SI model described above is depicted in Figure 1.

Figure 1: The SI epidemic model.

Initially, each member of the OSN is in a susceptible state (i.e. the user not adopt the behaviour A). After that, each user u chooses whether to remain susceptible or become infected based on the choices of his neighbours $\delta(u)$. Let $\sigma(u)$ the fraction of the infected u's neighbours, defined as $\sigma(u)=size(\{v\in\delta(u)|\ v\ infected\})/size(\delta(u))$ when $\delta(u)\neq 0$, 0 otherwise. The spreading rules of the model can be informally summarised as follows:

- A susceptible user whose neighbours are in a susceptible state ($\sigma(u)=0$) becomes infected with a fixed probability pInfect for some reason that depends on the real world, otherwise remains susceptible with probability (1−pInfect).
- A susceptible user who has at least one infected neighbour ($\sigma(u)>0$) becomes infected if the fraction of his infected neighbours is greater than (or equal to) a fixed threshold (i.e. $\sigma(u)\geq$threshold).
- An infected user remains infected for the entire simulation.

A more formal description of the rules described above is shown in Figure 2. The parameters that define the model are:

0≤pInfect≤1 The probability that a user, with no infected neighbour, will be infected. The same for all users.
0<threshold≤1 A homogeneous users'threshold which specifies the fraction of infected neighbours that each user must have in order to become infected.

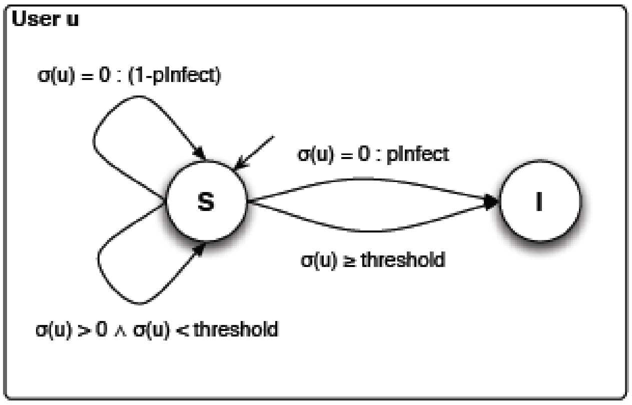

Figure 2: Events on each OSN's user.

We focus on a fixed OSN graph of 9 users arranged in a 3 by 3 grid, where only direct neighbours have a friendship relationship. Therefore, in our case, analysis of different network topologies requires different models. Figure 3 shows the OSN graph described above. Each user is identified by a progressive number indicating his position in the grid.

To simplify descriptions and help focus, ignoring details that distract from the essence of the problem we assume that the network is synchronous, reliable and without collisions. The model proceeds in rounds: at one point in time all users will choose the forward rule and apply it at the same time. All users update their state synchronously.

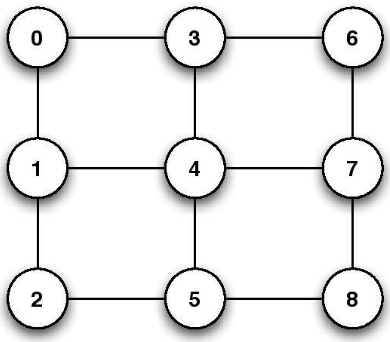

Figure 3: Online social network topology

We focus on the use of Discrete-time Markov chains (DTMC). Each user of the OSN is defined by a PRISM's *module* which carries out the rules depicted in Figure 2, based on both the local state of the current user and the local state of his neighbours. The module defines the user's state through a local variable that takes value 0 when the user's state is S and value 1 when the user's state is I.

For instance, in the network shown in Figure 3 the fraction of the infected v_1's neighbours $\sigma(v_1)$ is equal to $(state\_v_0+state\_v_4+state\_v_2)/3$. Table 1 shows the model type, the names of the modules and the name of the variables in each module.

Each rule of the modules is synchronized with a label and to each state of the model is assigned a unit cost/reward which aims to measure the amount of elapsed rounds.

Table 1: Information about the model

Type:	DTMC
Modules:	Node0 Node1 Node2 Node3 Node4 Node5 Node6 Node7 Node8
Variables:	infected0 infected1 infected2 infected3 infected4 infected5 infected6 infected7 infected8
States:	512
Transitions:	2011
Transition matrix:	1378 nodes
Reachability:	3 iterations

Table 1 shows statistics for the DTMC model we have built, where *pInfect* and *threshold* are respectively equal to 0.01 and 0.5. The tables include: the number of states and transitions in the DTMC representing the model, the number of nodes in the transition matrix of the model and the number of iterations required to find the reachable states (which is performed via a fix point algorithm).

As mentioned above above, the speed and the probability of infection cascade are the main local/global properties that we want to take into account in the formal analysis of the system. In order to test the number of infected users we define the following formula:

formula AllInfected = infected0 + infected1 + infected2 +
infected3 + infected4 + infected5 +
infected6 + infected7 + infected8;

2.1 Properties

Making use of PRISM as our model checker, we identified seven main properties to be verified as follows:

- With probability 1, eventually all users are infected

  P >=1 [true U AllInfected =9]

- The actual probability that eventually all users are infected

  P=? [F AllInfected =9]

- The actual probability that, eventually, a xed number of users (x) are infected

  P=? [F AllInfected =x]

- The actual probability that, eventually, a xed user is infected (for each user)

  P=? [F infected0 =1]

- The expected number of rounds taken to reach, from the initial state, a state where all users are infected

  R{" rounds "}=? [F AllInfected =9]

- The expected number of rounds taken to reach, from the initial state, a state where a xed number of users are infected

  R{" rounds "}=? [F AllInfected =x]

- The expected number of rounds taken to reach, from the initial state, a state where a xed user is infected (for each user)

  R{" rounds "}=? [F infected0 =1]

The results obtained here relate to verifying the properties on two specific instances of the model the first one with pInfect=0.01 and threshold=0.50 (referred to as Instance1) and the second with threshold=0.51 and pInfect=0.01 (referred to as Instance2). Instance1 satisfies the property 1 and consequently, the actual probability returned by properties 2 and 4 is 1.0.

For the Instance2 we find that the property 1 is not true. As a result, in the instance of the model there are some configurations that do not trigger a complete cascade of the new behaviour. Specifically, the actual probability that eventually, all users of the Instance2 are infected (property 2) is equal to 0.367.

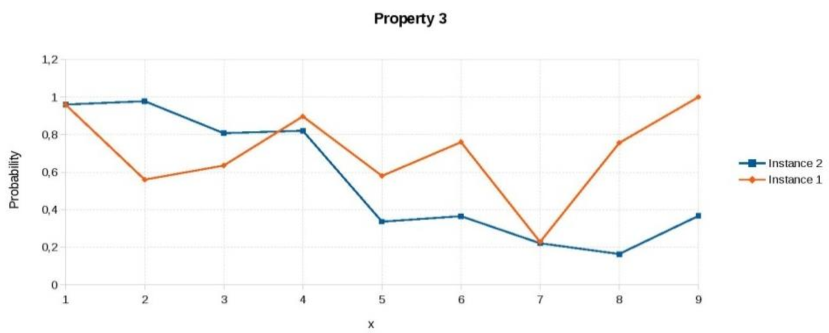

Figure 4: Probability that *x* users are infected

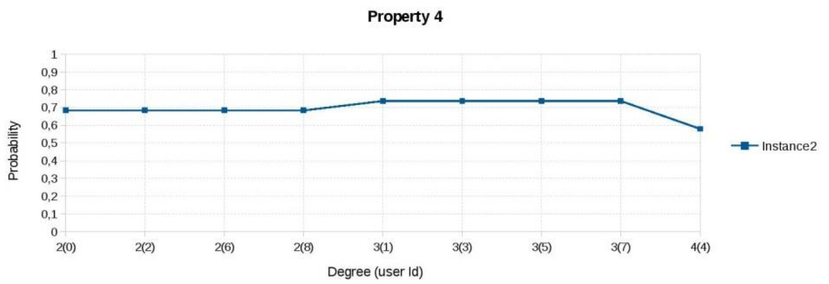

Figure 5: Probability that users are infected, related to their degree $\delta(u)$.

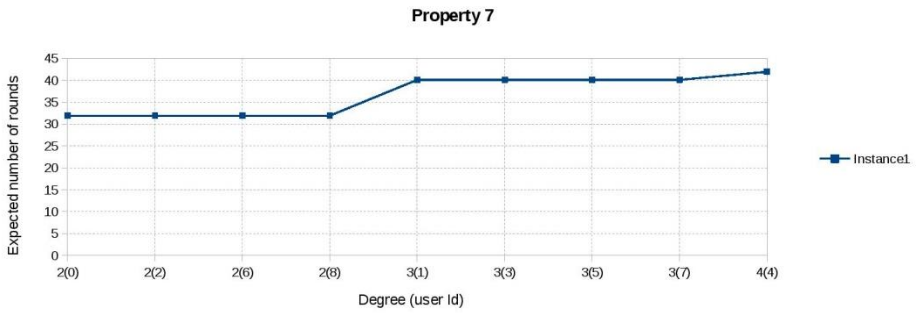

Figure 6: The expected number of rounds taken by a user *u* before his infection, related to their degree δ(*u*)

3. CONCLUSION

In this short paper we explored diffusion models in online social networks. Using Prism as a model checker, we checked our model considering the properties as defined in section 2 on two specific instances of the model the first one with pInfect=0.01 and threshold=0.50 (referred to as Instance1) and the second with threshold=0.51 and pInfect=0.01 (referred to as Instance2). Instance1 satisfies the property 1 and consequently, the actual probability returned by properties 2 and 4 is 1.0.

Chart in Figure 4 compares, for the two instances, the actual probability that a fixed number of users (x) are infected. Chart in Figure 5 shows the probability of infection of each user of the Instance2, related to his degree (property 4). As long as the probability of infecting all users of the Instance2 is less than 1, the expected number of rounds returned by properties 5, 6 and 7 is returned by property is equal to infinity (∞). For Instance1, the expected number of rounds taken to reach a state where all users are infected is 46.76 (property 5 or 6). Chart in Figure 6 shows the expected number of rounds taken by each user u to reach a state where he is infected, compared with his degree δ(u). Finally, we have seen how the expected number of rounds taken by a user, before his infection, increases with his degree.

REFERENCES

[1] Kwiatkowska, M., Norman, G., & Parker, D. (2011, January). PRISM 4.0: Verification of probabilistic real-time systems. In *Computer Aided Verification*(pp. 585-591). Springer Berlin Heidelberg.
[2] Kwiatkowska, M., Norman, G., & Parker, D. (2012). The PRISM benchmark suite. In *9th International Conference on Quantitative Evaluation of SysTems*(pp. 203-204).
[3] Hinton, A., Kwiatkowska, M., Norman, G., & Parker, D. (2006). PRISM: A tool for automatic verification of probabilistic systems. In *Tools and Algorithms for the Construction and Analysis of Systems* (pp. 441-444). Springer Berlin Heidelberg.
[4] Kempe, D., Kleinberg, J., & Tardos, É. (2005). Influential nodes in a diffusion model for social networks. In *Automata, languages and programming* (pp. 1127-1138). Springer Berlin Heidelberg
[5] Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and methods in social network analysis* (Vol. 28). Cambridge university press
[6] Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137-146). ACM

BIBLIOGRAPHY OF AUTHORS

Andrea de Salve was born in Casarano (Lecce, Italy). He holds both a bachelor and masters degrees in Computer Science from the University of Pisa. He is currently a PhD candidate in the department with a research focus on P2P systems and its application to decentralization of the OSNs, Complex network analysis and Web Mining

Patrick Mukala is currently a PhD Candidate in Computer Science at the University of Pisa in Italy. Holding a Masters degree in Software Engineering as well as a Masters degree in Information Networks from the Tshwane University of Technology in Pretoria (South Africa), his doctoral work spans from process and data mining in FLOSS repositories to complex networks and web mining.