The OSNSim Standardization Proposal for Input Parameters in Online Social Networks Simulations

Rebecca M. Alonso, Paula Luciene O. Libardi, and André Franceschi de Angelis[1]
University of Campinas
School of Technology, Limeira, SP, Brazil
Concrete Mathematics Laboratory

Abstract

The Online Social Networks (OSNs) have attracted the attention of the research community due to their current importance in the people's lives. One of the ways to investigate them is by using computational simulations, but often the selection of the input parameters of a simulator undergoes a combinatorial explosion problem. Furthermore, currently there is not a common frame to compare OSNs simulations, so each one uses his/her own set of values. Here, we propose the *OSNSim Standard* applicable to the choice of simulations input parameters, targeting to settle a reference in this field. We conducted a series of empirical testes using an OSN simulator to find initial ranges for the composition of the standard. We grouped and extended those ranges in a progressive sequence of scenarios focusing on a general solution. We ended up with 12 scenarios, each one composed by 72 input combinations, from small to large-scale simulations. The proposed standard is suitable as a general reference, avoids combinatorial explosions in the parameters selection, and provides a comparison frame for correlated researches.

Keywords: Complex Networks, Simulation, Social Network Analyses, Computational Data Analysis and Simulation in General Sciences.

1. INTRODUCTION

Nowadays, Online Social Networks (OSNs) are very important systems that permit communication among people all around the world. The biggest networks have million or even billion of users that daily exchange text messages, posts, files, and videos. The main players in this market keep a huge amount of data about people and business, and they need to manage the dynamic operation and the high complexity of their systems. The scale

[1]E-mail Corresponding Author: andre@ft.unicamp.br

of such systems arises some questions, because a number of tasks become impossible to be accomplished by hand or even by supervised procedures.

One of these tasks is the identification of death users whose profiles stay active in the OSNs registers. More details about this challenge are found in [1, 2, 3]. Currently, we are researching methods for the automatic unsupervised identification of the deads in OSNs as a special case of partition of network nodes. In [4], we developed a program named Demortuos [5] to simulate the operation of common OSNs and to test candidate algorithms of partition (Fig. 1). As a consequence of its operation, we faced a combinatorial explosion of the parameters that was quite difficult to manage. Although we have found particular answers to our needs, it was not enough as a broad solution, because the number of combinations of input ranges makes the comparison among different works very difficult.

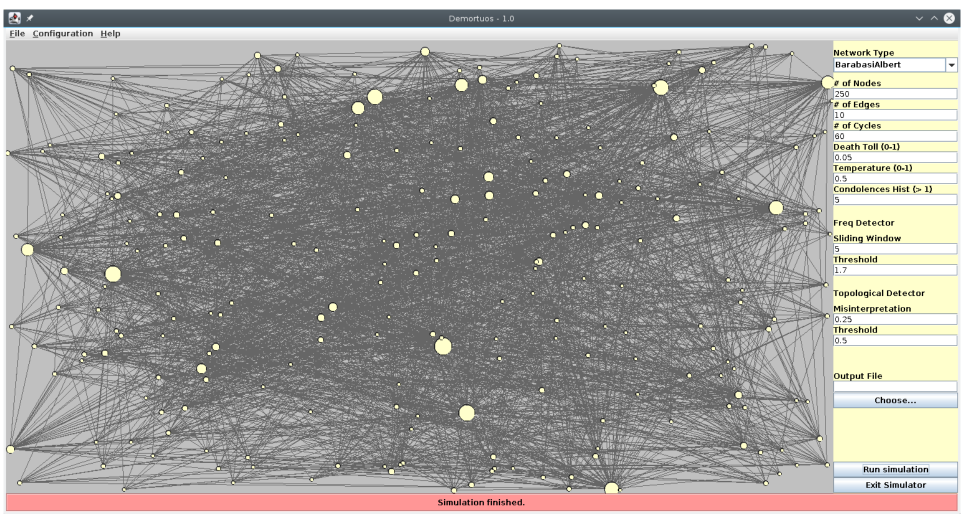

Figure 1 - A screenshot of *Dermortuos* program; a simulated network of 250 nodes and 10 edges per node is plotted in the drawing area; the parameters are inside the boxes at left.

Consequently, we dedicated our attention to define a standard set of simulation parameters to be used by any researcher to test his/her simulations according to a comparable frame. We separated the simulation parameters of the OSNs from those of the algorithms of partition, because the latter are test-dependent specialized sets, out of the scope of this work. Then, we only focused on the input values needed to execute the simulations.

The process to create the OSNSim standard started from a number of empirical tests helped by Demortuos program. In an intermediate step, we reached to a set of parameters close to our model, but those parameters could not be taken as general ones. Following the research development, that initial set was extended to deal with a broader series of configurations, covering about 800 parameters combinations.

Finally, we grouped these combinations in 12 scenarios to organize the standard proposal. This organization permits the application of the standard according to the interests of the researchers and to the available computational resources. Furthermore, it provides a stable frame for comparisons and avoids the combinatorial problem in OSNs simulations.

2. METHODS

In a nutshell, we observed the combinatorial explosion problem we had to face, executed a series of empirical tests to identify good parameters combinations, extended the initial set to encompass a broader joint of configurations, and grouped the selected parameters ranges to organize the standard. As the OSNSim proposal was developed after Demortous, it could benefit of our previous experience as the initial basis. All tests were executed in general-use personal computers with at least 4 GB RAM memory, running Linux or Windows operating systems.

The parameters ranges used in [4] were used in our first trials, since that work found acceptable sets. From them, we carried out a series of empirical tests to find the suitable parameter ranges for general simulations and posterior data analysis. We selected seven input parameters, looking for a reduced number of entries, and eliminating those variables that uniquely concerned the algorithms of partition. The parameters are the following: Network model, Order of the network; Average node degree; Simulation cycles; Mortality rate; Inactivity rate; and Temperature.

The first parameter is the model of the network. As there are lots of models, we restrict the standard proposal to two of the most used ones, the random graphs after [6], and scale-free networks, according to [7], in order to obtain a comparable frame. It means that any standard-compatible study must simulate at least one of this network models. However, simulating the both models can enlarge the comparison possibilities. It is convenient to note that a simulation of an OSN usually is based on some abstract model. Although any model could be welcomed, actual maps of networks are not properly covered by our proposed standard because they have fixed properties. Thus, it is necessary that the model be able to create network realizations from a set of values. Major reviews about network models are found in [8, 9].

The next two parameters are mandatory to most network models, even if a chosen model does not directly mention them. These numbers provide the basic recipe to build a network realization, plus an idea about the size of the simulated network, its possibilities, and its limitations. The standardized values range from five thousand to one million nodes because small sizes difficult the data analyses whereas bigger sizes are beyond the computational resources of most facilities. Average node degree ranges from 10, that is intended to support operation test of simulators, to 2.5 thousands, an upper limit about the half of today allowed friends in FaceBook.

Each of the simulation cycle represents one physical-world day in time. Since most of the interest in OSNs simulation is about the time evolution of the network, there are three standardized values in OSNSim, related to short, middle, and long-term runs. As some algorithms need a historical data series to work on smaller intervals would be impractical. We expected that simulation software be able to output intermediate results and, thus, a 360-cycle run covers all periods stated in the standard.

The rates of mortality and inactivity are of special interest to study the behavior of the OSNs in presence of the profiles of dead and inactive users, but are not mandatory for all researches. Then, simulations that are independent of these rates can ignore them. Although we initially intended to use real or near-to-real values for these two rates, we ended up with rounded values as follows. The mortality rate in the world undergoes variations according to the country and epoch. Currently annual rate in USA is 821.5 deaths per 100,000 inhabitants (0.82%) [10], whereas in Brazil it is about 0.6% [11]. The OSNSim states three values: 0.500 (50.0%), 0.050 (5.0%), and 0.005 (0.5%). The two first are intended to stress the simulation and the last one is near real rates. The use of a set of power of ten facilitates further comparisons. In our tests, rates from 0.005 to 0.008 have leaded us to indistinguishable results. The inactivity rates were chosen in a somewhat arbitrary way and ranges from 5.0% to 50.0%.

Temperature ranges from 0.0 to 1.0 and indicates the activity level of the network in an implementation-dependent fashion. For instance, Demortuos increases the number of exchanged messages among the users as the temperature increases.

We adopted a general approach to the problem by grouping some parameters ranges in scenarios that could effectively be general, even if our

Table 1: Proposed standard - Part I

Parameter	Scenario 01	Scenario 02	Scenario 03	Scenario 04	Scenario 05	Scenario 06
Model	RN; SF	RN; SF	RN; SF	RN; SF	RN; SF	RN; SF
Order	5K	5K	5K	20K	20K	50K
Average node degree	10; 100	10; 100	500	100	500	500
Cycles	30; 90; 360	30; 90; 360	30; 90; 360	30; 90; 360	30; 90; 360	30; 90; 360
Mortality rate	0.050; 0.500	0.005; 0.050	0.005; 0.050	0.005; 0.050	0.005; 0.050	0.005; 0.050
Inactivity rate	0.50	0.05	0.05; 0.20	0.05; 0.20	0.05; 0.20	0.05; 0.20
Temperature	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0
Combinations	72	72	72	72	72	72

Table 2: Proposed standard - Part II

Parameter	Scenario 07	Scenario 08	Scenario 09	Scenario 10	Scenario 11	Scenario 12
Model	RN; SF	RN; SF	RN; SF	RN; SF	RN; SF	RN; SF
Order	100K	100K	1M	1M	1M	1M
Average node degree	100	500	100	500	1k	2.5K
Cycles	30; 90; 360	30; 90; 360	30; 90; 360	30; 90; 360	30; 90; 360	30; 90; 360
Mortality rate	0.005; 0.050	0.005; 0.050	0.005; 0.050	0.005; 0.050	0.005; 0.050	0.005; 0.050
Inactivity rate	0.05; 0.20	0.05; 0.20	0.05; 0.20	0.05; 0.20	0.05; 0.20	0.05; 0.20
Temperature	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0	0.0; 0.5; 1.0
Combinations	72	72	72	72	72	72

computational environment or software tools were not able to cope with them. The OSNSim proposal is presented in the next section.

3. THE STANDARD

Here we present the OSNSim standardization proposal to OSNs simulations. As previous stated, there are scenarios that group some parameters ranges. In order to follow this standard, a set of scenarios of interest must be elected to be simulated. For each scenario, all the parameters range must be covered by the simulation to reach a comparable set of results. The scenarios and its parameters ranges are presented in the Tables 1 and 2.

Scenario 1 is intended only to help in the program tests, since it specifies small networks and unreal rates of mortality and inactivity. More realistic parameters values, planned to be employed in actual simulations are from Scenario 2 on. Scenarios 2 to 6 are feasible in general-use computers, but the following ones need more powerful devices or more sophisticated implementations to deal with memory exhaustion. Next we briefly discuss the meaning of each parameter.

- **Network model**: the abstract model that generates the network realizations, i.e., that is used to define the topology of a set of simulated networks. Here, RN refers to Random Graphs [6] and SF stands for Scale-Free Networks according to [7]. It is possible to simulate only one of the models, according to the researcher's needs.

- **Order**: the number of *nodes* in the network. Most of the static network models generate realizations with an exactly order and it will be easy to adjust the simulation to the standard. Conversely, dynamic models grow the network along the time and it is necessary to know when the standardized value will be considered. Thus, the simulation of the network operation must only be started when the topological generator reaches the desired order. It is recommended that this size be kept unchanged until the end of the simulation.

- **Average node degree**: the average number of neighbors of the nodes. This is a necessary piece of information for most of the network models, even indirectly, as in the cases where the connection probability among the nodes is the construction parameter for the model. The conversion from average degree to the specific needed construction parameter of a given model must be carried out to adapt the simulation to the standard.

- **Cycles**: the number of simulated physical-world days for the network operation.

- **Inactivity rate**: the rate of users that become inactive during the simulation. This rate must be uniformly distributed along the simulation to account for that users that keep their profiles in the OSN without frequent access to their accounts. This measure is intended only to be included in those researches in which the inactive users needed to be considered in the operation simulation.

- **Mortality rate**: the annual mortality rate expected to the overall population. This measure is intended only to be included in those researches in which the dead users needed to be considered in the operation simulation.

- **Temperature**: an implementation-dependent parameter that measures the level of network activity. This measure indicates how much the people in the OSN are active in terms of message exchange probability. A "cold" network will present a lower rate of messages along the time, whereas a "hot" one will have the opposite behavior. This parameter takes values in the interval 0.0 to 1.0.

4. CONCLUSION

We faced the combinatorial explosion of input parameters in OSNs simulations. As no one is able to investigate all possible combinations of input values, even for a small set of parameters, we proposed the OSNSim standardization to provide a common frame to OSNs simulations. The standard is a set of parameters ranges that must be considered in a given OSN simulation in order to make its results comparable. Furthermore, we grouped some ranges of parameters in scenarios, in order to permit better organization and application of the standard, according to the interests and computational resources of the researchers. The OSNSim helps the treatment of the combinatorial problem and makes it possible to compare different researches about social networks by means of the common standard.

Now, we are modifying the Demortous program to automatize the simulation of scenarios from 1 to 6 as a support for our current researches. We plan to develop a new simulator with a more sophisticated and flexible architecture, able to deal with the other scenarios as well.

References

[1] McCALLIG, D. Facebook after death: an evolving policy in a social network. International Journal of Law and Information Technology. 222 2014. 33p.

[2] VISWANATH, B.; MISLOVE, A.; CHA, M.; GUMMADI, K.P. On the evolution of user interaction in Facebook. Proc. Workshop on Online Social Networks. 2009. 5p.

[3] TARUN, P. How to Hack into Facebook without being a Hacker. WWW2013 Companion. Rio de Janeiro 2013. The rise of the dead: How many ghosts are on Facebook?. 2010. http://blog.1000memories.com/22-therise-of-the-dead-how-many-ghosts-are-on-facebooK

[4] LIBARDI, P.L.O. Detecção Computacional de Falecidos em Redes Sociais Online. (Dissertação de Mestrado) Faculdade de Tecnologia. Unicamp. Jan. 2015. 75p.

[5] Revista da Propriedade Industrial 2355. 02.07.15 BR 51 2015 000664 9 DEMORTUOS André Franceschi de Angelis e Paula Luciene Oliveira Libardi

[6] ERDÖS, P.; RÉNYI, A., On random graphs. Publ. Math. Debrecen. 6 1959. 7p.

[7] BARABÁSI, A.L., ALBERT, R. Emergence of Scaling in Random Networks. Science. 1999. 4p. DOI: 10.1126/science.286.5439.509

[8] NEWMAN, M.E.J., The structure and function of complex networks. 2003. SIAM REVIEW Vol. 45, No. 2, pp. 167256.

[9] BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D.U, Complex Networks: Structure and Dynamics. Physics Reports. 2006.

[10] National Center for Health Statistics. Deaths and Mortality. http://www.cdc.gov/nchs/fastats/deaths.htm

[11] IBGE. Taxa Bruta de Mortalidade por mil habitantes Brasil 2000 a 2015. http://brasilemsintese.ibge.gov.br/populacao/taxas-brutas-de-mortalidade.html