Scale-free network topologies with clustering similar to online social networks

Imre Varga[1] and Gergely Kocsis[1]

[1] Faculty of Informatics, University of Debrecen, Kassai 26, Debrecen, H4028 Hungary
$\{varga.imre, kocsis.gergely\}$ $@inf.unideb.hu$

Abstract

In this paper we propose a novel method to model real online social networks (OSNs) where our growing scale-free networks have tunable clustering coefficient. Models which based on purely preferential attachment are not able to describe high clustering coefficient of social networks. Beside the attractive popularity our model is based on the fact that if a person knows somebody, probably knows several individuals from his/her acquaintanceship as well. The topological properties of these complex systems were studied and it was found that it is very similar to OSNs together with the cliques are relevant in our networks.

Keyword: scale-free network, clustering coefficient, social network

1 Introduction

Recent research's showed that social networks are different from other complex networks in some sense. The biggest difference is in average clustering coefficient. In social networks there is a high probability that two friends of a given individual will also be friends of each other thus the clustering coefficient is high. Opposite to non-social networks, where these triangles are rare. Most of growing scale-free network models result low clustering coefficients [1], [2]. There are some trials to create scale-free networks with tunable clustering, but in these models the desired value of clustering coefficient is not widely adjustable or it determines other properties of the networks (for example average degree). Avoiding this problem we want to create a model for OSNs in which we can set the average clustering coefficient in a broad range without significant affecting other properties of the network.

2 Model

In order to achieve our goal we generalized the well-known BA model [1] and the Holme-Kim model [3] modifying the linking method. Our growing networks start from a small network. Then we start to grow the network by adding new nodes to it step by step. When a new node joins it is attached by links to existing nodes. These vertices are chosen by two different ways. *i)* π number of nodes are chosen based on pure preferential attachment. *ii)* the new node is linked with same probability to ν number of neighbors of each vertices chosen in the previous step. The basic idea of this two-step linking is that to have a popular friend is advantageous and then one gets to know some acquaintances of the popular friend. At this point our simulation results show that we are able to adjust the average clustering coefficient $\langle C\rangle$ by π and ν. However the value of these parameters determines the average degree $\langle k\rangle$ of nodes as well. In order to model different real networks we must tune $\langle C\rangle$ and $\langle k\rangle$ independently. That is why after the growing period the system undergoes a destroying procedure a so called attack process where certain amount of independently chosen nodes are removed [2]. The strength of the attack η can be described by the ratio of removed nodes and the original node number N. So the model has four parameters: π, ν, N and η.

3 Properties of the networks

Large number of computer simulation is carried out to analyze the properties of the generated networks and to determine the effect of the model parameters. We found that the average shortest path length grows logarithmically with the number of nodes N, indicating 'small-word' property of our system. Increasing ν leads to larger diameter due to the presence of cliques. The degree distribution obeys power law with exponent $\gamma = 2.895 \pm 0.038$. So the networks remain scale-free similar to BA model despite the different linking method and the attack process. The most interesting feature of our graphs can be seen if we analyze their average clustering coefficient $\langle C\rangle$. When a network is growing $\langle C\rangle$ is decreasing tends to finite values, opposite to BA networks. This can be written in the following functional form $\langle C\rangle \propto N^{-3/4} + C_\infty$, where we are able to determine the value of C_∞ as a function of parameters by fitting the simulation results. Practically speaking our linking method makes us able to generate large scale-free networks with different discrete values of average clustering coefficient in a wide range between 0 and 0.75. The finite C_∞ indicates that in our generalized networks cliques remains dominant at any system sizes.

The attack process has only a negligible effect to $\langle C \rangle$, but applying it the average degree $\langle k \rangle$ is freely tunable. Due to the destroyer method these quantities are tunable independently, however it has a side effect the presence of clusters. Since this a scale-free network in spite of large number of clusters generally most of the nodes belongs to a so called giant component same as in case of real OSNs. In BA networks (special case of our model, $\nu = 0$) the giant component is always larger than in generalized networks ($\nu > 0$) at a given link number after the same attack. This shows that BA networks are strongly connected while our system is a weakly connected set of densely linked group of nodes. So all of our detailed numerical results indicate the importance of cliques in our model.

We have got a large Facebook sample [4] and we managed to find a parameter set where the generated network is very similar to this real system. Both the average clustering coefficient and the average degree show less than 7% difference beside similar functional form of the degree distribution and cluster size distribution. Therefore networks based on this model are good candidates of further researches. We use these networks to study information spreading on OSNs [5].

Acknowledgment

The publication was supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

References

[1] A. Barabási and R. Albert, *Emergence of scaling in random networks*, Science 286 pp. 509-512 (1999).

[2] I. Varga, A. Németh and G. Kocsis, *A novel method of generating tunable underlying network topologies for social simulation*, in Proc. 4th IEEE CogInfoComm, pp. 71-74 (2013).

[3] P. Holme and B. J. Kim, *Growing scale-free networks with tunable clustering*, Physical Review E, **65**, p. 026107 (2002).

[4] M. Gjoka, M. Kurant, C. T. Butts and A. Markopoulou, *Walking in Facebook: A Case Study of Unbiased Sampling of OSNs*, in Proc. of IEEE INFOCOM '10, pp. 1-9 (2010).

[5] G. Kocsis and I. Varga, *Investigation of spreading phenomena on social networks*, submitted to Infocommunications Journal (2014).