

# Novel model of social networks with tunable clustering coefficient

Imre Varga, Gergely Kocsis

University of Debrecen, Faculty of Informatics  
{varga.imre,kocsis.gergely}@inf.unideb.hu

## Abstract

We propose a novel method to generate scale-free networks with discretely tunable clustering coefficient in order to model real social networks. Recently several methods were introduced to generate networks with power law degree distribution. Most of them are based on preferential attachment, but in these networks the average clustering coefficient is low opposite to the real social networks. Beside the attractive popularity our model is based on the fact that if a person knows somebody, probably knows several people from his/her acquaintanceship as well. The topological properties of these networks were studied and it was found that in these networks the cliques are relevant independently from the system size.

*Keywords:* clustering coefficient, scale-free network, social network

*MSC:* 05C82, 90B15, 91D30

## 1. Introduction

Networks are in the focus of scientific research from computer science through chemistry and biology to social sciences. In the last years it was showed out that social networks have different properties than other networks. As Newman and Park [1] evinced the main difference is in the clustering coefficient. On the one hand in the network of humans friends of a person are often friends of each other as well. This implies a high clustering coefficient. On the other hand in non-social network these kind of triplets are infrequent.

Several network models were developed in the last decades however most of them cannot describe the clustering of social networks. The so called small-world networks developed by Watts and Strogatz [2] cannot give back power law degree distribution. Growing scale-free network models lead to low clustering [3, 4, 5]. Some scale-free models with tunable clustering coefficient [6, 7, 8] has only restricted

capacity. In this work we represent a novel model of social networks with tunable clustering coefficient.

## 2. The graph generation model

We generalized the linking method of the model of Holme and Kim [6] which based on preferential attachment. In the initial state we have a fully connected graph of  $N_0$  vertices so all of them have  $N_0 - 1$  edges. Then we add more and more new vertices one after the other. Each new node gets edges following two different rules in a loop. First we connect the new vertex to a popular one based on preferential attachment, so the linking probability of vertices is proportional to their degree. Then  $\nu$  number of neighbors of formerly chosen one are linked to the new vertex with the same probability. After that this two-step method is repeated  $\pi$  times. Further vertices are joining till their number  $N$  reaches the desired value. With this two-step method we try to catch the following. Individuals prefer to know popular people and after a while they will know the friends of popular ones. During the growing period each new node has  $m = \pi(1 + \nu)$  edges to other vertices. In case of  $\nu = 0$  we get back the BA-model, while in case of  $\nu = 1$  we get the Holme model. Our model has three input parameters: the number of vertices  $N$ , the number of popular neighbors of a new vertex  $\pi$  and the number of linked neighbors of each large-degree vertex  $\nu$ .

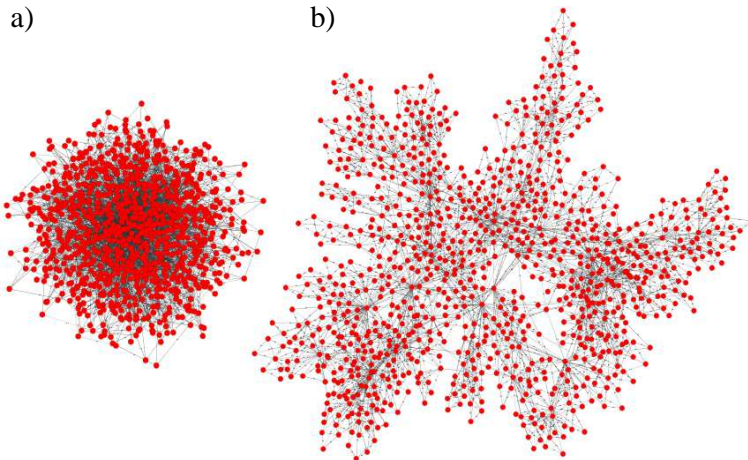


Figure 1: The graphs of our model can be very different even if the number of nodes ( $N = 1000$ ) and links ( $m = 3$ ) are the same. On the left the graph is a BA network ( $\pi = 3, \nu = 0$ ). On the right a completely different graph of our model ( $\pi = 1, \nu = 2$ ) is presented using the same representation technique.

## 2.1. Properties of generated graphs

As it can be seen in Figure 1 this small change in the generation method leads to large differences in the network properties opposite to BA-model.

The average shortest path length  $\langle L \rangle$  is an important property of graphs. In our undirected graphs  $\langle L \rangle$  is small compared to the number of nodes  $N$ . We found that  $\langle L \rangle$  increase proportionally to  $\ln N$ , so these graphs have small world property. The coefficient of this proportionality depends on the parameters  $\pi$  and  $\nu$ . BA-like networks ( $\nu = 0$ ) have smaller average shortest path length than networks with high value of  $\nu$ . This indicates that the generalized graphs contains networks of small strongly connected groups of vertices. By increasing  $\nu$  at the same value of  $m$  cliques are getting more and more important. As it is expected larger number of links leads to graphs with smaller  $\langle L \rangle$ , where the average shortest path length obeys power law decay with parameter dependent exponent. Large number of simulation shows that the average shortest path length has the following functional form

$$\langle L \rangle \propto m^{-F(\pi, \nu)} \ln N. \quad (2.1)$$

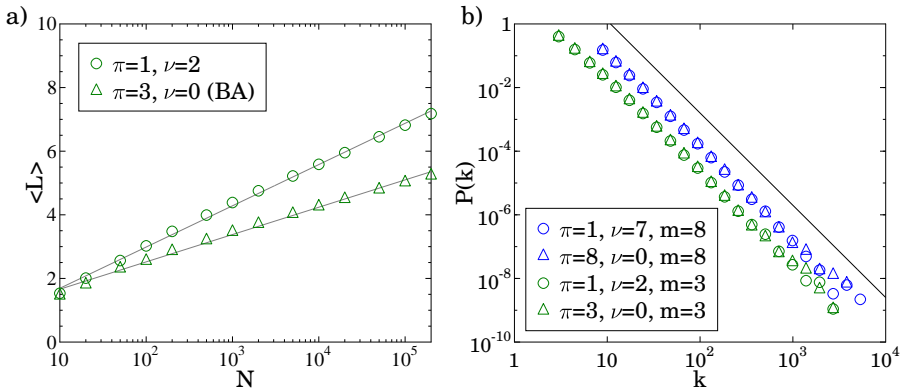


Figure 2: a) The average shortest path length  $\langle L \rangle$  as a function of system size  $N$ . In case of same density the BA-like graphs are smaller than generalized graphs. Grey lines indicate fits with Equation 2.1. b) All the graphs generated by our method have power law degree distribution. The same amount of links results the same curve independently from  $\pi$  and  $\nu$ . The exponent of the solid line is 2.9 as in BA model.

Initially vertices have the same amount of edges, but finally their degree vary in a wide range. The growing algorithm exactly determines the average degree of vertices  $\langle k \rangle = 2m = 2\pi(1 + \nu)$ . The degree distribution can be described by a power law functional form

$$P(k) \propto k^{-\gamma}, \quad (2.2)$$

so our graphs are scale-free. The exponent  $\gamma$  is independent from  $m$ ,  $\pi$  and  $\nu$ , its value is close to 2.9 as expected for BA model. (See Figure 2b.)

Study of  $\langle L \rangle$  predicts cliques. To get a better characterization of cliques in these complex systems we calculated the clustering coefficient of vertices. Local clustering coefficient  $C$  of a vertex can be expressed by the existing number of edges between neighbors of this vertex divided by  $k(k-1)/2$ , where  $k$  is the degree of the vertex. In a general case  $C$  is proportional to  $1/k$ , which indicates small degree vertices are mainly members of cliques while hubs of the networks connect them together.

The average clustering coefficient  $\langle C \rangle$  is the most important quantity of our graphs. In larger networks  $\langle C \rangle$  is smaller and smaller, this can be described by the following functional form

$$\langle C \rangle - C_\infty \propto N^{-3/4}, \quad (2.3)$$

where  $C_\infty = 0$  in case of BA networks. In other cases it means that  $\langle C \rangle$  tends to finite values, not to zero. (See Figure 3.) This indicates that when  $\nu = 0$  in a large graphs cliques are negligible, while in our generalized graphs they remain dominant at any system sizes. Large number of simulations were performed to discover how

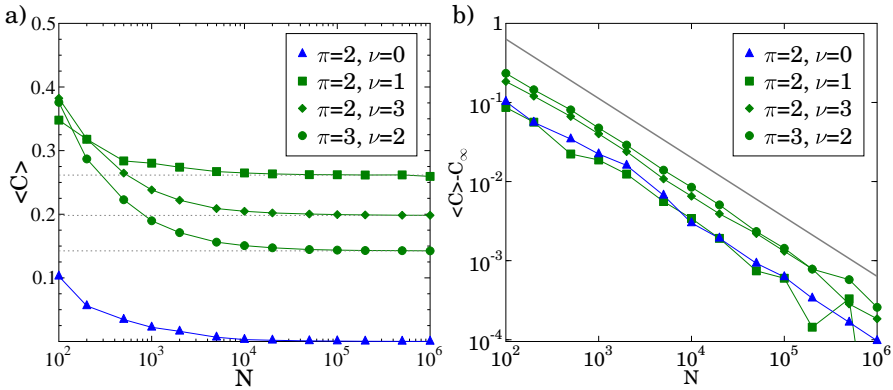


Figure 3: a) Increasing number of vertices  $N$  in the system leads to decreasing average clustering coefficient  $\langle C \rangle$ , but it tends to zero only in case of pure preferential attachment. When parameter  $\nu$  is not zero in a large network the value of  $\langle C \rangle$  is constant. Values of  $C_\infty$  which is determined by  $\pi$  and  $\nu$ , are indicated by dashed lines. b) According to Equation 2.3 on log-log scale we get straight line with exponent  $-0.75$ .

the constant value of  $C_\infty$  depends on the input parameters. We found that

$$C_\infty \propto \pi^{-A} e^{-B\nu}, \quad (2.4)$$

if  $\pi > 1$  and  $\nu > 0$ , where  $A$  and  $B$  are constants. In our graphs more links result in smaller  $C_\infty$ , where both types of linking method (so  $\pi$  and  $\nu$ ) has influence on

the value of  $C_\infty$ , but they act in different ways. 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, however smaller values are more common. If we have for example maximum 15 edges to each new vertex ( $m = 15$ ) we can create graphs with 45 different values of  $C_\infty$ .

### 3. Real online social networks

We managed to get data sets of two online social networks [9]. These networks are location-based social networks namely Brightkite and Gowalla. Its degree distribution can be characterized by a power law regime, so they are scale-free network. The quite high average clustering coefficients indicate the presence of cliques of users.

If we want to find an appropriate input parameter set to model the real network we have to take into account both  $\langle C \rangle$  and  $\langle k \rangle$  because they are not independent. This is a small disadvantage of our model. We tried to find not only the best fit of average clustering coefficient, but the possible best fit of  $\langle C \rangle$  and  $\langle k \rangle$  together.

The Gowalla sample has almost 200.000 nodes where an average vertex has  $\langle k \rangle = 19.3$  connections. The average clustering coefficient is  $\langle C \rangle = 0.1963$ . In the most similar model graph generated by values  $\pi = 2$  and  $\nu = 3$  resulted the average clustering coefficient  $\langle C \rangle = 0.1989$ , so the difference between the real and the model is just 1.3%.

The Brightkite sample contains almost 60.000 users where the average degree is  $\langle k \rangle = 14.7$ . It contains several clusters of vertices, but the largest one contains the 97% of nodes, so it is similar to our connected system. The average clustering coefficient is  $\langle C \rangle = 0.1408$  in Brightkite. The best input parameter set to create a model of Brightkite is  $\pi = 3$  and  $\nu = 2$ . In this model system  $\langle C \rangle = 0.1436$ , so the difference from the real network is just 2.0%.

### 4. Summary

In conclusion, we proposed a simple model for generating graphs with power law degree distribution and tunable average clustering coefficient. The method generalizes the Barabasi-Albert model where the edges of a new vertex play different roles. Some edges based on preferential attachment while others obey the so called 'friend of my friend is my friend' philosophy. We showed that in these graphs the cliques have very important role which cannot be described by the original BA model. That is why our model can be an excellent candidate of modeling online social networks. We managed to create graphs with properties very similar to real online social networks.

**Acknowledgements.** 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.

The computations were performed on the HPC centre provided by the Hungarian National Information Infrastructure Development Institute at the University of Debrecen.

## References

- [1] M. E. J. Newman and J. Park, *Why social networks are different from other types of networks*, Physical Review E, **68**, p. 036122 (2003)
- [2] D.J. Watts and S.H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (6684), pp. 440–442 (1998)
- [3] A. Barabasi and R. Albert, *Emergence of scaling in random networks*, Science **286** pp. 509-512 (1999)
- [4] H. Y. Lee and Hoi Yeoung Chan and Pak Ming Hui, *Scale-free networks with tunable degree distribution exponents*, Physical Review E, **69** p. 067102, (2004)
- [5] I. Varga, A. Németh and G. Kocsis *A novel method of generating tunable underlying network topologies for social simulation*, Proceedings of 4th IEEE International Conference on Cognitive Infocommunicaitons, pp. 71-74 (2013)
- [6] P. Holme and B. J. Kim *Growing scale-free networks with tunable clustering*, Physical Review E, **65**, p. 026107 (2002)
- [7] M. E. J. Newman, *Random graphs with clustering*, Physical Review Letters, **103**, p. 058701 (2009)
- [8] Lenwood S. Heath, Nidhi Parikh *Generating random graphs with tunable clustering coefficients*, Physica A **390** pp. 4577-4587 (2011)
- [9] E. Cho, S. A. Myers, J. Leskovec *Friendship and mobility: User movement in location-based social networks*, In Proc. Int. Conf. on Knowledge Discovery and Data Mining, pp. 1082-1090 (2011)