

RESEARCH

Open Access



# Friendship paradox in growth networks: analytical and empirical analysis

Sergei P. Sidorov\* , Sergei V. Mironov and Alexey A. Grigoriev

\*Correspondence:  
sidorovsp@info.sgu.ru  
Saratov State University,  
Saratov, Russia 410012

## Abstract

Many empirical studies have shown that in social, citation, collaboration, and other types of networks in real world, the degree of almost every node is less than the average degree of its neighbors. This imbalance is well known in sociology as the friendship paradox and states that your friends are more popular than you on average. If we introduce a value equal to the ratio of the average degree of the neighbors for a certain node to the degree of this node (which is called the 'friendship index', FI), then the FI value of more than 1 for most nodes indicates the presence of the friendship paradox in the network. In this paper, we study the behavior of the FI over time for networks generated by growth network models. We will focus our analysis on two models based on the use of the preferential attachment mechanism: the Barabási–Albert model and the triadic closure model. Using the mean-field approach, we obtain differential equations describing the dynamics of changes in the FI over time, and accordingly, after obtaining their solutions, we find the expected values of this index over iterations. The results show that the values of FI are decreasing over time for all nodes in both models. However, for networks constructed in accordance with the triadic closure model, this decrease occurs at a much slower rate than for the Barabási–Albert graphs. In addition, we analyze several real-world networks and show that their FI distributions follow a power law. We show that both the Barabási–Albert and the triadic closure networks exhibit the same behavior. However, for networks based on the triadic closure model, the distributions of FI are more heavy-tailed and, in this sense, are closer to the distributions for real networks.

**Keywords:** Complex networks, Social networks, Friendship paradox, Network analysis, Preferential attachment model, Triadic closure model, Network models

## Introduction

One of the non-trivial heterogeneous properties of complex networks is captured by the phenomenon of the friendship paradox which states that people, on average, have fewer friends than their friends do. The friendship paradox has been the focus of various research studies related to social networks, including (Feld 1991; Eom and Jo 2014; Alipourfard et al. 2020; Fotouhi et al. 2015; Momeni and Rabbat 2016; Bollen et al. 2017; Jackson 2019; Higham 2018).

To quantify some characteristics of the friendship paradox for real networks, the paper (Pal et al. 2019) proposes a local network metric (called the “friendship index”,

*FI*). Its value for a node is defined as the ratio of the average degree of the node neighbors to the degree of this node. The metric contains important information associated with the friendship paradox. Firstly, it shows the ‘direction of influence’ which may be interpreted as an answer to whether the node is more or less popular than its neighbors. Secondly, it measures the ‘disparity’ compared to a network without the friendship paradox: if values of *FI* are higher than 1, the paradox is present. In the paper (Pal et al. 2019) the value of *FI* is aggregated over the whole network, and this aggregate measure is examined theoretically and experimentally. Other measures and approaches for analysis of vertex popularity compared to its neighbors were developed in papers (Eom and Jo 2014; Momeni and Rabbat 2016; Fotouhi et al. 2015; Lee et al. 2019), e.g. a binary measure was used for analyzing the friendship paradox based on the comparison between a node’s degree to both the mean and median of neighbors’ degrees.

Theoretical properties of *FI* were studied in Pal et al. (2019). Friendship index quantifies the local degree asymmetry as it shows the differences in the structure of node degree distributions at a local level. For example, the value of the index will be equal to 1 in a graph in which degrees of all vertices are equal to each other (e.g. a complete, lattice or cyclic graphs). At the same time, for a star-type graph (one central and  $n$  peripheral vertices) the value of this index for the central vertex will be equal to  $\frac{1}{n}$ , while for peripheral vertices it will be equal to  $n$ . Differences in the value of this index for different vertices indicate the unevenness of the local structures of the graph in terms of node degree distributions.

This paper (Section “[Friendship index distribution in real complex networks: examples](#)”) provides empirical evidence of the friendship paradox in four undirected and two directed real networks:

- the product network based on *Customers Who Bought This Item Also Bought* feature of Amazon [the network was introduced in Yang and Leskovec (2013)];
- the network of phone calls [data was taken from Song et al. (2010)];
- the network based on 2019 snapshot of a number of Github users (Rozemberczki et al. 2019a);
- the network based on a 2017 snapshot of Facebook users [data from Rozemberczki et al. (2019b)],
- the citation network of APS journals [from Redner (2004)];
- the web graph, based on web pages and hyperlinks between them (Rossi and Ahmed 2015).

We show that all these networks exhibit the following features: the *FI* distributions are heavy-tailed and the vast majority of nodes have the friendship index value greater than 1.

It should be noted that all of these networks are growth networks. It is well-known that many phenomena exhibited by such networks can be partially explained by the use of preferential attachment mechanism. Moreover, their local clustering is higher in comparison with random or BA networks. One of the mechanisms that allows this to be explained is triadic closure.

Therefore, the main research question of this paper is the following: does the use of preferential attachment and/or triadic closure mechanisms lead to the emergence of the friendship paradox phenomenon in growth networks? This paper examines the analytical properties of the friendship index in random networks generated by two models:

- the Barabási–Albert model,
- the triadic closure model by Holme and Kim (2002).

The participation of triadic closure in the formation of real complex networks was initially described in Rapoport (1953). Triadic closure is a generating mechanism that explains community creation in a variety of complex networks (Stolov et al. 2000; Mack 2001), including knowledge, citation and research collaboration (Jin et al. 2001; Wu and Holme 2009; Juhasz and Lengyel 2017; Golosovsky and Solomon 2017; Muppidi and Reddy 2020; Ren et al. 2012; Carayol et al. 2019), affiliation (Brunson 2015), social (Fang and Tang 2015; Bianconi et al. 2014; Huang et al. 2015, 2018; Linyi and Shugang 2017; Huang et al. 2015), mobile network (Zhou et al. 2018), social interaction (Song et al. 2019; Li et al. 2013; Chen and Poquet 2020; Wharrie et al. 2019; Yin et al. 2019), financial (Souza and Aste 2019) and WWW-based networks (Louzoun et al. 2006).

One of the accountable models utilizing the triadic closure mechanism was examined by Holme and Kim (2002). It is an expansion of the Barabási–Albert model of preferential attachment (Barabási and Albert 1999) and uses the idea of triadic closure previously specified in papers (Rapoport 1953; Stolov et al. 2000; Mack 2001). This model generates networks with heavy-tailed degree distributions (as BA model does), but with a much higher clustering similar to real-world networks. In accordance with the model, a newborn node obtains a link with one of the already existing nodes chosen with the probability proportional to its degree (i.e. using *preferential attachment* mechanism). Each of the leftover  $m - 1$  edges is linked with a probability  $p$  to a randomly taken neighbor of the node that has obtained the preferentially attached link (it is called *triad formation* step), while the newborn node links with a probability  $1 - p$  to a random node of the whole network chosen with use of PA mechanism. It has been shown in Holme and Kim (2002) that the model may produce networks with various levels of clustering by selecting  $p$  and  $m$ . On the other hand, their degree distributions follow a power law with exponent  $\gamma = -3$  for any  $p$ , i.e. it is the same pattern as in the BA model.

A different version of triadic closure model was studied in Bianconi et al. (2014) (each newborn node is linked to a random vertex of the network selected with uniform probability). Extensions of the triadic closure mechanism can be found in works (Itzhack et al. 2010) and (Brot et al. 2015).

To answer the main research question of this paper, we first find the dynamics of the friendship index expected values for every single node  $v_i$  in the networks generated in accordance with both models (Sections “Dynamics of the friendship index in Barabási–Albert model” and “Dynamics of the friendship index of a node”, respectively). In addition, we study how the friendship index values are distributed in such networks. Results indicate that there is a clear presence of the friendship paradox for networks evolved in accordance with both models. The results also show that the FI values are decreasing over time for all nodes in both models. However, for networks constructed based on the

triadic closure model, this decrease occurs at a much slower rate than for the Barabási–Albert graphs. Moreover, the FI distributions for networks of both types are heavy-tailed and comparable with the empirical FI distributions for real networks. However, for networks based on the triadic closure model, their FI distributions are more heavy-tailed and, in this sense, are closer to the FI distributions for real networks.

In contrast with the paper (Pal et al. 2019) which examines the behavior of the FI value averaged over all nodes, this paper examines the dynamics of the FI for any node of the Barabási–Albert network. This paper expands the paper (Sidorov et al. 2021) in which the simplest version of BA model was examined (with one attached link to a new node at each iteration).

### Notations, definitions and methodology

A complex network at iteration  $t$  is represented by graph  $G(t)$  which can be defined as a pair  $(V(t), E(t))$  where  $V(t) = \{v_1, \dots, v_t\}$  is a set of vertices representing nodes of complex network and  $E(t)$  is a set of edges representing connections (links) between the network nodes, i.e.  $E(t) = \{(v_i, v_j) | v_i, v_j \in V(t)\}$ . If  $(v_i, v_j) \in E(t)$  then  $v_i$  and  $v_j$  are neighbors. In this paper, index  $i$  will denote the iteration at which node  $v_i$  appeared.

Let  $d_i(t)$  denote the degree of node  $v_i$  at iteration  $t$ . Then  $s_i(t)$  would denote the total degree of all neighbors of node  $v_i$  at given iteration  $t$ :

$$s_i(t) = \sum_{j: (v_i, v_j) \in E(t)} d_j(t).$$

Let  $\alpha_i(t)$  denote the average degree of all neighbors of node  $v_i$  (at iteration  $t$ ), i.e. the ratio of the sum of the degrees of the node  $v_i$  neighbors to the number of its neighboring vertices:

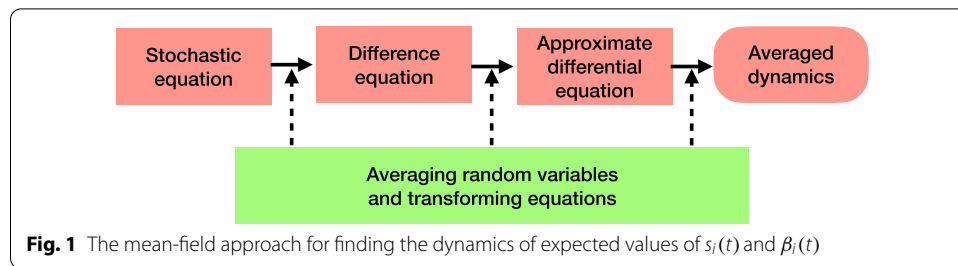
$$\alpha_i(t) = \frac{s_i(t)}{d_i(t)}.$$

The *friendship index* of node  $v_i$  (at iteration  $t$ ) is defined in Pal et al. (2019) as follows:

$$\beta_i(t) = FI_i(t) = \frac{\alpha_i(t)}{d_i(t)} = \frac{s_i(t)}{d_i^2(t)}.$$

Local characteristics of node  $v_i$ , such as the sum of the degrees of its neighbors  $s_i(t)$ , or its friendship index  $\beta_i(t)$ , may change at each subsequent iteration, depending on whether the newborn node chooses this node or its neighbors. The trajectories of these local characteristics over time are described by stochastic Markov nonstationary processes. These processes are Markov processes, since their behavior depends only on the network state at current iteration, and does not depend on its states at previous times. On the other hand, the processes under consideration are non-stationary due to the fact that their characteristics change with the growth of the network scale. We are interested in finding the expected values of these processes at any given time.

We will use standard scheme, called the mean-field method, summarized in Fig. 1. First, we should obtain a stochastic equation describing the dynamics of the studied quantity in the interval between two adjacent iterations  $t$  and  $t + 1$ . This stochastic



**Table 1** Networks and their characteristics of friendship index distribution,  $\beta$  is the friendship index,  $\gamma$  is the power law exponent

Network	$ V $	$ E $	$\beta > 1, \%$	$\gamma$	Directed
Amazon products	335,000	950,000	80.5	2.33	–
Phone calls	37,000	92,000	78.6	2.7	–
Github users	37,700	289,000	99.0	1.11	–
Facebook artists	50,500	819,000	93.4	1.6	–
Literature citations	449,700	4,689,500	84.3	1.34	+
Web hyperlinks	875,000	5,105,000	75.4	1.28	+

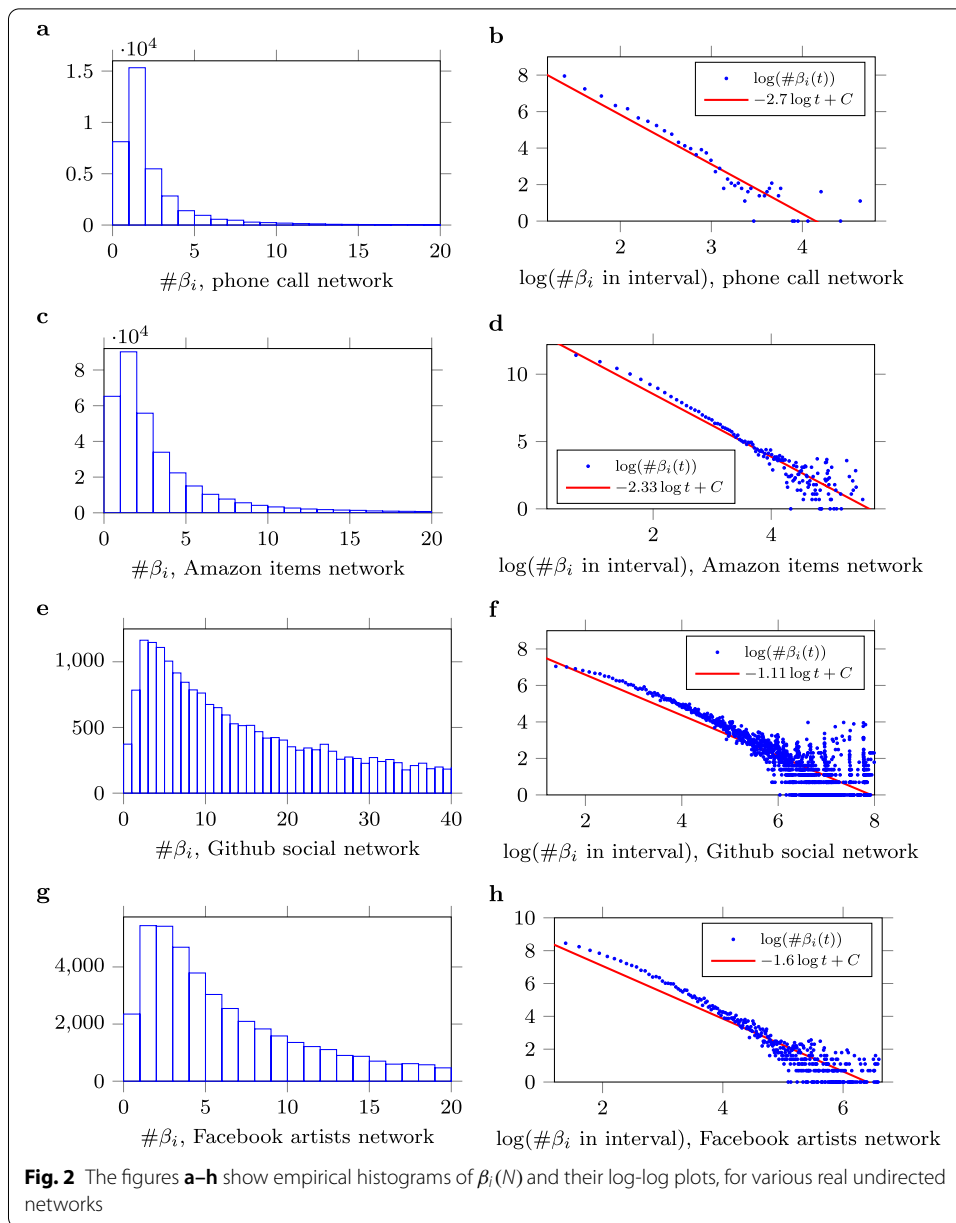
equation should then be transformed into a difference equation that describes the change in the quantity under study, taken at times  $t$  and  $t + 1$ . The next step is to replace the difference equation with its approximate analogue—the corresponding differential equation, the solution of which describes the dynamics of the quantity under investigation. In the course of this reasoning, we calculate (if possible) the averaged (expected) values of the random variables which are present in these relations.

**Friendship index distribution in real complex networks: examples**

Let us briefly demonstrate the presence of friendship paradox in real networks. Six real networks with varying characteristics are chosen for the analysis. These networks represent Github users (Rozemberczki et al. 2019a), phone calls (Song et al. 2010), Amazon products (Yang and Leskovec 2013), Facebook users (Rozemberczki et al. 2019b), citations (Redner 2004) and web hyperlinks (Rossi and Ahmed 2015). Network statistics are shown in Table 1.

We build histograms to observe the friendship index distribution. For each network we use linear and log-log plots, the latter is required as friendship index in real networks follows the power law, which can be observed on the log-log plots. From the obtained results we calculate what percentage of  $\beta$  is greater than 1, which is directly related to the detection of friendship paradox. The power-law exponents  $\gamma$  were obtained fitting linear regression to logarithmically binned data points taken from the decreasing parts of histograms.

Figure 2a, b show the friendship index distribution of  $\beta$  for real network of phone calls from Song et al. (2010). Nodes are represented by a sample of cell phone users, which are connected if they called each other during the observed period. The network is of size  $|V| = 37,000$  and  $|E| = 92,000$ . The amount of nodes with  $\beta > 1$  is 78.6%. The histogram for the real network and its log-log variant are shown in Fig. 2a,



**b**, respectively. It should be noted that the empirical distribution for the real network is asymmetric, heavy-tailed, and has a large variance ( $\mathbb{E}(\beta_i) = 2.3$ ,  $\text{VAR}(\beta_i) = 9.4$ ). The results also indicate that the overwhelming majority of nodes in the network has the friendship index greater than 1, which means that there is a remarkable friendship paradox in it.

Figure **2c, d** show the distribution of  $\beta$  for real product network based on *Customers Who Bought This Item Also Bought* feature of Amazon. The network was introduced in Yang and Leskovec (2013), 2012. Nodes are products that are connected by undirected edges if they are frequently co-purchased.

Figure 2e, f show the distribution of friendship index for network based on 2019 snapshot of a number of Github users. Individuals represented in nodes are connected if there are mutual follower relationships between them.

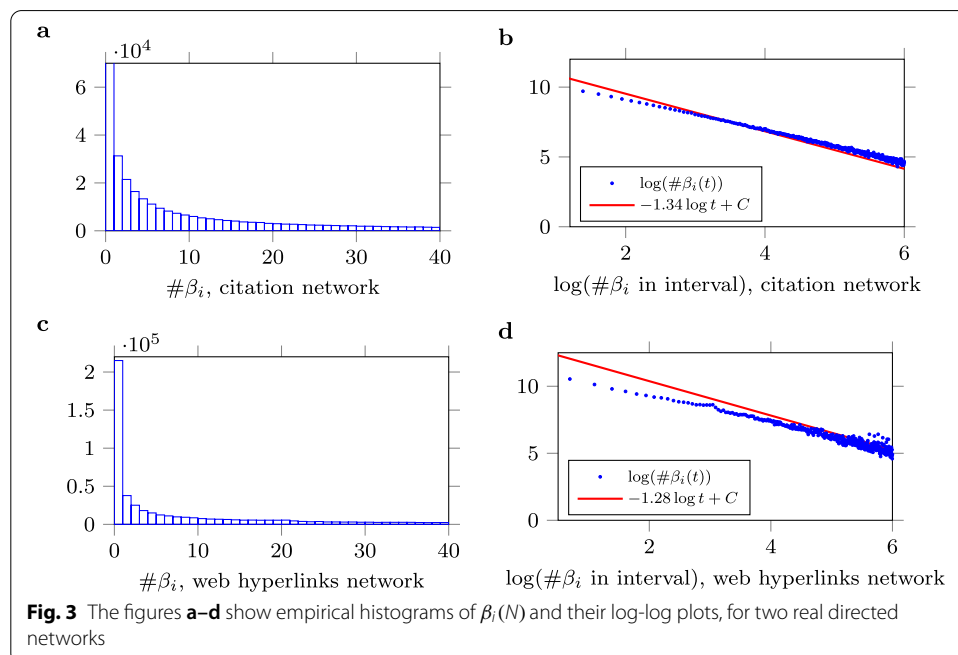
Figure 2g, h show empirical histograms of  $\beta_i(N)$  and its log-log plot, respectively, for real network based on a 2017 snapshot of users of Facebook (Rozemberczki et al. 2019b). Nodes are blue verified Facebook pages which are connected if there are mutual likes among them.

Similar results are obtained for directed networks. The main difference for them is that friendship index calculation is based on in-degree of vertices. We plot  $\beta$  distributions for directed networks on linear and log-log plots as seen in Fig. 3. Unlike undirected networks, the large number of vertices are with  $\beta$  in the range of 0 to 1. However, due to the fact that the distributions are heavy-tailed, the percentages of nodes with  $\beta > 1$  remain high enough to undoubtedly consider that the networks are under the effect of friendship paradox. Network characteristics and results are shown in Table 1.

Figure 3a, b show the distribution of  $\beta$  in a citation network of the APS journals. Nodes are papers, and node  $v_i$  is connected with a directed link with  $v_j$  if paper  $v_i$  cites paper  $v_j$ .

Figure 3c, d show the empirical histograms of  $\beta$  in a sparse directed network of web pages. Node  $v_i$  is connected with a directed link to  $v_j$  if page  $v_i$  has a hyperlink to page  $v_j$ .

It can be seen that the majority of nodes in all networks have friendship index value greater than 1, which implies the existence of friendship paradox. Furthermore, all networks have a distinguishable slope in their distributions of  $\beta$ , however the length of the tail may differ. We may see that all  $\beta$  distributions follow the power law.



## Dynamics of the friendship index in the Barabási–Albert growth networks

### Preliminary analysis

In the Barabási–Albert model (Barabási and Albert 1999) network evolves according to the following rules. Let  $m$  be the number of attached links for each newborn node, then at each iteration  $t$ :

1. Single node  $v_t$  is added;
2. One link is added to the graph, and the new node is connected by this link to one of existing nodes  $v_i$  with a probability proportional to the degree of this node  $d_i(t)$ . This process is repeated  $m$  times over the iteration.

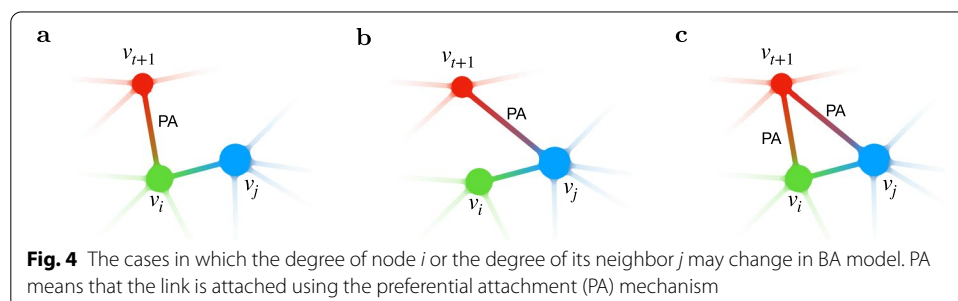
As a result of each of the  $m$  attachments, the value  $s_i(t)$  may increase in three cases only:

- the new node  $v_{t+1}$  (of degree  $m$ ) connects with node  $v_i$  on iteration  $t + 1$ , and thus the value of  $s_i(t)$  is increased by  $m$ , while  $d_i(t)$  is increased by 1:  $d_i(t + 1) = d_i(t) + 1$  (Fig. 4a);
- the new node  $v_{t+1}$  connects with one of the neighbors of node  $v_i$  by one of its  $m$  edges, and then the value of  $s_i(t)$  is increased by 1, while  $d_i(t)$  remains the same:  $d_i(t + 1) = d_i(t)$  (Fig. 4b);
- the new node  $v_{t+1}$  is connected by one of its  $m$  edges with a neighbor of node  $v_i$ , while one of the other edges links to the node  $v_i$ . Then the total degree of neighbors of node  $v_i$  is increased by  $m + 1$  and the degree of node  $v_i$  is increased by 1 (Fig. 4c).

It should be noted that the probability of the third case is an order of magnitude closer to zero compared to the probabilities of the first two cases, so we exclude this case from the analysis so as not to clutter the derivation of the equations.

In order to obtain stochastic relations describing the dynamics of these processes in time, we will introduce auxiliary indicator variables that characterize random events associated with whether node  $v_i$  will be selected by a new node (variable  $\xi_{i,l}^{(t+1)}$ ), or whether a neighbor of  $v_i$  will be selected (variable  $\eta_{i,l}^{(t+1)}$ ):

- Let the random variable  $\xi_{i,l}^{(t+1)} = 1$  if node  $v_i$  is chosen by node  $v_{t+1}$  to be connected at iteration  $t + 1$  with one of its links  $l \in \{1, \dots, m\}$ , and  $\xi_{i,l}^{(t+1)} = 0$ , otherwise.
- Let the random variable  $\eta_{i,l}^{(t+1)} = 1$  if the new node  $v_{t+1}$  connects to one of the nodes already linked to node  $v_i$  using one of  $m$  generated links  $l \in \{1, \dots, m\}$ , and  $\eta_{i,l}^{(t+1)} = 0$ , otherwise.





For further convenience we shall write  $\xi_i^{(t+1)} = \sum_{l=1}^m \xi_{i,l}^{(t+1)}$  and  $\eta_i^{(t+1)} = \sum_{l=1}^m \eta_{i,l}^{(t+1)}$ .

For simplicity, we assume that all  $m$  links at iteration  $t + 1$  are drawn simultaneously and independently of each other. However, in this case, it is possible that node  $v_i$  will be selected two or more times. The probability that node  $v_i$  will be chosen  $k$  times during the iteration is proportional to  $\left(\frac{d_i(t)}{2mt}\right)^k$ . Although this value is not equal to zero, it is an order of magnitude less than the probability of choosing node  $v_i$  once in the series of  $m$  experiments, and therefore we can exclude these cases from the analysis. Since node  $v_i$  has  $m$  chances to be chosen at iteration  $t + 1$  with the probability proportional its degree  $d_i(t)$ , we get

$$\mathbb{E}(\xi_i^{(t+1)}) = m \frac{d_i(t)}{2mt} = \frac{d_i(t)}{2t} \tag{1}$$

and

$$\mathbb{E}(\eta_i^{(t+1)}) = m \sum_{j:(v_j, v_i) \in E(t)} \frac{d_j(t)}{2mt} = m \frac{1}{2mt} s_i(t) = \frac{s_i(t)}{2t}. \tag{2}$$

**Dynamics of the total degree of node neighbors in Barabási–Albert model**

Let us calculate how the value of  $s_i(t)$  changes after a new node is added at iteration  $t + 1$ . The analysis of Section “Preliminary analysis” gives

$$\Delta s_i(t + 1) = s_i(t + 1) - s_i(t) = m \xi_i^{(t+1)} + \eta_i^{(t+1)}. \tag{3}$$

Then from (1) and (2) we get the linear nonhomogeneous differential equation of first order (as an approximation to the difference equation (3)):

$$\frac{ds_i(t)}{dt} = m \frac{d_i(t)}{2t} + \frac{s_i(t)}{2t}, \tag{4}$$

Solving the equation (see “Appendix A”), we get that the expected value of  $s_i(t)$  is

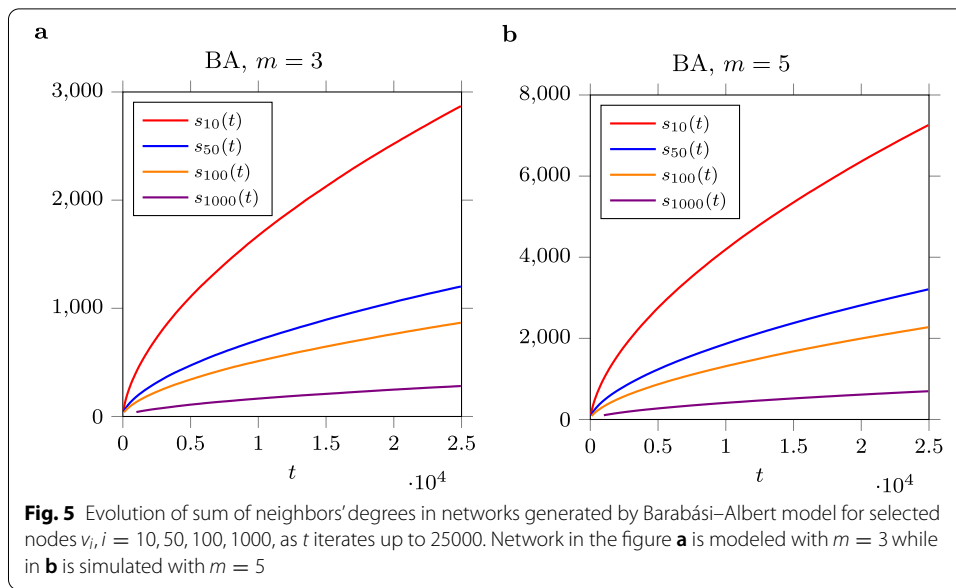
$$\bar{s}_i(t) = \mathbb{E}(s_i(t)) = \frac{m^2}{2} \left(\frac{t}{i}\right)^{\frac{1}{2}} (\log t + c(i)). \tag{5}$$

The expected initial value of  $s_i(t)$  at moment  $t = i$  is

$$\mathbb{E}(s_i(i)) = \sum_{j=1}^{i-1} P(v_i, v_j) d_j(i) = m^2 \sum_{j=1}^{i-1} \frac{\binom{i}{j}^{\frac{1}{2}}}{2i} \binom{i}{j}^{\frac{1}{2}} = \frac{m^2}{2} \sum_{j=1}^{i-1} \frac{1}{j} \sim \frac{m^2}{2} \log i, \tag{6}$$

where  $P(v_i, v_j)$  denotes the probability that at the moment of its appearance node  $v_i$  will be linked with node  $v_j$ .

Figure 5 presents the evolution of the  $s_i(t)$  averaged over 100 independent simulations. The empirical behavior of  $s_i(t)$  are indistinguishable from predictions of Eq. (5).



**Dynamics of the friendship index in Barabási–Albert model**

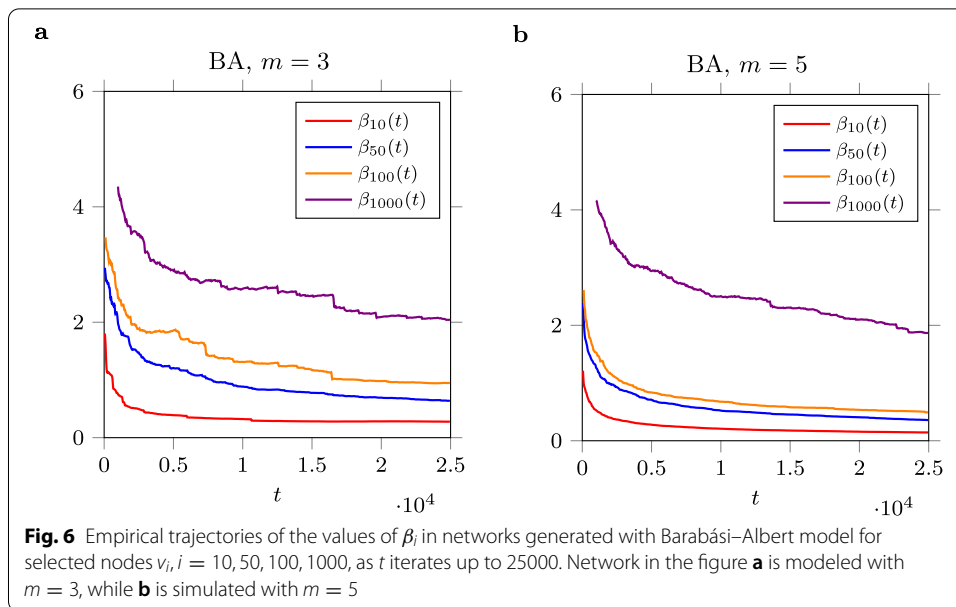
In order to measure the dynamics of  $\beta_i(t)$  we may calculate the changes in  $\beta_i(t)$  after adding a new node at iteration  $t + 1$ . Note that it would be incorrect to find  $\mathbb{E}(\beta_i(t))$  as the ratio of the expected value of the sum of degrees of neighbors of node  $v_i$  to the square of the expected number of its neighboring vertices, since

$$\bar{\beta}_i(t) := \mathbb{E}(\beta_i(t)) = \mathbb{E}\left(\frac{s_i(t)}{d_i^2(t)}\right) \neq \frac{\mathbb{E}(s_i(t))}{\mathbb{E}(d_i^2(t))}.$$

Let  $\xi_i^{(t+1)}, \eta_i^{(t+1)}$  be defined as above. Then we have

$$\begin{aligned} \Delta\beta_i(t + 1) &= \beta_i(t + 1) - \beta_i(t) \\ &= \xi_i^{(t+1)} \frac{s_i(t) + m}{(d_i(t) + 1)^2} + \eta_i^{(t+1)} \frac{s_i(t) + 1}{d_i^2(t)} \\ &\quad + \left(1 - \xi_i^{(t+1)} - \eta_i^{(t+1)}\right) \frac{s_i(t)}{d_i^2(t)} - \frac{s_i(t)}{d_i^2(t)} \\ &= \xi_i^{(t+1)} \left(\frac{s_i(t) + m}{(d_i(t) + 1)^2} - \frac{s_i(t)}{d_i^2(t)}\right) \\ &\quad + \eta_i^{(t+1)} \left(\frac{s_i(t) + 1}{d_i^2(t)} - \frac{s_i(t)}{d_i^2(t)}\right) \\ &= \xi_i^{(t+1)} \left(\frac{m}{(d_i(t) + 1)^2} - 2\frac{d_i(t)}{(d_i(t) + 1)^2}\beta_i(t) - \frac{1}{(d_i(t) + 1)^2}\beta_i(t)\right) \\ &\quad + \eta_i^{(t+1)} \frac{1}{d_i^2(t)}. \end{aligned} \tag{7}$$

Since  $\mathbb{E}(\xi_i^{(t+1)}) = \frac{d_i(t)}{2t}$  and  $\mathbb{E}(\eta_i^{(t+1)}) = \frac{s_i(t)}{2t}$  we get



$$\Delta\beta_i(t + 1) = \frac{md_i(t)}{2t(d_i(t) + 1)^2} - \beta_i(t) \left( \frac{d_i^2(t)}{t(d_i(t) + 1)^2} + \frac{d_i(t)}{2t(d_i(t) + 1)^2} - \frac{1}{2t} \right).$$

From the asymptotic  $\frac{k^2}{(k+1)^2} \sim 1 - \frac{2}{k+1}$  we can get the following first order linear nonhomogeneous differential equation (as an approximation of the difference equation (7)):

$$\frac{d\beta_i(t)}{dt} \sim \frac{md_i(t)}{2t(d_i(t) + 1)^2} - \beta_i(t) \left( \frac{1}{2t} - \frac{3d_i(t)}{2t(d_i(t) + 1)^2} \right), \tag{8}$$

Solving Eq. (8) it can be seen (“Appendix B”) that the expected value of  $\beta_i(t)$  asymptotically follows

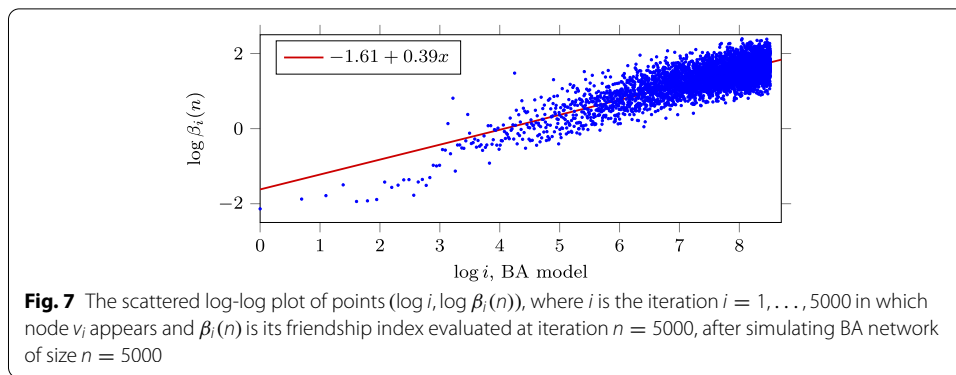
$$\bar{\beta}_i(t) := \mathbb{E}(\beta_i(t)) \sim \frac{1}{2} \left( \frac{i}{t} \right)^{\frac{1}{2}} \log t + o(t^{-\frac{1}{2}}), \tag{9}$$

where the little-o notation  $f(t) = o(g(t))$  means that  $g(t)$  grows much faster than  $f(t)$ . It follows from (6) that the initial value of  $\bar{\beta}_i$  is

$$\bar{\beta}_i(i) = \frac{\mathbb{E}(s(i))}{d_i^2(i)} = \frac{1}{m^2} \frac{m^2}{2} \sum_{j=1}^{i-1} \frac{1}{j} \sim \frac{1}{2} \log i,$$

since  $d_i^2(i) = m^2$ .

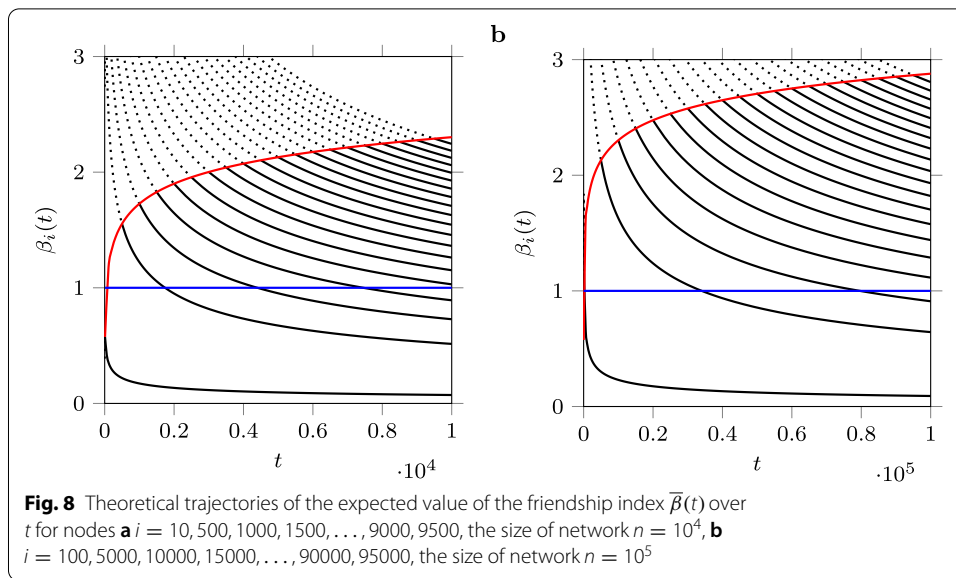
Eq. (9) describes the trajectory of the expected value of  $\beta_i(t)$  with the increase in the number of iterations  $t$ . It should be noted that the equation describes the trajectory of the *expected* value of the friendship index, while the actual trajectory in every particular simulation may differ significantly from the theoretical one. However, by averaging ones obtained from a large number of simulations, we can get a curve that lies close to the curve predicted by the equation (9). Such averaged trajectories for some nodes are shown in Fig. 6. The figure presents the dynamics of the friendship index  $\beta_i(t)$  for nodes



$v_i, i = 10, 50, 100, 1000$ , averaged over 100 simulated networks of size  $|V| = 25,000$ . It can be seen that  $\beta_i(t)$  are gradually decreasing to 0 with the growth of the network.

The analysis of Eq. (9) allows us to draw the following conclusions regarding the behavior of the friendship index over time:

- 1 The trajectory of the expected value of  $\beta_i(t)$  lies on the curve (9), and with the increase in the number of iterations  $t$ , i.e. with the growth of the network, the value of  $\bar{\beta}_i(t)$  decreases to zero with the rate  $\frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} \log t$  depending on the number of iterations  $t$ . In other words, the value of  $\beta_i(t)$  tends to zero for all nodes  $i$ , as  $t \rightarrow \infty$ .
- 2 For node  $v_i$  that appears on iteration  $i$ , the expected initial value of  $\beta_i(i)$  is  $\frac{1}{2} \log i$ .
- 3 For each node  $v_j$  that is added after node  $v_i$ , the expected value of the friendship index is higher than the expected value of the friendship index for node  $v_i$ , i.e.  $\bar{\beta}_j(t) > \bar{\beta}_i(t)$  if  $j > i$ . Thus, the more popular a node is, the lower the value of its friendship index is. The scattered plot of points  $(\log i, \log \beta_i(n)), i = 1, \dots, 5000$ , obtained for simulated BA network of size  $n = 5000$  with  $m = 4$ , is shown in Fig. 7. It presents the empirical evidence of the fact that the earlier a node appears, the lower its friendship index is, and vice versa.
- 4 Figure 8a, b show the theoretical trajectories of the friendship index for some nodes of a network built according to the Barabási–Albert (BA) model with  $m = 4$  over iterations. The starting points of each individual trajectory lie on the curve  $0.5 \log t$ . Each trajectory is shown as a solid line, while the dotted line corresponds to those portions of the same curve where the corresponding node has not yet existed. The density of the values of  $\beta_i(t)$  at moment  $t$  increases from 0 to  $0.5 \log t$ .
- 5 Having solved the equation  $\frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} \log t = 1$ , we can find the index  $i^* = i^*(t) = 4 \frac{t}{\log^2 t}$  for which the value of  $\bar{\beta}_{i^*}(t) = 1$  at iteration  $t$ . Thus, for every node  $v_j$  that appeared earlier, i.e.  $j < i^*$ , the inequality  $\bar{\beta}_j(t) < 1$  holds, while for any node  $v_l$  that appeared later, i.e.  $l > i^*$ , the relation  $\bar{\beta}_l(t) > 1$  holds. The fraction of nodes, for which the expected value of the friendship index is less than 1, is equal to  $\frac{i^*(t)}{t} = 4 \frac{1}{\log^2 t}$ . Therefore, with an increase in the number of iterations  $t$ , the fraction of such nodes decreases and tends to zero. On the other hand, the proportion of nodes, for which the expected value of the friendship index is greater than 1, tends to 1, i.e. the friendship paradox will affect almost all nodes as the network grows. If one compares



**Table 2** Characteristics of friendship index distributions in synthetic networks generated with Barabási–Albert model

$ N $	$ E $	$\beta > 1, \%$
5,000	25,000	91.591
10,000	50,000	91.961
15,000	75,000	92.287
25,000	125,000	92.516
50,000	250,000	92.672

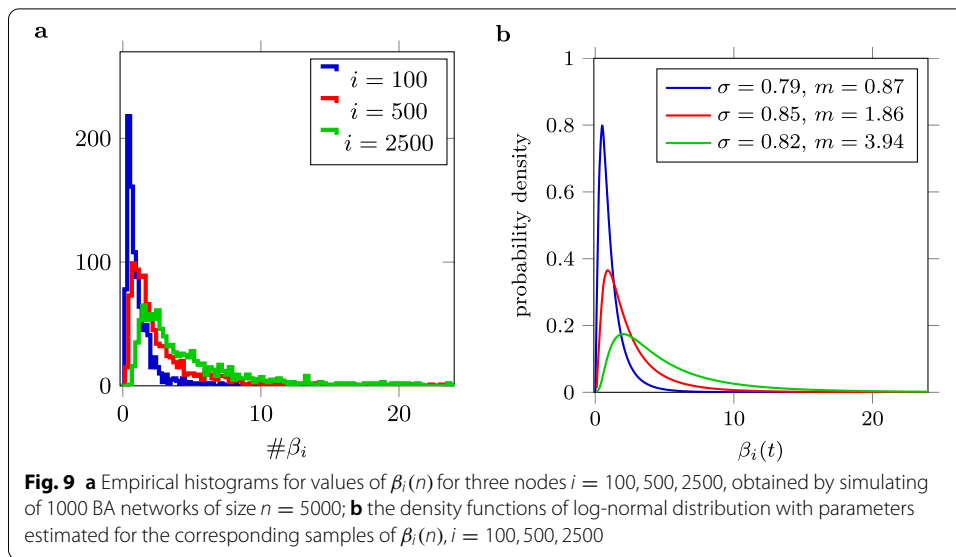
Fig. 16a ( $n = 10^4$ ) and b ( $n = 10^5$ ), one can see that the share of nodes with  $\bar{\beta}_i(n) < 1$  decreases.

We verify our results by running experiments in which we create synthetic networks and obtain  $\beta$  distribution over all nodes in the network. Table 2 shows the averaged results obtained from 100 individual runs per each network with different parameters. We sum all nodes from the intervals of size 1 and divide them by the number of experiments. Then we calculate the percentage of nodes that did not fall into 0 to 1 interval. It can be seen that with the network growth the percentage of nodes with  $\beta$  exceeding 1 steadily increases with the growth of the network.

**FI distribution in BA networks**

Note that Eq. (9) gives the expected value of  $\beta_i(t)$  at moment  $t$ , and does not tell anything about how random variable  $\beta_i(t)$  is distributed at iteration  $t$ .

To understand how much the empirical values of  $\beta_i(t)$  obtained by simulations are scattered around its mean  $\bar{\beta}_i(t)$ , how skewed the values  $\beta_i(t)$  are, i.e. how the values  $\beta_i(t)$  obtained by simulations are distributed, we generate  $N = 1000$  networks of the same



size  $n = 5000$ , and for each of 1000 networks we find empirical values of the friendship index  $\beta_i(n)$  for their nodes at moment  $n$ . Figure 9a shows the corresponding histograms obtained for three nodes  $i = 100, i = 500$  and  $i = 2500$ . Simulations show that the distributions of  $\beta_i(t)$  are heavy tailed.

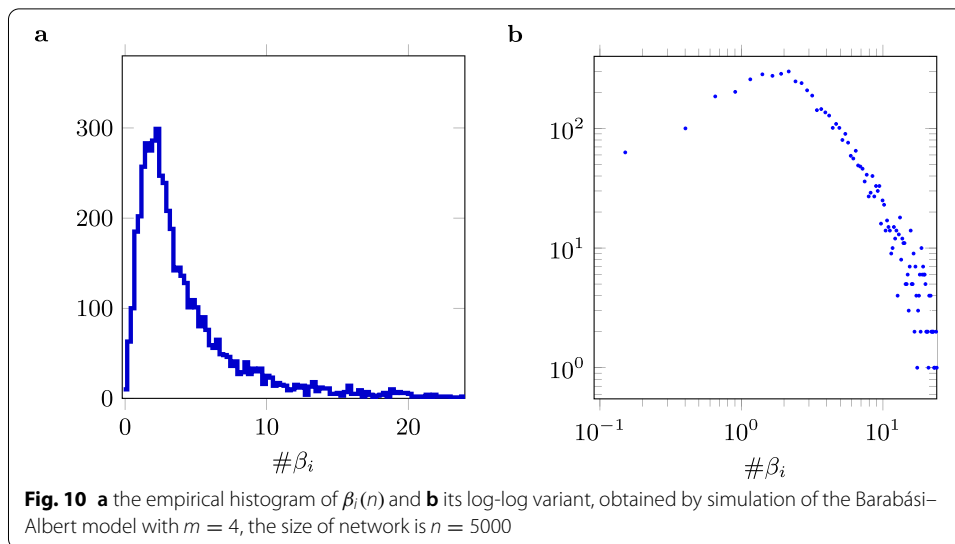
We use the Kolmogorov–Smirnov (K–S) test to compare 1000-length samples, obtained for nodes  $i = 1, \dots, 5000$ , with log-normal probability distribution. The null hypothesis is that the samples are drawn from the (reference) log-normal distribution. The general formula for the probability density function of the (standard) log-normal distribution is

$$f(x) = \frac{e^{-\frac{\log^2 \frac{x}{m}}{2\sigma^2}}}{x\sigma\sqrt{2\pi}}, \quad x > 0, \sigma > 0, m > 0,$$

where  $\sigma$  is the shape parameter,  $m$  is the scale parameter.

We let the threshold value for  $p$ -value (the significance level of the K–S test) at 0.05. Results of the Kolmogorov–Smirnov test show that  $p$ -values are more than 0.05 for all  $i$ -samples, they increase with growth of  $i$ , e.g.  $p$ -value for  $i = 500$  is equal to 0.16 and  $p$ -value for  $i = 2500$  is equal to 0.29. We may conclude that the observed samples are sufficiently consistent with the null hypothesis and that the null hypothesis may not be rejected. Figure 9b presents the density functions of the log-normal distribution with estimated parameters for corresponding nodes  $i = 100, 500, 2500$ .

To construct a histogram of FI for the Barabási–Albert network, we generate a network of size  $n = 5000$ , and then we obtain values over each of intervals of length 0.25. The obtained histogram (Fig. 10a) and its log-log variant (Fig. 10b) show that the FI distribution in the simulated network has heavy tail and a huge variance. The results also indicate that the overwhelming majority of nodes in the modeled BA networks



have the friendship index greater than 1, which means that there is a remarkable friendship paradox.

Moreover, we conduct the Kolmogorov–Smirnov test to verify the null hypothesis (that the samples are drawn from the (reference) log-normal distribution). Results of the K–S test show that  $p$ -value is equal to 0.1313. Thus, observed data are sufficiently consistent with the null hypothesis, i.e. the null hypothesis may not be rejected at significance level 0.05.

### Triadic closure model analysis

#### Triadic closure model

The triadic closure model by Holme and Kim (2002) generates growth networks as follows. At each iteration  $t$ :

1. One node  $v_t$  is added;
2.  $m$  links are added to the network, which connect the new node  $v_t$  and  $m$  other nodes as follows:
  - (a) the first link at iteration  $v_t$  is connected to node  $v_i$  using preferential attachment (probability of being connected to node  $v_i$  is proportional to its degree)
  - (b) other remaining  $m - 1$  links connect the new node  $v_t$  as follows:
    - (b1) with probability  $p$ , triad formation occurs, which means that the link is attached to an arbitrary neighbor of the node  $v_i$ ;
    - (b2) with probability  $1 - p$ , the link is connected to one node of the network using preferential attachment.

The main difference between the BA and triadic closure models is that the latter creates many more triangles during the network growth process. The degree evolution of every node over time, as described in Holme and Kim (2002), follows

$$k_i(t) = m \left( \frac{t}{i} \right)^\beta \tag{10}$$

with the same exponent  $\beta = \frac{1}{2}$  for all nodes. Furthermore, it was shown in Holme and Kim (2002) that the degree distributions of networks generated by the triadic closure model follows the power law with exponent  $\gamma = 3$  that does not depend on  $p$  and  $m$ . Throughout the rest of this section we will assume  $p \neq 0$  and  $m \geq 2$ .

If new vertex  $v_{t+1}$  chooses a neighbor  $v_j$  of vertex  $v_i$  at step (a), and then new vertex  $v_{t+1}$  links by remaining  $m - 1$  edges to its neighbors (with probability  $p$ ) at step (b1), so the probability that they are common neighbors of node  $v_i$  is very high. This is due to the fact that the triadic closure model generates many triangles as the network grows, which leads to a greater local and global clustering. To analyze the growth of networks, we need certain facts about the behavior of the clustering coefficient over time. The average value of the clustering coefficient at iteration  $t$  is defined as follows

$$\theta(t) := \frac{1}{t} \sum_{i=1}^t \theta_i(t),$$

where  $\theta_i(t)$  denotes the local clustering coefficient of node  $v_i$  at iteration  $t$ . In ‘‘Appendix C’’, we obtain equations describing the dynamics of the clustering coefficient and show that for networks generated by the triadic closure model, its value tends to some constant value  $\theta$  with an increase in the number of iterations. Note that this fact was established empirically in the work (Holme and Kim 2002). Therefore, we approximate  $\theta(t)$  by  $\theta$  in our analysis.

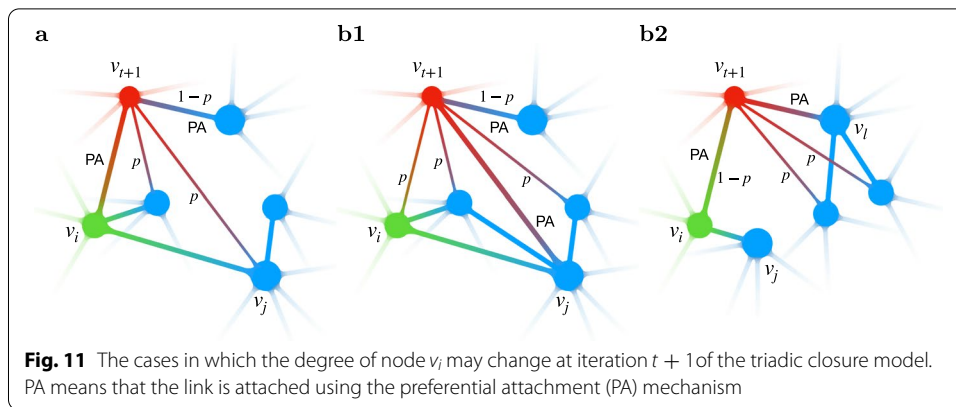
**Preliminary analysis**

In order to obtain stochastic relations describing the dynamics of processes  $s_i(t)$  and  $\beta_i(t)$  in time, we need auxiliary indicator variables that characterize random events associated with whether this node will be selected by a new node at steps (a), (b1) or (b2) (variables  $\zeta_i^{(a,t+1)}$ ,  $\zeta_{i,l}^{(b1,t+1)}$  and  $\zeta_{i,l}^{(b2,t+1)}$ , respectively):

$$\begin{aligned} \zeta_i^{(a,t+1)} &= \begin{cases} 1, & \text{if node } v_{t+1} \text{ links to node } v_i \text{ at step (a) of iteration } t + 1, \\ 0, & \text{otherwise,} \end{cases} \\ \zeta_{i,l}^{(b1,t+1)} &= \begin{cases} 1, & \text{if } l\text{-th edge, } l = 1, \dots, m - 1, \text{ links node } v_{t+1} \text{ with node } v_i \\ & \text{at step (b1) of iteration } t + 1, \\ 0, & \text{otherwise,} \end{cases} \\ \zeta_{i,l}^{(b2,t+1)} &= \begin{cases} 1, & \text{if } l\text{-th edge, } l = 1, \dots, m - 1, \text{ links node } v_{t+1} \text{ with node } v_i \\ & \text{at step (b2) of iteration } t + 1, \\ 0, & \text{otherwise.} \end{cases} \\ \zeta_i^{(b1,t+1)} &:= \sum_{l=1}^{m-1} \zeta_{i,l}^{(b1,t+1)}, \quad \zeta_i^{(b2,t+1)} := \sum_{l=1}^{m-1} \zeta_{i,l}^{(b2,t+1)}. \end{aligned}$$

We have





$$\mathbb{E}(\zeta_i^{(a,t+1)}) = \frac{d_i(t)}{2|E(t)|},$$

$$\mathbb{E}(\zeta_i^{(b1,t+1)}) = \sum_{(v_i, v_j) \in E(t)} \frac{d_j(t)}{2|E(t)|} \frac{1}{d_j(t)} p(m-1) = \frac{d_i(t)}{2|E(t)|} p(m-1),$$

$$\mathbb{E}(\zeta_i^{(b2,t+1)}) = \sum_{j=1}^{|V(t)|} \frac{d_j(t)}{2|E(t)|} \frac{d_i(t)}{2|E(t)|} (1-p)(m-1) = \frac{d_i(t)}{2|E(t)|} (1-p)(m-1),$$

where  $|E(t)|$  is the number of edges in the network after  $t$  iterations.

Let nodes  $v_j$  and  $v_i$  be neighbors, i.e.  $(v_j, v_i) \in E(t)$ . The probability that a randomly chosen neighbor of node  $v_j$  is also the neighbor of node  $v_i$  can be approximated by the value of the averaged clustering coefficient  $\theta(t)$  obtained in “Appendix C” (see Eq. (30)).

Let us study how the values of  $d_i(t)$  and  $s_i(t)$  change during the iteration  $t + 1$ . To do this, let us consider two groups of situations: situations in which a new node is attached to node  $v_i$  at iteration  $t + 1$ , and situations where a new node is attached to a neighbor of node  $v_i$ , but not to itself.

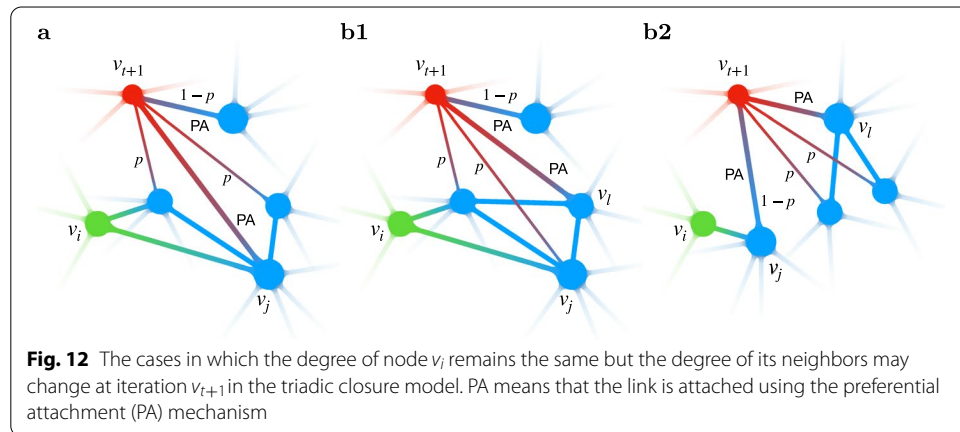
First, we examine the cases in which the degree of node  $v_i$  changes as the result of linking the new node  $v_{t+1}$  to node  $v_i$ . It may occur at steps (a), (b1) and (b2). Let us consider them separately:

- If new node  $v_{t+1}$  joins node  $v_i$  at step (a) (see Fig. 11a), then the total degree of all neighbors for the node  $v_i$  will increase by
  - $m$ , since this node will obtain the new neighbor with a degree of  $m$ , while its other neighbors will not change their degrees;
  - $p(m - 1)$ , since we should add, on average,  $p(m - 1)$  edges to the neighbors of the node  $v_i$  as a result of the choice of the node  $v_i$  at step (a).

Therefore, in this case we have  $s_i(t + 1) - s_i(t) = m(1 + p(m - 1))$ . The degree of node  $v_i$  will be increased by 1, i.e.  $d_i(t + 1) - d_i(t) = 1$ .

**Table 3** The cases in which the degree of node  $v_i$  changes

	Step (a)	Step (b1)	Step (b2)
$\Delta s_i(t + 1)$	$m + p(m - 1)$	$m + p(m - 2)\theta$	$m$
$\Delta d_i(t + 1)$	1	1	1
Probability	$\frac{d_i(t)}{2 E(t) }$	$\frac{d_i(t)}{2 E(t) }p(m - 1)$	$\frac{d_i(t)}{2 E(t) }(1 - p)(m - 1)$



- In order for node  $v_i$  to be selected at step (b1), it is necessary that at step (a) one of the neighbors  $v_j$  of node  $v_i$  was selected, and then the node  $v_i$  would be selected as a neighbor of the node  $v_j$  at step (b1) (see Fig. 11b1). In this case, the degree of node  $v_i$  will increase by 1, while the total degree of its neighbors  $s_i(t)$  will increase by
  - $m$  as a result of joining node  $v_{t+1}$  of degree  $m$  to the node  $v_j$ ;
  - $p(m - 2)\theta$  as a result of linking the newborn node  $v_{t+1}$  to  $p(m - 2)\theta$  (in average) common neighbors of nodes  $v_j$  and  $v_i$  at step (b1). After one of the  $m - 1$  edges is drawn to node  $v_i$ , it remains to attach  $m - 2$  nodes.

Thus,  $s_i(t + 1) - s_i(t) = m + p(m - 2)\theta$  and  $d_i(t + 1) - d_i(t) = 1$  in this case.

- If node  $v_i$  was selected at step (b2) (see Fig. 11b2), then the node degree increases by 1, while the total degree of its neighbors will increase by  $m$  as a result of joining the new neighbor  $v_{t+1}$  of degree  $m$ .

Denote  $\Delta s_i(t + 1) := s_i(t + 1) - s_i(t)$ ,  $\Delta d_i(t + 1) := d_i(t + 1) - d_i(t)$ .

All these cases are present in Table 3 with corresponding probabilities.

Now let us consider those cases in which no links are drawn to node  $v_i$  as a result of the iteration  $t + 1$ , but nevertheless one of its neighbors increases its degree. This is possible in three cases:

- If one of the neighbors  $v_j$  of the node  $v_i$  is selected at step (a), and the node  $v_i$  is not chosen at step (b1) (see Fig. 12a), then the degree of node  $v_i$  will not change, while the overall degree of its neighbors  $s_i(t)$  will increase by
  - 1 as a result of joining node  $v_{t+1}$  to the node  $v_j$ ;
  - $p(m - 1)\theta$  as a result of linking the newborn node  $v_{t+1}$  to common neighbors of nodes  $v_j$  and  $v_i$  at step (b1).

The probability of this case is

$$\sum_{(v_j, v_i) \in E(t)} \frac{d_j(t)}{2|E(t)|} \left(1 - \frac{p(m - 1)}{d_j(t)}\right) = \frac{s_i(t)}{2|E(t)|} - \frac{d_i(t)}{2|E(t)|} p(m - 1).$$

- If one of the neighbors  $v_l$  of the node  $v_j$  (which in turn is the neighbor of the node  $v_i$ ) is selected at step (a), and then the node  $v_j$  is chosen at step (b1) (see Fig. 12b1), then the degree of node  $v_i$  does not change, while the total degree of its neighbors will increase by
  - 1 as a result of joining node  $v_{t+1}$  of to the node  $v_j$ ;
  - $p(m - 2)\theta^2$  as a result of linking the newborn node  $v_{t+1}$  to common neighbors of nodes  $v_l, v_j$  and  $v_i$  at step (b1).

The probability of this case is

$$\sum_{j: (v_j, v_i) \in E} \sum_{l: (v_l, v_j) \in E, l \neq i} \frac{d_l(t)}{2|E(t)|} \frac{1}{d_l(t)} p(m - 1) = \frac{s_i(t) - d_i(t)}{2|E(t)|} p(m - 1).$$

- If one of the neighbors  $v_l$  of the node  $v_j$  (the neighbor of the node  $v_i$ ) is chosen at step (a), and then the node  $v_j$  is selected at step (b2) (see Fig. 12b2), then the degree of the node  $v_i$  also does not change, while the total degree of its neighbors will increase by 1. The probability of this case is

$$\sum_{j: (v_j, v_i) \in E} \sum_{l=1}^{|V(t)|} \frac{d_l(t)}{2|E(t)|} \frac{d_j(t)}{2|E(t)|} (1 - p)(m - 1) = \frac{s_i(t)}{2|E(t)|} (1 - p)(m - 1).$$

For clarity, we present all these three cases in Table 4.

### Dynamics of the total degree of the node neighbors

Combining all the cases, the expected value of  $\Delta s_i(t + 1) = s_i(t + 1) - s_i(t)$  is equal to

**Table 4** The cases in which the degree of node  $v_i$  does not change

	Step (a)	Step (b1)	Step (b2)
$\Delta s_i(t + 1)$	$1 + p(m - 1)\theta$	$1 + p(m - 2)\theta^2$	1
$\Delta d_i(t + 1)$	0	0	0
Probability	$\frac{s_i(t)}{2 E(t) } - \frac{d_i(t)}{2 E(t) } p(m - 1)$	$\frac{s_i(t) - d_i(t)}{2 E(t) } p(m - 1)$	$\frac{s_i(t)}{2 E(t) } (1 - p)(m - 1)$

$$\begin{aligned}
 \mathbb{E}(\Delta s_i(t+1)) &= (m + p(m-1)) \frac{d_i}{2|E|} \\
 &+ (m + p(m-2)\theta) \frac{d_i}{2|E|} p(m-1) + m \frac{d_i}{2|E|} (1-p)(m-1) \\
 &+ (1 + p(m-1)\theta) \left( \frac{s_i}{2|E|} - \frac{d_i}{2|E|} p(m-1) \right) \\
 &+ (1 + p(m-2)\theta^2) \frac{s_i - d_i}{2|E|} p(m-1) + \frac{s_i}{2|E|} (1-p)(m-1) \\
 &= d_i \frac{m}{2|E|} \left( m - \frac{p(m-1)}{m} (1 + p\theta + p(m-2)\theta^2) \right) \\
 &+ s_i \frac{m}{2|E|} \left( 1 + \frac{p(m-1)}{m} (\theta + p^2(m-2)\theta^2) \right).
 \end{aligned} \tag{11}$$

Since  $\mathbb{E}(|E(t)|) = mt$ , we obtain the linear nonhomogeneous differential equation of first order corresponding to the difference equation (11):

$$\frac{ds_i(t)}{dt} = C_1(p, m) \frac{d_i(t)}{t} + C_2(p, m) \frac{s_i(t)}{t}, \tag{12}$$

where

$$C_1(p, m) = \frac{1}{2} \left( m - \frac{p(m-1)}{m} (1 + p\theta + p(m-2)\theta^2) \right), \tag{13}$$

$$C_2(p, m) = \frac{1}{2} \left( 1 + \frac{p(m-1)}{m} (\theta + p(m-2)\theta^2) \right). \tag{14}$$

The solution of Eq. (12) can be found in ‘‘Appendix D’’, where it is shown (see Eq. (35)) that the expected value of  $s_i(t)$  follows

$$\bar{s}_i(t) = \mathbb{E}(s_i(t)) = c_1 t^{\frac{1}{2}} + c_2 t^{C_2(p,m)}, \tag{15}$$

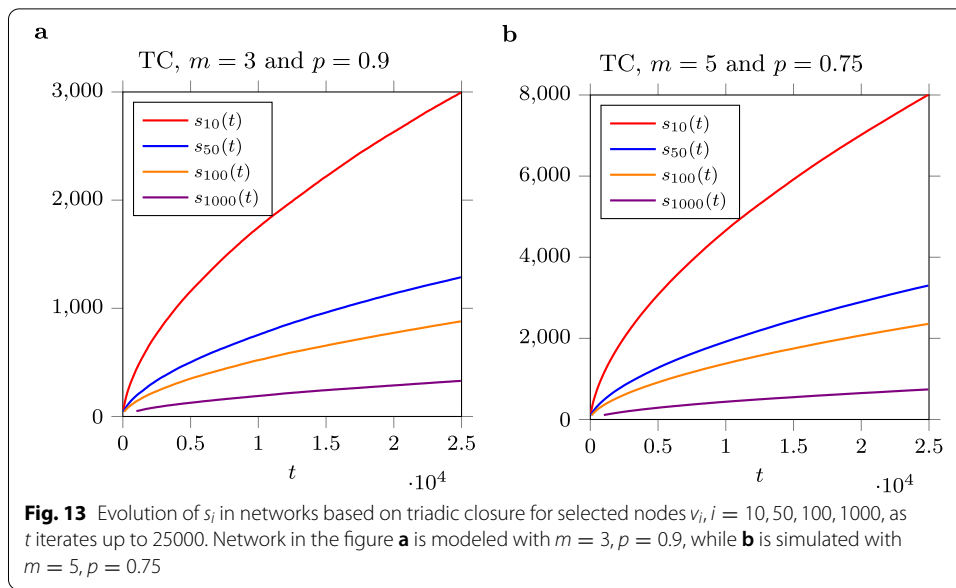
where  $c_1, c_2$  do not depend on  $t$ .

Eq. (15) shows that there is a huge difference between the BA model (Eq. 5) and the triadic closure model with respect to the dynamic behavior of the total degree of all neighbors of a node.

Figure 13 presents the evolution of the  $s_i(t)$  over  $t$  averaged over 100 independent simulations. The figures show that the empirical behavior of  $s_i(t)$  is indistinguishable from predictions of Eq. (15).

### Dynamics of the friendship index of a node

In this subsection we will find the equation describing the time evolution of  $\beta_i(t)$ , i.e. the average value of the friendship index for node  $v_i$ . The following difference equation for  $\Delta\beta_i(t+1) := \beta_i(t+1) - \beta_i(t)$  follows from the analysis of Section ‘‘Preliminary analysis’’:

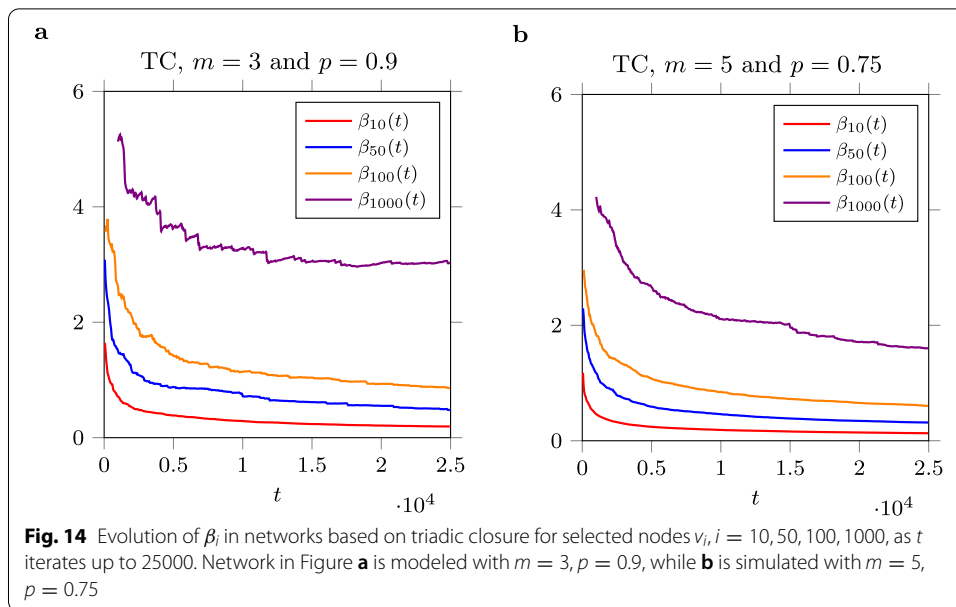


$$\begin{aligned}
 \mathbb{E}(\Delta\beta_i(t+1)) &= \left( \frac{s_i + m + p(m-1)}{(d_i + 1)^2} - \frac{s_i}{d_i^2} \right) \frac{d_i}{2|E|} \\
 &+ \left( \frac{s_i + m + p(m-2)\theta}{(d_i + 1)^2} - \frac{s_i}{d_i^2} \right) \frac{d_i}{2|E|} p(m-1) \\
 &+ \left( \frac{s_i + m}{(d_i + 1)^2} - \frac{s_i}{d_i^2} \right) \frac{d_i}{2|E|} (1-p)(m-1) \\
 &+ \left( \frac{s_i + 1 + p(m-1)\theta}{d_i^2} - \frac{s_i}{d_i^2} \right) \left( \frac{s_i}{2|E|} - \frac{d_i}{2|E|} p(m-1) \right) \\
 &+ \left( \frac{s_i + 1 + p(m-2)\theta^2}{d_i^2} - \frac{s_i}{d_i^2} \right) \frac{s_i - d_i}{2|E|} p(m-1) \\
 &+ \left( \frac{s_i + 1}{d_i^2} - \frac{s_i}{d_i^2} \right) \frac{s_i}{2|E|} (1-p)(m-1) \\
 &\sim \frac{m}{2|E|} \frac{s_i}{d_i^2} \left( -1 + \frac{p(m-1)}{m} (\theta + p(m-2)\theta^2 + \frac{d_i}{(d_i+1)^2}) \right) \\
 &+ \frac{md_i}{2|E|(d_i+1)^2} \left( m - \frac{p(m-1)}{m} (1 + p\theta + p(m-2)\theta^2) \right).
 \end{aligned} \tag{16}$$

It follows from  $|E(t)| \sim mt$ ,  $d_i(t)/(d_i + 1) \sim 1$  and  $d_i(t)/(d_i + 1)^2 \sim 1/d_i$  that the difference equation (16) may be approximated by the following linear nonhomogeneous differential equation of first order:

$$\frac{d\beta_i(t)}{dt} \sim \frac{\beta_i}{t} \left( C_2(p, m) - 1 + \frac{1}{d_i(t)} \right) + \frac{C_1(p, m)}{td_i(t)}, \tag{17}$$

where  $C_1(p, m)$  and  $C_2(p, m)$  are defined in Eqs. (13).



The solution of Eq. (17) can be found in “Appendix E”, where it is proved (see Eq. (41)) that the expected value of friendship index  $\beta_i(t)$  follows

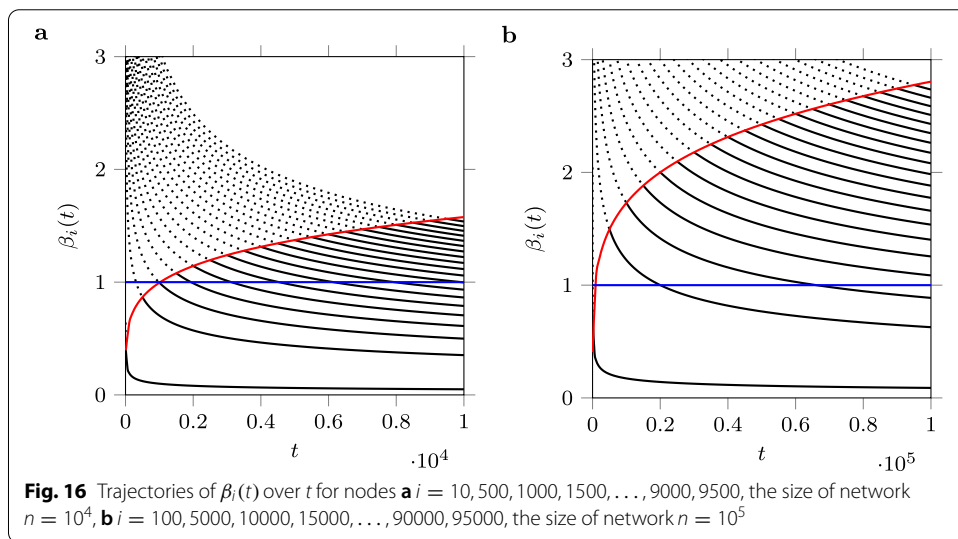
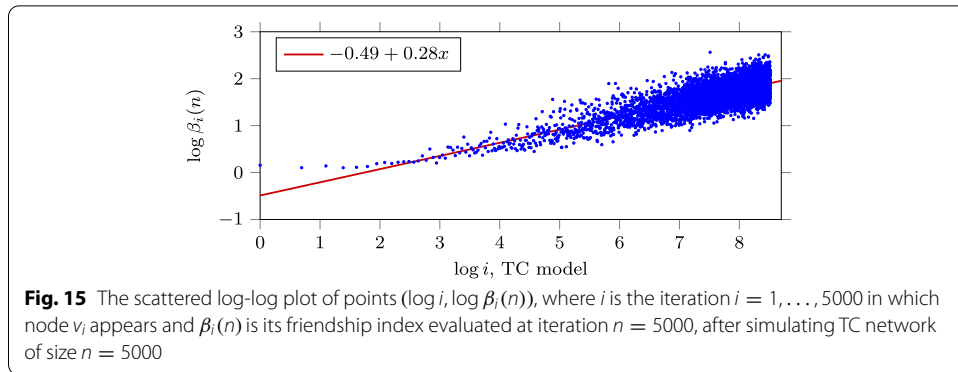
$$\bar{\beta}_i(t) := \mathbb{E}(\beta_i(t)) \sim \frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} t^{C_2(p,m)-\frac{1}{2}} + o(t^{-\frac{1}{2}}). \tag{18}$$

Thus, the FI dynamics of every single node follows the power law with exponent  $C_2(p, m) - 1$ , i.e. the decrease of FI over time for the triadic closure model is much slower than for the BA model.

Equation (18) depicts the dynamics of the expected value of  $\beta_i(t)$  with the growth of the network built in accordance to the TC model. It should be mentioned that the equation describes the trajectory of the *expected* values of the FI, while in each separate simulation the path can deviate significantly from the theoretical curve of Eq. (18). However, if we average the paths obtained from a large number of simulations, we can get a trajectory that is located close to the curve predicted by Eq. (18). The curves for several nodes are shown in Fig. 14. The figure exhibits trajectories of the friendship index  $\beta_i(t)$  for nodes  $v_i, i = 10, 50, 100, 1000$ , for a simulated network of size  $|V| = 25,000$ . It can be seen that  $\beta_i(t)$  are gradually decreasing to 0 with the growth of the network.

The analysis of Eq. (18) allows us to obtain the following conclusions regarding the dynamics of FI over time in the TC model:

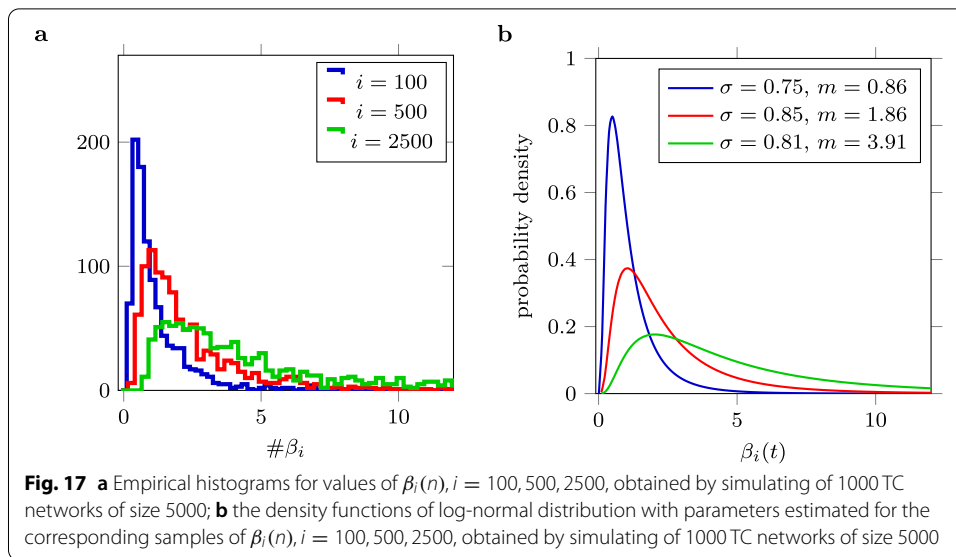
1. The curve depicting the expected value of  $\beta_i(t)$  follows Eq. (18), and as  $t$  tends to  $\infty$ , i.e. with the growth of the network, this value decreases to zero with the rate  $\frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} t^{C_2(p,m)-\frac{1}{2}}$  depending on the number of iterations  $t$ . The value of  $\beta_i(t)$  tends to zero for all nodes  $i$ , as  $t \rightarrow \infty$ ;
2. For node  $v_i$  that appears on iteration  $i$ , the expected initial value of  $\beta_i(i)$  is  $\frac{1}{2} i^{C_2(p,m)-\frac{1}{2}}$ . It means that initial FI values for new nodes tends to  $\infty$  as  $t \rightarrow \infty$ ;



3. For each node  $v_j$  that appeared after the iteration  $i$ , i.e.  $j > i$ , the value of  $\bar{\beta}_j(t)$  will be higher than the value of  $\bar{\beta}_i(t)$  for node  $v_i$ . The scattered plot of points  $(\log i, \log \beta_i(n))$ ,  $i = 1, \dots, 5000$ , obtained for simulated TC network of size  $n = 5000$  with  $m = 4$ , is shown on Fig. 15. It presents the empirical evidence of the fact that the earlier a node appeared, the lower its friendship index is, and vice versa.
4. Figure 16a, b show the theoretical trajectories of the friendship index for different nodes of a network built according to TC model with  $m = 3, p = 0.9$  over iterations. The starting points of each individual trajectory lies on the curve  $0.25t^{0.2}$ . Each trajectory is shown as a solid line, while the dotted line corresponds to those portions of the same curve where the corresponding node has not yet existed. The density of the values of  $\beta_i(t)$  at moment  $t$  increases from 0 to  $0.5 \log t$ .
5. Solving the equation  $\frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} t^{C_2(p,m)-\frac{1}{2}} = 1$ , we find index

$$i^* = i^*(t) = 4t^{2(1-C_2(p,m))} = 4t^{\frac{p(m-1)}{m}(\theta+p(m-2)\theta^2)},$$

for which the value of  $\bar{\beta}_{i^*}(t) = 1$  at iteration  $t$ . Thus, for every node  $v_j$  that appeared earlier, i.e.  $j < i^*$ , the inequality  $\bar{\beta}_j(t) < 1$  holds, while for any node  $v_l$  that appeared later, i.e.  $l > i^*$ , the relation  $\bar{\beta}_l(t) > 1$  holds. The share of nodes for which the



expected value of the friendship index is less than 1 is equal to  $\frac{i^*(t)}{t} = 4m^{1-2C_2(p,m)}$ . Therefore, with an increase in the number of iterations  $t$ , the fraction of such nodes decreases and tends to zero. On the other hand, the proportion of nodes for which the expected value of the friendship index is greater than 1 tends to 1, i.e. the friendship paradox will affect almost all nodes as the network grows. This effect is clearly seen on Fig. 16a ( $n = 10^4$ ) and b ( $n = 10^5$ ), where the share of nodes with  $\bar{\beta}_i(n) < 1$  shrinks.

It should be noted that Eq. (18) estimates the *expected* value of  $\beta_i(t)$  at iteration  $t$ , and does not represent information about how random variable  $\beta_i(t)$  is distributed at moment  $t$ .

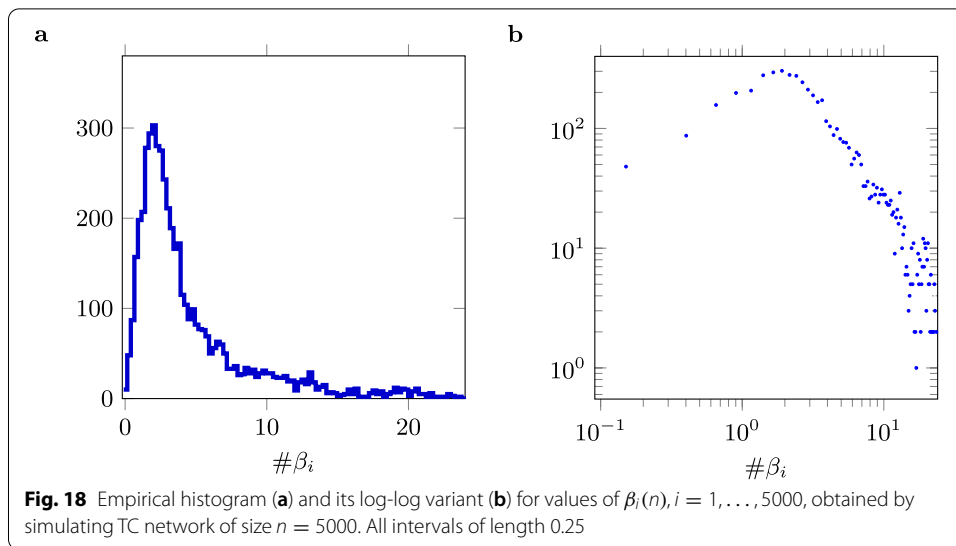
To figure out how the values  $\beta_i(n)$  are distributed for a fixed  $i$  at moment  $n$ , we generated  $N = 1000$  TC networks of the same size  $n = 5000$ , and for each of 1000 networks we calculate empirical values of the friendship index  $\beta_i(n)$  for their nodes. Figure 17a shows the corresponding histograms obtained for three nodes  $i = 100$ ,  $i = 500$  and  $i = 2500$ . Simulations show that the distributions of  $\beta_i(t)$  are heavy tailed.

We use the Kolmogorov–Smirnov test to compare 1000-length samples, obtained for nodes  $i = 1, \dots, 5000$ , with log-normal probability distribution. The null hypothesis is that the samples are drawn from the (reference) log-normal distribution.

We set the significance level of the K–S test at 0.05. Results of the Kolmogorov–Smirnov test show that  $p$ -values are more than 0.05 for all  $i$ -samples, they increase with growth of  $i$ , e.g.  $p$ -value for  $i = 500$  is equal to 0.12 and  $p$ -value for  $i = 2500$  is equal to 0.31. We may conclude that the observed samples are sufficiently consistent with the null hypothesis, and we can not reject the null hypothesis. Figure 17b presents the density functions of the log-normal distribution with estimated parameters for corresponding nodes  $i = 100, 500, 2500$ .

To construct a histogram of FI for the TC network, we generate a network of size  $n = 5000$ , and then we obtain values over each of intervals of length 0.25. The





**Table 5** Characteristics of friendship index distributions in synthetic networks generated with triadic closure model

$ M $	$ E $	$\beta > 1, \%$
5,000	15,000	91.678
5,000	15,000	92.721
10,000	30,000	91.883
10,000	50,000	93.093
15,000	45,000	92.019
15,000	75,000	93.280
25,000	75,000	92.006
25,000	125,000	93.307
50,000	150,000	92.086
50,000	250,000	93.466

obtained histogram (Fig. 18a) and its log-log variant (Fig. 18b) show that the FI distribution in the simulated TC network has heavy tail and a huge variance. The results also indicate that the vast majority of nodes in the modeled TC networks have the friendship index greater than 1, which means that there is a significant presence of the friendship paradox.

Moreover, we carry out the K–S test to verify that the samples are drawn from the (reference) log-normal distribution (the null hypothesis). Results of the K–S test show that  $p$ -value is equal to 0.1331. Thus, the null hypothesis may not be rejected at significance level 0.05.

Similarly to networks generated with Barabási–Albert model, we verify our results by running experiments in which we create synthetic networks, based on triadic closure model. From them we obtain  $\beta$  distribution over all nodes in the network. Table 5 shows the averaged results obtained from 100 individual runs per each network with different parameters. The percentage of nodes with  $\beta$  exceeding 1 grows with the network size.

**Table 6** Simulated networks based on Barabási–Albert and triadic closure model.  $\gamma$  is the power law exponent

Network	$ V $	$ E $	$\beta > 1, \%$	$\gamma$
BA, $m = 3$	335,000	1,005,000	91.1	2.55
TC, $m = 3, p = 0.9$	335,000	1,005,000	92.5	2.23
BA, $m = 3$	50,000	150,000	91.034	2.32
TC, $m = 3, p = 0.75$	50,000	150,000	92.086	2.22
BA, $m = 5$	50,000	250,000	92.672	2.78
TC, $m = 5, p = 0.75$	50,000	250,000	93.466	2.69

### The comparison of friendship index distribution in BA and TC networks

In this section we take a look at the friendship index distribution in networks created with Barabási–Albert and triadic closure models of the same sizes. We build histograms to observe the friendship index distribution in networks generated by BA and TC models with different model parameters (the size of network  $|V|$ , the number of attached links  $m$  (which results in the amount of edges  $|E|$ ), the probability  $p$  (only in TC model)). This section presents empirical results indicating that friendship index distributions are heavy tailed for both models. Their FI distributions, which can be observed on the log-log plot, are approximately linear on a part of the range. Network statistics are shown in Table 6. The power-law exponents  $\gamma$  are obtained fitting linear regression to logarithmically binned data points taken from the decreasing parts of histograms.

To construct a histogram for both Barabási–Albert and triadic closure network, we generated 100 networks with 50,000 nodes and 250,000 edges, corresponding BA and TC networks were created with parameter  $m = 5$ . The obtained values were averaged over each of intervals of length 1. The obtained histogram and its log-log version are shown in Fig. 19a–d show that FI distributions in the modeled networks have heavy tails and huge variances. Moreover, the friendship index for the majority of nodes in the modeled networks is greater than 1, which means that there is a friendship paradox in both networks. It should be noted that the exponent  $\gamma$  for TC network is less than the exponent obtained for BA graph, which means that the TC networks are more heavy tailed than BA graph in the sense of FI distribution. We also constructed 100 networks with parameter  $m = 3$  as seen in Fig. 20. There are no significant differences in the behaviour of the distribution for networks with different  $m$  parameter.

We also demonstrate how a single large simulated network with 335,000 nodes and 1,005,000 edges behaves. Corresponding BA and TC networks were created with parameter  $m = 3$ . The obtained histogram and its log-log version are shown in Fig. 21a–d, which show that FI distributions in the modeled networks have heavy tails and a huge variance (e.g.  $\mathbb{E}(\beta_i) = 13.16$ ,  $\text{VAR}(\beta_i) = 439.8$  for BA network).

It can be seen that the majority of nodes in both BA and TC simulated networks have friendship index greater than 1, which is a sign of the presence of friendship paradox. Moreover, all distributions of  $\beta$  have a distinguishable slope, however the length of the tail may differ for the networks of the same size. We may see that all  $\beta$  distributions follow the power law. For convenience, results for all networks are gathered in

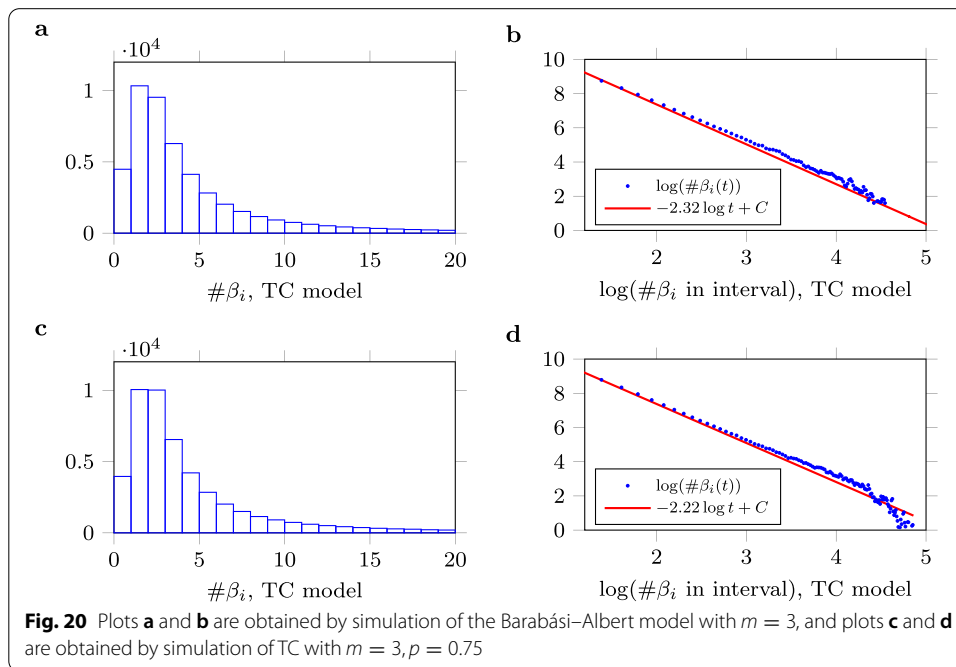
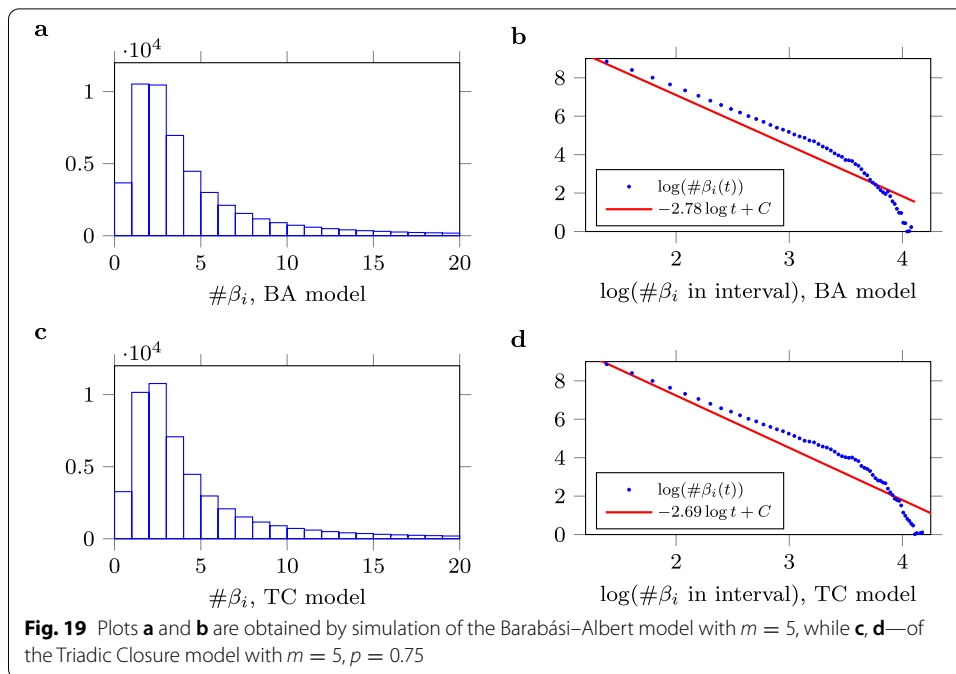
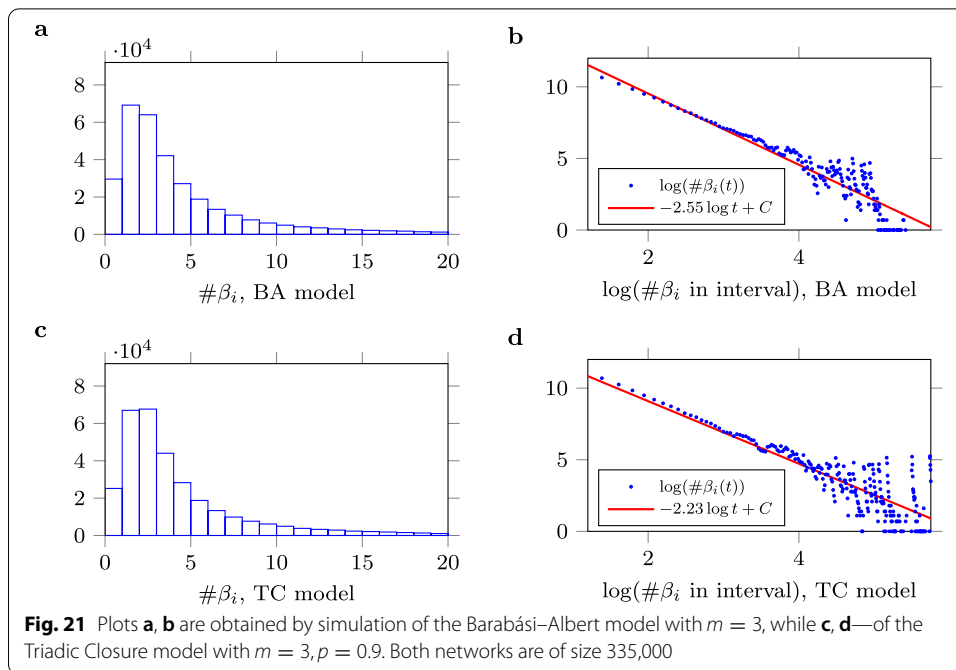


Table 6. In addition, the experiments show that by changing the parameters  $m$  (for BA model) and  $m, p$  (for TC model), it is possible to adjust the FI distributions of the generated networks, although the general behaviour remains the same.



## Conclusion

This paper studies the properties of the friendship index dynamic behavior for growth networks generated by the Barabási–Albert and the triadic closure models. The results were obtained using mean-field methods and rate equations. They show that there is an undoubted presence of the friendship paradox in such types of networks. The paper extends the analysis of paper (Sidorov et al. 2021) in which the simplest version of BA model was examined (with one attached link to a new node at each iteration).

Several real networks were analyzed in the paper. It turned out that their FI distributions have heavy tails. The same picture is observed for networks generated from BA and TC models. We studied the distribution of the friendship index for an arbitrary vertex  $i$  of the network at iteration  $t$ : first, we analytically found the expected value of the friendship index  $\beta_i(t)$  at moment  $t$ ; second, we have shown (using simulations) that the friendship index of any node at iteration  $t$  is a log-normally distributed random variable. In addition, experiments have shown that the distributions of the friendship index in both types of networks are also close to the log-normal distribution, i.e. have fat tails.

However, in the case of the triadic closure model, the tails of the FI distributions are heavier, which makes this model closer to reproducing the properties of real networks. It is also interesting that for the analyzed real networks, the exponent of the power-law is between 2 and 3, so although the average value of the friendship index can be easily found, the standard deviation is too large. Thus, the paper reveals the scale-free property of real networks with respect to the FI values of their nodes.

Thus, this work shows that the use of such network growth mechanisms as PA or PA+TC leads to the appearance of the friendship paradox in the generated networks. It should be noted that for the formation of growth networks, there are many other

mechanisms, which describe well the phenomena arising in real ones. In this regard, it would be interesting to study the quantitative and qualitative features of the friendship paradox in networks generated using these mechanisms.

**Appendix A: The solution of Eq. (4)**

The solution of Eq. (4) has the form  $s_i(t) = u(t)v(t)$ , where  $v(t)$  follows

$$\frac{dv(t)}{dt} = v(t) \frac{1}{2t} \tag{19}$$

and  $u(t)$  satisfies the differential equation

$$\frac{du(t)}{dt} v(t) = m \frac{d_i(t)}{2t}. \tag{20}$$

The solution of (19) is

$$v(t) = t^{\frac{1}{2}}. \tag{21}$$

Then the solution of (20) is

$$u(t) = m \int \frac{d_i(t)}{2t^{\frac{3}{2}}} dt + C. \tag{22}$$

Therefore,

$$s_i(t) = mt^{\frac{1}{2}} \left( \int \frac{d_i(t)}{2t^{\frac{3}{2}}} dt + C \right).$$

Note that at each moment  $t$ ,  $d_i$  is a random value with a probability density function  $\kappa_i(x)$  s.t.  $\bar{k}_i(t) := \mathbf{E}(d_i(t)) = \int x\kappa_i(x)dx = m\left(\frac{t}{i}\right)^{\frac{1}{2}}$  and  $\int \kappa_i(x)dx = 1$ . Then the expected value of  $s_i(t)$  is

$$\begin{aligned} \bar{s}_i(t) &= \mathbf{E}(s_i(t)) = m \int t^{\frac{1}{2}} \left( \int \frac{x}{2t^{\frac{3}{2}}} dt + C \right) \kappa_i(x) dx \\ &= mt^{\frac{1}{2}} \left( \int \left( \int x\kappa_i(x) dx \right) \frac{1}{2t^{\frac{3}{2}}} dt + C \left( \int \kappa_i(x) dx \right) \right) \\ &= m^2 t^{\frac{1}{2}} \left( \int \frac{1}{2i^{\frac{1}{2}} t} dt + C \right) = \frac{m^2}{2} \left( \frac{t}{i} \right)^{\frac{1}{2}} (\log t + c(i)). \end{aligned} \tag{23}$$

**Appendix B: The solution of Eq. (8)**

The solution of (8) is  $\bar{\beta}_i(t) = u(t)v(t)$ , where  $v(t)$  follows the differential equation

$$\frac{dv(t)}{dt} = -v(t) \left( \frac{1}{2t} - \frac{3d_i(t)}{2t(d_i(t) + 1)^2} \right), \tag{24}$$

and  $u(t)$  satisfies

$$\frac{du(t)}{dt}v(t) = m \frac{d_i(t)}{2t(d_i(t) + 1)^2}. \tag{25}$$

The solution of (24) is

$$v(t) = t^{-\frac{1}{2}} \exp\left(\frac{3}{2} \int \frac{d_i(t)}{t(d_i(t) + 1)^2} dt\right). \tag{26}$$

Then the solution of (25) is

$$u(t) = m \int \frac{d_i(t)}{2t^{\frac{1}{2}}(d_i(t) + 1)^2} \exp\left(-\frac{3}{2} \int \frac{d_i(t)}{t(d_i(t) + 1)^2} dt\right) dt + C. \tag{27}$$

Then

$$\begin{aligned} \beta_i(t) &= u(t)v(t) = mt^{-\frac{1}{2}} \exp\left(\frac{3}{2} \int \frac{d_i(t)}{t(d_i(t) + 1)^2} dt\right) \\ &\quad \times m \left( \int \frac{d_i(t)}{2t^{\frac{1}{2}}(d_i(t) + 1)^2} \exp\left(-\frac{3}{2} \int \frac{d_i(t)}{t(d_i(t) + 1)^2} dt\right) dt + C \right) \\ &\sim mt^{-\frac{1}{2}} \left( 1 + \frac{3}{2} \int \frac{d_i(t)}{t(d_i(t) + 1)^2} dt \right) \\ &\quad \times m \left( \int \frac{d_i(t)}{2t^{\frac{1}{2}}(d_i(t) + 1)^2} \left( 1 - \frac{3}{2} \int \frac{d_i(t)}{t(d_i(t) + 1)^2} dt \right) dt + C \right). \end{aligned} \tag{28}$$

Note that at each iteration  $t$ ,  $d_i(t)$  is a random value with a p.d.f.  $\kappa_i(x)$  s.t.  $\bar{k}_i(t) := \mathbf{E}(d_i(t)) = \int x\kappa_i(x)dx = \left(\frac{t}{i}\right)^{\frac{1}{2}}$  and  $\int \kappa_i(x)dx = 1$ . Since  $\int \frac{x\kappa_i(x)dx}{(x+1)^2} \sim \int \frac{\kappa_i(x)dx}{x} \sim \frac{1}{m} \left(\frac{i}{t}\right)^{\frac{1}{2}}$ , it can be shown that the expected value of  $\beta_i(t)$  asymptotically follows

$$\bar{\beta}_i(t) := \mathbf{E}(\beta_i(t)) \sim \frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} \log t + o(t^{-\frac{1}{2}}). \tag{29}$$

**Appendix C: The dynamics of local clustering coefficient for the triadic closure model**

Let  $S_i(t)$  denote the number of triads in which the node  $v_i$  is one of the vertices at the iteration  $t$ . To find the values of the local clustering coefficients for nodes of a network constructed using the triadic closure model, we derive the equation characterizing the dynamics of the local clustering coefficient  $C_i(t)$  of node  $v_i$ , defined by

$$\theta_i(t) = \frac{2S_i(t)}{d_i(t)(d_i(t) - 1)},$$

over time.

Note that at the time of its appearance at the iteration  $i$  the degree of node  $v_i$  is  $m$ , while the expected number of triads, in which this new node is participated, will be  $\bar{S}_i(i) = (m - 1)p$ . Therefore,  $\bar{\theta}_i(i) = \frac{2p}{m}$ .

The change in the value of  $\theta_i$  at iteration  $t + 1$  may occur in three cases only:

- if the node  $v_i$  is selected in step 2(a) then the expected number of newly formed triads is  $p(m - 1)$ , i.e.  $S_i(t + 1) = S_i(t) + p(m - 1)$ , while the degree of node  $v_i$  increases by one, i.e.  $d_i(t + 1) = d_i(t) + 1$ . Thus, with probability  $\frac{d_i}{2|E(t)|}$  the difference  $\Delta\theta_i(t + 1) = \theta_i(t + 1) - \theta_i(t)$  will be

$$\frac{2(S_i(t) + p(m - 1))}{(d_i(t) + 1)d_i(t)} - \frac{2S_i(t)}{d_i(t)(d_i(t) - 1)};$$

- if the node  $v_i$  is selected in step 2(b1), then the number of triads  $S_i(t)$  and the degree of node  $v_i$  increase both by one and the difference  $\Delta C_i(t + 1)$  will be

$$\frac{2(S_i(t) + 1)}{(d_i(t) + 1)d_i(t)} - \frac{2S_i(t)}{d_i(t)(d_i(t) - 1)};$$

The probability of this event is

$$\sum_{l:(v_l, v_i) \in E} \frac{d_l}{2|E(t)|} \frac{1}{d_l} (m - 1)p = (m - 1)p \frac{\bar{d}_i}{2|E(t)|},$$

where  $|E(t)|$  denotes the number of edges in the network after iteration  $t$ .

- if the node  $v_i$  is selected in step 2(b2), then the number of triads  $S_i(t)$  remains the same, while  $d_i(t)$  increases by one. Thus, the difference  $\Delta C_i(t + 1)$  will be

$$\frac{2S_i(t)}{(d_i(t) + 1)d_i(t)} - \frac{2S_i(t)}{d_i(t)(d_i(t) - 1)};$$

the probability of this event is

$$\sum_{l:(v_l, v_i) \in E} \frac{d_l}{2|E(t)|} \frac{1}{d_l} (m - 1)(1 - p) = (m - 1)(1 - p) \frac{\bar{d}_i}{2|E(t)|}.$$

Combining all three cases, performing simple algebraic transformations and using  $|E(t)| \sim mt$  and  $d_i(t) \sim m(t/i)^{\frac{1}{2}}$ , we get the approximate differential equation

$$\frac{d\theta_i(t)}{dt} = -\frac{\theta_i(t)}{t} + \frac{2p(m - 1)i^{\frac{1}{2}}}{m^2 t^{\frac{3}{2}}}$$

Then, taking the initial condition  $\theta_i(i) = \frac{2p}{m}$  into account, we get its solution

$$\bar{\theta}_i(t) \sim \frac{2p(m - 1)}{m^2} \left(\frac{i}{t}\right)^{\frac{1}{2}} + \frac{2pi}{m^2 t}. \tag{30}$$

Then the average value of the clustering coefficient at iteration  $t$  is

$$\theta = \theta(t) := \frac{1}{t} \sum_{i=1}^t \bar{\theta}_i(t) \sim \frac{2p(m-1)}{m^2} \frac{1}{t^{\frac{3}{2}}} \sum_{i=1}^t i^{\frac{1}{2}} + \frac{2p}{m^2 t^2} \sum_{i=1}^t i = \frac{4pm-1}{3m^2} + o(t^{-1}),$$

i.e.  $\theta$  does not depend on  $t$ . However, it may be tuned to with the use of parameters  $p$  and  $m$ , i.e.  $\theta = \theta(p, m)$ . This facts were observed in paper (Holme and Kim 2002).

**Appendix D: The solution of Eq. (12)**

The solution of Eq. (12) can be written as  $s_i(t) = u(t)v(t)$ , where  $v(t)$  is the solution of

$$\frac{dv(t)}{dt} = C_2(p, m)v(t) \frac{1}{t} \tag{31}$$

while  $u(t)$  follows the differential equation

$$\frac{du(t)}{dt} v(t) = C_1(p, m) \frac{d_i(t)}{t}. \tag{32}$$

The solution of (31) is

$$v(t) = t^{C_2(p,m)}. \tag{33}$$

Then the solution of (32) is

$$u(t) = C_1(p, m) \int \frac{d_i(t)}{t^{1+C_2(p,m)}} dt + C. \tag{34}$$

Thus,

$$s_i(t) = t^{C_2(p,m)} \left( C_1(p, m) \int \frac{d_i(t)}{t^{1+C_2(p,m)}} dt + C \right).$$

At every iteration  $t$ ,  $d_i$  takes a random values according to a probability density function  $\kappa_i(x)$  satisfying  $\bar{k}_i(t) := \mathbf{E}(d_i(t)) = \int x\kappa_i(x)dx = m\left(\frac{t}{i}\right)^{\frac{1}{2}}$  and  $\int \kappa_i(x)dx = 1$ . The expected value of  $s_i(t)$  can be found as the expectation over the probability density function  $\kappa_i(x)$  as follows:

$$\begin{aligned} \bar{s}_i(t) &= \mathbf{E}(s_i(t)) = \int t^{C_2(p,m)} \left( C_1(p, m) \int \frac{x}{t^{1+C_2(p,m)}} dt + C \right) \kappa_i(x) dx \\ &= t^{C_2(p,m)} \left( C_1(p, m) \int \left( \int x\kappa_i(x) dx \right) \frac{1}{t^{1+C_2(p,m)}} dt + C \left( \int \kappa_i(x) dx \right) \right) \\ &= t^{C_2(p,m)} \left( C_1(p, m) \int \frac{m}{i^{\frac{1}{2}} t^{\frac{1}{2} + C_2(p,m)}} dt + C \right) = c_1 t^{\frac{1}{2}} + c_2 t^{C_2(p,m)}, \end{aligned} \tag{35}$$

where  $c_1, c_2$  do not depend on  $t$ .



**Appendix E: The solution of Eq. (17)**

The solution of Eq. (17) can be written as  $\bar{\beta}_i(t) = u(t)v(t)$  with  $v(t)$  defined as the solution of the differential equation

$$\frac{dv(t)}{dt} = v(t) \left( \frac{C_2(p, m) - 1}{t} + \frac{1}{td_i(t)} \right), \tag{36}$$

and  $u(t)$  satisfied the equation

$$\frac{du(t)}{dt}v(t) = \frac{C_1(p, m)}{td_i(t)}. \tag{37}$$

The solution of (36) is

$$v(t) = t^{C_2(p, m)-1} \exp \left( \int \frac{1}{td_i} dt \right). \tag{38}$$

Then the solution of (37) is

$$u(t) = C_1(p, m) \int \frac{1}{t^{C_2(p, m)}d_i(t)} \exp \left( - \int \frac{1}{td_i(t)} dt \right) dt + C. \tag{39}$$

Then

$$\begin{aligned} \beta_i(t) = u(t)v(t) &= C_1(p, m)t^{C_2(p, m)-1} \exp \left( \int \frac{1}{td_i(t)} dt \right) \\ &\times \left( \int \frac{1}{t^{C_2(p, m)}d_i(t)} \exp \left( - \int \frac{1}{td_i(t)} dt \right) dt + C \right) \\ &\sim C_1(p, m)t^{C_2(p, m)-1} \left( 1 + \int \frac{1}{td_i(t)} dt \right) \\ &\times \left( \int \frac{1}{t^{C_2(p, m)}d_i(t)} \left( 1 - \int \frac{1}{td_i(t)} dt \right) dt + C \right). \end{aligned} \tag{40}$$

The value of  $d_i(t)$  is randomly distributed according to the p.d.f.  $\kappa_i(x)$  such that  $\bar{k}_i(t) := E(d_i(t)) = \int x\kappa_i(x)dx = m\left(\frac{t}{i}\right)^{\frac{1}{2}}$  and  $\int \kappa_i(x)dx = 1$ . The expected value of  $\beta_i(t)$  is the expectation over the p.d.f.  $\kappa_i(x)$ , as follows: Since  $\int \frac{\kappa_i(x)dx}{x} \sim \frac{1}{m} \left(\frac{i}{t}\right)^{\frac{1}{2}}$ , the expected value of  $\beta_i(t)$  asymptotically follows

$$\bar{\beta}_i(t) := E(\beta_i(t)) \sim ct^{C_2(p, m)-1} + o(t^{-\frac{1}{2}}),$$

where  $c$  does not depend on  $t$ . It follows from the initial condition  $\bar{\beta}_i(t) = \frac{1}{2}t^{C_2(p, m)-\frac{1}{2}}$  that

$$\bar{\beta}_i(t) := E(\beta_i(t)) \sim \frac{1}{2} \left(\frac{i}{t}\right)^{\frac{1}{2}} t^{C_2(p, m)-\frac{1}{2}} + o(t^{-\frac{1}{2}}). \tag{41}$$

**Abbreviations**

PA: Preferential attachment; BA: Barabási–Albert; TC: Triadic closure; FI: Friendship index.

**Acknowledgements**

The authors are very grateful to referees for their valuable comments and helpful suggestions.

**Authors' contributions**

SS and SM conceived the idea presented in the paper. SS and SM developed the theory. SM and AG performed the necessary experiments. AG and SM verified the methods and results. All authors read and approved the final manuscript.

**Funding**

This work was supported by the Ministry of Science and Higher Education of the Russian Federation in the framework of the basic part of the scientific research state task, project FSRR-2020-0006.

**Availability of data and materials**

The algorithms and obtained data that support the findings of this study are freely available on GitHub: <https://github.com/Alexflames/local-degree-asymmetry/tree/main>

**Declarations****Competing interests**

The authors declare that they have no competing interests.

Received: 24 February 2021 Accepted: 12 June 2021

Published online: 07 July 2021

**References**

- Alipourfard N, Nettasinghe B, Abeliuk A, Krishnamurthy V, Lerman K (2020) Friendship paradox biases perceptions in directed networks. *Nat Commun* 11(1):707. <https://doi.org/10.1038/s41467-020-14394-x>
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512. <https://doi.org/10.1126/science.286.5439.509>
- Bianconi G, Darst RK, Iacovacci J, Fortunato S (2014) Triadic closure as a basic generating mechanism of communities in complex networks. *Phys Rev E Stat Nonlinear Soft Matter Phys* 90(4):042806. <https://doi.org/10.1103/PhysRevE.90.042806>
- Bollen J, Gonçalves B, van de Leemput I, Ruan G (2017) The happiness paradox: your friends are happier than you. *EPJ Data Sci* 6(1):1–17. <https://doi.org/10.1140/epjds/s13688-017-0100-1>
- Brot H, Muchnik L, Louzoun Y (2015) Directed triadic closure and edge deletion mechanism induce asymmetry in directed edge properties. *Eur Phys J B* 88(12):1. <https://doi.org/10.1140/epjb/e2014-50220-4>
- Brunson JC (2015) Triadic analysis of affiliation networks. *Network Sci* 3(4):480–508. <https://doi.org/10.1017/nws.2015.38>
- Carayol N, Bergé L, Cassi L, Roux P (2019) Unintended triadic closure in social networks: the strategic formation of research collaborations between French inventors. *J Econ Behav Organ* 163:218–238. <https://doi.org/10.1016/j.jebo.2018.10.009>
- Chen B, Poquet O (2020) Socio-temporal dynamics in peer interaction events. In: Proceedings of the tenth international conference on learning analytics & knowledge, pp 203–208. Association for Computing Machinery, New York, NY, United States. <https://doi.org/10.1145/3375462.3375535>
- Eom Y-H, Jo H-H (2014) Generalized friendship paradox in complex networks: the case of scientific collaboration. *Sci Rep* 4(1):4603. <https://doi.org/10.1038/srep04603>
- Fang Z, Tang J (2015) Uncovering the formation of triadic closure in social networks. In: Proceedings of the 24th international conference on artificial intelligence. IJCAI'15, pp 2062–2068. AAAI Press, Palo Alto, California, USA
- Feld SL (1991) Why your friends have more friends than you do. *Am J Sociol* 96(6):1464–1477
- Fotouhi B, Momeni N, Rabbat MG (2015) Generalized friendship paradox: an analytical approach. In: Aiello LM, McFarland D (eds) *Social informatics*. Springer, Cham, pp 339–352
- Golosovsky M, Solomon S (2017) Growing complex network of citations of scientific papers: modeling and measurements. *Phys Rev E* 95(1):012324. <https://doi.org/10.1103/PhysRevE.95.012324>
- Higham DJ (2018) Centrality-friendship paradoxes: when our friends are more important than us. *J Complex Netw* 7(4):515–528. <https://doi.org/10.1093/comnet/cny029>
- Holme P, Kim BJ (2002) Growing scale-free networks with tunable clustering. *Phys Rev E* 65(2):026107. <https://doi.org/10.1103/PhysRevE.65.026107>
- Huang H, Tang J, Liu L, Luo J, Fu X (2015) Triadic closure pattern analysis and prediction in social networks. *IEEE Trans Knowl Data Eng* 27(12):3374–3389. <https://doi.org/10.1109/TKDE.2015.2453956>
- Huang H, Dong Y, Tang J, Yang H, Chawla NV, Fu X (2018) Will triadic closure strengthen ties in social networks? *ACM Trans Knowl Discov Data* 12(3):1–25. <https://doi.org/10.1145/3154399>
- Itzhack R, Muchnik L, Erez T, Tsaban L, Goldenberg J, Solomon S, Louzoun Y (2010) Empirical extraction of mechanisms underlying real world network generation. *Physica A* 389(22):5308–5318. <https://doi.org/10.1016/j.physa.2010.07.011>
- Jackson MO (2019) The friendship paradox and systematic biases in perceptions and social norms. *J Polit Econ* 127(2):777–818. <https://doi.org/10.1086/701031>
- Jin EM, Girvan M, Newman MEJ (2001) Structure of growing social networks. *Phys Rev E* 64(4):046132. <https://doi.org/10.1103/PhysRevE.64.046132>
- Juhász S, Lengyel B (2017) Creation and persistence of ties in cluster knowledge networks. *J Econ Geogr* 18(6):1203–1226. <https://doi.org/10.1093/jeg/lbx039>
- Lee E, Lee S, Eom Y-H, Holme P, Jo H-H (2019) Impact of perception models on friendship paradox and opinion formation. *Phys Rev E* 99(5):052302. <https://doi.org/10.1103/PhysRevE.99.052302>
- Li M, Zou H, Guan S, Gong X, Li K, Di Z, Lai C-H (2013) A coevolving model based on preferential triadic closure for social media networks. *Sci Rep* 3(1):1. <https://doi.org/10.1038/srep02512>

- Linyi Z, Shugang L (2017) The node influence for link prediction based on triadic closure structure. In: 2017 IEEE 2nd information technology, networking, electronic and automation control conference (ITNEC), pp 761–766. <https://doi.org/10.1109/ITNEC.2017.8284836>
- Louzoun Y, Muchnik L, Solomon S (2006) Copying nodes versus editing links: the source of the difference between genetic regulatory networks and the WWW. *Bioinformatics* 22(5):581–588. <https://doi.org/10.1093/bioinformatics/btk030>
- Mack G (2001) Universal dynamics, a unified theory of complex systems. *Emergence, life and death. Commun Math Phys* 219(1):141–178. <https://doi.org/10.1007/s002200100397>
- Momeni N, Rabbat M (2016) Qualities and inequalities in online social networks through the lens of the generalized friendship paradox. *PLoS ONE* 11(2):0143633. <https://doi.org/10.1371/journal.pone.0143633>
- Muppidi S, Reddy KT (2020) Co-occurrence analysis of scientific documents in citation networks. *Int J Knowl Based Intell Eng Syst* 24(1):19–25. <https://doi.org/10.3233/KES-200025>
- Pal S, Yu F, Novick Y, Bar-Noy A (2019) A study on the friendship paradox-quantitative analysis and relationship with assortative mixing. *Appl Netw Sci* 4(1):71. <https://doi.org/10.1007/s41109-019-0190-8>
- Rapoport A (1953) Spread of information through a population with socio-structural bias: I. The assumption of transitivity. *Bull Math Biophys* 15(4):523–533. <https://doi.org/10.1007/BF02476440>
- Redner S (2004) Citation statistics from more than a century of physical review. *Phys Today* 58:9
- Ren F-X, Shen H-W, Cheng X-Q (2012) Modeling the clustering in citation networks. *Physica A* 391(12):3533–3539. <https://doi.org/10.1016/j.physa.2012.02.001>
- Rossi RA, Ahmed NK (2015) The network data repository with interactive graph analytics and visualization. In: AAAI. <http://networkrepository.com>
- Rozemberczki B, Allen C, Sarkar R (2019a) Multi-scale attributed node embedding. [arXiv: abs/1909.13021](https://arxiv.org/abs/1909.13021)
- Rozemberczki B, Davies R, Sarkar R, Sutton C (2019b) Gemsec: Graph embedding with self clustering. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining. *ASONAM '19*, pp 65–72. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3341161.3342890>
- Sidorov S, Mironov S, Malinskii I, Kadomtsev D (2021) Local degree asymmetry for preferential attachment model. In: *Complex networks & their applications IX*, pp 450–461. Springer, Cham. [https://doi.org/10.1007/978-3-030-65351-4\\_36](https://doi.org/10.1007/978-3-030-65351-4_36)
- Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021. <https://doi.org/10.1126/science.1177170>
- Song T, Tang Q, Huang J (2019) Triadic closure, homophily, and reciprocation: an empirical investigation of social ties between content providers. *Inf Syst Res* 30(3):912–926. <https://doi.org/10.1287/isre.2019.0838>
- Souza TTP, Aste T (2019) Predicting future stock market structure by combining social and financial network information. *Physica A* 535:122343. <https://doi.org/10.1016/j.physa.2019.122343>
- Stolov Y, Idel M, Solomon S (2000) What are stories made of?—Quantitative categorical deconstruction of creation. *Int J Mod Phys C* 11(04):827–835. <https://doi.org/10.1142/S0129183100000699>
- Wharrie S, Azizi L, Altmann EG (2019) Micro-, meso-, macroscales: the effect of triangles on communities in networks. *Phys Rev E* 100(2):1. <https://doi.org/10.1103/PhysRevE.100.022315>
- Wu Z-X, Holme P (2009) Modeling scientific-citation patterns and other triangle-rich acyclic networks. *Phys Rev E* 80(3):037101. <https://doi.org/10.1103/PhysRevE.80.037101>
- Yang J, Leskovec J (2013) Defining and evaluating network communities based on ground-truth. *Knowl Inf Syst* 42(1):181–213. <https://doi.org/10.1007/s10115-013-0693-z>
- Yin H, Benson AR, Leskovec J (2019) The local closure coefficient: A new perspective on network clustering. In: *WSDM'19: Proceedings of the twelfth ACM international conference on web search and data mining*, pp 303–311. ACM, New York, NY, USA. <https://doi.org/10.1145/3289600.3290991>
- Zhou L-k, Yang Y, Ren X, Wu F, Zhuang Y (2018) Dynamic network embedding by modeling triadic closure process. In: *The thirty-second AAAI conference on artificial intelligence (AAAI-18)*, pp 571–578

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---