

Research

Open Access



A study on the friendship paradox – quantitative analysis and relationship with assortative mixing

Siddharth Pal¹, Feng Yu^{2*} , Yitzchak Novick^{2,3}, Ananthram Swami⁴ and Amotz Bar-Noy²

*Correspondence:

fyu@gradcenter.cuny.edu

²City University of New York, 10016
New York, USA

Full list of author information is
available at the end of the article

Abstract

The friendship paradox is the observation that friends of individuals tend to have more friends or be more popular than the individuals themselves. In this work, we first study local metrics to capture the strength of the paradox and the direction of the paradox from the perspective of individual nodes, i.e., an indication of whether the individual is more or less popular than its friends. These local metrics are aggregated, and global metrics are proposed to express the phenomenon on a network-wide level. Theoretical results show that the defined metrics are well-behaved enough to capture the friendship paradox. We also theoretically analyze the behavior of the friendship paradox for popular network models in order to understand regimes where friendship paradox occurs. These theoretical findings are complemented by experimental results on both network models and real-world networks. By conducting a correlation study between the proposed metrics and degree assortativity, we experimentally demonstrate that the phenomenon of the friendship paradox is related to the well-known phenomenon of assortative mixing.

Keywords: Friendship paradox, Assortativity, Assortative mixing, Network analysis, Network models, Homophily

Introduction

Topologies of complex networks have been known to exhibit non-trivial heterogeneous properties such as heavy-tailed degree distributions (Barabási and Albert 1999) and assortative mixing (Newman 2002), to name a few. A similar non-trivial network heterogeneity is captured by the phenomenon of the friendship paradox. The friendship paradox was first studied by Scott Feld in the context of social networks (Feld 1991). It states that the average number of friends of the collection of friends of individuals in a social network will be higher than the average number of friends of the collection of the individuals themselves. The phenomenon results from a sampling bias that causes nodes to be counted in proportion to their degree, and it extends beyond popularity to other individual traits such as amount of content in social networks (Hodas et al. 2013), number of coauthors, citations and publications in scientific collaboration networks (Eom and Jo 2014), etc. This is known as the generalized friendship paradox and has been attributed to a correlation between these desirable traits and degree (Eom and Jo 2014; Fotouhi et al. 2014) although it has been proven to exist in absence of a significant correlation as well

(Momeni and Rabbat 2016). The phenomenon has a measurable impact on people's psyches because people tend to evaluate success in regard to a given trait through comparison to the social network around them (Bollen et al. 2017; Jackson 2016); an overshadowing effect of a collection of nodes on another is significant. The friendship paradox has been studied in the context of both online (Bagrow et al. 2017; Bollen et al. 2017; Hodas et al. 2013) and offline social networks (Feld 1991). The phenomenon is described as a global, network-wide characteristic. It is natural to ask if the friendship paradox holds at a more local level, e.g., at the level of individual nodes in a network (Hodas et al. 2013; Momeni and Rabbat 2016). In this work, we mathematically quantify the friendship paradox so as to understand its strength both locally and globally in the network.

Inspired by the friendship paradox, we first study a local network metric for a node, which we call the "friendship index" (FI). For a vertex v , it is defined as $FI(v) = \frac{\sum_{(v,u) \in E} d_u}{d_v^2}$, which is the ratio of the average degree of the neighbors of the vertex to its own degree. FI captures two significant characteristics related to the friendship paradox. First, it captures the 'direction of influence', i.e., it indicates whether, in accordance with the friendship paradox, the vertex is outperformed by its neighbors in popularity, or if it is one of the vertices for which the paradox does not hold because it is more popular than its neighbors. Second, it quantifies the 'discrepancy' compared to a scenario where the friendship paradox does not exist; FI values further from 1 indicate an extreme manifestation of the paradox whereas FI values closer to 1 would indicate the vertex experiences the paradox less severely. Notice that this measure is truly local. The node's FI value is determined entirely by its own degree and the degrees of its neighbors, the structure of network as a whole does not influence a node's FI . This distinguishes FI from local assortativity (Thedchanamoorthy et al. 2014; Piraveenan et al. 2010), a node metric that captures a relative value normalized by other such values in the network. Having defined the local metric, we propose aggregating the indices over all vertices in order to quantify the strength of the overall network-wide phenomenon of friendship paradox by taking an arithmetic mean, the log of the geometric mean, and the harmonic mean. We explore these measures theoretically and experimentally, noting their ability to capture network characteristics. We compare and contrast them and ultimately find value in the arithmetic mean and the log of the geometric mean, while discarding the harmonic mean because drastically different graphs have equivalent harmonic means. Consistent with the nature of the friendship paradox, we prove a significant lower bound on these means that indicates the average FI is always greater than or equal to 1, meaning nodes feel outperformed by their neighbors on average. Because of this bound, we can say the aggregates indicate a stronger occurrence of the friendship paradox as they increase. Unlike FI , where the distance from 1 can be in either the positive or negative direction, the aggregate measures only increase with the strength of the paradox. We explore real-world networks that are categorized by their type, and find some consistencies in our aggregate measures within the individual categories.

A number of existing works have analysed the friendship paradox from different perspectives. The comparison between a node's degree to both the mean and median of neighbors' degrees has been used in the literature, primarily as a binary measure of whether a node is more or less popular compared to its neighbors (Jo and Eom 2014; Momeni and Rabbat 2016; Momeni and Rabbat 2018; Lee et al. 2019), while some (Hodas et al. 2013; Jackson 2016) have also used its absolute value to indicate severity of

the paradox. Hodas et al. (2013) uses the binary measure obtained through the mean of neighbor's degrees to investigate the friendship paradox in Twitter network. Jo and Eom (2014) characterized the paradox holding probability of individual nodes, and studied its behavior for network models with tunable degree-degree and degree-attribute correlations. A significant finding was that the paradox holding probability may depend on the assortativity of the network, which acts as a strong motivation for our work. Momeni and Rabbat (2016) used both the mean and median based binary measures to investigate the prevalence of friendship paradox in Twitter network. Lee et al. (2019) studied three perception models of the friendship paradox - mean-based, median-based and fraction-based binary measures for the friendship paradox aggregated across the entire network. These measures were analysed in configuration models with tunable assortativity, and the impact of the perception model was studied on opinion formation. Our contribution has been to investigate the measure based on the mean of neighbors' degrees as a network metric by studying its properties in popular network models, and illustrating its relationship with local and global measures of assortativity.

The phenomenon of assortativity, or assortative mixing captures the preference for a network's nodes to attach to others that are similar in some way. In this work, we focus on degree assortativity (Newman 2002) where the measure of similarity is in terms of the nodes' degrees. Although assortativity was proposed as a global measure, local versions have also been proposed more recently (Piraveenan et al. 2010; Thedchanamoorthy et al. 2014). It is worth mentioning that these local assortativity measures are examples of node measures that are not truly local, because they are influenced by the global assortativity of the graph. Jo and Eom (2014), Lee et al. (2019), and Momeni and Rabbat (2018) conduct studies on networks that vary in assortativity and explore its impact on the friendship paradox. In this work, we further investigate the relationship between the FI of nodes and their local assortativity measures. We observe that *FI* values close to 1.0 indicate strong local assortativity, while *FI* values further from 1.0 indicate weak assortativity. We also explore the relationship between our network-wide aggregate measures and network assortativity. Similar to the results of the local measures, when the aggregate values are extreme they indicate disassortativity and when they are moderate they indicate assortativity. These consistencies between the measures are revealed through theoretical arguments and experimental results. We conclude that the friendship index captures some information about assortative mixing in the network. We consider canonical graphs and common graph models and find our aggregate measures produce well behaved functions that smoothly reflect small adjustments in the graphs where the function of network assortativity has a far less smooth curve. The present work is an extension of our previous work (Pal et al. 2018); here we present complete and detailed proofs of results partially proven in Pal et al. (2018) along with new theoretical findings on the proposed metrics in networks models, such as Erdos-Renyi and Barabasi-Albert graphs. Additionally, a more thorough simulation study on network models is presented here.

The paper is organized as follows. In "Preliminary" section we define the local and global notions of friendship index and assortativity. These metrics are then studied for various canonical graphs such as regular graphs, star graphs and complete bipartite graphs. In "Theoretical results on friendship index" section, we highlight some theoretical

results of the friendship index, followed by experimental results on network models and real networks in “[Experimental results](#)” section.

Preliminary

Friendship index

Consider a graph $G = (V, E)$, with V and $E \subseteq V \times V$ being the set of nodes and edges respectively. For each node $i \in V$, let d_i denote the degree of node i and S_i denote the sum of i 's neighbors' degrees. Therefore, we have

$$S_i = \sum_{v \in N_i} d_v,$$

where N_i denotes the set of neighbors of node i . The friendship index (FI) of node i is defined as

$$\begin{aligned} FI(i) &= \frac{S_i}{d_i^2}, \text{ if } d_i > 0 \\ &= 1, \text{ if } d_i = 0. \end{aligned} \quad (1)$$

The friendship index of a node is the average degree of its neighbors normalized by its own degree. Therefore, the FI of a node will be less than 1 if its own degree is larger in comparison to its neighbors' degrees, and greater than 1 otherwise.

Another way to understand the friendship index is through the uniform and neighborhood node sampler (Leskovec and Faloutsos 2006; Momeni and Rabbat 2018). For a graph $G = (V, E)$, we define the following node sampling techniques:

a) Uniform sampler ρ - A uniformly sampled node is chosen uniformly at random among the nodes in the graph. Therefore, $\rho \sim \mathcal{U}(V)$.

b) Edge sampler σ - An edge is sampled uniformly at random, then one of its endpoints are chosen uniformly at random. For an uniformly chosen edge $(\sigma_1, \sigma_2) \sim \mathcal{U}(E)$, $\sigma \sim \mathcal{U}(\sigma_1, \sigma_2)$.

c) Neighborhood sampler μ - First define $\mu(i) \sim \mathcal{U}(N_i)$ to be a node sampled uniformly at random among the neighbors of node i . If node i has no neighbors then set $\mu = i$. Next, we define $\mu(\rho)$ to be a node that is sampled randomly from the neighbors of an uniformly sampled node ρ . Operationally, first sample $\rho \sim \mathcal{U}(V)$ and $\mu \sim \mathcal{U}(N_\rho)$; However if $N_\rho = \emptyset$, then set $\mu = \rho$.

We can equivalently define the friendship index as follows,

$$\begin{aligned} FI(i) &= \frac{\mathbb{E}[d_{\mu(i)}]}{d_i}, \text{ if } d_i > 0 \\ &= 1, \text{ if } d_i = 0. \end{aligned} \quad (2)$$

It is easy to see that the friendship index for regular graphs will always be 1 because all degrees are the same. This is a local network metric concerning a particular node. We could either aggregate the local measure to obtain a global measure or directly work with the local FIs $\{FI(i), i \in V\}$.

The local measures of the friendship paradox can be extended to the entire network in different ways. The following global notions of FI can be defined for a given graph G :

- 1 Arithmetic FI (AFI) - We define AFI as the average, or arithmetic mean, of all the local FIs

$$AFI(G) = \frac{1}{|V|} \sum_{i \in V} FI(i) \quad (3)$$

Observe that

$$AFI(G) = \frac{1}{|V|} \sum_{i \in V} \frac{\mathbb{E}[d_{\mu(i)}]}{d_i} = \mathbb{E} \frac{d_{\mu(\rho)}}{d_\rho} \quad (4)$$

with the caveat that $\frac{\mathbb{E}d_i}{d_i}$ is set to 1 if $d_i = 0$. Therefore, AFI is the expected ratio between the degrees of a neighborhood sampled node $\mu(\rho)$ and that of the uniformly sampled node ρ .

- 2 Geometric FI (GFI) - This metric is defined as the logarithm of the geometric mean of all the local FIs

$$GFI(G) = \frac{1}{|V|} \sum_{v \in V} \log FI(v)$$

While the linearity of expectation allows the alternative representation for AFI through the notion of sampling, we do not have an equivalent expression for GFI. This leads to greater mathematical tractability in analysing AFI of network models compared to GFI or other binary measures based on individual nodes involving Heaviside step functions. However, GFI has other benefits over AFI that will become apparent later.

- 3 Harmonic FI (HFI) - This metric is defined as the harmonic mean of all the local FIs

$$HFI(G) = \frac{1}{\frac{1}{|V|} \sum_{v \in V} \frac{1}{FI(v)}}$$

Again for HFI, we do not have an alternative representation like (4).

We will give theoretical results on AFI and experimental results on both AFI and GFI. In what follows we explain the relative advantages of considering AFI and GFI.

Global and local assortativity

Newman (2002) defines assortativity as a measure of the similarity of degree in adjacent vertices.¹ Newman formally quantifies assortativity as a Pearson correlation coefficient of the degrees of the two vertices attached to every edge. This can be expressed as

$$r = \frac{\mathbb{E}[d_{\sigma_1} d_{\sigma_2}] - \mathbb{E}[d_\sigma]^2}{\mathbb{E}[d_\sigma]^2 - \mathbb{E}[d_\sigma]^2} \quad (5)$$

where (σ_1, σ_2) is a uniformly sampled edge and σ is a node sampled according to the random edge sampling. This difference in sampling nodes introduces inherent differences between the friendship index and assortativity.

The assortativity can also be defined in terms of the network degree distribution \mathbf{p} , excess or remaining degree distribution \mathbf{q} and the link distribution $e_{j,k}$. Note that the distribution \mathbf{p} is the degree distribution of a uniformly sampled node ρ . The distribution \mathbf{q} is related to the excess degree of a node arrived through the random edge sampling procedure, given by

¹This is a specific type of homophily, using degree as a trait.

$$q(k) = \frac{(k+1)p(k+1)}{\sum_{j=1}^{|V|} jp(j)}. \quad (6)$$

We also define the joint probability distribution of the remaining degrees of the two nodes on either end of a randomly chosen link as $e_{j,k}$. We can equivalently define assortativity as

$$r = \frac{1}{\sigma_q^2} \left[\sum_{jk} jk (e_{j,k} - q(j)q(k)) \right] \quad (7)$$

where σ_q is the standard deviation of the distribution \mathbf{q} .

While this definition of assortativity is at the global level, several efforts have been made to quantify the assortativity of the network at a local level (Piraveenan et al. 2010; Thedchanamoorthy et al. 2014). Specifically, we first consider the definition of local assortativity presented by Piraveenan et al. (2010). This definition of local assortativity, which we call the p -local assortativity is defined as

$$r_p(i) = \frac{d_i \left((d_i - 1)\bar{k} - \mu_q^2 \right)}{2|E|\sigma_q^2}, \quad i \in V \quad (8)$$

where, \bar{k} is the average remaining degree of the node's neighbors, μ_q and σ_q are the mean and standard deviation of the distribution \mathbf{q} . It also follows that the p -local assortativities of all the nodes sum up to the network assortativity r .

Note that $r_p(i)$ is positive and large if $(d_i - 1)\bar{k} \gg \mu_q^2$. This happens when a high degree node is connected to other nodes which have relatively high degree or if a node with degree in a medium range has neighbors with large degrees. And, $r_p(i)$ is large in magnitude and negative if $(d_i - 1)\bar{k} \ll \mu_q^2$, which can happen when a node with low degree is connected to low degree nodes. While in such cases the local assortativity should be positive, the p -local assortativity will have a negative value.

We define another notion of local assortativity due to Thedchanamoorthy et al. (2014), which has been argued to be closer to the fundamental notion of assortativity compared to that proposed by Piraveenan et al. (2010). This concurs with our analysis in the previous paragraph that the p -local assortativity does not always capture what a local measure of assortativity should. For these reasons, we only show results for the local assortativity due to Thedchanamoorthy et al.

The local assortativity measure calculates the 'average neighbor difference,' a direct indicator of a node's dissortativity, given by

$$\delta_i = \frac{1}{d_i} \sum_{v \in N_i} |d_i - d_v|, \quad (9)$$

which is then scaled by the sum of the neighbor differences across all the nodes, to obtain the normalized neighbor differences $\bar{\delta}_i$. We therefore have,

$$\bar{\delta}_i = \frac{\delta_i}{\sum_{j \in V} \delta_j}.$$

The final assortativity of a node, which we call the T-local assortativity (TLA) is obtained by

$$r_T(i) = \lambda - \bar{\delta}_i \quad (10)$$

where $\lambda = \frac{r+1}{|V|}$, such that the sum of all the local assortativities yield the global assortativity r .

Observe that if δ_i is small, then the degrees of node i and its neighbors are very close. Therefore, the local measure $FI(i)$ should be close to 1. However, if δ_i is large then at least one neighbor of i has degree very different from i . Therefore, it is expected that the $FI(i)$ would be far from 1. If it is much less than 1, then the degree of i is more than its neighbors' on average; while if it is greater than 1, then the degree of i is less than its neighbors' on average. Therefore FI captures 'direction of influence' in scenarios of local disassortativity. If the TLA at node i is close to 1, i.e., degree of node i is similar to its neighbors, then $FI(i)$ would be close to 1; and if TLA is small and close to 0, i.e., the degree of node i is significantly different from its neighbors, then $FI(i)$ is expected to be away from 1. Therefore, $|FI(i) - 1|$ is expected to be anticorrelated with TLA, i.e., as $|FI(i) - 1|$ increases, TLA is expected to reduce.

Results for special graphs

- 1 ***d*-regular graphs** - Since all nodes have the same degree d , $FI = 1$ for all nodes. Therefore $AFI = 1$, $GFI = 0$ and $HFI = 1$. A clique of n nodes is a $n - 1$ -regular graph. Thus, the result also holds for cliques. Therefore, for d -regular graphs where the friendship paradox does not occur, we get the minimum values for AFI and GFI (see Theorems 1 and 2). Also, the assortativity $r = 1$ for regular graphs, and all local assortativity values will be positive and equal ($r_T = \frac{2}{n}$).
- 2 **Star graphs** - Consider a star graph with n nodes. The FI for the center of the star will be $\frac{1}{n-1}$, and $n - 1$ for all leaves. One can show that the FI values for the center and the leaves of a star are the maximum and minimum possible values for a graph with the given number of nodes. Thus the friendship paradox is rather severe for a star graph. We calculate, $AFI = n - 2 + \frac{1}{n-1}$, $GFI = \frac{n-2}{n} \log n$ and $HFI = 1$. An advantage of considering GFI is that the maximum and minimum possible values of FIs have the same absolute value in the logarithmic scale. Surprisingly, HFI for a star is the same as that of a d -regular graph. The assortativity $r = -1$, while all the local assortativities r_T will be the minimum value of -1 .
- 3 **Complete bipartite graphs** - Consider a bipartite graph with x nodes on one side and y nodes on the other. This is a generalization of the star graph ($x = 1, y = n$). Let us denote this as *Bipartite*(x, y), with the set of x nodes on one side being denoted as X , and the set of y nodes on the other side denoted as Y . For vertex $v \in X$, $FI(v) = \frac{x}{y}$ and for $v \in Y$, $FI(v) = \frac{y}{x}$. Therefore, if $x > y$ then nodes in X will have $FI > 1$ and nodes in Y will have $FI < 1$. The $AFI = 1 + \frac{(x-y)^2}{xy}$ which is always greater than 1 as long as $x \neq y$. Furthermore, $\frac{\partial AFI}{\partial y} = \frac{(y^2 - x^2)}{xy^2}$ implying that $\frac{\partial AFI}{\partial y} > 0$ for $y > x$ and $\frac{\partial AFI}{\partial y} < 0$ for $y < x$. Therefore, with x kept constant, the AFI increases when y is varied away from x . We also have $GFI = \frac{1}{x+y} \log \left(\frac{x^x y^y}{x^y y^x} \right) > 1$ for $x \neq y$. We calculate, $\frac{\partial GFI}{\partial y} = \frac{2xy \log(\frac{y}{x}) + (y^2 - x^2)}{(y+x)^2 y}$, which again implies that the GFI will increase as y is varied away from x . Therefore, both AFI and GFI are well-behaved for the class of complete bipartite graphs in that both of the measures

increase as y is varied away from x indicating more prominence of the friendship paradox. Interestingly, for $x \neq y$, assortativity $r = -1$, while for $x = y$ assortativity $r = 1$ by definition, indicating that the variation in assortativity is rather sharp. We also compute $HFI = 1$ for all positive values of x and y .

We find HFI to be equal to 1 for widely different classes of graphs ranging from d -regular graphs to star graphs. Therefore in later sections, we will only continue investigation into AFI and GFI owing to their more well-behaved nature. Also from the star graph, we observed that in the logarithmic scale, the maximum and minimum values of the FI are equidistant from 0, the $\log FI$ value where friendship paradox is not observed. Therefore in the “[Experimental results](#)” section, we study the correlation between $|\log FI|$ and TLA, because $|\log FI|$ better captures the deviation from the scenario where friendship paradox does not occur.

Theoretical results on friendship index

A lower bound on global measures of fi

We first state a simple result that will be used in this section.

Lemma 1 *For any positive number x, y and monotonically increasing functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$ such that $f(x), g(x) \geq 0$ for $x \geq 1$, we have*

$$\frac{f(x)}{g(y)} + \frac{f(y)}{g(x)} \geq \frac{f(x)}{g(x)} + \frac{f(y)}{g(y)} \quad (11)$$

Proof For $x, y \geq 1$, we have

$$\begin{aligned} & \left(\frac{f(x)}{g(y)} + \frac{f(y)}{g(x)} \right) - \left(\frac{f(x)}{g(x)} + \frac{f(y)}{g(y)} \right) \\ &= \frac{(f(x) - f(y))(g(x) - g(y))}{g(x)g(y)} \geq 0. \end{aligned} \quad (12)$$

□

With $f(x) = x$, and $g(x) = x^2$, the above lemma leads to the following corollary.

Corollary 1 *For any edge $e(i, j)$, we have $\frac{d_i}{d_j^2} + \frac{d_j}{d_i^2} \geq \frac{1}{d_i} + \frac{1}{d_j}$*

Theorem 1 *For all graphs G , $AFI(G) \geq 1$. $AFI(G) = 1$ only for the class of graphs where each connected component is a regular graph.*

Proof For a graph G , we have

$$\begin{aligned} AFI(G) &= \frac{1}{n} \sum_{i \in V} \frac{S_i}{d_i^2} \\ &= \frac{1}{n} \sum_{i \in V} \sum_{j \in N_i} \frac{d_j}{d_i^2} \end{aligned} \quad (13)$$

Observe that for each edge (i, j) , the term $\frac{d_j}{d_i^2}$ appears when summing over i and $\frac{d_i}{d_j^2}$ appears when summing over j . Therefore, rather than the double sum in (13), we can have

$$\begin{aligned} AFI(G) &= \frac{1}{n} \sum_{(i,j) \in E} \left(\frac{d_j}{d_i^2} + \frac{d_i}{d_j^2} \right) \\ &\geq \frac{1}{n} \sum_{(i,j) \in E} \left(\frac{1}{d_i} + \frac{1}{d_j} \right) \quad (\text{Using Corollary 1}) \end{aligned} \quad (14)$$

$$= \frac{1}{n} \sum_{i \in V} \frac{d_i}{d_i} = 1 \quad (15)$$

where the last step follows by noticing that for each node i , the term $\frac{1}{d_i}$ appears for exactly d_i edges. Also, equality in (14) occurs if and only if all the nodes on either side of an edge have the same degree, i.e., all nodes in a connected component have the same degree. \square

With $f(x) = \log x$, and $g(x) = x$, Lemma 1 leads to the following corollary.

Corollary 2 For any edge $e(i, j)$, we have $\frac{\log d_i}{d_j} + \frac{\log d_j}{d_i} \geq \frac{\log d_i}{d_i} + \frac{\log d_j}{d_j}$

Theorem 2 For all graphs G , $GFI(G) \geq 0$. $GFI(G) = 0$ only for the class of graphs where each connected component is a regular graph.

Proof Using the expression for GFI, we obtain for a graph G

$$\begin{aligned} GFI(G) &= \frac{1}{n} \sum_{i \in V} \log \frac{S_i}{d_i^2} \\ &= \frac{1}{n} \sum_{i \in V} \log \frac{1}{d_i} + \frac{1}{n} \sum_{i \in V} \log \frac{S_i}{d_i} \\ &= \frac{1}{n} \sum_{i \in V} \log \frac{1}{d_i} + \frac{1}{n} \sum_{i \in V} \log \left(\frac{1}{d_i} \sum_{j \in N_i} d_j \right). \end{aligned} \quad (16)$$

Applying Jensen's inequality in (16), we obtain

$$\begin{aligned} GFI(G) &\geq \frac{1}{n} \sum_{i \in V} \log \frac{1}{d_i} + \frac{1}{n} \sum_{i \in V} \sum_{j \in N_i} \frac{1}{d_i} \log d_j \\ &= \frac{1}{n} \sum_{i \in V} \log \frac{1}{d_i} + \sum_{(i,j) \in E} \left(\frac{1}{d_i} \log d_j + \frac{1}{d_j} \log d_i \right) \\ &\geq \frac{1}{n} \sum_{i \in V} \log \frac{1}{d_i} + \sum_{(i,j) \in E} \left(\frac{1}{d_i} \log d_i + \frac{1}{d_j} \log d_j \right) \quad (\text{Using Corollary 2}) \\ &= \frac{1}{n} \sum_{i \in V} \log \frac{1}{d_i} + \frac{1}{n} \sum_{i \in V} \log d_i = 0 \end{aligned} \quad (17)$$

where the last step is obtained by noticing that for each node i , the term $\frac{1}{d_i} \log d_i$ appears for exactly d_i edges. Again, equality occurs if and only if both nodes on either side of an edge have the same degree. \square

Theorem 1 has been proven by Jackson (2016)(Lemma 1), while Theorem 2 is a new result. Theorems 1 & 2 show that both the AFI and GFI will be at their minimum values for the class of regular graphs, where the friendship paradox does not occur. For any other graph, the AFI and GFI will be strictly greater than their minimum values, thereby

suggesting the occurrence of the friendship paradox due to an imbalance between neighbor degrees. This is in agreement with the observation due to Feld (1991), where he noted that the mean number of friends of friends is always greater than the mean number of friends, i.e.,

$$\frac{\sum_i d_i^2}{\sum_i d_i} \geq \frac{1}{n} \left(\sum_i d_i \right) \quad (18)$$

with equality taking place in (18) only for regular graphs where the empirical degree distribution has zero variance. This demonstrates a well-behaved property for both of the proposed global metrics.

Erdos-Renyi graphs

We study the behavior of the friendship index for the class of Erdos-Renyi (ER) graphs (Erdos and Rényi 1960), a basic graph model for capturing random connections in networks. Consider a sequence of ER graphs $\{\mathbb{G}_1, \mathbb{G}_2, \dots\}$, such that $\mathbb{G}_n = (V_n, \mathbb{E}_n)$ with V_n being the vertex set $\{1, 2, \dots, n\}$, and \mathbb{E}_n being the random edge set with every edge occurring with probability p . We define the following notation: For nodes i and j , let $\{i \sim j\}$ denote the event of i being connected to j . Let $D_{n,i}$ and $\mathcal{N}_{n,i}$ denote the degree and the set of neighbors of node i in \mathbb{G}_n . In a random graph setting, the expected friendship index of node i in \mathbb{G}_n can be expressed as

$$\mathbb{E}[FI_n(i)] = \mathbb{E} \left[\frac{S_{n,i}}{D_{n,i}^2} \mathbf{1}[D_{n,i} > 0] + \mathbf{1}[D_{n,i} = 0] \right]. \quad (19)$$

Although we will primarily focus on the expected behavior of the FI of a node, it is worth noting that the AFI will exhibit similar behavior, because

$$\mathbb{E}[AFI_n] = \mathbb{E} \left[\frac{1}{n} \sum_{v \in V} FI_n(v) \right] = \mathbb{E}[FI_n(i)], \quad i \in V. \quad (20)$$

with, AFI_n being the AFI of graph \mathbb{G}_n .

Case 1: np constant

Considering the first term in (19), we have

$$\begin{aligned} \mathbb{E} \left[\frac{S_{n,i}}{D_{n,i}^2} \mathbf{1}[D_{n,i} > 0] \right] &= \mathbb{E} \left[\mathbf{1}[D_{n,i} > 0] \frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right] \\ &= \mathbb{E} \left[\mathbf{1}[D_{n,i} > 0] \frac{1}{D_{n,i}^2} \mathbb{E} \left[\sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \mid D_{n,i} = d_i \right] \right] \\ &= \mathbb{E} \left[\mathbf{1}[D_{n,i} > 0] \frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} \mathbb{E} \left[1 + \sum_{k \in V \setminus \{i,j\}} \mathbf{1}[j \sim k] \mid D_{n,i} = d_i \right] \right] \\ &= \mathbb{E} \left[\mathbf{1}[D_{n,i} > 0] \frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} \mathbb{E} \left[1 + \sum_{k \in V \setminus \{i,j\}} \mathbf{1}[j \sim k] \right] \right] \\ &= (1 + (n-2)p) \mathbb{E} \left[\frac{1}{D_{n,i}} \mathbf{1}[D_i > 0] \right]. \end{aligned} \quad (21)$$

As $n \rightarrow \infty$, $D_{n,i} \Rightarrow_d \Pi(c)$ under the regime $np \rightarrow c$. We know that if $X_n \Rightarrow_d X$, then $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ for bounded Lipschitz functions f . The mapping $f: x \rightarrow \frac{1}{x}$, $x \in \mathbb{N}$ is bounded by 1, and Lipschitz because $|f(x) - f(y)| = \left| \frac{1}{x} - \frac{1}{y} \right| \leq |x - y|$. Using this result in (21), we obtain

$$\mathbb{E}[FI_n(i)] \rightarrow (1+c)\mathbb{E}\left[\frac{1}{\Pi(c)}\mathbf{1}[\Pi(c) > 0]\right] + \mathbf{P}[\Pi(c) = 0] \quad (22)$$

under the regime $np \rightarrow c$, with $c > 0$. We set the limit as

$$\begin{aligned} \beta(c) &= (1+c)\mathbb{E}\left[\frac{1}{\Pi(c)}\mathbf{1}[\Pi(c) > 0]\right] + \mathbf{P}[\Pi(c) = 0] \\ &= (1+c)\sum_{d=1}^{\infty} \frac{1}{d}e^{-c}\frac{c^d}{d!} + e^{-c}. \end{aligned} \quad (23)$$

Simple bounds on Friendship Index: Using (22), we first try bounding the FI, since the expression cannot be exactly computed.

Lemma 2 For sequence of Erdos-Renyi graphs $\{\mathbb{G}_n, n = 1, 2, \dots\}$

$$\max\left(1, e^{-c}\left[1 + (1+c)\left(c + \frac{c^2}{4}\right)\right]\right) \leq \lim_{n \rightarrow \infty} \mathbb{E}[FI_n(i)] \leq 1 + c - ce^{-c} \quad (24)$$

under the regime $np \rightarrow c$.

Proof Using the limiting expression of FI, we obtain for $c > 0$

$$\begin{aligned} (1+c)\mathbb{E}\left[\frac{1}{\Pi(c)}\mathbf{1}[\Pi(c) > 0]\right] + \mathbf{P}[\Pi(c) = 0] &= (1+c)\sum_{d=1}^{\infty} \frac{1}{d}e^{-c}\frac{c^d}{d!} + e^{-c} \\ &\leq (1+c)\sum_{d=1}^{\infty} e^{-c}\frac{c^d}{d!} + e^{-c} = 1 + c - ce^{-c}. \end{aligned} \quad (25)$$

For obtaining the lower bound we only use the first term from the infinite sum

$$(1+c)\mathbb{E}\left[\frac{1}{\Pi(c)}\mathbf{1}[\Pi(c) > 0]\right] + \mathbf{P}[\Pi(c) = 0] \geq (1+c)\left(ce^{-c} + \frac{c^2}{4}e^{-c}\right) + e^{-c}, \quad (26)$$

and also note that $\mathbb{E}[FI_n(i)] \geq 1$ since $\mathbb{E}[FI_n(i)] = \mathbb{E}[AFI_n] \geq 1$ using Theorem 1. \square

Theorem 3 For a sequence of Erdos-Renyi graphs $\{\mathbb{G}_n, n = 1, 2, \dots\}$ under the regime $np \rightarrow c$, and node i , we have (a) $\lim_{n \rightarrow \infty} \mathbb{E}[FI_n(i)] \rightarrow 1$ as $c \rightarrow 0$; (b) $\lim_{n \rightarrow \infty} \mathbb{E}[FI_n(i)] \rightarrow 1$ as $c \rightarrow \infty$; (c) there exists $\delta > 0$ such that for all $c \in (0, \delta)$, $\lim_{n \rightarrow \infty} \mathbb{E}[FI_n(i)] > 1$.

Proof Part (a) of Theorem 3 can be proved using the upper bound of Lemma 2 and also by using the lower bound $\mathbb{E}[FI_n(i)] = \mathbb{E}[AFI_n] \geq 1$.

Lemma 2 does not help in determining the behavior of β for c large, because the upper and lower bounds diverge as c increases. Therefore, we study the behavior of $\beta: \mathbb{R} \rightarrow \mathbb{R}$

over the range of c analytically. For $c \in (0, C)$, with $C > 0$, the infinite sum $\beta(c)$ converges uniformly on the specified range. Thus, differentiating $\beta(c)$ w.r.t. c yields

$$\begin{aligned}\beta'(c) &= -c \sum_{d=1}^{\infty} \frac{1}{d} e^{-c} \frac{c^d}{d!} + \frac{1+c}{c} (e^c - 1) e^{-c} - e^{-c} \\ &= -\frac{c}{1+c} \beta(c) + \left(1 + \frac{1}{c}\right) (1 - e^{-c}) - e^{-c}\end{aligned}\quad (27)$$

For c large and $\beta(c) > 1$, we have $\beta'(c) < 0$. Furthermore for c large, $\beta'(c) \approx -\beta(c) + 1$ implying that $\lim_{c \rightarrow \infty} \beta(c) = 1$. This to Theorem 3(b).

Part(c) follows from the lower bound in Lemma 2 and observing that the lower bound is greater than 1 for $c \in (0, \delta)$ (for some $\delta > 0$).

□

Therefore, the result indicates that for c small, the effect of the friendship paradox is weak when measured in terms of AFI. This is probably because the graph is so sparse that there are a lot of isolated nodes and edges, and the paradox does not occur for the two entities. This is true for c large as well because the graph becomes very dense. Also, from Theorem 3(c), we prove that the AFI is strictly greater than 1 for a certain range of parameter c suggesting that the friendship paradox is observed even in large Erdos-Renyi graphs.

Note that the average degree of the ER graph in this regime is $(n-1)p$ which converges to c . Theorem 3 can also be thought to imply that in random ER graphs, as the average degree of the graph becomes small or gets large, the friendship paradox becomes weaker. For c large, the growth in every node's degree and their neighbors' degrees are such that the friendship paradox is not observed at any chosen node. This is rather intuitive, by recalling that FI is defined as the ratio between the degree of a node's random neighbor and its own degree. If the degrees get large in an ER graph where the degree rvs are uncorrelated, then the sampling bias is expected to be small, which is exactly what we observe.

Case 2: p constant

We also address the question whether the friendship paradox is observed when the probability of connection p is kept constant as n goes to infinity. We return to the expression of the friendship index

$$FI_n(i) = \left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] + \mathbf{1}[D_{n,i} = 0] \quad (28)$$

As n gets large, the second term in (28) behaves as

$$\mathbf{1}[D_{n,i} = 0] \longrightarrow_n 0 \text{ a.s.} \quad (29)$$

We define the edge rvs as $\chi_{i,j}$, where $\chi_{i,j} = 1$ if the edge exists between nodes i and j , 0 otherwise. We write the first expression in (28) as follows

$$\begin{aligned}
& \left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \\
&= \frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} \left[1 + \sum_{\ell \in V \setminus \{i,j\}} \chi_{\ell,j} \right] \mathbf{1}[D_{n,i} > 0] \\
&= \frac{1}{D_{n,i}} \mathbf{1}[D_{n,i} > 0] + \left[\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} \sum_{\ell \in V \setminus \{i,j\}} \chi_{\ell,j} \right] \mathbf{1}[D_{n,i} > 0] \\
&= \frac{1}{D_{n,i}} \mathbf{1}[D_{n,i} > 0] + \left[\frac{n}{D_{n,i}} \cdot \frac{1}{n D_{n,i}} \sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j} \right] \mathbf{1}[D_{n,i} > 0], \tag{30}
\end{aligned}$$

where we have defined,

$$\Omega_{i,j} = \sum_{\ell \in V \setminus \{i,j\}} \chi_{\ell,j},$$

as the excess degree of node j with respect to the edge (i, j) .

Lemma 3

$$\frac{\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j}}{(n-2)D_{n,i}} \mathbf{1}[D_{n,i} > 0] \longrightarrow_n p \text{ a.s.} \tag{31}$$

Proof We use the Borel-Cantelli lemma to show the a.s. convergence result. For ease of exposition the detailed proof is shown in Appendix “[Proof of lemma 3](#)”. \square

Furthermore, we have

$$\frac{n}{D_{n,i}} \mathbf{1}[D_{n,i} > 0] \longrightarrow_n \frac{1}{p} \text{ a.s.} \tag{32}$$

and

$$\frac{1}{D_{n,i}} \mathbf{1}[D_{n,i} > 0] \longrightarrow_n 0 \text{ a.s.} \tag{33}$$

The convergence results (31)-(33) when applied to (30) yields

$$\left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \longrightarrow_n 1 \text{ a.s.} \tag{34}$$

which together with (29) leads to Theorem 4.

Theorem 4 For sequence of Erdos-Renyi graphs $\{\mathbb{G}_n, n = 1, 2, \dots\}$

$$FI_n(i) \longrightarrow_n 1 \text{ a.s.} \tag{35}$$

under the regime p constant.

Also, using the dominated convergence theorem we can also prove convergence in expectation. Details provided in Appendix “[Proof of theorem 5](#)”.

Theorem 5 For sequence of Erdos-Renyi graphs $\{\mathbb{G}_n, n = 1, 2, \dots\}$

$$\mathbb{E}[FI_n(i)] \longrightarrow_n 1 \quad (36)$$

under the regime p constant. Therefore,

$$\mathbb{E}[AFI_n] \longrightarrow_n 1.$$

Theorems 4 & 5 together imply that the friendship paradox is not observed for large ER graphs when p is kept constant. This suggests that the degree imbalances in every node's neighborhood vanish for large graphs under the mentioned regime. One can show that with p held constant, degree of individual nodes diverge in the large graph limit. The intuition for this result is similar to that for the np constant regime with c large, although the two regimes are very different mathematically.

Barabasi-Albert graphs

We analyze the behavior of the AFI in Barabasi-Albert graphs (1999), a well known model for understanding power-law behavior in networks. The sequence of BA graphs $\{\mathbb{G}_0, \mathbb{G}_1, \dots\}$ are such that \mathbb{G}_0 is the initial graph, and the graph grows from \mathbb{G}_t to \mathbb{G}_{t+1} by adding a node labelled $t + 1$ and an edge preferentially between $t + 1$ and a node Q_t chosen from the graph \mathbb{G}_t preferentially on the basis of degree.

The joint degree distribution $p_T(k, \ell)$ is defined as the joint probability of having an edge between two nodes with degrees k and ℓ in graph \mathbb{G}_T . This can be written as

$$p_T(k, \ell) = \frac{1}{T} \sum_{t=1}^T \mathbf{P}[D_{T,Q_t} = k, D_{T,t} = \ell] \quad (37)$$

where $D_{T,t}$ and D_{T,Q_t} are the degrees of nodes t and Q_t in graph \mathbb{G}_T respectively.

The limiting joint degree distribution exists and is defined as (Fotouhi and Rabbat 2013),

$$\lim_{T \rightarrow \infty} p_T(k, \ell) = p(k, \ell) = \frac{4}{k(k+1)\ell(\ell+1)} \left[1 - 6 \frac{\binom{k+\ell-2}{l-1}}{\binom{k+\ell+2}{l+1}} \right]. \quad (38)$$

Let AFI for the graph \mathbb{G}_T be defined as AFI_T . We actually prove that AFI_T diverges with respect to T .

Theorem 6 For sequence of Barabasi-Albert graphs $\mathbb{G}_T, T = 1, 2, \dots$, the expected values of AFIs, $\mathbb{E}[AFI_T]$ diverges.

Proof The following proof uses the fact that the sum of FIs for all the nodes is greater than the sum of the FIs of nodes with degree 1. Using this insight we have

$$\begin{aligned} \mathbb{E}[AFI_T] &\geq \frac{1}{T} \sum_{t=1}^T \mathbb{E}[D_{T,Q_t} \mathbf{1}[D_{T,t} = 1]] \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{d=1}^T d \cdot \mathbf{P}[D_{T,Q_t} = d, D_{T,t} = 1] \end{aligned} \quad (39)$$

We continue by exchanging the order of the finite double sum and only summing for degrees $d \leq M$ for some fixed positive integer M

$$\begin{aligned}\mathbb{E}[AFI_T] &\geq \sum_{d=1}^T \frac{1}{T} \sum_{t=1}^T d \cdot \mathbf{P}[D_{T,Q_t} = d, D_{T,t} = 1] \\ &\geq \sum_{d=1}^M d \frac{1}{T} \sum_{t=1}^T \mathbf{P}[D_{T,Q_t} = d, D_{T,t} = 1] \\ &= \sum_{d=1}^M dp_T(d, 1)\end{aligned}\quad (40)$$

Using (38) with $k = d$ and $\ell = 1$, we obtain on letting T go to infinity

$$\begin{aligned}\lim_{T \rightarrow \infty} \mathbb{E}[AFI_T] &\geq \sum_{d=1}^M d \cdot \frac{2}{d(d+1)} \left[1 - \frac{12}{(d+2)(d+3)} \right] \\ &= \sum_{d=1}^M \frac{2}{d+1} \left[1 - \frac{12}{(d+2)(d+3)} \right]\end{aligned}\quad (41)$$

Observe that the sum in (41) diverges as M is taken to infinity, and therefore the result follows. \square

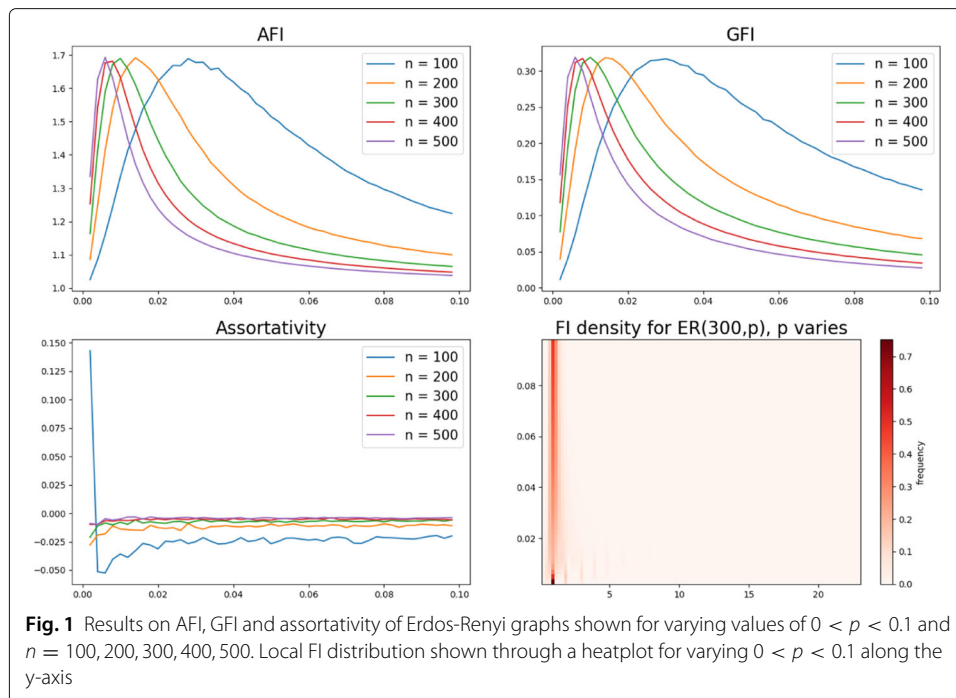
The above result suggests that the friendship paradox keeps growing as the size of the graph grows. This is potentially due to the increasing influence of hubs as the size of the network grows. This is in stark contrast to the behavior of assortativity, which is known to converge to 0 as the size of the BA graph grows (Newman 2002).

Experimental results

Network models

We study three network models - Erdos-Renyi (ER) graphs for modeling random connections; Barabasi-Albert (BA) graphs for modeling power-law behavior; and Watts-Strogatz (WS) graphs known for capturing small-world networks. We study how the global metrics like AFI, GFI, and assortativity behave as the parameters of the models are varied; and connect the experimental results with our theoretical findings on ER and BA graphs obtained in the previous section. We also study the correlation between the local FI metrics and local assortativity (TLA). Specifically, we compute the Pearson, Spearman, and Kendall-Tau correlation coefficients between $|\log FI|$ and TLA because the intuition developed in “[Theoretical results on friendship index](#)” section suggests that $|\log FI|$ and TLA must be negatively correlated – Any deviation from the scenario where the friendship paradox does not occur is best captured by $|\log FI|$, and this measure is expected to increase as the local assortativity (TLA) reduces.

Erdos-Renyi graphs (1960). In Fig. 1, we show the AFI, GFI and assortativity plots for $n = 100, 200, 300, 400, 500$ and varying $0 \leq p \leq 1$. We observe that the phenomenon of the friendship paradox is minimal for both small and large values of p , with the global FI indices reaching a maximum somewhere in between. The intuition behind this finding is that as the probability of connection, p , is varied keeping n fixed, three regimes occur – (a) for small values of p , the graph mostly consists of isolated nodes and edges, which leads to a low value of AFI and GFI; (b) for large values of p , the graph is very close to a



clique, which again exhibits low friendship paradox, and therefore low values of global FI metrics; (c) the global FI metrics reaches the maxima for intermediate values of p . From Theorems 3 & 4 we know that friendship paradox occurs in the regime np constant and does not occur for p constant. And from our experiments we infer that the friendship paradox is observed in the intermediate range of p , which shrinks and gets closer to 0 as the size of the graph grows. This corresponds to the np constant regime, where the friendship paradox is known to occur. Furthermore, we observe that for any fixed value of sufficiently large p such that the realized graphs are dense enough ($p > 0.04$ in Fig. 1), increasing the size of the graph n , reduces the AFI. This validates the theoretical result Theorem 4, which states that the AFI should converge to 1 as n grows large keeping p fixed.

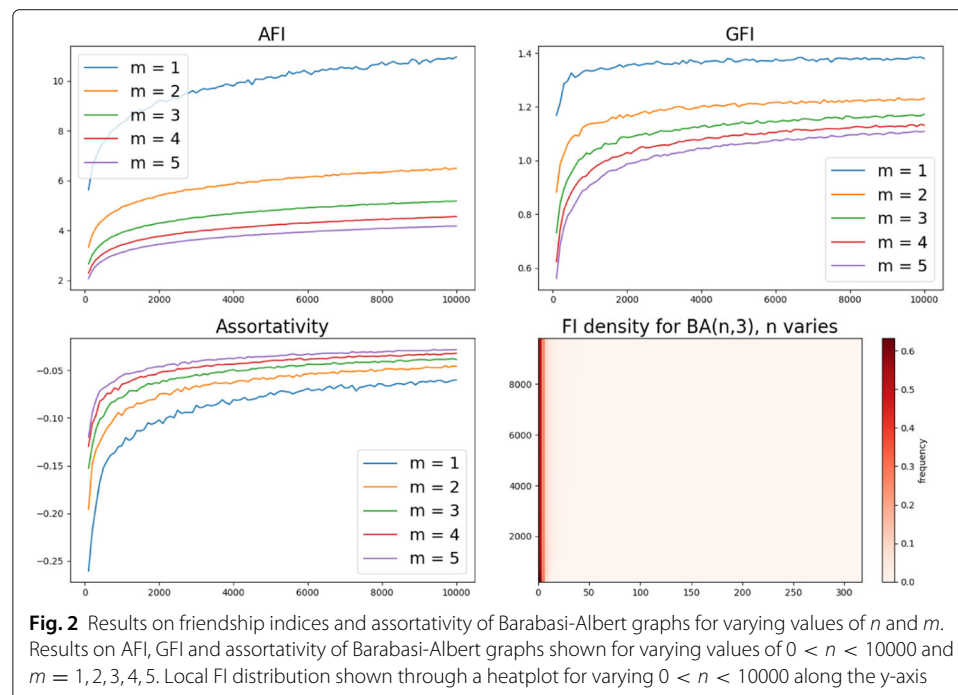
On the other hand, assortativity is observed to be very close to 0 for almost all values of n and p . Thus even for the class of ER graphs, the global FIs exhibit behavior somewhat different from assortativity. We also observe from the heatplot that the local FI distribution peaks at 1 for all values of p , with the peak being very prominent for low values of p , less prominent as p increases, and again becomes prominent as p increases even further. This agrees with the plots on the global metrics of AFI and GFI. In Fig. 4, we show the correlation coefficient between the $|\log FI|$ values and the T-local assortativities of individual nodes. We observe that as p increases, the correlation coefficients start from high negative values close to -1 , then reduce in magnitude, and then again approach -1 as p is further increased. This implies that for sparse and dense ER graphs ($p < 0.8$), the $|\log FI|$ is strongly anticorrelated to TLA. This can be explained by our previous plots which show that the friendship paradox is strong only for a range of values of p , while it is weak for small and large values of p . Our interpretation is that for this particular range of p values where the friendship paradox exists, the anticorrelation between $|\log FI|$ and

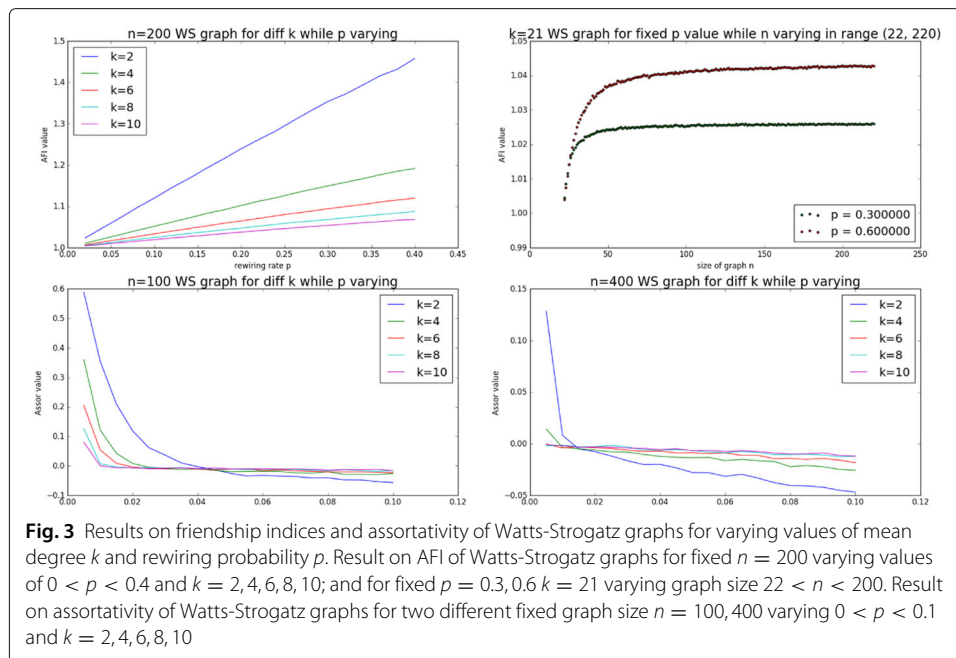
TLA reaches a minima. However, when p is further increased beyond 0.8, the correlation coefficients become unstable because both the $|\log FI|$ and the TLA values approach 0.

Barabasi-Albert graphs (Barabási and Albert 1999). In Fig. 2, we observe that the AFI and GFI increase with number of nodes n , keeping number of edges m fixed, which agrees with the theoretical result Theorem 6, and decrease with increasing m for fixed values of n . The intuition behind this is that with fixed m , higher values of n lead to greater occurrence of the friendship paradox, because the disparity in the degrees of the hubs and leaves or low degree nodes become greater. This is evident by looking at star graphs and how the global FIs increase with graph size. Increasing m and keeping n fixed leads to lower global FIs, because the low degree nodes now have larger degrees to begin with. The assortativity approaches 0 with increasing m for reasons similar to that of the FIs. However, assortativity approaches 0 with increasing n , because of the limiting behavior of BA graphs. The heatmap of the local FI distribution indicate a peak at 1, thereby agreeing with the plots of AFI and GFI.

We also show the correlation coefficient plot in Fig. 4 between $|\log FI|$ and TLA for BA graphs with $n = 1000$ and varying m . We observe the two measures to be anticorrelated, with small and larger values of m leading to stronger anticorrelation.

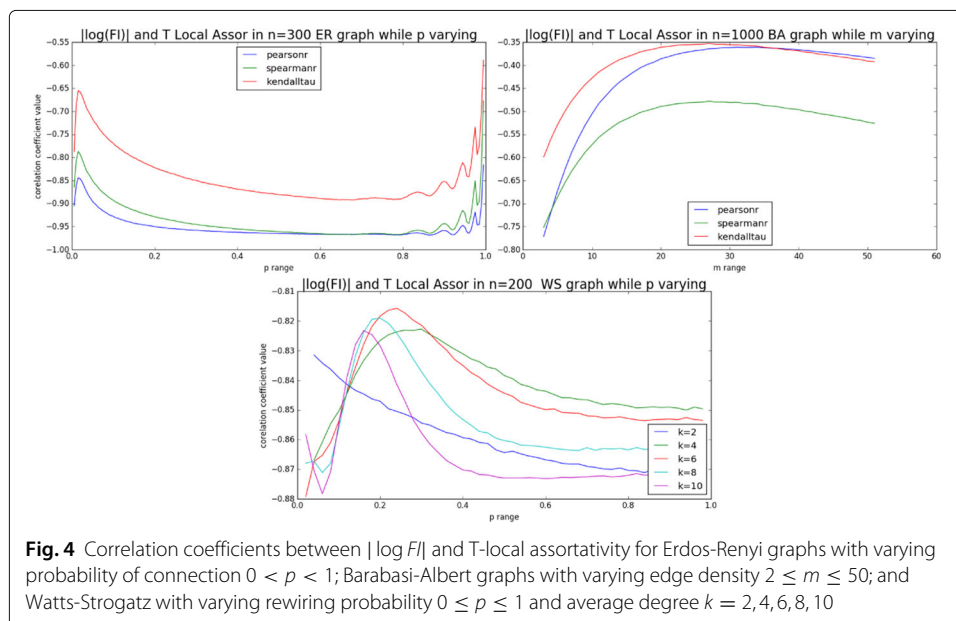
Watts-Strogatz graphs (1998). In Fig. 3 we show the behavior of AFI and assortativity as the parameters of the network models such as the size of the graph n , probability of rewiring p or average degree k are changed. We observe that with graph size fixed at $n = 200$ nodes, increasing p increases the AFI; and, increasing k for fixed p reduces the AFI. The reason for this behavior is that with increasing p , the graphs become more heterogeneous and hence the friendship paradox strengthens. Also, fixing p and increasing k reduces the AFI because larger average degree weakens the effect of the friendship





paradox. We also observe that increasing n keeping p and k fixed increases the AFI which saturates beyond a certain point.

From Fig. 3, we also observe that increasing p reduces the assortativity; While for a fixed p , increasing k gets the assortativity to be closer to 0. The reason for this observation is similar to that stated in the previous paragraph, in that increasing p makes the graph more heterogeneous and leads to the formation of hubs, while fixing p and increasing k reduces the influence of hubs. We also observe that for a larger value of n , the graphs become dissortative for smaller values of p , the reason being that hubs form for smaller values of p in larger graphs. Also Fig. 4 show strong anti-correlation between $|\log FI|$ and TLA for the different parametric settings being considered.



While we study the behavior of local and global metrics on network models, we also observe that the FI is related to assortativity in that $|\log FI|$ is anti-correlated to TLA for most regimes in the three network models being studied. This strengthens the argument that the phenomenon of friendship paradox and assortative mixing are related.

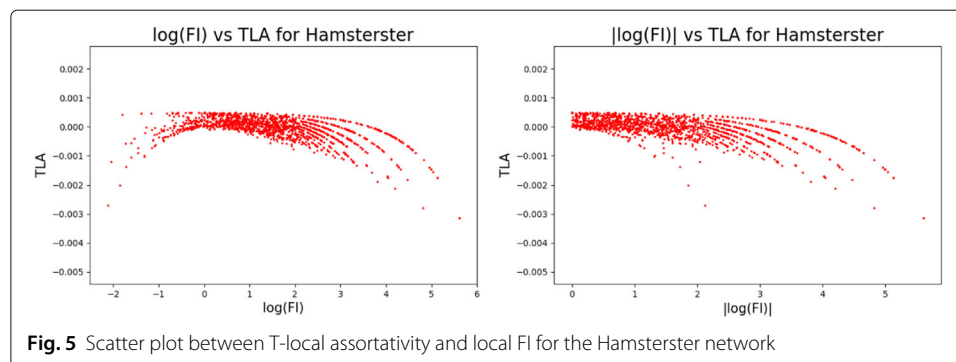
Real world networks

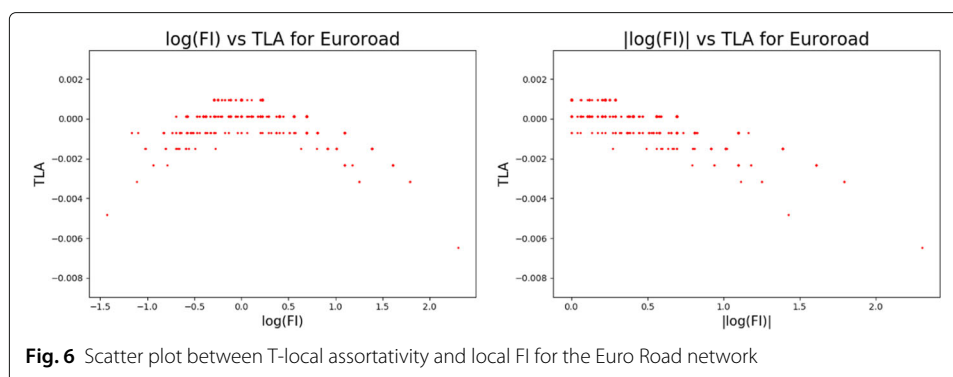
Description of networks. We consider Social(S) networks like the Hamsterster, Brightkite, Douban, Gowalla and Hyves datasets; Human Social(HS) networks of Jazz musicians and the Zachary karate club; Human contact(HC) network at the ACM Hyper-text conference held in Turin and the INFECTIOUS:STAY AWAY exhibit at Dublin, both in 2009; Computer(C) networks formed by Route views and the Internet topology; and, Infrastructure(I) networks like the Power grid and the Euroroad datasets. All of the network datasets were taken from the Koblenz network collection (Kunegis 2013).

Study on the local metrics. In Figs. 5 and 6, we show the scatter plot between the local FI values and the T-local assortativities for the Hamsterster and the EuroRoad networks respectively. We observe that when T-local assortativity is positive the $\log FI$ is close to 0, while for negative local assortativity the $\log FI$ could be either more negative or positive. This gives rise to a bell-shaped curve between local assortativity and $\log FI$, and negative correlation between local assortativity and $|\log FI|$ as observed in the scatter plots. Therefore, if the local assortativity is positive then the friendship paradox would be low, leading to $\log FI$ being close to 0. However, if the local assortativity is negative, then there would be a significant occurrence of the friendship paradox, thereby leading to either more negative or positive values of $\log FI$, and hence higher $|\log FI|$. This justifies the significant negative correlation between TLA and $|\log FI|$ as observed in Table 1.

Study on the global metrics. We consider 13 real networks of different types. We report basic network parameters such as number of nodes and edges, global metrics such as assortativity r and the proposed metrics AFI and GFI. We also report the Pearson correlation coefficient ρ_p and Spearman's rank correlation coefficient ρ_s between the local FI measure in absolute value, $|\log FI|$, and local assortativity r_T .

We observe that while most of the social networks have negative assortativity, all of them exhibit the friendship paradox very strongly. On the other hand, the human social and contact networks exhibit the friendship paradox very weakly partly due to the small size of the network. Computer networks like the Route views and Internet topology show negative assortativity and a strong friendship paradox while infrastructure networks show positive assortativity and a weak friendship paradox. Experimental results from Momeni





and Rabbat (2018) also show that networks with high assortativity (Collaboration in Figure 2 of the paper) have many nodes that do not exhibit friendship paradox while most nodes exhibit the paradox for networks with negative assortativity (Friendster). Lee et al. (2019) also observed that the median-based and the fraction-based perception models correlate negatively with assortativity, which does not quite hold for the mean-based model. Note that the aggregate measures used by Lee et al. (2019) are based on binary measures corresponding to each node, which is substantially different from our setting. As argued previously, we also observe that local assortativity is negatively correlated to $|\log FI|$ very strongly for most considered networks.

Concluding remarks

We first propose a metric called the friendship index that captures the phenomenon of the friendship paradox, locally, from the perspective of an individual node. We then aggregate these metrics to globally capture the network-level friendship paradox. The arithmetic mean of the FI values, AFI , is found to have operational significance in sampling, and is also found to be mathematically suited for analysis of network models. On the other hand, the geometric mean, GFI , is found to better adjust the range of FI values about the FI value for which the friendship paradox does not occur. We lower bound the global metrics, AFI and GFI , and show that the lower bound is achieved only for the class of regular

Table 1 Assortativity and FIs on real networks and the correlations between their local measures

Metric Level	Network Parameter		Global Metric			Local Metric	
Network	nodes	edges	r	AFI	GFI	ρ_p	ρ_s
Hamsterster (S)	2426	16631	-0.0847	12.6416	1.6430	-0.6360	-0.5942
Brightkite (S)	58228	214078	0.0108	24.6750	1.9888	-0.6046	-0.6464
Douban (S)	154908	327162	-0.1803	39.1273	3.3163	-0.5003	-0.7024
Gowalla (S)	196591	950327	-0.0293	235.496	2.4101	-0.7440	-0.7661
Hyves (S)	1402673	2777419	-0.0234	786.311	3.6955	-0.6591	-0.8775
Jazz musicians (HS)	198	2742	0.0202	2.7522	0.5199	-0.6578	-0.6745
Zachary (HS)	34	78	-0.4756	3.6701	0.8985	-0.9098	-0.8803
Hypertext (HC)	113	2196	-0.1226	2.4755	0.3643	-0.9319	-0.9339
Infectious (HC)	410	2765	0.2258	2.1345	0.4262	-0.6997	-0.6204
Route views (C)	6474	13895	-0.1818	176.460	3.7155	-0.8304	-0.9279
Internet topology (C)	34761	107720	-0.2149	303.464	4.3231	-0.7901	-0.8967
Power grid (I)	4941	6594	0.0035	2.1782	0.4792	-0.8315	-0.8030
Euroroad (I)	1174	1417	0.1267	1.4438	0.2289	-0.8256	-0.7721

graphs where the friendship paradox does not occur, thereby showing that the proposed aggregate measures are well-behaved. We theoretically demonstrate that large random (ER) graphs with average degree kept constant exhibit the friendship paradox. However, if the connection probability is kept constant, then large ER graphs do not exhibit the paradox anymore. The results indicate that the sampling bias disappears for large random graphs as the average degree of nodes gets large or very small. We also theoretically show that the paradox exists for the BA graph and gets stronger as the size of the graph grows. This behavior is very different from global assortativity which is known to converge to 0 for both the ER and BA graphs. This is because the friendship index measures the sampling bias by choosing a random neighbor instead of a random node, and the assortativity only compares the degrees of the two endpoints of a random edge. Nevertheless, the two phenomena are not unrelated.

In fact, the local FI measure can shed light on the local assortativity of the network. Experimental results on network models and real world networks suggest that the local FI measures and local assortativity (TLA) are closely related. High values of TLA would mean that the node is connected to similar nodes and the $|\log FI|$ measure would be close to 0, while small values of TLA would mean that the node is connected to dissimilar nodes and $|\log FI|$ would be greater. More broadly, the FI measure captures the imbalance between a node and its neighbors' degrees along with the direction of imbalance, while assortativity only serves as a measure for indicating imbalance between two nodes connected by a random edge. In conclusion, although the friendship paradox and assortativity measures have very different functional forms due to their widely different motivations, they are nonetheless related to each other. Future work could focus on finding theoretical relationships between the two concepts in general graphs, or in certain classes of graphs where the analysis is tractable.

Appendix

Proof of lemma 3

To prove Lemma 3, we use the Borel-Cantelli lemma (Billingsley 2008) which we state below.

Lemma 4 *Let E_1, E_2, \dots be a sequence of events in some probability space. If the sum of the probabilities of $\{E_n\}$ is finite,*

$$\sum_{n=1}^{\infty} \mathbf{P}[E_n] < \infty$$

then the probability that infinitely many of them occur is 0, i.e.,

$$\mathbf{P} \left[\limsup_{n \rightarrow \infty} E_n \right] = \mathbf{P} \left[\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} E_k \right] = 0$$

For a fixed $\epsilon > 0$ and $n = 1, 2, \dots$, we define the following event,

$$E_{n,\epsilon} = \left\{ \left| \frac{\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j}}{(n-2)D_{n,i}} \mathbf{1}[D_{n,i} > 0] - p \right| > \epsilon \right\}.$$

We will use bounding argument on these events and apply the Borel-Cantelli lemma to prove a.s. convergence.

Using Markov's inequality, we have the upper bound

$$\begin{aligned}
 \mathbf{P}[E_{n,\varepsilon}] &= \mathbf{P}\left[\left|\frac{\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j}}{(n-2)D_{n,i}} \mathbf{1}[D_{n,i} > 0] - p\right| > \varepsilon\right] \\
 &\leq \frac{1}{\varepsilon^4} \mathbb{E}\left[\left(\frac{\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j}}{(n-2)D_{n,i}} \mathbf{1}[D_{n,i} > 0] - p\right)^4\right] \\
 &= \frac{1}{\varepsilon^4} \mathbb{E}\left[\left(\frac{\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j}}{(n-2)D_{n,i}} - p\right)^4 \mathbf{1}[D_{n,i} > 0]\right] + \frac{p^4}{\varepsilon^4} \mathbf{P}[D_{n,i} = 0] \\
 &= \frac{1}{\varepsilon^4} \mathbb{E}\left[\frac{1}{n^4 D_{n,i}^4} \left(\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j} - (n-2)D_{n,i}p\right)^4 \mathbf{1}[D_{n,i} > 0]\right] + \frac{p^4}{\varepsilon^4} (1-p)^{n-1}
 \end{aligned} \tag{42}$$

The second term in (42) goes to 0 exponentially fast. The first term can be written as

$$\begin{aligned}
 &\frac{1}{\varepsilon^4} \mathbb{E}\left[\frac{1}{n^4 D_{n,i}^4} \left(\sum_{j \in \mathcal{N}_{n,i}} \Omega_{i,j} - (n-2)D_{n,i}p\right)^4 \mathbf{1}[D_{n,i} > 0]\right] \\
 &= \frac{1}{\varepsilon^4} \mathbb{E}\left[\frac{1}{n^4 D_{n,i}^4} \left(\sum_{j \in \mathcal{N}_{n,i}} [\Omega_{i,j} - (n-2)p]\right)^4 \mathbf{1}[D_{n,i} > 0]\right] \\
 &= \frac{1}{\varepsilon^4} \mathbb{E}\left[\frac{1}{n^4 D_{n,i}^4} \mathbb{E}\left[\left(\sum_{j \in \mathcal{N}_{n,i}} [\Omega_{i,j} - (n-2)p]\right)^4 \middle| \mathcal{N}_{n,i} = N_i\right] \mathbf{1}[D_{n,i} > 0]\right].
 \end{aligned} \tag{43}$$

We work with the term in the inner expectation,

$$\begin{aligned}
 &\mathbb{E}\left[\left(\sum_{j \in \mathcal{N}_{n,i}} [\Omega_{i,j} - (n-2)p]\right)^4 \middle| \mathcal{N}_{n,i} = N_i\right] \\
 &= \mathbb{E}\left[\sum_{q \in \mathcal{N}_{n,i}} \sum_{r \in \mathcal{N}_{n,i}} \sum_{s \in \mathcal{N}_{n,i}} \sum_{t \in \mathcal{N}_{n,i}} (\Omega_{i,q} - (n-2)p) (\Omega_{i,r} - (n-2)p) \right. \\
 &\quad \left. \times (\Omega_{i,s} - (n-2)p) (\Omega_{i,t} - (n-2)p) \middle| \mathcal{N}_{n,i} = N_i\right] \\
 &= \sum_{q \in \mathcal{N}_{n,i}} \sum_{r \in \mathcal{N}_{n,i}} \sum_{s \in \mathcal{N}_{n,i}} \sum_{t \in \mathcal{N}_{n,i}} \mathbb{E}\left[(\Omega_{i,q} - (n-2)p) (\Omega_{i,r} - (n-2)p) \right. \\
 &\quad \left. \times (\Omega_{i,s} - (n-2)p) (\Omega_{i,t} - (n-2)p) \middle| \mathcal{N}_{n,i} = N_i\right] \\
 &= \sum_{q \in \mathcal{N}_{n,i}} \sum_{r \in \mathcal{N}_{n,i}} \sum_{s \in \mathcal{N}_{n,i}} \sum_{t \in \mathcal{N}_{n,i}} \mathbb{E}\left[(\Omega_{i,q} - (n-2)p) (\Omega_{i,r} - (n-2)p) \right. \\
 &\quad \left. \times (\Omega_{i,s} - (n-2)p) (\Omega_{i,t} - (n-2)p) \right]
 \end{aligned} \tag{44}$$

For a given choice of q, r, s, t , we define

$$\Omega'_{i,z} = \sum_{\ell \neq i, q, r, s, t} \chi_{z,\ell}, \text{ for } z = q, r, s, t$$

The first term in the product can be written as

$$\left(\left(\Omega'_{i,q} - (n-5)p \right) + (\chi_{q,r} - p) + (\chi_{q,s} - p) + (\chi_{q,t} - p) \right).$$

Observe that, the terms $(\chi_{q,z} - p)$, $z = r, s, t$ is mutually independent with respect to terms $(\Omega'_{i,z} - (n-5)p)$, $z = r, s, t$. Therefore, such terms will not contribute. The product between terms of the form $(\chi_{q,z} - p)$, $z = q, r, s, t$ will be finite and bounded by a constant say c_0 . Therefore, we can only consider the product of the terms $(\Omega'_{i,z} - (n-5)p)$, $z = q, r, s, t$. There are several types of products that have to be considered separately – (a) For $q \neq r \neq s \neq t$, the contribution is 0 because all the terms are mutually independent; (b) For $q = r, r \neq s \neq t$, the terms $(\Omega'_{i,z} - (n-5)p)$ for $z = s, t$ are mutually independent with respect to $z = q$. Hence, such terms also do not have any contribution; (c) For $q = r \neq s = t$, the two terms $(\Omega'_{i,z} - (n-5)p)^2$, $z = q, s$ will have a contribution, which will be discussed later; (d) For $q = r = s \neq t$, term $(\Omega'_{i,t} - (n-5)p)^2$ is independent of the other three terms. The expectation of this term is 0, therefore leading to no contribution. (e) For $q = r = s = t$, the contribution of such terms need to be considered.

Using the above insights and continuing from (44), we have

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j \in N_i} [\Omega_{i,j} - (n-2)p] \right)^4 \middle| \mathcal{N}_{n,i} = N_i \right] &= \sum_{q \in \mathcal{N}_{n,i}} \mathbb{E} \left[\left(\Omega'_{i,q} - (n-5)p \right)^4 \right] \\ &+ \sum_{q \in \mathcal{N}_{n,i}} \sum_{s \in \mathcal{N}_{n,i}} \mathbb{E} \left[\left(\Omega'_{i,q} - (n-5)p \right)^2 \right] \mathbb{E} \left[\left(\Omega'_{i,s} - (n-5)p \right)^2 \right] \end{aligned} \quad (45)$$

We write

$$\begin{aligned} \mathbb{E} \left[\left(\Omega'_{i,q} - (n-5)p \right)^2 \right] &\approx \mathbb{E} \left[\left(\sum_{\ell \neq i, q, s} \chi_{q,\ell} - p \right)^2 \right] \\ &= \sum_{\ell \neq i, q, s} \mathbb{E} \left[(\chi_{1,2} - p)^2 \right] = nc_0, \end{aligned} \quad (46)$$

and

$$\begin{aligned} \mathbb{E} \left[\left(\Omega'_{i,q} - (n-5)p \right)^4 \right] &\approx \mathbb{E} \left[\left(\sum_{\ell \neq i, q, s} \chi_{q,\ell} - p \right)^4 \right] \\ &= \sum_{\ell \neq q, s} \sum_{\ell' \neq q, s, \ell} \mathbb{E} \left[(\chi_{q,\ell} - p)^2 \right] \mathbb{E} \left[(\chi_{q,\ell'} - p)^2 \right] + \sum_{\ell \neq q, s} \mathbb{E} \left[(\chi_{q,\ell} - p)^4 \right] \\ &\approx n^2 c_0^2 + nc_1 \end{aligned} \quad (47)$$

where $c_0 = \mathbb{E} \left[(\chi_{1,2} - p)^2 \right]$, and $c_1 = \mathbb{E} \left[(\chi_{1,2} - p)^4 \right]$. Substituting (46) and (47) into (45), we obtain

$$\mathbb{E} \left[\left(\sum_{j \in N_i} [\Omega_{i,j} - (n-2)p] \right)^4 \middle| \mathcal{N}_{n,i} = N_i \right] = D_{n,i}^2 n^2 c_0^2 + D_{n,i} (n^2 c_0^2 + nc_1) \quad (48)$$

Substituting (48) to (42), we obtain

$$\begin{aligned} \mathbf{P}[E_{n,\varepsilon}] &= \frac{1}{\varepsilon^4} \mathbb{E} \left[\frac{1}{n^4 D_{n,i}^4} \cdot (D_{n,i}^2 n^2 c_0^2 + D_{n,i} (n^2 c_0^2 + n c_1)) \mathbf{1}[D_{n,i} > 0] \right] + \frac{p^4}{\varepsilon^4} (1-p)^{n-1} \\ &\leq \frac{2c_0^2}{\varepsilon^4 n^2} + \frac{p^4}{\varepsilon^4} (1-p)^{n-1} \end{aligned} \quad (49)$$

Observe that,

$$\sum_{n=1}^{\infty} \mathbf{P}[E_{n,\varepsilon}] < \infty \quad (50)$$

holds and therefore the result follows using Borel-Cantelli Lemma.

Proof of theorem 5

We will use bounded convergence theorem (Billingsley 2008) to prove convergence in expectation of the FIs.

Recall that

$$FI_n(i) = \left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] + \mathbf{1}[D_{n,i} = 0] \quad (51)$$

Since $\mathbf{1}[D_{n,i} = 0] \leq 1$, we only consider the first term in (51). In order to apply Chernoff-Hoeffding bound we consider the event $\{D_{n,i} < n(p - \delta)\}$, and fix $\delta > 0$. We have the following decomposition

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \right] \\ &= \mathbb{E} \left[\left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \mathbf{1}[D_{n,i} > n(p - \delta)] \right] \\ &\quad + \mathbb{E} \left[\left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \mathbf{1}[D_{n,i} < n(p - \delta)] \right] \end{aligned} \quad (52)$$

The first term in (52) is upper bounded as

$$\mathbb{E} \left[\left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \mathbf{1}[D_{n,i} > n(p - \delta)] \right] \leq \frac{1}{n^2(p - \delta)^2} \cdot n^2 = \frac{1}{(p - \delta)^2}, \quad (53)$$

where the upper bound follows by noting that the sum of the degree of neighbors of i is bounded by n^2 . Furthermore, we also have the upper bound

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{D_{n,i}^2} \sum_{j \in \mathcal{N}_{n,i}} D_{n,j} \right) \mathbf{1}[D_{n,i} > 0] \mathbf{1}[D_{n,i} < n(p - \delta)] \right] &\leq n^2 \mathbf{P}[D_{n,i} < n(p - \delta)] \\ &\leq n^2 e^{-nD(p||p-\delta)}, \end{aligned} \quad (54)$$

where $D(p||p - \delta)$ is the KL divergence between Bernoulli rvs with parameters p and $p - \delta$ and the final step is obtained by Chernoff-Hoeffding bound. Since $n^2 e^{-nD(p||p-\delta)} \rightarrow 0$ as $n \rightarrow \infty$, there exists a constant c_δ such that

$$n^2 e^{-nD(p||p-\delta)} \leq c_\delta, n = 1, 2, \dots \quad (55)$$

Therefore using the upper bounds (53) and (55), we obtain that

$$\mathbb{E}[FI_n(i)] \leq 1 + c_\delta + \frac{1}{(p - \delta)^2}, \quad n = 1, 2, \dots \quad (56)$$

Since all the rvs $FI_n(i)$ are bounded, the bounded convergence theorem yields the result.

Abbreviations

AFI: Arithmetic friendship index; BA: Barabasi-Albert; ER: Erdos-Renyi; FI: Friendship index; GFI: Geometric friendship index; HFI: Harmonic friendship index; TLA: Thedchanamoorthy's local assortativity

Acknowledgements

SP acknowledges that his prior discussions with Prof. Armand Makowski at University of Maryland College Park was helpful in connecting the definition of friendship index and assortativity with that of various sampling methods.

Authors' contributions

SP worked on conceptualization, methodology, theoretical analysis, experimental design, writing, review and editing, and acquisition of funding for the paper. FY worked on conceptualization, methodology, software, and theoretical analysis. YN worked on conceptualization, experimental analysis, and writing. AS worked on conceptualization, supervision of research, and review and editing. ABN worked on conceptualization, supervision of research and acquisition of funding. All authors read and approved the final manuscript.

Funding

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

Availability of data and materials

All of the network datasets were taken from the Koblenz network collection, which is available at <http://konect.uni-koblenz.de/>.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Raytheon BBN Technologies, 10 Moulton Street, 02138 Cambridge, USA. ²City University of New York, 10016 New York, USA. ³Touro College and University System, 10001 New York, USA. ⁴U.S. Army Research Lab, 20783 Adelphi, USA.

Received: 2 April 2019 Accepted: 23 August 2019

Published online: 23 September 2019

References

- Bagrow JP, Danforth CM, Mitchell L (2017) Which friends are more popular than you?: Contact strength and the friendship paradox in social networks. In: Proceedings of the IEEE/ACM Int'l Conference on Advances in Social Networks Analysis and Mining 2017. ACM, New York. pp 103–108
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
- Billingsley P. (2008) Probability and measure. Wiley, New York
- 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):4
- Eom Y-H, Jo H-H (2014) Generalized friendship paradox in complex networks: The case of scientific collaboration. *Sci Rep* 4:4603
- Erdos P, Rényi A (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5(1):17–60
- Feld SL (1991) Why your friends have more friends than you do. *Am J Sociol* 96(6):1464–77
- Fotouhi B, Rabbat MG (2013) Degree correlation in scale-free graphs. *Eur Phys J B* 86(12):510
- Fotouhi B, Momeni N, Rabbat MG (2014) Generalized friendship paradox: An analytical approach. In: Int Conf Soc Inform. Springer, Cham. pp 339–352
- Hodas NO, Kooti F, Lerman K (2013) Friendship paradox redux: Your friends are more interesting than you. *Int AAAI Conf Web Soc Med (ICWSM)* 13:8–10
- Jackson MO (2016) The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy* 127(2):777–818
- Jo H-H, Eom Y-H (2014) Generalized friendship paradox in networks with tunable degree-attribute correlation. *Phys Rev E* 90(2):022809
- Kunegis J (2013) KONECT: The Koblenz network collection. <http://konect.uni-koblenz.de/>
- 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
- Leskovec J, Faloutsos C (2006) Sampling from large graphs. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York. pp 631–636

- Momeni N, Rabbat M (2016) Qualities and inequalities in online social networks through the lens of the friendship paradox. *PloS one* 11(2):e0143633
- Momeni N, Rabbat M (2018) Effectiveness of alter sampling in social networks. *arXiv:1812.03096v2*
- Newman ME (2002) Assortative mixing in networks. *Phys Rev Lett* 89(20):208701
- Pal S, Yu F, Novick Y, Swami A, Bar-Noy A (2018) Quantifying the strength of the friendship paradox. In: *International Workshop on Complex Networks and Their Applications*. Springer, Cham. pp 460–472
- Piraveenan M, Prokopenko M, Zomaya AY (2010) Classifying complex networks using unbiased local assortativity. In: *Conf Artif Life*. pp 329–336
- Thedchanamoorthy G, Piraveenan M, Kasthuriratna D, Senanayake U (2014) Node assortativity in complex networks: An alternative approach. *Procedia Comput Sci* 29:2449–2461
- Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440

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)