The friendship paradox in real and model networks

GEORGE T. CANTWELL[†]

*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA and Department of Physics,
University of Michigan, Ann Arbor, MI 48109, USA*
[†]*Corresponding author. Email: gcant@umich.edu*

ALEC KIRKLEY

Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

AND

M. E. J. NEWMAN

*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA and Center for the Study of
Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA*

Edited by: Ernesto Estrada

The friendship paradox is the observation that the degrees of the neighbours of a node in any network
will, on average, be greater than the degree of the node itself. In common parlance, your friends have
more friends than you do. In this article, we develop the mathematical theory of the friendship paradox,
both in general as well as for specific model networks, focusing not only on average behaviour but also
on variation about the average and using generating function methods to calculate full distributions of
quantities of interest. We compare the predictions of our theory with measurements on a large number of
real-world network datasets and find remarkably good agreement. We also develop equivalent theory for
the generalized friendship paradox, which compares characteristics of nodes other than degree to those of
their neighbours.

Keywords: friendship paradox; generalized friendship paradox; random graphs; real-world networks;
generating functions; assortative mixing.

1. Introduction

It may appear to you that your friends are more popular than you are, and if so, you may well be right. It
is however not necessarily your fault. It is, rather, a natural consequence of network structure. Feld [1]
has shown that in any network the average degree (i.e. the number of neighbours) of the neighbour of a
node is strictly greater than the average degree of nodes in the network as a whole. Applied to networks
of friendship, this implies that on average your friends have more friends than you do. This phenomenon
is known as the *friendship paradox*.

A related phenomenon, the *generalized friendship paradox*, describes similar behaviour with respect
to other attributes of network nodes [2]. Are your friends richer than you, for instance, or smarter, or more
attractive? Generalized friendship paradoxes arise when such attributes are correlated with node degree.
If richer people are on average also more popular, then wealth and popularity will be positively correlated

and hence the tendency for your friends to be more popular than you could mean they are also richer. Examples of generalized friendship paradoxes occur for instance with citation counts in collaboration networks (your collaborators have more citations than you do) [3] and viral content dissemination in online social networks (your online friends receive more viral content than you do) [4].

Understood as a mathematical statement about network averages, the friendship paradox occurs in all networks. How the effect is manifested, however, depends on the details of a network's structure. For any given person, we can measure the difference between how popular their friends are and how popular they are. The results of Feld [1] tell us that this difference must be positive on average, but to accurately characterize the effect we should consider the entire distribution of differences. How large will the differences be? Is their average driven by a handful of outliers? How much variation is there? For how many people will the difference be positive? The answers to such questions all depend on the specific details of the network under study.

In this article, we develop the theory of the friendship paradox in both its original and generalized versions. We derive expressions for the full distribution of differences in three progressively more complex network models and find that the observed distributions in real-world networks are in good agreement with our theoretical results.

2. The friendship paradox

Informally, the friendship paradox states that people's friends tend to be more popular than they themselves are. Stated a little more precisely, nodes in a network tend to have a lower degree than their neighbours do. Consider an undirected network of n nodes, labelled by integers $i = 1, \ldots, n$. For any given node i we can compute the difference Δ_i between the average of its neighbours' degrees and its own degree:

$$\Delta_i = \frac{1}{k_i} \sum_j A_{ij} k_j - k_i, \tag{2.1}$$

where A_{ij} is an element of the adjacency matrix and $k_i = \sum_j A_{ij}$ is the degree of node i. (The value of Δ_i is undefined for nodes with degree zero, since they have no neighbours. We will assume that there are no such nodes in the network, or that all of them have been removed before the analysis.)

One statement of the friendship paradox is that the average of Δ_i across all nodes is greater than zero[1], which can be proven as follows. We write the average as

$$\frac{1}{n} \sum_{i=1}^n \Delta_i = \frac{1}{n} \sum_i \left(\frac{1}{k_i} \sum_j A_{ij} k_j - k_i \right) = \frac{1}{n} \sum_{ij} \left(A_{ij} \frac{k_j}{k_i} - A_{ij} \right) = \frac{1}{n} \sum_{ij} A_{ij} \left(\frac{k_j}{k_i} - 1 \right). \tag{2.2}$$

---

[1] Here, we use the arithmetic mean of Δ_i but this is not the only option for precisely formulating the claim that 'people's friends tend to be more popular than they themselves are'. One could, for example, consider the fraction of nodes for which $\Delta_i > 0$. Another alternative is to consider the average degree of nodes at the end of randomly chosen edges and show that this is greater than the average degree of nodes chosen uniformly at random. This latter approach, however, only corresponds to the English sentence 'people's friends have more friends than them' when nodes' neighbours are sufficiently representative of the full population—see [1] for a discussion of this point.

Exchanging the summation variables i and j and adding the result to Eq. (2.2), this result can also be written

$$\frac{1}{n} \sum_{i=1}^{n} \Delta_i = \frac{1}{2n} \sum_{ij} A_{ij} \left(\frac{k_j}{k_i} + \frac{k_i}{k_j} - 2 \right) = \frac{1}{2n} \sum_{ij} A_{ij} \left(\sqrt{\frac{k_j}{k_i}} - \sqrt{\frac{k_i}{k_j}} \right)^2 \geq 0. \tag{2.3}$$

The exact equality holds only when $k_i = k_j$ for all pairs of neighbouring nodes, that is, when the network is a regular graph (or, technically, when every component is a regular graph). In all other cases, the average difference between the mean degree of a node's neighbours and its own degree is strictly greater than zero.

For the generalized friendship paradox, which considers attributes other than degree, one can define an analogous quantity $\Delta_i^{(x)}$ for an attribute x according to

$$\Delta_i^{(x)} = \frac{1}{k_i} \sum_j A_{ij} x_j - x_i, \tag{2.4}$$

which measures the difference between the average of the attribute for node i's neighbours and the value for i itself. When the average of this quantity over all nodes is positive one may say that the generalized friendship paradox holds.[2] In contrast to the case of degree, this is not always true—the value of $\Delta_i^{(x)}$ can be zero or negative—but we can write the average as

$$\frac{1}{n} \sum_i \Delta_i^{(x)} = \frac{1}{n} \sum_i \left(\frac{1}{k_i} \sum_j A_{ij} x_j - x_i \right) = \frac{1}{n} \sum_i \left(x_i \sum_j \frac{A_{ij}}{k_j} - x_i \right), \tag{2.5}$$

where the second line again follows from interchanging summation indices. Defining the new quantity

$$\kappa_i = \sum_j \frac{A_{ij}}{k_j} \tag{2.6}$$

and noting that

$$\frac{1}{n} \sum_i \kappa_i = \frac{1}{n} \sum_{ij} \frac{A_{ij}}{k_j} = \frac{1}{n} \sum_j \frac{1}{k_j} \sum_i A_{ij} = 1, \tag{2.7}$$

we can then write

$$\frac{1}{n} \sum_i \Delta_i^{(x)} = \frac{1}{n} \sum_i x_i \kappa_i - \frac{1}{n} \sum_i x_i \frac{1}{n} \sum_i \kappa_i = \text{Cov}(x, \kappa). \tag{2.8}$$

---

[2] Again, one need not use the arithmetic mean of $\Delta_i^{(x)}$ and there are multiple ways to make precise the claim that 'people's friends tend to have more x than they themselves do'. One approach, as before, is to consider the average value of x at the end of randomly chosen edges and compare this to the average value of x for uniformly chosen nodes. This corresponds to considering $\text{Cov}(x, k)$, and is equivalent to the definition of the 'network level' generalized friendship paradox in [3]. But, for the same reason as for the friendship paradox [1, 5], this only really corresponds to the plain English statement of the generalized friendship paradox when nodes' neighbours are sufficiently representative of the population.

Thus, we will have a generalized friendship paradox in the sense defined here if (and only if) x and κ are positively correlated.

This is perhaps not exactly the result one might have anticipated when, in the introduction, we argued that generalized friendship paradoxes arise when an attribute x is positively correlated with degree—it turns out that correlation with the more complex quantity κ is the crucial behaviour. Degree, however, is always positively correlated with κ since, combining Eqs (2.3) and (2.8), we have $\mathrm{Cov}(k, \kappa) = (1/n) \sum_i \Delta_i \geq 0$. While it is not mathematically guaranteed, in practice we therefore expect properties that correlate with degree to also correlate with κ and hence in such situations we can reasonably expect to observe a generalized friendship paradox.

3. Models

While these results are illuminating, there is more to be said on this topic. We have shown that the network average of Δ_i will always be greater than zero, but we have not said by how much or what the value depends on. Nor have we considered how Δ_i varies across a network. As an example (albeit a somewhat contrived one), consider a 1000-node network consisting of a complete graph with one edge removed. In such a network almost all of the nodes—998 of them—have a degree *larger* than the average of their neighbours, but there are two outliers that substantially skew the distribution so that over the whole network nodes are still less popular than their neighbours on average. In this case, therefore, the average is not very informative. To understand the friendship paradox fully, we need to move beyond statements about averages.

In doing so, however, the specific structure of the network becomes important, which can lead to a more subtle picture of the effect (see e.g. [6]). A common way to study the effects of structure is to examine the behaviour of model networks and this is the approach we take in the remainder of this article, considering the friendship paradox for three successively more complex network models, the Poisson random graph (sometimes called the Erdős–Rényi model) [7–9], the configuration model [10, 11], and a more sophisticated model that incorporates degree correlations.

3.1 *Poisson random graph*

The Poisson random graph is the simplest of random graph models and one of the most widely studied. In this model, one takes n nodes and connects each pair independently with some probability p. When the network is large and sufficiently sparse—when $n \to \infty$ and $p = \lambda/(n-1)$ with λ growing slower than n so that $p \to 0$—the degrees are Poisson distributed with mean λ, meaning that the probability p_k of a node having degree k is

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{3.1}$$

Let us compute the complete distribution within this model of the quantity Δ_i defined in Eq. (2.1). To compute this distribution, we note that

$$P(\Delta) = \sum_k p_k P(\Delta|k), \tag{3.2}$$

where $P(\Delta|k)$ denotes the probability that a node has value Δ given that it has degree k. For given Δ and k, Eq. (2.1) tells us that the sum of the neighbouring degrees of the node is $\sum_j A_{ij} k_j = k\Delta + k^2$. Let

us denote this sum by K. Then, we have

$$P(\Delta|k) = P(K = k\Delta + k^2|k). \tag{3.3}$$

We know that K is an integer with $K \geq k$, since each of the node's k neighbours necessarily has degree at least 1. Hence the allowed values of Δ that satisfy $K = k\Delta + k^2$ must be rational numbers of the form $\Delta = 1 - k + m/k$, where m is a non-negative integer.

To evaluate (3.3), we need to know the distribution of the sum $K = \sum_j A_{ij}k_j$ of neighbour degrees. While nodes in general follow the degree distribution p_k, the degrees of neighbours follow a modified distribution. A neighbour, by definition, is a node arrived at by following an edge and each node of degree k is at the end of k edges, so the degree distribution for nodes at the ends of edges goes not as p_k but as kp_k, which after appropriate normalization gives a probability distribution q_k for neighbour degrees of the form

$$q_k = \frac{kp_k}{\sum_j jp_j}. \tag{3.4}$$

For the Poisson degree distribution of Eq. (3.1), we then find that

$$q_k = \frac{k\lambda^k e^{-\lambda}}{\lambda k!} = \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}. \tag{3.5}$$

In other words, q_k is also a Poisson distribution, but shifted by one, meaning that $k - 1$ is a Poisson variable with mean λ.

We can write the sum K as

$$K = \sum_j A_{ij}k_j = k_i + \sum_j A_{ij}(k_j - 1). \tag{3.6}$$

But $\sum_j A_{ij}(k_j - 1) \sim \text{Poisson}(k_i\lambda)$, since the sum of the Poisson variables $(k_j - 1)$ is itself a Poisson random variable with mean a factor of k_i greater. Hence,

$$P(\Delta|k) = P(K = k\Delta + k^2|k) = \frac{(k\lambda)^{k\Delta + k^2 - k}}{(k\Delta + k^2 - k)!}e^{-k\lambda}. \tag{3.7}$$

For each value of k, this gives us a distribution over a discrete set of equally spaced values $\Delta = 1 - k + m/k$ with $m = 0 \ldots \infty$, and the full distribution, Eq. (3.2), is a linear combination of an infinite number of such distributions.

Figure 1 shows an example of the resulting distribution of Δ computed from Eq. (3.2) for the case $\lambda = 8$. The individual discrete distributions are clearly visible as sets of equally spaced peaks and the overall resulting distribution is quite complicated—both widely dispersed and jagged, even for this simple network model. Note also that a significant fraction of nodes have $\Delta < 0$, meaning that they do not satisfy the traditional definition of the friendship paradox—they have more friends than their average neighbour does.

It is also possible to calculate the mean and variance of Δ for the Poisson random graph. We find that

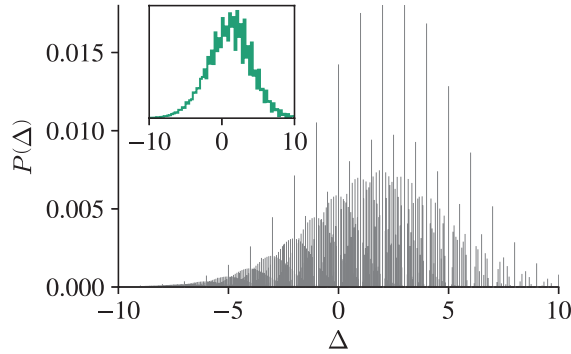$$E[\Delta] = 1; \qquad \text{Var}(\Delta) = \lambda(1 + E[k^{-1}]). \tag{3.8}$$

FIG. 1. The distribution of the quantity Δ, defined in Eq. (2.1), for a random graph with mean degree 8. The main panel shows the probability mass function while the inset shows a traditional histogram of the distribution.

(Note that again we remove any nodes of degree zero before computing $E[k^{-1}]$.) While the expected value of Δ is always 1, the variance is larger than the mean degree λ, meaning that more and more nodes have $\Delta < 0$ as λ becomes large, with the fraction tending to a half. For example, in Fig. 1, where the mean degree is 8, around 35% of nodes have a degree larger than that of their average neighbour. This rises to 44% when the mean degree is 64, and 49% for a mean degree of 1024. For large λ therefore, no meaningful 'friendship paradox' applies. Here, as is often the case, looking only at the average value of the distribution is misleading when there is a large variation.

3.2 *The configuration model*

The random graph of the previous section is in many respects not a realistic model. In particular, as we have noted, it has a Poisson degree distribution, which is very different from the broad degree distributions seen in typical real-world networks [12, 13]. We can address this shortcoming by using a more sophisticated random graph model that allows for arbitrary degree distributions, the so-called configuration model [10, 11]. In this model, one fixes the degree of each of the nodes and then draws a network at random from the set of all networks with the given degrees.

Calculations on networks such as the configuration model can be greatly simplified by using generating function methods [5, 14]. For instance, many properties of the model can be expressed in terms of the generating function for the degree distribution p_k, defined by

$$f(z) = \sum_{k=0}^{\infty} p_k z^k. \tag{3.9}$$

We employ this approach here too, but there is a catch, in that generating functions are normally applied to distributions over integer quantities, like the degree, but the quantity Δ, whose distribution is our main focus here, can take non-integer values. To allow for this, we make use of the (two-sided) Laplace transform

$$F(s) = \int_{-\infty}^{\infty} p(x)\, e^{-sx}\, dx, \tag{3.10}$$

which is the standard extension of the generating function to a variable x on the real line.

However, the distribution of Δ is not continuous-valued either. It is non-zero on a (potentially) dense set of rational values but zero everywhere else—see Fig. 1. To allow for this, we consider functions $p(x)$ that are equal to a discrete sum of Dirac delta functions. For the degree distribution p_k, for instance, we would define

$$p(x) = \sum_{k=0}^{\infty} p_k \delta(x - k). \tag{3.11}$$

With this definition Eqs. (3.9) and (3.10) are essentially equivalent, since

$$F(s) = \int_{-\infty}^{\infty} \sum_k p_k \delta(x - k)\, e^{-sx}\, dx = \sum_k p_k e^{-sk} = f(e^{-s}), \tag{3.12}$$

but Eq. (3.10) also allows for quantities like Δ that have non-integer values.

The Laplace transform has several properties that will be useful for our purposes. First, if x is a random variable whose distribution has Laplace transform $F_x(s)$, then the Laplace transform for the distribution of $ax + b$ is

$$F_{ax+b}(s) = e^{-bs} F_x(as). \tag{3.13}$$

Second, if x and y are independent random variables, then the Laplace transform for their sum is the product of the Laplace transforms for x and y alone:

$$F_{x+y}(s) = F_x(s) F_y(s). \tag{3.14}$$

These two results now allow us to calculate $F_\Delta(s)$, the Laplace transform for $P(\Delta)$. We follow essentially the same logic as we did for the Poisson random graph: we consider the Laplace transform for the distribution of Δ for fixed degree, then we average over degree.

If node i has degree k_i, then from Eq. (2.1) Δ_i is

$$\Delta_i = \sum_j A_{ij} \left(\frac{k_j}{k_i} - 1 \right). \tag{3.15}$$

Each k_j is the degree of a neighbour node which, as before, is a random quantity distributed according to $q_k \propto k p_k$. Let $G(s)$ be the Laplace transform for this distribution:

$$G(s) = \sum_{k=0}^{\infty} q_k e^{-sk}. \tag{3.16}$$

Then from Eq. (3.13) the Laplace transform for $k_j/k_i - 1$ is $e^s\, G(s/k_i)$ and, since each of the k_i non-zero terms in the sum of Eq. (3.15) is independent, the Laplace transform for the full sum is $[e^s G(s/k_i)]^{k_i}$ by Eq. (3.14).

This is for a node of degree k_i. Since the Laplace transform is linear, we can now simply average over degree to compute the transform for the full distribution of Δ:

$$F_\Delta(s) = \sum_k p_k\, e^{sk}\, G(s/k)^k. \tag{3.17}$$

Given any degree distribution p_k we can use this equation to compute $F_\Delta(s)$. Inverting the Laplace transform then gives us the density function $\rho(x)$ of Δ itself:

$$\rho(x) = \sum_\Delta P(\Delta)\delta(x - \Delta) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_\Delta(is) e^{isx} \, ds, \tag{3.18}$$

where the sum over Δ is a sum over all rational numbers—all possible values of Δ.

Since $P(\Delta)$ is a rather complicated object, it is in practice simpler to integrate $\rho(x)$ to compute the probability that Δ falls between any two values—in other words a histogram of Δ. In fact, we can do something more sophisticated: we can use any kernel we like to calculate a kernel density estimate of the distribution of Δ. For a general kernel function $\kappa(x)$ with Laplace transform $F_\kappa(s)$, we have

$$\rho_\kappa(x) = \sum_\Delta P(\Delta)\kappa(x - \Delta) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_\Delta(is)F_\kappa(is) e^{isx} \, ds. \tag{3.19}$$

A conventional histogram is equivalent to using a rectangular ('top hat') kernel, but for our figures we use a smoother double-exponential kernel (also known as a Laplace distribution):

$$\kappa(x) = \frac{1}{2b}e^{-|x|/b}, \tag{3.20}$$

where the parameter b sets the width of the distribution. (We arbitrarily pick $b = 1/3$.) The Laplace transform for this choice of $\kappa(x)$ is

$$F_\kappa(s) = \frac{1}{1 - b^2s^2}. \tag{3.21}$$

Figure 2 shows the distribution of Δ computed in this way for configuration models with the truncated power-law degree distribution

$$p_k \propto k^{-\alpha}e^{-\beta k} \tag{3.22}$$

and three different choices of the parameters α and β. Each example has the same mean degree but, as the figure shows, the distributions of Δ are quite different.

3.3 *Random graphs with degree correlations*

The configuration model of the previous section improves on the Poisson random graph by allowing arbitrary distributions of node degrees, but like the random graph, it lacks any correlation or assortativity between the degrees of adjacent nodes. Such correlation is common in real-world networks [15–17] and will clearly impact friendship paradox phenomena.

One can create a model network with degree correlations by fixing not only the degree distribution p_k as in the configuration model, but the joint distribution of adjacent degrees Q_{jk}, which is the fraction of edges that join nodes of degrees j and k [15]. Note that the distribution of degrees at the end of an edge is then given by $q_k = \sum_j Q_{jk}$, so that fixing Q_{jk} also fixes the degree distribution.
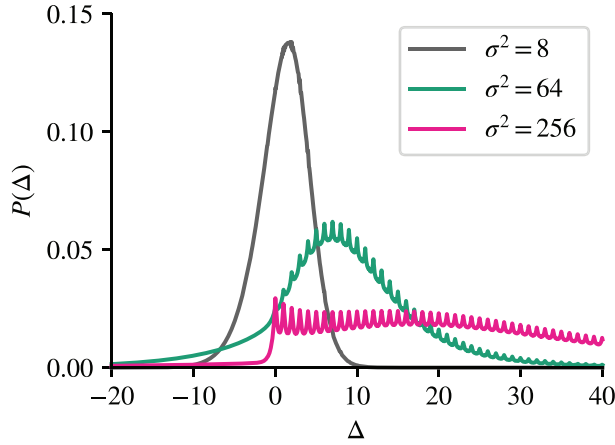
FIG. 2. The distribution of Δ_i for three configuration models with the truncated power-law degree distribution $p_k \propto k^{-\alpha} e^{-\beta k}$ and different choices of the parameters α and β. Δ_i can take only rational values, but for clarity we show a kernel density estimate of the distribution, calculated from Eq. (3.19) with a Laplace distribution kernel. The values of α and β were chosen so that the mean degree is always 8, while the variance is 8, 64 or 256 as indicated.

The calculation of the distribution of Δ proceeds in a similar manner to that for the configuration model. If we follow an edge that begins at a node of degree k, it will end up at a node of degree j with probability Q_{jk}/q_k, meaning that the Laplace transform for the degrees of neighbours is

$$G_k(s) = \sum_j \frac{Q_{jk}}{q_k} e^{-sj}. \tag{3.23}$$

Note that this function depends on the degree k of the node at which we started. Nevertheless, the calculation proceeds essentially as before. The equivalent of Eq. (3.17) is

$$F_\Delta(s) = \sum_k p_k \, e^{sk} \, G_k(s/k)^k, \tag{3.24}$$

and the distribution of Δ can be calculated from $F_\Delta(s)$ using Eq. (3.19).

Degree correlations are commonly quantified using an assortativity coefficient r, defined as the Pearson correlation coefficient of degrees across edges [15]. In terms of the quantities defined here,

$$r = \frac{\sum_{jk} jk(Q_{jk} - q_j q_k)}{\sigma_q^2}, \tag{3.25}$$

where σ_q^2 is the variance of the distribution q_k. To study how the effects of the friendship paradox vary with varying r, it is convenient to define a model network that allows us to adjust r, in effect a correlated version of the configuration model, parameterized by its degree distribution and a single extra parameter controlling the assortativity. This means choosing a suitable value of Q_{jk}, which we do by maximizing
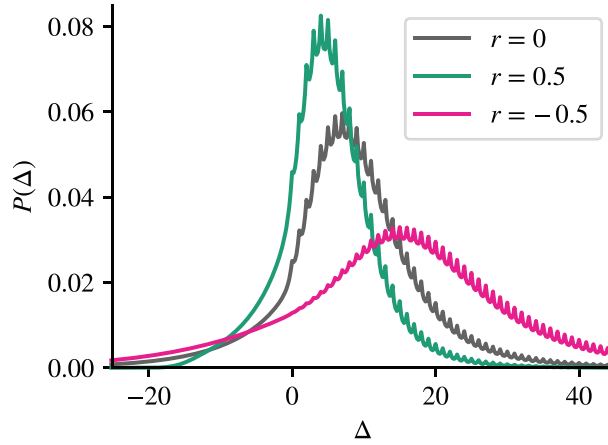
FIG. 3. Distribution for Δ in a random graphs with p_k defined by Eq. (3.22). The mean degree $\langle k \rangle = 8$ and $\sigma_k^2 = 64$. Three different values of γ were chosen so that the assortativity coefficient r takes the values $0, 0.5$ and -0.5.

the entropy

$$S(Q) = -\sum_{jk} Q_{jk} \log Q_{jk} \qquad (3.26)$$

subject to the constraints $\sum_j Q_{jk} = q_k = kp_k/\sum_j jp_j$ for all k and $\sum_{jk} jk(Q_{jk} - q_jq_k)/\sigma_q^2 = r$. The maximum entropy distribution is often considered to be the least biased choice for a given set of constraints, meaning that it makes no assumptions other than those implied by the constraints themselves.

The maximum entropy solution for Q_{jk} in this case is

$$Q_{jk} = \frac{e^{\gamma jk}}{Z_j Z_k} q_j q_k \,, \qquad (3.27)$$

where γ is a Lagrange multiplier whose value controls the assortativity and the Z_k are normalizing constants that satisfy the equations

$$Z_k = \sum_j \frac{q_j e^{\gamma jk}}{Z_j}, \qquad (3.28)$$

which we solve numerically by iteration. This model defines an ensemble of random networks with a desired level of assortativity and allows us to study the generic effects of assortativity on network properties, including the friendship paradox.

Figure 3, for example, shows the probability distribution of Δ for fixed degree distribution and three different choices of the assortativity coefficient r. Note how the average value of Δ decreases as the networks become more assortative, but so too does the variance, leading to a more complex picture. This is a good example of why we should be wary of conclusions based on the average value of Δ alone.

Figure 4 sheds more light on this point, showing the average value along with the expected fraction $P(\Delta > 0)$ of nodes with positive Δ, both as a function of r. Both of these quantities can be viewed as
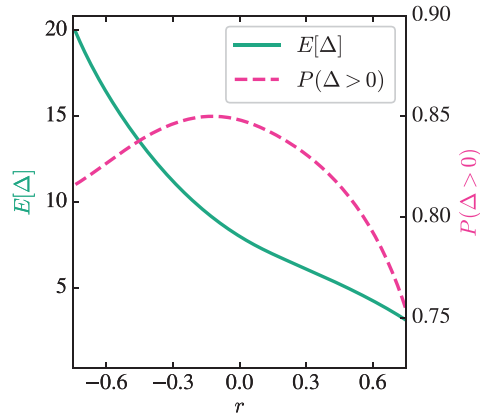
FIG. 4. Two measures of the magnitude of the friendship paradox as a function of assortativity. The degree distribution is that of Eq. (3.22) with α and β chosen so that $\langle k \rangle = 8$ and $\sigma_k^2 = 64$.

measures of the strength of the friendship paradox, but they behave in different ways. While $E[\Delta]$ does indeed decrease monotonically with assortativity, the fraction of nodes with positive Δ—in effect, the fraction of nodes that display the classic friendship paradox behaviour—peaks at small negative r, and has lower values for both large positive and large negative r.

3.4 *Comparison with real-world networks*

How good a guide are these model calculations to the behaviour of real-world networks? To shed light on this question, we compare the mean and variance of Δ in 32 real-world social networks with values calculated from the assortative network model of the previous section with the same r. (The mean and variance in the model can be computed from the first and second derivatives of Eq. (3.24).) The results are shown in Fig. 5 and, as we can see, there is remarkably good agreement between theoretical and empirical results—we find R^2 values of 0.93 and 0.99 between theory and experiment for the mean and standard deviation of Δ, respectively. By comparison, the standard configuration model, which fixes the degree distribution only, gives R^2 values of 0.77 and 0.95. Thus, it would be fair to say that the distribution of Δ is fairly accurately captured by the degree distribution alone, but that the inclusion of assortativity results in a significant improvement.

3.5 *Generalized friendship paradox*

Before finishing let us return to the generalized friendship paradox. Recall that for any quantity x defined on the nodes of a network the quantity $\Delta_i^{(x)}$, Eq. (2.4), measures the difference between the average value of x at node i's neighbours and i's own value. We can compute the distribution of $\Delta^{(x)}$, for instance for the degree-correlated model of Section 3.3, using the Laplace transform formalism again. The argument differs from previous developments in some details but remains conceptually similar. One first writes the Laplace transform for the distribution of a node's value of x given its degree k

$$H_k(s) = \int_{-\infty}^{\infty} P(x|k)\, e^{-sx}\, dx, \tag{3.29}$$
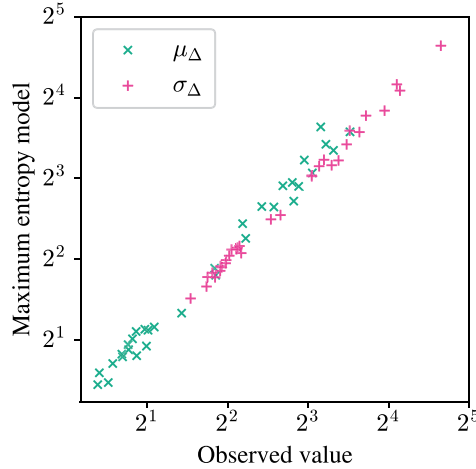
FIG. 5. Mean and standard deviation of Δ in 32 real-world social networks [18–33] compared to predictions from the maximum-entropy model, Eq. (3.27).

where $P(x|k)$ is the probability that a node of degree k has value x. Since the neighbours of a degree-k node have degree distributed as Q_{jk}/q_k, they have a distribution of x values with Laplace transform

$$G_k(s) = \sum_j \frac{Q_{kj}}{q_k} H_j(s). \tag{3.30}$$

Applying the key properties in Eqs (3.13) and (3.14), we then arrive at

$$F_{\Delta^{(x)}}(s) = \sum_k p_k G_k(s/k)^k H_k(-s), \tag{3.31}$$

and inverting $F_{\Delta^{(x)}}$ gives the distribution for $\Delta^{(x)}$. Performing calculations such as these allows one to analytically compare real networks to models, alleviating the need for simulations, such as those of [34].

As an example, we have tested Eq. (3.31) for normally distributed x linearly correlated with degree and find behaviour closely similar to that of Fig. 3. Behaviour like this could have an impact in any situation where the generalized friendship paradox has practical consequences. For example, it has been found that people's individual well-being can be substantially affected by the behaviour of their network neighbours. People whose acquaintances smoke are more likely to smoke themselves [35]. People whose friends appear to be better off than they are may develop a lower sense of self-worth [36, 37]. If people with more acquaintances tend to smoke more, or if well-off people with exciting lives have a lot of friends or followers on social media, then we may have a generalized friendship paradox in which you are most likely to have contact with precisely those people who would adversely affect you.

As another example, it has been shown that polling forecasts of election outcomes can be significantly improved by focusing not on how study participants say they will vote but on how they expect their acquaintances to vote [38], in part because this reduces variance in the estimates of outcomes. If voting intention is subject to the generalized friendship paradox however—if for instance partisan inclination is correlated with degree—then the tendency for one's friends to have a high degree will cause the resulting

sample of the population to be biased and introduce systematic errors [39]. The formalism developed here allows us to quantify these effects not only in terms of the average individual but in terms of the complete distribution of outcomes over the entire population.

4. Conclusions

In this article, we have quantified the friendship and generalized friendship paradoxes in terms of the difference Δ between the characteristics of a node and the average of the same characteristics for the node's neighbours. Previous studies have examined the mean of this difference but, as we have argued here, to get a full picture one must examine the complete distribution of values. We have performed theoretical calculations of this distribution for three classes of model networks, the Poisson random graph, the configuration model and a model of a random degree-assortative network. Among other things, our results indicate that the friendship paradox will tend to be strongest in networks with very heterogeneous degree distributions and negative assortativity. Conversely, the effects will tend to be muted when degrees are fairly homogeneous and the network is degree assortative. On the other hand, we have also seen that even in simple network models the distribution for Δ can be widely dispersed, meaning that the average value offers an incomplete description of the behaviour.

We have also compared our results with a selection of real-world networks, finding remarkably good agreement between theoretical predictions and empirical measurements, particularly in the case of the model that incorporates assortativity.

REFERENCES

1. FELD, S. (1991) Why your friends have more friends than you do. *Am. J. Sociol.,* **96**, 1464–1477.
2. JO, H.-H. & EOM, Y.-H. (2014) Generalized friendship paradox in networks with tunable degree-attribute correlation. *Phys. Rev. E,* **90**, 022809.
3. EOM, Y.-H. & JO, H.-H. (2014) Generalized friendship paradox in complex networks: the case of scientific collaboration. *Sci. Rep.,* **4**, 4603.
4. HODAS, N. O., KOOTI, F. & LERMAN, K. (2013) Friendship paradox redux: your friends are more interesting than you. *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media.* Palo Alto, CA: AAAI Press.
5. NEWMAN, M. (2018) *Networks*, 2nd edn. Oxford: Oxford University Press.

6. MOMENI, N. & RABBAT, M. (2016) Qualities and inequalities in online social networks through the lens of the generalized friendship paradox. *PLoS One* **11**, e0143633.

7. GILBERT, E. N. (1959) Random graphs. *Ann. Math. Stat.,* **30**, 1141–1144.

8. ERDŐS, P. & RÉNYI, A. (1960) On the evolution of random graphs. *Public. Math. Inst. Hungar. Acad. Sci.,* **5**, 17–61.

9. BOLLOBÁS, B. (2001) *Random Graphs*, 2nd edn. New York: Academic Press.

10. BOLLOBÁS, B. (1980) A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.,* **1**, 311–316.

11. NEWMAN, M. E. J., STROGATZ, S. H. & WATTS, D. J. (2001) Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E,* **64**, 026118.

12. BARABÁSI, A.-L. & ALBERT, R. (1999) Emergence of scaling in random networks. *Science,* **286**, 509–512.

13. AMARAL, L. A. N., SCALA, A., BARTHÉLEMY, M. & STANLEY, H. E. (2000) Classes of small-world networks. *Proc. Natl. Acad. Sci. USA,* **97**, 11149–11152.

14. WILF, H. (1994) *Generating Functionology*, 2nd edn. London: Academic Press.

15. NEWMAN, M. E. J. (2002) Assortative mixing in networks. *Phys. Rev. Lett.,* **89**, 208701.

16. NEWMAN, M. E. J. & PARK, J. (2003) Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122.

17. HU, H.-B. & WANG, X.-F. (2009) Disassortative mixing in online social networks. *Europhys. Lett.* **86**, 18003.

18. GLEISER, P. & DANON, L. (2003) Community structure in jazz. *Adv. Complex Syst.,* **6**, 565–573.

19. BOGUÑÁ, M., PASTOR-SATORRAS, R., DÍAZ-GUILERA, A. & ARENAS, A. (2004) Models of social networks based on social distance attachment. *Phys. Rev. E,* **70**, 056122.

20. NEWMAN, M. E. J. (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E,* **74**, 036104.

21. NEWMAN, M. E. J. (2001) The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA,* **98**, 404–409.

22. GUIMERÀ, R., DANON, L., DÍAZ-GUILERA, A., GIRALT, F. & ARENAS, A. (2003) Self-similar community structure in a network of human interactions. *Phys. Rev. E,* **68**, 065103.

23. HARRIS, K. M. & UDRY, J. R. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994–2008, Public Use Version 21 (2008).

24. MAIER, B. F. & BROCKMANN, D. (2017) Cover time for random walks on arbitrary complex networks. *Phys. Rev. E,* **96**, 042307.

25. KNUTH, D. E. (1993) *The Stanford GraphBase: A Platform for Combinatorial Computing.* Reading, MA: Addison-Wesley.

26. BEVERIDGE, A. & SHAN, J. (2016) Network of thrones. *Math Horizons,* **23**, 18–22.

27. WEEKS, M. R., CLAIR, S., BORGATTI, S. P., RADDA, K. & SCHENSUL, J. J. (2002) Social networks of drug users in high-risk sites: finding the connections. *AIDS Behav.,* **6**, 193–206.

28. LUSSEAU, D., SCHNEIDER, K., BOISSEAU, O. J., HAASE, P., SLOOTEN, E. & DAWSON, S. M. (2003) The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait? *Behav. Ecol. Sociobiol.,* **54**, 396–405.

29. FIRE, M., KATZ, G., ELOVICI, Y., SHAPIRA, B., & ROKACH, L. (2012) Predicting student exam's scores by analyzing social network data. *Active Media Technology* (R. Huang, A. A. Ghorbani, G. Pasi, T. Yamaguchi, N. Y. Yen, & B. Jin eds) Berlin: Springer, pp. 584–595.

30. TANG, J., SUN, J., WANG, C. & YANG, Z. (2009) Social influence analysis in large-scale networks. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09. New York, NY: ACM, pp. 807–816.

31. EDWARDS, G. (2015) FIFA networks. https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/fifa.

32. SEIERSTAD, C. & OPSAHL, T. (2011) For the few not the many? The effects of affirmative action on presence, prominence, and social capital of women directors in Norway. *Scand. J. Manag.,* **27**, 44–54.

33. MASTRANDREA, R., FOURNET, J. & BARRAT, A. (2015) Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS One,* **10**, e107878.

34. KOOTI, F., HODAS, N. O. & LERMAN, K. (2014) Network weirdness: exploring the origins of network paradoxes. *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. Palo Alto, CA: AAAI Press.

35. CRISTAKIS, N. A. & FOWLER, J. H. (2008) The collective dynamics of smoking in a large social network. *N. Engl. J. Med.,* **358**, 2249–2258.

36. KROSS, E., VERDUYN, P., DEMIRALP, E., PARK, J., LEE, D. S., LIN, N., SHABLACK, H., JONIDES, J. & YBARRA, O. (2013) Facebook use predicts declines in subjective well-being in young adults. *PLoS One,* **8**, e69841.

37. LERMAN, K., YAN, X. & WU, X.-Z. (2016) The majority illusion in social networks. *PLoS One,* **11**, e147616.

38. NETTASINGHE, B. & KRISHNAMURTHY, V. (2019) What do your friends think? Efficient polling methods for networks using friendship paradox. *IEEE Trans. Knowl. Data Eng.*, **33**, 1291–1305.

39. ROTHSCHILD, D. (2009) Forecasting elections: comparing prediction markets, polls, and their biases. *Public Opin. Q.,* **73**, 895–916.