Centrality-friendship paradoxes: when our friends are more important than us

DESMOND J. HIGHAM[†]

Department of Mathematics and Statistics, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, UK [†]Corresponding author. Email: d.j.higham@strath.ac.uk

Edited by: Michele Benzi

The friendship paradox states that, on average, our friends have more friends than we do. In network terms, the average degree over the nodes can never exceed the average degree over the neighbours of nodes. This effect, which is a classic example of sampling bias, has attracted much attention in the social science and network science literature, with variations and extensions of the paradox being defined, tested and interpreted. Here, we show that a version of the paradox holds rigorously for eigenvector centrality: on average, our friends are more important than us. We then consider general matrix-function centrality, including Katz centrality, and give sufficient conditions for the paradox to hold. We also discuss which results can be generalized to the cases of directed and weighted edges. In this way, we add theoretical support for a field that has largely been evolving through empirical testing.

Keywords: combinatorics; degree; eigenvector centrality; friendship paradox; graph theory; Katz centrality; matrix function; Perron-Frobenius theory; sampling bias; walk

1. Motivation

Consider the graph in Fig. 1. Imagine that the nodes represent people and the edges represent reciprocated friendships. Nodes 1, 2, 3, 4, 5, 6, 7, 8 have 4, 1, 1, 1, 3, 2, 3, 1 friends, respectively. So the average number of friends possessed by a node is $16/8 = 2$. Now look at the friends of each node. The four friends of node 1 possess {1, 1, 1, 3} friends. Similarly for nodes 2, 3, 4, 5, 6, 7, 8 we find {4}, {4}, {4}, {4, 2, 3}, {3, 3}, {2, , 3, 1} and {3}, respectively. So the average number of friends possessed by a friend is $42/16 = 2.625$, which is greater than 2. This effect—that *our friends have more friends than we do, on average*—was identified by Feld [1] and has become known as the *friendship paradox*. Feld showed that the friend-of-friend average always dominates the friend average, with equality if and only if all individuals have the same number of friends. The paradox is a classic example of sampling bias. In Fig. 1, node 1 has 4 friends and hence appears 4 times in the friend-of-friend sum, whereas node 2 only contributes its value 1 on a single occasion; in general, highly connected nodes have a greater influence on the sum.

The friendship paradox has motivated much activity in the social network literature and is also mentioned regularly in the wider media; see, for example, [2]. Researchers have measured the extent to which the discrepancy holds on real networks involving, for example, high school and university students [1, 3, 4], scientific coauthors [5], plants and pollinators [6] and users of social media [7–9]. (We mention that some of these studies also looked at individual-level analogues, such as 'what proportion of nodes have fewer friends than the average over their friends'? In this work, we focus exclusively on the global averages used in the original reference [1].)
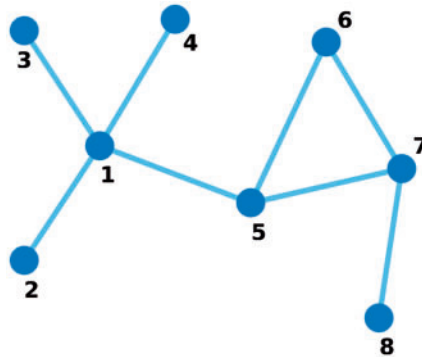
Fig. 1. Simple undirected network with eight nodes.

Extending this idea, Eom and Jo [5] looked at the case where each node may be quantified according to some externally derived attribute and studied the *generalized friendship paradox*: on average, do our friends have more of this attribute than us? They showed that the answer is yes for attributes that correlate positively with the number of friends and found the effect to hold empirically for certain scientific collaboration networks in the case where the attribute was publication or citation count. Similarly, Hodas *et al.* [8] made empirical studies of Twitter networks and tested for a friend activity paradox (are our friends more active than us?) and for a virality paradox (do our friends spread more viral content than us?).

Our aim here is to study the generalized friendship paradox in the case where the attribute is importance, as quantified by a network centrality measure. Aside from the fact that centrality is a fundamental and informative nodal property [10–12], we also note that centrality measures are defined explicitly in terms of the network topology, and hence there is potential to derive results that hold universally, or at least for some well-defined classes of network. This allows us to add further theoretical backing that complements the recent data-driven studies mentioned above. Our results also alleviate the need for certain experiments. For example, in [3] Grund tested the eigenvector centrality version of the generalized friendship paradox on two small-scale friendship networks; Theorem 3.1 in Section 3 shows that this paradox holds for *all* networks.

This manuscript is organized as follows. In Section 2, we set up some notation, formalize the friendship paradox and explain how it follows directly from the Cauchy–Schwarz inequality. We also define the generalized friendship paradox from [5] and show how it arises when the quantity of interest correlates with degree. The new material starts in Section 3, where we show that a paradox always holds for eigenvector centrality. In Section 4, we consider other types of network centrality based on matrix functions. Using a combinatorial result from [13] we show that the paradox holds for certain types of matrix function. We also derive and interpret sufficient conditions for general matrix functions defined through power series with non-negative coefficients, including the resolvent case corresponding to Katz centrality. Sections 3 and 4 deal with undirected, unweighted networks. In Section 5, we look at directed networks, where the picture is less straightforward. We discuss various paradoxes that arise from the use of out-degree and in-degree, and give some sufficient conditions for centrality-based analogues. We explain in Section 6 how all results extend readily to the case of non-negatively weighted networks. We conclude in Section 7 with an overview of the main results and an indication of possible future lines of pursuit.

2. Friendship paradox and generalized friendship paradox

Suppose A represents the adjacency matrix for an undirected, unweighted, network with n nodes. (Directed edges will be considered in Section 5 and weighted edges in Section 6.) So $A \in \mathbb{R}^{n \times n}$, with $a_{ii} = 0$ and with $a_{ij} = a_{ji} = 1$ if nodes i and j are connected. To avoid the trivial special case of an empty network, we assume at least one edge exists. Letting $\mathbf{1} \in \mathbb{R}^n$ denote the vector with all components equal to one, we may define the degree vector

$$\mathbf{d} = A\mathbf{1},$$

where d_i gives the degree of node i.

We will make use of the two-norm and the one-norm, which for a vector $\mathbf{x} \in \mathbb{R}^n$ are defined by

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top \mathbf{x}} \quad \text{and} \quad \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|,$$

respectively. We note that in many cases we will be dealing with a non-negatively valued vector $\mathbf{x}$, whence the one-norm reduces to the sum of the entries.

In this notation, the average degree over the nodes may be written

$$\frac{\|\mathbf{d}\|_1}{n}.$$

In the friendship paradox, we wish to compare this quantity with the average of the values that arise when we take each node, look at each of that node's neighbours, and record how many neighbours those neighbours have. When we do this count, each node i appears as a neighbour d_i times and each time it contributes d_i neighbours, so the count totals $\mathbf{d}^\top \mathbf{d}$. The number of terms in the count is twice the number of edges, which is $\|\mathbf{d}\|_1$. The overall friend-of-friend average is therefore

$$\frac{\mathbf{d}^\top \mathbf{d}}{\|\mathbf{d}\|_1}.$$

So the friendship paradox is equivalent to the inequality

$$\frac{\mathbf{d}^\top \mathbf{d}}{\|\mathbf{d}\|_1} - \frac{\|\mathbf{d}\|_1}{n} \geq 0. \tag{1}$$

To see why (1) is always true, we recall from the Cauchy–Schwarz inequality [14] that for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have

$$\mathbf{u}^\top \mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2. \tag{2}$$

Taking $\mathbf{u} = \mathbf{1}$, this implies

$$\|\mathbf{v}\|_1 \leq \sqrt{n}\, \|\mathbf{v}\|_2. \tag{3}$$

For $\|\mathbf{v}\|_1 \neq 0$, after squaring and rearranging we may write this inequality in the form

$$\frac{\mathbf{v}^\top \mathbf{v}}{\|\mathbf{v}\|_1} - \frac{\|\mathbf{v}\|_1}{n} \geq 0. \tag{4}$$

So we see that the friendship paradox inequality (1) is always satisfied.

Further, equality holds in the Cauchy–Schwarz inequality (2) if and only if $\mathbf{u}$ is a multiple of $\mathbf{v}$. So we have equality in the friendship paradox inequality (1) if and only if the network is regular—all nodes have the same degree.

To define the generalized friendship paradox [5] a non-negative quantity $x_i \geq 0$ is assigned to node i, and we compare the average over the nodes,

$$\frac{\|\mathbf{x}\|_1}{n},$$

with the average over neighbours of nodes,

$$\frac{\mathbf{x}^\top \mathbf{d}}{\|\mathbf{d}\|_1}.$$

The numerator $\mathbf{x}^\top \mathbf{d}$ in the latter quantity arises because in the overall sum each node i contributes its value x_i a total of d_i times. The denominator $\|\mathbf{d}\|_1$ arises because each edge is used twice. Hence, we may say that a generalized friendship paradox with respect to the quantity $\mathbf{x}$ arises if

$$\frac{\mathbf{d}^\top \mathbf{x}}{\|\mathbf{d}\|_1} - \frac{\|\mathbf{x}\|_1}{n} \geq 0. \tag{5}$$

Note that the original version (1) corresponds to the case where $\mathbf{x}$ is the degree vector.

Introducing the covariance between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ as

$$\mathrm{Cov}(\mathbf{u}, \mathbf{v}) = \frac{1}{n} \sum_{i=1}^{n} (u_i - \mu_u)(v_i - \mu_v),$$

where

$$\mu_u = \frac{1}{n} \sum_{i=1}^{n} u_i \quad \text{and} \quad \mu_v = \frac{1}{n} \sum_{i=1}^{n} v_i$$

denote the corresponding means, we may rewrite (5) as

$$\frac{1}{\mu_d} \mathrm{Cov}(\mathbf{d}, \mathbf{x}) \geq 0.$$

Hence, as indicated in [5], a generalized friendship paradox with respect to the quantity $\mathbf{x}$ arises if $\mathbf{x}$ is non-negatively correlated with degree.

Rather than considering an externally defined attribute, as was done in the tests of [5, 8], we will look at circumstances where $\mathbf{x}$ is a network centrality measure that quantifies the relative importance of each

node. In this way we can address, in the same generality as the original work [1], the question: *are our friends more important than us, on average?*

3. Eigenvector centrality paradox

In this section we consider the case of eigenvector centrality [10, 12, 15–17]. To make this centrality measure well-defined, we assume that the network is connected, and hence the symmetric matrix A is irreducible. From Perron–Frobenius theory [14], we know that A has a real, positive, dominant eigenvalue λ_1 that is equal to $\|A\|_2$, the matrix two-norm of A. The centrality measure is then given by the corresponding Perron–Frobenius eigenvector $\mathbf{x}$, which satisfies $A\mathbf{x} = \lambda_1 \mathbf{x}$ and has positive elements.

We have the following result.

THEOREM 3.1 Given any connected network, the generalized friendship paradox inequality (5) holds for eigenvector centrality, with equality if and only if the network is regular.

Proof. Here, $\mathbf{x}$ denotes the Perron–Frobenius eigenvector corresponding to the eigenvalue λ_1. From the definition of the subordinate matrix two-norm, we have

$$\lambda_1 = \|A\|_2 \geq \|A\frac{\mathbf{1}}{\sqrt{n}}\|_2 = \|\frac{\mathbf{d}}{\sqrt{n}}\|_2. \tag{6}$$

Using (3), this implies

$$\lambda_1 \geq \frac{1}{\sqrt{n}}\|\frac{\mathbf{d}}{\sqrt{n}}\|_1. \tag{7}$$

Hence,

$$\mathbf{d}^\top \mathbf{x} = \mathbf{1}^\top A\mathbf{x} = \mathbf{1}^\top \lambda_1 \mathbf{x} = \lambda_1 \|\mathbf{x}\|_1 \geq \frac{\|\mathbf{d}\|_1}{n}\|\mathbf{x}\|_1,$$

and we see that (5) always holds. Further, because $\mathbf{x}$ is the only eigenvector whose elements are all positive, we have equality in (6) if and only if $\mathbf{x}$ is a multiple of $\mathbf{1}$; that is, if and only if the network is regular. □

4. Matrix function centrality paradox

We now move on to the case where $\mathbf{x}$ is defined from a power series expansion

$$\mathbf{x} = \left(c_0 I + c_1 A + c_2 A^2 + \cdots\right)\mathbf{1}. \tag{8}$$

Here, we assume that $c_k \geq 0$ for all k and that these coefficients have been chosen in such a way that the series converges. Centrality measures of this type have been studied by several authors, see, for example, [10, 12, 17–23]. They can be motivated from the combinatoric fact that $(A^k)_{ij}$ counts the number of distinct walks of length k between i and j. Particular examples are

- *Katz centrality* [24], where $c_k = \alpha^k$. Here the real parameter α must be chosen such that $0 < \alpha < 1/\rho(A)$, where $\rho(A)$ denotes the spectral radius of A. In this case $\mathbf{x}$ solves the linear system $(I - \alpha A)\mathbf{x} = \mathbf{1}$.

- *total centrality* [18, 19], where $c_k = \beta^k/k!$ for some positive real parameter β. In this case the series converges for any β, and $\mathbf{x}$ may be written $\mathbf{x} = \exp(\beta A)\mathbf{1}$. Other factorial-based coefficients have also been proposed [20].

- *odd and even centralities* based on odd and even power series, such as those for sinh and cosh [23].

We also note that degree centrality, on which the original friendship paradox is based, corresponds to $c_1 = 1$ in (8) with all other coefficients equal to zero.

Our aim in this section is to look for conditions under which the generalized friendship paradox (5) holds for centrality measures of the form (8).

For $\mathbf{x}$ in (8) we have

$$\|\mathbf{x}\|_1 = c_0 n + c_1 \mathbf{1}^\top A \mathbf{1} + c_2 \mathbf{1}^\top A^2 \mathbf{1} + \cdots.$$

and

$$\mathbf{d}^\top \mathbf{x} = c_0 \mathbf{1}^\top A \mathbf{1} + c_1 \mathbf{1}^\top A^2 \mathbf{1} + c_2 \mathbf{1}^\top A^3 \mathbf{1} + \cdots.$$

So the generalized friendship paradox inequality (5) may be written

$$\left(c_0 \mathbf{1}^\top A \mathbf{1} + c_1 \mathbf{1}^\top A^2 \mathbf{1} + c_2 \mathbf{1}^\top A^3 \mathbf{1} + \cdots\right) - \left(c_0 n + c_1 \mathbf{1}^\top A \mathbf{1} + c_2 \mathbf{1}^\top A^2 \mathbf{1} + \cdots\right) \frac{\mathbf{1}^\top A \mathbf{1}}{n} \geq 0. \qquad (9)$$

By comparing terms in the two expansions, we arrive at the following sufficient condition.

THEOREM 4.1 The generalized friendship paradox inequality (5) holds for $\mathbf{x}$ in (8) if

$$\mathbf{1}^\top A^{k+1} \mathbf{1} \geq \mathbf{1}^\top A^k \mathbf{1} \frac{\mathbf{1}^\top A \mathbf{1}}{n}, \qquad (10)$$

for every $k \geq 1$ for which $c_k > 0$.

Proof. We see that the term on the left-hand side of (9) involving c_0 collapses to zero. Generally, we may obtain a sufficient condition by asking for each individual term involving c_k to be greater than or equal to zero, for all $k \geq 1$. This leads to (10). □

We note that the sufficient condition (10) has a simple combinatoric interpretation: *the total number of walks of length $k + 1$ must dominate the product of the total number of walks of length k and the average degree.*

To proceed we make use of the following result.

THEOREM 4.2 (Lagarias et al. [13]) For any positive integers r and s such that $r + s$ is even, we have

$$\mathbf{1}^\top A^{r+s} \mathbf{1} \geq \frac{\mathbf{1}^\top A^r \mathbf{1} \mathbf{1}^\top A^s \mathbf{1}}{n}. \qquad (11)$$

Proof. See [13, Theorem 1]. □

In words, Theorem 4.2 says that, for $r + s$ even, *the total number of walks of length $r + s$ dominates the product of the total number of walks of length r and the total number of walks of length s, scaled by the number of nodes, n.*

This theorem allows us to deal with odd power series:

THEOREM 4.3 The generalized friendship paradox inequality (5) holds for **x** in (8) in the case where $c_k = 0$ for k even, with equality if and only if the network is regular.

Proof. First, suppose the network is regular. Let deg denote the common degree, so that $A\mathbf{1} = \deg\mathbf{1}$. Since **1** is an eigenvector with positive entries, it must be the Perron–Frobenius eigenvector, so $\deg = \lambda_1 = \|A\|_2$. Then $\mathbf{1}^\top A^k \mathbf{1} = \mathbf{1}^\top \lambda_1^k \mathbf{1} = n\lambda_1^k$ for all $k \geq 0$. It follows that for each c_k, the term on the left-hand side of (9) collapses to zero, giving equality.

Now suppose that the network is not regular. On the left-hand side of (9), the coefficient c_1 receives the factor $\mathbf{d}^\top\mathbf{d} - \|\mathbf{d}\|_1^2/n$. This quantity relates to the original friendship paradox—see (1)—and is strictly positive by the Cauchy–Schwarz inequality. To deal with the remaining terms, it is then enough to show that (10) holds for odd $k > 1$. This is done by taking $r = k$ and $s = 1$ in (11). □

The next result focuses on Katz centrality. The $\alpha \to 0^+$ limit of Katz corresponds to degree. So it is intuitively reasonable that 'by continuity' the original friendship paradox should transfer to Katz for sufficiently small α. This theorem formalizes the argument.

THEOREM 4.4 Consider the case where $c_k = \alpha^k$ in (8). For any network there exists a value $\alpha^\star > 0$ such that the generalized friendship paradox inequality (5) holds for all parameter values $0 < \alpha < \alpha^\star$, with equality if and only if the network is regular.

Proof. First, suppose the network is regular. We are using deg to denote the common degree, so $\mathbf{d} = \deg\mathbf{1}$. It follows that $\mathbf{x} = \mathbf{1}/(1 - \alpha\deg)$ is the unique solution to the Katz centrality equation $(I - \alpha A)\mathbf{x} = \mathbf{1}$ for all $0 < \alpha < 1/\deg$. Because **x** is a multiple of the degree vector, we have equality in (5).

Now suppose that the network is not regular, so **d** is not a multiple of **1**. With $c_k = \alpha^k$, and α small, the left-hand side of (9) may be expanded as

$$\alpha \left(\mathbf{d}^\top\mathbf{d} - \frac{\|\mathbf{d}\|_1^2}{n} \right) + O(\alpha^2),$$

and we see that the factor in parentheses is strictly positive. □

5. Directed networks

In this section, we consider the case of unweighted directed networks, so A is no longer assumed to be symmetric. To be concrete when discussing results, we imagine that the network represents human–human follower relationships on a social media platform. So an edge from i to j, represented by $a_{ij} = 1$, indicates that *person i follows person j.*

We define the out-degree vector and in-degree vector by

$$\mathbf{d}^{\text{out}} = A\mathbf{1} \qquad \text{and} \qquad \mathbf{d}^{\text{in}} = A^\top\mathbf{1},$$

respectively. Hence, d_i^{out} counts the number of people that person i follows, and d_j^{in} counts the number of people who follow person j. Note that

$$\|\mathbf{d}^{\mathrm{out}}\|_1 = \|\mathbf{d}^{\mathrm{in}}\|_1 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}.$$

Our first observation is that the inequality (4) holds for any non-zero vector $\mathbf{v}$, with equality if and only if $\mathbf{v}$ is a multiple of $\mathbf{1}$. Hence we have

$$\frac{\mathbf{d}^{\mathrm{out}\top}\mathbf{d}^{\mathrm{out}}}{\|\mathbf{d}^{\mathrm{out}}\|_1} - \frac{\|\mathbf{d}^{\mathrm{out}}\|_1}{n} \geq 0 \quad \text{and} \quad \frac{\mathbf{d}^{\mathrm{in}\top}\mathbf{d}^{\mathrm{in}}}{\|\mathbf{d}^{\mathrm{in}}\|_1} - \frac{\|\mathbf{d}^{\mathrm{in}}\|_1}{n} \geq 0. \tag{12}$$

A little care is needed when interpreting these inequalities. The total $\mathbf{d}^{\mathrm{out}\top}\mathbf{d}^{\mathrm{out}}$ arises if we take each person in turn, look at the people who follow them, and record how much following these people do. (In this way, each person k shows up d_k^{out} times as a follower, each time contributing the number of people they follow, d_k^{out}.) Similarly, $\mathbf{d}^{\mathrm{in}\top}\mathbf{d}^{\mathrm{in}}$ arises if we take each person in turn, look at the people who they follow, and record how many times these people are followed. (In this way, each person k shows up d_k^{in} times as being followed, each time contributing the number of people who follow them, d_k^{in}.)

In words, it is always true that

(i) *our followers follow at least as many people as us, on average* (and there is equality if and only if everybody follows the same number of people), and

(ii) *the people we follow have at least as many followers as us, on average* (and there is equality if and only if everybody has the same number of followers).

From the discussion in Section 2, we also see that the in-out/out-in analogue of (12) and corresponding statements are valid only if $\mathrm{Cov}(\mathbf{d}^{\mathrm{out}}, \mathbf{d}^{\mathrm{in}}) \geq 0$. Simple examples where $\mathrm{Cov}(\mathbf{d}^{\mathrm{out}}, \mathbf{d}^{\mathrm{in}})$ is negative include the outward star graph where the only edges start at node 1 and end at nodes $2, 3, \ldots, n$, for which

$$\mathbf{d}^{\mathrm{out}} = \begin{bmatrix} n-1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{d}^{\mathrm{in}} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

and also the corresponding inward star graph. A strongly connected example has edges from node 1 to nodes $2, 3, \ldots, n$ and from node i to node $i+1$ for $i = 2, 3, \ldots, n-1$, plus an edge from node n to node 2 and an edge from node n back to node 1. Here, we have

$$\mathbf{d}^{\mathrm{out}} = \begin{bmatrix} n-1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{d}^{\mathrm{in}} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ \vdots \\ \vdots \\ 2 \end{bmatrix}. \tag{13}$$

In this case, $\mathbf{d}^{\text{out}\top}\mathbf{d}^{\text{in}} = 3n - 1$ and $\|\mathbf{d}^{\text{out}}\|_1 = \|\mathbf{d}^{\text{in}}\|_1 = 2n - 1$, so $\mathbf{d}^{\text{out}\top}\mathbf{d}^{\text{in}} - \|\mathbf{d}^{\text{out}}\|_1\|\mathbf{d}^{\text{in}}\|_1/n = -n + 3 - 1/n$, which is negative for $n \geq 3$. Hence, for such graphs it is **not true** that

(iii) *our followers have at least as many followers as us, on average*, or

(iv) *the people we follow are following at least as many people as us, on average.*

Taking a step back and comparing (i) and (ii) with (iii) and (iv), it is intuitively clear that the first two cases can contribute to a sampling bias, whereas the second two cannot. For example, in (i) by allowing a person to follow us we increase the number of people they follow, but in (iii) allowing a person to follow us has no effect on the number of people who follow them.

The reference [8] is unusual in that it tests the friendship paradox on directed networks. The authors consider the four versions (i)–(iv) and find that the paradox holds in each case for a large social network constructed from Twitter data. Our reasoning above shows that two of these versions will hold for *all* directed networks.

For an arbitrary network measure $\mathbf{x}$ the relevant inequality that describes the out-degree version of the generalized friendship paradox (5) is

$$\frac{\mathbf{d}^{\text{out}\top}\mathbf{x}}{\|\mathbf{d}^{\text{out}}\|_1} - \frac{\|\mathbf{x}\|_1}{n} \geq 0. \tag{14}$$

Similarly, the in-degree version is

$$\frac{\mathbf{d}^{\text{in}\top}\mathbf{x}}{\|\mathbf{d}^{\text{in}}\|_1} - \frac{\|\mathbf{x}\|_1}{n} \geq 0. \tag{15}$$

We now consider eigenvector centrality as our network measure. We assume that the network is strongly connected so that A is irreducible. In this directed case, we have potentially distinct left and right Perron–Frobenius eigenvectors, which we denote $\mathbf{x}_{\text{L}}$ and $\mathbf{x}_{\text{R}}$, respectively. Here, $A\mathbf{x}_{\text{R}} = \lambda_1\mathbf{x}_{\text{R}}$ and $\mathbf{x}_{\text{L}}^\top A = \lambda_1\mathbf{x}_{\text{L}}^\top$, with $\lambda_1 = \rho(A)$. Both vectors $\mathbf{x}_{\text{L}}$ and $\mathbf{x}_{\text{R}}$ have positive components.

The next result characterizes two cases.

THEOREM 5.1 For a strongly connected directed network the out-degree generalized friendship paradox inequality (14) holds for the case where $\mathbf{x} = \mathbf{x}_{\text{L}}$ if and only if

$$\lambda_1 \geq \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}}{n}. \tag{16}$$

Similarly, (16) also characterizes the in-degree generalized friendship paradox (15) where $\mathbf{x} = \mathbf{x}_{\text{R}}$.

Proof. When $\mathbf{x} = \mathbf{x}_{\text{L}}$ we have

$$\mathbf{d}^{\text{out}\top}\mathbf{x} = \mathbf{1}^\top A^\top\mathbf{x}_{\text{L}} = \mathbf{1}^\top\lambda_1\mathbf{x}_{\text{L}} = \lambda_1\|\mathbf{x}_{\text{L}}\|_1.$$

It follows that (14) reduces to (16).

The second statement may be proved by replacing A with $A^\top$. □

It is of interest to note that for our unsymmetric A, a classical result is that λ_1 lies between the minimum out-degree or in-degree and the maximum out-degree or in-degree [14, Theorem 8.1.22]. However, it is not true in general that λ_1 dominates the average in-degree (and hence average out-degree). An example is given by the strongly connected graph with adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

In this case λ_1 is the real root of $\lambda^3 - \lambda - 1$, which has the form

$$\lambda_1 = \frac{-1}{3}\left(C + \frac{3}{C}\right), \quad \text{for} \quad C = \sqrt[3]{\frac{-27 + \sqrt{621}}{2}},$$

where $\sqrt[3]{\cdot}$ denotes the real cube root. Here $\lambda_1 \approx 1.3247$, which is strictly below the average out/in degree of $4/3$. Hence, for such networks Theorem 5.1 shows that neither the out-degree generalized friendship paradox for $\mathbf{x} = \mathbf{x}_L$ nor the in-degree generalized friendship paradox for $\mathbf{x} = \mathbf{x}_R$ applies.

Matrix function-based centrality measures of the form (8) continue to make sense for directed networks. Here, the entry $(A^k)_{ij}$ counts the number of distinct directed walks of length k from i to j. We will focus on the Katz case, where

$$\mathbf{x} = \left(I + \alpha A + \alpha^2 A + \cdots\right)\mathbf{1}. \tag{17}$$

Let us first check how this measure correlates with in-degree. The relevant difference $\mathbf{d}^{\text{in}\top}\mathbf{x} - \|\mathbf{d}^{\text{in}}\|_1 \|\mathbf{x}\|_1/n$ from (15) then takes the form

$$\mathbf{1}^\top A \left(I + \alpha A + \alpha^2 A^2 + \cdots\right)\mathbf{1} - \mathbf{1}^\top\left(I + \alpha A + \alpha^2 A^2 + \cdots\right)\mathbf{1}\frac{\mathbf{1}^\top A\mathbf{1}}{n}.$$

In terms of powers of α, the zeroth-order term vanishes and the first-order term is

$$\alpha\left(\mathbf{1}^\top A^2\mathbf{1} - \frac{(\mathbf{1}^\top A\mathbf{1})^2}{n}\right). \tag{18}$$

In words, we are comparing the total number of directed walks of length two with the square of the total number of directed walks of length one, scaled by the number of nodes. This difference can be negative—for example, the graph that was used to give (13) produces $\mathbf{1}^\top A^2\mathbf{1} - (\mathbf{1}^\top A\mathbf{1})^2/n = -n + 3 - 1/n$. Hence, the corresponding generalized friendship paradox fails on this example for small α when $n \geq 3$.

We show next that is possible to prove something positive for the alternative out-degree case.

THEOREM 5.2 Consider the out-degree generalized friendship paradox inequality (14) in the Katz case (17). For any network there exists a value $\alpha^\star > 0$ such that the inequality holds for all parameter values $0 < \alpha < \alpha^\star$, with equality if and only if the network has a common out degree. Further, a sufficient condition for the inequality to hold for all $0 < \alpha < 1/\rho(A)$ is that for all $k \geq 1$

$$\mathbf{1}^\top A^\top A^k\mathbf{1} \geq \mathbf{1}^\top A^k\mathbf{1}\frac{\mathbf{1}^\top A\mathbf{1}}{n}. \tag{19}$$

Proof. The relevant difference is $\mathbf{d}^{\mathrm{out}\top}\mathbf{x} - \|\mathbf{d}^{\mathrm{out}}\|_1 \|\mathbf{x}\|_1/n$, which takes the form

$$\mathbf{1}^\top A^\top \left(I + \alpha A + \alpha^2 A^2 + \cdots\right)\mathbf{1} - \mathbf{1}^\top\left(I + \alpha A + \alpha^2 A^2 + \cdots\right)\mathbf{1}\frac{\mathbf{1}^\top A\mathbf{1}}{n}. \tag{20}$$

The sufficient condition (19) follows by considering powers of α.

Now, suppose the network has a common out degree, so $A\mathbf{1} = \deg\mathbf{1}$ for some value deg. Then $\mathbf{1}$ must be the Perron–Frobenius right eigenvector, so $A\mathbf{1} = \lambda_1\mathbf{1}$, where λ_1 is the Perron–Frobenius eigenvalue. In this case, $\mathbf{1}^\top A^\top A^k\mathbf{1} = n\lambda_1^{k+1} = \mathbf{1}^\top A^k\mathbf{1}\mathbf{1}^\top A\mathbf{1}/n$. So we have the stated equality.

Now, suppose the network does not have a common out degree, so $A\mathbf{1}$ is not a multiple of $\mathbf{1}$. For small α, the leading term in (20) may be written

$$\alpha\left(\mathbf{d}^{\mathrm{out}\top}\mathbf{d}^{\mathrm{out}} - \frac{\|\mathbf{d}^{\mathrm{out}}\|_1^2}{n}\right),$$

which is positive by Cauchy–Schwarz. Hence a suitable $\alpha^\star$ exists. $\qquad\square$

The Katz centrality measure (17) assigns to node i a weighted sum of all directed walks starting from node i. In a message-passing context, this measure rewards nodes that are able to **broadcast** information effectively. In the limit $\alpha \to 0$ this measure approaches (a shifted version of) the out-degree. As an alternative, we could replace A by $A^\top$ in (17). This measure assigns to node i a weighted sum of all directed walks finishing at node i, thereby rewarding nodes that are able to **receive** effectively. In the limit $\alpha \to 0$ this measure approaches (a shifted version of) the in-degree. Analogous versions of the conclusions that follow (18) and the statement of Theorem 5.2 are then valid with A replaced by $A^\top$ and 'in' and 'out' swapped.

We also note that part of Theorem 5.2 extends to the general power series centrality measure $\mathbf{x} = \left(c_0 I + c_1 A + c_2 A^2 + \cdots\right)\mathbf{1}$—requiring (19) to hold for every $k \geq 1$ for which $c_k > 0$ serves as a sufficient condition.

6. Weighted networks

The results in the previous sections, including [13, Theorem 1], do not require the network to be unweighted. So the conclusions extend to non-negatively weighted networks if we are willing to use the formulations (1) and (12) for the friendship paradox and (5), (14) and (15) for the generalized friendship paradox, with the degree vectors having the same definitions: $\mathbf{d} = A\mathbf{1}$, $\mathbf{d}^{\mathrm{out}} = A\mathbf{1}$ $\mathbf{d}^{\mathrm{in}} = A^\top\mathbf{1}$. In this extended setting, the degree vectors represent sums of weights rather than edge counts, and we note that the inequalities are invariant to positive rescaling of degree, so we may assume without loss of generality that $\mathbf{d}$, $\mathbf{d}^{\mathrm{out}}$ or $\mathbf{d}^{\mathrm{in}}$ have elements that sum to unity. Similarly, we may assume that $\mathbf{x}$ also sums to unity. The friendship and generalized friendship paradoxes then apply if 'average' is interpreted as 'weighted average'.

7. Discussion

The original friendship paradox has spawned a range of activity in quantitative network science, and it has been argued that the effect may explain reports of increasing levels of dissatisfaction in online social interaction [7] and may be systematically distorting our perceptions and behaviours [25]. It has also been shown that the paradox may be leveraged in order to detect the spread of information or disease, and to drive effective interventions [1, 6, 26–29]. Our main result, Theorem 3.1, shows that the paradox

holds with the same level of generality when we replace *number of friends* by *eigenvector centrality*. Hence, the effect is always present for an attribute that is widely adopted in social network analysis as a more sophisticated and informative measure of influence than degree centrality. Our work therefore adds support to the hypotheses in the social science literature relating the paradox to our observed attitudes and actions. Moreover, it has been found that centrality measures correlate strongly with infectiousness in studies of the spread of disease [30–32]; in particular, [33] reported that eigenvector centrality was the best among several options for characterizing superspreaders. So Theorem 3.1 provides further justification for techniques that exploit the friendship paradox for network monitoring and intervention.

It is of interest to note that the original friendship paradox is based on a purely *local* quantity—the number of immediate neighbours. Theorem 3.1 shows that the effect is also present for a *global* quantity that takes account of long range interactions. Indeed the walk-based Katz centrality measure, (8) with $c_k = \alpha^k$, interpolates between these two extremes: $\alpha \to 0$ from above reduces to degree and $\alpha \to 1/\rho(A)$ from below becomes eigenvector centrality [17, 19]. Theorem 4.4 shows that Katz maintains the paradox for sufficiently small α, but it is an open question as to whether there is an undirected network for which the paradox fails to hold for some $0 < \alpha < 1/\rho(A)$.

We note that [13] gives a concrete example of a connected network on which the total number of walks of length three is strictly less than the product of the total number of walks of length two and the average degree; so the inequality (11) is violated for $r = 2$ and $s = 1$. It follows that by taking $c_0 = c_2 = 1$ and the remaining c_k sufficiently small, we can construct a centrality measure (8) based on a power series with positive coefficients for which the generalized friendship paradox fails to hold. This raises the question of categorizing those power series that never give rise to such counterexamples. Theorem 4.3 shows that odd power series are one such class. We remark that numerical experiments (not described here) on the $r = 2, s = 1$ example from [13] indicated that it does not provide a counterexample for a Katz-friendship paradox—with $c_k = \alpha^k$, the α^2 term in the left-hand side of (9) makes a negative contribution, but the sum remains non-negative for all $0 < \alpha < 1/\rho(A)$.

In [13] it is also stated that for any given network the inequality (11), and hence the sufficiency condition (10), holds for large enough $r + s$. This is entirely consistent with the eigenvector result in Theorem 3.1—increasing α in Katz centrality emphasizes longer walks, and the $\alpha \to 1/\rho(A)^-$ limit corresponds to the eigenvector case [17, 19].

To the best of our knowledge, extensions of the friendship paradox to directed networks had only been studied empirically, as in [8]. In Section 5, we clarified that two of the four out-/in-degree versions always hold, while the other two may fail. For example, when we allow for a lack of reciprocation it remains the case that *the people we admire have more admirers than us, on average* (something many of us first discovered at high school), and, for the same reason, *people we hate are hated by more people than us, on average*. However, it is **not true** in general that *we admire/hate people who admire/hate more people than us, on average*.

We gave in Theorem 5.1 a spectral condition that determines whether the relevant eigenvector centrality maintains the generalized paradox for directed networks. In this unsymmetric setting, it would be of interest to find useful classes of network for which the spectral condition is satisfied, and also to identify power series centrality measures for which a generalized friendship paradox is always guaranteed, thereby extending the sufficiency result in Theorem 5.2.

REFERENCES

1. FELD, S. L. (1991) Why your friends have more friends than you do. *Am. J. Sociol.*, **96**, 1464–1477.
2. STROGATZ, S. (2012) *Friends you can count on*. New York: The New York Times.
3. GRUND, T. U. (2014) Why your friends are more important and special than you think. *Sociol. Sci.*, **1**, 128–140.
4. ZUCKERMAN, E. W. & JOST, J. T. (2001) What makes you think you're so popular? Self-evaluation maintenance and the subjective side of the "friendship paradox". *Soc. Psychol. Q.*, **64**, 207–223.
5. EOM, Y.-H. & JO, H.-H. Generalized friendship paradox in complex networks: the case of scientific collaboration. *Sci. Rep.*, **4**, 4603.
6. PIRES, M. M., MARQUITTI, F. M. & GUIMARÃES, P. R. (2017) The friendship paradox in species-rich ecological networks: implications for conservation and monitoring. *Biol. Conserv.*, **209**, 245–252.
7. BOLLEN, J., GONÇALVES, B., VAN DE LEEMPUT, I. & RUAN, G. (2017) The happiness paradox: your friends are happier than you. *EPJ Data Sci.*, **6**, 4.
8. 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 (ICWSM), 2013* (E. Kiciman, N. B. Ellison, B. Hogan, P. Resnick & I. Soboroff eds). *Honorable mention paper*. Cambridge, Massachusetts: AAAI Press.
9. KOOTI, F., HODAS, N. O. & LERMAN, K. (2014) Network weirdness: exploring the origins of network paradoxes. *International Conference on Weblogs and Social Media (ICWSM)* (E. Adar & P. Resnick eds), March 2014. Ann Arbor Michigan: AAAI Press.
10. ESTRADA, E. (2011) *The Structure of Complex Networks*. Oxford: Oxford University Press.
11. FREEMAN, L. C. (1979) Centrality networks: I. Conceptual clarifications. *Soc. Netw.*, **1**, 215–239.
12. NEWMAN, M. E. J. (2010) *Networks: An Introduction*. Oxford: Oxford University Press.
13. LAGARIAS, J. C., MAZO, J. E., SHEPP, L. A. & MCKAY, B. (1983) An inequality for walks in a graph. *SIAM Rev.*, **25**, 580–582.
14. HORN, R. A. & JOHNSON, C. R. (2013) *Matrix Analysis*, 2nd edn. Cambridge: Cambridge University Press.
15. BONACICH, P. (1972) Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.*, **2**, 113–120.
16. BONACICH, P. (1987) Power and centrality: a family of measures. *Am. J. Sociol.*, **92**, 1170–1182.
17. VIGNA, S. (2016) Spectral ranking. *Netw. Sci.*, **4**, 433–445.
18. BENZI, M. & KLYMKO, C. (2013) Total communicability as a centrality measure. *J. Complex Netw.*, **1**, 124–149.
19. BENZI, M. & KLYMKO, C. (2015) On the limiting behavior of parameter-dependent network centrality measures. *SIAM J. Matrix Anal. Appl.*, **36**, 686–706.
20. ESTRADA, E. (2010) *Generalized walks-based centrality measures for complex biological networks*. J. Theor. Biol., **263**, 556–565.
21. ESTRADA, E., HATANO, N. & BENZI, M. (2012) The physics of communicability in complex networks. *Phys. Rep.*, **514**, 89–119.
22. ESTRADA, E. & HIGHAM, D. J. (2010) Network properties revealed through matrix functions. *SIAM Rev.*, **52**, 696–671.
23. RODRÍGUEZ, J. A., ESTRADA, E. & GUTIÉRREZ, A. (2007) Functional centrality in graphs. *Linear Multilinear Algebra*, **55**, 293–302.
24. KATZ, L. (1953) A new index derived from sociometric data analysis. *Psychometrika*, **18**, 39–43.
25. JACKSON, M. O. *The friendship paradox and systematic biases in perceptions and social norms*, J. Political Economy, https://web.stanford.edu/~jacksonm/papersarticles.html.
26. CHRISTAKIS, N. A. & FOWLER, J. H. (2010) Social network sensors for early detection of contagious outbreaks. *PLoS One*, **5**, 0012948.
27. GARCIA-HERRANZ, M., MORO, E., CEBRIAN, M., CHRISTAKIS, N. A. & FOWLER, J. H. (2014) Using friends as sensors to detect global-scale contagious outbreaks. *PLoS One*, **9**, 0092413.
28. KUMAR, V., KRACKHARDT, D. & FELD, S. Network interventions based on inversity: leveraging the friendship paradox in unknown network structures, Working Paper, http://faculty.som.yale.edu/vineetkumar/.

29. KUMAR, V. & SUDHIR, K. Can friends seed more buzz and adoption? Working Paper, http://faculty.som.yale.edu/vineetkumar/.

30. DE, P., SINGH, A. E., WONG, T., YACOUB, W. & JOLLY, A. M. (2004) Sexual network analysis of a gonorrhoea outbreak. *Sex. Transm. Infect.*, **80**, 280–285.

31. GÓMEZ, J. M., NUNN, C. L. & VERDÚ, M. (2013) Centrality in primate–parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. *Proc. Natl. Acad. Sci. USA*, **110**, 7738–7741.

32. MANTZARIS, A. V. & HIGHAM, D. J. (2013) Dynamic communicability predicts infectiousness. *Temporal Networks* (P. Holme & J. Saramäki eds). Berlin: Springer, pp. 283–294.

33. SHAKARIAN, P., BHATNAGAR, A., ALEALI, A., SHAABANI, E. & GUO, R. (2015) *The SIR Model and Identification of Spreaders.* (S. Zdonik, S. Shekhar, X. Wu, L. C. Jain, D. Padua, X. S. Shen, B. Furht, V. S. Subrahmanian, M. Hebert, K. Ikeuchi, B. Siciliano, S. Jajodia & N. Lee eds) Springer Briefs in Computer Science. Switzerland: Springer, pp. 3–18, ISBN: 9783319231044.