The friendship paradox for weighted and directed networks

Kenneth S. Berenhaut\textsuperscript{a,1}, Hongyi Jiang\textsuperscript{a}

\textsuperscript{a}Department of Mathematics and Statistics, Wake Forest University, 1834 Wake Forest Road, Winston Salem, NC, USA

Abstract

This paper studies the friendship paradox for weighted and directed networks, from a probabilistic perspective. We consolidate and extend recent results of Cao and Ross and Kramer, Cutler and Radcliffe, to weighted networks. Friendship paradox results for directed networks are given; connections to detailed balance are considered.

Keywords: Friendship paradox, random walks, directed networks, weighted networks, configuration model, detailed balance

2000 MSC: 05C81, 05C07, 60J10, 60J15, 05C20, 91D30, 05C80

1. Introduction

The friendship paradox, introduced by Feld\cite{feld1981}, states roughly that, in a network scenario, one’s neighbours have (on average) more neighbours than oneself. The result has recently been employed to advantage in epidemic detection and more general sampling scenarios (see Cohen et al.\cite{cohen2001}, Christakis and Fowler\cite{christakis2007}, Garcia-Herranz et al.\cite{garcia-herranz2010}, Eom and Jo\cite{eom2010}, Kim et al.\cite{kim2011}, Herrera et al.\cite{herrera2012}, Singer\cite{singer2013}), and has received considerable attention from scientists across disciplines. For a recent discussion of societal welfare implications see Jackson\cite{jackson2014}.

Throughout, we will assume that $G = (V, \omega)$ is a directed weighted graph, or network, with a set of vertices or nodes, $V = \{v_1, v_2, \ldots, v_n\}$, and a weight function $\omega$ from $V \times V$ to the non-negative reals, $\mathbb{R}^+$. Such graphs arise in many physical, ecological, social, and economic studies where the weights represent varying tie strength, intensity or capacity (see for instance Newman\cite{newman2003}, Serrano et al.\cite{serrano2007}).

Email address: berenhks@wfu.edu (Kenneth S. Berenhaut)
We also consider a general node attribute function, \( f : V \to [0, \infty) \). For a node \( v \in V \), \( f(v) \) could denote, for instance, the in-degree or out-degree of \( v \) in \( G \), some other graph dependent measure such as betweenness or closeness (see for instance Opsahl et al. [18] and the references therein), or some attribute extraneous to \( G \) itself, such as biomass in a food web.

For \( v \in V \), define the in-degree and out-degree functions \( d_i \) and \( d_o \) from \( V \) to \( \mathbb{R}^+ \), via

\[
d_i(v) \triangleq \sum_{u \in V} \omega(u, v) \quad \text{and} \quad d_o(v) \triangleq \sum_{u \in V} \omega(v, u),
\]

respectively. If \( \omega \) is symmetric, i.e. \( \omega(v, w) = \omega(w, v) \) for all \( (v, w) \in V \times V \), we will on occasion refer to the degree function \( d = d_i = d_o \).

If \( \omega(v, w) \in \{0, 1\} \) for all \( v, w \in V \), \( G \) is an unweighted (possibly directed) graph, and we can write \( G = (V, E) \), where \( E = \{(v, w) \in V \times V : \omega(v, w) = 1\} \).

For the purposes of neighbour selection, as in Kramer et al. [19], it will be convenient to consider a time-homogeneous random walk \( X = (X_0, X_1, \ldots) \) on the graph \( G \) dictated by a transition matrix, \( P = [P_{i,j}] \), with

\[
P_{i,j} \triangleq \mathbb{P}(X_{l+1} = v_j | X_l = v_i) = \frac{\omega(v_i, v_j)}{d_o(v_i)},
\]

for \( 1 \leq i, j \leq n \) and \( l \geq 0 \). We will assume throughout that \( X_0 \) is uniformly selected from \( V \).

Recently, Cao and Ross [20] proved the following version of the friendship paradox (see also Jackson [9]).

**Theorem 1.** Suppose \( X = (X_0, X_1, \ldots) \) is a random walk on an unweighted, undirected graph \( G \). Then, the degree of \( X_1 \) is stochastically larger than the degree of \( X_0 \), i.e. for all \( t \in \mathbb{R} \)

\[
\mathbb{P}(d(X_1) \geq t) \geq \mathbb{P}(d(X_0) \geq t).
\]

In addition, Kramer et al. [19] proved the following result regarding random walks on unweighted, undirected graphs.

**Theorem 2.** Suppose \( X = (X_0, X_1, \ldots) \) is a random walk on an unweighted, undirected graph \( G \). Then, for \( k \geq 0 \), the degree of \( X_k \) is no less than the degree of \( X_0 \) in expected value, i.e.

\[
\mathbb{E}(d(X_k)) \geq \mathbb{E}(d(X_0)).
\]
For further recent work related to the friendship paradox, see for instance Eom and Jo [21], Jo and Eom [22], Lerman et al. [23], Momeni and Rabbat [24], Fotouhi et al. [25], Momeni and Rabbat [26], Kooti et al. [27], Hodas et al. [28], Momeni and Rabbat [29], Wu et al. [30], and for further information on random walks on graphs, see for instance Lovász [31], Aldous and Fill [32], Zhou and Lipowsky [33], Pons and Latapy [34], Solé-Ribalta et al. [35].

In what follows we will extend Theorems 1 and 2 in the context of weighted and directed graphs. The following simple lemma will be crucial, throughout.

Lemma 1. Suppose $X = (X_0, X_1, \ldots)$ is a time-homogeneous Markov chain on the state space $V$ with transition matrix $P = [P_{i,j}]$ and $X_0$ distributed uniformly on $V$. If $t \in \mathbb{R}$, and $f : V \to [0, \infty)$, then

$$
\Delta_t \overset{def}{=} \mathbb{P}(f(X_1) \geq t) - \mathbb{P}(f(X_0) \geq t) = \frac{1}{n} \sum_{1 \leq i,j \leq n} \left( P_{i,j} - P_{j,i} \right).
$$

(5)

Proof. For $t > 0$, define the sets

$$
S_i^+ = \{ v : f(v) \geq t \} \text{ and } S_i^- = \{ v : f(v) < t \},
$$

(6)

and note that $|S_i^+| + |S_i^-| = n$. We then have

$$
\mathbb{P}(f(X_1) \geq t) = \mathbb{P}(X_1 \in S_i^+, X_0 \in S_i^+) + \mathbb{P}(X_1 \in S_i^+, X_0 \in S_i^-) = \mathbb{P}(X_0 \in S_i^+, X_1 \in S_i^-) + \mathbb{P}(X_0 \in S_i^+, X_1 \in S_i^+).
$$

(7)

(8)

Taking a difference in (7) and (8) then gives

$$
\Delta_t = \mathbb{P}(X_1 \in S_i^+, X_0 \in S_i^-) - \mathbb{P}(X_0 \in S_i^+, X_1 \in S_i^-) = \frac{1}{n} \left( \sum_{f(v_i) < t \leq f(v_j)} P_{i,j} - \sum_{f(v_j) < t \leq f(v_i)} P_{j,i} \right).
$$

(9)

(10)

Despite its simplicity, Lemma 1 leads directly to several results.

Theorem 3. Suppose $X = (X_0, X_1, \ldots)$ is a time-homogeneous Markov chain on the state space $V$ with transition matrix $P = [P_{i,j}]$, and consider a function $f : V \to [0, \infty)$. If

$$
f(v_i) > f(v_j) \text{ implies } P_{i,j} \leq P_{j,i},
$$

(11)

then $\mathbb{P}(f(X_1) \geq t) \geq \mathbb{P}(f(X_0) \geq t)$, for all $t \in \mathbb{R}$. 

3
In essence, the requirement in (11) states that the flow of probability into nodes with higher \( f \)-value from nodes with lower \( f \)-value is greater than the flow in the opposing direction. One key instance when (11) holds is when \( f(v) \) is the degree of node \( v \) in an undirected graph, \( G \), and \( P \) is the transition matrix for a random walk on \( G \).

**Theorem 4.** Suppose \( X = (X_0, X_1, \ldots) \) is a time-homogeneous Markov chain on the state space \( V \) with transition matrix \( P = [P_{i,j}] \), and consider a function \( f : V \rightarrow [0, \infty) \). If detailed balance holds, i.e. for all \( 1 \leq i, j \leq n \),

\[
f(v_i)P_{i,j} = f(v_j)P_{j,i},
\]

then \( P(f(X_1) \geq t) \geq P(f(X_0) \geq t) \), for all \( t \in \mathbb{R} \).

**Proof.** Suppose \( 1 \leq i, j \leq n \) with \( f(v_i) > f(v_j) \geq 0 \). If \( P_{j,i} > 0 \), then

\[
\frac{P_{i,j}}{P_{j,i}} = \frac{f(v_j)}{f(v_i)} \leq 1,
\]

and hence (11) holds. Otherwise, \( P_{j,i} = P_{i,j} = 0 \), and the result follows. \( \square \)

Now, let \( F = [F_{i,j}] \) be a diagonal matrix with \( i \)-th diagonal entry \( F_{i,i} = f(v_i) \). If \( P = F^{-1}B \) for some symmetric matrix \( B = [B_{i,j}] \), then \( FP = B \) is symmetric and hence Equation (12) holds. Suppose that \( A = [A_{i,j}] \) is an \( n \times n \) matrix with \((i,j)\)-entry \( A_{i,j} = \omega(v_i, v_j) \) and that \( A \) is symmetric. For \( v \in V \), let \( f(v) = d(v) \) be the degree of \( v \) in \( G \) and set \( P = F^{-1}A \). The \( k \)-step transition matrix, \( P^k \), satisfies \( P^k = F^{-1}B \) where \( B = F(F^{-1}A)^k \) is symmetric. We immediately have the following extension of Theorems 1 and 2.

**Theorem 5.** *(Multistep friendship paradox)* Suppose \( (X_0, X_1, \ldots) \) is a random walk on an undirected graph \( G = (V, \omega) \). Then, for \( k \geq 1 \) the degree of \( X_k \) is stochastically larger than the degree of \( X_0 \).

We will employ Lemma 1 further in the next section. We now turn to consideration of directed networks.
2. Directed networks

The friendship paradox in directed graph scenarios has been considered in Hodas et al. [28], Kooti et al. [27], Garcia-Herranz et al. [4], Momeni and Rabbat [29]. It is quite clear that results such as those in Section 1 will not hold in general (for all \( t > 0 \)) for directed networks, as the next example illustrates.

![Figure 1: Two five-node directed networks. In-degrees are indicated adjacent to the corresponding nodes. The constants along edges in (a) indicate multiple edges.](image)

**Example 1.** Consider the 5-node directed graph depicted in Figure 1(a), with an in-degree sum of ten; note that there are two multi-edges of weight two and three, respectively. Here, the mean in-degree is two, while selecting a random \( v \in V \) and a random out-edge of \( v \) leads to a node with expected in-degree \( \frac{39}{20} < 2 \). Similarly for the 5-node graph in (b), the mean in-degree is \( 2.8 \) and selecting a random \( v \in V \) and a random out-edge of \( v \) leads to a node with expected in-degree \( \frac{41}{15} < 2.8 \). \( \square \)

In this section we will show that the friendship paradox holds, on average, for given degree sequences, in a certain sense. In particular suppose \( G = (V, \omega) \) is a directed graph with \( \omega(v,w) \in \mathbb{Z}^+ = \{0,1,2,3,\ldots\} \), for all \( (v,w) \in V \times V \). Now, consider the sequences of in-degree and out-degrees of \( G \), \( d_i = (d_i(v_1),d_i(v_2),\ldots,d_i(v_n)) \) and \( d_o = (d_o(v_1),d_o(v_2),\ldots,d_o(v_n)) \). The configuration (or matching) model provides a well-studied means to produce a random directed multi-edge graph with fixed degree sequences \( d_i \) and \( d_o \) (see for instance Milo et al. [36], Chen and Olvera-Cravioto [37]). In particular, assign to each node


\(v, d_i(v)\) in-stubs and \(d_o(v)\) out-stubs, and randomly pair in-stubs and out-stubs to create a (random) graph with degree sequences \(d_i\) and \(d_o\). For further discussion of configuration models, see for instance Molloy and Reed [38, 39], Bender and Canfield [40].

Now, as in (6), for \(t > 0\), set

\[
S^+_t \overset{\text{def}}{=} \{v : d_i(v) \geq t\} \quad \text{and} \quad S^-_t \overset{\text{def}}{=} \{v : d_i(v) < t\}.
\]

In addition define the degree sums

\[
I^+_t \overset{\text{def}}{=} \sum_{v \in S^+_t} d_i(v), \quad I^-_t \overset{\text{def}}{=} \sum_{v \in S^-_t} d_i(v),
\]

\[
O^-_t \overset{\text{def}}{=} \sum_{v \in S^-_t} d_o(v), \quad O^+_t \overset{\text{def}}{=} \sum_{v \in S^+_t} d_o(v),
\]

and \(M = I^+_t + I^-_t = O^+_t + O^-_t\). We will prove the following result.

**Theorem 6.** (Friendship paradox for directed networks) Suppose \(V = \{v_1, \ldots, v_n\}\) and the degree sequences \(d_i\) and \(d_o\) are fixed. Consider the ensemble of graphs, \(G\) of all graphs with these degree sequences. If \(G \in \mathcal{G}\) (with adjacency matrix \(A = A_G = [A_{i,j}]\)) is formed randomly via the configuration model, and \(P = P_G = [P_{i,j}]\) is the resulting (random) transition matrix for a random walk on \(G\), i.e.

\[
P_{i,j} = \frac{A_{i,j}}{d_o(v_j)},
\]

then for \(t \geq 0\), with \(S^-_t\) and \(S^+_t\) non-empty,

\[
\mathbb{E}_G (\mathbb{P}(d_i(X_1) \geq t) - \mathbb{P}(d_i(X_0) \geq t)) = \frac{|S^+_t||S^-_t|}{Mn} \left( \frac{I^+_t}{|S^+_t|} - \frac{I^-_t}{|S^-_t|} \right)
\]

\[
= \frac{I^+_t}{M} - \frac{|S^+_t|}{n} > 0,
\]

where \(\mathbb{E}_G\) indicates expected value with respect to the configuration model for the given degree sequences.

**Proof.** For \(v_j \in S^-_t\), set \(Y_{j,k} = 1/d_o(v_j)\) whenever the \(k\)-th stub outgoing from \(v_j\) attaches to a node in \(S^+_t\) and zero otherwise. Similarly for \(v_j \in S^+_t\), \(Y_{j,k} = 1/d_o(v_j)\)
whenever the $k$-th stub outgoing from $v_j$ attaches to a node in $S_t^-$ and zero otherwise. Then

$$
\mathbb{E}_G(Y_{j,k}) = \begin{cases} 
\frac{1}{d_o(v_j)}(I_t^+ / M) & \text{if } v_j \in S_t^- \\
\frac{1}{d_o(v_j)}(I_t^- / M) & \text{if } v_j \in S_t^+ \\
0 & \text{otherwise.}
\end{cases}
$$

(19)

Now, the random transition matrix, $P$ satisfies

$$
\sum_{d_i(v_i) < t \leq d_i(v_j)} P_{i,j} = \sum_{v_i \in S_t^-} \sum_{v_j \in S_t^+} \sum_{1 \leq k \leq d_o(v_i)} Y_{i,k},
$$

(20)

and employing (19) gives

$$
\mathbb{E}_G\left(\sum_{d_i(v_i) < t \leq d_i(v_j)} P_{i,j}\right) = \sum_{v_i \in S_t^-} \frac{I_t^+}{M} \cdot |S_t^-|,
$$

(21)

and similarly

$$
\mathbb{E}_G\left(\sum_{d_i(v_j) < t \leq d_i(v_i)} P_{i,j}\right) = \frac{I_t^-}{M} \cdot |S_t^+|.
$$

(22)

Employing Lemma 1 and Equations (21) and (22), we have

$$
\mathbb{E}_G(\mathbb{P}(d_i(X_1) \geq t) - \mathbb{P}(d_i(X_0) \geq t)) = \frac{1}{n} \left( \frac{I_t^+}{M} \cdot |S_t^-| - \frac{I_t^-}{M} \cdot |S_t^+| \right),
$$

(23)

and the theorem follows upon simplification, noting that $I_t^+ / |S_t^+|$ and $I_t^- / |S_t^-|$ are the mean in-degrees over the sets $S_t^+$ and $S_t^-$, respectively. □

Summing over $t \geq 1$ in (18) leads to the following corollary.

**Corollary 1.** Under the assumptions in the statement of Theorem 6, we have

$$
\mathbb{E}_G(\mathbb{E}(d_i(X_1)) - \mathbb{E}(d_i(X_0))) = \frac{n^{-1} \sum_{v \in V} (d_i(v) - m)^2}{m},
$$

(24)

where $m$ is the mean in-degree, given by $m \overset{\text{def}}{=} n^{-1} \sum_{v \in V} d_i(v)$.  

7
Proof. Summing over \( t \geq 1 \), we obtain

\[
\sum_{t \geq 1} I_t^+ = \sum_{t \geq 1} \sum_{v: d_i(v) \geq t} d_i(v) = \sum_{v \in V} \sum_{1 \leq t \leq d_i(v)} d_i(v) = \sum_{v \in V} d_i(v)^2 \tag{25}
\]

and

\[
\sum_{t \geq 1} |S_t^+| = \sum_{t \geq 1} \sum_{v: d_i(v) \geq t} 1 = \sum_{v \in V} \sum_{1 \leq t \leq d_i(v)} 1 = \sum_{v \in V} d_i(v). \tag{26}
\]

Hence, employing (18) gives

\[
\mathbb{E}_G (\mathbb{E}(d_i(X_1)) - \mathbb{E}(d_i(X_0))) = \mathbb{E}_G \left( \sum_{t \geq 1} \mathbb{P}(d_i(X_1) \geq t) - \sum_{t \geq 1} \mathbb{P}(d_i(X_0) \geq t) \right) = \sum_{t \geq 1} \left( \mathbb{E}_G (\mathbb{P}(d_i(X_1) \geq t) - \mathbb{P}(d_i(X_0) \geq t)) \right) = \sum_{t \geq 1} \left( \frac{I_t^+}{M} - \frac{|S_t^+|}{n} \right). \tag{27}
\]

Equations (25) and (26), and the fact that \( M/n = m \), then imply

\[
\mathbb{E}_G (\mathbb{E}(d_i(X_1)) - \mathbb{E}(d_i(X_0))) = \sum_{t \geq 1} \frac{I_t^+}{M} - \sum_{t \geq 1} \frac{|S_t^+|}{n} = \frac{1}{m} \left( \frac{1}{n} \sum_{v \in V} d_i(v)^2 - m^2 \right). \tag{28}
\]

The result follows. \( \square \)

Arguing similarly we also have the corresponding results to Theorem 6 and Corollary 1, when the in-degree and out-degree functions, \( d_i \) and \( d_o \), are swapped throughout.

Remark. From a sociological perspective, it may be appropriate to restrict the configuration model to attachments wherein out-stubs can only connect to in-stubs of other nodes (disallowing self-loops). Consideration of the friendship paradox under this constraint is work in progress; subtleties arise due to the non-uniform nature of stub attachments. For discussion of simple directed graphs in the context of the configuration model see for instance Chen and Olvera-Cravioto [37]. \( \square \)
Now, consider the ensemble of undirected multigraphs, (where $\omega$ is symmetric), and define

$$D_t^+ \overset{def}{=} \sum_{v \in S_t^+} d(v), \quad D_t^- \overset{def}{=} \sum_{v \in S_t^-} d(v),$$

(29)

and $M \overset{def}{=} D_t^+ + D_t^-$. Similar to above, the configuration model also provides a means to produce a random undirected multi-edge graph with degree sequence $d = (d(v_1), d(v_2), \ldots, d(v_n))$; see for instance Milo et al. [36], Chen and Olvera-Cravioto [37]. In particular, assign to each node $v$, $d(v)$ stubs, and randomly pair these stubs to create a (random) graph with degree sequence $d$.

The following result is proved similar to above; the $M - 1$ in equations (31) and (32) arises due to the fact that while loops are possible, in considering an edge incident to a single node, that edge cannot connect to itself.

**Theorem 7.** Suppose $V = \{v_1, \ldots, v_n\}$ and the degree sequence $d$ is fixed. Consider the ensemble of graphs, $G$, of all graphs with this degree sequence. If $G \in G$ (with adjacency matrix $A$) is formed randomly via the configuration model (for undirected graphs) and $P = [P_{i,j}]$ is the resulting (random) transition matrix for a random walk on $G$, i.e.

$$P_{i,j} = \frac{A_{i,j}}{d(v_i)},$$

(30)

then for $t \geq 0$, with $S_t^-$ and $S_t^+$ non-empty,

$$E_G(\mathbb{P}(d(X_1) \geq t) - \mathbb{P}(d(X_0) \geq t)) = \frac{|S_t^+||S_t^-|}{(M - 1)n} \left( \frac{D_t^+}{|S_t^+|} - \frac{D_t^-}{|S_t^-|} \right)$$

(31)

$$= \frac{D_t^+}{(M - 1)} - \frac{|S_t^+|}{n} \frac{M}{M - 1} > 0. \quad (32)$$

We now turn briefly to discussion of in-degree under stationary distributions.

**Example 1 (revisited).** Note that for a random walk on the directed graph in Figure 1(b), the stationary distribution, $\pi$, for the associated Markov chain is given by $\pi(1) = 1/4$ and $\pi(2) = \pi(3) = \pi(4) = \pi(5) = 3/16$. 

9
If \( Y \) is selected according to \( \pi \), we have
\[
P(d_i(Y) = 2) = 1/4, \quad P(d_i(X_0) = 2) = 1/5,
\]
and hence
\[
\frac{P(d_i(Y) = 2)}{P(d_i(X_0) = 2)} = \frac{5/4}{15/16} = \frac{1}{4} = \frac{P(d_i(Y) = 3)}{P(d_i(X_0) = 3)}.
\]
(33)
The inequality in (33) shows that, in general for directed graphs, \( d_i(Y) \) is not necessarily likelihood ratio larger than \( d_i(X_0) \).
□

For an unweighted, undirected graph, let \( E = \{(i, j) : \omega(v_i, v_j) = 1 \text{ and } i < j \} \) be the set of connected node-pairs. Then
\[
\frac{d(v)}{\sum_{u \in V} d(u)} = \frac{d(v)}{|V|} = \frac{d(v)}{|E|} \frac{1}{2},
\]
and hence a node may be selected according to a stationary distribution by first (uniformly) selecting an edge-pair and then selecting a node, \( Y \), from that pair.

Employing Proposition 1 in [20], \( d(Y) \) is likelihood ratio larger than \( d(X_0) \). Akin to this result, we have the following for weighted directed graphs.

**Proposition 1.** Suppose that the random pair \( \alpha \in V \times V \) is selected via
\[
P(\alpha = (u, v)) = \frac{\omega(u, v)}{\sum_{u, v \in V} \omega(u, v)},
\]
and for \( \alpha = (u, v) \), set \( Y_2 = v \) (i.e. the terminal node of the edge \( \alpha \)). Then, \( d_i(Y_2) \) is likelihood ratio larger than \( d_i(X_0) \).

**Proof.** Suppose \( x \in \{d_i(v) : v \in V \} \), and let \( \mathcal{A} = \{v : d_i(v) = x \} \). Then, \( P(d_i(X_0) = x) = |\mathcal{A}|/n \), while
\[
P(d_i(Y_2) = x) = \frac{x}{\sum_{u, v \in V} \omega(u, v)} |\mathcal{A}|.
\]
(36)
Hence
\[
\frac{P(d_i(Y_2) = x)}{P(d_i(X_0) = x)} = \frac{x}{\sum_{u, v \in V} \omega(u, v)} n
\]
(37)
is increasing in \( x \), and the result follows. □
Note that in the particular case of equal in- and out-degrees for each node (i.e. $d_i(v) = d_o(v)$ for all $v \in V$), with notation as in the statement of Proposition 1,

$$
P(Y_2 = v) = \frac{d_i(v)}{\sum_{u \in V} d_i(u)} = \frac{d_o(v)}{\sum_{u \in V} d_i(u)},
$$

i.e. $Y_2$ has distribution $\pi$, where $\pi$ is a stationary distribution for a random walk on the graph. Proposition 1 then implies that if a node $Y$ is selected according to $\pi$, $d_i(Y)$ is likelihood ratio larger than $d_i(X_0)$.


[29] N. Momeni, M. Rabbat, Qualities and inequalities in online social networks through the lens of the generalized friendship paradox, PLOS ONE 11 (2016) e0143633.


