Maximum Reciprocity in Multigraphs of Fixed Degree

Daniel Cirkovic, Sam Gailliot and Venkata Sameer Kumar Betana Bhotla

Abstract

Reciprocity, or the propensity for two nodes in a directed graph to form a pair of reciprocal edges, is an important and understudied behavior commonly exhibited in real-world networks. Unless the graph is assumed to be randomly generated, however, it is difficult to quantify the magnitude of reciprocity in a network. Jiang et al. [7] chooses to quantify reciprocity of a directed network by comparing it to the maximum achievable reciprocity in a network with the same degree sequence and they provide a greedy algorithm to estimate the maximum. In this paper, we extend their result to directed multigraphs, which are more realistic data models for social media networks. Our approach is based on a rewiring procedure which increases the reciprocity of simple triangular motifs within the graph, thus increasing the total reciprocity of the graph. We find that for graphs drawn from internet forums such as Mathoverflow our method increases reciprocity to near the upper bound.

1 Introduction

With the proliferation of social media in business, politics and personal relationships, social networks and how agents interact within them have become an increasingly important object of study. In recent years, researchers have tried painstakingly to find the underlying explanatory mechanisms that drive the interactions between people in a social network. Often, these mechanisms result in complex networks with interesting structures. For example, the Barabási-Albert [1] or Preferential Attachment model posits that new users in a network are more likely to connect to already popular nodes in a “rich get richer" fashion. Such a mechanism results in networks that exhibit power-law degree distributions [14].

Another relatively under-appreciated mechanism commonly exhibited by social networks is reciprocity, or the propensity of users to reply to messages sent to them. Research [15] has shown that reciprocity can often explain the dependence structure between the in- and out-degrees of social networks and that reciprocity may be able to be “learned" solely through the degree distribution. [3] found that the ratio between the in-degree and out-degree of a given node contributed to the probability of returning a reciprocal link. Those nodes with fewer in-degrees and large out-degrees had “lower status" and would likely reciprocate to nodes with “higher status". [15] views reciprocity as a generative mechanism that induces structure in the degree sequence, while [3] views the degree sequence as a predictor of reciprocity.

Contrary to [3] and [15], [7] views the degree sequence as a *constraint* to the total reciprocity of a network. In particular, [7] considers the problem of quantifying the magnitude of reciprocity in a social network. They lament that unless one assumed the observed network was generated according to some random mechanism á la [15], there is no discernible way to quantify the degree to which reciprocity is an influential aspect of the network structure. They thus consider all graphs with a fixed degree sequence and ask how far the observed reciprocity is from the maximum achievable reciprocity in such graphs.

Although [7] provides some very nice results for capturing the magnitude of reciprocity in simple directed graphs, for most social networks of interest, their results are quite limited. Namely, in social networks like Twitter or Facebook, users often send multiple messages to each other and themselves. These networks are thus *multigraphs*, or graphs that allow multiple edges between users and self-loops. For these situations, we prefer multigraphs over, say, collapsing edges into a weighted edges as they can potentially reveal temporal information about the connections, among other things [12]. In this report, we extend some of the results of [7] to the multigraph setting. Namely, we

- extend minor theoretical results in [7] to the multigraph setting.
- offer a new greedy algorithm for approximating the maximum reciprocity in multigraphs with a fixed degree sequence.

2 Preliminaries

Before presenting our results, we first establish notation and review some relevant materials from [7]. Let $G = (V, E)$ be a directed, simple graph. Later, we will relax this and allow G to be a multigraph. As usual, let $n = |V|$, $m = |E|$ and $\mathbf{A}$ be the adjancency matrix for G. When G is multigraph, $\mathbf{A}$ becomes a weighted adjancency matrix where $\mathbf{A}_{ij} = A_{ij}$ denotes the number of edges pointing from node i to node j. The number of self-loops pointing from node i to itself are thus given in entry A_{ii}. Regardless of if G is a multigraph or simple graph, we denote the in-degree and out-degree of a fixed node $i \in V$ as d_i^{in} and d_i^{out}. We thus define $\mathbf{d}^{\text{in}} = \begin{bmatrix} d_1^{\text{in}} & d_2^{\text{in}} & \cdots & d_n^{\text{in}} \end{bmatrix}^T$ and $\mathbf{d}^{\text{out}}$ in a similar manner. In both cases, $\mathbf{d}^{\text{in}}$ and $\mathbf{d}^{\text{out}}$ can be computed from $\mathbf{A}$ by taking its column- and row-sums, respectively.

Next, we clarify what we mean by a reciprocal edge. In the simple directed case, a *pair* of reciprocal edges is a set $\{(i, j), (j, i)\} \subset E$, while either $(i, j) \in E$ or $(j, i) \in E$ can be referred to as a reciprocal edge. We can identify a pair of reciprocal edges in $\mathbf{A}$ if $A_{ij} = A_{ji} = 1$. The same definitions hold in the multigraph case, except now $\min(A_{ij}, A_{ji})$ denotes the number of reciprocal $\{(i, j), (j, i)\}$ pairs for $i \neq j$. We explicitly exclude self-loops from being part of a reciprocal pair in the multigraph case, as not doing so would lead to uninteresting maximal reciprocity structures, as we will see later.

For both simple graphs and multigraphs with in-degrees $\mathbf{d}^{\text{in}}$ and out-degrees $\mathbf{d}^{\text{out}}$, we define the *maximal reciprocity problem* as

$$
\begin{aligned}
&\underset{\mathbf{A} \in (\mathbb{N} \cup \{0\})^{n \times n}}{\text{maximize}} \; \frac{2}{m} \sum_{i=1}^{n} \sum_{i<j\leq n} \min(A_{ij}, A_{ji}) \\
&\text{subject to} \; \sum_{i=1}^{n} A_{ij} = d_j^{\text{in}} \text{ and } \sum_{j=1}^{n} A_{ij} = d_i^{\text{out}} \;\; \forall i, j \in V
\end{aligned}
\tag{1}
$$

Here, $\rho(\mathbf{A}) := 2 \sum_{i=1}^{n} \sum_{i<j\leq n} \min(A_{ij}, A_{ji})$ counts the number of reciprocal edges and $\rho(\mathbf{A})/m$ is called the reciprocity of $\mathbf{A}$. Since $m = \sum_i d_i^{\text{in}} = \sum_i d_i^{\text{out}}$ and $\mathbf{d}^{\text{in}}$, $\mathbf{d}^{\text{out}}$ are fixed, it is equivalent to just maximize $\sum_{i=1}^{n} \sum_{i<j\leq n} \min(A_{ij}, A_{ji})$ in (1). We note that the objective is reminiscent of densest subgraph objectives.

Given the problem (1), it is relevant to ask what degree-sequences are realizable for a directed multigraph. For a directed simple graph, [7] defines *graphic sequences* $\mathbf{d} = (d_1, \ldots, d_n)$ as those for which there exists a simple graph G with degree sequence $\mathbf{d}$. They provide the theorem of Erdős and Gallai which helps identify graphic sequences. They similarly define graphic sequences for directed simple graphs, and provide the Fulkerson-Chen-Anstee theorem to identify graphic sequences $(\mathbf{d}^{\text{in}}, \mathbf{d}^{\text{out}})$. For directed multigraphs, characterizing observable degree-sequences seems like an open problem. For *undirected multigraphs with no self-loops*, we have the following theorem of Havel-Hakimi [5, 6]. We will call these degree sequences *multigraphic* sequence to differentiate from the graphic sequence for simple graphs.

Theorem 1. *A sequence of non-negative integers $d_1 \geq \ldots \geq d_n$ with $\sum_{i=1}^{n} d_i$ being even is multigraphic (with no self-loops) if and only if $d_1 \leq d_2 + \ldots + d_n$.*

3 Theory and a Greedy Algorithm

3.1 Some results motivating the algorithm

Since we would like to maximize $\rho(\mathbf{A})$ in (1), it would be sensible to find an upper bound for the objective. The below theorem is proved in [7], though our framework extends it to the multigraph case and proves it in a different manner. Though, we do admit that their proof is more intuitive.

Theorem 2. *Let $\mathbf{A}$ be an adjancency matrix for a directed simple/multigraph G and let $\mathbf{d}^{in}$ and $\mathbf{d}^{out}$ be its in- and out-degree sequences. Then*

$$\rho(\mathbf{A}) \leq \sum_{i=1}^{n} \min(d_i^{in}, d_i^{out})$$

Proof.

$$\rho(\mathbf{A}) = 2\sum_{i=1}^{n}\sum_{i<j\leq n} \min(A_{ij}, A_{ji}) \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \min(A_{ij}, A_{ji}) \leq \sum_{i=1}^{n} \min\left(\sum_{j=1}^{n} A_{ij}, \sum_{j=1}^{n} A_{ji}\right) = \sum_{i=1}^{n} \min\left(d_i^{\text{out}}, d_i^{\text{in}}\right)$$

□

[7] states that it is NP-complete to decide whether or not $\rho(\mathbf{A}) = \sum_{i=1}^{n} \min(d_i^{\text{in}}, d_i^{\text{out}})$ in the simple graph case, making (1) NP-hard. Their proof relies on mechanics we are not familiar with, so we assume that the problem is NP-hard in the multigraph case as well. This motivates the need for a greedy algorithm to approximate the objective (1). Before we present the greedy algorithm, we present a lemma which is not proved by [7], but motivates the rewiring algorithm.

Lemma 3. *Let $\mathbf{A}$ be a symmetric adjancency matrix with zero diagonal for a directed simple/multigraph G. Then $\mathbf{A}$ achieves the upper bound in Theorem 2.*

Proof. Note that symmetry of $\mathbf{A}$ implies $\mathbf{d}^{\text{in}} = \mathbf{d}^{\text{out}}$ and $\min(A_{ij}, A_{ji}) = A_{ij}$. Hence

$$\rho(\mathbf{A}) = 2\sum_{i=1}^{n}\sum_{i<j\leq n} \min(A_{ij}, A_{ji}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \min(A_{ij}, A_{ji}) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} = \sum_{i=1}^{n} d_i^{\text{out}} = \sum_{i=1}^{n} \min\left(d_i^{\text{out}}, d_i^{\text{in}}\right)$$

□

Note that symmetry is not necessary to achieve the upper bound. For example, the following adjacency matrix also achieves the upper bound since it has 6 reciprocal edges and $\sum_{i=1}^{n} \min\left(d_i^{\text{out}}, d_i^{\text{in}}\right) = 2 + 2 + 2 = 6$.

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}$$

Though, the lemma does seem to imply that in order to achieve the upper bound, it would be particularly amenable to rewire the adjacency matrix such that it is *close* to a symmetric matrix. In particular, to maximize the reciprocity of a graph with a fixed degree sequence, we would like a rewiring that maximizes the number of $(i, j) \in E$ such that $\min(A_{ij}, A_{ji}) > 0$, hence increasing $\rho(\mathbf{A}) = 2\sum_{i=1}^{n}\sum_{i<j\leq n} \min(A_{ij}, A_{ji})$. This is exactly what [7] does, though they focus on 3-paths. Since we are in the multigraph setting, we have added flexibility in how we can rewire our graph locally and still preserve the degree sequence. We note that this rewiring reminds us of notions of

graph distances, such as chemical distance [2], and wonder if rewiring $\mathbf{A}$ such that it is close to a symmetric matrix can be formalized/generalized in this manner.

Notice that based on our definition if a multigraph has self-loops then it will never achieve the upper bound. It is in our best interest to apply an rewiring scheme to remove all self-loops from the multigraph. This leads to the following lemma which gives us the necessary condition for achieving the upper bound for reciprocity in multigraphs.

Lemma 4. *A necessary condition for a directed multigraph G to achieve the upper bound in Theorem 2 is that both $\min(\mathbf{d}^{in}, \mathbf{d}^{out})$ and $\max(\mathbf{d}^{in}, \mathbf{d}^{out})$ must be multigraphic (with no self-loops), where the minimum and maximum are taken element-wise.*

Proof. Suppose that the upper bound is achieved. Let $\mathbf{A}_r$ be the symmetric matrix defined by $(A_r)_{ij} = \min(A_{ij}, A_{ji})$. Note that $\rho(\mathbf{A}_r) = \rho(\mathbf{A})$. Then $\sum_{j=1}^{n}(A_r)_{ij} = \sum_{j=1}^{n}\min(A_{ij}, A_{ji}) \leq \sum_{j=1}^{n} A_{ij} = d_i^{\text{out}}$. Similarly, $\sum_{j=1}^{n}(A_r)_{ij} \leq d_i^{\text{in}}$ and thus $\sum_{j=1}^{n}(A_r)_{ij} \leq \min(d_i^{\text{in}}, d_i^{\text{out}})$. Suppose towards a contradiction that $\sum_{j=1}^{n}(A_r)_{ij} < \min(d_i^{\text{in}}, d_i^{\text{out}})$ for at least one i. Then summing over i gives $\rho(\mathbf{A}) = \rho(\mathbf{A}_r) < \sum_{i=1}^{n}\min(d_i^{\text{in}}, d_i^{\text{out}})$, which is a contradiction to the achievability of the upper bound. Hence $\sum_{j=1}^{n}(A_r)_{ij} = \min(d_i^{\text{in}}, d_i^{\text{out}})\ \forall i \in \{1, \ldots, n\}$ and $\min(\mathbf{d}^{\text{in}}, \mathbf{d}^{\text{out}})$ is multigraphic (that is, $\mathbf{A}_r$ represents a undirected multigraph with the given degree-sequence).

Let $\mathbf{A}_c$ be the symmetric matrix defined by $(A_c)_{ij} = \max(A_{ij}, A_{ji})$. Since $\mathbf{A}$ achieves the maximal reciprocity.

$$\sum_{j=1}^{n}(\mathbf{A}_c)_{ij} = \sum_{j=1}^{n}(\mathbf{A}_c - \mathbf{A})_{ij} + \sum_{j=1}^{n}\mathbf{A}_{ij} = \sum_{j=1}^{n}(\mathbf{A}_c - \mathbf{A})_{ij} + \min(d_i^{\text{in}}, d_i^{\text{out}})$$

If $d_i^{\text{in}} < d_i^{\text{out}}$, then $\sum_{j=1}^{n}(\mathbf{A}_c - \mathbf{A})_{ij} = d_i^{\text{out}} - d_i^{\text{in}}$ since in order to create $\mathbf{A}_c$, we need to add in-edges so that the in-degree is increased from d_i^{in} to d_i^{out}. Hence $\sum_{j=1}^{n}(\mathbf{A}_c)_{ij} = d_i^{\text{out}} - d_i^{\text{in}} + d_i^{\text{in}} = d_i^{\text{out}}$. Similarly, if $d_i^{\text{out}} < d_i^{\text{in}}$, $\sum_{j=1}^{n}(\mathbf{A}_c)_{ij} = d_i^{\text{in}} - d_i^{\text{out}} + d_i^{\text{out}} = d_i^{\text{in}}$. Hence $\sum_{j=1}^{n}(\mathbf{A}_c)_{ij} = \max(d_i^{\text{in}}, d_i^{\text{out}})$ and $\mathbf{A}_c$ is multigraphic. $\square$

3.2 The greedy algorithm

We now present to a greedy algorithm used to approximate the maximal reciprocity. In particular, in our rewiring scheme, we focus on *unidirectional triangles* and *self-loops*. The only input for the algorithm is an adjacency matrix $\mathbf{A}$. The first step of the algorithm is to find unidirectional triangles. By a unidirectional directional triangle, we mean edges $(i, j), (j, k), (k, i)$ which are not already reciprocated. Let $A^{\star}_{ij}$ be the number of (i, j) edges which are not reciprocated in the triangle (i, j, k). Let $\eta = \min(A^{\star}_{ij}, A^{\star}_{jk}, A^{\star}_{ki})$ be the minimum number of unreciprocated edges pointing in one of the $(i, j), (j, k)$ or (k, i) directions. If $\eta > 1$, we perform the following rewirings

$$\begin{aligned}
A_{ji} &\leftarrow A_{ji} + (\eta - \eta \mod 2)/2 & \qquad A_{ij} &\leftarrow A_{ij} - (\eta - \eta \mod 2)/2 \\
A_{kj} &\leftarrow A_{kj} + (\eta - \eta \mod 2)/2 & \qquad A_{jk} &\leftarrow A_{jk} - (\eta - \eta \mod 2)/2 \\
A_{ik} &\leftarrow A_{ik} + (\eta - \eta \mod 2)/2 & \qquad A_{ki} &\leftarrow A_{ki} - (\eta - \eta \mod 2)/2
\end{aligned}$$

and in addition, if $\eta \mod 2 = 1$

$$\begin{aligned}
A_{ij} &\leftarrow A_{ij} - 1 & \qquad A_{ii} &\leftarrow A_{ii} + 1 \\
A_{ki} &\leftarrow A_{ki} - 1 & \qquad A_{kj} &\leftarrow A_{kj} + 1
\end{aligned}$$

The first rewiring implies that if the minimum number of edges in one side of the triangle is at least two, you should reverse the direction of approximately half of the edges. In practice this looks like rotating half of the edges counter-clockwise in the triangle. An example of such a rewiring is shown in Figure 1. The second rewiring indicates that if you are left with a unidirectional triangle such that the minimum number of edges on one side of the triangle is one, you should break two connections and rewire into a self-loop and a reciprocal edge. Such a rewiring is shown in Figure 2. This rewiring conveys the immense flexibility self-loops offer when rewiring; something not possible in [7].

Finally, after rewiring unidirectional triangles, we locally focus on self-loops. This rewiring is more obvious, if two nodes have self loops, we can delete the self-loops and create a reciprocal edge

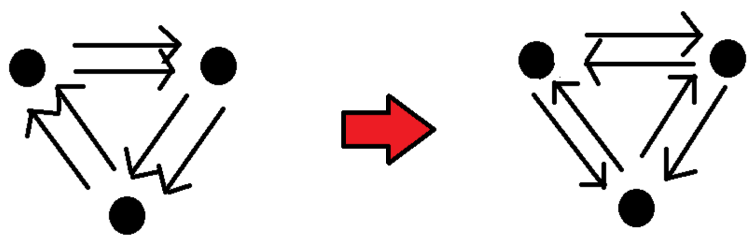

Figure 1: An example rewiring with $\eta = 2$.

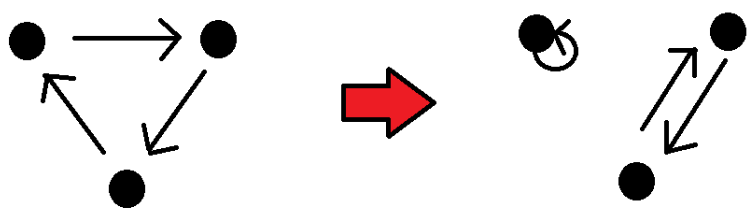

Figure 2: An example rewiring with $\eta \mod 2 = 1$.

between the nodes. We perform this step after rewiring the triangles, as rewiring the triangles can potentially create additional self-loops as seen in Figure 2. An example of this rewiring is given in Figure 3. These rewirings, although simple, must be carefully implemented in order to construct efficient code. We discuss these aspects in the following section.

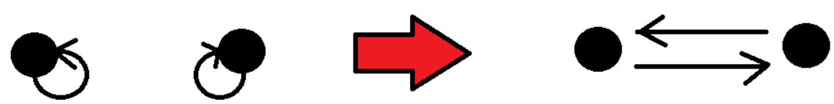

Figure 3: An example rewiring for two nodes with self-loops.

4 Experimental Results

In order to test the effectiveness of our new greedy algorithm in approximating (1), we run the algorithm on 8 datasets in total. 7 datasets are taken from the SuiteSparse Matrix Collection at https://sparse.tamu.edu/. We refer to them as College Msg [9], Epinions [11], EU Email [10], Facebook [13], Mathoverflow [10], Slashdot [8] and Twitter [4]. The final network is a randomly generated preferential attachment network, generating reciprocal edges with probability 0.2. All datasets are directed multigraphs. The output of the greedy algorithm on these datasets is displayed in Figure 4.

The rewired graph output by the greedy algorithm has variable effects on the datasets. For example, the reciprocity significantly increases for the Mathoverflow and Slashdot datasets. We believe that this is likely due to the forum-like structure of these websites where unidirectional triangles are more common. In the more social media-like datasets like Twitter and Facebook, the increase in reciprocity is minimal, indicating that perhaps the unidirectional triangle motif is not as prevalent in these networks. This opens the door for the exploration of more potential rewirings. Note that one

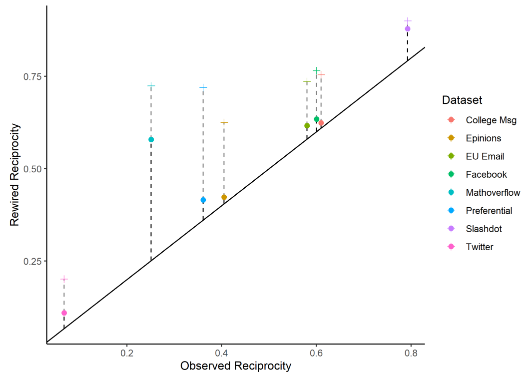

Figure 4: Emprical results of the greedy algorithm on 8 datasets. The + signs denote the upper bound given by (2) and the black line is the line $y = x$.

could add the rewirings in [7] and further increase reciprocity, though we wanted to isolate the effects of our rewiring scheme.

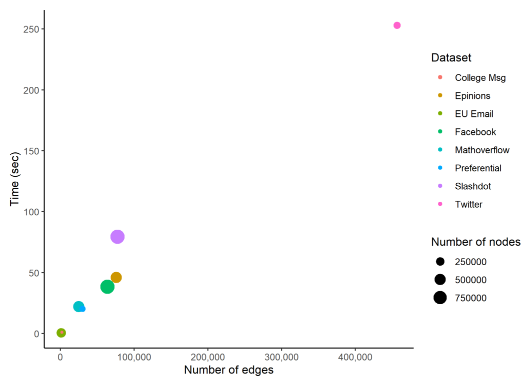

Figure 5: Empirical results of the time to execute the greedy algorithm. Larger points indicate graphs with a larger number of nodes.

Figure 4 displays the time to complete the greedy algorithm as a function of the number of edges and nodes in the network. Clearly, the number of edges is more impactful on time to completion of the algorithm, as more edges give rise to more potential self-loops and unidirectional triangles. The algorithm is relatively fast, given that finding unidirectional triangles in a network is no easy feat. We made this process faster by iterating over only the nonzero elements of $\mathbf{A}$ in order to find triangles (i, j, k). Also, in order to efficiently rewire self-loops as in Figure 3, we define $\mathbf{y}$ by $\mathbf{y}_i = A_{ii}$ and sort $\mathbf{y}$ in decreasing order. We then iterate through $\mathbf{y}$, connecting the the node with the i-th most self loops to nodes after it until we run out of self-loops.

5 Conclusion and Discussion

In this paper we have built on the work of Jiang et al. [7] by extending their methodology from simple directed graphs to directed multigraphs. We assumed that a source of lost reciprocity in real graphs is suboptimal triangular motifs. We provide an upper bound on reciprocity and a greedy algorithm which rewires suboptimal motifs to locally increase reciprocity. Unlike [7], we find that the observed reciprocity of real multigraphs is not necessarily close to the upper bound. Our results do indicate that our greedy algorithm works well for graphs drawn from forum communities, but only marginally increases the reciprocity for social media networks. This motivates the need to find other suboptimal motifs that leak reciprocity and that are common in those types of networks.

References

[1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[2] Jose Bento and Stratis Ioannidis. A family of tractable graph distances. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 333–341. SIAM, 2018.

[3] Justin Cheng, Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Predicting reciprocity in social networks. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 49–56. IEEE, 2011.

[4] Manlio De Domenico, Antonio Lima, Paul Mougel, and Mirco Musolesi. The anatomy of a scientific rumor. *Scientific reports*, 3(1):1–9, 2013.

[5] Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph. *Journal of the Society for Industrial and Applied Mathematics*, 10, 1962.

[6] Garth Isaak. Degree sequences for multitrees. *Lecture notes*, 2015.

[7] Bo Jiang, Zhi-Li Zhang, and Don Towsley. Reciprocity in social networks with capacity constraints. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 457–466, 2015.

[8] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.

[9] Pietro Panzarasa, Tore Opsahl, and Kathleen M Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932, 2009.

[10] Ashwin Paranjape, Austin R Benson, and Jure Leskovec. Motifs in temporal networks. In *Proceedings of the tenth ACM international conference on web search and data mining*, pages 601–610, 2017.

[11] Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *International semantic Web conference*, pages 351–368. Springer, 2003.

[12] Termeh Shafie. A multigraph approach to social network analysis. *Journal of Social Structure*, 16, 2015.

[13] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42, 2009.

[14] Phyllis Wan, Tiandong Wang, Richard A Davis, and Sidney I Resnick. Fitting the linear preferential attachment model. *Electronic Journal of Statistics*, 11(2):3738–3780, 2017.

[15] Tiandong Wang and Sidney I Resnick. Asymptotic dependence of in-and out-degrees in a preferential attachment model with reciprocity. *arXiv preprint arXiv:2108.03278*, 2021.