Indifferent Attachment: The Role of Degree in Ranking Friends

David Liben-Nowell, Carissa Knipe, and Calder Coalson
Department of Computer Science, Carleton College, Northfield, MN 55057
{dlibenno, knipec, coalsonc}@carleton.edu

***Abstract*—The MySpace social networking site allows each user to designate a small subset of her friends as "Top Friends," and place them in a rank-ordered list that is displayed prominently on her profile. By examining a large set of ≈11M MySpace users' choices of their #1 (best) and #2 (second-best) friends from historical crawl data from when MySpace was more popular than it now is, we discover that MySpace users were nearly indifferent to the popularity of these two friends when choosing which to designate as their best friend. Depending on the precise metric of popularity we choose, the fraction of users who select the more popular of these two individuals as their best friend wavers above and below 50%, and is always between 49.3% and 51.4%: that is, the popularity of the two candidates is essentially uninformative about which will be chosen as the best friend. Comparisons of other pairs of ranks within the Top Friends (e.g., #1-versus-#3, #2-versus-#3, . . .) also reveal no marked preference for a popular friend over a less popular one; in fact, there is some evidence that individuals tend to prefer less popular friends over more popular ones. To the extent that an individual's ranking decision in selecting between two close friends is a window into broader decisions about whom to befriend at all, these observations suggest that network-growth models based on a preference to befriend more popular individuals, like preferential attachment, may fail to capture important social behaviors. Positing individuals' tendency to attach to popular people may not suffice to explain the heavy-tailed degree distributions observed in real social networks.**

I. Introduction

Different social relationships have different priorities. Implicitly or explicitly, we all make daily decisions in which we choose one friend over another: we answer *that* email first; we find time for coffee with *this* person after telling another that we were too busy; we mention the job opportunity or free tickets to *that* friend instead of the other one. The way in which we prioritize one friend over another is an interesting, and important, question about our social relationships; it reflects the way in which a social network is used and constructed. Our prioritization decisions may also be a window into friend-making in general: the mechanisms by which Alice prioritizes her friend Bob over her friend Charlie may be very similar to the mechanisms by which Alice chooses to befriend Bob instead of befriending Charlie. (The intimacy of any social relationship falls on a continuum ranging from "best friend" via "distant friend" and "acquaintance" to "stranger," and both of Alice's decisions are comparisons of Bob to a reference point—Charlie or "friend"—on that continuum.)

As our daily lives have moved increasingly online over the last decade or so—and started to leave behind mineable data on social interactions—a voluminous body of computational research exploring the structural and behavioral properties of individuals embedded in social networks has blossomed. The preponderance of this research has treated ties in social networks as binary—we are friends; we are not friends—but a growing thread of this research has begun to consider the comparative strength of relationships.

The importance of relationship strength, and indeed the importance of *weak* relationships, has been studied in the social sciences for decades—most prominently in Mark Granovetter's "The Strength of Weak Ties" [12]. An expanding body of recent computational research has explored relationship strength, too. A nonexhaustive sampling of these papers follows. Onnela et al. [22] studied a massive dataset of weighted ties, constructed via the rate of interactions among mobile phone users. Adamic, Lauterbach, and various coauthors [1, 16, 24] have studied the role of friendship strength in determining levels of trust among members of the CouchSurfing community; for example, these authors showed that there is an inflationary effect in users' ratings of others if those ratings are made publicly and nonanonymously. Wuchty and Uzzi [25] compared self-reported "close relationships" with those inferred based on email-response times. Backstrom et al. [2] have studied how Facebook users distribute their "attention" (fraction of wall posts, photo comments, etc.) across their friends. Gilbert and Karaholios [10] and Xiang, Neville, and Rogati [26] constructed predictive models of users' perceptions of relationship strength, based on a collection of measures of profile similarity and interaction between users. Perhaps the work most similar to our own is by Kahanda and Neville [14], who attempt to classify edges in Facebook as "strong" or "weak" links (as denoted by the presence or absence of a friend in a Facebook application in which a user could list their "top" friends) using the same type of structural and transactional properties.

In this paper, we address a fine-grained question of prioritization among friends: rather than considering the *rating* of friendships on an absolute scale (from "distant" to "close"), we will consider the *ranking* of friendships on a relative scale. Specifically, we examine a feature of the MySpace online social networking site, called *Top Friends*. Each MySpace user may choose to select a subset of his or her friends to designate as Top Friends. The user puts these Top Friends into a totally ordered list, from #1 down through #k, where the cardinality k of the Top Friends list is chosen by the user. (Nearly all users choose $k \leq 40$; for some period of time predating the data acquired for the present work, MySpace fixed the cardinality $k = 8$, so disproportionately many users have 8 Top Friends. As a result, the feature is also sometimes called the "Top 8.") These Top Friends are displayed prominently on the user's profile; the remaining (non-Top) friends are accessible by clicking through to the list of “All Friends.” We are most interested in a user’s choice as to which of two individuals will be her #1-ranked friend and which will be her #2-ranked friend [8], though we also consider #i-versus-#j decisions for all other $i, j \in \{1, \ldots, 8\}$.

There is some evidence that suggests that people work hard to avoid public declarations of the ranking of their friendships—choosing weekly themes for their Top Friends or fake profiles with profile photos of Scrabble tiles that spell out, in order, a profane anti–“Top Friends” message; ranking top friends appears to be an angst-generating activity for many MySpace users [4]. This phenomenon is related to the fact that users give more generous (higher) ratings when their ratings are public and signed than when their ratings are private and anonymous [24]; MySpace users may work to avoid making their true rankings of friends known. But enough MySpace users (millions of users in our dataset) do provide rankings of real profiles that we can begin to pose, and answer, some interesting structural questions.

Indeed, the question of friendship ranking—as opposed to the question of friendship rating or of friendship existence—is a setting of scarce resources; after all, Alice can only have one #1-ranked friend. There is a minimal cost of accepting a distant acquaintance’s friend request in an online social network—perhaps just a slight risk to one’s reputation if the “friend” does something embarrassing, and the mildly mental taxation of having one more relationship to track [9, 11]. Similarly, there is little cost in the type of “grade inflation” in rating one’s friends observed in CouchSurfing [24]. But the scarcity of highly ranked friendship slots means that the ranking environment may shed a different light on potentially awkward social decisions.

The present work. This paper addresses the role of the *popularity* of Bob and Charlie when Alice chooses which of the two to prioritize over the other. Suppose that Bob is more popular than Charlie, as measured by degree in the social network. One can formulate intuitive arguments on both sides as to which of Bob or Charlie would be a better friend: Alice should tend to prefer Charlie (he’s a more “committed” friend because he has fewer distractions) or Bob (he’s a more “valuable” friend because he knows more people). Indeed, a number of “rich get richer” network-growth models designed to account for the empirically observed degree distribution of real social networks take this second view: most prominently, preferential attachment [3] posits that the probability of u being involved in a new friendship is linearly increasing in u’s current popularity.

Here we consider a large (≈11M-profile) sample of MySpace users, each of whom has selected a #1- and #2-ranked friend: a best friend and a second-best friend. Using several distinct but straightforward measures of degree, we compute the relative popularity of these two friends. What we observe, essentially, is that the popularity of these two candidates has nearly negligible predictive power in separating the #1- and #2-ranked friend. Depending on precisely which measure of popularity we use, we observe that the probability that the more popular candidate is chosen as the best friend wavers between being above 50% (as high as 51.4%) and below 50% (as low as 49.3%).

We also perform the analogous comparisons for individuals’ choice of #i-versus-#j-ranked friends for all other pairs of ranks $i, j \in \{1, \ldots, 8\}$. As in the #1-versus-#2 decision, the fraction of individuals who prefer the more popular candidate as the better-ranked (closer to #1) friend varies from slightly above 50% (as high as 51.8%) to somewhat further below 50% (as low as 46.1%).

At best, individuals exhibit a very mild preference for popularity in their choice of which of two friends to rank better; at worst, they are indifferent or even prefer an unpopular friend over one who is more popular. This lack of empirical support for individuals preferring the more popular candidate as a closer friend suggests one of two things. Either the reason for the heavy-tailed degree distribution seen in real social networks is subtler than the reinforcement-type mechanisms suggested by models based on a preference for the popular, or something about the way that we decide on the relative closeness of two close friends is fundamentally different from the way that we decide friend-versus-nonfriend.

II. The Data

We make use of a sample of the MySpace social network acquired using a cluster of desktop machines executing a parallelized BFS-style crawl over about five months in 2007–2008 [8]. We excluded profiles that were private, syntactically anomalous, had more than 20K total friends reported on their main profile page, or failed to list both an age and a sex. (Requiring both an age and a sex is a crude way of filtering out profiles of bands and other nonpersonal users from our data.) The resulting dataset contained the profiles of approximately 11 million MySpace users—10,989,190 users, to be precise. A partial BFS-style crawl is biased towards more “central” nodes; nodes in the sample may, e.g., have higher PageRank and higher degree than typical MySpace profiles [13, 21]. (Of course, there are also important differences between a typical MySpace user and a typical person in the real world, and—in this work as in all social-behavior research based on data from social networking sites, or indeed based on data from any kind of restricted population—we must be judicious in the generality of our conclusions.)

Because we focus here on the popularity of individuals, we consider several relevant measures of degree for a user u:

- The *listed degree* of u is the number of friends that are declared in u’s profile, as in “Alice has 150 friends.” We have found that this number was occasionally unreliably reported; MySpace profiles in our crawl occasionally seemed to declare fewer friends than were actually linked from the user’s profile.
- The *ranked outdegree* of u is the number of Top Friends chosen by u. (This quantity is most frequently 8, and ranges up to 40 except in rare cases.)

All of the other quantities are based on the number of people in our sample who claim u as a friend:

- The *(sample) total indegree* of u is the number of the ≈11M MySpace users who list u as a friend.
- The *(sample) rank$_k$ indegree* of u is the number of in-sample users who list u as their kth-best friend.

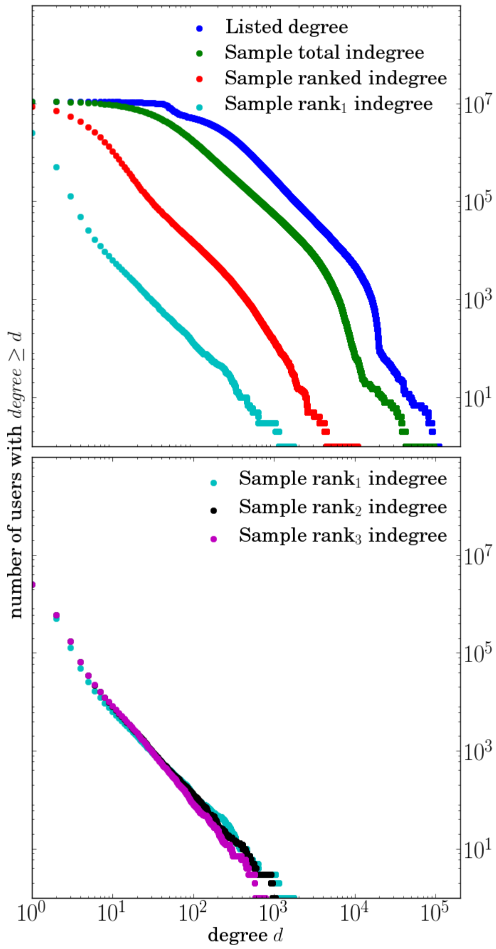

Fig. 1. Cumulative degree distributions in the ≈11M-person dataset, for various measures of degree.

- The *(sample) ranked indegree* of u is the number of users who include u anywhere in their Top Friends list, given by $\sum_k$(rank$_k$ indegree of u).

Figure 1 shows the cumulative degree distributions of the ≈11M MySpace users under six of these measures. We see that the listed degree and each of the sample indegrees (total, ranked, rank$_1$, rank$_2$, and rank$_3$) show a heavy-tailed degree distribution. (For our purposes here, we remain agnostic about the particular shape of these distributions; we make no claim about the quality of fit of a power-law model for these distributions [6].)

As one would expect, the rank$_1$, rank$_2$, and rank$_3$ indegrees are smaller in an absolute sense. For any fixed k, rank$_k$ links are a scarce resource; only one such outgoing link is possible per user, so there are only ≈11M such possible links in total. The remaining degree measures are effectively unlimited in the sense that each user can generate arbitrarily many outgoing links and nearly arbitrarily many outgoing ranked links (by lengthening her Top Friends list).

Aside from ranked outdegree, which seems qualitatively different from the other type of degree, all of these quantities seem to describe intuitively similar measures of popularity: each quantity is some version of how well-liked u is. (In contrast, a user u's ranked outdegree is a decision about how many Top Friends to rank—something about how expansive u's view of "closest friends" is.) Figure 2 shows the correlations across users among these various measures of degree. Indeed, the ranked outdegree is positively but weakly correlated with the other measures of popularity. The other totaled measures of degree—listed degree, sample ranked indegree, and sample total indegree—are all well-correlated with each other (all > 0.5). In the second part of Figure 2, we see that rank$_j$ indegree and rank$_k$ indegree are strongly positively correlated for every j, k. In fact, with one minor exception, for every $k \in \{1, \ldots, 8\}$ the rank$_k$ indegree is more strongly correlated with rank$_j$ indegree as $j < k$ gets closer and closer to k. (The lone exception is that rank$_4$ is slightly better correlated with rank$_8$ than rank$_5$ is.)

III. The Central Prediction Task

Let G_{MS} denote the MySpace social network, represented as a directed graph. Annotate each edge $u \to v$ with v's rank in u's Top Friends list (or "unranked" if v is a friend of u but not a Top Friend).

We focus on the following prediction task. Consider a focal individual u in G_{MS}. The user u names a best friend v_1 and a second-best friend v_2—the #1- and #2-ranked friends in u's Top Friends list, respectively. We erase the ranking labels from the edges from u to v_1 and v_2. The task is to predict which of $\{v_1, v_2\}$ is u's #1 friend. A predictor p may access all of the graph G_{MS}, aside from the two erased labels, in making its prediction. We say that p is *correct* on u if it identifies v_1 as u's #1 friend, *incorrect* on u if it identifies v_2 as u's #1 friend, and *nondispositive* if p cannot distinguish between v_1 and v_2. (A predictor may be nondispositive in the case of missing data or in the case that v_1 and v_2 have the same value under the predictor in question.)

Previous work on this prediction task, performed in collaboration with Peter DeScioli, Robert Kurzban, and Elizabeth Koch [8], has shown that MySpace users have a statistically significant preference for *homophily* [18] in choosing their best friends—that is, individuals tend to be more demographically similar to their #1 friend than to their #2 friend. In particular, over 56% of individuals have selected a best friend who is geographically closer than their second-best friend. (Note that this work ignored individuals on which geographic distance was nondispositive because of missing/malformed geographic data or ties in geographic distance.) A similar but substantially weaker homophilic effect holds for age: individuals tend to have best friends who are closer to their own age than their second-best friends are. In this previous work, we also identified another structural predictor that performed extremely well. Define the *relative rank* of u's friend v as the rank that v assigns to u in v's own Top Friends list. We showed that 68.8% of MySpace users selected a best friend who ranks u better than their second-best friend does. (Note again that users for which the predictor was nondispositive were ignored.) DeScioli and Kurzban [7] take an evolutionary psychological perspective on friendship, and argue for an alliance-based view of the function of friendship that anticipates the success of the relative rank predictor. Other recent work has built a prediction system for *when* a link will be reciprocated in Twitter [5] or BuzzNet [15].

		in-sample indegrees			
	ranked outdegree	ranked indegree	total indegree	rank_1	rank_2
listed degree	0.1653	0.5075	0.8586	0.2312	0.2934
ranked outdegree		0.1390	0.1183	0.0513	0.0620
ranked indegree			0.6695	0.6124	0.7118
total indegree				0.3186	0.4128
rank_1					0.6747

	rank_2	rank_3	rank_4	rank_5	rank_6	rank_7	rank_8	correlation with $\text{rank}_{1\text{–}8}$
rank_1	0.6747	0.5685	0.5253	0.4882	0.4478	0.4251	0.3918	
rank_2		0.6818	0.6319	0.5970	0.5723	0.5234	0.4697	
rank_3			0.6581	0.6324	0.6030	0.5788	0.5067	
rank_4				0.6429	0.6211	0.5891	0.5523	
rank_5					0.6268	0.5960	0.5429	
rank_6						0.6076	0.5625	
rank_7							0.5762	
rank_8								

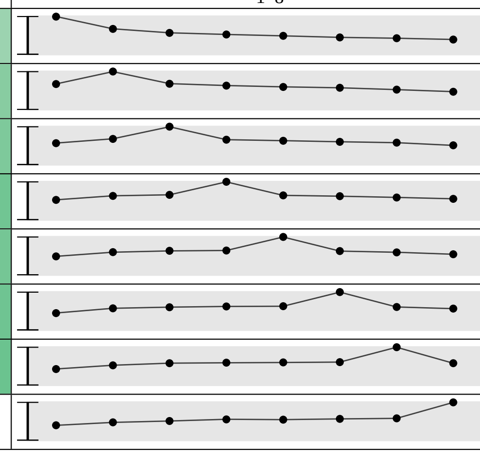

Fig. 2. Correlation of various pairs of degree measures across users. Cells are shaded in proportion to their correlation; for each k, the plot of correlations of rank_k with all other ranks is shown in the kth row.

IV. Using Popularity to Predict Preferences among Friends

One can imagine many ways of attacking the central prediction task described in the previous section. In this paper, we concentrate on purely *popularity-based* predictors. That is, to what extent does user u's decision about which of $\{v, w\}$ to choose as u's #1 friend correlate with the relative popularities of v and w?

To address this question, we filter the ≈11M MySpace profiles to identify all those focal individuals u for which u's #1- and #2-ranked friends also appear in the sample. We culled a population of ≈1.36M (precisely: 1,360,879) profiles using this filtering process.

For the purposes of our main prediction task, we must ensure that we do not "cheat" by baking the correct answer into the measure of popularity. In particular, when we refer to the rank_1 and rank_2 indegrees for the purposes of predicting a user u's #1 friend, we mean the rank_1 and rank_2 indegrees *excluding the ranked edge from* u. (Our other measures of popularity are affected equally by the edges from u to v and w, so excluding this edge makes no difference.)

A scatterplot displaying the popularities of each user's Top 2 friends, under four of the popularity measures discussed previously, is shown in Figure 3. To the extent that these measures of popularity are helpful in differentiating #1- and #2-ranked friends, we would see an asymmetry in the distributions. Specifically, if individuals tend to prefer friends who are more popular, then we would see more points below the diagonal line. What is perhaps visually striking about Figure 3 is that all four panels appear highly symmetric across the diagonal; not only is there no obvious preponderance of points on one side of the diagonal, but the points appear to be distributed close to symmetrically across that diagonal line.

To make this observation more precise, for each measure μ of popularity described previously, we compute the number of users u whose #1-ranked friend is more popular under μ than u's #2-ranked friend is (a *win* for μ); the number of users u whose #1- and #2-ranked friends are equally popular under μ (a *tie*); and the number of users u whose #2-ranked friend is more popular under μ (a *loss*). Figure 4 shows the results.

Because a random guess will be correct on 50% of its predictions, absolute deviation from 50% is the most interesting measure of success. Every popularity-based predictor has a success rate within 0.014 of 0.5; measured by deviation from 50%, the two most successful are rank_1 (51.4%) and rank_2 (49.3%) indegrees: there is a mild tendency for the #1-ranked friend to be ranked #1 more often by others, and for the #2-ranked friend to be ranked #2 more often by others. (See below for some discussion of the phenomenon that rank_1 indegree better predicts #1-rank and rank_2 indegree better predicts #2-rank.) All other measures of popularity perform between 49.5% and 50.2%.

Even the most informative measures give only weak information, and indeed the various measures of popularity even differ in directionality: four of the predictors (listed degree, total sample indegree, rank_1 indegree, and rank_8 indegree) say that individuals (weakly) prefer others who are *more* popular; the remaining eight predictors say that individuals (weakly) prefer others who are *less* popular. For the sake of comparison, two other predictors are displayed in Figure 4: the geographic distance predictor ("u prefers the friend who lives geographically closer to u") and the relative rank predictor ("u

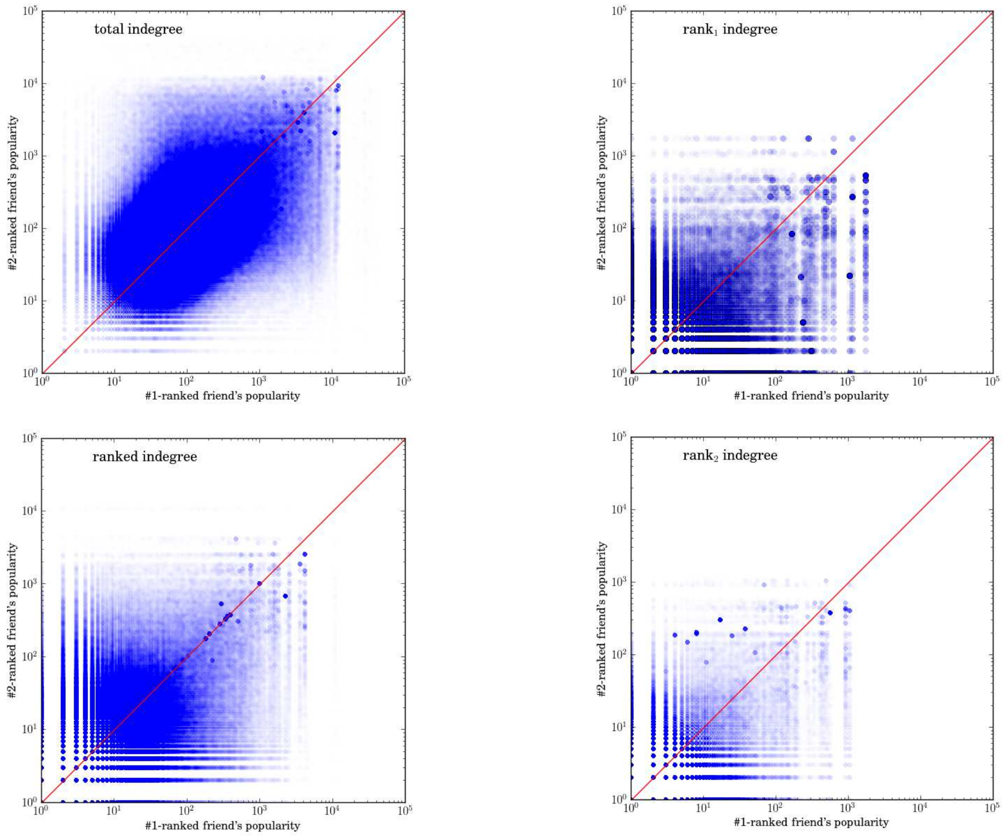

Fig. 3. The popularities of u's #1- and #2-ranked friends under four measures of popularity: total indegree; ranked indegree; $rank_1$ indegree; and $rank_2$ indegree. In each panel, one point is shown for each of the ≈1.36M users in the culled dataset.

measure	*wins*	*ties*	*losses*	$\frac{\text{wins}}{\text{wins+losses}}$
listed degree	672677	19723	668479	0.502
ranked outdegree	588922	172826	599131	0.496
ranked indegree	632726	85509	642644	0.496
total indegree	677825	9224	673830	0.502
$rank_1$ indegree	428626	526959	405294	0.514
$rank_2$ indegree	433075	483153	444651	0.493
$rank_3$ indegree	428523	495158	437198	0.495
$rank_4$ indegree	421599	514973	424307	0.498
$rank_5$ indegree	403552	551255	406072	0.498
$rank_6$ indegree	395258	564024	401597	0.496
$rank_7$ indegree	385062	587947	387870	0.498
$rank_8$ indegree	367007	628465	365407	0.501
geographic proximity [8]				0.564
relative rank [8]				0.689

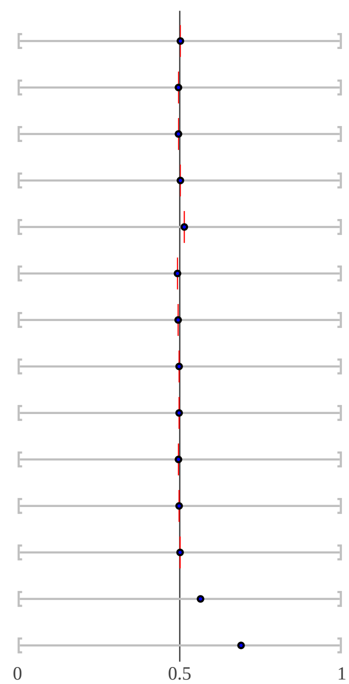

Fig. 4. The performance of each popularity-based predictor on the ≈1.35M culled MySpace users. A *win* for a predictor is an individual for whom $\text{degree}(\#1) > \text{degree}(\#2)$, where #1 and #2 are the best- and second-best friends, respectively. A *tie* is an individual for whom $\text{degree}(\#1) = \text{degree}(\#2)$. A *loss* is an individual for whom $\text{degree}(\#1) < \text{degree}(\#2)$. The smallest number of non-tied data points is $n = 732414$; viewing each of these predictors as a n-trial binomial distribution, the standard error for each of these measures is ≤ 0.0012, and 99.9% confidence intervals are shown as the small vertical red bars.

prefers the friend who ranks u better").

Beyond the "Top 2"

We have focused our discussion thus far on distinguishing #1-versus-#2-ranked friends, but the same calculations can be performed for any pair of ranks. For any two ranks i and $j > i$, we compute the fraction of focal individuals for whom friend #i is more popular than friend #j. (Our previous discussion was for $i = 1$ and $j = 2$.) Figure 5 displays the tables of results for eight predictors, omitting only the $\text{rank}_{5\text{–}8}$ indegree predictors, which are qualitatively similar to the rank_4 table. (As before, we must avoid baking the correct answer into the predictor: when predicting a user u's #k-versus-#j friends, we exclude the two edges from u when computing the rank_k and rank_j predictors.)

Figure 5 reveals that the qualitative pattern of the #1-versus-#2 comparison remains true for other pairs of ranks. For the broader degree measures (listed degree, ranked outdegree, total indegree, ranked indegree), the fraction of individuals who prefer the more popular candidate friend is generally close to or below 0.50, and only rarely greater than half. In each of these cases, the fraction of individuals whose best friend is more popular than their #j-friend decreases with j; for example, fewer than 47% of individuals have a best friend who is more popular than their $\#8$ friend under listed degree, total indegree, and ranked indegree.

We do see the hints of one exception to this trend: out of two candidate friends of an individual u, the one with a higher rank_i indegree *is* generally more likely to be u's $\#i$ friend. (Recall that when predicting u's #i friend, "rank_i indegree" means "rank_i indegree *aside from the edge from u herself*.") This phenomenon is visible in the rank-i row of the rank_i panels in Figure 5: for example, having a higher rank_2 indegree corresponds to a 51.3%–51.8% chance of being ranked #2 instead of being ranked $\{\#3, \#4, \#5, \#6\}$. One partial explanation for this observation is the "Top Friends"–avoiding tactics employed by some users that survived in the data set: choosing a "Top 8" whose profile pictures, letter-by-letter and in order, spelled out a particular four-letter word plus T+O+P+8. The profile with the "U" Scrabble tile—the second letter of this four-letter word—as its profile photo would have a high rank_2 indegree (and a low $\text{rank}_{i \neq 2}$ indegree); this profile would often be correctly predicted to be a #2 friend using the rank_2 indegree predictor. Still, this avoidance behavior appears to relatively rare, and even among the $\text{rank}_{i \in \{1,2,3,4\}}$ indegree predictors, there are slightly more rank pairs in which individuals prefer the less popular friend.

V. Discussion: Indifferent Attachment?

In the earliest days of the present era of computational social network research, Albert-László Barabási and Reka Albert [3] published an influential paper on degree distributions of social networks. Barabási and Albert made two key observations. First, they showed empirically that the degree distributions of real social networks are heavy tailed. (And, they argue, the form of the degree distribution is specifically well modeled by a power law, though Clauset, Shalizi, and Newman [6] raise some serious concerns about the quality of the power-law fit for this type of network data.) Second, Barabási and Albert proposed *preferential attachment (PA)* as a generative model of social network growth. (Both observations were presaged in the literature of other disciplines earlier; see the early work of Yule [27] and Simon [23], and the more recent survey by Mitzenmacher [19].) As other structural properties of social networks have been discovered, alternative generative models have been proposed. These models—e.g., community-guided attachment and forest-fire models [17]—do not seem to make as-obvious predictions about how preferences among individuals will be expressed; thus, we focus our discussion here network-formation models, like PA, with some form of popularity reinforcement—nodes with higher degree gain edges at a higher rate than nodes with lower degree.

Here is the preferential attachment model, in its basic form. We start with a small network, which grows by one node and one edge at each time step. At time step t, a new node u_t appears in the system, and it forms one edge from u_t to an existing node. More popular existing nodes are more likely to be chosen as u_t's neighbor; specifically, the probability that u_t chooses v is directly proportional to the current value of $\text{degree}(v)$. PA is a particular instantiation of what we might call the *preference for popularity*—that, given a choice between two candidate friends, the more popular candidate is the one more likely to be preferred. (Other nonlinear instantiations of this preference occur in other models.)

While the basic form of PA does not speak directly to rankings of friends, the underlying preference for popularity does make particular predictions about ranking. PA can be most straightforwardly adapted to the ranked setting by modeling a node as ranking its neighbors in the order in which edges formed. (So the #1-ranked friend for u is the friend u chose when joining the network; u's #2-ranked friend is the first node $v \neq u$ that chose u when v joined the network, u's #3-ranked friend is the second node that chose u upon joining, etc.) We simulated this Ranked-PA (RPA) network growth model for a 100,000-node network, and observed that friend ranking is much better predicted by popularity in RPA than in MySpace: over 95% of RPA nodes had a best friend that was more popular than their second-best friend, and between 59% and 61% of nodes had a #i-ranked friend more popular than their #$(i+1)$-ranked friend for $i \in \{2, 3, 4, 5\}$. (In PA, the age of a node and the node's degree are positively correlated; thus the edge formed earlier is more likely to have been formed from a neighbor that would eventually become more popular. The #1-ranked friend is special—it was chosen with explicit preference towards high degree, instead of by age—and so its popularity advantage over the #2-ranked friend is much higher than the advantage of #2 over #3 and the other pairs.)

The empirical and modeling observations of Barabási and Albert sparked a large body of literature, empirical and theoretical, that has made a great deal of progress in modeling and analyzing the structural properties of real-world social networks—particularly regarding a hypothesis of the origin of the apparently ubiquitous heavy-tailed degree distributions. But the results shown in Figures 4 and 5, coupled with the simulations of RPA, suggest that a preference for popularity may not provide a full explanation for empirically observed heavy-tailed degree distributions: when a user is choosing which of two friends she prefers, the popularity of the two candidates is at best essentially uninformative about which will

Outdegree

count ≥ 911176
std. error ≤ 0.0010

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.5016	0.4959	0.4898	0.4870	0.4816	0.4750	0.4684
rank2		0.5019	0.4968	0.4929	0.4865	0.4813	0.4740
rank3			0.4990	0.4952	0.4900	0.4837	0.4764
rank4				0.4989	0.4931	0.4871	0.4812
rank5					0.4987	0.4921	0.4857
rank6						0.4978	0.4899
rank7							0.4952

Rank$_1$ indegree

count ≥ 560348
std. error ≤ 0.0013

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.5140	0.5127	0.5104	0.5101	0.5057	0.4990	0.4894
rank2		0.5124	0.5133	0.5141	0.5115	0.5063	0.4954
rank3			0.5066	0.5076	0.5073	0.5008	0.4933
rank4				0.5050	0.5043	0.5001	0.4940
rank5					0.5053	0.5001	0.4965
rank6						0.5000	0.4957
rank7							0.4986

Ranked outdegree

count ≥ 818479
std. error ≤ 0.0011

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.4957	0.4918	0.4895	0.4897	0.4878	0.4860	0.4856
rank2		0.4992	0.4963	0.4950	0.4919	0.4901	0.4888
rank3			0.4991	0.4971	0.4952	0.4927	0.4910
rank4				0.4992	0.4969	0.4957	0.4932
rank5					0.4993	0.4966	0.4949
rank6						0.4992	0.4979
rank7							0.4991

Rank$_2$ indegree

count ≥ 583068
std. error ≤ 0.0013

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.4934	0.5015	0.5032	0.5045	0.5009	0.4943	0.4862
rank2		0.5132	0.5161	0.5175	0.5139	0.5092	0.5016
rank3			0.5079	0.5106	0.5078	0.5033	0.4966
rank4				0.5061	0.5044	0.5007	0.4949
rank5					0.5023	0.5000	0.4951
rank6						0.5014	0.4975
rank7							0.4983

Ranked indegree

count ≥ 872227
std. error ≤ 0.0011

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.4961	0.4894	0.4829	0.4794	0.4753	0.4681	0.4610
rank2		0.5037	0.4985	0.4953	0.4892	0.4829	0.4763
rank3			0.5014	0.4988	0.4933	0.4875	0.4826
rank4				0.5016	0.4968	0.4919	0.4868
rank5					0.5014	0.4954	0.4911
rank6						0.5001	0.4950
rank7							0.4996

Rank$_3$ indegree

count ≥ 583183
std. error ≤ 0.0013

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.4950	0.4882	0.4917	0.4929	0.4912	0.4854	0.4781
rank2		0.4990	0.5041	0.5070	0.5049	0.4991	0.4920
rank3			0.5075	0.5096	0.5088	0.5029	0.4973
rank4				0.5045	0.5045	0.5004	0.4961
rank5					0.5029	0.4994	0.4956
rank6						0.4985	0.4963
rank7							0.4997

Sample indegree

count ≥ 917024
std. error ≤ 0.0010

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.5015	0.4942	0.4876	0.4832	0.4773	0.4691	0.4615
rank2		0.5023	0.4960	0.4918	0.4846	0.4777	0.4696
rank3			0.4990	0.4950	0.4882	0.4804	0.4734
rank4				0.4990	0.4925	0.4860	0.4789
rank5					0.4989	0.4913	0.4840
rank6						0.4979	0.4891
rank7							0.4960

Rank$_4$ indegree

count ≥ 578356
std. error ≤ 0.0013

	rank2	rank3	rank4	rank5	rank6	rank7	rank8
rank1	0.4984	0.4899	0.4827	0.4855	0.4843	0.4788	0.4712
rank2		0.4999	0.4936	0.4966	0.4956	0.4907	0.4829
rank3			0.4974	0.5015	0.5008	0.4950	0.4903
rank4				0.5058	0.5035	0.4996	0.4950
rank5					0.5027	0.4993	0.4951
rank6						0.4992	0.4945
rank7							0.4980

Fig. 5. Results for other rank comparisons. In each panel, the i-versus-j cell displays the fraction of individuals u whose #ith and #j friends are ordered so that u prefers the one who is more popular. (The 1-versus-2 cells correspond to Figure 4.) Red-shaded cells indicate that more individuals prefer the less popular friend, and blue-shaded cells indicate that more individuals prefer the more popular friend; the darker the shading, the further from 0.50. The displayed count is $\min_{i,j}(wins + losses)$ for #i-versus-#j friends; an upper bound on the standard error, viewing each of these predictions as a binomial distribution, is shown as well.

be chosen by that user, and at worst she actually prefers the less popular friend. While the twelve measures of popularity that we consider here are strongly positively correlated, they are not perfectly aligned (Figure 2). But they are all aligned with respect to our central prediction task: each is at best only marginally informative about the preferences of individuals among their closest friends.

It may be the case that the analogy between choosing whether to befriend an individual and choosing whether to rank an individual highly is a weak one; those decisions may be made differently, and the results of this paper may not speak to the underlying "to friend or not to friend" decision. (It is an interesting direction for future research to assess to what extent deciding whether to befriend an individual and whether to rank an individual highly *are* similar or different. An understanding of why, and how, a pair of individuals decide to assign a "friend" label to their relationship is generally missing from the current computational literature on social networks—and this understanding is obviously crucial to properly interpreting what the edges in the network actually mean. Interactions in online social networking sites have some key differences from real-world interactions: the question "does our relationship rise to the level that deserves the 'friends' label?" is rarely explicitly called in the offline world, and the fact that it is continually raised by a social networking site may impose different or stronger social pressures in how we react online. Still, massive digital data presents a promising opportunity to better understand the friendship-or-no decisions.)

But to the extent that ranking decisions and befriending decisions are analogous, our observations suggest that network-growth models based on a preference for popularity miss some important behavioral properties; we will need a different explanation to account for empirically observed heavy-tailed degree distributions. And even if ranking and befriending decisions are fundamentally different, the heavy-tailed degree distribution for rank_1 indegree (Figure 1) seems to require an explanation fundamentally different from preferential attachment.

Mitzenmacher [20] argues compellingly that, as a research community, we must move toward the necessary future direction of *validation* (or, at least, invalidation) in research on heavy-tailed degree distributions. We hope that the present work can serve as a small step forward in that enterprise.

Acknowledgements

Thanks to Peter DeScioli, Elizabeth Koch, Robert Kurzban, Dave Musicant, Jeff Ondich, and Aaron Swoboda for helpful discussions and comments. This work was supported in part by NSF grant CCF-0728779.

References

[1] Lada A. Adamic, Debra Lauterbach, Chun-Yuen Teng, and Mark S. Ackerman. Rating friends without making enemies. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM'11)*, July 2011.

[2] Lars Backstrom, Eytan Bakshy, Jon Kleinberg, Thomas Lento, and Itamar Rosen. Center of attention: How Facebook users allocate attention across friends. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM'11)*, July 2011.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[4] Danah Boyd. Friends, Friendsters, and MySpace Top 8: Writing community into being on social network sites. *First Monday*, 11(12), December 2006.

[5] J. Cheng, D. Romero, B. Meeder, and J. Kleinberg. Predicting reciprocity in social networks. In *Proceedings of the 3rd IEEE Conference on Social Computing (SocialCom'11)*, 2011.

[6] Aaron Clauset, Cosma R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[7] P. DeScioli and R. Kurzban. The alliance hypothesis for human friendship. *PLoS ONE*, 4:e5802, 2009.

[8] Peter DeScioli, Robert Kurzban, Elizabeth N. Koch, and David Liben-Nowell. Best friends: Alliances, friend ranking, and the MySpace social network. *Perspectives on Psychological Science*, 6(1):6–8, January 2011.

[9] R.I.M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22:469, 1992.

[10] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI'09)*, pages 211–220, 2009.

[11] Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. Modeling users' activity on Twitter networks: Validation of Dunbar's number. *PLoS One*, 6(8):e22656, 2011.

[12] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[13] M. R. Henzinger, A. Heydon, M. Mitzenmacher, and M. Najork. On near-uniform URL sampling. *Computer Networks*, 33:295–308, 2000.

[14] Indika Kahanda and Jennifer Neville. Using transactional information to predict link strength in online social networks. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM'09)*, 2009.

[15] Juan Lang and Shyhtsun Felix Wu. Anti-preferential attachment: If I follow you, will you follow me? In *Proceedings of the 3rd IEEE Conference on Social Computing (SocialCom'11)*, 2011.

[16] Debra Lauterbach, Hung Truong, Tanuj Shah, and Lada Adamic. Surfing a web of trust: Reputation and reciprocity on CouchSurfing.com. In *Proceedings of the 2009 International Conference on Computational Science and Engineering*, pages 346–353, 2009.

[17] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

[18] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[19] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.

[20] Michael Mitzenmacher. Editorial: The future of power law research. *Internet Mathematics*, 2(4):525–534, 2006.

[21] M. Najork and J. L. Weiner. Breadth-first crawling yields high-quality pages. In *Proceedings of the 10th International Conference on World Wide Web (WWW'01)*, pages 114–118, 2001.

[22] J. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, May 2007.

[23] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[24] Chun-Yuen Teng, Debra Lauterbach, and Lada A. Adamic. I rate you. You rate me. Should we do so publicly? In *3rd Workshop on Online Social Networks (WOSN'10)*, June 2010.

[25] Stefan Wuchty and Brian Uzzi. Human communication dynamics in digital footsteps: A study of the agreement between self-reported ties and email networks. *PLoS ONE*, 6(11):e26972, 11 2011.

[26] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *19th International World Wide Web Conference (WWW'10)*, 2010.

[27] G. Yule. A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis. *Philosophical Transactions of the Royal Society of London*, 213:21–87, 1925.