Best Friends: Alliances, Friend Ranking, and the MySpace Social Network*

Peter DeScioli
Economic Science Institute
Chapman University

Robert Kurzban
Department of Psychology
University of Pennsylvania

Elizabeth N. Koch†
Department of Computer Science
University of Minnesota

David Liben-Nowell
Department of Computer Science
Carleton College

Abstract

Like many topics of psychological research, the explanation for friendship is at once apparently intuitive and yet simultaneously difficult to address empirically. The standard difficulties worsen when one seeks, as we do, to go beyond "obvious" explanations ("humans are social creatures") to ask deeper questions such as: what is the evolved function of human friendship? In recent years, however, a new window into human behavior has opened as a growing fraction of people's social activity has moved online, leaving a wealth of digital traces behind. One example is a feature of the MySpace social network that allows millions of users to rank their "Top Friends." Here we collect over 10 million people's friendship decisions from MySpace to test predictions made by hypotheses about human friendship. We find particular support for the alliance hypothesis, which holds that human friendship is caused by cognitive systems that function to create alliances for potential disputes. Because an ally's support can be undermined by a stronger outside relationship, the alliance model predicts that people will prefer partners who rank them above other friends. Consistent with the alliance model, we find that an individual's choice of best friend in MySpace is strongly predicted by how partners rank that individual.

*The authoritative version of this paper appears in *Perspectives on Psychological Science*, 6(1):6–8, January 2011. This work was supported in part by NSF grant CCF-0728779, by an NSF Graduate Research Fellowship to the third author, and by a predoctoral fellowship to the first author from the International Foundation for Research in Experimental Economics (IFREE). We thank Carol Drysdale and HongDa Tang for helpful comments.

†Work performed in part at Carleton College.

In recent years, hundreds of millions of people have left behind digital records of their social behavior as they interact via online communities like Facebook, MySpace, and Twitter. These digital traces form a massive naturalistic data source that can address a wide variety of psychological questions: *Alice is now in a relationship. Bob has 150 friends. Charlie has 12 new followers. Alice is now single; Bob likes this.* Our interests in this paper lie in explanations of human friendship, and MySpace has a feature that makes it particularly useful for research on close relationships. MySpace profiles contain a "Top Friends" list, in which individuals designate a subset of their friends as "Top Friends" and organize these friends in a ranked order. Through this feature, MySpace users have created a vast network of ranked friendships. This data source can be used to test fine-grained predictions about people's rankings of friends—and theories about why friendships exist at all.

Humans are unusual in that they form long-term, dyadic relationships with non-relatives, and the evolved function of friendship behavior remains unclear (Silk, 2003). The traditional theory is that friendship is a trade relationship in which people exchange goods and services to reap gains in trade (Trivers, 1971). Cognitive systems designed for trade should closely monitor benefits given and received (Cosmides and Tooby, 1992; Trivers, 1971). However, substantial evidence shows that friends cooperate without closely monitoring contributions (reviewed by Silk (2003)). For friend rankings, trade models predict that people will favor friends who generate more trade surplus than others, without special concern about friends' other friendships.

An alternative theory is the alliance hypothesis, which holds that friendship is caused by cognitive systems designed to assemble a group of allies for potential disputes (DeScioli and Kurzban, 2009). The value of an ally crucially depends on the partner's alliances with others because an individual cannot count on a partner for support when the partner has a stronger alliance with the individual's opponent (Snyder, 1997). This point leads to a key prediction of the alliance hypothesis: *an individual will favor those friends who rank fewer others above the individual.* A partner who ranks the individual first, as the best friend, is particularly valuable because that partner's alliance support cannot be undermined by an outside relationship.

We wrote software to collect a sample of ≈11M MySpace profiles, including demographic data (age, sex, geographic location; Figure 1) and each individual's rank-ordered "Top Friends." We focused on the best-friend network defined by the connections listed in the first-ranked position in each of these ≈11M lists. Of them, ≈3.5M named best friends also in the sample. In the best-friend network, we computed the most common *connected components*, subsets of individuals connected through best-friend links (Figure 2). Broadly, the most common component was a mutual pair of best friends, whereas stars, paths, and other network structures occurred much less frequently (see Supporting Analysis, p. 11).

A central prediction of the alliance hypothesis is that people will be very concerned about how their friends rank them among other friends. We tested this idea by looking at people's decisions about whom they rank first versus second in their "Top Friends." Specifically, we tested whether *relative rank*, defined as alter's rank of ego (relative to other alters' ranks of ego, DeScioli and Kurzban (2009)), predicts first-ranked friendship. We also looked

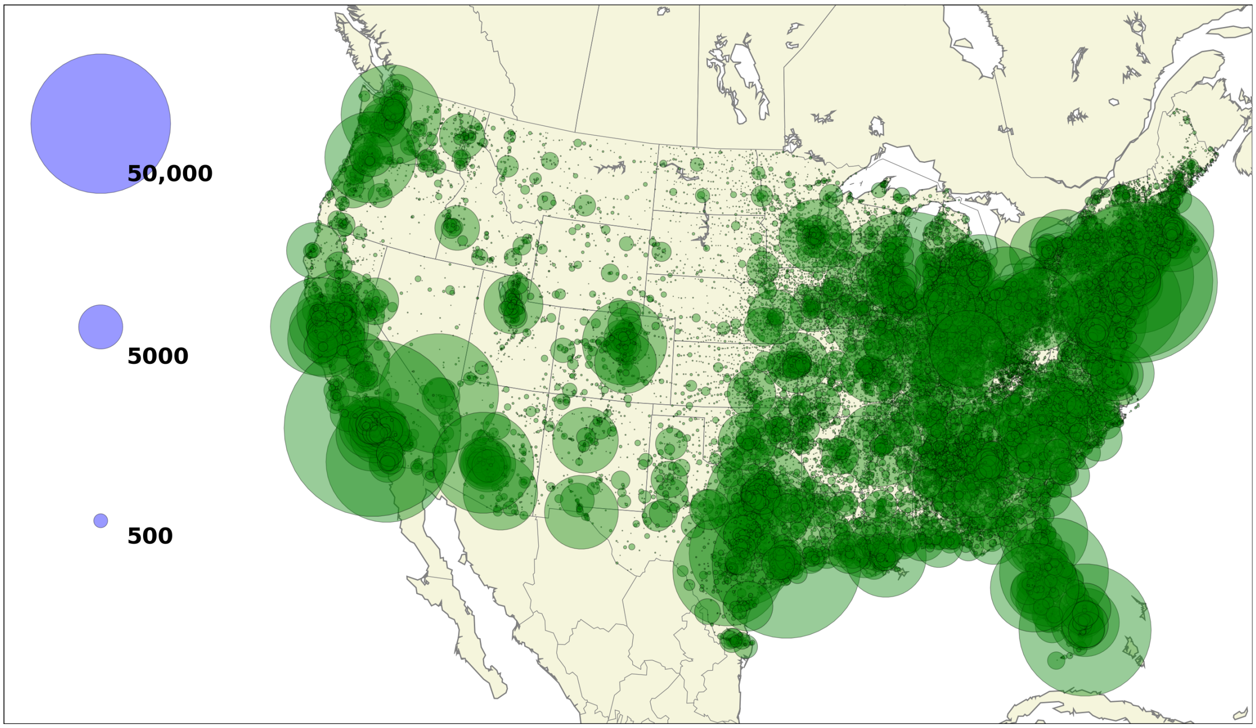

Figure 1: The geographic distribution of our sample from the MySpace network. We were able to determine the locations of 41.7% of crawled profiles in a United States Geological Survey (USGS) database of longitudes and latitudes of cities in the contiguous United States. For each longitude–latitude pair x, a circle centered at x with area proportional to the number of crawled MySpace users declaring their location as x is shown. For scale, circles of sizes corresponding to 500, 5000, and 50,000 people are shown on the left.

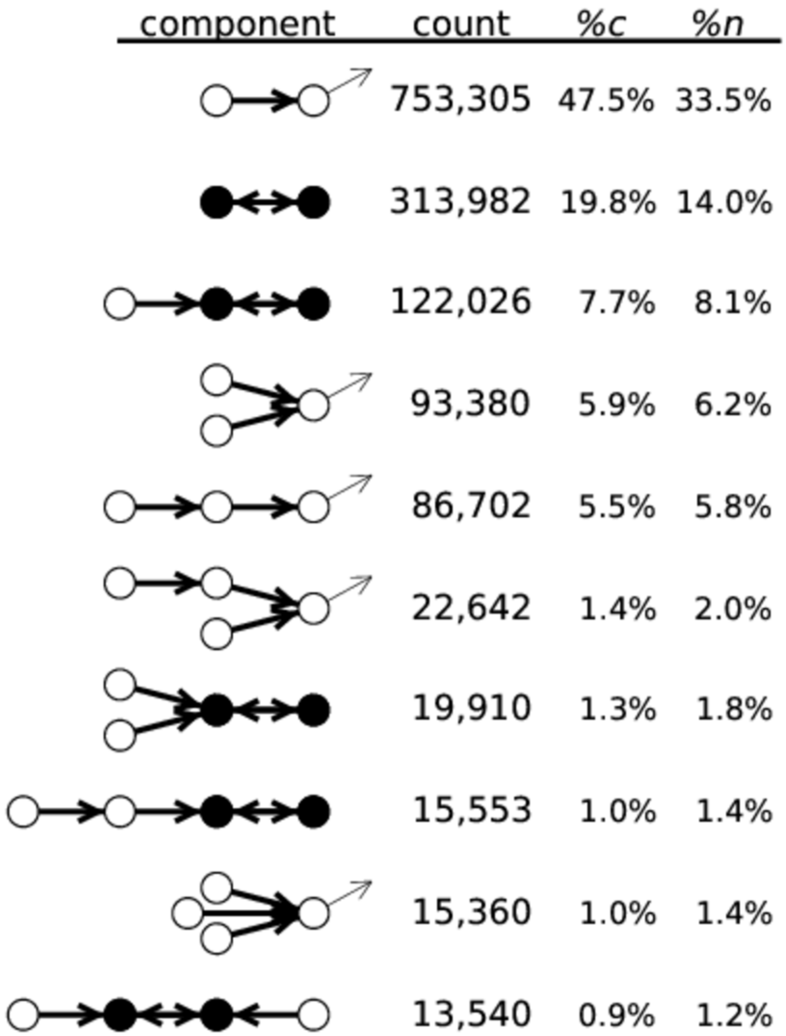

Figure 2: Distribution of connected components in the best-friend network. A person is represented by a circle, and an arrow leads from each person to that person's best friend. The figure shows the ten most frequent components, the number of occurrences, the percentage of components that are of this type (%c), and the percentage of all individuals who occur in this type of component (%n). Shaded nodes are part of the component's cycle (see text). Components without a cycle contain one person whose best friend was uncrawled or a non-person, indicated by an arrow that does not point to a circle. The ten depicted components account for 91.85% of 1,585,561 total components; 70.17% of the 3,445,329 total best friendships; and 75.38% of the 4,495,696 nonisolated people (individuals who list a best friend or are listed as a best friend).

Ego's first-ranked friend ...	All friends		Same-sex friends	
... ranks ego better	881,909	68.85	391,804	66.32
... is geographically closer	114,266	56.41	45,747	53.85
... is opposite sex of ego	534,587	56.85	—	
... is closer in age	1,151,751	50.71	457,005	50.93
... is older	1,199,391	51.36	475,208	51.99
... is ranked more often	1,354,710	$^{\text{ns}}$49.99	562,938	51.51
... is ranked #1 more often	914,703	51.83	381,867	53.36

Table 1: Predictors of ego's first- versus second-ranked friend. Percentages indicate the proportion of individuals whose first-ranked friends had a better predictor value than the individual's second-ranked friend. All percentages differ from chance (50%) at the $p < .001$ level unless labeled "ns." In the same-sex sample, egos are included only if ego, ego's first-ranked alter, and ego's second-ranked alter are all male or all female. In all cases, n denotes the number of individuals who had in-sample first- and second-ranked friends who differed in the predictor variable.

at demographic predictors based on the alter's age, sex, and geographic distance from ego. Finally, we tested two predictors based on network popularity: the number of other members (excluding ego) who declared the alter as their first-ranked friend, and the total number of people who ranked the alter somewhere in their "Top Friends" list.

For each predictor variable, we found the subset of egos for which the first- and second-ranked alters differed in the value of the variable (114,266 < all ns < 1,199,391). We analyzed whether these differences predict first- and second-ranked friendships (Table 1). Individuals showed a weak tendency to choose best friends who were older (51%, $p < .001$, binomial test) and closer in age (51%, $p < .001$). Individuals tended to choose best friends who were opposite sex (57%, $p < .001$), which might reflect romantic partners; however, when we repeated all comparisons using only same-sex friendships, we found the same qualitative patterns for all other variables in the same-sex-only sample (Table 1, Figure 3, Figure 4). The two measures of popularity showed little predictive power: the number of best-frienders had a significant but small effect (52%, $p < .001$), and the number of appearances in a "Top Friends" list was not significant (50%, $p = .76$).

Consistent with the importance of physical proximity (Festinger et al., 1950; Liben-Nowell et al., 2005), individuals chose geographically closer individuals as best friends (56%, $p < .001$). This result extends previous research by showing that geographic distance predicts not only the presence or absence of friendships but also fine-grained distinctions among individuals' highest-ranked friends.

Finally, the key alliance variable, alter's rank of ego, was the best predictor of first-ranked friendship: 69% of egos ($p < .001$) selected the alter who ranked ego better. This shows that knowing only how alters rank ego, researchers can distinguish first- and second-ranked friends with 69% accuracy. The relative rank variable predicted a significantly greater proportion of best friendships than the next best predictor, geographic proximity ($z = 84.44$, $p < .001$).

We conducted the same comparisons between first-ranked versus third- through eighth-

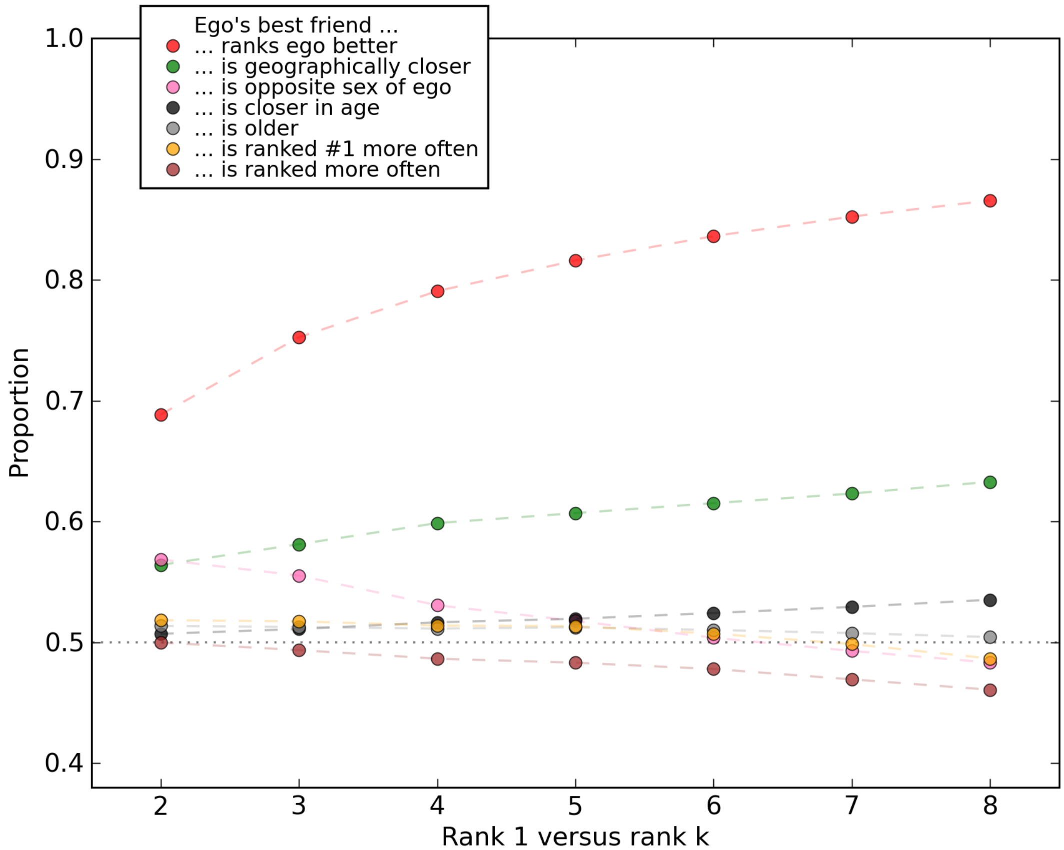

Figure 3: Predictors of first- versus kth-ranked friends. To compare best friends to kth-ranked friends ($k = 2, 3, \ldots, 8$), we found the subset of individuals whose 1st- and kth-ranked friends differed in the predictor variable. (Individuals are excluded if they did not have a person as their 1st- or kth-ranked friend, or if the predictor variable cannot distinguish those two friends because of a tie or missing data.) Each data point shows the proportion of individuals whose first-ranked friend had a better predictor value than the individual's kth-ranked friend. For each point: $88{,}204 \leq n \leq 1{,}354{,}710$. All values differ from chance (50%) at the $p < .001$ level except: 1 vs. 2, "ranked more often"; and 1 vs. 7, "ranked #1 more often."

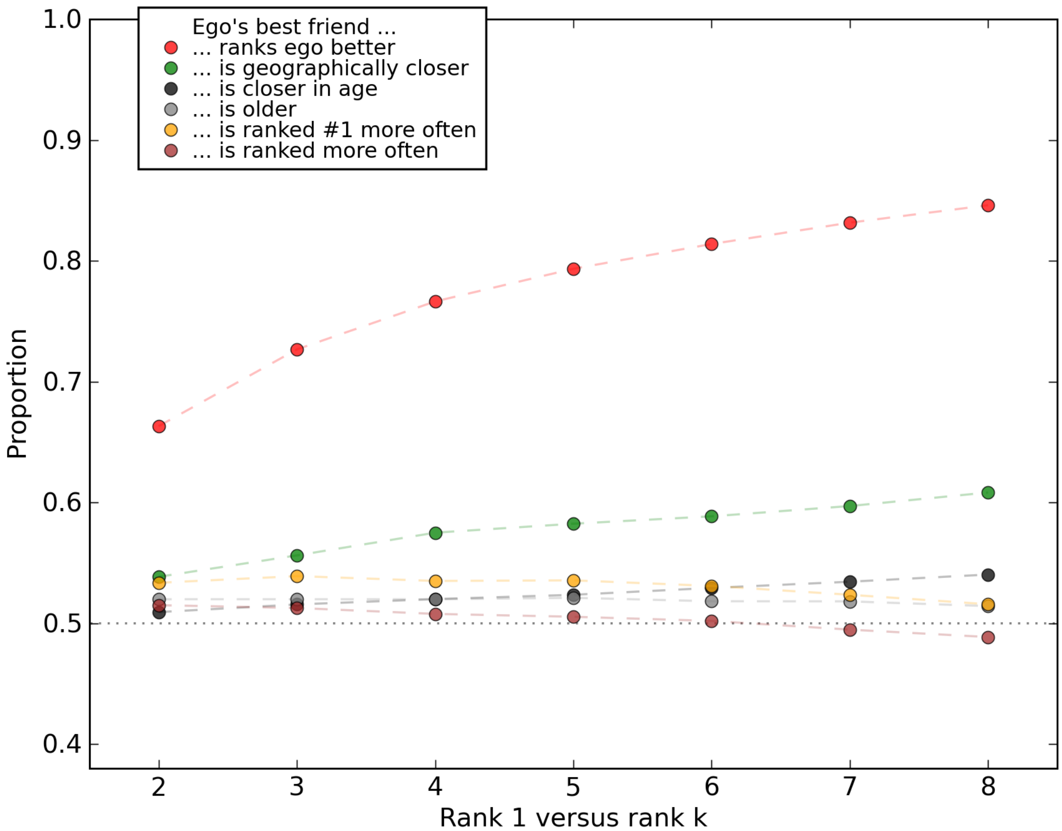

Figure 4: Predictors of first- versus kth-ranked friends restricted to same-sex friendships. Egos are included only if ego, ego's 1st-ranked alter, and ego's kth-ranked alter are all male or all female. All values differ from chance (50%) at the $p < .001$ level except: 1 vs. 6, "ranked more often."

ranked friends (Figure 3). The predictive performance of relative rank increased monotonically to 89% for the first-versus-eighth friend comparison. Geographic proximity also showed an increasing trend up to 63%.

In sum, we found that the key alliance variable—relative rank—strongly predicted whom people identified as their first-ranked friend rather than second-ranked friend: 69% of egos chose for first rank the alter who ranked ego better. This factor was considerably stronger than geographic proximity and also performed better than demographic predictors and variables measuring popularity.

Our findings provide converging evidence for the alliance hypothesis along with previous research (DeScioli and Kurzban, 2009). Previous work used survey methods to measure a dozen properties of participants' ten closest friendships. The study showed that participants' perceptions of how their friends rank them was the best predictor of their own friend rankings. This result held after controlling for all other measured variables, including similarity, received benefits, friendship duration, frequency of contact, and traits such as generosity and attractiveness.

The present results add to previous work by testing the alliance hypothesis on a dataset of real-world friendship decisions. In contrast to laboratory surveys, MySpace users make friendship decisions in a public forum where they have real-life consequences. These consequences can be observed by browsing MySpace profiles where people frequently discuss the prominently displayed friend rankings (usually asking a friend for a higher rank). Further, our MySpace dataset allows us to embed our observations in a network that includes "both sides" of friendships, whereas previous work looked at individuals' perceptions about friends' rankings. Finally, the MySpace network provides data from millions of individuals spanning a wide geographic range, providing sampling advantages over traditional methods.

This work shows how functional theories, traditional research, and massive datasets derived from traces of online behavior can be combined to generate insight into human social behavior. Because the human mind consists of mechanisms with evolved functions, functional theories can provide a powerful framework for generating testable hypotheses. Traditional research can test hypotheses using controlled laboratory methods, typically with samples in the hundreds. Online social networks can provide enormous datasets which are thousands of times larger and measure naturalistic behavior with real-world consequences. All three perspectives on the research question are valuable, and here, they converge to suggest that human friendship might be caused by cognitive systems that are functionally organized for alliance-building. If the alliance hypothesis is correct, then the same strategic dynamics that shape coalitions among nations might have shaped the evolved cognitive systems that explain why humans have friends and best friends.

References

Cosmides, L. and Tooby, J. (1992). Cognitive adaptations for social exchange. In Barkow, J., Cosmides, L., and Tooby, J., editors, *The adapted mind*, pages 163–228. Oxford University Press, New York.

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

Festinger, L., Schachter, S., and Back, K. (1950). *Social pressures in informal groups: A study of human factors in housing.* Harper, New York.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., and Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102:11623–11628.

Silk, J. (2003). Cooperation without counting: The puzzle of friendship. In Hammerstein, P., editor, *Genetic and cultural evolution of cooperation*, pages 37–54. MIT Press, Boston.

Snyder, G. H. (1997). *Alliance politics.* Cornell University Press, Ithaca.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 45:35–57.

Supporting Information

Supporting Methods

We crawled MySpace between October 2007 and February 2008 by using a cluster of desktop machines running a parallelized variant of breadth-first search (BFS). There are sampling biases in a partial BFS-style crawl of a massive-scale network such that individuals in our sample might have higher than average popularity or network "centrality" (Henzinger et al., 2000; Najork and Weiner, 2001). Our view of the network is also "blurry" because the crawl was performed over several months. We did not crawl profiles that were designated as private or profiles that had more than 20,000 friends listed, and we discarded profiles with syntactically anomalous data. For all crawled users, we collected the demographic data stored in their profile (age, sex, geographic location) and their list of top friends. We separated personal profiles from non-personal profiles (e.g., music bands) based on whether the profile listed age and sex. We crawled 10,979,634 personal profiles and 1,805,706 non-personal profiles; and we observed 7,755,385 private profiles. The ≈11M people in our sample were 46% female, had a mean (SD) age of 27 (16), and were from a wide range of geographic locations (Figure 1).

We examined the best-friend network, which is defined by the connections listed in the topmost position in each of the ≈11M individuals' "Top Friends" lists. In our sample of ≈11M personal profiles, 43% (n = 4,684,325)) identified best friends whose profiles we did not observe. Most of the remainder identified in-sample persons as best friends (n = 3,445,329; 32%), whereas the others chose non-persons such as music bands (n = 1,063,167; 10%), private profiles (n = 1,759,335; 16%), or profiles that we ignored due to syntactic anomalies (n = 27,278; 0.25%). Thus, most individuals (76% of those who chose an in-sample non-private best friend) filled their best-friend slot with a personal profile. We concentrate the present analysis on the ≈3.5M crawled individuals who identified a crawled person as their best friend.

Overview of Best Friends on MySpace

We use standard terminology in referring to a focal individual as "ego" and ego's ranked friends as "alters," noting that ranks are not necessarily symmetric, i.e., ego's rank of alter can differ from alter's rank of ego.

Individual egos tended to choose as a best friend an alter who was close in age, close in geography, same-sex, and who ranked ego first on the alter's "Top Friends" list. In our sample of ≈3.5M best friendships, the median absolute age difference was 2 years, interquartile range (IQR) = 1–6 years. The proportion of same-sex best friends was 55% for females and 58% for males. We computed geographic proximity for individuals in the contiguous United States by querying individuals' locations in a database (`geonames.usgs.gov`) to determine longitude and latitude (Figure 1). We located 41.7% (n = 4,578,321) of the crawled individuals (the remainder were outside the contiguous U.S. or did not list an official municipality). The median distance between ego and best friend was 4.16 km, IQR = 0–148 km ($n = 800{,}915$).

We looked at how ego, the "best-friender," was ranked by the alter whom ego ranked as best friend. For egos who were ranked by their best friend ($n = 2{,}225{,}212$), we computed the proportions of egos whose best friend ranked ego first through twelfth. The mode was at rank one, accounting for 48% of egos, another 17% of egos were ranked second, and across subsequent ranks the proportion of egos steadily decreased. We plotted the same proportions for the second- and third-ranked friends (Figure 5). The aggregate pattern suggests that individuals tend to choose best friends who rank them highly, especially alters who rank ego first.

Supporting Analysis

Prominent theories of friendship make different predictions about the network structure of best friendships. Some theories, such as alliance models and assortative models (e.g., homophily theory, McPherson et al. (2001)), predict that the network will be largely composed of pairs of mutual best friends, due to preferences for loyalty and exclusivity (in alliances) or preferences for similarity (in age, sex, geographic location, etc.). In contrast, other theories predict that the network should have many star structures, which contain multiple best-frienders organized around a central popular individual, due to preferences for specific traits such as intelligence, caring, or attractiveness (Dion and Berscheid, 1974; Langlois et al., 2000). A third possibility is that people do not view best friends as closer than other friends. Fiske (1992) argued, for instance, that people's closest relationships are based on a "principle of equivalence" which "makes it impossible to make graduated differentiations among people" (p. 716). If so, then individuals' rankings of close friends will be essentially arbitrary. This theory predicts best-friend networks that primarily contain relatively long paths of individuals, most of whom have a best-friender but choose someone else as best friend.

To test these ideas, we first looked at the distribution of best-friend slots in our sample. Overall, the 3,445,329 best-friend slots in our sample were distributed across 2,486,712 unique individuals. We computed the indegree distribution—that is, the proportion $N(k)$ of the ≈2.5M individuals with exactly k best-frienders, for each $k \geq 1$. Consistent with previous results (e.g., (Barabási and Albert, 1999)), this distribution was heavy-tailed. Most relevant here, the mode was at one and the distribution dropped off steeply. We observed $N(1) = 79.58\%$, $N(2) = 15.20\%$, $N(3) = 3.22\%$, $N(4\text{–}8) = 1.65\%$, and $N(9+) = 0.36\%$, showing that most people who were chosen as a best friend had only one best-friender. We also computed the proportion $T(k)$ of the ≈3.5M individuals whose best friend was chosen by exactly k best-frienders. These values were $T(1) = 57.44\%$, $T(2) = 21.94\%$, $T(3) = 6.96\%$, $T(4\text{–}8) = 5.76\%$, and $T(9+) = 7.90\%$, showing that most people chose best friends who were not chosen by others.

Next, we analyzed the network in terms of the graph-theoretic concept of a *connected component*—namely, a substructure of the network consisting of all people (nodes) connected

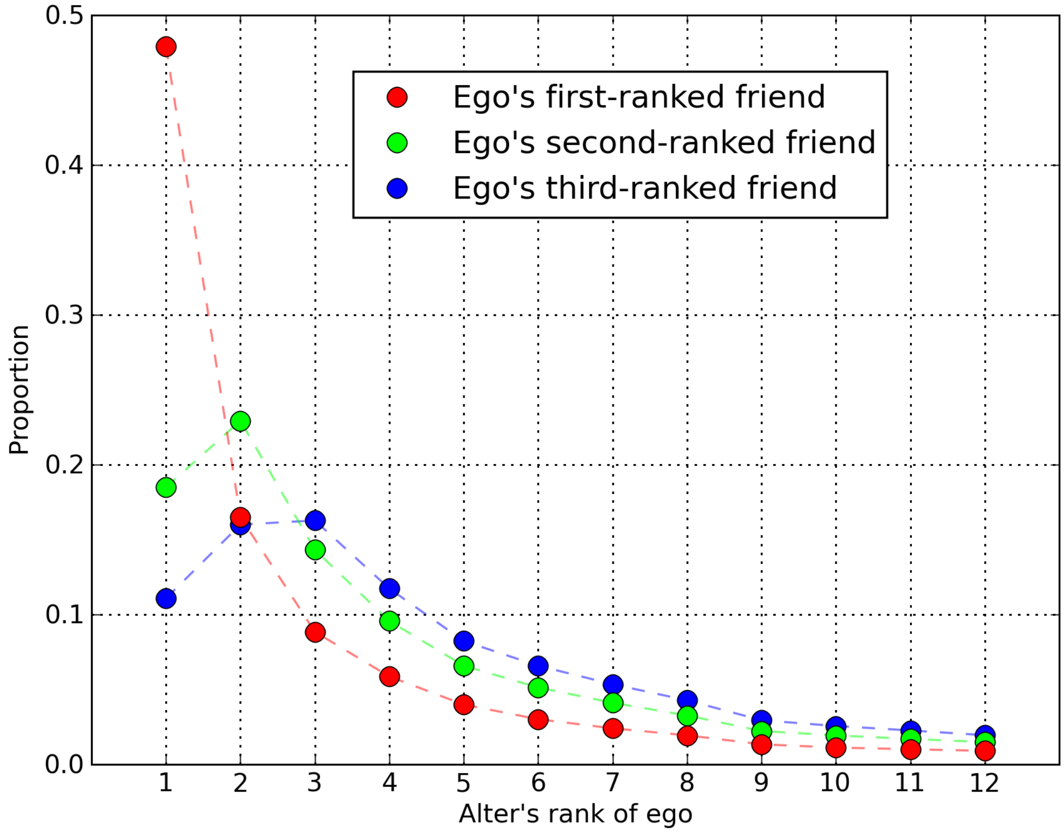

Figure 5: The distribution of the rank of ego by ego's first-ranked friend. Egos who were not ranked by their top-ranked friend were excluded. Of the remaining egos (n = 2,225,213), the proportion who were assigned each rank 1–12 by their top-ranked friend is displayed. Ranks greater than twelve are included in the proportion values but omitted from the figure. The analogous plots for ego's second-ranked friend (n = 1,985,729) and third-ranked friend (n = 1,776,654) are also shown.

to each other via any sequence of best-friend links (edges). The ≈3.5M best-friend links partitioned the ≈11M-person network into 1,585,561 connected components that contained 2+ people each; these components included a total of 4,495,696 individuals. All individuals have exactly one best friend, which implies that each component can contain only one *cycle*, a sequence of nodes such that following the directed edges from a given node leads through the other nodes and back to the original node. We characterized components according to the number of nodes in the component's cycle, the *cycle length*. In our sample, 66% of components (n = 1,050,367) did not contain a cycle because some people listed unobserved, private, or non-personal profiles as their top friend. Of the components that did contain a cycle (34%, n = 535,194), nearly all (n = 533,168) contained a cycle of length 2, a pair of mutual best friends. In comparison, 0.1% contained a cycle with 3 nodes and 0.007% contained a cycle with 4 or more nodes. Figure 2 shows the most common connected components. Isolated mutual pairs accounted for 20% of observed components and another 8% of components consisted of a mutual pair with a third individual pointing to one of the pair. These results show that mutual pairs occurred frequently relative to paths, stars, and other structures.

The patterns in the MySpace best-friend network are consistent with friendship models leading to networks composed primarily of mutual pairs. These models include the alliance hypothesis, in which individuals prefer high rank per se, and assortative models, in which friends' symmetric ranks occur as a byproduct of other preferences such as an attraction to similar others. In contrast, our results contradict other models in which best friends are chosen based on traits such as attractiveness, intelligence, or generosity. If humans chose best friends based on widely valued and easily observable characteristics, then this would lead to networks composed of star structures with particularly valued individuals attracting a number of best-frienders. Finally, our results contradict equivalence models in which individuals do not discriminate within their group of close friends. The MySpace network structure shows that humans, like several other species (Connor et al., 2001; de Villiers et al., 2003; Emery et al., 2007; Holekamp et al., 2007), have strong partner preferences which lead to mutual pairs in friendship networks.

Supporting References

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

Connor, R. C., Heithaus, M. R., and Barre, L. M. (2001). Complex social structure, alliance stability, and mating access in a bottlenose dolphin 'super-alliance'. *Proceedings of the Royal Society of London B*, 268:263–267.

de Villiers, M. S., Richardson, P. R. K., and van Jaarsveld, A. S. (2003). Patterns of coalition formation and spatial association in a social carnivore, the african wild dog (*Lycaon pictus*). *Journal of Zoology*, 260:377–389.

Dion, K. K. and Berscheid, E. (1974). Physical attractiveness and peer perception among children. *Sociometry*, 37:1–12.

Emery, N. J., Seed, A. M., von Bayern, A. M. P., and Clayton, N. S. (2007). Cognitive adaptations of social bonding in birds. *Philosophical Transactions of the Royal Society B*, 362:489–505.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99:689–723.

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

Holekamp, K. E., Sakai, S. T., and Lundrigan, B. L. (2007). Social intelligence in the spotted hyena (*Crocuta crocuta*). *Philosophical Transactions of the Royal Society B*, 362:523–538.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., and Smoot, M. H. M. (2000). Maxims or myths of beauty? a meta-analytic and theoretical review. *Psychological Bulletin*, 126:390–423.

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

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