All Friends are not Equal: Using Weights in Social Graphs to Improve Search

Sudheendra Hangal, Diana MacLean, Monica S. Lam, Jeffrey Heer
Computer Science Department
Stanford University
Stanford, CA 94305
{hangal,malcdi,lam,jheer}@cs.stanford.edu

Abstract

In conducting a search for a person in an online social network, most contemporary social networking sites return source-target paths based purely on degrees of separation. Not only does this fail to reflect social tie strength (how well two people know one another), but it also does not model asymmetry in social relationships (i.e. person A may pay attention to person B does not mean that person B will reciprocate). We show that search in social networks can be made more effective by incorporating weighted and directed *influence* edges in the social graph, thus capturing both tie strength and asymmetry. We study two large real-world networks, DBLP (a computer science bibliography) and the network formed by one month of Twitter retweet data. We model directed influence between individuals based on their interactions. Our experiments show that for these social networks, the best paths according to our metric are not necessarily the shortest paths: a longer path is better in 68% of searches in Twitter and 45% of searches in DBLP. Further, even when the path length is the same, we show that the best path is often significantly better than a random shortest path of that length.

Our results suggest that social network sites should consider incorporating weights into their search functionality to improve the quality of results.

1. Introduction

Online social networking sites, such as Facebook, LinkedIn and Flickr commonly model social relationships as binary; two people are either "friends" or they are not. This is a coarse approximation to real life, in which people maintain a large number of relationships with *varying* tie strength: close friends, casual acquaintances, family, colleagues, and so on.

Intuitively, several socially-oriented tasks (such as social search, community discovery etc.) could benefit from embedding tie strength into their algorithmic frameworks. In this paper, we explore this hypothesis in the context of *global social search*: the problem of finding the "best" path from one person to another in a social network given global network information.

One can imagine several real-life examples of global social search: Jack, who wants to work at Google, might consult his LinkedIn network to see if he knows anyone well-placed to recommend him. Or John, who has a crush on Mary, might consult Facebook to see whether they have any friends in common who could either arrange an introduction or put in a good word for John.

The problem of using binary relationships for global social search should be apparent to anyone who has conducted a search on the LinkedIn network to find a path to a target person. Since LinkedIn treats all relationships evenly, the search returns the shortest path to the target. This path tends to go through *weak* ties to highly connected people: typically those whose jobs involve some sort of professional networking (e.g. recruiters). However, a longer path through stronger ties may yield a superior result. For example, consider the following 2 paths from A to B: $P_1 =< A, C, D, B >$ and $P_2 =< A, E, B >$. If A and B are virtually strangers to E, but C and D are close friends, and family members of A and B respectively, then P_1 is more likely to yield an introduction of A to B than P_2. Without edge weights, the problem of selecting the right path is even worse when conducting a search not for a specific person, but for a generic attribute like "what's the best route in my social network to anyone employed at Amazon ?" This problem is easily extrapolated to our examples given above, as well as other scenarios, such as a site like thread.com that attempts matchmaking based on an underlying social graph.

1.1 Influence in Social Networks

We generalize the notion of tie strength in the social networking literature to a formal, quantitative measure of asymmetric "influence" in online social networks. If a social graph has information on directional interactions, we model the influence A has over B as the fraction of B's interactions that are with A. The influence of a path in the social graph is defined as the product of the influences of its edges. If A has high influence over B, then B is most likely to honor the request to forward the message towards its eventual destination. It would be most effective if the request were routed through the path that has the greatest influence.

That weak ties are extremely important in real-life social networks (e.g. in finding jobs) has been well accepted by sociologists since the 1970s [4, 5]. Therefore, a network like LinkedIn loses information when it asks users to accept links only with people they know well and disregard invitations from others. In spite of this recommendation, it is common social practice for LinkedIn users to connect with people they know only slightly. We argue that the value of online social networks is precisely that they have the capacity to capture ties of varying types and strengths, and that this capacity should be used effectively.

1.2 Contributions

In this paper, we make the following contributions: we define *influence* as an edge weight metric that is calculated based on relative fractions of interaction between two nodes. We define the "best" path between 2 people A and B as the most *influential*: that which optimizes the chance of A's message is delivered to B. Positing that the most influential paths between two nodes are not always the shortest paths, we conduct an experiment on two social networks (DBLP and Twitter retweets) in order to compare the relationship between path length and influence. We find that the most influential paths are often *not* the shortest, suggesting that the incorporation of edge weights may improve the performance of global social search. Furthermore, this approach is also useful in finding the best path among paths of the same length.

The rest of this paper is structured as follows. We first define our influence metric in Section 2, followed by a model for global social search in Section 3. We present our algorithm for finding the most influential path in a network in Section 4. Next, we describe our experiments and results in Section 5, and provide a broad discussion in Section 6. We compare with prior work on social search and inducing edge weights in Section 7 and conclude in Section 8.

2. Modeling Influence

The problem of global social search can be viewed as a problem of routing requests in a social network. Therefore, a natural optimization is to find the path to the target along which one has the most *influence*. As discussed above, the success of a search lies in finding a path such that each node has influence over the succeeding node.

2.1 Social Interactions

To model influence, we start with an estimation of tie strength. One simple way of estimating tie strength between two nodes is to count their mutual friends. This method is appealing as it relies solely on graph structure, and does not require further meta-information about each edge.

We chose to use a more sophisticated metric based on actual interactions between the individuals involved. In many graphs, such data about actual communication or collaboration between individuals is available (such as the number of Facebook Wall posts, or e-mail messages exchanged between two people.) As interaction involves some degree of time and effort on the part of the participants, the number of interactions is an informative measure of tie strength. One could also imagine combining mutual friends and communication frequency to get a better estimate of tie strength. In this paper, however, our model only uses interaction counts.

There are many different kinds of interactions. People working on a paper together must spend a nontrivial amount of time together, whereas a person can follow another on twitter and the followed has no knowledge of the follower. Interactions may be non-directional as in the case of writing a paper, and directed as in the case of twitter.

2.2 Asymmetry of Influence

Influence is asymmetric: A has high influence on B does not mean the vice versa is true. A simple scenario using Twitter illustrates the asymmetry. Consider two nodes: "Obama" and "Joe the Plumber". Joe likes to retweet Obama. In fact, he has retweeted Obama 1,456 times! Obama, on the other hand, has never referred to Joe in his tweets. Now, if Joe wanted an introduction to an acquaintance of Obama's, it might be a mistake to go through Obama: he has no influence over him. It would be easy for Obama, on the other hand, to get Joe to introduce him to one of his friends as Obama has high influence over Joe. In other words, if A retweets B, B has influence over A.

A second example using a co-authorship network illustrates that asymmetry is not limited to directed interactions. Consider a Ph.D. student and his thesis advisor. Because advisors are frequently co-authors on publications, the proportion of the student's publications that are shared with the advisor are high; the advisor has high influence on the student. For the same reasons, the proportion of the professor's publications that are shared with the student are low; the student has comparatively lower influence on his advisor. Thus, even though the interaction in this case is bidirectional, there is still a difference in influence between two collaborators. This difference can be attributed to the proportion of the collaboration to the total amount of work each collaborator performs.

A perhaps more intuitive way of conceptualizing influence is as the complement of personal investment. A person distributes personal investment amongst her acquaintances. For example, if person B invests a lot of his time in person A, then A has high influence over B.

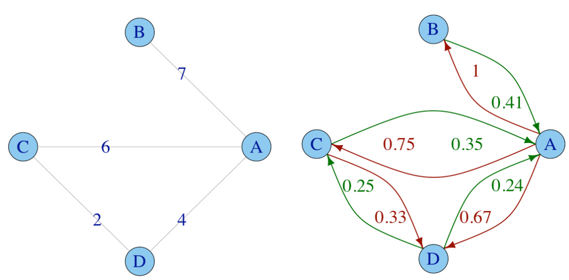
We model the influence from A to B, Influence(A, B), as the proportion of B's investments on A. Let Invests(B, A) be the investment B makes on A.

$$\text{Influence}(A, B) = \frac{\text{Invests}(B, A)}{\sum_X \text{Invests}(B, X)}$$

A non-directional interaction can simply be modeled as having two investments, one in each direction.

The analogy between influence and investment carries over well to real life. We have control over how we distribute our investments. We have less control, however, over who invest in us. In real life, influence is a quality that must be given by others.

The influence of an edge in a social graph is always between 0 and 1. This enables two edges in the graph be compared easily in the global search algorithm. Figure 1(a) depicts an undirected graph showing the investments as weights, and Figure 1(b) shows the same graph with edges weighted by influence.



(a) Before inferring influence (b) After inferring influence

Figure 1. This figure illustrates the results of allocating influence to the edges in an undirected network. An intuitive interpretation of this graph runs as follows: imagine that node A is an adviser, and nodes B, C, and D are her students. The edge weights in Figure 1(a) depict the number of co-authorships between node pairs. In Figure 1(b) we see that the adviser holds more influence over her students than her students hold over her. Moreover, student D, who has authored fewer papers than student C, is more influenced by student C because a larger proportion of his total publications involve student C.

2.3 Influential Individuals

A person is *influential* if he has high influences on many people. We define the *influence of a node* as the sum of the influences the node has on others. That is, the influence of a node A,

$$\text{Influence}(A) = \sum_X \text{Influence}(A, X)$$

3. Model for Social Search

We model the social search problem as one of finding the "strongest" path in a weighted and directed graph in which nodes represent people, and edge weights, ranging from $[0, 1]$ represent directed influence between nodes. A high influence from A to B corresponds with a high probability that B will forward A's message to the desired target, whether that be the end goal or another intermediary along the path. However, we also note that there is some decay factor associated with path length: similarly to the game of "broken telephone", the longer the path, the more likely a message will be dropped. The probability, then, that the message will make it from source to destination is simply the product of the weights of each edge in the path, multiplied by a discount factor that models decay according to path length. We set the discount factor to 0.95. We refer to this probability as the *strength* of the path and calculate it as follows: for a path P of length $|P|$ that contains edges edges $e_1, e_2, \ldots, e_n$, the strength is:

$$\text{S}(P) = (0.95)^{|P|} \prod \text{Influence}(e_i), e_i \in P.$$

Our goal is to find the path that maximizes this probability.

We note that there are several other intuitive definitions of path strength. One idea might be to impose an incremental decay on edge weight proportional to its distance from the source (that is, the decay factor decreases with each hop). Going beyond tie strength, we could also label edges type of relationship and then filter queries along specified labels. We do not deal with incremental decay or edge labels in this work, but note that they are both natural extensions of our above definition of path strength, which we chose for its simplicity and intuitive appeal.

4. Algorithm: Computing the Strongest Path

We compute the *shortest* path from A to B using Djikstra's algorithm. To compute the *strongest* path, we make a natural adaptation to the Djikstra algorithm. Given specific source and target nodes, we would like to find a path P from the source to the target that maximizes:

$$\prod \text{Influence}(e_i), e_i \in P$$

.

Therefore we would like to maximize

$$\sum \log(\text{Influence}(e_i)), e_i \in P$$

and therefore to minimize

$$\sum -\log(\text{Influence}(e_i)), e_i \in P$$

which leads us to minimize

$$\sum \log(1/\text{Influence}(e_i)), e_i \in P$$

Therefore given edge weights e_i between A and B, we can compute the strongest path by simply providing $\log(1/\text{Influence}(e_i))$ as the starting edge weights to the shortest path algorithm to.

5. Experiments

We evaluate the benefit of using weights during global social search on two large networks, DBLP and Twitter retweets. In both networks, influence may be inferred from interactions between individuals. We chose these datasets because they were large, realistic and provided a global view of the network needed for our quantitative evaluation of weighted paths. We specifically studied these two networks for their representative diversity: the DBLP data forms a dense network in which ties are precipitated from intense, real-life social interaction. The Twitter dataset, on the other hand, is directed and more sparse; furthermore, ties do not necessarily represent real life social interaction. Our expectation was that the use of such different networks in our experiments leads to richer feedback on our model assumptions and more representative experimental results.

5.1 DBLP Computer Science Bibliography

The DBLP dataset[1] includes approximately 2.06 million papers with 775,143 unique author names. We take the social graph $G = (V, E)$ where V is the set of all authors and $E = \{(v_i, v_j) : i \neq j, v_i, v_j \in V$ and v_i, v_j are co-authors on a paper$\}$. Considering only the giant component reduces this graph to approximately 1.78 million papers and 603,237 unique authors.

We use the number of papers on which both v_i and v_j are co-authors as a measure of investment. To induce a directed, influence-weighted graph, we create directed edges between each connected pair of authors by computing the proportion of shared papers between each pair relative to the total interactions of each author with all others (as discussed in Section 2). That is, if $\text{Papers}(v_i, v_j)$ is the number of papers co-authored by v_i and v_j, then

$$\text{Influence}(v_i, v_j) = \frac{\text{Papers}(v_j, v_i)}{\sum_{v_k} \text{Papers}(v_j, v_k)}$$

Our resulting graph contains approximately 4.06 million edges. Figure 2 depicts a histogram of the natural log node influence distribution.

In using the DBLP dataset, we make the underlying assumption that a large number of co-authorships between

[1] Available at http://dblp.uni-trier.de/xml/

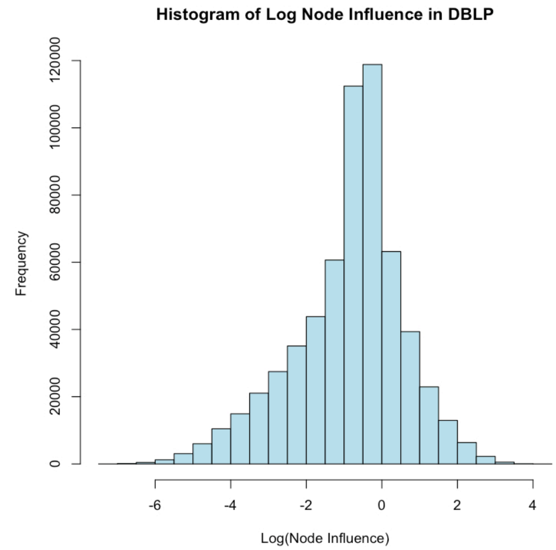


Figure 2. More nodes in the DBLP dataset have an influence < 1, which is congruent with the observation that influence in real life is asymmetric. A small number of nodes are highly influential - but not nearly as influential as those in the tails of the Twitter influence distribution, shown in Figure 3.

two actors indicates that they have a strong tie. We note that "weak" ties on DBLP are probably not all that weak: since writing a paper is a significant time investment, after all! Of course, low influence edges may be more indicative of non-overlapping research interests rather than weak social influence between two actors. However, for experimental purposes we restrict ourselves to the "DBLP universe" and simply influence computation using co-authorship count.

5.1.1 DBLP Results

To compute the effect of weights on shortest path between two nodes, we compute the probability of success along 2 different paths:
a) P_{SHORT} the shortest path based on the number of hops. Since there may be many paths with the same shortest length, we pick a random path from amongst all such paths.
b) P_{STRONG} the strongest path. Strongest path is computed using the weights model described in Section 4.

Once we pick a path P_{SHORT}, we compute the strength $\text{S}(P_{\text{SHORT}})$ along that path, and compare it with the strength of the strongest path in the weighted graph, $\text{S}(P_{\text{STRONG}})$.

We conducted trials on a set of 500 randomly selected source/destination pairs. Table 1 summarizes the results.

5.2 Twitter

Our Twitter dataset consists of 1 month's worth of tweets from Twitter. Considering only retweets (RTs) yields a directed graph, weighted by number of tweets from one user to another, comprising approximately 2.4 million unique

	All Pairs	$\lvert P_{\text{STRONG}}\rvert > \lvert P_{\text{SHORT}}\rvert$	$\lvert P_{\text{SHORT}}\rvert = \lvert P_{\text{STRONG}}\rvert$
Papers	500	215 (43.0%)	285 (57.0%)
Avg. $\lvert P_{\text{SHORT}}\rvert$	6.5	6.6	6.5
Avg. $\lvert P_{\text{STRONG}}\rvert$	7.0	7.8	6.5
Avg. improvement in influence $(\frac{\text{S}(P_{\text{STRONG}})}{\text{S}(P_{\text{SHORT}})})$	548	605	506

Table 1. Summary statistics for experiment results conducted on the DBLP dataset. $\text{S}(P_{\text{STRONG}})$ and $\text{S}(P_{\text{SHORT}})$ denote the strength of the strongest or shortest path, respectively. $\lvert P_{\text{STRONG}}\rvert$ and $\lvert P_{\text{SHORT}}\rvert$ denotes the length of the strongest or shortest path, respectively.

users and 8.85 million directed edges. Retaining only the giant connected component reduces this to 2.25 million unique users and 8.75 million directed edges. We induce influence weights over the edges as follows: if Retweets(B, A) is the number of times B retweeted A, then:

$$\text{Influence}(A, B) = \frac{\text{Retweets}(B, A)}{\sum_X \text{Retweets}(B, X)}$$

That is, influence is the proportion of interactions that a node directs to another to all of her outgoing interactions, which is congruent with the definition described in Section 2. Because the Twitter dataset is directed to begin with, there exist some nodes that never retweet: that is, they have no outgoing edges in the original graph. This means that not all nodes contribute to the total amount of influence in the graph. The resulting average influence per node is 0.6, with a high variance of 138.3. The maximum node influence is 8414.9. A histogram of the log node influence distribution is shown in Figure 3.

Note that the directed property of this graph is important: we cannot make the analogous assumption that we did for DBLP, namely that a large number of tweets between two users correlates with a strong tie strength. Interaction in Twitter is often asymmetric. However, we do note that it seems reasonable to assume that a large number of directed tweets from A to B indicates that B has high influence over A.

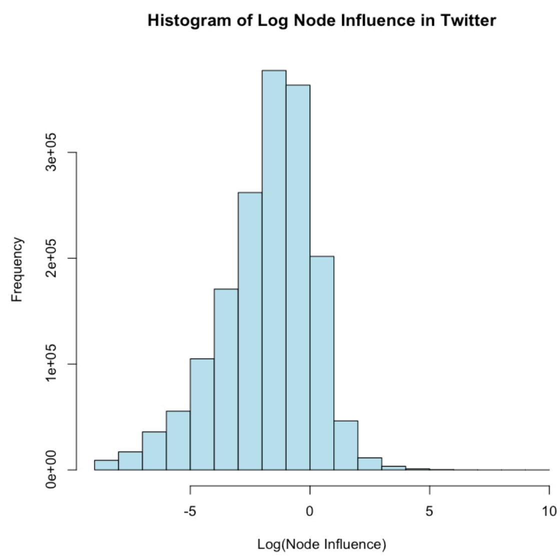


Figure 3. As with the DBLP dataset, most nodes in the Twitter dataset have a total influence < 1. However, note that a very small number of nodes have an extremely high influence (10, on a natural log scale). The two nodes with highest influence are “revrunwisdom”, a religious leader who tweets religious and spiritual quotations, and “tweetmeme”, which aggregates popular links on twitter.

5.2.1 Twitter Results

We conduct our experiment on the Twitter data identically to that on the DBLP data. That is, we compute the probability of success along 2 different paths:
a) P_{SHORT} the shortest path based on the number of hops.
b) P_{STRONG} the strongest path.

Again, we conducted our experiment on 500 randomly selected node pairs in the dataset. We summarize our results in Table 2.

6. Discussion

The hypothesis that utilizing edge weights may well improve global social search is reflected nicely in our results. We discuss this here and – noting that these results are still extremely preliminary – also discuss some limitations of our approach, suggesting improvements for future work.

67.8% of the best paths in the Twitter graph, and 43% of the best paths in the DBLP graph, are *longer* than the corresponding *shortest* paths. In the Twitter dataset these best, but longer, paths contain just over 2 hops more than the corresponding shortest paths, on average. In the DBLP dataset these best, but longer, paths contain, on average, just over 1 hop more than the shortest possible path.

For the remaining 32.2% and 57.0% of the paths in the Twitter and DBLP networks respectively, the strongest and shortest paths have the same length. As we choose the shortest path randomly from the set of all possible

	All Pairs	$\|P_{\text{STRONG}}\| > \|P_{\text{SHORT}}\|$	$\|P_{\text{SHORT}}\| = \|P_{\text{STRONG}}\|$
Retweets	500	339 (67.8%)	161 (32.2%)
Avg. $\|P_{\text{SHORT}}\|$	7.7	7.9	7.3
Avg. $\|P_{\text{STRONG}}\|$	9.2	10.1	7.3
Avg. improvement in influence ($\frac{S(P_{\text{STRONG}})}{S(P_{\text{SHORT}})}$)	35,081	30,360	45,021

Table 2. Summary statistics for experiment results conducted on the Twitter dataset. $S(P_{\text{STRONG}})$ and $S(P_{\text{SHORT}})$ denote the strength of the strongest or shortest path, respectively. $|P_{\text{STRONG}}|$ and $|P_{\text{SHORT}}|$ denotes the length of the strongest or shortest path, respectively.

shortest paths, this does not mean that they are the same path. Indeed, in the DBLP network, the strongest path is, on average, 506 times more influential than the shortest path. The corresponding statistic in the Twitter dataset is 45,201 times as influential.

Our results suggest that utilizing weights in global social search may improve search results. In the case in which the strongest and shortest paths are of equal length, there can be only gain in picking the more influential path. In the case in which the strongest path is longer than the shortest path, 1-2 extra hops seems a small price to pay for a significant improvement in path influence. Further, neither improvement can be made without considering edge weights. However, measuring the scale of the improvement proves tricky, as we discuss in the next paragraph.

We posit that the large discrepancy between the Twitter and the DBLP datasets is due to their fundamentally different structure. The property of influence seems intrinsic to a network such as Twitter, where interactions are driven by hype and popularity. In a co-authorship network, however, influence is a consequence of contribution, and so is more evenly distributed amongst nodes. This is well illustrated by the two histograms in Figures 2 and 3.

Our results also provide fodder for future model and experimental design improvements. We have run our experiments on two graphs, albeit very different ones. Running our experiments on a wider range of datasets would give us a broader understanding of the problem and solution spaces of influence weights.

A final caveat, often noted in social network analysis, is that tie strength in any one dataset is not representative of tie strength in real life, which can be interpreted in a variety of subjective ways. For example, one may direct tweets to colleagues at work much more often than to one's best friend back home. We believe, however, that proxies for tie strengths provide useful information that unweighted networks miss, and tie strengths often reflect the truth in a particular "universe" (e.g. the "Twitter universe", or the "DBLP universe"), where interactions do indicate better social paths for interactions within that universe.

7. Related Work

Much work has focused on the problem of search in social networks, and especially on the problem of *local* social search (that is, the search is conducted by nodes in the network, and they do not have a global view of the network). In a social search experiment, Dodds et al. asked people to forward a message through acquaintances to target persons they did not know [2]. They found that successful social searches did not require hubs as crucial relay points, but did rely heavily on professional ties; ties tended to be medium to weak in strength. Adamic et al. simulate similar "small world" experiments using email data and online social networks [1]. Both of these research branches are based on prior work by Watts et al. [13], which argues for social hierarchies as a framework for modeling social search. A common theme in all of this research is the notion of enriching tie strength with information such as geographical proximity, homophily etc. We are not the first to express frustration with the use of binary ties for social search.

In terms of global social search, Aardvark[2], a service that connects people with specific questions to the people most qualified, or most likely, to have an answer, internally employs a symmetric measure of *affinity* between users [6]. The affinity between users is calculated using a weighted cosine similarity over a number of features, including: vocabulary match, profile similarity, and social connectedness in real life. Aardvark's success is testimony to the efficacy of routing social requests using edge weights in global social search. Although our proposed method of weighted edges is much coarser, it has the advantage of being applicable to a wider array of datasets: those which have communication frequencies between node elements. Similarly, Facebook employs an internal measure of edge weight between friends, which is used to weight items when generating users' news feeds.

To date, most network metrics and methods (such as diameter, clustering algorithms etc.) assume binary edges. However, it seems that many of these could easily be adapted for use with weighted, directed edges. The problem remains to infer edge weights in social graphs. Communi-

[2] www.vark.com

cation frequency data, when available, is often used in the form of a threshold (e.g. define an edge between A and B if they have exchanged at least 5 messages) [7, 12].

However, some recent work has focused on inferring edge annotations based on structural and metadata properties of original datasets.

A step up from binary networks, in *signed* networks, edges may be either positive, negative or not present. Leskovec et al. have researched properties of signed networks as well as methods of predicting signs for online social networks [8, 9].

Some recent work has focused on inducing relationship tie strength from social network metadata. Gilbert and Karahalios present a successful method for predicting closeness from Facebook profile attributes, such as wall posts, messages exchanged etc. [3] Xiang et al. present a model for learning relationship strength based on communication activity and profile similarities in online networks such as Facebook and LinkedIn [14].

8. Conclusion

In this paper, we posited that utilizing edge weights in the problem of global social search could yield more effective results than a search based on binary ties alone. Not only do binary ties fail to capture relationship strength, but they also do not model relationship asymmetry. We presented *influence* as a generalized, asymmetric measure of tie strength between two nodes, defining influence of person A on person B as the proportional investment that B makes in A, and defined a method of calculating the most influential path from a source to a target node. We conducted an experiment designed to measure the relationship between path strength and path length on two datasets: the DBLP paper co-authorship network, and one month's worth of Twitter retweets (RTs). The results of our experiment showed that in many cases, the most influential path between two nodes is, on average, 1-2 hops longer than the shortest path between those nodes. Moreover, for cases in which the most influential path has the same length as the shortest path, choosing the most influential path is still more effective. We conclude that incorporating edge weights into global social search algorithms would be beneficial to online social networking sites.

Acknowledgments

We thank Jure Leskovec for discussions on this topic.

References

[1] L. Adamic and E. Adar. How to search a social network. *Social Networks*, 27(3):187–203, 2005.

[2] P. Dodds, R. Muhamad, and D. Watts. An experimental study of search in global social networks. *Science*, 301(5634):827, 2003.

[3] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 211–220, New York, NY, USA, 2009. ACM.

[4] M. Granovetter. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[5] M. S. Granovetter. *Getting a job*. Harvard University Press, 1974.

[6] D. Horowitz and S. Kamvar. The Anatomy of a Large-Scale Social Search Engine, 2010.

[7] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 915–924, New York, NY, USA, 2008. ACM.

[8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting Positive and Negative Links in Online Social Networks. *Arxiv preprint arXiv:1003.2429*, 2010.

[9] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. *Proc. 28th CHI*, 2010.

[10] M. E. J. Newman. Fast algorithm for detecting community structure in networks, September 2003.

[11] Ö. Şimşek and D. Jensen. Navigating networks by using homophily and degree. *Proceedings of the National Academy of Sciences*, 105(35):12758, 2008.

[12] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman. *Email as Spectroscopy: Automated Discovery of Community Structure within Organizations*, pages 81–96. Kluwer, B.V., Deventer, The Netherlands, The Netherlands, 2003.

[13] D. Watts, P. Dodds, and M. Newman. Identity and search in social networks. *Science*, 296(5571):1302, 2002.

[14] R. Xiang, J. Neville, and M. Rogati. Modeling Relationship Strength in Online Social Networks. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010*, 2010.