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

When searching for a person in an online social network, most contemporary networks return source-target paths, ranked only by degrees of separation. Not only does this fail to reflect social tie strength (how well two people know one another), it also does not model asymmetry in social relationships (i.e. just because one person pays attention to another, it does not mean the latter will reciprocate).

We propose 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. Formally, the influence a person A has over person B is defined as the proportion of B's investments B makes on A.

We study two large real-world networks, DBLP (a computer science bibliography network) and a network formed by one month of Twitter retweet data. Our experiments show that for these social networks, the best paths according to our influence metric are not necessarily the shortest paths: a longer path is better in 68% of searches in Twitter and 45% of searches in DBLP. Furthermore, even when the best and shortest path lengths are equal, we find that the best path is often better than a random shortest path of the same length by a significant margin.

1. INTRODUCTION

The popularity of online social networks has made them an important resource for social searches, in which the goal is to find a person, or a "chain" of people, who might pass on a recommendation for a particular job, or introduction, to a specific person. For example: 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 a network like Facebook to see whether they have any friends in common who could arrange an introduction. Contemporary social networks, such as Facebook, LinkedIn and others, commonly model social relationships as binary: two people are either "friends" or they are not. As such, the social search algorithms in these networks tend to return the shortest path between source and the target. This paper explores the hypothesis that social searches can be made more effective by taking into consideration the *influence* a person has over another, which is inherently asymmetric and has varying strength. We conduct our studies in the context of *global social search*, where we assume that all network information is available to the search algorithm.

The 4th SNA-KDD Workshop '10 (SNA-KDD'10), July 25, 2010 , Washington D.C. USA. Copyright 2010 ACM 978-1-4503-0225-8...$10.00.

1.1 Strength of Ties

In real life, people maintain a large number of relationships with *varying* tie strength: close friends, family, work colleagues, casual acquaintances, and so on. 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 an important benefit of online social networks is precisely their capacity to capture ties of varying types and strengths.

Degrees of separation are all we need when we ask how two people are connected out of curiosity; Erdos number or Bacon number are two well known examples of such queries. In social searches where we wish to find the best way to get to some target person, finding the shortest path is often insufficient.

This problem 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 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: $P_1 = \langle A, C, D, B \rangle$ and $P_2 = \langle A, E, B \rangle$. If A and B are virtually strangers to E, but C and D are close friends and are relatives of A and B, respectively, then P_1 is more likely to yield an introduction of A to B than P_2. Even when paths are of the same length, some may be preferable to others;

however the search returns, in several cases, what appears to be an unranked list of hundreds of paths. The problem of selecting the right path is even worse when conducting a search not for a specific person, but for a generic attribute such as "what is the best route in my social network to anyone employed at Twitter ?" This problem is easily extrapolated to our examples given above and other environments. One example is matchmaking sites that attempt to pair individuals based on an underlying social graph; such a site could use models for relationship weights between individuals to derive the best way to introduce users to each other.

1.2 Asymmetric Influence in Social Networks

In conducting a social search we have to consider, in addition to tie strength, relationship asymmetry. Let us illustrate asymmetry of relationships using a scenario in Twitter. 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. In addition, if Joe does not retweet anybody else, than Obama has 100% influence over Joe; whereas if Joe has a habit of retweeting everybody, then the influence Obama has on Joe is proportionally smaller.

Given a social graph, we model the influence A has over B as the fraction of B's actions due to A. The influence of a path in the social graph is correspondingly 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. In global social search, therefore, it would be most effective to route requests through the most influential path.

1.3 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 being 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 ranking paths of the same length.

The rest of this paper proceeds 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, providing a broad discussion in Section 5.3. We compare with prior work on social search and inducing edge weights in Section 6 and conclude in Section 7.

2. INFLUENCE

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 reasonable influence over the succeeding node.

2.1 Social Interactions

Social networks model several types of social interactions, from daily communications (e-mail) to co-authorship (DBLP) to professional acquaintances (LinkedIn). Moreover, interactions may be directed (Twitter, for example, in which someone being followed may have no knowledge of the follower) or undirected (co-authorship).

We model *influence* based on social interactions that require some kind of cost, or *investment* on behalf of the people involved. As interaction involves an investment of time and effort from participants, the number of interactions is an informative measure of tie strength. The key intuition behind this proposition is the reciprocity of investment and influence: if A invests time in B, then B must have some influence over A. In addition to its intuitive appeal, interaction-based influence data already exists in almost all online social networks. Examples of interaction data are Facebook Wall posts, and email messages exchanged between two people.

We note that there are several other mechanisms of deriving an influence metric for social network edges. A simple heuristic that can be used when interaction data is not available is to count the number of mutual friends that A and B have. More complicated methods estimating tie strength may be based on profile similarity, or detailed comparisons of a variety of social interactions [3, 13]. Similarly, in trust networks, influence can possibly be derived from trust metrics (if A trusts B, then B has influence over A.)

2.2 Asymmetry of Influence

Influence is often asymmetric: A has high influence on B does not mean the reverse is true. Asymmetry may also be present when interactions are undirected. Consider, for example, a co-authorship network. Because advisors are frequently co-authors on publications, the proportion of the student's publications that are shared with her advisor are high; the advisor has high influence on the student. However, the professor has many students, and therefore the proportion of the professor's publications that are shared with the student is low; this reflects the reality that the student has comparatively lower influence on her advisor. Thus, even though the interaction in this case is undirected (or bidirectional), there is still an asymmetry in influence between two collaborators.

2.3 Influential Ties

As mentioned above, 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 her time in person A, then A has high influence over B.

We now develop a quantitative definition of influence. We define 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 the result of 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 investment. We have less control, however, over who invest in us. As in real life, influence is a quality that must be given by others.

The influence of an edge in a social graph always lies 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.

2.4 Influential Individuals

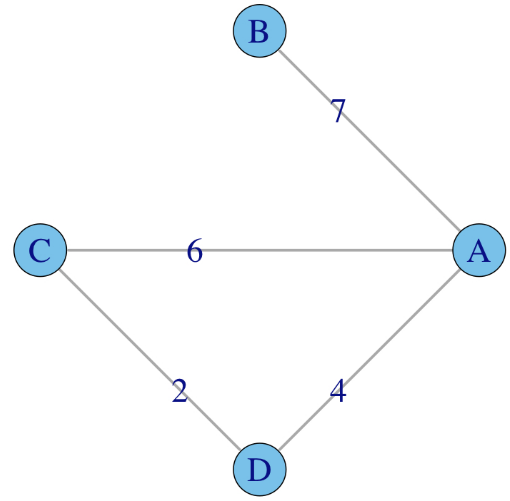
A person is *influential* if he or she has high influence on many people. We define the *influence of a node* as the sum of the influences the node has on others. That is, the influence of a node A,

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

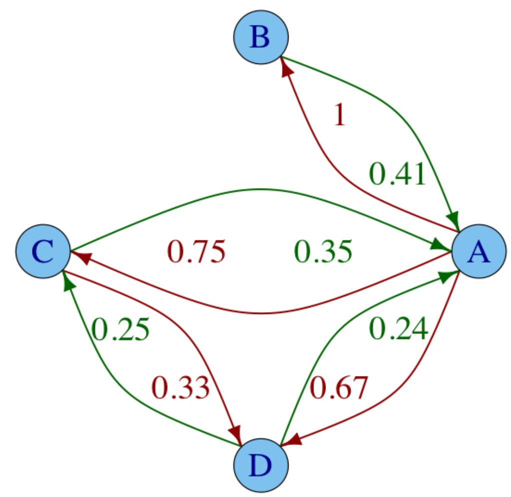
As reported in Section 5, we use this definition to plot the distribution of influence across different users in the network.

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. For a message to be routed along the chain of nodes, it is required for each of the intermediate nodes to participate. In addition, we also associate some penalty with path length, by assigning a constant discount factor to each link in the network. This factor dampens probabilities that a message will be passed along an edge. Similar to the game of "broken telephone", the longer the path, the more likely a message will be dropped. We define the influence of a path to be the product of the influence of each edge on the path, adjusted for the discount factor. We refer to the final influence as the *strength* of the path and calculate it as follows: for a path P of length $|P|$ that contains edges



(a) Before inferring influence



(b) After inferring influence

Figure 1: This figure illustrates the results of allocating influence to the edges in a network with undirected interactions, such as DBLP. 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 her total publications involve student C.

edges $e_1, e_2, \ldots, e_n$, and a discount factor x, the strength is:

$$\mathrm{S}(P) = \prod D \times \mathrm{Influence}(e_i), e_i \in P.$$

Our goal is to find the path that maximizes this probability. In our experiments, we set the discount factor D to 0.95.

As with edge influence, we note that there are several other potential definitions of path strength; we have merely picked one that is plausible and easy to model. One possibility 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 with types of relationship and then factor influence based on the type of query (e.g. prefer to propagate a professional inquiry through current and former co-workers.) We do not deal with incremental decay or edge labels in this work, but note that they are natural extensions of our definition of path strength above, which we chose for its simplicity and intuitive appeal.

4. ALGORITHM: COMPUTING THE STRONGEST PATH

We use Djikstra's algorithm to compute the shortest path between two nodes. To compute the *strongest* path between two nodes, we make a natural adaptation to this algorithm. Given specific source and target nodes and discount factor D, we would like to find a path P from the source to the target that maximizes:

$$\prod D \times \mathrm{Influence}(e_i), e_i \in P$$

Therefore we would like to maximize the logarithm of this metric:

$$\sum log(D) + log(\mathrm{Influence}(e_i)), e_i \in P$$

and therefore to minimize

$$-\sum log(D) + log(\mathrm{Influence}(e_i)), e_i \in P$$

which leads us to minimize

$$\sum log(1/D) + log(1/\mathrm{Influence}(e_i)), e_i \in P$$

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

5. EXPERIMENTS

We evaluate the benefit of incorporating influence weights into global social search on two large networks: DBLP and Twitter retweets. Our goal is to evaluate whether considering edge weights results in *better* paths than simply picking one of the shortest paths. In each network, influence may be inferred from interactions between individuals, as described above. Both datasets are large, contain realistic social data and provide a global view of the network, which is needed for our quantitative evaluation of weighted paths. We deliberately chose these two networks for their representative diversity: the DBLP data forms a dense network in which undirected ties are typically precipitated from intense, real-world social interaction. The Twitter dataset, on the other hand, is directed and more sparse; furthermore, ties do not necessarily represent real world social interaction[1]. Our expectation was that the use of such different networks in our experiments would lead to richer feedback on our model assumptions and structural properties of our influence metric.

5.1 DBLP Computer Science Bibliography

The DBLP dataset[2] 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 $\mathrm{Papers}(v_i, v_j)$ is the number of papers co-authored by v_i and v_j, then

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

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

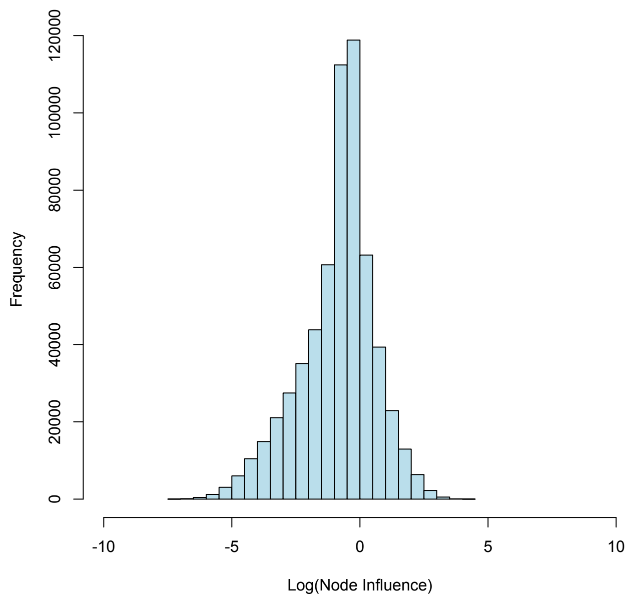
In using the DBLP dataset, we make the underlying assumption that a large number of co-authorships between two actors indicates that they have a strong tie. We note that "weak" ties on DBLP are probably not all that weak: writing a paper is a significant time investment after all! Of course, a low weight on an influence edge does not necessarily indicate weak *social* influence between the two actors in an absolute sense; for example, two colleagues in non-overlapping areas may know and respect each other without having co-authored papers together. However, for experimental purposes we restrict ourselves to the "DBLP world" and base influence edge weights on co-authorship counts.

To evaluate the effectiveness of using influence in global searches, we compute for each of the 500 randomly selected source-destination pairs

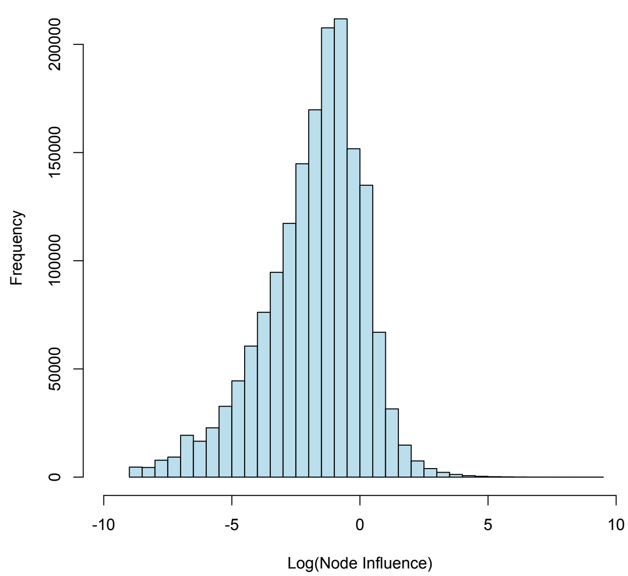
- $\mathrm{S}(P_{\mathrm{SHORT}})$, the strength of the shortest path based on the number of hops. Since there may be many paths with the same shortest length, we pick a random path from amongst all such paths, and

[1]In that people linked by a tie need not even know one another.

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



(a) DBLP



(b) Twitter

Figure 2: Log node influence in DBLP and Twitter

- S(P_{STRONG}), the strength of the strongest path. Strongest path is computed using the weights model described in Section 4.

Table 1 compares the paths computed and their respective strengths. The results are discussed in Section 5.3.

5.2 Twitter Retweets

Our Twitter dataset consists of 1 month's worth of tweets from Twitter. Considering only retweets yields a directed graph, comprising approximately 2.4 million unique users and 8.85 million directed edges. Retaining only the giant component reduces this to 2.25 million unique users and 8.75 million directed edges. We assign 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 (as described in Section 2). Since some people never retweet, they have no outgoing edges, and nobody has influence over them in this model. Using this model, a histogram of the log node influence distribution is shown in Figure 2(b). The average influence a person has in total is 0.6, with a high variance of 138.3. The maximum node influence is 8414.9. Table 2 shows the results of a similar experiment on Twitter retweets comparing the shortest and strongest paths of 500 randomly selected source and target pairs.

5.3 Discussion

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

From Figure 2, we see that most nodes in both datasets have influence below one, which is congruent with the observation that in real life influence is asymmetric. So a few nodes are highly influential, and there are many nodes that are influenced by others more than they influence anybody. We see that a handful of nodes in DBLP are highly influential, but not nearly as influential as those in the tails of the Twitter influence distribution. Indeed, some nodes in the Twitter dataset have an extremely high influence (8414.9, or about 10 on a natural log scale). The two nodes with highest influence in this dataset are "revrunwisdom", a religious leader who tweets religious and spiritual quotations, and "tweetmeme", which aggregates popular links on twitter.

67.8% of the *strongest* 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 two hops more than the corresponding shortest paths, on average. In the DBLP

	All	$\lvert P_{\text{STRONG}} \rvert > \lvert P_{\text{SHORT}} \rvert$	$\lvert P_{\text{SHORT}} \rvert = \lvert P_{\text{STRONG}} \rvert$
Node Pairs	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 ($\frac{S(P_{\text{STRONG}})}{S(P_{\text{SHORT}})}$)	548	605	506

Table 1: Summary statistics for experiment results conducted on the DBLP dataset. S(P_{STRONG}) and S(P_{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.

	All	$\lvert P_{\text{STRONG}} \rvert > \lvert P_{\text{SHORT}} \rvert$	$\lvert P_{\text{SHORT}} \rvert = \lvert P_{\text{STRONG}} \rvert$
Node Pairs	500	339 (67.8%)	161 (32.2%)
Avg. $\lvert P_{\text{SHORT}} \rvert$	7.7	7.9	7.3
Avg. $\lvert P_{\text{STRONG}} \rvert$	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_{STRONG}) and S(P_{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.

dataset these best, but longer, paths contain, on average, just over one hop more than the shortest possible path.

For the remaining 32.2% and 57% 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 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 a random shortest path. The corresponding statistic in the Twitter dataset is 45,201 times as influential. The precise factors of improvement are not so important; the key takeaway here is that weights can be used to select significantly better paths in the network. In the case in which the strongest path is longer than the shortest path, one or two extra hops seems a small price to pay for an improvement in path influence. And when there are many paths of the same length, the influence metric can be used to rank them and present them in an order that is presumably more relevant to the user than an unranked list.

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.

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 may not be representative of tie strength in real life; which itself 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 "world" (e.g. the "Twitter world", or the "DBLP world"), where interactions can indicate better social paths for interactions within that world.

6. 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. [12], 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[3], a service that connects people with specific questions to the people most

[3]www.vark.com

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. Similarly, Facebook is reported to use an affinity metric called EdgeRank to generate news feeds[4]; however, it is not public how the affinity metric is computed.

There have been several efforts to infer edge weights in social graphs. A common and simple way of converting communication frequency data into a binary edge weight is to use a threshold (e.g. define an edge between A and B if they have exchanged at least 5 messages) [8, 11]. Realistic weighted network data has been obtained in the context of mobile call graphs [10], though such detailed data is generally hard to come by. A step up from binary networks, in *signed* networks, edges may be either positive, negative or not present [9]. Other recent work has focused on inducing relationship tie strength from social network metadata. Gilbert and Karahalios present a 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 [13].

Outside the domain of search, there has been work on diffusion models in networks which employ measures of internode influence. Some of this work employs influence models similar to ours, but uses it to study how viral effects propagate over time [7]. It seems that many other network analysis techniques and studies could effectively be adapted for use with weighted, directed edges.

We find that generalized global social search has received relative little research consideration in the past. However, the inherent problems with binary ties has sparked healthy research interest in estimating tie strength; a natural next step is to develop social network analysis methods that incorporate these into social search and other problems.

7. 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 an 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. The results of our experiment showed that in many cases, the most influential path between two nodes is, on average, one or two 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, the strongest path may be much more effective than a random path. We therefore conclude that incorporating edge weights into global social search algorithms can be greatly beneficial to online social networks.

Acknowledgments

We thank Jure Leskovec for discussions on this topic. This research is supported in part by the NSF POMI (Programmable Open Mobile Internet) 2020 Expedition Grant 0832820 and a Stanford graduate student scholarship.

8. 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 *Proc. 27th CHI*. ACM, 2009.

[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. D. Kamvar. The anatomy of a large-scale social search engine. In *Proc. 19th WWW*. ACM, 2010.

[7] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. 9th KDD*. ACM, 2003.

[8] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *Proc. 17th WWW*. ACM, 2008.

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

[10] J. P. Onnela, A. L. Barabási, J. Hyvőnen, K. Kaski, J. Kertész, D. Lazer, J. Saramáki, and G. Szabó. Analysis of a large-scale weighted network of one-to-one human communication. In *New Journal of Physics 9, 179*, 2007.

[11] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman. *Email as Spectroscopy: Automated Discovery of Community Structure within Organizations*. Kluwer, B.V., 2003.

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

[13] R. Xiang, J. Neville, and M. Rogati. Modeling Relationship Strength in Online Social Networks. In *Proc. 19th WWW*. ACM, 2010.

[4]See Facebook F8 2010, Focus on Feed session.