Measuring Tie Strength in Implicit Social Networks*

Mangesh Gupte
Department of Computer Science
Rutgers University
Piscataway, NJ 08854
mangesh@cs.rutgers.edu

Tina Eliassi-Rad
Department of Computer Science
Rutgers University
Piscataway, NJ 08854
tina@eliassi.org

ABSTRACT
Given a set of people and a set of events attended by them, we address the problem of measuring *connectedness* or *tie strength* between each pair of persons. The underlying assumption is that attendance at mutual events gives an implicit social network between people. We take an axiomatic approach to this problem. Starting from a list of axioms, which a measure of tie strength must satisfy, we characterize functions that satisfy all the axioms. We then show that there is a range of tie-strength measures that satisfy this characterization.

A measure of tie strength induces a ranking on the edges of the social network (and on the set of neighbors for every person). We show that for applications where the ranking, and not the absolute value of the tie strength, is the important thing about the measure, the axioms are equivalent to a natural partial order. To settle on a particular measure, we must make a non-obvious decision about extending this partial order to a total order. This decision is best left to particular applications. We also classify existing tie-strength measures according to the axioms that they satisfy; and observe that none of the "self-referential" tie-strength measures satisfy the axioms. In our experiments, we demonstrate the efficacy of our approach; show the completeness and soundness of our axioms, and present Kendall Tau Rank Correlation between various tie-strength measures.

Author Keywords
Social networks, tie strength, axiomatic approach

ACM Classification Keywords
J.4 Computer Applications: Social and Behavioral Sciences [Sociology]

General Terms
Theory, Measurement, Experimentation

*This work was supported under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344. The current affiliation of the first author is Google.

WebSci 2012, June 22–24, 2012, Evanston, Illinois, USA.

INTRODUCTION
Explicitly declared friendship links suffer from a low signal to noise ratio (e.g. Facebook friends or LinkedIn contacts). Links are added for a variety of reasons like reciprocation, peer-pressure, etc. Detecting which of these links are important is a challenge.

Social structures are implied by various interactions between users of a network. We look at event information, where users participate in mutual events. Our goal is to infer the strength of ties between various users given this event information – i.e. the implicit social networks.

There has been a surge of interest in implicit social networks. We can see anecdotal evidence for this in startups like COLOR (`http://www.color.com`) and new features in products like Gmail. COLOR builds an implicit social network based on people's proximity information while taking photos.[1] Gmail's *don't forget bob* feature uses an implicit social network to suggest new people to add to an email given a existing list [16].

Consider people attending different events with each other. We define an *event* by the set of people that attend it. An event can represent the set of people who took a photo at the same place and time, like COLOR, or a set of people who are on an email, like in Gmail. Given the set of events, we would like to infer how *connected* two people are – i.e. we would like to measure the *strength of the tie* between people. All that is known about each event is the list of people who attended it. People attend events based on an implicit social network (with ties between pairs of people). We want to solve the inference problem of finding the weighted social network that gives rise to the set of observed events.

Given a bipartite (a.k.a. two-mode) graph, with people as one set of vertices and events as the other set, we want to infer the tie-strength between the set of people. Hence, in our problem, we do not even have access to any directly declared (i.e., explicit) social network between people. We want to infer the *implicit* social network based on the set of people who interact together at different events.

We start with a set of axioms and find a characterization of functions that could serve as a measure of tie strength, just given the event information. We do not end up with a single function that works best under all circumstances. In fact, we show that there are non-obvious decisions that need to be made to settle down on a single measure of tie strength.

[1] `http://mashable.com/2011/03/24/color/`

Moreover, we examine the case where the absolute value of the tie strength is not important, but its order is important (see Section "Tie Strength and Orderings"). We show that in this case the axioms are equivalent to a natural partial order on the strength of ties. We also show that choosing a particular tie strength function is equivalent to choosing a particular linear extension of this partial order (which yields a total order).

Our contributions are:

- *We present an axiomatic approach to the problem of inferring tie strength in implicit social networks.*
- *We characterize functions that satisfy all the axioms and show a range of measures that satisfy this characterization.*
- *We show that in ranking applications, the axioms are equivalent to a natural partial order. We demonstrate that to settle on a particular measure, we must make non-obvious decisions about extending this partial order to a total order, which is best left to the particular application.*
- *We classify measures found in prior literature according to the axioms that they satisfy.*
- *In our experiments, we illustrate the efficacy of our approach by demonstrating that event information produces useful predictions of tie-strength. We measure the soundness and completeness of our axioms. We also present the Kendall's Tau Rank Correlation between various tie-strength measures.*

The remainder of this paper is structured as follows: Related Work, Model, Axioms of Tie Strength, Measures of Tie Strength, Experiments, and Conclusions.

RELATED WORK

We split the related works into different subsections that emphasize particular methods/applications.

Strength of Ties: Granovetter [8] introduces the notion of *strength of ties* in social networks. He shows that weak ties are important for various aspects like spread of information in social networks. There have been various studies on identifying the strength of ties given different features of a graph. Gilbert & Karahalios [7] model tie strength as a linear combination of node attributes (e.g., intensity and intimacy) to classify ties in a social network as strong or weak. Weights on attributes enable them to find attributes that are most useful in making tie-strength predictions. Kahanda & Neville [11] take a supervised learning approach to the problem by constructing a predictor that determines whether a link in a social network is a strong tie or a weak tie. They report that *network transactional features*, which combine network structure with transactional features (such as the number of wall posting) form the best predictors.

Link Prediction: Adamic & Adar [1] consider the problem of predicting links between Web-pages of individuals, using information such as membership of mailing lists and use of common phrases on Web pages. They define a measure of similarity between users by creating a bipartite graph of users on the left and features (e.g., phrases and mailing-lists) on the right as $w(u, v) = \sum_{(i \text{ neighbor of } u\&v)} \frac{1}{\log |i|}$. Liben-Nowell & Kleinberg [14] formalize the problem of predicting which new interactions will occur in a social network given a snapshot of the current state of the network. They use many existing predictors of similarity between nodes (such as Adamic-Adar [1], SimRank[9], and Katz [13]); and generate a ranking of node-pairs, which were not connected by an edge. To measure the efficacy of various measures, they compare across different datasets. Their main finding is that there is enough information in the network structure that all the predictors handily beat the random predictor, but not enough that the absolute number of predictions is high. Allali et al. [2] address the problem of predicting links in a bipartite network. They define *internal links* as links between left nodes that have a right node in common (i.e. nodes that are at a distance of two from each other); then offer predictions that are only for internal links.

Email networks: Because of the ubiquitous nature of email, there has been a lot of work on various aspects of email networks. Roth et al. [16] discuss a way to suggest more recipients for an email given the sender and the current set of recipients. This feature has been integrated in the Google's popular Gmail service.

Axiomatic approach to Similarity: Altman & Tennenholtz [3] were one of the first to axiomatize graph measures. In particular, they studied axiomatizing PageRank. More recently, Jin et al. [10] present an axiomatic ranking of network role similarity. The closest in spirit to our work is the work by Lin [15] that defines an information-theoretic measure of similarity. This measure depends on the existence of a probability distribution on the features that define objects. While the measure of tie strength between people is similar to a measure of similarity, there are important differences. We do not have any probability distribution over events, just a log of the ones that occurred. More importantly, Lin [15] defines items by the attributes or features they have. Hence, items with the same features are identical. In our case, even if two people attend all the same events, they are not the same person, and in fact they might not even have very high tie strength depending on how large the events were.

MODEL

We model people and events as nodes and use a bipartite graph $G = (L \cup R, E)$ where the edges represent membership. The left vertices L correspond to people while the right vertices R correspond to events. Except for the set of people who attended the events, we ignore all other features/information about events such as timing, location, and the inherent importance of events. These are features that would be important to the overall goal of measuring tie strength between users, but in this work we focus on the task of inferring tie strength using only the graph structure. Hence, in our model an event is defined only by the set of people that attend it.

We shall denote users in L by small letters $(u, v, \ldots)$ and events in R by capital letters $(P, Q, \ldots)$. There is an edge between u and P if and only if u attended event P. Hence, our problem is to find a function on bipartite graphs that models *tie strength* between people, given this bipartite graph representation of people and events. Figure 1 shows an example of a bipartite person$\times$event graph.

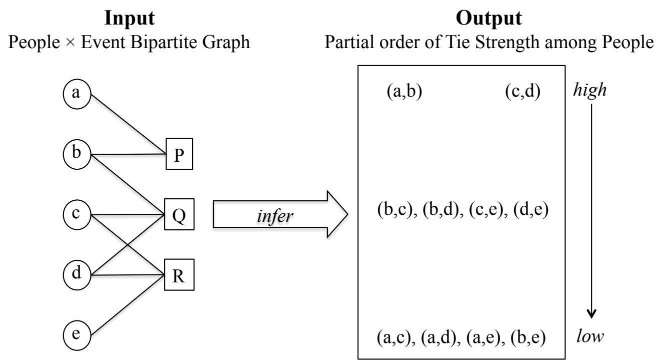

Figure 1. Given a bipartite person$\times$event graph, we want to infer the induced partial order of tie strength among the people.

We also introduce some notation. We shall denote the tie strength of u and v due to a graph G as $TS_G(u, v)$ or as $TS(u, v)$ if G is obvious from context. We shall also use $TS_{\{E_1,...,E_k\}}(u, v)$ to denote the tie strength between u and v in the graph induced by events $\{E_1, \ldots, E_k\}$ and users that attend at least one of these events. For a single event E, then $TS_E(u, v)$ denotes the tie strength between u and v if E where the only event.

We denote the set of natural numbers by $\mathbb{N}$. A sequence of k natural numbers is given by $(a_1, \ldots, a_k)$ and the set of all such sequences is $\mathbb{N}^k$. The set of all finite sequence of natural numbers is represented as $\mathbb{N}^* = \cup_k \mathbb{N}^k$

AXIOMS OF TIE STRENGTH

We now discuss the various intuitive properties that we want any measure of tie strength between two users u and v to satisfy. We formalize these intuitions as axioms that a measure must follow.

Intuition for Axiom 1: We focus our attention on the problem of inferring tie strength between every pair of vertices while actively ignoring information about what the events themselves might signify. Therefore, tie strength should not depend on the particular labels we give to events or people.

Axiom 1 (Isomorphism)**.** *Suppose we have two graphs G and H and a mapping of vertices such that G and H are isomorphic. Let vertex u of G map to vertex a of H and vertex v to b. Then $TS_G(u, v) = TS_H(a, b)$. Hence, the tie strength between u and v does not depend on the labels of u and v, only on the link structure.*

Intuition for Axiom 2: Since $TS(u, v)$ produces a real number, we would like to normalize the measure. There are many different ways to normalize the measure that would be mathematically consistent. We select one that is intuitively obvious. Our next axiom comes from the intuition that tie strength between two people who have never been at an event together is 0. Similarly, we choose as unit strength as the strength of a tie between two people who have only been at a single event with each other.

Axiom 2 (Baseline)**.** *If there are no events, i.e. the graph is empty, then the tie strength between every pair u and v is 0. $TS_\phi(u, v) = 0$. If there are only two people u and v and a single event which they attend, then their tie strength is 1. $TS_{\{u,v\}}(u, v) = 1$.*

Intuition for Axiom 3: We would like to reason about the effect of attending multiple events together on tie strength. Social psychology literature [17, 4] gives us the *familiarity effect*, which states that as individuals become more familiar with each other, they develop more positive feelings towards each other.

Axiom 3 (Frequency: More events create stronger ties)**.** *All other things being equal, the more events common to u and v, the stronger the tie strength between u and v. Formally, consider a graph $G = (L \cup R, E)$ and two vertices $u, v \in L$. Also, consider the graph $G' = (L \cup (R \cup P), E \cup P_{u,v,\ldots})$, where $P_{u,v,\ldots}$ is a new event which both u and v attend. Then, $TS_{G'}(u, v) \geq TS_G(u, v)$.*

Intuition for Axiom 4: The number of people at an event clearly affects the strength of ties between individuals. The intuition here is that smaller events are more intimate and hence create stronger individual ties.

Axiom 4 (Intimacy: Smaller events create stronger ties)**.** *All other things being equal, the fewer invitees there are to any particular event attended by u and v, the stronger the tie strength between u and v.*
Formally, consider a graph $G = (L \cup R, E)$ such that $P \in R$ and $(u, P), (v, P), (w, P) \in E$ for some vertex w. Also, consider the graph $G' = ((L \cup R), E - (w, P)))$, where the edge (w, P) is deleted. Then, $TS_G(u, v) \leq TS_{G'}(u, v)$.

Intuition for Axiom 5: The previous two axioms dealt with the strength of individual ties with a group. The next one deals with the total tie-strength "capital" generated by an event. The intuition here is that larger events create more "capital" than smaller events.

Axiom 5 (Popularity: Larger events create more ties)**.** *Consider two events P and Q. If the number of people attending P is larger than the number of people attending Q, then the total tie strength created by event P is more than that created by event Q. Formally,*
$|P| \geq |Q| \implies \sum_{u,v \in P} TS_P(u, v) \geq \sum_{u,v \in Q} TS_Q(u, v)$.

Intuition for Axiom 6: The next axiom captures the intuition of independence between individuals. It states that events at which an individual is not present, do not affect the tie strength of that individual with other individuals who were present at those events. While simple to state, this axiom is a very strong axiom and does not allow for indirect effects.

Axiom 6 (Conditional Independence of Vertices)**.** *The tie strength of a vertex u to other vertices does not depend on events that u does not attend; it only depends on events that u attends.*

Intuition for Axiom 7: The next axiom captures the intuition of independence between events. It states that the increase in tie strength due to an event depends only on the the event itself and on the existing tie strength between u and v and *not* on other events. This is another very strong axiom that does not allow for indirect effects.

Axiom 7 (Conditional Independence of Events)**.** *The increase in tie strength between u and v due to an event P does not depend on other events, just on the existing tie strength between u and v. Formally, $TS_{G+P}(u,v) = g(TS_G(u,v), TS_P(u,v))$ for some fixed monotonically increasing function g.*

Intuition for Axiom 8: Our final axiom states that tie-strength follows the law of diminishing returns (where adding an element to a smaller set helps more than adding it to a larger set).

Axiom 8 (Submodularity)**.** *The marginal increase in tie strength of u and v due to an event Q is at most the tie strength between u and v if Q was their only event. Formally, for a graph G, and every event Q, $TS_G(u,v) + TS_Q(u,v) \geq TS_{G+Q}(u,v)$.*

Discussion

These axioms give a measure of tie strength between nodes that is positive but unbounded.[2] Pairs of nodes that have a higher TS value are closer to each other than pairs that have lower TS value.

We get a sense of the axioms by applying them to Figure 1. Axiom 1 (Isomorphism) implies that $TS(b,c) = TS(b,d)$ and $TS(c,e) = TS(d,e)$. Axioms 2 (Baseline), 6 (Conditional Independence of Vertices), and 7 (Conditional Independence of Events) imply that $TS(a,c) = TS(a,d) = TS(a,e) = TS(b,e) = 0$. Axiom 4 (Intimacy) implies that $TS(a,b) \geq TS(d,e)$. Axiom 3 (Frequency) implies that $TS(c,d) \geq TS(d,e)$.

While each of the aforementioned axioms are fairly intuitive, they are hardly trivial. In fact, we shall see that various measures used in prior literature break some of these axioms. On the other hand, it might seem that satisfying all the axioms is a fairly strict condition. However, we shall see that even satisfying all the axioms is not sufficient to uniquely identify a measure of tie strength. That is, the axioms leave considerable space for different measures of tie strength.

One reason the axioms do not define a particular function is that there is an inherent tension between Axiom 3 (Frequency) and Axiom 4 (Intimacy). While both state ways in which tie strength becomes stronger, the axioms do not resolve which one dominates the other or how they interact with each other. This is a non-obvious decision that we feel is best left to the application in question. In Figure 1, we cannot tell by using just Axioms 1 through 8 which of $TS(a,b)$ and $TS(c,d)$ is larger. We discuss this further in Section "Tie Strength and Orderings."

Given that we cannot obtain a single measure of tie-strength that can be used in all applications, the allure of the axiomatic approach is to classify existing and future measures and to understand the exact trade-offs involved while choosing a particular measure within the constraints imposed by the particular application at hand.

Characterizing Tie Strength

In this section, we shall state and prove Theorem 6 that gives a characterization of all functions that satisfy the aforementioned axioms of tie strength. Axioms 1 through 8 do not uniquely define a function. In fact, one of the reasons that tie strength is not uniquely defined by the given axioms is that we do not have any notion for comparing the relative importance of number of events (frequency) versus the exclusivity of events (intimacy). For example, in terms of the partial order, it is not clear the tie strength of u and v when they attend two events with two people attending each event is stronger than or weaker than when u and v attend three events each of which have three people attending them.

We shall use the following definition for deciding how much total tie strength a single event generates, given the size of the event.

Notation 1. *If k people attend a single event, we shall denote the total tie-strength generated as $f(k)$.*

Lemma 2 (Local Neighborhood)**.** *The tie strength of u and v is affected only by the events that both u and v attend.*

Proof. Given a graph G and users u and v in G, let G^{-u} denote the graph obtained by deleting all events in which u does not participate. Similarly, let $G^{-u,v}$ be the graph obtained by deleting all events of G^{-u} in which v does not participate. By 6, tie strength of u only depends on events that u attends. Hence, $TS_G(u,v) = TS_{G^{-u}}(u,v)$. Also, tie strength of v only depends on events that v attends. Hence, $TS_G(u,v) = TS_{G^{-u}}(u,v) = TS_{G^{-u,v}}(u,v)$. This proves our claim. □

Lemma 3. *The tie strength between any two people is always non-negative and is equal to zero if they have never attended an event together.*

Proof. If two people have never attended an event together, then from Lemma 2 their tie strength remains unchanged if we delete all the events not containing either (which in this case is all the events). Then, Axiom 2 (Baseline) tells us that $TS(u,v) = 0$.

Suppose that for a graph $G = (V,E)$, $E = \{E_1, \ldots, E_k\}$. Then, Axiom 3 (Frequency) implies that $TS_G(u,v) \geq TS_{\{G-E_1\}}(u,v) \geq \cdots \geq TS_\phi(u,v) = 0$. Hence, the tie strength is always non-negative. □

Lemma 4. *If there is a single event with k people, the tie strength of each tie is equal to $\frac{f(k)}{\binom{k}{2}}$.*

Proof. By Axiom 1 (Isomorphism), it follows that the tie-strength on each tie is the same. Since the sum of all the ties is equal to $f(k)$, and there are $\binom{k}{2}$ edges, the tie-strength of each edge is equal to $\frac{f(k)}{\binom{k}{2}}$. □

[2] For a fixed graph G with k events, the tie-strength between any two nodes has an upper-bound of $\binom{n}{2}$.

Lemma 5. *The total tie strength created at an event E with k people is a monotone function $f(k)$ that is bounded by $1 \leq f(k) \leq \binom{k}{2}$*

Proof. By Axiom 4 (Intimacy), the tie strength of u and v due to E is less than an event where they are alone. $TS_E(u,v) \leq TS_{(u,v)}(u,v) = 1$, by Axiom 2 (Baseline). Summing up over all ties gives us that $\sum_{u,v} TS_E(u,v) \leq \binom{k}{2}$. Also, by Axiom 5 (Popularity), $f(k) \geq f(i) : \forall i < k$. In particular, $f(k) \geq f(1) = 1$. This proves the result. □

We are now ready to state the main theorem in this section.

Theorem 6. *Given a graph $G = (L \cup R, E)$ and two vertices u, v. We now characterize Axioms 1 through 8 in terms of a function on G. The tie-strength function TS follows Axioms 1 through 8 if and only if the function is of the form*

$$TS_G(u,v) = g(h(|P_1|), h(|P_2|), \ldots, h(|P_k|))$$

where $\{P_i\}_{1 \leq i \leq k}$ are the events common to both u and v, and $|P|$ denotes the number of people attending event P, $h : \mathbb{N} \to \mathbb{R}$ is a monotonically decreasing function bounded by $h(0) = 0, h(1) = 1, 1 \geq h(n) \geq \frac{1}{\binom{n}{2}}, n \geq 2$ and $g : \mathbb{N}^* \to \mathbb{R}$ is a monotonically increasing submodular function.

Proof. Given two users u and v, we use Axioms 1 through 8 to successively change the form of $TS_G(u,v)$ to first prove that its of the form described above. Let $\{P_i\}_{1 \leq i \leq k}$ be all the events common to u and v. Axiom 7 (Conditional Independence of Events) implies that $TS_G(u,v) = g(TS_{P_i}(u,v))_{1 \leq i \leq k}$, where g is a monotonically increasing submodular function. Given an event P, $TS_P(u,v) = h(|P|) = \frac{f(|P|)}{\binom{|P|}{2}}$. By Axiom 4 (Intimacy), h is a monotonically decreasing function. Also, by Lemma 5, f is bounded by $1 \leq f(n) \leq \binom{n}{2}$. Hence, h is bounded by $1 \geq h(n) \geq \frac{1}{\binom{n}{2}}$. This completes the first part of the proof. To prove the converse we need to verify that each axiom is satisfied if the tie-strength function follows the form described in the theorem. This is fairly straightforward to verify and we leave it out for brevity. □

Theorem 6 gives us a way to explore the space of valid functions for representing tie strength and for finding which work for a given application. In Section "Measures of Tie Strength," we shall look at popular measures of tie strength and show that most of them follow Axioms 1 through 8, and hence are of the form described by Theorem 6. We also describe the functions h and g that characterize these common measures of tie strength . While Theorem 6 gives a characterization of functions suitable for describing tie strength, they leave open a wide variety of functions. In particular, it does not give the comfort of having a single function that we could use. We discuss the reasons for this and what we would need to do to settle upon a particular function in the next section.

Tie Strength and Orderings

We begin this section with a definition of an ordering on elements in a set.

Definition 7 (Total Order)**.** *Given a set S and a binary relation $\leq_{\mathcal{O}}$ on S, $\mathcal{O} = (S, \leq_{\mathcal{O}})$ is called a total order if and only if it satisfies the following properties (i Total). for every $u, v \in S$, $u \leq_{\mathcal{O}} v$ or $v \leq_{\mathcal{O}} u$ (ii Anti-Symmetric). $u \leq_{\mathcal{O}} v$ and $v \leq_{\mathcal{O}} u \implies u = v$ (iii Transitive). $u \leq_{\mathcal{O}} v$ and $v \leq_{\mathcal{O}} w \implies u \leq_{\mathcal{O}} w$*

A total order is also called a linear order.

Consider a measure TS that assigns a measure of tie strength to each pair of nodes u, v given the events that all nodes attend in the form of a graph G. Since TS assigns a real number to each edge and the set of reals is totally ordered, TS gives a total order on all the edges upto equivalences. In fact, the function TS actually gives a total ordering of $\mathbb{N}^*$ upto equivalences. In particular, if we fix a vertex u, then TS induces a total order on the set of neighbors of u, given by the increasing values of TS on the corresponding edges.

The Partial Order on $\mathbb{N}^*$

We now review the definition of a partial order on elements of a set.

Definition 8 (Partial Order)**.** *Given a set S and a binary relation $\leq_{\mathcal{P}}$ on S, $\mathcal{P} = (S, \leq_{\mathcal{P}})$ is called a partial order if and only if it satisfies the following properties (i Reflexive). for every $u \in S, u \leq_{\mathcal{P}} u$ (ii Anti-Symmetric). $u \leq_{\mathcal{P}} v$ and $v \leq_{\mathcal{P}} u \implies u = v$ (iii Transitive). $u \leq_{\mathcal{P}} v$ and $v \leq_{\mathcal{P}} w \implies u \leq_{\mathcal{P}} w$*

The set S is called a partially ordered set or a poset.

Note the difference from a total order is that in a partial order not every pair of elements is comparable. We shall now define a natural partial order $\mathcal{N} = (\mathbb{N}^*, \leq_{\mathcal{N}})$ on the set $\mathbb{N}^*$ of all finite sequences of natural numbers. Recall that $\mathbb{N}^* = \cup_k \mathbb{N}^k$. We shall think of this sequence as the number of common events that a pair of users attend.

Definition 9 (Partial order on N^*)**.** *Let $a, b \in \mathbb{N}^*$ where $a = (a_i)_{1 \leq i \leq A}$ and $b = (b_i)_{1 \leq i \leq B}$. We say that $a \geq_{\mathcal{N}} b$ if and only if $A \geq B$ and $a_i \leq b_i : 1 \leq i \leq B$. This gives the partial order $\mathcal{N} = (\mathbb{N}^*, \leq_{\mathcal{N}})$.*

The partial order $\mathcal{N}$ corresponds to the intuition that more events create stronger ties and that smaller events create stronger ties. In fact, we claim that this is exactly the partial order implied by the Axioms 1 through 8. Theorem 11 formalizes this intuition along with giving the proof. What we would really like is a total ordering which would be equivalent to a measure of tie strength. Can we go from the partial ordering given by the Axioms 1 through 8 to a total order on $\mathbb{N}^*$? Theorem 11 also suggests ways in which we can do this.

Partial Orderings and Linear Extensions

In this section, we connect the definitions of partial order and the functions of tie strength that we are studying. First we start with a definition.

Definition 10 (Linear Extension)**.** *$\mathcal{L} = (S, \leq_{\mathcal{L}})$ is called the* linear extension *of a given partial order $\mathcal{P} = (S, \leq_{\mathcal{P}})$ if and only if $\mathcal{L}$ is a total order and $\mathcal{L}$ is consistent with the ordering defined by $\mathcal{P}$, that is, for all $u, v \in S$, $u \leq_{\mathcal{P}} v \implies u \leq_{\mathcal{L}} v$.*

We are now ready to state the main theorem which characterizes functions that satisfy Axioms 1 through 8 in terms of a partial ordering on $\mathbb{N}^*$.

Fix nodes u and v and let $P_1, \ldots, P_k$ be all the events that both u and v attend. Consider the sequence of numbers $(|P_i|)_{1\le i\le k}$ that give the number of people in each of these events. Without loss of generality assume that these are sorted in ascending order. Hence $|P_i| \le |P_{i+1}|$. We associate this *sorted sequence* of numbers with the tie (u, v). The partial order $\mathcal{N}$ induces a partial order on the set of pairs via this mapping. We also call this partial order $\mathcal{N}$. Fixing any particular measure of tie strength, gives a mapping of $\mathbb{N}^*$ to $\mathbb{R}$ and hence implies fixing a particular linear extension of $\mathcal{N}$, and fixing a linear extension of $\mathcal{N}$ involves making non-obvious decisions between elements of the partial order. We formalize this in the next theorem.

Theorem 11. *Let $G = (L \cup R, E)$ be a bipartite graph of users and events. For each pair of users, we associate a sequence $s \in N^*$. Given two users $(u, v) \in (L \times L)$, let $\{|P_i|\}_{1\le i\le k} \in R$ be the set of events common to users (u, v).*

We associate the sequence $n = (n_i)_{1\le i\le k} \in N^*$ with (u, v), where n_i is the sorted version of $\{|P_i|\}_{1\le i\le k} \in R$.

Through this association, the partial order $\mathcal{N} = (\mathbb{N}^*, \le_{\mathcal{N}})$ on finite sequences of numbers induces a partial order on $L \times L$ which we also call $\mathcal{N}$.

Let TS be a function that satisfies Axioms 1 through 8. Then TS induces a total order on the edges that is a linear extension of the partial order $\mathcal{N}$ on $L \times L$.

Conversely, for every linear extension $\mathcal{L}$ of the partial order $\mathcal{N}$, we can find a function TS that induces $\mathcal{L}$ on $L \times L$ and that satisfies Axioms 1 through 8.

Proof. Suppose we are given a tie strength function $TS : L \times L \to \mathbb{R}$. Hence, it gives a total order on the set of pairs of user. We want to show that if TS satisfies Axioms 1 through 8, then the total order is a linear extension of $\mathcal{N}$.

The characterization in Theorem 6 states that given a pair of vertices $(u, v) \in (L \times L)$, $TS(u, v)$ is characterized by the number of users in events common to u and v; and it can be expressed as $TS_G(u, v) = g(h(|P_i|))_{1\le i\le k}$ where g is a monotone submodular function and h is a monotone decreasing function. Since $TS : L \times L \to \mathbb{R}$, it induces a total order on all pairs of users. We now show that this total order is a consistent with the partial order $\mathcal{N}$.

Consider two pairs $(u_1, v_1), (u_2, v_2)$ with event profiles $x = (x_1, \ldots, x_X)$ and $y = (y_1, \ldots, y_Y)$. Suppose $x \ge_{\mathcal{N}} y$. We want to show that $TS(u_1, v_1) \ge TS(u_2, v_2)$. $x \ge_{\mathcal{N}} y$ implies that $X \ge Y$ and that $x_i \le y_i : \forall 1 \le i \le Y$.

$TS(u_1, v_1)$
$= g(h(x_1), \ldots, h(x_X))$
$\ge g(h(x_1), \ldots, h(x_Y))$ (because g is monotone and $X \ge Y$)
$\ge g(h(y_1), \ldots, h(y_B))$ (because g is monotone increasing, h is monotone decreasing, and $x_i \le y_i$)
$= TS(u_2, v_2)$

This proves the first part of the theorem.

For the converse, we are given a total ordering $\mathcal{L} = (\mathbb{N}^*, \le_{\mathcal{L}})$ that is an extension of the partial order $\mathcal{N}$. We want to prove that there exists a tie strength function $TS : L \times L \to \mathbb{R}$ that satisfies Axioms 1 through 8 and that induces $\mathcal{L}$ on $L \times L$. We shall prove this by constructing such a function. We shall define a function $f : \mathbb{N}^* \to \mathbb{Q}$ and define $TS_G(u, v) = f(a_1, \ldots, a_k)$, where $a_i = |P_i|$, the number of users that attend event P_i in G.

Define $f(n) = \frac{1}{n-1}$ and $f(\phi) = 0$. Hence, $TS_\phi(u, v) = f(\phi) = 0$ and $TS_{\{u,v\}}(u, v) = f(2) = \frac{1}{2-1} = 1$. This shows that TS satisfies Axiom 2. Also, define

$$f(\underbrace{1, 1, \ldots, 1}_{n}) = n$$

Since $\mathbb{N}^*$ is countable, consider elements in some order. If for the current element a under consideration, there exists an element b such that $a =_{\mathcal{N}} b$ and we have already defined $TS(b)$, then define $TS(a) = TS(b)$. Else, let $a_{glb} = argmax_e \{TS(e)$ is defined and $a \ge_{\mathcal{N}} e\}$ and let $a_{lub} = \text{argmin}_e \{TS(e)$ is defined and $a \le_{\mathcal{N}} e\}$. Since, at every point the sets over which we take the maximum of minimum are finite, both a_{glb} and a_{lub} are well defined and exist. Define $TS(a) = \frac{1}{2}(TS(a_{glb}) + TS(a_{lub}))$. □

In this abstract framework, an intuitively appealing linear extension is the random linear extension of the partial order under consideration. Polynomial time algorithms exist for this to calculation [12]. We leave the analysis of the analytical properties and its viability as a strength function in real world applications as an open research question.

In the next section, we turn our attention to actual measures of tie strength. We see some popular measures that have been proposed before as well as some new ones.

MEASURES OF TIE STRENGTH

There have been many measures of tie-strength discussed in previous literature. We review the most popular here and classify them according to the axioms that they satisfy. In this section, for a node u, the set of events that u attends (i.e. its neighborhood in the node×event graph) is represented by $\Gamma(u)$. For an event P, we denote by $|P|$ the number of people that attended event P.

Common Neighbors. This is the simplest measure of tie strength and is equal to the total number of common events that both u and v attended.

$$TS(u, v) = |\Gamma(u) \cap \Gamma(v)|$$

Jaccard Index. A more refined measure of tie strength is given by the Jaccard Index, which normalizes for how "social" u and v are

$$TS(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

Measures of Tie Strength \ Axioms	(1) Isomorphism	(2) Baseline	(3) Frequency	(4) Intimacy	(5) Popularity	(6) Cond. Indep. of Vertices	(7) Cond. Indep. of Events	(8) Submodularity	$g(a_1, \ldots, a_k)$ and $h(\|P_i\|) = a_i$ (From the characterization in Theorem 6)
Common Neighbors.	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ $h(n) = 1$
Jaccard Index.	✓	✓	✓	✓	✓	x	x	x	x
Delta.	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ $h(n) = \frac{1}{\binom{n}{2}}$
Adamic and Adar.	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ $h(n) = \frac{1}{\log n}$
Katz Measure.	✓	x	✓	✓	✓	✓	x	x	x
Preferential attachment.	✓	✓	x	✓	✓	✓	x	x	x
Random Walk with Restarts.	✓	x	x	x	✓	✓	x	x	x
Simrank.	✓	x	x	x	x	x	x	x	x
Max.	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \max_{i=1}^k a_i$ $h(n) = \frac{1}{n}$
Linear.	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ $h(n) = \frac{1}{n}$
Proportional.	✓	x	x	✓	x	✓	x	x	x
Temporal Proportional.	x	x	x	✓	✓	x	x	x	x

Table 1. Measures of tie strength and the axioms they satisfy

Delta. Tie strength increases with the number of events.

$$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\binom{|P|}{2}}$$

Adamic and Adar. This measure was introduced in [1].

$$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log |P|}$$

Linear. Tie strength increases with the number of events.

$$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{|P|}$$

Preferential attachment.

$$TS(u,v) = |\Gamma(u)| \cdot |\Gamma(v)|$$

Katz Measure. This was introduced in [13]. It counts the number of paths between u and v, where each path is discounted exponentially by the length of path. It is parametrized by a constant $0 \le \gamma \le 1$.

$$TS(u,v) = \sum_{q \in \text{ path between } u,v} \gamma^{-|q|}$$

Random Walk with Restarts. This gives a non-symmetric measure of tie strength. For each node u, we do a random walk centered at u defined as follows. At each step, the random walk jumps with probability α to the original node u. With probability $1-\alpha$, it picks a neighbor of the current node uniformly at random. α is called the restart probability. The tie strength between u and v is the stationary probability that we end at node v under this process.

Simrank. This captures the similarity between two nodes u and v by recursively computing the similarity of their neighbors. It is parametrized by a constant $0 \le \gamma \le 1$.

$$TS^{\gamma}(u,v) = \begin{cases} 1 & \text{if } u = v \\ \gamma \cdot \frac{\sum_{a \in \Gamma(u)} \sum_{b \in \Gamma(v)} TS(a,b)}{|\Gamma(u)| \cdot |\Gamma(v)|} & \text{otherwise} \end{cases}$$

Next, we introduce three new measures of tie strength. Following notation from Theorem 6, $g = \sum$ is at one extreme of the range of functions allowed by the Theorem and that is the default function used in the above measures. $g = \max$ is at the other extreme of the range of functions.

Max. Tie strength does not increase with number of events.

$$TS(u,v) = \max_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{|P|}$$

Proportional. Tie strength increases with number of events. People spend time proportional to their TS in an event. TS is the fixed-point of this equation:

$$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{\epsilon}{|P|} + (1-\epsilon) \frac{TS(u,v)}{\sum_{w \in \Gamma(u)} TS(u,w)}$$

Temporal Proportional. This is similar to Proportional, but with a temporal aspect. TS is not a fixed point, but starts

Dataset	# of People	# of Events
Southern Women	18	14
The Tempest	19	34
A Comedy of Errors	19	40
Macbeth	38	67
Reality Mining Bluetooth	104	326,248
Enron Emails	32,471	371,321

Table 2. Datasets used in our experiments

with a default value and is changed according to the following equation, where the events are ordered by time.

$$TS(u,v,t) = \begin{cases} TS(u,v,t-1) & \text{if } u \text{ and } v \text{ do not attend } P_t \\ \epsilon\frac{1}{|P_t|} + (1-\epsilon)\frac{TS(u,v,t-1)}{\sum_{w\in P_t} TS(u,w,t-1)} & \text{otherwise} \end{cases}$$

Table 1 provides a classification of all these tie-strength measures, according to which axioms they satisfy. If they satisfy all the axioms, then we use Theorem 6 to find the characterizing functions g and h. An interesting observation is that all the "self-referential" measures (such as Katz Measure, Random Walk with Restart, Simrank, and Proportional) fail to satisfy the axioms. Another interesting observation is in the classification of measures that satisfy the axioms. The majority use $g = \sum$ to aggregate tie strength across events. Per event, the majority compute tie strength as one over a simple function of the size of the event.

EXPERIMENTS

Our experiments answer the following questions:

Efficacy of our Approach: Can we predict useful tie strengths by using only event information and ignoring other information like timing, event type, and intensity?

Axiom Completeness: What is the *coverage* of our axioms? How much uncertainty in tie-strength rankings do our axioms remove?

Axiom Soundness: How different, in terms of conflicts in rankings, are our axioms from the tie-strength measures that do not satisfy them?

Rankings Correlation: How well do the rankings produced by different measures correlate with each other?

Data

Table 2 lists our datasets. We briefly describe each dataset next.

Davis Southern Women Social Events. We use the data collected by Davis et al. [5] on 18 Southern women and their attendances at 14 events from the 1930s.

Shakespearean Plays. We take three well-known plays by Shakespeare (Macbeth, The Tempest, and A Comedy of Errors) and create bipartite person×event graphs. The person-nodes are the characters in the play. Each event is a set of characters who are on the stage at the same time. We calculate the strength of ties between each pair of nodes. Thus without using any semantic information and even without analyzing any dialogue, we estimate how much characters interact with one another.

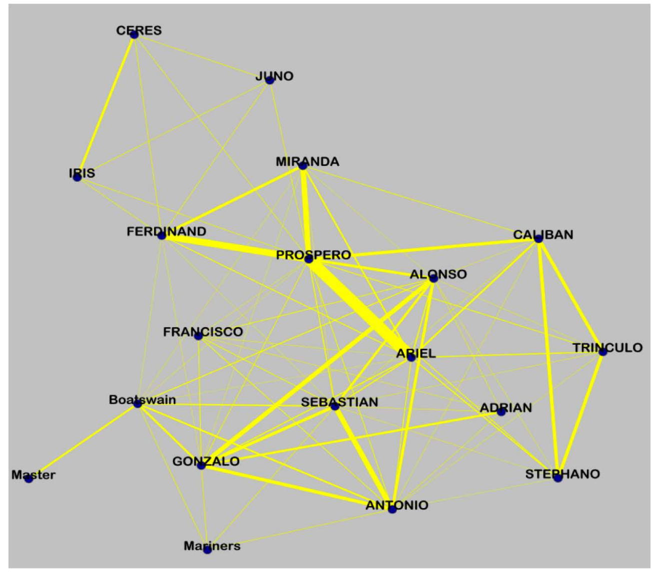

Figure 2. The inferred weighted social network between characters in The Tempest. The thicker an edge, the stronger the tie. Tie strength was calculated using the tie-strength measure Linear.

The Reality Mining Project [6]. The MIT Reality Mining project gave approximately one hundred smart phones to participants and logged information generated by these smart phones for 9 months. We use the bluetooth proximity data generated as part of this project. The bluetooth radio was switched on every five minutes and logged other bluetooth devices in close proximity. The people are the participants in the study and events record the proximity between people.

Enron Emails.[3] This dataset consists of emails from the Enron corporation. We consider all emails that occur between Enron addresses. Each email is an event and all the people copied on that email – i.e., the sender (from) and the receivers (to, cc and bcc) – are included in that event.

For brevity, we have omitted the plots showing the frequency distributions of the number of people at an event across our diverse data sets. As expected, these frequency distribution are very different across the various data sets. For Enron and Reality Mining, the distributions are scale free. For the Shakespearean Plays, the distributions are unimodal and bell-shaped. For the Southern Women data, the distribution is multimodal.

Efficacy of Tie Strength Measures

We start with the Shakespearean datasets. The aim is to infer strength of ties between characters of the play given only proximity information. We model each scene as an event. Concretely, an event is defined by the set of people on stage

[3]`http://www.cs.cmu.edu/~enron/`

at the same time. We obtain the strengths between characters from three Shakespearean plays using the tie-strength measure Linear. Note that the inference is only based on people occupying the stage at the same time and not on any semantic analysis of the text. Figure 2 shows the inferred weighted social network for The Tempest. For brevity, we have omitted the networks for Macbeth and A Comedy of Errors. The inferred weights (i.e., tie strengths) are consistent with the stories. For example, the highest tie strengths are between Ariel and Prospero in The Tempest. We also observed that the highest tie strengths were between Macbeth and Lady Macbeth in the play Macbeth and between Dromio of Syracuse and Antipholus of Syracuse in A Comedy of Errors. This experiment demonstrates that using *only* event information can capture the underlying tie strength between individuals.

Completeness of the Axioms

In Section "Axioms of Tie Strength," we discussed axioms governing tie strength and characterized the axioms in terms of a partial order in Theorem 11. We shall now conduct an experiment to determine the *completeness* of our set of axioms. Given a dataset, we measure *completeness* in terms of the number of tie-pairs that are ranked by the partial order. This will give us an empirical measure of how many tie-pairs are unresolved by a tie-strength function that satisfies Axioms 1 through 8.

We use Theorem 11 to conduct this experiment. For different datasets, we consider all possible rankings that satisfy our axioms by generating the partial order between all ties implied by Theorem 11.[4] We then calculate the percentage of ties that are comparable under this partial order. A high percentage will indicate that most ties are actually resolved by our axioms for real world datasets.

Each measure of tie strength gives a total order on the ties; and, hence resolves all the comparisons between pairs of ties. The number of tie-pairs which are left incomparable in the partial order gives a notion of the how much room the axioms leave open for different tie-strength functions to differ from each other.

Table 3 shows the percentage of all ties that are *not* resolved by the partial order (i.e., the percentage of the ties for which the partial order cannot tells us if one tie is greater or if they are equal); so a lower percentage is better. We observe that the partial order defined by our axioms does indeed resolve a very high percentage of the ties. Also, we see that our axioms resolve more ties in the scripted cleaner world of Shakespearean plays than in the real-world Reality Mining dataset.

Soundness of the Axioms

In the previous section, we looked at tie-strength functions that satisfy the measures of tie-strength, and measured the percentage of ties that were actually resolved by the axioms. In this section, we consider the issue of *soundness*. To empirically measure soundness, we look at measures of tie-strength

[4]The total number of tie pairs is $\binom{\binom{n}{2}}{2}$, where n = # people vertices. This means that for Enron Emails, the total number of tie pairs is $\binom{\binom{32471}{2}}{2} = 138,952,356,623,361,270$.

Dataset	Tie Pairs	Incomparable Pairs (%)
Southern Women	11,628	683 (5.87)
The Tempest	14,535	275 (1.89)
A Comedy of Errors	14,535	726 (4.99)
Macbeth	246,753	584 (0.24)
Reality Mining	13,794,378	1,764,546 (12.79)

Table 3. Number of ties *not* resolved by the partial order. The last column shows the percentage of tie pairs on which different tie-strength functions can differ.

that have been used previously in literature, and find *how much* they violate the axioms. To measure this, we use two implications of Theorem 11. First, our axioms are equivalent to the partial order on ties. Second, our axioms identify functions that do not obey the partial order. So, we use the proportion of tie-pairs in which the tie-strength order violates the partial order predicted by the axioms. We look at two tie-strength functions that do not obey the axioms: Jaccard Index and Temporal Proportional. Table 4 shows the number of tie-pairs that are actually in conflict. This experiment informs us about how far away a measure is from the axioms. We observe that for these datasets, Temporal Proportional agrees with the partial order more than the Jaccard Index. We also note that as the size of the dataset increases, the percentage of conflicts decreases.

Dataset	Tie Pairs	Jaccard (%)	Temporal(%)
S. Women	11,628	1,441 (12.39)	665 (5.72)
Tempest	14,535	488 (3.36)	261 (1.80)
Comedy	14,535	1,114 (7.66)	381 (2.62)
Macbeth	246,753	2,638 (1.07)	978 (0.40)
Reality	13,794,378	290,934 (2.11)	112,546 (0.82)

Table 4. Number of conflicts between the partial order and tie-strength functions: Jaccard Index and Temporal Proportional. The second and third columns show the percentage of tie-pairs in conflict with the partial order.

Measuring Correlation among Tie-Strength Functions

We want to measure how close different tie-strength functions are to each other. To do this, we calculate the correlation between the rankings generated by these functions. Figure 3 shows Kendall τ correlation coefficient for our datasets. We find that, depending on the data set, different measures of tie strength are correlated. For instance, in the "clean" world of Shakespearean plays Common Neighbor is the least correlated measure; while in the "messy" real world data from Reality Mining and Enron emails, Max is the least correlated measure. Moreover, we observe that Common Neighbor and Max are mostly uncorrelated ($-0.2 \leqslant \tau \leqslant 0.2$); and that Adamic-Adar, Delta, and Linear are highly positively correlated ($\tau > 0.6$) no matter the dataset.

CONCLUSIONS

We presented an axiomatic approach to the problem of inferring implicit social networks by measuring tie strength from bipartite person$\times$event graphs. We characterized functions that satisfy all axioms and demonstrated a range of measures that satisfy this characterization. We showed that in ranking

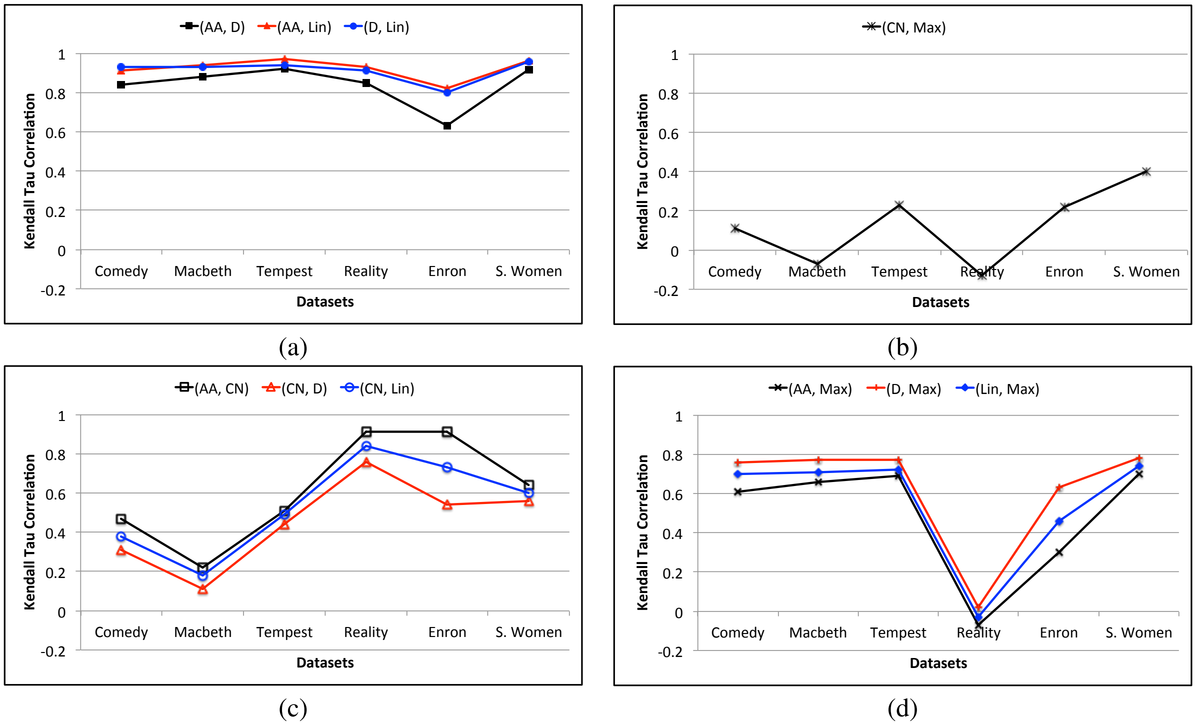

Figure 3. Measuring Kendall τ correlations on rankings produced by different tie-strength measures (that satisfy our axioms) on our datasets. The correlations highlight three groupings: {Adamic-Adar, Delta, Linear}, {Common-Neighbor}, {Max}. (a) Adamic-Adar (AA), Delta (D), and Linear (Lin) produce tie-strength rankings that are highly correlated ($0.6 < \tau$). (b) Common Neighbor (CN) and Max produce tie-strength rankings that are mostly uncorrelated ($-0.2 \leqslant \tau \leqslant 0.2$) or slightly correlated ($tau = 0.4$ for Southern Women). (c) Since Adamic-Adar, Delta, and Linear are highly correlated, their Kendall τ correlations with Common Neighbor are similar; and are all positively correlated. Enron has the most varied correlations from $\tau(CommonNeighbor, Delta) = 0.54$ to $\tau(AdamicAdar, CommonNeighbor) = 0.91$. (d) Again since Adamic-Adar, Delta, and Linear are highly correlated, their Kendall τ correlations with Max are similar. They are all positively correlated except on the Reality Mining data set, where their rankings are uncorrelated with Max. On the Enron data set, their ranking is positively correlated with Max but varied from $\tau(AdamicAdar, Max) = 0.3$ to $\tau(Delta, Max) = 0.65$.

applications, the axioms are equivalent to a natural partial order. We also demonstrated that to settle on a particular measure, we must make a non-obvious decision about extending this partial order to a total order, which is best left to the particular application. We classified measures found in prior literature according to the axioms that they satisfy. Finally, our experiments demonstrated the efficacy of our approach, the completeness and soundness of our axioms, and the Kendall τ correlation between various tie-strength measures.

REFERENCES

1. Adamic, L. A., and Adar, E. Friends and neighbors on the web. *Social Networks 25*, 3 (2003), 211–230.
2. Allali, O., Magnien, C., and Latapy, M. Link prediction in bipartite graphs using internal links and weighted projection. In *Proc. of the 3rd NetSciCom* (2011).
3. Altman, A., and Tennenholtz, M. Ranking systems: the pagerank axioms. In *Proc. of the 6th EC* (2005), 1–8.
4. Bornstein, R. Exposure and affect: Overview and meta-analysis of research. *Psychological Bulletin 106*, 2 (1989), 265–289.
5. Davis, A., Gardner, B., and Gardner, M. *Deep South*. University of Chicago Press, 1941.
6. Eagle, N., Pentland, A., and Lazer, D. Inferring social network structure using mobile phone data. *Proceedings of the National Academy of Sciences 106*, 36 (2009), 15274–15278.
7. Gilbert, E., and Karahalios, K. Predicting tie strength with social media. In *Proc. of the 27th CHI* (2009), 211–220.
8. Granovetter, M. The strength of weak ties. *American Journal of Sociology 78*, 6 (1973), 1360–1380.
9. Jeh, G., and Widom, J. Simrank: A measure of structural-context similarity. In *Proc. of the 8th KDD* (2002), 538–543.
10. Jin, R., Lee, V. E., and Hong, H. Axiomatic ranking of network role similarity. In *Proc. of the 17th KDD* (2011), 922–930.
11. Kahanda, I., and Neville, J. Using transactional information to predict link strength in online social networks. In *Proc. of the 3rd ICWSM* (2009).
12. Karzanov, A., and Khachiyan, L. On the conductance of order markov chains. *Order 8*, 1 (1991), 7–15.
13. Katz, L. A new status index derived from sociometric analysis. *Psychometrika 18*, 1 (1953), 39–43.
14. Liben-Nowell, D., and Kleinberg, J. The link prediction problem for social networks. In *Proc. of the 12th CIKM* (2003), 556–559.
15. Lin, D. An information-theoretic definition of similarity. In *Proc. of the 15th ICML* (1998), 296–304.
16. Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., Leiser, N., Matias, Y., and Merom, R. Suggesting friends using the implicit social graph. In *Proc. of the 16th KDD* (2010), 233–242.
17. Zajonc, R. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology 9*, 2 (1968), 1–27.