# An Axiomatic Approach to Tie-Strength Measures

**Mangesh Gupte** **Tina Eliassi-Rad**
Department of Computer Science
Rutgers University
Piscataway, NJ 08854
{mangesh, eliassi}@cs.rutgers.edu

## Abstract

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.

## 1 Introduction

Given a set of people and a set of events attended by them, we address the problem of measuring *tie strength* between various people. Mutual attendance at events implies an implicit social network between people. Our goal is to infer these implicit social networks by measuring tie-strength from bipartite person-event graphs. There has always been interest in inferring implicit social networks. Most recently, 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. Gmail's *don't forget bob* [4] feature uses an implicit social network to suggest new people to add to an email given a existing list. In our setting, people attend different events with each other. For example, an event can represent the set of people who took a photo at the same place and time, like COLOR. Given the set of events, we would like to infer how *connected* two people are. 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 present an axiomatic approach to the problem of inferring this implicit (weighted) social networks by measuring tie-strength from bipartite person-event graphs. We characterize functions that satisfy all the axioms and show a range of measures that satisfy this characterization. We classify measures found in prior literature according to the axioms that they satisfy. Lastly, we show empirically through the use of Kendall's Tau correlation whether a dataset is well-behaved, where we do not have to worry about which tie-strength measure to choose, or whether we have to be careful about the exact choice of measure.

## 2 Tie Strength

We model users and events as nodes and use a bipartite graph $G = (L \cup R, E)$ where the edges represent membership. The left vertices correspond to users while the right vertices correspond to events. We ignore any information other than the set of people who attended the events such as events' timing, location, importance, etc. 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. 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 events.

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 $TS_{\{E_1,\ldots,E_k\}}(u, v)$ 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 $\mathbb{N}^* = \cup_k \mathbb{N}^k$. We now discuss the axioms that various measures of tie strength between two users $u$ and $v$ must follow.

**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.

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

**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 of $u$ and $v$.
Given a graph $G = (L \cup R, E)$ and two vertices $u, v \in L$. 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 the $TS_{G'}(u, v) \geq TS_G(u, v)$.

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

**Axiom 5 (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 $Q$.
$|P| \geq |Q| \implies \sum_{u,v \in P} TS_P(u, v) \geq \sum_{u,v \in Q} TS_Q(u, v)$.

**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.

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

**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.
If $G$ is a graph and $Q$ is a single event, $TS_G(u, v) + TS_Q(u, v) \geq TS_{G+Q}(u, v)$.

These axioms provide a measure of tie strength between nodes that is positive but unbounded. Nodes that have a higher value are closer to each other than nodes that have lower value. While these axioms are pretty strong, they still encompass a wide variety of tie-strength functions. Theorem 1 gives a characterization of all functions that satisfy the axioms of tie strength. Axioms (1-8) do not uniquely define a function, and in fact, one of the reasons that tie strength is not uniquely defined up to the given axioms is that we do not have any notion for comparing the relative importance of number of events, Axiom 3 (Frequency: More events create stronger ties), versus the exclusivity of events, Axiom 4 (Intimacy: Smaller events create stronger ties) . In terms of the partial order, it is not clear whether $u$ and $v$ having two events in common with two people attending them is better than or worse than $u$ and $v$ having three events in common with three people attending them.

**Theorem 1.** *Given a graph $G = (L \cup R, E)$ and two vertices $u, v$, if the tie-strength function $TS$ follows Axioms (1-8), then the function has to be 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 $v$, $h : \mathbb{N} \to \mathbb{R}$ is a monotonically decreasing function bounded by $1 \geq h(n) \geq \frac{1}{\binom{n}{2}}$ and $g : \mathbb{N}^* \to \mathbb{R}$ is a monotonically increasing submodular function.*

*Proof.* Omitted for brevity. □

Theorem 1 gives us a way to explore the space of valid functions for representing tie strength and find which work well given particular applications. Table 1 classifies the various measures used in prior literature according to which axioms they satisfy. (See Appendix A for definitions.) If a measure satisfies all the axioms, we use Theorem 1 to find the characterizing functions $g$ and $h$.

| Axioms / Measures of Tie Strength | Axiom 1 (Isomorphism) | Axiom 2 (Baseline) | Axiom 3 (Frequency: More events create stronger ties) | Axiom 4 (Intimacy: Smaller events create stronger ties) | Axiom 5 (Larger events create more ties) | Axiom 6 (Conditional Independence of Vertices) | Axiom 7 (Conditional Independence of Events) | Axiom 8 (Submodularity) | $g(a_1, \ldots, a_k)$ and $h(\lvert P_i \rvert) = a_i$ (From the characterization in Theorem 1) |
|---|---|---|---|---|---|---|---|---|---|
| Common Neighbors. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | $g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ <br> $h(n) = 1$ |
| Jaccard Index. | ✓ | ✓ | ✓ | ✓ | ✓ | x | x | x | x |
| Delta. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | $g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ <br> $h(n) = \frac{1}{\binom{n}{2}}$ |
| Adamic and Adar. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | $g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ <br> $h(n) = \frac{1}{\log n}$ |
| 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. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | $g(a_1, \ldots, a_k) = \max_{i=1}^k a_i$ <br> $h(n) = \frac{1}{n}$ |
| Linear. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | $g(a_1, \ldots, a_k) = \sum_{i=1}^k a_i$ <br> $h(n) = \frac{1}{n}$ |
| Proportional. | ✓ | x | x | ✓ | x | ✓ | x | x | x |

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

## 3 Experiments

We conducted several experiments on 5 real data-sets (namely, Enron emails,[1] Reality Mining proximity data [2], and three Shakespearean plays). One of our experiments[2] measured the Kendall's $\tau$ coefficient between the measures of tie strength listed in Table 1 that satisfy Axioms 1 through 8 (see Figure 1). 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. Conducting such experiments can inform us of whether a dataset is well-behaved, where we do not have to worry about which tie-strength measure to choose, or whether we have to be careful about the exact choice of measure.

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

[2] For brevity, we have omitted experimental results on measuring the coverage of Axioms 1-8 by computing (a) the percentage of incomparable tie pairs and (b) the number of conflicts between the partial order and tie-strength functions that do *not* satisfy the axioms.

**A Comedy of Errors**

| | Adamic-Adar | Common Neighbor | Delta | Linear | Max |
|---|---|---|---|---|---|
| Adamic-Adar | 1.00 | 0.47 | 0.84 | 0.91 | 0.61 |
| Common Neighbor | 0.47 | 1.00 | 0.31 | 0.38 | 0.11 |
| Delta | 0.84 | 0.31 | 1.00 | 0.93 | 0.76 |
| Linear | 0.91 | 0.38 | 0.93 | 1.00 | 0.70 |
| Max | 0.61 | 0.11 | 0.76 | 0.70 | 1.00 |

**Macbeth**

| | Adamic-Adar | Common Neighbor | Delta | Linear | Max |
|---|---|---|---|---|---|
| Adamic-Adar | 1.00 | 0.22 | 0.88 | 0.94 | 0.66 |
| Common Neighbor | 0.22 | 1.00 | 0.11 | 0.18 | -0.07 |
| Delta | 0.88 | 0.11 | 1.00 | 0.93 | 0.77 |
| Linear | 0.94 | 0.18 | 0.93 | 1.00 | 0.71 |
| Max | 0.66 | -0.07 | 0.77 | 0.71 | 1.00 |

**Tempest**

| | Adamic-Adar | Common Neighbor | Delta | Linear | Max |
|---|---|---|---|---|---|
| Adamic-Adar | 1.00 | 0.51 | 0.92 | 0.97 | 0.69 |
| Common Neighbor | 0.51 | 1.00 | 0.44 | 0.49 | 0.23 |
| Delta | 0.92 | 0.44 | 1.00 | 0.94 | 0.77 |
| Linear | 0.97 | 0.49 | 0.94 | 1.00 | 0.72 |
| Max | 0.69 | 0.23 | 0.77 | 0.72 | 1.00 |

**Reality Mining**

| | Adamic-Adar | Common Neighbor | Delta | Linear | Max |
|---|---|---|---|---|---|
| Adamic-Adar | 1.00 | 0.91 | 0.85 | 0.93 | -0.07 |
| Common Neighbor | 0.91 | 1.00 | 0.76 | 0.84 | -0.13 |
| Delta | 0.85 | 0.76 | 1.00 | 0.91 | 0.02 |
| Linear | 0.93 | 0.84 | 0.91 | 1.00 | -0.03 |
| Max | -0.07 | -0.13 | 0.02 | -0.03 | 1.00 |

**Enron Emails**

| | Adamic-Adar | Common Neighbor | Delta | Linear | Max |
|---|---|---|---|---|---|
| Adamic-Adar | 1.00 | 0.91 | 0.63 | 0.82 | 0.30 |
| Common Neighbor | 0.91 | 1.00 | 0.54 | 0.73 | 0.22 |
| Delta | 0.63 | 0.54 | 1.00 | 0.80 | 0.63 |
| Linear | 0.82 | 0.73 | 0.80 | 1.00 | 0.46 |
| Max | 0.30 | 0.22 | 0.63 | 0.46 | 1.00 |

Figure 1: Kendall's $\tau$ coefficient for Shakespearean plays, the Reality Mining data and the Enron emails. The color scale goes from bright green (coefficient = 1) to bright red (coefficient = -1). In the Skakespearean plays, the least correlated measure is Common Neighbor (as indicated by the red cells in that column). In the real-world communication networks of Enron and Reality Mining, the least correlated measure is Max (again as indicated by the red cells in that column).

## 4 Conclusions

To recap, we presented an axiomatic approach to the problem of inferring implicit social networks by measuring tie strength from bipartite person-event graphs. We characterized functions that satisfy all the axioms and showed a range of measures that satisfy this characterization. We classified measures found in prior literature according to the axioms that they satisfy.

## Appendix A

There have been many measures of tie strength discussed in previous literature. We review them here. For an event $P$, we denote by $|P|$ the number of people in the event $P$. For a person $u$, we denote by $|\Gamma(u)|$ the number of events $u$ attended.

**Common Neighbors.** This is the simplest measure of tie strength, given by 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 gives importance to events that normalizes for how "social" $u$ and $v$ are

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

**Delta.**

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

**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 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 [3]. It counts the number of paths between $u$ and $v$ where each path is discounted exponentially by the length of path.

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

**Random Walk with Restarts.** This gives a non-symmetric measure of tie strength . For a node $u$, we either jump with probability $\alpha$ to a node $u$ and with probability $1 - \alpha$ to a neighbor of the current node. 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.

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

**Max.** Tie strength does not increases 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 a party. S is the fixed point of this set of equations:

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

## References


[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211–230, 2003.

[2] N. Eagle, A. Pentland, and D. Lazer. Inferring social network structure using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

[3] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[4] M. Roth, A. Ben-David, D. Deutscher, G. Flysher, I. Horn, A. Leichtberg, N. Leiser, Y. Matias, and R. Merom. Suggesting friends using the implicit social graph. In *Proceedings of the 16th Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 233–242, 2010.