Measuring Tie-Strength in Implicit Social Networks

Tina Eliassi-Rad

tina@eliassi.org

Joint with Mangesh Gupte (Rutgers → Google)
Problem Definition

• Given a bipartite graph with people as one set of vertices and events as the other set, measure tie strength between each pair of individuals

• Assumption
  – Attendance at mutual events implies an implicit weighted social network between people
Motivation

• Most real-world networks are 2-mode and are converted to a 1-mode (e.g., $AA^T$)

• Explicitly declared friendship links can suffer from a low signal-to-noise ratio (e.g., Facebook friends)

---

• **Challenge:** Detect which of links in the 1-mode graph are important

• **Goal:** Infer the *implicit weighted social network* from people’s participation in mutual events
Tie Strength

• A measure of tie strength induces
  – a ranking on all the edges, and
  – a ranking on the set of neighbors for every person

• Example of a simple tie-strength measure
  – Common neighbor measures the total number of common events to a pair of individuals
5. MEASURES OF TIE STRENGTH

In previous literature, there have been plenty of measures of tie strength discussed. We review these and classify them. In this section, we turn our attention to actual measures of tie strength, given by the total number of common events attended.

Tie strength increases with number of events. This is the simplest measure of tie strength, given by the total number of common events attended.

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

• There are many different measures of tie-strength
  1. Common neighbor
  2. Jaccard index
  3. Max
  4. Linear
  5. Delta
  6. Adamic and Adar
  7. Preferential attachment
  8. Katz measure
  9. Random walk with restarts
  10. Simrank
  11. Proportional
  12. ...

Which one should you choose?
• An axiomatic approach to the problem of inferring implicit social networks by measuring tie strength
• A characterization of functions that satisfy all our axioms
• Classification of prior measures according to the axioms that they satisfy
• Experiments
• Conclusions
Running Example

**Input**
People × Event Bipartite Graph

```
<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Output**
Partial Order of Tie Strength among People

```
(a,b) (c,d) high
(b,c), (b,d), (c,e), (d,e)
(a,c), (a,d), (a,e), (b,e) low
```

The arrows indicate the direction of the partial order.
Axioms

• Axiom 1: Isomorphism
• Axiom 2: Baseline
• Axiom 3: Frequency
• Axiom 4: Intimacy
• Axiom 5: Popularity
• Axiom 6: Conditional Independence of People
• Axiom 7: Conditional Independence of Events
• Axiom 8: Submodularity
Axiom 1: Isomorphism

• Tie strength between \( u \) and \( v \) is independent of the labels of \( u \) and \( v \)
Axiom 2: Baseline

• If there are no events, then tie strength between each pair $u$ and $v$ is 0

$$TS_{\emptyset}(u, v) = 0$$

• If there are only two people $u$ and $v$ and a single event $P$ that they attend, then their tie strength is at most 1

$$TS_{P}(u, v) \leq 1$$

– Defines an upper-bound for how much tie strength can be generated from a single event between two people
Axiom 3: Frequency & Axiom 4: Intimacy

• Axiom 3 (Frequency)
  – More events create stronger ties
  – All other things being equal, the more events common to $u$ and $v$, the stronger their tie-strength

• 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 their tie-strength
Axiom 5: Popularity

• Larger events create more ties
• Consider two events $P$ and $Q$
• If $|Q| > |P|$, then the total tie strength created by $Q$ is more than that created by $P$
Axioms 6 & 7: Conditional Independence of People and of Events

• Axiom 6: **Conditional Independence of People**
  – A node $u$’s tie strength to other people does **not** depend on events that $u$ does **not** attend

• 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 $v$
  
  \[
  TS_{(G+P)}(u, v) = g(TS_G(u, v), TS_P(u, v))
  \]
  
  – where $g$ is some monotonically increasing function
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, then

$$TS_{(G+Q)}(u, v) - TS_G(u, v) \leq TS_Q(u, v)$$
Example – Mapping to Axioms

Input
People × Event Bipartite Graph

Output
Partial order of Tie Strength

(a,b) (c,d)
(b,c), (b,d), (c,e), (d,e)
(a,c), (a,d), (a,e), (b,e)

Axiom 1 (Isomorphism)
Axiom 2 (Baseline) & Axiom 6 (Cond. Indep. of Vertices) & Axiom 7 (Cond. Indep. of Events)
Axiom 3 (Freq)
Axiom 4 (Intimacy) & Axiom 6 (Cond. Indep. of Vertices) & Axiom 7 (Cond. Indep. of Events)
Observations on the Axioms

• Our axioms are fairly intuitive

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A5: Popularity</td>
<td>A6: Cond. Indep. of people</td>
<td>A7: Cond. indep. of events</td>
<td>A8: Submodularity</td>
</tr>
</tbody>
</table>

• **But**, several previous measures in the literature break some of these axioms

• Satisfying all the axioms is **not** sufficient to uniquely identify a measure of tie strength
  
  – One reason: inherent tension between Axiom 3 (Frequency) and Axiom 4 (Intimacy)
Inherent Tension Between Frequency & Intimacy

• Scenario #1 (intimate)
  – Mary and Susan go to 2 parties, where they are the only people there.

• Scenario #2 (frequent)
  – Mary, Susan, and Jane go to 3 parties, where they are the only people there.

• In which scenario is Mary’s tie to Susan stronger?
Observations on the Axioms (cont.)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A5: Popularity</td>
<td>A6: Cond. Indep. of people</td>
<td>A7: Cond. indep. of events</td>
<td>A8: Submodularity</td>
<td></td>
</tr>
</tbody>
</table>

• Axioms are equivalent to a **natural partial order** on the strength of ties
  – Pertinent to ranking application

• Choosing a particular tie-strength function is equivalent to choosing a particular **linear extension** of this partial order
  – Non-obvious decision
Preamble to the Characterization Theorem

• Let $f(n) = \text{total tie strength generated in a single event with } n \text{ people}$
• If there is a single party with $n$ people, the tie strength of each tie is
  \[ \frac{f(n)}{\binom{n}{2}} \]
  – Based on Axiom 1 (Isomorphism)
• The total tie strength created at an event $P$ with $n$ people is a monotone function $f(n)$ that is bounded by
  \[ 1 \leq f(n) \leq \binom{n}{2} \]
  – Based on Axiom 2 (Baseline) and Axiom 4 (Intimacy) and Axiom 5 (Popularity)
Characterizing Tie Strength

**Theorem.** Given a graph $G = (L \cup R, E)$ and two vertices $u$ and $v$, if the tie-strength function $TS$ follows Axioms (1-8), then the function has to be of the form

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

- $\{P_i\}_{1 \leq i \leq k}$ are the events common to both $u$ and $v$
- $h$ is a monotonically decreasing function bounded by $1 \geq h(n) \geq \frac{1}{\binom{n}{2}}$, $n \geq 2$; $h(1) = 1$; $h(0) = 0$.
- $g$ is a monotonically increasing submodular function
Many Measures of Tie Strength

1. Common neighbor
2. Jaccard index
3. Max
4. Linear
5. Delta
6. Adamic and Adar
7. Preferential attachment
8. Katz measure
9. Random walk with restarts
10. Simrank
11. Proportional

\[ TS(u, v) = |\Gamma(u) \cap \Gamma(v)| \]

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

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

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

\[ TS(u, v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{(|P|)^2} \]

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

\[ TS(u, v) = |\Gamma(u)| \cdot |\Gamma(v)| \]

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

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

\[ TS(u, v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{\epsilon}{|P|} + (1 - \epsilon) \frac{TS(u, v)}{\sum_{w \in \Gamma(u)} TS(u, w)} \]
Non Self-Referential Tie Strength Measures

- **Common neighbor**
  - The total # of common events that both $u$ and $v$ attended

- **Jaccard Index**
  - Similar to common neighbor
  - Normalizes for how “social” $u$ and $v$ are

- **Adamic and Adar [2003], Delta, and Linear**
  - Tie strength increases with the number of events
  - Tie strength is 1 over a simple function of event size

- **Max**
  - Tie strength does not increase with the number of events
  - Tie strength is the maximum tie strength from all common events
Self-Referential Tie-Strength Measures

• **Katz measure [Katz, 1953]**
  – Tie strength is the number of paths between $u$ and $v$, where each path is discounted exponentially by the length of the path

• **Random walk with restarts**
  – A non-symmetric measure of tie strength
  – Tie strength is the stationary probability of a Markov chain process
  – With probability $\alpha$, jump to a node $u$; and with probability $1-\alpha$, jump to a neighbor of a current node.

• **Simrank [Jeh & Widom, 2002]**
  – Tie strength is captured by recursively computing the tie strength of neighbors

• **Proportional**
  – Tie strength increases with # of events
  – People spend time proportional to their tie-strength at a party
# Measures of Tie-Strength that Satisfy All the Axioms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A5: Popularity</td>
<td>A6: Cond. indep. of P</td>
<td>A7: Cond. indep. of E</td>
<td>A8: Submodularity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Neighbors</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>g(a₁, ..., aₖ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(</td>
<td>P_i</td>
<td>) = a_i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g(a₁, ..., aₖ) = Σa_i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delta</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>g(a₁, ..., aₖ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = Σa_i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = 2(n(n-1))⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adamic &amp; Adar</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>g(a₁, ..., aₖ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = (log(n))⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>g(a₁, ..., aₖ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = n⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>g(a₁, ..., aₖ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = n⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g(a₁, ..., aₖ) = max{a_i}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(n) = n⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Measures of Tie-Strength that Do **Not** Satisfy All the Axioms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaccard Index</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Katz Measure</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Preferential Attachment</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>RWR</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Simrank</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Proportional</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

The Jaccard Index satisfies all the axioms except for **A7: Cond. indep. of E**. It is used to measure the similarity between sets, calculated as:

\[ \text{Jaccard Index} = \frac{|A \cap B|}{|A \cup B|} \]

Katz Measure satisfies all the axioms except for **A4: Intimacy** and **A8: Submodularity**.

Preferential Attachment satisfies all the axioms except for **A7: Cond. indep. of E**.

RWR satisfies all the axioms except for **A8: Submodularity**.

Simrank satisfies all the axioms except for **A7: Cond. indep. of E**.

Proportional satisfies all the axioms except for **A7: Cond. indep. of E**.
Tie Strength and Orderings

- Let $TS$ be a function that satisfies Axioms 1-8

<table>
<thead>
<tr>
<th>(1) Isomorphism</th>
<th>(2) Baseline</th>
<th>(3) Frequency</th>
<th>(4) Intimacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5) Popularity</td>
<td>(6) Cond. indep. of P</td>
<td>(7) Cond. indep. of E</td>
<td>(8) Submodularity</td>
</tr>
</tbody>
</table>

- $TS$ induces a **total order** on the edges that is a linear extension of the partial order on the node-tie pairs
Theorem 11. Let $G = (L \cup R, E)$ be a bipartite graph of users and events. Given two users $(u, v) \in (L \times L)$, let $(|P_i|)_{1 \leq i \leq k} \in R$ be the set of events common to users $(u, v)$. Through this association, the partial order $\mathcal{N} = (\mathbb{N}^*, \leq_\mathcal{N})$ 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 (1-8). 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 (1-8).
## Data Sets

<table>
<thead>
<tr>
<th>Graphs</th>
<th># of People</th>
<th># of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Women</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>The Tempest</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>A Comedy of Errors</td>
<td>19</td>
<td>40</td>
</tr>
<tr>
<td>Macbeth</td>
<td>38</td>
<td>67</td>
</tr>
<tr>
<td>Reality Mining Bluetooth</td>
<td>104</td>
<td>326,248</td>
</tr>
<tr>
<td>Enron Emails</td>
<td>32,471</td>
<td>371,321</td>
</tr>
</tbody>
</table>
Degree Distributions

Enron & Reality Mining

Shakespeare’s Plays

Southern Women
Completeness of Axioms 1-8  
(Number of Ties \textbf{Not} Resolved by the Partial Order)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tie Pairs</th>
<th>Incomparable Pairs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Women</td>
<td>11,628</td>
<td>683 (5.87)</td>
</tr>
<tr>
<td>The Tempest</td>
<td>14,535</td>
<td>275 (1.89)</td>
</tr>
<tr>
<td>A Comedy of Errors</td>
<td>14,535</td>
<td>726 (4.99)</td>
</tr>
<tr>
<td>Macbeth</td>
<td>246,753</td>
<td>584 (0.23)</td>
</tr>
<tr>
<td>Reality Mining</td>
<td>13,794,378</td>
<td>1,764,546 (12.79)</td>
</tr>
</tbody>
</table>

- % of tie-pairs where different tie-strength functions can differ
  - Smaller is better
  - Generally, percentages are small
  - Large real-world networks have more unresolved ties

\[
\text{# of tie pairs} = \binom{n}{2}
\]
Take-away point #1

% of tie pairs on which different tie strength functions can differ is small.*

* This is for ranking application and tie strength functions satisfying the axioms.
Two Tie-Strength Functions that Do Not Satisfy the Axioms

• **Jaccard Index**
  – Normalizes for how “social” $u$ and $v$ are
  $$TS(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

• **Temporal Proportional**
  – Increases with number of events
  – People spend time proportional to their tie-strength in a party
  – Events are ordered by time
  $$TS(u, v, t) = \begin{cases} 
  TS(u, v, t - 1) & \text{if } u \text{ and } v \text{ do not attend } P_t \\
  \epsilon \frac{1}{|P_t|} + (1 - \epsilon) \frac{TS(u, v, t - 1)}{\sum_{w \in P_t} TS(u, w, t - 1)} & \text{otherwise}
  \end{cases}$$
Soundness of Axioms 1-8
(Number of Conflicts Between the Partial Order and Tie-Strength Functions Not Satisfying the Axioms)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tie Pairs</th>
<th>Jaccard (%)</th>
<th>Temporal (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Women</td>
<td>11,628</td>
<td>1,441 (12.39)</td>
<td>665 (5.72)</td>
</tr>
<tr>
<td>The Tempest</td>
<td>14,535</td>
<td>488 (3.35)</td>
<td>261 (1.79)</td>
</tr>
<tr>
<td>A Comedy of Errors</td>
<td>14,535</td>
<td>1,114 (7.76)</td>
<td>381 (2.62)</td>
</tr>
<tr>
<td>Macbeth</td>
<td>246,753</td>
<td>2,638 (1.06)</td>
<td>978 (0.39)</td>
</tr>
<tr>
<td>Reality Mining</td>
<td>13,794,378</td>
<td>290,934 (0.02)</td>
<td>112,546 (0.01)</td>
</tr>
</tbody>
</table>

- % of tie-pairs in conflict with the partial order
  - **Smaller is better**
  - Generally, percentages are small
  - They decrease as the dataset increases
More on Soundness

• **Question 1:**

  Are the number of conflicts, between the partial order and tie-strength functions not satisfying the axioms, **small** because most of the tie-strengths are zeros (sparsity of real graph)?

• **Answer:**
  
  • This is **partially true**.
  
  • For some pairs, the tie-strength being set to zero is caused by the axioms.
  
  • It may or may not be true that all these pairs have tie-strength zero in the actual function used.
    
    – For example, this won’t be true for some self-referential functions like Simrank, Random Walk with Restart, etc.
Even More on Soundness

- **Question 2:** How do the conflict numbers change if we only looked at tie pairs that have nonzero tie-strengths?
- **Answer:** The percentages go up but not by much.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tie Pairs</th>
<th>Tie Pairs (excluding TS=0)</th>
<th>Jaccard</th>
<th>Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Women</td>
<td>11,628</td>
<td>11,537</td>
<td>1,441</td>
<td>665</td>
</tr>
<tr>
<td>The Tempest</td>
<td>14,535</td>
<td>10,257</td>
<td>488</td>
<td>261</td>
</tr>
<tr>
<td>A Comedy of Errors</td>
<td>14,535</td>
<td>11,685</td>
<td>1,114</td>
<td>381</td>
</tr>
<tr>
<td>Macbeth</td>
<td>246,753</td>
<td>74,175</td>
<td>2,638</td>
<td>978</td>
</tr>
<tr>
<td>Reality Mining</td>
<td>13,794,378</td>
<td>12,819,272</td>
<td>290,934</td>
<td>112,546</td>
</tr>
</tbody>
</table>
Even More on Soundness
Take-away point #2
% of conflicts between our axioms and tie-strength functions not satisfying our axioms is small.*

* This is for ranking application.
Take-away point #1
% of tie pairs on which different tie-strength functions can differ is small.

Take-away point #2
% of conflicts between our axioms and tie-strength functions not satisfying our axioms is small.

Take-away point #3
If your application is ranking, just pick the most computationally efficient tie-strength measure (e.g. common neighbor).
# Tie Strength Measures Used in Rank Correlation Experiments

<table>
<thead>
<tr>
<th>Tie Strength Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Neighbor</td>
<td>$TS(u, v) =</td>
</tr>
<tr>
<td>Max</td>
<td>$TS(u, v) = \max_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{</td>
</tr>
<tr>
<td>Linear</td>
<td>$TS(u, v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{</td>
</tr>
<tr>
<td>Delta</td>
<td>$TS(u, v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{(</td>
</tr>
<tr>
<td>Adamic-Adar</td>
<td>$TS(u, v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log</td>
</tr>
</tbody>
</table>
Kendall $\tau$ Coefficient

• It is a measure of rank correlation
  – The similarity of the orderings of the data when ranked by each of the quantities

$$\tau = \frac{(\text{# of concordant pairs}) - (\text{# of discordant pairs})}{\frac{1}{2} n(n-1)}$$
Adamic-Adar, Delta, & Linear produce TS rankings that are highly correlated

A measure of tie strength induces TS $\tau_i$ satisfies $TS(u, v) = \sum a_i$.

Tie strength increases with number of events

There have been plenty of measures of tie strength discussed before as well as some new ones.

A1 A2 A3 A4 A5 A6 A7 A8 $g(a_1, ..., a_n)$ $h(|P|) = a_i$

<table>
<thead>
<tr>
<th>Function</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Neighbors</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Delta</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Adamic &amp; Adar</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Linear</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Max</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

$g(a_1, ..., a_n) = \sum a_i$

$h(n) = \begin{cases} 1 & \text{if } n = 1 \\ \frac{2(n(n-1))}{n^2} & \text{if } n > 1 \end{cases}$

$g(a_1, ..., a_n) = \sum a_i$

$\log(n)$

$g(a_1, ..., a_n) = \text{max}(a_i)$

$\sum a_i$
Common Neighbor & Max produce TS rankings that are mostly uncorrelated

<table>
<thead>
<tr>
<th>Problem Definition</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>( g(a_i, \ldots, a_k) )</th>
<th>( h({P_i}) = a_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Neighbors</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>( a_i )</td>
<td>1</td>
</tr>
<tr>
<td>Delta</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>( a_i )</td>
<td>( 2(n(n-1))^{-1} )</td>
</tr>
<tr>
<td>Adamic &amp; Adar</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>( a_i )</td>
<td>( (\log(n))^{-1} )</td>
</tr>
<tr>
<td>Linear</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>( a_i )</td>
<td>( n^{-1} )</td>
</tr>
<tr>
<td>Max</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>( \max(a_i) )</td>
<td>( n^{-1} )</td>
</tr>
</tbody>
</table>

Note: Kendall Tau Correlation.

Graphs: (CN, Max)
Take-away point #4
Kendall τ correlations on rankings produced by tie-strength functions (that satisfy our axioms) highlight three groups: (1) \{Adamic-Adar, Delta, Linear\}, (2) \{Common Neighbor\}, and (3) \{Max\}.
Scalability Issue

- # of tie pairs = \( \binom{n}{2} \)
- Enron has 32,471
- # of tie pairs in Enron \( \approx 138 \) quadrillion

\[ \binom{32471}{2} = 138,952,356,623,361,270 \]
- Ignore zero tie-strengths
Related Work

• Strength of ties
  – Spread of information in social networks [Granovetter, 1973]
  – Use external information to learn strength of tie
    • [Gilbert & Karahalios, 2009], [Kahanda & Neville, 2009]

• Very few axiomatic work approaches to graph measures
  – PageRank axiomatization [Altman & Tennenholtz, 2005]
  – Information theoretic measure of similarity [Lin, 1998]
    • Assumes probability distribution over events

• Link prediction
  – [Adamic & Adar, 2003]
  – [Liben-Nowell & Kleinberg, 2003]
  – [Sarkar, Chakrabarti, Moore, 2010 & 2011]
Conclusions

1. Presented an axiomatic approach to the problem of inferring implicit social networks by measuring tie strength
2. Characterized functions that satisfy all the axioms
3. Classified prior measures according to the axioms that they satisfy
4. Demonstrated coverage of axioms, conflict with axioms, and correlation among tie-strength measures
5. In ranking applications, the axioms are equivalent to a natural partial order
Take-away point #5

Axiomatic approaches to various measures on networks (such as tie-strength measures in this study) enable us to systematically study existing measures and characterize functions that satisfy our axioms.
Thank You!

Details @ http://eliassi.org/papers/gupte-websci12.pdf

A Comedy of Errors

The Tempest

Supported by LLNL & DTRA