RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

MLG 2013

Measuring Tie-Strength in Implicit Social Networks

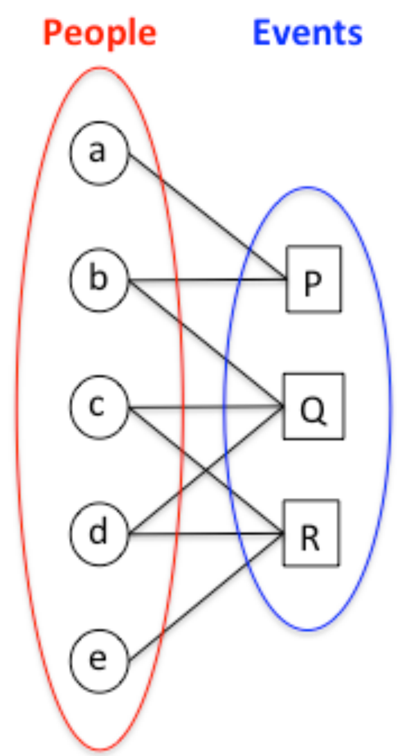
Tina Eliassi-Rad

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)

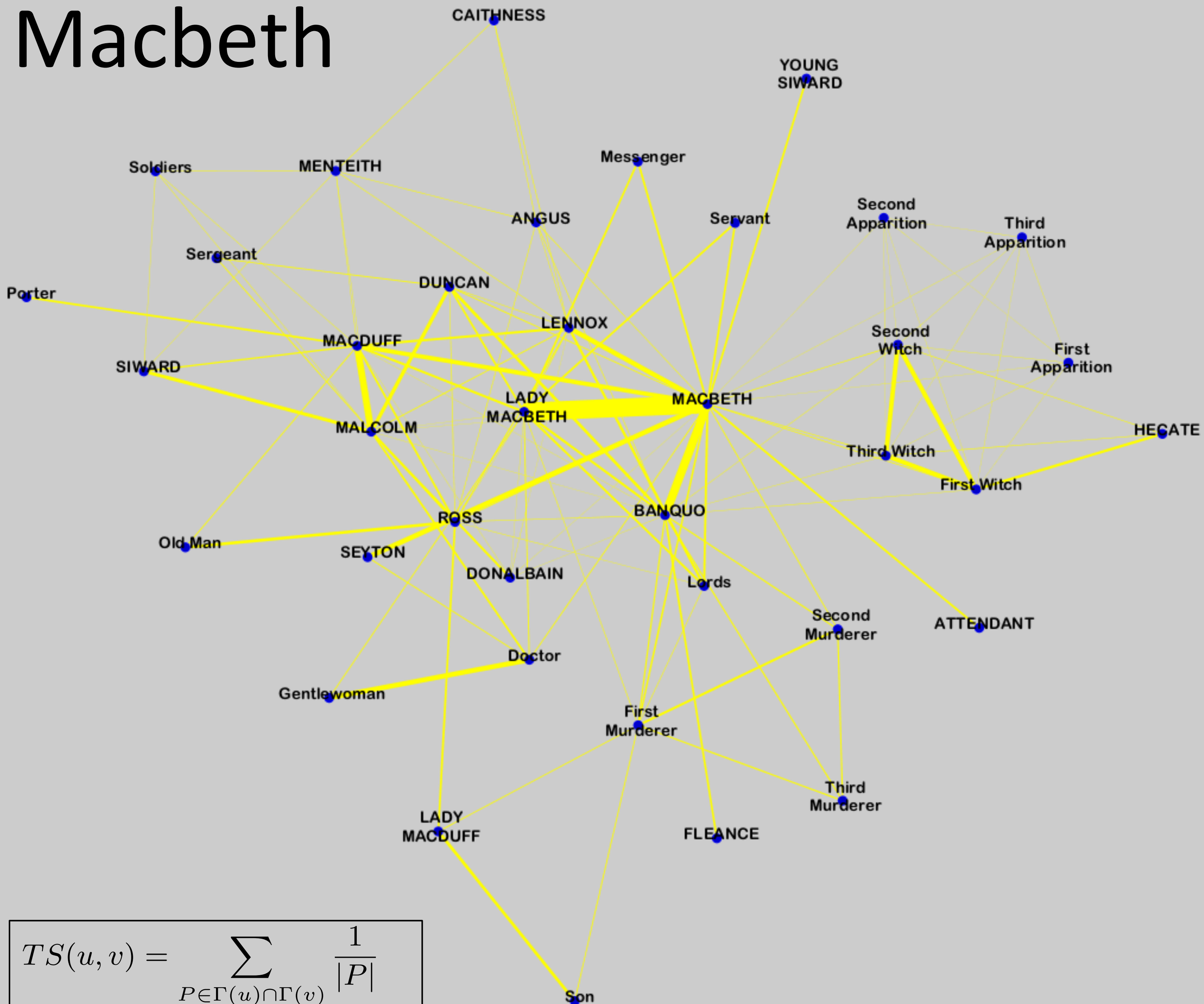
- **Challenge:** Detect which of links in the 1-mode graph are important
- **Goal:** Infer the implicit weighted social network from people's participation in mutual events

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

Macbeth



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

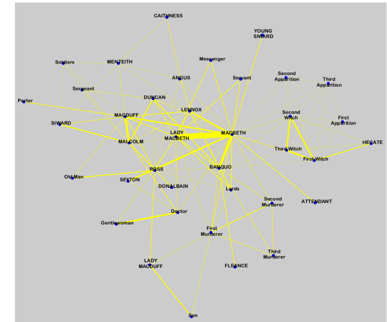
Decisions, Decisions

- 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?

Outline

- 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



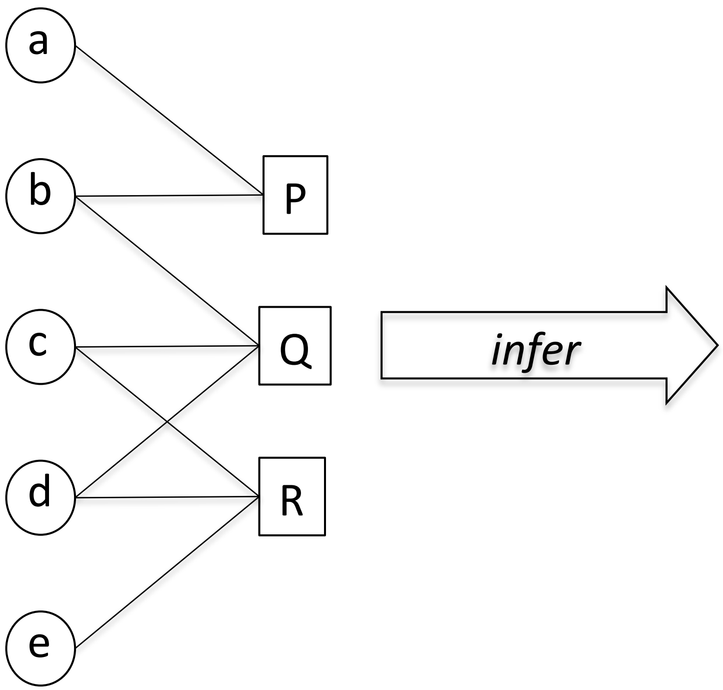
Running Example

Input

People × Event Bipartite Graph

Output

Partial Order of Tie Strength among People



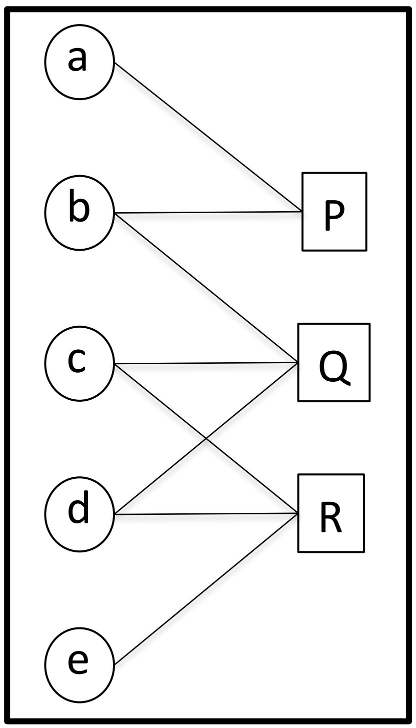
(a,b) (c,d)	*high*
(b,c), (b,d), (c,e), (d,e)	↓
(a,c), (a,d), (a,e), (b,e)	*low*

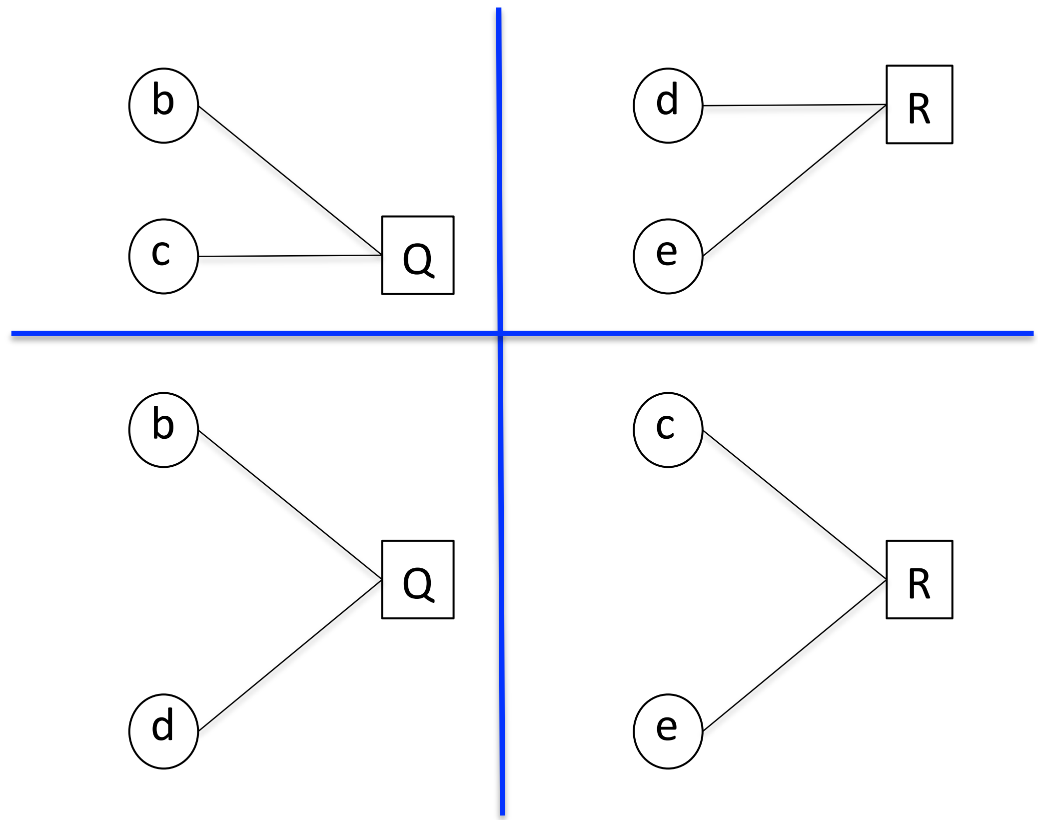
Axioms

- 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

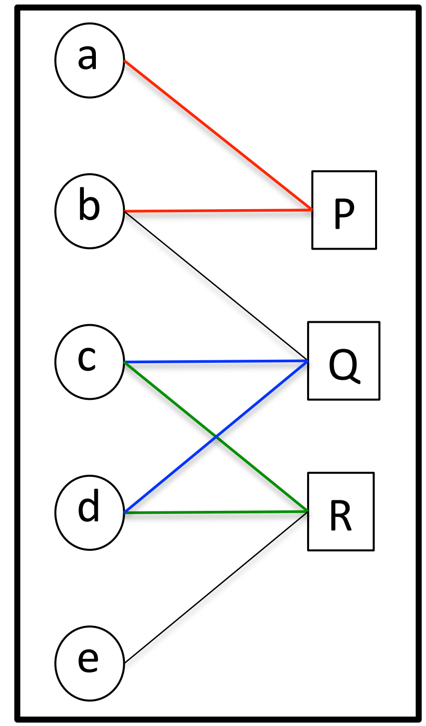
$$TS_{\emptyset}(u, v) = 0$$

- 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

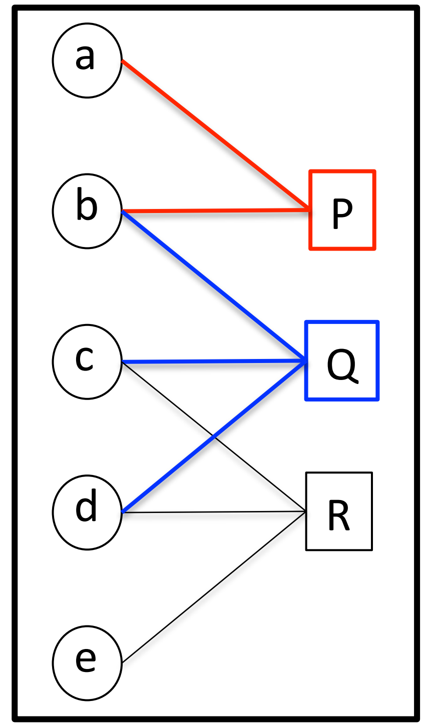
Axiom 3: Frequency & Axiom 4: Intimacy



- 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 $TS_{(G+Q)}(u, v) - TS_G(u, v) \leq TS_Q(u, v)$

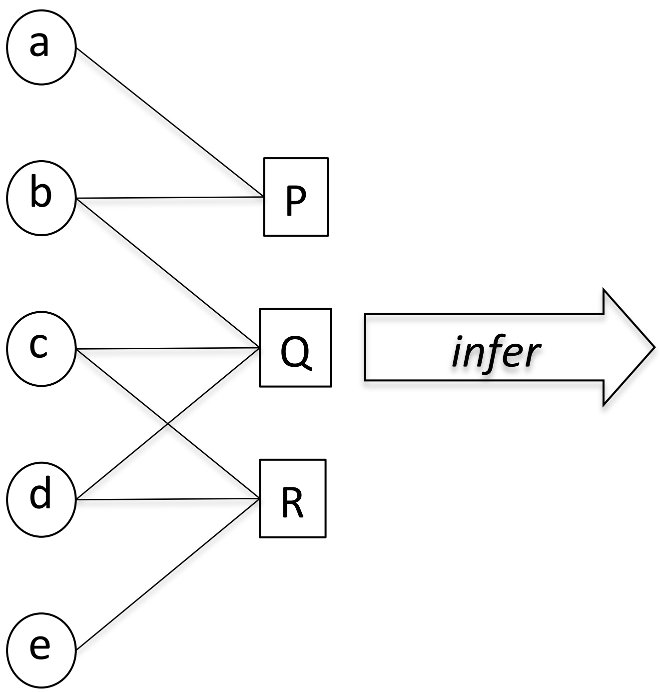
Example – Mapping to Axioms

Input

People × Event Bipartite Graph

Output

Partial order of Tie Strength



(a,b) (c,d)	*high* — *Axiom 4 (Intimacy) & Axiom 3 (Freq)*
(b,c), (b,d), (c,e), (d,e)	*Axiom 1 (Isomorphism)*
(a,c), (a,d), (a,e), (b,e)	*low* — *Axiom 2 (Baseline) & Axiom 6 (Cond. Indep. of Vertices) & Axiom 7 (Cond. Indep. of Events)*

Observations on the Axioms

- Our axioms are fairly intuitive

A1: Isomorphism	A2: Baseline	A3: Frequency	A4: Intimacy
A5: Popularity	A6: Cond. Indep. of people	A7: Cond. indep. of events	A8: Submodularity

- 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.)

A1: Isomorphism	A2: Baseline	A3: Frequency	A4: Intimacy
A5: Popularity	A6: Cond. Indep. of people	A7: Cond. indep. of events	A8: Submodularity

- 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*) = total tie strength generated in a single event with *n* 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

A way to explore the space of valid functions for representing tie strength and find which work given particular applications

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

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

- $\{P_i\}_{1 \le i \le k}$ 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

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

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

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

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

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

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

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

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

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

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

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 α, jump to a node *u*; and with probability 1-α, 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

A1: Isomorphism	A2: Baseline	A3: Frequency	A4: Intimacy
A5: Popularity	A6: Cond. indep. of P	A7: Cond. indep. of E	A8: Submodularity

	A1	A2	A3	A4	A5	A6	A7	A8	$g(a_1, \ldots, a_k)$ $h(\|P_i\|) = a_i$
Common Neighbors	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = 1$
Delta	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = 2(n(n-1))^{-1}$
Adamic & Adar	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = (\log(n))^{-1}$
Linear	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = n^{-1}$
Max	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \max\{a_i\}$ $h(n) = n^{-1}$

Measures of Tie-Strength that Do Not Satisfy All the Axioms

A1: Isomorphism	A2: Baseline	A3: Frequency	A4: Intimacy
A5: Popularity	A6: Cond. indep. of V	A7: Cond. indep. of E	A8: Submodularity

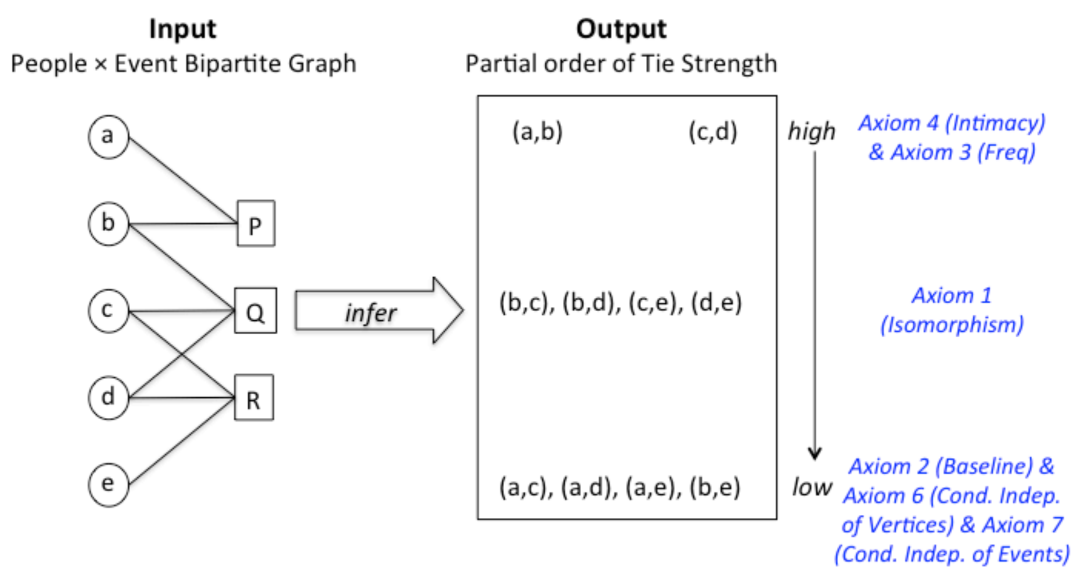
	A1	A2	A3	A4	A5	A6	A7	A8	$g(a_1, \ldots, a_k)$ $h(\lvert P_i \rvert) = a_i$
Jaccard Index	✓	✓	✓	✓	✓	✗	✗	✗	✗
Katz Measure	✓	✗	✓	✓	✓	✓	✗	✗	✗
Preferential Attachment	✓	✓	✗	✓	✓	✓	✗	✗	✗
RWR	✓	✗	✗	✗	✓	✓	✗	✗	✗
Simrank	✓	✗	✗	✗	✗	✗	✗	✗	✗
Proportional	✓	✗	✗	✓	✗	✓	✗	✗	✗

Tie Strength and Orderings

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

(1) Isomorphism	(2) Baseline	(3) Frequency	(4) Intimacy
(5) Popularity	(6) Cond. indep. of P	(7) Cond. indep. of E	(8) Submodularity

- *TS* induces a **total order** on the edges that is a linear extension of the partial order on the node-tie pairs



DETAILS

Tie Strength & Orderings

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 $L \times L$.

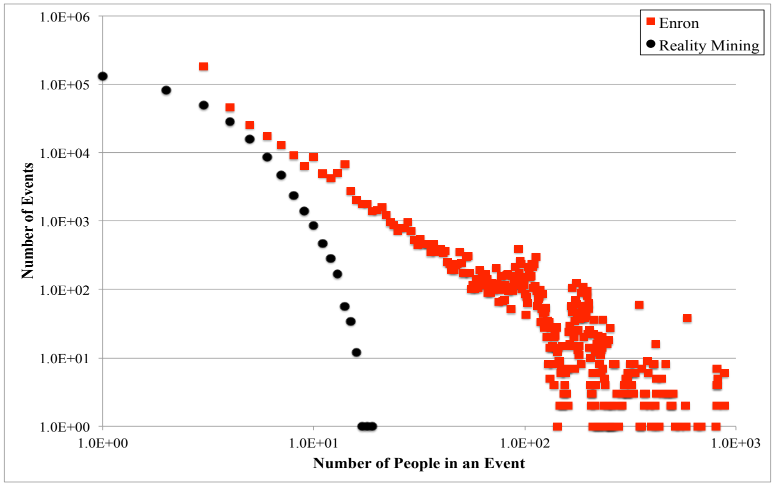
Conversely, for every linear extension $\mathcal{L}$ of the partial order $\mathcal{N}$, we can find a function TS that induces $\mathcal{L}$ on $L \times L$ and that satisfies Axioms (1-8).

Data Sets

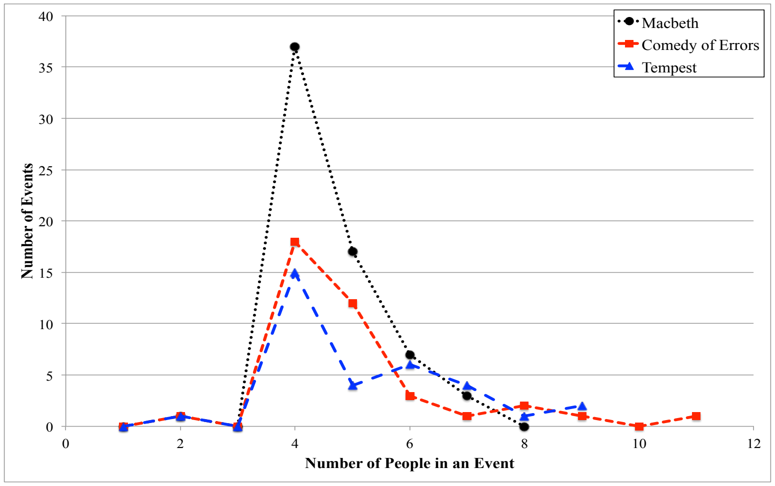
Graphs	# of People	# of Events
Southern Women	18	14
The Tempest	19	34
A Comedy of Errors	19	40
Macbeth	38	67
Reality Mining Bluetooth	104	326,248
Enron Emails	32,471	371,321

Degree Distributions

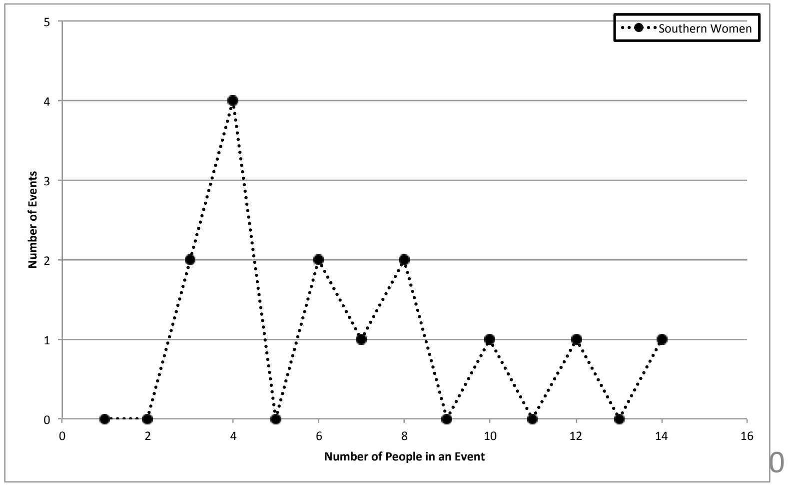
Enron & Reality Mining



Shakespeare's Plays



Southern Women



Completeness of Axioms 1-8

(Number of Ties **Not** Resolved by the Partial Order)

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 (0.23)
Reality Mining	13,794,378	1,764,546 (12.79)

- % 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

# of tie pairs = $\binom{\binom{n}{2}}{2}$

<u>Take-away point #1</u>

% 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)

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

- % 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)?

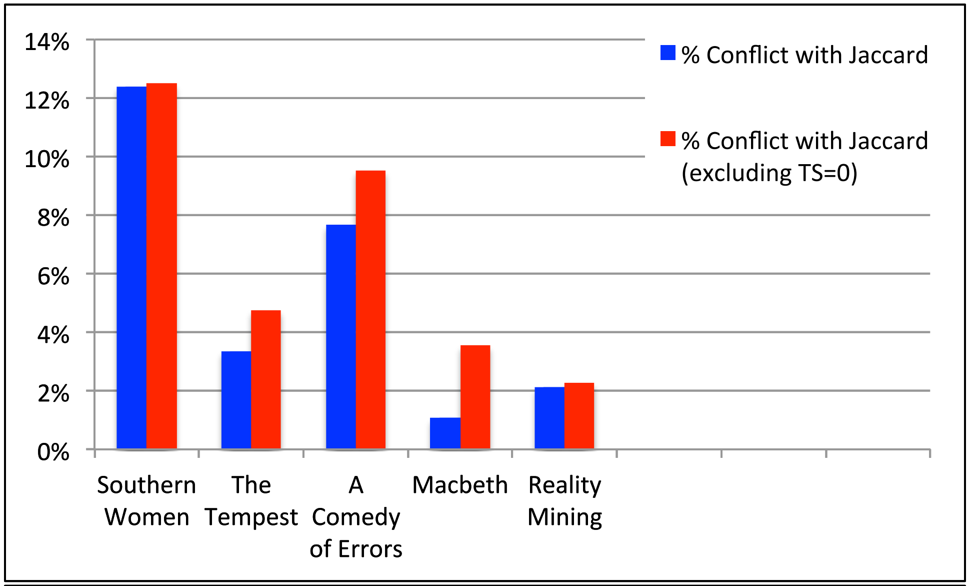
- **Answer:**
  - This is **partially true**.
  - For some pairs, the tie-strength being set to zero is caused by the axioms.
  - It may or may not be true that all these pairs have tie-strength zero in the actual function used.
    - For example, this won't be true for some self-referential functions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

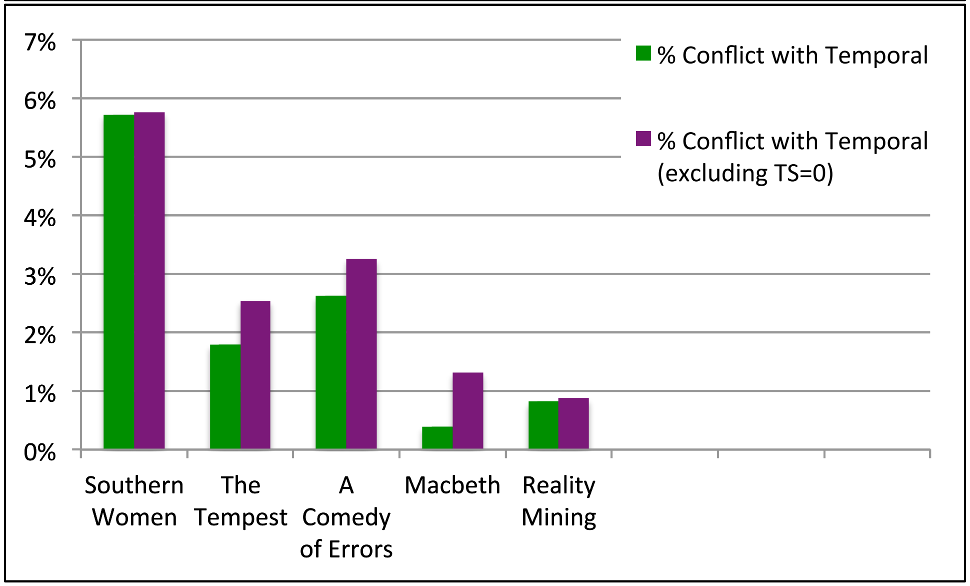
Even More on Soundness

- **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.

Dataset	Tie Pairs	Tie Pairs (excluding TS=0)	Jaccard	Temporal
Southern Women	11,628	11,537	1,441	665
The Tempest	14,535	10,257	488	261
A Comedy of Errors	14,535	11,685	1,114	381
Macbeth	246,753	74,175	2,638	978
Reality Mining	13,794,378	12,819,272	290,934	112,546

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

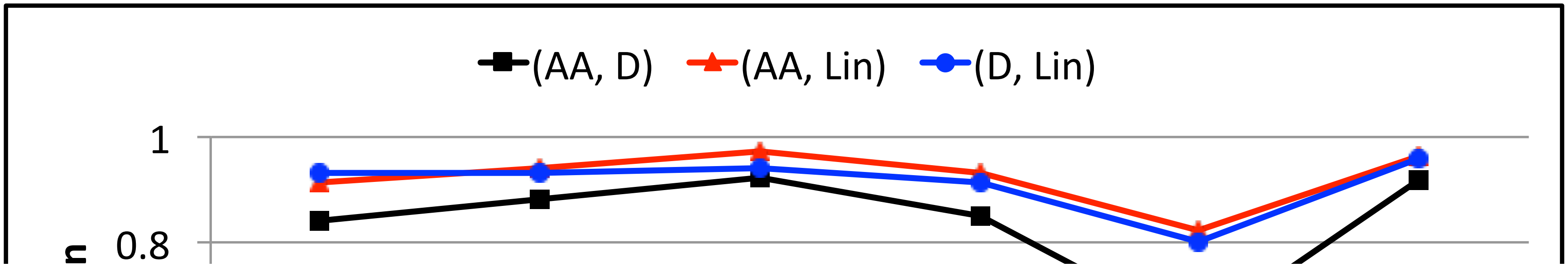
Tie Strength Measure	Formula
Common Neighbor	$TS(u,v) = \|\Gamma(u) \cap \Gamma(v)\|$
Max	$TS(u,v) = \max_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\|P\|}$
Linear	$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\|P\|}$
Delta	$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\binom{\|P\|}{2}}$
Adamic-Adar	$TS(u,v) = \sum_{P \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log \|P\|}$

Kendall τ Coefficient

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

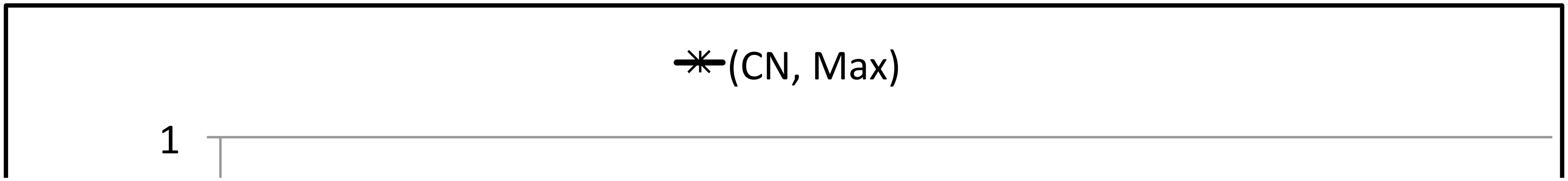
$$\tau = \frac{\text{(\# of concordant pairs) - (\# of discordant pairs)}}{\frac{1}{2}n(n-1)}$$

Adamic-Adar, Delta, & Linear produce TS rankings that are highly correlated



	A1	A2	A3	A4	A5	A6	A7	A8	$g(a_1, \ldots, a_k)$ $h(\lvert P_i \rvert) = a_i$
Common Neighbors	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = 1$
Delta	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = 2(n(n-1))^{-1}$
Adamic & Adar	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = (\log(n))^{-1}$
Linear	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = n^{-1}$
Max	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \max\{a_i\}$ $h(n) = n^{-1}$

Common Neighbor & Max produce TS rankings that are mostly uncorrelated



	A1	A2	A3	A4	A5	A6	A7	A8	$g(a_1, \ldots, a_k)$ $h(\lvert P_i \rvert) = a_i$
Common Neighbors	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = 1$
Delta	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = 2(n(n-1))^{-1}$
Adamic & Adar	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = (\log(n))^{-1}$
Linear	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \Sigma a_i$ $h(n) = n^{-1}$
Max	✓	✓	✓	✓	✓	✓	✓	✓	$g(a_1, \ldots, a_k) = \max\{a_i\}$ $h(n) = n^{-1}$

<u>Take-away point #4</u>

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

$$\binom{\binom{32471}{2}}{2} = 138,952,356,623,361,270$$

- 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

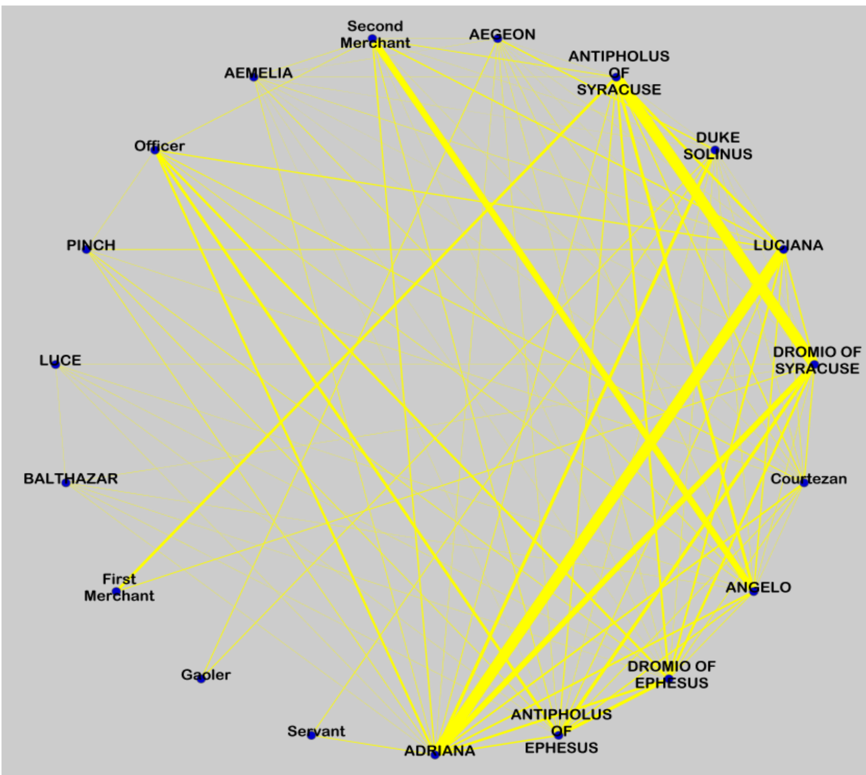
1. Presented an axiomatic approach to the problem of inferring implicit social networks by measuring tie strength
2. Characterized functions that satisfy all the axioms
3. Classified prior measures according to the axioms that they satisfy
4. Demonstrated coverage of axioms, conflict with axioms, and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

<u>Take-away point #5</u>

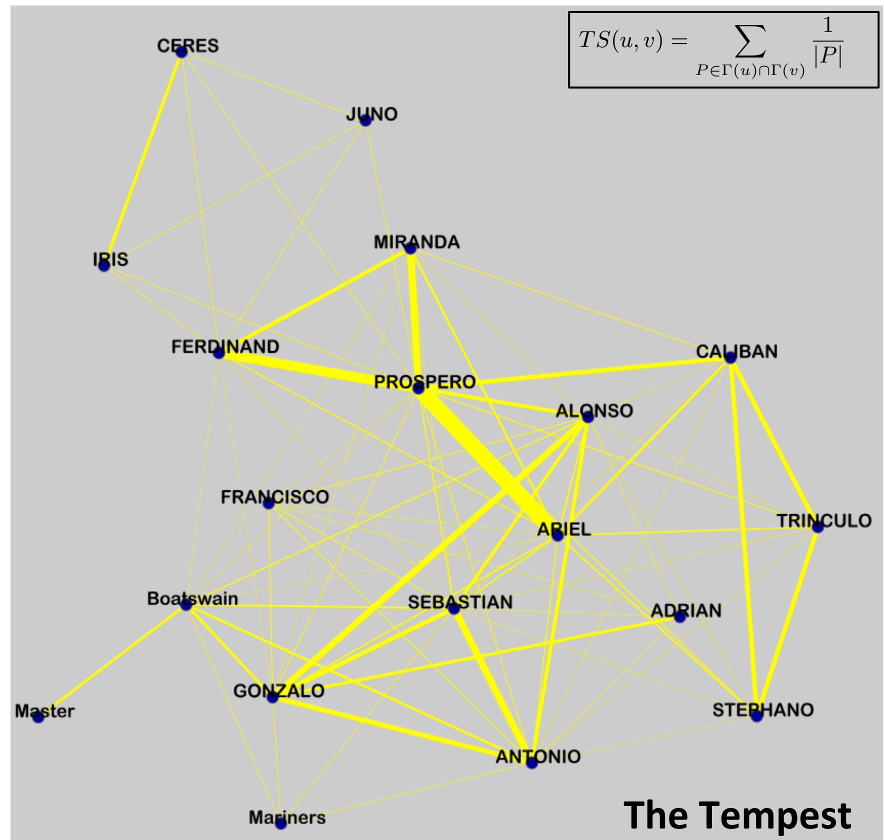
Axiomatic approaches to various measures on networks (such as tie-strength measures in this study) enable us to systematically study existing measures and characterize functions that satisfy our axioms.

Thank You!

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



A Comedy of Errors



The Tempest

Supported by LLNL & DTRA