Analytics & Predictive Models for Social Media:

Part 2: Rich Interactions

Jure Leskovec
Stanford University
Social Media: Interactions

- In Social Media users interact with one another and the content they both create and consume.
- Traditional social network analysis only distinguishes between pairs of people that are linked vs. not-linked.
- But, user interactions in social media are much richer.
Part 2 of the Tutorial: Outline

- How to learn to recommend/predict links in social networks?
- User interactions in social media:
  - Strength: strong vs. weak ties
  - Friends vs. Foes
  - Trust vs. Distrust
- How people evaluate one another and the content that is being produced by others?
Part 1: Information flow in networks

Part 2: Rich interactions
- 2.1: Recommending links in networks
- 2.2: Predicting tie strength
- 2.3: Predicting friends vs. foes
- 2.4: How do people evaluate others?
Part 1: Information flow in networks

Part 2: Rich interactions

2.1: Recommending links in networks
2.2: Predicting tie strength
2.3: Predicting friends vs. foes
2.4: How do people evaluate others?
Link prediction task:

- Given $G[t_0, t_0']$ a graph on edges up to time $t_0'$, output a ranked list $L$ of links (not in $G[t_0, t_0']$) that are predicted to appear in $G[t_1, t_1']$.

Evaluation:

- $n=|E_{new}|$: # new edges that appear during the test period $[t_1, t_1']$
- Take top $n$ elements of $L$ and count correct edges.
Predict links evolving collaboration network

### Core: Since network data is very sparse
- Consider only nodes with in-degree and out-degree of at least 3

<table>
<thead>
<tr>
<th></th>
<th>training period</th>
<th></th>
<th></th>
<th>Core</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>authors</td>
<td>papers</td>
<td>collaborations</td>
<td></td>
</tr>
<tr>
<td>astro-ph</td>
<td>5343</td>
<td>5816</td>
<td>41852</td>
<td>1561</td>
</tr>
<tr>
<td>cond-mat</td>
<td>5469</td>
<td>6700</td>
<td>19881</td>
<td>1253</td>
</tr>
<tr>
<td>gr-qc</td>
<td>2122</td>
<td>3287</td>
<td>5724</td>
<td>486</td>
</tr>
<tr>
<td>hep-ph</td>
<td>5414</td>
<td>10254</td>
<td>47806</td>
<td>1790</td>
</tr>
<tr>
<td>hep-th</td>
<td>5241</td>
<td>9498</td>
<td>15842</td>
<td>1438</td>
</tr>
</tbody>
</table>

[LibenNowell-Kleinberg '03]
Link prediction via proximity

- For every pair of nodes \((x, y)\) compute:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph distance</td>
<td>((\text{negated})) length of shortest path between (x) and (y)</td>
</tr>
<tr>
<td>Common neighbors</td>
<td>(</td>
</tr>
<tr>
<td>Jaccard’s coefficient</td>
<td>(</td>
</tr>
<tr>
<td>Adamic/Adar</td>
<td>(\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log</td>
</tr>
<tr>
<td>Preferential attachment</td>
<td>(</td>
</tr>
<tr>
<td>Katz(_\beta)</td>
<td>(\sum_{\ell=1}^{\infty} \beta^\ell \cdot</td>
</tr>
</tbody>
</table>

where \(\text{paths}^{(\ell)}_{x,y} := \{\text{paths of length exactly } \ell \text{ from } x \text{ to } y\}\)

- Sort the pairs by score and predict top \(n\) pairs as new links

\[ E^*_\text{new} := E_{\text{new}} \cap (\text{Core} \times \text{Core}) \]

\(\Gamma(x)\) … degree of node \(x\)
Link prediction via proximity

- Rank potential links \((x, y)\) based on:

\[
\begin{align*}
\text{hitting time} & \quad -H_{x,y} \\
\text{stationary-normed} & \quad -H_{x,y} \cdot \pi_y \\
\text{commute time} & \quad -(H_{x,y} + H_{y,x}) \\
\text{stationary-normed} & \quad -(H_{x,y} \cdot \pi_y + H_{y,x} \cdot \pi_x)
\end{align*}
\]

where \(H_{x,y} := \) expected time for random walk from \(x\) to reach \(y\)

\(\pi_y := \) stationary distribution weight of \(y\)

(proportion of time the random walk is at node \(y\))

---

\[ \Gamma(x) \ldots \text{degree of node } x \]
Results: Improvement over random
Results: Common neighbors

- Improvement over #common neighbors
How to learn to predict new friends in networks?

- Facebook’s People You May Know
- Let’s look at the data:
  - 92% of new friendships on FB are friend-of-a-friend
  - More common friends helps
Supervised Link Prediction

- Recommend a list of possible friends
- **Supervised machine learning setting:**
  - Training example:
    - For every node $s$ have a list of nodes she will create links to $\{v_1, \ldots, v_k\}$
  - Problem:
    - For a given node $s$ learn to rank nodes $\{v_1, \ldots, v_k\}$ higher than other nodes in the network
- **Supervised Random Walks** based on word by Agarwal & Chakrabarti
How to combine node/edge attributes and the network structure?

Learn how to often visit green nodes:

- Learn params. $w$ of a func. $f_w(u,v)$ that assigns weights to edges $(u,v)$ based on:
  - Profile of user $u$, profile of user $v$
  - Interaction history of $u$ and $v$

- Do a Personalized PageRank on this weighted graphs with a teleport set $\{s\}$

- Rank nodes by their visiting prob. $p_u$
Supervised Random Walks

- Let $s$ be the center node
- Let $f_w(u,v)$ be a function that assigns a weight to each edge:
  
  \[ a_{uv} = f_w(u,v) = \exp(-w^T\Psi_{uv}) \]
  
  - $\Psi_{uv}$ is a feature vector
    - Features of node $u$
    - Features of node $v$
    - Features of edge $(u,v)$
  
  - $w$ is the parameter vector we want to learn
- Do a Personalized PageRank from $s$ where transitions are according to edge weights
- How to learn $f_w(u,v)$?
Random walk transition matrix:

\[ Q'_{uv} = \begin{cases} \frac{a_{uv}}{\sum_w a_{uw}} & \text{if } (u, v) \in E, \\ 0 & \text{otherwise} \end{cases} \]

PageRank transition matrix:

\[ Q_{ij} = (1 - \alpha)Q'_{ij} + \alpha 1(j = s) \]

- with prob. \( \alpha \) jump back to \( s \)

Compute PageRank vector: \( p = p^T Q \)

Rank nodes by \( p_u \)
Each node $u$ has a score $p_u$

Destination nodes $D = \{v_1, \ldots, v_k\}$

No-link nodes $L = \{\text{the rest}\}$

What do we want?

$$
\min_{w} F(w) = \|w\|^2
$$

such that

$$\forall d \in D, l \in L : p_l < p_d$$

Hard constraints, make them soft
Making constraints soft

- Want to minimize:

\[ \min_w F(w) = ||w||^2 + \lambda \sum_{ld} h(p_l - p_d) \]

- **Loss:** \( h(x) = 0 \) if \( x < 0 \), \( x^2 \) else
How to minimize $F$?

$$\min_w F(w) = \|w\|^2 + \lambda \sum_{ld} h(p_l - p_d)$$

$p_l$ and $p_d$ depend on $w$

- Given $w$ assign edge weights $a_{uv} = f_w(u, v)$
- Using transition matrix $Q = [a_{uv}]$
- Compute PageRank scores $p_u$
- Rank nodes by the PageRank score

Want to set $w$ such that $p_l < p_d$
How to minimize F?

- Take the derivative!

\[
\frac{\partial F}{\partial w} = 2w + \sum_{l,d} \frac{\partial h(p_l - p_d)}{\partial w}
\]

- We know:

\[p = p^T Q\] i.e. \[p_u = \sum_j p_j Q_{ju}\]

- So:

\[
\frac{\partial p_u}{\partial w} = \sum_j Q_{ju} \frac{\partial p_j}{\partial w} + p_j \frac{\partial Q_{ju}}{\partial w}
\]

- Solve by power iteration!
To optimize $F$, use gradient based method:

- Pick a random starting point $w_0$
- Compute the personalized PageRank vector $p$
- Compute gradient with respect to weight vector $w$
- Updated $w$
  - Optimize using quasi-Newton method
Facebook Iceland network
- 174,000 nodes (55% of population)
- Avg. degree 168
- Avg. person added 26 new friends/month

For every node $s$:
- Positive examples:
  - $D=\{\text{new friendships of } s \text{ created in Nov '09}\}$
- Negative examples:
  - $L=\{\text{other nodes } s \text{ did not create new links to}\}$
- Limit to friends of friends
  - on avg. there are 20k FoFs (max 2M)!
Node and Edge features for learning:

- **Node:**
  - Age
  - Gender
  - Degree

- **Edge:**
  - Age of an edge
  - Communication,
  - Profile visits
  - Co-tagged photos

**Baselines:**

- Decision trees and logistic regression:
  - Above features + 10 network features (PageRank, common friends)

**Evaluation:**

- AUC and precision at Top20
## Results: Facebook Iceland

- **Facebook:** predict future friends
  - Adamic-Adar already works great
  - Logistic regression also strong
  - SRW gives slight improvement

<table>
<thead>
<tr>
<th>Learning Method</th>
<th>AUC</th>
<th>Prec@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Walk with Restart</td>
<td>0.81725</td>
<td>6.80</td>
</tr>
<tr>
<td>Adamic-Adar</td>
<td>0.81586</td>
<td>7.35</td>
</tr>
<tr>
<td>Common Friends</td>
<td>0.80054</td>
<td>7.35</td>
</tr>
<tr>
<td>Degree</td>
<td>0.58535</td>
<td>3.25</td>
</tr>
<tr>
<td>DT: Node features</td>
<td>0.59248</td>
<td>2.38</td>
</tr>
<tr>
<td>DT: Network features</td>
<td>0.76979</td>
<td>5.38</td>
</tr>
<tr>
<td>DT: Node+Network</td>
<td>0.76217</td>
<td>5.86</td>
</tr>
<tr>
<td>DT: Path features</td>
<td>0.62836</td>
<td>2.46</td>
</tr>
<tr>
<td>DT: All features</td>
<td>0.72986</td>
<td>5.34</td>
</tr>
<tr>
<td>LR: Node features</td>
<td>0.54134</td>
<td>1.38</td>
</tr>
<tr>
<td>LR: Network features</td>
<td>0.80560</td>
<td>7.56</td>
</tr>
<tr>
<td>LR: Node+Network</td>
<td>0.80280</td>
<td>7.56</td>
</tr>
<tr>
<td>LR: Path features</td>
<td>0.51418</td>
<td>0.74</td>
</tr>
<tr>
<td>LR: All features</td>
<td>0.81681</td>
<td>7.52</td>
</tr>
<tr>
<td>SRW: one edge type</td>
<td>0.82502</td>
<td>6.87</td>
</tr>
<tr>
<td>SRW: multiple edge types</td>
<td>0.82799</td>
<td>7.57</td>
</tr>
</tbody>
</table>
Results: Co-authorship

- Arxiv Hep-Ph collaboration network:
  - Poor performance of unsupervised methods
  - Logistic regression and decision trees don’t work to well
  - SRW gives 10% boost in Prec@20

<table>
<thead>
<tr>
<th>Learning Method</th>
<th>AUC</th>
<th>Prec@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Walk with Restart</td>
<td>0.63831</td>
<td>3.41</td>
</tr>
<tr>
<td>Adamic-Adar</td>
<td>0.60570</td>
<td>3.13</td>
</tr>
<tr>
<td>Common Friends</td>
<td>0.59370</td>
<td>3.11</td>
</tr>
<tr>
<td>Degree</td>
<td>0.56522</td>
<td>3.05</td>
</tr>
<tr>
<td>DT: Node features</td>
<td>0.60961</td>
<td>3.54</td>
</tr>
<tr>
<td>DT: Network features</td>
<td>0.59302</td>
<td>3.69</td>
</tr>
<tr>
<td>DT: Node+Network</td>
<td>0.63711</td>
<td>3.95</td>
</tr>
<tr>
<td>DT: Path features</td>
<td>0.56213</td>
<td>1.72</td>
</tr>
<tr>
<td>DT: All features</td>
<td>0.61820</td>
<td>3.77</td>
</tr>
<tr>
<td>LR: Node features</td>
<td>0.64754</td>
<td>3.19</td>
</tr>
<tr>
<td>LR: Network features</td>
<td>0.58732</td>
<td>3.27</td>
</tr>
<tr>
<td>LR: Node+Network</td>
<td>0.64644</td>
<td>3.81</td>
</tr>
<tr>
<td>LR: Path features</td>
<td>0.67237</td>
<td>2.78</td>
</tr>
<tr>
<td>LR: All features</td>
<td>0.67426</td>
<td>3.82</td>
</tr>
<tr>
<td>SRW: one edge type</td>
<td>0.69996</td>
<td>4.24</td>
</tr>
<tr>
<td>SRW: multiple edge types</td>
<td>0.71238</td>
<td>4.25</td>
</tr>
</tbody>
</table>
Part 1: Information flow in networks

Part 2: Rich interactions
- 2.1: Recommending links in networks
- 2.2: Predicting tie strength
- 2.3: Predicting friends vs. foes
- 2.4: How do people evaluate others?
Tie strength

“The strength of a tie is a (probably linear) combination of the amount of TIME, the emotional INTENSITY, the INTIMACY (mutual confiding), and the reciprocal SERVICES which characterize the tie.”

[Grannovetter]

- Gilbert & Karahalios surveyed 35 Facebook users to label 2,184 friendships (links)

- Describe each link by 70+ features
Five aspects of tie strength

[Image of Facebook profile]

- How strong is your relationship with this person?
  - barely know them
  - very close

- How would you feel asking this friend to loan you $100 or more?
  - would never ask
  - very comfortable

- How helpful would this person be if you were looking for a job?
  - no help at all
  - very helpful

- How upset would you be if this person unfriended you?
  - not upset at all
  - very upset

- If you left Facebook for another social site, how important would it be to bring this friend along?
  - would not matter
  - must bring them!
## Five aspects of tie strength

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>0–1 Scale</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>How strong is your relationship?</td>
<td></td>
<td>0.411</td>
</tr>
<tr>
<td>How comfortable asking for loan?</td>
<td></td>
<td>0.076</td>
</tr>
<tr>
<td>How helpful if looking for job?</td>
<td></td>
<td>0.362</td>
</tr>
<tr>
<td>How upset if unfriended?</td>
<td></td>
<td>0.552</td>
</tr>
<tr>
<td>How important to bring friend?</td>
<td></td>
<td>0.324</td>
</tr>
</tbody>
</table>

- Survey results for the 2,184 Facebook friendship links
Attributes of the friendship

- **Intensity**
  - wall words exchanged
  - friend-initiated wall posts
  - part.-initiated wall posts
  - inbox messages together
  - inbox thread depth
  - part.’s status updates
  - friend’s status updates

- **Intimacy**
  - participant’s friends
  - friend’s friends
  - days since last comm.
  - wall intimacy words
  - inbox intimacy words
  - together in photo
  - miles between hometowns

[Gilbert-Karahalios, '09]
Attributes of the friendship

- **Social Distance**
  - age difference
  - # occupations difference
  - educational difference
  - political difference

- **Structural**
  - mutual friends
  - groups in common
  - Cosine similarity of interests

- **Reciprocal services**
  - Links exchanged by wall
  - Applications in common

- **Emotional support**
  - Positive emotion words
  - Negative emotion words
Train a linear (regression) model

Results for the “How strong is your relationship?”

[Gilbert-Karahalios, ’09]
Results: Most predictive features

- Days since last communication: -0.762
- Days since first communication: 0.755
- Intimacy × Structural: 0.4
- Wall words exchanged: 0.299
- Mean strength of mutual friends: 0.257
- Educational difference: -0.223
- Structural × Structural: 0.195
- Reciprocal Serv. × Reciprocal Serv.: -0.19
- Participant-initiated wall posts: 0.146
- Inbox thread depth: -0.137
- Participant’s number of friends: -0.136
- Inbox positive emotion words: 0.135
- Social Distance × Structural: 0.13
- Participant’s number of apps: -0.122
- Wall intimacy words: 0.111
Part 1: Information flow in networks

Part 2: Rich interactions

2.1: Recommending links in networks
2.2: Predicting tie strength
2.3: Predicting friends vs. foes
2.4: How do people evaluate others?
So far we viewed links as **positive** but links can also be **negative**

**Question:**
- How do edge signs and network interact?
- How to model and predict edge signs?

**Applications:**
- **Friend recommendation**
  - Not just whether you know someone but what do you think of them
Each link $A \rightarrow B$ is explicitly tagged with a sign:

- **Epinions**: Trust/Distrust
  - Does $A$ trust $B$’s product reviews? (only positive links are visible)

- **Wikipedia**: Support/Oppose
  - Does $A$ support $B$ to become Wikipedia administrator?

- **Slashdot**: Friend/Foe
  - Does $A$ like $B$’s comments?
Consider edges as undirected

- **Start with intuition** [Heider ‘46]:
  - Friend of my friend is my friend
  - Enemy of enemy is my friend
  - Enemy of friend is my enemy

- Look at connected triples of nodes:

  **Balanced**
  - Consistent with “friend of a friend” or “enemy of the enemy” intuition

  **Unbalanced**
  - Inconsistent with the “friend of a friend” or “enemy of the enemy” intuition
Theory of Status

- **Status theory** [Davis-Leinhardt ‘68, Guha et al. ’04, Leskovec et al. ‘10]
  - Link \( A \rightarrow B \) means: \( B \) has higher status than \( A \)
  - Link \( A \rightarrow B \) means: \( B \) has lower status than \( A \)
  - Based on signs/directions of links from/to node \( X \) make a prediction
- Status and balance can make different predictions:

![Diagram of status and balance examples](image-url)
The Plan

- How do these two theories align with ways people create links:
  - Not just “which is right” but how are aspects of each reflected in the data
  - Provide insights into how these linking systems are being used

- Outline:
  - Study links as undirected: Balance theory
  - Study links as directed and evolving: Status theory
  - Predicting signs of edges
Undirected Links: Balance

- Consider networks as undirected
- Compare frequencies of signed triads in real and shuffled data
  - 4 triad types $t$:
    - ![Diagrams of triad types]
  - **Surprise** value for triad type $t$:
    - Number of std. deviations by which number of occurrences of triad $t$ differs from the expected number in shuffled data

Real data

Shuffled data
**Undirected Links: Balance**

- **Surprise values:**
  - *i.e.*, z-score
  - (deviation from random measured in the number of std. devs.)

<table>
<thead>
<tr>
<th>Triad</th>
<th>Epin</th>
<th>Wiki</th>
<th>Slashdot</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="+" alt="Balanced" /></td>
<td>1,881</td>
<td>380</td>
<td>927</td>
</tr>
<tr>
<td><img src="+" alt="Unbalanced" /></td>
<td>249</td>
<td>289</td>
<td>-175</td>
</tr>
<tr>
<td><img src="+" alt="Unbalanced" /></td>
<td>-2,105</td>
<td>-573</td>
<td>-824</td>
</tr>
<tr>
<td><img src="+" alt="Unbalanced" /></td>
<td>288</td>
<td>11</td>
<td>-9</td>
</tr>
</tbody>
</table>

- **Observations:**
  - Strong signal for balance
  - Epinions and Wikipedia agree on all types
  - Consistency with Davis’s ['67] weak balance
Evolving directed networks

- Our networks are really directed
  - trust, opinion (friendship)
- How many Δ are now explained by balance?
  - Half (8 out of 16)
- Is there a better explanation?
  - Yes. Theory of Status.
Alternate theory: Status

- Links are directed and created over time
- Status theory [Davis-Leinhardt ‘68, Guha et al. ’04, Leskovec et al. ‘10]
  - Link $A \rightarrow B$ means: B has higher status than A
  - Link $A \leftarrow B$ means: B has lower status than A
- Status and balance can give different predictions:

![Diagram 1](image1)

- Balance: +
- Status: –

![Diagram 2](image2)

- Balance: +
- Status: –
Evolving Directed Networks

- Links are directed
- Links are created over time
  - X has links to/from A and B
  - Now, A links to B
- To compare balance and status we need to formalize:
  - Links are embedded in triads – provides context for signs
  - Users are heterogeneous in their linking behavior
Link contexts:

- A contextualized link is a triple \((A,B;X)\) such that directed \(A-B\) link forms after there is a two-step semi-path \(A-X-B\).

- \(A-X\) and \(B-X\) links can have either direction and either sign: 16 possible types.
Different users make signs differently:
- Generative baseline (frac. of + given by A)
- Receptive baseline (frac. of + received by B)

How do different link contexts cause users to deviate from baselines?

Surprise: How much behavior of A/B deviates from baseline when they are in context
Two basic examples:

- More **negative** than gen. baseline of A
- More **negative** than rec. baseline of B

There are two diagrams showing relationships between nodes A, X, and B.
Consistency with Status

- Determine node status:
  - Assign X status 0
  - Based on signs and directions of edges set status of A and B

- Surprise is **status**-consistent, if:
  - **Gen.** surprise is status-consistent if it has same sign as status of B
  - **Rec.** surprise is status-consistent if it has the opposite sign from the status of A

- Surprise is **balance**-consistent, if:
  - If it completes a balanced triad

Status-consistent if:
Gen. surprise < 0
Rec. surprise < 0
Status vs. Balance (Epinions)

Results for Epinions: Trust vs. Distrust

<table>
<thead>
<tr>
<th>$t_i$</th>
<th>count</th>
<th>$P(+)</th>
<th>s_{out}</th>
<th>s_{in}</th>
<th>$B_{out}$</th>
<th>$B_{in}$</th>
<th>$S_{out}$</th>
<th>$S_{in}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>178,051</td>
<td>0.97</td>
<td>95.9</td>
<td>197.8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_2$</td>
<td>45,797</td>
<td>0.54</td>
<td>-151.3</td>
<td>-229.9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>○</td>
</tr>
<tr>
<td>$t_3$</td>
<td>246,371</td>
<td>0.94</td>
<td>89.9</td>
<td>195.9</td>
<td>✓</td>
<td>✓</td>
<td>○</td>
<td>✓</td>
</tr>
<tr>
<td>$t_4$</td>
<td>25,384</td>
<td>0.89</td>
<td>1.8</td>
<td>44.9</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_5$</td>
<td>45,925</td>
<td>0.30</td>
<td>18.1</td>
<td>-333.7</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_6$</td>
<td>11,215</td>
<td>0.23</td>
<td>-15.5</td>
<td>-193.6</td>
<td>○</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_7$</td>
<td>36,184</td>
<td>0.14</td>
<td>-53.1</td>
<td>-357.3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_8$</td>
<td>61,519</td>
<td>0.63</td>
<td>124.1</td>
<td>-225.6</td>
<td>✓</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_9$</td>
<td>338,238</td>
<td>0.82</td>
<td>207.0</td>
<td>-239.5</td>
<td>✓</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_{10}$</td>
<td>27,089</td>
<td>0.20</td>
<td>-110.7</td>
<td>-449.6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_{11}$</td>
<td>35,093</td>
<td>0.53</td>
<td>-7.4</td>
<td>-260.1</td>
<td>○</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_{12}$</td>
<td>20,933</td>
<td>0.71</td>
<td>17.2</td>
<td>-113.4</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_{13}$</td>
<td>14,305</td>
<td>0.79</td>
<td>23.5</td>
<td>24.0</td>
<td>○</td>
<td>○</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$t_{14}$</td>
<td>30,235</td>
<td>0.69</td>
<td>-12.8</td>
<td>-53.6</td>
<td>○</td>
<td>○</td>
<td>✓</td>
<td>○</td>
</tr>
<tr>
<td>$t_{15}$</td>
<td>17,189</td>
<td>0.76</td>
<td>6.4</td>
<td>24.0</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>✓</td>
</tr>
<tr>
<td>$t_{16}$</td>
<td>4,133</td>
<td>0.77</td>
<td>11.9</td>
<td>-2.6</td>
<td>✓</td>
<td>○</td>
<td>✓</td>
<td>○</td>
</tr>
</tbody>
</table>

Number of correct predictions: 8 7 14 13
Status seems to fail on triad types where A and B have both low status relative to X

- Mistakes for gen. surprise (a) and (b) are “duals” of mistakes for receptive surprise (c) and (d)
- Type (a) is one of most basic cases of balance theory
Global Structure of Signed Nets

- Intuitive picture of social network in terms of densely linked clusters
- How does structure interact with links?
- Embeddedness of link (A, B): number of hared neighbors
Global structure: Embeddedness

- Embeddedness of ties:
  - Embedded ties tend to be more positive

- A natural connection to closure based social capital [Coleman ‘88]

- Public display of signs (votes) in Wikipedia further strengthens this
Edge sign prediction problem
- Given a network and signs on all but one edge, predict the missing sign

Machine Learning formulation:
- Predict sign of edge (u,v)
- Class label:
  - +1: positive edge
  - -1: negative edge
- Learning method:
  - Logistic regression

\[
P(+|x) = \frac{1}{1 + e^{-(b_0 + \sum_i^n b_i x_i)}}
\]

Dataset:
- Original: 80% +edges
- Balanced: 50% +edges

Evaluation:
- Accuracy and ROC curves

Features for learning:
- Next slide
For each edge \((u,v)\) create features:

- **Triad counts (16):**
  - Counts of signed triads edge \(u \rightarrow v\) takes part in

- **Degree (7 features):**
  - Signed degree:
    - \(d_{out}^+(u), \ d_{out}^-(u), \ d_{in}^+(v), \ d_{in}^-(v)\)
  - Total degree:
    - \(d_{out}(u), \ d_{in}(v)\)
  - Embeddedness of edge \((u,v)\)
Edge sign prediction

- Error rates:
  - Epinions: 6.5%
  - Slashdot: 6.6%
  - Wikipedia: 19%

- Signs can be modeled from local network structure alone
  - Trust propagation model of [Guha et al. ‘04] has 14% error on Epinions

- Triad features perform less well for less embedded edges

- Wikipedia is harder to model:
  - Votes are publicly visible
Do people use these very different linking systems by obeying the same principles?

- How generalizable are the results across the datasets?
  - Train on row “dataset”, predict on “column”

<table>
<thead>
<tr>
<th></th>
<th>Epinions</th>
<th>Slashdot</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epinions</td>
<td>0.9342</td>
<td>0.9289</td>
<td>0.7722</td>
</tr>
<tr>
<td>Slashdot</td>
<td>0.9249</td>
<td>0.9351</td>
<td>0.7717</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>0.9272</td>
<td>0.9260</td>
<td>0.8021</td>
</tr>
</tbody>
</table>

Almost perfect generalization of the models even though networks come from very different applications.
Balance and Status: Complete model

<table>
<thead>
<tr>
<th>Feature</th>
<th>Bal</th>
<th>Stat</th>
<th>Epin</th>
<th>Slashd</th>
<th>Wikip</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>1</td>
<td>1</td>
<td>-0.2</td>
<td>0.02</td>
<td>-0.2</td>
</tr>
<tr>
<td>●+→●+→●</td>
<td>-1</td>
<td>0</td>
<td>-0.5</td>
<td>-0.9</td>
<td>-0.4</td>
</tr>
<tr>
<td>●+→●-→●</td>
<td>-1</td>
<td>0</td>
<td>-0.4</td>
<td>-1.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>●-→●+→●</td>
<td>1</td>
<td>-1</td>
<td>-0.7</td>
<td>-0.6</td>
<td>-0.8</td>
</tr>
<tr>
<td>●+→●+→●</td>
<td>1</td>
<td>0</td>
<td>0.3</td>
<td>0.4</td>
<td>0.05</td>
</tr>
<tr>
<td>●+→●-→●</td>
<td>-1</td>
<td>1</td>
<td>-0.01</td>
<td>-0.1</td>
<td>-0.01</td>
</tr>
<tr>
<td>●-→●+→●</td>
<td>-1</td>
<td>-1</td>
<td>-0.9</td>
<td>-1.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>●-→●-→●</td>
<td>1</td>
<td>0</td>
<td>0.04</td>
<td>-0.07</td>
<td>-0.03</td>
</tr>
<tr>
<td>●+→●+→●</td>
<td>1</td>
<td>0</td>
<td>0.08</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>●+→●-→●</td>
<td>-1</td>
<td>-1</td>
<td>-1.3</td>
<td>-1.1</td>
<td>-0.4</td>
</tr>
<tr>
<td>●-→●+→●</td>
<td>-1</td>
<td>1</td>
<td>-0.1</td>
<td>-0.2</td>
<td>0.05</td>
</tr>
<tr>
<td>●+→●+→●</td>
<td>1</td>
<td>0</td>
<td>0.08</td>
<td>-0.02</td>
<td>-0.1</td>
</tr>
<tr>
<td>●+→●-→●</td>
<td>1</td>
<td>-1</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.01</td>
</tr>
<tr>
<td>●-→●+→●</td>
<td>-1</td>
<td>0</td>
<td>-0.05</td>
<td>-0.3</td>
<td>-0.02</td>
</tr>
<tr>
<td>●+→●-→●</td>
<td>-1</td>
<td>0</td>
<td>-0.04</td>
<td>-0.3</td>
<td>0.05</td>
</tr>
<tr>
<td>●-→●-→●</td>
<td>1</td>
<td>1</td>
<td>-0.02</td>
<td>0.2</td>
<td>-0.2</td>
</tr>
</tbody>
</table>
Both theories agree well with learned models

Further observations:
- Backward-backward triads have smaller weights than forward and mixed direction triads
- Balance is in better agreement with Epinions and Slashdot while Status is with Wikipedia
- Balance consistently disagrees with “enemy of my enemy is my friend”
Part 1: Information flow in networks

Part 2: Rich interactions

2.1: Recommending links in networks
2.2: Predicting tie strength
2.3: Predicting friends vs. foes
2.4: How do people evaluate others?
Recommendations & Opinions

- What do people think about our recommendations and opinions?

Amazon.com for Dummies (Paperback)
by Nora Friedman (Author) "No one (except maybe Amazon.com founder Jeff Bezos) ever imagined that one day there would be a way that you could buy everything from books..." (more)

Available from these sellers.

12 new from $3.13 15 used from $2.93

4 of 14 people found the following review helpful:

problems with navigating amazon.com?, November 18, 2005
By Gary Kuhlman "speedk0re" (Irvine, CA USA) - See all my reviews

ok so i've never read this book, but if you need a book to navigate amazon.com, then you should just give me your money instead. i mean, i know it's hard to type a word and press enter, and then press buy; i think the real difficulty of amazon.com is how the author managed to write XXX pages about navigating amazon.com. having said that, it almost makes me want to buy this book, so i'm changing my 1 Star to 2.

Help other customers find the most helpful reviews | Report this | Permalink
Was this review helpful to you? Yes No

[Danescu et al., 2009]
People find conforming opinions more helpful
Positive reviews are more helpful
As variance of star ratings increases the two camps emerge.

It is not about review quality but whether we agree with it or not.
So far we examined the aggregate behavior:
- We observed what kinds of reviews the population as a whole find helpful

What about behavior of individuals?

Voting for admin promotion on Wikipedia
- Vote is public and consequential
  - Do you support B to become Wikipedia administrator?
- Data:
  - 2,800 elections between 2004-08
    - 44% resulted in promotion
  - 8,200 users
  - 114,000 votes (evaluations)
    - 80% of votes are positive
Voter (evaluator) V evaluates candidate C

Prob. of positive evaluation of V as a function of status difference: $S_V - S_C$

Hypothesis: Monotonically decreases

```
P(support vote)          
\[\begin{array}{c}
(S_V < S_C) & 10 \\
(S_V = S_C) & 0 \\
(S_V > S_C) & -10 \\
\end{array}\]
```
Effects of Status

- Prob. of positive evaluation of V as a function of status difference: $S_V - S_C$
  - Two notions of status:
    - Number of edits of a user
    - Number of barnstars

- Observations:
  - V is especially negative when status equals: $S_V = S_C$
  - “Mercy bounce” for $S_V > S_C$
Prob. of positive evaluation as a function of prior interaction of V and C

Observation:
Prior interaction increases prob. of a positive evaluation
Aggregate response function:

- How does the probability of a positive evaluation depend on the fraction of positive evaluations so far?

![Graphs showing different response functions: Linear, Diminishing returns, Threshold.](image)
Thresholds and diversity of voters

- **Aggregate response function:**
  - Baseline: If voter were to flip a coin then $f(x) = x$

- **Observation:**
  - Voters more inclined to express opinion when it goes against the prevailing opinion
  - Consistent with [Wu-Huberman]
Thresholds and diversity of voters

- **Personal response functions:**
  - How does prob. of voter V positively evaluating C depend on frac. of positive evaluations so far?
  - Enough data that we can build models of individuals

![Graph showing probability of voting positively vs. fraction of support votes at time of vote](image)

11 users that took part in >400 elections
- Personal response functions:
  - Average is close to baseline but individual variation in shape of response function is large

28 users that took part in >300 elections
Personal response functions:

- Over time voters become more conservative
  - Response functions shift down and left

78 users that took part in >200 elections
Social media sites are governed by (often implicit) user evaluations.

Wikipedia voting process has an explicit, public and recorded process of evaluation.

Main characteristics:

- Importance of relative assessment: Status
- Importance of prior interaction: Similarity
- Diversity of individuals’ response functions

How to explain low aggregate evaluations given by users to others of similar status?
New Setting: Stack Overflow

- Stack Overflow:
  - Q&A website for questions about programming
  - Users can up/down vote others questions/answers
  - Questions are annotated with tags describing relevant topics
- Complete data:
  - 1.1M questions, 3.2M answers, and 7.5M votes (1.8M on questions, 5.7M on answers)
  - 93.4% of the votes are positive
For a fixed status difference curves are flat

- Prob. of positive evaluation does not depend on target status

- Prob. of positive eval. increases in similarity both in topic (tag) and social circle (eval. similarity)
Evaluators with lower status than the target tend to be more similar than when they have higher status.
How to explain low aggregate evaluations given by users to others of similar status?

- This cannot be due to colleagues being overly tough on each other since the similarity tends to increase the positivity of evaluations, not decrease it.

Two possible explanations:

- (A) Low-status targets tend to garner evaluations that depend mainly on the target’s absolute status
- (B) Evaluations of higher-status targets are functions of the differential status between the pair
Remember:

- Evaluation behavior is different for low status targets
- Most evaluators have low status

Model (explains StackOverflow):

- Evaluators do not discriminate among high-status targets
  - For high-status (>\(T_{\text{min}}\)) target vote + with const prob
- But for low status targets they are increasingly negative as the target’s status decrease
  - Vote + with prob. \(\propto \frac{T}{T_{\text{min}}}\)
In Wikipedia the dip at zero persists even when removing low status targets.

Model ingredients:

- High similarity between evaluator and target yields more positive votes.
- Selection bias effect: Higher-status evaluators are more similar to the target than lower-status evaluators are.
Based on only who showed to up to evaluate predict the outcome of the Wiki election

<table>
<thead>
<tr>
<th>Number of votes</th>
<th>$E$</th>
<th>Relative gain over LogReg</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>71.4%</td>
<td>12.8%</td>
</tr>
<tr>
<td>10</td>
<td>75.0%</td>
<td>23.8%</td>
</tr>
<tr>
<td>all</td>
<td>75.6%</td>
<td>25.7%</td>
</tr>
</tbody>
</table>

**Method:**
- Divide the Status-Similarity space, each cell prob. + vote

**Baseline:**
- Guessing gives 50% accuracy
- Logistic Regression based on the target status (67% acc)
Conclusions: Status and Similarity

- Going beyond unlabeled edges
- Networks are globally organized based on status
- People use signed edges consistently regardless of a particular application
  - Near perfect generalization across datasets
- Assessing tie strength
- Effects of Status and Similarity
  - Implicit user evaluations
Tutorial Outline

- Part 1: Information flow in networks
- Part 2: Rich interactions
- Conclusion and reflections
Messages arriving through networks from real-time sources requires new ways of thinking about information dynamics and consumption

“Tell me about X” vs. “Tell me what I need to know now”
Diffusion of Topics and Sentiment
  - How news cascade through on-line networks
  - Do we need new notions of rank/importance?

Incentives and Diffusion
  - Using diffusion in the design of on-line systems
  - Connections to game theory

When will one cascade overtake the other?
Part 1: Opportunities

- A number of novel opportunities:
  - Predictive modeling of the spread of new ideas and behaviors
  - Opportunity to design systems that make use of diffusion process

- Applications:
  - Search
  - Real-time search
  - Social search
Part 2: Rich Interactions

- Links are more than just links
  - Strengths
  - Sentiment
  - They reveal what we think of others

- Main characteristics:
  - Importance of relative assessment: Status
  - Importance of prior interaction: Similarity
Part 2: Connections

- Don’t predict just who we link to but also what we think of them
- Evaluations range from evaluating a person to the content they produced
- Different dimensions of the evaluation:
  - Is the content technically correct?
  - Do I agree/disagree with the answer?
Part 2: Opportunities

- Composition of an audience can tell us something about the audience’s reaction
  - Predict outcomes simply from the statuses and similarities of the users who show up to provide evaluations, without ever seeing the values of the evaluations themselves
  - Connections to collaborative filtering

- Design reputation systems that account for status and similarity and encourage interaction