How Users Evaluate Each Other in Social Media

Jure Leskovec
Stanford University
Including joint work with Ashton Anderson, Dan Huttenlocher, Dan Jurafsky, Jon Kleinberg, and Julian McAuley

Recommendations

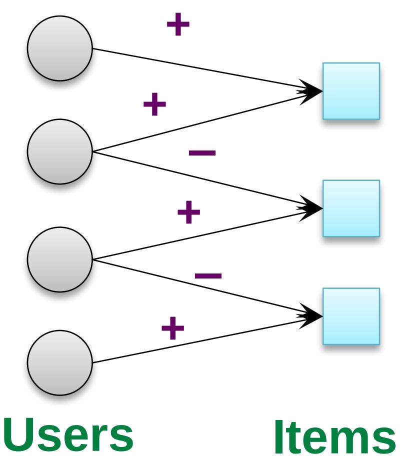
- Recommender Systems drive the Web!
- Anything can be recommended:
  - Advertising messages
  - Investment choices
  - Restaurants
  - News articles
  - Music tracks
  - Movies
  - TV programs
  - Books
  - Clothes
  - Tags
  - Future friends (Social network sites)
  - Courses in e-learning
  - Online mates (Dating)
  - Supermarket goods
  - Drug components
  - Research papers
  - Citations
  - Code modules
  - Programmers

Evaluations Drive RecSys

- **Success of recommender systems heavily depends on people expressing their attitudes and opinions**
  - Through **consumption**:
    - Buying
    - Clicking
  - Through **actions**:
    - Rating a product
    - Pressing a “like” button
  - Through **text**:
    - Writing a comment, a review

A Common View

- The most common and traditional form of evaluations: **Users evaluate items**
  - Movies, books, music, products, …



- **Traditional view of Recommender Systems:** Systems then attempt to predict how much you may like a certain product
  - **Collaborative filtering** [Resnick et al. '94]
  - **Latent space models** [Koren-Bell-Volinsky '09]

The Social Transformation

Online friendships
[Ugander-Karrer-Backstrom-Marlow, '11]

CouchSurfing activity
[Lauterbach-Truong-Shah-Adamic, '09]

- **Social Transformation of Computing**
  - Technological networks intertwined with social
- **Profound transformation in:**
  - How information is produced and shared
  - How people interact and communicate
  - The scope of CS as a discipline

Social Recommendations

- **A different view of Recommender Systems:** Systems that help people find information that will interest them, by facilitating social and conceptual connections
- **Recommendations in online communities**
- **In communities people express opinions:**
  - **About other community members**
  - **About content created by other members of the community**

User-User Evaluations

- **Many on-line settings where one person expresses an opinion about another**
  (or about another's content)
  - **I trust you** [Kamvar-Schlosser-Garcia-Molina '03]
  - **I agree with you** [Adamic-Glance '04]
  - **I vote in favor of admitting you into the community** [Cosley et al. '05, Burke-Kraut '08]
  - **I find your answer/opinion helpful**
    [Danescu-Niculescu-Mizil et al. '09, Borgs-Chayes-Kalai- Malekian-Tennenholtz '10]

Natural Analogies

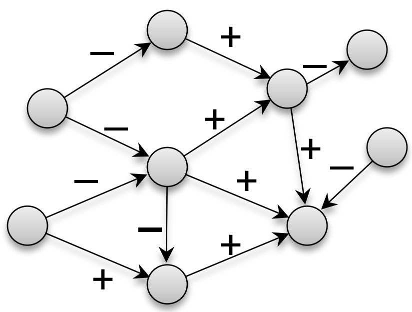
- **Natural analogies to how evaluation works in scientific communities:**
  - Acceptance of papers to conferences and journals
  - Funding of grant proposals
  - Who gets hired, who receives awards, …

U-U Evaluations: Some Issues

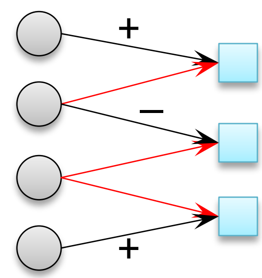
- **Need to understand ways in which humans evaluate each other**
  - What factors play role?
  - What biases arise?
- **New forms of evaluations & feedback**
  - Allowing for interactions between users
  - Computing composite opinion of a community
  - Using audience composition as a way to extract (implicit) evaluations

This Talk: Setting

- **People evaluate each other:**



Direct

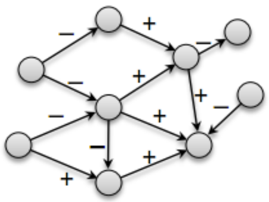


Indirect

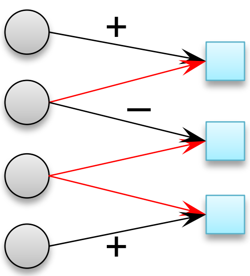
  - **Direct:** User to user [ICWSM '10]
  - **Indirect:** User to content (created by another member of a community) [WSDM '12]
- **Where online does this explicitly occur on a large scale?**

This Talk: Data

- **Wikipedia adminship elections**
  - Support/Oppose (120k votes in English)
  - 4 languages: EN, GER, FR, SP



- **Stack Overflow Q&A community**
  - Upvote/Downvote (7.5M votes)



- **Epinions product reviews**
  - Ratings of others' product reviews (13M)
    - 5 = positive, 1-4 = negative

This Talk: Overview

- **Questions:**
  - **1) Factors:** What ingredients/factors lead people when they evaluate each other?
  - **2) Synthesis:** How do we create a composite description that accurately reflects cumulative opinion of the community?
  - **3) Implicit feedback:** How to use audience composition as a way to extract evaluations?

Human Evaluations

- **What drives human evaluations?**

- How do properties of **evaluator A** and **target B** affect A's vote?
  - **Status** and **Similarity** are two fundamental drivers behind human evaluations

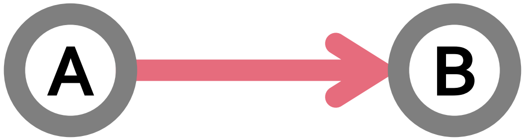
Definitions

[WSDM '12]

- **Status**
  - Level of recognition, merit, achievement, reputation in the community
    - Wikipedia: # edits, # barnstars
    - Stack Overflow: # answers
- **User-user Similarity**
  - Overlapping topical interests of A and B
    - Wikipedia: Cosine similarity of the articles edited
    - Stack Overflow: Cosine similarity of users evaluated

Relative vs. Absolute Assessment

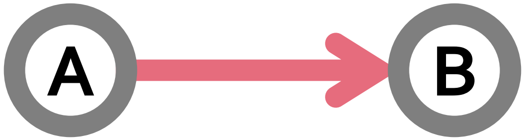
- How do properties of **evaluator A** and **target B** affect A's vote?



- Two natural (but competing) hypotheses:
  - **(1)** Prob. that B receives a positive evaluation depends primarily on the characteristics of B
    - There is some objective criteria for user B to receive a positive evaluation

Relative vs. Absolute Assessment

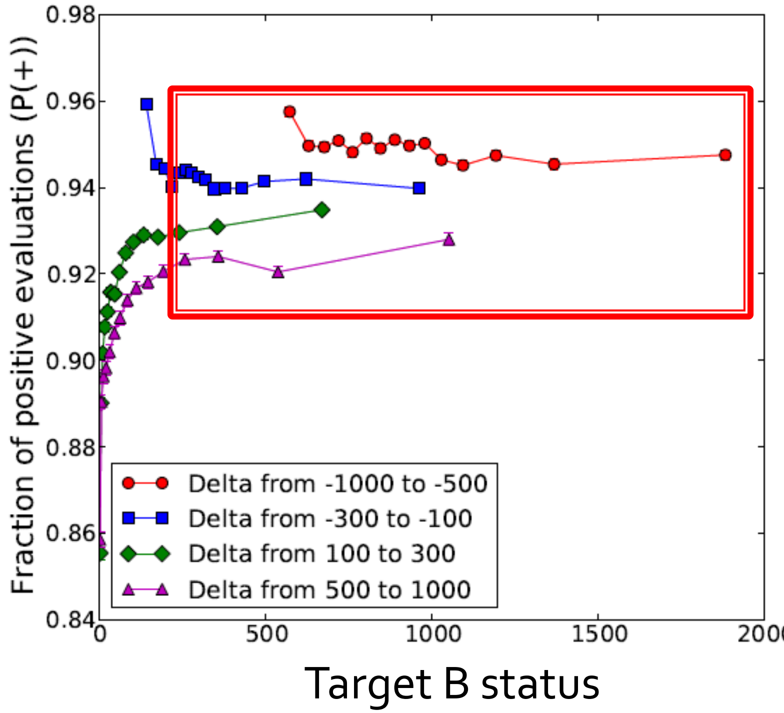
- How do properties of **evaluator A** and **target B** affect A's vote?



- Two natural (but competing) hypotheses:
  - **(2)** Prob. that B receives a positive evaluation depends on relationship between the characteristics of A and B
    - User A compares herself to user B and then makes the evaluation

Effects of Status

- **How does status of B affect A's evaluation?**
  - Each curve is fixed status difference: $\Delta = S_A - S_B$
- **Observations:**
  - **Flat curves:** Prob. of positive eval. P(+) doesn't depend on B's status
  - **Different levels:** Different values of Δ result in different behavior

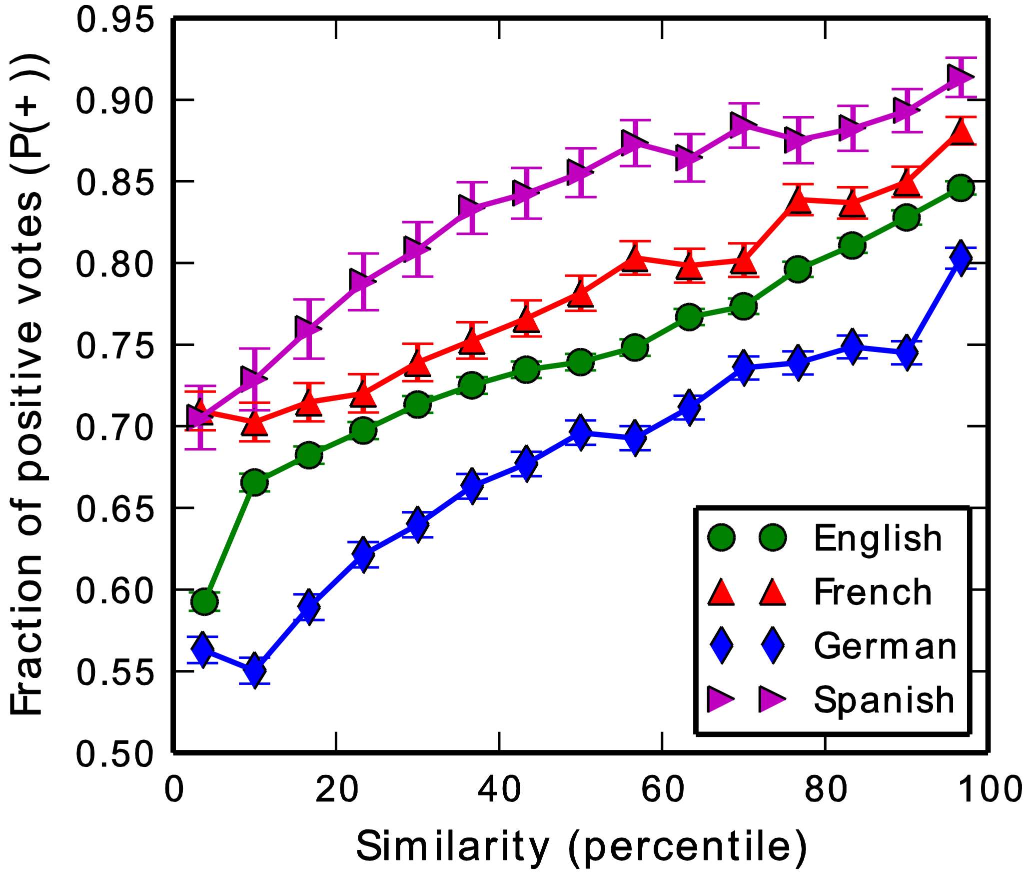


Status **difference** remains salient even as A and B acquire more status

Effects of Similarity

- How does prior interaction shape evaluations? 2 hypotheses:
  - (1) Evaluators are more supportive of targets in their area
    - “The more similar you are, the more I like you”
  - (2) More familiar evaluators know weaknesses and are more harsh
    - “The more similar you are, the better I can understand your weaknesses”

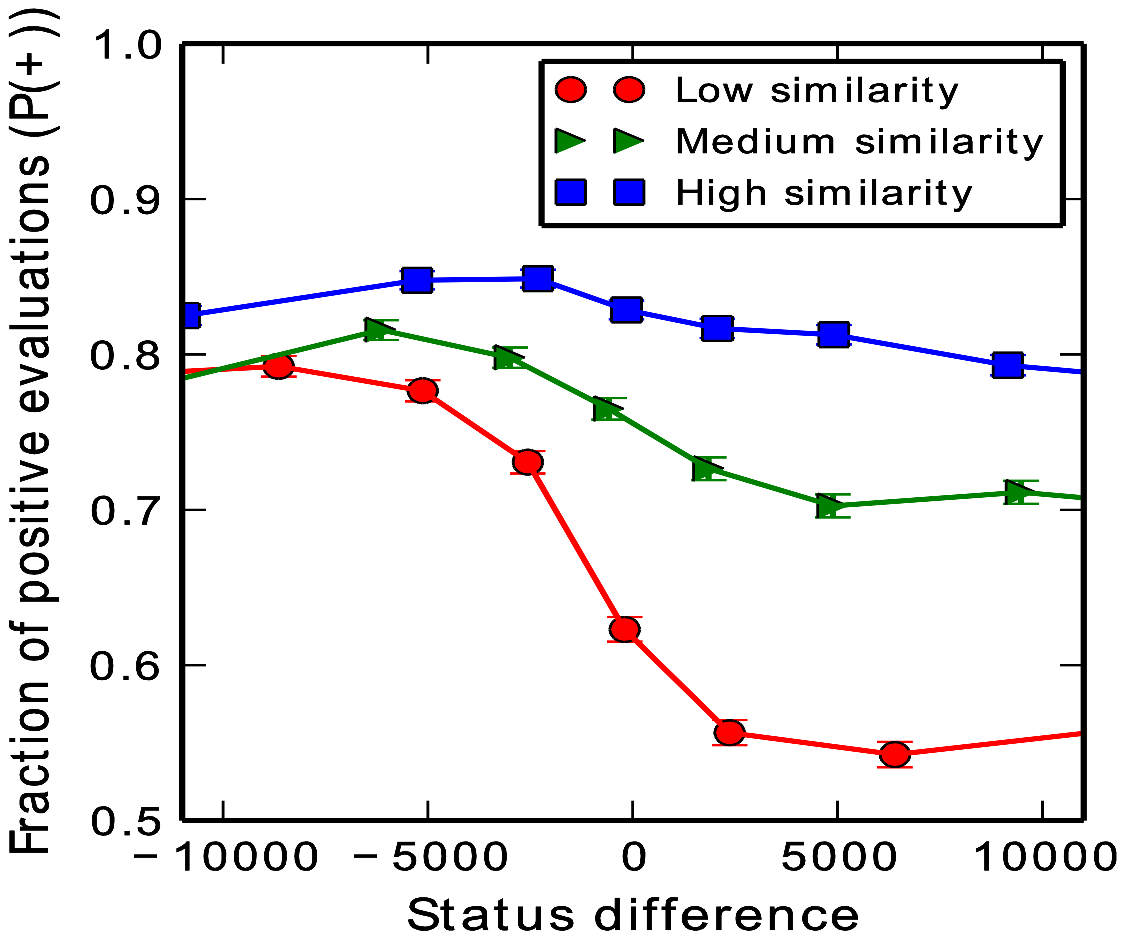
Effects of Similarity



Prior interaction/ similarity boosts positive evaluations

[WSDM '12]

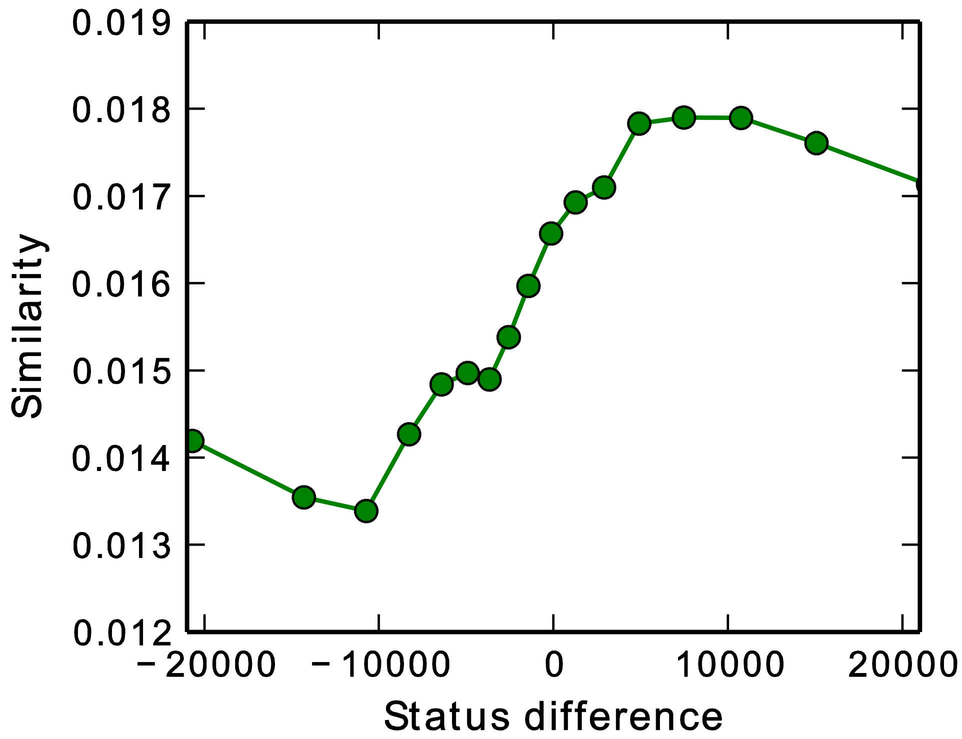
Status & Similarity



Status is a proxy for quality when evaluator does not know the target

Status & Similarity

- **Who shows up to evaluate?**

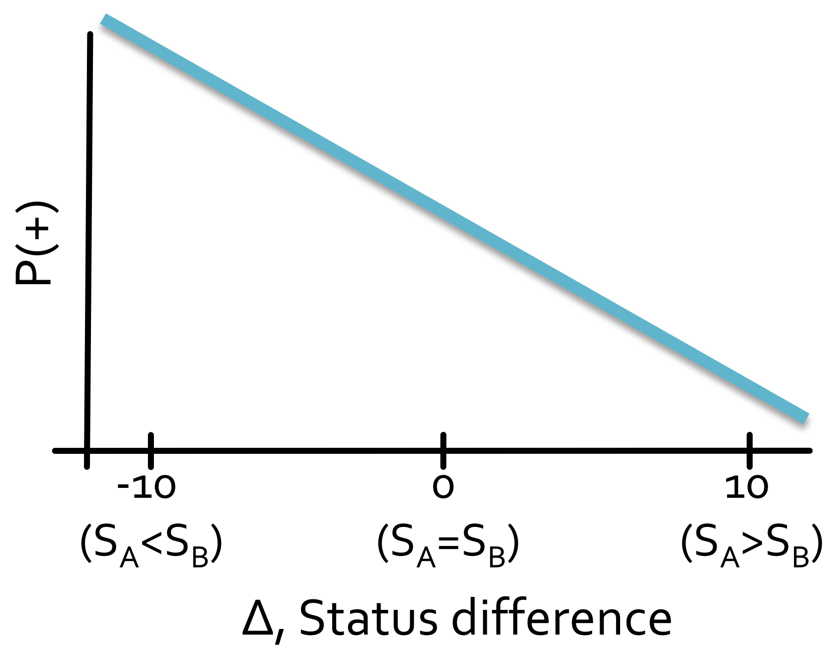


Elite evaluators vote on targets in their area of expertise

- Selection effect in who gives the evaluation
  - If $S_A > S_B$ then A and B are highly similar

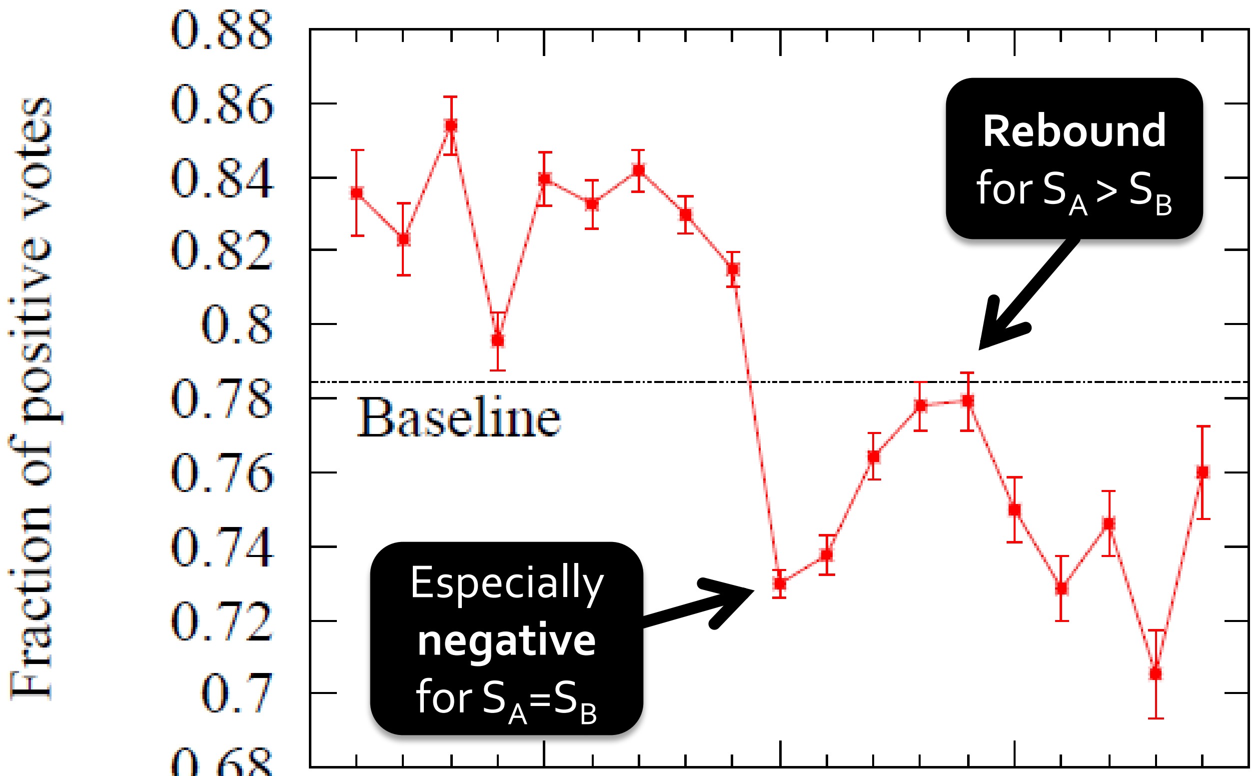
A Puzzle

- **What is P(+) as a function of ? = S_A-S_B?**
  - Based on findings so far: **Monotonically decreasing**



A Puzzle: The Mercy Bounce

- **What is P(+) as a function of ? = S_A-S_B?**

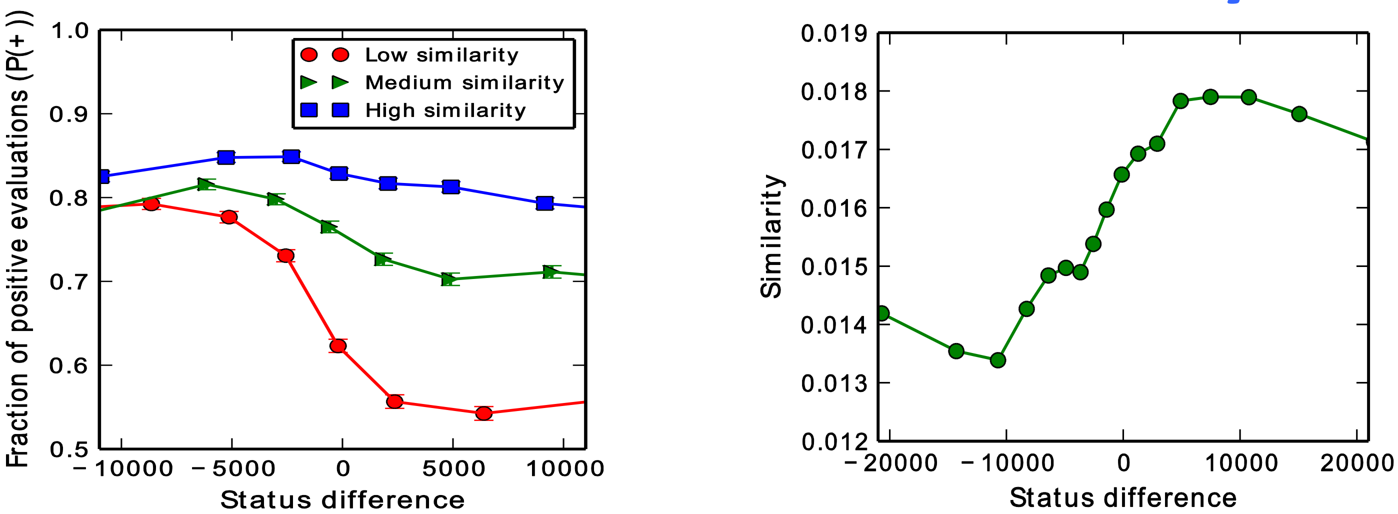


How can we explain this?

[WSDM '12]

The Mercy Bounce

- **Why low evals. of users of same status?**
  - Not due to users being tough on each other
  - **But due to the effects of similarity**



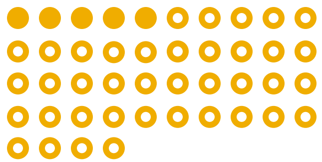
- **So:** High-status evaluators tend to be more favorably disposed

Aggregating Evaluations

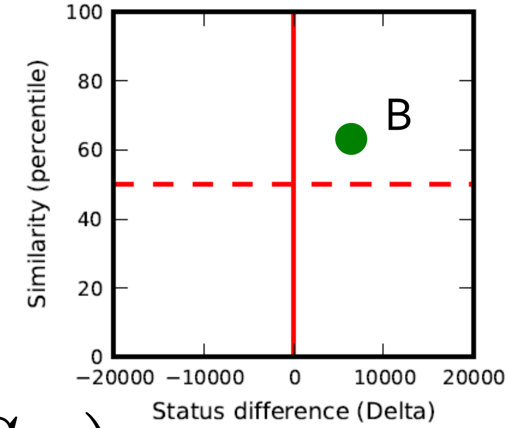
- **So far:** Properties of individual evaluations
- **But:** Evaluations need to be "summarized"
  - Determining rankings of users or items
  - Multiple evaluations lead to a group decision
- **How to aggregate user evaluations to obtain the opinion of the community?**
  - Can we guess community's opinion from a small fraction of the makeup of the community?

Ballot-blind Prediction

- **Predict Wikipedia adminship election results without seeing the votes**
  - Observe identities of the first *k* (=5) people voting (but *not* how they voted)
  - Want to predict the election outcome
    - Promotion vs. no promotion
- **Why is it hard?**
  - Don't see the votes (just voters)
  - Only see first 5 voters (out of ~50)



Ballot-blind: The Model



- **Want to model prob. user *A* votes + in election of user *B***
- **Our model:**

$$P(A=+|B) = P_A + d(\Delta_B, S_B)$$

  - P_A ... empirical fraction of + votes of A
  - *d(S,Δ)* ... avg. deviation in fraction of + votes
    - When *A*s evaluate *B* from a particular *(S,Δ)* quadrant, how does this change their behavior
- Predict 'elected' if: $\sum_{i=1}^{k} P(A_i=+|B) > w$

Ballot-blind Prediction

- Based on only who showed to vote predict the outcome of the election

Number of voters seen	Accuracy
5	71.4%
10	75.0%
all	75.6%

  - Other methods:

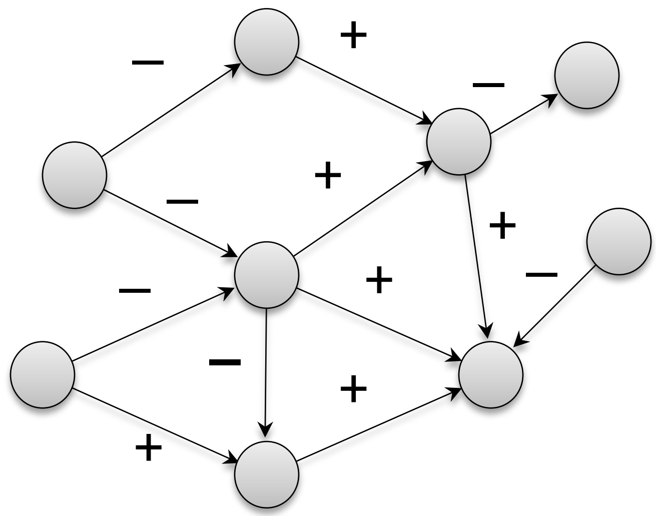
Theme: Learning from implicit feedback

Audience composition tells us something about their reaction

Evaluations in a Context

[WWW '10]

- Evaluations form a signed network



  - Network provides a context in which signed edges are formed
- **What can we say about the edges?**

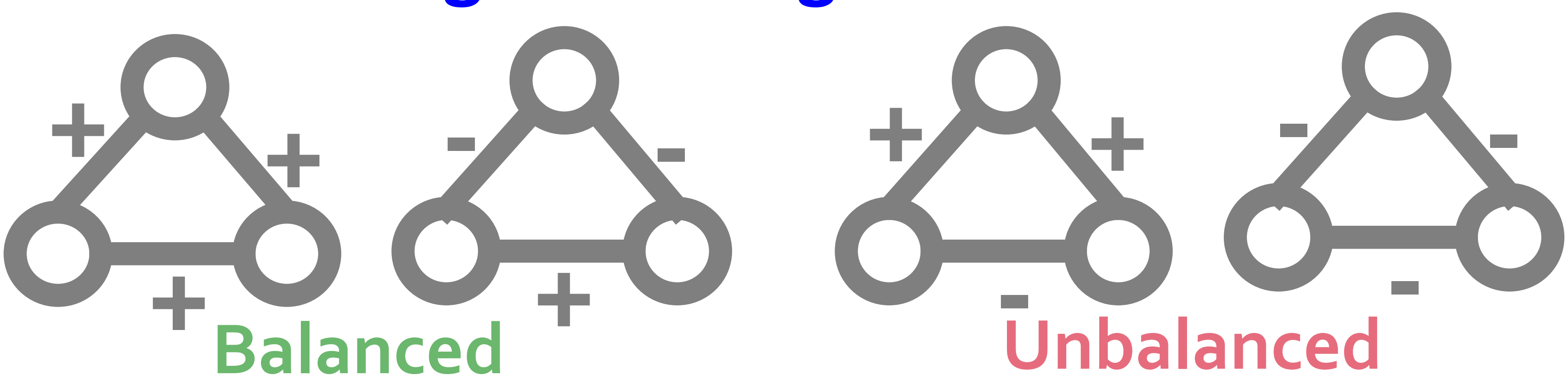
Status Theory

[CHI '10]

- **Status in a network** [Davis-Leinhardt '68]
  - $A \xrightarrow{+} B$ :: B has **higher** status than A
  - $A \xrightarrow{-} B$ :: B has **lower** status than A
    - (Note the notion of status is now implicit)
  - Apply this principle transitively over paths
    - Can replace each $A \xrightarrow{-} B$ with $A \xleftarrow{+} B$
    - Obtain an all-positive network with same status interpretation

Structural Balance

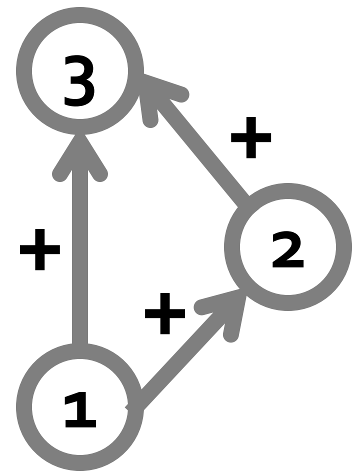
- **Start with the intuition** [Heider '46]
  - The friend of my friend is my friend
  - The enemy of enemy is my friend
  - The enemy of friend is my enemy
  - The friend of my enemy is my enemy
- **Look at signed triangles:**

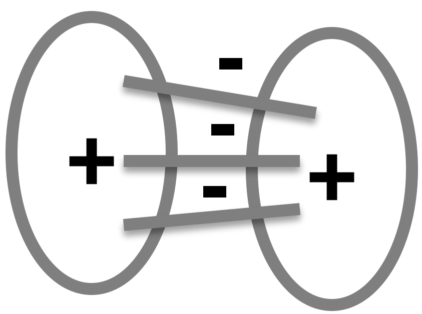


Status vs. Balance

At a global level:

- **Status ⇒ Hierarchy**
  - All-positive directed network should be (approximately) **acyclic**
- **Balance ⇒ Coalitions**
  - Balance ignores directions and implies that subgraph of negative edges should be (approximately) **bipartite**





[CHI '10]

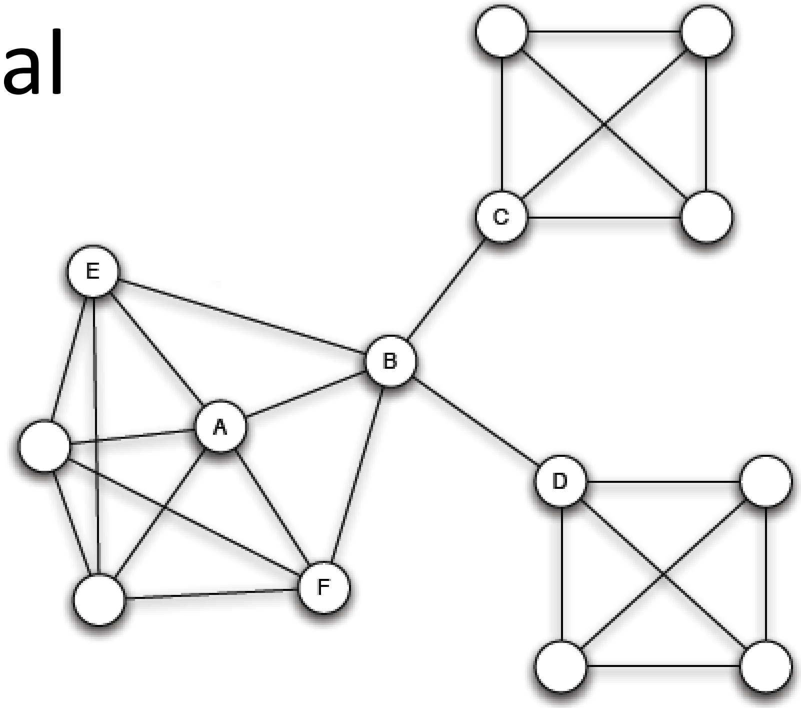
Status vs. Balance

- Aggregate tendency toward Status

- Theories are at work at different levels:
  - Balance more applicable on reciprocated links
- Design implication:
  "I agree with you" vs. "I respect you."

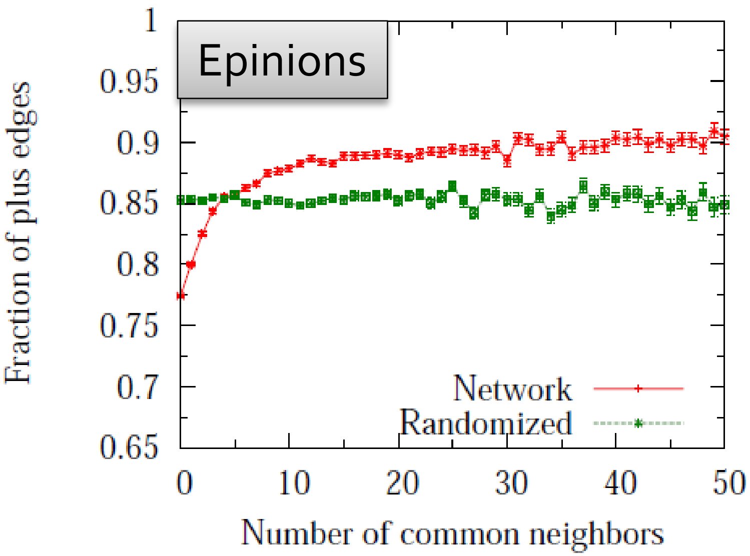
Global Structure

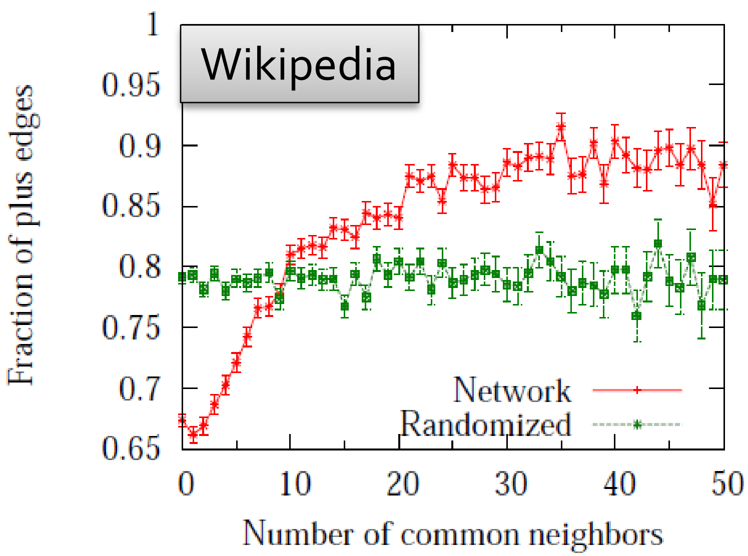
- Intuitive picture of social network in terms of densely linked clusters
- **How do link structure and signs interact?**
- **Embeddedness** of an edge (A,B): number of shared neighbors



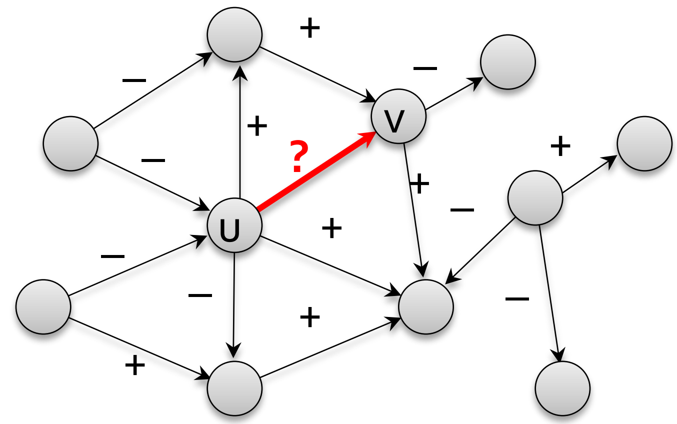
Embeddedness

- **Embeddedness of ties:**
  - Embedded ties tend to be more positive
- A natural connection to triadic closure based **social capital** [Coleman '88]
  - Public display of signs (votes) in Wikipedia further strengthens this

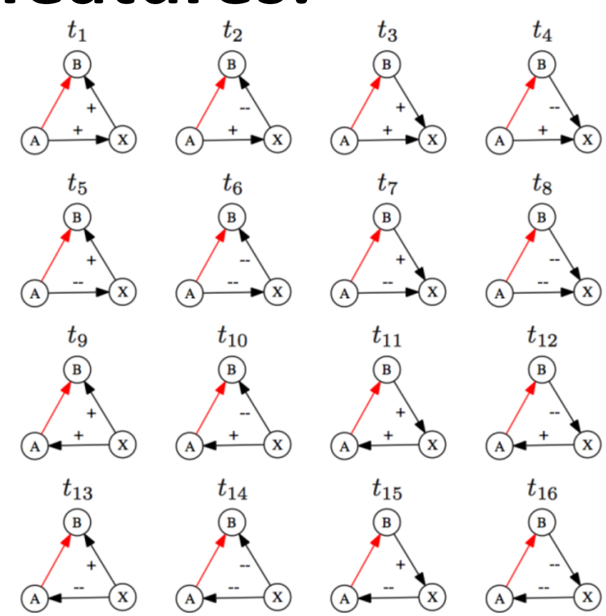




Application: Predicting Signs



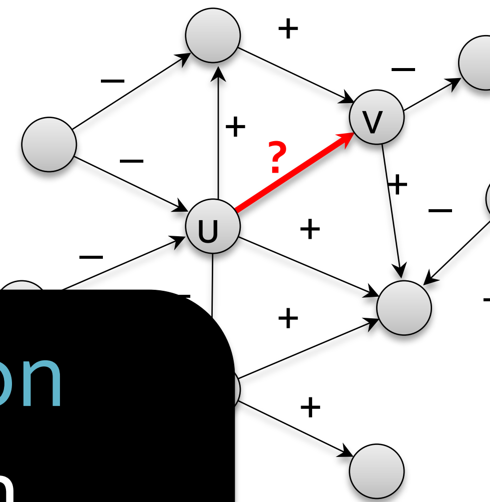
- **How will A evaluate B?**
  - **Predicting edge signs**
- **Model:**
  - Count the triads in which edge A ⟶ B is embedded: **16 features:**
  - Train Logistic Regression
  - **Predictive accuracy: ~90%**
- Evaluations can be modeled from the local network structure alone!



[WWW '10]

Application: Predicting Signs

- **How generalizable are the results across the datasets?**
  - Epinions: Trust/Distrust



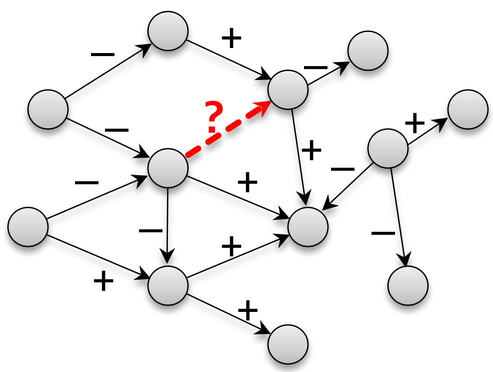
Nearly perfect generalization of the models even though evaluations have very different meaning

All23	Epinions	Slashdot	Wikipedia
Epinions	0.9342	0.9289	0.7722
Slashdot	0.9249	0.9351	0.7717
Wikipedia	0.9272	0.9260	0.8021

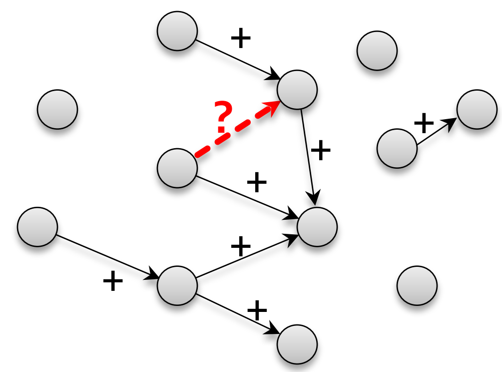
[WWW '10]

Negative information helps?

- Suppose we are only interested in predicting whether there is a positive edge or no edge
- **Does knowing negative edges help? YES!**



Vs.



Features	Epinions	Slashdot	Wikipedia
Positive edges	0.5612	0.5579	0.6983
Positive and negative edges	0.5911	0.5953	0.7114

Conclusions and Reflections

- **General challenge:** In many situations, opinions and evaluations are expressed, but the underlying principles driving them may not be obvious
- **Basic models provide a vocabulary for dissecting the fundamental ingredients**
  - Relative assessment: **Status**
  - Prior interaction: **Similarity**

Conclusions and Reflections

- **Dimensions of an opinion:**
  - Status vs. Similarity
  - Agreement with the statement vs. Statement is technically correct
- **On-line domains:** People are applying multiple dimensions of evaluation, but the interfaces they use collapse them to a single dimension

Future Directions

- **How communities form collective judgments in social applications?**
- **Model outcomes of group decisions from small set of evaluations**
  - **Predict outcomes without explicit user feedback**
  - Audience composition predicts audience's reaction

Conclusion

- **Evaluations create incentives** (and sometimes unfair evaluations can produce better outcomes)
  - **Status and reputation mechanisms**
- **Trust issues:** Why should I trust another user, or the community as a whole?

An opportunity to understand the range of forces at work, and use this to inform the design of new applications

THANKS!

Data + Code:

http://snap.stanford.edu

References

- Signed Networks in Social Media by J. Leskovec, D. Huttenlocher, J. Kleinberg. *CHI*, 2010.
- Predicting Positive and Negative Links in Online Social Networks by J. Leskovec, D. Huttenlocher, J. Kleinberg. *WWW*, 2010.
- Governance in Social Media: A case study of the Wikipedia promotion process by J. Leskovec, D. Huttenlocher, J. Kleinberg. *ICWSM*, 2010.
- Effects of User Similarity in Social Media by A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec. *WSDM*, 2012.
- Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow by A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec. *KDD*, 2012.
- Supervised Random Walks: Predicting and Recommending Links in Social Networks by L. Backstrom, J. Leskovec. *WSDM*, 2011.