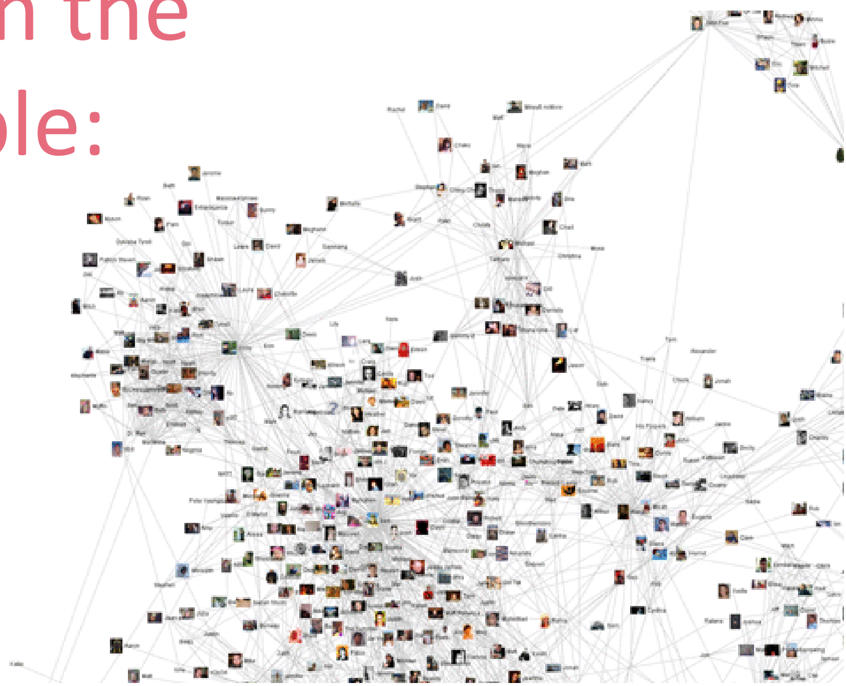
How People Evaluate One Another in Social Media

Jure Leskovec
Computer Science, Stanford

Social Networks and Social Media

Science advances when the invisible becomes visible:



- Social interaction is leaving digital traces on-line
- Can we recognize fundamental patterns of human behavior from raw digital traces?

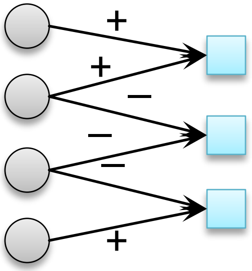
People have Opinions

People express positive and negative attitudes/opinions:

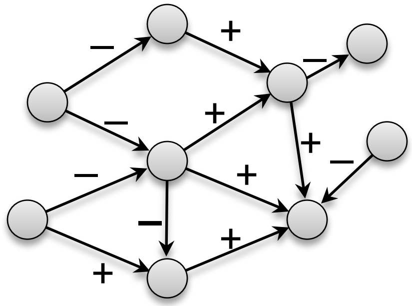
- Through actions:
  - Rating a product
  - Pressing "like" button
- Through text: Sentiment analysis [Pang-Lee '08]
  - Writing a comment, a review

People Express Opinions

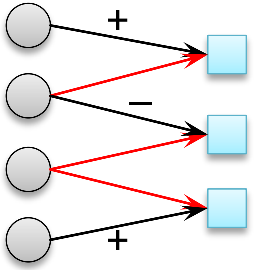
- About items:
  - Movie and product reviews



- About other users:
  - Online communities

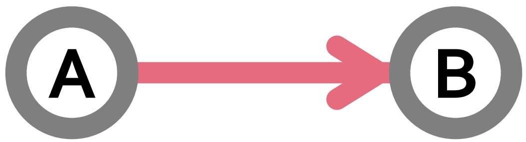


- About items created by others:
  - Q&A websites

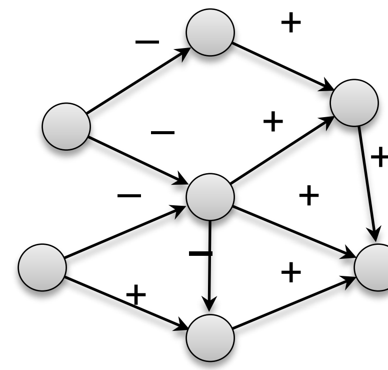


This talk: Users evaluating others

- Any user A can evaluate any user B:



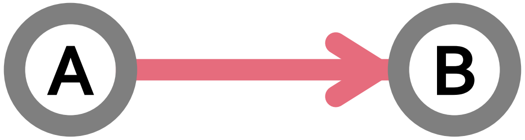
  - Positive (+) vs. negative (–) evaluation



- In what (online) settings does this process naturally occur at large scale?
  - Epinions: Trust/Distrust (1M evals)
    - Does A trust B's product reviews?
  - Wikipedia: Support/Oppose (150k votes)
    - Does A support B to become Wiki admin?
  - Stackoverflow: Up/down vote (6M votes)
    - Does A think B contributed a good answer?

Relative vs. Absolute Assessment

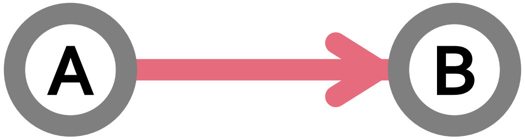
- How do properties of **evaluator A** and **target B** affect A's vote?



- 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 a user 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 characteristics of A and B
    - **Similarity:** Prior interaction between A and B
    - **Status:** A compares status of B to her own status

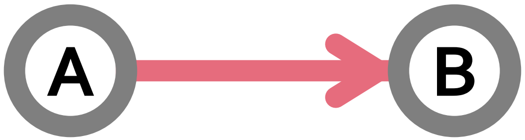
Status (level of contribution)

Three ways to quantify status S:

- Total number of edits of a user:
  - The more edits the user made the higher status she has
- Total number of answers of a user:
  - The more answers given by the user the higher status she has

Status: How to model?

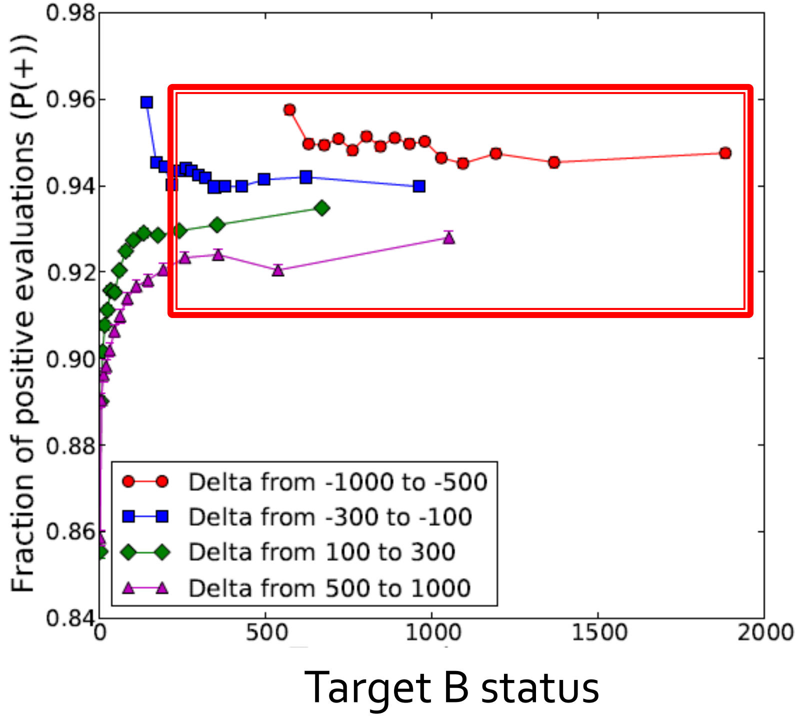
- How does the prob. of A evaluating positively depend on the status of A and status of B?



  - Model it as a function of status S_A of A and S_B of B separately?
  - Model as the status difference S_A-S_B?
  - Model as the status ratio S_A/S_B?

Status: Relative Assessment (1)

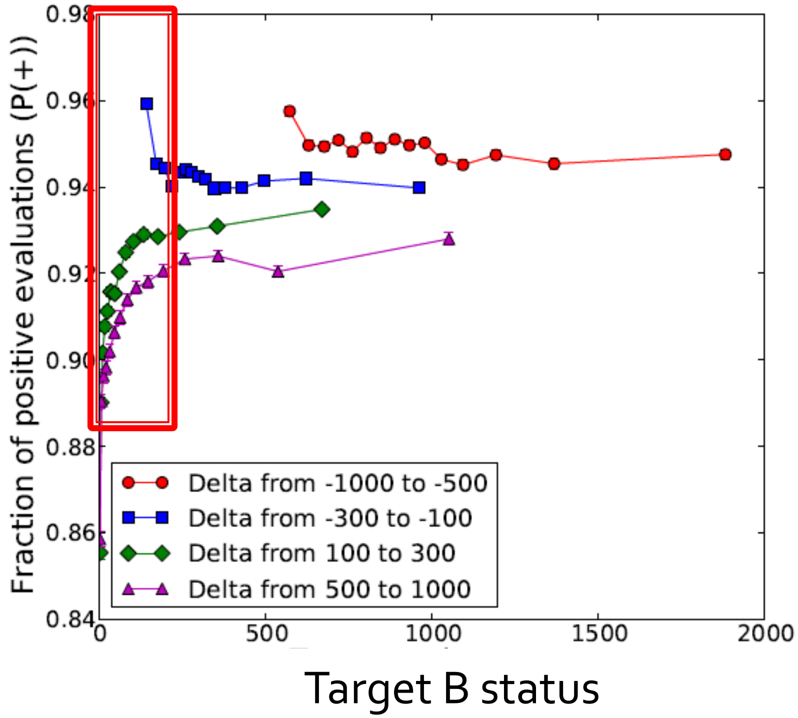
- How does status of B affect A's evaluation?
  - Each curve is fixed status difference: $\Delta = S_A - S_B$
- Observations:
  - Flat curves: Prob. of positive evaluation doesn't depend on B's status
  - Different levels: Different values of Δ result in different behavior



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

Status: Relative Assessment (2)

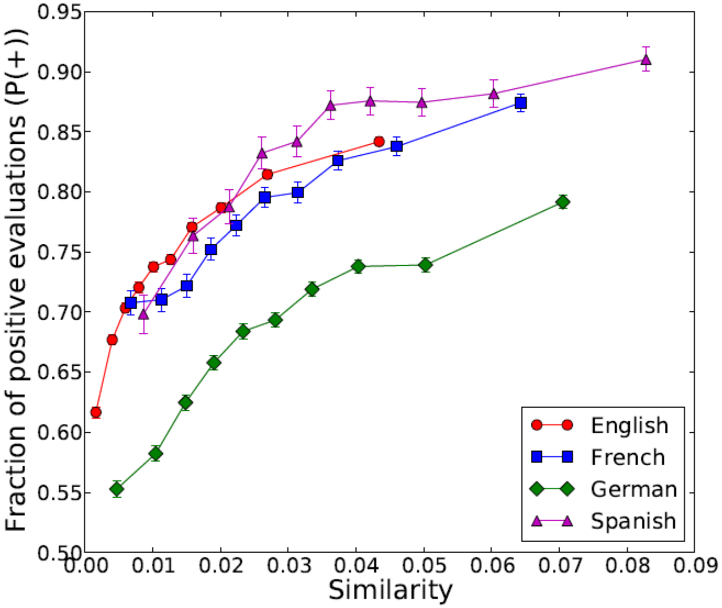
- How does status of B affect A's evaluation?
  - Each curve is fixed status difference: $\Delta = S_A - S_B$
- Observations:
  - Below some threshold targets are judged based on their absolute status
    - And independently of evaluator's status



Low-status targets are evaluated based on absolute status

Effects of Similarity

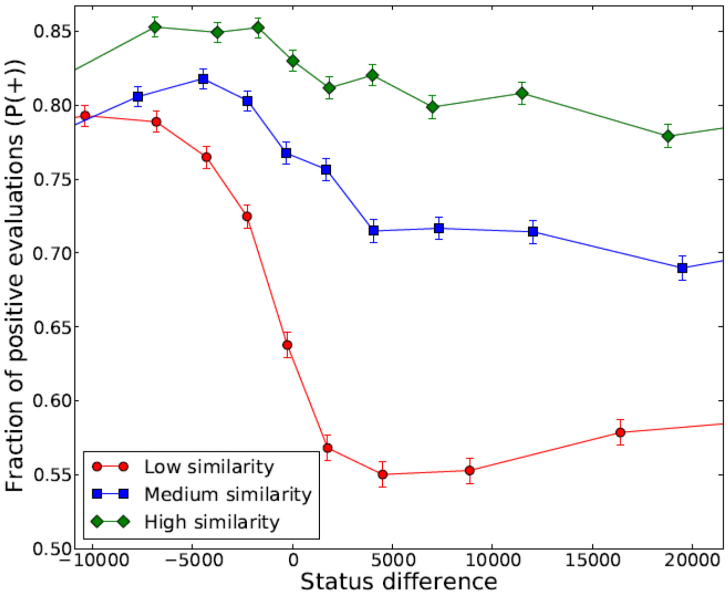
- How does prior interaction shape evaluations?
  - (1) Evaluators are more supportive of targets in their area
  - (2) More familiar evaluators know weaknesses and are more harsh
- Observation:
  - Prior interaction/similarity increases prob. of a positive evaluation



Prior interaction/ similarity boosts positive evaluations

Relating Status and Similarity (1)

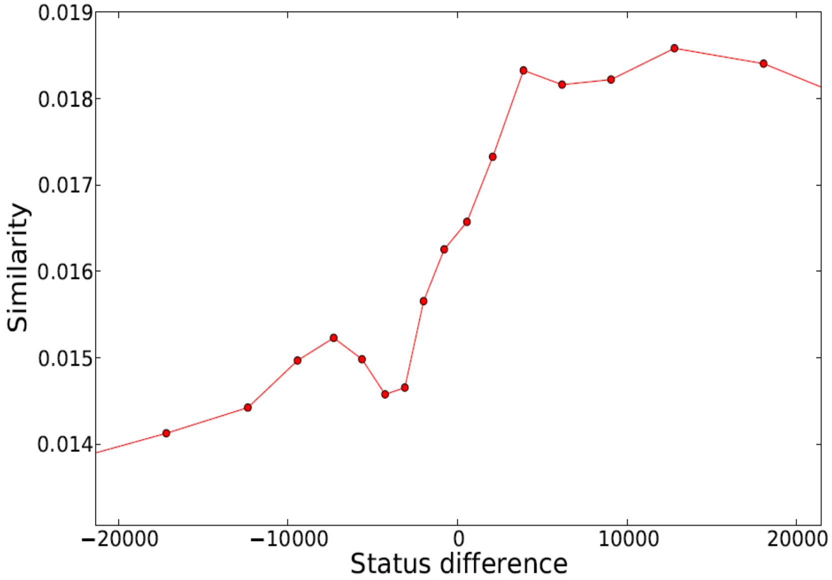
- Observation:
  - Evaluation depends less on status when evaluator A is more informed
- Consequence:
  - Evaluators use status as proxy for quality in the absence of direct knowledge of B



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

Relating Status and Similarity (2)

- Observation:
  - Evaluators with higher status than the target are more similar to the target

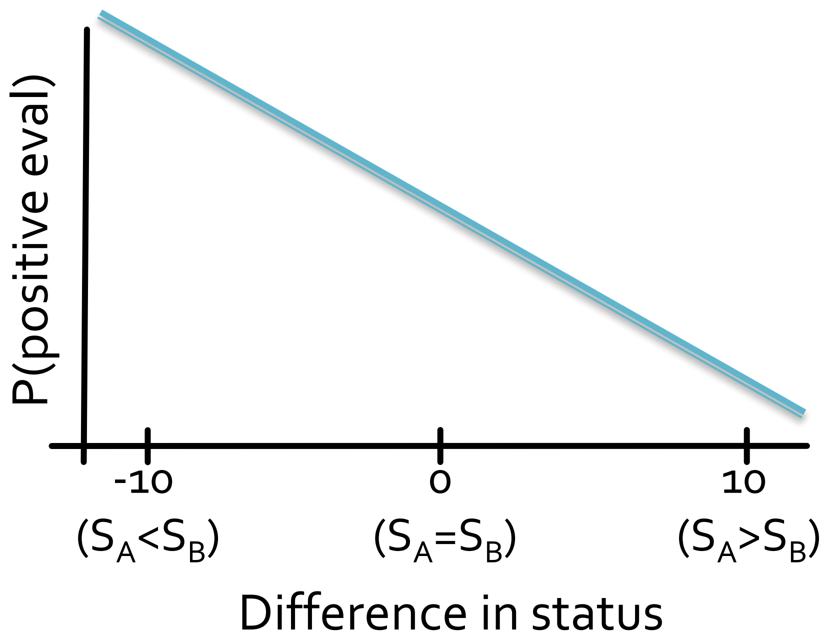


- Selection bias:
  - High-status evaluators are more similar to the target

Elite evaluators vote on targets in their area of expertise

Puzzle: Status

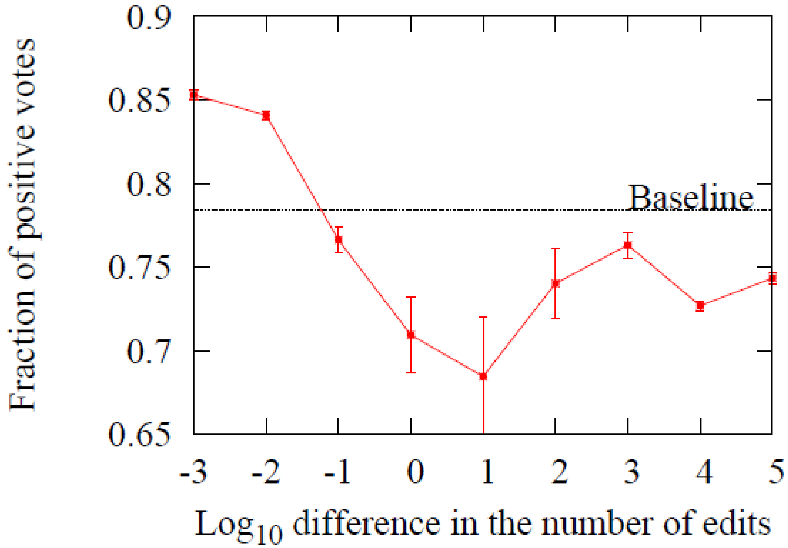
- Evaluator A evaluates target B
- Prob. of positive evaluation of A as a function of status difference: $\Delta = S_A - S_B$
  - Hypothesis: Monotonically decreases

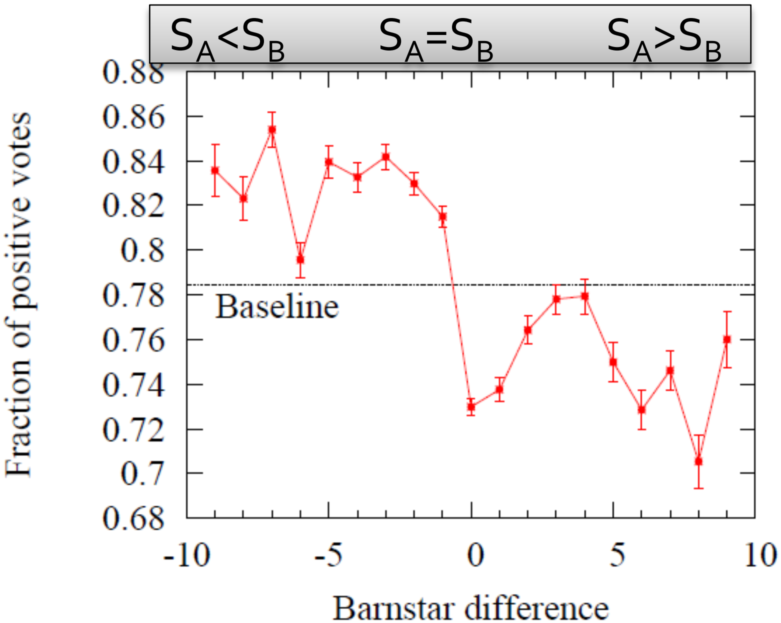


Puzzle: Status

- Prob. of positive evaluation of B as a function of status difference: $\Delta = S_A - S_B$
- Observations:
  - A is especially negative when status equals: $S_A=S_B$
  - "Mercy bounce" for $S_A>S_B$

How to explain the bounce?





Why most harsh at zero difference?

How to explain low aggregate evaluations given by users to others of same status?

- Not due to users being tough on each other
  - Similarity increases the positivity of evaluations

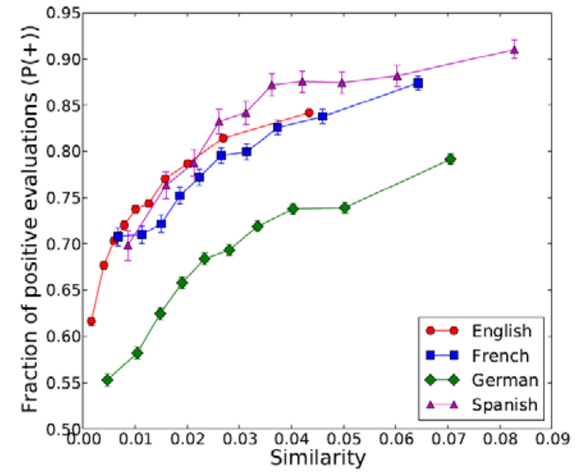
Possible (but wrong) explanation:

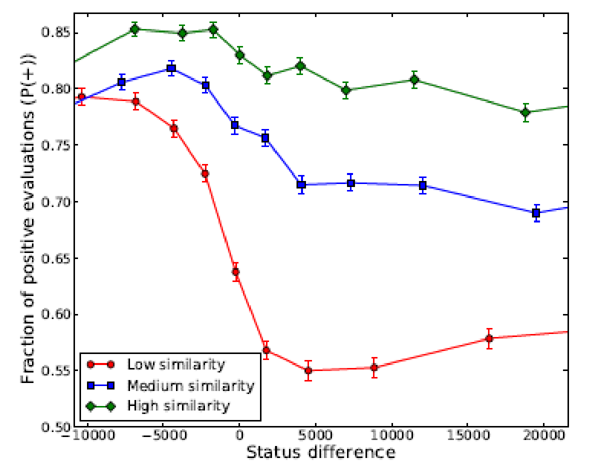
- Most targets have low status (small $\Delta > 0$)
- Low-status targets are judged on abs. status
  - The rebound persists even for high-status targets

Explanation: Differential Status

Model ingredients:

- Similarity:
  - Highly similar users are more positive
- Selection bias:
  - High-similarity users are overrepresented among high-status evaluators

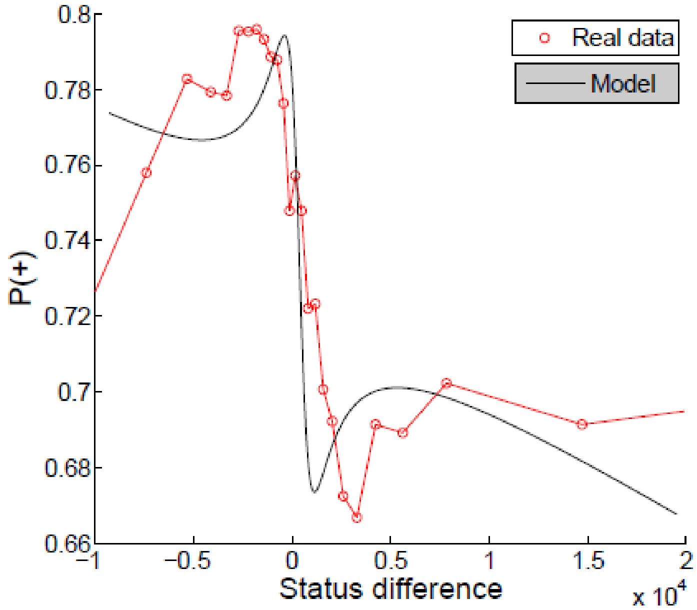




Explanation: Differential Status

- The rebound not the effect of harshness of same status evaluators…

  … but a combination of
- how low-status users are evaluated
- who shows up to evaluate users



Application: Predicting outcomes

- Predict the outcome using only properties of evaluators without looking at their votes
  - Wikipedia: Based on only who showed to up to vote predict the outcome of the election
- Simple model:
  - Target status
  - Evaluator status
  - Similarity

Application: Ballot-blind prediction

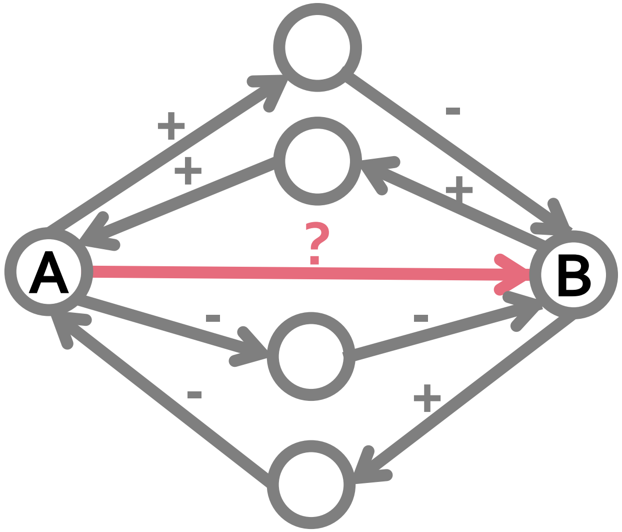
- Based on only who showed to up to evaluate predict the outcome of the Wiki election

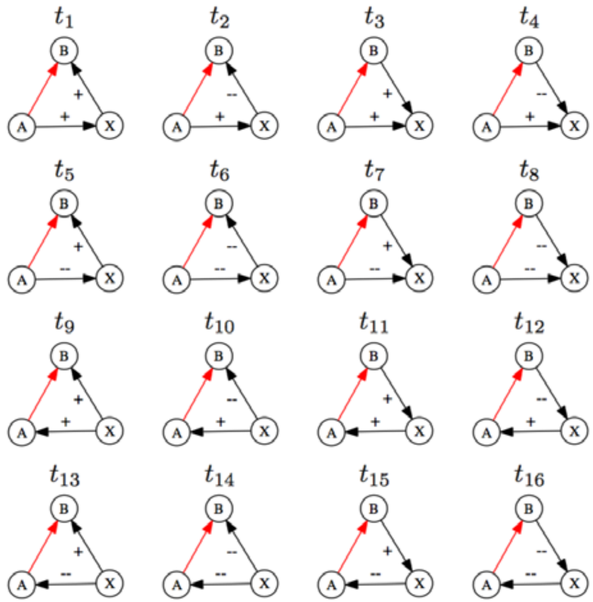
Number of votes	E	Relative gain over LogReg
5	71.4%	12.8%
10	75.0%	23.8%
all	75.6%	25.7%

- Method:
  - Divide the Status-Similarity space, each cell prob. + vote
- Baseline:
  - Guessing gives 50% accuracy
  - Logistic Regression based on the target status (67% acc)

Application: Predicting evaluations

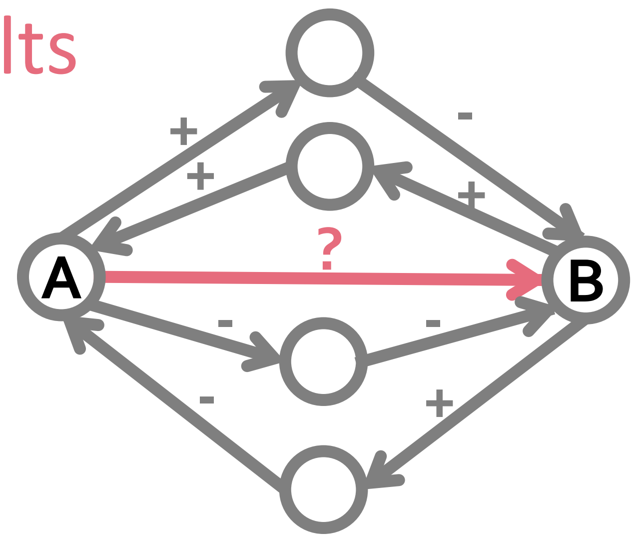
- How will A evaluate B?
- Model:
  - Count the triads in which edge A → B is embedded
  - Predictive accuracy: ~95%
- Evaluations can be modeled from local network structure alone!





Application: Predicting evaluations

- How do people evaluate in different contexts? How generalizable are the results across the datasets?
  - <u>Wikipedia:</u> Support/Oppose
  - <u>Epinions:</u> Trust/Distrust
  - <u>Stackoverflow:</u> Up/Down vote



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

Conclusions

- Social media sites are governed by (often implicit) user evaluations
- Wikipedia voting process has an explicit, public and recorded process of evaluation
  - Similarly, Epinions and Stackoverflow
- Main characteristics:
  - Importance of relative assessment: **Status**
  - Importance of prior interaction: **Similarity**

Conclusion and reflections

- Online social systems are globally organized based on status

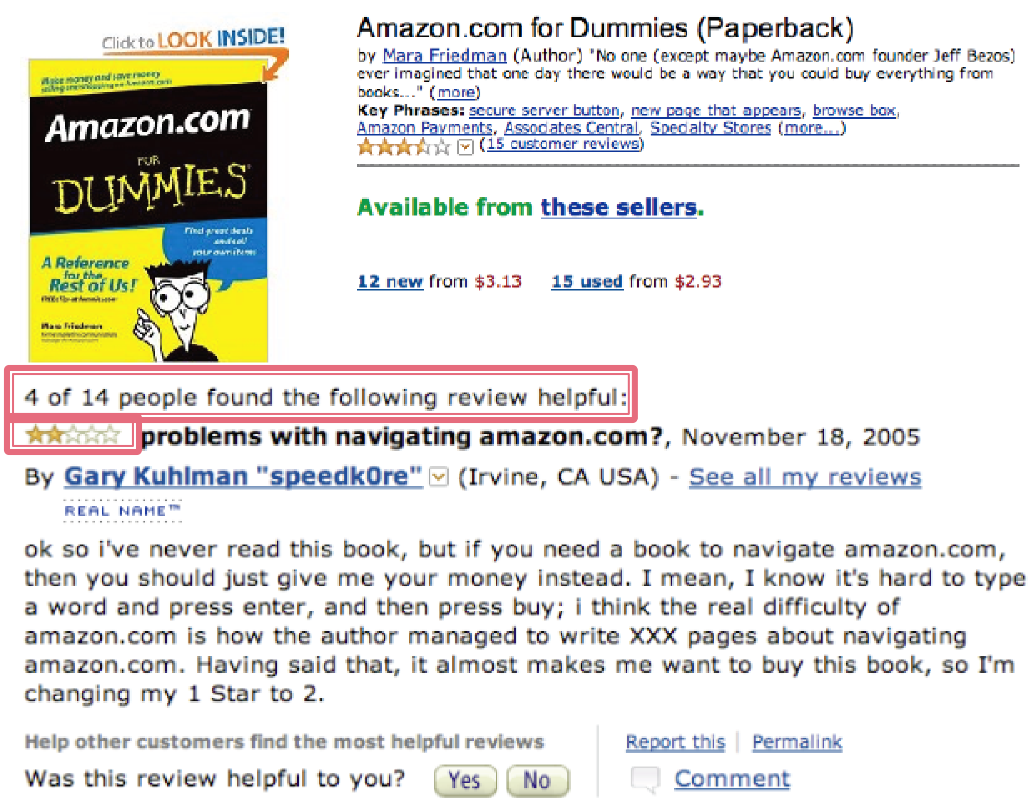
- Users use evaluations consistently regardless of a particular application
  - Near perfect generalization across datasets

- What kinds of opinions do people find helpful?

[Danescu et al., 2009]

What do people find helpful?

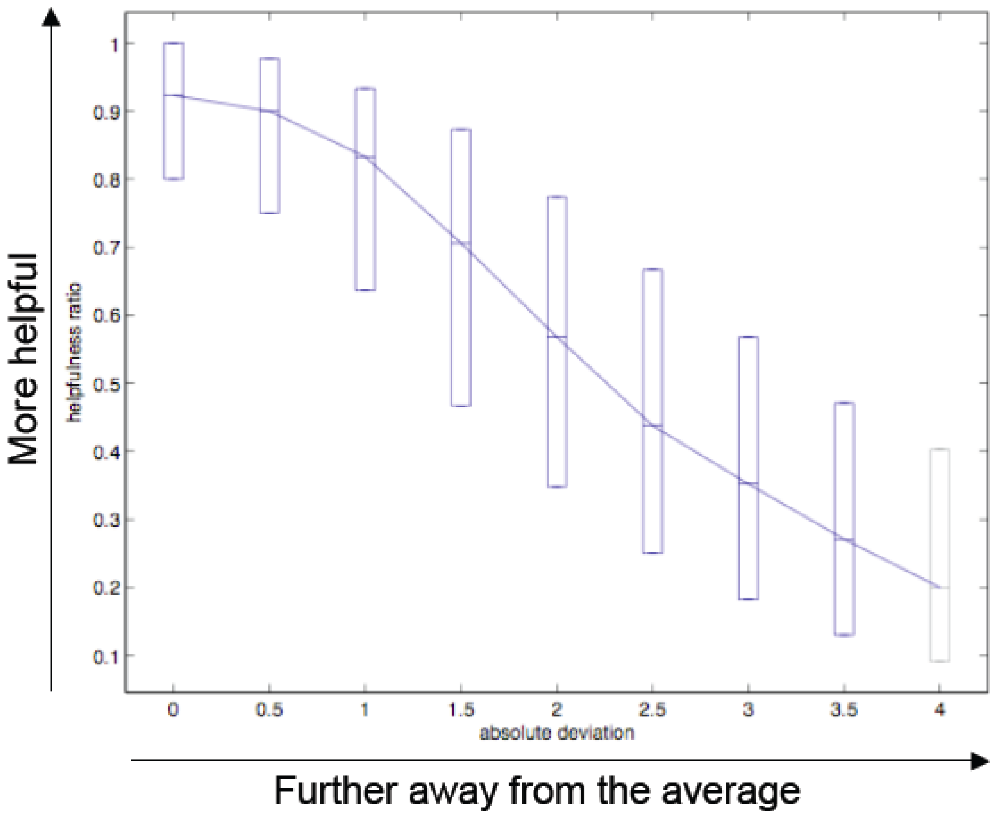
- What do people think about our recommendations and opinions?



[Danescu et al., 2009]

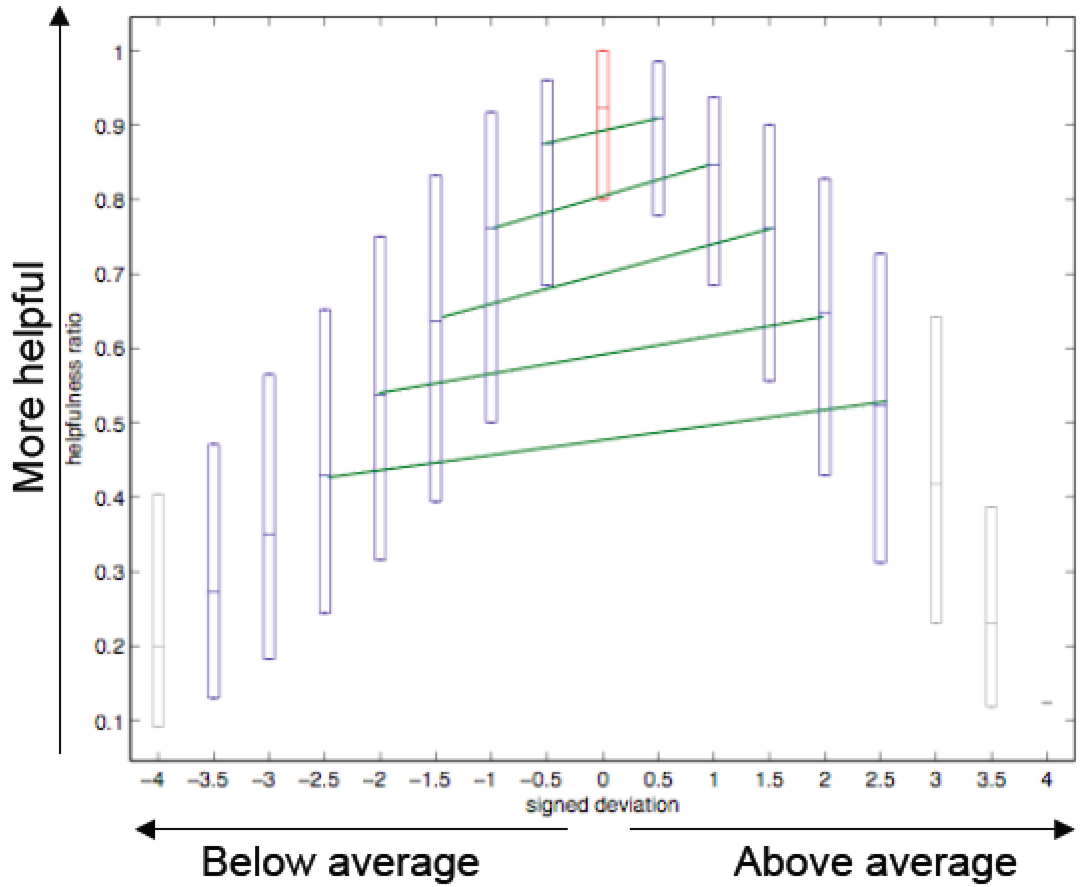
Review helpfulness: Conformity

- People find conforming opinions more helpful



Review helpfulness: Deviation

- Positive reviews are more helpful



Future Directions

- Predict the outcome of group evaluations from small set of evaluations
  - Seeing just a few votes, what’s the final outcome
- Predicting outcomes without explicit user feedback
  - Based on who shoed up, predict outcome

Future Directions

- Understanding the dimensions of the opinion:
  - Status vs. Similarity
  - Agreement with the statement vs. Statement is technically correct
- Status and reputation mechanisms
  - What reputation/merit mechanisms should we build into the social systems to achieve desirable behavior?

http://snap.stanford.edu

THANK YOU!