LECTURE 12:HOW DO PEOPLE EVALUATE OTHERS?

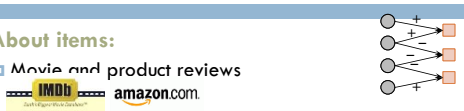
People Express Opinions

In many online applications users express positive and negative attitudes/opinions:

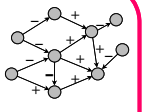
- Through **actions**:
  - Rating a product
  - Pressing a "like" button
- Through **text**:
  - Writing a comment, a review
- **Success of these online applications is built on people expressing opinions**
  - Recommender systems
  - Wisdom of the Crowds
  - Ranking

People & Evaluations

- **About items:**
  - Movie and product reviews



- **About other users:**
  - Online communities



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

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]

Evaluations: Some Issues

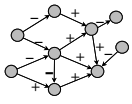
Some of the central issues:

- **Factors:**
  What factors drive one's evaluations?

- **Synthesis:**
  How do we create a composite description that accurately reflects cumulative opinion of the community?

Evaluations: the 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?**

Evaluations: the Data

7

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

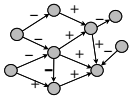
- **Stack Overflow Q&A community**
  - Upvote/Downvote (7.5M votes)

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

The New Setting

8

- **Relation to the previous class:**
  We still talk about one person evaluating the other via a +/- evaluation



So far we focused on evaluations in the context of a network

Now we focus on a single evaluation (without the context of a network)

Human Evaluations

9

- **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

10

- **Status:**
  (note status is now explicit, and not implicitly determined by the network!)
  - 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:** Similarity of the articles edited
    - **Stack Overflow:** Similarity of users evaluated

Relative vs. Absolute Assessment

11

- 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

12

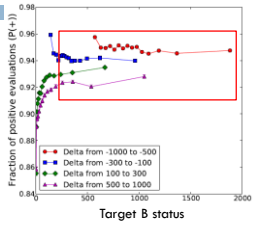
- How do properties of **evaluator A** and **target B** affect A's vote?

- 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

13

- 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

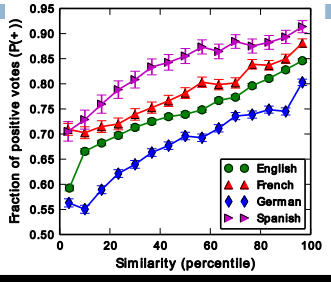
Effects of Similarity

14

- 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

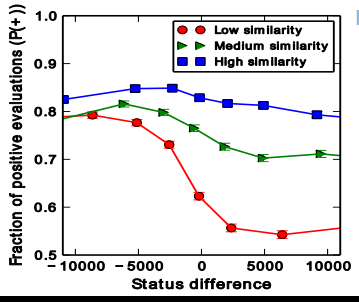
15



Prior interaction/ similarity boosts positive evaluations

Status & Similarity

16

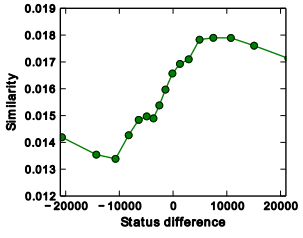


Status is a proxy for quality when evaluator does not know the target

Status & Similarity

17

- Who shows up to evaluate?



Elite evaluators vote on targets in their area of expertise

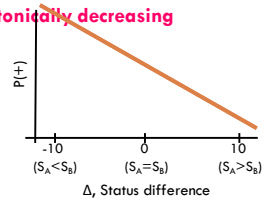
- Selection effect in who gives the evaluation
  - If $S_A > S_B$ then A and B are highly similar

A Puzzle

18

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

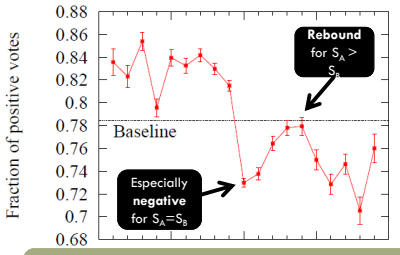
Monotonically decreasing



A Puzzle: The Mercy Bounce

19

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

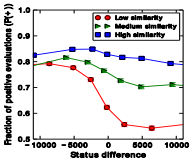


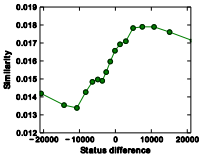
How can we explain this?

The Mercy Bounce

20

- **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

21

- **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

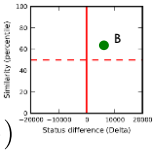
22

- **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

23

- **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
  - $d(S,\Delta)$ … 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

24

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

Number of votes	E
5	71.4%
10	75.0%
all	75.6%

- **Other methods:**
  - Guessing gives 52% accuracy
  - Logistic Regression on status and similarity features: 67%
  - If we see the first *k* votes 85% (gold standard)

Theme: Learning from implicit feedback
Audience composition tells us something about their reaction

Summary

25

- 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
- **Application:** Ballot-blind prediction

Important Points

26

- Online social systems are globally organized based on **status**
- **Similarity** plays important role
- Audience composition helps predict audience's reaction
- **What kinds of opinions do people find helpful?**