

Networks with Signed Edges

CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

<http://cs224w.stanford.edu>



Announcement: Course Project

- **Project is a substantial part of the class**
 - Students put significant effort and great things have been done
- **Types of projects:**
 - **(1)** Analysis of an interesting dataset with the goal to develop a (new) model or an algorithm
 - **(2)** A test of a model or algorithm (that you have read about or your own) on real & simulated data.
 - Fast algorithms for big graphs. Can be integrated into SNAP.
- **Other points:**
 - The project should contain some mathematical analysis, and some experimentation on real or synthetic data
 - The result of the project will typically be an 8 page paper, describing the approach, the results, and related work.
 - **Come to us if you need help with a project idea!**

Announcement: Project Proposal

Project proposal: 3-5 pages, teams of up to 3 students

- **Project proposal has 3 parts:**
 - **(0) Quick 200 word abstract**
 - **(1) Reaction paper/Related work (2-3 pages):**
 - Read 3 papers related to the project/class
 - Do reading beyond what was covered in class
 - Think beyond what you read. Don't take other's work for granted!
 - **2-3 pages:** Summary (~1 page), Critique (~1 page)
 - **(2) Proposal (1-2 pages):**
 - Clearly define the problem you are solving.
 - **How does it relate to what you read for the Reaction paper?**
 - What data will you use? **(make sure you already have it!)**
 - Which algorithm/model will you use/develop? **Be specific!**
 - How will you evaluate/test your method?

See <http://cs224w.stanford.edu/info.html> for detailed instructions and examples of previous proposals

Announcement: Project Proposal

- **Logistics:**
 - **1) Register your group at:**
<http://bit.ly/1BNiHae>
 - **2) Submit PDF on GradeScope**
AND at <http://snap.stanford.edu/submit/>
 - **Due in 7 days: Thu Oct 16 at 9:30am!**
 - If you use a late period it will be added to all team members
- **If you need help/ideas/advice come to office hours/email us**

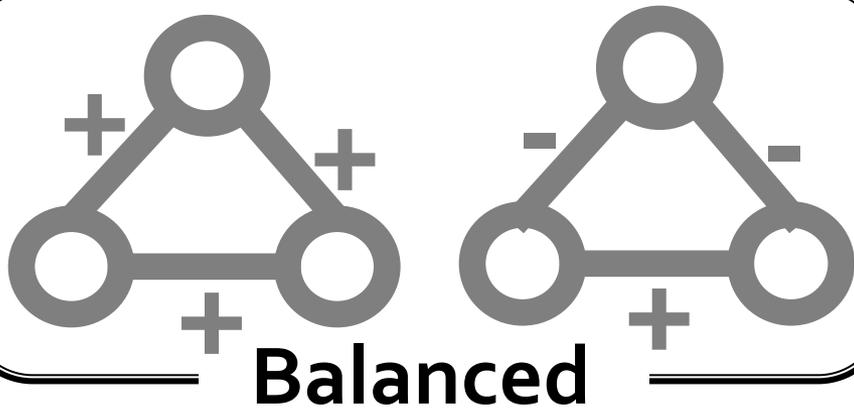
Interesting Datasets

- **Some example datasets:**
 - Books reviews + ratings
 - Townsquared: business page links and graph between Silicon Valley businesses
 - Github dataset
 - Stanford MOOC forums
 - Google Local
 - Flickr network
 - Beer review datasets

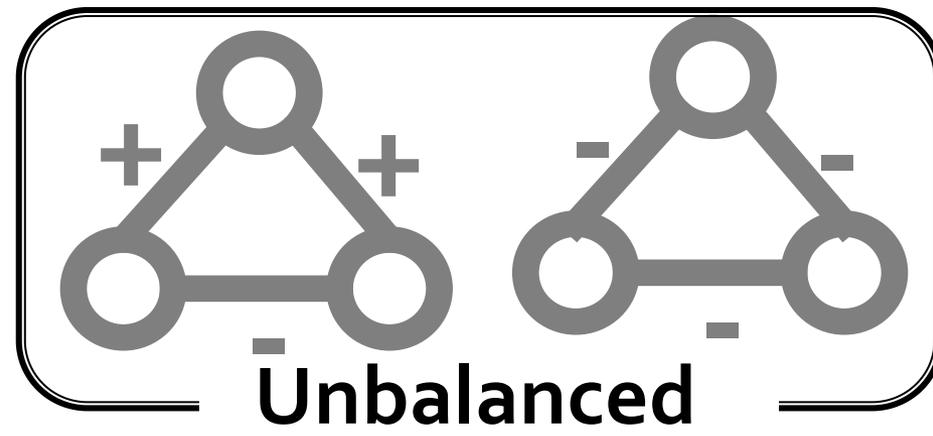
More at: <http://cs224w.stanford.edu/resources.html>

RECAP: Structural Balance

- **Start with the 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:



Consistent with "friend of a friend" or "enemy of the enemy" intuition



Inconsistent with the "friend of a friend" or "enemy of the enemy" intuition

Balance in General Networks

So far we talked about complete graphs

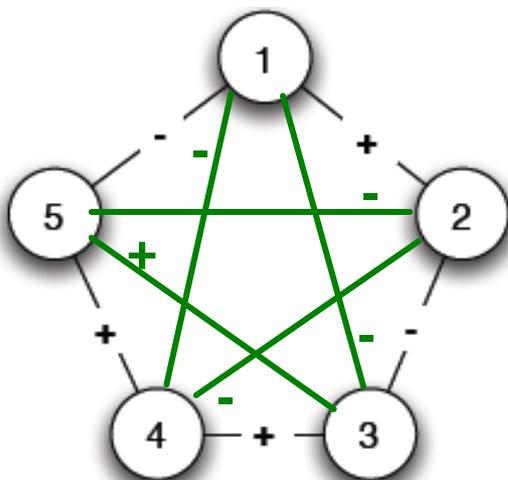
Def 1: Local view

Fill in the missing edges to achieve balance

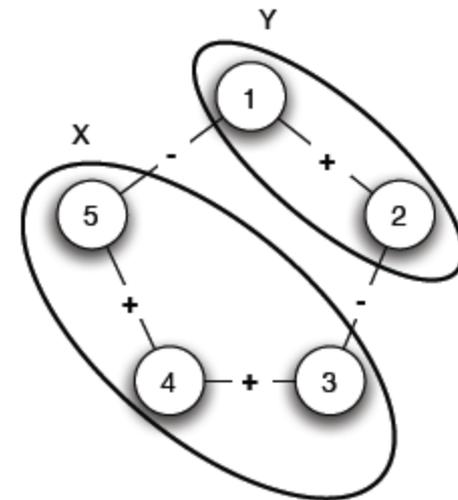
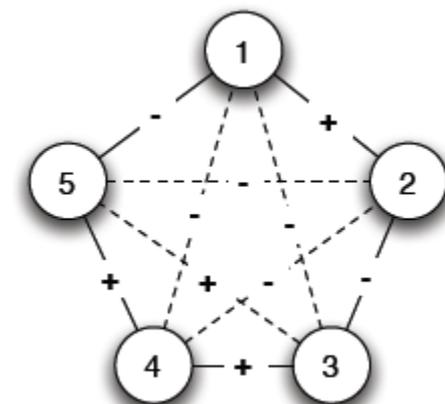
Def 2: Global view

Divide the graph into two coalitions

The 2 definitions are equivalent!

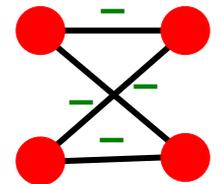


Balanced?

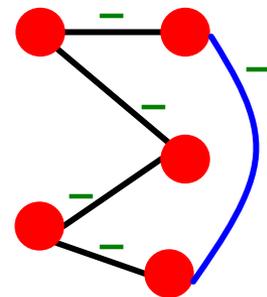


Is a Signed Network Balanced?

- Graph is **balanced** if and only if it contains **no cycle with an odd number of negative edges**
- **How to compute this?**
 - Find connected components on + edges
 - If we find a component of nodes on +edges that contains a – edge \Rightarrow **Unbalanced**
 - For each component create a super-node
 - Connect components A and B if there is a negative edge between the members
 - Assign super-nodes to sides using BFS

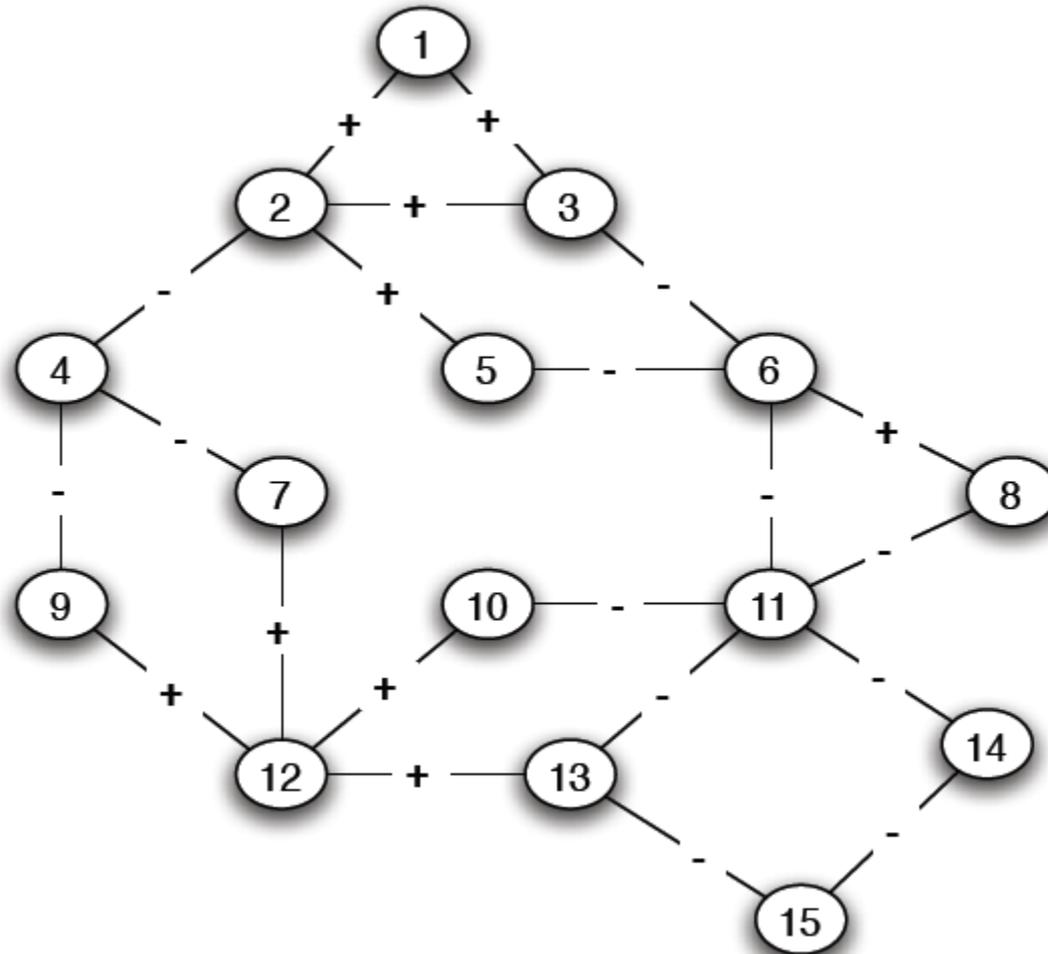


Even length cycle

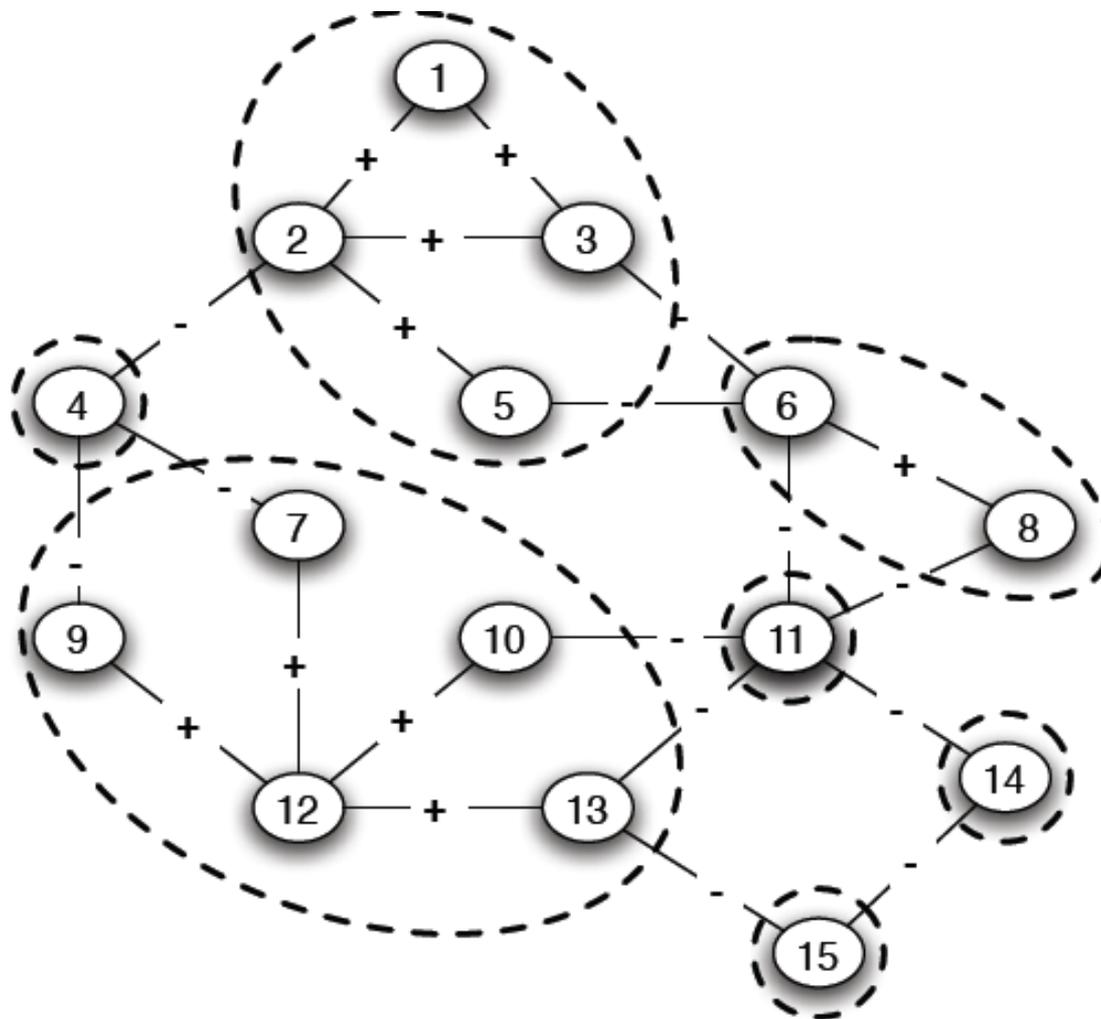


Odd length cycle

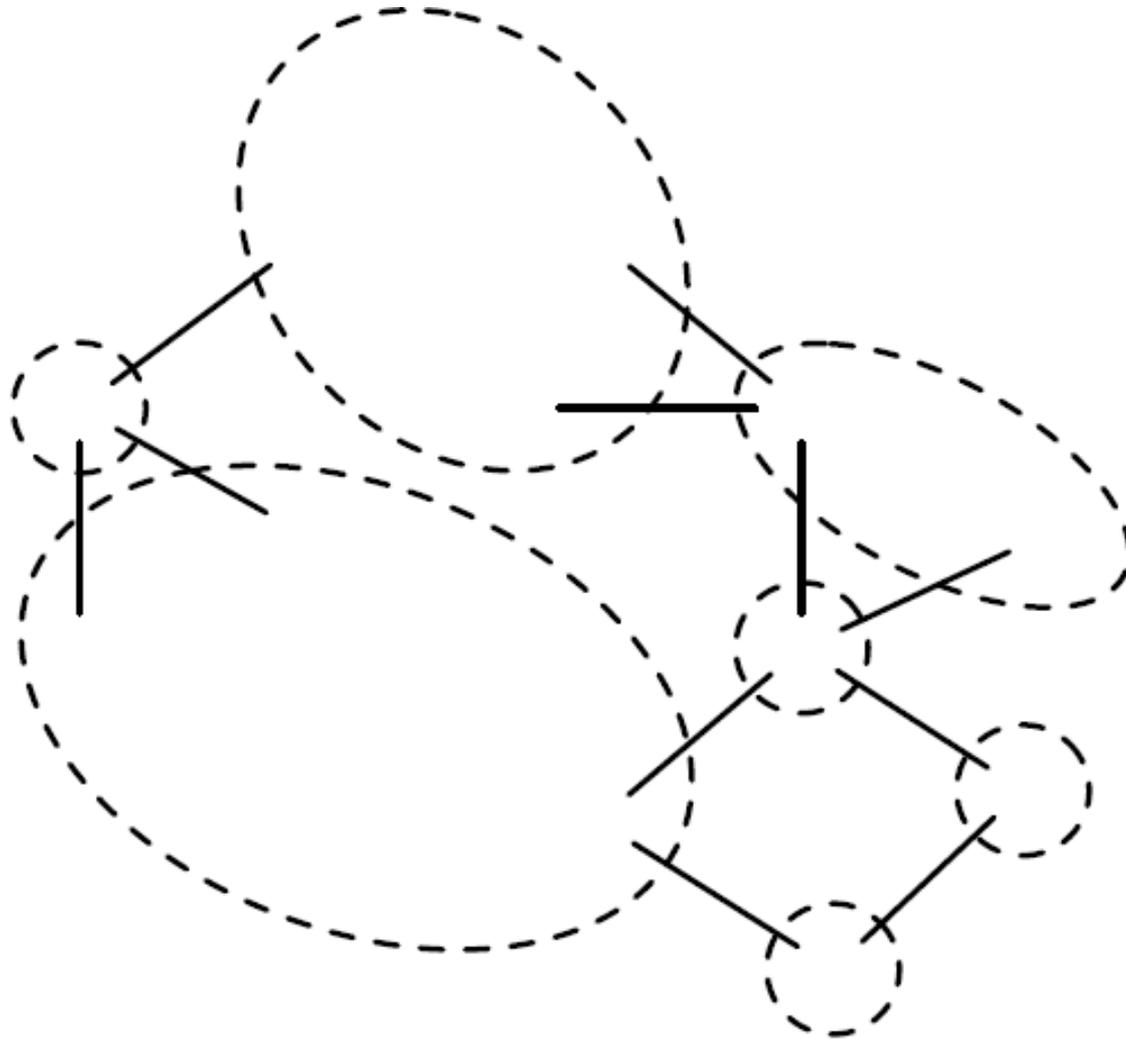
Signed Graph: Is it Balanced?



Positive Connected Components

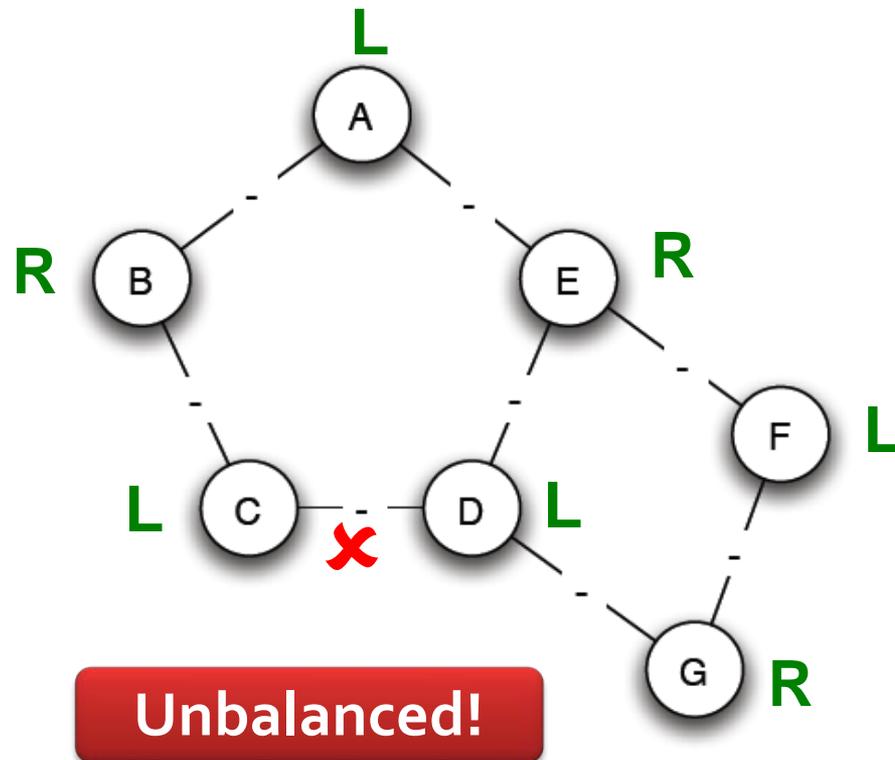


Reduced Graph on Super-Nodes



BFS on Reduced Graph

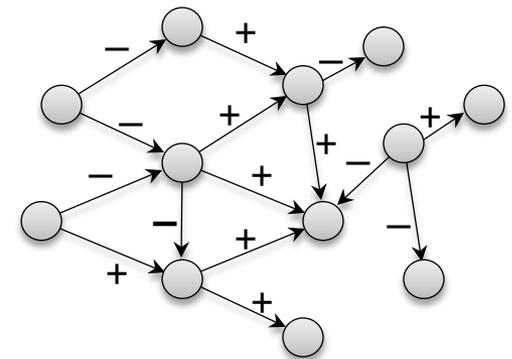
- Using BFS assign each node a **side**
- Graph is **unbalanced** if any two connected super-nodes are assigned the **same side**



Exploring Real Data

Real Large Signed Networks

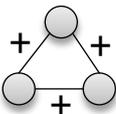
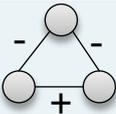
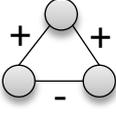
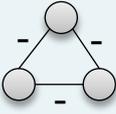
- Each link $A \rightarrow B$ is explicitly tagged with a sign:
 - **Epinions**: Trust/Distrust
 - Does A trust B's product reviews?
(only positive links are visible to users)
 - **Wikipedia**: Support/Oppose
 - Does A support B to become Wikipedia administrator?
 - **Slashdot**: Friend/Foe
 - Does A like B's comments?
 - **Other examples**:
 - Online multiplayer games



	Epinions	Slashdot	Wikipedia
Nodes	119,217	82,144	7,118
Edges	841,200	549,202	103,747
+ edges	85.0%	77.4%	78.7%
- edges	15.0%	22.6%	21.2%

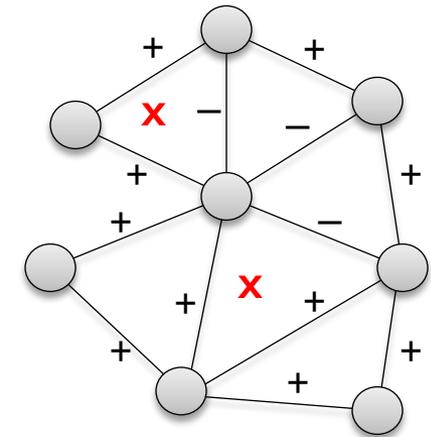
Balance in Our Network Data

- **Does structural balance hold?**
 - Compare frequencies of signed triads in real and “shuffled” signs

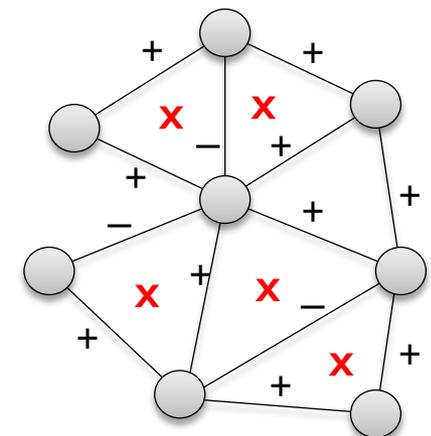
	Triad	Epinions		Wikipedia		Consistent with Balance?
		P(T)	$P_0(T)$	P(T)	$P_0(T)$	
Balanced		0.87	0.62	0.70	0.49	✓
		0.07	0.05	0.21	0.10	✓
Unbalanced		0.05	0.32	0.08	0.49	✓
		0.007	0.003	0.011	0.010	✗

P(T) ... fraction of a triads

$P_0(T)$... triad fraction if the signs would appear at random



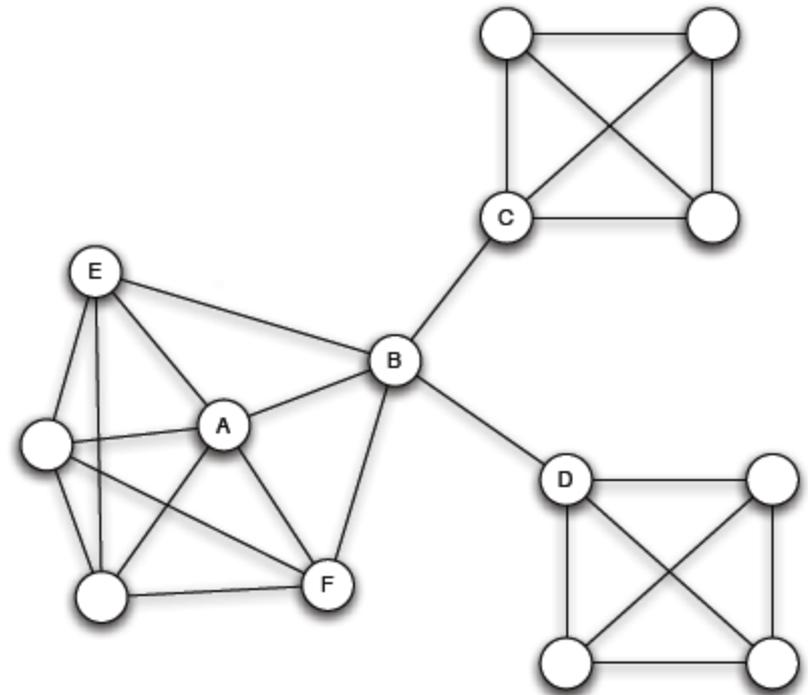
Real data



Shuffled data

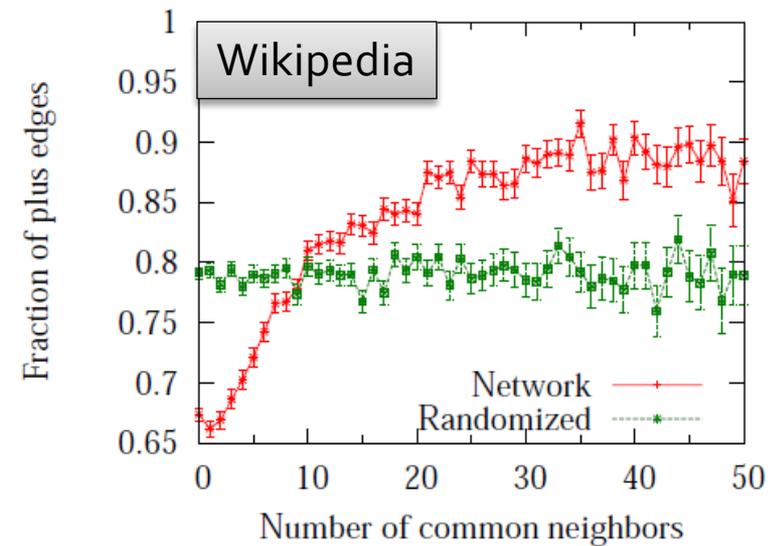
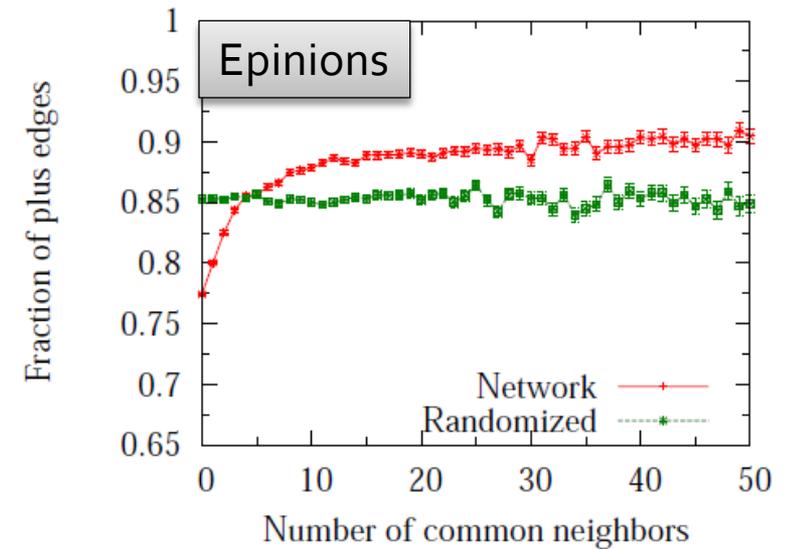
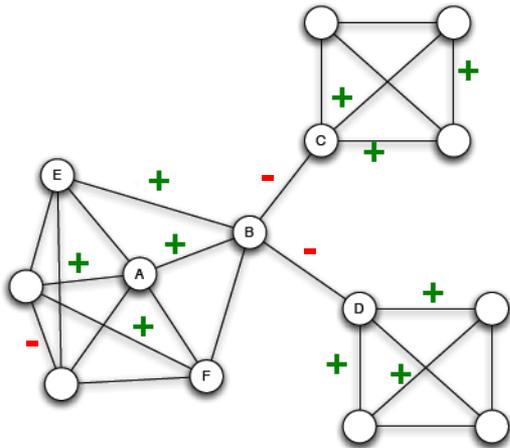
Global Structure of Signed Nets

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



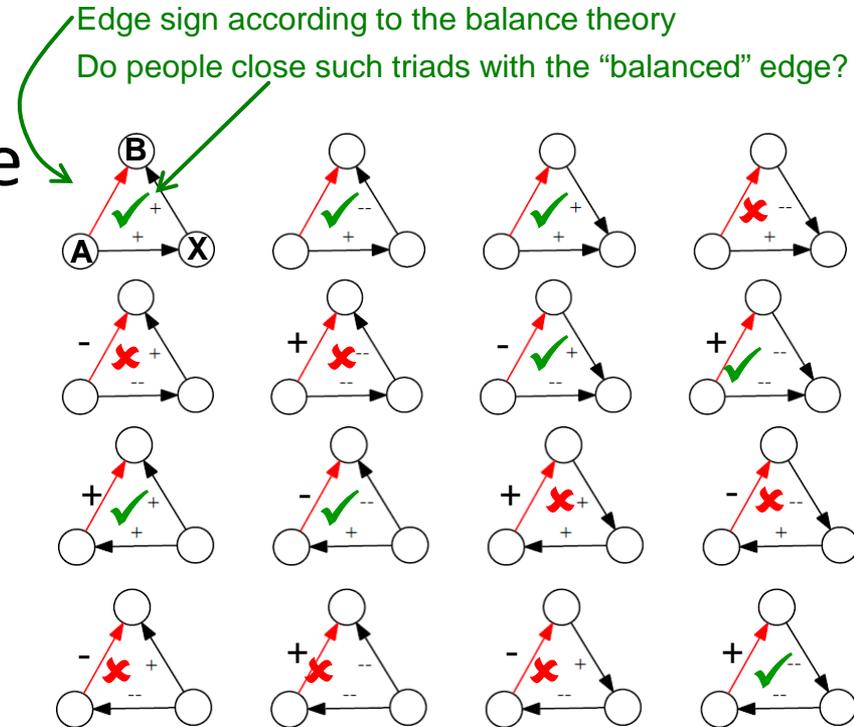
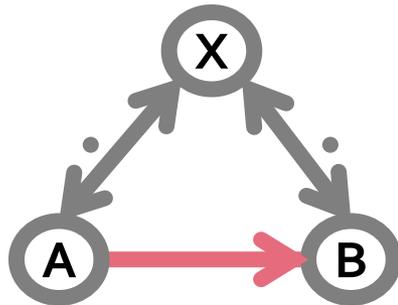
Global Factions: Embeddedness

- **Embeddedness of ties:**
 - Positive ties tend to be **more** embedded



Evolving Directed Networks

- **New setting:** Links are **directed**, created over time
 - Node **A** links to **B**
 - Directions and signs of links from/to X provide context



16 signed directed triads

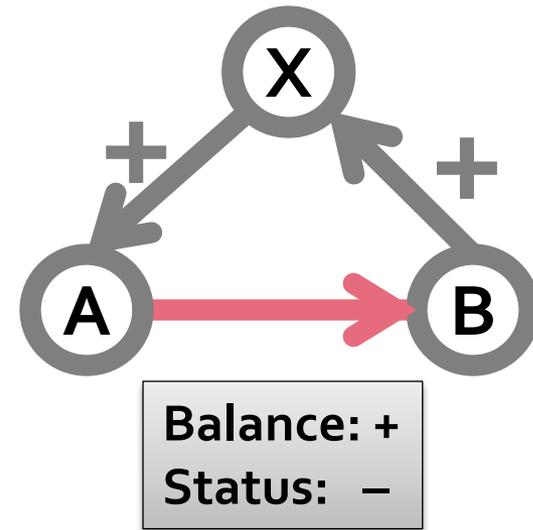
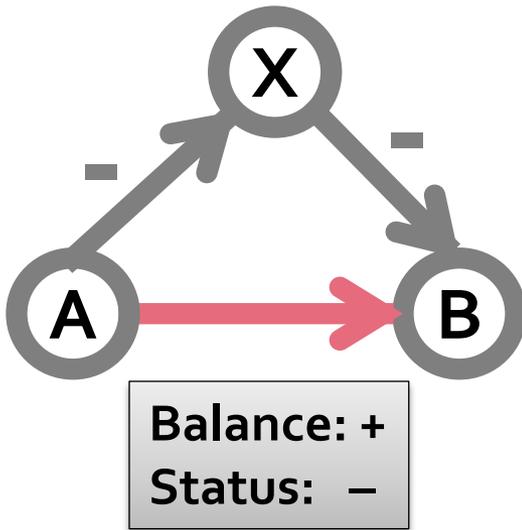
(in directed networks people traditionally applied balance by ignoring edge directions)

- How many \triangle are now explained by balance?
 - **Only half** (8 out of 16)

Alternate Theory: Status

- **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 and governed by the network (rather than the number of edits)
- **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

Status vs. Balance



Status and balance give
different predictions!

Status vs. Balance

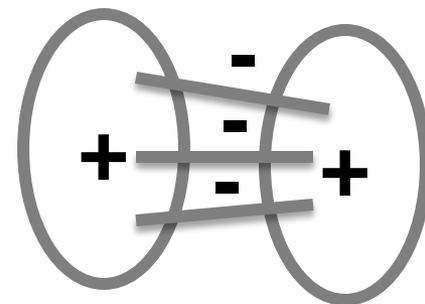
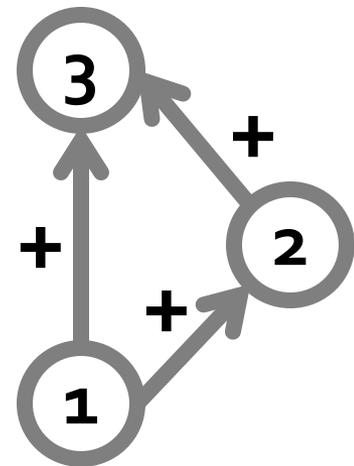
At a global level (in the ideal case):

- **Status \Rightarrow Hierarchy**

- All-positive directed network should be approximately **acyclic**

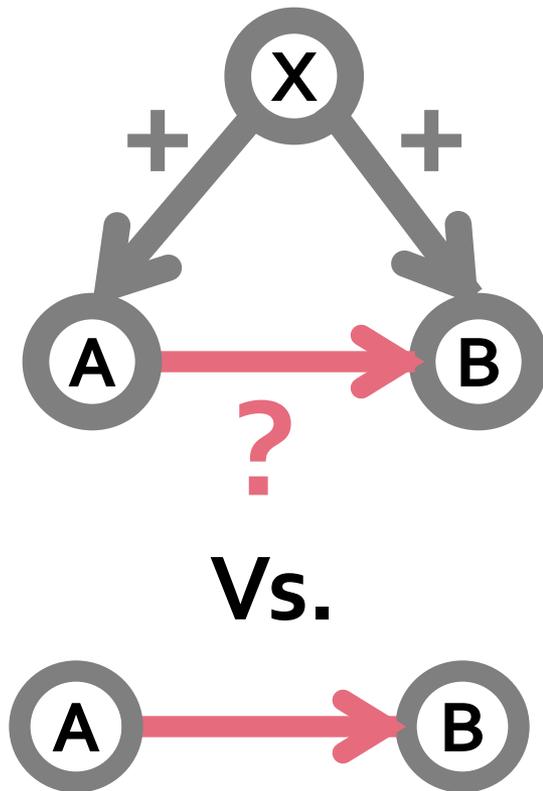
- **Balance \Rightarrow Coalitions**

- Balance ignores directions and implies that subgraph of negative edges should be approximately **bipartite**



Theory of Status

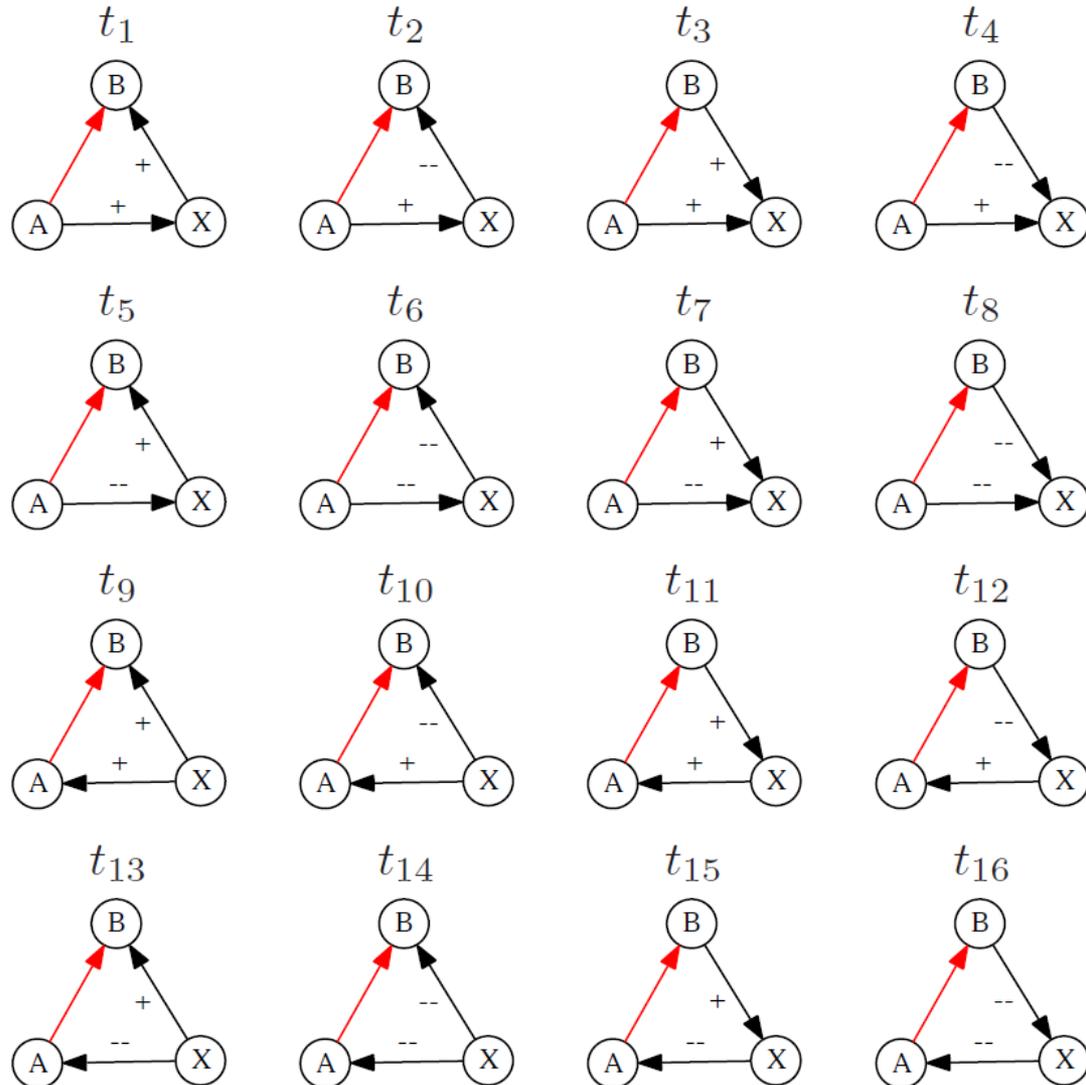
- Edges are **directed**:
 - X has links to A and B
 - Now, A links to B (triad A-B-X)
 - How does sign of $A \rightarrow B$ depend signs from/to X?
 $P(A \overset{+}{\rightarrow} B \mid X)$ vs. $P(A \rightarrow B)$
- We need to formalize:
 - 1) Links are **embedded in triads**:
Triads provide **context for signs**
 - 2) Users are **heterogeneous** in their **linking behavior**



1) Context: 16 Types

- Link $A \rightarrow B$ appears in context X :
 $A \rightarrow B \mid X$

- 16 possible contexts:



Note: Context of a link is uniquely determined by the directions and signs of links from/to X

2) Heterogeneity in linking behavior

- Users differ in frac. of + links they give/receive
- For a user U:
 - Generative baseline: Frac. of + given by U
 - Receptive baseline: Frac. of + received by U

Basic question:

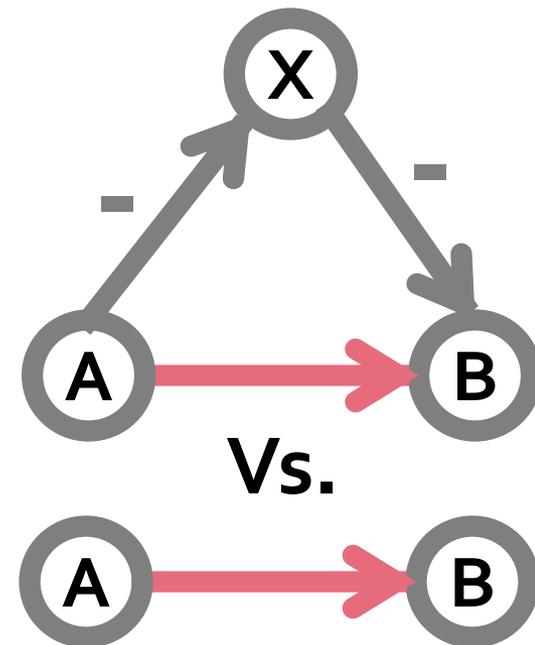
- How do different link contexts cause users to deviate from their baselines?
 - Link contexts as modifiers on a person's predicted behavior
 - Def: Surprise: How much behavior of A/B deviates from his/her baseline when A/B is in context X

Computing Surprise

- **Surprise:** How much behavior of user **deviates** from **baseline** in **context X**

- **Baseline:** For every user A_i :
 $p_g(A_i)$... **generative baseline** of A_i
 - Fraction of times A_i gives a plus
- **Context:** $(A_1, B_1 | X_1), \dots, (A_n, B_n | X_n)$
 ... all instances of triad context **X**
 - (A_i, B_i, X_i) ... an instance where when user A_i links to user B_i the triad of type **X** is created.
 - Say k of those triads closed with a plus
 - k out of n times: $A_i \xrightarrow{+} B_i$

Context X:

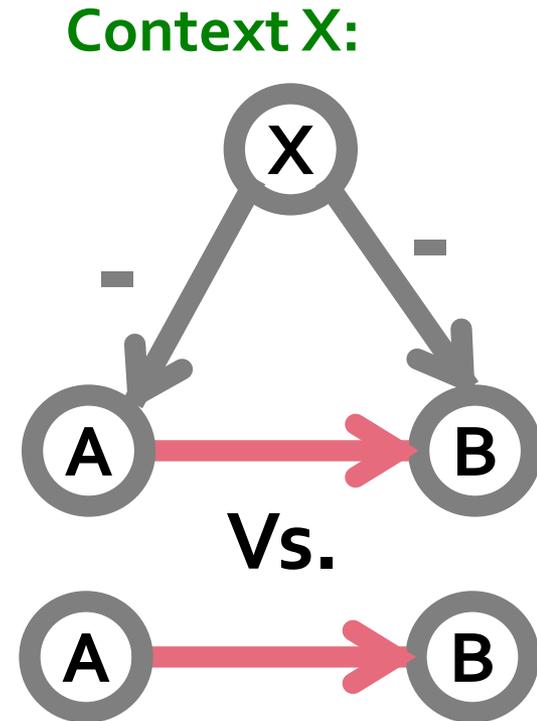


Computing Surprise

- **Surprise:** How much behavior of users **deviates** from **baseline** when in **context X**
- **Generative surprise of context X:**

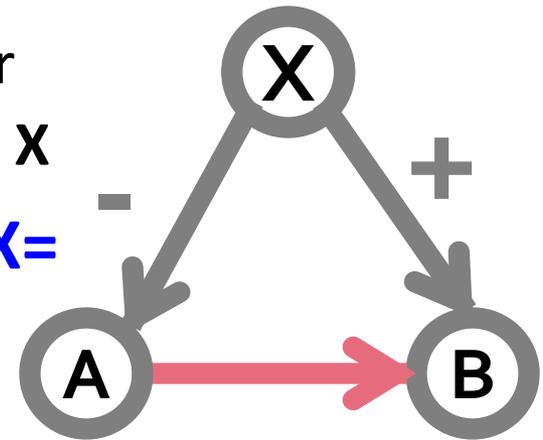
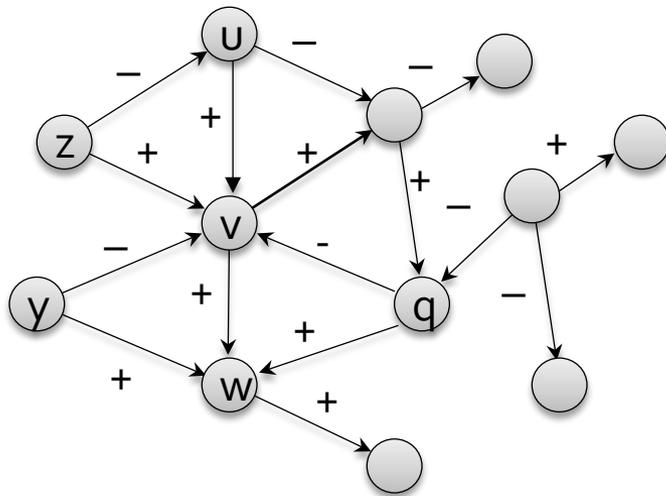
$$s_g(X) = \frac{k - \sum_{i=1}^n p_g(A_i)}{\sqrt{\sum_{i=1}^n p_g(A_i)(1 - p_g(A_i))}}$$

- $p_g(A_i)$... generative baseline of A_i
- **Context X:** $(A_1, B_1 | X_1), \dots, (A_n, B_n | X_n)$
- k of instances of triad X closed with a plus edges
- Receptive surprise is similar, just use $p_r(A_i)$



Example: Computing Surprise

- **Surprise:** How much behavior of user **deviates** from **baseline** when in **context X**
 - **Generative surprise of context X=**



$$s_g(X) = \frac{k - \sum_{i=1}^n p_g(A_i)}{\sqrt{\sum_{i=1}^n p_g(A_i)(1 - p_g(A_i))}}$$

We have 3 triads of context X: (z,u,v), (y,v,w), (q,v,w)

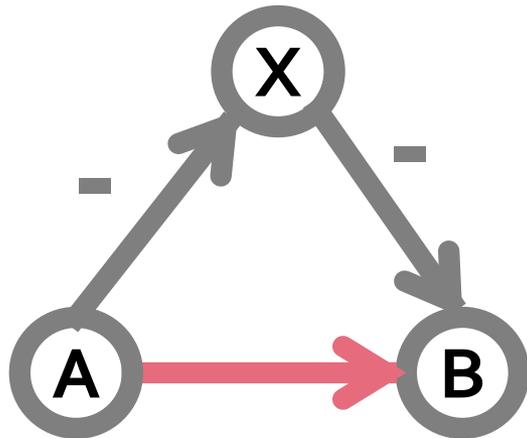
They all close with a plus: So $k=3$

$$P_g(u)=1/2=0.5 \quad P_g(v)=2/2=1$$

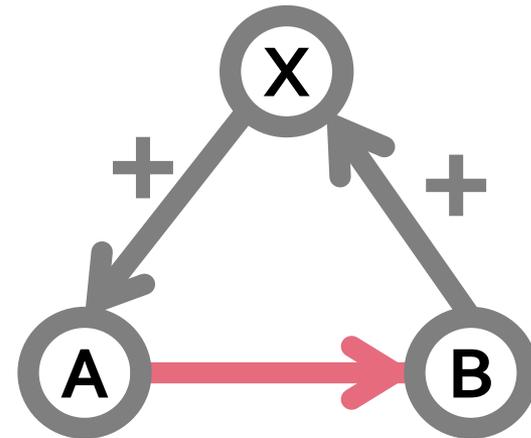
$$S_g(X)=(3-2.5)/\sqrt{(0.5*0.5+1*0+1*0)} = 1$$

Status: Two Examples

- Assume status is at work
- What happens?



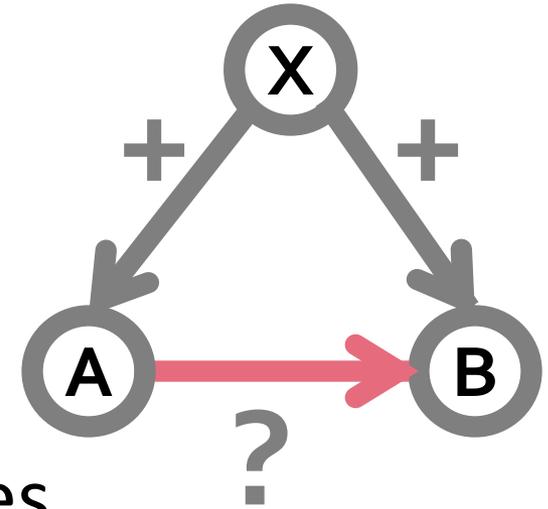
Gen. surprise of A: –
Rec. surprise of B: –



Gen. surprise of A: –
Rec. surprise of B: –

Joint Positive Endorsement

- X positively endorses A and B
- Now A links to B

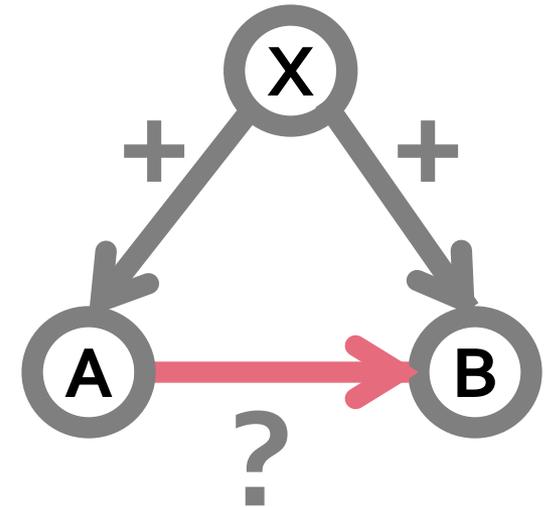


A puzzle:

- In our data we observe:
Fraction of positive links deviates
 - Above generative baseline of A: $S_g(X) > 0$
 - Below receptive baseline of B: $S_r(X) < 0$
- Why?

A Story: Soccer Team

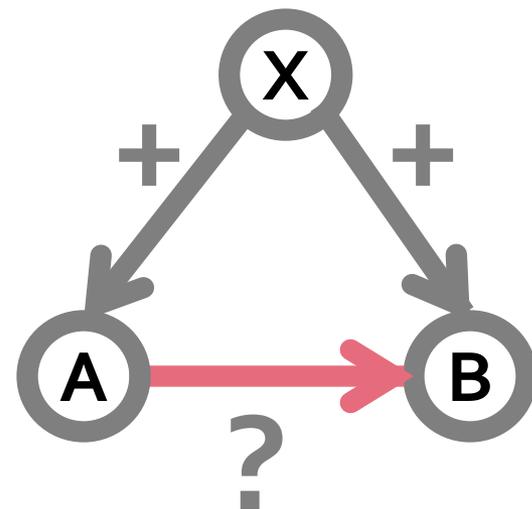
- **Ask every node: How does skill of B compare to yours?**
 - Build a signed directed network
- We haven't asked **A** about **B**
- But we know that **X** thinks **A** and **B** are both better than him
- **What can we infer about A's answer?**



A Story: Soccer Team

- **A's viewpoint:**

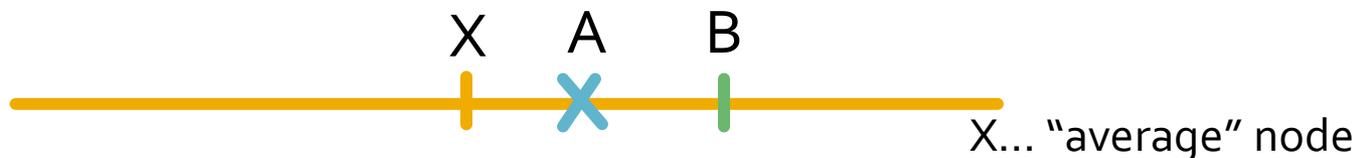
- Since **B** has positive evaluation, **B** is high status
- Thus, evaluation **A** gives is **more likely to be positive** than the baseline



How does A evaluate B?

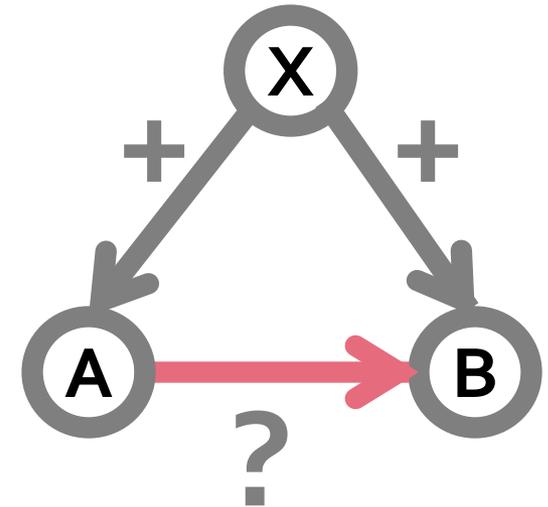
A is evaluating someone who is better than avg.

→ A is **more positive than average**



A Story: Soccer Team

- **B's viewpoint:**
 - Since **A** has positive evaluation, **A** is high status
 - Thus, evaluation **B** receives is **less likely to be positive** than the baseline



How is B evaluated by A?

B is evaluated by someone better than average.

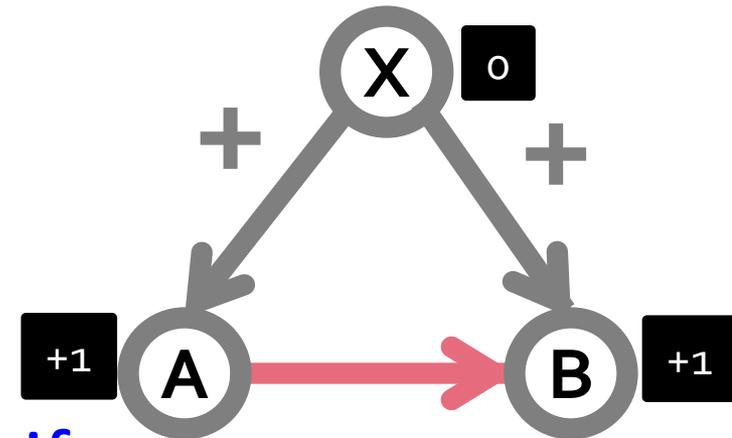
→ They will be **more negative to B than average**

Surprise of $A \rightarrow B$ deviates in different directions depending on the viewpoint!

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

Status-consistent if:

Gen. surprise > 0

Rec. surprise < 0

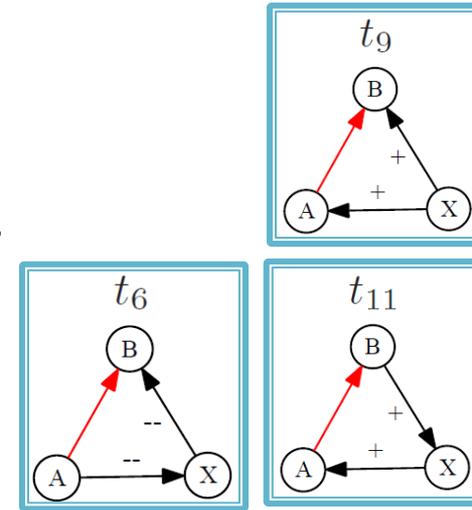
- Surprise is **balance-consistent**, if:

- If it completes a balanced triad

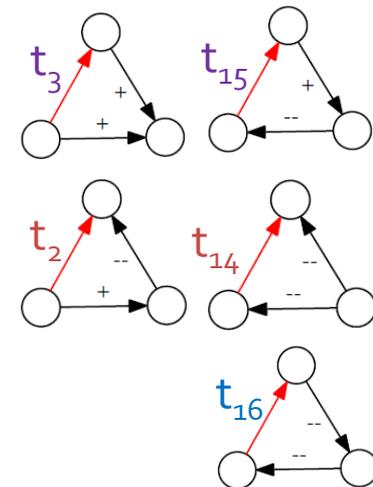
Status vs. Balance (Epinions)

Predictions by status and balance:

t_i	count	$P(+)$	$S_g(t_i)$	$S_r(t_i)$	B_g	B_r	S_g	S_r
t_1	178,051	0.97	95.9	197.8	✓	✓	✓	✓
t_2	45,797	0.54	-151.3	-229.9	✓	✓	✓	●
t_3	246,371	0.94	89.9	195.9	✓	✓	●	✓
t_4	25,384	0.89	1.8	44.9	○	○	✓	✓
t_5	45,925	0.30	18.1	-333.7	○	✓	✓	✓
t_6	11,215	0.23	-15.5	-193.6	○	○	✓	✓
t_7	36,184	0.14	-53.1	-357.3	✓	✓	✓	✓
t_8	61,519	0.63	124.1	-225.6	✓	○	✓	✓
t_9	338,238	0.82	207.0	-239.5	✓	○	✓	✓
t_{10}	27,089	0.20	-110.7	-449.6	✓	✓	✓	✓
t_{11}	35,093	0.53	-7.4	-260.1	○	○	✓	✓
t_{12}	20,933	0.71	17.2	-113.4	○	✓	✓	✓
t_{13}	14,305	0.79	23.5	24.0	○	○	✓	✓
t_{14}	30,235	0.69	-12.8	-53.6	○	○	✓	●
t_{15}	17,189	0.76	6.4	24.0	○	○	●	✓
t_{16}	4,133	0.77	11.9	-2.6	✓	○	✓	●
Number of correct predictions					8	7	14	13



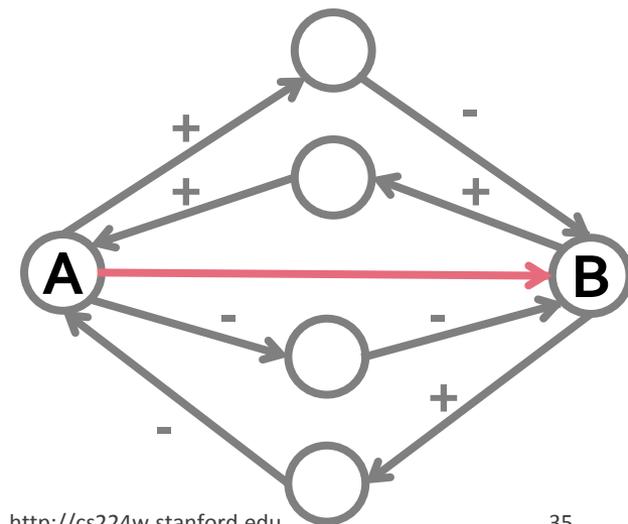
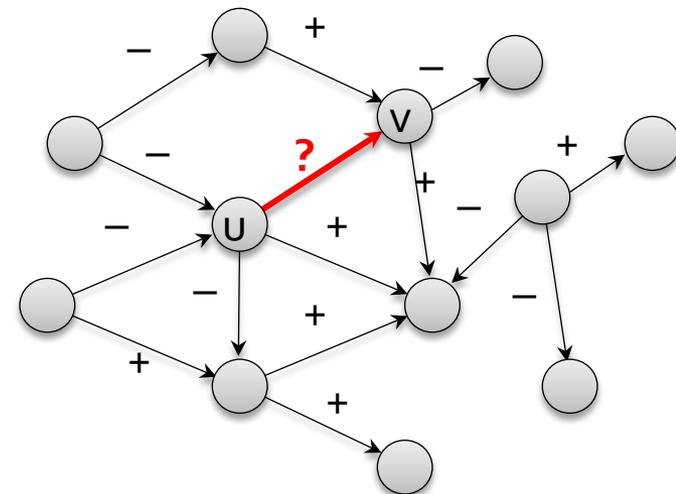
Mistakes:



Predicting Edge Signs

Edge sign prediction problem

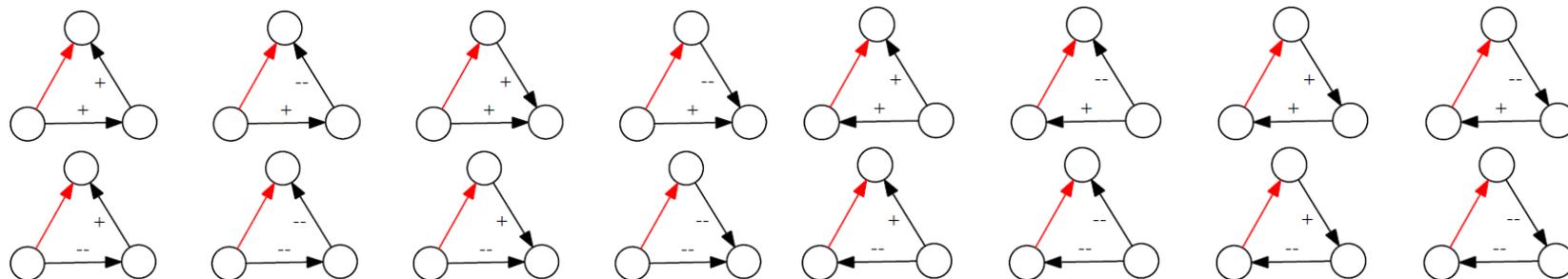
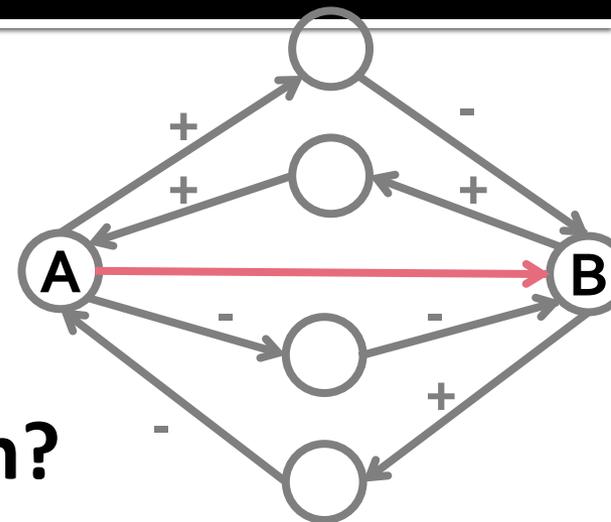
- Given a network and signs on all but one edge, predict the missing sign
- **Friend recommendation:**
 - Predicting whether you know someone vs. Predicting what you think of them
- **Setting:**
 - Given edge (A,B) , predict its sign:
 - Let's look at signed triads (A,B) belongs to:



Features for Learning

For the edge (A,B) we examine
Its network context:

- In what types of triads does our red-edge participate in?



- Each triad then “votes” and we determine the sign

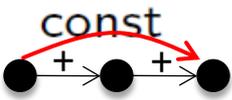
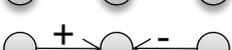
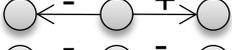
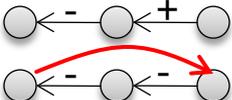
Balance and Status: Complete Model

Triad	Bal
	1
	-1
	-1
	1
	1
	-1
	-1
	1
	1
	-1
	-1
	1
	1
	-1
	-1
	1

Balance and Status: Complete Model

Triad	Bal	Stat
	1	1
	-1	0
	-1	0
	1	-1
	1	0
	-1	1
	-1	-1
	1	0
	1	0
	-1	-1
	-1	1
	1	0
	1	-1
	-1	0
	-1	0
	1	1

Balance and Status: Complete Model

Triad	Bal	Stat	Epin	Slashd	Wikip
	1	1	-0.2 0.5	0.02 0.9	-0.2 0.3
	-1	0	-0.5	-0.9	-0.4
	-1	0	-0.4	-1.1	-0.3
	1	-1	-0.7	-0.6	-0.8
	1	0	0.3	0.4	0.05
	-1	1	-0.01	-0.1	-0.01
	-1	-1	-0.9	-1.2	-0.2
	1	0	0.04	-0.07	-0.03
	1	0	0.08	0.4	0.1
	-1	-1	-1.3	-1.1	-0.4
	-1	1	-0.1	-0.2	0.05
	1	0	0.08	-0.02	-0.1
	1	-1	-0.09	-0.09	-0.01
	-1	0	-0.05	-0.3	-0.02
	-1	0	-0.04	-0.3	0.05
	1	1	-0.02	0.2	-0.2

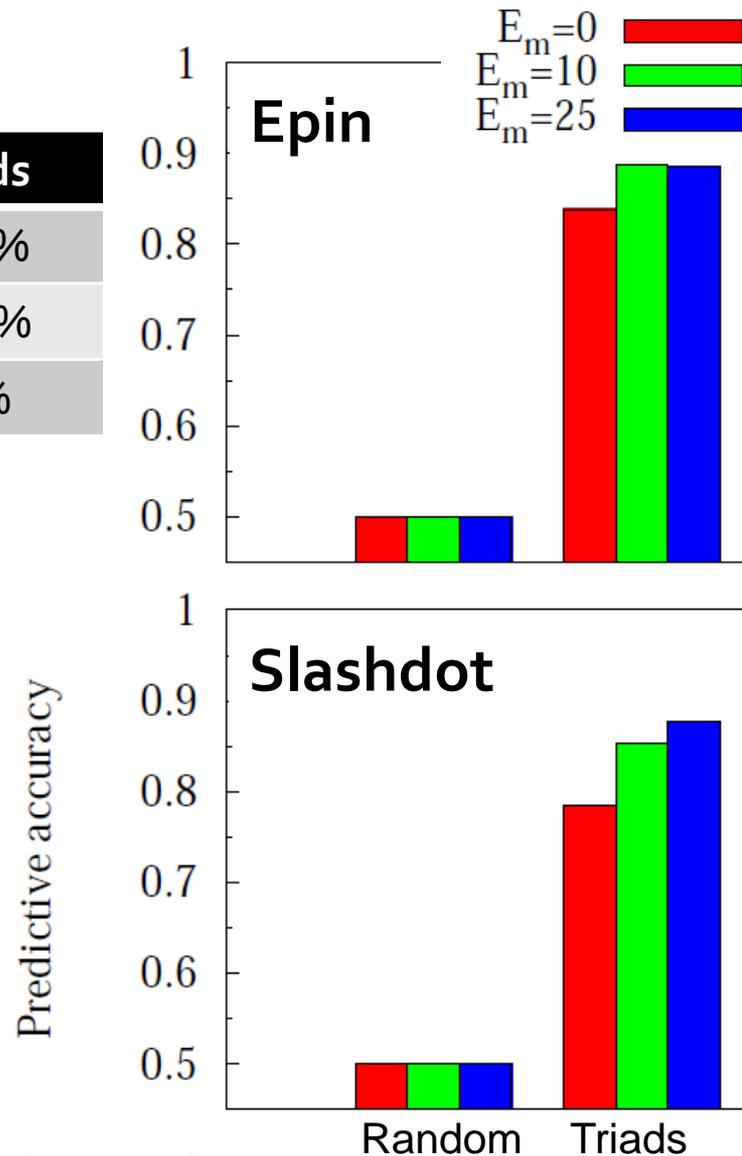
Edge Sign Prediction

Prediction accuracy:

	Balance	Status	Triads
Epinions	80%	82%	93.5%
Slashdot	84%	72%	94.4%
Wikipedia	64%	70%	81%

Observations:

- Signs can be modeled from local network structure alone!
- Triad counts perform less well for less embedded edges (E_m)
 - Wikipedia is harder to model:
 - Votes are publicly visible



Generalization

- Do people use these very different linking systems by obeying the same principles?
 - How generalizable are the results across the datasets?

Train on row, test on column	Epinions	Slashdot	Wikipedia
Epinions	0.9342	0.9289	0.7722
Slashdot	0.9249	0.9351	0.7717
Wikipedia	0.9272	0.9260	0.8021

- Nearly **perfect generalization** of the models even though networks come from very different applications!

Summary: Signed Networks

- Signed networks provide insight into how social computing systems are used:
 - Status vs. Balance
 - Role of embeddedness and public display
 - More evidence that networks are globally organized based on status
- Sign of relationship can be reliably predicted from the local network context
 - ~90% accuracy sign of the edge
 - People use signed edges consistently regardless of particular application
 - Near perfect generalization of models across datasets



What about the effect of evaluations on the target T?

CS224W: Social and Information Network Analysis
Jure Leskovec, Stanford University
<http://cs224w.stanford.edu>



Facebook privacy now defaults to friends only



By **Doug Gross**, CNN

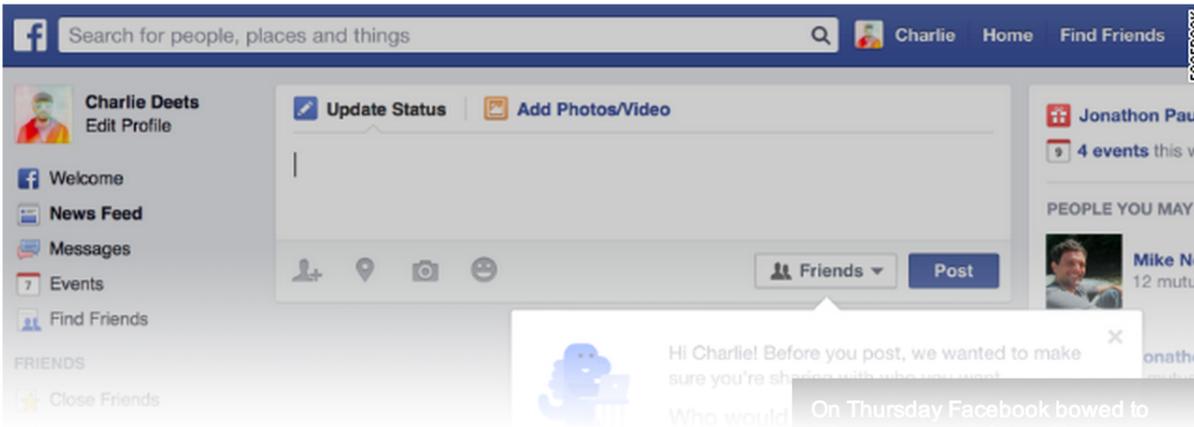
updated 3:39 PM EDT, Thu May 22, 2014 | Filed under: [Social Media](#)

SHARE THIS



- Print
- Email
- More sharing

Recommend 235



More from CNN Video:



Serial killer's daughter recalls horrors



Widow hugs husband's killer in court

35 Comments

CNN

Login

Sort by Best

Share Favorite



Join the discussion

Sort by Best

Share Favorite



Join the discussion...



Tom · 7 hours ago

If you're posting something to facebook, it shouldn't be anything you wouldn't print and tape to the front door of a local grocery store.

21 ^ | v · Reply · Share



ccw101 → Tom · 7 hours ago

I hate Facebook for the fact the only person you have control over is yourself. I have seen full grown adults get angry at their own children and rip them a new one on their Facebook home page!

If adults can be so ST***id then what do kids do?

Facebook is scary. And has given people the opportunity to use it to cause home break in's, ruined reputations, fights , suicides etc.

8 ^ | v · Reply · Share



IAmNotATroll → ccw101 · 7 hours ago

Come now, I thoroughly enjoy watching my in-laws publicly argue and shred each other to pieces over Facebook.

8 ^ | v · Reply · Share



Furby → IAmNotATroll · 6 hours ago

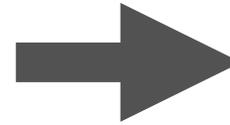
I had a distant cousin try to blackmail my mom on FB publically. Me and her became real close after that - and not in the way you want to get close to someone. Some people are just plain dumb

How do people react to evaluations they receive?

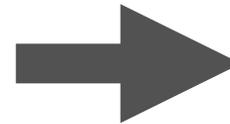
How does positive/negative feedback influence subsequent user behavior?



Positively
Evaluated



Negatively
Evaluated



Evaluations can affect

Post quality (How well you write)

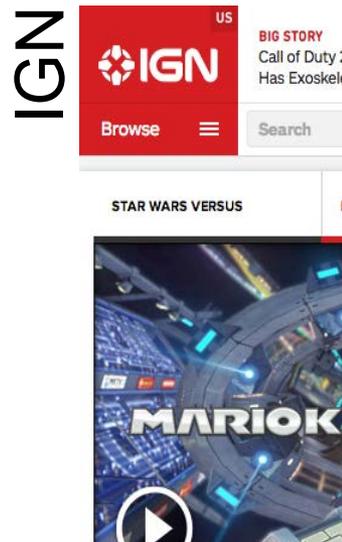
Community bias (How people perceive you)

Posting frequency (How regularly you post)

Voting behavior (How you vote on others)

Four large comment-based news communities with

1.2M articles, 1.8M registered users, 42M posts, 140M votes, 1 year



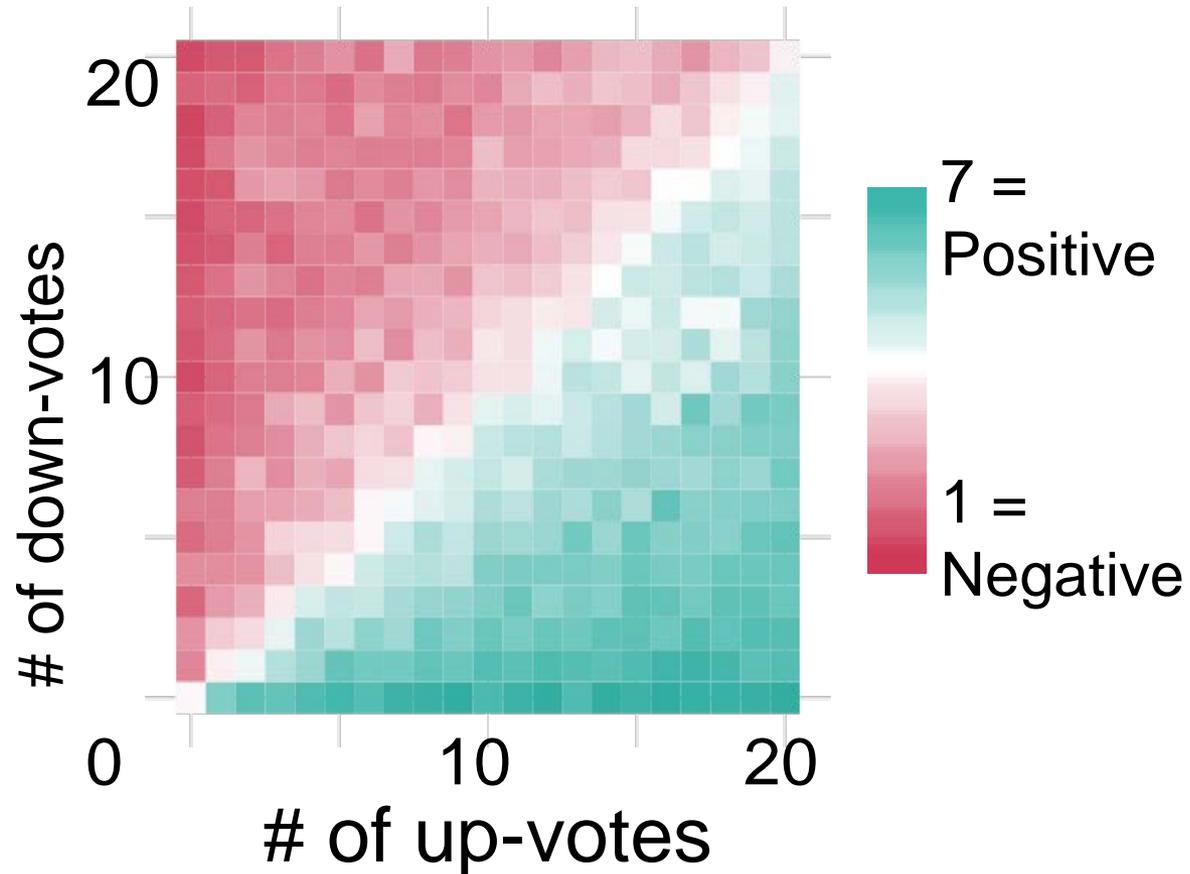
How do we measure community feedback?

Number of up-votes

Up-votes minus Down-votes

Fraction of up-votes

User ratings were independent of the total number of votes



Fraction of up-votes: $R^2=0.92$

What happens after you
give a user a positive or a
negative

**Wish:
Randomized
controlled trial**

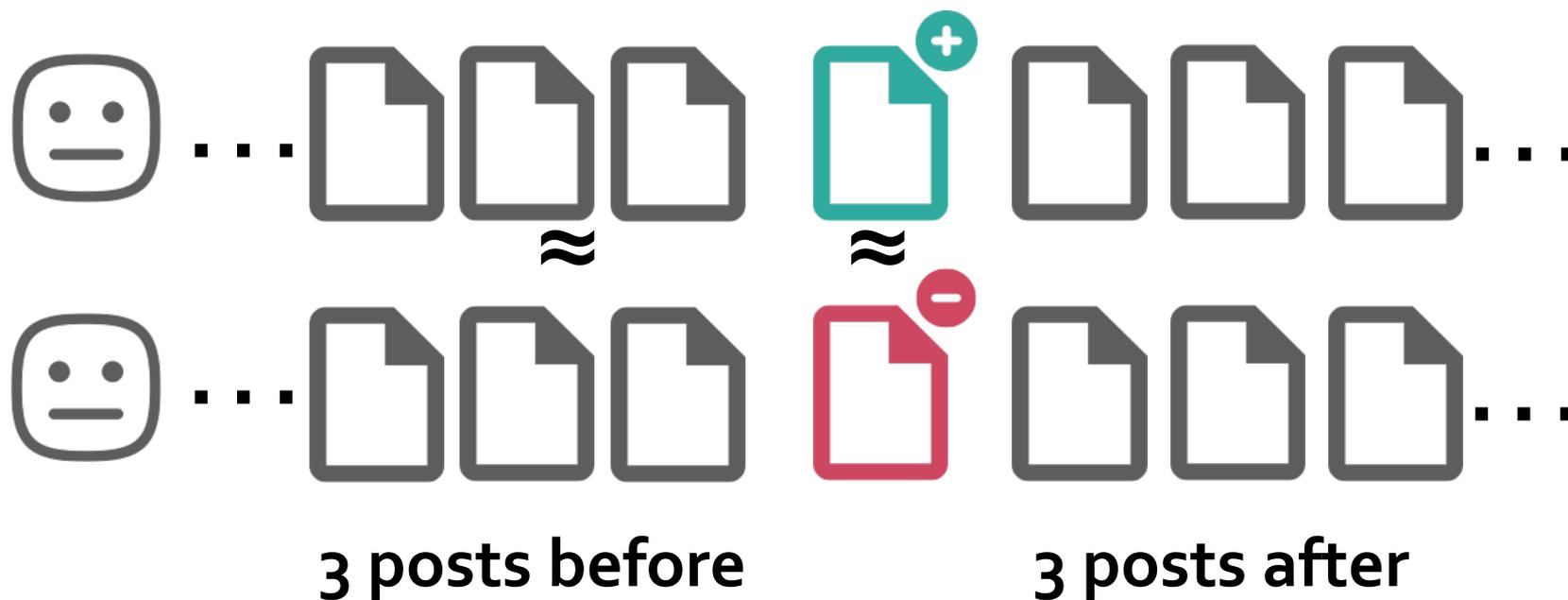


...



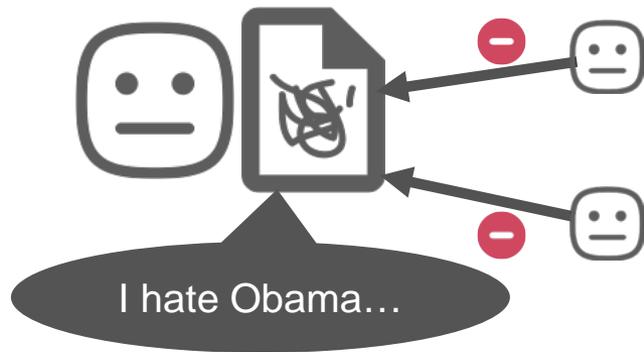
...

Compare similar pairs of users who were evaluated differently on similar content

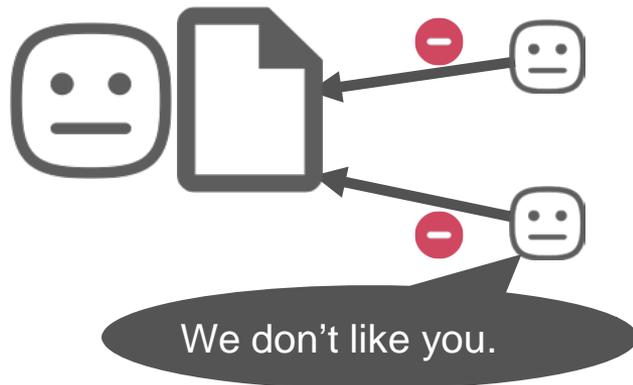


Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects.

How much of a future evaluation can be explained by textual, or community effects?



Down-voting because of **post content**



Down-voting because the **community** dislikes the author

Textual Effects

Text quality drops significantly after a negative evaluation, but does not change after a positive evaluation

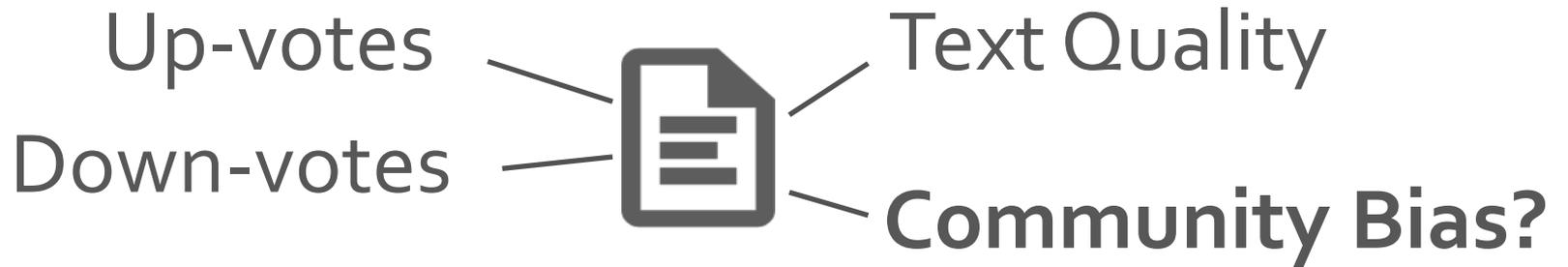
$p < 0.05$ in all communities

To learn more about these types of effects, see Kanouse, D. E., & Hanson Jr, L. R. (1987). Negativity in evaluations.

Evaluations can affect

Community bias (How people perceive you)

**How does community
perception of a user change
after an evaluation?**



Actual Evaluation $P/(P+N)$ 0.9

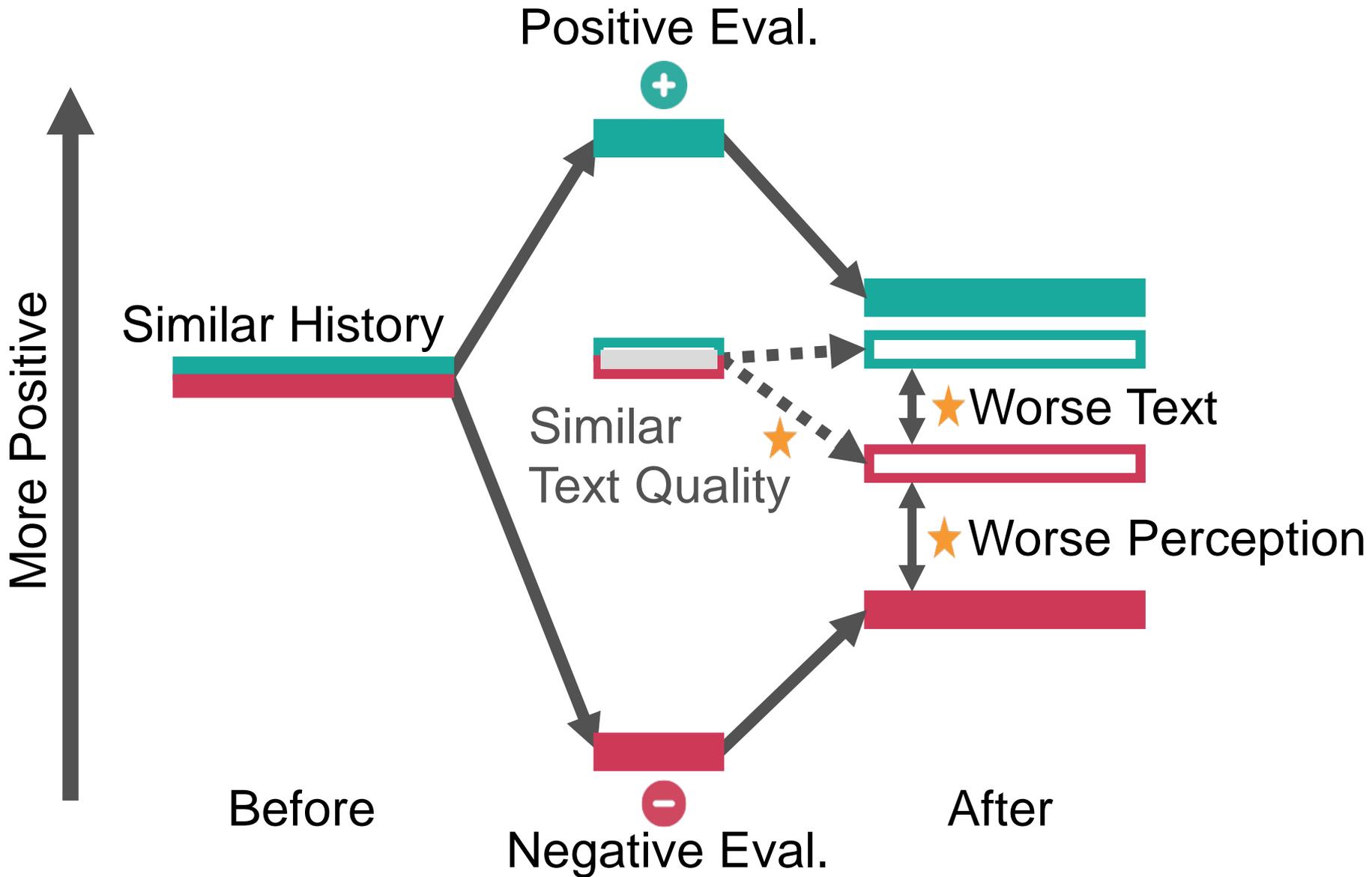
Text Quality 0.8

**Community Bias $0.9-0.8$
 $= +0.1$**

Community Effects

Posts made after a negative evaluation were perceived worse than those made after a positive evaluation

$p < 0.05$ in all communities

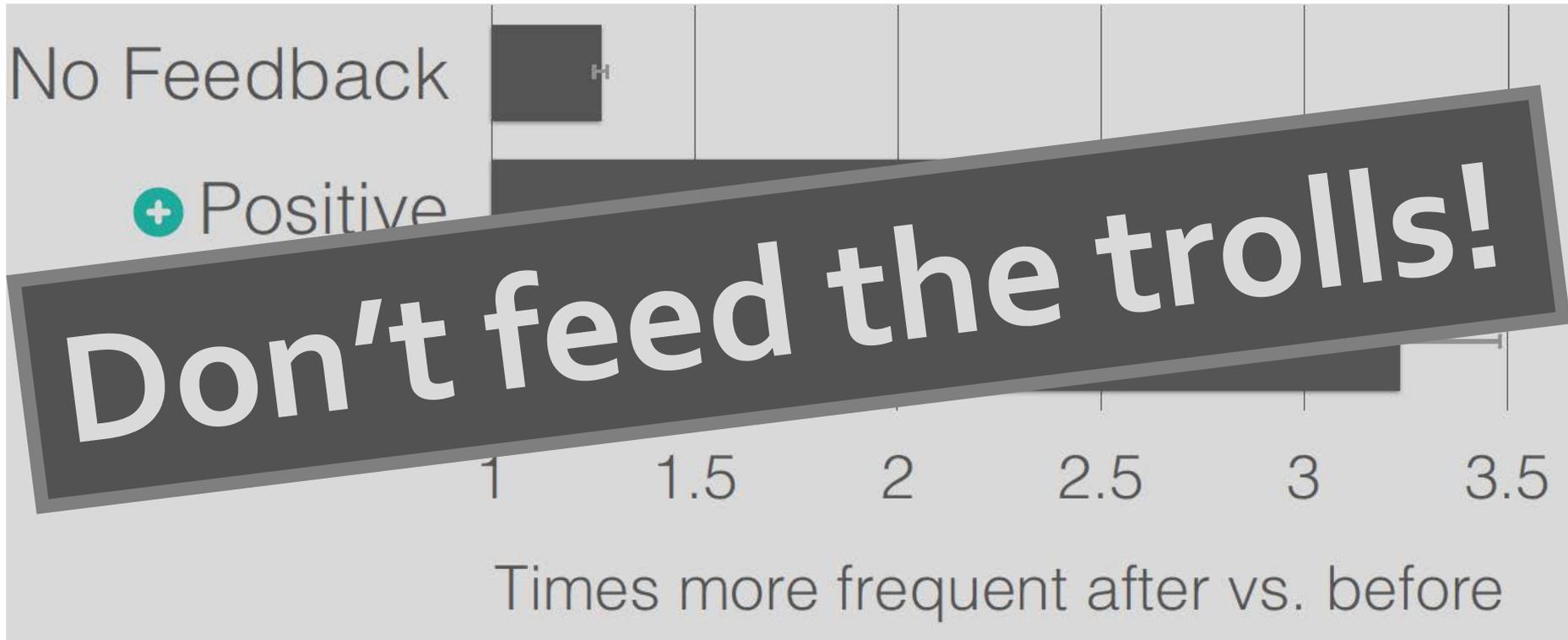


Evaluations can affect

Posting frequency (How regularly you post)

**Does feedback regulate
post *quantity*?**

Users who receive negative feedback post more frequently

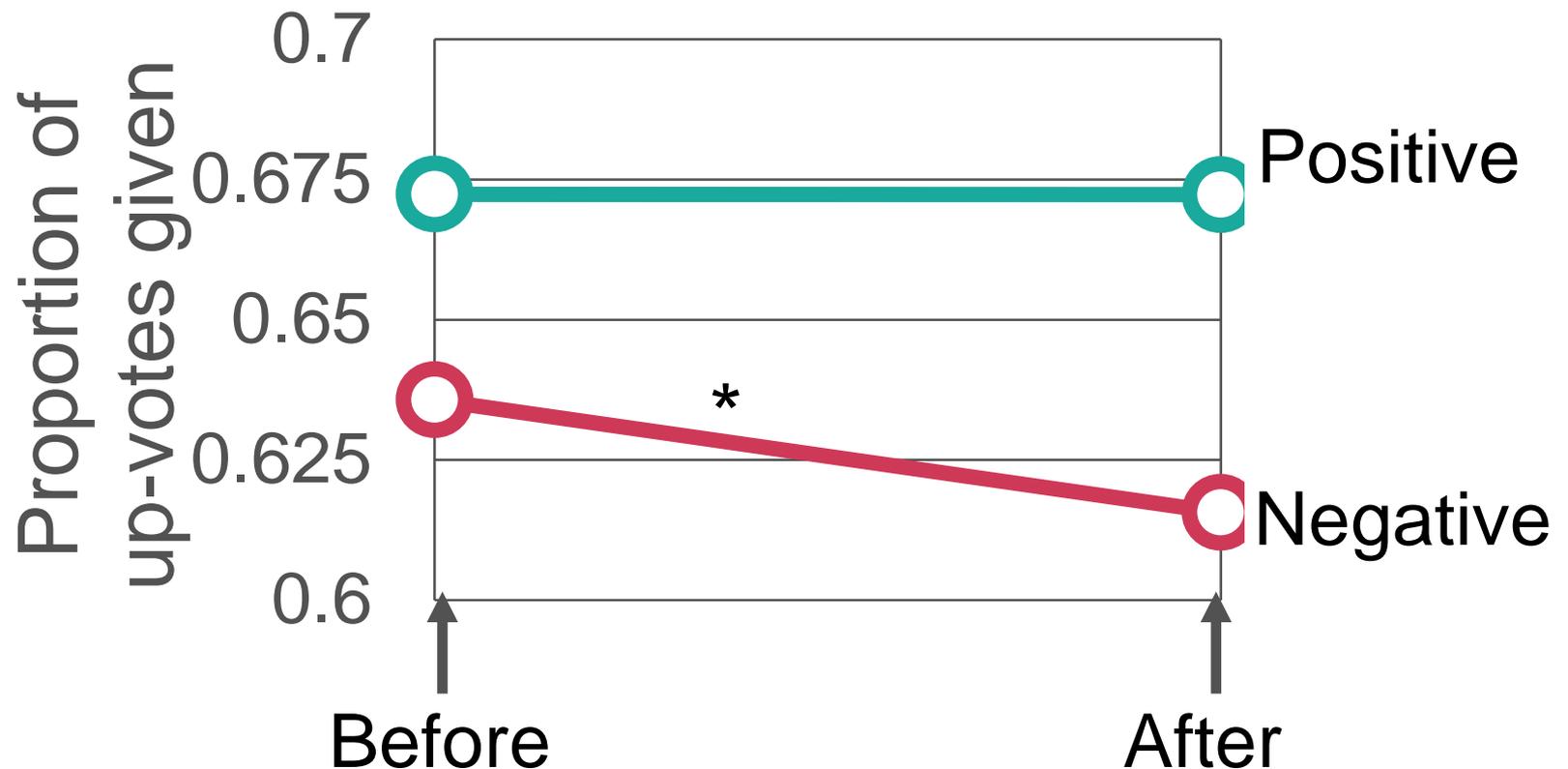


Evaluations can affect

Voting Behavior (How you vote on others)

**Does feedback result in
subsequent backlash?**

Users who receive negative feedback are more likely to down-vote others



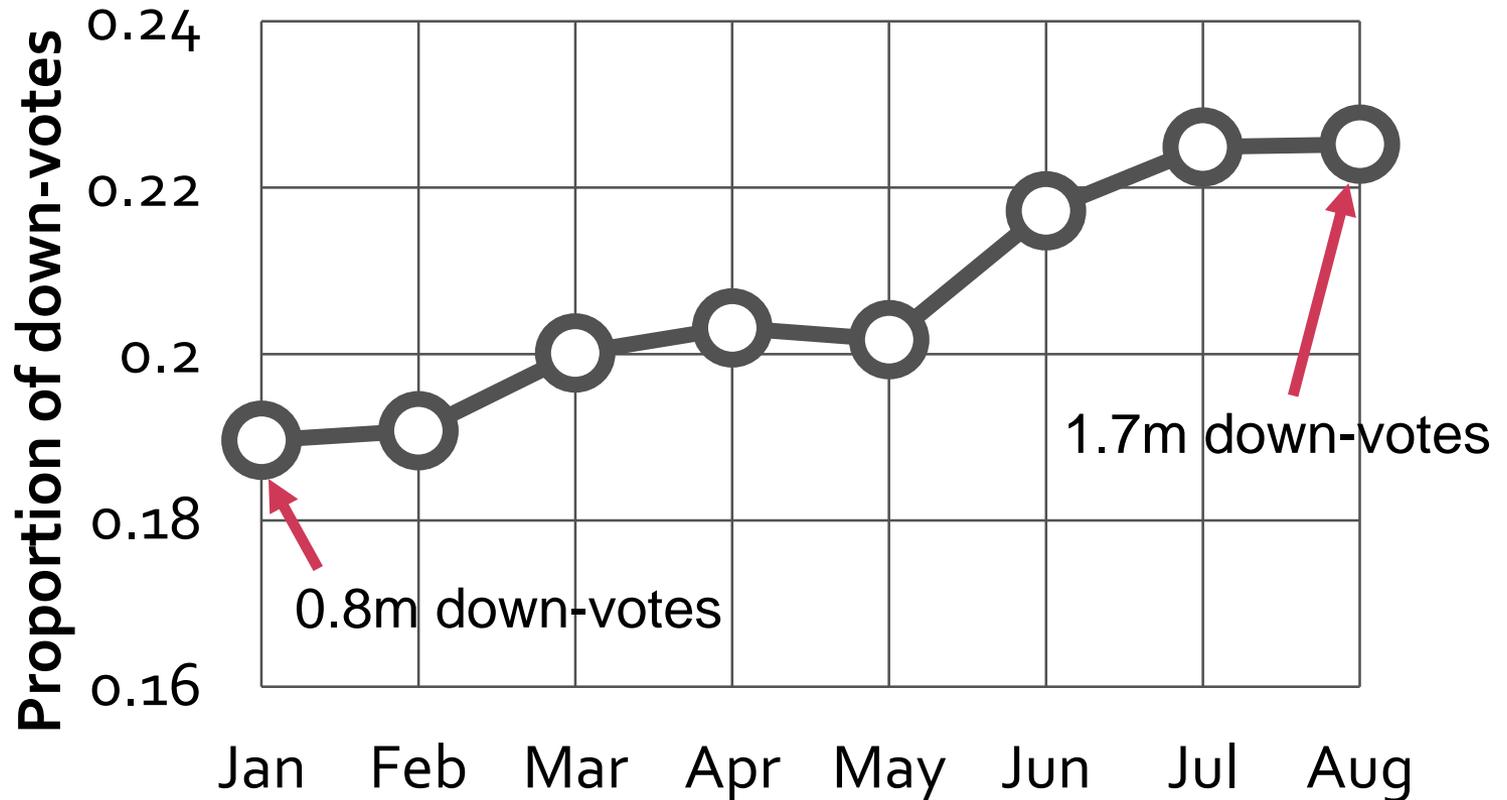
Conclusion

Negatively-evaluated users write worse (and more!), are themselves evaluated worse by the community, and evaluate other community members worse.

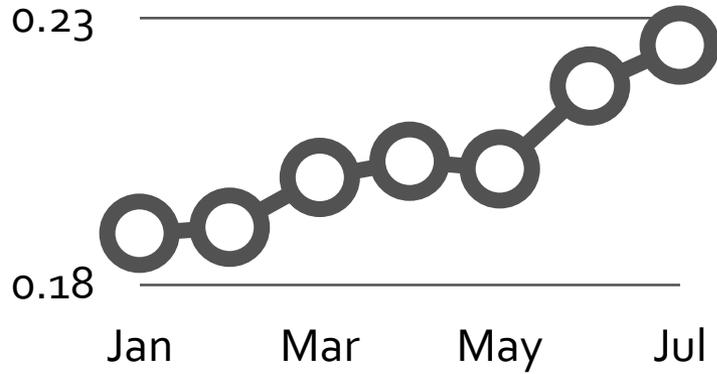
Positively-evaluated users, on the other hand, don't do any better.

**Is there a downward spiral
in online communities?**

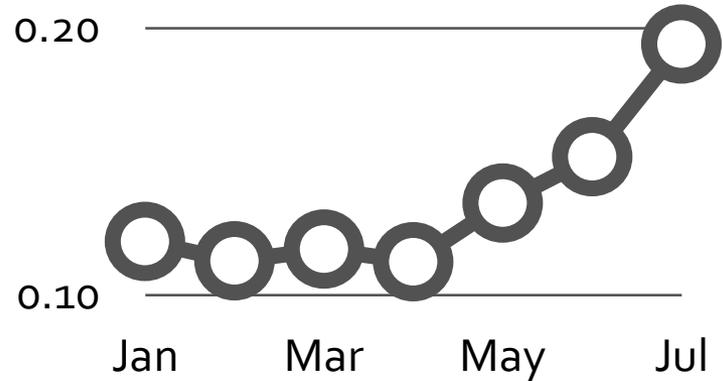
The proportion of down-votes is increasing over time



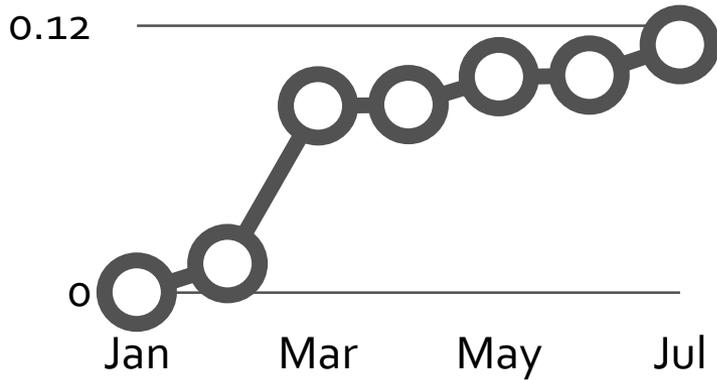
CNN



IGN



Breitbart



allkpop

