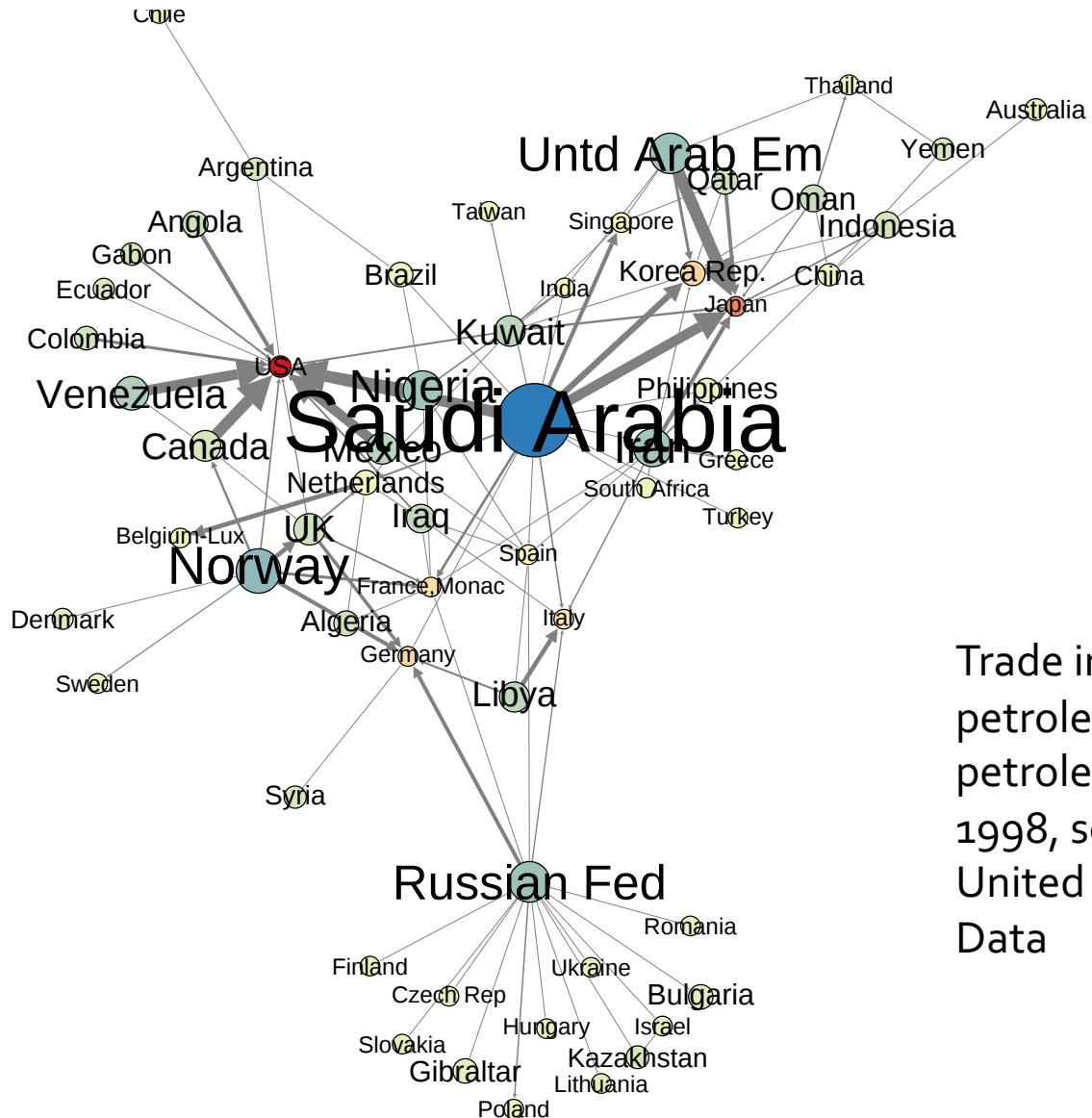


# **Applications in Social Network Analysis: Centrality**

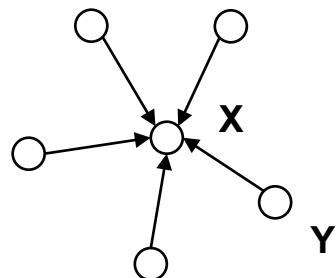
# Quick Topic: Centrality



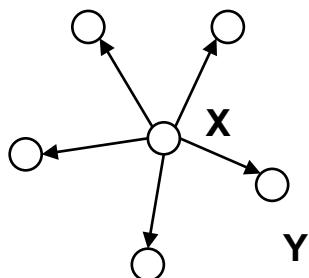
Trade in crude  
petroleum and  
petroleum products,  
1998, source: NBER-  
United Nations Trade  
Data

# Centrality

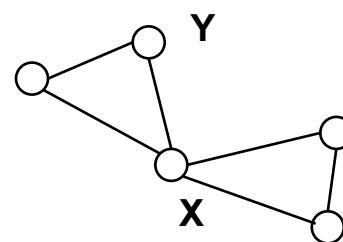
In each of the following networks, X has higher centrality than Y according to a particular measure



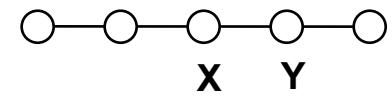
indegree



outdegree



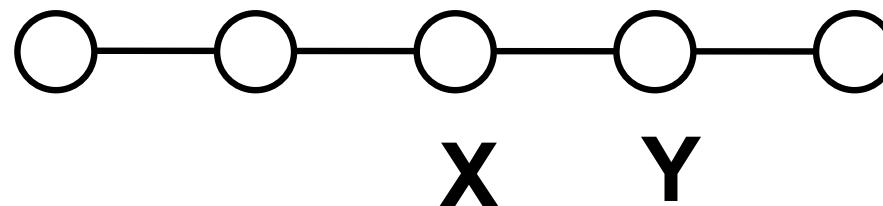
betweenness



closeness

# Betweenness: Capturing Brokerage

- **Intuition:** How many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?



# Betweenness: Definition

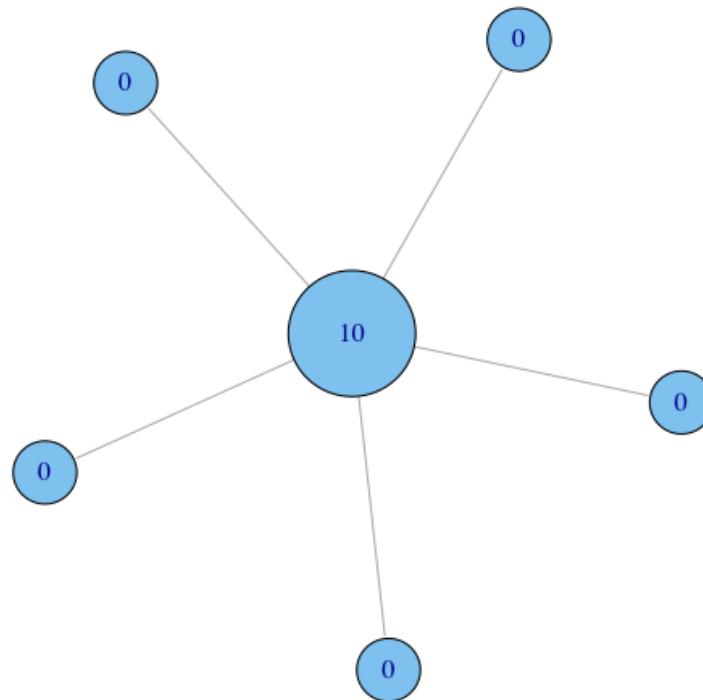
$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where  $\sigma_{st}(v)$  = the number of shortest paths  $s - t$  through node  $v$

$\sigma_{st}$  = the number of shortest paths from  $s$  to  $t$ .

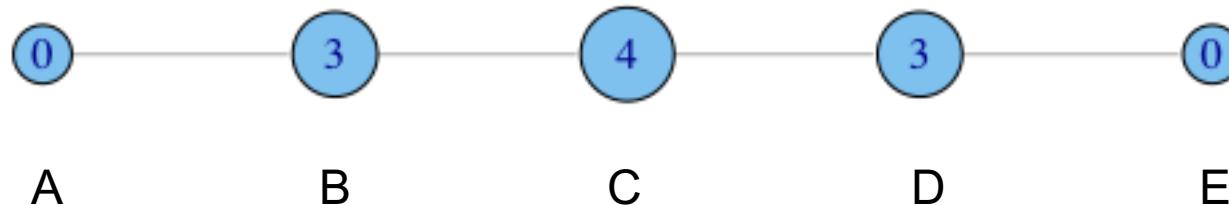
# Betweenness: Example (1)

- Non-normalized version of betweenness centrality (numbers are centralities):



# Betweenness: Example (2)

- Non-normalized version:



- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices  
(A,D),(A,E),(B,D),(B,E)
- Note that there are no alternate paths for these pairs to take, so C gets full credit

# Closeness Centrality

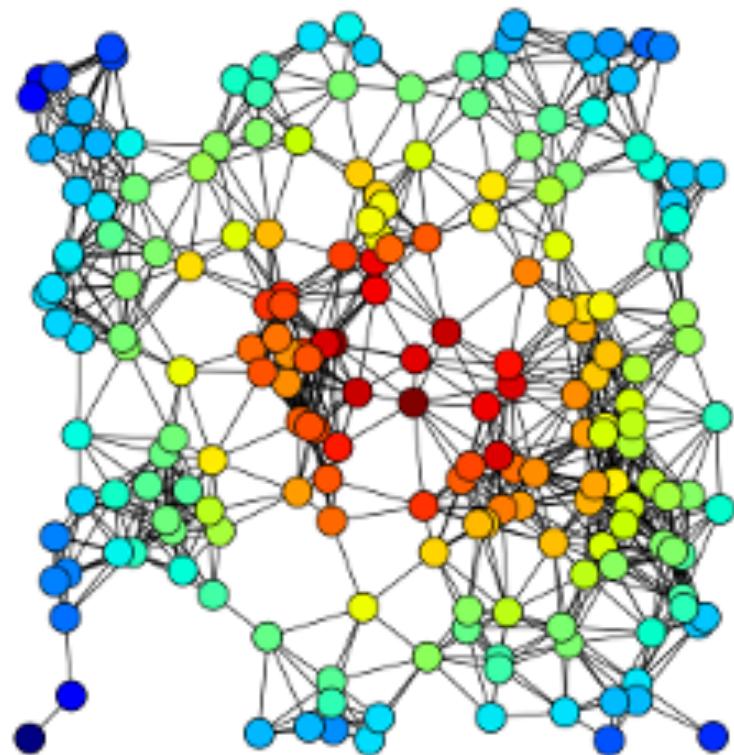
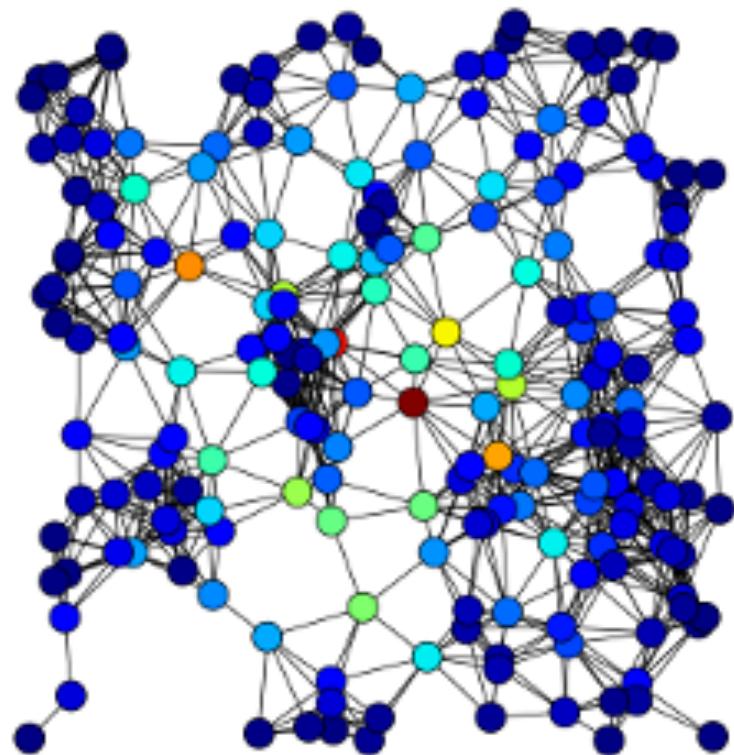
- **Closeness Centrality:** Reciprocal of the mean average shortest path length from node  $x$  to all other nodes in the graph  $y$ .

$$\frac{N}{\sum_y d(y, x)}.$$

- **Farness centrality:** Avg. shortest path length from node  $x$  to all other nodes

(we assume graph is connected)

# Closeness vs. Betweenness



- Betweenness (left), Closeness (right)

# **Human Evaluations and Signed Networks**

# High-level Overview of the Lecture

- We will talk about human behavior online
- We will try to understand how people express opinions about each other online
  - We will use data and network science theory to model factors around human evaluations
  - This will be an example of **Computational Social Science** research
    - We are making social science constructs quantitative and then use computation to measure them

# How the Class Fits Together

## Observations

Small diameter,  
Edge clustering

Patterns of signed  
edge creation

Viral Marketing, Biogosphere,  
Memetracking

Scale-Free

Densification power law,  
Shrinking diameters

Strength of weak ties,  
Core-periphery

## Models

Erdős-Renyi model,  
Small-world model

Structural balance,  
Theory of status

Independent cascade model,  
Game theoretic model

Preferential attachment,  
Copying model

Microscopic model of  
evolving networks

Kronecker Graphs

## Algorithms

Decentralized search

Models for predicting  
edge signs

Influence maximization,  
Outbreak detection, LIM

PageRank, Hubs and  
authorities

Link prediction,  
Supervised random walks

Community detection:  
Girvan-Newman, Modularity

# People Express Opinions

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

- Through actions:
  - Rating a product/person
  - 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
  - Sharing economy

amazon.com.



NETFLIX



lyft

YouTube



yelp

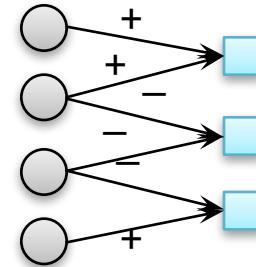
airbnb

IMDb  
Earth's Biggest Movie Database™

# People & Evaluations

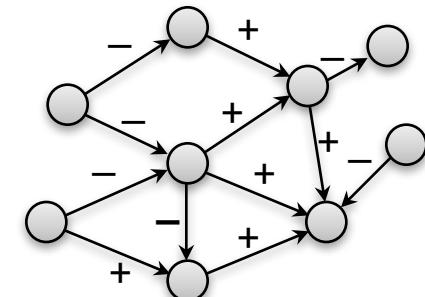
## ■ About items:

- Movie and product reviews



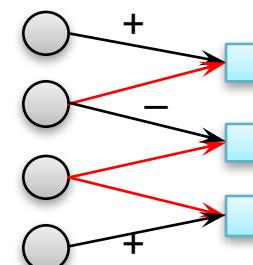
## ■ About other users:

- Online communities



## ■ About items created by others:

- Q&A websites



# User-User Evaluations

- Many online 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:

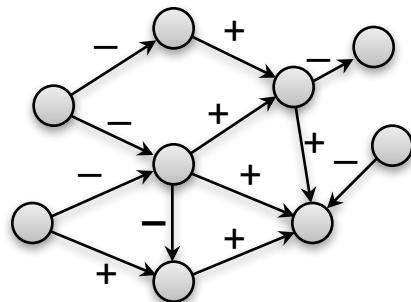
- **Factors:**

**What factors drive one's evaluations?**

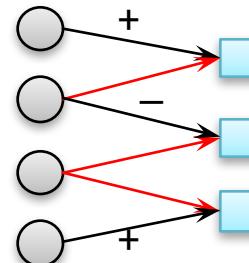
- **Synthesis:**

**How do we create a composite description  
that accurately reflects aggregate opinion of  
the community?**

# Evaluations: The Setting



Direct

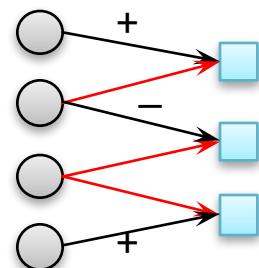
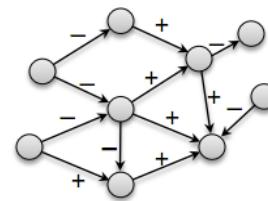


Indirect

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

# Evaluations: The 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

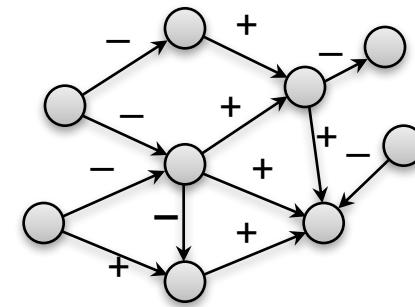


# Two ways to look at this

- There are two ways to look at this:  
**One person evaluates the other via a positive/negative evaluation**



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



Then we will focus on evaluations in the context of a network

# 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

## ■ **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:** Similarity of the articles edited
  - **Stack Overflow:** 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

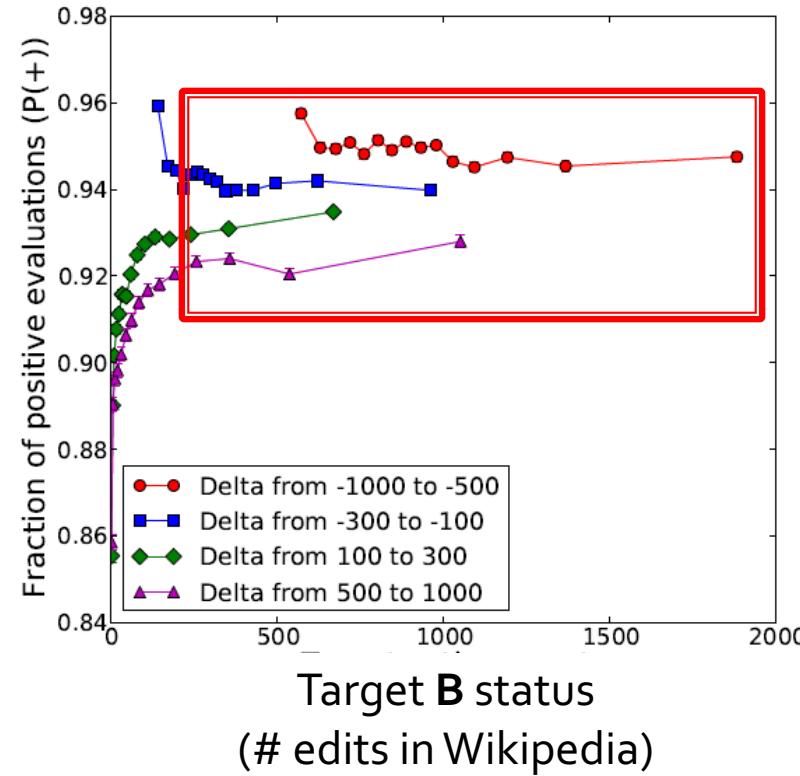
- How do properties of **evaluator A** and **target B** affect A's vote?



- **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: Wikipedia

- How does status of B affect A's evaluation?
  - Each curve is a 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  $\Delta$  result in different behavior

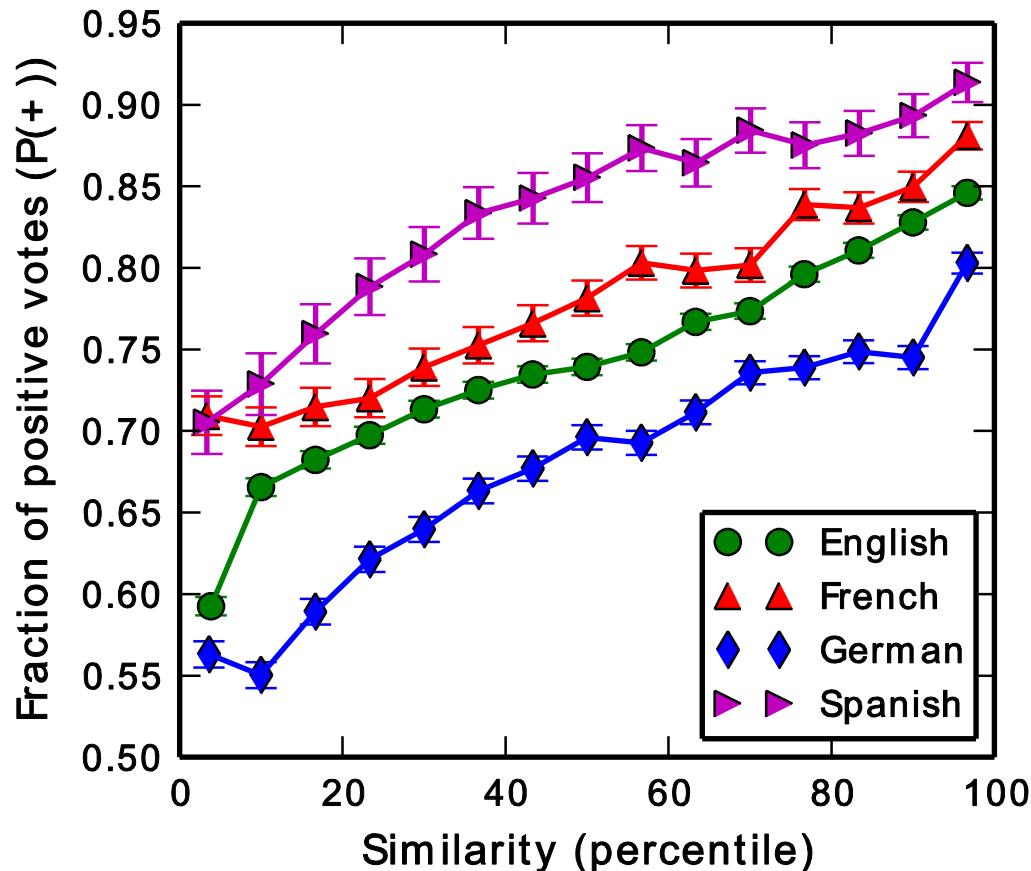


We keep increasing status of B, while keeping the status difference ( $S_A - S_B$ ) fixed

# 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”

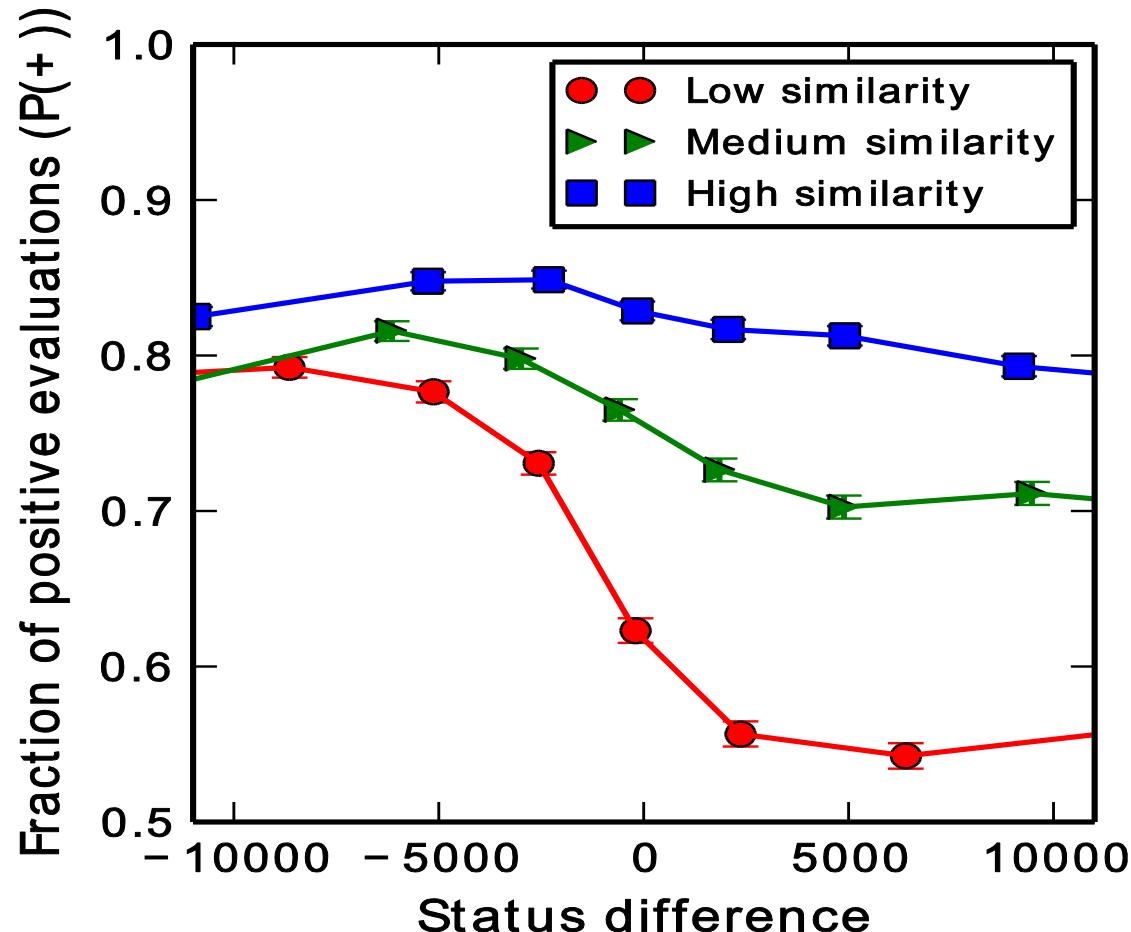
# Effects of Similarity



**Similarity:** For each user create a set of words of all articles she edited. The similarity is then the Jaccard similarity between the two sets of words. Then sort the user pairs by similarity and bucket them into percentiles.

Prior interaction/ similarity boosts positive evaluations

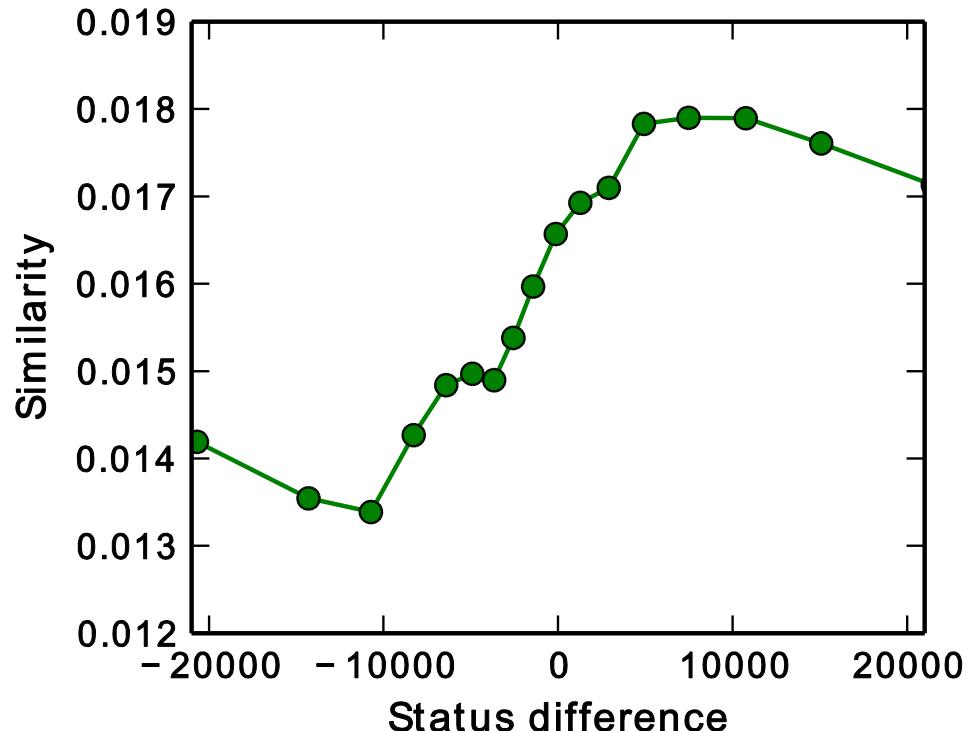
# Status & Similarity



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

# Status & Similarity

## ■ Who shows up to evaluate?

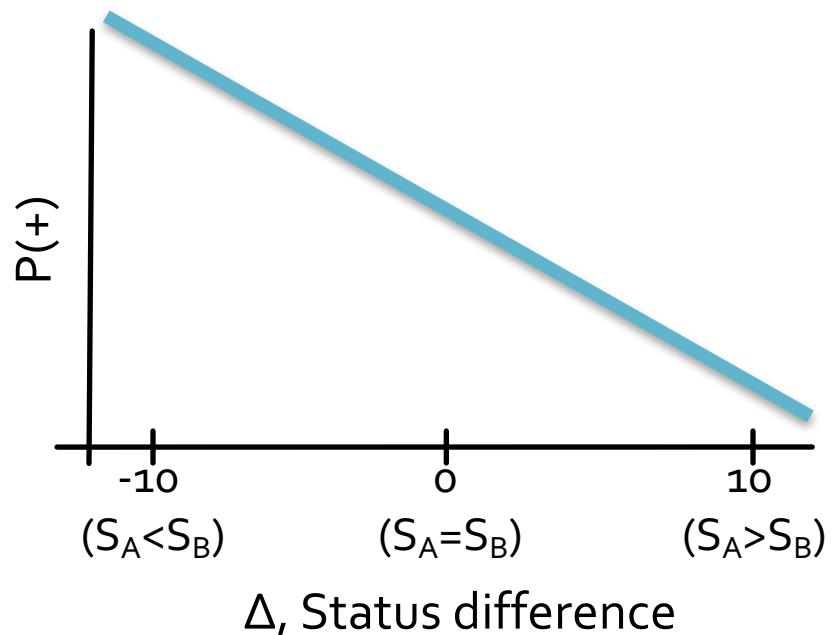


Elite evaluators  
vote on targets in  
their area of  
expertise

- Selection effect in who gives the evaluation
  - If  $S_A > S_B$  then A and B are more likely to be similar

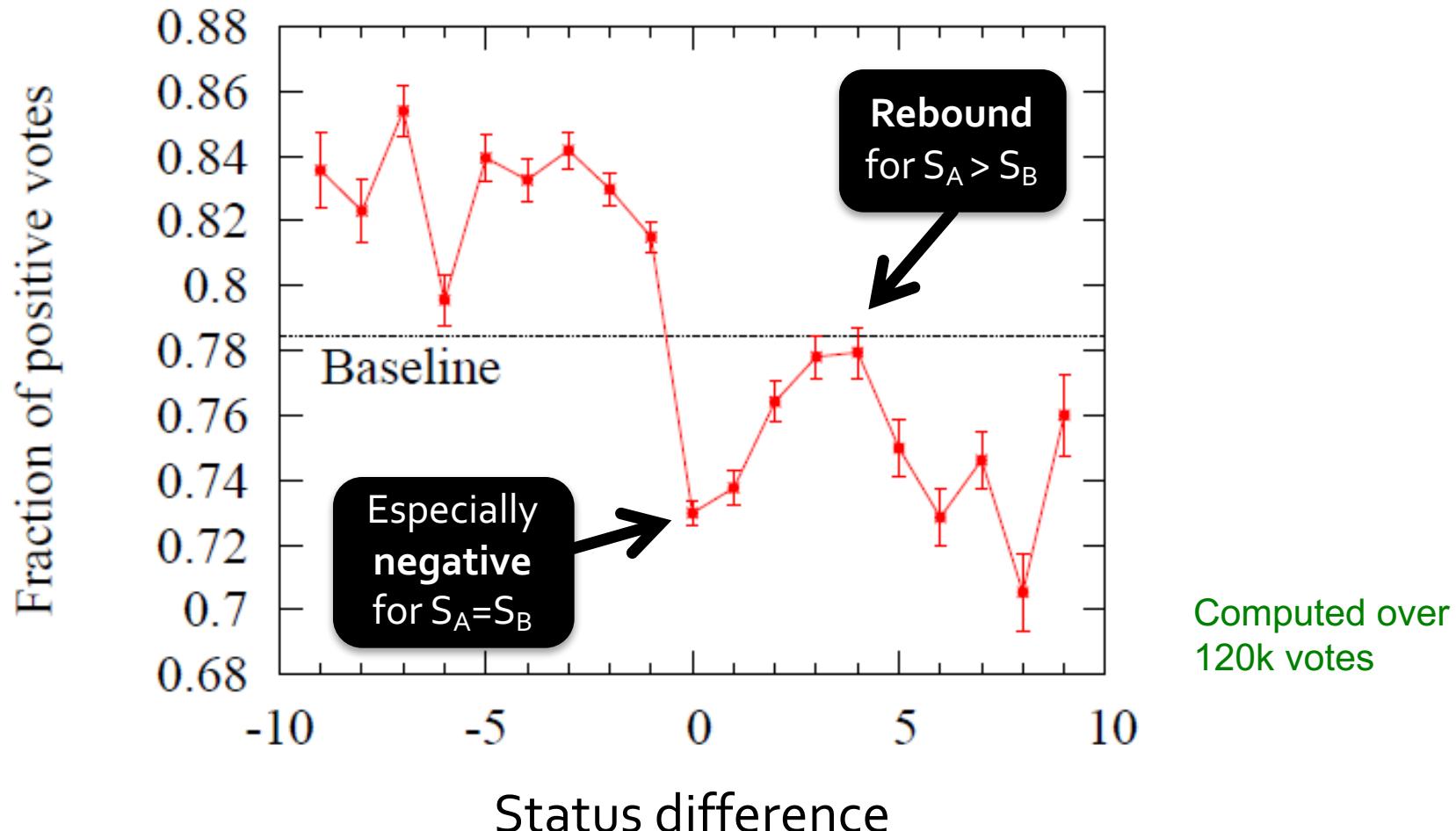
# A Puzzle

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



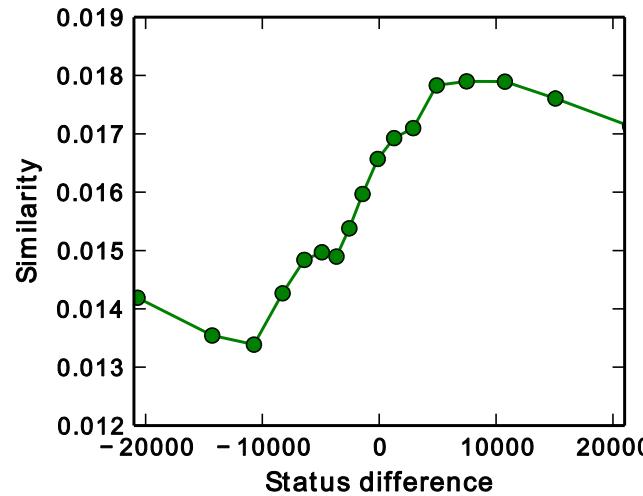
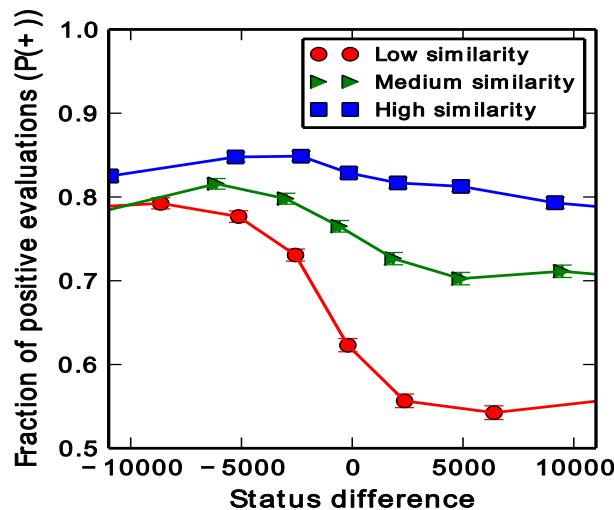
# A Puzzle: The Mercy Bounce

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



# 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



**Explanation:** For negative status difference we have low similarity people which behave according to the red curve on the left plot. As status difference increases the similarity also increases (green curve). For positive status difference, similarity is high, and evaluations follow the blue curve (left). By having a particularly weighted combination of red, green, and blue curve we observe the “mercy bounce” from the previous slide.

- So we get the “mercy” bounce due to uneven mixing of votes

# 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)



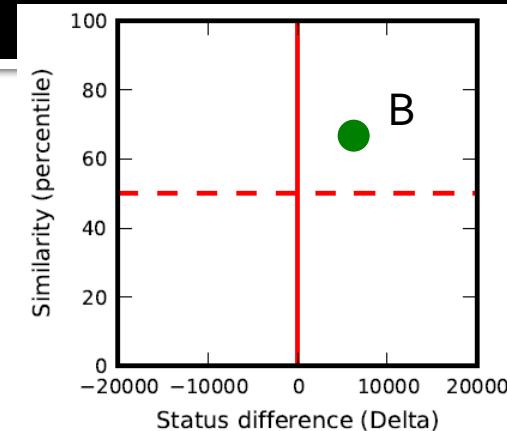
# Ballot-blind: The Model

- Want to model prob. user  $A$  votes + in election of user  $B$

- Our model:

$$P(A = +|B) = P_A + d(S_A - S_B, \text{sim}(A, B))$$

- $P_A$  ... empirical fraction of +votes of A
- $d(\text{status}, \text{similarity})$  ... avg. deviation in frac. of +votes
  - When  $A$  evaluates  $B$  from a particular  $(\text{status}, \text{similarity})$  quadrant, how does this change their behavior on average?
    - Note:**  $d(\text{status}, \text{similarity})$  only takes 4 different values (based on the quadrant in the  $(\text{status}, \text{similarity})$  space). Value computed empirically.
- 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:
  - Guessing gives 52% accuracy
  - Logistic Regression on status and similarity features: 67%
  - If we see the first  $k=5$  votes 85% (gold standard)

Theme: Learning from implicit feedback

Audience composition tells us something about their reaction

# Summary so far

- 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

- **Status seems to be salient feature**
- **Similarity** also plays important role
- Audience composition helps predict audience's reaction

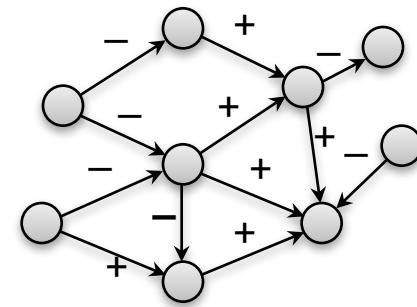
**Evaluations happen in the  
context of a network!**

# Two ways to look at this

- There are two ways to look at this:  
**One person evaluates the other via a positive/negative evaluation**



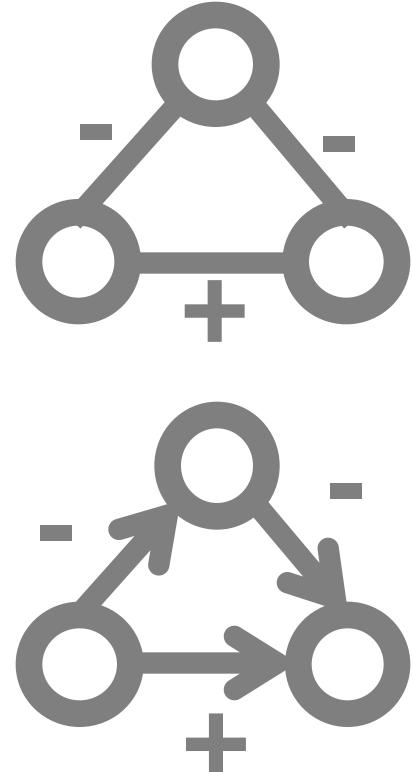
So far we focused on a single evaluation (without the context of a network)



Now we will focus on evaluations in the context of a network

# Signed Networks

- Networks with positive and negative relationships
- Our basic unit of investigation will be **signed triangles**
- First we talk about **undirected** networks then **directed**
- **Plan:**
  - **Model:** Consider two soc. theories of signed nets
  - **Data:** Reason about them in large online networks
  - **Application:** Predict if A and B are linked with + or -

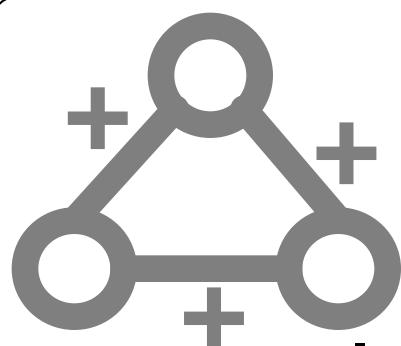


# Signed Networks

- Networks with **positive** and **negative** relationships
- Consider an undirected complete graph
- Label each edge as either:
  - **Positive**: friendship, trust, positive sentiment, ...
  - **Negative**: enemy, distrust, negative sentiment, ...
- Examine triples of connected nodes A, B, C

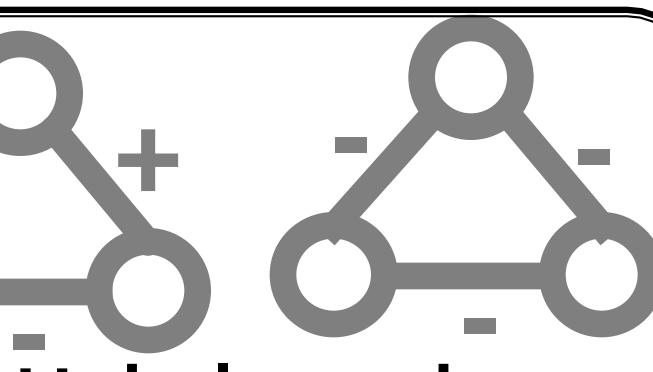
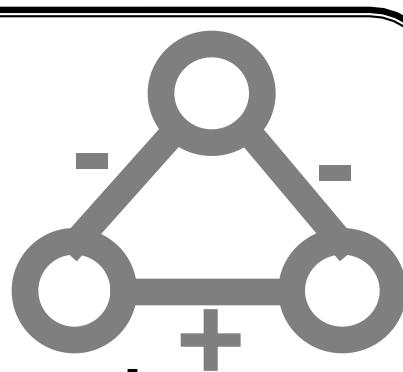
# Theory of Structural Balance

- 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:



Balanced

Consistent with “friend of a friend” or  
“enemy of the enemy” intuition

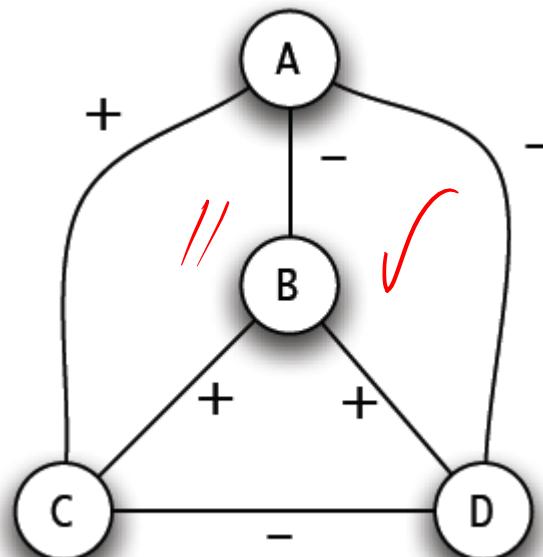


Unbalanced

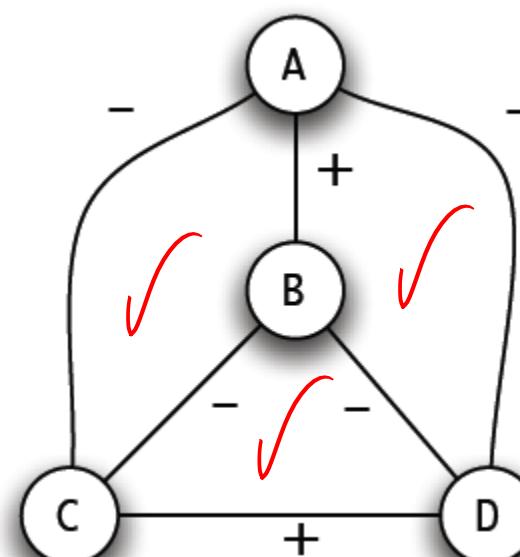
Inconsistent with the “friend of a friend”  
or “enemy of the enemy” intuition

# Balanced/Unbalanced Networks

- Graph is balanced if every connected triple of nodes has:
  - All 3 edges labeled +, or
  - Exactly 1 edge labeled +



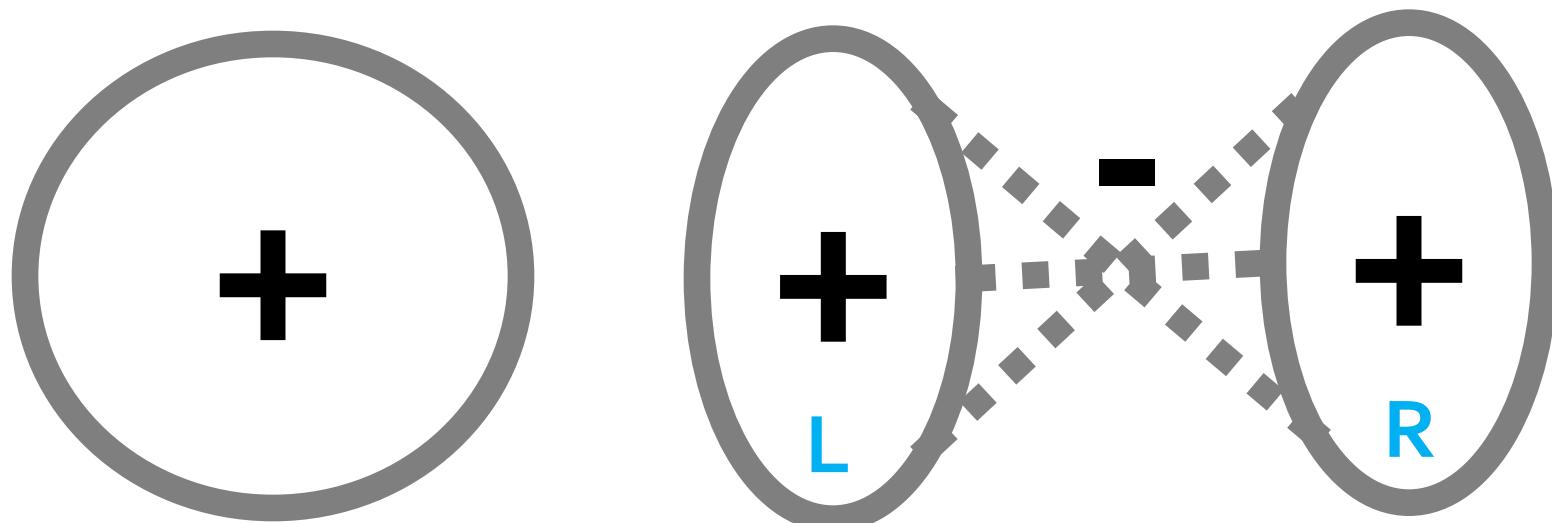
Unbalanced



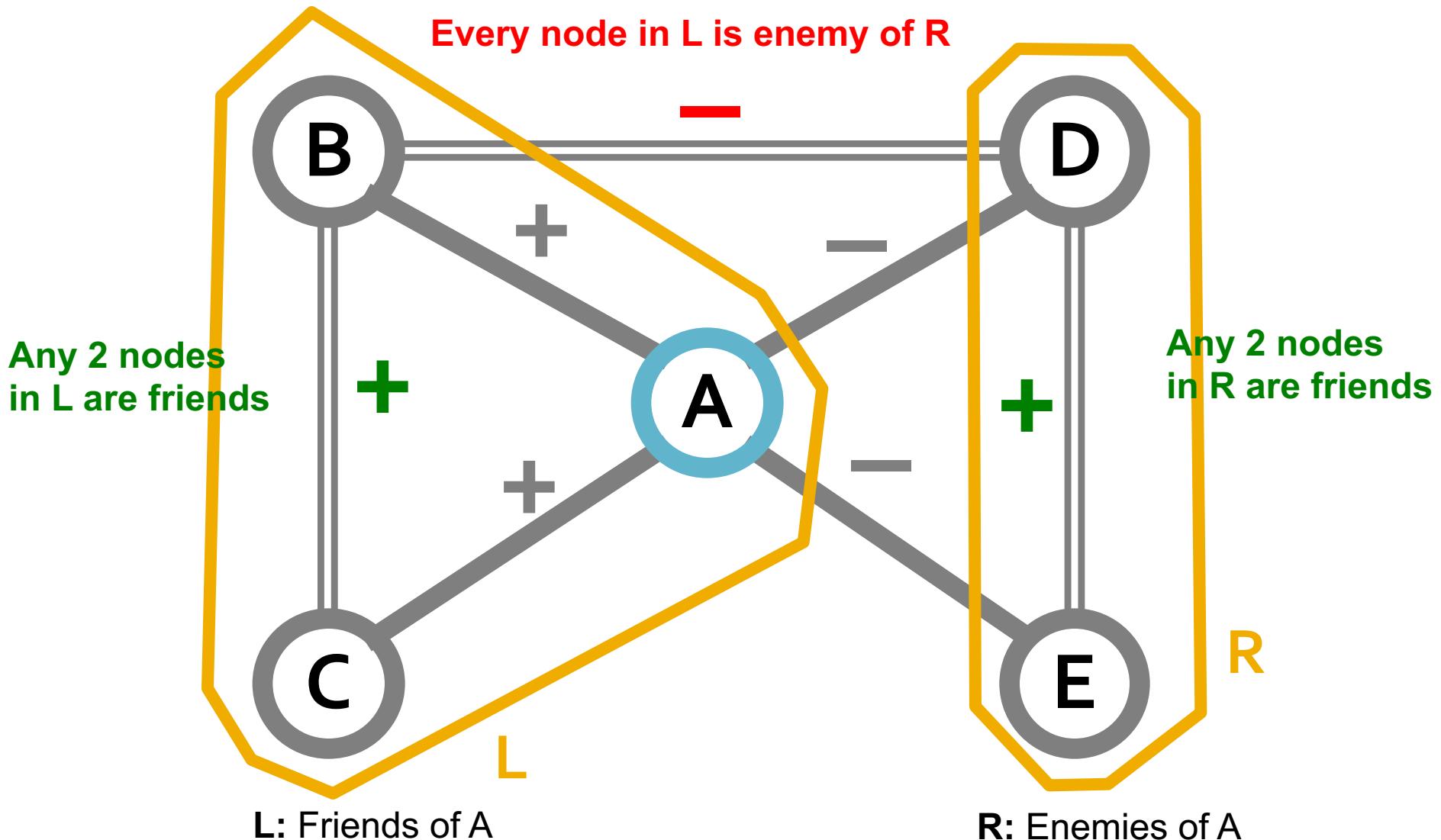
Balanced

# Local Balance → Global Fractions

- **Balance implies global coalitions** [Cartwright-Harary]
- Fact: If all triangles are balanced, then either:
  - The network contains only positive edges, or
  - Nodes can be split into 2 sets where negative edges only point between the sets



# Analysis of Balance: Coalitions



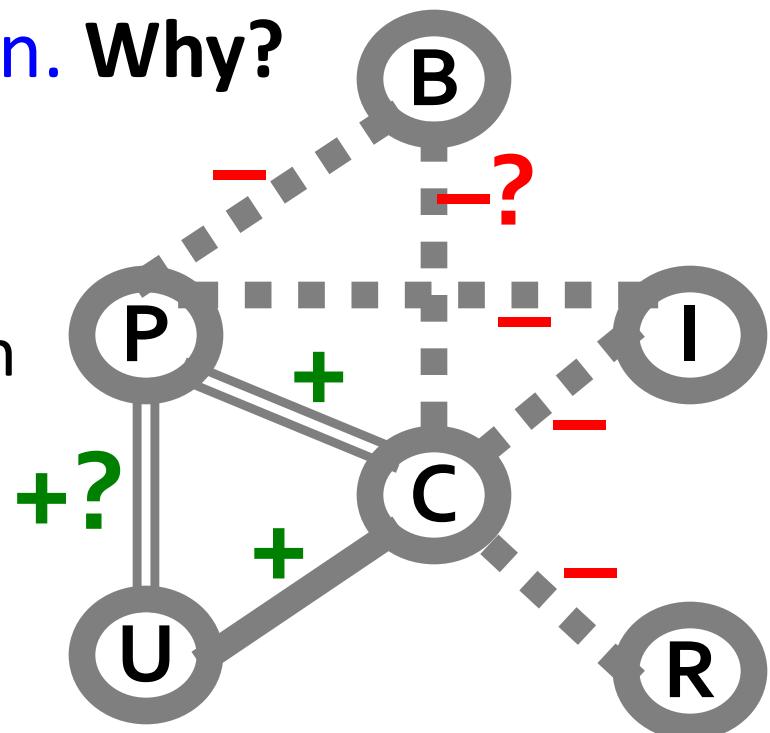
# Example: International Relations

- **International relations:**

- **Positive** edge: alliance
- **Negative** edge: animosity

- Separation of Bangladesh from Pakistan in 1971: US supports Pakistan. Why?

- USSR was enemy of China
- China was enemy of India
- India was enemy of Pakistan
- US was friendly with China
- China vetoed  
Bangladesh from U.N.



# Balance in General Networks

- So far we talked about complete graphs

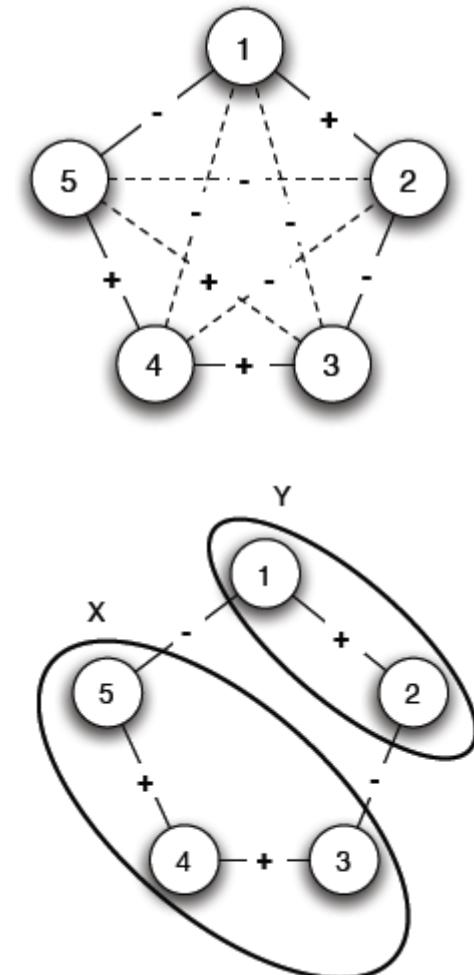
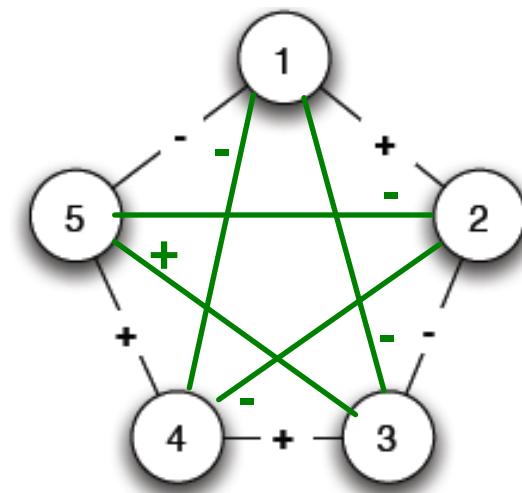
## Def 1: Local view

Fill in the missing edges to achieve balance

## Def 2: Global view

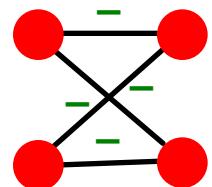
Divide the graph into two coalitions

The 2 definitions are **equivalent!**

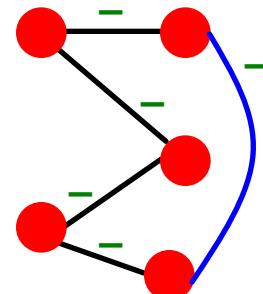


# Is a Signed Network Balanced?

- 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

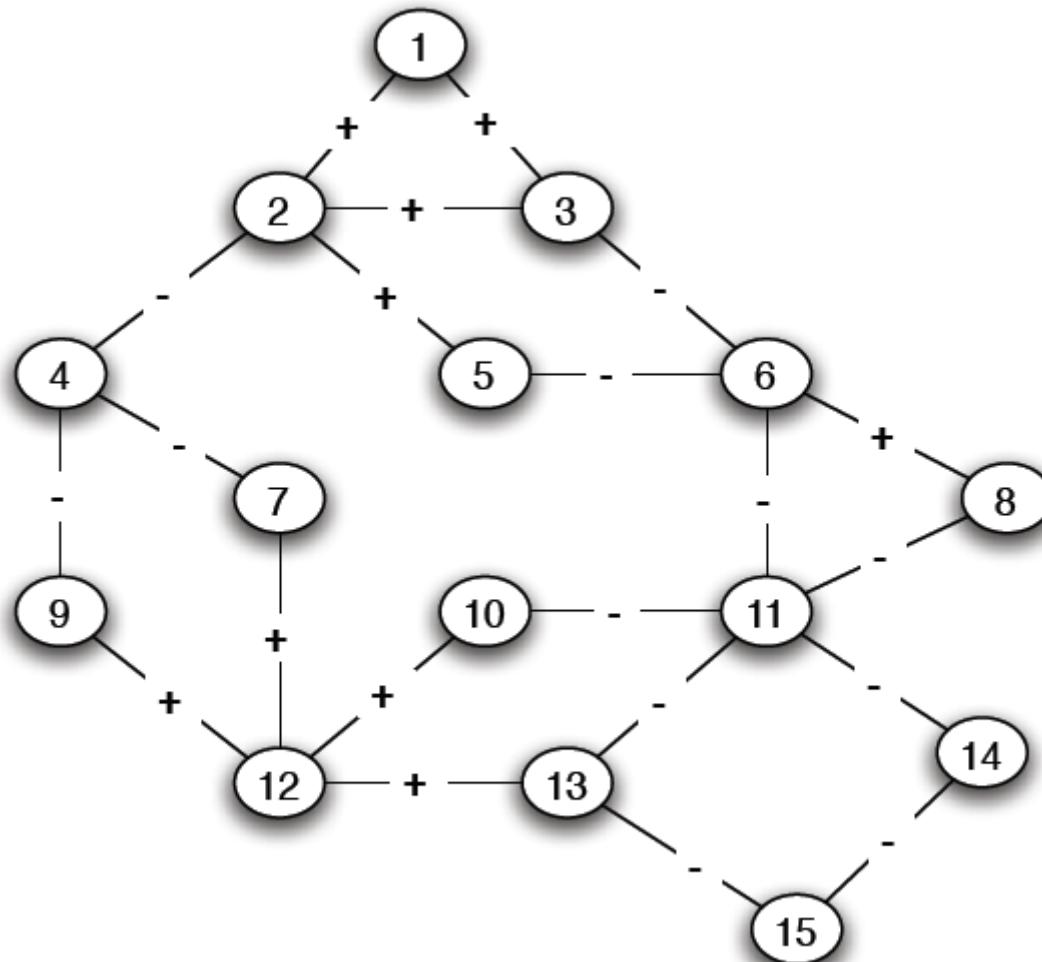


Even length cycle

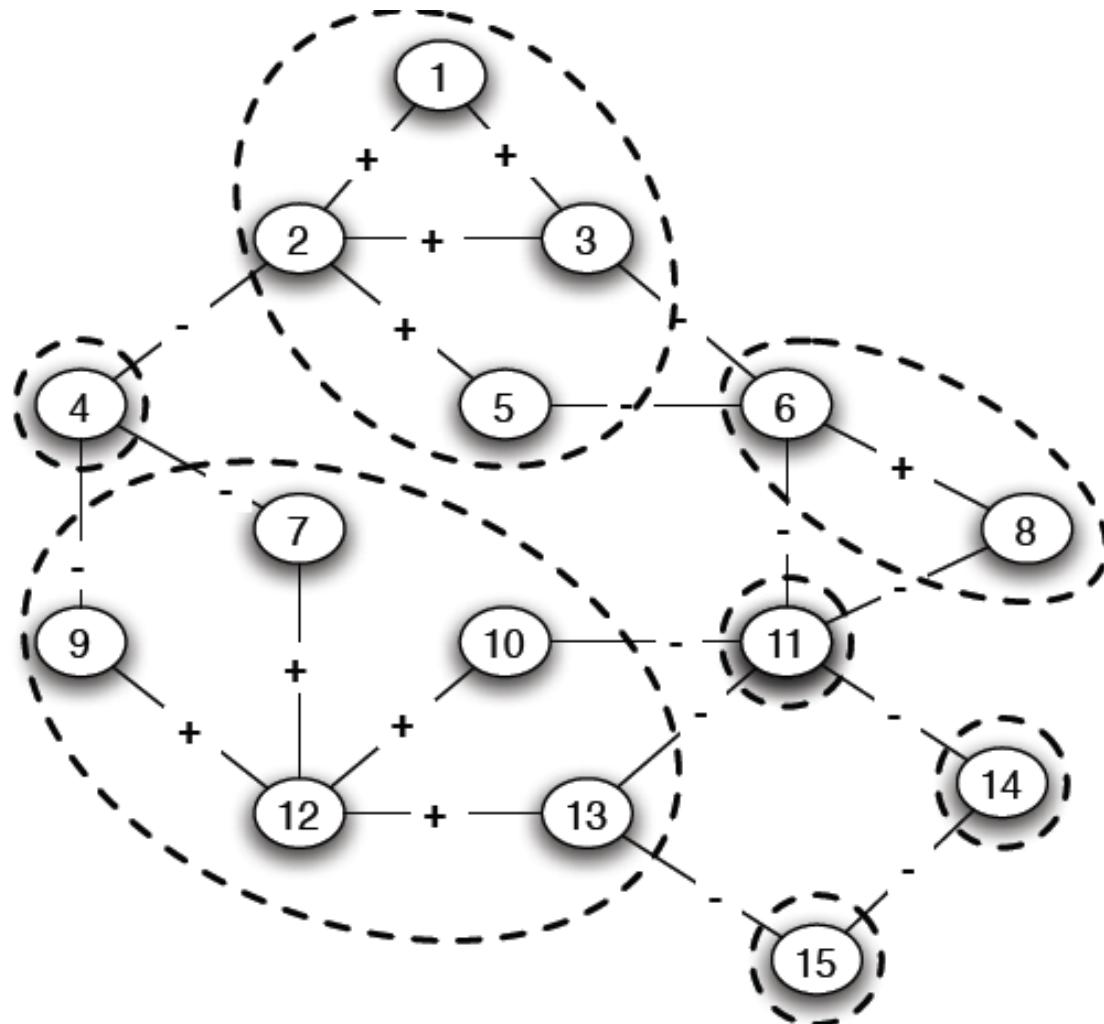


Odd length cycle

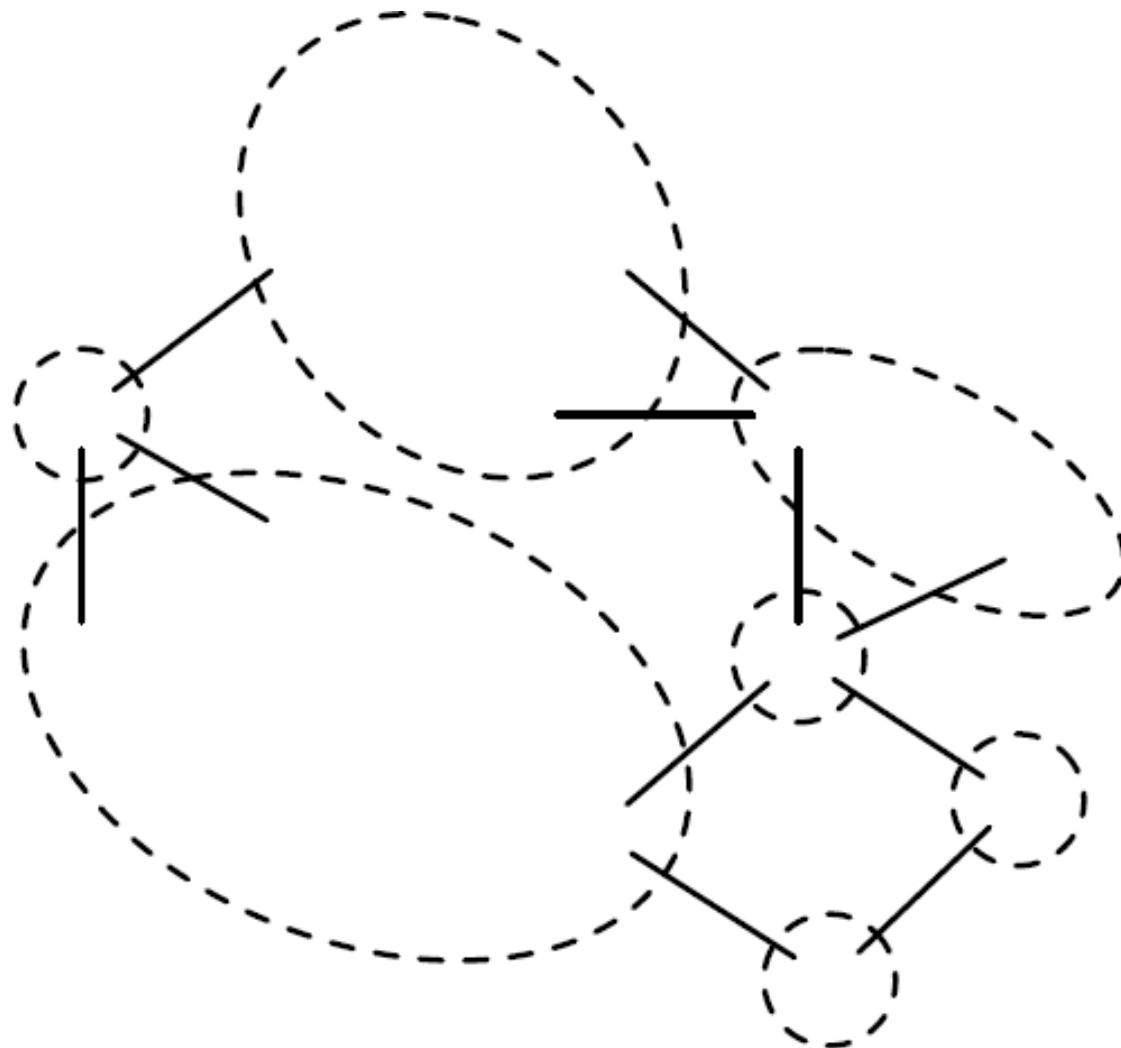
# Signed Graph: Is it Balanced?



# 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**

