

Signed networks: Structural Balance and Status

CS 322: (Social and Information) Network Analysis
Jure Leskovec
Stanford University

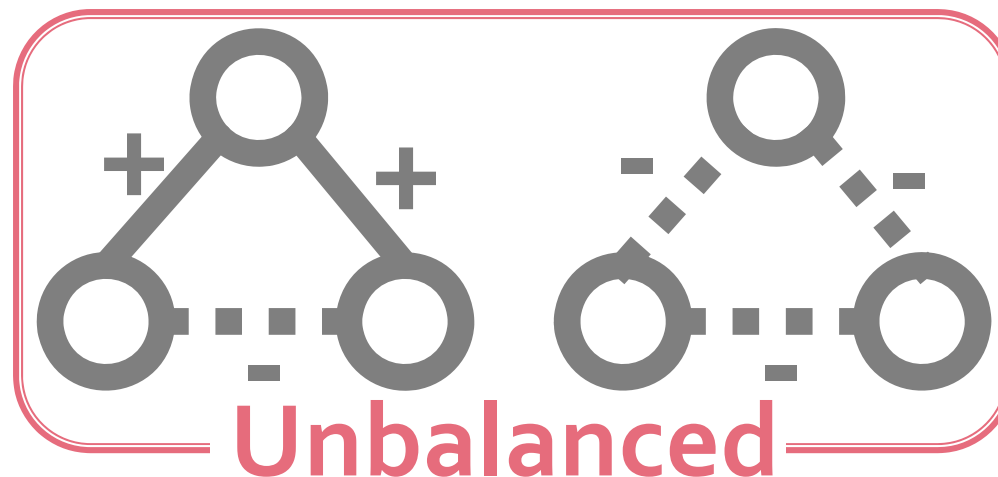


Signed networks

- 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

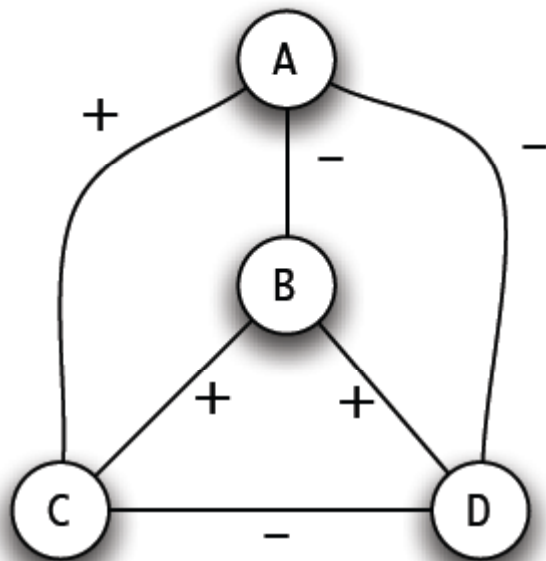
Theory of Structural Balance

- Three-Node Signed Triads [Heider '46]

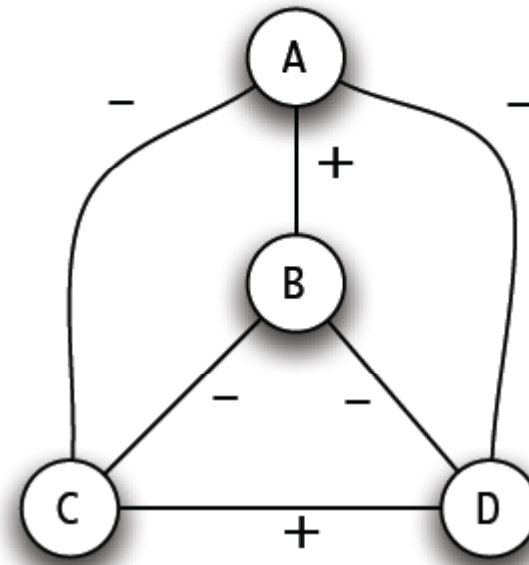


Balanced/unbalanced networks

- Graph is **balanced** if every connected triple of nodes has all 3 edges labeled +, or else exactly 1 edge is labeled +.



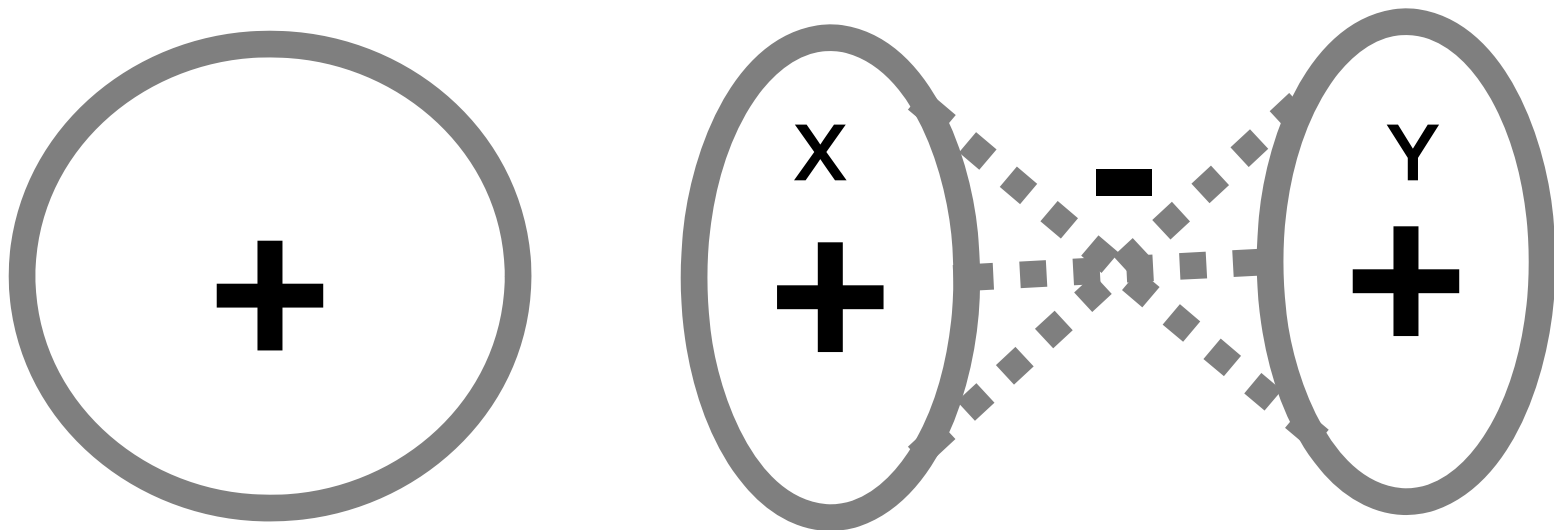
Unbalanced



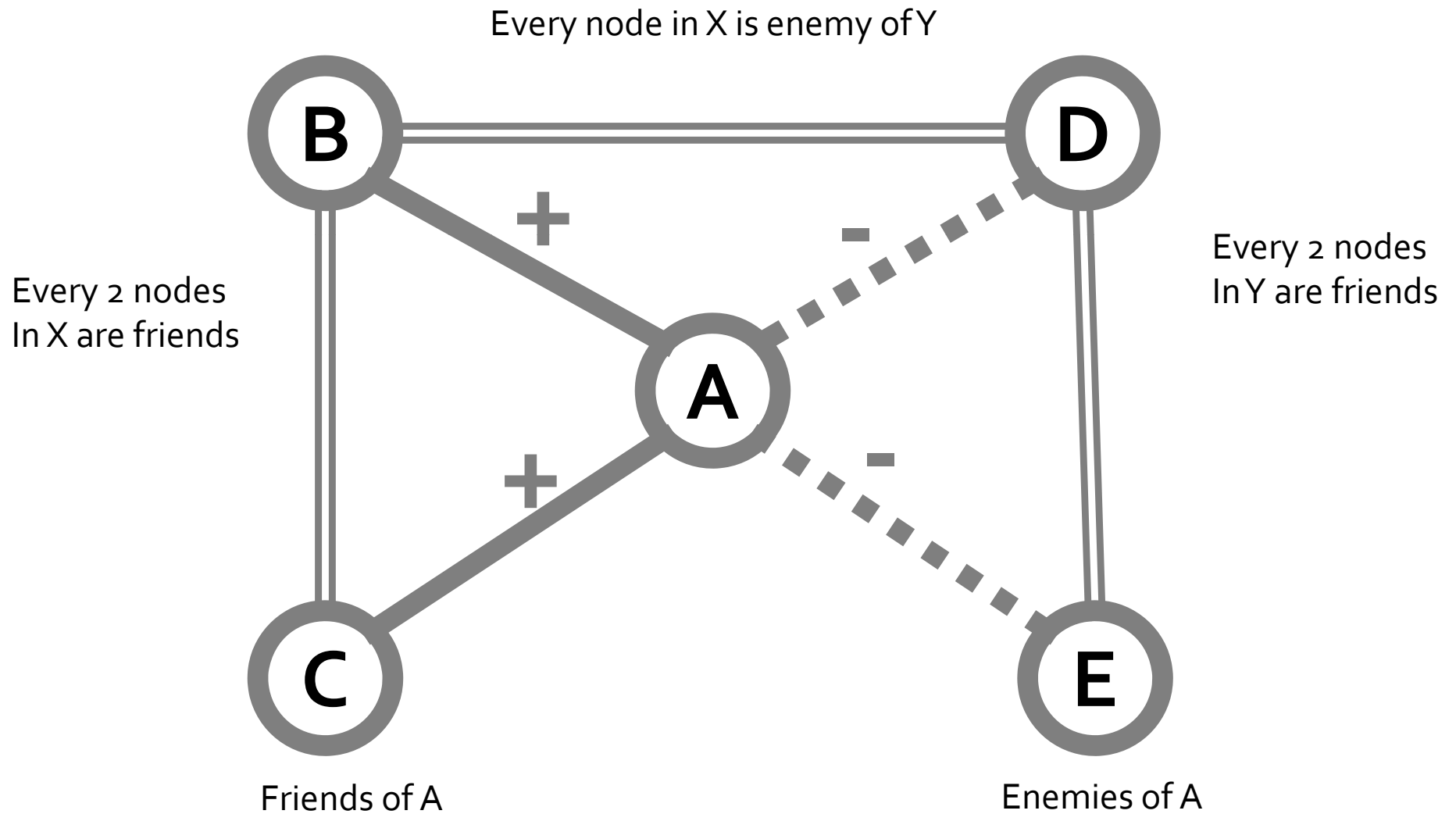
Balanced

Local balance \rightarrow Global factions

- Balance implies global coalitions [Cartwright-Harary]
 - If all triangles are balanced, then either:
 - The network contains only positive edges, or
 - Nodes can be split into 2 factions linked by negative edges

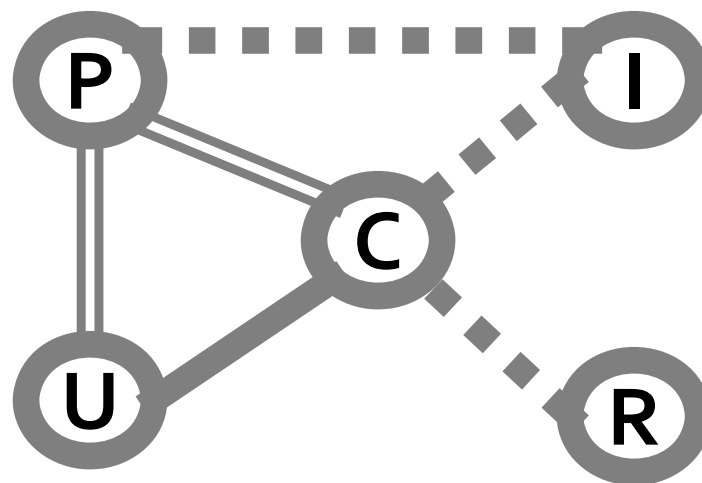


Analysis of balance

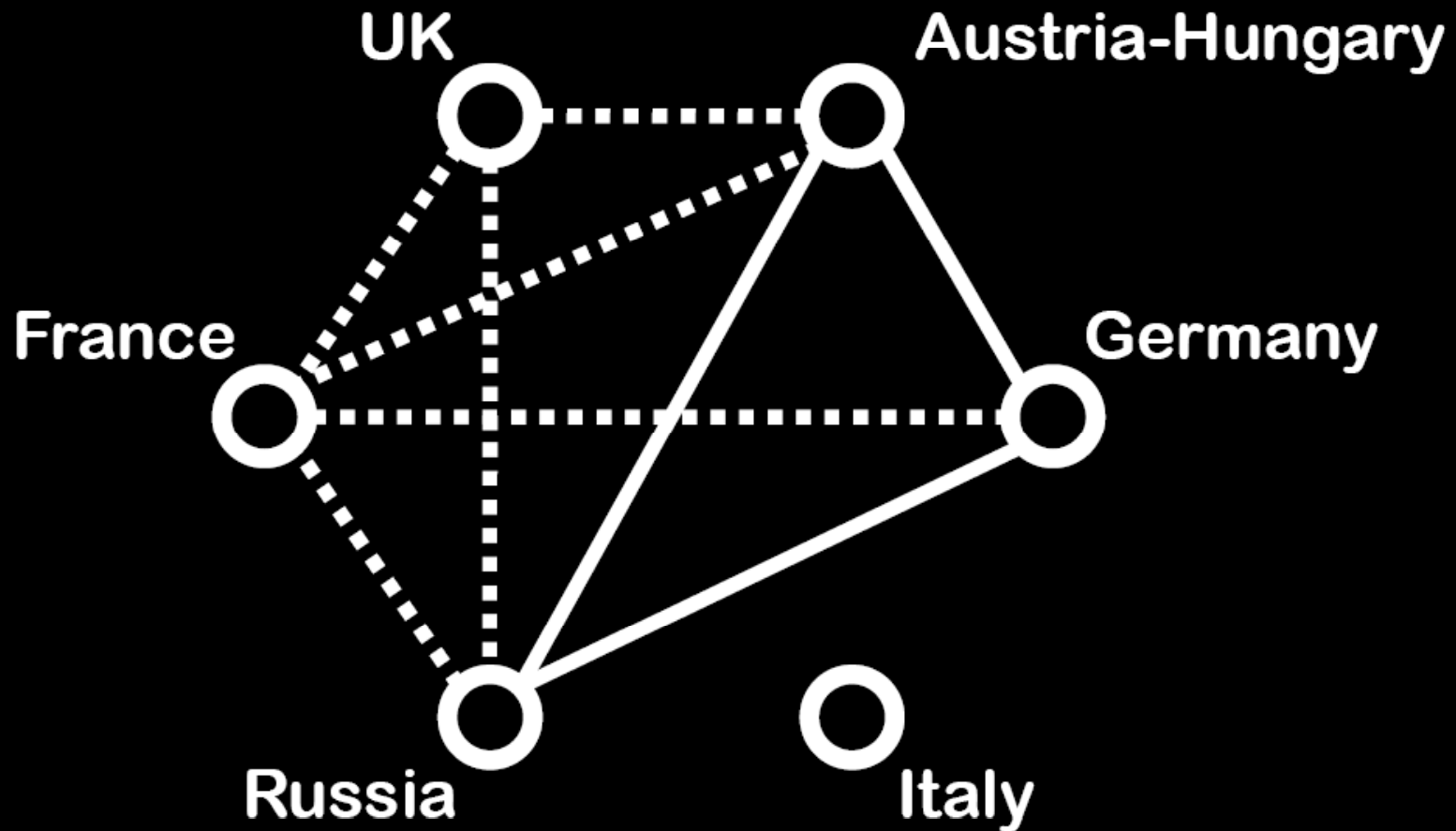


Example

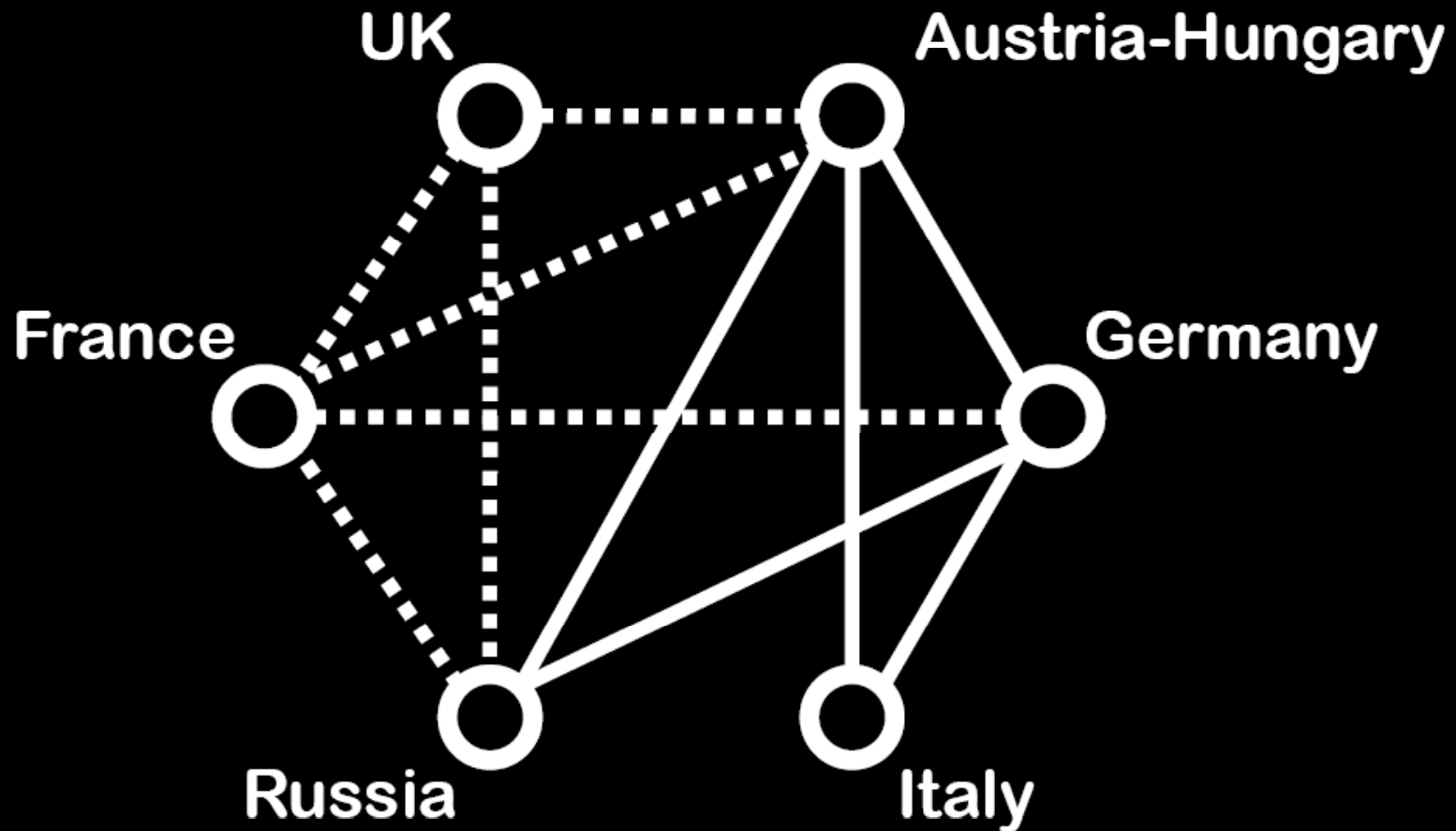
- International relations:
 - **Positive** edge: alliance
 - **Negative** edge: animosity
- Separation of Bangladesh from Pakistan in 1972: **US supports Pakistan**. Why?
 - **USSR** was enemy of **China**
 - **China** was enemy of **India**
 - **India** was enemy of **Pakistan**
 - **US** was friendly with **China**
 - **China vetoed Bangladesh**



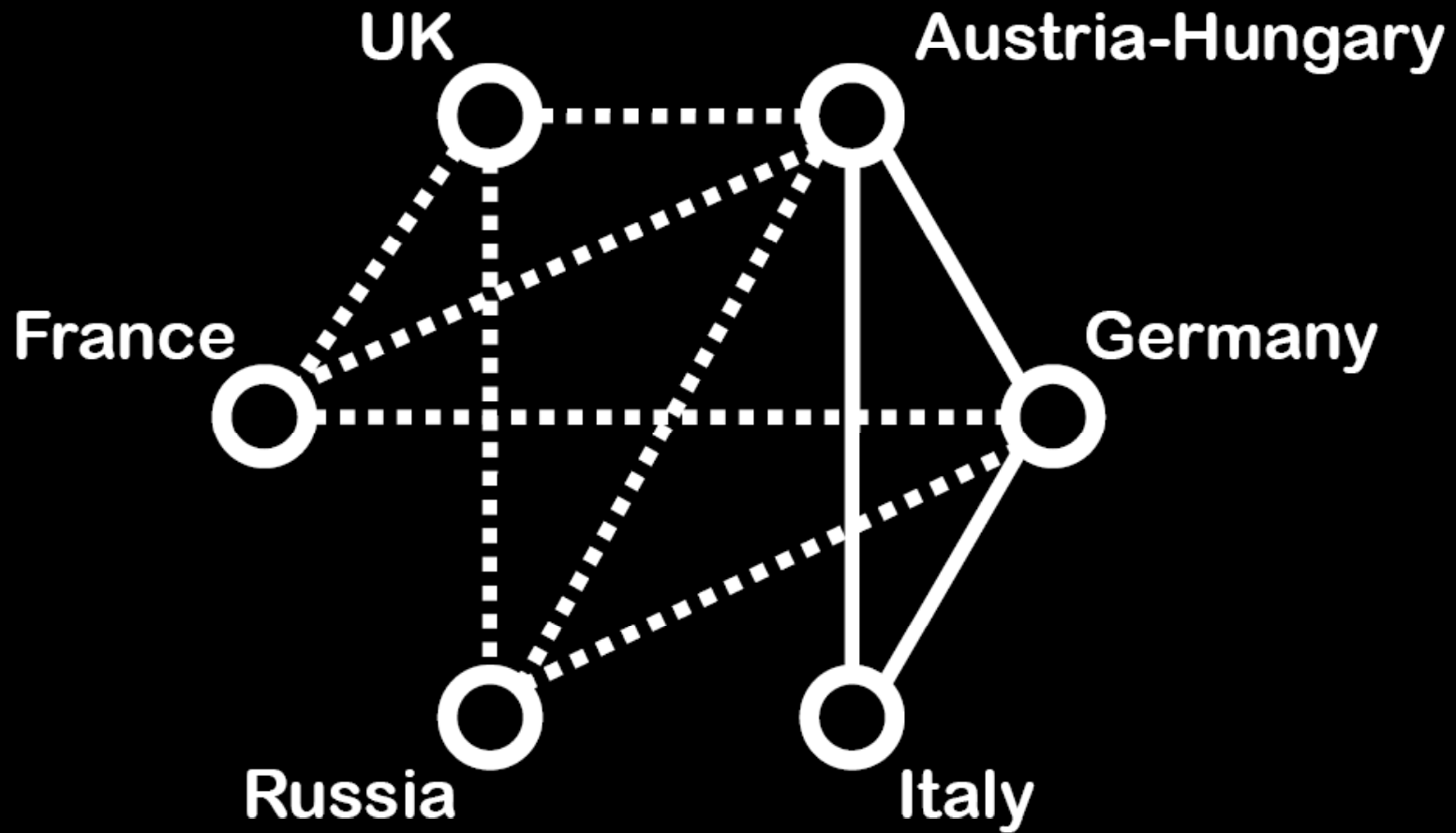
1872-1881



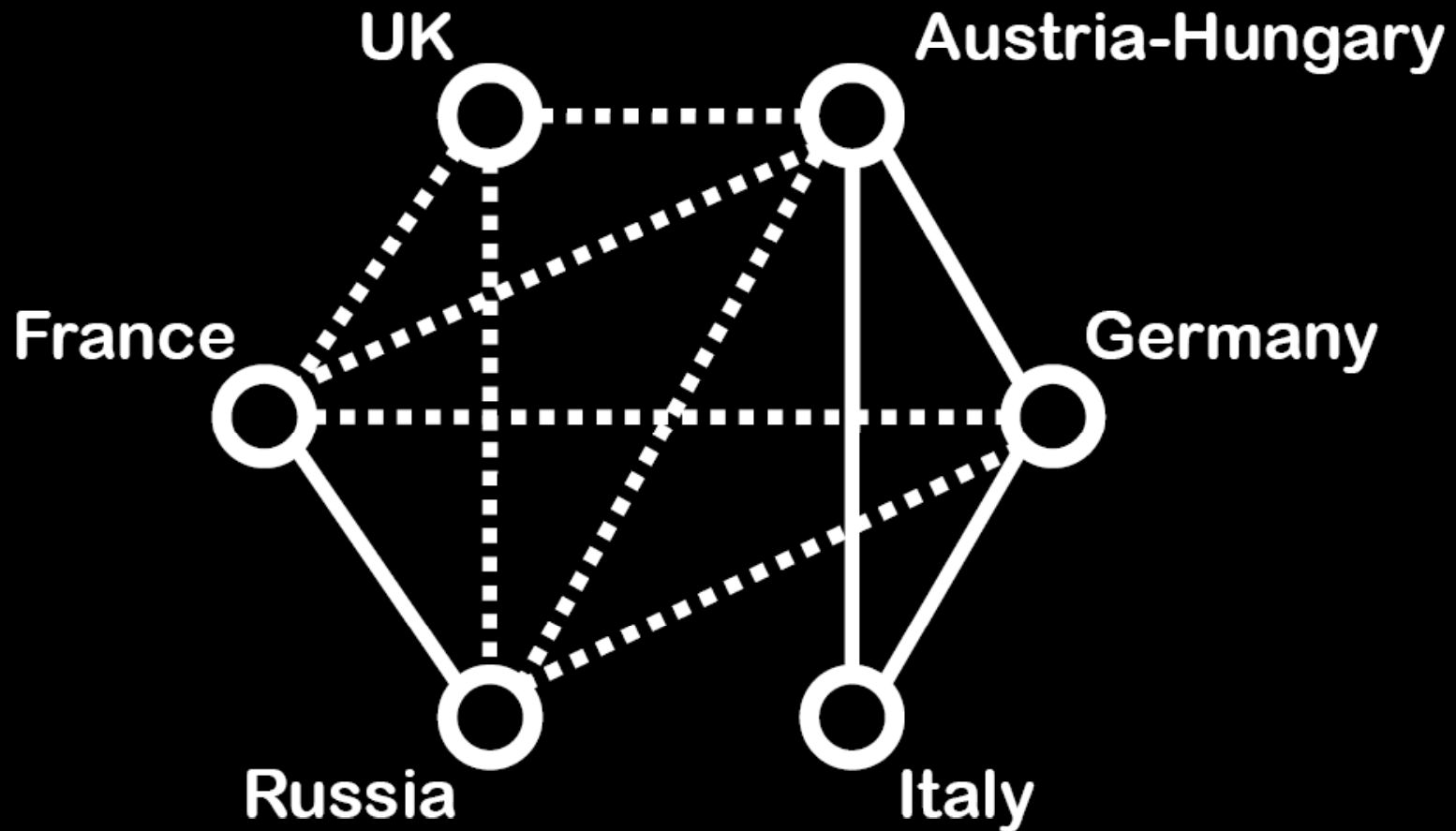
1882



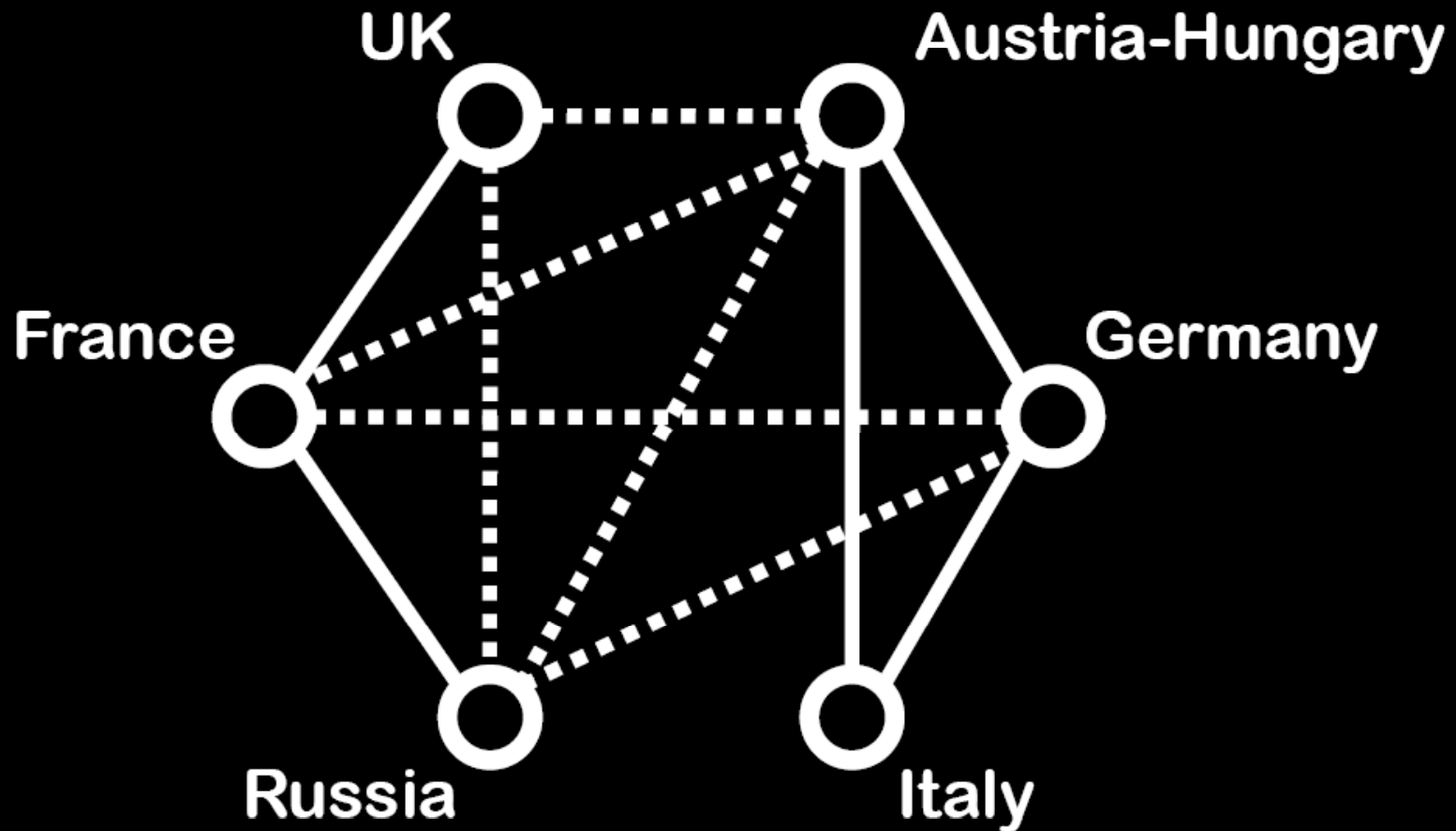
1890



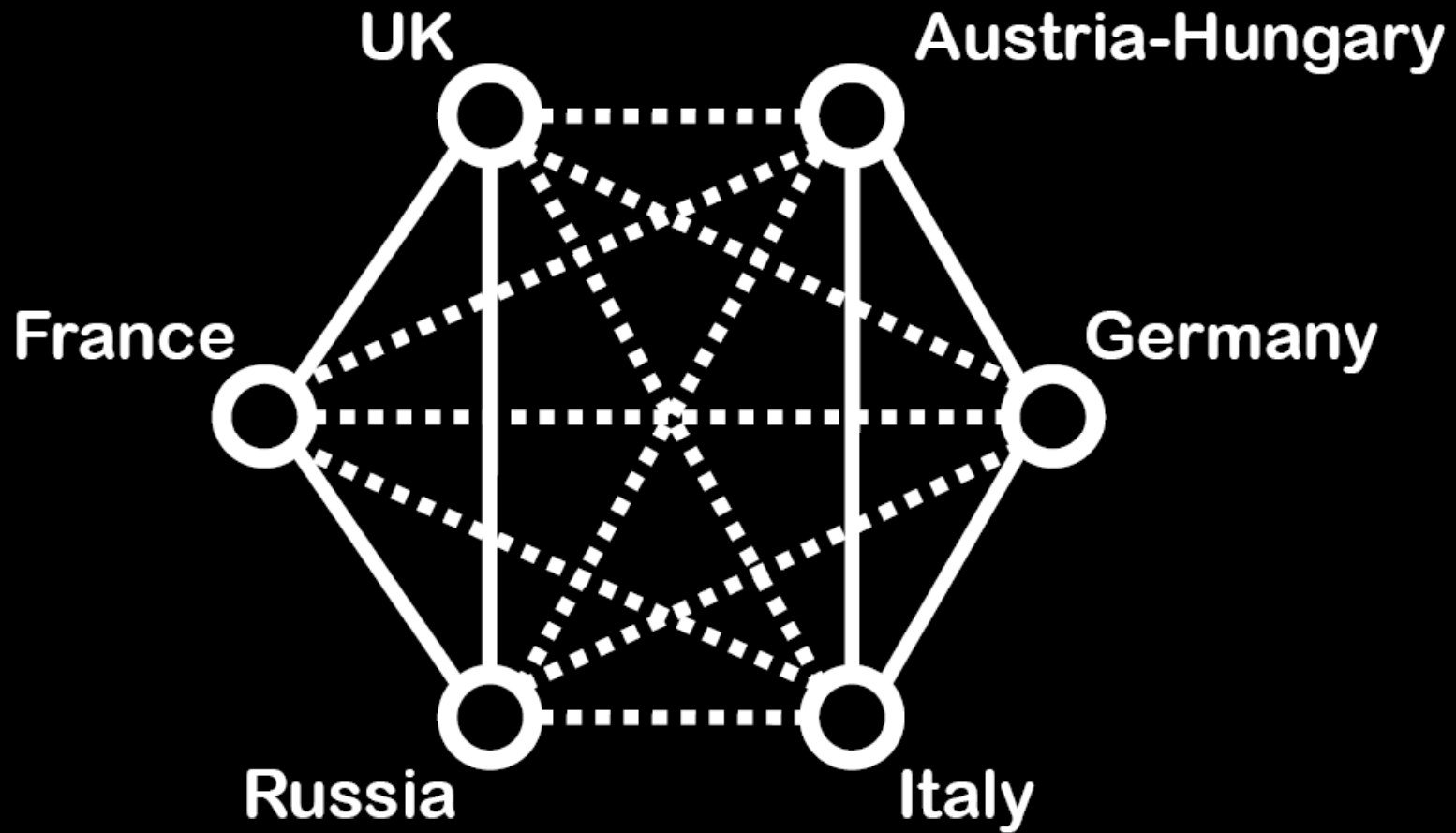
1891-1894



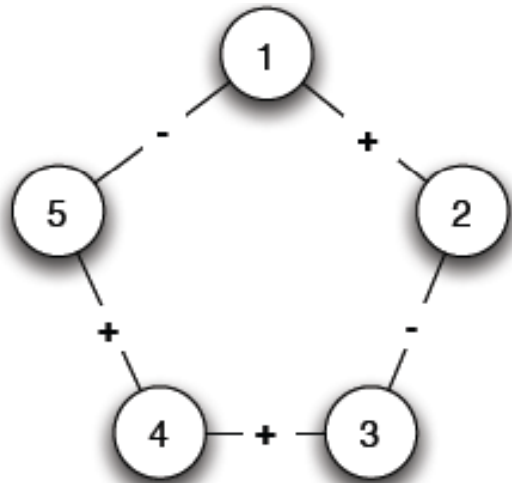
1904



1907

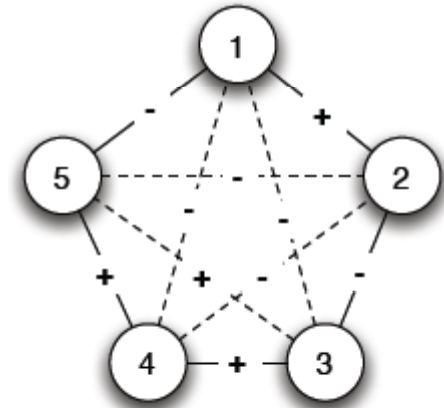


Balance in general networks

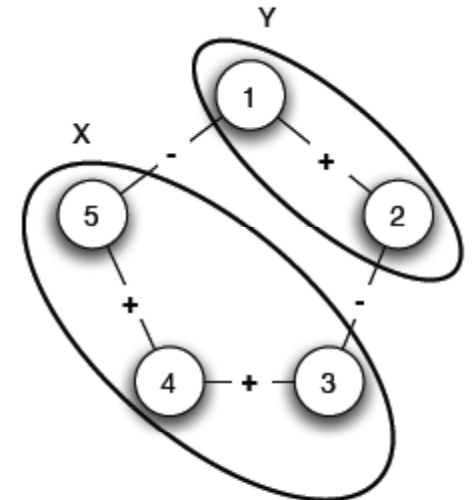


Balanced?

- Def 1: Local view
 - Fill in the missing edges to achieve balance



- Def 2: Global view
 - Divide the graph into two coalitions

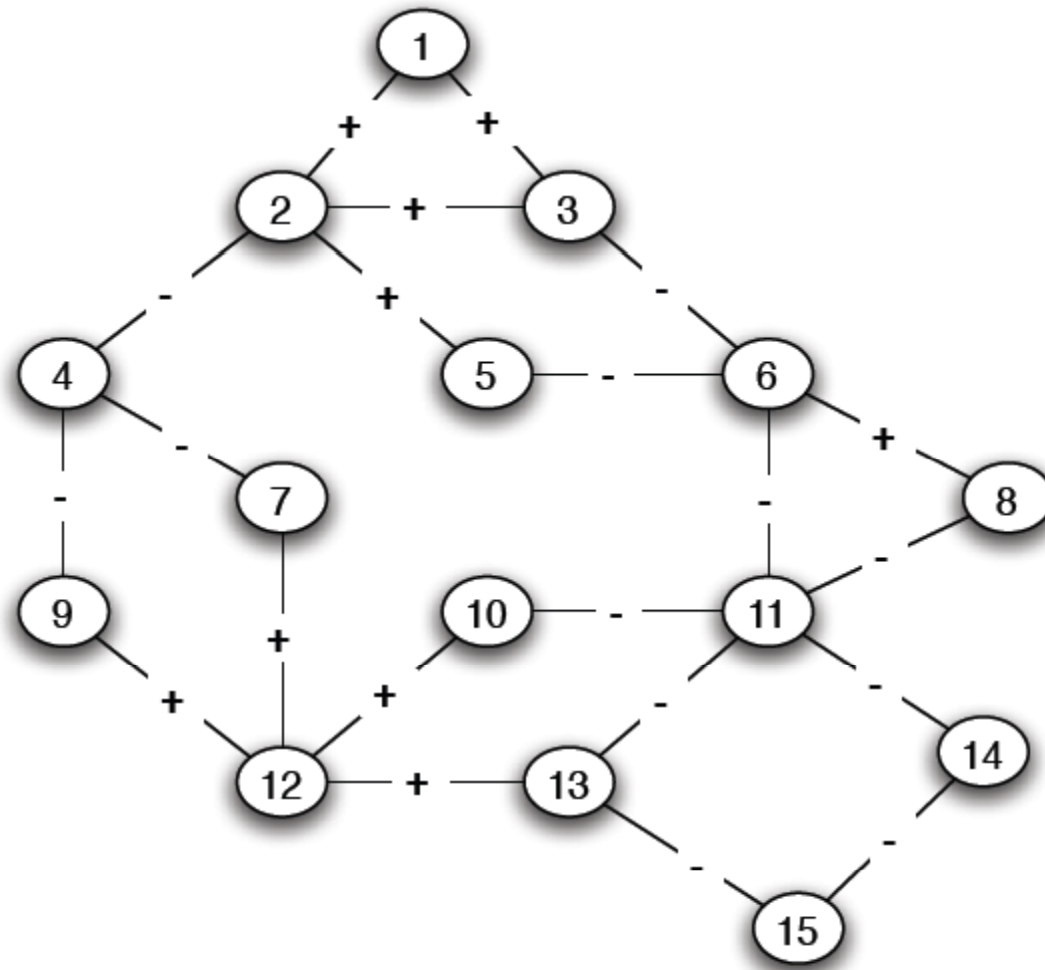


- Defs are equivalent!

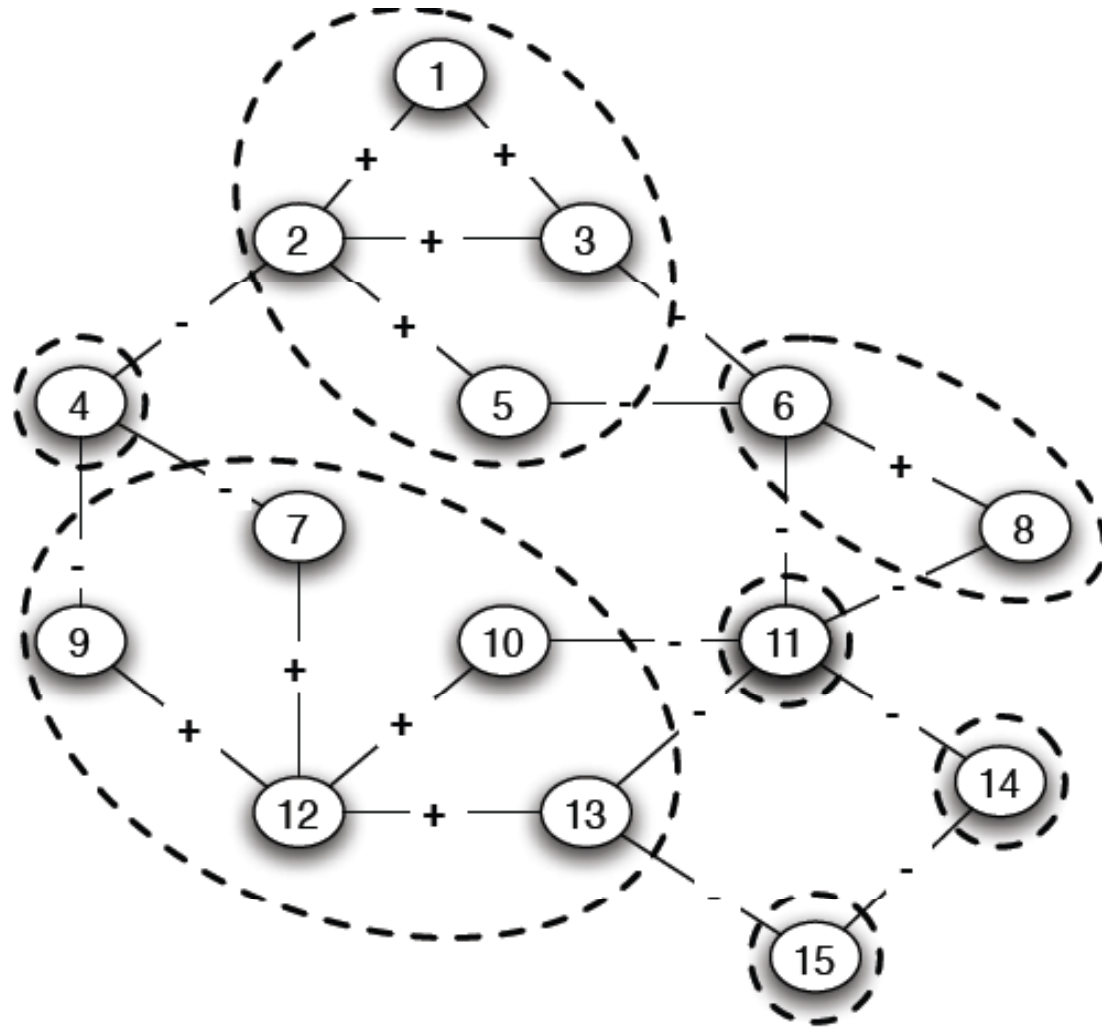
Is a signed network balanced?

- A 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
 - 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

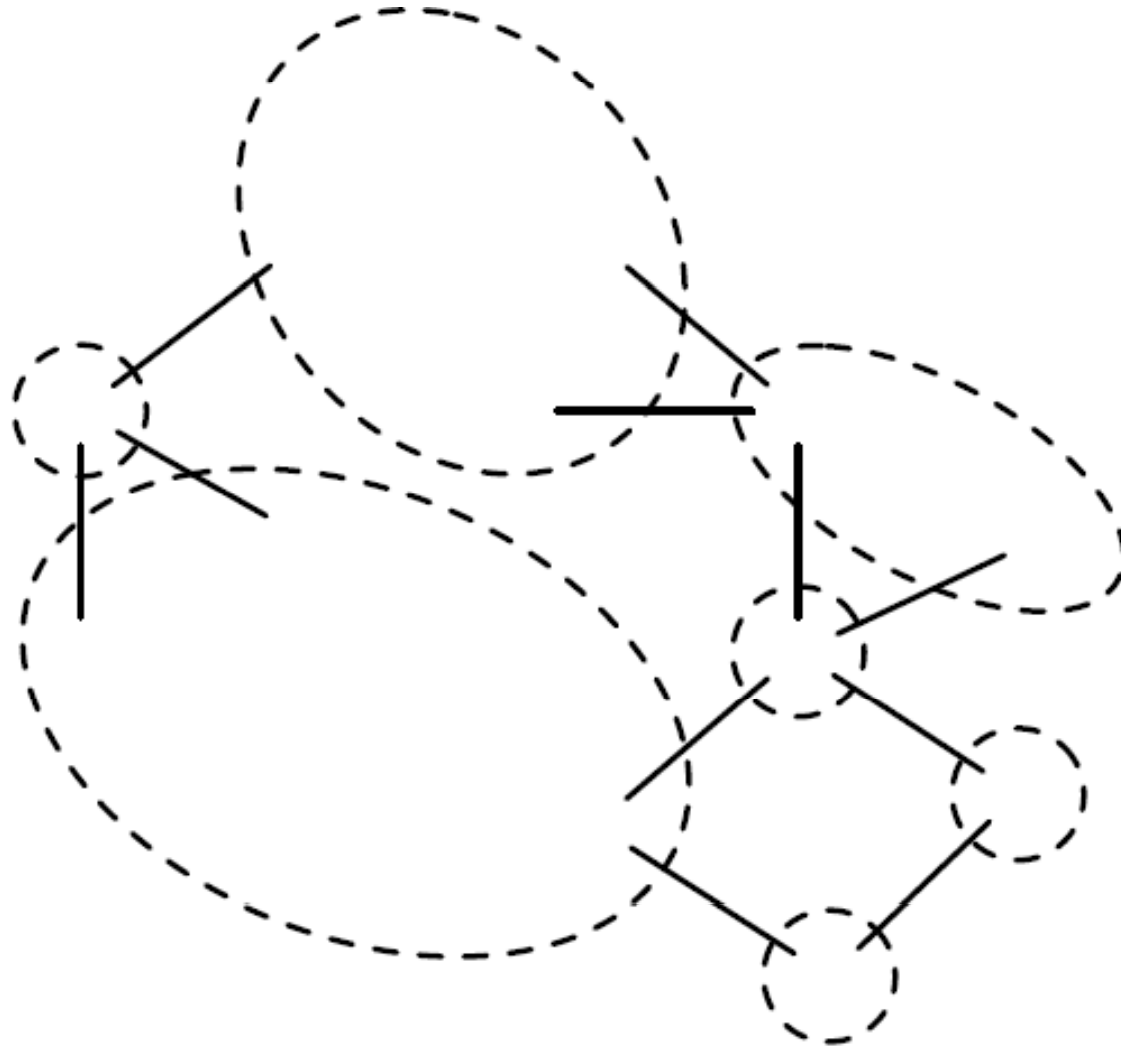
Signed Graph



Positive connected components

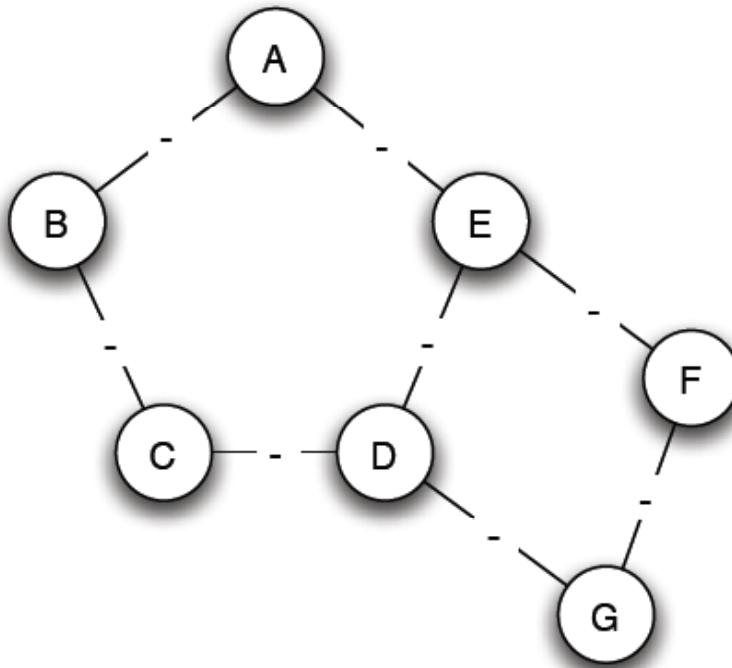


Reduced graph on super nodes



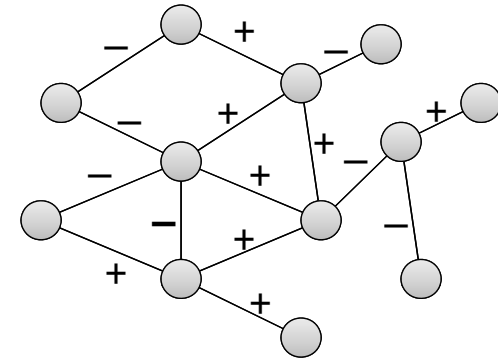
BFS on reduced graph

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



Real Large Signed Networks

- Each edge has a **sign** (+ or -)
- **Meaning of signs can be:**
 - Support/Oppose (Wikipedia)
 - Trust/Distrust (Epinions)
 - Friend/Foe (Slashdot)

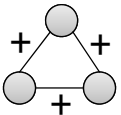
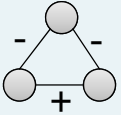
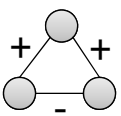
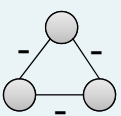


| | Epinions | Slashdot | Wikipedia |
|---------|----------|----------|-----------|
| Nodes | 119,217 | 82,144 | 7,118 |
| Edges | 841,200 | 549,202 | 103,747 |
| + edges | 85.0% | 77.4% | 78.7% |
| - edges | 15.0% | 22.6% | 21.2% |

- **Questions:**
 - How do edge signs and network structure interact?
 - **What theories explain signs of edges?**
 - Can we accurately predict signs of edges?

Balance in our network data

- Does structural balance hold?

| Triad | Epinions | | Wikipedia | | Balance |
|---|----------|----------|-----------|----------|---------|
| | P(T) | $P_o(T)$ | P(T) | $P_o(T)$ | |
|  | 0.87 | 0.62 | 0.70 | 0.49 | ✓ |
|  | 0.71 | 0.055 | 0.21 | 0.10 | ✓ |
|  | 0.05 | 0.32 | 0.08 | 0.49 | ✓ |
|  | 0.007 | 0.003 | 0.011 | 0.010 | ✗ |

P(T) ... probability of a triad
 $P_o(T)$... triad probability if the signs would be random

Global factions: Structure

| | Size | | Clustering | | Component | |
|--------------|---------|---------|------------|-------|-----------|-------|
| | Nodes | Edges | Real | Rnd | Real | Rnd |
| Epinions: - | 119,090 | 123,602 | 0.012 | 0.022 | 0.308 | 0.334 |
| Epinions: + | 119,090 | 717,027 | 0.093 | 0.077 | 0.815 | 0.870 |
| Slashdot: - | 82,144 | 124,130 | 0.005 | 0.010 | 0.423 | 0.524 |
| Slashdot: + | 82,144 | 425,072 | 0.025 | 0.022 | 0.906 | 0.909 |
| Wikipedia: - | 7,115 | 21,984 | 0.028 | 0.031 | 0.583 | 0.612 |
| Wikipedia: + | 7,115 | 81,705 | 0.130 | 0.103 | 0.870 | 0.918 |

- **Clustering:**

- +net: more clustering than baseline
- -net: less clustering than expected

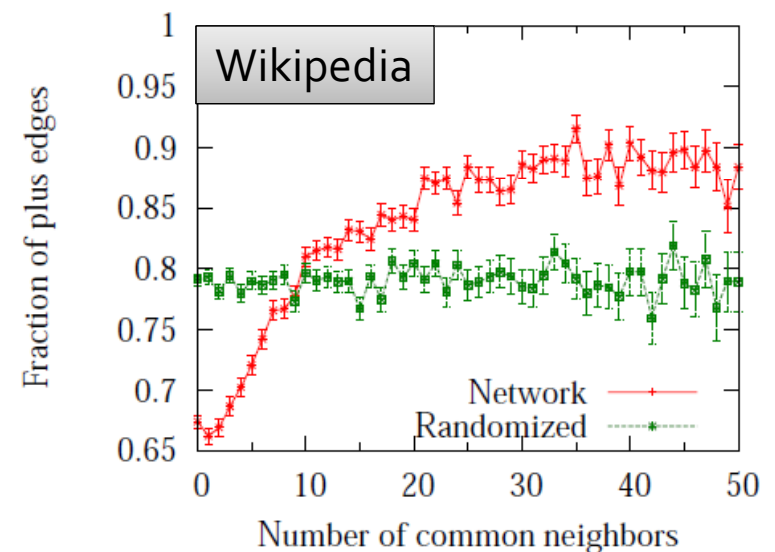
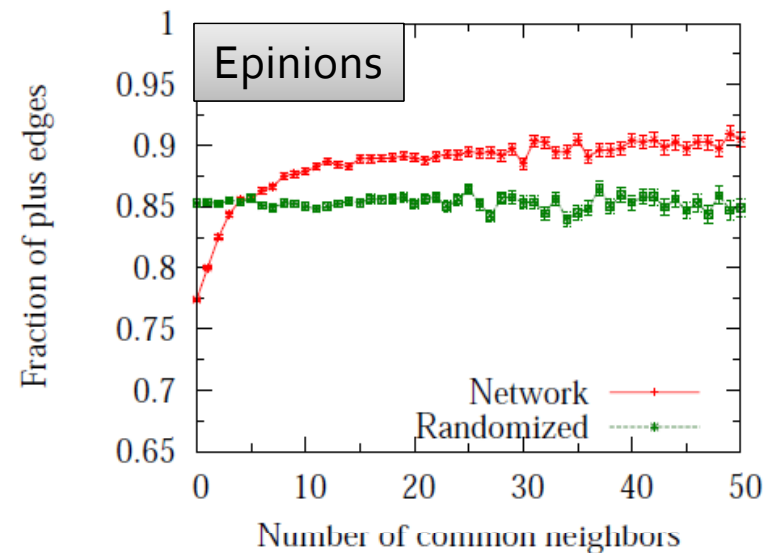
- **Size of connected component:**

- +/-net: smaller than expected

Suggests the network with factions

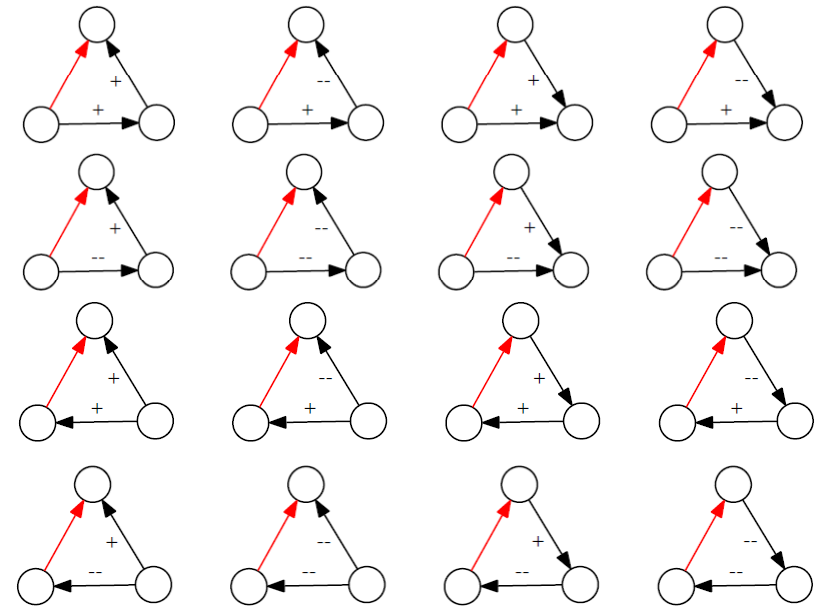
Global factions: Embeddedness

- Embeddedness of ties:
 - Positive ties tend to be **more** embedded
- Positive ties tend to be more **clumped together**
- Public display of signs (votes) in Wikipedia further attenuates this



Directed edges

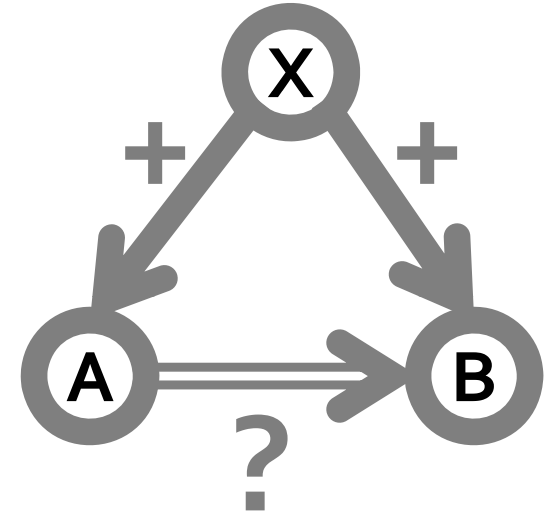
- But, our networks are really **directed**
 - trust, opinion (, friendship)
- How many \triangle are now explained by balance?
 - **Only half** (8 out of 16)
- Can we do better?
 - Yes. **Theory of Status.**



16 * 2 signed directed triads

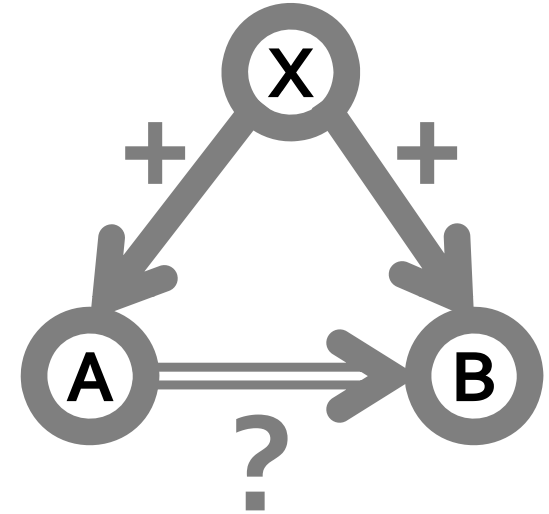
Theory of Status

- Edges are **directed**
- Edges are **created over time**
 - X has links to A and B
 - Now, A links to B (triad A-B-X)
 - How does sign of A-B depend signs of X?
- Different users make signs differently:
 - **Generative** baseline (prob. of A giving a +)
 - **Receptive** baseline (prob. of A receiving a +)



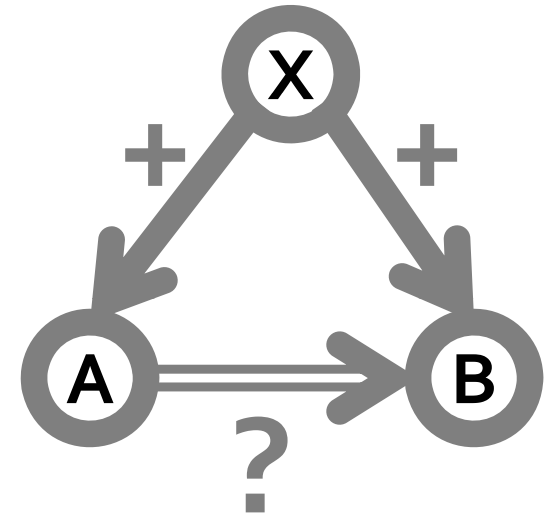
Joint positive endorsement

- X positively endorses A and B, then the link $A \rightarrow B$ is
 - More likely to be positive than generative baseline of A
 - Less likely to be positive than receptive baseline of B
- Why?



Example: Soccer team

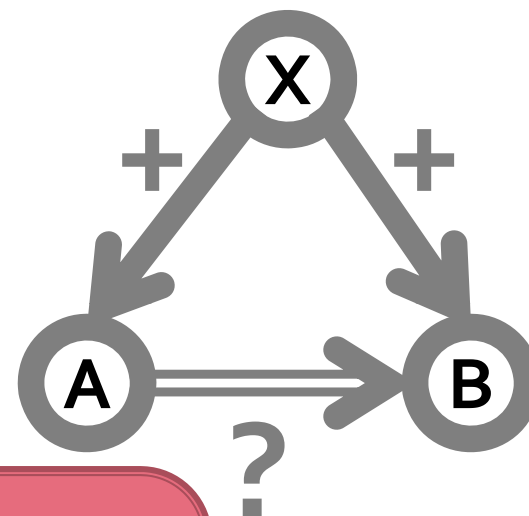
- Ask A: How does skill of B compare to yours?
 - Build a signed directed network
- Don't know what A thinks of B (don't know sign of $A \rightarrow B$)
- Can we infer the sign based on opinion of X?
- Conjecture 1 (generative hypothesis)
 - Since B has positive evaluation, B is high status
 - Thus, evaluation from A should be more likely to be positive than A evaluating a random player.



Example: Soccer team

- Conjecture 2 (receptive)

- Since A has positive evaluation, A is high status
- Thus, sign from A should be

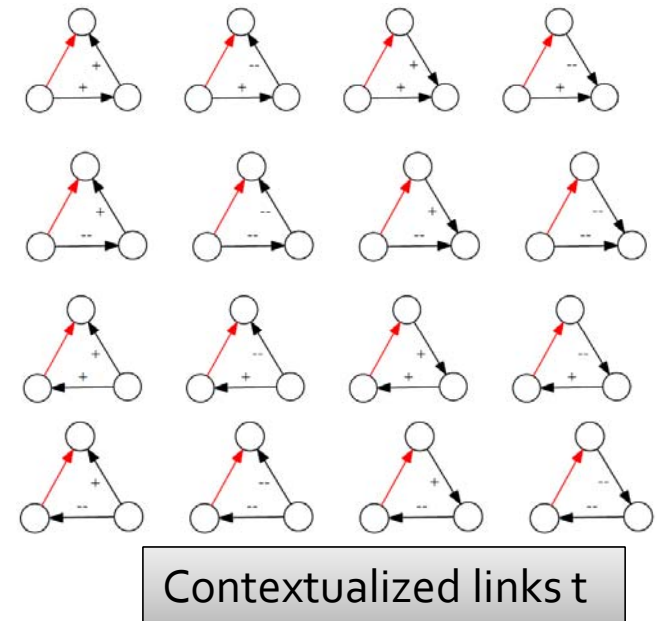


Sign of $A \rightarrow B$ deviates in different directions depending on the viewpoint!

- Thus, evaluation from A should be more likely to be positive than A evaluating a random player.

Contextualized links

- 16 possible contexts of a new link:
- Surprise:
 - How much behavior of A/B deviates from baseline based on the context t:
 - From viewpoint of A: Out-surprise $s_{out}(t)$
 - From viewpoint of B: In-surprise $s_{in}(t)$
 - More precisely:
 - $s_{out}(t) = \sum_A (t_A - d_A p_A) / \sqrt{\sum_A d_A p_A (1 - p_A)}$

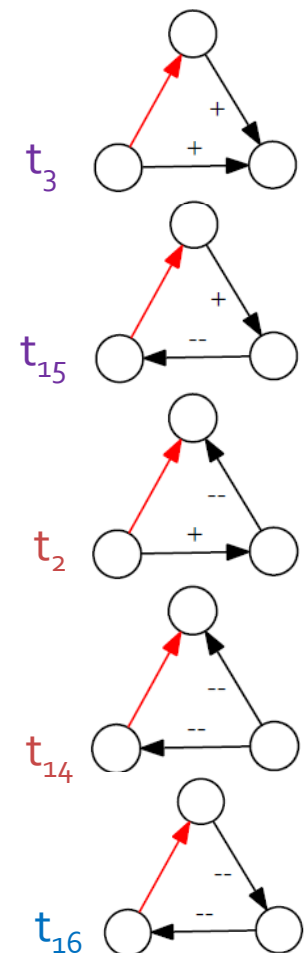


d_A ... out-degree of A
 p_A ... prob of A giving +
 t_A ... #of + of A in context t

Status vs. Balance (Epinions)

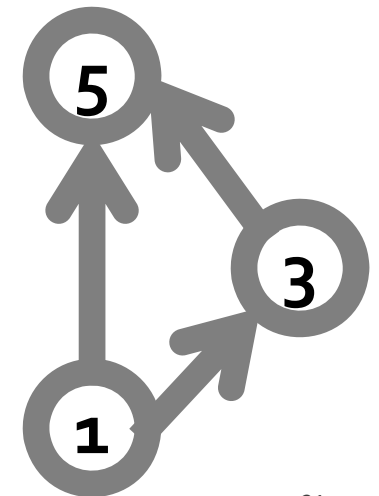
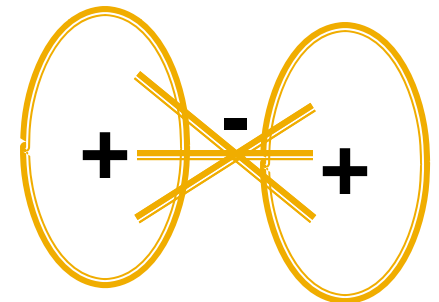
Predictions on 16 contextualized links:

| t_i | count | $P(+)$ | s_{out} | s_{in} | B_{out} | B_{in} | S_{out} | S_{in} |
|-------------------------------|---------|--------|-----------|----------|-----------|----------|-----------|----------|
| 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 |



From Local to Global Structure

- Both theories make predictions about the global structure of the network
- **Structural balance – Factions**
 - Put nodes into groups such that the number of in group “+” and between group “-” edges is maximized
- **Status theory – Global Status**
 - Flip direction and sign of negative edges
 - Assign each node a unique status value so that most edges point from low to high



From Local to Global Structure

- Fraction of edges that satisfy Balance and Status?

| Balance | Epinions | Slashdot | Wikipedia |
|----------------|----------|----------|-----------|
| Network | 0.8344 | 0.8105 | 0.7809 |
| Permuted | 0.8562 | 0.7779 | 0.7866 |
| Rewired | 0.8993 | 0.8310 | 0.8316 |
| Status | Epinions | Slashdot | Wikipedia |
| Network | 0.7905 | 0.8221 | 0.8538 |
| Permuted | 0.7241 | 0.7568 | 0.7767 |
| Rewired | 0.6377 | 0.6644 | 0.6321 |

Permuted: keep network structure, shuffle signs
Rewired: rewire edges, keep signs

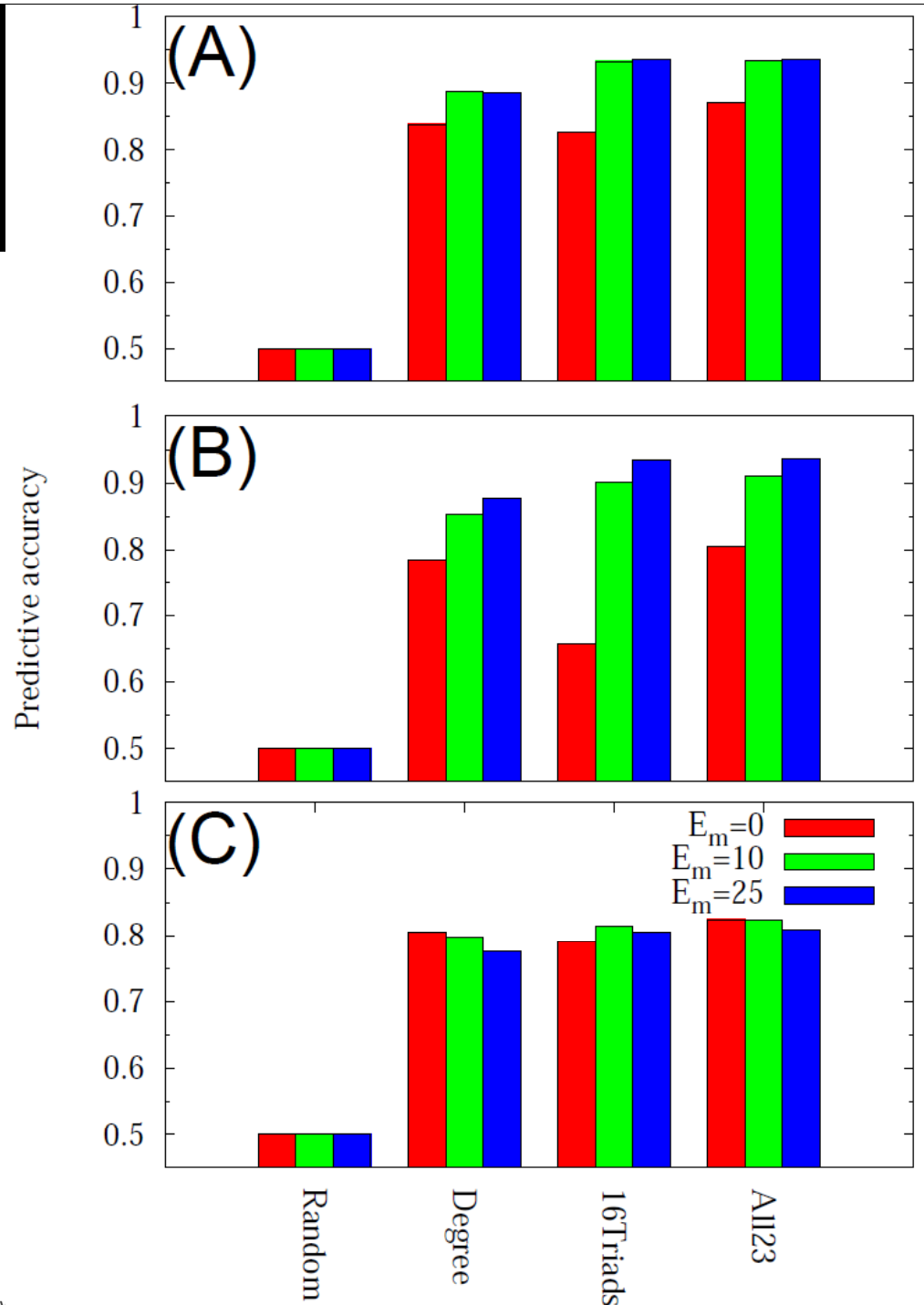
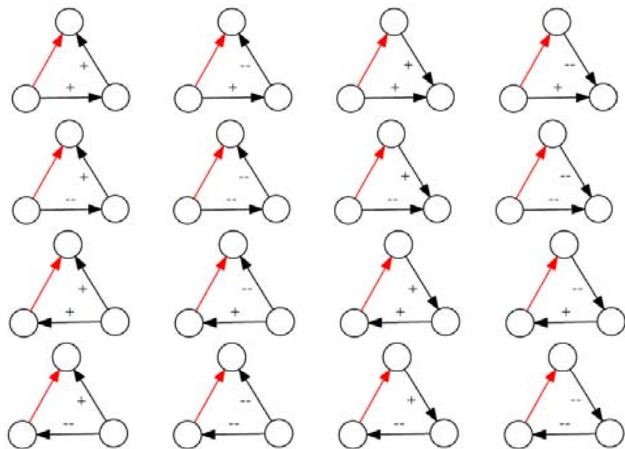
Predicting edge signs

- Problem formulation
 - Predict the sign of edge $A \rightarrow B$
 - Given signs of all other edges in the network
- Given an edge $A \rightarrow B$ create a set of features describing the context
 - Positive/negative in/out degree of A/B
 - Counts of various signed triads A \rightarrow B takes part in
- Use logistic regression to predict the edge sign:
$$P(+|x) = \frac{1}{1 + e^{-(b_0 + \sum_i^n b_i x_i)}}$$

Example:

Features for learning

- Degree:
 - 6 degree features
 - embeddedness
- 16Triads:



Generalization

- How general are the results across datasets?
 - Train on row “dataset”, predict on “column”

| 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 |

- Almost perfect generalization of the models even though networks come from very different applications