Networks with Signed Edges

CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

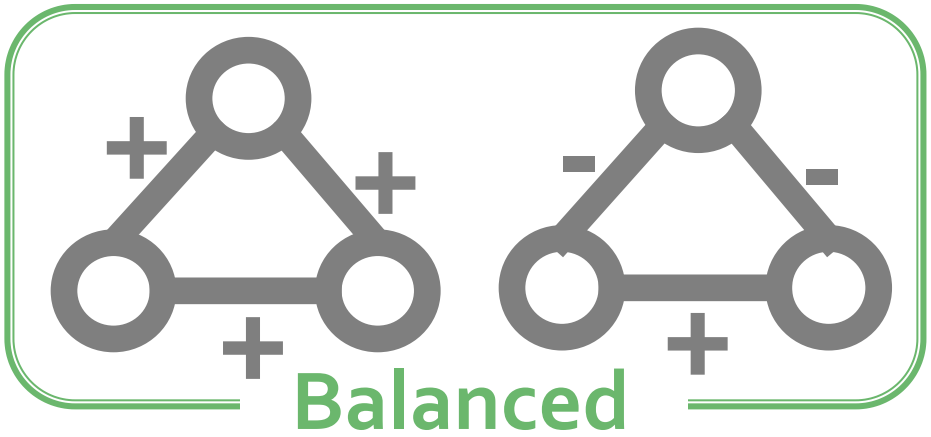
http://cs224w.stanford.edu

Signed Networks

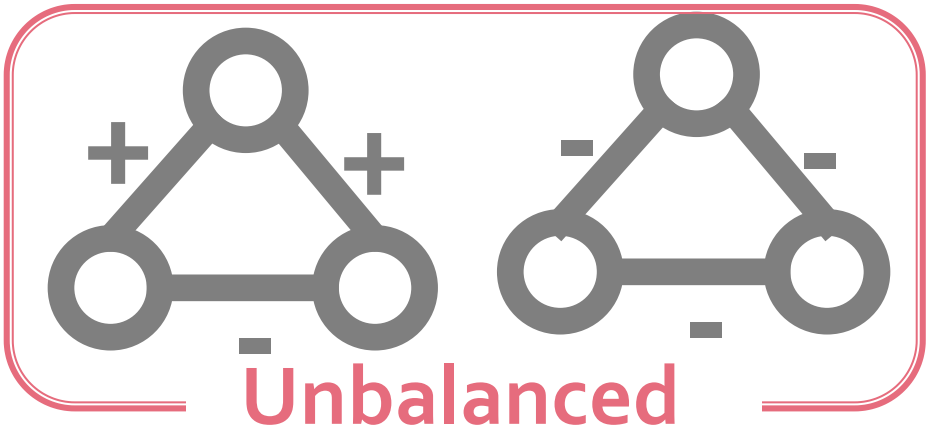
- Networks with positive and negative relationships
- Consider an **undirected complete graph**
- Label each edge as either:
  - **Positive:** friendship, trust, positive sentiment, …
  - **Negative:** enemy, distrust, negative sentiment, …
- Examine triples of connected nodes A, B, C

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:



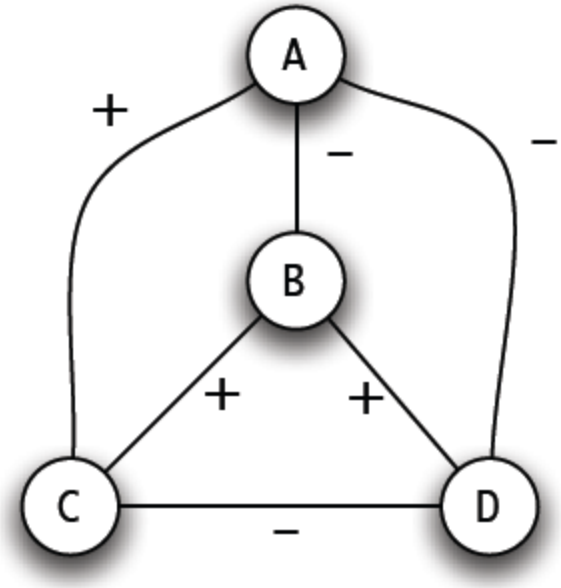
Consistent with "friend of a friend" or "enemy of the enemy" intuition



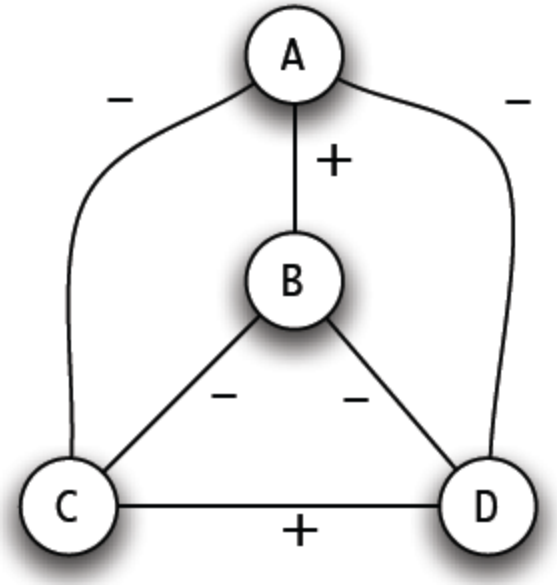
Inconsistent with the "friend of a friend" or "enemy of the enemy" intuition

Balanced/unbalanced networks

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



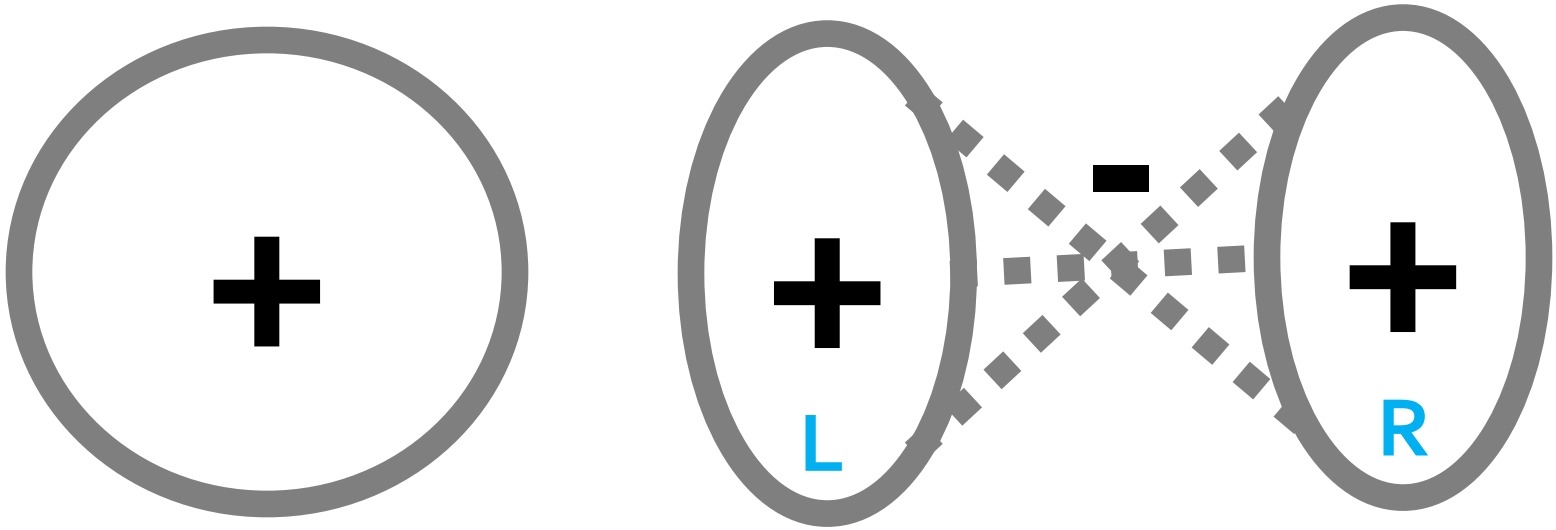
Unbalanced



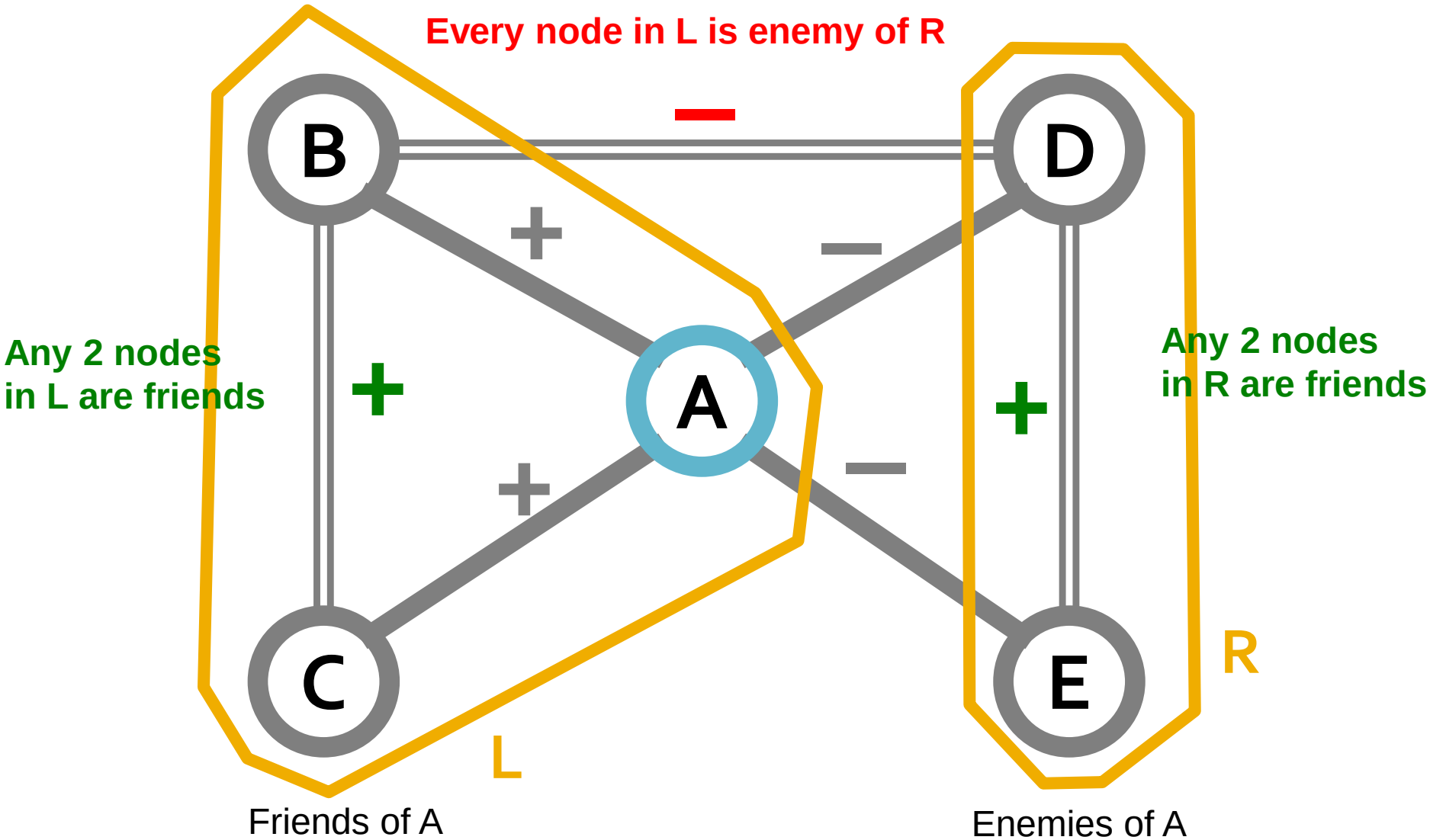
Balanced

Local Balance → Global 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 sets where negative edges only point between the sets

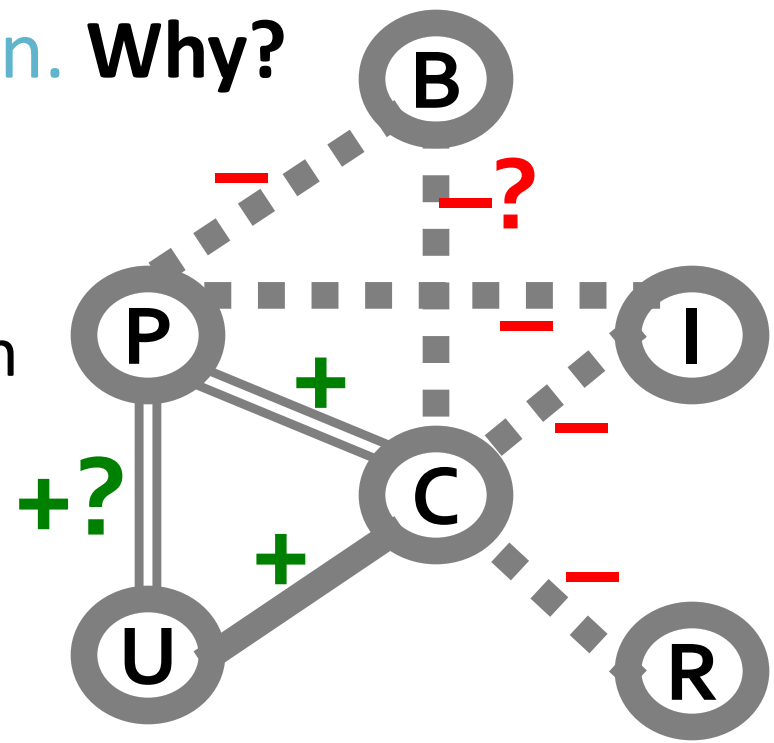


Analysis of Balance

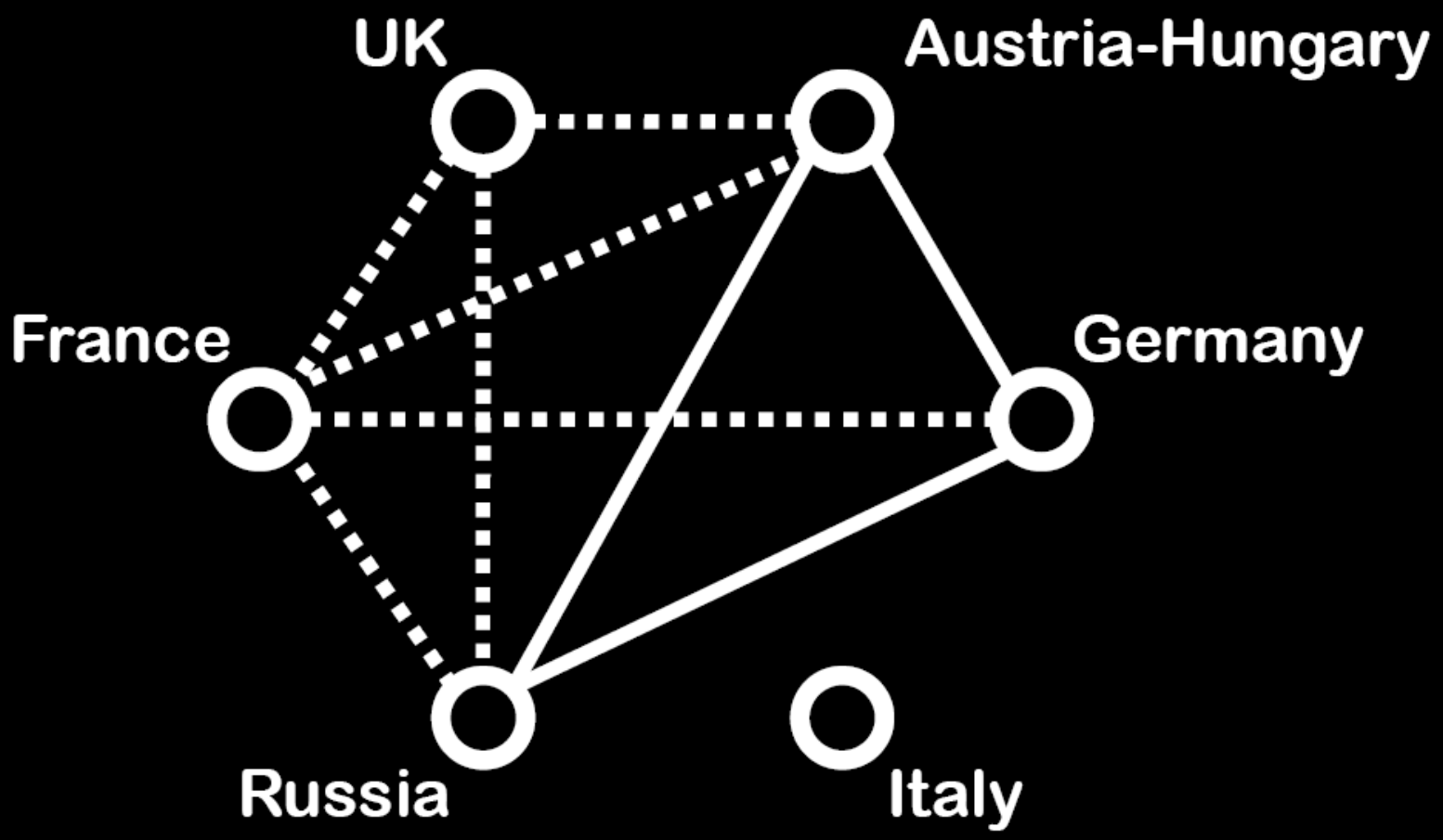


Example: International Relations

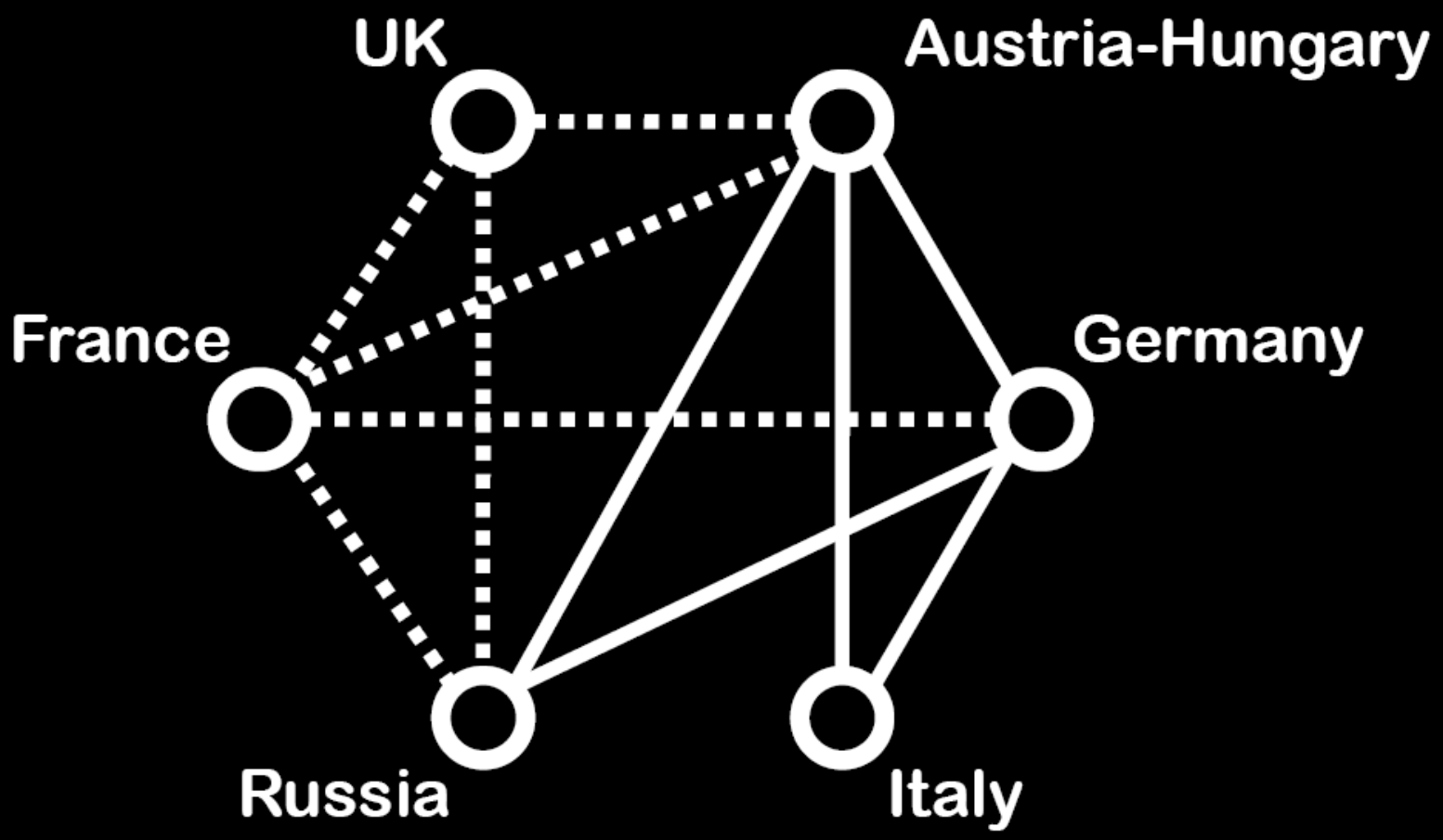
- **International relations:**
  - **Positive** edge: alliance
  - **Negative** edge: animosity
- Separation of Bangladesh from Pakistan in 1971: US supports Pakistan. **Why?**
  - USS**R** was enemy of **C**hina
  - **C**hina was enemy of **I**ndia
  - **I**ndia was enemy of **P**akistan
  - **U**S was friendly with **C**hina
  - China vetoed Bangladesh from U.N.



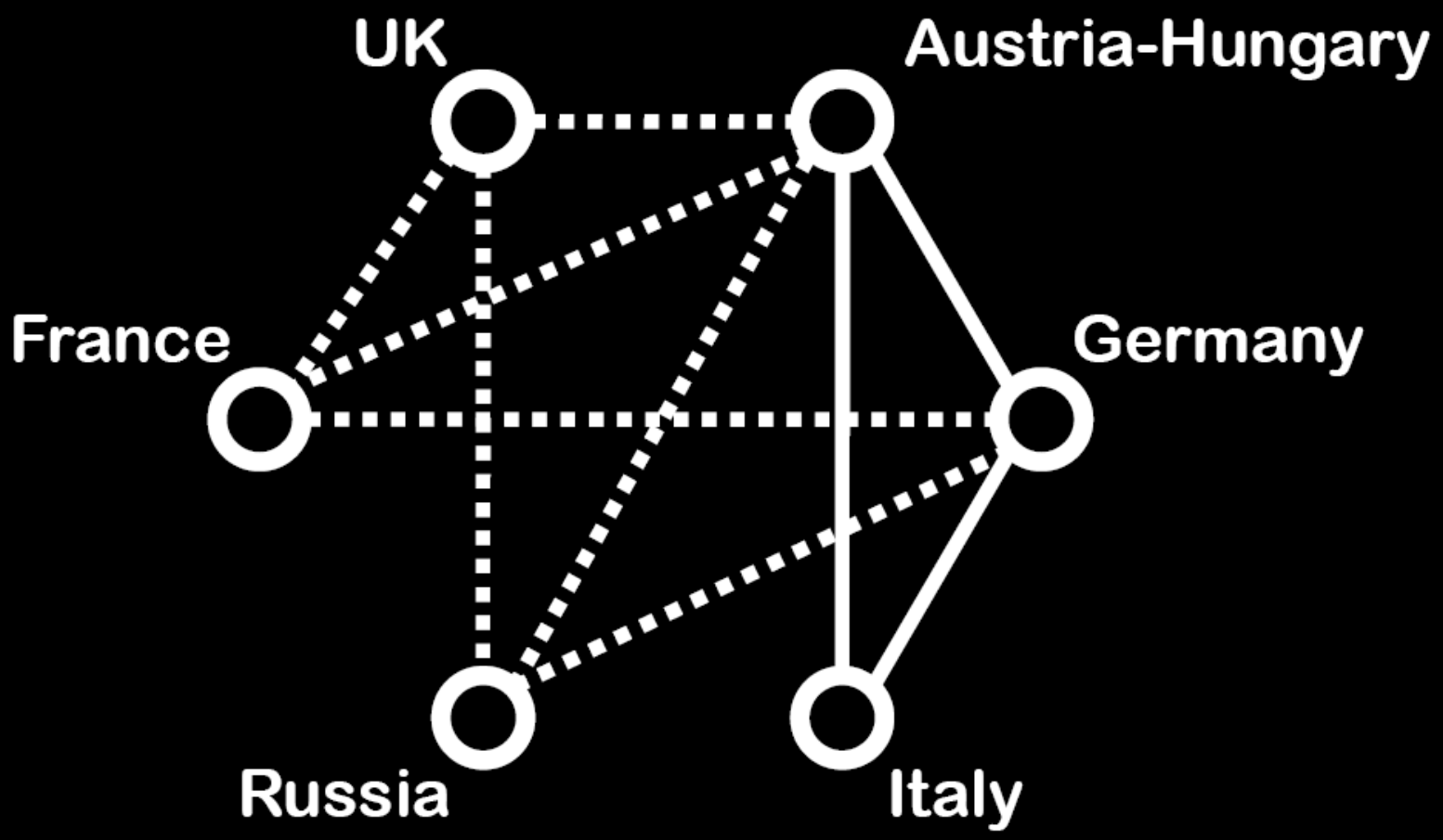
1872-1881



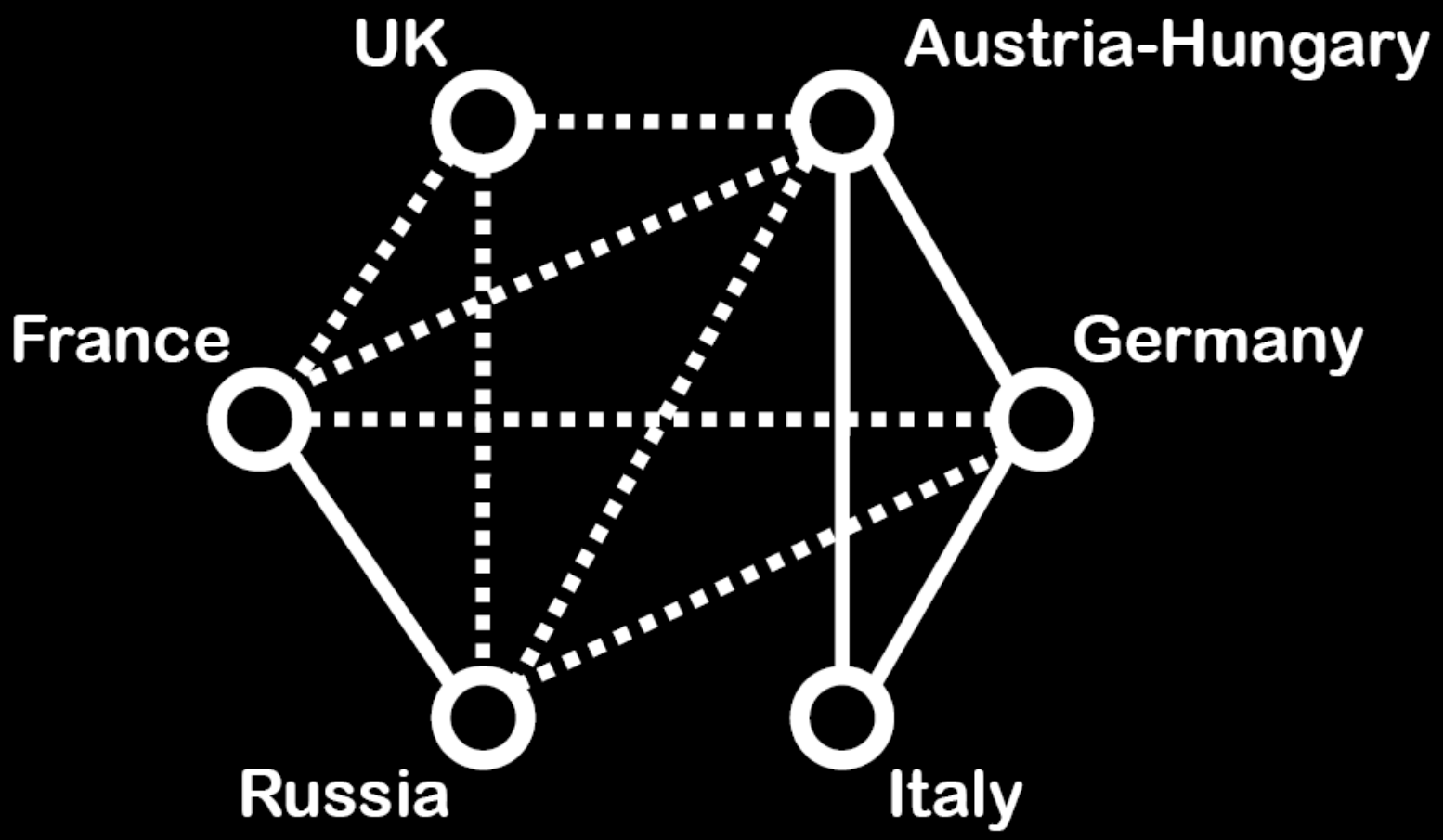
1882



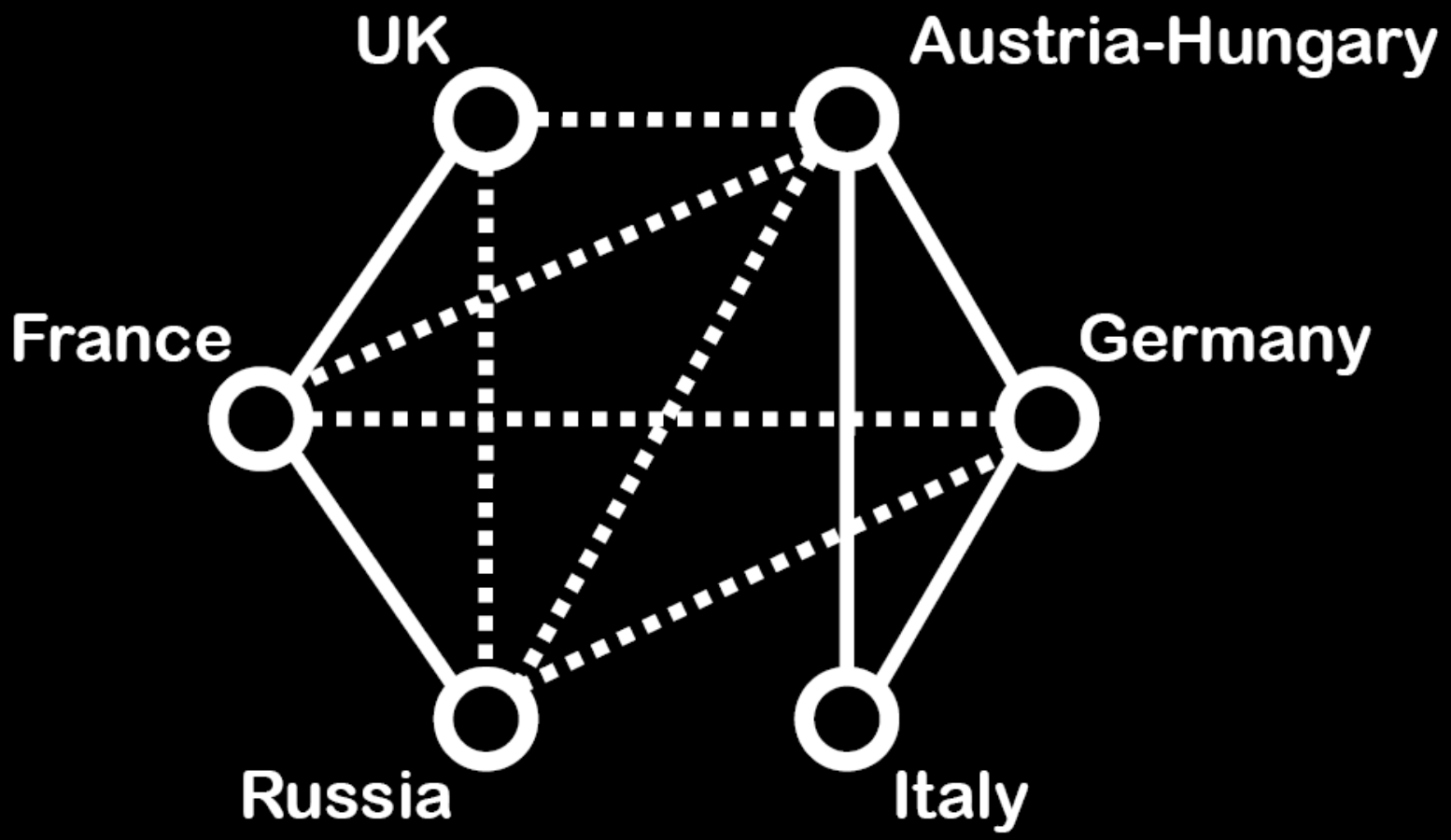
1890



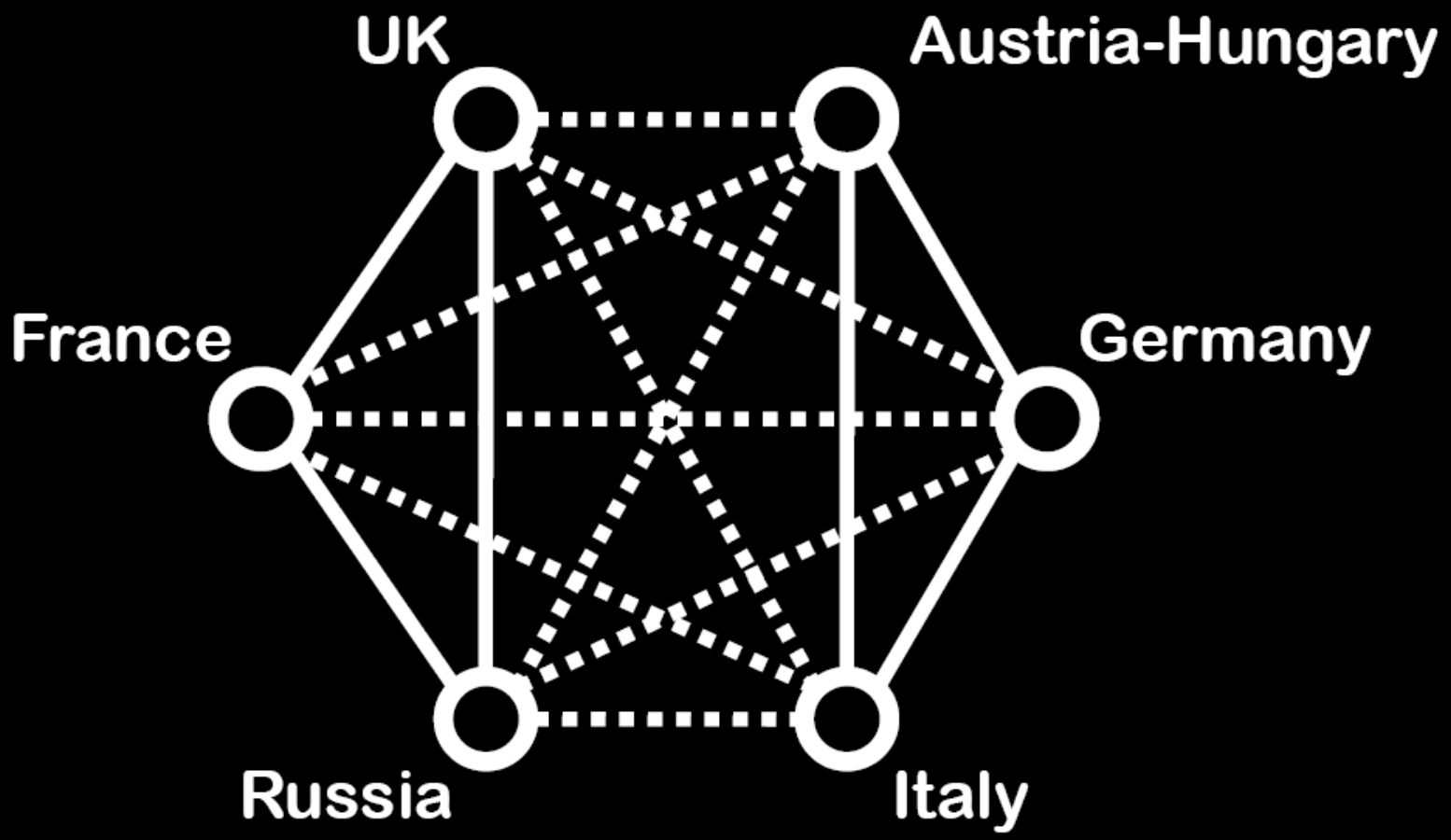
1891-1894



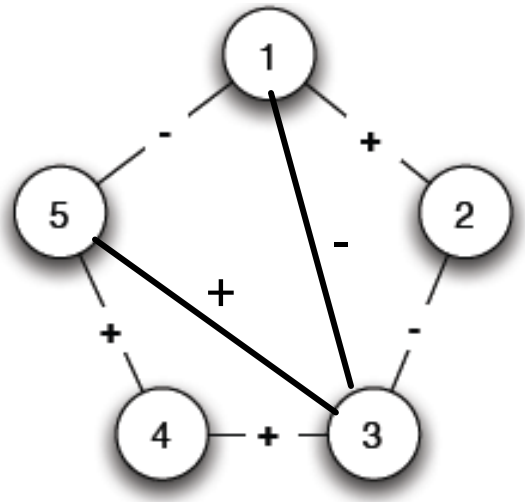
1904



1907

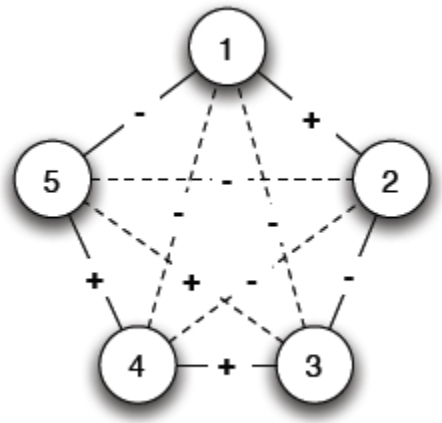


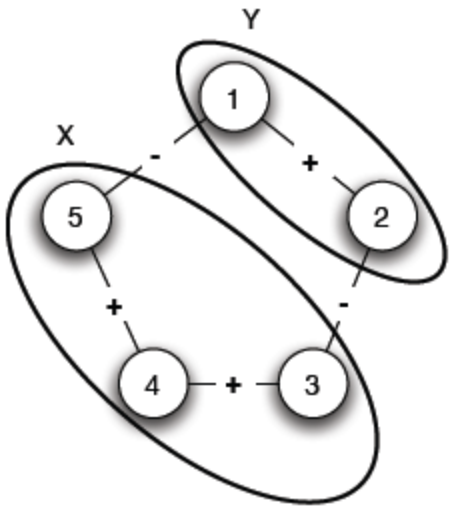
Balance in General Networks



Balanced?

- **Def 1: Local view**
  - Fill in the missing edges to achieve balance
- **Def 2: Global view**
  - Divide the graph into two coalitions
- The 2 defs. are **equivalent!**

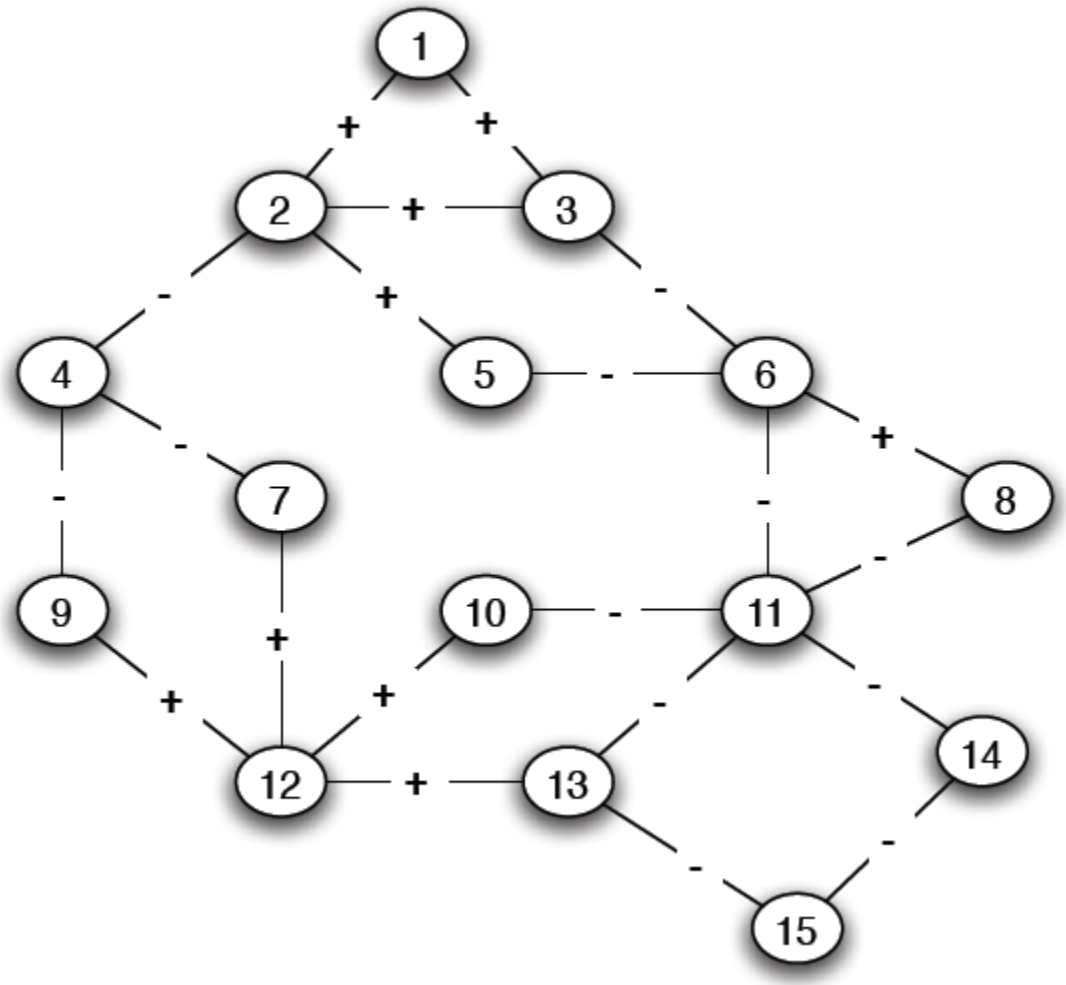




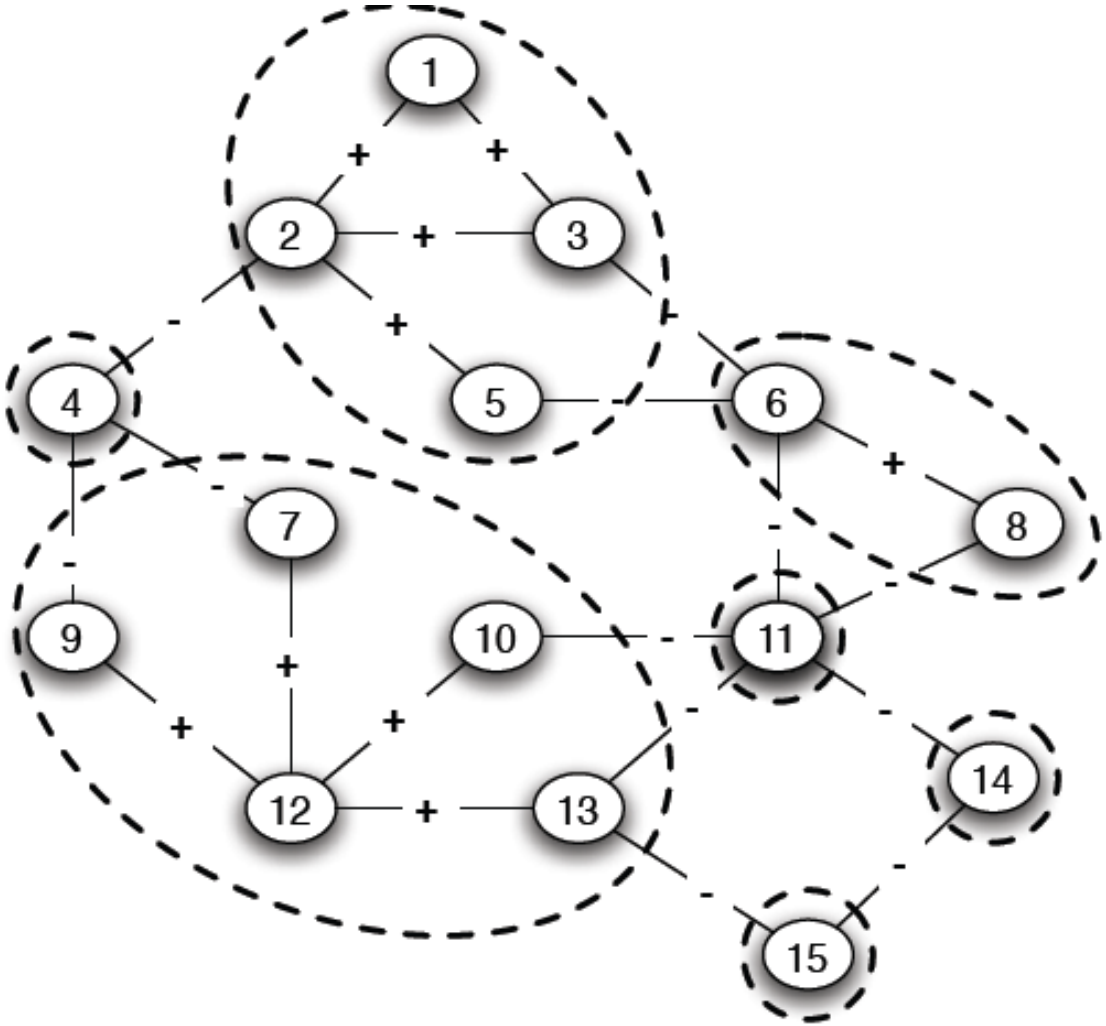
Is a Signed Network Balanced?

- 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
  - 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

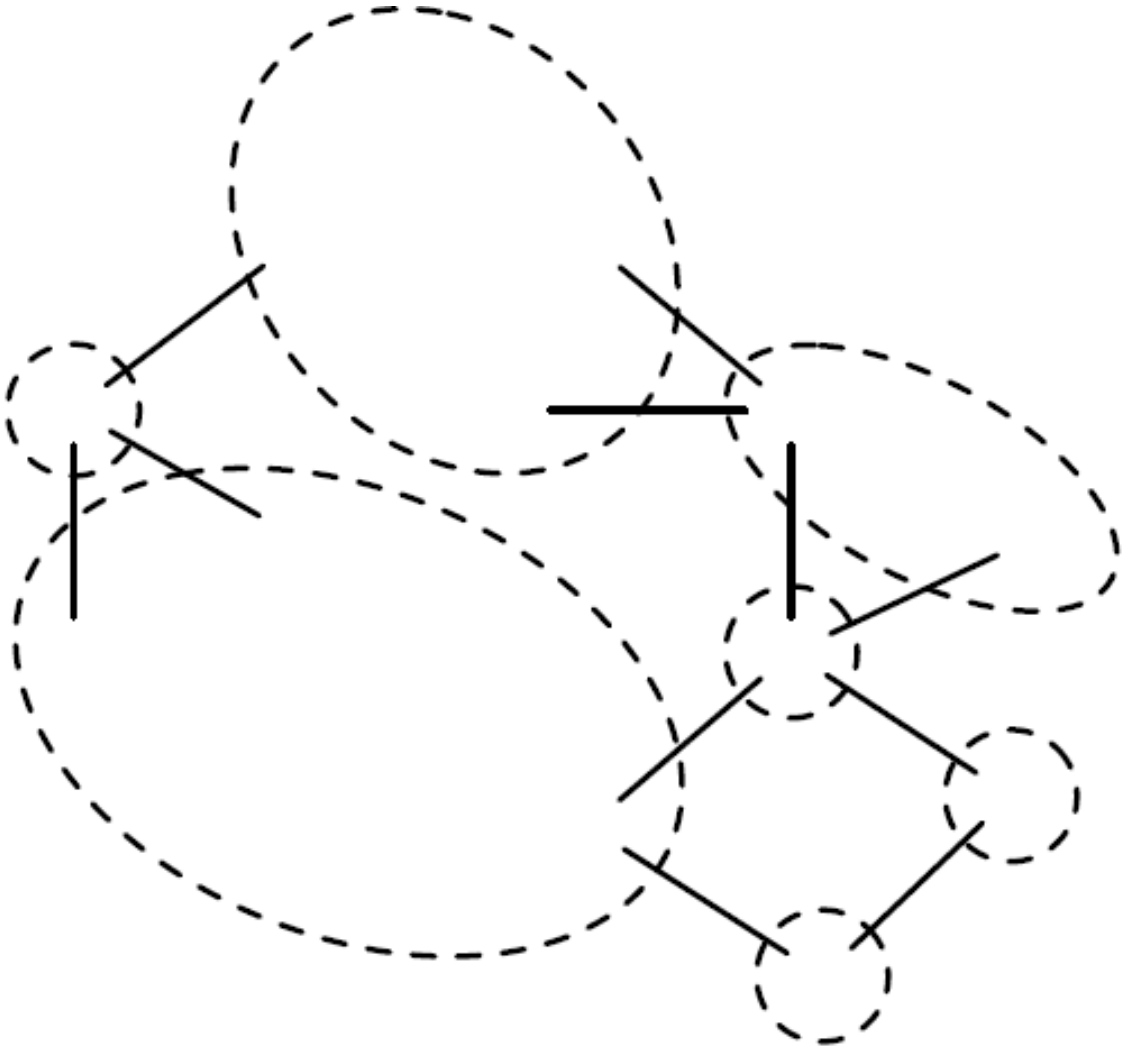
Signed Graph: Is it Balanced?



Positive Connected Components

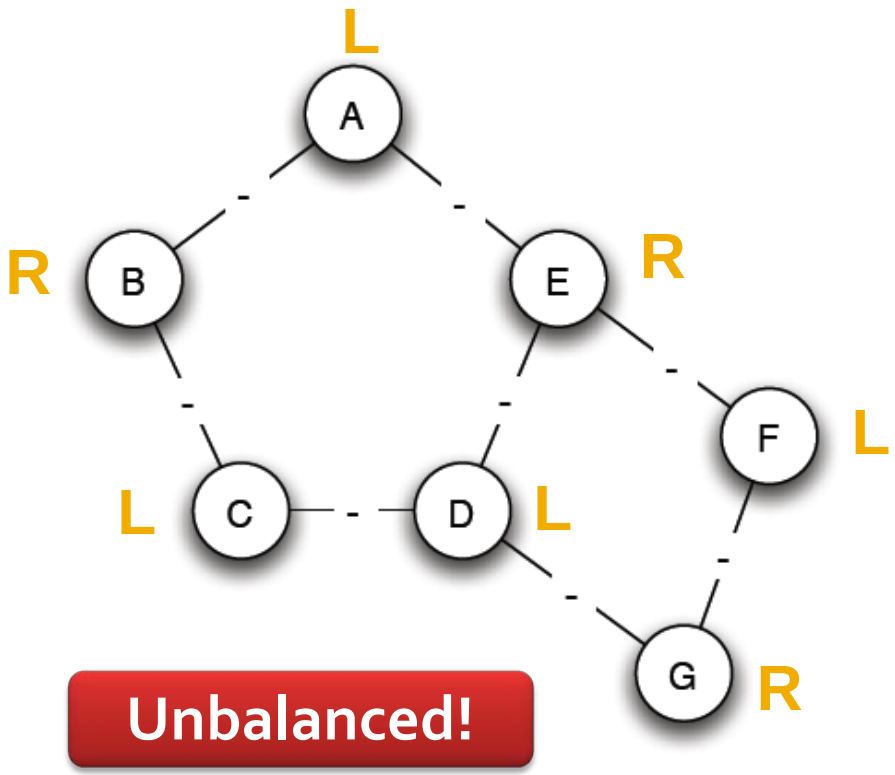


Reduced Graph on Super-Nodes



BFS on Reduced Graph

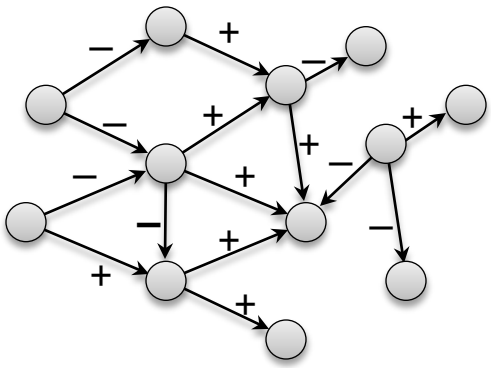
- Using BFS assign each node a **side**
- Graph is **unbalanced** if any two super-nodes are assigned the **same side**



[CHI '10]

Real Large Signed Networks

- Each link A→B is **explicitly** tagged with a sign:
  - Epinions: Trust/Distrust
    - Does A trust B's product reviews?
      (only positive links are visible)
  - 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

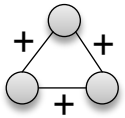
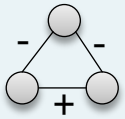
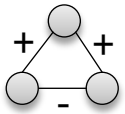
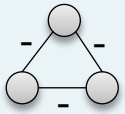


	Epinions	Slashdot	Wikipedia
Nodes	119,217	82,144	7,118
Edges	841,200	549,202	103,747
+ edges	85.0%	77.4%	78.7%
− edges	15.0%	22.6%	21.2%

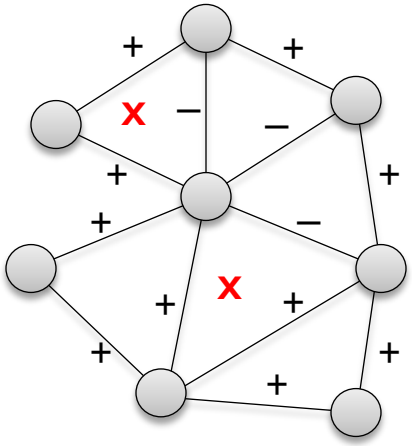
[CHI '10]

Balance in our network data

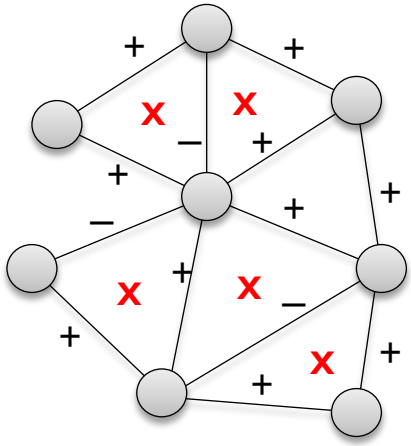
- **Does structural balance hold?**

Triad	Epinions		Wikipedia		Balance
	P(T)	$P_0(T)$	P(T)	$P_0(T)$	
	0.87	0.62	0.70	0.49	✓
	0.07	0.05	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
$P_0(T)$… triad probability if the signs would be random



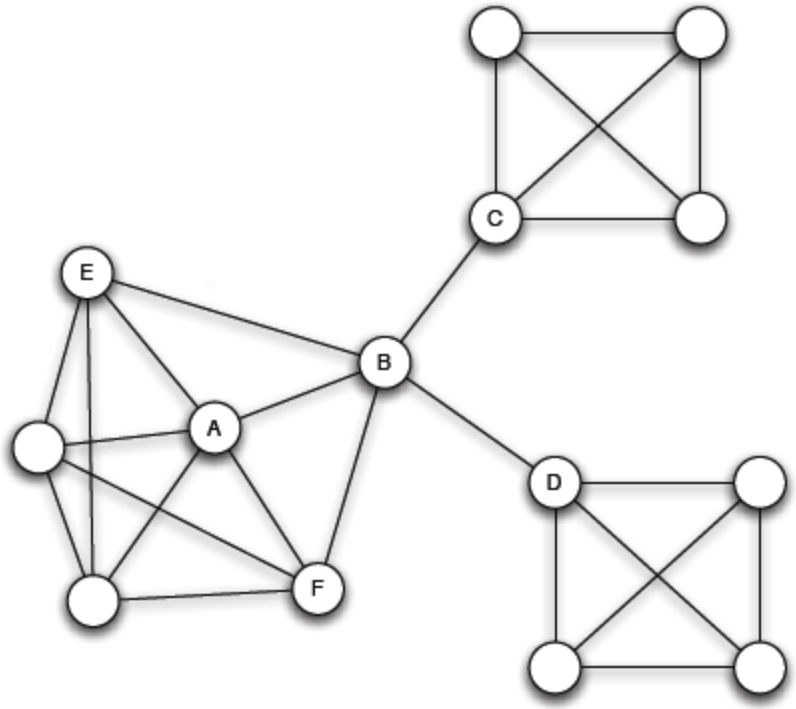
Real data



Shuffled data

Global Structure of Signed Nets

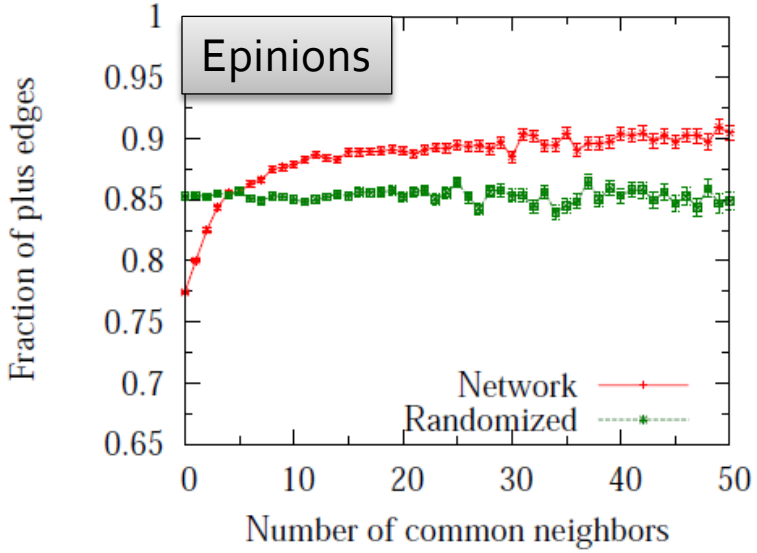
- Intuitive picture of social network in terms of densely linked clusters
- **How does structure interact with links?**
- **Embeddedness of link (A,B):** Number of shared neighbors

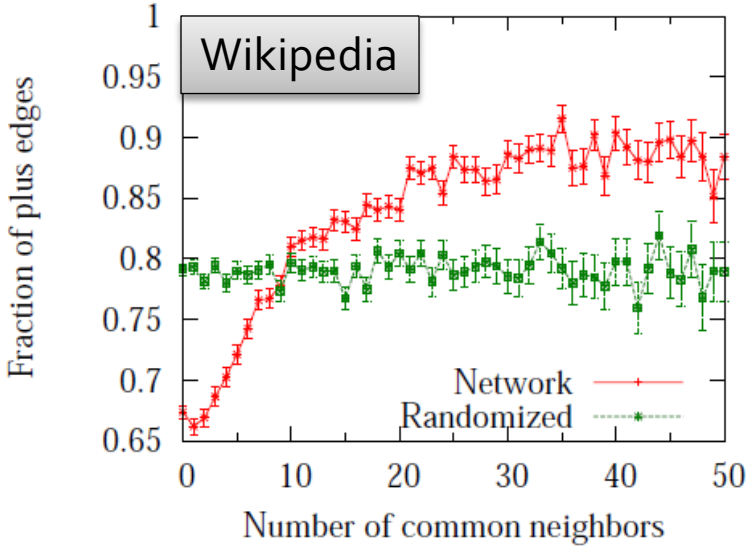


[CHI '10]

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



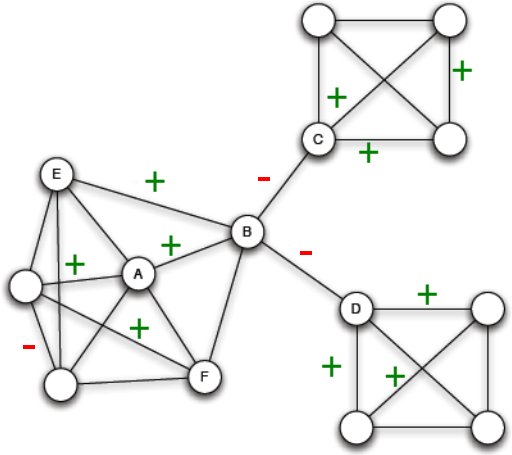


[CHI '10]

Global Structure of Signed Nets

	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

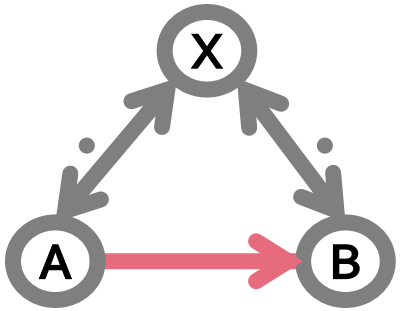
- **Clustering:**
  - +net: More clustering than baseline
  - –net: Less clustering than baseline
- Size of connected component:
  - +/–net: Smaller than the baseline

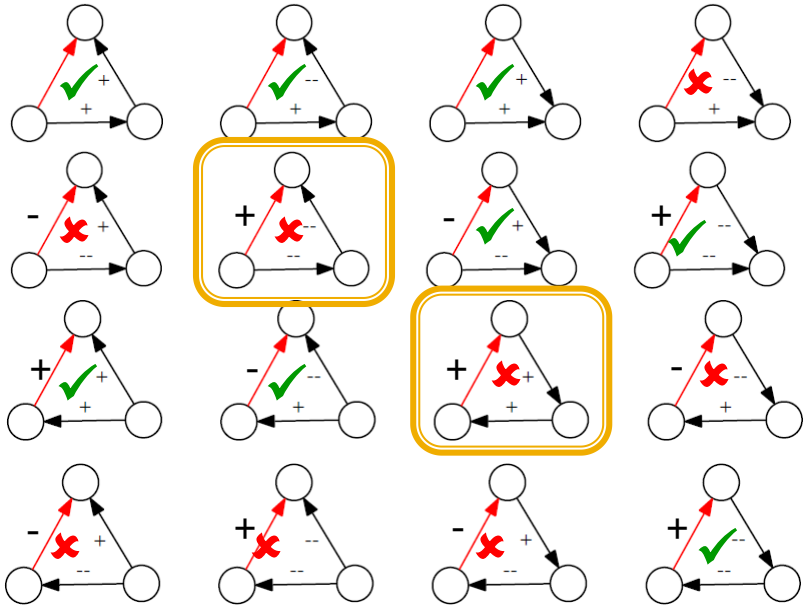


[CHI '10]

Evolving Directed Networks

- **New setting:**
  - Links are **directed** and **created over time**



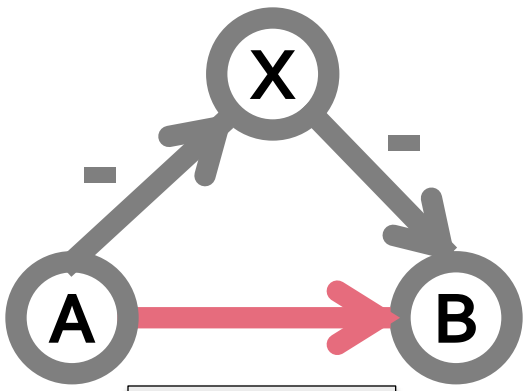


16 *2 signed directed triads

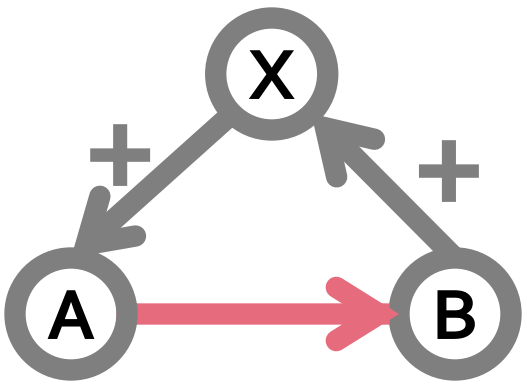
- **How many △ are now explained by balance?**
  - **Only half** (8 out of 16)
- **Is there a better explanation?** Yes. **Status.**

Alternate Theory: Status

- Links are **directed and created over time**
- **Status theory** [Davis-Leinhardt '68, Guha et al. '04, Leskovec et al. '10]
  - Link A $\xrightarrow{+}$ B means: B has **higher** status than A
  - Link A $\xrightarrow{-}$ B means: B has **lower** status than A
- **Status and balance give different predictions:**



Balance: +
Status: –

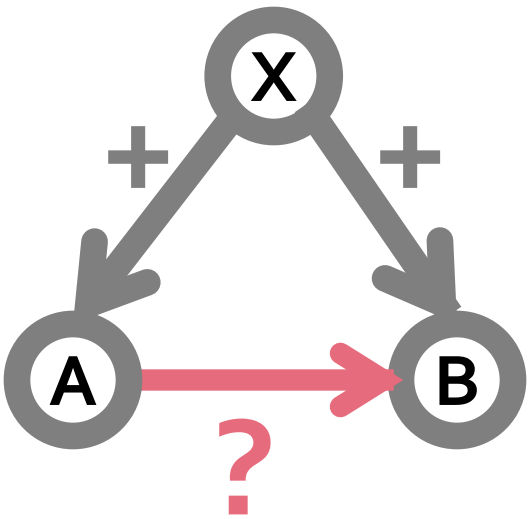


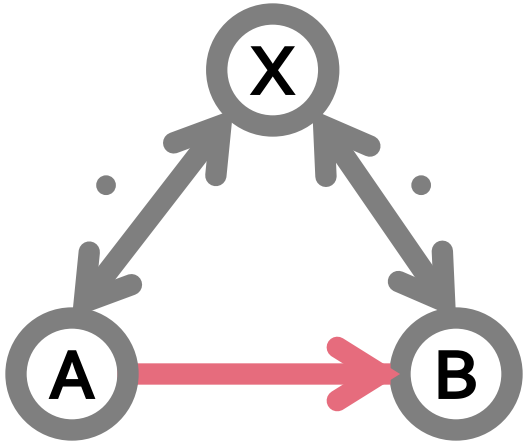
Balance: +
Status: –

[CHI '10]

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?**
- We need to formalize:
  - Links are embedded in triads:
    - Provides **context for signs**
  - Users are **heterogeneous** in their linking behavior

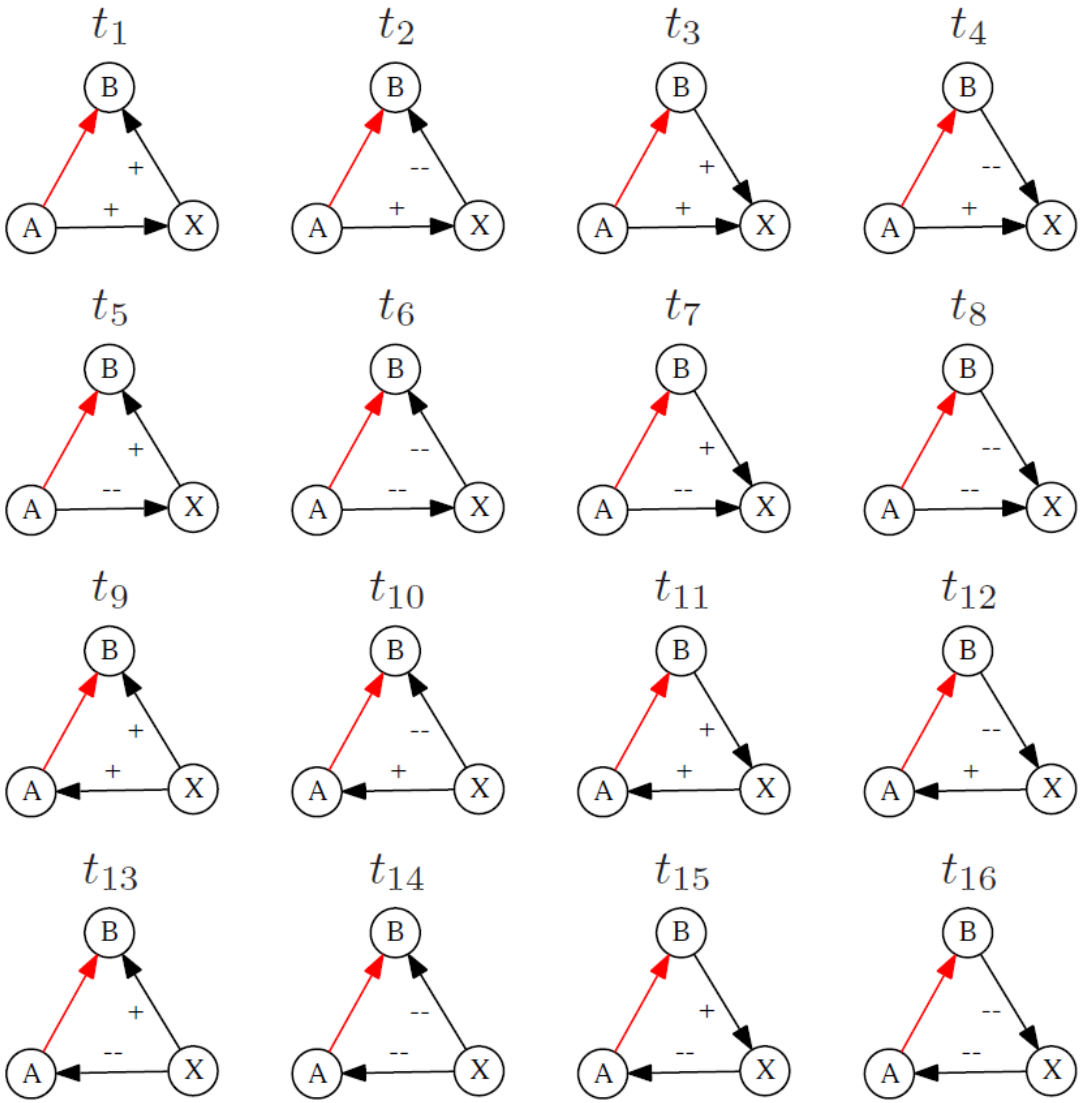




[CHI '10]

16 Types of Context

- Link (A,B) appears in the **context** (A,B; X)
- **16 different contextualized links:**

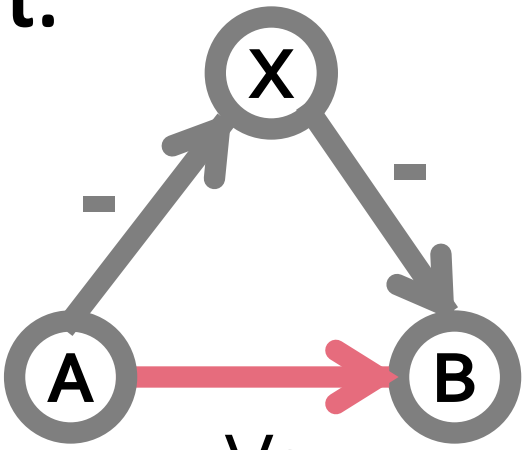


[CHI '10]

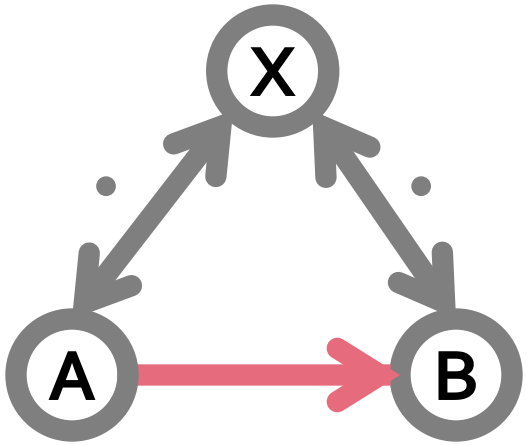
Generative (Receptive) Surprise

- **Surprise:** How much behavior of user **deviates** from **baseline** in **context t:**
  - $(A_1, B_1; X_1), \ldots, (A_n, B_n; X_n)$ … instances of contextualized link *t*
  - *k* of them closed with a plus
  - $p_g(A_i)$… generative baseline of A_i
    - empirical prob. of A_i giving a plus
- Then: **generative surprise of triad type *t*:**

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

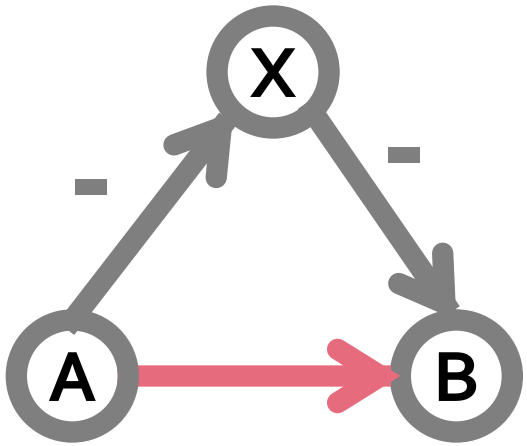


Vs.

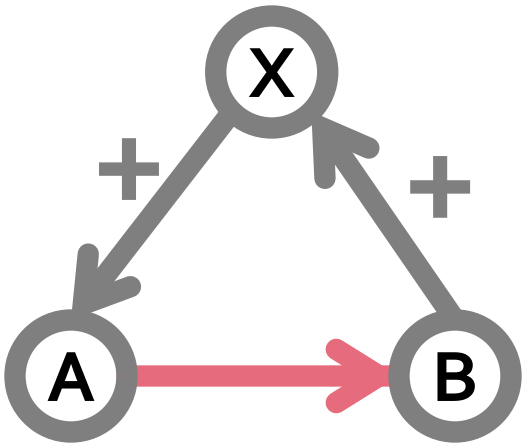


Status: Two Examples

- **Two basic examples:**



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

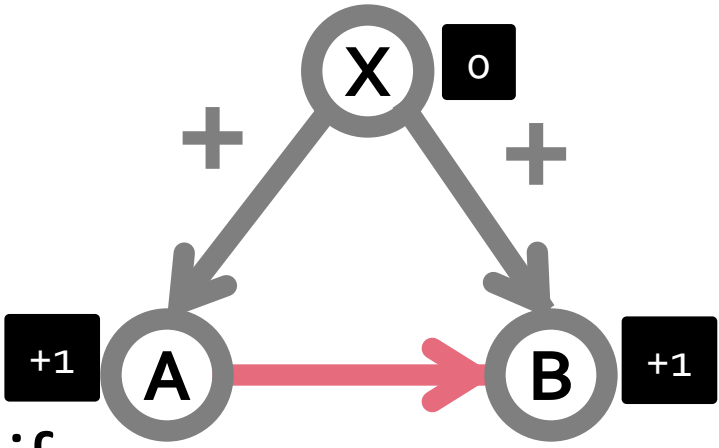


Gen. surprise of A: —
Rec. surprise of B: —

[CHI '10]

Consistency with Status

- **Determine node status:**
  - Assign X status 0
  - Based on signs and directions of edges set status of A and B



Status-consistent if:
Gen. surprise > 0
Rec. surprise < 0

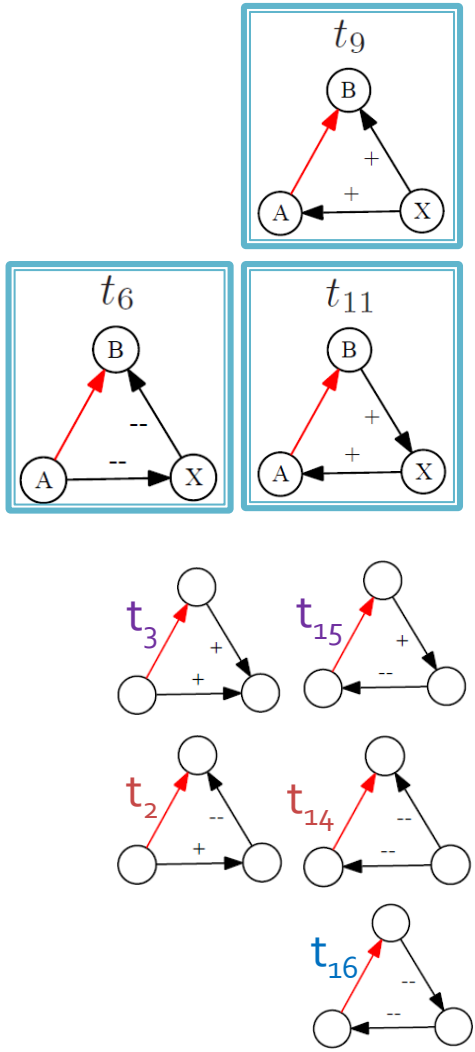
- Surprise is **status**-consistent, if:
  - **G**en. surprise is status-consistent if it has **same** sign as status of B
  - **R**ec. surprise is status-consistent if it has the **opposite** sign from the status of A
- Surprise is **balance**-consistent, if:
  - If it completes a balanced triad

[CHI '10]

Status vs. Balance (Epinions)

- **Predictions:**

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

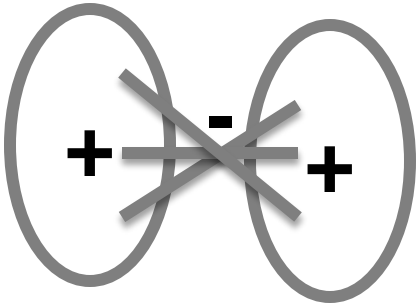


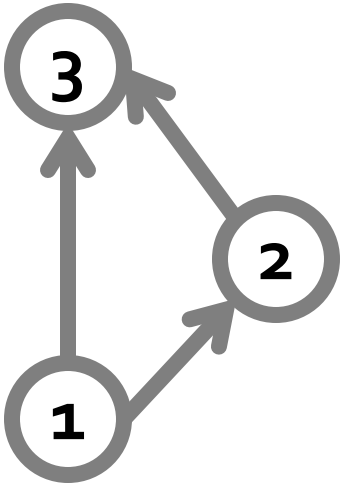
[WWW '10]

From Local to Global Structure

- **Both theories make predictions about the global structure of the network**

- Structural balance – Factions
  - Find coalitions
- Status theory – Global Status
  - Flip direction and sign of minus edges
  - Assign each node a unique status so that edges point from low to high





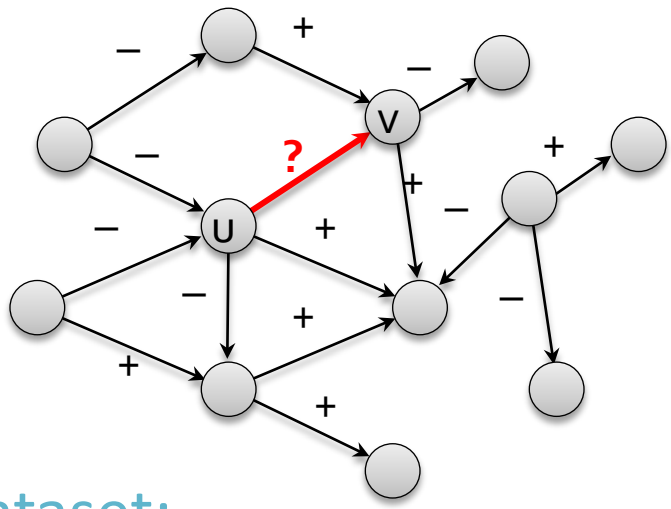
[WWW '10]

From Local to Global Structure

- **Fraction of edges of the network that satisfy Balance and Status?**
- **Observations:**
  - No evidence for global balance beyond the random baselines
    - Real data is 80% consistent vs. 80% consistency under random baseline
  - Evidence for global status beyond the random baselines
    - Real data is 80% consistent, but 50% consistency under random baseline

[WWW '10]

Predicting Edge Signs



Edge sign prediction problem

- Given a network and signs on all but one edge, predict the missing sign

Machine Learning Formulation:

- Predict sign of edge (u,v)
- Class label:
  - +1: positive edge
  - -1: negative edge
- Learning method:
  - Logistic regression

$$P(+|x) = \frac{1}{1 + e^{-(b_0 + \sum_i^n b_i x_i)}}$$

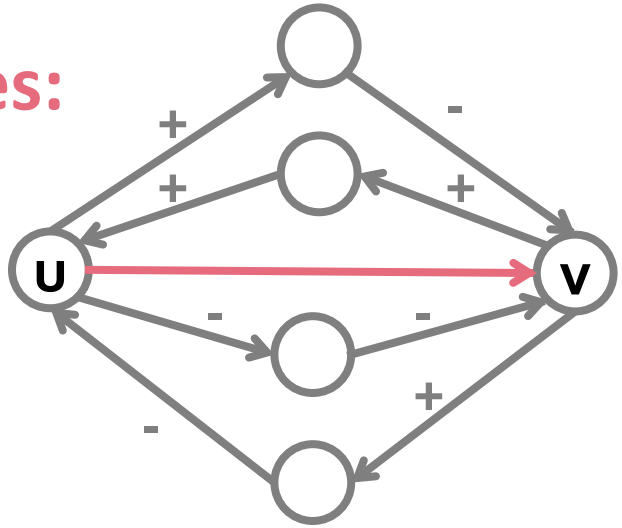
- Dataset:
  - Original: 80% +edges
  - Balanced: 50% +edges
- Evaluation:
  - Accuracy and ROC curves
- Features for learning:
  - Next slide

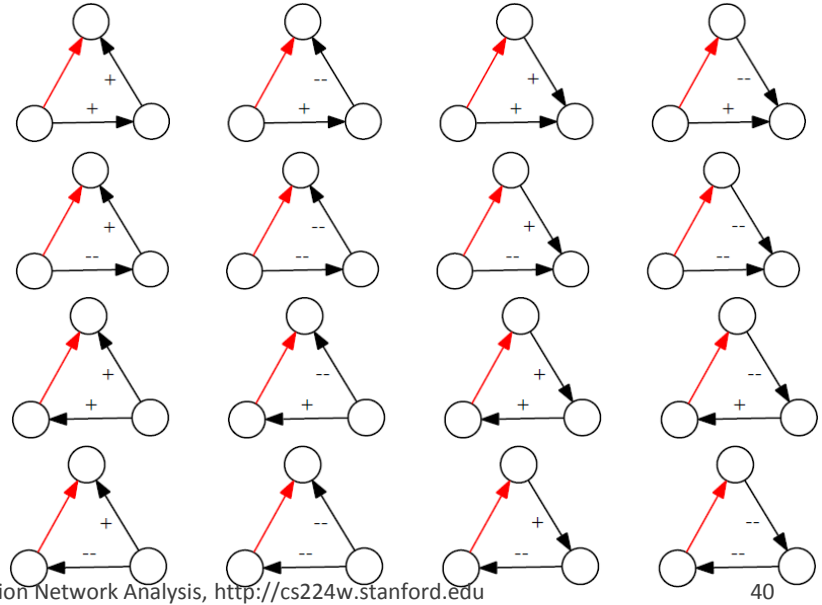
[WWW '10]

Features for Learning

For each edge (u,v) create features:

- Triad counts (16):
  - Counts of signed triads edge u→v takes part in
- Node degree (7 features):
  - Signed degree:
    - $d^{+}_{out}(u)$, $d^{-}_{out}(u)$, $d^{+}_{in}(v)$, $d^{-}_{in}(v)$
  - Total degree:
    - $d_{out}(u)$, $d_{in}(v)$
  - Embeddedness of edge (u,v)

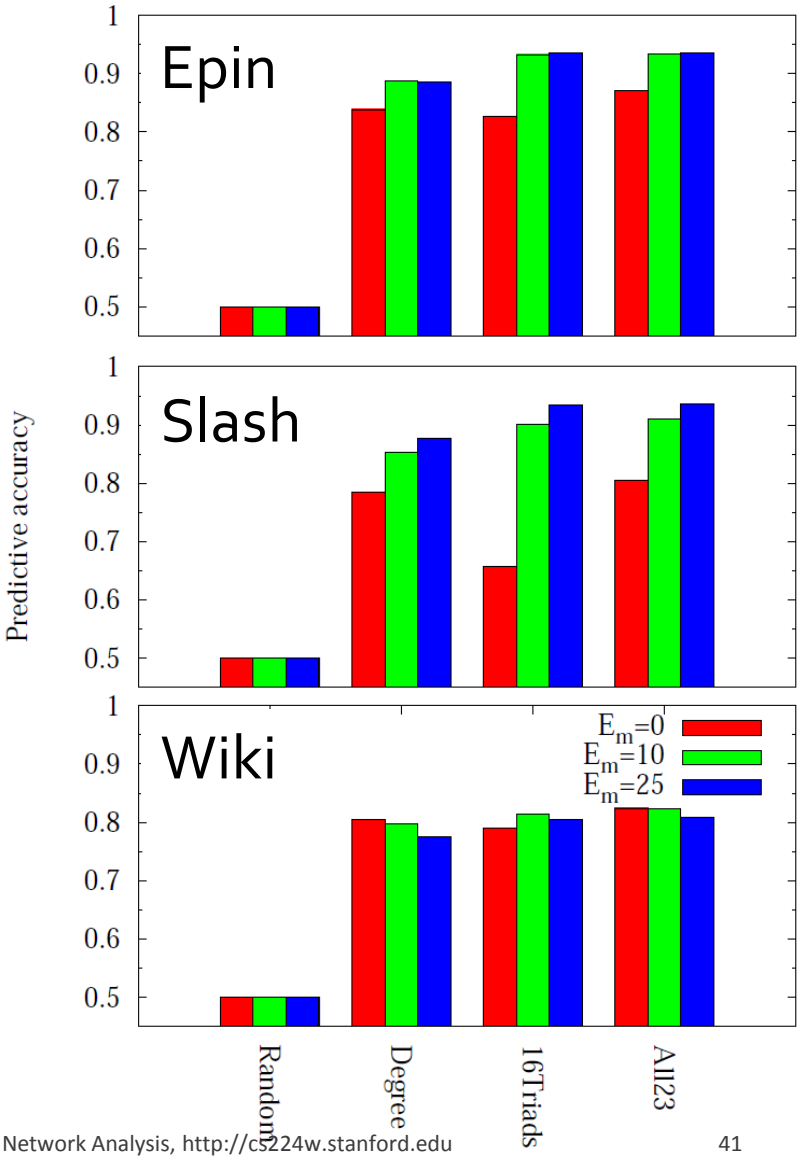




[WWW '10]

Edge sign prediction

- Classification Accuracy:
  - Epinions: 93.5%
  - Slashdot: 94.4%
  - Wikipedia: 81%
- Signs can be modeled from local network structure alone
  - Trust propagation model of [Guha et al. '04] has 14% error on Epinions
- Triad features perform less well for less embedded edges
- Wikipedia is harder to model:
  - Votes are publicly visible



Balance and Status: Complete Model

Feature	Bal	Stat	**Epin**	**Slashd**	**Wikip**
const			-0.2	0.02	-0.2
$\bullet \xrightarrow{+} \bullet \xrightarrow{+} \bullet$	1	1	**0.5**	**0.9**	**0.3**
$\bullet \xrightarrow{+} \bullet \xrightarrow{-} \bullet$	-1	0	**-0.5**	**-0.9**	**-0.4**
$\bullet \xrightarrow{-} \bullet \xrightarrow{+} \bullet$	-1	0	**-0.4**	**-1.1**	**-0.3**
$\bullet \xrightarrow{-} \bullet \xrightarrow{-} \bullet$	1	-1	**-0.7**	**-0.6**	**-0.8**
$\circ \xrightarrow{+} \circ \xleftarrow{+} \circ$	1	0	0.3	0.4	0.05
$\circ \xrightarrow{+} \circ \xleftarrow{-} \circ$	-1	1	-0.01	-0.1	-0.01
$\bullet \xrightarrow{-} \bullet \xleftarrow{+} \bullet$	-1	-1	**-0.9**	**-1.2**	**-0.2**
$\circ \xrightarrow{-} \circ \xleftarrow{-} \circ$	1	0	0.04	-0.07	-0.03
$\circ \xleftarrow{+} \circ \xrightarrow{+} \circ$	1	0	0.08	0.4	0.1
$\bullet \xleftarrow{+} \bullet \xrightarrow{-} \bullet$	-1	-1	**-1.3**	**-1.1**	**-0.4**
$\circ \xleftarrow{-} \circ \xrightarrow{+} \circ$	-1	1	-0.1	-0.2	0.05
$\circ \xleftarrow{-} \circ \xrightarrow{-} \circ$	1	0	0.08	-0.02	-0.1
$\circ \xleftarrow{+} \circ \xleftarrow{+} \circ$	1	-1	-0.09	-0.09	-0.01
$\circ \xleftarrow{+} \circ \xleftarrow{-} \circ$	-1	0	-0.05	-0.3	-0.02
$\circ \xleftarrow{-} \circ \xleftarrow{+} \circ$	-1	0	-0.04	-0.3	0.05
$\circ \xleftarrow{-} \circ \xleftarrow{-} \circ$	1	1	-0.02	0.2	-0.2

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 "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

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

Concluding Remarks

- **Signed networks provide insight into how social computing systems are used:**
  - Status vs. Balance
  - Different role of reciprocated links
  - Role of embeddedness and public display
- Sign of relationship can be reliably predicted from the local network context
  - ~90% accuracy sign of the edge

Concluding Remarks

- More evidence that networks are globally organized based on status
- People use signed edges consistently regardless of particular application
  - Near perfect generalization of models across datasets
- Many further directions:
  - Status difference of nodes A and B [ICWSM '10]: