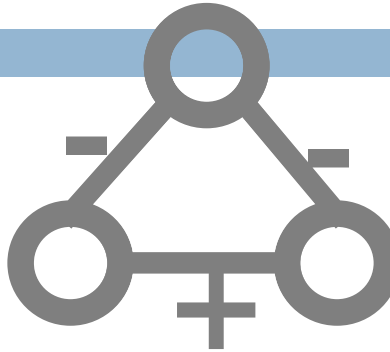
LECTURE 11:NETWORKS WITH SIGNED EDGES

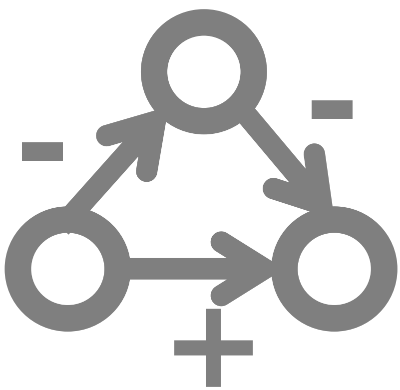
COMP4641: Social Information Network Analysis and Engineering

Wednesday March 24th 2015

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 for this lecture:**
  - **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

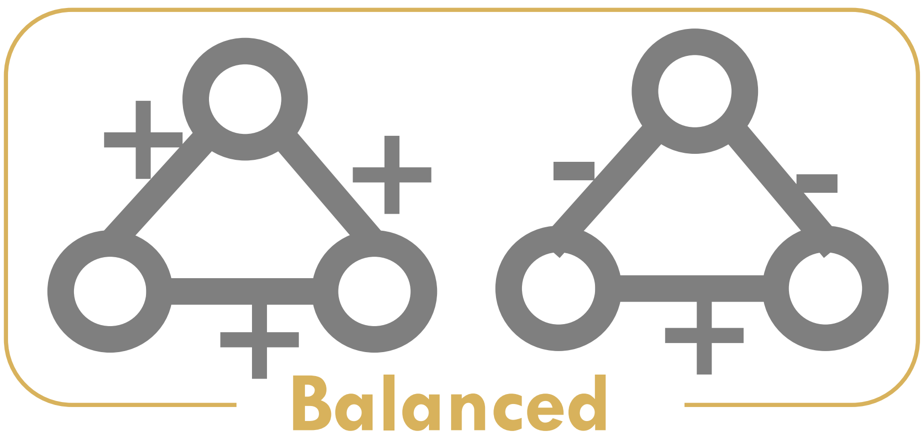
- Networks with positive and negative relationships

- Consider an [undirected complete graph]
- Label each edge as either:
  - **Positive:** friendship, trust, positive sentiment, …
  - **Negative:** enemy, distrust, negative sentiment, …

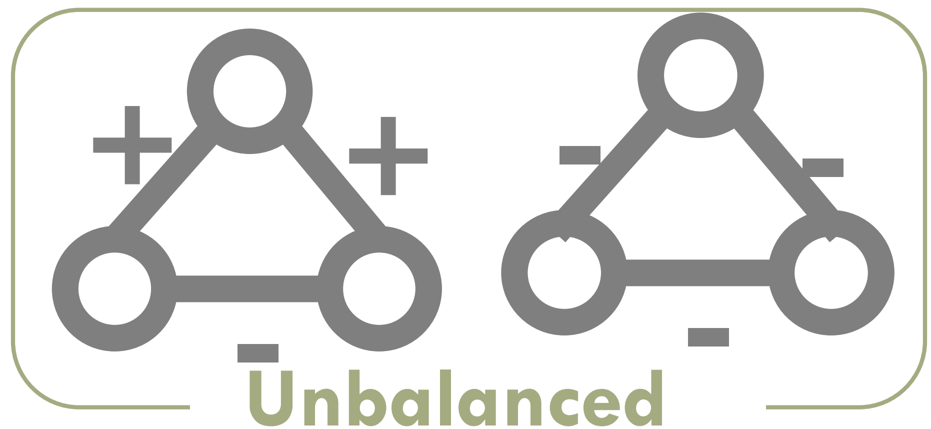
- Examine triples of connected nodes A, B, C

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:



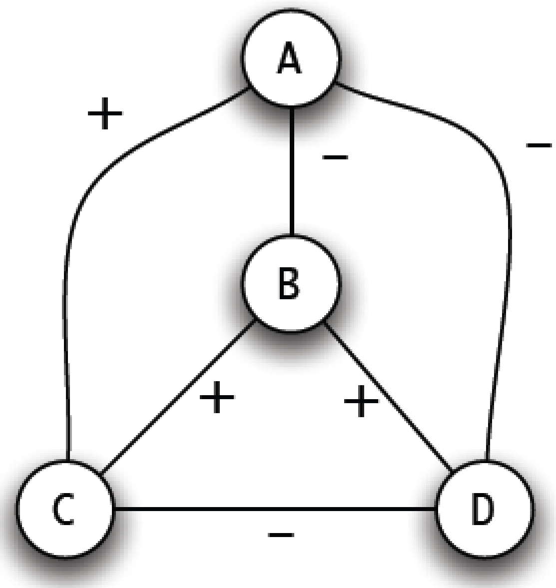
Consistent with "friend of a friend" or "enemy of the enemy" intuition



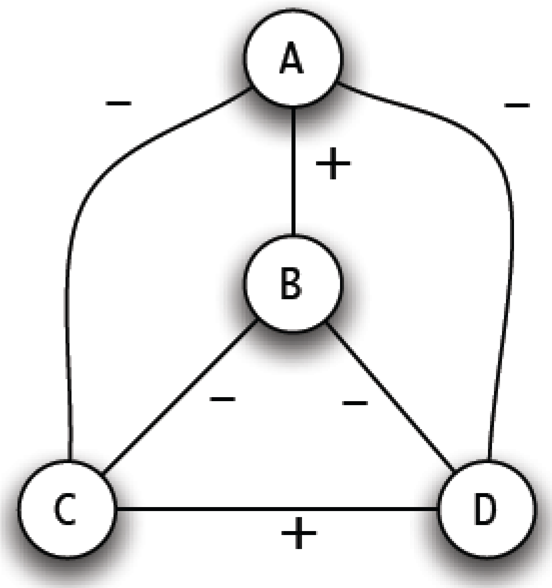
Inconsistent with the "friend of a friend" or "enemy of the enemy" intuition

Balanced/Unbalanced Networks

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



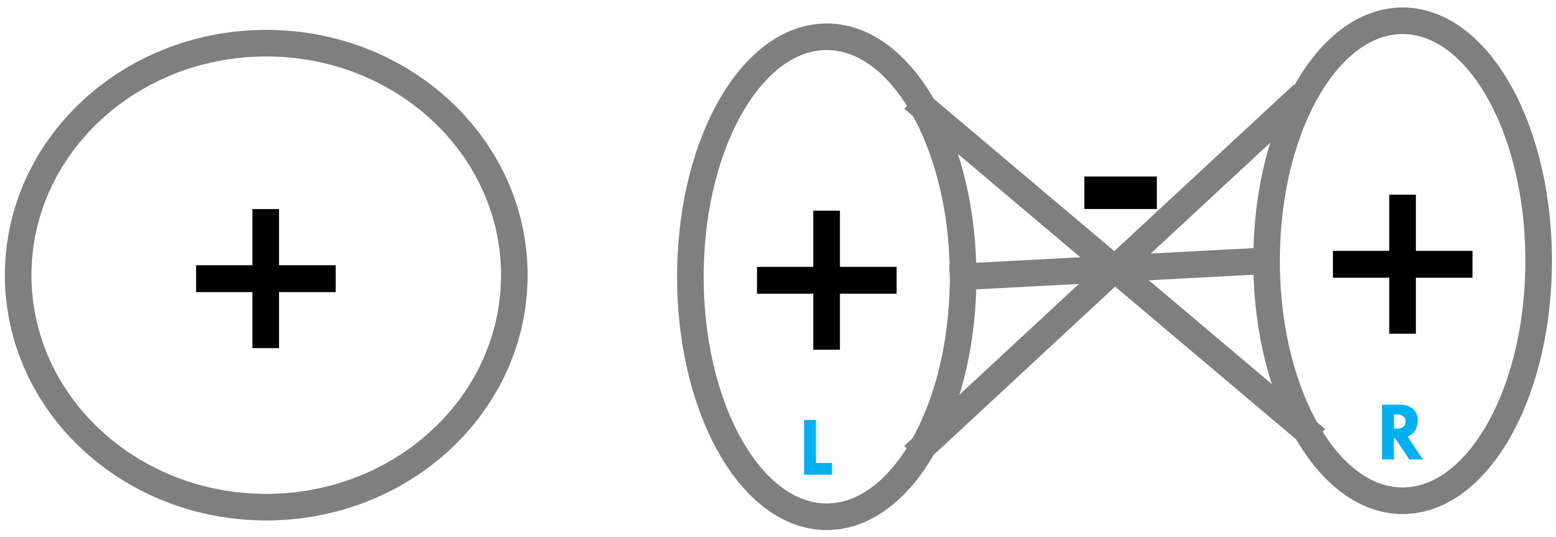
Unbalanced



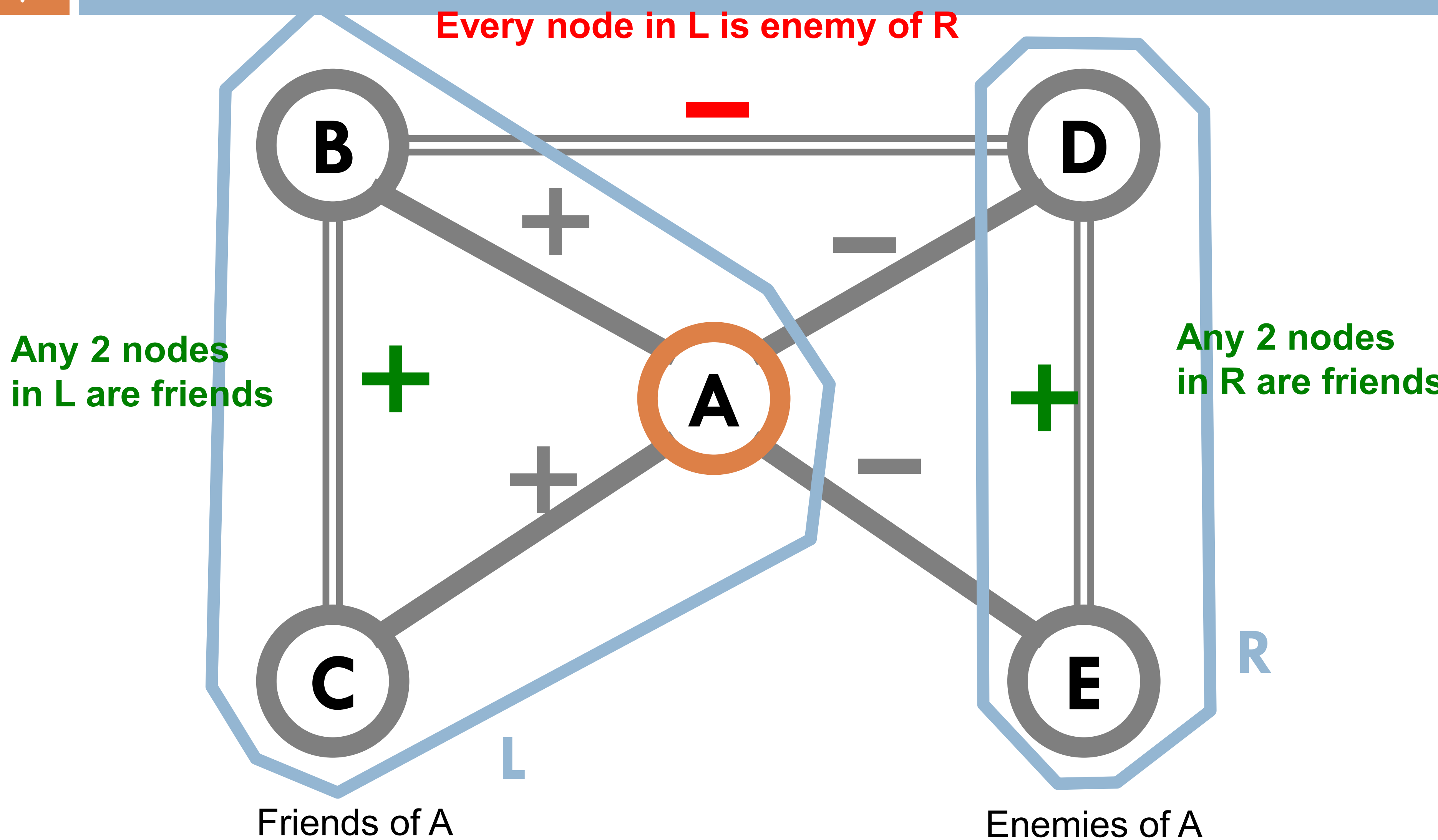
Balanced

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

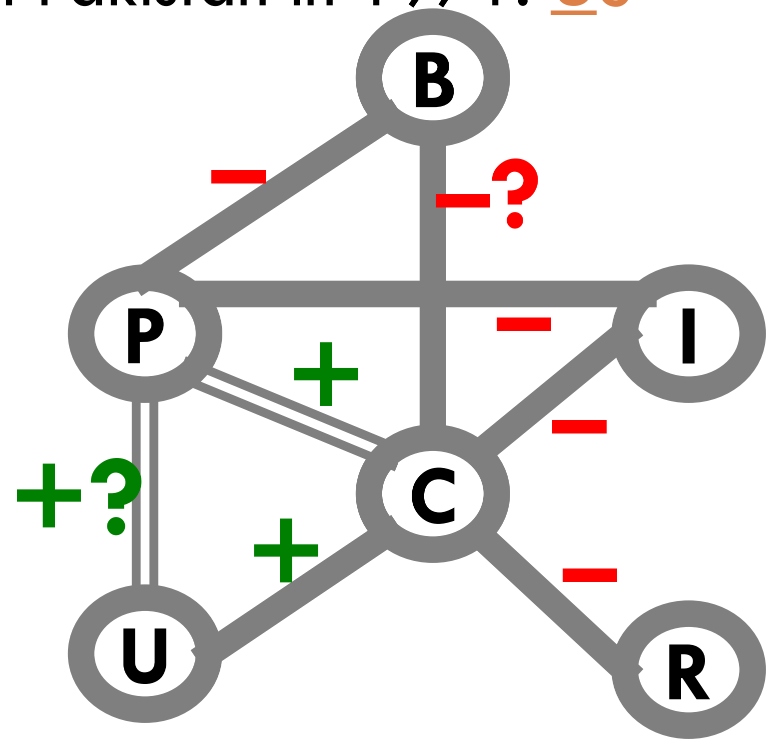


Analysis of Balance

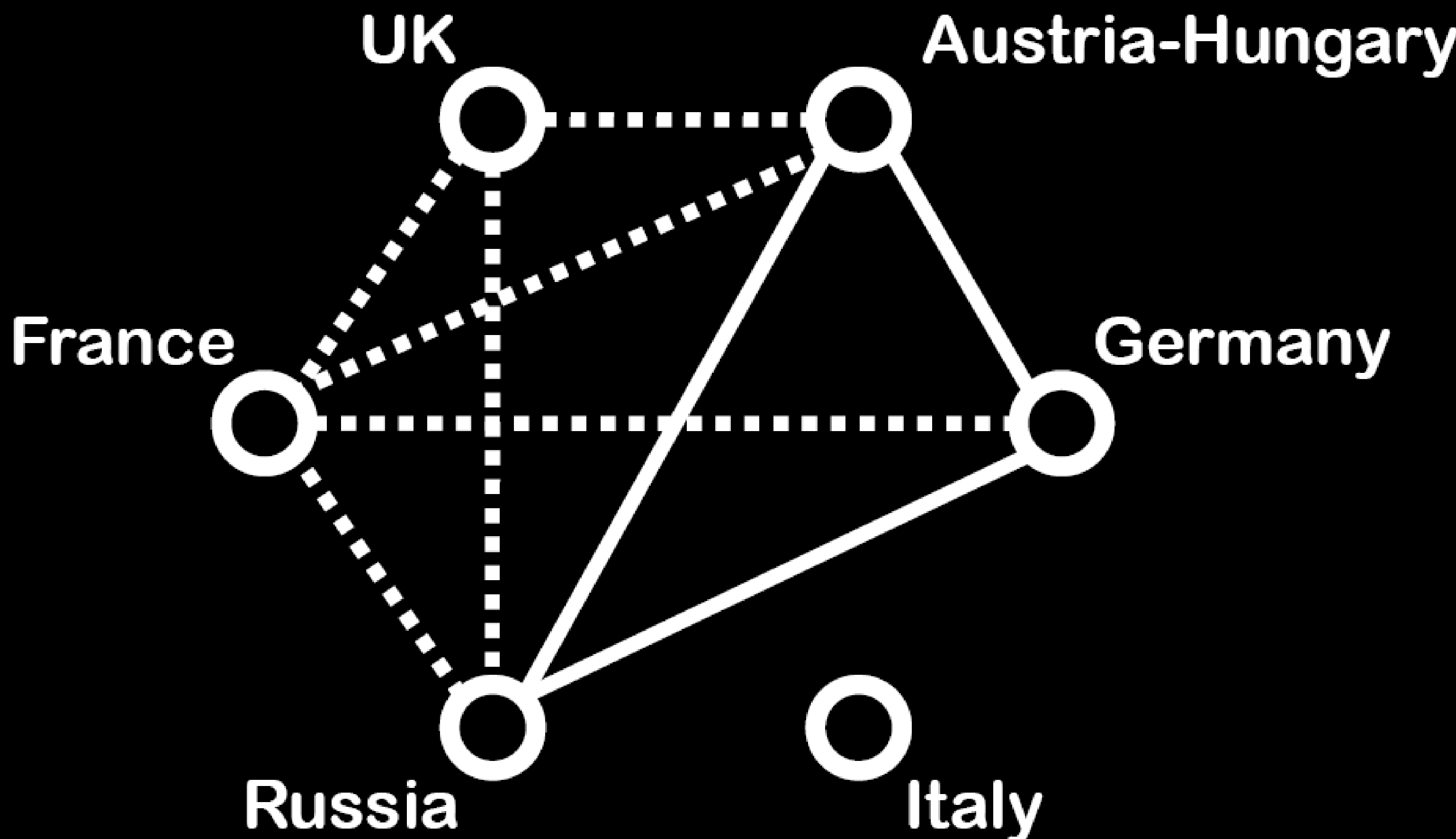


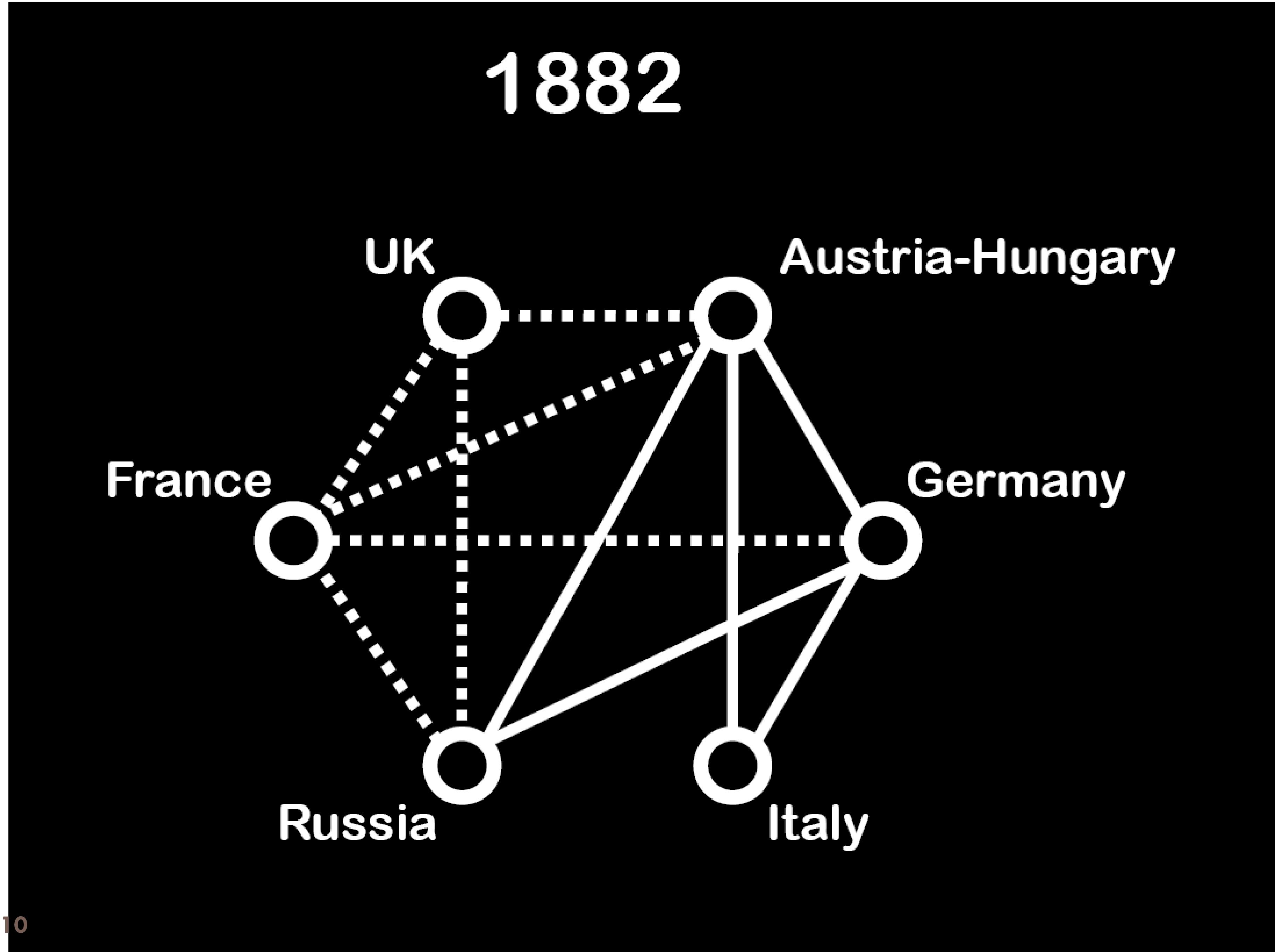
Example: International Relations

- **International relations:**
  - Positive edge: alliance
  - Negative edge: animosity
- Separation of Bangladesh from Pakistan in 1971: US supports Pakistan. **Why?**
  - 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.

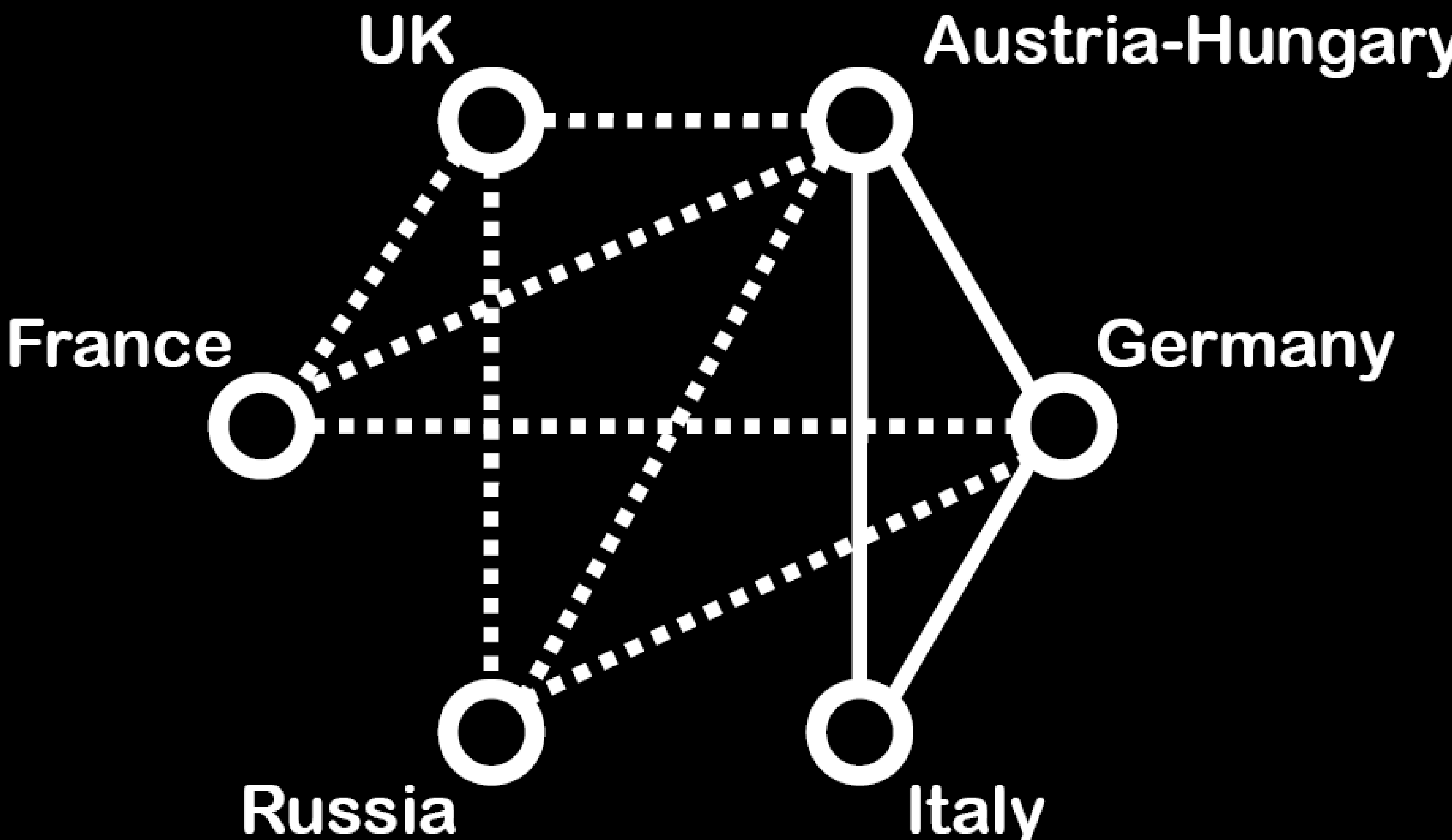


1872-1881

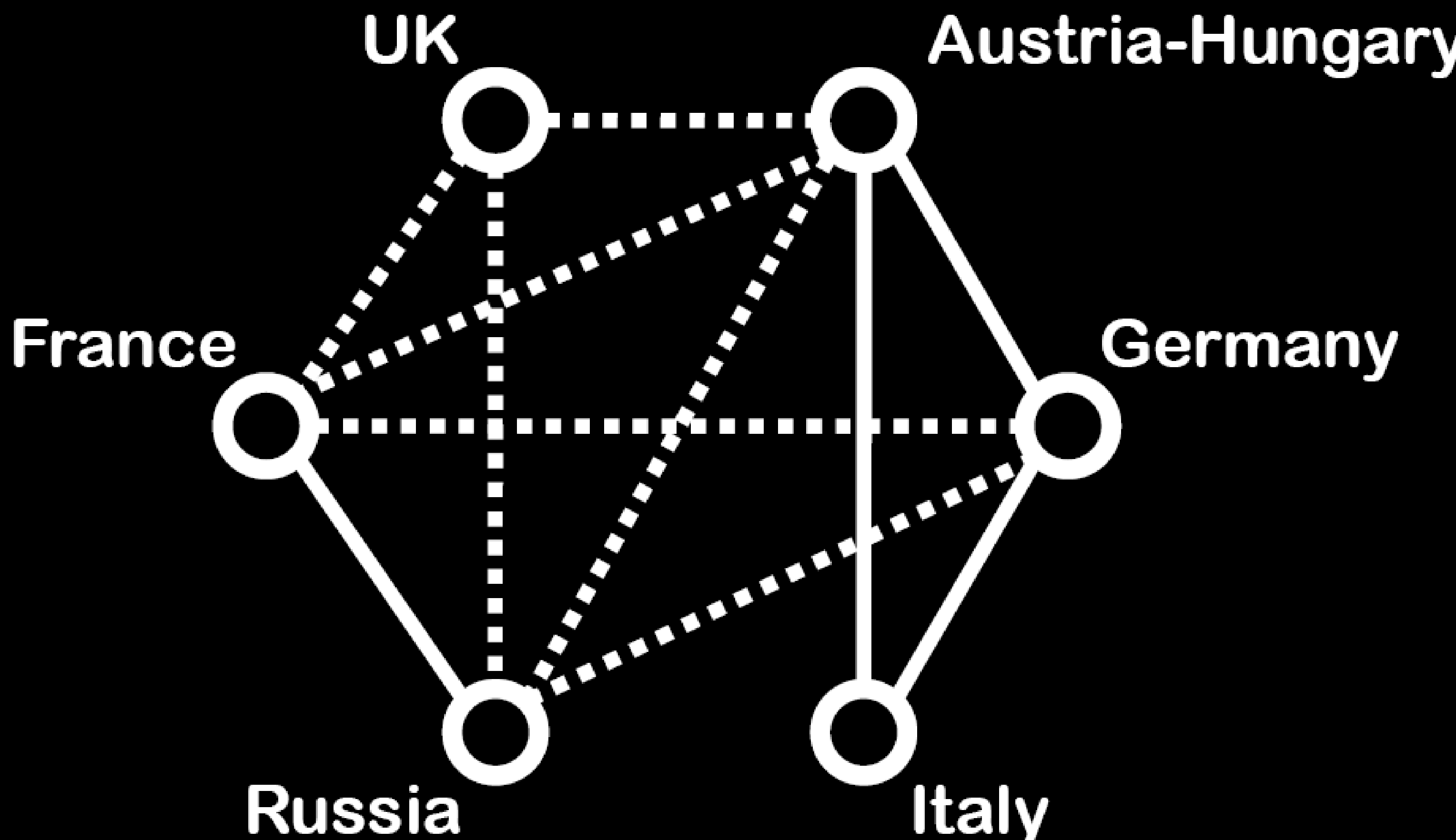



1882
UK
Austria-Hungary
France
Germany
Russia
Italy

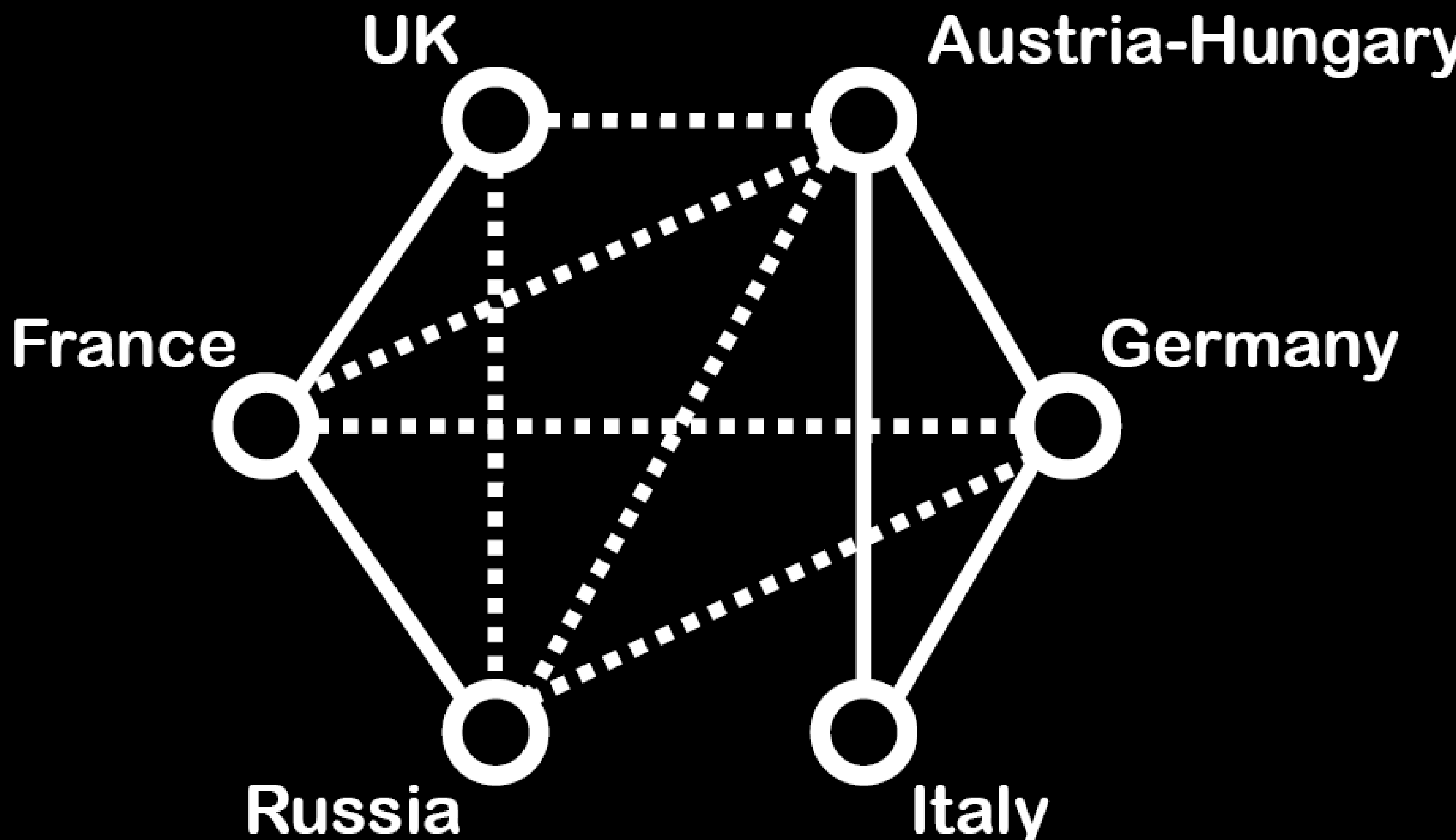
1890

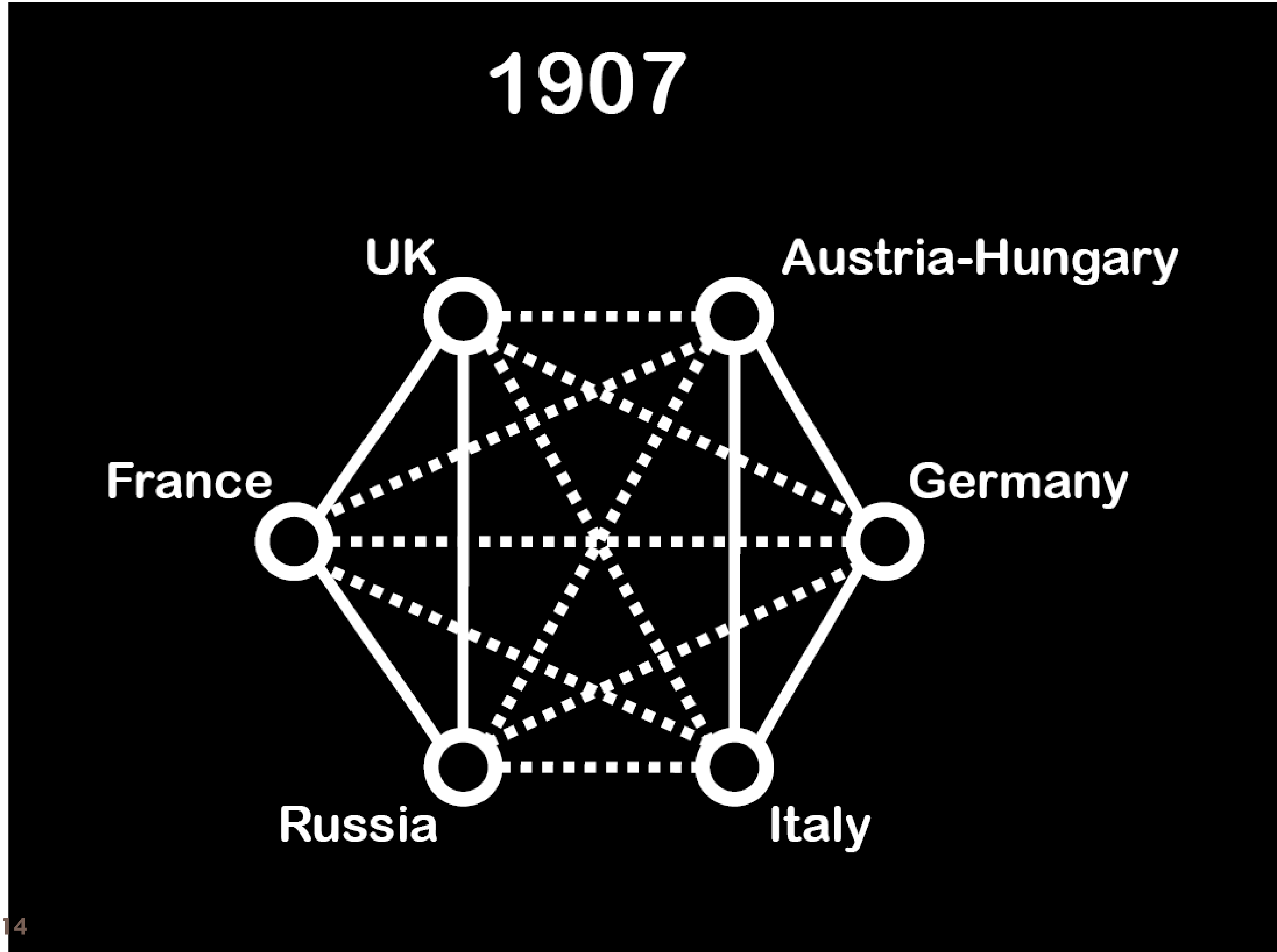


1891-1894



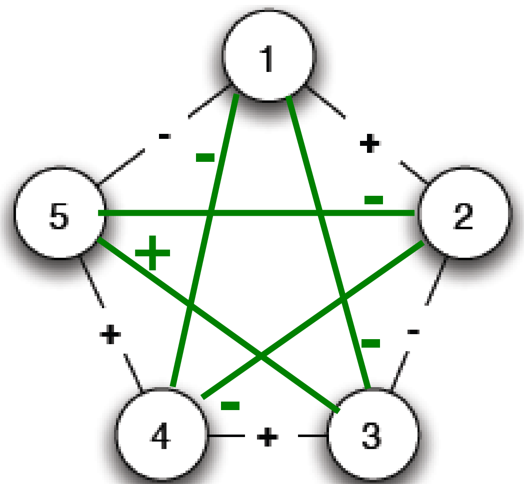
1904




1907
UK
Austria-Hungary
France
Germany
Russia
Italy

Balance in General Networks

- So far we talked about complete graphs



Balanced?

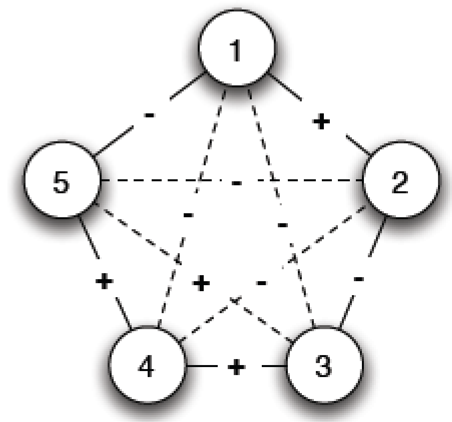
Def 1: Local view

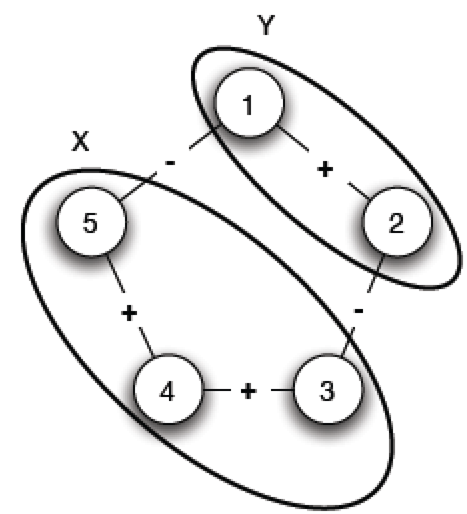
Fill in the missing edges to achieve balance

Def 2: Global view

Divide the graph into two coalitions

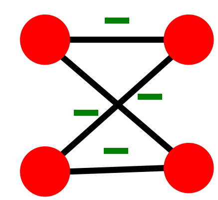
The 2 definitions are equivalent!



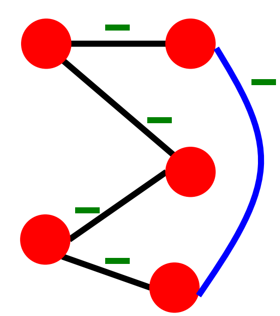


Is a Signed Network Balanced?

- 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 ⇒ **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

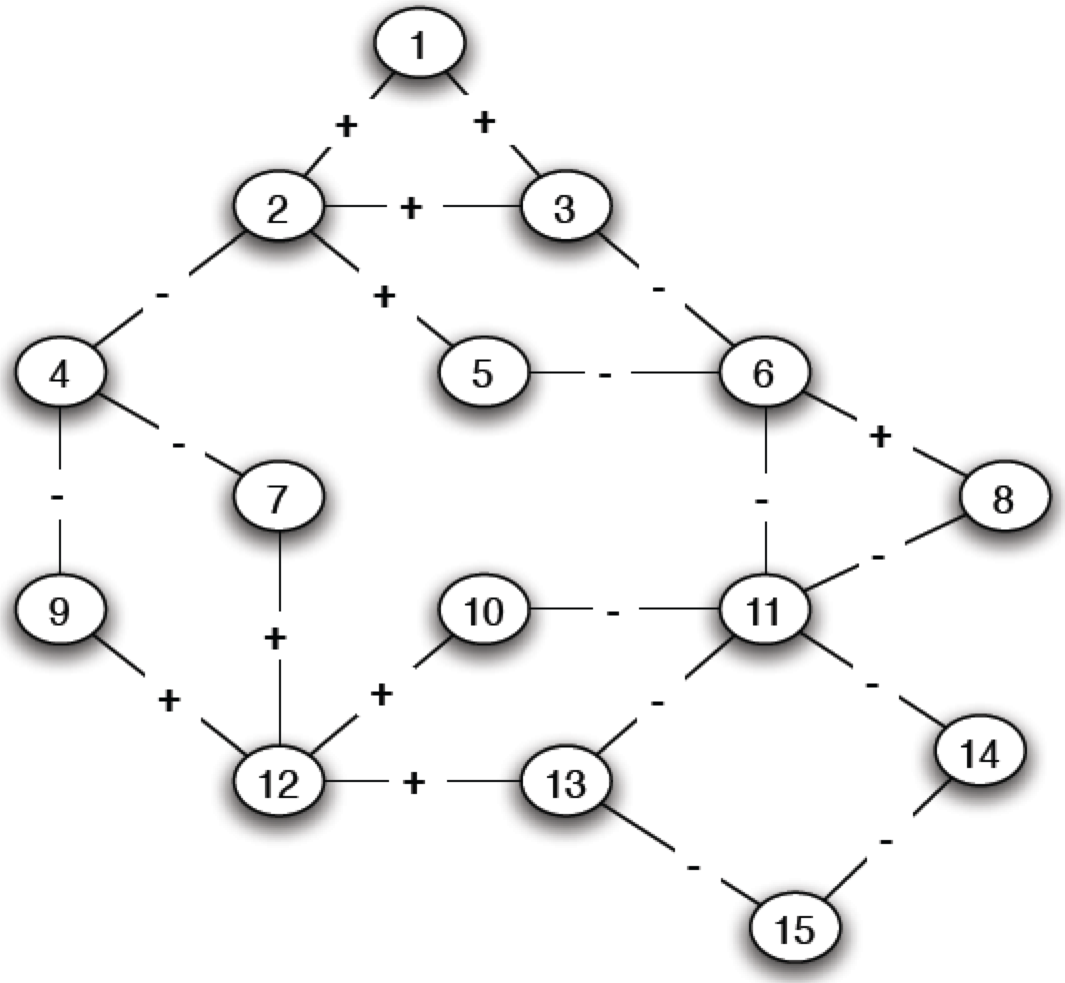


Even length cycle

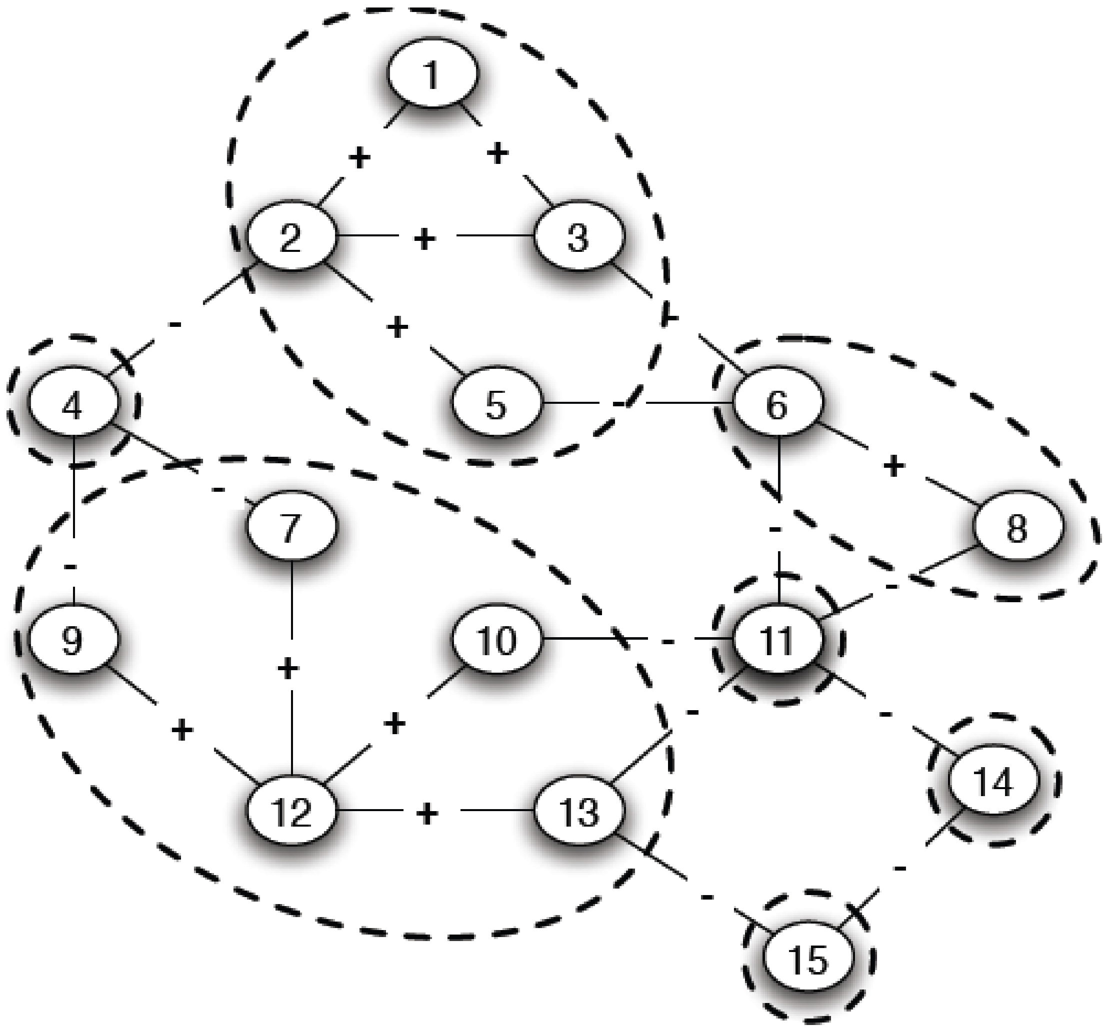


Odd length cycle

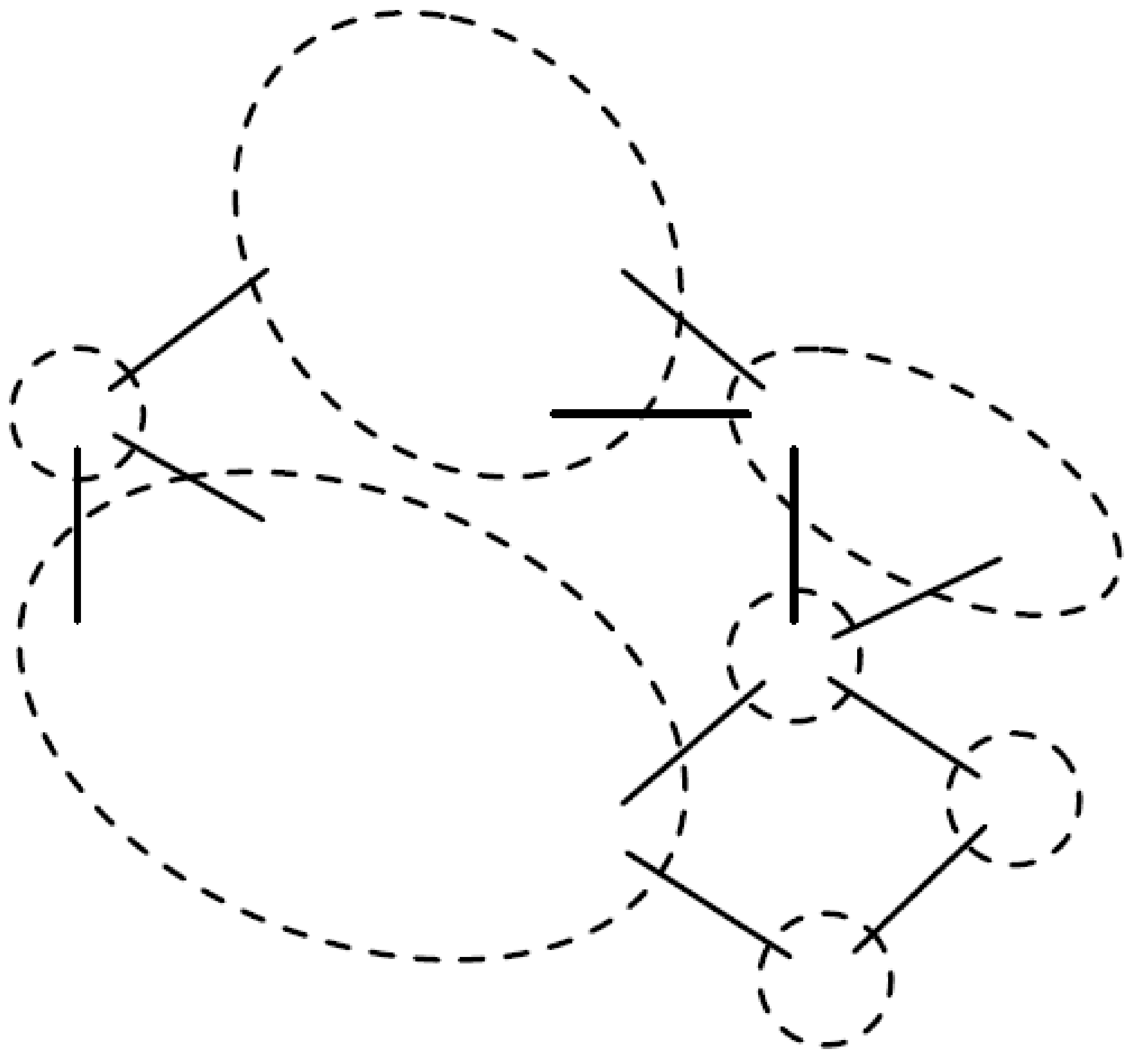
Signed Graph: Is it Balanced?



Positive Connected Components

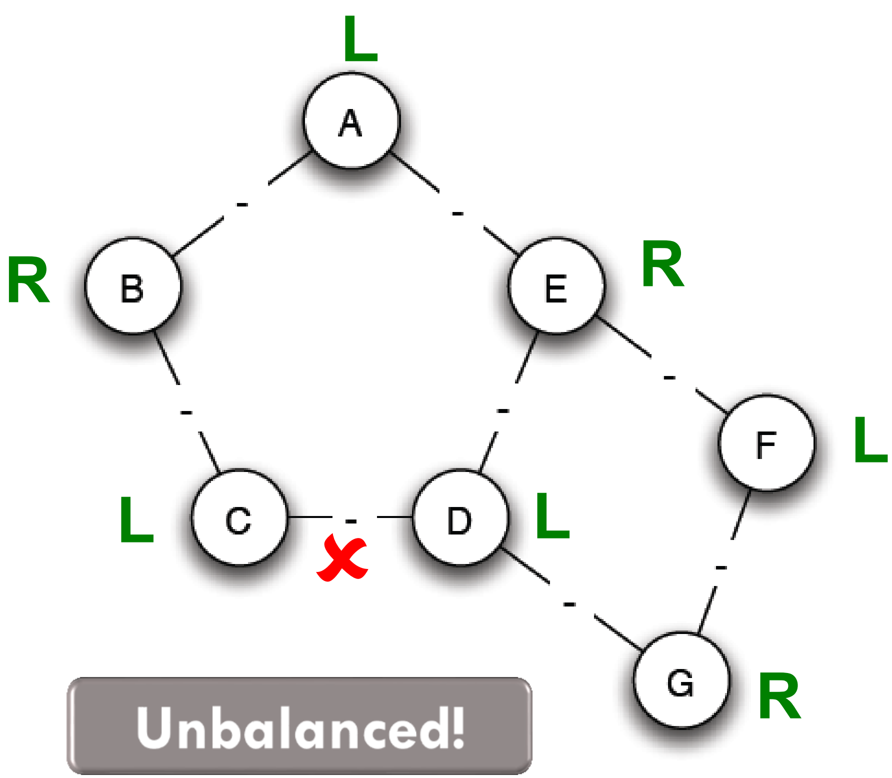


Reduced Graph on Super-Nodes



BFS on Reduced Graph

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

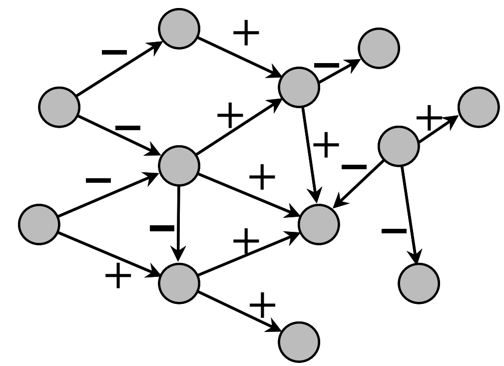


EXPLORING REAL DATA

19-Mar-15

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%

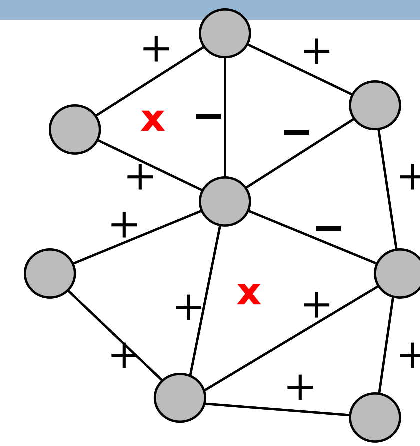
Balance in Our Network Data

- **Does structural balance hold?**
  - Compare frequencies of signed triads in real and "shuffled" data

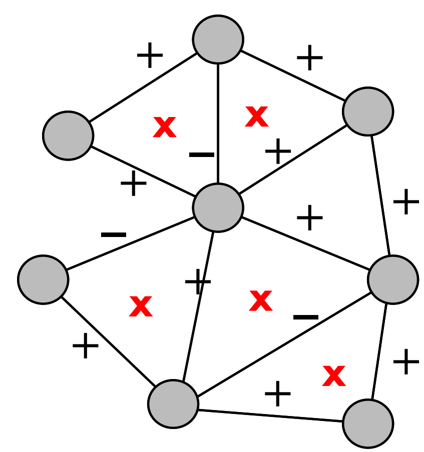
	Triad	Epinions		Wikipedia		Consistent with Balance?
		P(T)	P_0(T)	P(T)	P_0(T)	
Balanced	+ + +	0.87	0.62	0.70	0.49	✓
Balanced	- - +	0.07	0.05	0.21	0.10	✓
Unbalanced	+ + -	0.05	0.32	0.08	0.49	✓
Unbalanced	- - -	0.007	0.003	0.011	0.010	✗

P(T) … fraction of a triads

P_0(T)… triad fraction if the signs would be random



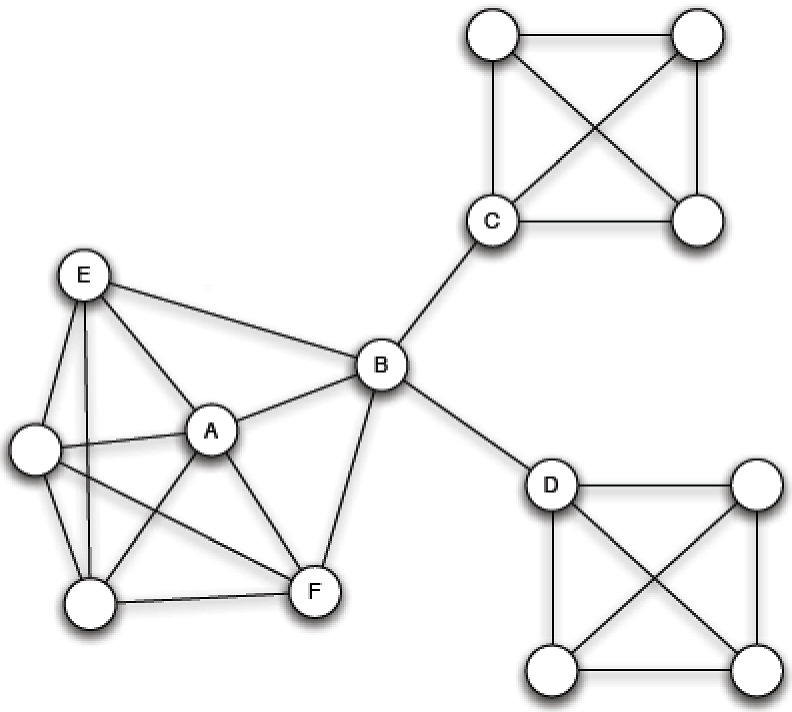
Real data



Shuffled data

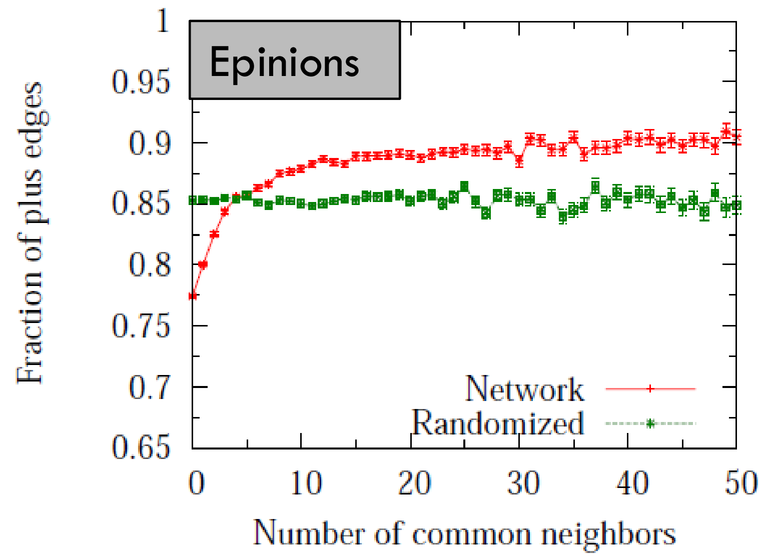
Global Structure of Signed Nets

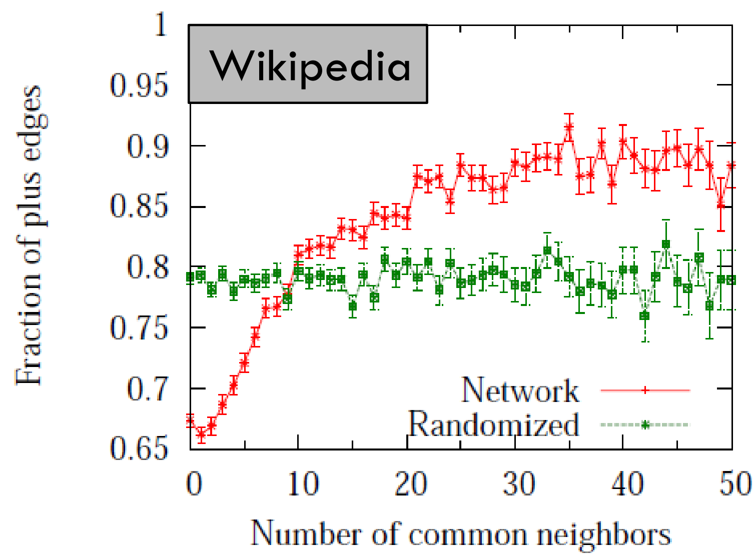
- Intuitive picture of social network in terms of densely linked clusters
- **How does structure interact with links?**
- **Embeddedness of link (A,B):** Number of shared neighbors



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

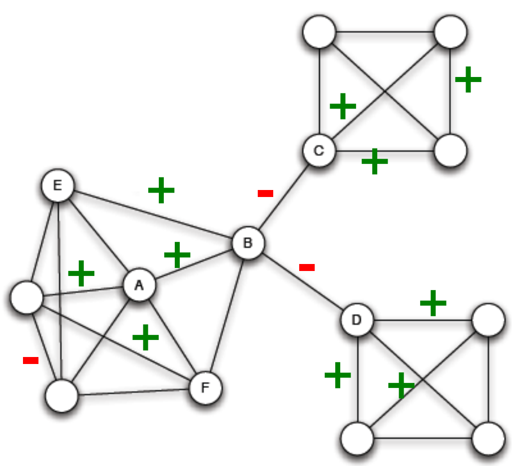




Global Structure of Signed Nets

	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

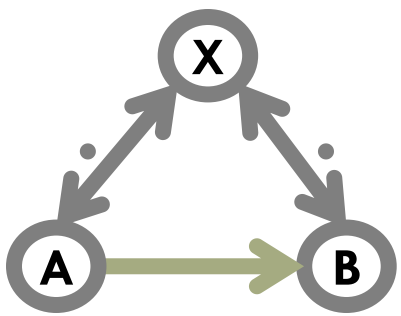
- **Clustering:**
  - +net: More clustering than baseline
  - –net: Less clustering than baseline
- **Size of max. component:**
  - +/–net: Smaller than the baseline



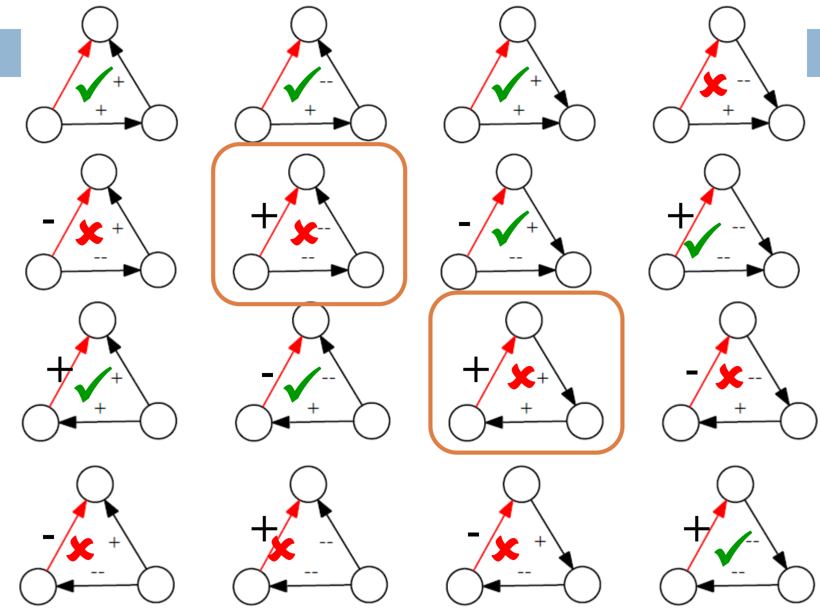
Evolving Directed Networks

- **New setting:**

  Links are **directed** and **created over time**



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



16 signed directed triads

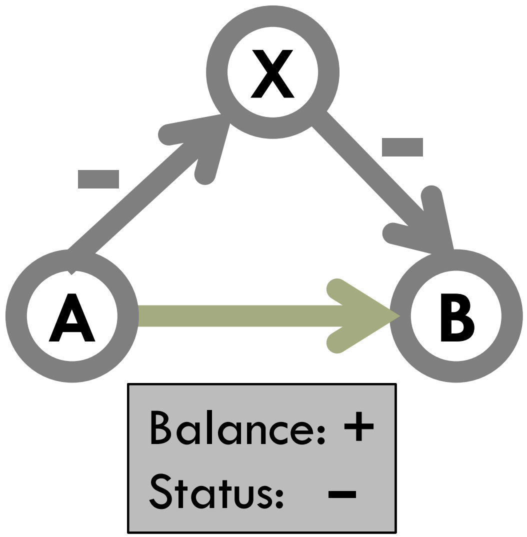
(in directed networks people traditionally applied balance by ignoring edge directions)

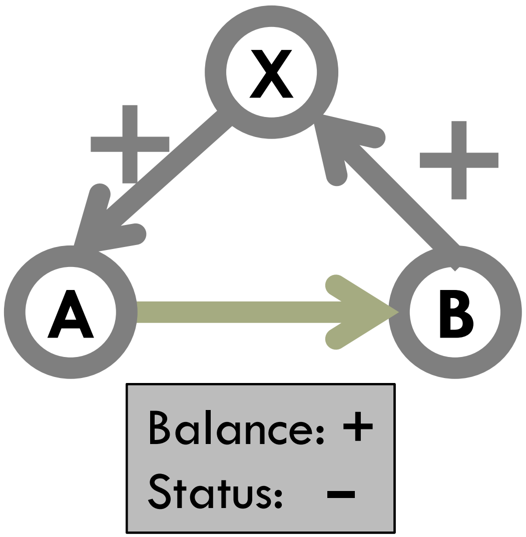
Alternate Theory: Status

- **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)

  - **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



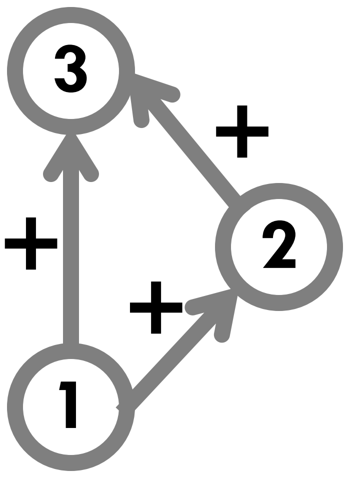


Status and balance give different predictions!

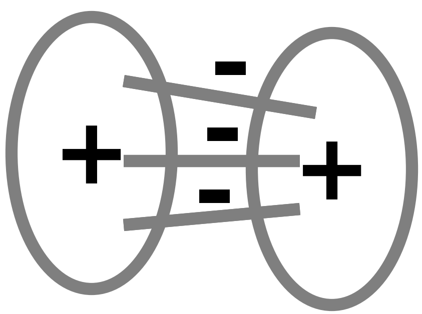
Status vs. Balance

At a global level:

- **Status ⇒ Hierarchy**
  - All-positive directed network should be (approximately) **acyclic**

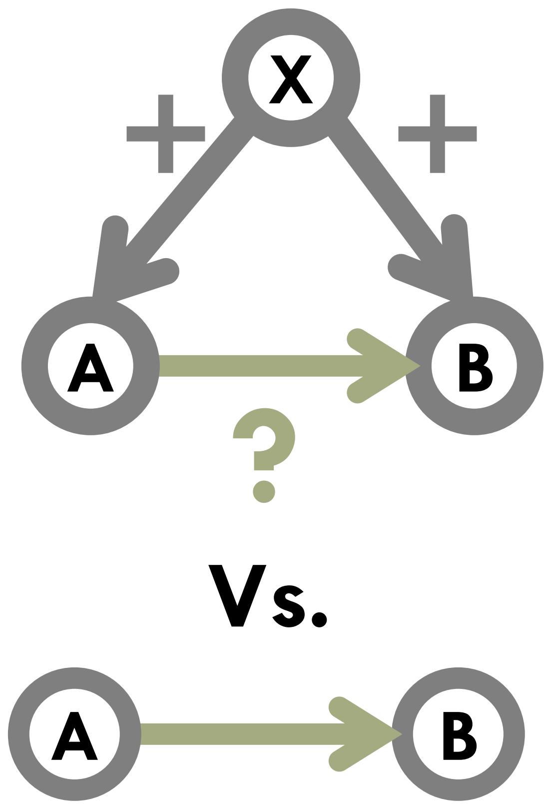


- **Balance ⇒ Coalitions**
  - Balance ignores directions and implies that subgraph of negative edges should be (approximately) **bipartite**



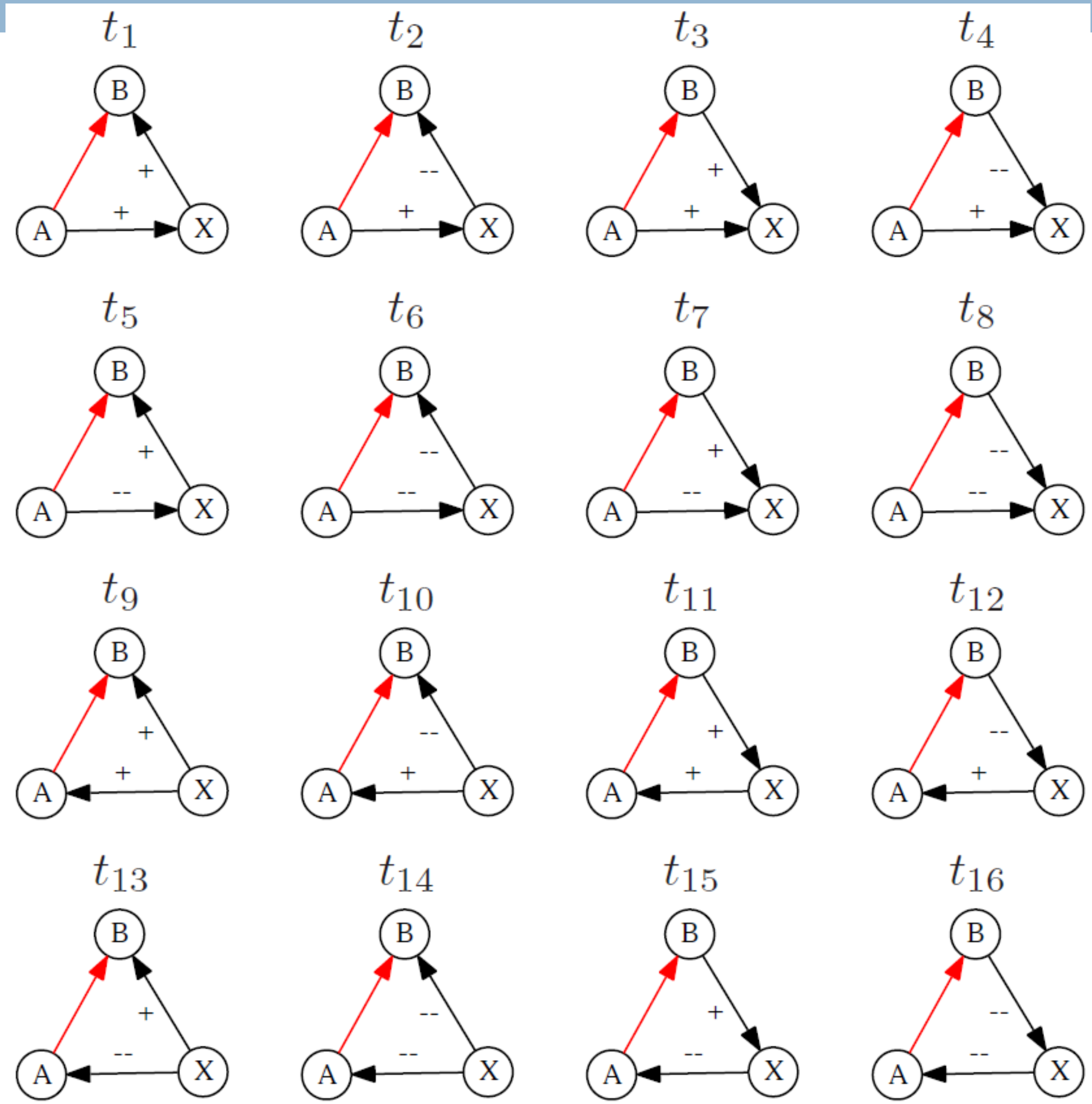
Theory of Status

- Edges are **directed and 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 from/to X?**
    $P(A \xrightarrow[+]{} B \mid X)$ vs. $P(A \xrightarrow[+]{} B)$
- **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→B appears in context X: A→B | X**
- **16 possible contexts:**



2) Heterogeneity in linking behavior

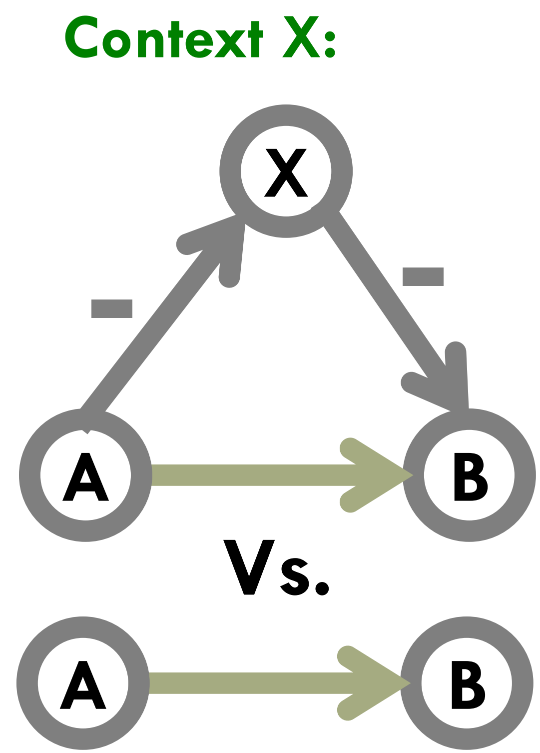
- **Users differ in frac. of + links they give/receive**
- **For a user U:**
  - **Generative baseline:** Frac. of + given by U
  - **Receptive baseline:** Frac. of + received by U

Basic question:

- How do **different link contexts** cause users to **deviate from their baselines**?
  - Link contexts as modifiers on a person's predicted behavior
  - **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:** $(A_1, B_1 | X_1), \ldots, (A_n, B_n | X_n)$
    … 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: $A_i \xrightarrow[+]{} B_i$



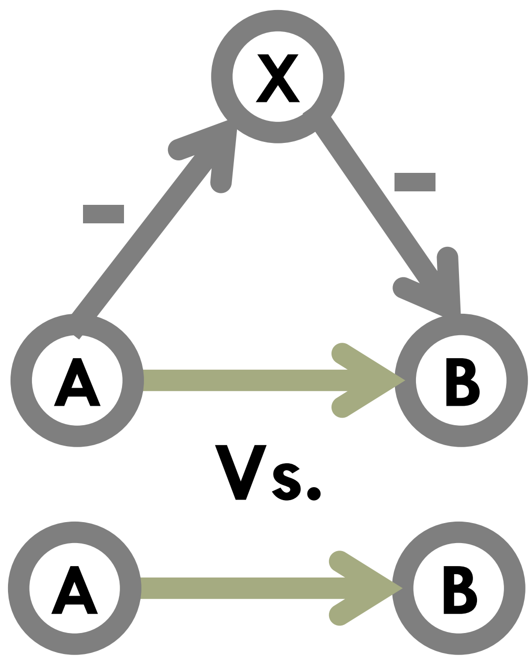
Computing Surprise

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

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

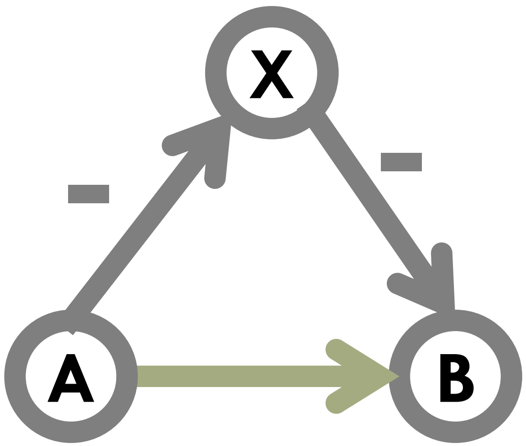
    - $p_g(A_i)$ … **generative baseline** of A_i
    - **Context X:** $(A_1, B_1 \mid X_1), \ldots, (A_n, B_n \mid X_n)$
    - *k* of instances of triad *X* closed with a plus edges
  - Receptive surprise is similar, just use $p_r(A_i)$

Context X:

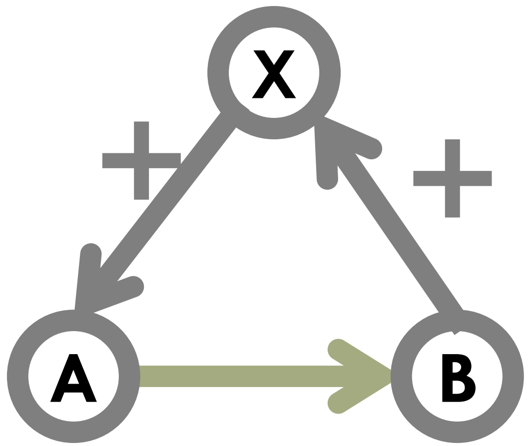


Status: Two Examples

- Assume status is at work
- What happens?



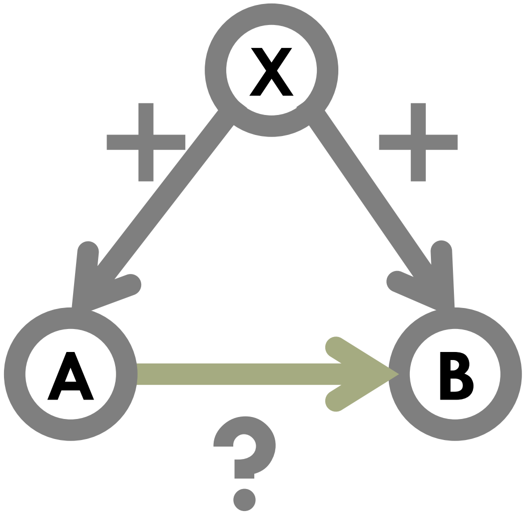
Gen. surprise of A: –
Rec. surprise of B: –



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

Joint Positive Endorsement

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

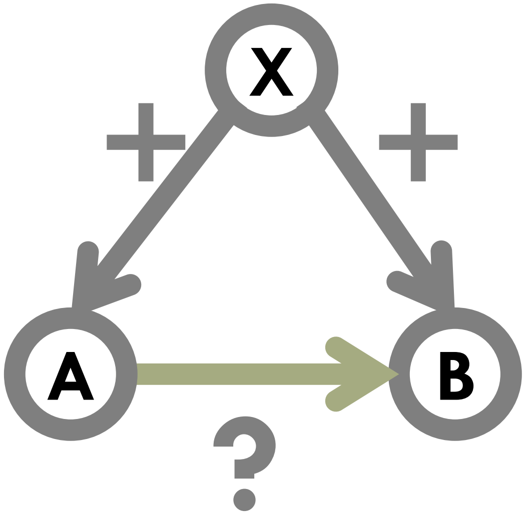


A puzzle:

- 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

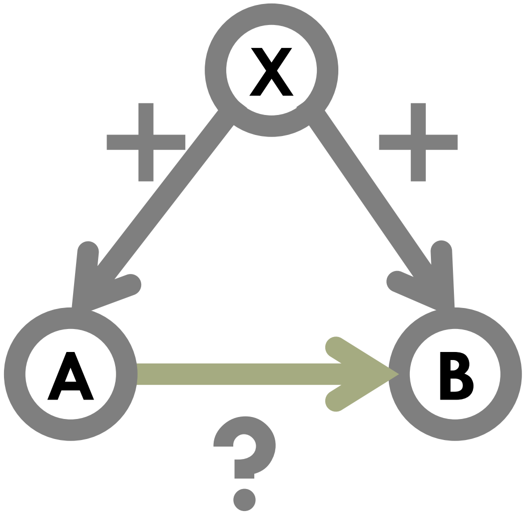


- We haven't asked A about B
- But we know that X thinks A and B are both better than him

- **What can we infer about A's answer?**

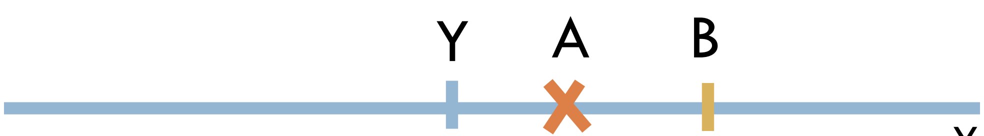
A Story: Soccer Team

- **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



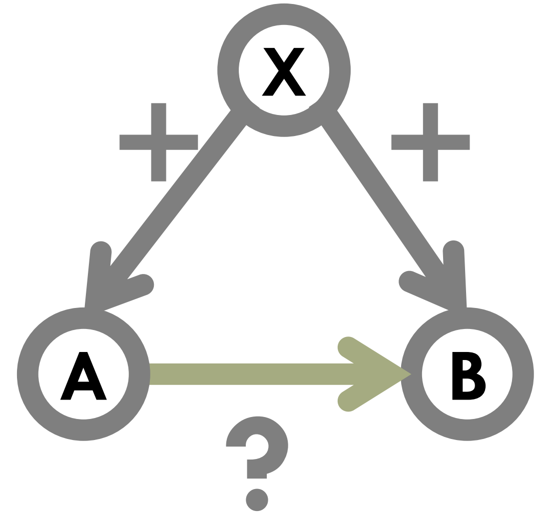
How does A evaluate B?

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



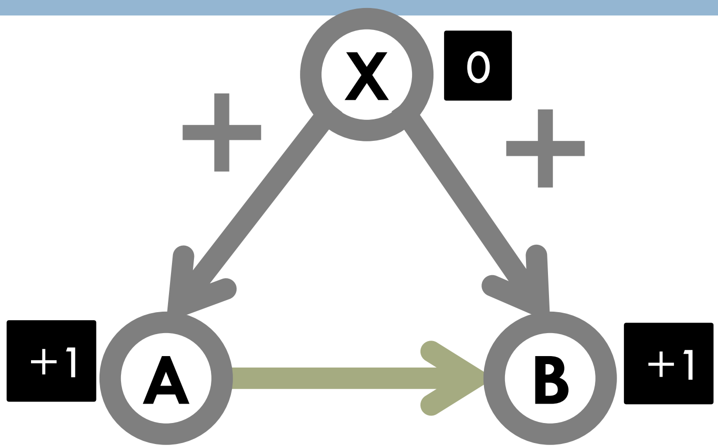
How is B evaluated by A?

B is evaluated by someone better than average.
→ They will be more negative to B than average

Sign of A→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



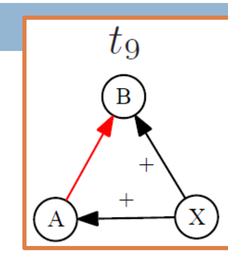
Status-consistent if:
Gen. surprise > 0
Rec. surprise < 0

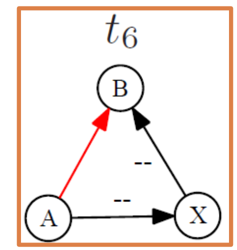
- Surprise is **status**-consistent, if:
  - Gen. surprise is status-consistent if it has **same** sign as status of B
  - Rec. surprise is status-consistent if it has the **opposite** sign from the status of A
- Surprise is **balance**-consistent, if:
  - If it completes a balanced triad

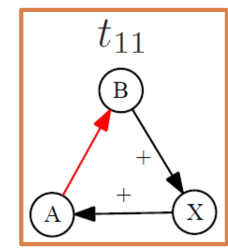
Status vs. Balance (Epinions)

- **Predictions:**

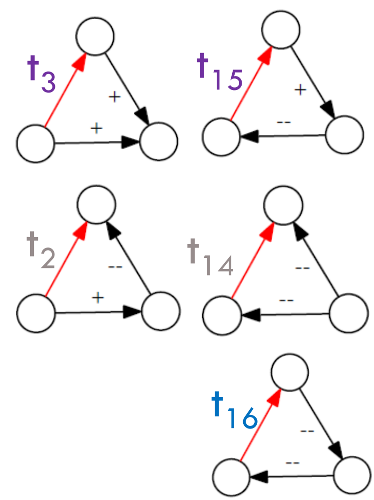
t_i	count	$P(+)$	$S_g(t_i)$	$S_r(t_i)$	B_g	B_r	S_g	S_r
t_1	178,051	0.97	95.9	197.8	✓	✓	✓	✓
t_2	45,797	0.54	-151.3	-229.9	✓	✓	✓	●
t_3	246,371	0.94	89.9	195.9	✓	✓	●	✓
t_4	25,384	0.89	1.8	44.9	○	○	✓	✓
t_5	45,925	0.30	18.1	-333.7	○	✓	✓	✓
t_6	11,215	0.23	-15.5	-193.6	○	○	✓	✓
t_7	36,184	0.14	-53.1	-357.3	✓	✓	✓	✓
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







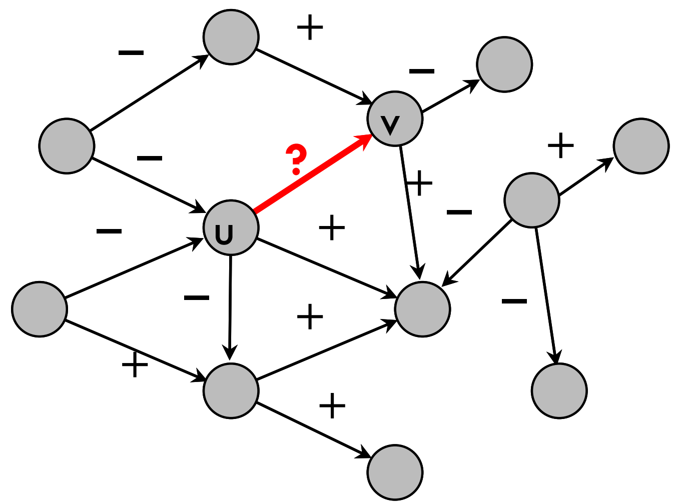
Mistakes:



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

Summary

- **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