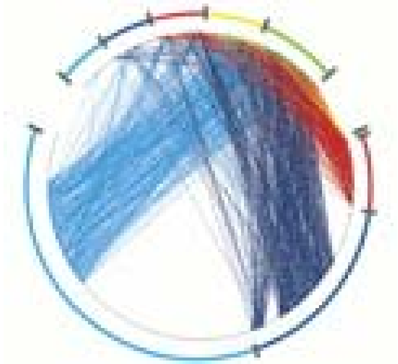
Lecture 17: Signed Graphs & Social Balance



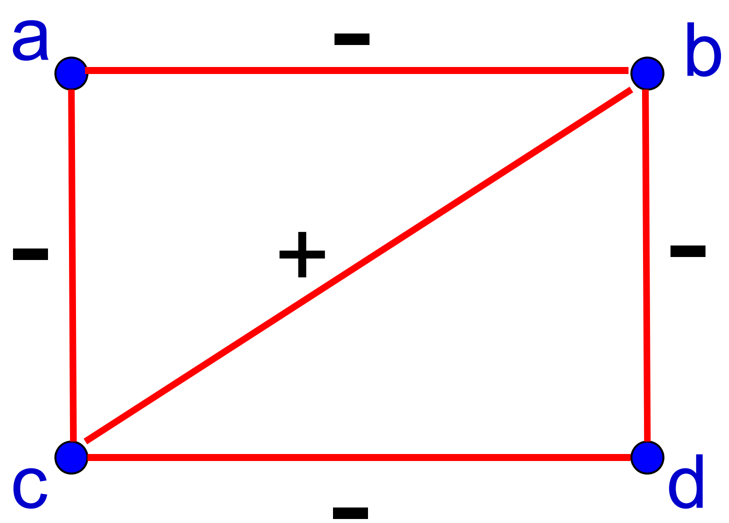
Positive or negative links

- Signe graphs
- 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 and C

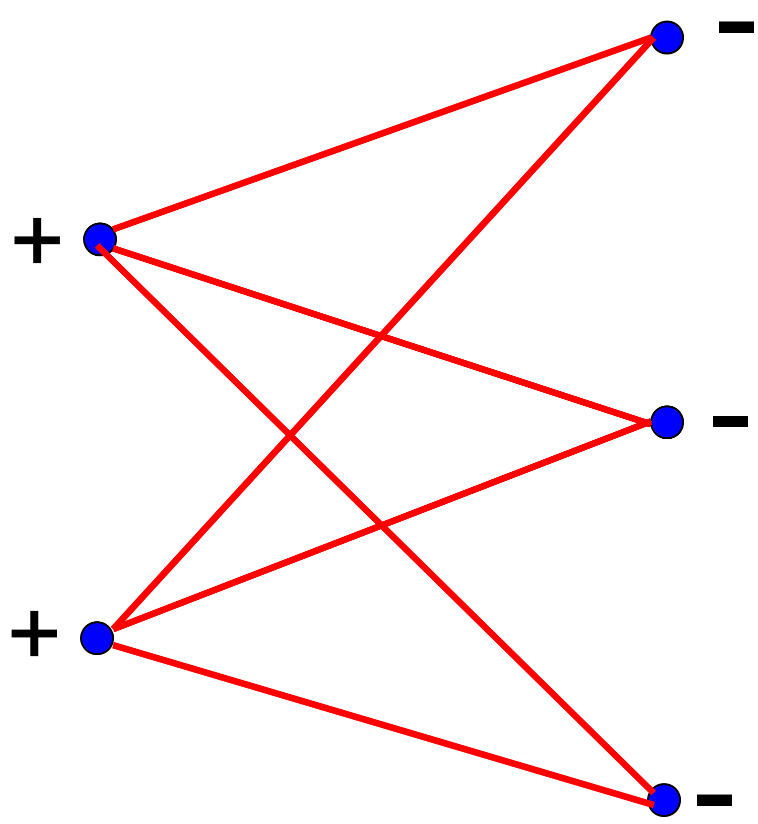
Signed and marked graphs

- Data in the social sciences can often be modeled using a **signed graph:** A graph where every edge has a sign + or –.
- Less widely used in the social sciences is a **marked graph**, where every vertex has a sign + or –.
- A signed graph is **balanced** if every cycle has an even number of – signs.
- A marked graph is **consistent** if every cycle has an even number of – signs.

Balanced signed graphs



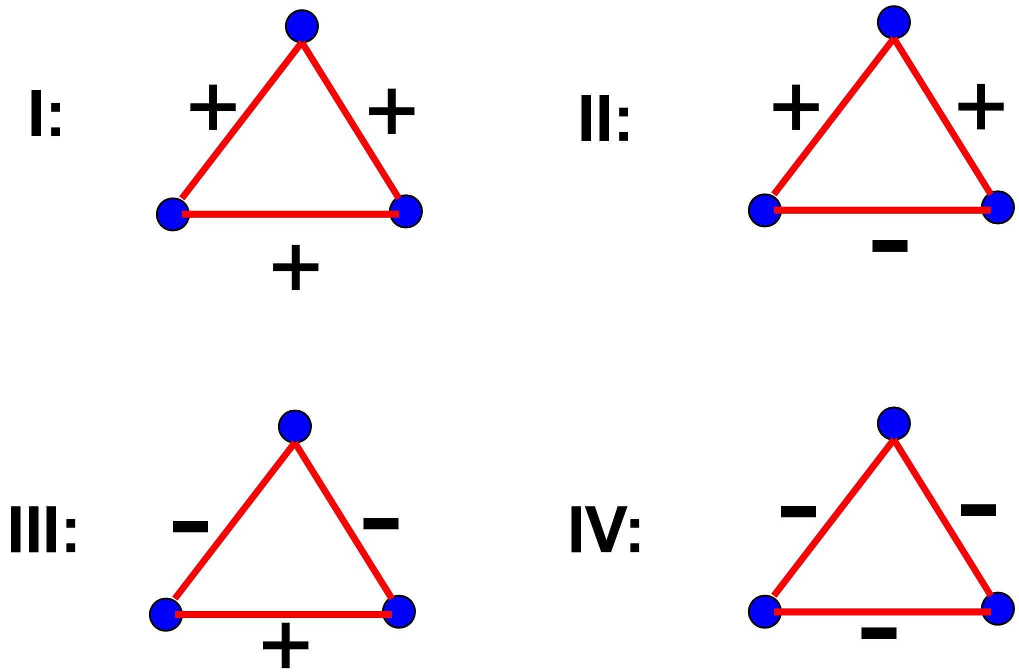
Balanced marked graphs



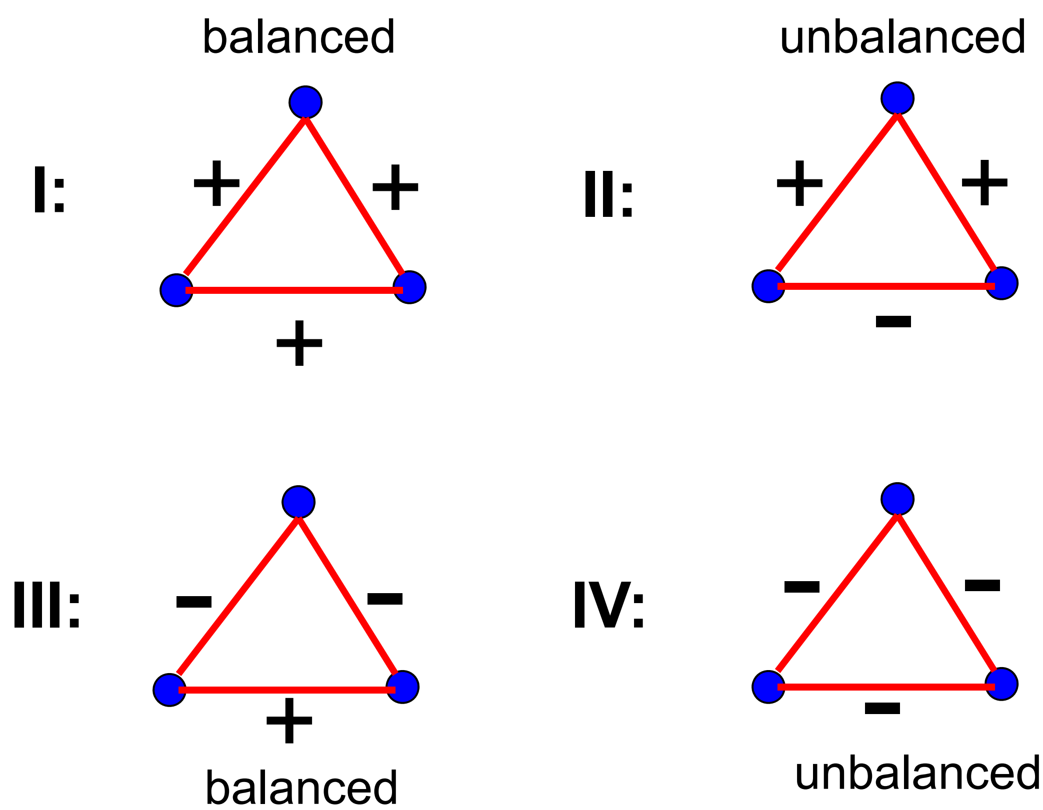
Signed and marked graphs

- We will speak of the **sign of a path or cycle** as being + if it has an even number of – signs, and – otherwise .
- So: a signed graph is balanced iff every cycle is +.
- A marked graph is consistent iff every cycle is +.
- Small group is “balanced” if it works well together, lacks tension.
- Balanced signed graphs introduced as model for balanced small groups by Cartwright and Harary in early 1950s.
- Evidence that small group is balanced iff its corresponding signed graph is balanced.

Heider's experiments



Heider’s experiments



Justification

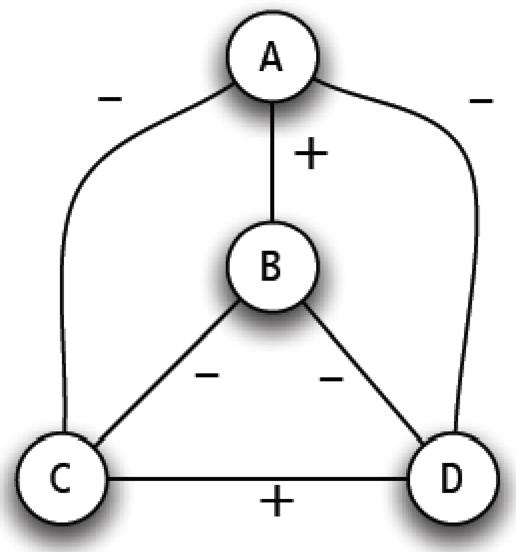
- Each pair of people are either friends or enemies (+ or -)
- Given a set of people A, B, and C, having three pluses among them is a very natural situation
  - It corresponds to three people who are mutual friends.
  - This situation is structurally balanced.
- Having a single plus and two minuses in the relations among the there people is also very natural
  - It means that two of the three are friends, and they have a mutual enemy.
  - This situation is structurally balanced.

Justification

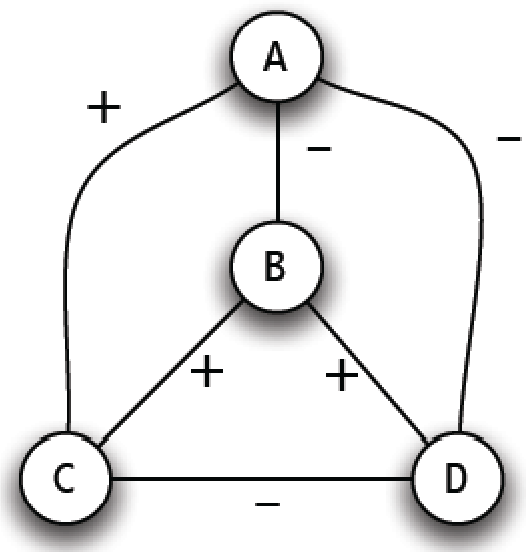
- A triangle with two pluses and one minus corresponds to a person A who is friends with each of B and C, but B and C don't get along with each other.
  - There would be implicit forces pushing A to try to get B and C to become friends (thus turning the B-C edge label to +); or else for A to side with one of B or C against the other (turning one of the edge labels out of A to a -).
  - This situation is structurally unbalanced.
- There are sources of instability in a situation where each of A, B, and C are mutual enemies
  - There would be forces motivating two of the three people to “team up" against the third (turning one of the three edge labels to a +).
  - This situation is structurally unbalanced.

Balanced/unbalance networks

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



Balanced



Unbalanced

Balanced/unbalance networks

- Theorem (Harary 1954): A signed graph G is balanced iff the set of vertices of G can be partitioned into two disjoint sets such that each + edge joins vertices in the same set and each – edge joins vertices in different sets.
- This can be made into a linear time algorithm to check for balance (Maybee and Maybee 1983, Hansen 1978).

Balanced/unbalance networks

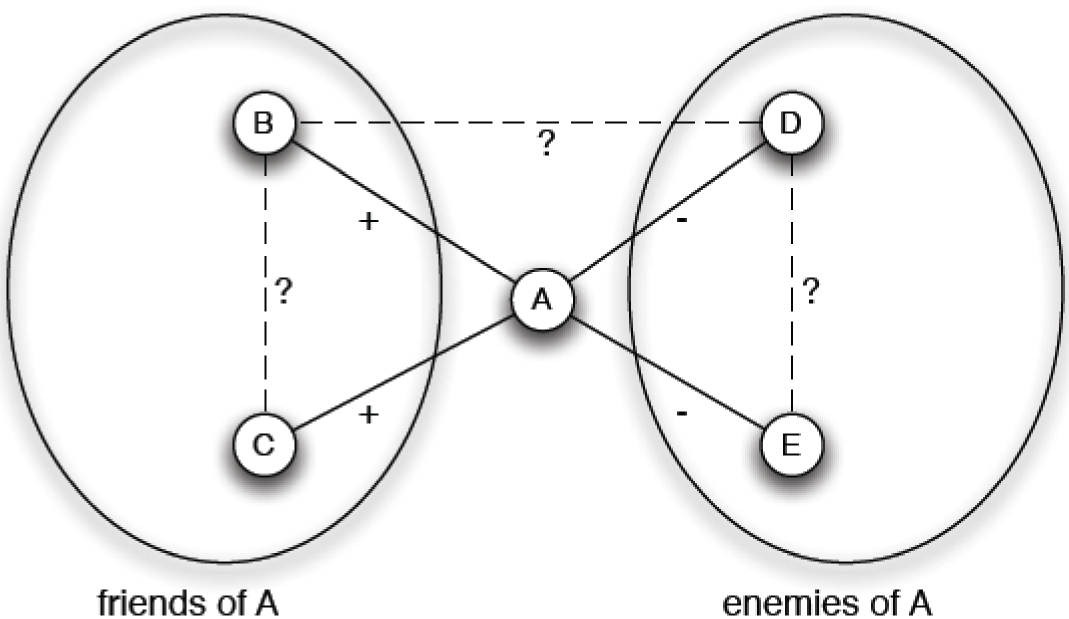
- Cartwright-Harary Theorem: If a labeled complete graph is balanced, then either all pairs of nodes are friends, or else the nodes can be divided into two groups, X and Y , such that every pair of people in X like each other, every pair of people in Y like each other, and everyone in X is the enemy of everyone in Y .

Proof

- Suppose we have a labeled complete graph that is balanced.
- If it has no negative edges at all, then everyone is friends, and the proof is complete.
- Otherwise, there is at least one negative edge.
- Let's divide the graph into X and Y , with complete antagonism between them.
- Let's pick any node in the network (A) and consider things from A's perspective:
  - Every other node is either a friend of A or an enemy of A.
  - Thus, natural candidates to try for the sets X and Y would be to define X to be A and all its friends, and define Y to be all the enemies of A.

Proof

- Recall what we need to show in order for these two sets X and Y to satisfy the conditions of the claim:
  i. Every two nodes in X are friends.
  ii. Every two nodes in Y are friends.
  iii. Every node in X is an enemy of every node in Y.



Proof

- i)
  - A is friends with every other node in X.
  - Let’s consider two other nodes in X (let's call them B and C).
  - We know that A is friends with both B and C, so if B and C were enemies of each other, then A, B, and C would form a triangle with two + labels that is a violation of the balance condition.
  - We know the network is balanced, this can't happen, so it must be that B and C in fact are friends.

Proof

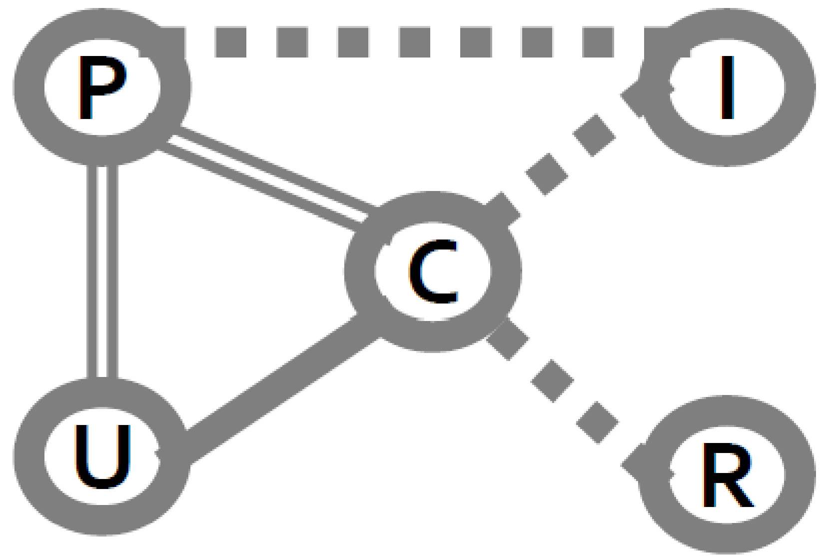
- ii)
  - Consider any two nodes in Y (let's call them D and E).
  - We know that A is enemies with both D and E, so if D and E were enemies of each other, then A, D, and E would form a triangle with no + labels that is a violation of the balance condition.
  - We know the network is balanced, this can‘t happen, so it must be that D and E in fact are friends.
  - Since D and E were the names of any two nodes in Y , we have concluded that every two nodes in Y are friends.

Proof

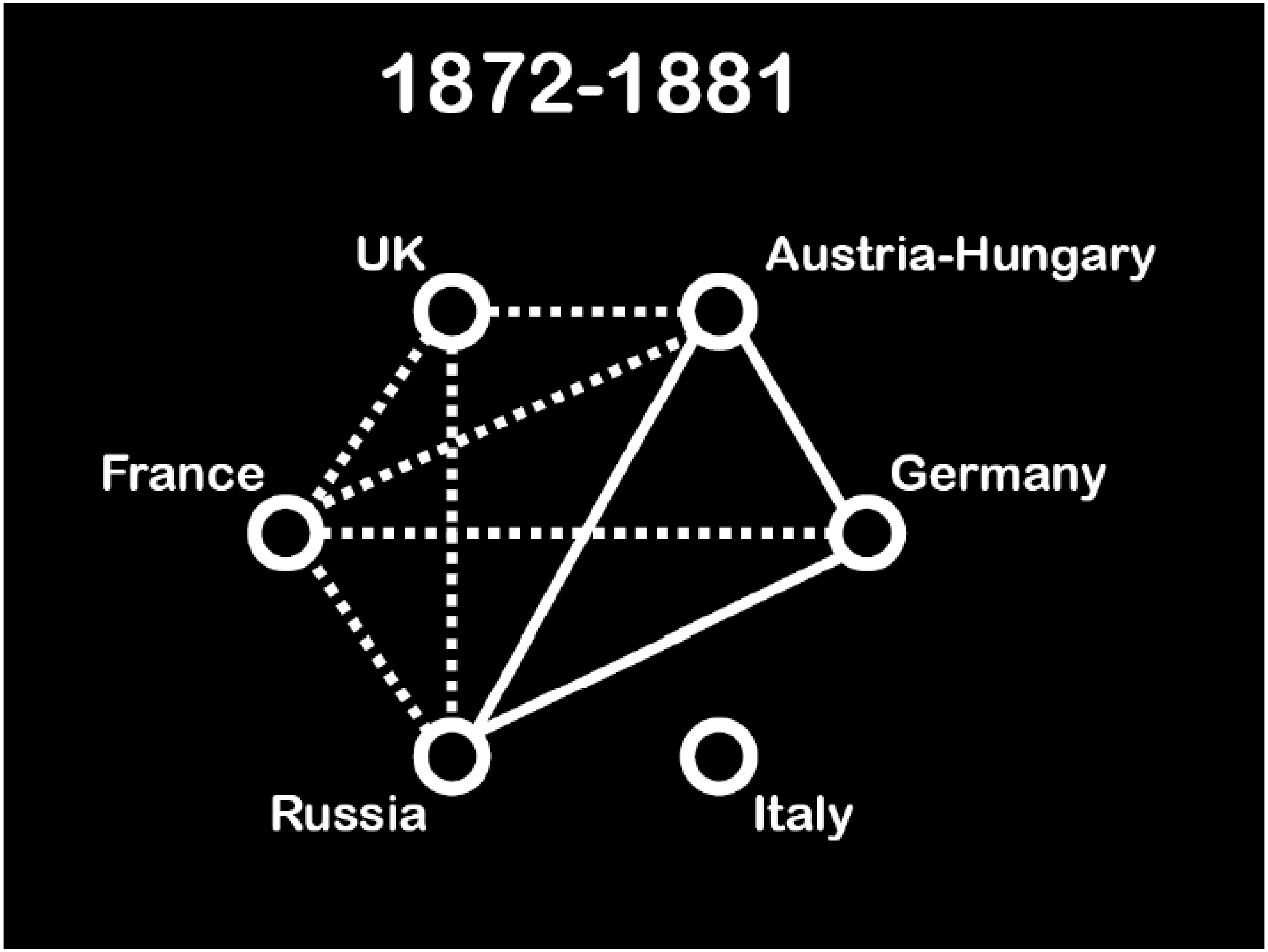
- iii)
  - Consider a node in X (call if B) and a node in Y (call it D).
  - We know A is friends with B and enemies with D, so if B and D were friends, then a, B, and D would form a triangle with two + labels that is a violation of the balance condition.
  - We know the network is balanced, this can't happen, so it must be that B and D in fact are enemies.
  - Since B and D were the names of any node in X and any node in Y , we have concluded that every such pair constitutes a pair of enemies.

Example: International relations

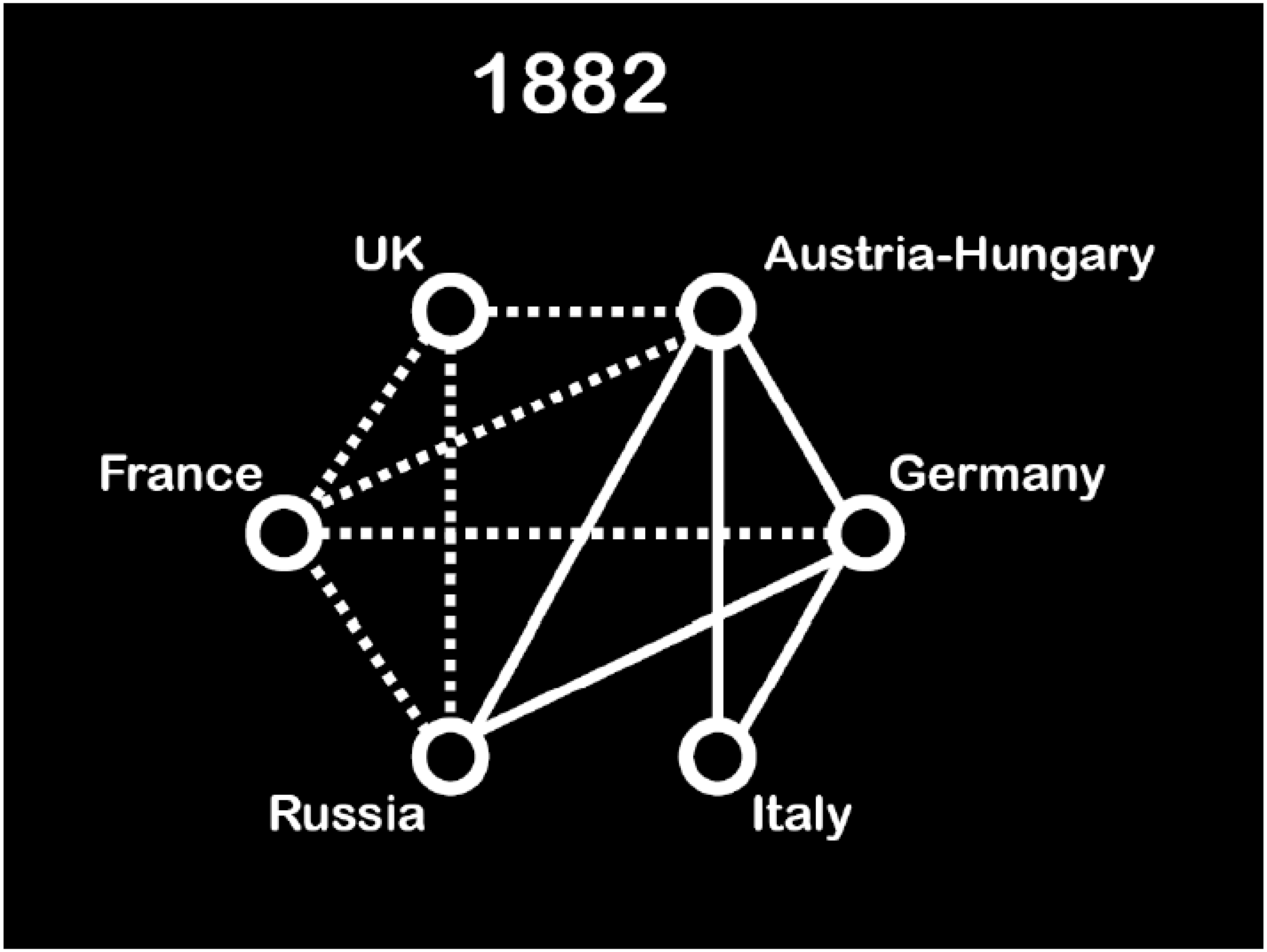
- **Positive edge: alliance**
- **Negative edge: animosity**
- Separation of Bangladesh from Pakistan in 1972: US supports Pakistan. Why?
  - USSR was enemy of China
  - China was enemy of India
  - India was enemy of Pakistan
  - US was friendly with China
  - China vetoed Bangladesh



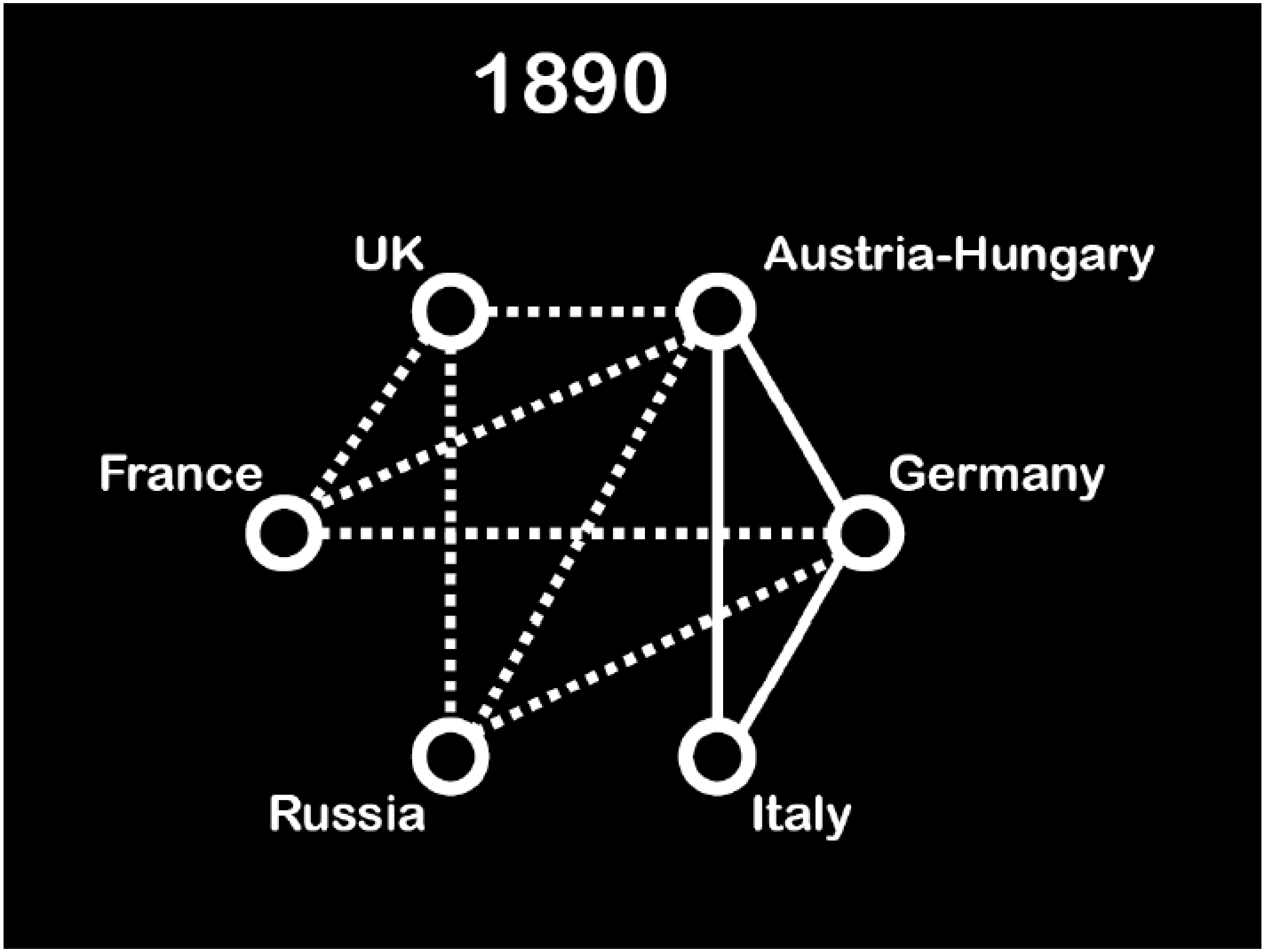
Another example



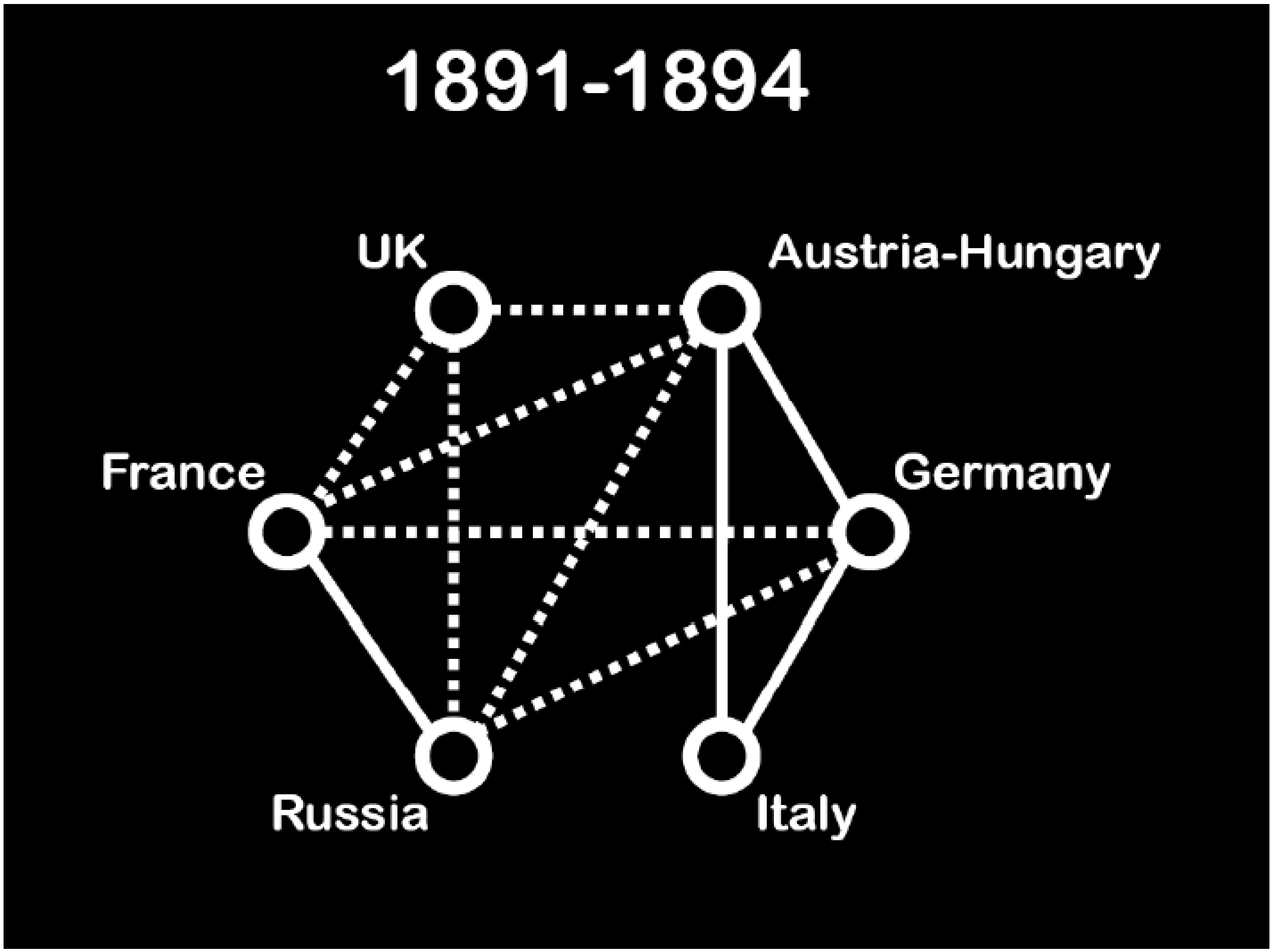
Another example



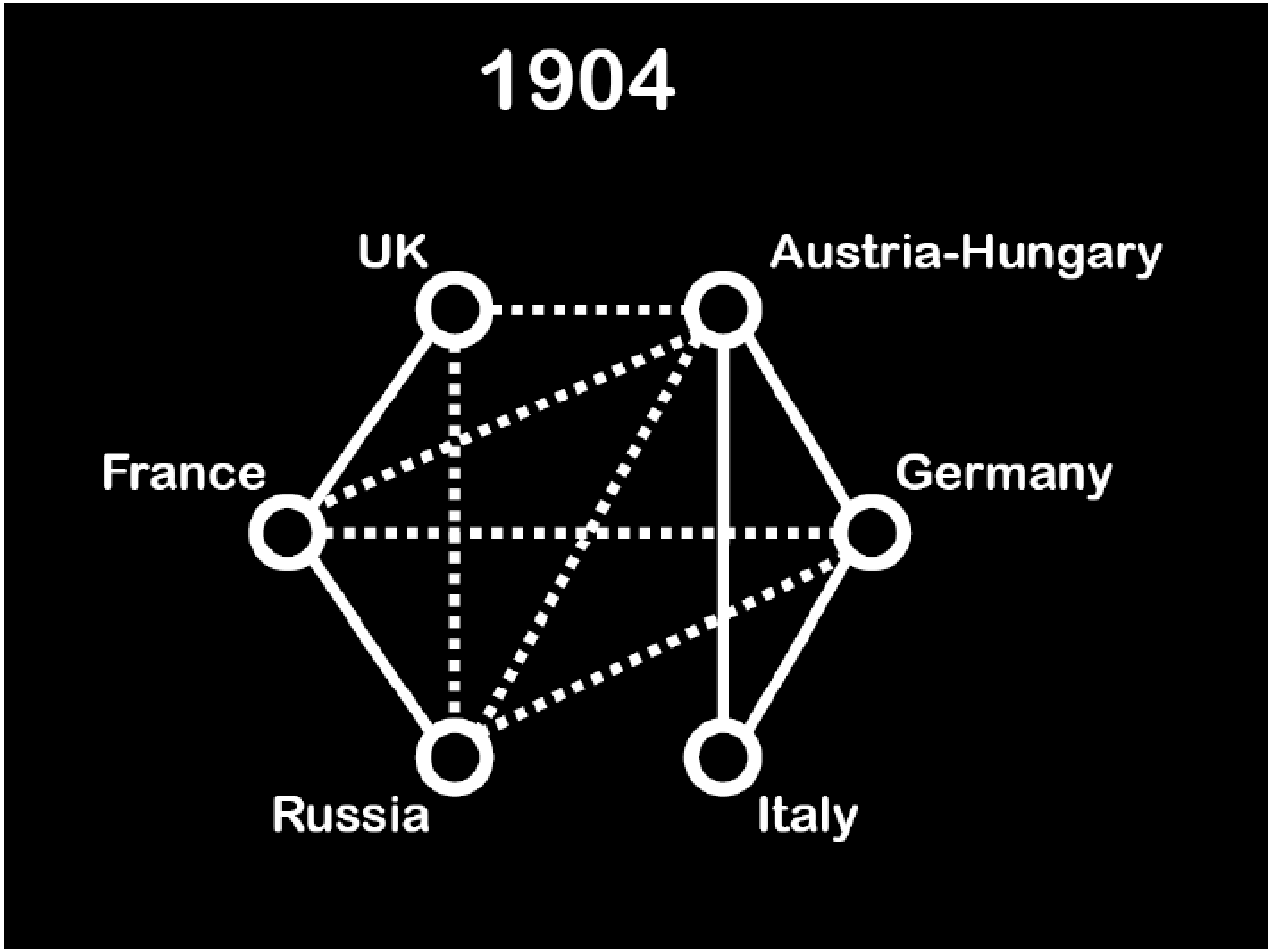
Another example



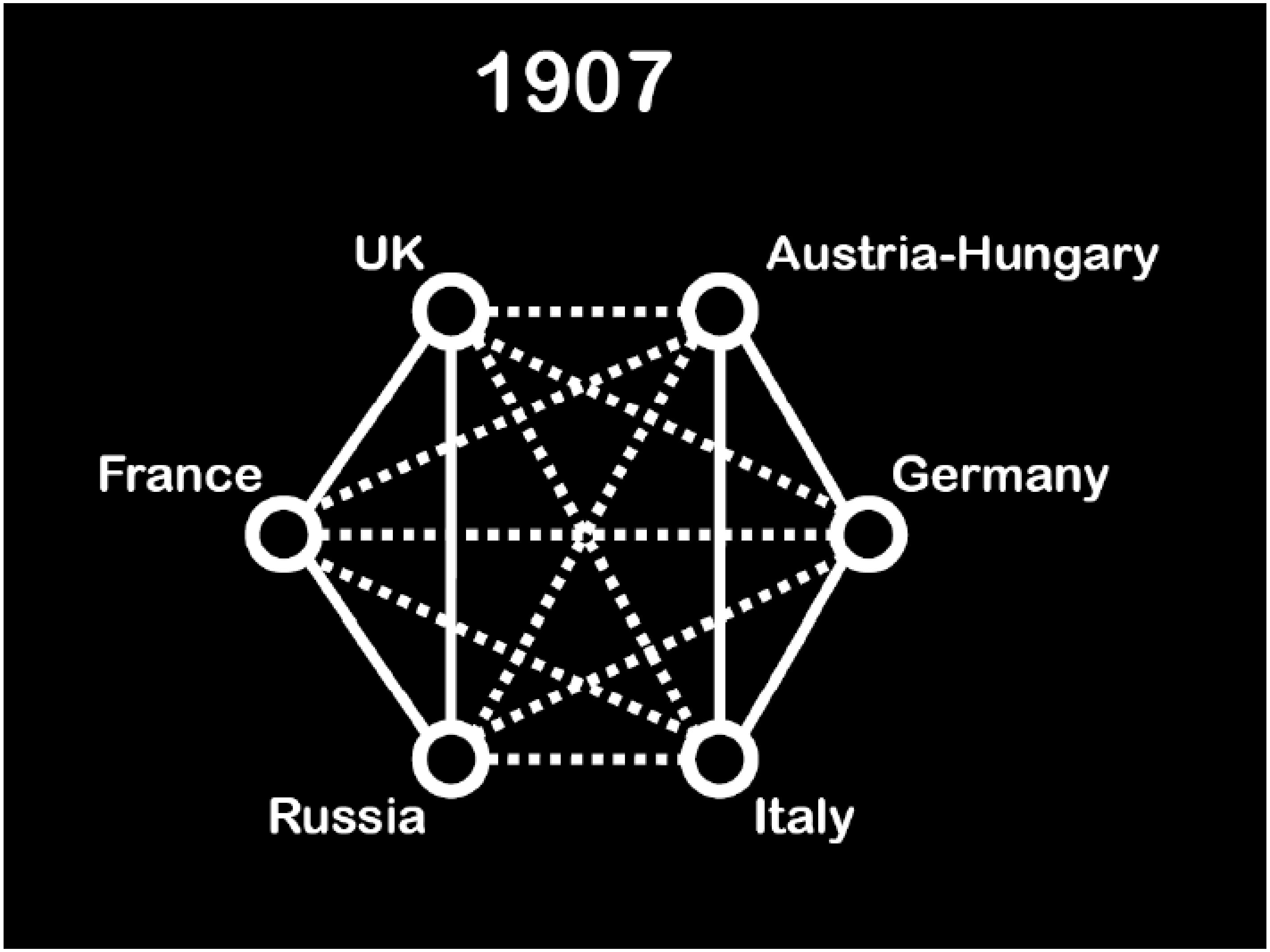
Another example



Another example



Another example



Another example

These results reinforces the fact that structural balance is not necessarily a good thing: since its global outcome is often two implacably opposed alliances, the search for balance in a system can sometimes be seen as a slide into a hard-to-resolve opposition between two sides.

Weakly balanced networks

- We will say that a complete graph, with each edge labeled by + or -, is weakly balanced if:
- There is no set of three nodes such that the edges among them consist of exactly two positive edges and one negative edge.
- Note that in this case, the triangular with all negative edges is allowed.

Weakly balanced networks

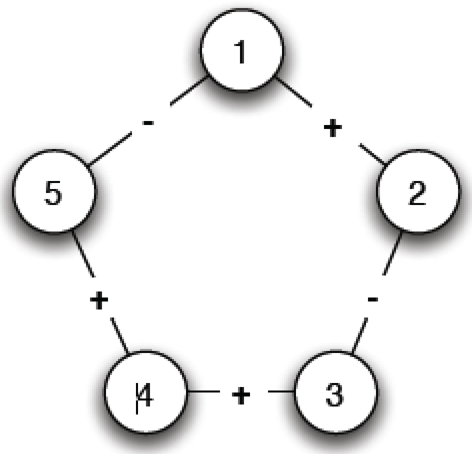
- <u>Theorem</u>: If a labeled complete graph is weakly balanced, then its nodes can be divided into groups in such a way that every two nodes belonging to the same group are friends, and every two nodes belonging to different groups are enemies.
- Proof is similar to the previous case.

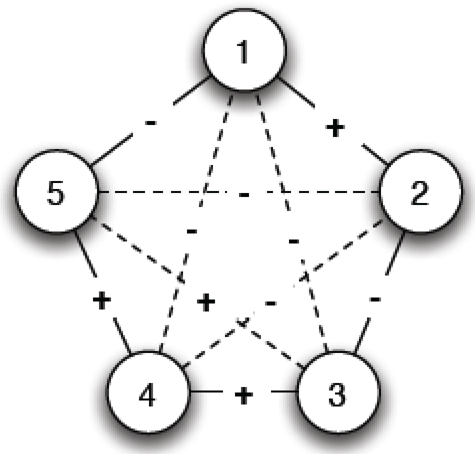
So,

- Our formulations so far is applied only to complete graphs.
  - We require that each person know and have an opinion (positive or negative) on everyone else.
  - What if only some pairs of people know each other?
- The elegant Cartwright-Harary Theorem showing that structural balance implies a global division of the world into two factions, only applies to the case in which every triangle is balanced.
  - Can we relax this to say that if most triangles are balanced, then the world can be approximately divided into two factions?

Structural balance in arbitrary networks

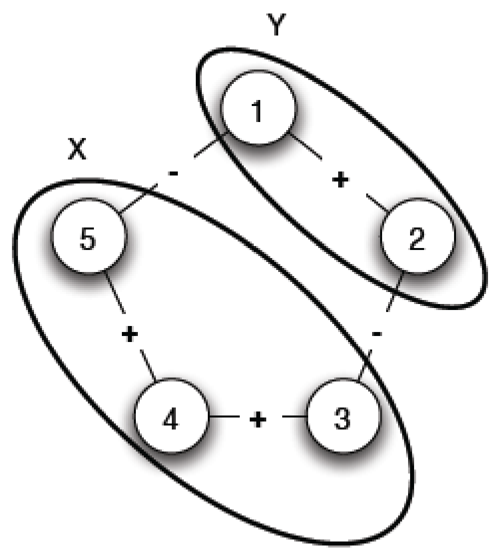
- Definition 1: Local view
  - A (non-complete) graph is balanced if it can be "completed" by adding edges to form a signed complete graph that is balanced.





Structural balance in arbitrary networks

- Definition 2: <u>Global view</u>
  - Viewing structural balance as implying a division of the network into two mutually opposed sets of friends

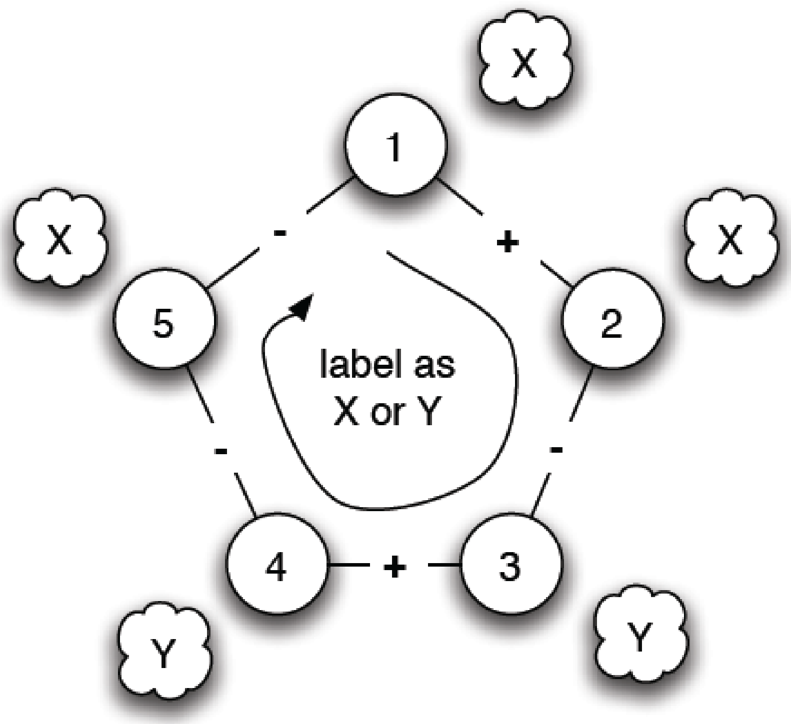


Structural balance in arbitrary networks

- These two ways of defining balance are equivalent
- An arbitrary signed graph is balanced under the first definition iff it is balanced under the second definition.
- If a signed graph is balanced under the first definition, then after filling in all the missing edges appropriately, we have a signed complete graph to which we can apply the Cartwright-Harary Theorem.
- This gives us a division of the network into two sets X and Y that satisfies the properties of the second definition.
- On the other hand, if a signed graph is balanced under the second definition, then after finding a division of the nodes into sets X and Y , we can fill in positive edges inside X and inside Y , and fill in negative edges between X and Y.

So, we can state that

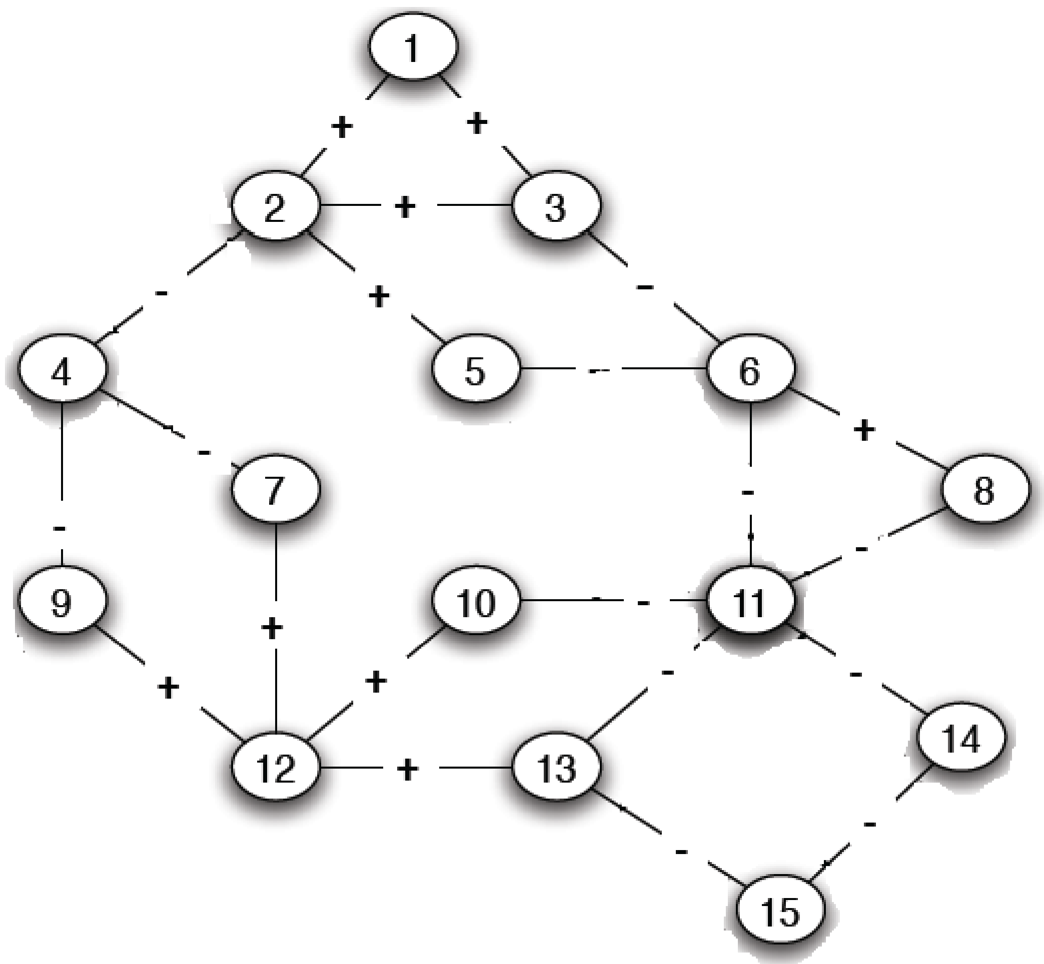
- A signed graph is balanced if and only if it contains no cycle with an odd number of negative edges.
  - If we pick one of the nodes and try to place it in X, then following the set of friend/enemy relations around the cycle will produce a conflict by the time we get to the starting node.



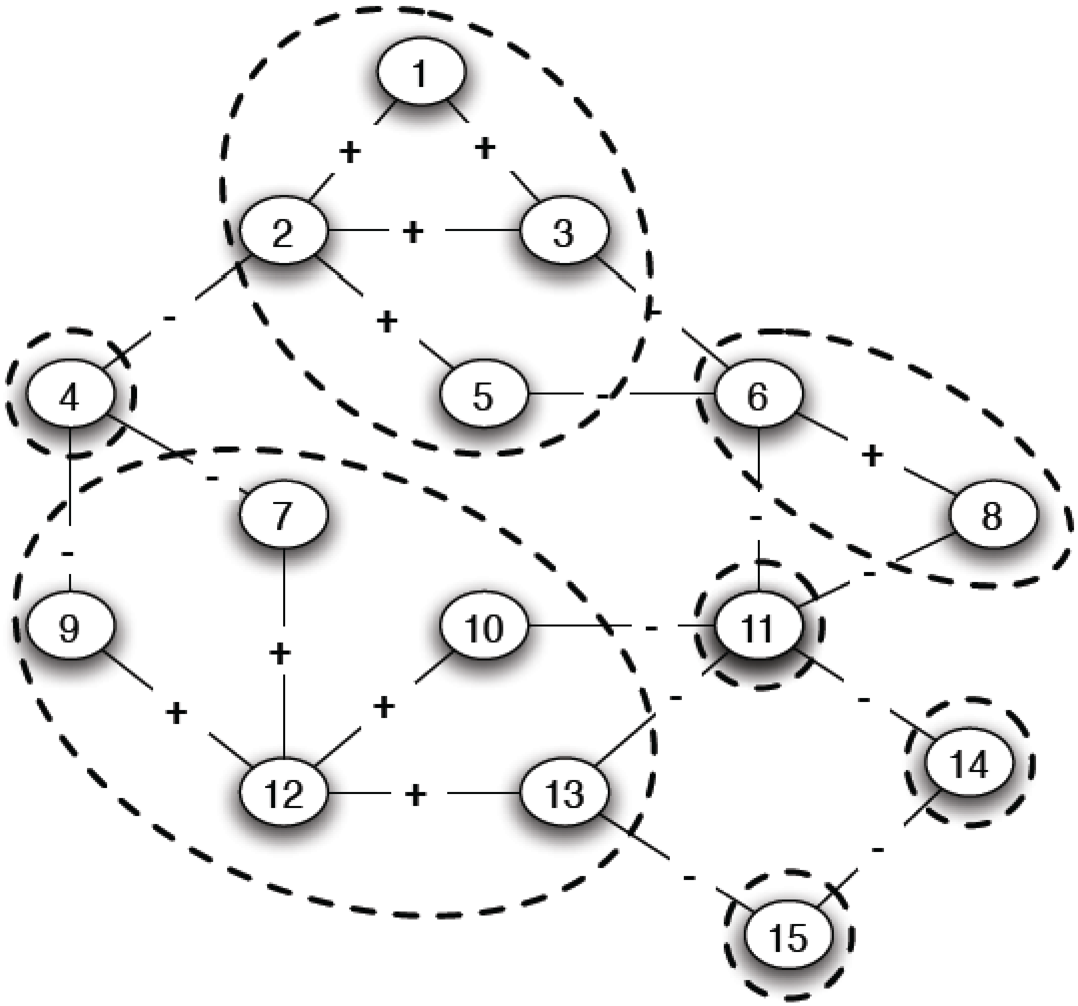
How to compute this?

- Find connected components on positive edges
- For each component create a supernode
- Connect components A and B if there is a negative edge between the members
- Assign supernodes to sides using BFS
- Graph is unbalanced if any two supernodes are assigned the same side
  - Remind that only negative edges are between the supernodes, since the positive edges have been merged before to form the supernodes

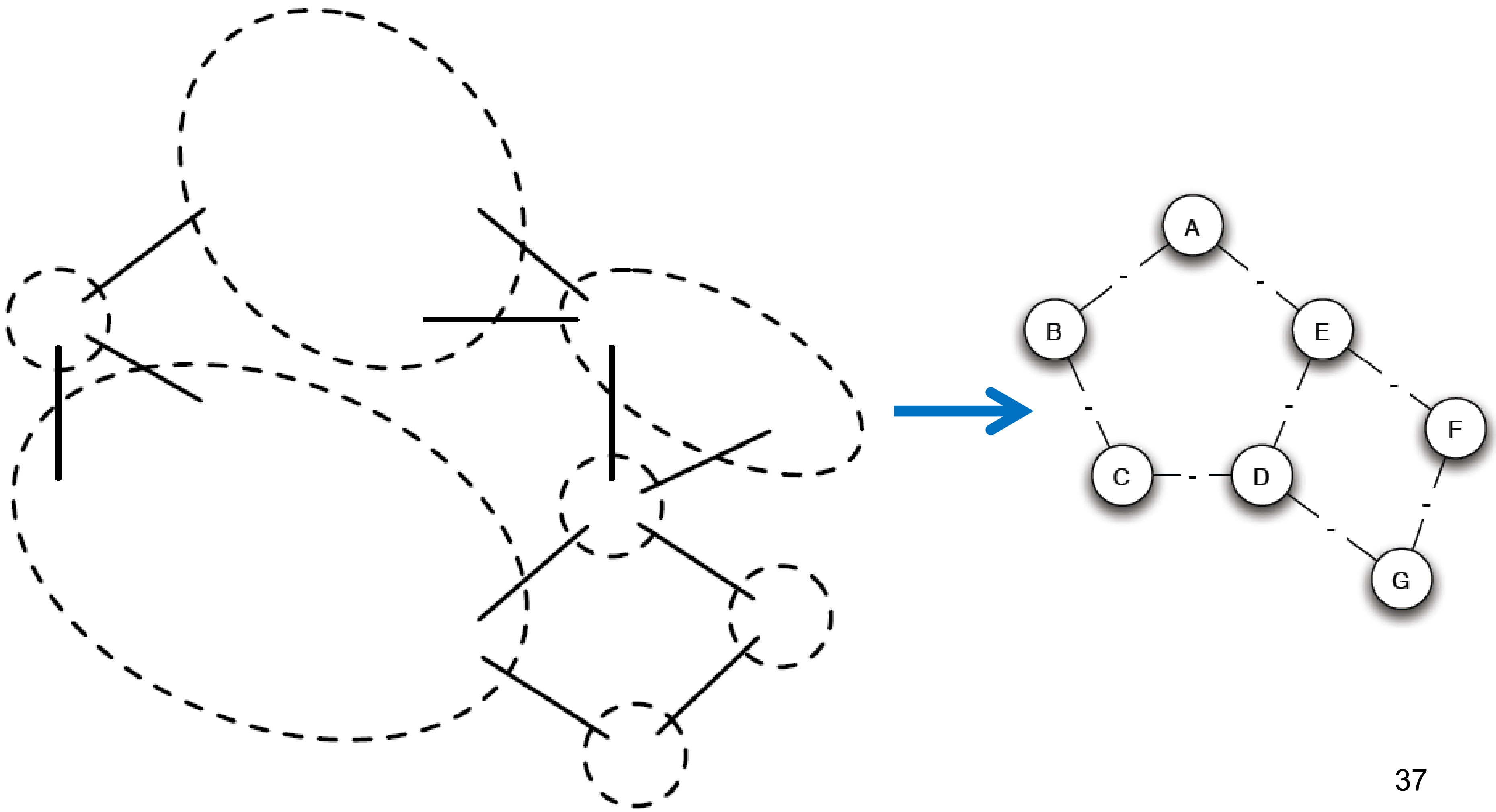
How to compute this?



How to compute this?



How to compute this?

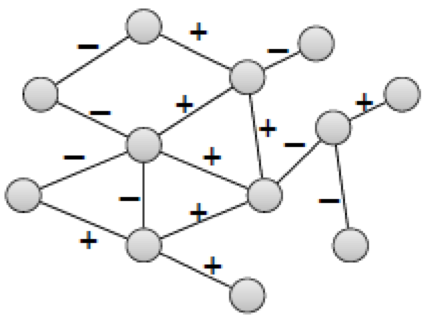


Approximately balanced networks

- <u>Claim</u>: Let ε be any number such that $0 \leq \varepsilon < 1/8$, and define $\sigma = \varepsilon^{1/3}$. If at least $1- \varepsilon$ of all triangles in a labeled complete graph are balanced, then either
- a) There is a set consisting of at least $1-\sigma$ of the nodes in which at least $1-\sigma$ of all pairs are friends, or else
- b) The nodes can be divided into two groups, X and Y , such that
  - (i) at least $1-\sigma$ of the pairs in X like each other,
  - (ii) at least $1-\sigma$ of the pairs in Y like each other, and
  - (iii) at least $1-\sigma$ of the pairs with one end in X and the other end in Y are enemies.

Real large signed networks

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



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

	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%

Source, L. Leskovec et al

Does structural balance hold?

- P(T): Probability of a triad
- $P_0(T)$: Triad probability if the signs would be random

Triad	Epinions		Wikipedia		Balance
	P(T)	$P_0(T)$	P(T)	$P_0(T)$	
+ + +	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	×

Source, L. Leskovec et al

Global factions: structure

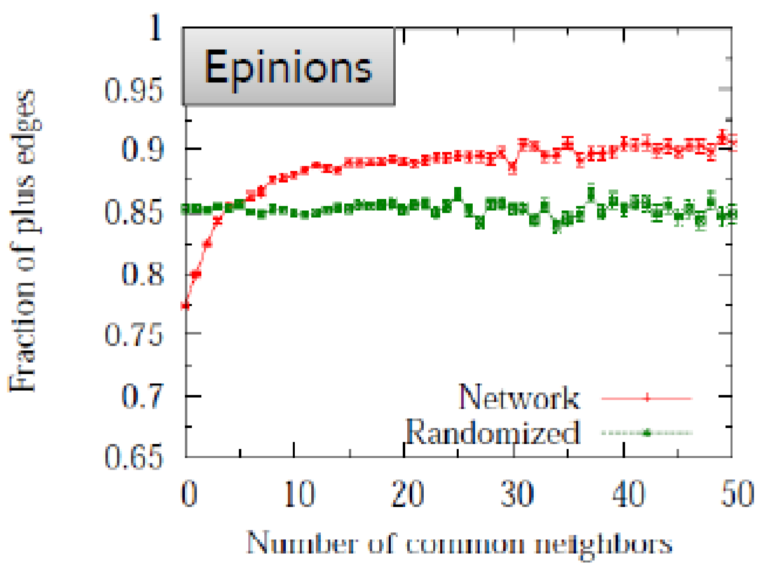
- P(T): Probability of a triad
- Clustering:
  - +net: more clustering than baseline
  - –net: less clustering than expected
- Size of connected component:
  - +/–net: smaller than expected

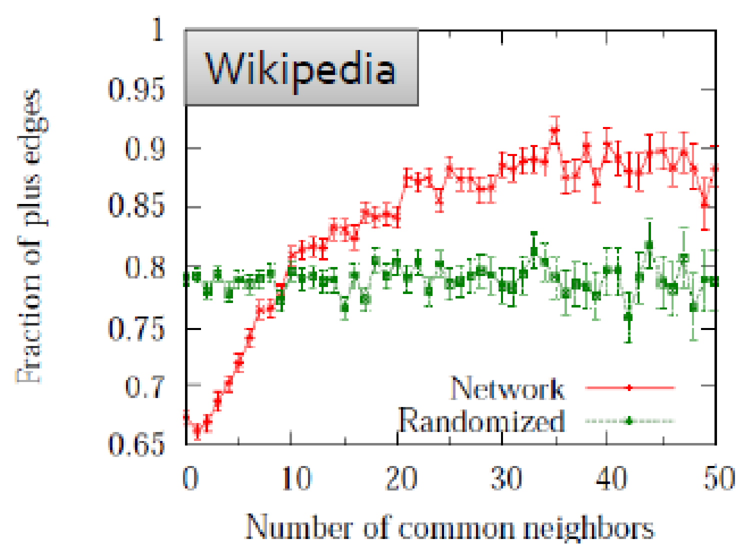
	Size		Clustering		Component	
	Nodes	Edges	Real	Rnd	Real	Rnd
Epinions: −	119,090	123,602	0.012	0.022	0.308	0.334
Epinions: +	119,090	717,027	0.093	0.077	0.815	0.870
Slashdot: −	82,144	124,130	0.005	0.010	0.423	0.524
Slashdot: +	82,144	425,072	0.025	0.022	0.906	0.909
Wikipedia: −	7,115	21,984	0.028	0.031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

Source, L. Leskovec et al

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

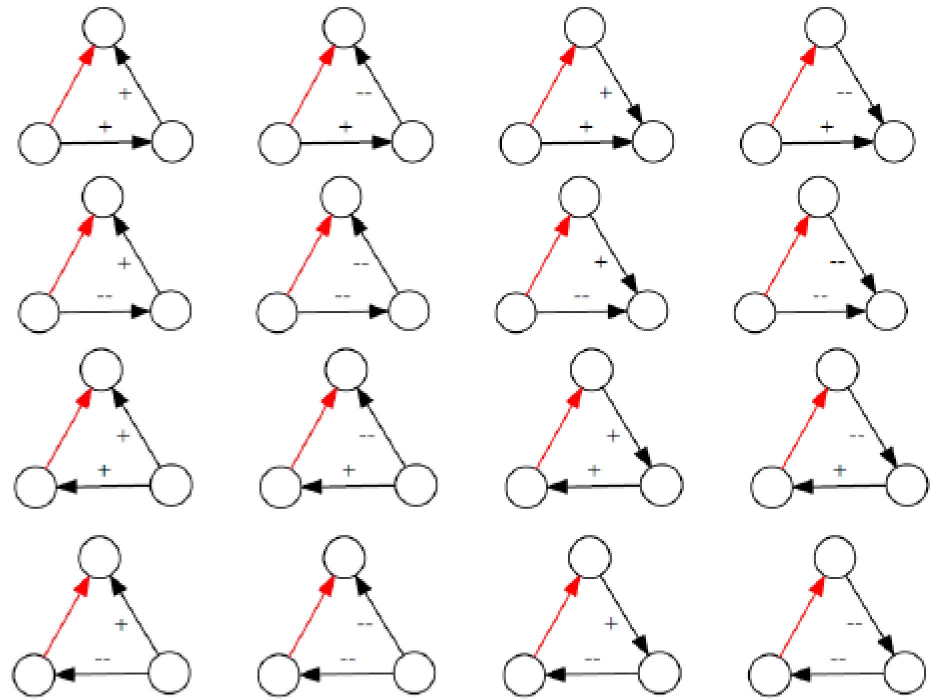




Source, L. Leskovec et al

Directed edges

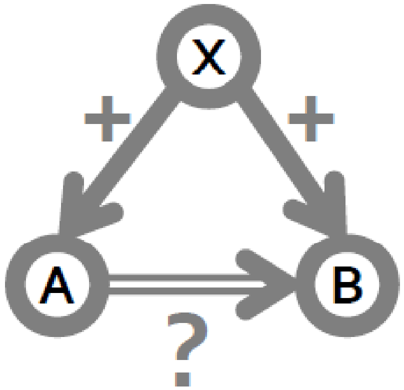
- But, the networks are really directed
  - trust, opinion (friendship)
- How many Δ are now explained by balance?
  - Only half (8 out of 16)
- Can we do better?
  - Yes. **Theory of Status.**



Source, L. Leskovec et al

Theory of Status

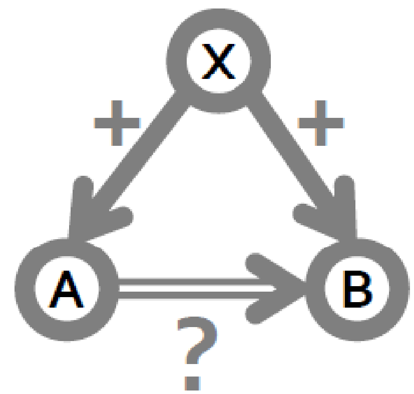
- Edges are directed
- Edges are created over time
  - X has links to A and B
  - Now, A links to B (triad A-B-X)
  - How does sign of A-B depend on signs of X?
- Different users make signs differently:
  - Generative baseline (probability of A giving a +)
  - Receptive baseline (probability of A receiving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



Source, L. Leskovec et al

Joint positive endorsement

- X positively endorses A and B
- Then, the link A-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 → B)
- Can we infer the sign based on opinion of X?

Source, L. Leskovec et al

Example: soccer team

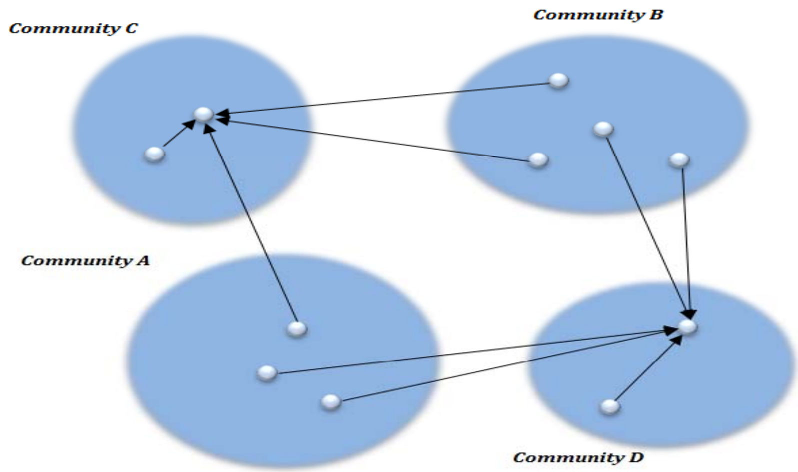
- Conjecture (generative hypothesis):
  - Since B has positive evaluation, B is high status
  - Thus, evaluation from A should be more likely to be positive than A evaluating a random player.
- Conjecture 2 (receptive):
  - Since A has positive evaluation, A is high status
  - Thus, sign from A should be less likely to be positive than an evaluation received by B from a random player

Sign of A → B deviates in different directions depending on different viewpoints.

Source, L. Leskovec et al

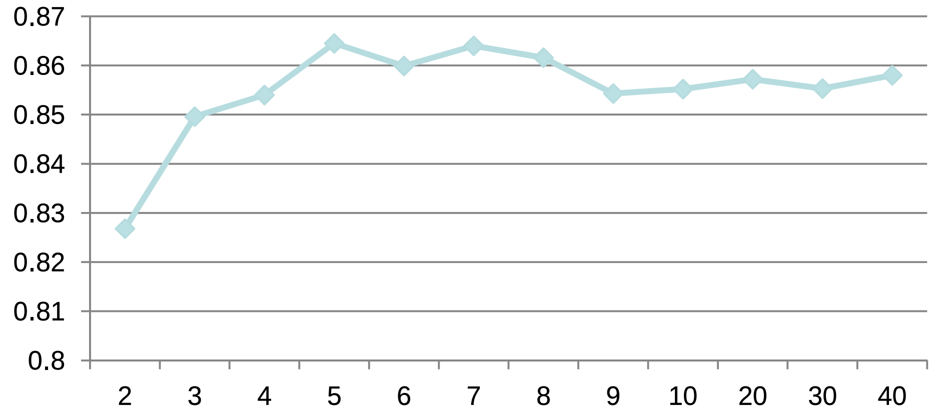
Clustering and sign prediction

- Clustering the sign graph
- Minimizing the positive inter-cluster links
- Minimizing the negative intra-cluster links
- Applying collaborative filtering on the clustered network
- No machine learning approach
- highly scalable

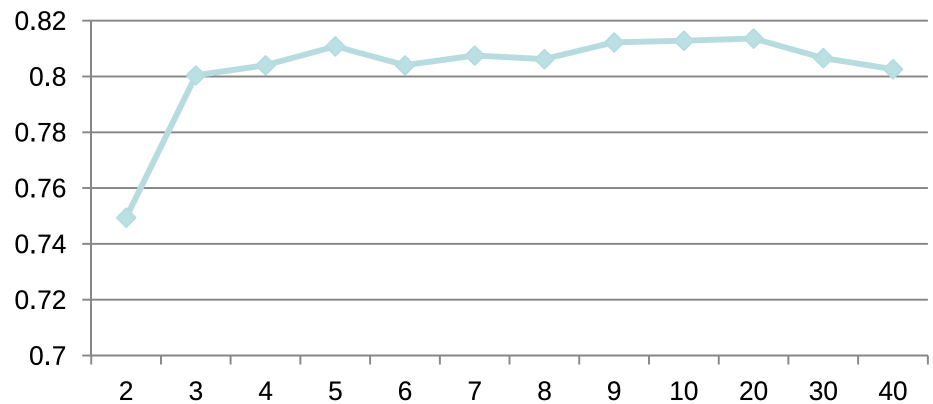


Source, Javari and Jalili, ACM Transactions on IST, 2014

Clustering and sign prediction



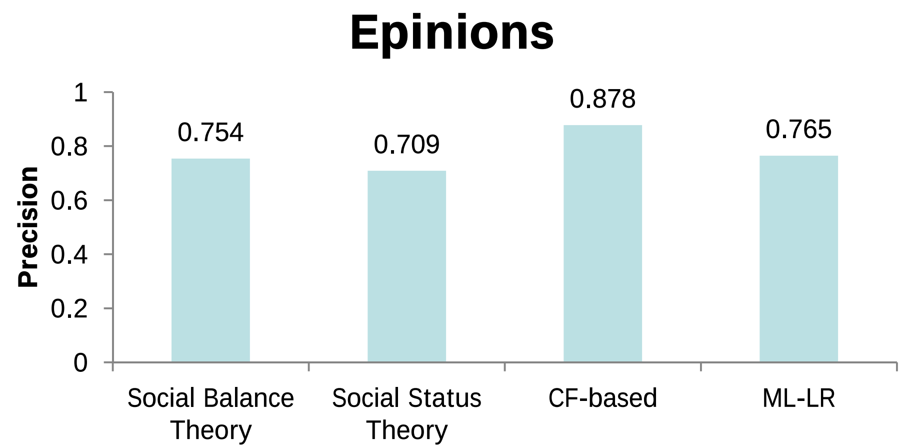
Precision as a function of the number of clusters for Epinions dataset

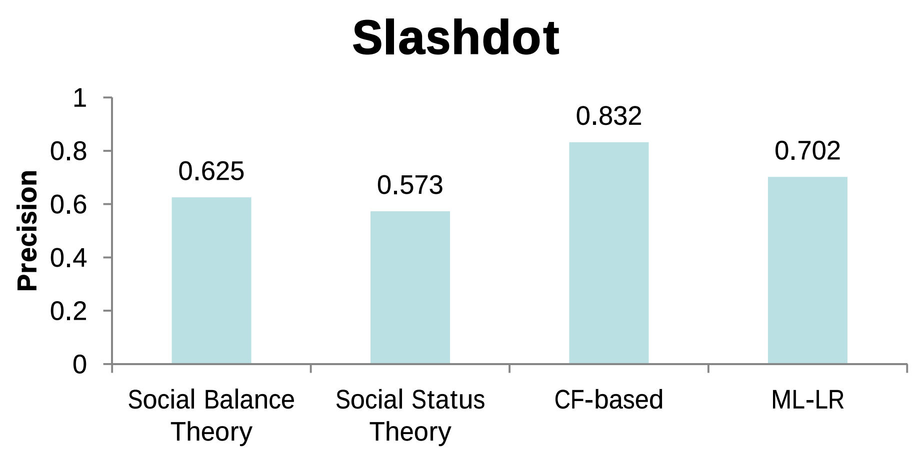


Precision as a function of the number of clusters for Slashdot dataset

Source, Javari and Jalili, ACM Transactions on IST, 2014

Clustering and sign prediction





Source, Javari and Jalili, ACM Transactions on IST, 2014

Ranking nodes in signed networks

$$\begin{cases} h_i^{(+)} = \sum\limits_{j \in IN_i^{(+)}} a_j^{(+)} \quad ; \quad a_i^{(+)} = \sum\limits_{j \in OUT_i^{(+)}} h_j^{(+)} \\ h_i^{(-)} = \sum\limits_{j \in IN_i^{(-)}} a_j^{(-)} \quad ; \quad a_i^{(-)} = \sum\limits_{j \in OUT_i^{(-)}} h_j^{(-)} \end{cases}$$

$$\begin{cases} a^{(+)}(t+1) = A^{+T} A^{+} a^{(+)}(t); a^{(-)}(t+1) = A^{-T} A^{-} a^{(-)}(t) \\ h^{(+)}(t+1) = A^{+} A^{+T} h^{(+)}(t); h^{(-)}(t+1) = A^{-} A^{-T} h^{(-)}(t) \end{cases}$$

HITS

Source, Shahriairi and Jalili, SNAM, 2014

Ranking nodes in signed networks

$$\begin{cases} h_i(t+1) = \dfrac{\sum_{j\in IN_i^{(+)}} a_j(t) - \sum_{j\in IN_i^{(-)}} a_j(t)}{\sum_{j\in IN_i^{(+)}} a_j(t) + \sum_{j\in IN_i^{(-)}} a_j(t)} \\ \\ a_i(t+1) = \dfrac{\sum_{j\in OUT_i^{(+)}} h_j(t) - \sum_{j\in OUT_i^{(-)}} h_j(t)}{\sum_{j\in OUT_i^{(+)}} h_j(t) + \sum_{j\in OUT_i^{(-)}} h_j(t)} \end{cases}$$

Modified HITS

Source, Shahriairi and Jalili, SNAM, 2014

Ranking nodes in signed networks

$$PR_i^{+}\left(t+1\right)=\alpha\sum_{j\in IN_i}\frac{PR^{+}{}_{j}\left(t\right)}{|OUT_j^{(+)}|}+(1-\alpha)\frac{1}{N}$$

$$PR_i^{-}\left(t+1\right)=\alpha\sum_{j\in IN_i}\frac{PR^{-}{}_{j}\left(t\right)}{|OUT_j^{(-)}|}+(1-\alpha)\frac{1}{N}$$

$$PR=PR^{+}-PR^{-}$$

Modified PageRank

Source, Shahriairi and Jalili, SNAM, 2014

Ranking nodes in signed networks

$$Rep_i = \frac{\sum_{j \in IN_i^{(+)}} R_j - \sum_{j \in IN_i^{(-)}} R_j}{\sum_{j \in IN_i^{(+)}} R_j + \sum_{j \in IN_i^{(-)}} R_j} \qquad Opt_i = \frac{\sum_{j \in OUT_i^{(+)}} R_j - \sum_{j \in OUT_i^{(-)}} R_j}{\sum_{j \in OUT_i^{(+)}} R_j + \sum_{j \in OUT_i^{(-)}} R_j}$$

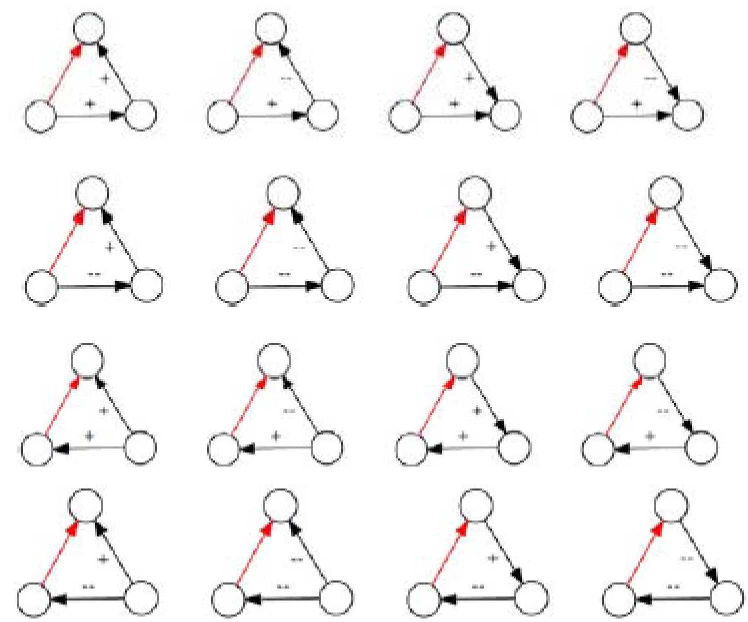
- Reputation and Optimism for predicting the signs
- Features for the nodes
- Use of machine learning approach
- Resulting in good performance

Source, Shahriairi and Jalili, SNAM, 2014

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:

$s_{out}(t) = \sum_A (t_A - d_A p_A) / \sqrt{\sum_A d_A p_A (1-p_A)}$



Contextualized links t

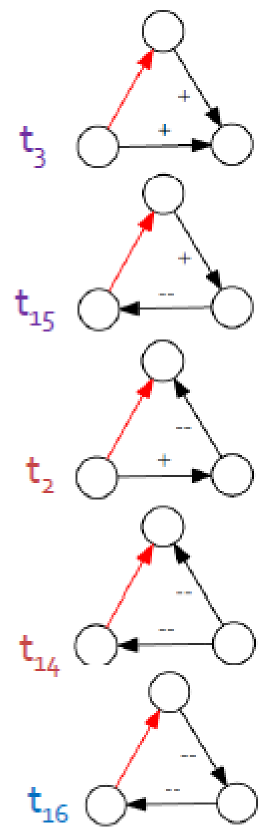
d_A: out-degree of A
p_A: probability of A giving +
t_A: number of + of A in context t

Source, L. Leskovec et al

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



Source, L. Leskovec et al

Readings

- "Networks, Crowds, and Markets" by Easley and Kleinberg (Chapter 5).
- Javari, A., & Jalili, M. (2014). Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links. ACM Transactions on Intelligent Systems and Technology, 5(2).
- Shahriari, M., & Jalili, M. (2014). Ranking nodes in signed social networks. Social Networks Analysis and Mining, 4(1), 172.