Social and Information Networks

CSCC46H, Fall 2022

Lecture 4

Prof. Ashton Anderson
ashton@cs.toronto.edu

Logistics

A1 due next week on MarkUs, last time submissions will be accepted is Friday at 10am ET.

First letter of last name A–J? First blog post due next Friday at 5pm.
https://cmsweb.utsc.utoronto.ca/c46blog-f22/

CSCC46 Piazza created

Today

Signed networks
Homophily and Friendship Paradox

Positive and Negative Relationships

So far, edges mostly interpreted positively

—Friendship
—Interaction
—Collaboration

But relationships can be **negative** too

—Dislike
—Bad interaction
—Enemy

Network Representation

How would you model this?

Signed Networks

Networks with positive and negative relationships

Consider an undirected complete graph
Label each edge as either:

Positive: friendship, trust, positive sentiment, …

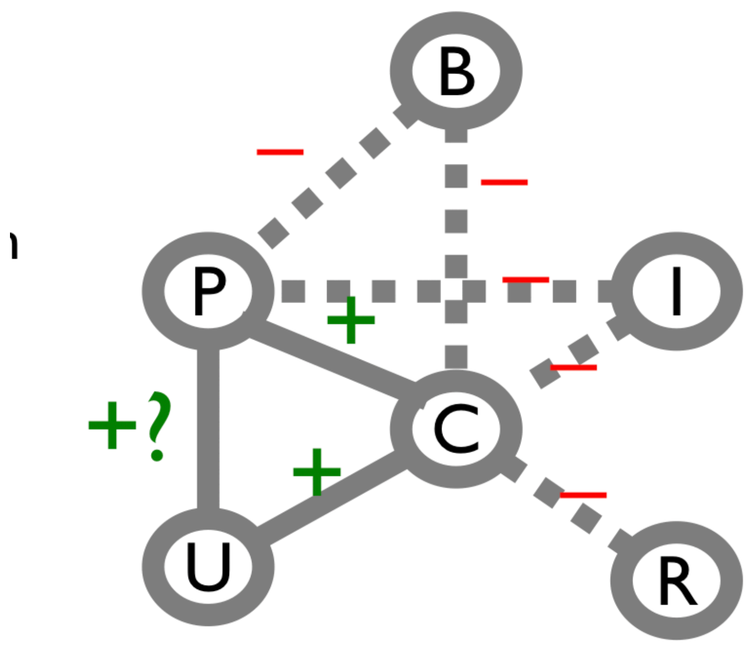
Negative: enemy, distrust, negative sentiment, …

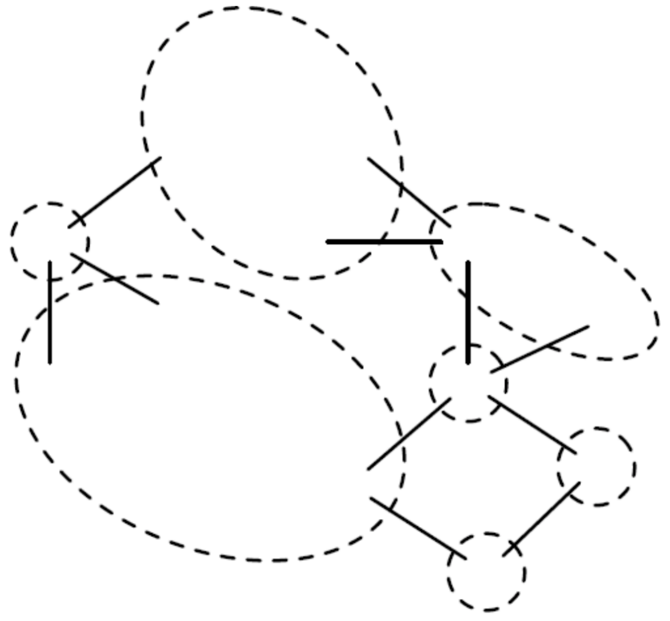
Questions about Signed Networks

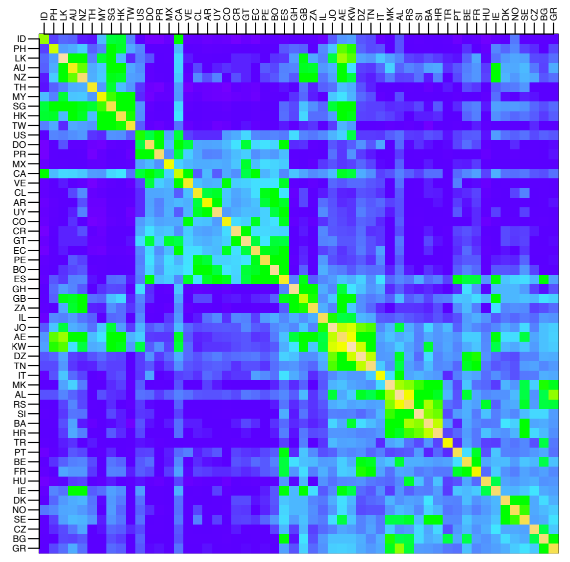
What are the typical patterns of interaction in signed networks?

How do we reason about local and global structure of positive and negative interactions?

What are the patterns in empirical data?



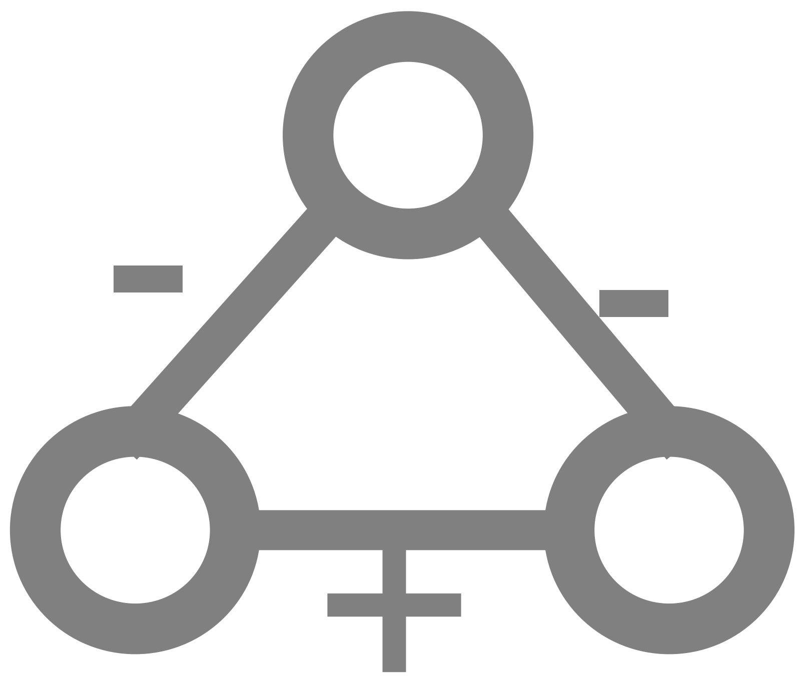




Signed Networks

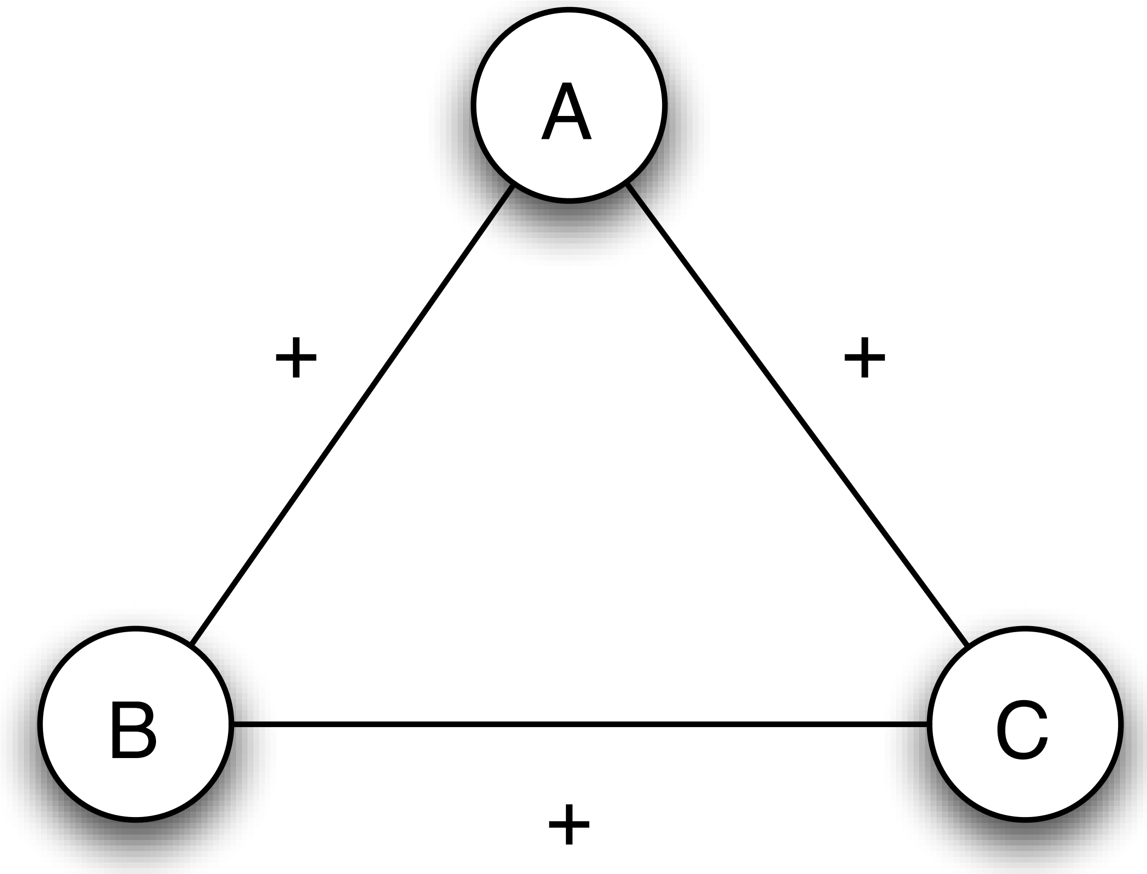
Networks with positive and negative relationships

Our basic unit of investigation will be **signed triangles**

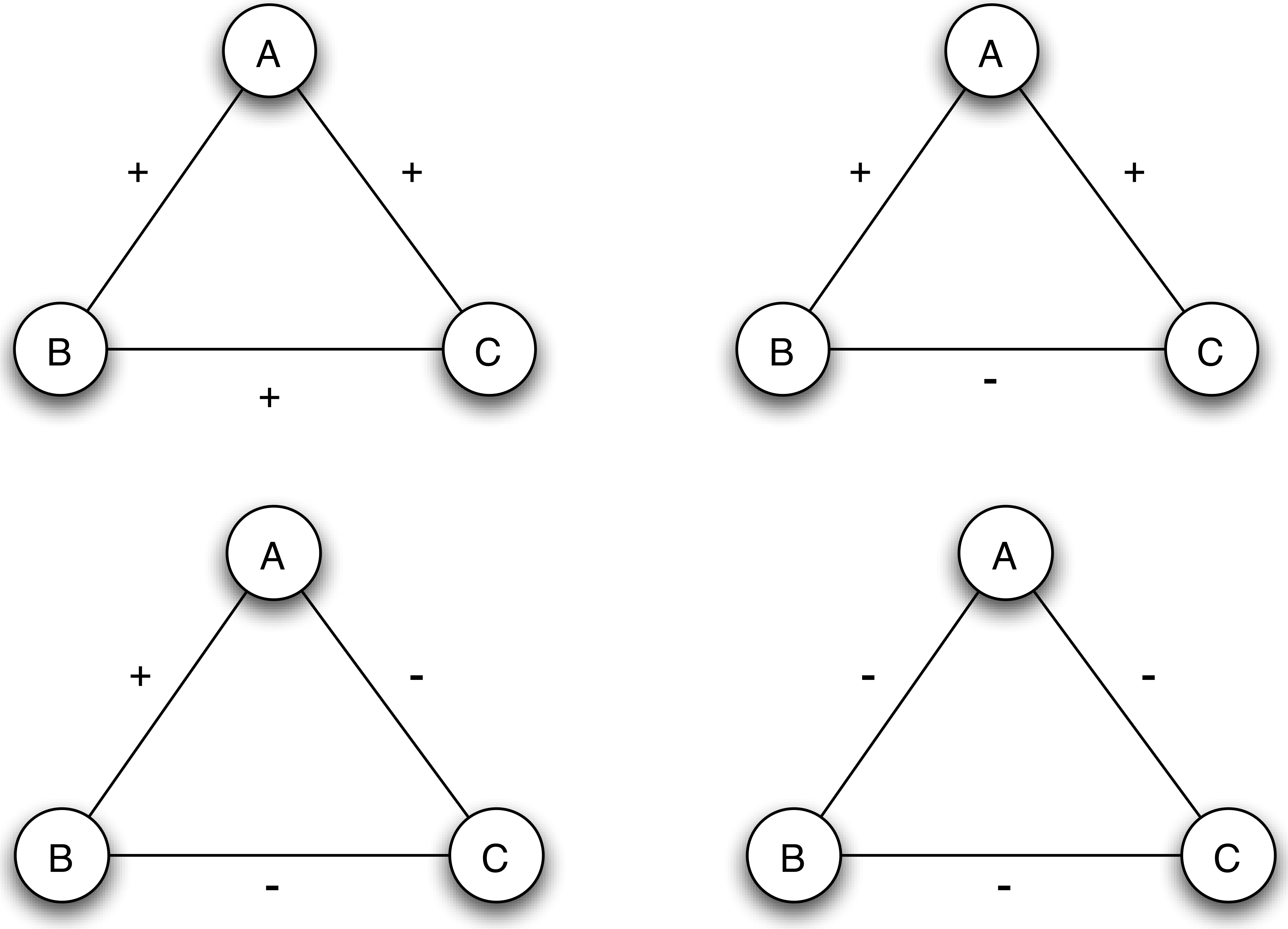


Focus on undirected networks

Structural Balance

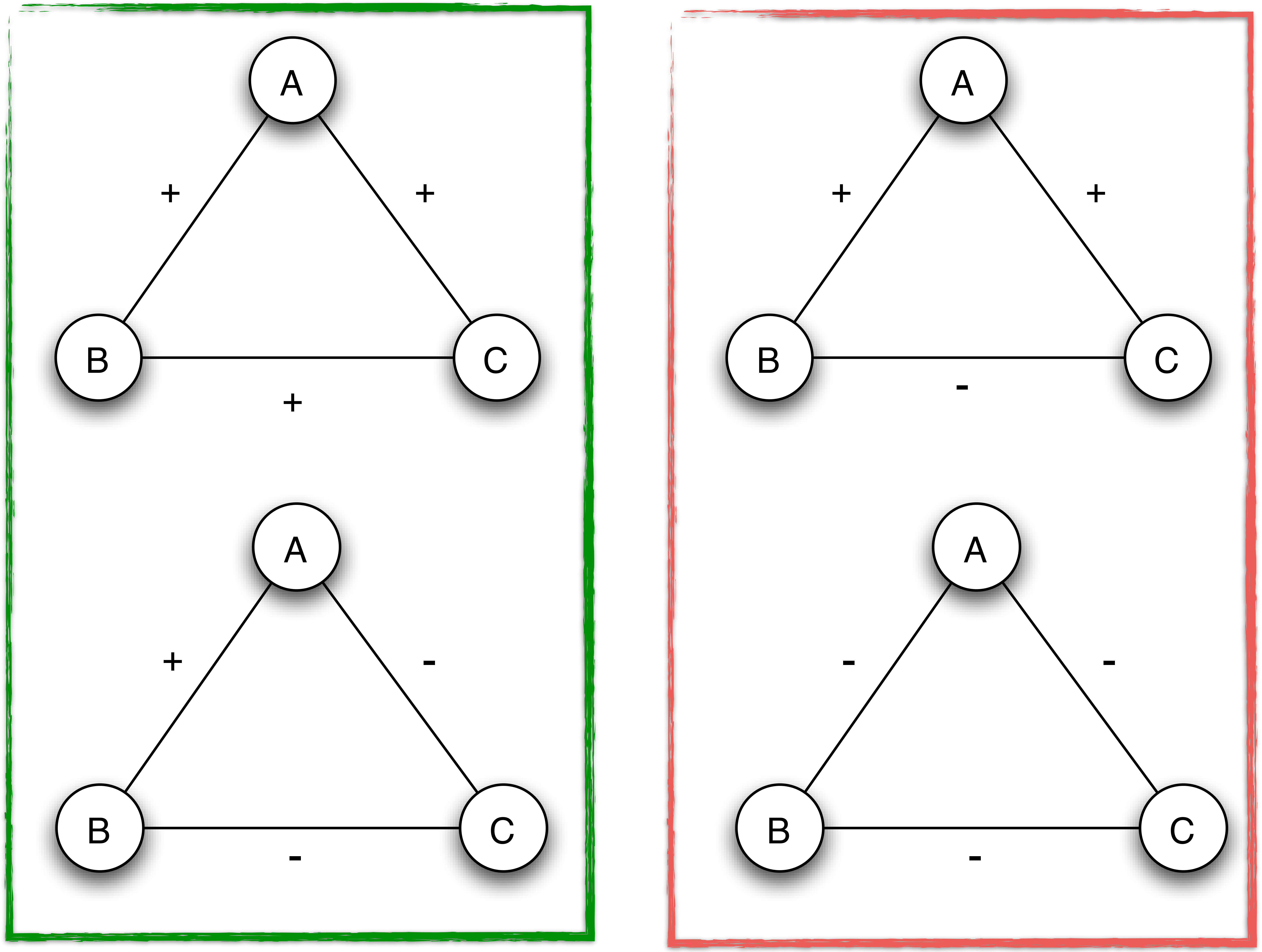


Structural Balance



Four signed triads: which are **stable**?

Structural Balance



Four signed triads: which are **stable**?

Theory of Structural Balance

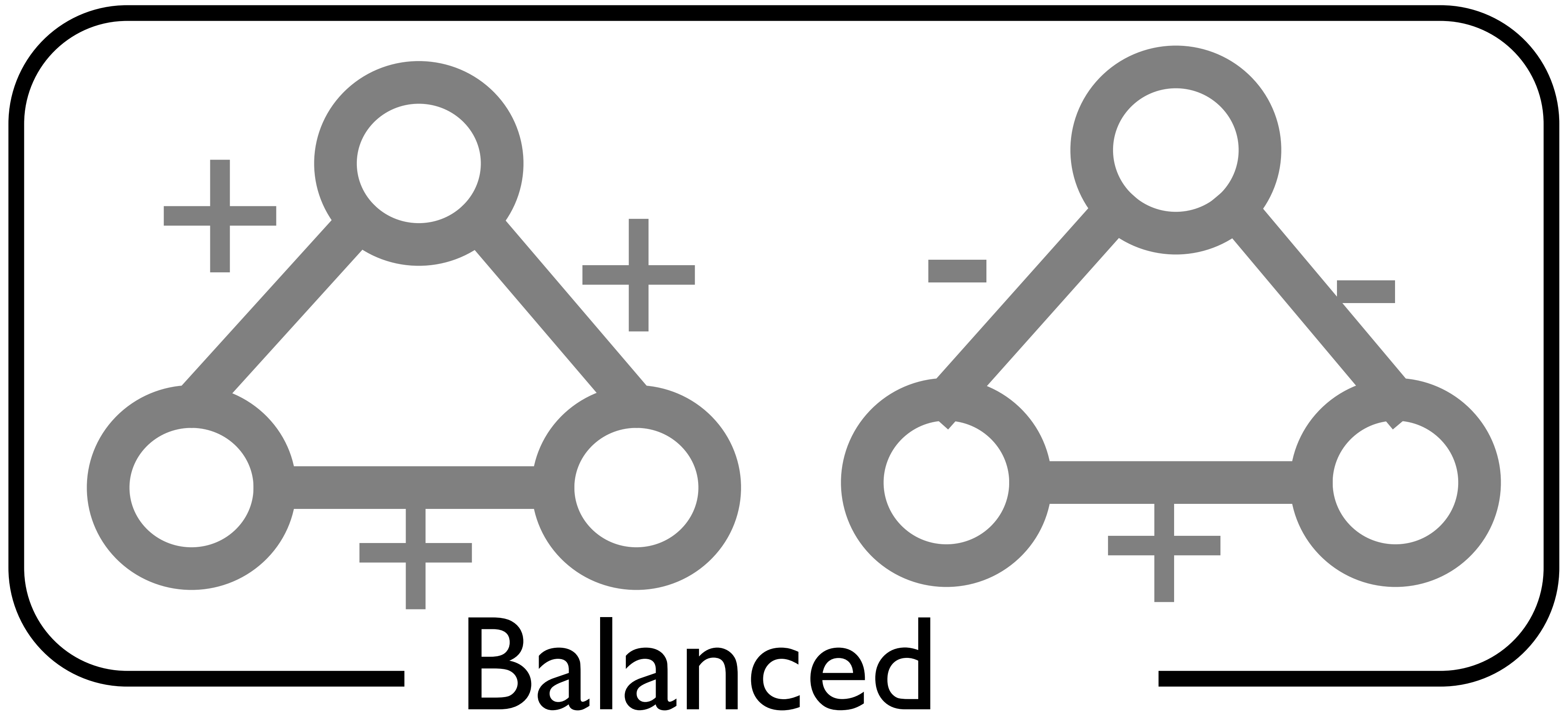
Start with the intuition [Heider '46]:

Friend of my friend is my friend

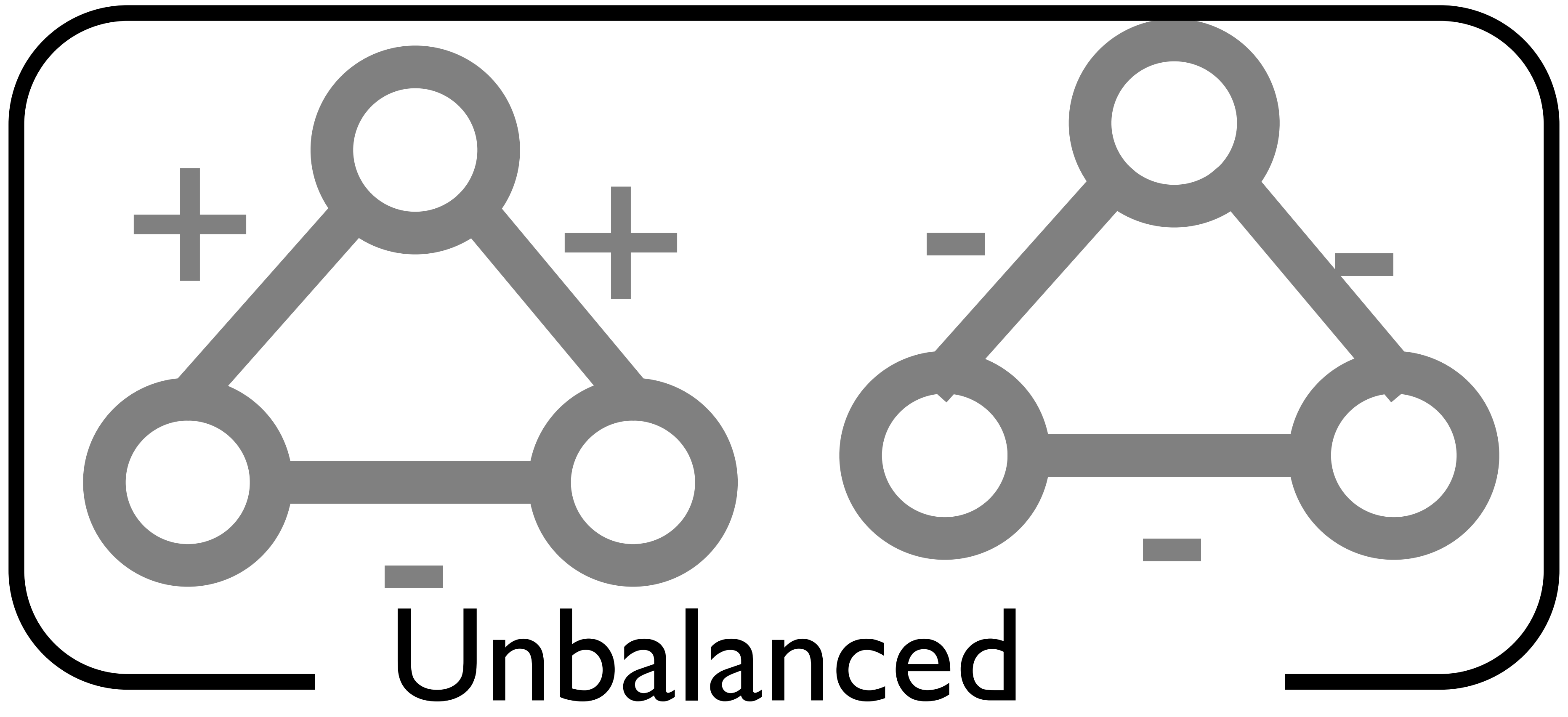
Enemy of enemy is my friend

Enemy of friend is my enemy

Look at connected triples of nodes:

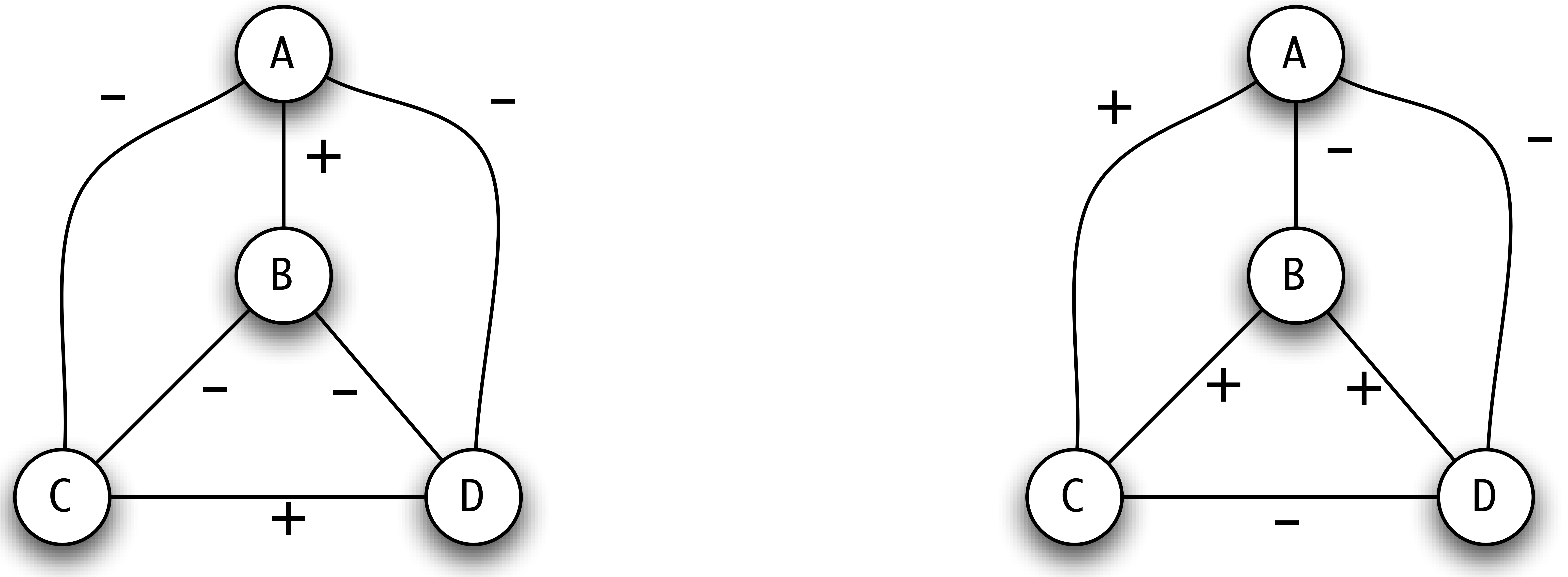


Consistent with "friend of a friend" or "enemy of the enemy" intuition



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

Structural Balance

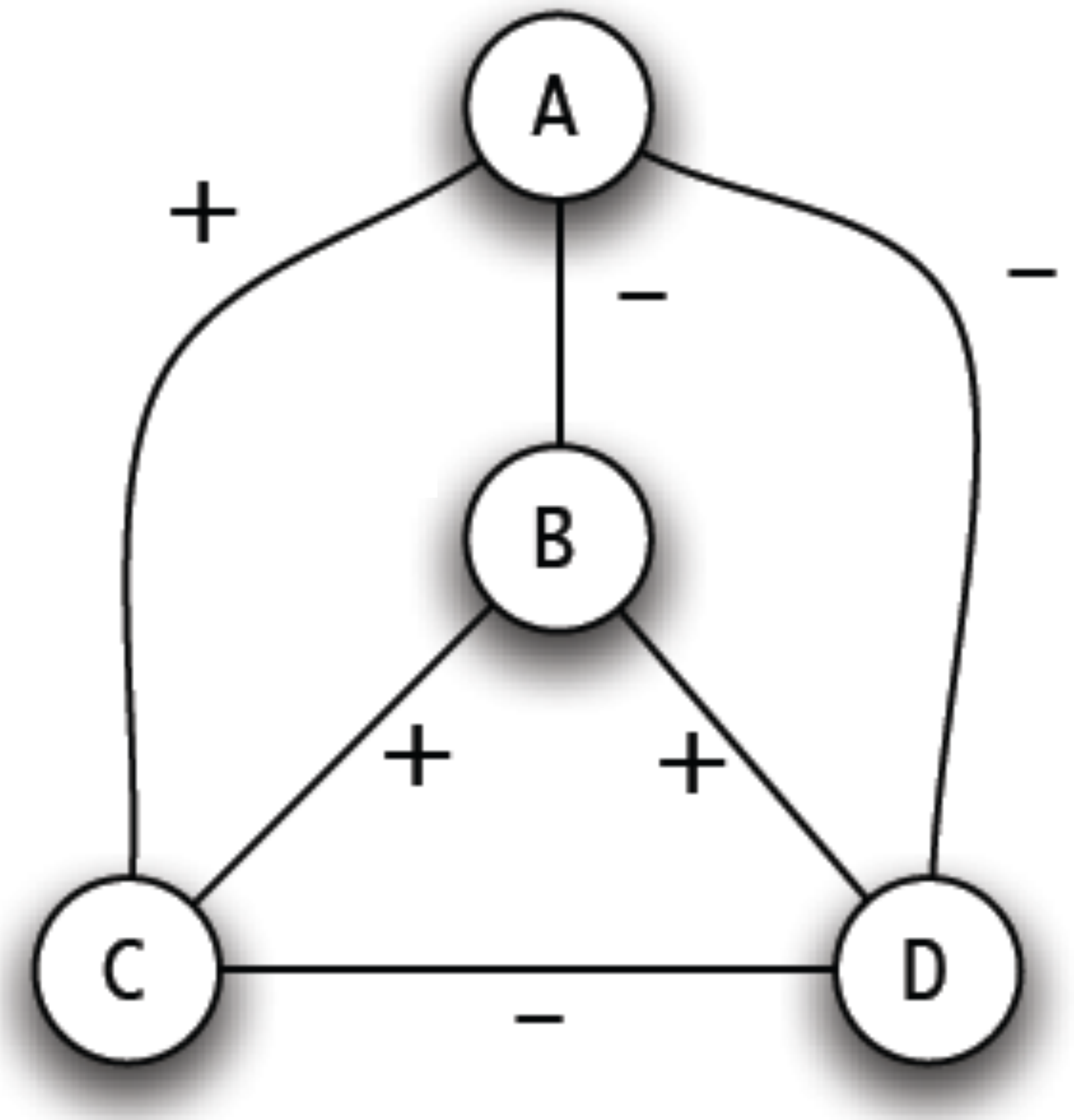


Which network is balanced?

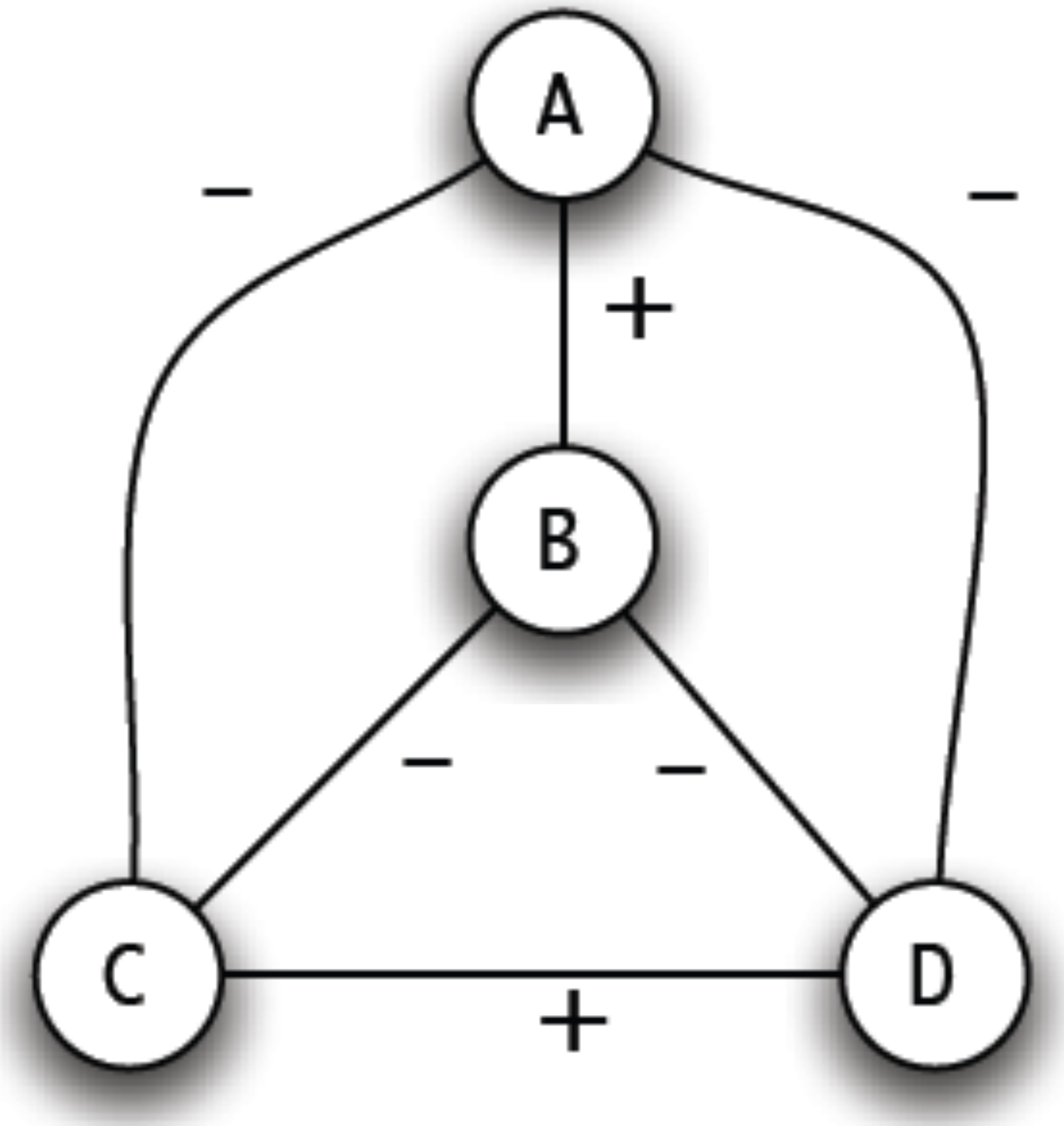
Balanced/Unbalanced Networks

Define: A complete graph is *balanced* if every connected triple of nodes has:

All 3 edges labeled + **or** Exactly 1 edge labeled +

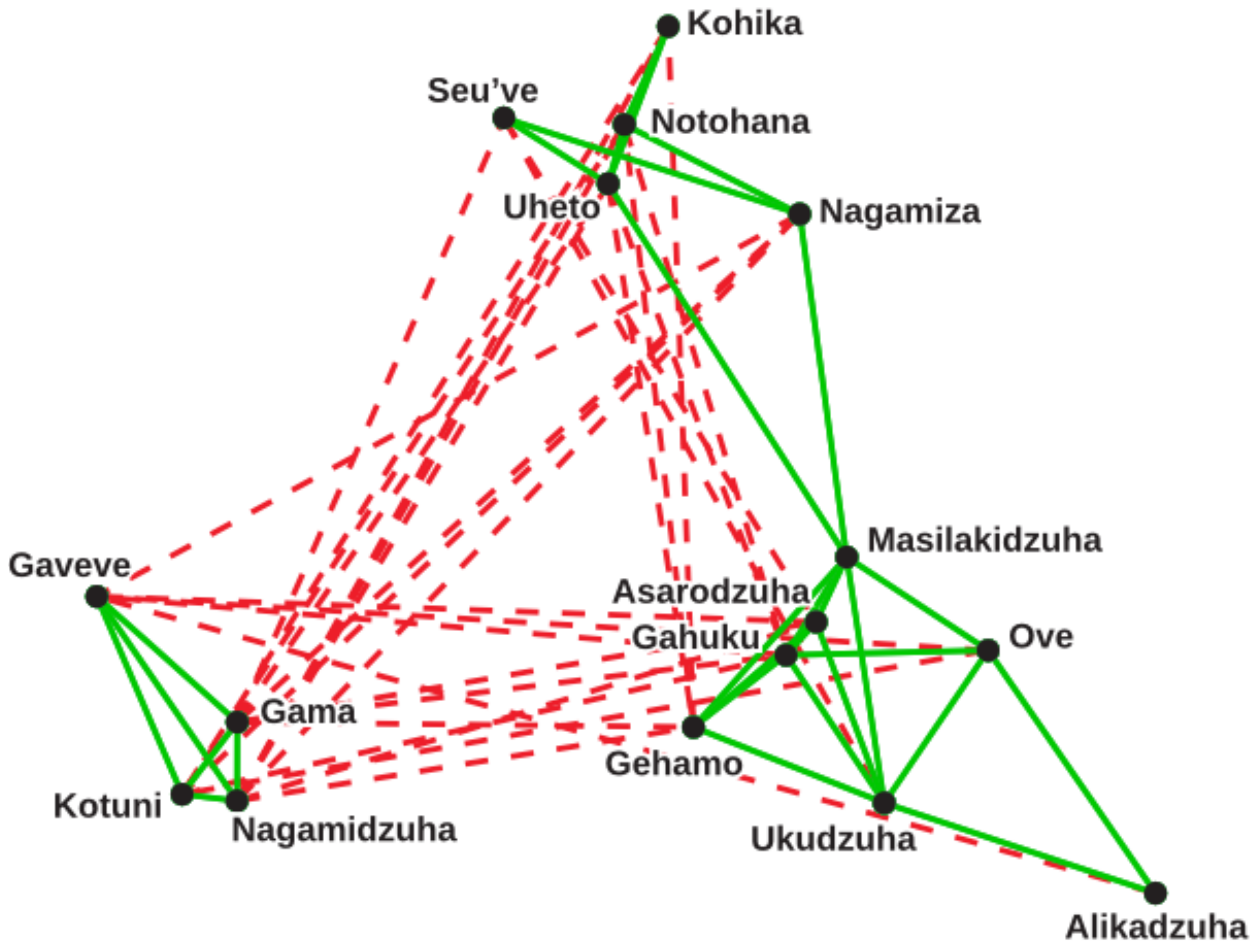


Unbalanced

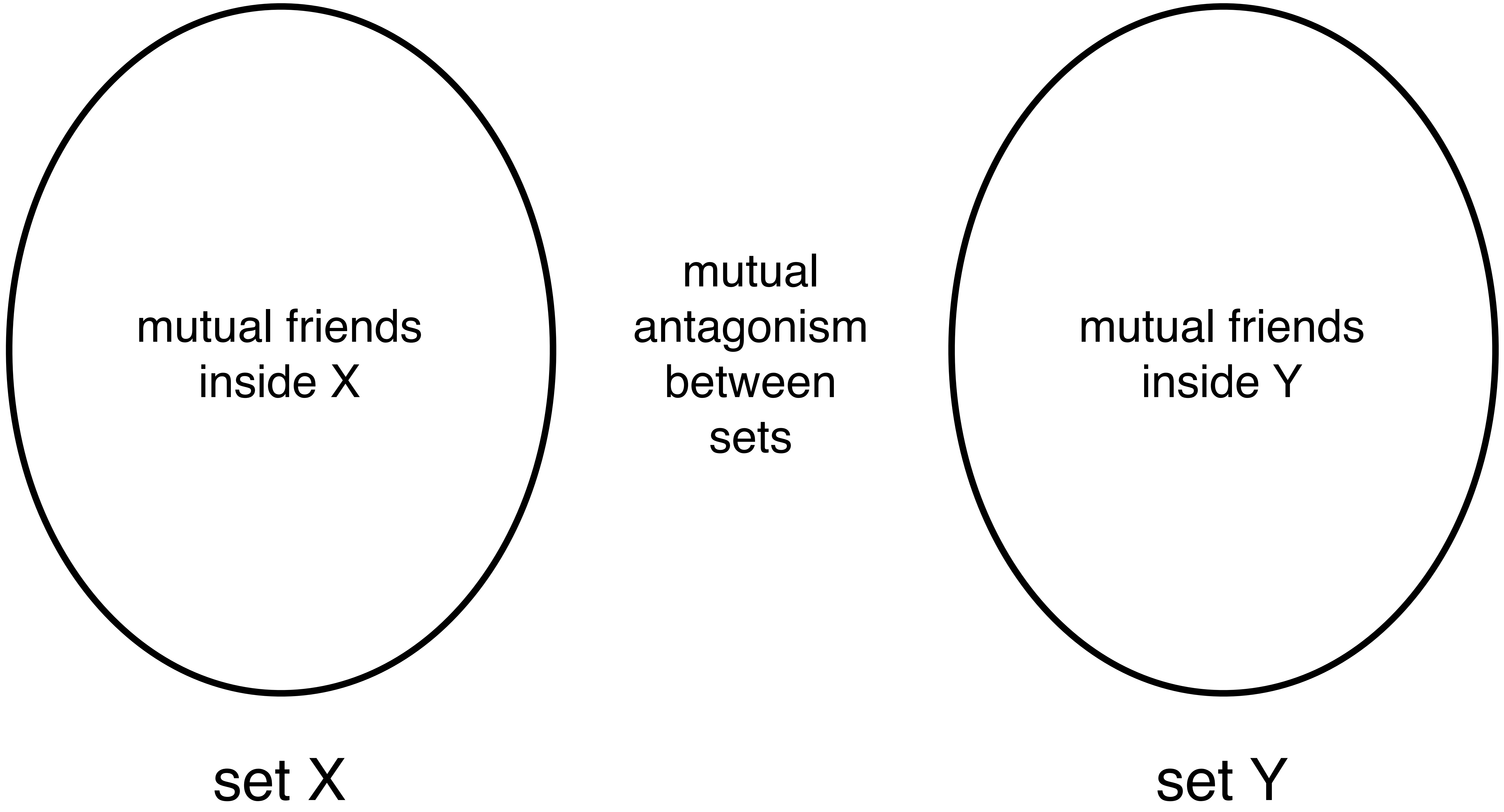


Balanced

The Tribes of Eastern Central Highlands of New Guinea



How general is this?



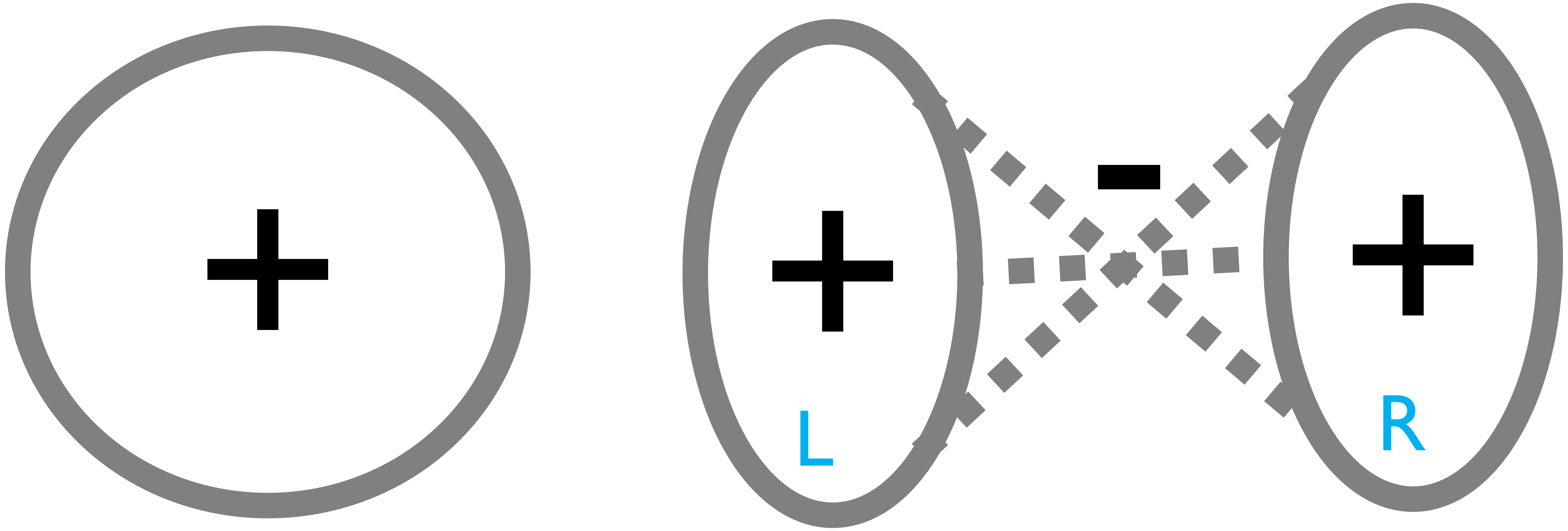
Local Balance → Global Factions

The Balance Theorem: Balance implies global coalitions
[Cartwright-Harary]

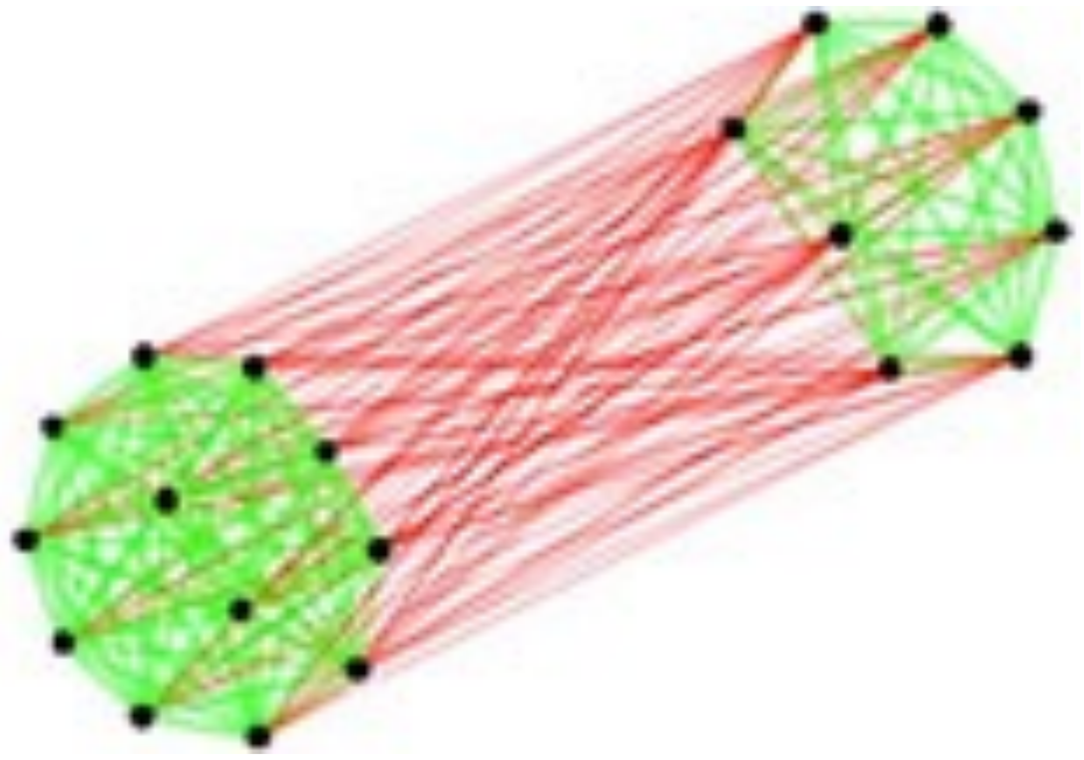
If **all triangles are balanced**, then either:

A) The network contains only positive edges, or

B) The network can be split into two factions: Nodes can be split into 2 sets where negative edges only point between the sets



Balance Theorem



Global coalitions => balance

Straightforward

Every complete graph that looks like "this" is balanced

Balance => Global coalitions

Less straightforward

Every complete graph that's balanced looks like "this"?

Balance Theorem

Global coalitions => balance:

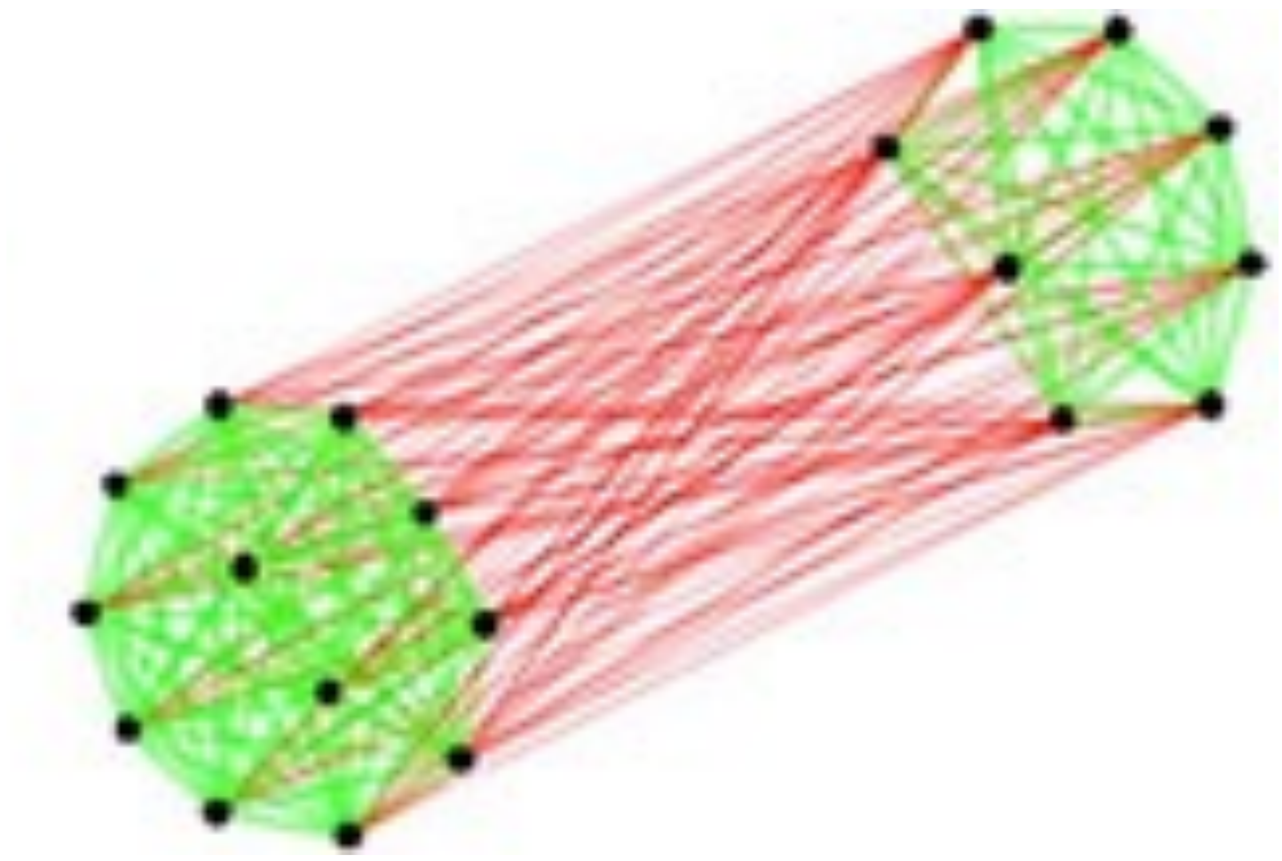
Any triangle is one of two types:

A) All 3 nodes in one of the partitions

B) 2 nodes in one partition, 1 in the other

A): all 3 edges are + → balanced

B): 2 nodes in one partition are +,
other 2 edges are - → balanced



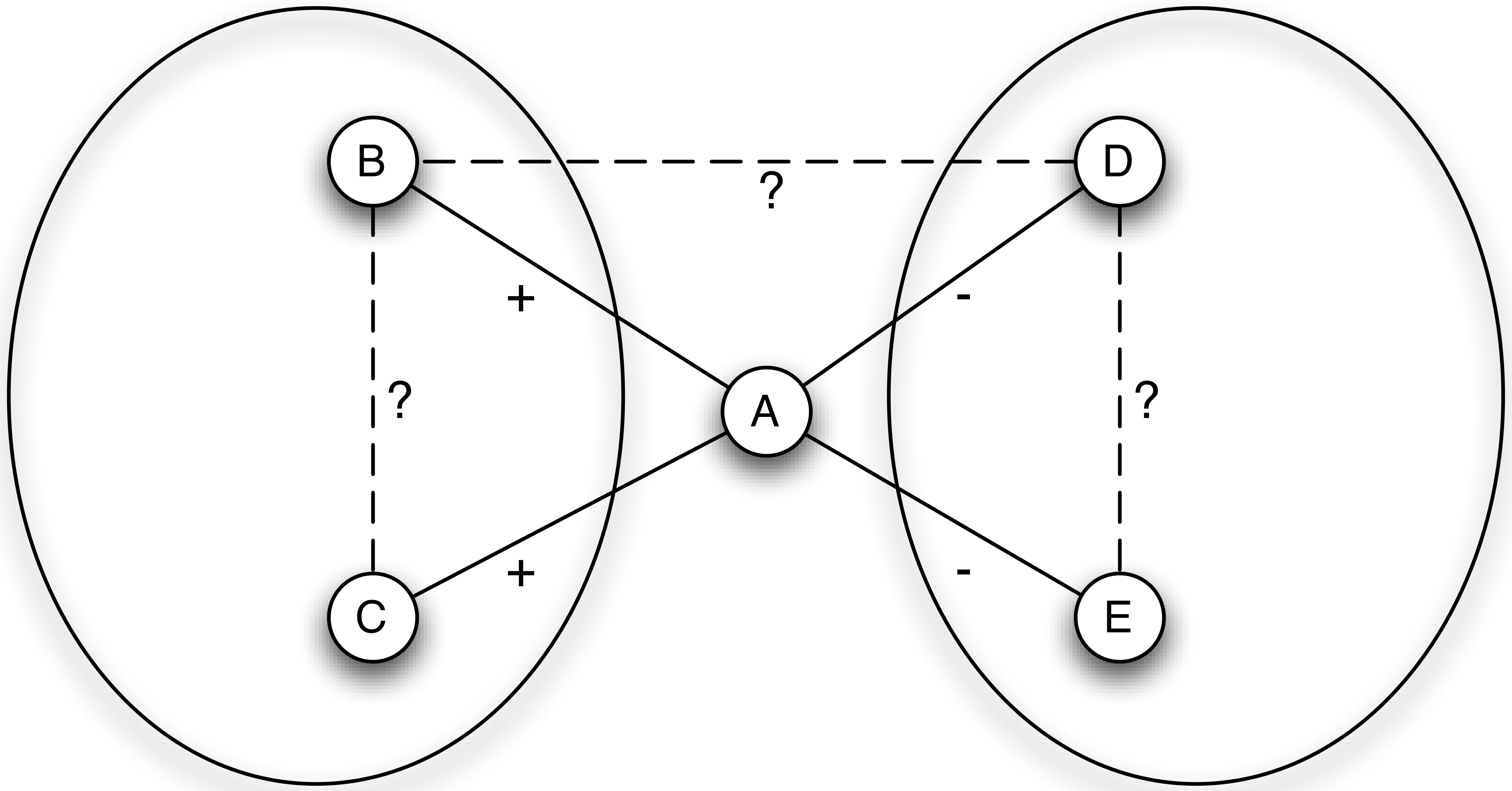
Proof of Balance Theorem

Balance => Global coalitions:

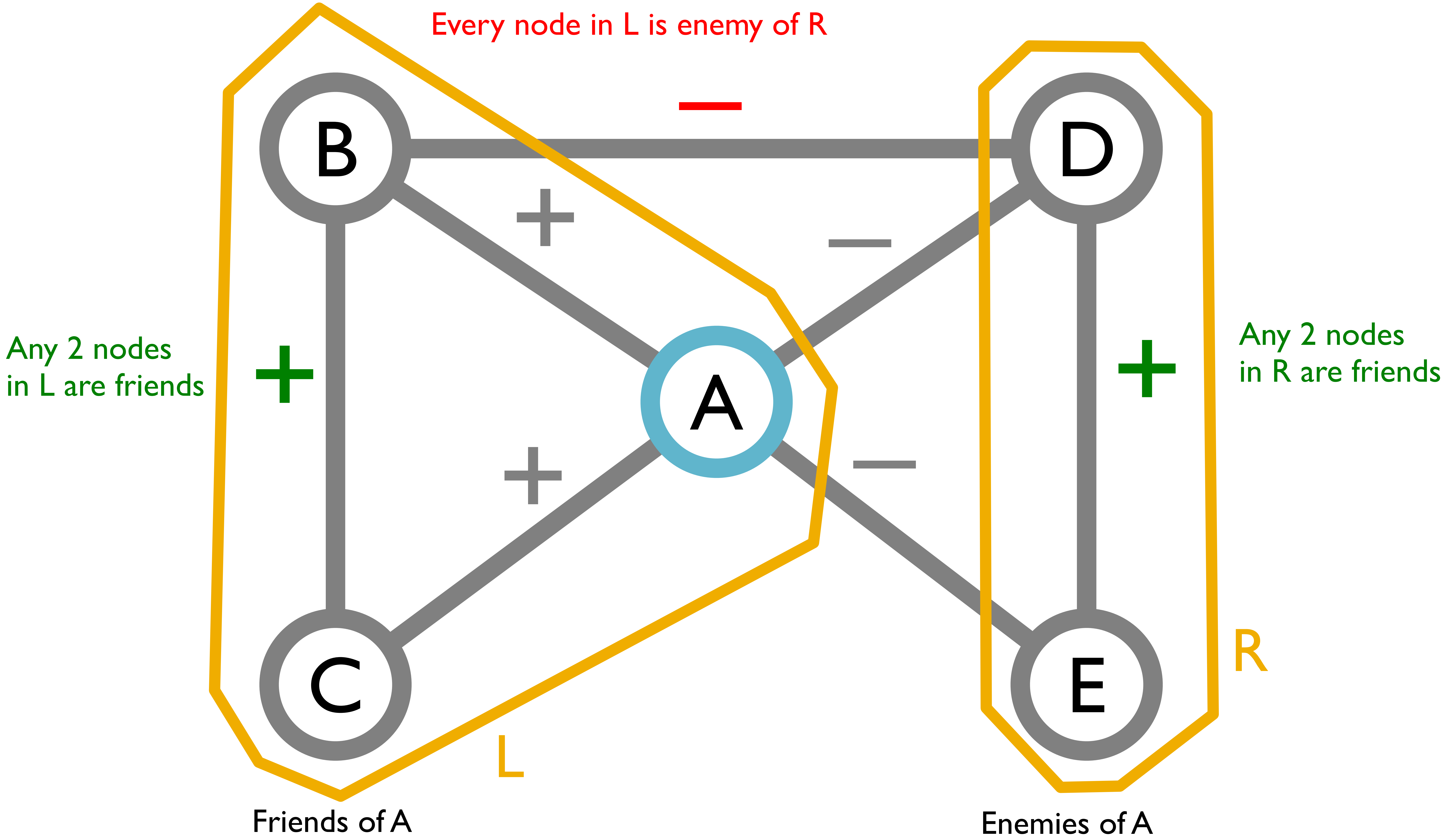
Pick a node **A**.

Because it's a complete graph, **A** is either friends or enemies with each person.

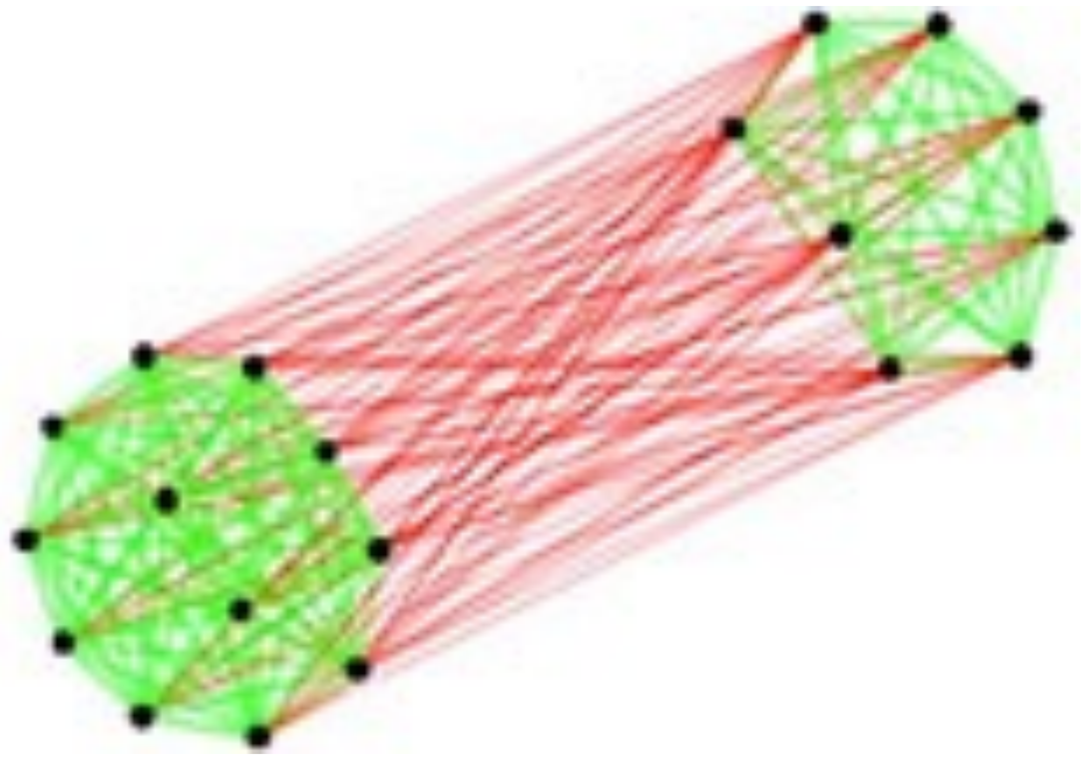
Now **check 3 cases**:



Proof of Balance Theorem



Balance Theorem



Global coalitions => balance

Straight-forward

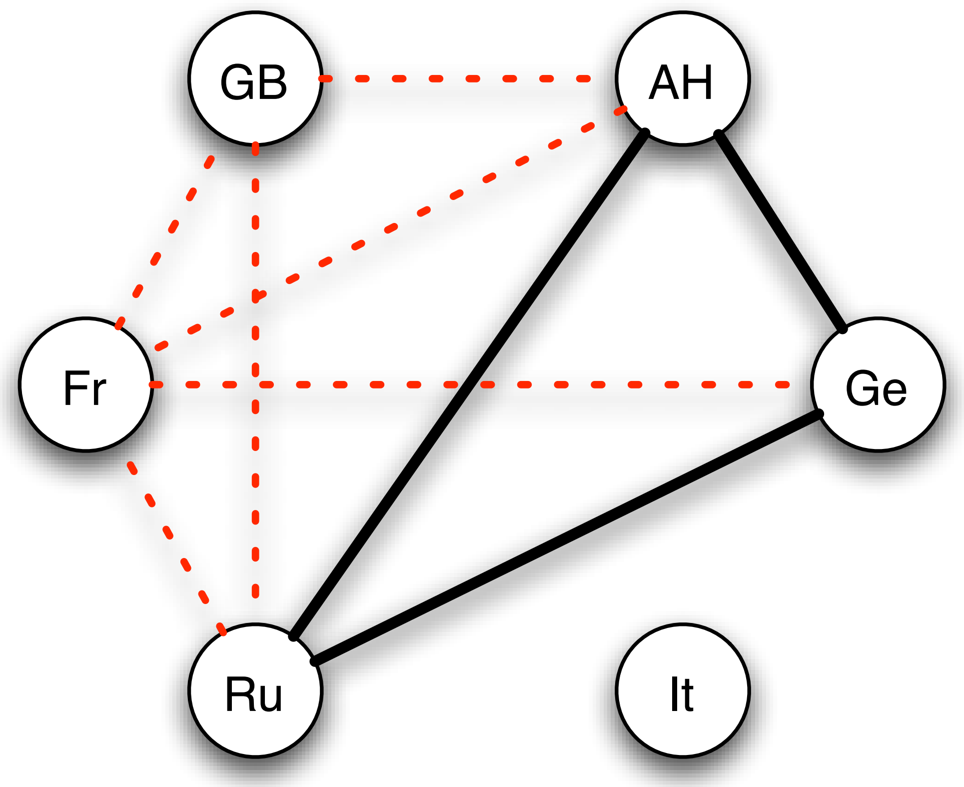
Every complete graph partitioned into two friendly coalitions that dislike either other is balanced

Balance => Global coalitions

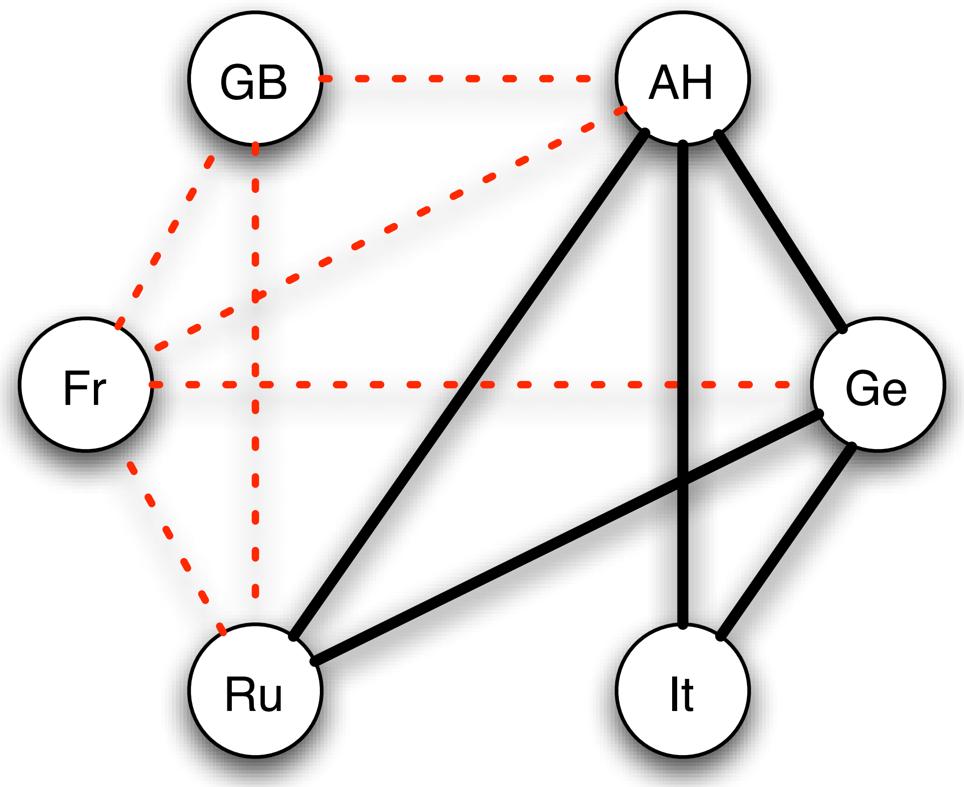
Less straight-forward

Every complete graph that's balanced can be partitioned into two friendly coalitions that dislike either other

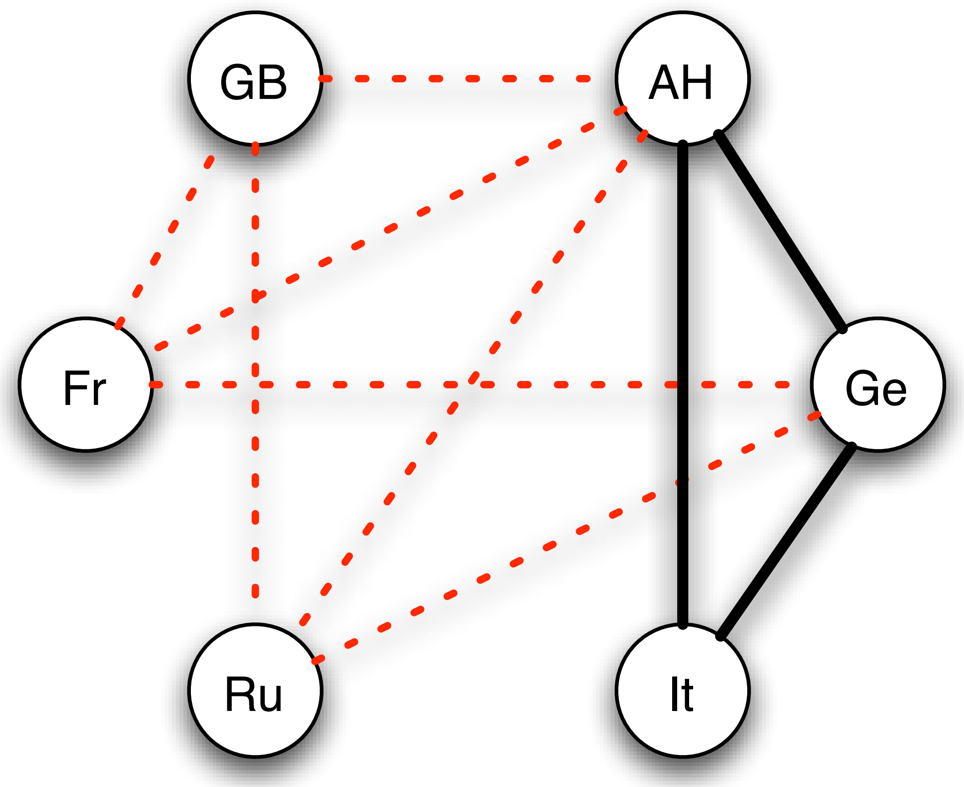
European alliances, pre-WWI



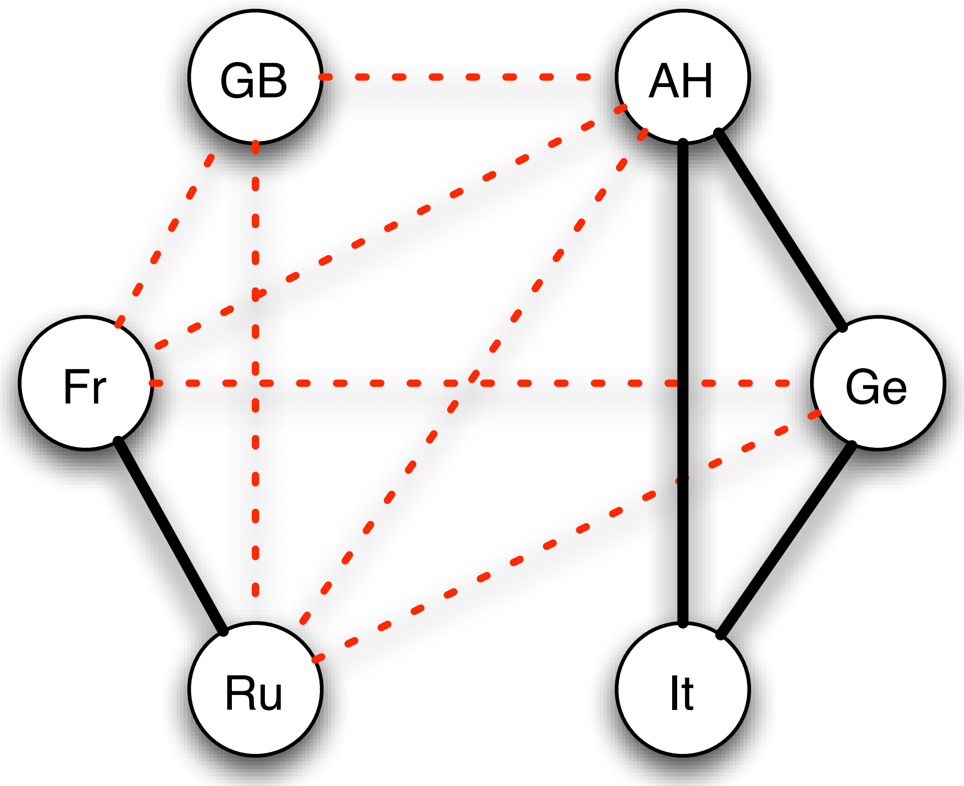
(a) *Three Emperors' League 1872–81*



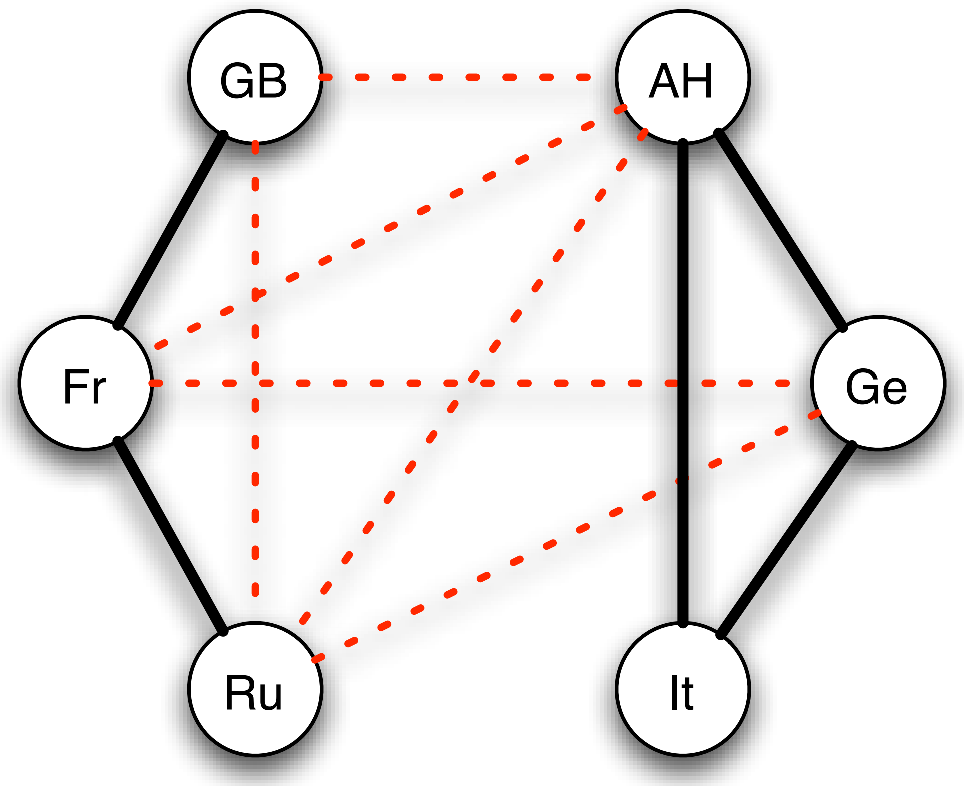
(b) *Triple Alliance 1882*



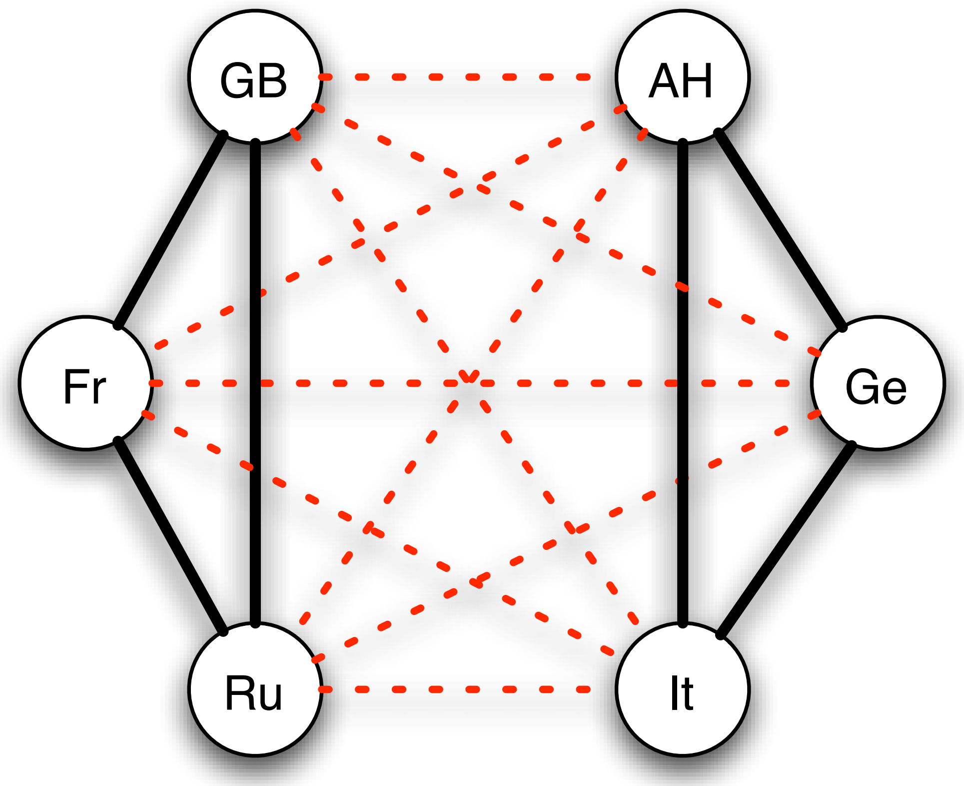
(c) *German-Russian Lapse 1890*



(d) *French-Russian Alliance 1891–94*



(e) *Entente Cordiale 1904*



(f) *British Russian Alliance 1907*

Example: International Relations

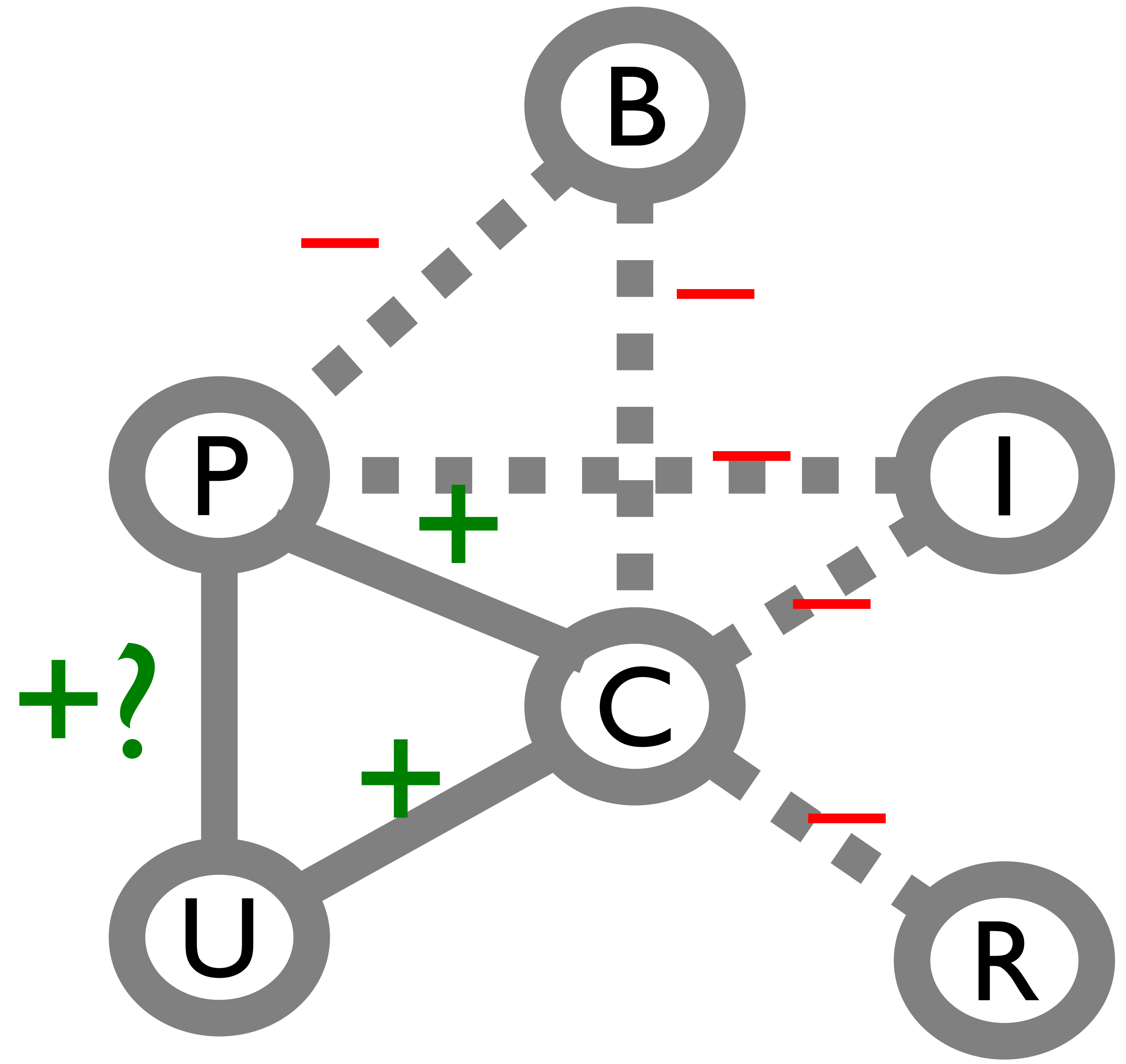
International relations:

Positive edge: alliance

Negative edge: animosity

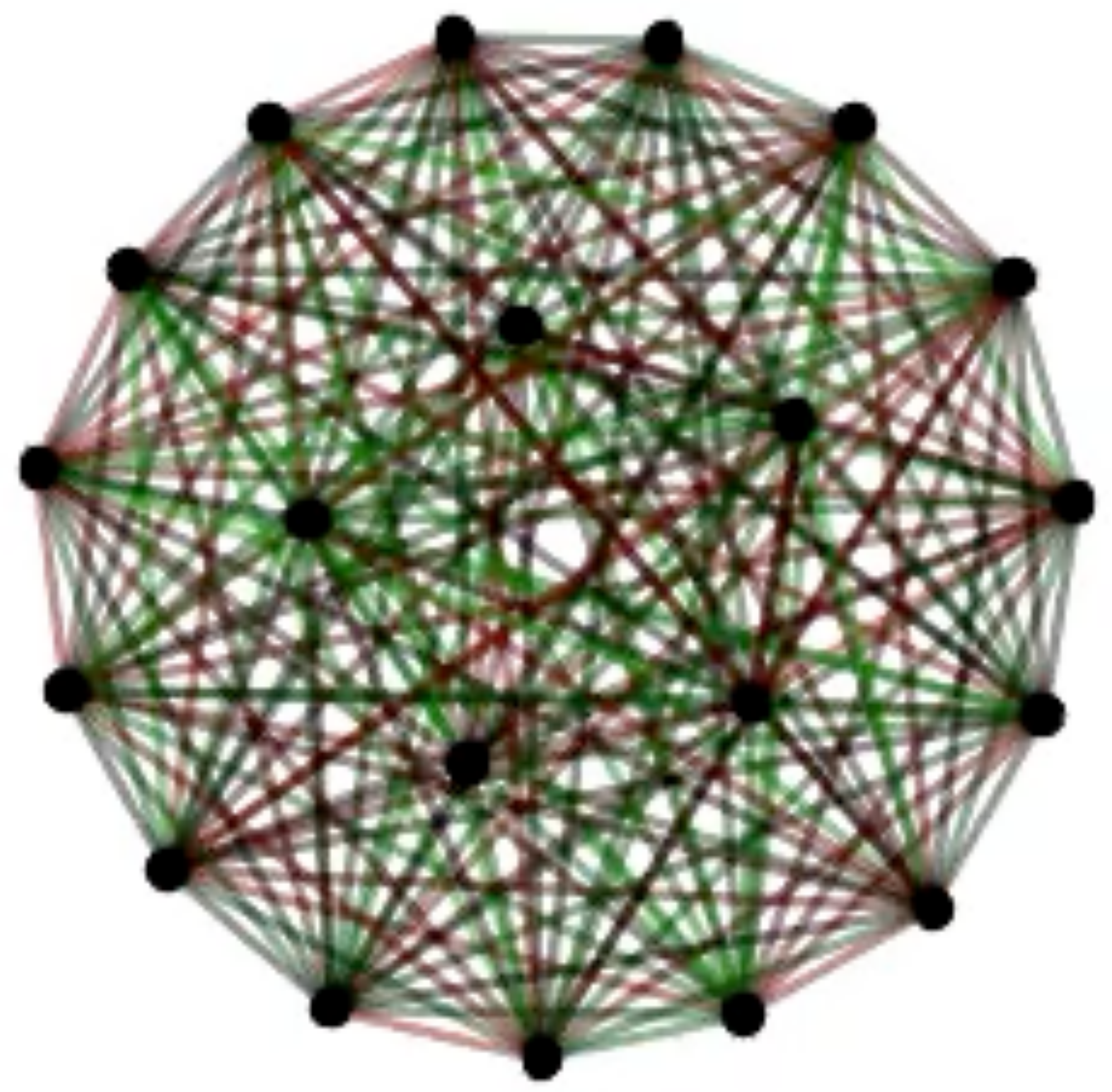
Separation of Bangladesh from Pakistan in 1971:
US supports **P**akistan. **Why?**

- USS**R** was the enemy of **C**hina
- **C**hina was the enemy of **I**ndia
- **I**ndia was the enemy of **P**akistan
- **U**S was friendly with **C**hina
- **C**hina vetoed **B**angladesh from U.N.

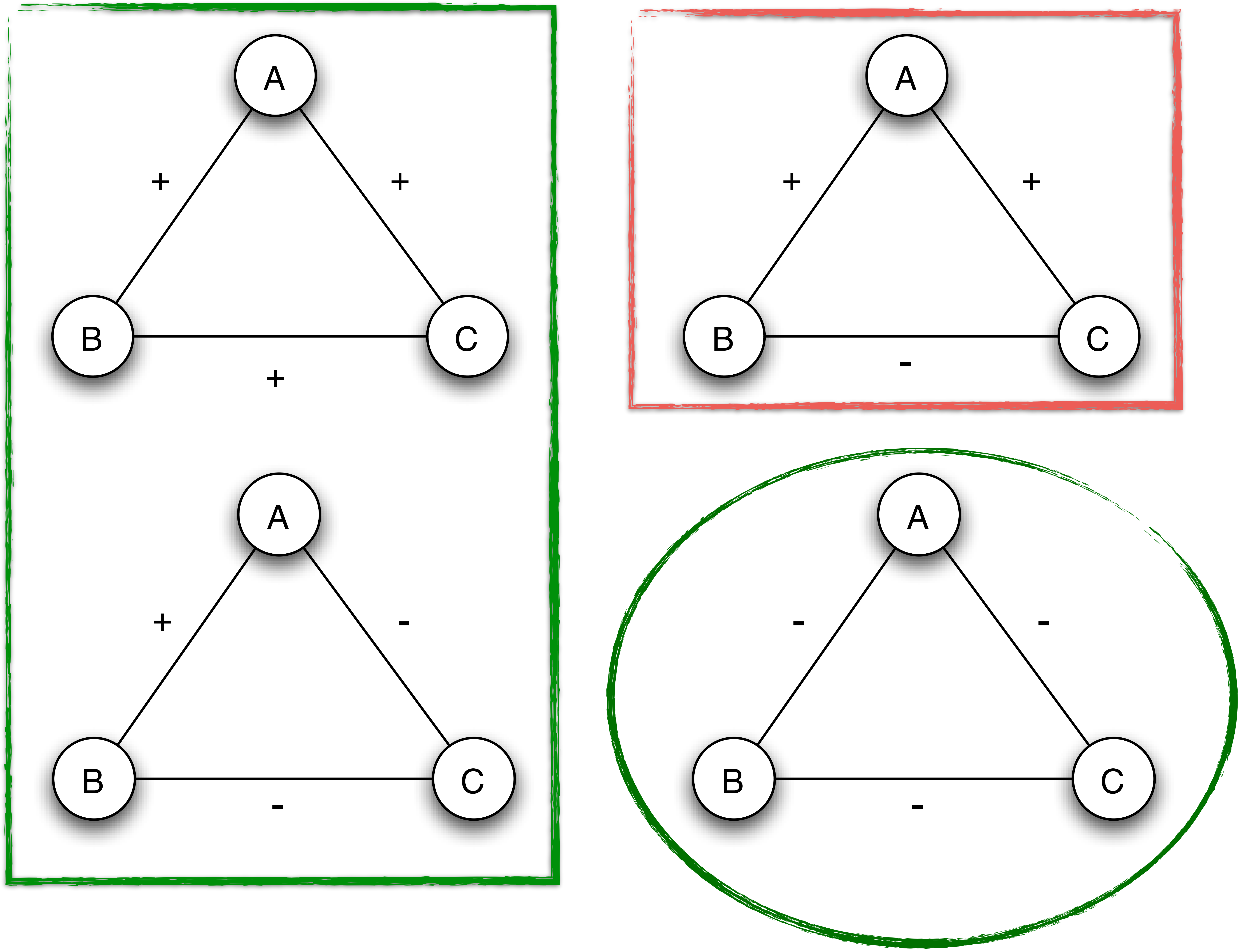


Dynamic Model of Structural Balance

In a simple model of edge evolution in signed networks,
all end states are balanced [Marvel et al., PNAS 2011]



Structural Balance



What if we allow three mutual enemies?

Weak Structural Balance → Many Global Factions

Define: A complete network is *weakly balanced* if there is no triangle with exactly 2 positive edges and 1 negative edge.

Characterization of Weakly Balanced Networks:
If a labeled complete graph is weakly balanced, then its nodes can be **partitioned**

(divided into groups such that two nodes belonging to the same group are friends, and every two nodes belonging to different groups are enemies)

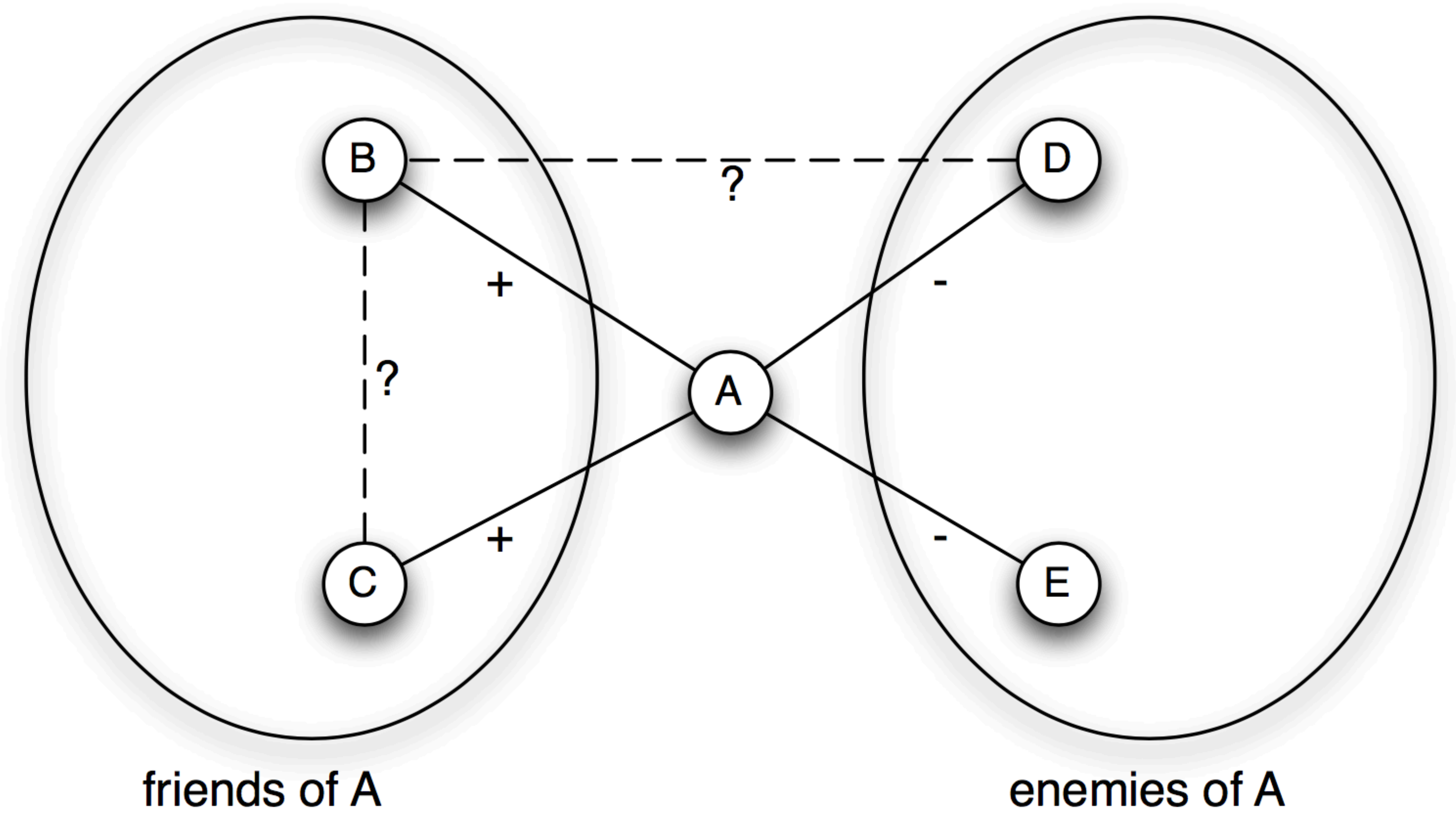
Global picture: same thing as before, but with many factions, not necessarily two

Proof of Characterization

Pick a node **A**.

Because it's a complete graph, **A** is either friends or enemies with each person.

Now check 2 cases:

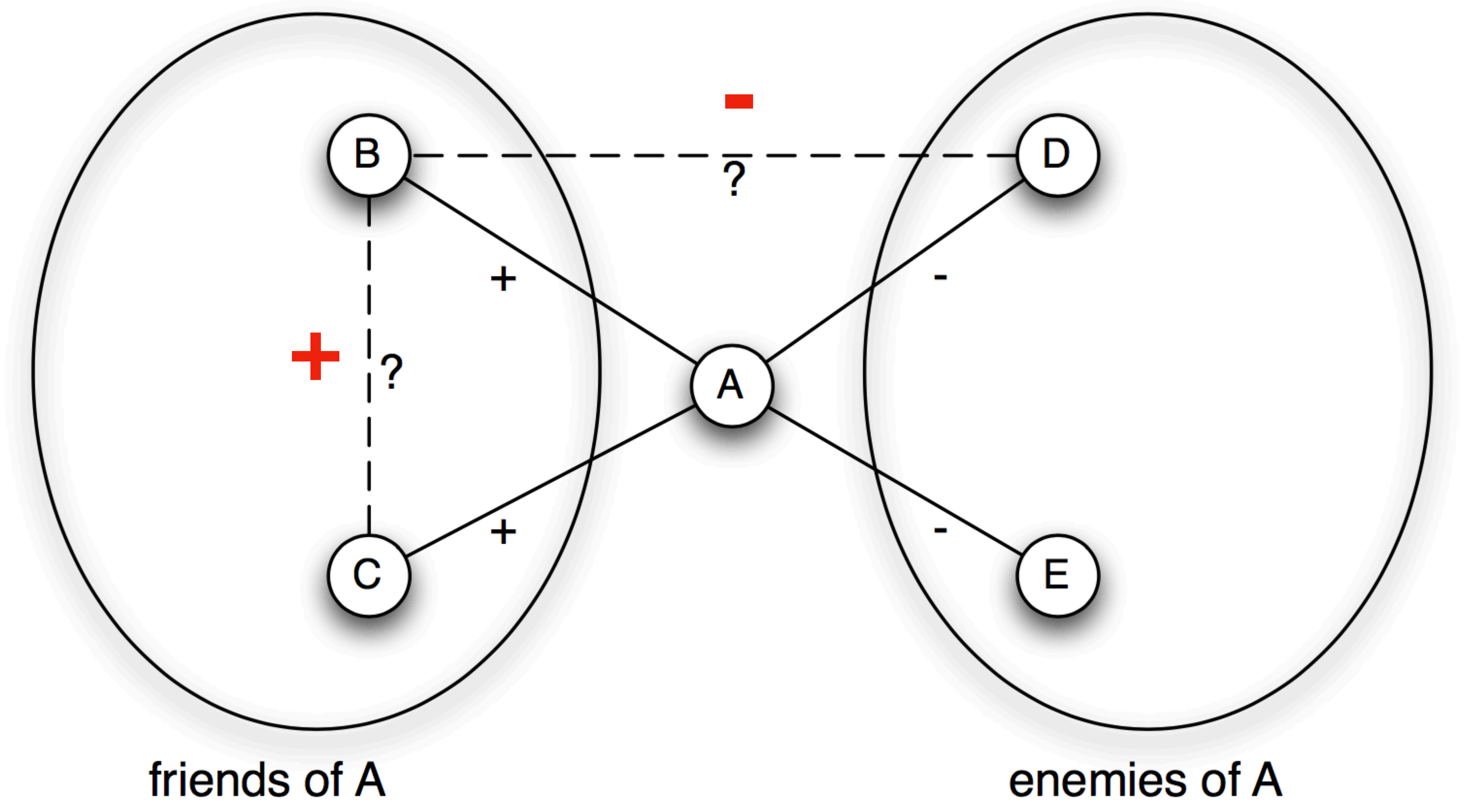


Proof of Characterization

All of A's friends are friends with each other and are enemies with all of A's enemies

Remove A and his friends from the graph and **recurse**!

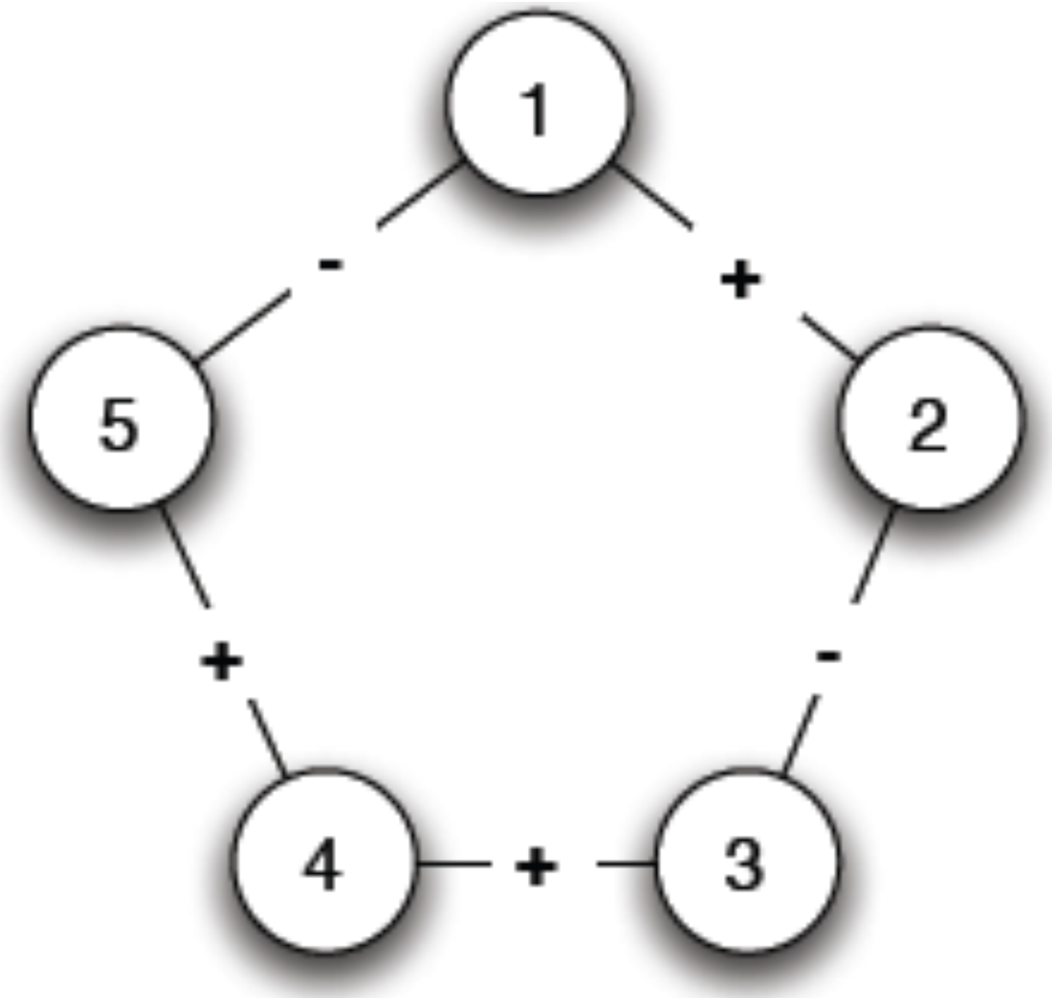
Graph still weakly balanced, find a second group, same argument applies, recurse until we've found all factions



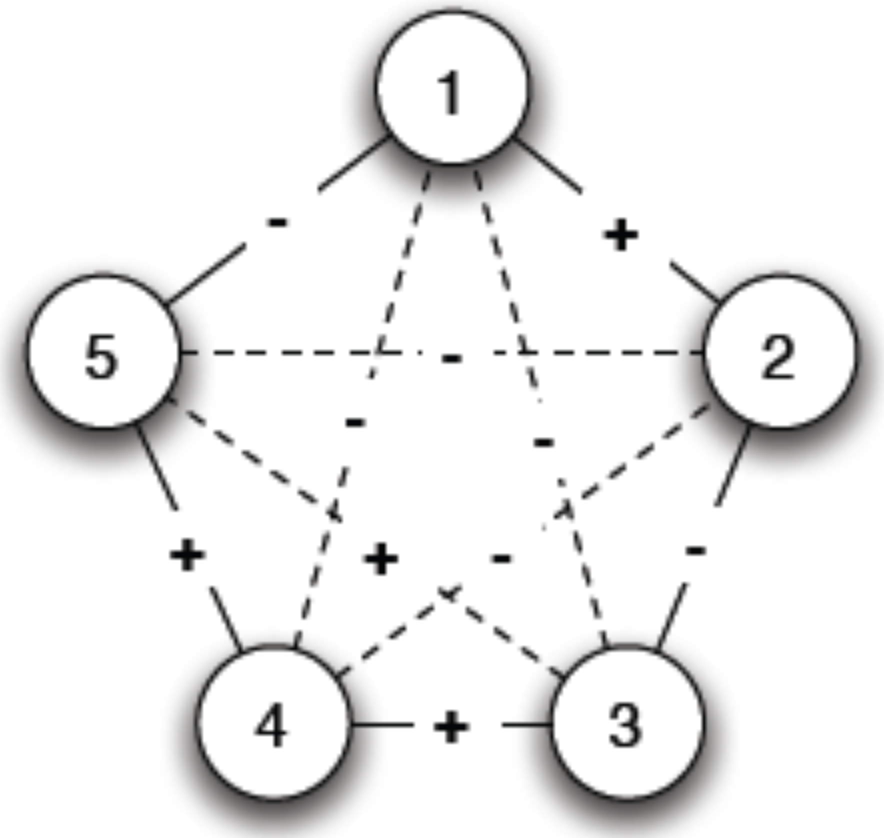
Balance in General Networks

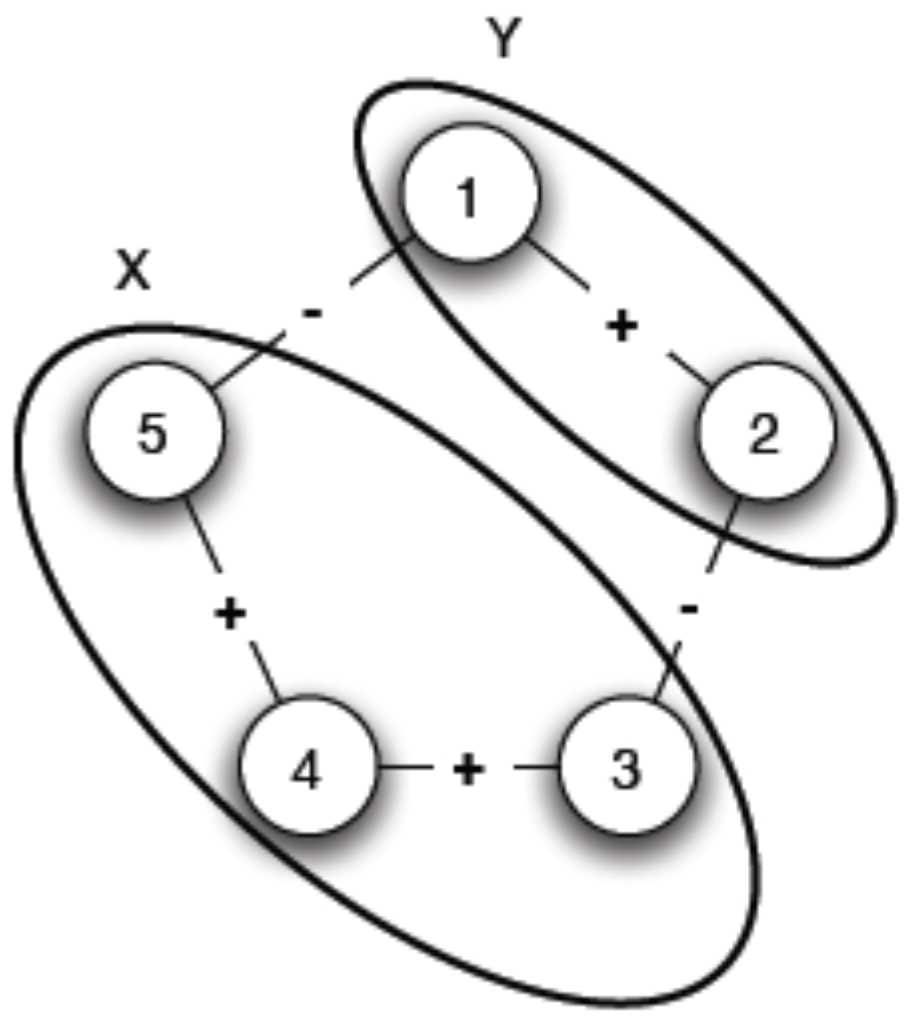
So far we've talked about complete graphs

What about incomplete graphs?

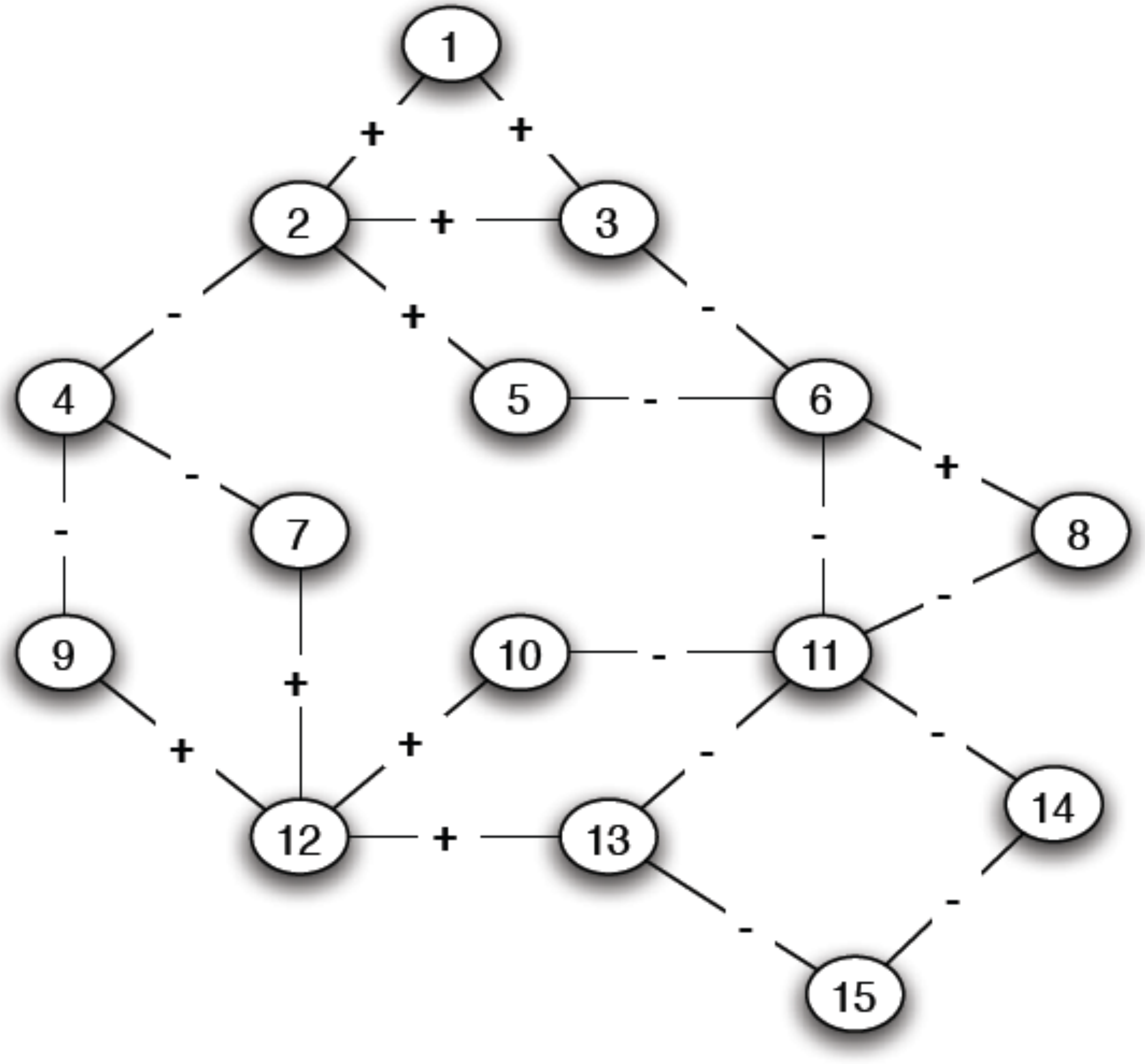


Balanced?





Signed Graph: Is it Balanced?

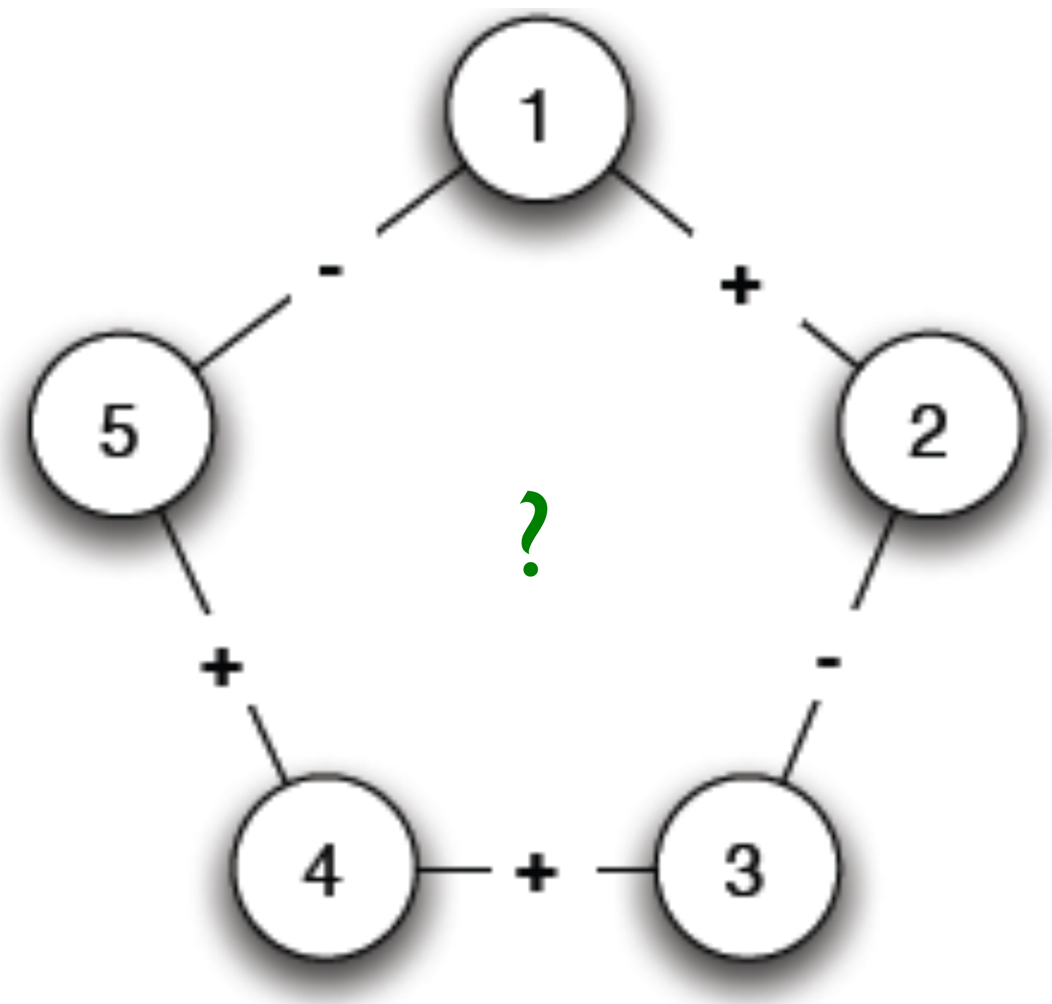


Balance in General Networks

So far we talked about complete graphs

Def 1: Local view

Fill in the missing edges to achieve balance

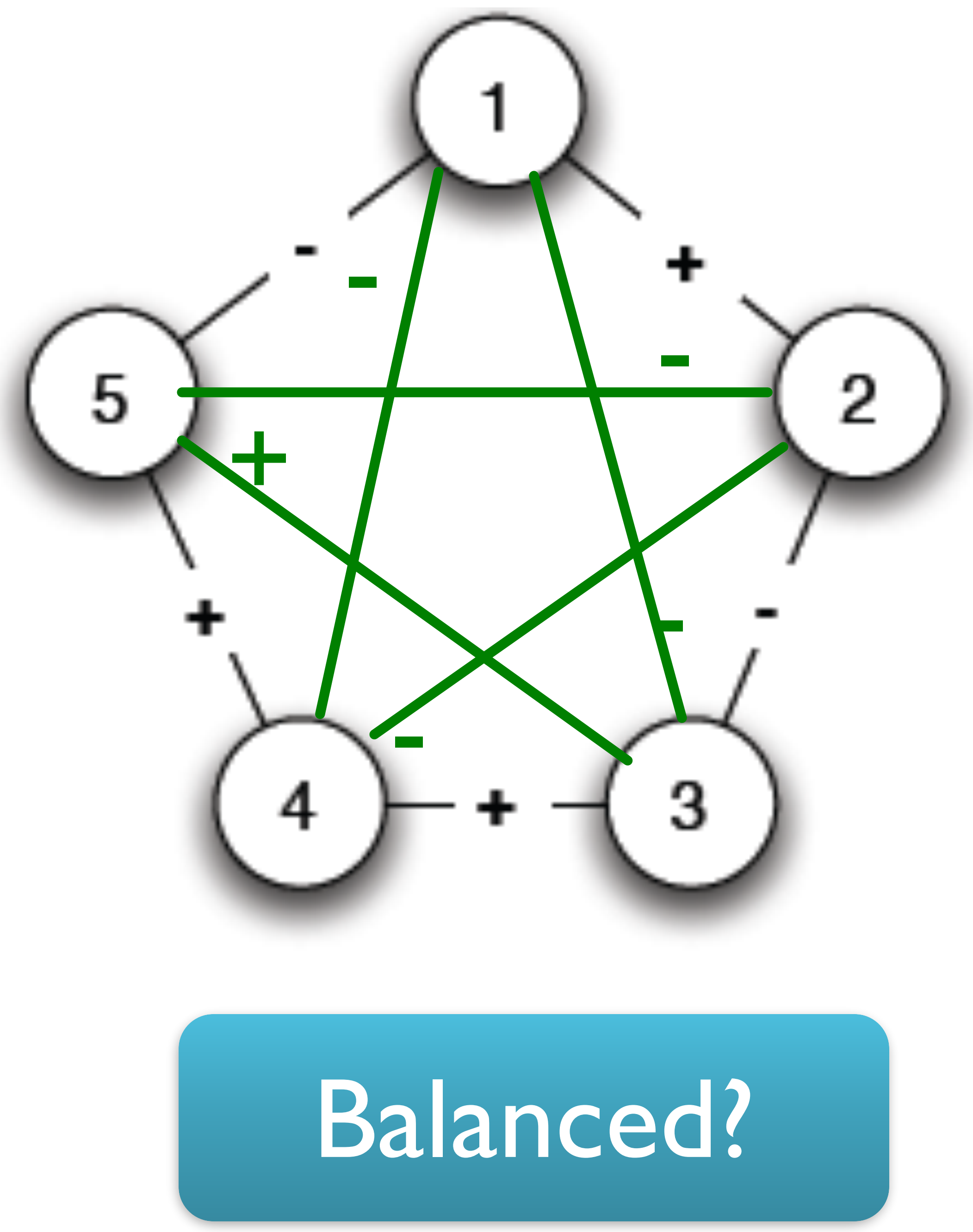


Balanced?

If the graph is "Balance-able", then call it balanced

Balance in General Networks

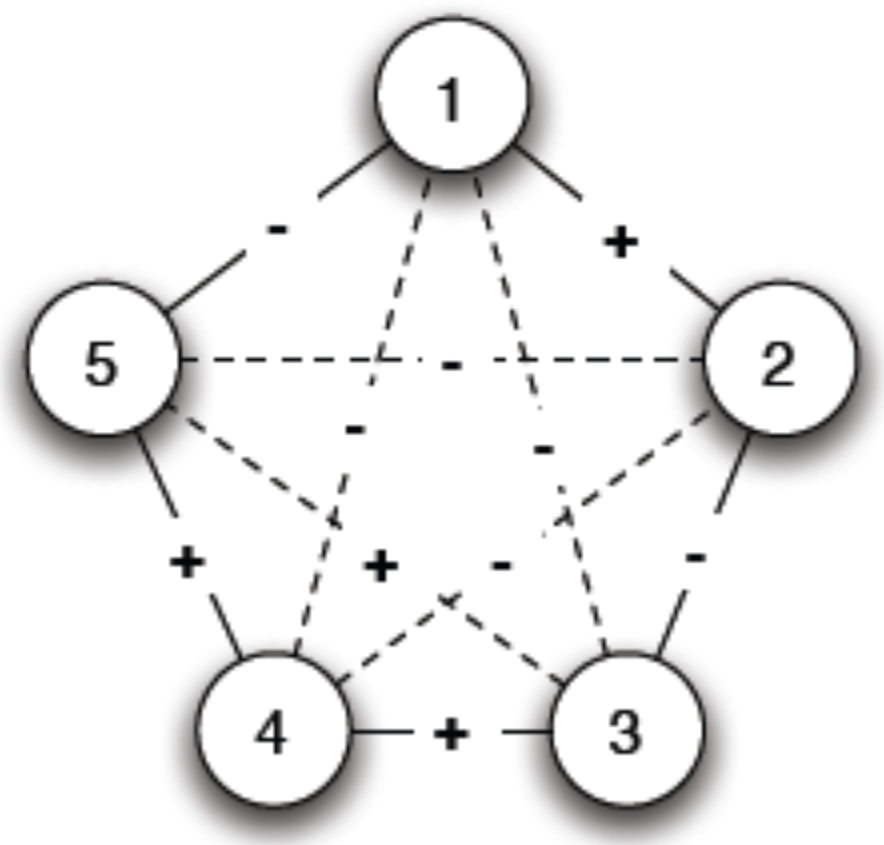
So far we talked about complete graphs



Def 1: Local view

Fill in the missing edges to achieve balance

If the graph is "Balance-able", then call it balanced



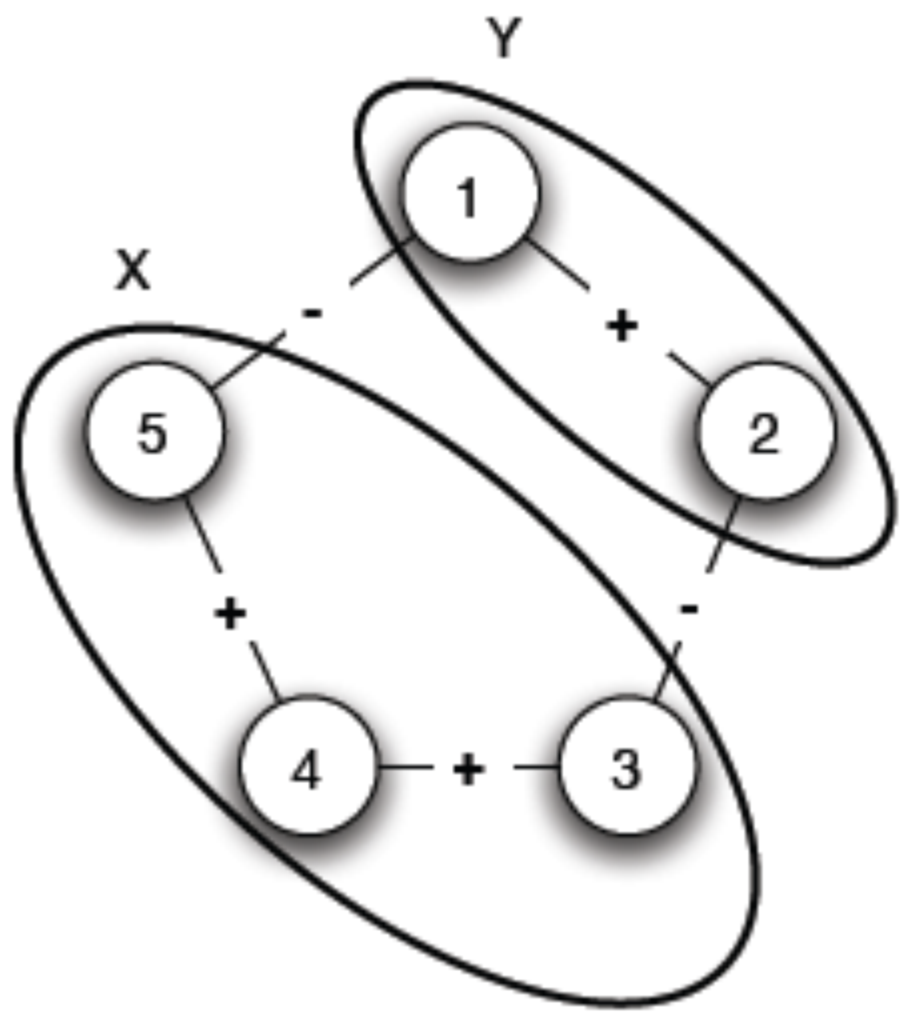
Balance in General Networks

So far we talked about complete graphs

Def 2: Global view

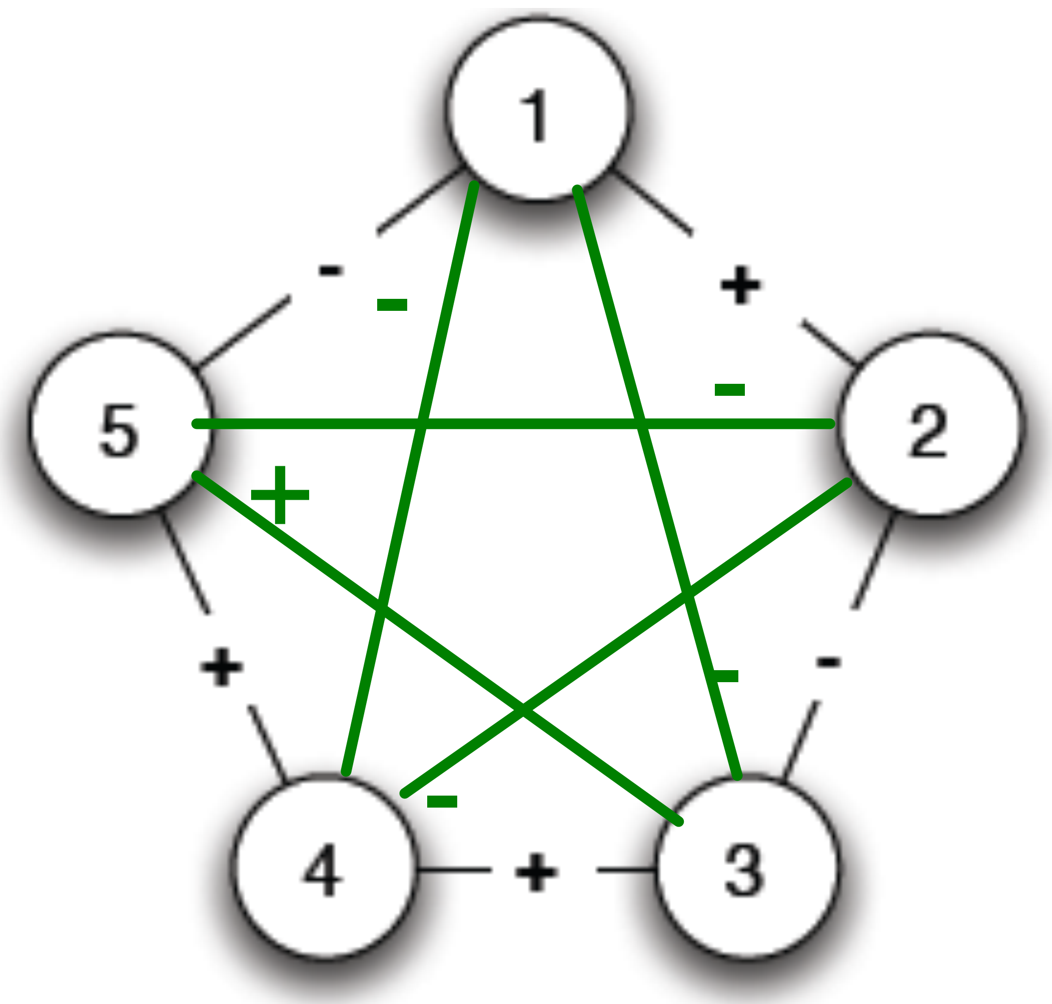
Divide the graph into two coalitions

If you can separate the graph into coalitions as before, call it balanced



Balance in General Networks

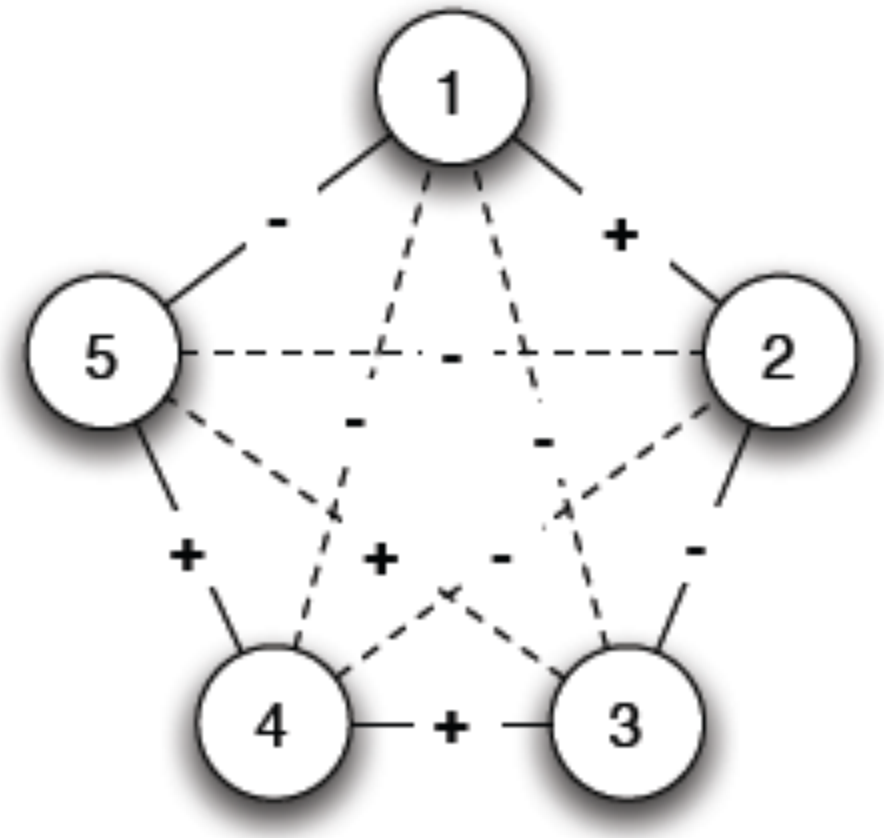
So far we talked about complete graphs



Balanced?

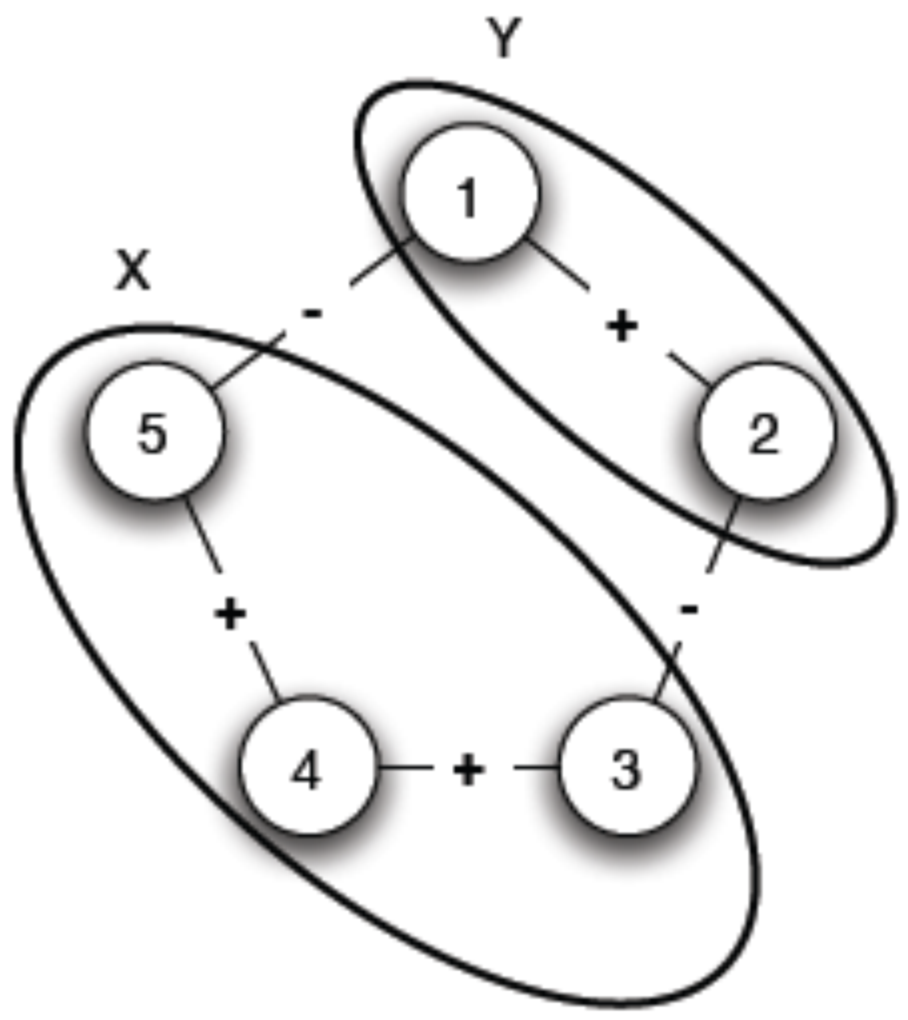
Def 1: Local view

Fill in the missing edges to achieve balance



Def 2: Global view

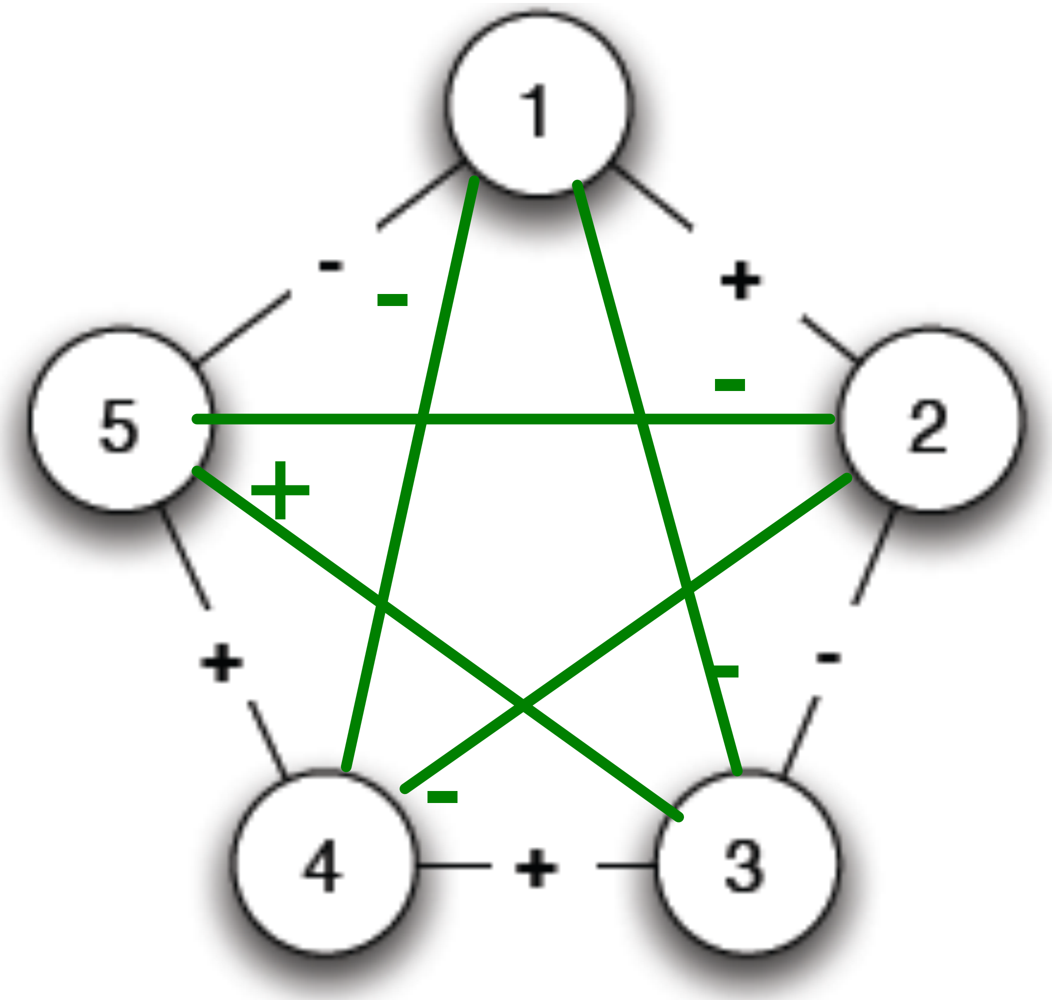
Divide the graph into two coalitions



The 2 definitions are equivalent!

Balance in General Networks

<u>Claim</u>: in general (not necessarily complete) networks, the local and global definitions of balance are equivalent

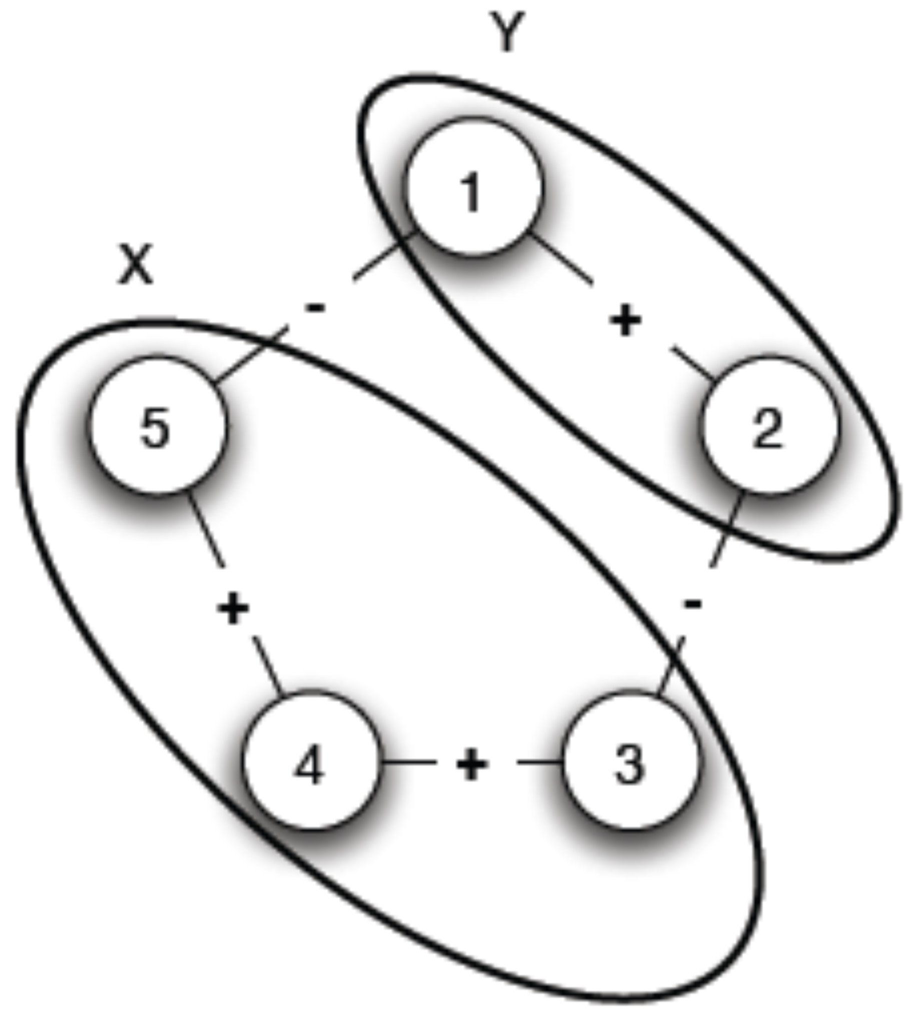


Def 1: Local view

Fill in the missing edges to achieve balance

Def 2: Global view

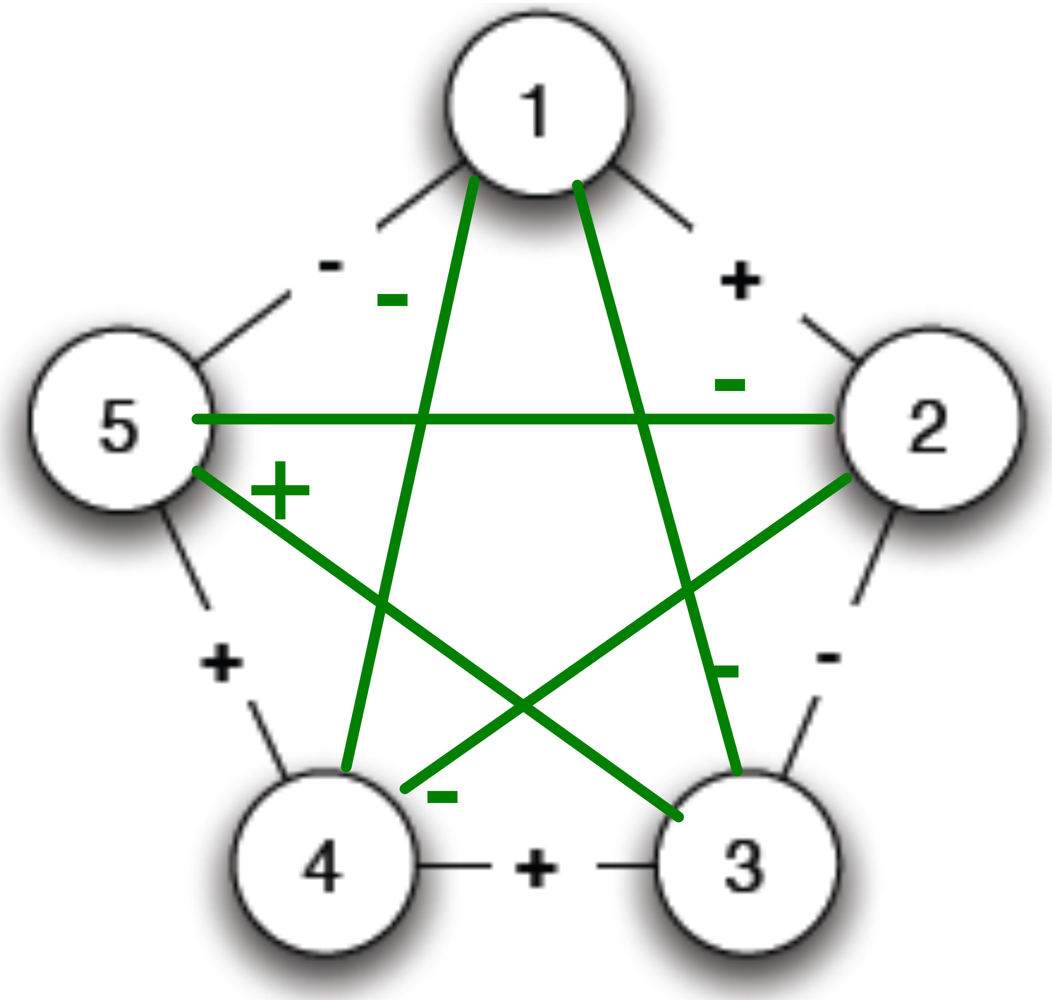
Divide the graph into two coalitions



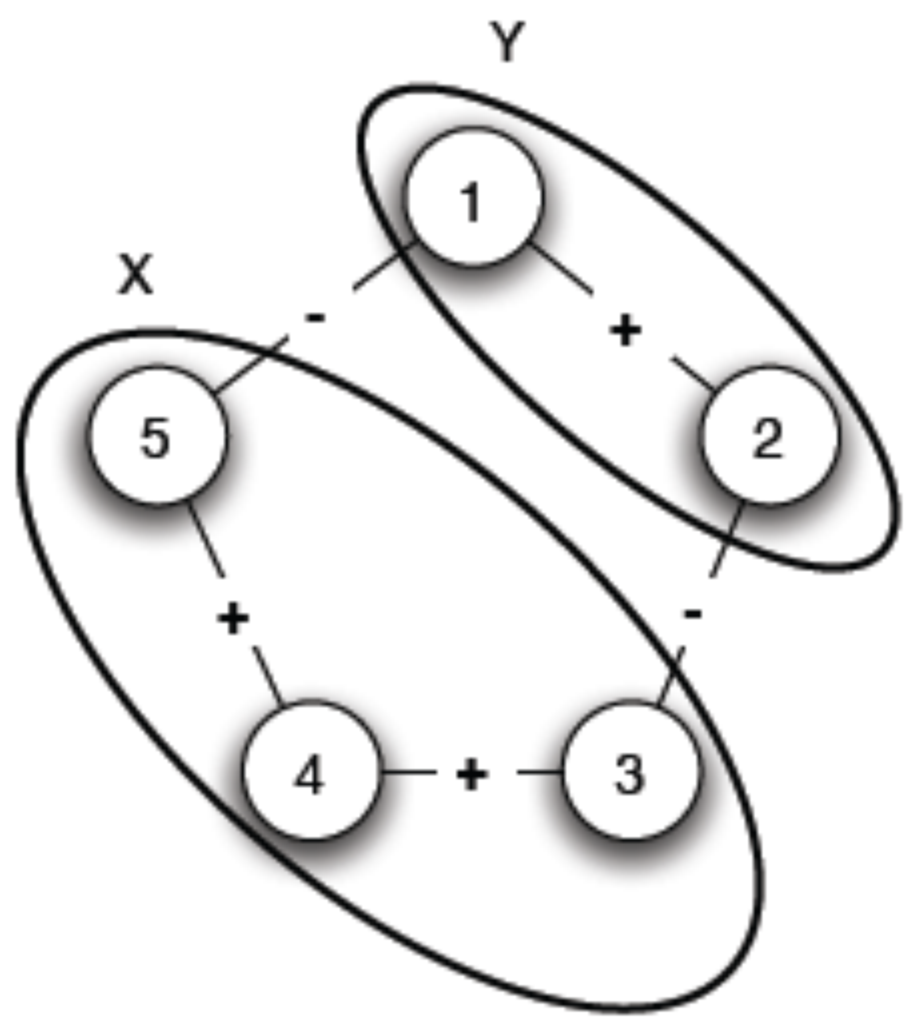
Balance in General Networks

Actually easy to see:

Local => global: (if you can fill in edges such that the resulting complete graph is balanced, then it can be divided into coalitions)



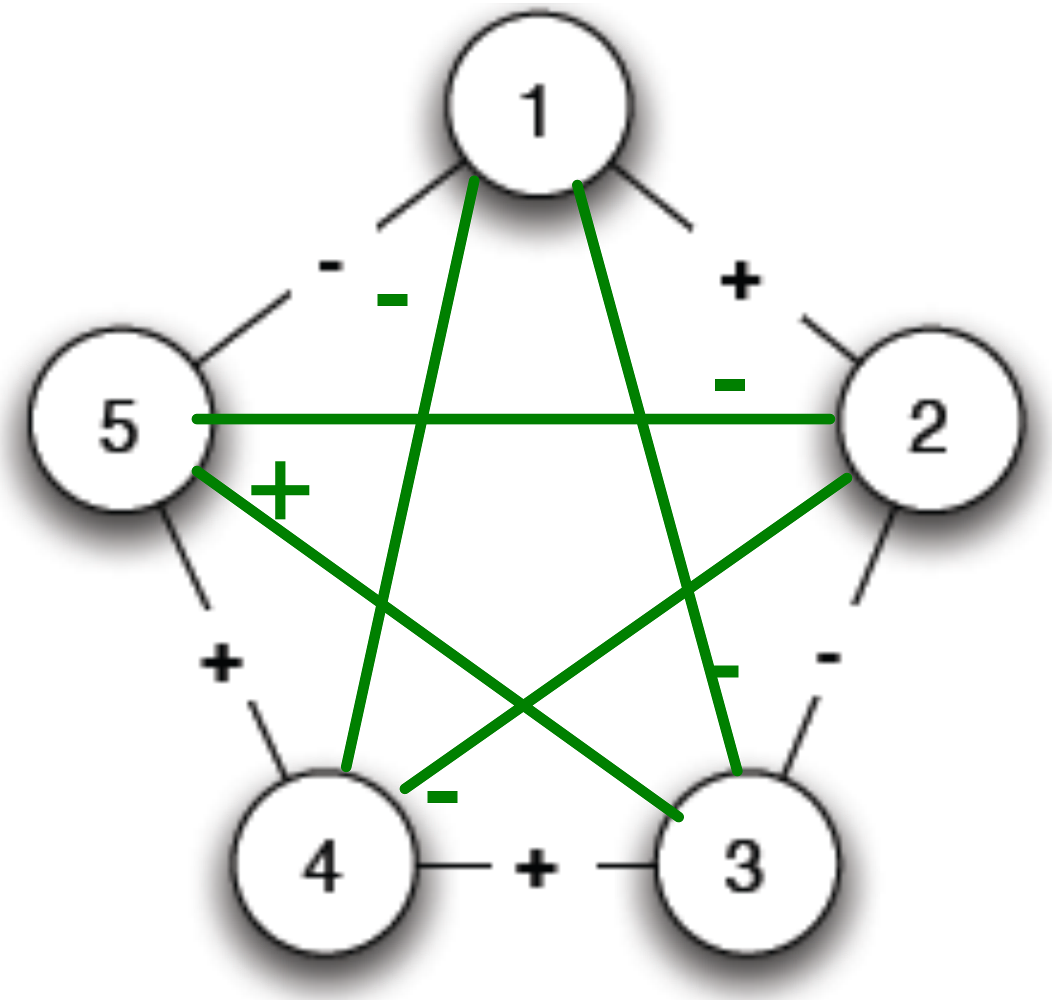
After filling in, we have a complete network as before, the Balance Theorem applies



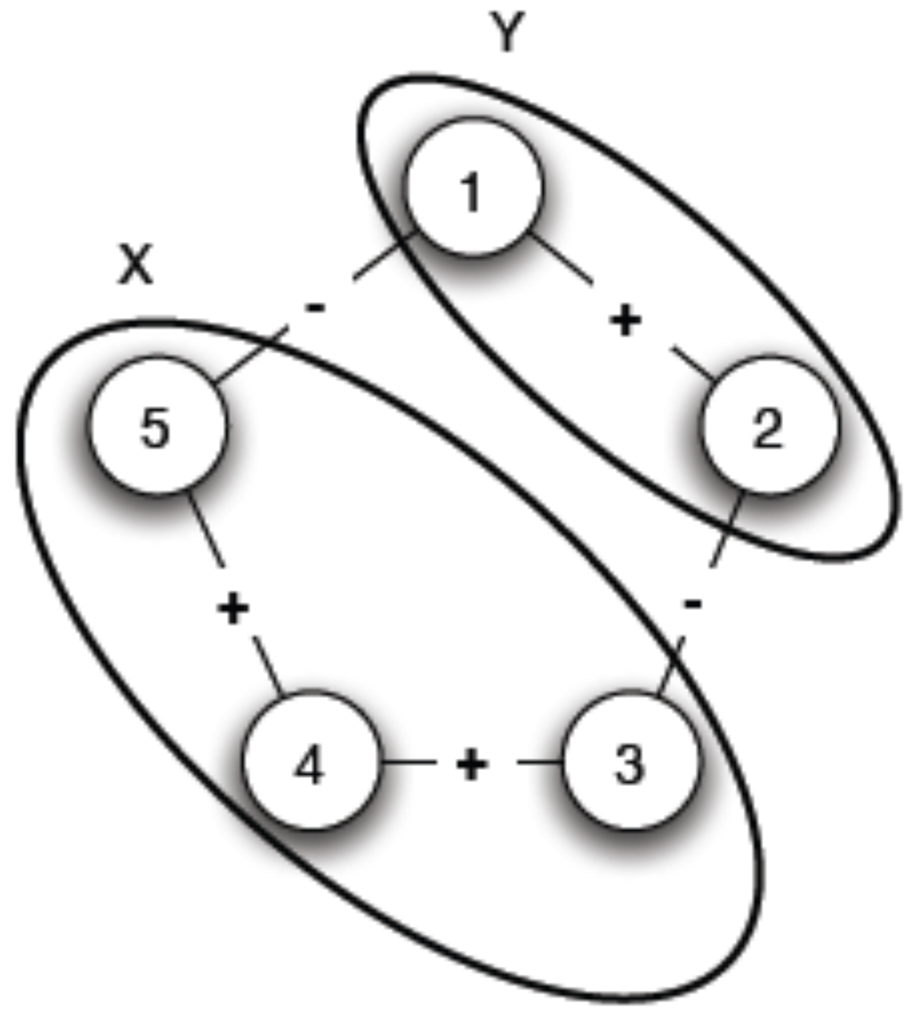
Balance in General Networks

Actually easy to see:

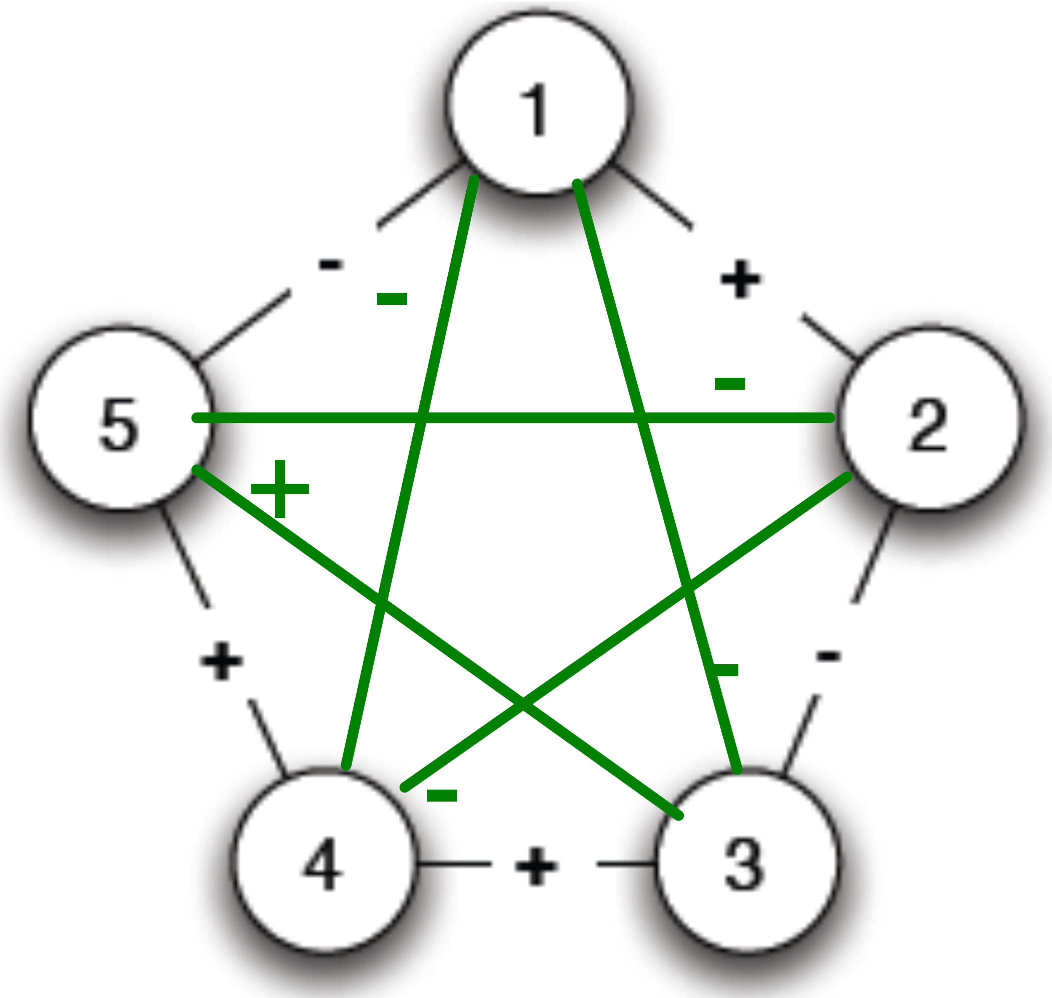
Global => local: (if the graph can be divided into coalitions, then you can fill in edges that results in a complete balanced graph)



Fill in edges within and between coalitions as before: positive edges within the coalitions and negative edges between them



Balance in General Networks

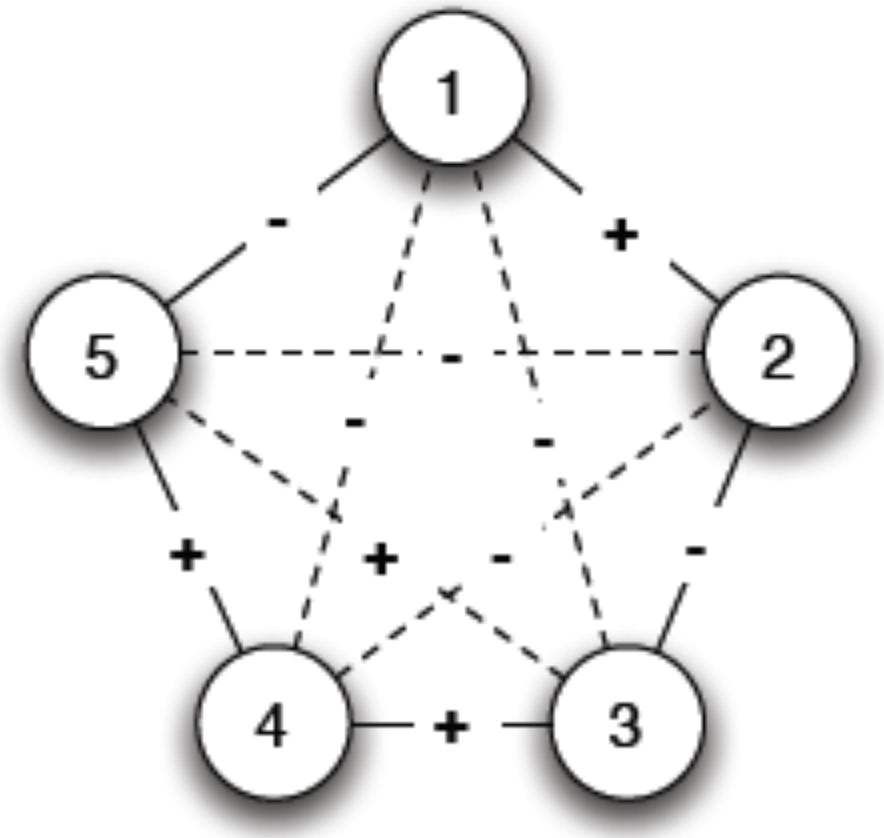


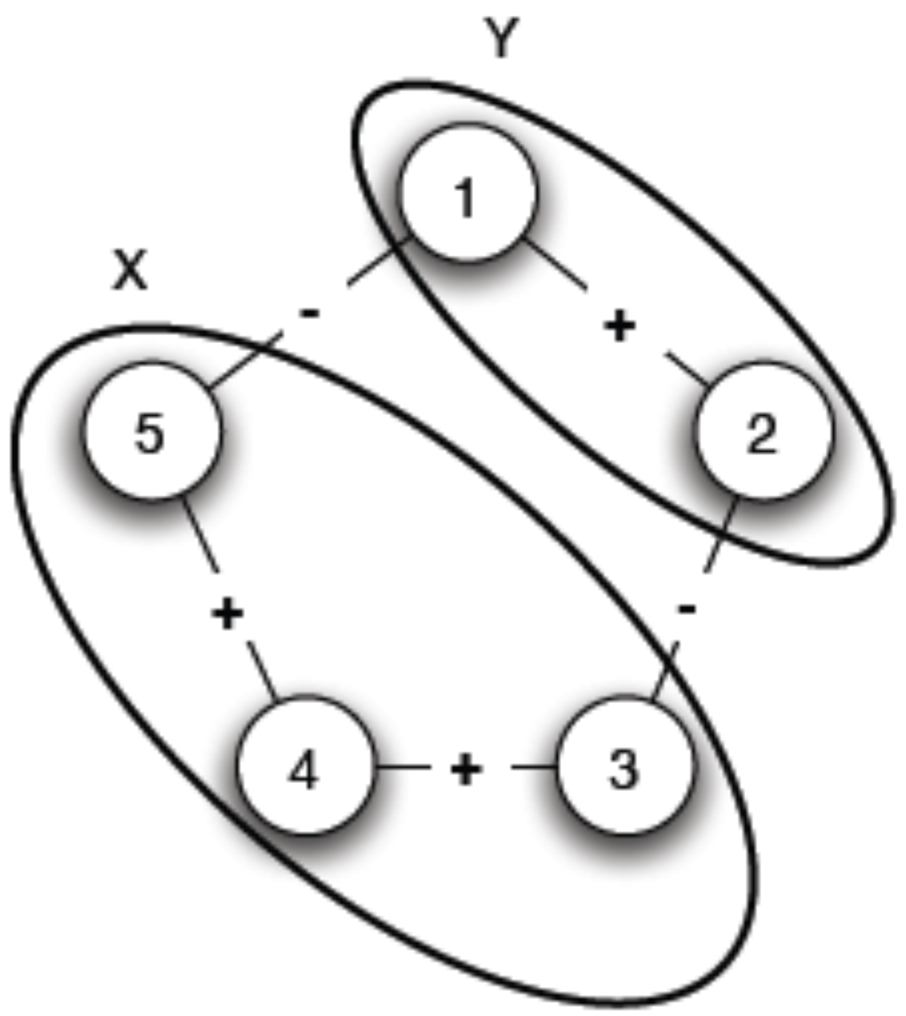
Actually easy to see:

Local => global: after filling in, result in complete network as before

Global => local: fill in edges within and between coalitions as before

Done!

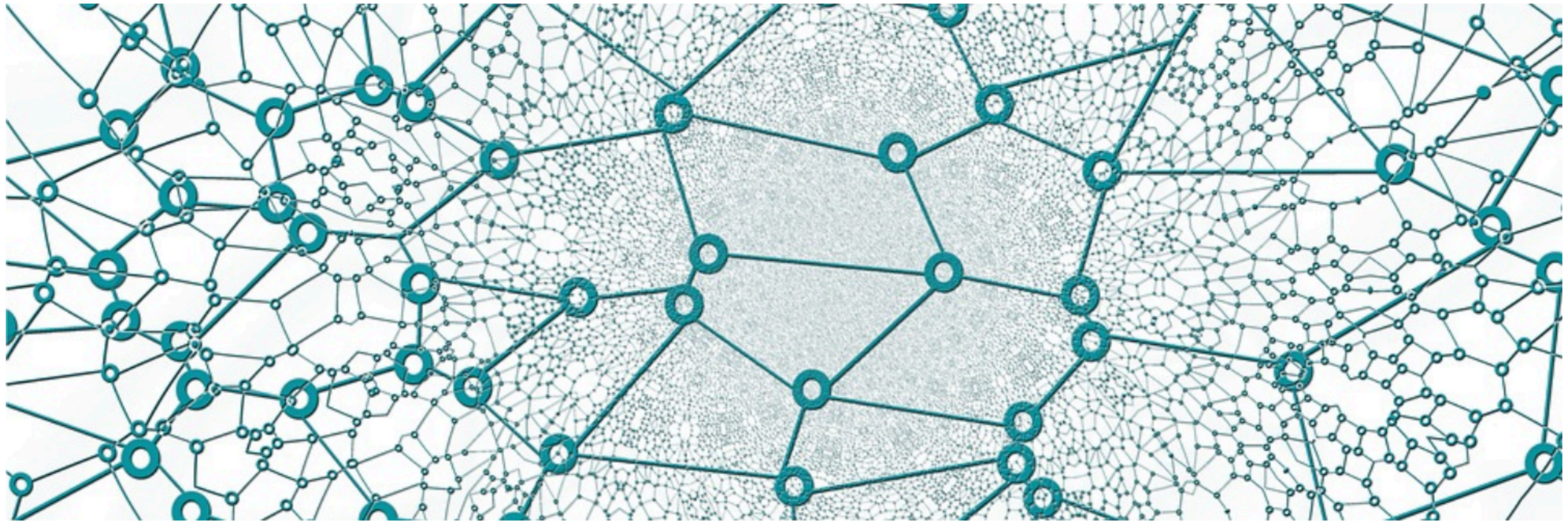




Balance in General Networks

We have a natural definition for **balance** in general signed networks

"**Natural**" because we arrived at it **two different ways** that **turn out to be equivalent**

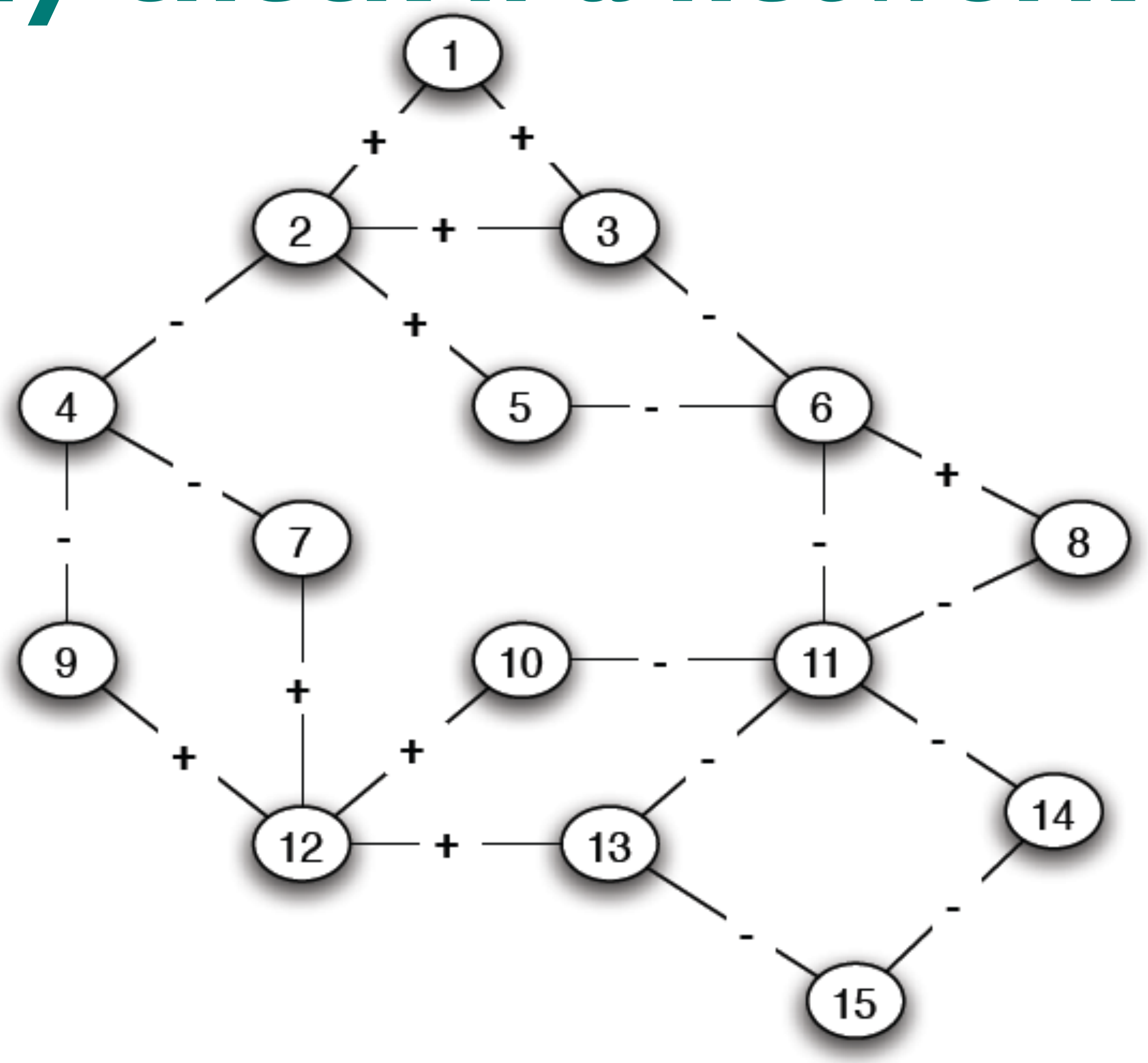


Balance in General Networks

We have a natural definition for **balance** in general signed networks

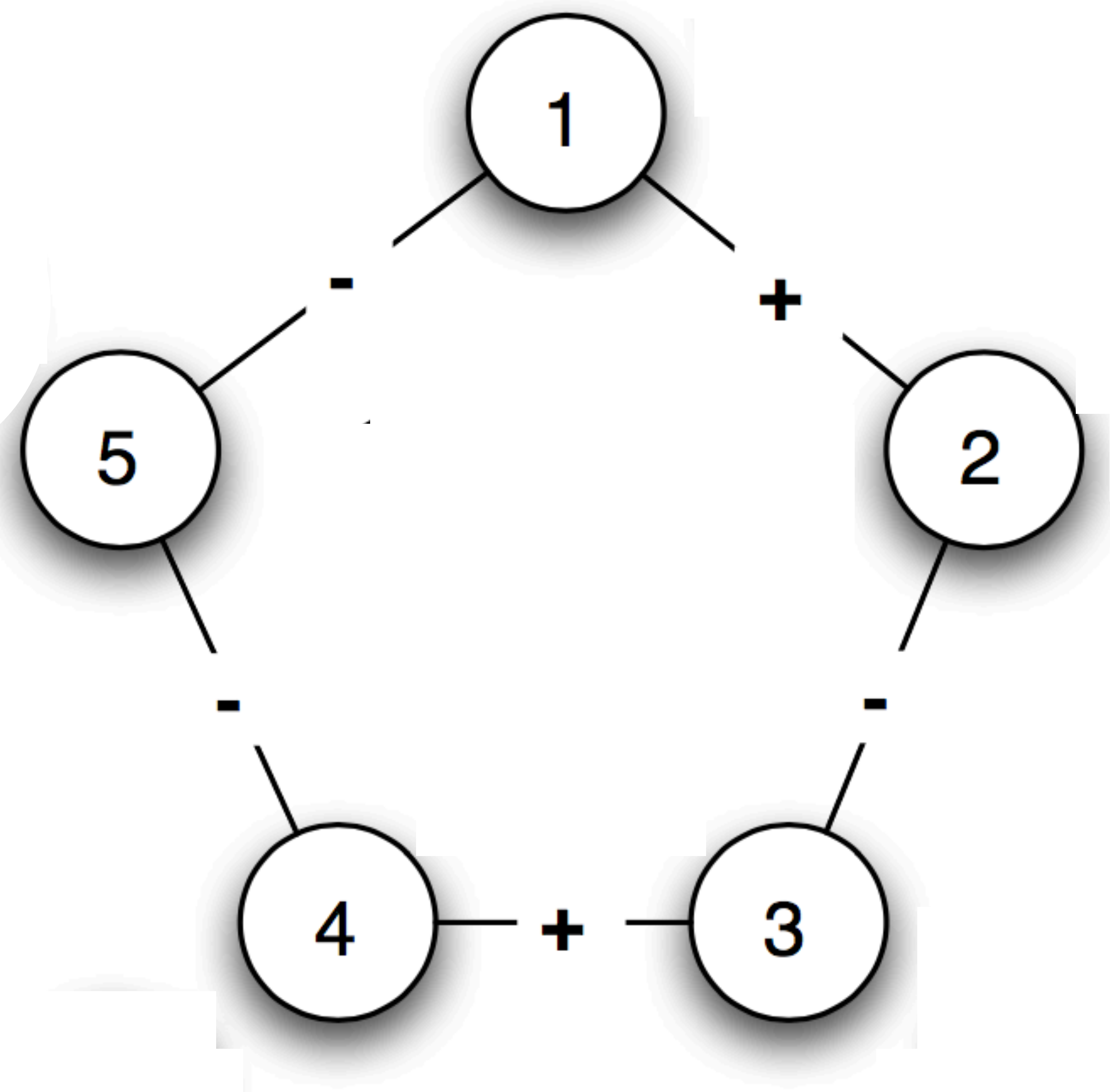
"**Natural**" because we arrived at it **two different ways** that **turn out to be equivalent**

But, there's a problem: **how to actually check if a network is balanced in this way?**



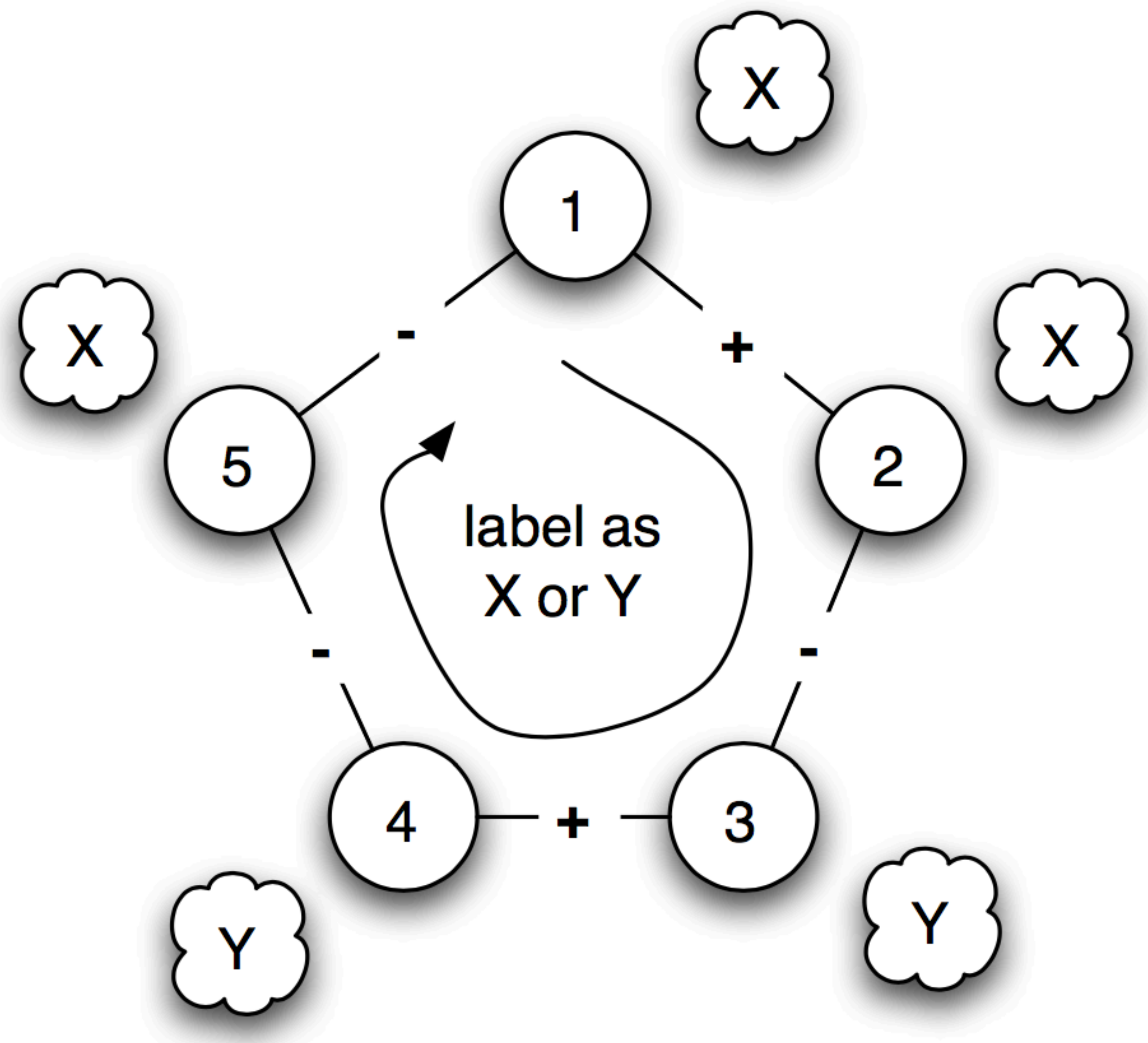
Balance in General Networks

Why isn't this graph balanced?



Balance in General Networks

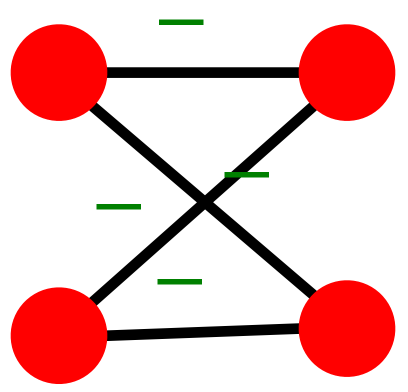
Why isn't this graph balanced?



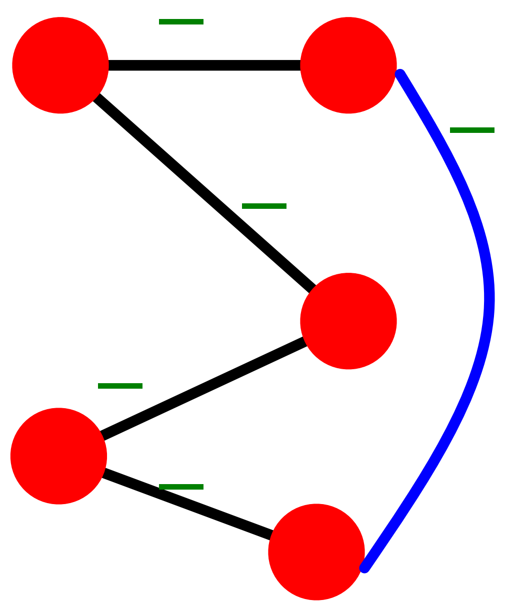
Walk around a cycle, every time we see a negative edge we have to switch coalitions

Is a Signed Network Balanced?

Theorem: Graph is **balanced** if and only if it contains **no cycle with an odd number of negative** edges
[Harary 1953, 1956]



Even length cycle

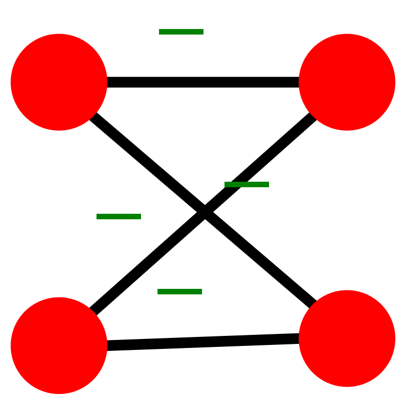


Odd length cycle

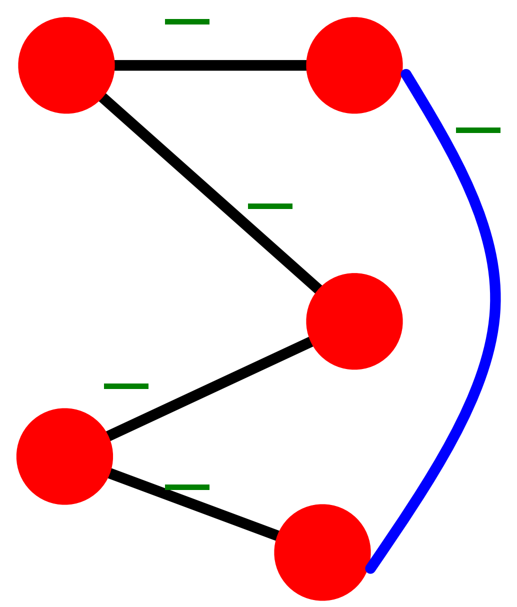
Is a Signed Network Balanced?

Theorem: Graph is **balanced** if and only if it contains **no cycle with an odd number of negative** edges
[Harary 1953, 1956]

This theorem is saying that the only way a graph can be unbalanced is if there is a cycle with an odd number of negative cycles. That's the only possible problem!



Even length cycle

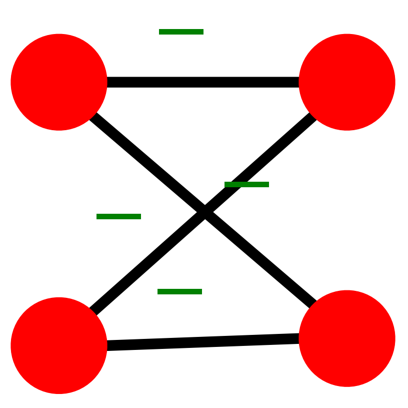


Odd length cycle

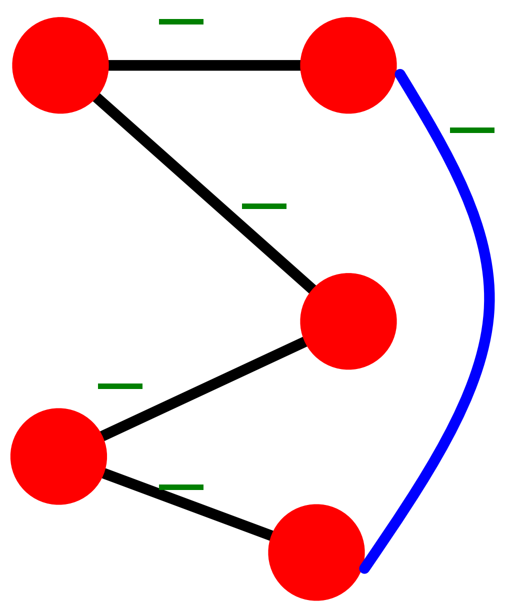
Is a Signed Network Balanced?

Theorem: Graph is **balanced** if and only if it contains **no cycle with an odd number of negative** edges [Harary 1953, 1956]

Proof: We will show that every graph is either balanced or contains a cycle with odd number of negative edges (i.e. a *constructive* proof).



Even length cycle

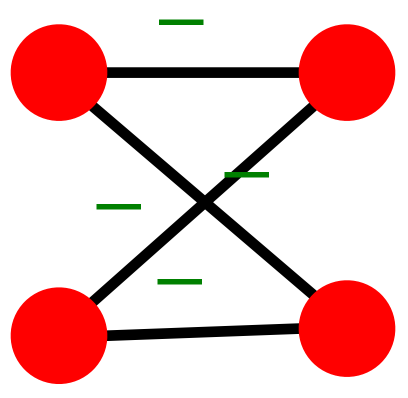


Odd length cycle

Is a Signed Network Balanced?

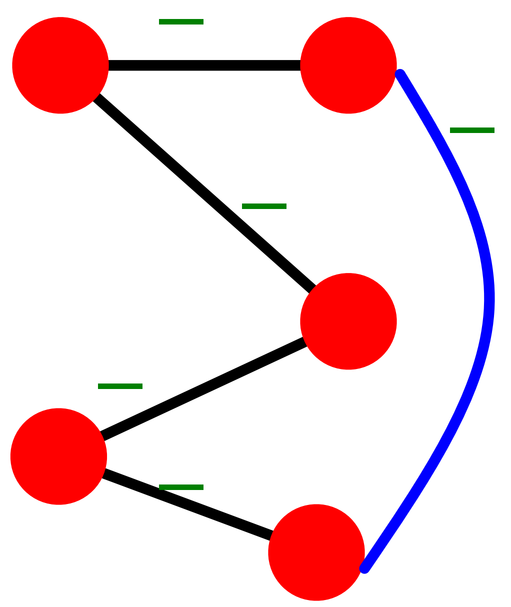
Theorem: Graph is **balanced** if and only if it contains **no cycle with an odd number of negative** edges
[Harary 1953, 1956]

Proof by algorithm: We will do this by actually constructing an algorithm that either outputs a division into coalitions or a cycle with odd number of negative edges



Even length cycle

Because these are the only two outcomes, this **proves the claim**

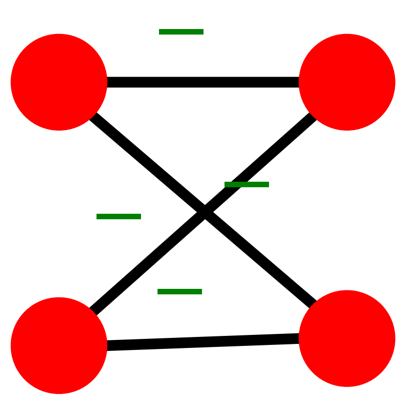


Odd length cycle

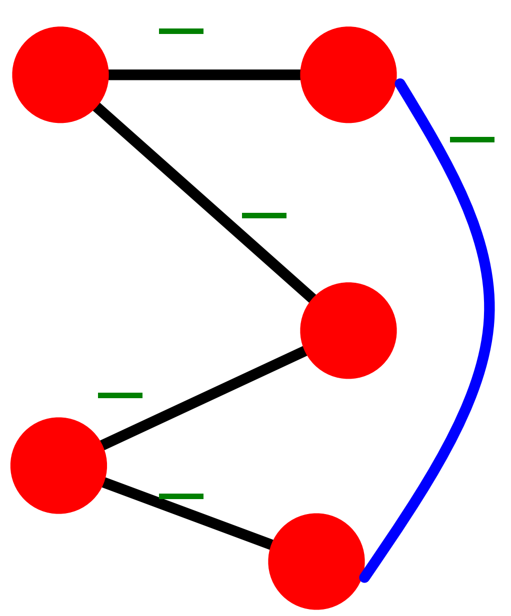
Is a Signed Network Balanced?

Theorem: Graph is **balanced** if and only if it contains **no cycle with an odd number of negative** edges [Harary 1953, 1956]

Proof sketch: Our algorithm will try to assign nodes to coalitions such that the graph is balanced. We will reason that the only way it can fail is if there is a cycle with an odd number of negative edges.



Even length cycle



Odd length cycle

Is a Signed Network Balanced?

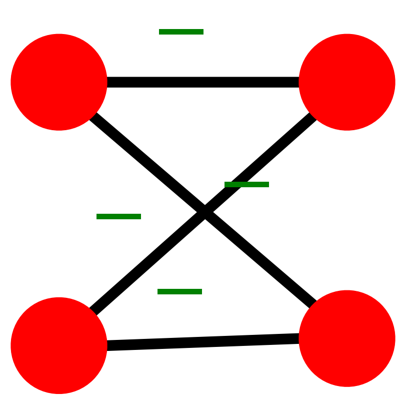
Signed graph algorithm:

Step 1: Find connected components on + edges and for each component create a super-node

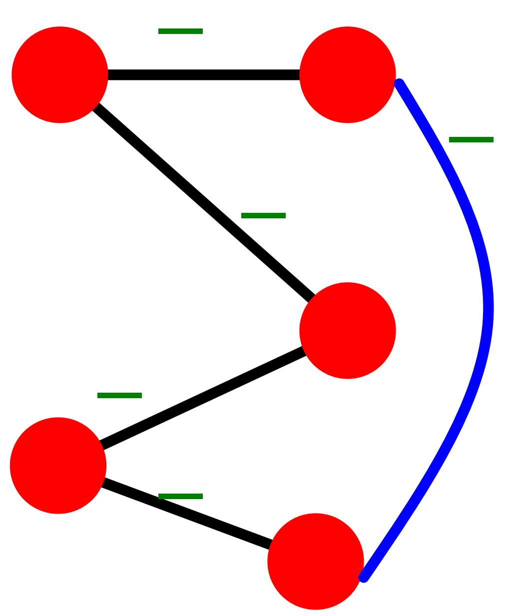
- Since nodes connected by a + edge must be in same coalition
- If any – edge in the super node, done (cycle with 1 negative edge)

Step 2: Connect components A and B if there is a negative edge between the members

- Note there are only negative edges pointing out of a super-node (otherwise should've connected the two super-nodes that have a positive edge)



Even length cycle

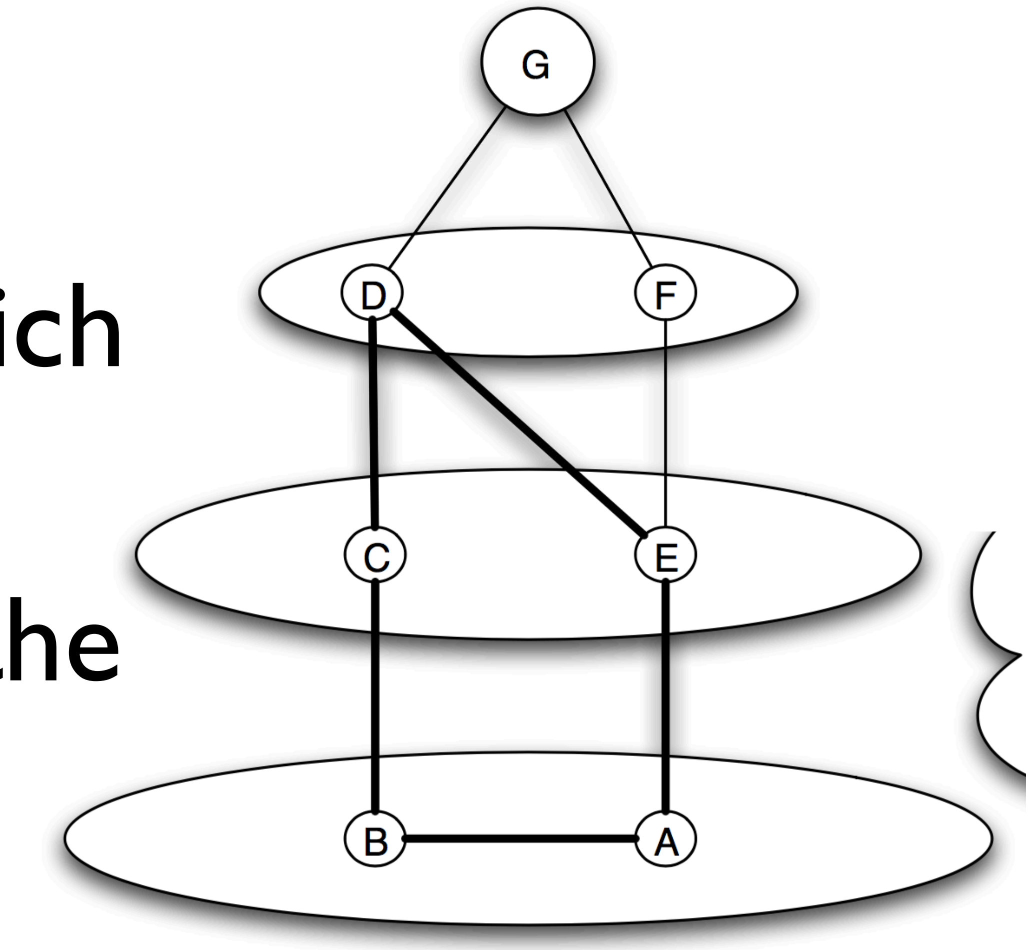


Odd length cycle

Is a Signed Network Balanced?

Signed graph algorithm

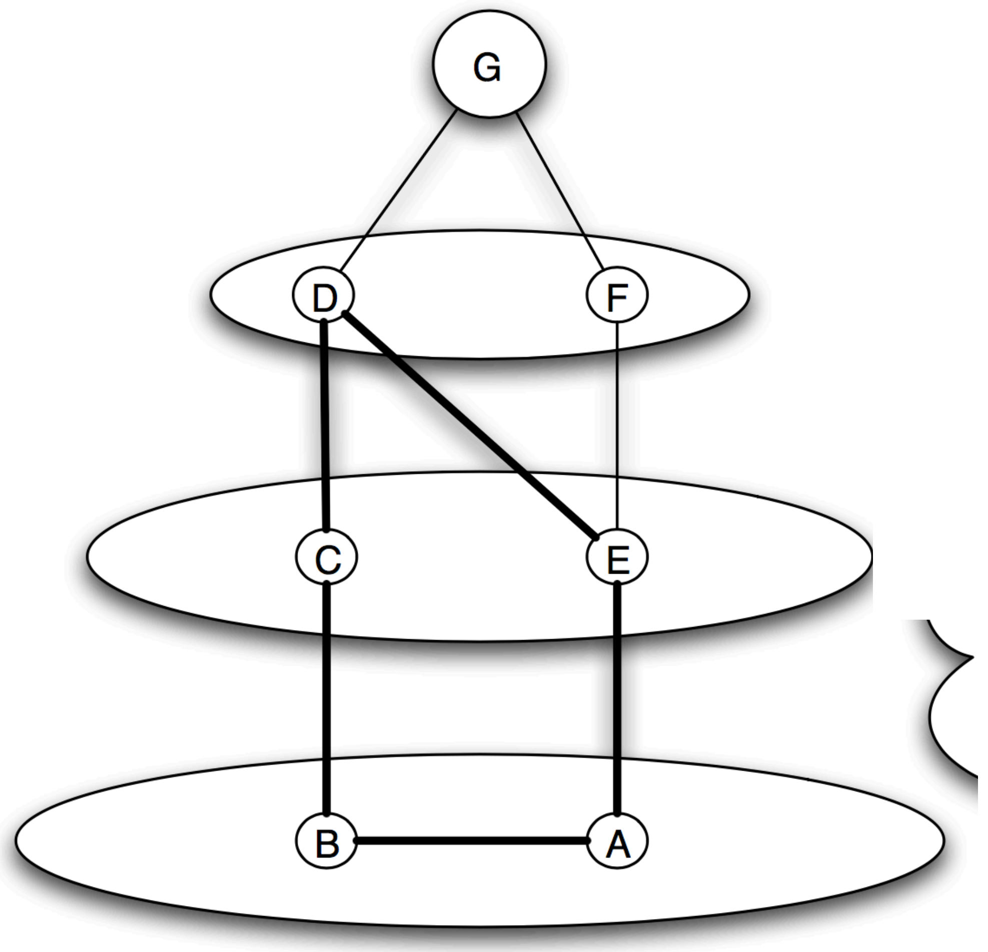
- Now we have a graph on super-nodes joined by negative edges
- Just need to consistently assign super-nodes to coalitions X and Y
- BFS starting at any node in the super-node graph (which only has – edges)
- Produces a set of layers of increasing distances from the root
- Call all even layers X and odd layers Y
- If edges are only between adjacent layers (not within-layer), then all – edges point between X and Y, balanced!
- Otherwise, within-layer edge A-B. Cycle G-A-B-G has length 2k+1, therefore it's odd, therefore unbalanced!



Is a Signed Network Balanced?

Two outcomes:

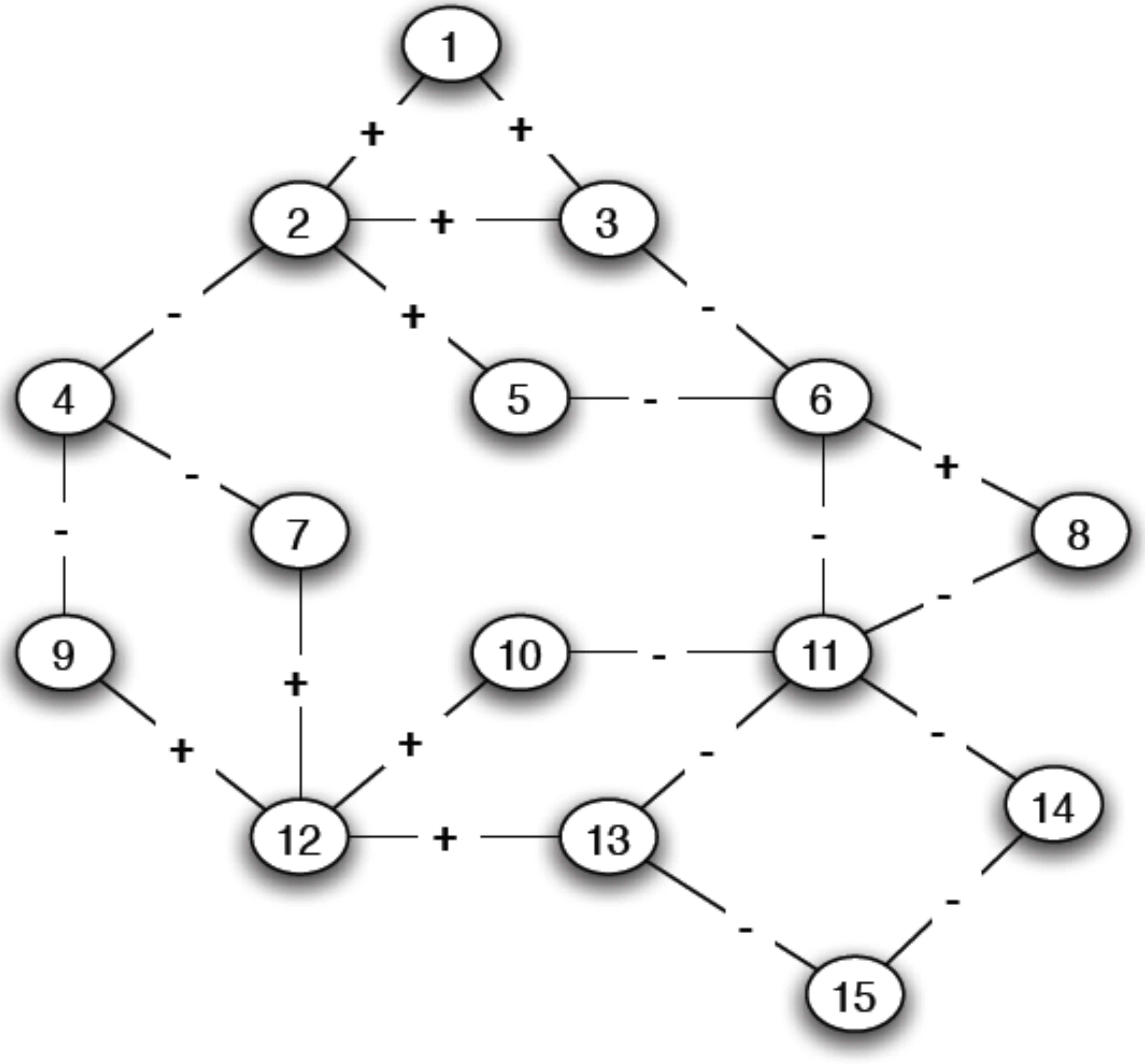
1) label each super-node as either X or Y, in such a way that every edge has endpoints with opposite labels. Then we can create a balanced division of the original graph, by labeling each node the way its supernode is labeled in the reduced graph.



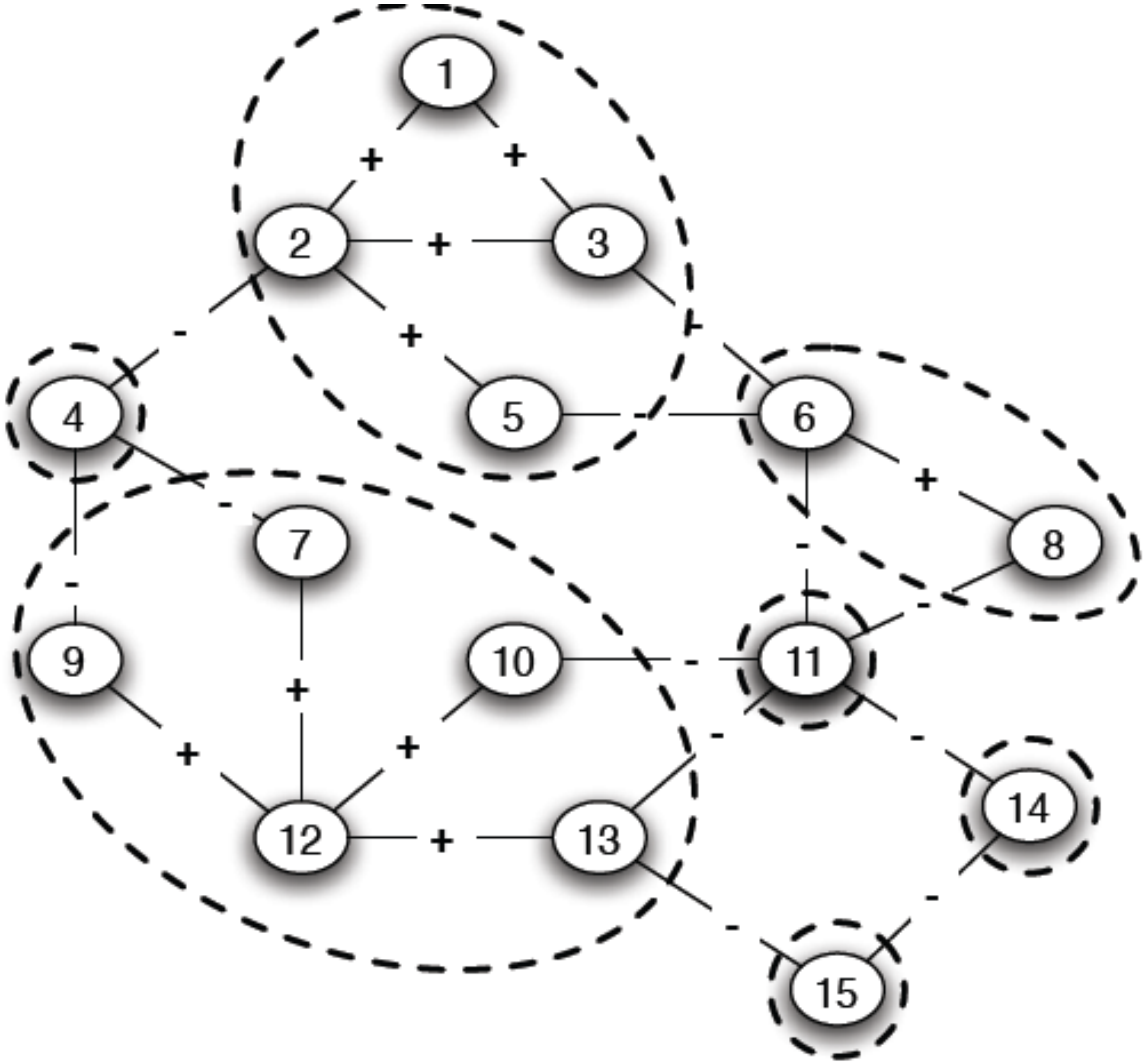
2) find a cycle in the original graph that has an odd number of negative edges

Simply "stitch together" these negative edges using paths consisting entirely of positive edges that go through the insides of the supernodes

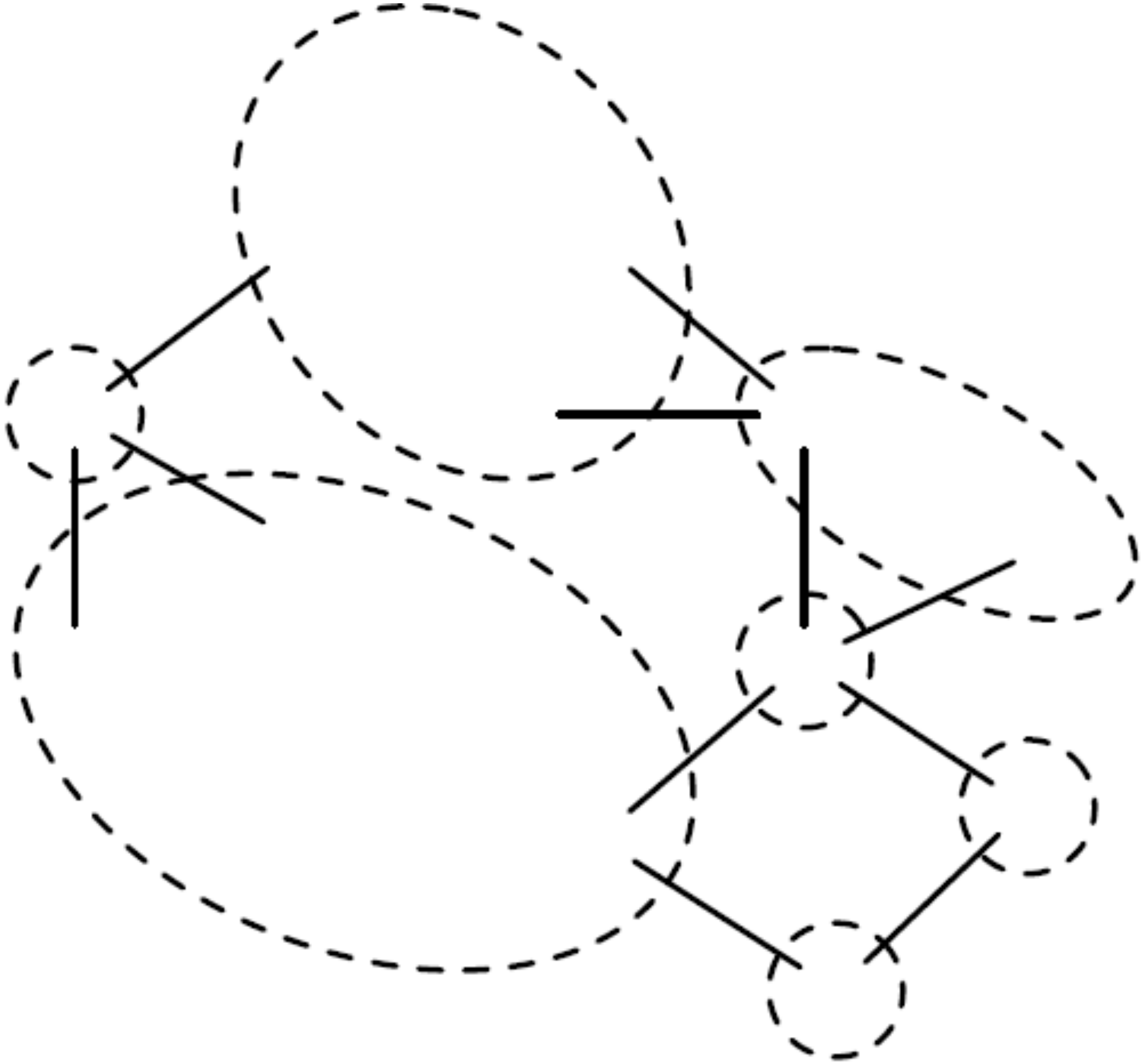
Signed Graph: Is it Balanced?



Positive Connected Components



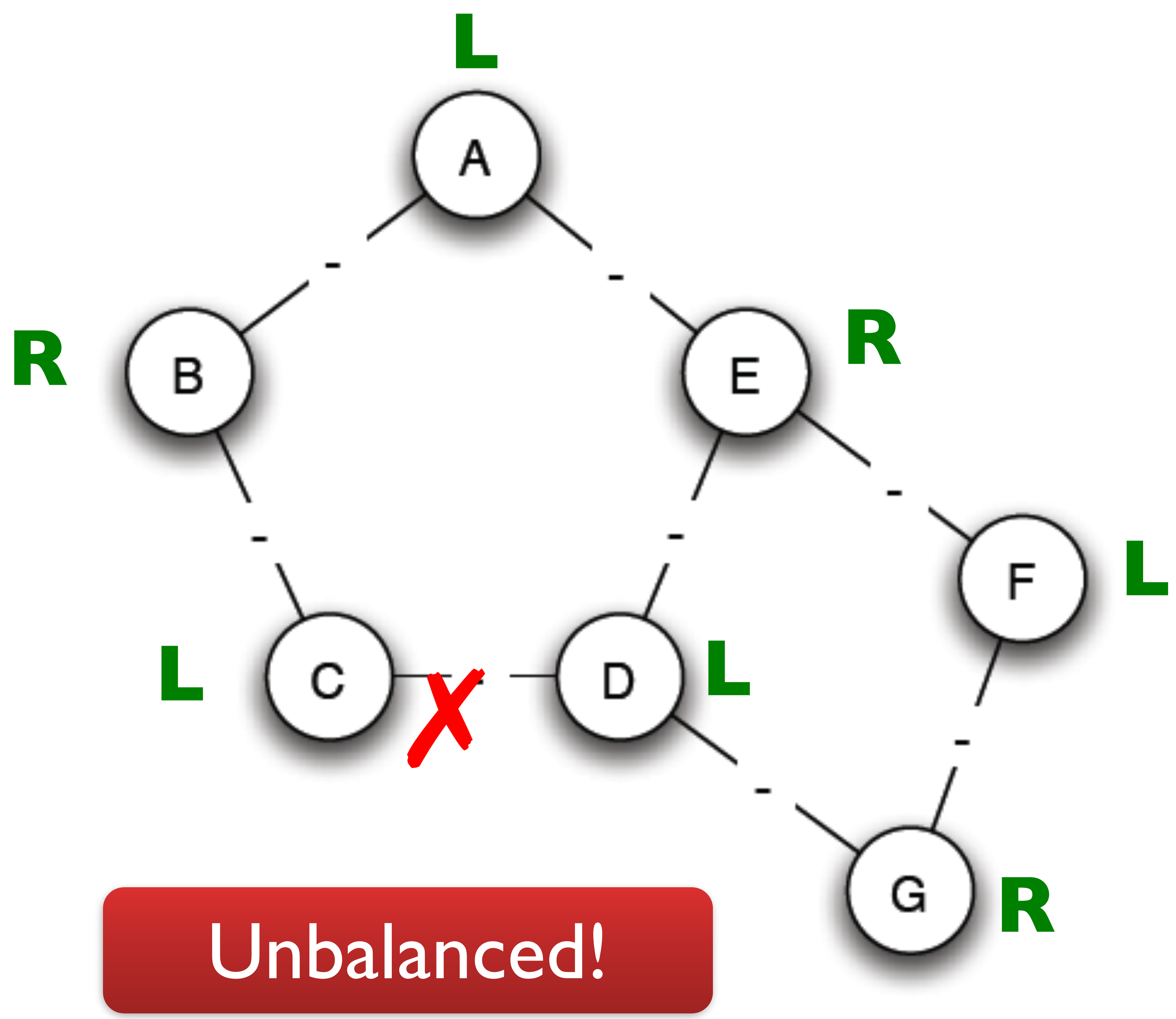
Reduced Graph on Super-Nodes



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**



Where Do Signed Edges Come From?

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

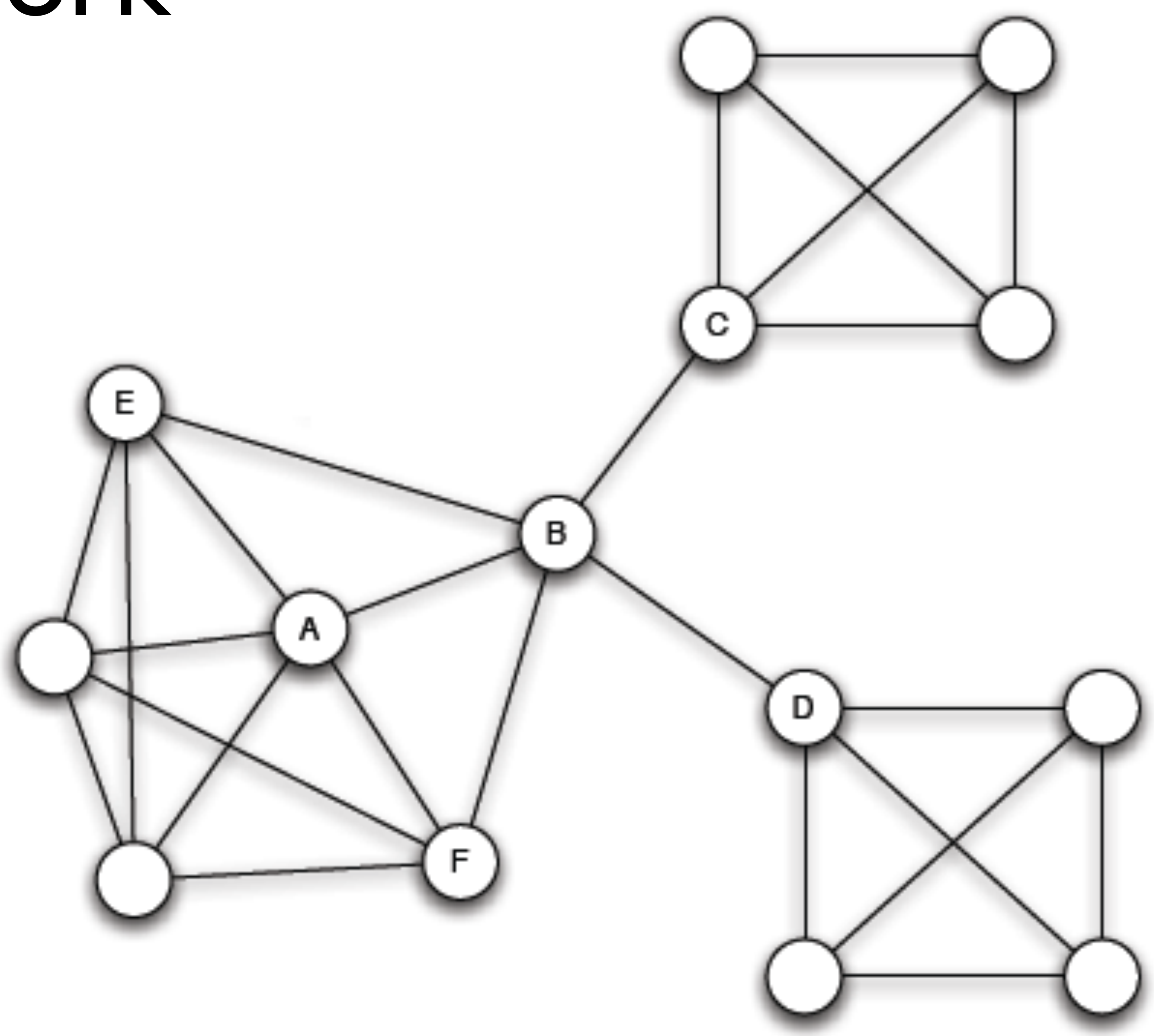
- Through **actions**:
  - Rating a product/person
  - Pressing a "like" button
- Through **text**:
  - Writing a comment, a review
- **Success of these online applications is built on people expressing opinions**
  - Recommender systems
  - Wisdom of the Crowds
  - Sharing economy

Global Structure of Signed Nets

Intuitive picture of social network in terms of densely linked clusters

How does structure interact with links?

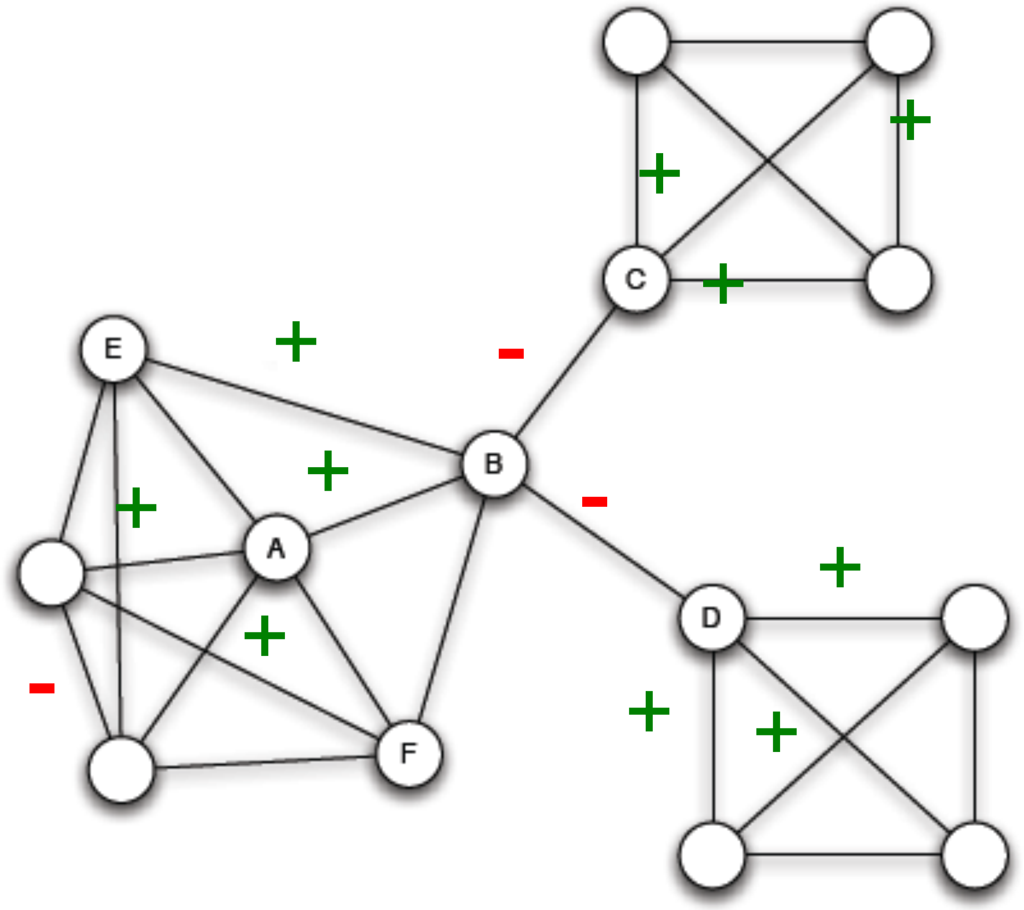
Embeddedness of link (A,B): Number of shared neighbors

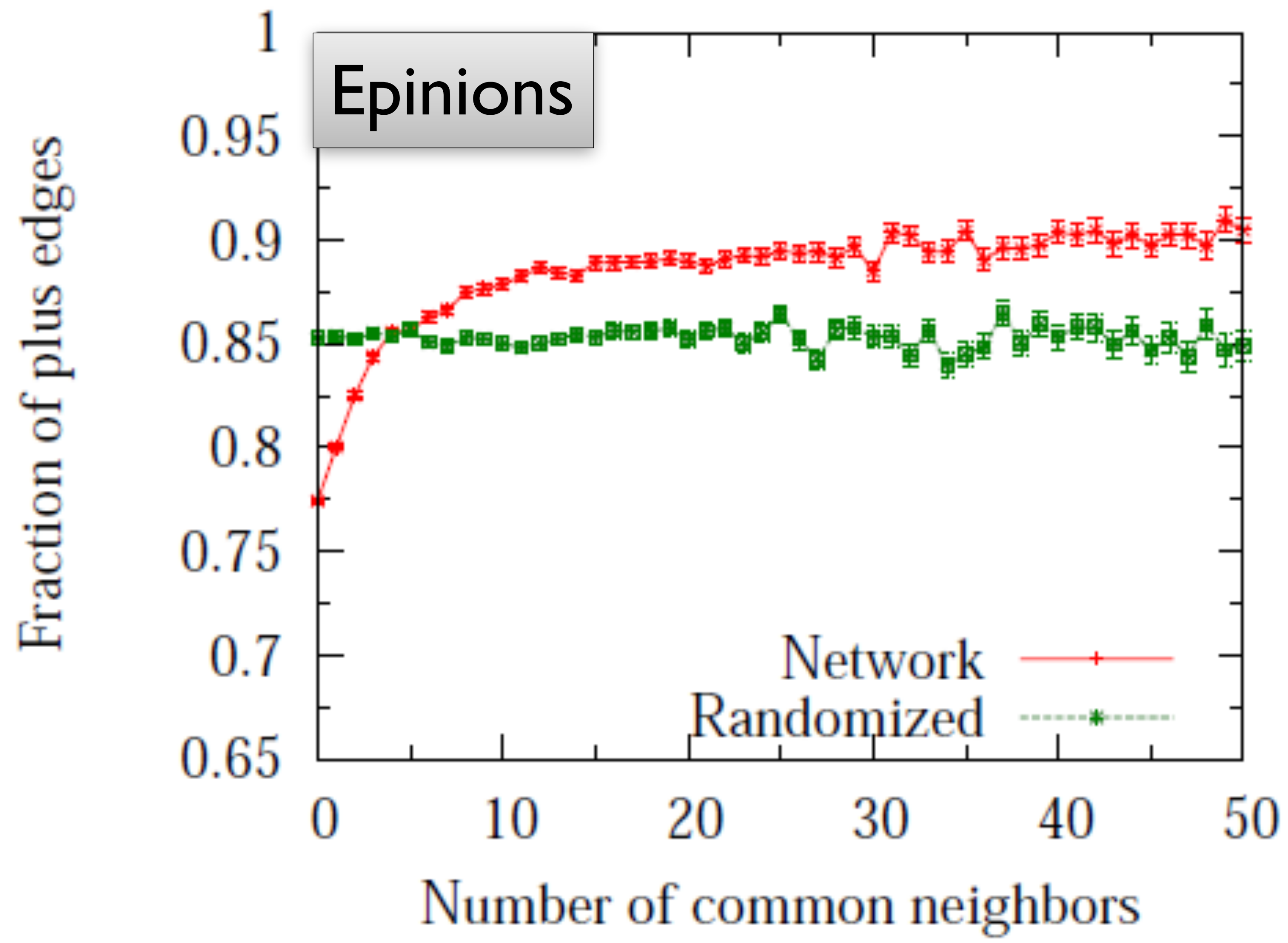


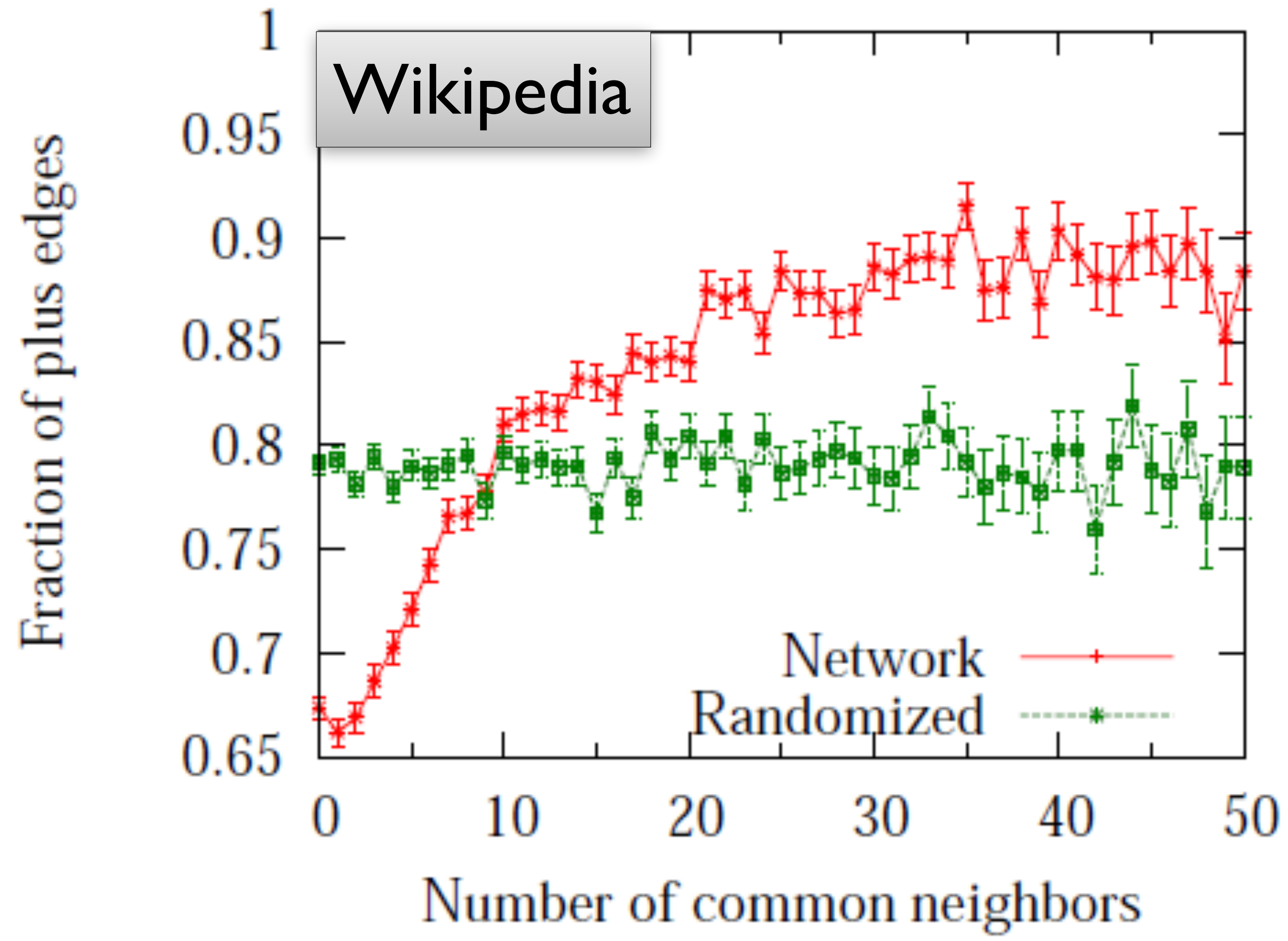
Global Factions: Embeddedness

Embeddedness of ties:

Positive ties tend to be **more** embedded







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 to users)

Wikipedia: Support/Oppose

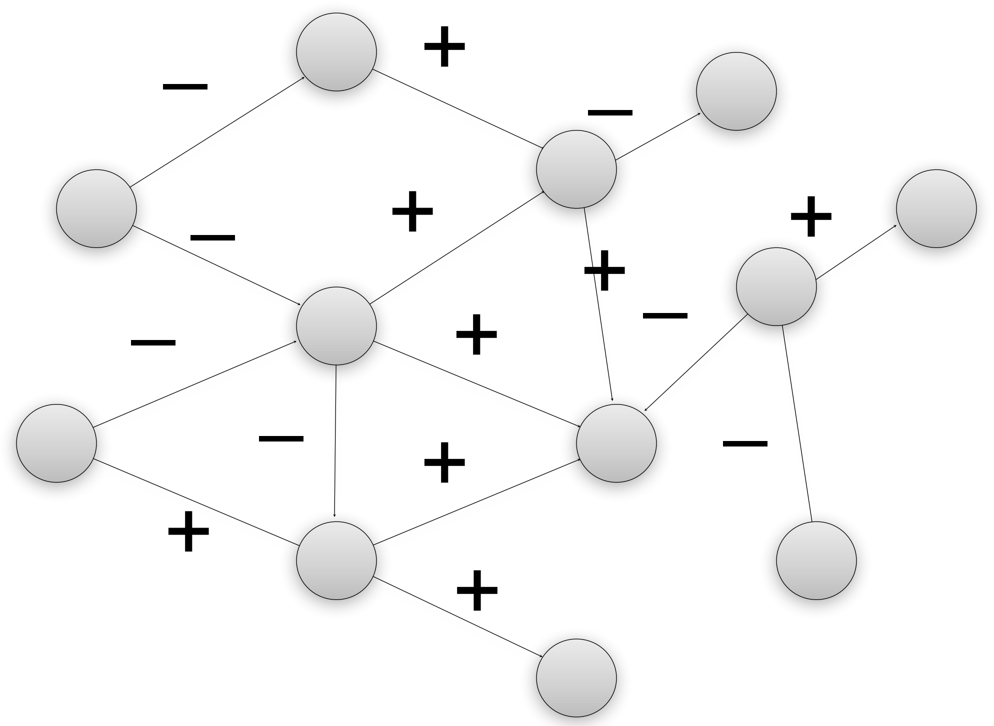
Does A support B to become Wikipedia administrator?

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

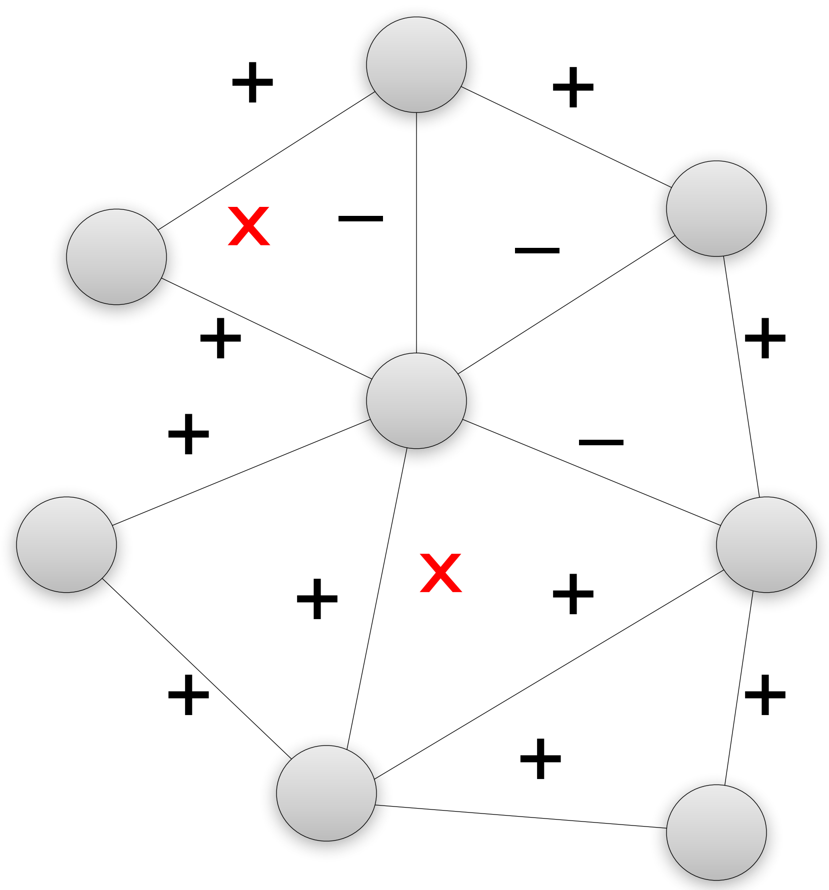
Does structural balance hold?

Compare frequencies of signed triads in real and "shuffled" signs

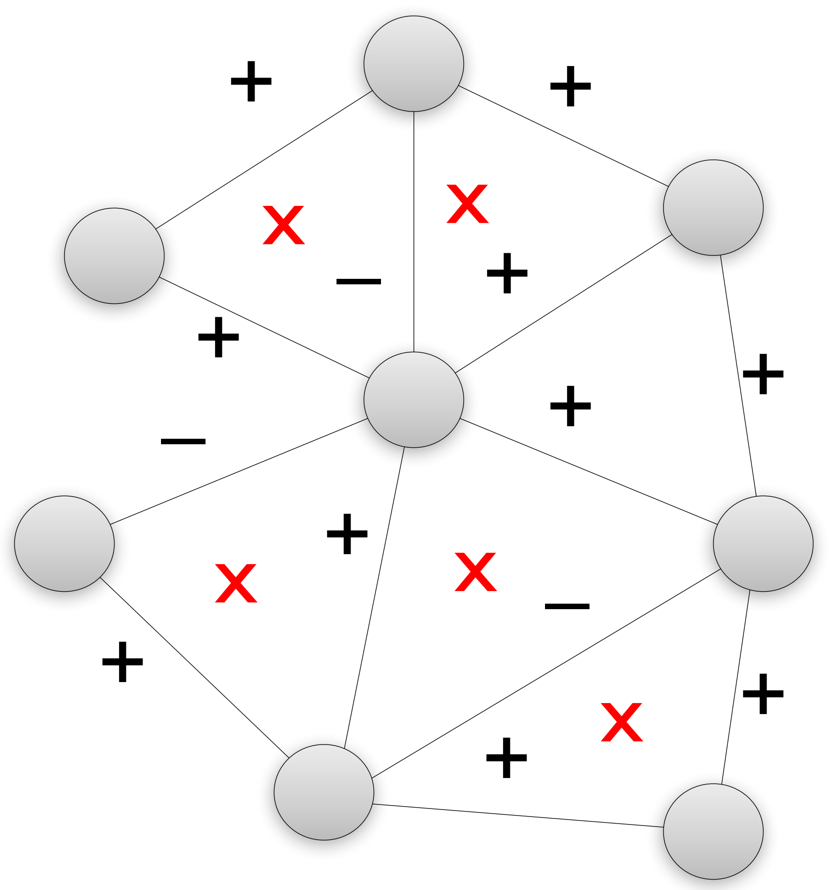
	Triad	Epinions		Wikipedia		Consistent with Balance?
		P(T)	P_0(T)	P(T)	P_0(T)	
Balanced	+ + +	0.87	0.62	0.70	0.49	✓
Balanced	- - +	0.07	0.05	0.21	0.10	✓
Unbalanced	+ + -	0.05	0.32	0.08	0.49	✓
Unbalanced	- - -	0.007	0.003	0.011	0.010	✗

P(T) … fraction of a triads

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



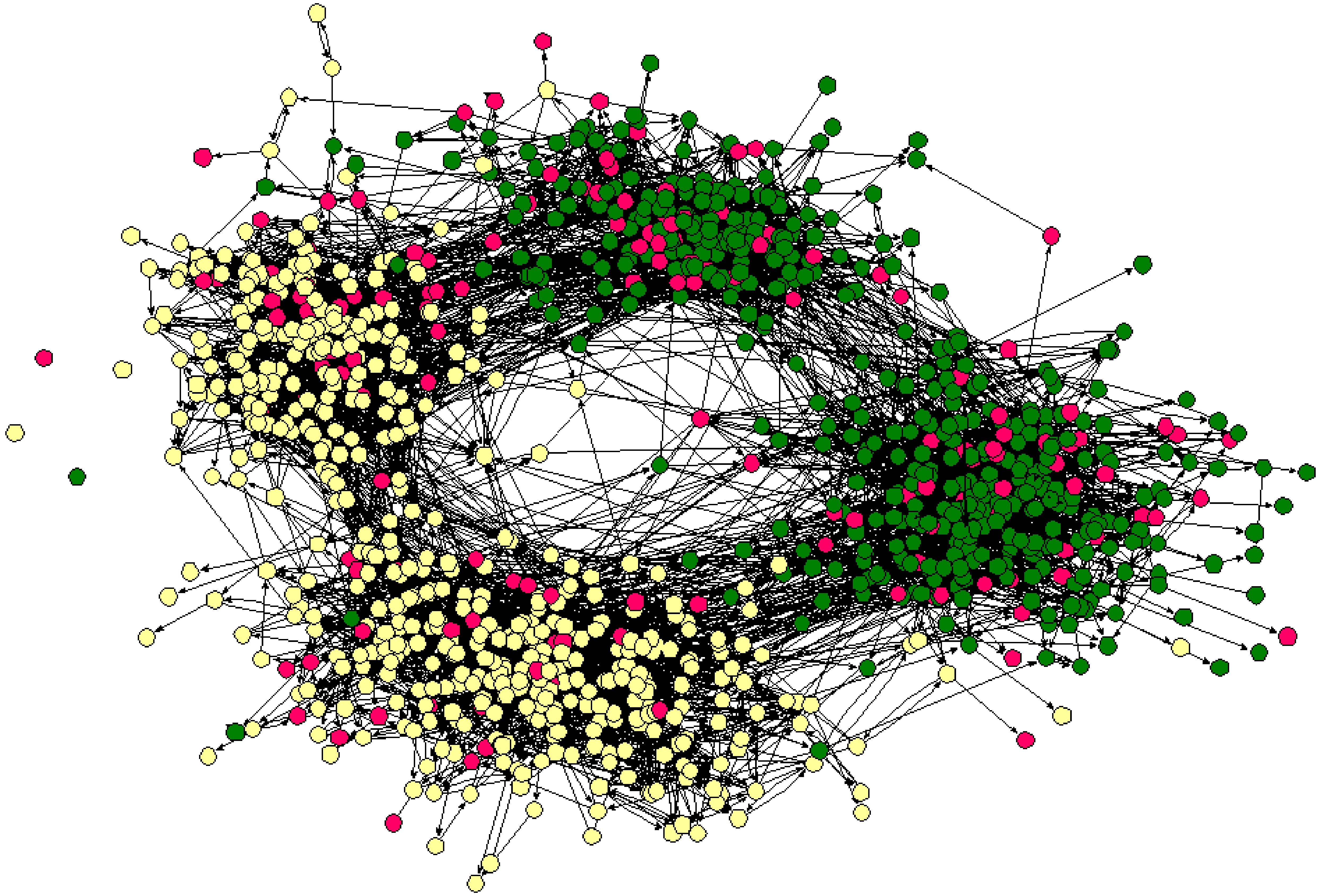
Real data



Shuffled data

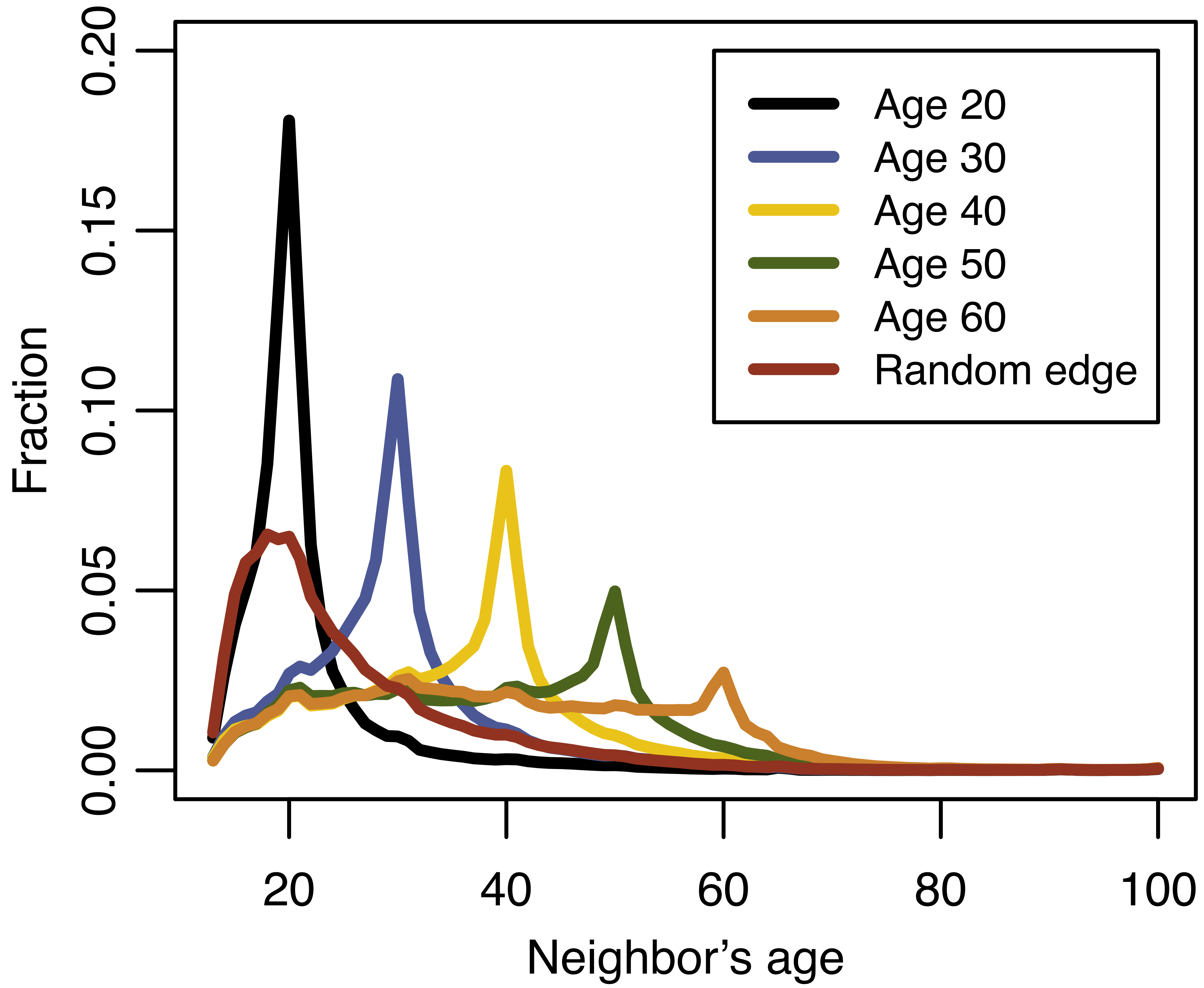
Homophily
“Birds of a Feather Flock Together”

Homophily



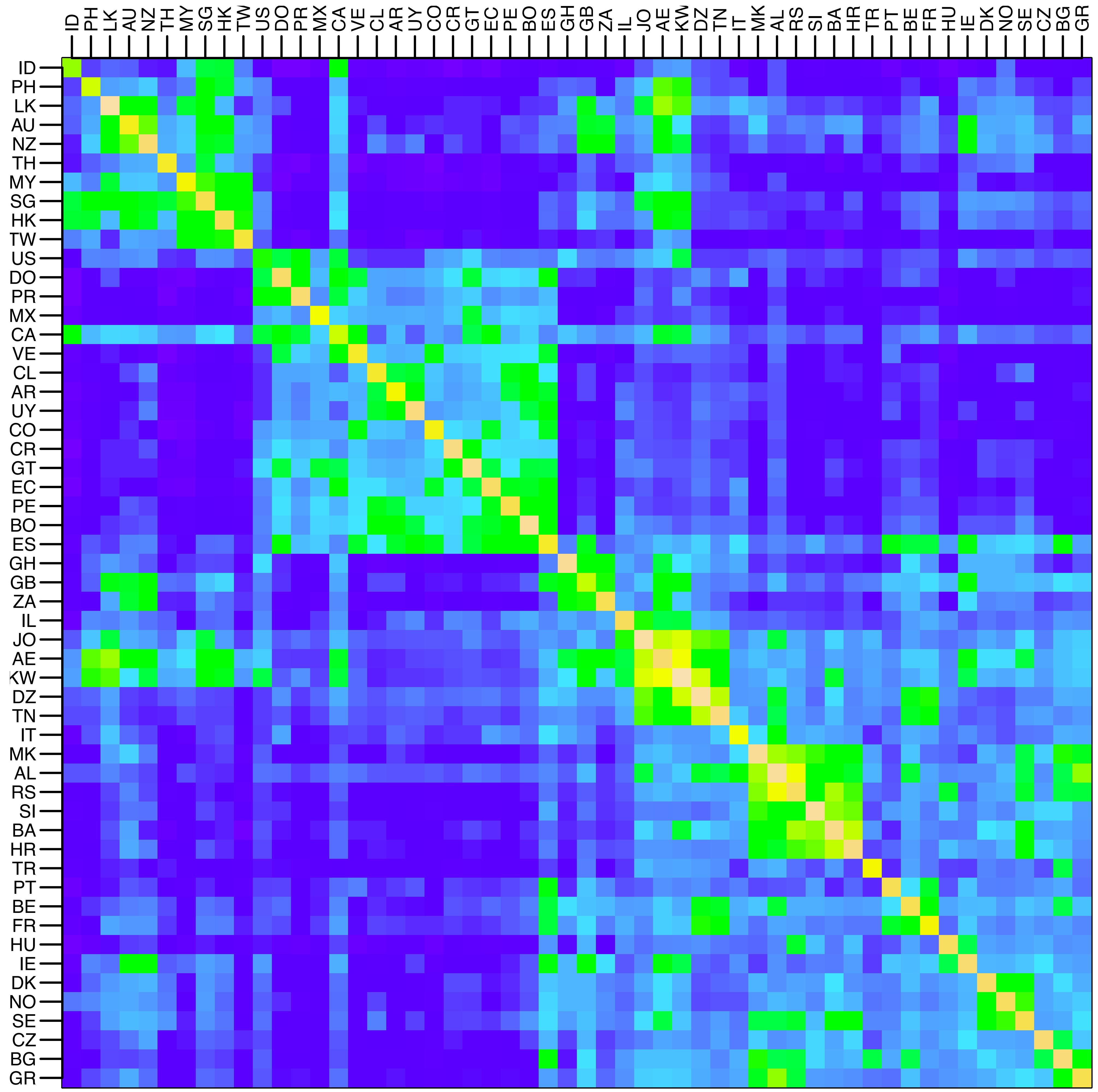
- US middle school + high school
- node color = self-identified race

Homophily: Age



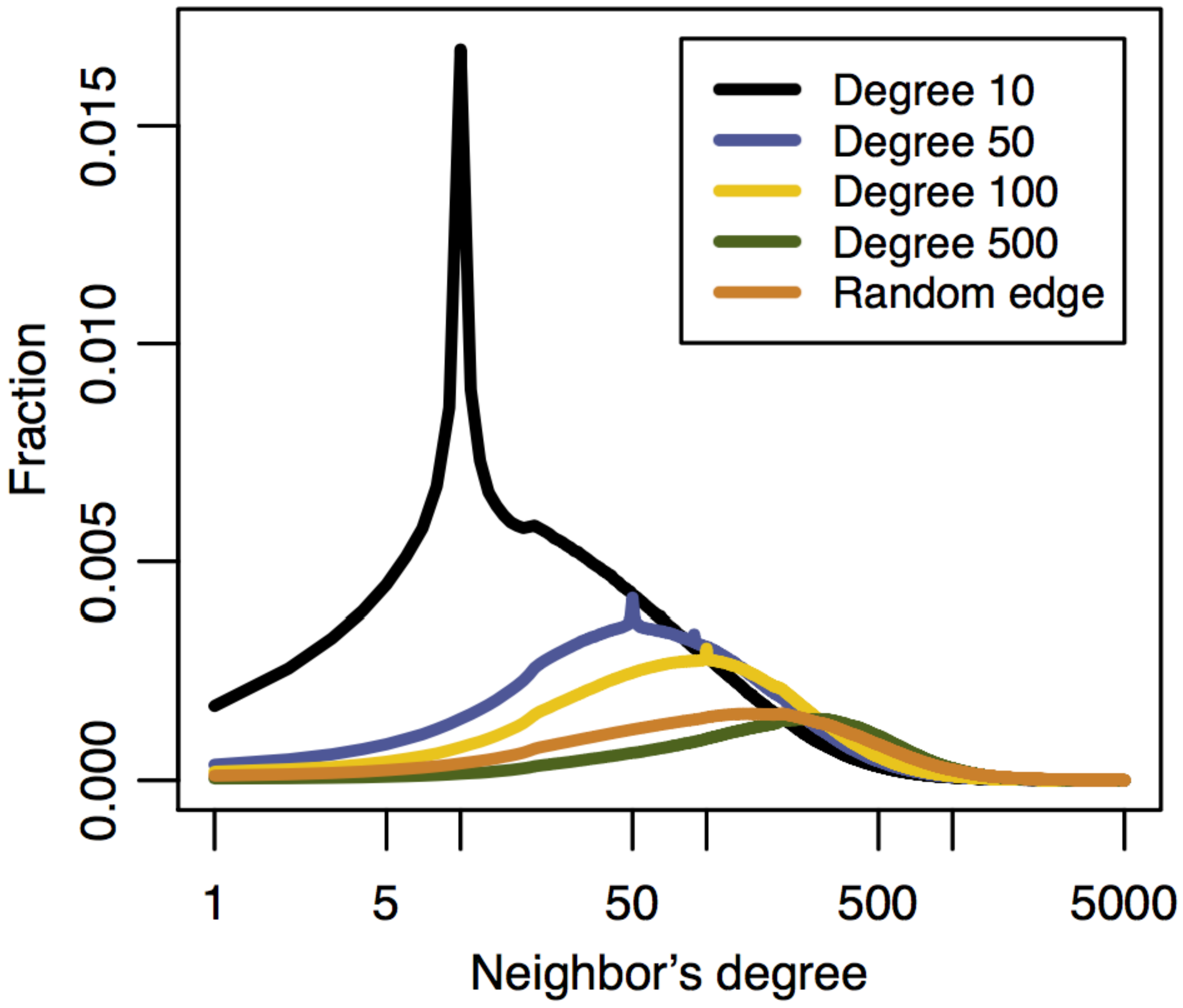
- Facebook friendship network, 2011

Homophily: Nationality



- Facebook friendship network, 2011

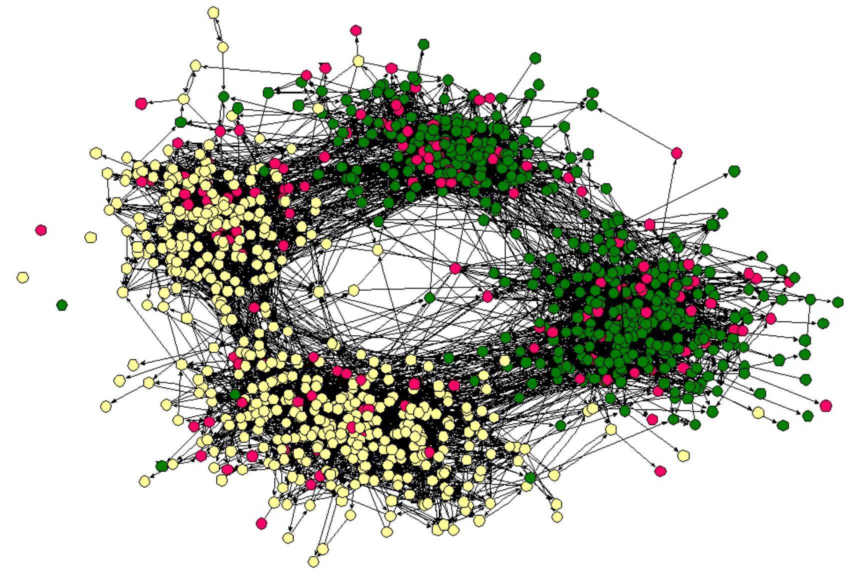
Homophily: Friend count



- Facebook friendship network, 2011

Homophily

- Connections don't form uniformly at random
- **Null model**: what if they were forming at random?
- **Measuring homophily**: are there fewer connections between nodes across traits than you'd expect at random?
- **Homophily test:** If the fraction of cross-gender edges is significantly less than at random, then there is evidence of homophily.



Homophily

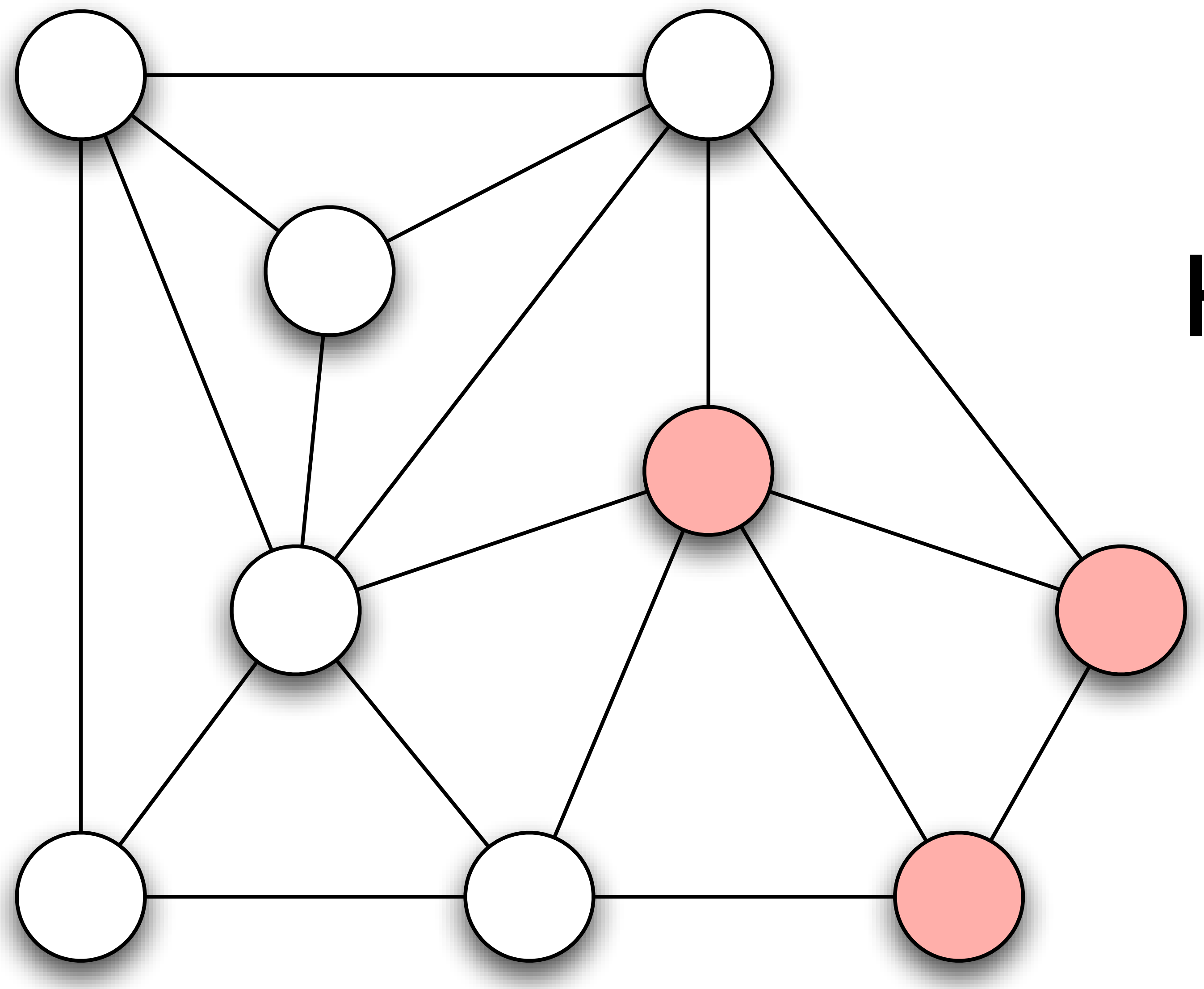
p = Probability that a node is white

q = Probability that a node is red

Prob an edge is between two white nodes?

Prob an edge is between two red nodes?

Prob an edge is between 1 red, 1 white?



Homophily test:

Homophily

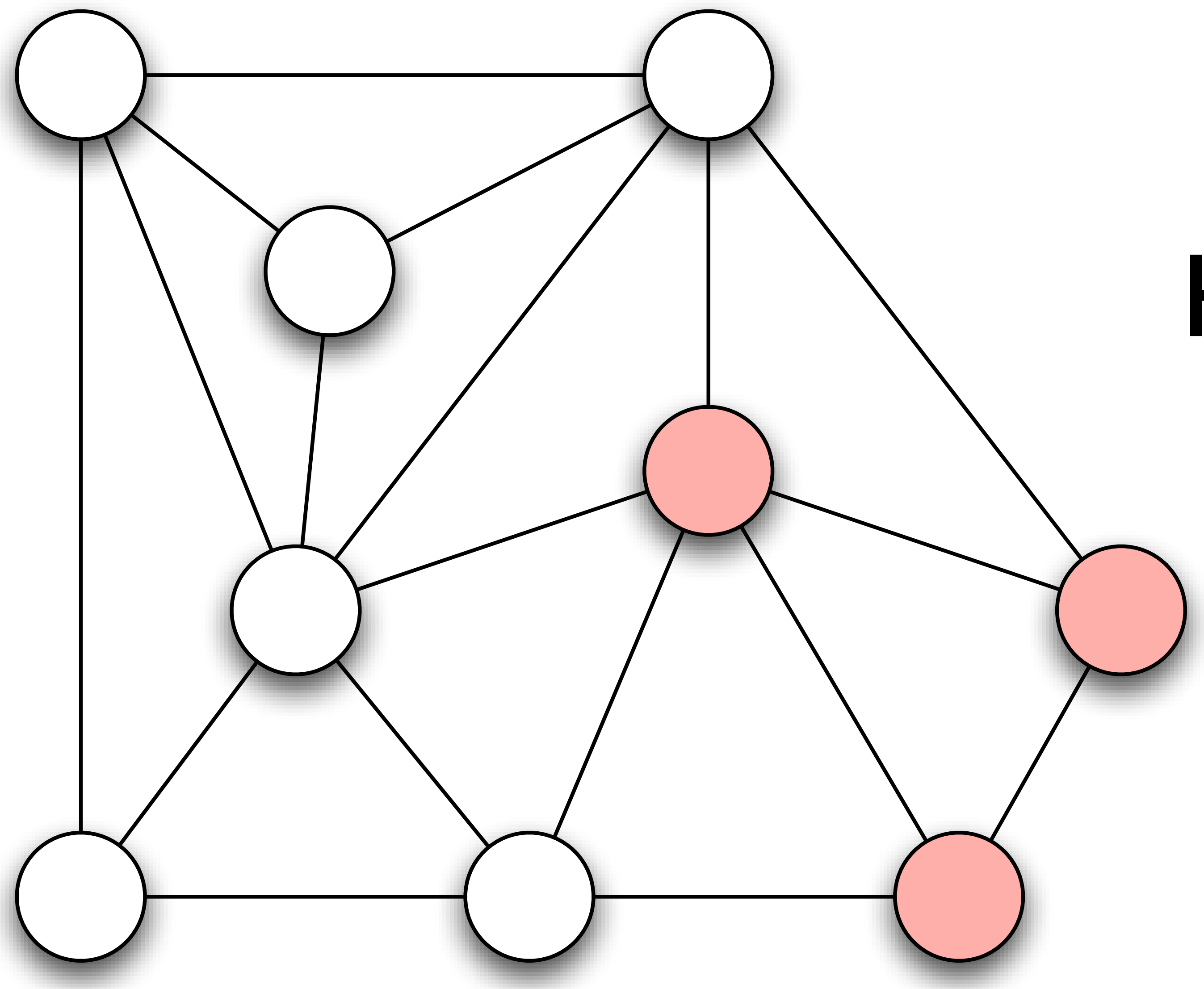
p = Probability that a node is white 6/9=2/3

q = Probability that a node is red 3/9=1/3

Prob an edge is between two white nodes? p^2

Prob an edge is between two red nodes? q^2

Prob an edge is between 1 red, 1 white? $2pq$



Homophily test: $2pq = 4/9 = 8/18$

Observed: 5/18

The Friendship Paradox

Friendship paradox

Your friends probably have **more friends** than you do

Friendship paradox

Average degree <= Average friend degree

Friendship paradox

- **Facebook friend graph (2012):**
  - 720M people, 70B edges
  - Average Facebook user number of friends: 190
  - Average friend's number of friends: 635
  - User's friend count was lower than the average of their friends' friend counts 93% of the time
  - ???

Friendship paradox

- **Consider an example:**
  - Two buses to school
    - One big one with 90 students
    - One small one with 10 students

  - Average bus size = 50
  - This is misleading…

Friendship paradox

- **Consider an example:**
  - Two buses to school
    - One big one with 90 students
    - One small one with 10 students
  - Average bus size = 50
  - What about average *bus-rider experience*?

Friendship paradox

- **From students' point of view:**
  - How packed is your bus?
    - 90 students say 90
    - 10 students say 10

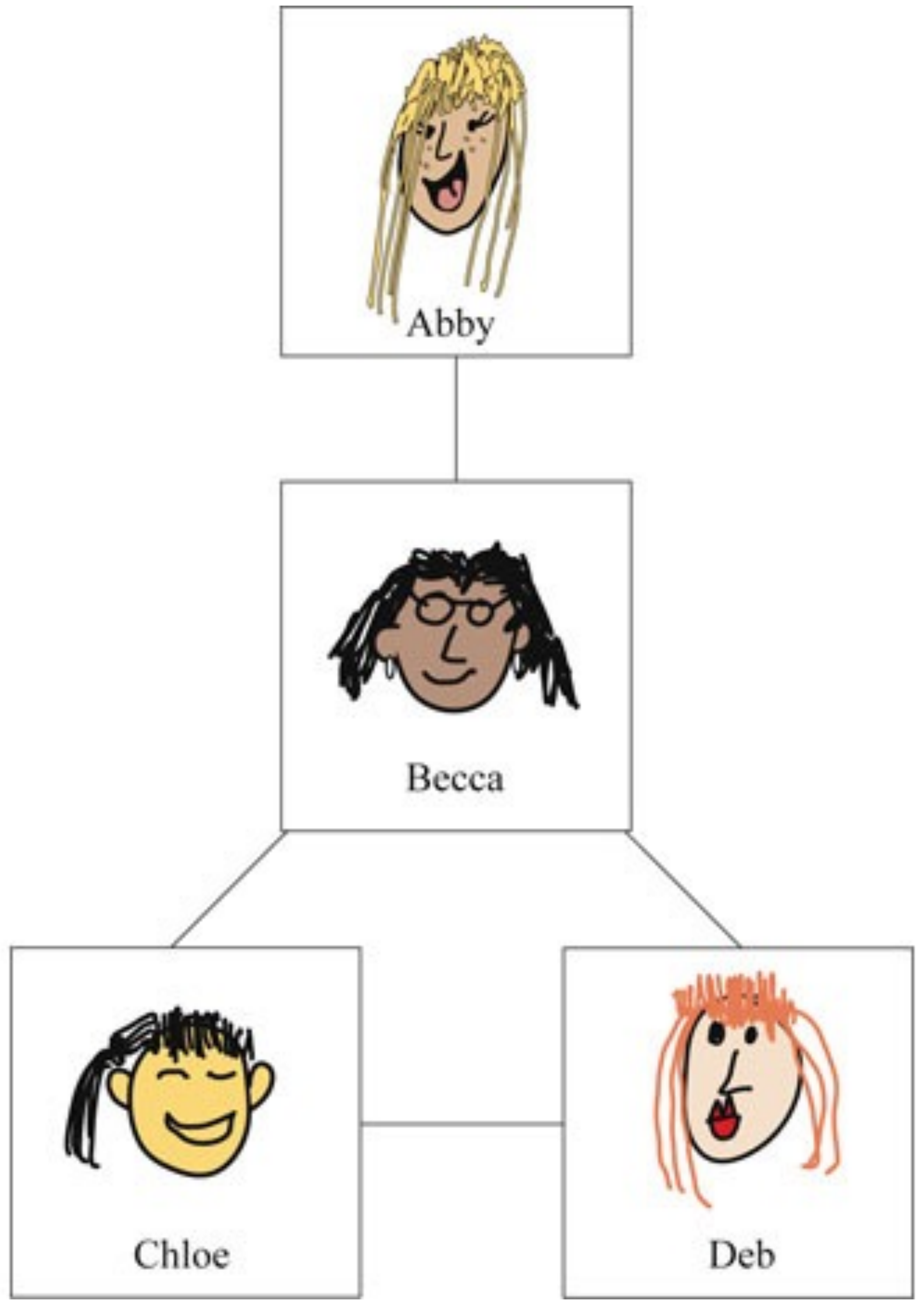
Average *bus-rider experience* =

[(90*90)+(10*10)]/100 = 82

Friendship paradox

- Friend counts: 1, 3, 2, 2.
- Average friend count:

- Average friend count of a friend:



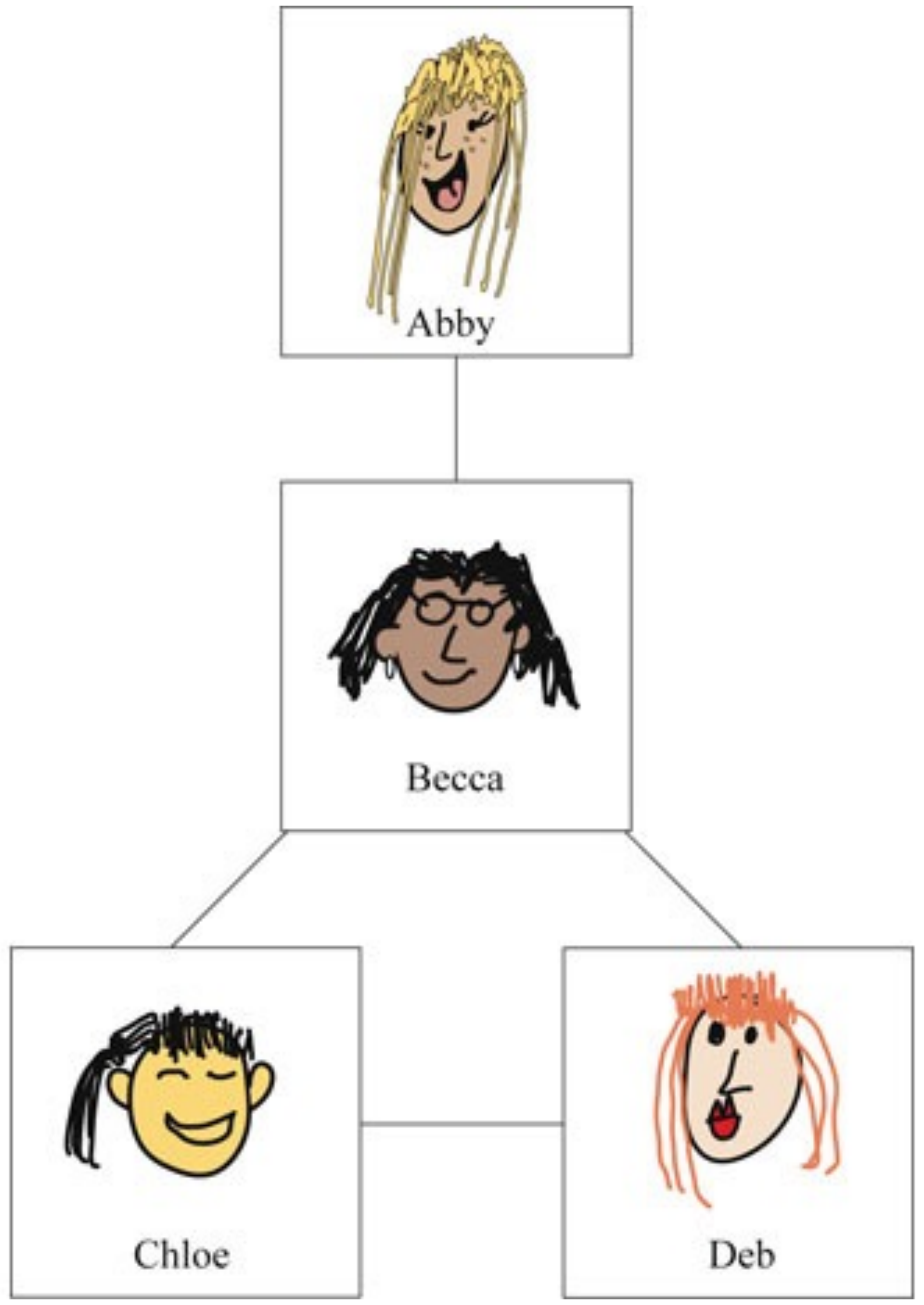
Friendship paradox

- Friend counts: 1, 3, 2, 2.
- Average friend count: 8/4=2
- Average friend count of a friend:

A: 3, avg = 3
B: 1, 2, 2, avg = 5/3
C: 3, 2, avg = 2.5
D: 3, 2, avg = 2.5

Avg friend of friends = 2.4166 > 2

B mentioned 3 times, A only 1

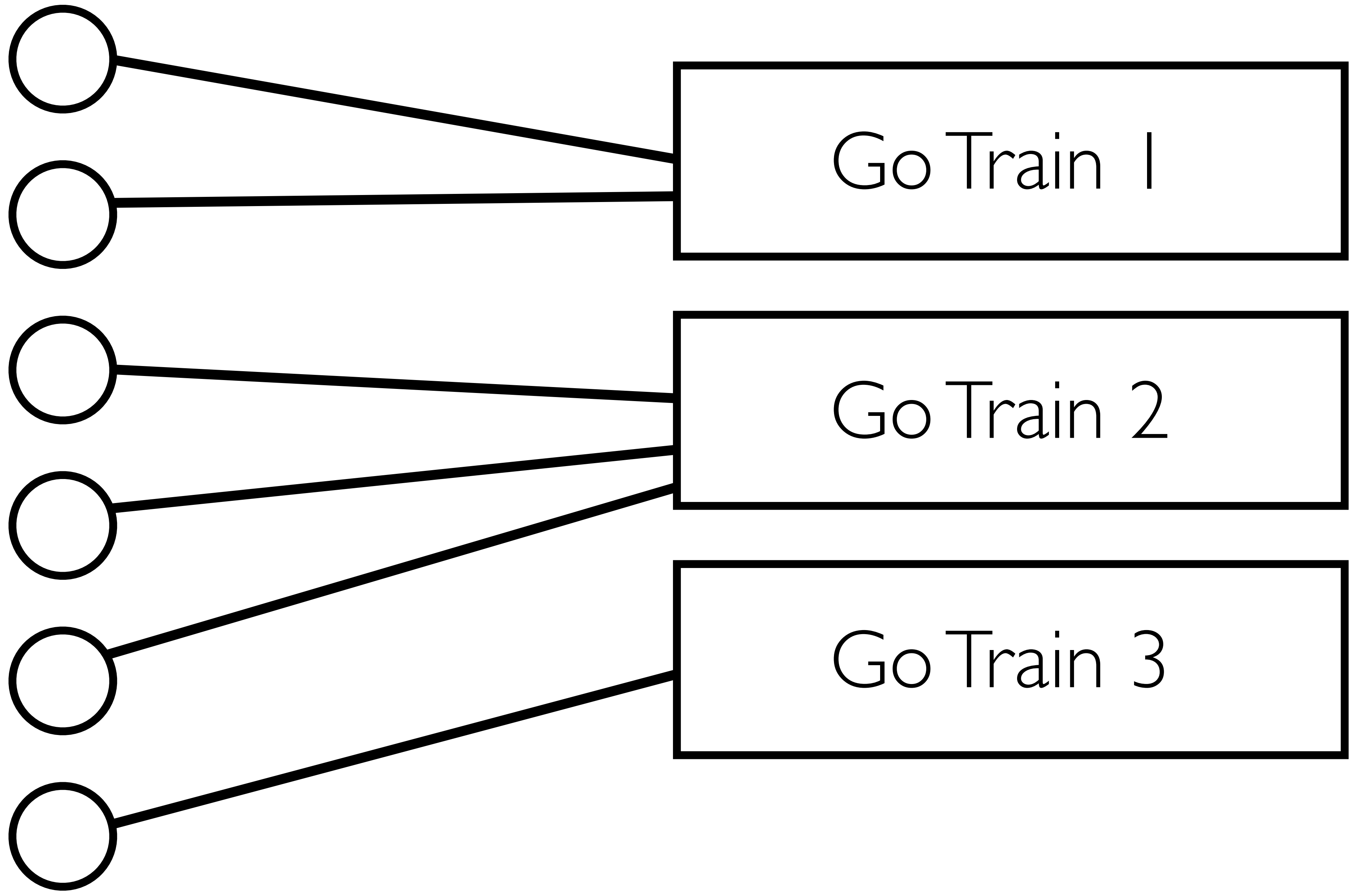


"Average friend-experience" vs. average friends

"Friendship paradox"

- Avg friend count person ≤ Avg friend count of friend

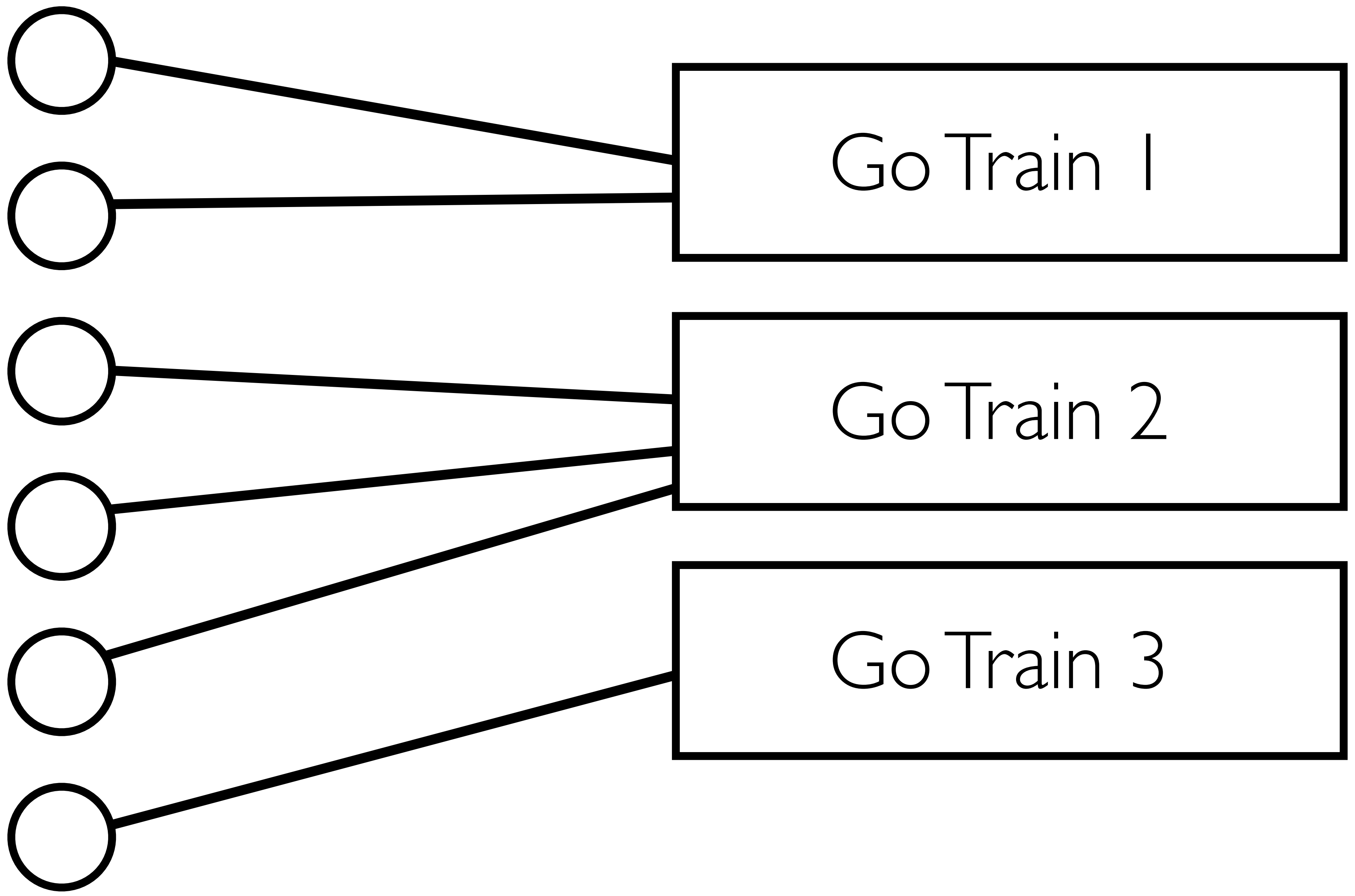
- Avg # on a train ≤ Avg # on "train experience"

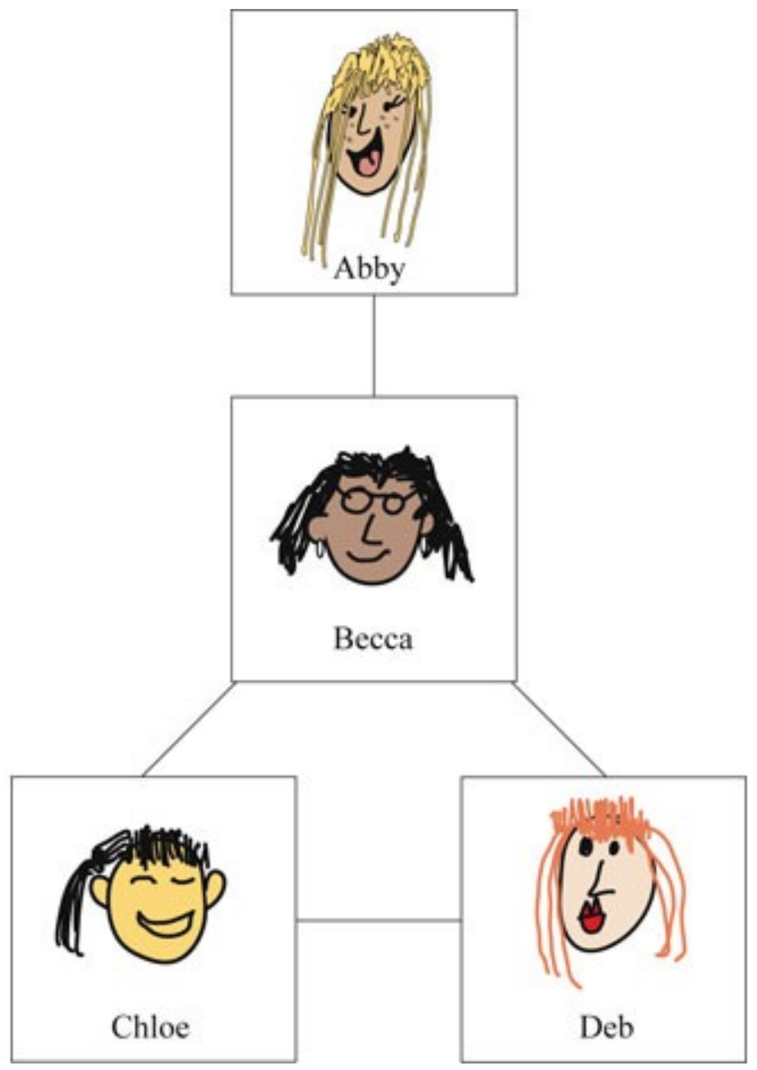


“Friendship paradox”

- Avg friend count person ≤ Avg friend count of friend

- Avg # on a train ≤ Avg # on “train experience”





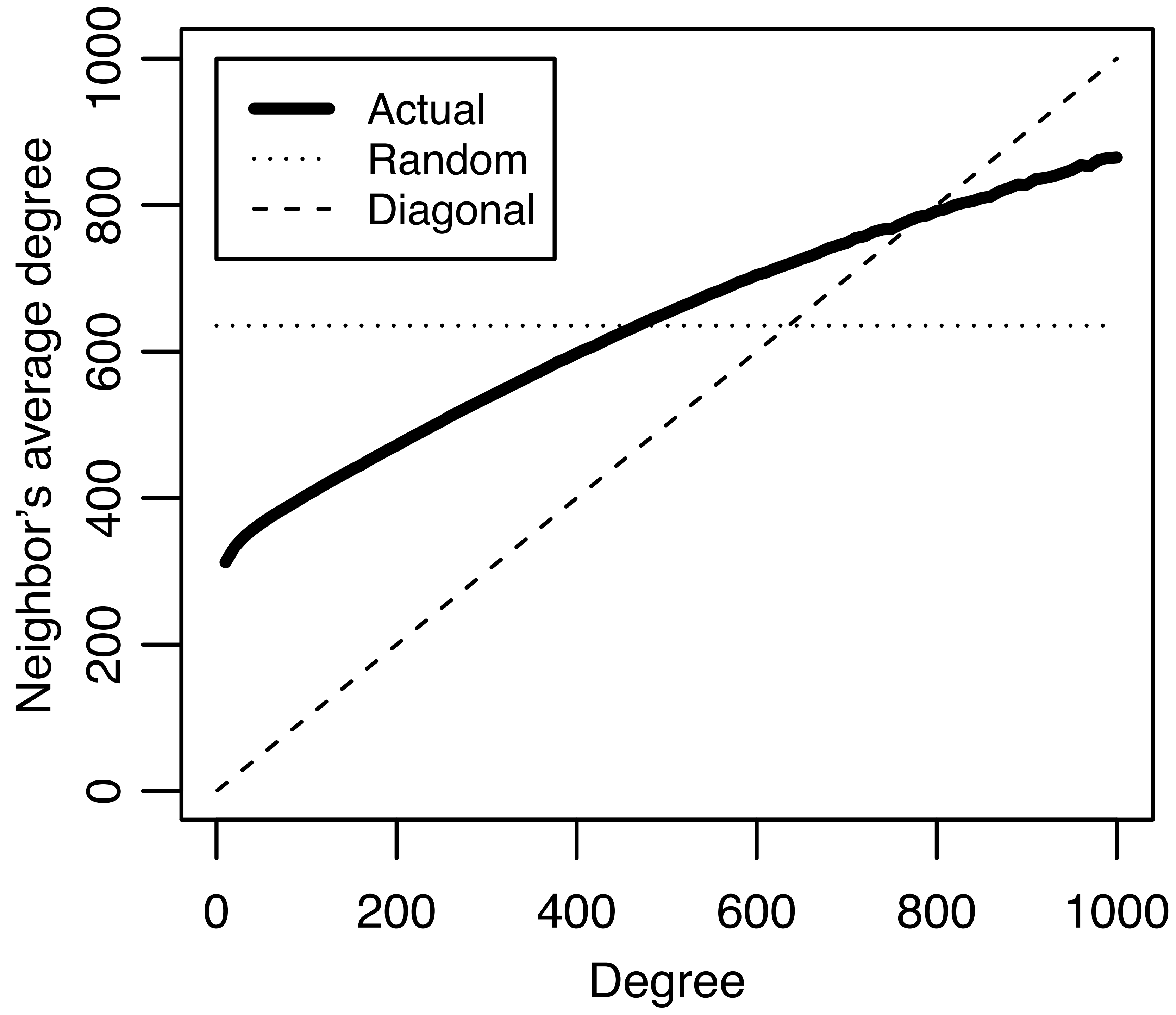
- Basic principle: weighted averages

"Friendship paradox"

- Friend average = $\frac{\text{Weighted average}}{\text{Average}}$

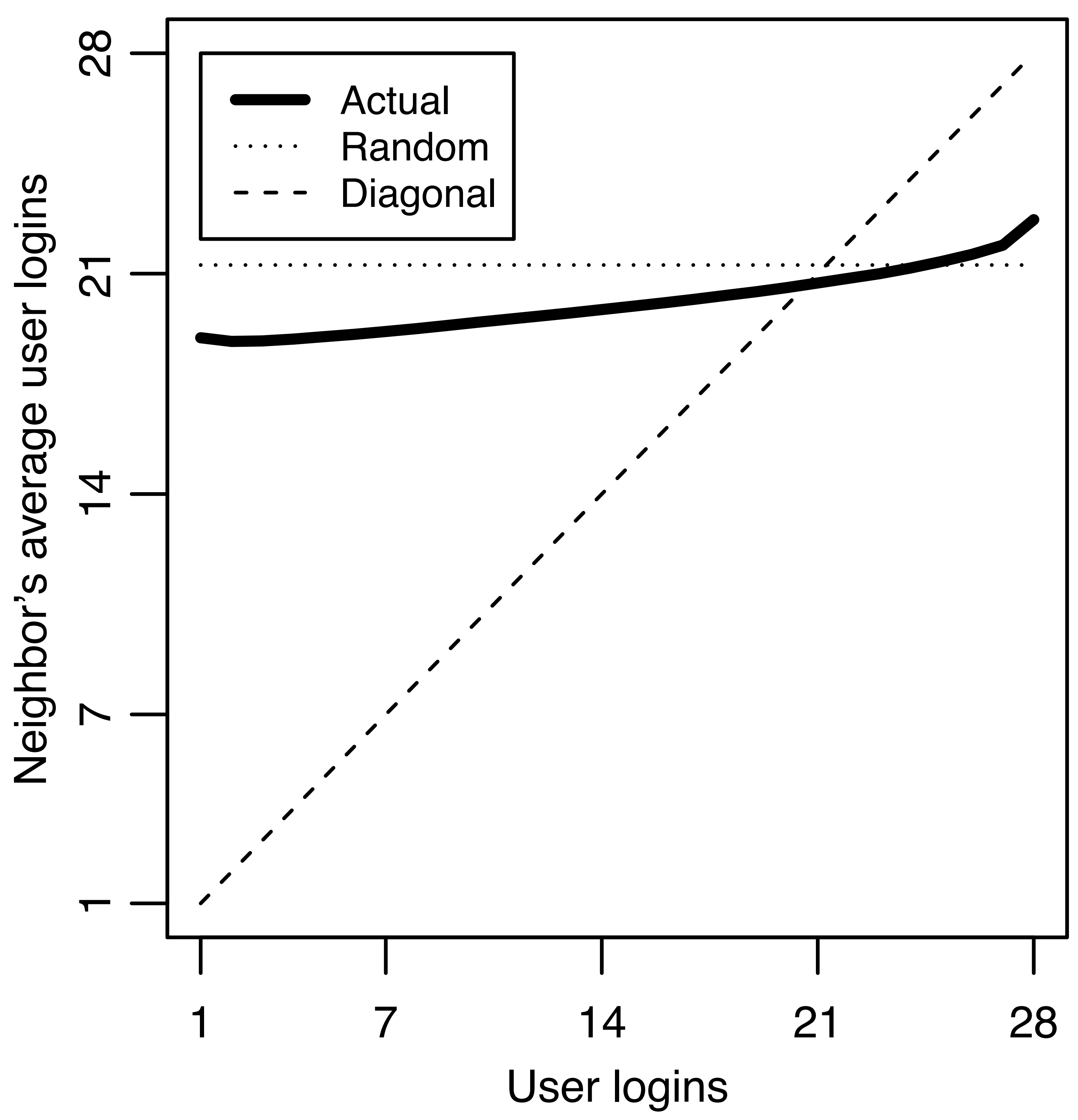
- Friend average = Average + $\frac{\text{Variance}}{\text{Average}}$

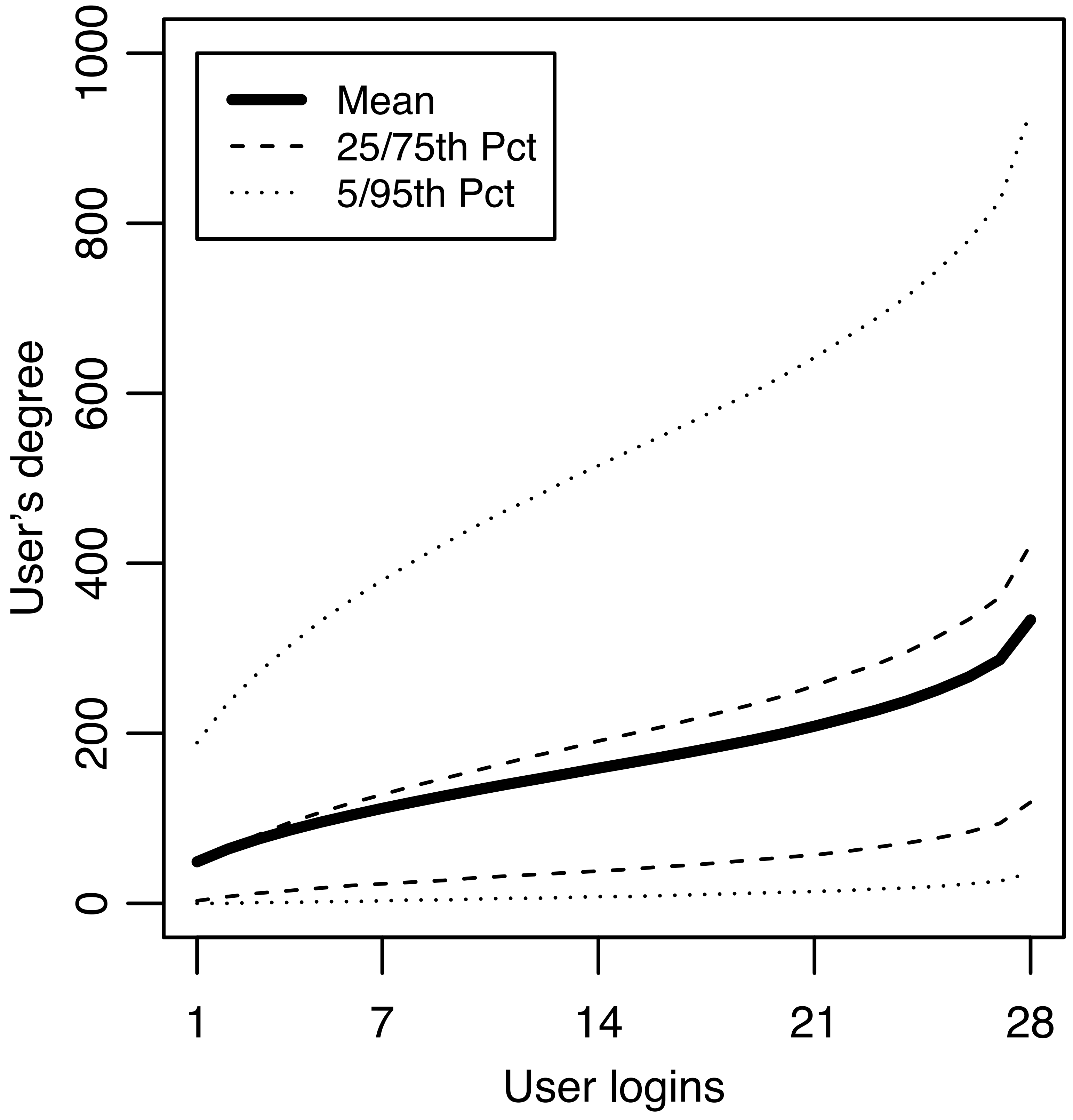
Friendship paradox on FB



Corollary paradoxes

- "Your friends log in more than you" (and more)





Friendship paradox

- Not a social fact!
  - It's a mathematical fact
  - Applies to virtually any network
  - But it has social implications…
    - Web pages you link to probably have more links
    - People you high-five probably high-five more people than you
    - Etc etc

Friendship paradox

- **Application: Disease outbreak**
  - **Many diseases spread via social networks**
  - **Model: immunize random friends of random people instead of random people**
  - **With random people: need to immunize 80-90% of population**
  - **With random friends of random people: only immunize 20-40% of population**
  - **We'll study contagion in later weeks**