Online Social Networks and Media

Network Ties

Introduction

- How simple processes at the level of *individual nodes* and links can have complex effects at the *whole population*

- How *information flows* within the network

- How *links/ties* are formed and the distinct roles that structurally different nodes play in link formation

Assortativity

similar nodes are connected with each other more often than with dissimilar nodes

Why are friendship networks assortative (similar)?

- (Social) Influence (or, socialization): an individual (the influential) affects another individual such that the influenced individual becomes more similar to the influential figure

- Selection (Homophily): similar individuals become friends due to their high similarity

- Confounding: the environment's effect on making individuals similar/Surrounding context: factors other than node and edges that affect how the network structure evolves (for instance, individuals who live in Russia speak Russian fluently)

Mutable & immutable characteristics

Influence vs Homophily

- Individuals already linked together change the values of their attributes

- Connections are formed due to similarity

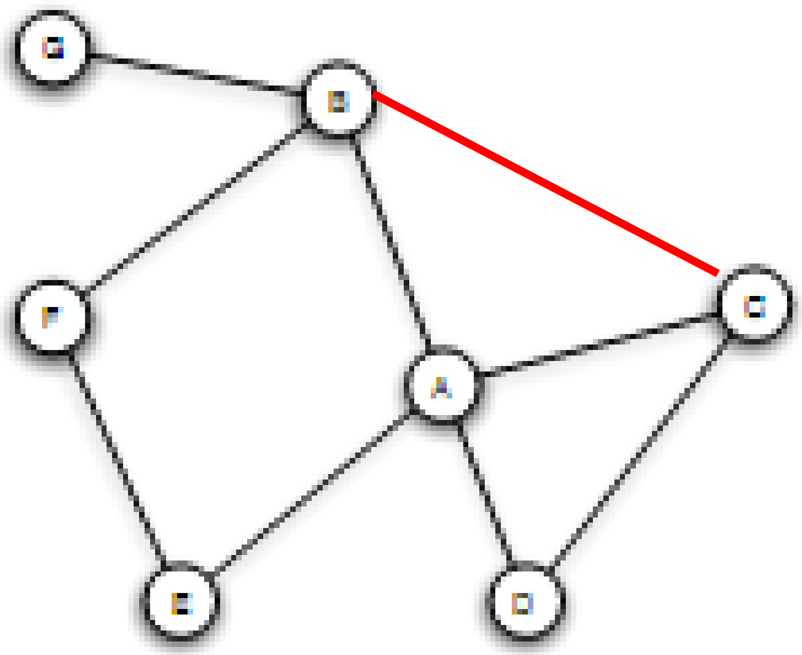
Influence vs Homophily

Which social force (influence or homophily) resulted in an assortative network?

STRONG AND WEAK TIES

Triadic Closure

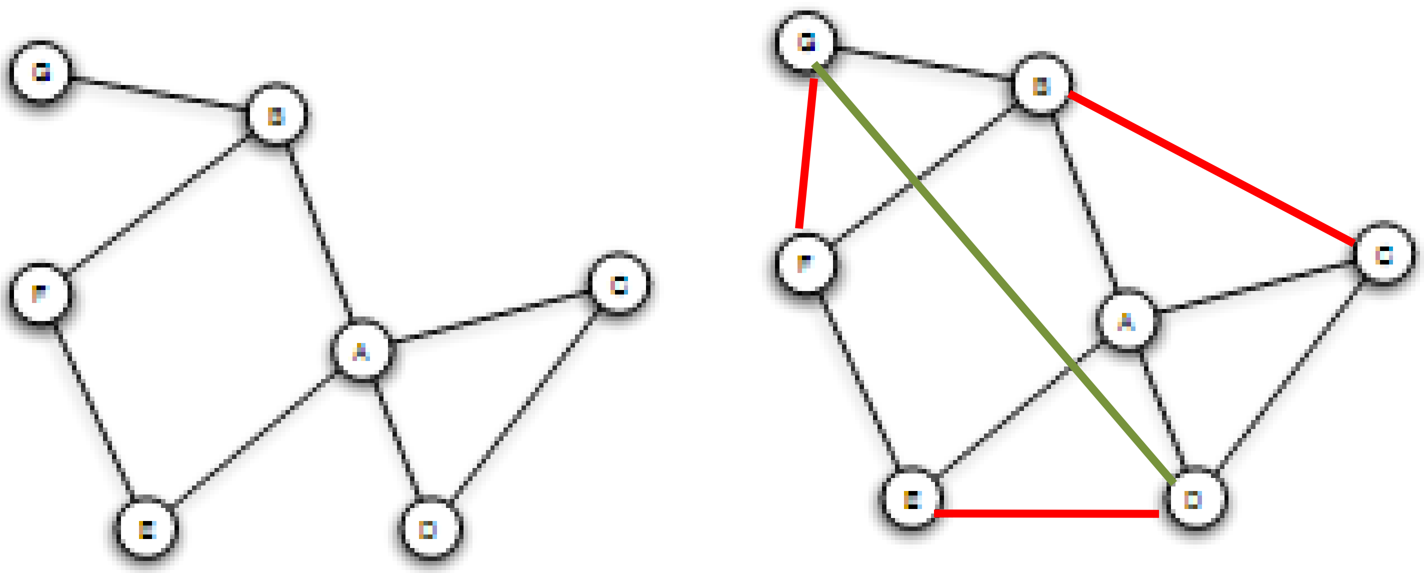
If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future



Triangle

Triadic Closure

Snapshots over time:



Clustering Coefficient

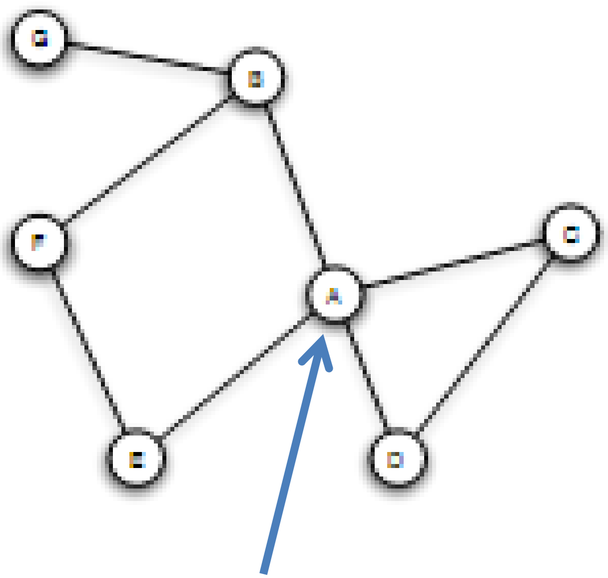
(Local) clustering coefficient for a node is the probability that two randomly selected friends of a node are friends with each other *(form a triangle)*

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} \qquad e_{jk} \in E, u_i, u_j \in N_i, k \text{ size of } N_i, N_i \text{ neigborhood of } u_i$$

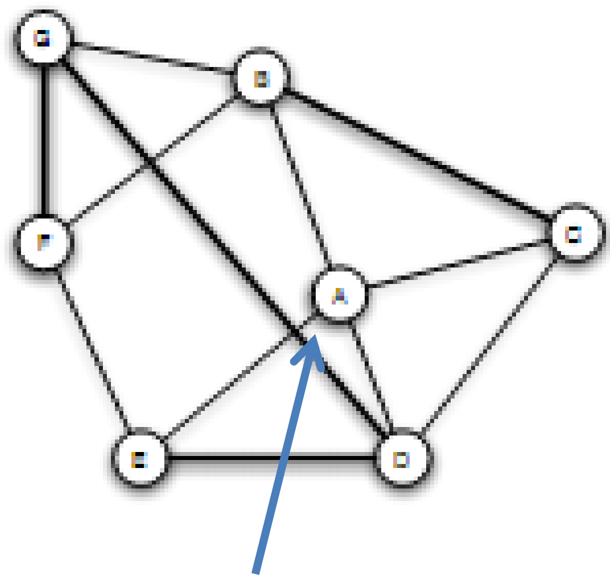
Fraction of the friends of a node that are friends with each other (i.e., connected)

$$C^{(1)} = \frac{\sum_i \text{triangles centered at node } i}{\sum_i \text{triples centered at node } i}$$

Clustering Coefficient



1/6

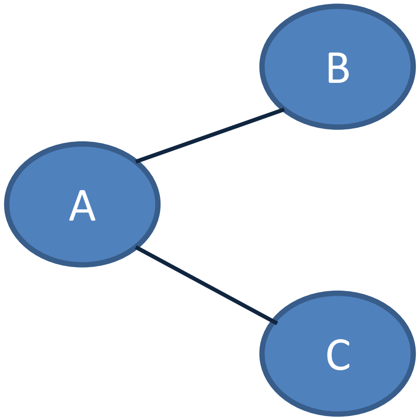


1/2

Ranges from 0 to 1

Triadic Closure

If A knows B and C, B and C are likely to become friends, but WHY?



1. Opportunity
2. Trust
3. Incentive of A (latent stress for A, if B and C are not friends, dating back to social psychology, e.g., relating low clustering coefficient to suicides)

The Strength of Weak Ties Hypothesis

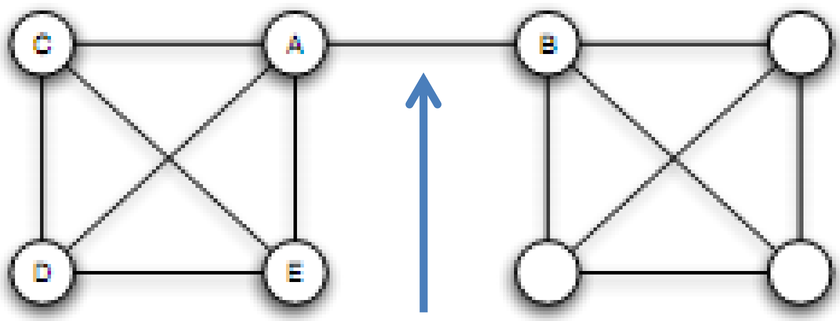
Mark Granovetter, in the late 1960s

Many people learned information leading to their current job *through personal contacts*, often described as *acquaintances* rather than *closed friends*

Two aspects

- Structural
- Local (interpersonal)

Bridges and Local Bridges



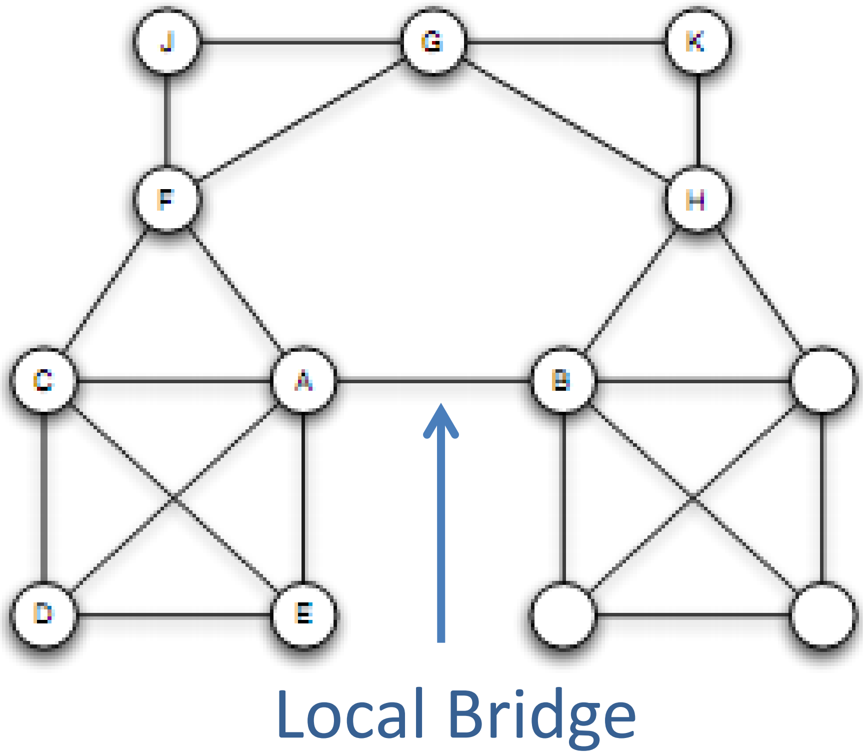
Bridge
(aka cut-edge)

An edge between A and B is a *bridge* if deleting that edge would cause A and B to lie in two different components

AB the only “route” between A and B

extremely rare in social networks

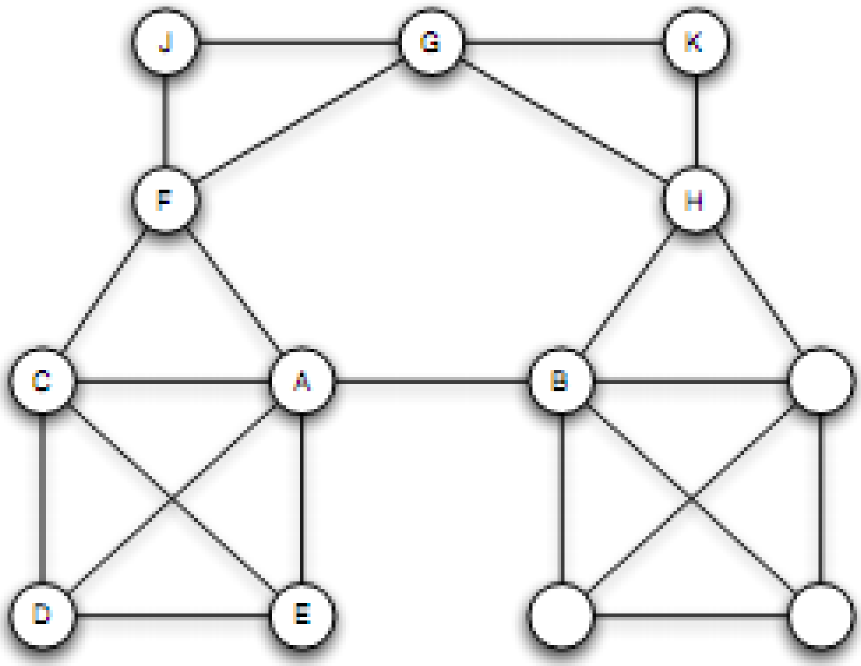
Bridges and Local Bridges



An edge between A and B is a *local bridge* if deleting that edge would increase the distance between A and B to a value strictly more than 2

Span of a local bridge: distance of the its endpoints if the edge is deleted

Bridges and Local Bridges

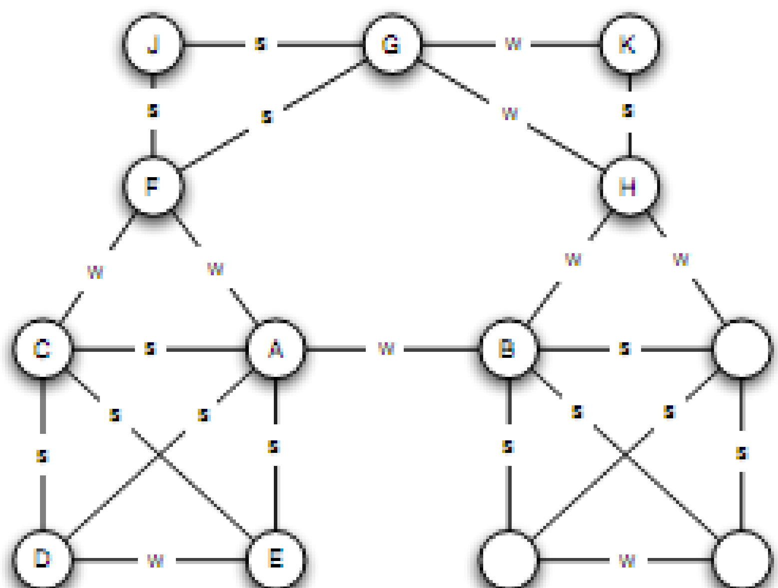


An edge is a local bridge, if an only if, it is not part of any triangle in the graph

The Strong Triadic Closure Property

- Levels of strength of a link
- Strong and weak ties
- May vary across different times and situations

Annotated graph

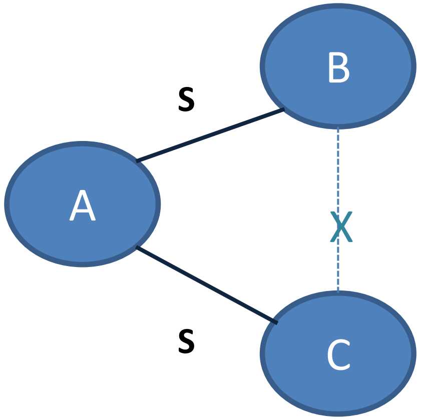


The Strong Triadic Closure Property

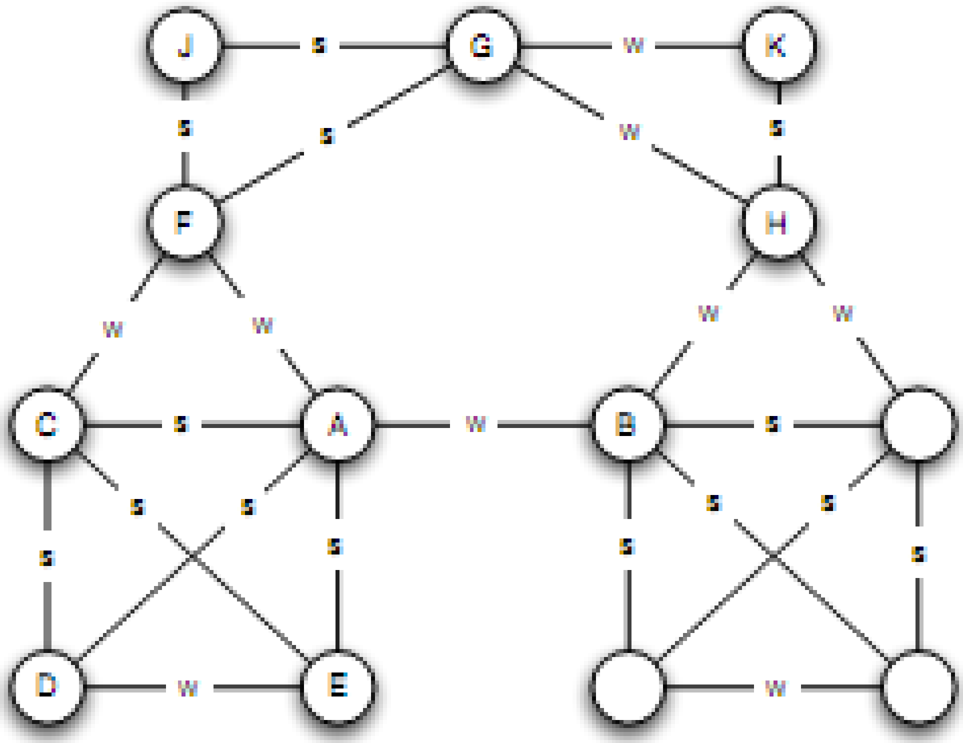
If a node A has edges to nodes B and C, then the B-C edge is <u>especially likely</u> to form if both A-B and A-C are <u>strong ties</u>

A node A violates the Strong Triadic Closure Property, if it has strong ties to two other nodes B and C, and there is no edge (strong or weak tie) between B and C.

A node A satisfies the Strong Triadic Property if it does not violate it



The Strong Triadic Closure Property



Local Bridges and Weak Ties

Local distinction: weak and strong ties ->

Global structural distinction: local bridges or not

Claim:

If a node A in a network *satisfies the Strong Triadic Closure* and is involved in *at least two strong ties*, then any *local bridge* it is involved in must be a *weak tie*

Proof: by contradiction

Relation to job seeking?

The role of simplifying assumptions:

- Useful when they lead to statements robust in practice, making sense as qualitative conclusions that hold in approximate forms even when the assumptions are relaxed
- Stated precisely, so possible to test them in real-world data
- A framework to explain surprising facts

Tie Strength and Network Structure in Large-Scale Data

How to test these prediction on large social networks?

Tie Strength and Network Structure in Large-Scale Data

Communication network: "who-talks-to-whom"
Strength of the tie: time spent talking during an observation period

Cell-phone study [Omnela et. al., 2007]

"who-talks-to-whom network", covering 20% of the national population

- Nodes: cell phone users
- Edge: if they make phone calls to each other in both directions over 18-week observation periods

<u>Is it a "social network"?</u>
Cells generally used for personal communication + no central directory, thus cell-phone numbers exchanged among people who already know each other
Broad structural features of large social networks (*giant component*, 84% of nodes)

Generalizing Weak Ties and Local Bridges

So far:

✓ Either weak or strong

✓ Local bridge or not

Tie Strength: Numerical quantity (= number of min spent on the phone)

Quantify "local bridges", how?

Generalizing Weak Ties and Local Bridges

Bridges

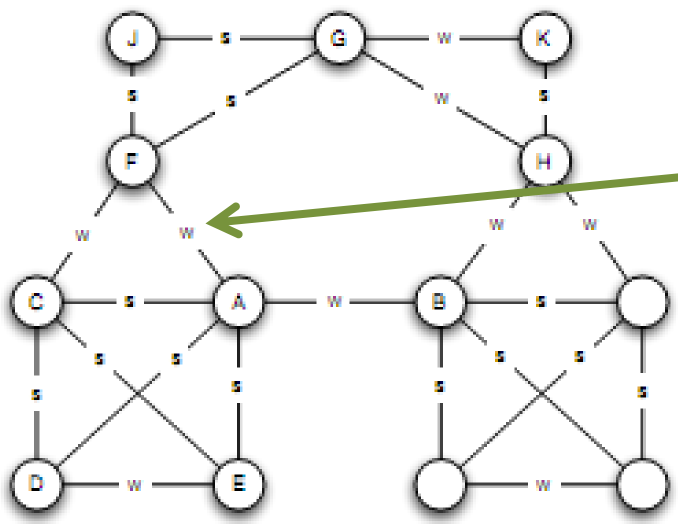
"almost" local bridges

Neighborhood overlap of an edge e_{ij}

(*) In the denominator we do not count A or B themselves

$$\frac{|N_i \cap N_j|}{|N_i \cup N_j|}$$

Jaccard coefficient



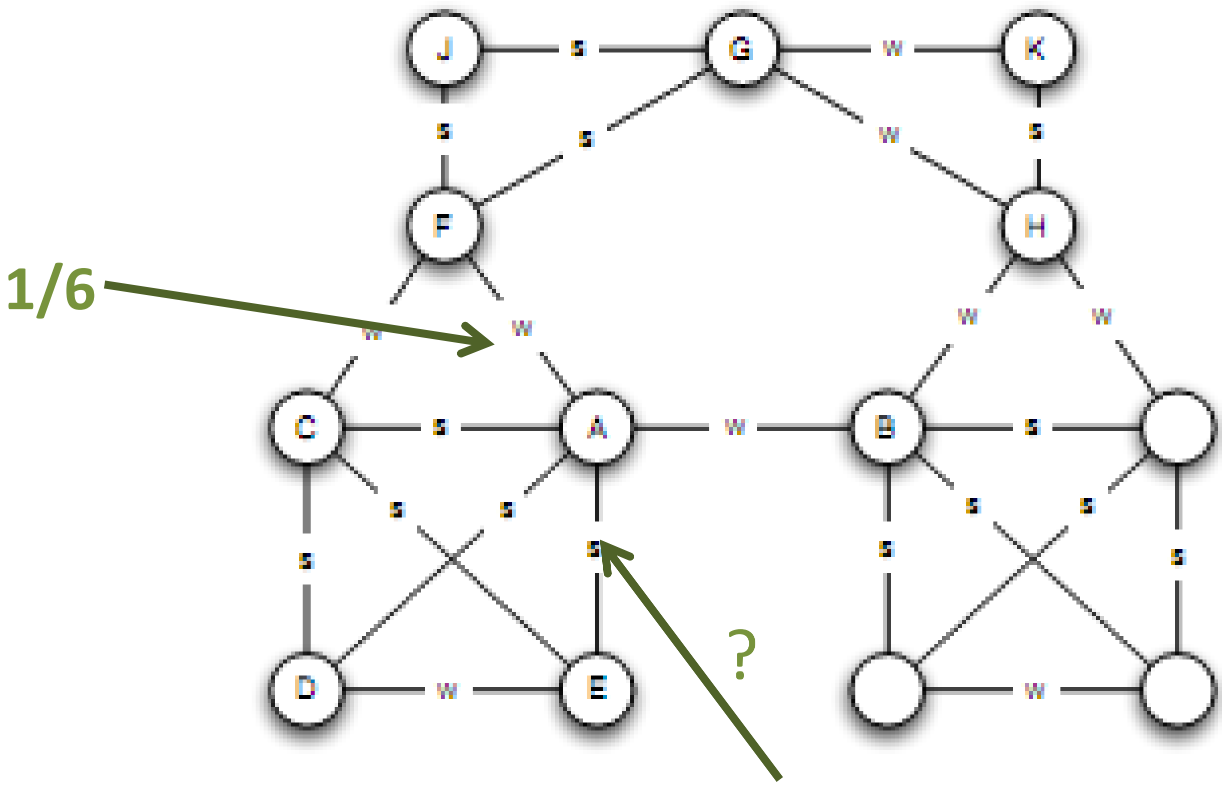
A: B, E, D, C
F: C, J, G

1/6

When is this value 0?

Generalizing Weak Ties and Local Bridges

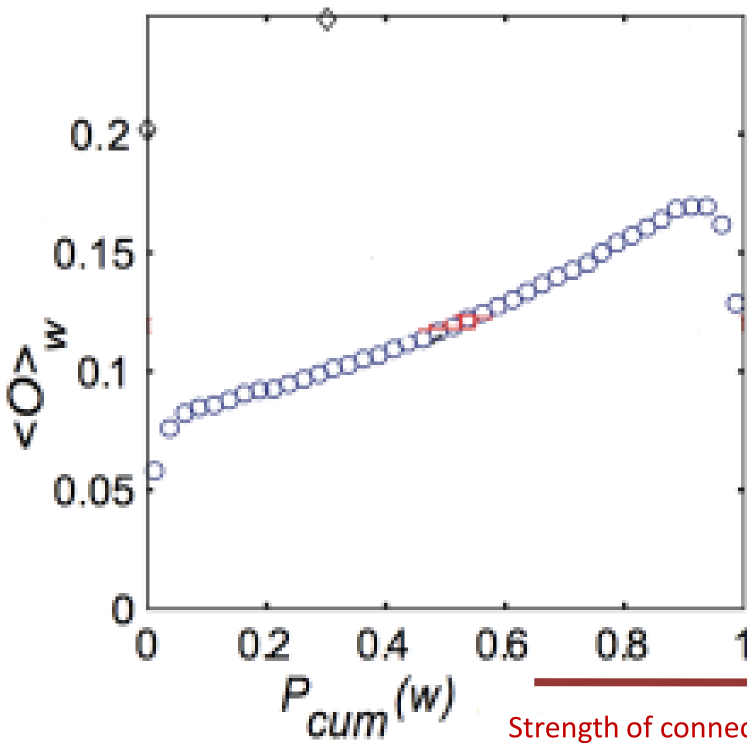
Neighborhood overlap = 0: edge is a local bridge
Small value: “almost” local bridges



Generalizing Weak Ties and Local Bridges:
Empirical Results

How the neighborhood overlap of an edge depends on its strength

(Hypothesis: the strength of weak ties predicts that neighborhood overlap should grow as tie strength grows)



(*) Some deviation at the right-hand edge of the plot

sort the edges -> for each edge at which percentile

Strength of connection (function of the percentile in the sorted order)

Generalizing Weak Ties and Local Bridges:
Empirical Results

How to test the following global (macroscopic) level hypothesis:

Hypothesis: weak ties serve to link different tightly-knit communities that each contain a large number of stronger ties

Generalizing Weak Ties and Local Bridges: Empirical Results

Delete edges from the network one at a time

- Starting with the strongest ties and working downwards in order of tie strength
    - giant component shrank steadily

-Starting with the weakest ties and upwards in order of tie strength
    - giant component shrank more rapidly, broke apart abruptly as a critical number of weak ties were removed

Social Media and Passive Engagement

People maintain large explicit lists of friends

Test:

How *online activity* is distributed across *links of different strengths*

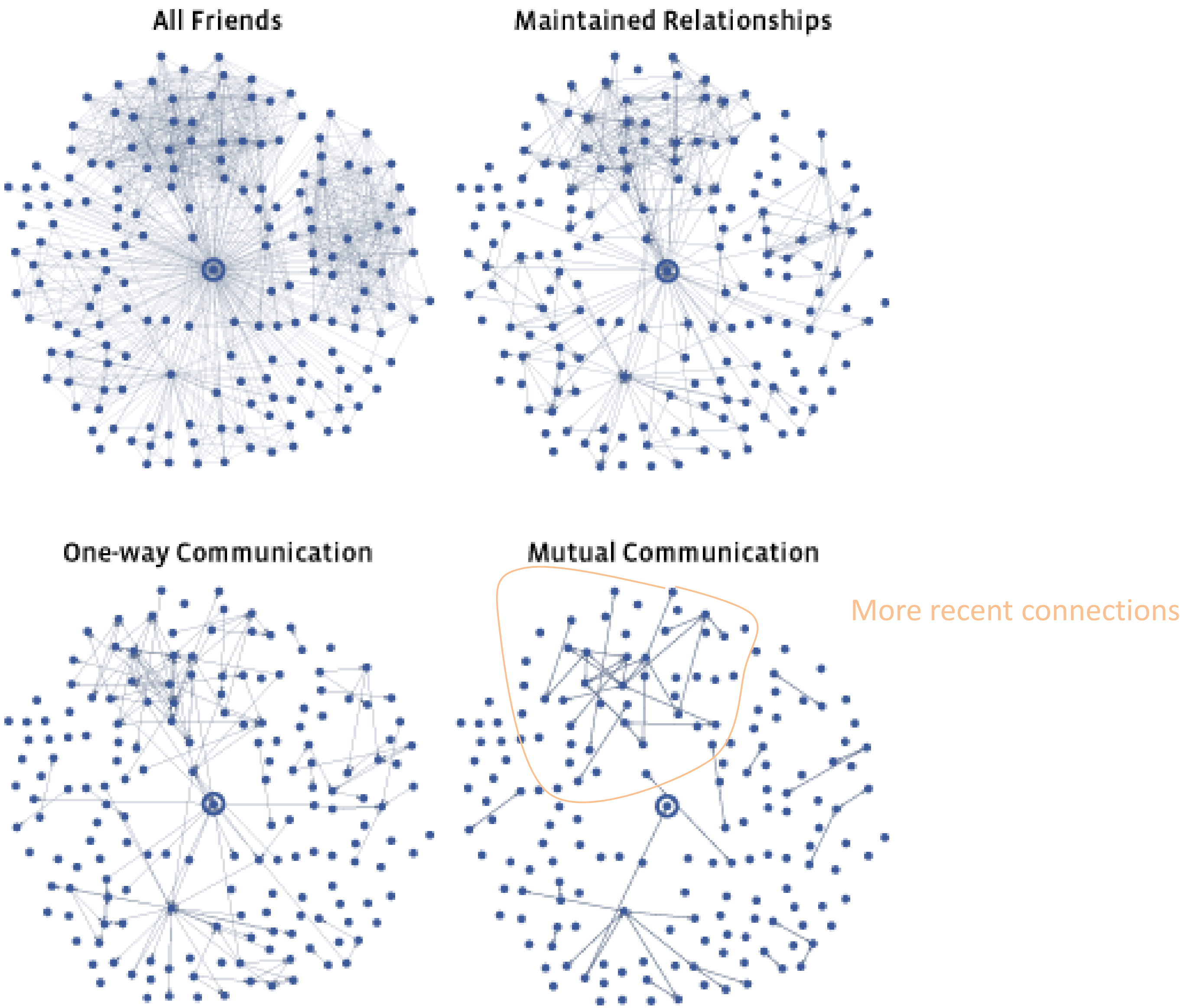
Tie Strength on Facebook

Cameron Marlow, et al, 2009
At what extent each link was used for social interactions

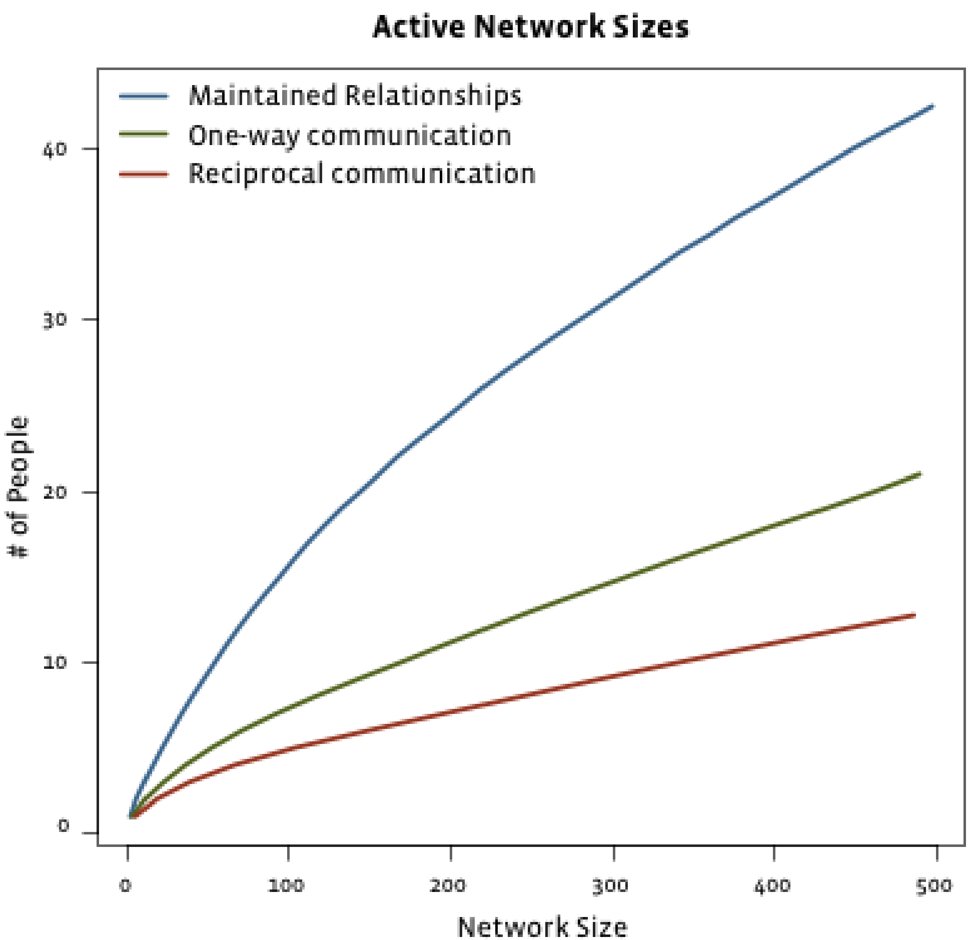
Three (not exclusive) kinds of ties (links)

1. Reciprocal (mutual) communication: both send and received messages to friends at the other end of the link
2. One-way communication: the user send one or more message to the friend at the other end of the link
3. Maintained relationship: the user followed information about the friend at the other end of the link (click on content via News feed or visit the friend profile more than once)

Tie Strength on Facebook



Tie Strength on Facebook



Total number of friends

Even for users with very large number of friends

- actually communicate : 10-20
- number of friends follow even passively <50

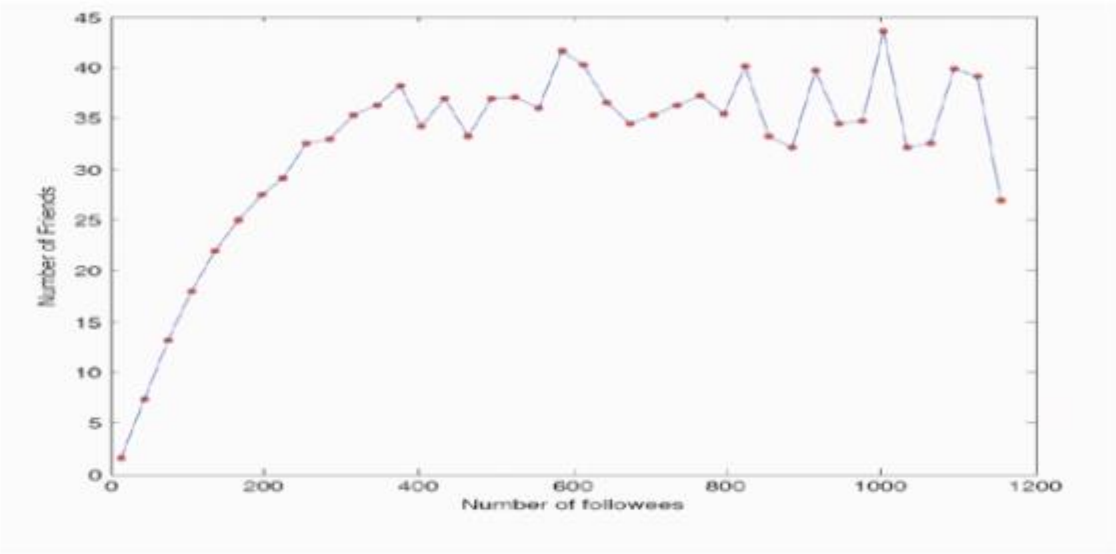
Passive engagement (keep up with friends by reading about them even in the absence of communication)

Tie Strength on Twitter

Huberman, Romero and Wu, 2009

Two kinds of links

- Follow
- Strong ties (friends): users to whom the user has *directed at least two messages* over the course if the observation period



Social Media and Passive Engagement

- Strong ties require continuous investment of time and effort to maintain (as opposed to weak ties)

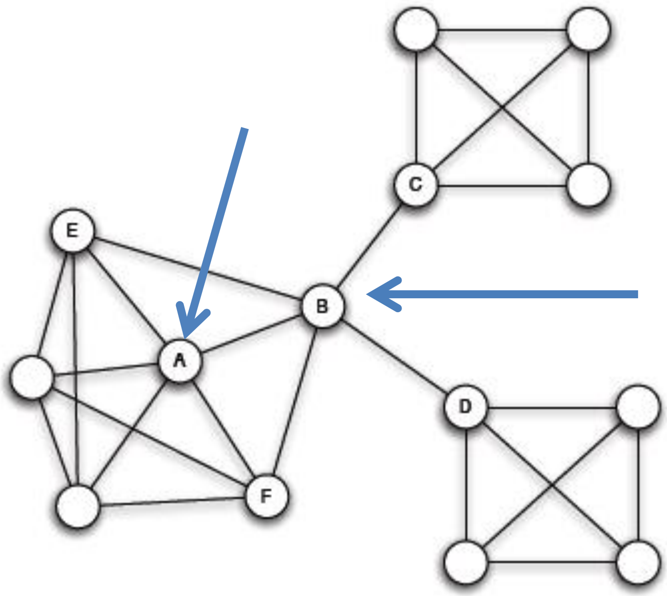
- Network of strong ties still remain sparse

- How different links are used to convey information

Closure, Structural Holes and Social Capital

Different roles that *nodes* play in this structure

Access to edges that span different groups is not equally distributed across all nodes

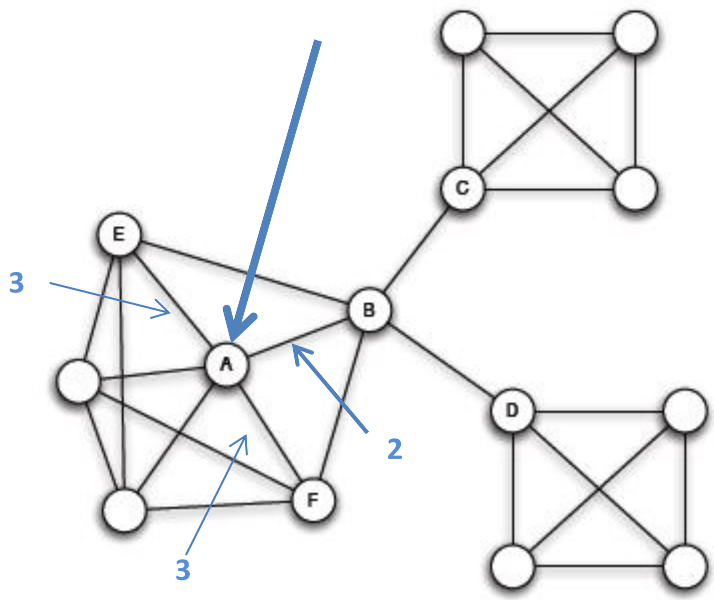


Embeddedness

A has a large clustering coefficient

- Embeddedness of an edge: number of common neighbors of its endpoints (neighborhood overlap, local bridge if 0)

For A, all its edges have significant embeddedness



(sociology) if two individuals are connected by an embedded edge => trust

- "Put the interactions between two people on display"

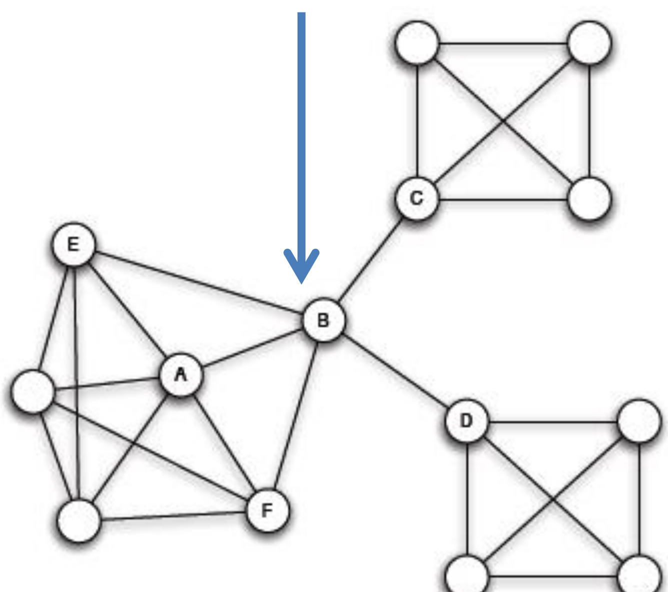
Structural Holes

(sociology) B-C, B-D much riskier, also, possible contradictory constraints
Success in a large cooperation correlated to access to local bridges

B “spans a structural hole”

- B has access to information originating in multiple, non interacting parts of the network
- An amplifier for creativity
- Source of power as a social “gate-keeping”

Social capital



MORE ON LINK FORMATION:

AFFILIATIONS AND MEASUREMENTS

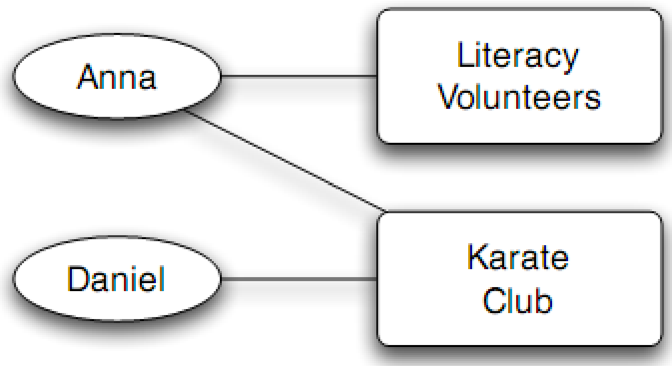
Affiliation

A larger network that contains both people and context as nodes

foci

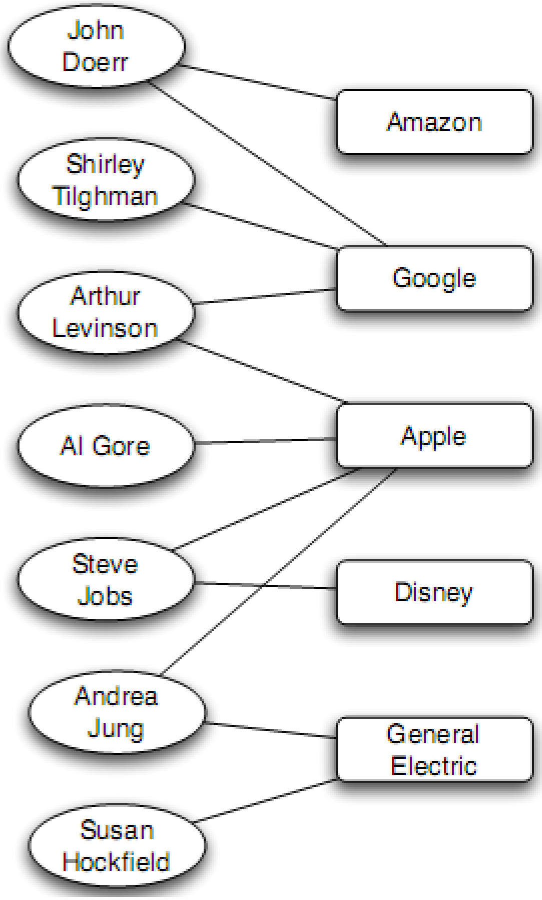
Affiliation network:

A bipartite graph
A node for each person and a node for each focus
An edge between a person A and focus X, if A participates in X



Affiliation

Example:
Board of directors



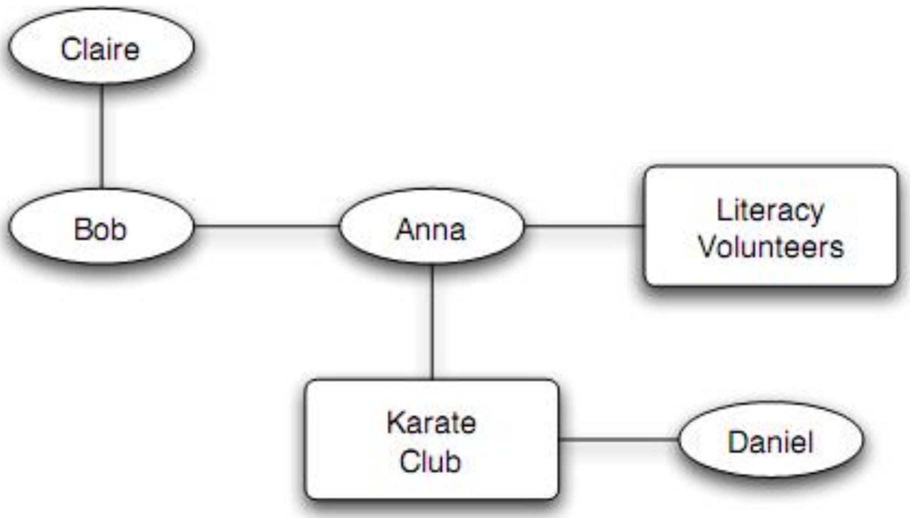
- Companies implicitly links by having the same person sit on both their boards
- People implicitly linked by serving together on a aboard
- Other contexts, president of two major universities and a former Vice-President

Co-evolution of Social and Affiliation Networks

Social Affiliation Network

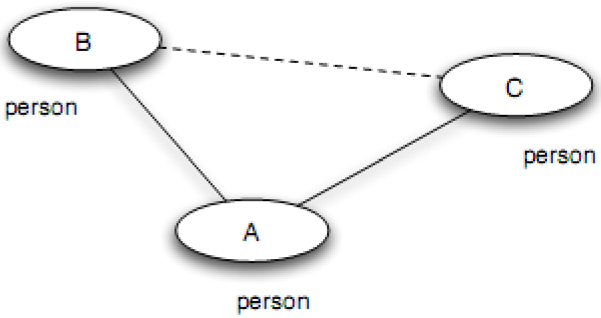
Two type of edges:

1. *Friendship:* between two people
2. *Participation:* between a person and a focus

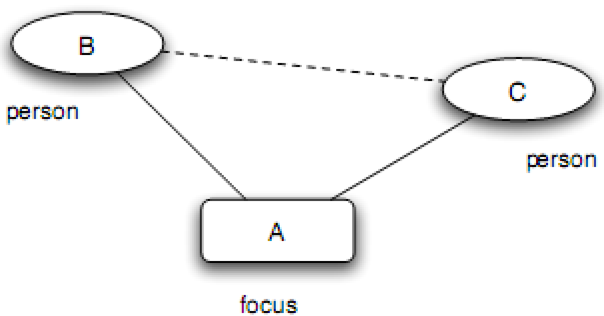


- Co-evolution reflect the interplay of selection and social influence: if two people in a shared focus opportunity to become friends, if friends, influence each other foci.

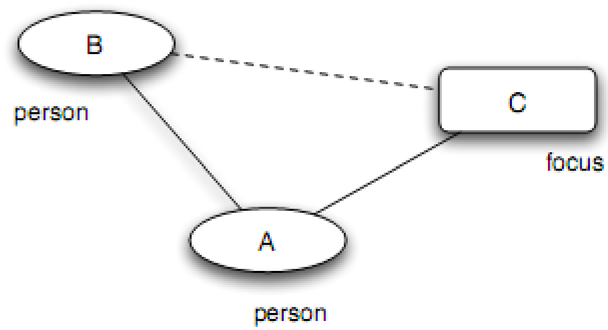
Co-evaluation of Social and Affiliation Networks: Closure process



Triadic closure: (two people with a friend in common - A introduces B to C)

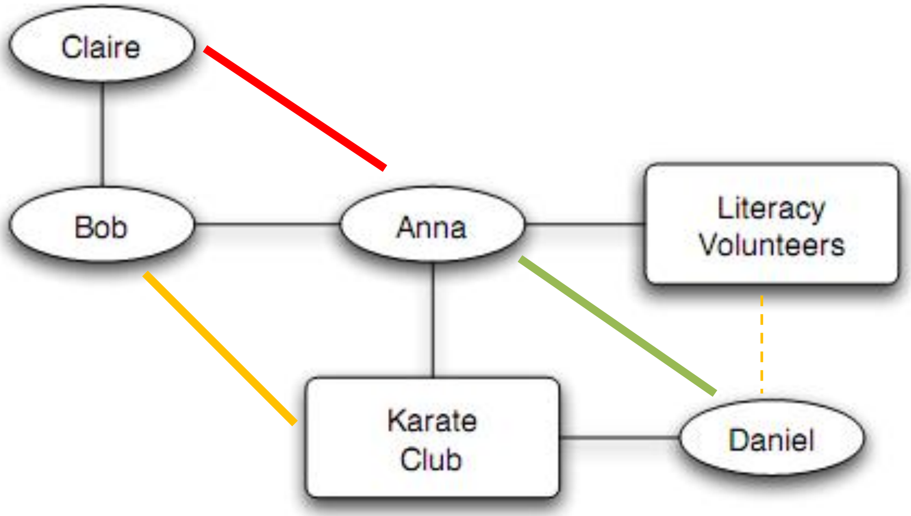


Focal closure: (two people with a focus in common - focus A introduces B to C) (selection)

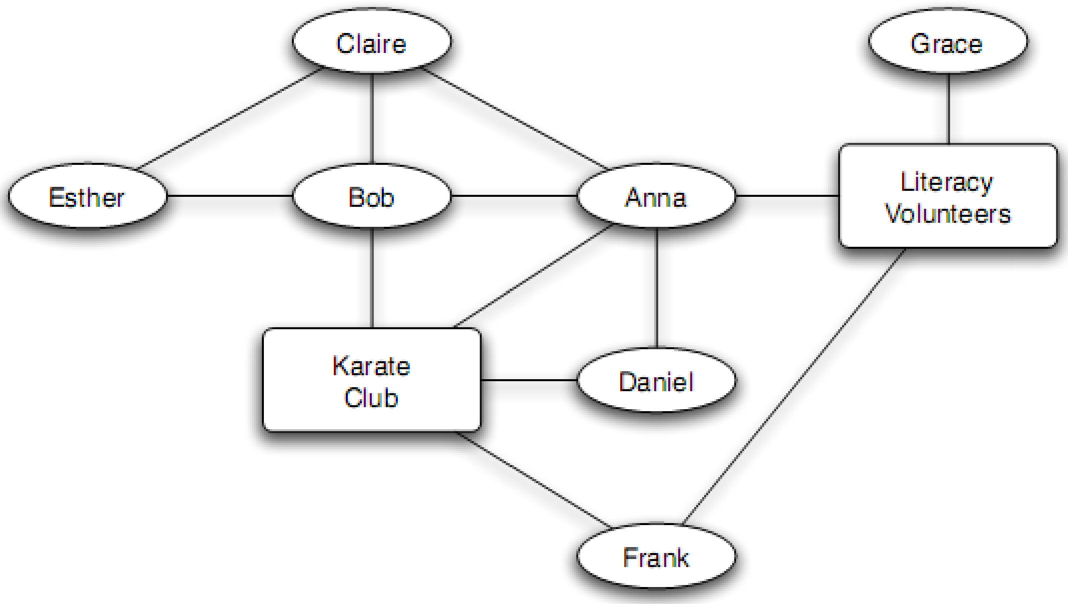


Membership closure: (a person joining a focus that a friend is already involved in - A introduces focus C to B) (social influence)

Co-evaluation of Social and Affiliation Networks



Example



Claire
Grace
Esther
Bob
Anna
Literacy Volunteers
Karate Club
Daniel
Frank

Tracking Link Formation in Online Data:
triadic closure

Triadic closure:

- How much more likely is a link to form between two people if they have *a friend* in common
- How much more likely is a link to form between two people if they have *multiple* friends in common?

Tracking Link Formation in Online Data:
triadic closure

Take two snapshots of the network at different times:

I. For each *k*, identify all pairs of nodes that have exactly *k* friends in common in the first snapshot, but who are not directly connected

II. Define *T(k)* to be the fraction of these pairs that have formed an edge by the time of the second snapshot

III. Plot *T(k)* as a function of *k*

T(0): rate at which link formation happens when it does not close any triangle

T(k): the rate at which link formation happens when it does close a triangle (*k* common neighbors, triangles)

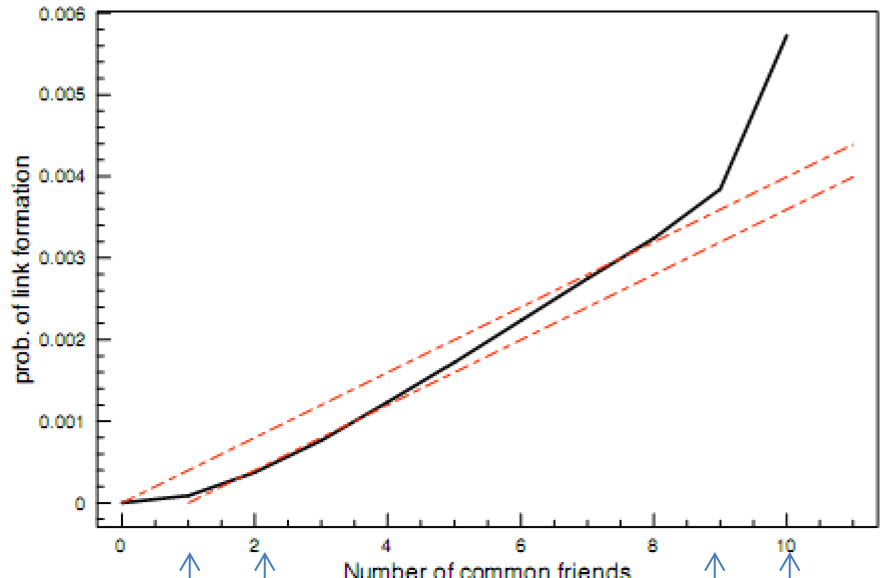
Tracking Link Formation in Online Data: triadic closure

E-mail ("who-talks-to-whom" dataset type)
Among 22,000 undergrad and grad students (large US university)
For 1-year

Network evolving over time

- At each instance (snapshot), two people join, if they have exchanged e-mail in each direction at some point in the past 60 days
- Multiple pairs of snapshots ->
- Built a curve for T(k) on each pair, then average all the curves

Snapshots – one day apart (average probability that two people form a link per day)



From 0 to 1 to 2 friends

✓ Almost linear

From 8 to 9 to 10 friend (but occurs on a much smaller population)

Having two common friends produces significantly more than twice the effect compared to a single common friend

Tracking Link Formation in Online Data: triadic closure

Baseline model:
Assume triadic closure:
Each common friend two people have gives them an independent probability *p* of forming a link each day

For two people with *k* friend in common,
Probability *not* forming a link on any given day

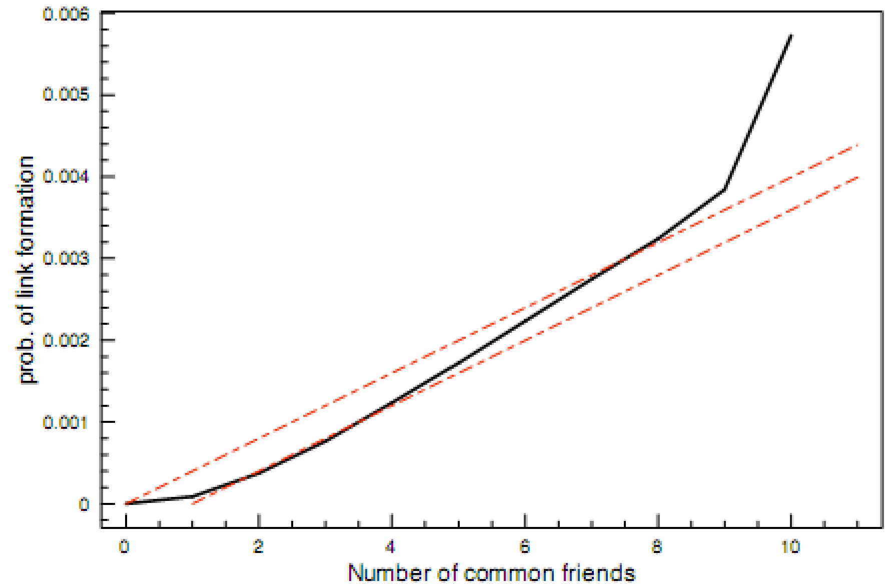
$$(1-p)^k$$

Probability *forming a link* on any given day

$$T_{baseline1}(k) = 1 - (1-p)^k$$

Given the small absolute effect of the first common friend in the data

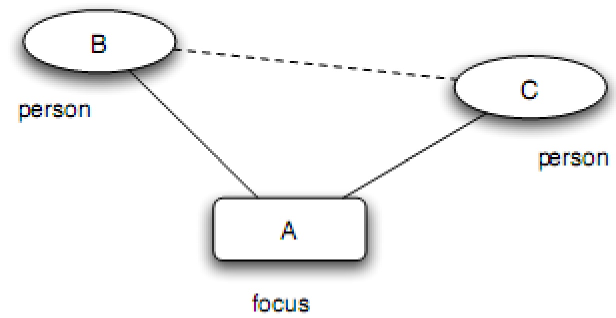
$$T_{baseline2}(k) = 1 - (1-p)^{k-1}$$

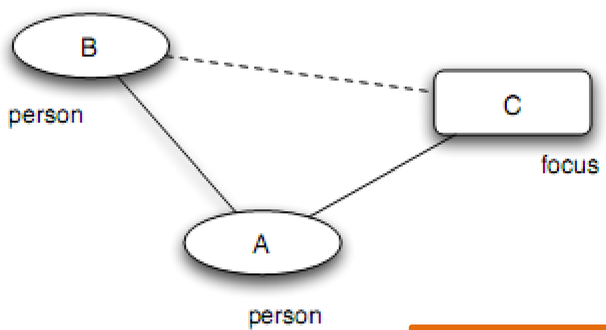


Qualitative similar (linear), but independent assumption too simple

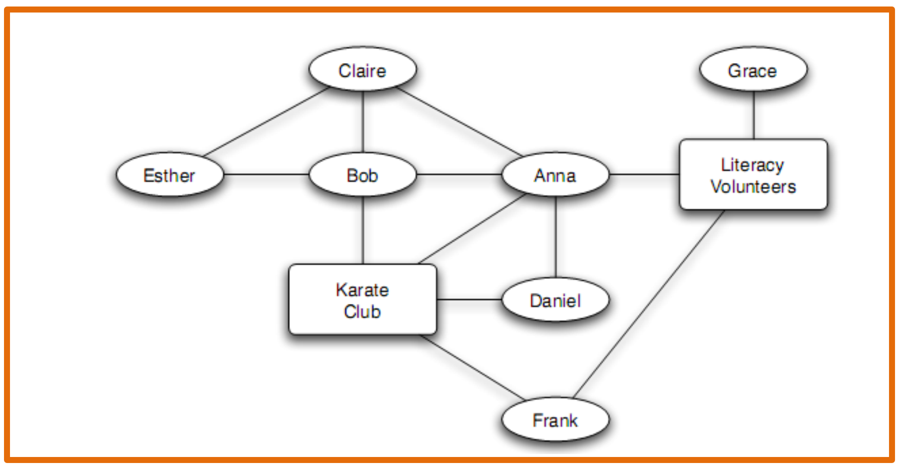
Tracking Link Formation in Online Data: focal and membership closure

Focal closure: what is the probability that two people form a link as a function of the *number of foci* that are jointly affiliated with





Membership closure: what is the probability that a person becomes involved with a particular focus as a function of the *number of friends* who are already involved in it?

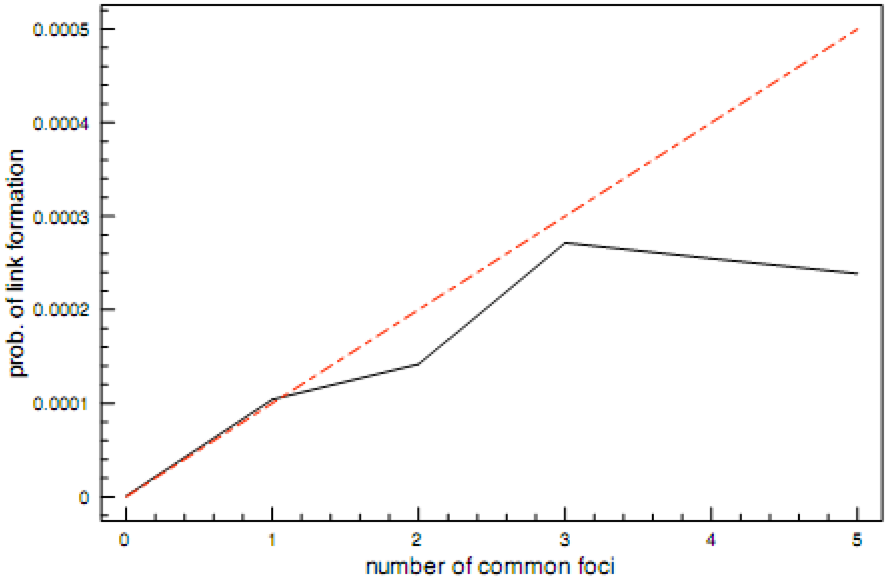


Tracking Link Formation in Online Data: focal closure

E-mail ("who-talks-to-whom" dataset type)

Use the *class schedule* of each student

Focus: class (common focus – a class together)



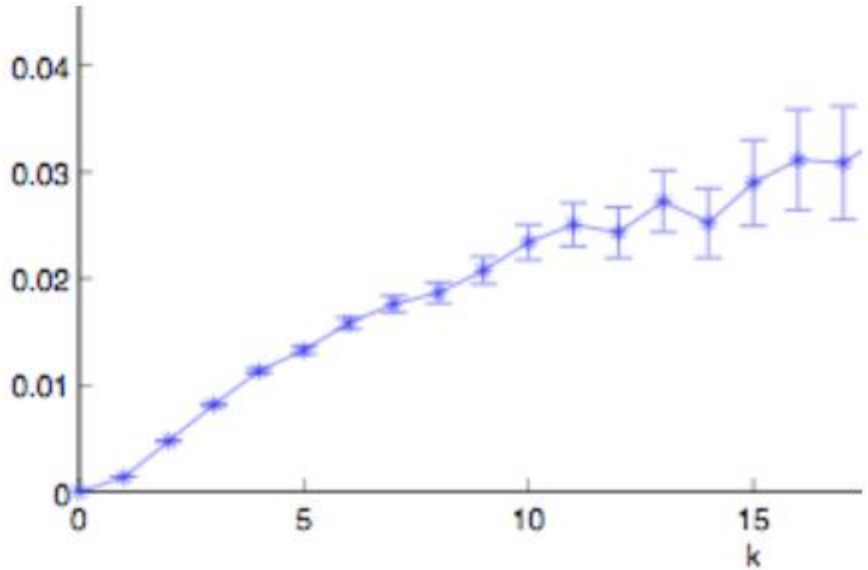
A single shared class same effect as a single shared friend, then different

Subsequent shared classes after the first produce a diminishing returns effect

Tracking Link Formation in Online Data: membership closure

Node: Wikipedia editor who maintains a user account and user talk page
Link: if they have communicated by one user writing on the user talk page of the other
Focus: Wikipedia article
Association to focus: edited the article

Again, an initial increasing effect: the probability of editing a Wikipedia article is more than twice as large when you have two connections into the focus than one



✓ Also, multiple effects can operate simultaneously

POSITIVE AND NEGATIVE TIES

Structural Balance

What about negative edges?

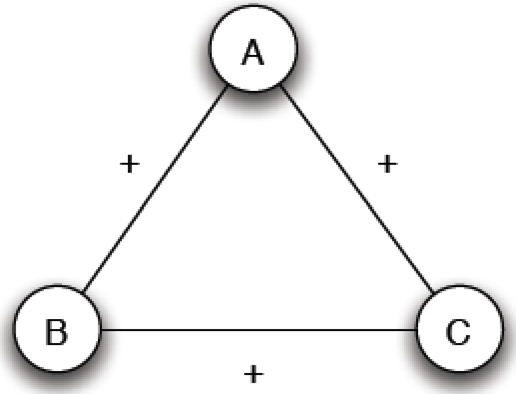
Initially, a complete graph (or clique): every edge either + or -

Let us first look at individual triangles

- Lets look at 3 people => 4 cases
- See if all are equally possible (local property)

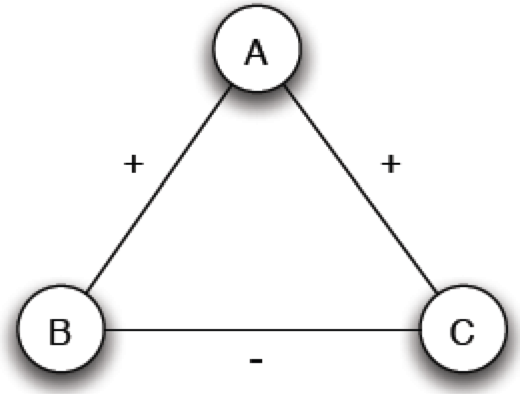
Structural Balance

Case (a): 3 +



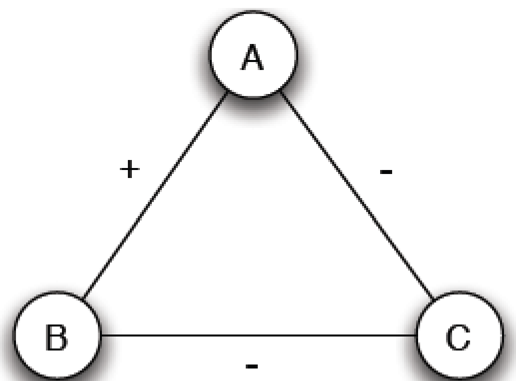
Mutual friends

Case (b): 2 +, 1 -



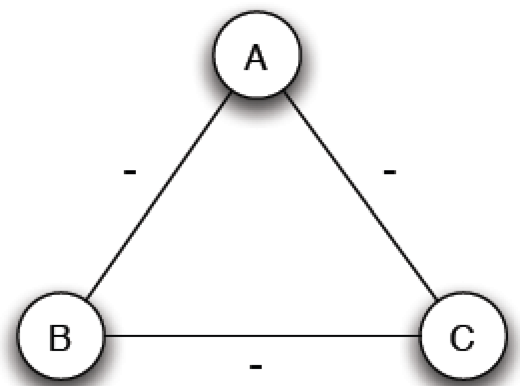
A is friend with B and C, but B and C do not get well together

Case (c): 1 +, 2 -



A and B are friends with a mutual enemy

Case (d): 3 -

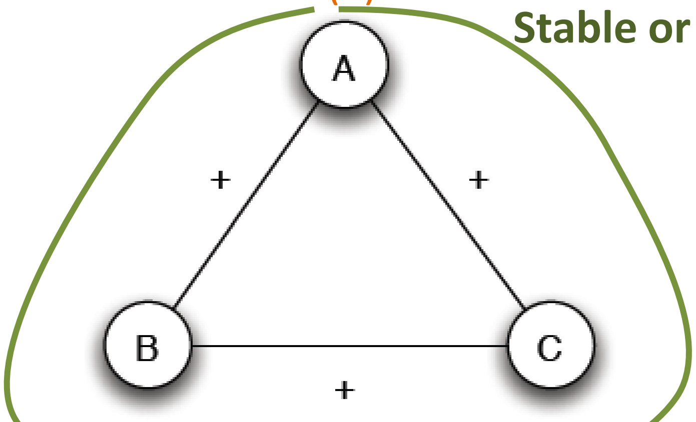


Mutual enemies

Structural Balance

Case (a): 3 +

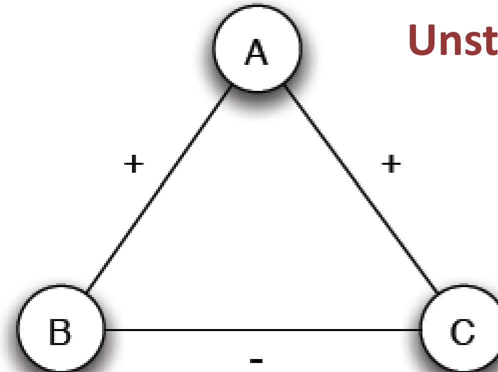
Stable or balanced



Mutual friends

Case (b): 2 +, 1 -

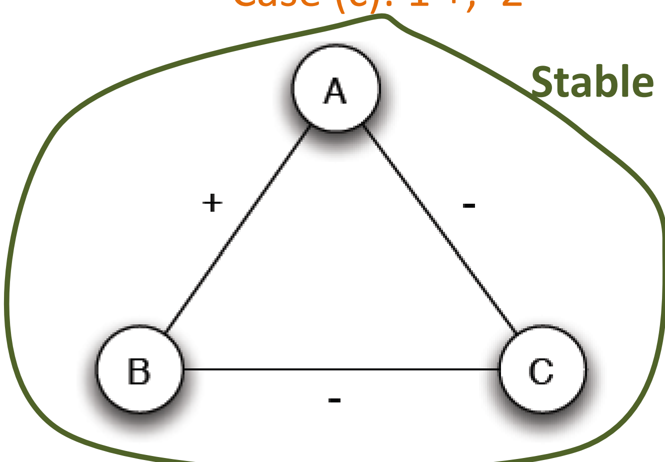
Unstable



A is friend with B and C, but B and C do not get well together
Implicit force to make B and C friends (- => +) or turn one of the + to -

Case (c): 1 +, 2 -

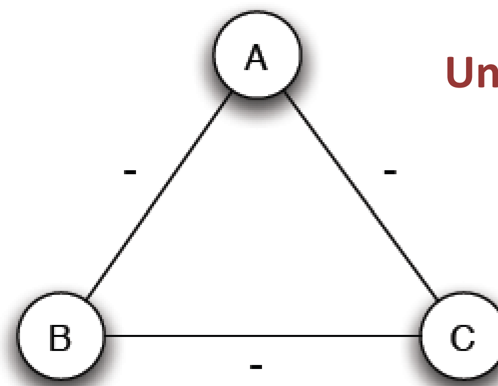
Stable or balanced



A and B are friends with a mutual enemy

Case (d): 3 -

Unstable

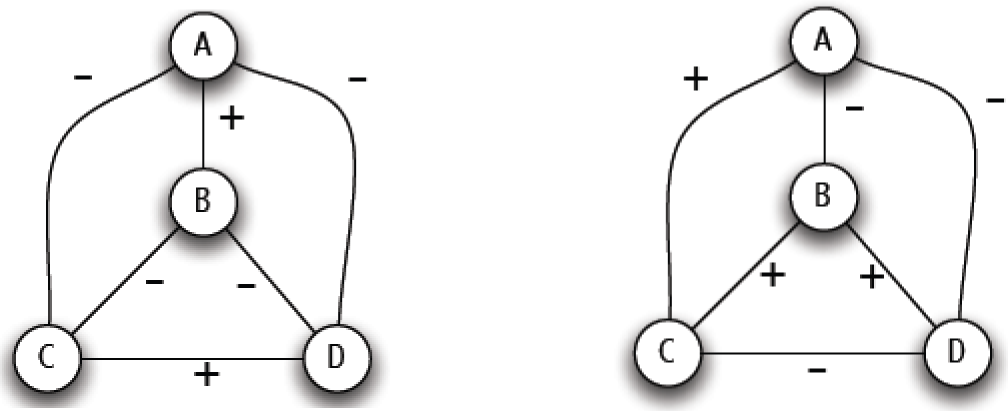


Mutual enemies
Forces to team up against the third (turn 1 – to +)

Structural Balance

A labeled complete graph is balanced if every one of its triangles is balanced

Structural Balance Property: For every set of three nodes, if we consider the three edges connecting them, either all three of these are labeled +, or else exactly one of them is labeled – (odd number of +)

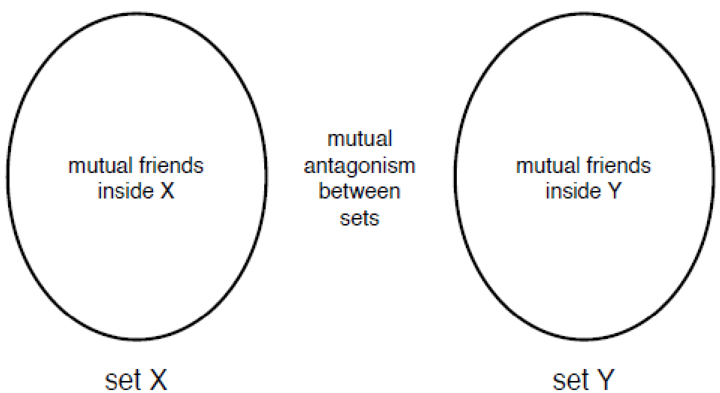


What does a balanced network look like?

The Structure of Balanced Networks

Balance Theorem: If a labeled *complete* graph is balanced,
(a) all pairs of nodes are friends, or
(b) the nodes can be divided *into two groups* X and Y, such that every pair of nodes in X like each other, every pair of nodes in Y like each other, and every one in X is the enemy of every one in Y.

From a local to a global property

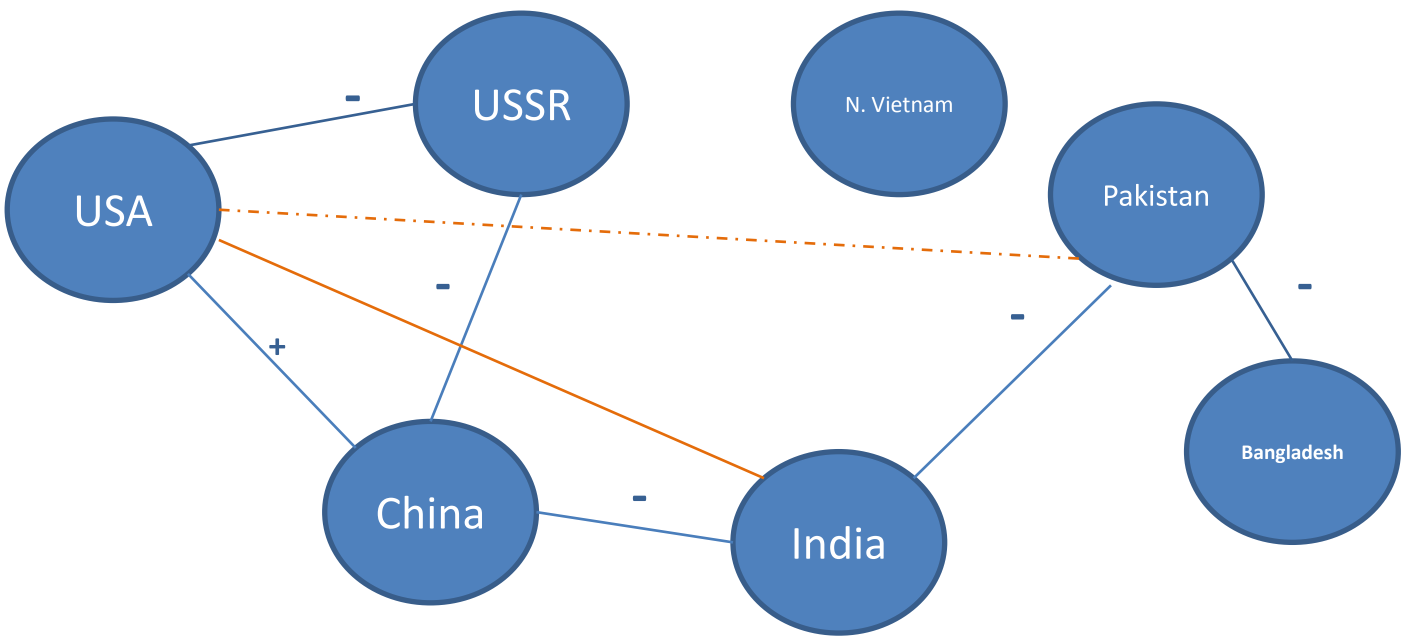


Proof ...

Applications of Structural Balance

- ✓ How a network evolves over time
- ✓ Political science: International relationships (I)

The conflict of Bangladesh's separation from Pakistan in 1972 (1)

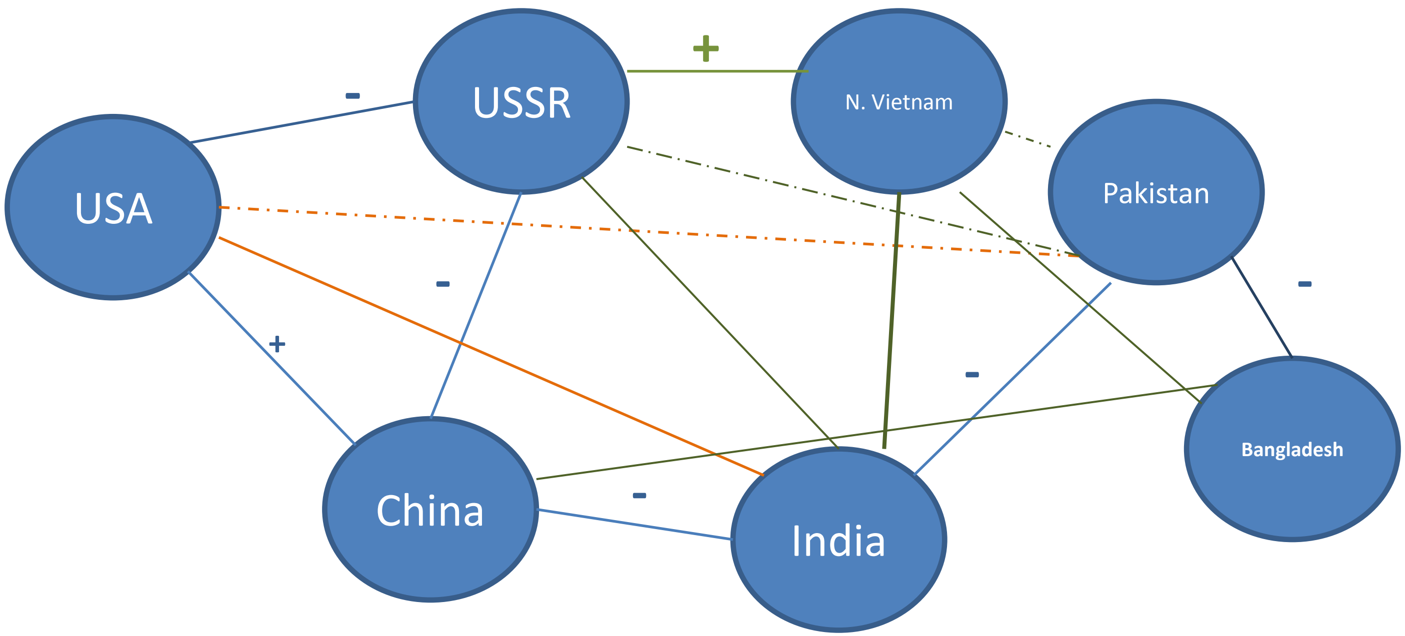


USA support to Pakistan?

Applications of Structural Balance

✓ International relationships (I)

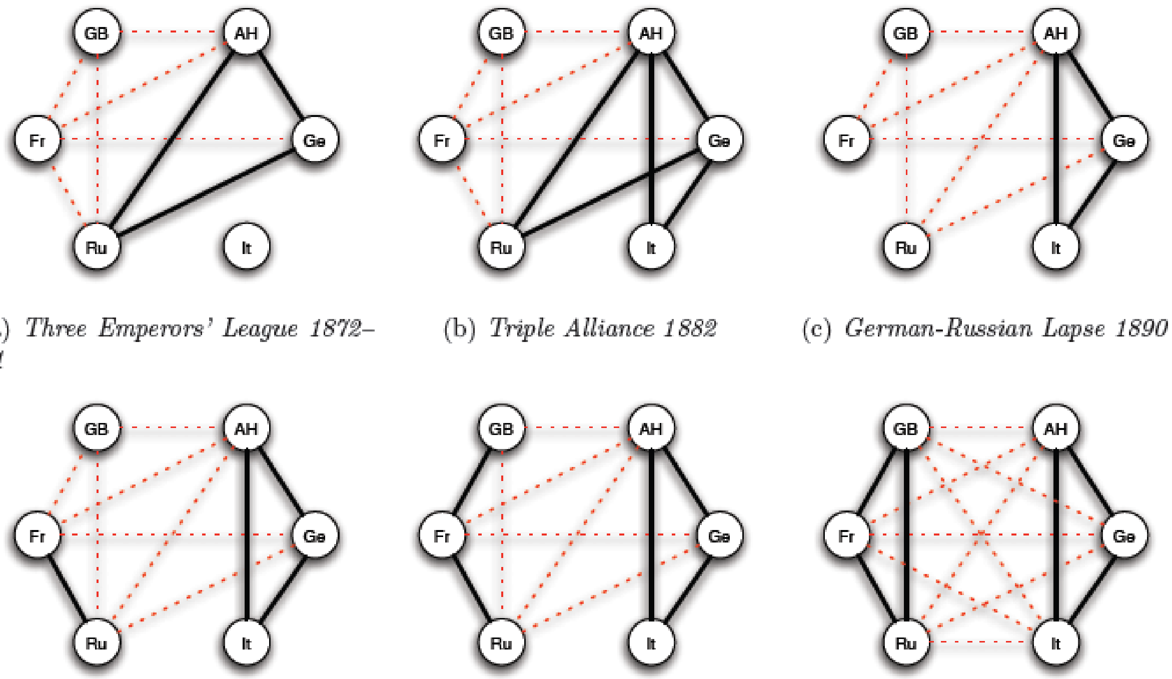
The conflict of Bangladesh's separation from Pakistan in 1972 (II)



China?

Applications of Structural Balance

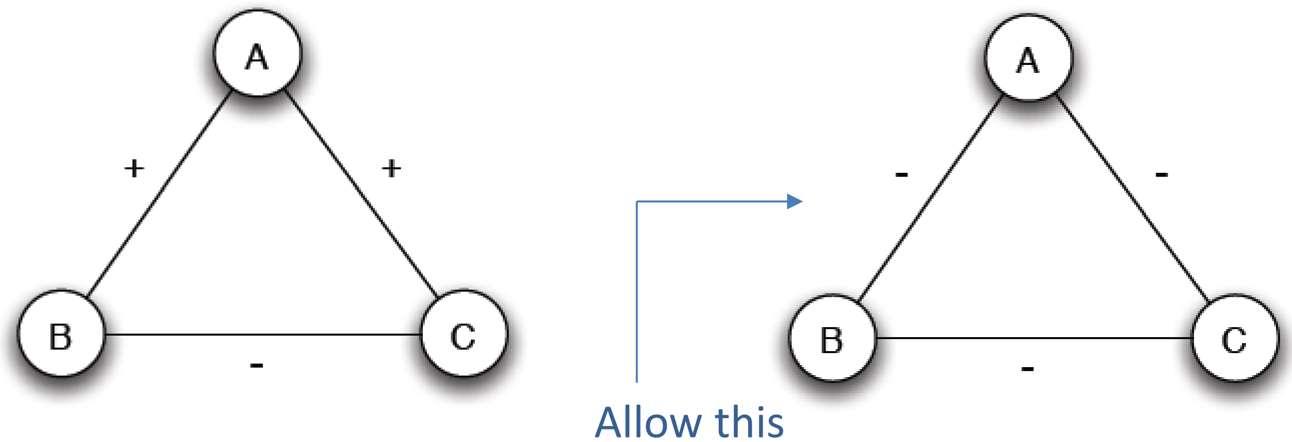
✓ International relationships (II)



(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*

Figure 5.5: The evolution of alliances in Europe, 1872-1907 (the nations GB, Fr, Ru, It, Ge, and AH are Great Britain, France, Russia, Italy, Germany, and Austria-Hungary respectively). Solid dark edges indicate friendship while dotted red edges indicate enmity. Note how the network slides into a balanced labeling — and into World War I. This figure and example are from Antal, Krapivsky, and Redner [20].

A Weaker Form of Structural Balance



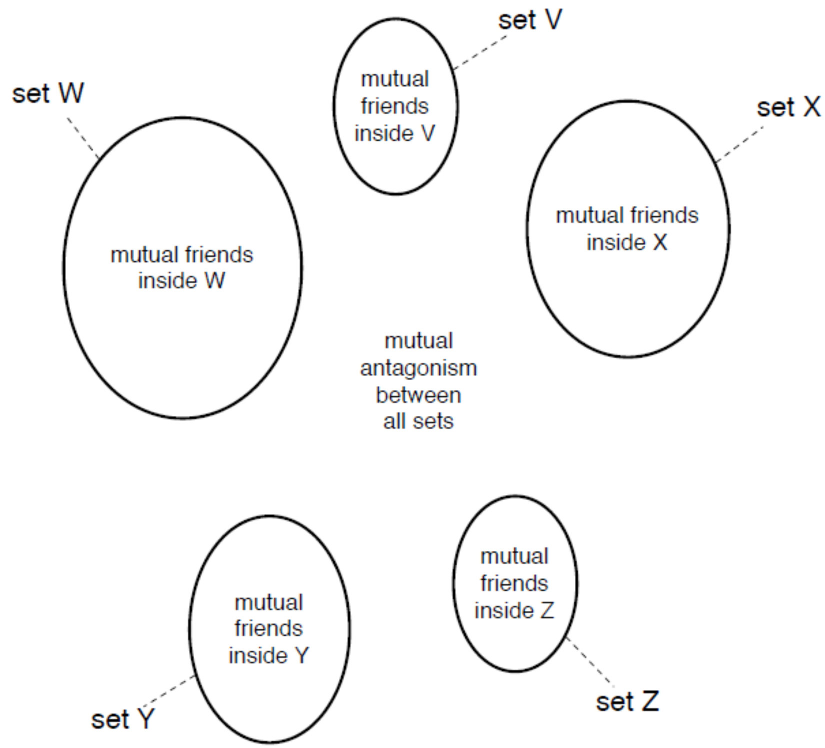
Weak Structural Balance Property: There is no set of three nodes such that the edges among them consist of exactly two positive edges and one negative edge

A Weaker Form of Structural Balance

Weakly Balance Theorem: If a labeled complete graph is weakly balanced, 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 …

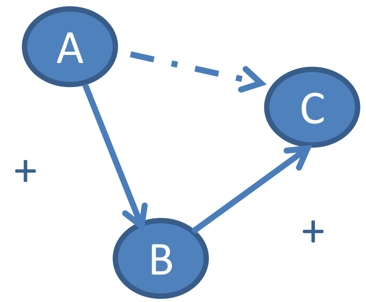
A Weaker Form of Structural Balance



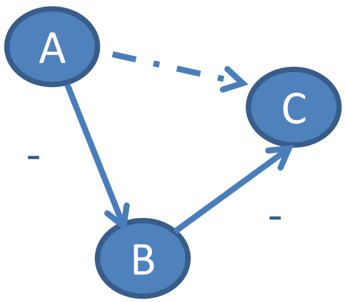
Trust, distrust and online ratings

Evaluation of products and trust/distrust of other users

Directed Graphs



A trusts B, B trusts C, A ? C



A distrusts B, B distrusts C, A ? C

If distrust enemy relation, +

A distrusts means that A is better than B, -

Depends on the application

Rating political books or

Consumer rating electronics products

Generalizing

1. Non-complete graphs

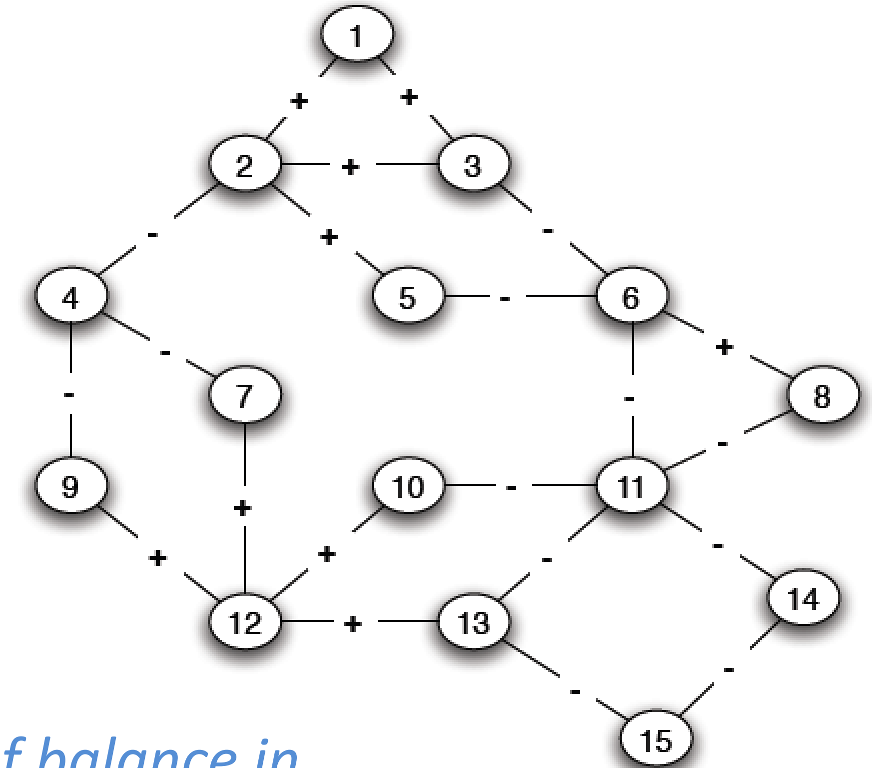
2. Instead of all triangles, “most” triangles, approximately divide the graph

We shall use the original (“non-weak” definition of structural balance)

Structural Balance in Arbitrary Graphs

Thee possible relations

- Positive edge
- Negative edge
- Absence of an edge

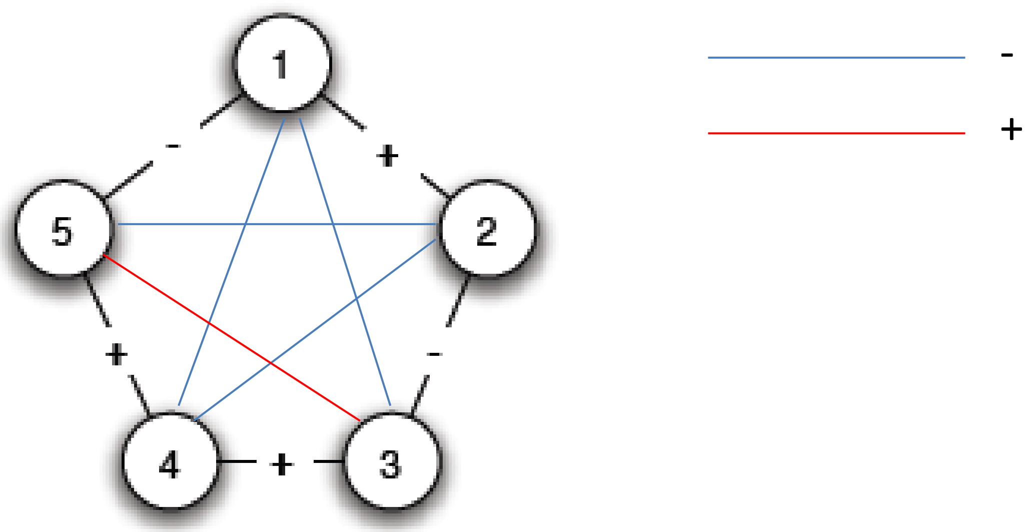


What is a good definition of balance in a non-complete graph?

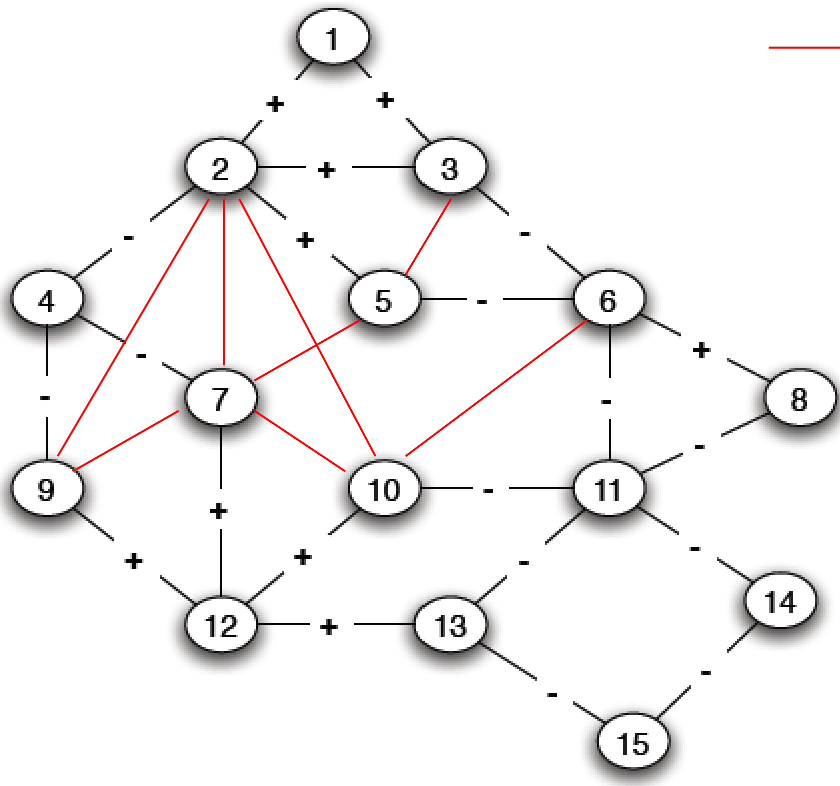
Balance Definition for General Graphs

1. Based on triangles (local view)
2. Division of the network (global view)

A (non-complete) graph is balanced if it can be completed by adding edges to form a signed complete graph that is balanced



Balance Definition for General Graphs

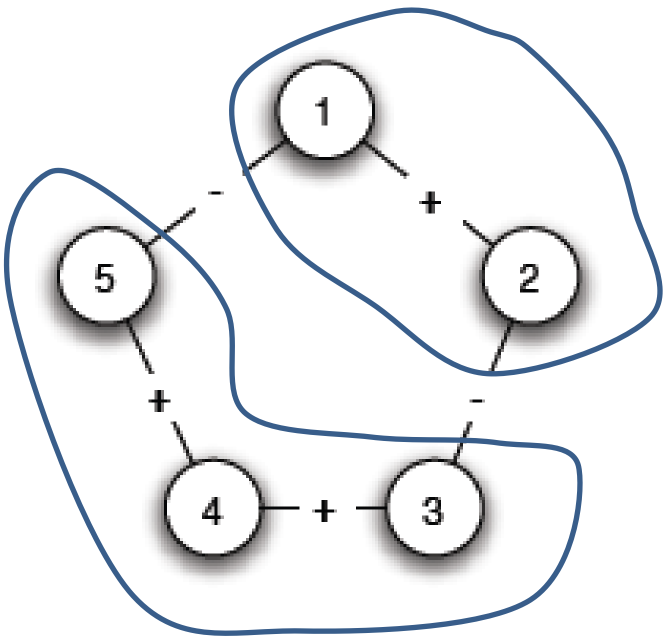


+

Balance Definition for General Graphs

1. Based on triangles (local view)
2. Division of the network (global view)

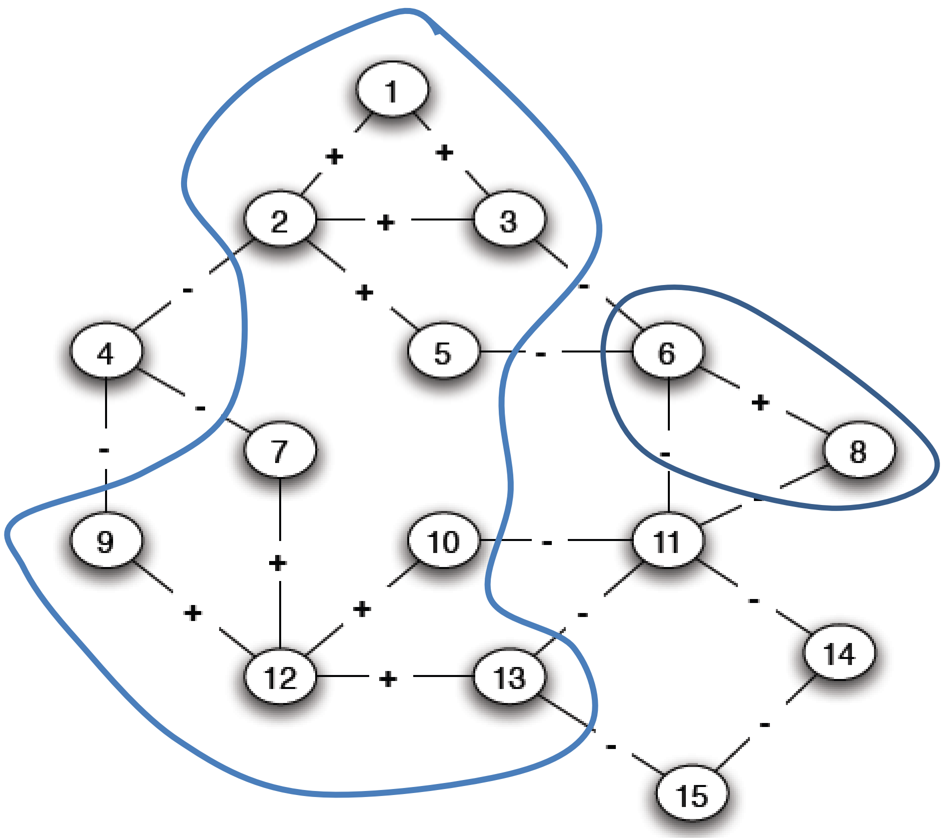
A (non-complete) graph is balanced if it possible to divide the nodes into two sets X and Y, such that any edge with both ends inside X or both ends inside Y is positive and any edge with one end in X and one end in Y is negative



The **two definition** are **equivalent**:
An arbitrary signed graph is balanced under the first definition, if and only if, it is balanced under the second definitions

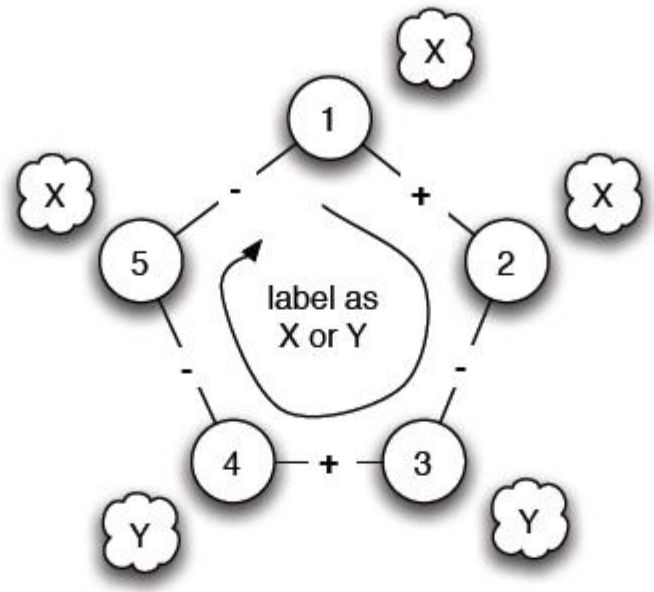
Balance Definition for General Graphs

Algorithm for dividing the nodes?



Balance Characterization

What prevents a network from being balanced?



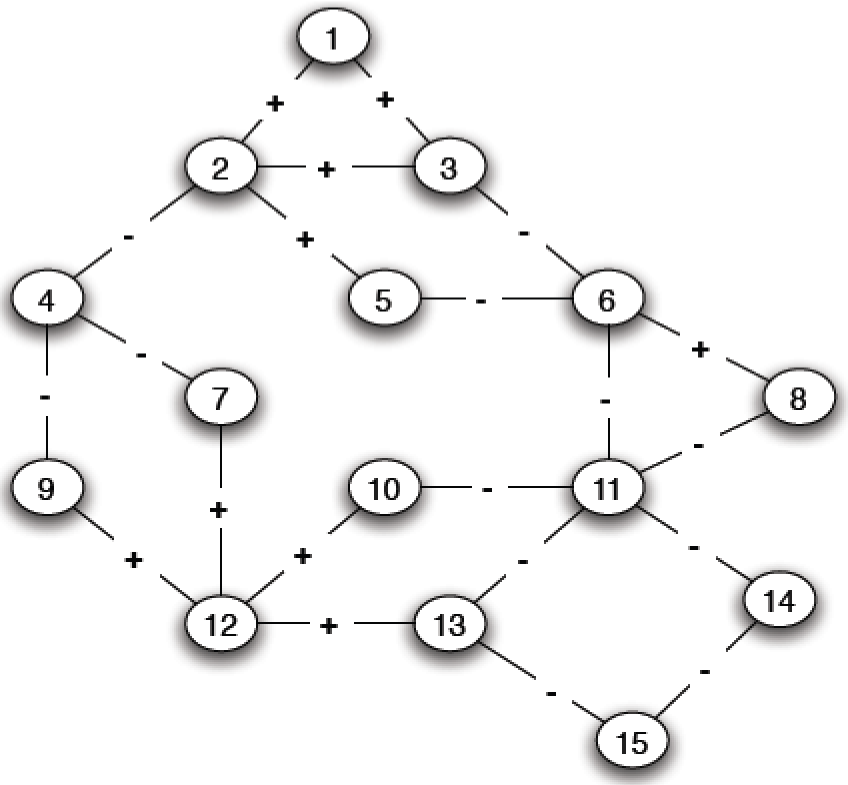
- Start from a node and place nodes in X or Y
- Every time we cross a negative edge, change the set

Cycle with odd number of negative edges

Balance Definition for General Graphs

Cycle with odd number of - => unbalanced

Is there such a cycle with an odd number of -?



Balance Characterization

Claim: A signed graph is balanced, if and only if, it contains no cycles with an odd number of negative edges

(proof by construction)

Find *a balanced division:* partition into sets X and Y, all edges inside X and Y positive, crossing edges negative

Either succeeds or Stops with a cycle containing an odd number of -

Two steps:

1. Convert the graph into a reduced one with only negative edges
2. Solve the problem in the reduced graph

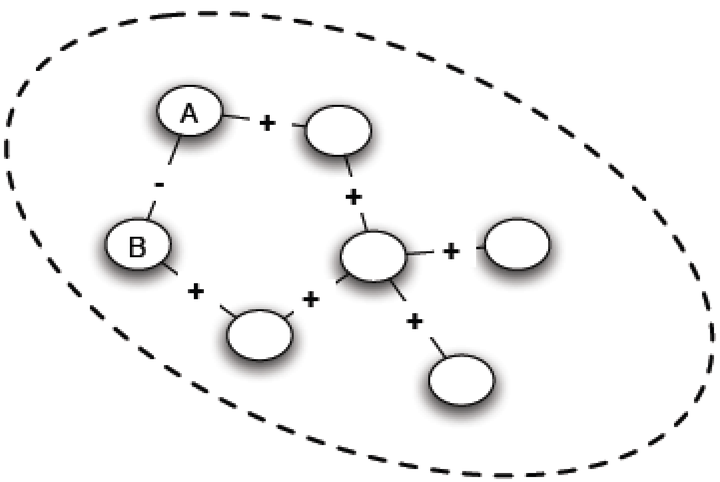
Balance Characterization: Step 1

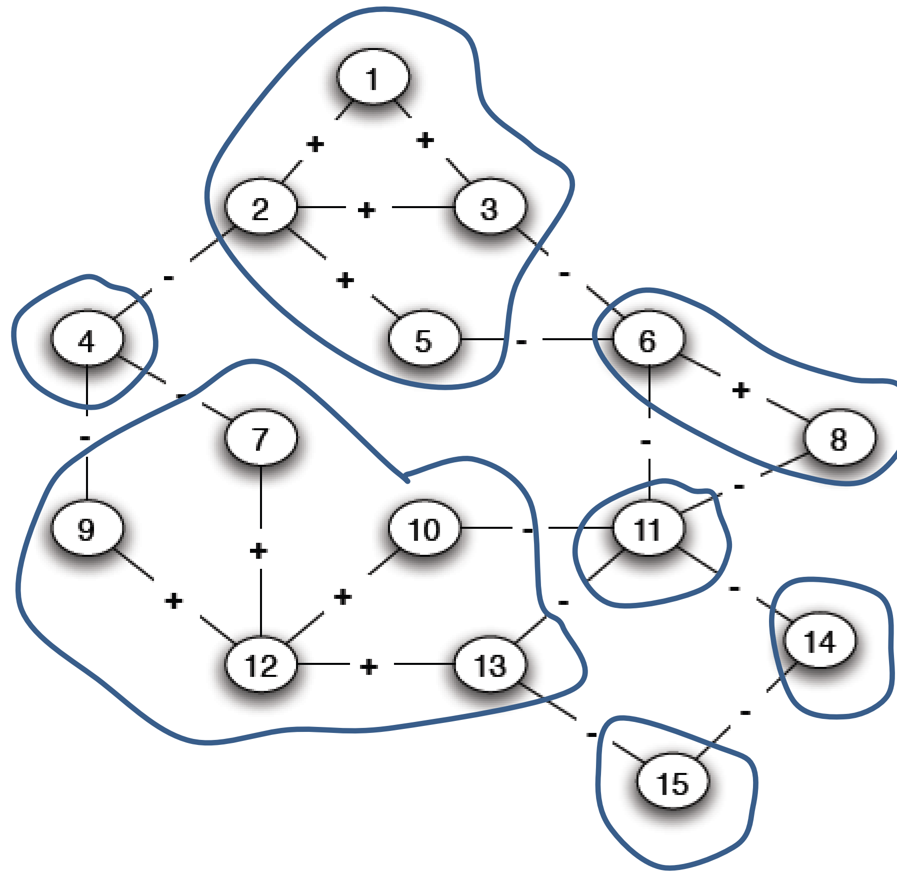
a. Find *connected components* (supernodes) by considering only positive edges

b. Check: Do supernodes contain a negative edge between any pair of their nodes

(a) Yes -> odd cycle (1)

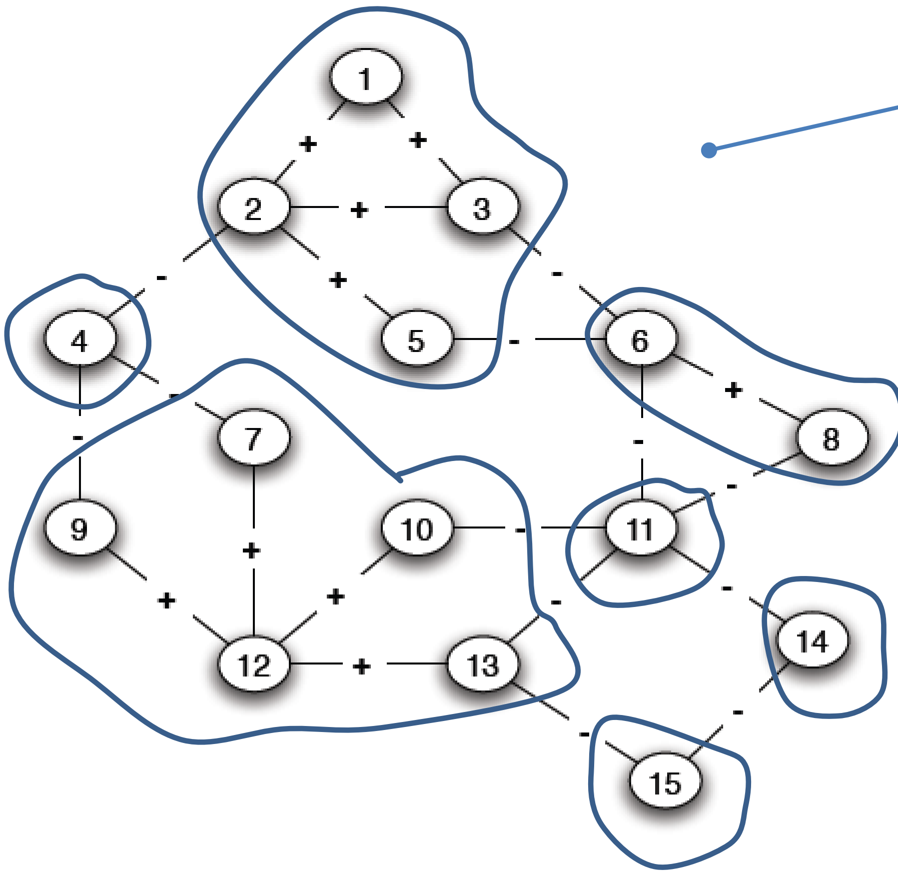
(b) No -> each supernode either X or Y

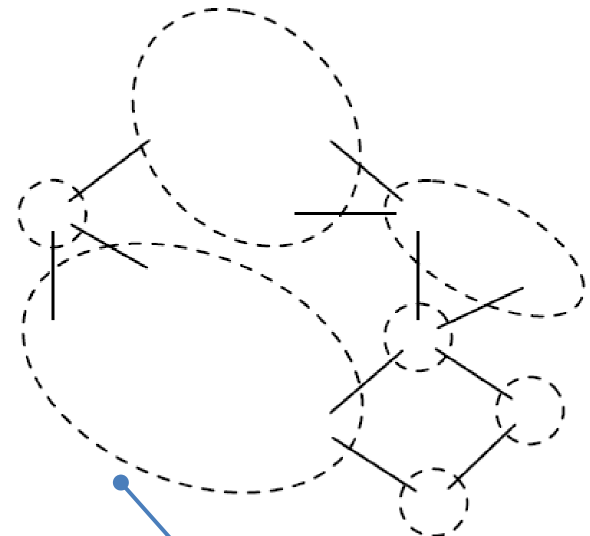


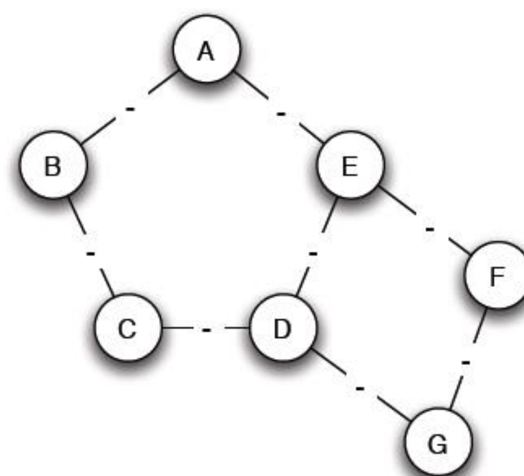


Balance Characterization: Step 1

3. Reduced problem: a node for each supernode, an edge between two supernodes if an edge in the original







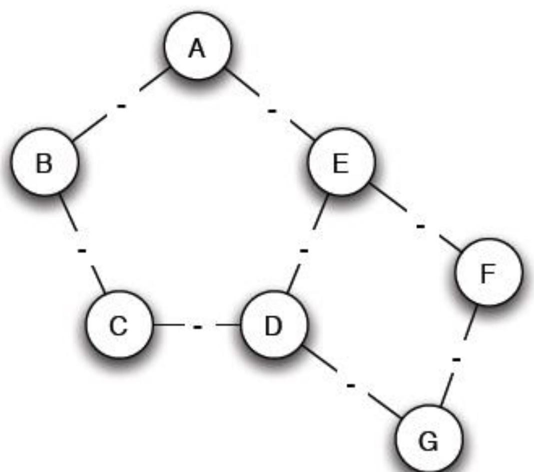
Balance Characterization: Step 2

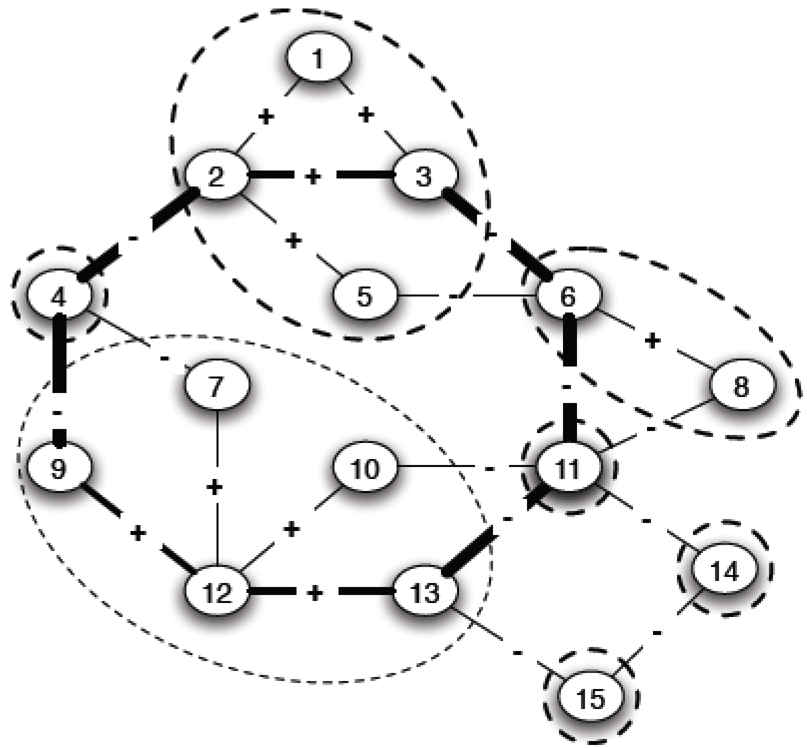
Note: Only negative edges among supernodes

Start labeling by either X and Y

If successful, then label the nodes of the supernode correspondingly

✓ A cycle with an odd number, corresponds to a (possibly larger) odd cycle in the original





Balance Characterization: Step 2

Determining whether the graph is bipartite (there is no edge between nodes in X or Y, *the only edges are from nodes in X to nodes in Y*)

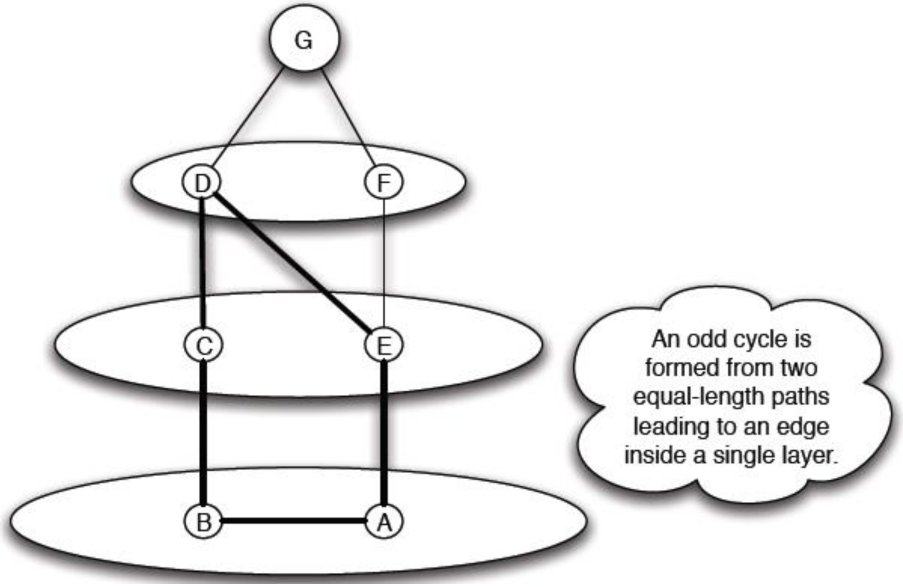
Use Breadth-First-Search (BFS)

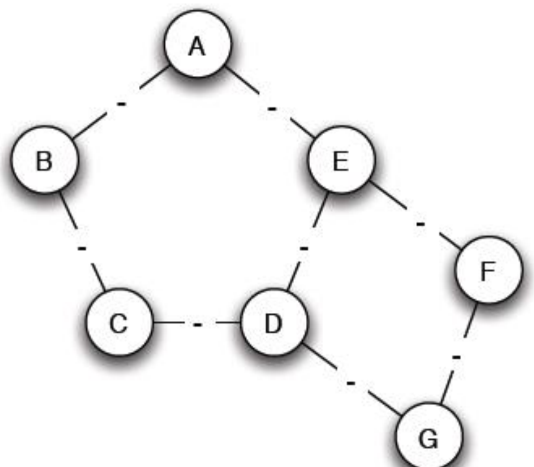
Two type of edges: (1) between nodes in adjacent levels (2) between nodes in the same level

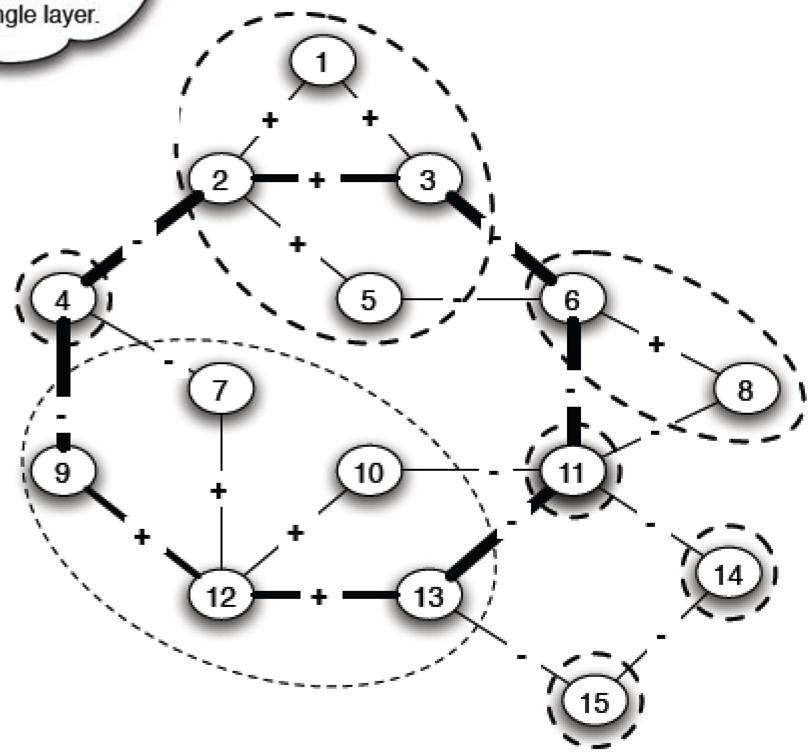
If only type (1), alternate X and Y labels at each level

If type (2), then odd cycle

Balance Characterization







Generalizing

1. Non-complete graphs

2. Instead of all triangles, “most” triangles, approximately divide the graph

Approximately Balance Networks

a complete graph (or clique): every edge either + or -

Claim: If all triangles in a labeled complete graph are balanced, than either
(a) all pairs of nodes are friends or,
(b) the nodes can be divided into two groups X and Y, such that
  (i) every pair of nodes in X like each other,
  (ii) every pair of nodes in Y like each other, and
  (iii) every one in X is the enemy of every one in Y.

Not all, but most, triangles are balanced

Claim: If *at least 99.9%* of all triangles in a labeled compete graph are balanced, then either,
(a) There is a set consisting of *at least 90%* of the nodes in which *at least 90%* of all pairs are friends, or,
(b) the nodes can be divided into two groups X and Y, such that
  (i) *at least 90%* of the pairs in X like each other,
  (ii) *at least 90%* of the pairs in Y like each other, and
  (iii) *at least 90%* of the pairs with one end in X and one in Y are enemies

Approximately Balance Networks

Claim: If *at least 99.9%* of all triangles in a labeled complete graph are balanced, then either,

(a) There is a set consisting of *at least 90%* of the nodes in which *at least 90%* of all pairs are friends, or,

(b) the nodes can be divided into two groups X and Y, such that

  (i) *at least 90%* of the pairs in X like each other,

  (ii) *at least 90%* of the pairs in Y like each other, and

  (iii) *at least 90%* of the pairs with one end in X and one in Y are enemies

Claim: Let ε be any number, such that $0 \leq \varepsilon < 1/8$. 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* $1\text{-}\delta$ of the nodes in which *at least* $1\text{-}\delta$ of all pairs are friends, or,

(b) the nodes can be divided into two groups X and Y, such that

  (i) *at least* $1\text{-}\delta$ of the pairs in X like each other,

  (ii) *at least* $1\text{-}\delta$ of the pairs in Y like each other, and

  (iii) *at least* $1\text{-}\delta$ of the pairs with one end in X and one in Y are enemies

$$\delta = \sqrt[3]{\varepsilon}$$

Approximately Balance Networks

Basic idea – find a "good" node A (s.t., it does not belong to too many unbalanced triangles) to partition into X and Y

Pigeonhole principle: if *n* items are put into *m* *pigeonholes* with *n* > *m*, then at least one pigeonhole must contain more than one item

Counting argument based on pigeonhole: compute the average value of a set of objects and then argue that there must be at least one node that is equal to the average or below (or equal and above)

References

Networks, Crowds, and Markets (Chapter 3, 4, 5)