Online Social Networks and Media

Team Formation in Social Networks

Network Ties

Thanks to Evimaria Terzi

ALGORITHMS FOR TEAM FORMATION

Team-formation problems

- Given a task and a set of experts (organized in a network) find the subset of experts that can effectively perform the task
- Task: set of required skills and potentially a budget
- Expert: has a set of skills and potentially a price
- Network: represents strength of relationships

Security expert

Electronics expert

Insider

Mechanic

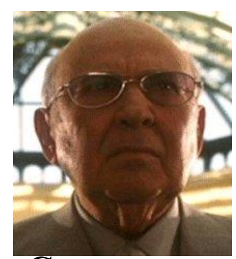
Organizer

Pick-pocket thief

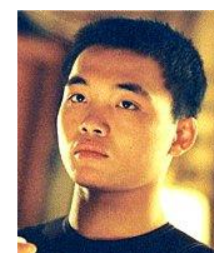
Co-organizer

Mechanic

Explosives expert

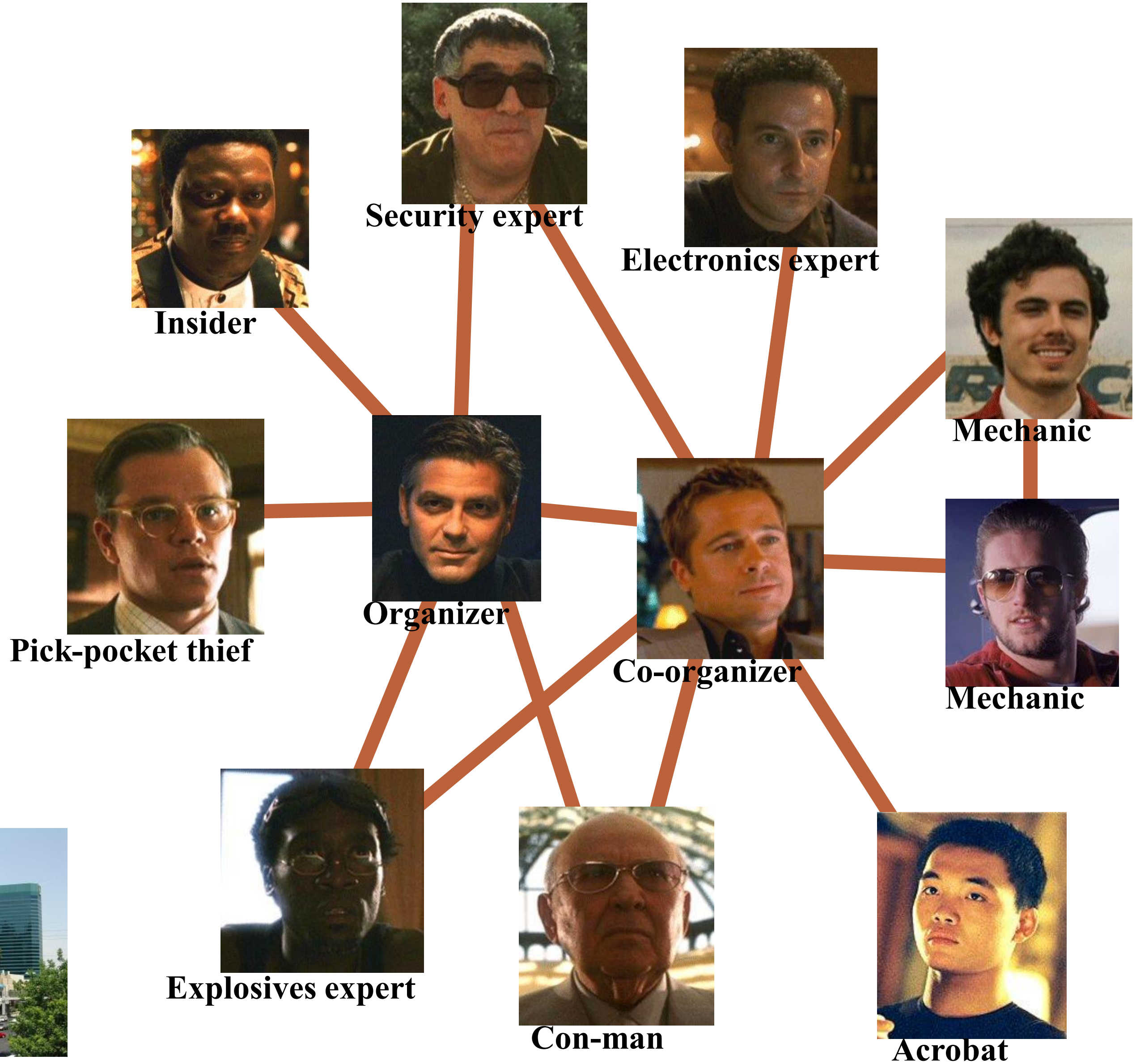


Con-man



Acrobat


george clooney
matt damon
andy garcia
brad pitt
julia roberts
OCEAN'S ELEVEN
Place your bets.


Insider
Security expert
Electronics expert
Mechanic
Pick-pocket thief
Organizer
Co-organizer
Mechanic
Explosives expert
Con-man
Acrobat

Applications

- Collaboration networks (e.g., scientists, actors)
- Organizational structure of companies
- LinkedIn, UpWork, FreeLance
- Geographical (map) of experts

Simple Team formation Problem

- Input:
  - A task T, consisting of a set of skills
  - A set of candidate experts each having a subset of skills

T = {algorithms, java, graphics, python}

Alice	Bob	Cynthia	David	Eleanor
{algorithms}	{python}	{graphics, java}	{graphics}	{graphics,java,python}

- Problem: Given a task and a set of experts, find the smallest subset (team) of experts that together have all the required skills for the task

Set Cover

- The Set Cover problem:
  - We have a universe of elements $U = \{x_1, \dots, x_N\}$
  - We have a collection of subsets of U, $\boldsymbol{S} = \{S_1, \dots, S_n\}$, such that $\bigcup_i S_i = U$
  - We want to find the smallest sub-collection $\boldsymbol{C} \subseteq \boldsymbol{S}$ of $\boldsymbol{S}$, such that $\bigcup_{S_i \in \boldsymbol{C}} S_i = U$
    - The sets in $\boldsymbol{C}$ cover the elements of U

Coverage

- The Simple Team Formation Problem is a just an instance of the Set Cover problem
  - Universe U of elements = Set of all skills
  - Collection $\boldsymbol{S}$ of subsets = The set of experts and the subset of skills they possess.

T = {algorithms, java, graphics, python}

Alice	Bob	Cynthia	David	Eleanor
{algorithms}	{python}	{graphics, java}	{graphics}	{graphics,java,python}

Complexity

- The Set Cover problem are NP-complete
  - What does this mean?
  - Why do we care?

- There is no algorithm that can guarantee finding the best solution in polynomial time
  - Can we find an algorithm that can guarantee to find a solution that is close to the optimal?
  - Approximation Algorithms.

Approximation Algorithms

- For a (combinatorial) minimization problem, where:
  - X is an instance of the problem,
  - OPT(X) is the value of the optimal solution for X,
  - ALG(X) is the value of the solution of an algorithm ALG for X

  ALG is a good approximation algorithm if the ratio of ALG(X)/OPT(X) and is bounded for all input instances X
    - We want the ratio to be close to 1

- Minimum set cover: input X = (U,**S**) is the universe of elements and the set collection, OPT(X) is the size of minimum set cover, ALG(X) is the size of the set cover found by an algorithm ALG.

Approximation Algorithms

- For a minimization problem, the algorithm ALG is an α-approximation algorithm, for $\alpha > 1$, if for all input instances X,

$$ALG(X) \leq \alpha OPT(X)$$

- In simple words: the algorithm ALG is at most α times worse than the optimal.

- α is the approximation ratio of the algorithm – we want α to be as close to 1 as possible
  - Best case: $\alpha = 1 + \epsilon$ and $\epsilon \to 0$, as $n \to \infty$ (e.g., $\epsilon = \frac{1}{n}$)
  - Good case: $\alpha = O(1)$ is a constant (e.g., $\alpha = 2$)
  - OK case: $\alpha = \mathrm{O}(\log n)$
  - Bad case $\alpha = \mathrm{O}(n^{\epsilon})$

A simple approximation ratio for set cover

- Any algorithm for set cover has approximation ratio $\alpha = |S_{max}|$, where S_{max} is the set in $\boldsymbol{S}$ with the largest cardinality

- Proof:
  - $OPT(X) \geq N/|S_{max}| \Rightarrow N \leq |S_{max}|OPT(X)$
  - $ALG(X) \leq N \leq |S_{max}|OPT(X)$

- This is true for any algorithm.
- Not a good bound since it may be that $|S_{max}| = O(N)$

An algorithm for Set Cover

- What is the most natural algorithm for Set Cover?

- Greedy: each time add to the collection $\boldsymbol{C}$ the set S_i from $\boldsymbol{S}$ that covers the most of the remaining uncovered elements.

The GREEDY algorithm

GREEDY(U,S)

X= U

C = {}

while X is not empty do

For all $S_i \in \boldsymbol{S}$ let $\text{gain}(S_i) = |S_i \cap X|$

Let S_* be such that $gain(S_*)$ is maximum

C = C U {S_*}

X = X\ S_*

S = ***S***\ S_*

The number of elements covered by S_i not already covered by C.

Greedy is not always optimal

Alice
C, C++, Unix

Eleanor
Python, Joomla

Required Skills
C, C++, Unix, php, Java, Python, Joomla

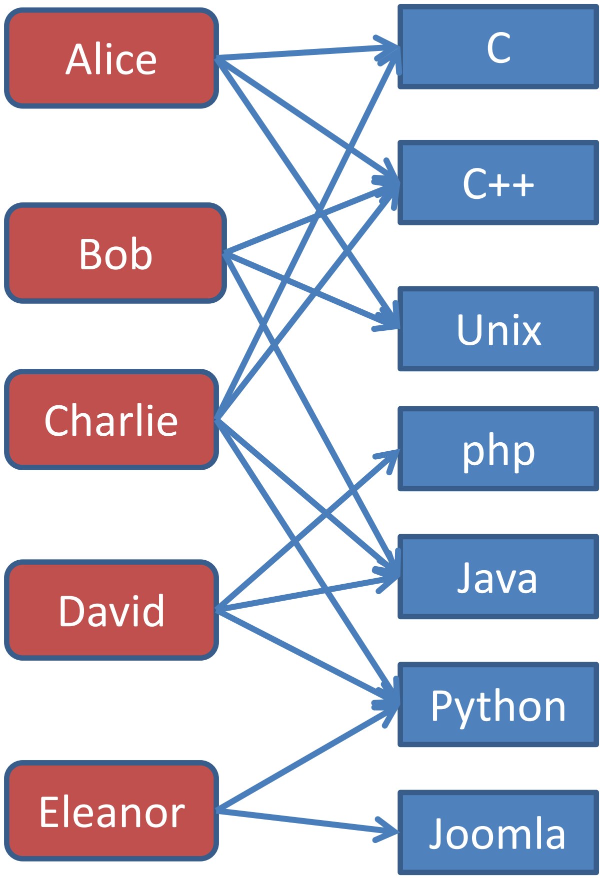
Bob
C++, Unix, Java

David
php, Java, Python

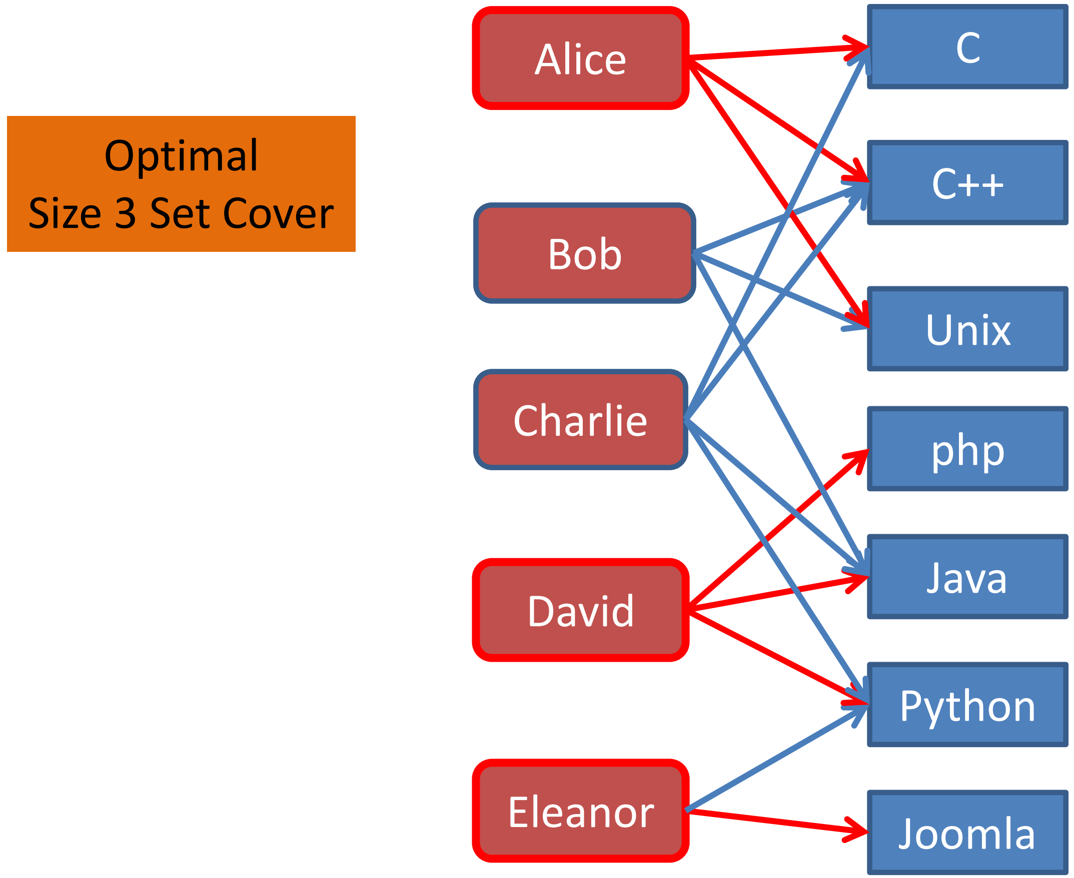
Charlie
C, C++, Java, Python

Greedy is not always optimal

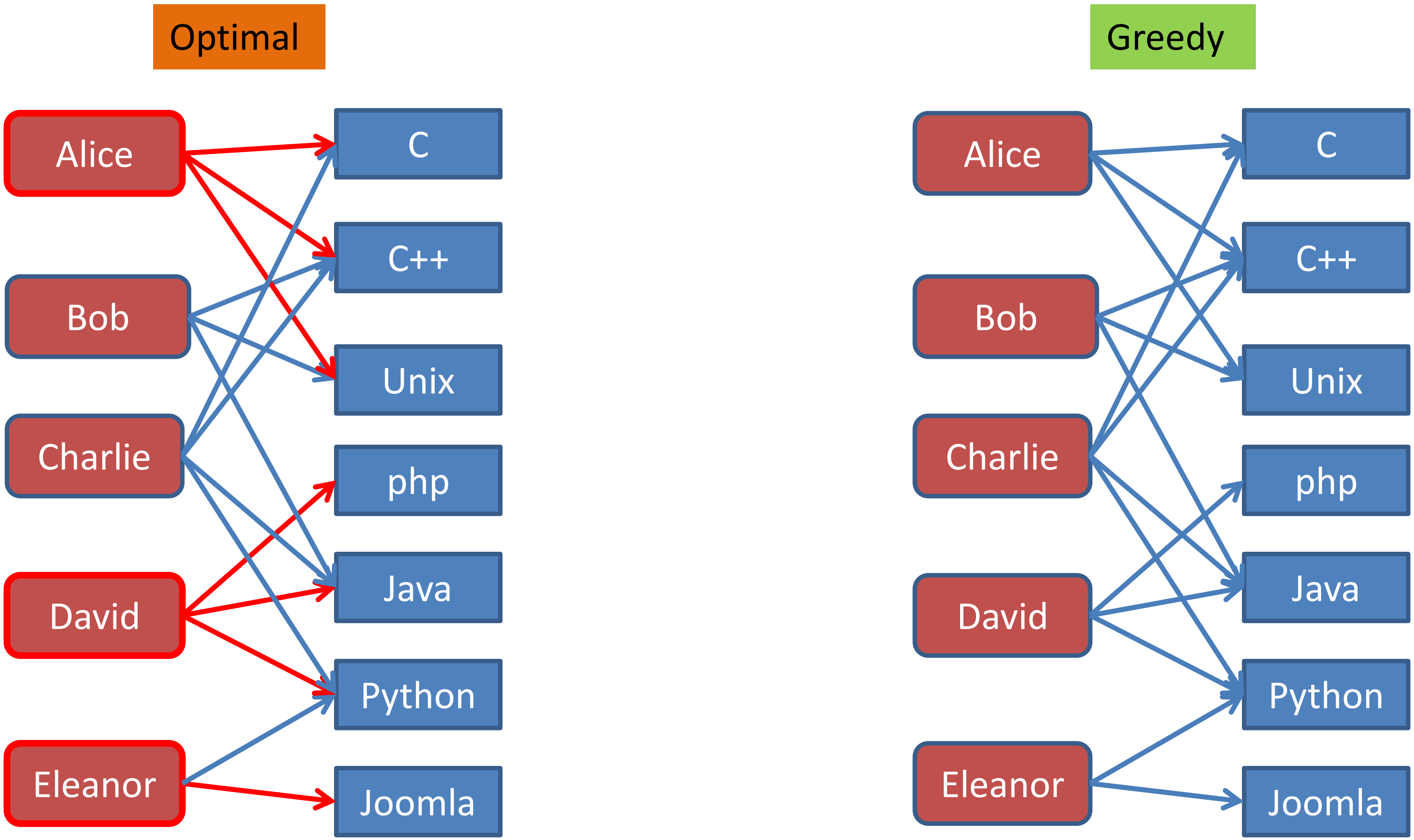
A different representation



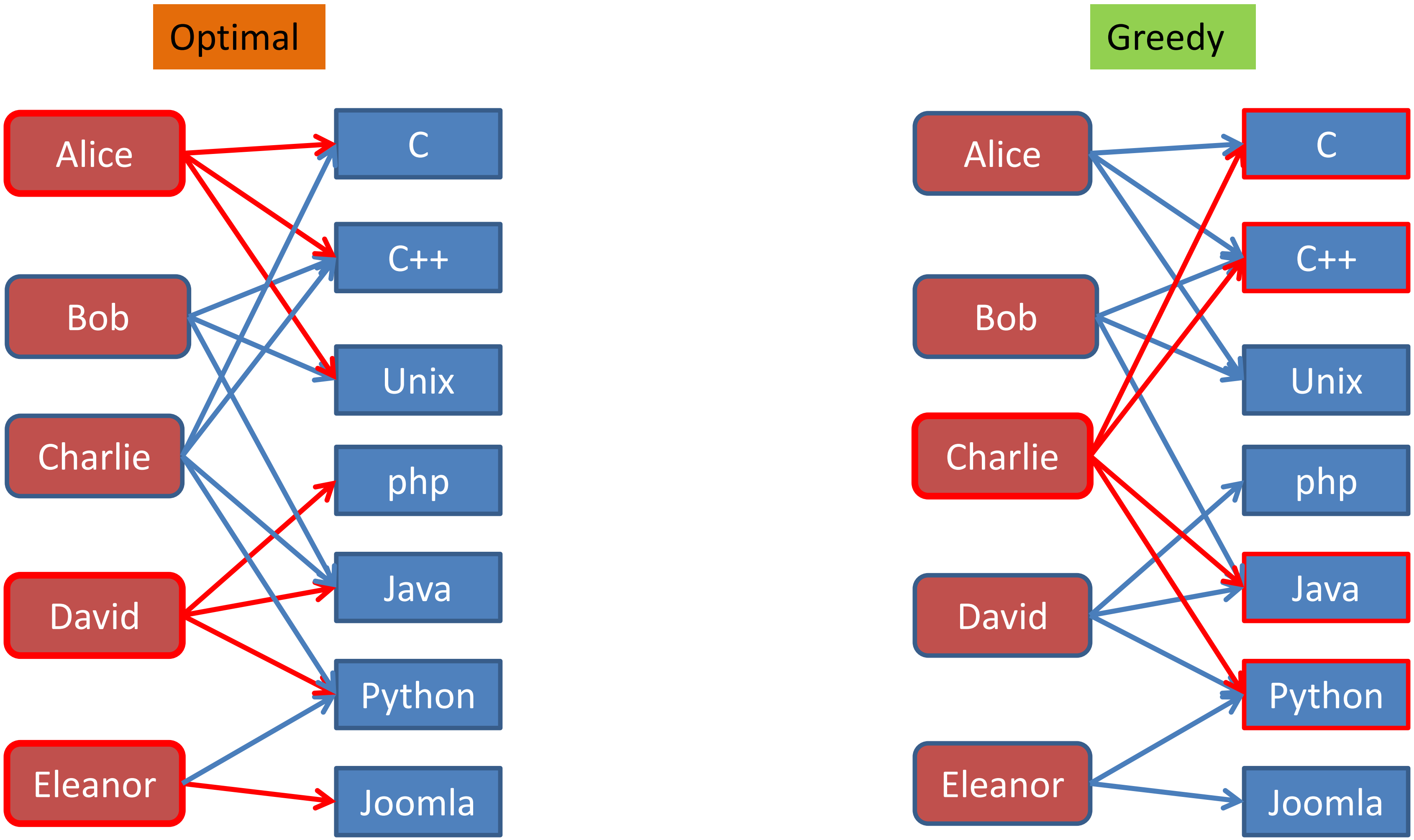
Greedy is not always optimal



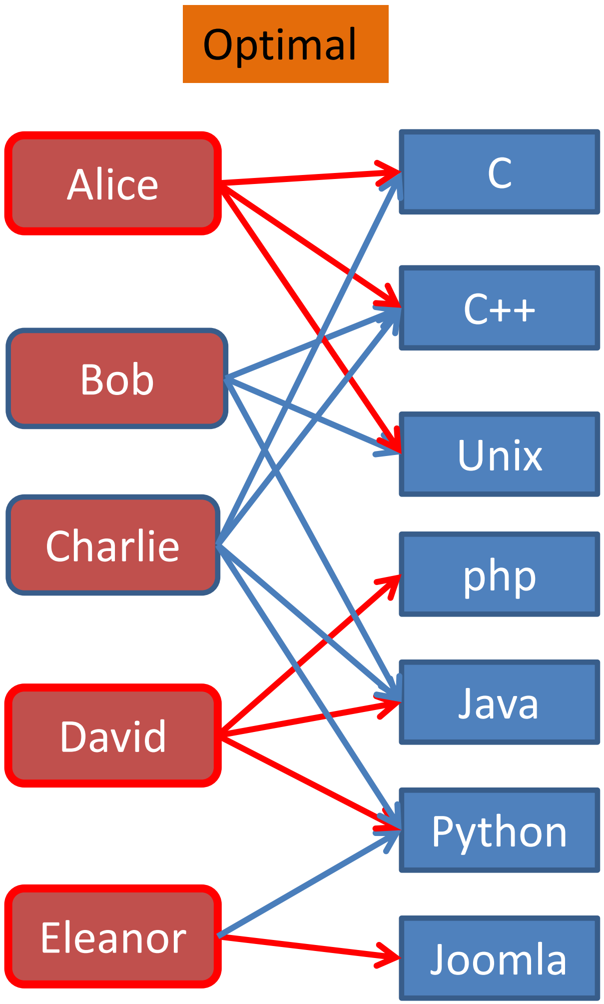
Greedy is not always optimal

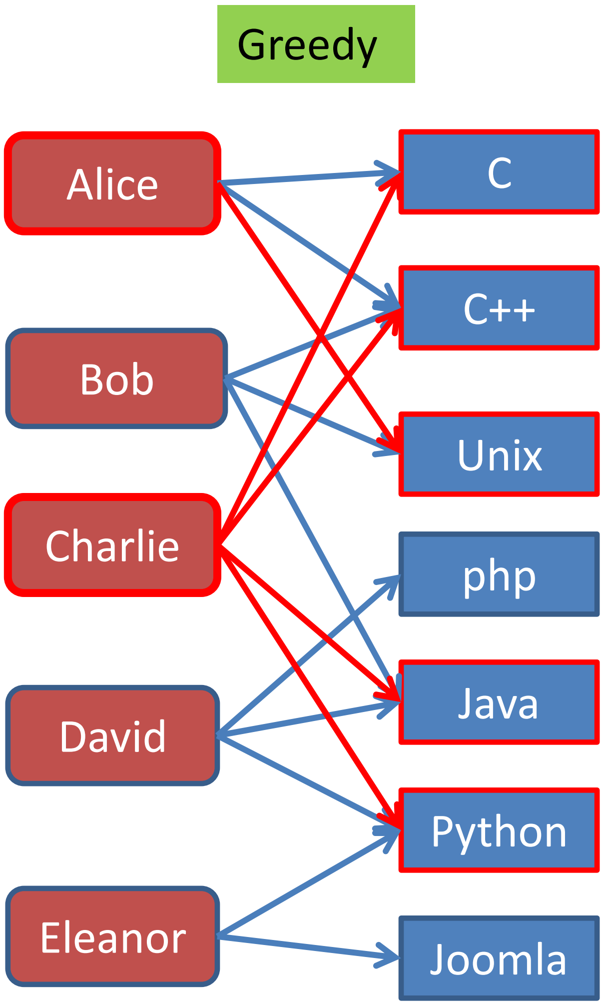


Greedy is not always optimal

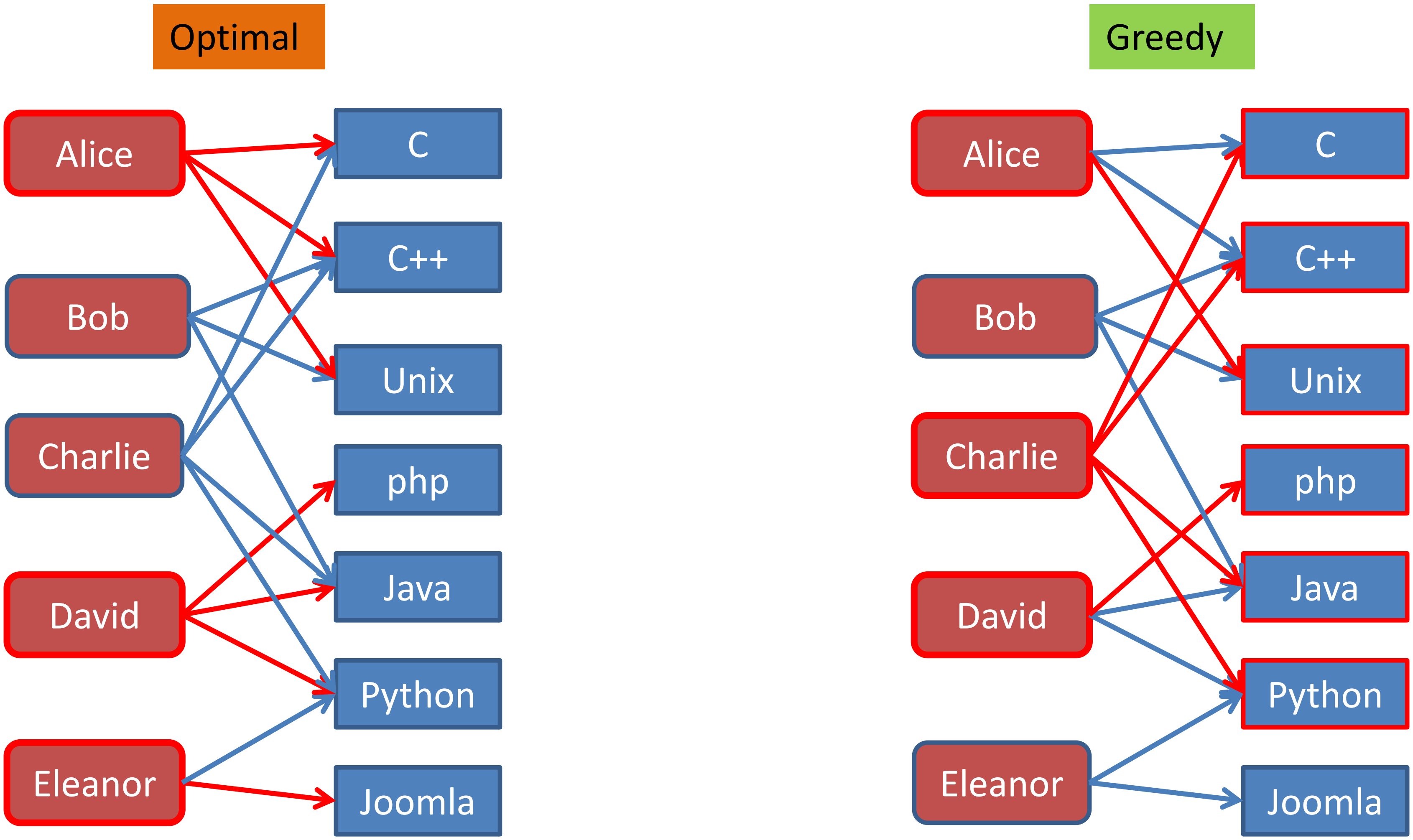


Greedy is not always optimal

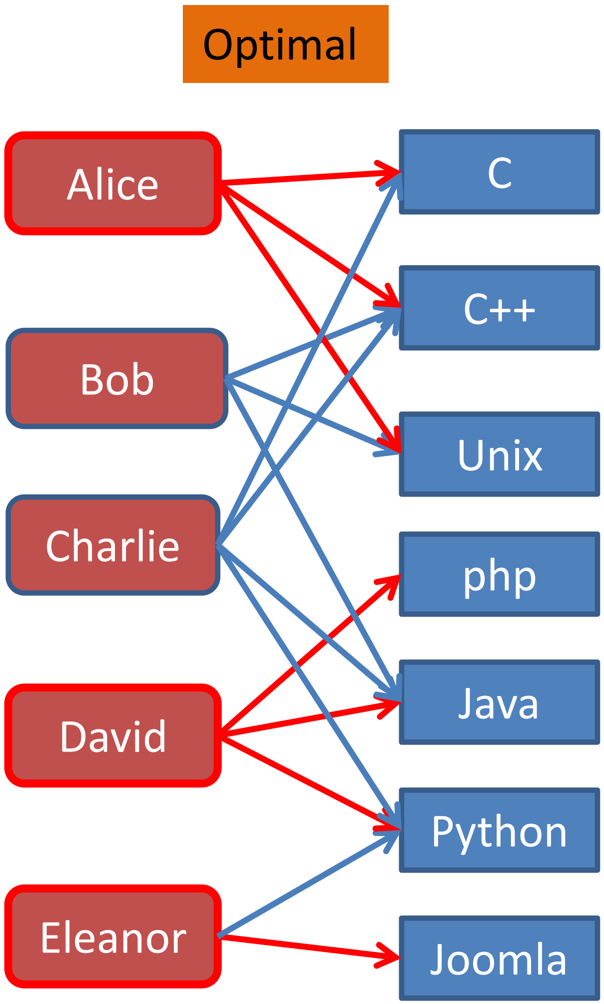


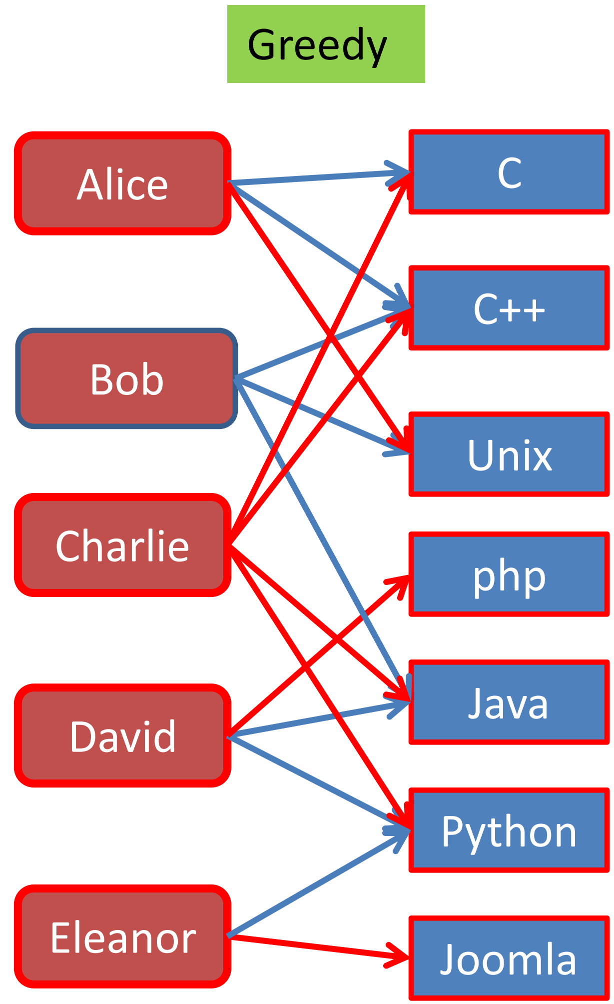


Greedy is not always optimal

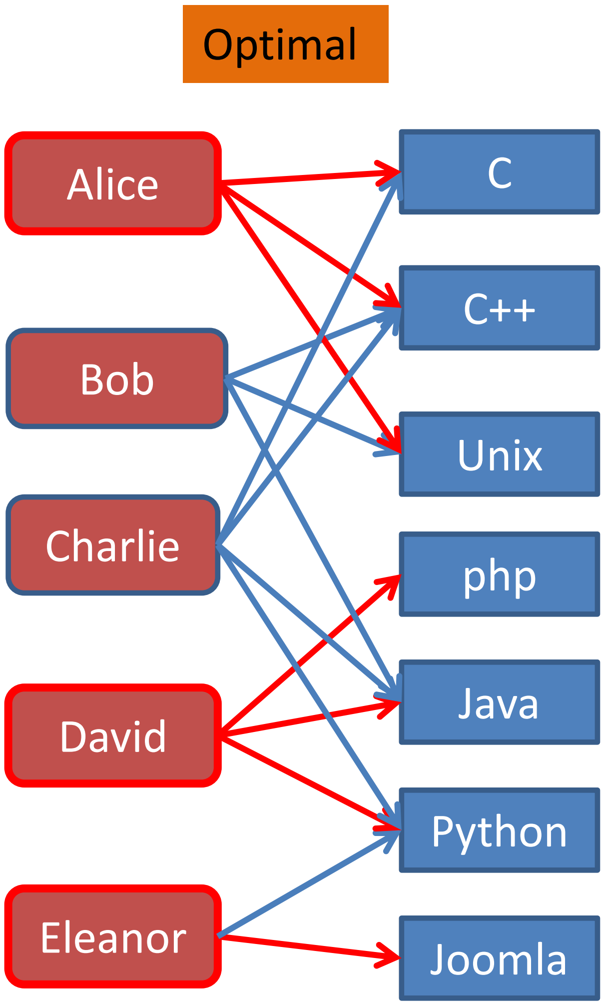


Greedy is not always optimal

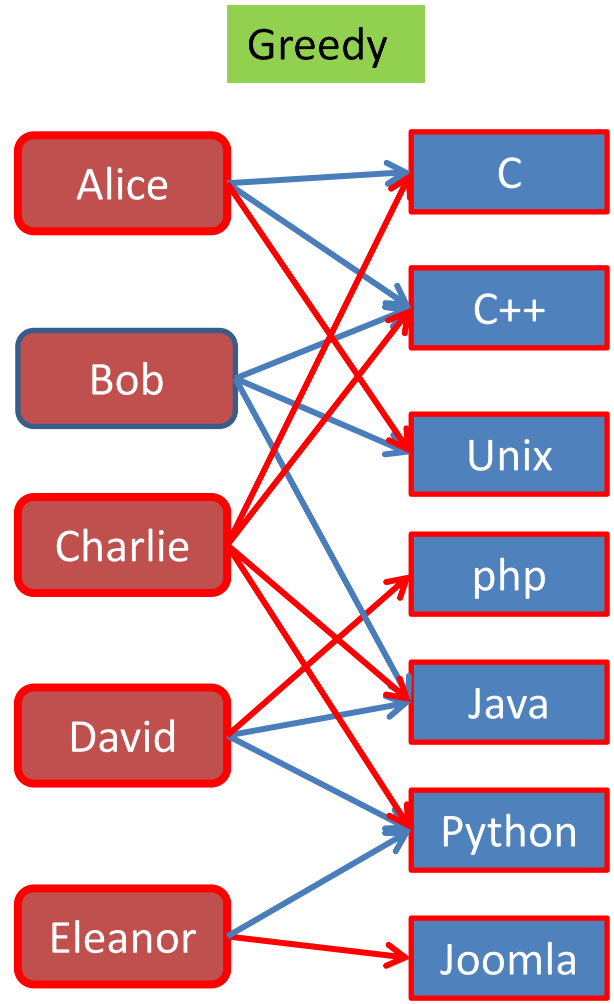




Greedy is not always optimal



- Selecting Charlie is useless since we still need Alice and David
- Alice and David cover a superset of the skills covered by Charlie



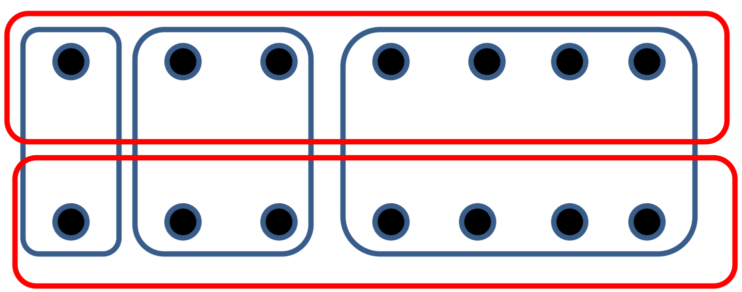
Approximation ratio of GREEDY

- Good news: GREEDY has approximation ratio:

$$\alpha = H(|S_{\max}|) = 1 + \ln|S_{\max}|, \qquad H(n) = \sum_{k=1}^{n} \frac{1}{k}$$

$$GREEDY(X) \le (1 + \ln|S_{\max}|)OPT(X),$$ for all X

- The approximation ratio is tight up to a constant
  - Tight means that we can find a counter example with this ratio



OPT(X) = 2
GREEDY(X) = logN
α=½logN

Team formation in the presence of a social network

- Given a task and a set of experts organized in a network find the subset of experts that can effectively perform the task

- Task: set of required skills

- Expert: has a set of skills

- Network: represents strength of relationships

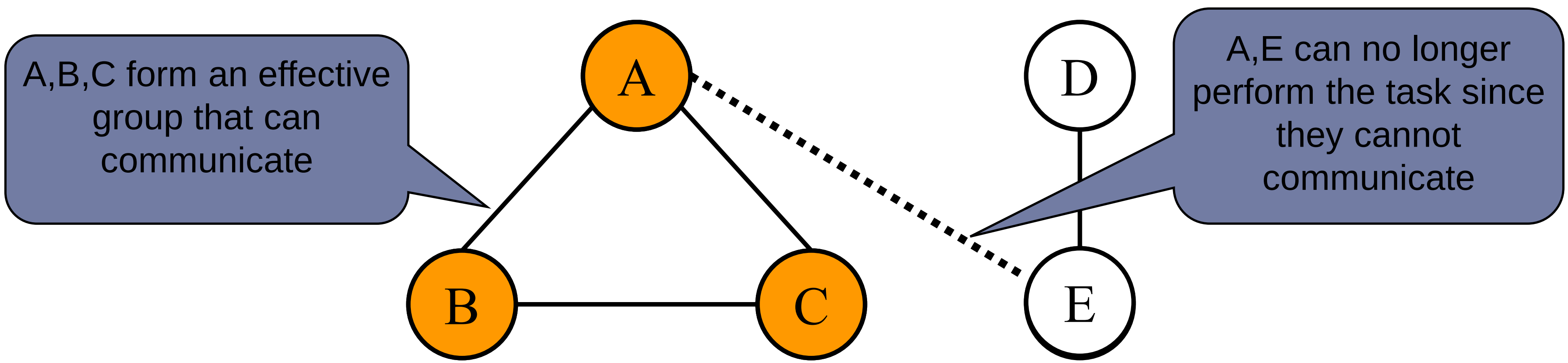
- Effectively: There is good communication between the team members
  - What does good mean? E.g., all team members are connected.

Coverage is NOT enough

T={**algorithms**,**java**,**graphics**,**python**}

Alice	**B**ob	**C**ynthia	**D**avid	**E**leanor
{algorithms}	{python}	{graphics, java}	{graphics}	{graphics,java,python}

Alice and Eleanor are the smallest team that covers all skills

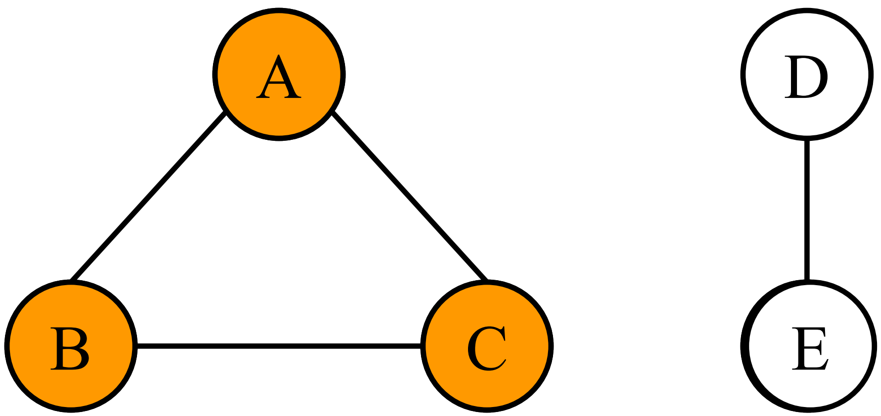


Communication: the members of the team must be able to efficiently communicate and work together

How to measure effective communication?

The longest shortest path between any two nodes in the subgraph

- Diameter of the subgraph defined by the group members

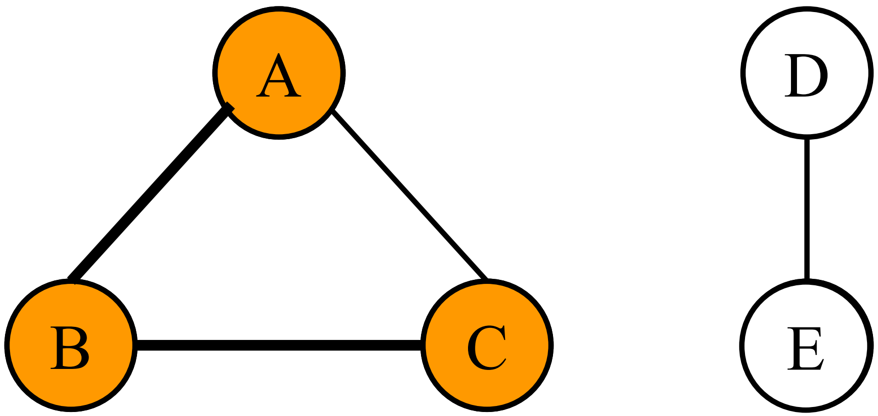


diameter = 1

How to measure effective communication?

The total weight of the edges of a tree that spans all the team nodes

- MST (Minimum spanning tree) of the subgraph defined by the group members



MST = 2

Problem definition (MinDiameter)

- Given a task and a social network G of experts, find the subset (team) of experts that can perform the given task and they define a subgraph G' in G with the minimum diameter.

- Problem is NP-hard
- Equivalent to the Multiple Choice Cover (MCC)
  - We have a set cover instance $(U, \mathbf{S})$, but we also have a distance matrix D with distances between the different sets in $\mathbf{S}$.
  - We want a cover that has the minimum diameter (minimizes the largest pairwise distance in the cover)

The RarestFirst algorithm

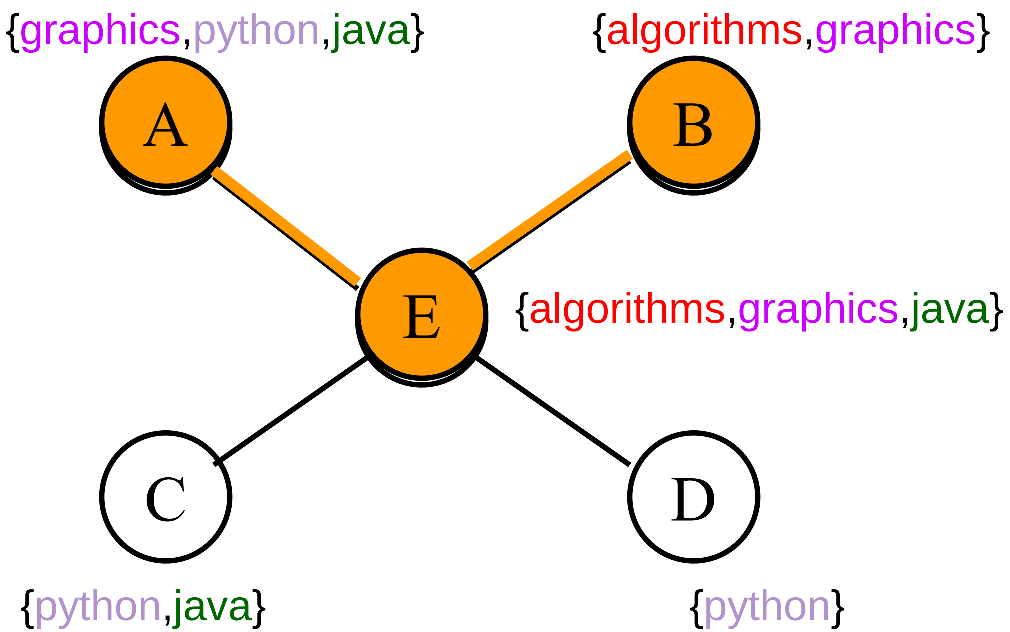
- Compute all shortest path distances in the input graph G and create a new complete graph G_C
- Find Rarest skill α_{rare} required for a task
- S_{rare} = group of people that have α_{rare}
- Evaluate star graphs in G_C, centered at individuals from S_{rare}
- Report cheapest star

Running time: Quadratic to the number of nodes

Approximation factor: 2xOPT

The RarestFirst algorithm

T={algorithms,java,graphics,python}



Skills:

algorithms

graphics

java

python

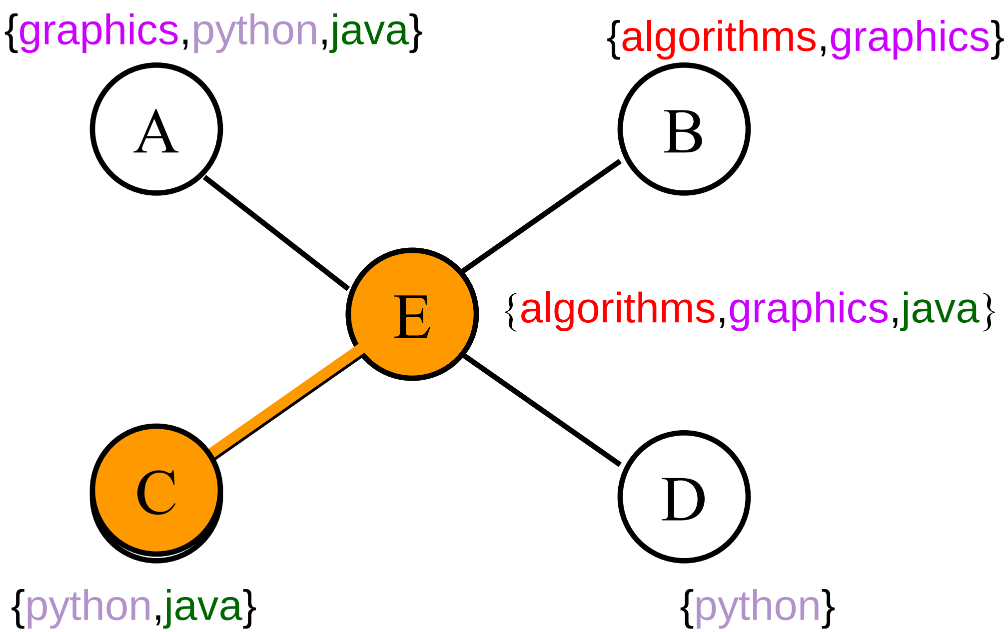
α_{rare} = algorithms

S_{rare} ={Bob, Eleanor}

Diameter = 2

The RarestFirst algorithm

T={algorithms,java,graphics,python}



Skills:

algorithms

graphics

java

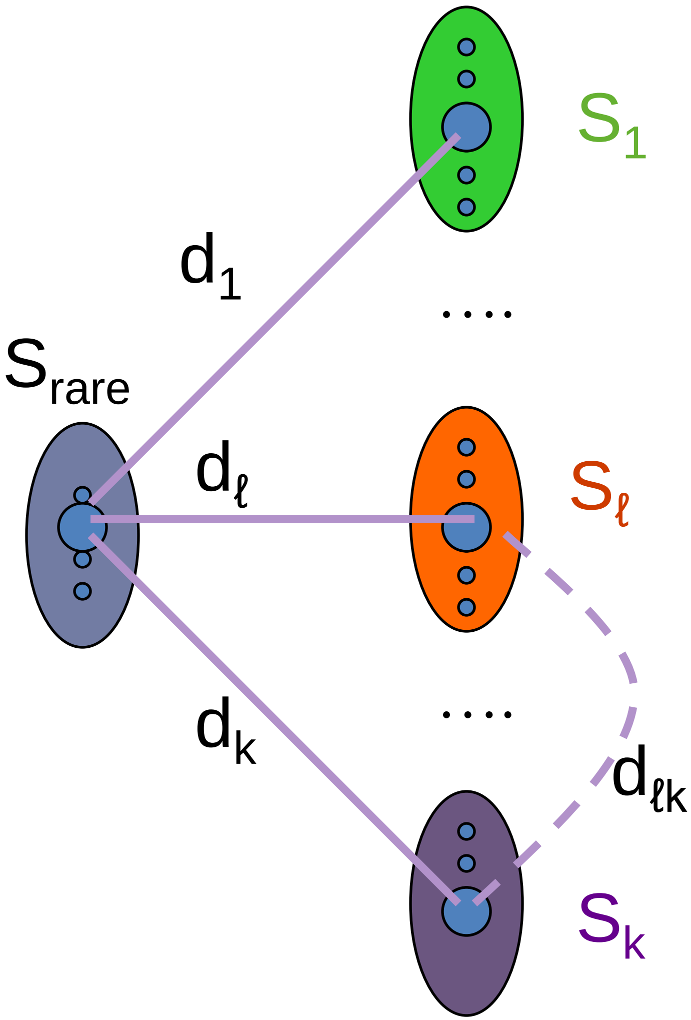
python

α_{rare} = algorithms

S_{rare} ={Bob, Eleanor}

Diameter = 1

Analysis of RarestFirst



- The diameter is
  - either $D = d_k$, for some node k,
  - or $D = d_{\ell k}$ for some pair of nodes ℓ, k
- Fact: $OPT \geq d_k$
- Fact: $OPT \geq d_\ell$
- $D \leq d_{\ell k} \leq d_\ell + d_k \leq 2*OPT$

Problem definition (MinMST)

- Given a task and a social network G of experts, find the subset (team) of experts that can perform the given task and they define a subgraph G' in G with the minimum MST cost.

- Problem is NP-hard
- Follows from a connection with Group Steiner Tree problem

The SteinerTree problem

- Graph G(V,E)

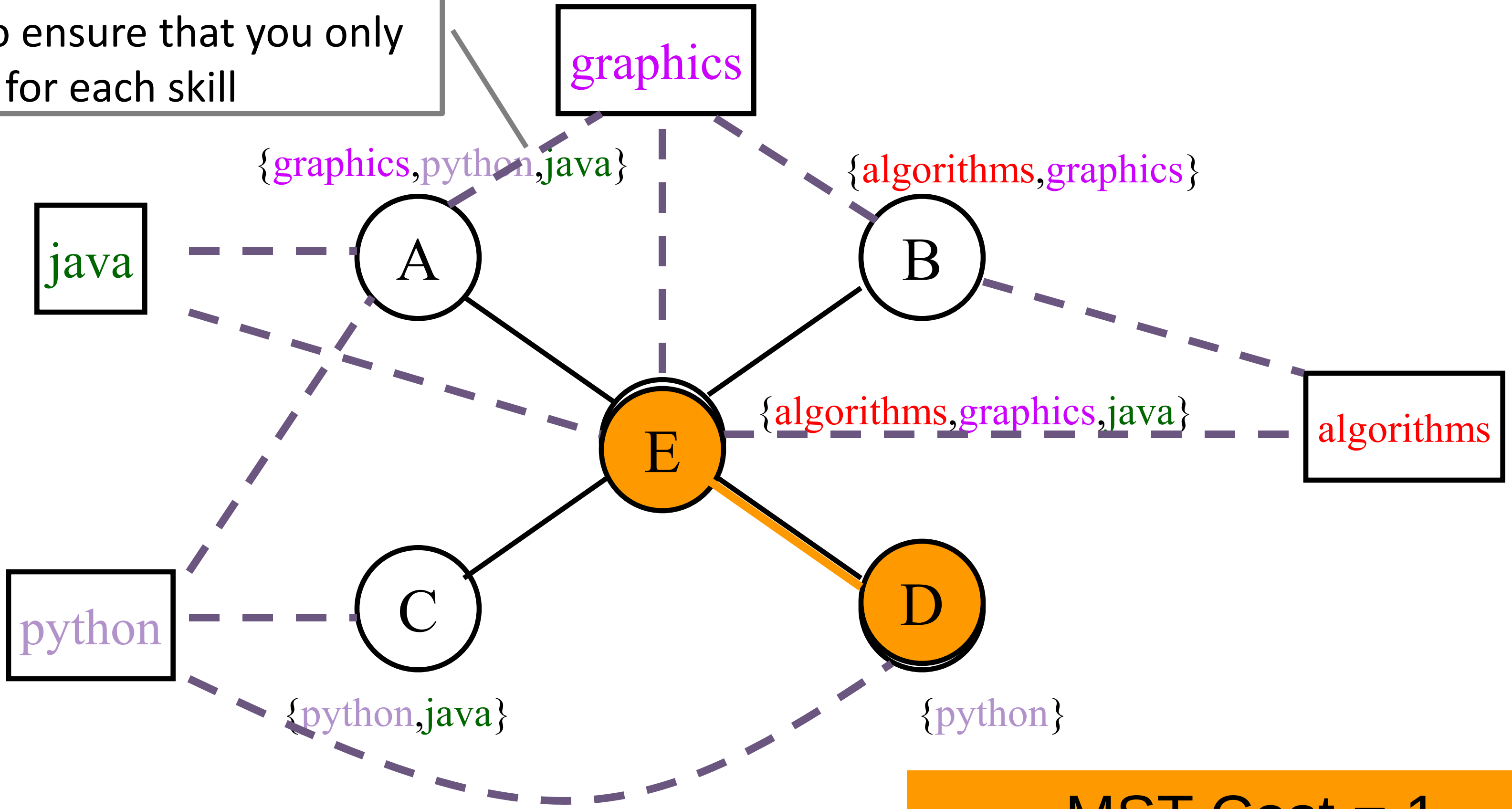
Required vertices

- Partition of V into V = {R,N}
- Find G' subgraph of G such that G' contains all the required vertices (R) and MST(G') is minimized
  - Find the cheapest tree that contains all the required nodes.

The EnhancedSteiner algorithm

Put a large weight on the new edges (more than the sum of all edges) to ensure that you only pick one for each skill

T={**algorithms**,**java**,**graphics**,**python**}

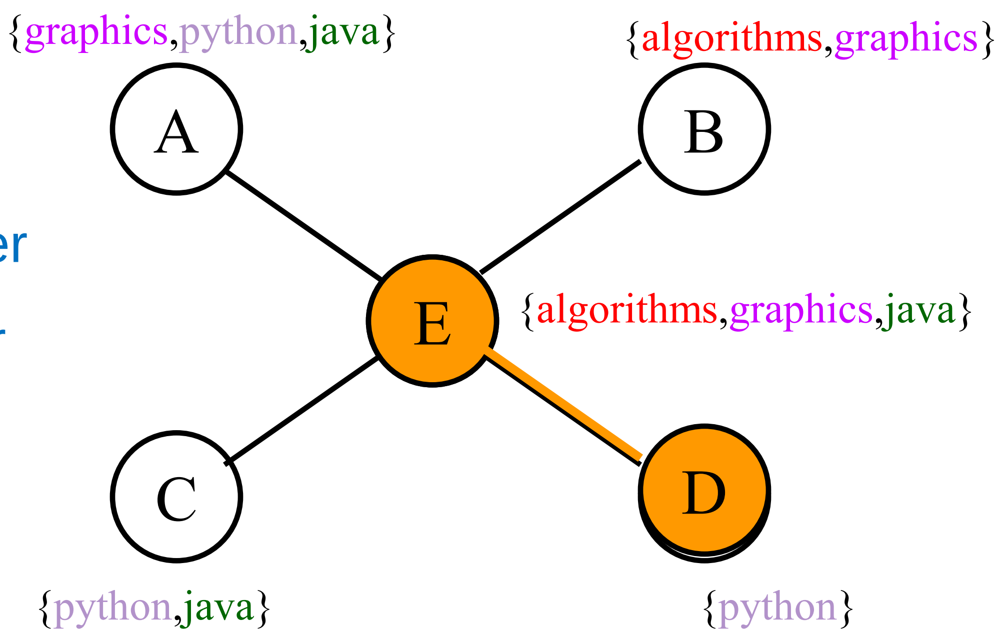


MST Cost = 1

The CoverSteiner algorithm

T={**algorithms**,**java**,**graphics**,**python**}

1. Solve SetCover
2. Solve Steiner

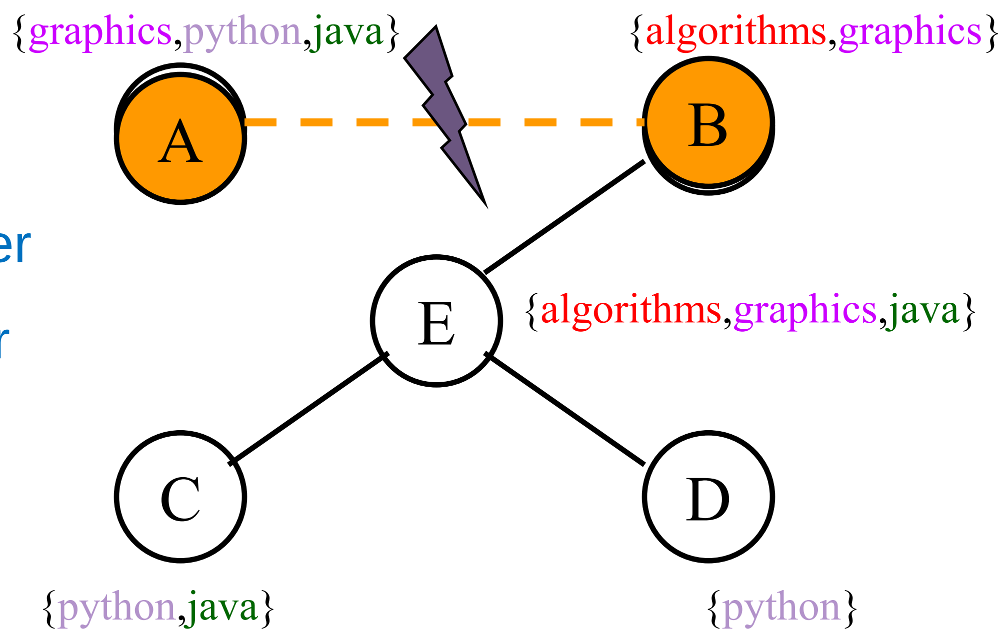


MST Cost = 1

How good is CoverSteiner?

T={**algorithms**,**java**,**graphics**,**python**}

1. Solve SetCover
2. Solve Steiner



MST Cost = Infty

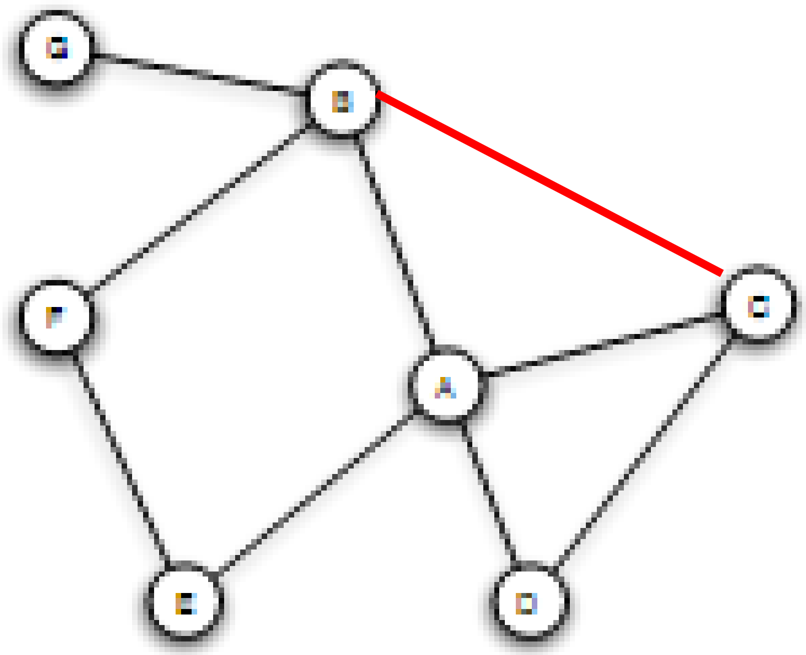
References

Theodoros Lappas, Kun Liu, Evimaria Terzi, Finding a team of experts in social networks. KDD 2009: 467-476

STRONG AND WEAK TIES

Triadic Closure

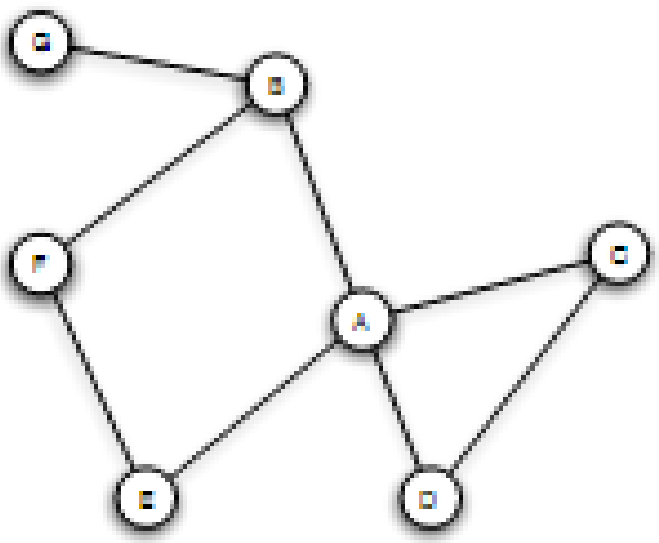
If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future

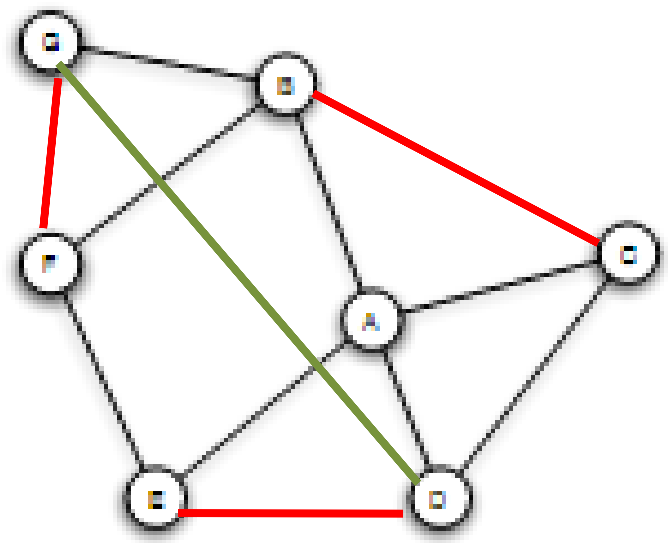


Triangle

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)*

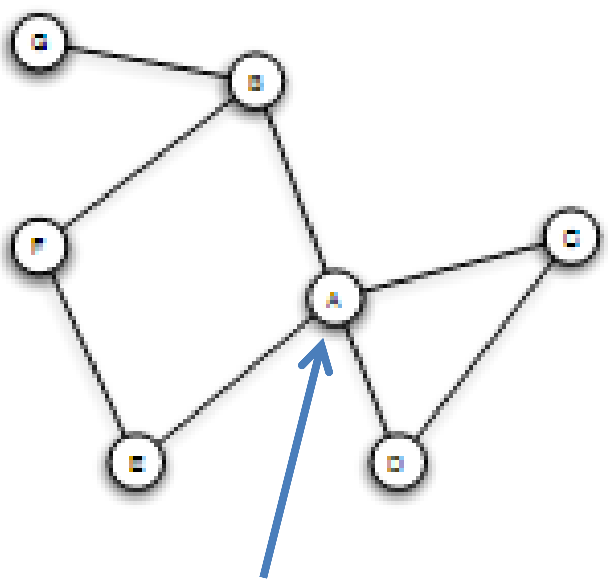
$$C_i = \frac{2 |\{e_{jk}\}|}{k_i(k_i - 1)}$$

$e_{jk} \in E, u_i, u_j \in N_i, k$ size of $\mathrm{N_i}$, $\mathrm{N_i}$ neighborhood of u_i

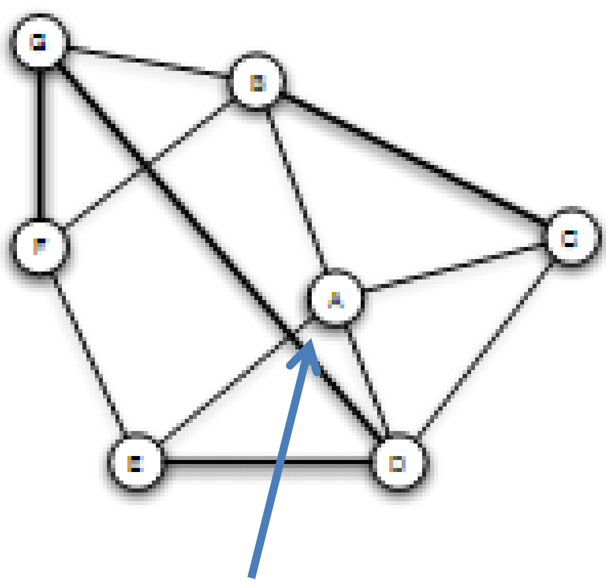
Fraction of the friends of a node that are friends with each other (i.e., connected)

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

Clustering Coefficient



1/6

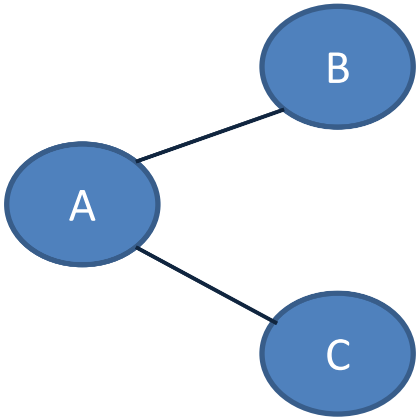


1/2

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

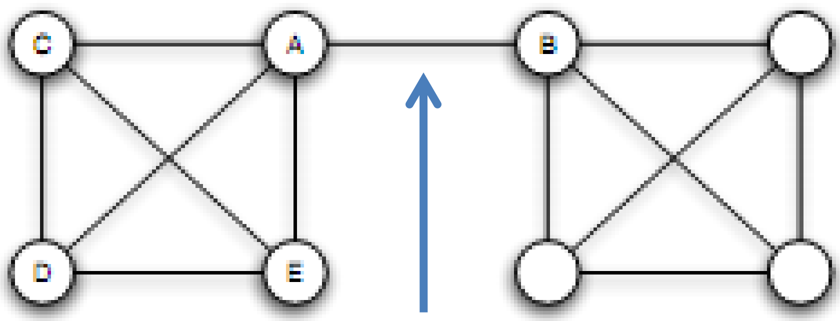
Mark Granovetter, in the late 1960s

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



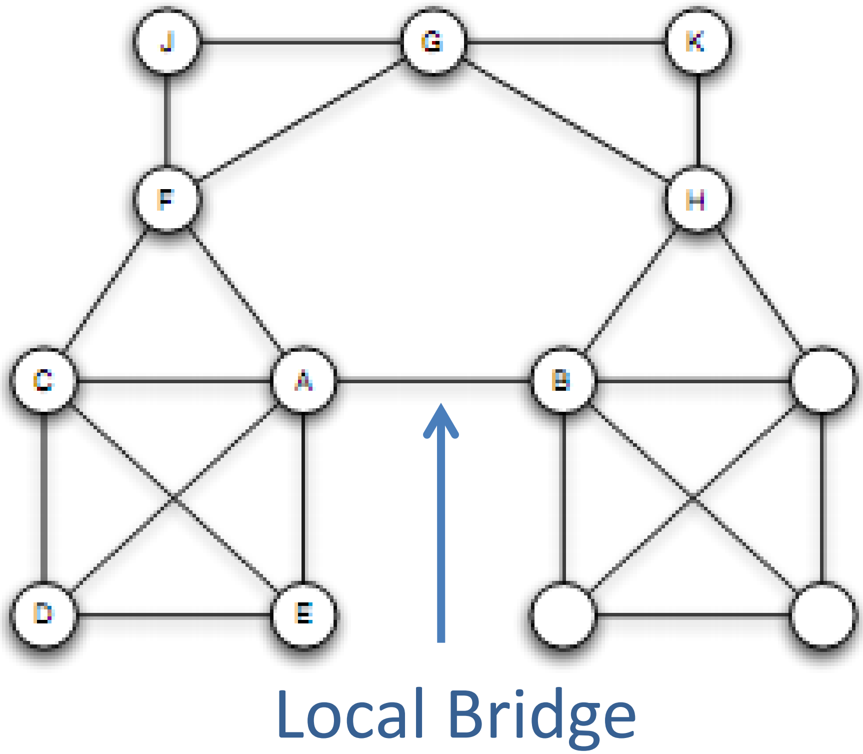
Bridge
(aka cut-edge)

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

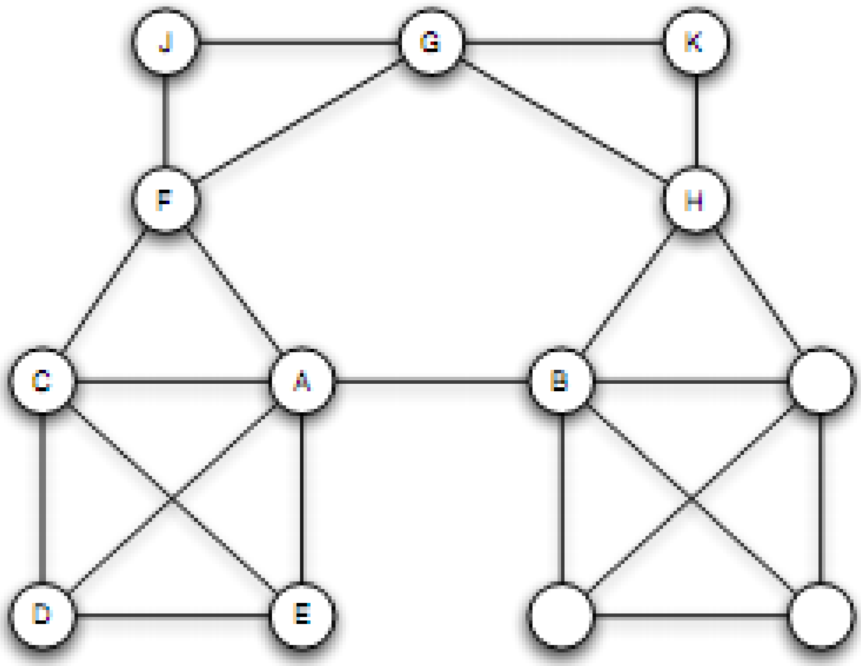
Bridges and Local Bridges



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

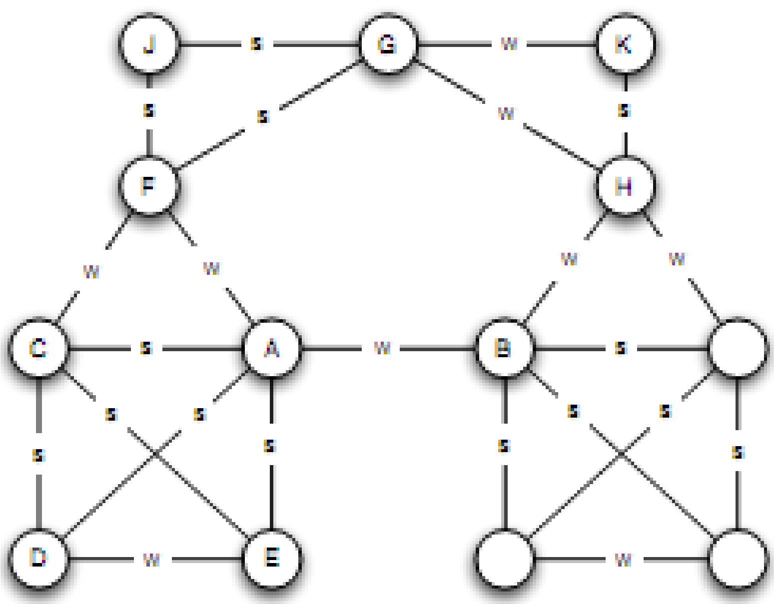


An edge is a local bridge, if an only if, it is not part of any triangle in the graph

The Strong Triadic Closure Property

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

Annotated graph

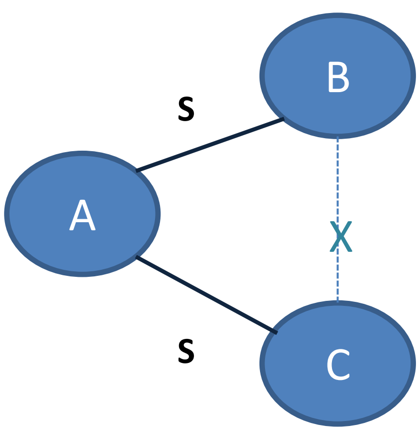


The Strong Triadic Closure Property

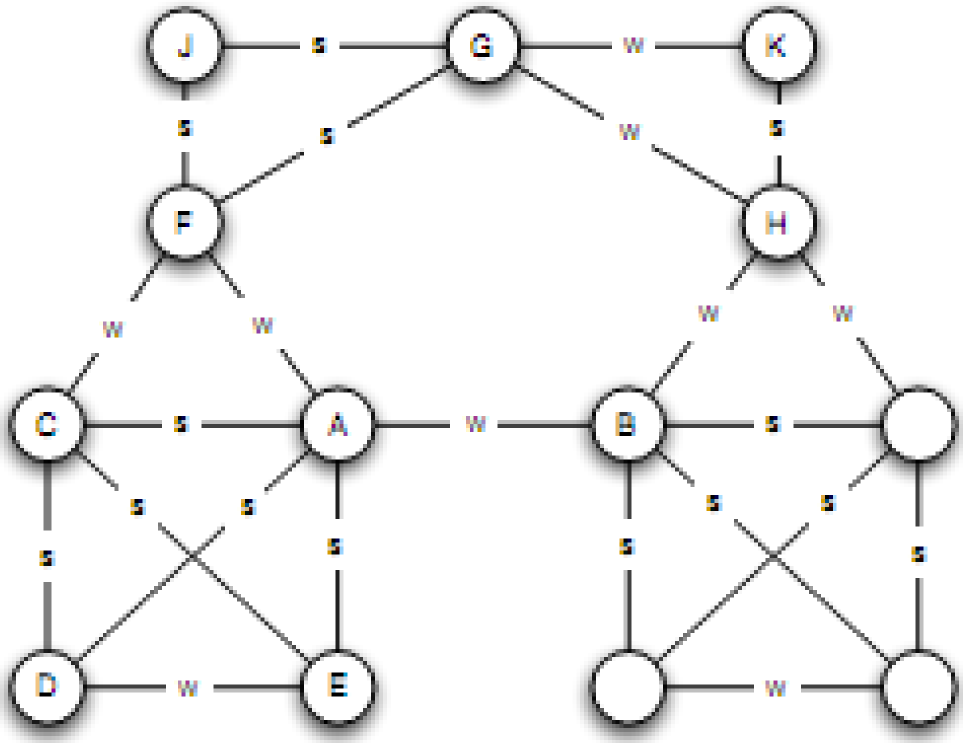
If a node A has edges to nodes B and C, then the B-C edge is <u>especially likely</u> to form if both A-B and A-C are <u>strong ties</u>

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

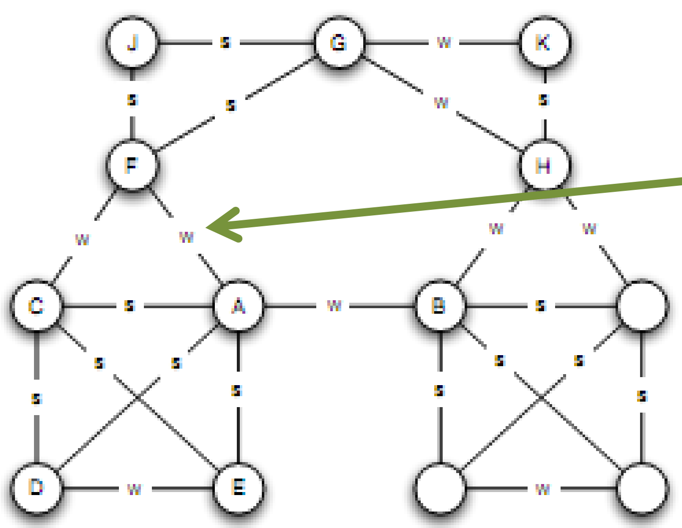
"almost" local bridges

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

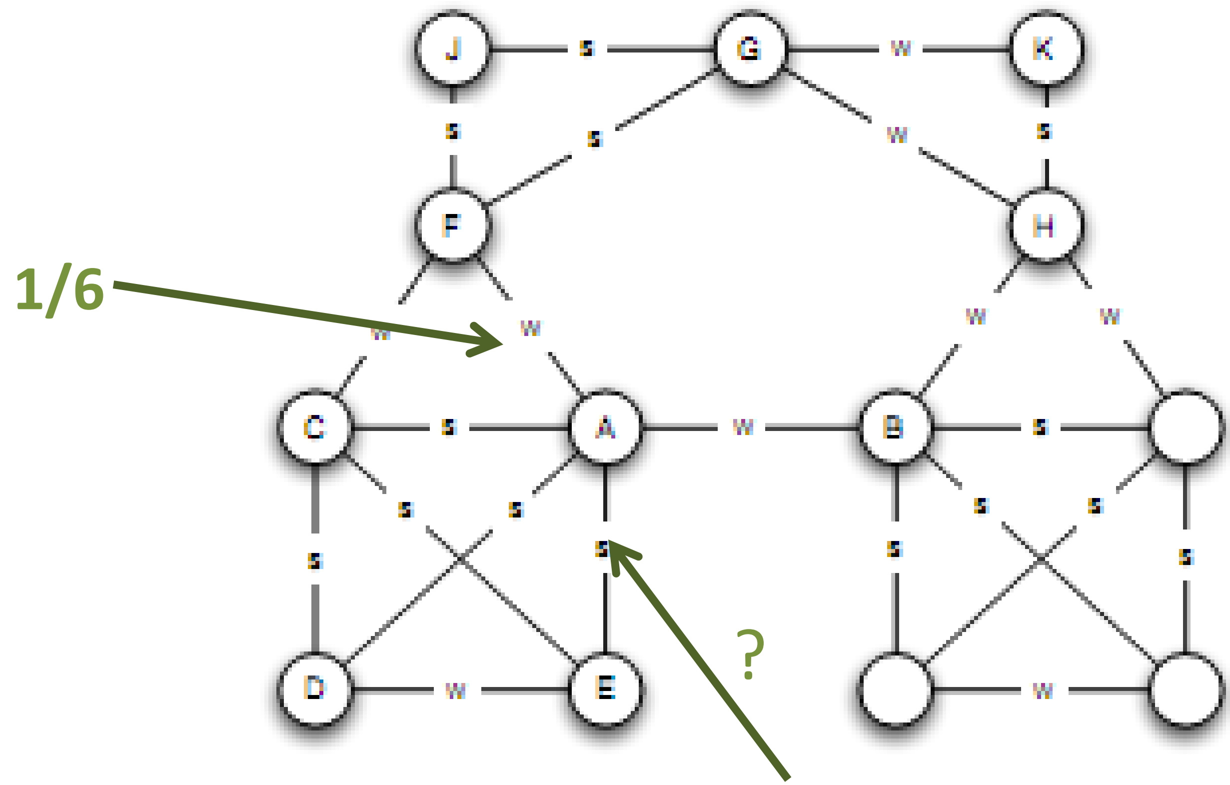
F: C, J, G

1/6

When is this value 0?

Generalizing Weak Ties and Local Bridges

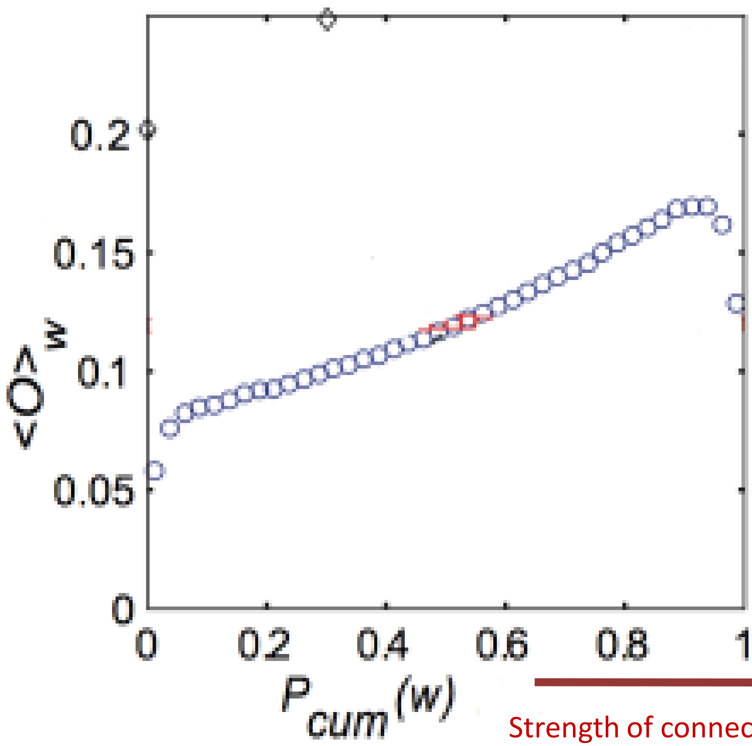
Neighborhood overlap = 0: edge is a local bridge
Small value: “almost” local bridges



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*

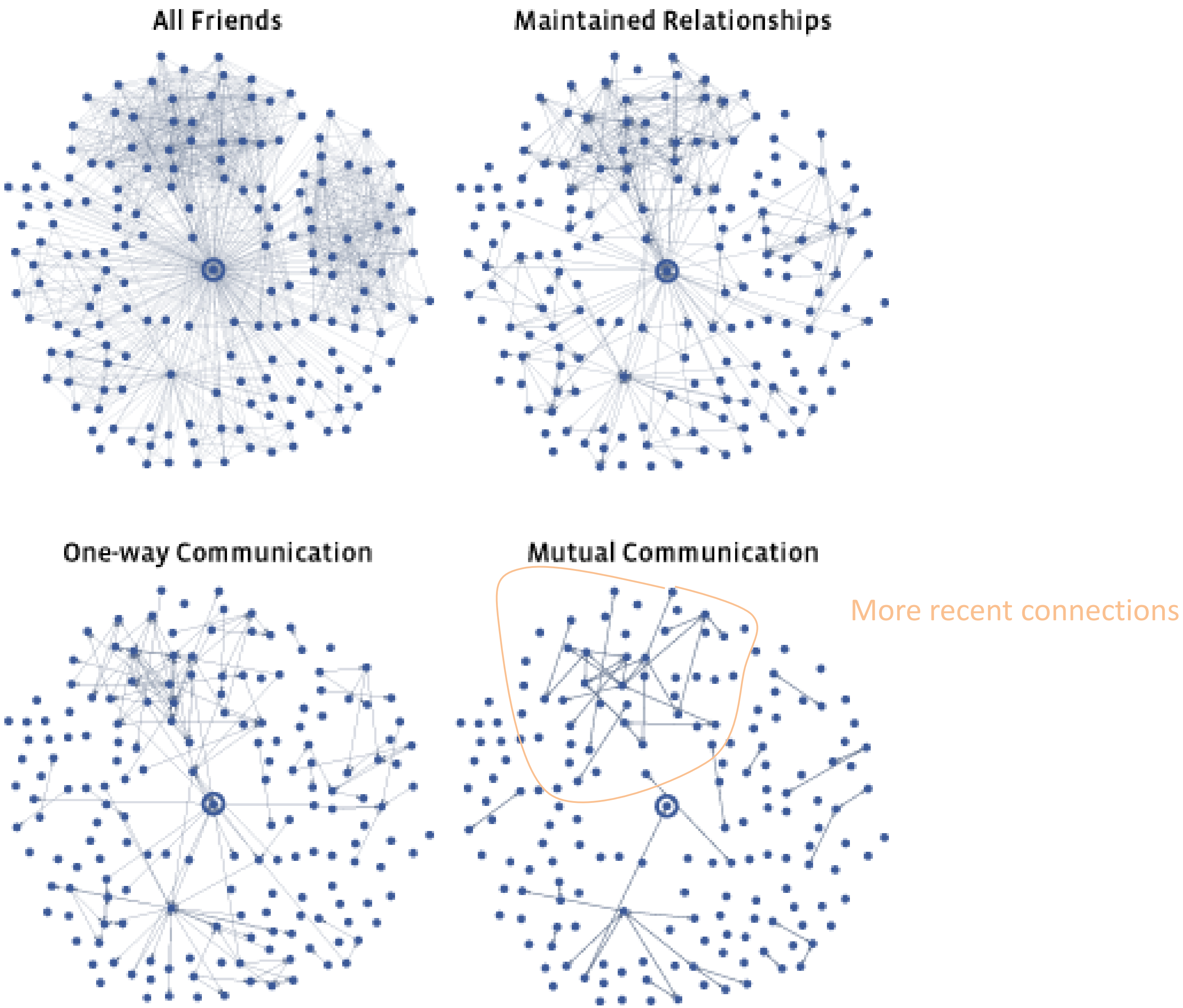
Tie Strength on Facebook

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

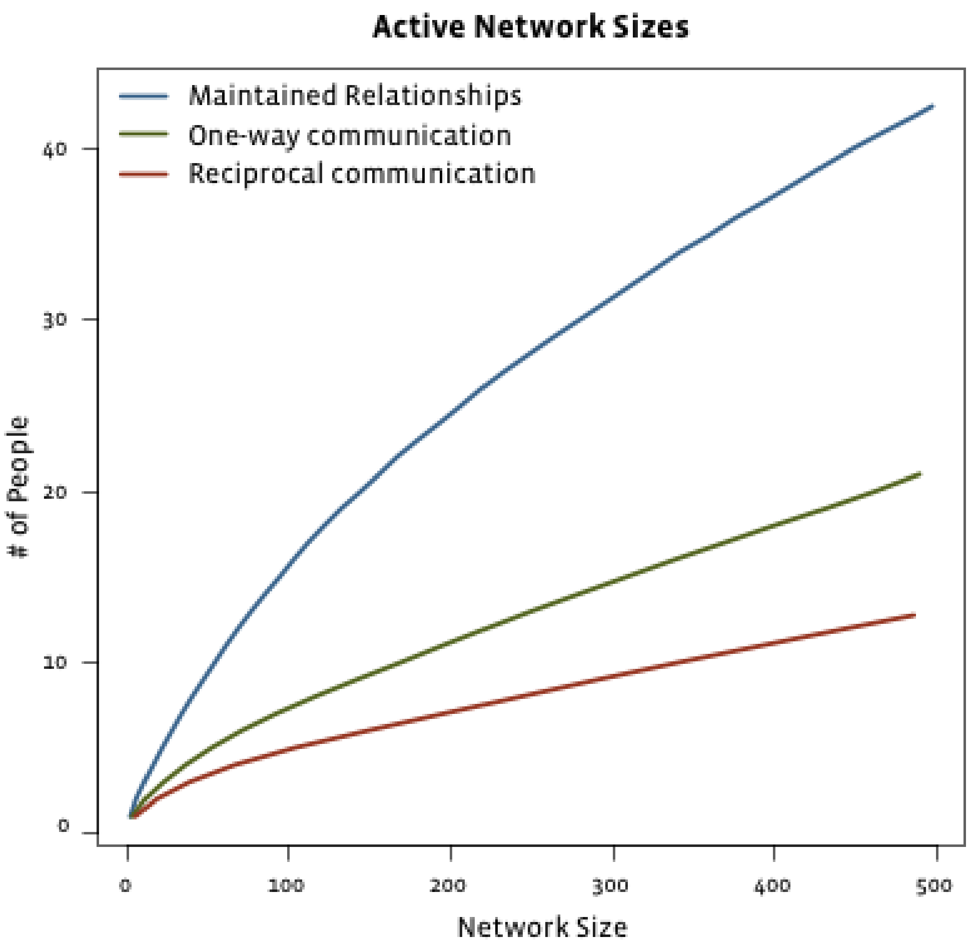
Three (not exclusive) kinds of ties (links)

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

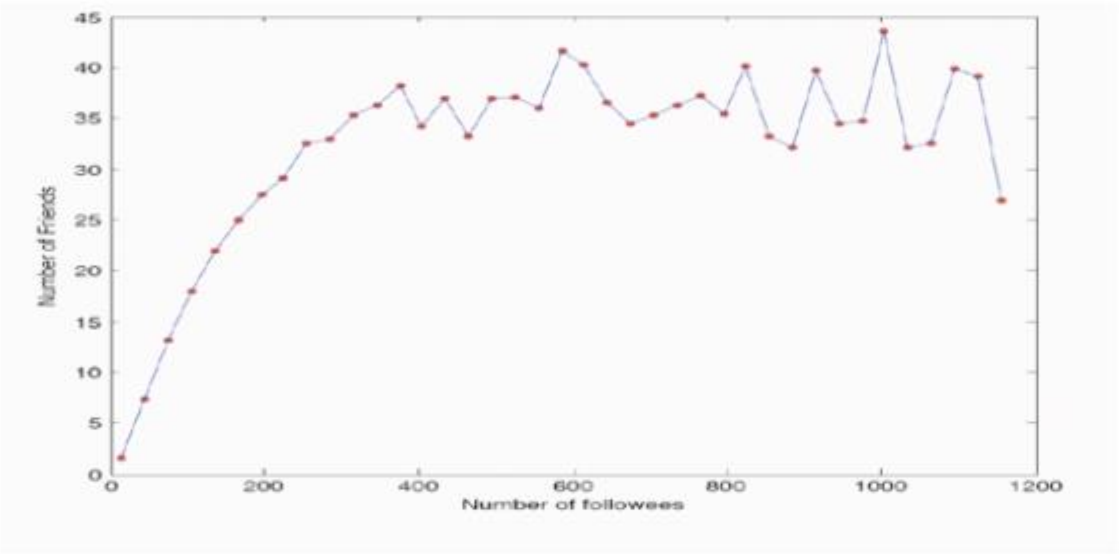
Passive engagement (keep up with friends by reading about them even in the absence of communication)

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)

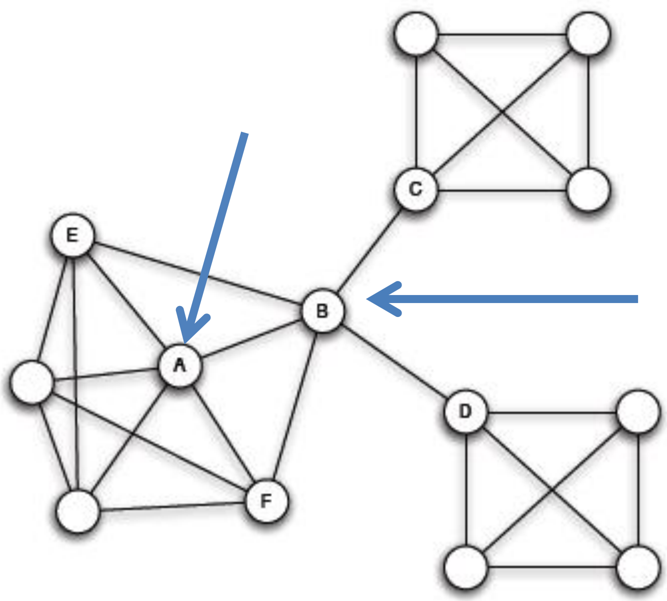
- Network of strong ties still remain sparse

- 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

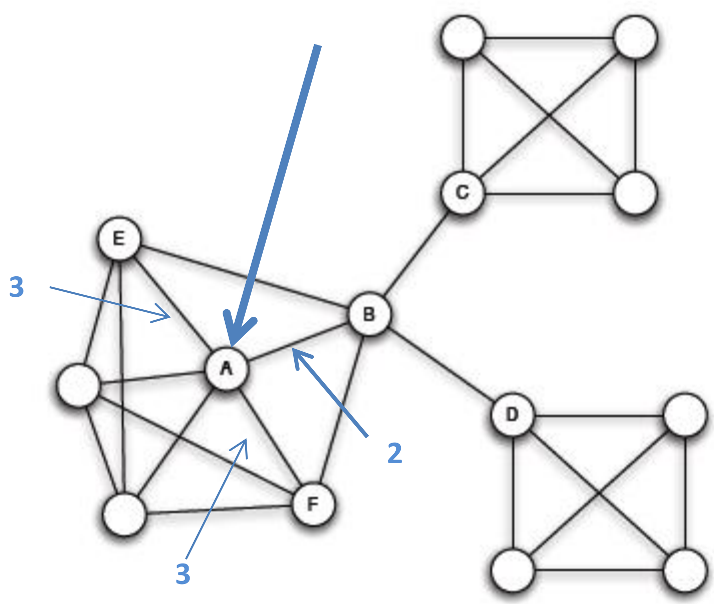


Embeddedness

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"

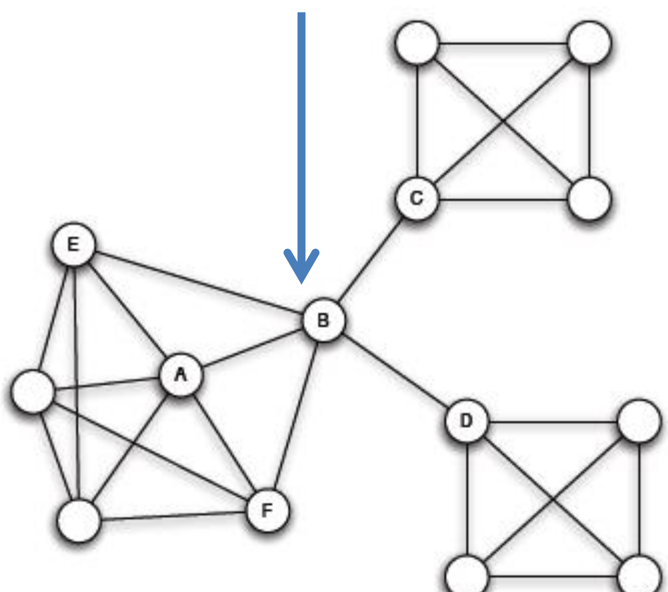
Structural Holes

(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"

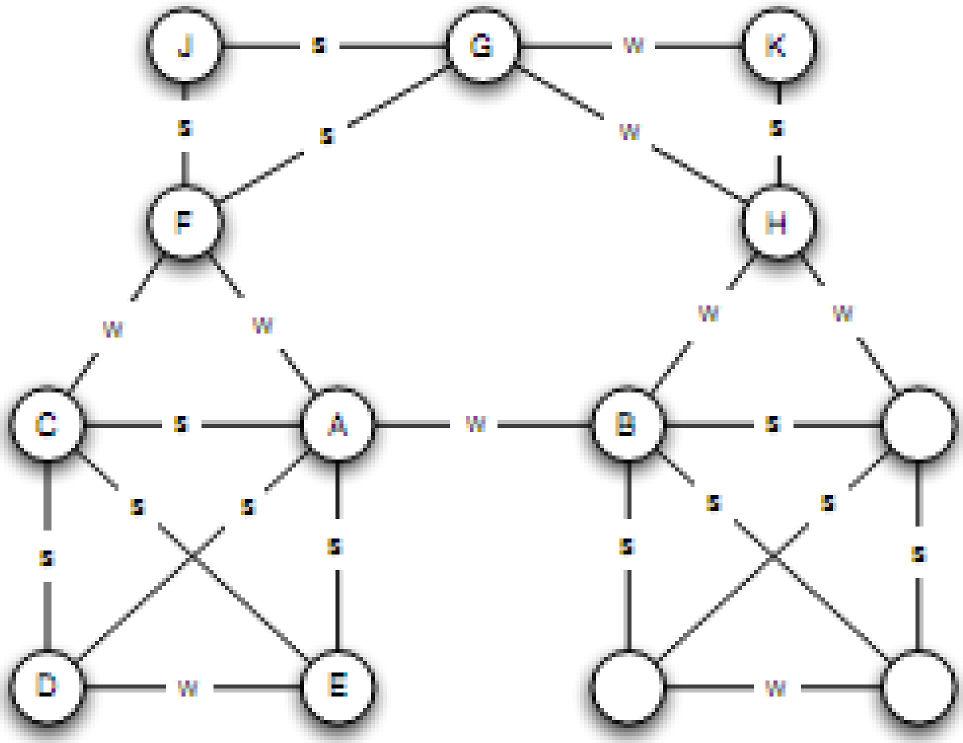
- B has access to information originating in multiple, non interacting parts of the network
- An amplifier for creativity
- Source of power as a social "gate-keeping"

Social capital



ENFORCING STRONG TRIADIC CLOSURE

The Strong Triadic Closure Property



If we do not have the labels, how can we label the edges so as to satisfy the Strong Triadic Closure Property?

Problem Definition

- Goal: Label (color) ties of a social network as Strong or Weak so that the Strong Triadic Closure property holds.

- MaxSTC Problem: Find an edge labeling (S, W) that satisfies the STC property and maximizes the number of Strong edges.

- MinSTC Problem: Find an edge labeling (S, W) that satisfies the STC property and minimizes the number of Weak edges.

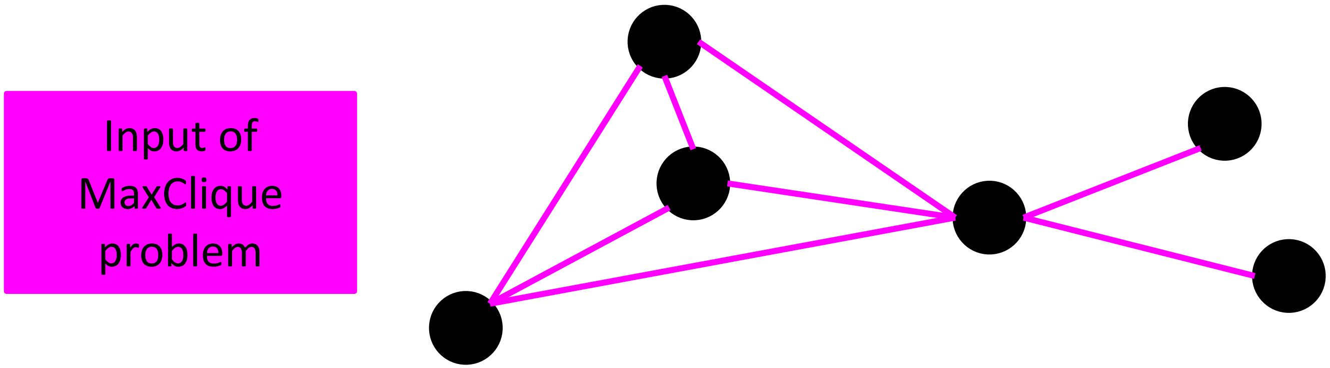
Complexity

- Bad News: MaxSTC and MinSTC are NP-hard problems!
  - Reduction from MaxClique to the MaxSTC problem.

- MaxClique: Given a graph $G = (V, E)$, find the maximum subset $V \subseteq V$ that defines a complete subgraph.

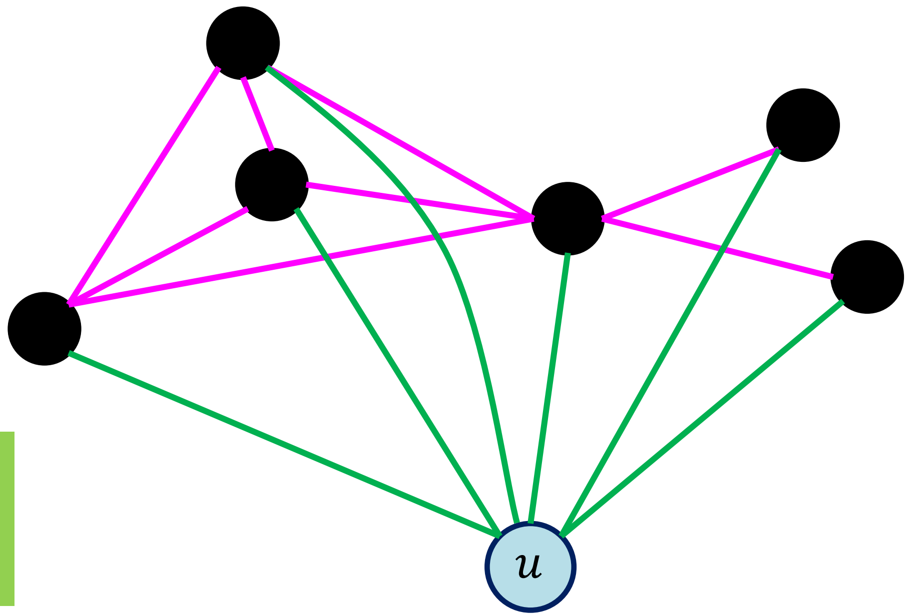
Reduction

- Given a graph G as input to the MaxClique problem



Reduction

- Given a graph G as input to the MaxClique problem
- Construct a new graph by adding a node u and a set of edges $\boldsymbol{E_u}$ to all nodes in G



MaxEgoSTC is at least as hard as MaxSTC

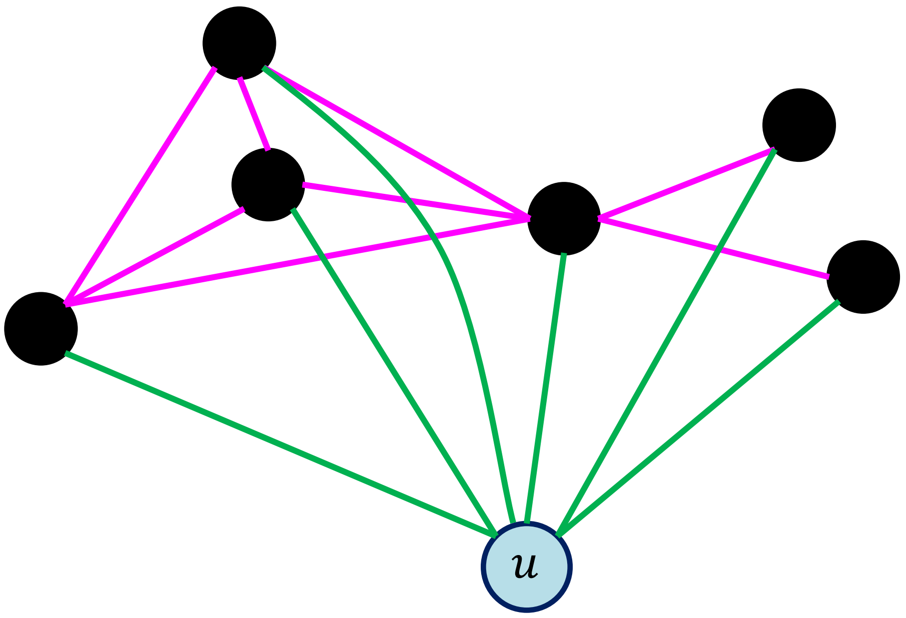
The labelings of pink and green edges are independent

MaxEgoSTC: Label the edges in $\boldsymbol{E_u}$ into Strong or Weak so as to satisfy STC and maximize the number of Strong edges

Reduction

- Given a graph G as input to the MaxClique problem
- Construct a new graph by adding a node u and a set of edges $\boldsymbol{E_u}$ to all nodes in G

Input to the MaxEgoSTC problem



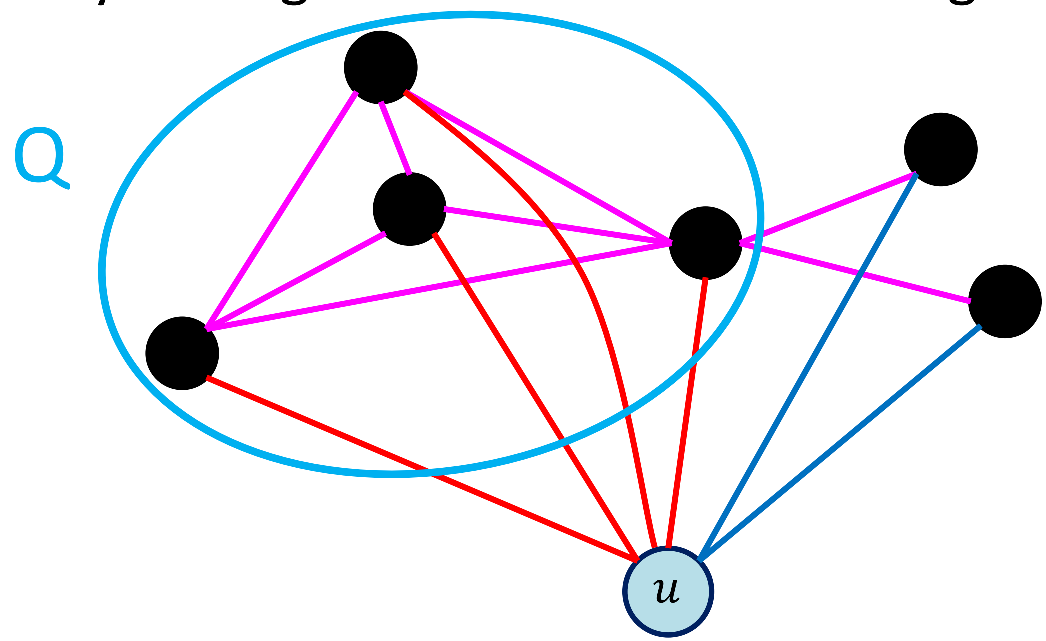
MaxEgoSTC: Label the edges in $\boldsymbol{E_u}$ into Strong or Weak so as to satisfy STC and maximize the number of Strong edges

Reduction

- Given a graph G as input to the MaxClique problem
- Construct a new graph by adding a node u and a set of edges $\boldsymbol{E_u}$ to all nodes in G

Find the max clique Q in G

Maximize Strong edges in $\boldsymbol{E_u}$



MaxEgoSTC: Label the edges in $\boldsymbol{E_u}$ into Strong or Weak so as to satisfy STC and maximize the number of Strong edges

Approximation Algorithms

- Bad News: MaxSTC is hard to approximate.
- Good News: There exists a 2-approximation algorithm for the MinSTC problem.
  - The number of weak edges it produces is at most two times those of the optimal solution.

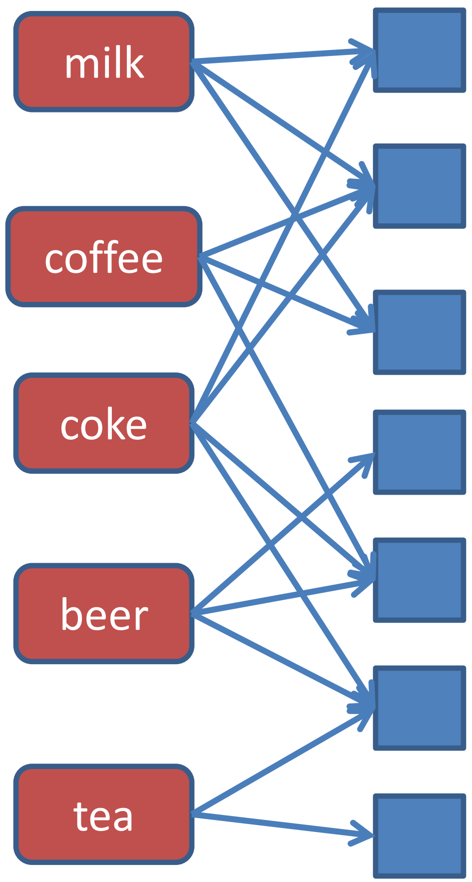
- The algorithm comes by reducing our problem to a coverage problem

Set Cover

- The Set Cover problem:
  - We have a universe of elements $U = \{x_1, \dots, x_N\}$
  - We have a collection of subsets of U, $\boldsymbol{S} = \{S_1, \dots, S_n\}$, such that $\bigcup_i S_i = U$
  - We want to find the smallest sub-collection $\boldsymbol{C} \subseteq \boldsymbol{S}$ of $\boldsymbol{S}$, such that $\bigcup_{S_i \in \boldsymbol{C}} S_i = U$
    - The sets in $\boldsymbol{C}$ cover the elements of U

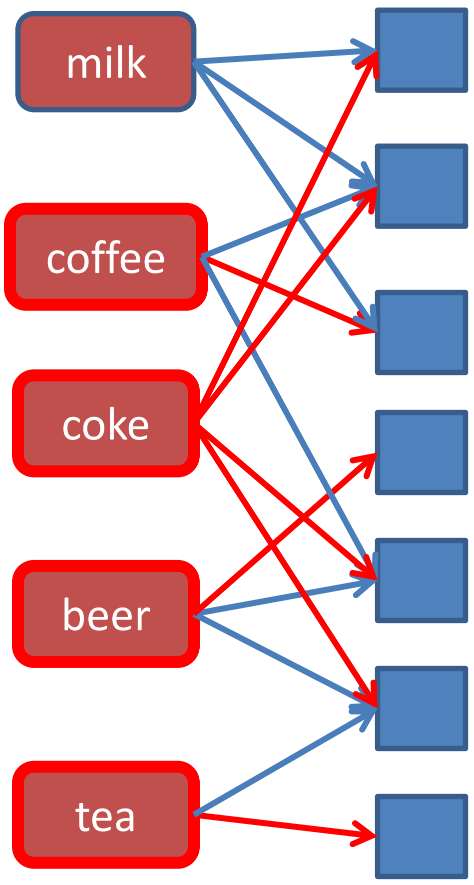
Example

- The universe U of elements is the set of customers of a store.
- Each set corresponds to a product p sold in the store: $S_p = \{customers\ that\ bought\ p\}$
- Set cover: Find the minimum number of products (sets) that cover all the customers (elements of the universe)



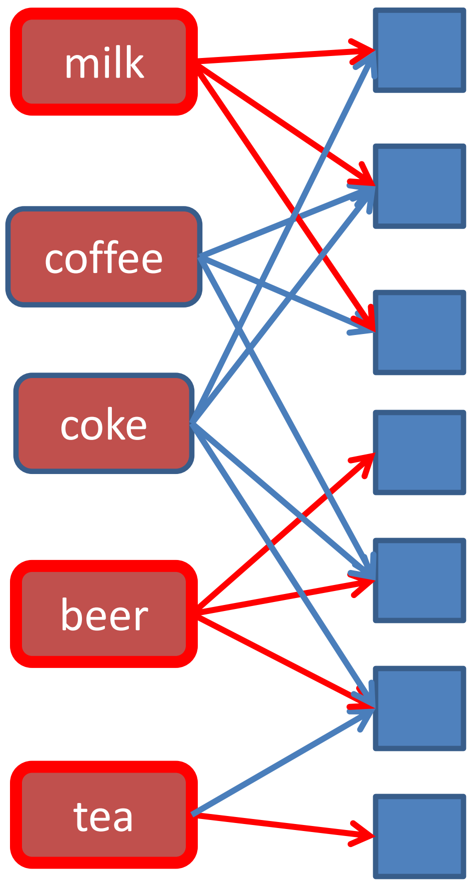
Example

- The universe U of elements is the set of customers of a store.
- Each set corresponds to a product p sold in the store:
  $S_p = \{customers\ that\ bought\ p\}$
- Set cover: Find the minimum number of products (sets) that cover all the customers (elements of the universe)



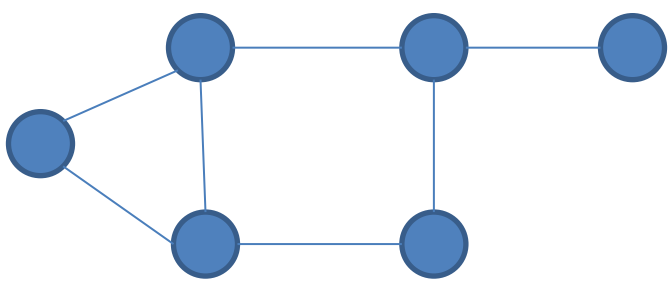
Example

- The universe U of elements is the set of customers of a store.
- Each set corresponds to a product p sold in the store:
  $S_p = \{customers\ that\ bought\ p\}$
- Set cover: Find the minimum number of products (sets) that cover all the customers (elements of the universe)



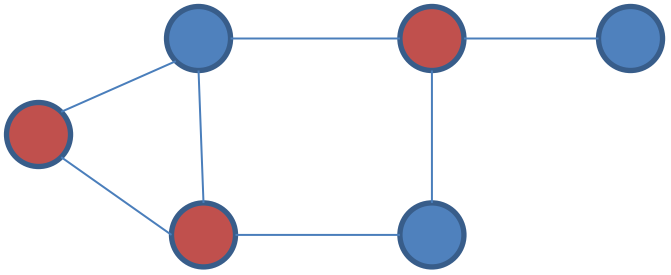
Vertex Cover

- Given a graph $G = (V, E)$ find a subset of vertices $S \subseteq V$ such that for each edge $e \in E$ at least one endpoint of e is in S.
  - Special case of set cover, where all elements are edges and sets the set of edges incident on a node.
    - Each element is covered by exactly two sets



Vertex Cover

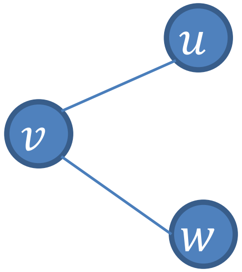
- Given a graph $G = (V, E)$ find a subset of vertices $S \subseteq V$ such that for each edge $e \in E$ at least one endpoint of e is in S.
  – Special case of set cover, where all elements are edges and sets the set of edges incident on a node.
    - Each element is covered by exactly two sets



MinSTC and Coverage

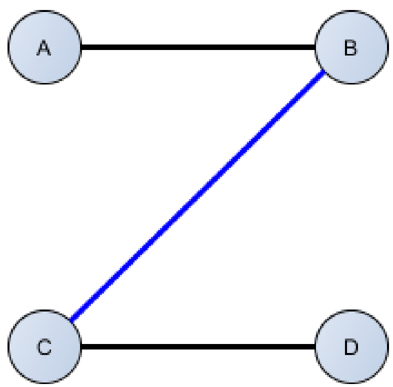
- What is the relationship between the MinSTC problem and Coverage?

- Hint: A labeling satisfies STC if for any two edges (u, v) and (v, w) that form an open triangle at least one of the edges is labeled weak



Coverage

- Intuition
  - STC property implies that there cannot be an open triangle with both strong edges
  - For every open triangle: a weak edge must cover the triangle

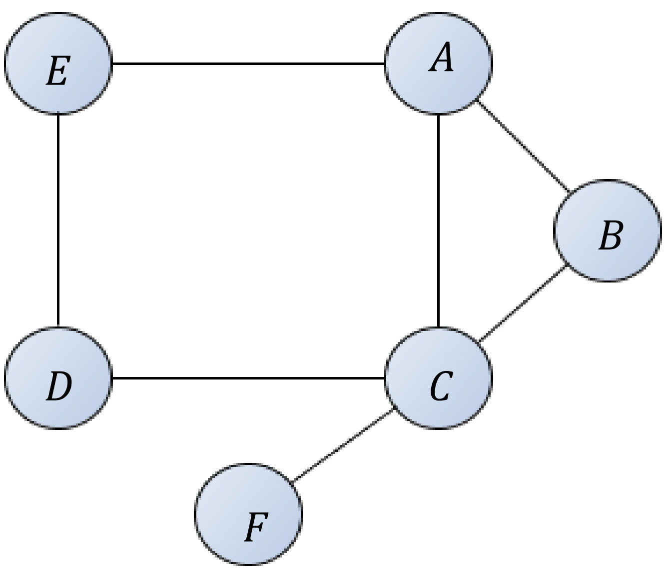


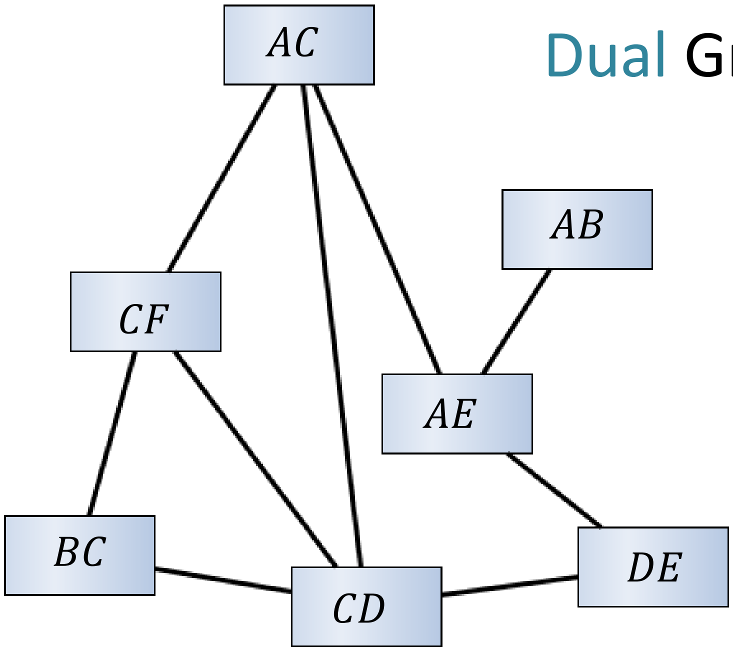
  - MinSTC can be mapped to the Minimum Vertex Cover problem.

Dual Graph

- Given a graph G, we create the dual graph D:
  - For every edge in G we create a node in D.
  - Two nodes in D are connected if the corresponding edges in G participate in an open triangle.

Initial Graph G

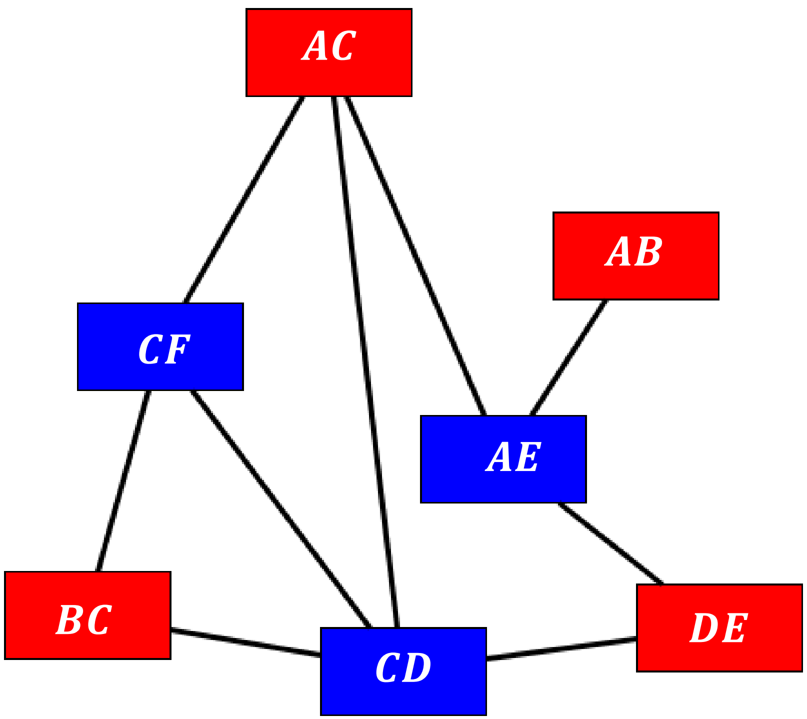


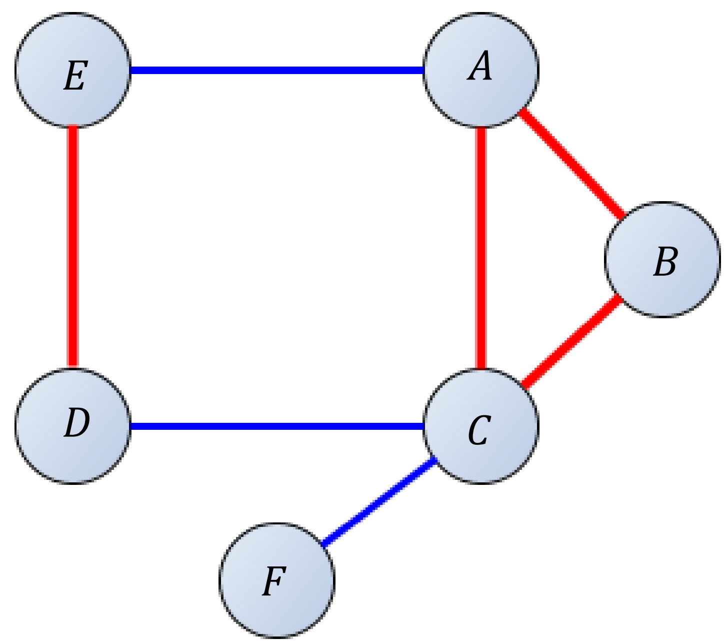


Dual Graph D

Minimum Vertex Cover - MinSTC

- Solving MinSTC on G is reduced to solving a Minimum Vertex Cover problem on D.





Approximation Algorithms

Approximation algorithms for the Minimum Vertex Cover problem:

<table>
<tr>
<td>

Maximal Matching Algorithm

- Output a maximal matching
  - Maximal Matching: A collection of non-adjacent edges of the graph where no additional edges can be added.

Approximation Factor: 2

</td>
<td>

Greedy Algorithm

- Greedily select each time the vertex that covers most uncovered edges.

Approximation Factor: log n

</td>
</tr>
</table>

Given a vertex cover for dual graph D, the corresponding edges of G are labeled Weak and the remaining edges Strong.

Experiments

- Experimental Goal: Does our labeling have any practical utility?

Datasets

- Actors: Collaboration network between movie actors. (IMDB)
- Authors: Collaboration network between authors. (DBLP)
- Les Miserables: Network of co-appearances between characters of Victor Hugo's novel. (D. E. Knuth)
- Karate Club: Social network of friendships between 34 members of a karate club. (W. W. Zachary)
- Amazon Books: Co-purchasing network between books about US politics. (http://www.orgnet.com/)

Dataset	Number of Nodes	Number of Edges
Actors	1,986	103,121
Authors	3,418	9,908
Les Miserables	77	254
Karate Club	34	78
Amazon Books	105	441

Comparison of Greedy and MaximalMatching

	Greedy			Maximal Matching	
	Strong	Weak		Strong	Weak
Actors	11,184	91,937		8,581	94,540
Authors	3,608	6,300		2,676	7,232
Les Miserables	128	126		106	148
Karate Club	25	53		14	64
Amazon Books	114	327		71	370

Measuring Tie Strength

- Question: Is there a correlation between the assigned labels and the empirical strength of the edges?
- Three weighted graphs: Actors, Authors, Les Miserables.
  - Strength: amount of common activity.

Mean activity intersection for Strong, Weak Edges

	Strong	Weak
Actors	1.4	1.1
Authors	1.34	1.15
Les Miserables	3.83	2.61

- The differences are statistically signicant

Measuring Tie Strength

- Frequent common activity may be an artifact of frequent activity.
- Fraction of activity devoted to the relationship
  - Strength: Jaccard Similarity of activity

$$\text{Jaccard Similarity} = \frac{\text{Common Activities}}{\text{Union of Activities}}$$

Mean Jaccard similarity for Strong, Weak Edges

	Strong	**Weak**
Actors	0.06	0.04
Authors	0.145	0.084

- The differences are statistically signicant

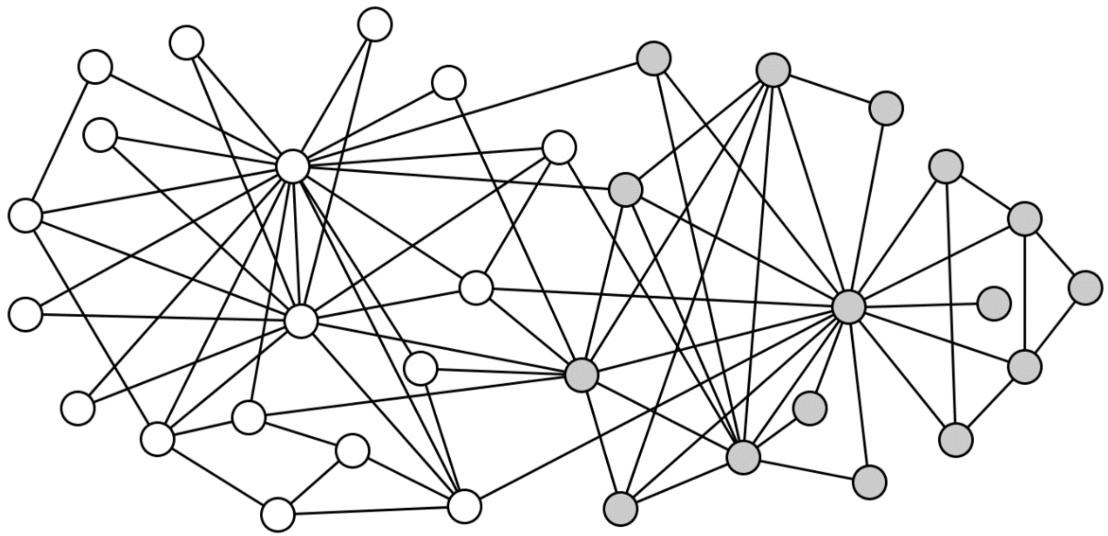
The Strength of Weak Ties

- [Granovetter] People learn information leading to jobs through acquaintances (Weak ties) rather than close friends (Strong ties).

- [Easly and Kleinberg] Graph theoretic formalization:
  – Acquaintances (Weak ties) act as bridges between different groups of people with access to different sources of information.
  – Close friends (Strong ties) belong to the same group of people, and are exposed to similar sources of information.

Datasets with known communities

- Amazon Books
  - US Politics books : liberal, conservative, neutral.
- Karate Club
  - Two fractions within the members of the club.

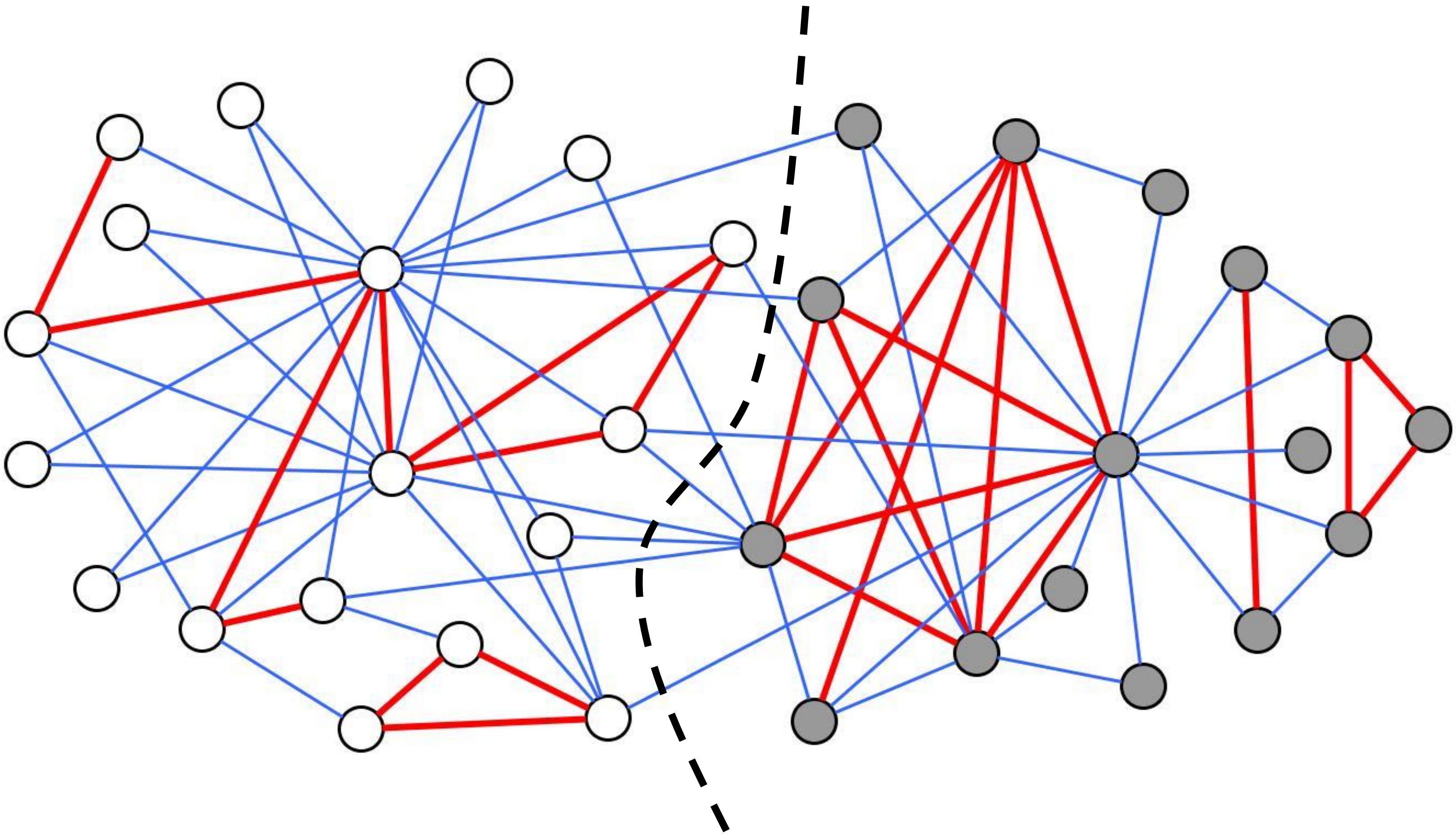


Weak Edges as Bridges

- Edges between communities (inter-community) ⇒ Weak
  - R_W = Fraction of inter-community edges that are labeled Weak.
- Strong ⇒ Edges within the community (intra-community).
  - P_S = Fraction of Strong edges that are intra-community edges

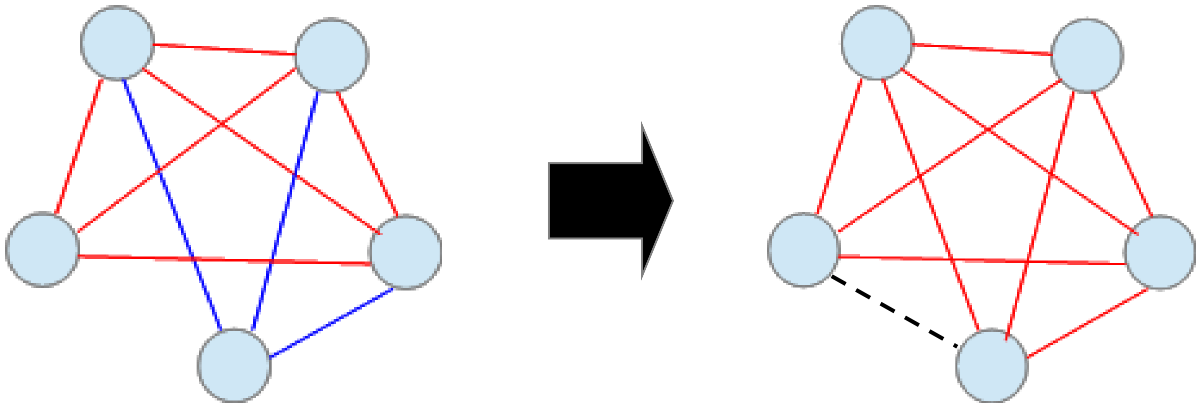
	P_S	R_W
Karate Club	1	1
Amazon Books	0.81	0.69

Karate Club graph



Extensions

- Allow for edge additions



  - Still a coverage problem: an open triangle can be covered with either a weak edge or an added edge
- Allow k types of strong of edges
  - Vertex Coloring of the dual graph with a neutral color
  - Approximation algorithm for k=2 types, hard to approximate for k > 2

POSITIVE AND NEGATIVE TIES

Structural Balance

What about negative edges?

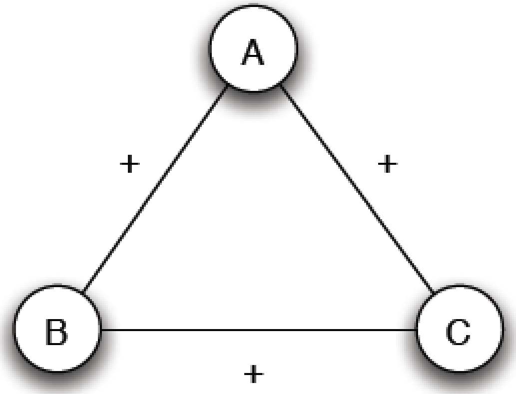
Initially, a complete graph (or clique): every edge either + or -

Let us first look at individual triangles

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

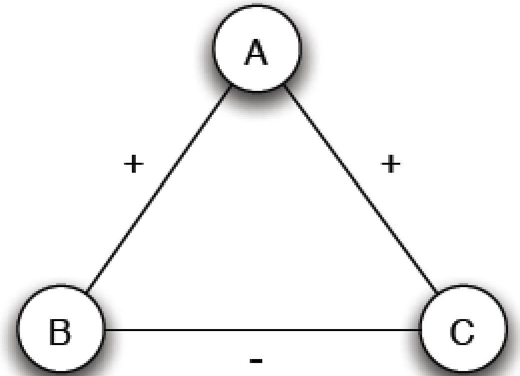
Structural Balance

Case (a): 3 +



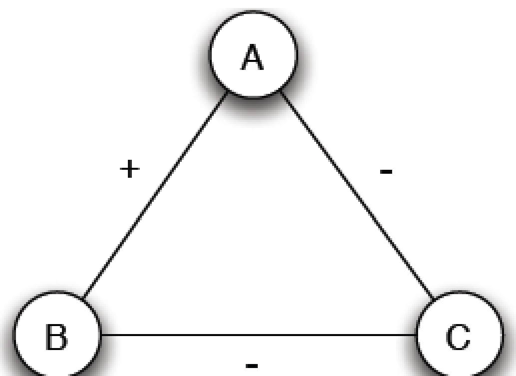
Mutual friends

Case (b): 2 +, 1 -



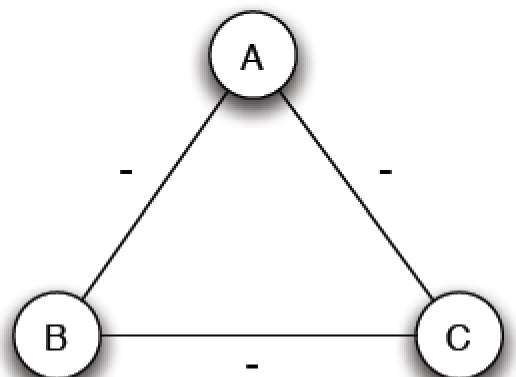
A is friend with B and C, but B and C do not get well together

Case (c): 1 +, 2 -



A and B are friends with a mutual enemy

Case (d): 3 -

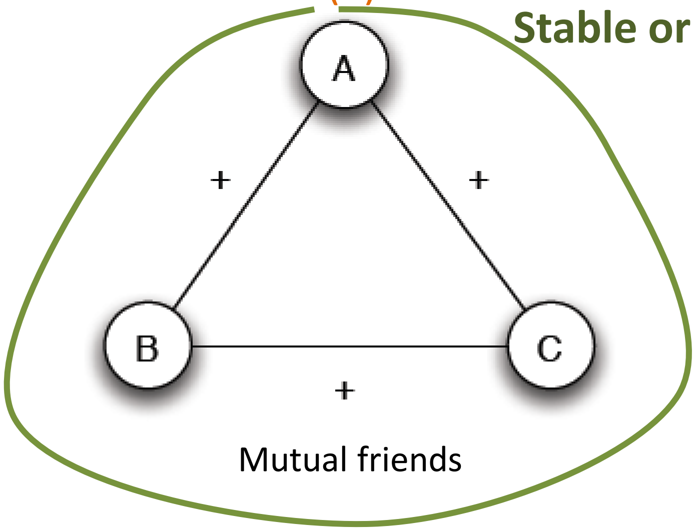


Mutual enemies

Structural Balance

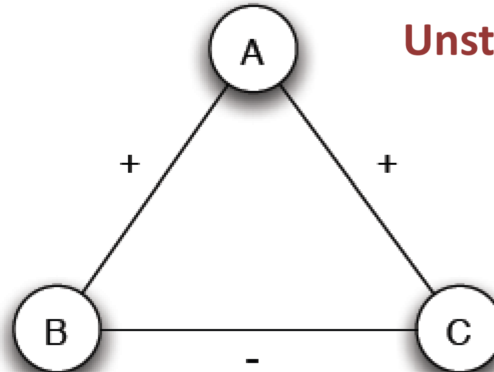
Case (a): 3 +

Stable or balanced



Case (b): 2 +, 1 -

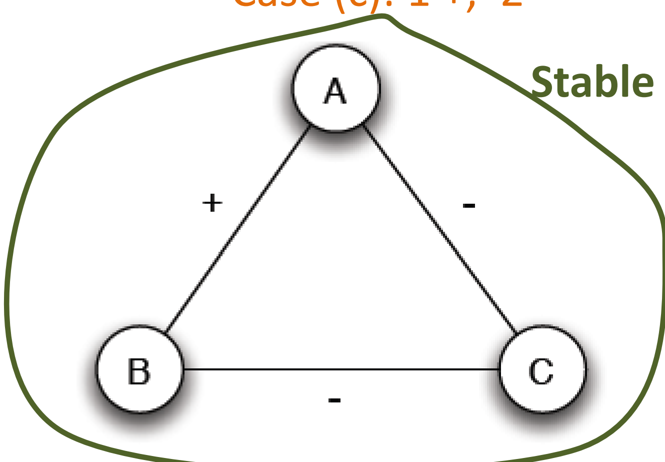
Unstable



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 -

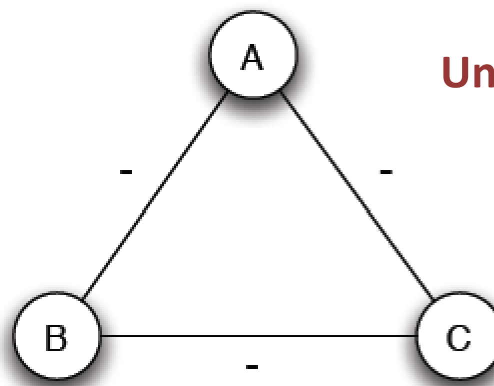
Stable or balanced



A and B are friends with a mutual enemy

Case (d): 3 -

Unstable

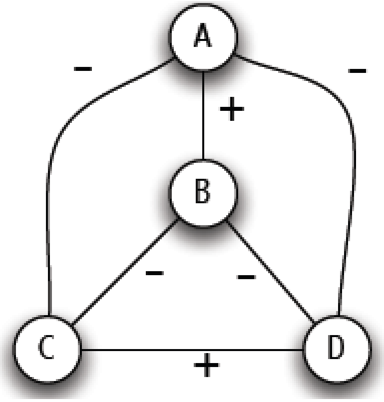


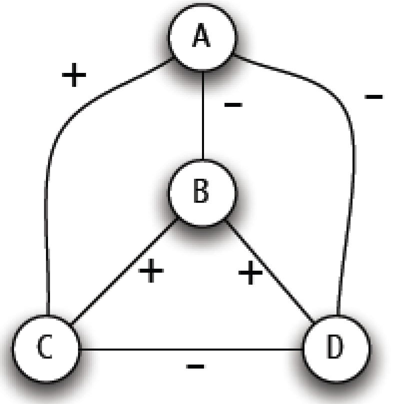
Mutual enemies
Forces to team up against the third (turn 1 – to +)

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?

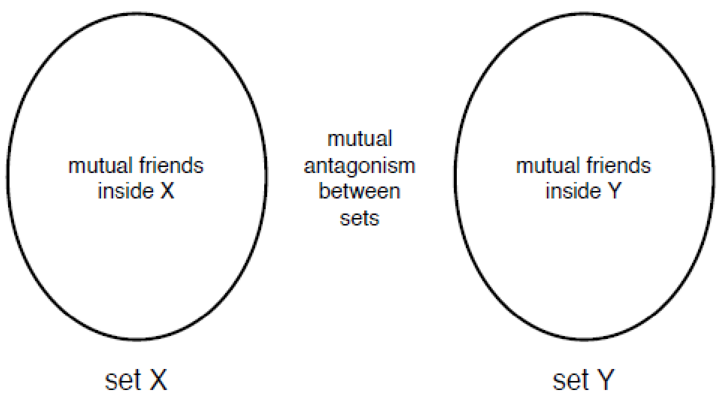
The Structure of Balanced Networks

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

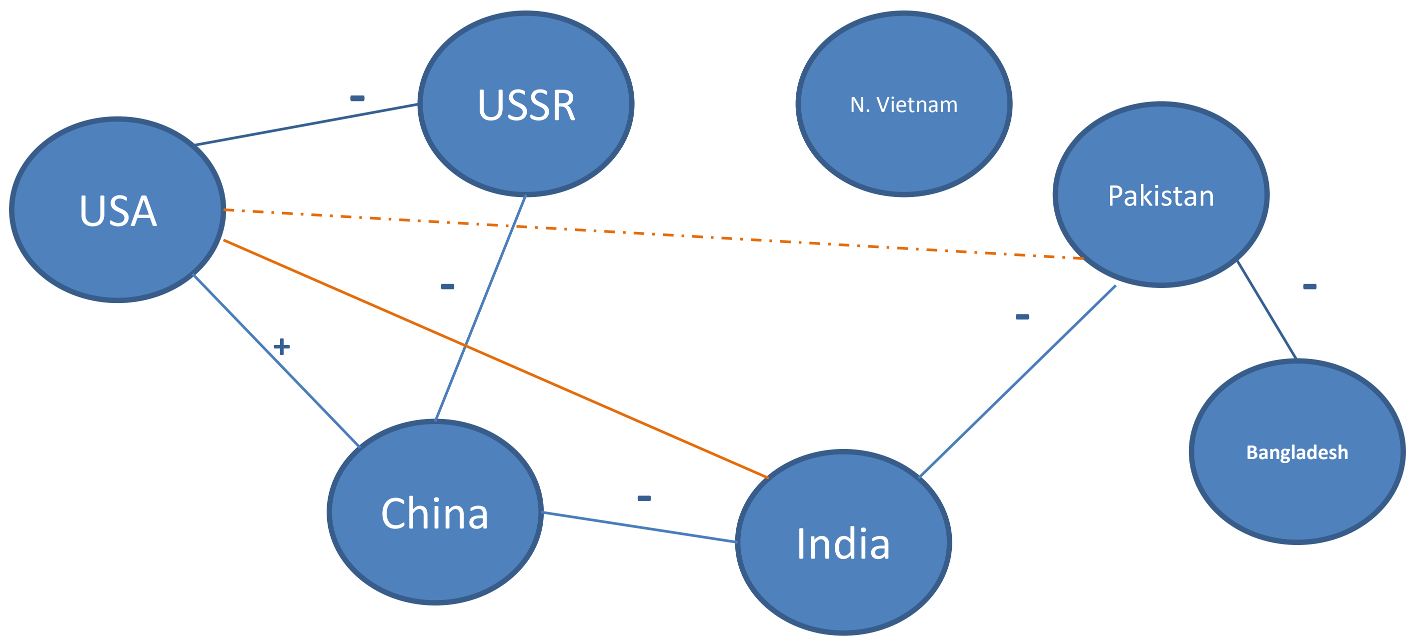


Proof ...

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)

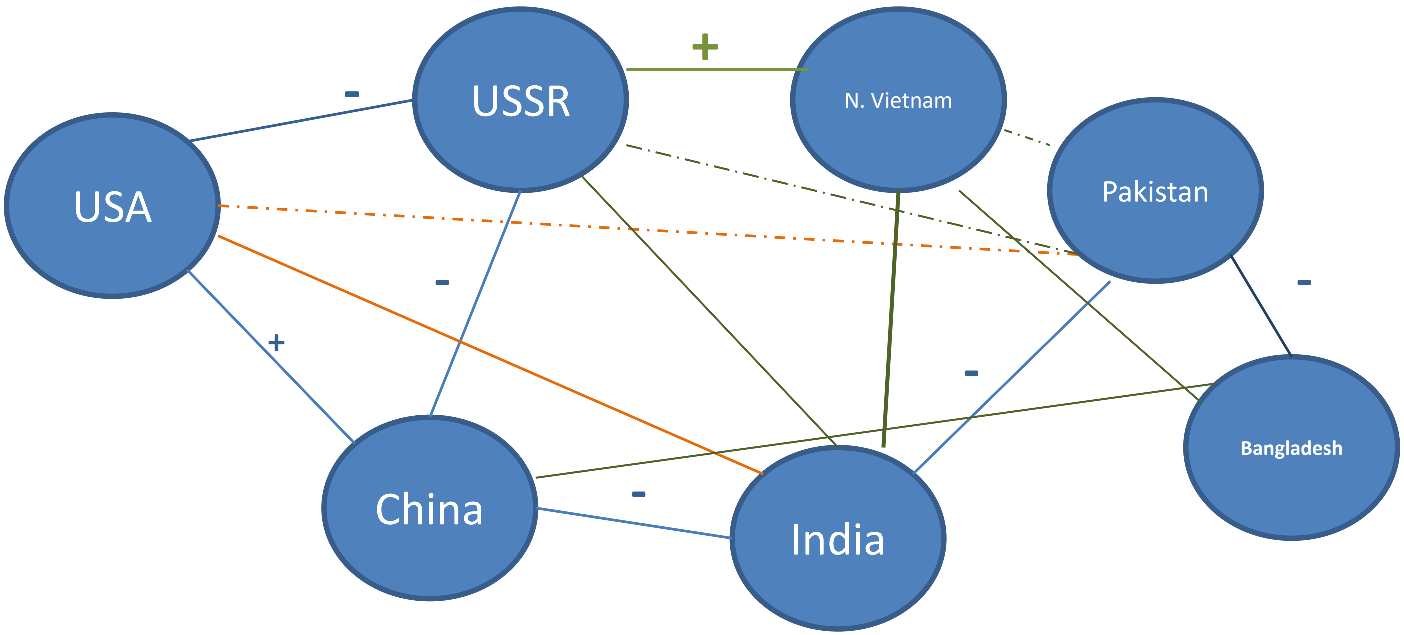


USA support to Pakistan?

Applications of Structural Balance

✓ International relationships (I)

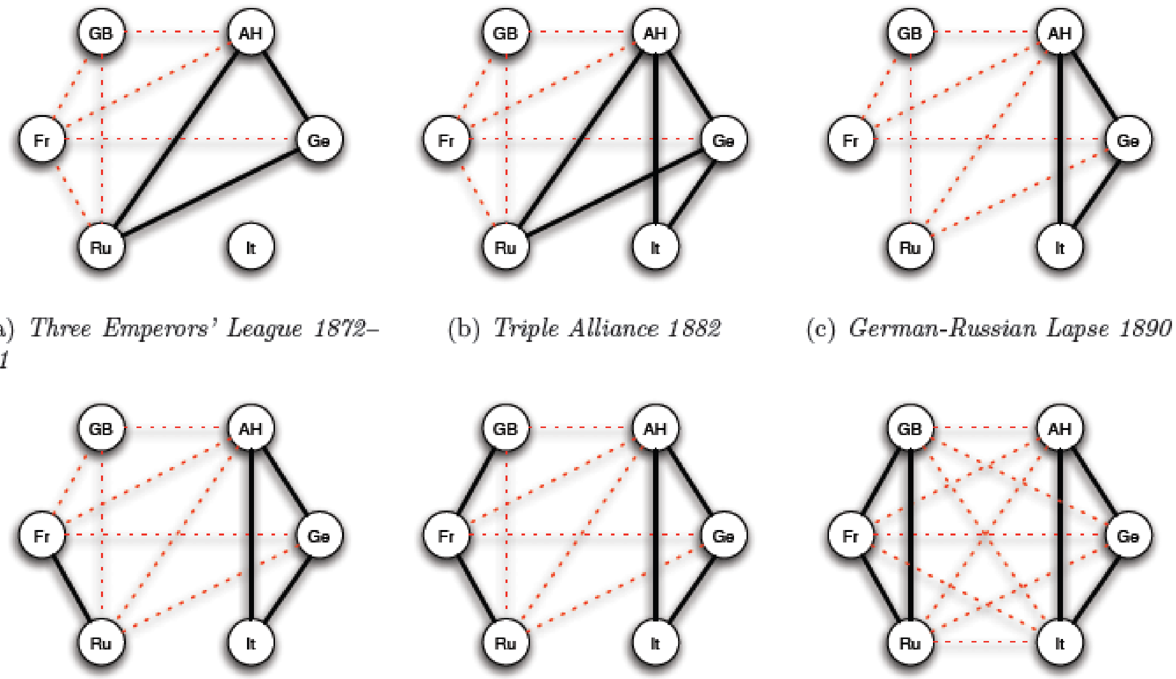
The conflict of Bangladesh's separation from Pakistan in 1972 (II)



China?

Applications of Structural Balance

✓ International relationships (II)

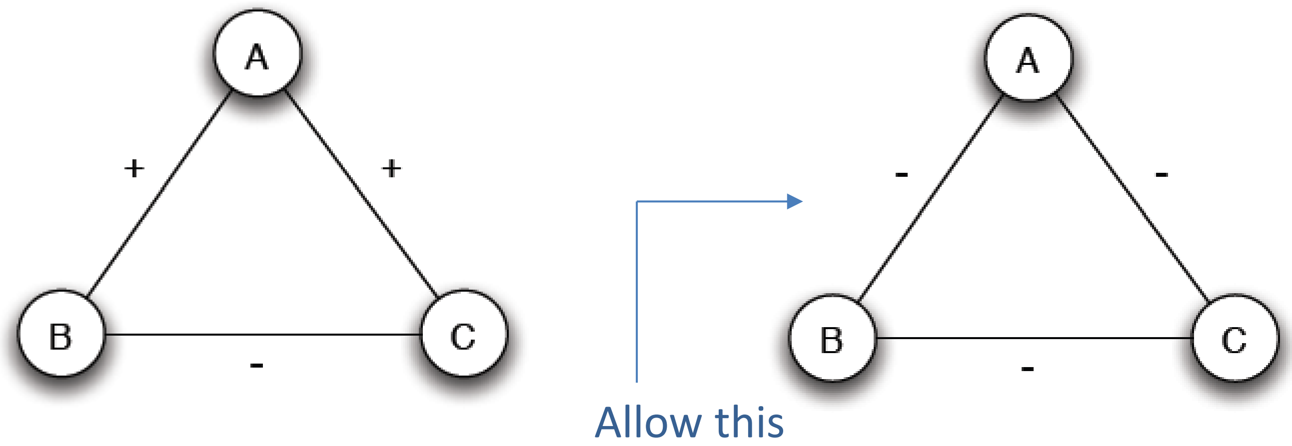


(a) *Three Emperors' League 1872–81* (b) *Triple Alliance 1882* (c) *German-Russian Lapse 1890*

(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



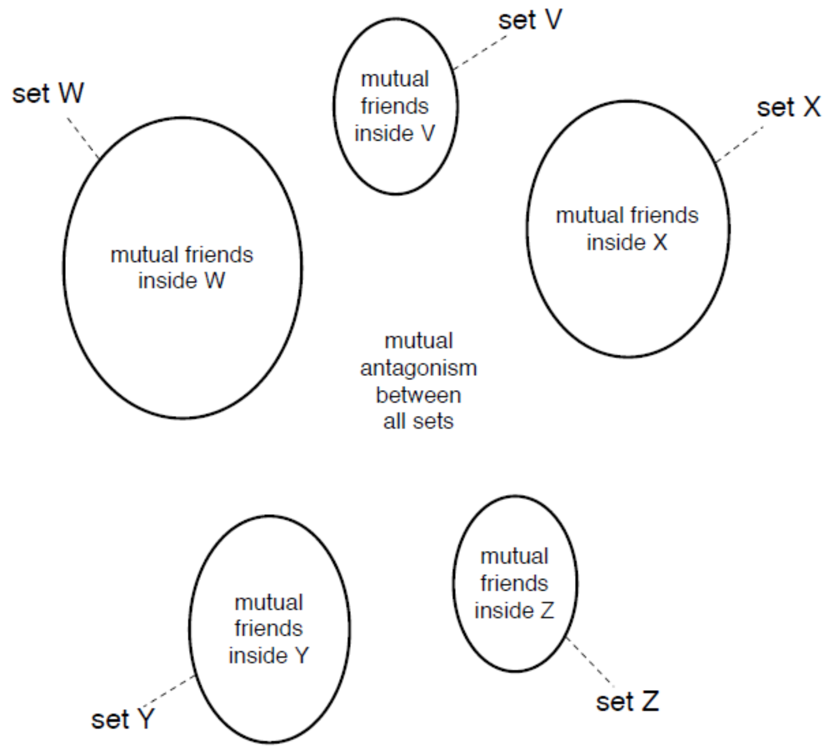
Weak Structural Balance Property: There is no set of three nodes such that the edges among them consist of exactly two positive edges and one negative edge

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 …

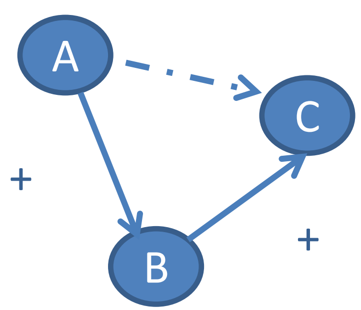
A Weaker Form of Structural Balance



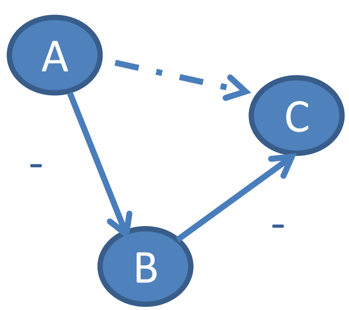
Trust, distrust and directed graphs

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

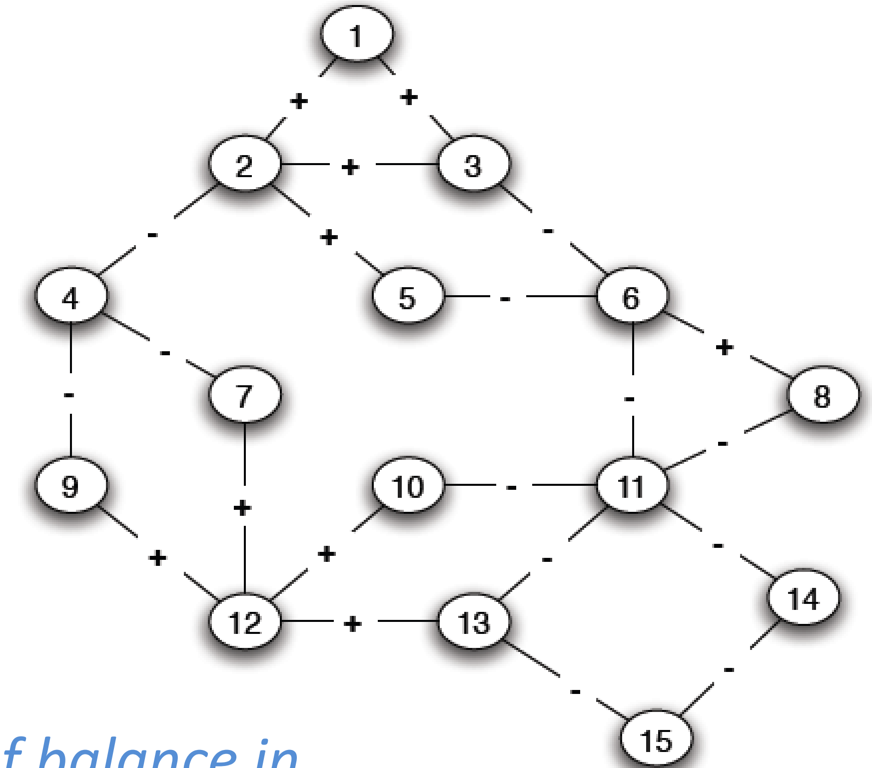
2. Instead of all triangles, “most” triangles, approximately divide the graph

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

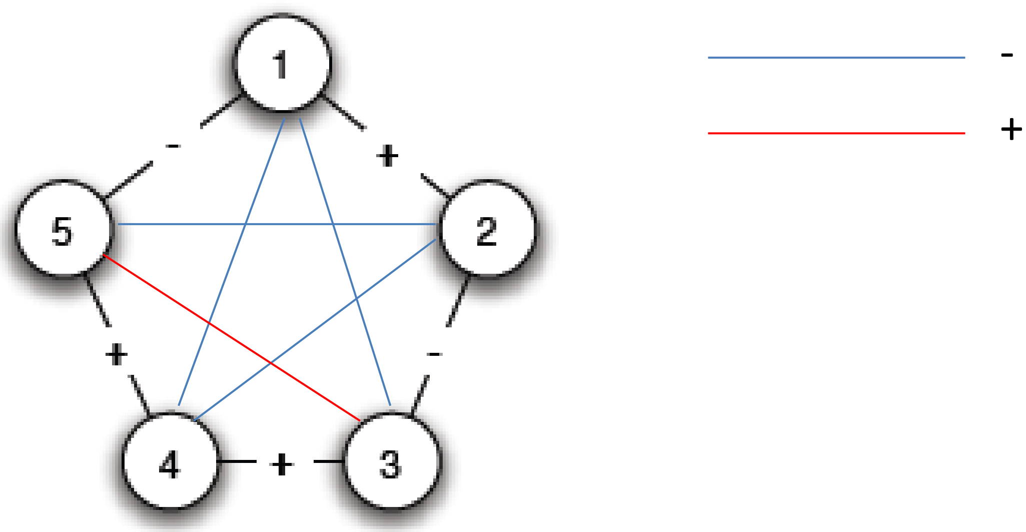


What is a good definition of balance in a non-complete graph?

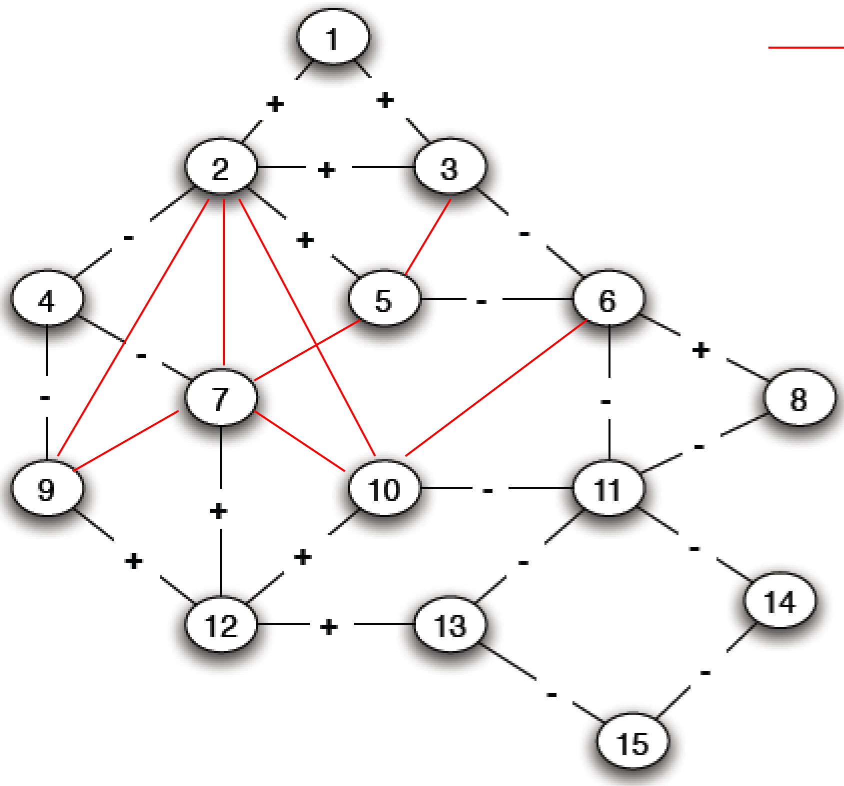
Balance Definition for General Graphs

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



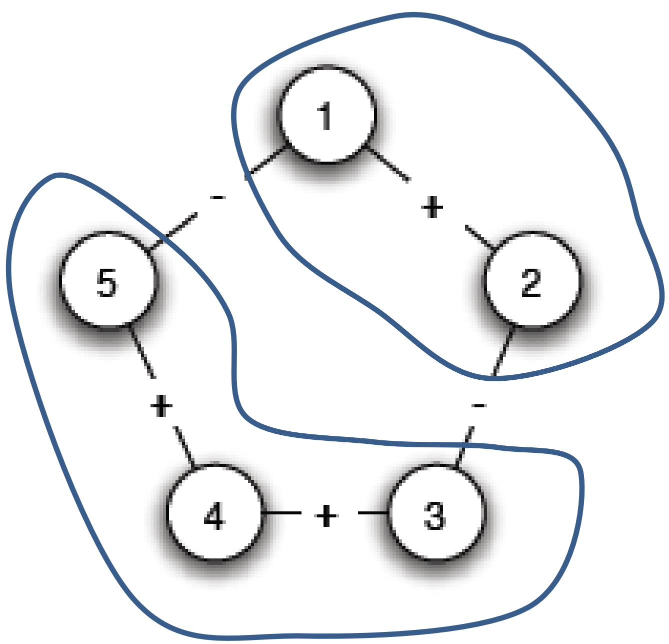
Balance Definition for General Graphs



Balance Definition for General Graphs

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

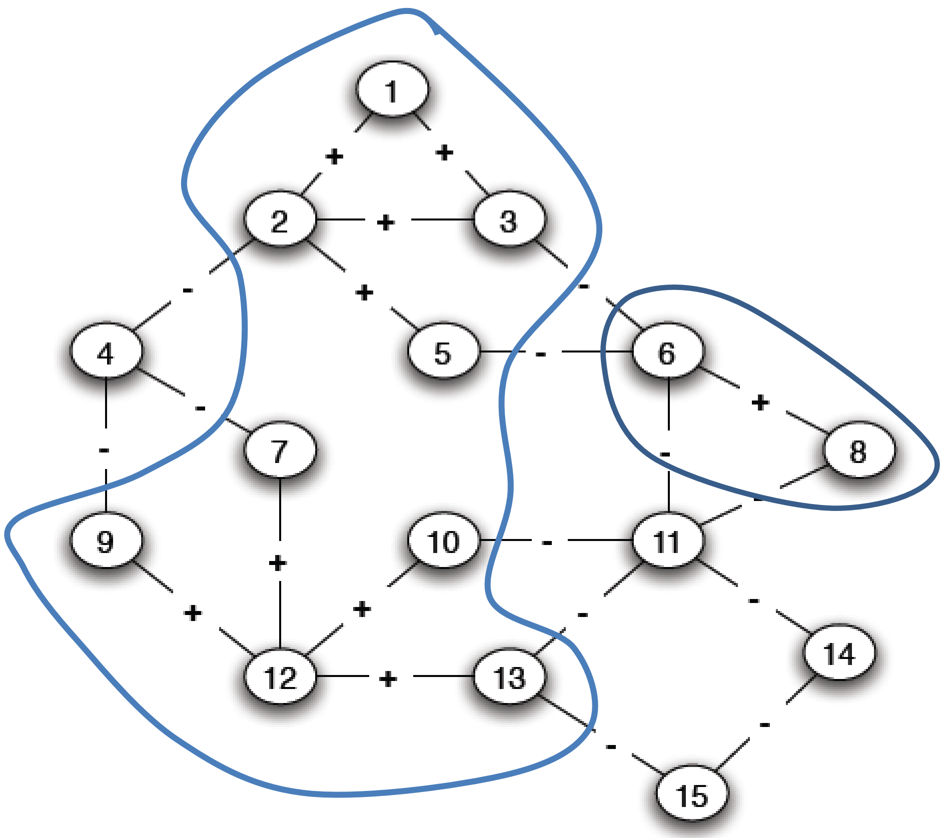
A (non-complete) graph is balanced if it possible to divide the nodes into two sets X and Y, such that any edge with both ends inside X or both ends inside Y is positive and any edge with one end in X and one end in Y is negative



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

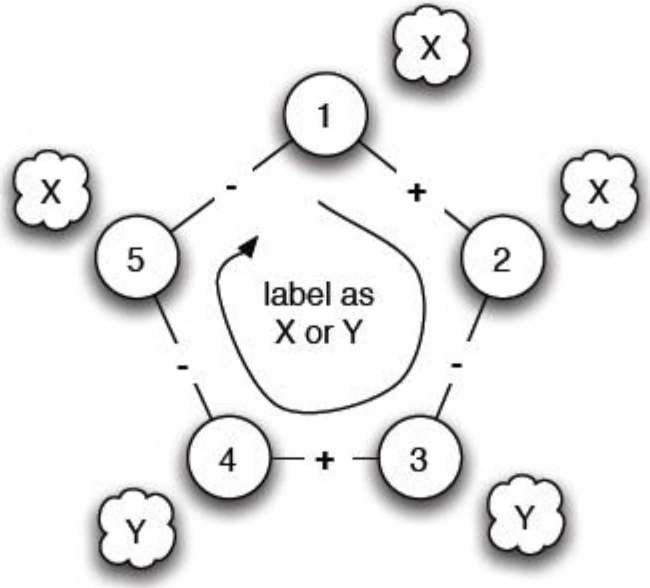
Balance Definition for General Graphs

Algorithm for dividing the nodes?



Balance Characterization

What prevents a network from being balanced?



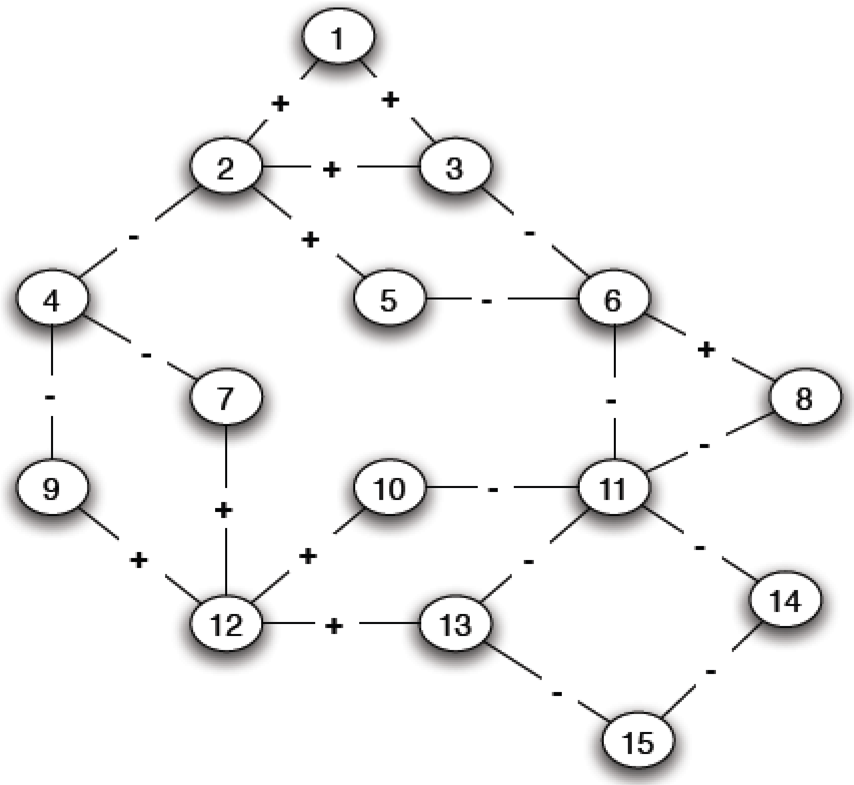
- Start from a node and place nodes in X or Y
- Every time we cross a negative edge, change the set

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

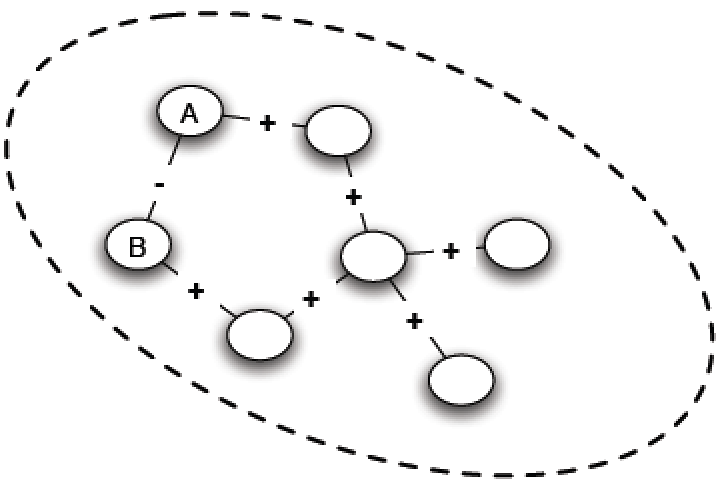
Balance Characterization: Step 1

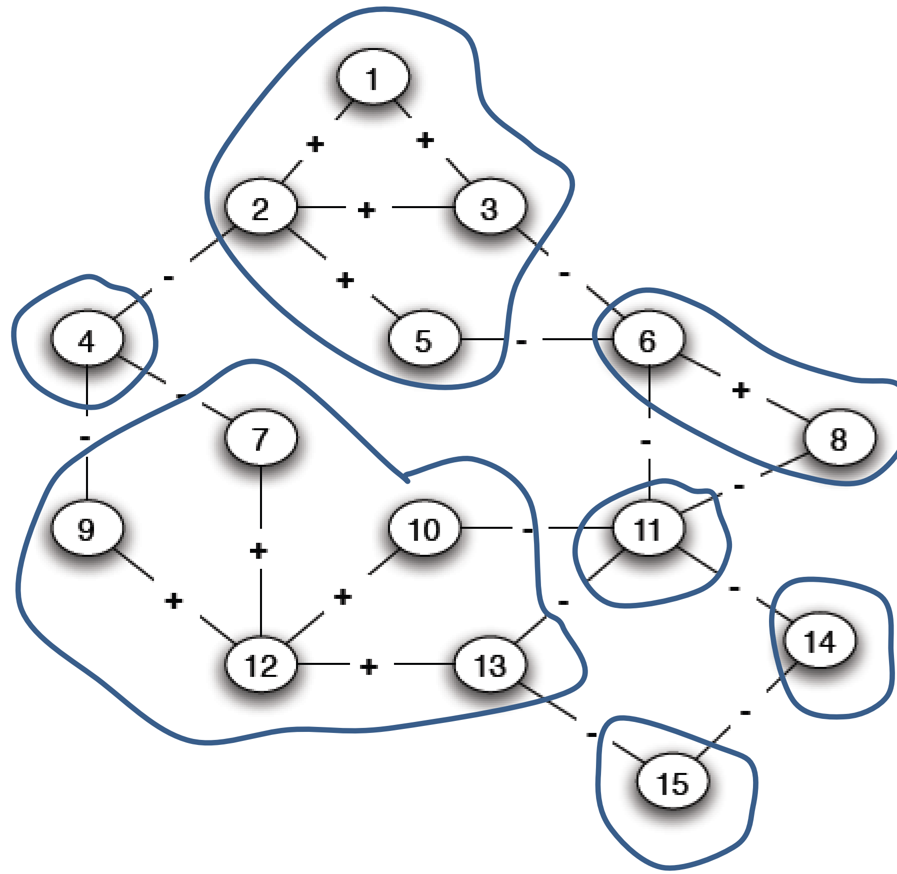
a. Find *connected components* (supernodes) by considering only positive edges

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

(a) Yes -> odd cycle (1)

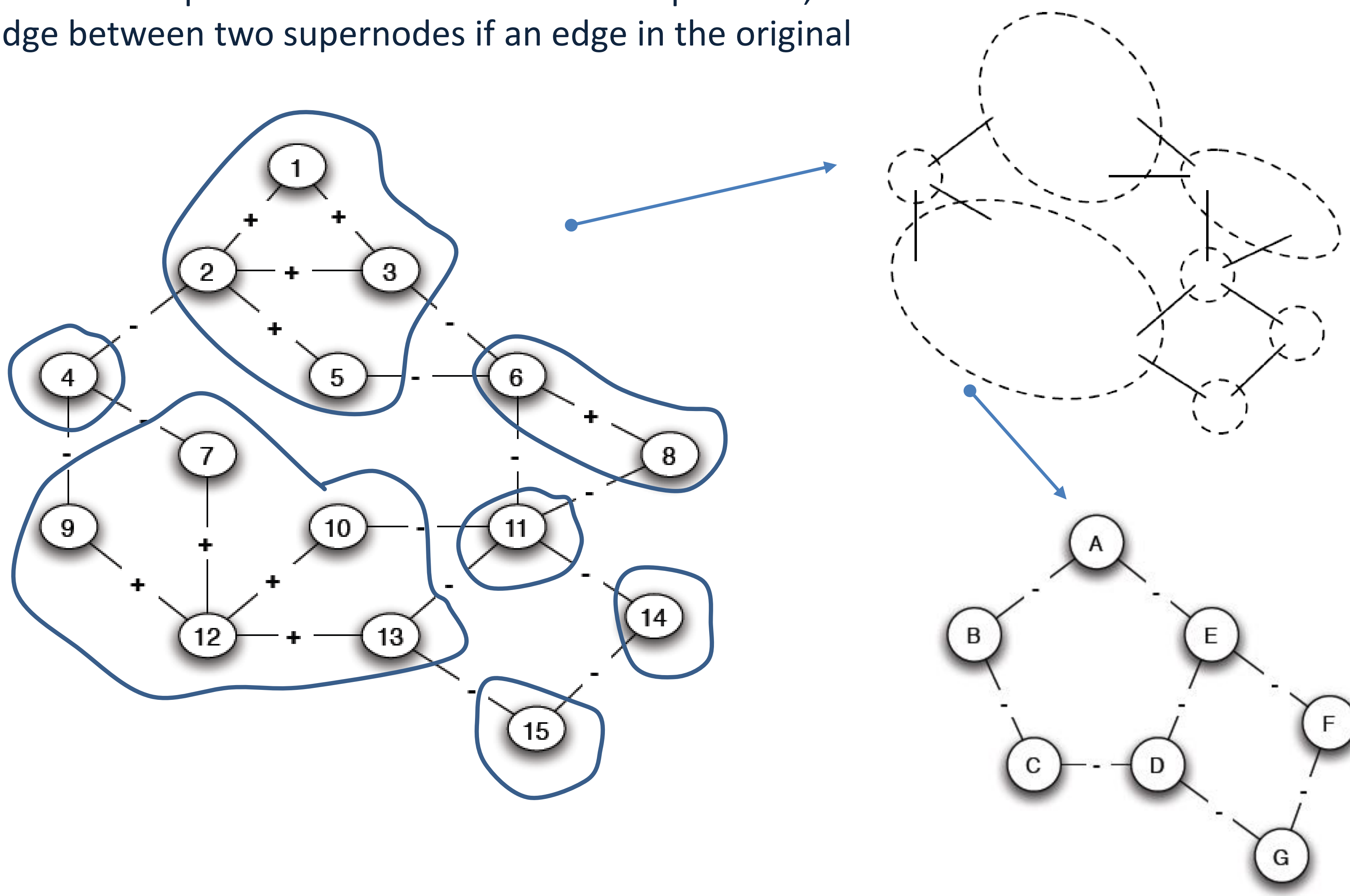
(b) No -> each supernode either X or Y





Balance Characterization: Step 1

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



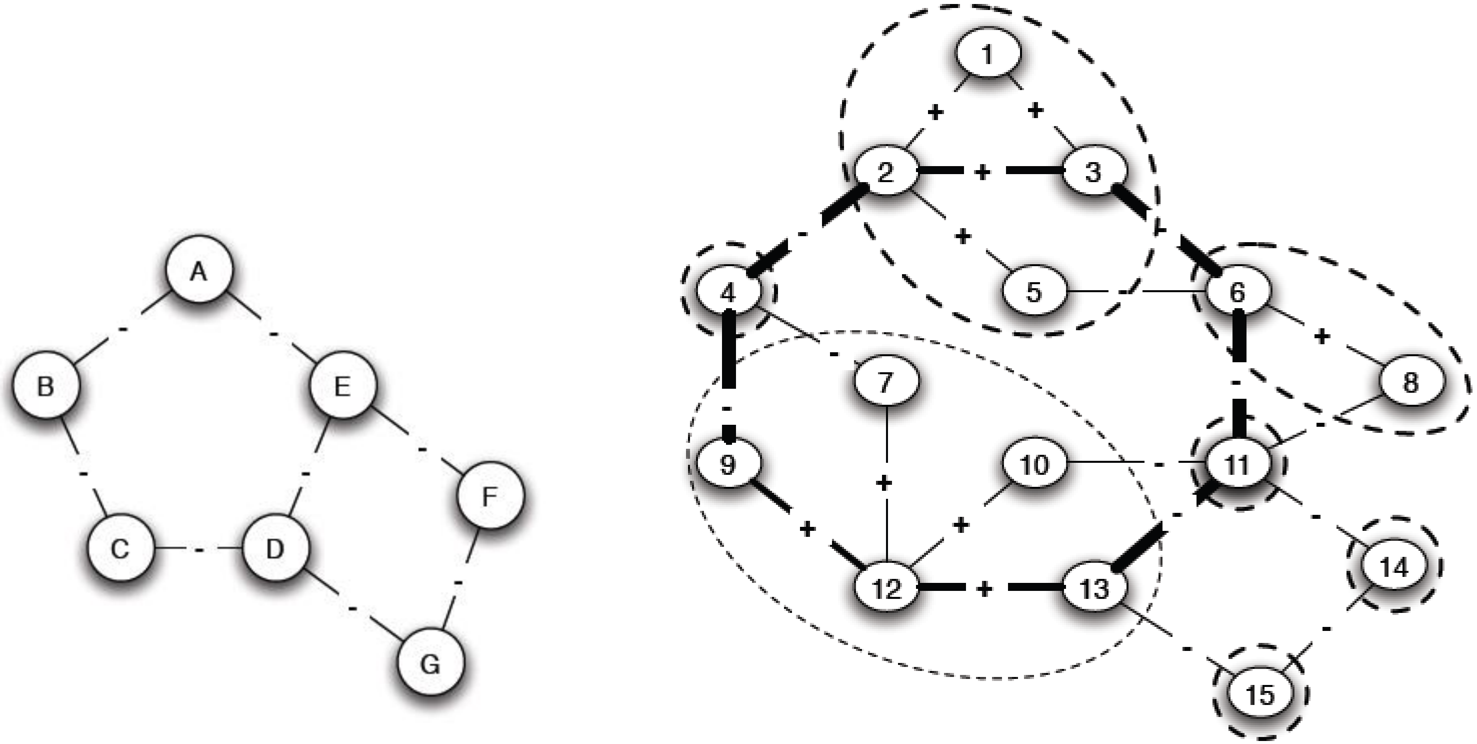
Balance Characterization: Step 2

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*)

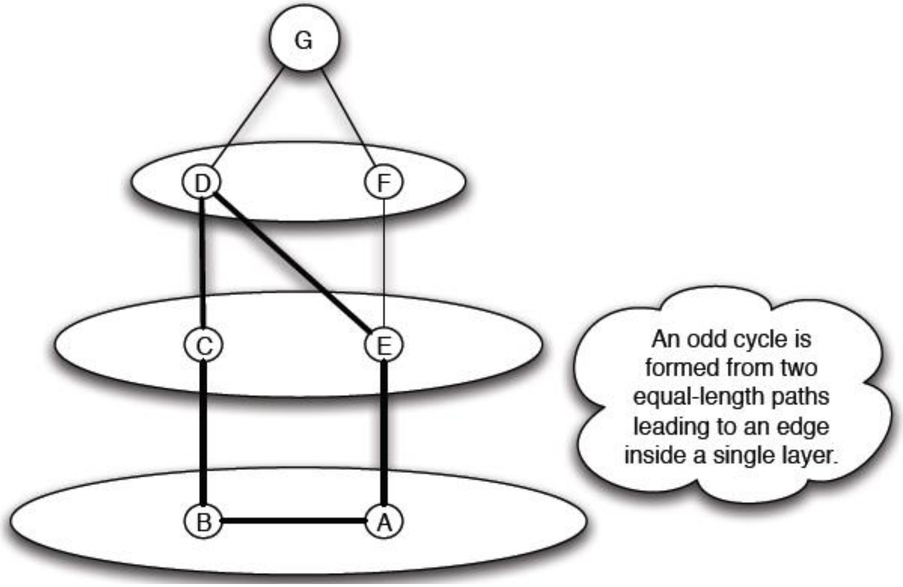
Use Breadth-First-Search (BFS)

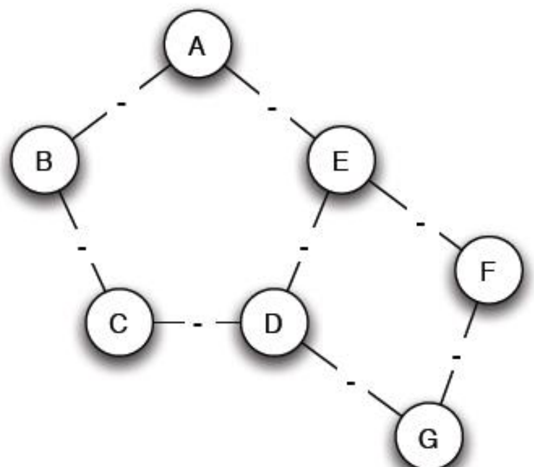
Two type of edges: (1) between nodes in adjacent levels (2) between nodes in the same level

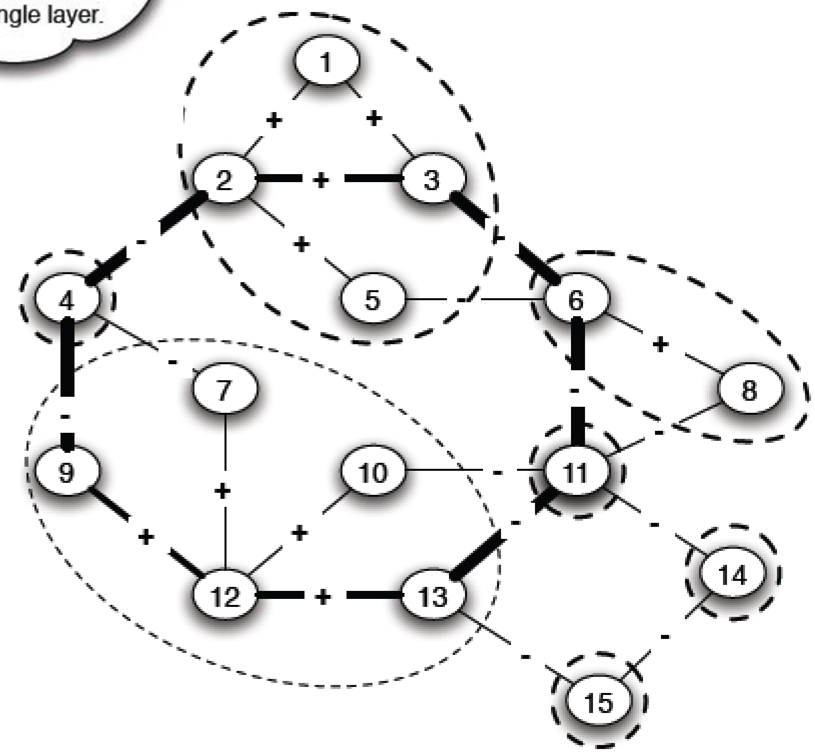
If only type (1), alternate X and Y labels at each level

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 $\boxed{\text{all}}$ triangles in a labeled complete graph are balanced, than either

(a) $\boxed{\text{all}}$ pairs of nodes are friends or,

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

- (i) $\boxed{\text{every}}$ pair of nodes in X like each other,
- (ii) $\boxed{\text{every}}$ pair of nodes in Y like each other, and
- (iii) $\boxed{\text{every}}$ 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* $1-\delta$ of the nodes in which *at least* $1-\delta$ of all pairs are friends, or,

(b) the nodes can be divided into two groups X and Y, such that
   - (i) *at least* $1-\delta$ of the pairs in X like each other,
   - (ii) *at least* $1-\delta$ of the pairs in Y like each other, and
   - (iii) *at least* $1-\delta$ of the pairs with one end in X and one in Y are enemies

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

References

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

S. Sintos, P. Tsaparas, Using Strong Triadic Closure to Characterize Ties in Social Networks. ACM International Conference on Knowledge Discovery and Data Mining (KDD), August 2014