

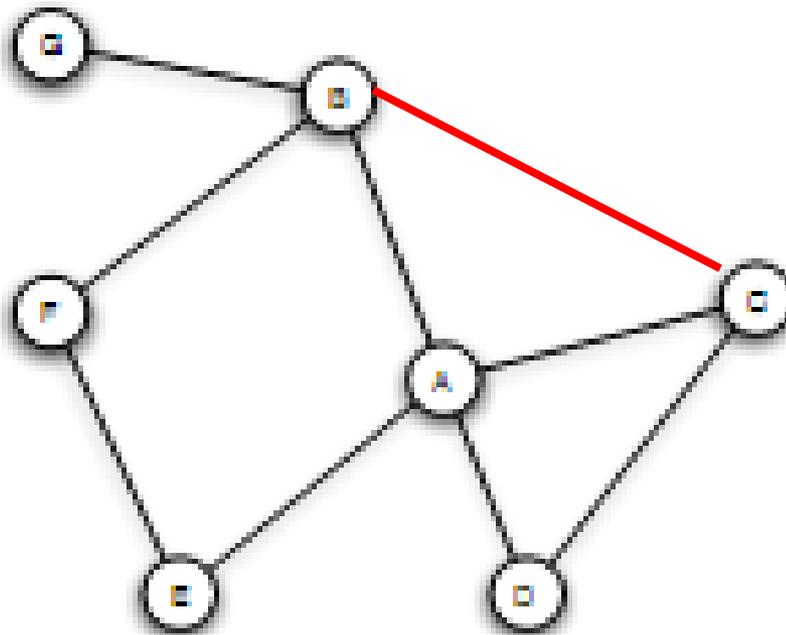
Online Social Networks and Media

Strong and Weak Ties

STRONG AND WEAK TIES

Triadic Closure

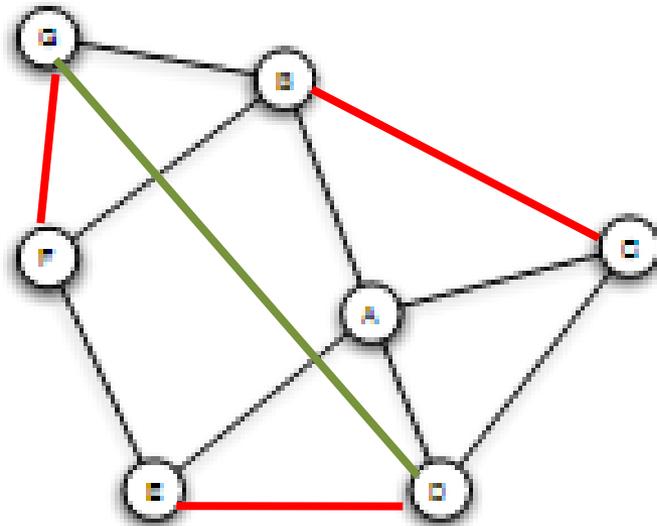
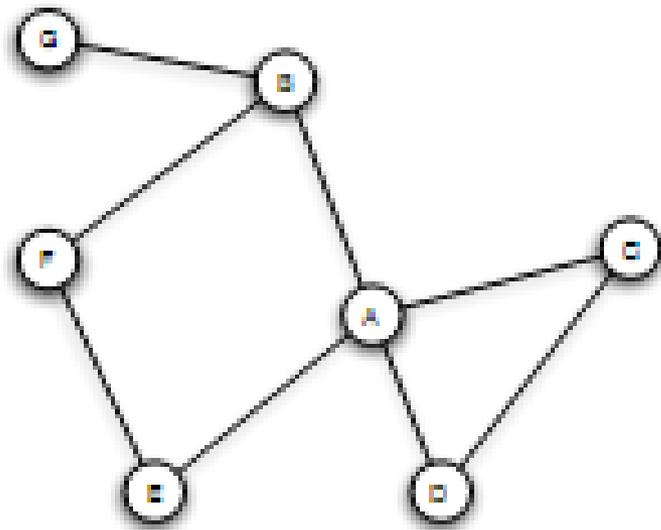
If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future



Triangle

Triadic Closure

Snapshots over time:



Clustering Coefficient

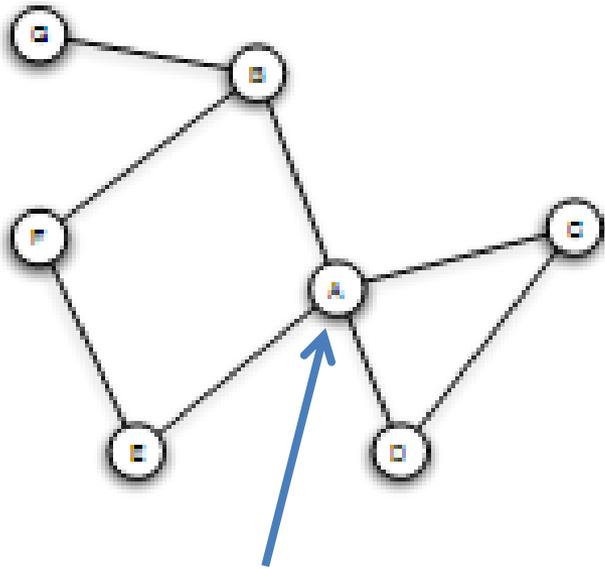
(Local) clustering coefficient for a node is the probability that two randomly selected friends of a node are friends with each other (*form a triangle*)

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

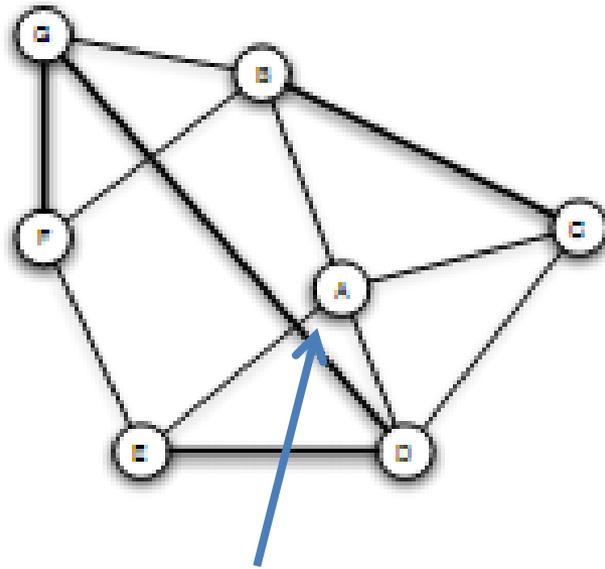
Fraction of the friends of a node that are friends with each other (i.e., connected)

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

Clustering Coefficient



$1/6$

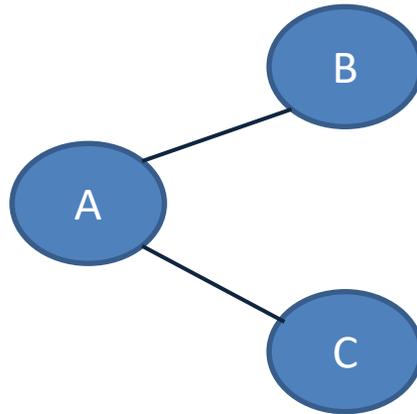


$1/2$

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

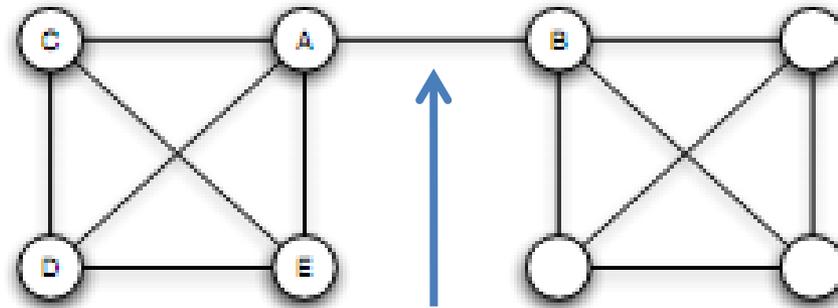
Mark Granovetter, in the late 1960s

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



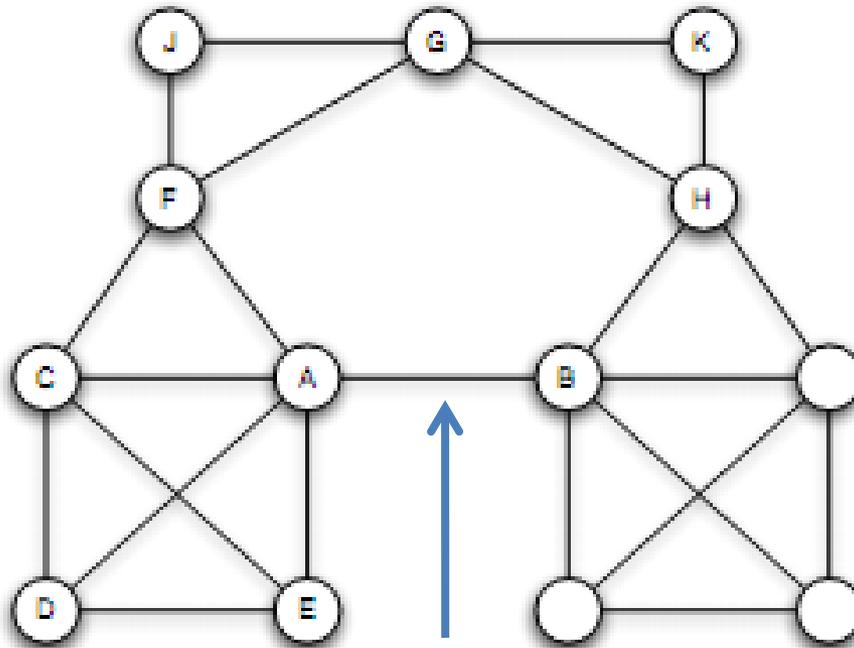
Bridge
(aka cut-edge)

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

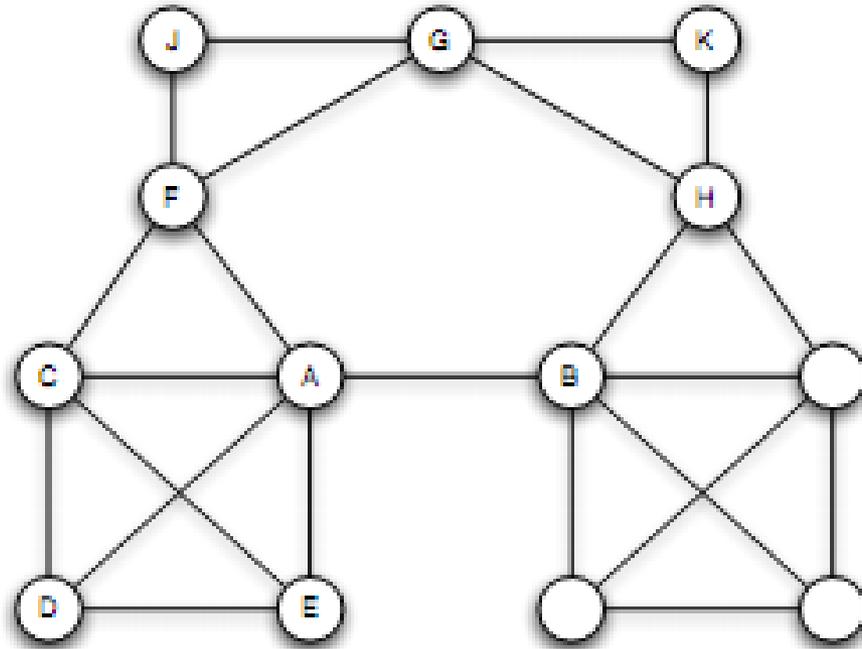


Local Bridge

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

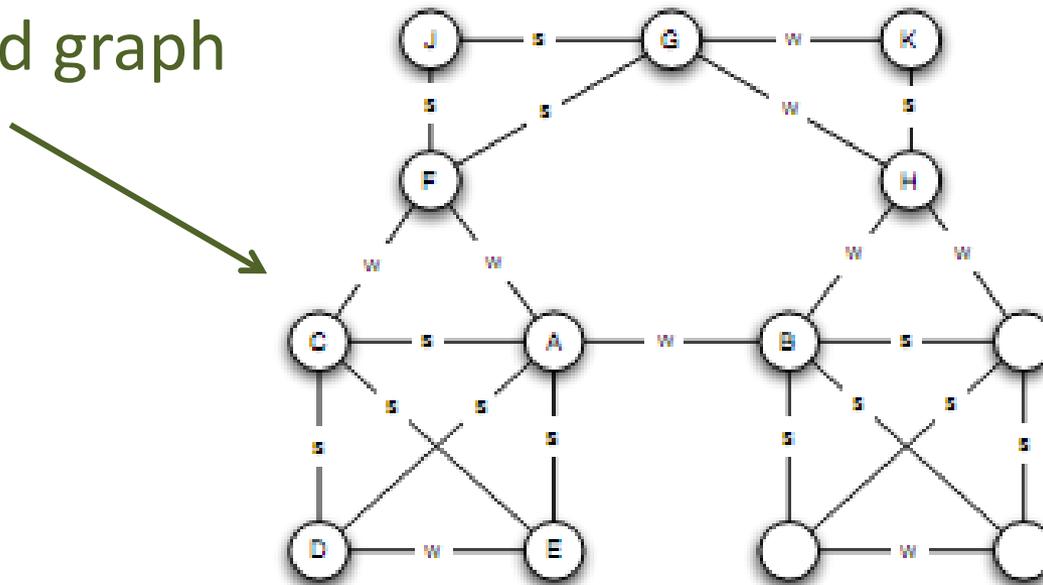


An edge is a local bridge, if and only if, it is not part of any **triangle** in the graph

The Strong Triadic Closure Property

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

Annotated graph

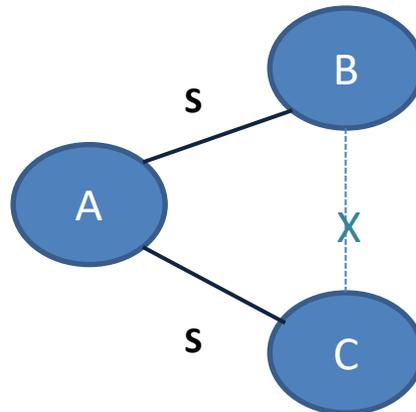


The Strong Triadic Closure Property

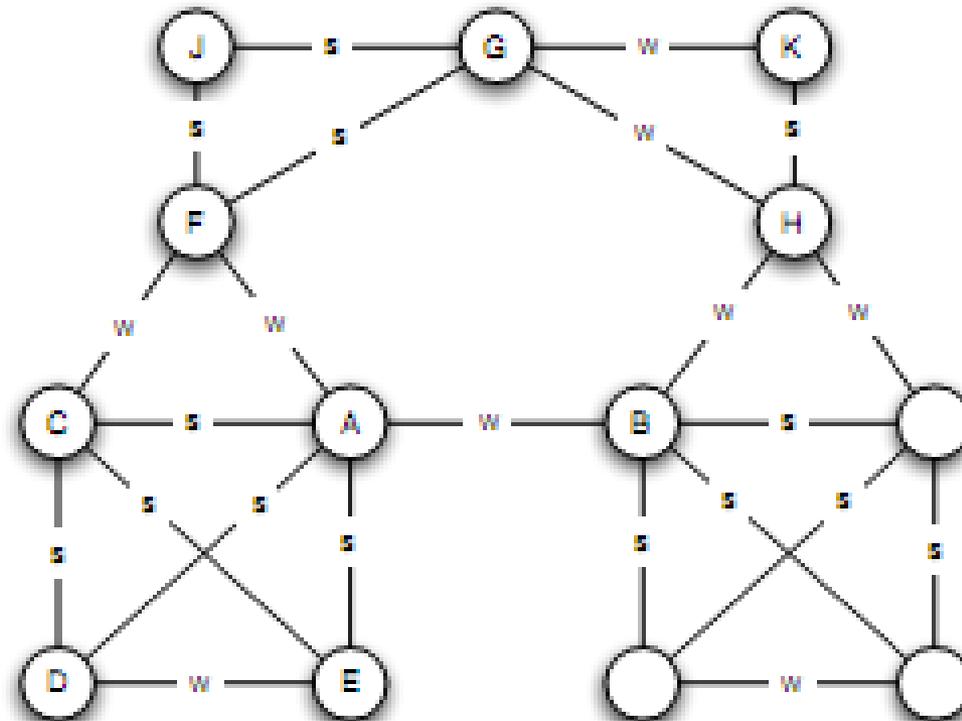
If a node A has edges to nodes B and C, then the B-C edge is especially likely to form if both A-B and A-C are strong ties

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?

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

Is it a “social network”?

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

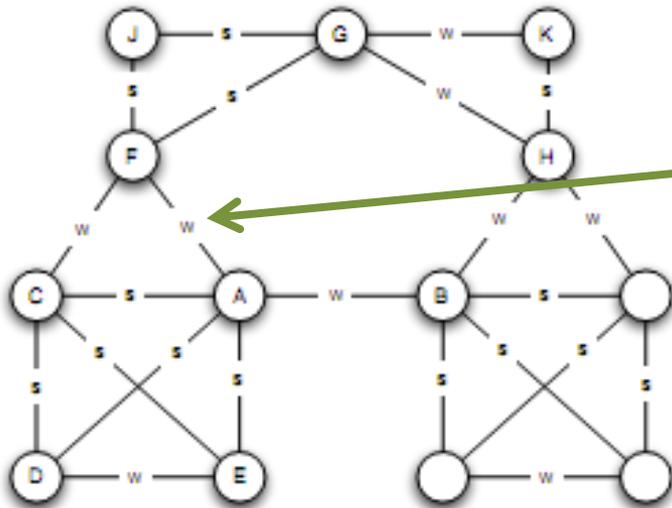
“almost” local bridges

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

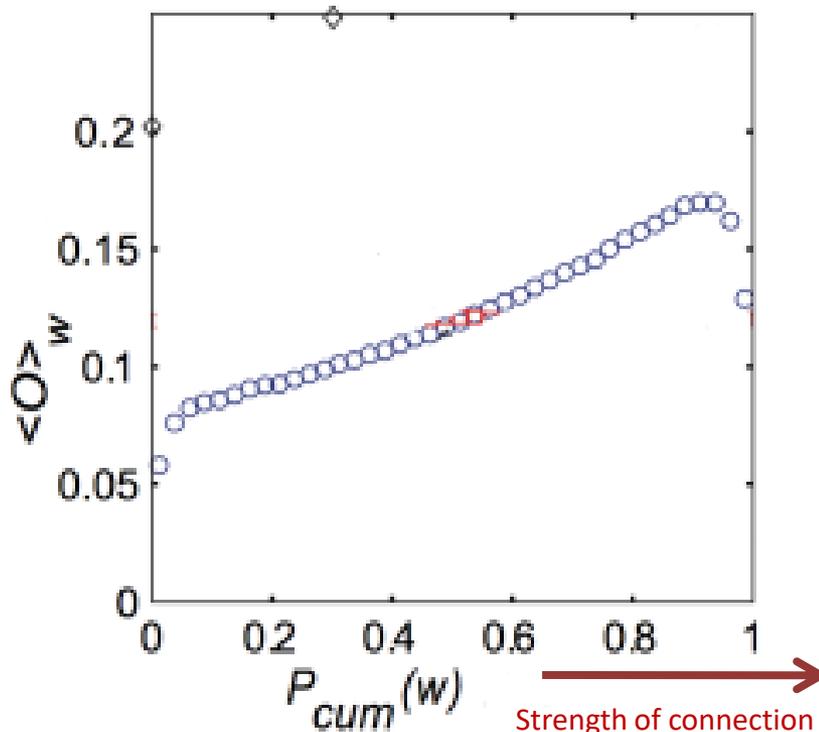
1/6

When is this value 0?

Generalizing Weak Ties and Local Bridges: 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

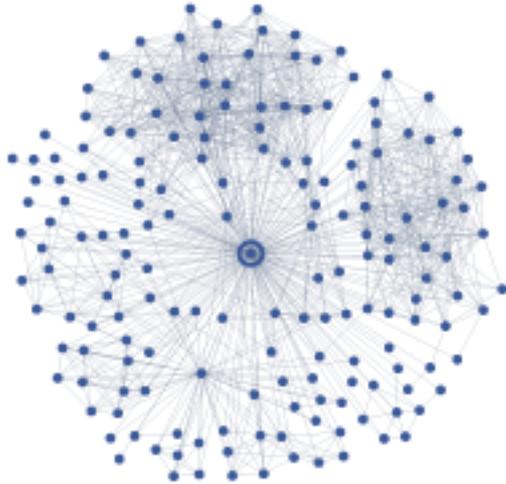
At what extent each link was used for social interactions

Three (not exclusive) kinds of ties (links)

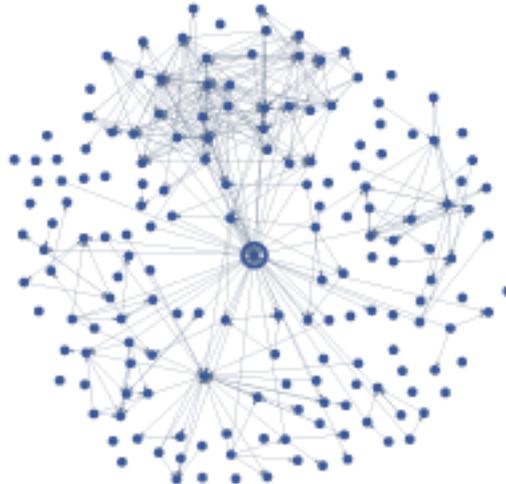
1. **Reciprocal (mutual) communication**: both send and received messages to friends at the other end of the link
2. **One-way communication**: the user send one or more message to the friend at the other end of the link
3. **Maintained relationship**: the user followed information about the friend at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

Tie Strength on Facebook

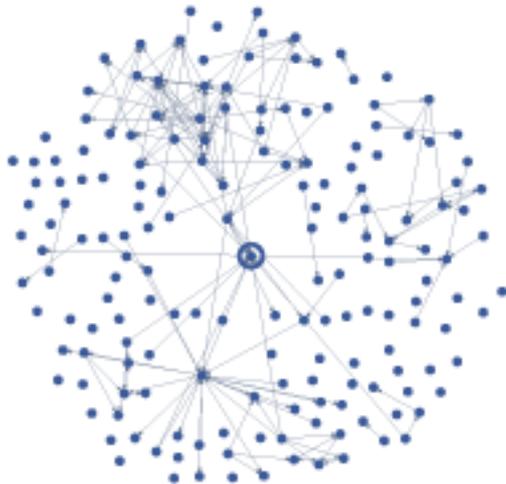
All Friends



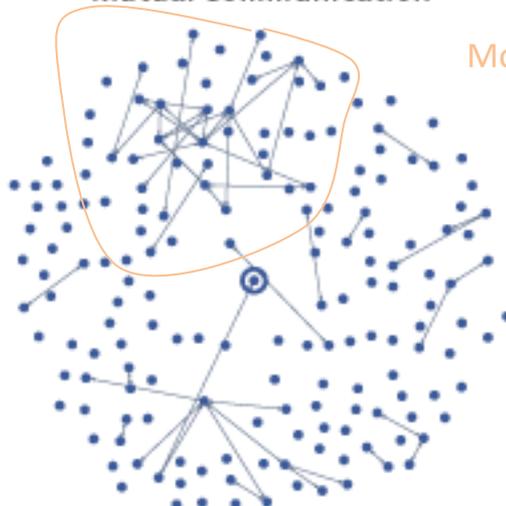
Maintained Relationships



One-way Communication

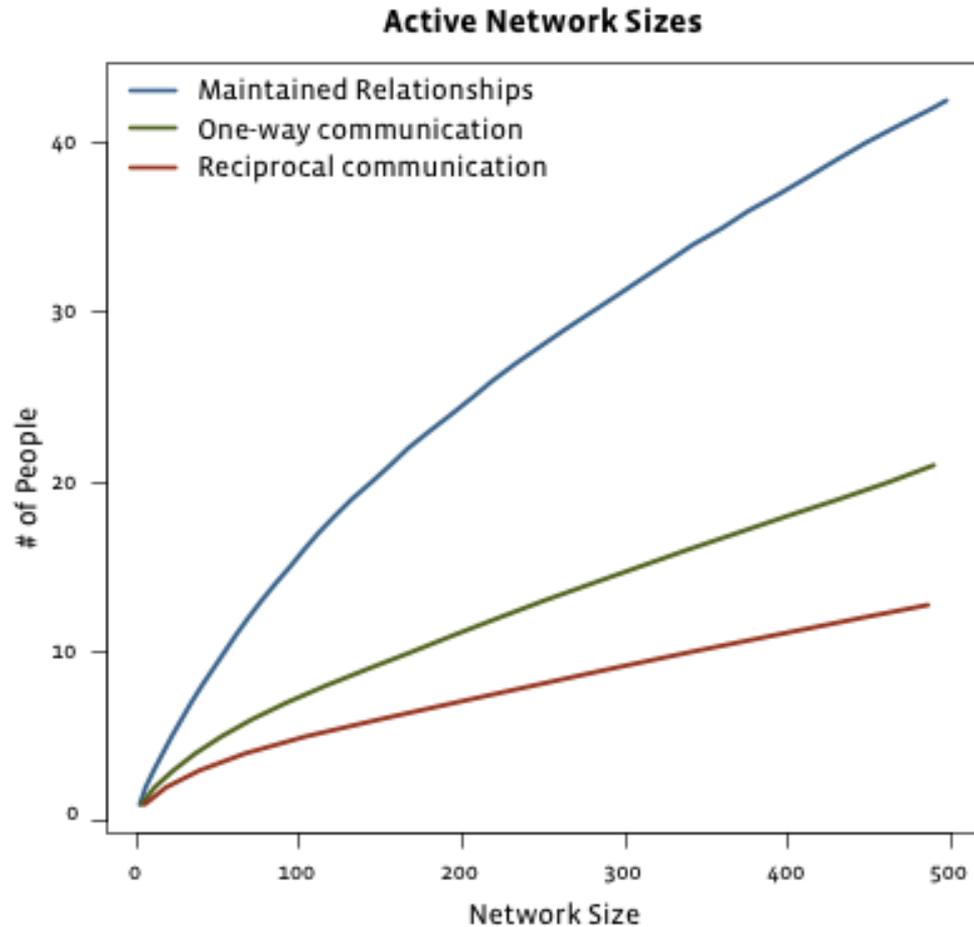


Mutual Communication



More recent connections

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

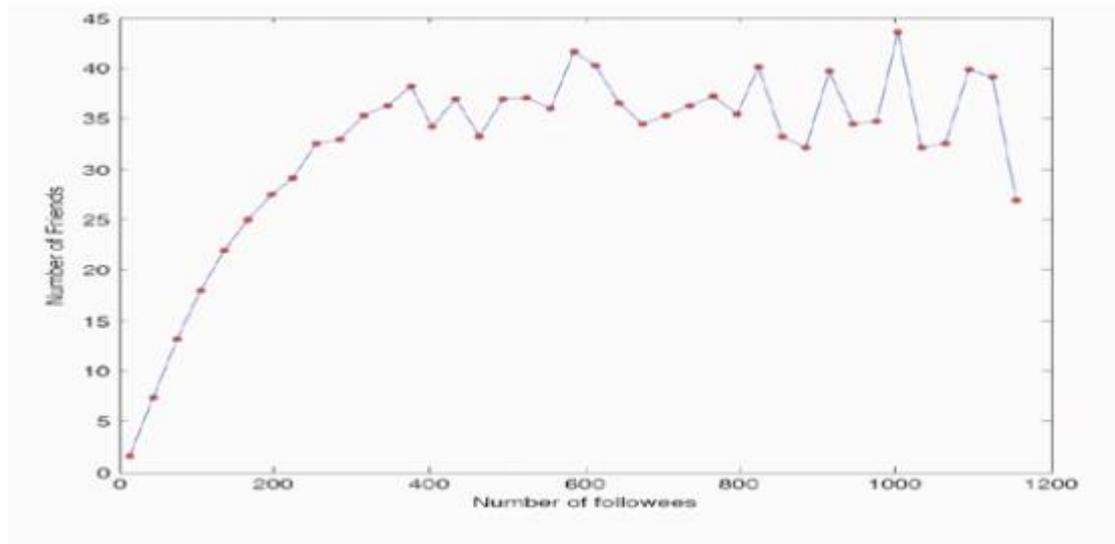
Passive engagement (keep up with friends by reading about them even in the absence of communication)

Tie Strength on Twitter

Huberman, Romero and Wu, 2009

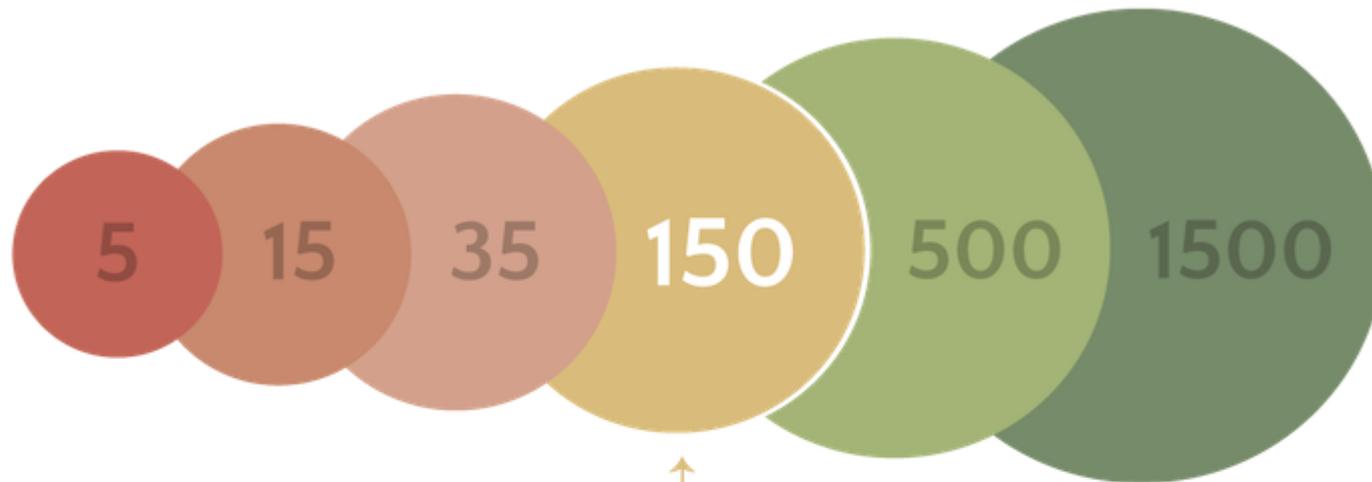
Two kinds of links

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



Dunbar's number

- A biologically determined limit on the number of relationships a person can maintain



Dunbar's Number

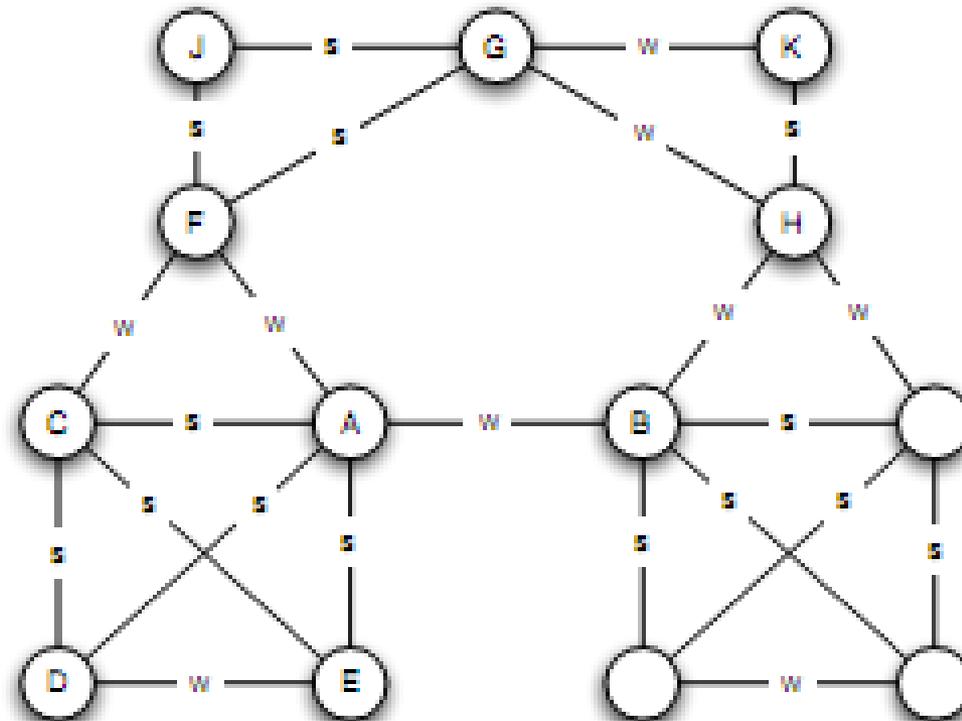
the max number of relationships a person can maintain

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

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.

Approximation Algorithms

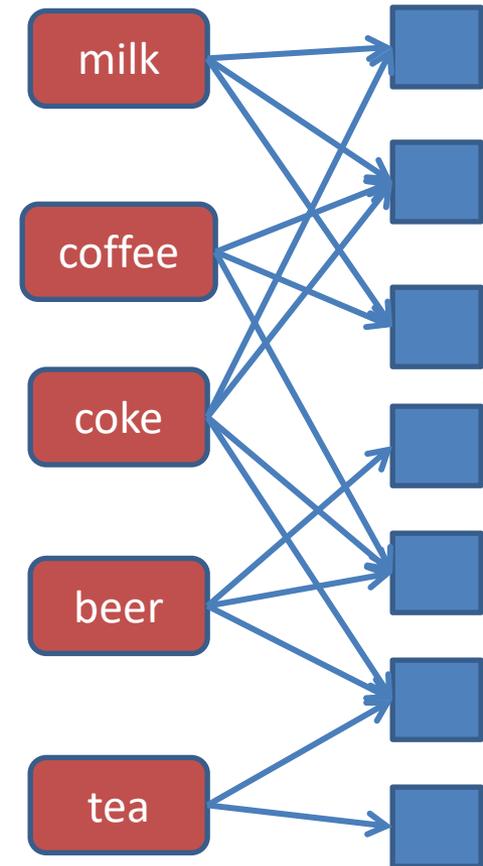
- **Bad News:** **MaxSTC** is hard to approximate.
- **Good News:** There exists a **2-approximation** algorithm for the **MinSTC** problem.
 - The number of weak edges it produces is at most two times those of the optimal solution.
- The algorithm comes by reducing our problem to a **coverage** problem

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 , $S = \{S_1, \dots, S_n\}$, such that $\bigcup_i S_i = U$
 - We want to find the smallest sub-collection $C \subseteq S$ of S , such that $\bigcup_{S_i \in C} S_i = U$
 - The sets in C cover the elements of U

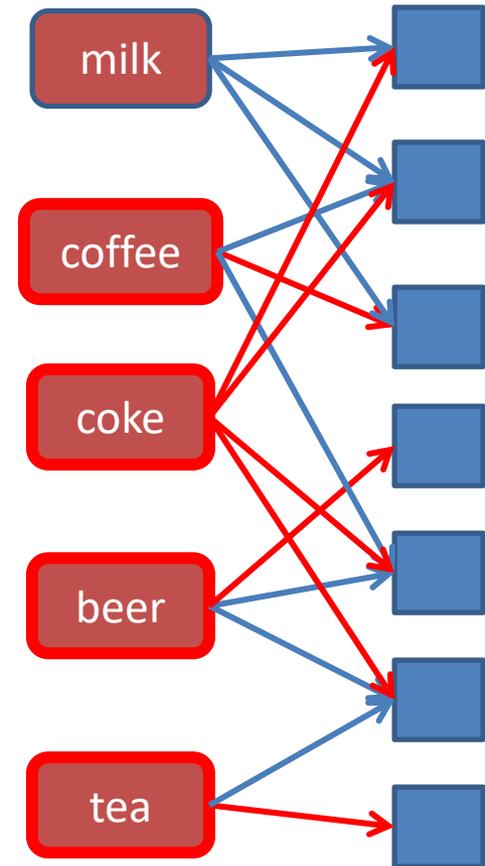
Example

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



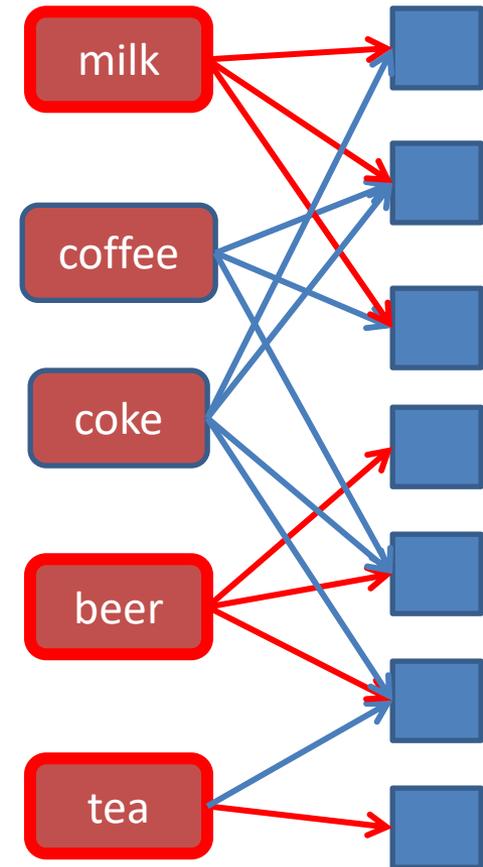
Example

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



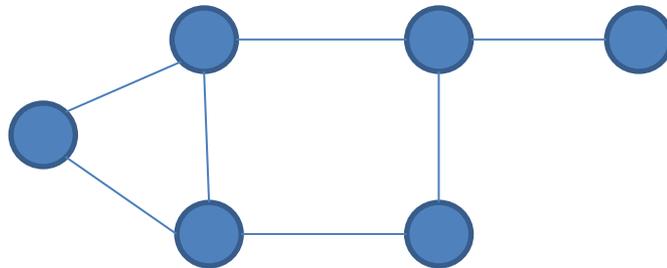
Example

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



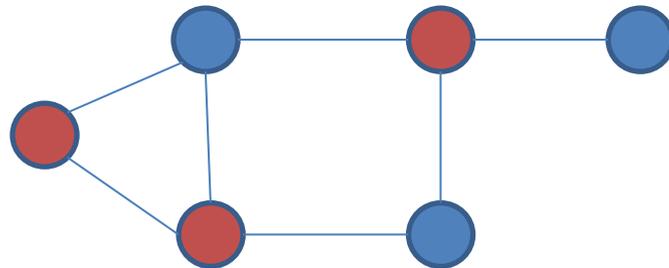
Vertex Cover

- 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 the elements are edges and a set is defined for each node, as the set of edges incident on a node.
 - Each element is covered by exactly two sets



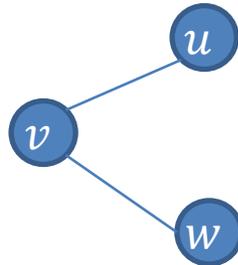
Vertex Cover

- 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 the elements are edges and a set is defined for each node, as the set of edges incident on a node.
 - Each element is covered by exactly two sets



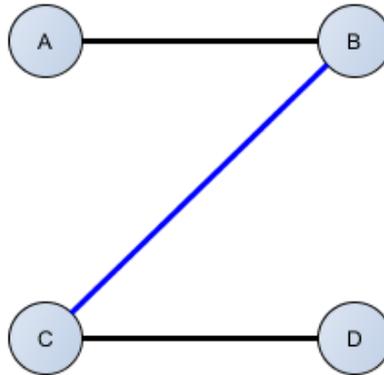
MinSTC and Coverage

- 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 open triangle

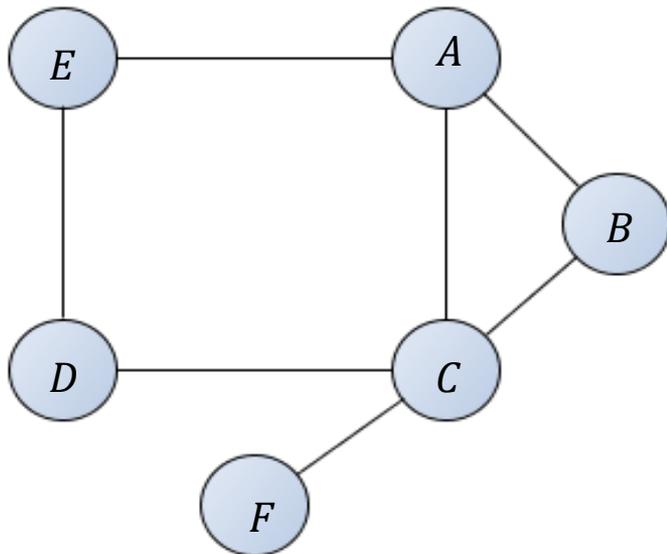


- **MinSTC** can be mapped to the **Minimum Vertex Cover** problem.

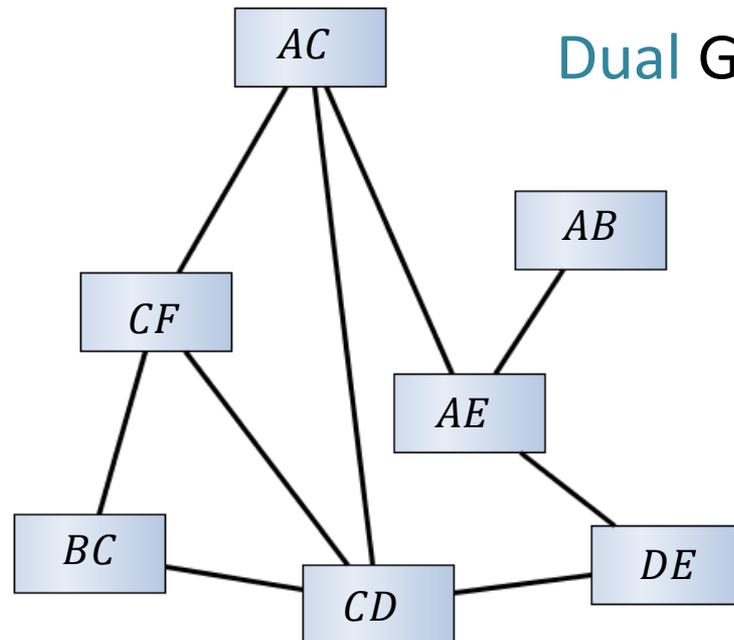
Dual Graph (Gallai graph)

- 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

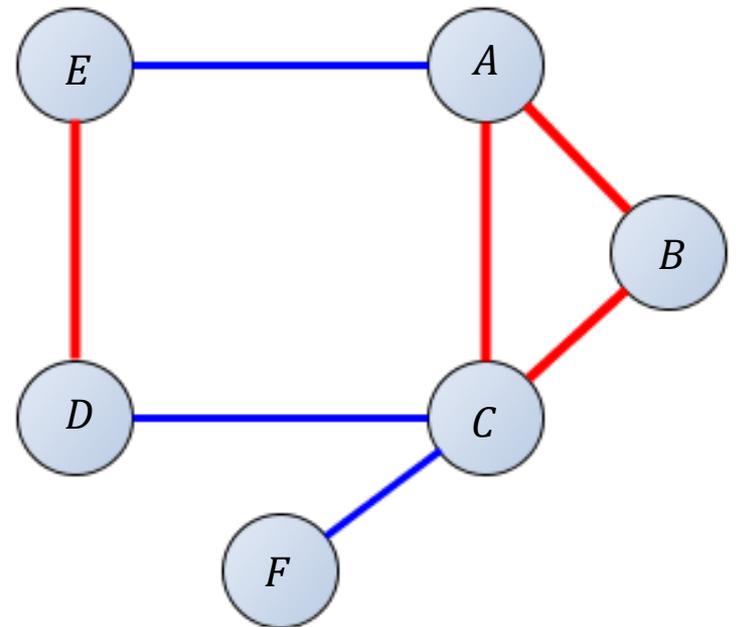
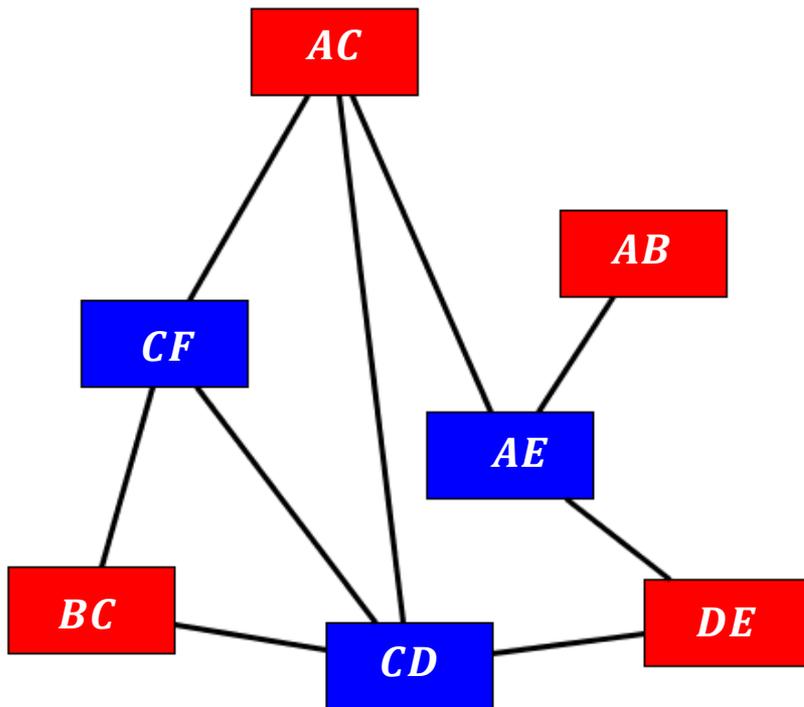


Dual Graph D



Minimum Vertex Cover - MinSTC

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



Complexity

- The **Vertex Cover** problem are **NP-complete**
- 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 **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 $\epsilon \rightarrow 0$, as $n \rightarrow \infty$ (e.g., $\epsilon = \frac{1}{n}$)
 - Good case: $\alpha = O(1)$ is a constant (e.g., $\alpha = 2$)
 - OK case: $\alpha = O(\log n)$
 - Bad case $\alpha = O(n^\epsilon)$

Approximation Algorithms

Approximation algorithms for the **Minimum Vertex Cover** problem:

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**

Greedy Algorithm

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

Approximation Factor: **$\log n$**

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

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 **common activity** 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 significant**

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 significant**

The Strength of Weak Ties

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

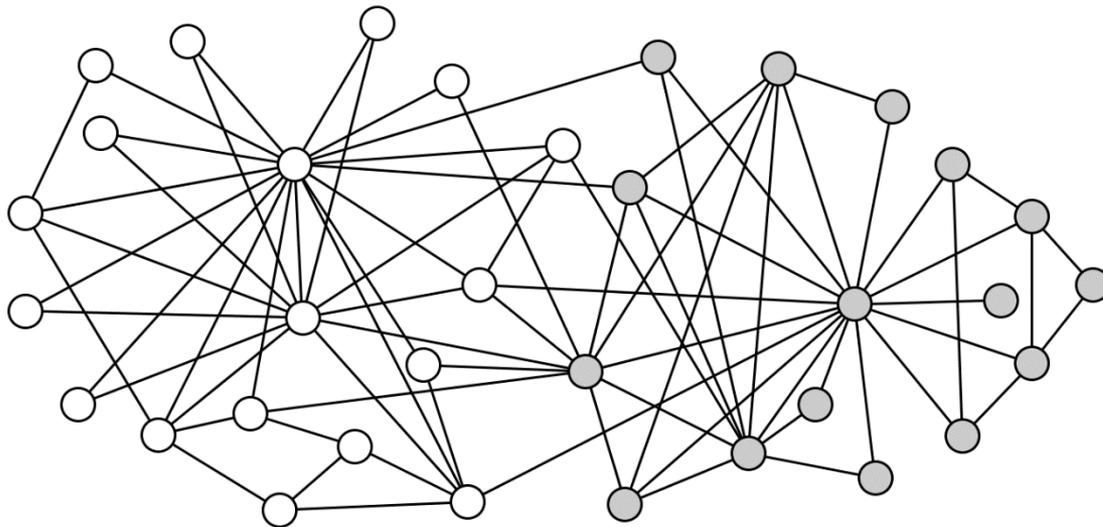
Datasets with known communities

- Amazon Books

- US Politics books : liberal, conservative, neutral.

- Karate Club

- Two fractions within the members of the club.

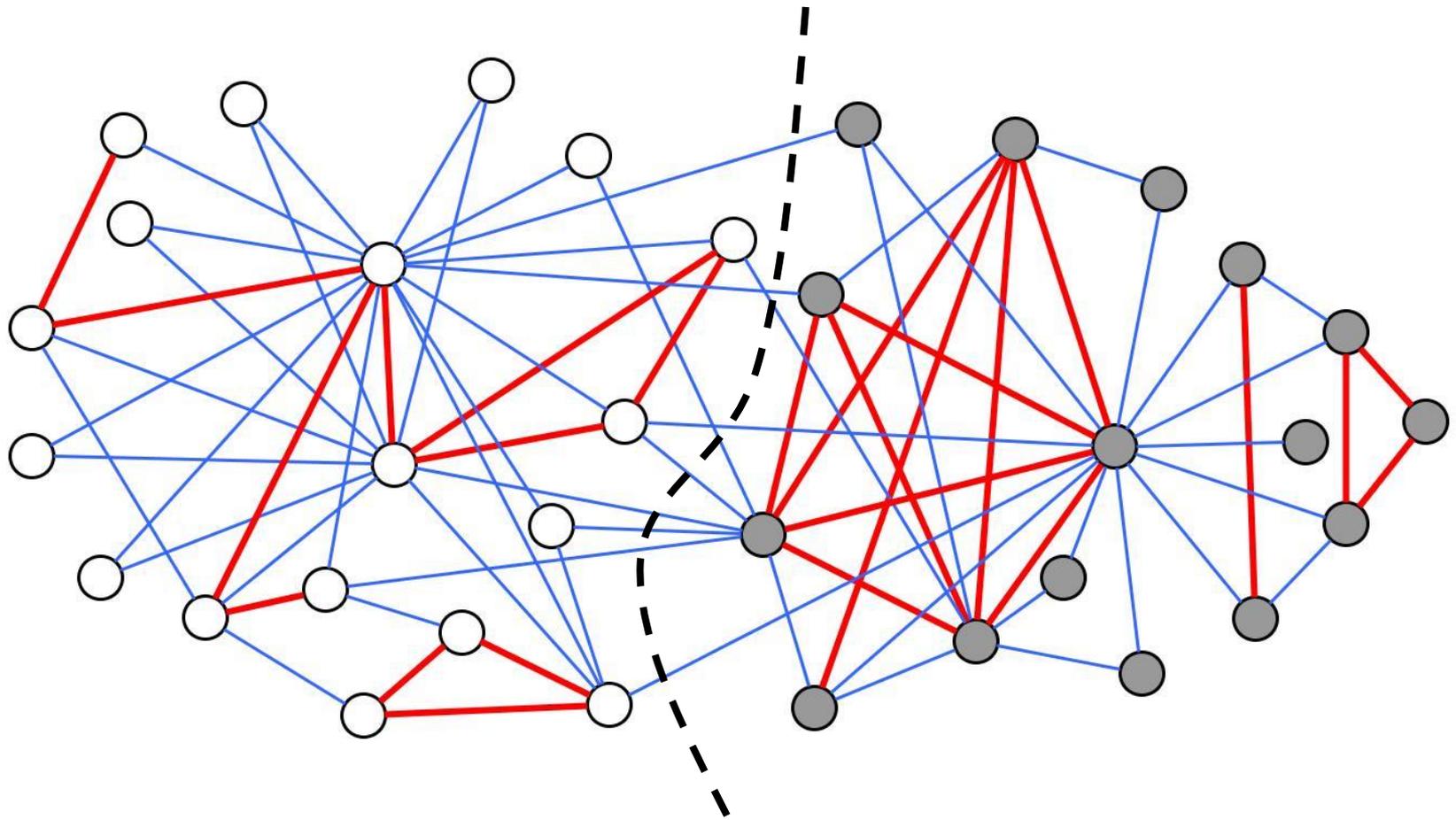


Weak Edges as Bridges

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

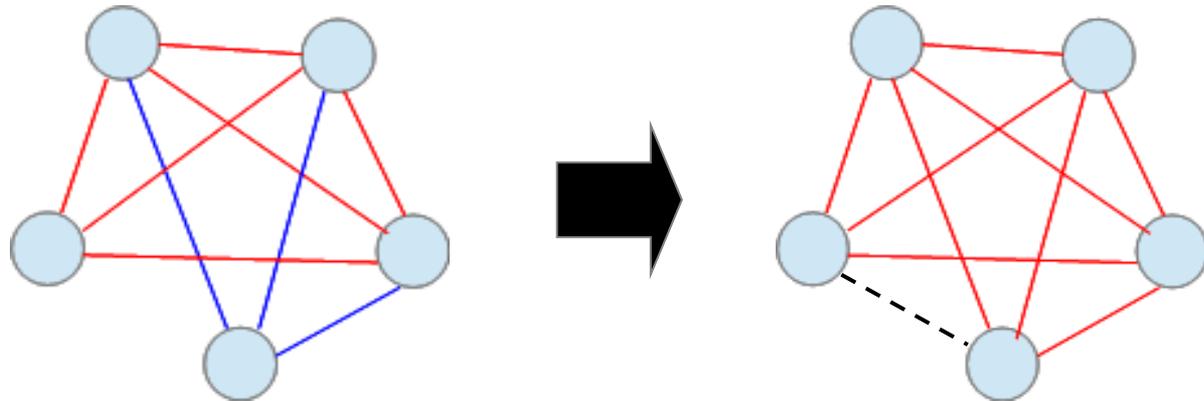
	P_S	R_W
Karate Club	1	1
Amazon Books	0.81	0.69

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**

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