Corso di laurea magistrale in Ingegneria Informatica

Social Networks Analysis
A.A. 2021/22

Local Behaviors and global Phenomena

Local Behaviours and Global Phenomena

- Networks play a powerful role as a bridge between the local and the global
  - offer explanations of how simple processes at the level of individual nodes and links can have complex cumulative effects that ripple through a population as a whole

- In this lection, we consider some fundamental social network issues
  - how information flows through a social network?
  - how different nodes can play structurally distinct roles in this process?
  - How can we recognize them?
  - how these structural considerations shape the evolution of the network itself over time?

GRANOVETTER'S EXPERIMENT (1970)

- As part of his Ph.D. thesis research, Mark Granovetter interviewed people who had recently changed employers to learn how they discovered their new jobs
  - In most of the cases through personal contacts
  - More often described as acquaintances rather than close friends

- Striking conclusion
  - Friends should be more motivated in helping you when you're between two jobs

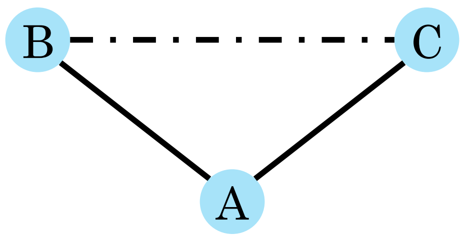
- why are distant acquaintances more helpful than your friends in obtaining crucial information leading to your new job?

Granovetter's Hypothesis

- There are two different kinds of friendships
- At a local level
  - Friendship can be strong or weak
- At a global level
  - Friendship can tie nodes that are close or far in the network
- Granovetter's hypothesis (Strength of Weak Ties)
  - Information travels over links that tie distinct parts of the network
  - These links tend to represent weak ties

Triadic Closure

- **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



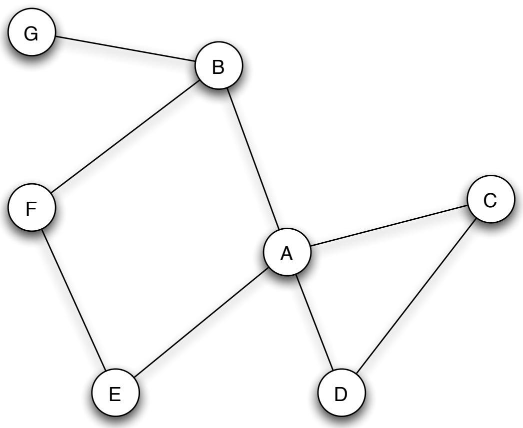
B and C tend to become friends since they have a common friend A

- Sociological motivations
  - More opportunities to meet and spend time together
  - They tend to trust each other on the basis of the common friend
  - the common friend A has incentives to bring B and C together so that she has not to split among the two

Clustering Coefficient

- The basic role of triadic closure in social networks has motivated the formulation of simple social network measures to capture its prevalence

- The **clustering coefficient** of a node A is defined as the probability that two randomly selected friends of A are friends with each other
  - i.e., the fraction of pairs of A's friends that are connected to each other by edges

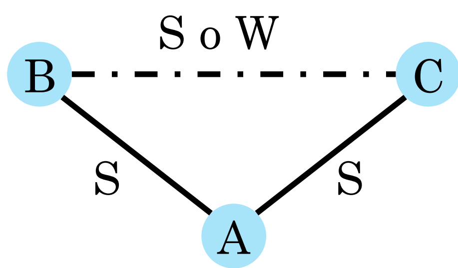


Clustering coefficient of A = 1/6
Clustering coefficient of D = 1

- the more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be

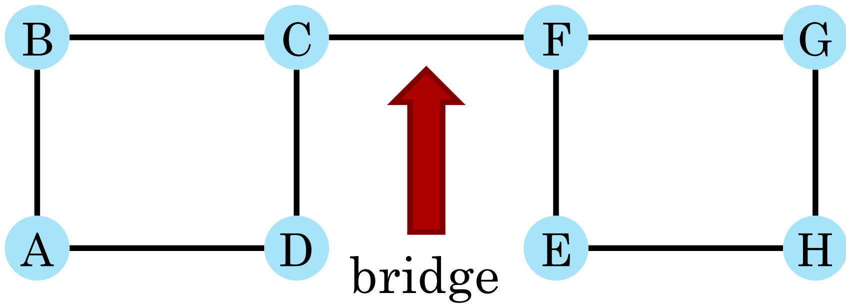
STRENGTH OF A TIE

- According to Granovetter, the interpersonal ties can be of two types
  - Strong = friendship
  - Weak = acquaintanceship
  - Local property
- Strong Triadic Closure
  - Granovetter operating hypothesis
  - If A has strong ties with B and C then with high probability there is a tie between B and C
    - Can be both strong and weak



BRIDGE

- A bridge is the unique link connecting two distinct components of the network
  - Removing the link the two components are disconnected
  - Global property of the network



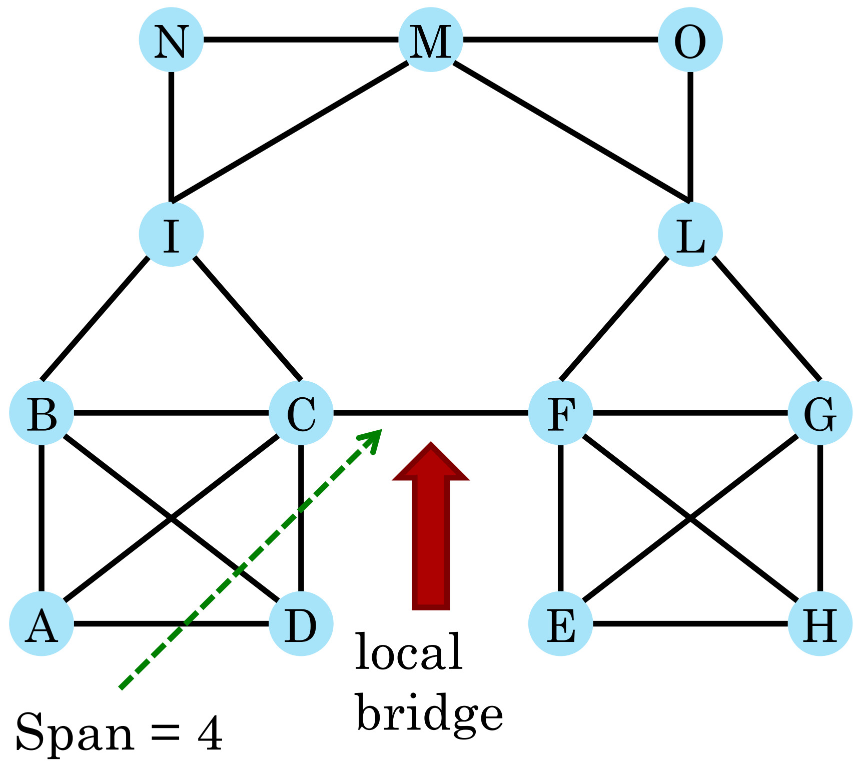
- A bridge does not form a side of any triangle in the graph
  - In a social network bridges are rare

LOCAL BRIDGE

- A local bridge is a link that reduces the distance between its two endpoints
  - Removing it, the distance between its endpoints strictly increases
  - Plays a role similar to the bridge
  - Global property of the network

The span of a local bridge is the distance its endpoints would be from each other if the edge were deleted

- An edge is a local bridge if it is not part of any triangle in the graph
- The higher the span of the link the more crucial is the link for the coesion of the network

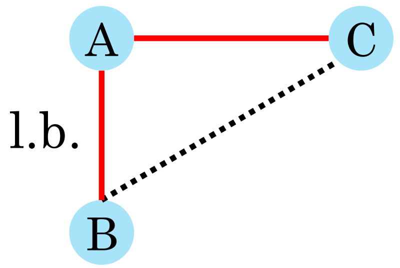


Strength of Weak Ties

Assume Strong Triadic Closure holds

The **Strength of Weak Ties** claims

- If a node A is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie



- Let AB be a local bridge
- Let AB and AC be strong ties
- BC cannot exist since AB cannot be part of a triangle
- BC has to exist by STC – CONTRADICTION!!
- Then, AB cannot be a a strong tie

- Then
  - If the network has a sufficiently high number of strong ties then each local bridge is a weak tie
  - Crucial information travels along weak ties

- This argument connects the local property of tie strength and the global property of serving as a local bridge

Tie Strength and Network Structure in Large-Scale Data

- For many years after Granovetter's initial work these predictions remained relatively untested on large social networks
  - due to the difficulty in finding data that reliably captured the strengths of edges in large-scale, realistic settings.

- Nowadays we have a lot of large-scale "who-talks-to-whom" data-sets
  - Facebook, social networks, phone records, emails

- who-talks-to-whom" data-sets contain
  - the network structure of communication among pairs of people
  - a measure of the strength of the tie (e.g., the total time that two people spend talking to each other)

- Experiment by Onnela et al. (2007)
  - Data maintained by a cell-phone provider that covered roughly 20% of the USA population
  - Period of observation = 18 weeks
  - There is an edge joining two nodes if they made phone calls to each other in both directions in the observation period
  - The strength of the tie is measured by the conversation minutes

Generalizing the Notions of Local Bridges and Weak Ties

- The Granovetter's hypothesis is based on two strong dichotomies
  - Ties can be either strong or weak
  - Links can be local bridges or not

- When examining real data it is useful to have definitions that exhibit smoother gradations

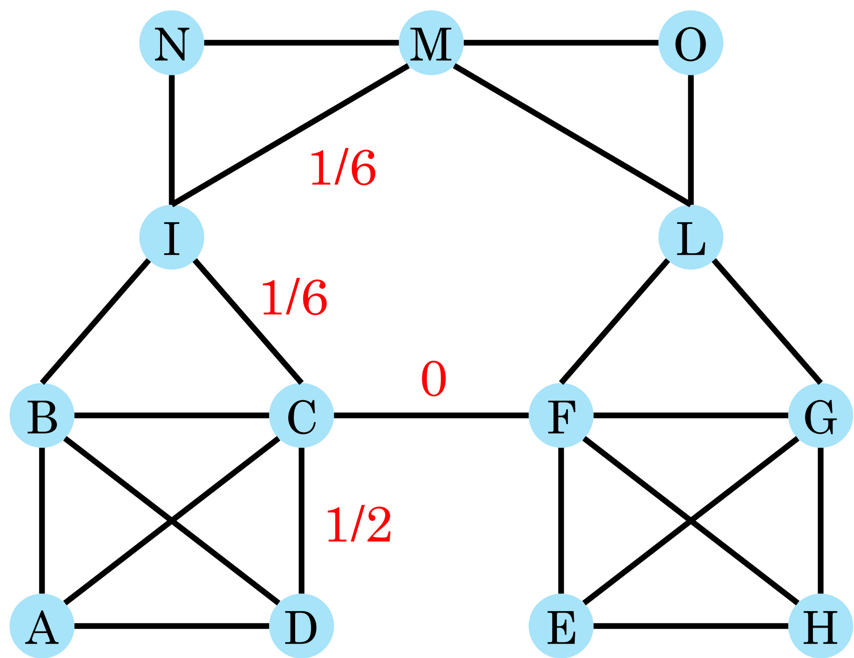
- We can introduce quantities for
  - measuring the strength of the tie
  - Measuring how "similar" is a link to a local bridge

Measuring the Strength of the Tie

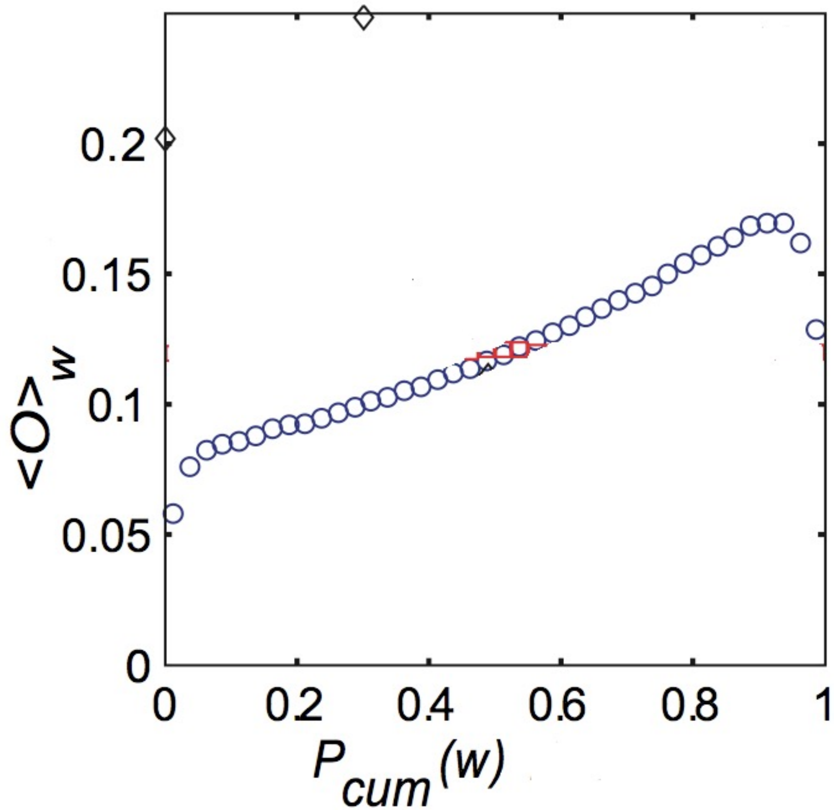
- In the Annela experiment the strength of an edge is a numerical quantity,
  - the total number of minutes spent on phone calls between the two ends of the edge
  - Measures with good approximation how "close" are the two nodes
- We can sort all the edges by tie strength and have statistics on the strength of ties
  - E.g., the percentile an edge occupies in the list of the links ordered by strength

NEIGHBORHOOD OVERLAP

- The similarity of a link to a local bridge is measured by the neighborhood overlap
- The neighborhood overlap of A and B (NO(A, B)) is
  - NO(A, B) = #(nodes that are neighbors of both A and B)/ #(nodes that are neighbors of either A or B)
  - If AB is a local bridge → NO(A, B) = 0
  - If NO(A, B) is small the link AB is an "almost" local bridge



EMPIRICAL RESULTS ON STRENGTH OF TIES AND NEIGHBORHOOD OVERLAP



A plot of NO of edges as a function of their percentile in the sorted order of all edges by tie strength

- The experiment confirms the Granovetter prediction that NO (global) increases with the strength of the ties (local)
- Can we also confirm the hypothesis that crucial information mainly travels along weak ties?

Network Structure

- Granovetter predicted that a social network consists of different tightly-knit communities, each one containing a large number of stronger ties, that are connected through weak ties
  - Weak ties play the crucial role of maintaining the communities connected each other

- The experiment (Annela et al.)
  1. Delete edges one at time in no-increasing order of strength
     - The size of the giant component decreases gradually
  2. Delete edges one at time in no-decreasing order of strength
     - The size of the giant component decreases more rapidly, and its remnants broke apart abruptly once a critical number of weak ties had been removed

- Conclusion
  - the weak ties provide the crucial connective structure for holding together disparate communities, and for keeping the global structure of the giant component intact

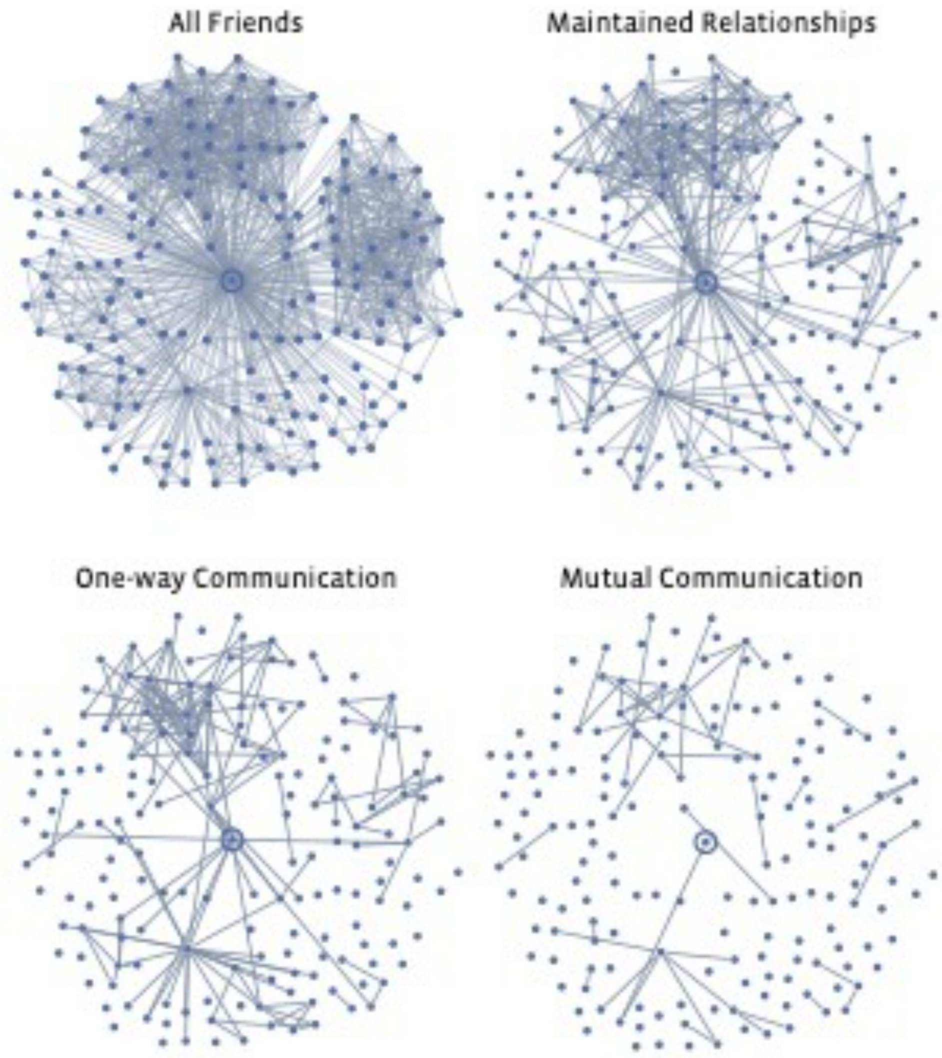
Strength of Ties in Online Social Networks

- An increasing amount of social interactions are moving on-line
  - Is this changing the way in which we maintain and access our social networks ?
  - First studies appeared at early 90s with the emergence of Internet

- In online social networks people maintain much larger explicit lists of friends in their profiles
  - They can maintain relations with people, scattered all over the world

- Several experiments run on social media as Facebook (Cameron et al, 2009) and Twitter (Huberman et al, 2009) proved that users of online social networks tend to have more but weaker social relations
  - Information spreads very fast
  - passive engagement
    - A user follows a relation only through notifications but with no direct relations
  - strong ties are limited in number since they require the continuous investment of time and effort to maintain

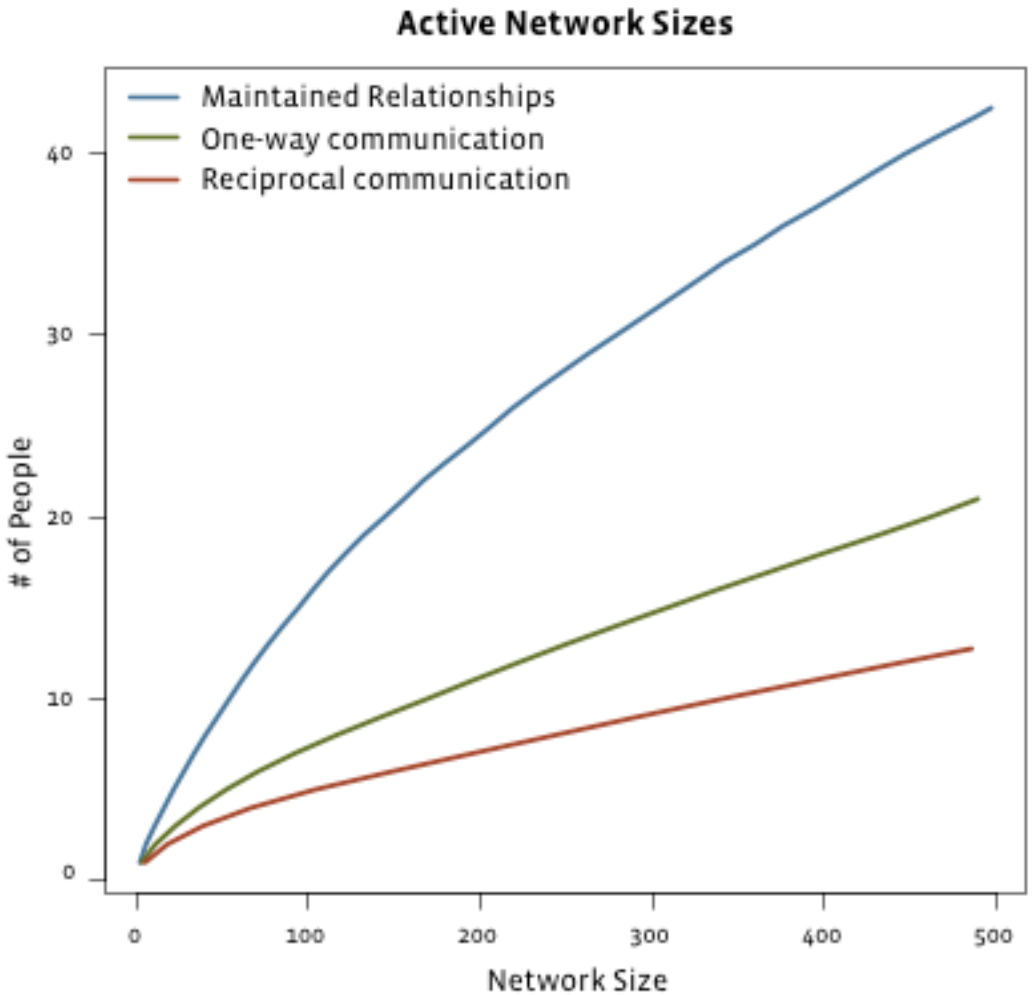
Strength of Ties in Facebook



Marlow et al (2009)

- Stratification of links of a Facebook user according to their use

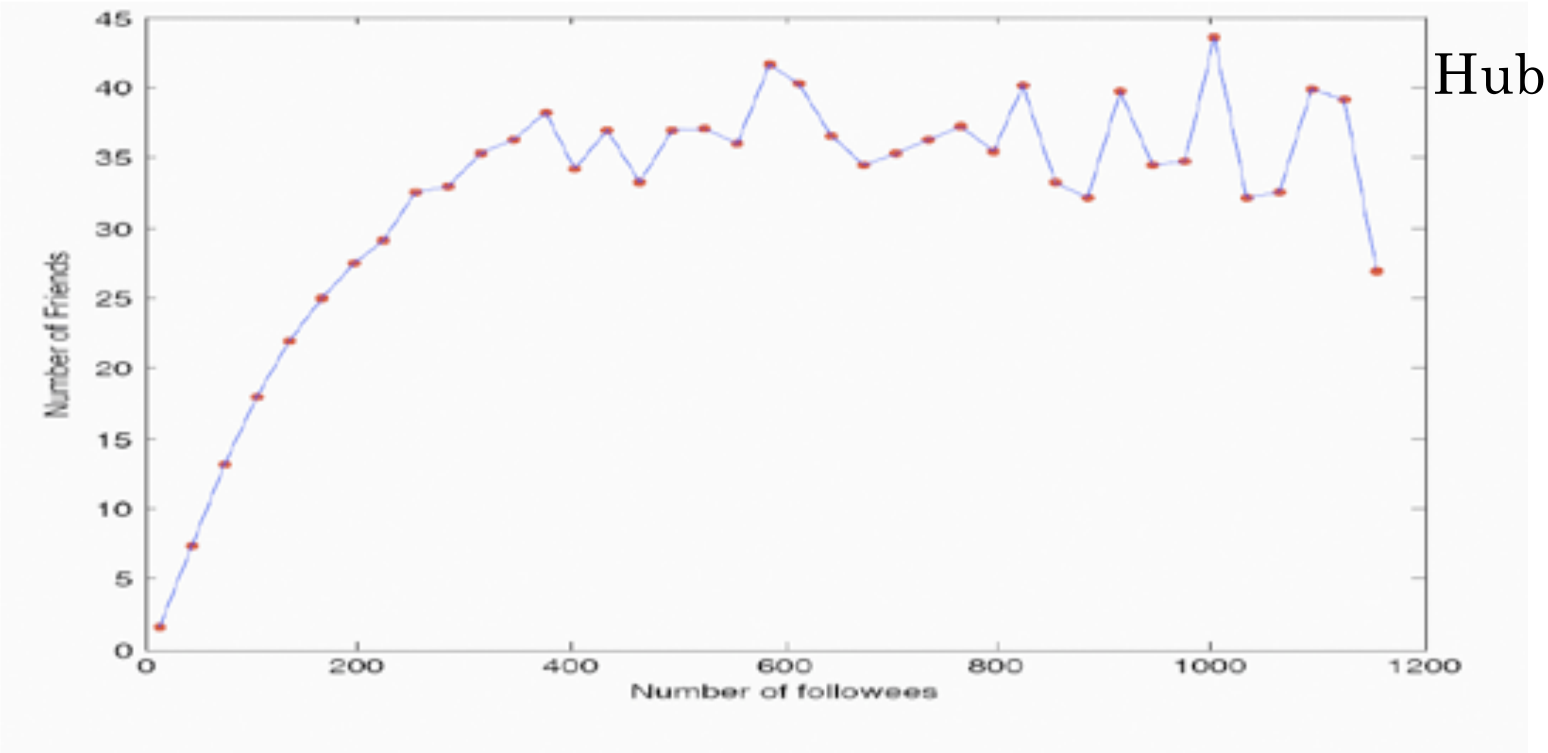
Strength of Ties in Facebook



- The number of links corresponding to maintained relationships, one-way communication, and reciprocal communication as a function of the total neighborhood size for users on Facebook

Strength of Ties in Twitter

- In Twitter we can distinguish two types of ties
  - Weak ties: a user declares to follow another user
  - Strong ties: a user publishes public messages specifically intended for a particular user



Huberman et al (2009)

- Even for users with over 1000 followes no more than 50 strong ties

Roles of Nodes in a Social Network

- Previous discussion suggests a general view of social networks in terms of tightly-knit groups and the weak ties that link them
  - Our analysis focused on the role of edges
- Can we recognize different roles that nodes play in these networks?
  - some nodes are positioned at the interface between multiple groups, with access to boundary-spanning edges
  - others are positioned in the middle of a single group
  - They experiment different experiences about the social network

Embeddedness and Structural Holes

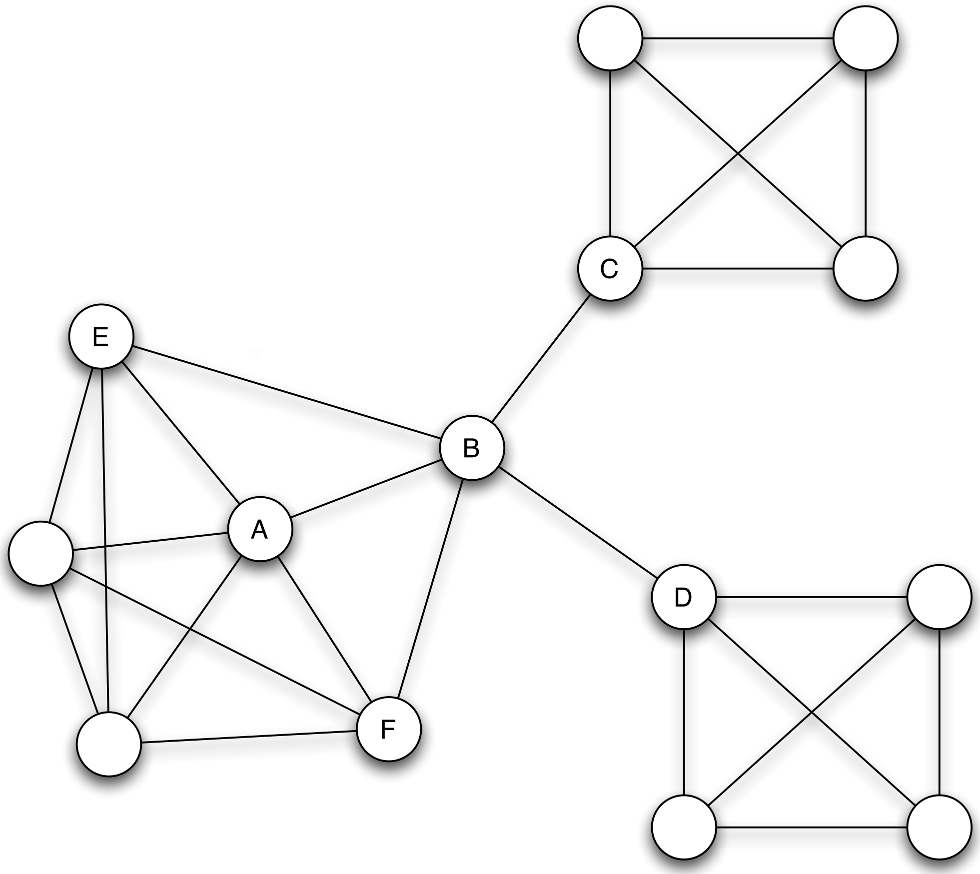
- Embeddedness of an edge AB = #(common neighbors of A and B)
  - Numerator of the Neighborhood Overlp function
    - A local bridge has zero embeddedness
  - Trust among A and B increases with the embeddedness ot the edge connecting them

- A node adiacent to edges with high embeddedness has easier relations since she has larger trust in her neighbours and a better reputation

- A structural hole describes the absence (or scarsity) of connections between two different parts of the netowrk
  - Such holes are crossed by the local bridges

- Several empirical experiments showed that nodes adiacent to structural holes have a positional advantage
  - They have a control on the flow of information
  - They early access to information originating in multiple, non-interacting parts of the network

AN EXAMPLE



- A is adiacent to all edges with high embeddedness
  - She's embedded in a tightly-knit community
- B, C e D are adiacent to a structural hole
  - They control all the interactions among the different communities

Node Centrality Measures

- Node (edge) centrality measures are used to measure the importance of each node with respect to a given process

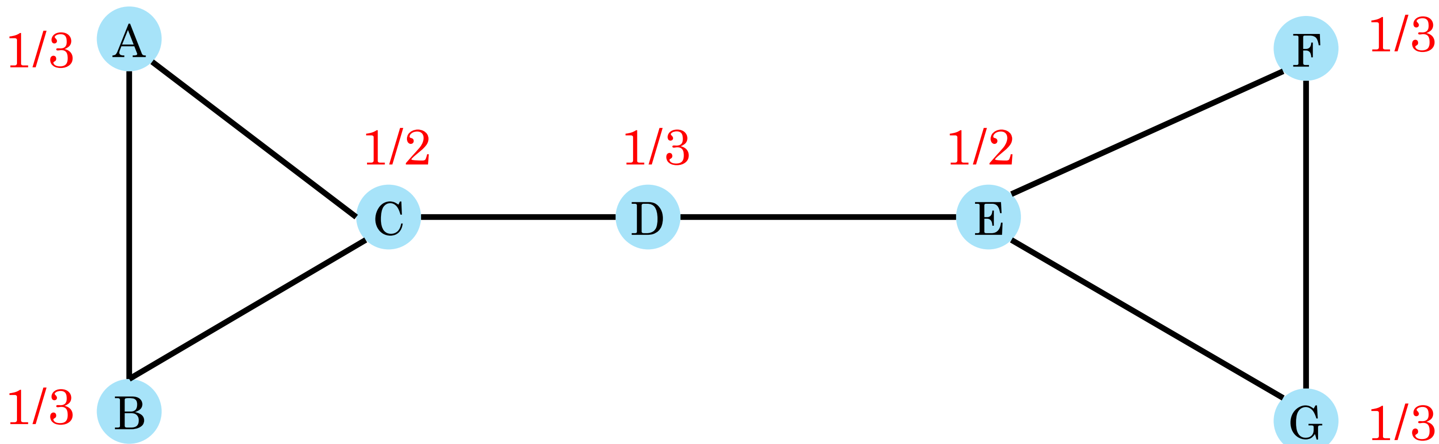
- Different centrality measures are used to deal with different processes in the network

- The more popular measures are
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality
  - Eigenvector centrality

- All measures are normalized to give a value in [0, 1]
  - They measure how a node is "important"

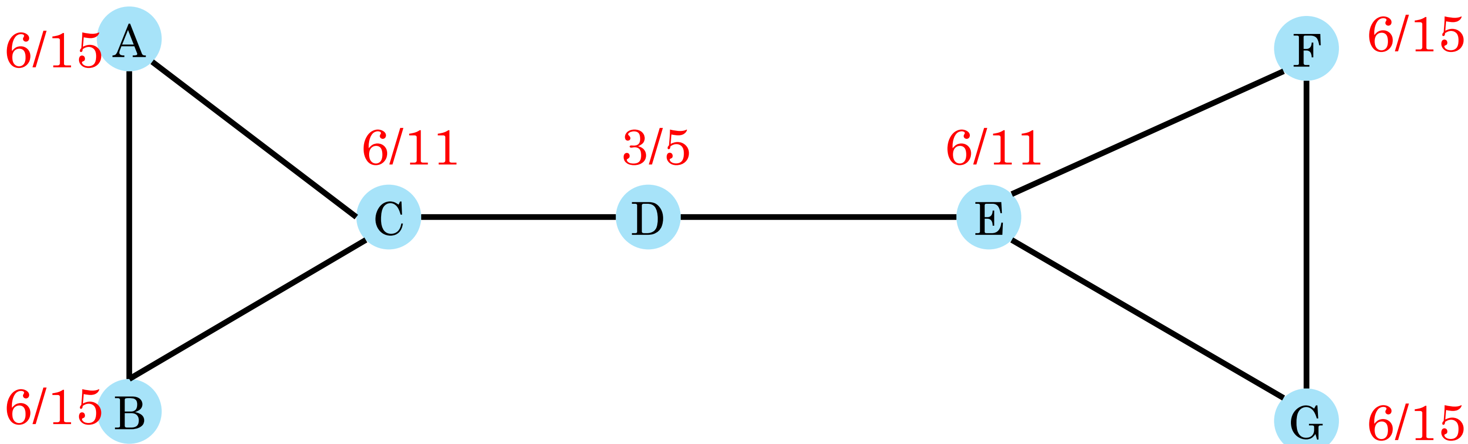
DEGREE CENTRALITY

- Degree centrality of node z
  - degree(z)/(n-1)
  - Measures the importance of a node in terms of the number of its relations
  - Simple but very rudimental



Closeness Centrality

- Closeness centrality of node z
  - $(n\text{-}1)/\sum_{u \neq z} d(u,z)$
  - Inverse of the average distance
  - Measures how fast a node can reach (or is reached) by all the other nodes in the network



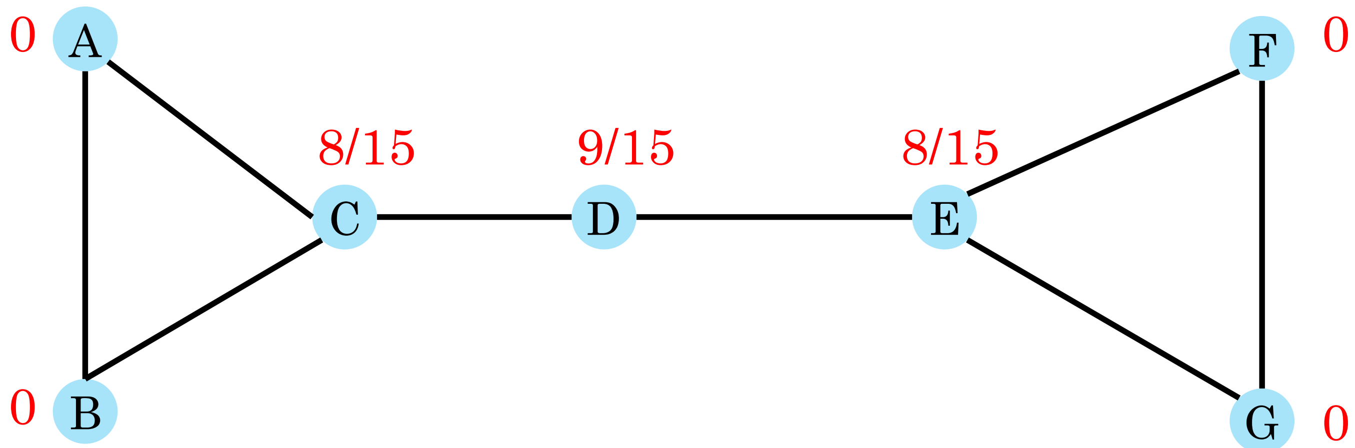
Closeness Centrality with a Decay Parameter

- We could want to assign larger weight to close nodes

- Closeness centrality with decay parameter δ, with $0 < \delta < 1$, of node z
  - $\sum_{u \neq z} \delta^{\mathrm{d}(u,z)}$
  - More weight to closer nodes
  - For $\delta \rightarrow 1$ tends to the size of the connected component of z
  - Per $\delta \rightarrow 0$ tends to the degree of z

BETWEENESS CENTRALITY

- Betweenness centrality of node z
  - $2 / (n-1)(n-2) \sum_{u \neq v,\, z \neq u,v} P_z(u,v) / P(u,v)$
  - $P(u,v)$ = # shortest paths between u and v
  - $P_z(u,v)$ = # shortest paths between u and v going through z
  - Measures how much z is crucial for the transmission of information between each pair of nodes in the network



Eigenvector Centrality

- The Eigenvector Centrality of node z
  - measures the relevance of z in terms of the relevance of its neighbors

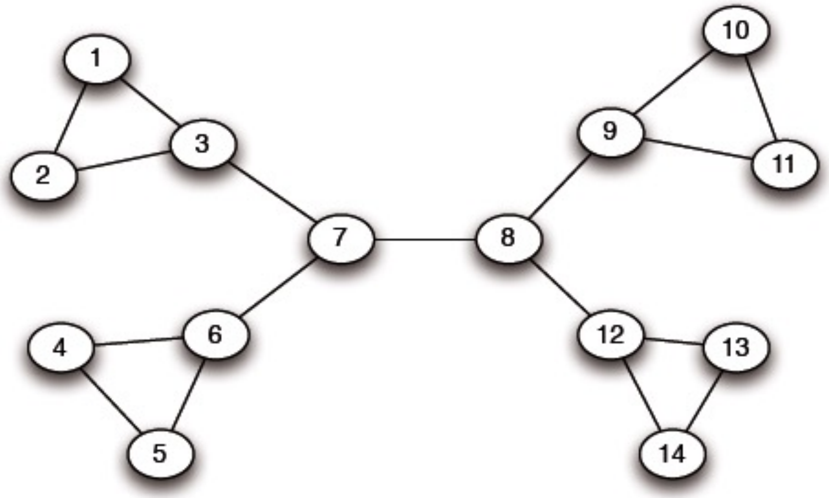
- What does it recall?

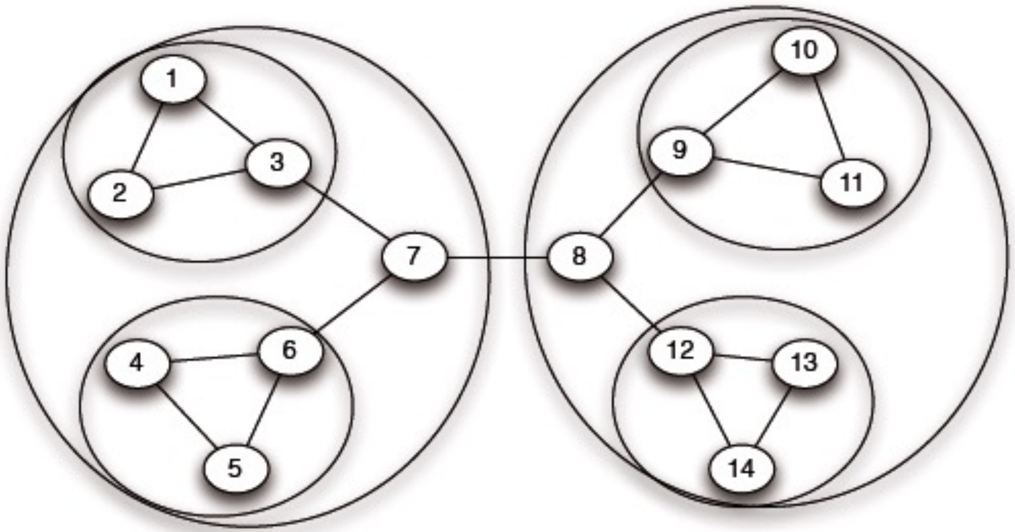
- $C_z(G) = \sum_{u \neq z} G_{u,z}\, C_u(G)$
  - In a matricial form $\lambda C(G) = G \times C(G)$
    - C(G) is an eigenvector of G and λ is its corresponding eigenvalue
    - We get the greatest eigenvalue that for our networks is always no-negative

GRAPH PARTITIONING

- From the previous discussion we obtain a model of social network consisting of different tightly-knit communities that are connected through local bridges
  - We take this definition intentionally informal
- How can we identify the communities and partition the graph into its components?
  - Crucial for understanding the network structure and unveil dynamic behaviors of the network
    - i.e., marketing, voting, artificial intelligence, bioinformatics, etc.
  - No trivial algorithmic problem
- Several heuristic methods proposed for graph partition and community detection
  - Agglomerative methods
    - Start by indivduals and try to agglomerate them in groups
  - Divisive methods
    - Start from the network as a whole and try to split is in highly connected subnetworks that are sparsly connected each other

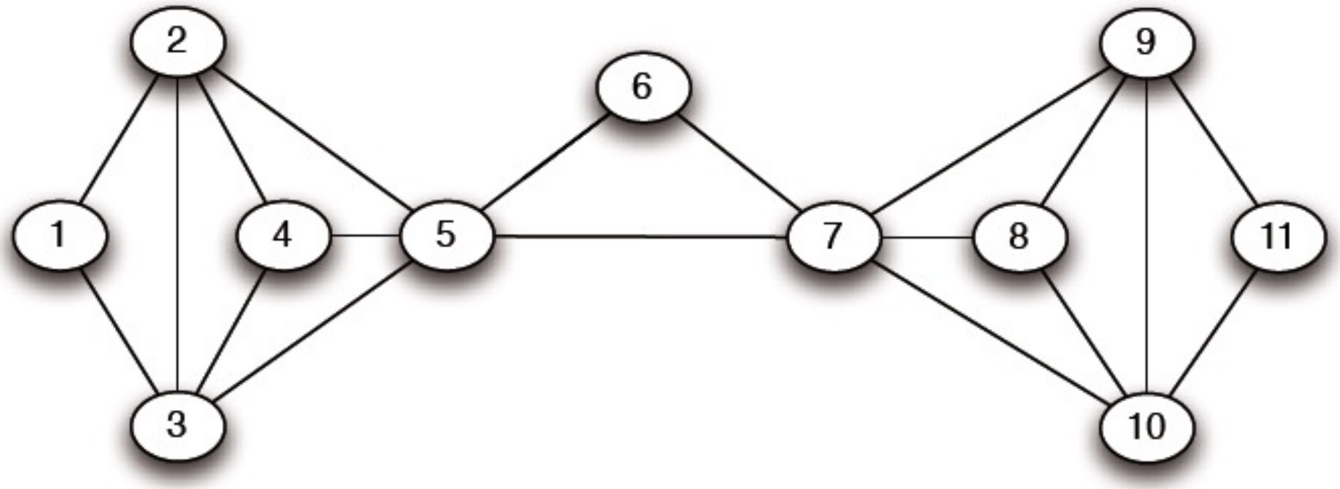
An Example of Graph Partitioning in Communities





Graph Partitioning by Division

- Is it sufficient to identifiy the local bridges?
  - No



- A possible solution
  - Compute the betweenness of each edge
  - On edges with the highest betweenness travels most of the information
    - They are local bridges

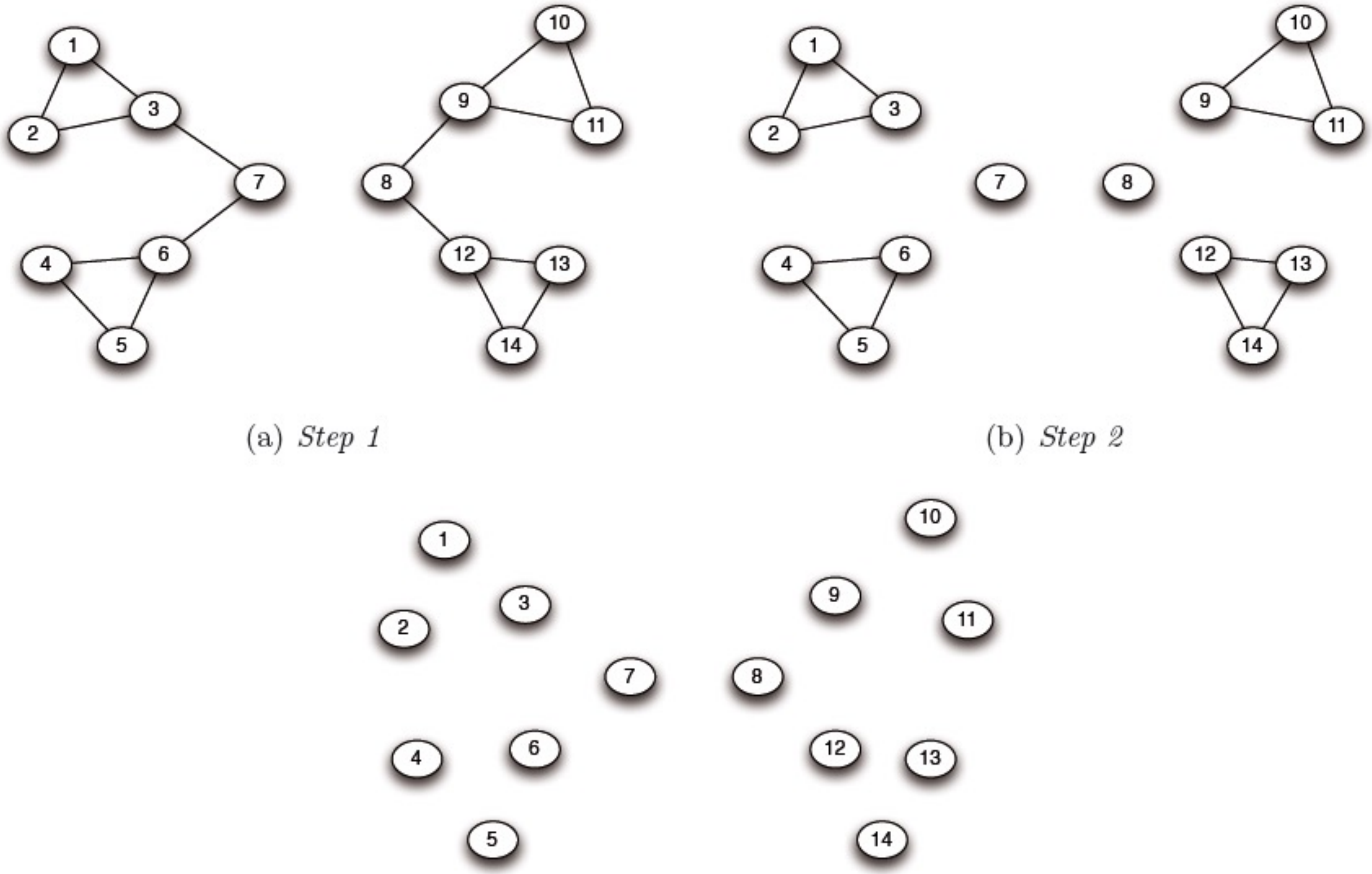
Graph Partitioning by Division

Girvan-Newman Algorithm

1. Search for the edges with the highest betweenness and remove them from the graph
   - If the graph splits in different components you have found the communities
2. Recompute the betweenness on each component and go back to step 1
   - Each component splits in sub-components nested in the father component
3. Continues until all edges have been removed

- This algorithm can be adapted to compute also node betweenness

An Example



(a) *Step 1* (b) *Step 2*

(c) *Step 3*

How to Compute the Edge Betweenness

- The Girman-Newman algorithm has to recompute at each round the betweenness of all the remaining edges
  - Betweenness depends on the number of shortest paths between each pair of nodes
  - How can we do efficently?

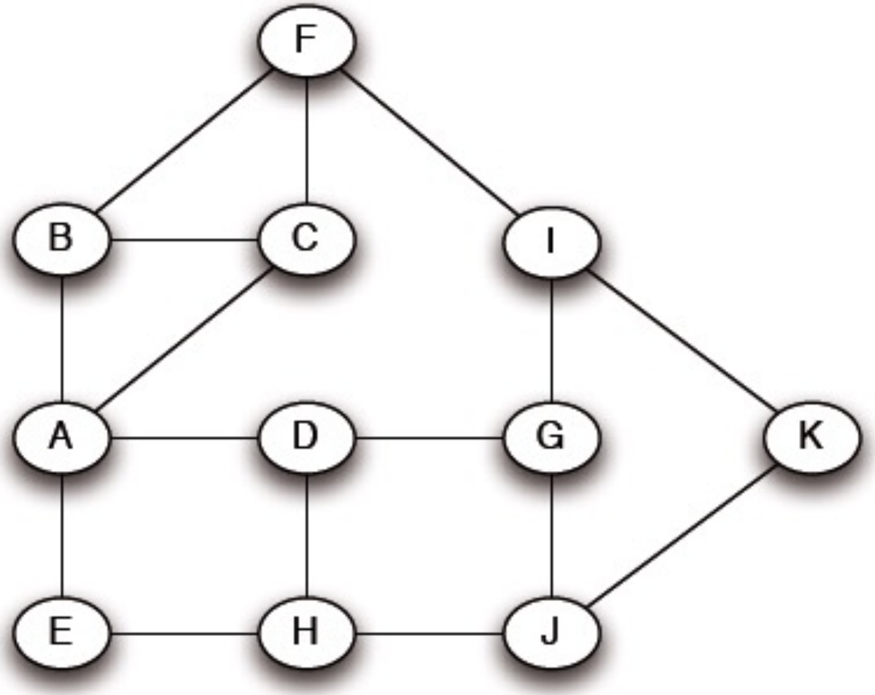
- Idea
  - Compute a BFS from each starting node u
  - Compute how a unit of flow starting from u spreads in the network

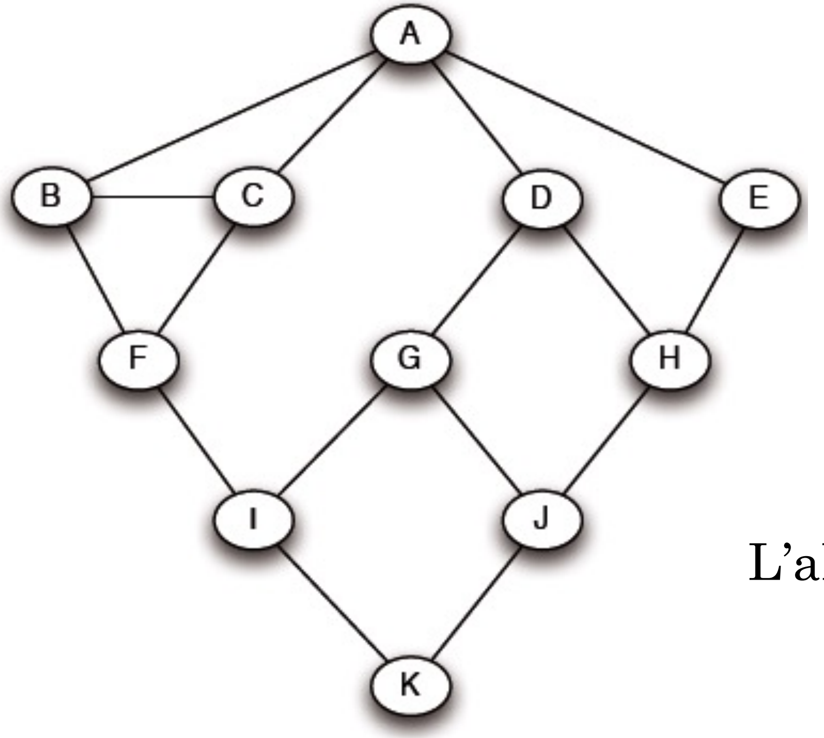
An Algorithm to Compute the Betweenness

- The algorithm
  1. **Run a BFS from each source node u and compute the shortest path tree**
  2. **For each node *v*, compute how many shortest paths there are between *u* and *v***
     - If node v si at the k-th level of the BFS tree rooted at u all the shortest paths from u to v have length k and go through fathers of v *i*n the BFS tree
     - The number of shortest paths to v is given by the sum of the shortest paths from u to the fathers of v
  3. **Compute how much flow goes though each edge in the graph**
     - Bottom up
     - At each node assign a unit of flow starting from that node and all the flow coming from its children
     - Each node distributes equally all its flow among its fathers

An Example

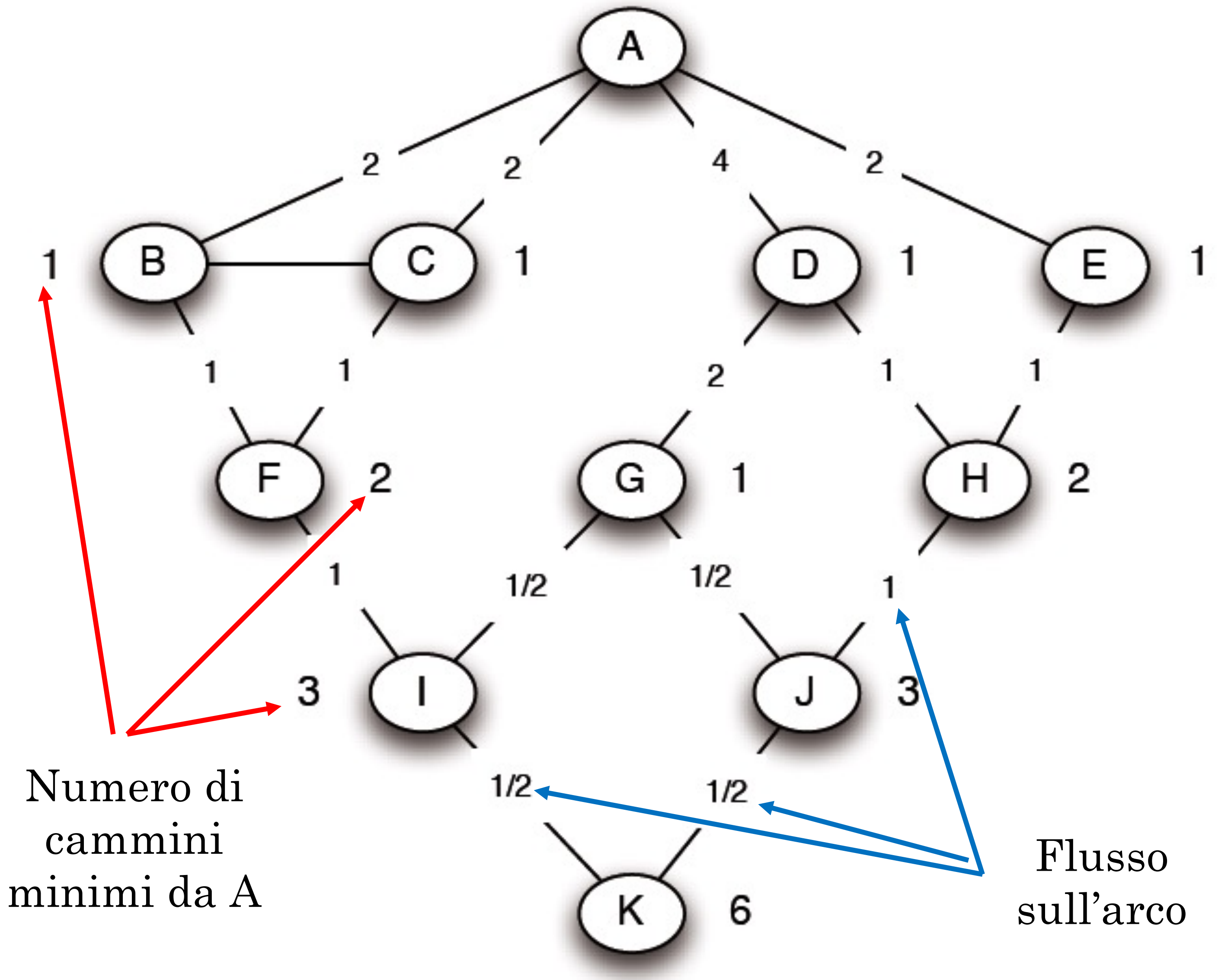
La rete





L'albero della BFS da A

AN EXAMPLE - CONTD



How Efficient is Girman-Newman?

- At each round the Girman-Newmann algorithm computes the betweenness of each remaining pair of edges
  - Not scalable for networks of high dimension
- Alternative approaches
  - Compute an approximation of the betweenness
  - Use alternative algorithms for graph paritioning, based on agglomerative methods
- Hot research area