# Social Network Analysis

Sakthi Balan M

LNMIIT

# Social Network Analysis

## Network Measures

LNMIIT

**Some pictures have been used from the following books:**
**(This is only for learning purpose)**

**Social Media Mining by Zafaranai et al,**
**Network Science by Barabasi**
**Networks, Crowd, Behaviour by Kleinberg et al**

# Network Measures — Objective

- Who are the central figures (influential individuals) in the network?
- What interaction patterns are common in friends?
- Who are the like-minded users and how can we find these similar individuals?
- To answer the above questions we need:
  - to quantify centrality
  - to quantify the level of interaction
  - to quantify similarity among individuals

LNMIIT

# Centrality

**Centrality defines how important a node is within a network**

- Degree Centrality
- Eigenvector Centrality
- Katz Centrality
- PageRank
- Betweeness Centrality
- Closeness Centrality
- Group Centrality

LNMIIT

# Degree Centrality

- degree centrality measures node with respect to their degree
  - node having more connections implies it is important
  - $C_d(v_i) = d_i$ for an undirected graph
  - For a directed graph
    - $C_d(v_i) = d_i^{in}$ (Prestige) or $C_d(v_i) = d_i^{out}$ (Gregariousness)
    - Another definition is $C_d(v_i) = d_i^{in} + d_i^{out}$

LNMIIT

# Degree Centrality

- Normalising degree centrality

$$C_d^{norm}(v_i) = \frac{d_i}{n-1} \qquad C_d^{max}(v_i) = \frac{d_i}{max_j d_j} \qquad C_d^{sum}(v_i) = \frac{d_i}{\sum_j d_j} = \frac{d_i}{2m}$$

# Eigenvector Centrality

- Philosophy is *having more important friends provides a stronger signal*

- Tries to generalize degree centrality of a node by using the importance of its neighbor nodes (or incoming neighbors in directed graphs)

$$c_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} A_{j,i} c_e(v_i)$$

Let $\mathbb{C} = (c_e(v_1), c_e(v_2), \ldots, c_e(v_n))^T$

then $\lambda \mathbb{C}_e = A^T \mathbb{C}_e$ this basically means

$\mathbb{C}_e$ **is the eigen vector of** $A^T$ **and** $\lambda$ **is the eigen value**

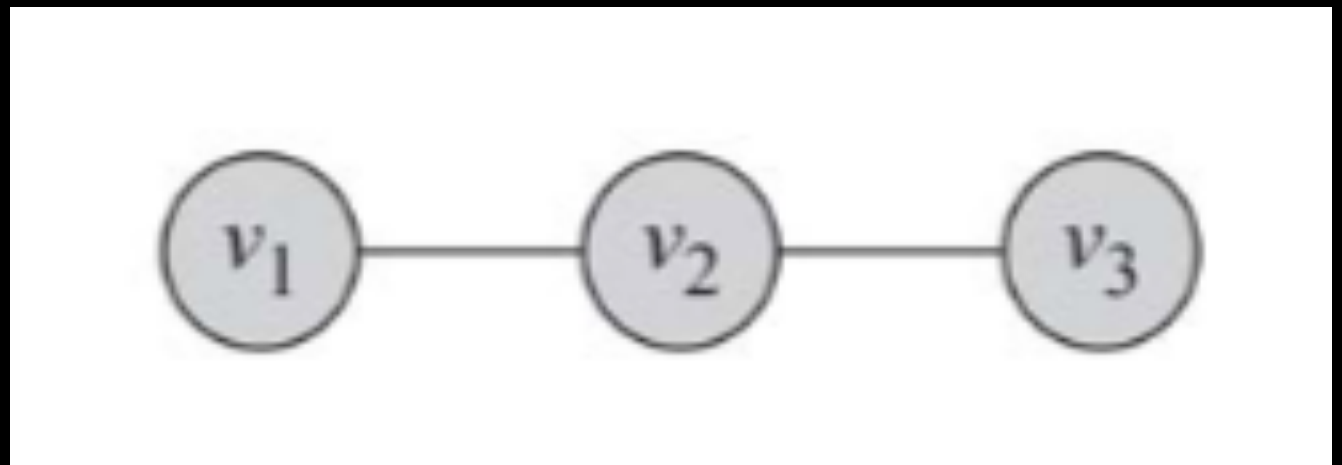
LNMIIT

# Eigenvector Centrality

**Theorem 3.1** (Perron-Frobenius Theorem). *Let $A \in \mathbb{R}^{n \times n}$ represent the adjacency matrix for a [strongly] connected graph or $A : A_{i,j} > 0$ (i.e. a positive n by n matrix). There exists a positive real number (Perron-Frobenius eigenvalue) $\lambda_{\max}$, such that $\lambda_{\max}$ is an eigenvalue of A and any other eigenvalue of A is strictly smaller than $\lambda_{\max}$. Furthermore, there exists a corresponding eigenvector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ of A with eigenvalue $\lambda_{\max}$ such that $\forall v_i > 0$.*

LNMIIT

# Find the eigenvector centrality for the following graph




**Example 3.2.** *For the graph shown in Figure 3.2(a), the adjacency matrix is*

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{3.10}$$

*Based on Equation 3.9, we need to solve* $\lambda \mathbf{C}_e = A\mathbf{C}_e$, *or*

$$(A - \lambda I)\mathbf{C}_e = 0. \tag{3.11}$$

*Assuming* $\mathbf{C}_e = [u_1\ u_2\ u_3]^T$,

$$\begin{bmatrix} 0-\lambda & 1 & 0 \\ 1 & 0-\lambda & 1 \\ 0 & 1 & 0-\lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3.12}$$

*Since* $\mathbf{C}_e \neq [0\ 0\ 0]^T$, *the characteristic equation is*

$$\det(A - \lambda I) = \begin{vmatrix} 0-\lambda & 1 & 0 \\ 1 & 0-\lambda & 1 \\ 0 & 1 & 0-\lambda \end{vmatrix} = 0, \tag{3.13}$$

*or equivalently,*

$$(-\lambda)(\lambda^2 - 1) - 1(-\lambda) = 2\lambda - \lambda^3 = \lambda(2 - \lambda^2) = 0. \tag{3.14}$$

*So the eigenvalues are* $(-\sqrt{2}, 0, +\sqrt{2})$. *We select the largest eigenvalue:* $\sqrt{2}$. *We compute the corresponding eigenvector:*

$$\begin{bmatrix} 0-\sqrt{2} & 1 & 0 \\ 1 & 0-\sqrt{2} & 1 \\ 0 & 1 & 0-\sqrt{2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3.15}$$

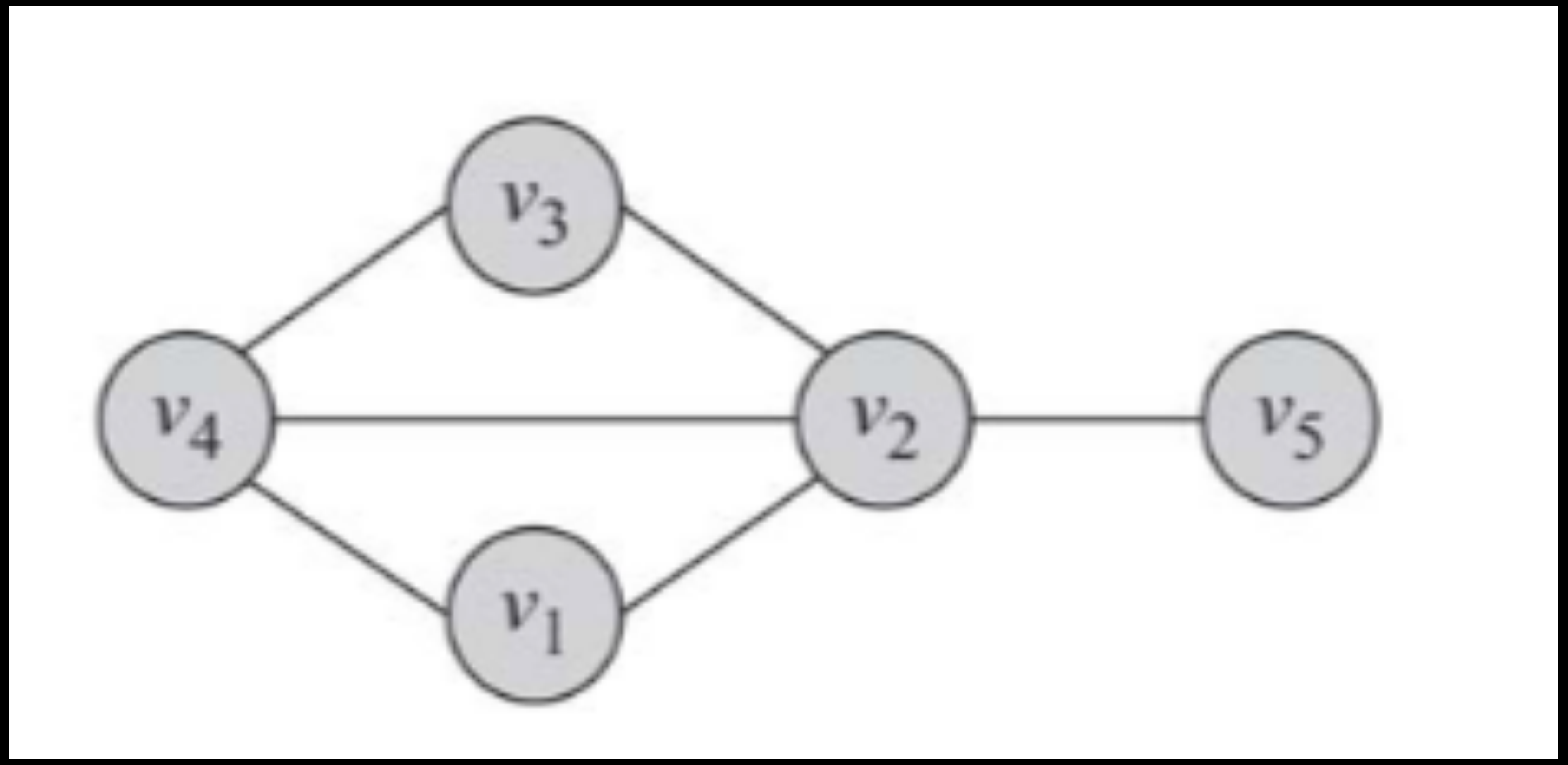
*Assuming* $\mathbf{C}_e$ *vector has norm 1, its solution is*

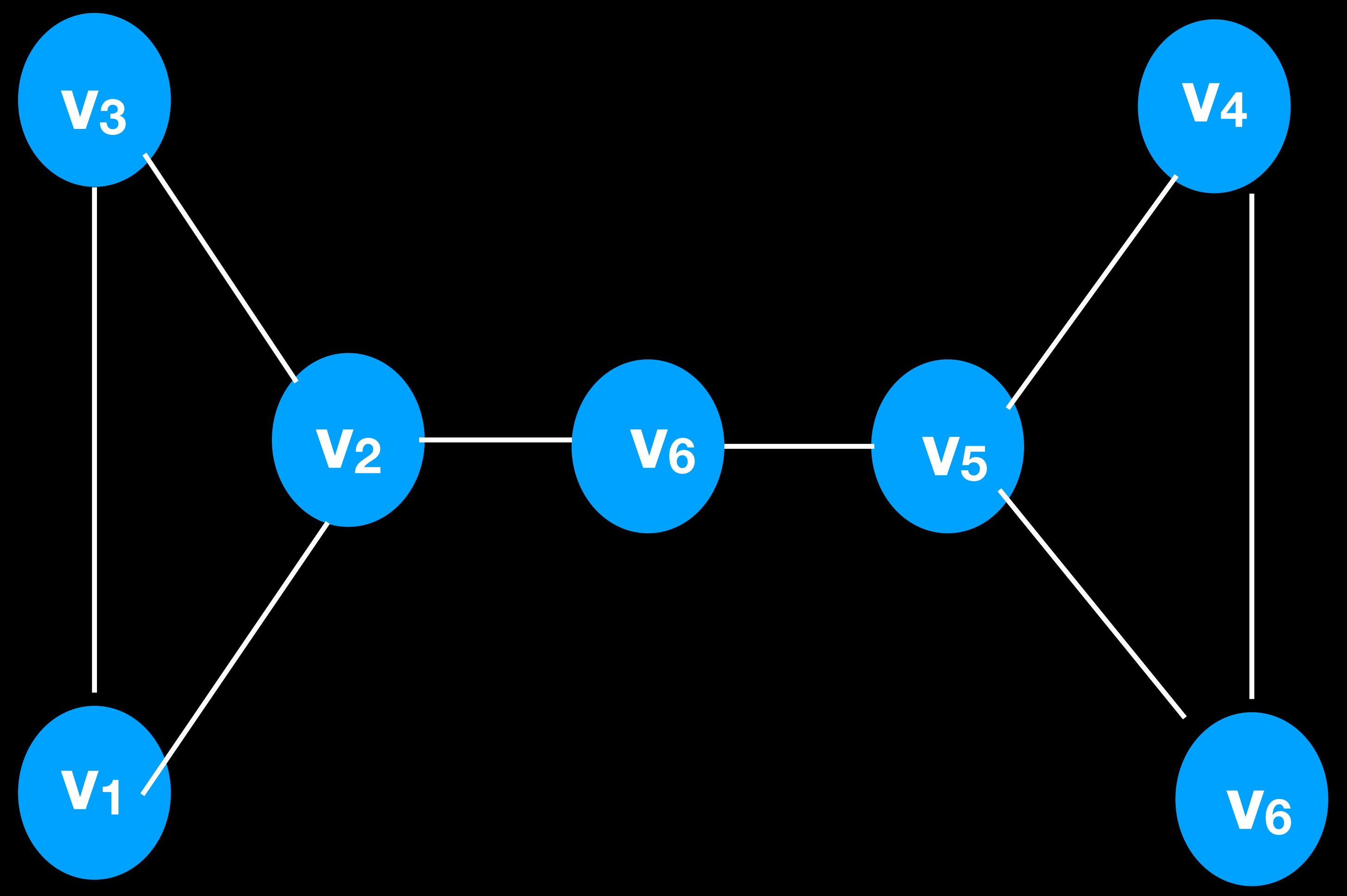
$$\mathbf{C}_e = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ \sqrt{2}/2 \\ 1/2 \end{bmatrix}, \tag{3.16}$$

*which denotes that node* $v_2$ *is the most central node and nodes* $v_1$ *and* $v_3$ *have equal centrality values.*

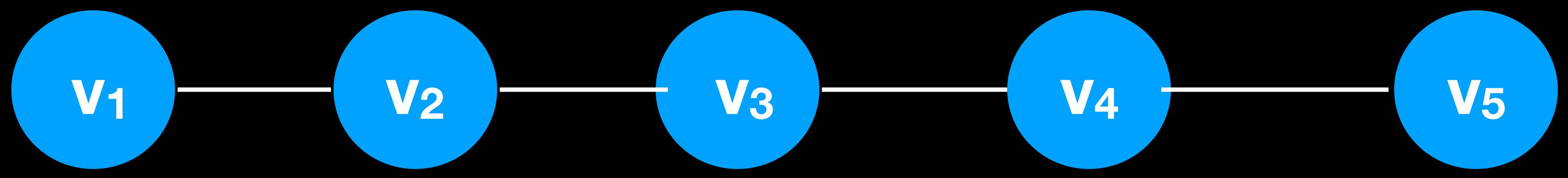
LNMIIT

Find the eigenvector centrality for the following graph




How about degree centrality and eigenvector centrality for the following graphs



# Katz Centrality

- Eigenvector Centrality is passed on to the nodes only when there are outgoing edges.

- In special cases when a node is in a directed acyclic graph, centrality becomes zero, even though the node can have many edges connected to it.

$$C_{katz}(v_i) = \alpha \sum_{j=1}^{n} A_{j,i} C_{katz}(v_i) + \beta$$

$$\mathbb{C}_{katz} = \alpha A^T \mathbb{C}_{katz} + \beta 1 \longrightarrow \mathbb{C}_{katz} = \beta(I - \alpha A^T)^{-1} \cdot 1$$

$$\mathbb{C}_{katz} = \beta(I - \alpha A^T)^{-1} \cdot 1$$

LNMIIT

# Katz Centrality

$$\mathbb{C}_{katz} = \beta(I - \alpha A^T)^{-1} \cdot 1$$

**When** $\alpha = 0$ **the eigenvector centrality is nullified**

**When** $det(I - \alpha A^T) = 0$ **centrality value diverges**

**When** $\alpha = 1/\lambda$ **the value diverges. Good value is** $\alpha < 1/\lambda$

# PageRank

- Katz centrality encounters some challenges.
  - Once a node becomes an authority (high centrality), it passes all its centrality along all of its out-links.
  - Not everyone known by a well known person is well known.

$$C_p(v_i) = \alpha \sum_{j=1}^{n} A_{j,i} \frac{C_p(v_j)}{d_j^{out}} + \beta \rightarrow C_p = \alpha A^T D^{-1} C_p + \beta 1 \rightarrow C_p = \beta(I - \alpha A^T D^{-1})^{-1} \cdot 1$$

**Here *D* is a diagonal matrix with entries of degrees in its main diagonal**

LNMIIT

# Betweenness Centrality

- How important nodes are in connecting other nodes
  - compute the number of shortest paths between other nodes that pass through $v_i$

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_s t}$$

How will you find the shortest path between various vertices?

**Normalisation of** $C_b(v_i)$

$$C_b^{norm}(v_i) = \frac{C_b(v_i)}{2\binom{(n-1)}{2}}$$

Will you use Dijsktra's Algorithm?

LNMIIT

# Closeness Centrality

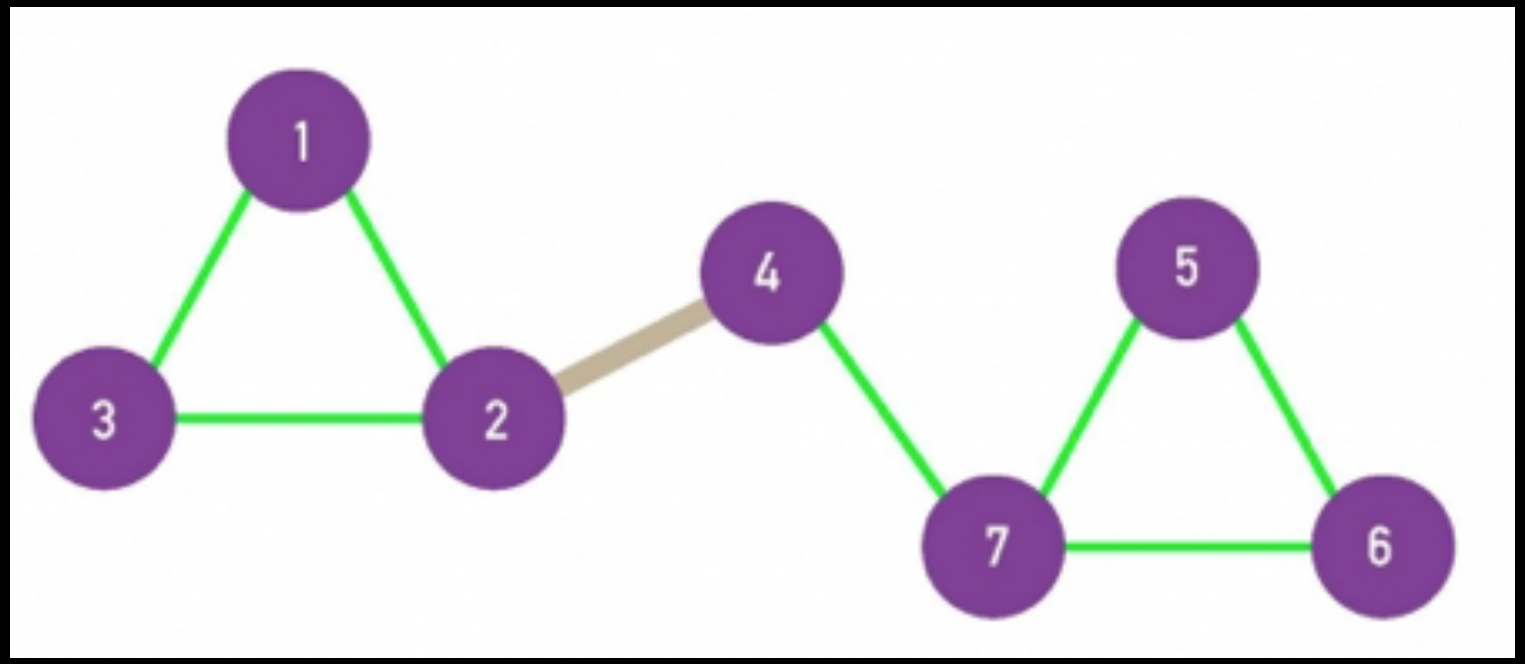
- Philosophy here is that the nodes that are central are the ones from where we can reach other nodes more quickly

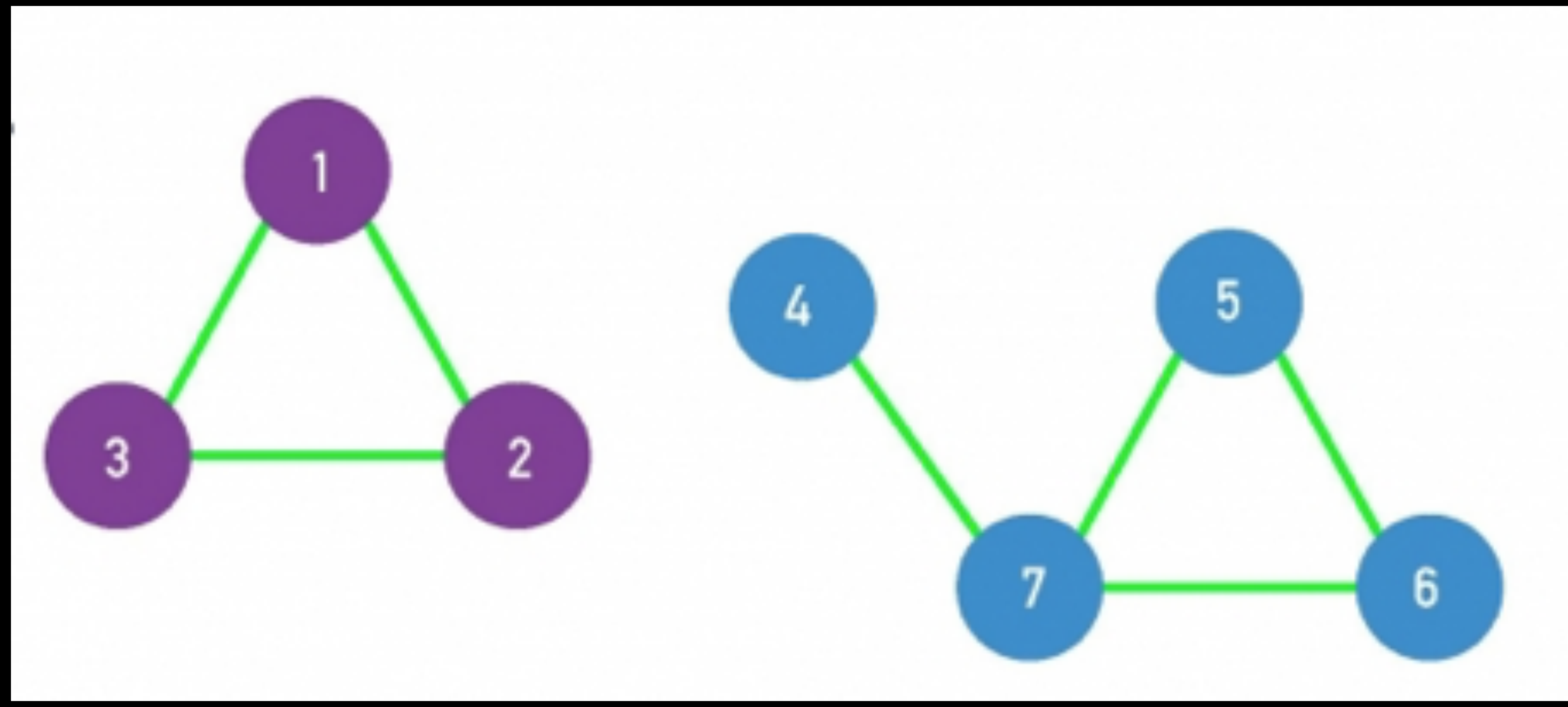
$$C_c(v_i) = \frac{1}{\bar{l}_{v_i}} \quad \textbf{where} \quad \bar{l}_{v_i} = \frac{1}{n-1} \sum_{v_i \neq v_j} l_{i,j}$$

LNMIIT

# Group Centrality Measures

# Connectedness - revisit

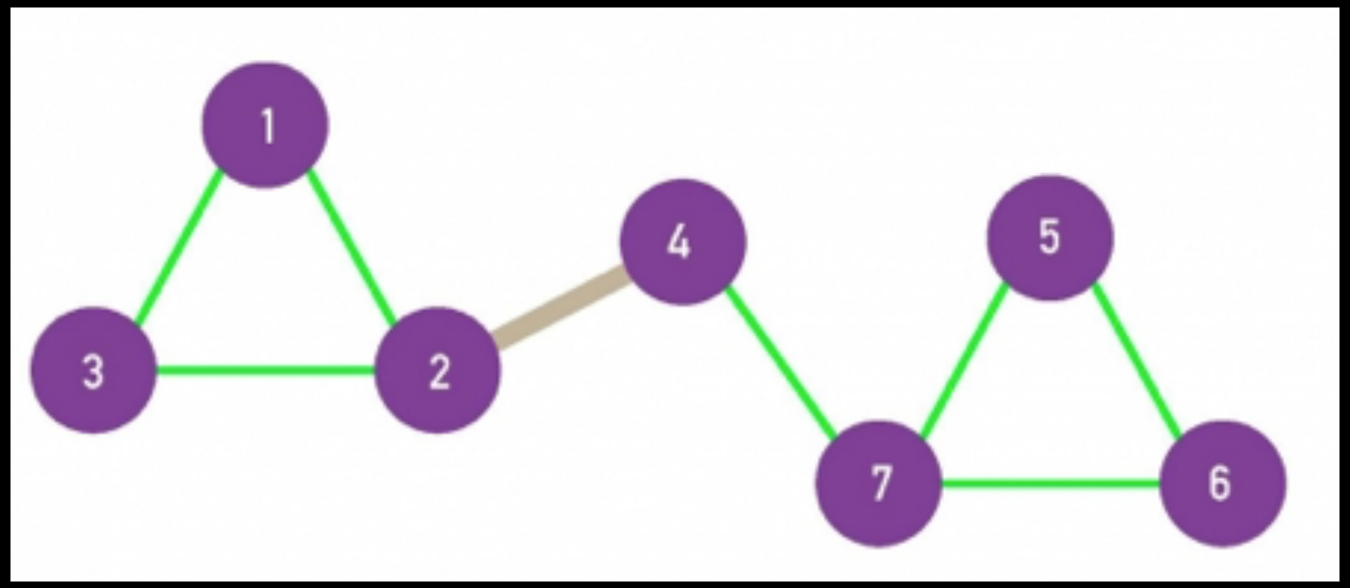


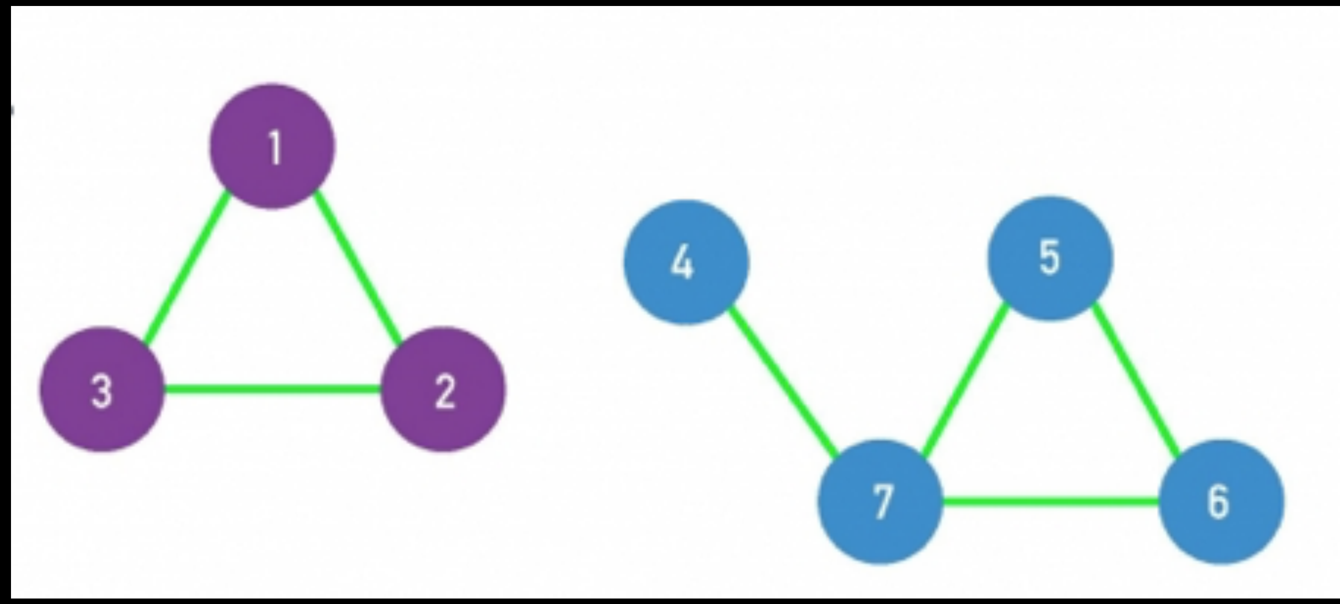

**How will you find if a graph is connected or not?**

**If not connected how will you find all the connected components of the graph?**

# Connectedness - revisit




$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$




$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

# Transitivity

- *A friend of my friend is my friend*

LNMIIT

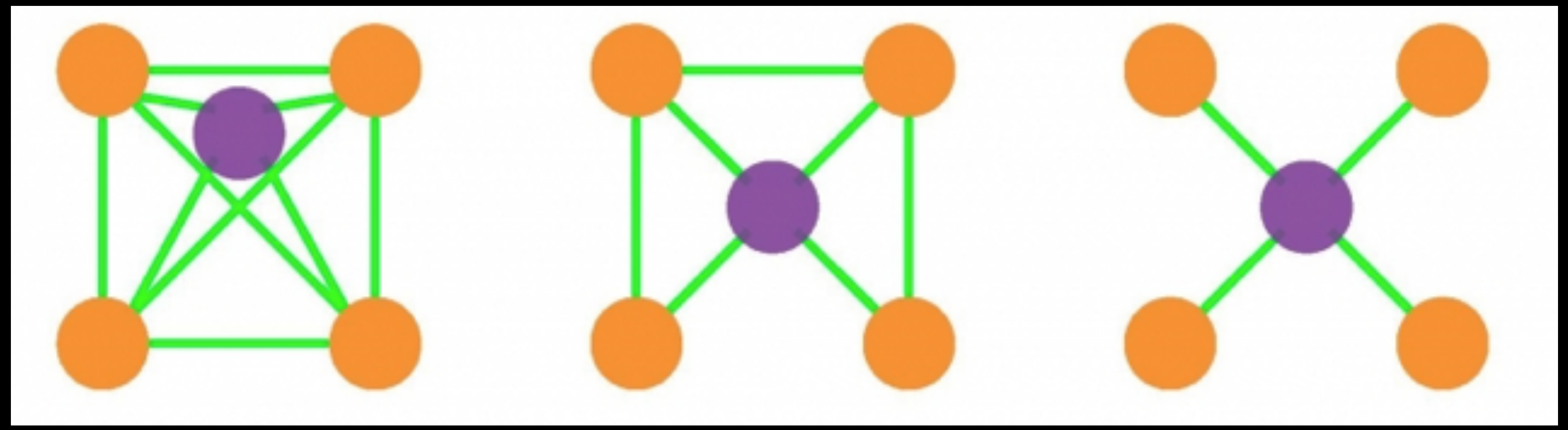
# Clustering Coefficient

- Measures the neighbors level of connectivity of a given node. For a node $i$ with degree $k_i$ the local clustering coefficient

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

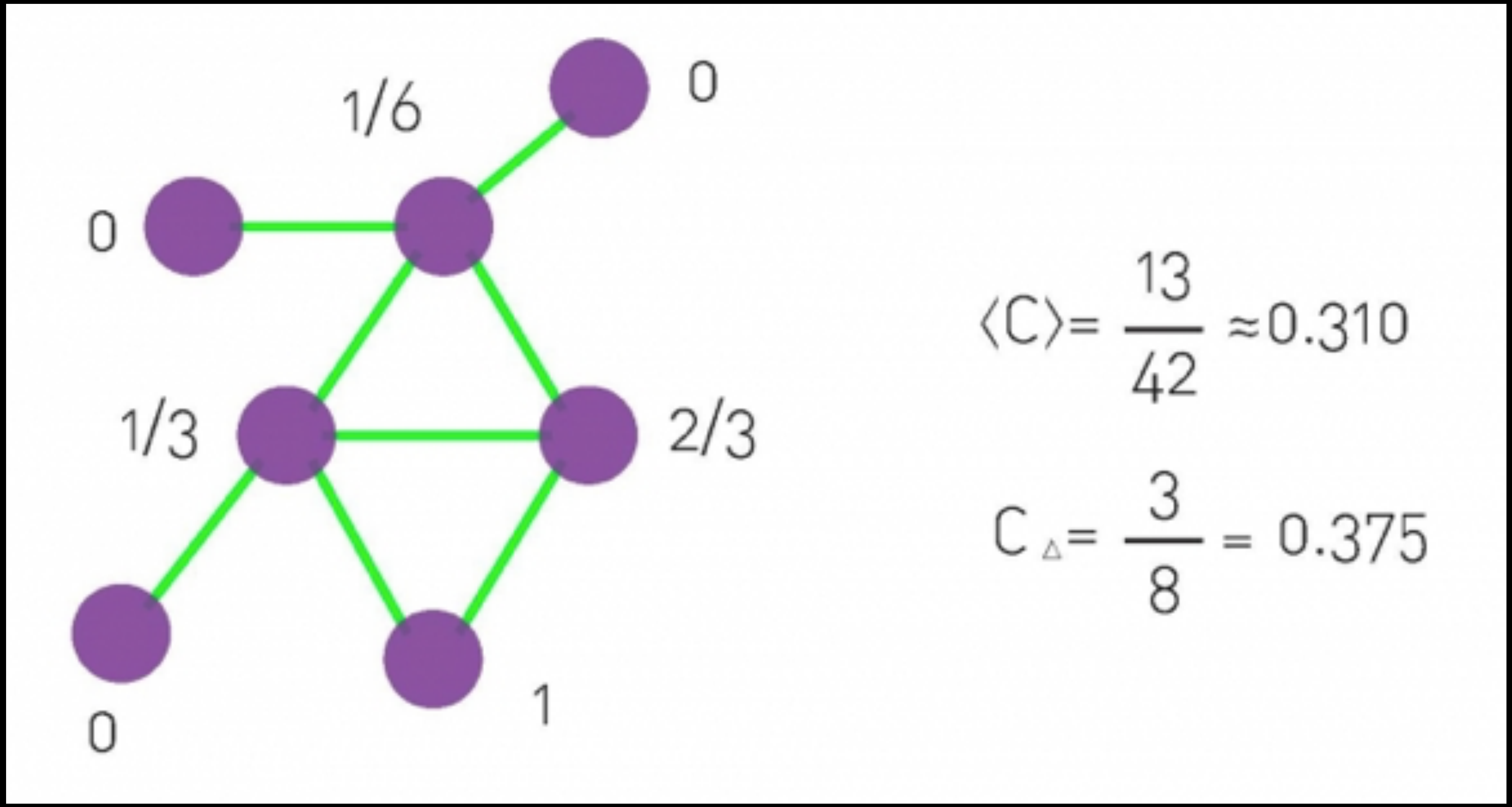
***$L_i$* represents the number of links between the *$k_i$* neighbors of node *$i$*.**

**Note that *$C_i$* is between 0 and 1**



LNMIIT

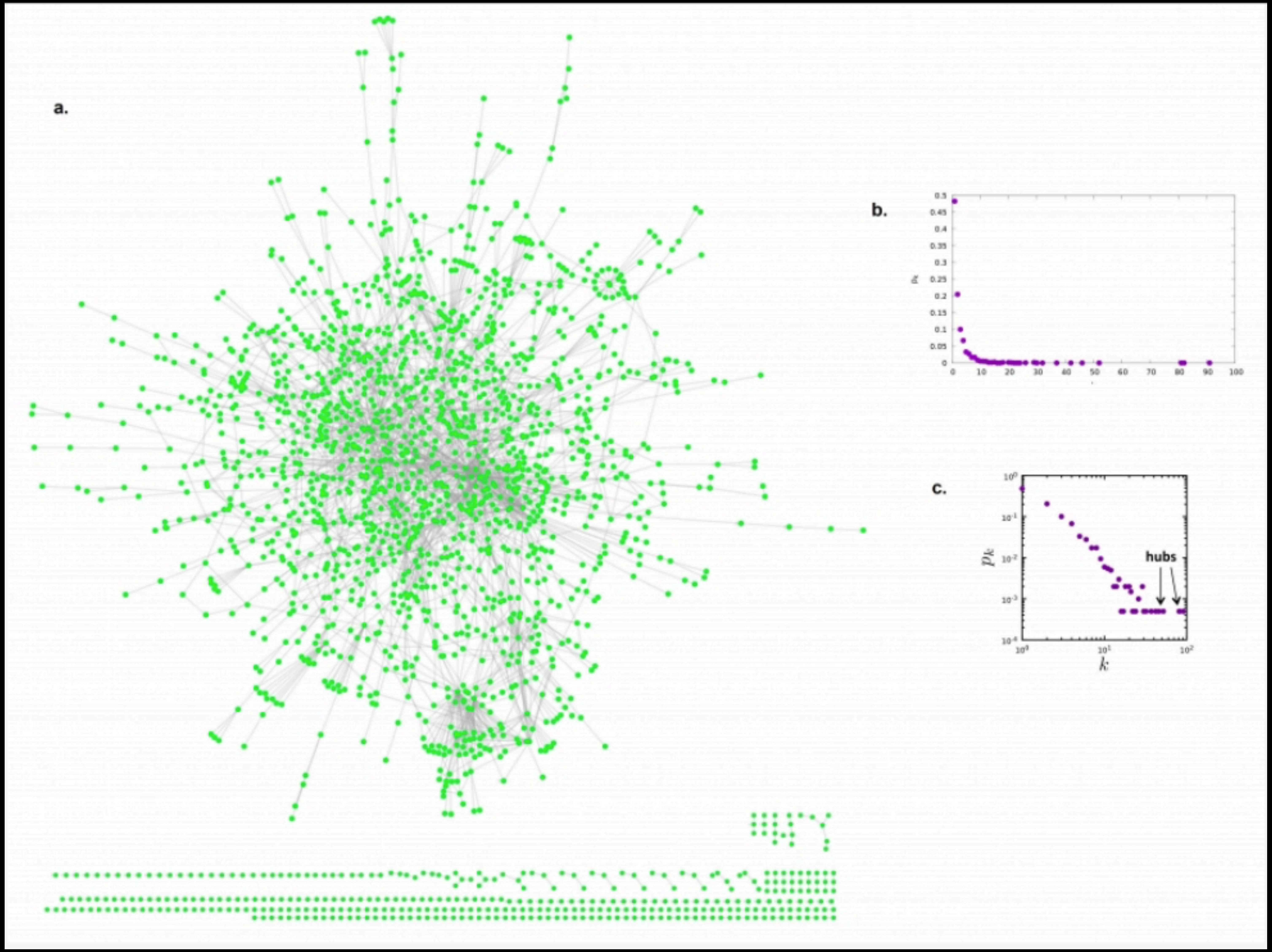
# Clustering Coefficient



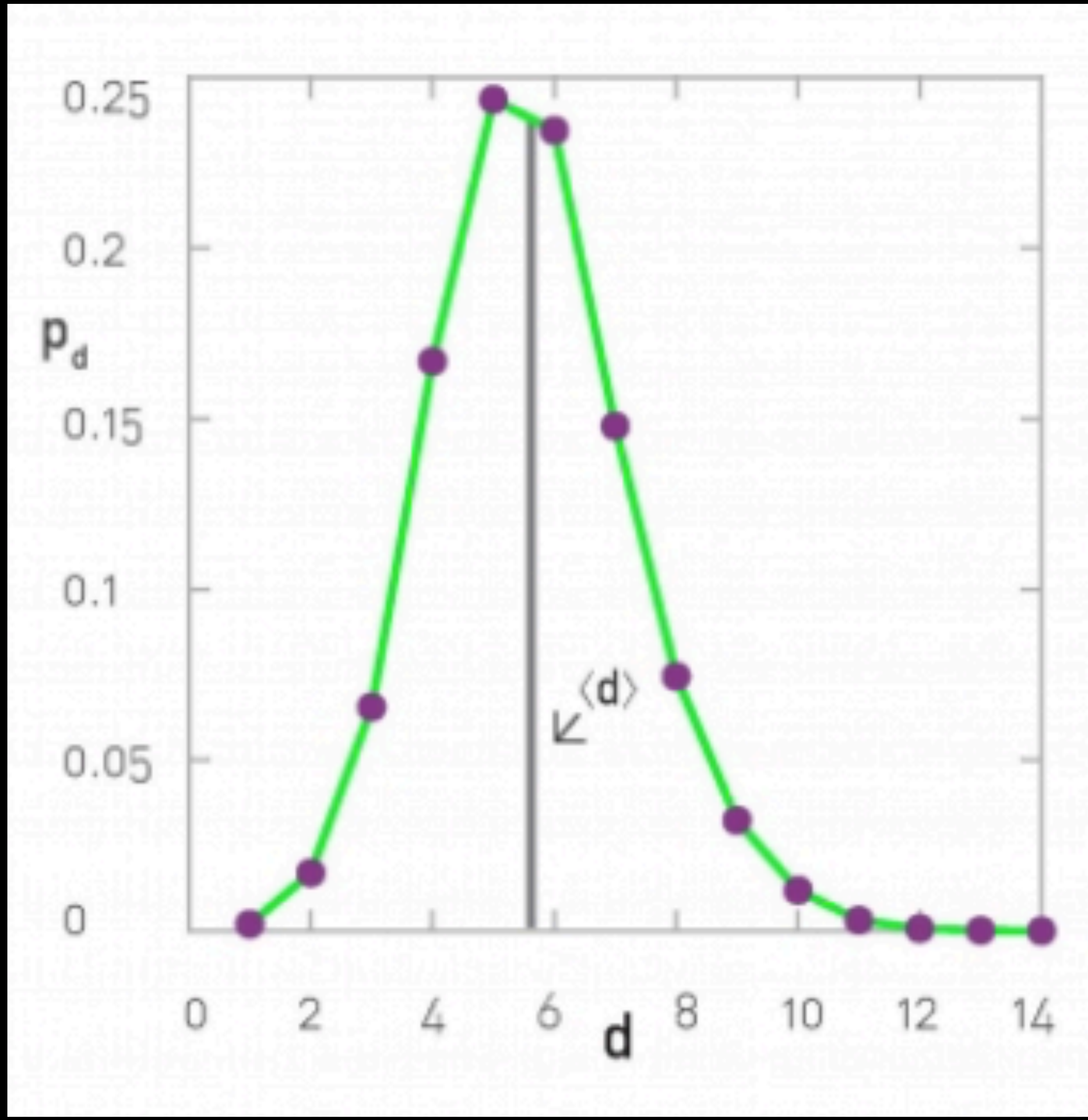

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

LNMIIT

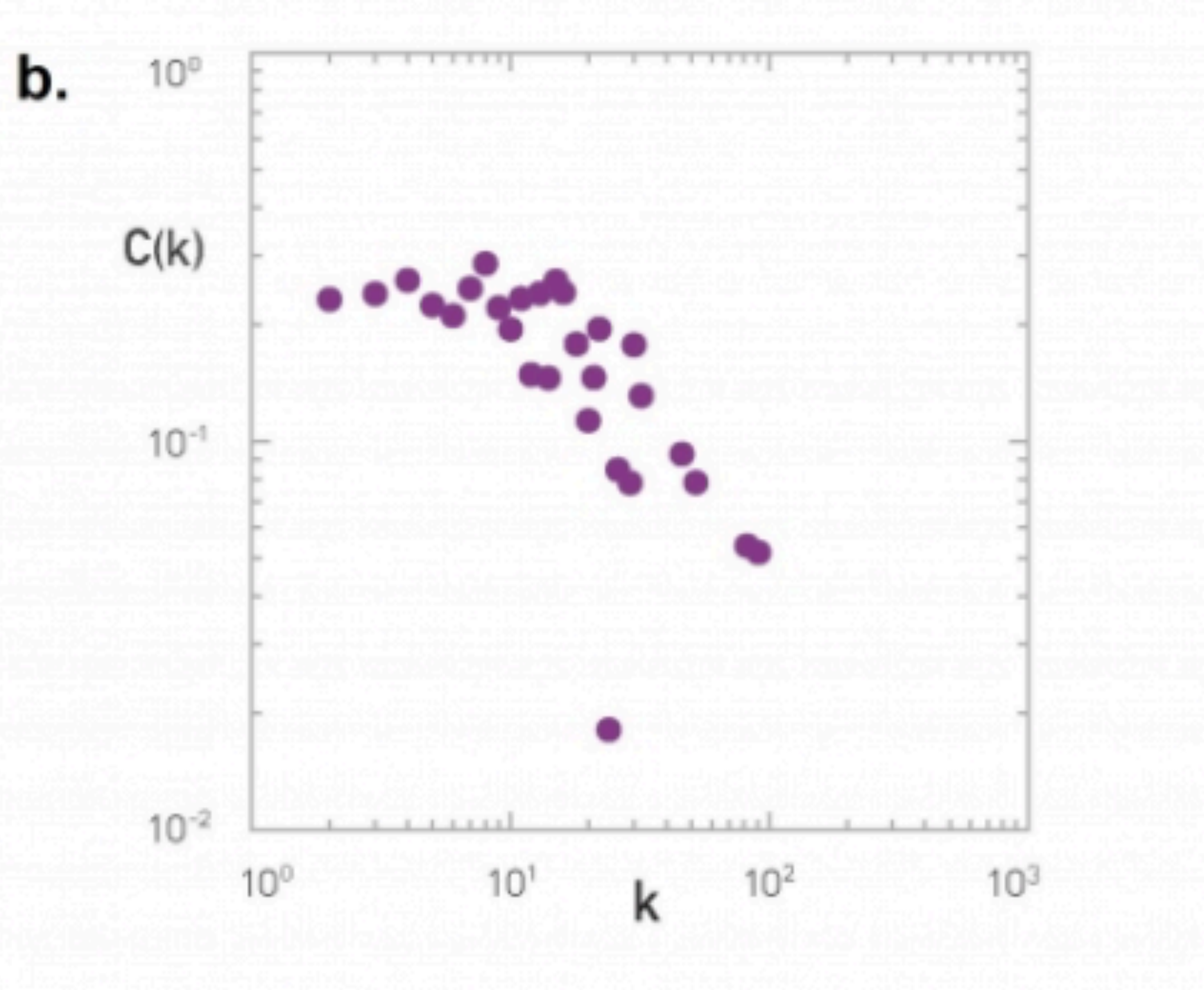
# A layout of the protein interaction network of yeast



# A layout of the protein protein interaction (PPI) network of yeast




The distance distribution $p_d$ for the PPI network




Plot for *C(k)* versus *k*

LNMIIT

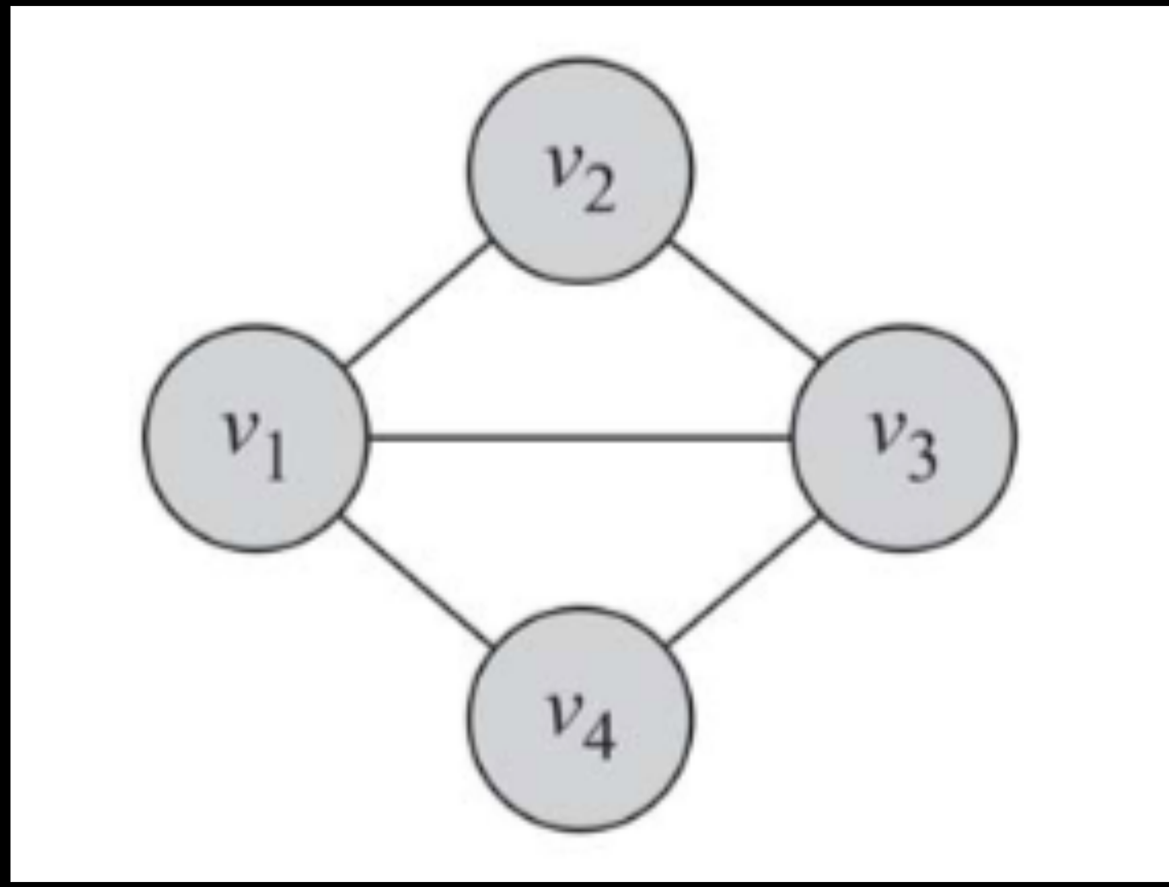
# Clustering Coefficient

- Network's global clustering is defined as

  $C_{\Delta}$ = 3 x number of triangles / number of connected triples

- Global clustering measures the transitivity in the graph

- Transitivity - *A friend of my friend is my friend*




**Note that Global clustering coefficient and average clustering coefficient are different!**

# Reciprocity

- Closed loops of length 2 that can only happen in a directed graph
- *If you are my friend, me too*

$$R = \frac{\sum_{i,j,i<j} A_{i,j} A_{j,i}}{m/2}$$

$$R = \frac{2}{m} \times \frac{1}{2} Tr(A^2)$$

$$R = \frac{2}{m} \sum_{i,j,i<j} A_{i,j} A_{j,i}$$

$$R = \frac{1}{m} \times Tr(A^2)$$

LNMIIT

# Balance and Status

**Balance**

- A signed that represents qualitative relationships of nodes in a social network such as friends or foes
  - a positive edge from node $v_1$ to $v_2$ denotes that $v_1$ considers $v_2$ as a friend and a negative edge denotes that $v_1$ assumes $v_2$ is an enemy.

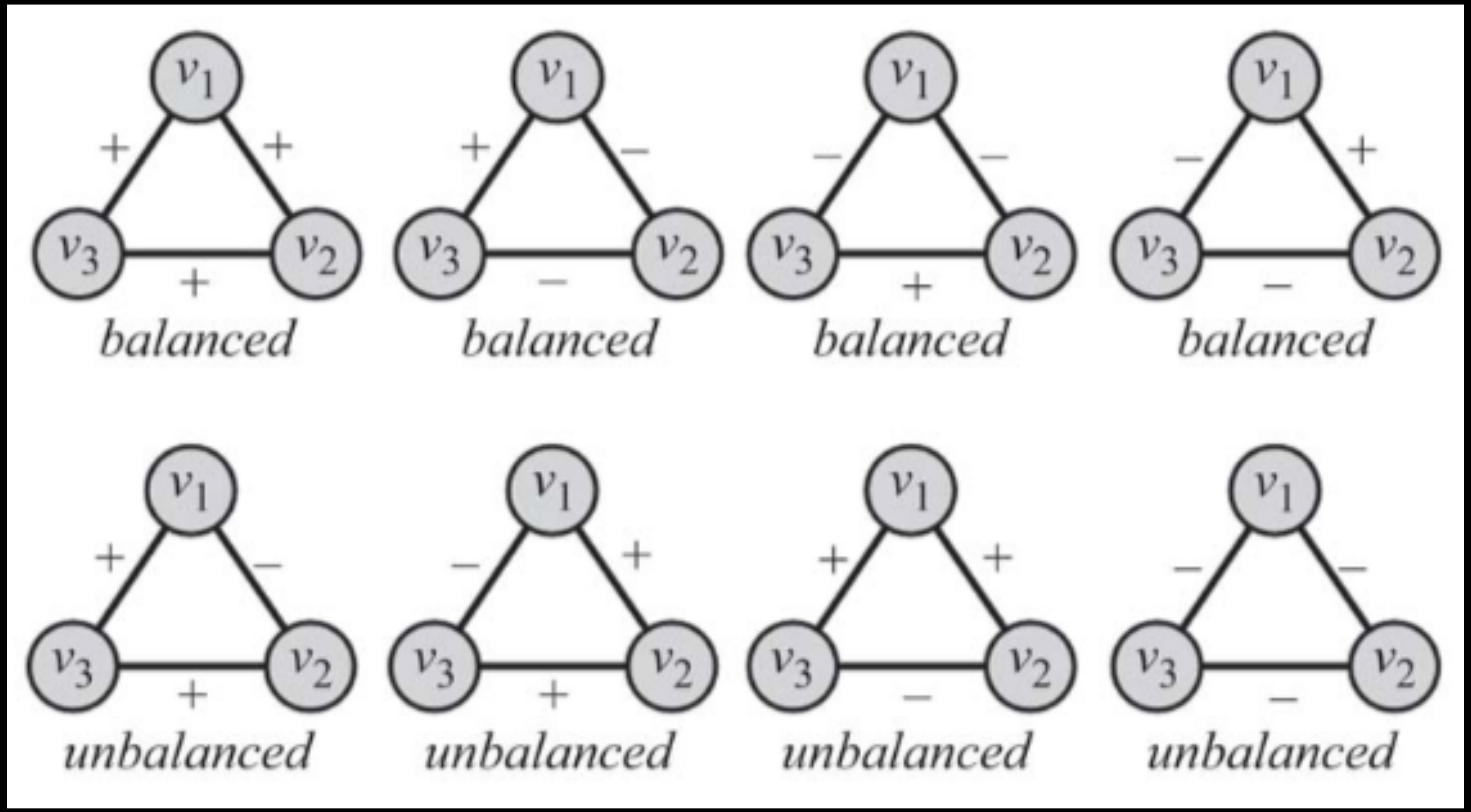
**Status**

- A directed graph that represents the status of nodes — a directed edge from $v_1$ to $v_2$ says that $v_1$ consider $v_2$'s status higher than its own.

LNMIIT

# Social Balance

- This discusses consistency in friend/foe relationships among individuals in a network
- Social balance theory says friend/foe relationships are consistent or not
- How many (in)consistent triads are there? Is the whole network consistent or not?




**The friend of my friend is my friend**

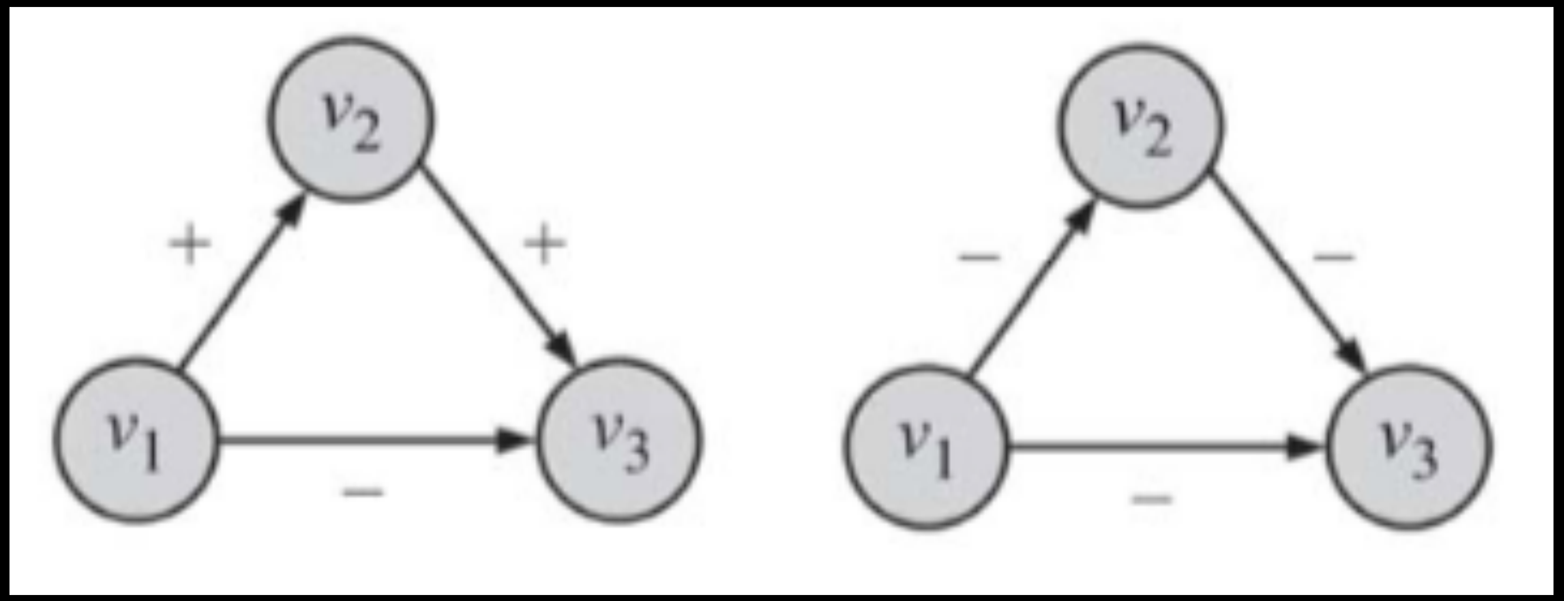
**The friend of my enemy is my enemy**

**The enemy of my enemy is my friend**

**The enemy of my friend is my enemy.**

# Social Status Theory

- If *X* has a higher status than *Y* and *Y* has a higher status than *Z*, then *X* should have a higher status than *Z*




**In a cycle of n nodes, where n – 1 consecutive edges are positive and the last edge is negative, social status theory considers the cycle balanced**

# Similarity

- Network Similarity
    - Structural equivalence
    - Regular equivalence
- Content Similarity

# Structural Equivalence

- Look at the neighborhood shared by two nodes
  - The size of this neighborhood defines how similar two nodes are
  - For instance, two brothers have in common sisters, mother, father, grandparents, and so on

$$\sigma_{jaccard}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|}$$

$$\sigma(v_i, v_j) = |N(v_i) \cap N(v_j)| \rightarrow$$

$$\sigma_{cosine}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)||N(v_j)|}}$$

# Structural Equivalence

**A more interesting way of measuring the similarity between vi and vj is to compare σ(vi, vj) with the expected value of σ(vi, vj) when nodes pick their neighbors at random.**

$$\sigma_{significance}(v_i, v_j) = \sum_k A_{i,k}A_{j,k} - \frac{d_i d_j}{n}$$

$$= \sum_k A_{i,k}A_{j,k} - n\frac{1}{n}\sum_k A_{i,k}\frac{1}{n}\sum_k A_{j,k}$$

$$= \sum_k A_{i,k}A_{j,k} - n\bar{A}_i\bar{A}_j$$

$$= \sum_k (A_{i,k}A_{j,k} - \bar{A}_i\bar{A}_j)$$

$$= \sum_k (A_{i,k}A_{j,k} - \bar{A}_i\bar{A}_j - \bar{A}_i\bar{A}_j + \bar{A}_i\bar{A}_j)$$

$$= \sum_k (A_{i,k}A_{j,k} - A_{i,k}\bar{A}_j - \bar{A}_i A_{j,k} + \bar{A}_i\bar{A}_j)$$

$$= \sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j)$$

Covariance bet $A_i$ and $A_j$

$$\sigma_{pearson}(v_i, v_j) = \frac{\sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j)}{\sqrt{\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\sum_k (A_{j,k} - \bar{A}_j)^2}}$$

LNMIIT

# Regular Equivalence

LNMIIT

# Ties and Triadic Closures

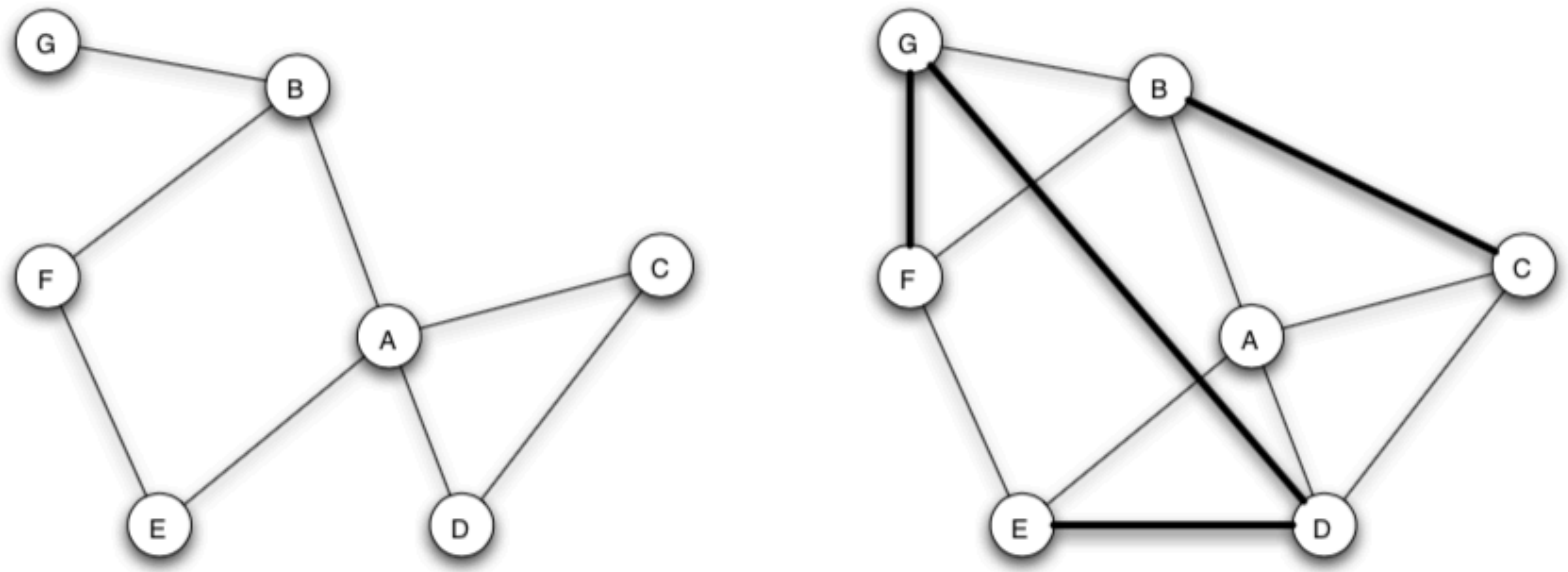
- Mark Granovetter in late 1960s found acquaintances bring new ideas and concepts than close friends!

- This finding links two different perspectives on distant friendships — one structural and other interpersonal

- 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

# Triadic Closure



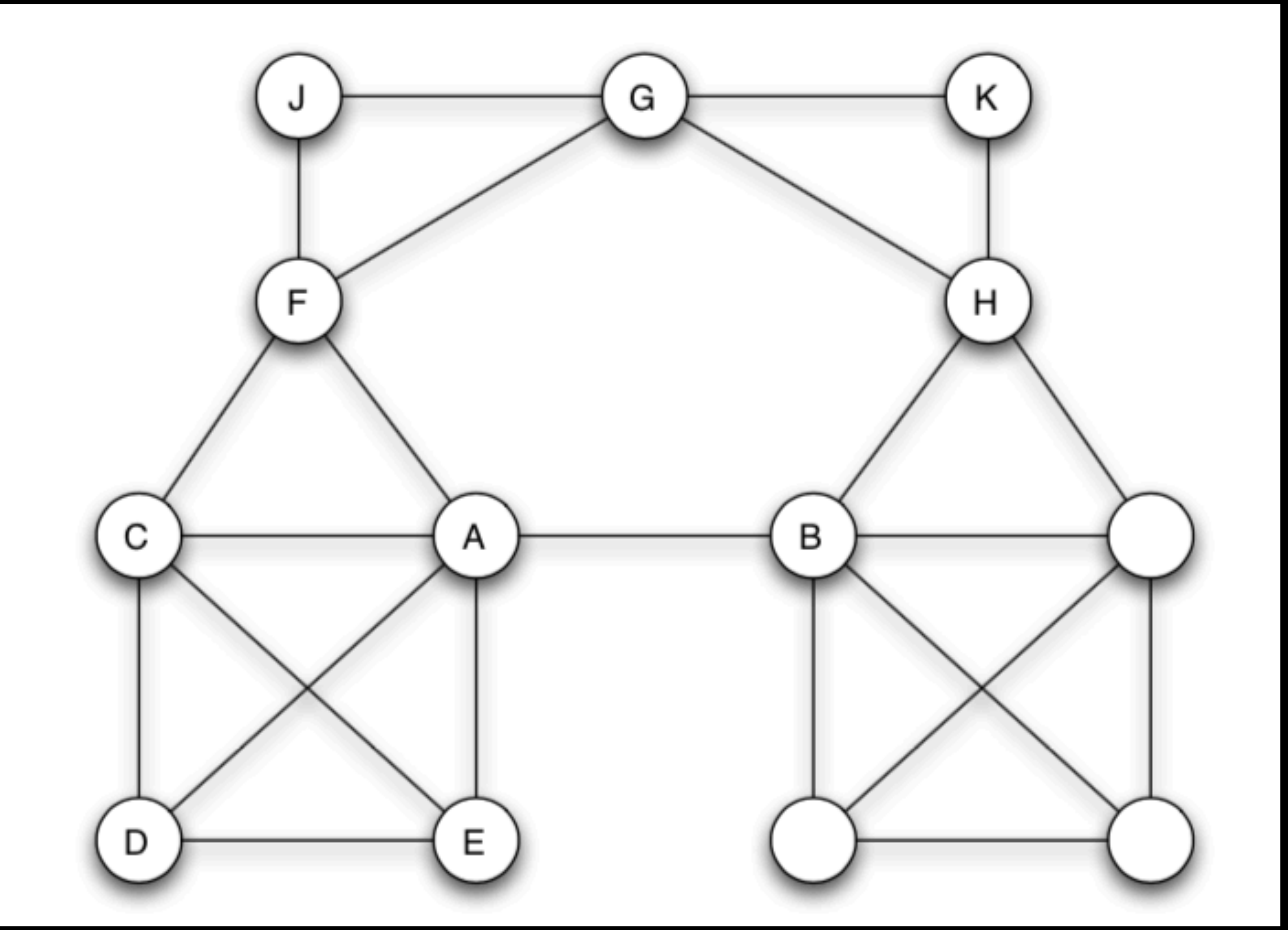

(a) *Before new edges form.*

(b) *After new edges form.*

# Reasons for Triadic Closure

- Reasons for Triadic Closure
    - Opportunity a common friend A gives for friends B and C to meet and make friendship
    - B and C have a common friend A so trust factor is taken care
    - Incentive that A brings — latent stress if B and C are not friends

# Bridges and Local Bridges

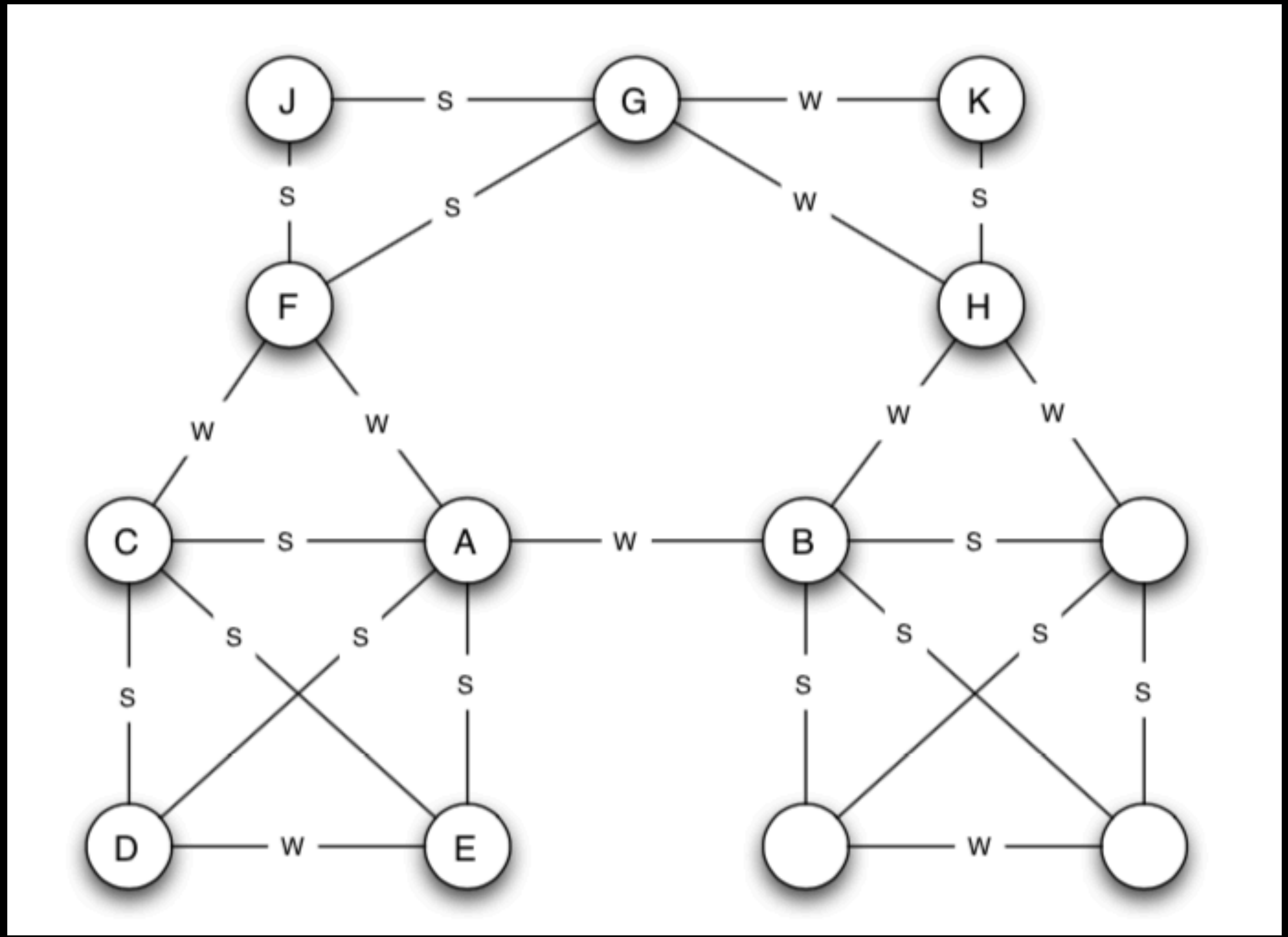



- We say that an edge joining two nodes A and B in a graph is a bridge if deleting the edge would cause A and B to lie in two different components.
- But bridges are presumably extremely rare in real social networks.
- We say that an edge joining two nodes A and B in a graph is a local bridge if its endpoints A and B have no friends in common — in other words, if deleting the edge would increase the distance between A and B to a value strictly more than two.

LNMIIT

# Strong Triadic Closure Property

- Two types of ties
  - strong ties (the stronger links, corresponding to friends)
  - weak ties (the weaker links, corresponding to acquaintances)




LNMIIT

# Strong Triadic Closure Property

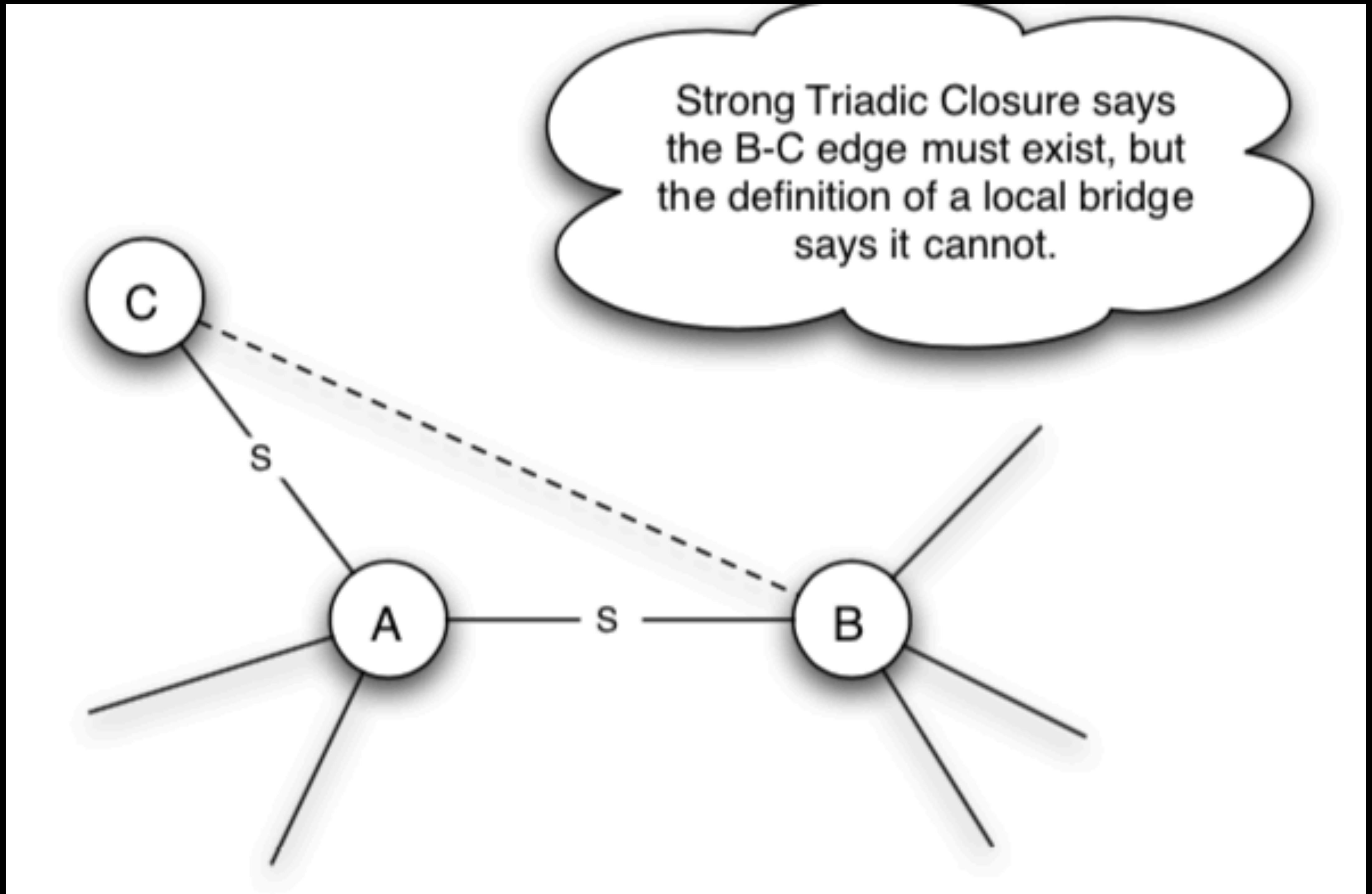
- If a node A has edges to nodes B and C, then the B-C edge is especially likely to form if A’s edges to B and C are both strong ties.
- More stronger property: We say that a node A violates the Strong Triadic Closure Property if it has strong ties to two other nodes B and C, and there is no edge at all (either a strong or weak tie) between B and C. We say that a node A satisfies the Strong Triadic Closure Property if it does not violate it.
- Strong Triadic Closure Property is too extreme for us to expect it hold across all nodes of a large social network

# Local Bridges and Weak Ties

- Connect a local (weak tie) and a global property (local bridge) — can we?
  - Claim: If a node A in a network satifies the Strong Triadic Closure Property and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie.

LNMIIT

# Local Bridges and Weak Ties



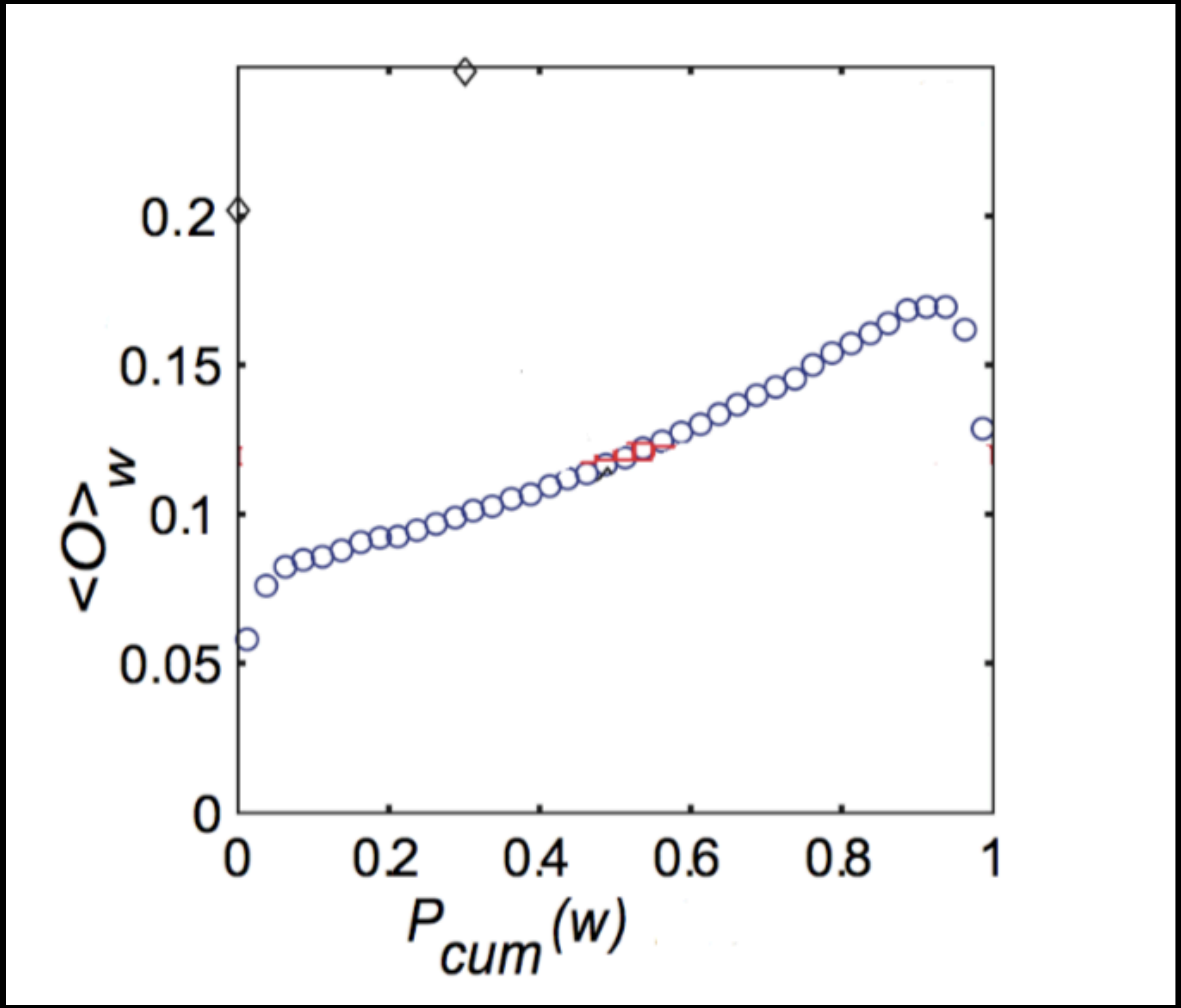

- consider a node A that satisfies the Strong Triadic Closure Property and is involved in at least two strong ties
- suppose A is involved in a local bridge — say, to a node B — that is a strong tie
- since A is involved in at least two strong ties, and the edge to B is only one of them, it must have a strong tie to some other node, which we'll call C.
- Since the edge from A to B is a local bridge, A and B must have no friends in common, and so the B-C edge must not exist. But this contradicts Strong Triadic Closure
- This implies a local bridge has to be a weak tie

LNMIIT

# Generalizing the Notions of Weak Ties and Local Bridges

- We define the neighborhood overlap of an edge connecting A and B to be the ratio
  - Let number of nodes who are neighbors of both A and B = n
  - Let number of nodes who are neighbors of at least one of A or B = m
  - Measure of neighbourhood ratio = n/m
- What will be the neighbourhood ratio for a bridge?
- The edges with very small neighborhood overlap are called as “almost” local bridges.
-

LNMIIT




Tie Strength and Neighborhood Overlap

**Empirical results show the validity of theoretical formulations**

LNMIIT

# Empirical Results

- Weak ties serve to link together different tightly-knit communities that each contain a large number of stronger ties.

- Onnela et al. provided an indirect analysis to address this question

  - They first deleted edges from the network one at a time, starting with the strongest ties and working downward in order of tie strength. The giant component shrank steadily as they did this, its size going down gradually due to the elimination of connections among the nodes.

  - They then tried the same thing, but starting from the weakest ties and working upward in order of tie strength. In this case, they found that the giant component shrank more rapidly, and moreover that its remnants broke apart abruptly once a critical number of weak ties had been removed.

  - This is consistent with a picture in which the weak ties provide the more crucial connective structure for holding together disparate communities, and for keeping the global structure of the giant component intact.

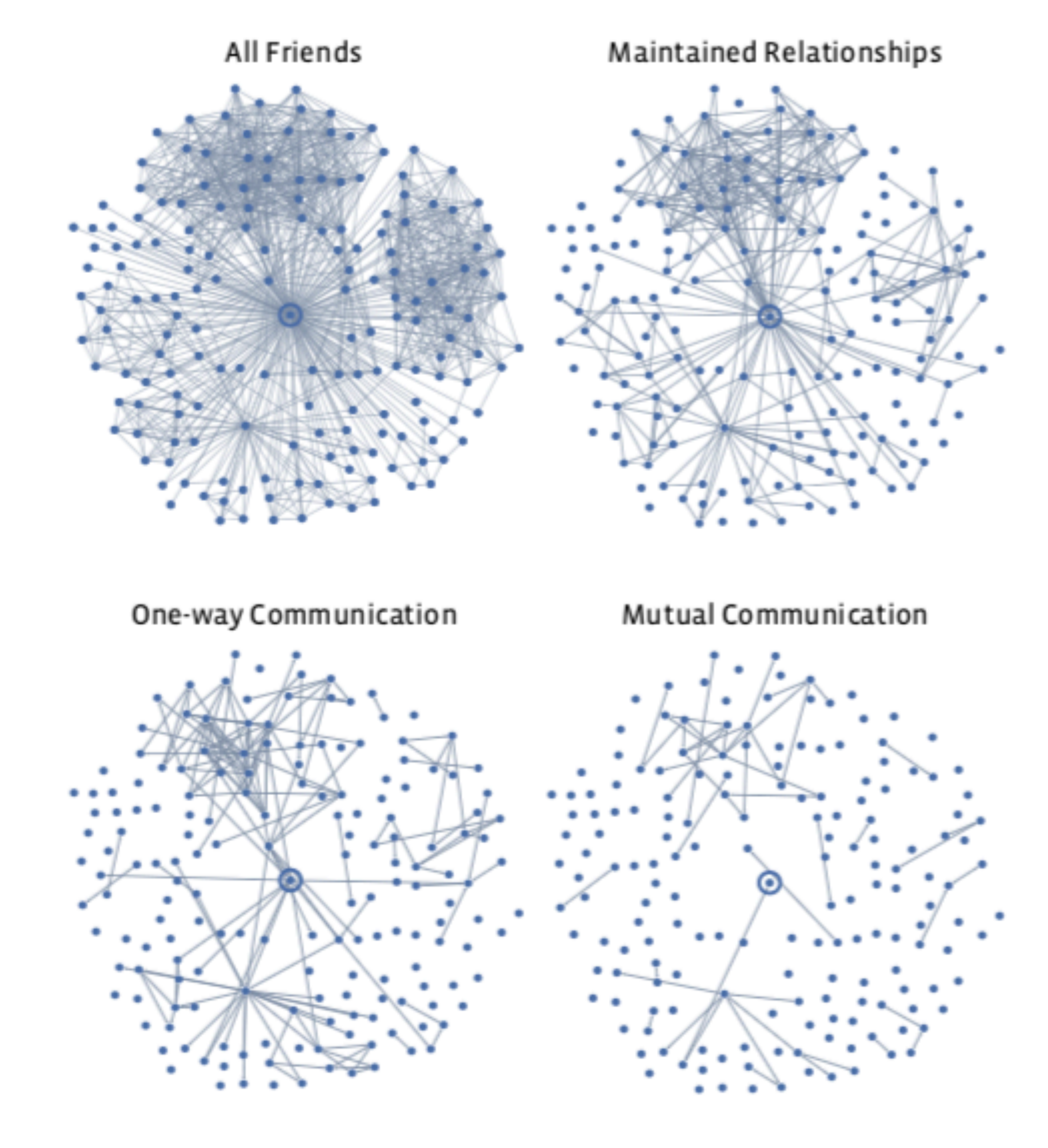
# Tie Strength on Facebook

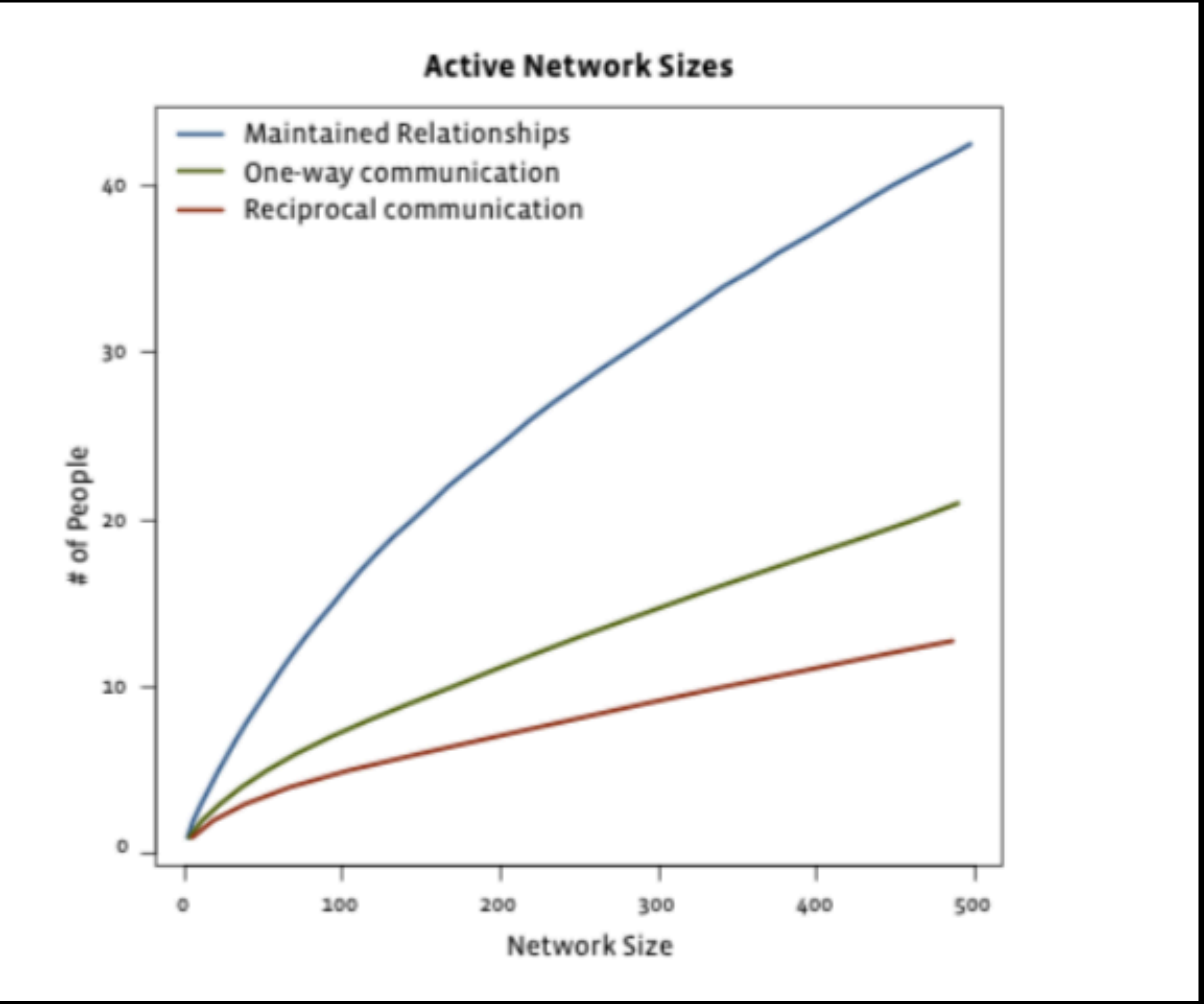
- At Facebook, Cameron Marlow and his colleagues analyzed the friendship links reported in each user's profile
    - what extent each link was actually used for social interaction, beyond simply being reported in the profile
    - where are the strong ties among a user's friends?

# Tie Strength on Facebook

- Three categories of links based on usage over a one-month observation period.
  - A link represents reciprocal (mutual) communication, if the user both sent messages during the observation period.
  - A link represents one-way communication if the user sent one or more messages to the friend.
  - A link represents a maintained relationship if the user followed information about the friend, whether or not actual communication took place.

LNMIIT


All Friends
Maintained Relationships
One-way Communication
Mutual Communication



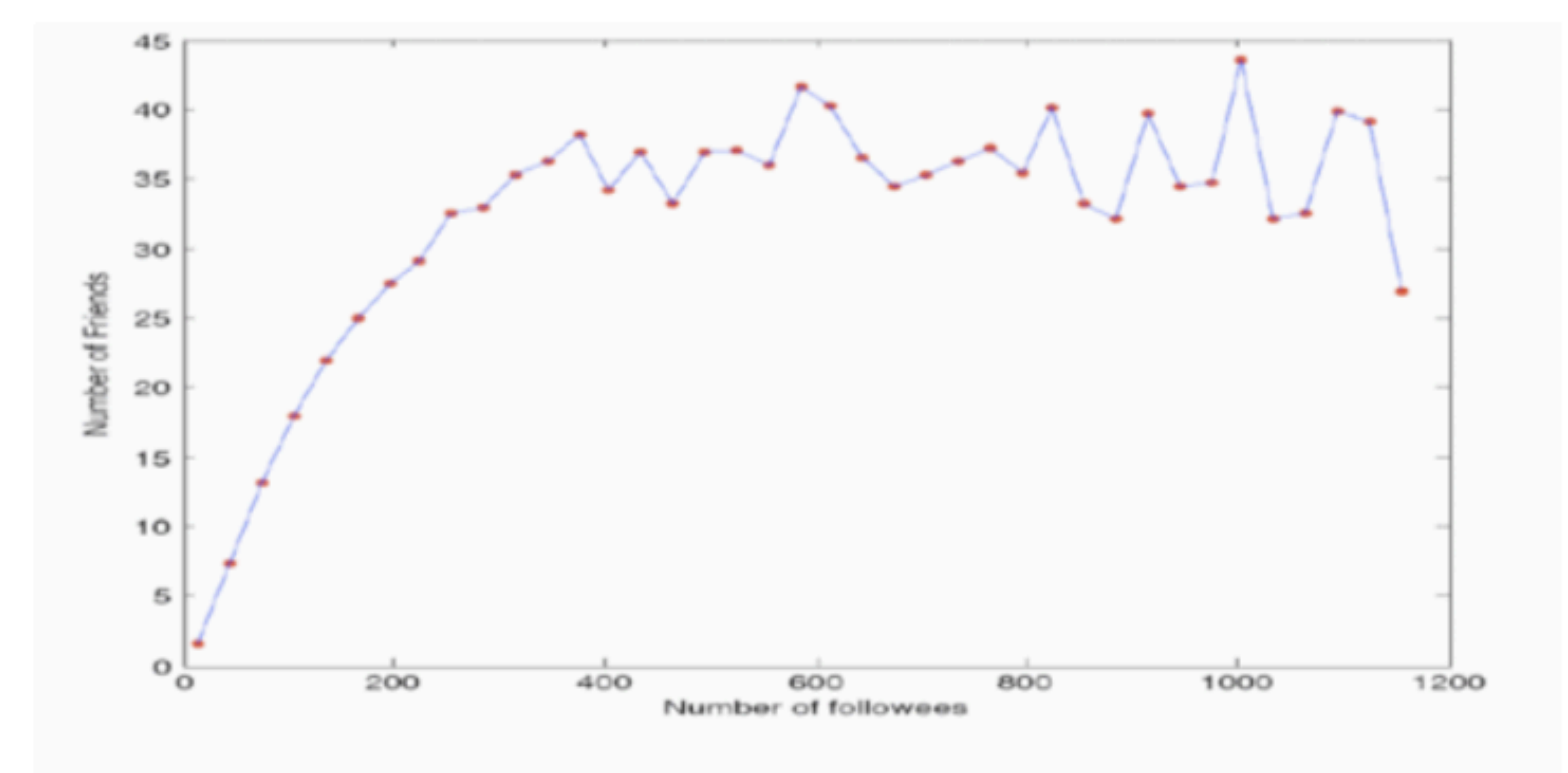
Active Network Sizes
Maintained Relationships
One-way communication
Reciprocal communication
# of People
Network Size
0
10
20
30
40
0
100
200
300
400
500


LNMIIT

# Tie Strength on Twitter

- Huberman, Romero, and Wu analyzed the relative abundance of these two kinds of links on Twitter

    - for each user find her “followees”, and then define her strong ties to consist of the users to whom she had directed at least two messages over the course of an observation period.

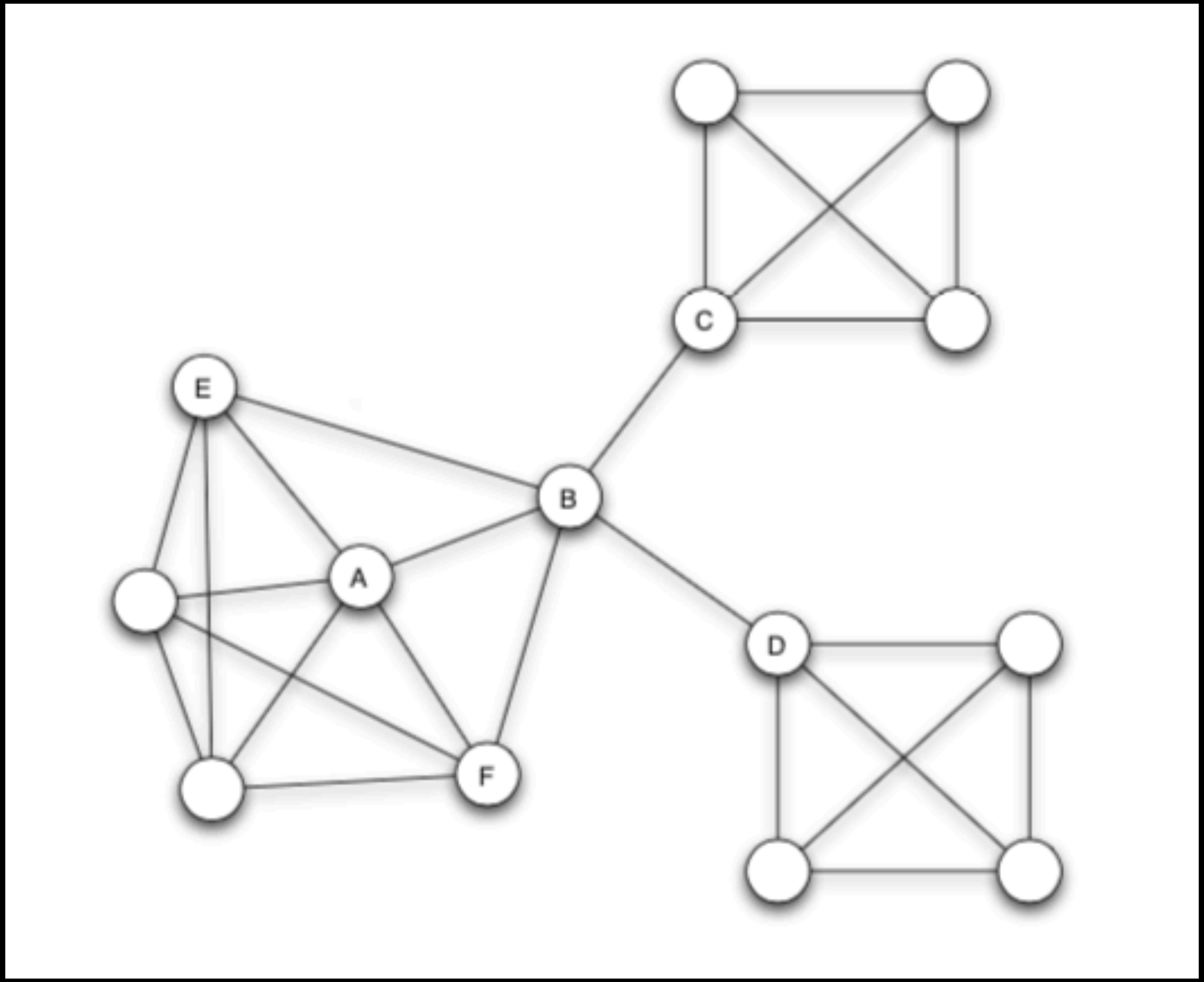



LNMIIT

# Explanation for Strong and Weak Ties

- Each strong tie requires the continuous investment of time and effort to maintain
    - So naturally it is restricted by the hours available in a day.
- The formation of weak ties is governed by much milder constraints
    - They need to be established at their outset but not necessarily maintained continuously.

# Embeddedness and Structural Holes




- A and B have two common neighbours
- So AB has embeddedness value 2
- What is the embeddedness value of BC?

B has early access to information originating in multiple, non-interacting parts of the network

B's position at the interface of three also gives her the opportunity for creating novel ideas by combining these disparate sources of information in new ways

B with a source of power in the organization. Certain people in this situation might try to prevent triangles from forming around the local bridges they're part of !

Preventing triangles means flow of info between groups takes more time to communicate!