Graph Theory and Algorithms

Ph.D. Course – Marco Viviani

Complex Networks
(June 18, 2021 / 15:00-17:00)

TABLE OF CONTENTS

1

Introduction

Basic Concepts

Interaction Networks

- **Interaction networks** are "complex" systems (from the Latin *cum* (together) - *plexus* (intertwined), "intertwined together").
  - A complex system is composed of several parts connected to each other and "intertwined" with each other so that;
  - the result is different from the sum of the parts.

- The behavior of a complex system cannot be deduced from the analysis, however accurate, of the elements that compose it: instead, it is necessary to observe the interactions between them.

Emerging Behaviors

- Simple entities interacting with each other and with the surrounding environment can in fact give rise to non-trivial macroscopic behaviors called **emerging behaviors**.

- Emerging behavior is a collective phenomenon: that is, it occurs **spontaneously** and not thanks to a centralized organization.

Network Topology

- The structure (**topology**) of the contact network is crucial in determining collective behavior.
- The study of the **topological properties** of networks allows us to understand:
  - how these properties affect the emerging behaviors and dynamics of complex systems;
  - how the network of interactions itself can in turn modify and readjust itself in an adaptive way.

Consequences related to the topology
(Example)

- The outcome of a **spreading dynamic of a pathogen** is heavily influenced by the interaction network of the population in which the infectious agent spreads.

- The specific characteristics of social networks, especially the **high heterogeneity**, strongly strengthen the incidence of infection, and radically change the epidemiological framework compared to that classically adopted in describing the spread of diseases.

2

Regular and Random Networks

Examples of Regular and ER Networks

Regular Networks and Random Networks

- Classical graph theory, prior to that on complex networks, mainly deals with **two types of networks**:
  - Regular networks
  - Random networks
    - Halfway between graph theory and probability theory

Regular Networks

- In **regular networks** each node is connected to a fixed number of nodes.

- They have **regular patterns** within the structure.
  - They may or may not be k-regular graphs;
  - Entropy is 0 or very close to zero.
    - By entropy we mean the degree of randomness in a graph structure.

Regular Networks ... Cont'd

- Some of these networks are characterized by **strong aggregation**: nodes connected to a given node tend to be connected to each other.
- In other words, there is a high local density of connections, measured by the **clustering coefficient**.

Clustering Coefficient

- The **clustering** (or **aggregation**) **coefficient** is the measure of the degree to which the nodes of a graph tend to be connected to each other.

- Three possibilities to calculate the clustering coefficient:
  - Local clustering coefficient.
  - Average clustering coefficient.
  - Global clustering coefficient.

Local Clustering Coefficient

- Given $N(v)$ the set of neighbors of v, the **local clustering coefficient** $cc(v)$ of a vertex v is given by the number of edges between the members of $N(v)$ divided by the number of potential edges between them.

- **Directed graph**:

$$cc(v) = \frac{||N(v)||}{k(k-1)}$$

Maximum number of potential edges between the vertices in $N(v)$ in a directed graph

$$k = |N(v)| = d(v)$$

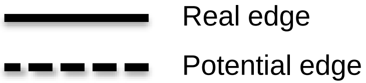
- **Undirected graph**:

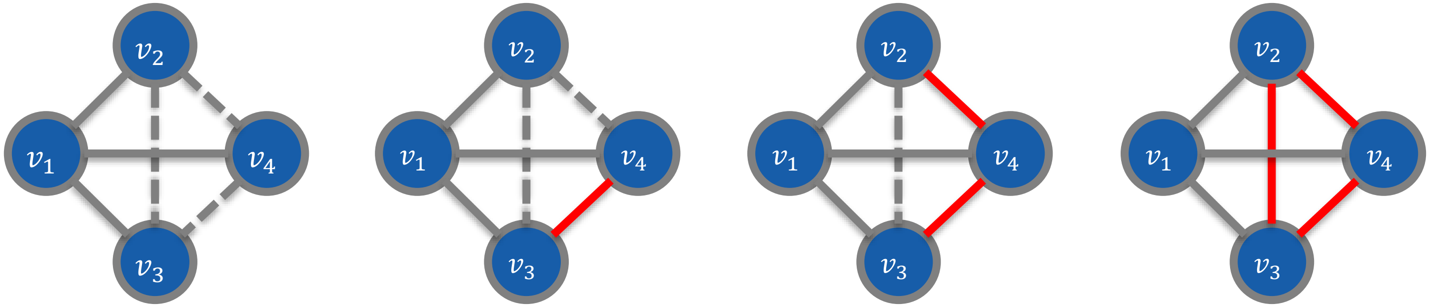
$$cc(v) = \frac{2||N(v)||}{k(k-1)}$$

In an undirected graph the maximum number of potential edges between the neighbors of v is $\frac{k(k-1)}{2}$

Local Clustering Coefficient (Examples)

$$cc(v) = \frac{2||N(v)||}{k(k-1)}$$

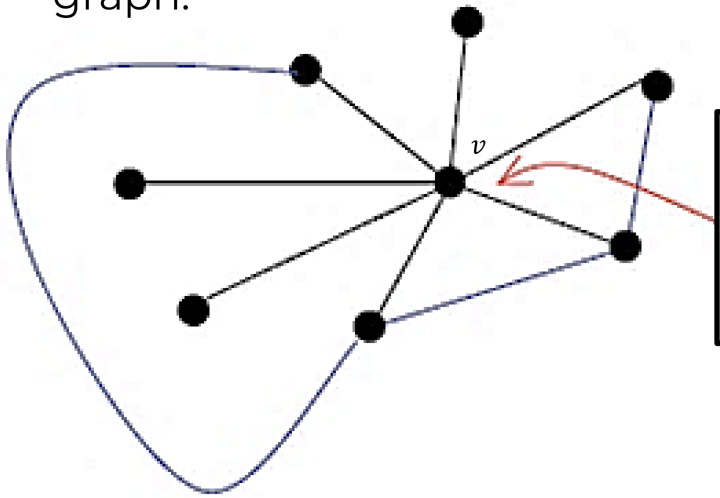




$$cc(v_1) = \frac{2*0}{3*2} = \frac{0}{6} = 0 \qquad cc(v_1) = \frac{2*1}{3*2} = \frac{1}{3} \qquad cc(v_1) = \frac{2}{3} \qquad cc(v_1) = \frac{6}{6} = 1$$

Exercise

- Calculate the local clustering coefficient of node v in the following graph:



$$cc(v) = \frac{2||N(v)||}{k(k-1)} = ?$$

$$cc(v) = \frac{2 * 3}{7 * 6} = \frac{6}{42} = \frac{1}{7} = 0{,}14$$

Average Clustering Coefficient

- The **average clustering coefficient** $cc(G)$ of a graph G is given by the average of the clustering coefficients for each single node of the graph.

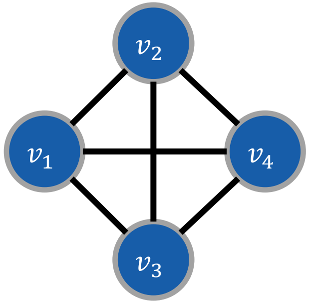
- Formally:

$$cc(G) = \frac{1}{|V|} \sum_{i=1}^{n} cc(v_i)$$

Average Clustering Coefficient (Examples)

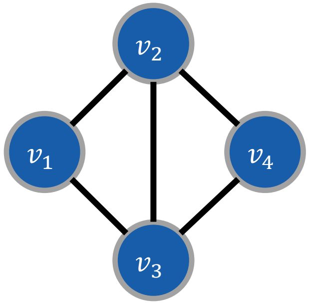
- $cc(G_1) = \frac{1}{4}(1 + 1 + 1 + 1) = 1$

- $cc(G_2) = \frac{1}{4}\left(1 + \frac{2}{3} + \frac{2}{3} + 1\right) = \frac{5}{6} = 0{,}8\overline{3}$
  - $cc(v_1) = \frac{2*1}{2*1} = 1$
  - $cc(v_2) = \frac{2*2}{3*2} = 2/3$
  - $cc(v_3) = \frac{2*2}{3*2} = 2/3$
  - $cc(v_4) = \frac{2*1}{2*1} = 1$



G_1

$$cc(G) = \frac{1}{|V|}\sum_{i=1}^{n} cc(v_i)$$



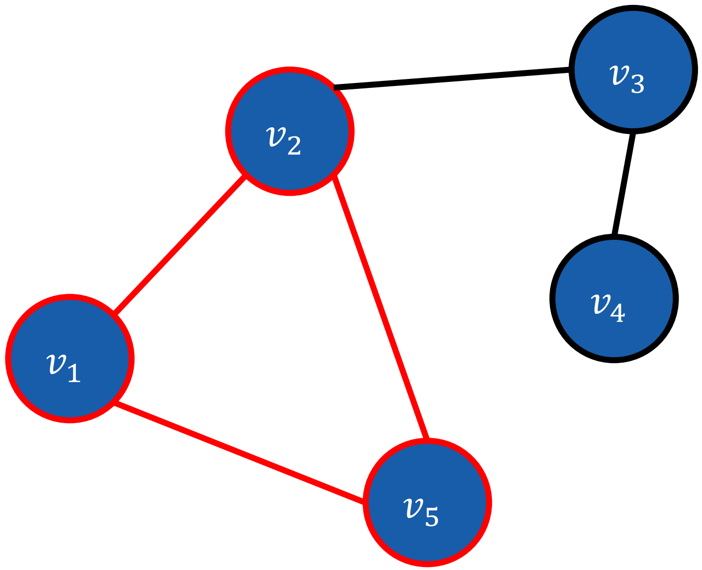
G_2

Global Clustering Coefficient

- The concept of **global clustering coefficient** (a.k.a. **transitivity**) is based on triples (triads) of vertices.
  - Open triplet: three nodes connected by two edges.
  - Closed triplet: three nodes connected by three edges.
- Each triple is **centered** around a vertex.
- A **triangle** consists of three closed triples centered on the same three nodes that compose them.

Triangle (Example)

Vertex	Triplets centered around the vertex
v_1	$\langle v_1, v_2, v_5 \rangle$
v_2	$\langle v_1, v_2, v_3 \rangle$ $\langle v_1, v_2, v_5 \rangle$ $\langle v_2, v_3, v_5 \rangle$
v_3	$\langle v_2, v_3, v_4 \rangle$
v_4	—
v_5	$\langle v_1, v_2, v_5 \rangle$



Global Clustering Coefficient / Definition

- The **global clustering coefficient** $cc_{\triangle}(G)$ of a graph G is calculated as the number of closed triples (or 3 times the number of triangles) divided by the total number of triples (open and closed ones)
- Formally:

$$cc_{\triangle}(G) = \frac{3 * n_{\triangle}(G)}{n_{\wedge}(G)} = \frac{\sum_{i=1}^{n}(cc(v_i) * \omega_i)}{\sum_{i=1}^{n} \omega_i}$$

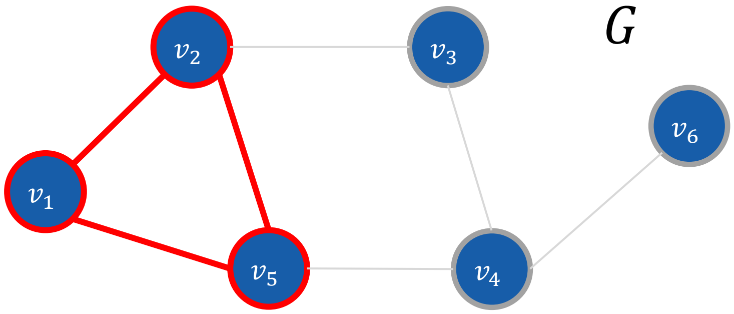
Number of triangles in the graph

Total number of triples (open and closed) in the graph

Number of triples in which the node v_i is central («weight» of the node v_i)

Global Clustering Coefficient (Example – 1)

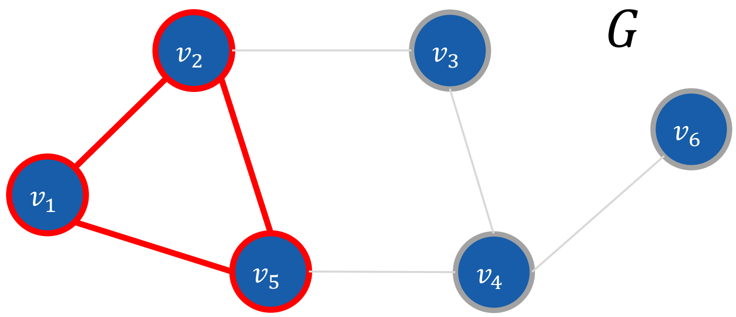
Vertex	Triplets centered around the vertex	Weight (ω_i)
v_1	$\langle v_1, v_2, v_5 \rangle$	1
v_2	$\langle v_1, v_2, v_3 \rangle$ $\langle v_1, v_2, v_5 \rangle$ $\langle v_2, v_3, v_5 \rangle$	3
v_3	$\langle v_2, v_3, v_4 \rangle$	1
v_4	$\langle v_3, v_4, v_5 \rangle$ $\langle v_3, v_4, v_6 \rangle$ $\langle v_4, v_5, v_6 \rangle$	3
v_5	$\langle v_1, v_2, v_5 \rangle$ $\langle v_1, v_4, v_5 \rangle$ $\langle v_2, v_4, v_5 \rangle$	3
v_6	—	0



$$cc_{\triangle}(G) = \frac{3 * n_{\triangle}(G)}{n_{\wedge}(G)} = \frac{3 * 1}{11} = \frac{3}{11}$$

Global clustering coefficient (Example – 2)

Vertex	Weight (ω_i)	$cc(v_i)$
v_1	1	$2 * 1/2 * 1 = 1$
v_2	3	$2 * 1/3 * 2 = 1/3$
v_3	1	$2 * 0/2 * 1 = 0$
v_4	3	$2 * 0/3 * 2 = 0$
v_5	3	$2 * 1/3 * 2 = 1/3$
v_6	0	0



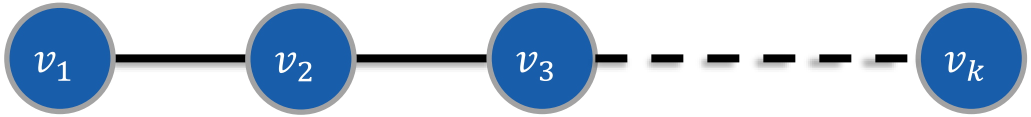
$$cc(v) = \frac{2||N(v)||}{k(k-1)}$$

$$cc_{\triangle}(G) = \frac{\sum_{i=1}^{n}(cc(v_i) * \omega_i)}{\sum_{i=1}^{n} \omega_i} = \frac{(1 * 1) + \left(\frac{1}{3} * 3\right) + (0 * 1) + (0 * 3) + \left(\frac{1}{3} * 3\right) + 0}{11} = \frac{3}{11}$$

Examples of Regular Networks

Linear Network (1)

- A **linear network** is a linear sequence L of connected vertices:
- $L = v_1, e_{12}, v_2, e_{23}, v_3, \dots, v_{k-1}, e_{(k-1)k} v_k$



- Order: $|L| = k$
- Size: $||L|| = k - 1$
- Degree: $1 \leq d(L) \leq 2$
- Diameter: $diam(L) = k - 1$
- Clustering coefficient: $cc(L) = 0$

Examples of Regular Networks

Linear Network (2)

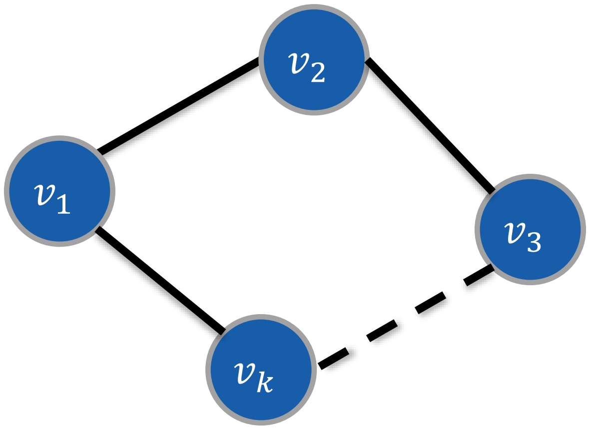
- Degree of connectivity: not biconnected
- Adjacency matrix: diagonals

0	1	0	0	…	0	0
1	0	1	0	…	0	0
0	1	0	1	…	0	0
0	0	1	0	…	0	0
…	…	…	…	…	…	…
0	0	0	0	…	0	1
0	0	0	0	…	1	0

Examples of Regular Networks

Ring Network (1)

- A **ring network** A is a network topology in which each node is connected to exactly two other nodes, forming a single continuous path, a ring



$$A = v_1, e_{12}, v_2, e_{23}, v_3, \ldots, v_k, e_{k1}, v_1$$

Examples of Regular Networks

Ring Network (2)

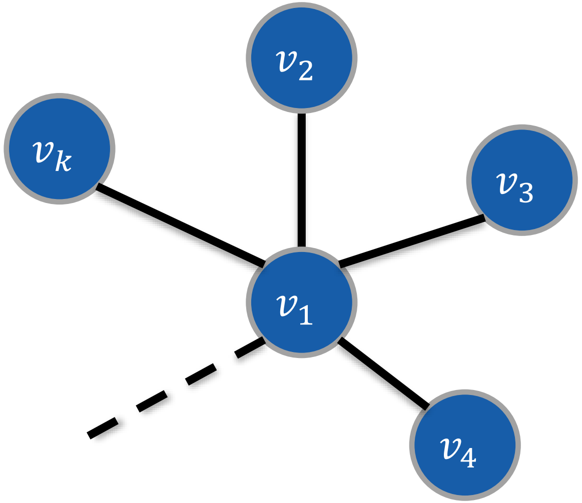
- Order: $|A| = k$
- Size: $||A|| = k$
- Degree: $d(A) = 2$
- Diameter: $diam(A) = k/2$
- Clustering coefficient: $cc(A) = 0$
- Degree of connectivity: biconnected
- Adjacency matrix: diagonals + angles

0	1	0	0	…	0	1
1	0	1	0	…	0	0
0	1	0	1	…	0	0
0	0	1	0	…	0	0
…	…	…	…	…	…	…
0	0	0	0	…	0	1
1	0	0	0	…	1	0

Examples of Regular Networks

Star Network (1)

- A **star network** is a tree S with a single vertex of maximum degree



$$S = v_1, e_{12}, v_2, v_1, e_{13}, v_3, \dots, v_1, e_{1k}, v_k$$

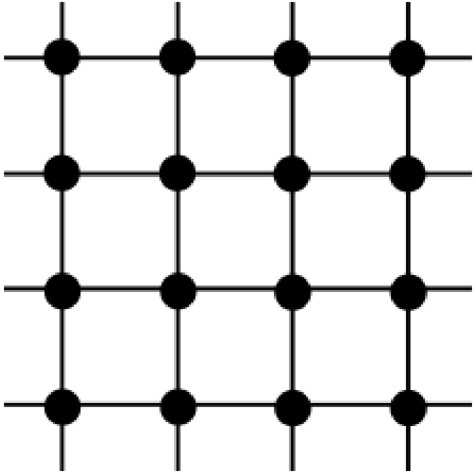
Examples of Regular Networks
Star Network (2)

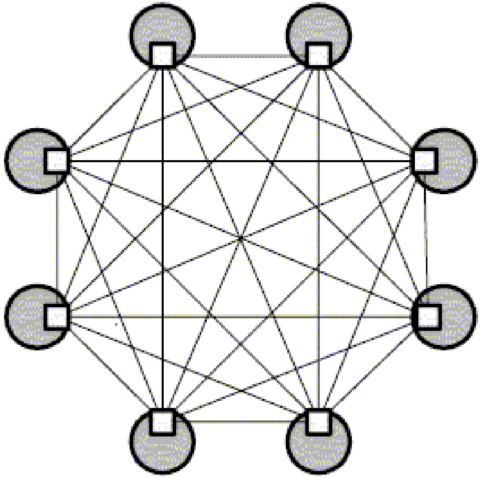
- Order: $|S| = k$
- Size: $||S|| = k - 1$
- Degree: $d(S) = 1$ or $d(S) = k - 1$
- Diameter: $diam(S) = 2$
- Clustering coefficient: $cc(S) = 0$
- Degree of connectivity: point of articulation on internal vertex
- Adjacency matrix: first row and column

0	**1**	**1**	**1**	…	**1**	**1**
1	0	0	0	…	0	0
1	0	0	0	…	0	0
1	0		0	…	0	0
…	…	…	…	…	…	…
1	0	0	0	…	0	0
1	0	0	0	…	0	0

Examples of regular networks

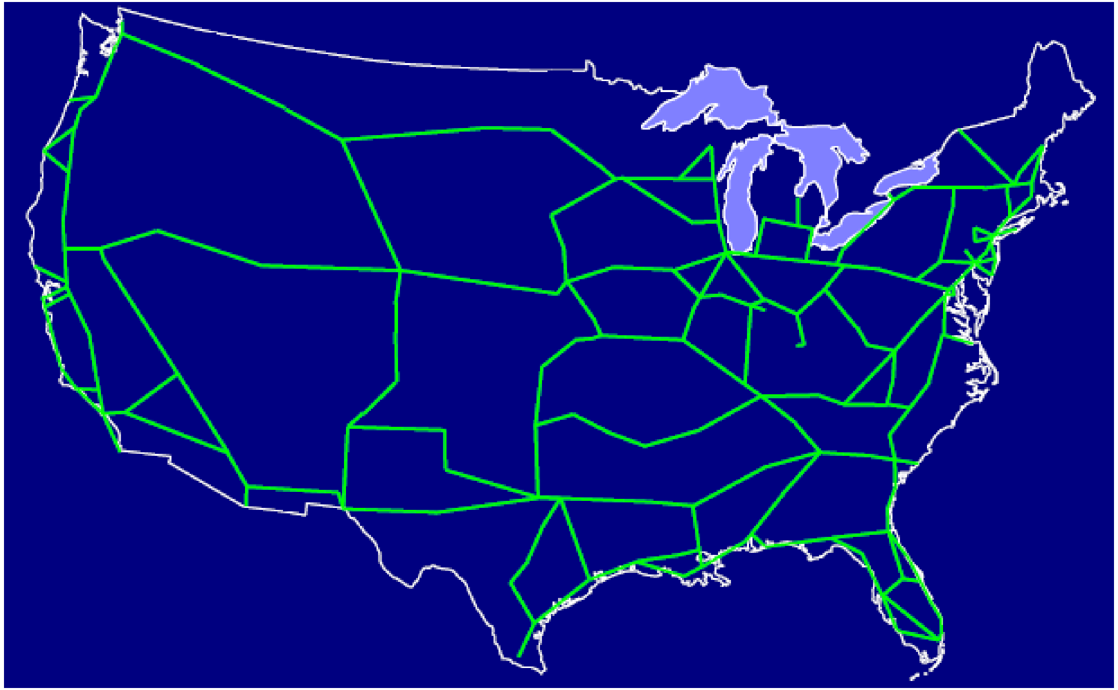
Manhattan Grid and Full Mesh





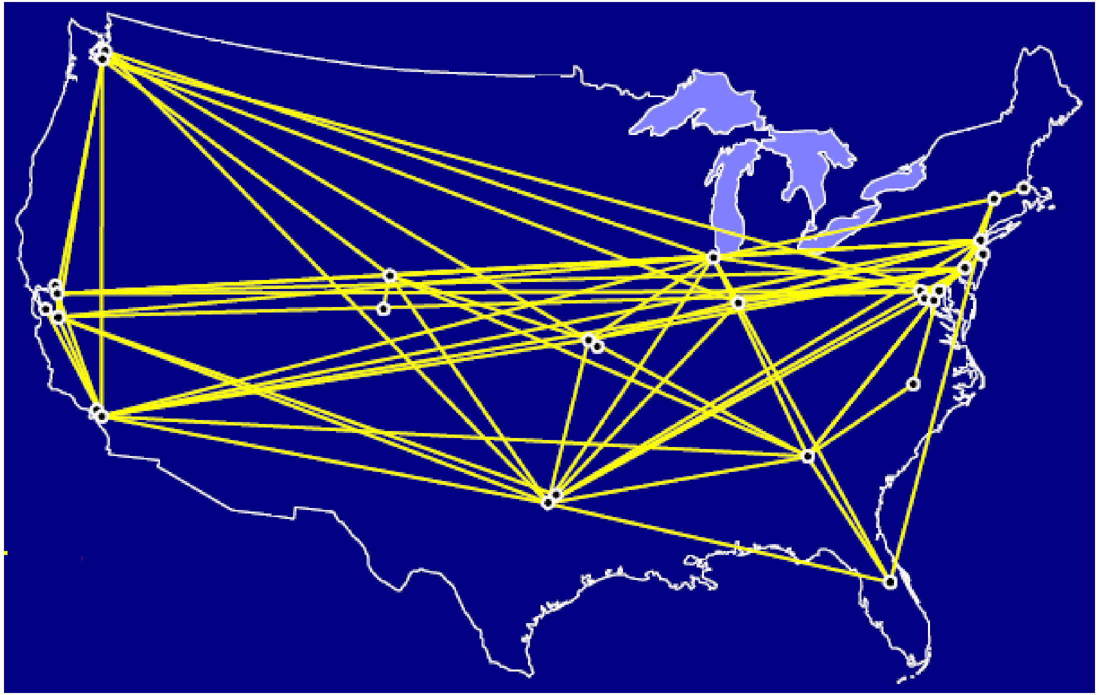
Example of Networks with Grid Topology

Optical Fiber



Example of Networks with Mesh Topology

IP Routing



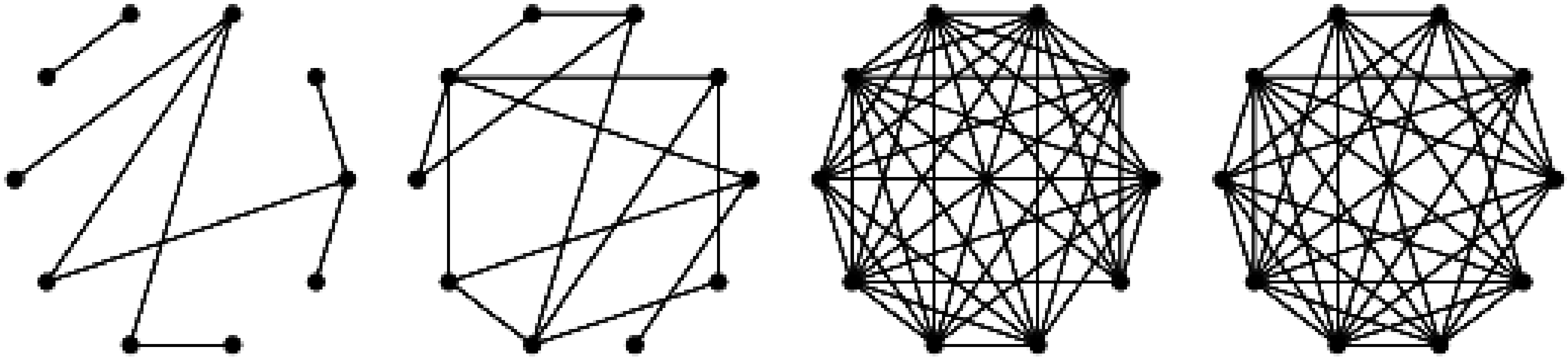
Random Networks (or Graphs)
ER Graphs (1)

- Initially studied by **Erdős** and **Rényi** in 1959.
  - Pairs of nodes are randomly connected by a given number of connections.
  - Two nodes are connected by a certain probability p.
- All nodes have approximately the same number of neighbors, which differs slightly from the average value.

Erdős, P.; Rényi, A. (1959). "On Random Graphs. I". Publicationes Mathematicae. **6**: 290–297

Random Networks (or Graphs)

ER Graphs (2)



3

Complex Networks

Complex Network Theory

- Regular networks **do not resemble** interaction networks at all.
- Random networks **capture some aspect** of it.
- **Complex networks of interactions** have substantially different characteristics from both classes of graphs.
- The attempt to formulate models capable of reproducing the properties of real interaction networks has produced the evolution of the classical graph theory into the modern **theory of complex networks**.

Complex Network Theory … Cont'd

- These networks show **non-intuitive characteristics** and can be made up of millions of units communicating with each other.
- **Mathematical methods** based on graph theory are therefore used to extract information from complex networks in a synthetic and objective way.

Instantiations of Complex Networks
Connections on the Personal Network

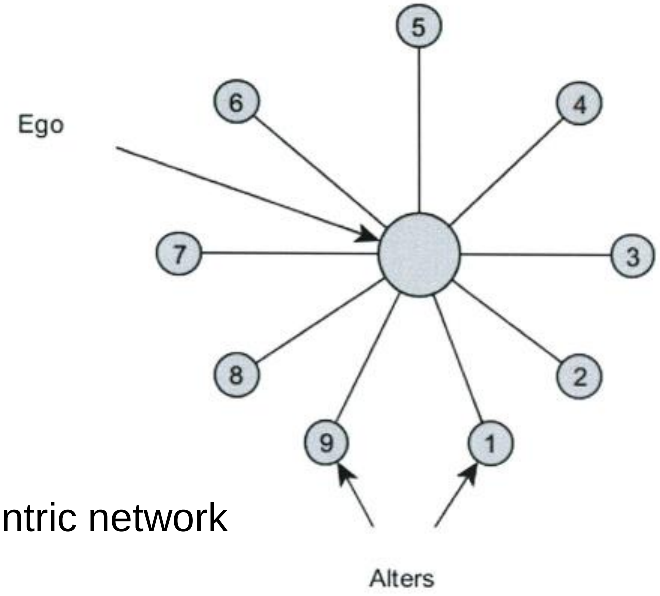
- (Online) social networks are **complex networks**.
- The **personal network of contacts*** is usually composed of a "first order" area (relationships that the individual has directly), a "second order" area (contacts made through an intermediary), and so on...

*Mitchell, J.C. (ed.) 1969 Social Networks in Urban Situations. Analyses of Personal Relationships in Central African Towns, Manchester, UK, Manchester University Press

Networks of First, Second, ..., Order

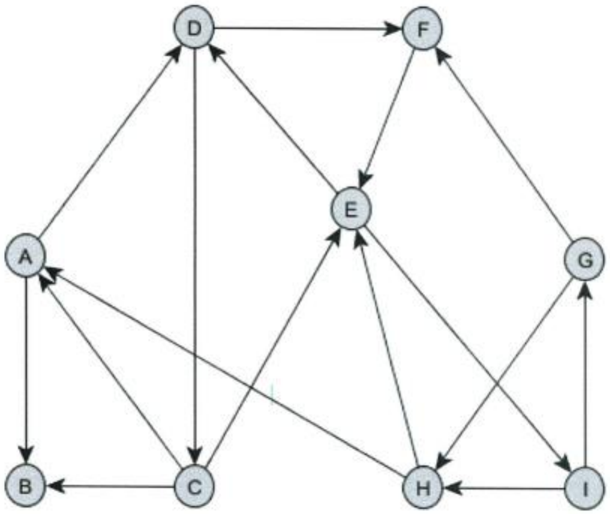
- The first-order network is generally called the **ego-centric network** (or ego-network).
- Networks that move away from the first order are **called socio-centric networks**.
- Both types of networks can be studied using **Social Network Analysis** techniques to grasp a large number of clues about social networks.

«Ego-centric» VS «Socio-centric»



Ego-centric network

Socio-centric network



«Ego-centric» VS «Socio-centric» ... Cont'd

- **Ego-centric analysis**:
  - It focuses on the network surrounding a specific node.
- **Socio-centric analysis**:
  - It concerns the study of "complete" networks.
- Data relating to "ego-centric" networks can be identified in a "socio-centric" network by selecting a node and examining its neighbors and the connections between them.
  - The most important property of social data is that they are often based on cultural values and symbols, and are generated on the basis of motivations, meanings and typification.

Characteristics of Complex Networks

- **Several phenomena** are related to the theory of complex networks:
    1. Small-world ("six degrees of separation")
    2. Clustering
    3. Strength of weak ties
    4. Scale invariance (scale-free)

1. The «Small-world» Theory

La teoria del mondo piccolo (o dei piccoli mondi)

- The concept of **small-world** coined by Milgram* is linked to studies on the personal network.

- In order to define the concept of "small-world", Miligram carried out an experiment that remained famous and subsequently repeated in various social networks.

*S. Milgram. The small world problem. Psychology Today, 1:61–67, 1967

The Milgram Experiment (Definition)

- Milgram randomly chose a sample of Americans (in the cities of Omaha, Nebraska and Wichita, Kansas) and asked them to deliver a message to a stranger (a stockbroker who worked in Boston and resided in Sharon, Massachusetts) of whom they only knew the name and a few other details but not the address.
- The messages had to be sent using only one's own network of acquaintances, based on considerations on which could be the highest probability of reaching the recipient.

The Milgram Experiment (Results)

- Results: The packets took on average only between 5 and 7 steps to reach the recipient.

Length distribution of the completed chains in Milgram small-world experiment.

Chain length	2	3	4	5	6	7	8	9	10
Number of completed chains	2	4	9	8	11	5	1	2	2

The Milgram Experiment (Considerations)

- The results of the experiment show how **short paths** (*cammini brevi*) exist between individuals in large social networks.

- More importantly, the experiment shows how these short paths can be found by ordinary people.
  - People rarely have more than local knowledge of their network;
  - One can get acquainted with friends and friends of friends;
  - It will be difficult to have knowledge of the entire path of individuals between themselves and an arbitrary "target" user.

An Experiment on the «Information Network»

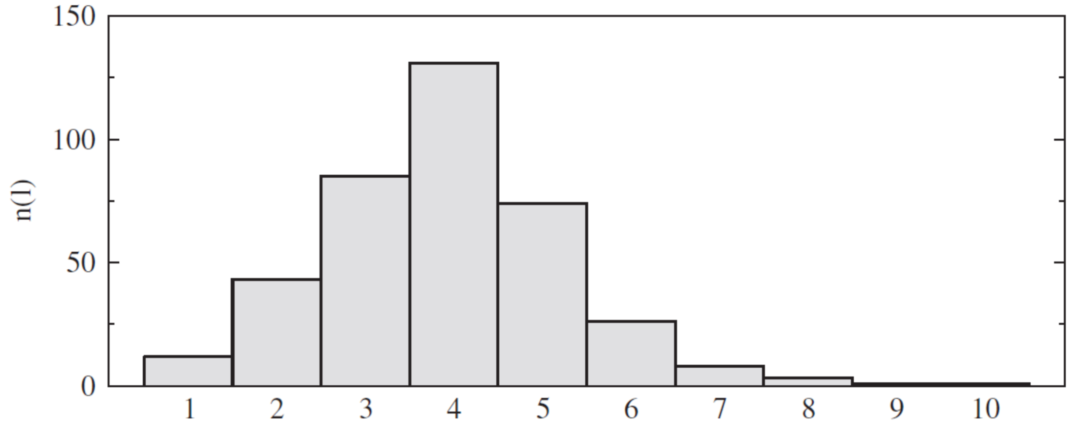
- In addition to social networks, other examples of a small-world are the **World Wide Web** and the communication network formed by the **exchange of e-mails**.

- **Experiment**: in 2003 more than 60,000 emails were sent by 166 participants in the experiment, trying to reach 18 "target" users from 13 different countries by forwarding the received email only to acquaintances.

https://www.newscientist.com/article/dn4037-email-experiment-confirms-six-degrees-of-separation/

Results

- Out of 24,163 chains, only 384 were completed.
- The distribution of completed chains versus the number of steps required.



Results … Cont'd

- Reasons for choosing the next recipient:

Declared reasons for choosing the next recipient at each step of the chain of the email small-world experiment.

Step	Location	Travel	Family	Work	Education	Friends	Cooperative	Other
1	33	16	11	16	3	9	9	3
2	40	11	11	19	4	6	7	2
3	37	8	10	26	6	6	4	3
4	33	6	7	31	8	5	5	5
5	27	3	6	38	12	6	3	5
6	21	3	5	42	15	4	5	5
7	16	3	3	46	19	8	5	0

The Number of Erdős

- **Paul Erdős**, one of the fathers of the theory of random networks, wrote scientific articles in various areas of mathematics, in collaboration with other famous mathematicians.

- It is possible to calculate a number that identifies the distance between Erdős from other mathematicians, the so-called Erdős number.

How to Compute the Number of Erdős

- The direct co-authors of Erdős have number 1.
- Co-authors of a co-author of Erdős have number 2.
- And so on…
- How far are you from Erdős?? → https://mathscinet.ams.org/mathscinet/collaborationDistance.html

Example

Search MSC | **Collaboration Distance** | Current Journals | Current Publications

MR Erdos Number = 5

Marco Viviani	coauthored with	Gabriella Pasi
Gabriella Pasi	coauthored with	Krassimir Todorov Atanassov
Krassimir Todorov Atanassov	coauthored with	Anthony G. Shannon
Anthony G. Shannon	coauthored with	Krishnaswami Alladi
Krishnaswami Alladi	coauthored with	Paul Erdős[1]

Change First Author | **Change Second Author** | **New Search**

The «Oracle of Bacon»

- In a January 1994 interview, actor Kevin Bacon claimed to have worked with all the actors in Hollywood or, at least, with someone who had acted in them together.
- It allows to calculate the degrees of separation between Kevin Bacon (or another actor/actress) and any other actor or actress.
- Based on «[The Internet Movie Database](https://www.imdb.com)».

Example

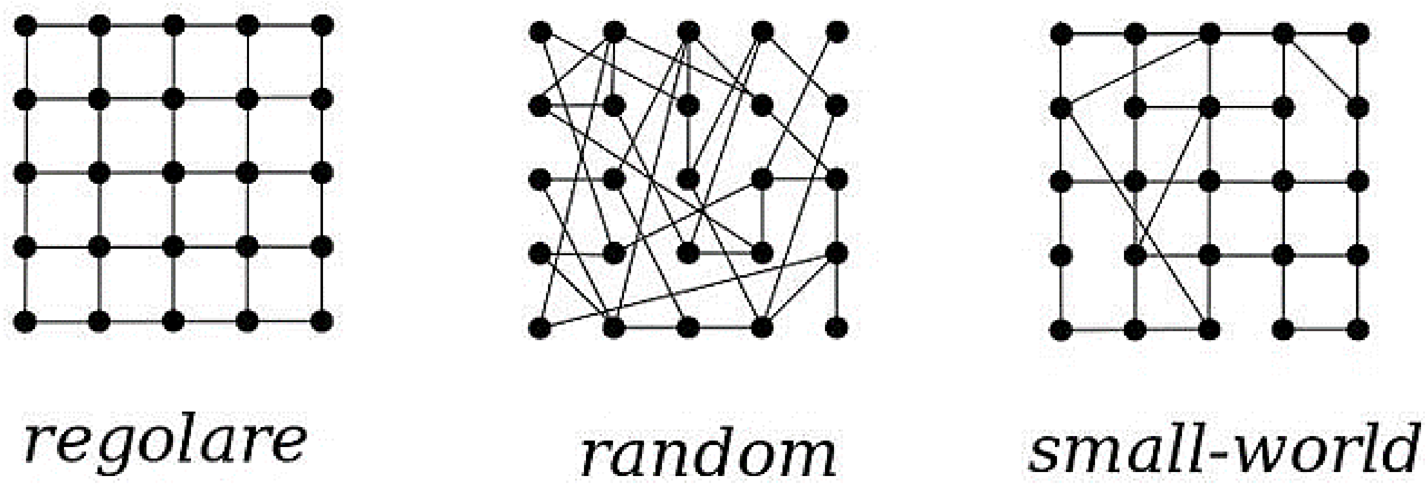
Anne Hathaway

Lino Banfi

Conclusions on the «Small-world» Theory

- Today we know that this property is not peculiar to social networks: practically all real networks of interactions in complex systems have the characteristic of small-world.
- This property makes networks very efficient in terms of **information propagation** speed (and more)
  - Infectious diseases, for example, spread over a small-world network much more easily than over a regular network.

Motivations Illustrated Graphically



2. Clustering

- There is a tendency in social networks to create **clusters**:
  - In social networks there are often communities of individuals all or almost all in relation to each other.

- This property already known as **transitivity** in sociology (Wassermann and Faust, 1994)* is called clustering and has been quantified by the so-called clustering coefficient (Watts and Strogatz, 1998)**, which essentially measures how many of the friends of a given individual are also in turn friends with each other.

*Wasserman, Stanley, and Katherine Faust. Social network analysis: Methods and applications. Vol. 8. Cambridge university press, 1994
**Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world'networks." Nature 393.6684 (1998): 440

Clustering and Small-world

- Random networks **resemble** interaction networks w.r.t. the small-world property.

- However, they **differ** in another important respect:
  - The clustering coefficient of a typical interaction network is usually much larger than that of the corresponding random network (i.e., a network with the same number of nodes and ties, but where each tie connects a pair of nodes chosen at random).

Concrete Examples

k = degree L = path lenght C = clustering coefficient

Characteristic path length and clustering coefficient of three real networks (two social networks and a biological one) and of the corresponding ER random graphs with same number of nodes and links.

	N	$\langle k \rangle$	L	L^{ER}	C	C^{ER}
Movie actors	225226	73.71	3.65	2.87	0.79	0.00033
ArXiv	44337	10.79	5.99	4.50	0.62	0.00024
C. elegans	279	16.39	2.44	2.01	0.34	0.06

3. The Strength of Weak Ties

- The **strength of weak ties** (Mark Granovetter, 1977*: the strength of a tie is given by the (probably linear) combination of the amount of time, emotional intensity, intimacy (mutual confidence) and the exchange of services that characterizes the tie.

*Granovetter, Mark S. "The strength of weak ties."
Social networks. 1977. 347-367

3. The Strength of Weak Ties ... Cont'd

- Granovetter defines people's "**strong ties**" as those that unite them to primary networks (family, institutions, organizations), since they constitute rather compact networks of belonging.

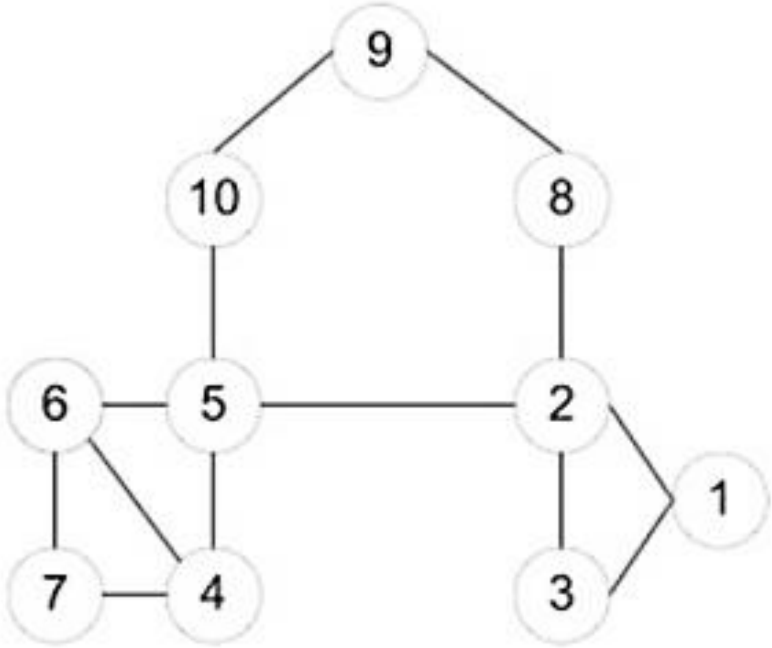
- He calls "**weak ties**" those that characterize the informal networks of people, which, in terms of social integration and increase in social capital, are often more important than strong ties, functioning as bridges between different segments of the social network.
  - These types of ties are crucial, for example, when looking for a new job.

Measuring the Strength of Ties

- The sociological concept of "**tie strength**" can be measured through Social Network Analysis techniques in different ways:
  - By identifying "shortcut bridges" (*ponti scorciatoia*).
  - The neighborhood overlap assessment.
  - Evaluation of user activities.

Shortcut Bridges

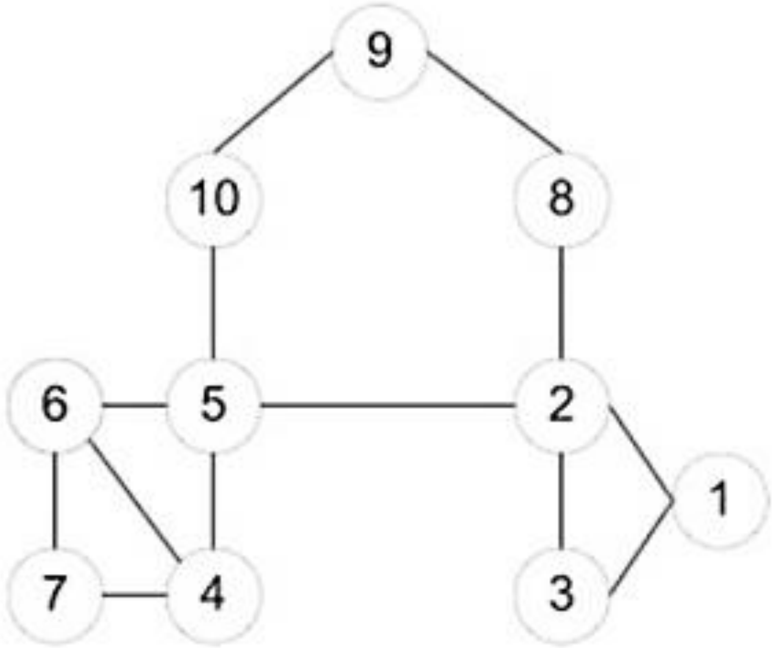
- **Bridges** are very rare in real social networks
- Alternatively, you can "loosen" the definition, and check if the distance between two nodes increases when the arc is removed, which in this case is a "shortcut"
- The greater the distance, the "weaker" the type of bond



Shortcut Bridges (Example)

- $d(2,5) = 4$
  if $e = (2,5)$ is removed
- $d(5,6) = 2$
  if $e = (5,6)$ is removed

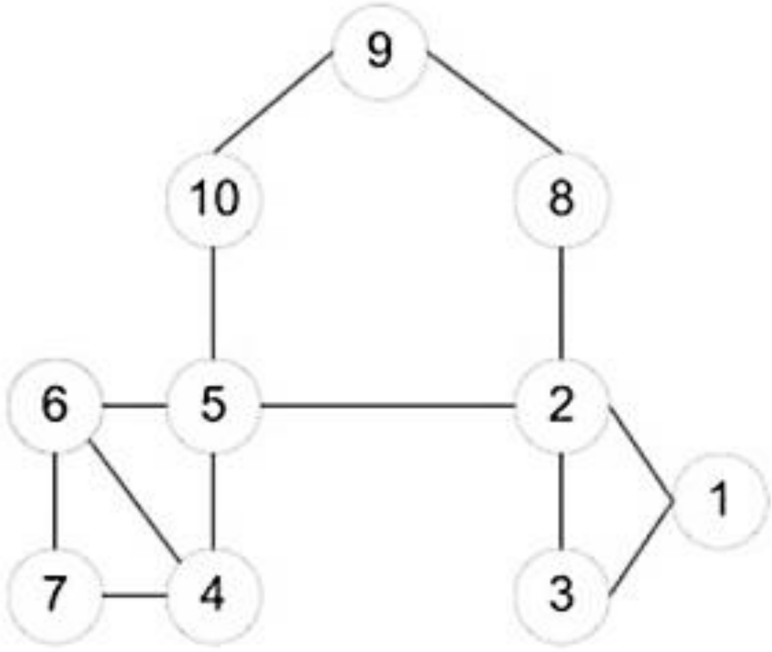
$e = (5,6)$ is stronger than $e = (2,5)$



Neighborhood Overlap

- The strength of a tie can also be measured by **how overlapping** the neighbors of two nodes v_i and v_j are
- This overlap $O(v_i, v_j)$ can be defined as:

$$O(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)| - 2}$$



Neighborhood Overlap (Example)

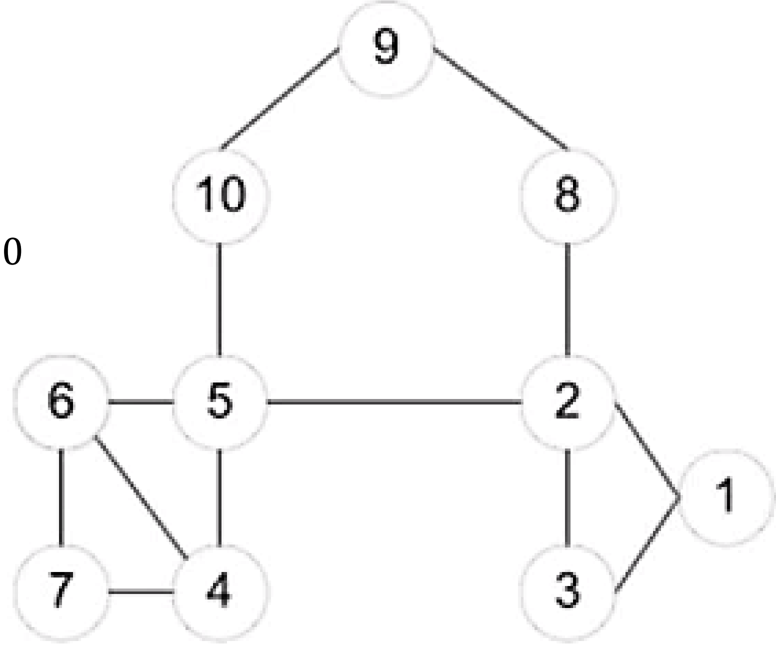
- $O(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)| - 2}$

- $O(2,5) = \frac{|\{1,3,5,8\} \cap \{2,4,6,10\}|}{|\{1,3,5,8\} \cup \{2,4,6,10\}| - 2} = \frac{0}{6} = 0$

- $O(5,6) =$

$$\frac{|\{2,4,6,10\} \cap \{4,5,7\}|}{|\{1,3,5,8\} \cup \{4,5,7\ \}| - 2} =$$

$$= \frac{|\{4\}|}{|\{1,3,4,5,7,8\}| - 2} = \frac{1}{4}$$

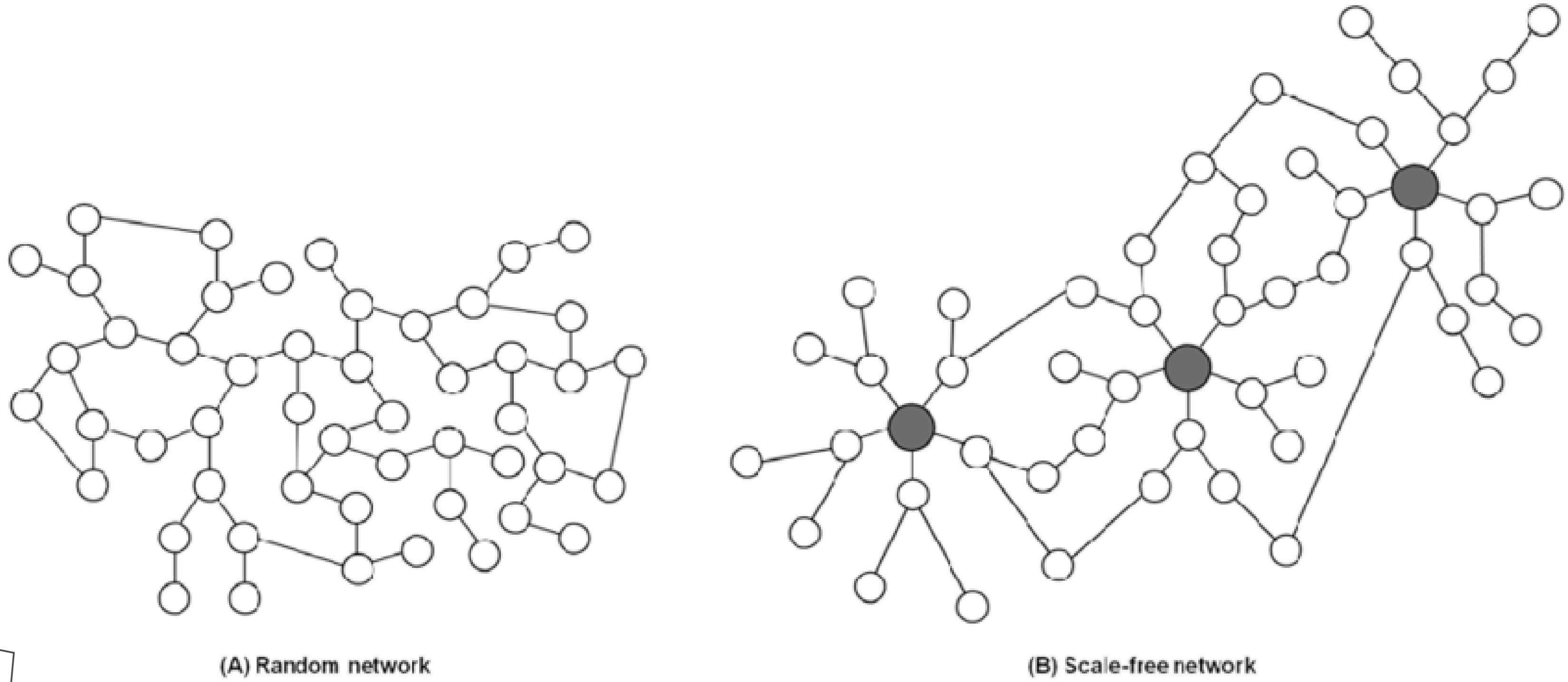


4. Scale Invariance

- **Scale invariance** in a network (scale-free network) occurs when in a graph the relationship between the number of nodes and the number of their connections is of a negative exponential type, and therefore invariant under changes in scale.*

*Barabási, Albert-László, Réka Albert, and Hawoong Jeong. "Mean-field theory for scale-free random networks." *Physica A: Statistical Mechanics and its Applications* 272.1-2 (1999): 173-187

Scale-free Network (Example)



Scale-free Network (Hub)

- **Assumption**: When a node needs to establish a new connection, it prefers to do so with a node that already has many connections.
  - This leads to exponential growth as the number of connections in the network increases.

- Nodes of this type are called **hubs**.

- This mechanism is very resistant compared to other growth mechanisms of the network and often leads to the preservation of the property of the invariance of scale.

Scale Invariance and «Small-world»

- The **presence of hubs** is the basis of the small-world effect.
- In this sense, the hubs have the function of connecting areas of the graph that would otherwise be separate.