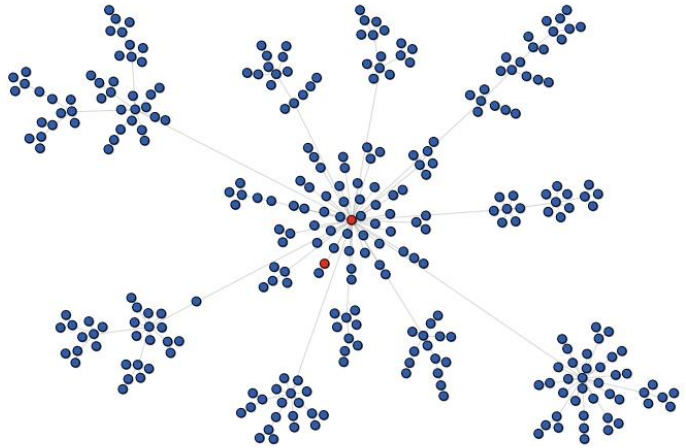
Topological Characteristics of biological networks

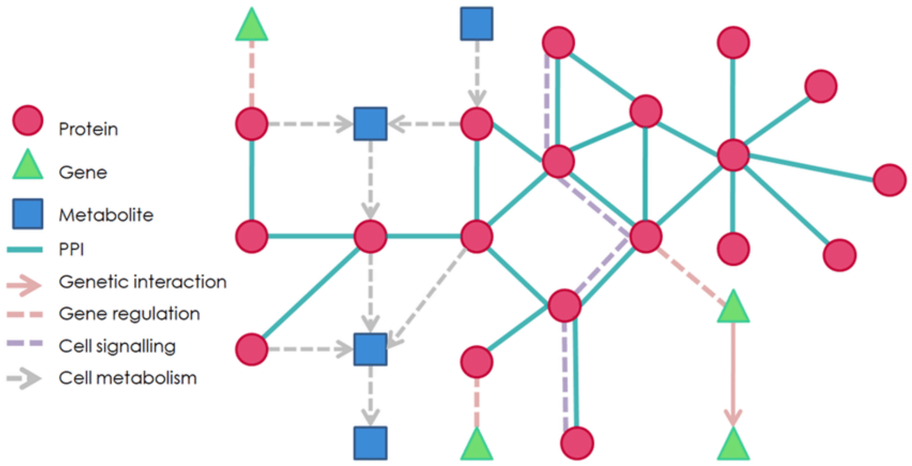
Sadegh Sulaimany



Biological network analysis course

Agenda

- Important measures for network type determination
  1. *Average path length* (L)
  2. *Clustering coefficient* (CC)
  3. *Degree distribution* (P_k)
  4. *Rich club coefficient*
  5. *Modularity*
- Network types
  1. Random
  2. Small-world
  3. Scale-free

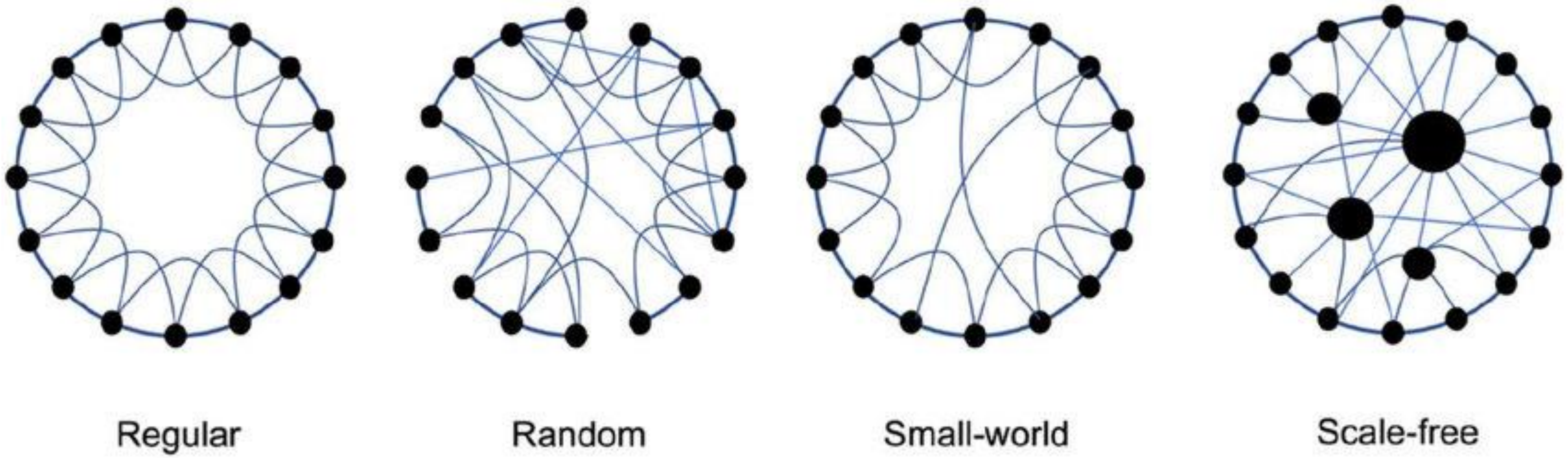


Introduction

- Unlike conventional graphs
  - Real-world networks exhibit non-trivial topological properties, such as
    - varying degree distributions
    - high or low clustering coefficients
      degree assortativity
    - Low average path length
  - Accordingly, networks are classified into different models

Introduction

- How to distinguish different network types?
  - Describe and compare the complex networks
    - By set of summary statistics
      or quantitative performance measures



Most important measures for network type

› Play a key role in network analysis
  1. *Average path length*
  2. *Clustering coefficient*
  3. *Degree distribution*
  4. *Rich club coefficient*
  5. *Modularity*

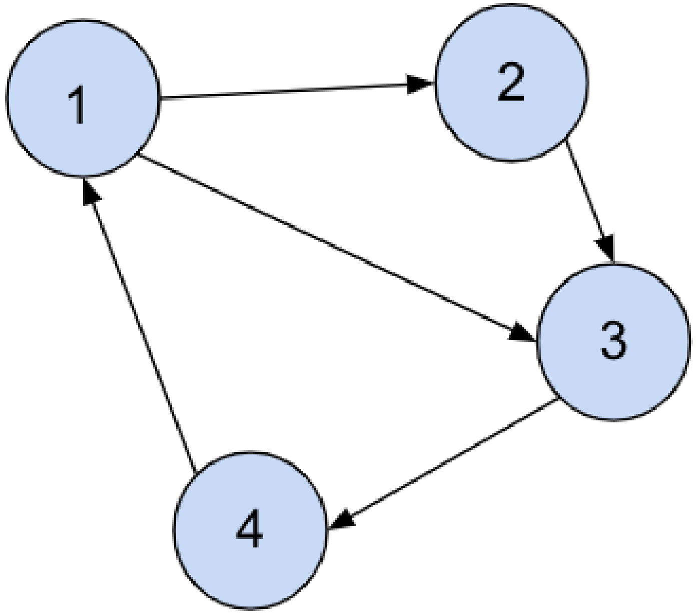
Average path length

- How complex real-world networks are "wired" and evolving
- Moreover, the it is a measure of network size
- It indicates the rate of (quick) transfer of information throughout the network
  - Mean distance of all possible shortest path (*di,j*) followed between any two nodes, *vi* and *vj*

for $\mathcal{G}(\mathcal{V}, \mathcal{E})$

$$L = \frac{1}{|\mathcal{V}|(|\mathcal{V}| - 1)} \sum_{i=1}^{|\mathcal{V}|} \sum_{j=1}^{|\mathcal{V}|} d_{i,j}$$

Average path length



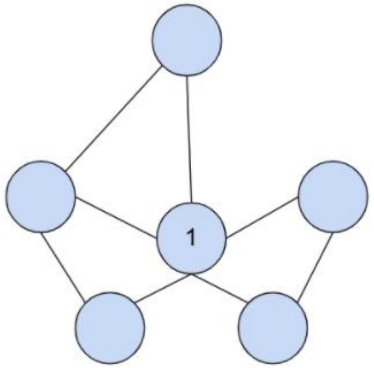
Shortest path:
P(1,2) = 1
P(1,3) = 1
P(1,4) = (1,3) + (3,4) = 2
P(2,1) = (2,3) + (3,4) + (4,1) = 3
P(2,3) = 1
.....

Average path length:
(1 + 1 + 2 + 3 + 1 + ...) / (number of all shortest path)

- Most of the real-world networks have a small average path length

Clustering coefficient

- A measure of affinity (likelihood), to which nodes in a network tends to create tightly connected group with each others
- Two types
  - Local: for one node
    - the ratio of total number of edges that are present among the neighbors of a node *vi*
      to the total number of possible edges that could exist among the neighbors of *vi*
  - Global: for whole network
    - Mean of local clustering coefficients of all nodes



Clustering coefficient

– Local CC

› For directed graph

$$\mathtt{CC}_{v_i} = \frac{\sum_{j=1}^{|N_{v_i}|} \lambda(v_i, v_j)}{|N_{v_i}|(|N_{v_i}| - 1)},$$

$$\text{where, } \lambda(v_i, v_j) = \begin{cases} 1, & \text{if } (v_i, v_j) \text{ is connected, } \forall v_j \in N_{v_i}, i \neq j \\ 0, & \text{otherwise.} \end{cases}$$

$$N_{v_i} = \{v_k | e(i,k) \in \mathcal{E} \vee e(k,i) \in \mathcal{E}\}$$

Clustering coefficient

› Local CC

$$\mathrm{CC}_{v_i} = \frac{2 \times \sum_{j=1}^{|N_{v_i}|} \lambda(v_i, v_j)}{|N_{v_i}|(|N_{v_i}| - 1)}$$

› Global CC

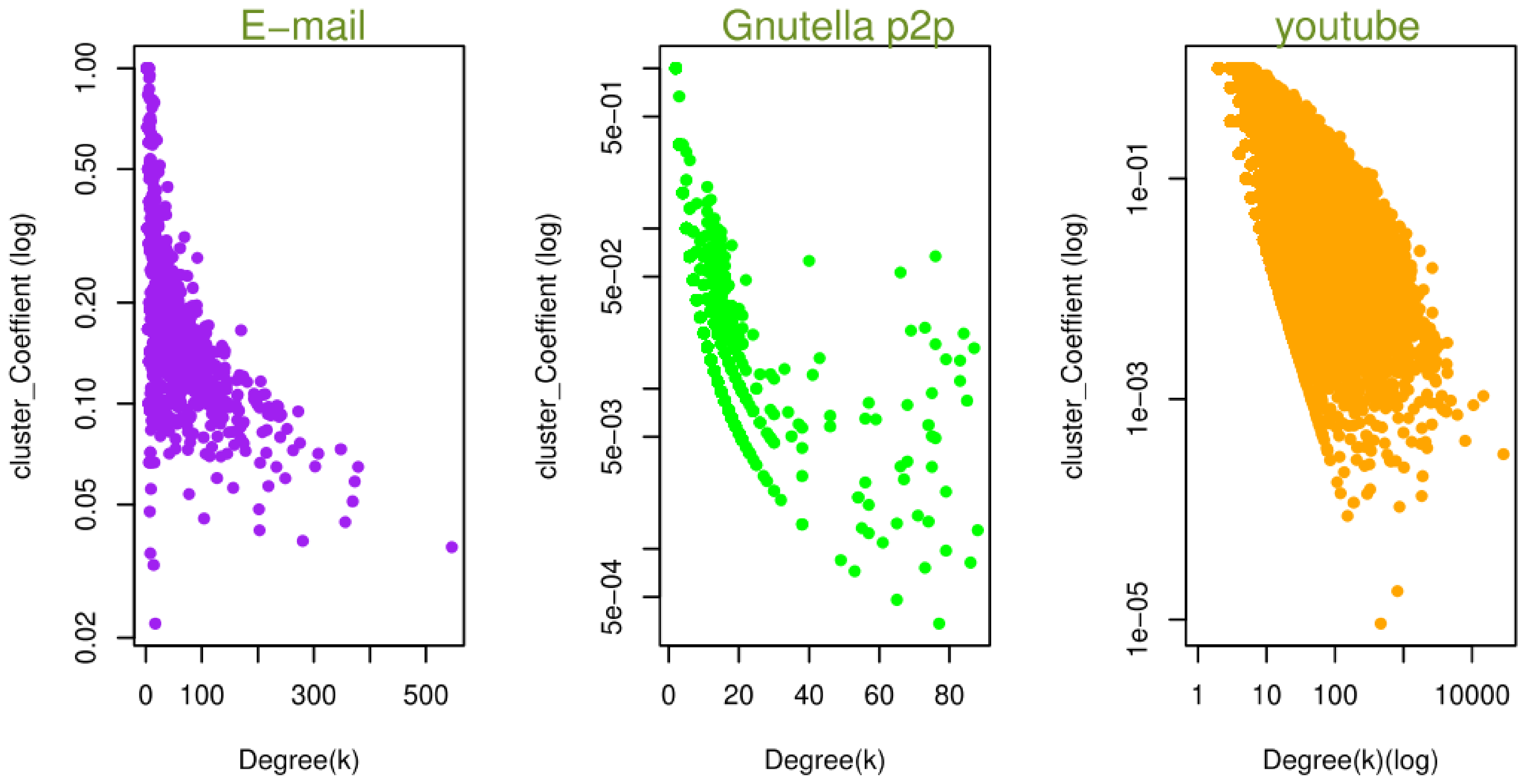
$$\overline{\mathrm{CC}} = \frac{\sum_{i=1}^{\mathcal{V}} \mathrm{CC}_{v_i}}{\mathcal{V}}$$

Clustering coefficient

- Clustering coefficient value can be between [0, 1]
  - the higher the value, the higher the tendency the network has to organize dense groups inside it

- What about Clustering coefficient for directed, weighted bipartite and … Networks?

Clustering coefficient



Degree distribution (P_k)

- Is the probability that a node chosen randomly has a degree k
- So
  - P_k of a network is the fraction of nodes (n) having degree $k(n_k)$

$$P_k = \frac{n_k}{n}$$

  - For a large network with $\mathcal{V}$ nodes
    - where average degree is far less than the number of nodes

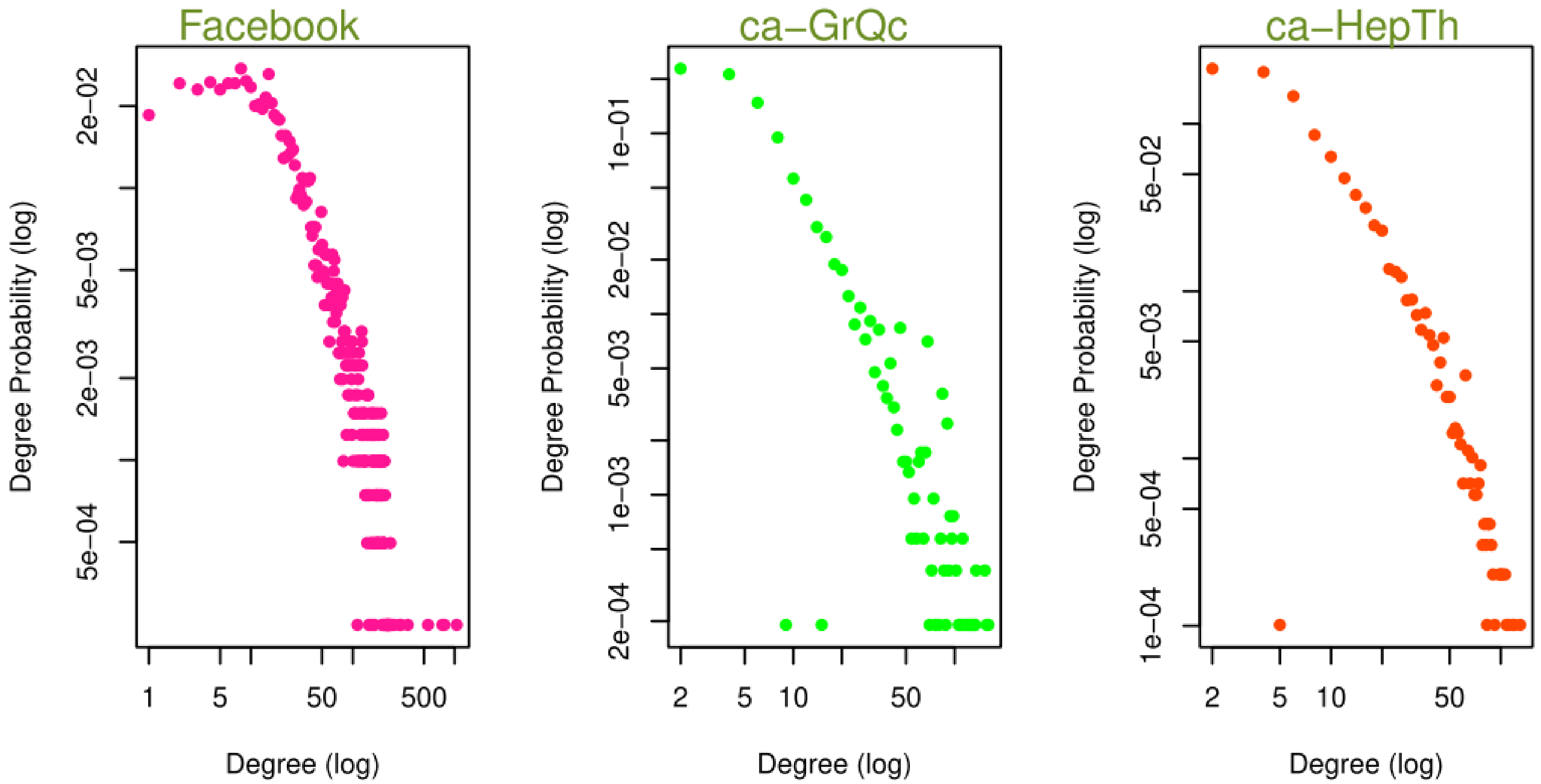
$$\langle k \rangle \ll |\mathcal{V}|$$

    - Degree distribution approximately follows the Poisson distribution:

$$P_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Degree distribution (P_k)

› Example



Degree distribution

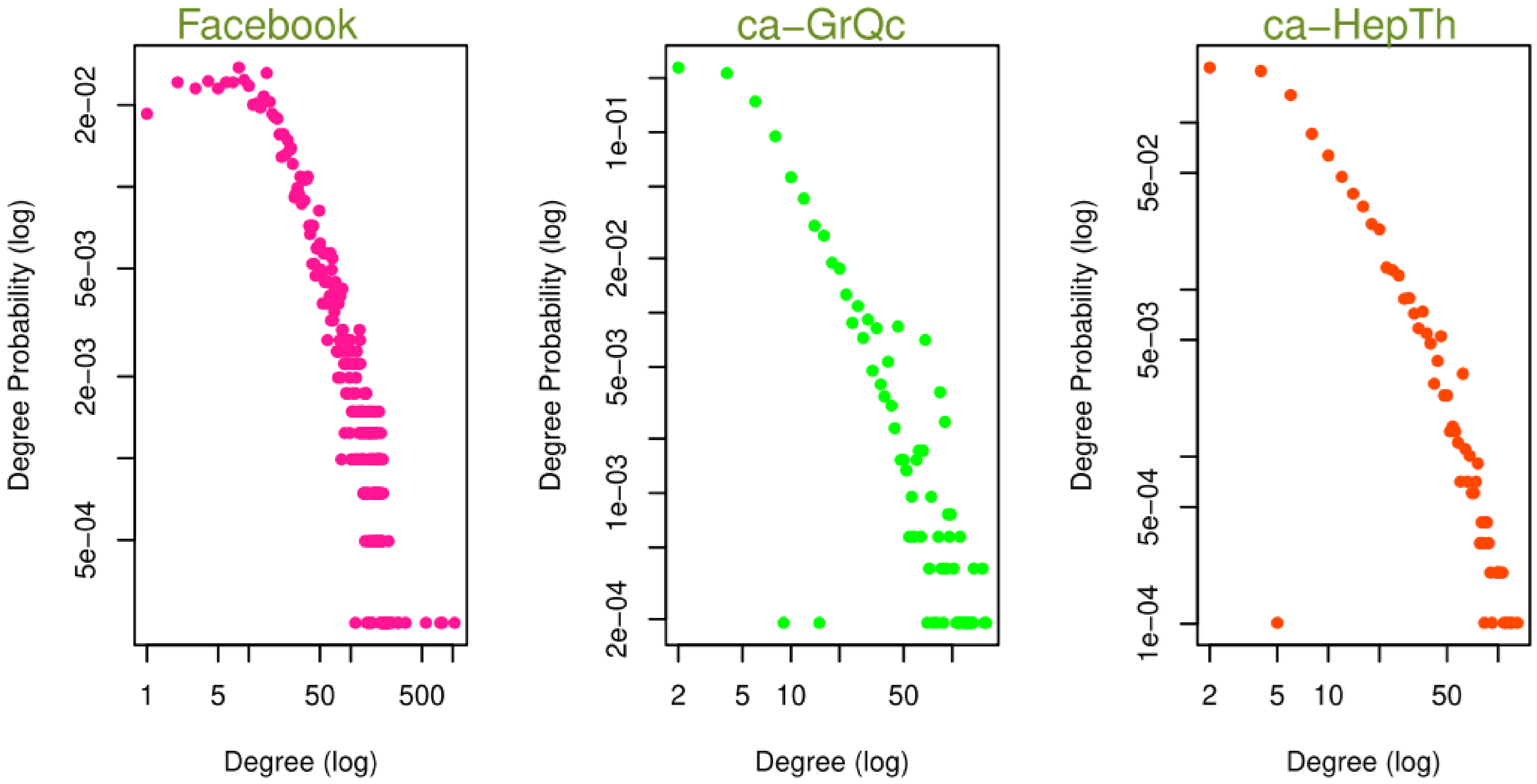
- For real-world networks
  - like the Internet, social network, etc.
    - follow the *power law* property
    - γ a constant between 2 and 3

$$P_k \sim k^{-\gamma}$$

    - *power law distribution*
    - or *scale-free distribution*
      - The shape of the distribution does not change with scale

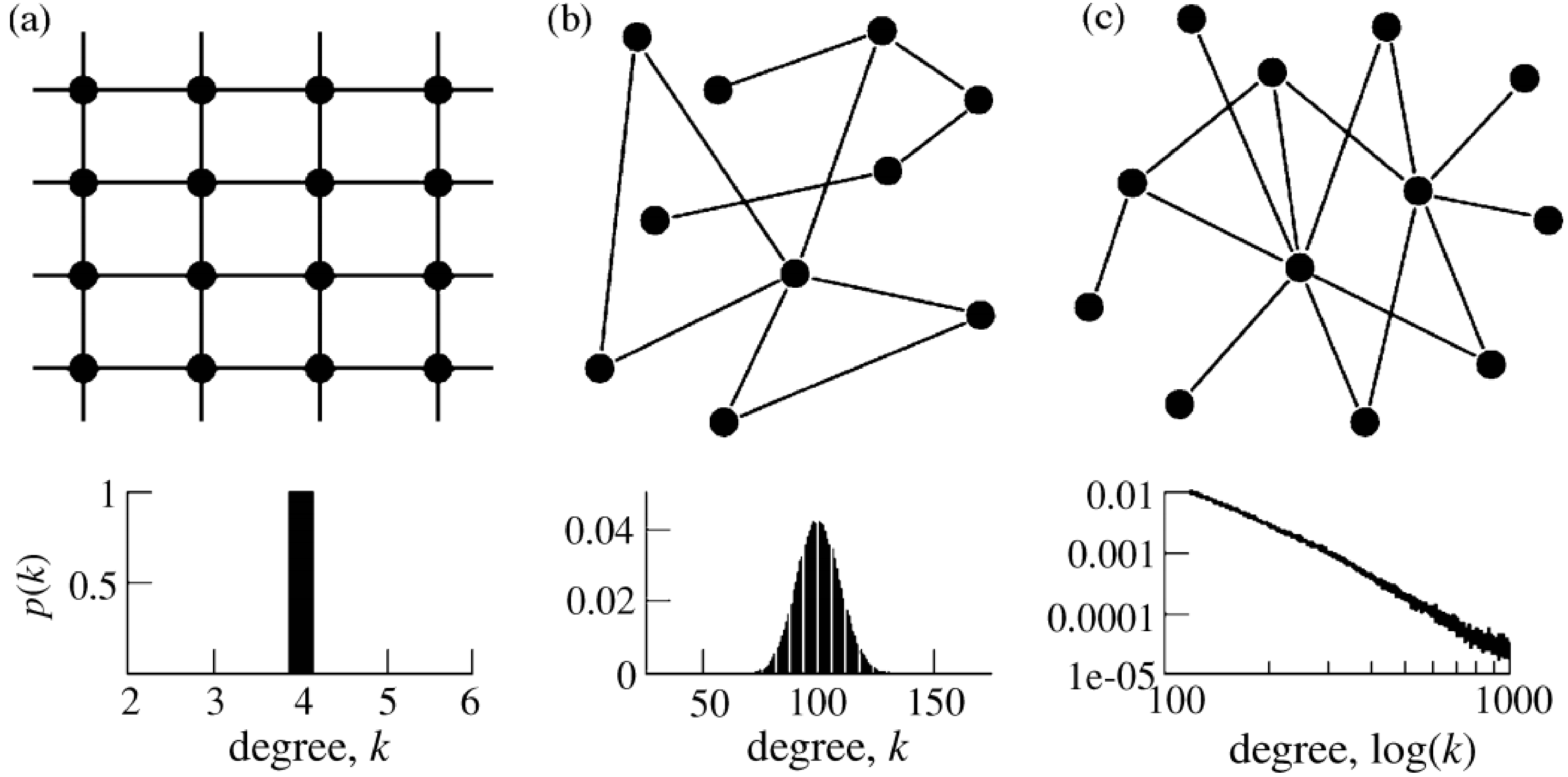
Degree distribution

– Power law distribution examples



Degree distribution

– Degree distribution in different networks



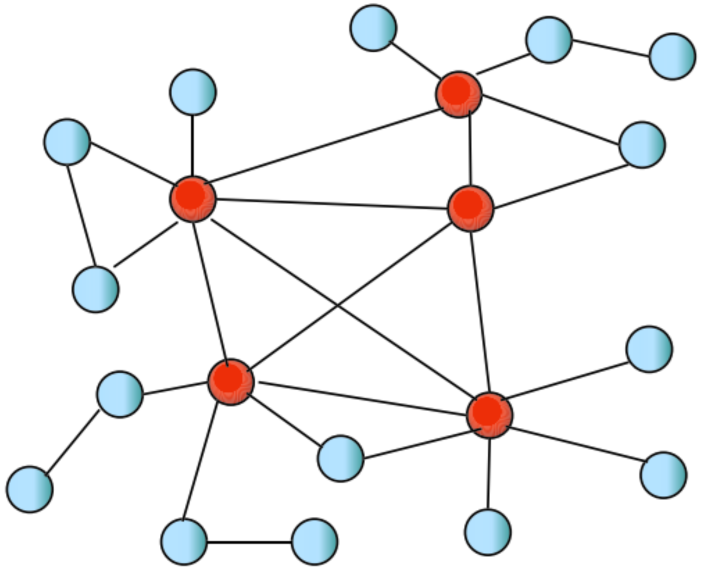
Rich club coefficient

- Introduced by Zhou and Mondragon in the context of the Internet topology, 2014
  - › Tendency of high-degree nodes (i.e., the hubs) in the network, to be very well-connected to other hub nodes
    - Rich = hub, club = subgraph
    - Measure of connectedness density within the *club*

$$\Phi(r) = \frac{2E(r)}{r(r-1)}$$

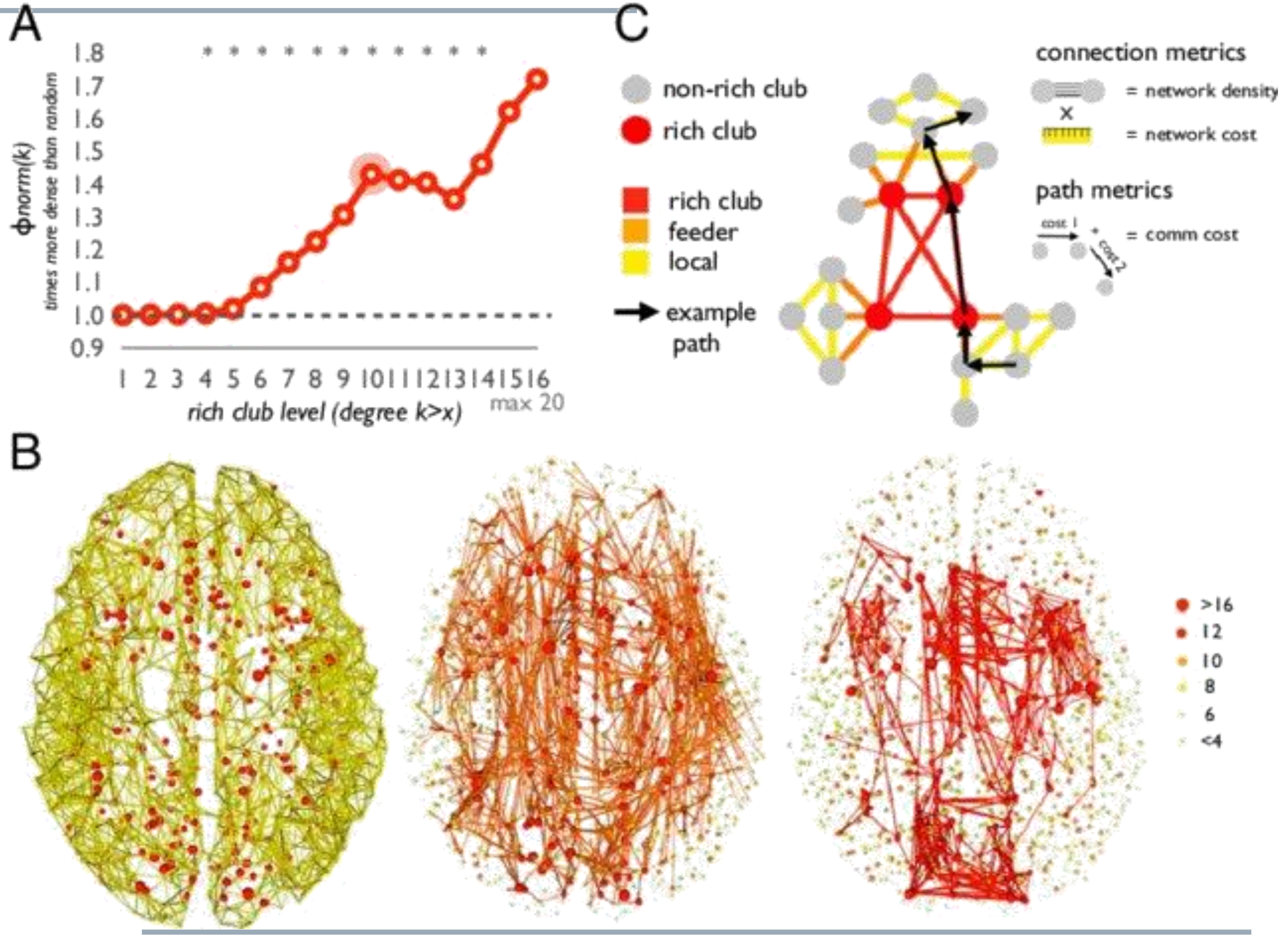
- *r* is the rank of node in case of its degree
- *E(r)* is the total number of links between *r* hub nodes

This measure explicitly reflects
how densely connected are the nodes within a network



Rich club coefficient

› Example



Modularity

- *Strength* and *quality* of partition of a network into communities or clusters
  - How much the Dense subgraph
    - A*ij* is the adjacency matrix
    - M the total number of edges
    - P*ij* expected number of edges between vertices i and j
    - $\delta_{C_i,C_j} = 1$ if nodes i and j are in the same community (C*i*=C*j*), otherwise 0

$$Q = \frac{1}{2M}\sum_{i,j\in V}[A_{ij} - P_{ij}]\delta_{C_i,C_j}$$

Characteristics of various real-world networks. Each network has N number of nodes, E number of edges, average degree ⟨K⟩, average path length L, maximum distance (diameter) L_{max}, and clustering coefficient, C.

Network	$\mathcal{N}$	$\mathcal{E}$	⟨K⟩	L	L_{max}	C
Zachary karate club	34	78	4.5	2.44	5	0.25
Dolphins	62	159	5.12	3.45	8	0.309
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	0.59
Protein Interactions	2,018	2,930	2.9	5.61	14	0.023
Power Grid	4,941	6,594	2.67	18.99	46	0.107
Science Collaboration	23,133	93,439	8.08	5.35	15	0.32
arXiv hep-ph	28,093	3,148,447	327.26	2.83	9	0.280
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	0.15
Email	57,194	103,731	1.81	5.88	18	0.032
Facebook friendships	63,731	817,035	25.64	4.31	15	0.148
Internet	192,244	609,066	6.34	6.98	26	0.24
DBLP co-authorship	317,080	1,049,866	6.62	6.75	23	0.306
WWW	325,729	1,497,134	4.6	11.27	93	0.11
Amazon (MDS)	334,863	925,872	5.52	11.73	47	0.205
Citation Network	449,673	4,707,958	10.43	11.21	42	0.43
Actor Network	702,388	29,397,908	83.71	3.91	14	0.79
LiveJournal	4,847,571	68,475,391	28.25	5.48	20	0.118

Network models

› Almost all large-scale networks share some common topological properties

– such as *scale-free distributions*, *low average path length*, and *strong community structure*

1. Random network

› Erd˝os-Rényi (ER) Model

– a specified probability describes the existence of an edge between each couple of nodes

– $\mathcal{G}(\mathcal{V}, P)$ random graph with *v* nodes and existence probability P for each edge

› ER has low clustering coefficient

› Average path length growth logarithmically with size of network

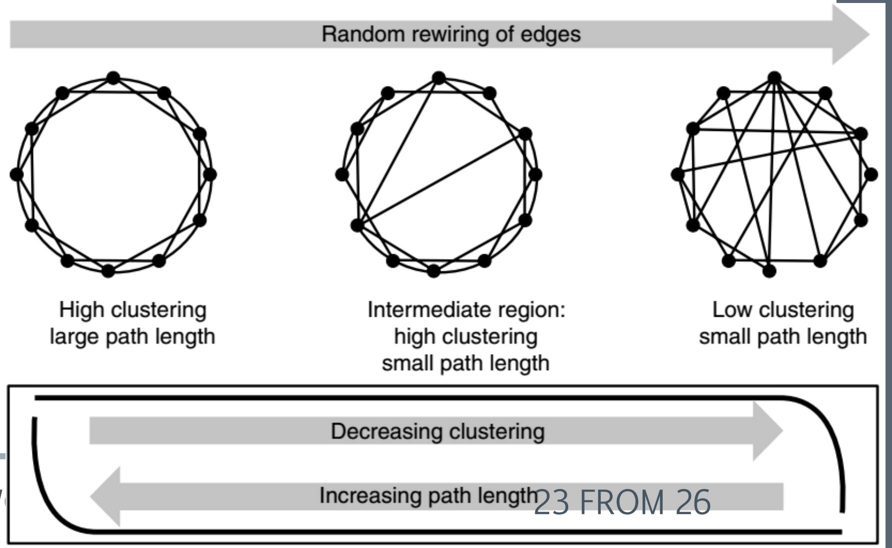
› Degree distribution is binomial!

$$\langle L \rangle \sim ln(|\mathcal{V}|)$$

$$P_k = \binom{|\mathcal{V}|-1}{k} P^k (1-P)^{|\mathcal{V}|-1-k}$$

Network models

2. Small-world networks
   - any two individuals in the world can reach out from another through a path of six or fewer acquaintances between them
   - *"small-world effect"* or *"six degree of separation"*
   - Example:
     - brain network
   - High clustering coefficient, Low average path length
     - Information spreading is fast
     - Different with regular or lattice, network?
       - Degree distribution
     - Watts and Strogatz
       - Artificial small-world network
         - Between Regular and Random networks



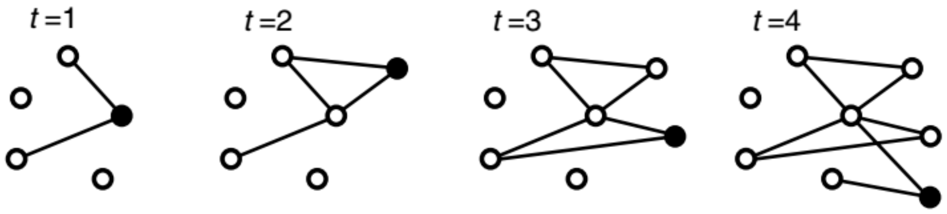
Network models

3. Scale-free networks

- all nodes have a roughly equal number of connections
- degree distribution for such networks follow Poisson distribution
- large number of nodes have a low degree, and there exist few high-degree (hub) nodes

- Example
  - › Metabolic networks
  - › phenomenon of *preferential attachment bias* is known as *“rich-get-richer”*

- Simulation
  - › Barabási Albert
    - – Growth and Preferential attachment



Thinking

- What is the time complexity of finding the small-world property of a network?
- How can we find the small-world or scale-free property of a network?
- Is it possible for a network to be simultaneously small-world and scale fee?
- Is it possible to find clustering coefficient for hypergraphs (metabolic networks)?

Question?

Bioinformation.ir

info@Bioinformation.ir

Social & Biological Network Analysis Laboratory