Ανάλυση Κοινωνικών Δικτύων Και Εφαρμογές

Μετρικές Ανάλυσης Ι

Συμεών Παπαβασιλείου (papavass@mail.ntua.gr)
Βασίλειος Καρυώτης (vassilis@netmode.ntua.gr)

07 Απριλίου, 2022

Outline

- Types of metrics
  1. Connections
  2. Distributions
  3. Segmentations
- Degree distribution (type 2)
- Strength (type 2)
- Average path length (type 2)
- Clustering coefficient (type 3)

Social & Complex Network Analysis

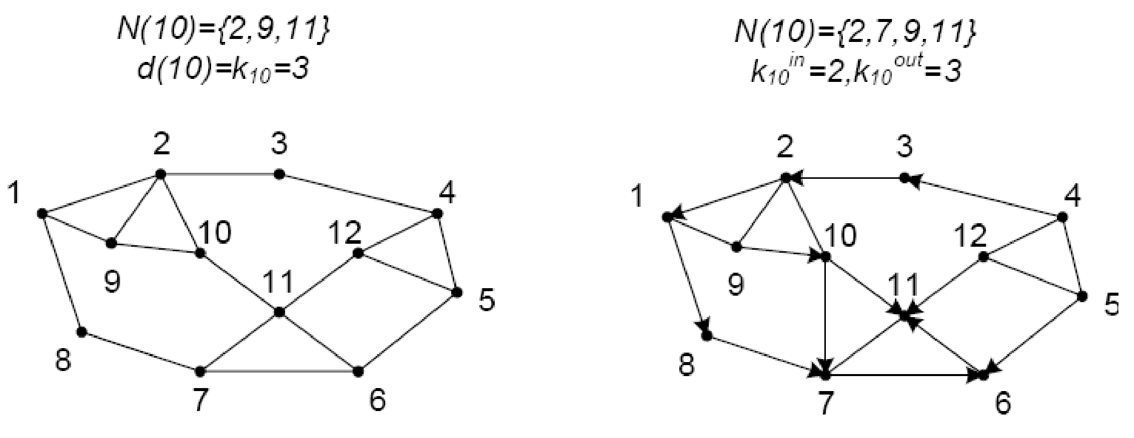
- Social network analysis (SNA)
  - investigate social structures using network & graph theories
- Complex network analysis (CNA)
  - Study the structure of networks with various methodologies

SNA and CNA include:

- SNA metrics
  - modeling
- Visualization
  - Inspection & monitoring
- Network (bigdata) analytics
  - Discovery of hidden correlations & exploitation

Degree Distribution

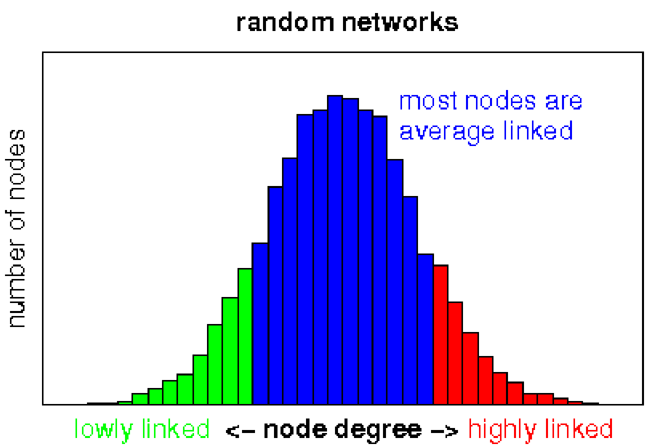
- Node degree (graph-theoretic metric)
  - Undirected: number of neighbors
  - Directed: number of in-/out-neighbors
- Degree distribution
  - Deterministic: for each node → degree value
  - Probabilistic: for each node → prob. giving degree value

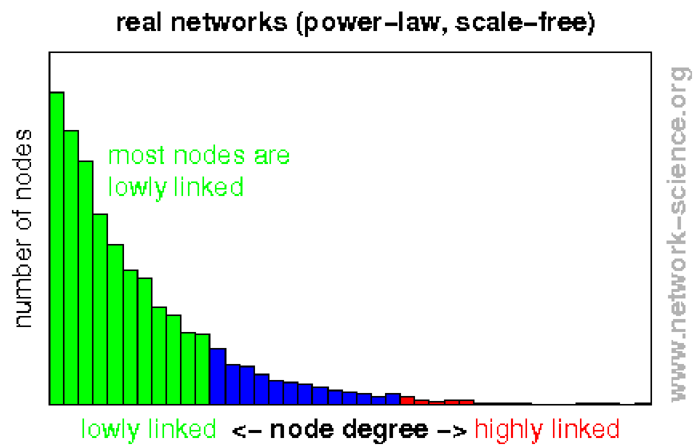


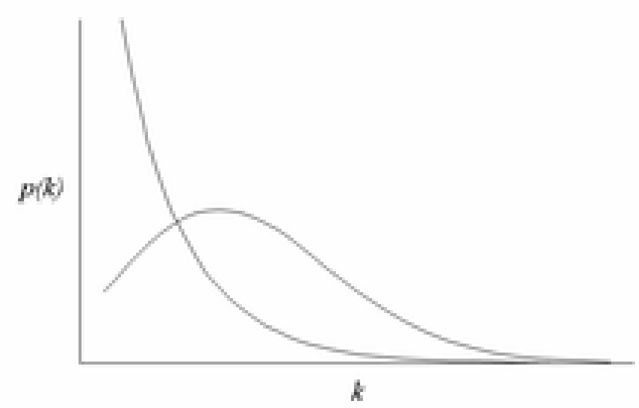
(a) undirected graph   (b) directed graph

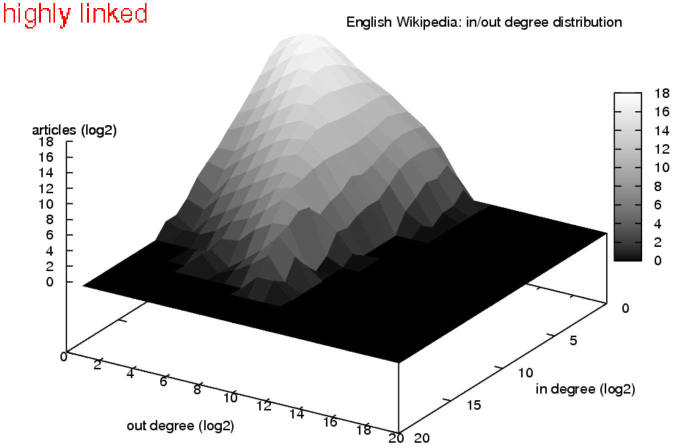
Examples of Degree Distributions

- Different types of networks → different degree distributions
  - Conversely: different degree distributions → network characterization
  - In fact, degree distribution determines uniquely network topology









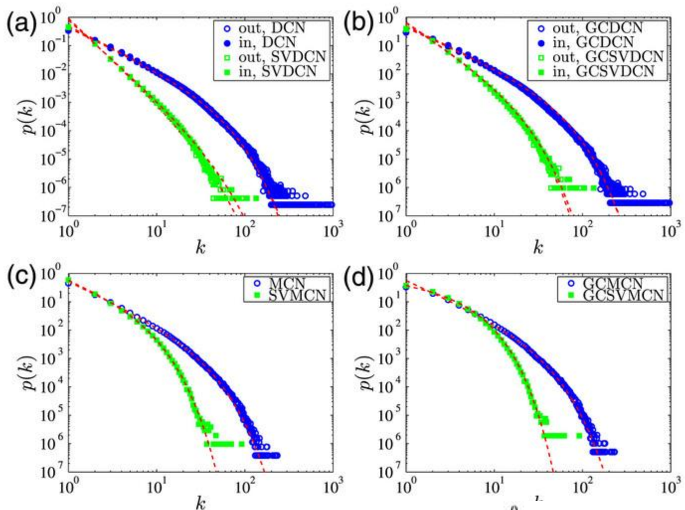
Strength Distribution

- Similarly to degree distribution → weighted networks
  - Deterministic
  - Probabilistic

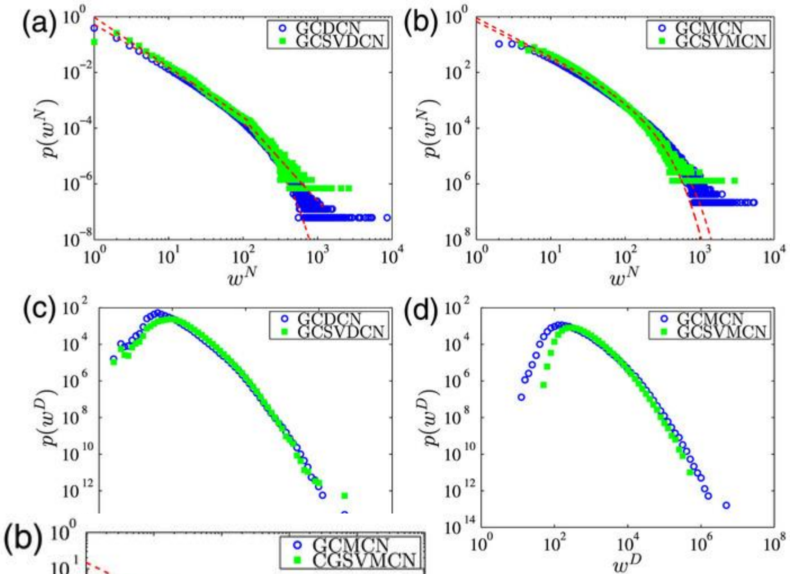
- Does not uniquely determine network topology (nodes, edge structure) !

- But provides an accurate picture of structure-flow combination

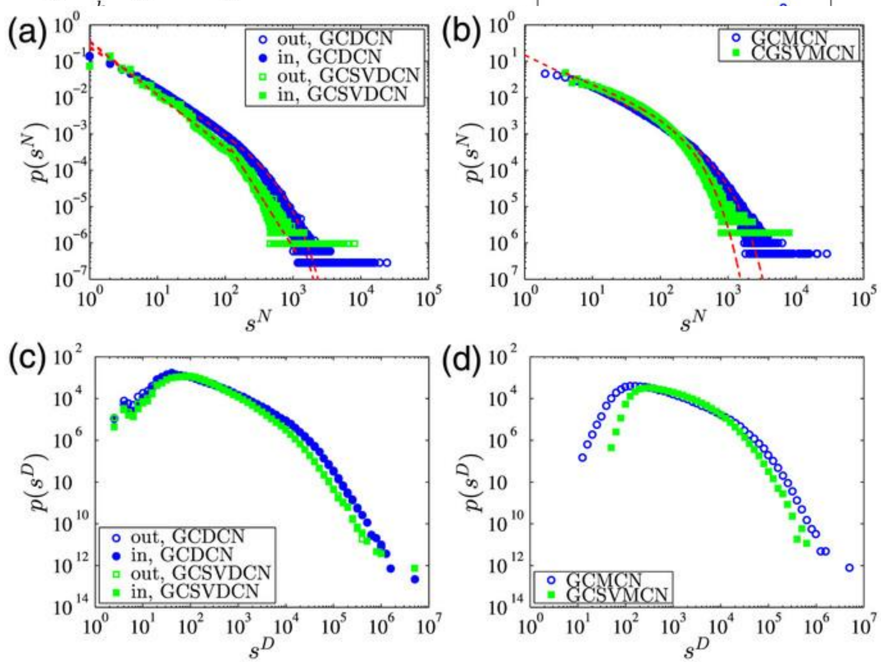
Examples of Node Strength Distributions



Node degree distribution



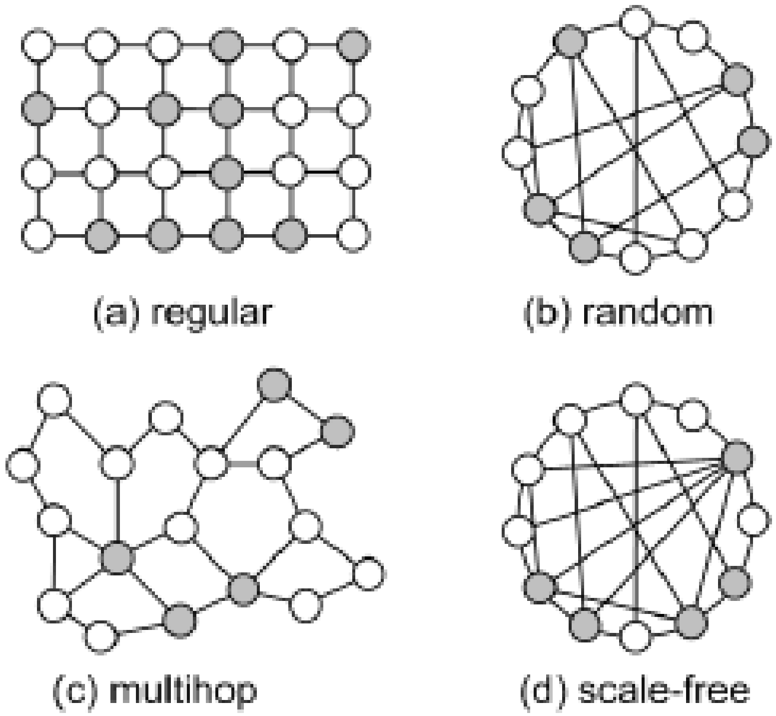
Edge weight distribution



Node strength distribution

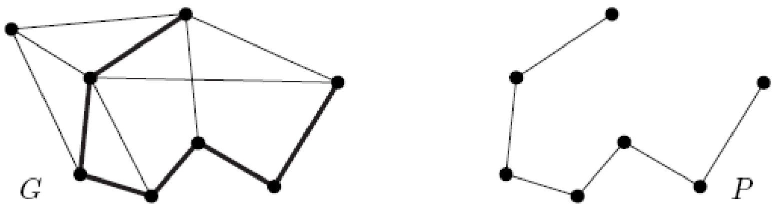
Average Path Length – Intuition

- Different structure leads to different path lengths
- The more clustered the topology, the less the average path length



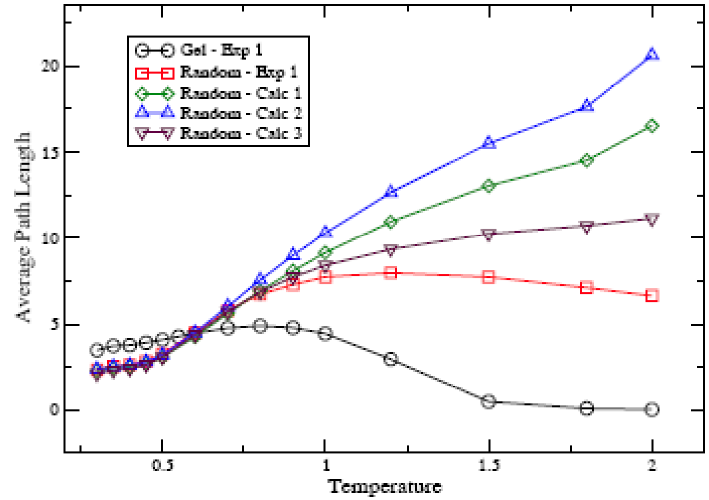
Average Path Length

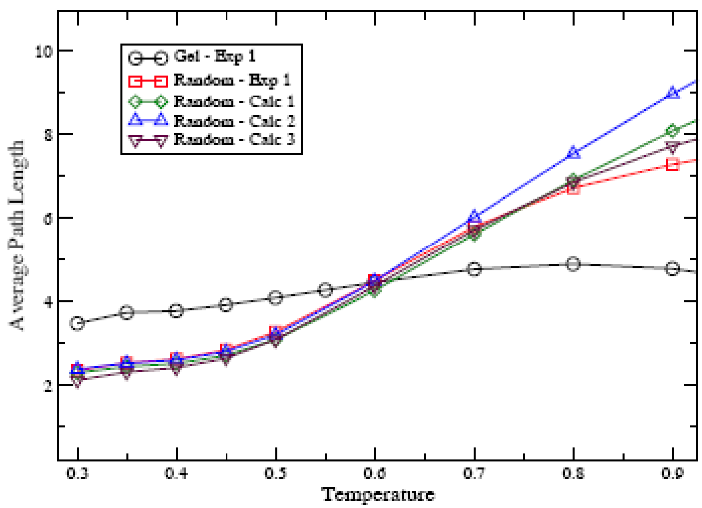
- Path $\rightarrow$ sequence of vertices traversed in a network
- Geodesic path $\leftrightarrow$ shortest path (topology)
  - shortest path through a network from a vertex to another
- Network diameter $\rightarrow$ length of longest geodesic
- Definition of average path length: $l_G = \frac{1}{n*(n-1)} * \sum_{i,j} d(v_i, v_j)$
  - Un-weighted graph
  - Total # nodes is *n*,
  - $d(v_i,v_j)$ geodesic length of v_i from v_j



- The actual path length experienced on average by a user
- The lower L_G is, the better it is in general
  - Information dissemination
  - Lower cost

Examples of Average Path Length





Clustering Coefficient (CC)

- Measure quantifying the ‘grouping’ of the network

- Network type
  - Directed
  - Undirected

  determines the CC definition to be used

- Results can vary according to the employed definition for the same network
  - Need to specify the CC definition used in each case

Clustering Coefficient – Undirected Network

- Measure of degree $\rightarrow$ graph nodes tend to cluster together
  - Global

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$

  - Local

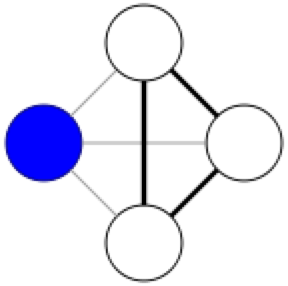
$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E$$

    - quantifies how close its neighbors are to being a clique (complete graph)
    - Degree of node *i* is denoted by k_i
- (Network-wide) Average Clustering Coefficient
  - Average of the local clustering coefficients of all vertices

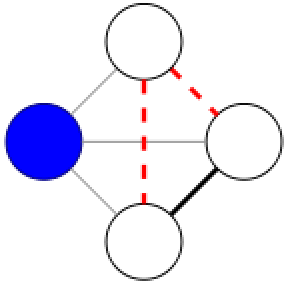
$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i.$$

Clustering Coefficient – Example

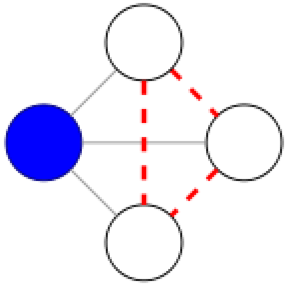
- Local clustering coefficient (LCC) of blue node → proportion of connections among its neighbors actually realized, compared with all possible connections
- The light blue node has three neighbors → max. 3 connections among them
- Top: all three possible connections are realized → LCC = 1
- Middle: only one connection is realized, 2 connections missing → LCC = 1/3
- Bottom: none of possible connections among the neighbors of blue node are realized → LCC = 0



c = 1



c = 1/3



c = 0

Next in SNA ....

- SNA metrics II
  - Centralities
- Complex networks
  - Regular
  - Random
  - Random geometric
  - Scale-free
  - Small-world

+ more applications & examples ....