Fundamental Characteristics of Networks
Models of Random Graphs
Network Application Diagnostics B2M32DSA

Radek Mařík

Czech Technical University
Faculty of Electrical Engineering
Department of Telecommunication Engineering
Prague CZ

October 8, 2017

Outline

The Network Perspective [Weh13]

Mainstream Social Science

- Society is a set of independent individuals.
- Individuals are the unit of analysis, treated as bundles of attributes.

Complex Network Analysis (CNA)

- **Relations** (dyads, triads) are the unit of analysis.
- Actions of **actors** are interdependent.
- **Static**: Structure is (first of all) thought to be a stable pattern.
- **Dynamic**: Choices/actions result in structures, but structures shapes decisions and actions, i.e. processes take place on networks.

Networks Focused on Relations [Weh13]

RELATIONS MATTER!

Contrasted with both an *atomistic* perspective or a *whole-group* perspective

Social Network Analysis (SNA)

- Humanities and social science
- Activities and structures tied with people
  - Shopping basket analysis, targeted advertising
  - Enterprise processes analysis(people cooperation, good distribution)

Complex Network Analysis (CNA)

- Uses the same method as SNA
- Applied to all domains of human acting
- Biology, military, computer network, citations, telecommunication

Network Properties [Weh13]

- A graph $\mathcal{G}$ can be represented as sets or with matrices.
- Properties of vertices $\mathcal{P}$ and lines $\mathcal{W}$ can be measured in different scales:
  - *numerical* (mapped to real numbers),
  - *ordinal* (categorical value with an order), and
  - *nominal* (categorical value with no natural ordering).
- The size of a network/graph is expressed by two numbers:
  - number of vertices $N = |\mathcal{V}|$
  - number of lines $M = |\mathcal{L}|$.

How to Analyze Complex Networks [Erc15]

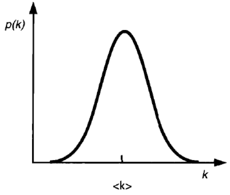
- Determination of what properties to search for.
- Which nodes of the complex networks are more important than others.
- Which groups of nodes are more closely related to each other.
- To see if some subgraph pattern is repeating itself significantly
    - an indication of a fundamental network functionality

Typical Characteristics of Complex Networks [Erc15, Weh13]

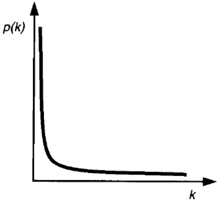
- *Local (node) view*
  - **Degree Heterogeneity**
    - Actors differ in the number of ties they maintain.
    - Centrality measures help to identify prominent actors.
    - Presumption is that nodes or edges that are (in some sense) in the middle of a network are important for the network's function.
  - **Bridges and Small Worlds**
    - New information arrives over weak ties (Granovetter) or bridges (Burt).
    - Bridges tend to be short cuts in the networks,
    - . . . are responsible for short average path lengths.
- *Global (community, structure, network) view*
  - Networks often have dense subgraphs.
  - Community detection helps to find them.
  - **Clusters**
  - **Modularity**
    - Based on a different null models.

Degree Heterogeneity [Weh13]

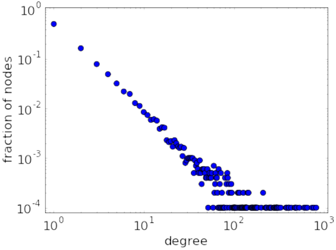
- Not all nodes show the same activity (degree) in networks.
- Some nodes show an astounding activity.
- Degree is most of all a question of tie formation cost.
  - Preferential attachment
  - Fitness model



Gaussian



Skewed Distributions



Vertex Degree Statistics [Erc15]

Theorem 1 (Theorem 4.1 [Erc15], p.64)

For any graph $G(V, E)$, the sum of the degrees of vertices is twice the number of its edges, stated formally as follows:

$$\sum_{v \in V} k(v) = 2M \tag{1}$$

where $k(v)$ is the degree of vertex x.

- The **average degree of a graph**

$$\bar{k} = \langle k \rangle = \frac{1}{N} \sum_{v \in V} k(v) = \frac{2M}{N} \tag{2}$$

Degree Variability [Erc15]

- The **degree variance** $\sigma(G)$ of a graph $G(E, V)$

$$\sigma(G) = \frac{1}{N-1} \sum_{v \in V} (k(v) - \bar{k})^2 \tag{3}$$

- The **mean** of absolute distance between node degrees and the average degress of a graph G

$$\tau(G) = \frac{1}{N} \sum_{v \in V} |k(v) - \bar{k}| \tag{4}$$

Graph Density [Die05, Weh13, Erc15]

- The **density** ρ of a graph is the proportion of present lines to the maximum possible number of lines.
- A **complete graph** is a graph with maximum density.
- There are $\binom{N}{2} = N(N-1)/2$ possible lines (unordered pairs).
- The **graph (edge) density** for *undirected simple* graphs

$$\rho_G = \frac{2|E|}{|V||V|-1)} = \frac{2M}{N(N-1)} = \frac{\bar{k}}{(N-1)} \tag{5}$$

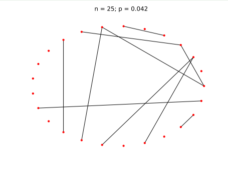
  - for large networks where $N >> 1$, $\rho = \bar{k}/N$
- The **graph (edge) density** for *directed simple* graphs

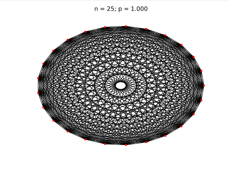
$$\rho_{\vec{G}} = \frac{|E|}{|V||V|-1)} = \frac{M}{N(N-1)} \tag{6}$$

Graph Sparsity [Die05, Erc15]

- The network is called **dense**
  - if ρ does not change significantly as $N \to \infty$ [Erc15], p. 65
  - the number of edges is about quadratic in their number of vertices, i.e. $|E| \approx |V|^2$ [Die05], p. 163
- The network is called **sparse**
  - if $\rho \to 0$ as $N \to \infty$ [Erc15], p. 65
  - the number of edges is about linear in their number of vertices, i.e. $|E| \approx \alpha|V|$ [Die05], p. 164 or $|E| \to$ const. as $N \to \infty$ [New10]
- A dramatic impact on processing of graphs.

A sparse graph and a dense graph with $N = 25$

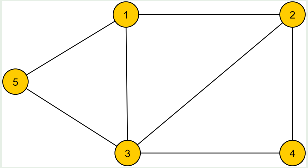


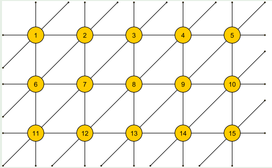


Degree Sequence [Erc15]

- The **degree sequence** of a graph G is the listing of the degrees of its vertices, usually in descending order.
- In **regular graphs** each vertex has the same degree.

Degree Sequence $[4, 3, 3, 2, 2]$





Degree Distribution [Erc15]

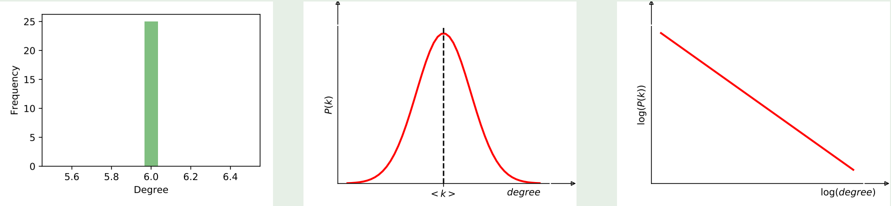
Definition 1 (Definition 3 [Erc15], p.65)

The degree distribution $P(k)$ of degree k in a graph G is given as the fraction of vertices with the same degree to the total number of vertices as below.

$$P(k) = \frac{n_k}{N} \tag{7}$$

where n_k is the number of vertices with degree k.

Degree distributions of regular, random, small-world graphs

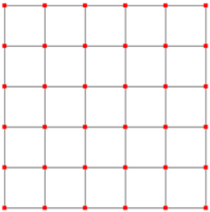


Random Graphs

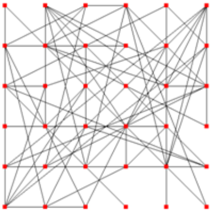
- Basic idea
  - Edges are added at random between a fixed number N of vertices
  - Each instance is a snapshot at a particular time of a stochastic process, starting with unconnected vertices and for every time unit adding a new edge
- Four basic models of complex networks
  - **Regular lattices** (meshes) and trees
  - **Erdös-Renyi Random Graphs** (ER)
    - A disconnected set of nodes that are paired with a uniform probability.
  - Watts-Strogatz Models [WS98] (WS, SW)
    - **Small-world networks**
    - Connections between the nodes in a regular graph were rewired with a certain probability
  - Barabási-Albert Model [BAJ99] (BA, SF)
    - **Scale-free networks** characterized by a highly heterogeneous degree distribution, which follows a "power-law"

$$P(k) \sim k^{-\gamma}$$

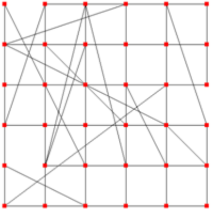
Complex Network Models [GDZ+15]



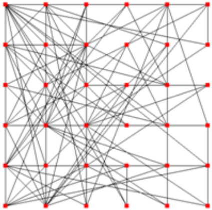
(a) Regular lattice ($p = 0$)



(b) Random network ($p = 1$)

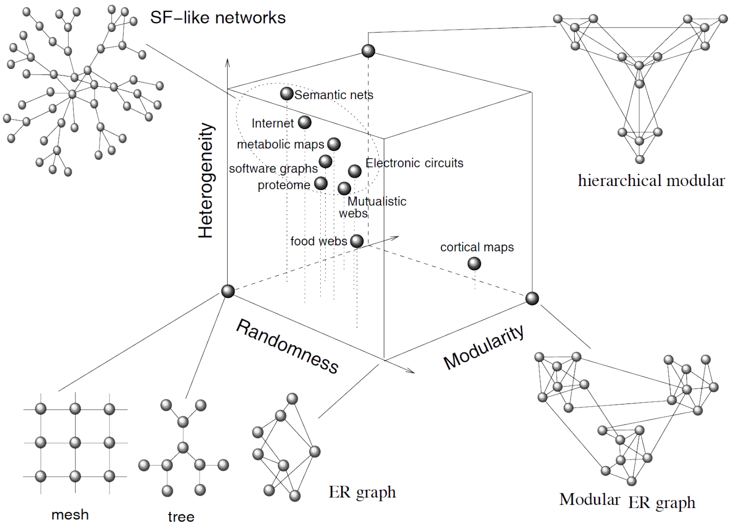


(c) Small-world ($p = 0.01$)



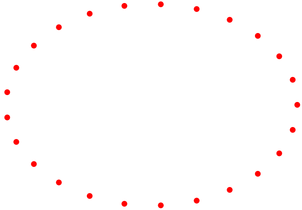
(d) Scale-free ($n_0 = 3$, $m_0 = 3$)

Zoo of Complex Networks [SV04]

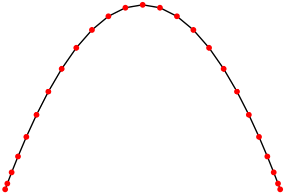


Basic Topologies of Graphs I

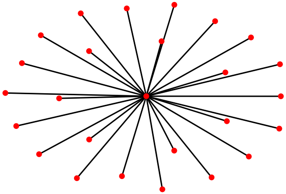
Empty graph: n = 25; m = 0



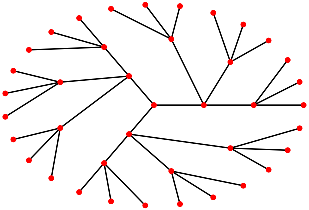
Path graph: n = 25; m = 24



Star graph: n = 26; m = 25

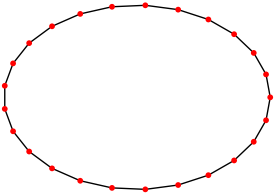


Tree graph: n = 40; m = 39

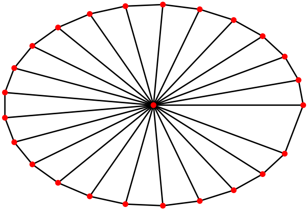


Basic Topologies of Graphs II

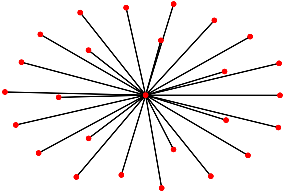
Cycle graph: n = 25; m = 25



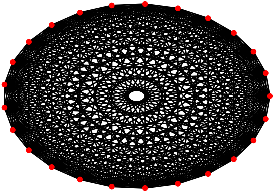
Wheel graph: n = 25; m = 48



Star graph: n = 26; m = 25

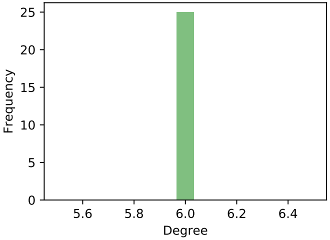


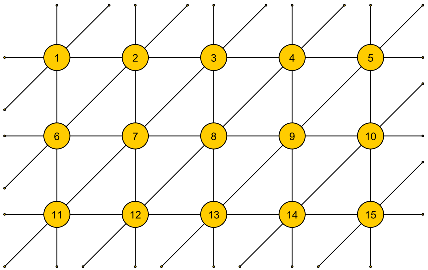
Complete graph: n = 25; m = 48



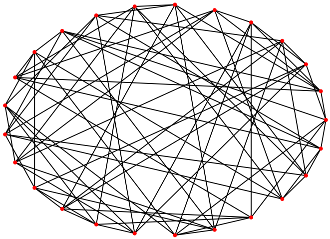
Regular Graph [Erc15]

- All vertices have the same degree.

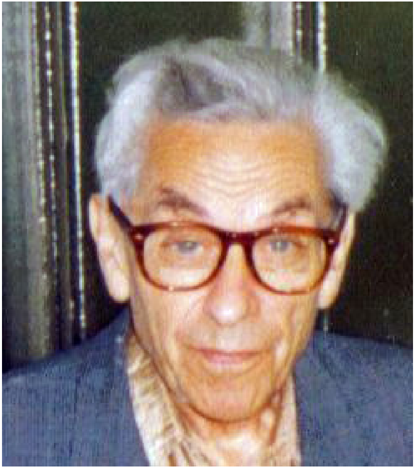




n = 25; d = 6



The Erdös and Renyi Model



Paul Erdös
(1913-1996)

Alfréd Rényi
(1921-1970)

Classical Random Graph (ER-model) [New10, Erc15]

- Proposed by Erdös and Renyi
- Let $G(V, E)$ be a simple graph with n vertices and m edges
- The propability to have an edge between any pair of nodes is distributed uniformly at random.

$$p = \frac{2M}{N(N-1)}$$

- The degree distribution of ER-model is binomial
  - A given vertex is connected with independent probability p to each of the $N-1$ other vertices.
  - The probability of being connected to a particular k other vertices and not to any of the others $p^k(1-p)^{N-1-k}$.
  - There are $\binom{N-1}{k}$ way to choose those k other vertices.
  - The total probability of being connected to exactly k others is

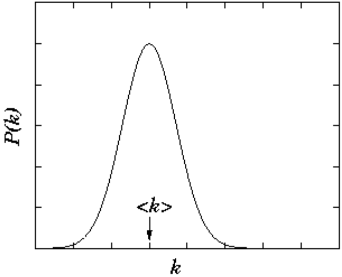
$$p_k = p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

ER-model Properties [New10, Erc15, EA15]

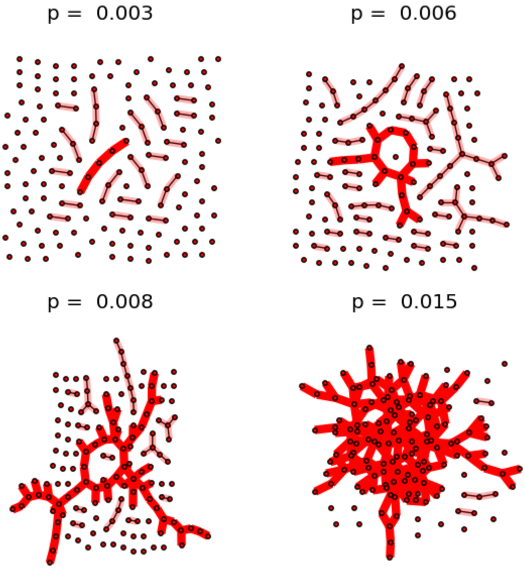
- It does not represent many real complex networks.
- It exhibits
  - homogeneous degree distribution.
  - a small diameter

- Approaching Poisson distribution as $N \to \infty$

$$P(k) \sim e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



ER-model. Giant Component [HSS08, New10]



Six Degree of Separation - Milgram Experiment 1967

- Random people from Nebraska were to send a letter (via intermediaries) to a stock broker in Boston.
- Could only send to someone with whom they were on a first-name basis.
- Among the letters that found the target, the average number of links was **six**.

  six degree of separation [Erc15]

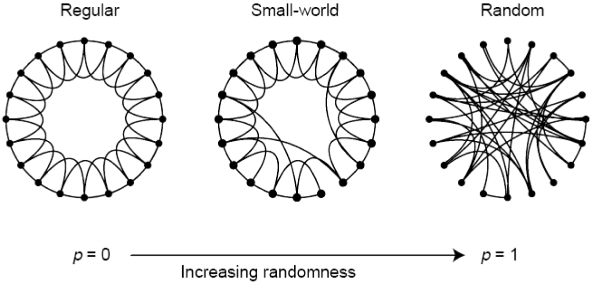
Stanley Milgram (1933 - 1984)

The Watts-Strogatz Model

Duncan J. Watts
(born 1971)

Steven Strogatz
(born 1959)

The Watts-Strogatz Small World Model



- A simple model for interpolating between regular and random networks
- Randomness controlled by a single tuning parameters

The Model

- Take a regular clustered network
- Rewire the endpoint of each link to a random node with probability p

Small World Model - Properties [Erc15, EA15]

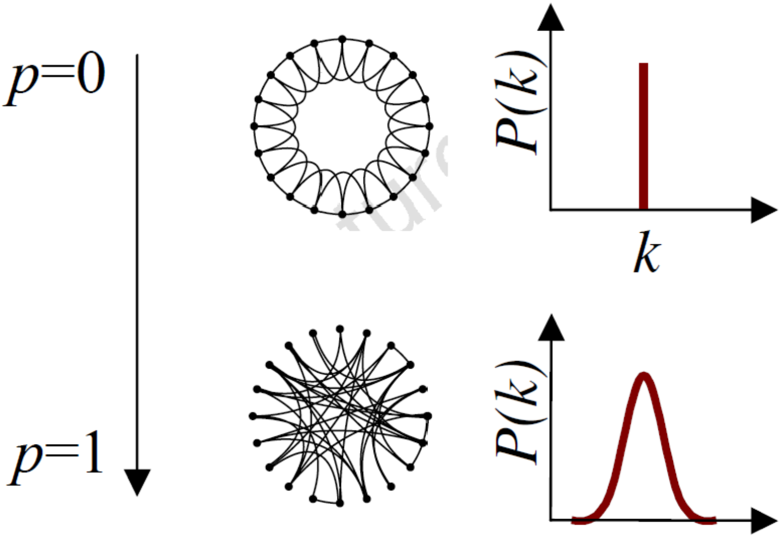
The Watts-Strogatz Model [WS98]

- Starting from the circulant network with n nodes connected to k neighbors.
- The diameter of the network increases with the logarithms of the network order:

$$d \approx \log N \text{ as } N \rightarrow \infty$$

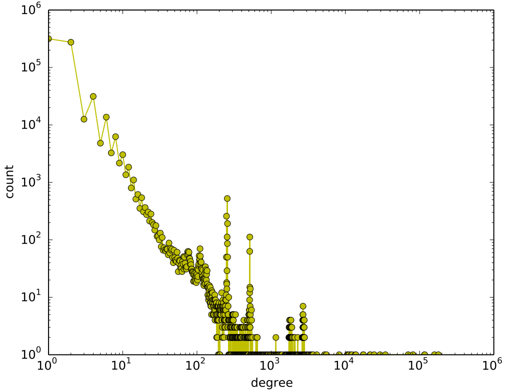
- A high local clustering
  - The starting is a ring topology which each node is connected to its closest $k/2$ left neighbors and $k/2$ right neighbors

Small World Model - Degree Distributions [Erc15, EA15]



Real-world Networks with Fat-tail Distributions [Erc15, EA15]

- Many networks in the real-world have a fat-tailed degree distribution.
- Many real-life complex networks dynamically grow and change by adding and removing nodes and edges.
- Free-scale IP2IP network



The Barabási and Albert Model

Albert-László Barabási
(born 1967)

Réka Albert
(born 1972)

Scale-Free (BA) Network [BAJ99, Erc15, EA15]

Node Degree Distribution

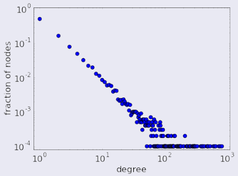
- a heavy-tailed distribution
- follows a power law (asymptotically)

$$P(k) \sim k^{-\gamma}$$

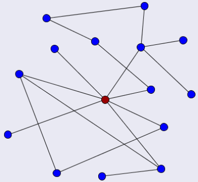
Assumptions:

- Preferential attachment
- Fitness model

Degree Distribution



Small network hub



Barabási-Albert Model [BAJ99, Erc15, EA15]

The outline of the model:

- Begin with a small number, m_0, of nodes.
- At each step, add a new node v to the network, and connect it to $m \leq m_0$ of the existing nodes $u \in V$ with probability

$$p_{uv} = \frac{k_u}{\sum_{w \in V} k_w}$$

Algorithm 1 BA_Generator

```
Input: G(V, E), V_new ... new vertices to joined to G
m_0 ← |E|
for all v ∈ V_new do
    V ← V ∪ {v}
    for m = 0; m ≤ m_0; m++ do
        attach v to u ∈ V with probability P_uv = k_u / Σ_{w∈V} k_w
    end for
end for
```

Scale-Free (BA) Network - Properties [BAJ99, Erc15, EA15]

- **Scale-free property**, c is a constant

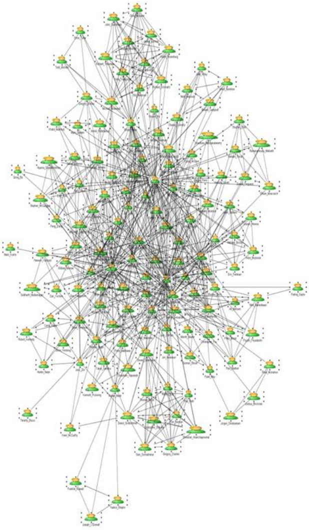
$$p(k) = Ak^{-\gamma}$$
$$p(ck) = A(ck)^{-\gamma} = c^{-\gamma}p(k)$$

- The intercept and the slope is preserved on a logarithmic scale

$$\ln p(k) = -\gamma \ln k + \ln A$$
$$\ln p(ck) = -\gamma \ln(ck) + \ln A = -\gamma \ln(k) + \ln A - \gamma \ln(c)$$

- Degree distribution follows power law, with the exhibition of very few high degree nodes and many low degree nodes. $P(k) \sim k^{-3}$
- The average clustering coefficient of these networks is low due to the large number of low-degree nodes. $C \sim N^{-0.75}$
- The average diameter is low due to the clustering of nodes around the high-degree nodes. $\ell \sim \frac{\ln N}{\ln \ln N}$

Example - Collaboration of People on Projects



Assortativity [New02, New03b]

- the presence of non trivial correlations in network connectivity pattern.
- **Assortative mixing**, or **assortativity**, or **homophily** in SNA (CZ asortativní párování) (i.e., "love of the same") is the tendency of agents to associate and bond with similar others.
  - as in the proverb "birds of a feather flock together"
- **Disassortative mixing** is a bias in favor of connections between dissimilar nodes.
- **Degree correlations** . . . assortativity regarding to node degree.
- **Assortativity coefficient**: vertex is labeled with a scalar value or an enumerative/categorical value (e.g., shape, color) [New02, New03a].

Rich Club [ZM04, CFSV06]

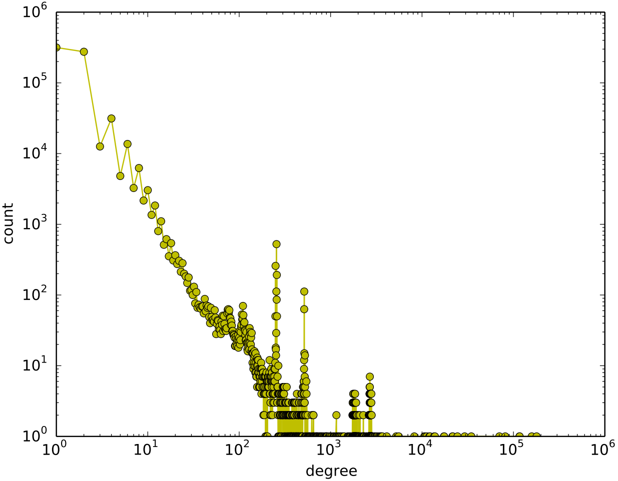
- **Rich-club phenomenon**: Hubs (nodes of high degree) tend to connect to other hubs (rich tends to connect to other rich)
- **Rich-club coefficient** . . . the fraction between the *actual* and the potential number of edges among $V_{>k}$.

$$\Phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}$$

where

- $V_{>k}$ is the set of vertices with degree larger than k,
- $N_{>k}$ is the number of such vertices, and
- $E_{>k}$ is the number of edges among such vertices.

Real-world Networks with Fat-tail Distributions



Summary

- Complex networks basic characteristics
- Topological forms
- Random Network Models
    - Classical Erdos-Renyi model
    - Small world model
    - Scale-free model
- Rich club detection

Competencies

- Describe the network perspective approach to problem solutions.
- What are the typical characteristics of complex networks?
- Describe the meaning of degree hetergeneity.
- Define graph density and sparsity.
- Define graph degree distribution and show some its typical examples.
- List the four basic models of complex networks and their characteristics.
- List basic graph topologies.
- Describe Erdos-Renyi graph model.
- Describe Watts-Strogatz graph model.
- Describe Barabasi-Albert graph model and its scale-free property.
- What is the meaning of “the rich-club phenomenon”.

References I

[BAJ99] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1):173 – 187, 1999.

[CFSV06] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks. *Nat Phys*, 2:110–115, 2006.

[Die05] Reinhard Diestel. *Graph Theory*. Springer, 2005.

[EA15] Ernesto Estrada and Philip A.Knight. *A First Course in Network Theory*. Oxford University Press, 2015.

[Erc15] Kayhan Erciyes. *Complex Networks, An Algorithmic Perspective*. CRC Press, 2015.

[GDZ+15] Dehua Gao, Xiuquan Deng, Qiuhong Zhao, Hong Zhou, and Bing Bai. Multi-agent based simulation of organizational routines on complex networks. *Journal of Artificial Societies and Social Simulation*, 18(3):17, 2015.

[HSS08] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.

[New02] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, Oct 2002.

[New03a] M. E. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, February 2003.

[New03b] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, March 2003.

[New10] M. Newman. *Networks: an introduction*. Oxford University Press, Inc., 2010.

[SV04] Ricard V. Solé and Sergi Valverde. *Information Theory of Complex Networks: On Evolution and Architectural Constraints*, pages 189–207. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[Weh13] Stefan Wehrli. Social network analysis, lecture notes, December 2013.

References II

[WS98] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of /'small-world/' networks. *Nature*, (393):440–442, June 1998.

[ZM04] Shi Zhou and R. J. Mondragon. The rich-club phenomenon in the internet topology. *IEEE Communications Letters*, 8(3):180–182, March 2004.