

Lecture 4: Network Biology: Network Topologies / Parameters

Dynamics of Complex Systems: Week 3

Karthik Raman

Department of Biotechnology



Indian Institute of Technology Madras

<https://home.iitm.ac.in/kraman/lab/>

May 23, 2017

Key Concepts

- ▶ Node/Edge/Edge Weight
- ▶ Density
- ▶ Degree
- ▶ Shortest path/geodesic
- ▶ Diameter
- ▶ Characteristic path length
- ▶ Degree distribution
- ▶ Clustering coefficient
- ▶ Closeness centrality
- ▶ Betweenness centrality
- ▶ Edge betweenness
- ▶ Connected component
- ▶ Strongly connected component in directed graphs
- ▶ Acyclic graphs
- ▶ Motifs

Network Analysis: Network Parameters

- **Connectivity (or) Degree** k , represents number of links the node has to other nodes in the network. For directed networks, $k = k_{in} + k_{out}$
- **Clustering Coefficient** Notion of how connected the neighbours of a given node are (*cliquishness*).

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

- **Characteristic Path Length** Number of edges in the shortest path between two vertices, averaged over all pairs of vertices; measures typical separation between two vertices in the network. Represents the network's overall navigability.
- **Betweenness** Centrality measure of a vertex within a graph.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

σ_{st} — number of shortest paths from s to t ; $\sigma_{st}(v)$ — number of shortest paths from s to t that pass through vertex v . Similarly, edge betweenness.

Network Biology Overview

- ▶ Concept of networks: Nodes, edges, directed/un-directed/weighted networks, adjacency matrix, shortest path
- ▶ Network parameters: degree [distribution], diameter, characteristic path length, clustering coefficient, measures of centrality (betweenness)
- ▶ Topologies: random, small-world, scale-free/power-law

Further reading: **Barabasi AL & Oltvai ZN** (2004) *Nature Reviews Genetics* **5**:101–113

Why do we need network models?

- ▶ Create models that mimic real-world networks, based on theory
- ▶ Or rather, discover generative models for real-world networks, based on theoretical principles
- ▶ ‘Bin’ our observations on real-world networks into the *canonicals*
- ▶ Improve understanding of real-world networks
- ▶ And also the ability to analyse/simulate them, predict behaviours ...

Random networks

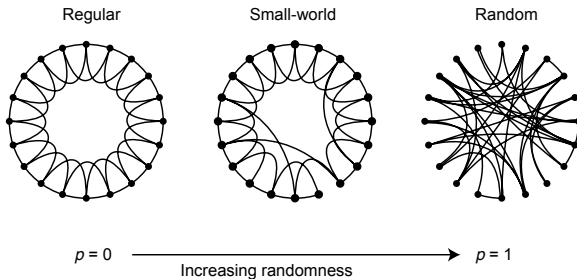
Erdős–Rényi (ER) random networks

ER random network model

- ▶ Starts with N nodes and
 - ▶ connects each pair of nodes with a probability p ,
 - ▶ which creates a network with approximately $pN(N - 1)/2$ randomly placed links
-
- ▶ Node degrees follow a *Poisson* distribution — most nodes have approximately the same number of links
 - ▶ Characteristic path length is proportional to the logarithm of the network size $L \sim \log N$
 - ▶ $C(k)$ (clustering coefficient of nodes with k links) is independent of k

Small-world networks

Watts-Strogatz model: **Watts DJ & Strogatz SH (1998) *Nature* 393:440–442**



For a large region between $p = 0$ and $p = 1$, we find

- ▶ high clustering, and
- ▶ low path lengths

In both ER random graph and the small-world model, connectivity distribution peaks at an average value and decays exponentially. Such networks are called “exponential networks” or “homogeneous networks”, because each node has about the same number of links.

Small-world networks

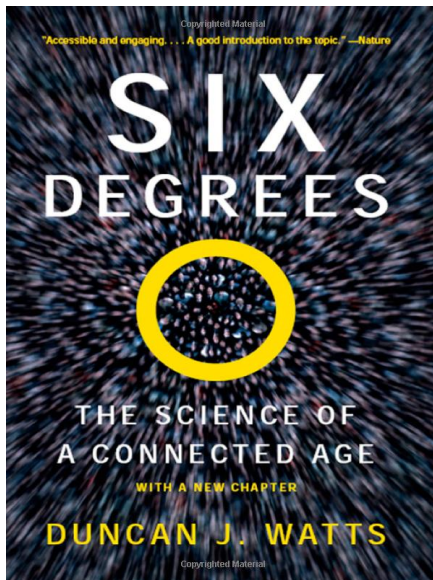
- ▶ Small-world networks have a characteristic path length of the same order as random networks ($L \gtrsim \log N$), but have $C \gg C_{\text{rand}}$

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
<i>C. elegans</i>	2.65	2.25	0.28	0.05

Characteristic path length L and clustering coefficient C for three real networks, compared to random graphs with the same number of vertices (n) and average number of edges per vertex (k). (Actors: $n = 225,226$, $k = 61$. Power grid: $n = 4,941$, $k = 2.67$. *C. elegans*: $n = 282$, $k = 14$.) The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component¹⁶ of this graph, which includes $\sim 90\%$ of all actors listed in the Internet Movie Database (available at <http://us.imdb.com>), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction. We treat all edges as undirected and unweighted, and all vertices as identical, recognizing that these are crude approximations. All three networks show the small-world phenomenon: $L \gtrsim L_{\text{random}}$ but $C \gg C_{\text{random}}$.

Small-world networks



Power-law networks

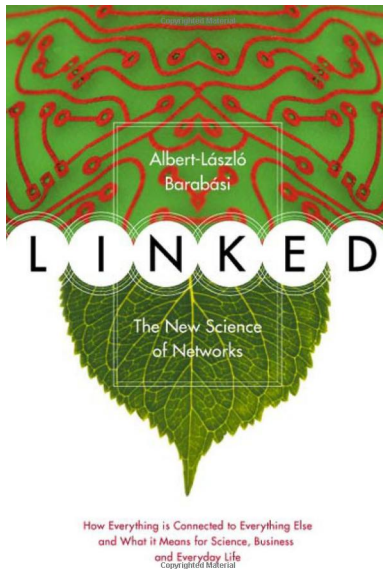
Barabasi–Albert power-law model: **Barabási AL & Albert R (1999) *Science* 286:509–512**

- ▶ Characterised by a power law degree distribution

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

- ▶ Such networks can be generated using the “preferential attachment”, or “rich get richer” model
- ▶ Networks generated by this model have a power-law degree distribution characterised by $\gamma = 3$
- ▶ Scale-free networks with $2 < \gamma < 3$, a range commonly observed in many biological networks, are *ultra-small*, with a characteristic path length $L \sim \log \log N$, significantly smaller than that of random networks ($\log N$)

Power-law networks

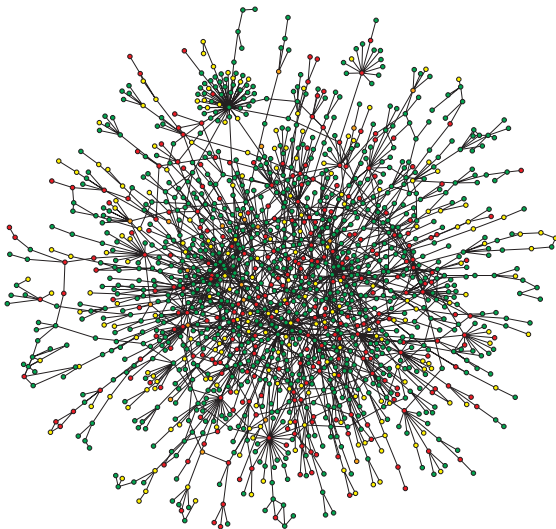


Small-world vs. Power-law

Small-world	Power-law
egalitarian	hubs and resources
no growth	preferential attachment
local clusters and distant links	power law
under attack: deteriorates	under random attack: stable
no particular targets	under targeted attack: weak
Small average node-to-node distance	

Yeast Protein–protein Interaction Network

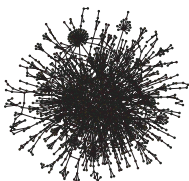
Jeong H et al. (2001) *Nature* 411:41–42



Largest cluster, containing $\approx 78\%$ of all proteins, is shown. Colour of a node signifies the phenotypic effect of removing the protein: red, lethal; green, non-lethal; orange, slow growth; yellow, unknown

Yeast Protein–protein Interaction Network

Jeong H et al. (2001) *Nature* 411:41–42



- ▶ Although proteins with ≤ 5 links constitute about 93% of the total proteins, only $\approx 21\%$ of them were essential
- ▶ On the other hand, only 0.7% of the proteins have > 15 links, but single deletion of 62% of these proves lethal
- ▶ This implies that highly connected proteins with a central role in the architecture of the network are *three times more likely to be essential* than proteins with only a small number of links to other proteins
- ▶ This observation holds across many organisms — for degree and betweenness centrality^a
- ▶ Biological networks are also found to be *disassortative*: mix with different types of nodes (in terms of degree)

^aRaman K et al. (2014) *Systems and Synthetic Biology* 8:73–81

Perturbations to Networks

- ▶ Networks can be perturbed the removal of nodes/edges
- ▶ Typical analysis: disrupt a node and its corresponding edges
- ▶ Networks of different topologies vary in their resilience to various types of perturbations

Quantifying disruptions

- ▶ How do we quantify damage/disruption to a network?
- ▶ Shortest paths disrupted?
- ▶ Pairwise disconnectivity index: fraction of those initially connected pairs of vertices in a network, which become disconnected if vertex v is removed from the network^a

^a Potapov A et al. (2008) *BMC Bioinformatics* **9**:227+

