Biological Networks

Latifa LECHLECH

Abstract

In this particular programming assignment, our objective revolved around delving into the intricate concepts of Biological networks. We aimed not only to comprehend these networks but also to explore the algorithms specifically designed for their creation and analysis.

1 Introduction

The computational study of biological systems is characterized by analyzing the relationships between biological factors rather than focusing solely on the biological factors themselves. Biological functions stem from a complex, redundant, and robust network of interactions among molecules. This network processes information, makes sophisticated decisions, and thrives in a dynamic environment.

The term "biological networks" describes the interdependent systems found in living things; these systems are sometimes shown as intricate webs made up of molecules, genes, proteins, or other biological components. comprehension of diverse biological processes, such as cellular activities, signaling routes, and connections between multiple parts inside an organism, requires a comprehension of these networks. There are several types of biological networks:

- Gene Regulatory Networks (GRNs): GRNs illustrate how genes control the expression of other genes by connecting genes and their regulatory components. They shed light on the mechanisms of genetic regulation and the interactions between various genes throughout cellular functions.
- Protein-Protein Interaction (PPI) Networks: These networks show the physical interactions that take place between proteins inside a cell. Comprehending PPI networks facilitates the interpretation of cellular mechanisms, discernment of functional modules, and anticipation of protein activities.
- Metabolic Networks: These networks describe the chemical processes that take place within cells or organisms They demonstrate how metabolites move through linked enzymatic processes, giving insight into the transformation and use of chemicals in biological systems.
- Signaling Networks: involve the transmission of information within and between cells. Signaling pathways transmit messages from external stimuli, including growth factors or hormones, to intracellular elements, therefore initiating distinct physiological reactions. Frequently, they entail a

series of molecular interactions that result in modifications to gene expression, metabolism, or cell activity. The main roles of signaling pathways are information transmission and cellular response coordination, while they can also influence metabolic processes by controlling enzyme activity or gene expression.

- Neural Networks: These networks are used in neuroscience to depict the connections that exist between brain neurons. They are essential to comprehending behavior, learning, and brain function.
- Phylogenetic networks: These illustrate the divergences and ancestral linkages that exist between various species or groups of animals in their evolutionary relationships.

The many biological networks mentioned below are models that are used for investigating and analyzing various biological systems, from chemical interactions inside cells to higher-order functions like brain activity and links between species in evolution.

2 Representation of biological systems as networks

Representing biological systems as networks, often termed as network biology or systems biology, involves using graph theory to model the interactions within biological entities.

A simple graph G consists of a non-empty finite set $V(G)$ of elements called vertices (or nodes), and a finite set $E(G)$ of distinct unordered pairs of distinct elements of V(G) called edges. We call $V(G)$ the vertex set and $E(G)$ the edge set of G. An edge v, w is said to join the vertices v and w, and is usually abbreviated to vw.[16]

In the context of representing biological systems:

- Nodes represent biological entities such as molecules (e.g., proteins, genes, metabolites), cells, tissues, organs, or organisms.
- Edges represent the interactions or relationships between these biological entities. For instance, in PPI networks it represents the interactions between proteins. These interactions can take various forms, such as physical binding, enzymatic reactions, signaling cascades, or complex formation. interactions between proteins. These interactions can take various forms, such as physical binding, enzymatic reactions, signaling cascades, or complex formation.

2.1 Examples

2.1.1 Gene circuitry in sea urchin development

Davidson and his team utilized a combination of experimental techniques and computational modeling to unravel the intricacies of GRNs in sea urchin embryos. Some of their significant contributions include experimental Approaches, computational modeling, establishing the Sea Urchin model and identification of Key regulatory genes.

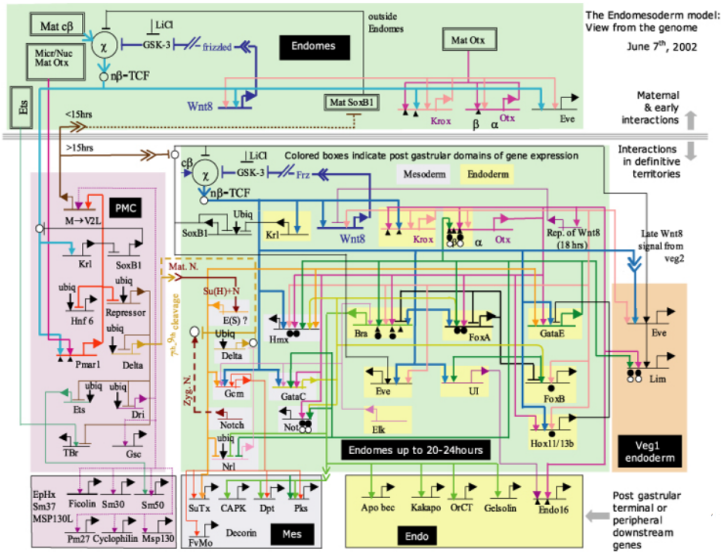


Figure 1: Gene regulatory network for the early development of the sea urchin embryo The circuit [15]

2.1.2 Gene homology network

2.1.3 Yeast interaction network

was presented by H. Jeong, S.P. Mason, A.-L. Barabasi, and Z.N. Oltvai in their ground-breaking 2001 publication that was published in Nature. Their study, "Lethality and centrality in protein networks," revealed surprising organizing principles inside cellular networks, which profoundly altered our knowledge of biological systems. The link between a protein's centrality in a network and its need for an organism's survival was investigated in this study. Through analysis, the Saccharomyces cerevisiae (baker's yeast) protein-protein interaction network, scientists were able to show that highly linked proteins, or hubs, were more likely to be critical to the organism's survival. This discovery disproved accepted wisdom regarding network resilience and offered vital background information for comprehending the interaction between biological functionality and network topology. It also shed light on the importance of core proteins for preserving cellular viability and function.

In this graph (see 2.1.3) representation nodes represent proteins and physical interactions are represented by edges

2.2 Types of Networks

2.2.1 Exponential Networks

Exponential networks are characterized by a rapid increase in connections between nodes following an exponential pattern. Within these networks, the number of connections grows swiftly as new nodes join, resulting in a densely connected structure. Notably, there's a prevalent 'average connectivity' among

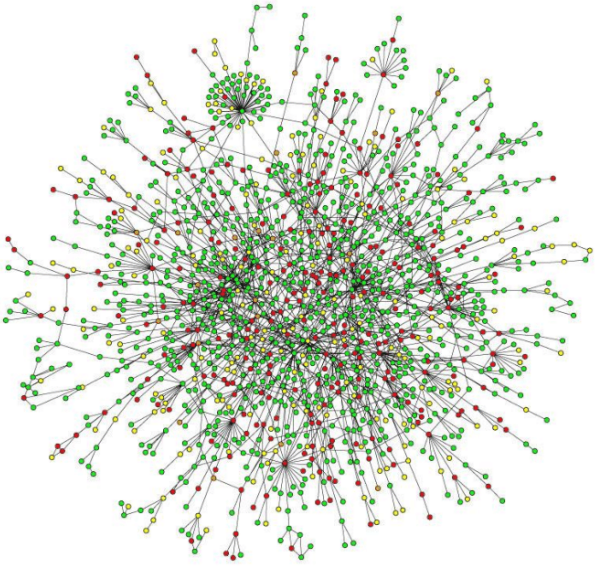


Figure 2: 2 Network diagram showing a map of protein-protein interactions in a yeast (Saccharomyces cerevisiae) cell. This cluster includes 78 percent of the proteins in the yeast proteome. The color of a node represents the phenotypic effect of removing the corresponding protein (red, lethal; green, nonlethal; orange, slow growth; yellow, unknown) [2]

nodes, where most nodes share a similar number of links, and only a small fraction deviates from this average degree.

2.2.2 Scale-Free Networks

Networks with **power-law distributions** a called scale-free [5] due to their consistent distribution across all scales. This distribution, is illustrated by::

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

where γ is the exponent defining the power-law distribution.It has distinct effects based on its magnitude:

- For $\gamma < 2$: The number of links grows faster than the number of nodes and they naturally possess the small world property, because the diameter increases by the logarithm of the size of the network and the clustering coefficient is finite.[14]
- For $\gamma = 2$: the emergence of a "hub-and-spoke" pattern occurs. Where a few key nodes play a central role in the overall connectivity of the network.
- For $2 < \gamma < 3$: the network maintains a power-law distribution of degrees, and it retains a hierarchical structure of hubs.
- For $\gamma \geq 3$ the network displays a heavily skewed distribution towards a few highly connected hubs, resulting in an absence of many typical characteristics.

In these networks, a small number of nodes referred to as **hubs** have a disproportionately large number of connections as compared to the other nodes in the network, whereas the majority of nodes have a comparatively low number of connections.

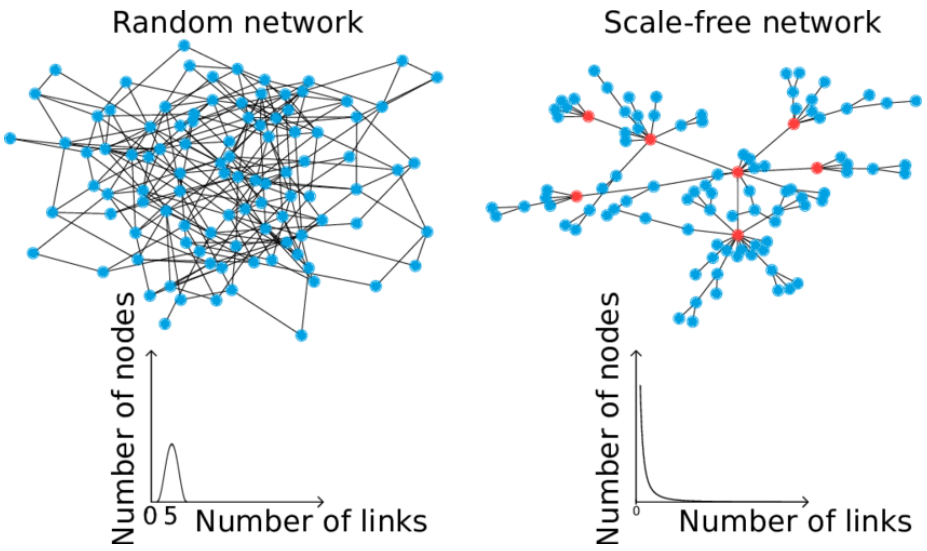
2.2.3 Random Networks

A random graph model is a mathematical representation of a network in which nodes are connected by random processes. Typically, nodes in a random graph are connected with a certain probability, producing a network structure with random or probabilistic characteristics. The degree distribution often approximates a Poisson distribution when the number of nodes in the network is sufficiently large.

3 Network proporities

Network properties refer to the characteristics and behaviors exhibited within different types of networks. These properties can include:

- **Connectivity**: refers to the degree to which nodes within a network are connected to each other.



- **Centrality** indeed measures the relative importance of nodes within a network. A node's significance inside a network may be measured using a metric called **betweenness centrality**, which takes into account how well it influences or facilitates information flow between other nodes. Several shortest pathways in the network connect nodes with high betweenness centrality. Since they serve as vital links or bridges between various network nodes, nodes with high betweenness centrality have a substantial impact on information flow. It is computed by taking the percentage of all pairs of nodes' shortest routes that pass through a certain node, indicating how crucial that node is to the network's ability to communicate and interact. The betweenness centrality of a node v is then defined by the following formula:

$$BC(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- **Transitivity or Clustering Coefficient**:The clustering coefficient reveals how much a network's nodes prefer to group together. A high clustering coefficient indicates that there is a greater likelihood of nodes in the network creating groups or clusters through connections with their neighbors.

  The clustering coefficient C is defined as the average probability that two neighbors of a node are connected to each other. It is calculated as follows:

$$C_v = \frac{2E_v}{d_v(d_v - 1)}$$

  where E_v is the number of edges between neighbors of v. d_v is the number of neighbors of node v. v is a node in the network.

- **Community structure** In Fig. 8 we show a visualization of the friendship network of children in a US school taken from a study by Moody [13]

  The figure was generated using a "spring embedding" algorithm, employing linear springs between vertices and a first-order energy minimization to relax the system.

- **Scale-Free Nature**: Some networks exhibit a scale-free topology.As a result, a hierarchical structure is created in which a small number of nodes significantly influence the behavior of the entire network.

- **Small-world Phenomenon**: assumes that all individuals are related to all others indirectly through a finite number of intermediates. Milgram popularized the concept. who performed a well-known experiment involving letter delivery between strangers in different regions of the United States. Using only local knowledge of immediate acquaintances, each participant was asked to forward the letter to a well-known acquaintance selected to bring it closer to the ultimate destination.[6]The experiment demonstrated that in large social networks, there are short path between five and six, in Milgram's experiment and that people can locate these paths by utilizing only local knowledge. Most node pairs within a 'small world' network appear to be connected by a short path. Where the geodesic distance is calculated as follow:

  $$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i>j} d_{ij}$$

  where d_{ij} is the distance between node i and j.

- **Resilience and Robustness**: The ability of a network to maintain its structure and functionality despite random failures or deliberate attacks. Random choice in network robustness involves randomly removing nodes or links, and simulating natural failures. On the other hand, attacks aim to interfere with the operation of the network by specifically targeting high-degree or particular nodes. See Figure 4.1.2 where blue plot refers to random choice and red refers to Directed attacks.

- **Dynamics and Evolution**: Networks can change over time, with nodes being added or removed, links changing or rewiring, and the overall structure evolving. Comprehending the evolution of networks has significant importance in many domains such as social dynamics, diseases, and technical advancements.

4 Graph models

4.1 Random graphs

4.1.1 Erdös–Rényi (ER) model

The Erdős–Rényi model is a mathematical concept used to generate random graphs. These models are named after Hungarian mathematicians Paul Erdős

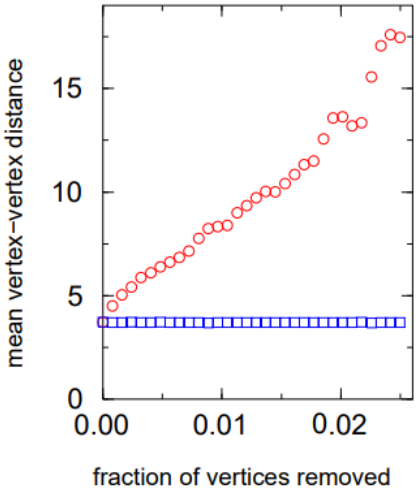


Figure 3: Mean vertex–vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices (circles), then distance increases sharply. After Albert et al.

and Alfréd Rényi, who introduced one of the models in 1959.[4] [8] It connects nodes (vertices) within a graph by either establishing connections based on probability p. As a result, this graph generates approximately $\frac{pN(N-1)}{2}$ randomly distributed edges.

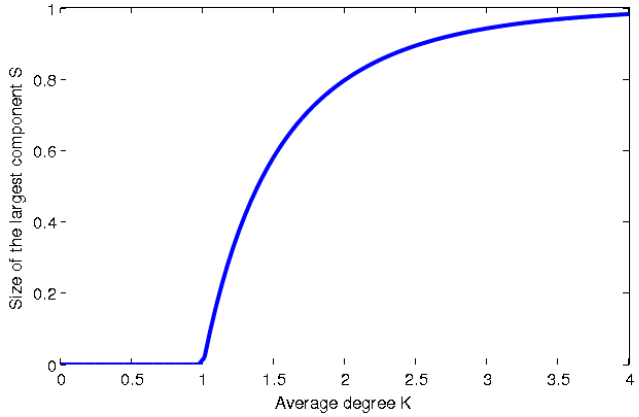
Over degree k, the Erdos-Renyi Model has a binomial distribution.

$$P(k) = \begin{pmatrix} n-1 \\ k \end{pmatrix} p^k (1-p)^{n-1-k}$$

The Erdős–Rényi model possesses several key properties:

- **Randomness**
- **Phase Transition**: at m = n/2, a significant change takes place. At this moment, a number of clusters spontaneously collapse, creating the "giant component," which is one huge component. Compared to the other components, this massive component has a substantially bigger collection of nodes, usually O(n) nodes, whereas the second-largest component has about O(log n) nodes. In the field of statistical physics, this phenomenon is known as percolation.
- **Degree Distribution**: follows a Poisson distribution. As it can represent approximately the binomial distribution in the scenario when $np \to \lambda$, we assign $p = \frac{\lambda}{n}$ and claim the final outcome that the Erdos–Renyi random graph's degree distribution is Poisson with parameter λ:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-k}$$



- **Clustering coefficient** is constant and low. This means that nodes in this random graph model are less likely to form clusters or triangles compared to many real-world networks.

$$C(v) = \frac{e(v)}{\frac{1}{2}k(v)(k(v)-1)} = \frac{p \times \frac{1}{2}k(v)(k(v)-1)}{\frac{1}{2}k(v)(k(v)-1)} = p = \frac{\bar{k}}{n-1} \approx \frac{\bar{k}}{n}$$

  And the clustering coefficient for a random graph $G(n,p)$ becomes

$$\tilde{C}(G) = \frac{1}{n}\sum_{v \in N} C(v) = p = \frac{\bar{k}}{n-1} \approx \frac{\bar{k}}{n}$$

- **Short Average Path Length** are often short, this characteristic is a key feature of small-world networks.

$$h \approx \log_{\bar{k}}(n) = \frac{\log(n)}{\log(\bar{k})} = \frac{\log(n)}{\log(p(n-1))}$$

In bioinformatics networks often exhibit non-random, specific interactions and complex structures, the reason why the Erdős–Rényi (ER) Model might not fully mirror the complexities found in biological systems due to its assumption of random and uniform connections between nodes.

4.1.2 Barabási–Albert model

The Barabási–Albert (BA) model is a mathematical model used to explain the emergence of complex networks, particularly scale-free networks. It was proposed by Albert-László Barabási and Réka Albert in 1999. This model uses a preferential attachment mechanism to create random scale-free networks.

- Starting with a small network
- Place a new node next to the network
- Draw a fixed number m links to the exisiting nodes, with probability of linking to a node proportional to that node's degree.

$$\pi_i(k_i) = \frac{k_i}{\sum_{i=1}^{N-1} k_i}$$

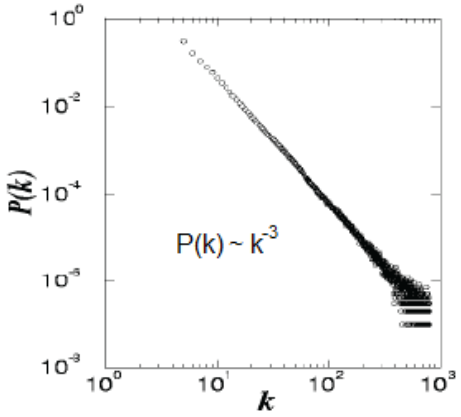


Figure 4: Power law degree distribution of a Barabasi Albert network model Barabasi and Albert

- Repeat steps 2-3 at each time step

The main Barabási–Albert model properties are:

- **Sparse graph**: the number of edges is significantly smaller compared to the maximum possible number of edges between the nodes.
- **Having multiple cliquish**: a cliquish refers to subsets of nodes within a graph where every node is directly connected to every other node in that subset.
- **Degree Distribution**: follows a power law distribution. It is a **Scale-Free** model
$$P(k) = k^{-3}$$
- **Small Average Path Length**: the average distance between two nodes in a BA graph increases approximately logarithmically with the size of the network.
$$l \sim \frac{ln}{lnlnN}$$
- **Clustering coefficient**: An analytical result for the clustering coefficient of the BA model [12] was obtained by Klemm and Eguíluz and proven by Bollobás. [1] A mean-field approach to study the clustering coefficient was applied by Fronczak, Fronczak and Holyst. [11]

  In small-world networks, clustering is independent of system size; this behavior is nonetheless different from theirs. Within hierarchical networks, a power-law also drives clustering as a function of node degree.
$$C(k) = k^{-1}$$
  This result was obtained analytically by Dorogovtsev, Goltsev and Mendes. [9]

- **Growth and Preferential Attachment**: The model describes how networks develop over time, mirroring the processes of growth observed in many actual networks, including the development of social networks and the World Wide Web.
- **Robustness**: The BA model produces scale-free networks that are resilient to random node failures but vulnerable to deliberate attacks on highly linked nodes. Because there are hubs to sustain connectivity, removing random nodes is less likely to upset the overall structure. Targeted hub attacks, however, may severely damage the integrity of the network.

While the Barabási-Albert (BA) Model is valuable in understanding scale-free networks and preferential attachment, its limitations arise in bioinformatics due to oversimplified assumptions. The broad and complex connections between biological components like genes or proteins, which may not always exactly follow preferred attachment principles, are not well captured by the rigorous commitment to a preferential attachment that the BA model requires in biological systems. The application of the BA model in bioinformatics to capture the subtle complexity of biological interactions and network formations is limited since biological networks frequently display multiple types of connection and regulatory processes that reach beyond preferential attachment.

4.2 Hierarchical network model

Iterative approaches for building networks that can replicate both the high node clustering and the special characteristics of the scale-free topology. These traits are frequently seen in nature, including biology, language, and certain social networks.

The Hierarchical network model properties are:

- **Degree distribution** follows the power law.

  $$P(k) = C.k^{-\gamma}$$

  where c is a constant and is the degree exponent. In most real world networks exhibiting scale-free properties lies in the interval [2,3].[7] As a specific result for hierarchical models it has been shown that the degree exponent of the distribution function can be calculated as

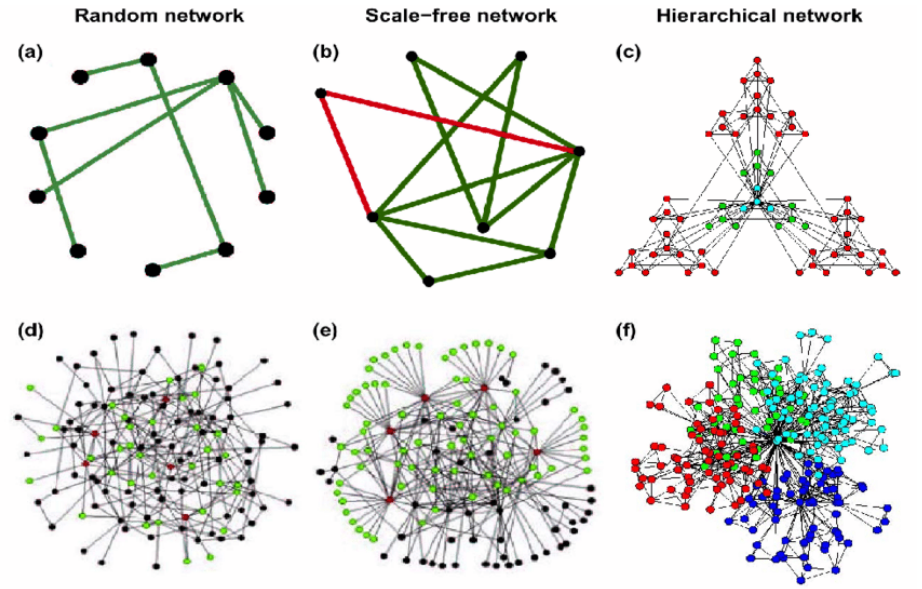
  where M represents the replication factor of the model

- **Clustering coefficient**: the clustering coefficient can be expressed as a function of the degree

  $$C(k) \sim k^{-\beta}$$

  It has been analytically shown that in deterministic scale-free networks the exponent β takes the value of 1. [10]

Hierarchical Network Models face limitations in bioinformatics due to the dynamic and interconnected nature of biological systems, making it challenging to create accurate representations that encompass the complex relationships across multiple levels.



[3]

5 Conclusion:

The computational analysis of biological data is a common topic in bioinformatics. Network models such as Hierarchical, Barabási-Albert, and Erdős–Rényi (ER) are frequently used to represent and comprehend biological systems.

In bioinformatics, the "best" model differs according on the particular biological system or phenomena being studied. Various models address different areas, including metabolic pathways, protein interactions, and gene regulation. The model must correspond to the goals of the research and properly and efficiently capture the intricacies of the biological system under study in order to be selected.

References

[1] CiteSeerX.

[2] Figure 1.2 network diagram showing a map of protein-protein...

[3] Figure 3. graphical representation of three network models: (a) and (d)...

[4] On random graphs. i. | semantic scholar.

[5] Rev. mod. phys. 74, 47 (2002) - statistical mechanics of complex networks.

[6] Small-world phenomenon - an overview | ScienceDirect topics.

[7] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. 286(5439):509–512. Publisher: American Association for the Advancement of Science.

[8] Béla Bollobás. *Random graphs*. Number 73 in Cambridge studies in advanced mathematics. Cambridge University Press, 2nd ed edition.

[9] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Pseudofractal scale-free web. 65(6):066122. Publisher: American Physical Society.

[10] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Pseudofractal scale-free web. 65(6):066122. Publisher: American Physical Society.

[11] Agata Fronczak, Piotr Fronczak, and Janusz A. Holyst. Mean-field theory for clustering coefficients in barabasi-albert networks. 68(4):046126.

[12] Konstantin Klemm and Víctor M. Eguíluz. Growing scale-free networks with small-world behavior. 65:057102. ADS Bibcode: 2002PhRvE..65e7102K.

[13] Manuel Lima. visualcomplexity.com | school friendship network.

[14] Avi Ma'ayan. Introduction to network analysis in systems biology. 4(190):tr5.

[15] Paola Oliveri and Eric H Davidson. Gene regulatory network controlling embryonic specification in the sea urchin. 14(4):351–360.

[16] Robin J. Wilson. *Introduction to graph theory*. Prentice Hall, 4. ed., [nachdr.] edition.