Hierarchy and Specificity: An Attempt at Generating Realistic Social Networks

Syed Mohammad Ali
York University
North York, Canada
saadali1@my.yorku.ca

Abstract

Using the work of predecessors such as Erdős–Rényi, Barabási–Albert, and Watts-Strogatz and some innovation to generate a new algorithm for realistic social networks. An attempt to create a new foundational model that can simultaneously possess small short-path lengths, high clustering, and power law distributivity which reflects realistic social networks. The goal is to make the algorithm as simple with as few parameters as possible, which allows it to be foundational like its predecessors, although there are modern-day approaches able to possess all 3 foundational properties. This is done through the use of the UALIS model that uses the Colour Combination Algorithm; nodes have specificity alongside preference when deciding which nodes to link to. For comparison, the foundational models were generated with similar parameters and a real network from Stanford SNAP's database was used. The results were that 2 out of 3 of the foundational properties were reflected within the graph generated by the algorithm: high clustering and small short-path lengths. Power-law distribution could not be attained however there are promising results within different iterations of the code that show that it is plausible. Next steps would be to refine the parameterization and generalize the algorithm, as well as optimize the algorithm to produce all 3 of the foundational properties. Other steps include finalizing the list of key elements that result in a realistic social network. A theoretical algorithm discussed in this paper that builds off of the UALIS model and is a complex version of the simple algorithm is the humanity graph, which focuses on real-time social networks instead of snapshots, and creates a foundation for generating networks of agentic AI for problem-solving and observing the evolution of social networks.

Keywords

ualis, colour combination collision, network science, graph theory, network algorithms, community detection, network modeling, realistic social networks

1 Introduction

1.1 Motivation

The search for realistic network models has advanced considerably since the 1900s. With random models, small-worlds, and scale-free networks, a sizable foundation has been created for modeling realistic networks. Modern approaches include: Kronecker, hyperbolic, and reconstructive models for generating realistic networks. Using insights from the literature as well as some innovation, this paper attempts to formulate algorithms for generating realistic social networks that possess high clustering, small short-path lengths, and power-law distributivity. Elements of structure and hub formation will also be examined to better match real world network data. Possible applications include the creation of artificial intelligence social networks. An attempt at generating realistic social networks has difficulties with the following: matching all 3 foundational properties, while working on any sized graph.

1.2 Research Questions / Objectives

(1) Is there another simple generative algorithm that models the 3 real-world network foundational features?
(2) What are the possible key elements in social network algorithms?

1.3 Literature Review

Simple generative models have been able to partially model the three foundational features of real-world networks; power-law distribution, high clustering, and small average short-path lengths. Erdős–Rényi [1] were able to form random networks that possessed small average short-path lengths while Watts–Strogatz [2] generated small-world networks which possessed high clustering alongside small average short-path lengths. Barabási-Albert [3] managed to create a power-law distribution and small average short-path lengths through their preferential attachment model, which was further worked upon by Holme-Kim [4] to allow for tunable clustering. This created the first model that was able to possess all three properties.

These properties, however, are only the foundation. Other properties of note in realistic networks are community structuring, assortativity, temporal dynamics, multiplexity, and network geometry. Complex models that emerged after sought to also fit some of these other properties into their models. Bianconi-Barabási [7] created the fitness model which helped in explaining heterogenous hub formations for community structure. Temporal evolution was captured well by Leskovec et al. [8] through their forest fire algorithm where edges grow faster than nodes. The Kronecker graph model [9] generates synthetic networks that closely match real world networks and is used today as a benchmark. Two modern approaches with tangible applications are hyperbolic models and reconstructive models. Krioukov et al. [10] discuss hyperbolic models that use distance and popularity within a hyperbolic geometric space in order to link nodes, which is used in modeling communication networks today. Li et al. [11] worked on the reconstruction of real networks, which allowed for the formation of a generative model that exactly matches a real network's historical data.

2 Methodology

To address the research questions, an algorithm will need to be formulated from scratch. This means that It will need to be defined, mathematically formulated, and then tested. This will allow for a comparison against established models as well as real world data. The established models we will look at are: Erdős–Rényi, Barabási–Albert, and Watts–Strogatz. Each of them show different foundational properties such as small average shortest path length, high clustering coefficient, and power-law degree distribution.

2.0.1 Erdős–Rényi (ER) Graph. The Erdős–Rényi model $G(n, p)$ [1] generates a graph with n nodes where each pair of nodes is connected independently with probability p. Formally, for nodes i and j, an edge exists with probability:

$$P((i, j) \in E) = p, \quad 1 \le i < j \le n$$

This model produces graphs with a *binomial degree distribution*, which approaches a Poisson distribution for large n. ER graphs typically exhibit low clustering and short average path lengths, resembling random networks.

2.0.2 Barabási–Albert (BA) Graph. The Barabási–Albert model [3] generates *scale-free networks* using a preferential attachment mechanism. Starting from a small connected network of m_0 nodes, nodes are added sequentially. Each new node connects to $m \le m_0$ existing nodes with probability proportional to their degree:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

where k_i is the degree of existing node i. This process produces a *power-law degree distribution* $P(k) \sim k^{-\gamma}$ with $\gamma \approx 3$, capturing hub formation and heavy-tailed connectivity typical of many real-world networks.

2.0.3 Watts–Strogatz (WS) Graph. The Watts–Strogatz model [2] generates *small-world networks* with high clustering and short average path lengths. Starting from a regular ring lattice of n nodes where each node is connected to k nearest neighbors, each edge is rewired with probability β:

$$P(\text{rewire edge}) = \beta$$

For $\beta = 0$, the network is regular; for $\beta = 1$, it becomes a random graph. Intermediate values of β produce small-world networks exhibiting both local clustering and short global distances.

2.1 Simple Algorithm: Colour Combination Collision; The UALIS model

Using the foundational models as reference, the UALIS algorithm aims to work in a similar manner where only a few parameters need to be set in order to generate a graph that possesses some or all of the foundational properties of high clustering, power-law distribution, and small average short-path lengths. The hope is also that the algorithm is able to perform closed-form calculations similar to the Erdős–Rényi and Barabási-Albert models. The algorithm goes as follows: in reality, humans are very complex. It is not possible to get a one hundred percent accurate graph of social connections because things always change with time and distance, among other factors. However, we can attempt to breakdown the ability to form connections with others. The simplest understanding of that ability is that there are X number of factors involved in whether two humans will form a connection with each other. Since we are not able to universally quantify every single aspect of whether a connection will form between two humans—as for each human these factors may differ—we can use generics in order to encapsulate the chance that two humans (nodes) form a connection. The UALIS algorithm takes inspiration from its predecessors, the 3 foundational models.

From Erdős–Rényi, it takes on the probabilistic nature of node connections in forming a graph. From Watts-Strogatz, it takes on the small-world aspect in that similarly coloured nodes will end up clustering together, with bridges being formed from complimentary coloured nodes and other coloured nodes from random chance. From Barabási-Albert, nodes are added to the graph over time, with a chance of edge formations to each of the pre-existing nodes within the graph. It takes on the preferential attachment notion and turns it into another form, called specificity. The probabilistic nature of node connections in forming a graph can be observed by looking at a colour wheel. Each node is assigned a colour in the colour wheel, with an equally probable chance. If two nodes have the same colour, they receive a certain probabilistic chance to form a connection, which will be determined experimentally.

Algorithm 1 UALIS Colour Combination Collision

Require: Number of nodes n, colour arity c
Ensure: Graph $G = (V, E)$

Initialize empty graph G
weights $\leftarrow$ chance of picking an arity
colour_list $\leftarrow$ sequence of number of colours in each arity
for $i = 0$ to $n - 1$ **do**
  Add node i with attributes $(\mathit{color}, \mathit{arity})$
  for all $j \in V(G)$ where $j \neq i$ **do**
    if $\mathit{color}(i) = \mathit{color}(j)$ **then**
      $p \leftarrow 0.736$
    else if $\mathit{wheel}(i) < \mathit{wheel}(j)$ **then**
      $p \leftarrow [0.13, 0.047, 0.023, 0.013][\mathit{wheel}(i)]$
    else
      $p \leftarrow 0.01$
    end if
    if Random$() < p$ **then**
      Add edge (i, j) to G
    end if
  end for
end for
return G

The specificity is generated from new colours being added to the colour wheel over time by combining the different colours, leading to the primary colours being able to connect with similar probability to any of the colours that are similar to them, whereas the secondary colours are more specified, and the tertiary colours moreso, as seen in figure 1. With the broadness of the primary colours, it will allow for hub formations to naturally occur, while the specificity of other colours will allow for a graph that is not complete, which would not be realistic. This is similar to humans in that when we are younger we are willing and more open to forming

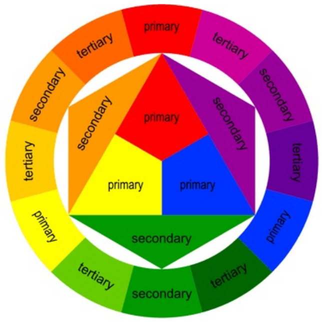

Figure 1: Colour Wheel for Edge Formations

connections with other, and as we age we become more specific and choosy on which connections we want to form. Of course, there are caveats such as workplace relations and other instances, which is why there is a damping effect similar to PageRank scoring on forming connections with colours that are not similar in nature as well. Colours that are complimentary in nature would also possess the same possibility of forming an edge connection as colours that are the same. For the purposes of this paper, 'complimentary' will be considered any colour who's arity is less than another (primary < secondary < tertiary <...< n-ary), so a tertiary colour would have a chance of connecting with a primary colour, but not vice versa. Within another possible experimental set-up, this would be done by creating pairings between all colours that also gain a probability of forming a connection

For determining the probability of each type of connection, the idea is to work backwards from a power-law equation in order to create a set of probabilities that will result in a graph possessing power-law distribution. This introduces some mathematical complexity when working with different sized graphs and factoring the weights for arity and the probabilities for each type of connection. To generalize the results would be difficult. To recap, the primary coloured nodes are able to form the most connections, and as colour specificity increases, the probability of edge formation decreases. The number of colours are selected in triplets, as that is usually the case for colours (red, yellow, blue) in the colour wheel. There is also an equivalent chance of edge formation between complimentary colours, and small chances of edge formation with other colours. For the purposes of this paper, we will work on colours up to quinternary, a total of 48 colours in total.

2.2 Complex Algorithm: The Humanity Graph

This algorithm is purely theoretical, and not included within the rest of this paper. It is only discussed here as a foundation for potential next steps regarding this research, as well as a tool for building agentic AI networks. Using the foundational properties, modern approaches, and other crucial realistic network properties, this algorithm aims to emulate real-world social networks as an expanded form of the simple algorithm, with added constraints and parameters.

To build off of the simple algorithm, it is necessary to understand a few things. Social networks are highly complex, and also change over time. This introduces the notion of multiplex edges as well as edge weights when generating realistic social networks. The "plexity" of edges determine the type of relationship: relations such as co-worker, friend, family, etc. The weights determine the closeness of that relationship and/or how important that relationship is to a node. These weights adjust over time, and can also go to zero, for relations that end. Rather than a snapshot of a social network, this algorithm aims to create a dynamic social network, where each node is represented by an agentic AI, similar to a Sims world or other such multiplayer games. This provides each node with reasoning, and allows for communication to occur within a network in its entirety, as well as on a communal scale (Ex. to see a community of workers in a workplace, show only nodes and their associated edge for that community).

The second thing needed for the complex algorithm is cascading behavior, which has effects on popular nodes such as influencers, celebrities, and politicians, but also within more local hubs as well. When an action is made by one of these popular nodes, it can result in a dramatic impact across a network, such as a loss or gain of followers on a social media (Take, for example, Logan Paul's vlog of a suicide). This can be seen on a smaller scale within friend groups and workplaces as well; when rumours spread about a friend they can be removed from a friend group, when a co-worker is fired then they are not a part of the work group. Cascading behavior works along the community network that connects nodes, such as a friend group or workplace, and passes a weight that is simply another node, and is either positive or negative. This is analogous to someone spreading a rumor, and each individual node that hears it responds in a different fashion based on who they are, their relation with that node, and whether to pass that information along or not. Another example would be influencing a node so that their colour changes (likes, behavior, dislikes, etc).

Another component of the complex algorithm would be the introduction of more colours that represent "eccentric" nodes, such as those with antisocial behavior or those with highly social behavior. These can be represented by black and white, or no colour and shifting colour. The other big change would be grouped introduction, similar to real life, where the "first-contact" between many nodes is primarily through schooling (the actual first-contact usually being family members), resulting in the formation of new communities and influences on nodes. This can also mean that instead of introducing specificity of nodes from the start, the specificity is acquired over time, where each node begins as one of a set of colours, such as the primary colours, which then has the specificity change over time based on their interactions, changing to a secondary colour and then to a tertiary colour and so on, or remaining as a primary colour, or a secondary colour, etc.

The final thing required in the complex algorithm is hierarchy. This directly effects the influence nodes have on each other, and can be seen in many real-life scenarios such as an employer and employee, a parent and child, etc. These are also dynamic relations, which means the hierarchy is also adjusted over time; if an employee no longer works for an employer, the hierarchy is broken and inexistent. The impact of hierarchy is very important in determining a node's next action; an employee works on what their employer tells them to, a social group does an activity usually based on group consensus (wants and needs of the group are higher in hierarchy than the individual), etc. Although there are caveats, the use of specificity and hierarchy provide structure on node edge formation and node influence, among others.

Although there may be other things required in this algorithm, the above-mentioned parameters are a good baseline for developing a realistic social network that evolves over time. This can then be used to create a network of agentic AI, each set as a node in a network, and each given a colour, i.e personality, that simulates the real world through a text-based approach through LLMs (large language models) such as ChatGPT and Google Gemini. They can also be given varying levels of intelligence in their large language models in the form of different weightings within their transformer matrices or different temperature levels. This provides a two-pronged approach; observations on realistic social networks, and also the potential for unique ideas to be created by one out of many of the agentic AI, similar to the real world. Necessity is the mother of invention, and if a social network is generated that emulates real life, where there is consequences and rewards and a highly varied number of individuals that each are adapting to their surroundings (Ex. humans restructure their thoughts, the LLMs restructure their transformer matrices), then agentic AI can provide unique and wholly new ideas and innovations.

3 Evaluation

Multiple cycles of parameterization and algorithm formulation were conducted, with 11 generated versions of the simple algorithm and its corresponding graphs to observe whether any of the foundational properties can be attained. This included changing weights, probabilities, and the logic of the algorithm. For the purpose of this paper, the iteration with the best results were kept, and the rest of the iterations can be seen in the code file. Similar to Watts-Strogatz small-world model [2], it is possible that the algorithm generated may behave on a spectrum that possesses foundational properties within certain parameter bounds, such as node size. These results were also compared against the baseline models of Erdős–Rényi, Barabási–Albert, Watts-Strogatz, and real world datasets with similar node and edge values to see whether any insights or patterns can be determined.

3.1 Measurements

To evaluate the viability of the simple algorithm, the following network properties will be measured: *degree distribution*, *clustering coefficient*, and *average shortest path length*. The *community structure* will also be examined to compare hub formations between real-world data and the generated networks.

Degree Distribution. The degree distribution $P(k)$ represents the probability that a randomly chosen node has degree k:

$$P(k) = \frac{N_k}{N}$$

where N_k is the number of nodes with degree k and N is the total number of nodes in the graph.

Clustering Coefficient. The clustering coefficient C_i for a node i measures the density of connections among its neighbors:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

where E_i is the number of edges between the k_i neighbors of node i. The average clustering coefficient C of the network is then:

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i$$

Average Shortest Path Length. The average shortest path length L is the mean of the shortest path distances between all pairs of nodes:

$$L = \frac{1}{N(N-1)}\sum_{i \neq j} d(i,j)$$

where $d(i,j)$ is the length of the shortest path between nodes i and j.

Community Structure. The quality of community structure is quantified using the modularity Q:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

where A_{ij} is the adjacency matrix, k_i and k_j are the degrees of nodes i and j, m is the total number of edges, and $\delta(c_i, c_j)$ equals 1 if nodes i and j belong to the same community and 0 otherwise.

3.2 Data Sources

- Real network 1: SNAP social network ego-Facebook dataset; an undirected network consisting of 4039 nodes and 88234 edges. Each node is a profile, and each edge is a friendship.
- Synthetic networks 1,2, and 3: generated using Python NetworkX for the following foundational models: Erdős–Rényi, Watts-Strogatz, and Barabási-Albert. Each of the graphs uses 4000 nodes and the Erdős–Rényi graph uses a p value of 0.1, the Watts-Strogatz uses a k value of 4 and beta of 0.3, while the Barabási-Albert uses an m value of 2.

3.3 Results

The UALIS model was able to show 2 of the 3 foundational properties; an average short path of 1.98 and a clustering coefficient of 0.398 (Table 1), for a node size of 4000. There was a previous graph generation using 1000 nodes with showed even more promising results, but do not match the data set we are trying to fit to in this paper, though it should be noted by the reader in case they wish to generalize the results.

The UALIS model was unable to achieve a power-law distribution (Figure 2), while the Barabási–Albert and Facebook Social Circles graphs were able to (Figure 4 and 6). For community detection, the modularity of The UALIS model shows strong community structure, similar to the facebook social circle's graph, with modularities of 0.643 and 0.777 respectively. (Table 2).

4 Conclusion

The UALIS model Colour Combination Collision algorithm was able to generate some of the foundational properties, although with some caveats. The probabilities and weightings used in this paper were hard-coded, so the results cannot be translated to graphs of any size. Originally, this algorithm was tried with a graph of size 1000 node, with even better results and a log-log degree distribution that was close to the power-law. The results shown here show

Table 1: Comparison of Graph Properties Across Models

Graph	Nodes	Edges	Avg. Degree	Max	Min	Avg. Path	Clustering
UALIS	4000	366,694	183.35	683	100	1.98	0.398
Erdős–Rényi	4000	799,690	399.85	480	323	1.90	0.100
Barabási–Albert	4000	7,996	3.998	149	2	4.63	0.010
Watts–Strogatz	4000	8,000	4.00	9	2	7.58	0.179
Facebook Social Circles	4039	88,234	43.69	1045	1	3.69	0.606

Table 2: Community Detection Results for Each Graph

Graph	Number of Communities	Modularity
Colour Combination Collision	15	0.643
Erdős–Rényi	3	0.033
Barabási–Albert	31	0.531
Watts–Strogatz	44	0.708
Facebook Social Circles	13	0.777

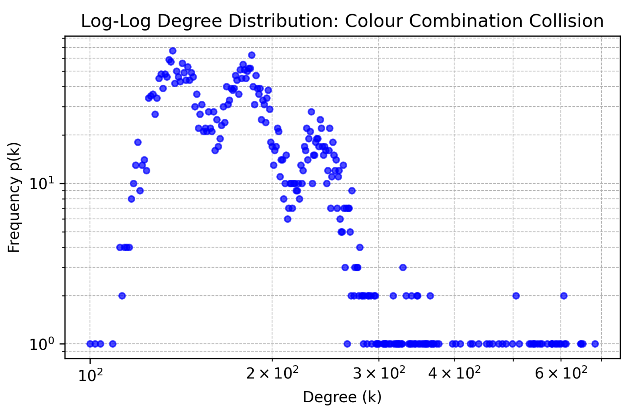

Figure 2: Log–log degree distribution of Colour Combination Collision network.

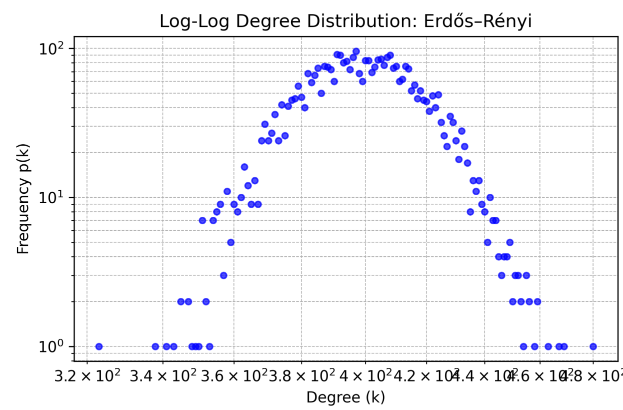

Figure 3: Log–log degree distribution of Erdős–Rényi network.

that the possibility of this algorithm in generating realistic social networks that possess the three foundational properties as well as realistic hub formations and community structure. There are multiple research directions that can stem from these results.

Important considerations in this algorithm are: the arity number, which is set in this model but should be determined mathematically based on node size. The weighting of each arity; how many of each arity's colours are added into a node.

Based off of the initial results of this research, it appears possible to create a foundational model needing only a few parameters instantiated in order to create a realistic social network possessing the three foundational properties. The key elements to consider are the number of colours associated with specificity (Some say there are 16 personality types, some use O.C.E.A.N for determining personality) as well as getting a better estimate of a generalized probability for two specific nodes to form a link. Another key element would be a distance factor reflected by the numbering of nodes, where distance reflects the probability of two nodes meeting, and then applying the probability of two nodes forming a link, which could result in similar structure to a Watts-Strogatz graph.

Each of the above, if worked out, could potentially realize an algorithm that matches all 3 foundational properties, while also enabling hub formations. The next steps in this research would be to use generalized parameters, working on the algorithm to determine better probability equations for each type of edge formation, and also working on the complex version of this algorithm.

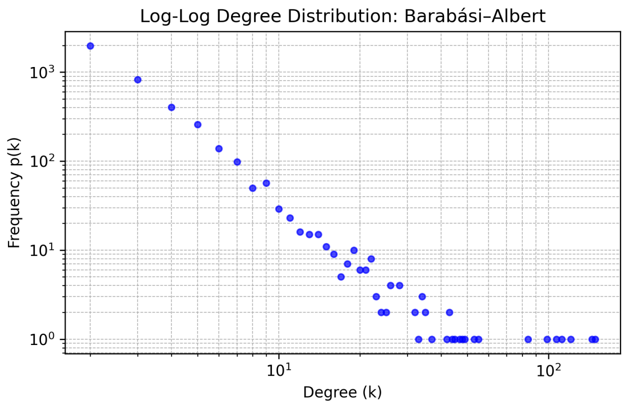

Figure 4: Log–log degree distribution of Barabási–Albert network.

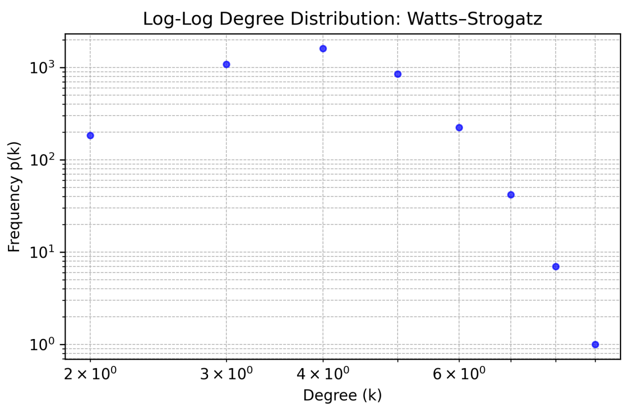

Figure 5: Log–log degree distribution of Watts–Strogatz network.

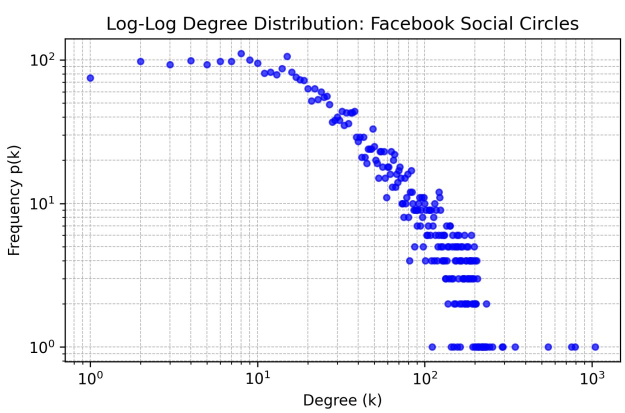

Figure 6: Log–log degree distribution of Facebook Social Circles network.

Acknowledgments

ChatGPT for helping in the literature review, providing a template for this project proposal, as well as project topic ideas. Also for providing quick answers to questions on network science, published papers, and other topics. There was also aid in generating parts of the code used in this paper, as well as parts of this paper itself. Full transcripts of the conversations can be provided upon request.

Reproducibility

All source code and scripts needed to reproduce these experiments are available at: https://github.com/sali446/EECS4414.

References

[1] Erdős, P., & Rényi, A. (1959). *On Random Graphs I*. Publicationes Mathematicae, Debrecen. – Introduced the Erdős–Rényi (ER) random graph model; baseline for random networks, short paths but low clustering and Poisson degree distribution.

[2] Watts, D. J., & Strogatz, S. H. (1998). *Collective dynamics of 'small-world' networks*. Nature, 393(6684), 440–442. – Proposed the Watts–Strogatz (WS) model; captures high clustering and small-world path lengths, but degree distribution is narrow.

[3] Barabási, A.-L., & Albert, R. (1999). *Emergence of scaling in random networks*. Science, 286(5439), 509–512. – Introduced preferential attachment (BA model); produces scale-free networks with hubs, low clustering, and short paths.

[4] Holme, P., & Kim, B. J. (2002). *Growing scale-free networks with tunable clustering*. Physical Review E, 65(2), 026107. – Extension of BA with triad formation; achieves power-law degree distribution, high clustering, and short paths.

[5] Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (2000). *Structure of growing networks with preferential linking*. Physical Review Letters, 85(21), 4633–4636. – Generalized preferential attachment; explains how tuning attachment rules affects degree distributions and clustering.

[6] Krapivsky, P. L., Redner, S., & Leyvraz, F. (2000). *Connectivity of growing random networks*. Physical Review Letters, 85(21), 4629–4632. – Nonlinear preferential attachment; flexible modeling of degree distribution and connectivity.

[7] Bianconi, G., & Barabási, A.-L. (2001). *Competition and multiscaling in evolving networks*. Europhysics Letters, 54(4), 436–442. – Fitness model: node attachment depends on both degree and intrinsic fitness; explains heterogeneous hub formation.

[8] Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). *Graphs over time: densification laws, shrinking diameters and possible explanations*. KDD '05. – Introduced Forest Fire model; captures temporal evolution, densification, clustering, and heavy-tailed degrees.

[9] Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., & Ghahramani, Z. (2010). *Kronecker graphs: An approach to modeling networks*. Journal of Machine Learning Research, 11, 985–1042. – Kronecker graph model; recursive generative model producing realistic large-scale networks with clustering, heavy tails, and short paths.

[10] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., & Boguñá, M. (2010). *Hyperbolic geometry of complex networks*. Physical Review E, 82(3), 036106. – Hyperbolic network model; nodes embedded in hyperbolic space; naturally generates power-law degree distribution, high clustering, and small-world paths.

[11] Li, X., Zhang, Y., & Wang, J. (2024). *Reconstructing the Evolution History of Networked Complex Systems*. Scientific Reports, 14, 11234. – Reconstruction model; reverse-engineers network evolution from snapshots, producing networks matching clustering, degree distribution, and path lengths.