Proceedings of the 3rd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/100/2025.22144

Maximum Likelihood Estimation for Erdős-Rényi Graphs in Social Network Analysis

Ziang Li

Shanghai Southwest Weiyu Middle School, Shanghai, China
lxu.e23sh3@ceibs.com

Abstract: This paper explores the application of the Erdős-Rényi graph model and maximum likelihood estimation in social network analysis, presenting a robust framework for understanding complex network structures. Focused on the statistical characterization of network connections through simple probabilistic methods, the study utilizes the Erdős-Rényi model to predict and analyze the dynamics within social networks. Fundamental mathematical principles such as matrix theory and Bernoulli distribution are employed to facilitate a comprehensive analysis of adjacency matrices, allowing for precise probability calculations of edge formations between nodes. Despite the model's broad applicability in predicting social interactions and refining social media algorithms, it encounters limitations due to its assumptions of uniform randomness and equal probability of edge formation. These constraints highlight the need for model enhancements to more accurately mirror real-world complexities such as community structures and variable connectivity. The paper underscores the potential of extending these theoretical models to accommodate non-uniform connection probabilities and additional social variables, thus improving their predictive accuracy and practical relevance in social network studies.

Keywords: Social network analysis, Maximum likelihood estimation, Bernoulli distribution

1. Introduction

With the development of science and technology, the information is increasingly complex and diverse. Sometimes in real life, when conducting research, a lot of data will always dazzle researchers [1,2]. This case was exactly true in social network situation. When calculating some social situations such as friendship, huge amounts of information is needed. Without a thorough understanding and estimation of the data, the realization of the study must be limited. Social Networks model relationships between individuals. Social media platforms like Facebook, Instagram, and Twitter are full of social networks. People's connections, followers, and following relationships form complex network structures. The application of social network is numerous: including information dissemination, which can help us predict how quickly and widely information will spread; influence analysis, which can can identify influential individuals in a social network; and marketing campaign, which can help companies promote products more effectively by finding influential people. Such broad application of social network analysis requires data and estimation. Luckily, maximum likelihood estimation is a great idea. By determining the likelihood of the particular social network, the relationship as well as the probability of forming relationship is distinctly presented.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/100/2025.22144

2. Overview of Erdős-Rényi Graphs

2.1. Fundamentals and Parameters

Erdős-Rényi (ER) graphs are a fundamental model for studying random graphs. The graphs named after mathematicians Paul Erdős and Alfréd Rényi, who introduced the model in the late 1950s, provide a simple yet powerful framework for understanding the properties of random networks. The ER model is fundamental to graph theory and finds applications in a variety of fields, including computer science, biology, sociology, and physics.Despite their simplicity, they provide valuable insights into the nature of random networks and serve as a baseline for more complicated models. Understanding ER graphs is a good starting point for investigating the wide and diverse field of network science [3,4]. The Erdős-Rényi (ER) graph model is defined by two primary parameters: the number of vertices nn and the edge probability p (in the G(n,p) model) or the number of edges M (in the G(n,M) model).These factors play an important role in establishing the structure and attributes of the generated random graph. Also, the increase in p led to the increase in probability of connection, as seen in the picture. In this picture, the blue point is the representation of each individual node, and the green lines are the interconnection--that is, the interrelationship--between each individuals.

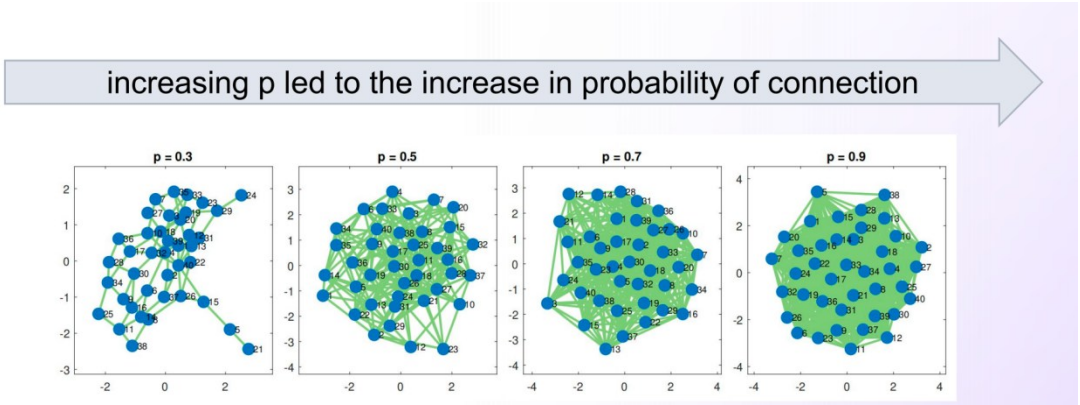

Figure 1: The increase in p led to the increase in probability of connection. (Picture credit : Original)

2.2. Mathematical Contextualization

In order to gain a thorough understanding, the formula that foster the model is required. The matrix is needed to construct a real situation. Define this matrix to X, and i and j are each individual elements in this matrix. (It is a square matrix, in other words, the number of rows is equal to the number of columns, $i = j$).

$$A_{n\times n} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{m1} & X_{m2} & \cdots & X_{n\times n} \end{bmatrix} = X_{ij}] \tag{1}$$

In the case of Bernoulli distribution, Each X_{ij} follows a Bernoulli distribution with parameter p, where p is the probability of having an edge between any two distinct vertices. The Bernoulli distribution is a discrete probability distribution that takes value 1 with probability p and value 0 with probability $1 - p$.This means that for each pair of vertices (i, j), there is a probability p that they are connected by an edge and a probability $1 - p$ that they are not. Recall the Bernoulli distribution is the binomial distribution for $n = 1$. That is, only one event test is conducted first, and the probability

DOI: 10.54254/2753-8818/100/2025.22144

of the event occurring is p, and the probability of not occurring is $1-p$. This is the simplest distribution, any random phenomenon with only two outcomes follows the 0-1 distribution [5].

2.3. Mechanisms and Dynamics

After that, the Joint PMF of the Adjacency Matrix is successfully calculated,

$$\mathrm{P}(X_{ij}=1)=p \tag{2}$$

$$\mathrm{P}(X_{ij}=0)=1-p \tag{3}$$

$$f_X(\mathrm{x};\mathrm{p})=\prod_{i=1}^{N}\prod_{j=1}^{N}p^{x_{ij}}\,(1-p)^{1-x_{ij}} \tag{4}$$

This expression represents the probability of observing a specific adjacency matrix with entries X_{ij}. It is the product of the probabilities of each edge being present or absent, raised to the power of their respective observed values. According to the Bernoulli distribution, the p is the probability. and the PMF is the product of these p and $1-p$. The likelihood function $\mathrm{L}(p|\mathrm{X})$ is essentially the joint PMF $\mathrm{fX}(x;p)$, but viewed as a function of the parameter p given the observed data X. To make this function easier to work with, especially for optimization purposes, we take the logarithm of the likelihood function. This transforms the product of probabilities into a sum of logarithms, which is mathematically more convenient. Recall that $\log a \times \log b = \log(a+b)$. In conclusion, the log-likelihood function is a key tool in the Erdős-Rényi graph model. It transforms the joint PMF of the adjacency matrix into a sum of logarithms, making it easier to work with for parameter estimation [6]. By maximizing the log-likelihood, we can estimate the edge probability. In order to locate and estimate the parameter, looking for the p that makes our observed data most probable is necessary. To find the value of pp that maximizes the log-likelihood function, We take the derivative of the log-likelihood function with respect to p. This derivative tells us how the log-likelihood changes as p changes.

$$\frac{d}{dp}logL(p|X)=\sum_{i=1}^{N}\sum_{j=1}^{N}\{\frac{x_{ij}}{p}-\frac{1-x_{ij}}{1-p}\} \tag{5}$$

To find the maximum, we set the derivative equal to zero and solve for p. This equation can be simplified and solved to find the value of p that maximizes the log-likelihood. The last step is to use some algebra and calculation. The detail of this step is shown in the picture below.

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\{\frac{x_{ij}}{p}-\frac{1-x_{ij}}{1-p}\}=0 \tag{6}$$

Substitute $\sum_{i=1}^{N}\sum_{j=1}^{N}X_{ij}$ with S:

$$\frac{S}{p}-\frac{N^2-S}{1-p}=0 \tag{7}$$

Finally, we rearrange and simplify the terms.

$$\hat{p}_{ML}=\frac{S}{N^2} \tag{8}$$

Refer back,

$$\sum_{i=1}^{N}\sum_{j=1}^{N}X_{ij}=S \tag{9}$$

DOI: 10.54254/2753-8818/100/2025.22144

2.4. Computational Validation

After the explanation of the formula part of the graph, let's put this program in the computer: And here's the output of this program, as shown in Figure 1:

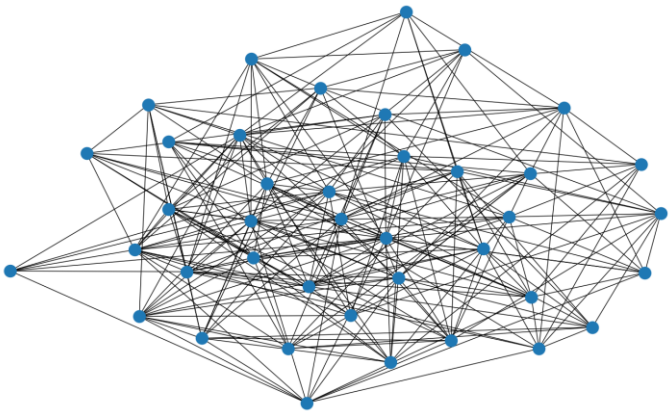

Figure 2: The output of the Graph. (Picture credit : Original)

This Erdős-Rényi graph successfully presented the individual node as well as its interconnections. Also, the number output are the calculated probability p that represent the possibility of forming interconnections [7,8]. In this program, the original p is 0.3 and N is 5. Noting that N is the number of individuals. Interestingly, the increase in N led to the increase in the error, that is, the difference between the original value p and the calculated p. The reason is consistent with the law of large numbers. As the sample size increases, the average of the results obtained from the sample (sample mean) gets closer to the expected value (population mean). This is a fundamental principle in probability theory known as the Law of Large Numbers. Essentially, larger samples tend to provide a more reliable estimate of the population parameters. Specifically, Sampling error is the variability or disparity between the sample statistic and the population parameter. A greater sample size reduces sampling error since it is more likely to reflect the total population. This reduced variability leads to more exact estimations. In conclusion, the larger the N, the smaller the error.

3. Practical Applications of Erdős-Rényi Graphs

3.1. Case Studies

The application of this graph is numerous. It can be applied into any case study. In this case, let's assume the prediction of friendship in a high school social network [9]. Granted, some students were reluctant to tell their information about friendship, so some sample in the data are missing. Here's the information:

Total possible friendship pairs: 1225 pairs

Observed pair: 300 pairs

Missing data: 200 pairs

There are 200 pairs missing from the data, in other words, whether or not these friendship are forming can not be determined. Therefore, the Maximum Likelihood Estimation is needed.

Since the 200 pairs are missing from the data, just ignore it. The final sample size is 1225-200=1025 pairs of possible friendship. After putting 1025 in the denominator, the numerator is the observed pair is 300. Just do a little calculation, the final result is about 29.2%. Now, applying this percentage into the missing sample, that is, the 200 pairs of unknown friendship. 200×29.2%=58.

DOI: 10.54254/2753-8818/100/2025.22144

Now, it is the number of the pairs of friendship in the missing data. Finally, the total number of friendship is 300+58=358.

By using the computer program, the case is represented clearly [10]. As shown in Figure 2. The red lines are the missing data that we finally calculated.

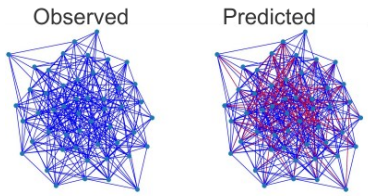

Figure 3: The observed and predicted cased represented. (Picture credit : Original)

3.2. Benefits

The most obvious benefit of this model is Simplicity and Tractability. ER graphs are straightforward to generate and analyze. They are defined by two parameters: the number of nodes n and the probability p of an edge existing between any two nodes. The simplicity of the model makes it easier to derive analytical results, such as the expected number of edges, degree distribution, and connectivity properties. In other words, this program can be applied to any data without any consideration or alteration. Such as social network (ER graphs can model random interactions in social networks, helping to study phenomena like rumor spreading or opinion dynamics.), biological network (In order to find non-random patterns, they are used to compare random networks with actual biological networks (such as gene regulatory networks)), communication network (simulate random connections in wireless sensor networks or the internet, aiding in the study of robustness and failure tolerance.), and Epidemiology (To investigate the transmission of disease and vaccination tactics, it can simulate random encounters in communities.).

When comparing networks in the actual world, ER graphs are used as a baseline or null model. Researchers can find non-random aspects like clustering, community structure, or scale-free behavior by contrasting an ER graph with the characteristics of a real network (such as social networks or networks of protein interactions). Furthermore, theories on whether observable network features are the result of random chance or statistical significance are frequently tested using ER graphs.

ER graphs help researchers understand the role of randomness in network formation. For example, they can be used to study how random connections influence phenomena like information diffusion, disease spread, or network resilience. ER graphs also exhibit interesting phase transitions, such as the sudden emergence of a giant connected component when p exceeds a critical threshold. This property is useful for studying connectivity and percolation in networks.

3.3. Challenges

Despite its many benefits and status as a fundamental model in random graph theory, Erdős-Rényi (ER) graphs have serious drawbacks when it comes to simulating real-world networks. These drawbacks result from the extremely simplistic assumption of randomness that underlies ER graphs, which frequently falls short of capturing the intricacy and organization of real-world systems.

The most obvious limitation is Lack of Degree Heterogeneity. Many real-world networks, such as social networks, the internet, and biological networks, exhibit heavy-tailed degree distributions (e.g., scale-free networks), where a few nodes (hubs) have very high degrees, while most nodes have low degrees. However, in ER graphs, the degree distribution follows a Poisson distribution, meaning most

DOI: 10.54254/2753-8818/100/2025.22144

nodes have degrees close to the average, and extreme values (very high or very low degrees) are rare. This makes ER graphs unsuitable for modeling networks with hubs or highly connected nodes.

This graph also had no Community Structure. Many networks, such as social networks, biological networks, and the internet, have community structure, where nodes form densely connected groups with sparse connections between groups. On the other hand, ER graphs lack any inherent community structure because edges are formed independently and uniformly at random. This makes them unsuitable for studying phenomena like community detection or modular organization in networks.

In conclusion, the major drawback of the graph is Overly Simplistic Assumption of Randomness. Real-world networks often exhibit non-random patterns, such as preferential attachment (e.g., new nodes are more likely to connect to highly connected nodes) or hierarchical organization. Granted, The assumption of uniform randomness in ER graphs fails to capture these mechanisms, which are critical for understanding the formation and evolution of many real-world networks. In this graph, the same probability of each connection is assumed, and the situation is much more complex in real life. For example, the personality can affect the result. Some people are introverted, while others are extroverted. This introverted may be likely to work alone. In this way, the probability of forming connections between people is greatly reduced. On the other hand, the extroverted are likely to communicate with others as well as forming interconnections. Fittingly, the p--that is, the possibility of forming connections--is larger than the previous one.

4. Conclusion

In this paper, the use of the Erdős-Rényi graph model in conjunction with maximum likelihood estimation was extensively discussed to illustrate its applications in social network analysis. This approach underscores the utility of the Erdős-Rényi model to analyze complex social structures through relatively simple probabilistic methods. The inclusion of mathematical techniques, such as matrix representations and the Bernoulli distribution, provided a foundational understanding necessary to navigate the complexities of social connections.

The derivation of the maximum likelihood estimation through a detailed analysis of adjacency matrices and probability distributions offered a quantitative method to examine the probability of connections within a network. This methodology is particularly useful in a variety of applications, from predicting social interactions in educational settings to enhancing the accuracy of friend recommendation systems in social media platforms. By applying these statistical methods, the model can predict and analyze the dynamics of social networks with a significant degree of accuracy.

However, despite its broad applicability and foundational importance in network theory, the Erdős-Rényi model faces limitations due to its inherent assumptions of randomness and uniform probability of edge formation. Real-world networks often display a higher degree of complexity, including features such as community structures and networks with nodes of varying degrees of connectivity which are not adequately modeled by the Erdős-Rényi approach. Future studies might focus on extending this model to include more complex scenarios, such as those involving non-uniform connection probabilities and additional social factors like personality traits and common interests, which could affect the likelihood of forming connections.

References

[1] *Chan, S. (2021). Introduction to probability for Data Science. https://doi.org/10.3998/mpub.12387745.*

[2] *Da F Costa, L., Rodrigues, F. A., Travieso, G., & Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. Advances in Physics, 56(1), 167–242.*

[3] *Erdős, L., Knowles, A., Yau, H., & Yin, J. (2013b). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. The Annals of Probability, 41(3B).*

DOI: 10.54254/2753-8818/100/2025.22144

[4] *Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. Demography, 46(1), 103–125.*

[5] *Zhu, X., Liu, Z., Cambria, E., Yu, X., Fan, X., Chen, H., & Wang, R. (2025). A client–server based recognition system: Non-contact single/multiple emotional and behavioral state assessment methods. Computer Methods and Programs in Biomedicine, 260, 108564.*

[6] *Wang, R., Zhu, J., Wang, S., Wang, T., Huang, J., & Zhu, X. (2024). Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking. International Journal of Multimedia Information Retrieval, 13(4), 39.*

[7] *Wang, F., Ju, M., Zhu, X., Zhu, Q., Wang, H., Qian, C., & Wang, R. (2025). A Geometric algebra-enhanced network for skin lesion detection with diagnostic prior. The Journal of Supercomputing, 81(1), 1-24.*

[8] *Zhao, Z., Zhu, X., Wei, X., Wang, X., & Zuo, J. (2021, June). Application of Workflow Technology in the Integrated Management Platform of Smart Park. In 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (Vol. 4, pp. 1433-1437). IEEE.*

[9] *Zhang, Y., Zhao, H., Zhu, X., Zhao, Z., & Zuo, J. (2019, October). Strain Measurement Quantization Technology based on DAS System. In 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (pp. 214-218). IEEE.*

[10] *Squartini, T., & Garlaschelli, D. (2011). Analytical maximum-likelihood method to detect patterns in real networks. New Journal of Physics, 13(8), 083001.*