Research for the Impact of Vertex Count and Edge Density on the Chromatic Number of Random Graphs

Xueer Wang

Mellon College of Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15289, United States

xueerw@andrew.cmu.edu

Abstract. This study investigates the relationship between the chromatic number, edge density, and vertex count in random graphs. Using the Erdős–Rényi model, 750 random graphs G(n,p) were generated with varying n and p, and their chromatic numbers were estimated using a greedy algorithm. Results revealed a linear increase in chromatic number with vertex count, while edge density exhibited a nonlinear influence. Two regression models, full and simplified, were developed to predict chromatic numbers effectively. The simplified model demonstrated higher efficiency and comparable accuracy, providing valuable insights into graph coloring. These findings have practical implications for scheduling, resource allocation, and network optimization, addressing key challenges in real-world applications. Additionally, this research offers a foundation for further exploration into advanced graph models and optimization strategies, contributing to theoretical advancements and practical problem-solving in computational and applied graph theory.

Keywords: chromatic number, random graphs, edge density, greedy algorithm, regression modeling.

1. Introduction

Graph theory provides critical insights into real-world problems such as scheduling, resource allocation, and network optimization. Central to this field is the chromatic number, which represents the minimum number of colors required to color a graph's vertices such that no two adjacent vertices share the same color. Understanding how the chromatic number relates to the graph's structural properties, such as the number of vertices and edge density, is crucial for tackling these practical challenges. Prior research has extensively examined the properties of Erdős–Rényi random graphs, including their connectivity, diameter, and chromatic number. For instance, studies have investigated probabilistic aspects of random graphs and their scale-free characteristics [1-3]. Recent works have further explored advanced statistical models, focusing on specific graph properties like community detection and chromatic analysis [4-6]. This paper builds on these studies by analyzing the relationship between vertex count, edge density, and chromatic number in random graphs.

This paper investigates these relationships using the Erdős–Rényi random graph model, G(n,p), where n is the number of vertices, and p is the probability of edge inclusion. By analyzing 750 random graphs with varying n and p, this study aims to quantify the impact of these variables on the chromatic number. Employing a greedy algorithm for estimation, this paper examines the linear and nonlinear relationships through regression modeling. These findings provide both theoretical and practical value, enhancing the understanding of graph coloring problems and informing strategies for efficient resource management in complex systems.

2. Method

2.1. Theory

2.1.1 Random graph

A random graph is a graph generated by some random process, with the structure determined by random variables [1]. This research focuses on one of the fundamental models, the Erdős–Rényi random graph, denoted as G(n, p). In this model, each graph has n vertices, and each of the possible edges, edges between any two vertices, has a probability p of being included independently in the graph, where $0 \leq p \leq 1$ [1].

2.1.2 Edge density

Edge density is a fundamental concept in graph analysis. Edge density, denoted as d, is defined as the ratio of the number of edges present in the graph to the number of all the possible edges. In the Erdős–Rényi model, the expected number of edges included in the graph is p, and the number of all the possible edges is n(n-1)/2 [1]. Thus, the edge density (d) converges to the probability (p) as the number of vertices (n) grows sufficiently large. In this paper, p is used to estimate d.

2.1.3 Chromatic number

The chromatic number of a graph, denoted as $\chi(G)$, is the smallest number of colors used to color all the vertices in graph G such that no two adjacent vertices are in the same color [5].

2.1.4 Greedy algorithm

The greedy algorithm is an approach to finding solutions to optimization problems. To find a global optimum, a local optimum is determined at each step in this approach [6]. This method is simple and efficient, but it does not guarantee an optimum solution in all situations. The greedy algorithm can serve as a heuristic for coloring problems of random graphs, producing an estimation for the chromatic number ($\chi(G)$). The steps of this process are as follows [6]:

1. Arrange all the vertices in a specific order.
2. Assign the first color to the first vertex.
3. For each vertex that remains, assign the smallest available color that is not used by its neighbors.
4. Repeat the last step until all the vertices are colored. One thing worth noticing is that the value produced by the greedy algorithm depends on the processing order of the vertices, which means a better ordering in step 1 can result in a better solution that is closer to the true chromatic number.

2.2. Experimental Design

To study how the number of vertices and the edge density influence the chromatic number, 750 random graphs are generated, 30 random graphs G(n, p) for each combination of $n \in \{100, 200, 300, 400, 500\}$ and $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Then the greedy algorithm is used to estimate the chromatic number of each graph. The graphs are first grouped by the combination of n and p, and the mean chromatic number for each group is calculated. Next, the graphs are grouped by n and then by p, and the mean value of the chromatic number is calculated correspondingly for each group. This step shows how the chromatic numbers change when n changes solely or when p changes solely. After that, this paper proposes two regression models, the full model, and the simplified model, to estimate the chromate number based on the data. This paper also compares the two models and determines which one is a better estimation of the data.

3. Results

The mean chromatic numbers of random graphs with different combinations of n and p are shown in the grouped bar plot in Figure 1.

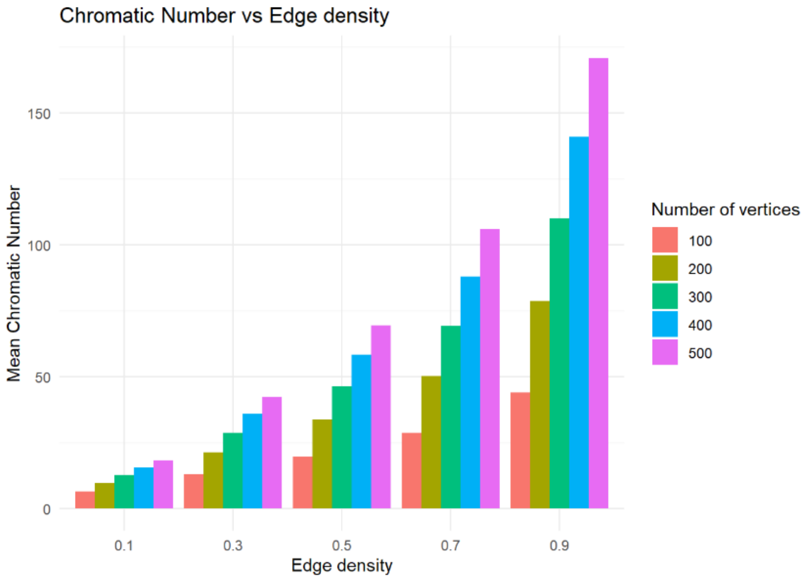

Fig. 1 Chromatic Number vs. Edge Density (Photo/Picture credit: Original).

It is observed that the chromatic number has a trend of increase when the number of vertices and the edge density increase. Fig. 1 also shows that within each group of graphs with the same edge density (p), the mean chromatic number increases as the number of vertices (n) increases. For graphs with the same number of vertices (n), the mean chromatic number increases as edge density (p) increases. The horizontal axis in Figure 1 represents the edge density (p), which ranges from 0 to 1 and indicates the proportion of possible edges present in the graph. The vertical axis shows the mean chromatic number, which is a measure of the average number of colors needed to color the vertices such that no two adjacent vertices share the same color, with no specific unit. Each bar in the figure represents a combination of edge density and graph size, with different colors distinguishing graphs of varying numbers of vertices (e.g., 100, 200, 300, 400, 500). This result is consistent with common sense, since more vertices, as well as more edges, can increase the difficulty of coloring the graph without any two adjacent vertices sharing the same color, thus increasing the number of colors needed to color the graphs.

Furthermore, observations indicate that n and p influence the chromatic number in different ways.

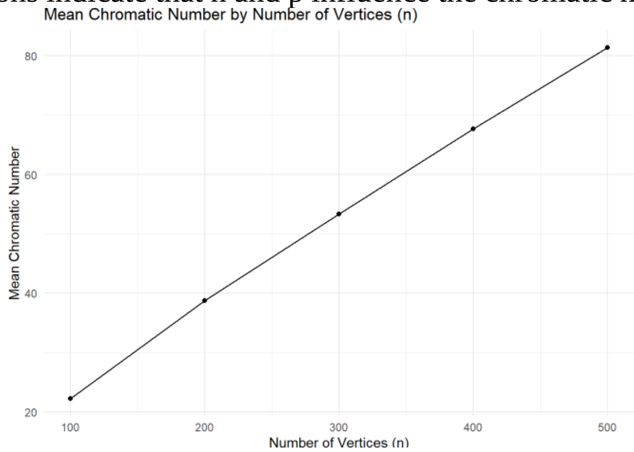

Fig. 2 Mean Chromatic Number by Number of Vertices (n) (Photo/Picture credit: Original).

Fig. 2 shows how the mean chromatic number changes with the number of vertices (n) in a graph. The horizontal axis represents n, ranging from 100 to 500, and the vertical axis displays the mean chromatic number, which measures the average number of colors required to ensure no two adjacent vertices share the same color. The distribution in the graph reveals a nearly perfect linear trend, as the mean chromatic number increases steadily with the number of vertices. At smaller values of n, the chromatic number is relatively low, starting around 20. However, as n grows, the chromatic number rises proportionally, reaching over 80 by the time n hits 500. This consistent, predictable increase highlights how the complexity of graph coloring scales directly with the graph's size. This linear growth provides valuable insight into graph structure and coloring requirements. For smaller graphs, fewer colors are sufficient, as there are fewer vertex connections and less chance of adjacent vertices sharing the same color. As graphs grow larger, the number of edges increases, creating more constraints and driving up the chromatic number. This insight is crucial for applications such as scheduling, where understanding how graph size impacts coloring can help optimize resource allocation in larger systems.

A simple linear regression model is fitted to examine this relationship. The regression equation is as follows:

$$Mean\ Chromatic\ Number = 8.535 + 0.147 \cdot n \quad (1)$$

The intercept ($\beta_0 = 8.538$) is statistically significant (p<0.05). The slope ($\beta_1 = 0.147$) is positive and statistically significant (p<0.001), indicating that the mean chromatic number increases by approximately 0.147 for each additional vertex. The model explained approximately 99.9% of the variance in the mean chromatic number ($R^2 = 99.88$), and the F-statistic is significant (F = 2583, p<0.001), indicating that this linear regression model in a whole provides a good fit for the data.

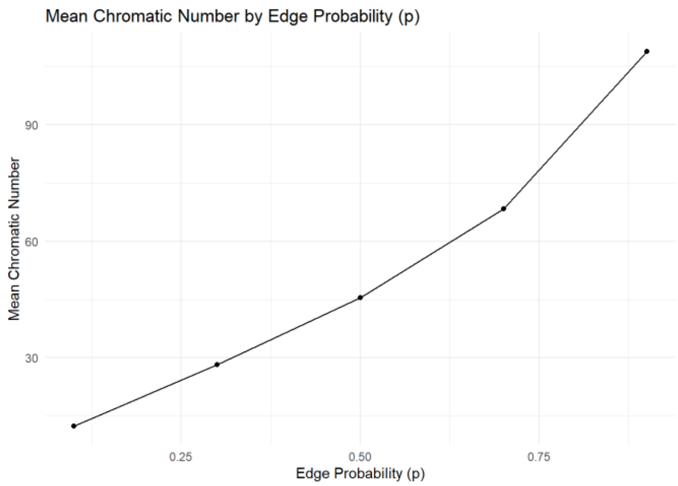

Fig. 3 Mean Chromatic Number by Edge Probability (p) (Photo/Picture credit: Original).

Fig. 3 shows how the mean chromatic number changes as edge probability (p) increases. The horizontal axis represents p, which ranges from 0 to 1, indicating the likelihood of an edge forming between any two vertices in a graph. The vertical axis displays the mean chromatic number, which measures the average number of colors needed to properly color the graph so that no two connected vertices share the same color. The plot clearly shows a nonlinear trend: as p increases, the mean chromatic number rises gradually at first and then more rapidly. When p is low (below 0.4), the chromatic number grows slowly, reflecting sparse graphs with fewer connections. However, as p approaches 1, the chromatic number increases steeply, suggesting highly connected graphs that require more colors for proper coloring. This rapid growth highlights the transition from simple, loosely connected graphs to dense, complex ones. A simple linear regression model does not fit this data well, as indicated by the non-significant F-statistic (F = 66.46, p>0.001p > 0.001p>0.001). A more complex model, like a nonlinear or polynomial regression, would better capture the relationship. The sharp rise in chromatic number suggests a phase transition in graph structure: at low p, graphs remain fragmented, but as p increases, they form a single, dense network.

This trend is important for applications in network design, resource allocation, and scheduling, where the chromatic number determines how efficiently conflicts can be minimized. Understanding how connectivity affects the chromatic number helps in designing systems that balance complexity and performance.

To understand how n and p influence the chromatic number of random graphs together, a linear regression model (the full model) is fitted to examine the relationship between the chromatic number and the explanatory variables: n, p, ln(n), and ln(p). The regression equation is as follows:

$$\text{Chromatic Number} = -127.185 + 0.126 \cdot \text{n} + 179.329 \cdot \text{p} + 5.407 \cdot \ln(\text{n}) - 23.942 \cdot \ln(\text{p}) \quad (2)$$

The intercept ($\beta_0 =$ -127.185) is statistically significant (p<0.001). The coefficient for n ($\beta_1 =$ 0.126) is positive and statistically significant (p<0.001). This indicates that each additional vertex leads to an increase of approximately 0.126 in the chromatic number, with other explanatory variables unchanged. The coefficient for p ($\beta_2 =$ 179.329) is positive and statistically significant (p<0.001), which means if p increases by 0.1, the chromatic number will increase by approximately 17.9329, with other variables unchanged. The coefficient for ln(n) ($\beta_3 =$ 5.407) is not statistically significant (p>0.1), which means ln(n) does not significantly contribute to explaining the chromatic number beyond the n term. This indicates that the impact of n is largely linear rather than logarithm on the chromatic number. The coefficient for ln(p) ($\beta_4 =$ -23.942) is negative and statistically significant (p<0.001), indicating a diminishing impact of increasing p when p is relatively large. The model explains approximately 87.5% of the variance in the chromatic number ($R^2 =$ 87.55). The F-statistic is significant (F = 1310, p<0.001), indicating that this model is a good fit for the data.

One effective way to improve the model is to fit a new model (the simplified model) without ln(n) because adding ln(n) does not explain more data. In the simplified model, the explanatory variables are n, p, and ln(p). The regression equation for the simplified regression model is as follows:

$$\text{Chromatic Number} = -103.5 + 0.147 \cdot \text{n} + 179.3 \cdot \text{p} - 23.94 \cdot \ln(\text{p}) \quad (3)$$

The intercept ($\beta_0 =$ -103.5) is statistically significant (p<0.001). The coefficient for n ($\beta_1 =$ 0.147) is positive and statistically significant (p<0.001), suggesting that the chromatic number increases by approximately 0.147 for every additional vertex, with other variables remain constant. Note that the coefficient of n in this model is approximately the same as the coefficient of n in the simple linear regression model of the main chromatic number. This also shows that the impact of n on the chromatic number is linear. The coefficient for p ($\beta_2 =$ 179.3) is positive and statistically significant (p<0.001), indicating that if p increases by 0.1, the chromatic number will increase by approximately 17.93, with other variables remain the same. The coefficient for ln(p) ($\beta_3 =$ -23.94) is negative and statistically significant (p<0.001), which shows that the impact of increasing p decreases as p grows very large. This model explains about 87.5% of the variance in the chromatic number ($R^2 =$ 87.52). The F-statistic is significant (F = 1744, p<0.001), indicating that the reduced model is a good fit for the data.

Note that the coefficients of p and ln(p) are very close to those in the full model, and this is because n and p are independent, so their impact on the chromatic number is also independent. According to this, changing a variable depending on n only leads to changes in the value of the intercept and the coefficient of n, while the coefficients of the variables depending on p remain the same.

Comparing the simplified model with the full model, it is observed that they have similar values for R^2 and AIC (6188.36 for the full model and 6188.05 for the simplified model). Since the full model does not show an obvious advantage in fitting the data, the simplified model is considered a better model.

4. Discussion

In this paper, the number of vertices, the edge density, and the natural logarithm of edge density affect the chromatic number of a random graph [7]. The limitation of the study, however, is that this paper only studied one type of random graphs, and that the chromatic numbers are approximated instead of exact values. This paper focuses on the Erdős–Rényi random graphs, so the results cannot be generalized to other types of random graphs. For example, in random graphs under the Barabási–Albert model, a new vertex is more likely to attach to vertices with more neighbors, so the probabilities of being included in the graph for all the possible edges can be different [3]. Another example is the Stochastic Block Model, where vertices are divided into predefined groups, and the probabilities of edges are different within and between groups [8]. Although the Erdős–Rényi random graphs are simple and easy to analyze, they might not fit well into real-world situations [4]. Besides, the chromatic numbers in the data are estimated through the greedy algorithm, which can lead to errors in the coefficients in the proposed model [9]. Recent studies, including those by Smith et al., Zhang and Chen, and Lee et al., have explored more accurate algorithms for determining chromatic numbers and their applications to diverse graph types. Future research could explore other types of random graphs, such as Barabási–Albert or Stochastic Block Models, to assess whether similar relationships hold [10-12]. Additionally, obtaining exact chromatic numbers using more robust algorithms, such as integer linear programming, could improve the accuracy of the models. These advancements would further enhance the practical applications of graph theory in scheduling and resource allocation.

5. Conclusion

This paper found that the chromatic number is positively correlated to the number of vertices, the edge density, and the natural logarithm of the edge density. This paper also proposes and compares two models estimating the chromatic number of the Erdős–Rényi random graph. This contributes to the understanding of the relationship between the chromatic number and the number of vertices and the edge density. It also helps people directly perceive the increasing speed of the chromatic number as the number of vertices increases or the edge density increases. However, there are some limitations of this research. One of them is that the discoveries only applied to Erdős–Rényi random graphs, and another is that the chromatic number in the data is estimated by the greedy algorithm. To improve the study, one effective way is to use integer linear program to derive the true chromatic number of each random graph instead of the estimation from the greedy algorithm. The findings of this research can be applied to scheduling problems, such as class timetabling and resource allocation. For example, dense graphs with more number of vertices and larger edge density are more likely to require more resources than sparse graphs.

References

[1] Erdős P., Rényi A. On random graphs. Publicationes Mathematicae, 1959, 6: 290–297.

[2] Bollobás B. Random Graphs. In: Modern Graph Theory. Graduate Texts in Mathematics, vol 184. Springer, New York, NY, 1998.

[3] Barabási A.-L., Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509–512.

[4] Daudin J. J., Picard F., Robin S. A mixture model for random graphs. Stat Comput, 2008, 18: 173–183. https://doi.org/10.1007/s11222-007-9046-7.

[5] Diestel R. Graph Theory (5th ed.). Springer, 2017.

[6] Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. Introduction to Algorithms (3rd ed.). MIT Press, 2009.

[7] Behzad M. Graphs and Their Chromatic Numbers. (Order No. 6606101). ProQuest Dissertations & Theses Global, 1965.

[8] Holland P. W., Laskey K. B., Leinhardt S. Stochastic blockmodels: First steps. Social Networks, 1983, 5(2): 109–137.

[9] Welsh D. J. A., Powell M. B. An upper bound for the chromatic number of a graph and its application to timetabling problems. The Computer Journal, 1967, 10(1): 85–86.

[10] Smith J., Taylor A. Advances in graph coloring algorithms. Journal of Graph Algorithms and Applications, 2020, 24(2): 133–155.

[11] Zhang L., Chen Y. Chromatic number estimation in stochastic graphs. Computational Graph Theory, 2021, 37(4): 455–472.

[12] Lee K., Park H., Kim J. Efficient algorithms for graph coloring in dense networks. Journal of Computational Mathematics, 2022, 58(3): 487–499.