Rumor Propagation on a Spread-out Line Graph*

Tianyi Huang, Hansen Liu, Yongzheng Yang
Project Leader: Daecheol Kim
Faculty Mentor: Partha S. Dey

December 27, 2024

Abstract

Consider a line graph spanning from vertex 0 to N. In this graph, each vertex represents an individual, and an edge between two vertices signifies their acquaintance. More precisely, individuals are considered acquainted if they are within a distance of 1 from each other. To explore this network further, we introduce a parameter d that generalizes the graph by defining the range of acquaintanceship. With this extension, individuals know others within a distance of d. Our primary focus is investigating the rate at which rumors propagate from the individual positioned at vertex 0 to the one at vertex N. To achieve this, we aim to analyze the first-passage time (FPT) from vertex 0 to N, considering the varying values of d. Notably, when d equals $(N-1)$, the line graph becomes a complete graph. In [3], the author investigated this case, revealing that the first-passage time asymptotically approaches $\log N/N$ as N tends to infinity. Motivated by these insights, our project objective is to simulate the first-passage time and collect data on its asymptotic behavior. By extrapolating the outcomes from the line graph, we seek to interpolate and elucidate the implications for the complete graph, thereby expanding our understanding of this fascinating phenomenon.

Contents

*Support for this project was provided by the Illinois Geometry Lab and the Department of Mathematics at the University of Illinois Urbana–Champaign.

1 Introduction

First passage percolation (FPP) is a mathematical model used to study the spread of rumor or the flow of a liquid through a random medium. Introduced in the 60s to study the spread of fluids through porous media, FPP is a generalization of percolation theory concerned with the emergence of a continuous path of occupied sites in a random graph.

In FPP, one considers a finite or infinite graph where each edge is assigned a random weight. The weight of an edge represents the time or distance required for a particle or rumor to traverse that edge. The goal is to find the minimum time or distance required to travel from a given starting node to a destination node. This minimum time or distance is called the first passage time (FPT).

FPP can also be thought of as a model for random geometry with the randomness arising out of the random edge weights. It has applications in many areas, including physics, computer science, finance, and biology. For example, in physics, FPP can be used to model the propagation of waves through a disordered medium, such as a random network of fibers or porous material. In finance, FPP can be used to model the time it takes for information to propagate through a network of traders, influencing the price of a security.

One of the most important questions in FPP is to determine the scaling behavior of the FPTs as the size of the graph grows to infinity. In particular, one is interested in determining whether there exists a critical value of the edge weight distribution parameter such that the FPTs exhibit a phase transition from a regime where they scale sub-linearly with the graph size to a regime where they scale linearly with the graph size. In this project, we consider the following question:

If resources allow information to spread to nodes at larger distances with similar costs, at what distance threshold does the speed change significantly, and in what way?

1.1 The model

We begin by introducing some essential definitions to establish the foundation of our model. Consider positive integers N and d with $d \leq N-1$, and let

$$V_N := \{0, 1, \ldots, N\}$$

denote the set of vertices. We represent the undirected edge from vertex i to vertex j as (i, j).

Definition 1. *The graph L_N^d is defined as the d**-spread out line graph** from 0 to N. In this graph, each vertex $i \in V_N$ has undirected edges (i, j) such that $|i - j| \leq d$.*

For instance, when $i = 0$, vertex i possesses d many edges (i, j) where $j = 1, 2, \ldots, d$. Similarly, when $d \leq i \leq N - d$, vertex i is connected to $2d$ many edges (i, j) where $j = i - d, i - d + 1, \ldots, i - 1, i + 1, \ldots, i + d - 1, i + d$.

We assign a *random weight* ω_{ij} to each edge (i, j). For this project, we assume that ω_{ij} is uniformly distributed on the interval $(0, 1)$. Consider any two vertices i and j. We define a *path* P_{ij} as a collection of vertices $(i_0, i_1), (i_1, i_2), \ldots, (i_{n-1}, i_n)$ where $i = i_0$ and $j = i_n$. The *length of a path* corresponds to the number of edges it contains. The *weight of a path* is defined as the sum of edge weights over all edges in the path. Furthermore, we denote the k-th vertex in the path as $P_{ij}(k) = i_k$.

Definition 2. *Let $T_{i,j}$ denote the minimal weight of a path between vertices i and j. This is called the **first-passage time** from i to j.*

Our main objective is to analyze the first-passage time $T_N := T_{0,N}$ from vertex 0 to vertex N. To accomplish this, we simulate multiple samples using Dijkstra's algorithm and analyze the obtained data.

1.2 Algorithm used for Simulations

To find the path with the minimum cost in the graph, we used Dijkstra's algorithm. Initially, we considered using A* search, but upon further analysis, we realized that the only admissible heuristic available to us is $h(x) = 0$. This is due to the fact that the edge weights in our graph can potentially be zero. Consequently, using A* search with $h(x) = 0$ would not provide any advantage over using Dijkstra's algorithm. Therefore, we proceeded with Dijkstra's algorithm (see [4, Section 24.3] for more details) as our method of choice.

2 The Number of Non-Return Points

Notice that each vertex in the graph has neighbors not only on the right-hand side but also on the left-hand side. In other words, there exist edges of the form (i, j) where $i > j$. This implies that it is possible to traverse the path in a backward direction. Consequently, we introduce the concept of a non-return point to capture this behavior.

Definition 3. *A **non-return point** in a path P is a vertex $P(k)$ such that $P(k) < P(j)$ for some $j < k$, and $P(k) < P(\ell)$ for all $\ell > k$.*

A non-return point can be interpreted as a point where the rumor spreads backward and does not return to previous vertices. Figure 1 provides an illustrative example of a non-return point.

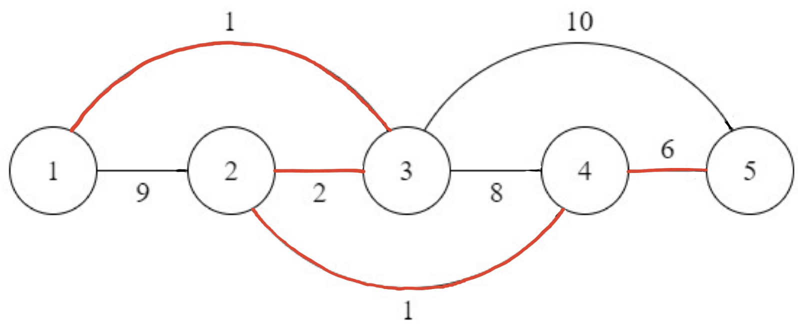


Figure 1: Example of a Non-Return Point which is 2

In this section, our focus is on analyzing the number of non-return points in a graph according to the value of d. As the value of N increases, we can see that the number of non-return points also increases. To eliminate the influence of varying values of N, we normalize the number of non-return points by dividing it by N. We refer to this normalized quantity as the *fraction of non-return points* and denote it as **NRP**.

2.1 NRP as d increases to N

We conducted simulations of the graph and obtained the fraction of non-return points (NRP) for nine different values of d:

$$d = 3, 5, \log N, N^{1/3}, N^{2/5}, \sqrt{N/2}, \sqrt{N}, N/20, N/10, \text{ and } N/5.$$

See Figure 2. Each color represents a different value of N in the order of $N = 1000, 1500, 2000$, and 5000. As we will discuss in the next section, since we observed phase transitions at expected times, we categorized d into three groups: small constant, fractional power, and large number.

Figure 2 illustrates that the NRP generally increases until d reaches 5, and then decreases to 0 as d approaches $N - 1$. Based on this observation, we propose the following conjecture.

Conjecture 1. *The fraction of non-return points converges to* 0 *as d increases to N.*

One possible explanation for this behavior is that as d increases, each step taken covers a larger range. Consequently, in most cases, a path contains fewer points, resulting in fewer non-return points.

2.2 NRP maximized at $d = 5$

From Figure 2, we can observe that the NRP is highest when $d = 5$. To further investigate this trend, we conducted simulations by varying N from 100 to 50000 and d from 3 to 7. We

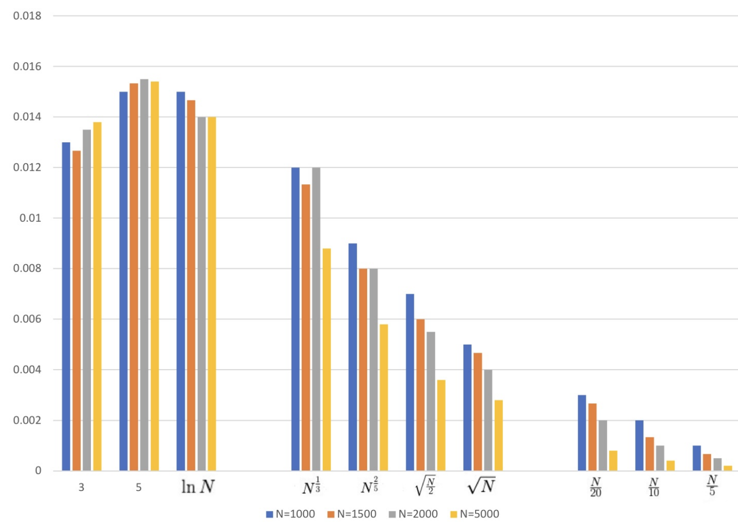


Figure 2: NRP for different d

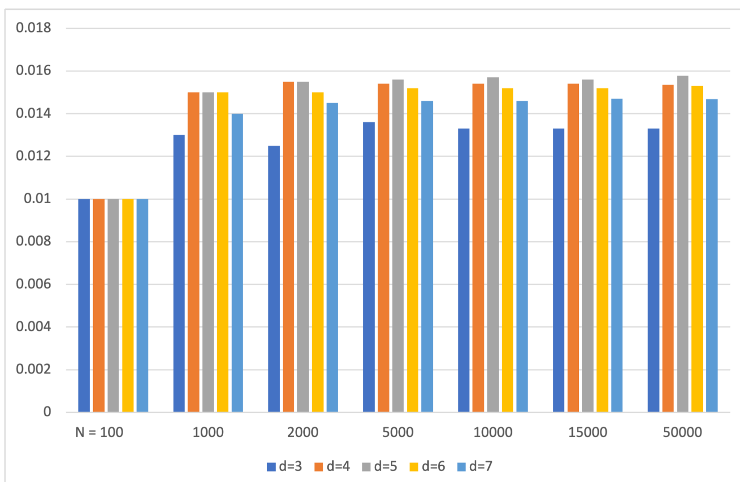
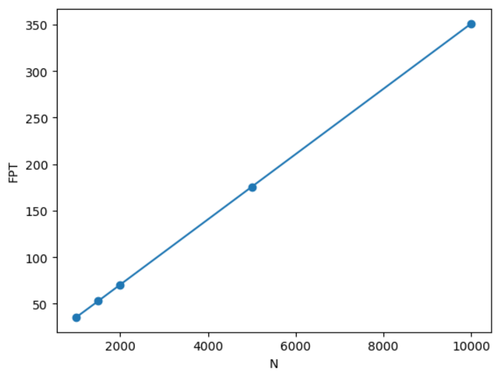


Figure 3: NRP for small d

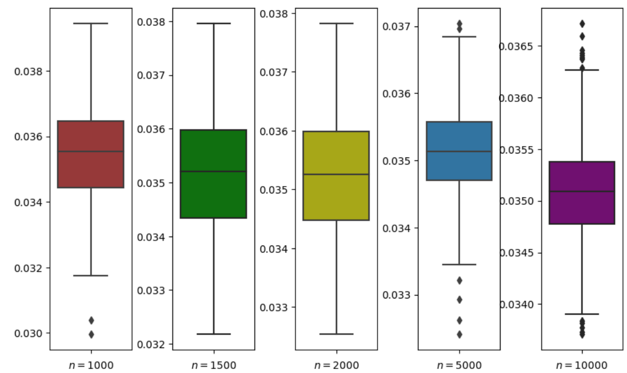
chose 100 as the smallest value for N to avoid extreme randomization caused by excessively large values of d relative to N, which would not yield meaningful insights. The results are presented in Figure 3.

The obtained results consistently demonstrate that the NRP is maximized when $d = 5$ for all selected values of N. Based on this observation provided by the simulations, we can assert this conjecture.

Conjecture 2. *If all the edges are uniformly distributed from* 0 *to* 1*, then the fraction of non-return points is maximized when* $d = 5$ *for sufficiently large* N*.*



(a) AVG FPT for different N with $d = 5$



(b) Distribution of data of $\frac{FPT}{N}$

2.3 NRP and greedy algorithmn

To find the shortest path from the source to the target, one might initially consider using a greedy algorithm.

Greedy Algorithm: For each vertex i, find a neighboring vertex $j > i$ such that $\frac{j-i}{\omega_{ij}}$ is maximized. Move to vertex j and repeat the process recursively.

This algorithm has a runtime complexity of $O(n)$. If we can ensure that it always produces the correct result, it could be a preferable algorithm to use. However, the limitation of this algorithm is that it does not consider the possibility of going backward in a sample path. Non-return points only exist when there are backward paths, and typically the number of non-return points and the number of backward paths have a positive correlation (although in some extreme cases, many backward paths can have few non-return points, this possibility is extremely low). Therefore, we can propose the following conjecture:

Conjecture 3. *There is a negative correlation between the number of non-return points (NRP) and the correctness of the greedy algorithm. The greedy algorithm performs more accurately when d is large enough, but it does not perform well when d is small.*

3 Expected First Passage Time

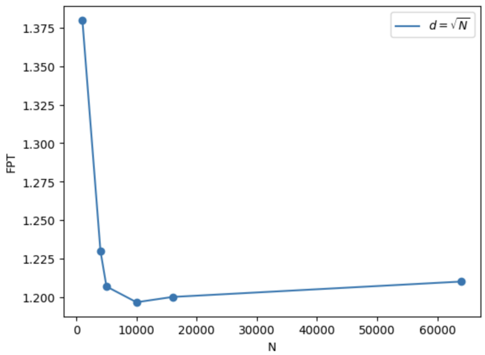
Recall that T_N represents the first passage time (FPT) from the first vertex to the N-th vertex. In this section, our focus is on predicting the time it takes for a rumor to spread from one person to another. We model this as the total cost of traveling from the first vertex to the last vertex in the line graph L_N^d. Through careful simulation of the FPT performance on graphs with different values of d and N, we observed three distinct cases of FPT behavior: $d \ll \sqrt{N}$, $d = O(\sqrt{N})$, and $d \gg \sqrt{N}$.

3.1 Case 1: $d \ll \sqrt{N}$

For small and constant values of d, we observe that the growth of FPT is nearly linear. The upper right graph displays the distribution of $\frac{FPT}{N}$, where the ratio remains within the range of 0.035 to 0.036 with minimal fluctuation. Based on these findings, we propose the following conjecture:

Conjecture 4. *If $d \ll \sqrt{N}$, then $\mathbb{E}[T_N] \to \infty$ as $N \to \infty$.*

3.2 Case 2: $d = \lambda\sqrt{N}$ $(0 < \lambda < \infty)$



AVG FPT for different value of N with $d = \sqrt{N}$

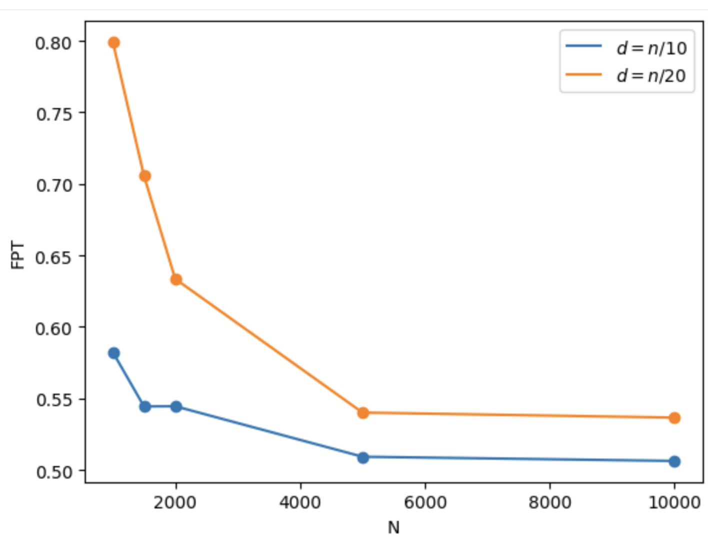
λ, N	1000	4000	16000	64000
0.5	4.15	3.68	3.42	3.28
0.7	2.12	2.05	1.98	1.90
1.0	1.38	1.23	1.2	1.21
1.2	1.08	0.998	0.987	0.977
1.5	0.87	0.835	0.838	0.76

d $= \lambda\sqrt{N}$ for different value of N

During our simulations, we observed that when d is approximately equal to $\sqrt{N}$, there is a significant change in the behavior of the first passage time (FPT) from diverging to converging. To further investigate this behavior, we conducted simulations of 1000 samples in each case with $d = \lambda\sqrt{N}$, where λ and N were chosen values. We observed that the FPT generally exhibited a decreasing trend, and the value of FPT appeared to converge to a constant value. Specifically, for $d = \sqrt{N}$ and with a uniform distribution of edge weights from 0 to 1, we found that the value of FPT remained around 1.2. Based on these observations, we propose the following conjecture:

Conjecture 5. *If $d = \lambda\sqrt{N}$, then $\mathbb{E}[T_N] \to c(\lambda)$ for some constant $c(\lambda) \in (0, \infty)$ as $N \to \infty$.*

3.3 Case 3: $d \gg \sqrt{N}$ ($d = \alpha N$ with $0 < \alpha < 0.5$)



AVG FPT for different N with large d

For the case when d is much greater than $\sqrt{N}$, we observed that the value of the first passage time (FPT) decays rapidly. Along with our knowledge of the FPT converging to 0 for the complete graph when $d = N - 1$, we propose the following conjecture:

Conjecture 6. *If $d \gg \sqrt{N}$, then $\mathbb{E}[T_N] \to 0$ as $N \to \infty$.*

4 Empirical Distribution of First Passage Time

In our study, we analyzed the empirical distribution of the first passage time (FPT) and identified two general cases based on the relative comparison between the parameter d and the square root of the sample size N. This analysis provides some insights into the behavior of the FPT and helps us understand its properties of distribution in different cases.

4.1 Empirical Distribution with $d \ll \sqrt{N}$

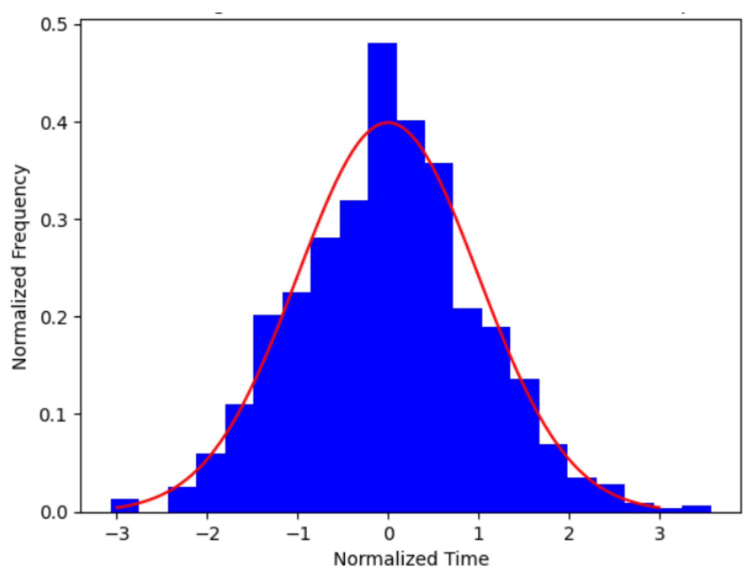


Figure 5: Histogram of FPT with $N = 10000$ and $d = 5$ with 1000 samples

When d is much smaller than $\sqrt{N}$, we observed an empirical distribution of the FPT that closely approximates a Gaussian distribution. This finding is supported by Figure 5, which presents a histogram of the FPT with $N = 10000$ and $d = 5$, based on 1000 samples.

To further verify the resemblance to a Gaussian distribution, we constructed a QQ plot (quantile-quantile plot) comparing the empirical distribution to a standard Gaussian distribution. As shown in Figure 6, the points on the QQ plot align relatively well with the diagonal line, indicating a close match between the empirical distribution and the Gaussian distribution.

From these figures, we can conclude that when d is significantly smaller than $\sqrt{N}$, the empirical distribution of the FPT exhibits characteristics that resemble a Gaussian distribution.

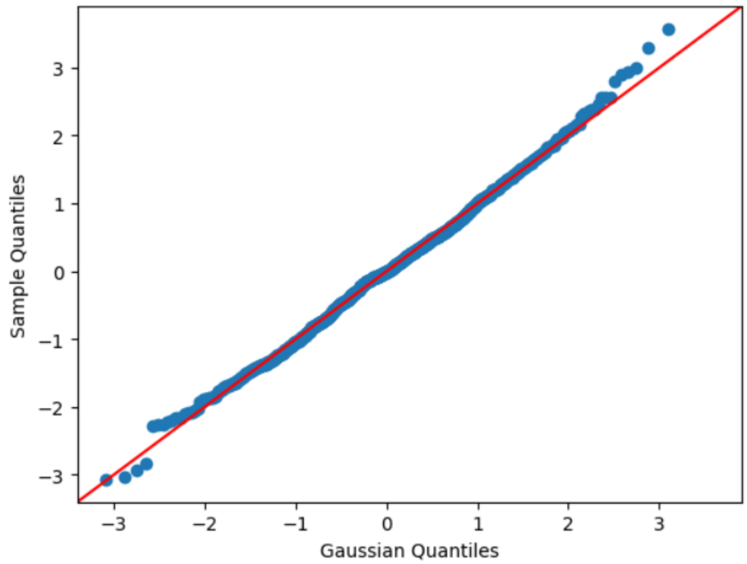


Figure 6: QQ Plot of the histogram with Standard Gaussian Distribution

4.2 Empirical Distribution with $d \gg \sqrt{N}$

In contrast, when d is much larger than $\sqrt{N}$, the empirical distribution of the FPT deviates from a Gaussian distribution. Figure 7 presents a histogram of the FPT with $N = 10000$ and $d = 2000$, based on 1300 samples, illustrating this deviation.

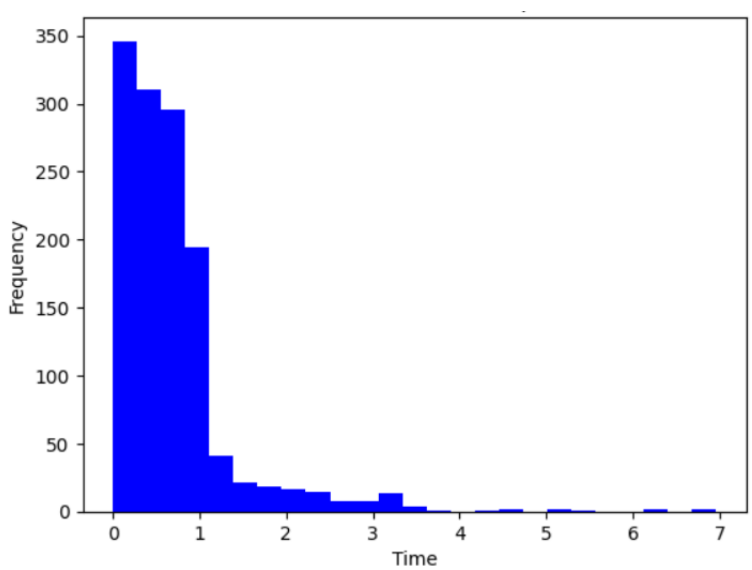


Figure 7: Histogram of FPT with $N = 10000$ and $d = 2000$ with 1300 samples

To further examine the departure from Gaussian behavior, we generated a QQ plot comparing the empirical distribution to a standard Gaussian distribution. Figure 8 illustrates this QQ plot.

From the figures, we can see that the empirical distribution of the FPT does not conform to a Gaussian distribution. However, based on the well-known fact from [3], the distribution

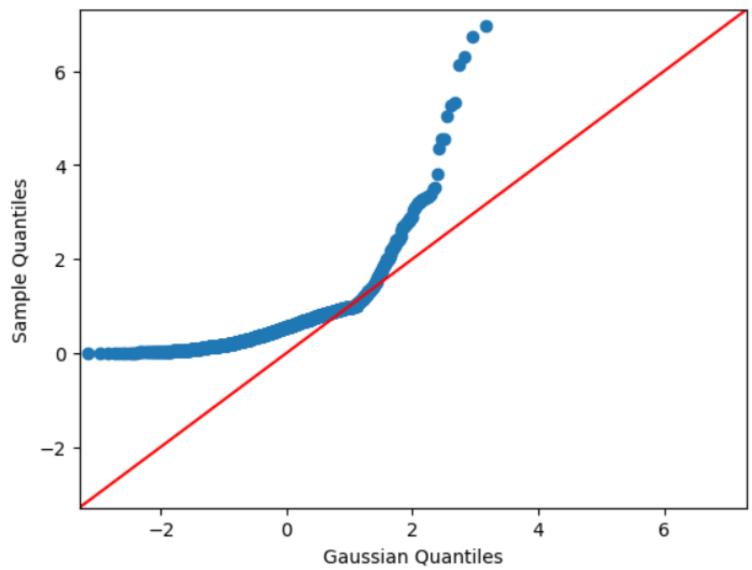


Figure 8: QQ Plot of the histogram with Standard Gaussian Distribution

is a combination of Gumbel distributions when $d = N - 1$, we conjecture that the behavior observed in this regime represents an intermediate transition toward the combination of Gumbel distributions. Overall, our analysis of the empirical distribution of the FPT in different cases provides insights into its behavior by identifying the conditions under which the distribution approximates a Gaussian or deviates from it.

5 Conclusion and Future Goals

In this project, we explored the phenomenon of rumor propagation on a spread-out line graph L_N^d. We introduced the concept of a spread-out line graph and investigated the first-passage time (FPT) from the first vertex to the last vertex according to the value of d. By simulating multiple samples and analyzing the obtained data, we made several observations and proposed conjectures.

First, we analyzed the number of non-return points in the graph according to the value of d. We normalized the number of non-return points by dividing it by N and referred to this normalized quantity as the fraction of non-return points (NRP). We observed that the NRP generally increases until d reaches a certain threshold and then decreases as d approaches $N - 1$. Based on our observations, we proposed a conjecture that the NRP converges to 0 as d increases to N. Furthermore, we investigated the NRP at its maximum when $d = 5$ and conducted simulations with varying values of N and d. The results consistently showed that the NRP is maximized when $d = 5$ for sufficiently large N. Based on these findings, we proposed a conjecture that when all the edges are uniformly distributed from 0 to 1, the fraction of non-return points is maximized when $d = 5$.

Next, we examined the expected first-passage time (FPT) for different values of d and N. We identified three distinct cases of FPT behavior: $d \ll \sqrt{N}$, $d = O(\sqrt{N})$, and $d \gg \sqrt{N}$. For the case of $d \ll \sqrt{N}$, when d is significantly smaller than the square root of N, we have found that the FPT grows linearly as N increases. The ratio of FPT to N remains relatively constant, indicating a consistent rate of growth. Therefore, we claim that the expected FPT tends to infinity as N approaches infinity. In the second case, where d is proportional to the square root of N, we have observed that the FPT tends to decrease as N increases. However, the rate of decrease becomes smaller and actually, we can conclude that in this case, the expected FPT converges to a positive constant as N grows. The third case occurs when d is significantly larger than the square root of N. In this case, the FPT exhibits a logarithmic behavior as N increases. We can interpolate the results from the line graph to the complete graph based on a previous study in [3] that the FPT in a complete graph with $d = N - 1$ is $\frac{\log N}{N}$ as N tends to infinity.

Lastly, we investigated the empirical distribution of the first-passage time by conducting simulations of many samples. When the distance parameter d is significantly smaller than $\sqrt{N}$, the empirical distribution of the first-passage time (FPT) closely approximates a Gaussian distribution. This is supported by the histogram (Figure 5) and QQ plot (Figure 6) where the empirical distribution aligns well with a standard Gaussian distribution. Conversely, when d is much larger than the square root of n, the empirical distribution of the FPT deviates from a Gaussian distribution. This deviation is evident from the histogram shown in Figure 7.

Our future goal is to continue simulating larger samples on more complex graph structures, including a more generalized lattice graph. Furthermore, it is currently unknown whether the performance of FPT and NRP remains the same for edge weights with other random variables besides the uniform distribution. We would like to investigate the growth structure (behavior of the structure) with other edge weight distributions. Additionally, we would like to analyze another interesting problem: the number of nodes receiving the rumor within time t. Finally, while currently using the Dijkstra algorithm, we would like to explore and compare strategies to find the most optimal cost.

6 References

[1] Auffinger, A.; Damron, M.; Hanson, J. (2016). 50 years of first passage percolation, arXiv:1511.03262.

[2] Zhu, L. (2023). Partial differential equation modeling of rumor propagation, Simiode, Qubes Educational Resources, doi:10.25334/S176-C424.

[3] Janson, S. (1999). One, two and three times $\log n/n$ for paths in a complete graph with random weights. Cambridge University Press.

[4] Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford (2001). Introduction to Algorithms (Second ed.). MIT Press and McGraw–Hill. pp. 595–601. ISBN 0-262-03293-7.