Random Walk on a Network

The Impact of Random Diffusion Processes in a Non-Spatial Domain

Nathan Pacey

Email: 18njp@queensu.ca

SN: 20153310

April 7, 2024

Final Report for ENPH 317 Mathematical Methods in Physics II

1.0 Introduction

The notion of randomness permeates various branches of mathematical physics, particularly in the study of diffusion. Conventionally, diffusion is discussed in spatial terms, referring to the process by which particles spread from a region of higher concentration to one of lower concentration, often following Brownian motion [1].

This principle, while extensively studied in physical environments, also translates into the abstract domain of networks, where the concept of space is replaced by the structure of interconnections. In a network, a random walk is a sequence of steps from one node to another, with each step determined probabilistically by the network's adjacency properties. The network comprises nodes, representing entities, connected by edges that symbolize the potential pathways for movement or information transfer [2]. The random walk on such a network becomes a powerful tool for modeling and predicting behavior in complex systems, where the spatial constraints are less defined, but the connections and their strengths are crucial.

The mathematics of a random walk on a network is founded on the probabilities associated with each step from one node to its connected peers. These probabilities are often derived from the degree of each node—how many connections it has—and the overall topology of the network. The focus on the convergence to a steady state, or equilibrium, is of particular interest [3] [2]. The steady state is a condition where the probabilities of being at each node no longer change significantly with additional steps. Mathematically, this state is characterized by a probability distribution across the nodes that remains constant over time.

Measuring the convergence to a steady state solution provides profound insights into the network's dynamics. It allows us to quantify the efficiency of information spread, identify bottlenecks or critical nodes within the network, and assess the resilience of the system to changes or failures among its constituents. In the broader context, understanding the steady state behavior of random walks on networks helps us to dissect and influence a range of phenomena, from epidemic spreading to communication patterns in social media networks [3]. This project is geared towards examining how a random walk evolves over a network and what the characteristics of the steady state tell us about the underlying system, with a particular emphasis on the mathematical framework that governs these processes.

2.0 Problem Definition

This project investigates the dynamics of a random walk on a network, where the nodes symbolize individuals and the edges represent the connections between them. Using an established network, characterized by an adjacency matrix that maps the friendships within the community, the objective is to determine the steady-state probabilities for each individual to possess a metaphorically passed box. To achieve this, we will develop a mathematical model of the random walk, grounded in the connectivity data provided by the adjacency matrix [4].

$$A_{ij} = \begin{cases} 1 & if\ person\ i\ and\ j\ are\ friends \\ 0 & if\ person\ i\ and\ j\ are\ not\ friends \end{cases} \tag{1}$$

The simulation component of this project will pinpoint when the network reaches a steady state and what the associated probability distribution looks like. This involves calculating the probabilities $p_i(t)$ that individual i has the box at time t, with the transition probabilities being influenced by the individual's number of connections within the network, indicated by k_j [4].

$$p_i(t) = \sum_j \frac{A_{ij}}{k_j} p_j(t-1) \tag{2}$$

This equilibrium is mathematically expressed through the convergence of p_i (t) as t approaches infinity, where the network's topology—the arrangement and weighting of its connections—dictates the final probability distribution [4].

Following this analysis, the investigation aims to use the model to inform strategies for navigating the interpersonal dynamics within the network. A scenario is considered where an individual might desire or disdain possession of the "box." The task is to analytically deduce which connections within the network would increase or decrease the chances of an individual receiving the "box," implying a strategy for whom to befriend or avoid, respectively. This requires a nuanced understanding of how individual positions and connections within the network's topology impact the overall probability distribution in the random walk's steady state. This study will explore the interplay between the network structure and the resulting dynamics of the random walk, using the graph Laplacian [4].

$$L = D - A \tag{3}$$

With the degree matrix D, described by the degree of connections k_j along the diagonal elements of node j [4].

$$D = diag(k_1, k_2, k_3, \dots, k_N) \tag{4}$$

Note that this means that the probability equation can be simplified.

$$p_i(t) = \sum_j D_{jj}^{-1} A_{ij}\, p_j(t-1) \tag{5}$$

Which is known as the discrete-time Markov chain. Where the eigenvectors of the Laplacian illuminate the steady-state characteristics. This work aims not only to find a solution for this particular network model but also to contribute to the broader understanding of how network configurations can influence the outcome of such random processes.

3.0 Methodology and Analysis

3.1 Network Visualization and Friendship Degrees

To visualize the network, the Python libraries matplotlib and networkx were utilized. Networkx facilitated the creation and manipulation of the network graph, while matplotlib provided the tools to display it, with nodes color-coded to reflect their degree of connectivity representing friendships within the network.

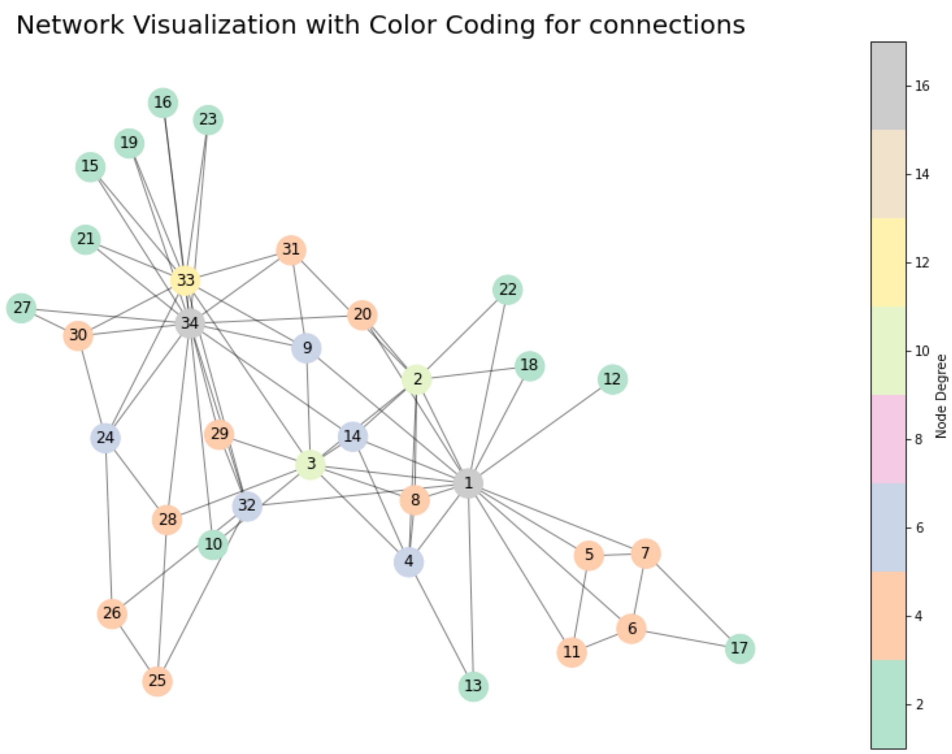

Figure 1: Network visualization including colorized nodes representing the number of connections or degrees each node has.

A histogram can be employed to visualize the connection degree of each node, which is important for quantitatively understanding the frequency distribution of connections within the network and identifying key nodes that may significantly influence the network's dynamics.

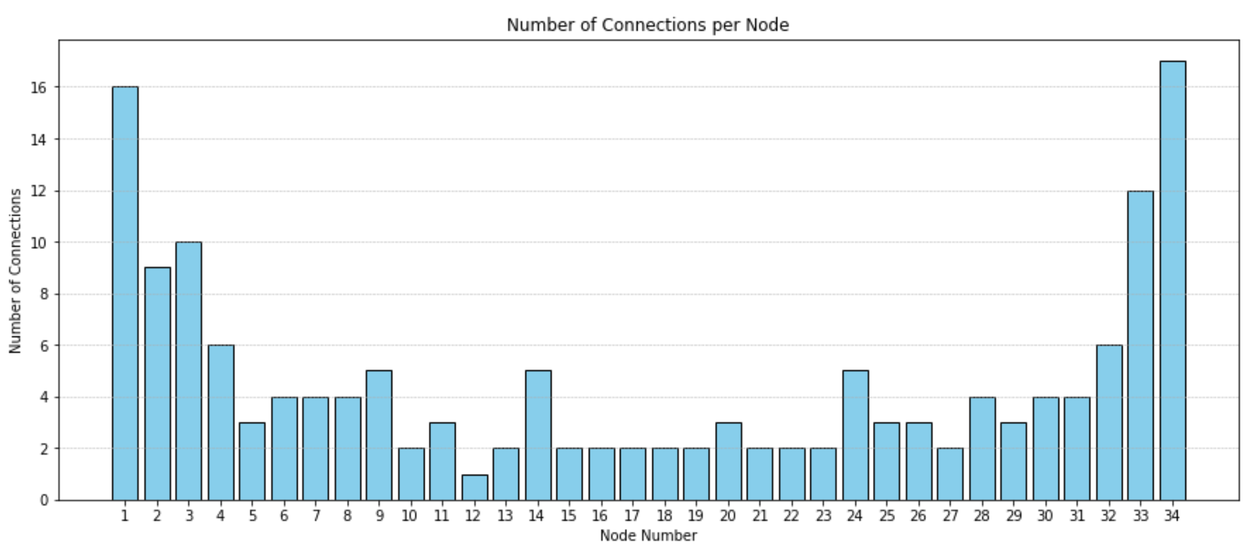

Figure 2: Histogram plot of the number of connections (friends) each node in the network has.

The probability of a node being the end-point of a random walk at steady state is directly correlated with its number of connections; hence, the steady-state solution typically mirrors the histogram's structure, underscoring the connection degree's influence on network behavior.

3.2 Steady State Probability Distribution Methodology

The steady-state probability distribution of the network was computed through a discrete-time Markov chain model using a Python script. This model was realized by first constructing an adjacency matrix A, where entries Aij are set to 1 to represent a connection between nodes i and j, and 0 otherwise. In parallel, a degree matrix D was assembled, with diagonal entries corresponding to the number of connections, or degrees, each node has.

Initialization of the simulation involved setting up a probability vector p, with all elements at zero except for the starting node, which is set to one to indicate the initial location of the box. The script then enters an iterative loop to simulate the Markov process. During each iteration, it updates the probability vector according to the Markov chain equation as described by equation 5.

The loop continues until the simulation reaches what is determined as a steady state—specifically when the relative change in probabilities between successive iterations is less than or equal to 1%. This threshold confirms that the probabilities have settled, and no significant shifts are observed, indicating that the network has stabilized to an equilibrium (steady state).

At this point, the steady-state probability distribution is determined by taking the last column of the probability matrix p, representing the long-term probabilities of the 'box' being at each node. The resulting steady-state distribution provides critical insights into the dynamics of the network, revealing the nodes' long-term behavior and their likelihoods of possessing the box when the system's random walk has equilibrated.

3.3 Steady State Analysis

A visualization was created to map the steady-state probabilities against the nodes' degrees from a randomly chosen starting point, serving to evaluate the precision with which the probability distribution aligns with the nodes' connective extents.

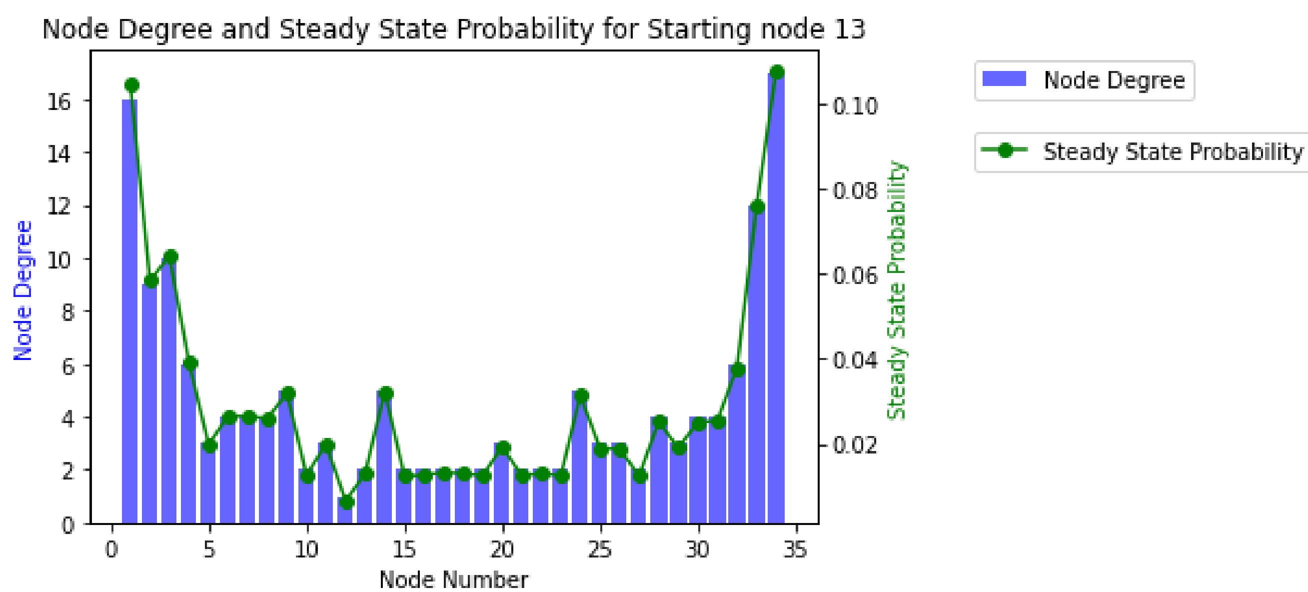

Figure 3: Steady state probability distributions for a random walk with an iterative threshold of 1% plotted against the number of node connections.

To capture the nuances of the network's behavior, a visualization was constructed to display the steady-state probabilities for each potential starting node, reflecting the subtle deviations in each node's chance of receiving the box. These variations, stem from the iteration accepted error margins (threshold). By plotting the probability distributions with both larger and smaller threshold values, one can observe the convergence toward a single steady-state solution. This approach demonstrates how varying levels of precision affect the time it takes for the system to reach equilibrium.

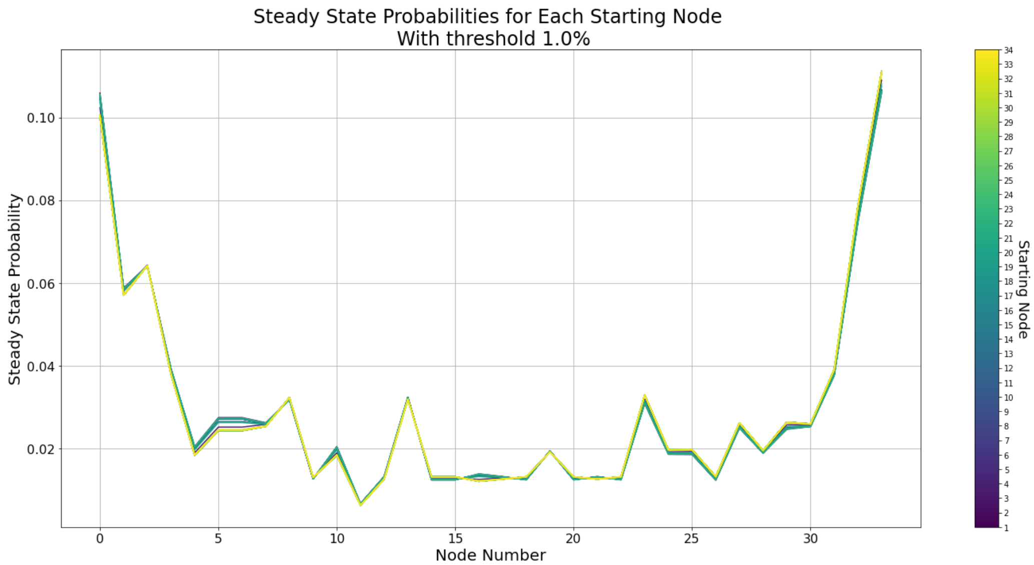

Figure 4: Steady state probability distribution of each node receiving the box over each starting node with an iterative threshold of 1%.

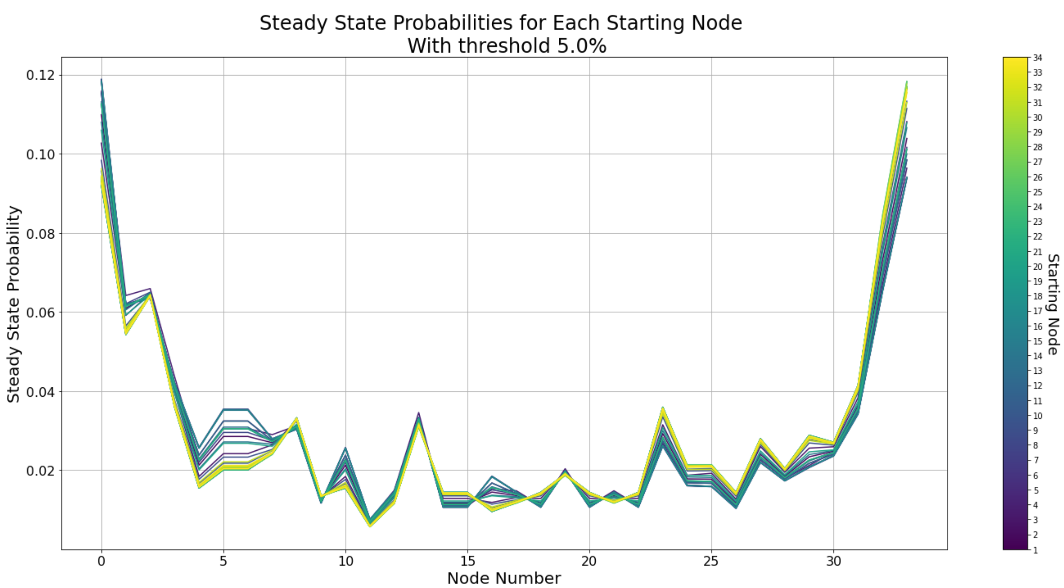

Figure 5: Steady state probability distribution of each node receiving the box over each starting node with an iterative threshold of 5%.

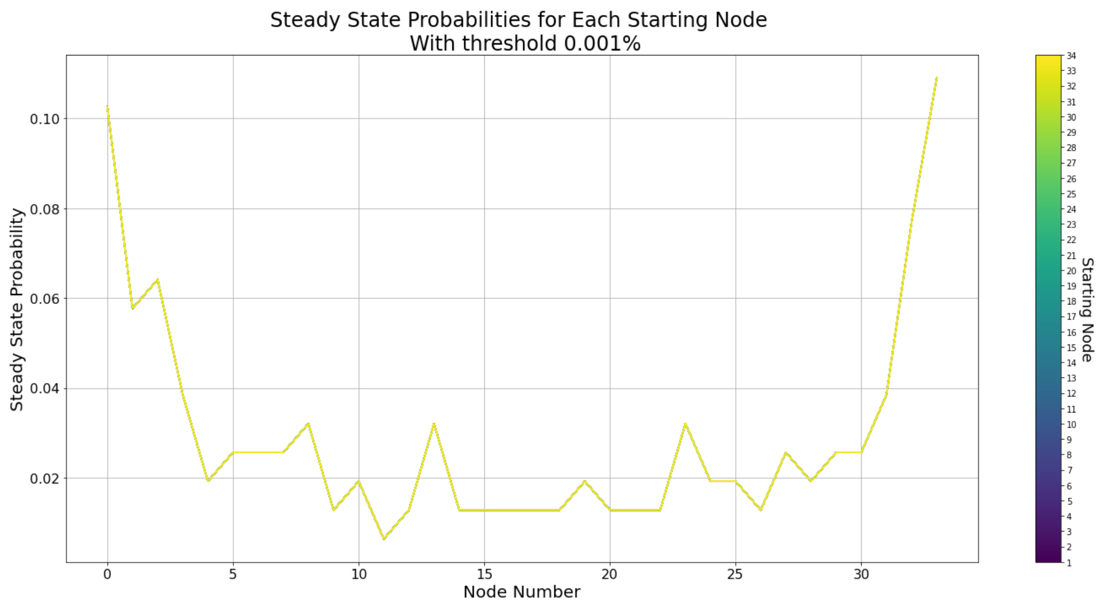

Figure 6: Steady state probability distribution of each node receiving the box over each starting node with an iterative threshold of 0.001%.

As the threshold approaches zero, the differences in probabilities originating from different starting nodes decrease, and the probabilities converge to a single steady-state solution. This convergence is effectively illustrated when using a very small threshold, such as 0.001%, to approximate the steady state.

By reducing the threshold, the steady-state approximation becomes more precise and can be graphically compared with each node's degree of connections. This visualization demonstrates that as the threshold decreases, the steady-state probabilities align more closely with the nodes' connectivity, confirming the intuitive correlation between a node's number of connections and its likelihood of holding the 'box' in the steady state.

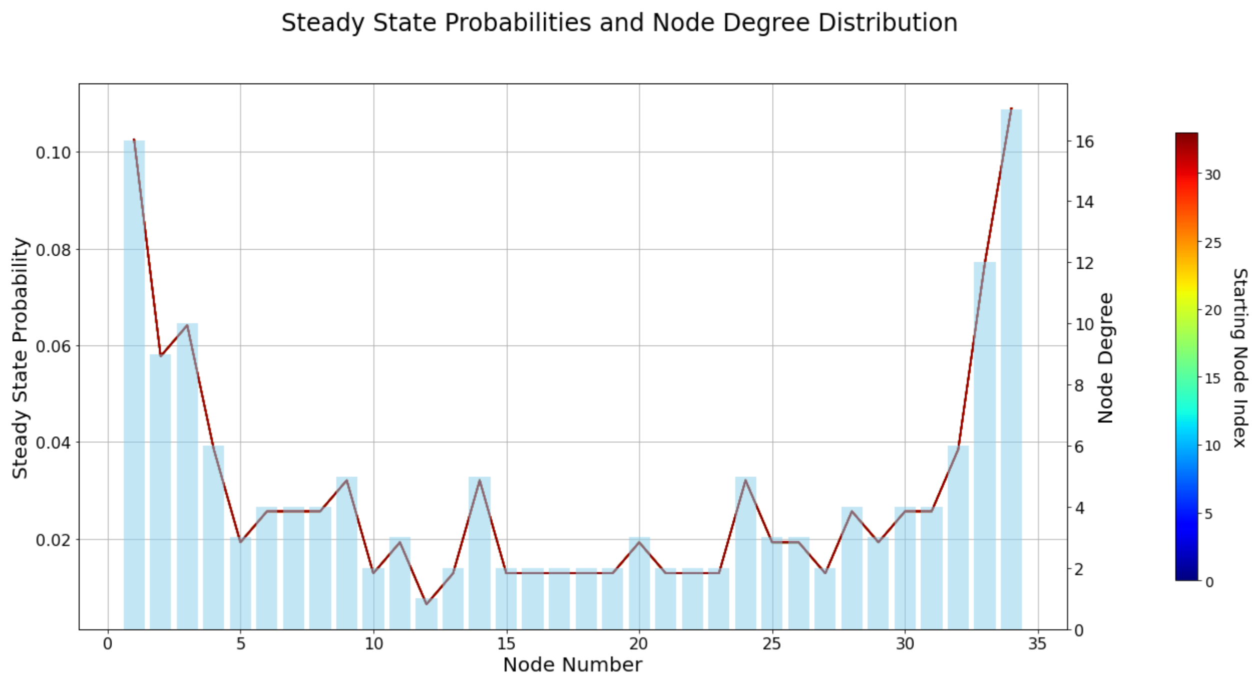

Figure 7: Steady state probability distribution of each node receiving the box over each starting node with an iterative threshold of 0.001% against the histogram of the degree of connections for each node.

The graph thus reveals the increased accuracy of the model in reflecting the network's structure as the threshold approaches a smaller value.

3.4 Laplacian Relationship to Probability Distribution

The Fiedler vector is a fundamental concept in spectral graph theory where the magnitude of the Fiedler value itself is indicative of the graph's algebraic connectivity: the larger it is, the more connected the graph is overall. The Fiedler vector is the eigenvector corresponding to the second smallest eigenvalue (known as the Fiedler value) of the network's Laplacian matrix L, which is defined by equation 3 [5].

The components of the Fiedler vector provide insight into the graph's connectivity and are instrumental in identifying potential clusters or partitions within the network. Networks with well-defined clusters will have a Fiedler vector with components that take on distinctly different values for nodes in different clusters. This characteristic makes the Fiedler vector especially useful in tasks like network community detection, where the goal is to determine groups of nodes more densely connected internally than with the rest of the network [6].

For this network, the Fiedler vector for each node was determined to be the following where the order is ascending from node 1 to 34.

```
Fiedler vector:
 [-0.11213743 -0.04128789  0.02321896 -0.05549978 -0.28460453 -0.32372722
 -0.32372722 -0.052586    0.05160128  0.09280089 -0.28460453 -0.21099295
 -0.1094613  -0.01474197  0.16275078  0.16275078 -0.42276533 -0.10018142
  0.16275078 -0.01363713  0.16275078 -0.10018142  0.16275078  0.15569457
  0.15302558  0.16096292  0.18710956  0.12766405  0.0951523   0.16765031
  0.07349964  0.09875343  0.13034546  0.11890326]

Nodes with positive Fiedler vector components:
 [ 3  9 10 15 16 19 21 23 24 25 26 27 28 29 30 31 32 33 34]
Nodes with negative Fiedler vector components:
 [ 1  2  4  5  6  7  8 11 12 13 14 17 18 20 22]
```

Figure 8: The Fiedler values of each node in the network along with arrays of the two groups of nodes as defined by these values (positive and negative Fiedler vector groups).

The network graph can be effectively segmented into two separate clusters by identifying nodes with positive and negative components of the Fiedler vector. The network was plotted, distinctly visualizing nodes with positive Fiedler vector components in red and negative ones in blue, thereby elucidating the network's clustered structure.

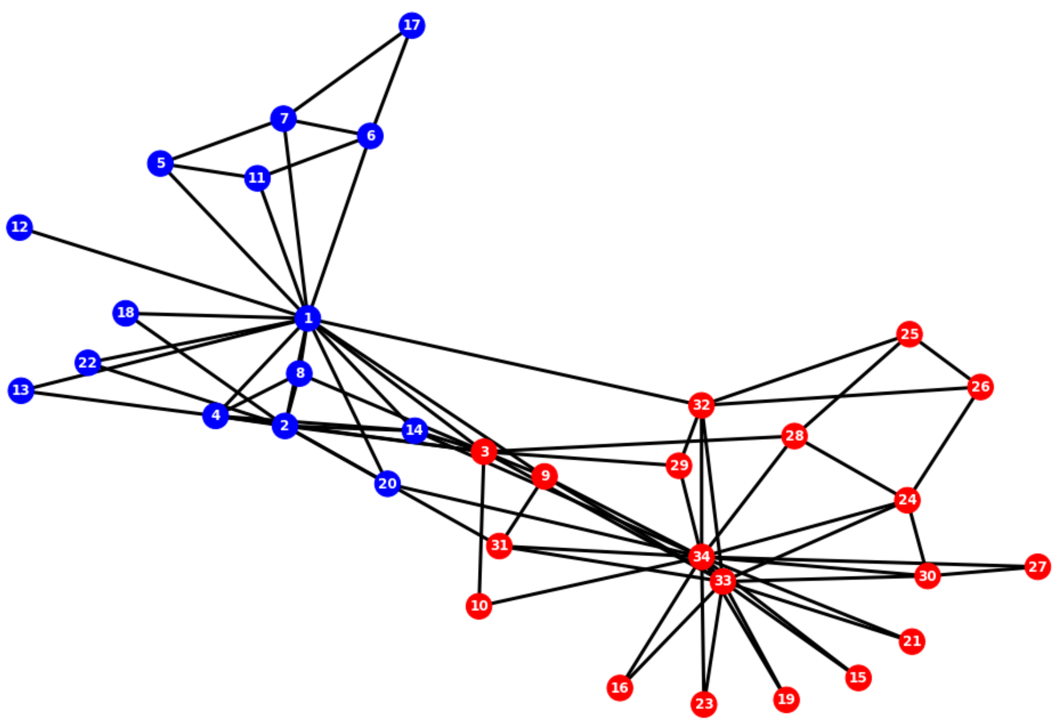

Figure 9: Network visualization with specific clusters as identified by positive Feidler value nodes in red and negative value nodes in blue.

3.5 Probability Impacts of the Fiedler Vector

For Further reinforcing the idea that Fiedler vectors delineate groups of nodes with unique characteristics and geometrical relationships, these groupings become evident when the threshold is adjusted. Nodes inclined towards directing the box to node 1, and those favoring node 34, align with the nodes associated with negative and positive Fiedler vector components, respectively. This delineation was highlighted by plotting the steady-state probability distributions for each set of Fiedler vector nodes at a threshold of 10%.

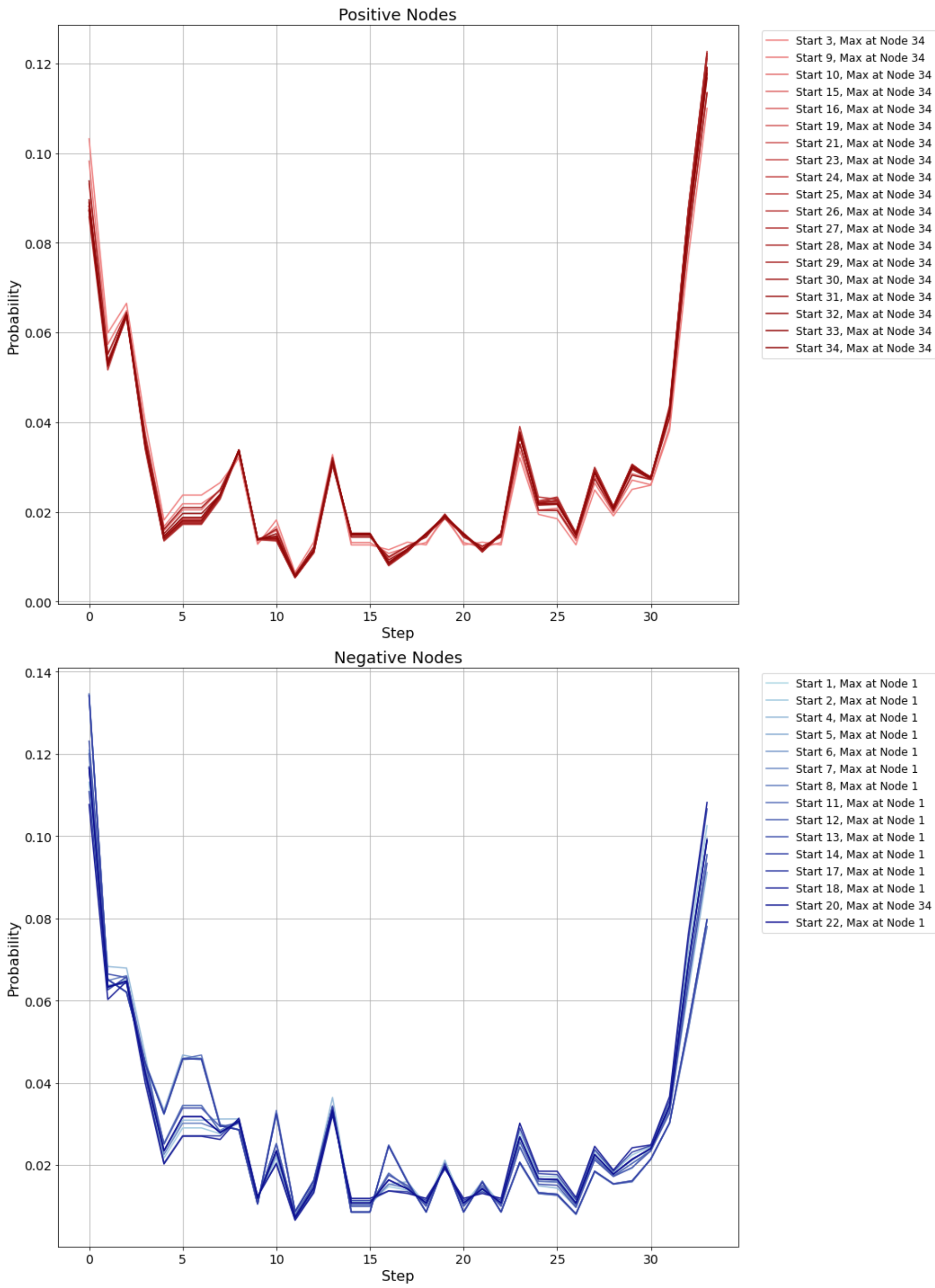

Figure 10: Positive and negative Fielder vector nodes probability distribution for each starting node with a threshold of 10%.

This outcome is anticipated because raising the threshold reduces the number of iterations in the random walk, making the results more sensitive to the immediate connections of each node, essentially emphasizing the local network structure around each starting point.

It's important to highlight that the clear grouping of nodes, as categorized by the Fiedler vector, emerges not solely at high thresholds but also in the specific probability magnitudes for each node, particularly noticeable for nodes 1 and 34. This suggests that the Fiedler vector's role in defining distinct node clusters is evident across different aspects of the network's dynamics, including the individual probabilities of nodes. This was shown for a threshold of 0.001% across each node.

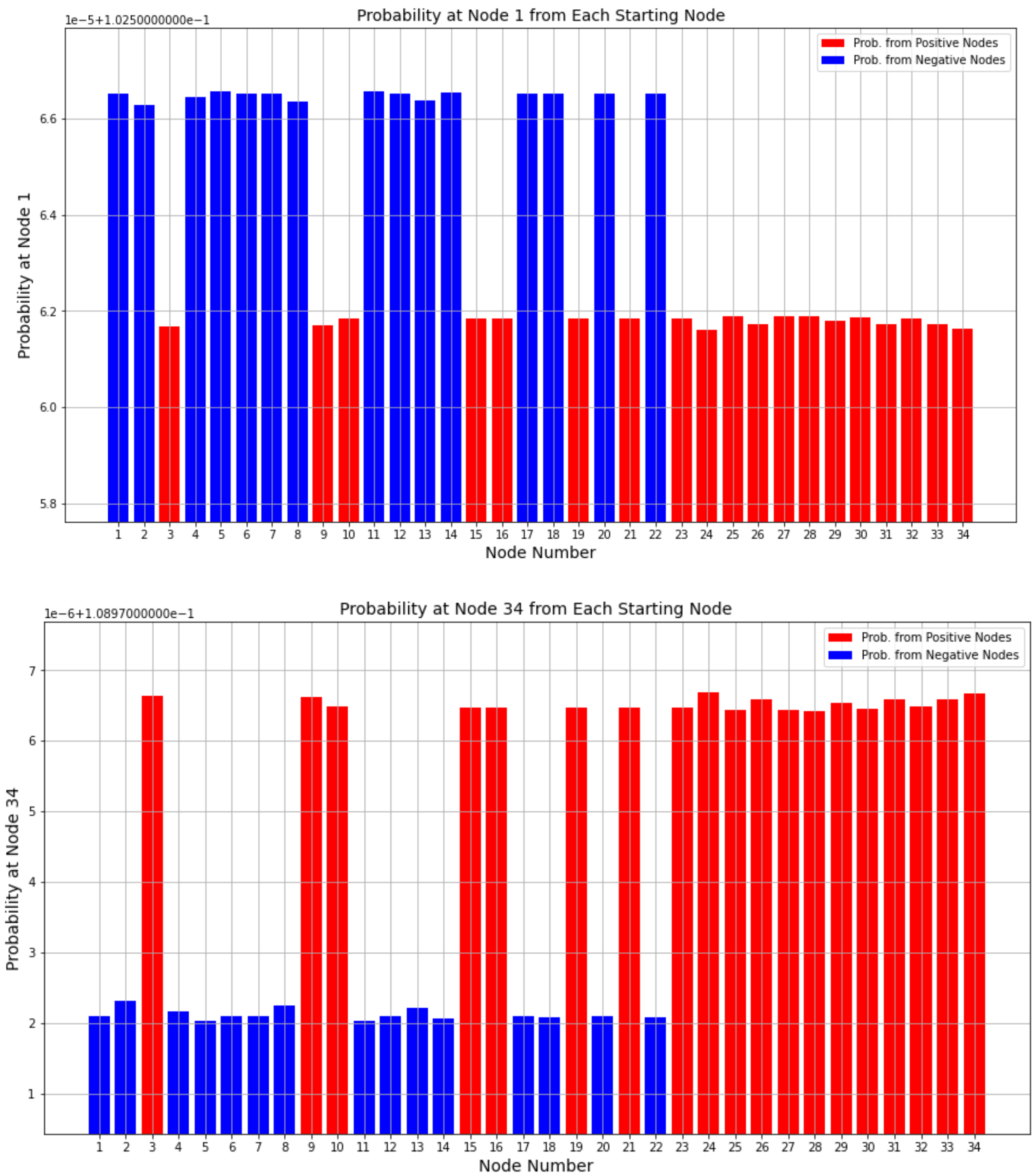

Figure 11: Magnitude of probability distribution for nodes 1 and 34 with positive and negative Fielder vector nodes specified.

4.0 Results

Our analysis revealed that as the threshold approaches zero, the steady-state solution for each starting node converges to a singular probability distribution across all starting points. This distribution is found to be directly proportional to the degree of connectivity each node possesses, implying that nodes with a higher number of connections are more likely to end up with the box in a myriad of random walk scenarios. Consistent with expectations, node 34 emerges as the most probable endpoint due to its superior connectivity. The derivation of the Laplacian and subsequent computation of Fiedler vectors shed light on the network's inherent structure, segmenting it into two distinct clusters based on the sign of the Fiedler vectors. This bifurcation is evident in the probability distributions for nodes within each cluster, influenced by their network positions and connections. A notable pattern observed with a lower threshold is the alignment of nodes with negative Fiedler vectors towards giving node 1 a higher probability of receiving the 'box', whereas nodes with positive Fiedler vectors predominantly favor node 34. This delineation is congruent with the visual network representation, where nodes with negative Fiedler vectors predominantly connect to node 1, and those with positive vectors to node 34.

Therefore, in general if a new node is added to the network and wants to receive the contents of the box connecting (or befriending) node 34 is the optimal choice. However, if the new node is being added and already is friends with negative fielder starting nodes it would be beneficial to connect with node 1. The same but opposite is true is the node wants to avoid receiving the contents of the box.

4.0 Conclusion

This report has explored random walks on a network, focusing on how the structure of connections influences the probability distribution at steady state. The primary goal was to understand how the random walk's outcome depends on the network's topology, particularly the number of connections each node has. Using the Laplacian matrix and Fiedler vectors, distinct clusters were identified within the network, which helped in understanding the network's geometry. The analysis revealed that nodes with more connections, like node 34, are more likely to be the endpoint of the random walk. This aligns with the intuitive understanding that more connected nodes have a higher chance of receiving the box. Furthermore, by varying the threshold for steady-state convergence, it was observed that the starting node's influence diminishes as the threshold decreases, leading to a convergence towards a consistent probability distribution across all nodes. This distribution closely correlates with each node's degree of connectivity. Based on the analysis and impact of the structural geometry of the network as highlighted by the Fielder vectors a new node looking to receive the box should become connected to highly connected nodes such as 34 or 1. Conversely if the new node does not want to receive the box it would be best to connect with lower probability (and connected) nodes such as node 12. Overall, this study sheds light on the significance of network topology in random walks and provides insights that could be valuable in various applications, from optimizing information flow in social networks to understanding spread mechanisms in epidemiological models.

References

[1] Complexity Explorer, "Random Walks Tutorial: Brownian Motion," Youtube, 3 April 2019. [Online]. Available: https://www.youtube.com/watch?v=89LP5pHWxM8&ab_channel=ComplexityExplorer. [Accessed 1 April 2024].

[2] N. Masuda, "Random walks and diffusion on networks," *Physics Reports,* pp. 1-58, 2017.

[3] L. Speidal, "Steady state and mean recurrence time for random walks on stochastic temporal networks," *Physicsal Review,* vol. 91, no. 012806, pp. 1-6, 2015.

[4] G. V. Anders, "Resource: Networks," Queen's University, Kingston, 2024.

[5] E. Weisstein, "Fiedler Vector," Wolfram, [Online]. Available: https://mathworld.wolfram.com/FiedlerVector.html. [Accessed 4 April 2024].

[6] J. Demmel, "CS 267: Notes for Lecture 23, April 9, 1999. Graph Partitioning, Part 2," UC Berkeley, [Online]. Available: https://people.eecs.berkeley.edu/~demmel/cs267/lecture20/lecture20.html. [Accessed 4 April 2024].