

NETWORK ANALYSIS

Assignment 2 Report
SIR analysis on rat brain

Szymon ZINKOWICZ
Matricola number: 5181814

May 8, 2024

University of GENOVA
Instructor: Professor MARINA RIBAUDO

Abstract

This study explores the dynamics of infection spread on a biological network using the Susceptible-Infected-Recovered (SIR) model derived from the mouse-kasthuri-graph-v4 dataset. The primary aim was to investigate the effects of varying the disease transmission probability (p) and recovery probability (q), alongside different initial infection strategies, within a computational network analysis context. The network, selected for its biological relevance and complexity, consists of 1,029 nodes and 1,559 edges, simulating synaptic connections within a rat's brain.

Our findings indicate that increases in the transmission probability (p) significantly accelerated the infection spread, leading to higher peaks of infected nodes, albeit contained within shorter durations. Conversely, elevating the recovery probability (q) delayed the peak infection times, extending the disease's presence within the network. Among the infection strategies tested, those based on node closeness and degree centrality were most effective in spreading the infection rapidly and extensively, suggesting that these metrics are critical in destabilizing network structures under epidemic scenarios. The random infection strategy, however, resulted in the longest recovery times.

The disease did not persist, consistent with the SIR model's predictions, and eventually died out. The network's structure, characterized by 20 connected components compared to a similar-sized random network's 53, suggests a more integrated but segmented topology, which might have influenced the spread and containment of the disease.

Given the scope and results of this simulation, performed as part of a Computer Science network analysis class, we demonstrate the utility and relevance of biological networks in computational studies, shedding light on the pivotal roles of network topology and node centrality in disease dynamics.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 4 |
| 1.1 | Methods | 4 |
| 1.2 | Dataset Description | 4 |
| 1.2.1 | Source and Collection | 5 |
| 1.2.2 | Characteristics | 5 |
| 2 | Model Implementation and Configuration | 5 |
| 2.1 | SIR Model Basics | 5 |
| 2.2 | Model Parameters | 5 |
| 2.3 | Infection Duration | 6 |
| 2.4 | Initial Conditions and Infection Strategies | 6 |
| 2.5 | Simulation Procedure | 6 |
| 2.6 | Data Collection and Analysis | 7 |
| 3 | Experimental Design and Parameters | 7 |
| 3.1 | Overview | 7 |
| 3.2 | Parameter Variations | 7 |
| 4 | Experiment Setup | 7 |
| 4.1 | Rationale and Expectations | 8 |
| 5 | Comparison of Different Probabilities | 9 |
| 5.1 | Experiments 1-4: Varying Transmission Probability (p) | 9 |
| 5.2 | Experiments 5-9: Varying Recovery Probability (q) | 10 |
| 5.3 | Comparison of of infections strategies | 11 |
| 6 | Analysis and Interpretation | 13 |
| 6.1 | Comparative Analysis | 13 |
| 6.2 | Role of Network Topology | 13 |
| 6.3 | Significance of Node Centrality | 13 |
| 6.4 | Practical Applications | 14 |

1 Introduction

1.1 Methods

This study employs the Susceptible-Infected-Recovered (SIR) model on the "mouse-kasthuri-graph-v4" dataset, a detailed representation of the brain's network, to simulate and analyze the dynamics of disease spread.

The dataset provides a biological network framework that allows us to explore how structural characteristics influence epidemic processes. The following methods and analyses were specifically adapted to assess the impact of network properties on disease transmission and recovery:

- **SIR Model Implementation:** Utilizing the well-established epidemiological SIR model to simulate disease spread across the network, emphasizing the impact of node states and transmission dynamics.
- **Parameter Variation:** Systematic adjustment of the disease transmission probability (p) and recovery probability (q), along with various initial infection strategies, to evaluate their effects on epidemic outcomes.
- **Centrality Measures:** Analysis of betweenness, closeness, and eigenvector centrality to identify critical nodes influencing the speed and spread of the infection.
- **Community Detection:** Investigation of how communities within the network affect the containment or propagation of the disease.
- **Network Resilience:** Evaluating how network structure, including connectivity and clustering coefficients, influences the robustness and susceptibility of the network to epidemics.
- **Infection Strategy Analysis:** Comparing different strategies for initiating the infection (based on centrality measures and random selection) to determine their effectiveness in different scenarios.

1.2 Dataset Description

The "mouse-kasthuri-graph-v4" dataset, derived from the somatosensory cortex of a mouse brain, serves as the basis for our network analysis. This dataset is part of the broader efforts by Neurodata's Kasthuri project to provide high-resolution electron microscopy data of neural connectivity, facilitating the exploration of complex biological networks.

2. Model Implementation and Configuration

1.2.1 Source and Collection

Sourced from the Kasthuri lab and available through networkrepository.com, this dataset allows us to map a real biological network onto a computational model to study the spread of infectious diseases.

1.2.2 Characteristics

The dataset features a network graph with the following properties:

- **Number of Nodes:** 1029, representing various cellular components within the brain.
- **Number of Edges:** 1559, indicative of the synaptic connections between these components.
- **Average Degree:** Approximately 3.0301, reflecting the average number of connections per node.
- **Network Type:** Unweighted and undirected, suitable for fundamental network analysis.

This detailed structural and functional representation provides a unique platform for studying the dynamics of disease spread and assessing the impact of network topology on epidemiological outcomes.

2 Model Implementation and Configuration

2.1 SIR Model Basics

The SIR (Susceptible-Infected-Recovered) model divides the population into three compartments: Susceptible (S), Infected (I), and Recovered (R). Each node in the network represents an individual in one of these states. The model simulates the progression of disease through these states based on probabilistic transitions influenced by network interactions among individuals.

2.2 Model Parameters

The dynamics of the SIR model are primarily governed by two parameters:

- **Transmission Probability (p):** Represents the likelihood that the disease will be transmitted between a susceptible and an infected individual during a contact. This probability was varied to observe its effect on the speed and extent of the epidemic spread.

2. Model Implementation and Configuration

2.3 Infection Duration

- **Recovery Probability (q):** Indicates the chance that an infected individual recovers and moves to the recovered state in any given time step. This parameter was also adjusted to see how it affects the duration and resolution of the epidemic.

2.3 Infection Duration

To simplify the model, it was assumed that each infected individual remains in the infected state for a minimum time, t_I , before they have a chance to recover. This duration mimics the infectious period of a disease and varies based on real-world data or hypothetical scenarios.

2.4 Initial Conditions and Infection Strategies

The initial condition of the network was set with all nodes in the susceptible state, except for a small number predefined as infected:

- **Number of Initially Infected (i_0):** A critical parameter that affects the outbreak's potential. Different values were tested to understand threshold effects.
- **Selection of Initially Infected Nodes:** Nodes were selected based on several strategies:
 - **Random Selection:** Nodes are chosen randomly, representing a scenario where disease introduction is sporadic and unstructured.
 - **Centrality-Based Selection:** Nodes with high betweenness, closeness, or degree centrality were chosen to simulate scenarios where highly connected or strategically positioned individuals are the initial carriers.

2.5 Simulation Procedure

Each simulation run involved iterating through discrete time steps. At each step:

- Infected nodes attempted to infect their susceptible neighbors with a probability pp .
- After completing the minimum infectious period t_I , infected nodes had a chance to recover based on the probability q .

4. Experiment Setup

2.6 Data Collection and Analysis

- The network state was updated at each step, tracking transitions from S to I, and I to R.
- The process continued until no infected nodes remained, or a maximum number of time steps was reached.

2.6 Data Collection and Analysis

Throughout the simulation, data were collected on infection rates, recovery rates, and the duration of the epidemic. This data allowed for analysis of the impact of various parameters and infection strategies on epidemic dynamics.

3 Experimental Design and Parameters

3.1 Overview

A series of controlled experiments were conducted to investigate the effects of varying key parameters of the SIR model on the dynamics of disease spread within the network. The experiments were designed to systematically vary the disease transmission probability (p), the recovery probability (q), and the duration of infection (t_I) while observing the resultant epidemic outcomes. Each experiment started with a fixed initial number of infected nodes (i_0) to maintain consistency across different trials.

3.2 Parameter Variations

The experiments were structured into two main groups to isolate the effects of changing transmission and recovery probabilities:

4 Experiment Setup

Here we detail the configuration of each experiment:

4. Experiment Setup

4.1 Rationale and Expectations

| Experiment ID | Transmission Probability (p) | Recovery Probability (q) |
|---------------|----------------------------------|------------------------------|
| exp11 | 0.1 | 0.3 |
| exp12 | 0.2 | 0.3 |
| exp13 | 0.3 | 0.3 |
| exp14 | 0.4 | 0.3 |
| exp15 | 0.5 | 0.3 |
| exp16 | 0.5 | 0.3 |
| exp21 | 0.3 | 0.1 |
| exp22 | 0.3 | 0.2 |
| exp23 | 0.3 | 0.3 |
| exp24 | 0.3 | 0.4 |
| exp25 | 0.3 | 0.5 |
| exp26 | 0.3 | 0.7 |
| exp27 | 0.3 | 0.9 |

Table 1: Summary of Experiments

- a. **Duration of Infection (tI) and Initial Infected (i0):** These parameters were kept constant throughout the experiments, with tI set to 1 and i0 set to 5.
- b. **Varying Transmission Probability (p):** This series of experiments focused on understanding how changes in the likelihood of disease transmission affect the spread and severity of the epidemic. The transmission probability was varied from 0.1 to 0.5, keeping the recovery probability and the duration of infection constant.
- c. **Varying Recovery Probability (q):** This set of experiments was designed to assess the impact of the recovery rate on the disease's duration and the recovery timeline within the network. Here, the transmission probability was held constant while the recovery probability was varied significantly.

4.1 Rationale and Expectations

The rationale behind varying these parameters was to explore how different rates of transmission and recovery influence the epidemiological characteristics of the network, such as peak infection rates, total number of infections, and the speed of epidemic resolution. These experiments are expected to provide insights into the resilience of the network against infectious diseases and the effectiveness of potential interventions.

5 Comparison of Different Probabilities

5.1 Experiments 1-4: Varying Transmission Probability (p)

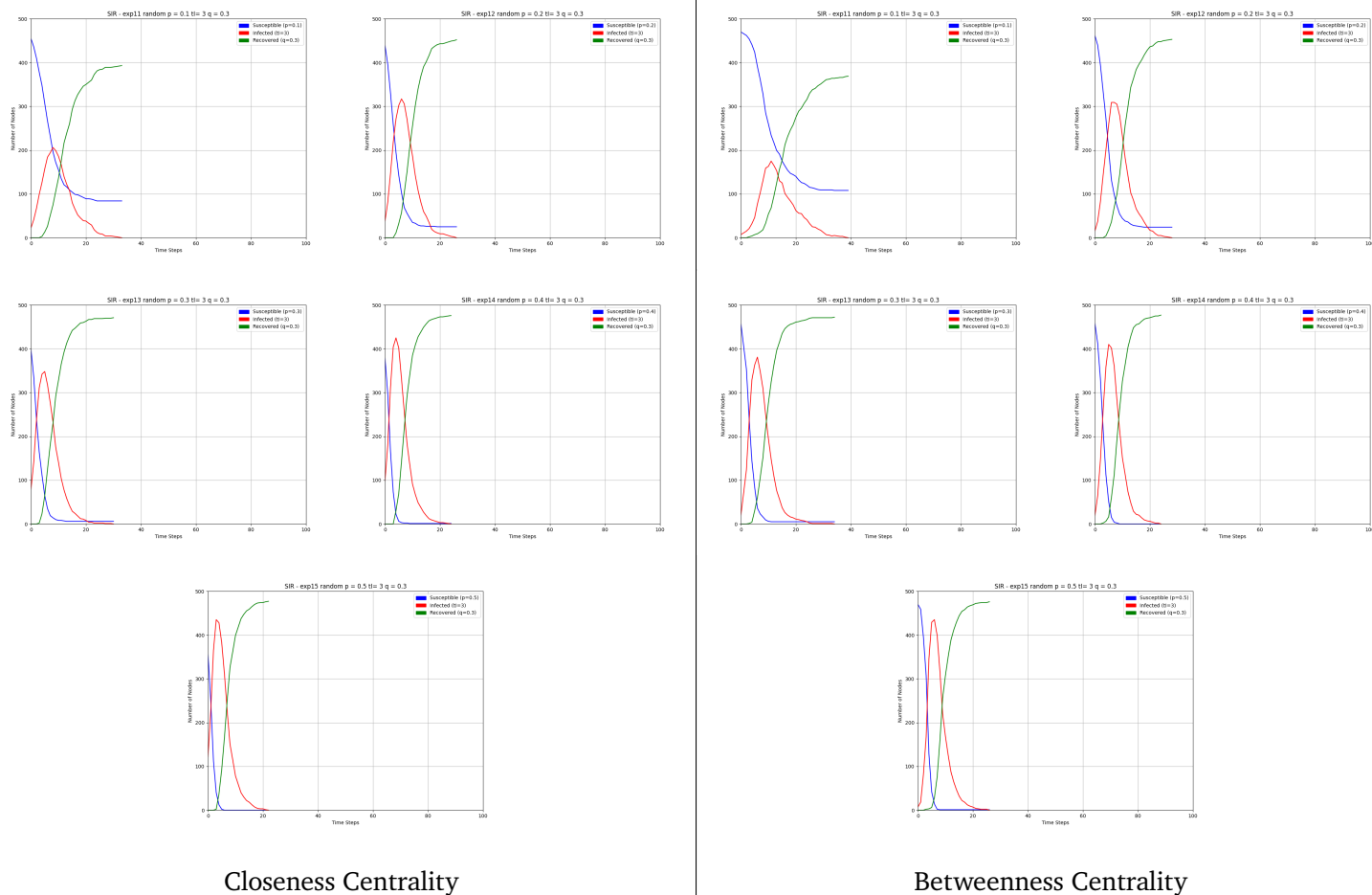


Figure 1: Comparison of Transmission Probabilities: Closeness and Betweenness Centrality

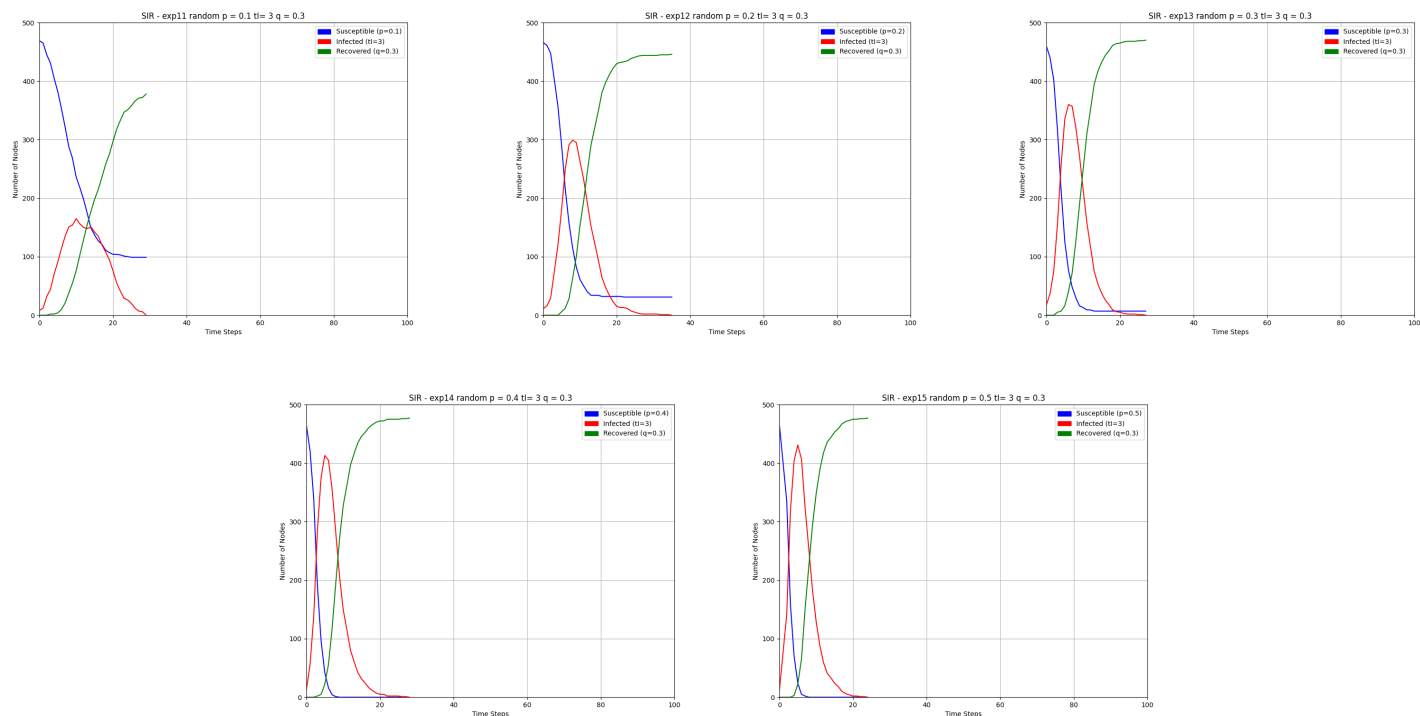


Figure 2: Comparison of Transmission Probabilities: Random Centrality

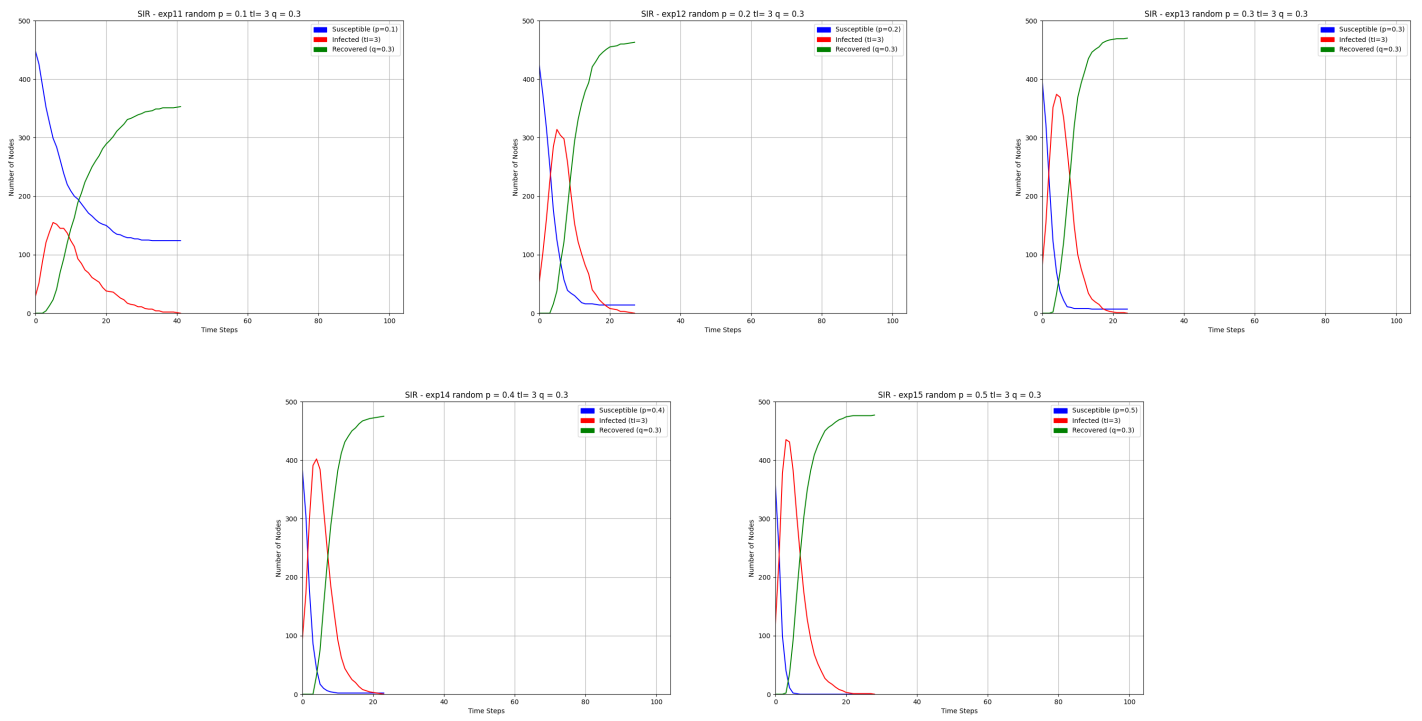


Figure 3: Comparison of Transmission Probabilities: Degree Centrality

5.2 Experiments 5-9: Varying Recovery Probability (q)

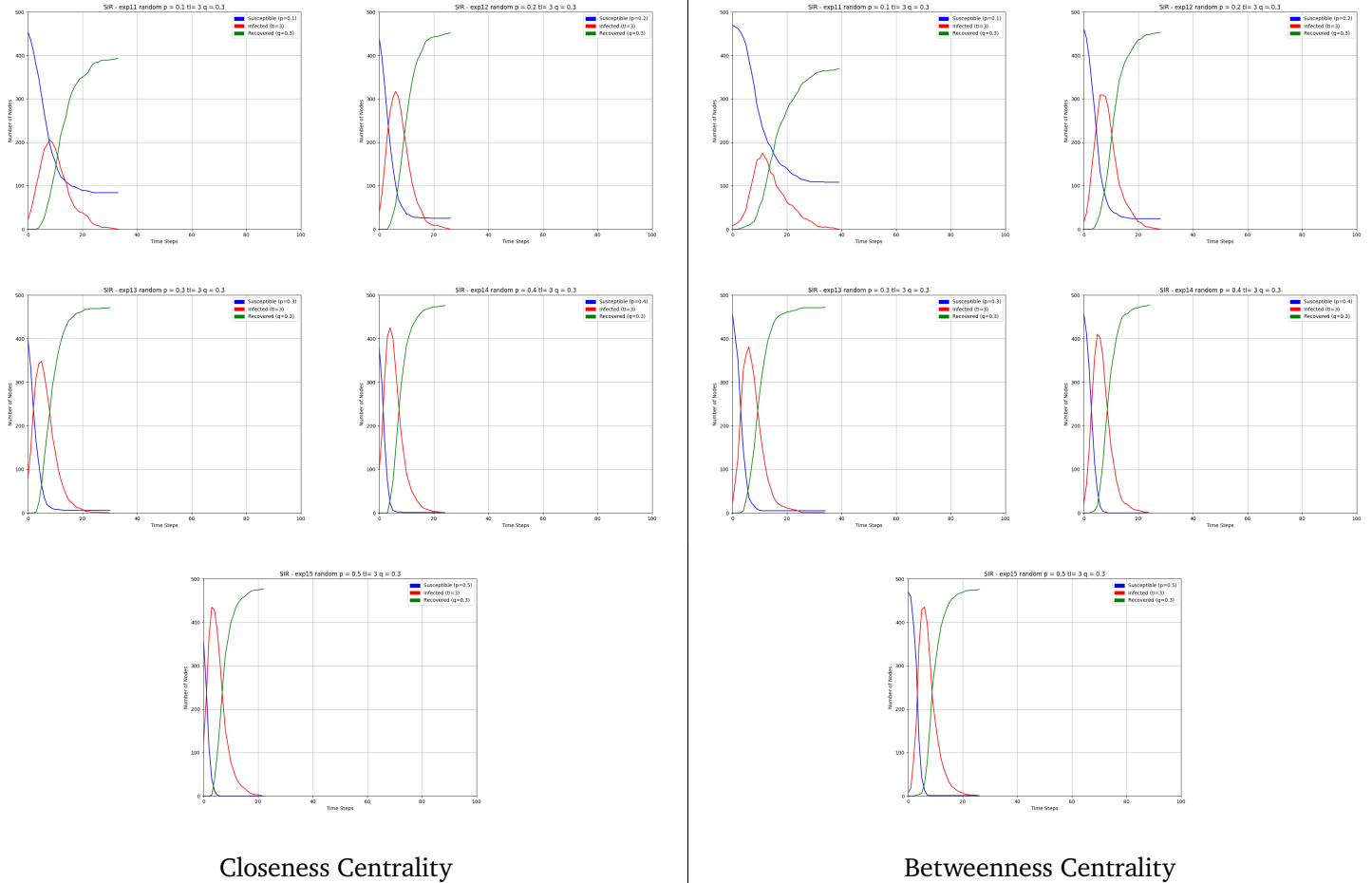


Figure 4: Comparison of Transmission Probabilities: Closeness and Betweenness Centrality

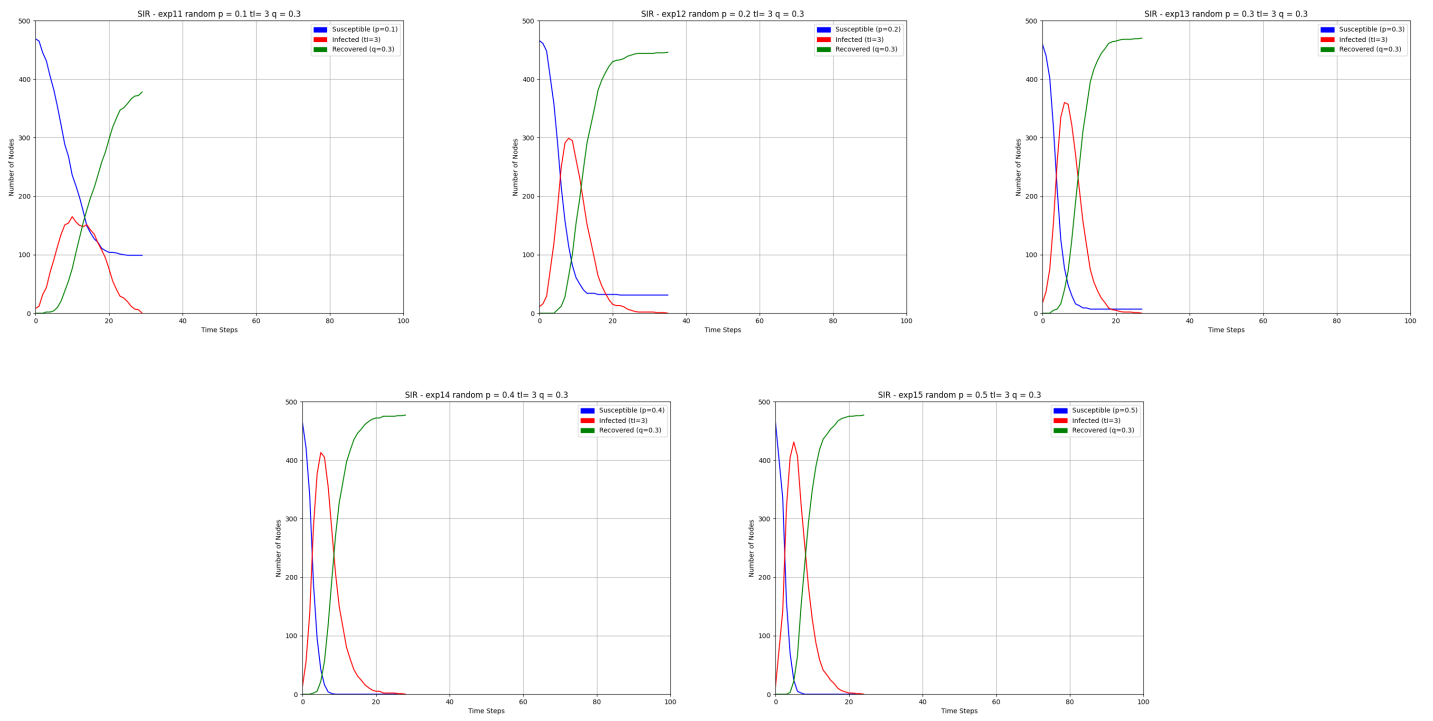


Figure 5: Comparison of Transmission Probabilities: Random Centrality

5.3 Comparison of of infections strategies

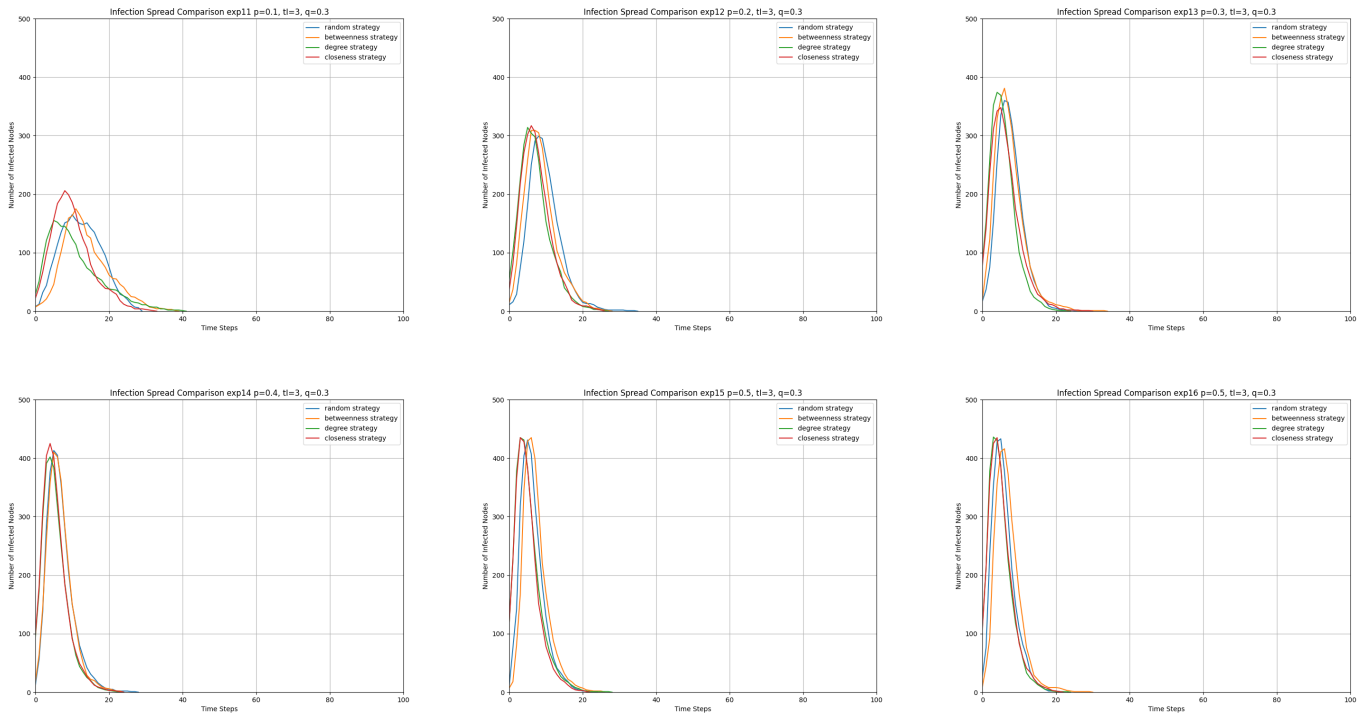


Figure 6: Comparison of infection strategies: Changing transmission probability

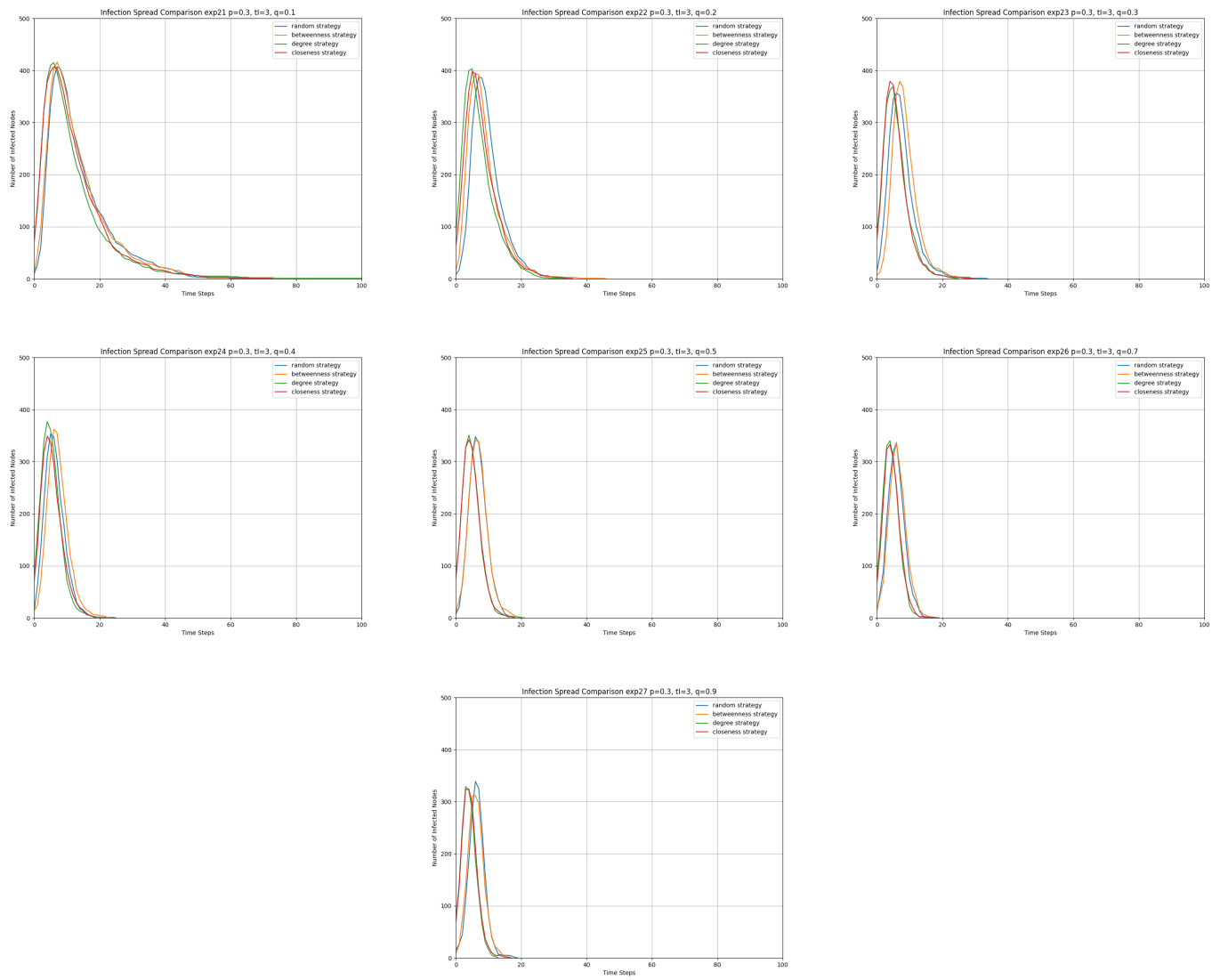


Figure 7: Comparison of infection strategies: Changing recovery probability

6 Analysis and Interpretation

6.1 Comparative Analysis

The simulation data clearly shows that as the transmission probability (p) was increased, the peak infection rates also increased across all strategies. This observation is consistent with epidemiological models, where a higher probability of disease transmission leads to more rapid and widespread infection. Also because of higher and faster peak of infections, the infection died out in significantly less steps from **40** to **30**. Notably, the closeness strategy consistently resulted in the highest peak infection, indicating that nodes with high closeness centrality are crucial in spreading the infection more efficiently across the network. Conversely, the degree strategy resulted in the lowest peaks, suggesting that while highly connected nodes are influential, their impact may be localized depending on the network structure.

As for the recovery probability (q), increasing q led to a reduction in the initial spike of infections, confirming the expected inverse relationship between recovery rate and infection spread. A faster recovery rate tends to stabilize the network more quickly, reducing the number of simultaneously infected nodes and potentially shortening the epidemic's duration from almost **100** to over **20**.

6.2 Role of Network Topology

Regarding the initial infection strategies, the results were nuanced. The betweenness and random strategies resulted in similar timing for infection spikes, but the betweenness strategy was less effective in claiming more nodes than the random strategy. This could suggest that while betweenness centrality identifies crucial nodes in terms of network flow, it does not always correlate with the most effective spread of infectious diseases, possibly due to the layout of the remaining network structure or the specific interactions among nodes. Both degree and closeness strategies were more effective, particularly in terms of reaching earlier and higher spikes in infections, confirming assessment that strategies leading to faster and more extensive disruptions are preferable for understanding potential vulnerabilities in network structures.

6.3 Significance of Node Centrality

The variations in effectiveness among the centrality measures highlight the complex role of node properties in disease dynamics. Closeness centrality's success in achieving higher peaks more quickly suggests that the ability of

6. Analysis and Interpretation

6.4 Practical Applications

nodes to infect others across shorter path lengths is crucial in densely connected networks. On the other hand, the relative underperformance of betweenness centrality could indicate its lesser relevance in networks where disease spread does not hinge on specific "bridge" nodes.

6.4 Practical Applications

- Closeness Centrality: Was effective in causing quicker and higher peaks of infection, supporting the theory that nodes which can reach other nodes more quickly (lower path lengths) are critical in the spread of infections.
- Betweenness Centrality: Despite theoretically being considered crucial for controlling information or disease flow through a network, it was less effective than random selection in some cases. This suggests that the importance of betweenness centrality may be context-dependent, influenced by the specific structure and connectivity of the network rather than universally applicable.