Viral evolution within heterogeneous human contact patterns

Sudarshan Anand*
sanand315@gatech.edu
School of CSE
Georgia Institute of Technology
Atlanta, GA

Rohini Janivara
rjanivara3@gatech.edu
School of Biological Sciences
Georgia Institute of Technology
Atlanta, GA

Alejandro Danies Lopez
adanieslopez3@gatech.edu
School of CSE
Georgia Institute of Technology
Atlanta, GA

Abstract

Understanding the interplay between network structure and viral evolution is essential for predicting disease spread, particularly for rapidly mutating pathogens. Small-world networks, which are characterized by clusters of local connections with occasional long-range links, resemble real-world interaction patterns and transport networks. Such structures are pivotal in determining the likelihood of a localized outbreak escalating into a global pandemic. This research focuses on simulating the spread of an evolving virus within these networks, taking into account mutation-driven changes in transmissibility. By analyzing how network heterogeneity affects the fixation of new strains and the evolution of disease transmissibility, we can gain insights into how viral mutations interact with network structures to promote or hinder epidemic outbreaks. This approach informs strategic interventions and public health policies for outbreak containment in a globally connected world.

Keywords

Epidemiology, Contact Networks, Viral Evolution, Heterogeneity

1 Introduction

Infectious disease modeling is essential for understanding disease spread, especially as we face new and evolving pathogens. With pathogens that mutate rapidly, there is an added layer of complexity, particularly when multiple strains coexist or compete within a population. These interactions, including co-infection and mutation-driven transmissibility changes, complicate efforts to predict disease dynamics and control measures. Many pathogens, especially those that can mutate and adapt over time, cause infections with diverse effects, such as varying lethality and immune responses. While biological insights into host-pathogen interactions help to clarify the severity and progression of disease, complementary epidemiological models that integrate mutation, strain competition, and network effects are vital for predicting and controlling disease outbreaks effectively.

Our work expands on these efforts by developing an epidemiological model that incorporates both viral mutation and competitive dynamics between multiple strains within diverse human contact networks. By using a SISD (Susceptible-Infected-Susceptible-Dead) framework, we simulate disease spread with sequential genetic mutation-driven strain evolution, allowing us to capture how increased transmissibility impacts the emergence and dominance of viral strains. Our approach introduces viral strains V_1, V_2 and V_3 with each V_i arising from V_{i-1} through genetic mutation. Each of these differs in transmissibilities, enabling the study of complex interactions such as cross-infection, strain replacement, and the conditions under which one strain may outcompete another.

In addition to the biological dynamics of infection and mutation, the structure of human contact networks characterized by heterogeneous connectivity and clustering strongly influences the patterns of disease spread. Real-world social interactions vary widely, with some individuals having significantly more contacts than others, leading to hubs of transmission that can accelerate the spread. To capture these dynamics, we simulate disease spread on networks with a range of structural properties, including varying degree distributions, clustering coefficients, and connectivity patterns, and examine how these factors affect strain competition and mutation outcomes.

By integrating both the biological complexity of mutation and competition between viral strains, as well as the heterogeneity of human contact patterns, our model provides a more comprehensive and realistic framework for understanding and predicting the spread of mutating pathogens. This dual focus allows us to test key assumptions about viral evolution and network effects, contributing insights that can inform public health policies managing epidemics of rapidly evolving infectious agents.

2 Response to comments on Milestone

(1) **Specify what you expect out of the project (the hypothesis):** The overall expectations from our project were two-fold:
- When considering strains each formed from sequential genetic mutations of the previous only varying in transmissibility (mutated one is more transmissible than its parent), in case of homogeneous mixing, the most transmissible strain persists in the population eventually, and the time from which it takes over depends on the mutation rate (μ).
- Unlike in the case of homogeneous mixing, heterogeneity in topology would provide a selective advantage to one of the strains over the other and not necessarily to the most transmissible strain.

We aim to answer a few key questions with regard to the following aspects:
- **Viral evolution:**
  - How do sequential genetic mutations, which increase transmissibility ($\beta_1 < \beta_2 < \beta_3$), shape the competition between viral strains?
  - What role does the mutation rate (μ) play in the timing and dominance of emergent strains?

*All authors contributed equally to this research.

DS for EPI, Fall 2024, Atlanta, GA, USA

- **Disease spread dynamics:**
  - How do cross-infections between individuals carrying different strains alter the spread and persistence of each strain?
  - What are the implications of recovery (γ) and mortality (δ) rates on strain replacement dynamics?
- **Network topologies:**
  - Does limiting long-range connections (regular lattice) delay the spread of the disease through the network?
  - Does having higher clustering localize the strains to specific clusters and thus favor co-existence over dominance?
  - Does having hub nodes, which are potential superspreaders level the playing field for the competition?
  - How does real-world network calibration impact the predictability and control of disease spread?
- **Metrics for epidemic analysis:**
  - What is the relationship between network structure and the percolation threshold (P_c) for epidemic onset?
  - How does $R_{0,effective}$ capturing heterogeneity via degree distribution, compare across network types?
  - How do peak infection time (P_t) and proportion (P_I) vary under different scenarios?
  - How do mutation and contact network properties jointly influence epidemic size (E)?

(2) ***Need to look at more realistic networks and datasets:*** We addressed this by calibrating our model based on Influenza infections [1] that provided details on which cases were caused by which strain of the virus. As for the networks, we calibrated our model on the US High School student-teacher-staff contact network[2] from a Stanford study on tracing spread od flu due to proximity contact. More details in the section on calibration (Section 4.3).

3 Related work

Most mechanistic epidemiological models assume homogeneity in human contact patterns. Real-world interactions tend to be heterogeneous. Lloyd-Smith et. al., [3] emphasize the utility of incorporating population level heterogeneity while modeling disease spread beyond standard average measures such as R_0, showing that such average estimates obscure impactful dynamics caused by variations in individual infectiousness. Their model takes this variation into account and finds significant differences in outbreak dynamics. Their results highlight the role of superspreaders in contagion, improving predictions of outbreak dynamics and implementing targeted interventions; in particular, they show that considering non-uniform distributions of transmissibility levels, in the form of non-homogeneous network structures, is key to more accurate modeling.

To understand the dynamics of competing viral strains, researchers study their *antigenic diversity*[1] Additional evidence for the relevance of clustering variations in human contact patterns is provided by Buckee et. al., [4]. They examine the effects of interaction between network communities on antigenic diversity apart from discussing various levels of *cross-immunity* (immunity to one strain implies partial immunity to other strains). The key finding of this paper is that the contact across communities has more pronounced effects than simply the structure of the different communities. Well-connected communities promote accelerated evolution of pathogens by providing a higher potential for competitive co-existence.

Even in scenarios with only one disease strain, as studied by Boots M., and Sasaki A., [5], interactions across communities were found to influence the pathogen's attributes. Networks with more global transmission result in the pathogen developing higher transmission rates and virulence than networks with primarily local transmission, reinforcing the importance of incorporating heterogeneous networks in models.

Along similar lines Rüdiger, et. al [6] study the infection dynamics of mutating pathogens in small-world networks, delineating the effect of network structure on pathogen evolution by characterizing the impact of long-range links on the spread of the disease. They show that the fraction of long-range links drives viral evolution and increases the risk of the infection transforming into an epidemic, noting that a higher mutation rate leads to a higher probability of a nearly network-wide spread, but a higher fraction of long-range links lowers this probability.

Heterogeneity of contact networks can also alter pathogenic evolution through interaction effects. Leventhal et. al., [7] explore this direction, considering two strains with different infection rates but the same recovery rate. With perfect cross-immunity and no co-infection, they monitor the *Fixation probability* (P_{fix}), which is the proportion of simulations where the new strain drives the resident strain to extinction. They monitor this value in two settings of network heterogeneity - degree distribution and the *clustering coefficient* (Definition in Section 4). They found that contact across communities had more of an effect than community structures, consistent with findings from Buckee et. al. [4].

Modeling infection dynamics more accurately involves recognizing relevant disease states like susceptible, infected, recovered, and deceased. Sebbagh et. al. [8] model the spread of COVID-19 in Algeria by implementing an SIRD model to incorporate data on recovery and death. These additional states prove to be highly informative in surveillance and disease prediction especially in the presence of competing pathogens.

In light of dealing with competition between two strains and the stochastic effects of the same on the infection dynamics, Vafadar et. al., [9] utilizes a computational model inspired by Competitive Exclusion Principle (CEP) [10] [2]. By incorporating stochasticity, they elucidate the viral co-infection dynamics within a host and how this could help block the infection by the fatal virus.

While many of these models explore the dynamics of disease spread in heterogeneous contact networks, which are closer to real-world scenarios, a critical assumption in them is that of pathogen evolution, restricting the viral diseases that they can model. Diseases like influenza and SARS-CoV progress through mutating viruses with large inter-strain lethality differences, complicating treatment

[1] **antigenic diversity of a pathogen population:** The variety of different surface proteins (antigens) that pathogens, like viruses or bacteria, can express. These allow pathogens to evade the immune system of their hosts, making it harder for the body to recognize and fight them off.

[2] CEP states that: which in essence states that *"Two species competing for the same limited resources cannot coexist in the same ecological niche for long"*

processes. Thus, to extend this model, we aim to incorporate competition among strains both in an individual and across the contact network and elucidate their utility in modeling disease spread in heterogeneous networks.

4 Methodology

4.1 Network generation

Homogeneous networks: We generate two homogeneous networks - a complete graph on 1000 nodes and a 4-regular graph on 1000 nodes. These are used to isolate and study the effects of structural heterogeneity.

Degree distribution: From Walker, et. al [11], we infer that higher variance of node degrees make the network more heterogeneous. So, we generate 3 networks each with average degree $k = 4$ and variance (σ_k^2) values 4.46, 8.76 and 30.35. Since we want to independently tune standard deviation and mean, we choose to sample the degrees from a gamma distribution. To discretize, we first sample degrees from a gamma distribution with mean $(k-1)$ and standard deviation σ_k, rounding to the nearest integer and adding 1. We verified numerically that this generated degree sequences with the desired properties. Once the degree sequence is formed, we form the connected network using the Havel-Hakimi algorithm [12], which greedily picks the node with the lowest degree d and connects to d vertices having the highest remaining degree.

Clustering-coefficient: The clustering coefficient ϕ is also known as the *global clustering coefficient* or *transitivity*[13]. If A is the adjacency matrix of the graph G, we can calculate $\phi(A)$ as follows:

$$\phi(A) = \frac{trace(A^3)}{||A^2|| - trace(A^2)}$$

To generate networks with different clustering coefficients, we utilized the Watts-Strogatz model that arranges the n nodes in a ring and connects each node u to the k nearest neighbors. Then, each edge is rewired with a probability p (for each edge (u, v) is replaced with (u, w) with a probability p, where w is a uniform random choice of existing node). We tune the clustering coefficient using this rewiring probability p. We note that as p increases, the clustering coefficient ϕ decreases, indicating increasing heterogeneity in the network structure (Refer to Figure 3). We chose to use p values 0.05, 0.1 and 0.5 for which we got corresponding ϕ values to be 0.49, 0.34 and 0.05.

4.2 Modeling the SISD Epidemic Spread with Mutation

We employ an SISD (Susceptible-Infected-Susceptible-Dead) model that incorporates viral mutation, allowing for the evolution of virus transmissibility over time:

- **State Initialization**: 95% of the nodes start in the susceptible (S) state, with 5% randomly chosen node set to infected (I_1) with Virus 1. No nodes are initially in the dead (D) state.
- **Virus Strain Characteristics and Transmission Parameters**: We define three viral strains (Virus 1, Virus 2, Virus 3), each with unique transmission probabilities $\beta_1 = 0.115$, $\beta_2 = 0.129$, and $\beta_3 = 0.144$, respectively. The mutation probability $\mu = 0.0005$ per time step governs the transition between viral strains, simulating evolutionary changes in transmissibility. These values were obtained from calibration on real-world datasets, which is further explained in the next section.
- **Transition Rules**:
  - A susceptible node (S) can become infected by Virus 1, Virus 2, or Virus 3, depending on the infection probability β_i of the virus i it contacts.
  - Mutations allow Virus 1 to evolve into Virus 2 and Virus 2 into Virus 3, with each mutation increasing transmissibility.
  - Nodes infected by Virus j (I_j) can be re-infected by any of the other Virus k ($I_k, k \neq j$). The latest infecting strain replaces the previous resident strain in the individual in case of co-infection.
  - All infected states (I_1), (I_2), (I_3) have a fixed probability $\gamma = 0.05$ of recovery back to S and a probability $\delta = 0.0005$ of transitioning to the D state.
- **Mutation Dynamics**: Each infected node has a mutation probability μ at each time step, dictating its likelihood of transitioning from Virus 1 to Virus 2, or from Virus 2 to

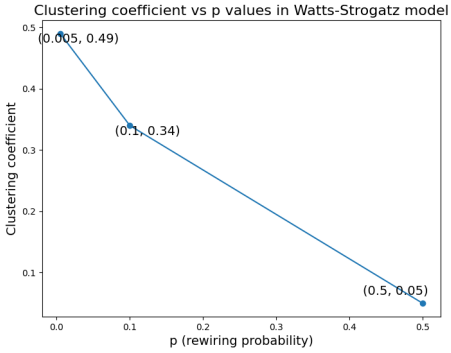

Figure 1: Variation of clustering coefficient with rewiring probability

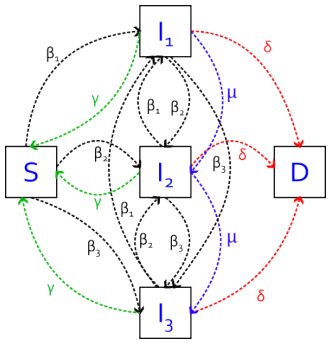

Figure 2: Modeling SISD dynamics

Virus 3. This mutation allows the pathogen to evolve and potentially increase its overall transmissibility.
- **Simulation**: The model is simulated for 1500 time steps, tracking the states (S, I_1, I_2, I_3, and D) of all nodes over time, and the results are averaged over 50 runs.

4.3 Calibration

In order to elucidate our choice of parameters, we calibrated our model on real-world data. We tried multiple network datasets (listed below) as well as different virues (listed below). Here, we summarize our observations from the same. For the same, we tried out different networks and time series data related to different diseases:

(1) For the disease infections data, we tried the following datasets:
   - **COVID-19 cases with geography[14]:** We obtained the weekly COVID cases till July 2024 with state level granularity. However, since our focus was to look at the interaction of the different strains of a virus, this dataset was not sufficient given that this did not give strain-specific information.
   - **Dengue phylogenics[15][16]:** We obtained the weekly dengue cases across the 50 US states between 2010 and 2013, and the phylogenetic tree for the various dengue strains that persisted during this period. However, due to lack of information to map the strain to the emergence of the case, we did not carry on with this dataset.
   - **Influenza strains [1]:** This dataset by WHO provides the statistics of the weekly number of tests out of which how many tested positive for influenza (along with information of which of the strains A or B caused it) and how many tested negative. We obtained the data for the whole of the US starting from Oct 2016 to May 2024.
(2) For the networks, we tried the following datasets:
   - **Synthetic population of Norway[17]:** This consists of a synthetic population of Norway ($\approx$ 4 million) with the demographic attributes, location of households, activity times etc. We tried constructing the contact network - however, we could not proceed due to a lack of description of each column, the relations with other columns. So, we tried to form the most intuitive connections by connecting individuals in the same household. Doing so made the graph very sparse. So, we include it in the future work to construct the network using a clustering algorithm like DBSCAN based on latitude, longitude and activity location with additional information obtained from the authors of the dataset.
   - **Oregon Router networks[18]:** This is a temporal network of routers comprising the Internet grouped into subgraphs called Autonomous Systems (AS). This dataset contains 9 snapshots of the network each taken once a week from March 31st 2001 to May 26th 2001. We chose to use the network with largest number of nodes ($\approx$ 11k nodes) and hence after analysis, we chose the network of May 26th and considered it as a static graph for our calibration. However, the loss value was extremely large and the optimization never converged. This is reasonable since we try to calibrate the parameters using human contact and infections data on a contact network of routers, for which the contact patterns will be vastly different. So, we next looked for human contact networks.
   - **High school contact network[2]:** This network represents contact pattern among students, teachers and staff in a high school in the USA. This was constructed to trace the flu spread due to close proximity interactions. This network perfectly fit our needs for the following reasons - it is a reasonably large network (788 nodes, 118k edges), it models influenza spread and in the US. So, this was the network that closely related to our time series data and thus, we calibrated on this dataset.

4.4 Simulating Virus Evolution and Outcome Measures

To assess how mutation and network heterogeneity affect viral competition and the spread of infection, we define the following outcome measures:
- **Effective** R_0: The R_0 of each of the viral strains is calculated at the time at which they first appear in the simulation. For the first strain, this is simply $\frac{\beta_1}{\delta+\gamma+\mu}$. For the second, we must also consider the outflow caused by nodes infected by strain 2 then being infected by strain 1: $\frac{\mu*I_1+\beta_2*(S+I_1)/N}{(\beta_1*I_1/N)+\delta+\gamma+\mu}$. Similarly, strain 3 also considers the outflows caused by strains 1 and 2. Finally, we adjust for the network heterogeneity by dividing R_0 by $\lambda_c = \frac{\langle k\rangle}{\langle k^2\rangle-\langle k\rangle}$, where $\langle k\rangle$ and $\langle k^2\rangle$ are the first and second moments of the network degrees, respectively.
- **Probability of fixation**: The proportion of runs in which all of the preceding strains of the disease have died out at the end of the simulation.
- **Epidemic size**: The proportion of all nodes not in the susceptible state by the end of the simulation.
- **Peak infection**: The highest value of $I_1+I_2+I_3$ throughout the simulation.
- **Peak time**: The time at which peak infection was reached.
- **Effective percolation threshold**: The value of $beta$ at which effective R_0 is equal to one. To find it, we simply divide each beta by the respective effective R_0.

5 Results

5.1 Homogeneous network

After running the SISD simulations on the homogeneous network with uniform degree distribution of $k = 4$ as seen in Figure 3, we see that increased transmissibility alone, with a constant rate of re-infection and fatality, provides selective advantage for the viral strains. The number of infections with virus 1 peak and reduce around the time the first mutation enters the contact network and starts spreading, a similar pattern is observed when virus 3 arises and increases its spread rate, thus reducing the number of infections with virus 2. The number of deceased individuals seems to increase at a constant rate.

The probability of fixation of strain 3 in these runs is 44%. This shows that although not deterministic, strain 3 being the most

competitive viral strain drives the other two to extinction. The selection exponent for strain 3 in this homogeneous network is 2.598. The overall percolation threshold with all the three viral strains is similar to that of virus 2 and 3 alone. This ties into our experimental design of beta ratios that allow strains 2 and 3 to be greater than the critical range, with the beta for strain 1 low enough not to ensure deterministic fixation.

Because individuals with earlier strains can contract the mutated versions of the disease, R_0 values vary depending on how many strains are present. With only strain 1, its R_0 is equal to 1.597. When strain 2 appears, it has an R_0 of 1.796, and strain 1's goes down to 0.57; with all three strains, the R_0 of strains 1, 2, and 3 are 0.333, 0.600, and 2.000 respectively, highlighting the impact of more contagious strains on the transmissibility of their competitors.

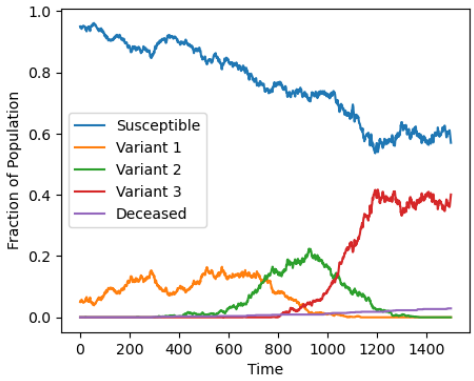

Figure 3: An example run that represents the proportion of individuals susceptible to all three strains, infected with each of the three strains, and deceased from infection

5.2 Heterogeneous Networks

As described previously, we considered two dimensions of heterogeneity: degree variance and clustering coefficient. We ran the simulation on each of the heterogeneous networks we generated, and observed the outcomes for each of the metrics defined above.

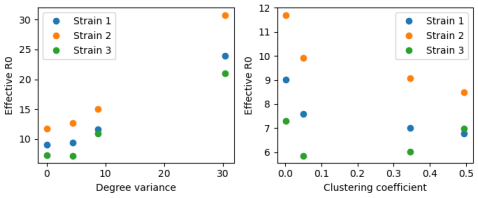

Figure 4: Change in effective R_0 with heterogeneity

Effective R_0 is perhaps the metric that was most evidently affected by both forms of heterogeneity. There are clear trends indicating a positive correlation between degree variance and effective R_0, suggesting that this variance causes the emergence of potential superspreaders with many connections, increasing transmissibility potential. Clustering, coefficient, on the other hand, seems to have a negative correlation with effective R_0; this makes sense when considering more clustered networks are more likely to have bridges that cut off sections of the network when disconnected, which would be the case when a node enters the D state and becomes unable to spread the disease.

Notably, strain 2 showed higher effective R_0 even though strain 3 has a higher transmissibility. This is likely due to the fact that R_0 was calculated at the moment each new strain first appeared. Strain 3 has a lower susceptible population to infect once it appears than earlier strains, and it must also compete with its predecessors; meanwhile, strain 2 only competes with strain 1 at first and has a higher available population to infect.

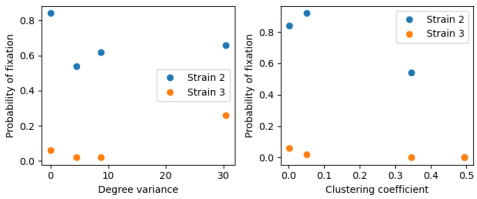

Figure 5: Change in probability of fixation with heterogeneity

Degree variance did not appear to influence the probability of fixation of different strains in any meaningful way. Clustering coefficient did seem to have a stronger effect, particularly with the highest values: with our highest clustering coefficient, none of the strains achieved fixation, meaning that none of the previous strains fully died out. This is in part because the evolution of the disease was much slower in highly clustered networks, but also because, as discussed previously, clustering can lead to isolated regions of the network. If pools of different strains are cut of from each other, this essentially ends competition between them and instantiates smaller networks with each one that then develop independently. It is therefore evident that clustering promotes coexistence between strains, and reduces the probability of one taking over the rest through winner-takes-all dynamics.

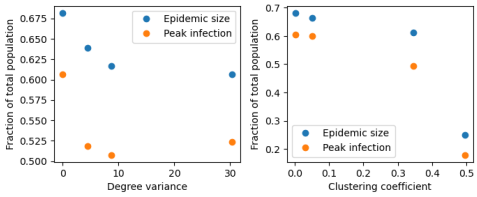

Figure 6: Change in epidemic size and peak infection with heterogeneity

Epidemic size and peak infection show a very tight correlation, which makes sense. With constant mortality and recovery rates across diseases and simulations, it is reasonable to expect such a relationship. They both seem to vary inversely with both heterogeneity measures, weakly with degree variance and somewhat more strongly with clustering coefficient. The latter is coherent with the other results we have observed thus far; high clustering

coefficients create more isolation between segments of the network, increasing the likelihood of the disease being cut off from potential hosts. This lowers the potential reach of the epidemic, as well as its expected peak.

The fact that degree variance shows the opposite relationship with epidemic size and peak infection that it does with effective R_0 could stem from several factors. It might point towards a weakness of the effective R_0 calculation, which, at least for the first strain, is theoretical and deterministic given initial conditions. This might be decoupled from empirical results under certain scenarios, though the discrepancy here is too weak to make definitive judgements. Another possible explanation is simply that the resolution of our experiment is to low. Were we to have had more time to run additional simulations on networks with different degree variances, we might have observed a different relationship between it and epidemic size and peak infection. These are all potential avenues to explore in further research.

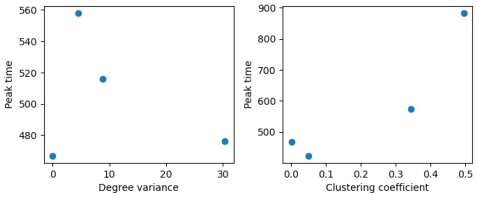

Figure 7: Change in peak time with heterogeneity

Peak time did not appear to exhibit a meaningful relationship with degree variance. However, an increase in clustering coefficient did delay the onset of the epidemic's peak; this can be attributed to networks with a higher clustering coefficient having effective R_0 values which are lower, but still greater than one. The result is a smaller-scale outbreak that nonetheless doesn't quickly die out.

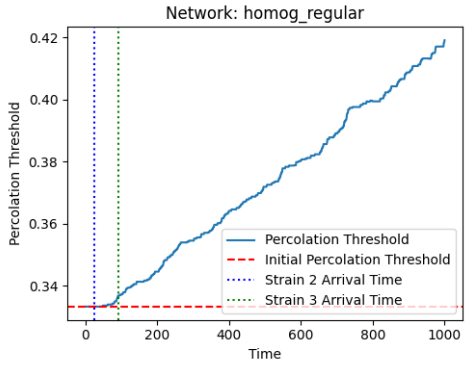

Figure 8: Change in percolation threshold as disease spread introduces heterogeneity within the homogeneous lattice network model through mortality

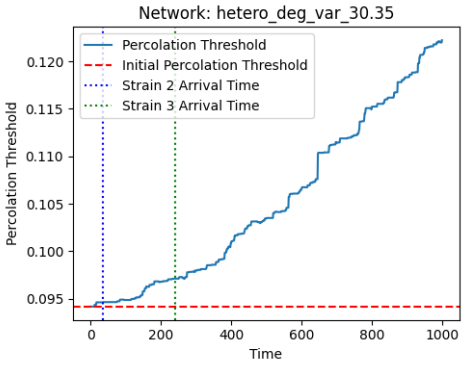

Figure 9: Change in percolation threshold as disease spread increases heterogeneity within the heterogeneous (degree variance of 30.35) network through mortality

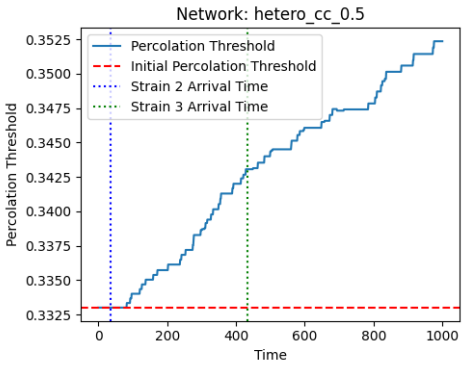

Figure 10: Change in percolation threshold as disease spread increases heterogeneity within the heterogeneous network (clustering coefficient of 0.75) through mortality

Figures 8 show us that mortality-induced heterogeneity in homogeneous lattice networks, and similar exacerbation of heterogeneity induced by mortality in heterogeneous models in figure 9 and 10, show that as a high mortality disease evolves and progresses through the network, more substructures are created, thereby influencing disease spread. The mortality-exacerbated heterogeneity induces a slowing down of disease spread in highly clustered networks (Figuee 10) in the case of removed bridge nodes that can result in network fragmentaion and prove to be protective of subsections of the populations that do not interact with the rest of the individuals.

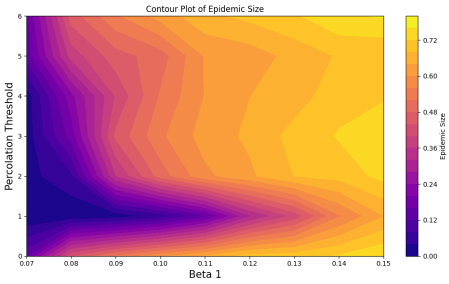

Figure 11: Contour plot depicting changes in the epidemic size with percolation threshold and transmissibility of the first strain

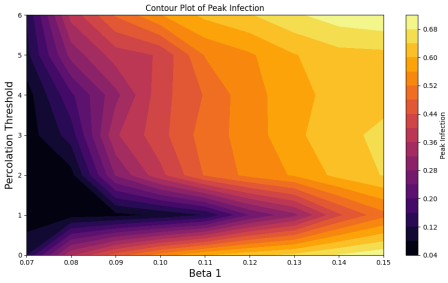

Figure 12: Contour plot depicting changes in the peak infection with percolation threshold and transmissibility of the first strain

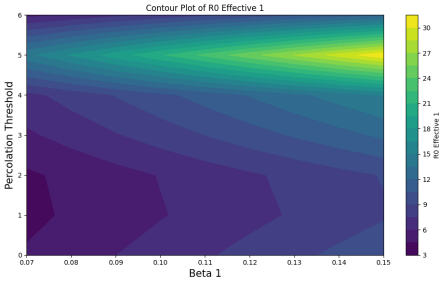

Figure 13: Contour plot depicting changes in the effective reproductive number with percolation threshold and transmissibility of the first strain

Within our operating parameter space (based on ODE calibration with a real-world U.S. high school dataset), we notice that increasing transmissibilities for a given percolation threshold increases the epidemic potential (Figure 11) and peak infection (Figure 12). In addition, beyond a critical value of the percolation threshold, a highly transmissible virus has high epidemic potential even when the percolation threshold is low (Figure 13). This means that the network heterogeneity does not interfere with theoretical predictions in cases of highly transmissible viruses.

6 Conclusions

In conclusion, our study demonstrates that the heterogeneity of human contact patterns and network topology play a pivotal role in shaping the spread and evolution of viral pathogens. Specifically, we found that variability in contact patterns can significantly slow down both viral evolution and the spread of disease. This finding underscores the importance of network fragmentation as a barrier to rapid transmission, ultimately impeding viral adaptation and reducing epidemic size.

Our analysis also shows that networks with high variance in degree distributions tend to exhibit earlier infection peaks and reduced epidemic potential, suggesting that such network structures may facilitate more rapid but contained outbreaks. Conversely, networks with high clustering coefficients tend to experience delayed peaks and larger epidemic sizes, emphasizing the role of tightly-knit communities in sustaining viral transmission over extended periods.

Moreover, our findings regarding the percolation threshold reinforce the importance of network connectivity in determining epidemic magnitude. Lower percolation thresholds correlate with larger epidemic sizes and higher infection peaks, particularly when transmissibility is high, highlighting the conditions under which widespread outbreaks are most likely to occur. We also see that concurrently with theory, highly transmissible viruses have very high epidemic potential, despite network topology.

Finally, we identified a dynamic feedback loop between viral evolution and network heterogeneity, driven by patient mortality. Increased virulence results in higher mortality, which in turn amplifies network heterogeneity by reducing the number of active nodes and fragmenting the network. This interplay complicates efforts to control disease spread and underscores the need for strategies that address both the biological and structural factors influencing epidemic dynamics. Without higher resolution for inferring human interaction patterns, inferred disease parameters tend to be inflated.

Our project provides valuable insights for public health strategies aimed at containing outbreaks, particularly in the context of rapidly mutating pathogens. Interventions that reduce network connectivity, limit clustering, or target key nodes within heterogeneous networks could be highly effective in mitigating the spread of infectious diseases in a globally connected world.

7 Scope for Future work

(1) **Identifying Policy Implications and Control Strategies Based on Network Structure** Translating the findings into actionable strategies for public health policy is a key goal of this research:
- Surveillance to Estimate Lead Time: Surveillance of randomly selected nodes within the network to calculate lead time, the period before a potential outbreak where warning signs may be detectable.
- Estimating Notification Time: Incorporate mortality data with infection surveillance of the same random nodes to estimate the optimal notification time, when public health authorities should issue an alert based on early signals.
- Determining Optimal Announcement Timing (t_critical): Calculate the latest time, t_critical, to release a public

safety announcement to reduce transmission rates effectively. Such an announcement would reduce the contact rates β enough to lower the effective reproduction number of all strains below 1, aiming for a 70% reduction in the infected proportion of the population by encouraging preventive behaviors.

Acknowledgments

We thank Dr. Aditya Prakash and Harshavardhan Kamarthi for their helpful discussions and valuable feedback.

References

[1] Global Influenza Programme. Influenza laboratory surveillance information. URL https://worldhealthorg.shinyapps.io/flunetchart/.

[2] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W. Feldman, and James H. Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010. doi: 10.1073/pnas.1009094108. URL https://www.pnas.org/doi/abs/10.1073/pnas.1009094108.

[3] S. O. Lloyd-Smith, S. J. Shcreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066): 355–359, 11 2005. ISSN 1476-4687. doi: 10.1038/nature04153. URL https://doi.org/10.1038/nature04153.

[4] Caroline Buckee, Leon Danon, and Sunetra Gupta. Host community structure and the maintenance of pathogen diversity. *Proceedings of the Royal Society B: Biological Sciences*, 274(1619):1715–1721, 2007. doi: 10.1098/rspb.2007.0415. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2007.0415.

[5] M. Boots and A. Sasaki. 'small worlds' and the evolution of virulence: infection occurs locally and at a distance. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1432):1933–1938, 10 1999. doi: 10.1098/rspb.1999.0869. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1999.0869.

[6] Sten Rüdiger, Anton Plietzsch, Francesc Sagués, Igor M. Sokolov, and Jürgen Kurths. Epidemics with mutating infectivity on small-world networks. *Scientific Reports*, 10(1):5919, Apr 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-62597-5. URL https://doi.org/10.1038/s41598-020-62597-5.

[7] Gabriel E. Leventhal, Alison L. Hill, Martin A. Nowak, and Sebastian Bonhoeffer. Evolution and emergence of infectious diseases in theoretical and real-world networks. *Nature Communications*, 6(1):6101, 01 2015. doi: 10.1038/ncomms7101. URL https://doi.org/10.1038/ncomms7101.

[8] Abdennour Sebbagh and Sihem Kechida. Ekf-sird model algorithm for predicting the coronavirus (covid-19) spreading dynamics. *Scientific Reports*, 12(13415), 12 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-16496-6. URL https://doi.org/10.1038/s41598-022-16496-6.

[9] Safar Vafadar, Maryam Shahdoust, Ata Kalirad, Pooya Zakeri, and Mehdi Sadeghi. Competitive exclusion during co-infection as a strategy to prevent the spread of a virus: A computational perspective. *PLOS ONE*, 16(2):1–18, 02 2021. doi: 10.1371/journal.pone.0247200. URL https://doi.org/10.1371/journal.pone.0247200.

[10] Garrett Hardin. The competitive exclusion principle. *Science*, 131(3409):1292–1297, 1960. doi: 10.1126/science.131.3409.1292. URL https://www.science.org/doi/abs/10.1126/science.131.3409.1292.

[11] Sara Imari Walker and Cole Mathis. *Network Theory in Prebiotic Evolution*, pages 263–291. Springer International Publishing, Cham, 2018. ISBN 978-3-319-93584-3. doi: 10.1007/978-3-319-93584-3_10. URL https://doi.org/10.1007/978-3-319-93584-3_10.

[12] S. L. Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph. i. *Journal of the Society for Industrial and Applied Mathematics*, 10 (3):496–506, 1962. doi: 10.1137/0110037. URL https://doi.org/10.1137/0110037.

[13] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 6 1998. ISSN 1476-4687. doi: 10.1038/30918. URL https://doi.org/10.1038/30918.

[14] CDC-INFO. Covid-19 case surveillance public use data with geography, 02 2021. URL https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4/about_data.

[15] GenBank. Real-time tracking of dengue virus evolution. URL https://nextstrain.org/dengue/all/genome.

[16] CDC. Dengue-historical data (2010-2013). URL https://www.cdc.gov/dengue/data-research/facts-stats/historic-data.html.

[17] Abhijin Adiga, Hannah Baek, Stephen Eubank, Przemyslaw Porebski, Madhav Marathe, Henning Mortveit, Samarth Swarup, Mandy Wilson, and Dawen Xie. Synthetic population for nor, April 2022. URL https://doi.org/10.5281/zenodo.6503475.

[18] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, page 177–187, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 159593135X. doi: 10.1145/1081870.1081893. URL https://doi.org/10.1145/1081870.1081893.