

Outbreak Diversity Through Contact Networks!

By Arjun Peroor and Pranav Kaliaperumal

Background

Humans are very social creatures. Every day, you see humans interact with at least one other human. It's a way of life for us. The more people we talk to, the more connections we build, resulting in a happier life. However, one unfortunate downside to interactions is the spread of diseases. It's sad and true, and with what happened recently with the novel Coronavirus in 2019, it is important to bring this up. After that entire experience with COVID-19 in 2020, I would imagine people would like to be more prepared for any potential future diseases that may come underway. Knowing how quickly a disease can spread or how dangerous it is will be extremely useful for the public. One metric is the reproductive number(R_0) to give a good sense of how quickly a disease can spread. Unfortunately, it is not the best metric to go off because it treats everyone the same and assumes that all people have the same number of connections but in reality, we know that everyone is different. Some people know more people than others. For context, an R_0 greater than 1 normally means that an outbreak is likely, and if it is less than 1 it means that it will die out fast. For example, if you measure the reproductive number in a hospital, it will be high because there are a lot of people in close contact and if we apply it to the general population it may seem like the disease is more contagious than it is. So it's important to understand how the disease spreads and how it varies from person to person and we can see that through contact networks.

Introduction

For this paper, we are reproducing the results of another paper called Network

Theory and SARS: Predicting Outbreak Diversity written by Meyers, Pourbohloul, Newman, Skowronski, and Brunham which focused on 3 contact networks and simulated data for the city of Vancouver, British Columbia. They were simulating data about Severe Acute Respiratory Syndrome(SARS), a disease that originated in China in 2002 and spread to parts of Canada in 2003. So this begs the question: How do the different types of contact networks affect the spread of diseases in general?

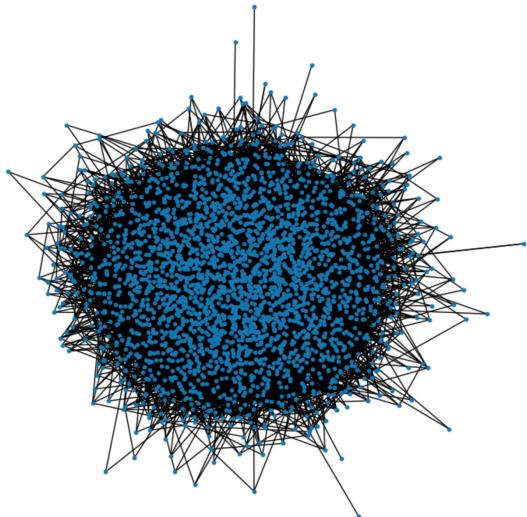
Methods and Results

To reproduce the results of the paper, we used Python and Networkx to simulate the data in Vancouver, British Columbia. To reproduce the results from the paper, we needed to make some assumptions along with what it provided us. What the paper provided were common places where people meet. The places where households, schools, hospitals, workplaces, and other public places. Connection probabilities were given for each setting. Households were 1 which makes sense because people in houses are close to one another so it is guaranteed that they will make close contact with one another. For schools and hospitals, it was given to be 0.3 for probability connections. Workplaces were at 0.03 and other public places were set at 0.003. They also gave us information that there were 1000 households which yielded about 2600 people. In addition to the information they gave us, we had to make a few assumptions. We made it so that households varied from 1-5 people and we assigned a probability of that household having that many people for each number from 1-5. We also assigned random ages to

people and gave each person an ID so they could be identified and then we assigned them to each setting based on age. For school people people were of the age 5 to 21. For work, people were 22 to 65, and hospital people were less than equal to 4 and greater than 65 years old. And then we assigned any random age to other public places. Households also did not have a set age bracket because people of any age are seen in houses. For every 30 students, in school, an adult(teacher) was assigned. For hospital wards, for every ward, which had 20 people in it, 30% of the people were set as Nurses, and 20% were set as Doctors. Then we had every person as a node and we graphed it in 3 different Networks. For the Poisson network, we used Erdos Reyi Graph which was a function of k_x . For Power Law, we used the Barabasi Albert graph which was also a k_x function and for the Urban graph, we just graphed 'G' normally.

Let's take a look at the results:

Poisson Network Structure

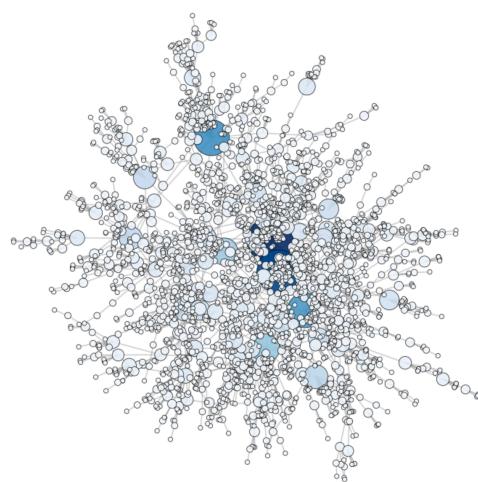


This is the Poisson distribution, which has a homogeneous structure. Everything is clustered in the middle except for a few

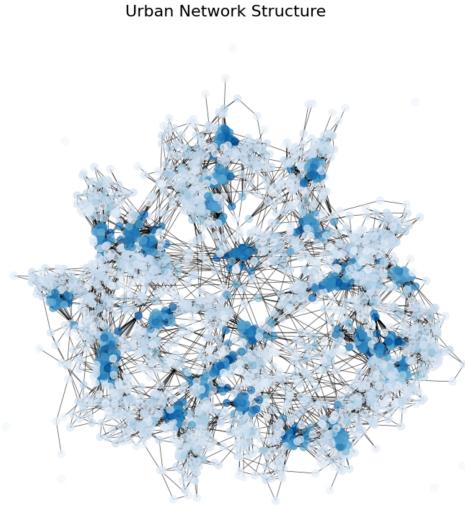
outliers. This network would work well for the basic reproductive number because it is homogeneous and assumes that all people have a similar number of connections. However, we know that it is not indicative of how it is in real life.

The next network we'll take a look at is the Power Law.

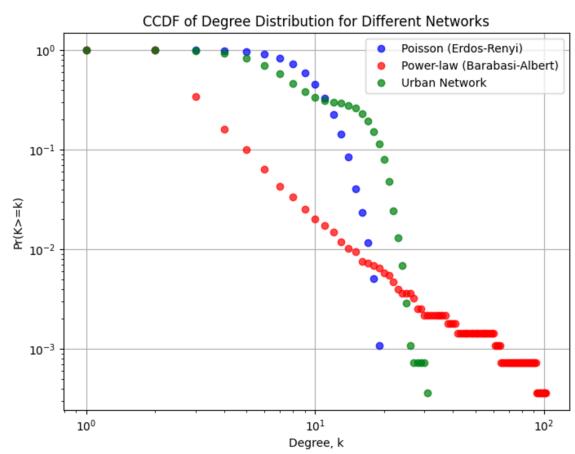
Power-Law Network



This network is much different than the Poisson one as it is less homogeneous and has more hubs present. The larger the node the darker it is that it has more degrees/connections. However, outside of those hubs of superspreaders, the rest of the nodes are shown to have very few connections. The Power Law network shows many people with one or two degrees if they are not part of a hub. That also is not very representative of the real world. Let's take a look at the Urban Network now:



This network has clusters and hubs present but also shows that the nodes that are not part of the clusters have a decent amount of connections rather than just one or two. This network displays more of a balance. Not everyone has the same amount of connections but not everyone who isn't a superspreaders only knows one or two people. This shows how important looking at different contact networks is because the Poisson Network will have a disease spread very quickly while the Power Law will have that same disease die out rather quickly. It is important to strike a balance in these networks and we are hopeful the Urban Network does just that.



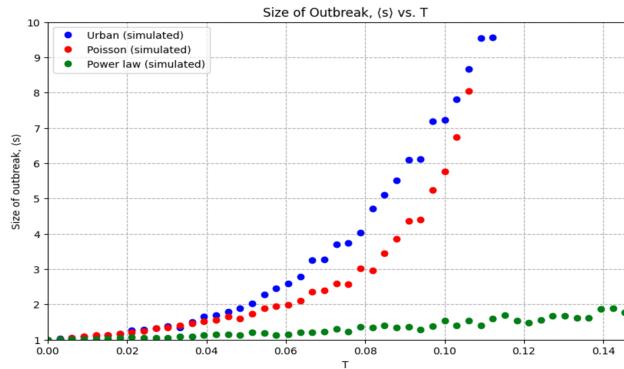
Looking at the degree distributions the one that sticks out is the Power law because of its heavy tail which results from the nodes that aren't part of hubs having very few connections as stated above. As the degree increases, the probability of having higher connections decreases for all the networks, hence explaining the decay for all of them.

As stated earlier in the paper, people would love to know how deadly a disease can turn out to be and understand the limits of the reproductive number, we can look at a different metric. Transmissibility. The equation given to us in the paper is:

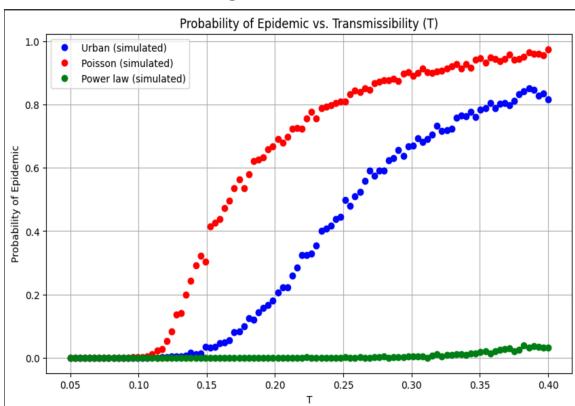
$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle},$$

(Meyers, Pourbohloul, Newman, Skowronski, Brunham, 2004)

Transmissibility looks at the average probability that an infectious individual will spread to a new person, rather than looking at the total number of new cases. As a result, this was used in the SARS paper to predict how quickly and how much a disease spreads. They had 2 graphs. In the first graph, they looked at a small outbreak's size given modest Transmissibility numbers. Then they had another graph that looked at the probability that an epidemic would start given higher numbers of Transmissibility. So we attempted to reproduce these results using a SIR model to calculate the outbreak size where we assigned a random person to be infected and after it infected a neighbor, we iterated through those neighbors of the infected ones. The simulation looked at the average outbreak size of every situation:

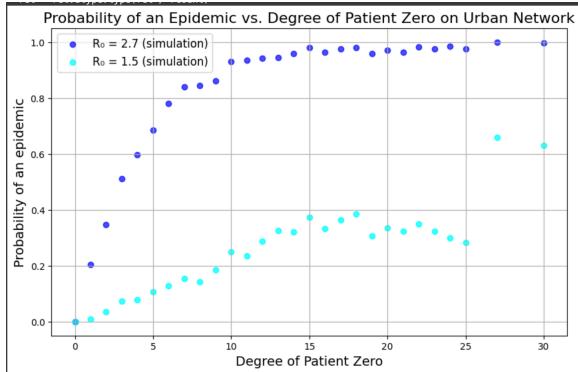


We expected the Poisson Network to be the highest of them all but with the sample size being small we weren't shocked with the Urban Network being ahead initially because it is possible that some parts of the urban ones could have been highly connected but as you can see as the values of T increase, Poisson's slope gets steeper and they intersect at $T=0.1$. The Power Law network is more susceptible to killing any chances of an outbreak because the connections between the hubs are so thin so once a hub gets infected, there's not a great chance it spreads anywhere else and it'll die out. Then we increased the Transmissibility range to 0.4 and observed the trends. We once again did a SIR simulation that followed a similar process but this time just returned the probability of an epidemic rather than average size and we defined an epidemic as if more than 5% of the population gets infected.



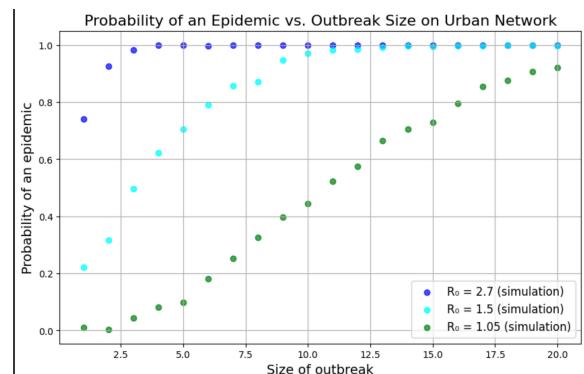
The graph above confirms our beliefs from earlier that the Poisson Network would have the highest chance of starting an epidemic because it is so homogenous and clusters everyone together. The Power Law Network showed similar results to the previous graph when T values were lower as the chance of an Epidemic is much lower because of the abundance of low-degree nodes. Urban is seen to be in between the 2 other networks illustrating the balance it has in terms of having superspreaders and low degree nodes. The graph above illustrates how differently each contact network produces the results of the same outbreak.

Then, we did a deep dive on the Urban Network to see how much the initial stages of a disease can impact its epidemiology. The paper spoke on how for most communities, the introduction of one case may or may not be enough to spark an epidemic but we can question which factors make either outcome favorable of whether an epidemic is likely to start or not. To find out we ran a simulation on the Urban Network we formed to predict the likelihood of an epidemic given the initial degree of patient 0. This time, we used the basic reproductive number because we wanted to observe how two identical communities can experience different outbreaks if the initial contact patterns differ from one another. So we decided to plot two sets of plots where we set one of the reproductive numbers to 2.7 and the other to 1.5 and observed the probability of an epidemic occurring given the degree of the first patient.



It selects a random node and we simulate the spread of the infection and once again set the threshold to where at least 5% of the population has to be infected in order it for be to considered an epidemic. As you can see if the person has a lot of connections, there is a higher chance of starting an epidemic. Despite the limitations of the Basic Reproductive Number, it is still helpful in terms of allowing us to see how important the initial stages of spreading are. If a person spreads, on average, the disease to an average of 2.7 in a susceptible population, then the chances of an outbreak will certainly be much higher for those who spread 1.5. This highlights that even in the same setting, the results can vary based on someone's spreading potential. In addition to studying the results of the initial degree of a random person, we also looked at how the initial outbreak affects epidemic probability. The paper looked at when the disease is near or at the epidemic threshold compared to diseases that are well above the

threshold.



Once again we assumed that an epidemic would take place when at least 5% of the population is infected. The simulation was similar to the previous except this time it was based on the initial size of the outbreak. Intuitively the results show that the more initial cases there are, there's higher the chance of an epidemic and as stated in the paper, the threat of an outbreak is overwhelming for diseases much higher than the epidemic threshold and the difference is noticeable especially early.

Discussion

We chose to work on this because diseases have been a large problem to the general population for all of human history and we have finally reached a point where we have sufficient technology to analyze, predict, and control these disease outbreaks before they decimate the population. Whether it's a deadly disease like the plague, or a less severe disease like the flu, being able to influence the spread of diseases would be revolutionary to everyone going forward. This specific study on SARS was a great look at different factors that affect an outbreak. These observations can be extrapolated to larger communities and even other diseases in the future. There would be many adjustments that would

have to be made to do this, especially for other diseases, but the overall methods of prevention and overall trends should stay somewhat similar.

The importance of outbreak prevention was highlighted recently with the COVID outbreak. There were many scientists and disease prevention experts with varying opinions over the course of the pandemic but ultimately the strategies that were implemented were not very effective. We think the main reason for this was that a lot of the COVID decision-making was a compromise between statistically advised decisions and what people would perceive as a good decision since these two were often the direct opposite of each other. Thus, this study and these discoveries won't be the cure to all future outbreaks, both due to the actual limitations of the calculations and due to how people in power would generally never make a decision based purely on statistics. Despite this, having this information for more major diseases on a national and even global scale would be a huge improvement to the overall health of the world.

In our project, we tested the three different contact networks mentioned earlier in order to see which one was the best to represent these outbreaks. The Poisson network is useful for modeling and analyzing theoretical disease outbreaks because it captures the random, independent nature of interactions between individuals in a population but it is not the best for real world applications as it tends to be an overestimate for the spread of a disease. On the other hand, due to the way the Power-Law Network works it tends to be a drastic underestimate of the spread of a disease. Since it works with hubs, where

one person is spreading it to many other people, it can be very inaccurate for predicting the spread of a disease since if that hub is cut off from the disease it automatically assumes everyone connected to that hub can't get the disease. This isn't accurate as it is not a realistic representation of people's social interaction since most people socialize with a variety of different groups on a daily basis. In terms of disease prevention, the Poisson would be more useful in general since it helps more to have an overestimate than an underestimate of a disease since there is less harm in taking too many safety measures than in taking not enough precautions. However, overall the Urban Network was the best for predicting the outbreaks. Since it takes both social and spacial interactions into account, it gives the most accurate prediction of the spread of an outbreak. Thus, using the Urban Network on a larger scale could drastically help reduce the spread of future outbreaks.

Some of the challenges we faced were making sure that all of our contact networks were realistic and would give us useful data. Since all of our data was simulated, we had to make sure that it wasn't unrealistic and that it was reasonable enough to make relevant conclusions from. Additionally, the Poisson network was also hard to read because of how many data points were on it so it was difficult to know if our simulation recreations were accurate when we first began the project.

Bibliography

L.A. Meyers et al., **Network theory and SARS: predicting outbreak diversity.** *J. Theoretical Biology* **232**(1), 71-81,
2025,<https://www.sciencedirect.com/science/article/pii/S0022519304003510>

Paul L. Delamater, Erica J. Street, Timothy F. Leslie, Y. Tony Yang, Kathryn H. Jacobsen,
Complexity of the Basic Reproduction Number (R_0), Volume 25, Number 1, 2019,
https://wwwnc.cdc.gov/eid/article/25/1/17-1901_article

Clauset, Aaron, Lecture Notes 7, Epidemiology I: models and networks, 2024,
https://aaronclauset.github.io/courses/3352/csci3352_2024_L7.pdf

Clauset, Aaron, Lecture Notes 8, Epidemiology II: models and networks 2024,
https://aaronclauset.github.io/courses/3352/csci3352_2024_L8.pdf