

February
2021

Epidemic Spreading Analysis Using Social Networks

DIGITAL MEDIA AND SOCIAL
NETWORKS

ECS637U/ECS757P

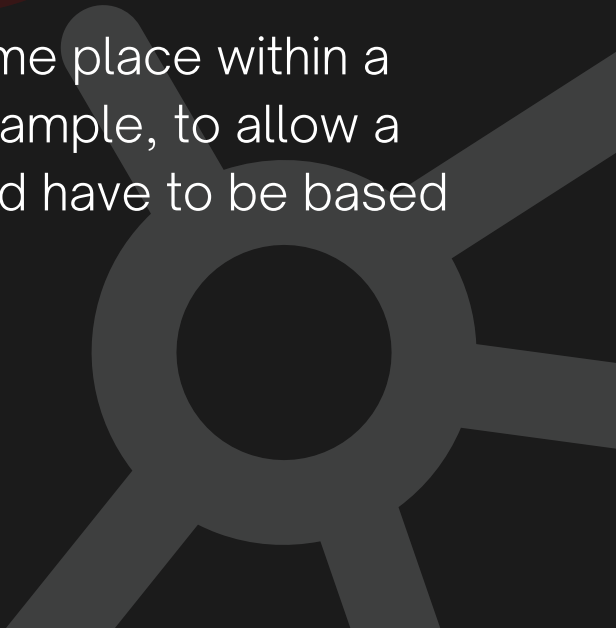


Context

Assessments of epidemic spreading in a population usually focus on large scale analyses, e.g. movements of people between cities or countries. However, it is also interesting to elucidate how spreading works at a small scale (i.e. within cities or communities). To combat an epidemics, it is critical to take into account the structure of the social network and how it supports or slows down spreading, and to model the effects of local changes in the network (such as lockdown).

This project will rely on subsets of datasets available from the social network app Gowalla. Gowalla was a location-based social networking service. The first large-scale dataset is a friendship network, where each node is a user and each edge denotes friendship. The second large-scale dataset is a record of 6,442,890 checks-ins (including geographical coordinates) of these users over the period of Feb. 2009 - Oct. 2010. The project will be based on excerpts of these datasets, on a much smaller scale.

By combining these two dataset sources, one can simulate a hypothetical epidemic spreading between friends who meet in a same place within a given time window, and see how the network structure influence this spreading. Then, specific measures can be proposed: for example, to allow a maximum travel distance within the city, or a maximum of friends one could visit. The analysis and subsequent measures proposed have to be based on network key features such as eigenvector centrality, node removal analysis, or other methods covered in the module.

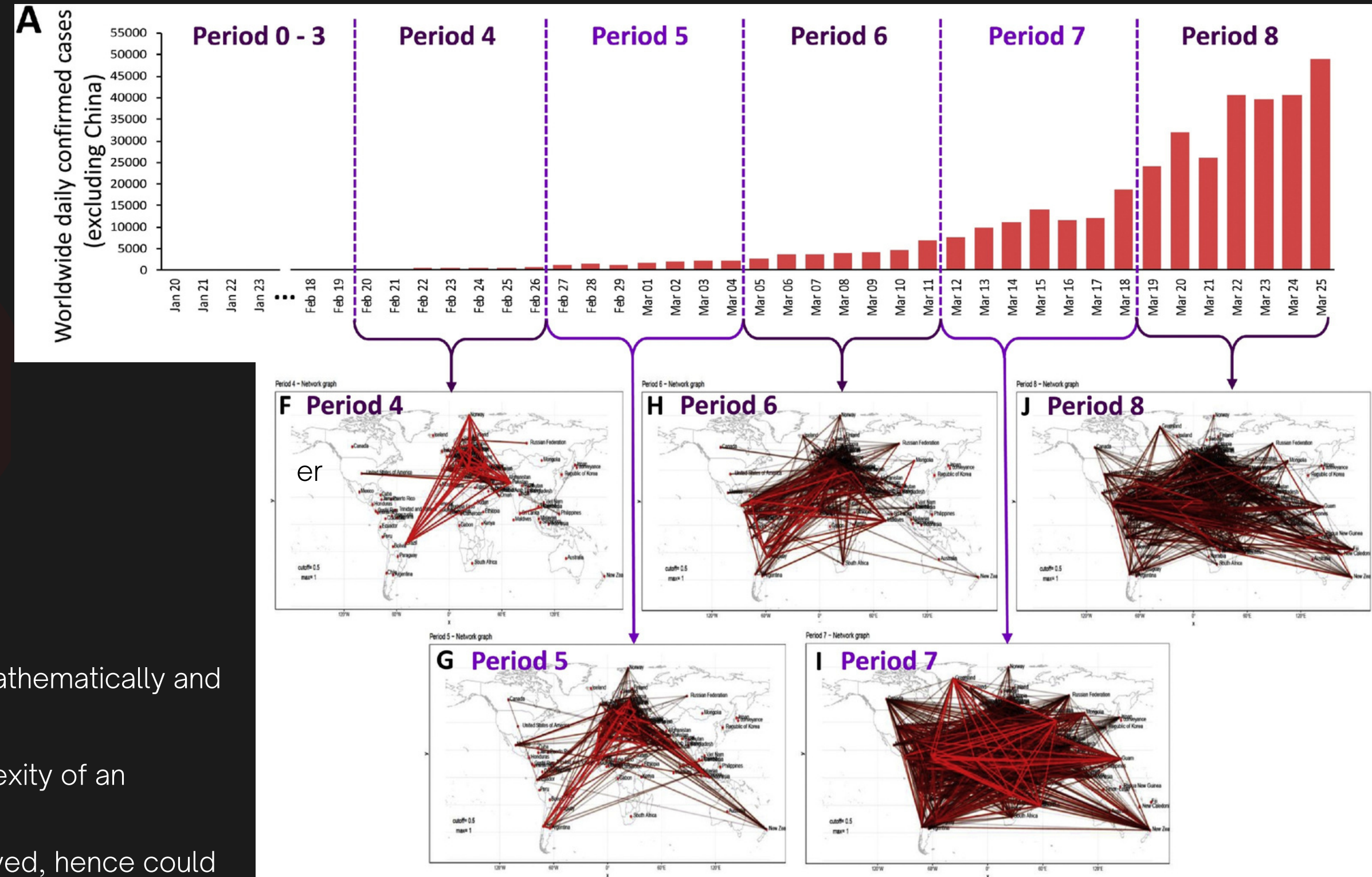


Motivation

- Growing attention in estimating epidemic spread after Coronavirus outbreak.
- Graphs are extremely flexible for representing such complex processes. They capture interactions among the communities.
- Network centrality measures are effective in predicting the spread of diseases.
- Study the network to identify the hotspots and communities impacting them to define approaches to control the spread.

Challenges

- Modelling transmission through networks is mathematically and computationally challenging.
- Difficult to understand the probabilistic complexity of an epidemic spread.
- A small part of the network is generally observed, hence could be misleading.
- The coarseness of the data being gathered is determined by the geographical locations. As a result, anyone suitable consistent approach not available



Ref: <https://www.sciencedirect.com/science/article/pii/S1201971220303179>

ABOUT THE DATASET

DATASET 1: FRIENDSHIP NETWORK

Friendship network defines the relationship (edges) between people (nodes) identifying various communities. This data would be used for targeted community based interventions.

- Identify the Giant component in the network.
- Undirected graph created with nodes as people and edges as their relationship. No relationship strength considered
- Partition the data with respect to modularity for overall max modularity value.

Nodes :	10001
Edges:	140869
Closeness Centrality :	2.897
Betweenness Centrality :	5
Modularity:	0.480

DATA PRE-PROCESSING

STATISTICS

DATASET 2: CHECKINS DATASET

Checkins dataset provides an overview of the location where people usually interact. This will be used to identify targeted hotspot-based interventions.

- Two datasets have been sampled:
 - Checkins data for the whole of September by location
 - Checkins data for a September evening to track interactions within a span of 4 hours.
- Two graphs created:
 - Directed Graph with nodes representing people pointing outwards to nodes representing location for September. Edge weight is the frequency of the checkins per location per person.
 - Undirected Graph, with nodes, representing people and locations, for checkins over an evening in 4 hours period

Nodes:	112501
Edges:	138552
Average Degree:	2.463
Network Diameter:	22
Modularity:	0.481

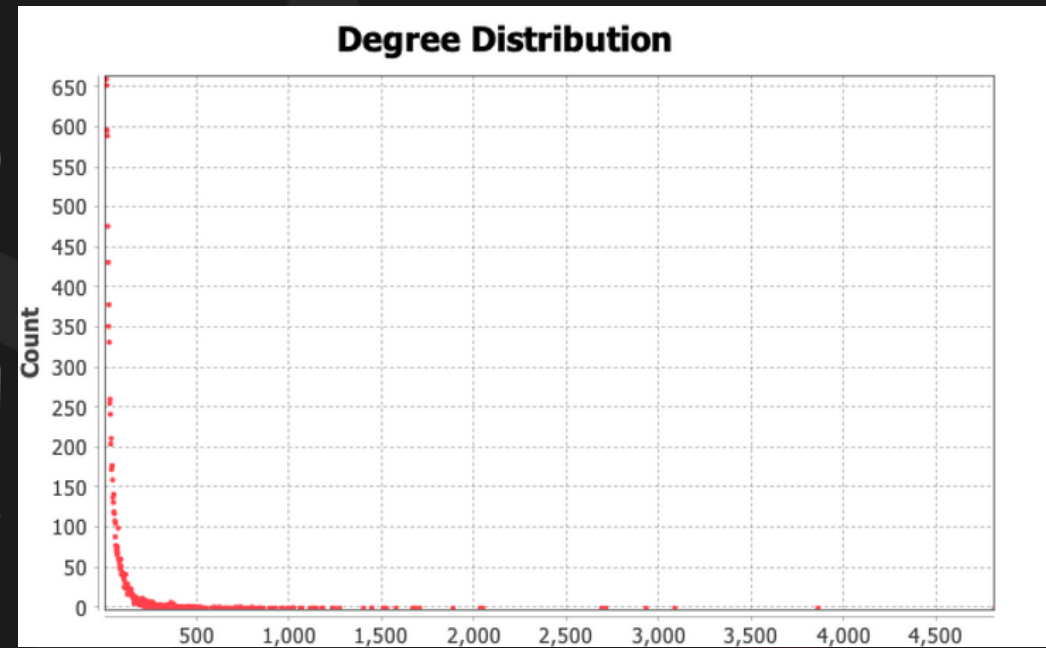
September Checkins

Nodes:	1970
Edges:	1440
Average Degree:	1.462
Network Diameter:	8
Modularity:	0.989

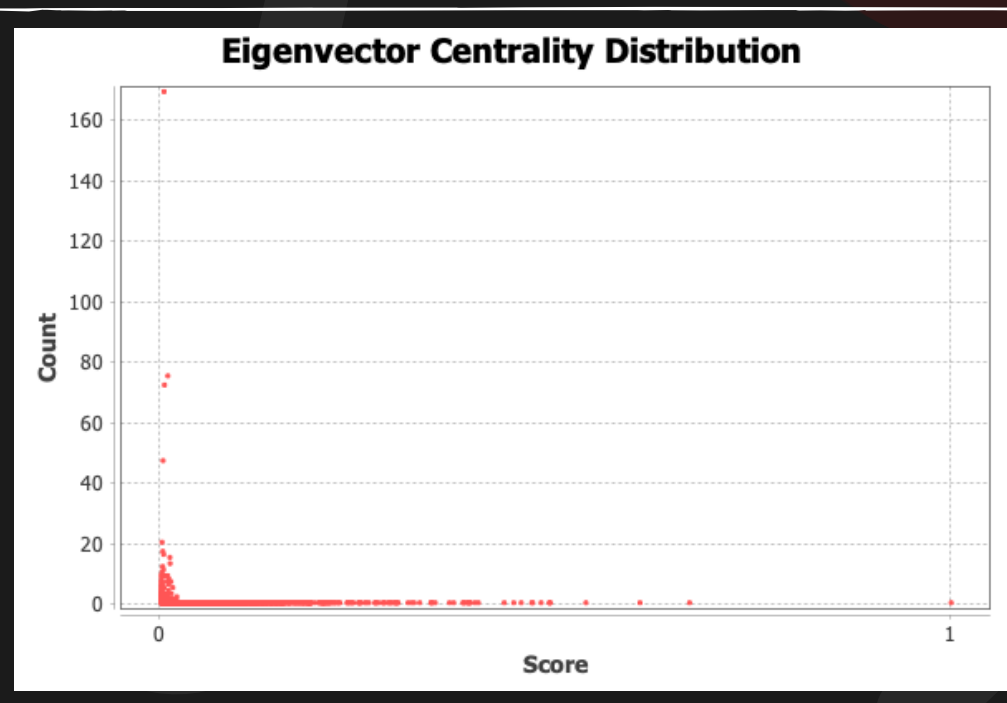
Evening Checkins

NETWORK STATISTICS

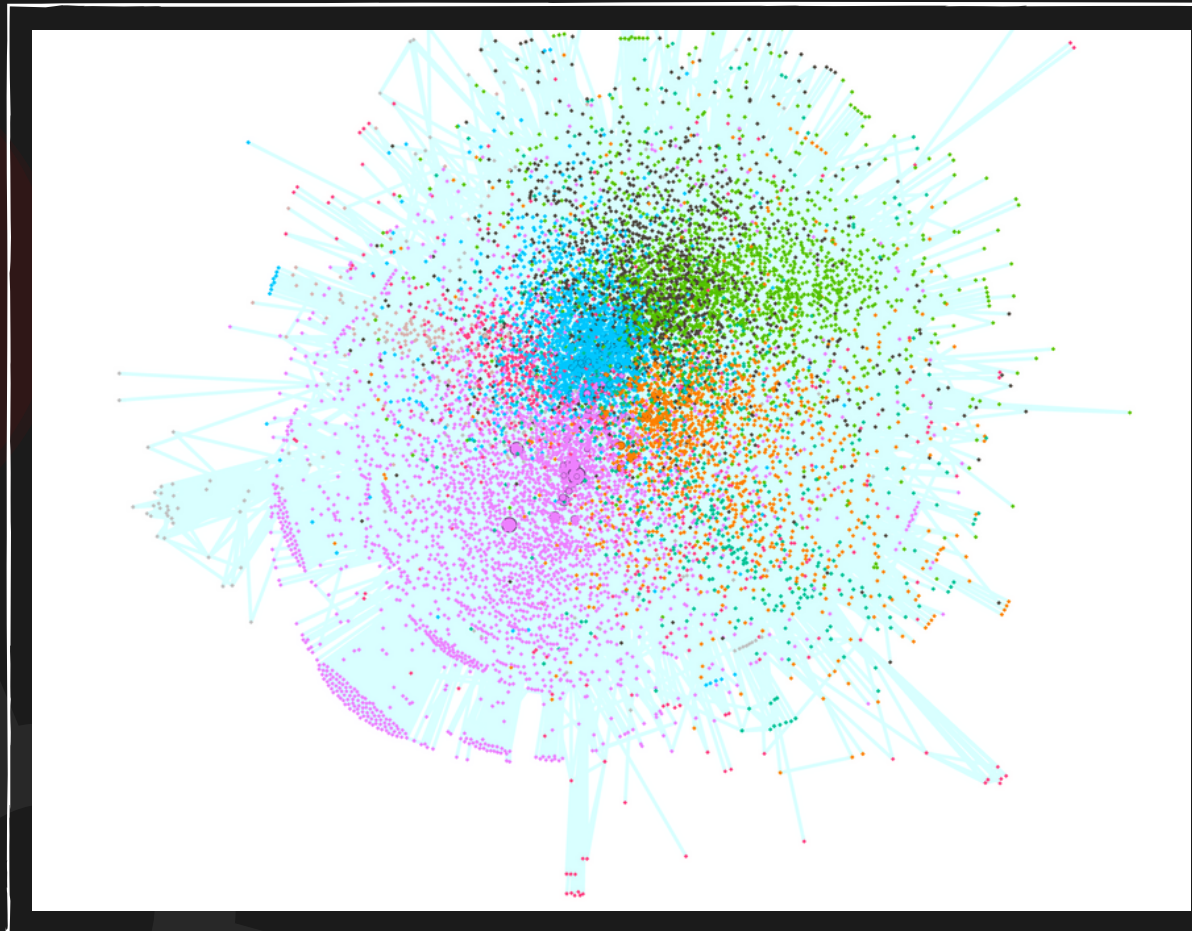
Friendship Network



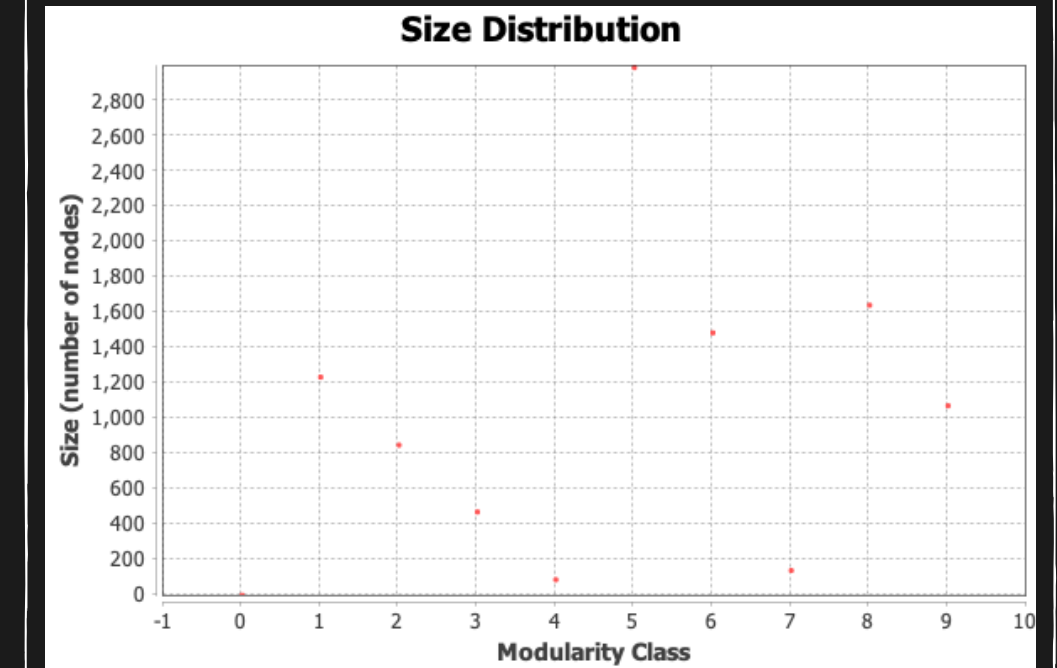
50% of the nodes have a degree less than 12.
Also, 80% nodes have a degree less than 50.



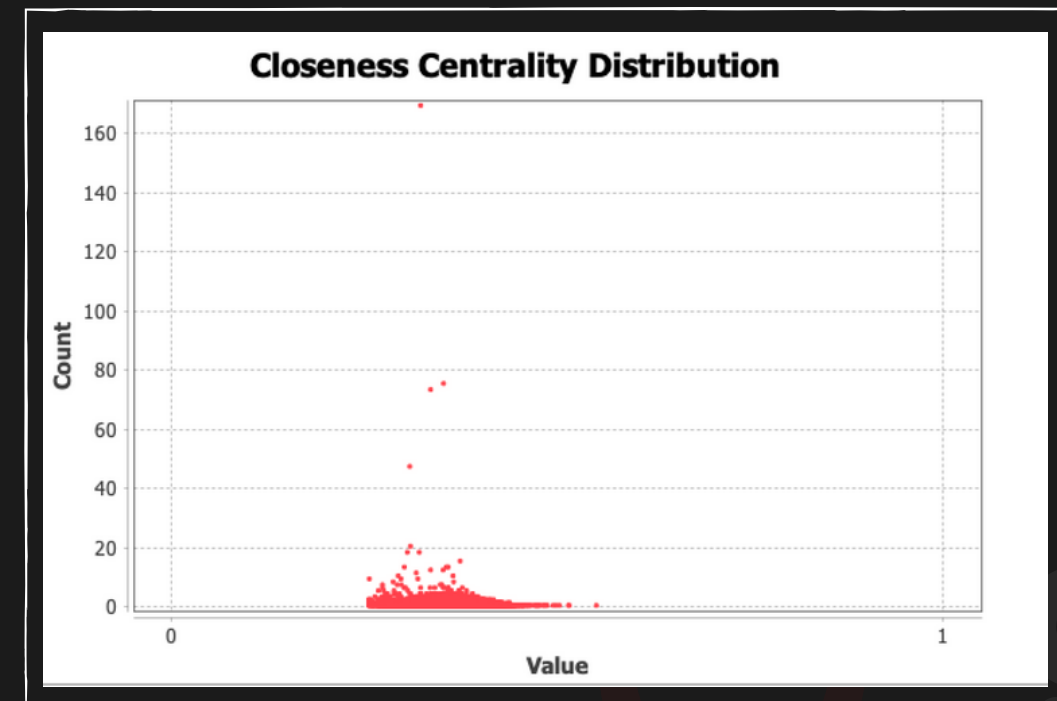
50% of nodes have eigenvector centrality less than 0.03
Less than 5% of the nodes have a value of more than 0.1
No. of Iterations: 100



5 Major Communities can be clearly visualise.



Modularity 0.48 No. of Communities 10
5 communities with more than 1K people.

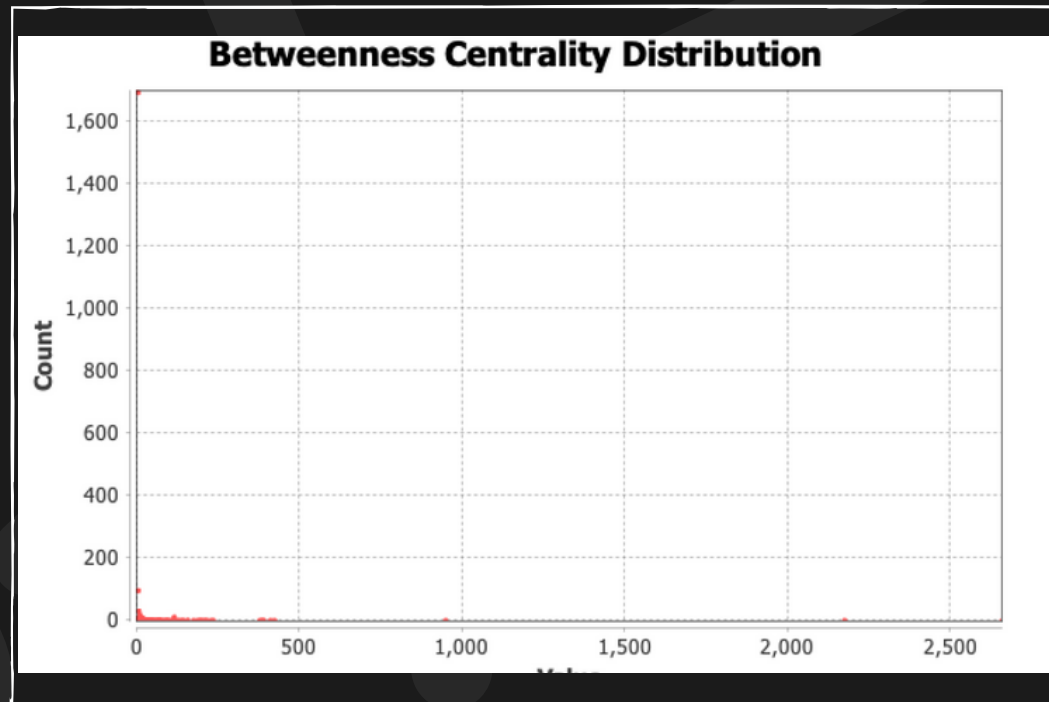


Average path length 2.89
Network Diameter 4

NETWORK STATISTICS

Checkins Network

September Evening Checkins



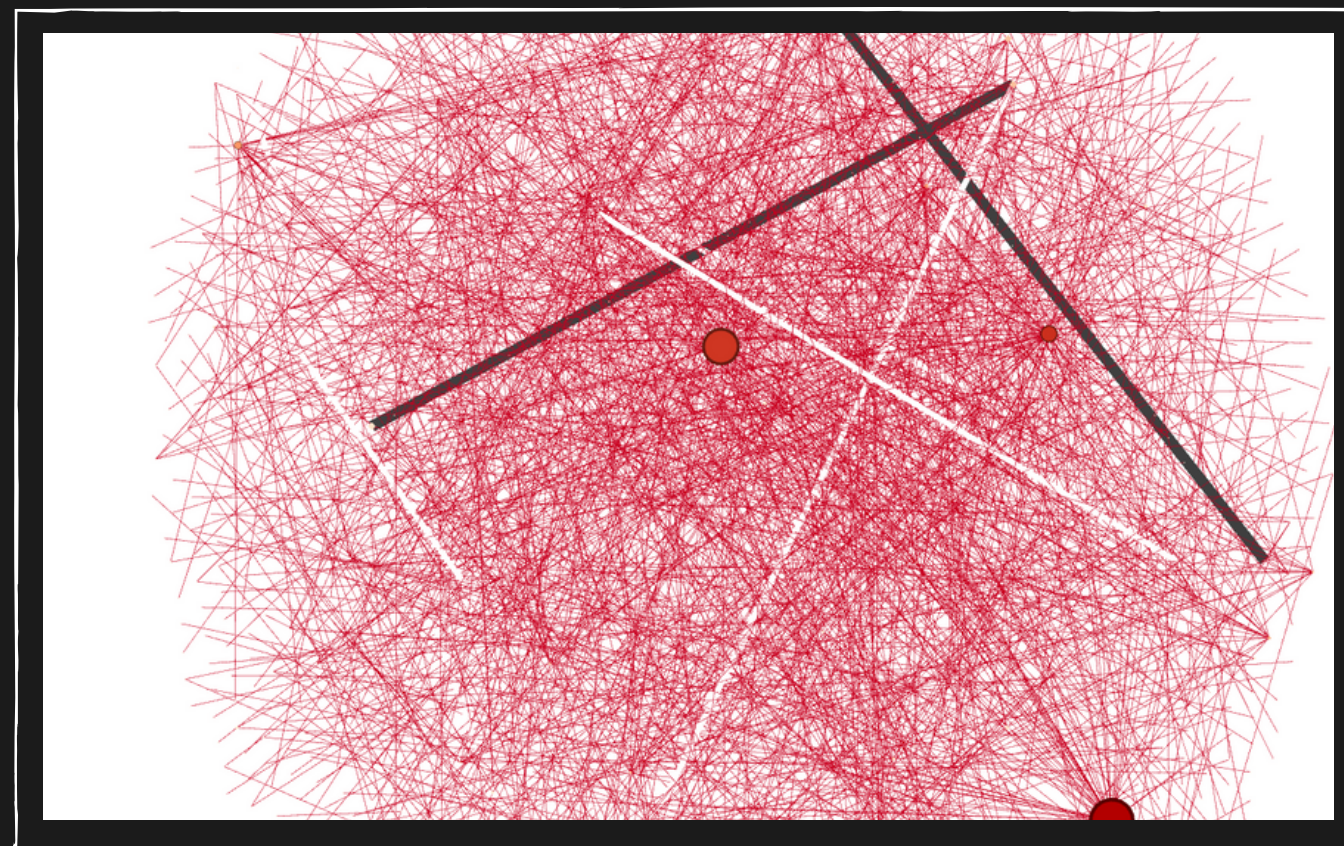
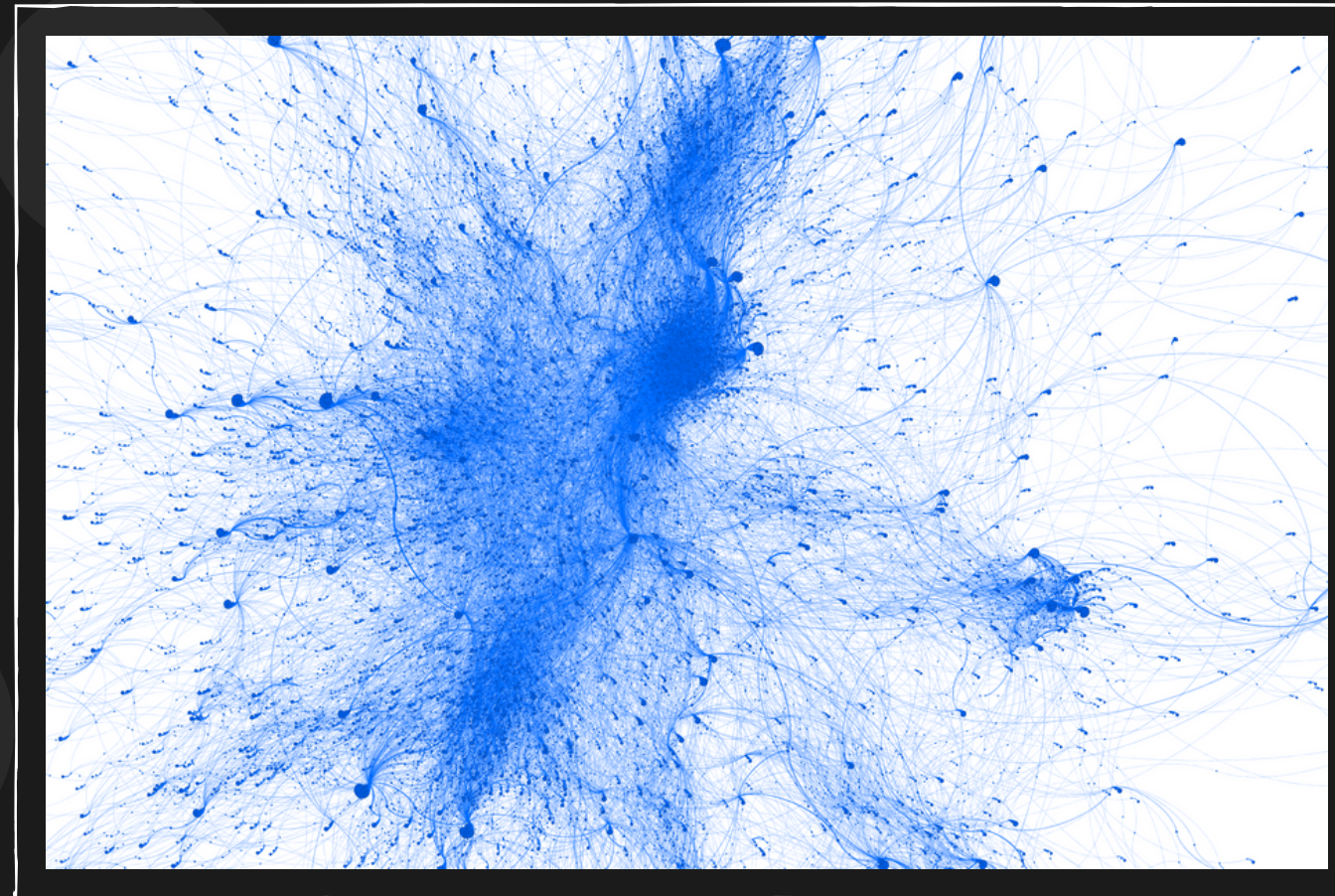
1.5% locations have betweenness more than 100

Average Path Length: 2.4

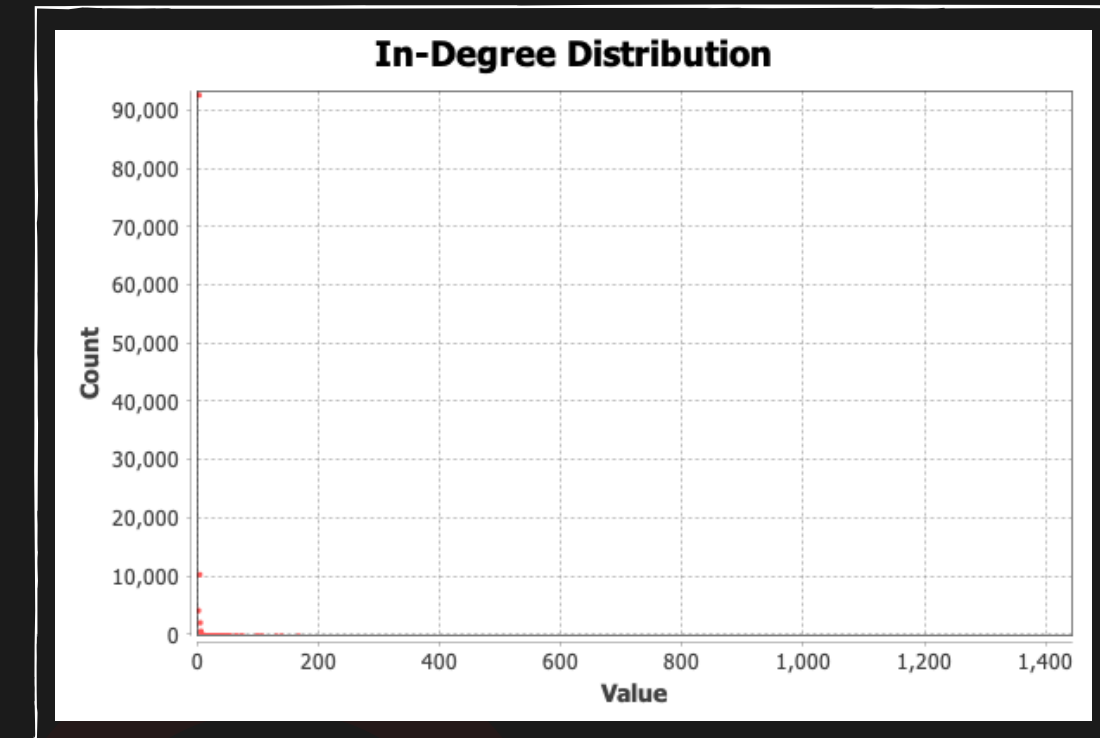
Approximately, 10% of people have a betweenness more than 100

The large size nodes have large betweenness centrality hence maybe the hotspot.

Dense clouds may represent these hotspots(most visited location).



September Checkins



60% of people have an outdegree less than 18, 80% of the people have an outdegree less than 50.

Set of 13 locations which have indegree greater than 50

Captures the local hotspots which are the location visited the most. In the distribution we can see certain nodes having in-degree above 100, these could be the locations which are potential hotspot for spreading diseases.

APPROACH

Epidemic Simulation in Community

Using SIR (Susceptible, Infectious, Recovered) model to simulate epidemic spread.

Identify potential hotspot locations and persons with maximum betweenness and degree centrality

Simulate epidemic spread considering a probability p to infect over coming in contact

Centrality with Epidemic Spread

Identify the nodes with highest centrality over the friendship network

Consider betweenness, eigenvector, degree, closeness and other centrality measures

Simulate spread after removing 5%,10%,20% of nodes with highest centrality

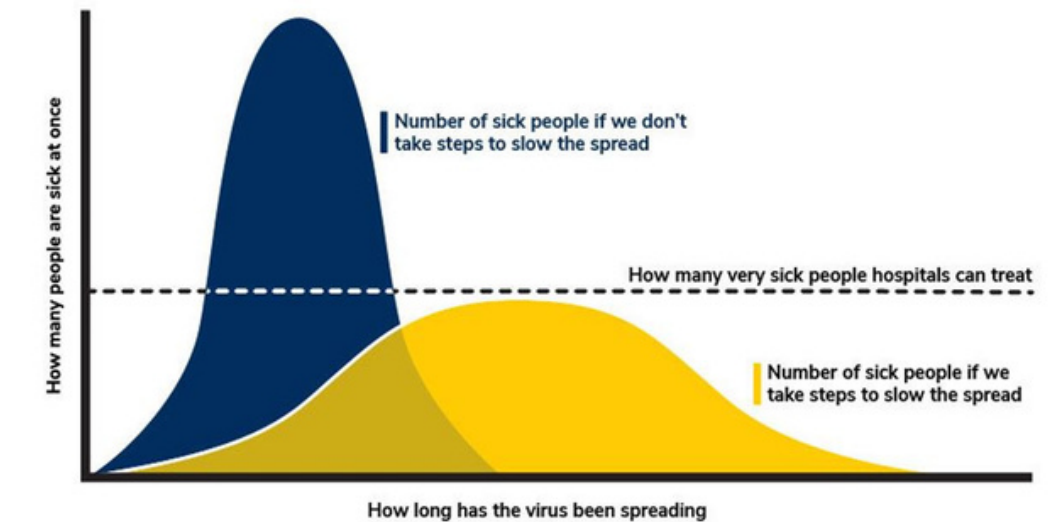
Community with Epidemic Spread

Simulate epidemic spread within a community assuming epidemic triggers from a **single person** there

Identify key hotspots and persons removing which (whom) would slow the spread **within** community

Simulate epidemic spread between communities assuming epidemic triggers from a **single community**

Identify key hotspots and persons removing which (whom) would slow the spread **between** communities



Ref:<https://link.springer.com/article/10.1007/s13755-020-00121-9>

Key Metrics

Epidemic Simulation

- Size of Spread/ Final Size
- Time for spread of infection with number infected at a time
- R number by time

Removing nodes with highest centrality

- How Size of Spread is reduced
- Time taken to spread increases
- R number along with max infected at a point is reduced

Community with Epidemic Spread

- Size of Spread within a community is reduced
- Reduce time to spread between communities
- Overall R number impact with max infected at a point

The background is a dark gray with abstract geometric patterns. It includes several concentric circles of varying sizes and colors (dark gray and dark red). There are also lines radiating from central points, creating a star-like or network-like structure. A thin white horizontal line is positioned near the top right, and another is near the bottom left, both with small white squares at their ends.

Thank You..