Epidemics Graph Neural Network Node Classification Project Definition

Jaykumar Patel
The University of Texas at Austin
110 Inner Campus Drive
Austin, TX 78712
patel.jay4802@utexas.edu

Afnan Mir
The University of Texas at Austin
110 Inner Campus Drive
Austin, TX 78712
afnanmir@utexas.edu

1. Overview of the Project

When COVID-19 first appeared, one of the key ways that its spread was mitigated was through contact tracing. Contact tracing is the process of "identify[ing] individuals who have possibly come in close contact with an infected person while that person was the carrier of the viral pathogens" [?].

However, the pandemic revealed that the COVID-19 disease can spread faster than manual contact tracing. Thus, the objective of this project is to automate contact tracing by incorporating machine learning using Graph Neural Networks (GNNs). This automation will allow quicker contact tracing, and possibly lead to a greater mitigation of the spread of COVID-19 when compared to manual contact tracing.

2. Objectives and Deliverables

First, we will analyze the networks to determine their properties. Then, to better understand network behavior, we will run simulation to see how COVID-19 spreads through the network.

The main goal of contact tracing is to find people who may be infected due to their contact with other infected people. Similarly, this project will aim to classify infected nodes in a graph, where nodes represent people, and links represent contact between people. This network will be formed by utilizing real contact information from mobile devices. In order to predict infections within a network, this project will implement and analyze GNNs, specifically link prediction.

The dataset we will be using is the "foursquare" mobility dataset that consists of location visit logs in Austin, TX from 2019 to 2021. The visit log includes information about when and where a device was, as well as how long it was there.

3. Tasks and Timeline

Firstly, we will collect and clean the data. This includes reformatting and filtering the data. We will analyze the dataset to get a better understanding of its properties (such as degree, betweenness centrality, clustering coefficient, etc.). We plan to complete this by October 10th, 2023.

Then we plan to run simulations on the network to get a better understand of how the virus flows. We plan to complete this by November 14th, 2023.

Finally, we will utilize GNNs to perform link prediction to see how the network evolves over time. We plan to complete this by November 30th, 2023.

We will pair program; thus, the member contribution will be 50/50.

4. Conclusion and References

In conclusion, the project aims to utilize GNNs to make contact tracing more efficient and accurate. We will analyze the dataset to understand its properties, we will run simulations to see how the virus spreads through the network, and we will train GNNs to perform link prediction.

We will start by researching GNNs. Particularly, we will look at an article by Neptune.ai on the application of GNNs and a DGL tutorial on link prediction [?] [?].

References

[1] Link Prediction using Graph Neural Networks x2014; DGL 0.8.2post1 documentation — docs.dgl.ai. https://docs.dgl.ai/en/0.8.x/tutorials/blitz/$4_link_predict.html$.

[2] A. Menzli. Graph neural network and some of gnn applications: Everything you need to know, September 11 2023.

[3] M. Shahroz, F. Ahmad, M. S. Younis, N. Ahmad, M. N. K. Boulos, R. Vinuesa, and J. Qadir. Covid-19 digital contact tracing applications and techniques: A review post initial deployments, 2021.