Temporal Analysis of Epidemiology indicators and Air Travel Data for Covid-19

Sumit Purohit* Filipp Shelobolin † Lawrence B. Holder ‡ George Chin§

Abstract

Coronavirus Disease 2019 (Covid-19) is an ongoing outbreak and the latest threat to global health. It is imperative to understand the implications of social interaction on Covid-19 indicators in order to help formulate policies and guidelines by governments and local authorities. We present a case-study of curating state-level Covid-19 indicators such as Active Cases, Deaths, Hospitalization Rate, etc. for the United States. We also curate open source domestic US air travel data and present its impact on Covid-19 indicators. We perform a time-series analysis of the dataset using Independent Temporal Motif (ITeM) to find weekly trends in the data. We publish the dataset and the results for further exploration by the research community.

1 Data Collection

One of the major contributions of this work is a dynamic, heterogeneous, attributed graph that describes various real-world indicators and travel patterns from January 2020 to August 2020. The dataset is divided into two groups to describe global data and domestic US data. Global Covid-19 data model each country as a node in the graph and air-travel between two countries is modeled as dynamic edges. Similarly, the domestic US dataset models each US state as a node and air-travel between them as dynamic edges. This work uses different indicators curated for the Covid19 dataset extracted from sources such as US Census, Google geocoding, JHU C19 database, etc. The edge data for global passenger flow is simulated from a model developed by Mao et al. [4]. It uses a Poisson regression model to predict monthly passenger volumes between directly connected airports.

We use The Bureau of Transportation Statistics (BTS) [1] to collect open-source domestic US air-travel data. We also use Google Covid-19 open data repository [2] and Geocoding API to extract temporal data for each US state. We also open source the code to set up the repository and provide instructions to daily update the dataset. All the data and the data curation pipeline are available at https://github.com/fshelobolin/C19DynamicGraph.

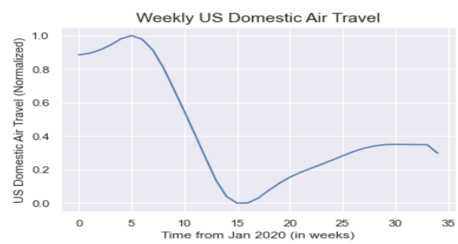

Figure 1: Domestic US air-travel trend From Jan 2020-Aug 2020

2 Independent Temporal Motif (ITeM)

Independent Temporal Motif (ITeM) is an edge-disjoint temporal motif [6] used to model transactional networks [5]. We use ITeM to observe temporal patterns in the Covid-19 travel graph. The ITeM provides insight into the temporal evolution of the travel graph, such as its rate of growth, neighborhood, and the change in the travel pattern of a state over time when modeled as a node in the graph.

3 Temporal Analysis using ITeM

We create a dynamic graph to model air travel between different US states. Each US state is modeled as a node in the graph. We create edges using the total air-travel passenger count. We create a logarithmic bin to reduce the number of edges between any two states for a given day, without the loss of overall travel trends as shown in the Figure 1. We also create a sequence of weekly temporal graphs to measure temporal trends in the Covid-19 air-travel dataset. We generate a weekly average for all Covid-19 indicators. As shown in Figure 2, we qualitatively observe three different classes of real-world Covid-19 indicators. The “Hospitalization Rate”

*Pacific Northwest National Laboratory
†Carnegie Mellon University
‡Washington State University, Pullman WA
§Pacific Northwest National Laboratory

and the absolute number of “People Hospitalized” in the US averaged over a week show opposite trends. In contrast, indicators such as “Active Cases”, “Confirmed Cases”, “Deaths” show a slowly-increasing trend over the time window. The “Mortality Rate” does show an initial sharp upward and slowly decreasing trend similar to “Hospitalization Rate” but in contrast to the other indicators. We also compute ITeM distribution for each week and a scalar value for each distribution using Principle Component Analysis (PCA) to compare temporal trends measured using ITeM with the rest of the indicators. As shown in Figure 2, ITeM shows a similar temporal trend as “Hospitalization Rate”.

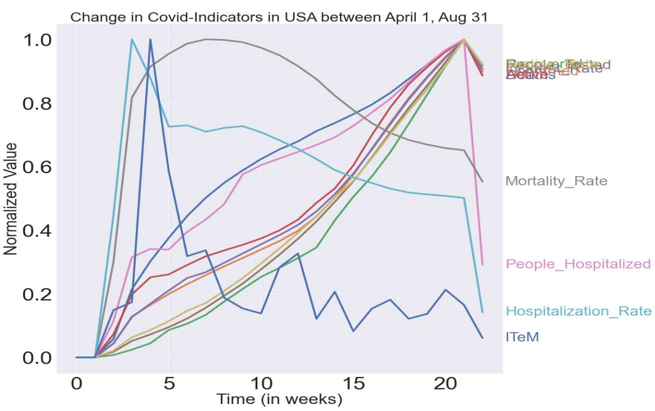

Figure 2: Weekly Real-world Indicators and ITeM trends from April 2020-Aug 2020

To quantify ITeM similarity with the rest of the indicators, we use Dynamic Time Warping (DTW) [3]. DTW measures the similarity (or distance) between two time-series by computing optimal matches between them. We compute the pair-wise distance between all real-world indicators and ITeM distribution for each week from April 2020 to August 2020. We used the degree distribution of the travel graph as the baseline similarity with each indicator. As shown in Figure 3, degree distribution does not show any variation in the similarity with any of the real-world indicators. The overall similarity is as good as with a random time series. In contrast, ITeM distribution in Figure 4 shows an interesting trend as it shows high similarity with average travel and ”Hospitalization Rate”. Specifically, it shows that the ITeM-based time series has the most similarity (or lowest distance) with ”Hospitalization Rate” that corroborate our initial observation as shown in Figure 2.

4 Conclusion

Covid-19 is an ongoing challenge to global health and poses various open research questions to understand

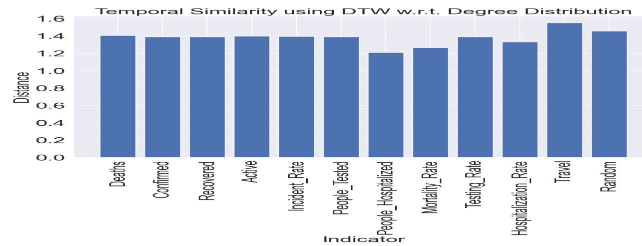

Figure 3: Covid19 Real-world Indicator Similarity with Degree Distribution

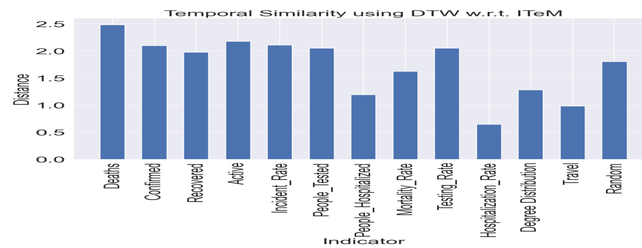

Figure 4: Covid19 Real-world Indicator Similarity with ITeM Distribution

the dynamics and nature of the spread. Graph-based temporal analysis can be used to gain insight into social interaction patterns and their impact on real-world Covid-19 indicators. We present an open-source dynamic graph representation of Covid-19 indicators and travel data. We use ITeM to explain such interaction patterns using domestic US air travel and show its correlation with the indicators.

References

[1] Bureau of transportation statistics. https://www.bts.gov/tsar.

[2] Google covid-19 open data repository. https://cloud.google.com/blog/products/data-analytics/publicly-available-covid-19-data-for-analytics.

[3] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA:, 1994.

[4] Liang Mao, Xiao Wu, Zhuojie Huang, and Andrew J Tatem. Modeling monthly flows of global air travel passengers: An open-access data resource. *Journal of Transport Geography*, 48:52–60, 2015.

[5] Sumit Purohit, Lawrence B Holder, and George Chin. Temporal graph generation based on a distribution of temporal motifs. In *14th International Workshop on Mining and Learning with Graphs*, 2018.

[6] Sumit Purohit, Lawrence B. Holder, and George Chin. Item: Independent temporal motifs to summarize and compare temporal networks, 2020.