
Temporal Graph Analysis for Outbreak Pattern Detection in COVID-19 Contact Tracing Networks

Dario Antweiler

Fraunhofer Center for Machine Learning
Fraunhofer IAIS
Sankt Augustin, Germany
dario.antweiler@iais.fraunhofer.de

Pascal Welke

University of Bonn
Germany
welke@cs.uni-bonn.de

Problem statement A main challenge for local healthcare authorities in the ongoing COVID-19 pandemic is tracking and tracing infections and high risk exposure contacts [1]. Accurate analysis of the corresponding data can improve decisions on adapting social restrictions and declaring quarantines to stop the pandemic from spreading. Early detection of outbreaks are vital for a comprehensive public health strategy, especially with the median serial interval for COVID-19 being shorter than the incubation period [2], i.e. people being infectious before developing symptoms in contrast to most other infectious diseases [3]. Contact tracing data can be represented as a temporal graph. We propose a framework consisting of temporal graph analysis methods detecting outbreak patterns in these networks to help the local healthcare authorities manage the situation by identifying persons most at risk of spreading the virus further. We assess our framework on a real world dataset on contact tracing including more than 10k persons.

We collaborate with Germany’s largest local healthcare authority in Cologne, covering a population of over a million. The data provided results in a temporal graph consisting of nodes (positively tested cases and their contacts) and temporal edges (infection and exposure events) which changes over time as more people get infected and report contacts, that had a high risk exposure as seen in Fig. 1.

Method We let the user select a number of connected components from the network, that she deems interesting in her current analysis question to allow an intuitive querying of the data. Our system then simultaneously identifies similar components in the network as well as time frames containing a large number of such components. Additionally, the framework identifies and returns the features which are most relevant to that classification decision, detecting recurring temporal patterns. Finally, the identified components and their corresponding time periods are displayed as an interactive network visualization supporting decision making on resource allocation.

The features under investigation are jointly selected with public healthcare experts. One of the proposed component metrics is as follows. Given a component of the network C , we define the risk factor $r(C) = a_{\text{med}}/s_{\text{med}}$ as the quotient between the median number of infections per index case a_{med} and the median serial interval between subsequent generations of infections s_{med} . Examining components with high risk scores reveals fast-moving components with a large number of added infections, making the monitoring of included contacts especially important.

We plan to evaluate the framework against related methods for persistence mining [4], motif mining [5] and frequent subgraph mining [6]. We also plan to get feedback from subject matter experts in our partnering local healthcare authority. Further research will be dedicated to improving detection models by introducing additional collected data on local infrastructure, geographic distribution and medical records.

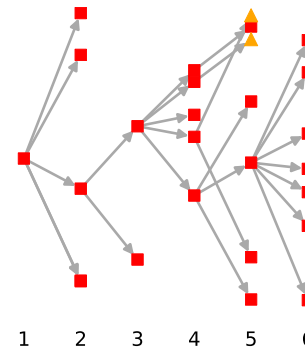


Figure 1: Component of the COVID-19 contact tracing network showcasing positive cases per generation

Broader Impact

This work is aimed at supporting governmental and healthcare officials in their informed decision making with regard to spreading and containing infectious diseases. The described framework works with anonymized data, its results are open for interpretation and should be examined cautiously by subject matter experts.

Acknowledgments and Disclosure of Funding

This work was developed by the Fraunhofer Center for Machine Learning within the Fraunhofer Cluster for Cognitive Internet Technologies and has been partially funded by the Federal Ministry of Education and Research of Germany as part of the competence center for machine learning ML2R (01IS18038C).

References

- [1] WHO Global. (2020). Contact tracing in the context of COVID-19. Online: <https://www.who.int/publications/i/item/contact-tracing-in-the-context-of-covid-19>
- [2] Ma, S., Zhang, J., Zeng, M., Yun, Q., Guo, W., Zheng, Y., Zhao, S., Wang, M., & Yang, Z. (2020). Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. medRxiv.
- [3] Robert Koch-Institut (RKI). (2020). SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19) (German).
- [4] Belth, C., Zheng, X., & Koutra, D. (2020). Mining Persistent Activity in Continually Evolving Networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 934–944). Association for Computing Machinery.
- [5] Ashwin Paranjape and Austin R. Benson and Jure Leskovec (2017). Motifs in Temporal Networks. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, February 6–10, 2017 (pp. 601–610). ACM.
- [6] Ehab Abdelhamid and Mustafa Caim and Mohammad Sadoghi and Bishwaranjan Bhattacharjee and Yuan-Chi Chang and Panos Kalnis (2018). Incremental Frequent Subgraph Mining on Large Evolving Graphs. In 34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16–19, 2018 (pp. 1767–1768). IEEE Computer Society.