Mobility data improve forecasting of COVID-19 incidence using Graph Neural Networks

Simon Witzke[1], Noel Danz[1], Katharina Baum[1,2], Bernhard Y. Renard[1]
[firstname.lastname]@hpi.de
[1]Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Germany
[2]Department of Mathematics and Computer Science, Free University Berlin, Germany

ABSTRACT

The COVID-19 pandemic has had a considerable global impact over the last few years. Many efforts were made to understand and estimate its development. The availability of large amounts of data, including mobility data, has led to numerous Graph Neural Networks (GNN) being proposed to leverage this data and forecast case numbers for the short-term future. However, information about trend developments, especially where trends reverse directions, is crucial in informing decisions. GNNs may be able to use information from regions where trends change first to improve predictions for locations with delays. We consider the first omicron wave in Germany at the end of 2021 and compare a heterogeneous GNN using mobility data with a model without spatial information. We observe that, for this period, mobility data significantly improve forecasts and specifically that improvements occur earlier in time. Using GNNs and mobility data enables leveraging information from counties affected earlier to improve forecasts for counties affected later. We conclude that such performance improvements could be transferred to counties with earlier change points by also including neighboring states in the graph structure. Further, we emphasize the need for systematic contextual evaluation of GNN-based models for forecasting pandemic trends.

KEYWORDS

mobility data, trend estimation, graph neural networks, covid-19

1 INTRODUCTION

Spreading from Wuhan, China, in late 2019, the COVID-19 pandemic has held humanity in its grasp until recently [29]. The pandemic has had drastic consequences, with estimates of almost fifteen million excess death only in 2020 and 2021 [16] and considerable economic and social damages [4]. The global scale of the pandemic led to large amounts of data on different modalities related to epidemic spread being shared, such as mobility and sequencing data. These have been made available to support the development of forecasting methods intended to inform decision makers concerning potential interventions [17, 18]. Human mobility is a central driver in the geographical spread of epidemics caused by air-borne diseases [3], enabling the virus to travel between regions and, in the case of COVID-19, rapidly infecting most of the world. Schlosser et al.[25] have shown that lockdowns strongly impacted mobility structures during the first COVID-19 wave in Germany and that the associated reduction in mobility can slow the virus' geographical spread.

Recent advances in deep graph learning have led to Graph Neural Networks (GNNs) gaining popularity in domains as diverse as traffic forecasting [9] or computational chemistry [21]. Human mobility between geographical regions can naturally be represented as graphs, where nodes represent locations, such as counties, and edges movements between them. Consequently, numerous approaches that try leveraging the power of GNNs to forecast COVID-19-related metrics, such as cases, deaths, and hospitalizations, have been proposed [7, 8, 10, 19]. These approaches have shown promising results in providing insights into the short-term development of the COVID-19 pandemic. However, informing decision makers about a trend forecast rather than exact numbers might be more beneficial, as they are strong indicators of a need for interventions, and their interpretation is straightforward. For this purpose, systematically evaluating GNN-based methods' ability to correctly forecast trends is essential. Accurate forecasts are especially relevant for phases with change points, where locations successively experience a change in their trend, such as the peak of a wave.

There are secondary time series modalities, such as Google search trends and smart body temperature sensors. These modalities potentially reflect changes in trends faster than case numbers. This has been successfully leveraged by Kogan et al.[12] and Stolerman et al.[26] to develop early-warning systems in the United States that detect such trend signals up to weeks in advance. Similarly, GNNs may utilize nodes with leading time series to improve forecasts for nodes with lagging time series by passing information via the underlying graph, i.e., information from locations where changes occur earlier might be beneficial for forecasting locations where similar changes are delayed.

In this work, we investigate whether mobility data can improve forecasts of 14-day linear trends of the COVID-19 incidence. We evaluate county-level forecasts of a heterogeneous GNN for locations experiencing a change point during the second half of the first omicron wave at the end of 2021 in Germany [15], where cases are beginning to decline. We further analyze whether our GNN can utilize information from counties with leading changes for forecasting counties that experience similar changes later. Finally, we discuss the implications for developing and evaluating future GNN-based methods for pandemic forecasting.

2 MATERIALS AND METHODS

2.1 Graph Construction

Inspired by Kapoor et al.[10], we construct heterogeneous spatio-temporal graph samples with distinct edge types for spatial and temporal connections. We design each graph sample to contain 15 weighted mobility subgraphs, representing movements between the 400 German counties as nodes at successive points in time, $t-14, ..., t$. We use spatial edges to express these mobility graphs. The directed but unweighted temporal edges then link each county

at a time point $t-14, ..., t$ to its representations on up to seven previous days, connecting the spatial components of the graph.

We use mobility data [13, 23] to build the spatial edges. The used dataset contains the daily movements of nearly one million mobile phone users in Germany and is non-public due to privacy concerns. The number of mobile phones sending location information varies daily, so we normalize the movements by the daily device count and then re-scale all movements with the average daily device count. We find that the daily mobility networks' adjacency matrices are primarily symmetric, i.e., the opposing edges are highly similar. Therefore, we convert the directed into undirected graphs by summing the weights of the edges in both directions. Finally, we denoise the mobility graphs by removing 30% of the non-zero edges with the lowest edge weights, where edges on the thresholding boundary are removed randomly.

The node features of our graph consist of dynamic and static features. We obtain data on the COVID-19 case numbers from the Robert Koch Institute [22] and aggregate the data on the county level, resulting in a total of 400 time series. Countering reporting inaccuracies, we calculate the county-level 7-day incidence, a right-aligned 7-day moving sum normalized by the county population and then scaled by 100,000. Each node at time t has the 7-day incidence of the previous seven days until day $t-6$ as node features. Additionally, we include a cyclical sine/cosine encoding [28] for the weekday and month. Lastly, we use the population density of each county as the only static feature. We collect the census data, such as population size and population density, from the German Federal Office of Statistics [14].

As prediction targets, we obtain 14-day trends in the COVID-19 incidence by linear approximations. A linear approximation has the advantage that it allows us to estimate the strength of a trend and not only its direction compared to converting the problem to a classification task. For this purpose, we smooth the 7-day incidence to remove remaining artifacts, using a center-aligned 7-day moving average. For each county, we perform a linear regression on this smoothed time series with the time series values at time points $t+1, ..., t+14$ as the dependent variable and the number of days from time t into the future as the independent variable. We then use the slope of this regression, representing a linear trend of the COVID-19 incidence over the next 14 days, as the target for our forecasts.

2.2 Graph Neural Network

Our GNN is similar to the network used by Kapoor et al.[10] and based on Kipf and Welling's[11] graph convolutional layer. We extend this architecture by using relational graph convolutional layers (R-GCN), an extension for heterogeneous graphs proposed by Schlichtkrull et al.[24] that allows feature updates via multiple edge types, where each edge type has its own set of learned parameters. First, the node features are passed through an initial encoding layer followed by a dropout with a probability of 0.2. Next is a three-layer GNN, each with a dropout probability of 0.5. Like Kapoor et al.[10], we add skip-connections and concatenate the output of the initial encoding layer to the output of each R-GCN layer to preserve local information and counter over-smoothing. Lastly, we use a multi-layer perceptron with a single hidden layer to produce the final prediction. We note that for each graph sample, we only use the embeddings of the most recent spatial subgraph to obtain a single forecast for all 400 counties. All layers have 32 hidden units and use a ReLU as the non-linear activation function, except for the last linear layer, which has 16 hidden units. The output layer uses no activation function, allowing positive and negative trend predictions. We implement our GNN in PyTorch [20] and PyTorch Geometric [5].

2.3 Training setup

We use a mean squared error (MSE) regression loss and an ADAM optimizer with a learning rate of 1.33e−4 and weight decay of 1e−5. We employ a batch size of 128 and train for a maximum of 250 epochs with early stopping, with a patience of 10 epochs without improvement.

We adopt a rolling-origin evaluation approach [27] where we extend the training set by the test sample of the previous iteration. We test from November 10, 2021, until December 19, 2021, with all previous data being used for training and validation. Our validation set consists of the day after the last training sample and is used for early stopping and model selection. We always have a 17-day gap between the validation and test samples to avoid information leakage to the test sample while also mimicking a real-world situation where we use all the available data to make a forecast.

To counter the sparseness of training data and avoid conditioning our model too strongly on periods that contain limited information, such as summer periods with low incidences, we oversample the training set by multiplicating specific samples. We convert the global German COVID-19 incidence time series into a discrete probability distribution where each sample is assigned a probability. We then draw from this distribution with replacement. We use an oversampling rate of 10.

2.4 Evaluation Scenario

While we train our models using an MSE regression loss, this metric is not optimal for evaluating our models' performance. Different counties experience the considered phase of the pandemic differently and a metric dependent on the range of the trend values could bias our evaluation.

Therefore we evaluate the models' performance using the Mean Absolute Percentage Error (MAPE) (Appendix A.1) and the symmetric Mean Absolute Percentage Error (sMAPE) (Appendix A.2). Further, while MAPE and sMAPE provide insight into the error in the magnitude of the trend, we are also interested in the model's ability to predict the direction of the trend. For this purpose, we evaluate our models with an adaption of the Mean Directional Accuracy (MDA) (Appendix A.3).

To investigate if our models can leverage mobility data to improve predictions in counties with lagging change points, we consider the first omicron wave at the end of 2021, from November 10 to December 19. We extract the date on which each county's corresponding smoothed COVID-19 7-day incidence time series has its maximum, i.e., its peak. We consider this the point when the trend will likely change from positive to negative as the incidence begins to decline.

We use a 7-day moving window to evaluate how the prediction performance develops as more counties reach their peak. For each window, we collect all counties that have their peak inside this window. We then compute all metrics for these counties using the forecast and ground truth of their respective peak date.

We conduct additional experiments with the same evaluation setup but replace the adjacency matrices of the mobility subgraphs with identity matrices to verify that difference in performance can be accounted to the mobility data. Thus, we train models with the same number of parameters but do not include spatial information.

3 RESULTS

For all experiments, there is a clear performance improvement as more counties reach their peak over time that is consistent across all metrics. This improvement is more pronounced for models with mobility data than those without spatial information (see Figure 1). To verify that our findings that models with mobility data perform better than models without spatial information are significant, we conduct paired one-tail Wilcoxon signed-rank tests with significance level $\alpha = 0.05$ for all metrics. After correcting for multiple testing using the Benjamini-Hochberg method [1], we find that for MAPE ($p\text{-}value \approx 0.021$), sMAPE ($p\text{-}value \approx 2.738\text{e}{-6}$), and MDA ($p\text{-}value \approx 6.661\text{e}{-6}$) the mobility-conditioned models significantly outperform the models without spatial information.

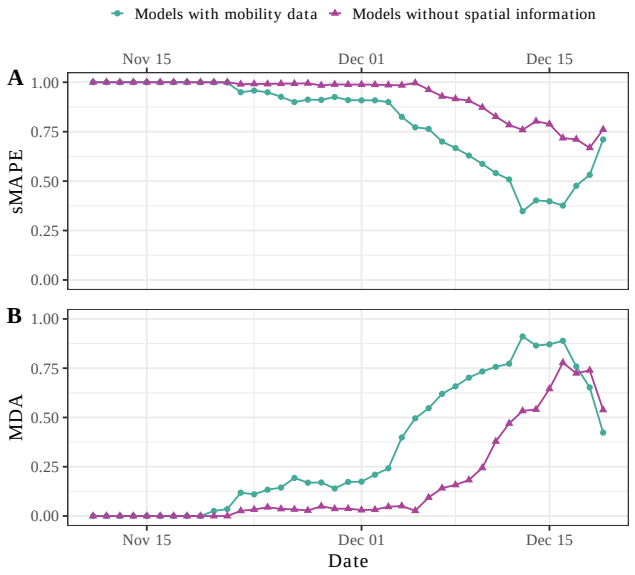

Figure 1: (A) sMAPE (lower is better) for peaks in 7-day moving windows. The performance improves over time for both experiments before declining. The effect occurs earlier and is greater for models with mobility data. (B) The MDA (higher is better) almost mirrors the sMAPE's behavior. This suggests that while more recent training data improve predictions, this effect is amplified by mobility data.

Figure 1 (A, B) clearly shows that the improvements in sMAPE and MDA happen earlier and are more extreme for the models with mobility data. This difference indicates that the improvements cannot solely be attributed to the fact that the models have seen more recent and relevant data and are therefore conditioned better. Furthermore, due to the 17-day gap to avoid information leakage, the model is unlikely to have seen any recent negative trends for a county before its peak during training. However, as earlier counties are already past their peak and are experiencing decreasing incidences, they can share this information with counties where peaks occur later.

4 DISCUSSION AND CONCLUSION

We find that mobility data significantly improve forecasting performance compared to experiments without spatial information. We have two hypotheses for our observations. Firstly, the structural information in the mobility networks and their variation over time might lead to improved predictions. Secondly, our GNN model can pick up information from counties that experience changes, such as beginning downtrends in incidences, earlier and use them for forecasts of counties where these changes occur delayed. With our current experimental setup, we are unable to disentangle these hypotheses. However, further experiments, for example, using static spatial connections, could provide insights.

Counties that are the first to experience a change in trend seem unable to benefit from mobility data. However, these counties might be of the highest interest as changes occur earlier and are likely more vital indicators of the need for interventions. Therefore it could be valuable to include additional nodes representing neighboring states in our graph to leverage potentially leading information from them.

Our analysis suggests that systematically analyzing models' capabilities of making accurate trend forecasts during times of interest is highly valuable. Different components, such as the magnitude and direction of a trend, are relevant for providing a holistic understanding in an epidemiological context. It could be helpful to extend evaluations by applying post-hoc explainability methods for graph-based models to understand better how the models make their predictions. Such explanations could provide insights for epidemiologists to construct hypotheses regarding the pandemic's current state and spreading behavior.

We showed the capabilities of a heterogeneous spatio-temporal GNN in leveraging mobility data to improve forecasts for counties with lagging time series directly after a change in trend. We suggest that including more global information via nodes representing other states could extend this effect to leading counties where changes occur first. Currently, we evaluate single rolling-origin evaluation experiments for the change point of the COVID-19 pandemic in Germany. To substantiate our findings, we will consider different phases of the pandemic, including change points with a switch to upward trends. Furthermore, we will run experiments repeatedly to verify the robustness of our results and establish confidence bounds.

ACKNOWLEDGMENTS

This work was supported by the German BMWK through the DAKI-FWS project [01MK21009E to B.Y.R.].

REFERENCES

[1] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 1, 289–300. Retrieved May 30, 2023 from https://www.jstor.org/stable/2346101.

[2] Oliver Blaskowitz and Helmut Herwartz. 2009. On economic evaluation of directional forecasts, (Oct. 29, 2009). Publisher: Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät. DOI: 10.18452/4217.

[3] Dirk Brockmann and Dirk Helbing. 2013. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342, 6164, (Dec. 13, 2013), 1337–1342. DOI: 10.1126/science.1245200.

[4] Orestis Delardas, Konstantinos S. Kechagias, Pantelis N. Pontikos, and Panagiotis Giannos. 2022. Socio-economic impacts and challenges of the coronavirus pandemic (COVID-19): an updated review. *Sustainability*, 14, 15, (Aug. 6, 2022), 9699. DOI: 10.3390/su14159699.

[5] Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with PyTorch geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*. arXiv, (Apr. 25, 2019). DOI: 10.48550/arXiv.1903.02428.

[6] Benito E Flores. 1986. A pragmatic view of accuracy measurement in forecasting. *Omega*, 14, 2, (Jan. 1986), 93–98. DOI: 10.1016/0305-0483(86)90013-7.

[7] Cornelius Fritz, Emilio Dorigatti, and David Rügamer. 2022. Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly COVID-19 cases in germany. *Scientific Reports*, 12, 1, (Mar. 10, 2022), 3930. DOI: 10.1038/s41598-022-07757-5.

[8] Junyi Gao, Rakshith Sharma, Cheng Qian, Lucas M Glass, Jeffrey Spaeder, Justin Romberg, Jimeng Sun, and Cao Xiao. 2021. STAN: spatio-temporal attention network for pandemic prediction using real-world evidence. *Journal of the American Medical Informatics Association*, 28, 4, (Apr. 1, 2021), 733–743. DOI: 10.1093/jamia/ocaa322.

[9] Weiwei Jiang and Jiayun Luo. 2022. Graph neural network for traffic forecasting: a survey. *Expert Systems with Applications*, 207, (Nov. 2022), 117921. arXiv: 2101.11174[cs]. DOI: 10.1016/j.eswa.2022.117921.

[10] Amol Kapoor, Xue Ben, Luyang Liu, Bryan Perozzi, Matt Barnes, Martin Blais, and Shawn O'Banion. 2020. Examining COVID-19 forecasting using spatio-temporal graph neural networks. (July 6, 2020). arXiv: 2007.03113[cs]. DOI: 10.48550/arXiv.2007.03113.

[11] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net. Retrieved May 19, 2023 from https://openreview.net/forum?id=SJU4ayYgl.

[12] Nicole E. Kogan et al. 2021. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Science Advances*, 7, 10, (Mar. 5, 2021), eabd6989. DOI: 10.1126/sciadv.abd6989.

[13] [n. d.] Kontaktindex von NET CHECK – kontaktindex von NET CHECK. Retrieved May 25, 2023 from https://contactindex.netcheck.de/.

[14] 2021. Kreisfreie städte und landkreise nach fläche, bevölkerung und bevölkerungsdichte am 31.12.2020. de. https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/04-kreise.html. Accessed: 2021-10-25. (Sept. 2021).

[15] Benjamin F Maier, Angelique Burdinski, Marc Wiedermann, Annika H Rose, Frank Schlosser, Matthias An Der Heiden, Ole Wichmann, Thomas Harder, and Dirk Brockmann. 2023. Modeling the impact of the omicron infection wave in germany. *Biology Methods and Protocols*, 8, 1, (Jan. 10, 2023), bpad005. DOI: 10.1093/biomethods/bpad005.

[16] William Msemburi, Ariel Karlinsky, Victoria Knutson, Serge Aleshin-Guendel, Somnath Chatterji, and Jon Wakefield. 2023. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, 613, 7942, (Jan. 5, 2023), 130–137. DOI: 10.1038/s41586-022-05522-2.

[17] Anatol-Fiete Näher et al. 2023. Secondary data for global health digitalisation. *The Lancet Digital Health*, 5, 2, (Feb. 2023), e93–e101. DOI: 10.1016/S2589-7500(22)00195-9.

[18] Nuria Oliver et al. 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6, 23, (June 5, 2020), eabc0764. DOI: 10.1126/sciadv.abc0764.

[19] George Panagopoulos, Giannis Nikolentzos, and Michalis Vazirgiannis. 2021. Transfer graph neural networks for pandemic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35, 6, (May 18, 2021), 4838–4845. Number: 6. DOI: 10.1609/aaai.v35i6.16616.

[20] Adam Paszke et al. 2019. PyTorch: an imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Number 721. Curran Associates Inc., Red Hook, NY, USA, (Dec. 8, 2019), 8026–8037. Retrieved May 26, 2023 from.

[21] Patrick Reiser et al. 2022. Graph neural networks for materials science and chemistry. *Communications Materials*, 3, 1, (Nov. 26, 2022), 93. DOI: 10.1038/s43246-022-00315-6.

[22] Robert Koch-Institut. 2022. SARS-CoV-2 Infektionen in Deutschland, (Aug. 2022). DOI: 10.5281/zenodo.6994808.

[23] Sten Rüdiger, Stefan Konigorski, Alexander Rakowski, Jonathan Antonio Edelman, Detlef Zernick, Alexander Thieme, and Christoph Lippert. 2021. Predicting the SARS-CoV-2 effective reproduction number using bulk contact data from mobile phones. *Proceedings of the National Academy of Sciences of the United States of America*, 118, 31, (Aug. 3, 2021), e2026731118. DOI: 10.1073/pnas.2026731118.

[24] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling relational data with graph convolutional networks. (Oct. 26, 2017). arXiv: 1703.06103[cs,stat]. DOI: 10.48550/arXiv.1703.06103.

[25] Frank Schlosser, Benjamin F. Maier, Olivia Jack, David Hinrichs, Adrian Zachariae, and Dirk Brockmann. 2020. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences*, 117, 52, (Dec. 29, 2020), 32883–32890. DOI: 10.1073/pnas.2012326117.

[26] Lucas M. Stolerman, Leonardo Clemente, Canelle Poirier, Kris V. Parag, Atreyee Majumder, Serge Masyn, Bernd Resch, and Mauricio Santillana. 2023. Using digital traces to build prospective and real-time county-level early warning systems to anticipate COVID-19 outbreaks in the united states. *Science Advances*, 9, 3, (Jan. 20, 2023), eabq0199. DOI: 10.1126/sciadv.abq0199.

[27] Leonard J. Tashman. 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16, 4, (Oct. 2000), 437–450. DOI: 10.1016/S0169-2070(00)00065-0.

[28] Sean J. Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician*, 72, 1, (Jan. 2, 2018), 37–45. DOI: 10.1080/00031305.2017.1380080.

[29] World Health Organization. 2023. Statement on the fifteenth meeting of the IHR (2005) emergency committee on the COVID-19 pandemic. (May 2023). Retrieved May 26, 2023 from https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic.

A EVALUATION METRICS

A.1 Mean Absolute Percentage Error

The Mean Absolute Percentage Error (MAPE) [6]:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{|y_i|},$$

where n is the number of counties, is a relative error independent of the range of values. The MAPE is highly susceptible to observations close to zero causing the metric to explode. A smaller MAPE is better.

A.2 symmetric Mean Absolute Percentage Error

The symmetric Mean Absolute Percentage Error (sMAPE) [6]:

$$sMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{|\hat{y}_i| + |y_i|},$$

where n is the number of counties, is another relative metric that takes on values between 0 and 1 and is therefore protected against exploding values. A smaller sMAPE is better.

A.3 Mean Directional Accuracy

We use an adaption of the Mean Directional Accuracy (MDA) [2]. As we only forecast a single value, the MDA can be simplified, yielding the rate at which the models can identify the trend correctly:

$$MDA = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{sign(\hat{y}_i)=sign(y_i)},$$

where n is the number of counties and $\mathbf{1}$ the indicator function. A larger MDA is better.