TrendGNN: Towards Understanding of Epidemics, Beliefs, and Behaviors

Mulin Tian
Electrical and Computer Engineering
University of Southern California
Los Angeles, CA, USA
mulintia@usc.edu

Ajitesh Srivastava
Electrical and Computer Engineering
University of Southern California
Los Angeles, CA, USA
ajiteshs@usc.edu

Abstract

Epidemic outcomes have a complex interplay with human behavior and beliefs. Most of the forecasting literature has focused on the task of predicting epidemic signals using simple mechanistic models or black-box models, such as deep transformers, that ingest all available signals without offering interpretability. However, to better understand the mechanisms and predict the impact of interventions, we need the ability to forecast signals associated with beliefs and behaviors in an interpretable manner. In this work, we propose a graph-based forecasting framework that first constructs a graph of interrelated signals based on trend similarity, and then applies graph neural networks (GNNs) for prediction. This approach enables interpretable analysis by revealing which signals are more predictable and which relationships contribute most to forecasting accuracy. We believe our method provides early steps towards a framework for interpretable modeling in domains with multiple potentially interdependent signals, with implications for building future simulation models that integrate behavior, beliefs, and observations.

1 Introduction

Understanding the interplay between behaviors, beliefs, and epidemic outcomes remains a fundamental challenge, requiring comprehensive models to guide effective public health interventions. Capturing these dynamics is particularly difficult, as the signals are high-dimensional, noisy, and interdependent across regions and populations. However, many existing models rely on overly simplistic assumptions about these mechanisms ([20, 13, 4, 12, 3]). Further, existing forecasting methods [2, 15] are focused on epidemic signals and do not provide any insight into how these might influence (or predict) behaviors and beliefs, and vice versa.

In this work, we present a graph-based framework that models these relationships explicitly. We construct graphs to represent dependencies among features and states, and then apply GraphSAGE to leverage these structured relationships for multi-week forecasting tasks. Our study shows that both the use of graph neural networks and the choice of graph construction strategy are critical for improving mid- to long-term forecasting. Beyond accuracy, the graph-based perspective provides interpretability by highlighting which signals and dependencies matter most, offering a principled path toward simulation models that integrate behavior, beliefs, and epidemic outcomes.

Our contributions We introduce a graph-based forecasting framework that combines trend similarity graphs with GraphSAGE, and systematically compare different graph construction strategies to demonstrate their critical impact on forecasting accuracy across multiple signals including epidemics, behavior, and beliefs. We show that the framework provides interpretability by revealing influential signals and dependencies, offering insights that can inform simulation models for public health.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time Series for Health (TS4H).

2 Background and Related Works

Incorporating Human Behavior Existing works that combine epidemic and behavior modeling take a very simplistic view of both aspects [7, 1, 11, 10] with one of more of the following drawbacks. (i) They use a simple epidemic model (like SIR) that does not represent the complexities of real-world. (ii) They use simple behavioral models where individuals assess the cost of following interventions at an individual level. (iii) They present simulations without rigorously grounding their assumptions and results with real-world observations. We take a data-driven approach to understand how various signals are related to each other, which can then be used in mechanistic modelling.

Epidemic Forecasting Early approaches to epidemic forecasting relied on statistical and compartmental models such as ARIMA [18] and SEIR [8], which capture short-term dynamics but struggle with complex, high-dimensional signals. More recent methods apply machine learning and deep learning models, including recurrent neural networks [17] and Transformer [21], to integrate diverse behavioral and epidemiological data. While these black-box models can achieve competitive accuracy, especially for short-term horizons, they often lack interpretability and do not explicitly model relationships among signals—limitations that motivate graph-based approaches. A number of methods are used in forecast hubs [2, 5] but they only target epidemic outcomes (e.g., deaths and hospitalization) and not behavior, which is necessary to understand the effect of interventions.

Graph-based Time Series Forecasting Graph neural networks (GNNs) have recently been applied to multivariate time series forecasting in domains such as traffic, climate, and energy, where graph structures capture dependencies among signals. Methods such as GCN [9] and GAT[22] leverage spatial or relational information to improve prediction accuracy beyond purely sequential models. GraphSAGE[6], in particular, has been used to efficiently aggregate information from local neighborhoods and scale to larger graphs, making it well-suited for dynamic, high-dimensional forecasting tasks. However, existing work often assumes a fixed or externally provided graph, with limited attention to how the graph itself should be constructed from data – a key focus of our study.

Signal Similarity and Graph Construction Graph construction is crucial for forecasting performance. Traditional approaches often rely on lagged correlations, which capture delayed dependencies among signals. More recent methods, such as DTW+S (Dynamic Time Warping with Shapelets) [19], emphasize local trend similarity by aligning shapelet-based representations of time series. These strategies highlight the importance of choosing meaningful similarity measures, though their relative effectiveness in epidemic forecasting remains underexplored.

3 Method

Our goal is to build a framework that reveals dependencies across multiple signals and to show its usefulness through a forecasting task. To demonstrate this we choose a collection of a number of signals available during COVID-19.

Dataset To validate our approach, we use a subset of the COVID-19 Trends and Impact Survey (CTIS) [16]. The survey was designed to collect large-scale, timely data on the spread and impact of COVID-19 across the United States. This subset contains data from **26 U.S. states** covering a period of **402 days**, from May 20, 2020 to June 25, 2022. For each state, we include **26 COVID-19 related signals**, after removing signals that were not fully available in that period. All signals were rescaled using min–max normalization to lie within the interval [0, 1]. Each signal falls into one of the following four categories (Detailed category grouping and signal explanation are in Appendix A):

- **Health indicators** derived from self-reported symptoms of influenza-like and COVID-like illness.
- **Behavioral indicators** such as mask-wearing and social distancing.
- **Testing and vaccination indicators** based on respondent reports.
- **Demographics and beliefs** including attitudes toward COVID-19 treatment and prevention.

Architecture Figure 1 illustrates the architecture of our TrendGNN workflow. For each state, we generate a similarity matrix using DTW+S and sparsify it by retaining the top five most correlated connections for each signal. The resulting matrices are aggregated into a block-diagonal structure, with each block corresponding to one state. This block-diagonal matrix is then used to construct a unified graph encompassing all nodes across states. Finally, the constructed graph is combined with data sampled using a rolling window of size four and fed into the network for training.

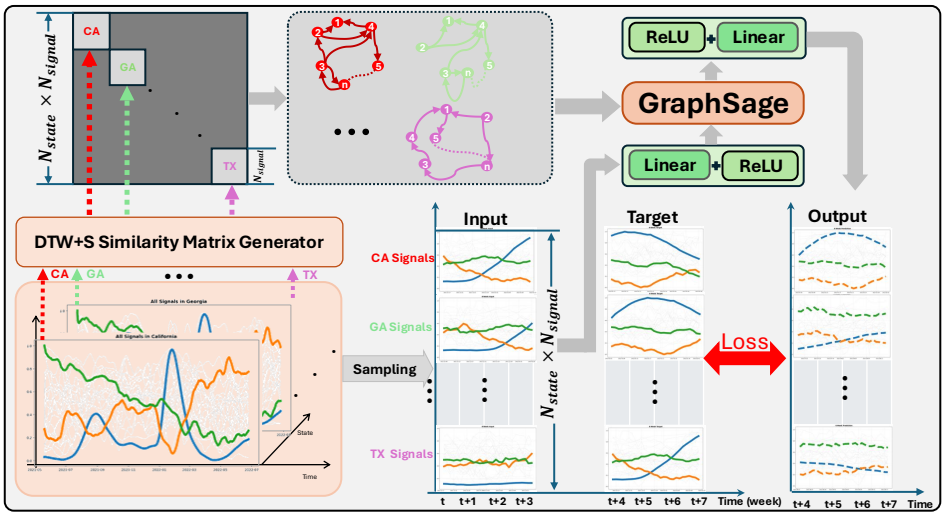

Figure 1: Overview of the TrendGNN pipeline. The dark gray block denotes a block-diagonal matrix, where each light gray sub-block on the diagonal is an $N_{\text{signal}} \times N_{\text{signal}}$ similarity matrix for one state (computed using DTW+S). Off-diagonal regions are set to zero, and stacking across N_{state} yields a final square matrix of size $(N_{\text{signal}} \times N_{\text{state}})^2$. The Input, Output, and Target in the diagram each have size $(N_{\text{state}} \times N_{\text{signal}}) \times$ window, where the window length is four weeks as shown.

Retrospective Training Setup For the training and evaluation, we adopt a progressive rolling-window strategy to split the dataset into training and testing segments – this mimics the real-world setup, where every week we are interested in forecasting the outcomes over the next weeks based on the data available so far. Specifically, τ_{start} indicates the number of initial weeks used for training. For example, if $\tau_{\text{start}} = 20$, the model is trained on the first 20 weeks of data and evaluated on the subsequent 4 weeks. As τ_{start} increases, the training window gradually expands while the test set moves forward accordingly.

4 Experiments

Our goal is to show that having a graph of temporal signals improves our ability to forecast all the signals. Specifically, we wish to demonstrate that (1) graph-based methods are better (particularly for longer horizons), and (2) having a "good" graph of signals is advantageous. We chose forecasting of a four-week horizon, which is typical in epidemic forecasting [2].

Baselines We used several baselines to compare our approach against. Consistent with our approach, all models take the past four weeks of data as input and predict the subsequent four weeks. In addition, to reduce the influence of randomness, each model was trained five times, and the final results are presented as the mean +/- standard deviation.

Non-GNN Methods. The following baselines do not use the graph obtained form the signals: (i) *Flat Baseline:* directly uses the most recent week's values as predictions for each task; (ii)*ARIMA:* a traditional statistical model widely used in time-series forecasting; and (iii) *Transformer:* a modern deep learning approach that captures long-range temporal dependencies.

GNN Methods. To identify the impact of the graph, we construct the graphs with the following strategies: (i) *Random Graph:* for each state, we generate a directed graph among its 26 signals by randomly assigning edges, ensuring that each node has on average five outgoing edges; (ii)*Fully Connected Graph:* for each state, we construct a fully connected directed graph among the 26 signals. (iii) *Lagged Correlation Graph:* similar to our approach except that DTW+S is replaced by lagged correlation; and (iv) *DTW+S Graph:* our proposed approach

Results Our experiments demonstrate that graph-based models consistently outperform both traditional models and simple baselines, particularly for medium- to long-term horizons (2–4 weeks)

Table 1: This table reports the prediction errors of different models on 1–4 week-ahead forecasting tasks, measured by Mean Absolute Error (MAE). **B**, **DB**, **H**, and **TV** are abbreviations for the four categories introduced in 3. **AVG** denotes the average error across these four categories. For each column, the worst-performing value is highlighted in red, while the best and second-best values are marked in bold and underlined, respectively.

MAE	1-Week-Ahead					2-Week-Ahead				
	B	DB	H	TV	AVG	B	DB	H	TV	AVG
Baseline	**0.0729**	0.1148	0.0501	0.1108	0.0872	0.1082	0.1376	0.0898	0.1306	0.1166
ARIMA	0.0741	0.1059	0.0504	0.1046	**0.0837**	0.1098	0.1209	**0.0894**	0.1220	0.1105
Sage (Random)	0.0816	0.1064	0.0620	0.1021	0.0880	0.1117	0.1131	0.1006	0.1084	0.1084
Transformer	0.0780	0.1108	0.0629	0.1069	0.0896	0.1044	0.1195	0.0912	0.1148	0.1075
Sage (Full)	0.0837	**0.1055**	0.0646	0.1013	0.0888	0.1118	**0.1125**	0.1014	0.1080	0.1084
Sage (Lagged)	0.0792	0.1067	0.0609	**0.1011**	0.0870	0.1035	0.1127	0.0975	**0.1066**	0.1051
Sage (DTW+S)	0.0785	0.1064	0.0622	0.1013	0.0871	**0.1026**	**0.1125**	0.0966	0.1067	**0.1046**
MAE	3-Week-Ahead					4-Week-Ahead				
	B	DB	H	TV	AVG	B	DB	H	TV	AVG
Baseline	0.1339	0.1417	0.1250	0.1343	0.1338	0.1606	0.1434	0.1553	0.1367	0.1490
ARIMA	0.1370	0.1279	0.1233	0.1299	0.1295	0.1624	0.1333	0.1516	0.1365	0.1459
Transformer	0.1263	0.1200	0.1168	0.1148	0.1195	0.1537	0.1249	0.1414	0.1212	0.1353
Sage (Random)	0.1239	0.1128	0.1212	0.1070	0.1162	0.1315	0.1139	0.1347	0.1077	0.1219
Sage (Full)	0.1225	0.1124	0.1203	0.1072	0.1156	0.1277	0.1132	0.1330	0.1072	0.1203
Sage (Lagged)	**0.1126**	**0.1115**	0.1162	0.1055	0.1115	0.1217	0.1135	0.1294	0.1069	0.1179
Sage (DTW+S)	**0.1126**	0.1117	**0.1160**	**0.1050**	**0.1113**	**0.1194**	**0.1129**	**0.1276**	**0.1059**	**0.1165**

in Table 1. Among graph constructions, semantically meaningful approaches such as DTW+S and lagged correlation yield clear gains over random or fully connected graphs. While ARIMA performs slightly better for the 1-week-ahead case, the best and second-best results across 2–4 week forecasts are consistently obtained with DTW+S and lagged-correlation-based graphs. DTW+S is especially effective, as it captures temporal misalignments and non-linear similarities, which are common in COVID-related indicators, while lagged correlation is limited to linear dependencies with fixed shifts. In addition, we report MAE values with standard deviations aggregated by signal categories, and include the corresponding table in the appendix for reference.

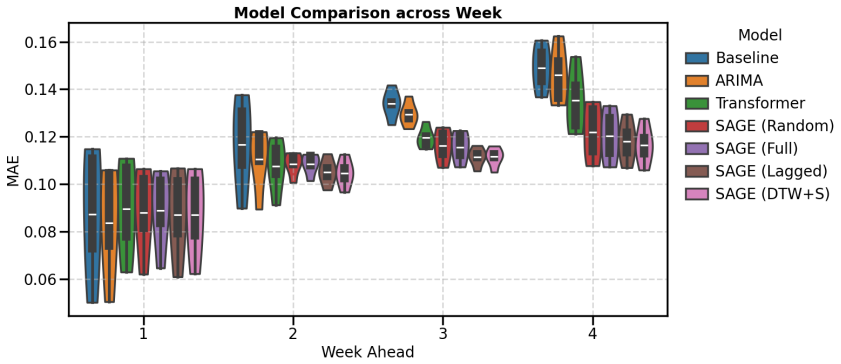

Figure 2: Mean absolute error (MAE) distributions of different models across 1–4 week-ahead forecasting tasks.

5 Graph Explanation

We envision that our framework will assist with designing models that better capture the underlying mechanisms. As an early step, it is important to identify which relationships are critical in driving the behavior of the signals. While we used DTW+S to select top relationships based on trends, we use the CF-GNNExplainer [14] framework on our forecasting model to further identify which signals are critical in impacting the predictive behavior of the GNNs. Unlike the original formulation, which identifies the minimal edge changes that flip a node classification, our variant seeks the smallest perturbations that brings the MAE at or below the MAE using random graph. For each signal, we applied CF-GNNExplainer [14] across all states and weeks. We then counted how often the most

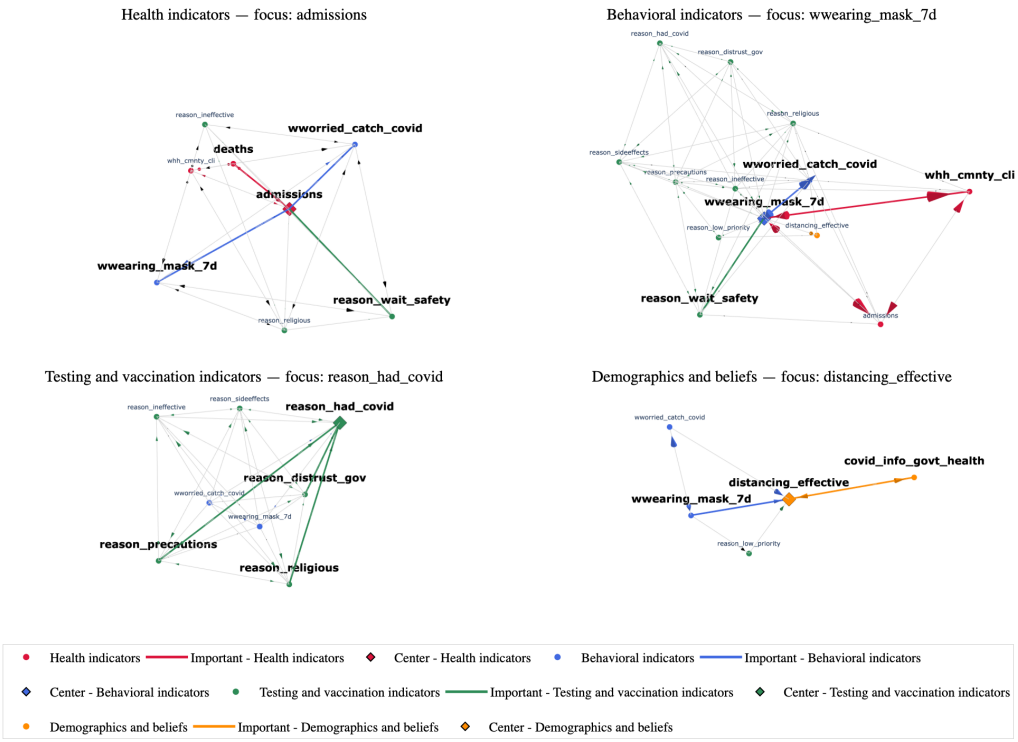

Figure 3: 3D visualization of signal graphs across four categories of indicators. Each subplot corresponds to one category. Nodes are colored by category, with the center node (focus signal) highlighted. Thick edges denote important connections.

important edges appeared for that signal and selected the relatively most critical edges as the final results. In Figure 3, for each category we chose one representative signal and visualized a subgraph from the DTW+S-constructed graph together with its most important edge.

Health indicators, *hospitalization* were most influenced by mortality outcomes, perceptions of vaccine safety, and worries about catching COVID-19, showing how both epidemiological severity and behavioral attitudes influence admissions.

Behavioral indicators, *mask-wearing* was strongly shaped by safety concerns, local illness reports, and infection worries, highlighting its dual role as both a preventive action and a proxy for perceived community risk.

Testing and vaccination indicators, *beliefs about not needing a vaccine due to prior infection* were primarily driven by religious reasons, distrust of government, and alternative precautionary attitudes, revealing how prior experience and beliefs drive hesitancy.

Demographics and beliefs, perceived *effectiveness of distancing* was largely influenced by protective behaviors and trust in government information, illustrating the interplay between institutional trust and adherence to preventive measures.

6 Conclusion

Beyond accuracy, the graph view highlights which signals are more predictable and reveals the dependencies that drive improvements, offering a more interpretable perspective compared to black-box models. These findings underline the importance of constructing meaningful, temporally aligned graphs rather than relying on arbitrary connectivity. Future directions include incorporating lag-aware causal structures to disentangle true drivers from correlations, as well as enhancing edge- and subgraph-level interpretability to support policy decisions and simulation modeling of behavioral and epidemic dynamics.

Acknowledgements

This work is supported by the Centers for Disease Control and Prevention's (CDC) Center for Forecasting and Outbreak Analytics (CFA) through the Insight Net award.

References

[1] Folashade B Agusto, Igor V Erovenko, Alexander Fulk, Qays Abu-Saymeh, Daniel Romero-Alvarez, Joan Ponce, Suzanne Sindi, Omayra Ortega, Jarron M Saint Onge, and A Townsend Peterson. To isolate or not to isolate: The impact of changing behavior on covid-19 transmission. *BMC Public Health*, 22(1):1–20, 2022.

[2] Johannes Bracher, Evan L Ray, Tilmann Gneiting, and Nicholas G Reich. Evaluating epidemic forecasts in an interval format. *PLoS computational biology*, 17(2):e1008618, 2021.

[3] Jiangzhuo Chen, Anil Vullikanti, Joost Santos, Srinivasan Venkatramanan, Stefan Hoops, Henning Mortveit, Bryan Lewis, Wen You, Stephen Eubank, Madhav Marathe, et al. Epidemiological and economic impact of covid-19 in the us. *Scientific reports*, 11(1):20451, 2021.

[4] Jessica T Davis, Matteo Chinazzi, Nicola Perra, Kunpeng Mu, Ana Pastore y Piontti, Marco Ajelli, Natalie E Dean, Corrado Gioannini, Maria Litvinova, Stefano Merler, et al. Cryptic transmission of sars-cov-2 and the first covid-19 wave. *Nature*, 600(7887):127–132, 2021.

[5] Centers for Disease Control, Prevention, et al. Flusight: flu forecasting, 2019.

[6] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018.

[7] KM Ariful Kabir and Jun Tanimoto. Evolutionary game theory modelling to represent the behavioural dynamics of economic shutdowns and shield immunity in the covid-19 pandemic. *Royal Society open science*, 7(9):201095, 2020.

[8] William Ogilvy Kermack and Anderson Gray McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 138(834):55–83, 1932.

[9] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[10] Ioannis Kordonis, Athanasios-Rafail Lagos, and George P Papavassilopoulos. Dynamic games of social distancing during an epidemic: Analysis of asymmetric solutions. *Dynamic Games and Applications*, pages 1–23, 2022.

[11] CAK Kwuimy, Foad Nazari, Xun Jiao, Pejman Rohani, and C Nataraj. Nonlinear dynamic analysis of an epidemiological model for covid-19 including public behavior and government action. *Nonlinear Dynamics*, 101:1545–1559, 2020.

[12] Joseph C Lemaitre, Kyra H Grantz, Joshua Kaminsky, Hannah R Meredith, Shaun A Truelove, Stephen A Lauer, Lindsay T Keegan, Sam Shah, Josh Wills, Kathryn Kaminsky, et al. A scenario modeling pipeline for covid-19 emergency planning. *Scientific reports*, 11(1):7534, 2021.

[13] Benjamin P Linas, Jade Xiao, Ozden O Dalgic, Peter P Mueller, Madeline Adee, Alec Aaron, Turgay Ayer, and Jagpreet Chhatwal. Projecting covid-19 mortality as states relax nonpharmacologic interventions. In *JAMA Health Forum*, volume 3, pages e220760–e220760. American Medical Association, 2022.

[14] Ana Lucic, Maartje ter Hoeve, Gabriele Tolomei, Maarten de Rijke, and Fabrizio Silvestri. Cf-gnnexplainer: Counterfactual explanations for graph neural networks, 2022.

[15] Evan L Ray, Nutcha Wattanachit, Jarad Niemi, Abdul Hannan Kanji, Katie House, Estee Y Cramer, Johannes Bracher, Andrew Zheng, Teresa K Yamana, Xinyue Xiong, and others. Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the US. *medRxiv*, 2020. Publisher: Cold Spring Harbor Laboratory Press.

[16] Joshua A. Salomon, Alex Reinhart, Alyssa Bilinski, Eu Jing Chua, Wichada La Motte-Kerr, Minttu M. Rönn, Marissa B. Reitsma, Katherine A. Morris, Sarah LaRocca, Tamer H. Farag, Frauke Kreuter, Roni Rosenfeld, and Ryan J. Tibshirani. The us covid-19 trends and impact survey: Continuous real-time measurement of covid-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proceedings of the National Academy of Sciences*, 118(51):e2111454118, 2021.

[17] Robin M. Schmidt. Recurrent neural networks (rnns): A gentle introduction and overview, 2019.

[18] Sima Siami-Namini and Akbar Siami Namin. Forecasting economics and financial time series: Arima vs. lstm, 2018.

[19] Ajitesh Srivastava. Aligning time-series by local trends: Applications in public health. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27):28405–28412, Apr. 2025.

[20] Brad Suchoski, Steve Stage, Heidi Gurung, and Prasith Baccam. Gpu accelerated parallel processing for large-scale monte carlo analysis: Covid-19 parameter estimation and new case forecasting. *Frontiers in Applied Mathematics and Statistics*, 8:818016, 2022.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[22] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.

A Explanation

A.1 Dataset Explanation

1. **new_time_value:** Date in format (Year-Month-Day).
2. **smoothed_waccept_covid_vaccine_no_appointment:** Estimated percentage of respondents who would definitely or probably choose to get vaccinated, if a vaccine were offered to them today, among respondents who have not yet been vaccinated and do not have an appointment to do so.
3. **smoothed_wbelief_created_small_group:** Estimated percentage of people who believe that the statement "COVID-19 was deliberately created by a small group of people who secretly manipulate world events" is definitely or probably true.
4. **smoothed_wbelief_distancing_effective:** Estimated percentage of respondents who believe that social distancing is either very or moderately effective for preventing the spread of COVID-19.
5. **smoothed_wbelief_govt_exploitation:** Estimated percentage of people who indicate that the statement "The COVID-19 pandemic is being exploited by the government to control people" is definitely or probably true.
6. **smoothed_wdelayed_care_cost:** Estimated percentage of respondents who have ever delayed or not sought medical care in the past year because of cost.
7. **smoothed_wdontneed_reason_dont_spend_time:** Estimated percentage of respondents who say they don't need to get a COVID-19 vaccine because they don't spend time with high-risk people, among respondents who provided at least one reason for why they believe a COVID-19 vaccine is unnecessary.
8. **smoothed_wdontneed_reason_had_covid:** Estimated percentage of respondents who say they don't need to get a COVID-19 vaccine because they already had the illness, among respondents who provided at least one reason for why they believe a COVID-19 vaccine is unnecessary.
9. **smoothed_wdontneed_reason_precautions:** Estimated percentage of respondents who say they don't need to get a COVID-19 vaccine because they will use other precautions, such as a mask, instead, among respondents who provided at least one reason for why they believe a COVID-19 vaccine is unnecessary.
10. **smoothed_whesitancy_reason_cost:** Estimated percentage of respondents who say they are hesitant to get vaccinated because they are worried about the cost, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not" when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.
11. **smoothed_whesitancy_reason_distrust_gov:** Estimated percentage of respondents who say they are hesitant to get vaccinated because they don't trust the government, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not" when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.
12. **smoothed_whesitancy_reason_ineffective:** Estimated percentage of respondents who say they are hesitant to get vaccinated because they don't know if a COVID-19 vaccine will work, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not" when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.
13. **smoothed_whesitancy_reason_low_priority:** Estimated percentage of respondents who say they are hesitant to get vaccinated because they think other people need it more than they do, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not" when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.
14. **smoothed_whesitancy_reason_religious:** Estimated percentage of respondents who say they are hesitant to get vaccinated because it is against their religious beliefs, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not"

when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.

15. **smoothed_whesitancy_reason_sideeffects:** Estimated percentage of respondents who say they are hesitant to get vaccinated because they are worried about side effects, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not" when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.
16. **smoothed_whesitancy_reason_wait_safety:** Estimated percentage of respondents who say they are hesitant to get vaccinated because they want to wait to see if the COVID-19 vaccines are safe, among respondents who answered "Yes, probably", "No, probably not", or "No, definitely not" when asked if they would get vaccinated if offered (item V3). This series of items was shown to respondents starting in Wave 8.
17. **smoothed_whh_cmnty_cli:** Estimated percentage of people reporting illness in their local community, as described in this link, including their household.
18. **smoothed_wrace_treated_fairly_healthcare:** Estimated percentage of respondents who somewhat or strongly agree that people of their race are treated fairly in a healthcare setting.
19. **smoothed_wspent_time_indoors_1d:** Estimated percentage of respondents who "spent time indoors with someone who isn't currently staying with you" in the past 24 hours.
20. **smoothed_wtrust_covid_info_friends:** Estimated percentage of respondents who trust friends and family to provide accurate news and information about COVID-19.
21. **smoothed_wtrust_covid_info_govt_health:** Estimated percentage of respondents who trust government health officials to provide accurate news and information about COVID-19.
22. **smoothed_wtrust_covid_info_religious:** Estimated percentage of respondents who trust religious leaders to provide accurate news and information about COVID-19.
23. **smoothed_wwearing_mask_7d:** Estimated percentage of people who wore a mask for most or all of the time while in public in the past 7 days; those not in public in the past 7 days are not counted.
24. **smoothed_wworried_catch_covid:** Estimated percentage of respondents worrying either a great deal or a moderate amount about catching COVID-19.
25. **deaths_incidence_num:** Number of new confirmed deaths due to COVID-19, daily.
26. **confirmed_admissions_covid_1d_7dav:** Sum of adult and pediatric confirmed COVID-19 hospital admissions occurring each day.
27. **smoothed_wpublic_transit_1d:** Estimated percentage of respondents who "used public transit" in the past 24 hours.

A.2 Signal Grouping

The 26 signal in our dataset can be grouped into four major categories:

Health indicators. `whh_cmnty_cli`, `deaths`, `admissions`.

Behavioral indicators. `wwearing_mask_7d`, `wspent_time_indoors_1d`, `wpublic_transit`, `wworried_catch_covid`.

Testing and vaccination indicators. `vaccine_no_appointment`, `reason_dont_spend_time`, `reason_had_covid`, `reason_precautions`, `reason_cost`, `reason_distrust_gov`, `reason_ineffective`, `reason_low_priority`, `reason_religious`, `reason_sideeffects`, `reason_wait_safety`.

Demographics and beliefs. `created_small_group`, `distancing_effective`, `govt_exploitation`, `care_cost`, `treated_fairly_healthcare`, `covid_info_friends`, `covid_info_govt_health`, `covid_info_religious`.

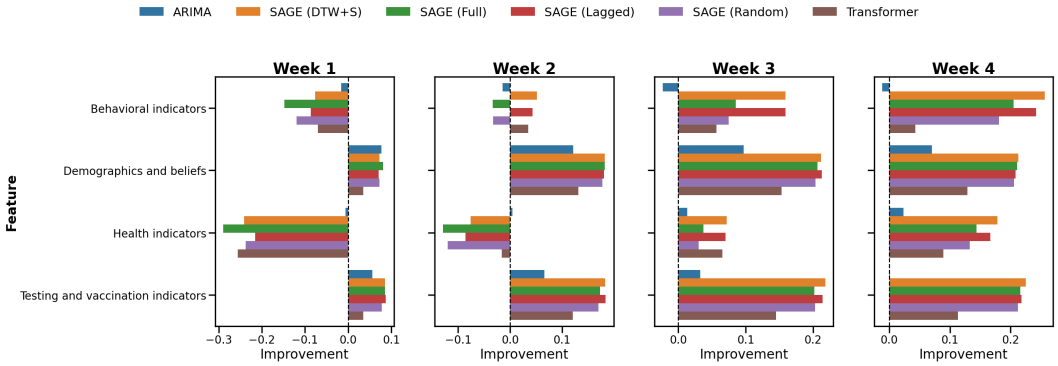

Figure 4: Distribution of relative improvements in MAE for different models over the baseline across 1–4 week-ahead forecasting tasks. The x-axis shows relative improvement, the y-axis shows signal categories, and colors denote models. Positive values indicate gains over the baseline, negative values indicate degradation.

Table 2: Category-level forecasting performance. Mean Absolute Error (MAE) ± standard deviation, aggregated over signals within each category. The worst-performing models are highlighted in red, while the best and second-best results are indicated by **bold** and underline, respectively.

MAE	1-Week-Ahead					2-Week-Ahead				
	B	DB	H	TV	AVG	B	DB	H	TV	AVG
Baseline	**0.0729 ± NA**	0.1148 ± NA	0.0501 ± NA	0.1108 ± NA	0.0872 ± NA	0.1082 ± NA	0.1376 ± NA	0.0898 ± NA	0.1306 ± NA	0.1166 ± NA
ARIMA	0.0741 ±NA	0.1059 ±NA	0.0504 ±NA	0.1046 ± NA	**0.0837 ±NA**	0.1098 ±NA	0.1209 ±NA	**0.0894 ±NA**	0.1220 NA	0.1105 ± NA
Sage (Random)	0.0816 ± 0.0013	0.1064 ± 0.0004	0.0620 ± 0.0015	0.1021 ± 0.0005	0.0880 ± 0.0006	0.1117 ± 0.0030	0.1131 ± 0.0006	0.1006 ± 0.0010	0.1084 ± 0.0004	0.1084 ± 0.0010
Transformer	0.0780 ± 0.0009	0.1108 ± 0.0005	0.0629 ± 0.0001	0.1069 ± 0.0005	0.0896 ±0.0004	0.1044 ± 0.0013	0.1195 ± 0.0009	0.0912 ± 0.0006	0.1148 ± 0.0005	0.1075 ± 0.0005
Sage (Full)	0.0837 ±0.0013	**0.1055 ± 0.0005**	0.0646 ±0.0021	0.1013 ± 0.0005	0.0888 ± 0.0009	0.1118 ±0.0016	**0.1125 ± 0.0006**	0.1014 ± 0.0013	0.1080 ± 0.0004	0.1084 ± 0.0001
Sage (Lagged)	0.0792 ± 0.0005	0.1067 ± 0.0006	0.0609 ± 0.0022	**0.1011 ± 0.0006**	0.0870 ± 0.0008	0.1035 ± 0.0019	0.1127 ± 0.0003	0.0975 ± 0.0031	**0.1066 ± 0.0002**	0.1051 ± 0.0008
Sage (DTW+S)	0.0785 ± 0.0019	0.1064 ± 0.0003	0.0622 ± 0.0017	0.1013 ± 0.0007	0.0871 ± 0.0007	**0.1026 ± 0.0025**	**0.1125 ± 0.0011**	0.0966 ± 0.0029	0.1067 ± 0.0008	**0.1046 ± 0.0004**
MAE	3-Week-Ahead					4-Week-Ahead				
	B	DB	H	TV	AVG	B	DB	H	TV	AVG
Baseline	0.1339 ± NA	0.1417 ± NA	0.1250 ±NA	0.1343 ±NA	0.1338 ± NA	0.1606 ± NA	0.1434 ± NA	0.1553 ± NA	0.1367 ± NA	0.1490 ± NA
ARIMA	0.1370 ± NA	0.1279 ± NA	0.1233 ± NA	0.1299 ± NA	0.1295 ± NA	0.1624 ± NA	0.1333 ± NA	0.1516 ± NA	0.1365 ± NA	0.1459 ± NA
Transformer	0.1263 ± 0.0012	0.1200 ± 0.0008	0.1168 ± 0.0035	0.1148 ± 0.0002	0.1195 ± 0.0013	0.1537 ± 0.0025	0.1249 ± 0.0007	0.1414 ± 0.0024	0.1212 ± 0.0009	0.1353 ± 0.0012
Sage (Random)	0.1239 ± 0.0050	0.1128 ± 0.0007	0.1212 ± 0.0021	0.1070 ± 0.0005	0.1162 ± 0.0015	0.1315 ± 0.0039	0.1139 ± 0.0007	0.1347 ± 0.0013	0.1077 ± 0.0012	0.1219 ± 0.0008
Sage (Full)	0.1225 ± 0.0039	0.1124 ± 0.0009	0.1203 ± 0.0017	0.1072 ± 0.0005	0.1156 ± 0.0011	0.1277 ± 0.0026	0.1132 ± 0.0007	0.1330 ± 0.0034	0.1072 ± 0.0005	0.1203 ± 0.0015
Sage (Lagged)	**0.1126 ± 0.0030**	**0.1115 ± 0.0007**	0.1162 ± 0.0030	0.1055 ± 0.0006	0.1115 ± 0.0012	0.1217 ± 0.0023	0.1135 ± 0.0010	0.1294 ± 0.0021	0.1069 ± 0.0007	0.1179 ± 0.0004
Sage (DTW+S)	**0.1126 ± 0.0033**	0.1117 ± 0.0008	**0.1160 ± 0.0022**	**0.1050 ± 0.0007**	**0.1113 ± 0.0007**	**0.1194 ± 0.0044**	**0.1129 ± 0.0006**	**0.1276 ± 0.0048**	**0.1059 ± 0.0008**	**0.1165 ± 0.0017**