9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Mobility Networks and Pandemic Analytics: How much can we gain by using complex network models?

Keywords: Complex Networks, Geospatial Modeling, Behavioral Response, Mobility Data, Pandemic Modeling

Extended Abstract

Social interaction networks play a crucial role in disease transmission and evaluating pandemic policies [2]. However, obtaining accurate data on physical contacts in a population is challenging. To estimate interaction patterns, researchers often use alternative data sources like surveys or co-location data from mobile devices. Mobility flow data between meta-populations has been used as a proxy measure for social interaction in large-scale empirical studies [1]. Despite this, the conditions and mechanisms by which mobility flow data can enhance pandemic prediction, particularly in network models, are not well understood.

This research aims to investigate the benefits of incorporating mobility networks into pandemic modeling and to identify when what, and how much they can improve pandemic prediction. We examine the role of mobility networks compared to simple mobility intensity and use changes in mobility measures as proxies for changes in behavioral responses. Specifically, we build dynamic, town-level mobility networks in Massachusetts (MA) using mobility data from the first 15 months of the COVID-19 pandemic to answer the following questions: (1) What network parameters are the most predictive? (2) How much improvement in predictive power can be achieved under varying conditions? (3) When during the pandemic do network models offer the most benefits?

The computational process is presented in a three-step diagram in Fig.1. Initially, we utilize block-level mobility data from Safegraph Co. to produce directed, weighted networks, which are then transformed into directed Town-level mobility networks. In the second step, we create our panel data by including network measures, especially different centrality measures, and town-level mobility inflow/outflow of each town, both captured by the town-level network, along with COVID-confirmed cases, provided by the MA COVID-19 Reporting. In the third step, we determine our model specification by selecting a limited number of features and appropriate interaction terms to input into regression models, which are used to predict the number of confirmed cases in each town in MA. The *adjusted* R^2 is our primary metric for assessing the prediction model's performance. With the addition of the mobility data, We will evaluate the improvement of the prediction model in two different dimensions: granularity (i.e., local versus holistic) and time (i.e., multiple prediction windows or different phases/waves of the pandemic).

We examined the advantages of mobility flow networks by distinguishing between local and systemic (also known as holistic) measures. In essence, local measures primarily rely on node degree (in-degree/out-degree/weighted degree) instead of other more systemic measures, such as various centrality measures for weighted directional networks (betweenness, eigenvector), to evaluate how a town is affected by changes in the mobility flow between pairs of other towns. Our analysis reveals that among all network measures, betweenness centrality is the most predictive mobility parameter across the entire pandemic period. This is not surprising since this parameter shows how central a given town is for connecting other towns together (think of a hub airport and its centrality with respect to other airports). Consequently, our studies also show that towns with higher betweenness centrality have higher uptake of cases when the state-level cases increase (hence the interaction term). Moreover, the predictive power of these mobility parameters is sustained and can extend the forecasting window by a few weeks (up to 4), as Fig.2 shows.

Our next objective is to investigate the impact of network measures during various stages of the COVID-19 pandemic, particularly when there are significant behavioral responses during a specific phase. The behavioral response becomes especially important when we have partial observability for our disease data. This is because full observability of the states of node and edge at every time step degenerates the network information to only local (neighborhood-level) data, and results in a spatial similar to Markov assumption. However, in reality, we often have partial observability, particularly for attributes like the number of confirmed cases, due to delays in testing and infection. We thus hypothesize that the significance of systemic measures grows as behavioral responses to the pandemic intensify. To assess this hypothesis, we analyzed data from multiple pandemic phases, such as early versus later waves, and steady versus fluctuating periods (e.g. the ramp-up vs. ramp-down of each wave). The result is in support of our hypothesis and shows that the network parameter has more improvement in the predictive power in the first wave than in the second wave (Tab: 1). One of the factors is the rapid changes in the mobile network due to the intensified public behavioral response to the pandemic during the first wave. This response led to a significant modification in the public's mobility pattern compared to the pre-pandemic period. This mobility pattern change not only affected the amount of mobility but also altered the structure of the mobility network, resulting in higher fluctuation in the network index of towns during the first wave. This factor could explain part of the variation in the confirmed cases during this period. In contrast, during the second wave, the adjustment of the network structure was lower due to the moderate public behavioral response. Because of this, we can assume that the change in the mobility pattern was already reflected in the simple in-flow/out-flow of mobility data for each town, rather than the mobility network structure, and further inclusion of centrality measures does not provide significant extra benefit. Additionally, a higher improvement in prediction accuracy was observed during the fluctuating periods compared to the steady phase (see Table 2).

To summarize, our research findings indicate that incorporating mobility measures can significantly enhance the accuracy of regression models and account for an additional 75% of the unexplained variation in case numbers. Even a single mobility measure, coupled with appropriate interaction terms in a panel-regression model, can considerably improve the predictive power. Moreover, when significant behavioral responses occur, incorporating more holistic network measures can further increase the model's predictive ability. Mobility parameters can also increase the forecasting window by up to four weeks, particularly during the initial pandemic phase, with this effect diminishing as people adapt to safer mobility behaviors (behavioral learning) and as intervention policies change.

References

[1] Alberto Aleta et al. "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19". In: *Nature Human Behaviour* 4.9 (2020), pp. 964–971.

[2] Qingtao Cao and Babak Heydari. "Micro-level social structures and the success of COVID-19 national policies". In: *Nature Computational Science* 2.9 (2022), pp. 595–604.

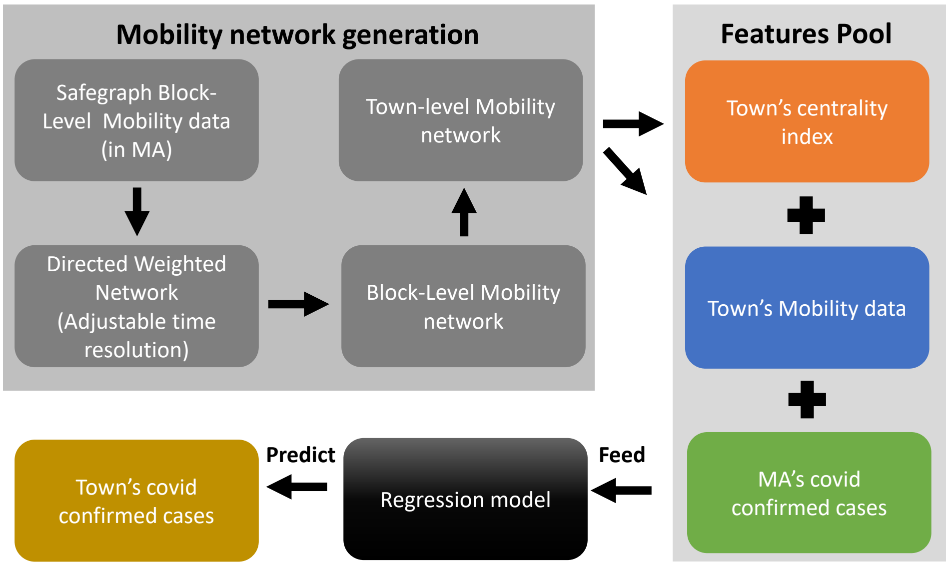

Figure 1: The diagram of the network generation and the regression model.

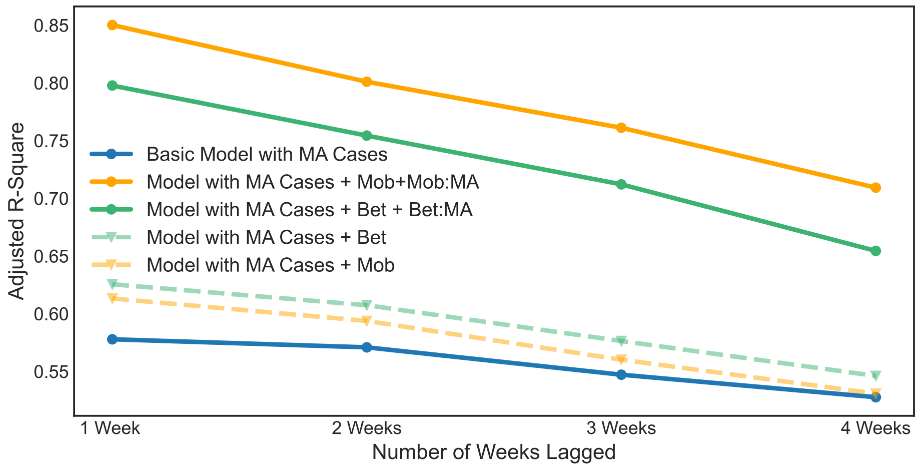

Figure 2: The trajectory of the adjusted R-square of the basic model and developed model with interaction term in different lagged weeks.

Table 1:

hline	1st wave Basic Model	1st waves mobility:MA	1st wave Bet_wei:MA	1st wave Combined	2nd wave Basic Model	2nd waves mobility:MA	2nd wave Bet_wei:MA	2nd wave Combined
MA	0.305***	-0.246***	0.117***	-0.023	0.264***	-0.045***	0.152***	0.007
	(0.061)	(0.055)	(0.024)	(0.065)	(0.035)	(0.009)	(0.016)	(0.013)
inweightt		-0.008		0.002		-0.035***		-0.027***
		(0.007)		(0.011)		(0.005)		(0.005)
bet_wei			-440.730	-258.214			-1374.337***	-152.404
			(496.339)	(429.348)			(296.176)	(418.564)
inweightt:MA		0.000***		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)		(0.000)
bet_wei:MA			15.571***	11.182***			8.764***	3.107***
			(1.661)	(3.810)			(0.580)	(0.677)
R-squared	0.509	0.813	0.848	0.858	0.805	0.899	0.890	0.902
R-squared Adj.	0.446	0.789	0.829	0.840	0.793	0.893	0.883	0.896
Observation	2754	2754	2754	2754	5508	5508	5508	5508
df_model	313	315	315	317	313	315	315	317

Table 2:

	Two waves Basic Model	Two waves inweightt:MA	Two waves bet:MA	Two waves Combined	Steady phase Basic Model	Steady phase mobility:MA	Steady phase bet:MA	Steady phase Combined
MA	0.299***	-0.075***	0.173***	-0.002	0.438***	-0.279***	0.209***	-0.125**
	(0.037)	(0.011)	(0.018)	(0.017)	(0.071)	(0.087)	(0.053)	(0.056)
inweightt		-0.018***		-0.011**		-0.006***		-0.003***
		(0.004)		(0.005)		(0.001)		(0.001)
bet_wei			-1960.407***	-929.734***			-674.592***	-440.298***
			(194.076)	(266.103)			(211.484)	(150.857)
inweightt:MA		0.000***		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)		(0.000)
bet_wei:MA			10.230***	4.131***			20.604***	9.093***
			(0.793)	(0.768)			(3.794)	(2.435)
R-squared	0.673	0.863	0.848	0.871	0.742	0.868	0.859	0.877
R-squared Adj.	0.661	0.858	0.841	0.866	0.720	0.856	0.847	0.866
Observation	8263	8263	8263	8263	4006	4006	4006	4006
df_model	313	315	315	317	313	315	315	317