Computational Economics
https://doi.org/10.1007/s10614-022-10278-y

Spatial Interactions and the Spread of COVID-19: A Network Perspective

Cui Zhang[1] · Dandan Zhang[1]

Accepted: 16 April 2022

Abstract
We use unique data on the travel history of confirmed patients at a daily frequency across 31 provinces in China to study how spatial interactions influence the geographic spread of pandemic COVID-19. We develop and simultaneously estimate a structural model of dynamic disease transmission network formation and spatial interaction. This allows us to understand what externalities the disease risk associated with a single place may create for the entire country. We find a positive and significant spatial interaction effect that strongly influences the duration and severity of pandemic COVID-19. And there exists heterogeneity in this interaction effect: the spatial spillover effect from the source province is significantly higher than from other provinces. Further counterfactual policy analysis shows that targeting the key province can improve the effectiveness of policy interventions for containing the geographic spread of pandemic COVID-19, and the effect of such targeted policy decreases with an increase in the time of delay.

Keywords Spatial interactions · Spillovers · Social network · Externalities · Pandemic COVID-19

1 Introduction

The phenomenon of the COVID-19 pandemic and its causes have attracted a great deal of research attention. Previous research has examined the reproduction rate of the COVID-19 pandemic (Goenka et al., 2021; Guimarães, 2021; Sherwani et al., 2021; Wu et al., 2020), mortality caused by the COVID-19 pandemic (Aslam et al., 2021a, b; Jha et al., 2022; Konstantinoudis et al., 2022; Kung et al., 2021; Li et al., 2020), and the socioeconomic determinants of the spread of pandemic COVID-19

✉ Cui Zhang
zzzhangcui@126.com

[1] School of Economics, Jinan University, Guangzhou, People's Republic of China

Published online: 06 July 2022

(Basu & Sen, 2022; Desbordes, 2021; Fang et al., 2020; Qiu et al., 2020). However, the existing studies on the spatial interactions in pandemic spatial transmission are extremely limited. As emphasized by Brockmann and Helbing (2013) and Jackson et al. (2017), the spread of epidemics are complex, network-driven dynamic processes, which has inherent spatial interaction patterns that shape the impact of policies. Abstracting from spatial interactions comes with the risk of severely misunderstanding pandemic diffusion and its causes.

The purpose of our paper is to fill this research gap by integrating the network perspective into the pandemic field and empirically study how spatial interactions influence the geographic spread of the COVID-19 pandemic. We begin our analysis by constructing a network representation of disease spatial transmission system based on unique data on the travel history of confirmed patients at a daily frequency across 31 provinces in China, which allows us to track the spatial spread of pandemic COVID-19. Then we apply recent methods from social network econometrics to empirically assess the role of spatial interaction in the spread of pandemic COVID-19. A major econometric challenge in this estimation framework is to overcome the endogeneity of the disease transmission network. We address this fundamental identification problem by developing and simultaneously estimating a structural model of dynamic disease transmission network formation and spatial interaction using a Bayesian approach, which is a sophisticated approach to deal with the endogeneity of network and selection bias (Goldsmith-Pinkham & Imbens, 2013; Han et al., 2021; Hsieh & Lee, 2016; Patacchini et al., 2017).

We find a positive and significant spatial interaction effect that drives the interdependence of regional new confirmed cases of the COVID-19 pandemic. The existence of such spatial interaction effect implies that adopting coordinated policies across regions is crucial to slowing the spread of the pandemic, which sheds light on the design of containment schemes and mitigation policies by multilateral organizations. Moreover, there exists heterogeneity in this interaction effect. The spatial spillover effect from the source (Hubei Province) is significantly higher than from other provinces, which suggests the dominant status of the source region in the spread of the pandemic. In terms of the disease transmission network formation, we find that the movement of infected people is more likely to occur between adjacent provinces, provinces with different observed characteristics, and provinces that were connected to each other in the previous period. Provinces with more new cases in the previous period would receive fewer infected people from other provinces in the current period.

Our paper contributes to the literature on the COVID-19 pandemic in three ways. First, we exploit a new source of data, which reports not only the number of confirmed COVID-19 patients, but also their detailed travel history at a daily frequency across geographical locations in China. The use of travel history is novel in the pandemic literature, which has typically used the (inverse) spatial distance (e.g., Qiu et al., 2020) and the flow of travelers by road or by air (e.g., Fang et al., 2020; Tatem & Smith, 2010) as measures of disease connections between places. However, epidemiologists have long known infected individuals contribute to the spread of pandemics dating back to Kermack and McKendrick (1927). Hence, compared to the existing studies, the travel history of confirmed COVID-19 patients used in our paper

can provide more direct evidence of actual spatial interactions between places' COVID-19 (e.g., which places' COVID-19 is connected to each other through the movement of infected people), and a realistic representation of real-world pandemic transmission networks.

Second, as mentioned before, while recent research has empirically studied the socioeconomic determinants of the spread of pandemics, it provides little direct evidence on the spatial interactions in pandemic spatial transmission. Our study is the first to assess the role of spatial interactions in the spread of pandemic COVID-19 from a network perspective. A network approach can be instrumental in capturing the intricate structure of linkages between regions as well as what externalities the disease risk associated with a single place may create for the entire country. A better understanding of spatial interactions may facilitate the design of effective strategies for pandemic containment.

Meanwhile, in terms of methodology, by developing and simultaneously estimating a structural model of dynamic disease transmission network formation and spatial interaction, we capture many dynamic aspects in this structural model, such as the persistence of disease incidence, the persistence of disease links between places and transitivity, and the effect of previous disease incidence on current disease transmission network formation, which can provide valuable insights into the dynamics of spatial interaction and disease diffusion.

A third contribution of the paper is to empirically determine the optimal policy for preventing the spread of pandemics. This is an important but under-explored area, especially when medical resources are scarce and externalities through spatial spillovers abound (Bouveret & Mandel, 2021). We use a particular counterfactual study called key player analysis to demonstrate how targeting can improve the effectiveness of policy interventions for containing the geographic spread of COVID-19. The key player approach is particularly important in providing practical guidance for the design of policies, as it takes into account both the network structure and the intensity of interactions (Lee et al., 2021).

The counterfactual policy analysis shows that due to the spatial interaction effect, the source region is the first node in COVID-19 transmission network that needs to be targeted and gets priority when allocating our scarce medical resources in order to optimally contain the geographic spread of the pandemics. There can be large potential gains from targeting interventions to the province which is likely to be a bad seed for spatial diffusion. Moreover, earlier implementation of this targeted policy can have a large effect on reducing aggregate new cases, which highlights the importance of a timely, accurate, and focused response in containing the pandemic.

The paper proceeds as follows. Section 2 presents our data and describes the interprovincial COVID-19 propagation network. Section 3 presents a structural model of dynamic disease transmission network formation and spatial interaction. The main empirical findings are given in Sect. 4. We conduct a Monte Carlo simulation study to show the performance of our estimation method in Sect. 5. In Sect. 6, we conduct a counterfactual policy analysis. Section 7 concludes.

2 Data and Interprovincial COVID-19 Propagation Network

2.1 Data and Sample Selection

Consistent with the existing studies (Muniz-Rodriguez et al., 2020; Qiu et al., 2020), our study is based on 31 provinces in Mainland China. Data on confirmed COVID-19 cases in Mainland China by province come from 31 provincial health commissions. The COVID-19 outbreak first emerged in Wuhan, Hubei Province in China. January 19, 2020, is the first day that COVID-19 cases were reported outside of Hubei Province. Figure 1 displays the time series pattern of daily new confirmed cases in Mainland China over the period January 19–March 9. Since Hubei Province started to include clinically diagnosed cases into new confirmed cases on February 12, the number of new confirmed cases reached a peak on this day. Therefore, in order to keep a consistent statistical criterion, we choose January 19–February 11 as our sample period.[1]

The key measurement challenge for the empirical literature studying the spatial spread of diseases is the difficulty of observing which regions' diseases are connected to each other. Traditionally, researchers have used the (inverse) spatial distance between regions (e.g., Qiu et al., 2020) and the flow of travelers by road or by air between regions (e.g., Fang et al., 2020; Tatem & Smith, 2010) to study the structures of real-world disease transmission networks, which is a rough estimation of the disease connections between places. In contrast, we introduce a new data set based on the travel history of confirmed patients to capture how two regions are connected through the movement of infected people, which have been identified as an important source for the spread of communicable diseases by epidemiologists. Thus, the data provide a more direct way to assess the disease contagion risk that one place poses to other places through infected people flows and allow us to analyze the spatial and temporal diffusion of COVID-19 within a country.

Specifically, when the earliest cases of COVID-19 were detected in Wuhan, Hubei Province, the virus spread rapidly to other provinces aided by the Chinese spring festival travel. As case numbers grew, provincial health commissions in China

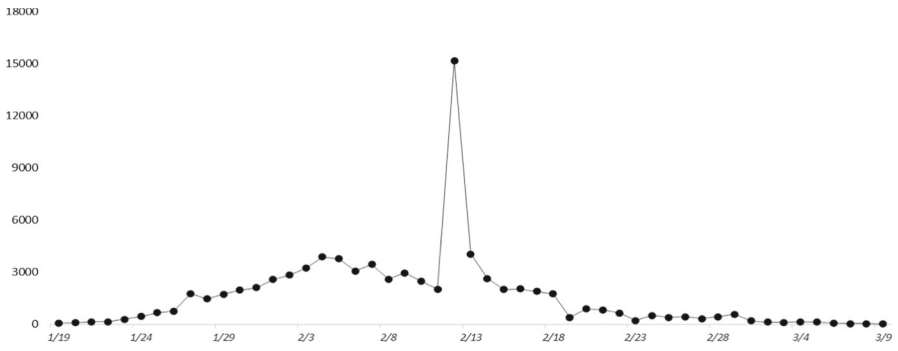

Fig. 1 Daily new confirmed cases in Mainland China

[1] We also conduct a robustness check on the selection of the sample period in Sect. 4.3.

reported a detailed log of anonymized COVID-19 cases' movements down to the day on their websites for a massive contact-tracing. Therefore, we manually collect the travel history of confirmed cases from 31 provincial health commissions, which contains 6162 individual observations on interprovincial migration.

We aggregate the individual-level observations to the province level based on the places that infected people move from and to. Our final confirmed cases' migration data set covers 22,320 province-pair observations for 31 Chinese provinces in the sample period.[2]

2.2 Interprovincial COVID-19 Propagation Network

Our data provide valuable information on the disease transmission network and allow us to track the spatial spread of COVID-19.

Specifically, Fig. 2 depicts the interprovincial COVID-19 propagation network on January 29, 2020, which is the first day that the epidemic spread to every province of Mainland China. Each circle represents a province and each arrow captures the directed movement of confirmed cases between two provinces (i.e. flowing out of province A to province B). The size of each circle corresponds to a province's new confirmed cases. As shown in Fig. 2, provinces with more ties to Hubei as well as other provinces have more new confirmed COVID-19 cases, revealing the nature of the spatial interaction between places: that is, one province's level of epidemic is affected by the epidemic in other provinces through infected people flows. Such spatial interaction can function as a potential propagation mechanism of COVID-19 throughout the country.

3 The Empirical Model

In the network effects literature, the spatial autoregressive (SAR) model is normally used as a spatial interaction model. Based on this model, we further develop a spatial dynamic panel data model to empirically investigate spatial interactions and the spread of COVID-19 across regions from a dynamic perspective:

$$Y_{i,t} = \rho Y_{i,t-1} + \lambda \sum_{j=1}^{J} g_{ij,t} Y_{j,t} + X_{i,t}\beta_1 + \frac{1}{g_{i,t}} \sum_{j=1}^{J} g_{ij,t} X_{j,t}\beta_2 + \alpha_t + \mu_{i,t}, \qquad (1)$$

where $Y_{i,t}$ and $Y_{j,t}$ denote the numbers of new confirmed cases of COVID-19 in provinces i and j on date t. $g_{ij,t}$ is the cumulative number of confirmed cases from province j to province i on date t, and $g_{i,t} = \sum_{j=1}^{J} g_{ij,t}$. Network (spatial weights or adjacency) matrix $G_t = \left[g_{ij,t}\right]$ summarizes the connections between provinces on date t. Since the network links have no mandatory reciprocity, G_t is not symmetric. The coefficient λ captures the spatial interaction (spillover) effect in the spread of

[2] If we aggregate the individual-level observations to the city level, it would consist of 3,171,168 city-pair observations for 364 Chinese cities in the sample period, which would be huge computation burden in the following estimation.

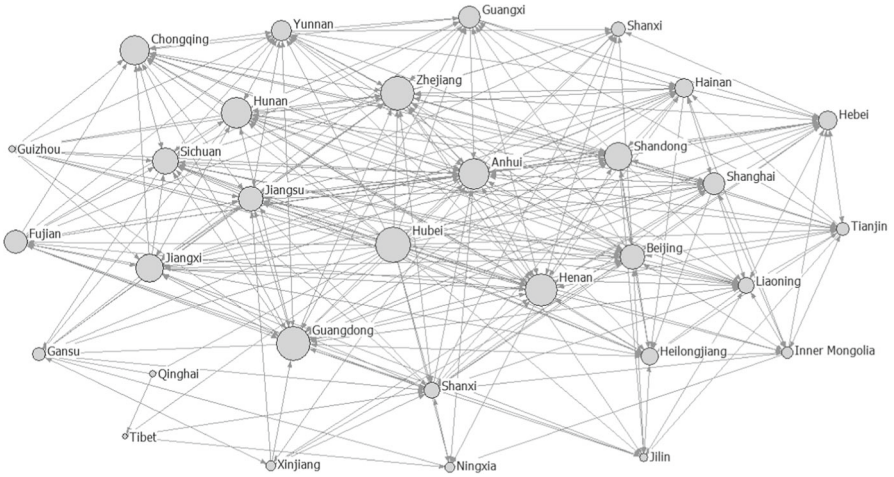

Fig. 2 Propagation network of COVID-19 across provinces on January 29, 2020

COVID-19, that is, how one province's level of epidemic is affected by the epidemic in other provinces through infected people flows, which is the key parameter of interest.

$X_{i,t}$ and $X_{j,t}$ control for observable province-specific variables which the literature has identified as potentially influencing the spread of diseases (Adams et al., 2003; Oster, 2005; Adda, 2016; Fang et al., 2020; Qiu et al., 2020). These variables include contemporaneous temperature and humidity as well as number of hospitals, population density, per capita GDP, unemployment rates, age structure as the fraction of the population above age 65, number of colleges and universities, passenger turnover through rail, airport, ferry and road services, the share of trade (exports and imports) in GDP. Data on these variables come from China Meteorological Administration and *the China Statistical Yearbook 2019*. Further we include time fixed effects α_t and $\mu_{i,t}$ is the random error term. In the social network literature, β_1 and β_2 represent the own and contextual effects of exogenous characteristics, respectively. Table 1 presents the summary statistics for all variables.

A major econometric challenge is to obtain consistent estimate of the key parameter λ. If the formation process of the time varying network matrix G_t involves unobserved provincial attributes that correlate with $\mu_{i,t}$, the resulting estimate for λ would be biased and inconsistent. A sophisticated approach to deal with this potential network endogeneity problem is to model the network formation process and then simultaneously estimate the network formation and the outcome equation (Goldsmith-Pinkham & Imbens, 2013; Han et al., 2021; Hsieh & Lee, 2016; Patacchini et al., 2017).

Specifically, let us consider a dynamic disease transmission network formation model based on homophily from observed and unobserved characteristics. Since the disease transmission network is directional, we have:

Table 1 Summary statistics

Variable	Observations	Mean	Standard deviation	Min	Max
Panel A. Province-pair level					
Cumulative number of cases from one province to another (ln)	22,320	0.52	0.94	0.00	6.40
Panel B. Province level					
Number of new confirmed cases (ln)	744	2.20	1.55	0.00	8.06
Temperature	744	2.21	8.56	−22.24	22.78
Humidity (ln)	744	4.20	0.27	3.11	4.58
Number of hospitals (ln)	744	10.10	0.84	8.42	11.35
Age structure (%)	744	11.27	2.46	5.68	15.16
GDP per capita (ln)	744	11.01	0.38	10.35	11.85
Population density (ln)	744	7.90	0.39	7.04	8.61
Unemployment rates (%)	744	3.11	0.57	1.40	4.00
Number of colleges and universities (ln)	744	4.26	0.74	1.95	5.12
Passenger turnover (ln)	744	6.31	0.91	3.85	7.64
Trade share (%)	744	23.63	25.13	1.68	104.08

$$g_{ij,t} = \alpha_0 + \left(X_{i,t} - X_{j,t}\right)\alpha_x + \alpha_y\left(Y_{i,t-1} - Y_{j,t-1}\right) + \alpha_d D_{ij,\mathrm{t}} + \alpha_g g_{ij,\mathrm{t}-1} + \alpha_f F_{ij,t-1} + \alpha_v\left|v_{i,t} - v_{j,t}\right| + \varepsilon_{ij,t}, \quad (2)$$

where $\left(X_{i,t} - X_{j,t}\right)$ is the difference in observable characteristics between provinces i and j and $\left(Y_{i,t-1} - Y_{j,t-1}\right)$ is the difference between the lagged outcomes, which captures homophily effects. $D_{ij,\mathrm{t}}$ denotes whether province i is adjacent to province j. $g_{ij,\mathrm{t}-1}$ captures the persistence of past network links and $F_{ij,t-1}$ is the transitivity of past network links, that is, whether the two provinces share links in common in the previous period. By introducing the persistence and the transitivity of past network links, our model captures interdependencies in link formation across different periods (Graham, 2016; Han et al., 2021). $v_{i,t}$ accounts for all province i's unobservable attributes and $\left|v_{i,t} - v_{j,t}\right|$ measures the difference in the unobserved characteristics between provinces. This is a standard model of homophily (Fafchamps & Gubert, 2007).

By introducing this province-specific unobserved component $v_{i,t}$, we generalize the spatial interaction model in Eq. (1) to

$$Y_{i,t} = \rho Y_{i,t-1} + \lambda\sum_{j=1}^{J} g_{ij,t}Y_{j,t} + X_{i,t}\beta_1 + \frac{1}{g_{i,t}}\sum_{j=1}^{J} g_{ij,t}X_{j,t}\beta_2 + \beta_3 v_{i,t} + \alpha_t + \mu_{i,t}. \quad (3)$$

Equations (2) and (3) form a structural model of dynamic disease transmission network formation equation and spatial interaction equation. As the disease

transmission network formation is explicitly modeled, the potential network endogeneity problem can be addressed.

Consistent with the literature, we use the Bayesian method to estimate it, which has advantages over traditional methods (Goldsmith-Pinkham & Imbens, 2013; Hsieh and Lin, 2021). Specifically, we specify the prior distributions of the unknown parameters and the unobserved variables in the models. Given the prior distributions and the likelihood functions of the models, we derive the posterior distributions of the parameters and simulate random draws from them by the Markov Chain Monte Carlo (MCMC) sampling to obtain the parameter estimates. Details on the Bayesian estimation procedure can be found in "Appendix".

4 Results

4.1 Spatial Interactions in the Spread of COVID-19

We run the MCMC algorithm for 80,000 iterations and drop the first 15,000 draws for burn-in and keep every 20th of the remaining draws to compute the posterior mean as a point estimate and posterior variance for each parameter. Following Hsieh et al. (2019), the trace plot of draws for the spatial interaction (spillover) effect parameter λ is depicted in Fig. 3 to check the convergence of the MCMC algorithm. The trace plot of MCMC draws for λ and its posterior distribution in the upper and middle panels show that the MCMC draws are stable and have good variations. The autocorrelation function (ACF) plotted in the bottom panel indicates that the

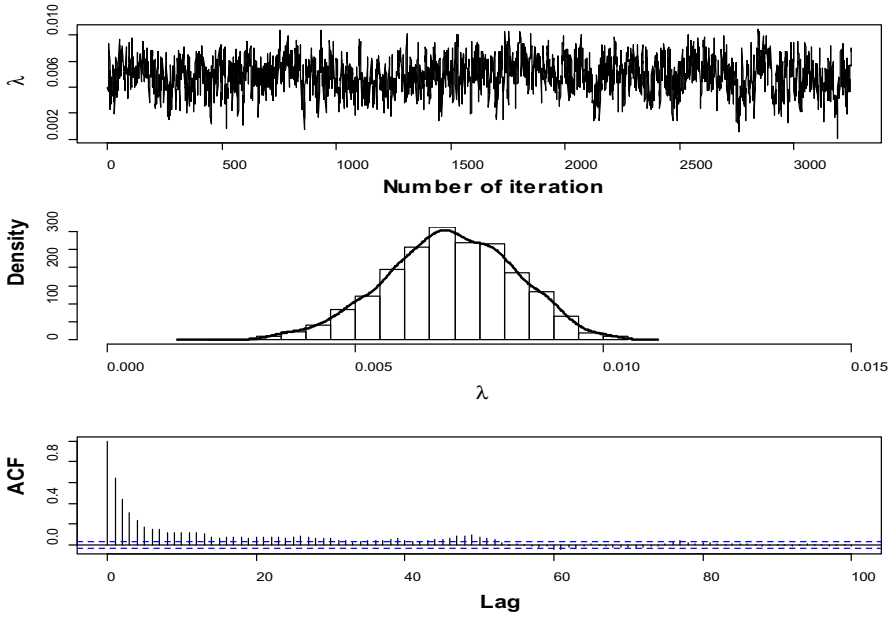

Fig. 3 Trace plot for MCMC draws of λ

correlation among draws decline gradually over iterations. And the draws also pass the convergence diagnostic test of Geweke (1992) with a p-value of 0.6836, which further confirms the convergence of the MCMC algorithm.

Table 2 presents the estimation results. Column (1) shows the basic estimates for Eq. (1) without considering the network endogeneity issue. Column (2) incorporates the network formation model.

Several results from this table should be highlighted. First and foremost, compared to the existing studies without considering spatial interactions (Adams et al., 2003; Oster, 2005; Adda, 2016; Fang et al., 2020; Qiu et al., 2020; Desbordes, 2021), throughout the models, the estimated spatial spillover effects are always positive and statistically significant, suggesting the existence of spatial interactions between regional new confirmed cases of COVID-19. In other words, one province's level of epidemic is affected by the epidemic in other provinces. This result means that abstracting from such spatial interaction effect comes with the risk of severely misunderstanding the propagation mechanism of the pandemic. Furthermore, we also conduct a Monte Carlo simulation to show that the estimates without considering spatial interactions are severely biased in Sect. 5.

In terms of the design of effective strategies for containment of the pandemic, this empirical result shows that there exists a fundamental externality in containing the pandemic. That is, due to the spatial interactions in disease spatial transmission, a region's policy interventions can affect the epidemic in other regions. Hence, to contain the geographic spread of the pandemic, local governments need to collaborate and policy interventions across regions should be coordinated. That is to say, effective containment and mitigation strategies largely rest on multilateral collaboration.

Second, the coefficients of the lagged dependent variable also seem robust and statistically significant. This result suggests the time persistent nature of COVID-19.

Third, for own characteristics, local hospitals significantly decrease the number of new cases, while increases in the number of colleges and universities significantly increase the number of new cases. For contextual characteristics, passenger turnover in other provinces have a positive and statistically significant effect on the number of new cases in the focal province, while the share of elderly people and hospitals in other provinces have a negative and statistically significant effect on the number of new cases in the focal province.

Fourth, the parameters from the network formation model show that heterophily on observed factors are important for interprovincial COVID-19 transmission network formation. Specifically, larger differences in unemployment rate, passenger turnover and trade share result in more movement of infected people between two provinces. Meanwhile, smaller difference in unobserved characteristic result in more movement of infected people between two provinces. And infected people are more likely to move between adjacent provinces ($D_{ij,t}$). In terms of the dynamics of the disease transmission network formation, provinces with more new cases in the previous period ($Y_{i,t-1} - Y_{j,t-1}$) would receive fewer infected people from other provinces in the current period. The movement of infected people between two provinces is less likely to occur, if they were connected to one common province in

Table 2 Spatial interactions and the spread of COVID-19

	(1) Outcome equation without link formation		(2) Outcome equation with link formation	
Spatial spillover effect (λ)	0.0065***	(0.0013)	0.0069***	(0.0013)
$Y_{i,t-1}$	0.6605***	(0.0291)	0.6426***	(0.0290)
Own effects				
Temperature	0.0007	(0.0053)	0.0039	(0.0053)
Humidity	0.0664	(0.1225)	0.0211	(0.1249)
Number of hospitals	−0.1071	(0.0666)	−0.1500**	(0.0710)
Age structure	−0.0009	(0.0161)	0.0040	(0.0167)
GDP per capita	0.0749	(0.0610)	0.0784	(0.0625)
Population density	0.0537	(0.0735)	0.1008	(0.0796)
Unemployment rates	0.0265	(0.0617)	0.0149	(0.0660)
Number of colleges and universities	0.4083***	(0.1492)	0.4236***	(0.1535)
Passenger turnover	−0.0252	(0.0978)	−0.0238	(0.1016)
Trade share	−0.0013	(0.0019)	−0.0022	(0.0020)
Contextual effects				
Temperature	−0.0033	(0.0156)	−0.0060	(0.0158)
Humidity	0.1286	(0.2481)	0.1527	(0.2521)
Number of hospitals	−0.9249***	(0.3332)	−1.2853***	(0.3824)
Age structure	−0.1457**	(0.0665)	−0.1739**	(0.0713)
GDP per capita	0.0805	(0.2367)	0.1188	(0.2528)
Population density	−0.0274	(0.3285)	0.0761	(0.3354)
Unemployment rates	0.3287	(0.2950)	0.5148	(0.3191)
Number of colleges and universities	−0.5298	(0.4116)	−0.0702	(0.4647)
Passenger turnover	1.4755***	(0.4864)	1.5117***	(0.5058)
Trade share	−0.0025	(0.0074)	−0.0035	(0.0077)
$v_{i,t}$			0.0241***	(0.0086)
Network formation				
Temperature			0.0000	(0.0001)
Humidity			−0.0021	(0.0022)
Number of hospitals			−0.0006	(0.0014)
Age structure			0.0001	(0.0003)
GDP per capita			−0.0035	(0.0022)
Population density			−0.0014	(0.0014)
Unemployment rates			0.0019*	(0.0011)
Number of colleges and universities			−0.0016	(0.0021)
Passenger turnover			0.0030*	(0.0016)
Trade share			0.0001**	(0.0000)
$D_{ij,t}$			0.0177***	(0.0019)

Table 2 continued

	(1) Outcome equation without link formation	(2) Outcome equation with link formation	
$Y_{i,t-1} - Y_{j,t-1}$		−0.0015***	(0.0005)
$F_{ij,t-1}$		−0.0118***	(0.0036)
$g_{ij,t-1}$		1.0037***	(0.0009)
$v_{i,t} - v_{j,t}$		−0.0031***	(0.0003)
Time fixed effects	Yes	Yes	
Observations (network model)		21,390	
Observations (outcome equation)	713	713	

Standard deviations of the posterior draws are in parentheses

***, **, and *Highest posterior density range does not cover zero at 99%, 95%, and 90% levels

the previous period ($F_{ij,t-1}$). The movement of infected people between two provinces is more likely to occur, if they were connected to each other in the previous period ($g_{ij,t-1}$) which reflects the persistence of past network links.

4.2 Heterogeneity in Spatial Interactions

The earliest cases of COVID-19 in China were detected in Hubei Province, and most of the early cases outside Hubei can be traced to previous contacts with persons in Hubei. Hubei has been at the center of China's outbreak. Meanwhile, all cities in Hubei Province were consecutively placed under lockdown with traffic ban for all residents after the Wuhan lockdown starting on January 23, 2020 (Qiu et al., 2020). Therefore, to capture heterogeneous spatial spillover effects from Hubei Province and all other provinces, we modify the outcome Eq. (3) as follows:

$$Y_{i,t} = \rho Y_{i,t-1} + \gamma \sum_{j \neq Hubei} g_{ij,t} Y_{j,t} + \delta g_{i,Hubei,t} Y_{Hubei,t} + X_{i,t}\beta_1 + \frac{1}{g_{i,t}} \sum_{j=1}^{J} g_{ij,t} X_{j,t} \beta_2 + \beta_3 v_{i,t} + \alpha_t + \mu_{i,t}, \quad (4)$$

where $Y_{Hubei,t}$ is the number of new confirmed cases of COVID-19 in Hubei on date t. $g_{i,Hubei,t}$ is the cumulative number of confirmed cases from Hubei in province i on date t.

Table 3 shows the estimation results. Column (1) shows the basic estimates for Eq. (4) without considering the network endogeneity issue. Column (2) incorporates the network formation model. We find that the estimated spatial spillover effects from Hubei (δ) and the spatial spillover effects from other provinces except Hubei (γ) are positive and statistically significant, consistent with the results from Table 2.

Moreover, the spatial spillover effects from Hubei are significantly higher than from other provinces, suggesting the dominant status of the source region in the

geographic spread of COVID-19 in China. An important implication of this finding is that policy interventions targeting the source region would be most effective due to its influential status. This is consistent with the result of Hollingsworth et al. (2006) on the global diffusion of SARS in 2003, which strongly suggests that intervention in the source region is crucial: if the source epidemic is controlled before there are thousands of cases, travel restrictions during the containment phase may have a large impact on the probability that an infected individual travels out of the source area and potentially seeds a new outbreak.

4.3 Sensitivity Analysis

First, as mentioned before, the connections between provinces on date t are based on the network matrix $G_t = \left[g_{ij,t}\right]$, where $g_{ij,t}$ is the cumulative number of confirmed cases from province j to province i on date t. To test its robustness, we convert this network matrix into a 0–1 network matrix. Specifically, $g_{ij,t} = 1$ if province i has one or more confirmed cases from province j and 0 otherwise on date t. As shown in column (1) of Table 4, the coefficient of spatial spillover effect is positive and statistically significant, which is consistent with the baseline results of Table 2, suggesting the robustness of our main results.

Second, we exclude Hubei Province which is the source region in the geographic spread of COVID-19 in Mainland China. The aim is to check whether the source region generates the correlations of spatial interaction with the spread of COVID-19. This turns out not to be the case. As shown in column (2) of Table 4, our main results remain robust to the exclusion of Hubei Province.

Table 3 Heterogeneous spatial spillover effects

	(1) Outcome equation without link form		(2) Outcome equation with link form	
Spatial spillover effect from Hubei (δ)	0.0093*	(0.0057)	0.0138**	(0.0065)
Spatial spillover effect from other provinces except Hubei (γ)	0.0057**	(0.0022)	0.0051**	(0.0024)
$Y_{i,t-1}$	0.6559***	(0.0305)	0.6271***	(0.0329)
Own effects	Yes		Yes	
Contextual effects	Yes		Yes	
Network formation	No		Yes	
Time fixed effects	Yes		Yes	
Observations (network model)			21,390	
Observations (outcome equation)	713		713	

Standard deviations of the posterior draws are in parentheses

***, **, and *Highest posterior density range does not cover zero at 99%, 95%, and 90% levels

Third, we restrict the sample to the before Hubei Province lockdown period (i.e., before January 28, 2020) to check the robustness of the sample period selection. Column (3) of Table 4 presents the estimates for this subsample. It also indicates a positive and significant effect of spatial spillover effect, which suggests our main results remain robust to this subsample.

5 Simulation Study

We also conduct a Monte Carlo simulation study to investigate the performance of our estimation method.

In this simulation, we consider 40 nodes (J=40) and 25 time periods (T=25) to investigate the finite sample performance of our estimation algorithm. We set the number of Monte Carlo repetitions at 50. Following Han et al. (2021), our data generating process (DGP) runs as follows: we generate individual exogenous variables $\{X_{i,t}\}_{i=1}^{J}$ for the network formation and the spatial interaction model from the normal distribution $\mathrm{N}(0,4)$. We generate dyadic exogenous variable $\{D_{ij,\mathrm{t}}\}_{i,j=1}^{J}$ by drawing two variables, namely, u_1 and u_2 from the uniform distribution $U(0,1)$ and setting $D_{ij,\mathrm{t}} = 1$ if both u_1 and u_2 are above 0.7 or below 0.3; and setting $D_{ij,\mathrm{t}} = 0$ otherwise. The individual effects are simulated from $\mathrm{N}(0,0.05)$. We generate time effects with α_1 set to 0 and $\{\alpha_t\}_{t=2}^{T}$ simulated from $\mathrm{N}(0,1)$. The latent variable $v_{i,0}$ is drawn from $\mathrm{N}(0,1)$ and we iteratively generate $v_{i,t} = v_{i,t-1} + \mu_v$ with μ_v drawn from $\mathrm{N}(0,1)$. We simulate the initial outcome $\{Y_{i,0}\}_{i=1}^{J}$ from $\mathrm{N}(0,1)$ and the initial network G_0 from Eq. (2). Thereafter, we follow the network formation and spatial

Table 4 Sensitivity analysis

	(1) Alternative network matrix	(2) Excluding the source region	(3) Before lockdown
Spatial spillover effect (λ)	0.0131***	0.0118***	0.0065**
	(0.0024)	(0.0014)	(0.0027)
$Y_{i,t-1}$	0.6382***	0.5842***	0.5920***
	(0.0311)	(0.0319)	(0.0553)
Own effects	Yes	Yes	Yes
Contextual effects	Yes	Yes	Yes
Network formation	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Observations (network model)	21,390	20,010	8370
Observations (outcome equation)	713	690	279

Standard deviations of the posterior draws are in parentheses

***, **, and *Highest posterior density range does not cover zero at 99%, 95%, and 90% levels

interaction models to generate networks G_t and outcome Y_t for $t \geq 1$. The preceding steps apply to all relevant time periods.

Table 5 presents the simulation results. The true values of the DGP parameters are listed in the first column for comparison. We use the posterior mean of MCMC draws as the point estimate for each parameter and report the mean and the standard deviation across repetitions in column (2) of Table 5. As shown in Table 5, the Bayesian MCMC estimation approach in Sect. 4 can effectively recover the true parameters from the model of Eqs. (2) and (3). The simulation demonstrates that the MCMC approach performs well in estimation.

Meanwhile, column (3) of Table 5 shows the results without considering network matrix, which is the standard linear regression model used in the existing studies on the determinants of the spread of pandemics (Adams et al., 2003; Oster, 2005; Adda, 2016; Fang et al., 2020; Qiu et al., 2020; Desbordes, 2021). The estimates in this model are severely biased because of the omissions of spatial interactions (i.e., network matrix).

The above experiment results are based on 50 repetitions. The values shown in the table are average and standard deviation of the point estimates across repetitions. The point estimates from the mean of MCMC posterior draws.

6 Counterfactual Policy Analysis

In this section, we use a particular counterfactual study called key player analysis to demonstrate how targeting can improve the effectiveness of policy interventions for containing the geographic spread of COVID-19. A crucial aspect of network analysis

Table 5 Simulation results

Parameters	(1)	(2)			(3)		
	DGP	With considering network matrix			Without considering network matrix		
		Mean	Bias	Standard deviation	Mean	Bias	Standard deviation
λ	0.0500	0.0497	−0.0003	0.0014			
ρ	0.3000	0.2999	−0.0001	0.0029	0.3438	0.0438	0.0347
β_1	1.0000	1.0021	0.0021	0.0181	0.9281	−0.0719	0.3561
β_2	1.0000	1.0001	0.0001	0.0030			
β_3	1.0000	1.0170	0.0170	0.0539			
α_0	0.0200	0.0202	0.0002	0.0017			
α_d	0.0100	0.0099	−0.0001	0.0030			
α_x	−0.0010	−0.0010	0.0000	0.0005			
α_y	−0.0010	−0.0010	0.0000	0.0000			
α_g	0.0500	0.0484	−0.0016	0.0057			
α_v	0.0100	0.0103	0.0003	0.0007			
σ_μ^2	1.0000	1.0121	0.0121	0.0882	814.6740	813.6740	2548.6430
σ_ε^2	0.0500	0.0501	0.0001	0.0003			

is the identification of the key player which holds a vital position in a network and should be targeted by the planner so that, once removed from the network generates the highest decrease in total activity (Zenou, 2016). In our paper, the key player is the province that should get priority when allocating scarce medical resources or be quarantined in order to optimally contain the geographic spread of COVID-19.

Ballester et al. (2006) propose a measure which takes into account both the network topology and the intensity of social interactions in determining the identity of the key player (Lee et al., 2021).

More formally, let $Y(g) = \sum_{t=2}^{T} \sum_{i=1}^{J} Y_{i,t}$ denote the total level of epidemic in COVID-19 transmission network g for the whole sample period and $g^{[-i]}$ denote the network obtained by removing the ith province from network g. Then the key player in network g is defined as

$$i^* = \text{argmax}\left\{\text{Y(g)} - \text{Y}\left(\text{g}^{[-\text{i}]}\right)|\text{i} = 1, \ldots, \text{n}\right\} = \text{argmin}\left\{\text{Y}\left(\text{g}^{[-\text{i}]}\right)|\text{i} = 1, \ldots, \text{n}\right\}. \tag{5}$$

And the reduction in total new cases is calculated as

$$\left(Y(g) - Y\left(g^{[-i]}\right)\right)/Y(g).$$

With the estimates of Model 2 reported in Table 2, we calculate the formula of the key player for each province. The results for the key player analysis are reported in Table 6, where top ten provinces are ranked according to the reduction in aggregate new cases when they are removed from the network (e.g., by quarantine or focusing medical resources on them). Such a ranking can guide a national policymaker that wants to minimize total new cases by implementing public health measures and allow the policymaker to identify which provinces should get priority when allocating scarce medical resources or be quarantined.

The key province is Hubei, which is the source region and center of China's outbreak. Our results suggest that, without this province, total new cases would be 87.70% lower for the whole sample period. The second and third highest ranked provinces are Guangdong and Henan. Removing them reduces the total new cases by 13.32% and 7.38%, respectively.

We also consider the case of heterogeneous spillovers by using the estimates of Model 2 reported in Table 3. As shown in Table 6, the key province is Hubei, consistent with the homogeneous case. Removing the key player reduces the total new cases by 91.91%, which is slightly higher than the homogeneous case.

To evaluate the efficiency of the key-player-targeted policy, we also consider a random target policy. Specifically, for each province i in the COVID-19 network g, the reduction in total new cases when i is the random selection with uniform probability for all provinces in the network is defined as:

$$\frac{1}{J}\sum_{i=1}^{J}\left[Y(g) - Y\left(g^{[-i]}\right)\right]/Y(g).$$

Table 6 Key player ranking

Province	Homogeneous spatial spillovers		Province	Heterogeneous spatial spillovers	
	Total new cases reduction (%)	Rank		Total new cases reduction (%)	Rank
Hubei	−87.70	1	Hubei	−91.91	1
Guangdong	−13.32	2	Guangdong	−11.01	2
Henan	−7.38	3	Jiangxi	−5.85	3
Jiangxi	−7.21	4	Henan	−5.56	4
Jiangsu	−6.60	5	Jiangsu	−5.53	5
Anhui	−6.44	6	Hunan	−5.21	6
Hunan	−6.33	7	Anhui	−5.10	7
Zhejiang	−4.39	8	Zhejiang	−3.89	8
Fujian	−2.97	9	Fujian	−2.51	9
Shanxi	−2.67	10	Tianjin	−2.27	10

The differences between the impacts of the two policies are remarkable. Specifically, we find that removing a random province reduces the total new cases only by 4.84%, while removing the key player (Hubei) reduces the total new cases by 87.70% in the homogeneous case and by 91.91% in the heterogeneous case as shown in Table 6. This suggests that the key-player-targeted policy is more effective in containing the geographic spread of COVID-19 as it takes full advantage of the spatial interaction effect. In other words, targeting key players generates large multiplier and amplifying effects as opposed to the random target policy (Lee et al., 2021): targeting the key player, not only reduces its level of epidemic, but also the epidemic in its directly connected regions, which, in turn, affects that in their directly connected regions, and so forth. In a random target policy, these chain effects are similar but the target is on the "wrong" province. Overall, the result suggests that there can be large potential gains from targeting interventions to provinces which are likely to be bad seeds for spatial diffusion, which has important implications for the optimal allocation of scarce medical resources during a pandemic. That is, the optimal allocation is to first exclusively focus medical resources on key provinces rather than splitting medical resources equally across regions in order to optimally contain a pandemic.

In addition, to assess the effect of key-player-targeted policy over time, we consider a 1-, 2-, and 3-week delay on this policy, respectively. As shown in Table 7 based on the homogeneous spillover case, a one-week delay in removing the key player (Hubei) would reduce the total new cases by 75.82%, which is less effective than the no delay case. And removing the key player (Hubei) could reduce the total new cases only by 47.02% and 10.05% for a 2- and 3-week delay respectively, suggesting the effect of key-player-targeted policy decreases with an increase in the time of delay. An important implication of this finding is that the speed of policy response is the crucial factor in containing the pandemic.

7 Conclusion

Because viruses are a major threat to human health and impose a cost on society, understanding global disease dynamics has become a major twenty first-century challenge (Adda, 2016; Brockmann & Helbing, 2013). This paper analyzes the role of spatial interactions in determining the spatial and temporal evolution of COVID-19 from a network perspective.

We use data on actual spatial interactions, as measured by the directed movement of confirmed cases between two provinces in China, in combination with data on each province's new confirmed COVID-19 cases to construct an interprovincial COVID-19 propagation network. By developing and simultaneously estimating a structural model of dynamic disease transmission network formation and spatial interaction using a Bayesian approach, we find that spatial interactions have a pronounced effect on the spread of COVID-19. We further find that the spatial spillover effect from Hubei is significantly higher than from other provinces, suggesting the dominant status of the source region in the spread of disease. We also find that the movement of infected people between two provinces is more likely to occur, if they are adjacent provinces, and if they were connected to each other in the previous period which reflects the persistence of past network links. Provinces with more new cases in the previous period would receive fewer infected people from other provinces in the current period. The Monte Carlo simulation results show that our MCMC approach performs well in estimation, and the estimates without considering spatial interactions are severely biased.

We then conduct a counterfactual policy analysis showing that in the presence of spatial interaction effect, targeting the source region would significantly improve the effectiveness of policy interventions for containing the geographic spread of COVID-19. This is due to the multiplier and amplifying effects generated by the key player in the network. Moreover, the effect of key-player-targeted policy decreases with an increase in the time of delay.

Table 7 Total new cases reduction for a 1-, 2-, and 3-week delay (%)

	No delay	One-week delay	Two-week delay	Three-week delay
Hubei	−87.70	−75.82	−47.02	−10.05
Guangdong	−13.32	−11.54	−7.01	−1.21
Henan	−7.38	−6.53	−3.87	−0.63
Jiangxi	−7.21	−6.17	−3.61	−0.57
Jiangsu	−6.60	−5.70	−3.32	−0.59
Anhui	−6.44	−5.62	−3.23	−0.53
Hunan	−6.33	−5.32	−3.03	−0.43
Zhejiang	−4.39	−4.09	−2.58	−0.47
Fujian	−2.97	−2.53	−1.39	−0.19
Shanxi	−2.67	−2.27	−1.26	−0.18

Crises like the COVID-19 pandemic occur with regularity. Past examples include MERS, SARS, and 2009 H1N1 pandemics. And there is substantial evidence that the rate of emergence of novel diseases is increasing (Dobson et al., 2020). Understanding the forces at work is of high policy importance. Our results suggest that spatial interactions between regions strongly influence the duration and severity of a pandemic. Within this context, to contain the geographic spread of the pandemic, local governments need to collaborate and policy interventions across regions should be coordinated. Effective containment and mitigation strategies largely rest on multilateral collaboration. In addition, our counterfactual analysis shows early intervention in the source region can lead to a much higher total new cases reduction. Therefore, compared to dividing the medical resources equally across regions, allocating scarce medical resources to the source region can have an immediate and pronounced impact.

Although our study focuses on the role of spatial interactions in the spread of pandemic COVID-19 in China, our estimated results can have general implications to other countries in their fight against COVID-19. Moreover, our estimation methodology can also be applied to education, crime, R&D, scientific co-authorship, and other fields where interaction effects are present (Bramoullé et al., 2020).

While our analysis fills a research gap in the field of the determinants of the spread of pandemics across time and space, limitations also apply to our research that future research can address. First, because of data constraints, we only examine the effect of spatial interactions on the spread of COVID-19 in the short run, and future research could investigate this issue in a long-run perspective once the data become available. Second, COVID-19 data may be imprecise, incomplete, and uncertain due to some reasons, such as the incompleteness of our observations, measurements, and estimations, or due to the existing disturbances and uncertainties in the statistical processes (Chen et al., 2017). Neutrosophic statistics is the extension of classical statistics and is applied when the data is coming from a complex process or from an uncertain environment (Aslam, 2021a, 2021a; Aslam et al., 2019, 2020; Chen et al., 2017; Sherwani et al., 2021; Smarandache et al., 2018). Hence, future research can extend our work by applying neutrosophic statistics to the statistical analysis of COVID-19 data.[3]

Appendix: Bayesian Estimation

Let $\theta = \left(\Psi', \beta', \beta_3, \beta_4, \sigma_\mu^2\right)'$ with $\Psi = (\lambda, \rho)'$, $\beta = (\beta_1{}', \beta_2{}')'$ and $\Gamma = \left(\Phi', \sigma_\varepsilon^2\right)'$ with $\Phi = \left(\alpha_0, \alpha_x{}', \alpha_y, \alpha_d, \alpha_g, \alpha_f, \alpha_v\right)'$. Following Han et al. (2021), our prior distributions are

$$\Phi \backsim N_{6+k}(\Phi_0, P_0),$$

$$\beta \backsim N_{2k}(\beta_0, B_0),$$

[3] We appreciate the insights provided by an anonymous reviewer on this issue.

$$\beta_3 \backsim N(\beta_{30}, B_{30}),$$

$$\beta_4 \backsim N(\beta_{40}, B_{40}),$$

$$\sigma_\mu^2 \backsim \mathrm{IG}\left(\frac{\mathrm{a}_\mu}{2}, \frac{\mathrm{b}_\mu}{2}\right),$$

$$\sigma_\varepsilon^2 \backsim \mathrm{IG}\left(\frac{\mathrm{a}_\varepsilon}{2}, \frac{\mathrm{b}_\varepsilon}{2}\right),$$

$$\lambda \backsim U\left(-\frac{1}{\Delta_G}, \frac{1}{\Delta_G}\right),$$

$$\rho \backsim U(-1 + |\lambda|\Delta_G, 1 - |\lambda|\Delta_G),$$

where $\Delta_G = max_{t=1,2,\ldots,\mathrm{T}} \| G_t \|_\infty$ due to the stability conditions: $S_t(\lambda) = I_J - \lambda G_t$ is invertible and $\| S_t^{-1}(\lambda)\rho I_J \|_\infty < 1$, $N_l(\cdot)$, $IG(\cdot)$ and $U(\cdot)$ are respectively the multivariate normal, inverse gamma and uniform distribution, and β_4 is the coefficient of time fixed effects α_t. We choose the hyperparameters in the prior distributions as follows: $\Phi_0 = 0$, $P_0 = 100I_{6+k}$, $\beta_0 = 0$, $B_0 = 100I_{2k}$, $\beta_{30} = 0$, $B_{30} = 100$, $\beta_{40} = 0$, $B_{40} = 100$, $a_\mu = 3$, $b_\mu = 2$, $a_\varepsilon = 3$, $b_\varepsilon = 2$.

The latent variables V_t are modeled by a Markov process with the initial distribution

$$\mathrm{P}\left(\mathrm{V}_1|\sigma_{\mathrm{v}_0}^2\right) = \prod_{\mathrm{i}=1}^{\mathrm{J}} \phi\left(\mathrm{v}_{\mathrm{i}1}|0, \sigma_{\mathrm{v}_0}^2\right)$$

and the transition equation

$$\mathrm{P}\left(\mathrm{V}_\mathrm{t}|\mathrm{V}_{\mathrm{t}-1}, \sigma_\mathrm{v}^2\right) = \prod_{\mathrm{i}=1}^{\mathrm{J}} \phi\left(\mathrm{v}_{\mathrm{it}}|\mathrm{v}_{\mathrm{i,t}-1}, \sigma_\mathrm{v}^2\right)$$

for $t = 2, 3, \ldots, T$, where ϕ denotes the normal density and $\sigma_{v_0}^2$ and σ_v^2 are normalized to be 1.

The MCMC sampling procedure combines the Gibbs sampling and the Metropolis–Hastings (M–H) algorithm. It consists of the following steps:

Step 1: Sample v_{i1} from $\mathrm{P}\left(\mathrm{v}_{\mathrm{i}1}|\mathrm{Y}_1, \mathrm{G}_1, \mathrm{v}_{-\mathrm{i},1}, \mathrm{v}_{\mathrm{i}2}, \theta, \Gamma\right) \propto \phi(\mathrm{v}_{\mathrm{i}1}) \times \phi(\mathrm{v}_{\mathrm{i}2}|\mathrm{v}_{\mathrm{i}1}) \times \mathrm{P}(\mathrm{Y}_1, \mathrm{G}_1|\mathrm{V}_1, \theta, \Gamma)$ for $i = 1, 2, \ldots, \mathrm{J}$, using the M–H algorithm.

Step 2: Sample v_{it} from $\mathrm{P}\left(\mathrm{v}_{\mathrm{it}}|\mathrm{Y}_\mathrm{t}, \mathrm{G}_\mathrm{t}, \mathrm{v}_{\mathrm{i,t}+1}, \mathrm{v}_{-\mathrm{i,t}}, \mathrm{v}_{\mathrm{i,t}-1}, \theta, \Gamma\right) \propto \phi\left(\mathrm{v}_{\mathrm{i,t}+1}|\mathrm{v}_{\mathrm{it}}\right) \times \phi\left(\mathrm{v}_{\mathrm{it}}|\mathrm{v}_{\mathrm{i,t}-1}\right) \times \mathrm{P}(\mathrm{Y}_\mathrm{t}, \mathrm{G}_\mathrm{t}|\mathrm{V}_\mathrm{t}, \theta, \Gamma)$ for $i = 1, 2, \ldots, J$ and $t = 2, \ldots, T-1$, using the M–H algorithm.

Step 3: Sample v_{iT} from $\mathrm{P}\left(\mathrm{v}_{\mathrm{iT}}|\mathrm{Y}_\mathrm{T}, \mathrm{G}_\mathrm{T}, \mathrm{v}_{-\mathrm{i,T}}, \mathrm{v}_{\mathrm{i,T}-1}, \theta, \Gamma\right) \propto \phi\left(\mathrm{v}_{\mathrm{iT}}|\mathrm{v}_{\mathrm{i,T}-1}\right) \times \mathrm{P}(\mathrm{Y}_\mathrm{T}, \mathrm{G}_\mathrm{T}|\mathrm{V}_\mathrm{T}, \theta, \Gamma)$ for $i = 1, 2, \ldots, \mathrm{J}$, using the M–H algorithm.

Step 4: Sample Ψ from $\mathrm{P}\left(\Psi|\{\mathrm{Y_t}\},\{\mathrm{G_t}\},\{\mathrm{V_t}\},\beta,\beta_3,\beta_4,\sigma_\mu^2\right) \propto \mathrm{u}(\Psi) \times \mathrm{P}(\{\mathrm{Y_t}\}|\{\mathrm{G_t}\},\{\mathrm{V_t}\},\Psi,\beta,\beta_3,\beta_4,\sigma_\mu^2)$ using the M–H algorithm, where $u(\cdot)$ is the uniform density function.

Step 5: Sample β_3 from $\mathrm{P}\left(\beta_3|\{\mathrm{Y_t}\},\{\mathrm{G_t}\},\{\mathrm{V_t}\},\Psi,\beta,\beta_4,\sigma_\mu^2\right) \propto \phi\left(\mathrm{T}_{\beta_3},\Sigma_{\beta_3}\right)$, where $\Sigma_{\beta_3} = \left(\mathrm{B}_{30}^{-1} + \sigma_\mu^{-2}\sum_{t=1}^{T} V_t' V_t\right)^{-1}$ and $T_{\beta_3} = \Sigma_{\beta_3}\left(\mathrm{B}_{30}^{-1}\beta_{30} + \sigma_\mu^{-2}\sum_{t=1}^{T} V_t'(S_t(\lambda)Y_t - \rho Y_{t-1} - \boldsymbol{X}_t\beta - d_t\beta_4)\right)$ with $\boldsymbol{X}_t = \left(X_t, \overline{G}_t X_t\right)$ and $\overline{G}_t$ denotes the row-normalized G_t. We restrict $\beta_3 > 0$ to pin down the signs of V_t's.

Step 6: Sample β from $\mathrm{P}\left(\beta|\{\mathrm{Y_t}\},\{\mathrm{G_t}\},\{\mathrm{V_t}\},\Psi,\beta_3,\beta_4,\sigma_\mu^2\right) \propto \phi_{2\mathrm{k}}\left(\mathrm{T}_\beta,\Sigma_\beta\right)$, where $\phi_{2k}(\cdot)$ is the multivariate $2k$-dimensional normal density function, $\Sigma_\beta = \left(B_0^{-1} + \sigma_\mu^{-2}\sum_{t=1}^{T}\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1}$ and $T_\beta = \Sigma_\beta\left(B_0^{-1}\beta_0 + \sigma_\mu^{-2}\sum_{t=1}^{T}\boldsymbol{X}_t'(S_t(\lambda)Y_t - \rho Y_{t-1} - V_t\beta_3 - d_t\beta_4)\right)$.

Step 7: Sample σ_μ^2 from $\mathrm{P}\left(\sigma_\mu^2|\{\mathrm{Y_t}\},\{\mathrm{G_t}\},\{\mathrm{V_t}\},\Psi,\beta_3,\beta_4,\beta\right) \propto \imath\gamma\left(\frac{\mathrm{a}_\mu^{\mathrm{new}}}{2},\frac{\mathrm{b}_\mu^{\mathrm{new}}}{2}\right)$, where $\imath\gamma(\cdot)$ is the inverse gamma density function, $a_\mu^{new} = a_\mu + T\cdot J$ and $b_\mu^{new} = b_\mu + \sum_{t=1}^{T}\mu_t'\mu_t$.

Step 8: Sample β_4 from $\mathrm{P}\left(\beta_4|\{\mathrm{Y_t}\},\{\mathrm{G_t}\},\{\mathrm{V_t}\},\Psi,\beta,\beta_3,\sigma_\mu^2\right) \propto \phi\left(\mathrm{T}_{\beta_4},\Sigma_{\beta_4}\right)$, where $\Sigma_{\beta_4} = \left(B_{40}^{-1} + \sigma_\mu^{-2}\sum_{t=1}^{T} d_t' d_t\right)^{-1}$ and $T_{\beta_4} = \Sigma_{\beta_4}\left(B_{40}^{-1}\beta_{40} + \sigma_\mu^{-2}\sum_{t=1}^{T} d_t'(S_t(\lambda)Y_t - \rho Y_{t-1} - \boldsymbol{X}_t\beta - V_t\beta_3)\right)$.

Step 9: Sample Φ from $\mathrm{P}\left(\Phi|\{\mathrm{G_t}\},\{\mathrm{V_t}\},\sigma_\varepsilon^2\right) \propto \phi_{6+\mathrm{k}}(\mathrm{T}_\Phi,\Sigma_\Phi)$, where $\Sigma_\Phi = \left(\mathrm{P}_0^{-1} + \sigma_\varepsilon^{-2}\sum_{t=1}^{T}\chi_t'\chi_t\right)^{-1}$, $T_\Phi = \Sigma_\Phi\left(\mathrm{P}_0^{-1}\Phi_0 + \sigma_\varepsilon^{-2}\sum_{t=1}^{T}\chi_t'\tilde{G}_t\right)$ with

$$\chi_t = \begin{pmatrix} 1, \left(X_{1,t} - X_{2,t}\right), \left(Y_{1,t-1} - Y_{2,t-1}\right), D_{12,\mathrm{t}}, g_{12,\mathrm{t}-1}, F_{12,t-1}, \left|v_{1,t} - v_{2,t}\right| \\ 1, \left(X_{1,t} - X_{3,t}\right), \left(Y_{1,t-1} - Y_{3,t-1}\right), D_{13,\mathrm{t}}, g_{13,\mathrm{t}-1}, F_{13,t-1}, \left|v_{1,t} - v_{3,t}\right| \\ \vdots \\ 1, \left(X_{i,t} - X_{j,t}\right), \left(Y_{i,t-1} - Y_{j,t-1}\right), D_{ij,\mathrm{t}}, g_{ij,\mathrm{t}-1}, F_{ij,t-1}, \left|v_{i,t} - v_{j,t}\right| \\ \vdots \\ 1, \left(X_{J,t} - X_{J-1,t}\right), \left(Y_{J,t-1} - Y_{J-1,t-1}\right), D_{J,J-1,\mathrm{t}}, g_{J,J-1,\mathrm{t}-1}, F_{J,J-1,t-1}, \left|v_{J,t} - v_{J-1,t}\right| \end{pmatrix}$$

for $i,j = 1,2,\ldots,\mathrm{J}, i \neq j$, and let $G_{it\backslash i} = \left(g_{i1,t}, g_{i2,t}, \ldots g_{i,i-1,t}, g_{i,i+1,t}, \ldots, g_{iJ,t}\right)$ for $i = 1,2,\ldots,\mathrm{J}$, $\tilde{G}_t = \left(G_{1t\backslash 1}, G_{2t\backslash 2}, \ldots, G_{Jt\backslash J}\right)'$.

Step 10: Sample σ_ε^2 from $\mathrm{P}\left(\sigma_\varepsilon^2|\{\mathrm{G_t}\},\{\mathrm{V_t}\},\Phi\right) \propto \imath\gamma\left(\frac{\mathrm{a}_\varepsilon^{\mathrm{new}}}{2},\frac{\mathrm{b}_\varepsilon^{\mathrm{new}}}{2}\right)$, where $a_\varepsilon^{new} = a_\varepsilon + T\cdot J$ and $b_\varepsilon^{new} = b_\varepsilon + \sum_{t=1}^{T}\varepsilon_t'\varepsilon_t$.

We collect the draws from iterating the above steps and compute the posterior mean and the posterior standard deviation as our estimation results.

Acknowledgements This paper has benefited from helpful comments of Chih-Sheng Hsieh and Xiaoyi Han. Financial support from National Natural Science Foundation of China (71974076) and Guangdong Natural Science Foundation of China (2019A1515011923) is gratefully acknowledged.

Funding This study was funded by National Natural Science Foundation of China (71974076), Guangdong Natural Science Foundation of China (2019A1515011923) and the Fundamental Research Funds for the Central Universities (19JNKY07).

Declarations

Conflict of interest The authors Cui Zhang and Dandan Zhang of the paper entitled "Spatial interactions and the spread of COVID-19: A network perspective", declare that there is no conflict of interest.

References

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics, 112* (1), 3–56.

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *The Quarterly Journal of Economics, 131*(2), 891–941.

Aslam, M. (2021). Analyzing wind power data using analysis of means under neutrosophic statistics. *Soft Computing, 25*(10), 7087–7093.

Aslam, M., Arif, O. H., & Sherwani, R. A. K. (2020). New diagnosis test under the neutrosophic statistics: An application to diabetic patients. *BioMed Research International*. https://doi.org/10.1155/2020/2086185

Aslam, M., Bantan, R. A., & Khan, N. (2019). Design of a new attribute control chart under neutrosophic statistics. *International Journal of Fuzzy Systems, 21*(2), 433–440.

Aslam, M., Rao, G. S., Saleem, M., Sherwani, R. A. K., & Jun, C. H. (2021). Monitoring mortality caused by COVID-19 using gamma-distributed variables based on generalized multiple dependent state sampling. *Computational and Mathematical Methods in Medicine*. https://doi.org/10.1155/2021/6634887

Ballester, C., Calvó-Armengol, A., & Zenou, Y. (2006). Who's who in networks. Wanted: The key player. *Econometrica, 74*(5), 1403–1417.

Basu, S., & Sen, S. (2022). COVID 19 pandemic, socio-economic behaviour and infection characteristics: An inter-country predictive study using deep learning. *Computational Economics*. https://doi.org/10.1007/s10614-021-10223-5

Bouveret, G., & Mandel, A. (2021). Social interactions and the prophylaxis of SI epidemics on networks. *Journal of Mathematical Economics, 93*, 102486.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics, 12*, 603–629.

Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science, 342*(6164), 1337–1342.

Chen, J., Ye, J., Du, S., & Yong, R. (2017). Expressions of rock joint roughness coefficient using neutrosophic interval statistical numbers. *Symmetry, 9*(7), 123.

Desbordes, R. (2021). Spatial dynamics of major infectious diseases outbreaks: A global empirical assessment. *Journal of Mathematical Economics, 93*, 102493.

Dobson, A. P., Pimm, S. L., Hannah, L., Kaufman, L., Ahumada, J. A., Ando, A. W., Bernstein, A., Busch, J., Daszak, P., Engelmann, J., Kinnaird, M. F., Li, B. V., Loch-Temzelides, T., Lovejoy, T., Nowak, K., Roehrdanz, P. R., & Vale, M. M. (2020). Ecology and economics for pandemic prevention. *Science, 369*(6502), 379–381.

Fafchamps, M., & Gubert, F. (2007). Risk sharing and network formation. *The American Economic Review, 97*(2), 75–79.

Fang, H., Wang, L., & Yang, Y. (2020). Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China. *Journal of Public Economics, 191*, 104272.
Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian Statistics, 4*, 641–649.
Goenka, A., Liu, L., & Nguyen, M. H. (2021). SIR economic epidemiological models with disease induced mortality. *Journal of Mathematical Economics, 93*, 102476.
Goldsmith-Pinkham, P., & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business and Economic Statistics, 31*(3), 253–264.
Graham, B. S. (2016). *Homophily and transitivity in dynamic network formation* (No. w22186). National Bureau of Economic Research.
Guimarães, L. (2021). Antibody tests: They are more important than we thought. *Journal of Mathematical Economics, 93*, 102485.
Han, X., Hsieh, C. S., & Ko, S. I. (2021). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business and Economic Statistics, 39*(1), 120–135.
Hollingsworth, T. D., Ferguson, N. M., & Anderson, R. M. (2006). Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine, 12*(5), 497–499.
Hsieh, C. S., König, M. D., & Liu, X. (2019). A structural model for the coevolution of networks and behavior. *Review of Economics and Statistics, 104*(2), 1–41.
Hsieh, C. S., & Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics, 31*(2), 301–319.
Hsieh, C. S., & Lin, X. (2021). Social interactions and social preferences in social networks. *Journal of Applied Econometrics, 36*(2), 165–189.
Jackson, M. O., Rogers, B. W., & Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature, 55*(1), 49–95.
Jha, P., Deshmukh, Y., Tumbe, C., Suraweera, W., Bhowmick, A., Sharma, S.,...,Brown, P. (2022). COVID mortality in India: National survey data and health facility deaths. *Science.* https://doi.org/10.1126/science.abm5154.
Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London: Series A Containing Papers of a Mathematical and Physical Character, 115*(772), 700–721.
Konstantinoudis, G., Cameletti, M., Gómez-Rubio, V., Gómez, I. L., Pirani, M., Baio, G.,...,Blangiardo, M. (2022). Regional excess mortality during the 2020 COVID-19 pandemic in five European countries. *Nature Communications, 13*(1), 1–11.
Kung, S., Doppen, M., Black, M., Hills, T., & Kearns, N. (2021). Reduced mortality in New Zealand during the COVID-19 pandemic. *The Lancet, 397*(10268), 25.
Lee, L. F., Liu, X., Patacchini, E., & Zenou, Y. (2021). Who is the key player? A network analysis of juvenile delinquency. *Journal of Business and Economic Statistics, 39*(3), 849–857.
Li, J., Wang, L., Guo, S., Xie, N., Yao, L., Cao, Y.,...,Sun, D. (2020). The data set for patient information based algorithm to predict mortality cause by COVID-19. *Data in Brief, 30*, 105619.
Muniz-Rodriguez, K., Chowell, G., Cheung, C. H., Jia, D., Lai, P. Y., Lee, Y., & Fung, I. C. H. (2020). Doubling time of the COVID-19 epidemic by province, China. *Emerging Infectious Diseases, 26*(8), 1912.
Oster, E. (2005). Sexually transmitted infections, sexual behavior, and the HIV/AIDS epidemic. *The Quarterly Journal of Economics, 120*(2), 467–515.
Patacchini, E., Rainone, E., & Zenou, Y. (2017). Heterogeneous peer effects in education. *Journal of Economic Behavior and Organization, 134*, 190–227.
Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *Journal of Population Economics, 33*(4), 1127–1172.
Sherwani, R. A. K., Shakeel, H., Saleem, M., Awan, W. B., Aslam, M., & Farooq, M. (2021). A new neutrosophic sign test: An application to COVID-19 data. *PLoS ONE, 16*(8), e0255671.
Smarandache, F., Khalid, H. E., & Essa, A. K. (2018). Neutrosophic logic: The revolutionary logic in science and philosophy. In *Infinite study.*
Tatem, A. J., & Smith, D. L. (2010). International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proceedings of the National Academy of Sciences of USA, 107*(27), 12222–12227.

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet, 395*(10225), 689–697.

Zenou, Y. (2016). Key players. *The Oxford Handbook of the Economics of Networks, 4*, 244–274.