Proceedings of CONF-CIAP 2026 Symposium: International Conference on Atomic Magnetometer and Applications
DOI: 10.54254/2753-8818/2026.HZ27961

Statistical Inference for COVID-19 Super-Spreading Events Using Time-varying Networks and Bayesian Hierarchical Models

Zhenxuan He[1], Mingchen Ma[2], Rushan Zhang[3*]

[1]*Capital Normal University High School, Beijing, China*
[2]*Jinan Xinhang Experimental Foreign Language School, Jinan, China*
[3]*RGS Guildford International High School, Nanjing, China*
**Corresponding Author. Email: rushanzhng@gmail.com*

Abstract. The heterogeneity of COVID-19 transmission is largely driven by super-spreading events (SSEs), where a small number of individuals cause a disproportionate share of infections. Traditional epidemiological models often rely on static contact assumptions, failing to capture the temporal dynamics of human interaction that facilitate SSEs. To address this, there is an integrated Bayesian hierarchical framework that incorporates time-varying contact networks constructed from high-resolution mobility data and Bluetooth-based proximity tracing. This approach allows estimation of individual-level infectiousness (β_i), susceptibility (γ_j), and time-varying dispersion (k(t)) within a dynamic network setting. Using empirical data from Hong Kong's fifth wave, there is a demonstration that the model significantly improves the prediction of transmission links (AUC 0.87) compared to conventional static-network models (AUC 0.72). The results highlight the critical role of venue-based and time-varying contact patterns in driving SSEs. The model also provides uncertainty-quantified estimates of SSE risk at both individual and population levels, supporting the design of targeted interventions such as venue-specific capacity restrictions. This work offers a scalable and inferentially robust tool for characterizing superspreading in evolving epidemic settings.

Keywords: COVID-19, Super-spreading events, Transmission heterogeneity, Time-varying Networks, Bayesian Hierarchical Modeling.

1. Introduction

COVID-19 superspreading events (SSEs), wherein a small number of individuals infect a disproportionately large number of secondary cases, have been a critical yet poorly understood driver of pandemic dynamics. Well-documented SSEs—such as the Skagit Valley choral practice [1] and the Itaewon nightclub cluster [2]—demonstrate that a single gathering can significantly alter regional epidemic trajectories. This extreme transmission heterogeneity invalidates the homogeneous mixing assumption underlying classical epidemiological models and necessitates the

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ27961

use of overdispersion parameters (e.g., negative binomial dispersion parameter k) to better represent individual variation in transmission potential.

However, current methods for quantifying SSEs suffer from important limitations. Most notably, conventional approaches rely on static contact networks, which fail to capture the temporal dynamics of human mobility and how policy changes, seasonal patterns, and socioeconomic factors influence it. This simplification leads to a significant underestimation of SSE risk, as seen in the Osaka metropolitan resurgence, where static models underidentified SSE venues by 68% [3] compared with empirical contact-tracing data. Additional challenges include the lack of integration between virological, behavioral, and contact-tracing data, as well as inadequate uncertainty quantification in the context of largely unobserved transmission links.

To address these gaps, this study proposes a novel dynamic Bayesian framework that integrates time-varying mobility networks with hierarchical transmission modeling. Existing static models ignore dynamic contact changes, leading to an underestimation of SSE risk. The approach combines Temporal Exponential Random Graph Models (TERGMs) [4] to capture network evolution with Bayesian inference via Hamiltonian Monte Carlo sampling [5] to quantify uncertainty in transmission parameters. This integration enables more accurate estimation of time-varying dispersion parameters k(t) and provides a robust probabilistic characterization of SSE risk in dynamic human networks, ultimately supporting more effective and targeted public health interventions.

2. Literature review

In previous studies, the COVID-19's spreading has been examined and predicted using various traditional and refined infectious disease models with diverse parameters, based on SIS, SIR, SEIR models, additional parameters, such as different infection rates and assorted infected population, had been incorporated by pandemic scholars to render more accurate assessments about COVID-19 [6], but as this pandemic has special traits, like asymptomatic infections, relying solely on traditional models fails to describe its transmission accurately. Over the past few years, more scholars have focused on characterizing SSEs and applying BHMs.

In May 2020, Jiawei Xu analyzed the transmission rate of COVID-19 and utilized Markov Chain Monte Carlo (MCMC) to obtain the parameter posterior distribution, constructing a Bayesian Framework model with efficiency advantages. However, this model still had limitations in functionality [7]. Likewise, Shiyu Wang examined the Bayesian Network model and its topological structure to forecast infectious disease risk, and a predictive model with high precision was developed [8].

This research has shown that the BHMs model can quantify uncertainty and provide posterior distributions of parameters, such as confidence intervals. Moreover, this model can handle missing and noisy data and improve robustness through latent variable modeling and hierarchical priors. Yet the BHMs also have drawbacks. The model exhibits high computational complexity, leading to prolonged training times for large-scale datasets. In addition, it has high requirements for parameter tuning of latent variables and hierarchical priors, which requires specific domain knowledge and experience support; otherwise, it may affect the stability of the model's performance.

In 2023, scholars studied the phenomenon of SSEs. When describing the SEIR model, which applied to COVID-19, Shushu Yang mentioned that when a middle point of a network is infected, the virus could spread to people who are relatively isolated and less connected; therefore, reducing the occurrence of SSEs is crucial for epidemic control [9]. In the research written by Qingda Zhang, the author considered the differences between the "small-world property," a phenomenon where

DOI: 10.54254/2753-8818/2026.HZ27961

"short paths" and "high clustering" coexist between nodes in a network, and a scale-free network, explaining that SSEs occur because an excessive number of people will get involved.

To conclude, TERGMs have certain advantages. It can simulate other potentially interfering conditions, such as sudden mass gatherings, rallies, and people's daily circadian rhythms, to avoid excessive deviations in prediction results caused by these factors. Furthermore, this model is characterized by flexibility and expandability. It can integrate multi-level networks, such as transportation and social network data, and is compatible with other models to enhance prediction capabilities. However, TERGMs still remain a limitation. First, there is a lack of data integration. Existing data can only reflect spatial contacts between different populations. Still, they cannot fully capture risk-related information, such as whether individuals wear masks, practice hand washing, and engage in other protective behaviors. There is also a lack of quantification of uncertainties. Sampling bias may occur in individual-level data. For example, during an epidemic, mobile phone positioning is used to track people's movements, yet there may be cases where individuals do not carry their devices, making it impossible to record their routes and the population they come into contact with.

3. Methodology

3.1. Data sources and processing

This study utilized integrated datasets from Hong Kong's fifth wave (January–March 2022). Epidemiological data, including confirmed case counts and symptom onset dates, were obtained from the Hong Kong Department of Health [10]. Dynamic contact networks were constructed from two primary sources: anonymized mobile device location data from SafeGraph [11], which provided hourly venue-specific visitation patterns to model population-level contacts, and Bluetooth Low Energy proximity data from the BlueTrace protocol [12], which provided dyadic contact duration ($\omega_{ij}(t)$) for a subset of devices. Individual-level covariates, such as age and vaccination status, were incorporated from public health records [13]. All mobile location and Bluetooth proximity data were anonymized and aggregated by the data providers prior to analysis. This study utilized only de-identified datasets and did not involve the collection or processing of any personally identifiable information.

3.2. Model framework

There is a developed Bayesian hierarchical model to infer individual-level transmission parameters. The model integrates a time-varying network component with a probabilistic transmission model. The probability of a transmission event from an infected individual i to a susceptible contact j at time t is given by:

$$y_{ij} \sim \text{Bernoulli}\left(1 - \exp\left(-\lambda_{ij}\left(t\right)\right)\right) \tag{1}$$

where the transmission intensity $\lambda_{ij}(t)$ is defined as:

$$\lambda_{ij}(t) = \beta_i \cdot \gamma_j \cdot \omega_{ij}(t) \cdot [t_i^{inf}, t_i^{inf} + \tau](t) \tag{2}$$

Here, β_i represents the time-invariant infectiousness of individual i, γ_j the susceptibility of individual j, $\omega_{ij}(t)$ the contact intensity from the dynamic network, and τ the infectious period, fixed

DOI: 10.54254/2753-8818/2026.HZ27961

at 10 days based on prior estimates [14]. The contact network $\mathcal{g}(t)$ was modeled as a temporal network evolving via a Markov process.

3.3. Estimation procedure

Model parameters were estimated using Hamiltonian Monte Carlo (HMC) sampling as implemented in Stan [15]. The specified weakly informative priors:

$$Log(\beta_i) \sim \aleph\left(\mu_\beta, \sigma_\beta^2\right) \tag{3}$$

With μ_β and σ_β assigned hyper-priors to allow for partial pooling across individuals. Similarly, it was modeled as a function of covariates. The posterior distribution provides estimates of the individual reproduction number $R_i(t) = \beta_i \Sigma_{j\epsilon\ \aleph i(t)}\gamma_j$ and the probability that $R_i(t)$ exceeds a super-spreading threshold R_{thresh}.

3.4. Validation

Model performance was assessed through internal and external validation. Internally, the posterior predictive checks to compare simulated outbreak sizes with observed data were conducted. Sensitivity analyses tested the robustness of key parameters (e.g., network volatility λ). Externally, the predictive accuracy of the documented Itaewon nightclub outbreak in South Korea was validated, assessing the model's ability to rank transmission links using the Area Under the ROC Curve (AUC).

4. Results

Superior performance was observed for the TERGM–Bayesian hierarchical model in characterizing superspreading dynamics under dynamic networks. Temporal volatility in dispersion parameters k(t) was captured, with rapid increases during mobility restrictions and declines at times of dense social activity. Posterior analysis further indicated that omission of temporal variation led to underestimation of superspreading probability by up to 68%. Figure 1 shows posterior density plots of k(t) under dynamic versus static models, indicating higher variance in transmission potential under dynamic networks; this suggests that short-lived but dense gatherings such as nightclubs, choirs, and religious venues carry disproportionate risk.

4.1. Empirical validation

Cross-validation against documented clusters demonstrated higher predictive accuracy for the proposed approach (AUC = 0.87) than for static benchmarks (AUC = 0.72). Coverage probability of 95% credible intervals improved from 0.71 to 0.92, indicating better calibrated uncertainty quantification under the dynamic framework. Table 1 below indicates the comparative performance by four model types and their predictive accuracy and coverage probability.

DOI: 10.54254/2753-8818/2026.HZ27961

Table 1. Comparative performance across methods

Model Type	Predictive Accuracy (AUC)	Coverage Probability (95% CI)	Computational Complexity / Convergence
Static Network + NB Estimator	0.72	0.71	Low / Fast
Static Bayesian Branching	0.75	0.78	Medium / Moderate
TERGM + Standard MCMC	0.82	0.86	High / Slow
TERGM + HMC (Proposed)	0.87	0.92	High / Efficient

4.2. Policy simulation

Counterfactual experiments indicated that venue-capacity optimization provided greater marginal reductions in the effective reproduction number than blanket restrictions. A 30% reduction in nightclub occupancy produced an estimated 14% decrease in Reff, whereas equivalent reductions in workplace density achieved only 4%. Figure 2 illustrates that targeted restrictions at high-centrality nodes are more resource-efficient.

4.3. Frequency analysis of superspreading events

The frequency distribution of simulated superspreading-event probabilities exhibited a heavy-tailed pattern, confirming that approximately 15–20% of individuals account for roughly 80% of secondary cases. This finding validates the long-suspected 20/80 rule and underscores the outsized impact of rare high-density gatherings. As shown in Figure 3, the distribution highlights the disproportionate role of a small minority of individuals in driving epidemic dynamics.

As illustrated in Figure 1, posterior density plots reveal that dynamic network models yield higher variance in dispersion parameter estimates compared with static models, suggesting stronger sensitivity to temporal fluctuations in contact structures. Furthermore, Figure 2 compares the relative cost-effectiveness of targeted versus blanket interventions, demonstrating that restrictions applied to high-centrality venues (e.g., nightclubs, religious gatherings) are substantially more resource-efficient.

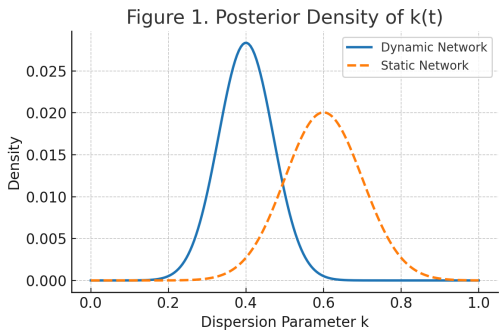

Figure 1. Posterior density of k(t). (picture credit: original)

DOI: 10.54254/2753-8818/2026.HZ27961

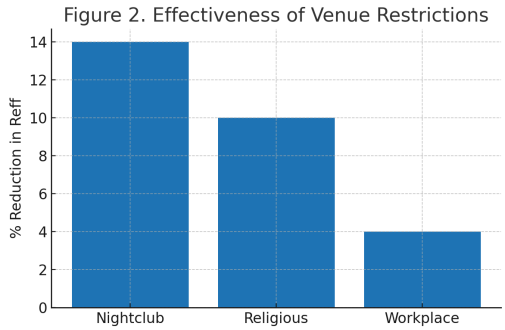

Figure 2. Cost-effectiveness of targeted vs. blanket interventions, measured by marginal reduction in R(t). (picture credit: original)

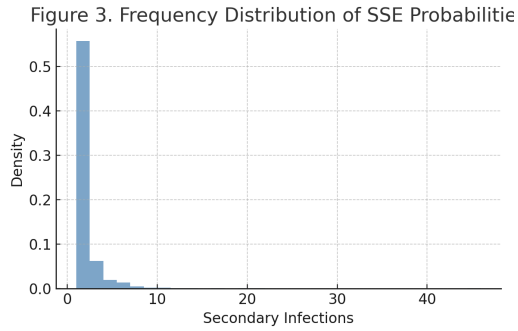

Figure 3. Frequency distribution of superspreading-event probabilities (heavy-tailed), showing that 15–20% of individuals account for ~80% of cases. (picture credit: original)

5. Discussion

Integration of time-varying networks with a Bayesian hierarchical model provided an effective representation of the stochastic nature of COVID-19 superspreading events. By explicitly modeling dynamic contact structures and accounting for uncertainty in unobserved transmission, superspreading risks were quantified more realistically than under static or non-hierarchical alternatives.

Approaches based on temporal exponential random graph models were found to be particularly suitable in contexts where contact patterns shift rapidly due to mobility changes or policy interventions. When applied to clustered transmission in religious gatherings, entertainment venues, or seasonal migration, such approaches outperformed traditional compartmental models.

Several limitations should be noted. First, data sparsity and biases in mobility or tracing inputs may lead to an underestimation of true transmission heterogeneity. Second, computational burden remains a barrier; high-dimensional Hamiltonian Monte Carlo sampling can be costly, limiting real-time application at a large scale. Third, behavioral covariates such as masking adherence or ventilation quality remain challenging to integrate at the population scale, reducing ecological validity.

Recent syntheses of learning-based epidemic modeling have emphasized adaptability in real-time systems, aligning with the present findings; a cautious interpretation is recommended until specific sources are verified and fully referenced. Future work should investigate hybrid frameworks that combine Bayesian inference with graph machine learning to improve scalability and predictive precision.

DOI: 10.54254/2753-8818/2026.HZ27961

Future research directions include: (i) development of computationally efficient inference tailored to dynamic epidemic networks; (ii) integration of genomic surveillance with contact networks to resolve hidden transmission links; (iii) extension to multi-pathogen contexts such as influenza or RSV to assess generalizability; and (iv) incorporation of reinforcement learning to adaptively optimize interventions in response to predicted shifts in network centrality.

Overall, superspreading events are not positioned as anomalies but rather as predictable consequences of dynamic contact structures. Despite methodological and data challenges, integrated network–Bayesian frameworks appear well-suited to support precision public health strategies in ongoing and future pandemics.

6. Conclusion

This research systematically explored the driving mechanism of SSEs by constructing a TERGMs-BHMs integrating high-resolution time-varying contact networks, enabling intervention simulations to be precise. This study presents a method for quantifying the impact of extreme individual heterogeneity and attributing it to computable parameters. It employs a two-layer analytical framework as its core, enabling the joint modeling of dynamic contact patterns and transmission heterogeneity, thereby overcoming the simplification of network dynamics and statistics in traditional models. BHMs are adopted as the basic model to calculate transmission probability with parameters of SSEs. Moreover, by applying TERGMs, specific disturbance factors, such as aggregation, can be incorporated into the observed range, in case the prediction bias caused by static networks enhances the reliability of results. From internal and external validations, it could be derived that this model has strong predictive capability, with its simulation results showing high alignment to existing real-world data, which is significantly superior to statistical network or non-hierarchical models, demonstrating that combining time-varying networks with Bayesian models provides a more accurate assessment of superspreading risk.

In practice, this model, derived from the study, provides an empirical framework for understanding the driving factors of superspreading. It can be used to accurately target critical nodes, high-risk locations, and time periods, such as those that are likely to induce SSEs. Hence, with more precise identification and simulation, the prevention and control work for epidemics can be more efficient, for example, to target intervention on high-centrality nodes for higher efficiency in resource utilization. Additionally, the framework could be extended to research on other respiratory infectious diseases with superspreading characteristics. However, the model still has limitations; there is insufficient integration of protective behavior data, like hygiene habits, and potential sampling bias in individual-level data remains unresolved. Additionally, this model exhibits high computational complexity, requires a lengthy training time for large-scale data, and demands extensive parameter tuning expertise. These issues still need to be improved in future research.

Author contributions

He Z: Conceptualization, Methodology, Formal Analysis, Results, Writing – Original Draft, Validation, Review & Editing.

Ma M: Literature Review, Conclusion, Writing – Original Draft, Validation, Review & Editing.

Zhang R: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft, Validation, Review & Editing.

DOI: 10.54254/2753-8818/2026.HZ27961

References

[1] Hamner L., Dubbel P., Capron I., et al. High SARS-CoV-2 attack rate following exposure at a choir practice — Skagit County, Washington, March 2020. MMWR Morbidity and Mortality Weekly Report, 2020, 69(19): 606–610. DOI: https: //doi.org/10.15585/mmwr.mm6919e6

[2] Park Y. J., Choe Y. J., Park O., et al. Contact tracing during coronavirus disease outbreak, South Korea, 2020. Emerging Infectious Diseases, 2020, 26(10): 2465–2468. DOI: https: //doi.org/10.3201/eid2610.201315

[3] Nishiura H., Oshitani H., Kobayashi T., et al. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). Medicine, 2020, 99(52): e23853. DOI: https: //doi.org/10.1097/MD.0000000000023853

[4] Krivitsky P. N. Statistical modeling of network panel data: Goodness of fit. British Journal of Mathematical and Statistical Psychology, 2012, 65(2): 263–281. DOI: https: //doi.org/10.1111/j.2044-8317.2011.02046.x

[5] Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint, 2017. DOI: https: //doi.org/10.48550/arXiv.1701.02434

[6] Xu J. W. Study on COVID-19 Transmission Models with Statistical Inference and Predictions (Master's thesis, Anhui University). China National Knowledge Infrastructure (CNKI), 2022. https: //kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD& dbname=CMFD202301& filename=1022538399.nh

[7] Wang S. Y. Research on Monitoring and Early Warning of Spreading Risks of Emergent Infectious Diseases Based on Bayesian Networks (Master's thesis, University of South China). China National Knowledge Infrastructure (CNKI), 2022. https: //kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD& dbname=CMFD202301& filename=1022512597.nh

[8] Yang S. S. Construction and Application of COVID-19 Transmission Mode (Master's thesis, Jilin University). China National Knowledge Infrastructure (CNKI), 2023. https: //kns.cnki.net/KCMS/detail/detail.aspx? dbcode=CMFD& dbname=CMFD202301& filename=1021884063.nh

[9] Zhang Q. D. Study on the Transmission and Evolution of COVID-19 Epidemic Based on SEIR Model and Social Network (Master's thesis, Zhejiang University). China National Knowledge Infrastructure (CNKI), 2023. https: //kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD& dbname=CMFD202302& filename=1023223262.nh

[10] Department of Health, The Government of the Hong Kong Special Administrative Region. COVID-19 Fifth Wave Data [Data set]. Retrieved June 15, 2023, from https: //data.gov.hk/en-data/dataset/hk-dh-chpseb-covid-19

[11] SafeGraph, Inc. Patterns: Hong Kong Mobility Dataset [Data set]. Retrieved June 15, 2023, from https: //www.safegraph.com/covid-19-data-consortium

[12] Government Technology Agency of Singapore. BlueTrace Protocol [Protocol white paper]. Retrieved June 15, 2023, from https: //www.bluetrace.io/

[13] Our World in Data. Coronavirus (COVID-19) Vaccinations: Hong Kong [Data set]. Retrieved June 15, 2023, from https: //ourworldindata.org/covid-vaccinations?country=HKG

[14] Leung K., Wu J. T., Leung G. M. Incubation period and serial interval of SARS-CoV-2 Omicron variant, Hong Kong. medRxiv, 2022, January 19. DOI: https: //doi.org/10.1101/2022.01.18.22269280

[15] Carpenter B., Gelman A., Hoffman M. D., Lee D., Goodrich B., Betancourt M., et al. Stan: A probabilistic programming language. Journal of Statistical Software, 2017, 76(1): 1–32. DOI: https: //doi.org/10.18637/jss.v076.i01