

Overcoming Challenges for Estimating Virus Spread Dynamics from Data

Damir Vrabac, Philip E. Paré, Henrik Sandberg, and Karl H. Johansson*

Abstract—In this paper we investigate estimating the parameters of a discrete time networked virus spread model from time series data. We explore the effect of multiple challenges on the estimation process including system noise, missing data, time-varying network structure, and quantization of the measurements. We also demonstrate how well a heterogeneous model can be captured by homogeneous model parameters. We further illustrate these challenges by employing recent data collected from the ongoing 2019 novel coronavirus (2019-nCoV) outbreak, motivating future work.

I. INTRODUCTION

Understanding virus spread dynamics over network systems is a vital problem, punctuated by the ongoing outbreak of the 2019 novel coronavirus (2019-nCoV) originating in Wuhan, China [1]. The data available from the outbreak is limited and not clean; further efforts are required to effectively combine the real data with mathematical models in order to be able to implement, and justify implementing, preventative measures. The model considered here is the susceptible-infected-susceptible (SIS) model, where each node can transition from being infected to healthy according to a healing rate δ and can become infected according to an infection rate β scaled by the node's connections to its infected neighbors and their infection levels.

Discrete time SIS models have been studied in the literature [2]–[9]. The authors of [2] introduce a discrete-time homogeneous¹ virus spread model over an undirected graph and provide a condition that ensures convergence to the healthy state, that is, where the virus is eradicated. In [3], the authors explore in more detail the model from [2]. In [4], the authors consider the healthy and the endemic (nonzero) states of several models, providing a sufficient condition for global stability of the endemic state for the model in [2], [3] and existence, uniqueness, and stability conditions for special cases of the endemic state. In [5], [6], existence and stability of the healthy and endemic states are provided for a more general model than those considered before. We employ the same model here.

In [5]–[8] the identification of the discrete time spread process we consider here was investigated. In [5], [6], conditions for recovering the the spread parameters from data are presented and real data is used to validate the model. However, several strong assumptions are made for the identification results, such as no noise in the system or measurements and full knowledge of the graph structure. In

[8], the assumption on the knowledge of the graph structure is relaxed (however assumptions are still made on the graph such that it be undirected) and, in addition to recovering the homogeneous spread parameters, the authors studied recovering the network structure of the model. In [5]–[7], [10], validation work was carried out using real data. In [10], Ebola data was used to fit a continuous time SID/S (where ‘D’ stands for dead or removed) model. Similar ideas were recently applied to the 2019-nCoV data [11].

In this paper we explore how parameter estimation techniques perform when the assumptions on the data are not met. These challenges include additive system noise, missing data, time-varying network structure, and quantized measurements. We also demonstrate, when there is not enough data to learn heterogeneous spread parameters, how well the spread process can be captured by homogeneous parameters. Finally, as a case study, we employ real data from the ongoing outbreak of 2019-nCoV in Wuhan, Hubei, China collected by JHU CSSE [12] from the World Health Organization [1], the CDC [13], the National Health Commission of the PRC, and DXY.cn; see Figure 1. We estimate homogeneous spread parameters of the outbreak in China and use the resulting model to project when the outbreak may have begun. While insightful, the results show that, given the limited, corrupted data available at the beginning of an outbreak, further work is required to overcome all of the challenges necessary to effectively estimate virus spread dynamics from data.

In Section II, we present the model and several assumptions required for the model to be well defined. In Section III we provide several results on estimating the spreading parameters of the model from time series data, improving on the results in [5], [6]. The results in Section III appear in [14] and were derived from the efforts in this paper. In Section IV, we explore via simulations the effect of system noise, missing data, time-varying network structure, and quantization of the measurements. We also illustrate how well a heterogeneous model can be captured by homogeneous model parameters. Some of these ideas are also explored in [15] in the competing virus setting. In Section V we apply the results to the 2019-nCoV dataset.

A. Notation

Given a vector function of continuous time x , \dot{x} indicates the time-derivative. Given a vector function of discrete time x^k , k is the time index. Given a vector $x \in \mathbb{R}^n$, the 2-norm is denoted by $\|x\|$ and the transpose by x^\top . The vector of all equal zeros is denoted by $\mathbf{0}$. Given two vectors $x^1, x^2 \in \mathbb{R}^n$, $x^1 > x^2$ indicates each element of x^1 is greater than or equal to the corresponding element of x^2 and $x^1 \neq x^2$, and

* The authors are all with the Division of Decision and Control Systems at KTH Royal Institute of Technology and Philip E. Paré can be contacted at (philipar@kth.se).

¹Homogeneous means β and δ are the same for every node in the network.



Fig. 1: [12] Spread of 2019-nCoV per province in China, as of January 30, 2020.

$x^1 \gg x^2$ indicates each element of x^1 is strictly greater than the corresponding element of x^2 . Given a matrix $A \in \mathbb{R}^{n \times n}$, the spectral radius is $\rho(A)$. Also, a_{ij} indicates the i, j^{th} entry of the matrix A , and $\|A\|_F$ indicates the Frobenius norm of A . The notation $\text{diag}(\cdot)$ refers to a diagonal matrix with the argument(s) on the diagonal; the argument can be a vector x or its elements x_i . For $n \in \mathbb{Z}^+$, $[n] := \{1, \dots, n\}$.

II. MODEL OF SPREADING PROCESS

The state of the model x_i is the adoption or infection level of the i th node. The state of can be considered as the probability of node i having adopted or being infected, or the proportion of subpopulation i that has adopted or is infected. For each node i the dynamics evolve as

$$x_i^{k+1} = x_i^k + h \left((1 - x_i^k) \beta_i \sum_j a_{ij} x_j^k - \delta_i x_i^k \right), \quad (1)$$

where k is the time index, $\beta_i > 0$ is the susceptibility rate, a_{ij} are the non-negative, edge weights between the nodes, $\delta_i > 0$ is the healing rate, for node i , and $h > 0$ is the sampling parameter. In matrix form, (1) becomes

$$x^{k+1} = x^k + h((I - X^k)BA - D)x^k, \quad (2)$$

where $X^k = \text{diag}(x^k)$, $B = \text{diag}(\beta_i)$, A is the matrix of a_{ij} , and $D = \text{diag}(\delta_i)$.

For the model to be well-posed we need the following assumptions.

Assumption 1. For all $i \in [n]$, we have $x_i^0 \in [0, 1]$.

Assumption 2. For all $i \in [n]$, we have $\delta_i \geq 0$, $\beta_i \geq 0$, and, for all $j \in [n]$, $a_{ij} \geq 0$.

Assumption 3. For all $i \in [n]$, we have $h\delta_i \leq 1$ and $h \sum_{j=1}^n \beta_i a_{ij} \leq 1$.

Lemma 1. [6] For the system in (2), under the conditions of Assumptions 1, 2, and 3, $x_i^k \in [0, 1]$ for all $i \in [n]$, and $k \geq 0$.

Lemma 1 implies that the set

$$\mathcal{D} = \{x \mid \mathbf{0} \leq x \leq \mathbf{1}\} \quad (3)$$

is positively invariant with respect to the system defined by (2). Since x_i denotes the probability of infection of individual i , or the fraction of group i infected, and $1 - x_i$ denotes the probability of individual i being healthy, or the fraction of group i that is healthy, it is natural to assume that their initial values are in the interval $[0, 1]$, since otherwise the values will lack any physical meaning for the epidemic model considered here. Therefore, we focus on the analysis of (2) only on the domain \mathcal{D} .

We need an assumption to ensure *non-trivial* spread.

Assumption 4. We have $B, A \neq 0$, $h \neq 0$, and $n > 1$.

III. LEARNING SPREAD PARAMETERS

In this section, we clearly lay out the assumptions and the identification techniques for the model of spreading processes. We present several results on learning the spread parameters of the model in (1) from data.

Theorem 1. Consider the model in (2) under Assumptions 1-4 with homogeneous spread, that is, β and δ are the same for all nodes. Assume that A, x^k , for all $k \in [T] \cup \{0\}$, and h are known. Then, β and δ can be identified uniquely if and only if $T > 0$, and there exist $i, j \in [n]$ and $k_1, k_2 \in [T-1] \cup \{0\}$ such that

$$x_i^{k_1} g_j(x^{k_2}) \neq x_j^{k_2} g_i(x^{k_1}), \quad (4)$$

where $g(x^k) := (I - X^k)Ax^k$.

Proof: Since x^k , for all $k \in [T-1] \cup \{0\}$, and A are known, using (2) we can construct the matrix Φ , defined as,

$$\Phi := \begin{bmatrix} (I - X^0)Ax^0 & -x^0 \\ \vdots & \vdots \\ (I - X^{T-1})Ax^{T-1} & -x^{T-1} \end{bmatrix}. \quad (5)$$

Therefore, since we also know x^T and h , we can rewrite (2) as

$$\begin{bmatrix} x^1 - x^0 \\ \vdots \\ x^T - x^{T-1} \end{bmatrix} = h\Phi \begin{bmatrix} \beta \\ \delta \end{bmatrix}. \quad (6)$$

Since $n > 1$, Φ has at least two rows. By the assumption that there exist $i, j \in [n]$ and $k_1, k_2 \in [T-1] \cup \{0\}$ such that (4) holds, Φ has column rank equal to two, with two unknowns. Therefore there exists a unique solution to (6) using the inverse or pseudoinverse.

If there do not exist $i, j \in [n]$ and $k_1, k_2 \in [T-1] \cup \{0\}$ such that (4) holds, then Φ has a nontrivial nullspace. Therefore (8) does not have a unique solution. \square

Now we present two corollaries where $h\beta$ and $h\delta$, denoted by β_h and δ_h , respectively, can be recovered.

Corollary 1. Consider the model in (2) under Assumptions 1-4 with homogeneous spread. Assume that A and x^k , for all $k \in [T] \cup \{0\}$, are known. Then, β_h and δ_h can be identified uniquely if and only if $T > 0$ and there exist $i, j \in [n]$ and $k_1, k_2 \in [T-1] \cup \{0\}$ such that $x_i^{k_1} g_j(x^{k_2}) \neq x_j^{k_2} g_i(x^{k_1})$.

Proof. Since h factors out of the right hand side of (6) and is nonzero by Assumption 4, even if h is not known, a scaled version of β and δ can be recovered exactly. Therefore the proportion of the two parameters can be found. \square

If the assumption is made that the underlying spread process is heterogeneous, we have a similar condition.

Theorem 2. Consider the model in (1) under Assumptions 1-4. Assume that x^k , for all $k \in [T-1] \cup \{0\}$, x_i^T , A , and h are known. Then, the parameters of the spreading process for node i can be identified uniquely if and only if $T > 1$, and there exist $k_1, k_2 \in [T-1] \cup \{0\}$ such that

$$x_i^{k_1} (1 - x_i^{k_2}) \sum_{j=1}^n a_{ij} x_j^{k_2} \neq x_i^{k_2} (1 - x_i^{k_1}) \sum_{j=1}^n a_{ij} x_j^{k_1}. \quad (7)$$

Proof. Since x^k , for all $k \in [T-1] \cup \{0\}$, and A are known, we can construct the matrix Φ_i , defined as,

$$\Phi_i := \begin{bmatrix} (1 - x_i^0) \sum_{j=1}^n a_{ij} x_j^0 & -x_i^0 \\ \vdots & \vdots \\ (1 - x_i^{T-1}) \sum_{j=1}^n a_{ij} x_j^{T-1} & -x_i^{T-1} \end{bmatrix}.$$

Then, since we also know x_i^T and h , we have

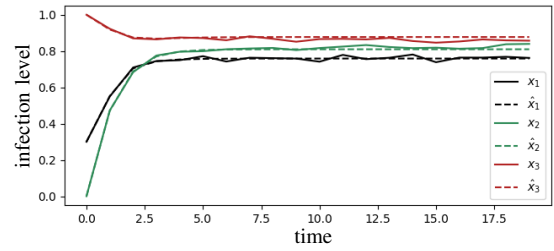
$$\begin{bmatrix} x_i^1 - x_i^0 \\ \vdots \\ x_i^T - x_i^{T-1} \end{bmatrix} = h \Phi_i \begin{bmatrix} \beta_i \\ \delta_i \end{bmatrix}. \quad (8)$$

Since $T > 1$, Φ_i has at least two rows. By the assumption that there exist $k_1, k_2 \in [T-1] \cup \{0\}$ such that (7) holds, Φ_i has column rank equal to two, with two unknowns. Therefore there exists a unique solution to (8) using the inverse or pseudoinverse.

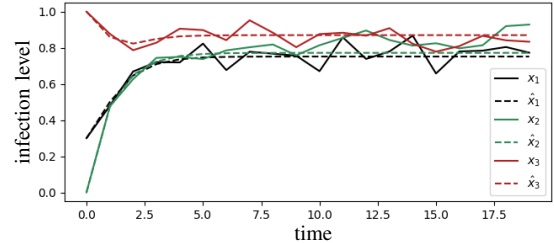
If there do not exist $k_1, k_2 \in [T-1] \cup \{0\}$ such that (7) holds, then Φ_i has a nontrivial nullspace. Therefore (8) does not have a unique solution. \square

IV. SIMULATIONS

In this section, we present a number of illustrative simulations and a corollary based off those simulations. Unless otherwise stated, we use the following parameters in our simulations



(a) $\sigma = 0.01$



(b) $\sigma = 0.05$

Fig. 2: Simulation of the epidemic states of a heterogeneous system with additive i.i.d. Gaussian noise and recovered states using recovered heterogeneous spread parameters.

$$\begin{aligned} x_0 &= [0.3 \ 0.0 \ 1.0]^T, \\ \beta &= [1.5 \ 1.1 \ 0.8], \\ \delta &= [0.8 \ 0.4 \ 0.2], \text{ and} \\ A &= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \end{aligned}$$

We estimate the spread parameters using the pseudo-inverse with either (6) or (8) to recover homogeneous or heterogeneous spread parameters, $\hat{\beta}, \hat{\delta}$ or $\hat{\beta}_i, \hat{\delta}_i$, respectively. The recovered states, \hat{x} , are simulated as

$$\hat{x}_i^{k+1} = \hat{x}_i^k + h \left((1 - \hat{x}_i^k) \hat{\beta}_i \sum_j a_{ij} \hat{x}_j^k - \hat{\delta}_i \hat{x}_i^k \right), \quad (9)$$

and unless otherwise stated, is assumed the a_{ij} 's and the initial condition are known.

A. Recovering from Noisy Data

Additive system noise can be used to take into account the effect of external impacts on the spread of a virus [16]. To investigate the usefulness of the discrete-time spread model for applications, we ran simulations with additive i.i.d. Gaussian system noise. Figure 2 shows simulations of a heterogeneous system with the states generated by

$$x_i^{k+1} = x_i^k + h \left((1 - x_i^k) \beta_i \sum_j a_{ij} x_j^k - \delta_i x_i^k \right) + \varepsilon_i, \quad (10)$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ are i.i.d for all $i \in [n]$ and a floor and ceiling function are used to ensure that each state remains in $[0, 1]$. The simulations in Figure 2 demonstrate that even in the presence of noise, the behavior of the spread process can be captured quite well.

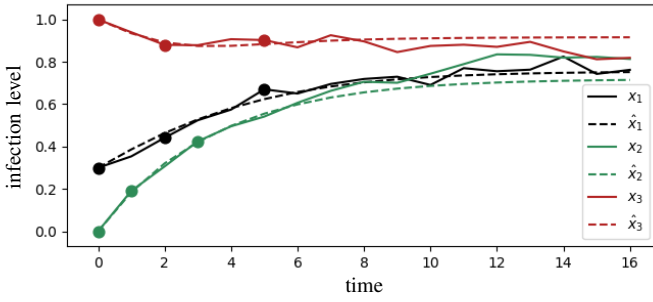


Fig. 3: Simulating data loss in the epidemic states of a heterogeneous system with additive i.i.d. Gaussian noise and a variance of 0.03 . Recovering heterogeneous parameters using the data points marked by circles.

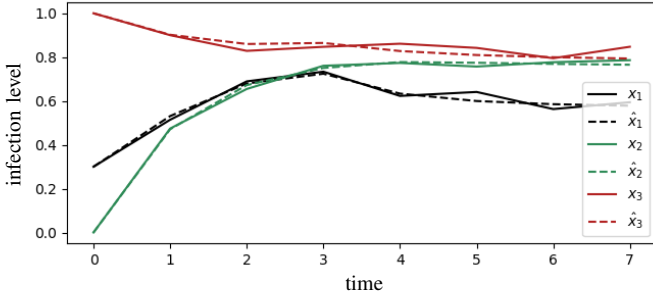


Fig. 4: Simulation of the epidemic states of a heterogeneous system with time-varying graph and additive i.i.d. Gaussian noise with a variance of 0.03, and recovered states using recovered heterogeneous spread parameters.

B. Recovering with Data Loss

In most cases when applying the model to a real world dataset, the dataset has some data loss and therefore it is important to investigate how well the recovery process performs in that case. To test the effect of data loss on the model, we simulated a dataset, removed some of the data, and used linear interpolation to fill in the missing data points. The model did very well recovering the states that were generated with no data loss, x , by using the interpolated dataset considering the number of data points that were used for recovery. The simulation is illustrated in Figure 3. We also see that the recovery becomes less accurate as σ increases and that the recovery is significantly worse for $\sigma = 0.1$ compared to the case where $\sigma = 0.03$. However, the decreased performance is due more to the increase in the variance of the noise than the data loss. Therefore, we conclude that the recovery process is relatively insensitive to data loss compared to the noise, indicating that the recovery process can be effective for systems where lots of data loss may occur as long as the system noise is not too large.

C. Recovering with Time-varying Graph

In a human social network, the interconnection structure is often dynamic and usually changes from day to day, week to week or month to month [16]. If one is able to measure the changes in the network, the recovered states, \hat{x} , can be improved upon by incorporating this information.

Therefore, there are cases where it might be more interesting to use a dynamic graph rather than a static one. With a time-varying graph the SIS model with noise in (10) becomes

$$x_i^{k+1} = x_i^k + h \left((1 - x_i^k) \beta_i \sum_j a_{ij}^k x_j^k - \delta_i x_i^k \right) + \varepsilon_i \quad (11)$$

where $\varepsilon_i \in \mathcal{N}(0, \sigma^2)$ are i.i.d. for all $i \in [n]$. We use (8) for recovery but incorporating the time-varying network structure information by modifying Φ_i to become

$$\begin{bmatrix} h(1 - x_i^0) \sum_{j=1}^n a_{ij}^0 x_j^0 & -h x_i^0 \\ \vdots & \vdots \\ h(1 - x_i^{T-1}) \sum_{j=1}^n a_{ij}^{T-1} x_j^{T-1} & -h x_i^{T-1} \end{bmatrix}. \quad (12)$$

This formulation leads to the following corollary of Theorem 2.

Corollary 2. Consider the model in (1) under Assumptions 1-4 with a possible time-varying network, denoted by A^k . Assume that x^k , for all $k \in [T-1] \cup \{0\}$, x_i^T , A^k , and h are known. Then, the parameters of the spreading process for node i can be identified uniquely if and only if $T > 1$, and there exist $k_1, k_2 \in [T-1] \cup \{0\}$ such that

$$x_i^{k_1} (1 - x_i^{k_2}) \sum_{j=1}^n a_{ij}^{k_2} x_j^{k_2} \neq x_i^{k_2} (1 - x_i^{k_1}) \sum_{j=1}^n a_{ij}^{k_1} x_j^{k_1}.$$

Proof. The result follows directly the same as the proof of Theorem 2 after replacing Φ_i in (8) with (12). \square

We simulated this system in Figure 4 where the edge between Agent 1 and Agent 2 was removed from the complete graph at time step $k = 3$. From the simulations we see that the recovery works very well even when there is noise in the system, assuming we know the graph structure exactly for each time step. Consequently, the simulations indicate that the model can capture the progress of the states well, even when the network graph changes over time, which is common in human social networks, and this observation is important for many practical applications. However, it may be difficult to accurately measure time-varying graphs in some cases.

D. Recovering with Quantized States

When collecting data, it is often binned, allowing for respondents to choose a certain level of pain or sickness, such as from one to ten. To further investigate the usefulness of the SIS model in applications we performed simulations recovering the spread parameters from observed states that had been quantized. The simulations were carried out by generating states, x using (10) which were quantized into five evenly distributed values between 0 and 1, \tilde{x} . The quantized states, \tilde{x} were then used in (8) to recover the spreading parameters. Clearly, the accuracy of the recovered states, \hat{x} depends on the number of quantized states that are chosen,

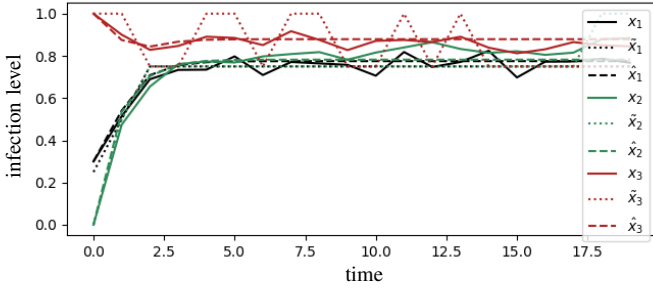


Fig. 5: Simulation of the SIS model with quantized states which were evenly distributed between 0 and 1. The variance of the noise in x was 0.003.

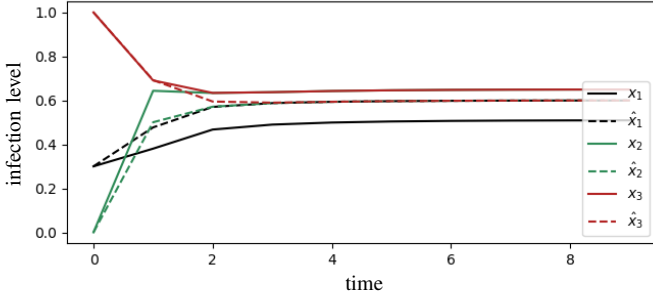


Fig. 6: Simulation of the epidemic states of a heterogeneous system with additive i.i.d. Gaussian noise and a variance of 0.03 and recovered states using a homogeneous approximation of the system.

however, here we show, in Figure 5, that five quantized states makes the recovered states fairly accurate and might be enough for many applications.

E. Heterogeneous Spread Approximated by Homogeneous Model

At times there may not be rich enough data to satisfy the requirements of Theorem 2 or it may be assumed incorrectly that a heterogeneous system is homogeneous. To understand how accurately a heterogeneous system can be approximated by a homogeneous model we use $T = 4$ to learn homogeneous spread parameters. A simulation of a system with heterogeneous disease parameters, x , and recovered states, \hat{x} , is shown in Figure 6. The recovered states, \hat{x} , was generated by recovering a homogeneous approximation of the spread parameters, $\hat{\beta}$ and $\hat{\delta}$. To generate x we use $\beta = [0.4 \ 0.8 \ 0.8]$ and $\delta = [0.5 \ 0.5 \ 0.5]$. The learned parameters are

$$\begin{bmatrix} \hat{\delta} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} 0.6223 \\ 0.4980 \end{bmatrix}. \quad (13)$$

One can see that the error between the recovered states, \hat{x}_1 and the original system, x_1 , is higher than the errors for Agent 2 and 3. The decreased accuracy of \hat{x}_1 can be explained by the difference in magnitude of β_1 from the infection rates of the other agents. The simulations suggest that in many cases we may be able to effectively approximate a heterogeneous system with homogeneous spread parameters, and that the individual characteristics of each node

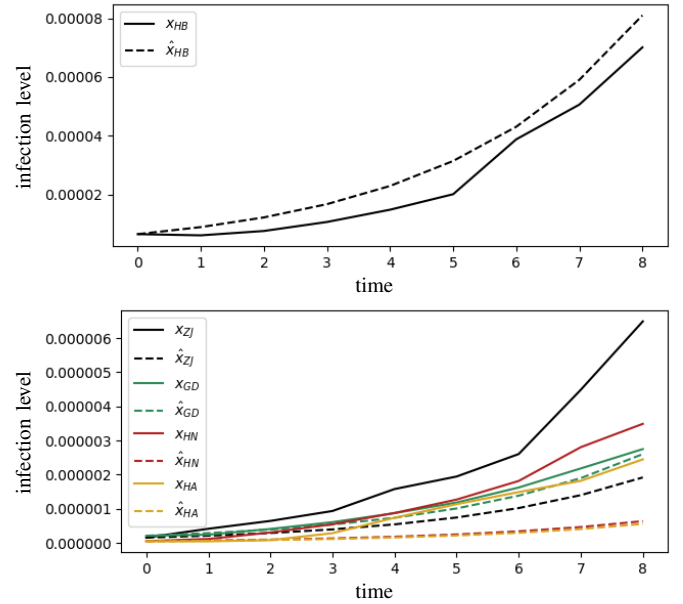


Fig. 7: Data fitting results using the parameters in (14), which were estimated with constraints on the parameters to assure Assumptions 2 and 3 were met.

could be approximated by some sort of average susceptibility and healing strength. The spreading would then be fully determined by the graph structure and the two estimated homogeneous disease parameters.

V. CORONAVIRUS DATASET

In this section we illustrate the challenges of using real data by implementing the estimation technique on the 2019-nCoV dataset collected in [12], from January 22, 2020 until January 30, 2020. We treat the subpopulation of each Chinese provincial-level entity (including Hong Kong, Taiwan, and Macau) as a node in the network. The state of the model is the proportion of current confirmed cases of 2019-nCoV in each subpopulation for each day. The number of current confirmed cases is calculated by subtracting the number of recovered cases from the total number of confirmed cases. While it is clear that the confirmed cases are being under-reported, the number of recovered cases is drastically below what we believe to be the real number. This deficiency presents difficulties when trying to accurately estimate the healing rate.

We factor the adjacency matrix into several parts: 1) self loops, that is, interactions inside each province, 2) passenger trains between provinces (scaled by speed: 3 High-Speed Railway, 2 Rapid Railway, 1 Regular Train [17]), 3) borders (reflecting travel between provinces on roads), and 4) flights (number of daily flights between the capitals of each node [18]). To each of the four adjacency matrices we assign an infection parameter, labeled β_s , β_t , β_b , and β_f , respectively.

Estimating the parameters using (6) resulted in parameter values that did not satisfy Assumptions 2 and 3, which results in the states leaving the sensible domain of the model, as defined in (3). Therefore we found parameters to minimize

the error in (6) using a convex solver, while constraining the parameters to meet Assumptions 2 and 3 and to keep δ small, giving

$$\begin{bmatrix} \hat{\beta}_s \\ \hat{\beta}_t \\ \hat{\beta}_b \\ \hat{\beta}_f \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} 0.372 \\ 4.64 \times 10^{-9} \\ 1.46 \times 10^{-9} \\ 1.17 \times 10^{-10} \\ 5.00 \times 10^{-5} \end{bmatrix}. \quad (14)$$

The overall error, measured as the Frobenius norm of the difference between the simulations using (14) and the real data, is 2.73×10^{-5} . This error is skewed by the small magnitude of the states, given the large population sizes. Dividing by the Frobenius norm of the data increases the error to 0.262, showing that the system has approximately 26% error. To illustrate the accuracy of the estimation technique we depict the times series from the data (x) and the simulations (\hat{x}) for the provinces with the highest number of confirmed cases: Hubei (HB), Zhejiang (ZJ), Guangdong (GD), Hunan (HN), and Henan (HA); see Figure 7. In order to estimate when the outbreak began, we used the parameters from (14) and initialized the system with one resident of Hubei infected. We then found the day when \hat{x}_{HB} was equal to double of the initial x_{HB} from January 22, 2020, therefore assuming that half of the cases were not reported/confirmed. Subtracting the number of days until that point indicated that, according to this model, the first human became sick in December.

The results do not capture the behavior of the outbreak very well using homogeneous parameters. The results from Section IV-E indicate that the heterogeneous parameters are most likely very different between provinces. Further, the limited number of datapoints restrict our ability to learn all of the heterogeneous spread behaviors. As noted earlier, another major challenge of this dataset is the incompleteness of the measurements. In order to correctly estimate the spreading parameters, better measurements of the confirmed and recovered cases are required. Alternatively, new data processing techniques can be developed in future work to deal with these issues.

VI. CONCLUSION

In this paper we presented how to employ time series data to estimate the spreading parameters of a discrete time networked virus spread model. We explored the effect of system noise, missing data, time-varying network structure, and quantization of the measurements. We presented a corollary as a result of the simulations. We also illustrated how well the behavior of a heterogeneous model can be captured by a homogeneous model, which is useful if there is not enough data to learn the heterogeneous spreading parameters. Finally, we highlighted the practical importance of this work by studying the ongoing coronavirus outbreak in China. Through our estimation of the virus spread dynamics we projected that the outbreak may have begun in December. For future work, we would like to further develop data processing techniques for this type of data and these models. We also

hope to further analyze this outbreak as more, and cleaner, data becomes available.

REFERENCES

- [1] World Health Organization, “Novel coronavirus (2019-nCoV),” <https://www.who.int/westernpacific/emergencies/novel-coronavirus>, accessed: 2020-01-27.
- [2] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, “Epidemic spreading in real networks: An eigenvalue viewpoint,” in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems*. IEEE, 2003, pp. 25–34.
- [3] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, “Epidemic thresholds in real networks,” *ACM Transactions on Information and System Security (TISSEC)*, vol. 10, no. 4, p. 1, 2008.
- [4] H. J. Ahn and B. Hassibi, “Global dynamics of epidemic spread over complex networks,” in *Proceedings of the 52nd IEEE Conference on Decision and Control (CDC)*, 2013, pp. 4579–4585.
- [5] P. E. Paré, B. E. Kirwan, J. Liu, T. Başar, and C. L. Beck, “Discrete-time spread processes: Analysis, identification, and validation,” in *Proceedings of the American Control Conference*, 2018.
- [6] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, , and T. Başar, “Discrete time virus spread processes: Analysis, identification, and validation,” *IEEE Transactions on Control Systems Technology*, vol. 28, no. 1, pp. 79–93, 2020.
- [7] A. Melo, C. L. Beck, J. I. Pena, and P. E. Paré, “Knowledge transfer from universities to regions as a network spreading process,” in *Proceedings of the IEEE International Systems Engineering Symposium (ISSE)*, 2018.
- [8] B. Prasse and P. Van Mieghem, “Network reconstruction and prediction of epidemic outbreaks for NIMFA processes,” *arXiv preprint arXiv:1811.06741*, 2018.
- [9] —, “The viral state dynamics of the discrete-time NIMFA epidemic model,” *arXiv preprint arXiv:1903.08027*, 2019.
- [10] N. Slighton, J. M. Rico, E. Kallfelz, J. Qi, and C. G. Brinton, “A network-driven approach to modeling the spread of Ebola-type epidemics,” in *52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018.
- [11] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell, “Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions,” *medRxiv*, 2020.
- [12] L. Gardner, “2019-nCoV global cases,” 2020, <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>.
- [13] Center for Disease Control, “2019 novel coronavirus (2019-nCoV),” <https://www.cdc.gov/coronavirus/>, accessed: 2020-01-27.
- [14] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, and T. Başar, “Erratum to analysis, identification, and validation of discrete-time epidemic processes,” *IEEE Transactions on Control Systems Technology*, 2020, to appear.
- [15] P. E. Paré, D. Vrabac, H. Sandberg, and K. H. Johansson, “Analysis, online estimation, and validation of a competing virus model,” in acceptance to the *American Control Conference*, 2020, arXiv:2001.10532 [eess.SY].
- [16] P. E. Paré, C. L. Beck, and A. Nedić, “Epidemic processes over time-varying networks,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1322–1334, 2018.
- [17] K. Peng, “China railway map,” <http://cnrail.geogv.org/enus/>, accessed: 2020-01-28.
- [18] O. T. Network, “Air travel schedule,” http://jipiao.oklx.com/cn.airfield_schedule.aspx, accessed: 2020-01-29.