Network Inference-Based Prediction of Epidemics: A Case Study on Mexican State COVID-19 Infection Counts

BEITSKE FLAKE, University of Twente, The Netherlands

With the world recently having suffered from the global COVID-19 pandemic, it created a necessity to predict the spread of this virus and that of possible future epidemics. Predicting the spread of the virus, and understanding the way a virus interacts within individuals of a population, can contribute to its understanding as well as the effectiveness of counter measures. In an attempt to make these predictions it is possible to use the existing COVID-19 data, including time series of infection counts gathered during the pandemic. A method to infer an interaction network from this data, and make predictions on the future dynamics of this network, is the Network Inference-based Prediction Algorithm (NIPA). This paper aims to infer the COVID-19 interaction network from the daily infection data of the states of Mexico using NIPA. The SIR (Susceptible Infected Removed) epidemic model is applied to capture the dynamics of the COVID-19 spread within each state. We exploit the inferred interaction network in an attempt to estimate the interaction patterns between states, and compare those with the observations from past COVID-19 outbreaks. Finally, we assess the results produced by the inferred infection matrix, and explain how they reflect on different aspects of a virus spreading in the real world, such as via international visitors and tourism.

Additional keywords and phrases: Epidemic Predictions, Network Inference-based Prediction Algorithm (NIPA), SIR compartmental model, COVID-19, Interaction Networks, Time Series

1 INTRODUCTION

Around November 2019, the corona virus or COVID-19 emerged in Hubei, China, and quickly spread to every continent [1]. COVID-19 was declared as a pandemic by the World Health Organisation in March 2020. In the aim to contain the spread, be it nationwide or worldwide, it became necessary to develop predictive models and other forecasting methods to predict the spread of this virus. These many epidemic forecasting methods done in earlier studies include deep learning models, neural network powered models, compartmental models and many more epidemic models or spread phenomena forecasting algorithms [2–5], such as the Susceptible-Infected-Susceptible (SIS) model and the Susceptible-Infected-Recovered (SIR) model [2, 5, 6].

In 2020 Prasse and Van Mieghem researched a new way to predict the dynamics of a network apart from the network topology. They proposed the Network Inference-based Prediction Algorithm (NIPA): a method to infer a network of interactions from time series data, resulting in highly accurate predictions of the network dynamics in their test cases [5]. Six models on dynamic networks were studied as an 'interaction function' for the adjacency matrix, including Susceptible-Infected-Susceptible epidemics (SIS), which can describe the spreading phenomena of epidemics [6].

TScIT 39, July 7, 2023, Enschede, The Netherlands

Building on the network reconstruction done with the SIS model, a case study was made for the Chinese province Hubei by Prasse et al. where they applied a "network-based SIR epidemic model to predict the outbreak of the COVID-19 virus for each city" [7]. NIPA could successfully forecast the spread of COVID-19 in the province, but the underlying infection matrix could not be inferred [5, 8].

A comparison of the accuracy of NIPA against NIPA variations and other prediction algorithms was made on two case studies: the spread of COVID-19 in cities of the province Hubei, China, and in provinces in the Netherlands [8]. It was concluded that the original NIPA performed better in prediction accuracy than any of the other compared algorithms.

Considering the promising results of the NIPA case studies [7, 8], a next step in the research of NIPA's prediction accuracy is following the procedure described by Prasse, Achterberg et al. on a COVID-19 infection network that has not been studied yet with this method. Therefore, a new case study will be conducted in this paper on the infection counts of the states of Mexico. This case study investigates the following research questions:

RQ1 *Can the NIPA algorithm be used to infer an infection matrix of Mexico's COVID-19 time series?*

RQ2 *To what extend does the inferred Mexican COVID-19 infection matrix generated by the NIPA algorithm in **RQ1** reflect on the true data from the COVID-19 virus spread in Mexico?*

As such, the purpose of this research is to apply the NIPA algorithm to infection count time series from Mexico, to find if an infection matrix can be inferred for the COVID-19 spread in the states of Mexico. This will be done by following and replicating the procedures and methods followed by Prasse et al. as described in their paper [7].

This paper is structured as followed. Firstly, some background on epidemic predictions and network inference algorithms will be provided in section 2. In section 3, the methods and approach of replicating the SIR based NIPA procedure and inferring the infection matrix will be described. The results of the case study predictions are presented in section 4. The discussion in section 5 will describe the performance and limitations of the algorithm, including potential future work. Finally, section 6 concludes this paper.

2 RELATED WORK

This research focuses on the fields of epidemic predictions and spread phenomena, using mathematical optimisation algorithms. In order to gather the related literature Mendeley, Scopus, Google Scholar and IEEE were used.

In 1927, Kermack and Mckendrick mathematically described the progress of an epidemic in a homogeneous population in the form of a mathematical investigation [9]. They described several aspects of the epidemic dynamics in regard to the population including infectivity rate, transmission rate, population density and susceptibility of the population.

Their work eventually led to the Susceptible-Infected-Recovered (SIR) epidemic model [10, 11], in which the dynamics of a spreading virus within a population of individuals can be described, where these individuals were divided into three compartments of the population. The spread of COVID-19 can be described with this model as well, albeit not exactly [8, 12]. One of the aspects that was found lacking was accurate predictions over longer periods of prediction times. This is partly because the SIR model evolves in discrete time, whereas the COVID-19 pandemic evolves in continuous time. Another problem found with the model was that it is unable to describe phenomena like lockdowns or the availability of vaccinations. However, several studies show that the SIR model can still be applied to COVID-19 with appropriate parameter selection including the usage of optimisation algorithms, the realisation of introduced model errors and considering different scenarios [2, 8, 13].

In the field of making epidemic predictions from inferred interactions networks we already described the work of Prasse and Van Mieghem. They concluded that with the right methodology, predictions about the general dynamics of a network could be made on an estimated network generated from the original, true network [5]. Even when the estimated network bares no topological similarity to the true network, the predictions on the dynamics were found to be accurate. The research managed to apply the SIS epidemic model as a dynamic modelling function of the NIPA method, but it was suggested that observing a sufficiently great number of time series for a virus like COVID-19 might not be viable. However, based on the work with NIPA and COVID-19 provided by Achterberg et al. [8], we will make the assumption that it is possible to observe enough time series in the Mexico data set to reconstruct the inferred adjacency matrix.

3 METHODOLOGY & APPROACH

This section describes the selected approach and underlying principles to this research.

3.1 Definitions of algorithms

3.1.1 SIR model. The Susceptible-Infected-Recovered (SIR) epidemic model describes the behaviour and spread of an infectious virus within a population of individuals [7, 14, 15]. Each individual from the population can be divided in one of three states, otherwise called compartments, at a point in time:

Susceptible The individual is healthy and not yet infected, but they could become infected. As the virus spreads, the individual may become infectious over time.

Infectious The individual has been infected, and has now become infectious to other susceptible individuals.

Recovered These individuals have recovered from the virus, either by removal (being immune, resistant or having received a cure), or having died.

The contact amongst the individuals of each fraction influences the spread and cure of the virus to the other fractions and individuals with a certain probability. These network characteristics can be expressed as parameters β (infection probability) and δ (removal or curing probability), as illustrated in fig. 1 [7, 15].

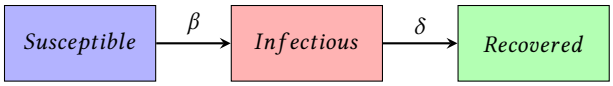

Fig. 1. SIR epidemic model transition graph

For every group i at every point in discrete time $k \in \mathbf{N}$, we can denote the three components as fractions of susceptible $S_i[k]$, infectious $I_i[k]$ and recovered $R_i[k]$ individuals, for which holds: $S_i[k] + I_i[k] + R_i[k] = 1$. In our case study, we denote a group as the population of a Mexican state or region. Following the group-based, discrete time SIR epidemic model of Prasse and Van Mieghem [7, 16], we denote the 3 x 1 viral state vector for every state i at time k as:

Definition 1. *For every region i, the viral state*

$$v_i[k] = \begin{pmatrix} S_i[k] \\ I_i[k] \\ R_i[k] \end{pmatrix} = (S_i[k], I_i[k], R_i[k])^T$$

evolves over discrete time $k = 1, 2, ..., n$.

At any time k, an individual of region i can change from compartment $S_i[k]$ to compartment $I_j[k]$, which denotes the infected fraction of individuals of region j. This transition can happen with probability $\sum_{j=1}^{N} \beta_{ij} I_j[k]$. Therefore, the viral state $v_i[k]$ evolves over discrete time according to:

Definition 2. *SIR epidemic model [7, 15, 16]. For every region i, the viral state $v_i[k]$ evolves in discrete time $k = 1, 2, \ldots$ according to:*

$$I_i[k+1] = (1 - \delta_i) I_i[k] + (1 - I_i[k] - R_i[k]) \sum_{j=1}^{N} \beta_{ij} I_j[k],$$

$$R_i[k+1] = R_i[k] + \delta_i I_i[k],$$

$$S_i[k] = 1 - I_i[k] - R_i[k].$$

Here, β_{ij} denotes the infection probability from region i to region j, and δ_i denotes the curing probability of region i.

Neither the curing probabilities δ_i nor the infection probabilities β_{ij} are known for the COVID-19 epidemic, so we consider no a priori knowledge on both of them in this research. Using NIPA and LASSO, estimations $\hat{\delta}_i$ and $\hat{\beta}_{ij}$ of the unknown spreading parameters δ_i and β_{ij} will be made to reconstruct the infection network, based on the reported number of infected individuals.

3.1.2 NIPA algorithm. NIPA, or the Network Inference-based Prediction Algorithm, is used to process the SIR time series and produce the estimates $\hat{\delta}_i$ and $\hat{\beta}_{ij}$ of the unknown spreading parameters. The SIR time series $v_i[1], ..., v_i[n]$ are obtained by processing raw data of the confirmed number of infection counts of every Mexican state. NIPA exists of three steps [5, 7, 8]:

(1) **Data preprocessing** The raw data of confirmed numbers of infected individuals are processed to obtain an SIR time series $v_i[1], ..., v_i[n]$ of the viral state for every region i. We denote the discrete time as $k = 1, 2, ...n$, where n is the total number of observed days, and the first reported infection case as day $k = 1$.

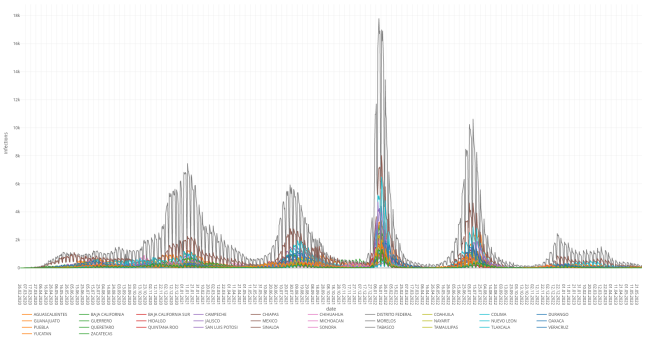

Fig. 2. Reported infections per day per Mexican state

(2) **Network inference** Based on the time series $v_i[1], ..., v_i[n]$ obtained by the data processing, the estimates $\hat{\delta}_i$ and $\hat{\beta}_{ij}$ of the unknown spreading parameters δ_i and β_{ij} are obtained. These estimates are obtained by network inference, where the goal is to estimate and construct the infection matrix B of infection probabilities from the SIR viral state observations. The details of our network inference approach are illustrated in section 3.4. The network or true adjacency matrix B is inferred by the LASSO [5, 8, 17].

(3) **Iterating SIR model** Finally, the estimates $\hat{\delta}_i$ and $\hat{\beta}_{ij}$ from the inferred infection matrix result in an SIR model. This model is iterated for future times k to predict the evolution and thus the spread of the virus.

Pseudo-code for the NIPA algorithm is provided in Appendix A.

3.2 Infection counts data set

For the raw infection counts data, we are working with the daily reported infection numbers for every Mexican state provided by Mexican research center Conacyt [18]. The reported infections from the data set start at 26-02-2020 and end on 15-05-2023, covering the 32 states of Mexico over a 3 year time period. Therefore, the initial time $k = 1$ corresponds to February 26, 2020. The data set also contains the population size p_i of every state. Additional visualisation of the data set is provided in Appendix B.

3.3 Preprocessing data

The next step is obtaining the reported fraction of infections time series in every region $i = 1, ..., N$ from the reported infected individuals $N_{rep,i}[k]$, needed for the viral state vector from definition 1. We obtain the fraction of infected individuals $I_{rep,i}[k]$ in region i at time k as follows:

$$I_{rep,i}[k] = N_{rep,i}[k]/p_i$$

As shown in figure 2, the data exhibits fluctuations around every 7 day cycle. We make the assumption that this is caused by fewer reports of COVID-19 infections during the weekend, as every dip is on average around weekend days. To compensate for these fluctuations, we apply a rolling average with a 7 day period on the reported

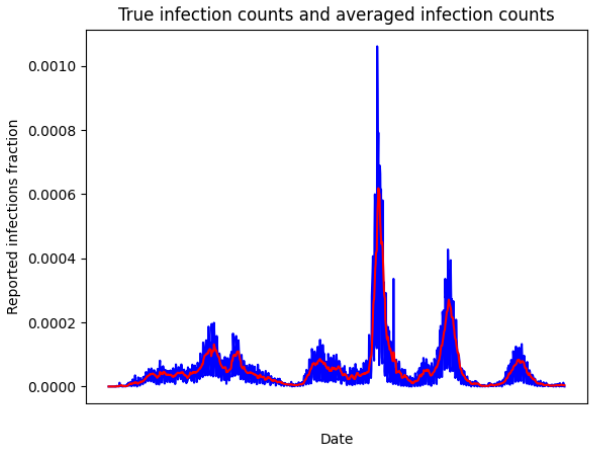

Fig. 3. The infection counts graph for the state Aguascalientes over time, with the true data points in blue and the averaged data points in red.

time series, using the `rolling mean` command from the Python library Pandas. The effect is illustrated in figure 3.

Based on the reported number of infection counts $N_{rep,i}[k]$, our goal is to obtain an SIR viral state vector for every region i as defined in (1). Because $S_i[k] = 1 - I_i[k] - R_i[k]$ and the fraction of infectious individuals $I_i[k]$ follows from $I_{rep,i}[k]$ [7], it suffices to determine the fraction $R_i[k]$. However, $R_i[k]$ is not known. We can assume that at the initial time $k = 1$ $R_i[1] = 0$ holds. And from definition 2 we can calculate $R_i[k]$ for any time $k \geq 2$ as long as the curing probability δ_i is known, which it is not.

Hence, following the procedure of Prasse et al. [7], we consider 50 equidistant candidate values for δ_i, ranging from $\delta_{min} = 0.01$ to $\delta_{max} = 1$. The set of candidate values is defined as $\Omega = \{\delta_{min}, ..., \delta_{max}\}$, and for every candidate value $\delta_i \in \Omega$ the fraction $R_i[k]$ follows from (2), leading to 50 potential sequences $R_i[1], ..., R_i[n]$. The curing probability δ_i and its corresponding sequence $R_i[1], ..., R_i[n]$ is estimated as the element in Ω that resulted in the best fit of the SIR model (2), and therefore the first step in the network inference phase.

3.4 Network inference

Our goals is to infer the adjacency matrix, otherwise called infection matrix, from the observed infection counts for all the regions N. As the infection probability β_{ij} specifies the contacts of individuals between region i and region j, the contact network or infection adjacency matrix is given by the following $N \times N$ matrix:

Definition 3. $$B = \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{N1} & \beta_{N2} & \dots & \beta_{NN} \end{pmatrix}$$

Prasse et al. observe from the SIR equations in (2) that even though β_{ij} appears linearly, the SIR state variables S_i, I_i and R_i do not [5, 7]. Therefore, from (2), the infection probabilities β_{ij} satisfy:

Definition 4. $V_i = F_i \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iN} \end{pmatrix}$ *for all regions i = 1, ..., N.*

The $(n-1) \times 1$ vector V_i and the $(n-1) \times N$ matrix F_i are given by:

Definition 5. $V_i = \begin{pmatrix} I_i[2] - (1-\delta_i) I_i[1] \\ \vdots \\ I_i[n] - (1-\delta_i) I_i[n-1] \end{pmatrix}$

and

$$F_i = \begin{pmatrix} S_i[1]I_i[1] & \dots & S_i[1]I_N[1] \\ \vdots & \ddots & \vdots \\ S_i[n-1]I_i[n-1] & \dots & S_i[n-1]I_N[n-1] \end{pmatrix}$$

3.4.1 Least absolute shrinkage and selection operator (LASSO). As this paper takes a network inference approach based on the research done by Prasse and Van Mieghem [5], we apply a variation of the LASSO to the linear system as described in (4) [7, 17]. For each given row i, we solve the LASSO to find the set $\beta_{i1}, ..., \beta_{iN}$ that minimizes the quadratic error of the linear system (4):

Definition 6. $\min_{\beta_{i1},\dots,\beta_{iN}} \left\| V_i - F_i \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iN} \end{pmatrix} \right\|_2^2 + \rho_i \sum_{j=1, j\neq i}^{N} \beta_{ij}$

The sum in the objective of (6) is an λ_1-norm regularisation term to avoid overfitting [7, 17]. The regularisation parameter in the LASSO is given by ρ_i. To determine the correct regularisation parameter, we consider 100 candidate values, specified by the set $\Theta_i = \{\rho_{min,i}, ..., \rho_{max,i}\}$. For every value of $\rho_i \in \Theta_i$, we compute the Mean Squared Error $MSE(\delta_i, \rho_i)$ by 3-fold-cross-validation.

The rows of V_i and F_i are divided into a training set $F_{i,train}, V_{i,train}$ and a test set $F_{i,test}, V_{i,test}$. The test set is set to be 30% of the original data set. Under these parameters, we compute the infection probability solution $\beta_{i1}, ..., \beta_{iN}$ to the LASSO (6) on the training set of every fold $F_{i,train}, V_{i,train}$. This yields a $MSE(\delta_i, \rho_i)$ that equals:

Definition 7. $MSE(\delta_i, \rho_i) = \left\| V_i - F_i \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iN} \end{pmatrix} \right\|_2^2$

The final estimate $\beta_{i1}(\delta_i), ..., \beta_{iN}(\delta_i)$ for the infection probabilities of infection matrix B (3) is obtained by solving the LASSO (6) on the whole matrix F_i and vector V_i. To solve the LASSO, the `Scikit-learn` library's `linear_model Lasso` is used with the regularisation parameter ρ_i as the `alpha` value [19].

4 RESULTS

We used the preprocessed data set as described in section 3, with $N = 32$ states, $n = 875$ days and $k = 1$ set to February 26, 2020. Running the NIPA algorithm over this dataset to infer the adjacency matrix B took ≈ 8.5 hours. The candidate values for δ_i and ρ_i were set as described in section 3. The algorithm presented us the following estimates $\beta_{i1}(\delta_i), ..., \beta_{iN}(\delta_i)$ with $N = 32$:

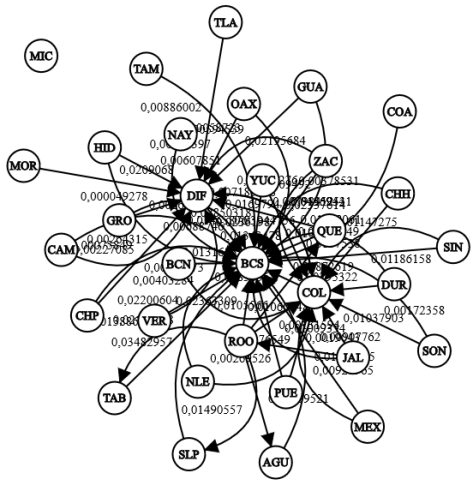

Fig. 4. Infection interaction network between all the states of Mexico, with the infection probability β_{ij} added as weight to the edges.

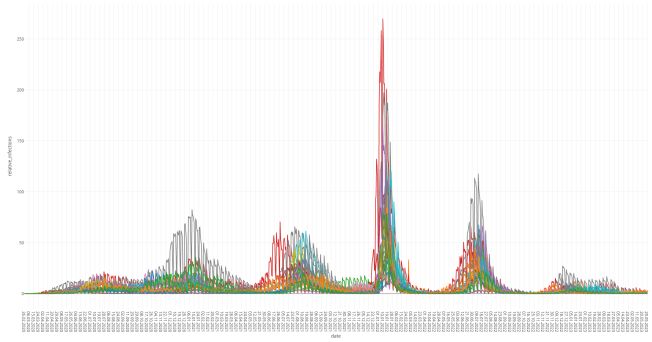

Fig. 5. Relative infection counts over time for each mexican state.

$$B = \begin{pmatrix} 0 & 0 & 0,01539715 & \dots & 0 \\ 0 & 0 & 0,01316153 & \dots & 0 \\ 0 & 0 & 0,17103391 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0,0169794 & \dots & 0 \end{pmatrix}$$

where B is an 32×32 matrix. The full estimated rows are presented in a heatmap in figure 6.

We put these results back into the context of the infection network, where the infection probabilities in infection matrix B form the weight of the edge between two region nodes i and j. It is possible for a node to have an edge directed to itself. The network is shown in figure 4.

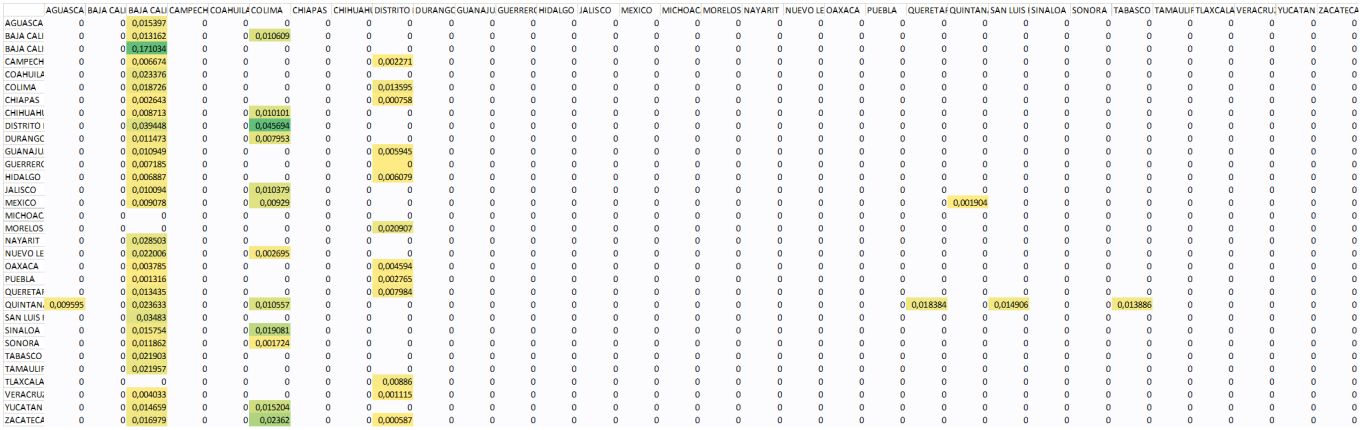

Fig. 6. Heatmap of all estimated infection probabilities $\beta_{i1}, ..., \beta_{iN}$ for each state i.

5 DISCUSSION & FUTURE WORK

Some observations can be made from the inferred interaction network in figure 4, as well as the results in the infection matrix B in figure 6. First of all, the three states Baja California Sur, Colima and Distrito Federal have by far the most incoming edges from other states, with Baja California Sur having incoming infection rates from almost every other state, including itself. Furthermore, the state Quintana Roo has 6 outgoing edges to other states, which is more than all the other states. If we compare these weights to the true, relative data in figure 5, we observe that all these aforementioned states have reported relatively high infection counts over time. This observation therefore supports the assumption that high numbers of infection counts in a region are reflected in the generated infection matrix, and thus in the final interaction network.

Moreover, the aforementioned states have a high touristic attraction, as well as international airports. Even though local governments in Mexico took measures like suspending flights and restricting the tourism industry over the course of the pandemic [20], there are also many cases reported where there were little restrictions to (international) tourism [21]. Moreover, there were cases where these restrictions were present locally, but business and tourism still thrived, albeit at a lower scale. It was also shown that international tourism increased noticeably at the end of 2020 [21, 22], and that touristic visitors from the US continued to arrive at Mexican touristic destinations over the course of the pandemic. Therefore, we put forward that the shown infection interactions from states like Baja California Sur and Quintana Roo might be explained by the interactions and dynamics of (international) visitors during COVID-19, therefore reflecting on the true situation.

Another observation is that most of these states have a 0 infection probability weight calculated for the other states, and that all the weights of the infection matrix B from the results are very small. This might be explained by the observation that we begin with very small fractions caused by small infection numbers in large state populations. Even though an infection matrix could be inferred for these fractions, a solution for these small fractions might have to be found in the future.

Furthermore, it was expected that most states would have an infection interaction > 0 with themselves, but in the results this is only the case for Baja California Sur. This might also be explained by the observation that we are working with very small fractions, and that these interaction parameters therefore cannot be fully estimated by the LASSO.

6 CONCLUSION

In this research, we attempted to infer the infection matrix B from the Mexican COVID019 infection count time series using NIPA. We were able to infer an infection matrix from this data, although a lot of the infection probabilities in the matrix were 0. We attribute these results to the small numbers we are working with during the network inference. More research is needed to find out how to infer these very small fractions with better results.

Furthermore, from the data that the infection network provided us, it could be concluded that the observed patterns were a reflection of Mexico's true interactions between regions. Therefore, we suggest that NIPA is still a promising way to infer the dynamics of a virus on a larger set of time series.

ACKNOWLEDGEMENT

We would like to thank Alberto Garcia-Robledo and Mahboobeh Zangiabady for all their hard work and useful contributions during the time of this research.

REFERENCES

[1] V. J. Munster, M. Koopmans, N. van Doremalen, D. van Riel, and E. de Wit, "A novel coronavirus emerging in china - key questions for impact assessment," *The New England journal of medicine*, vol. 382, pp. 692–694, 2 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31978293/

[2] I. Rahimi, F. Chen, and A. H. Gandomi, "A review on covid-19 forecasting models," *Neural Computing and Applications*, pp. 1–11, 2 2021. [Online]. Available: https://link.springer.com/article/10.1007/s00521-020-05626-8

[3] M. Wieczorek, J. Siłka, and M. Woźniak, "Neural network powered covid-19 spread forecasting model," *Chaos, Solitons & Fractals*, vol. 140, p. 110203, 11 2020.

[4] R. G. da Silva, M. H. D. M. Ribeiro, V. C. Mariani, and L. dos Santos Coelho, "Forecasting brazilian and american covid-19 cases based on artificial intelligence coupled with climatic exogenous variables," *Chaos, Solitons & Fractals*, vol. 139, p. 110027, 10 2020.

[5] B. Prasse and P. V. Mieghem, "Predicting dynamics on networks hardly depends on the topology," 5 2020.

Algorithm 1 `Network Inference-based Prediction Algorithm (NIPA)`

Input: reported fraction of infections $\mathcal{I}_{\text{rep},i}[1], ..., \mathcal{I}_{\text{rep},i}[n]$ for all cities i; prediction time n_{pred}
Output: predicted fraction of infections $\hat{\mathcal{I}}_i[n+1], ..., \hat{\mathcal{I}}_i[n+n_{\text{pred}}]$ for all cities i

Step 1 – Data preprocessing

$\mathcal{I}_{\text{rep},1}[17] \leftarrow (\mathcal{I}_{\text{rep},1}[16] + \mathcal{I}_{\text{rep},1}[18])/2$
$\mathcal{I}_i[1], ..., \mathcal{I}_i[n] \leftarrow$ `smoothdata`$(\mathcal{I}_{\text{rep},i}[1], ..., \mathcal{I}_{\text{rep},i}[n])$ for all $i = 1, ..., N$
$\mathcal{I}[k] \leftarrow (\mathcal{I}_1[k], ..., \mathcal{I}_N[k])^T$ for all $k = 1, ..., n$

Step 2 – Network inference

for $i = 1, ..., N$ **do**
  $\mathcal{R}_i[1] \leftarrow 0$
  for $\delta_i \in \Omega$ **do**
    $\mathcal{R}_i[k] \leftarrow \mathcal{R}_i[k-1] + \delta_i \mathcal{I}_i[k-1]$ for all $k = 2, ..., n$
    $\mathcal{S}_i[k] \leftarrow 1 - \mathcal{I}_i[k] - \mathcal{R}_i[k]$ for all $k = 1, ..., n$
    $v_i[k] \leftarrow (\mathcal{S}_i[k], \mathcal{I}_i[k], \mathcal{R}_i[k])^T$ for all $k = 1, ..., n$
    $(\beta_{i1}(\delta_i), ..., \beta_{iN}(\delta_i), \text{MSE}(\delta_i)) \leftarrow$ `Network inference`$(\delta_i, v_i[1], ..., v_i[n], \mathcal{I}[1], ..., \mathcal{I}[n])$
  end for
  $\hat{\delta}_i \leftarrow \underset{\delta_i \in \Omega}{\text{argmin}}\, \text{MSE}(\delta_i)$
  $(\hat{\beta}_{i1}, ..., \hat{\beta}_{iN}) \leftarrow \beta_{i1}(\hat{\delta}_i), ..., \beta_{iN}(\hat{\delta}_i)$
end for

Step 3 – Iterating SIR model

for $i = 1, ..., N$ **do**
  $\hat{\mathcal{I}}_i[n] \leftarrow \mathcal{I}_i[n]$
  $\hat{\mathcal{R}}_i[1] \leftarrow 0$
  $\hat{\mathcal{R}}_i[k] \leftarrow \hat{\mathcal{R}}_i[k-1] + \hat{\delta}_i \mathcal{I}_i[k-1]$ for all $k = 2, ..., n$
end for
for $k = n+1, ..., n+n_{\text{pred}}$ **do**
  for $i = 1, ..., N$ **do**
    $\hat{\mathcal{I}}_i[k] \leftarrow (1 - \hat{\delta}_i)\hat{\mathcal{I}}_i[k-1] + (1 - \hat{\mathcal{I}}_i[k-1] - \hat{\mathcal{R}}_i[k-1]) \sum_{j=1}^{N} \hat{\beta}_{ij} \hat{\mathcal{I}}_j[k-1]$
    $\hat{\mathcal{R}}_i[k] \leftarrow \hat{\mathcal{R}}_i[k-1] + \hat{\delta}_i \hat{\mathcal{I}}_i[k-1]$
  end for
end for

Fig. 7. NIPA algorithm [7]

Algorithm 2 `Network inference`

Input: curing probability δ_i; viral state $v_i[k]$ for $k = 1, ..., n$; infection state vector $\mathcal{I}[k]$ for $k = 1, ..., n$
Output: infection probability estimates $\beta_{i1}(\delta_i), ..., \beta_{iN}(\delta_i)$; mean squared error $\text{MSE}(\delta_i)$
Compute V_i and F_i by (5) and (6)
$\rho_{\max,i} \leftarrow 2\|F_i^T V_i\|_\infty$
$\rho_{\min,i} \leftarrow 10^{-4} \rho_{\max,i}$
$\Theta_i \leftarrow 100$ logarithmically equidistant values from $\rho_{\min,i}$ to $\rho_{\max,i}$
for $\rho_i \in \Theta_i$ **do**
  estimate $\text{MSE}(\delta_i, \rho_i)$ by 3-fold cross validation on F_i, V_i and solving (7) on the respective training set
end for
$\rho_{\text{opt},i} \leftarrow \underset{\rho_i \in \Theta_i}{\text{argmin}}\, \text{MSE}(\delta_i, \rho_i)$
$(\beta_{i1}(\delta_i), ..., \beta_{iN}(\delta_i)) \leftarrow$ the solution to (7) on the whole data set F_i, V_i for $\rho_i = \rho_{\text{opt},i}$
$\text{MSE}(\delta_i) \leftarrow \text{MSE}(\delta_i, \rho_{\text{opt},i})$

Fig. 8. Network Inference algorithm [7]

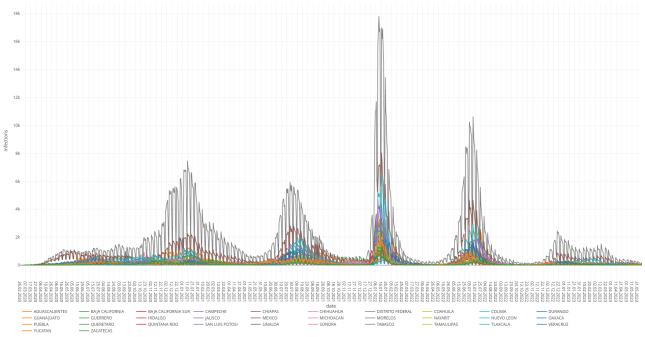

Fig. 9. Absolute infection counts

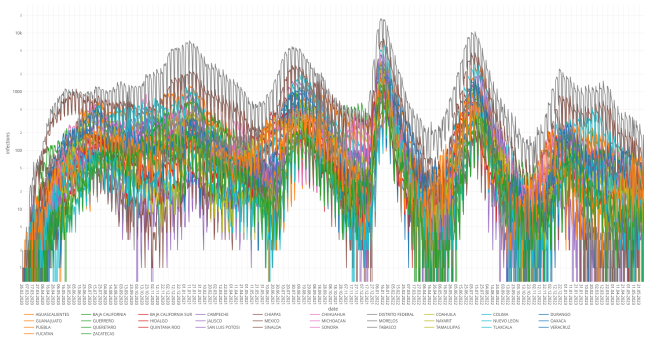

Fig. 10. Absolute infection counts per 100.000 people

[6] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. 17, 2009.
[7] B. Prasse, M. A. Achterberg, L. Ma, and P. V. Mieghem, "Network-inference-based prediction of the covid-19 epidemic outbreak in the chinese province hubei," *Applied Network Science*, vol. 5, pp. 1–11, 12 2020. [Online]. Available: https://appliednetsci.springeropen.com/articles/10.1007/s41109-020-00274-2
[8] M. A. Achterberg, B. Prasse, L. Ma, S. Trajanovski, M. Kitsak, and P. V. Mieghem, "Comparing the accuracy of several network-based covid-19 prediction algorithms," *International Journal of Forecasting*, vol. 38, pp. 489–504, 4 2022.
[9] W. . Kermack and A. G. Mckendrick, "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*, vol. 115, pp. 700–721, 8 1927. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118
[10] J. Satsuma, R. Willox, A. Ramani, B. Grammaticos, and A. S. Carstea, "Extending the sir epidemic model," *Physica A: Statistical Mechanics and its Applications*, vol. 336, pp. 369–375, 5 2004.
[11] V. Capasso and G. Serio, "A generalization of the kermack-mckendrick deterministic epidemic model," *Mathematical Biosciences*, vol. 42, pp. 43–61, 11 1978.
[12] S. Moein, N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, and Y. Gheisari, "Inefficiency of sir models in forecasting covid-19 epidemic: a case study of isfahan," *Scientific Reports 2021 11:1*, vol. 11, pp. 1–9, 2 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-84055-6
[13] I. Cooper, A. Mondal, and C. G. Antonopoulos, "A sir model assumption for the spread of covid-19 in different communities," *Chaos, Solitons & Fractals*, vol. 139, p. 110057, 10 2020.
[14] W. Just, S. Ahn, and D. Terman, "Neuronal networks: A discrete model," 2013. [Online]. Available: http://dx.doi.org/10.1016/B978-0-12-415780-4.00006-5
[15] M. Youssef and C. Scoglio, "An individual-based approach to sir epidemics in contact networks," *Journal of Theoretical Biology*, vol. 283, pp. 136–144, 8 2011.
[16] B. Prasse and P. V. Mieghem, "Network reconstruction and prediction of epidemic outbreaks for general group-based compartmental epidemic models," *IEEE Transactions on Network Science and Engineering*, vol. 7, pp. 2755–2764, 10 2020.
[17] T. Hastie, R. T. Martin, and W. Hastie, *Statistical Learning with Sparsity The Lasso and Generalizations Statistical Learning with Sparsity*. CRC press, 2015.
[18] C. CentroGeo, "Covid-19 tablero méxico - conacyt - centrogeo - geoint - datalab," 5 2023. [Online]. Available: https://datos.covid-19.conacyt.mx/#DownZCSV
[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[20] L. Quintana-Romero, M. Ángel Mendoza-González, and J. Álvarez García, "Covid-19 and tourism in mexico: Economic impacts and prospects," *Pandemics and Travel: COVID-19 Impacts in the Tourism Industry*, pp. 173–191, 1 2021.
[21] O. Cruz-Milan and S. Lagunas-Puls, "Effects of covid-19 on variations of taxpayers in tourism-reliant regions: The case of the mexican caribbean," *Journal of Risk and Financial Management 2021, Vol. 14, Page 578*, vol. 14, p. 578, 12 2021. [Online]. Available: https://www.mdpi.com/1911-8074/14/12/578/htmhttps://www.mdpi.com/1911-8074/14/12/578
[22] Y. Yang, B. Altschuler, Z. Liang, and X. R. Li, "Monitoring the global covid-19 impact on tourism: The covid19tourism index," *Annals of Tourism Research*, vol. 90, p. 103120, 9 2021. [Online]. Available: /pmc/articles/PMC8453610/https:

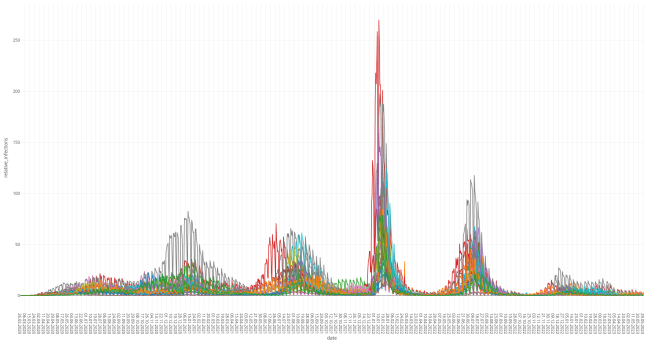

Fig. 11. Relative infection counts

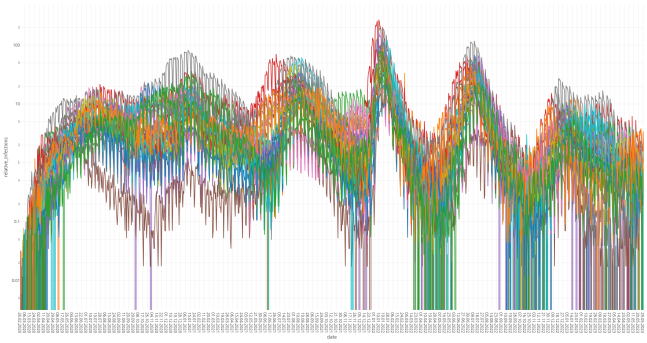

Fig. 12. Relative infection counts per 100.000 people

//www.ncbi.nlm.nih.gov/pmc/articles/PMC8453610/

A NIPA ALGORITHM PSEUDO-CODE

Both algorithms from the pseudocode figures 7 and 8 were provided by the procedure in the research from Prasse et al. [7]. The LASSO was done with the Python library `scikit-learn LASSO` [19]. Line 4 of the NIPA algorithm was exchanged by the `rolling mean` Pandas function.

B VISUALISATION OF THE INFECTION COUNTS DATA SET

These files are a visualisation of the data set in the form of absolute and relative numbers on both a linear and logarithmic scale.