Analyzing spatial mobility patterns with time-varying graphical lasso: Application to COVID-19 spread

Iván L. Degano | Pablo A. Lotito

1CEMIM, Facultad de Ciencias Exactas y Naturales, UNMdP, Mar del Plata, Argentina
2PLADEMA, Facultad de Ciencias Exactas, UNCPBA, Tandil, Argentina

Abstract
This work applies the time-varying graphical lasso (TVGL) method, an extension of the traditional graphical lasso approach, to address learning time-varying graphs from spatiotemporal measurements. Given georeferenced data, the TVGL method can estimate a time-varying network where an edge represents a partial correlation between two nodes. To achieve this, we use a COVID-19 data set from the Argentine province of Chaco. As an application, we use the estimated network to study the impact of COVID-19 confinement measures and evaluate whether the measures produced the expected result.

1 | INTRODUCTION

Spatial information often has an underlying network structure associated with it. Modeled as a graph, consisting of sets of nodes and edges, it is a fundamental tool to describe the relationship among georeferenced entities. Introducing a graph representation enables us to analyze georeferenced data on networks in many practical applications such as social, natural, and engineered systems. Examples include social networks, international relationships, airport networks, and transportation networks. This research field, together with its applications, is called spatial networks (see Barthélemy, 2011 and the references therein).

Many temporal networks are obtained as a result of the mobility of individuals. Examples include social contact networks of humans (Zhang et al., 2015), where contact is defined by physical proximity, animal networks due to animal trades (Chen et al., 2014), and transportation networks (Sun et al., 2013). Spatial constraints regulate...
mobility. For example, individuals tend to visit nearby rather than remote sites. Extending spatial and temporal networks to spatiotemporal networks is an exciting line of research.

Following this line of work, the metapopulation model constitutes a powerful framework that studies the impact of human mobility on population dynamics, such as epidemic processes (Colizza, Pastor-Satorras, & Vespignani, 2007; Colizza & Vespignani, 2008; Hufnagel, Brockmann, & Geisel, 2004). In brief, the model assumes a network of subpopulations, not of individuals. A subpopulation hosts individuals, and they can move from one subpopulation to another, in accordance with some rule. A metapopulation network can be seen as a time-varying network because the human behavior induces recurrent mobility patterns back and forth; for example, Balcan and Vespignani (2011) assume individuals commute to an adjacent community and come back to their home community. But it can also be treated as a temporal graph, as mobility patterns could change over time. Liu, Baronchelli, and Perra (2013), Nadini et al. (2018), and Ren and Wang (2014) studied this topic, modeling the epidemic dynamics over a time-varying community network structure.

Epidemic processes are probably the most studied dynamical processes on networks, both for static and temporal networks. In contrast to the wealth of theory available for static networks (Pastor-Satorras et al., 2015), studies of epidemic processes on temporal networks are more often conducted numerically, presumably because of the problem’s difficulty. A range of theoretical, numerical, and empirical results on temporal network epidemiology can be found in Masuda and Holme (2013).

In the past decade, numerous machine-learning algorithms have been introduced for analyzing spatial data on a priori known graphs. However, there are often settings where the graph is not readily available. Then the data structure has to be estimated for effective representation, processing, analysis, or visualization of the data. For example, there exist many algorithms to detect community structures in a network, that is, groups of nodes such that the nodes are densely connected within the same group and relatively sparsely connected across different groups (Greene, Doyle, & Cunningham, 2010; Palla, Barabási, & Vicsek, 2007; Peixoto & Rosvall, 2017, 2019). Another crucial task is to infer a graph topology that describes the characteristics of data observations, hence capturing the underlying relationship between these entities. The goal is to obtain all the edges connecting those nodes. In the case of temporal networks, the main idea is to assume that the multiple graphs share the same vertex set while being allowed to have distinct edge sets; for an overview, see Dong et al. (2019), Giannakis, Shen, and Karanikolas (2018), and Mateos et al. (2019).

To this end, many approaches have been proposed in the literature. Zhou, Lafferty, and Wasserman (2010) have proposed a method to track the network structure that can be switched between several states focused on a kernel method. Alternatively, some works have proposed different approaches to infer dynamic networks from observations within different time windows, with a penalty term imposed on the similarity between consecutive networks to be inferred. For example, Kalofolias et al. (2017) use a smoothness penalty term while Kolar et al. (2010) and Gibberd and Nelson (2014) implement an $\ell_1$ fused penalty. On a related note, the information cascade-based network inference (Baingana & Giannakis, 2017; Baingana, Mateos, & Giannakis, 2014; Gomez-Rodriguez, Leskovec, & Krause, 2012) assumes a viral spreading process over the nodes and aims to infer the links based on the node infection times.

Recently, Hallac et al. (2017) introduced a model named time-varying graphical lasso (TVGL) to infer dynamic networks from multivariate time series data, estimating a sequence of precision matrices from the observed data. They cast the problem in terms of estimating a sparse time-varying inverse covariance matrix, which reveals a dynamic network of interdependencies between the entities. Whereas previous works such as Gibberd and Nelson (2014), Kalofolias et al. (2017), and Kolar et al. (2010) allow a unique type of temporal variation between graphs, TVGL can model many different types of time evolution such as smoothly varying networks, rare but large-scale shifts, or a single node rewiring its connections. Since dynamic network inference is a computationally expensive task, they derive a scalable message-passing algorithm using the alternating direction method of multipliers (ADMM, Boyd et al., 2011), to solve this problem efficiently. This algorithm can incorporate various penalty types and take advantage of known properties and solution methods to derive closed-form updates.

This article presents the TVGL method as a tool for learning time-varying graphs from georeferenced real data; in particular, we use the TVGL algorithm to estimate the partial correlation involving spatiotemporal epidemiological data. At this level, a single node represents an entire city or district, and links represent dependencies between these
In this work, we use the TVGL algorithm and the reported cases of COVID-19 to estimate a spatiotemporal network that models similar outbreak patterns between different cities in the Argentine province of Chaco. Knowing the structure of this network can help design control strategies to prevent the spread of the virus or understand the epidemiological situation of a region. Most importantly, the resulting graph can show the behavioral influence between people in the different cities of the province, given some clues as to the mobility of the disease (and of the population), considering only the daily detected cases of COVID-19. The algorithm also allows us to detect sudden shifts in the network that we can link to events that happened in the region.

In the remainder of the article we first introduce the TVGL in Section 2. We then apply the TVGL framework in Section 3 for COVID-19 data sets from the Argentine province of Chaco to investigate the dynamic relationships between COVID-19 cases in each locality and the effect of lockdown politics. Some discussions and concluding remarks are given in Section 4.

2 | METHOD

Correlation networks are an approach to analyzing multivariate time series as networks. In a correlation network, two time series corresponding to two nodes are connected if they are strongly correlated according to a particular criterion. The graphical lasso method solves this problem. It infers the inverse covariance matrix $\Sigma^{-1}$ under the assumption that the observation of the multivariate time series at each discrete time $t$ follows a multivariate Gaussian distribution, which is independent of $t$ (Friedman, Hastie, & Tibshirani, 2007; Yuan & Lin, 2007). The inverse covariance matrix has the following interpretation: if $\Sigma_{ij}^{-1} = 0$, the corresponding nodes $i$ and $j$ are conditionally independent of each other, given the values on the other nodes. Moreover, the inverse covariance matrix elements are proportional to the partial correlation coefficients between a pair of nodes.

Within a single multivariate time series, the partial correlation itself may change over time. In this case, it is appropriate to consider temporal correlation networks. This setup has been explored by Hallac et al. (2017), who provided an extension of the graphical lasso to the case of time-varying correlational structure, called the time-varying graphical lasso. The method is able to provide relationships between entities in a time-varying context, even those that cannot be distinguished with the naked eye, and it is helpful to understand a large universe of data.

This section introduces the TVGL model and describes the selection of the parameters involved in the model.
2.1 | Time-varying graphical lasso

The TVGL was defined by Hallac et al. (2017) to infer dynamic networks. Let us assume that for each location, we have a sequence of observations at times \(0 \leq t_1 \leq \cdots \leq t_T\). At each time \(t_i\), there are \(n_i \geq 1\) different observation vectors for the \(p\) locations in study. We assume that the observations are sampled from a distribution \(x \sim N(0, \Sigma(t))\).

We aim to estimate the underlying covariance matrix \(\Sigma(t)\) which changes over time. To do so, Hallac et al. (2017) set up a sequence of graphical lasso models to infer sparse inverse covariance matrices \(\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_T)\) at times \(t_1, t_2, \ldots, t_T\), where \(\Theta_t = \Sigma(t)^{-1}\). This sequence gives a temporal network in the snapshot representation. For these graphical lasso problems, each one is coupled with others in a chain to penalize deviations in the estimations. The TVGL method approximates the covariance matrices \(\Sigma(t)\) by solving the convex optimization problem:

\[
\min_{\Theta_1, \ldots, \Theta_T} \sum_{t=1}^{T} -l_t(\Theta_t) + \lambda \|\Theta_t\|_{2,1} + \beta \sum_{t=2}^{T} \psi(\Theta_t - \Theta_{t-1})
\]

where \(l_t(\Theta_t) = n_t (\log \det \Theta_t - \text{Tr}(S_t \Theta_t))\)

with \(\Theta_t \in S_{++}^p\) (symmetric positive-definite), \(\lambda\) and \(\beta\) are non-negative model parameters and \(\|\Theta_t\|_{2,1}\) denotes the off-diagonal \(\ell_1\) seminorm, which enforces the elementwise sparsity for the solution of \(\Theta_t\) (the seminorm is the same as the norm except that the value may be equal to zero even when \(\Theta_t\) is not the zero matrix). In other words, \(\Theta_t\) will have many zeros when \(\lambda\) is set to an appropriately large value. \(S_t\) is the empirical covariance matrix, and \(n_t\) is the number of observations.

The elements of each inverse covariance matrix \(\Theta_t\) are proportional to the partial correlation coefficients between a pair of nodes (Lauritzen, 1996), that is, the method can measure the relationship between two variables given all the other variables in the data set. If we denote by \(R_{ij}\) the matrix of partial correlations at time \(i\), then the \(jk\)th element of \(R_{ij}\) is given by:

\[
R_{ijk} = \frac{-\Theta_{ijk}}{\sqrt{\Theta_{ij} \Theta_{kk}}} \tag{3}
\]

That is to say, the partial correlation is proportional to the negative of the inverse covariance matrix’s corresponding element. Then a zero value element of \(\Theta_t\) indicates conditional independence between the corresponding variables.

\(\psi(\Theta_t - \Theta_{t-1})\) is another convex penalty function, minimized at \(\psi(0)\). In this way, the penalty function \(\psi\) allows us to enforce different behaviors of the network over time. That is to say, if we have prior knowledge about how the connectivity of the network will change over time, we can achieve it through the penalty function \(\psi\). Because we would like to analyze relationships using georeferenced COVID-19 daily cases, we will assume at first that they change in a smooth way. Hence, the Laplacian penalty, defined by:

\[
\psi(X) = \sum_{ij} X_{ij}^2 \tag{4}
\]

is employed, where \(X_{ij}\) refers to the \(ij\)th element in the matrix \(\Theta_{k+1} - \Theta_k\). Another useful choice for \(\psi\) is:

\[
\psi(X) = \sum_{ij} |X_{ij}| \tag{5}
\]

which encourages the networks adjacent in time to be elementwise close to each other.
2.2 | The proposed algorithm

To solve the TVGL formulation defined by Equation (1), the ADMM (Boyd et al., 2011) is used. The ADMM is a well-established distributed convex optimization approach. By applying the ADMM, the original problem can be divided into a series of subproblems, and the globally optimal solution can be achieved. To do so, TVGL rewrites the terms in the form of proximal operators (Parikh & Boyd, 2014), which are defined for a matrix \( A \in \mathbb{R}^{m \times n} \) and a real-valued function \( f \) as:

\[
\text{prox}_{\eta f}(x) = \arg\min_{x \in \mathbb{R}^m} \left( f(x) + \frac{1}{2\eta} \|x - A\|_F^2 \right)
\]  

Writing the problems in this form, TVGL takes advantage of well-known properties in the area of convex analysis such as a closed-form update formula for each of the ADMM subproblems. Through introducing a consensus variable \( Z = \{Z_0, Z_1, Z_2\} \), where \( Z_0 = (Z_{1,0}, Z_{2,0}, ..., Z_{T,0}) \), \( Z_1 = (Z_{1,1}, Z_{2,1}, ..., Z_{T-1,1}) \), \( Z_2 = (Z_{2,2}, Z_{3,2}, ..., Z_{T,2}) \). Equation (1) can be represented as:

\[
\begin{align*}
\min & \quad \sum_{i=1}^{T} i_i (\Theta_i) + \lambda_i \|\Theta_i\|_{\text{vec}} + \beta \sum_{i=2}^{T} \Psi (Z_{i,2} - Z_{i-1,1}) \\
\text{subject to} & \quad Z_{i,0} = \Theta_i, \Theta_i \in S_{++}^p \text{ for } i = 1, 2, ..., T \\
& \quad (Z_{i-1,1}, Z_{i,2}) = (\Theta_{i-1}, \Theta_i) \text{ for } i = 1, 2, ..., T
\end{align*}
\]  

By applying the scaled dual variable \( U \), which has a similar format to the consensus variable \( Z \), the subproblems described in Equation (7) can be solved using Algorithm 1. By separating the problem described in Equation (1) into two blocks, \( \Theta \) and \( Z \), the algorithm converges to the global optimum solutions as guaranteed by this ADMM approach. More details of the algorithm can be found in Hallac et al. (2017).

**Algorithm 1:** TVGL algorithm using the ADMM.

1. Initialize \( \Theta^0, Z^0, U^0 \in \mathbb{R}^{p \times p} \);
2. Input regularization parameter \( \rho \in \mathbb{R} \);
3. for \( k = 0 \) to convergence do
   1. \( \Theta^{k+1} := \arg\min_{\Theta \in S_{++}^p} \mathcal{L}_\rho (\Theta, Z^k, U^k) \);
   2. \( Z^{k+1} = \begin{bmatrix} Z_{0}^{k+1} \\
                                Z_{1}^{k+1} \\
                                Z_{2}^{k+1} \end{bmatrix} := \arg\min_{Z_0, Z_1, Z_2} \mathcal{L}_\rho (\Theta^{k+1}, Z, U^k) \);
   3. \( U^{k+1} = \begin{bmatrix} U_{0}^{k+1} \\
                                U_{1}^{k+1} \\
                                U_{2}^{k+1} \end{bmatrix} := \begin{bmatrix} U_{0}^{k} \\
                                U_{1}^{k} \\
                                U_{2}^{k} \end{bmatrix} + \begin{bmatrix} \Theta^{k+1} - Z_{0}^{k+1} \\
                                                  (\Theta_{1}^{k+1}, ..., \Theta_{T-1}^{k+1}) - Z_{1}^{k+1} \\
                                                  (\Theta_{2}^{k+1}, ..., \Theta_{T}^{k+1}) - Z_{2}^{k+1} \end{bmatrix} \)
4. end
### 2.3 Parameter tuning

The parameter selection has an impact on the performance of the approach. The TVGL model defined in Equation (1) includes three key parameters: (1) the number of observation vectors within a segment $n_i$; (2) the sparsity level of the network $\lambda$; and (3) the correlated level between adjacent networks $\beta$. Note that the parameter $n_i$ controls the ability of the TVGL to estimate time-varying changes. Small values of $n_i$ can better estimate the dynamic information. $\lambda$ determines the sparsity level of the network. The smaller the $\lambda$ value is, the better the network matches with the empirical data. However, it will lead to dense networks which are difficult to interpret. $\beta$ affects how strongly the adjacent precision matrix estimations should be. As the value of $\beta$ increases, the estimation of adjacent networks will become smoother over time.

Since we want to analyze the time-varying mobility pattern using the COVID-19 daily cases data set, the first parameter selection step is to determine which terms affect the representation of the time-varying information. That is, how to select the $n_i$ in the TVGL model. To better describe the dynamics using the TVGL, the parameter $n_i$ should be as small as possible. However, by setting the $n_i$ to be 1, the desired relationships between locations cannot be obtained. This is because, on average, COVID-19 symptoms take 5–6 days from infection with the virus to the appearance of the first symptoms, according to the World Health Organization. Thus, in this study, the $n_i$ was set to 7 for all $i$.

The regularization terms $\lambda$ and $\beta$ are optimized using the Akaike information criterion (AIC) (Hastie, Tibshirani, & Friedman, 2009). Here, the AIC is defined by:

$$
AIC(x) = \sum_{i=1}^{T} \left( -\log|\hat{\Theta}_i(x)| + \text{Tr}(S_i\hat{\Theta}_i(x)) \right) + K 
$$

where $x$ is the parameter to be optimized (i.e., $\lambda$ and $\beta$), and $\hat{\Theta}_i(x)$ is the estimated precision matrix for the time-stamp $t_i$ obtained under a specific $x$. The component $S_i$ is the empirical matrix, and $K$ is the total number of non-zero entries in $\hat{\Theta}(x)$. We evaluate Equation (8) for the Laplacian penalty using 15 values of $\lambda$ and 15 values of $\beta$ for the COVID data set. Data used in this work were collected through the Argentine Integrated Health Information System (SISA). This contains nearly 5,100 georeferenced cases detected in the Argentine province of Chaco from March to August 2020. For this data set, we obtain the level map in Figure 1. Then, the minimum value is obtained when $\lambda$ is set to 2.85, while $\beta$ is set to 1.

![AIC](image.png)

**Figure 1** The Akaike information criterion (AIC) level map using 15 values of $\lambda$ and 15 values of $\beta$ for the COVID data set. The minimum value of AIC is found at the lower level of the graph.
3 | RESULTS

In this section we estimate a spatiotemporal network from the daily detected cases of COVID-19 in the Argentine province of Chaco, located in the north of the country, using the TVGL model defined in Equation (1). Argentina is one of the countries most severely affected by the pandemic, and Chaco is one of the provinces with the most significant difficulties in limiting the spread of the disease (Rodriguez, 2020). To contain and mitigate the effects of the COVID-19 epidemic, Chaco implemented early measures to restrict individual mobility and promote social distancing. The government adopted an increasing number of measures, including school and university closures, limits placed on large social gatherings, closure of bars and restaurants, and national stay-at-home orders. On March 18, 2020, all non-essential businesses and services were closed, effectively putting the province under lockdown (Página 12, 2020). These measures were gradually relaxed during the following months. The resulting time-varying province graph shows how the spread of the virus from one location is related to all the others, indicating that there could be people in transit between both cities and how it varies in time.

By analyzing the dynamic connectivity through precision matrices extracted by the TVGL approach, we can visualize the connections between the cities' epidemiological situation. A correlation between two cities suggests that could be people in transit between both localities, which means we have available a sketch of the mobility patterns between the cities in the province. It is worth mentioning that other factors could imply a correlation in epidemiological data. This approach will at least allow us to affirm that when there is no correlation, there is also no mobility. Besides, we use time-varying inference to detect significant events, such as an increase in new cases detected, lockdown impositions, or reopening of activities, that could affect the entire network's topology.

Specifically, we apply the TVGL method to the daily COVID-19 detected data to illustrate meaningful insights, analyzing two networks: on the one hand, an entire province network; and the other hand, a regional network. By examining the historical COVID-19 data, we carry out three experiments to analyze the epidemiological relationships between different locations. In the first place, we capture the states of the correlation in a time-varying graph for the province case, assuming the correlation between the data changes only slightly over time. Our inferred graphical representation then shows how the epidemiological situation of a city affects other cities and how it varies with time. Second, we look for events that occasionally occur, which can cause a sudden shift in the province network structure. To this end, we study a temporal dynamic where single nodes may occasionally reweigh all their connections at once. Finally, we use the time-varying inference to detect significant events that affect the entire regional network, not only the individual neighborhoods within it.

We then describe the procedure we used to build the time-varying graph.

1. We obtain the coordinates from the data. Then the data are preprocessed to classify them by cities and dates. Finally, we count the accumulated daily cases per city in the provincial case or by census tract for the Gran Resistencia region. We obtain the number of weeks between March 7 and August 31, namely, 25. The number of cities is 53 for the provincial experiment, and 414 census tract nodes in the Gran Resistencia study. We consider positive COVID-19 cases each day.

2. We run the TVGL algorithm on the COVID-19 data set. We must choose which penalty function we will implement according to the goal of the experiment. For instance, if we want the network's topology to remain unchanged over a short time horizon, we use a Laplacian penalty. With this procedure, we obtain 25 precision matrices \( \Theta_t \). As mentioned before, the \( ij \)th coordinate of the matrix \( \Theta_t \) indicates that must be an edge between node \( i \) and node \( j \).

3. With the precision matrices, we build the desired networks. We compare adjacent networks to reveal any change in the network configuration that could mean that a significant event happened at this time.
3.1 Results for Chaco province COVID-19 data set

Dynamic networks were estimated using the TVGL approach. We analyze time-varying mobility using the TVGL model with the number of observations $n_i = 7$ for all $i$, resulting in 25 precision matrices. To better compare the relationships within the province, we divided the study into two separate networks. One network contains the provincial case, where we consider the capital city, Resistencia, and its nearest neighboring cities, Barranqueras, Fontana, and Puerto Vilelas, as a unique node. On the other hand, we study the Gran Resistencia region network, which includes the cities of Resistencia, Barranqueras, Fontana, and Puerto Vilelas. Figure 2 shows both situations in terms of the nodes considered in this work.

By applying the TVGL approach, we capture six mobility states for each network. Figure 3 shows the states estimated by the TVGL approach for Chaco province. The inferred characteristics of the learned graph are as follows:

![Chaco Province](image1)
![Gran Resistencia region](image2)

**Figure 2** Left: Chaco province with 53 nodes, each representing a city. Right: Gran Resistencia region with 414 nodes defined by the COVID-19 data set.

![Week 5: 11 April to 18 April](image3)
![Week 10: 16 May to 23 May](image4)
![Week 15: 20 June to 27 June](image5)

![Week 20: 25 July to 01 August](image6)
![Week 23: 15 August to 22 August](image7)
![Week 24: 22 August to 29 August](image8)

**Figure 3** Visualization of the graph learned from the Chaco COVID-19 data. The transparency level of the edges indicates the magnitude of the partial correlation between nodes according to Equation (3).
• In the first weeks (weeks 5 and 10 in Figure 3), nodes were not connected even if the distances were short. This is reasonable because the lockdown was more rigid at that time and mobility between cities was not allowed.

• In the weeks that followed (weeks 15 and 20 in Figure 3), some nodes in the province’s interior were connected to the Gran Resistencia region. Significantly, the province nodes’ interior may have connected to the Gran Resistencia region despite being far away. It is seen that the increase in infected cases in the city of Resistencia is related to the growth of cases in other towns in the interior, indicating the movement of people between Resistencia and towns such as Presidencia Roque Saenz Peña and General José de San Martín, to name a few.

• In the last weeks of the analysis (weeks 23 and 25 in Figure 3), the learned graphs show locally connected structures, that is, there is a relationship between the growth of infected cases of neighboring towns. That suggests people in transit between closest towns, for example, between Presidencia Roque Saenz Peña, Tres Isletas, Avía Terai, and Juan José Castelli.

Generally, mobility patterns maintain the same behavior every non-holiday week, so we expect the correlation network between cities to change only slightly over time. However, a lockdown situation can cause a sudden shift in the network structure. We look at the dependencies of these localities while using a perturbed node penalty, defined by:

$$\psi(X_i) = \min_{V: V + V^T = X_i} \sum_j \| [V]_j \|_2$$

where $X_i = \Theta_{i+1} - \Theta_i$ and $[V]_j$ is the $j$th column of a square matrix $V$ of dimension $p$. Taking the Euclidean norm of the columns of $V$ forces entire columns to be selected, while imposing $V + V^T = X_i$ ensures $X_i$ is symmetric. Hence, the penalty in Equation (9) would produce a temporal dynamic where single nodes may occasionally reweight all their connections at once. It typically reflects that something happened to affect just one city while leaving the remainder of the network unchanged. We solve the TVGL optimization problem with a perturbed node penalty and discover that four events stood out as having the most significant single-node effect on the Chaco localities dynamics.

After running the TVGL method, we plot the temporal deviation, $\| \Theta_i - \Theta_{i-1} \|_F$, in Figure 4. We discover that there are four large spikes in the temporal deviation score. The first one occurs during the week starting June 13, the second one the following week, the third one during the week starting July 11, and the last one during the week starting July 25. These represent significant ‘shifts’ in the network at those specific times.

**Figure 4** The plot of our estimated province network’s temporal deviation detects four significant shifts in June and July
We can show the network changes in the week starting June 13 in Figure 5, where we see that the Gran Resistencia region adds two new links: Presidencia Roque Saenz Peña and Tres Isletas. We examined the media to understand what may have caused this shift, and we found during that week, the government announced the closure of almost all activities in the Gran Resistencia area, except those considered essential such as the gas station, food stores, and pharmacies, due to the high number of new cases detected at these three nodes (Télam, 2020a). Similarly, the mayors of Presidencia Roque Saenz Peña and Tres Isletas also announced similar measures after detecting new COVID-19 cases during this week (Diario Norte, 2020c; Datachaco.com, 2020).

In the week starting June 20, we can see in Figure 5 that the native community Colonia Aborigen was the new node added to the Gran Resistencia region, inducing a significant shift. We found that during that week, an important number of cases originated with a health professional who did not obey the physical distancing measures (Chaina, 2020).

The third significant shift occurred during the week starting July 11. Unlike the two previous shifts, in this case, the connection between Gran Resistencia and El Sauzalito vanished, as we see in Figure 6. We examined the media, and we found that this week there was a decrease in the daily COVID-19 cases detected in this city (Chaco Día por Día, 2020b). Finally, it is interesting to note that the fourth shift occurred during the week starting July 25, coinciding with the week where some commercial activities reopened, and the government allowed population mobility (Télam, 2020b). Furthermore, this week there was a record of daily COVID-19 provincial cases detected (Diario Norte, 2020b), many of these in the province’s interior, in towns such as Saenz Peña, La Verde, Castelli,
and El Sauzalito, among others (Chaco Día por Día, 2020a). We can see the change in the network configuration in Figure 7.

3.2 | Results for Gran Resistencia region data set

We now use time-varying inference to detect significant events that affected the entire network, not just single entities within it. We run an experiment similar to the previous example, except that we now utilize the georeferenced data set for the Gran Resistencia area, which includes the capital city Resistencia, and its nearest neighboring cities, Barranqueras, Fontana, and Puerto Vilelas. In this case, we preprocess the data to classify it by census tract and dates.

Since we are focusing on event detection, we look for the maximum temporal deviation with an $\ell_1$ penalty defined by:

$$\psi(\Theta_{i+1} - \Theta_i) = \sum_{ij} |\Theta_{i+1,j} - \Theta_{ij}|$$

(10)
This point in time represents the most significant “shock” to the network, the time-stamp where the TVGL model detected a significant and sudden change.

We plot the temporal deviation in Figure 8. We discover that a shift occurred between May 2 and 16 and also between June 13 and 20. Again, we examined the media to understand what may have caused this shift, and we saw that for the first case, many COVID-19 cases were detected in the Toba neighborhood, an area of the Gran Resistencia marked by social and economic vulnerability (Diario Norte, 2020a). For the other detected shift, we also found that a lockdown was imposed on June 14, 2020.

Our results imply that, even though detected cases of COVID-19 continue to grow, there were tangible long-term effects on the correlation network between neighborhoods in the Gran Resistencia area.

4 | CONCLUSIONS

Of particular note, the methods to infer a time-varying graph have been widely applied to estimate dynamic networks in many areas, such as information flows, gene regulation, brain functional connectivity, and financial interactions (Dong et al., 2019; Giannakis, Shen, & Karanikolas, 2018; Mateos et al., 2019). However, to the best of our knowledge, there is no actual application of such a class of models to estimate a dynamic network of georeferenced data. Thus, it may be of interest to extend some spatial networks (Barthélemy, 2011) to spatiotemporal networks (Aggarwal & Subbian, 2014; Holme, 2015; Kempe, Kleinberg, & Kumar, 2002). Although methods such as those of Gibberd and Nelson (2014), Kalofolias et al. (2017) and Kolar et al. (2010) offer a solution to solve this problem, they allow a unique type of temporal variation to describe the connectivity dynamics, and they have computationally expensive algorithms. In contrast, the TVGL defined by Hallac et al. (2017) can assess the problem with five different temporal variations and a simpler algorithm. Hence, we applied the TVGL in this work to the daily COVID-19 georeferenced detected cases. We performed several experiments using the TVGL model for COVID-19 data from the Argentine province of Chaco. We then optimized the key parameters and applied them to analyze the behavioral influences between people in different regions that lead to epidemic spread. In particular, the network transition behaviors were assessed, revealing the impact of the different public measures. Our experiments illustrated that the TVGL model captures the time-varying connectivity patterns between the data. We found that the TVGL can accurately represent the dynamic changes in the general politics of population mobility. For the dynamic networks estimated by the TVGL, we found that the correlated nodes developed in almost all consecutive states for both cases. This finding suggests that the spread of the disease, and probably the population mobility, developed grew up gradually in line with the relaxation of the measures taken by the government.

In particular, the model was used to investigate the correlation between outbreak patterns of different cities. We could observe the dynamics of the learned networks. More specifically, connectivity between cities in the province network and connectivity between the neighborhoods in the Gran Resistencia network helped to identify large-impact events. This fact may indicate that the TVGL model allows us to quantify the impact of public measures, such as lockdowns. In addition, note that many studies have examined population mobility using mobile phone data (Askitas, Tatsiramos, & Verheyden, 2020; Badr et al., 2020; Benzell, Collis, & Nicolaides, 2020; Chang et al., 2021; Duque et al., 2020; Jia et al., 2020; Lai et al., 2020; Schlosser et al., 2020) to understand epidemic spatial dynamics and quantify the impact of social distancing measures. We think the learned graphs can assist in the inference of the underlying network structure when mobility data are not available. We also believe that these graphs can be used as the underlying network of a metapopulation model, along the lines of Block et al. (2020), Calvetti et al. (2020), Chang et al. (2021), Duque et al. (2020) and Li et al. (2020).

One potential limitation in our framework is that we cannot ensure that the inferred graph precisely represents the population mobility. As we mentioned, the model only allows us to affirm that when there is no correlation between nodes, there is also no mobility. These limitations offer some exciting lines of work beyond those discussed in this article. For instance, one could combine the inferred graphs with cell phone data to validate...
the population mobility. Another interesting question to work on is how we can incorporate the learned graph in a metapopulation model to study the spread of the disease. Finally, it is of interest to enlarge the time horizon to incorporate data from the COVID-19 second wave.

In conclusion, we applied an approach named the time-varying graphical lasso (Hallac et al., 2017) to investigate mobility patterns using COVID-19 daily case data. The results demonstrated that the TVGL model reveals the partial correlation between the localities’ detected cases. Furthermore, the general policy on population mobility was tracked using the TVGL model with a perturbed node penalty and $\ell_1$ penalty. We showed that TVGL gives a compact representation of the connections between the georeferenced detected cases during an epidemic, suggesting that could be people in transit in a specific region. There were some interesting findings based on the results assessed by TVGL. In particular, by analyzing the obtained networks for Chaco, we observed no connections between localities in the first weeks of analysis due to the imposed lockdown. Then the connections between localities increased: between Resistencia and interior towns at first, and then among interior towns. This result demonstrates that by imposing a lockdown, the partial correlations between cases detected over a small region such as a province are null, but gradually the connections grow, indicating that there could be people in transit between the localities. In summary, the proposed TVGL model is a powerful tool for learning a georeferenced time-varying graph using epidemic data and could help in inferring the underlying network structure when mobile phone data are not available to infer the population mobility.

ORCID

Iván L. Degano

https://orcid.org/0000-0002-6557-7500

REFERENCES

Aggarwal, C., & Subbian, K. (2014). Evolutionary network analysis. ACM Computing Surveys, 47(1), 1–36.
Aleta, A., Martín-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., ..., Moreno, Y. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nature Human Behaviour, 4, 964–971. https://doi.org/10.1038/s41562-020-0931-9
Askutas, N., Tatsiramos, K., & Verheyden, B. (2020). Lockdown strategies, mobility patterns and COVID-19. Covid Economics: Vetted and Real-Time Papers, 23, 263–302.
Badr, H., Hongru, D., Marshall, M., Squire, M., & Garden, L. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. The Lancet Infectious Diseases, 20(11), 1247–1254. https://doi.org/10.1016/S1473-3099(20)30553-3
Baingana, B., & Giannakis, G. B. (2017). Tracking switched dynamic network topologies from information cascades. IEEE Transactions on Signal Processing, 65(4), 985–997. https://doi.org/10.1109/TSP.2016.2628354
Baingana, B., Mateos, G., & Giannakis, G. B. (2014). Proximal-gradient algorithms for tracking cascades over social networks. IEEE Journal of Selected Topics in Signal Processing, 8(4), 563–575. https://doi.org/10.1109/JSTSP.2014.2317284
Balcan, D., & Vespignani, A. (2011). Phase transitions in contagion processes mediated by recurrent mobility patterns. Nature Physics, 7(7), 581–586. https://doi.org/10.1038/nphys1944
Barabási, A. L. (2016). Network science. Cambridge, UK: Cambridge University Press.
Barthélémy, M. (2011). Spatial networks. Physics Reports, 499(1–3), 1–101. https://doi.org/10.1016/j.physrep.2010.11.002
Benzell, S. G., Collis, A., & Nicolaides, C. (2020). Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. Proceedings of the National Academy of Sciences of the United States of America, 117(26), 14642–14644. https://doi.org/10.1073/pnas.2008025117
Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nature Human Behavior, 4, 588–596. https://doi.org/10.1038/s41562-020-0898-6
Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3(1), 1–122.
Buckee, C. O., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., & Schroeder, A. (2020). Aggregated mobility data could help fight COVID-19. Science, 368, 145–146.
Calvetti, D., Hoover, A. P., Rose, J., & Somersalo, E. (2020). Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19. Frontiers in Physics, 8, 261. https://doi.org/10.3389/fphy.2020.00261
Chaco Día por Día. (2020a). Aseguran que el coronavirus se está expandiendo hacia el interior del Chaco. Retrieved from https://www.chacodiapordia.com/2020/07/31/aseguran-que-el-coronavirus-se-esta-expandiendo-hacia-el-interior-del-chaco/.

Chaco Día por Día. (2020b). Covid-19: dieron de alta a 15 pacientes positivos en El Sauzalito. Retrieved from https://www.chacodiapordia.com/2020/07/08/covid-19-dieron-de-alta-a-15-pacientes-positivos-en-el-sauzalito/

Chaina, P. (2020). Coronavirus: Alarma en una comunidad indígena. Retrieved from https://www.pagina12.com.ar/275438-coronavirus-alarma-en-una-comunidad-indigena

Chang, S., Piersson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. Nature, 589, 82–87.

Chen, S., White, B. J., Sanderson, M. W., Amrine, D. E., Ilany, A., & Lanzas, C. (2014). Highly dynamic animal contact network and implications on disease transmission. Scientific Reports, 4, 4472.

Colizza, V., Pastor-Satorras, R., & Vespignani, A. (2007). Reaction-diffusion processes and metapopulation models in heterogeneous networks. Nature Physics, 3(4), 276–282.

Colizza, V., & Vespignani, A. (2008). Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. Journal of Theoretical Biology, 251(3), 450–467. doi:10.1016/j.jtbi.2007.11.028

Datachaco.com. (2020). Desde el municipio confirman la circulación viral en Tres Isletas. Author. Diario Norte. (2020a). El coronavirus golpea en los barrios: el Toba entre los más complicados. Retrieved from https://www.diaronorte.com/191699-el-coronavirus-golpea-en-los-barrios-el-toba-entrelosmas-complicados

Diario Norte. (2020b). Récord de contagios: 93 nuevos casos de covid-19 en el Chaco. Retrieved from https://www.diaronorte.com/194652-record-de-contagios-93-nuevos-cases-de-covid-19-en-el-chaco

Diario Norte. (2020c). Sáenz Peña vuelve a la fase 1 del aislamiento. Retrieved from https://www.diaronorte.com/193236-saenz-pena-vuelve-a-la-fase-1-delaislamiento

Dong, X., Thanou, D., Rabbat, M., & Frossard, P. (2019). Learning graphs from data: A signal representation perspective. IEEE Signal Processing Magazine, 36(3), 44–63. https://doi.org/10.1109/MSP.2018.2887284

Ducharme, J. (2020). World Health Organization declares COVID-19 a pandemic: Here’s what that means. Time, March 11. Retrieved from https://time.com/5791661/who-coronavirus-pandemicdeclaration/

Duque, D., Morton, D. P., Singh, B., Du, Z., Pasco, R., & Meyers, L. A. (2020). Timing social distancing to avert unmanageable COVID-19 hospital surges. Proceedings of the National Academy of Sciences of the United States of America, 117(33), 19873–19878. https://doi.org/10.1073/pnas.2009033117

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature, 584, 257–261. https://doi.org/10.1038/s41586-020-2405-7

Friedman, J., Hastie, T., & Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Gianakis, G. B., Shen, Y., & Karanikolas, G. V. (2018). Topology identification and learning over graphs: Accounting for nonlinearities and dynamics. Proceedings of the IEEE, 106(5), 787–807. https://doi.org/10.1109/JPROC.2018.2804318

Gibberd, A. J., & Nelson, J. D. B. (2014). High dimensional changepoint detection with a dynamic graphical lasso. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy (pp. 2488–2684). Piscataway, NJ: IEEE.

Gomez-Rodriguez, M., Leskovec, J., & Krause, A. (2012). Inferring networks of diffusion and influence. ACM Transactions on Knowledge Discovery from Data, 5(4), 1–37. https://doi.org/10.1145/2067637.2086741.

Greene, D., Doyle, D., & Cunningham, P. (2010). Tracking the evolution of communities in dynamic social networks. In Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining, Odense, Denmark (pp. 176–183). New York, NY: ACM. https://doi.org/10.1109/ASONAM.2010.17

Hallac, D., Park, Y., Boyd, S., & Leskovec, J. (2017). Network inference via the time-varying graphical lasso. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada (pp. 205–213). New York, NY: ACM.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning, data mining, inference, and prediction. New York, NY: Springer.

Hufnagel, L., Brockmann, D., & Geisel, T. (2004). Forecast and control of epidemics in a globalized world. Proceedings of the National Academy of Sciences of the United States of America, 101(42), 15124–15129. https://doi.org/10.1073/pnas.0308344101

Holme, P. (2015). Modern temporal network theory: A colloquium. European Physical Journal B, 88(9), 234.

Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. Nature, 582, 389–394. https://doi.org/10.1038/s41586-020-2284-y

Kalofolias, V., Loukas, A., Thanou, D., & Frossard, P. (2017). Learning time varying graphs. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA (pp. 2826–2831). Piscataway, NJ: IEEE.
Kempe, D., Kleinberg, J., & Kumar, A. (2002). Connectivity and inference problems for temporal networks. *Journal of Computer and System Sciences, 64*(4), 820–842. https://doi.org/10.1006/jcss.2002.1829

Kolar, M., Song, L., Ahmed, A., & Xing, E. P. (2010). Estimating time-varying networks. *Annals of Applied Statistics, 4*(1), 94–123. https://doi.org/10.1214/09-AOAS308

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., & Tatem, A. J. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature, 585*, 410–413. https://doi.org/10.1038/s41586-020-2293-x

Lauritzen, S. L. (1996). Graphical models. Oxford, UK: Clarendon Press.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science, 368*, 489–493. https://doi.org/10.1126/science.abb3221

Liu, S.-Y., Baronchelli, A., & Perra, N. (2013). Contagion dynamics in time-varying metapopulation networks. *Physical Review E, 87*(3), 032805. https://doi.org/10.1103/PhysRevE.87.032805

Masuda, N., & Holme, P. (2013). Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Reports, 5*, 6.

Mateos, G., Segarra, S., Marques, A. G., & Ribeiro, A. (2019). Connecting the dots: Identifying network structure via graph signal processing. *IEEE Signal Processing Magazine, 36*(3), 16–43.

Nadini, M., Sun, K., Ubaldi, E., Starnini, M., Rizzo, A., & Perra, N. (2018). Epidemic spreading in modular time-varying networks. *Scientific Reports, 8*(1), 2352. https://doi.org/10.1038/s41598-018-20908-x

Página 12. (2020). Coronavirus: Medidas drásticas en el Chaco. Author. Retrieved from https://www.pagina12.com.ar/253626-coronavirus-medidas-drasticas-en-el-chaco

Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*(7136), 664–667.

Parikh, N., & Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in Optimization, 1*(3), 127–239. https://doi.org/10.1561/2400000003

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics, 87*(3), 925–979. https://doi.org/10.1103/RevModPhys.87.925

Peixoto, T. P., & Rosvall, M. (2017). Modelling sequences and temporal networks with dynamic community structures. *Nature Communications, 8*(1), 582. https://doi.org/10.1038/s41467-017-00148-9

Peixoto, T. P., & Rosvall, M. (2019). Modelling temporal networks with Markov chains, community structures and change points. In P. Holme, & J. Saramäki (Eds.), *Temporal network theory* (pp. 65–81). Cham, Switzerland: Springer.

Ren, G., & Wang, X. (2014). Epidemic spreading in time-varying community networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 24*(2), 023116.

Rodríguez, J. (2020). Coronavirus en Chaco: Una provincia paralizada, con su gente en shock y frente al temor de una ola de contagios. Retrieved from https://www.clarin.com/sociedad/provincia-chac-situacion-critica-aislada-paralizada-te- menibre-circulacion-virus\_01_tmnEEiT.html3c/ExternalRef%3e

Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., & Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences of the United States of America, 117*(52), 32883–32890. https://doi.org/10.1073/pnas.2012326117

Sun, L., Axhausen, K. W., Lee, D.-H., & Huang, X. (2013). Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences of the United States of America, 110*(34), 13774–13779. https://doi.org/10.1073/pnas.1306440110

Télam. (2020a). Chaco anunció medidas mucho más restrictivas para mitigar los efectos del coronavirus. Retrieved from https://www.telam.com.ar/notas/202006/476532-capitanich-anunciamedidas-mucho-mas-restrictivas-para-mitigar-los-efectos-del-coronavirus-en-chaco.html3c/ExternalRef%3e

Télam. (2020b). Chaco inicia la primera fase de reapertura progresiva de actividades. Retrieved from https://www.telam.com.ar/notas/202007/494118-chaco-coronavirus.html3c/ExternalRef%3e

Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika, 94*(1), 19–35. https://doi.org/10.1093/biomet/asm018

Zhang, Y., Li, X., Xu, J., & Vasilakos, A. (2015). Human interactive patterns in temporal networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 45*(2), 214–222. https://doi.org/10.1109/TSMC.2014.2360505

Zhou, S., Lafferty, J., & Wasserman, L. (2010). Time varying undirected graphs. *Machine Learning, 80*(2–3), 295–319. https://doi.org/10.1007/s10994-010-5180-0

---

**How to cite this article:** Degano, I. L., & Lotito, P. A. (2021). Analyzing spatial mobility patterns with time-varying graphical lasso: Application to COVID-19 spread. *Transactions in GIS, 25*, 2660–2674. [https://doi.org/10.1111/tgis.12799](https://doi.org/10.1111/tgis.12799)