

RESEARCH

Open Access



Network-based time series modeling for COVID-19 incidence in the Republic of Ireland

Stephanie Armbruster¹ and Gesine Reinert^{2*}

*Correspondence:
reinert@stats.ox.ac.uk

¹ Department of Biostatistics,
Harvard University, 655
Huntington Avenue, Boston, MA
02115, USA

² Department of Statistics,
University of Oxford, 24-29 St
Giles, Oxford OX1 3LB, UK

Abstract

Network-based time series models have experienced a surge in popularity over the past years due to their ability to model temporal and spatial dependencies, arising from the spread of infectious disease. The generalised network autoregressive (GNAR) model conceptualises time series on the vertices of a network; it has an autoregressive component for temporal dependence and a spatial autoregressive component for dependence between neighbouring vertices in the network. Consequently, the choice of underlying network is essential. This paper assesses the performance of GNAR models on different networks in predicting COVID-19 cases for the 26 counties in the Republic of Ireland, over two distinct pandemic phases (restricted and unrestricted), characterised by inter-county movement restrictions. Ten static networks are constructed, in which vertices represent counties, and edges are built upon neighbourhood relations, such as railway lines. We find that a GNAR model based on the fairly sparse Economic hub network explains the data best for the restricted pandemic phase while the fairly dense 21-nearest neighbour network performs best for the unrestricted phase. Across phases, GNAR models have higher predictive accuracy than standard ARIMA models which ignore the network structure. For county-specific predictions, in pandemic phases with more lenient or no COVID-19 regulation, the network effect is not quite as pronounced. The results indicate some robustness to the precise network architecture as long as the densities of the networks are similar. An analysis of the residuals justifies the model assumptions for the restricted phase but raises questions regarding their validity for the unrestricted phase. While generally performing better than ARIMA models which ignore network effects, there is scope for further development of the GNAR model to better model complex infectious diseases, including COVID-19.

Keywords: Network-based time series, COVID-19, Spatial models, Networks

Mathematics Subject Classification: 62M10, 05C82, 91D30

Introduction

In recent years, statistical models which incorporate networks and thereby acknowledge spatial dependencies when predicting temporal data have experienced a surge in popularity (e.g., Knight et al. 2019, 2016; Urrutia et al. 2022). Knight et al. (2016) developed

a generalised network autoregressive (GNAR) time series model which incorporates a secondary dependence in addition to standard temporal dependence. The secondary dependence is captured in a network. In Knight et al. (2016), the proposed network-based time series model is leveraged to predict mumps incidence across English counties during the British mumps outbreak in 2005. As graph to be associated with the Mumps network time series, Knight et al. (2016) chose a “county town” for each county and connected all towns which were less than a radius of a fixed number of kilometers away from each other.

Similar to mumps, COVID-19 is a highly infectious disease spread by direct contact between people (Nouvellet 2021). Human movement networks have been extensively relied upon to explain COVID-19 patterns (e.g. Jia 2020; Kraemer 2020; Li et al. 2021; Mo 2021; Nouvellet 2021; Sun et al. 2021; Wu et al. 2020). Therefore, it is a natural conjecture that such movement networks may help predict the spread of COVID-19. To investigate, this paper

- Fits GNAR models to predict the weekly COVID-19 incidence for all 26 counties in the Republic of Ireland, exploring different network constructions;
- Assesses the prevalence of a network effect in COVID-19 incidence in Ireland and the suitability of GNAR models to predict epidemic outbreaks as complex as COVID-19;
- Investigates the influence of changes in inter-county mobility, due to COVID-19 restrictions, on the performance of GNAR models as well as on the model parameters and hyperparameters.

The GNAR model is chosen because multivariate time series are often modelled by vector autoregressive (VAR) models. General VAR models are very flexible but require a large number of parameters to be estimated. The GNAR model which we employ here is a special case of a VAR model. It reduces the number of parameters to be estimated by restricting attention to edges in a network; in the case of a complete graph, the VAR model and the GNAR model coincide. Our overview of network-based time series models, given in Supplementary Material B.2, concludes that many network-based time series models can be conceptualized as a special case of the GNAR model, or are more restrictive with respect to the temporal-spatial dependencies they can model. Moreover, as a VAR-type model, the GNAR model inherits the well-understood VAR model framework, including parameter estimation via least squares, and model selection based on the BIC; for a survey see for example Lütkepohl (2005). These methods yield confidence sets for parameter estimation, which can inform analysis as well as policy development in a quantitative fashion. In contrast, deep learning approaches such as developed in Park et al. (2024) for predicting rental and return patterns at bicycle stations do not come with such theoretical guarantees.

In addition to a distance-based network as chosen in Knight et al. (2016), in this paper we explore a collection of network models which are motivated by potential movements of individuals. The movement networks are constructed according to general approaches based on statistical definitions of neighbourhoods as well as approaches specific to the infectious spread of the COVID-19 virus. In abuse of notation, we call these

networks *COVID-19 networks*, although they are only meant to reflect possible transmission routes of the disease. For each network, we select the best performing hyperparameter values, to predict COVID-19 incidence by a GNAR model, using the Bayesian Information Criterion (BIC). By splitting the available Irish data into two phases of the pandemic, restricted and unrestricted, we are able to investigate the potential change in the temporal and spatial dependencies in COVID-19 incidence between the two phases.

Overall, our findings are that while there is a clear network effect, the performance of the optimal GNAR model varies little across different network architectures of similar network density. GNAR models indicate higher predictive accuracy than ARIMA models on a country level, since they account for inter-county dependencies. On an individual county level, the variability of predictive performance is high, resulting in similar performance of ARIMA and GNAR models for some counties, while for others the GNAR model consistently outperforms the ARIMA model. The GNAR model seem better suited to predictive COVID-19 incidence in restricted pandemic phases than in unrestricted pandemic phases; the latter may be related to some of the model assumptions possibly requiring an adaptation as well as an increase in noise during unrestricted pandemic phases and high fluctuation of COVID-19 case numbers. Moreover, the classical VAR model, which is the GNAR model on the complete graph, does not perform as well as the GNAR model with an underlying network that has fewer edges, illustrating the value of using a GNAR model.

This paper is organised as follows. Section 2 introduces the data set. The methodology for network construction and for network-based time series modeling is described in Sect. 3. Section 4 provides an exploratory data analysis, while the model fit is shown in Sect. 5.1. The conclusions for the different pandemic phases are found in Sect. 5.2. The results are discussed in Section 6. The Supplementary Material includes a literature review on alternative models for incorporating temporal and spatial dependencies, visualisations of the COVID-19 networks, as well as details on the performance of GNAR models for predicting COVID-19 incidence.

The data and code are provided at https://github.com/stephanieArmbru/Case_study_GNAR_COVID_Ireland.git.

The Irish COVID-19 data set

By March 2023, the Republic of Ireland (abbreviated in this paper as *Ireland*) had recorded a total of 1.7 million confirmed COVID-19 cases and 8,719 deaths since the beginning of the pandemic (Health Protection Surveillance Centre 2022a). The Health Protection Surveillance Centre identified four main variants of concern for the COVID-19 virus in Ireland (Health Protection Surveillance Centre 2022b), in addition to the original variant: Alpha from 27.12.2020, Delta from 06.06.2021, Omicron I from 13.12.2021, and Omicron II from 13.03.2022 (Figure 1a in Health Protection Surveillance Centre (2022b)). The open data platform (Government of Ireland 2022; Ordnance Survey Ireland 2024) by the Irish Government provides weekly updated multivariate time series data on confirmed daily cumulative COVID-19 cases for all 26 Irish counties, starting from the beginning of the pandemic in February 2020. A COVID-19 case is attributed to

the county in which the patient has their primary residence.¹ To our knowledge, spatially more detailed COVID-19 data is not available for Ireland. The limited granularity makes it difficult to implement fine-scale spatial models. The GNAR model was originally demonstrated using county-level data, indicating its potential for modeling infectious diseases at lower resolutions. The cumulative case count is given for 100,000 inhabitants. Age profiles vary across counties and COVID-19 infection rates are age dependent. Hence, the cumulative case count is adjusted for age distribution according to the 2016 census of Ireland, to ensure inter-county comparability (Central Statistics Office 2016). In our dataset, the first COVID-19 case was registered in Dublin on 02.03.2020 and the last reported date is 23.01.2023, spanning a total of 152 weeks (Ordnance Survey Ireland 2024). From 20.03.2020 onward, COVID-19 cases were recorded in every Irish county. The daily COVID-19 data is aggregated to a weekly level to avoid modelling artificial weekly effects (Kubiczek and Hadasik 2021; Sartor 2020). Due to delayed reporting during winter 2021/22, the weekly COVID-19 incidences from 12.12.2021 to 27.02.2022 are averaged over a window of 4 weeks (Health Protection Surveillance Centre 2021, 2022c; Wei 2006).

The main COVID-19 regulations restricting physical movement and social interaction between Irish counties (Brennan 2022; Loughlin 2022; McQuinn et al. 2021) are used to naturally split the data into five sequential subsets, where the COVID-19 incidence is small at the beginning of the pandemic and shows a clear increasing trend over time. We denote by $\bar{\sigma}$ the average standard deviation in COVID-19 incidence across the 26 Irish counties within the considered data subset. The splits are

- (1) Start of the pandemic, with gradually stricter movement restrictions and lockdowns (*restricted* phase), from 27.02.2020 for 25 weeks ($\bar{\sigma} = 19.17$);
- (2) County-specific movement restrictions (*less restricted* phase), from 18.08.2020 for 18 weeks ($\bar{\sigma} = 35.05$);
- (3) Level-5 lockdown (*restricted* phase), with inter-county travel restrictions, from 26.12.2020 for 20 weeks ($\bar{\sigma} = 148.79$);
- (4) Allowance of non-essential inter-county travel (*less restricted* phase), from 10.05.2021 for 43 weeks ($\bar{\sigma} = 173.17$);
- (5) End of all restrictions (*unrestricted* phase), from 06.03.2022 for 46 weeks ($\bar{\sigma} = 101.46$).

The datasets are grouped to avoid small sample sizes. Time gaps in either dataset are inserted as missing data. The groups are

- *Restricted* dataset: dataset 1 and 3; representing pandemic situations with strict COVID-19 restrictions, including an inter-county travel ban (Brennan 2022); contains 45 weeks of observation ($\sigma_r = 99.27$)

¹ This attribution is only partially reliable due to a lack of validation during infection surges (Health Protection Surveillance Centre 2022a).

- *Unrestricted* data set: dataset 2, 4 and 5; representing periods with fewer or no regulations, in particular no inter-county travel limitations (McQuinn et al. 2021); contains 107 weeks of observation ($\sigma_{ur} = 129.62$)

Methodology

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ denote a fixed, simple, undirected, unweighted network with vertex set \mathcal{V} containing N vertices and edge set \mathcal{E} ; an edge between vertices i and j is denoted by $i \sim j$. The neighbourhood of a subset of vertices $A \subset \mathcal{V}$ is defined as the set of neighbours outside of A to the vertices in A , $N(A) = \bigcup_{i \in A} \{j \in \mathcal{V} \setminus A : i \sim j\}$. The set of r^{th} -stage neighbours, or the r^{th} -stage neighbourhood, for vertex $i \in A$ is defined recursively as $N^{(0)}(i) = \{i\}$ and

$$N^{(r)}(i) = N\left(N^{(r-1)}(i)\right) \setminus \bigcup_{q=1}^{r-1} N^{(q)}(i).$$

COVID-19 networks: constructions and properties

The key to the GNAR model is the network. The true network underlying the data generating process (in this case, who infected whom in the spread of the disease) is usually unknown. Ideally, expert knowledge can be leveraged to build a network that captures the relationship between vertices, representing the subjects of interest. Networks to model the spread of an infectious disease, such as COVID-19, are frequently modelled off of human mobility patterns, which are considered to have a shaping influence on disease spread (e.g. Colizza et al. 2006; Jia 2020; Li et al. 2021; Mo 2021; Sun et al. 2021). To our knowledge, detailed information on weekly population flow between Irish counties is unavailable. Hence, we construct implicit COVID-19 transmission networks (COVID-19 networks hereafter) based on geographical approaches, in line with Knight et al. (2016).

In the *Railway-based network*, an edge is established between two counties if there exists a direct train link between the respective county towns (without change of trains) and the county towns are closest to each other on this train connection. The *Queen's contiguity network* connects each county to the counties it shares a border with (Sawada 2022). The *Economic hub network* adds an additional edge between each county and its nearest economic hub to the Queen's contiguity network: Dublin, Cork, Limerick, Galway or Waterford (Gardham 2022). To measure the distance to the nearest economic hub we use the *Great Circle distance* $d_C(i, j)$, the shortest distance between two points on the surface of a sphere (Weisstein 2002). For two points i, j with latitude δ_i, δ_j and longitude λ_i, λ_j on a sphere of radius $r > 0$,

$$d_C(i, j) = r \cdot \cos^{-1}\left(\cos(\delta_i) \cdot \cos(\delta_j) \cdot \cos(\lambda_i - \lambda_j) - \sin(\delta_i) \cdot \cos(\delta_j)\right).$$

The *K-nearest neighbours network* (KNN) connects a vertex with its K nearest neighbours with respect to d_C (Bivand 2022; Eppstein et al. 1997). The *distance-based neighbour network* (DNN) constructs an edge between counties if their Great Circle distance d_C lies within a certain range $[l, r]$; this construction is similar to the one used in Knight et al. (2016). For the COVID-19 network, we set $l = 0$ and consider r a hyperparameter, chosen large enough to ensure that no vertex is isolated. The maximum value for r is

determined by the largest distance between any two vertices, for which it returns a fully connected network (Bivand et al. 2013).

In addition to these geographical networks, the *Delaunay triangulation* constructs geometric triangles between vertices such that no vertex lies within the circumsphere of any constructed triangle (Chen and Xu 2004), thus ensuring that there are no isolated vertices. The *Gabriel*, *Sphere of Influence network* and *Relative neighbourhood* are obtained from the Delaunay triangulation network by omitting certain edges. In a *Gabriel network*, vertices x and y in Euclidean space are connected if they are *Gabriel neighbours*; that is,

$$d(x, y) \leq \min \left(\sqrt{d(x, z)^2 + d(y, z)^2} \mid z \in \mathcal{V} \right)$$

where $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ denotes the Euclidean distance. In a *Sphere of Influence network* (SOI), long distance edges in the Delaunay triangulation network are eliminated and only edges between *SOI neighbours* are retained, as follows. For $x \in \mathcal{V}$ and d_x the Euclidean distance between x and its nearest neighbour in \mathcal{V} , let C_x denote the circle centred around x with radius d_x . For $y \in \mathcal{V}$, the quantities d_y and C_y are defined analogously. Vertices x and y are SOI neighbours if and only if C_x and C_y intersect at least twice, preserving the symmetry property of the Delaunay triangulation (Bivand et al. 2013). The *Relative neighbourhood network* only retains edges between relative neighbours,

$$d(x, y) \leq \min \left(\max(d(x, z), d(y, z)) \mid z \in \mathcal{V} \right) .$$

The Relative neighbourhood network is contained in the Delaunay triangulation, SOI and Gabriel network, and is the sparsest of the four networks (Bivand et al. 2013). Finally, the *Complete network* represents the homogeneous mixing assumption, where all countries are connected (Bansal et al. 2007).

Figure 1 shows the Economic hub network and the KNN network ($k = 11$ and $k = 21$) for Ireland. Figures of the other networks are found in the Supplementary Material A; network summaries are provided in Table 1.

The networks are created based on the literature on spatial modelling; Bivand et al. (2008, pp. 239–251) suggests the Delaunay network and its variants, Queen's contiguity network, distance-based networks, and K -nearest neighbour networks. In De Souza et al. (2021) it was found that infrastructure has an effect on the spread of COVID-19 in Brazil; here we use the railway network as infrastructure network. The Economic hub network is motivated by the idea of including a proxy of commuter flows in the network construction, as commuting to work has been shown in Mitze and Kosfeld (2022) to be related to the spread of COVID-19 in Germany.

Generalised network autoregressive models

Network-based time series models incorporate non-temporal dependencies in the form of networks in addition to temporal dependencies commonly modeled in time series models (Knight et al. 2019, 2016; Zhu et al. 2017). In contrast to standard time series methodology and spatial models (Box et al. 2015; Hamilton 2020; Wei 2006),

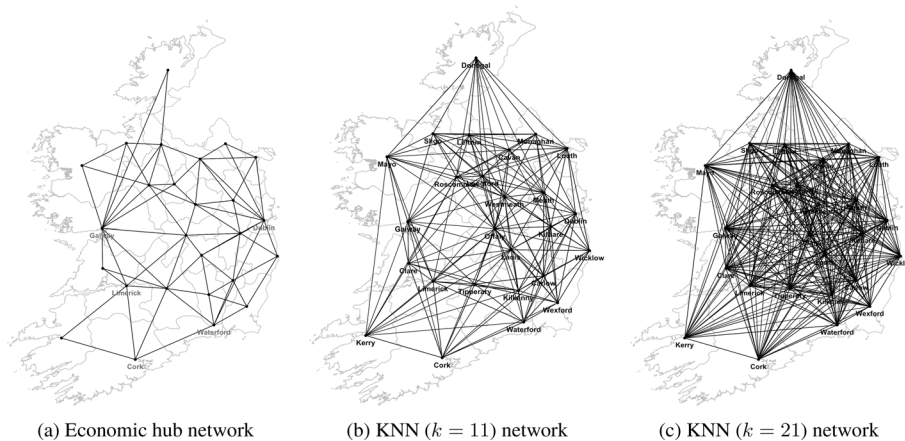


Fig. 1 Map of Ireland and COVID-19 networks; economic hub towns marked in blue

network-based time series models are not limited to geographic relationships but can incorporate any generic network. As COVID-19 is an infectious disease with spatial spreading behaviour, warranting the construction of networks based on spatial information, we use terms relating to spatial dependence in our exposition. Other types of dependence could easily be incorporated in the model through networks which reflect the hypothesised dependence.

The *global α generalised network autoregressive models* GNAR ($p-s_1, \dots, s_p$) models the observation $X_{i,t}$ for a vertex i at time t as the weighted linear combination of an autoregressive component of order p and a network neighbourhood autoregressive component of a certain order, also called neighbourhood stage; for $i = 1, \dots, p$, the entry s_i gives the largest neighbourhood stage considered for vertex i when regressing on up to p past values. In our analysis, $X_{i,t}$ denotes the 1-lag difference in weekly COVID-19 incidence over time t for county i . The effect of neighbouring vertices depends on some weight $\omega_{i,q}$. A GNAR model GNAR($p-s_1, \dots, s_p$) has the following form,

$$X_{i,t} = \sum_{j=1}^p \left(\alpha_j X_{i,t-j} + \sum_{r=1}^{s_j} \sum_{q \in N^{(r)}(i)} \beta_{j,r} \omega_{i,q} X_{q,t-j} \right) + \varepsilon_{i,t} \tag{1}$$

where $\varepsilon_{i,t} \sim N(0, \sigma_i^2)$ are uncorrelated.² As weights $\omega_{i,q}$ we choose the normalised inverse shortest path length (SPL) weight, where $d_{i,q}$ denotes the shortest path length (SPL) (Knight et al. 2016); in connected networks, $1 \leq d_{i,q} < \infty$ for $i \neq q$. For $i \in \mathcal{V}$ and $q \in N^{(r)}(i)$, we thus take, $\omega_{i,q} = d_{i,q}^{-1} / (\sum_{k \in N^{(r)}(i)} d_{i,k}^{-1})$. If the network is dynamic over time instead of static, the weights can be constructed to depend on time (Knight et al. 2016).

The general GNAR model relies on vertex specific coefficients α_{ij} , instead of vertex unspecific autoregressive coefficients, α_j , indicating vertex specific temporal dependence. This modification is comparable to including individual random effects in regression models.

² We define $\sum_{r=1}^0 (\cdot) := 0$.

To fit a GNAR model, we must choose two hyperparameters, the lag p , or α -order, and the vector of neighbourhood stages, $s = (s_1, \dots, s_p)$, also called β -order. They can be determined either through expert knowledge, e.g. on the spread of infections, or through a criterion-based search (Knight et al. 2016). The model coefficients are computed via Estimated Generalised Least Squares (EGLS) estimation (Knight et al. 2016; Leeming 2019; Lütkepohl 1991).³

GNAR model selection and predictive accuracy

For our analysis of the Irish COVID-19 data, model selection, i.e. the choice of α - and β -order, is performed by minimizing the *Bayesian Information Criterion* (BIC) (Knight et al. 2016). The BIC avoids overfitting by penalizing the observed likelihood L by the dimensionality of the required parameters (Schwarz 1978). For a sample X of size n and a parameter θ of dimension k , $\text{BIC}(k, n) = k \log(n) - 2 \cdot \log(L(X; \theta))$.

The GNAR package assumes Gaussian errors (R Package Documentation 2022); under this assumption, the BIC is consistent. This assumption could be weakened; it can be shown that the BIC is consistent for the GNAR model (1) if the error term is i.i.d. with bounded fourth moments (Leeming 2019; Lütkepohl 2005; Lv and Liu 2014).

The predictive accuracy of a GNAR model is measured by the *mean absolute scaled error* (MASE), due to its insensitivity towards outliers, its scale invariance and its robustness (Hyndman and Koehler 2006). MASE is defined for each county i as the ratio of absolute forecasting error $\hat{\varepsilon}_{i,t} = |X_{i,t} - \hat{X}_{i,t}|$ divided by the mean absolute error between true and a naive 1-lag random walk forecast for the entire observed time period $[1, T]$ (Hyndman and Koehler 2006; Urrutia et al. 2022);

$$|q_{i,t}| = \frac{|X_{i,t} - \hat{X}_{i,t}|}{\frac{1}{T-1} \sum_{l=2}^T |X_{i,l} - X_{i,l-1}|}.$$

Data exploration

The weekly incidence differences

The GNAR model requires stationary data. Stationary data has no trend over time and is homogeneous, i.e. has time-independent variance (Knight et al. 2016). The weekly COVID-19 count is clearly not stationary, as it shows an increasing trend in Fig. 2. To remove any linear trend, we perform 1-lag differencing on the weekly COVID-19 incidence for the 26 Irish counties, resulting in the incidence difference, (*1-lag*) *COVID-19 ID*, between two subsequent weeks (Montgomery et al. 2015). We assess the stationarity by applying a Box-Cox transformation to each data subset. Figure 12 in the Supplementary Material C, $\hat{\lambda}$ indicate that no further transformation is required.

Constructed networks

For the COVID-19 KNN network, neighbourhood sizes sequencing from $k = 1$ to the fully connected network, $k = 25$, with step size 2, are considered. The minimal distance

³ Additional information in Supplementary Material B.3.

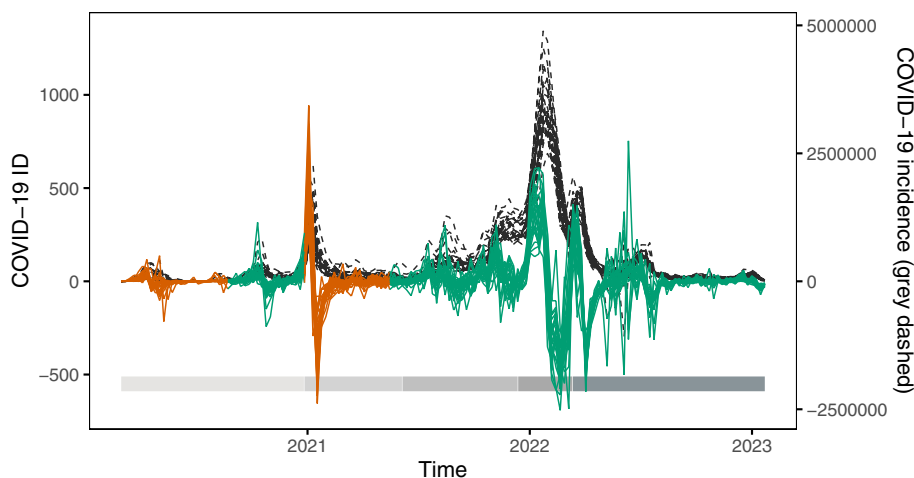


Fig. 2 Weekly (1-lag) COVID-19 incidence difference (ID) and weekly COVID-19 incidence (dark grey dashed lines) for 100,000 inhabitants, from the start of the pandemic in March 2020 to mid January 2023, red for restricted phases, green for unrestricted; predominant COVID-19 virus variants shown by color scale at the bottom, in order: Original, Alpha, Delta, Omicron I and Omicron II

for the COVID-19 DNN network measures 90.3 km, between Kerry and Cork, and the maximal value 338.5 km, between County Cork and Donegal. The KNN and DNN network parameters are chosen to minimise the BIC of the associated GNAR model. For the restricted pandemic phase, the KNN network has $k = 11$ and the DNN network $d = 325$. For the unrestricted pandemic phase, it is $k = 21$ and $d = 325$.

There is considerable variability in network characteristics, Table 1, in particular regarding the network density. The KNN and DNN networks for the abovementioned hyperparameters have much larger average degree than the other networks; the sparsest network is the Relative neighbourhood network. Consequently, the SPL is shortest in the denser DNN and KNN networks. The Railway-based network has the longest average SPL due to its vertex chains and the low number of shortcuts between counties. For the Economic hub network, the introduction of shortcuts to the economic hubs leads to a decrease in average SPL, i.e. the disease spreads quicker, compared to the Queen's contiguity network. The Gabriel network is sparser than the SOI network, with slightly longer average SPL. Deleting long edges in the Delaunay triangulation network to obtain the SOI network decreases the average degree and the average local clustering coefficient, but increases the average SPL. The Queen's network, the Economic hub network, the Delaunay network, the Gabriel network, and the SOI network show small world behaviour, i.e. high clustering with short SPL. To assess small world behaviour, the average SPL and average local clustering for a network is compared to a Bernoulli Random graph $G(n, m)$ with identical size n and number of edges m as the network under investigation. The Railway network has much larger average SPL than $G(n, m)$, while the dense KNN and DNN networks have almost the same average SPL and local clustering coefficient as the $G(n, m)$ network.

Although there are differences in the detailed summary statistics, the networks can be clustered according to density and average local clustering coefficient; we use the k-means algorithm, running 10 randomisations to ensure robustness over a range of $k = 1, \dots, 10$.

Table 1 Overview of network characteristics for the Railway-based, Queen's contiguity, Economic (Eco.) hub, KNN ($k = 11$ and $k = 21$), DNN ($d = 325$), Delaunay triangulation, Gabriel, SOI, Relative neighbourhood (Rel. neigh.) network; including average (av.) degree, average (av.) shortest path length (SPL), average (av.) local clustering (Clust.)

Metric	Railway	Queen	Eco. hub	KNN ($k = 11$)	KNN ($k = 21$)	DNN ($d = 325$)	Delaunay	Gabriel	SOI	Rel. neigh.
Av. degree	5	4.38	5.38	13.46	23.54	24.85	5.15	3.69	4.15	2.31
Av. SPL	4.15	2.74	2.27	1.46	1.06	1.01	2.41	3.02	2.86	3.99
Av. local clust.	0.2	0.51	0.6	0.75	0.95	0.99	0.49	0.31	0.51	0
BRG Av. SPL	2.08	2.24	2.05	1.46	1.06	1.01	2.1	2.24	2.28	3.38
BRG Av. local clust.	0.14	0.18	0.21	0.54	0.94	0.99	0.17	0.16	0.19	0.09

The average shortest path length and average local clustering coefficient for a Bernoulli Random graph $G(n, m)$, with $n = 26$, abbreviated BRG, for each network is also reported

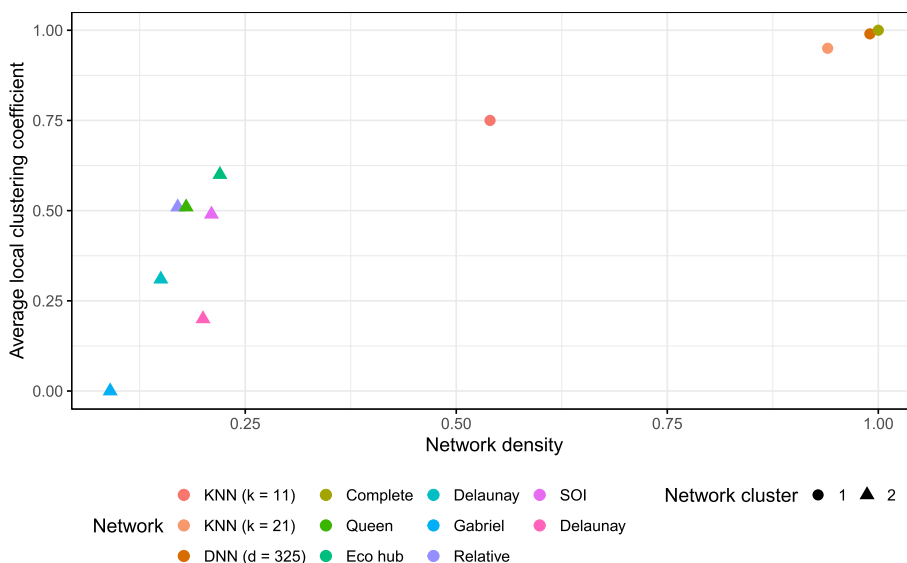


Fig. 3 Network density against average local clustering coefficient for all 10 considered COVID-19 networks; 2 clusters detectable

The corresponding elbow plot implies two clusters of COVID-19 networks. As evident in Fig. 3, one cluster has high density and high average local clustering coefficient, while the second cluster has low density and low to medium average local clustering coefficient.

Spatial effects

Intuitively, if spatial correlation is present in a network, the closer in SPL two vertices are, the more highly correlated their COVID-19 incidences. Moran’s I quantifies spatial correlation by estimating the average weighted correlation across space (Cliff and Ord 1981; Moran 1950; Zhou and Lin 2008). Let $t \in T$ and $x_i^{(t)}$ denote the COVID-19 ID for county i at time t ,

$$I^t = \frac{\sum_{i=1}^N \sum_{j=1, i \neq j}^N w_{ij} \cdot (x_i^{(t)} - \bar{x}^{(t)})(x_j - \bar{x}^{(t)})}{W_0 \cdot \frac{1}{N} \sum_i (x_i^{(t)} - \bar{x}^{(t)})^2}$$

where $W_0 = \sum_{i,j=1}^N w_{ij}$ for normalisation. For non-neighbours, the weights are zero, i.e. $\forall r : j \notin N^{(r)}(i) : w_{ij} = 0$. Here we use weights $w_{ij} = e^{-d_{ij}}$ where d_{ij} denotes the SPL between vertex i and j (Coscia 2021). The spatial dependency between counties varies strongly over time for every network, see Fig. 4 and Figure 11 in Supplementary Material A.1. Peaks in Moran’s I coincide with peaks in the 1-lag COVID-19 ID at the beginning of the pandemic as well as during the winters 2020/21 and 2021/22. The introduction of restrictive regulations, e.g. lockdowns, shows a decreasing trend in Moran’s I while the ease of restrictions from summer 2021 onward has lead to an increasing trend in Moran’s I. This indicates a network effect in the data, which is associated with the inter-county mobility, and becomes particularly evident after the official end of pandemic restriction in March 2022. To further assess the presence of non-linear spatial correlation, we also apply Moran’s I to the ranks of the COVID-19 ID at each time point t over the duration of data observation. The rank-based Moran’s I follows the same trajectory, with less extreme peaks, as evident in Fig. 5.

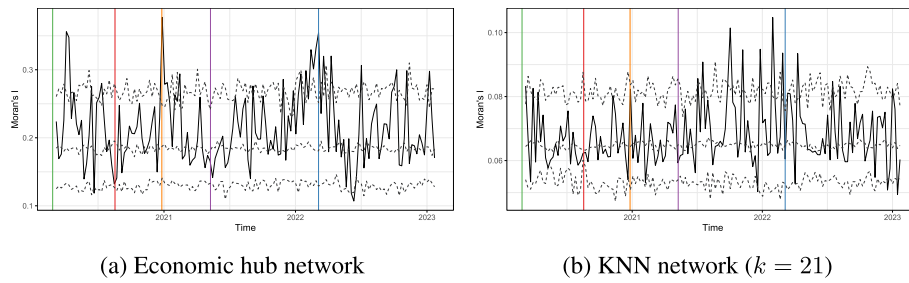


Fig. 4 Moran's I across time, with weights based on SPL; main COVID-19 regulations by the Irish Government indicated by vertical lines; in order: initial lockdown, county-specific restrictions, Level-5 lockdown, allowance of inter-county travel, official end of all restrictions; 95% credibility interval in grey dashed

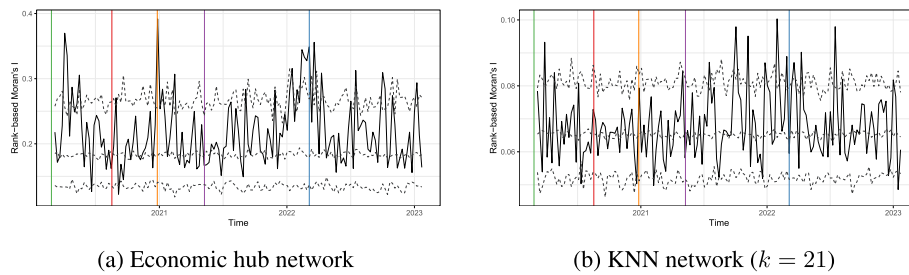


Fig. 5 Rank-based Moran's I across time for the optimal network for the restricted and unrestricted pandemic phase; main COVID-19 regulations by the Irish Government indicated by vertical lines; in order: initial lockdown, county-specific restrictions, Level-5 lockdown, allowance of inter-county travel, official end of all restrictions; 95% credibility interval in grey dashed

To statistically assess spatial dependence we perform a permutation test for the standard as well as the rank based Moran's I by permuting the COVID-19 cases between counties $R = 100$ times and computing Moran's I. A date-specific 95% credibility interval $\forall t = 1, \dots, T : [m_{t,l}, m_{t,u}]$ based on empirical quantiles ($q = 0.025, 0.5, 0.975$) is constructed. Under the null hypothesis, assuming no correlation between network structure and COVID-19 incidence, 5% ($0.05 \cdot T \approx 8$) of observed Moran's I values m_t over time $t = 1, \dots, T$ are expected to lie outside the time dependent 95% credibility interval, $[m_{t,l}, m_{t,u}]$. If the proportion $N_m = T^{-1} \sum_t \{\mathbb{I}(m_t > m_{t,u}) + \mathbb{I}(m_t < m_{t,l})\}$ of rejected tests over time is greater than expected under the null, we conclude that the network distance has an effect on the correlation between COVID-19 incidence.⁴ With exponential SPL weights, the proportion of rejected tests for the restricted and unrestricted data set are: Railway-based network $N_m = (0.25, 0.142)$, Queen's contiguity network $N_m = (0.227, 0.217)$, Economic hub network $N_m = (0.25, 0.179)$, KNN network ($k = 11$ for restricted, $k = 21$ for unrestricted phase) $N_m = (0.159, 0.151)$, DNN network ($d = 325$ for both phases) $N_m = (0.068, 0.094)$, Delaunay triangulation network $N_m = (0.205, 0.189)$, Gabriel network $N_m = (0.114, 0.132)$, SOI network $N_m = (0.182, 0.198)$, Relative neighbourhood network $N_m = (0.159, 0.189)$. The proportions indicate a significant spatial correlation. Depending on the network, the proportion for either the restricted or the unrestricted data set is larger. The rank-based

⁴ The constructed test is not a proper significance test in the statistical sense, given the dependence between tests over time. It rather provides a rough intuition regarding the spatial correlation in COVID-19 incidence assuming different underlying networks.

Table 2 Overview over the best performing model and network for restricted and unrestricted pandemic phases

Data subset	Network	GNAR model	BIC	$\bar{\varepsilon}$ (sd)	Av. MASE (sd)
<i>Restricted</i>	Eco. hub	GNAR(7-3,1,1,0,0,0,0)	58.91	-1.60 (16.51)	0.90 (0.73)
<i>Unrestricted</i>	KNN-21	GNAR(7-1,1,1,1,0,0,0)	190.07	-13.46 (14.67)	0.94 (0.74)
<i>Restricted</i>	ARIMA		534.58	-0.78 (19.08)	1.22 (1.34)
<i>Unrestricted</i>	ARIMA		670.27	-16.65 (18.50)	1.21 (0.96)

average residual $\bar{\varepsilon}$ and average (av.) MASE indicated for the predicted 5 weeks at the end of the observed time period across all counties, 11.04.2021–09.05.2021 for the restricted dataset and 25.12.2022–23.01.2023 for the unrestricted dataset; standard deviation (sd) across counties in brackets

Moran's I permutation test obtains similar proportions of rejected tests, indicating a significant nonlinear spatial correlation (data not shown).

Results

GNAR model fitting

To assess the benefit of accounting for network effects, the GNAR model is compared to a standard county-specific ARIMA model which is allowed to include seasonal effects.⁵ The GNAR model allows for more flexible spatial-temporal dependencies than other network-based time series models as detailed in Supplementary Material B.2, including the ARIMA models, with a network specific selection of α - and β -order. The models are selected by choosing the model with the lowest BIC.

On average, the ARIMA model achieves a $BIC = 1846.88$ on the entire data set, $BIC = 534.58$ for the restricted data set and $BIC = 670.27$ for the unrestricted data set. Optimal⁶ GNAR models for each COVID-19 network achieve much lower BIC. For the restricted phase, the best phase-specific GNAR model yields $BIC = 58.91$ on the Economic hub network, and for the unrestricted phase $BIC = 190.07$ on the KNN ($k = 21$) network, see Table 2. When fitted on the entire data set, the GNAR-5-11110 model with the KNN network ($k = 21$)⁷ achieves the lowest $BIC = 193.95$.⁸ All BICs are considerably smaller than those obtained from the ARIMA fit, with the minimal BIC for the entire data set much larger than the BIC for the restricted phase, and also larger than the BIC for the unrestricted phase, thus justifying the use of GNAR models, as well as the split of the data.

The nature of the virus suggests that the transmission of COVID-19 between Irish counties may depend strongly on the population flow between counties (Lotfi et al. 2020). Protective COVID-19 restrictions taken by the Irish Government restricted and at times forbade inter-county travel in Level 3-5 lockdowns (Department of the

⁵ Fitted for each county individually, the ARIMA models might differ in orders and parameters.

⁶ *Optimal* describes the best performing combination of α - and β -order as well as global- α setting which obtain the minimal BIC value. The a-priori range of α -order spans $\{1, \dots, 7\}$. The maximum lag to consider follows from Schwert's rule (Ng and Perron 1995), applied to the minimum number of weeks across the individual five datasets ($w = 18$). The possible choices for the β -order are listed in Supplementary Material B.4. The maximum neighbourhood stage that can be included in the GNAR model is determined by the smallest maximum SPL across most networks, which is 5. For the complete network, only 1st stage neighbourhood can be modelled, while for the Economic hub network the maximum neighbourhood stage is 4.

⁷ From hereon referred to as the KNN network.

⁸ For more detail on fitting a GNAR model to each COVID-19 network on the entire data set, see the Supplementary Material D.1.

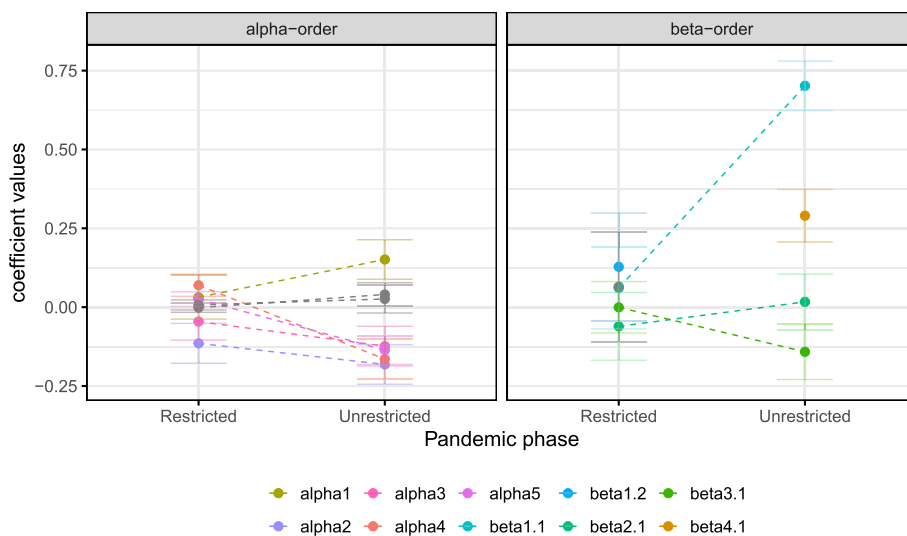


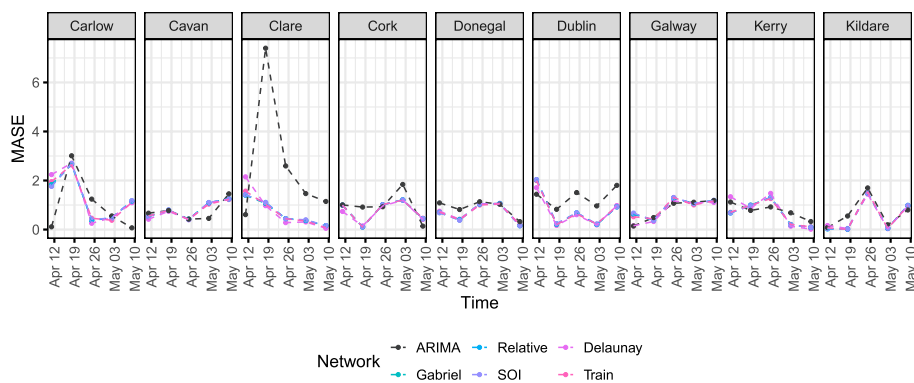
Fig. 6 Development of GNAR model coefficients for the restricted and unrestricted pandemic phase; restricted phase with Economic hub network, unrestricted phase with KNN ($k = 21$) network

Taoiseach 2020; McQuinn et al. 2021). As supported by the positive and negative trends in Moran’s I, the spatial dependence of COVID-19 incidence across counties is likely to have decreased during lockdowns and increased during periods in which inter-county travel was allowed (Wang 2022). This provides additional subject specific motivation to train a pandemic phase specific GNAR model.

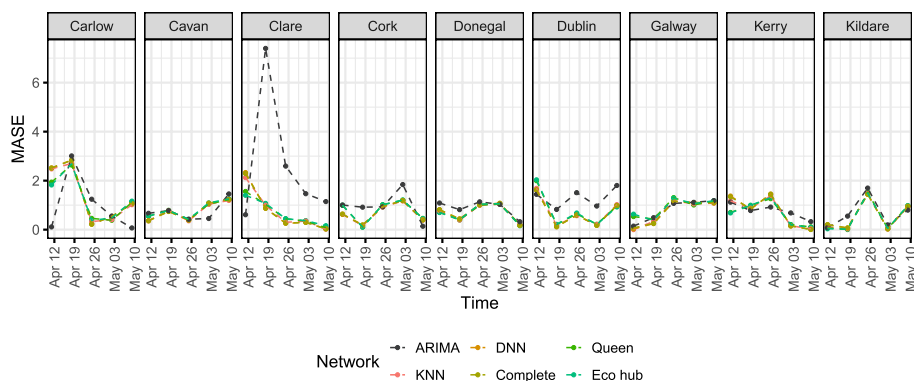
Pandemic phases

Table 2 summarises the optimal GNAR models and COVID-19 network for the restricted and unrestricted data set. For both phases, the best performing GNAR model select an autoregressive component of order 7. The average MASE are smaller for the restricted than the unrestricted pandemic phases, implying that GNAR models are more suited to predicting periods with strict regulations than periods with fewer or no restrictions. The variance for residuals and MASE is smaller for the GNAR model than the ARIMA model. The optimal network for the unrestricted pandemic phase is much denser than the optimal network for the restricted phase. As evident from Tables 3, 4 in Supplementary Material D.1, the BIC values for the optimal GNAR model lie within the range [58.91, 68.36] for the restricted data set and within the range [190.07, 192.56] for the unrestricted data set. Figure 13 in Supplementary Material D.1 illustrates that denser networks perform better for the unrestricted data set while sparse networks achieve lower BIC for the restricted data set, on average.

A decrease in inter-county dependence due to COVID-19 restrictions should result in decreasing values for the β -coefficients in the GNAR model. This hypothesis can only be partially verified, see Fig. 6. The absolute value of β -coefficients increases from the restricted to the unrestricted phase, implying increased spatial dependence after COVID-19 restrictions have been eased or lifted. We note that the GNAR model picks up a decrease in temporal dependence in COVID-19 ID. As a disease spreads



(a) Delaunay triangulation, Gabriel, Relative neighbourhood, SOI and Railway-based network



(b) KNN ($k = 11$), DNN ($d = 325$), Complete, Queen’s contiguity and Economic hub network

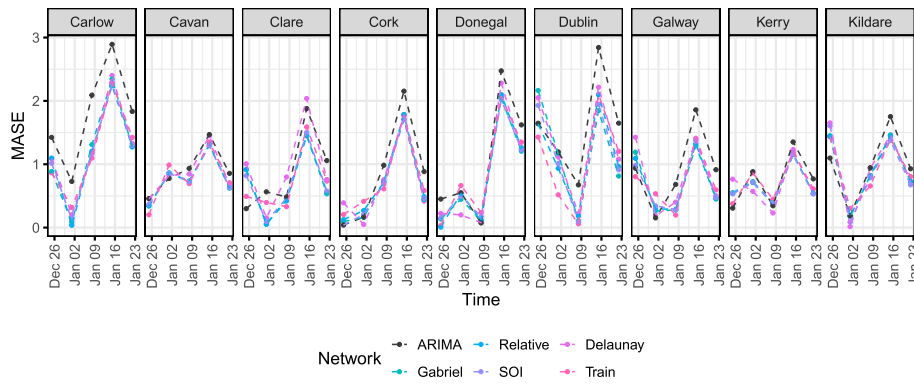
Fig. 7 MASE values for the restricted pandemic phase, for 9 selected counties

more freely due to lenient or no restrictions, it has been observed in other data studies that case numbers can grow more erratic and become less dependent on historic data (Firth 2020; Kraemer 2020). This effect, in addition to peaks and high volatility in COVID-19 ID observed during pandemic phases with less strict regulations, might contributed to the negative α -coefficient values for the unrestricted data set. In general, an increase in noise during unrestricted pandemic phases might contribute to the decrease in predictive performance of the GNAR model.

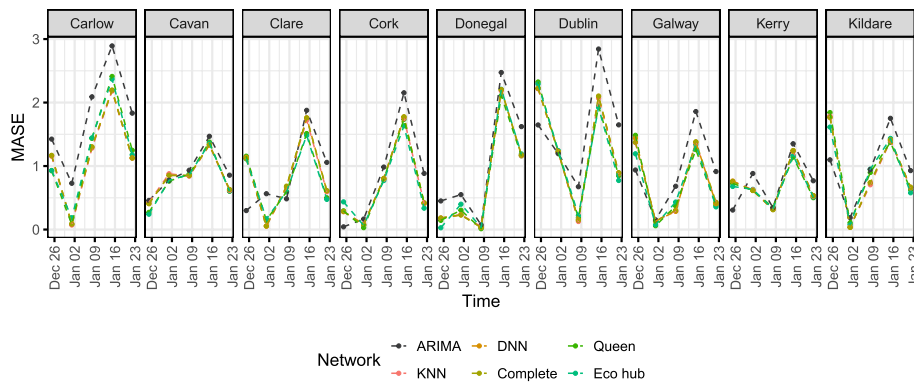
Identical observations can be made when considering how the coefficients develop between the restricted and unrestricted phase for the GNAR model that is optimal for the entire data set, namely, GNAR(5-1,1,1,1,0) with the KNN ($k = 21$) network; the β -coefficients increase in absolute value for the unrestricted phase compared to the restricted phase, see Fig. 14 in Supplementary Material D.2.

The predictive accuracy for both datasets is comparable and varies from county to county, see Figs. 7 and 8 for 9 example counties; MASE values for the remaining counties follow similar patterns.

For the restricted phase, GNAR models achieve lower MASE than the ARIMA models except for counties Cavan, Galway, Leitrim, Louth, Sligo, Tipperary, Roscommon, for which the ARIMA model performs equally well. For the unrestricted phase, the ARIMA model predicts particularly poorly for counties Carlow, Kilkenney, Louth and Waterford,



(a) Delaunay triangulation, Gabriel, Relative neighbourhood, SOI and Railway-based network



(b) KNN ($k = 21$), DNN ($d = 325$), Complete, Queen's contiguity and Economic hub network

Fig. 8 MASE values for the unrestricted pandemic phase, for 9 selected counties

and for the restricted phase for counties Cavan, Clare, Limerick and Wexford. We note that Cavan, Leitrim and Sligo have a border with Northern Ireland, which could introduce some confounding factors.

For the restricted phase, the predicted COVID-19 ID values differ more strongly between the GNAR model and the ARIMA model, see Figure 16 in Supplementary Material D.3. For the unrestricted phase, the GNAR and ARIMA models estimate roughly the same trajectory while the former achieves smaller residuals for most counties.

Assessing the model assumptions

The above models assume that the observations follow a Gaussian i.i.d. error structure. To assess this assumption, we test whether the residuals $\hat{\epsilon}_{i,t}$ follow a normal distribution with a county-specific Kolmogorov-Smirnov test, aggregated over time. In contrast, we obtain a majority of significant p -values across counties for the unrestricted phase (the counts of p -values are $\#p \leq 0.025 = 21$, $\#p > 0.025 = 5$),⁹ raising doubts about the Gaussianity assumption in the unrestricted phase. We obtain primarily insignificant

⁹ Insignificant p -values were established in counties Donegal, Dublin, Kilkenny, Laois, Offaly and Sligo.

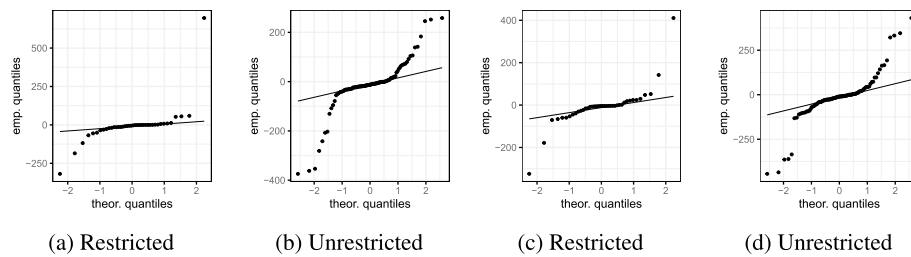


Fig. 9 QQ-plot for the residuals from the best performing GNAR model and network (Economic hub network for restricted phase, KNN ($k= 21$) network for unrestricted phase) for restricted and unrestricted pandemic phase for county Dublin (left) and Donegal (right)

p -values across counties for the restricted phase ($\#p \leq 0.025 = 8, \#p > 0.025 = 18$)¹⁰, so that the Gaussian assumption is not rejected.

Table 5 in Supplementary Material D.4 details the average MASE, average residual and p -value for each county, resulting for the two optimal GNAR models for the restricted and unrestricted data set. The Gaussian nature of residuals indicate suitability of the GNAR model to model restricted pandemic phases and ensure consistency in coefficient estimates. For the unrestricted phase, the Gaussianity in the model assumptions could not be statistically verified. These conclusions are supported by the county-specific QQ-plots in Supplementary Material D.4. The hyperparameters α - and β -order were set data-adaptively to minimize the BIC criterion, which assumes Gaussian error. Under the assumption of Gaussian error, the BIC would be minimal for the correct higher order network dependence. As we examined a large range of β -order choices, this deviance from Gaussianity leads us to propose investigating alternative error structures as future work (Fig. 9).

The GNAR model further assumes that the errors are uncorrelated. To assess this assumption, the residuals are investigated according to their temporal as well as spatial autocorrelation using the Ljung-Box test and Moran’s I based permutation test (Ljung and Box 1978; Moran 1950). The former concludes significant temporal correlation for short-term lags in the GNAR residuals for each county. Thus, there is evidence that the GNAR model insufficiently accounts for temporal dependence in COVID-19 incidence in subsequent weeks. The residuals show remaining spatial autocorrelation. The Moran’s I based permutation test counts $N_m = 8$ Moran’s I values outside the corresponding 95% credibility interval (expected $0.05 \cdot 45 \approx 2$) for the restricted phase and $N_m = 16$ for the unrestricted phase (expected $0.05 \cdot \dots \cdot 107 \approx 5$). The reduction in spatial correlation for the restricted phases and the Economic hub network is greater ($N_m = 10$ on COVID-19 cases to $N_m = 8$ for residuals) than for the unrestricted phases and the KNN network ($N_m = 16$ for COVID-19 cases and residuals). We conclude that there is evidence that the GNAR model may not sufficiently incorporate the spatial relationship in COVID-19 case numbers across counties. These possible violations of the model assumptions have to be taken into account when interpreting the model fit.

¹⁰ The Kolmogorov-Smirnov test was significant for counties Donegal, Galway, Kilkenny, Laois, Limerick, Offaly, Sligo and Wicklow.

Discussion

In general, a network model could be a powerful tool to inform the spread of infectious diseases, see for example (Britton 2019; Overton 2020). In this paper, we modelled the COVID-19 incidence across the 26 counties in the Republic of Ireland by fitting GNAR models, leveraging different networks to represent spatial dependence between the counties. While we do not assume that the disease only spreads along the network, we consider the edges to represent the main trajectory of the infection. The analysis shows that there is a clear network effect, but networks of similar density perform similarly in predictive accuracy. GNAR models perform better on data collected during pandemic phases with inter-county movement restrictions than data gathered during less restricted phases. Sparse networks perform better for the restricted data set, while denser networks achieve lower BIC for the unrestricted data set.

There are some caveats relating to the model. First, the time series for the restricted phase and for the unrestricted phase are actually concatenated time series; Fig. 2 shows the original time series. The concatenation was carried out because of data availability, but it is possible that it obfuscates some potentially interesting phase-specific signals. As seen in Fig. 2, even after differencing the time series do not display clear stationarity. Further, COVID-19 is subject to “seasonal” effects, e.g. systematic reporting delays due to weekends and winter waves (Kubiczek and Hadasik 2021; Nichols 2021; Sartor 2020). The GNAR model does not have a seasonal analogue which can incorporate seasonality in data, like SARIMA for ARIMA models (Shumway et al. 2000). Future work could introduce a seasonal component to the GNAR model, improving its applicability to infectious disease modelling. There may be other spatio-temporal patterns such as non-linear effects which the GNAR model currently does not include. Moreover, the COVID-19 pandemic had a strong influence on mobility patterns (COVID-19 Community Mobility Report 2022; Manzira et al. 2022), in particular due to restrictions of movement and an increased apprehension towards larger crowds. Considering only static networks may introduce a bias to the model (Bansal et al. 2007; Mo 2021; Perra 2012). Future work could therefore explore how GNAR models can include dynamic networks to incorporate a temporal component of spatial dependency. Alternative weighting schemes for GNAR models could be investigated to account for differences in edge relevance across time and network.

Regarding the theory of GNAR models, alternative error distributions, such as a Poisson distributed error term as in Armillotta and Fokianos (2021), could be explored given the indication of non-Gaussian residuals for the unrestricted pandemic phase. The stability of parameter estimation in GNAR models also warrants further investigation. The network constructions themselves could also be refined. Simulations have shown that GNAR models are sensitive to network misspecifications. Omitting edges may result in bias in the GNAR coefficients. While this paper has carried some robustness analysis regarding network choice, future analysis could focus on more content-based approaches to constructing networks, e.g. building a network based on the intensity of inter-county trade, computed according to the gravity equation theory (Chaney 2018). Many researchers have successfully modelled the initial spread of COVID-19 from Wuhan across China based on detailed mobility patterns, e.g. Jia (2020); Kraemer (2020). Finally, our statistical analysis did not include information about the dominant strain.

With COVID-19 being an evolving disease, different strains may display different transmission patterns. If more detailed data become available then this question would also be of interest for further investigation. However, this study has illustrated that it may be of benefit to use a GNAR model for the spread of an infectious disease, in particular during movement restrictions, so that the spread is mainly local. It has also detailed a range of possible network choices, and it has provided a set of tests to assess the performance of the model and its fit.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-024-00634-2>.

Supplementary Material.

Acknowledgements

The authors would like to thank Emma Davies-Smith for insightful feedback, and Marina Knight, Guy Nason and Matt Nunes for helpful discussions. Moreover, we thank the anonymous referees for their helpful suggestions.

Author contributions

All authors wrote and reviewed the manuscript. A.S. wrote code for the data exploration and analysis.

Funding

G.R. acknowledges support from EPSRC Grant Nos. EP/T018445/1, EP/X002195/1, EP/W037211/1, EP/V056883/1, and EP/R018472/1.

Availability of Data and materials

Data is available on the GitHub repository for the analysis as well as on the Open Data Platform by the Government of the Republic of Ireland: <https://data.gov.ie/dataset/covid-19-hpsc-county-statistics-historic-data1>.

Declarations

Ethics approval and Consent to participate

Not applicable.

Consent for publication

All authors give consent for publication.

Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2024 Accepted: 5 June 2024

Published online: 14 June 2024

References

- Armilotta M, Fokianos K (2021) Poisson network autoregression. arXiv preprint [arXiv:2104.06296](https://arxiv.org/abs/2104.06296)
- Bansal S, Grenfell B, Meyers LA (2007) When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface* 4:879–891
- Bivand R (2022) R Packages for analyzing spatial data: a comparative case study with areal data. <https://doi.org/10.1111/gean.12319>. Accessed: 2022-07-22
- Bivand RS, Pebesma EJ, Gómez-Rubio V, Pebesma EJ (2008) Applied spatial data analysis with R. Springer, Berlin
- Bivand R, Edzer P, Virgilio G-R (2013) Applied spatial data analysis with R <https://r-spatial.github.io/spdep/MISCs/nb.html>. Accessed: 2022-07-22. New York
- Box G, Jenkins G, Reinsel G, Ljung G (2015) Time series analysis: forecasting and control. Wiley, Chichester
- Brennan C (2022) A year with Covid in Ireland - timeline of incredible lockdowns, cases and deaths, pub closures and disasters <https://www.irishmirror.ie/news/irish-news/year-covid-Ireland-timeline-incredible-23585166>. Accessed 22/07/22
- Britton T et al (2019) Stochastic epidemic models with inference. Springer, Berlin
- Central statistics office (2016) Census Forms <https://www.cso.ie/en/census/2016censusforms/>. Accessed: 2022-07-06
- Chaney T (2018) The gravity equation in international trade: an explanation. *J Polit Econ* 126:150–177
- Chen L, Xu J-C (2004) Optimal delaunay triangulations. *J Comput Math* 22:299–308
- Cliff AD, Ord K (1981) Spatial processes: models & applications. Pion, London
- Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci* 103:2015–2020
- Coscia M (2021) Pearson correlations on complex networks. *J Compl Netw* 9:cnab036

- COVID-19 Community Mobility Report (2022) Mobility changes - Ireland 21 July 2022 https://www.gstatic.com/covid19/mobility/2022-07-21_IE_Mobility_Report_en-GB.pdf. Accessed: 2022-07-25
- De Souza C, Machado MF, da Silva Junior AG, Nunes BEBR, do Carmo RF (2021) Airports highways and COVID-19: an analysis of spatial dynamics in Brazil. *J Trans Health* 21:101067. <https://doi.org/10.1016/j.jth.2021.101067>
- Department of the Taoiseach (2020) Resilience and Recovery 2020-2021: Plan for Living with COVID-19 <https://www.gov.ie/en/publication/e5175-resilience-and-recovery-2020-2021-plan-for-living-with-COVID-19/>. Accessed: 2022-06-23
- Eppstein D, Paterson M, Yao F (1997) On nearest-neighbor graphs. *Discr Comput Geometry* 17:263–282
- Firth JA et al (2020) Combining fine-scale social contact data with epidemic modelling reveals interactions between contact tracing, quarantine, testing and physical distancing for controlling COVID-19. *MedRxiv*, 2020–05
- Gardham R (2022) The five largest cities in Ireland (and their investment strengths) <https://www.investmentmonitor.ai/analysis/Ireland-largest-cities-investmentdublin>. Accessed: 2022-07-06
- Government of Ireland (2022) Ireland's COVID-19 Data Hub <https://COVID-19.geohive.ie>. Accessed 23/01/23
- Hamilton JD (2020) Time series analysis. Princeton University Press, Princeton, NJ
- Health Protection Surveillance Centre (2021) First year of the COVID-19 pandemic in Ireland <https://www.hpsc.ie/az/respiratory/coronavirus/novelcoronavirus/casesinireland/COVID-19annualreports/>. Accessed: 2022-06-23
- Health protection surveillance centre (2022) Epidemiology of COVID-19 in Ireland - Dashboard <https://epi-COVID-19-hpscireland.hub.arcgis.com>. Accessed: 2023-03-20
- Health Protection Surveillance Centre (2022) Summary of COVID-19 virus variants in Ireland <https://www.hpsc.ie/az/respiratory/coronavirus/novelcoronavirus/surveillance/summaryofCOVID-19virusvariantsinireland/>. Accessed: 2022-07-22
- Health protection surveillance centre (2022) Weekly report on the epidemiology of COVID-19 in Ireland- Week 24, 2022 <https://www.hpsc.ie/az/respiratory/coronavirus/novelcoronavirus/surveillance/epidemiologyofCOVID-19inirelandweeklyreports/>. Accessed: 2022-06-28
- Hyndman R, Koehler A (2006) Another look at measures of forecast accuracy. *Int J Forecast* 22:679–688
- Jia J et al (2020) Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* 582:389–394
- Knight M et al (2016) Modelling, detrending and decorrelation of network time series. *arXiv preprint arXiv:1603.03221*
- Knight M et al (2019) Generalised network autoregressive processes and the GNAR package. *arXiv preprint arXiv:1912.04758*
- Kraemer M et al (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368:493–497
- Kubiczek J, Hadasik B (2021) Challenges in reporting the COVID-19 spread and its presentation to the society. *J Data Inform Quality (JDIQ)* 13:1–7
- Leeming K (2019) New methods in time series analysis: univariate testing and network autoregression modelling PhD thesis, University of Bristol
- Li T, Rong L, Zhang A (2021) Assessing regional risk of COVID-19 infection from Wuhan via high-speed rail. *Transp Policy* 106:226–238
- Ljung GM, Box GE (1978) On a measure of lack of fit in time series models. *Biometrika* 65:297–303
- Lotfi M, Hamblin M, Rezaei N (2020) COVID-19: transmission, prevention, and potential therapeutic opportunities. *Clin Chim Acta* 508:254–266
- Loughlin E (2022) Timeline of a pandemic: How COVID-19 changed our way of life <https://www.irishexaminer.com/news/arid-40790595.html>. Accessed 22/07/22
- Lütkepohl H (1991) Introduction to multiple time series analysis eng. ISBN: 9783540531944, Springer, Berlin and London
- Lütkepohl H (2005) New introduction to multiple time series analysis. Springer, Berlin
- Lv J, Liu JS (2014) Model selection principles in misspecified models. *J Royal Stat Soci: Series B: Stat Methodol* 76(1):141–167
- Manzira C, Charly A, Caulfield B (2022) Assessing the impact of mobility on the incidence of COVID-19 in Dublin City. *Sustain Cities Soc* 80:103770
- McQuinn C, Roche B, Cullen P (2021) Ireland reopens: inter-county travel resumes as hairdressers and non-essential retail return <https://www.irishtimes.com/news/politics/Ireland-reopens-inter-countytravel-resumes-as-hairdressers-and-non-essential-retail-return-1.4559937>. Accessed: 2023-02-02
- Mitze T, Kosfeld R (2022) The propagation effect of commuting to work in the spatial transmission of COVID-19. *J Geogr Syst* 24:5–31
- Mo B et al (2021) Modeling epidemic spreading through public transit using time-varying encounter network. *Transport Res Part C: Emerg Technol* 122:102893
- Montgomery DC, Jennings CL, Kulahci M (2015) Introduction to time series analysis and forecasting. Wiley, Chichester
- Moran P (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23
- Ng S, Perron P (1995) Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *J Am Stat Assoc* 90:268–281
- Nichols GL et al (2021) Coronavirus seasonality, respiratory infections and weather. *BMC Infect Dis* 21:1–15
- Nouvellet P et al (2021) Reduction in mobility and COVID-19 transmission. *Nat Commun* 12:1090
- Ordnance Survey Ireland (2024) COVID-19 HPSC county statistics historic data <https://data.gov.ie/dataset/covid-19-hpsc-county-statistics-historic-data1>. Accessed 24/03/25
- Overton CE et al (2020) Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. *Infect Dis Modell* 5:409–441
- Park S, Kwon Y, Soh H, Lee MJ, Son S-W (2024) Enhancing demand prediction in open systems by cartogram-aided deep learning. *arXiv preprint arXiv:2403.16049*
- Perra N et al (2012) Random walks and search in time-varying networks. *Phys Rev Lett* 109:238701
- R Package Documentation (2022) GNAR source code <https://rdrr.io/cran/GNAR/f/>. Accessed: 2022-07-14
- Sartor G et al (2020) COVID-19 in Italy: considerations on official data. *Int J Infect Dis* 98:188–190

- Sawada M (2022) Global spatial autocorrelation indices - Moran's I, Geary's C and the General Cross-Product Statistic <http://www.lpc.uottawa.ca/publications/moransi/moran.htm>. Accessed: 2022-06-28
- Schwarz G (1978) Estimating the dimension of a model. *Annals Stat* 6(2):461–464
- Shumway R, Stoffer D, Stoffer D (2000) *Time series analysis and its applications*. Springer, New York
- Sun X, Wandelt S, Zhang A (2021) On the degree of synchronization between air transport connectivity and COVID-19 cases at worldwide level. *Transp Policy* 105:115–123
- Urrutia P et al (2022) SARS-CoV-2 Dissemination using a network of the US counties in operations research forum 3, pp. 1–23
- Wang Y et al (2022) Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models. *BMC Infect Dis* 22:1–12
- Wei W (2006) *Time series analysis*. Addison-Wesley, Redwood City, CA
- Weisstein EW (2002) Great circle. <https://mathworld.wolfram.com/>. Accessed: 2022-07-06
- Wu J, Leung K, Leung G (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395:689–697
- Zhou X, Lin H (2008) *Moran's I* 725–725. Springer, Boston, MA
- Zhu X, Pan R, Li G, Liu Y, Wang H (2017) Network vector autoregression. *Ann Stat* 45:1096–1123

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.