OPEN

Risk evaluation at municipality level of a COVID-19 outbreak incorporating relevant geographic data: the study case of Galicia

Alejandro Carballosa¹, José Balsa-Barreiro^{2,3}, Adrián Garea¹, David García-Selfa^{1,4}, Ángel Miramontes² & Alberto P. Muñuzuri¹✉

The COVID-19 pandemic was an inevitable outcome of a globalized world in which a highly infective disease is able to reach every country in a matter of weeks. While lockdowns and strong mobility restrictions have proven to be efficient to contain the exponential transmission of the virus, its pervasiveness has made it impossible for economies to maintain this kind of measures in time. Understanding precisely how the spread of the virus occurs from a territorial perspective is crucial not only to prevent further infections but also to help with policy design regarding human mobility. From the large spatial differences in the behavior of the virus spread we can unveil which areas have been more vulnerable to it and why, and with this information try to assess the risk that each community has to suffer a future outbreak of infection. In this work we have analyzed the geographical distribution of the cumulative incidence during the first wave of the pandemic in the region of Galicia (north western part of Spain), and developed a mathematical approach that assigns a risk factor for each of the different municipalities that compose the region. This risk factor is independent of the actual evolution of the pandemic and incorporates geographic and demographic information. The comparison with empirical information from the first pandemic wave demonstrates the validity of the method. Our results can potentially be used to design appropriate preventive policies that help to contain the virus.

As of Nov 15th, 2020, the SARS-Cov-2 or COVID-19^{1,2} has left in Galicia, an autonomous community in Spain, a total of 43,714 cases, 5783 hospitalized, 661 in ICU and 1084 deaths³. Comparatively, these represent 2.9% of the total cases and 2.6% of the deceased nationwide. The ratio of infected to 100,000 inhabitants is 161.9 and the death ratio is 4 per 100,000 inhabitants. The fatality rate is 2.5 deaths for every 100 confirmed cases and 40.1 deaths for every 100,000 inhabitants. The first of these two ratios is in accordance with the rest of Spain and the second in the order of half (compared to Spain). The incidence of the so-called *first wave*, as of June 30th, showed a total of 10,853 cases and 619 deaths in this region. The fatality rate during this first wave was 5.9 deaths per 100 confirmed cases and 22.3 deaths per 100,000 inhabitants. The seroprevalence study carried out by the Galician government during the first wave estimates that only 1.15% of its population had passed the virus at that time⁴. Most recent studies estimate that this rate would reach 4.5% in December 2020, that means, lower than half than for the whole Spain⁵. Therefore, Galicia was one of the least affected regions in Spain.

By means of differential equations and a few tools of complex networks, we have simulated the geographical spread of the virus among the different municipalities (third-level national administration entities) that make up the region (Galicia). Some authors have already presented their concerns on the long term predictability of epidemiological models when describing large real-world systems and have compared it to the process of weather forecasting, where knowing the precise values of several variables only allows for short-term forecasting in matter of a few days^{6–8}. Accurately modelling all the small details that concerns the spreading of a virus, from biological features to social interaction, is a hugely difficult task that one model alone cannot comprehend and thus model

¹Group of Nonlinear Physics, Faculty of Physics, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain. ²Group of Territorial Analysis, Institute IDEGA, University of Santiago de Compostela, Santiago de Compostela, Spain. ³MIT Media Lab, Massachusetts Institute of Technology, 75 Amherst St, Cambridge, MA 02139, USA. ⁴CESGA (Supercomputing Center of Galicia), Avda. de Vigo s/n, Campus Sur, 15705 Santiago de Compostela, Spain. ✉email: alberto.perez.munuzuri@usc.es

predictions should be treated carefully. One important point to address is that the mobility of people across the country results in the sudden and unpredictable emergence of virus outbreaks and so understanding the mobility patterns of human population has been the focus of many previous studies in order to better understand the virus spreading process^{9–14}. Another remark to consider is that each population is geographically unique, has its own history, economic resources, and demographic circumstances. Unveiling which communities have a potentiality of becoming hotspots of the epidemic and which ones do not is crucial for understanding the large-scale behavior of the virus spread and therefore for policy making regarding any kind of mobility restrictions^{15,16} as well as provisions in terms of medical reinforcements. Based on this, the different municipalities that conform the Galician territory are considered as nodes of human population in the abstraction of our mathematical model. Links connecting these nodes shape the structure of a network and act as the different means of transportation between the localities, each one possessing a real number or “weight” that accounts for the flux of pendulum movements of commuting among the nodes. The larger the population density or the shorter the time distance between nodes, the stronger will be the weight of the link connecting them. Using this approach, we are able to differentiate between the most populated areas that act as hubs for human and economic activity, their surrounding localities that usually serve as dormitory towns and the rural areas, less populated and quiet. Thus, one should think of this set of nodes and links as the skeleton of the model accounting for the spatial distribution of population across the territory, while more layers of complexity are required for a more precise approximation to the real picture.

The use of complex networks for accommodating infective transmission dynamics is an extensive field of research within the complex systems science, being a topic covered in several books and reviews^{17–20}. A usual approach here is to consider each node of the network as a single individual and the edges between them as face-to-face interactions that occur daily and model the infective transmission of the disease as a stochastic process that occurs with a certain probability through the edges of the network. This scheme that traces the mobility of the individuals²¹, allows for simulations of entire populations that scale down to the individual interactions with their closest circles and local neighborhoods, being these interactions usually modelled by adding geographical and demographic data to the nodes and simulating the daily interaction patterns²². With this information at hand, one can effectively simulate the outcome of vaccination schemes targeting specific nodes or group of nodes^{23,24}, or analyze how non-pharmaceutical interventions such as social distancing or mobility restrictions can hinder the spread of the viral disease^{25,26}. Note that these strategies that aim to restrain the diffusion of the epidemic dynamics usually imply altering the connectivity pattern of the network, either by blocking nodes that act as hubs or super spreaders within the network or by cutting down the number of interactions and/or edges between the nodes. A comparison between the effectiveness of these methods and others has been carried out by²⁷.

A different scheme considers nodes as entities where multiple individuals can be located, acting as subgroups or subpopulations of a larger whole. Note that in this way one can attribute certain geographical characteristics to the nodes to naturally include the spatial distribution of a set of localities within a territory. This approach is known as *metapopulation network models*^{10,28,29} and here the infective transmission process occurs locally, where one could think of individuals as particles interacting with each other, while these are allowed to simultaneously travel through the edges of the network according to a set of rules and diffusing the disease through multiple nodes. These network models also allow the possibility of considering a deterministic nature to the infection process instead of the stochastic one, exploiting the features of the classic family of SIR-like models³⁰ with an added spatial dimension¹⁶. The SIR model distributes the population into three exclusive compartments: *Susceptible*, *Infected* and *Recovered*. Individuals from the first two groups are allowed to interact increasing the net amount of infective cases in time. Note, however, that these models assume that the whole population is homogeneously distributed in space, an assumption that can be overcome with the inclusion of the metapopulation network. With this consideration in mind, one can simulate the dynamics of the epidemics in a large heterogeneous population based on a computationally affordable way.

After reviewing the different methods and as we previously advanced, here we opted for the latter approach. It is known that deterministic epidemiological models have proven to be quite insightful in its simplest forms^{31–33}, while more elaborate variants have been successfully implemented in tackling the virulent spread of the current coronavirus disease 2019 (COVID-19) in the cases of the city of Wuhan^{34,35}, China³⁶, Italy³⁷ or Portugal³⁸, among others. Metapopulation network models with deterministic nature have also been applied to the dynamic modelling of COVID-19^{15,16} in order to better understand and assess the nation-wide spread of the virus and lockdown relaxation measures. Motivated by these works, we have adapted the model developed in³⁸ to the Galician network. It consists in a variation of the classic SIR model, with the inclusion of one compartment for Exposed E or yet asymptomatic individuals, and another compartment for Protected P individuals, introducing in the system the presence of non-pharmaceutical interventions. Rather than providing an optimal fit to gathered incidence data on active cases, our goals are closer to those of¹⁶, with the aim pointed towards the large-scale picture and finding which nodes are more likely to suffer an outbreak of the pandemic just by its situation on the network and socioeconomic reality, i.e. the weak spots in the network. In contrast with the cited works, where the adjacency matrix of the network was modelled based on actual commuting census data, our study addresses a suitable approach to reconstruct the locality interactions in the absence of such databases, incorporating further demographic data of interest such as age structure or harbor activity (relevant for Galicia). Given the outcomes of the simulations and a given set of considerations, we are able to associate what we shall name a risk factor to each one of the nodes.

As expected from the observed patterns of the virus spread¹⁶, the capital cities headed the rank while at the same time induced a strong influence on its closest neighboring nodes, especially in the case of the two most populated cities, Vigo and A Coruña. Enhancing the exposure to the virus on nodes with higher economic and human activity known beforehand helped to noticeably increase the risk factor on less populated nodes, while rural aged nodes with lesser human activity resulted on a small-to-none risk of suffering an outbreak. While the

calibration of the model was carried out following³⁸ with data at a regional level, a comparison of the results at node level with official data of cumulative cases showed a high degree of correlation.

An independent mathematical analysis based only on the linear stability of the model dynamics and the given network was also carried out, reassuring the performance of the model. This type of analysis based on the termed linear growth rates or growth factors, has been applied before in order to inspect if particular nodes of a complex network are more unstable than others, giving rise to node differentiation and predicting the emergence of mathematical abstractions such as the Turing instability^{39–41}. Extrapolating the approach to the infection transmissive dynamics, the information underlying in the network sufficed to achieve similar results to those of the detailed simulation based on numerical integration of the model.

Finally, notice that although the Galician territory has been considered as a study case, the developed approach to assign risk factors to each locality could be extrapolated to other regions, work which we intend to do in the future.

Study area. Our study area is centered in the region of Galicia, the northwest corner of Spain. It counts with 2.7 million inhabitants distributed in almost 30,000 km², showing a population density of 91.3 people per square Kilometer⁴². In administrative terms, this region is divided into four provinces (second-level territorial and administration units), three in contact with the coast and one inland. Galicia counts with 313 municipalities (third-level territorial and administration units), which correspond to local government units. The network of nodes in which we work in this paper is extracted from these territorial units counting with 313 nodes in total.

Traditionally, its spatial pattern of population distribution is characterized by high dissemination across the territory⁴³. In fact, this region concentrates almost 50% of all the singular population entities in Spain⁴⁴. This rate is about 10 times larger than its demographic weight at a national level. In the last decades, a decline of this traditional pattern of population distribution has been observed, being evident a tendency to the concentration of population. Currently, most of the population concentrates close to the so-called Atlantic Axis, the area surrounding the AP-9 motorway that links northern Galicia with Portugal. This area is located in the west side of the region, very close to the coastline. The major and most thriving cities in this region, five of the seven most populated ones, are located along this axis. Vigo and A Coruña are the two most important cities with approximately 300,000 inhabitants each. The rest of the major cities have populations of about 100,000, with the exception of Pontevedra and Ferrol, which are less populated. In Fig. 1 the seven major cities are also shown, namely, A Coruña, Ferrol, Lugo, Ourense, Pontevedra, Santiago and Vigo. From an urban perspective, this region is characterized by the lack of intermediate cities (between 30,000 and 100,000 people). In the last decades, some municipalities surrounding the metropolitan area of the most important cities (such as Narón, Arteixo, Ames, etc.) experienced the emergence of a growing group of intermediate cities⁴⁵.

The urban pattern of this region shows a hierarchy dominated by the western sector, which presents in the surroundings of the Atlantic Axis, positive demographic dynamics both from a quantitative (higher fertility rates, positive migratory balances) and qualitative perspective (less aged population and more dynamic demographic structures). Just the opposite occurs in the eastern sector where the two main urban references are the two province's capitals, which concentrate a large part of the socioeconomic activity of their provinces. The rest of the territory, which is more rural in nature, is articulated based on a group of head towns that dominate large spaces, whose influence goes beyond their administrative units.

Another important socio-economic factor in Galicia is its close relation with the sea. In fact, along its more than 1700 km of coastline, it has 128 ports. 122 ports are the responsibility of the regional government and 6 are managed by the Spanish government. The sea related activities in these ports is multiple and involves aspects ranging from purely fishing activities (deep-sea fishing activities, inshore, shellfish, commercial activities with loading and unloading of all kinds of merchandise), passengers (regular transport and cruises), recreational and tourist nautical activities. This generates a series of social and economic consequences with great significance for Galicia and, therefore, with important consequences within this pandemic context as we will explain below^{46,47}.

It is also noteworthy the age distribution in the region, in fact the life expectancy is one of the largest in the world (83.3 years in 2020) and the total population above retiring age is significantly large (25.5% in 2020). This is the second more aged region, with an *aging index* of 207.3% (rate estimated based on the number of people older than 64 years related to 100 people younger than 16 years), a value much higher than the national average (129.2%)⁴². Another important related factor is linked to the traditional structure of family units where most of the elders remain in their homes or in close-relatives homes. In fact, the rate of beds in nursing homes per 100 elderly people is significantly lower than in the rest of Spain⁴⁸. Thus, in order to keep simpler our model, we discard the effect of nursing homes in the evolution of the pandemic and instead increase the protection of aged municipalities.

Results

In this section, we present the territorial impact of the COVID-19 in the Galician region, a plausible model that simulated the infection process and two equivalent methods to estimate a local outbreak risk factor. In order to fit the parameters in our model we considered the cumulative incidence data officially gathered for the Galician region during the so-called first wave of the pandemic, this was, from March 1st to around June 30th, 2020. Note that this period of time corresponds with the first pandemic wave in Spain, the one that was less influenced by governmental interventions.

Territorial impact of COVID-19 disease. We considered data in this region from the start of the epidemic until June 30th, 2020. The Galician Health Service (SERGAS), the competent authority in health management in the region, provided this data³. In short, the database counts with 10,853 cases with their corresponding

Figure 1. Location and characterization of the Galician region. In colored background, distribution in provinces (second-level administrations) and municipalities (third-level administrations). The most populated cities and industrial hubs of each province appear highlighted in bold. Red lines represent the major road infrastructures, i.e., motorways in red and railways in yellow. Maps are generated using ArcGIS Desktop software version 10.8 (from ESRI, <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview>).

locations. The location of these cases allows us to evaluate the spatial distribution of the disease and to understand its spatial dynamic. Data are aggregated in municipalities and these are displayed in choropleth maps.

On the one hand, the geographical spread of the virus was clearly more significant in the western sector of the Atlantic Axis, especially in the areas close to the major cities. Figure 2a relates the number of reported cases with the total population. Reported cases are represented by proportional circles, whereas the total population is distributed in five defined intervals, ranging from purely rural to urban municipalities. A priori, it is observed a strong spatial correlation between the total population and the number of reported cases in each municipality. A more exhaustive analysis from a geographical perspective is shown in Fig. 2b, where we represent the incidence rate of the disease, i.e., the number of cases every 100,000 people. Here, we observe how the peri-urban areas surrounding the major cities acquire more prominence. It is explained because in these areas is where most of the people live (residential areas), but also where the industrial activities are mostly concentrated. Therefore, these spaces present a higher dynamism based on a higher number of social interactions.

Parameter estimation and model calibration. The mathematical equations describing the propagation of the epidemic are described in detail in the Methods section and in ref.³⁸. This model considers five distinct groups of population and the interaction among them and has been previously used to successfully describe the evolution of the pandemic elsewhere. Several parameters describe the interaction between the population categories and characterize the transitions from one category into another, being the most relevant the infective transmissive rate β , the fraction of susceptible individuals that enter and exit the protected compartment, p and m , and the parameter d that modulates the coupling between nodes of the network. This model considers several layers of information relevant to the study case (see “Methods” section for details), in particular, information of the net flow of passengers coming through the airports was included. Another layer includes an additional term in the infective transmissive rate, β , due to the port activity that it is so important in the coastal area of Galicia. The last layer of information includes an indication of the economic activity (measured via the average age of each location) as more economically active nodes are more prone to interact with others and, thus, have a negative impact in the spread of the disease.

Figure 2. Geographical spread of COVID-19 in Galicia at a municipal level: (a) Reported cases versus total population, and (b) Incidence rate. Maps are generated using ArcGIS Desktop software version 10.8 (from ESRI, <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview>).

The parameter estimation is extracted from the mathematical model by superposition of the infected and recovered/removed compartment at each time t of the simulation, establishing a straight-forward link between data and model and allowing us to fit the differential equations by means of a non-linear least squares minimization protocol³⁸. Since the available data is at regional level, the fitting process was performed considering all mentioned layers of complexity (see “Methods” section for details) and the sum of the compartments across all nodes in the network. I.e., the model was fitted to the big picture and not at the local level (single node level). Detailed data at the local level is usually hard to find and, in most of the cases, it is endowed with a larger error. Another important remark about the experimental data is that the imposition of a lockdown and subsequent restriction measures resulted in an irregular dataset that can only be well parametrized by time-variations of the transmissive infection rate, as already was pointed out by refs.^{15,16}. In the case at hand, the strong measures imposed by the national government set down a clear turning point on which the daily incidence of the virus started to decelerate and to slowly decrease in time. With this in mind, we have considered two different time regions with its own set of model parameters, separated by the time when the lockdown started. Specifically, during the fitting process we left free three of the model parameters, the infective transmission rate β , the network coupling d and the fraction of people that leave protective measures m . The time duration of the second region goes until the isolation measures were strongly relaxed and a state of “new normality” emerged during the summer 2020. The estimated curve of cumulative incidence is shown on Fig. 3 for both temporal situations, and the parameter values are presented on Table 1. The initially infected individuals were distributed along the regional capital cities for simplicity (30 initially infected individuals distributed along the major cities proportionally to their local population).

Figure 4 presents the average behavior over 1000 realizations of the same simulation with different randomly-chosen initial conditions. The red continuous line represents the mean value of the estimated cumulative infected for the whole region, while the red shadow accounts for the standard deviation over all simulations.

Estimation of the risk factor via numerical simulations. As explained in the introduction and in the Methods section, we performed multiple repetitions of the numerical simulation of the model with different randomly-chosen initial conditions (several initially infected cases randomly scattered along the different nodes in the network) in order to estimate some risk factor per node (see “Methods” section). After each simulation, we marked the locations (nodes in our network) with the larger infection peaks. Those nodes that appeared infected frequently during all the simulations were assigned with a higher value of the infection risk. On the contrary, nodes that almost never got infected during our simulations were assigned an almost null risk of infection. We

Figure 3. Estimated cumulative incidence compared to the real data. The fitting process of the model was separated on the two shown time-regions, one before the start of the lockdown and one after. The initial condition for the simulation was extracted from the first measured value of the dataset, and the number of cumulative cases distributed proportionally among the four capital cities of the provinces and the capital of the region (Santiago de Compostela).

Parameters	1st period (Mar. 1st–Mar. 16th, 2020)	2nd period (Mar. 16th–Jun. 30th, 2020)	References
β	1.0692	0.3735	Estimated
m	0.2882	0.01	Estimated
d	$1e-6$	$1e-6$	Estimated
p	Age dependent, see “Methods” section		
ϕ	$1/15d^{-1}$		Time until lockdown
w	$1/75d^{-1}$		Time until lifting the lockdown
μ	$1/30d^{-1}$		38
ν	0.15		38

Table 1. Summary of the parameter values used in the simulations presented along the text.

Figure 4. SAIRP network model with different random realizations of the initial state. The red line represents the mean value of the cumulative incidence over the whole number of carried simulations, while the red shadow represents the standard deviation of the sample.

have termed this value as the risk factor of the node, and we shall refer to it in this way from here on. The outcome of the risk factor estimation reproduces well the cumulative incidence of the Galician municipalities up to June 30th, as seen on Fig. 5a. Panel 5b represents the empirical number of cases per municipality (Fig. 2) on a scale from 0 to 1, normalized over the highest number of reported cases so that it can be directly compared with the risk factor used in the epidemiological model. A visual inspection of both panels demonstrates the qualitative validity of this technique.

Estimation of the risk factor via linear stability analysis. With the parameters presented on Table 1 and considering all layers of complexity, we computed the growth factors for each node of the network (see “Methods” section) resulting in the values shown in Fig. 5c. The stability analysis is done around the steady state that it is characterized by all individuals in the susceptible compartment, this is, $(\bar{S}_i, \bar{A}_i, \bar{I}_i, \bar{P}_i) = (N, 0, 0, 0)$. This

Figure 5. Spatial distribution of the risk factor (a), normalized cumulative incidence data (b), and growth factors (c) for each municipality. A value of 1 on the risk factor means that the node was found always among the top infected ones on each simulation. Following the same reasoning, a value of 0 means that that node was never found among the top ones. The ranking of the infective nodes was set with a total size of 30, and the cumulative incidence data was normalized over the highest value of reported cases. Maps are generated using ArcGIS Desktop software version 10.8 (from ESRI, <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview>).

Figure 6. Comparative of three different normalized indicators for the municipality nodes. The figure plots the normalized cumulative incidence data (blue), risk factors (red) and growth factors (green) versus the municipality population in semi-logarithmic scale.

is an unstable steady state (any perturbation will trigger the epidemic). In terms of stability analysis, this implies that for the parameters considered the epidemic will outbreak given any perturbation to the initial condition, and, thus, we expect the perturbations to grow on almost all nodes with positive growth factors. A normalization with the maximum value was done for the sake of comparison with the risk factor shown in the previous section. Essentially, the stability analysis outcome is qualitatively and quantitatively equivalent to the one calculated via numerical simulations, suggesting that hubs of human activity and nodes within the *Atlantic Axis* are the most likely to suffer hotspots during an epidemic. Note that the values depicted were obtained without solving the equations but just analyzing the structure of the network combined with the type of dynamics considered per node. Quantitative differences between the growth factors and the risk factors arise due to the risk factor being computed by means of discrete summing (how many times each node appeared on the simulation rankings) while the growth factor ranges in a continue set of values.

The distribution of the quantities analyzed against node's population is represented in Fig. 6, as a summary of the results presented in this section. Note that the risk factors calculated via growth factors are obtained considering only structural information of the system, i.e., dynamical equations and the different layers of structural information that were incorporated (mobility, age distribution, etc.).

Discussion and interpretation

We have modelled the geographical spread of the COVID-19 across the Galician territory during the first wave of the epidemic. With multiple random realizations of the initial seed of infected individuals, we have been able to estimate the risk factor characterizing each municipality (node in our extended network). Large values indicate a high risk to suffer an important outbreak of the virus. This risk was compared with the actual reported cases for each municipality during that time and the Spearman's rank correlation coefficient was computed⁴⁹. This coefficient gives us a measure of how strong the monotonic relationship between any two variables at hand is, making it ideal to compare the visible trends of our data. A positive correlation value of 0.7195 suggests that although not perfect, the model is able to capture at great extent the detailed risk for each node. An independent mathematical analysis based on the linear stability of the network dynamics was also carried out validating the previous results. The correlation value of this analysis compared to simulations is 0.9567, while the correlation with the reported data cases returned a value of 0.7421. Note that although these coefficients are merely orientative, they inform us if shared tendencies are present among the different outcomes. Keep in mind that we are actually comparing real cumulative incidence data (also endowed with some intrinsic error due to human manipulation) with the estimated risk based on our approach, not the cumulative incidence estimated with the model. Our risk assessment intends to give an orientative prediction on how likely, given similar circumstances but different initial conditions, the reported incidence would have been the same. Even more, this might become a tool for understanding the risk of a particular node and for evaluating different strategies in order to reduce it. For the most populated nodes ($\sim 10^5$ inhabitants), we found that the risk is high in most of all possible scenarios, while for medium populated nodes things could have been completely different in more than half, either by suffering stronger outbreaks or by not experiencing the virus at all. For a majority of nodes, mostly rural and low populated (Fig. 5), the risk of an outbreak was null or practically null and this can be easily contrasted with the reported cases. It is important to note that the risks factors were evaluated through intensive simulation running multitude of experiments while the growth factors were calculated once directly from the mathematical model. The large correlation between these two factors indicates their equivalency. Thus, instead of running costly simulations, one straightforward calculation can provide the same information and help evaluate the nodes that are more likely to suffer an outburst.

Concerning the two different ways to compute the risk factor, we observe that both cases reproduce the reality up to a good extent. The main discrepancy comes from the discrete number of realizations used to compute the risk factors. The growth factors (stability analysis) grow monotonically with the population while the risk factors (direct numerical simulation) concentrate mainly in the three clusters of values observed on Fig. 6, ~ 0 , ~ 0.3 and ~ 1 . It is important to remind that both methods designed to calculate the risk factors include several layers of information in order to improve the accuracy of the method. During the first pandemic wave, a complete national lockdown prevented that the Galician region would suffer from strong incidence (in comparison to the rest of

Model layers	Spearman		Pearson r^2		R^2	
	1st wave	2nd wave	1st wave	2nd wave	1st wave	2nd wave
Mobility	0.8572	0.9463	0.1283	0.1674	– 53.13	– 51.34
Ages	0.7463	0.8185	0.2332	0.3006	– 2.74	– 2.47
Multilayer	0.7421	0.8091	0.2397	0.3118	– 2.16	– 1.94

Table 2. Relevance of the different layers of information in the model for the first and second pandemic waves analyzed using different correlation parameters.

Figure 7. Effect of the different layers of information on the calculated growth factors. The blue crosses show the normalized cumulative incidence data of the second wave. Red dots show the growth factors calculated considering only the mobility information layer in the model. Green dots also include of the age-related economic activity and, finally, the black dots show the growth factors considering all the information layers (including port activity) described in the Method Section.

Spain or even Europe) and halted drastically the spread of the virus from the capital cities to the lower populated municipalities (see Figs. 5c and 6 for a visual inspection). This was not the case for the subsequent pandemic waves in which looser restrictions were imposed and the economy was reactivated, which led to higher rates of daily incidence⁵⁰. Table 2 presents several correlation factors between the growth factors and the experimental cumulative incidence calculated for the first and second waves and considering independently the different layers of information. The first piece of information that we observe is that all the correlation coefficients are improved in the second wave meaning that our model satisfactorily predicted the risk of the different municipalities.

In order to understand the importance of the different layers of information we need to consider other correlation coefficients besides the Spearman's already discussed. We have introduced the Pearson correlation coefficient for linear regression with zero intercept (values ranging from 0 to 1) and the coefficient of determination R^2 (with values from minus infinite to 1)⁴⁹. While the Spearman returns the highest correlation when considering the model with just the mobility network, this value is biased by the fact that the Growth Factors returned for this case a very similar not-null value for almost all low-populated municipalities (see Fig. 7). Thus, the monotonous relationship between risk factors and data is almost identical and the Spearman correlation is almost one, but what the risk factors would be telling us is that low populated municipalities would have an important risk of suffering a new outbreak, which both the empirical data and the other two metrics prove wrong. On the other hand, the Pearson correlation and the R^2 test are both based on the quantitative values of the data series and measure how much do the data series fit to a straight line. Note that the intercept of the linear fit must be fixed to zero in order to ensure that those municipalities with almost null cases also should have a very low risk factor. Otherwise, the Pearson coefficient in the mobility model would also return a misleading correlation value.

From Table 2, we observe that increasing the layers of information results in improved correlation with the empirical data for these two metrics. Note that the low values for the Pearson coefficient and the negative ones for the R^2 are just telling us that the relation between these two series is not linear, which we already expected. We should focus mainly on the relative differences among the outcomes. When only the mobility layer is considered, the correlations are significantly worse and including the economic activity via the age distribution results in an important improvement of the model. Finally, the rest of the information considering port activity slightly increases the correlations. Although this small difference could be telling us that this last layer of information is an unnecessary refinement, we are firm with its importance to accurately measure the risk of the Galician portuary municipalities, at least within the framework of the methodology presented here. For different approaches in which a specific infective transmission parameter β_i could be obtained directly for each municipality, e. g., fitting each subregion to its own data series, this layer of information could be discarded. The incoming of daily ships and transports to these cities and villages suppose a higher intrinsic risk of external contagions in comparison to central ones, and thus it is difficult to incorporate this information otherwise in our approach.

Focusing on territorial aspects, it is crucial not only for understanding the geographical patterns of the virus spread but also for policy making regarding the mobility restrictions of the population. We did not intend here

to postulate future scenarios of lockdown neither relaxation measures with the different possible outcomes, for that we refer the reader to a study such as¹⁵. Our main goal was to show the difference that geographical placement and demographic circumstances has on the impact of the virus spread, a future work concerning a larger portion of territory is intended.

Comparison of both maps suggests that the mathematical model was able to accurately capture which areas were the most affected. Highly populated hubs of socioeconomic activity contain most of the nodes with higher risk of heavy infection, increasing nearest less-populated localities risk to suffer an outbreak. The so-called *Atlantic Axis*, which refers to the western territorial area connecting the two main hubs, contains almost all the highlighted nodes on the map. It is clear how the infections are primarily focused on the main cities, but also in their suburban areas. Thus, we can see how the nodes surrounding the major cities show levels of infection extreme in relative terms (Fig. 2b). Our model also correctly estimates the higher risk levels in many intermediate nodes, such in the case of the so-called *head towns*, especially those located in both western provinces. This higher infection risk is due to the increase in the number of social interactions in these towns, which concentrate a relevant socioeconomic activity such as trading centers for large areas based on primary sector. In addition, the model differentiates some areas with singular local dynamics, such as the region termed *Costa da Morte*, situated on the far north-western coast and which presents a clear regressive demographic dynamic despite being close to the most dynamic axis in this region⁵¹. Therefore, the implemented model is able to capture the different local dynamics present in the system.

Several aspects must be considered from a territorial approach. First, remark again that our actual data was limited to the first wave and that the analysis of the second and successive waves, ranging from July 2020 to the present date of submission is also left for future work (although some hints were already included in order to prove the robustness of the method presented). Note that this first wave of the pandemic is also the less altered by political decisions. The complexity of the virus spread may be understood when it occurs in a hierarchical way, with a concentration of cases in the most populated cities during the first weeks. It is expected that the spreading process affects lower levels of the network hierarchically (less populated nodes) over time. In addition, it should be evaluated if areas where in theory people live more concentrated (cities, although it will depend on the building density) favor a major spread of the virus more than nodes with lower risk (villages), but where community life is closer with more strong family-based relationship structures.

A second fundamental factor is the unequal territorial impact of the epidemic in terms of fatalities. This is due to some regional characteristics that act as determinants⁵². It explains how the spatial distribution of fatalities may be different to the infection maps. Initially, the virus had circulated more in the western nodes, close to the major cities, due to a larger concentration of the population and the higher socioeconomic dynamism (more interactions) in all these areas. However, the territorial impact of the pandemic in terms of fatalities is probably more important in some eastern sectors where much of these municipalities present a weaker demographic structure in qualitative terms (more aged populations).

Along this paper we present a tool that we used to understand the different factors affecting the risk a given node presents to be infected. Here, we just tried to understand the first wave of the COVID-19 pandemic in Galicia (Spain) although we also present evidences of the accuracy of the method to explain the second wave. The developed tools can be potentially used as a workbench to test different policies and their implications in the risk factors at a node level in order to prepare for future COVID-19 waves and new pandemics.

Methods

Galician spatial network. Without entering in verbose definitions, a complex network⁵³ is a mathematical object represented by a graph, this is, a set of elements or *nodes* connected by links (also called *edges*). In the present case, the nodes are the different municipalities in the territory considered and the links establish the type of interactions between nodes. The interaction structure of the nodes within our particular complex network is given by the weighted adjacency matrix M , which we have constructed as a function of the council populations and the squared time distance between any two of them,

$$M_{ij} = \frac{N_i N_j}{t_{ij}^2} = w_{ij} \quad (1)$$

being N_i and N_j the population of any two locations i and j , t_{ij} the travel-time distance in car between them (that takes into account the real conditions of road infrastructures between nodes) and w_{ij} is the so-called weight of the edge connecting the nodes. With this kind of undirected symmetric adjacency matrix, we are able to build a heterogeneous network where each node has a unique set of weights or connectivity pattern. This will cause that each node experiences the dynamical process in a different way. The weights of the edges aim to model the flux of commuters from one node to another. In the absence of actual commuting census data, we considered more likely the exchange of travelers between heavily populated cities (the capital cities of the territories) and surrounding areas. According to these weights, we will speak about the node strength as the sum of all the weights of the connections of that node:

$$s_i = \sum_j^n M_{ij} = \sum_j^n w_{ij} \quad (2)$$

We have considered the time distance instead of the geometrical distance because of the Galician orography as well as the road infrastructures, that eventually might result in significantly larger travel times between neighboring locations than the expected from their relative physical distance⁵⁴. In the case of the Atlantic coast, where

the coast is segmented due to abundant estuaries, travel-times from one city to the other in opposed locations of the estuary takes longer by car than it would take on a boat (the straightforward direction). Of course, we have considered the time distance by car since it is a more likely option for travelling than the boat one. These time distances were estimated using the *Open Street Map API*⁵⁵, and all geographical and demographical information about each node is extracted from the Galician Institute of Statistics⁵⁶.

Epidemiological model. To describe the infection and transmission process of the COVID-19 pandemic, many different models have been considered and yet there is no clear consensus on how many compartments should the model contain. Increasing or reducing the number of compartments is often determined by the final goal of the research. With our aim at hand, we model the epidemic spread across the subpopulations with the SAIRP deterministic model, a variation of the classic SEIR model that while maintaining certain simplicity was successfully applied in the case study of Portugal³⁸. This model is composed of five differential equations that categorizes the human populations into five exclusive compartments, each governed by a set of transition rules. Basically, susceptible individuals S can become infected by interaction with either asymptomatic A or symptomatic infected I individuals. This process occurs with a transmission rate β , for interactions with individuals of both categories. The asymptomatic ones become symptomatic infected with a rate ν , while symptomatic become removed or recovered with rate μ . The distinctive feature of this model compared to a typical SEIR is the category of protected individuals P . A fraction p of susceptible individuals becomes protected with a transition rate ϕ , but at each time step a fraction m of the protected individuals becomes susceptible again with a transition rate w . We believe that this factor offers a suitable approach for modelling the increasing measures of non-pharmaceutical interventions used to control the propagation of the virus observed in the current times, such as social distancing, protective facial masks or mobility restrictions. The detailed equations are shown below.

Coupling with the complex network. The above-described epidemiological model is now considered for each of the nodes in our complex network. Thus, the population of each municipality (node) is subdivided in all five compartments, with all individuals belonging initially to the susceptible compartment. Then, an initial seed of infected individuals is introduced randomly in the network and the disease propagates from there both locally and globally through the nodes of the network via the connections. Mathematically, this means that for each one of the nodes, there are a set of five nonlinear dynamical equations. The equations in all the nodes are coupled via an additional term that accounts for the network interaction. The network term only contributes positively to the transmission of the disease and represents both the interaction between susceptible residents with commuting infected individuals and vice versa. The larger the weight of the edge, the larger the amount of these interactions and the stronger the influence that highly infected nodes exert on their connections. A parameter d was included in order to modulate the impact of these weights on the spreading dynamics. The model equations for any node i are presented in (3).

$$\begin{cases} \dot{S}_i = -\beta_i(1-p_i)\left(\frac{A_i}{N_i} + \frac{I_i}{N_i}\right)S_i - \phi p_i S_i + w m P_i - d\beta S_i \sum_{j=1}^n M_{ij}\left(\frac{A_j+I_j}{N_j}\right) \\ \dot{A}_i = \beta_i(1-p_i)\left(\frac{A_i}{N_i} + \frac{I_i}{N_i}\right)S_i - \nu A_i + d\beta S_i \sum_{j=1}^n M_{ij}\left(\frac{A_j+I_j}{N_j}\right) \\ \dot{I}_i = \nu A_i - \mu I_i \\ \dot{R}_i = \mu I_i \\ \dot{P}_i = \phi p_i S_i - w m P_i \end{cases} \quad (3)$$

The meaning of the parameters is described in the previous subsection. So far, the only difference between nodes lies in their total population N_i and their connectivity pattern given by the adjacency matrix M . Mobility is important but we also notice that nodes with the same number of individuals but different social and economic activities should behave differently. Just to name an example in the study case in Galicia, our model has to be able to differentiate a village on the countryside from one located on the shore with a harbor where there is usually more economic activity. In the following, we describe the different layers of information that we incorporated in the above model and that introduce the modulations on the β and p parameters for each node (hence, the subindex i in the reaction terms of the equations). For the network term coupling the municipalities, we leave the global infectivity β from the fitting, since using specific transmission rates among municipalities was not possible with the available data at the time.

Layers of information incorporated into the epidemic model. Apart from the mobility data already incorporated into the model via the adjacency matrix, there are other layers of information that are important to consider as they may play a major role in the evolution of a pandemic.

Three airports are located in the studied region of Galicia in the most populated cities of A Coruña, Vigo and Santiago de Compostela. The net flux of passengers is obtained monthly from the national air-traffic statistics⁵⁷. For simplicity, the flux was considered homogeneous in time and the flux of incoming travelers was considered to be half of each month's total. This number of travelers is incorporated into the local node population with a distribution within each category as expected from the daily distribution of infected at the country level.

The region studied is mostly surrounded by the ocean and its economy strongly depends on ocean-related activities, thus, human interactions in the coast locations are also important. In order to incorporate this effect in the model by considering an effective infective transmission rate, β , that it is modulated following,

Figure 8. Different levels of complexity on the network model. The left panels show the normalized values of the infection peak for each one of the nodes, distributed spatially. The right panels show the time evolution of the infected individuals for one specific node, Burela, situated on the northeastern coast. To start the simulation, an initial infected seed was placed into each node.

$$\beta_i = \beta + \beta_{port,i} \quad (4)$$

The additional term included, $\beta_{port,i}$, considers that the coastal cities are more prone to get infected due to interaction with the different ships arriving at the port. Here, it is important to distinguish the different types of economic activities in these coastal locations. Some of them have an intense economic activity while others are used mostly for leisure. In order to incorporate it into the model we considered the average age of each location, y_i , (extracted from demographic data) as an indicator of the level of economic activity in the node. Locations with a high value of the average age tend to be inhabited by retired senior citizens that use the port facilities mostly for leisure while younger populations are more involved in economic activities like fishing, trading, etc., that imply a higher risk of infection. All this is incorporated in the term,

$$\beta_{port,i}(y_i) = \begin{cases} \beta, & y_i < 18 \\ \beta - \beta \frac{(y_i-18)}{47}, & 18 > y_i > 65 \\ 0, & y_i > 65 \end{cases} \quad (5)$$

Note that we considered, as a first approximation that the infection rate linearly decreases with the average age of the location, with a saturation at the official retirement age (65 years). Of course, in the rest of localities the infective transmission rate remains $\beta_i = \beta$, this is, the one obtained from the regional fitting.

Next, we incorporated into the model information of the socioeconomic activity of each node, and we considered that those nodes that are more active have a higher rate of social mixing and interaction among individuals, situation that fosters the propagation of the virus. We can easily incorporate this rate of activity in the model by making the fraction of protected individuals p a function dependent of the mean age y_i at each node i ,

$$p_i(y_i) = \begin{cases} \frac{1}{65}y_i, & y_i < 65 \\ 1, & y_i \geq 65 \end{cases} \quad (6)$$

Note that retired population (above 65 years old) usually takes extreme protective measurements as they are the more vulnerable and this is reflected in the model equation.

The different layers of information considered here actually contribute unevenly to the infection curves. This can be observed in Fig. 8 where the local time evolution of the infected compartment is represented for a particular coastal node with the increasing layers of complexity from top to bottom. Left panels display the spatial distribution of the epidemic incidence. When the age or economic dynamicity is incorporated (central panel), the effect becomes more pronounced. Once the port activity is incorporated, the description of the epidemic in the coastal areas becomes more complete. Observing the panel on the bottom left, it can be grasped the emergence of two main hubs of infection, both on the extremes of the western coast and mainly influenced by the two most populated cities in the region, Vigo (south) and A Coruña (north). Since these cities conform the main hubs of economic and industrial activity it could be expected that their urban and suburban areas would suffer the worst outcome of the pandemic, and this is mainly what the network model captures. Eastern nodes

that are also highly affected by the pandemic (shown in brighter red on the map) belong to the capital cities of those provinces, and to high-activity harbor nodes on the north-eastern coast.

Regarding the simulations, we devised the following scheme to obtain a statistical representation of those nodes that present the higher risk of an outbreak according to their particular conditions. First, with all the individuals placed on the susceptible compartment, we initiate the simulation with a seed of n_I infected individuals distributed randomly (with uniform probability) among all the nodes. Then, we integrate Eq. (7) in time with a fourth order Runge–Kutta algorithm, until the dynamics of the infection have died out and all the individuals that were infected are placed on the removed compartment. As a measure of the impact of the disease on every node, we use the cumulative sum of symptomatic infected individuals. This quantity shows the impact of the infective transmission inside each node, giving us a tool to compare the evolution of the epidemic within the network. Finally, we repeat several times the simulation with different randomly-chosen initial seeds until we obtain enough statistical significance.

To assess the risk of each node to develop an outbreak, we proceed as follows. For each simulation, we sort the nodes as a function of this cumulative sum or total number of symptomatic infected individuals (from higher to lower) and keep for example, the first fifty nodes of each simulation (nodes with the larger number of symptomatic infected individuals). Following this example, if the number of simulations were to be 1000, we should have made a ranking of the nodes for each simulation and elaborate a table of 1000 rows and 50 columns. From this table, we count how many times each node appeared among the whole set of rankings, and we sort them again in terms of this counting. Those nodes with a higher value count will be those assigned with the highest risk. More precisely, we use this count as a measure or the *risk factor* by normalizing them to the number of carried simulations. If the risk factor equals 1, it means that the node appeared among the top infected nodes in all simulations, while a risk factor of 0 means that the node was never among them. Note that in this estimator the actual information on the size of the infection is discarded, since our aim is to assess if the municipality has a risk of being a hotspot rather than giving an accurate and quantitative prediction of the peak.

Growth factors and linear stability analysis. In this section, we explain the linear stability approach to estimate some risk factors without numerically simulating the set of Eq. (3). First, we perform a simplification to the set of Eq. (3) considering that the total population of any node N_i is constant over time, this is, $S_i(t) + A_i(t) + I_i(t) + R_i(t) + P_i(t) = N_i$ at any time t . Thus, the number of recovered individuals can be computed at any time $t > 0$ as the SAIP simplified model is, then, described by the following set of equations:

$$\begin{aligned} R(t) &= R(0) + \delta \int_0^t I(u) du \\ \begin{cases} \dot{S}_i = f(S_i, A_i, I_i, P_i) - d\beta S_i \sum_{j=1}^n M_{ij} \left(\frac{A_j + I_j}{N_j} \right) \\ \dot{A}_i = g(S_i, A_i, I_i, P_i) + d\beta S_i \sum_{j=1}^n M_{ij} \left(\frac{A_j + I_j}{N_j} \right) \\ \dot{I}_i = h(S_i, A_i, I_i, P_i) \\ \dot{P}_i = z(S_i, A_i, I_i, P_i) \end{cases} \end{aligned} \quad (7)$$

where f, g, h and z denote the nonlinear functions in Eq. (3). Let now us consider any of the possible steady states of the system and name it $(\bar{S}_i, \bar{A}_i, \bar{I}_i, \bar{P}_i)$. If we introduce in each node a small perturbation to the node's state

$$(S_i, A_i, I_i, P_i) = (\bar{S}_i, \bar{A}_i, \bar{I}_i, \bar{P}_i) + (\delta S_i, \delta A_i, \delta I_i, \delta P_i)$$

we can write the equations for the perturbations as,

$$\begin{cases} \dot{\delta S}_i = \left(f_S - d\beta \sum_{j=1}^n M_{ij} \left(\frac{\bar{A}_j + \bar{I}_j}{N_j} \right) \right) \delta S_i + f_A \delta A_i + f_I \delta I_i + f_P \delta P_i - d\beta \bar{S}_i \sum_{j=1}^n M_{ij} \left(\frac{\delta A_j + \delta I_j}{N_j} \right) \\ \dot{\delta A}_i = \left(g_S + d\beta \sum_{j=1}^n M_{ij} \left(\frac{\bar{A}_j + \bar{I}_j}{N_j} \right) \right) \delta S_i + g_A \delta A_i + g_I \delta I_i + g_P \delta P_i + d\beta \bar{S}_i \sum_{j=1}^n M_{ij} \left(\frac{\delta A_j + \delta I_j}{N_j} \right) \\ \dot{\delta I}_i = h_A \delta A_i + h_I \delta I_i \\ \dot{\delta P}_i = z_S \delta S_i + z_P \delta P_i \end{cases} \quad (8)$$

where subscripting represents the partial derivative over the variable (e.g., $f_S = \frac{\partial f}{\partial S}$). On the other hand, the perturbation of each one of the variables can be expanded over the set of eigenvectors of the Laplacian matrix of the network³⁹, in close analogy to the classical case of continuous media where non-uniform perturbations can be decomposed into a set of spatial Fourier modes that represent plane waves with different wavenumbers⁵⁸. In our case, since the Laplacian matrix is closely related to the adjacency matrix by the relation $L_{ij} = M_{ij} - \delta_{ij}k_i$, with δ_{ij} being the Kronecker delta and k_i the connectivity of node i , we follow the same approach as in³⁹ but with the adjacency matrix without any loss of generality. The eigenvalues Λ_α and eigenvectors $\Phi^\alpha = \{\phi_1^\alpha, \dots, \phi_n^\alpha\}$ of the adjacency matrix M_{ij} for each mode α are given by $\sum_{j=1}^n M_{ij} \phi_j^\alpha = \Lambda_\alpha \phi_i^\alpha$, with $\alpha = 1, \dots, N$. Expanding the perturbations δX_i over the set of adjacency eigenvectors

$$\delta X_i = \sum_{j=1}^n c_{\alpha} \exp(\lambda_{\alpha} t) \phi_i^{\alpha} \quad (9)$$

the set of Eq. (8) is transformed into a set of n independent equations for the different normal modes, resulting in the eigenvalue problem,

$$\lambda_{\alpha} \begin{pmatrix} 1 \\ B_{\alpha} \\ C_{\alpha} \\ D_{\alpha} \end{pmatrix} = \begin{pmatrix} f_S - d\beta \sum_{j=1}^n M_{ij} \left(\frac{\bar{A}_j + \bar{I}_j}{N_j} \right) & f_A - d\beta \bar{S}_i \Lambda_{\alpha} & f_I - d\beta \bar{S}_i \Lambda_{\alpha} & f_P \\ g_S + d\beta \sum_{j=1}^n M_{ij} \left(\frac{\bar{A}_j + \bar{I}_j}{N_j} \right) & g_A + d\beta \bar{S}_i \Lambda_{\alpha} & g_I + d\beta \bar{S}_i \Lambda_{\alpha} & 0 \\ 0 & h_A & h_I & 0 \\ z_S & 0 & 0 & z_P \end{pmatrix} \begin{pmatrix} 1 \\ B_{\alpha} \\ C_{\alpha} \\ D_{\alpha} \end{pmatrix}$$

The set of solutions λ_{α} for each mode α are called *growth factors* and from their definition on (9) it can be grasped that they give information on how fast any perturbation will drive the system towards or apart from the steady state. If we consider this state to be the one where all the population is in the susceptible compartment, it can be easily perturbed by introducing a small number of infected individuals. If for each node, there is at least one of the λ_{α} with a positive value, this means that the infective dynamics will develop in that node and that an outbreak of the pandemic will take place. On the other hand, a negative value implies that the perturbation will die out and the node would remain in its steady state. Furthermore, the value of the growth factor gives us a notion of how quickly an outbreak will evolve. Note that each node has a different set of values of the growth factor and the dynamics of that particular node will be controlled by the largest value of λ_{α} (in absolute value). Thus, we will record the maximum positive value of λ_{α} for each node and use as a reference of how likely is to become infected.

Received: 9 February 2021; Accepted: 11 October 2021

Published online: 28 October 2021

References

- Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England J. Med.* **382**, 1199–1207 (2020).
- Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *New England J. Med.* **382**, 727–733 (2020).
- <https://www.sergas.es/>, [Online].
- https://www.xunta.gal/hemeroteca/-/nova/104214/estudio-seroprevalencia-frente-covid-19-estima-que-galicia-estaria-inmunizada?langId=es_ES, [Online].
- Ministerio de Sanidad, Consumo y Bienestar Social, ESTUDIO ENE-COVID: CUARTA RONDA. ESTUDIO NACIONAL DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN ESPAÑA: <https://www.msbs.gob.es/gabinetePrensa/notaPrensa/pdf/15.12151220163348113.pdf>, (2020).
- Castro, M., Ares, S., Cuesta, J. A. & Manrubia, S. The turning point and end of an expanding epidemic cannot be precisely forecast. *Proc. Natl. Acad. Sci.* **117**, 26190–26196 (2020).
- Siegenfeld, A. F., Taleb, N. N. & Bar-Yam, Y. Opinion: What models can and cannot tell us about COVID-19. *Proc. Natl. Acad. Sci. USA* **117**, 16092–16095 (2020).
- Holmdahl, I. & Buckee, C. Wrong but useful: What Covid-19 epidemiologic models can and cannot tell Us. *New England J. Med.* **383**, 303–305 (2020).
- González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
- Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *J. Theor. Biol.* **251**, 450–467 (2008).
- Meloni, S. *et al.* Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.* **1**, 8 (2011).
- Balcan, D. *et al.* Modeling the spatial spread of infectious diseases: The GLObal Epidemic and Mobility computational model. *J. Comput. Sci.* **1**, 132–145 (2010).
- Ni, S. & Weng, W. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **79**, 016111 (2009).
- Dalziel, B. D., Pourbohloul, B. & Ellner, S. P. Human mobility patterns predict divergent epidemic dynamics among cities. *Proc. R. Soc. B Biol. Sci.* **280**, 20130763 (2013).
- Bertuzzo, E. *et al.* The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. *Nat. Commun.* **11**, 8 (2020).
- Calvetti, D., Hoover, A. P., Rose, J. & Somersalo, E. Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19. *Front. Phys.* **8**, 6 (2020).
- Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307 (2005).
- Barrat, A., Barthélemy, M. & Vespignani, A. Epidemic spreading in population networks. In *Dynamical Processes on Complex Networks* (eds Barrat, A. *et al.*) 180–215 (Cambridge University Press, 2008).
- Pastor-Satorras, R., Castellano, C., Miegheem, P. V. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- Agarwal, P., Nieto, J. J., Ruzhansky, M. & Torres, D. F. M. Analysis of Infectious Disease Problems (Covid-19) and Their Global Impact, Springer Nature Singapore Pte Ltd. (2020).
- Wang, D., Pedreschi, D., Song, C., Giannotti, F. & Barabási, A.-L. Human mobility, social ties, and link prediction. in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA (2011).
- Perez, L. & Dragicevic, S. An agent-based approach for modeling dynamics of contagious disease spread. *Int. J. Health Geogr.* **8**, 50 (2009).
- Martin, G., Marinescu, M.-C., Singh, D. E. & Carretero, J. Leveraging social networks for understanding the evolution of epidemics. *BMC Syst. Biol.* **5**, 14 (2011).
- Dezső, Z. & Barabási, A.-L. Halting viruses in scale-free networks. *Phys. Rev. E* **65**, 5 (2002).
- Hoertel, N. *et al.* A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nat. Med.* **26**, 1417–1421 (2020).
- Carballosa, A., Mussa-Juane, M. & Muñizuri, A. P. Incorporating social opinion in the evolution of an epidemic spread. *Sci. Rep.* **11**, 1772 (2021).
- Xiong, F. & Li, Z.-Y. Effective methods of restraining diffusion in terms of epidemic dynamics. *Sci. Rep.* **7**, 7 (2017).

28. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci.* **102**, 11157–11162 (2005).
29. Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276–282 (2007).
30. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Char.* **115**, 700–721 (1927).
31. Anderson, R. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, 1991).
32. Brauer, F. & Castillo-Chavez, C. *Mathematical Models in Population Biology and Epidemiology* (Springer, 2012).
33. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2007).
34. Lin, Q. *et al.* A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* **93**, 211–216 (2020).
35. Ndairou, F., Area, I., Nieto, J. J. & Torres, D. F. M. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons Fractals* **135**, 109846 (2020).
36. Ivorra, B., Ferrández, M. R., Vela-Pérez, M. & Ramos, A. M. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun. Nonlinear Sci. Numer. Simul.* **88**, 105303 (2020).
37. Giordano, G. *et al.* Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **26**, 855–860 (2020).
38. Silva, C. J., Cruz, C., Torres, D. F. M., Munuzuri, A. P., Carballosa, A., Area, I., Nieto, J. J., Fonseca-Pinto, R., da Fonseca, R. P., dos Santos, E. S., Abreu, W. & Mira, J. Optimal control of the COVID-19 pandemic: Controlled sanitary deconfinement in Portugal (2020).
39. Nakao, H. & Mikhailov, A. S. Turing patterns in network-organized activator–inhibitor systems. *Nat. Phys.* **6**, 544–550 (2010).
40. Mimar, S., Juane, M. M., Park, J., Muñuzuri, A. P. & Ghoshal, G. Turing patterns mediated by network topology in homogeneous active systems. *Phys. Rev. E* **99**, 062303 (2019).
41. Mussa Juane, M., García-Selfa, D. & Muñuzuri, A. P. Turing instability in nonlinear chemical oscillators coupled via an active medium. *Chaos, Solitons Fractals* **133**, 109603 (2020).
42. INE (2019): Cifras de población, [Online]. Available: <https://www.ine.es/>.
43. Bouhier, A. La Galice : Essai géographique d'analyse et d'interprétation d'un vieux complexe agraire, Yonnaise: La Roche-sur-Yon (Vendée), 1979.
44. N. P. d. P. C. p. u. p. Resultados, <https://www.ine.es/nomen2/index.do>.
45. R. C. a. P. D. Lois González, A Galicia urbana, Xerais de Galicia (2015), p. 755.
46. Xunta de Galicia, "Portos de Galicia," Xunta de Galicia, 2020. [Online]. Available: <https://portosdeg Galicia.gal/gl/web/portos-de-galicia/home>.
47. INE, "Estadística sobre el Transporte Marítimo de Mercancías y Pasajeros," 2020. [Online]. Available: <https://www.ine.es/dyngs/IOE/en/operacion.htm?numinv=22046>.
48. Abellán García, A., Aceituno Nieto, M. D. P. & Ramiro Fariñas, D. Estadísticas sobre residencias: Distribución de centros y plazas residenciales por provincia. Datos de abril de 2019. Informes Envejecimiento en red nº 24, 24 p., Madrid, 2019.
49. Kokoska, D. Z. A. S. *CRC Standard Probability and Statistics Tables and Formulae* (Chapman & Hall, 2000).
50. Güell, O. *Why is Spain Being Hit Again by the Coronavirus Pandemic?*, (2020)
51. Balsa-Barreiro, J. Insostenibilidad de modelos territoriales desde un punto de vista demográfico: A Costa da Morte (Galicia, España). *Papeles de Población* **19**(78), 167–206 (2014).
52. G. A. N. L. M. S. D. C. Mounir Amdaoud, Geography of COVID-19 outbreak and first policy answers in European regions and cities., <https://halshs.archives-ouvertes.fr/halshs-03046489> (2020).
53. Barabási, A.-L. *Network Science* (Cambridge University Press, 2016).
54. Balsa-Barreiro, J., Ambühl, L., Menéndez, M. & Pentland, A. Mapping time-varying accessibility and territorial cohesion with time-distorted maps. *IEEE Access* **7**, 41702–41714 (2019).
55. <https://www.openstreetmap.org>, [Online].
56. <https://www.ige.eu>, [Online].
57. <http://www.aena.es/csee/Satellite?pagename=Estadisticas/Home>, [Online].
58. Turing, A. M. The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **237**, 37–72 (1952).

Acknowledgements

The authors would like to thank the Galician Health Service, dependent on the Xunta de Galicia, for providing the data used to prepare this paper. This project has been financed by the Galician Innovation Agency (GAIN). Reported official cases related to the territorial spreading of COVID-19 in Galicia were provided by the Health Galician Service (SERGAS). This research is supported by Galician Innovation Office (GAIN), the Spanish Ministerio de Economía y Competitividad and European Regional Development Fund, research grant No. COV20/00617 and RTI2018-097063-B-I00 AEI/FEDER, UE; by Xunta de Galicia, Research Grant No 2021-PG036. AC, AG, DGS, APM are part of the CRETUS Strategic Partnership, AGRUP2015/02, supported by Xunta de Galicia. All these programs are co-funded by FEDER (UE). We also acknowledge support from the Portuguese Foundation for Science and Technology (FCT) within the Project n. 147. A substantial portion of the simulations were run at the Centro de Supercomputación de Galicia (Spain) and we acknowledge their support.

Author contributions

All authors contributed to the original idea and design of the manuscript. A.C. performed numerical simulations and data analysis. A.G. and D.G.S. contributed to the growth factors calculation. J.B.B. and A.M. contributed to the analysis of the results and the discussion. A.P.M. coordinated the project. A.C., J.B.B., A.M. and A.P.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.P.M.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021