scientific reports

Check for updates

OPEN

Assessing the impact of interregional mobility on COVID19 spread in Spain using transfer entropy

Miguel Ponce-de-Leon[1]✉, Camila Pontes[1], Alex Arenas[2,3] & Alfonso Valencia[1,4]

Human mobility played a key role in shaping the spatiotemporal dynamics of COVID19 transmission. This study employs Transfer Entropy (TE), an information-theoretic approach, to investigate the directional relationship between interregional mobility and COVID19 spread in Spain. Specifically, we use the mobility-associated risk time series, derived from phone-based origin–destination data and local infection prevalence, to estimate the flow of potentially infected individuals between regions. TE is then applied to measure the information flow from mobility-associated risk to regional case counts, enabling us to uncover spatio-temporal patterns of mobility-driven transmission. Using real-world data, we identified provinces that acted as outbreak drivers during the COVID19 pandemic in Spain and detected temporal shifts in the strength and direction of mobility's influence. Our findings align with key epidemiological events, such as the 2020 summer outbreak in Lleida linked to seasonal workers, and highlight the effects of non-pharmaceutical interventions, including bar closures in Catalunya, on transmission dynamics. Finally, we validated our approach using simulations from a metapopulation SIR model with known transmission pathways, showing that TE can recover mobility-induced transmission structure while reducing indirect or spurious associations. Altogether, our work provides a novel approach to study the effect of interregional mobility on epidemic spread and to uncover spatio-temporal patterns of mobility-driven transmission, offering valuable insights to inform the timing and regional targeting of non-pharmaceutical interventions.

Keywords Epidemic, Human mobility, Causal inference, Transfer entropy

The COVID19 pandemic emerged as an unprecedented global health crisis, challenging healthcare systems, economies, and societal norms worldwide. As nations endeavour to fortify their preparedness against future epidemics, integrating innovative data-driven strategies becomes imperative[1]. One promising approach is the use of anonymized phone-based mobility data (APMD) and reports of new infections to guide the design of public health responses such as non-pharmaceutical interventions (NPIs)[2–5]. Over the past decade, APMD applications have emerged as versatile tools, serving to model human mobility[6,7], analyze urban spatial structures[7,8], infer friendship networks[9], map dynamic populations[10], predict poverty and wealth[11], and aid disaster management[12]. In the realm of epidemiology, this data has been seamlessly integrated into various classes of epidemiological models to account for population mobility between regions[13–18].

While population mobility undoubtedly plays a role in the spread of infectious diseases, establishing a causal relationship or coupling between mobility patterns and regional disease outbreaks remains a challenging task. Several studies highlight that the relationship between mobility and the spread of COVID19 is highly context-dependent, shaped by factors such as pathogen-specific dynamics, geographic and demographic variability, and periods of behavioural or policy-driven changes. For instance, Perofsky *et* al. demonstrated that mobility positively predicted the transmission of endemic respiratory viruses but exhibited a negative or lagging correlation with SARS-CoV-2 during periods of stringent behavioural restrictions. The authors conclude that mobility is most predictive of respiratory virus transmission during significant behavioural shifts and the early stages of epidemic waves[19]. Similarly, Grantz *et* al. found that mobility reductions in highly populated counties were linked to declining infection rates during the early stages of the COVID19 pandemic, although this relationship was inconsistent across other regions and timeframes[20]. Another study integrating multiple mobility

[1]Barcelona Supercomputing Center, Barcelona, Spain. [2]Universitat Rovira i Virgili, Tarragona, Spain. [3]Pacific Northwest National Laboratory, Richland, WA 99354, USA. [4]ICREA, Passeig Lluís Companys 23 , 08010 Barcelona, Spain. ✉email: miguel.ponce@bsc.es

data sources over one year revealed that mobility does not consistently provide significant predictive power for COVID19 spread, underscoring the limitations of mobility data as a universal predictor[21]. These studies suggest that mobility has a stronger influence on transmission during periods of dramatic behavioural change or at the onset of epidemics. Nonetheless, finding a way to detect whether an increase in the number of cases in a region is likely to 'cause' an outbreak in a different region is critical to help in the development of effective NPIs[3,22]. We found that the big volumes of data in the form of time series generated during the COVID19 pandemic[23–27] constitute an excellent opportunity to develop, refine and validate methodologies aimed at elucidating potential causal relationships among various factors, including population mobility, emergence of novel variants, regional outbreaks, etc.

Causality inference is a complex and longstanding issue, drawing from various disciplines such as philosophy, economics, statistics, and biology[28]. Various definitions exist to describe the causal relationship between two variables X and Y[29]. A widely accepted definition posits that X causes Y if an intervention on X leads to a change in Y[28]. Leveraging this definition, causal relationships can be deduced from time series data using statistical and data-driven methodologies. These approaches, devoid of explicit mechanistic models of data generation, are commonly termed as "model-free" methods (for an extensive review, see[29,30]).

Model-free methods can be broadly categorised into two classes[31]. The first class comprises methods that rely on specific assumptions about the underlying processes generating the observed data. One prominent example is Granger Causality (GC), a statistical test assessing whether variable X enhances the predictive capability of variable Y[32,33]. While GC has found success across various domains, its utility is constrained by several restrictive assumptions, including linearity and stationarity, among others (see[31]] for a comprehensive review). Additionally, GC presupposes that certain properties of X can change independently of Y and that information flow is predominantly unidirectional. However, in complex dynamical systems, interactions between components can exhibit bidirectional tendencies, resulting in cyclic information flows[34,35]. Moreover, dependencies among system components can evolve over time and often manifest nonlinearly, rendering many of GC's assumptions invalid. In the context of epidemic processes, nonlinear dependencies and transient couplings between regions are anticipated due to the intricate interplay between daily population mobility and social interactions, which drive the spread of infection. Consequently, traditional causal inference methods such as GC encounter challenges in deciphering cause and effect within complex systems where causal relationships can become entangled[36].

The second class encompasses information-based methods that quantify the additional insights into the dynamics of Y gained from knowledge of X[29,37]. One notable example is Transfer Entropy (TE), a technique for assessing the directed (time-asymmetric) transfer of information between two stochastic processes[38]. TE serves as a generalisation of Granger Causality (GC), with both methods being entirely equivalent for Gaussian variables[39]. While TE is often described as a model-free or nonparametric method, this refers to the fact that it does not assume a predefined model structure for the underlying processes[40]. However, its application does involve methodological requirements, such as discretization of continuous data, approximate stationarity, and assumptions about the temporal dependencies e.g., the Markov order of the processes. Consequently, TE has found application in studying causal relationships within nonlinear systems across diverse domains such as neuroscience[37], complex systems[41,42], and economics and social science[43,44], among others.

TE has been used in the field of epidemic modelling to examine age-structured transmission dynamics during the autumn wave of the 2009 influenza pandemic in the USA[45]. More recently, TE was applied to quantify the relationship between mobility metrics and COVID19 cases and deaths in some European regions[21]. Specifically, it was assessed how each region's internal mobility influenced the number of COVID19 cases in that same region in a static manner, considering the overall correlation during the full period of the pandemic. Here, we present a different analysis that takes into account aspects such as the influence one region has over the others and the temporal dynamics of the pandemic.

In this study, we propose a data-driven framework to assess the role of interregional mobility in shaping epidemic dynamics using Transfer Entropy. Our approach focus on quantifying the directional information flow from the time series of inter regional mobility-associated risk (a proxy for the number of potentially infected individuals traveling between regions), to the time series of newly reported COVID19 cases. This enables us to examine whether interregional flows of individuals, weighted by infection prevalence at the source, are associated with subsequent daily reported cases at the destination region.

We apply this framework to the COVID19 pandemic in Spain, combining daily confirmed case counts with origin–destination mobility matrices derived from anonymized phone-based mobility data to compute MAR between provinces[27]. At the national scale, we characterize both global and local patterns of information flow across provinces and identify periods in which specific regions acted as major drivers of information transfer. We further demonstrate the applicability of our method to policy evaluation by analyzing the effect of a non-pharmaceutical intervention (NPI) implemented in Catalunya[4], on the coupling between mobility and disease spread. We show that the observed decrease in mobility-associated risk during the intervention period preceded a subsequent reduction in new cases, consistent with the expected impact of the NPI.

Finally, we validate our approach in controlled settings using epidemic simulations based on a metapopulation SIR model. These simulations confirm that TE can recover mobility-driven information flow under simplified and controlled scenarios and highlight the importance of accounting for the temporal structure of mobility networks when assessing causality in spatial epidemic spread. Altogether, our work provides a novel approach to study the effect of interregional mobility on epidemic spread and to uncover spatio-temporal patterns of mobility-driven transmission, offering valuable insights to inform the timing and regional targeting of non-pharmaceutical interventions.

Material and methods

Datasets

The COVID19 Flow-Maps are a cross-referenced geographical information system that includes time-series accounting for population mobility and daily reports of COVID19 cases in Spain at different scales of time and spatial resolution[27]. In this work, we have used different data sets from COVID19 Flow-Maps available at Zenodo. The datasets include daily reports on COVID19 cases[46], census population[47], and population mobility[48] at different levels of geographic resolution[49]. Since the different data records span different periods, we selected the maximum possible time interval that includes reported entries in all the datasets, including case reports, population, and mobility. We trimmed all the data records from 01-03-2020 to 01-05-2021, the maximum time range to which mobility and population data sets are available. Table 1 summarizes the different datasets used in this work, which are explained in the further subsections.

COVID19 case reports

We retrieved daily cases for all of Spain reported at the level of provinces and the level of Basic Health Areas (BHA) for Catalunya and Madrid[46]. Each record corresponds to a geo-referenced time series where each record has an associated date, the corresponding identifier of the layer, a code of the zone (*i.e.* a province or BHA) as well as the number of cases reported on that date (daily incidence). Then, for a given region $i \in L$ and a date $t \in T$, we denote the daily reported new cases as $C_i(t)$. To minimize the impact of artefacts in case reports (e.g., reporting delays or weekly patterns), we applied a 7-day rolling average to smooth the time series $C_i(t)$. In this context, by 'smoothing' we refer to the process of averaging the number of new daily COVID19 cases over a 7-day window to reduce short-term fluctuations and irregularities in the data, such as reporting delays or weekly cycles, by producing a more stable and consistent representation of the underlying trends in case counts. The cases dataset for Madrid is reported by day until the end of June and after by week. Therefore, to avoid artefacts due to the differences in the reporting frequency for the case of Madrid, we applied a 14-day moving average. Throughout this work, $C_i(t)$ denotes the smoothed time series of reported cases unless otherwise specified. We also calculate the 10-day cumulative incidence, denoted by $C_i^{10}(t)$, and use this quantity as a proxy of the number of infected individuals $I_i(t)$ at time t[27,50]. We also normalized $C_i(t)$ by the total population N_i and rescaled this value to express the incidences as new cases by 100,000 inhabitants and denote it as $\hat{C}_i(t) = C_i(t)/N_i \cdot 100,000$.

Population and daily mobility data

Mobility and population data records come from a study conducted by the Spanish Ministry of Transport, Mobility and Urban Agenda (MITMA, *Ministerio de Transportes, Movilidad y Agenda Urbana*) which analyses the mobility and distribution of the population in Spain from February 14th 2020 to May 9th 2021 (https://www.mitma.gob.es/ministerio/COVID19/evolucion-movilidad-big-data). The data records are based on a sample of more than 13 million anonymized mobile phone lines provided by a single mobile operator whose subscribers are evenly distributed and include two different mobility indicators. Daily origin-destination matrices (ODM) account for the number of trips between 2850 *mobility areas* that cover almost the entire territory of Spain, reconstructed by combining cell phone antenna coverage areas with districts and municipalities. In this work, we used ODMs[48] and population data[47], reported at the province and Basic Health areas levels. The raw data are obtained from a single mobile phone operator covering approximately 30% of the Spanish population and are processed using sample expansion methods to estimate mobility flows for the entire population, as detailed in the original dataset publication[27]). For mobility indicators and population, we will use the following notations: $M_{ij}(t)$ is the daily number of trips from zone *i* to *j* with both $i, j \in L$ and reported at date *t*, and will use N_i to refer to the total population for zone $i \in L$.

Mobility-associated risk

To assess the effect of population mobility on the spread and outbreaks of COVID19, we have previously introduced the Mobility Associated Risk (MAR)[27]. The MAR $R_{ij}(t)$ is an estimator of the potential number of infected individuals moving from zone *i* to *j* on the date *t* and is given by the following expression:

Region	Dataset	Layer	N° patches
Spain	COVID19 cases	Provinces	50
Spain	Population	Provinces	50
Spain	Mobility	Provinces	50†
Catalunya	COVID19 cases	BHA	374
Catalunya	Population	BHA	374
Catalunya	Mobility	BHA	374
Madrid	COVID19 cases	BHA	286
Madrid	Population	BHA	286
Madrid	Mobility	BHA	286

Table 1. Summary of the different datasets used in this work. †The cities of Ceuta and Melilla were excluded from the study because of the low coverage found in the phone-based mobility data.

$$R_{ij}(t) = \frac{I_i(t)}{N_i}\ M_{ij}(t) \tag{1}$$

, where the first term on the right side represents the density of infected individuals in the source zone *i* per the total number of inhabitants N_i and $M_{ij}(t)$, is the number of trips from region *i* to *j* at time *t*. $R_{ij}(t)$ is the estimated number of infected individuals commuting between regions *i* and *j* at time *t*. For a given date *t*, the values $R_{ij}(t)$ can be arranged in a $|L| \times |L|$ matrix $R(t)$ which represents a temporal directed weighted network where the nodes correspond to the different zones $i, j \in L$ and the flow corresponds to the estimated number of infected individuals commuting between regions *i* and *j* at time *t*. When working with simulated data, I_i is directly calculated through numerical integration of the model equations (see Eq. 8). Therefore, to obtain the mobility-associated risk $R_{ij}(t)$ between *i* and *j* for the simulated data I_i is divided by the total population N_i at region or patch *i* and then, multiply this value by the number of commuters $M_{ij}(t)$.

When dealing with real-world data, COVID19 incidence time series report the number of new daily cases reported $C_i(t)$ for a region *i*. The calculation of $R_{ij}(t)$, as defined in Eq. (1) requires the number of infected individuals $I_i(t)$ in the region *i* at time *t*, a quantity unknown in real-world scenarios. Hence, we used the accumulated incidence over the ten days $C_i^{10}(t)$ preceding the report as a proxy of the number of infected individuals $I_i(t)$ at time *t*. Then, we scale $C_i^{10}(t)$ by the total population N_i and use this quantity together with the trips $M_{ij}(t)$ and calculate $R_{ij}(t)$ from *i* to *j*[27].

Transfer entropy

Basic definitions

Transfer entropy (TE) is an information-theoretic measure that quantifies the directional flow of information between two time-dependent discrete random processes, $X(t)$ and $Y(t)$, defined over a discrete alphabet or set of symbols A[38]. Specifically, TE measures the reduction in uncertainty about the future value $y_{t+1} \in A$ when incorporating knowledge of both the past of $Y(t)$ and the past of $X(t)$, denoted $y_t^{(k)}$ and $x_t^{(l)}$, respectively. This approach provides insights into potential directed (non-symmetric) dependencies between variables and has been widely used to infer directional influences in time series. To compute TE, joint and conditional probabilities are estimated from the empirical frequencies of symbol sequences in the observed time series. The transfer entropy from X to Y is defined as:

$$TE_{X\to Y}(k,l) = \sum_{x,y} p(y_{t+1}, y_t^{(k)}, x_t^{(l)}) \log_2 \frac{p(y_{t+1}|y_t^{(k)}, x_t^{(l)})}{p(y_{t+1}|y_t^{(k)})} \tag{2}$$

where $y_t^{(k)} = (y_t, \ldots, y_{t-k+1})$ and $x_t^{(l)} = (x_t, \ldots, x_{t-l+1})$ represent the past states of Y and X up to orders k and l, respectively. Following the original formulation by Schreiber [38], we adopt the commonly used setting of $k = l = 1$ in our analyses.

Assumptions of transfer entropy analysis

Although Transfer Entropy (TE) is often described as a model-free method, its application to empirical time series involves several methodological assumptions. First, TE is typically defined for discrete symbolic sequences; thus, continuous-valued data must be transformed accordingly (see Discretisation below). Second, TE implicitly assumes that the underlying stochastic processes exhibit Markovian properties of a given embedding order. In our case, this assumption is justified by modelling epidemic dynamics at the metapopulation level, where transitions between states have been shown to follow Markovian approximations[51]. While we acknowledge that alternative non-Markovian approaches are gaining traction in the literature[52], we consider the Markov assumption both reasonable and empirically grounded for our chosen modelling framework. Finally, TE analysis requires approximately stationary time series. When this condition is not met, a common approach is to compute first-order differences, which often results in a stationary series. In our analysis of TE between mobility and new case reports, we apply this differencing procedure to ensure stationarity in all the time series.

Calculation of transfer entropy between mobility-associated risk and incidence time series

In this work, we aim to assess the impact of mobility on the dynamics of the epidemic spread of COVID19. We hypothesise that if an outbreak in a target region *j* is triggered by infected individuals arriving from a source region *i*, there might be a detectable flow of information from the time series of mobility-associated risk $R_{ij}(t)$ (see Eq. 1) to the new infections in region *j* (scaled by 100.000 inhabitants) denoted by $\hat{C}_i(t)$. Both $R_{ij}(t)$ and $\hat{C}_i(t)$ are continuous-valued time series: the former represents the estimated flux of infected individuals between regions (combining trip volume and the density of infected individuals), while the latter corresponds to the daily new cases per 100,000 inhabitants. To quantify the directional relationship between them, we measure the transfer entropy (TE) from $R_{ij}(t)$ to $\hat{C}_j(t)$, as defined in Eq. 2.

Time series stationarity. Since the TE framework requires the input time series to be stationary, we first tested the stationarity of both $R_{ij}(t)$ and $\hat{C}_i(t)$ using the Augmented Dickey-Fuller (ADF) test. In most cases, neither time series could be assumed stationary (see Supplementary Table 1). To address this, we applied a first-order differencing transformation to both series to achieve stationarity:

$$\begin{aligned} \Delta R_{ij}(t) &= R_{ij}(t) - R_{ij}(t-1) \\ \Delta \hat{C}_j(t) &= \hat{C}_j(t) - \hat{C}_j(t-1) \end{aligned} \tag{3}$$

We then applied the ADF test to the transformed time series $\Delta R_{ij}(t)$ and $\Delta \hat{C}_j(t)$ and found that all of them were stationary, with p-values below the 0.05 significance level, except for one case: the mobility-associated risk $R_{ij}(t)$ between the provinces of Zaragoza and Badajoz (see Supplementary Table 1).

Discretization of continuous time series. In addition to the stationary condition, the framework of Transfer Entropy also requires the input data to be discrete. A common method for transforming a continuous-valued series $X(t)$ into a symbolically encoded series $S(t)$ involves partitioning the values into a limited number of bins[43,53]. This method enables each value in the time series $X(t)$ to be assigned an integer label $(-1, 0, 1)$

$$S(t) = \begin{cases} -1 & \text{for } X(t) \leq q_1 \\ 0 & \text{for } q_1 < X(t) < q_2 \\ 1 & \text{for } X(t) \geq q_2 \end{cases} \tag{4}$$

The definitions of q_1 and q_2 can be done by either defining a set of uniform set of bins or by calculating quantiles from the Kernel Density Estimation of the distribution of observed values. In our case, the time series $\Delta R_{ij}(t)$ and $\Delta \hat{C}_j(t)$ units in the number of individuals, and therefore, a natural choice for partitioning the observations into discretized values is to set $q_1 = -q$ and $q_2 = q$ with $q = 1$. While we test the sensitivity of the results to the selection of q, the results presented in the Results section are for $q = 1$. Finally, we will use $X_{ij}(t) \in \{-1, 0, 1\}$ and $Y_j(t) \in \{-1, 0, 1\}$, with $i, j \in L$ and $t \in T$ to refer to the transformed time-series of $\Delta R_{ij}(t)$ and $\Delta \hat{C}_j(t)$, respectively. Then, to estimate the effect of the population mobility from *i* to *j* in the number of newly reported cases we will measure the transfer entropy between $X_{ij}(t)$ and $Y_j(t)$ that for simplicity will be denoted as $TE_{i\to j}$.

Considering incubation periods and reporting delays. To account for the incubation period and delays in case reports, we introduce a time delay δ (days) between the mobility-associated time series $X_{ij}(t)$ and the new case series $Y_j(t + \delta)$. Previous studies have reported incubation periods ranging from 5 to 6 days, with most symptomatic cases developing within 11 days[54,55].

Global and local transfer entropy. For each pair of regions *i* and *j*, we compute the global transfer entropy $TE_{i\to j}$ using the full-length time series $X_{ij}(t)$ and $Y_j(t + \delta)$. To capture temporal variations in information flow, we also compute a local (dynamic) transfer entropy $TE_{i\to j}(t)$ using a sliding window approach. Specifically, we define a window of size ω days and slide it across the time series in daily increments. For each time point *t*, we compute the TE using the mobility-associated risk time series $X_{ij}(s)$ over the interval $s \in [t, t + \omega]$, and the delayed incidence time series $Y_j(s)$ over $s \in [t + \delta, t + \omega + \delta]$. The resulting TE value is assigned to the right edge of the window, i.e., to time $t + \omega + \delta$. This assignment convention ensures consistent alignment with the most recent time point included in the estimation window, which is particularly relevant for retrospective or real-time monitoring applications. We selected $\omega = 28$ days to smooth out short-term fluctuations due to weekly reporting artifacts while preserving sufficient temporal granularity. The lag $\delta = 7$ days accounts for typical delays between mobility-induced exposure and case reporting, and both parameters are used consistently across the synthetic simulations and the real COVID19 data analysis.

Testing the effect of mobility. To investigate the role of mobility, we compute both global and local transfer entropy using the differenced time series of infections $Y_i(t)$ and $Y_j(t)$, thereby excluding the influence of mobility. This allows us to quantify the information transfer between the epidemic dynamics of regions *i* and *j* independently of mobility data. To avoid confusion, we explicitly denote this case-only transfer entropy as $TE^{\text{cases}}_{i\to j}$.

Effective transfer entropy

Some authors have claimed that transfer entropy estimates are likely to be biased due to small sample effects and thus propose a modified version of the transfer entropy, called the Effective Transfer Entropy[56]. Given a source and target time series $X(t)$ and $Y(t)$, respectively, the effective transfer entropy $ETE_{X\to Y}$ is computed by subtracting from the $TE_{X\to Y}$ (Eq. 2) the term $TE_{X_{shuffled}\to Y}$ as follows:

$$ETE_{X\to Y}(k, l) = TE_{X\to Y} - TE_{X_{shuffled}\to Y} \tag{5}$$

where $TE_{X_{shuffled}\to Y}$ is the transfer entropy calculated using a shuffled version of the time series of the variable $X(t)$. This definition is an attempt to reduce both the statistical dependencies between the time series $X(t)$ and $Y(t)$ and the time series dependencies of $X(t)$. This procedure's base relies on generating a new time series by a random realignment of the values from the time series $X(t)$ to obtain the shuffled time series $Z(t)$ and measuring the *TE* between the shuffled variable $Z(t)$ and $Y(t)$. This procedure is repeated *n* times and averaged to obtain $TE_{X_{shuffled}\to Y}$[56].

Directionality index

An essential property of Transfer Entropy is its asymmetry, i.e. $TE_{X\to Y} \neq TE_{Y\to X}$, this property allows quantifying the directional coupling between components of complex systems. Therefore, the difference between $TE_{X\to Y}$ and $TE_{Y\to X}$ can provide information regarding the dominant direction of the information flow[53]. The directionality index, $DI_{X\to Y}$, measures the asymmetry in the transfer of information between two processes and is given by:

$$DI_{X\to Y} = TE_{X\to Y} - TE_{Y\to X} \tag{6}$$

A positive value of $DI_{X\to Y}$ means process $X(t)$ is transferring a net flow of information to process $Y(t)$, whereas a negative value of $DI_{X\to Y}$ means that the net flow of information goes from $Y(t)$ to $X(t)$. Therefore, we use the net directional flow of information, $DI^{+}_{X\to Y}$ given as follows:

$$DI^{+}_{X\to Y} = \begin{cases} DI^{+}_{X\to Y}, & \text{if } DI^{+}_{X\to Y} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

to measure the dominant direction of the information flow between two components $X(t)$ and $Y(t)$.

Epidemic simulations

We conducted epidemic spread simulations using a Susceptible-Infected-Recovered (SIR) compartmental model within a metapopulation framework. In this model, a population of N individuals resides across a set L of regions connected by a network of daily commuters, who may travel to a different region for work each day. Here, N_{ij} denotes the fraction of individuals living in the region i who commute daily to j for work, and $N_i = \sum_j N_{ij}$ represents the total population residing in i (Fig. 1). Moreover, we use $S_{ij}(t)$ and $I_{ij}(t)$ to denote the number of susceptible and infected individuals living in region i and working in region j at time t, respectively.

If the first infected individuals are introduced in a region i, the epidemic will spread to other areas via infected daily commuters or through susceptible individuals who commute to region i and become infected in their work region and subsequently carry the infection back to their home region. To simulate infection during working time in the region i, we draw the number of new infections from a binomial distribution based on the total pool of susceptible individuals $S_i = \sum_j S_{ji}$, which includes individuals living and working in i (S_{ii}) and all other individuals working in i. The force of the infection is defined as $\lambda_i^{\text{work}} = (\frac{\beta}{2}\frac{I_{ii}+\sum_j I_{ji}}{N_{ii}+\sum_j N_{ji}})$ where β (days^{-1}) is the disease transmission rate and the terms I_{ii} and I_{ji} represent the infected individuals living and working in i those commuting from j to i, respectively. A similar approach is used to simulate infection events at home locations for the remainder of the day. Here, the number of new infections is drawn from a binomial distribution based on the total pool of susceptible individuals $S_i = \sum_j S_{ij}$ with a force $\lambda_i^{\text{home}} = \frac{\beta}{2}(\frac{I_{ii}+\sum_j I_{ij}}{N_{ii}+\sum_j N_{ij}})$. Below, there is a set of differential equations that govern the dynamics of the systems:

$$\begin{aligned} \frac{dS_{ij}}{dt} &= -S_{ij}(\lambda_i^{\text{work}} + \lambda_i^{\text{home}}) \\ \frac{dI_{ij}}{dt} &= -\mu I_{ij} + S_{ij}(\lambda_i^{\text{work}} + \lambda_i^{\text{home}}) \\ \frac{dZ_{ij}}{dt} &= \mu I_{ij} \end{aligned} \tag{8}$$

To avoid notational confusion with the mobility-associated risk metric $R_{ij}(t)$, we denote the number of recovered individuals living in region i and working in region j as $Z_{ij}(t)$, instead of the usual $R_{ij}(t)$ used in standard SIR-type models. Further details on this simulation approach can be found in previous works[13,14]. For

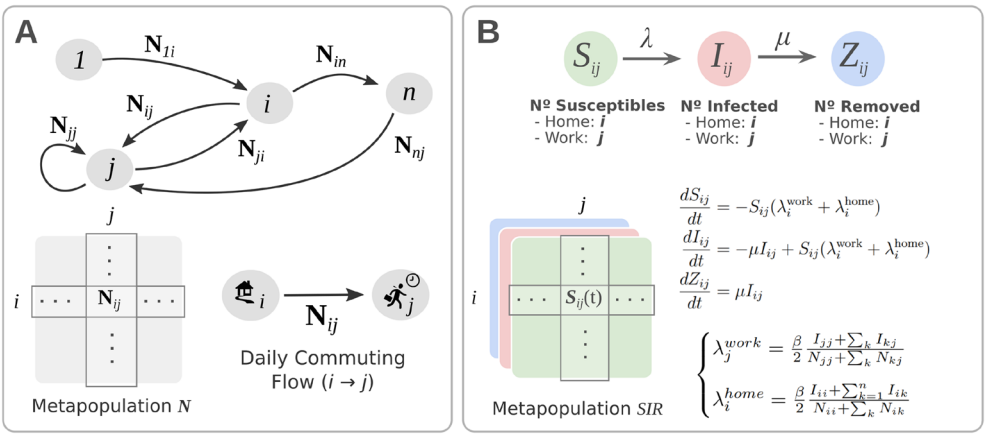

Fig. 1. SIR metapopulation model. Panel **A** represents a metapopulation composed of n zones or regions connected through a network of daily commuters. For two areas i and j, N_{ij} denotes the total population that lives in zone i and works in region j. Panel **B** shows a schematic representation of the SIR model in a metapopulation. S, I and R denote the three different stages of disease and $S_{ij}(t)$, $I_{ij}(t)$, $Z_{ij}(t)$ denote the number of susceptible, infected and removed individual living in zone i and working in zone j at time t, respectively. At any given time t, the state of the model and any given time can be arranged as a tensor of dimension $n \times n \times 3$.

our study, we conducted numerical simulations using EpiCommute[14], a Python implementation of the approach developed by Tizzoni et al.[13], available at https://github.com/franksh/EpiCommute/.

We evaluated three metapopulation structures of increasing complexity: a simple chain, a ring, and an extended star network (Supplementary Fig. 21). In each case, we assumed populations of equal size ($N = 100,000$ individuals) and a fixed number of commuters ($N_{ij} = 5,000$) if regions *i* and *j* are connected. We also considered a realistic scenario, modelling the metapopulation as the provinces of Spain, connected by recurrent mobility patterns inferred from anonymized phone-based mobility data (see the subsection on Datasets for further details). Since the mobility in the SIR models is static, we simulated two scenarios, considering two different recurrent mobility matrices obtained from the average mobility over one week before the lockdown (from *14-03-2020* to *21-03-2020*) and during the lockdown (from *14-04-2020* to *21-04-2020*). The corresponding mobility matrices were computed by averaging daily flows over one week before the lockdown (March 14–21, 2020) and one week after (April 14–21, 2020).

For each metapopulation model, we simulated the epidemic dynamics using a basic reproductive number of $R_0 = 2.5$, which represents a moderate level of transmissibility consistent with COVID19 estimates, typically ranging between 2.43 and 3.10 [57]. The average infectious period was set to $\mu^{-1} = 5$ days, in line with prior studies [14]. Due to the stochastic nature of the simulations, we performed 1,000 independent runs per scenario and averaged the results across all replicates. Each simulation was initialized with three infected individuals introduced into a single subpopulation; for the Spain model, this seed population was located in Madrid. To analyze the simulated outbreaks, we extracted time series representing both mobility-associated infection risk and daily incidence. Specifically, we used $I_{ij}(t)$ to denote the number of infected individuals residing in region *i* and traveling to region *j* at time *t*. This quantity reflects the actual number of infectious travelers and was used instead of a proxy for mobility-associated risk. Daily new infections in each region *j*, denoted $C_j(t)$, were computed by taking the difference in the number of susceptible individuals between consecutive time steps:

$$C_j(t) = \sum_k S_{jk}(t) - \sum_k S_{jk}(t-1)$$

Ones we obtained the time series we compute the first order difference and discretize the data following the approach used for the real datasets (see Calculation of Transfer Entropy between mobility-associated risk and incidence time series) and subsequently computed both global and dynamic effective transfer entropy (ETE) and directional influence (DI) using $\omega = 28$ (days) and $\delta = 7$ (days).

Results and discussion

The results of this study are structured to systematically explore the interplay between population mobility and the spread of infectious diseases, using Transfer Entropy (see Fig. 2). The first section analyses the relationship between mobility patterns and COVID19 case data across Spanish provinces, identifying regions that significantly influenced the epidemic's progression. The second section focuses on the dynamic nature of these interactions, using local or dynamic TE to reveal temporal changes in information flow. In the subsequent section, the impact of non-pharmaceutical interventions is assessed, with the closure of bars and restaurants in Catalunya serving as a case study to illustrate how mobility reductions significantly correlate with the infection trends. Finally, simulated epidemic scenarios validate the framework, demonstrating the critical role of mobility data in avoiding spurious correlations and accurately capturing the drivers of disease spread.

Measuring information flow between mobility patterns and epidemic spread

We used the proposed approach to investigate the role of population mobility in the spread of COVID19 in Spain from 1 March 2020 to 1 May 2021. Phone-based anonymised mobility data was used to compute the mobility-associated risk $R_{ij}(t)$ between regions and the time series of daily reported cases by 100,000 inhabitants $\hat{C}_i(t)$ for the different regions *i* (see Fig. 2A-B and Datasets for more details). We first focus the analysis at the national level, considering as regions the fifty provinces of Spain and excluding the autonomous cities of Ceuta and Melilla (see Supplementary Fig. S5 and Material and Methods Sections). We measured the global $TE_{i\to j}$ between each pair of provinces in Spain (see Fig. 2C). To discretise the time series of mobility-associated risk and daily reported COVID19 cases, we set $q = 1$ (see Calculation of Transfer Entropy between mobility-associated risk and incidence time series for further details on the calculation). The results show that Madrid is the province that exhibits the highest $TE_{i\to j}$, which is expected due to fact that Madrid's high population density, its central location and Madrid city is Spain's Capital (see Fig. 3A and C and Supplementary Table S1).

Furthermore, we analysed the directional information difference $DI^{+}_{i\to j}$, and total directional information DI_i, among provinces (Fig. 3B). The resulting patterns differ markedly from those seen in the analysis of total transfer entropy $TE_{i\to j}$ (Fig. 3A). For instance, Madrid, which shows the highest value of total TE_i, shows a slightly negative value for the DI_i, indicating that the information received by Madrid from other provinces exceeds the amount it transmits to them (Fig. 3B). Comparing province rankings based on TE_i and DI_i reveals substantial changes, highlighting potential asymmetries in information transfer between neighbouring regions. To further assess this asymmetry, we compared $TE_{i\to j}$ and $TE_{j\to i}$, finding a significant correlation ($R = 0.62$ and *p*-value$= 10^{-20}$) (see Supplementary Fig. S6A), suggesting that information flow between provinces is relatively balanced. However, notable exceptions exist, such as between A Coruña and Lugo, or Asturias and Lugo, where one province exhibits a substantially stronger influence over the other. These cases illustrate instances of pronounced directional influence between regions.

For each source province, we estimate the distribution of the maximum values $\max_j TE_{ij}$ for each province *i* (see Supplementary Fig. s7). We calculate the 10% percentile of these maximum values and use it as a threshold

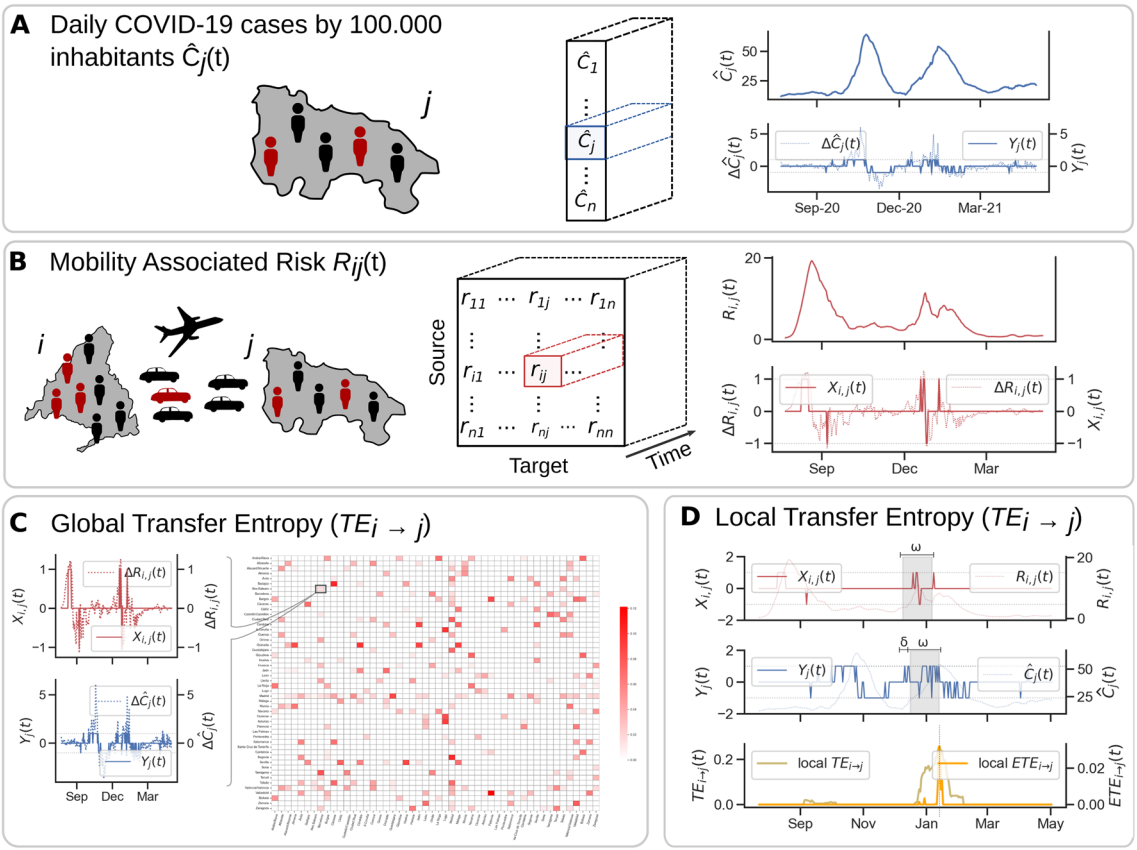

Fig. 2. Measuring transfer entropy between time series of phone-based mobility data and incidence to explore patterns of disease spread. Panel **A** displays the COVID19 incidence $\hat{C}_j(t)$ for region *j*, representing new daily cases per 100,000 inhabitants. This panel also includes the first-order difference $\Delta\hat{C}_j(t)$ and the resulting time series $Y_j(t)$, derived by discretising $\Delta\hat{C}_j(t)$. Panel **B** shows the mobility-associated risk between source region *i* and target region *j* ($R_{ij}(t)$), together with the first-order difference $\Delta R_{ij}(t)$ and the resulting time series $X_{ij}(t)$, derived by discretizing $\Delta R_{ij}(t)$. Panel **C** illustrates the calculation of the global Transfer Entropy $TE_{i\rightarrow j}$, which quantifies the information flow between the complete time series $X_{ij}(t)$ and $Y_j(t)$ to examine the influence of source region *i* on the case dynamics in target region *j*. Panel **D** illustrates the calculation of the local or dynamic Transfer Entropy $TE_{i\rightarrow j}(t)$ using sliding windows of size ω days to account for the local changes in the flow of information, considering the values of $X_{ij}(t)$ in the time interval $[t, t+\omega]$ and the values of $Y_j(t)$ in the time interval $[t+\delta, t+\delta+\omega]$. The base map was retrieved from CNIG https://centrodedescargas.cnig.es/CentroDescargas/ under CC-BY 4.0 Licence https://www.ign.es/resources/licencia/CondicionesLicenciaUso_IGN.pdf.

to filter out small interactions. After applying this filter, we represent the $TE_{i\rightarrow j}$ as a directed network on the map, finding that the flow of information primarily occurs between neighbouring provinces (Fig. 3A). Interestingly, the results show almost no Transfer Entropy for certain provinces, including the Canary Islands, the Balearic Islands, and Soria.

To investigate the contribution of mobility in the observed patterns of information flow, we computed the $TE^{cases}_{i\rightarrow j}$, i.e., the information flow between the time series of cases without considering the mobility (see Calculation of Transfer Entropy between mobility-associated risk and incidence time series). We found significant transfer entropy between almost every pair of provinces (Fig. 3D), suggesting that incorporating mobility helps filter out interactions that may otherwise be confounded or indirect when using case data alone (see Fig. 3C). It is important to note that the presence of significant TE between case time series, even when mobility data is not explicitly considered, does not necessarily indicate misleading or invalid results. Instead, it may reflect dependencies arising from unobserved or indirect mobility pathways, regional synchrony, or shared external influences. Nevertheless, by incorporating direct mobility measurements in the form of the mobility-associated risk, we aim to better resolve the specific pathways through which inter-regional influence occurs, thereby enhancing the interpretability of the observed information flow and grounding it in known mechanisms of disease transmission. We also compared the asymmetry between $TE^{cases}_{i\rightarrow j}$ and $TE^{cases}_{j\rightarrow i}$ and observed a remarkable reduction in the correlation coefficient ($R = 0.19$), emphasizing the critical role of mobility in shaping these dynamics (see Supplementary Fig. S6B). This decrease in symmetry suggests that case-only time series reflect more indirect or confounded dependencies, whereas mobility-informed TE captures more

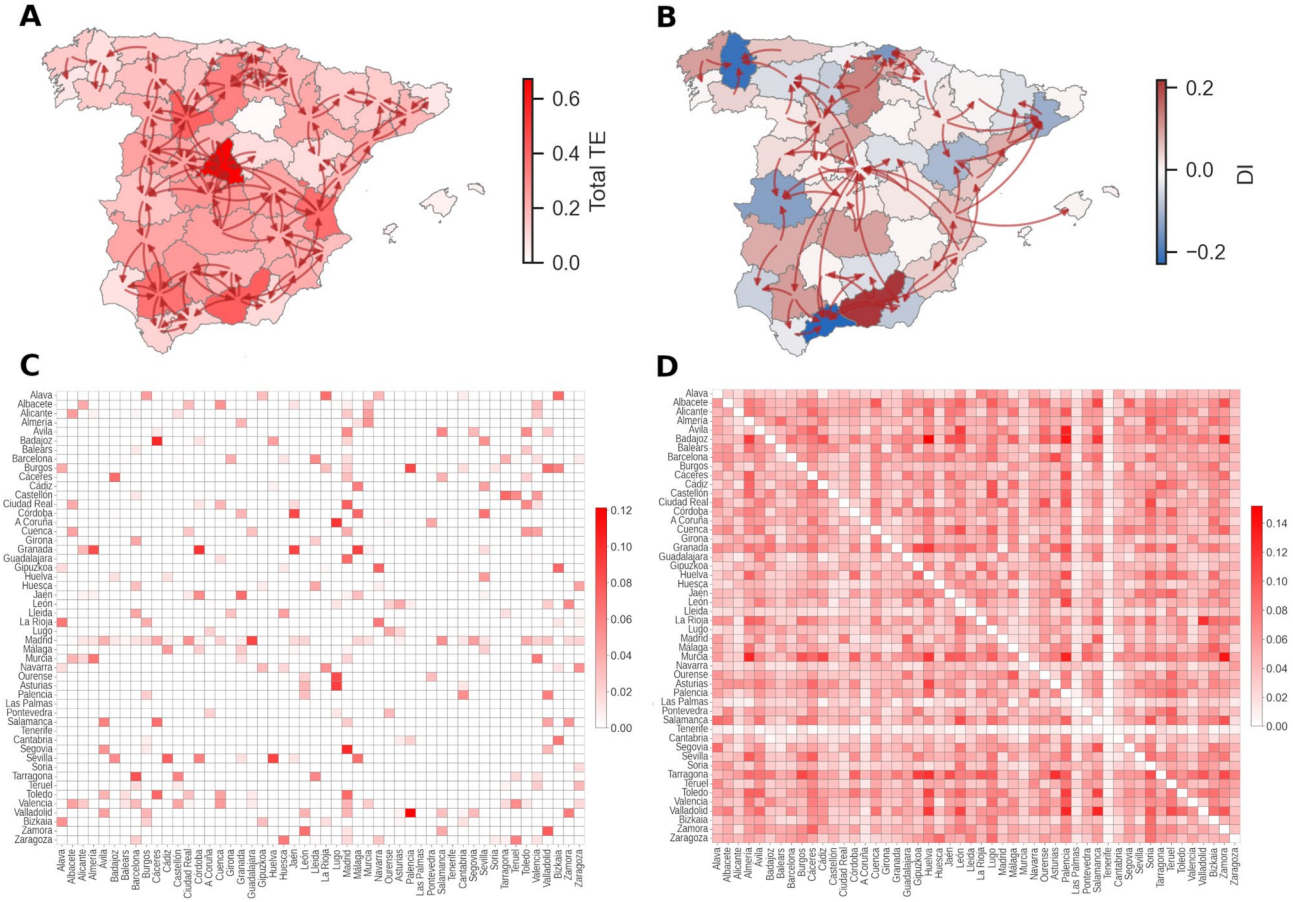

Fig. 3. Transfer entropy $TE_{i\to j}$ between provinces. Panel **A** shows the $TE_{i\to j}$ between provinces represented by arrows, and the colour of each patch represents the total TE_i transferred by each province *i*. Panel **B** shows the $DI^{+}_{i\to j}$ between provinces represented by arrows, and the colour of each patch represents the total DI_i transferred by each province *i*. In panels **A** and **B**, values were filtered using the 10% percentile from the distribution of the $\max_j\ TE_{ij}$ and $\max_j\ DI^{+}_{i\to j}$, respectively, as explained in the main text. Panels **C** and **D** show the Transfer Entropy $TE_{i\to j}$ between provinces of Spain, considering and not considering mobility, respectively. TE calculations were performed setting $\delta = 7$ (days) and $q = 1$. The Canary Islands are excluded from Panels **A** and **B** due to the absence of measured values for $TE_{j\to i}$ and $DI^{+}_{i\to j}$. The base map was retrieved from CNIG https://centrodedescargas.cnig.es/CentroDescargas/ under CC-BY 4.0 Licence https://www.ign.es/resources/licencia/Condiciones_licenciaUso_IGN.pdf.

balanced, bidirectional influence patterns, as expected from the generally reciprocal nature of human mobility between regions

We further examined the relationship between inter-provincial distances and $TE_{i\to j}$, finding a negative correlation $R = -0.56$ where $TE_{i\to j}$ decays with distance (see Supplementary Fig. S8A). This result is expected since the average number of trips M_{ij} also shows a negative relation with distance (see Supplementary Fig. S8C). If the analysis is performed between distance and $TE^{cases}_{i\to j}$, we found that the correlation coefficient drops to $R = -0.24$ (see Supplementary Fig. S8B). To address potential biases in Transfer Entropy calculations arising from the potentially limited sample size, we conducted a complementary analysis using Effective Transfer Entropy $ETE_{i\to j}$. The results from $ETE_{i\to j}$ were consistent with those from $TE_{i\to j}$ (see Fig. 3C and Supplementary Fig. S9), as expected given the strong correlation between $TE_{i\to j}$ and $ETE_{i\to j}$ (see Supplementary Fig. S10). We also computed the $ETE^{cases}_{i\to j}$ and found that even with the correction, there is still information flow between almost every pair of provinces (see Supplementary Fig. S11).

Finally, we performed a sensitivity analysis and confirmed that the observed TE and ETE patterns are robust to changes in the delay parameter δ days and the discretisation threshold q (see Supplementary Fig. S1, S2, S3 and S4), with optimal and stable results obtained around $\delta = 7$ days and $q \approx 1$ (see Supplementary section *Sensitivity Analysis of TE and ETE to Parameters δand q*). Overall, these results suggest that incorporating mobility data into TE analysis not only improves the interpretability of inter-regional influence but also helps eliminate misleading correlations present in case-only analyses.

Temporal dynamics of mobility-driven information flow

Having established the existence of directional information flow between provinces through global TE measures, we next sought to explore how these relationships evolved by analyzing the dynamic (local) TE. To capture these temporal changes, we measured the dynamic, or local, $TE_{i\to j}(t)$ using a day sliding window of length ω days with a lag of δ days between the time series to account for the virus's incubation period and symptom onset delay (see Fig. 2D and the subsection Calculation of Transfer Entropy between mobility-associated risk and incidence time series from Material and Methods). We computed both the standard and effective versions of local $TE_{i\to j}(t)$, using $\omega = 28$ days, $\delta = 7$ days, and $q = 1$, consistent with the parameters for global $TE_{i\to j}$. The resulting TE values are assigned to the end of the window, i.e., time $t + \omega$, to reflect that the measured TE value is linked to the full time interval rather than a precise time point.

Figure 4 presents the time profiles of local $TE_{i\to j}(t)$ calculated between Barcelona and the Balearic Islands as well as in the opposite direction. The results indicate that, for most of the study period, no significant information flow exists between the two provinces. However, in late October, there is a notable peak in information flow from Barcelona to the Balearic Islands, and another peak occurs in February in the reverse direction. Interestingly, both peaks correspond to an increase in $R_{ij}(t)$, followed by a rise in new cases in the receiving region. Additionally, the timing of these peaks is consistent between both TE and ETE measurements.

To assess consistency, we compared global and dynamic $TE_{i\to j}$ by taking the time average of dynamic $TE_{i\to j}(t)$. The results show a high correlation between global and dynamic TE measurements, with correlation coefficients of 0.86 for standard TE and 0.76 for effective ETE (see Supplementary Fig. S12). This strong correlation indicates that while the global TE still captures the total information flow, the dynamic TE provides insight on the temporal profile for the information flow. This also indicates that the size of the sliding windows, i.e. $\omega = 28$ days, provides a sample size sufficient to capture stable patterns of information transfer. Additionally, we compared the time-averaged local $TE_{i\to j}$ and dynamic $ETE_{i\to j}$, finding a robust correlation of $R = 0.94$ between these measures (see Supplementary Fig. S13).

We then used our approach to analyze the spreading patterns of the COVID19 pandemic in Spain during the first three waves. Specifically, we focus on detecting which provinces drive the spread of cases and which ones are driven during the different phases of the studied period. We first calculated the directionality index $DI_{i\to j}$ between all the provinces to measure the coupling between the different regions. Then, to evaluate the role of each province we aggregated the total $DI_{i\to j}$ for each province on each day and used these values to identify which provinces were drivers ($DI_i > 0$) and which ones were being driven ($DI_i < 0$) during the different phases of the pandemic. Fig. 5A shows the DI_i profile for each province, ordered in decreasing order by the total DI_i aggregated across time, and thus main drivers are placed on the top and provinces that were mainly driven are placed at the bottom (Fig. 5A).

The results indicate that Madrid was the primary driver of COVID19 cases throughout the pandemic, followed by Álava and Zaragoza (Fig. 5A). Conversely, the Canary and Balearic Islands, along with Soria, were among the most influenced regions. This is expected, as the islands' isolation reduces their likelihood of being significant drivers, and Soria is Spain's least populated province. Notably, among the top three driver regions, Zaragoza appears as a key province connecting Barcelona and Madrid, and it played a crucial role as a driver

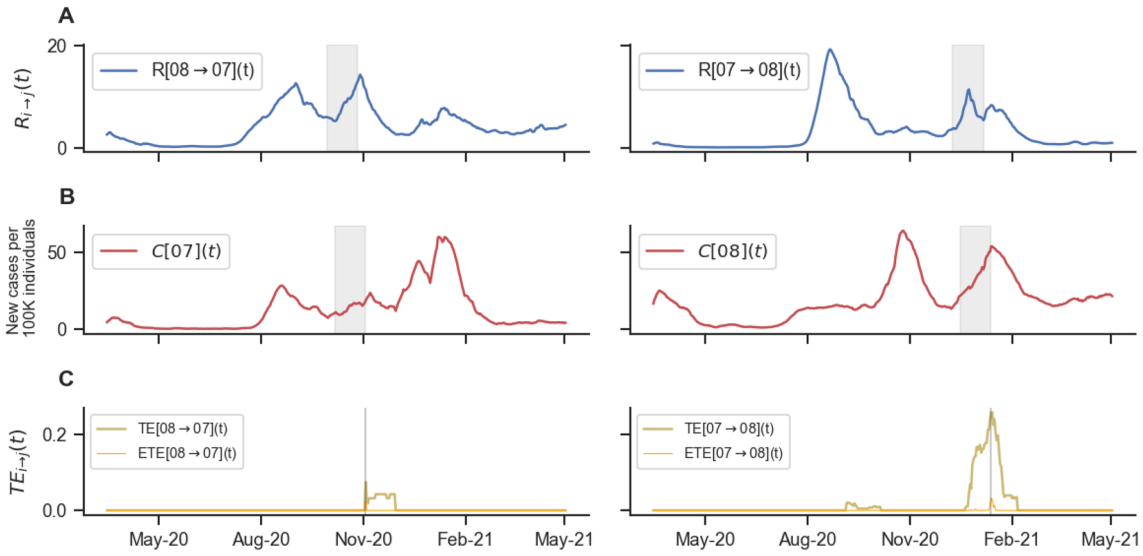

Fig. 4. Dynamics $TE_{i\to j}$ profiles. Panel **A** shows $R_{ij}(t)$ between Barcelona and Baleares on the left and in the opposite direction on the right. Panel **B** shows the normalized new cases for Barcelona (left) and Baleares (right). Panel **C** shows the $TE_{i\to j}$ and $ETE_{i\to j}$ from Barcelona to Baleares (left) and from Baleares to Barcelona. The shaded areas in panels **A** and **B** represent the sliding windows for which the TE reaches its maximum value. Transfer Entropy was computed setting $\delta = 7$ days and $q = 1$; for the dynamic TE, we set the length of the sliding window ω to 28 days. $ETE_{i\to j}$ was computed used $n = 500$ shuffles on $X_{ij}(t)$.

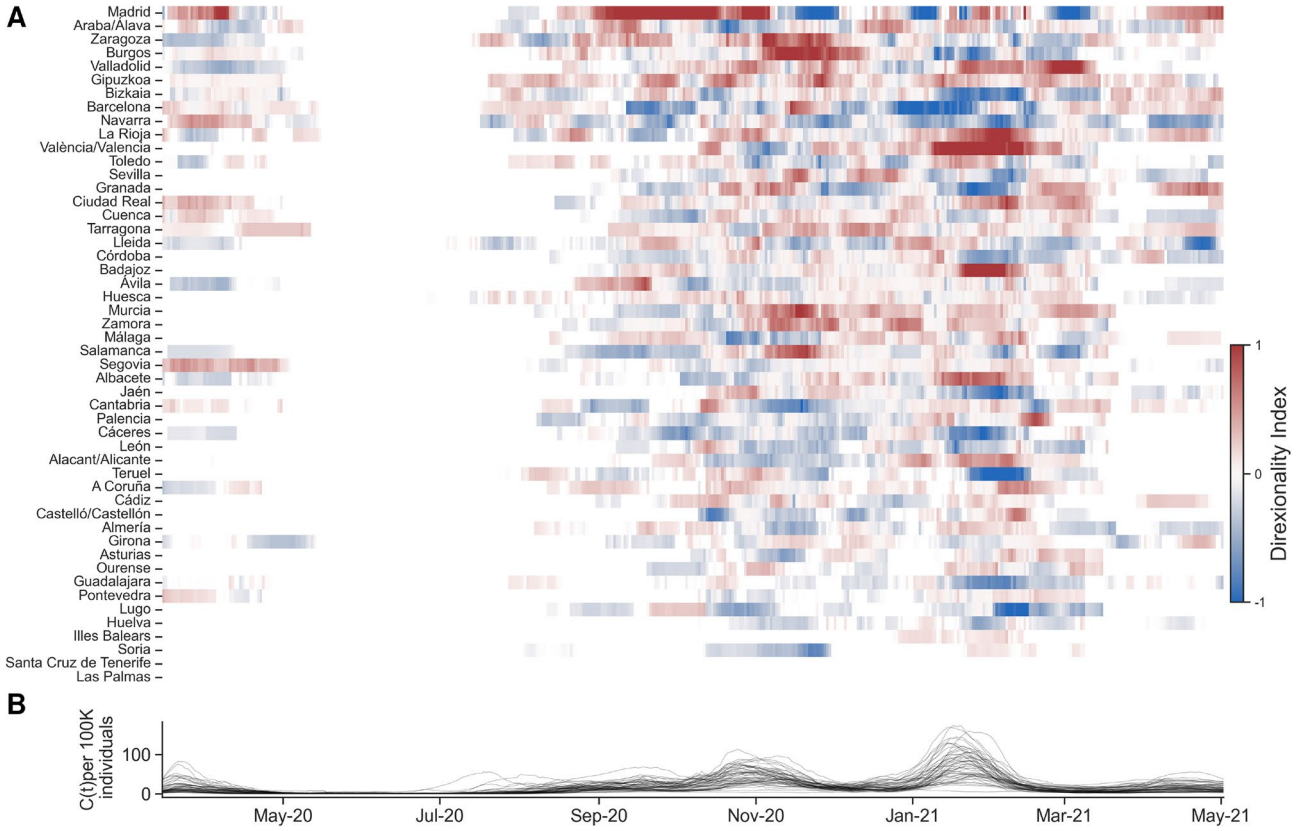

Fig. 5. Dynamics of the COVID19 pandemic in Spain. Panel **A** shows the average normalized directionality index (DI_i) for each province, reflecting the net information flow during the pandemic. Provinces are sorted by their total DI_i values. Transfer entropy ($TE_{i\to j}$) was computed using a sliding window of $\omega = 28$ days, with a step size of $\delta = 7$ days, and embedding dimension $q = 1$. Panel **B** shows the number of daily reported COVID19 cases per 100,000 inhabitants over the same period, with each curve representing a different province. See Supplementary Fig. S14 for individual time series by province.

during the second wave, which we will discuss further in this section. The results also show that, although Madrid was the main driver overall, its role shifted between pandemic waves (Fig. 5A and B). During the second wave, Madrid acted as a strong driver, whereas in the third wave, it was more of a driven region. This aligns with Spain's seasonal mobility patterns, where many people travel out of Madrid during the summer and return during the Christmas season.

Focusing on the beginning of the pandemic, our analysis identifies Álava, along with Madrid, as one of the main drivers of COVID19 spread in Spain. This finding aligns closely with the early pandemic timeline in the country[58]. As reported by El País, a significant superspreading event occurred in late February 2020, when a funeral held in Vitoria (the capital of Álava) became the largest recorded outbreak at that time. More than 60 attendees at the ceremony were infected, marking a major early milestone in the spread of the virus in Spain[59].

Given the likely influence of population movement at the onset of the second COVID19 wave in Spain, we decided to analyze this period in greater detail. At the beginning of Summer, our results show that Lleida started acting as an important driver, probably driving an outbreak of cases in Zaragoza, through Huesca, and later in Madrid. This event is known to have happened due to the arrival of seasonal workers who went to Lleida and Huesca and then got back to Zaragoza and Madrid[60]. In Fig. 5B, the peak of cases observed in July corresponds to Lleida[61]. The spreading pattern of this event can be seen in Fig. 6, which shows all Spanish provinces coloured by daily COVID19 incidence and the significant net flow of information $DI^{+}_{i\to j}$ between them. Our results showed how the local $TE_{i\to j}(t)$ can help uncover spatio-temporal spreading patterns of an epidemic process such as the COVID19 pandemic.

Evaluating the impact of non-pharmaceutical interventions: the case of closing bars in Catalunya

In previous work, we analyzed the policy's impact on population mobility and observed a significant reduction in movements during this period, unlike Madrid, which served as a control case where no such policy was applied. Moreover, the reduction in mobility in Catalunya was found to correlate significantly with a decrease in new infections[4]. In this section, we extended our analysis using the mobility-based TE framework to study the relationship between mobility and the rate of new cases, leveraging population mobility and COVID19 case data reported at a finer spatial resolution (see Materials and Methods). Specifically, TE was calculated at the

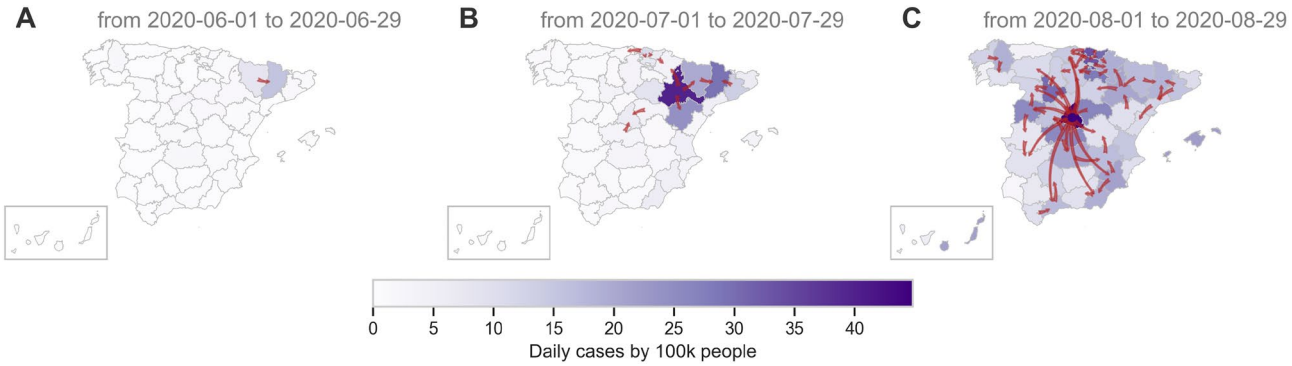

Fig. 6. Transfer of entropy between Spanish provinces during the Summer. The maps show Spanish provinces coloured by the number of daily COVID19 cases by 100k inhabitants. Red arrows indicate the net flow of information $DI^{+}_{i\to j}$ between provinces during 28 days starting at different dates. The base map was retrieved from CNIG https://centrodedescargas.cnig.es/CentroDescargas/ under CC-BY 4.0 Licence https://www.ign.es/resources/licencia/Condiciones_licenciaUso_IGN.pdf.

Basic Health Areas (BHAs) level for Catalunya and Madrid. We first examined the effects of using this higher spatial resolution on previously observed patterns of mobility and disease spread. Subsequently, we applied the mobility-based TE framework to evaluate the specific impact of the policy.

In a previous section (*Measuring Information Flow Between Mobility Patterns and Epidemic Spread*), we demonstrated a significant negative correlation between Transfer Entropy (TE) and geographic distance across pairs of provinces (see Supplementary Fig. S8). In contrast, at the Basic Health Area (BHA) level, while the total number of trips still exhibits a significant negative correlation with distance for both Catalunya and Madrid, the relationship between TE (or ETE) and geographic distance does not show a significant correlation (see Supplementary Figs. S15 and S16). One possible explanation for this discrepancy is that intra-province movement faced fewer restrictions during the pandemic, as such mobility was primarily work-related. In contrast, inter-province movement was more often associated with leisure. Additionally, we analyzed the relationship between the average values of dynamic TE/ETE over time and global TE/ETE in Catalunya and Madrid. We find significant correlations for both cases (see Supplementary Figs. S17 and S18), in agreement with the results observed at the province level (see Supplementary Fig. S12).

In the following, we analyzed the dynamic patterns $R_{ij}(t)$ and $\hat{C}_j(t)$ around the time when the policy of closing bars and restaurants was applied in Catalunya. Two weeks after the application of the policy in Catalunya, the daily incidence $\hat{C}(t)$ and the mobility $M(t)$ for this region changed their trends and started decreasing (see Supplementary Fig. S19A and B). We also calculated the rate of change for the daily incidence and found that a week after the policy was applied, there was an abrupt deceleration in the rate of change of new cases in Catalunya, whereas in Madrid, there were no changes (see Supplementary Fig. S19C). We calculate the correlation between mobility and the rate of change for the daily incidence in Catalunya and Madrid, for the whole of 2020 (Fig. 7A-B), and for the period when the policy was applied in Catalunya (Fig. 7C-D). The results showed a moderate correlation when the whole year was considered with $R = 0.25$ and $R = 0.14$ for Catalunya and Madrid, respectively (Fig. 7A-B). However, when the correlation is calculated throughout the policy we found a very strong correlation between mobility and the rate of change for the daily incidence with an $R = 0.84$ for Catalunya ($p = 5e - 11$), while for Madrid the correlation was close to zero and statistically not significant (Fig. 7C-D).

Finally, we analyzed the $ETE_{ij}(t)$ patterns in Catalunya during the period around the NPI application. As shown in Fig. 8, when the policy was implemented, the average values of both $R_{ij}(t)$ and $\hat{C}_j(t)$ were rising in Catalunya, indicating there was a COVID19 outbreak starting in different BHAs. In contrast, Madrid had already experienced a recent outbreak $R_{ij}(t)$, and $\hat{C}_j(t)$ were gradually declining. In both regions, it could be hypothesized that a reduction in $R_{ij}(t)$ might cause a subsequent decrease in $\hat{C}_j(t)$, potentially leading to an increase in $ETE_{ij}(t)$. However, a significant increase in the average $ETE_{ij}(t)$ was only observed for Catalunya. This suggests that during the NPI period, the reduction in mobility, reflected by a decrease in $R_{ij}(t)$, could be causally linked to the observed decline in $\hat{C}_j(t)$ in many BHAs. In Madrid, by contrast, no significant increase in the average $ETE_{ij}(t)$ was detected, indicating that the decline in $\hat{C}_j(t)$ likely preceded the reduction in $R_{ij}(t)$ and was not directly related, even though the average trends may appear correlated.

A closer examination of the dynamic ETE patterns among BHAs during this period reveals notable shifts in areas on the outskirts of Barcelona. The results also showed that many BHAs in the sanitary regions in the outskirts of Barcelona began exhibiting information transfer during a time that coincided with a decline in the number of new COVID19 cases across several areas (see Supplementary Fig. S20). This aligns with the patterns observed in Fig. 8, suggesting a temporal association between reduced mobility and the observed decrease in case numbers during the NPI period in Catalunya. Taken together, the results discussed in this section demonstrate that dynamic ETE patterns may serve as a useful signal for identifying periods when changes in population mobility were closely associated with shifts in disease transmission dynamics.

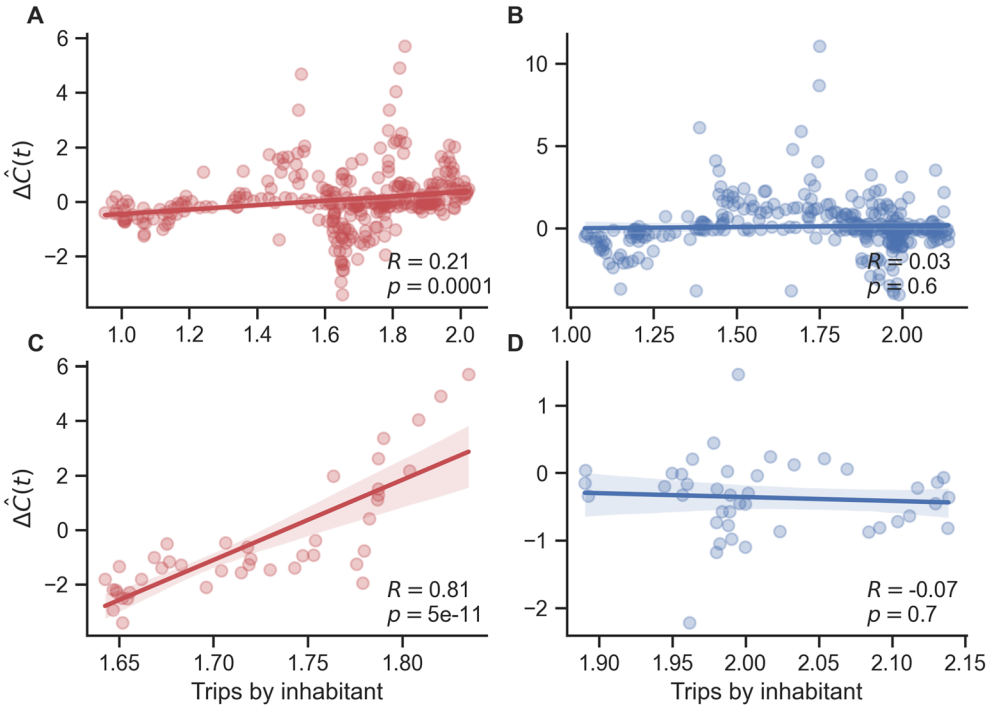

Fig. 7. Correlations between the number of trips $M(t)$ per inhabitant and the rate of change of $\Delta\hat{C}_i(t)$. Panels A and B show the relations between the number of trips $M(t)$ per inhabitant and the rate of change of daily reported new cases $\Delta\hat{C}_i(t)$ for Catalunya and Madrid, respectively, considering the period between March 2020-03-01 and 2021-02-01. Panels C and D show the relations between the number of trips $M(t)$ per inhabitant and the rate of change of $\Delta\hat{C}_j(t)$ for Catalunya and Madrid respectively, considering the period in which the policy of closing bars and restaurants was applied in Catalunya; the period of the application of the policy spanned between 2020-10-16 and 2020-11-26. The rate of change of $\Delta\hat{C}_i(t)$ was calculated using a first-order finite difference approximation.

Recovering disease spreading patterns with mobility-based transfer entropy.

To assess whether our empirical findings could be replicated in a controlled scenario, we simulated an epidemic using a metapopulation SIR model (Fig. 1A-B). Further insight into the simulation approach can be found in Epidemic Simulations. By simulating a simple model, we were able to isolate and examine the mechanisms by which mobility structures shape TE patterns and improve the interpretation of our real-world results. We initiated outbreaks in a single location and evaluated how inter-regional mobility influenced the spread of infection to neighboring areas. We used metapopulation networks of increasing complexity–ranging from a simple linear chain, to a ring, a extended star topology (Supplementary Fig. S21), and finally a model of Spain based on empirical mobility data between provinces.

By focusing on simplified scenarios where the relationship between mobility and transmission dynamics can be explicitly traced, we identified distinct conditions under which significant TE is observed between patches. We present results for the case of the metapopulation model with a extended star-like mobility network previously proposed[62], with the initial outbreak seeded in one of the peripheral nodes (Fig. 1A). The epidemic was simulated for 90 days using parameters representative of COVID19: a basic reproduction number $R_0 = 2.5$ and an average recovery time of $\mu^{-1} = 5$ days (see Materials and Methods for details). In this scenario, infections first peak at the origin node and then propagate sequentially to neighboring and downstream nodes (Fig. 1C).

We first analyzed the global effective transfer entropy between each pair of time series X_{ij} and Y_j, corresponding encoded mobility-associated risk flows from i to j, and the encoded normalized incidence time series of j. Although the underlying mobility network is static, the mobility-associated risk signal $R_{ij}(t)$ is time-dependent, as it combines fixed commuting flows with dynamically changing infection prevalence. As such, it serves as a meaningful, time-varying proxy for potential exposure between regions, allowing TE to detect dynamic influence patterns even when the contact structure remains fixed. The results show that global ETE patterns qualitatively reproduce the structure of the underlying mobility network (Fig. 9B). However, different values of the encoding parameter q led to different TE signatures. When $q = 1.0$, we observed frequent back-transfer of entropy from the target to the source patches. This effect reflects the synchronous timing of infection events across regions in the model, which assumes instantaneous infection and reporting. Since the delay δ between infection and case count is only one day, many regions become synchronized, resulting in reciprocal TE. In contrast, using $q = 0.5$ mitigated this back-transfer effect (Supplementary Fig. S25 and S26). With a lower q, the initial rise in R_{ij} preceded the increase in $\hat{C}_j$ by a greater margin, more closely capturing the signal

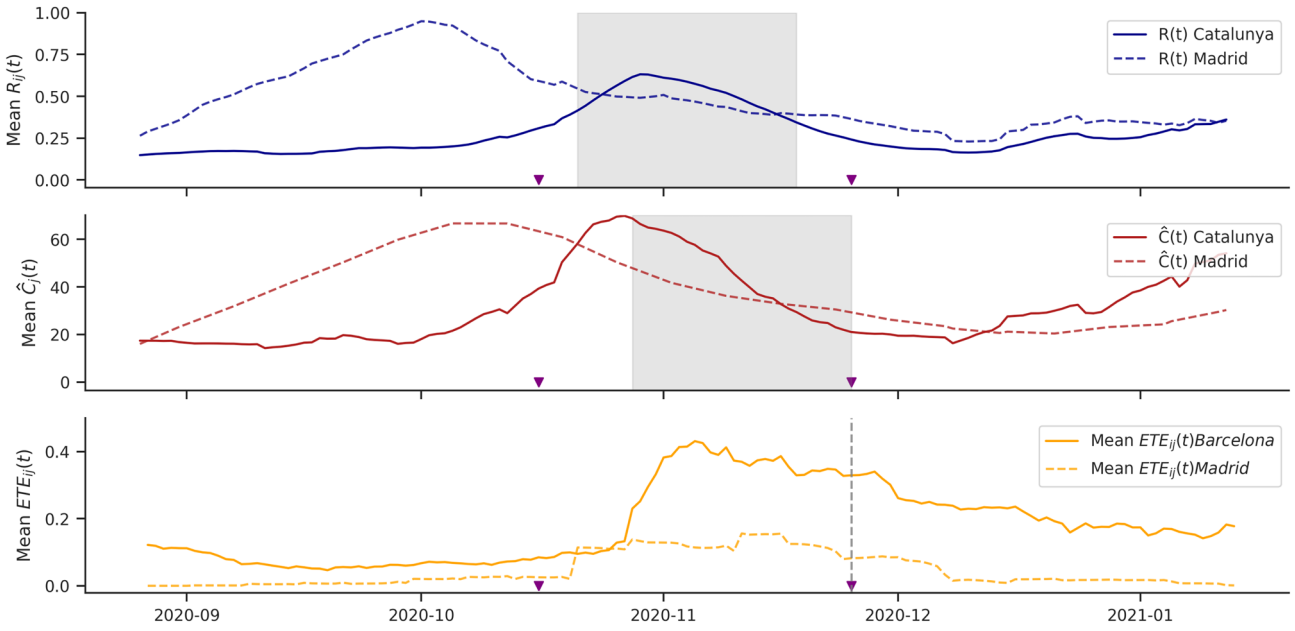

Fig. 8. Impact of the temporary closure of bars and restaurants in Catalunya. The panel shows the evolution over time of the mobility-associated risk $R_{ij}(t)$, the number of new COVID19 cases per 100,000 inhabitants $\hat{C}_j(t)$, and the effective transfer entropy $ETE_{ij}(t)$, all averaged over the basic health areas of Madrid and Catalunya. Purple arrows indicate the start and end dates of the bar and restaurant closure policy in Catalunya (October 16 to November 25, 2020). The grey shaded area represents the sliding window used to compute $TE_{ij}(t)$ and the $ETE_{ij}(t)$ at the end of the intervention period.

between mobility-associated risk and cases. As such, we selected $q = 0.5$ for the subsequent TE analyses, as it yielded results more consistent with the mechanistic transmission pathways.

To evaluate the effect of the mobility on the spreading between regions, we measured the dynamic effective transfer entropy $ETE_{i\to j}(t)$ between every pair of regions without considering the mobility as we did with the real datasets (see Measuring Information Flow Between Mobility Patterns and Epidemic Spread). For each scenario, we calculated the effective TE using the parameters $\omega = 28$ days, $\delta = 1$ days, and $q = 0.5$. We compared the networks of information flow obtained accounting and without accounting for mobility. We found that when the effective TE is measured considering the mobility, the influence network recovers the structure of the metapopulation through which the epidemic spreads (Fig. 9D). However, if the effective $TE_{i\to j}$ is measured only considering the time-series of cases, the network of influences contains many indirect influences (Supplementary Fig. S22). For instance, on day 36, we can observe that the network of influences inferred without considering mobility shows that regions *M*3 and *M*9 influence distant regions *M*8 and *M*4, respectively, which are not connected through the mobility network. On the contrary, when the effective $TE_{i\to j}$ is measured considering the cases at the source region together with the mobility patterns, we can see how the influence network changes but is always constrained by the structure of the metapopulation. These results are in agreement with those found when analysing the information flow using real data for Spain (see section Measuring Information Flow Between Mobility Patterns and Epidemic Spread). Furthermore, similar results were found for other commuting network structures, including the simple chain and the ring (see Supplementary Fig. S23 and S24). Hence, these results show that if mobility is not considered, the TE measures information flow between regions that are not connected through daily commuters, indicating spurious interactions.

Figure 9D illustrates two distinct scenarios where dynamic entropy transfer between patches is evident. At the beginning of the epidemic, an increase in risk in M2 precedes and likely signals subsequent rises in cases within M1, M3, and M9, highlighting an entropy transfer originating from M2 around D30. Later, around D54, a different pattern emerges: a decrease in risk in M2 appears to anticipate similar declines in other patches, showing another instance of entropy transfer from M2 to M1, M3, and M9. These results show the adaptability of the dynamic ETE, which responds to various epidemic phases, revealing a complex and evolving influence network (Fig. 9C and D).

We also conducted a SIR-based simulation using real population data and inter-province mobility matrices from Spain during the COVID19 pandemic. Because the SIR model does not incorporate dynamic mobility, we considered two static scenarios: one representing pre-lockdown conditions and the other post-lockdown. We calculated both the dynamic and the global ETE and DI between the provinces as described in Epidemic Simulations in the Material and Methods. The results show a marked decline in total global ETE in the lockdown scenario compared to the pre-lockdown case (see Supplementary Fig. S27), indicating a significant reduction in inter-provincial information flow due to restricted mobility. In addition to this overall decrease, we observed notable shifts in the roles of specific provinces. For example, Zaragoza and Toledo exhibited a substantial drop in incoming ETE during the lockdown, whereas others, like Burgos, became more prominent receivers of

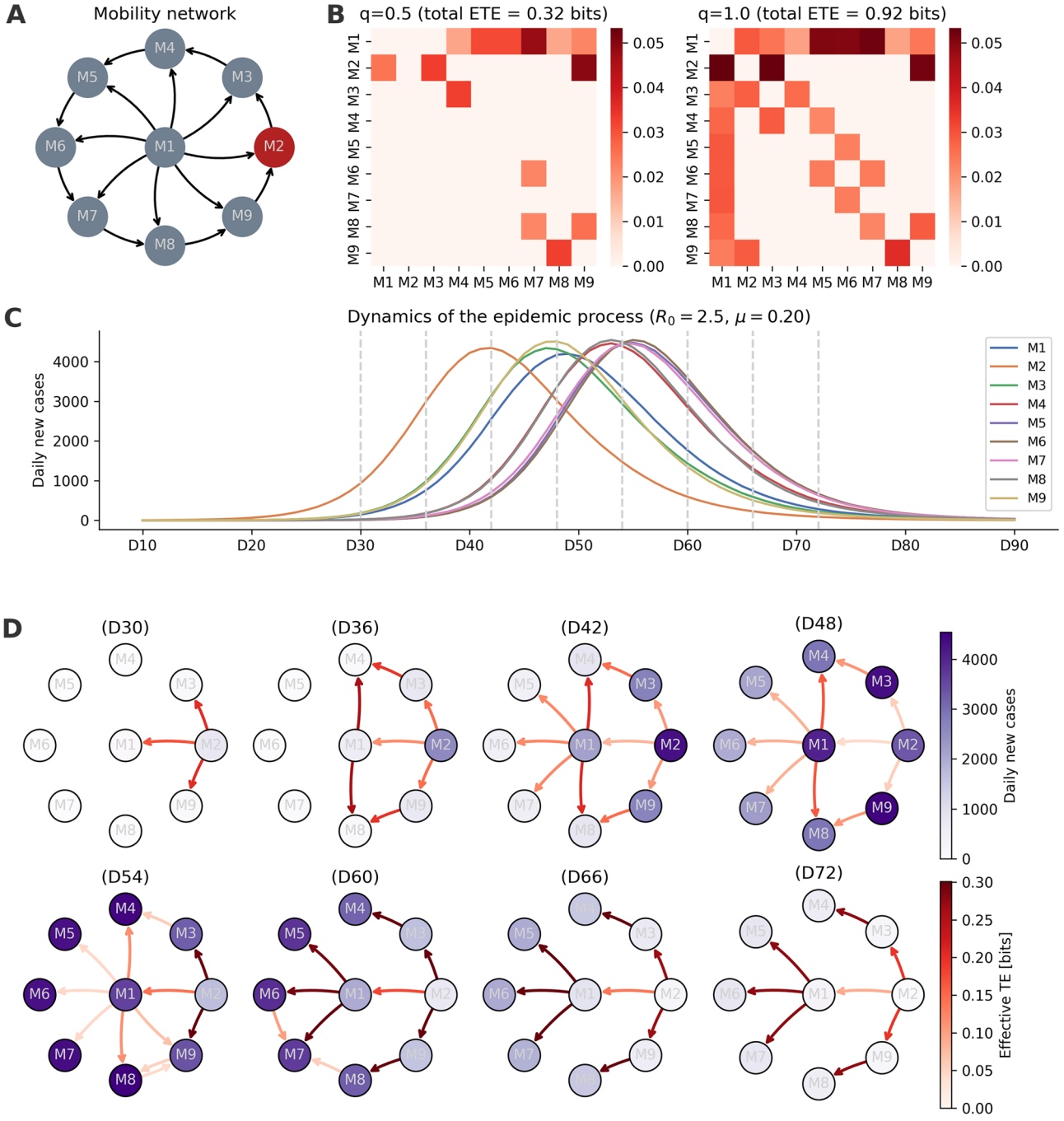

Fig. 9. Application of Transfer Entropy to infer a causal relationship between population mobility and the spread of an epidemic in a metapopulation SIR model. Panel A shows the structure of the commuting network, each node has a population of 100, 000 inhabitants, and the number of daily commuters is 5000 in each arrow. Panel B shows the global ETE between each pair of nodes for two different values of the embedding parameter, $q = 0.5$ and $q = 1.0$. Panel C shows the average over 1, 000 simulations of the normalized number of cases as a function of time (in days) obtained using an infection rate $R_0 = 2.5$ and a recovery rate $\mu^{-1} = 5$. Panel D shows the net value of $ETE_{i \to j}$ transferred at different days throughout the epidemic and the number of daily new cases in each node. The selected days shown in D are indicated in C as dashed gray lines. ETE parameters $\delta = 1$ days, $\omega = 28$ days and $q = 0.5$ were used. The red node in A is the region where the epidemic started, and the initial number of infected individuals was set to $I_0 = 3$.

information flow. These changes are mirrored in the patterns of dynamic DI over time (Supplementary Figs. S28 and S29), highlighting how mobility restrictions reshaped the epidemic influence network. Altogether, these findings reinforce the relevance of mobility data in uncovering potential causal pathways of epidemic spread and emphasize the value of dynamic information-theoretic measures in characterizing evolving transmission dynamics.

We also examined the relationship between inter-provincial distances and $TE_{i\to j}$ in both the pre-lockdown and lockdown scenarios. A significant negative correlation was observed in both cases–$R = -0.19$ ($p = 10^{-7}$) pre-lockdown and $R = -0.18$ ($p = 10^{-5}$) during lockdown–indicating that transfer entropy tends to decrease with increasing distance between provinces (Supplementary Fig. S30A and B). These findings are consistent with those obtained from the real-world data (Supplementary Fig. S8). To test whether this spatial dependency arises from mobility structure or is an artifact of the simulation, we conducted additional experiments under a homogeneous mixing assumption. In these simulations, all outgoing trips from each province were redistributed uniformly across other regions, thereby removing the natural correlation between mobility and geographic distance. Under this setup, the correlation between inter-provincial distances and $TE_{i\to j}$ became statistically non-significant (Supplementary Fig. S30C), supporting the idea that the spatial pattern of information flow is driven by structured mobility patterns rather than random mixing.

Finally, we compared the global ETE values obtained from real-world data with those derived from simulations. The overall Pearson correlation was $R = 0.255$ with a highly significant p-value of 10^{-38} (Supplementary Fig. S31). While modest, this correlation reflects the fundamental differences between simulated and real epidemic dynamics. In the simulations, mobility is static, and each province experiences only a single epidemic wave. In contrast, the real-world epidemic involves dynamic changes in mobility and multiple waves of transmission. Despite these simplifications, the simulations capture essential features of the observed TE patterns, providing valuable insights into the mechanisms underlying epidemic spread. Taken together, these results confirm that the observed TE signals are not merely statistical artifacts of noisy real-world data, but can mechanistically arise from structured, mobility-driven transmission processes. The use of synthetic outbreaks enabled us to isolate and scrutinize the causal pathways of epidemic spread, demonstrating how TE can capture influence patterns aligned with known transmission routes. By comparing networks with and without incorporating mobility data, we identified the conditions under which spurious or indirect interactions may appear. This reinforces the validity of our methodological framework and highlights the value of mobility-informed TE analysis in detecting meaningful, time-resolved influence patterns in epidemic dynamics.

Conclusions

In this study, we introduced a mobility-informed Transfer Entropy (TE) framework to explore the directional relationships between population mobility and infectious disease spread. By applying this methodology to both synthetic epidemic simulations and real-world COVID-19 data, we demonstrated that incorporating explicit mobility patterns helps reduce spurious associations and enhances the detection of potentially meaningful information flows within complex epidemiological systems.

A key challenge in inferring causal relationships in such systems is disentangling the impact of mobility from confounding factors such as non-pharmaceutical interventions, behavioural adaptations, and reporting delays. Our findings showed that conventional TE analyses based solely on incidence time series are prone to false positives, capturing apparent dependencies that do not reflect true transmission pathways. By integrating mobility data coupled with the epidemic state of the source region, our approach reduces the likelihood of such artifacts and offers a more nuanced understanding of disease dynamics–particularly in identifying regions that may act as transmission hubs.

Nevertheless, we emphasize that TE, while useful for detecting directional dependencies, does not by itself establish causality. Rather, it offers a probabilistic measure of information transfer that must be interpreted in light of epidemiological context and other sources of evidence. Our framework addresses some limitations of conventional TE applications, but it remains subject to challenges such as unobserved confounders and temporal aggregation effects. The interpretation of TE signals–whether they reflect propagation, suppression, or stochastic coincidence–requires careful domain-specific analysis.

Despite these caveats, the temporal and spatial patterns we identified were consistent with known features of the COVID-19 epidemic in Spain, lending support to the utility of our method. By integrating mobility data into information-theoretic analysis, our framework contributes to the broader development of data-driven tools for real-time epidemic assessment. Future work should aim to combine such approaches with complementary causal inference methods to strengthen the identification of transmission drivers and inform more targeted public health interventions.

Data availability

The datasets generated and/or analysed during the current study are available in the Zenodo repository, https://doi.org/10.5281/zenodo.12207212, https://zenodo.org/communities/flow-maps/.

Received: 18 February 2025; Accepted: 21 August 2025
Published online: 26 August 2025

References

1. Pascal P. K., et al. Enhancing global preparedness during an ongoing pandemic from partial and noisy data. *PNAS Nexus*, **2**(6), pgad192, (2023).
2. Oliver, N. et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**(23), eabc0764 (2020).

3. Perra, N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Phys. Rep.* **913**, 1–52 (2021).
4. Smith, M, Ponce-de-Leon, M & Valencia, Al. Evaluating the policy of closing bars and restaurants in Cataluña and its effects on mobility and COVID19 incidence. *Sci. Rep.* **12**(1), 9132 (2022).
5. Mitjà, O. et al. Experts' request to the Spanish Government: Move Spain towards complete lockdown. *The Lancet* **395**(10231), 1193–1194 (2020).
6. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**(7196), 779–782 (2008).
7. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A Tale of many cities: Universal patterns in human urban mobility. *PLoS ONE* **7**(5), e37027 (2012).
8. Louail, T. et al. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**(1), 5276 (2014).
9. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci.* **106**(36), 15274–15278 (2009).
10. Deville, P. et al. Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci.* **111**(45), 15888–15893 (2014).
11. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**(6264), 1073–1076 (2015).
12. Yabe, T. et al. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **94**, 101777 (2022).
13. Tizzoni, M. et al. On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **10**(7), e1003716 (2014).
14. Schlosser, F. et al. Covid-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc. Natl. Acad. Sci.* **117**(52), 32883–32890 (2020).
15. Arenas, A. et al. Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions. *Phys. Rev. X* **10**(4), 041055 (2020).
16. Ferrari, A. et al. Simulating SARS-CoV-2 epidemics by region-specific variables and modeling contact tracing app containment. *npj Digit. Med.* **4**(1), 1–8 (2021).
17. Balcan, D. et al. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *J. Comput. Sci.* **1**(3), 132–145 (2010).
18. Mazzoli, M. et al. Interplay between mobility, multi-seeding and lockdowns shapes COVID-19 local impact. *PLoS Comput. Biol.* **17**(10), e1009326 (2021).
19. Perofsky, A. C. et al. Impacts of human mobility on the citywide transmission dynamics of 18 respiratory viruses in pre- and post-COVID-19 pandemic years. *Nat. Commun.* **15**(1), 4164 (2024).
20. Grantz, K. H. et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat. Commun.* **11**(1), 4961 (2020).
21. Delussu, F., Tizzoni, M. & Gauvin, L. The limits of human mobility traces to predict the spread of covid-19: A transfer entropy approach. *PNAS nexus* **2**(10), 302 (2023).
22. Brauner, J. M. et al. Inferring the effectiveness of government interventions against COVID-19. *Science* **371**(6531), eabd9338 (2021).
23. Alamo, T., Reina, D. G., Mammarella, M. & Abella, A. Covid-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic. *Electronics* **9**(5), 827 (2020).
24. Harrison, P. W. et al. The COVID-19 Data Portal: Accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. *Nucleic Acids Res.* **49**(W1), W619–W623 (2021).
25. Latif, S. et al. Leveraging data science to combat COVID-19: A comprehensive review. *IEEE Trans. Artif. Intell.* **1**(1), 85–103 (2020).
26. Shuja, J., Alanazi, E., Alasmary, W. & Alashaikh, A. COVID-19 open source data sets: A comprehensive survey. *Appl. Intell.* **51**(3), 1296–1325 (2021).
27. Leon, Miguel P.-de et al. Covid-19 flow-maps an open geographic information system on covid-19 and human mobility for spain. *Sci. data* **8**(1), 1–16 (2021).
28. Judea, P. *Causality*. Cambridge University Press, Cambridge, 2 edition, 2009.
29. Alex, E. Y. and Wenying, S.. Data-driven causal analysis of observational biological time series. *eLife* **11**, e72518, (2022).
30. Ashley, R. C., Sarah K. H., Elaine, L., Daniel, M., and Joshua, S. W. A Primer for Microbiome Time-Series Analysis. *Frontiers in Genetics*, **11**, (2020).
31. Shojaie, A. & Fox, E. B. Granger Causality: A Review and Recent Advances. *Annu. Rev. Stat. Appl.* **9**(1), 289–319 (2022).
32. Clive, W. J. G. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, (1969).
33. Granger, C. W. J. Testing for causality: A personal viewpoint. *J. Econ. Dyn. Control* **2**, 329–352 (1980).
34. Galea, S., Riddle, M. & Kaplan, G. A. Causal thinking and complex system approaches in epidemiology. *Int. J. Epidemiol.* **39**(1), 97–106 (2010).
35. Sugihara, G. et al. etecting causality in complex ecosystems. *Science* **338**(6106), 496–500 (2012).
36. Daniel, H., Erik, L., Maik, S., & Klaus, R. P. Topological Causality in Dynamical Systems. *Phys. Rev. Lett.* **119**(9), 098301, (2017).
37. Vicente, R., Wibral, M., Lindner, M. & Pipa, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **30**(1), 45–67 (2011).
38. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**(2), 461 (2000).
39. Barnett, L., Barrett, A. B. & Seth, A. K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Phys. Rev. Lett.* **103**(23), 238701 (2009).
40. Lungarella, M., Ishiguro, K., Kuniyoshi, Y. & Otsu, N. Methods for quantifying the causal structure of bivariate time series. *Int. J. Bifurcation Chaos* **17**(03), 903–921 (2007).
41. Fatimah Abdul Razak and Henrik Jeldtoft Jensen. Quantifying 'Causality' in Complex Systems: Understanding Transfer Entropy. *PLOS ONE*, 9(6):e99462, June 2014.
42. Porfiri, M. & Ruiz Marín, M. Inference of time-varying networks through transfer entropy, the case of a Boolean network model. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**(10), 103123 (2018).
43. Thomas, D. & Franziska, J. P. Using transfer entropy to measure information flows between financial markets. *Stud. Nonlinear Dyn. Econom.* **17**(1), 85–102 (2013).
44. Borge-Holthoefer, J. et al. The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Sci. Adv.* **2**(4), e1501158 (2016).
45. Kissler, S. M., Viboud, C., Grenfell, B. T. & Gog, J. R. Symbolic transfer entropy reveals the age structure of pandemic influenza transmission from high-volume influenza-like illness data. *J. R. Soc. Interface* **17**(164), 20190628 (2020).
46. [dataset], Miguel, P.-de L. et al. COVID19 Flow-Maps Daily Cases Dataset. *Zenodo* https://doi.org/10.5281/zenodo.4634868, (2021).
47. [dataset], Miguel Ponce-de, L et al. COVID19 Flow-Maps Population Dataset (Maestra 2). *Zenodo* https://doi.org/10.5281/zenodo.4635258 (2021).
48. [dataset], Miguel Ponce-de, L et al. COVID19 Flow-Maps Daily-Mobility Dataset (Maestra 1). *Zenodo* https://doi.org/10.5281/zenodo.4634895 (2021).
49. [dataset], Miguel Ponce-de, L. et al. COVID19 Flow-Maps GeoLayers Dataset. *Zenodo* https://doi.org/10.5281/zenodo.4634662, (2021).

50. Hamada, S. B., Hongru, D., Maximilian, M., Ensheng D., Marietta, M. S. & Lauren, M. G. Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *Lancet Infect. Dis.* **20**(11), 1247–1254, (2020).
51. Granell, C., Gómez, S., Gómez-Gardeñes, J. & Arenas, A. Probabilistic discrete-time models for spreading processes in complex networks: A review. *Ann. Phys.* **536**(10), 2400078 (2024).
52. Feng, M., Cai, S.-M., Tang, M. & Lai, Y.-C. Equivalence and its invalidation between non-markovian and markovian spreading dynamics on complex networks. *Nat. Commun.* **10**(1), 3748 (2019).
53. Nicoló, A. C. & Paolo, P. Effective transfer entropy to measure information flows in credit markets. *Stat. Methods Appl.* **31**(4), 729–757, (2022).
54. Lauer, S. A. et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**(9), 577–582 (2020).
55. Cheng, Ch. et al. The incubation period of COVID-19: A global meta-analysis of 53 studies and a Chinese observation study of 11 545 patients. *Infect. Dis. Poverty* **10**(05), 1–13 (2021).
56. Marschinski, R. & Kantz, H. Analysing the information flow between financial time series. *Eur. Phys. J. B - Condensed Matter and Complex Systems* **30**(2), 275–281 (2002).
57. D'Arienzo, M. & Coniglio, A. Assessment of the sars-cov-2 basic reproduction number, r0, based on the early phase of covid-19 outbreak in italy. *Biosaf. Health* **2**(2), 57–59 (2020).
58. Oriol, G. Álava ya sufre una mayor incidencia del virus que Lombardía al decretarse la cuarentena. https://elpais.com/sociedad/2020-03-09/alava-ya-sufre-una-mayor-incidencia-del-virus-que-lombardia-al-decretarse-la-cuarentena.html,(2020).
59. Oriol, G. Un funeral en Vitoria causó más de 60 infectados. https://elpais.com/sociedad/2020-03-06/mas-de-60-personas-se-contagiaron-a-la-vez-en-un-funeral-en-vitoria.html, (2020).
60. Jessica, M. E. & Virginia, L. El brote entre temporeros de Aragón deja una cuarta comarca en la fase 2. https://elpais.com/sociedad/2020-06-23/el-brote-entre-temporeros-de-aragon-deja-una-cuarta-comarca-en-la-fase-2.html (2020)
61. Coronavirus: Aíslan cincuenta temporeros en un pueblo de Lleida por un positivo. https://www.lavanguardia.com/local/lleida/20200603/481590717165/aislan-cincuenta-temporeros-seros-lleida-positivo-covid-19-coronavirus.html, (2020).
62. Gómez-Gardeñes, J., Soriano-Paños, D. & Arenas, A. Critical regimes driven by recurrent mobility patterns of reaction-diffusion processes in networks. *Nat. Phys.* **14**(4), 391–395 (2018).

Acknowledgements

This work has received funding from the Horizon 2020 project CREXDATA (ID: 101092749) and the MePre-CiSa project of the UNICO I+D Cloud program, co-financed by the Ministry for Digital Transformation and of Civil Service and the EU-Next Generation EU as financing entities, within the framework of the PRTR and the MRR. CP was supported by the fellowship "Juan de la Cierva - Formación" from the Ministry of Education and Science of Spain (ID: FJC2021-046655-I). AA acknowledges support from the Spanish Ministerio de Ciencia e Innovación (PID2021-128005NB-C21), and the Joint Appointment Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Author contributions

M.P., C.P., A.A. and A.V. contributed to the conception of the work. M.P. and C.P. contributed to the acquisition, analysis, or interpretation of data. M.P. implemented the code used in the work. M.P. and C.P. have drafted the work or substantively revised it. All authors have approved the submitted version and are personally accountable for their own contributions.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-17218-4.

Correspondence and requests for materials should be addressed to M.P.-d.-L.