# Associations between mobility patterns and COVID-19 deaths during the pandemic: A network structure and rank propagation modelling approach

Furxhi Irini [a,b], Arash Negahdari Kia [b,*], Darren Shannon [b], Tim Jannusch [b,c], Finbarr Murphy [b,a], Barry Sheehan [b]

[a] Transgero Limited, Newcastle West, Limerick, Ireland
[b] Kemmy Business School, University of Limerick, Ireland
[c] Institut for Insurance Studies, TH, Köln, Germany

## ABSTRACT

*Background:* From February 2020, both urban and rural Ireland witnessed the rapid proliferation of the COVID-19 disease throughout its counties. During this period, the national COVID-19 responses included stay-at-home directives issued by the state, subject to varying levels of enforcement.

*Methods:* In this paper, we present a new method to assess and rank the causes of Ireland COVID-19 deaths as it relates to mobility activities within each county provided by Google while taking into consideration the epidemiological confirmed positive cases reported per county. We used a network structure and rank propagation modelling approach using Personalised PageRank to reveal the importance of each mobility category linked to cases and deaths. Then a novel feature-selection method using relative prominent factors finds important features related to each county's death. Finally, we clustered the counties based on features selected with the network results using a customised network clustering algorithm for the research problem.

*Findings:* Our analysis reveals that the most important mobility trend categories that exhibit the strongest association to COVID-19 cases and deaths include *retail and recreation* and *workplaces*. This is the first time a network structure and rank propagation modelling approach has been used to link COVID-19 data to mobility patterns. The infection determinants landscape illustrated by the network results aligns soundly with county socio-economic and demographic features. The novel feature selection and clustering method presented clusters useful to policymakers, managers of the health sector, politicians and even sociologists. Finally, each county has a different impact on the national total.

## 1. Introduction

As of January 16, 2021, more than two million people have died from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), widely known as COVID-19 [1]. The new infectious disease emerged in Wuhan (Hubei Province, China) and was first discovered as a cluster of pneumonia cases of unknown cause in December 2019 [2]. Since then, the disease spread rapidly throughout the world, obliging the WHO to declare COVID-19 as a global pandemic in March 2020 [3]. To this date, COVID-19 has affected more than 200 countries and territories, with western Europe particularly severely impacted [1].

Human mobility is among the key drivers fostering the rapid spread of disease [4,5]. Millions of people travel and commute every day within

and across countries, regions and cities. For this reason, local and national governments have implemented non-pharmaceutical interventions (NPIs) as a strategy to slow down and prevent COVID19 transmission [6–8]. Hale, Petherick et al. [9] studied the variation of government responses to COVID-19. Typical measures include: stay at home orders, travel bans, cancellations of public gatherings, school closures and other interventions to contain the virus's spread. In addition to these individual measures, China and Italy were among the first to enact a national-wide lockdown to curb the infection curve [10].

According to Perumal, Curran et al. [11]; COVID-19 was first discovered in Ireland on the February 17, 2020. A middle-aged lady returned from Northern Italy and suffered from general symptoms such as cough, followed by sweats, fever, nausea or chest pain. Nine days

later, on February 26, 2020, SARS-Coronavirus-2 RNA was detected. Since then, the total number of people infected with COVID-19 in Ireland increased to 147,613, with a 14-day case notification rate per 100,000 inhabitants of more than 1200 in week one of 2021 [12]. The latest numbers from the European Centre for Diseases Prevention and Control revealed that more than 2300 people have already died from the disease in Ireland. Among those, 80-year-old or older age groups continue to be impacted the hardest by the virus [13]. Throughout the pandemic, a pressing issue has been on preventing excess pressure on the health system and deciding on the optimum level of social restrictions to prevent the spread of COVID-19 while maintaining a high level of social utility. Quick Interventions using machine learning methods on COVID19 datasets are very important to reduce the spread speed [14]. To that end, this research adds to the current debate and provides a better understanding of the inter-relationship between the changes of mobility patterns and the epidemiological confirmed COVID-19 cases per county with the total number of deaths.

Coupled with confirmed COVID-19 cases and the total numbers of deaths from Ireland, we analyse publicly available geo-located smart-phone data provided by Google COVID-19 Mobility Reports (GCMR). This data can be used to compare pre-pandemic (baseline) to activity throughout the pandemic to better understand mobility patterns by inferring the location visited by individuals. A particular advantage of the GCMR is that the android operating system has more than a 57% market share among smartphone users, the largest market share in Ireland [15]. For this reason, the provided data likely captures the movements of more individuals than any other provider.

## 2. Literature review

Google community mobility data have been previously used in many other studies [16–21] and highlighted as one of the best data sources to analyse mobility patterns [22]. Huynh [18] examined the role of the cultural dimension in practising social distancing across the world by associating Google data with cultural factors. Tamagusko and Ferreira [21] used Google data for the Portuguese population's mobility patterns during the pandemic to understand the impact of social distance measures on disseminating COVID-19.

Diverse algorithms have been exploited on COVID-19 data, such as Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbours (kNN), Neural Networks (NN), Decision Trees (DTs). These tools have been used to predict various COVID-19 outcomes, such as the number of confirmed positive cases [23], as well as drug development, false news prediction, and vaccine discovery [24]. Bryant and Elofsson [16,17] built a Bayesian model to estimate the number of deaths on a given day based on changes in the basic reproductive number (R0) as a result of changes in mobility patterns. However, whereas Bryant and Elofsson [16,17] compute the importance of each category for eleven countries, in this study, we determine the importance of mobility factors within the counties of a single country. In addition, Ilin, Annan-Phan et al. [19] showed that data made available by Google, Facebook, and other providers could be used to evaluate the effectiveness of non-pharmaceutical interventions and forecast the spread of COVID-19 b y using statistical models. Sulyok and Walker [20]; using Google data for various countries, suggested that data related to community mobility are of use for COVID-19 modelling, using Bayesian Information Criteria as a model selection method. These sources of data have been highlighted as one of the best data sources to analyse mobility patterns [22]. Nevertheless, to our knowledge, this data has not been used in an Irish context, nor has it been used as a means of linking mobility patterns with COVID-19 transmission data using a rank propagation approach.

Health, societal, and economic factors have previously formed discussions in the literature regarding the spread of COVID-19. Bloemers and Montesanti [25] examined common keywords associated with industrial workers in Vietnam who contracted COVID-19 as a means of proposing preventative solutions. They identified that underlying health problems, combined with workers' perceived economic vulnerability, were associated with a high risk of spread. Ronchi and Lovreglio [26] propose an *ex-post* simulation framework to assess viral transmission patterns against confinement and interaction dynamics within buildings. Their research allows for appropriate parameters to be put in place in order to limit the spread of COVID-19 or future pandemics. Mohammadi, Chowdhury et al. [27] propose a similar approach, from the perspective of pedestrians interacting on sidewalks rather than in buildings.

The proliferation of COVID-19 has severely impeded the construction sector. Araya [28] took advantage of agent-based modelling (ABM) to simulate the spread of disease among construction workers to allow the successful realisation of construction projects throughout the current pandemic. One advantage of ABM is that it illustrates complex systems that consist of individual elements or agents. Rules that govern interactions and behaviours are defined to model the agents in a construction project. Araya [28] categorised activities as low, medium, and high risk for workers in their model. Through the heterogeneity of the agents and the possibility of interacting within the same environment, the system behaviour appears during simulations. Their results indicated that a spread of COVID-19 in a construction project may reduce the workforce by up to 90% in high-risk environments.

Bar-Yossef and Mashiach [29]; Barr [30] added to the scientific discussion about COVID-19 evidence that crowded environments with possibly infected people significantly increase infection risk. Accordingly, high numbers of people in queues and stores produce an environment with an elevated risk of contagion. In an attempt to protect both consumers and workers while securing cost-effectiveness, Perlman and Yechiali [31] constructed and analysed two non-classical multi-server queueing models. In this context, the researchers constructed a measure to estimate the mean infection risk for a customer, proportional to the second factorial moment of the number of clients who congregate in a space. Finally, a game-theoretic model has been employed to investigate equilibrium strategies regarding capacity and number of workers in the store. By their model, Perlman and Yechiali [31] allow the protection of customers and the workforce in a store while at the same time cutting down the related costs of the safety measures.

In conjunction with the retail sector, the latest research of [32] argues that the willingness of shoppers to return to shopping depends on their feeling of safety, which is strongly related to measures like providing hand sanitiser, limiting the number of people in stores and monitoring social distancing. With the aim to solve the issue of social distancing [33], exploit theoretical knowledge in different areas of research such as crowd science [34] or operational research of ergonomics [35]. With this theoretical foundation, the researchers developed an extensive method to determine the minimum amount of space an individual needs for social distancing in closed and open commercial spaces. Both static (e.g. queueing) and dynamic (e.g. walk freely) environments are considered in their model.

[36] find that people are more risk-averse when information is positively framed and *vice versa*. They also that high emotional people are more willing to comply with preventive health behaviors when information is framed positively (vs. negatively) [37]. show that human and organisational factors (HOFs) play an important role in preventing and controlling epidemics (PCE). They converted a classification system into a Bayesian Network to analyse the COVID-19 outbreak in China. They, therefore, provide a risk assessment model for epidemics and pandemics. Santamaria, Sermi et al. [38] produce a mobility indicator derived from anonymised and aggregated mobile positioning data. This indicator captures information on EU mobility patterns. That relates to our research in a way that the indicator is used to study the impact of COVID-19 confinement measure on mobility in Europe. They go on to show that these confinement measures can explain a large proportion of the change in mobility patterns.

These prior studies highlight complex interrelationships among different causes of viral transmissions and death-rate fluctuations that

have been hitherto unconsidered as part of a single methodological network. One modelling approach that addresses this issue is the use of a network structure. The modelled data and concepts from data can be thought of as nodes and links between nodes in network structures. The influence each node has on other nodes can be calculated with label propagation algorithms [39,40]. A famous label propagation algorithm is the Personalised PageRank (PPR) presented by Refs. [41–43].

However, this methodology has not been applied before in the case of COVID-19 statistical data and changes in mobility patterns. Within the data set, the rank propagation algorithm, PPR, is used to investigate the importance of different causes in the values of Ireland's confirmed death time series (the importance of a node from the perspective of another). PPR has been used in many different practical applications, such as offering Twitter user recommendations of whom to follow; in graph-theoretic problems such as community detection [44]. As part of the modelling approach, we calculated an individual ranking PPR score per mobility trend and did this per each Ireland county. These ranking scores are linked with mobility patterns, confirmed cases and deaths. The PPR scores are then normalised and divided by the count of categories to be transformed into relative prominent factors and absolute importance measures. This is a representative approach since all network nodes are valued simply by occurrence [41]. Once the scores have been estimated, hierarchical clustering is performed on the counties based on the results from the PPR. The concept of coupling PPR with clustering methodologies has been demonstrated before [45–47].

In the following sections, we describe the data used and the methodological approach applied to measure the effect of mobility patterns and confirmed cases per county on the total number of country's deaths throughout the COVID-19 pandemic. Afterwards, we present our results and argue that *retail and recreation* and *workplaces* appear systematically as the most critical attributes across all counties. We stress the significance of assessing the mobility impact at the subnational level. We conclude the paper with a short discussion of this research's main findings and highlight the limitations of our approach.

## 3. Materials and methods

In this section, we introduce the data collection and the methodology that we use in our article. Specifically, we first describe the data collection (section 3.1) and pre-processing analysis (section 3.2), and secondly, the algorithmic approach we employ to estimate inter-relationships among the data (section 3.3). Finally, we describe the attribute importance analysis and the clustering of our results (section 3.4).

### 3.1. Data collection

#### 3.1.1. COVID-19 community Mobility Reports

The freely available Google LLC [48] COVID-19 Community Mobility Reports (GCMR) were extracted, which offers an approximation to the changes in mobility due to different social distancing measures. The GCMR accounts for the percentage changes in Google Maps users' mobility compared to a baseline period before the pandemic (from January 3 to February 6, 2020) in various categories. The movement change trend categories include *grocery and pharmacy, parks, retail and recreation, transit stations, workplaces* and *residential* (Table 1). By calculating a set of seven baseline weekdays using the median value for each individual weekday during the 5-week baseline period, the data accounts for weekly movement seasonality. For any given data date, the daily relative change is valued as the percentage change with respect to the corresponding baseline weekday.GCMR for Ireland provides county-level information. Ireland is divided into twenty-six counties: Galway, Leitrim, Mayo, Roscommon and Sligo (Connacht province), Carlow, Dublin, Kildare, Kilkenny, Laois, Longford, Louth, Meath, Offaly, Westmeath, Wexford and Wicklow (Leinster province), Clare, Cork, Kerry, Limerick, Tipperary and Waterford (Munster province) and

**Table 1**
Granular definition of mobility patterns, according to Google LLC [48]].

| Mobility Pattern | Definition | Category |
| --- | --- | --- |
| Grocery & Pharmacy | Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies. | G |
| Parks | Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens. | P |
| Transit stations | Mobility trends for places like public transport hubs such as subway, bus, and train stations. | T |
| Retail & recreation | Mobility trends for places like restaurants, cafes, shopping centres, theme parks, museums, libraries, and movie theatres. | R |
| Workplaces | Mobility trends for places of work. | W |

Cavan, Donegal, and Monaghan (Ulster province).

#### 3.1.2. COVID-19 confirmed cases and deaths

Epidemiological data from Ireland's national statistical office, the Central Statistics Office (CSO), which includes weekly number data on confirmed cases and deaths for each IE county (26 counties), were extracted for the period 05-03-2020 to 07-12-2020.[1]

Each weekly data was extracted from a series of published information bulletins produced by the CSO -COVID-19 Insight Bulletins: Deaths and Cases.[2] Those bulletins aim to provide insights into those who have either died from or contracted COVID-19, combining data from the Health Protection Surveillance Centre Computerized Infectious Disease Reporting (CIDR), the HSE's Swiftcare (A2i) and COVID Care Tracker (CCT) systems. The statistics have been originally compiled from a broad range of sources, including the Central Statistics Office, the Central Bank of Ireland, Government Departments and bodies and other international sources.

### 3.2. Data pre-processing

A combination of GCMR and CSO Weekly Profile of COVID-19 confirmed cases and deaths is used for this research. The GCMR *Residential* category shows a change in duration (time users spent at home, using the home addresses provided to or estimated by Google Maps) —the other categories measure a change in total visitors [49]. As people spend most of the day at places of residence, the ability for variation is not significant.[3] The residential data were excluded from the GCMR as expressing a different metric and due to strong dependence on other variables used in the Google dataset.

The spread of COVID-19 mainly occurs through social contacts between infectious and susceptible individuals [50]. Hence, a reduction in mobility should lead to a reduction in social contacts, then in the infection spread and ultimately, in the COVID-related mortality [6]. However, this process requires time; as expected, the reduction in physical mobility observed today possibly has an impact on the infection spread and the related mortality in the coming weeks. This calls for the introduction of a time lag in the mobility data, which corresponds to the amount of time necessary for the change in mobility to have an impact on mortality.

We assume that there is a 1-week lag from the cause (alterations of movement trend categories percent change from baseline) to having a confirmed case. We also suspect that there is a 2-week lag from a

---

confirmed case to a confirmed death. These assumptions are based on WHO information that states: "*the incubation period of COVID-19, which is the time between exposure to the virus and symptom onset, is on average 5–6 days, but can be as long as 14 days. Thus, quarantine should be in place for 14 days from the last exposure to a confirmed case.*" The 2-week lag from a confirmed case to a confirmed death is also supported by visually inspecting the Irelands total number of deaths and cases (Fig. 1).

Furthermore, the daily mobility data did not align with the weekly CSO statistics. We amended our data in order to align the daily mobility patterns with the sum of the weekly CSO statistics, under the assumption that weekly totals serve as a proxy for the daily COVID-19 trends throughout the (appropriately-lagged) week.

Our focus in this network analysis is to measure the relative prominence of within-Ireland mobility regarding infection dynamics. Each factor's prominence is measured relative to its association with the number of deaths recorded in Ireland (following a 3-week lag). Given that, the emphasis is placed on *within*-Ireland factors (using factors recorded from each county), we removed 'Total Ireland Cases' from our analysis to avoid redundancy. We did not include in the final dataset the cumulative variable of Ireland. Instead, we used the individual data of counties. We conducted a preliminary examination that included Ireland as a node in the network analysis; this inclusion did not affect the final results (data not shown).

Up to December 2020, Ireland had a COVID-19 mortality rate of 2.8% (based on 76,000 cases). There is a relatively low number of deaths reported weekly on a per-county level, showing that a number of sparsely-populated or small counties did not record deaths each week. Therefore, we used the total number of deaths for Ireland, one death time series (Ireland's Confirmed Deaths) instead of using all death time series for all counties.

Our model assumes that there are correlation patterns among mobility and confirmed cases of different counties. This correlation may or may not reflect a causality relationship. However, this assumption is helpful in network modelling. Correlational networks have been useful in finding complex patterns in different research domains, especially in finance [51] and biology [52]. The algorithms used to extract results from the network models filter low correlation relations with methods like maximum spanning trees [42] or value them as much as the correlation value [53]. Another benefit of assuming interconnection between different mobility and case variables of different counties using correlation is the correlation's noise reduction effect [54]. The COVID data, especially the confirmed cases data, suffer an inherent problem of having noise due to lack of confidence in test methods, data gathering problems in local hospitals and laboratories, data aggregation by local authorities, and final data confirmation by governmental bodies. Using correlation network modelling and another smoothing method that we use in our approach reduces the effect of noisy data on results.

We refer to the mobility data in the dataset: as the Causes and the confirmed COVID-19 cases as the Cases. Time series with no data are excluded from Causes and Cases data. Regarding Causes, specifically, the following variables were eliminated for all the datasets because of missing data: [*Leitrim_transit_stations, Longford_parks, Longford_transit_stations. Monaghan_parks, Roscommon_transit_stations*].

A 7-day moving average, $x'_n = (x_n + x_{n-1} + x_{x-2} + \ldots + x_{n-6})/7$, is used in each time series of the dataset to reduce the noise by smoothing the data time curve. Each $x_i$ is a daily value in a time series and $x'_n$ is the converted value in the smoothed time series. The date range of Causes in the dataset is from March 05, 2020 to November 16, 2020, corresponding to a death reporting from March 26, 2020 to December 07, 2020. Even after data processing and noise reduction, a total of 1643 out of 39,064 values (~4%) of the time series elements were not available in the final dataset. We used the implementation of the correlation function in the Pandas package in Python, which excludes the missing values from its calculation.

### 3.3. Modelling approach

For modelling the relationship between variables, a correlation network structure is used. In our network structure, each time series is a node in the network, and the link between each pair of nodes (where there is a link) displays their correlation. As mentioned above, the dataset is divided into Causes (changes in mobility patterns), confirmed counties Cases, and Ireland's Confirmed Deaths. After building the network structure of relationships among Ireland Death variables, a rank propagation algorithm (Personalised PageRank) is used to score variables according to their importance in Ireland Death's fluctuation. This phase is similar to graph-based semi-supervised learning methods where the label of one node is propagated to all the other nodes in a graph structure [55]. Then, each county's rank scores are normalised and divided by the number of categories (calculation of relative prominent factor). The presented feature selection method uses categories with relative prominent factors greater than one. Finally, the counties with the same important selected features are put in the same clusters. A general schema of the methodology is presented in Fig. 2.
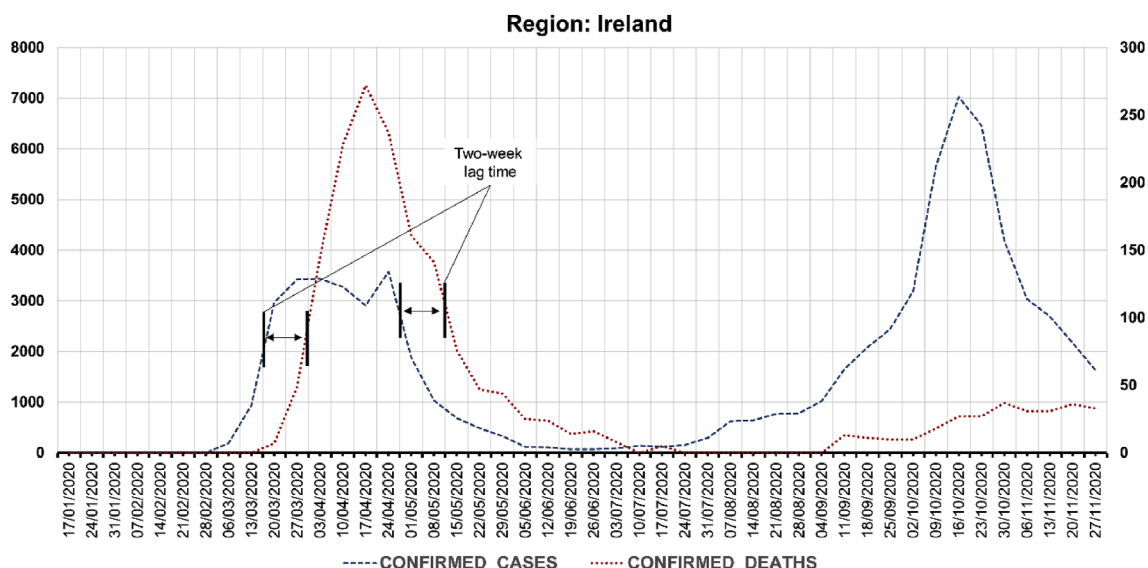


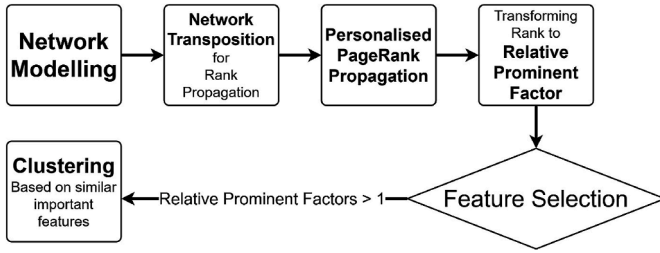**Fig. 1.** Ireland confirmed cases and deaths.

**Fig. 2.** General schema of the research methodology.

The final structure of the network is shown in Fig. 3. There are two fully connected networks for Causes and Cases and a directed connection from all the nodes in the Causes network to nodes in the Cases network. Finally, there is a directed connection between all nodes of Causes and Cases to the Confirmed Deaths node.

We label the weight of a link between each pair of Causes (or pair of Cases) $i$ and $j$ as $w(i, j)$.

For calculating the weights of the links, first, the correlation between node $i$ and $j$ is converted to distance, $d_{i,j}$ using Eq. (1) [56]:

$$d_{i,j} = \sqrt{2 \times (1 - Correlation_{i,j})} \qquad (1)$$

Then the weight between each pair of nodes (time series) is calculated as in Eq. (2) [55]:

$$w_{i,j} = e^{-\frac{d(i,j)}{26^2}} \qquad (2)$$

Due to the 1-week, 2-week, and 3-week lags, there will only be 1-way directed links from each cause to each case nodes (time series) in the network. There will also be only 1-way directed links from each case times series to Ireland's Confirmed Deaths' node and from Causes to Deaths' node.

To determine which nodes (time series) have a higher impact on Ireland's daily COVID-19 deaths, we should propagate the rank from Ireland's Confirmed Deaths node to all other nodes in the network. For this reason, we have to use the transpose network, which is the same as our network but with all directions reversed, Fig. 4.

After building the transpose network structure, we use a personalised PageRank centrality measurement [41] to investigate the importance of different Causes in Ireland Confirmed Deaths values. The personalised PageRank formula is presented in Eq. (3). In this equation, $P$ is the PageRank vector with $n$ elements, one for each of the time series (nodes in the network), $M$ is the Markov transition matrix extracted from the adjacency matrix of the network structure, $d$ is the damping factor set to 0.15 according to Ref. [41]; $n$ is the total number of time series (total number of nodes in the network), and $t$ indicates iteration count.
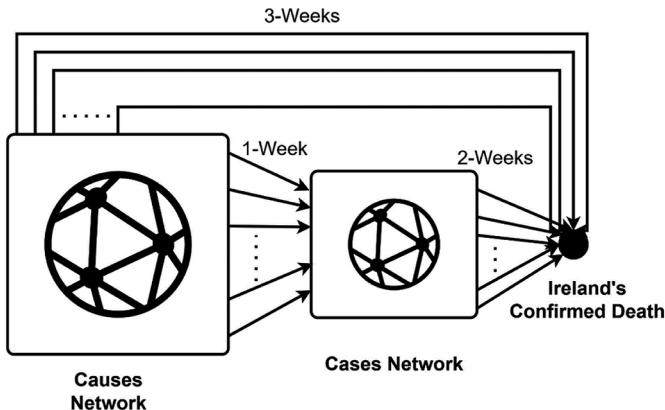


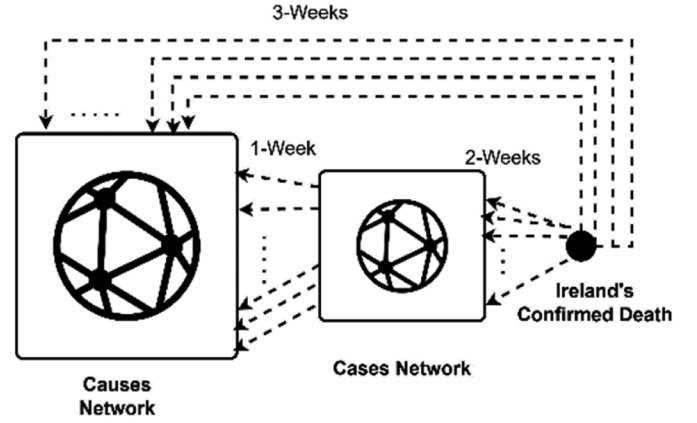**Fig. 3.** Network structure used to discover the Cause importance.



**Fig. 4.** Transpose Network for Personalised PageRank Propagation from Deaths time series to other time series (nodes).

$$\begin{aligned} P_t &= dMP_{t-1} + (1-d)/nP_0 \\ P_0 &= [0, 0, \ldots, 1](1\ for\ Deaths\ node,\ 0\ for\ other\ nodes) \\ |P_0| &= n,\ (n:\ number\ of\ nodes\ in\ network\ structure) \end{aligned} \qquad (3)$$

After some iterations, the algorithm converges when $|P_t - P_{t-1}| < epsilon$, where *epsilon* is the convergence threshold in Eq. (3) and the ranking scores for each node has been estimated. Then the nodes (times series) are arranged in descending order according to their Personalised PageRank scores.

### 3.4. Attribute analysis and clustering

From the modelling approach, a Personalised PageRank (PPR) score is calculated per each county individually, and linked with mobility patterns, confirmed cases and deaths. PPR indicates the absolute significance of the structures within the data, which unsurprisingly highlight cases and deaths as significant nodes of interest. Mobility patterns are relatively under-considered in the PPR network. As such, we transformed the PPR scores for the mobility patterns into *relative prominent factors*, such that relatively significant factors can be identified. We carry this out by:

1. Excluding 'county cases' and 'Ireland deaths' from the sum of PPR scores (which would otherwise equate to 1)
2. Normalising the PPR score of the mobility patterns associated with each county by dividing individual PPR scores by the sum of the remaining PPR scores to form individual Prominence Scores:

$$Prominence\ Score_i = \frac{PPR_i}{\sum PPR}$$

3. Dividing each Prominence Score by expected prominence scores, using the naïvely-informed approach. Under this approach, we would expect each factor remaining in the network to have an equally-likely prominence score (i.e. $E(Prominence\ Score_i) = \frac{1}{n}$). The relative prominence of each mobility pattern node within the structure can then be determined by measuring the observed value against the expected prominence score:

$$Relative\ Prominence\ Score_i = \frac{Prominence\ Score_i}{E(Prominence\ Score_i)} = Prominence\ Score_i \times n$$

4. Any variable with a relative prominent score >1 has a relatively-significant impact on the number of COVID-19 deaths, and

indicates mobility pattern nodes that were over-represented (i.e. prominent) within the data.

The underlying idea is to identify prominent 'nodes' (factors) in a network by identifying nodes that are supported, or linked to, other prominent nodes [29,57]. The network in our analysis considers each county individually, linked with mobility patterns and confirmed cases. PPR provides an indication of relative significance of the compared entities; each PageRank score indicates the likelihood of a random node being associated with another. The higher the likelihood, the more prominent the node in the network. In order to get an absolute measure of each node's prominence in the network, we compare the PageRank score of each node against a naïvely-informed score. The naïvely-informed score is calculated as the total PageRank scores divided by the number of factors in the network (125 for mobility factors). In other words, this score provides an indication of the 'expected average' prominence of a factor relative to its subgroup. We thereafter transformed the PPR into relative prominent scores by normalising and dividing the count of categories in each subgroup. This operation meant that the 'expected average' for each subgroup (mobility patterns) is 1. Any factor with a relative prominence score >1 is more connected with Ireland's number of deaths figure than a naïvely-informed connection.

Once the significance is estimated, the counties are grouped into different categories to facilitate visualisation of the results and discussion. According to the distance of their scores, we group the counties in two ways; according to the number of dominant mobility categories in each county. We sorted the most significant mobility categories per county according to their prominent factor scores, and we gathered the counties in groups of the same number and same dominant mobility categories. For instance, two counties with only retail and recreation and workplace scoring above 1, belong to the same group.

Additionally, we cluster the counties hierarchically according to the average Euclidean distances of their mobility scores into nine clusters [58,59]. Our hierarchical clustering is like an agglomerative ("bottom-up") type of clustering method. It begins with regarding each element as a separate cluster and then merge them into larger clusters successively. Specifically, in each particular step of hierarchical clustering, it finds the pair of clusters with the smallest distance and then merges them into a new parent cluster [60]. We used Euclidean clustering distance (square root of the sum of square distances). The step is repeated until only one cluster formed, and the results can be described in a dendrogram [61]. The clustering was implemented on the ClustVis [62].

## 4. Results

To maintain alignment between the mobility data and our expectations for COVID-19 transmission despite having a fixed period of data (as detailed in section 3.2), we employed a 'moving window' approach during the data preparation phase. Our daily mobility patterns totalled 257 observations from each locality, from March 5, 2020, to November 16, 2020. Assuming an average incubation period of 7 days from exposure to the presentation of symptoms, we associated the mobility patterns with average cases from March 12, 2020 to November 23, 2020. Assuming a further period from initial presentation of symptoms to death of 14 days, we associate mobility patterns and cases to weekly average 'deaths' data from March 26, 2020 to December 7, 2020. Furthermore, we employed a stepwise moving average of 7 days on the 'COVID-19 cases' and 'COVID-19 deaths' in order to reduce noise and extrapolate weekly data to daily data.

Mobility pattern data was available for all 26 counties in the Republic of Ireland along with five distinct indicators (Table 1). Combining these data with the cases associated with each county, and the total deaths recorded in Ireland, a total of 157 nodes in the network could be made available in the network. However, mobility pattern values pertaining to five indicators in sparsely-populated counties contained little-

to-no data; these factors were removed from the dataset.

Therefore, the final dataset consists of 257 daily data for 152 time series, including 125 time series of Google Mobility data, 26 time series of confirmed cases for counties with reported data, and one time series of the daily confirmed COVID-19 deaths of Ireland. In order to demonstrate the significance of the variables in the ranking score, we transformed the variables into relative prominent factors. Any variable with a prominent factor value > 1 has an impact on death number more significant compared to the naïve approach, as explained in section 3.4.

In this section, we show the results of the methods we employ in this paper. Specifically, we demonstrate (i) the most significant causes that influence Ireland's number of deaths, (ii) the primary causes that influence Ireland's number of deaths on a county level, and (iii) the clusters that appear to have similarities among the counties based on the impact on Ireland's number of deaths.

There is no ground truth about the ranking of important features in the dataset or anywhere else to have a test set to evaluate the error using famous criteria like accuracy, precision, or recall. In our Personalised PageRank algorithm, the value of 1 is assigned to the Death node and is propagated through the network structure. So the source of the error will not be what is propagated through the network but how it is propagated. The means of the propagation are the weights of the link. However, as mentioned before, the weight of the links can be a source of error due to the way they are calculated. These weights, as mentioned in the methodology, are calculated by lead-lag correlation functions. The correlation function uses the GCMR and CSO time series (noisy data), which are the base source of the error in our methodology. This noise is produced due to failure to record the right information or record the information correctly, information aggregation from different local sources like local hospitals, lack of data gathering at weekends and holidays. Some parts of these noises are correlated to each other, and some are not. These noises propagate into the correlation table, to the graph's adjacency matrix and finally to the PageRank results. To reduce the noise, we used a 7-day moving average smoothing function.

Another aspect related to the error is the PageRank algorithm which converges after iterations. A stable state defines convergence in the results, which should happen after some iterations. We used a value of $10^{-6}$ as the epsilon value. This value has been used in different researches and is one parameter that can control the final error of PageRank results. If the distance between the results of the two last iterations is less than the defined epsilon, then the convergence error is tolerable, and the ranks are reported. Therefore this epsilon number should be reported.

### 4.1. Attribute analysis and clustering

The mobility patterns account for a cumulative 19.5% of the prominence in the whole network structure, including Cases and Ireland deaths nodes. Based on 125 distinct mobility-pattern factors, mobility nodes with a PageRank score greater than 0.15% have a higher relative prominence than expected from a naïvely-informed approach. In contrast, mobility nodes with a PageRank score below 0.15% have a lower relative prominence than expected. We rebase the PageRank scores such that the expected average is 1. Therefore, mobility nodes greater than 0.15% are now 'relative prominence scores' that are greater than 1, and mobility nodes less than 0.15% are now 'relative prominence scores' less than 1. Table 2 shows the top factors (all variables with prominent factor > 1) that influence the Ireland Covid-19 Number of deaths.

It is clear from Table 2 that the mobility categories *retail and recreation* (Reta.) and *workplaces* (Work.) are the most important ones. Reta. And Work are the most frequent variables that appear in the top table, having relative prominent factors >1, appearing at 33% and 38% of the cases, respectively. Movement trend categories: *grocery and pharmacy* (Groc.) appears less often, at 15% of all cases. A specific county mobility

**Table 2**

Top causes that influence the Ireland Death numbers for 2020. Movement Trend Categories 1) Groc: *grocery and pharmacy percent change from baseline,* 2) Park: *parks percent change from baseline,* 3) Reta: *retail and recreation percent change from baseline,* Tran: *transit stations percent change from baseline,* 4) Work: *workplaces percent change from baseline.* RPF: *relative prominent factor.*

| County | RPF | County | RPF | County | RPF |
|---|---|---|---|---|---|
| Meath_Parks | 1.848 | Westmeath_Groc. | 1.404 | Sligo_Parks | 1.225 |
| Donegal_Work. | 1.516 | Limerick_Work. | 1.397 | Limerick_Parks | 1.215 |
| Kerry_Work. | 1.514 | Galway_Work. | 1.390 | Dublin_Work. | 1.210 |
| Wexford_Work. | 1.504 | Waterford_Reta. | 1.389 | Monaghan_Work. | 1.208 |
| Cork_Reta. | 1.489 | Kilkenny_Reta. | 1.387 | Louth_Tran. | 1.205 |
| Kildare_Reta. | 1.482 | Laois_Reta. | 1.380 | Tipperary_Parks | 1.204 |
| Louth_Reta. | 1.472 | Dublin_Reta. | 1.376 | Limerick_Groc. | 1.194 |
| Limerick_Reta. | 1.467 | Tipperary_Work. | 1.366 | Monaghan_Groc. | 1.161 |
| Louth_Work. | 1.464 | Offaly_Reta. | 1.366 | Wexford_Parks | 1.156 |
| Meath_Work. | 1.463 | Wexford_Reta. | 1.355 | Waterford_Groc. | 1.152 |
| Mayo_Work. | 1.457 | Kildare_Work. | 1.341 | Sligo_Groc. | 1.151 |
| Tipperary_Reta. | 1.451 | Mayo_Reta. | 1.338 | Donegal_Reta. | 1.143 |
| Waterford_Work. | 1.445 | Kilkenny_Work. | 1.329 | Cavan_Work. | 1.138 |
| Kildare_Parks | 1.437 | Carlow_Reta. | 1.322 | Sligo_Work. | 1.132 |
| Galway_Reta. | 1.431 | Sligo_Reta. | 1.321 | Roscommon_Work. | 1.124 |
| Wicklow_Reta. | 1.424 | Westmeath_Work. | 1.320 | Kerry_Reta. | 1.109 |
| Meath_Reta. | 1.420 | Cork_Groc. | 1.319 | Cork_Tran. | 1.097 |
| Wicklow_Work. | 1.419 | Clare_Work. | 1.318 | Carlow_Work. | 1.084 |
| Louth_Parks | 1.416 | Laois_Work. | 1.279 | Longford_Work. | 1.040 |
| Cork_Work. | 1.414 | Clare_Reta. | 1.276 | Longford_Groc. | 1.030 |
| Westmeath_Reta. | 1.411 | Offaly_Work. | 1.257 | Carlow_Groc. | 1.027 |
| Cavan_Reta. | 1.404 | Galway_Groc. | 1.234 | Kilkenny_Groc. | 1.024 |

category-*parks* appears the most significant (Meath_Parks) when comparing the entire country's variables. Very few other *parks* appear in the top results (11% approximately). Movement trend category: *transit* stations (Tran.) appears least often (less than 5%).

For the following five locations, Wicklow, Cavan, Laois, Dublin and Offaly, a similar structure is observed at county level, corresponding to
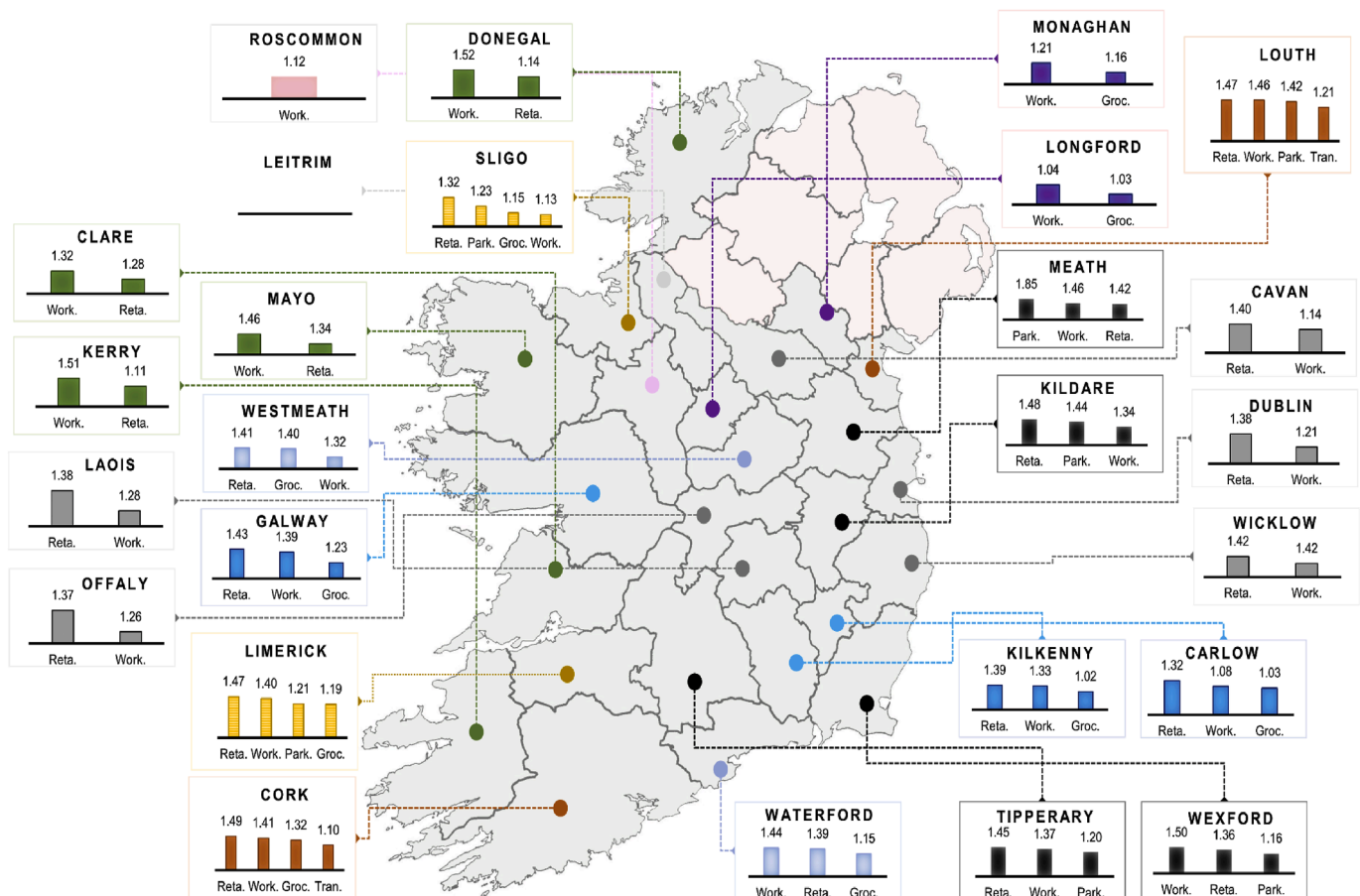


**Fig. 5.** Ireland map divided by counties. The most significant categories are shown per county (categories with a relative prominent factor >1). County colors demonstrate similarity. *Movement Trend Categories: Groc: grocery and pharmacy percent change from baseline, Park: parks percent change from baseline, Reta: retail and recreation percent change from baseline, Tran: transit stations percent change from baseline, Work: workplaces percent change from baseline.*

the same top two most important variables in descending order without the appearance of a third one that ranks higher than the naïve approach (*retail and recreation* and *workplaces*) (grey color, Fig. 5).

Waterford, Mayo, Kerry and Clare manifest identical associations to countries cases and deaths, with *workplaces* appearing first, followed by *retail and recreation* (green color, Fig. 5). Analogous results appear in Longford and Monaghan with *grocery and pharmacy* following the *workplaces* (purple color, Fig. 5).

Westmeath and Waterford consist of the same important attributes; however, *retail and recreation* appear to have a higher ranking in Westmeath than *workplaces* in Waterford (light blue color, Fig. 5).

Sligo and Limerick demonstrate similar results, with *retail and recreation* ranking first. *Workplaces, grocery and pharmacy* and *parks* follow next with a slight difference in the ranking (yellow color, Fig. 5). Comparably, Cork and Louth show similar ranking order with *transit stations* appearing uniquely as important in only those two cases (orange color, Fig. 5).

Kilkenny, Galway and Carlow reach equivalent ranking scores, with *retail and recreation* scoring first coming by *workplaces* and *grocery and pharmacy* (vivid blue color, Fig. 5).

Tipperary, Kildare, Wexford and Meath follow similar patterns even though the ranking is slightly different in all the cases individually. In general, *retail and recreation, workplaces* and *parks* appear as the top three significant variables (black color, Fig. 5).

Exceptions from the above ranking orders include Roscommon, where only *workplaces* appear as significant (pink color, Fig. 5) and Leitrim showing no relatively significant impact to Ireland death number by any category.

*Retail and recreation* and *workplaces* appear systematically as the most important causes across all counties. Five distinct segments of counties can be formed (

Fig. 6) depending on the number of their dominant categories. Ten segments are distinguishable when top categories are checked to be the same.

1. A segment in which four important categories (causes) appear as the most important attributes including *retail and recreation, workplace, parks, grocery and pharmacy* or *transit stations* (Cork, Sligo, Limerick and Louth). Sligo and Limerick have the same top four categories.
2. A segment in which three important categories appear as the most important attributes, including *retail and recreation, workplace, parks*

or *grocery and pharmacy* (Kilkenny, Galway, Carlow, Waterford, Westmeath, Tipperary, Kildare, Wexford and Meath).
3. A segment in which two important categories appear as the most important attributes including *retail and recreation, workplace,* or *grocery and pharmacy* (Wicklow, Dublin, Laois, Offaly, Cavan, Clare, Mayo, Donegal, Kerry, Longford and Monaghan).
4. A segment in which one category appears as the most important, including *workplaces* (Roscommon).
5. A segment in which neither category appears as significant (Leitrim).

In hierarchical clustering, twenty-five counties were used as the initial clusters to form eventually eight groups; Leitrim was excluded as it scored below one in all mobility categories and was assigned a cluster on its own (the ninth one) Fig. 7. Not surprisingly, the clustering based on the Euclidean distance of the scores produces clusters with the same objects exactly as the segmentation based on the dominant categories. There are three large clusters and six small ones.

## 5. Discussion

The COVID-19 pandemic has generated an abundance of research following the outbreak. Within a few months, more than a thousand studies have already appeared in the scientific literature [63,64]. The findings from Ref. [63] highlight that a holistic approach and a collective research initiative is required better explain and reduce the safety impacts of this crisis, whose implications reach beyond the bio-medical risks.

The World Health Organization is concerned about the current epidemiological situation of the subsequent pandemic wave and sets the main objective of establishing and maintaining indicators to more effectively support governments' decisions to minimise pandemic losses [65]. In this paper, we introduce a methodology to shed light on the pandemic related data. Our results are derived from ensembled COVID-19 confirmed cases and deaths in combination with mobility pattern changes to reveal significant correlations between mobility changes and Ireland's deaths in a quantified manner. Importantly, our results revealed the most impactful counties and the most important attributes at the county level to Ireland's death number. The analysis allowed us to visualise the importance and cluster the counties based on the scores derived from the Personalised PageRank algorithm.

In Table 2, we showed the top causes of COVID-19 deaths in Ireland using Personalised PageRank scoring. One of the most notable points is
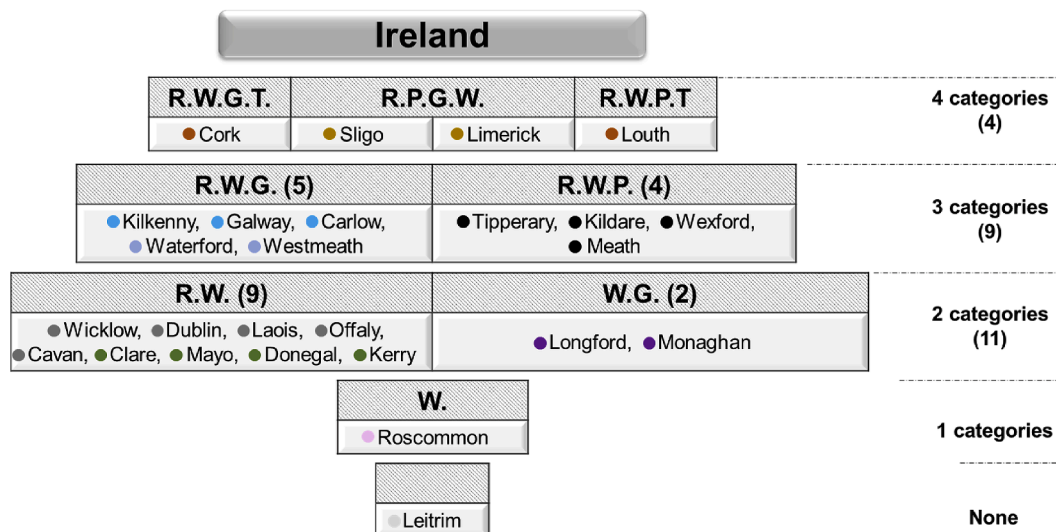


**Fig. 6.** Clustering of counties based on the number of significant variables that affect the country's number of deaths. *Movement Trend Categories: G: grocery and pharmacy percent change from baseline, P: parks percent change from baseline, R: retail and recreation percent change from baseline, T: transit stations percent change from baseline, W: workplaces percent change from baseline.*
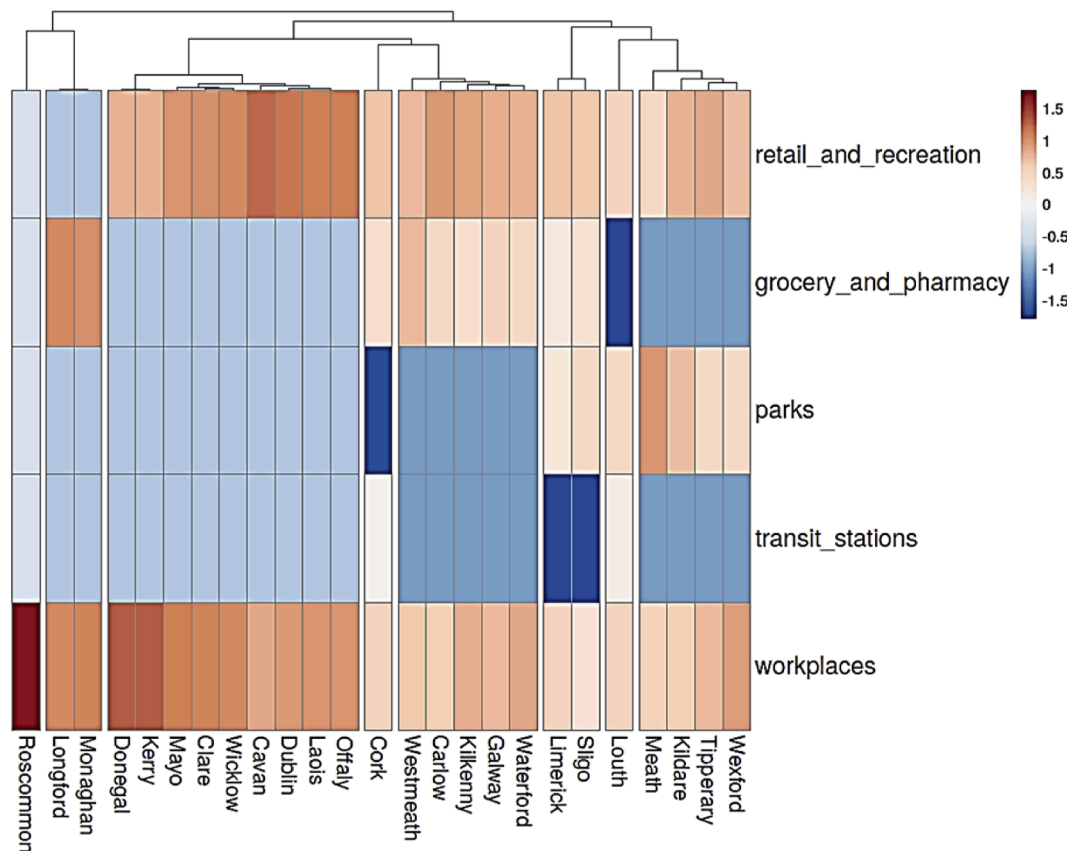
**Fig. 7.** Hierarchical Clustering of counties based on the RPF scores regarding the country's number of deaths. Leitrim (the ninth cluster) is not shown due to zero scores. The color scale refers to standardised values.

the limited occurrence of the movement trend category: *grocery and pharmacy,* which appears in just 15% of all cases. Given the obvious requirement for most citizens to shop for staples, this is surprising. We suggest that this may be as a result of heightened awareness of social distancing by shoppers and a prior culture of hygiene by food retailers. Equally, food retailers would suffer significant local and reputational damage if they were perceived as being less than committed to adhering to safety guidelines. An interesting study conducted by Ref. [66] results on aerosol and droplet dispersion physics related to the simulated transmission in a supermarket. The authors mention that the number of inhaled aerosols is very small for all the customers if coughing persons are not present. This research demonstrates that even with a high consumers count, the exposure will be reasonably low for a given person on a regular basis.

Table 2 also shows that *Retail and recreation,* and *workplaces* are the most frequent variables that appear in the ranking score having a relative prominent factor >1, at frequencies of 33% and 38%, respectively. This signals that this factor was prominent in contributing to the number of COVID-19 deaths in Ireland relative to other mobility patterns across Ireland. In contrast to grocery shopping, we suggest that non-food retail and recreation locations do not have a strong emphasis on hygiene prior to COVID-19 and are likely to have occasional visitors with lower standards and expectations in terms of the location's adherence to safety protocols. Similarly, we suggest that the familiarity of workers in *workplaces* and both the difficulty of social distancing and the cultural awkwardness of adhering to social distancing might be factors in the prominence of these locations in our findings. Furthermore, *Retail and recreation* and *workplaces* mobility were only allowed during times of lowered lockdown restrictions. This suggests that loosening restrictions had a direct causal impact on the COVID-19 related deaths.

At a county level, the map in Fig. 5 is instructive. For five locations

Wicklow, Cavan, Laois, Dublin, Offaly (Fig. 5, grey color), we observe a similar pattern structure. That is, *retail and recreation* and *workplaces* are the only important variables in a descending order. We observe a similar pattern for Donegal, Mayo, Kerry and Clare, where *workplaces* appear first, followed by *retail and recreation* (Fig. 5, green color). The former group is significantly more urban than the latter suggesting the higher prevalence of retail and recreation facilities in more densely populated locations correlates strongly with a greater incidence of COVID-19. For both of these two groups, the absence of *grocery and pharmacy* is notable. By contrast, in Longford and Monaghan, *grocery and pharmacy* are followed by the *workplaces* (Fig. 5, purple color). These counties are sparsely populated with relatively low industry and agriculture being an important economic activity but an isolated occupation. Therefore, *grocery and pharmacy* are the locations of most social interaction.

For Westmeath and Waterford (Fig. 5, light blue color), while both contain the same important attributes, their order of importance differs. In particular, the relative importance of the *workplaces* mobility attribute ranks with top importance for Waterford while bottom for Westmeath. Similarly, ranking order differs slightly for both Limerick and Sligo (Fig. 5, yellow color). Both counties exhibit all four significant mobility attributes, with *retail and recreation* ranking most prominently. However, changes in workplace mobility behaviors represent considerably less significance for Sligo. We consider this difference a result of rural (Sligo) versus urban (Limerick) employment profiles. Cork and Louth (Fig. 5, orange color) also demonstrate four significant categories and similar ranking orders, uniquely displaying the *transit stations* mobility attribute as significant. For Louth, this is unsurprising given its commuter town status to Ireland's capital city and centre of industry. Based on the relative prominence scores (Table 2), the four counties (Cork, Limerick, Louth, Sligo) had prominence scores that were greater than expected from a naïvely-informed perspective for four out of the

five mobility categories. Intriguingly, all four counties have direct railway networks with Dublin, the central hub of the virus in Ireland. However, the movement trend category: *transit stations* does not appear as often as the other variables (less than 5%). This is consistent with a relatively low level of public transport use and a relatively rural population.

*Retail and recreation*, *workplaces*, and *grocery and pharmacy* mobility behaviors had significant consequences for Galway, Kilkenny and Carlow (Fig. 5, vivid blue color). Similarly, Tipperary, Kildare, Wexford and Meath (Fig. 5, black color) include both *retail and recreation* and *workplaces* within their three significant attributes, with *parks* replacing *grocery and pharmacy* as a variable of importance.

Outliers in the analysis include Roscommon (Fig. 5, pink color), where only the *workplaces* mobility attribute was determined significant, and Leitrim, where no mobility attributes of significance were observed as having a relatively prominent association with COVID-19 deaths in Ireland. This may be due to low population densities resultant of Leitrim and Roscommon being ranked the most rural counties in Ireland according to the CSO Census 2016.

*Retail and recreation* category dominates the top determinants of infection casualties. The COVID-19 pandemic has forced numerous businesses to limit the number of shoppers inside the store to minimise infection rates. Yet, confusion about what steps have to be taken is hindering the efforts of retail traders, and there is an immediate necessity to comprehend how distancing can be secure. Ntounis, Mumford et al. [33] estimated the amount of space needed to safely social distance in various retail environments when there are people present. The proposed method can work as a step forward in understanding the very practical problem of capacity, which can allow retail spaces to operate safely and minimise the risk of virus transmission. Perlman and Yechiali [31] proposed a new method by which to quantify the risk of an infected customer when queuing outside the store, when shopping, and while checking out. In their attempts to protect both consumers and workers while reducing related costs, the results are valuable and applicable to governments and companies alike. We recommend those approaches to be investigated for counties where *Retail and recreation* ranks the highest.

Regarding the more recurrent determinant, the *workplaces*, some activities are impossible to perform virtually, such as construction activities. The construction area is a vital section of countries' economies. Araya [28] developed an agent-based approach to model the spread of the virus on construction workers finding that the workforce from a construction project may be reduced by 30%–90% due to the spread of COVID-19. The main way managers in problematic counties may have to reduce the spread of COVID-19 among construction workers is to classify the activities involved in construction projects as low, medium, and high risk regarding the spread and maximise those classified activities as low-risk. Counties where workplaces are ranked very high, should investigate COVID-19.

Another study conducted in Vietnam showed that underlying health problems, including respiratory system problems, were common in industrial *workplaces*. Many people suffered from occupational diseases or wounds and injuries from work, where the highest infection rate was among industrial works. They also showed more severe symptoms when infected [67]. We support the authors' suggestion that measures such as having COVID-19 testing centres at or near industrial sites that can effectively control the emergence of the virus in industrial sites not only support the overall effort of combating the disease but also help to protect this vulnerable population from more suffering in the future.

Regarding confined *workplaces*, another study showed how crowd modelling could be used to assess occupant exposure [26]. The adopted policies in buildings and social distancing in a disease can be different. These policies are more related to the macroscopic analysis of the infection rather than safety evaluation at the building-level. The proposed model allows to investigate the exposure according to microscopic analysis of workers' movements. This helps the policymakers to make better decisions when using buildings in a pandemic.

Taken collectively, our results suggest that *Retail and recreation* and *workplaces* are those locations where government interventions should be most directed. We included above a number of proposed safety assessments and management methodologies that could be used by policymakers before including additional directives, advertising, inspections and possible sanctions for non-compliance are issued.

One county mobility category-*parks* appears as the most significant (*Meath_Parks*). Very few other *parks* appear in the top results (11% approximately). Public parks in Dublin are not detailed in the top results suggesting that Parks are not a significant location for COVID-19 transmission. In rural counties, parks are uncommon, and the definition of parks is ambiguous. We suggest this is the result of a number of outlying cases in a rural county.

Urban agglomeration appeared in our study to be of the most impactful sites. Mohammadi, Chowdhury et al. [27] developed a risk estimation model considering the problem of physical distancing requirements for urban sidewalk pedestrians. This is why the urban designers and policymakers want to force a 2-m distance among pedestrians to avoid the virus spread. The study demonstrated the application of the proposed approach in a simulated virtual walkway environment. Such an approach could help health policymakers find out the best methods to enforce social distancing among urban sidewalk pedestrians.

Regarding the individualistic behaviour during the pandemic, a study showed that people are less (vs. more) willing to take risks when information is positively (negatively) framed, irrespective of disease type, although they are generally more risk-averse in real pandemics [36]. Furthermore, people high (vs. low) in emotionality are more willing to comply with preventive health behaviours when information is framed positively (vs. negatively), but only in the case of a real disease. These findings provide a range of insights into the design and management of health recommendations aimed at promoting public health. We suggest linking preventive suggestions to country individuality and local pride for establishing more effective and affecting communication.

A study using Bayesian Networks found a lack of responsibility and consciousness among the most likely risk factors for COVID-19 within the public sectors include [37]. County limited and specific preventing measurements and relevant information flow regarding the pandemic would enforce a sense of responsibility and shared consciousness in the community.

Instead of using the Google mobility data, mobile positioning data could be used to measure the effect of COVID-19 restriction measures on human mobility. Santamaria, Sermi et al. [38] carried out spatio-temporal harmonisation to a mobility indicator that captures insights into the mobility patterns of the EU population and can provide information regarding the effect of COVID-19 restriction measure on mobility in the EU.

Our findings suggest that even in the case of an island country, the significance of COVD-19 determinants shows geographical inhomogeneity. Confinement measures should be applied accordingly. Although interventions regarding mobility in *parks* and *retail and recreation* are advisable in the cases of Kildare or Wexford, they would have little impact in Longford and Monaghan. The clustering presented suggests similar infection spread management among counties inside a group and differentiated between the groups. As those measures put major economic and psychological stress on communities, resolving and targeting them as necessary only on locally impactful sectors will restrain costs, back measures public acceptance, enhance compliance and result in overall safety improvement.

This study has some limitations.

- First, we focused on quantifying the relationship between mobility patterns, cases and the total number of deaths; the number of deaths

per county did not allow for such an analysis to occur because of missing values and/or very low number in some counties.

- Secondly: there is a significant uncertainty surrounding the number of infections and deaths: data gathering problems in local hospitals and laboratories, data aggregation by local authorities, and final data confirmation by governmental bodies.

- Relying on a fixed baseline period ignores the yearly seasonality of movement, which may be affected by weather patterns, national holidays, vacation periods, etc. Important bank holidays, extreme weather events, or other major events can significantly affect the mobility relative change. Nonetheless, it is worth noting, though, that no evidence of any major events in Ireland that could have systematically biased the Google mobility data during the baseline period.

- The baseline week is based on the median value, which would be largely unaffected by short-lived temporary fluctuations in absolute mobility values. The baseline period is not affected by restrictions on movement, which were first introduced in late March.

- While we create relative prominence factors based on mobility patterns and COVID-19 deaths in this analysis, this does not implicitly indicate that the factors are associated with COVID-19 deaths. Although 'Ireland COVID-19 deaths' is the central node and the subject of interest, PageRank scores base their prominence on the support of, or links to, other nodes. Rather than identifying clear patterns in how each county's mobility patterns affected COVID-19 deaths, the structure we generated identified mobility patterns ('retail and recreation', 'workplace') that are generally prominent in a network structure containing COVID-19 deaths.

- There is no analysis to our knowledge investigating whether Google mobility data are representative of the Irish population. Probably the data correspond mostly to specific cohorts who are more likely to have an electronic trail, that is, the younger and working parts of the population, which is, however, also the more mobile ones.

- The mobility data utilised here have some uncertainties and lack details that may not represent the exact behaviour of a wider population. However, they are the best openly available data source for tracking a population's movement offering the unique opportunity to examine the relationship between mobility and disease incidence [20].

## 6. Conclusions and future perspectives

Human mobility is among the key drivers fostering the rapid spread of disease. This is the first time that rank propagation modelling has been coupled with a clustering approach and used to analyse the transmission patterns of COVID-19. We calculated an individual ranking score from an ensemble of COVID-19 cases and deaths in combination with mobility pattern changes to reveal correlations between mobility changes and Ireland's deaths in a quantified manner. The graphical analyses demonstrate that mobility channels with looser restrictions had a causal impact on COVID-19 related deaths. The mobility categories exhibiting the strongest association to COVID-19 cases and deaths are *retail and recreation* and *workplaces*. Workplaces were highlighted as a significant transmission channel in all counties within the Republic of Ireland (with the exclusion of Leitrim). In addition, workplaces in rural areas were a stronger transmission channel than workplaces in urban areas. We can infer that positioning testing centres near industrial sites may effectively detect the emergence of the virus spread. Given the causal link established between mobility and COVID-19 deaths, the level of restrictions that were put in place may have been required to prevent further deaths due to COVID-19. Our findings suggest that even in the case of an island country, the significance of COVID-19 determinants shows geographical inhomogeneity. We suggest identifying groups of counties of similar infection spread patterns and determinants and differentiate infection management among the groups. Focusing on locally impactful sectors limits major economic and psychological stress

on communities and costs. Furthermore, county limited and county specific information and preventing measurements would enforce a sense of responsibility and shared consciousness in the community and back measures public acceptance, enhance compliance and result in overall safety improvement. We conclude by encouraging private companies such as Google to continue sharing data to foster academic research in areas of public interest.

Further research could include different PageRank family algorithms on the pandemic datasets. In this research, we used Personalised PageRank to study the effect of different features on one feature with the assumption that there is an underlying network model in our data space. Personalised PageRank is an efficient algorithm regarding the time order complexity that can converge in a few number of iterations. According to the study of Page and Brin in 1999, PageRank can converge in 52 iterations for a network of Web Pages with 332 million links with different in-degrees and out-degrees [41]. For the use of other algorithms in the PageRank family, modifications of the research problem has to be formed. For instance, HITS and SALSA algorithms try to partition the nodes of a network into two categories of hubs and authorities, where hubs are the nodes that have an influence on other nodes and authorities are nodes that are affected. Changing the research problem to the discovery of effective and under influence features in the general data space of the pandemic, HITS and SALSA algorithm - which is an optimised version of HITS in some applications - can be used [68]. SimRank finds similar nodes in a graph considering the whole graph structure and interrelationships among nodes. Using SimRank, the research problem swaps into finding features with similar behaviour in the pandemic data space.

## Author contributions

**Furxhi Irini**: Conceptualisation; Formal analysis; Visualisation; original draft; review & editing. **Arash N. Kia**: Conceptualisation; Formal analysis; Methodology; Coding, review & editing. **Darren Shannon:** Data curation; Formal analysis; review & editing. **Tim Jannusch**: Visualization; review & editing. **Finbarr Murphy:** review & editing. **Barry Sheehan**: Data curation; review & editing.

## Funding

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

## References

[1] Who. WHO coronavirus disease (COVID-19) dashboard. 2021.

[2] Lu H, et al. Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. J Med Virol 2020;92(4):401–2.

[3] Who. WHO announces COVID-19 outbreak a pandemic. 2020.

[4] Wilson ME. Travel and the emergence of infectious diseases. Emerg Infect Dis 1995; 1(2):39.

[5] Tatem AJ, et al. Global transport networks and infectious disease spread. Adv Parasitol 2006;62:293–343.

[6] Basellini U, et al. Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales. French Institute for Demographic Studies; 2020.

[7] Brauner JM, et al. Inferring the effectiveness of government interventions against COVID-19. Science 2020:eabd9338.

[8] Seale H, et al. Improving the impact of non-pharmaceutical interventions during COVID-19: examining the factors that influence engagement and the impact on individuals. BMC Infect Dis 2020;20(1):607.

[9] Hale T, et al. Variation in government responses to COVID-19. 2020. Blavatnik school of government working paper 31.

[10] Bonaccorsi G, et al. Economic and social consequences of human mobility restrictions under COVID-19. Proc Natl Acad Sci Unit States Am 2020;117(27): 15530–5.

[11] Perumal V, et al. First case OF COVID-19 IN Ireland. Ulster Med J 2020;89(2). 128-128.

[12] ECDC. European centre for disease prevention and control: COVID-19 situation update for the EU/EEA. 2021. as of week 1 2021.

[13] CSO. In: Central statistics office Ireland: COVID-19 deaths and cases; 2020. Series 18.

[14] Oulefki A, et al. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. Pattern Recogn 2020:107747.

[15] StatCounter. Mobile operating system market share Ireland. 2020.

[16] Bryant P, Elofsson A. Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. medRxiv; 2020.

[17] Bryant P, Elofsson A. Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. PeerJ 2020;8. e9879-e9879.

[18] Huynh TLD. Does culture matter social distancing under the COVID-19 pandemic? Saf Sci 2020;130:104872.

[19] Ilin C, et al. Public mobility data enables COVID-19 forecasting and management at local and global scales. medRxiv; 2020. 2020.2010.2029.20222547.

[20] Sulyok M, Walker M. Community movement and COVID-19: a global study using Google's Community Mobility Reports. Epidemiol Infect 2020;148. e284-e284.

[21] Tamagusko T, Ferreira A. Data-driven approach to understand the mobility patterns of the Portuguese population during the COVID-19 pandemic. Sustainability 2020;12(22):9775.

[22] Drake TM, et al. The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK. The Lancet Digital Health; 2020.

[23] Saba T, et al. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. Microsc Res Tech 2021;84:1462–74.

[24] Ahmad A, et al. The number of confirmed cases of Covid-19 by using machine learning: methods and challenges. Archives of Computational Methods in Engineering; 2020.

[25] Bloemers M, Montesanti A. The FAIR funding model: providing a framework for research funders to drive the transition toward FAIR data management and stewardship practices. Data Intell 2020;2(1–2):171–80.

[26] Ronchi E, Lovreglio R. EXPOSED: an occupant exposure model for confined spaces to retrofit crowd models during a pandemic. Saf Sci 2020;130:104834.

[27] Mohammadi A, et al. Developing levels of pedestrian physical distancing during a pandemic. Saf Sci 2021;134:105066.

[28] Araya F. Modeling the spread of COVID-19 on construction workers: an agent-based approach. Saf Sci 2021;133:105022.

[29] Bar-Yossef Z, Mashiach L-T. Local approximation of pagerank and reverse pagerank. In: Proceedings of the 17th ACM conference on Information and knowledge management; 2008.

[30] Barr GD. The Covid-19 Crisis and the need for suitable face masks for the general population. Chin J. Med. Res 2021;3:28–31.

[31] Perlman Y, Yechiali U. Reducing risk of infection – the COVID-19 queueing game. Saf Sci 2020;132:104987.

[32] Springboard AM. Re-opening UK retail post COVID - an analysis of shopper concerns and preferences. 2021. Retrieved, https://www.spring-board.info/insight s/report/re-opening-uk-retail-post-covid-19. [Accessed 23 April 2021].

[33] Ntounis N, et al. How safe is it to shop? Estimating the amount of space needed to safely social distance in various retail environments. Saf Sci 2020;132:104985.

[34] Still K, et al. Place crowd safety, crowd science? Case studies and application. J Place Manag Dev 2020;13(4):385–407.

[35] Dyckhoff H. A typology of cutting and packing problems. Eur J Oper Res 1990;44 (2):145–59.

[36] Otterbring T, et al. Replication and extension of framing effects to compliance with health behaviors during pandemics. Saf Sci 2021;134:105065.

[37] Fu L, et al. Human and organisational factors within the public sectors for the prevention and control of epidemic. Saf Sci 2020;131:104929.

[38] Santamaria C, et al. Measuring the impact of COVID-19 confinement measures on human mobility using mobile positioning data. A European regional analysis. Saf Sci 2020;132:104925.

[39] Xing Y, et al. A node influence based label propagation algorithm for community detection in networks. Sci World J 2014;2014:627581.

[40] Zhang X-K, et al. Label propagation algorithm for community detection based on node importance and label influence. Phys Lett 2017;381(33):2691–8.

[41] Page L, et al. The PageRank citation ranking: bringing order to the web. Stanford InfoLab; 1999.

[42] Wang G-J, et al. Statistical properties of the foreign exchange network at different time scales: evidence from detrended cross-correlation coefficient and minimum spanning tree. Entropy 2013;15(5):1643–62.

[43] Kloumann IM, et al. Block models and personalised PageRank. Proc Natl Acad Sci Unit States Am 2017;114(1):33–8.

[44] Vial D, Subramanian V. On the role of clustering in personalised PageRank estimation. ACM Trans. Model. Perform. Eval. Comput. Syst. 2019;4(4). Article 21.

[45] Graham FC, Tsiatas A. Finding and visualising graph clusters using PageRank optimisation. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010.

[46] Tabrizi A, S, et al. Personalized PageRank Clustering: a graph clustering algorithm based on random walks. Phys Stat Mech Appl 2013;392(22):5772–85.

[47] Naik A, et al. Scalable Twitter user clustering approach boosted by Personalized PageRank. Int J Data Sci Anal 2018;6(4):297–309.

[48] Google LLC. Google COVID-19 community mobility Reports. 2021.

[49] Lapatinas A. The effect of COVID-19 confinement policies on community mobility trends in the EU. Luxembourg: JRC; 2020.

[50] Nishi A, et al. Network interventions for managing the COVID-19 pandemic and sustaining economy. Proc Natl Acad Sci Unit States Am 2020;117(48):30285–94.

[51] Namaki A, et al. Network analysis of a financial market based on genuine correlation and threshold method. Phys Stat Mech Appl 2011;390(21–22): 3835–41.

[52] McDonough JE, et al. Gene correlation network analysis to identify regulatory factors in idiopathic pulmonary fibrosis. Thorax 2019;74(2):132–40.

[53] Kia AN, et al. Network-based direction of movement prediction in financial markets. Eng Appl Artif Intell 2020;88:103340.

[54] Benesty J, et al. On the importance of the Pearson correlation coefficient in noise reduction. IEEE Trans Audio Speech Lang Process 2008;16(4):757–65.

[55] Zhou D, et al. Learning with local and global consistency. Adv Neural Inf Process Syst 2003;16:321–8.

[56] Chi KT, et al. A network perspective of the stock market. J Empir Finance 2010;17 (4):659–67.

[57] Perry BL, et al. Co-prescription network reveals social dynamics of opioid doctor shopping. PloS One 2019;14(10):e0223849.

[58] Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. WIREs Data Min Knowl Discov 2012;2(1):86–97.

[59] Zarikas V, et al. Clustering analysis of countries using the COVID-19 cases dataset. Data in Brief 2020;31:105787.

[60] Gu Y, Wang C. A study of hierarchical correlation clustering for scientific volume data. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010.

[61] Liu X, et al. A correlation-matrix-based hierarchical clustering method for functional connectivity analysis. J Neurosci Methods 2012;211(1):94–102.

[62] Metsalu T, Vilo J. ClustVis: a web tool for visualising clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res 2015;43(W1): W566–70.

[63] Haghani M, et al. The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: a scientometric analysis and scoping review. Saf Sci 2020;129:104806.

[64] Glendon AI. Safety Science directions: the journal. Saf Sci 2021;135:105127.

[65] Varotsos CA, et al. Diagnostic model for the society safety under covid-19 pandemic conditions. Safety Science; 2021. p. 105164.

[66] Vuorinen V, et al. Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. Saf Sci 2020;130:104866.

[67] Tran BX, et al. Characterise health and economic vulnerabilities of workers to control the emergence of COVID-19 in an industrial zone in Vietnam. Saf Sci 2020; 129:104811.

[68] Najork MA. Comparing the effectiveness of HITS and SALSA. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management; 2007.