



OPEN ACCESS

EDITED BY

Naoki Masuda,
University at Buffalo, United States

REVIEWED BY

Prosenjit Kundu,
Dhirubhai Ambani Institute of Information
and Communication Technology, India
Matjaž Perc,
University of Maribor, Slovenia

*CORRESPONDENCE

Tuomas Takko,
✉ tuomas.takko@aalto.fi

RECEIVED 05 January 2023

ACCEPTED 24 May 2023

PUBLISHED 07 June 2023

CITATION

Takko T, Bhattacharya K and Kaski K
(2023), Modelling exposure between
populations using networks of mobility
during COVID-19.
Front. Phys. 11:1138323.
doi: 10.3389/fphy.2023.1138323

COPYRIGHT

© 2023 Takko, Bhattacharya and Kaski.
This is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Modelling exposure between populations using networks of mobility during COVID-19

Tuomas Takko^{1*}, Kunal Bhattacharya^{1,2} and Kimmo Kaski¹

¹Aalto University School of Science, Department of Computer Science, Espoo, Finland, ²Aalto University School of Science, Department of Industrial Engineering and Management, Espoo, Finland

The use of mobile phone call detail records and device location data for the calling patterns, movements, and social contacts of individuals, have proven to be valuable for devising models and understanding of their mobility and behaviour patterns. In this study we investigate weighted exposure networks of human daily activities in the capital region of Finland as a proxy for contacts between postal code areas during the pre-pandemic year 2019 and pandemic years 2020, 2021 and early 2022. We investigate the suitability of gravity and radiation type models for reconstructing the exposure networks based on geo-spatial and population mobility information. For this we use a mobile phone dataset of aggregated daily visits from a postal code area to cellphone grid locations, and treat it as a bipartite network to create weighted one mode projections using a weighted co-occurrence function. We fit a classical gravity model and a radiation model to the averaged weekly and yearly projection networks with geo-spatial and socioeconomic variables of the postal code areas and their populations. We also consider an extended gravity type model comprising of additional postal area information such as distance via public transportation and population density. The results show that the co-occurrence of human activities, or exposure, between postal code areas follows both the gravity and radiation type interactions, once fitted to the empirical network. The effects of the pandemic beginning in 2020 can be observed as a decrease of the overall activity as well as of the exposure of the projected networks. These effects can also be observed in the network structure as changes towards lower clustering and higher assortativity. Evaluating the parameters of the fitted models over time shows on average a shift towards a higher exposure of areas in closer proximity as well as a higher exposure towards areas with larger population. In general, the results show that the postal code level networks changed to be more proximity weighted after the pandemic began, following the government imposed non-pharmaceutical interventions, with differences based on the geo-spatial and socioeconomic structure of the areas.

KEYWORDS

collective human mobility, social physics, data-driven modelling, complex networks, COVID-19

1 Introduction

Studies on human mobility using large scale and longitudinal datasets from real world systems are useful for understanding human behaviour, designing techno-social systems, as well as forecasting different crisis scenarios like pandemics. Alongside with improved methods for obtaining novel data from mobile phones or other digital devices, human mobility as long range migration and short range trips has been an active field of research.

The use of mobile phone records data, social media data and other mobile applications has made it possible to capture large and longitudinal datasets of human locations, trajectories and behavioural patterns [1–3]. This data has motivated the modelling of various aspects of human mobility, such as patterns of their behaviour and trajectories [4, 5], co-locations [6, 7], agent-based systems [8–11], and migration [3, 12–16]. Networks are naturally useful for modelling and analysing mobility patterns, as the locations and trips, or entities and social ties, can be considered as nodes and edges, respectively (e.g., [17]). Networks of telecommunication and proximity have shown characteristic structural properties being present in human networks, such as small-worldness and the presence of communities [18].

Over the years numerous studies have demonstrated the relationship between the mobility and distance, such that the number of trips between two locations has a negative correlation with the distance between them. Similarly, when modelling the mobility between two regions, the population sizes have been shown to correlate with the number of trips between the said locations [15, 19]. These gravity type models, inspired by Newton's law of gravity, have also been applied to communications [19], social connections and trade [20]. Another commonly used model for describing human mobility is the radiation model, which considers the number of opportunities as the distance variable such that the number of opportunities within the radius of the distance between two locations denominates the probability of an action to another point [21]. Such models have been utilized to describe the use of services, job seeking and migrating. It should also be noted that there are numerous other external variables affecting the human mobility in addition to the commonly used population sizes and distances, such as available means and cost of transportation or government imposed restrictions [22, 23].

With the emergence of the SARS-CoV-2 pandemic in Finland in early 2020, the authorities imposed non-pharmaceutical interventions (NPIs) in order to contain the spread of the virus and keep the demand for healthcare and hospitalisations on a sustainable level [24]. This global pandemic sparked vast interest in studying different models of epidemic spreading [25] and efficiency of strategies regarding the restrictions [26, 27] as data on human behaviour, epidemiological outcomes and NPIs became available from numerous sources. The NPIs included such methods as social distancing [28], contact tracing [29] as well as restrictions like school closings and remote work guidelines [27, 30], which affect the overall daily behaviour of people, should they comply with them. In addition to the restrictions imposed by the government, the self-imposed social distancing of health-aware people had an effect on their mobility as well as on the spread of the virus. The changes in human mobility and behaviour, caused by restrictions or general awareness, and their effects on the spreading of the virus have been studied at the level of countries [31–34] and at a more microscopic level such as cities and grids [35–38]. Worth noticing is that the dynamics of epidemic spreading have been shown to be more complex than just the contacts caused by activity [39] and the estimated epidemiological effects of different NPIs have been shown to be model dependent [40].

Yet, utilizing mobile phones as sources of mobility and contact data for researchers and decision makers is highly necessary [41,

42]. The data obtainable from mobile phones ranges from call detail records (CDR) and GPS traces from applications to Bluetooth-based proximity capturing. While the most commonly used data source for human mobility has been the CDR data, many studies have used the activity-type or location aggregated data from companies such as SafeGraph, Google, Apple or Meta (see [7, 31–33, 37, 43, 44]) to estimate the effect of mobility in various countries and regions on the spread of COVID-19 and on NPIs. For instance, Fritz and Kauermann [44] investigated the effects of the co-location rates and friendship networks on the epidemiological outcomes in Germany using mobility data from Facebook users. The authors utilized regression models to investigate the connection of mobility, social ties and the spread of the virus. From the analysis, the authors found that the co-location of an area was the primary aspect in the transmission rate of disease, reinforcing the importance of restrictions limiting mobility between areas. In addition to the mobility and transportation, there are density, demographic, environment and infrastructure related factors that have been shown to have significance in the spreading of the virus (see [45]), such as the population size and population density of urban areas [46, 47] or household size [48].

In this study we investigate the human behaviour before and during the COVID-19 pandemic in Finland using the daily activities as a proxy for estimating the possible contacts between the populations living in different postal code areas. The overarching objective of this study is to use modelling to show large scale changes in exposure induced by human mobility during the pandemic using postal area level networks. Our first research objective is to use the aggregated daily activity data for creating models of possible exposure networks (or contact networks). The second objective is to quantify the changes in this system during the first 2 years of the pandemic in comparison to the preceding prepandemic year 2019. Finally, we aim to show a relationship between the government imposed NPIs and the exposure networks, using data from the Oxford COVID-19 Government Response Tracker (OxCGRT) [49].

In contrast to previous studies utilizing mobility and short term activity data, in this study we seek to obtain new insight to the relationship between the activity and NPIs by using postal code and cell grid level activity data with daily time resolution from a type of dataset that is not commonly used. This data contains more micro level activity than the community level reports used in some studies, but less information than the individual mobility traces used in some other studies such as [7]. The data we use is obtained from Telia, a telecommunications company operating in Finland, and it contains aggregated daily activity data of mobile phone users. Similar data has been used in other studies such as [21, 38, 50, 51], but not in the same type of modeling paradigm. The methodology of using gravity and radiation type models for predicting the movement of people between two areas with varying population, number and diversity of services [52] has been investigated with different datasets of CDR or GPS locations in the form of origin-destination networks [53, 54]. Our study considers a bidirectional relationship between postal code areas, where the populations in these areas do not require visiting the other's geographical area to form an edge. Thus, the bidirectional edges represent the interaction of exposure in a broader setting than directed trips between an origin area and a destination area.

The rest of the manuscript is divided into four sections, Methods, Results, Discussion, and Conclusions. First, we present the daily activity data along with the pre-processing steps done to make the data into sets of networks for periods of time. From these networks we construct two types of models, a gravity-type model and a radiation-type model, by using postal code area level data. In the Results section we present the findings on the fitting of the two classical models and the changes in the resulting networks over the course of years 2019–2021. In the Discussion section we consider the implications of our findings and discuss potential future research. In the Conclusions section we present overall concluding remarks.

2 Materials and methods

In this section, we describe the empirical dataset and explain the construction of the weighted bipartite projections, the measures used to investigate the temporal changes in the network and finally the proposed model for predicting the likelihood of social connections between areas using geospatial and socioeconomic information.

2.1 Data

The data used in this study is obtained from Telia, the second largest telecommunications company operating in Finland with a $\approx 30\%$ market share. This data, called “Crowd Insights” [55], contains aggregated human activities for each day in the format of home and grid location matrix. In this study we define an activity as a temporary stay in a location during a given day. The data is based on the phone location calculated from the signal to the cell towers and the data is extrapolated by the company from the population of customers to represent the whole population. We make the assumption that the observations from the population were done uniformly. The difference to the CDR data is that in order to record an activity, the user does not need to perform an interaction such as a call. Thus the data is not event-driven, but “sightings data” [41]. Each mobile phone user’s home location is marked with its postal code area, where the user stayed longest before 9 am, i.e., the place the user woke up in the morning. The activity of each user is recorded if the user spent at least 20 min during the day in that particular grid location. This data is then anonymised and aggregated for each postal code area, thus generating a matrix where during 1 day a postal code area has an amount of visits or mobile activities to each grid cell. The grid was defined by the telecom areas, ranging in size from 0.5 by 0.5–2 by 2 km depending on the density of users in the area. In order to protect the privacy of less populated areas, activities with less than 5 people are redacted from the dataset. As the data is aggregated, we are not considering the explicit flow of people between two areas or the trajectories of individual users. The dataset contains the activities during the periods 1.1.2019–23.9.2019 and 1.2.2020–31.3.2022, with some days being removed due to low signal quality. For reducing the possible inaccuracy caused by the incomplete yearly sets and the seasonal trends therein, we create uniformly sized subsets where each date is contained in the set of each year. By filtering these

subsets, we obtain 219 dates for each year from 2019 to 2021. These subsets are referenced as yearly sets in this manuscript.

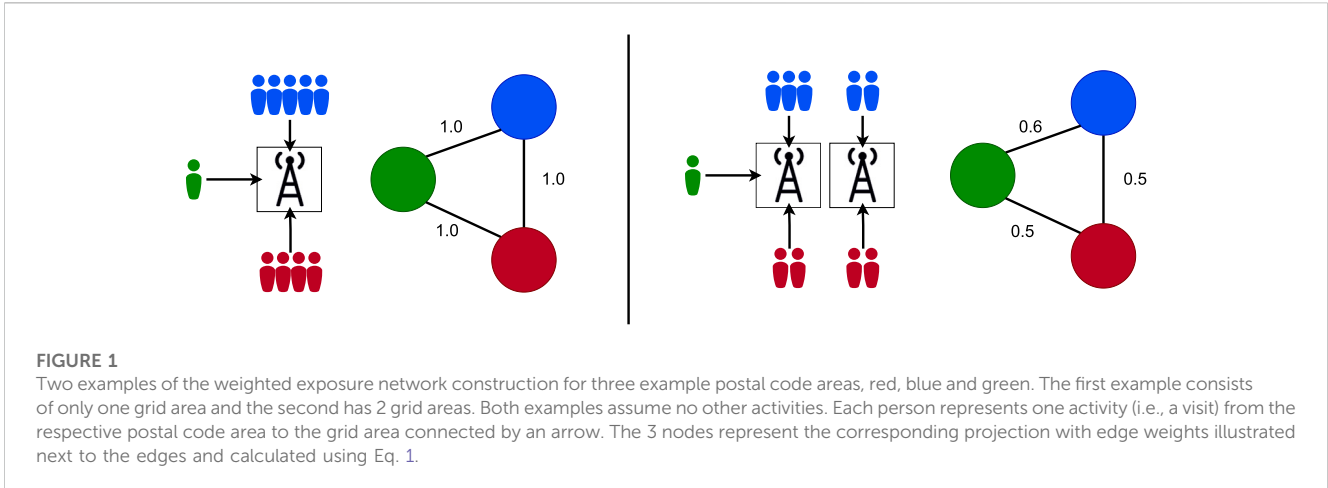
In this study we consider a subset of the data by choosing only the activities that are both stemming from and occurring within the capital area of Finland, meaning the municipalities of Helsinki, Vantaa, Espoo and Kauniainen. In other words, an accepted activity is performed by a person waking up in the chosen geographical area and performing the activity in the grids of the area. Thus, we consider the capital area as a separate subsystem from the rest of the country, even though the data contains people moving in and out of the system. Postal code areas in the chosen geographical area are highly populated, but also smaller in size than the average postal code areas in the country.

Filtering the data results in 167 postal code areas as home areas and 1444 grids as the areas for performing mobile activities. This data can be presented as a bipartite network, where for each day we consider the home areas and grid areas as separate classes of nodes and the links are the number of people who have performed the mobile activity. In the scope of this paper, we are investigating the difference in human behavior during the years 2019–2022 by constructing networks of the postal areas and analyzing it at the graph level using various time windows for the activity aggregation. The aim is to use the commonalities between activities of various areas as a proxy for contact as well as functional activities such as work and thus form a network of connected postal areas.

In addition to the activity data, we use socioeconomic data on the same postal code area level as variables in the model. The data is obtained from Statistics Finland [56] and is openly available. In this study we are not considering the changes in the socioeconomic values for each postal code area as the changes can be considered to be minor during the timespan of the study and the reporting of postal code areas changes between the years in the Statistics Finland data. As the final source of information we use the OxCGRT dataset [49] for obtaining a numerical index on the level of mobility and other activity related restrictions in Finland during the pandemic.

For the analysis of the daily contact networks between the postal code areas we calculate a set of graph theoretical properties with the intention of observing changes and quantify some of the characteristics of the networks. Daily population activities follow a weekly schedule as well as public holidays and holiday periods, resulting in relatively high variance throughout the year. We address this by analyzing the network properties on daily networks as moving averages as well as networks aggregated as averages of 7-day periods or as years. The dataset contains gaps, namely, the time between 24.09.2019 and 31.01.2020, which we address by comparing only the common dates during each year when doing comparisons between the years. Otherwise these empty periods of time are visualized as discontinuities.

For measuring the changes in the network and reinforcing the reasoning behind the models, we calculate the average clustering coefficient, assortativity of degree correlations and weighted degree distributions. Then we compare the resulting weighted networks to the distances and population sizes used in Eq. 2 and Eq. 3 without fitting the data beforehand. For the calculation of the average clustering coefficient and the degree assortativity we utilize the NetworkX library [57] in Python 3.7.



2.2 Models

In this section we describe first the process of constructing the networks of exposure from human mobility using the above described empirical data and then fitting two types of models to simulate the interactions between postal code areas, i.e., the nodes of the network. We use the edges of these projected networks for fitting a gravity type model and a radiation type model using postal code level population data. The changes in the structure of these networks over time are analysed using standard graph theoretical measures.

2.2.1 Bipartite network projection

The temporal activity data forms a bipartite network in the form of an adjacency matrix A of size 167×1444 , where each postal area (row) has a number of activities in each geolocated grid (column). From this matrix we calculate a weighted one mode projection by defining the edge weight function $W_{i,j}$ for the postal code areas i and j as the sum of products over each grid cell:

$$W_{i,j}(t) = \frac{\sum_g w_i^g(t) w_j^g(t)}{\sum_z w_i^z(t) \sum_z w_j^z(t)}, \quad (1)$$

where $w_a^g(t)$ is the number of activities, i.e., people visiting the grid cell g from the postal code area a at time t and the denominator is the sum of all activities from the corresponding postal code area. The multiplication ensures that both postal code areas, i and j , have activity in the particular grid cell during that day. The weight $W_{i,j}$ is bounded between 0 and 1, where 0 would indicate no co-occurring activities and value 1 would indicate full co-occurrence to a single grid cell, e. g., as depicted in Figure 1. A small value indicates a low probability for picking a random activity from each postal code area such that both of the activities were performed in the same grid cell. This way we obtain a weighted projection between the postal code areas that can be used as a proxy for contacts between the postal code areas. Examples of the edge weight function are illustrated in Figure 1. Using this method for the duration of the empirical data we obtain $S = 1031$ daily networks, in which the nodes are the only permanent structure.

We make two assumptions when considering these projections as exposure between postal code areas. First we assume that all the

grids are weighted uniformly, i.e., the grid size is not considered. We justify this with the fact that the vast majority of activities taking place in the grids with the smallest size and the largest grid cells are mostly recreational areas such as parks. Also, as the privacy of the mobile phone users is preserved by omitting activities made by less than 5 persons from a postal code area during the day, the effect of these less active grid locations is not significant. Secondly, we make the assumption that the activities to a grid are uniformly spread throughout the day across the postal code areas, i.e., the active population does not change during the day.

2.2.2 Modelling networks of exposure

The main research goal of this paper is to create a characteristic model for predicting and understanding the resulting bipartite network projections (Eq. 1) that represent the contacts between the postal code areas due to the co-occurrence of individual mobile activities or mobilities. For this we use two classical models: a gravity-type model and a radiation-type model [8, 11, 19, 21]. In case of the gravity-type model, we consider the exposure between populations in two postal code areas, $f_g(i, j) = W_{i,j}$, to be dependent increasingly on the size of the two populations and decaying on the distance between them, as follows

$$f_g(i, j) = C \frac{(p_i p_j)^B}{d_{i,j}^{B_d}}, \quad (2)$$

where C is a constant, B is the parameter specific exponent, p_i and p_j are the population variables of the postal code areas i and j , and $d_{i,j}$ is the “crow-flight” distance between the centroids of the land areas (excluding waters) of these postal areas, measured in meters. This is an assumption since the population is not evenly distributed and the real world distances can be longer due to the transportation infrastructure. As in literature we consider our gravity-type model symmetrical, i.e., the exposure or the individual mobility is the same from i to j as from j to i .

The second model we evaluate with the empirical networks (Eq. 1) is based on the radiation-type models, which in general are formulated as directed interactions. However, for the problem of this study we consider the interactions symmetrical as in the gravity-type model introduced above. The model is based on the notion of

TABLE 1 The fitted coefficients for the yearly average networks. The means are obtained by fitting the models for the full yearly averaged networks with postal code areas of population less than 500 being removed from the network. The number in brackets is the standard error for the coefficient. SMAPE and RMSE measure the error of the fitted model and lower number indicates a better fit.

		2019	2020	2021
Gravity	<i>C</i>	0.4949 (±8.3*10 ⁻²)	1.7864 (±9.2*10 ⁻²)	1.9716 (±9.8*10 ⁻²)
	<i>B_p</i>	0.0212 (±4.0*10 ⁻³)	0.0316 (±4.0*10 ⁻³)	0.0575 (±4.0*10 ⁻³)
	<i>B_d</i>	0.3913 (±2.0*10 ⁻³)	0.4980 (±3.0*10 ⁻³)	0.5412 (±3.0*10 ⁻³)
	SMAPE	0.1618	0.1788	0.1891
	RMSE	1.331*10 ⁻³	1.337*10 ⁻³	1.398*10 ⁻³
	<i>R</i> ²	0.6	0.665	0.674
Radiation	<i>C</i>	-0.9098 (±8.8*10 ⁻²)	0.3904 (±9.5*10 ⁻²)	0.4898 (±0.102)
	<i>B_p</i>	0.0608 (±4.0*10 ⁻³)	0.0816 (±4.0*10 ⁻³)	0.1117 (±4.0*10 ⁻³)
	<i>B_r</i>	0.1256 (±1.0*10 ⁻³)	0.1672 (±1.0*10 ⁻³)	0.1824 (±1.0*10 ⁻³)
	SMAPE	0.1810	0.1937	0.2055
	RMSE	1.145*10 ⁻³	1.358*10 ⁻³	1.401*10 ⁻³
	<i>R</i> ²	0.5	0.605	0.618
E.Gravity	<i>C</i>	0.8885 (±0.083)	2.513 (±0.0965)	2.7746 (±0.103)
	<i>B_p</i>	-0.0605 (±4.0*10 ⁻³)	-0.0092 (±5.0*10 ⁻³)	0.0190 (±5.0*10 ⁻³)
	<i>B_{p̂}</i>	0.0855 (±3.0*10 ⁻³)	0.0382 (±3.0*10 ⁻³)	0.0352 (±3.0*10 ⁻³)
	<i>B_d</i>	0.4774 (±9.0*10 ⁻³)	0.7085 (±0.01)	0.7815 (±0.011)
	<i>B_{d̂}</i>	0.4774 (±9.0*10 ⁻³)	0.3331 (±0.011)	0.3532 (±0.012)
	SMAPE	0.1465	0.1684	0.1804
	RMSE	1.240*10 ⁻³	1.297*10 ⁻³	1.364*10 ⁻³
	<i>R</i> ²	0.668	0.697	0.703

number of possible opportunities within the radius equal to the distance between two areas centered on the source area. Here the radiation model for exposure, $f_r(i, j) = W_{i,j}$, is constructed using the following modified form of the standard model:

$$f_r(i, j) = \frac{(p_i \cdot p_j)^{B_p}}{[(p_i + r_{d_{i,j}}) \cdot (p_j + r_{d_{j,i}}) \cdot (p_i + r_{d_{i,j}} + p_j) \cdot (p_i + r_{d_{j,i}} + p_j)]^{B_r}} \tag{3}$$

where $r_{d_{i,j}}$ is the sum of population variable within the radius $d_{i,j}$ from the area i excluding the populations in areas i and j , and B is the parameter used for fitting the model to the data. The population variable under radius, $r_{d_{i,j}}$, was obtained by calculating the fraction of the area under the radius $d_{i,j}$ for each postal code area and multiplying each population variable by the fraction of area covered. Here it is assumed that the population, services or workplaces are evenly distributed. We use this approximation due to lack of more detailed data. Prior to fitting the models, we filter out the postal code areas with less than 500 inhabitants, which from our empirical dataset removes only seven areas that are mostly uninhabited industrial zones.

Both the gravity and radiation type models are fitted to the weights of the edges of the empirical 7-day average networks and

the yearly average networks with a linear regression utilizing ordinary least squares (OLS) for obtaining the parameter values. To use a linear regression model we change the models in logarithmic form such that the gravity type model is of the form

$$\log(W_{i,j}(t)) = B_p \cdot \log(p_i \cdot p_j) - B_d \cdot \log(d_{i,j}) \tag{4}$$

while the radiation type model is of the form

$$\log(W_{i,j}(t)) = B_p \cdot \log(p_i \cdot p_j) - B_r \cdot \log[(p_i + r_{d_{i,j}}) \cdot (p_j + r_{d_{j,i}}) \cdot (p_i + r_{d_{i,j}} + p_j) \cdot (p_i + r_{d_{j,i}} + p_j)] \tag{5}$$

The statistics of the fitted exponents (B_p , B_p , B_d , B_r) are shown in Table 1 and in Figure 5 the values are visualized over the time period of our dataset.

After fitting and evaluating the gravity and radiation models, we investigate whether combining additional geospatial and population level variables to the gravity type model would give a better estimate of the edges of the network. These additional variables were chosen by evaluating the edges with largest error. The chosen variables are the population density, obtained from the socioeconomic Statistics Finland dataset [56], and a simple distance via public transportation (i.e., train and metro). The distance via public transportation was obtained by using the geographical coordinates of each train and metro station and

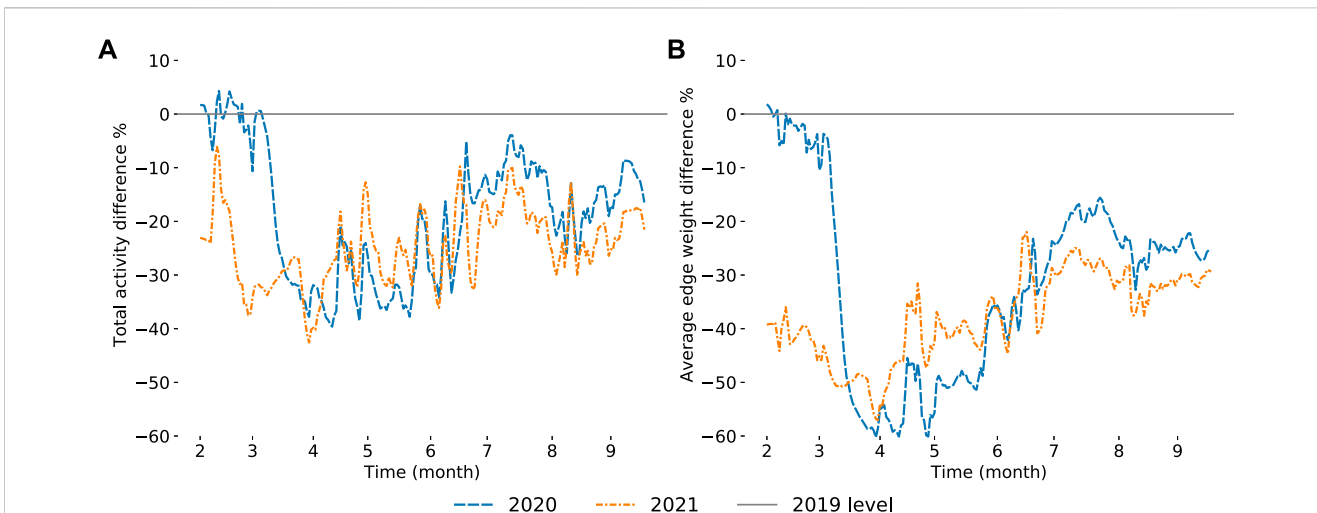


FIGURE 2
 Proportional percentage difference of the total activity in the empirical bipartite networks (A) and of the average edge weight in projected exposure networks (B) between the overlapping dates for the years 2020 and 2021 to the pre-pandemic year 2019 as 7-day moving averages. The differences to the pre-pandemic year are more pronounced in the projected exposure networks (panel B) than in the total activity (panel A) and show less fluctuations throughout the overlapping dates. The change in panel B indicates a topological change in the bipartite network, i.e., the places people visit. Self-loops were excluded from the projected exposure networks.

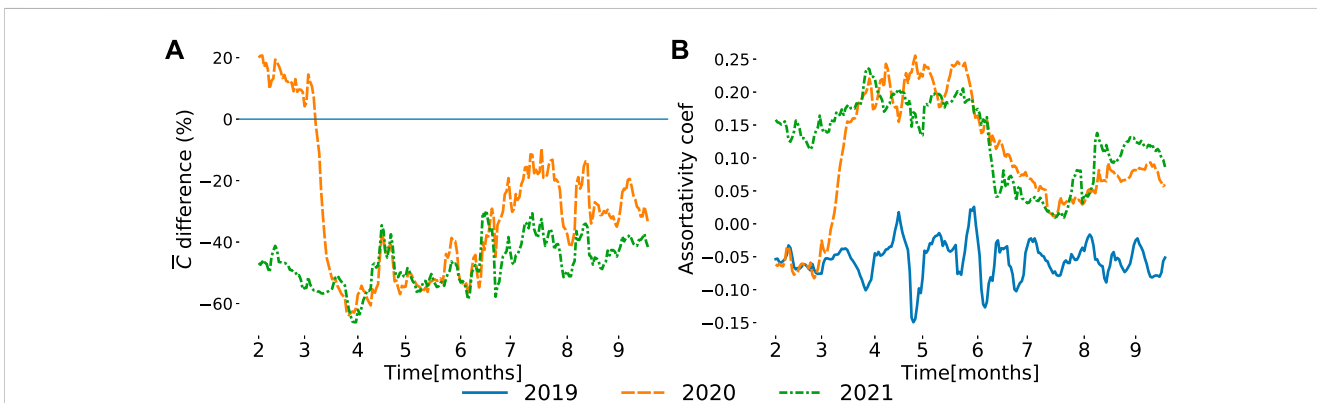


FIGURE 3
 7-day moving averages of measures of the projected networks during the yearly overlapping dates in the dataset: (A) Average weighted clustering coefficient (C) as proportional difference to the same time in the year 2019, (B) weighted degree assortativity correlation coefficient. The networks in both (A) and (B) are filtered by edge weight to have the 90-percentile of edges based on edge weight. The results indicate changes in the topology of the exposure networks, namely, in the edges between nodes of low weighted degree and high degree hubs, such as the postal code areas located in the proximity of central locations. This can be seen in the disassortativity during the pre-pandemic year changing towards assortativity during the pandemic.

their interconnections within the four municipalities. For each pair of postal code areas we calculate the distance from the centre of the area to the nearest station and then calculate the shortest path length via public transportation between the corresponding nearest stations. In the case both areas have the same nearest station, the distance is set as the distance between the two area centres. The final form of this extended model $f_e(i, j)$ is as follows

$$f_e(i, j) = C \frac{(p_i \cdot p_j)^{B_p} \cdot (\hat{p}_i \cdot \hat{p}_j)^{B_p}}{d_{i,j}^{B_d} \cdot \hat{d}_{i,j}^{B_d}}, \tag{6}$$

where \hat{p} is the population density and \hat{d} is the distance via public transport. The extended model was fitted similarly to Eq. 4 and Eq. 5.

3 Results

In this section, we will construct the projected empirical networks, analyze them in terms of general attributes and their time-dependent structural changes as well as fit the above introduced three models to the data. In total, the dataset contains 1031 days of individual mobile activities in the original grid cells which we project to daily postal area level weighted exposure networks using the formula in Eq. 1. The aggregated networks were constructed such that each edge weight was the mean of the corresponding edge within the chosen time frame. The weekly averaged networks were constructed as 7-day consecutive bins and the yearly averaged networks are

TABLE 2 Correlation coefficients between weekly networks, NPI indices and fitted model parameters. \bar{W} is the mean edge weight and $\sum_i \sum_g w_i^g$ is the total activity in the unprocessed bipartite network. The mean NPI stringency score is calculated from the subindices as described in Section 3.1.

	Variable	\bar{W}	$\overline{NPI_c}$
Data	\bar{W}		-0.8235
	$\sum_i \sum_g w_i^g$	0.7146	-0.6682
Gravity	B_p	-0.9206	0.6997
	B_d	-0.9859	0.8299
Radiation	B_p	-0.9474	0.7358
	B_r	-0.9851	0.8436
E.Gravity	B_p	-0.9710	0.8118
	$B_{\bar{p}}$	0.9052	-0.8147
	B_d	-0.9736	0.8267
	$B_{\bar{d}}$	-0.1005	0.0989

constructed using overlapping dates during the years 2019, 2020 and 2021 in the data (dates from 01.02. and 23.09.). With this binning we obtained 146 weekly networks and the yearly overlapping dates contain 219 days, which were used in the aggregated yearly networks.

In Figure 2 the total daily activity and the average edge weight of the pandemic years 2020 and 2021 are visualized in comparison to the same dates during pre-pandemic year 2019 as a moving average. Both these measures show similarities in terms of the overall difference to the pre-pandemic year. Both the seasonality and public holidays can be seen as peaks in the overall activity, whereas the same peaks are not as prominent in the average edge weight. Also worth noticing is the proportional difference, as the total activity at points exceeds the 2019 level, the average edge weight does not. Comparing the time series of the activity measure and the average edge weight to the average containment index from the NPI dataset [49] shows a negative correlation as would be expected if the population is compliant or epidemic aware. Interestingly the correlation for the average edge weight is stronger than for the total activity, as indicated by the correlation coefficients ≈ -0.76 and ≈ -0.56 , respectively. The correlation coefficient between the same measures is ≈ 0.72 . All correlation coefficients are shown in Table 2. The resulting aggregated yearly networks and the related weighted degree and edge weight distributions are shown in Figure 4. As can be expected from the proportional comparisons in Figure 2, the means of the distributions shift towards 0 during the pandemic. The highest weighted degree becomes also lower, i.e., ≈ 0.873 to ≈ 0.622 and ≈ 0.603 . For clarity of visualization, the yearly networks are depicted using the four strongest edges for each node with the node width and colour representing the weight in Figure 4. A visually noticeable difference is the decrease in the number of longer distance edges to the geographical center (central Helsinki) taking place after 2019. Utilizing the Clauset-Newman-Moore greedy modularity

maximization [58] for clustering the postal code areas to communities in the aggregated yearly networks using the 95-percentile of the edges based on weight shows that the number of detected communities declines from 16 during pre-pandemic 2019 to 4 and 5 during pandemic 2020 and 2021.

Next we measure the average clustering coefficient and degree assortativity in the filtered 90-percentile graphs. The time series are shown in Figure 3. The proportional difference in the weighted average clustering coefficient in Figure 3A shows similar change as the average edge weight. The Pearson weighted degree assortativity, depicted in Figure 3B, increases during the pandemic, which indicates that the edges in the filtered aggregate networks during the pandemic are more often found between postal code areas with similar weighted degree. This effect can also be seen in the network visualizations in Figure 4, where the strongest edges are no longer between the central area and areas further from it. Also, this correlates with the notion of larger communities detected during the pandemic. The result remains uniform when replacing the edge weight with 1 in the filtered network.

3.1 Evaluating the models

For the performance evaluation of our three models we first fit them to the aggregated yearly networks with the empirical data. Then we perform the evaluation by calculating the errors using symmetric mean absolute percentage errors (SMAPE) and root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum^n (W_f - W_o)^2}{n}}, SMAPE = \frac{1}{n} \sum^n \frac{|W_f - W_o|}{|W_f| + |W_o|}$$

where n is the number of edges, W_f is the predicted edge weight and W_o is the observed edge weight. The results for the fitted exponents and errors are shown in Table 1. The predicted values obtained for the radiation, gravity, and extended gravity models are plotted against the observed values for the year 2019, depicted in Figures 5A–C, respectively. As anticipated by the correlations to the unfitted models, the gravity model performs better than the radiation model in predicting the edge weights and the extended gravity model performs the best. Same can be seen in the models' R^2 values, which indicate the proportion of variance explained by the variables. All the three models have R^2 values over ≥ 0.5 in 2019 with interestingly an increase during the pandemic years. The difference in proportional error (SMAPE) between the three models remains consistent in all three aggregated yearly networks. The values of RMSE show that in the pre-pandemic year 2019 the radiation model shows the lowest error, but in the subsequent years the error shows an increasing trend for all three models. The error metrics over time for the weekly networks are shown in Supplementary Figure S2.

The fits of the models in Figure 5 illustrate the error. All three models overestimate the weakest edges and underestimate the strongest edges with varying magnitude, the radiation model having the largest error. While the absolute values between fitted exponents of the same model are not comparable due to the different units in the variables, the proportional differences show changes in

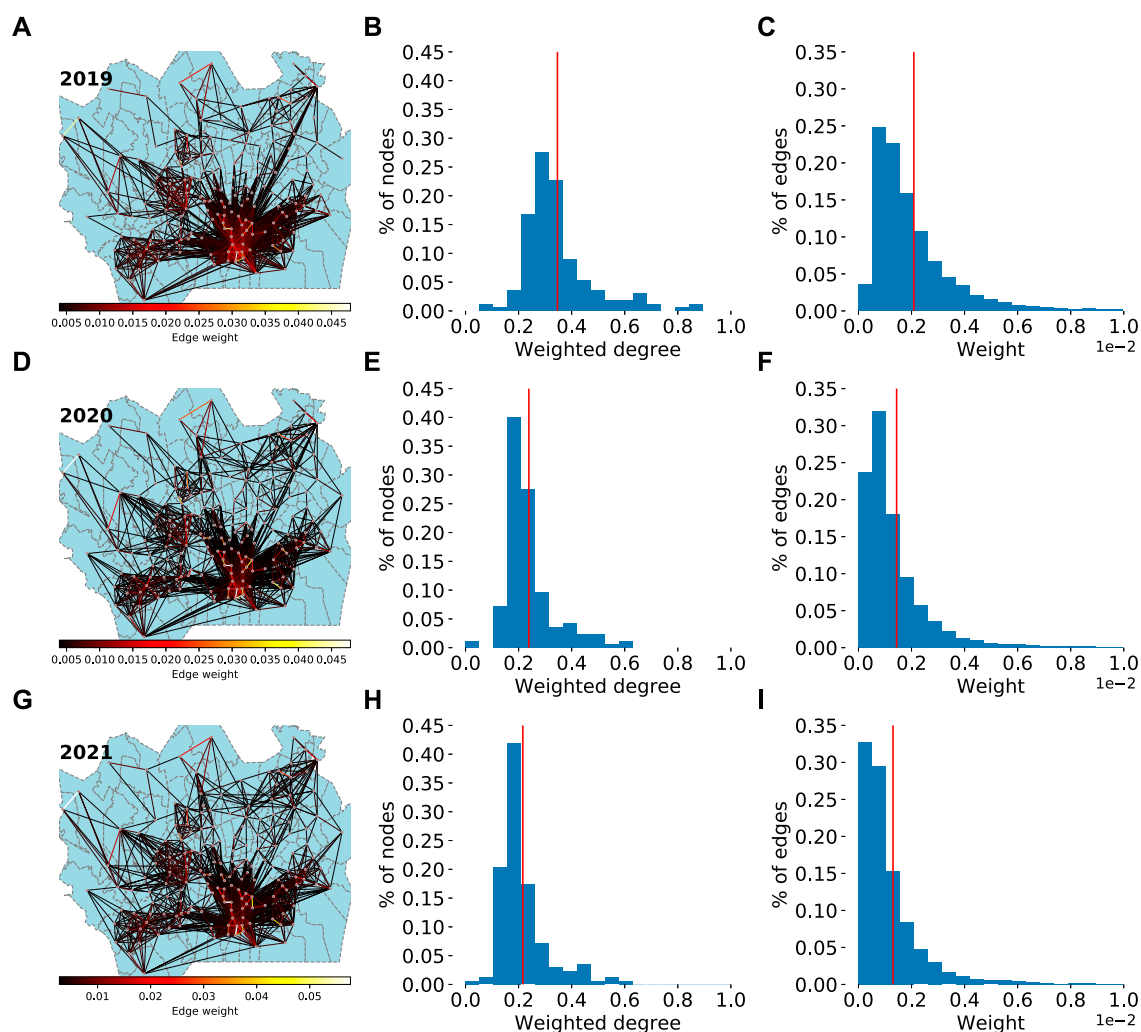
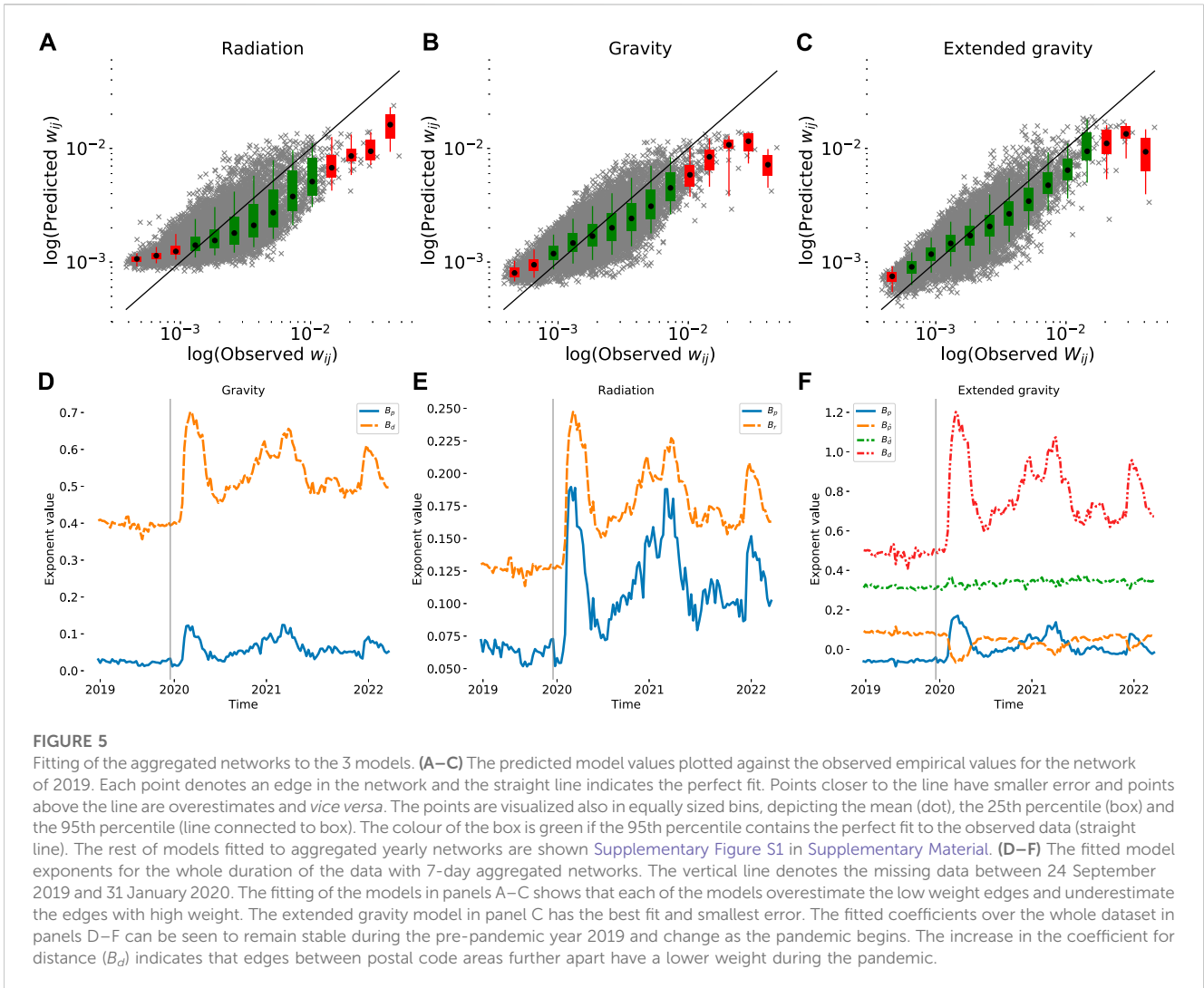


FIGURE 4

The yearly projected networks with the mean edge weights for the years 2019 (A–C), 2020 (D–F) and 2021 (G–I). In sub-figures A, D and G the network is shown as 90th percentile of the edges filtered by edge weight. The width and colour depict the weight of the edge and the node size depicts the weighted degree. The histograms in sub-figures B, E and H depict the weighted degree distribution of each yearly network. The histogram has 20 bins and the red vertical line depicts the mean. The rightmost histogram depicts the distribution of edge weights. The networks show that during the pandemic years the edges between the nodes in the city center and the nodes in in the surrounding areas are reconfigured to nodes in closer proximity of each area. For instance, the postal code areas in the upper right corner are more connected within themselves in 2021 when comparing to the pre-pandemic year 2019. The weighted degree and the average edge weight distributions can be seen to shift towards 0 during the pandemic.

the aggregated networks. The exponents of gravity model indicates a larger significance on the exponent for population B_p after 2019 as it increases in proportion to the distance exponent B_d . The distance exponent shows the previously discussed weakening of longer distance edges by increasing with each year. The fitted exponents of the radiation model show a similar change with the increasing weight of B_p . The exponent for population within radius, B_r , increases but not in proportion to B_p . When including the population density in the extended gravity model, the fitted population size exponent becomes negative in 2019 and increases during the pandemic while B_p decreases. B_d increases similar to the basic gravity model, decreasing the weight of long distance edges. The exponent for the distance via public transportation, B_j remains in the same order of similar magnitude throughout the 3 years, i.e., from prepandemic 2019 till pandemic 2020 and 2021.

After fitting and analyzing the yearly aggregated networks, we investigate the full duration of our dataset with the weekly aggregated networks. The fitted exponents over the course of the dataset are shown in Figure 5. The gravity model over time shows the two exponents B_d and B_p behaving similarly, while their relative difference changes. Similar shape can be seen in the case of radiation model but with more extreme relative changes. In the case of the extended gravity model the two additional variables show a negative relationship to exponents of the basic gravity model. The point in time when the exponent B_p becomes positive can be seen to occur at the beginning of the pandemic. Comparing the fitted exponents to the average edge weight over time shows high correlations as can be expected. Consequently, most of the exponents share a high correlation to the NPI restriction index. The outliers in terms of the correlation strength is the added



variable \hat{d} in the extended gravity model. These correlations are listed in [Table 2](#).

In order to investigate the effects of particular types of NPIs on the extended gravity model coefficients over time, we fit the 9 sub-indices related to the restriction stringency in the OxCGRT dataset [49] to each of the fitted exponents (B_p , B_d , B_s and B_a) of our extended gravity model (6) using an OLS regression, where the dependent variable is a model exponent and the independent variables are the set of stringency subindex scores:

$$B_i(t) = C + \beta_{H1} \cdot S_{H1}(t) + \dots + \beta_{C8EV} \cdot S_{C8EV}(t), \quad (7)$$

where $B_i(t)$ is the corresponding exponent from the extended gravity model at time t , β_i is the coefficient for the corresponding NPI index S_i and $S_i(t)$ is the stringency subindex score at time t . The NPI stringency subindices fitted to these exponents are: H1 Public information campaigns, C1M School closing, C2M Workplace closing, C3M Cancel public events, C4M Restrictions on gatherings, C5M Close public transport, C6M Stay at home requirements, C7M Restrictions on internal movement and C8EV International travel controls. The stringency subindex scores with values between 0 (no restrictions) and 100 (the most

stringent restrictions) are calculated from the OxCGRT data as instructed by the authors. For details on computing the subindex score, see equation 2 in the article by Hale et al; [49]. To match the weekly aggregation of the fitted exponents, we consider each weekly data point as the average stringency subindex score. The results show varying coefficients and p -values for each different NPI sub-index depending on the exponent they were fitted to. Across the four models fitted extended model's exponents, the stringency subindex for restrictions on gatherings has a constantly low p -value and standard error. In case of school and workplace closures, restrictions on internal movement and cancellation of public events also show significant p -values (< 0.05). In addition, the closures on public transport have an effect on B_d , as expected. The predicted values of each of the exponents of our extended gravity model are shown in [Supplementary Figure S3](#) and the related coefficients in [Supplementary Table S1](#).

Finally, comparing the predicted networks to the corresponding observed networks reveals the lacking aspect of gravity and radiation type models in the context of networks. The previously discussed errors and model fits indicate that the models capture the aspect of exposure between postal code areas. This can also be seen when

comparing the average edge weights of the predicted networks and observed ones, as the differences between the two values are of the order of 10^{-14} . However, a difference arises when comparing the average weighted clustering coefficient and the average weighted degree assortativity coefficient. The model produces networks with higher assortativity and clustering than observed in the empirical networks.

4 Discussion

In Finland the SARS-CoV-2 pandemic started in early 2020 having a visible effect on the activities of the population like their mobility, due to the government-imposed restrictions and self-imposed social distancing. This can be seen in [Figure 2](#) as a reduction in the overall activity in the bipartite networks and as a significant reduction in the average edge weight in the projected exposure networks during the pandemic. Notably, we can see that higher level of activity does not explicitly mean higher exposure between the postal code areas as implied by the Spearman correlation between the two values being ≈ 0.621 with p -value < 0.01 . Comparing the overall activity and edge weight to the NPIs in the form of an index, we see that the average edge weight has a higher correlation to the NPI stringency index than the activity has (-0.824 and -0.668). High correlation to the NPI stringency index could be seen as high compliance to the restrictions as well as general awareness of the public, resulting in lesser exposure. Moreover, the lesser amount of fluctuations, seasonal or other, in the average edge weight and the related fitted exponents of the models hint that the underlying behaviour remains stable during “normal” pre-pandemic times but changes once the restrictions are imposed. The changes in the network attributes such as assortativity depict the reorganization where instead of having more exposure to hubs of high weighted degree, the exposure is distributed more evenly across the postal code areas, as the assortativity shifts towards positive values. Similar trend can be seen in the weighted clustering coefficient, shifting to lower values when the pandemic sets on.

The proposed applications of gravity and radiation type models are shown to be sufficient in estimating the exposure between the postal code areas, with the application of gravity type model performing better than the radiation type model. As can be expected, the prediction error decreases once the gravity model is fitted with additional variables of population density, \hat{p} , and public transport distance $\widehat{d}_{i,j}$. In addition to changing exponent weights, the error also increases with increasing restrictions on human mobility (see [Supplementary Figure S2](#)). While the error values can be seen to increase, also the R^2 values show an increase, which we suspect being caused by a higher number of extreme outliers. Fitting the models over the full duration of our data shows that the exponents ([Figure 5](#)) correlate with the average edge weight as well as the NPI index. These changes in the exponents can be interpreted as shifting weight between distance and population size. Interestingly, in the extended gravity model the exponents for population size and population density show an anti-phase like behaviour. The significance of higher population size and density in resulting to higher exposure in the models shares similarities to the results

of the studies considering urban scaling and COVID-19, where the contact rate of the virus was shown to scale with population size and density in urban areas [[46](#), [47](#)]. In both gravity type models, the weight for longer distances decreases (see [Table 1](#)), which we also confirmed manually from the empirical data. The distances larger than the nearest vicinity show a comprehensive reduction and the longest distances remain at the low level. Of this finding the government imposed restrictions and guidelines such as social distancing and remote work could explain the lower weight for long distance edges as people eligible for remote working would not be commuting. Temporary closures of restaurants and other services can also have a similar effect as the number of trips decreases [[59](#)]. The results in both projected networks and in the fitted model exponents showed a high correlation with the general NPI stringency index, and when fitting the series of exponents of the extended gravity model to the subcategories of restrictions we see that certain restrictions affect certain exponents more than others. The coarse model in itself provides an estimate where the extended gravity model’s exponents would be given a certain set of restrictions. As mentioned before, using such exponents for the model yields a good estimate for the mean edge weight, but overestimates the network properties when compared to the observed networks.

The independent variables used in the extended gravity model, namely, population size, distance, population density and distance via public transport, share similarities and thus we tested for multicollinearity by using the variance inflation factors (see [Supplementary Table S2](#)). In the extended gravity model the variance inflation factors were found to be within acceptable bounds, i.e., in the range from 1.65 to 3.02, and indicating a moderate level of correlation. The exposure networks set certain limitations to the models applied. First of all the models cannot estimate the “self-loops” in the network as the distance in the denominator would be zero. However, the weight of these self-loops can be obtained from [formula 1](#). We have included a visualization similar to [Figure 3](#) of these self-loops over the duration of the dataset in the [Supplementary Material](#) of this manuscript (see [Supplementary Figure S4](#)). Similar to the average exposure (see [Figure 2B](#)) these values show a relative change when the pandemic begins, compared to the corresponding time in 2019. Interestingly, the trend is opposite, indicating that the probability of picking two activities from a postal code area increases as the exposure to other areas decreases due to people visiting fewer individual places and having less activity in general. In the visualized distributions of yearly aggregated self-loop weight the weights can be seen to shift towards higher exposure. Secondly, the exposure networks impose the limitation of using certain models. The values of exposure are bound between 0 and 1 and thus other regression models, such as generalized linear models like Poisson regression, commonly used for count variables are not readily applicable without transformations that are dependent on the data and as such lose interpretability. Testing the models with Breusch-Pagan test showed that heteroskedasticity is present in the three models. The formulation of the three models does not permit the inclusion of zero observations. Conceptually, in a network these zero observations mean that the corresponding edge does not exist. However, the number of such observations

in the data is low. The yearly aggregated networks contain 4, 11 and 9 zero observations for the years 2019, 2020 and 2021, respectively, out of the 13,861 possible edges.

The mobile phone location based dataset used in this study has two limitations that could cause some inaccuracies. Firstly, the data defines the home location of each person as the postal code area in which the person stayed longest before 9 a.m., i.e., the place the person woke up in the morning. This definition could misplace some groups of people like those staying in hotels, which could cause fluctuations in the number of activities *per capita* especially in the pre-pandemic times. Also, people working in night shifts would be misplaced during the working days. Secondly, the additional protection of individual privacy by filtering activities with less than five individuals from a postal code area creates inaccuracies in the projected networks, which is why in fitting the models we filter out the postal code areas with population less than 500. This filtering removes seven postal code areas, including three industrial areas with population sizes of less than 200, two sparsely populated areas with populations less than 270, a hospital, and a military area.

The dataset provides a comprehensive view of the population activities on daily aggregation level, but lacks in terms of temporal co-occurrence within the day. Datasets such as the one created by Iyer et al. [7] consider the aspect of co-occurrence of events within a minute of time. The usage of GPS traces also make it possible to consider uniformly sized grids, whereas the data used in our study contained grids of different size based on the location of telephone masts. The heterogeneous grid sizes as well as daily aggregation of the dataset can lead to some overestimation of the exposure in terms of people being in relative proximity at the same time, e.g., being able to transmit an infection. Similar overestimation can also be considered in urban areas, where the real life exposure is depending also on the two people being in the same room or building. While these are the limitations, the data being in an aggregate form can be considered to better protect the privacy of individual users.

The design of this study considered an area-wise limited system of people living in 4 municipalities in the capital yet most populated area of Finland. The limitation can cause inaccuracies with the radiation model in areas located in the outskirts of the investigated area, as the variable of people living within the radius ($r_{d_{i,j}}$) is not considering the populations outside of the system. For the radiation model, the values for $r_{d_{i,j}}$ were also calculated such that the distribution of population was assumed to be even across the postal code area. Another assumption was made for the population size data, as we only consider the numbers recorded in 2019. The changes between the beginning and the end of our dataset were not large enough to significantly change the results of the model. Due to this, we are not considering factors such as influx of new inhabitants or outflow of people to other municipalities. A potential limitation and inaccuracy arises from calculation of distances between postal code areas. The assumption of having the geo-spatial center of each postal code area the point of calculation disregards the population density and distribution within the area, which results in some level of error in distances especially between postal code areas sharing borders. Also, in reality the distances between postal code areas are not the same as the

calculated crow-flight distances, in which case a more realistic measure could be the travel time. Considering the travel time between locations would most likely improve the model, but such data would add to the model complexity as the travel time can vary due to, e.g., road closures or constructions and public transportation changes. Moreover, an average travel time weighted with the means of travel available to the local population could improve the results.

In the future we aim to conduct further analysis on the empirical dataset in terms of the time series of exposure and the socioeconomic aspects of each postal code area. Also, a detailed analysis of the NPI effects is warranted [59]. Intuitively, the method of constructing a weighted projection network based on data of daily visits to a set of locations and using the resulting links as a proxy for contact could be used in modelling spreading processes between the postal code areas with compartmental SEIRS-type models [25, 60–62].

5 Conclusion

Mobile phone based datasets have been shown to provide novel insights into human mobility and their social interactions. In this paper we have investigated the use of aggregated mobile phone location data as a source for gaining insight into the exposure between populations living in different postal code areas using co-occurring visits to mobile phone grid locations as a proxy. People visiting the same cellphone grid areas form a bipartite network, which we project using Eq. 1 resulting in a fully connected network with edge weights between 0 and 1. The resulting empirical networks show the changes in exposure stemming from co-occurring activities between pre-pandemic and pandemic times shifting towards lower clustering, higher assortativity and lower weight for edges with longer distances between populations and to higher populated areas. These changes along with high correlation with the NPI-indices reflect the adherence to government imposed restrictions, remote work as well as self-imposed social distancing of aware individuals. We showed that the weights of the exposure can be modelled using gravity and radiation type models with population variables and distances between the postal code areas. The simplistic models yield promising results at the level of average edge weights, but are found to be lacking in reproducing the assortativity and clustering of the empirical networks. Lastly, we conducted a brief investigation on modelling exposure for the whole duration of our dataset using the different NPI sub-categories for calculating the exponents of our best performing gravity model, showing a varying degree of weight for the NPI sub-categories depending on the gravity model exponent.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The mobility dataset, Crowd Insights, is commercial and property of Telia Company AB. The other datasets used for population level information are publicly

available and referenced in the manuscript. Requests to access these datasets should be directed to Website, <https://business.teliacompany.com/crowd-insights>.

Author contributions

TT, KB, and KK contributed to conception and design of the study. TT processed and analyzed the data and created the models. TT wrote the first draft of the manuscript and created the figures and tables. All authors contributed to the article and approved the submitted version.

Funding

TT acknowledges funding from the Vilho, Yrjö and Kalle Väisälä Foundation of the Finnish Academy of Science and Letters. This work was partly supported by NordForsk through the funding to The Network Dynamics of Ethnic Integration, project number 105147. KB and KK acknowledge support from EU HORIZON 2020 INFRAIA-2019-1 (SoBigData++) No. 871042.

References

- Blondel VD, Decuyper A, Krings G. A survey of results on mobile phone datasets analysis. *EPJ Data Sci* (2015) 4(10):10. doi:10.1140/epjds/s13688-015-0046-0
- Wang J, Kong X, Xia F, Sun L. Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD Explorations NewsL* (2019) 21(1):1–19. doi:10.1145/3331651.3331653
- Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, et al. Human mobility: Models and applications. *Phys Rep* (2018) 734:1–74. doi:10.1016/j.physrep.2018.01.001
- González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. *Nature* (2008) 453:779–82. doi:10.1038/nature06958
- Simini F, González MC, Maritan A, Barabási AL. A universal model for mobility and migration patterns. *Nature* (2012) 484:96–100. doi:10.1038/nature10856
- Riascos A, Mateos JL. Emergence of encounter networks due to human mobility. *PLoS ONE* (2017) 12(10):e0184532. doi:10.1371/journal.pone.0184532
- Iyer S, Karrer B, Citron DT, Kooti F, Maas P, Wang Z, et al. Large-scale measurement of aggregate human colocation patterns for epidemiological modeling. *Epidemics* (2023) 42:100663. doi:10.1016/j.epidem.2022.100663
- Nguyen AD, Sénac P, Ramiro V, Diaz M. In: J Domingo-Pascual, P Manzoni, S Palazzo, A Pont, C Scoglio, editors. *Steps - an approach for human mobility modeling. Networking 2011*. Berlin, Heidelberg: Springer Berlin Heidelberg (2011). p. 254–65. doi:10.1007/978-3-642-20757-0_20
- Wu H, Liu L, Yu Y, Peng Z, Jiao H, Niu Q. An agent-based model simulation of human mobility based on mobile phone data: How commuting relates to congestion. *ISPRS Int J Geo-Information* (2019) 8(7):313. doi:10.3390/ijgi8070313
- Sakamane P, Phithakkitnukoon S, Smoreda Z, Ratti C. Methods for inferring route choice of commuting trip from mobile phone network data. *ISPRS Int J Geo-Information* (2020) 9(5):306. doi:10.3390/ijgi9050306
- Jia T, Jiang B, Carling K, Bolin M, Ban Y. An empirical study on human mobility and its agent-based modeling. *J Stat Mech Theor Exp* (2012) 2012:P11024. doi:10.1088/1742-5468/2012/11/P11024
- Kang C, Liu Y, Guo D, Qin K. A generalized radiation model for human mobility: Spatial scale, searching direction and trip constraint. *PLoS ONE* (2015) 10(11):e0143500. doi:10.1371/journal.pone.0143500
- Yan XY, Wang WX, Gao ZY, Lai YC. Universal model of individual and population mobility on diverse spatial scales. *Nat Commun* (2017) 8:1639–9. doi:10.1038/s41467-017-01892-8
- Lenormand M, Bassolas A, Ramasco JJ. Systematic comparison of trip distribution laws and models. *J Transport Geogr* (2016) 51:158–69. doi:10.1016/j.jtrangeo.2015.12.008
- Prieto Curiel R, Pappalardo L, Gabrielli L, Bishop SR. Gravity and scaling laws of city to city migration. *PLoS ONE* (2018) 13(7):e0199892. doi:10.1371/journal.pone.0199892

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphy.2023.1138323/full#supplementary-material>

- Ren Y, Ercsey-Ravasz M, Wang P, González MC, Toroczkai Z. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nat Commun* (2014) 5:5347. doi:10.1038/ncomms6347
- Wang D, Pedreschi D, Song C, Giannotti F, Barabasi AL. *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*. New York, NY, USA: Association for Computing Machinery (2011). p. 1100–8. KDD '11. doi:10.1145/2020408.2020581
- Human mobility, social ties, and link prediction
- Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, et al. Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci* (2007) 104(18):7332–6. doi:10.1073/pnas.0610245104
- Krings G, Calabrese F, Ratti C, Blondel VD. Urban gravity: A model for inter-city telecommunication flows. *J Stat Mech Theor Exp* (2009) 2009:L07003. doi:10.1088/1742-5468/2009/07/L07003
- Bhattacharya K, Mukherjee G, Saramäki J, Kaski K, Manna SS. The international trade network: Weighted network analysis and modelling. *J Stat Mech Theor Exp* (2008) 2008:P02002. doi:10.1088/1742-5468/2008/02/P02002
- Kiashemshaki M, Huang Z, Saramäki J. Mobility signatures: A tool for characterizing cities using intercity mobility flows. *Front Big Data* (2022) 5:822889. doi:10.3389/fdata.2022.822889
- Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C. A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE* (2012) 7(5):e37027. doi:10.1371/journal.pone.0037027
- Souch JM, Cossman JS, Hayward MD. Interstates of infection: Preliminary investigations of human mobility patterns in the Covid-19 pandemic. *J Rural Health* (2021) 37:266–71. doi:10.1111/jrh.12558
- Tiirinki H, Tynkkynen LK, Sovala M, Atkins S, Koivusalo M, Rautiainen P, et al. Covid-19 pandemic in Finland—preliminary analysis on health system response and economic consequences. *Health Pol Technol* (2020) 9(4):649–62. doi:10.1016/j.hlpt.2020.08.005
- Gumel AB, Iboi EA, Ngonghala CN, Elbasha EH. A primer on using mathematics to understand Covid-19 dynamics: Modeling, analysis and simulations. *Infect Dis Model* (2021) 6:148–68. doi:10.1016/j.idm.2020.11.005
- Zhang M, Wang S, Hu T, Fu X, Wang X, Hu Y, et al. Human mobility and Covid-19 transmission: A systematic review and future directions. *Ann GIS* (2022) 28(4):501–14. doi:10.1080/19475683.2022.2041725
- Jordan E, Shin DE, Leekha S, Azarm S. Optimization in the context of Covid-19 prediction and control: A literature review. *IEEE Access* (2021) 9:130072–93. doi:10.1109/access.2021.3113812
- Soucy JPR, Sturrock SL, Berry I, Westwood DJ, Daneman N, MacFadden DR, et al. *Estimating effects of physical distancing on the covid-19 pandemic using an urban mobility index*. medRxiv [Preprint] (2020). doi:10.1101/2020.04.05.20054288

29. Barrat A, Cattuto C, Kivela M, Lehmann S, Saramaki J. Effect of manual and digital contact tracing on Covid-19 outbreaks: A study on empirical contact data. *J R Soc Interf* (2021) 18(178):20201000. doi:10.1098/rsif.2020.1000
30. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide Covid-19 government interventions. *Nat Hum Behav* (2020) 4:1303–12. doi:10.1038/s41562-020-01009-0
31. Bergman NK, Fishman R. *Correlations of mobility and covid-19 transmission in global data*. medRxiv [Preprint] (2020). doi:10.1101/2020.05.06.20093039
32. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on Covid-19 incidence and population mobility patterns using a multiple-event study. *Scientific Rep* (2021) 11:1972. doi:10.1038/s41598-021-81442-x
33. Sulyok M, Walker M. Community movement and Covid-19: A global study using google's community mobility reports. *Epidemiol Infect* (2020) 148:e284. doi:10.1017/S0950268820002757
34. Sadowski A, Galar Z, Walasek R, Zimon G, Engelseth P. Big data insight on global mobility during the Covid-19 pandemic lockdown. *J Big Data* (2021) 8:78. doi:10.1186/s40537-021-00474-2
35. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and Covid-19 transmission in the USA: A mathematical modelling study. *Lancet Infect Dis* (2020) 20(11):1247–54. doi:10.1016/S1473-3099(20)30553-3
36. Kim J, Kwan MP. The impact of the Covid-19 pandemic on people's mobility: A longitudinal study of the us from march to september of 2020. *J Transport Geogr* (2021) 93:103039. doi:10.1016/j.jtrangeo.2021.103039
37. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Brusky D, et al. Mobility network models of Covid-19 explain inequities and inform reopening. *Nature* (2021) 589:82–7. doi:10.1038/s41586-020-2923-3
38. Levin R, Chao DL, Wenger EA, Proctor JL. Insights into population behavior during the Covid-19 pandemic from cell phone mobility data and manifold learning. *Nat Comput Sci* (2021) 1:588–97. doi:10.1038/s43588-021-00125-9
39. Badr HS, Gardner LM. Limitations of using mobile phone data to model Covid-19 transmission in the USA. *Lancet Infect Dis* (2021) 21(5):e113. doi:10.1016/S1473-3099(20)30861-6
40. Chin V, Ioannidis JP, Tanner MA, Cripps S. Effect estimates of Covid-19 non-pharmaceutical interventions are non-robust and highly model-dependent. *J Clin Epidemiol* (2021) 136:96–132. doi:10.1016/j.jclinepi.2021.03.014
41. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the Covid-19 pandemic life cycle. *Sci Adv* (2020) 6(23):eabc0764. doi:10.1126/sciadv.abc0764
42. Grantz KH, Meredith HR, Cummings DA, Metcalf CJE, Grenfell BT, Giles JR, et al. The use of mobile phone data to inform analysis of Covid-19 pandemic epidemiology. *Nat Commun* (2020) 11:4961. doi:10.1038/s41467-020-18190-5
43. Shepherd HE, Atherden FS, Chan HMT, Loveridge A, Tatem AJ. Domestic and international mobility trends in the United Kingdom during the Covid-19 pandemic: An analysis of Facebook data. *Int J Health Geographics* (2021) 20:46. doi:10.1186/s12942-021-00299-5
44. Fritz C, Kauermann G. On the interplay of regional mobility, social connectedness and the spread of Covid-19 in Germany. *J R Stat Soc Ser A, (Statistics Society)* (2022) 185(1):400–24. doi:10.1111/rssa.12753
45. Alidadi M, Sharifi A. Effects of the built environment and human factors on the spread of Covid-19: A systematic literature review. *Sci Total Environ* (2022) 850:158056. doi:10.1016/j.scitotenv.2022.158056
46. Cardoso BHF, Goncalves S. Universal scaling law for human-to-human transmission diseases. *Europhysics Lett* (2021) 133:58001. doi:10.1209/0295-5075/133/58001
47. Ribeiro HV, Sunahara AS, Sutton J, Perc M, Hanley QS. City size and the spreading of Covid-19 in Brazil. *PLoS ONE* (2020) 15(9):e0239699. doi:10.1371/journal.pone.0239699
48. Thurner S, Klimek P, Hanel R. A network-based explanation of why most Covid-19 infection curves are linear. *Proc Natl Acad Sci* (2020) 117(37):22684–9. doi:10.1073/pnas.2010398117
49. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (oxford Covid-19 government response tracker). *Nat Hum Behav* (2021) 5:529–38. doi:10.1038/s41562-021-01079-8
50. Wetter E, Rosengren S, Torn F, et al. *Private sector data for understanding public behaviors in crisis: The case of covid-19 in Sweden*. Working paper. Stockholm School of Economics (2020). Available at: https://swoba.hhs.se/hastma/paper/hastma2020_001.1.pdf (Accessed 05 22, 2023).
51. Willberg E, Jarv O, Vaisanen T, Toivonen T. Escaping from cities during the Covid-19 crisis: Using mobile phone data to trace mobility in Finland. *ISPRS Int J Geo-Information* (2021) 10(2):103. doi:10.3390/ijgi10020103
52. Chong SK, Bahrami M, Chen H, Balciyos S, Bozkaya B, Pentland A. Economic outcomes predicted by diversity in cities. *EPJ Data Sci* (2020) 9:17. doi:10.1140/epjds/s13688-020-00234-x
53. Hancean MG, Slavinec M, Perc M. The impact of human mobility networks on the global spread of Covid-19. *J Complex Networks* (2020) 8:cnaa041. doi:10.1093/comnet/cnaa041
54. Lombardi A, Amoroso N, Monaco A, Tangaro S, Bellotti R. Complex network modelling of origin–destination commuting flows for the Covid-19 epidemic spread analysis in Italian lombardy region. *Appl Sci* (2021) 11(10):4381. doi:10.3390/app11104381
55. [Dataset] Telia Finland Oyj. *Crowd insights* (2019). Available at: <https://business.teliacompany.com/crowd-insights>.
56. [Dataset] Statistics Finland. *Paavo: Postal code area statistics* (2019). Available at: https://stat.fi/tup/paavo/index_en.html (Accessed 04 07, 2022).
57. Hagberg A, Swart P, Chult SD. *Exploring network structure, dynamics, and function using networkx*. Tech. rep. Los Alamos National Lab(LANL). NM (United States): Los Alamos (2008). Available at: <https://www.osti.gov/biblio/960616>.
58. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. *Phys Rev E* (2004) 70:066111. doi:10.1103/PhysRevE.70.066111
59. Tully MA, McMaw L, Adlakha D, Blair N, McAneney J, McAneney H, et al. The effect of different Covid-19 public health restrictions on mobility: A systematic review. *PLoS ONE* (2021) 16(12):e0260919. doi:10.1371/journal.pone.0260919
60. Paoluzzi M, Gnan N, Grassi F, Salvetti M, Vanacore N, Crisanti A. A single-agent extension of the sir model describes the impact of mobility restrictions on the Covid-19 epidemic. *Scientific Rep* (2021) 11:24467. doi:10.1038/s41598-021-03721-x
61. Barrio RA, Kaski KK, Haraldsson GG, Aspelund T, Govezensky T. A model for social spreading of Covid-19: Cases of Mexico, Finland and Iceland. *Physica A: Stat Mech its Appl* (2021) 582:126274. doi:10.1016/j.physa.2021.126274
62. Snellman JE, Barrio RA, Kaski KK, Käpylä MJ. Modelling the interplay between epidemics and regional socio-economics. *Physica A: Stat Mech its Appl* (2022) 604:127696. doi:10.1016/j.physa.2022.127696