Journal of Complex Networks (2020) **8**, cnaa041
doi: 10.1093/comnet/cnaa041

The impact of human mobility networks on the global spread of COVID-19

MARIAN-GABRIEL HÂNCEAN[†]
Department of Sociology, University of Bucharest, Bucharest, Romania
[†]Corresponding author. Email: gabriel.hancean@unibuc.ro

MITJA SLAVINEC
Faculty of Natural Sciences and Mathematics, University of Maribor, Maribor, Slovenia

AND

MATJAŽ PERC
Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan & Complexity Science Hub Vienna, Vienna, Austria

Edited by: Ernesto Estrada

[Received on 7 October 2020; editorial decision on 13 October 2020; accepted on 14 October 2020]

Human mobility networks are crucial for a better understanding and controlling the spread of epidemics. Here, we study the impact of human mobility networks on the COVID-19 onset in 203 different countries. We use exponential random graph models to perform an analysis of the country-to-country global spread of COVID-19. We find that most countries had similar levels of virus spreading, with only a few acting as the main global transmitters. Our evidence suggests that migration and tourism inflows increase the probability of COVID-19 case importations while controlling for contiguity, continent co-location and sharing a language. Moreover, we find that air flights were the dominant mode of transportation while male and returning travellers were the main carriers. In conclusion, a mix of mobility and geography factors predicts the COVID-19 global transmission from one country to another. These findings have implications for non-pharmaceutical public health interventions and the management of transborder human circulation.

Keywords: COVID-19 case importations; human mobility networks; incoming migration; inbound tourism; complex networks.

1. Introduction

The ongoing pandemic of coronavirus disease (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), originated in Wuhan, China. The SARS-CoV-2, a coronavirus that infects humans, spread, as of September 2020, in 216 countries and territories [1]. COVID-19 has proved to be a highly contagious pathogenic viral infection, globally amassing more than 30 million confirmed cases and giving rise to approximately 1 million deaths [1], since its detection in December 2019. Subsequent non-pharmaceutical interventions (NPIs) have been implemented by both authorities in China and worldwide to decrease the number of new daily infections and deaths, as well as to control the demand for healthcare and avoid the overwhelming of national medical systems. Essentially, many of these interventions have consistently restricted human mobility. Either with the purpose to reduce local transmissions,

for example, the case of physical distancing—closure of schools and universities, of workplaces, isolation, lockdown, ban of gatherings, etc. Or to decline regional, national and international transmissions, for example, the case of international and domestic travelling restrictions—travel bans, quarantine, border closure, etc. The quantitative investigation of the NPIs' efficiency and the effectiveness in response to COVID-19 is currently underway [2, 3]. On one hand, several studies argue that a predetermined change in the human mobility patterns is expected to delay [4, 5], mitigate [6] or reduce [7] the COVID-19 spread. Recent developments even suggest that individual-level travel history and demographic characteristics can be used to predict the spread and size of the COVID-19 in China [8]. On the other hand, some claim it is unclear whether the restricting movement has been a proportional response to the COVID-19 pandemic [9] and that further research on the topic is needed.

The importance of human mobility patterns in understanding and controlling the spread of biological viruses has already been documented [10–12]. Even more, it has been claimed that the diversity of human circulation can be reduced to simple stable patterns which, in turn, become relevant for phenomena such as epidemic prevention or response [13]. Our study starts from the already reported evidence showing the contribution of carriers (returned travellers and visitors) to the international spread of COVID-19, during its early stage [14–17]. A plethora of case studies illustrates that country-level COVID-19 dispersion was ignited by case importations [18–26], and international migration [27]. However, little is known on the country-to-country global spread of COVID-19, and a description is missing for the moment. Available studies have tended to focus on whether air travel volume data could predict COVID-19 outbreaks [28–30]. Additionally, this line of work has not assessed the effect of human mobility *per se,* in explaining the COVID-19 cases onset in each country worldwide.

In this article, the first objective is to cover the current gap in the literature by describing the early COVID-19 global transmission from one country to another. The second objective is to contribute to the understanding of the COVID-19 spread by examining the associated role of human mobility networks. To these objectives, our attention is not focused on the medical and biological features of COVID-19. But, we draw our investigation stemming from the theoretical reasoning that the spread of epidemics among humans is inherently a social network problem. Contact tracing, social interaction patterns and mobility fluxes feed epidemic forecast and modelling [10]. Consequently, we look at two forms of global human mobility: inflows of migration and tourism. We hypothesize that these factors (incoming migration—hypothesis 1) and (inbound tourism—hypothesis 2) have a positive impact upon the initial global spread of COVID-19 (the dependent variable). To test these hypotheses, we employ a socio-centric network analysis research design and fit exponential random graph models (ERGMs) [31]. In our analysis, the dependent variable is built after having collected a unique dataset on the first cases of COVID-19 officially reported by national authorities. Both the predictors of interest and dependent variable are input into the models as origin–destination matrices. In our statistical modelling, we also control for possible confounding variables, such as geographical (contiguity and sharing the same continent) and cultural proximity (sharing a common language). These variables, included as symmetric binary matrices, may affect both the migration [32, 33] and tourism [34–36] inflows as well as the initial COVID-19 spread [27]. All variables included in the ERGM models are real-world country-level network data.

2. Results

We start by reporting descriptive statistics on the individual COVID-19 case onset at a country-level. Throughout the 219 countries and territories, we observe 323 COVID-19 cases (some countries reported more than one patient as the first officially confirmed case), with an average age of 45.9 (95% CI 44.7–47.1, SD=16.5, SE=1.2, missing = 132). Among these, 94 (29.1%) are females, 156 (48.3%) are males,

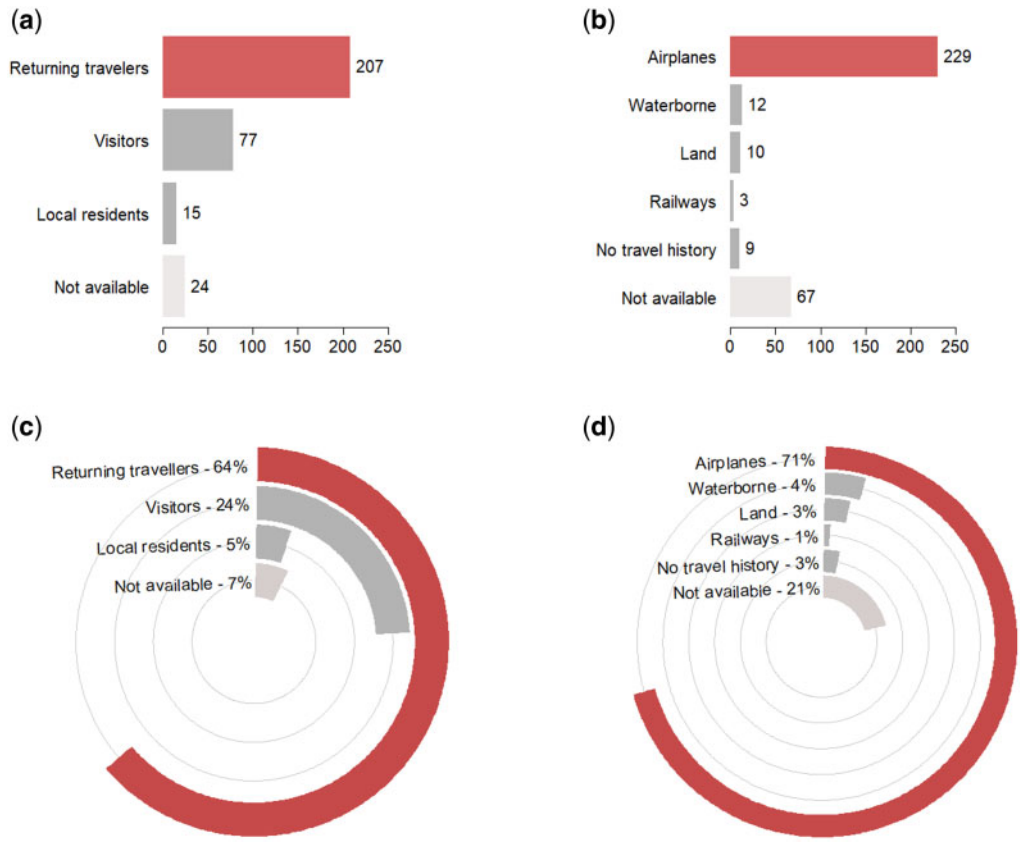

FIG. 1. Type of travellers and modes of transportation for the first 323 individual COVID-19 cases. Images (a) and (b) are counts, while (c) and (d) are percentages.

while for 73 (22.6%) patients there is no available evidence on the sex variable. In a 3-month time, COVID-19 spread in almost all the countries and territories of the globe. Specifically, approximately 67% of all first cases were officially announced during March 2020. Also, 95% of the countries and territories publicly confirmed their first cases between 1 January 2020 and 31 March 2020. Specifically, 12% in January, 18% in February and 65% in March. Put it differently, in 3 months, COVID-19 spread in 208 out of the 219 documented countries and territories.

Figure 1 displays the distribution of the first individual COVID-19 cases by type of travellers and by modes of transportation (where applicable). As shown in Fig. 1(a and c), of all the 323 documented cases, 64% ($n = 207$) are returning travellers, 24% ($n = 77$) visitors, 5% ($n = 15$) local residents and 7% ($n = 24$) are missing data (i.e. data not available in the country-level official reports). In terms of transportation modes (Fig. 1(b and d)), the airplanes were used by 71% ($n = 229$) of the individual cases, waterborne transport was used by 4% ($n = 12$), land transport (bus or car) was used by 3% ($n = 10$) and railways by 1% ($n = 3$). For 21% ($n = 67$) of the patients, the means of transportation was not public information (i.e. data not available). In the case of 3% ($n = 9$) of the cases, patients reported not having *a recent travel history during the past 20 days*. It is noteworthy that the percentage exceeds 100% as, in

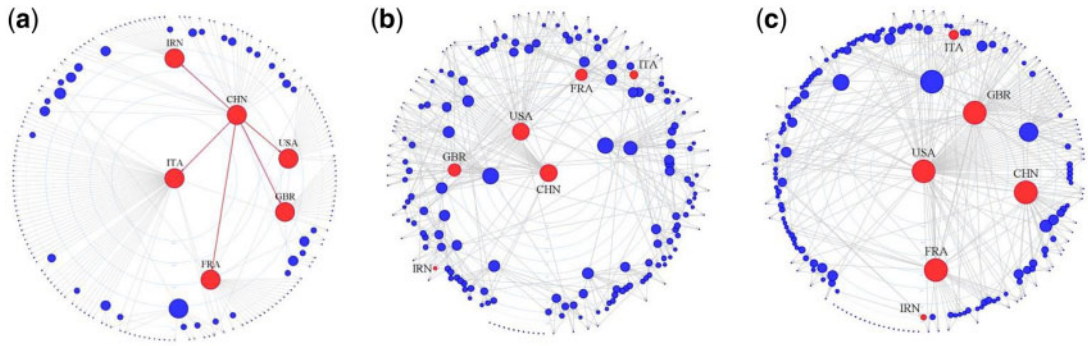

FIG. 2. Centrality layouts for visualizing the patterns of (a) COVID-19, (b) incoming migration and (c) inbound tourism ties among countries. Node size is proportional to out-degree centrality. Countries located more centrally in the pictures have higher centrality scores. Core-countries are marked with red. In (a), ties among core-countries are marked in red.

some cases ($n = 6$), multiple means of transportation were used, for example, railways and bus, airplane and bus or car, maritime and airplane.

Using out-degree centrality layouts, we illustrate three binary directed graphs: *the COVID-19 network* (203 nodes, 213 ties, density (D) = 0.005, out-degree centralization (*out-deg C*) = 0.249, in-degree centralization (*in-deg C*) = 0.015) (Fig. 2a), *the incoming migration network* (203 nodes, 571 ties, $D = 0.014$, *out-deg C* = 0.001, *in-deg C* = 0.105) (Fig. 2b) and *the inbound tourism network* (203 nodes, 519 ties, $D = 0.013$, *out-deg C* = 0.002, *in-deg C* = 0.321) (Fig. 2c). In Fig. 2, variations in country-level out-degree centrality scores (computed as percentage) is marked by both layout location (states with higher out-degrees are more central) and node size (the larger the node, the higher the centrality score). The centralization of the incoming migration network (both in-degree and out-degree) is influenced by the restrictions on the tie eliciting (only the first top three migration source countries were accounted in the analysis).

Fitting a core/periphery model [37, 38], we found that *China* (*CHN*), the *United States of America* (*USA*), *France* (*FRA*), *Italy* (*ITA*), *Great Britain* (*GRB*) and *Iran* (*IRN*) represent the core of the COVID-19 network. The core/periphery model has a fit (correlation) of 0.2, for an expected density of *core/periphery ties* and *periphery/core ties* of 0.5 (at five random starts and 100 iterations). In Fig. 2, we mark the six core-countries with red in all the three networks, for rapid identification of their centrality placement. All three networks fit power-law distributions [39]. Particularly, this is indicated by Kolmogorov–Smirnov (*KS*) statistic tests, that is, the COVID-19 network ($KS = 0.07$, $p = 0.99$), the incoming migration network ($KS = 0.05$, $p = 0.99$), the inbound tourism network ($KS = 0.04$, $p = 0.99$).

Figure 3(a–c) displays the *(common-)language-network* (203 nodes, 3194 edges, $D = 0.156$, *deg* C = 0.308), the *contiguity-network* (203 nodes, 301 edges, $D = 0.015$, *deg C* = 0.065) and the *(location-on-) same-continent network* (203 nodes, 4756 edges, $D = 0.232$, *deg C* = 0.031). The language-network has 40 components (the main component, *MC*, of 152 nodes, four components of triads, four dyads and 31 isolates), while the contiguity-network has 52 components (an MC of 128 nodes, another component of 23 nodes, 2 dyads and 48 isolates). The same-continent-network is made up of five components, corresponding to the geographical regions designed by the United Nations. Specifically, the countries and territories included in the analysis are spread across the regions in the following way: Africa—54 nodes, America—50 nodes, Asia—49, Europe—43 and Oceania—7. The groupings of countries and territories by regions can be scrutinized in Fig. 3c.

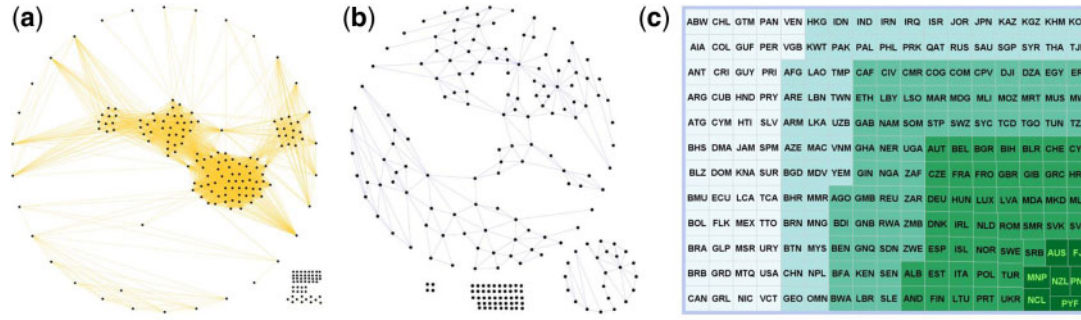

FIG. 3. Visualizations of (a) the common language network, (b) the contiguity network and (c) the location on the same continent network. In (c), countries sharing the same geographical region are marked by colour (we employed Unites Nations' region classification: America, Asia, Africa, Europe and Oceania). Also, the states included in the network are available by their ISO-alpha three codes (three-letter abbreviation).

TABLE 1 *QAP Pearson's correlations*

Network variables	1	2	3	4	5
1. COVID-19 network					
2. Migration	0.049				
	($p = 0.0002$)				
	SD = 0.0048				
3. Tourism	0.013	0.355			
	($p = 0.0144$)	($p = 0.0002$)			
	SD = 0.0046	SD = 0.0060			
4. Contiguity	0.065	0.406	0.340		
	($p = 0.0002$)	($p = 0.0002$)	($p = 0.0002$)		
	SD = 0.0053	SD = 0.0056	SD = 0.0058		
5. Common	0.038	0.102	0.101	0.097	
language	($p = 0.0002$)	($p = 0.0002$)	($p = 0.0002$)	($p = 0.0002$)	
	SD = 0.0095	SD = 0.0065	SD = 0.0097	SD = 0.0081	
6. Same continent	0.032	0.156	0.128	0.206	0.165
	($p = 0.0002$)	($p = 0.0002$)	($p = 0.0002$)	($p = 0.0002$)	($p = 0.0002$)
	SD = 0.0052	SD = 0.0056	SD = 0.0055	SD = 0.0070	SD = 0.0077

Correlations were performed based on 5000 permutations.

Table 1 reports the quadratic assignment procedure (QAP) Pearson's correlation coefficients computed for the squared adjacency matrices previously illustrated as networks in Figs. 2 and 3 [38]. All coefficients are positive and statistically significant. The pair-wise correlation measurements indicate rather weak relationships between the network variables.

Table 2 presents the most important results of our article: the estimates for the structural and relational attribute effects included in the ERGM models (see Section 4 for the definition of the effects). These models differ in the following way. Model 1 examines purely structural factors (*edges, activity spread, indegree* of 2, 1 and 0, *outdegree* of 2, 1 and 0, *twopaths*, and *isolates*). Model 2 examines the *incoming migration* and *inbound migration* effects (the predictors or the relational attribute effects of interest), controlling for structural effects. Model 3 (the full model) examines the predictors of interests, while

TABLE 2 *ERGM estimates of structural and relational attribute effects on the presence of a COVID-19 tie between a sending and a receiving country*

	Model 1			Model 2			Model 3		
	Est.	*p*-value	SE	Est.	*p*-value	SE	Est.	*p*-value	SE
Structural effects									
Edges	−1.70	0.0515	(0.87)	−2.09	0.0398	(1.02)	−2.36	0.0159	(0.98)
Activity spread†	−4.84	0.0001	(0.38)	−3.96	0.0001	(0.39)	−4.30	0.0001	(0.40)
Indegree (2)	0.28	0.8229	(1.26)	0.32	0.8211	(1.41)	0.31	0.8172	(1.35)
Indegree (1)	2.23	0.2768	(2.06)	2.33	0.3224	(2.35)	2.27	0.3141	(2.25)
Indegree (0)	−0.44	0.8841	(3.01)	−0.27	0.9365	(3.43)	−0.40	0.9041	(3.28)
Outdegree (2)	−1.26	0.0713	(0.70)	−0.89	0.2209	(0.72)	−1.06	0.1412	(0.72)
Outdegree (1)	−1.30	0.1039	(0.80)	−0.55	0.5050	(0.82)	−0.92	0.2648	(0.82)
Outdegree (0)	−0.59	0.5761	(1.05)	0.66	0.5442	(1.08)	0.03	0.9757	(1.09)
Twopath	−0.05	0.2356	(0.04)	−0.05	0.2621	(0.04)	−0.05	0.2335	(0.04)
Isolates	0.10	0.9076	(0.87)	0.10	0.9040	(0.84)	0.09	0.9107	(0.83)
Relational attribute effects									
Incoming migration				1.36	0.0001	(0.22)	0.97	0.0001	(0.23)
Inbound tourism				0.93	0.0001	(0.18)	0.77	0.0001	(0.16)
Contiguity							0.75	0.0057	(0.27)
Common language							0.73	0.0001	(0.14)
Same continent							0.37	0.0220	(0.16)
AIC	1631			1507			1476		
BIC	1717			1611			1605		
Log Likelihood	−805			−742			−723		

†Decay parameter fixed at 2, for activity spread (two-degree).

controlling both for the structural effects and for the other three relational attribute effects (*contiguity*, *common language* and *same continent*).

We notice, in Table 2, that *incoming migration* and *inbound tourism* have a statistically significant positive impact ($p = 0.0001$) in the process of COVID-19 tie formation (Models 2 and 3). Specifically, COVID-19 case importations are more likely from states that are important sources of migration and tourism. Model 2 illustrates an increase of 3.9 (Exp(1.36), $Est = 1.36$, $p = 0.0001$, SE=0.22) in the odds of receiving a COVID-19 tie from a country that is a top source of migration, and of 2.5 ($Est = 0.93$, $p = 0.0001$, $SE = 0.18$), from a country that is a top source of tourism. The probabilities of case importations from top sending migration and tourism countries are 0.80 (exp(1.36)/(1+exp(1.36)), and 0.72, respectively. Model 3 shows that the effects of incoming migration and inbound tourism remain statistically significant positive (p =0.0001) while controlling for geographical (contiguity and same continent) and cultural (common language) proximities. Precisely, the corresponding probabilities of a COVID-19 case importation are 0.73 (Est=0.97, $p = 0.0001$, $SE = 0.23$) and 0.68 (Est = 0.77, $p = 0.0001$, $SE = 0.16$). We also observe that geographical (*contiguity* and *same continent*) and cultural (*common language*) proximities are statistically significant predictors ($p = 0.0057$, $p = 0.0220$ and $p = 0.0001$, respectively) when accounting for COVID-19 ties between countries. Importing COVID-19 cases from neighbouring countries (*contiguity*) has a 0.68 probability (Est = 0.75, $p = 0.0057$, SE= 0.27), while COVID-19 contagion on the *same continent*, a probability of 0.59 (Est = 0.37, $p = 0.0220$, SE= 0.16). *Common language* gives an increase of 2.1 (Est=0.73, $p = 0.0001$, SE=0.14) in the odds of receiving a COVID-19 tie (the corresponding tie probability is 0.67).

Referring to the structural effects, the tendency of forming ties in the network (*edges*) has statistically significant negative estimates in Models 2 and 3 ($Est = -2.09$, $p = 0.0398$, $SE = 1.02$, and $Est = -2.36$,

$p = 0.0159$, SE= 0.98, respectively). This indicates that the number of ties in the observed network is lower than expected by random. *Activity spread* is a statistically significant negative effect in all the three models ($Est = -4.84$, $p = 0.0001$, SE = 0.38, $Est = -3.96$, $p = 0.0001$, SE= 0.39, and $Est = -4.30$, $p = 0.0001$, SE= 0.40, respectively). This suggests that the number of network structures with active nodes is lower than expected by chance alone. Put it differently, most countries have similar levels of activity in the network (i.e. sending COVID-19 ties).

The goodness of fit (GOF) statistics suggests that our models reproduce and fit the data well. Specifically, GOF statistics were less than 2.0 for features not explicitly modelled [40, 41]. Moreover, as shown in Table 2, the estimates do not exceed ± 5, while *standard errors* (S.E.) are no larger than 10 [42].

3. Discussion

We start this section by summarizing our main findings. Analysing data on COVID-19 importations in 203 countries, we show that the first cases were preponderantly males, returning travellers and that the most frequent mode of transportation was by air-flights. We also illustrate that the early country-to-country COVID-19 transmission network fits a power-law distribution and has five main global transmitters: China, France, Great Britain, Iran, Italy and the USA. Our statistical models indicate that incoming migration and inbound tourism are statistically significant positive predictors ($p = 0.0001$) for COVID-19 case importations. Specifically, a given country is more likely to receive COVID-19 cases from states that are among its top three sources of international migration and tourism. These predictors remain statistically significant ($p = 0.0001$) after controlling for contiguity, sharing the same continent and having a common language (which provides empirical support for our hypotheses).

Our work is in accord with simulation [29] and observational [28, 30] studies showcasing the risk of COVID-19 outbreak through air travel. Also, we are in good agreement with previous papers demonstrating that the COVID-19 global [43] or domestic prevalence follows a power-law distribution [44]. Furthermore, we found that the first COVID-19 cases across the globe were mostly males. Contrary to this specific finding, studies on the COVID-19 prevalence in various countries [45] argue males and females are equally likely to be diagnosed with COVID-19.

During the last decades, international tourism has increased, at an unprecedented level [46] and the global migration network (i.e. countries connected by migration flows) has evolved into a small-world type of structure, displaying a constant decrease in the average path length [47]. Put it differently, the global circulation of people is now more intense than ever, while distances have apparently melted. This entails unquestionable multiple social and economic benefits. However, at the same time, it seems to lay out various vulnerabilities, among which we can include the velocity of biological viruses' transmission. In this context, some studies have implied that the COVID-19 diffusion may be different from other pandemics given migrants' recent possibilities of travelling over long distances [48]. The role of migration corridors (origin and destination places socially connected by migrants) has been also emphasized as a key factor, in addition to the COVID-19 contiguity diffusion [27]. Moreover, it has been claimed that international tourists have even higher risk of spreading COVID-19 due to their unconstrained and infrequent movement patterns [48]. Evidence suggests that the movement of tourists should not be disregarded as a factor in COVID-19 diffusion while an abundant literature has so far documented case importations across the world [18–26]. Our study confirms the already reported evidence underlining the role of human mobility networks in the spread of biological viruses, in general. And, moreover, we give support to previous literature arguing the impact of migration and tourism on COVID-19, in particular.

Despite the preventive measures and NPIs implemented by governments worldwide or the efforts of providing effective medication [49], we show that COVID-19 spread, in a 3-month time interval, in

95% of the countries and territories that officially confirmed cases. A global diffusion of this magnitude may indicate that national states lacked an efficient coordination in controlling the country-to-country virus spread. A core group of countries seems to have involuntarily acted as hubs in the global virus transmission, confirming one of the well-documented features of complex networks, the preferential attachment [50]. The out-degree centrality of these countries in other global structures, such as migration and tourism networks, may have facilitated their acting as main global transmitters.

In line with our research objectives stated in the Section 1, we described how COVID-19 spread from one country to another and demonstrated the statistically significant effect of migration and tourism inflows ($p = 0.0001$). One possible implication refers to the utility that models predicting population movements [13, 51] have in preparing responses to future potential different pandemics. Or even possible next waves of COVID-19. Modelling global human mobility may be an important input in increasing the international coordination of national authorities for handling cross-border circulation. Additionally, we believe our study may be a valuable contribution to developing global strategies for the early prevention of pandemics. For instance, coordination of information exchange between neighbouring countries as well as between states that share cultural and migration corridors may be of vital importance for case importation detection. National early warning systems should be internationally synchronized and integrate customized real-time information about global mobility flows as well as about countries that act as global transmitters. Furthermore, international coordination of information streams on people's movement patterns may be a driver for increasing the consistency level of the governmental preventive measures. It may be the case that the existing uncoordinated variations in the responses of domestic authorities to the COVID-19 to have facilitated the global virus spread. Lastly, our work may be particularly useful for researchers aiming to compare COVID-19 to other biological viruses' non-medical spreading dynamics, such as severe acute respiratory syndrome (SARS) for example [48]. On top of that, we hope our results may be beneficial for policymakers and national authorities in their efforts of managing human mobility across borders, economic difficulties and migration-related issues. Rapid action is needed as the COVID-19 pandemic has heavily hit the tourism industry and the air-flight sector. And, perhaps, even more important, the global spread of COVID-19 has consistently negatively impacted upon migration, exacerbating social vulnerabilities and polarities, increasing stigmatization and exclusion and aggravating migrants' health vulnerabilities.

We are aware that our study may have at least two limitations. The first refers to the official country-level identification of the first COVID-19 cases. Our raw data come from national authorities' public announcements. Theoretically, that does not eliminate the possibility of COVID-19 existing in a country prior to being officially reported. The second concerns the network variables included in our statistical modelling as control variables. The measurement of geographical and cultural proximities was limited to contiguity, co-location on the same continent and language sharing. This limit is due to the fact that, generally, similar proxy variables at a global level are exceptionally scarce. Despite these limitations, it is noteworthy that the estimates of our statistical models hold for any global COVID-19 network that displays similar structural features.

To the best of our knowledge, this is the first study to document the global spread of COVID-19 outbreak and to assess the effects of human mobility networks. The evidence from this study shows that the global spread of the COVID-19 outbreak was not random but patterned by the incoming migration and inbound tourism, with geographical and cultural proximities being controlled for. Further research on this topic is needed to increase our understanding of how nowadays global human mobility routes impact upon virus spreading.

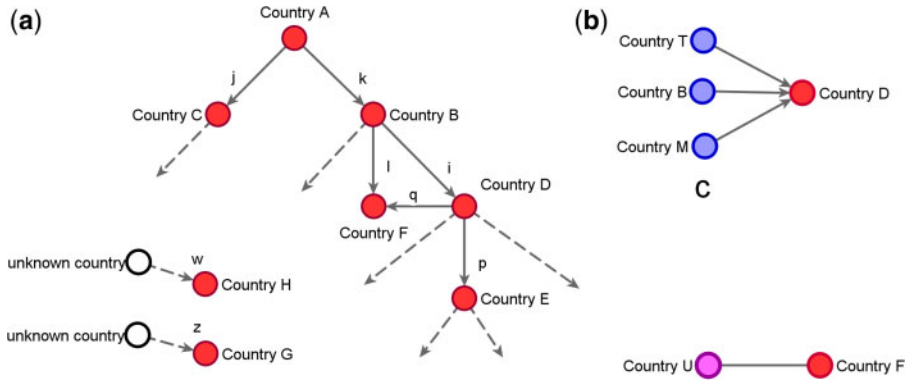

FIG. 4. The design of the COVID-19 country-level spreading network as well as of the other country-level networks used into the analysis. (a) Two countries share a COVID-19 connection (arrow) based on the travelling history of the carriers (small letters). Isolates (countries H and G) are countries for which the COVID-19 country source is unknown. (b) country D (a confirmed COVID-19 country) receives arrows from T, B and M—these are the top three sending migration/tourism sources. (c) a COVID-19 country (country F) is connected by an edge to another country if the two are contiguous/share the same continent/have a common language (at least 20% of the population in one country speak the language of the other country).

4. Methods

In this observational study, we collected empirical data on the global spread of COVID-19 outbreak at a country-level. Before starting the data collection process, we used the information from the World Health Organization [1] to identify the countries and territories with confirmed cases of COVID-19. Afterwards, for each state and territory in the list, we searched for available data on the first officially reported cases. We amassed a tally of 323 cases (patients), for which we identified related attribute information on *age, sex, travel mode*, *confirmation date*, *type of traveller* (e.g. visitor, returning traveller, etc.) and *travel information* (*target* and *source* countries). These country-level first cases of COVID-19 were documented by examining the evidence publicly shared by the national authorities in 219 countries and territories. For some states, the authorities announced more than one person as 'first case of COVID-19'. The available information on the travelling history of the first cases was used to build an origin-destination country matrix. This matrix was subsequently used to construct the COVID-19 global spreading nexus, that is, a binary directed graph. Specifically, in this network, an arrow from country *A* to country Billustrates that a patient *k*, returning traveller or visitor from *A*, was officially announced as the first COVID-19 case in country *B* (Fig. 4a). Moreover, the *isolates* in the network (for instance, countries H and G in Fig. 4a) are states for which information on the transmission route is not available due to various reasons (e.g. the patient did not report having a travel history during the last 15 days, national authorities did not provide travelling history information, etc.). According to our research design, the first cases of COVID-19 in a specific country are the patients who were officially identified 'as first COVID-19 cases' by domestic authorities.

In addition to the COVID-19 (country-level spreading) network, we built five supplementary networks: the (incoming) migration network, the (inbound) tourism network, the contiguity network, the same continent network and the common language network. Subsequently, we performed a list-wise deletion on the six network variables and kept for the analysis 203 (93%) out of the 219 documented countries and territories.

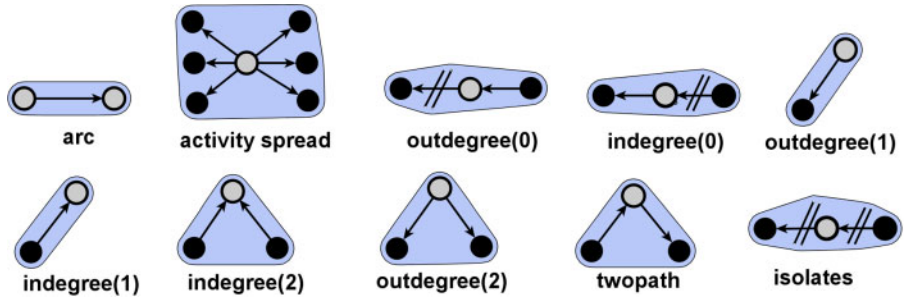

FIG. 5. Structural configurations included as effects in the ERGMs.

Triads similar to the one displayed as Fig. 4b were used to build the (incoming/ inflow) migration network and the (inbound) tourism network. Particularly, the top three migration sending countries and the top three tourism sending countries were identified for each of the states included in the COVID-19 network. We used the most recent available data on the global migration flows (2017), from the World Bank [52] and on the global tourism flows (2016), from World Tourism Organization [53].

In the *contiguity network*, two countries are connected by an edge if they share the same border. In the *common language network*, two states are linked if at least 20% of the population in one country speak the language of the other country. In the *same continent network*, two states are joined together if located on the same continent (we employed the geo-scheme advanced by the United Nations [54]). We used undirected dyads, similar to the one exhibited in Fig. 4c, to build these networks. The information on the geographical variables is publicly available from Centre D'Études Prospective Et D'Informations Internationales [55].

We used a suite of software tools for network descriptive statistics and visualization. Precisely, we run the *core/periphery* algorithm available with UCINET 6.0 [37, 38] for detecting the core-countries in the COVID-19 network. Also, we performed the *Kolmogorov–Smirnoff* test for power-law distribution fit, using the algorithm available with the NetIndices R-package [56]. For network visualizations, we employed the circular and treemap node layouts available in visone [57]. The dataset files used in this article as well as the code are available for replication [58].

Initially, we performed quadratic assignment procedure (QAP) correlations (with 5000 permutations) to infer association among network variables, at a dyadic level. QAP correlations were useful to identify multi-collinearity, but not to explain patterns of ties. Consequently, we modelled the COVID-19 network using exponential random graph models (ERGMs) [31]. Generally, ERGMs are statistical models for understanding how and why network ties arise in a specific network [41]. In this article, we were interested in examining the generative processes giving rise to the structures or tie patterns in the observed network of COVID-19 countries. In our ERGMs, we estimated two classes of predictors: structural effects and relational attribute effects. The coefficients estimated for each predictor define the probability of each edge in the COVID-19 network and the probability of the entire network [59, 60].

The *structural effects* refer to specific tie patterns or link configurations within the network [59, 60]. In Fig. 5, we visually display the structural effects introduced in our ERGMs. These images are built on similar visualizations available in the literature [41]. *Edges (arcs)* refer to the overall tendency for COVID-19 ties to form in the network. Generally, ERGMs include this term by default. *Activity spread* term controls for the COVID-19 network's tendency to have active nodes (countries that send many arrows).

The out-degree (0) or *sinks* parameter marks nodes with zero out-degree but positive in-degree. The *in-degree (0)* or *sources* parameter marks nodes with zero in-degree but positive out-degrees. The *isolates* term defines nodes with both in-degree and out-degree equal to zero [59, 60]. *Sinks, sources* and *isolates* are useful model parameters for controlling degree distributions in directed networks [40]. *Out-degree (1)*, (2) and (3) are related to the *activity spread*, while *in-degree (0), (1) and (2)* are related to *popularity spread,* as resulting from the network self-organization. *In-* and *out-degree* effects are indicative of the tendencies for centralization in the in- and out-degree distributions [41]. *Twopath* parameter accounts for the situations wherein actors who receive ties are also sending. Its model inclusion controls for the association between out-degree and in-degree.

The *relational attribute effects* in our ERGMs are predicated to affect the probability of a COVID-19 tie. The estimated effects were: ties with top three migration sending countries (*incoming migration*), ties with top three tourism sending countries (*incoming tourism*), ties with contiguous countries (*contiguity*), ties with same continent countries (*same continent*) and ties with countries wherein the language is spoken by 20% of the population (*common language 20%*).

We performed a diagnostic goodness of fit examination using the following procedure. We simulated 100 networks based on our model estimates. Then, we compared the simulated and the COVID-19 networks, examining the difference between the observed and mean scores on *dyad-wise shared partners, edge-wise shared partners, degree, in-degree* and *geodesic distances* [31, 40, 59–61]. In addition, we assessed whether the parameter estimates are in the ± 10 intervals and whether their standard errors do not exceed 5 [42].

Funding

Executive Agency for Higher Education, Research, Development and Innovation Funding (PN-III-P1-1.1-TE-2016-0362 to M.-G.H.); and The Slovenian Research Agency (J1-2457, J1-9112 and P1-0403 to M.P.).

Acknowledgements

We are grateful to Ms Laura Trandafir for figure formatting preparations.

References

1. World Health Organization. (2020) Coronavirus disease (COVID-19). Weekly Epidemiological Update.
2. Chinazzi, M. *et al.* (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, **368**, 395–400.
3. Zhao, S. *et al*. (2020) Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis. *J. Travel Med.*, **27**, 1–3.
4. Anzai, A. *et al*. (2020) Assessing the impact of reduced travel on exportation dynamics of novel coronavirus infection (COVID-19). *J. Clin. Med.*, **9**, 601.
5. Tian, H. *et al*. (2020) An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, **368**, 638–642.
6. Badr, H. S. *et al*. (2020) Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Dis*, **20**, 1247–1254.
7. Askitas, N., Tatsiramos, K. & Verheyden, B. (2020) Lockdown strategies, mobility patterns and COVID-19. *Lockdown Strategies, Mobility Patterns and COVID-19*. (Charles Wyplosz ed.), London, UK: The Centre for Economic Policy Research (CEPR), pp. 293–302.

8. Kraemer, M. U. G. *et al*. (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, **368**, 493–497.
9. Zhou, Y. *et al*. (2020) Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. *Lancet Digit. Health*, **2**, e417–e424.
10. Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J. & Vespignani, A. (2007) Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med.*, **4**, e13.
11. Eubank, S. *et al*. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature*, **429**, 180–184.
12. Hufnagel, L., Brockmann, D. & Geisel, T. (2004) Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. USA*, **101**, 15124–15129.
13. González, M. C., Hidalgo, C. A. & Barabási, A.-L. (2008) Understanding individual human mobility patterns. *Nature*, **453**, 779–782.
14. Gilbert, M. *et al*. (2020) Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *Lancet*, **395**, 871–877.
15. Giovanetti, M., Benvenuto, D., Angeletti, S. & Ciccozzi, M. (2020) The first two cases of 2019-nCoV in Italy: where they come from? *J. Med. Virol.*, **92**, 518–521.
16. Ghinai, I. *et al*. (2020) First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. *Lancet*, **395**, 1137–1144.
17. Hâncean, M.-G., Perc, M. & Lerner, J. (2020) Early spread of COVID-19 in Romania: imported cases from Italy and human-to-human transmission networks. *R. Soc. Open Sci.*, **7**, 200780.
18. Candido, D. D. S. *et al*. (2020) Routes for COVID-19 importation in Brazil. *J. Travel Med*., **27**, 1–3.
19. Day, M. (2020) Covid-19: Italy confirms 11 deaths as cases spread from north. *BMJ*, **368**, 1.
20. Escalera-Antezana, J. P. *et al*. (2020) Clinical features of the first cases and a cluster of Coronavirus Disease 2019 (COVID-19) in Bolivia imported from Italy and Spain. *Travel Med. Infect. Dis*, **2020**, 101653.
21. Pullano, G. *et al*. (2020) Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. *Eurosurveillance*, **25**, 2000057.
22. Rothe, C. *et al*. (2020) Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N. Engl. J. Med.*, **382**, 970–971.
23. Tuite, A. R., Ng, V., REES, E. & FISMAN, D. (2020) Estimation of COVID-19 outbreak size in Italy. *Lancet Infect. Dis*., **20**, 537.
24. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. (2020) A novel coronavirus outbreak of global health concern. *Lancet*, **395**, 470–473.
25. Wu, J. T., Leung, K. & Leung, G. M. (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, **395**, 689–697.
26. Zhuang, Z. *et al*. (2020) Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: a modelling analysis based on overseas cases and air travel data. *Int. J. Infect. Dis.*, **94**, 29–31.
27. Sirkeci, I. & Yucesahin, M. M. (2020) Coronavirus and migration: analysis of human mobility and the spread of Covid-19. *Migr. Lett.*, **17**, 379–398.
28. Bogoch, I. I. *et al*. (2020) Potential for global spread of a novel coronavirus from China. *J. Travel Med.*, **27**, 1–3.
29. Daon, Y., Thompson, R. N. & Obolski, U. (2020) Estimating COVID-19 outbreak risk through air travel. *J. Travel Med.*, **27**, 1–8.
30. Wells, C. R. *et al*. (2020) Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl. Acad. Sci. USA*, **117**, 7504–7509.
31. Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M. & Morris, M. (2008) ergm: a package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.*, **24**, p. nihpa54860.
32. Adserà, A. & Pytliková, M. (2015) The role of language in shaping international migration. *Econ. J.*, **125**, F49–F81.
33. Mayda, A. M. (2010) International migration: a panel data analysis of the determinants of bilateral flows. *J. Popul. Econ.*, **23**, 1249–1274.

34. Basala, S. L. & Klenosky, D. B. (2001) Travel-style preferences for visiting a novel destination: a conjoint investigation across the novelty-familiarity continuum. *J. Travel Res.*, **40**, 172–182.

35. Ng, S. I., Lee, J. A. & Soutar, G. N. (2007) Tourists' intention to visit a country: the impact of cultural distance. *Tour. Manag.*, **28**, 1497–1506.

36. Vanegas, J., Valencia, M., Restrepo, J. & Muñeton, G. (2020) Modeling determinants of tourism demand in Colombia. *Tour. Hosp. Manag.*, **26**, 49–67.

37. Borgatti, S. P. & Everett, M. G. (2000) Models of core/periphery structures. *Soc. Netw.*, **21**, 375–395.

38. Borgatti, S. P., Everett, M. G. & Freeman, L. C. (2002) *Ucinet 6 for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

39. Klaus, A., Yu, S. & Plenz, D. (2011) Statistical analyses support power law distributions found in neuronal avalanches. *PLoS One*, **6**. doi: 10.1371/journal.pone.0019779.

40. Robins, G., Pattison, P. & Wang, P. (2009) Closure, connectivity and degree distributions: exponential random graph (p*) models for directed social networks. *Soc. Netw.*, **31**, 105–117.

41. Lusher, D., Koskinen, J., & Robins, G. (2013) *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.

42. Kruse, H., Smith, S., van Tubergen, F. & Maas, I. (2016) From neighbors to school friends? How adolescents' place of residence relates to same-ethnic school friendships. *Soc. Netw.*, **44**, 130–142.

43. Blasius, B. (2020) Power-law distribution in the number of confirmed COVID-19 cases. *Chaos Interdiscip. J. Nonlinear Sci.*, **30**, 093123.

44. Manchein, C., Brugnago, E. L., da Silva, R. M., Mendes, C. F. O. & Beims, M. W. (2020) Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies. *Chaos Interdiscip. J. Nonlinear Sci.*, **30**, 041102.

45. THE SEX, GENDER AND COVID-19 PROJECT. (2020) Are men more at risk of infection? *Men, Sex, Gender and COVID-19* https://globalhealth5050.org/the-sex-gender-and-covid-19-project/men-sex-gender-and-covid-19/. (Last accessed: 29 September 2020).

46. Glaesser, D., Kester, J., Paulose, H., Alizadeh, A. & Valentin, B. (2017) Global travel patterns: an overview. *J. Travel Med.*, **24**, 1–5.

47. Davies, N. G. *et al*. (2020) Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health*, **5**, e375–e385.

48. Shi, Q., Dorling, D., Cao, G. & Liu, T. (2020) Changes in population movement make COVID-19 spread differently from SARS. *Soc. Sci. Med.*, **255**, 113036.

49. Estrada, E. (2020) Topological analysis of SARS CoV-2 main protease. *Chaos Interdiscip. J. Nonlinear Sci.*, **30**, 061102.

50. Perc, M. (2014) The Matthew effect in empirical data. *J. R. Soc. Interface*, **11**, 20140378–20140378.

51. Estrada, E. (2020) COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. *Phys. Rep.*, **869**, 1-51.

52. WORLD BANK. (2017) Migration and remittances data. https://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data (Last accessed: 25 September 2020).

53. WORLD TOURISM ORGANIZATION. (2018) *Yearbook of Tourism Statistics: Data 2012-2016, 2018 Edition*. Madrid: World Tourism Organization.

54. UNITED NATIONS STATISTICS DIVISION. (2020) Standard country or area codes for statistical use (M49). *Methodology* https://unstats.un.org/unsd/methodology/m49/ (Last accessed: 25 September 2020).

55. Mayer, T. & Zignago, S. (2011) *Notes on CEPII's Distances Measures: The GeoDist Database*. Paris:Centre D'Études Prospective Et D'Informations Internationales.

56. Kones, J. K., Soetaert, K., van Oevelen, D. & Owino, J. O. (2009) Are network indices robust indicators of food web functioning? A Monte Carlo approach. *Ecol. Model.*, **220**, 370–382.

57. Brandes, U. & Wagner, D. (2004) Analysis and visualization of social networks. *Graph Drawing Software* (M. Jünger & P. Mutzel eds) Berlin Heidelberg: Springer, pp. 321–340.

58. Hâncean, M.-G., Slavinec, M. & Perc, M. (2020) Data from: the impact of human mobility networks on the global spread of COVID-19. [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4071499.

59. MORRIS, M., HANDCOCK, M. S. & HUNTER, D. R. (2008) Specification of exponential-family random graph models: terms and computational aspects. *J. Stat. Softw.*, **24**, 1548–7660.

60. GOODREAU, S. M., HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T. & MORRIS, M. (2008) A statnet tutorial. *J. Stat. Softw.*, **24**, 1–27.

61. GOODREAU, S. M. (2007) Advances in exponential random graph (p*) models applied to a large social network. *Soc. Netw.*, **29**, 231–248.