



OPEN

Agent-based analysis of contagion events according to sourcing locations

Mijat Kustudic, Ben Niu & Qianying Liu

The first human infected with the Covid-19 virus was traced to a seafood market in Wuhan, China. Research shows that there are comparable types of viruses found in different and mutually distant areas. This raises several questions: what if the virus originated in another location? How will future waves of epidemics behave if they originate from different locations with a smaller/larger population than Wuhan? To explore these questions, we implement an agent-based model within fractal cities. Cities radiate gravitational social attraction based on their Zipfian population. The probability and predictability of contagion events are analyzed by examining fractal dimensions and lacunarity. Results show that weak gravitational forces of small locations help dissipate infections across country quicker if the pathogen had originated from that location. Gravitational forces of large cities help contain infections within them if they are the starting locations for the pathogen. Greater connectedness and symmetry allow for a more predictable epidemic outcome since there are no obstructions to spreading. To test our hypothesis, we implement datasets from two countries, Sierra Leone and Liberia, and two diseases, Ebola and Covid-19, and obtain the same results.

As the world struggles with the Covid-19 pandemic researchers are left asking numerous questions. These questions can be directed towards the past, for example, which species are the source of the virus¹, towards the present regarding what countermeasures should be implemented² or the future, by trying to predict the economic consequences of pandemics³.

To prevent similar events in the future we must understand how this virus jumped from bats to humans. The first infected human was traced back to a seafood market in Wuhan, China⁴. Research shows that a similar type of bat lives in Yunnan province and China shares 96% of its genetic sequence with SARS-CoV-2⁵. According to these facts, we must ask a question regarding the starting scenario of the virus: what if the virus originated from another location, with a smaller/larger population than Wuhan?

Research points to a possibility for managing the risk of pandemics following the extreme value theory (EVT) manifested through power laws⁶. It considers that extremes and not averages are fundamental sources of risk. An exploration of these phenomena has prompted the use of fractal geometry⁷ and fractal reaction principles⁸. In simple terms, fractals are pattern-like shapes that can be seen in snowflakes, lightning, clouds, and numerous plants such as broccoli or ferns. There is proof that this pandemic follows some fractal structure rules and shows a similar pattern in different regions of the world^{9,10}.

The first question in this paper examines disease dynamics and cross-country disease spreading when the disease originates from different sized source populations. The second question analyses event probability, predictability, and emergent behaviors of agents as they navigate across cities.

To answer these questions, we use two main methodological innovations and approaches. The first one comes from implementing an agent-based compartmental, and reservoir arrangement of fractal cities. All cities radiate gravitational social attraction based on their Zipfian distributed population. The second methodological innovation is based on implementing fractal dimensions and lacunarity for analyzing event probability and its consequences.

To test our hypothesis, we implement real datasets from two countries, Sierra Leone and Liberia, and two diseases, Ebola and Covid-19. Ebola started in the rural areas with a small populace¹¹ while Covid-19 started in China⁴ and was imported to the capital cities of respected countries. This difference in starting location makes for a perfect ground to test our hypothesis.

The remainder of the paper is structured as follows. Section 2 focuses on reviewing literature according to different disease modeling approaches. Section 3 describes the implemented framework based on multi-dimensional spatial distribution coupled with a compartmental aspect. Section 4 describes experimental results while Sect. 5

College of Management, Shenzhen University, Shenzhen 518060, China. email: dnuiben@gmail.com

shows how real diseases spread according to obtained datasets. Section 6 discusses and compares predicted with real disease dynamics. Section 7 discusses future research directions and concludes the paper.

Literature review

Disease modeling approaches. Disease modeling is a useful tool that can give insight into disease dynamics so that an effective response can be developed. The basic notions were defined by¹² through their compartmental model that uses linear equations. New and more flexible approaches often implement some variant of artificial intelligence one of them being agent-based modeling (ABM). The modeled disease can be unspecified¹³ or use exact pathogen characteristics, such as the case of Ebola³ or Covid-19¹⁴. Observations can be made according to the governmental action and individual reaction¹⁵, self-initiated preventive actions¹⁶, or according to disease-carrying vectors¹⁷. AB models have been used to define country-level epidemiological control and prevention measures¹⁸, recently a strategy developed in this manner² has been implemented in the United Kingdom. An important topic regarding Covid-19 is the research of the economic effects of crises caused by countermeasures such as lockdowns¹⁹. Research can go into such details to explore how better handwashing can significantly affect epidemiological outcomes^{16,20}.

Role of geographical factors on disease spreading. Epidemics can be viewed as diffusion waves and should be modeled similarly to other spread and change processes²¹. Therefore, the key factors for disease spreading can be narrowed down to the spatial (geographical) factors, human behavior, and time dependence factors²². When observing distances between cities or countries, a strong explanatory potential for their populations' behavior can be found in the models that are based on Newton's gravity law, where the sizes of these elements play the role of mass²³. Smaller cities are characterized by a smaller population, however, on average, all cities have similar sizes of social groups¹⁶. In general, large cities have a greater probability to become large spreaders compared to smaller locations because of their international connectedness and tight commuting relationships²⁴.

In standard economic practice, this form has been used for analyzing migration²⁵, urban population density²⁶, transportation²⁷, and so on. The gravitational approach is commonly implemented along with epidemiological models for disease prediction or analysis where the observation focus can be on different diseases, such as measles²⁸ and influenza²⁹ or to observe disease waves, and their spatial hierarchies of concentration^{30,31,32}. Analysis can be focused on urban disease spreading³³ and discovering possible solutions for its suppression¹⁸. Due to Covid-19's prevalence, containment needs to be done by implementing non-pharmaceutical measures such as lockdowns to prevent cross-country spreading².

Fractal patterns of diseases. One approach for managing limitations of modeling comes by observing event probabilities and their "contagion event sets"⁹ where the idea of fractality greatly helps. Due to the relative mathematical similarities with epidemiological modeling, pandemics caused by Covid-19 have conceived some fractal observations. Repeating patterns, power-law behaviors and scaling properties across different regions of the world have been noted¹⁰. Scalability is also explored through different sized social networks and clusters pointing out the fractal dimensions⁹. We note a definite lack of literature in this area requiring further exploration due to the present and future dangers of the Covid-19 pandemic as well as other possible ones.

An important class of models that introduces realism through heterogeneity is based on metapopulations³³ also referred to as fractality¹⁰. These models divide the network of agents into subpopulations of different sizes. This feature is important because it allows for the implementation of cities of different sizes which is an important aspect because it sets the limit on how large a fraction of the population a single individual might meet within a given period³⁴.

There are two main reasons why the epidemic spreading is likely to follow a fractal pattern, both in line with the main characteristics of fractals, recursiveness, and self-similarity. Recursiveness can be traced to the underlying framework of epidemics, namely the population, which tends to follow a power-law function in its self-organization³⁵. These aspects of human activity in urban centers have been analyzed in the science of cities that show how people live and interact in urban environments. The reason for recursiveness lies in the scalability of spreading where the transmission dynamics of the individual level (Fig. 1) is the same as the one in the sub-population, metapopulation, or city to city level¹⁰. Since all of these levels have the same transmission dynamics it leaves the scale of observation invariant, which is in line with the inherent behavior of cities³⁵.

Self-similarity patterns of the epidemic appear all over the world and have been identified in Romania, Italy, Spain, Germany³⁶, as well as in China, the USA, Brazil, and Europe¹⁰. This self-similarity of epidemics can be viewed as a useful feature because it allows researchers to assess the current condition and predict the next modifications in the epidemiological curves³⁶.

Framework description

The used framework consists of two parts, the spatial element, and the compartmental element. The spatial part is used to observe the movement and dynamics of the population and their cross-city behavior during epidemics. The compartmental element enables pathogen transmission and its observation.

Spatial dynamics of fractal cities. People can live in a city for a long time, meeting some inhabitants often and others, not at all¹⁰. This isolation feature of social clusters prevents the exponential spreading of diseases since greater numbers of infected do not directly correlate with a greater probability of getting infected. Therefore, the probability of infection refers more to social groups and not individuals because through an individual the entire group of closest contacts is in contact with another group.

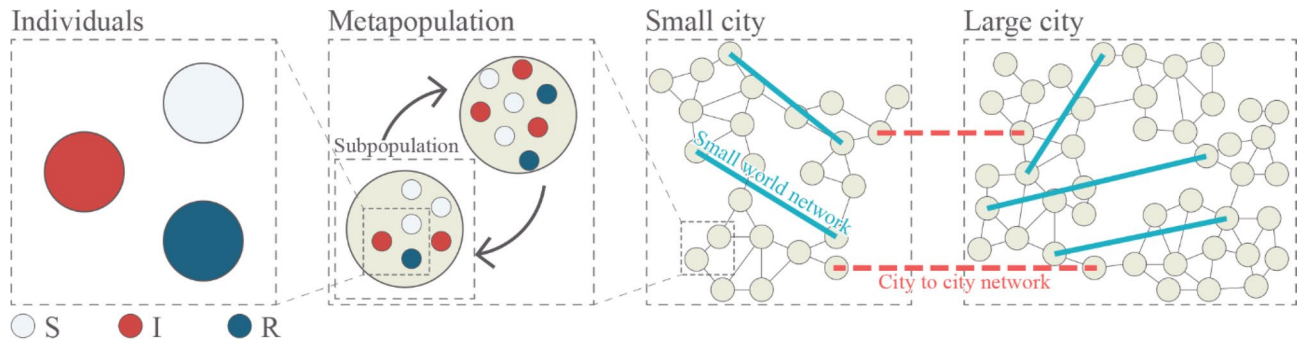


Figure 1. Shows different sized cities acting as reservoirs, they are distributed across a country. A bottom-up approach is obtained by implementing individuals, metapopulation, and subpopulations within cities. Small-world networks are implemented for obtaining circulatory dynamics.

Although the fractal scale range can be infinite, we define the lowest scale value as an individual agent. This agent is an integer and is in one of the compartmental states S, I or R . A higher scale of observation comes from incorporating the surrounding metapopulation and subpopulation of multiple cities within a country, as defined by^{10,33}. Agents are distributed across social groups of different sizes with the average one having 10 close individuals¹⁶. Figure 1 Shows the highly clustered network with a short average path length that is also highly intertwined hence it also can be defined as a network with a small-world structure (SWN)³⁷.

Cities are geographically and epidemiologically separate with the distance between them acting as a barrier inhibiting disease transmission. Since this observation coincides with the definition of epidemiological reservoirs³⁸ we use the same terminology. There are several population types based on the location of infected individuals. If there is an infected individual inside a city and the disease spreads within it, the population (city) is considered as host. Population from which the infection starts, meaning that it transmits the infection directly to another population, we define as a source population. The target population is the population of interest to the observer.

Compartmental organization. The compartmental organization is based on the SIR model¹² that consists of susceptible (S), infected (I), and recovered (R) individuals, where the S can become I while the R does not return to the previous stages ($S_t \rightarrow I_t \rightarrow R_t$). Models may incorporate more compartments (quarantine, treated, or vaccinated individuals) as well as recurring movement across compartments. The total population of each model is considered to be constant, as noted:

$$N = S_t + I_t + R_t \tag{1}$$

where the total population is N , note that we incorporate time t since the compartments may differ while still having susceptible (S_t), infected (I_t) and recovered (R_t) individuals. Formulas presented next define change between compartments:

$$S_t = -\beta S_t I_t \tag{2}$$

$$I_t = \beta S_t I_t - \gamma I_t \tag{3}$$

$$R_t = \gamma I_t \tag{4}$$

So that β shows the rate of infection, while γ is the removal rate of infected individuals. City observations are treated percentage-wise since commuting from (N_c^o) and to (N_c^d) a city due to mutual gravitational forces (F_c) changes the population of the city N_c but the country N stays constant. The city-specific susceptible population can be calculated as:

$$S_t = \sum_{c=1}^{cities} S_t^c \tag{5}$$

This formulation defines S_t as a function of all susceptible individuals spatially distributed across cities at a certain moment, other compartments use the same principle. To define and distribute the population size we use Zipf's law³⁹. To calculate the population of each city we use the following equation:

$$f(k; s, N) = \frac{1}{k^s H_{N,s}} \tag{6}$$

So that N is the total population size of a country, k is the rank of the city and s exponent represents the value that characterizes the distribution which is 1.07 according to³⁹. The calculation begins with the largest city and percolates to smaller cities so that the n th the city population is the $\frac{1}{n^s}$ of the largest city. Varying population sizes can depict different phenomena by implementing a multilayered observation instead of a node-based one.

For each shape

Divide the shape into boxes of size $r \times r$

For each box size r

Count the occupied boxes: N_r

Calculate the magnitude factor ($1/r$)

Implement equation X

End

End

Table 1. Pseudocode for the box-counting approach for calculating fractal dimensions.

Different population sizes are important for predicting population movement, as described by the law of demographic gravitation⁴⁰. The law is based on Newtonian principles of distance and mass and explains how cities have attraction forces that draw individuals to migrate or visit them, as shown by the following formula:

$$F_c = \frac{N_c^o * N_c^d}{d^2} \quad (7)$$

where F_c presents the force of attraction, N_c^o is the size of the city population of origin, N_c^d is the destination city population and d is the distance between the two cities. Cities interact via migrating agents and daily migrations do not influence their overall gravitational attraction.

Fractal dimensions. Natural and biological features are often fragmented implying the existence of a fractal dimension. Since for each naturally occurring fractal, there is a finite scaling range (zoom), the structure can become smooth (Euclidean) or rough and random (non-self-similar). In Euclidean n -space, a bounded set S can be considered statistically self-similar if S is the union of N_r non-intersecting subsets for a scaling factor r , each of which is of the form $r(S_n)$ where the N_r and S_n sets are congruent in distribution to S . We can use the fractal dimension to measure S the same way as we use a measurement tool in the Euclidean (discrete) space⁷, it is calculated:

$$FD = \frac{\log N_r}{\log \frac{1}{r}} \quad (8)$$

So that N_r is the number of self-similar (invariant) shapes and r is the corresponding scaling factor. On the Hausdorff Dimension scale (HD), a smooth line has a dimension of 1 which is a low value while the high values are present in the Mandelbrot or the Julia set with the complexity of 2. Thus, often the fractal dimension is $1 < HD < 2$.

Fractal dimension calculation. To understand the complexity of a shape we use fractal dimensions as a measure, for calculating it we will use the “Minkowski-Bouligand dimension,” or the “box-counting method,” its pseudocode is presented in Table 1. Calculations can be made with different sized boxes for more or less accurate representation potentially giving different results⁴¹. The number of boxes, for proportion, used in this paper is 160 obtaining a fractal dimension of 1.25 as in⁴¹. The calibration image is listed in references⁴².

With lower HD numbers we examine the speed and impact of the infection, a single wave it will be shown by lower numbers. If it is rebounding and multivalve it will show more complexity and thus greater numbers. Another dimension is observed through the total mean (cross country) HD where higher numbers show greater dynamics and more diversified results. To further differentiate results, we will calculate lacunarity which is a measure of the gap structure in patterns and coincides with abrupt declines in dispersal success on fractal landscapes⁴³. It can be also be viewed as a measure of “gappiness” and heterogeneity, higher numbers showing greater emptiness. Lower numbers imply a slow and steady infection while higher ones a dynamic and/or rebounding one.

$$L = \frac{1/MN \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(m, n)^2}{\left(1/MN \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} I(k, l)\right)^2} - 1 \quad (9)$$

It is defined in terms of the ratio of the variance over the mean value of the function where M and N are the sizes of the HD for the processed graph (image)⁴⁴. We will use it for analyzing the numbers of infected.

Experimental parameters. Table 2. shows experimental parameters that are applied for all scenarios. Implemented disease characteristics are based on¹⁵ with the infecting probability being 5% per contact and dis-

<i>c</i>	Number of cities	10
	Largest city population	10,000
<i>N</i>	Total Population	42,811
<i>k</i>	Zipf exponent	1.07
	Default nr. of connections	20

Table 2. Parameters for the simulations.

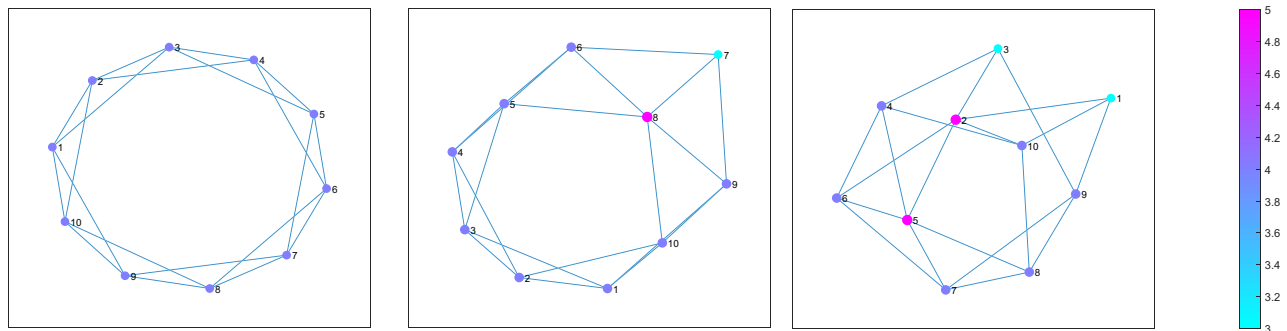


Figure 2. Spatial distributions of the three scenarios. Each scenario figure shows the sizes of the cities, by circles, and by the color bar on the right. Lines present the connections between cities that the agents may use for commuting.

ease duration 10 days. The population of the largest city is trickled down by using Eq. 7 to define the populations of other cities.

The number of simulation runs for each scenario (3) and each source (2), giving 6 in total, is 50 with each run having 350 iterations. For the spatial component, we use a 10-city setting. Since spatial organization and interconnectedness play an important role, we use three different scenarios as in Fig. 2 with the average network length of 1.5, 1.48, and 1.4 respectively.

The first scenario is based on an SWN with high interconnectedness of cities, as in³⁷. All cities are equally distant from one another, no city is better connected or presents a networking hub, the only difference is the population distribution. The second scenario has a semi-circular SWN orientation. Some cities are more isolated from the largest one, making them more difficult to reach. The third scenario has better connectedness across all city sizes and is more random in its orientation than the first two. Each city is easier to reach than in other cases.

Experimental results

To discover emergent behavior, we compare intra-scenario results. We check cross scenario results to see how disease spreading is affected by the interconnectedness of cities. We observe the maximum number of infected at a single moment which shows the burden on the health system. The standard deviation of infected population presents the dynamics of infecting. With greater incidence comes greater deviation. The total number of infected is a key indicator of the epidemic spreading, although alone it doesn't show the timeframe and dynamics.

Figure 3 is divided into two parts, the top part displays cumulative simulation results (each graph has 50) while the bottom contains the mean results of all simulations respectively. From left to right different scenarios are shown (1–3) while top to bottom is the sourcing positions of the pathogen, that being the smallest or largest city, respectively.

First scenario. We see that the situation is less favorable when sourcing from the smallest city, compared to the largest one. Furthermore, the mean HD shows that the dynamics is more predictable and “gravitates” towards the same outcome. Higher lacunarity numbers show the changing incidence rates indicating complexity and dynamics. There is a noticeable jump when infected advance from a small full reservoir to a new large one, similar to the “honeymoon effect” phenomenon⁴⁵ when effective control is used against an endemic infection resulting in an initial drop in prevalence to well below the endemic level. Afterward, it is followed by outbreaks that periodically increase prevalence above the endemic level as a consequence of a build-up of susceptible individuals⁴⁶.

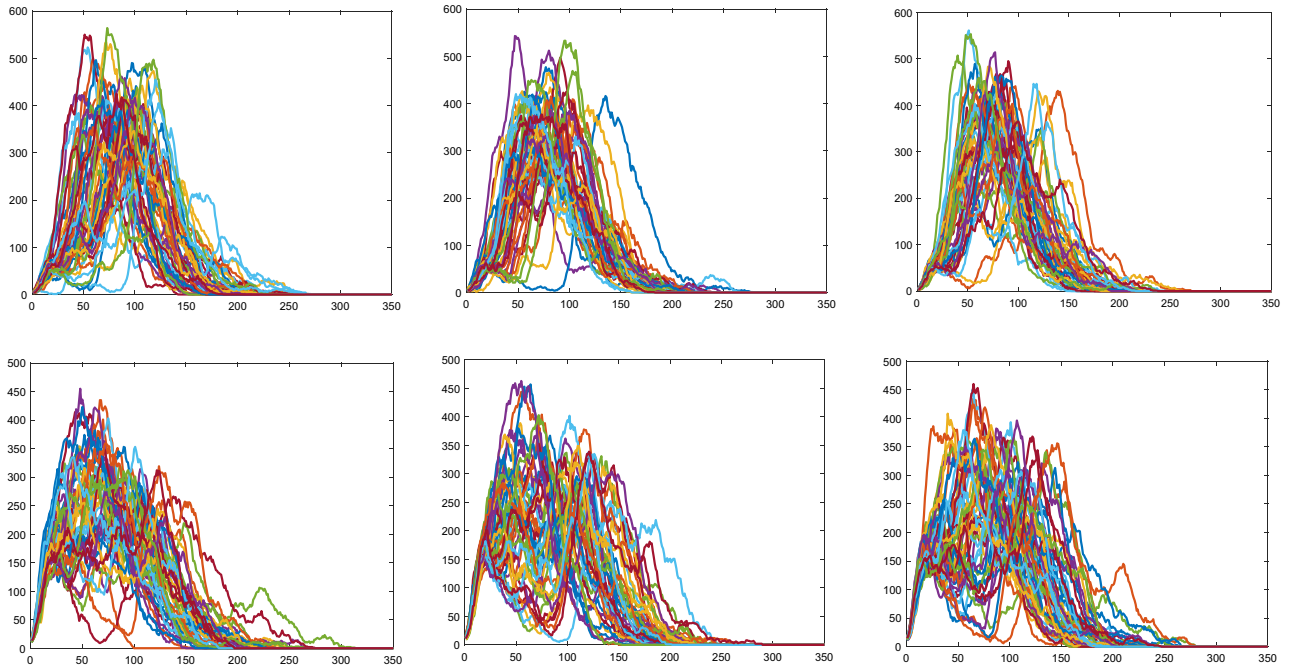
Greater infection intensity depletes the susceptible population eradicating the disease quicker. “In smaller populations, the number or density of infected hosts frequently falls to low levels, random extinction (fadeout) becomes inevitable, and the pathogen cannot persist”³⁸. Smaller sites are monitored and tested less often. The potential benefit of a disease spreading from a small location is that it is easier to control and put on lockdown blocking further propagation.

Cumulative results

Scenario 1.

Scenario 2.

Scenario 3.



Mean results

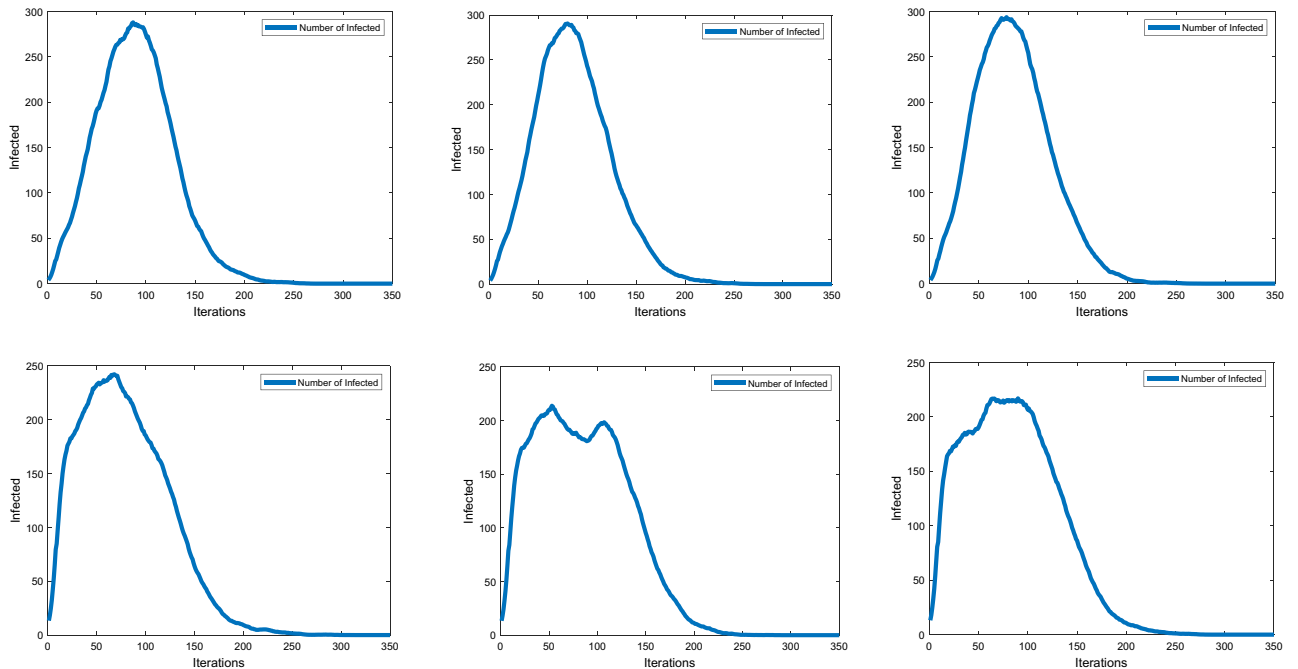


Figure 3. Shows Graphical results of simulations. Top to bottom for both cumulative and mean results are according to the sourcing (starting) positions of the pathogen, smallest and largest city, respectively.

Second scenario. We see that the smallest city as a source causes higher maximum infected, standard deviation, and the total number of infected. The reason for this comes from the gravitational pull F_c of a large city since $F_c^{Large} > F_c^{Small}$ when squared distances are adequately the same, according to (Eq. 7). It will attract

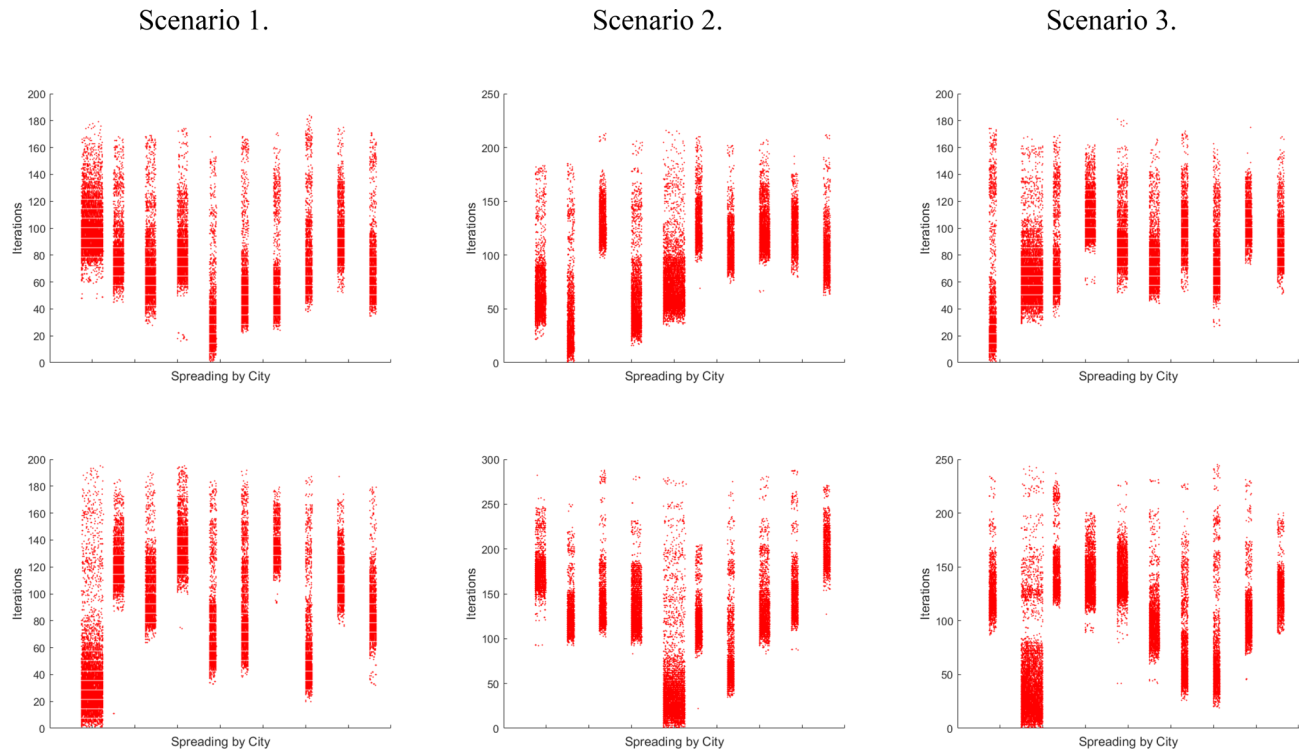


Figure 4. Shows the dynamics of simulations. Top to bottom results is according to the sourcing (starting) positions of the pathogen, smallest and largest city, respectively.

potential commuters from smaller cities around its vicinity and will keep the natives within it. Migrants from (other) smaller cities will return to their place of origin and locally spread the disease. The small reservoir is filled quicker so the infection spreads faster across smaller populations. This is combined with the higher probability of people moving from smaller to other cities, due to the small gravity, additionally increasing chances of spreading.

Third scenario. This scenario shows a lower number of max infected along with a lower standard deviation. Higher interconnectedness allows for disease to spread evenly across a country, so its moment of infection peaking is not concentrated. Although dispersed, it gives the same total number as in other scenarios. When the disease sources from a large city its population presents a large pool for potentially infected individuals. As its gravitational pull attracts individuals from other smaller locations they too participate in the social interactions of the large city and commute back to their place of origin which will help propagate the disease top-down. Due to the Zipfian distribution³⁹ (Eq. 6.) more small cities are making it more likely for people to commute from small to a large city.

Cross country comparison. Differences between scenarios are observed via the maximum number of infected at a single moment and the standard deviation values while lacunarity shows the influence of gravity. Unpredictable outcomes come from the second scenario because the lower network connectedness keeps the infection within a city. Higher connectedness helps spread the disease quicker³⁷ as in the third scenario. This combined with less isolation of other cities gives more predictable outcomes when the infection is sourced from the largest city because there are fewer boundaries for disease spreading so the outcomes are difficult to change and therefore influence.

Figure 4 shows simulation dynamics and cross-country disease spreading. A country is divided into 10 cities, with presenting the population, agents are tracked by their daily interactions. The results depict a single simulation.

In Table 3 When observing differences between HD in the 1st scenario we see lower values when the smallest city is the place of origin, pointing to the self-similarity of results. It means that this scenario is the most often repeated, and therefore probable. Greater lacunarity points to spikes of infected at a certain moment, they can be seen via the maximum infected at a single moment. Both are present in every scenario with the small city as the origin. This points to greater incidence due to the infection advancing from a small reservoir to a new large one⁴⁵, also observable through the standard deviation values. By performing the same simulations on more cities with different spatial orientations and population numbers we obtained the same results, note that initial scenario conditions are important.

Source City	Scenario 1		Scenario 2		Scenario 3	
	Small	Large	Small	Large	Small	Large
Maximum Infected	287.78	242.22	300.84	293.52	213.68	216.92
Standard deviation of Infected	100.0444	90.3565	105.3512	102.5832	84.2937	86.1186
Total Infected	27,471	27,465	27,552	27,609	27,656	27,585
Hausdorff Dimension	1.2374	1.2727	1.2365	1.24	1.279	1.2782
Lacunarity	2.1112	1.775	2.1042	2.0317	1.6787	1.6569

Table 3. Numerical results of simulations.

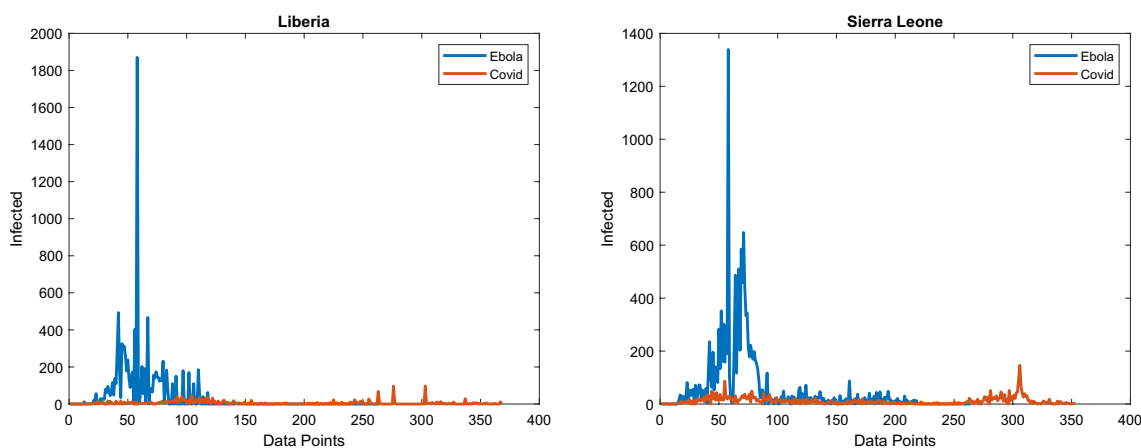


Figure 5. Shows the greater incidence of Ebola compared to Covid-19. Note that the epidemic of Ebola datapoints cutoff while the Covid-19 continue, the reason being is that the outbreak of Ebola lasted from 25th of March 2014, until the 30th of November 2015. Covid-19 epidemic datapoints start from the 15th of March 2020 until the 16th of March 2021, with an ongoing pandemic.

Comparison with real datasets

To test our hypothesis, we implement real datasets from two countries, Sierra Leone and Liberia, and two diseases, Ebola and Covid-19. Ebola started in the rural areas with a small populace while Covid-19 started in China was imported to their capital cities. This difference in starting location makes for a perfect ground for test our hypothesis. Our AB model observes cities while datasets observe districts. Characteristics of the two diseases will be examined in the discussion part.

The dataset used for Covid-19 in Liberia is based on the reports published by the National Public Health Institute of Liberia. For Sierra Leone, the data is according to the reports by The Ministry of Information and Communication. The data for Ebola is based on the Ebola World Health Organization (WHO) situation reports for respected countries. Figure 5 shows new cases of Covid-19 and Ebola in Liberia and Sierra Leone cumulatively.

Figure 6 observes the distribution of infected across country districts. Districts are sorted according to their population, with the least populated on the left, country capital cities are in the most populated districts. The starting point for Ebola was in a rural of Guinea in Guéckédou district¹¹, neighboring districts Lofa (Liberia) and Kailahun (Sierra Leone).

Covid-19 is much more present in the capital districts of Montserrado and Western Urban with high numbers appearing in their closest districts, Margibi and Western Rural respectfully. The Lofa district (Liberia) is fitting for disease transmission due to its proximity to the location of disease origin. Secondly, it has a well-developed road network further increasing dissemination²⁷.

Figure 7 shows that even though there are sparingly new cases in other districts the majority is located in the initial, largest, ones keeping to their gravitational attraction. Regarding Ebola, the reservoir nature of districts can be seen in Kailahun and Lofa. We see that the reservoir contains newly infected individuals for some time before the disease spreads to other districts.

Table 4. shows numerical standpoints of diseases. Our focus is on fractal characteristics for discerning dynamics enabling the prediction of future epidemiological curves³⁶. Low HD values of Ebola in both countries point to self-similarity and combined with high lacunarity point to high disease dynamics at a certain point, as seen in Fig. 3. For Ebola in both countries, the initial spike peaked with the maximum number of infected and was not repeated. Covid-19 has lower lacunarity values showing no significant spikes of infected. Higher HD values show an existing turmoil and dynamics of susceptibility to infection.

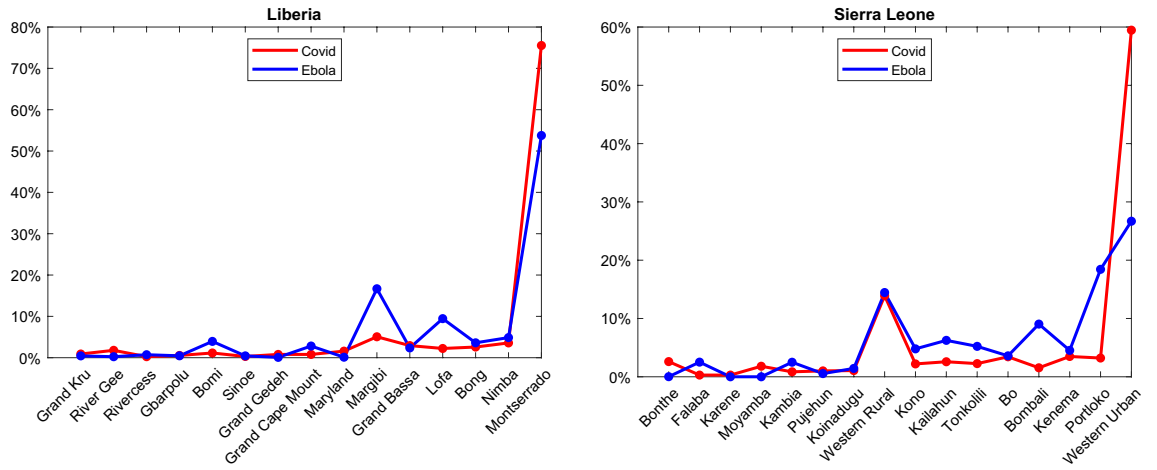
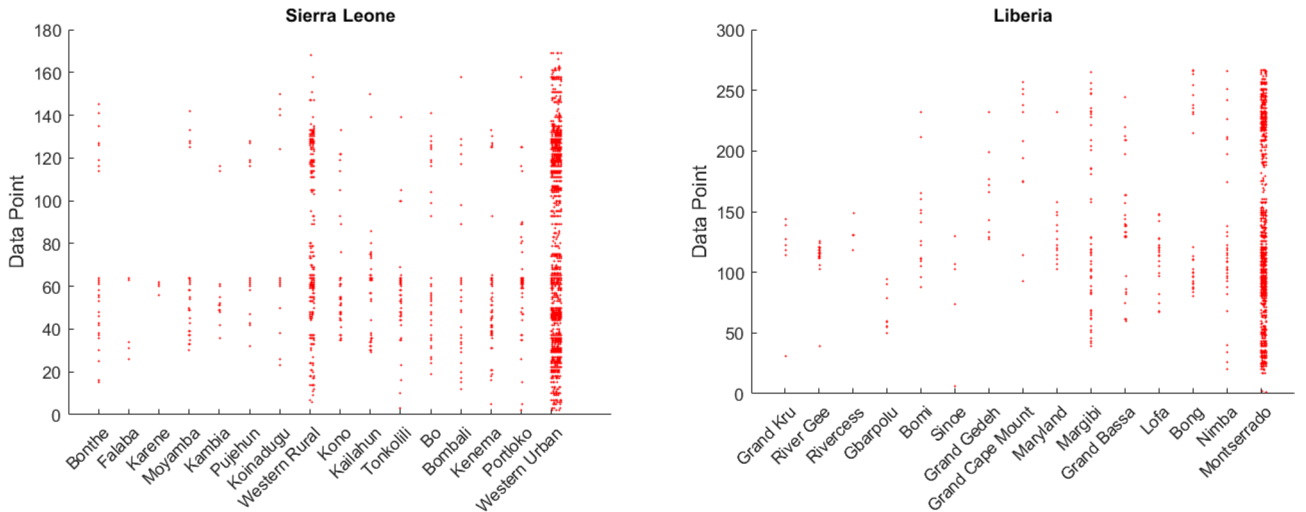


Figure 6. Observes the distribution of infected across country districts, the numbers are presented as a percentage of the total number of infected for easier comparison. The figure points to Ebola being more dispersed across districts, which is observable point by point and percentage-wise.

Covid-19



Ebola

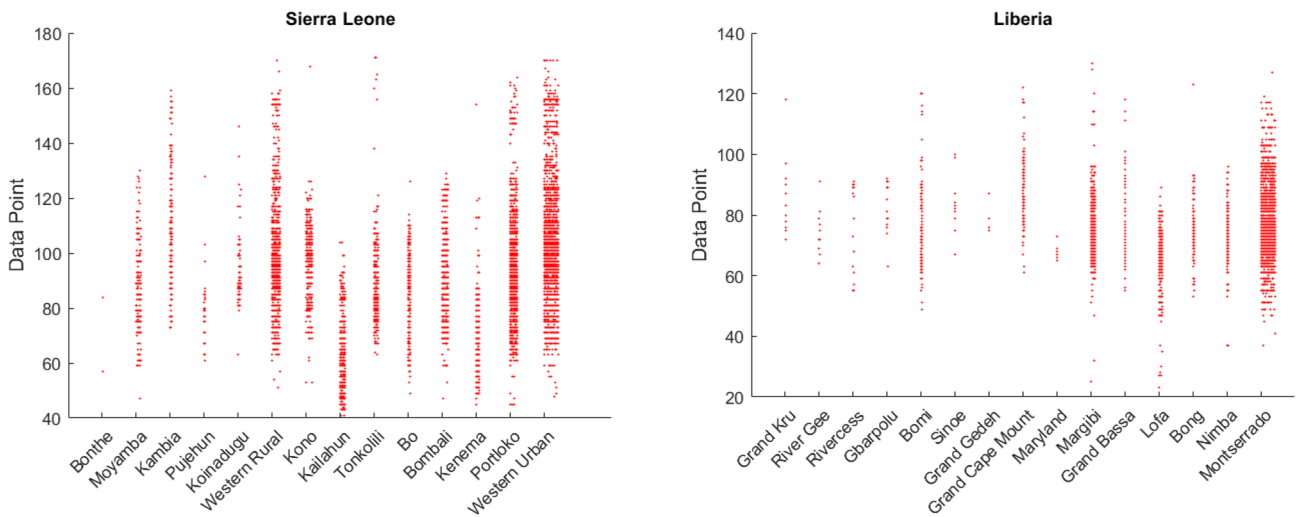


Figure 7. Shows the frequency of newly reported cases in Liberia and Sierra Leone across their districts.

Disease	Ebola		Covid-19	
	Liberia	Sierra Leone	Liberia	Sierra Leone
Maximum Infected	1870	1339	97	146
Standard deviation of Infected	137.24	127.91	10.725	14.794
Total Infected	10,678	14,124	2042	3947
Hausdorff Dimension	1.1403	1.1535	1.296	1.1309
Lacunarity	14.02	14.953	6.1716	3.6046

Table 4. Numerical results of diseases based on real datasets.

Discussion

Demographic gravitation has been used for analyzing disease spreading the focus was not on their potential for containment. Although diseases used for testing our hypothesis are not the same, they present the most documented and concentrated disease (epidemic) observation dynamics in recent decades.

Covid-19 and Ebola are Zoonoses, infectious diseases that originate from wildlife, which represents 60% of known emerging infectious diseases with their numbers growing fast³. Both of them most likely originated from bats^{5,11}. Differences between the observed diseases should have led to a different turn of events. Theoretically, these differences should enable Covid-19 to be more prevalent than Ebola because it is considerably more transmissible³. Being exposed to speaking or coughing is more common than to blood or secretions, which are necessary for the transmission of Ebola. Individuals might be asymptomatic and infectious while Ebola patients are not contagious until they develop symptoms⁴⁷. Greater lethality of Ebola³ should have slowed down its progress. Ebola aftermath shows that it has hit more districts and has spread quicker. Even though Covid-19 has started from the largest reservoir it still has not reached its prevalence as Ebola did.

Disease sourcing from rural areas points to behaviors described by our model, those being higher values of lacunarity and maximum infected indicating spikes of infected at a certain period. Ebola also has low HD values in both countries pointing to self-similarity, which is also evident in our simulations. Covid-19 is sourced from capital cities and has lower lacunarity values and no significant spikes of infection. Higher HD values show an existing turmoil and dynamics of susceptibility to infection. The majority of infected are located in the initial cities which keeps to our notion of their gravitational attraction.

Research shows that connectivity matters more than density in the spread of the Covid-19 pandemic^{24,48}. Large cities have a greater probability to become large spreaders compared to smaller locations, namely because of their international connectedness and tight commuting relationships²⁴. On the other hand, porous land borders, as in the observed countries⁴⁹, enable more population movement compared to airports.

In the case of this paper, both diseases have been introduced from abroad but to different sized locations. This puts high emphasis on connectivity and commuting, but the discussion regarding commuting must be linked with its direction. Demographic gravitation explains that large numbers of people act as an attractive force towards other people to migrate in that direction which is further strengthened by economic factors⁴⁸. This is in line with our gravitational observation where small locations that have less attractive force are not able to stop people migrating towards larger locations. Large locations keep individuals within them due to the same force. When daily commute is introduced individuals from smaller locations more often go to large ones than otherwise. All locations have similar sizes of social groups¹⁶ because they depend not on the size of the location but individuals. When an infected is introduced to a small reservoir, the reservoir is filled quicker because of the greater infection probability per capita, as can be seen in several locations^{50,51}. Now, this small reservoir has greater per capita infection probability, and individuals that are drawn by the greater attraction force of larger locations making them a more dangerous place for the infection to start from.

There are several limitations to our study. Disease spreading is influenced by other important factors that we did not describe. Two diseases under scrutiny are not the same and behave differently in different locations. Although the current situation points to our findings being correct, the Covid-19 pandemic is ongoing and the numbers will continue to change. Circumstances in Liberia and Sierra Leone do confirm our findings but they are relatively small-sized and disease dynamics might not fold the same way in other countries. Finally, our findings are based on numerous simulations which are averaged. Real events happen once and can come out of the most unpredictable sources. Nonetheless, our findings point in the same direction.

Conclusion

In this paper, we combine a spatial agent-based model with a compartmental (SIR) model to observe multi-city and cross-country epidemics. We observe that gravity maintains the infection inside the city when the sourcing position of the pathogen is the largest city. Sourcing from the smallest city quickly dissipates the infection across a country. To test our hypothesis, we implement datasets from two countries and two diseases. Disease sourcing from rural areas points to the same behaviors as described by our model.

Our future research will be based on implementing different scenarios regarding infection transmission rates and source population locations expanding them to cities of various sizes and levels of connectedness. We will continue to monitor the Covid-19 pandemic and will compare outcomes with our predictions.

Received: 14 October 2020; Accepted: 29 June 2021

Published online: 06 August 2021

References

1. Latinne, A., Hu, B., Olival, K. J., *et al.* Origin and cross-species transmission of bat coronaviruses in China. Preprint. bioRxiv (2020). doi:<https://doi.org/10.1101/2020.05.31.116061>
2. Ferguson, N. M., Laydon, D., & Nedjati-Gilani, G. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Imperial College COVID-19 Response Team (2020). <https://doi.org/10.25561/77482>
3. Stoop, N., Desbureaux, S., Kaota, A., Lunanga, E. & Verpoorten, M. Covid-19 vs. Ebola: Impact on households and small businesses in North Kivu, Democratic Republic of Congo. *World Dev.* **140**, 105352. <https://doi.org/10.1016/j.worlddev.2020.105352> (2020).
4. World Health Organization. Novel Coronavirus (2019-nCoV) situation reports. Technical Report 1–183, WHO, July 2020. URL <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. (2020) [retrieved March 30, 2021]
5. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273. <https://doi.org/10.1038/s41586-020-2012-7> (2020).
6. Cirillo, P. & Taleb, N. N. Tail risk of contagious diseases. *Nat. Phys.* **16**(6), 606–613 (2020).
7. Mandelbrot B. B. *Fractal Geometry of Nature*. (pp. 27–28) (1982). San Francisco, CA: Freeman
8. Kopelman, R. Fractal Reaction Kinetics. *Science (New York, N.Y.)* **241**, 1620–1626. <https://doi.org/10.1126/science.241.4873.1620> (1988).
9. Materassi, M. Some fractal thoughts about the COVID-19 infection outbreak. *Chaos Solitons Fract* **4**, 100032. <https://doi.org/10.1016/j.csf.2020.100032> (2019).
10. Abbasi, M. *et al.* Fractal signatures of the COVID-19 spread. *Chaos Solitons Fract.* **140**, 1. <https://doi.org/10.1016/j.chaos.2020.110119> (2020).
11. World Health Organization. <https://www.who.int/csr/disease/ebola/one-year-report/virus-origin/en/> (2015a) [retrieved March 30, 2021]
12. Kermack, W. O. & McKendrick, A. G. Contributions to the mathematical theory of epidemics-I. *Proc. R. Soc.* **115A**, 700–721 (1927).
13. Mei, S. *et al.* Simulating city-level airborne infectious diseases. *Comput. Environ. Urban Syst.* **51**, 97–105 (2015).
14. Fanelli, D. & Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy, and France. *Chaos Solitons Fract.* **134**, 109761. <https://doi.org/10.1016/j.chaos.2020.109761> (2020).
15. Lin, Q. *et al.* A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* **93**, 211–216. <https://doi.org/10.1016/j.ijid.2020.02.058> (2020).
16. Mao, L. Predicting Self-Initiated Preventive Behavior Against Epidemics with an Agent-Based Relative Agreement Model. *J. Artif. Soc. Soc. Simul.* **18**(4), 6. <https://doi.org/10.18564/jasss.2892> (2015).
17. Hunter, E., Mac, N. B. & Kelleher, J. A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology. *J. Artif. Soc. Soc. Simul.* **20**(3), 2. <https://doi.org/10.18564/jasss.3414> (2017).
18. Halloran, M. E. *et al.* Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci. USA* **105**(12), 4639–4644 (2008).
19. Manotosh, M. *et al.* A model-based study on the dynamics of COVID-19: Prediction and control. *Chaos Solitons Fract* <https://doi.org/10.1016/j.chaos.2020.109889> (2020).
20. Poletti, P., Caprile, B., Ajelli, M. & Pugliese, A. Merler S (2009): Spontaneous behavioural changes in response to epidemics. *J Theor Biol.* **260**(1), 31–40. <https://doi.org/10.1016/j.jtbi.2009.04.029> (2009).
21. Haggett, P. *The geographical structure of epidemics* 29–30 (Oxford University Press, 2000).
22. Gonçalves, S. Key factors in disease spreading: Spatial heterogeneity, time dependence, and human behavior. *Phys. Life Rev.* <https://doi.org/10.1016/j.plrev.2016.10.006> (2016).
23. Carrothers, V. A historical review of the gravity and potential concepts of human relations. *J. Am. Inst. Plan.* **22**, 94 (1958).
24. Hamidi, S., Sabouri, S. & Ewing, E. Does Density Aggravate the COVID-19 Pandemic?. *J. Am. Plann. Assoc.* **86**(4), 495–509. <https://doi.org/10.1080/01944363.2020.1777891> (2020).
25. Karemera, D., Oguledo, V. & Davis, B. A gravity model analysis of international migration to North America. *Appl. Econ.* **32**, 1745–1755 (2000).
26. Wang, F. & Guldmann, J. M. Simulating urban population density with a gravity-based model. *Socio-Econ. Plan. Sci.* **30**, 245–256 (2005).
27. Jung, W. S., Wang, F. & Stanley, H. E. Gravity model in the Korean highway. *Europhys. Lett.* **81**, 48005 (2008).
28. Xia, Y. C., Bjornstad, O. N. & Grenfell, B. T. Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *Am. Nat.* **164**, 267–281 (2004).
29. Li, X., Tian, H., Lai, D. & Zhang, Z. Validation of the gravity model in predicting the global spread of influenza. *Int. J. Environ. Res. Public Health* **8**, 3134–3143. <https://doi.org/10.3390/ijerph8083134> (2011).
30. Parker, J. & Epstein, J. M. A distributed platform for global-scale agent-based models of disease transmission. *ACM Trans. Model. Comput. Simul.* **22**(1), 2. <https://doi.org/10.1145/2043635.2043637>. PMID:24465120;PMCID:PMC3898773 (2011).
31. Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
32. Barrios, J. M. *et al.* Using the gravity model to estimate the spatial spread of vector-borne diseases. *Int. J. Environ. Res. Public Health* **9**(12), 4346–4364. <https://doi.org/10.3390/ijerph9124346> (2012).
33. Hackl, J. & Dubernet, T. Epidemic spreading in urban areas using agent-based transportation models. *Future Internet.* **11**, 92. <https://doi.org/10.3390/fi11040092> (2019).
34. Lund, H., Lizana, L. & Simonsen, I. Effects of city-size heterogeneity on epidemic spreading in a metapopulation: A reaction-diffusion approach. *J Stat Phys* **151**, 367–382. <https://doi.org/10.1007/s10955-013-0690-3> (2013).
35. Batty, M. & Longley, P. *Fractal Cities*, (18–19) Academic Press (1994).
36. Păcurar, C. & Necula, B. An analysis of COVID-19 spread based on fractal interpolation and fractal dimension. *Chaos Solitons Fract.* **139**, 110073. <https://doi.org/10.1016/j.chaos.2020.110073> (2020).
37. Watts, D. & Strogatz, S. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442. <https://doi.org/10.1038/30918> (1998).
38. Haydon, D. T., Cleaveland, S., Taylor, L. H. & Laurenson, M. K. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerg Infect Dis.* **8**(12), 1468–1473. <https://doi.org/10.3201/eid0812.010317> (2002).
39. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, 1949).
40. Stewart, J. Q. Demographic Gravitation: Evidence and Applications. *Sociometry* **11**(1/2), 31–58 (1948).
41. Mandelbrot, B. B. How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science* **156**(3775), 636–638. <https://doi.org/10.1126/science.156.3775.636>. PMID17837158 (1967).
42. Image link for proportion calibration: <http://www.aiecon.org/staff/shc/course/annga/RR/main/How%20Long%20is%20the%20Coast%20of%20Great%20Britain.files/Britain.gif>. Retrieved April 15, 2021
43. With, K. A. & King, A. W. Dispersal success on fractal landscapes: a consequence of lacunarity thresholds. *Landscape Ecol.* **14**, 73–82. <https://doi.org/10.1023/A:1008030215600> (1999).
44. Petrou, M. & Sevilla, P. G. *Image Processing: Dealing with Texture* (Wiley, 2006).

45. Hollingsworth, B., Okamoto, K., & Lloyd, A. After the Honeymoon, the Divorce: Unexpected Outcomes of Disease Control Measures Against Endemic Infections (2019). <https://doi.org/10.1101/608653>.
46. McLean, A. R. & Anderson, R. M. Mathematical modelling: A key to control of infectious diseases in man and animals. *Epidemiol. Infect.* **133**(Suppl. 1), S41–S43 (1988).
47. Dean, N. E., Halloran, M. E., Yang, Y. & Longini, I. M. Transmissibility and pathogenicity of ebola virus: A systematic review and meta-analysis of household secondary attack rate and asymptomatic infection. *Clin. Infect. Dis.* **62**(10), 1277–1286 (2016).
48. Sy, K. T., Martinez, M., Rader B., & White, L. Socioeconomic disparities in subway use and covid-19 outcomes in New York City. medRxiv: the preprint server for health sciences (2020). <https://doi.org/10.1101/2020.05.28.20115949>.
49. World Health Organization. <https://www.who.int/csr/disease/ebola/ebola-6-months/guinea/en/> (2015b) [retrieved March 30, 2021]
50. <https://www.nytimes.com/2020/11/12/us/coronavirus-crisis-united-states.html>. retrieved 17. 03. 2020
51. <https://www.latimes.com/california/story/2020-08-12/second-surge-california-coronavirus-deaths-nearly-double>. Retrieved 17. 03. 2020

Acknowledgements

The work described in this paper was supported by The Natural Science Foundation of China (Grant Nos.71971143, 71571120), Natural Science Foundation of Guangdong Province (2020A1515010749), Key Research Foundation of Higher Education of Guangdong Provincial Education Bureau (2019KZDXM030), Guangdong Province Soft Science Project (2019A101002075), Guangdong Province Educational Science Plan 2019 (2019JKCY010) and Guangdong Province Postgraduate Education Innovation Research Project (2019SFKC46).

Author contributions

M.K. conceived the study and conducted the analysis. Q.L. and B.N. interpreted the results. All authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.N.

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

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



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

© The Author(s) 2021