



Strong spatial embedding of social networks generates nonstandard epidemic dynamics independent of degree distribution and clustering

David J. Haw^{a,1} , Rachael Pung^{a,1}, Jonathan M. Read^b, and Steven Riley^{a,2}

^aMedical Research Council Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London W2 1PG, United Kingdom; and ^bCentre for Health Informatics Computing and Statistics, Lancaster Medical School, Lancaster University, Lancaster LA1 4YW, United Kingdom

Edited by Alan Hastings, University of California, Davis, CA, and approved August 5, 2020 (received for review July 31, 2019)

Some directly transmitted human pathogens, such as influenza and measles, generate sustained exponential growth in incidence and have a high peak incidence consistent with the rapid depletion of susceptible individuals. Many do not. While a prolonged exponential phase typically arises in traditional disease-dynamic models, current quantitative descriptions of nonstandard epidemic profiles are either abstract, phenomenological, or rely on highly skewed offspring distributions in network models. Here, we create large socio-spatial networks to represent contact behavior using human population-density data, a previously developed fitting algorithm, and gravity-like mobility kernels. We define a basic reproductive number R_0 for this system, analogous to that used for compartmental models. Controlling for R_0 , we then explore networks with a household-workplace structure in which between-household contacts can be formed with varying degrees of spatial correlation, determined by a single parameter from the gravity-like kernel. By varying this single parameter and simulating epidemic spread, we are able to identify how more frequent local movement can lead to strong spatial correlation and, thus, induce subexponential outbreak dynamics with lower, later epidemic peaks. Also, the ratio of peak height to final size was much smaller when movement was highly spatially correlated. We investigate the topological properties of our networks via a generalized clustering coefficient that extends beyond immediate neighborhoods, identifying very strong correlations between fourth-order clustering and nonstandard epidemic dynamics. Our results motivate the observation of both incidence and socio-spatial human behavior during epidemics that exhibit nonstandard incidence patterns.

random networks of reasonable size, the degree distribution is well-approximated by a Poisson in which variance is equal to the square of the mean. In contrast, for finite, scale-free networks, the offspring distribution is power-law-like, with a much higher variance. Further, distributions of second-order phenomena describe connections of length two. For example, the local clustering coefficient is a second-order property, defined to be the neighborhood density of a given node (5). For a limited set of network types, we can use analytical expressions for higher moments of the degree distribution to calculate key properties of their potential epidemics, such as the probability of epidemic establishment and cumulative incidence (6, 7). Although these higher-order moments are tractable for some special cases, they are seldom the primary target of theoretical studies. Semi-empirical networks that arise from detailed simulations (8) may have complex higher moments; however, their impact on epidemic dynamics is obscured by the variance of their offspring distribution (e.g., ref. 9). Here, we explicitly control our network-generation algorithm so as to have nontrivial higher-order structure, while maintaining a Poisson degree distribution and a prespecified clustering coefficient.

Epidemics can also be understood in terms of compartmental models, which are more tractable mathematically and are equivalent to large network models with very simple topologies (10). Key features of epidemic incidence curves are often explained by dynamics associated with these models (11, 12). Numerical solutions to multitype susceptible–infectious–removed-like compartmental models are easier to obtain than for many topologies of network and can explain the initial growth phase (13), the

networks | epidemics | subexponential | clustering

Epidemics are frequently conceptualized as resulting from the transmission of a pathogen across a network. Directly transmitted pathogens propagate through susceptible human populations and create directed infection trees with an offspring-like process (1). Each node may be a different type [e.g., children may be more infectious than adults (2)], and individuals with many contacts are more likely to cause infection than those with fewer contacts (3). Although difficult to observe, infection trees describe a real biological process: These pathogens do not reproduce outside of a human host, so the founding pathogen population for an infectee comes directly from their infector. Further, we can conceptualize that infection trees occur when a true offspring process is constrained to pass through a social network (4, 5), with infection occurring according to a specified probability when an edge exists between a susceptible and an infectious individual.

The properties of different contact network types can be described by distributions associated with their topology (5). First-order network properties are associated with first-order connections, as defined by the degree distribution. For finite

Significance

Epidemics are typically described using a standard set of mathematical models that do not capture social interactions or the way those interactions are determined by geography. Here, we propose a model that can reflect social networks influenced strongly by the way people travel, and we show that they lead to very different epidemic profiles. This type of model will likely be useful for forecasting.

Author contributions: D.J.H., J.M.R., and S.R. designed research; D.J.H., R.P., and S.R. performed research; D.J.H., R.P., and S.R. contributed new analytic tools; D.J.H., R.P., and S.R. analyzed data; and D.J.H., R.P., J.M.R., and S.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹D.J.H. and R.P. contributed equally to this work.

²To whom correspondence may be addressed. Email: s.riley@imperial.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1910181117/-DCSupplemental>.

First published September 8, 2020.

timing and amplitude of the peak (14), the epidemic duration (15), and the total number of cases (16). These models can efficiently describe many different types of complexity, such as age-specific susceptibility and transmissibility (17), behavioral risk groups (18), and, with increasing frequency, geographical location (19).

The basic reproductive number has been defined for both compartmental models and network models. For compartmental models, the reproduction number is conditional on the system having a well-defined period of exponential growth (20) and is defined as the average number of new infections generated by a typically infectious individual in an otherwise infectious population (20). The word “typically” is somewhat overloaded in this definition: During the exponential phase, a system with heterogeneous population will reach a steady-state distribution of infectives, corresponding to the eigenstate of the renewal process.

For network models, the basic reproduction number is most frequently defined as the expected ratio of cases between the first (seed) and second generations of infection. In homogeneous networks, this is equal to the product of the average degree and the probability of transmission per link per generation. However, many studies of epidemics on networks involve high-variance degree distributions (9, 21), and so this quantity must be modified to account for excess degree (21, 22). Here, we use R^* to denote the expected first-generation ratio if a network is homogeneous, defined to be the expected number of cases in the second generation divided by the number in the first generation. Our R^* is therefore consistent with ρ_0 , as defined in ref. 21, although we choose not to adjust for overdispersion, because we condition our network construction on this distribution having low variance.

The reproduction number for networks has also been defined to be more consistent with its definition for compartmental models. In ref. 23, R_* was defined as an asymptotic property of epidemics that were guaranteed to have an exponential phase when they occurred on infinitely large networks. We define our R_0 to be a finite-network approximation to this R_* in ref. 23. This R_0 is well-defined during periods of exponential growth.

Both compartmental and network models can be embedded in space (19). Each node can have a location in space, while each compartment can refer to a single unit of space. Node density can be assigned according to known population densities, and compartments can be assigned equal spatial areas, but different numbers of hosts. In general, the risk of infection passing between two people decreases as the distance between their home location increases. The propensity of nodes to form links across space or for infection to spread between compartments can be quantified by using mobility models borrowed from geography (24), such as the gravity and radiation models. Here, we are specifically interested in how the overall topology of a spatially embedded network model can be driven by different movement assumptions and, thus, drive the gross features of the epidemics that occur on the network.

Results

We used an existing variant of the Metropolis–Hastings algorithm (10) to create a spatially embedded bipartite network of homes and workplaces consistent with the population density of Monrovia, Liberia, and with three illustrative movement scenarios (SI Appendix, Fig. S1). An individual’s propensity to choose a given workplace was determined by the distance between their home and workplace and parameters of a gravity-like kernel. The kernel was inversely proportional to distance raised to the power α , with movement scenarios generated solely by changing the value of α : a control value $\alpha=0$ that removed the embedding and produced a *nonspatial* model; a *wide* ker-

nel with $\alpha=3$ typical of developed populations (10, 25); and a *highly local* kernel with $\alpha=6$ representing less-developed populations (SI Appendix, Fig. S1C compared with rural Huangshan in ref. 26). The resulting distributions of distances from home to work were driven strongly by our choice of α , with 95% of

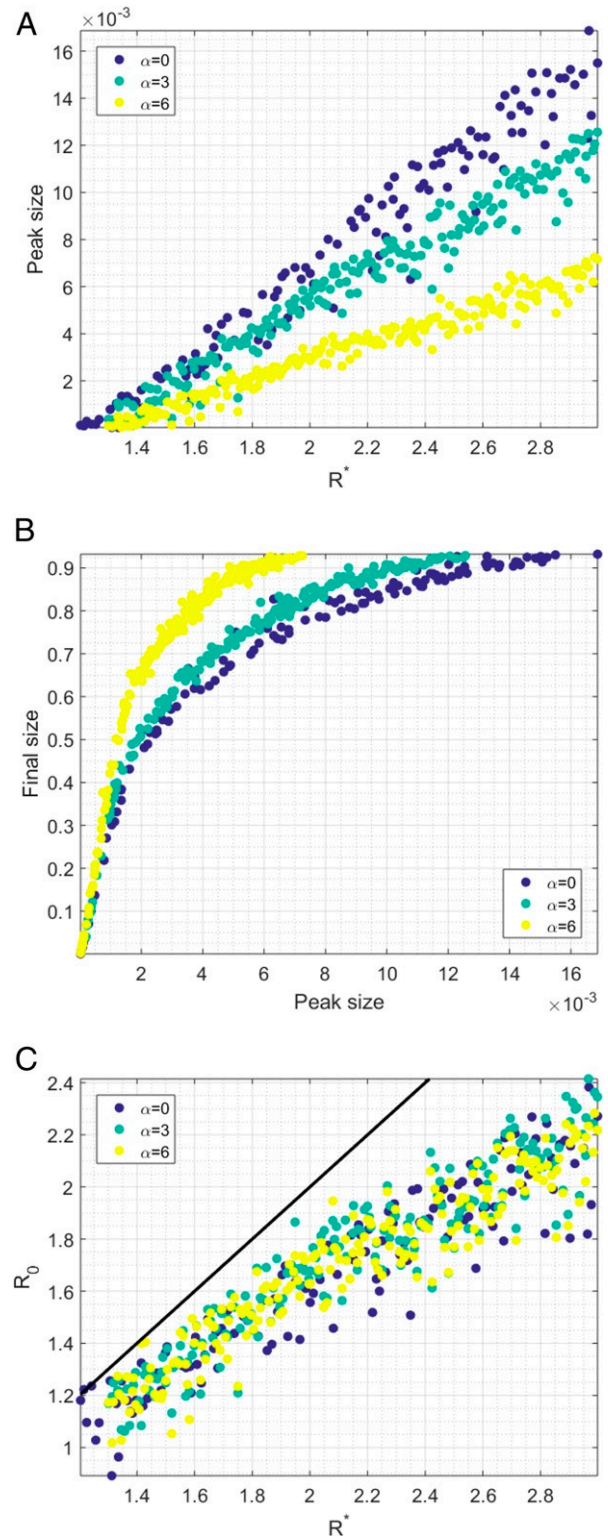


Fig. 1. For each set of parameters drawn from the Latin hypercube, and for $\alpha = 0, 3, 6$, we show relationships between R^* and peak size (A), peak size and final size (B), and R^* and R_0 (C) (with the line $R_0 = R^*$ shown in black).

journeys less than 24.12 km for $\alpha=0$; less than 12.91 km for $\alpha=3$; and less than 6.68 km for $\alpha=6$. Workplace links were dissolved into links between individuals in different households, resulting in a network of cliques (households) that were linked according to α .

The choice of movement kernel used to create the household-workplace networks affected gross features of simulated epidemics, even when controlling for other aspects of the network topology (Fig. 1). Unipartite contact networks between households were obtained from the bipartite network of households and workplaces and were dependent on three parameters: mean household size h , mean number of workplace links v , and probability of forming a link in the workplace p_w . The mean workplace size w and mean degree of the network were determined by these parameters: $w = v/p_w + 1$, $\langle k \rangle = h - 1 + v$. Across a broad range of plausible values for h , v , and p_w , very local movement ($\alpha=6$) produced later epidemics than did typical developed-population movement ($\alpha=3$) or spatially random mixing ($\alpha=0$; Fig. 1A). Similarly, time to extinction

was later for very local movement ($\alpha=6$) compared with more frequent, longer-distance movement ($\alpha=3$) or the absence of spatial embedding ($\alpha=0$). We calculated the coefficient of variation of the degree distribution $C_v^2 = \langle k^2 \rangle / \langle k \rangle - 1 \sim 0.1$ for each network, independently of α (21).

Each simulation was assigned a value of R^* , the average number of cases in the first generation per seed infection. For moderate to high values of the first-generation ratio R^* , there was very little difference in the final size of the outbreak for the different movement assumptions. However, for low values of $R^* < 1.8$, the average final size of the outbreak was substantially smaller for more local kernels. This was driven by a higher probability of extinction when more local movement was assumed. The difference in final size driven by α was no longer present when we controlled for extinction (SI Appendix, Fig. S2).

The choice of movement scenario had a substantial impact on peak incidence, even when R^* was high and there was little difference in the final sizes (Figs. 1B and 2, rows 1 and 2). For example, for parameters with first-generation ratios in the

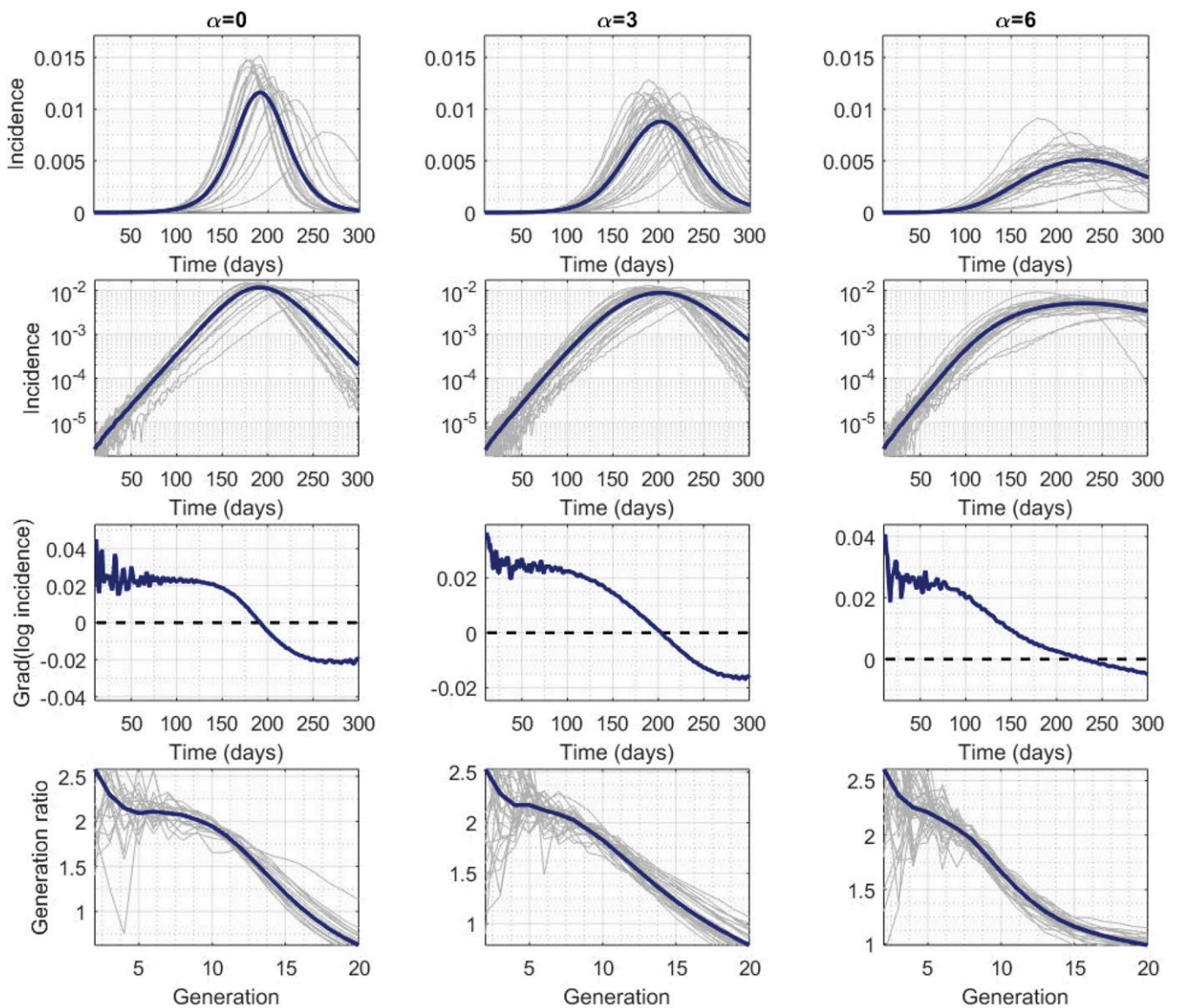


Fig. 2. Columns correspond to network structures with $\alpha=0, 3$, and 6 and simulations with $R_0 \in (2, 2.2]$. Exponential growth in real time is indicated by straight lines (second row) and horizontal lines (third row); horizontal lines in the bottom row indicate exponential growth by generation. SI Appendix, Figs. S3–S5 show results for a wider range of R_0 values for $\alpha=0, 3, 6$.

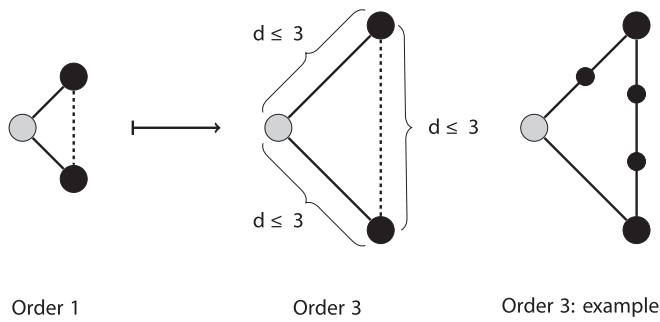


Fig. 3. A schematic showing the generalization of clustering coefficient CC^1 to higher orders CC^m : CC_i^m measures the density of paths of length $d \leq m$ between the up-to- m neighbors of node i (where node i is shown in gray).

range [1.8, 2.2], average peak daily incidence as a fraction of the total population was 6.5×10^{-3} for random spatial movement, 5.4×10^{-3} for movement assumptions typical of developed populations, and 3.0×10^{-3} when highly local movement was assumed. The relationship between peak height and first-generation ratio appeared to be strongly linear, with correlation coefficients 0.9778, 0.9826, and 0.9806 for $\alpha = 0, 3$, and 6, respectively.

The relationship between peak incidence and final size for the three movement scenarios illustrates further how clustering within the network directly affects gross features of an epidemic. Peak incidence was observed prior to final size during an epidemic. For the same peak height, local movement gave substantially larger final sizes. For peak daily incidences in the range $[3 \times 10^{-3}, 6.5 \times 10^{-3}]$, the final size of the outbreak was 68% when random spatial movement was assumed, 74% when movement was assumed to be typical of developed populations, and 84% when highly local movement was assumed.

For all movement scenarios, the basic reproductive number R_0 was smaller than the first-generation ratio R^* and different from the expected number of secondary cases generated by a single seed in an otherwise-susceptible population. The duration of the exponential phase can be seen when incidence is plotted on a log scale: A constant gradient of log incidence is evidence of exponential growth (Fig. 2, third row). However, in a network model with clearly defined generations, the generation ratio can also be used to define exponential growth: If the ratio of incidence between generation $n + 1$ and n is the same as the ratio between generations n and $n - 1$, then we can claim to have identified a period of exponential growth (*Materials and Methods* and Fig. 2). The value of that constant observed ratio is the basic reproductive number R_0 (20).

Incidence grew exponentially for a much shorter time for highly local movement than it did for a wider movement kernel, or for nonspatial networks, even when we controlled for R_0 to be within a narrow range (e.g., (2, 2.2]; Fig. 2). Despite this being a relatively large population, there was no obvious period of exponential growth when we assumed highly local movement. Therefore, given that the basic reproductive number is defined for a genuine renewal process—and its implied exponential growth (20)—it could be argued that R_0 does not exist for some of these networks for our model parameters. However, we did assign a value of R_0 for all simulations based on the most similar subset of consecutive early generations (*Materials and Methods*). The amplitude of the difference was not driven in any obvious way by the underlying assumptions used to create the networks. These patterns were not specific to the range of values for R_0 (*SI Appendix, Figs. S3–S5*).

Analysis of the higher-order structure of the networks suggests that movement scenarios were driving the observed

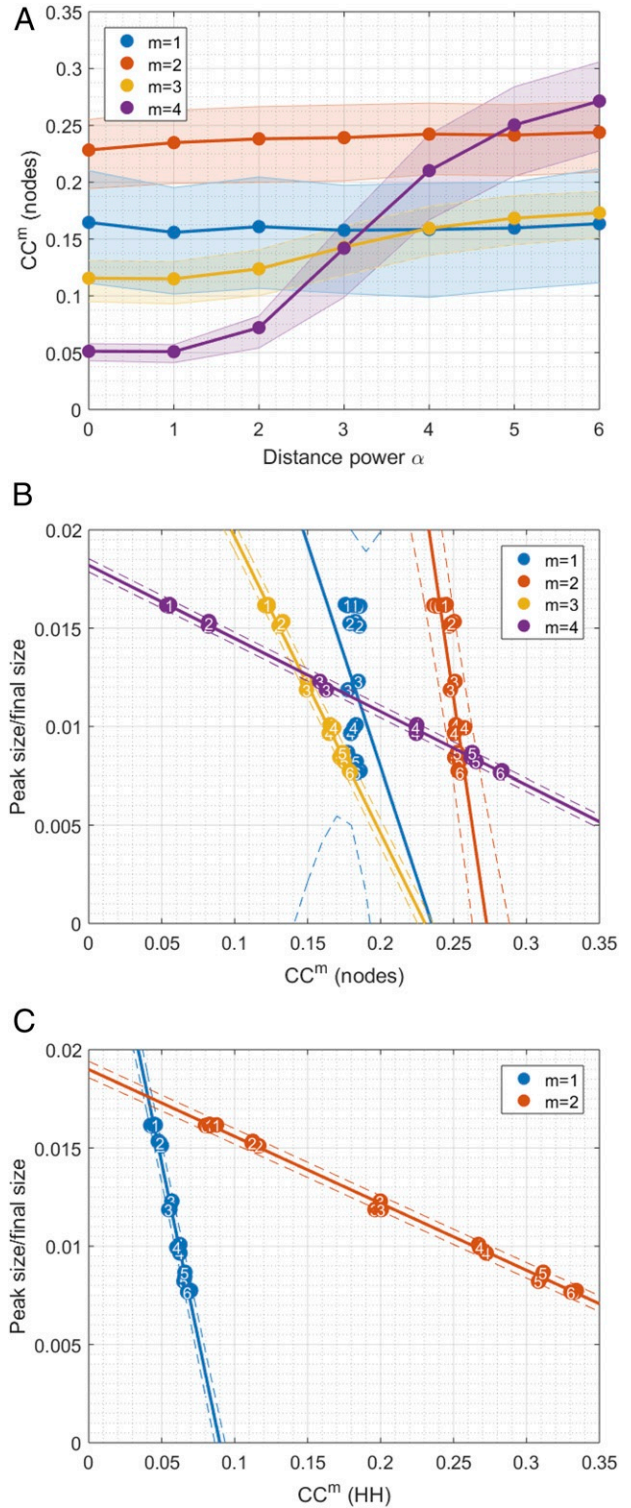


Fig. 4. (A) The 25th, 50th, and 75th percentiles of order- m clustering CC^m on networks constructed with different values of α and $h = 5, w = 50, p_w = 0.14, \langle k \rangle = 10$ and $R_0 \in [2, 2.2]$. Plot shows mean values over three different networks for each parameter set. (B) Using peak size as a crude metric for subexponential growth (given a fixed range for R_0), we see linear trends emerging with higher orders of clustering. Plot shows one point per network, with three networks generated for each parameter set, and the mean peak size over 10 independently simulated epidemics. All points are numbered with the corresponding value of α . (C) Similarly for the household-only networks. Solid lines show linear fits to data, and dotted lines show 95% CIs. Values of linear correlation coefficient and gradient of fits are given in *SI Appendix, Table S2*.

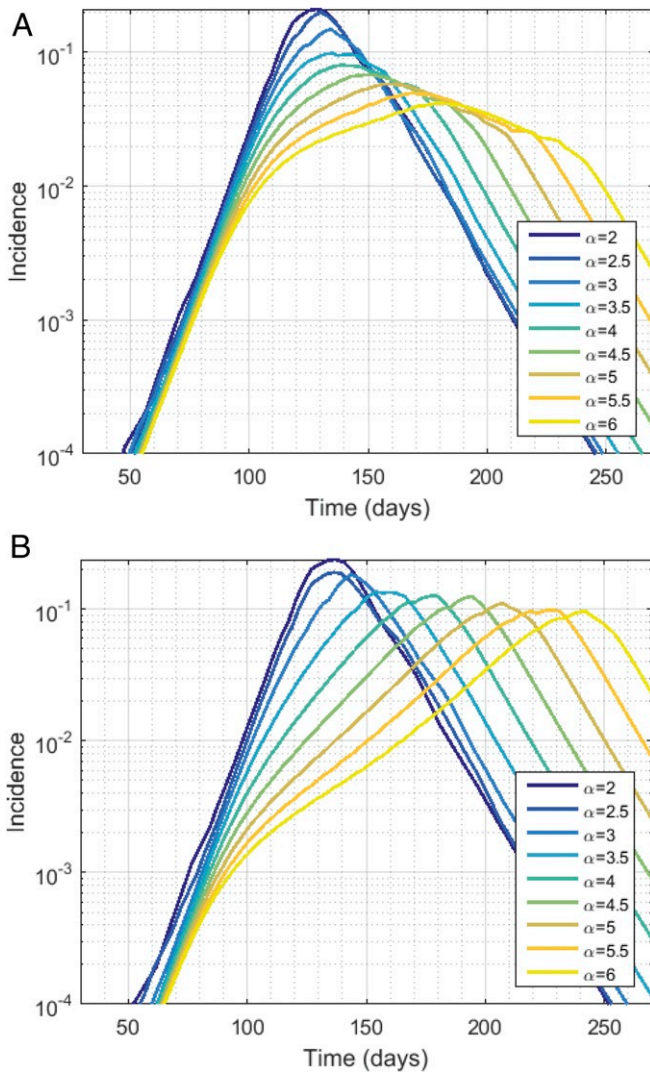


Fig. 5. Mean-field approximation with $R_0 = 2.2$, $\langle k \rangle = 10$, $h = 4$, using a 100×100 grid of uniformly spaced households. (A) Seeding in 10 randomly selected households (the same households are used in each simulation). (B) Seeding in the center only. Incidence is given as a proportion of the total population for α ranging from two to six. *SI Appendix, Fig. S10* shows time of peak incidence in the case $\alpha = 6$ seeded as above.

characteristics of epidemics, such as peak timing and attack rate via increased fourth-order clustering. We use the term first-order clustering for the quantity typically described as the local clustering coefficient (5): the link density of the immediate neighborhood of a given node. By extension, we defined order- m clustering coefficient to be the expected proportion of neighbors within m steps on the network who were also neighbors of each other within m steps (Fig. 3). We found no relationship between our assumed pattern of movement (α) and first- or second-order clustering coefficients. There was a weak relationship between α and third-order clustering and then a very strong relationship between α and fourth-order clustering. Patterns between epidemic properties and fourth-order clustering for individuals were similar to those between epidemic properties and second-order clustering of households, as would be expected, given the bipartite algorithm used to create individual-level networks.

Final size increased with spatial correlation, despite peak size displaying the opposite trend for controlled R^* or R_0 . There was a strong linear relationship between order- m clustering and peak

size/final size that could be explained by α , the strength of spatial embedding, when we control for R_0 (Fig. 4B). The gradient of the relationship decreased with order of clustering. Second-order household clustering showed the same relationship with peak size as did fourth-order individual clustering (Fig. 4C). These strong linear relationships only existed when we effectively controlled for R_0 , rather than R^* , and became less noisy when we reduced the interval used to define R_0 .

We conducted a number of sensitivity analyses for these network-simulation results. Analytic approximations for degree distribution $P(K = k)$ and expected first-order clustering $\langle CC^1 \rangle$ in our networks are given in *SI Appendix, Protocol S1* and are independent of α . We confirmed these relationships in *SI Appendix, Fig. S6* by computing these quantities on a set of networks that differ in α . *SI Appendix, Fig. S7* shows the relationship between α and clustering order 1 to 4 on networks generated by using a uniform population density. *SI Appendix, Fig. S8* shows the relationship between order- m clustering CC^m and peak size for different values of R_0 . *SI Appendix, Fig. S9* shows clustering orders 1 to 4 on networks with different h , w and p_w , and *SI Appendix, Fig. S10* provides an illustration of the relationship between higher-order clustering and rewiring probability on a commonly used network model with spatial embedding: the Watts–Strogatz Small World Network (5).

Finally, we mapped our network model onto a deterministic metapopulation framework so as to relate our simulations of incidence to prior analytic approximations of traveling spatial waves (see *SI Appendix, Protocol S1* for analytic construction). Fig. 5 shows the results of simulating on a grid of evenly spaced households of size $h = 4$, where a single continuous variable describes prevalence in each household, and spatial coupling between households used in the force of infection is exactly the kernel used in the construction of our spatially embedded networks. We simulate with randomly spaced seeds (as above) and with a central seed (the center-most four households), tracking global incidence and local time of peak incidence. The former case yielded global incidence curves similar to those generated in our network model (which was seeded similarly). The latter case allowed us to identify four distinct stages in the propagation of spatial waves that contribute to observed subexponential outbreak dynamics in more complex, network-based systems. *SI Appendix, Fig. S11* shows local peak timing in each case, and *SI Appendix, Fig. S12* shows simulation results in one spatial dimension with $\alpha = 6$ and $\alpha = 12$, alongside statistical properties of prevalence, which further clarify these growth phases (cf. figure legends for details and *SI Appendix, Protocol S1* for mathematical analysis).

Discussion

We have shown that nonstandard epidemic dynamics can arise from strongly spatially embedded social networks. Using a flexible algorithm of assigning individuals to households and then creating social networks with widely varying topologies, we can explain the absence of exponential growth and increased attack rate for a given peak height in terms of higher-order social structure, while maintaining a standard low-variance offspring distribution. We observed consistent patterns when we controlled for the basic reproductive number, as measured as directly as possible from a constant ratio of incidence between generations.

The algorithm we used (10) captures the key social contexts of home and workplace while using few parameters, which has allowed us to isolate specific relationships within the epidemic dynamics, across a broad range of network topologies. However, its simplicity is a potential limitation. Specifically, an individual only belongs to a single workplace (which may represent a school or social club). In reality, people will gather nonhousehold contacts from a variety of sources. Also, our networks are

not dynamic, which may limit the generalizability of the results to short-generation-time pathogens.

Accurate empirical data about higher-order social contacts would allow us to address some of these issues. There are a number of different approaches to gathering social-contact data, including contact diaries, mobile phone applications, and tag-based location tracking (27). Diary methods and current analytical approaches can provide accurate estimates of first-order moments [degree distribution (28)] and valuable insights into second-order moments [clustering (29)]. However, these data and current analytical approaches are limited for the estimation of higher-order moments. It seems likely that either high-resolution mobile-phone location data (30) or very-high-coverage tag-based studies will be needed to reveal these patterns (31). In addition, further work is needed on the use of algorithms similar to that used here to explicitly fit fully enumerated social networks to egocentric sample data from a subset of the population (or low-coverage nonegocentric data) (32).

Our results can be compared with other disease-dynamic models that produce nonstandard incidence profiles. Different functional forms have been suggested for the force-of-infection term in compartmental models that give polynomial growth in the early stages of an epidemic (20, 33). However, the key features of these model structures may be captured by a more straightforward underlying process (34). Faster-than-exponential growth can be achieved with very-high-variance offspring distributions, which have been inferred by diary studies of social contacts (9). There is also an extensive literature of much more abstract grid-based models of infectious disease that produce nonstandard epidemic dynamic because of very local spatial processes [cellular automata (35)]. We note that short periods of super-exponential growth were observed in our results for the simplified two-dimensional metapopulation example (Fig. 5B), arising from accelerating spatial waves of incidence, not driven by the variance of the offspring distribution.

Prospective forecasting of infectious-disease incidence during outbreaks (36) and seasonal epidemics (37) is an active area of public health research. Although nonmechanistic (38) and simple compartmental models (39, 40) have proven most reliable up to now, modern computing capacity enables studies to explore the possibility that incidence forecasts can be improved by the incorporation of realistic social-network topology (41, 42). For example, incidence of Ebola in West Africa in 2013 to 2016 and currently in Central Africa exhibits strong spatial clustering and highly nonstandard incidence dynamic, with short periods of exponential growth followed by low sustained peaks in incidence (43). Future forecasting studies should explore the possibility that

that sparse population density and short distances between contacts result in higher-order clustering in the social networks and the resulting nonstandard incidence profiles.

Materials and Methods

The Model. We simulated 10 independent epidemics for each of 200 parameter sets (h, v, ρ_w, R^*) drawn from a Latin hypercube, each seeded in 10 randomly selected individuals, and for each $\alpha = 0, 3, 6$. The ranges of values used in the Latin hypercube are given in *SI Appendix, Table S1*, and complete parameter sets for all networks are given in *SI Appendix, Table S1*. Our simulations allowed us to track disease incidence and disease generation of each infection.

We simulated an epidemic on the network to reflect the natural history of Ebola, with a latent period of 9.7 d and a serial interval of 15.3 d. The generation time was calibrated by varying the relative infectiousness of a short period before the onset of symptoms. Global transmissibility β was tuned to the value of R^* drawn from the Latin hypercube. For each time step, the probability of infection was calculated for each edge in the network. The algorithm progresses in real time with small time steps, so that it can be compared with results from compartmental models. Details of the network-simulation algorithm are given in ref. 10.

Assigning R_0 to each Simulation. For each simulation output, we calculated the mean reproductive ratio for each generation. For generations one to nine and for each possible consecutive string of three, four, or five values, we performed a linear regression fit. We defined R_0 as the mean reproductive ratio over the set of values for which the gradient of this fit was closest to zero (and all values that remained larger than one). This allowed us to assign a value R_0 to every simulation output.

Higher-Order Clustering. We computed our higher-order clustering coefficients on a subset of 1,000 nodes in each network, chosen at random. The algorithm involved storing the network structure as lists of neighbors for each node and performing an effective contact-tracing procedure. Though it is possible to compute these metrics for all nodes via successive multiplication of adjacency matrices, this procedure becomes computationally expensive in higher orders as networks become large.

Data Availability. The code to produce networks, the networks themselves, and code to analyze the networks have been deposited at Zenodo, <https://doi.org/10.5281/zenodo.3999974> (44), with more recent versions of the code available at GitHub, <https://github.com/c97sr/id.spatial.sim>.

ACKNOWLEDGMENTS. We thank Derek Cummings for useful discussions. This work was supported by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, also part of the EDCTP2 program supported by the European Union (Center MR/R015600/1) (to D.J.H. and S.R.); Wellcome Trust Investigator Award 200861/Z/16/Z (to S.R.); Wellcome Trust Collaborator Award 200187/Z/15/Z (to S.R.); National Institute for General Medical Sciences Grant MIDAS U01 GM110721-01 (to D.J.H. and S.R.); the National Institute for Health Research (UK, for Health Protection Research Unit funding) (S.R.); and the Centers for Disease Control and Prevention (D.J.H.).

- N. C. Grassly, C. Fraser, Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* **6**, 477–487 (2008).
- WHO Ebola Response Team *et al.*, Ebola virus disease among children in West Africa. *N. Engl. J. Med.* **372**, 1274–1277 (2015).
- K. O. Kwok *et al.*, Social contacts and the locations in which they occur as risk factors for influenza infection. *Proc. R. Soc. B Biol. Sci.* **281**, 20140709 (2014).
- S. P. Blythe, C. Castillo-Chavez, J. S. Palmer, M. Cheng, Toward a unified theory of sexual mixing and pair formation. *Math. Biosci.* **107**, 379–405 (1991).
- D. J. Watts, S. H. Strogatz, Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
- M. J. Keeling, The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. London Ser. B Biol. Sci.* **266**, 859–867 (1999).
- J. C. Miller, Bounding the size and probability of epidemics on networks. *J. Appl. Probab.* **45**, 498–512 (2008).
- S. Eubank *et al.*, Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
- Q. H. Liu *et al.*, Measurability of the epidemic reproduction number in data-driven contact networks. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12680–12685 (2018).
- S. Riley, N. M. Ferguson, Smallpox transmission and control: Spatial dynamics in Great Britain. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12637–12642 (2006).
- R. M. Anderson, R. M. May, B. Anderson, *Infectious Diseases of Humans: Dynamics and Control* (Wiley Online Library, New York, NY, 1992), Vol. 28.
- H. W. Hethcote, The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
- H. Nishiura, G. Chowell, M. Safan, C. Castillo-Chavez, Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theor. Biol. Med. Model.* **7**, 1 (2010).
- J. T. Wu, S. Riley, C. Fraser, G. M. Leung, Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med.* **3**, e361 (2006).
- M. Jesse, P. Ezanno, S. Davis, J. A. P. Heesterbeek, A fully coupled, mechanistic model for infectious disease dynamics in a metapopulation: Movement and epidemic duration. *J. Theor. Biol.* **254**, 331–338 (2008).
- D. J. Haw *et al.*, Differential mobility and local variation in infection attack rate. *PLoS Comput. Biol.* **22**, e1006600 (2019).
- J. Griffiths, D. Lowrie, J. Williams, Age-structured model for the AIDS epidemic. *Eur. J. Oper. Res.* **124**, 1–14 (2000).
- M. C. Boily, C. Lowndes, M. Alary, The impact of HIV epidemic phases on the effectiveness of core group interventions: Insights from mathematical models. *Sex. Transm. Infect.* **78**, i78–i90 (2002).
- S. Riley, K. Eames, V. Isham, D. Mollison, P. Trapman, Five challenges for spatial epidemic models. *Epidemics* **10**, 68–71 (2015).
- J. A. P. Heesterbeek, A brief history of R_0 and a recipe for its calculation. *Acta Biotheor.* **50**, 189–204 (2001).
- R. M. May, A. L. Lloyd, Infection dynamics on scale-free networks. *Phys. Rev. E* **64**, 066112 (2001).
- L. A. Meyers, Contact network epidemiology: Bond percolation applied to infectious disease prediction and control **44**, 63–86 (2007).

23. P. Trapman, On analytical approaches to epidemics on networks. *Theor. Popul. Biol.* **71**, 160–173 (2007).
24. P. Expert, T. S. Evans, V. D. Blondel, R. Lambiotte, Uncovering space-independent communities in spatial networks. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7663–7668 (2011).
25. J. Truscott, N. M. Ferguson, Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput. Biol.* **8**, e1002699 (2012).
26. T. Garske *et al.*, Travel patterns in China. *PLoS One* **6**, e16364 (2011).
27. J. M. Read, K. T. D. Eames, W. J. Edmunds, Dynamic social networks and the implications for the spread of infectious disease. *J. R. Soc. Interface* **5**, 1001–1007 (2008).
28. J. M. Read *et al.*, Social mixing patterns in rural and urban areas of southern China. *Proc. R. Soc. B Biol. Sci.* **281**, 20140268 (2014).
29. X. Xiao, A. J. van Hoek, M. G. Kenward, A. Melegaro, M. Jit, Clustering of contacts relevant to the spread of infectious disease. *Epidemics* **17**, 1–9 (2016).
30. E. Cho, S. A. Myers, J. Leskovec, “Friendship and mobility: User movement in location-based social networks” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11* (Association for Computing Machinery, New York, NY, 2011), pp. 1082–1090.
31. C. Cattuto *et al.*, Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One* **5**, e11596 (2010).
32. P. N. Krivitsky, M. Morris, Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Ann. Appl. Stat.* **11**, 427–455 (2017).
33. G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, Mathematical models to characterize early epidemic growth: A review. *Phys. Life Rev.* **18**, 66–97 (2016).
34. T. House, A general theory of early growth?: Comment on: “Mathematical models to characterize early epidemic growth: A review” by Gerardo Chowell *et al.* *Phys. Life Rev.* **18**, 109–111 (2016).
35. S. H. White, A. M. del Rey, G. R. Sánchez, Modeling epidemics using cellular automata. *Appl. Math. Comput.* **186**, 193–202 (2007).
36. M. I. Nelson *et al.*, Fogarty International Center collaborative networks in infectious disease modeling: Lessons learnt in research and capacity building. *Epidemics* **26**, 116–127 (2018).
37. M. Biggerstaff *et al.*, Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics* **24**, 26–33 (2018).
38. D. C. Farrow *et al.*, A human judgment approach to epidemiological forecasting. *PLoS Comput. Biol.* **13**, e1005248 (2017).
39. N. G. Reich *et al.*, A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3146–3154 (2019).
40. C. Viboud *et al.*, The RAPIDD Ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics* **22**, 13–21 (2017).
41. M. Tizzoni *et al.*, Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm. *BMC Med.* **10**, 165 (2012).
42. J. M. Drake *et al.*, Ebola cases and health system demand in Liberia. *PLoS Biol.* **13**, e1002056 (2015).
43. Ebola Outbreak Epidemiology Team, Outbreak of Ebola virus disease in the democratic republic of the Congo, April–May, 2018: An epidemiological study. *Lancet* **392**, 213–221 (2018).
44. D. J. Haw, R. Pung, J. M. Read, S. Riley, id_spatial_sim. Zenodo. <https://doi.org/10.5281/zenodo.3999974>. Deposited 25 August 2020.