

Research



Cite this article: Shridhar SV, Alexander M, Christakis NA. 2021 Characterizing super-spreaders using population-level weighted social networks in rural communities. *Phil. Trans. R. Soc. A* **380**: 20210123.
<https://doi.org/10.1098/rsta.2021.0123>

Received: 25 May 2021

Accepted: 7 July 2021

One contribution of 14 to a theme issue ‘Data science approach to infectious disease surveillance’.

Subject Areas:

biomedical engineering, computational biology, mathematical modelling

Keywords:

social networks, contagion, epidemic, agent-based network model

Author for correspondence:

Nicholas A. Christakis

e-mail: nicholas.christakis@yale.edu

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5673909>.

Characterizing super-spreaders using population-level weighted social networks in rural communities

Shivkumar Vishnempet Shridhar^{1,2},

Marcus Alexander² and Nicholas A. Christakis²

¹School of Engineering and Applied Science, and ²Yale Institute for Network Science, Yale University, 17 Hillhouse Ave, New Haven, CT 06520, USA

SVS, 0000-0001-6094-3041; NAC, 0000-0001-5547-1086

Sociocentric network maps of entire populations, when combined with data on the nature of constituent dyadic relationships, offer the dual promise of advancing understanding of the relevance of networks for disease transmission and of improving epidemic forecasts. Here, using detailed sociocentric data collected over 4 years in a population of 24 702 people in 176 villages in Honduras, along with diarrhoeal and respiratory disease prevalence, we create a social-network-powered transmission model and identify super-spreading nodes as well as the nodes most vulnerable to infection, using agent-based Monte Carlo network simulations. We predict the extent of outbreaks for communicable diseases based on detailed social interaction patterns. Evidence from three waves of population-level surveys of diarrhoeal and respiratory illness indicates a meaningful positive correlation with the computed super-spreading capability and relative vulnerability of individual nodes. Previous research has identified super-spreaders through retrospective contact tracing or simulated networks. By contrast, our simulations predict that a node’s super-spreading capability and its vulnerability in real communities are significantly affected by their connections, the nature of the interaction across these connections, individual characteristics (e.g. age and sex) that affect a person’s ability to disperse a pathogen, and also the intrinsic characteristics of the pathogen (e.g. infectious period and latency).

1. Introduction

Previous studies of biological and sociological features of human social interactions—including the evolutionary biology and genomics of social networks, their physiological implications and their possibly ancient heritage—suggest that natural selection has shaped social network structure and function [1–4]. One possible function of social networks relates to infectious diseases. Traditionally, the spread of infection in human communities has been analysed using compartmental epidemiological models like SIR/SEIR models [5]. However, these models generally assume uniform and fully mixed populations, which often lead to incorrect estimates of predicted infection counts and R_0 (basic reproduction number) [6,7].

Therefore, we need approaches for characterizing possible heterogeneity in transmission across individuals, which, in particular, is defined based on a host's social interactions (quantity and quality of interactions); their intrinsic ability to disperse a pathogen (e.g. based on their age and sex); and the transmission characteristics of the pathogens themselves (e.g. infectious periods and probability of transmission).

A previous study has reported a very well-illustrated characterization of heterogeneity in transmission based on variation across individuals in their ability to spread a pathogen [8]. And specific evidence from SARS outbreaks in 2003 (and for other outbreaks) has shown higher levels of super-spreading originating in older than younger individuals [9–11]. Other research has revealed how pathogens affect transmission, documenting that transmission is an essential combination of pathogen infectious period and its overlap with host social interactions [12–14].

In order to characterize the foregoing heterogeneity in transmission and accurately predict the trajectory of potential outbreaks, we collected sociocentric data from real networks. In this work, we ascertained social networks among 24 702 villagers in 176 villages in Honduras. In particular, this also includes the 'weight' of a tie (e.g. as measured by the frequency of interaction as well as the form in which pairs of individuals greet each other, signifying the type of physical contact). Therefore, we were able to synthesize a social-network-based transmission model where we implement a bottom-up approach by constructing the effective probability of passing on a pathogen from an ego to their alters. We perform Monte Carlo simulations to better predict a node's likelihood of passing on the infection, on the one hand, and of being infected, on the other hand, as a function of the node's individual and network attributes, as well as the node's pathogen-coupled attributes.

We chose a commonly studied and important infectious disease (diarrhoea) to evaluate our model. In Honduras, where our study of social networks is based, currently 2% of the deaths, i.e. 737 deaths per 100 000 population, can be attributed to diarrhoea-related causes [15]. Epidemic investigations in Ghana with a similar socioeconomic background as in Honduras have reported high $R_0 = 2.09$ in several diarrhoeal outbreaks [16–18]. Furthermore, we also used influenza A as an example. Prior empirical research has given us sufficient insight to characterize features of these pathogens that are likely to influence how the disease spreads through social interactions [12]. Our results show that network topology, quality of interaction and pathogen-specific dispersion play a crucial role in determining how vulnerable or super-spreading a node is.

2. Material and methods

We assessed social networks in a population of 24 702 people across 176 villages in the Western highlands of Honduras, as a part of a network-targeting public health intervention [19]. The total number of individuals who consented to participate and provided detailed social network, demographic, socio-economic and health data was 22 512 individuals in the most recent wave

3 (2019), 21 485 villagers in wave 2 (2018) and 24 702 villagers in wave 1 (2016). We performed sociocentric mapping of the study population using the photographic network census mobile app Trellis¹ (which we developed) to collect various ego-alter connections (using diverse name generators and villagers' photographs to verify the identity of social contacts) [19–22]. Using the Trellis platform, we also recorded the quality of interactions. We asked our respondents, 'Who do you spend free time with?' This was followed by 'How do you greet each other?' (characterizing levels of contact ranging from a smile, a bow/nod/wave, a verbal salute, a hand-shake/high-five, a pat on the back, a hug, or a kiss on the cheek). The respondents were also asked 'In the last month, how often did you see each other?' thus quantifying the frequency of their interaction. Therefore, these questions defining the levels of contact (salutations) and frequencies of contact add a qualitative dimension to the social network. Finally, we also used Trellis to collect information regarding subjects' history of diarrhoea and cough (respiratory related-infection).

We developed an agent-based model combining host, pathogen and (observed) social network characteristics. Each agent engages in social interaction with their alters, defined qualitatively (nature and frequency of contact) and quantitatively (number of connections). We then introduce the pathogen to these networks, which, with its own characteristics—such as infectious period, incubation time and transmission probability—and host-specific characteristics (related to intrinsic host ability to disperse a pathogen) transform these social edges into temporal probabilistic paths for disease transmission. Using temporally dynamic Monte Carlo simulations, we observe the epidemic evolution of the disease over a period of 100 days in 1-day intervals.

(a) Model parameters

The characteristic equation ($\rho(t)$, electronic supplementary material, table S1–S6 and highlighted as green in figure 1) captures the transmissibility of the pathogen during the infectious period [23], which for shigellosis (as a cause of diarrhoea) was 7 days, and, on each day, the probability of transmission varies depending on incubation time, symptom progression, and recovery during that time frame. For the disease transmission model, diarrhoea (e.g. due to shigellosis) was chosen as a model, due to its prominence in the developing world [16–18,24]. To decouple the effects of the dispersion factor of the pathogen, we also investigate the transmission of influenza A through our network. In contrast to shigella, influenza A has an infectious period of 15 days [25].

The next vital component is a set of social network attributes (factors highlighted as purple in figure 1). The number of connections a node has (degree) determines an equivalent number of possible routes a pathogen can transmit through (to and from). Based on the qualitative attributes of the social interaction (salutations), connections exhibiting a hand-shake/high-five, a pat on the back, a hug, or a kiss on the cheek were considered to be riskier behaviour, and therefore to lead to an increase in probability (p_s); the other salutations were considered as relatively safe. For frequency of contact, everyday contact was considered to have a probability of 1, and lesser frequencies were considered as an equivalent of their fractions (p_f).

To fully characterize the transmission model, we follow previous research [8] that has pointed out that airborne pathogen transmission is also affected by biological characteristics of the individuals spreading a pathogen (electronic supplementary material, tables S3–S6 and factors highlighted in red in figure 1). An individual can spread the same airborne pathogens differently depending on their breathing patterns, quantity of aerosols during exhalation, and host–pathogen interactions, which might affect viral shedding [8]. Thus, these factors may be allowed to vary with a person's age (p_a) [8,10] and sex (p_g) [11]. For instance, naturally, this factor related to an individual's propensity to disperse a pathogen is different for different pathogens [8]. For our influenza transmission model, the dispersion would account for all the airborne-related transmission factors (as above). Additionally, for diarrhoea, there is also evidence suggesting that individuals with different ages and sexes show varying severity in infection and also transmission [10,11]. Hence, to account for such factors in our model, without loss of generality, we also

¹For more information on Trellis, see <https://trellis.yale.edu>.

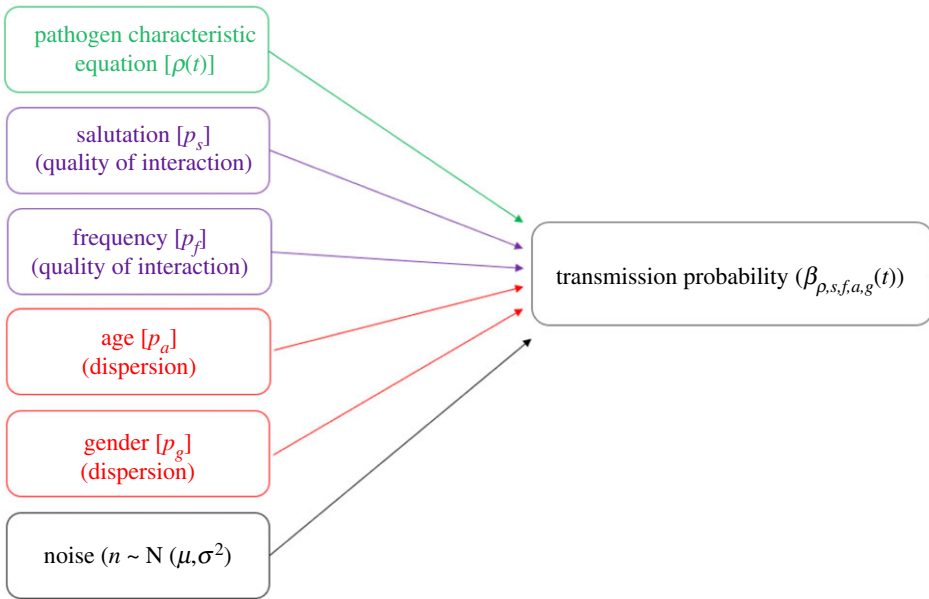


Figure 1. Schematic showing the factors going into the transmissibility or an individual node's effective transmission probability $\beta_{\rho,s,f,a,g}(t)$. The pathogen-characteristic equation (highlighted in green) is the pathogen's day-to-day transmission probability, decoupled from the human element in transmission. Salutations and frequency probabilities (highlighted in purple) are the qualitative sociocentric aspects of a node's transmission. Dispersion probabilities (highlighted in red) are the egocentric aspect of a node's propensity to transmit a pathogen. Noise (highlighted in black) was also included to account for variations in the node's interaction pattern. (Online version in colour.)

consider the age and gender of the individual transmitting the pathogen. All above mentioned model parameters can be found in electronic supplementary material, tables S1–S6.

With respect to the quality of interactions, we assume there is an inherent uncertainty in the frequency of interactions in the social network. For example, a person interacting with their alters once a week may vary this frequency in some weeks. Therefore, to account for this uncertainty, we introduce a small multiplicative Gaussian noise term, with $N(\mu, \sigma)$, where $\mu = 1$, $\sigma = 0.01$, so that there is a small probability that the person may interact more or less frequently than their reported interaction.

(b) Model formulation

An agent-based network model is implemented using Monte Carlo simulations for pathogen transmission. As described in figure 2, initially, the model begins by seeding a node with an infection. In the second step, the node's age and sex determine its dispersion factor (in the sense of the likelihood of conveying the pathogen to others). The degree, or the number of connections the node has, determines how many probabilistic transmission routes the pathogen can take. For each connection, the unique quality of interaction, i.e. the salutation and frequency of this edge, defines its corresponding probabilities. The characteristic equation of the disease, which varies from day to day, along with the noise, ultimately results in a combined transmission probability of $\beta_{\rho,s,f,a,g}(t)$, as shown in equation (2.1).

$$\beta_{\rho,s,f,a,g}(t) = \rho(t) \cdot p_s \cdot p_f \cdot p_a \cdot p_g \cdot n \quad (2.1)$$

The connected nodes have a random chance of getting infected with a probability $\beta_{\rho,s,f,a,g}(t)$. The $\beta_{\rho,s,f,a,g}(t)$ varies from day-to-day and also from edge-to-edge depending on the quality of interaction for the same transmitting node. Similarly, the neighbouring nodes also have a

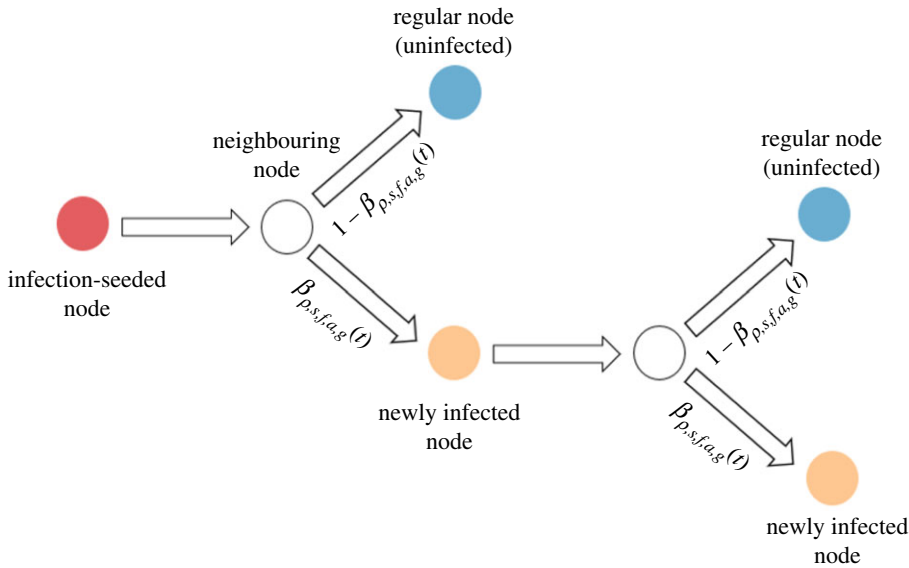


Figure 2. Diagrammatic representation of the agent-based model, similar to a chain reaction. An initially infection-seeded node would likely infect all the neighbouring nodes depending on its combined probability $\beta_{\rho,s,f,a,g}(t)$ unique to it and the interaction it has. A neighbouring node, if infected, would similarly give rise to a newer infection depending on it and its interaction. (Online version in colour.)

probability of not getting infected equivalent to $1 - \beta_{\rho,s,f,a,g}(t)$. If the neighbouring node is infected, the newly infected node will have a new $\beta_{\rho,s,f,a,g}(t)$, depending on its edges dispersion and the day. Upon infection, every infected node (including the seeded node) can only transmit infection for a period equivalent to the infectious period of the disease (7 days for shigella and 15 days for influenza A). Thus, this transmission process is finite, leading to the ultimate state of the network.

A single Monte Carlo simulation has 100 steps, each representing 1 day, netting a total of 100 days of transmission in a single simulation. This simulation is repeated 10 000 times for every infection-seeded node (red coloured node). Every node in the village also takes a turn in becoming an infection-seeded node. Therefore, for Hacienda San Juan, for example, which has 58 nodes, a total of 580 000 simulations were performed. Overall, for all 176 villages, 2 470 200 000 simulations were performed to characterize all possible transmission scenarios for shigellosis (diarrhoea). Furthermore, in addition to diarrhoeal system $[\rho_{\text{diarrhoea}}(t)]$, simulations were also repeated for influenza A transmission $[\rho_{\text{influenza}}(t)]$. Finally, we also repeated diarrhoeal simulations excluding the effect of quality of interaction $[\rho_{\text{uniform}}(t)]$.

Figure 3a shows one possible final state, where there were 12 new infections (yellow nodes) arising from the initial infection (red node) at the end of the 100-day Monte Carlo simulation. Figure 3b shows four possible final states, as an illustration, by repeating the simulation four times. The variation in the final states can be explained by the transmission probability $\beta_{\rho,s,f,a,g}(t)$ giving rise to differing numbers of infection-transmitting agents in each simulation.

(c) Outputs measured

The ultimate goal of these transmission simulations is to measure every node's vulnerability and also super-spreading capability. The vulnerability is measured by determining how many times a node transforms into a newly infected node (yellow node). The super-spreading capability is measured by the average number of new infections arising from the infection-seeded node in all of the 10 000 simulations.

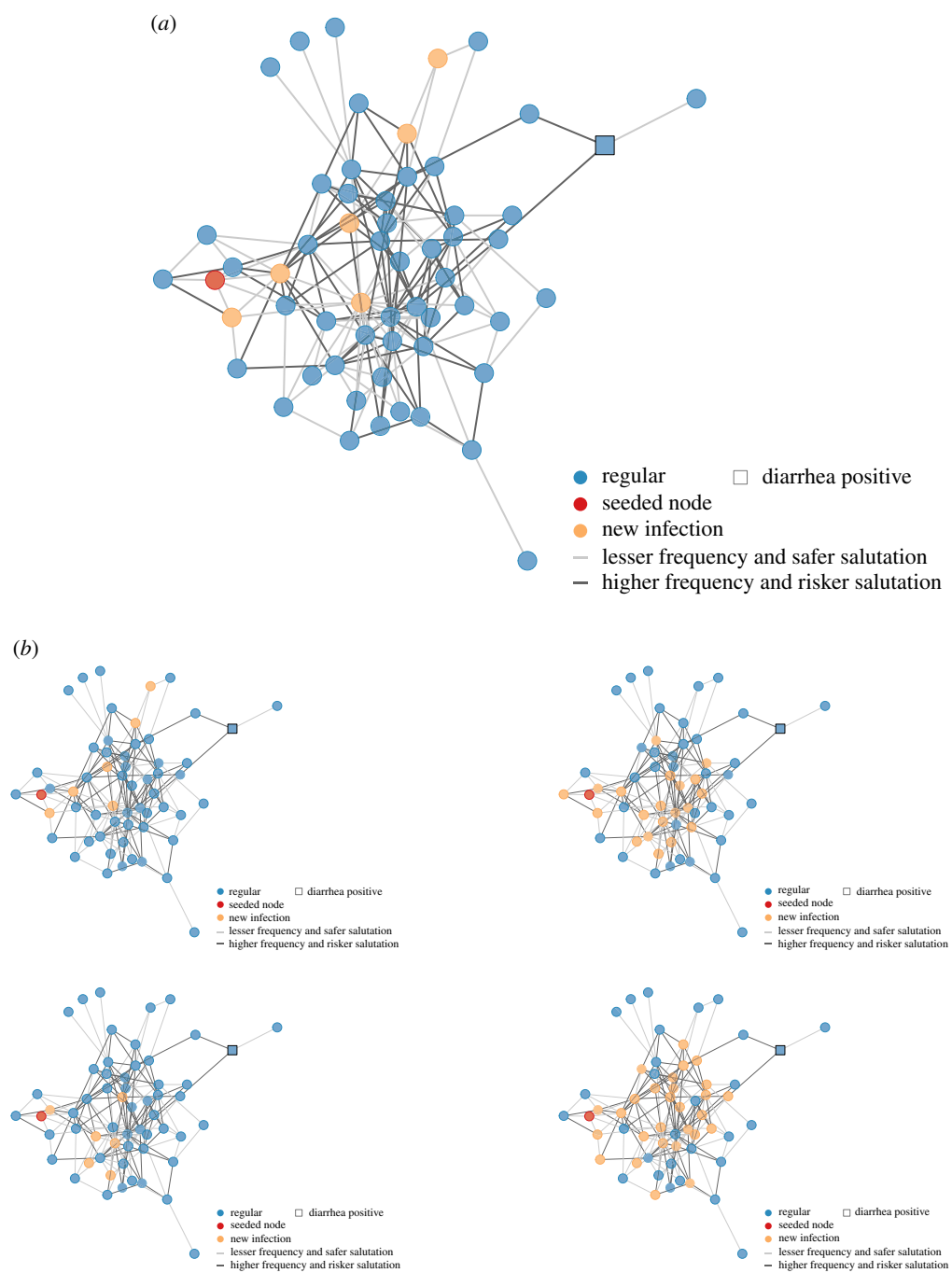


Figure 3. An example of the final state of a diarrhoeal simulation is shown in (a) at the end of the 100-day simulation period. After initially seeding the infection with (node red), based on agent-based transmission with transmissibility $\beta_{\rho,s,f,a,g}(t)$, this leads to newer infections. Repeating the simulation with the same infection-seeded node leads to very different final states as shown in (b), with different nodes being infected. (Online version in colour.)

The node's vulnerability will be shown as a relative measure, compared to the rest of the village in quintiles (relative vulnerability, RV). The node's super-spreading capability will be shown as relative measure in quintiles (relative super-spreading capability, RSS), and also as an absolute measure as percentage of total transmission in the village (super-spreading index, SSI).

To understand the importance of the quality of social connections (i.e. salutations and the frequencies of interactions), we repeated the simulations without considering the quality of interactions, obtaining new transmission probability of $\beta_{\rho,a,g}(t)$.

(d) Real infections from the survey network

In our Honduran social network, we also collect information on whether the individual respondent has had diarrhoea in the last four weeks. This is represented in figure 3 (as square nodes). Simultaneously, we also record if the respondent has been coughing for at least two weeks at the time of survey. Using these, we later analyse their correlations with tendencies to infect (super-spreading capability) and be infected (vulnerability).

3. Results

(a) Role of interaction quality in the transmission model

Figure 4*a–c* shows the infection counts RV, RSS and SSI of diarrhoeal simulations that include the quality of interactions, whereas panels (*d–f*) shows RV, RSS, SSI of diarrhoeal simulations without the considering the quality of interaction.

From this comparison, it becomes evident that the two systems are significantly different in RV (figure 4*g*, $p = 0.0305$) and in RSS (figure 4*h*, $p = 0.0014$) in the single illustrative village. This difference can also be visualized by figure 4*i*, which shows the relative rank difference in the super-spreading capability between the two systems (most nodes have a shift in ranking, visualized as difference between 4*b* and 4*e*). For example, in this village, only two nodes in the entire village of 58 nodes retained their relative super-spreading rank after interaction quality was taken into account. The relative rank difference can also be visualized in figure 4*j*, which shows that most nodes change ranking when including the quality of interaction in the model, regardless of their topological position in the network graph.

Essentially, removing the quality of interactions can give rise to a completely different set of super-spreaders (nodes in the top 5th quintile of super-spreading capability) and completely different indices of highly vulnerable individuals. Thus, the quality of interaction plays an important role in predicting super-spreading capability and vulnerability (see below for the analysis across all 176 villages).

(b) Role of pathogen

To consider the role of the pathogen in the transmission of infection, we perform analogous analyses with the influenza transmission model. Figure 5*a–c* shows the infection counts RV, RSS and SSI of diarrhoeal simulations, whereas panels (*d–f*) show these quantities for the influenza simulations in this illustrative village.

From this comparison, the influenza and diarrhoeal systems are different in vulnerability ($p = 0.05$) but not in super-spreading capability ($p = 0.284$). However, the relative rank of super-spreading capability is completely different for the two pathogens, with no villager retaining their same relative rank upon changing the pathogen. Thus, pathogen characteristics play a vital role in predicting the super-spreading capability and vulnerability.

(c) Aggregate analysis on all 176 villages in diarrhoeal simulations

An aggregate correlation plot can be visualized of all the social network factors, such as salutations, frequency, eigenvector centrality and degree, with the individual's diarrhoeal infection, and the RV, RSS and the absolute SSI. For the diarrhoeal model across all 176 villages (figure 6; electronic supplementary material, figure S2), the predicted RV, which is a measure of the number of times a node shows up as a new infection, shows a strong positive correlation of

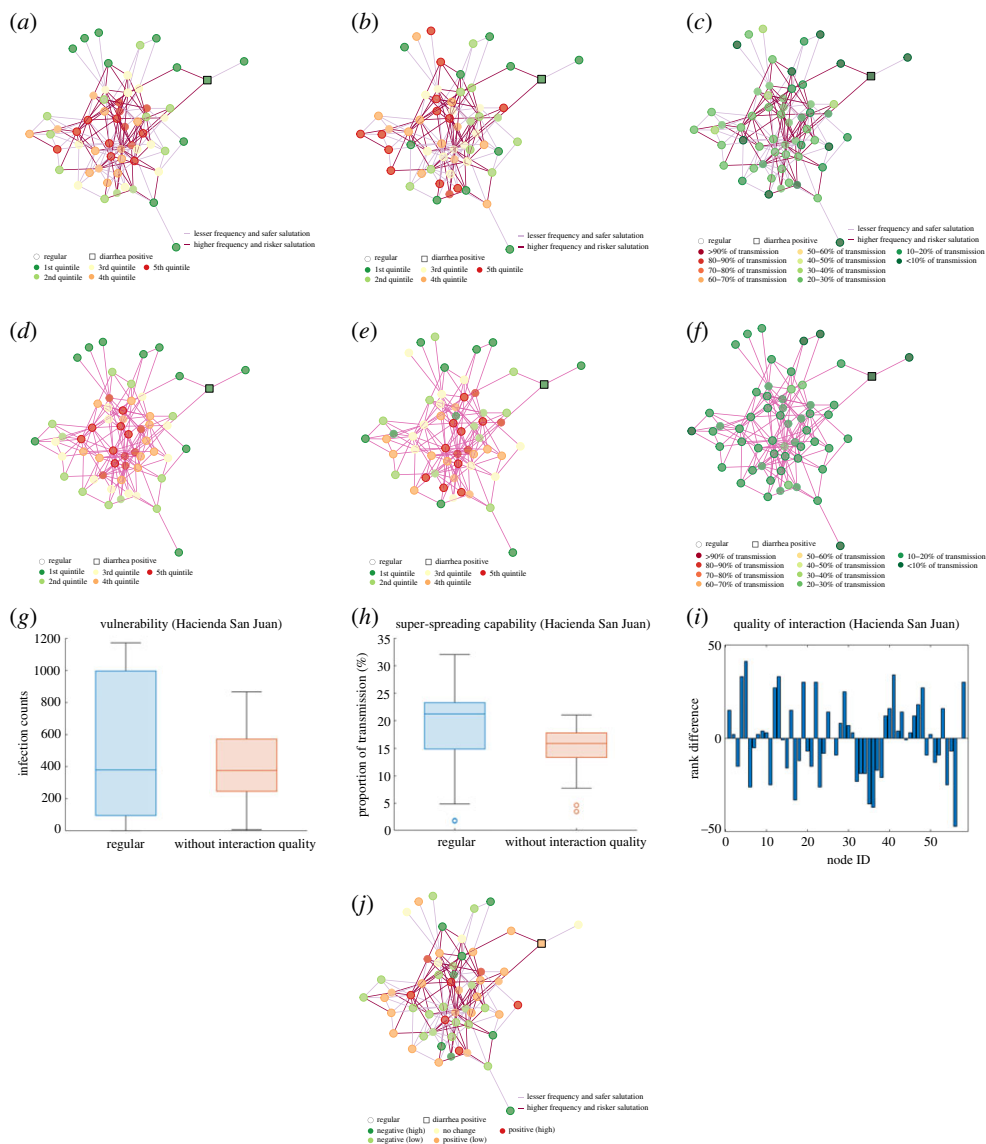


Figure 4. Comparison of diarrhoeal models with and without quality of interaction in terms of relative vulnerability (RV), relative super-spreading capability (RSS) and population-normalized super-spreading capability index (SSI) from all 10 000 simulations per node for the entire village of Hacienda San Juan. Panel (a–c) shows the RV, RSS and SSI for our regular social network, whereas panels (d–f) show RV, RSS and SSI without the quality of interaction (without salutations and frequency). Panels (g,h) shows the side-by-side differences in the RV and SSI, with and without the quality of interaction. Panel (i) is another way of visualizing this difference by looking at the relative ranking of the super-spreading capability, which can also be visualized at the network level in (j). (Online version in colour.)

0.30 (Pearson, $p = 6.13 \times 10^{-37}$) and 0.51 (Spearman, $p = 8.98 \times 10^{-64}$) with degree, and a positive correlation of 0.11 (Pearson, $p = 2.22 \times 10^{-18}$) with salutations, and a positive correlation of 0.17 (Pearson, $p = 1.15 \times 10^{-22}$) with the frequency of interaction. The RSS, which is a measure of average number of new infections caused by a node in all 10 000 simulations of the node (when seeded as an infected agent) reveals a positive correlation of 0.18 (Pearson, $p = 4.13 \times 10^{-25}$) and 0.30 (Spearman, $p = 1.52 \times 10^{-33}$) with degree, a positive correlation of 0.07 (Pearson, $p = 1.44 \times 10^{-4}$) with salutations, and a positive correlation of 0.12 (Pearson, $p = 2.12 \times 10^{-3}$)

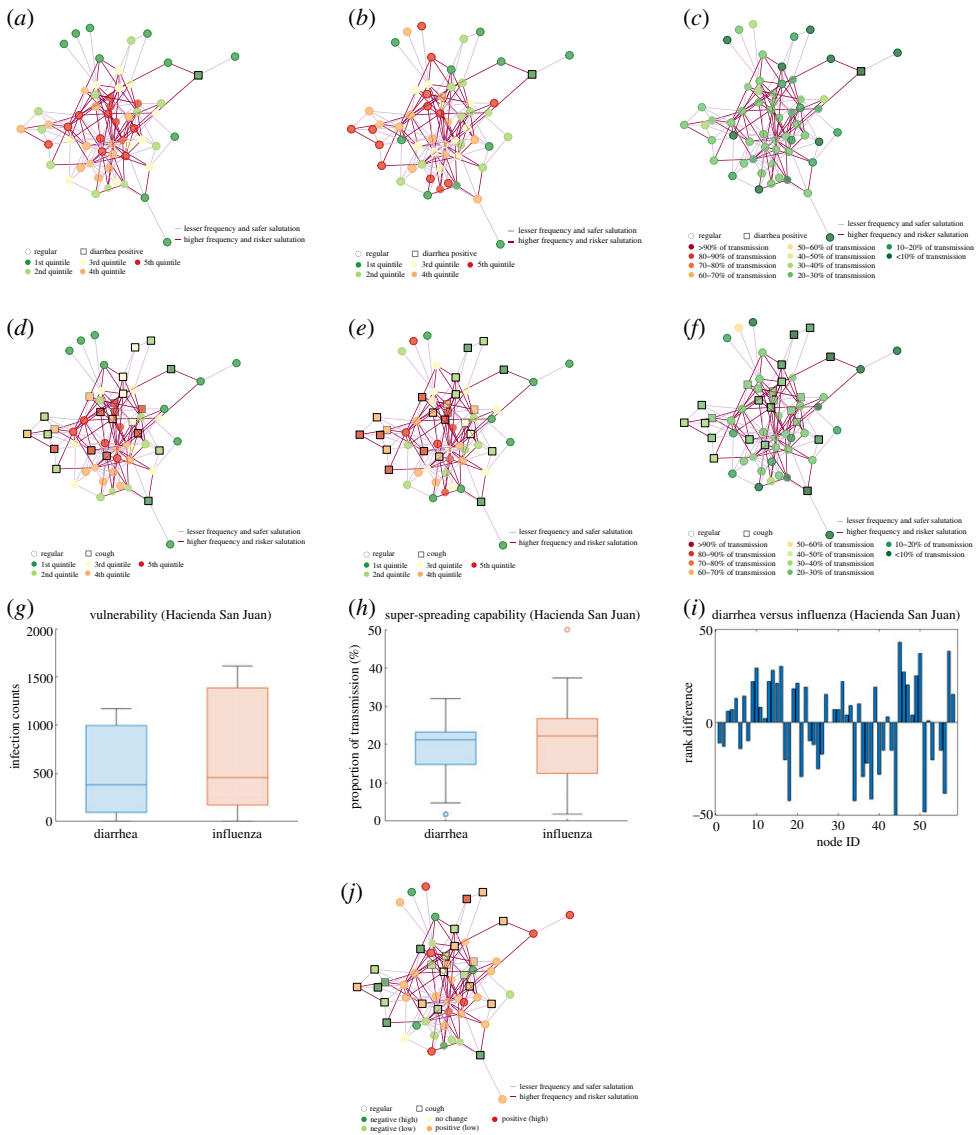


Figure 5. Comparison of diarrhoeal and influenza network in terms of RV, RSS and SSI for the village of Hacienda San Juan. Panels (a–c) shows the RV, RSS and SSI for our diarrhoeal network, whereas (d–f) show RV, RSS and SSI for the influenza network. (g) and (h) show the side-by-side differences in the RV and SSI for the influenza and the diarrhoeal network. Panel (i) is another way of visualizing this difference by looking at the relative ranking of the super-spreading capability with respect to their pathogen, which can also be visualized and the network level in (j). (Online version in colour.)

with frequency. The SSI, or the super-spreading index, which is the percent of village infected on average in all 10 000 simulations, is a village-population average metric, which shows a significant positive correlation of 0.25 (Pearson, $p = 4.44 \times 10^{-31}$) and 0.31 (Spearman, $p = 2.19 \times 10^{-39}$) with degree, a positive correlation of 0.10 (Pearson, $p = 4.84 \times 10^{-43}$) with salutation, and a positive correlation of 0.14 (Pearson, $p = 7.09 \times 10^{-19}$) with frequency of interaction. Moreover, eigenvector centrality shows a positive correlation of 0.14 with RV ($p = 8.117 \times 10^{-27}$), 0.06 with RSS ($p = 2.826 \times 10^{-13}$) and 0.322 with SSI ($p = 5.80 \times 10^{-19}$).

The number of diarrhoea-positive individuals (figure 6; electronic supplementary material, figures S3 and S4) in the preceding four weeks was 665 (2.69% of total respondents), 823

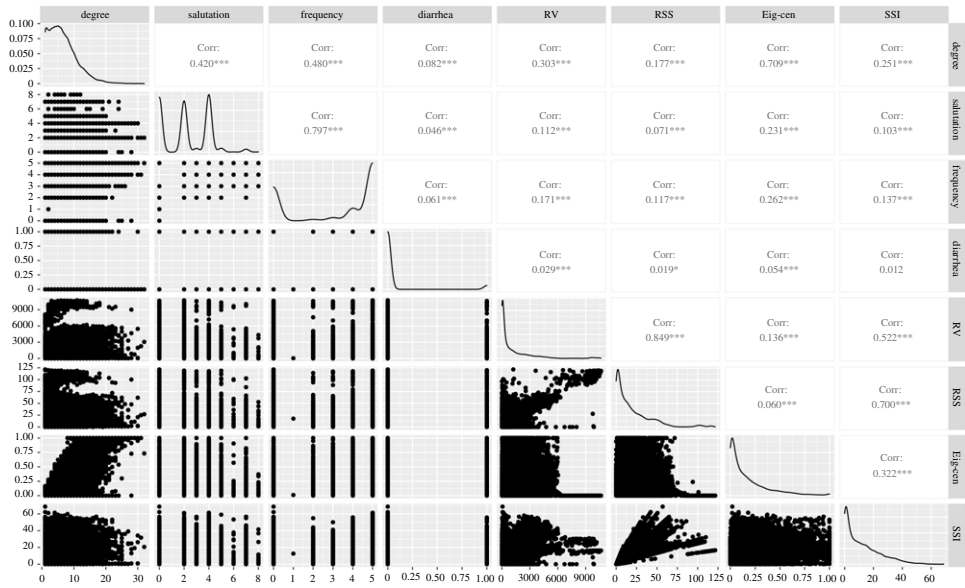


Figure 6. Combined correlation plot listing pairwise correlation between social network aspects, including: degree, salutations, frequency, and eigenvector-centrality; diarrhoea positive individuals; and the diarrhoeal-model predictions of every individual's RV, RSS and SSI.

(3.84%), 623(2.76%) in surveys waves 1, 2 and 3, respectively. Individuals with diarrhoeal infections showed a positive correlation of 0.08 (Pearson, $p = 5.81 \times 10^{-24}$) with degree, 0.05 (Pearson, $p = 2.69 \times 10^{-8}$) with salutations, 0.06 (Pearson, $p = 2.39 \times 10^{-13}$) with frequency, 0.03 (Pearson, $p = 5.3 \times 10^{-4}$) with RV, and 0.05 (Pearson, $p = 8.07 \times 10^{-11}$) with eigenvector centrality. Furthermore, the diarrhoeal-positive individuals also showed a strong dependence of $\chi^2 = 1627.4$ ($p = 2.2 \times 10^{-16}$) on salutations and $\chi^2 = 758.88$ ($p = 7.14 \times 10^{-11}$) on frequency of interaction. They also showed a positive correlation of 0.02 with RSS ($p = 0.021$). Thus, individuals with higher degree and/or those who engage in 'riskier' interactions (salutations and frequencies) are much more likely to be vulnerable to diarrhoeal infection, and slightly more likely to become super-spreaders of diarrhoeal infection.

By switching to the influenza model (electronic supplementary material, figure S1), where the only changed model parameters are the pathogen characteristic equation and infection probability with individuals' age and sex (related to their ability to disperse the pathogen), the individuals with persistent cough over two weeks showed a positive correlation of 0.12 with degree ($p = 5.77 \times 10^{-3}$), 0.12 with salutations ($p = 5.26 \times 10^{-3}$), 0.17 with frequency ($p = 7.93 \times 10^{-5}$). Moreover, they also showed a slightly less, but statistically significant positive correlation of 0.09 with RV ($p = 0.031$), 0.09 with RSS ($p = 0.047$), and 0.09 with eigenvector centrality ($p = 0.028$).

4. Discussion and conclusion

Location-tracking and phone-data surveillance or contact-tracing mechanisms are often considered as ground truth for any epidemic investigation [8,26–30]. But the nature of dyadic social interactions (figure 4), i.e. salutations and frequency, captures the quality of the relationship between nodes, thereby transforming regular social connections into weighted connections. Previous studies on weighted interactions in social networks have shown that weaker ties play an important role in preserving the local structure of the network [31]. Moreover, considering a social network with the quality of interaction has two significant advantages over location-tracking

or contact-tracing data. First, surveillance datasets have a fixed observation time-window, i.e. tracked data is recorded and analysed in a specific time frame [8,32]. In our model, the responding individual can assess their own quality of interaction which is typically not confined to the past few hours or weeks. Second, the quality of interaction can also be used in predicting future pathogen transmission; this cannot be as reliably forecast through location tracking or contact tracing or other conventional epidemiological investigation methods. Predicting individual transmission or infection risk without considering quality of interaction may thus lead to far less accurate predictions (as shown in figures 4 and 6).

When constructing a disease transmission model, dispersion is another key factor to be considered (figure 5). Dispersion can be defined here as an individual's transmitting capability, depending on their age, sex, breathing patterns, vocal activities, and host–pathogen interaction factors [8]. Hence, dispersion of a pathogen arises due to the unique interaction between the pathogen and the hosts' characteristics. As shown in figure 6 and electronic supplementary material, figure S1, switching the model from diarrhoea to influenza also significantly changes an individual's RV (which can also be understood as personal infection risk) and their RSS. Analogously, there are no unique super-spreaders for all pathogens. Rather, the characteristics of the pathogen also determine who is likely to be more vulnerable or more super-spreading within any population.

In sum, by considering the quality of social network interactions in addition to network structure and pathogen and host characteristics, we constructed a comprehensive social-network-based disease-transmission model with significantly higher accuracy in predicting an individual's vulnerability to infection and their spreading capability. Estimating these variables across all 176 villages reveals that the RV, RSS and SSI of all the villagers shows strong significant correlation with salutations, on par with the number of connections individuals may have. This further strengthens the need for considering the quality of interaction in any disease transmission model.

This general model can be applied to any contact-dependent communicable disease. The quality of social interaction can be incorporated to provide an estimate of future personal risk of either becoming infected or being a super-spreader. This model can be used even in the absence of location-tracked pathogen transmission data available for a population, or even before the pathogen has a chance to invade a region, in order to identify vulnerable individuals based on structural data combined with interaction data alone.

Ethics. All participants provided informed consent at the time of data collection. Ethics approval was obtained from the Institutional Review Board (protocol no. 1506016012) at Yale University in New Haven, USA.

Data accessibility. Additional figures and tables have been included as a supplement, and the source code is available at: https://github.com/shiv-y23/Super-spreader_paper.git.

Authors' contributions. S.V.S., M.A. and N.A.C. conceived the idea. S.V.S. developed the model and performed the simulations. M.A. and N.A.C. refined the model and its implementation. All authors analysed the results together, and drafted and edited the manuscript. N.A.C. secured funding. All authors approved the manuscript.

Competing interests. We declare we have no competing interests.

Funding. The Bill and Melinda Gates Foundation, Tata Sons Limited, Tata Consultancy Services Limited, Tata Chemicals Limited (grant no. OPP1098684) and the NOMIS foundation partially funded this project.

Acknowledgements. We specially appreciate the efforts of Rennie Negron and Thomas Keegan who managed several field operations and data collection in Honduras, and Mark McKnight and Wyatt Israel, who developed the Trellis software. We thank all our study participants in Honduras.

References

1. Fowler JH, Dawes CT, Christakis NA. 2009 Model of genetic variation in human social networks. *Proc. Natl Acad. Sci. USA* **106**, 1720–1724. (doi:10.1073/pnas.0806746106)
2. Christakis NA, Fowler JH. 2014 Friendship and natural selection. *Proc. Natl Acad. Sci. USA* **111**(Suppl. 3), 10796–10801. (doi:10.1073/pnas.1400825111)
3. Nishi A, Alexander M, Fowler JH, Christakis NA. 2020 Assortative mating at loci under recent natural selection in humans. *BioSystems* **187**, 104040. (doi:10.1016/j.biosystems.2019.104040)

4. Kim DA, Benjamin EJ, Fowler JH, Christakis NA. 2016 Social connectedness is associated with fibrinogen level in a human social network. *Proc. R. Soc. B* **283**, 20160958. (doi:10.1098/rspb.2016.0958)
5. Newman MEJ. 2002 Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128. (doi:10.1103/PhysRevE.66.016128)
6. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. 2005 Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* **232**, 71–81. (doi:10.1016/j.jtbi.2004.07.026)
7. Keeling MJ, Eames KTD. 2005 Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307. (doi:10.1098/rsif.2005.0051)
8. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
9. Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, Schuchat A. 2004 Superspreading SARS Events, Beijing, 2003. *Emerg. Infect. Dis.* **10**, 256. (doi:10.3201/eid1002.030732)
10. Hattis RP, Halstead SB, Herrmann KL, Witte JJ. 1973 Rubella in an immunized island population. *JAMA: The Journal of the American Medical Association* **223**, 1019–1021. (doi:10.1001/jama.1973.03220090039008)
11. Britton T, Ball F, Trapman P. 2020 A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846–849. (doi:10.1126/science.abc6810)
12. Colman E, Spies K, Bansal S. 2018 The reachability of contagion in temporal contact networks: how disease latency can exploit the rhythm of human behavior. *BMC Infect. Dis.* **18**, 1–10. (doi:10.1186/s12879-018-3117-6)
13. Hamburger M, Green MJ. 1946 The problem of the dangerous carrier of hemolytic streptococci. *J. Infect. Dis.* **79**, 68–81. (doi:10.1093/infdis/79.1.33)
14. Chin WCB, Bouffanais R. 2020 Spatial super-spreaders and super-susceptibles in human movement networks. *Sci. Rep.* **10**, 1–9. (doi:10.1038/s41598-020-75697-z)
15. World Health Organization (WHO). 2018 See www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease.
16. Qureshi S, Atangana A. 2020 Fractal-fractional differentiation for the modeling and mathematical analysis of nonlinear diarrhea transmission dynamics under the use of real data. *Chaos, Solitons Fractals* **136**, 109812. (doi:10.1016/j.chaos.2020.109812)
17. Department of Statistics and Censuses (Honduras) 2019 See <http://ghdx.healthdata.org/organizations/national-institute-statistics-honduras>.
18. Ghana statistical service. 2020 See www.statsghana.gov.gh/.
19. Shakya HB *et al.* 2017 Exploiting social influence to magnify population-level behaviour change in maternal and child health: study protocol for a randomised controlled trial of network targeting algorithms in rural Honduras. *BMJ Open* **7**, e012996. (doi:10.1136/bmjopen-2016-012996)
20. Shakya HB, Christakis NA, Fowler JH. 2017 An exploratory comparison of name generator content: data from rural India. *Soc. Netw.* **48**, 157–168. (doi:10.1016/j.socnet.2016.08.008)
21. Perkins JM, Subramanian SV, Christakis NA. 2015 Social networks and health: a systematic review of sociocentric network studies in low- and middle-income countries. *Soc. Sci. Med.* **125**, 60–78. (doi:10.1016/j.socscimed.2014.08.019)
22. Lungeanu A, McKnight M, Negron R, Munar W, Christakis NA, Contractor NS. 2021 Using Trellis software to enhance high-quality large-scale network data collection in the field. *Soc. Netw.* **66**, 171–184. (doi:10.1016/j.socnet.2021.02.007)
23. Chen Q *et al.* 2020 Epidemiological characteristics and transmissibility of shigellosis in Hubei Province, China, 2005–2017. *BMC Infect. Dis.* **20**, 1–3. (doi:10.1186/s12879-020-04976-x)
24. Woroszyło C, Choi B, Profitós JH, Lee J, Garabed R, Rempala GA. 2018 Modeling household transmission dynamics: application to waterborne diarrheal disease in Central Africa. *PLoS ONE* **13**, e0206418. (doi:10.1371/journal.pone.0206418)
25. Nishiura H, Wilson N, Baker MG. 2009 Quarantine for pandemic influenza control at the borders of small island nations. *BMC Infect. Dis.* **9**, 1–14. (doi:10.1186/1471-2334-9-27)
26. Bradshaw WJ, Alley EC, Huggins JH, Lloyd AL, Esvelt KM. 2021 Bidirectional contact tracing could dramatically improve COVID-19 control. *Nat. Commun.* **12**, 1–9. (doi:10.1038/s41467-020-20325-7)

27. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. 2020 Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* **5**, e452–e459. (doi:10.1016/S2468-2667(20)30157-2)
28. Gilbert M *et al.* 2010 Flying over an infected landscape: distribution of highly pathogenic avian influenza H5N1 risk in South Asia and satellite tracking of wild waterfowl. *EcoHealth* **7**, 448–458. (doi:10.1007/s10393-010-0672-8)
29. Chin WCB, Bouffanais R. 2020 Spatial super-spreaders and super-susceptibles in human movement networks. *Sci. Rep.* **10**, 1–19. (doi:10.1038/s41598-020-75697-z)
30. Onnela JP, Arbesman S, González MC, Barabási AL, Christakis NA. 2011 Geographic constraints on social network groups. *PLoS ONE* **6**, e16939. (doi:10.1371/journal.pone.0016939)
31. Onnela JP *et al.* 2007 Structure and tie strengths in mobile communication networks. *Proc. Natl Acad. Sci. USA* **104**, 7332–7336. (doi:10.1073/pnas.0610245104)
32. Cauchemez S *et al.* 2011 Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc. Natl Acad. Sci. USA* **108**, 2825–2830. (doi:10.1073/pnas.1008895108)