

Predicting and containing epidemic risk using friendship networks

Lorenzo Coviello, Massimo Franceschetti, Manuel García-Herranz, Iyad Rahwan

Abstract—Physical encounter is the most common vehicle for the spread of infectious diseases, but detailed information about said encounters is often unavailable because expensive, unpractical to collect or privacy sensitive. The present work asks whether the friendship ties between the individuals in a social network can be used to successfully predict and contain epidemic risk. Using a dataset from a popular online review service, we build a time-varying network that is a proxy of physical encounter between users and a static network based on their reported friendship – the encounter network and the friendship network. Through computer simulation, we compare infection processes on the resulting networks and show that friendship provides a poor identification of the individuals at risk if the infection is driven by physical encounter. This result is not driven by the static nature of the friendship network opposed to the time-varying nature of the encounter network, as a static version of the encounter network provides more accurate prediction of risk than the friendship network. Despite this limit, the information enclosed in the friendship network can be leveraged for monitoring and containment of epidemics. In particular, we show that periodical and relatively infrequent monitoring of the infection on the encounter network allows to correct the predicted infection on the friendship network and to achieve satisfactory prediction accuracy. In addition, the friendship network contains valuable information to effectively contain epidemic outbreaks when a limited budget is available for immunization.

I. INTRODUCTION

The forecast and mitigation of epidemics is a central theme in public health [17], [25], [26], [27], [34], [42], [56]. Events such as the recent ebola epidemic constantly drive the attention and resources of governments, institutions such as the World Health Organization, and the research community [31], [35], [51], [55], [58], [71]. The study of infectious processes on real-world networks is of interests to diverse disciplines, and similar models have been proposed to characterize the spread of information, behaviors, cultural norms, innovation, as well as the diffusion of computer viruses [30], [54], [69], [72], [84]. Therefore, epidemiologists, computer scientists and social scientist have joint forces in the study of these phenomena. Due to the impossibility to study the spread of infectious diseases through controlled experiments, modeling efforts have prevailed [36], [47], [50], [68], [69]. Recently, advancements in computation tools determined the emergence of data-driven simulations in the study of dynamical processes [86].

The spread of an infection over a real-world network is determined by the interplay of two processes: the dynamics *of* the network, whose edges change over time according to the encounters between individuals; and the dynamics *on* the network, whose nodes can infect each other after an encounter.

When the two dynamics operate at comparable time scales, the time-varying nature of the network cannot be ignored [33], [41], [48], [70], [76] and specifically devised control strategies are necessary [53]. Aggregating the dynamics of the edges into a static version of the network can provide useful insights [21] but can introduce bias [39], [70]. Empirical work suggests that bursty activity patterns slow down spreading [46], [80], [85], but temporal correlations seem to accelerate the early phase of an epidemic [45], [74]

Physical encounter is the most common vehicle for the spread of infectious diseases (as in the case of airborne diseases), and detailed information about said encounters is fundamental to monitor and contain outbreaks. Various sources of data can serve as a proxy of physical encounter – checkins on social networking platforms [14], [63], [64], traffic records [5], [82], [83], phone call records [32], [38], [65], wearable sensor data [12], [37], [43], [66], [75], [80], geographical and non-geographical information shared online [4], [13], surveys and diaries of daily contact [20], [59], [60], and multiplex data [81]. However, pervasive and detailed information is rarely available and might be expensive and unpractical to collect (as in the case of sensor technologies [12], [75], [80]), prone to errors (as in the case of survey data [19], [73]), and privacy-sensitive [2], [7], [8], [18], [49], [52], [77], [88]. Researchers have to rely on the information in their possession, and in this paper we consider friendship ties between the users of an online social network. Location data from cell phone records and online social networks has shown that social relationships can partially explain the patterns of human mobility [15]. Contact tracing based on phone communication activity has been proposed as a method to reduce the final size of epidemic outbreaks [23]. Networks generated from wearable sensor measurements, diaries of daily contacts, online links and self-reported friendship present similar structural properties [57], but contacts of short duration recorded by wearable sensors might not be reported in surveys [78]. In general, it is not clear whether and within which limits friendship can be considered a reliable proxy of physical encounter, as a process spreading from an initial seed, or “patient zero”, can only reach the nodes within its *set of influence* through paths that respect time ordering [40].

Given an infection spreading over a social network via physical encounter, the present work asks whether the friendship ties between the same individuals successfully predict who is at risk, and how this information can be leveraged to monitor and contain epidemic outbreaks. Using the Yelp

Dataset Challenge dataset¹, we build a time-varying network that is a proxy of physical encounter between users and a static network based on their reported friendship – the encounter network and the friendship network. For comparison, we also consider a static version of the encounter network, in which temporal information is ignored. Through computer simulation, we compare the evolution of Susceptible-Infected (SI) processes [3] on the different networks, in terms of the sets of infected individuals. First, given epidemic processes spreading independently on the two networks but initiated at the same seed, we analyze infections from a microscopic point of view. Despite distance on the friendship network is correlated to risk of infection on the encounter network, the set of nodes infected on the friendship network is not a good approximation of those infected in the encounter network. Then, we consider how this limit can be overcome. On the one hand, we show that periodical yet relatively infrequent observation of the real infection allows to reach and maintain high prediction accuracy using the friendship network. On the other hand, we show that friendship ties allow to effectively allocate a limited immunization budget, in order to reduce the risk of an outbreak.

II. DATASET

The Yelp Dataset Challenge dataset (Round 5) consists in 1,569,264 reviews and 495,107 tips to 61,184 businesses (in 10 cities around the world) posted by 366,715 users over more than 10 years. Within this period, we consider 1,469 consecutive days ranging from 1/1/2011 to 1/8/2015 (period of observation T), as reviews and tips before year 2011 are less numerous. Each review and tip includes the user who posted it, the reviewed business, and the date it was posted. Yelp users can form friendship ties between each other, and the list of friends of each user is included in the dataset. The encounter network (SI Section 2 for details) is a sequence of networks $\{N_E(t), t \in T\}$, where $N_E(t)$ is the undirected network of encounters on day t . We assume that two users x, y , encountered on day t , and $(x, y) \in N_E(t)$, if they posted reviews or tips to the same business on the same day (133,038 users had at least one encounter during T). For node s , let $t_0(s)$ be the smallest t such that s is connected in $N_E(t)$ (first encounter). The static version of the encounter network N_S ignores the temporal information and $(x, y) \in N_S$ if $(x, y) \in N_E(t)$ for some t (we consider the 113,187 nodes in the largest connected component). The friendship network N_F is the undirected, static network of friendship ties (168,923 nodes in the largest connected component).

III. INFECTION DYNAMICS

We consider Susceptible-Infected processes [3], in which nodes do not recover (SI Section 3 for details). Given a temporal network $\{N(t), t \in T\}$, let $I(t)$ be the set of infected nodes at time t . Each simulation is initiated at a single seed $I(0) = \{s\}$, selected uniformly at random from the nodes

¹The dataset for ‘‘Round 5’’ of the challenge is considered: www.yelp.com/dataset_challenge.

in $N_S \cap N_F$ such that $t_0(s) \leq 500$, unless differently stated (to allow the infection to start early enough on the encounter network). At time t , each node x not in $I(t - 1)$ becomes infected with probability $\max\{\beta d_x(t), 1\}$, where $0 \leq \beta \leq 1$ is the infection rate and $d_x(t)$ is the number of neighbors of x in $N(t)$ that are in $I(t - 1)$. For the encounter network, $N(t) = N_E(t)$ and the infection spreads from time $t_0(s)$ until $\max T$. For the static (resp. friendship) network $N(t) = N_S$ (resp. N_F) for all t , and the infection spreads from time zero until all nodes are infected. We consider different values of β for the different networks, due to their different connectivity (usually $\beta = 0.5$ on the encounter network, and $\beta = 0.01$ on the other networks).

IV. RESULTS

Throughout, we assume that *real* infections spread on the encounter network. Infection spreading on the friendship network and the static version of the encounter network are *predicted* infections. Given an infection spreading on the encounter network, a node’s risk of becoming infected is correlated to its distance on the friendship network from the infected seed. Figure 1 plots the probability of nodes’ infection as a function of this distance, for different values of the infection rate β (10,000 simulations for each β). In order to evaluate how accurately the friendship network predicts epidemic risk at a microscopic level, we consider infection processes initiated at seeds present in both networks and spreading independently of each other, and compare the sets of infected nodes (SI Section 5 for details). The unpredictability of epidemic risk is due to the structural properties of the different networks as well as to the randomness of the infection processes. Therefore, for each of 5,000 seeds, we consider two independent infection processes on the encounter network, one on the friendship network and one on the static version of the encounter network (indexed by E_1 , E_2 , F and S , respectively). For target size m and infection A initiated at seed s , let $t_A(m; s)$ be the first time at which at least m nodes are infected (the quantity might be undefined). Let $I_A(m; s)$ be the corresponding infected set, whose size is at least m when defined. Consider two infections A and B initiated at the same seed s and possibly spreading on two different networks. If $I_A(m; s)$ and $I_B(m; s)$ are defined, their Jaccard similarity is given by

$$J_{A,B}(m; s) = \frac{|I_A(m; s) \cap I_B(m; s)|}{|I_A(m; s) \cup I_B(m; s)|},$$

These measures allow to characterize how accurately the friendship network and the static network predict epidemic risk on the encounter network (see SI for details and analyses involving different metrics). For each of 5000 selections of a random single seed, two simulations on the encounter network, one on the static network and one on the friendship network are run independently. Results are shown in Figure 2 and SI Section 5. The left panel shows the metrics $J_{E_1, E_2}(m; s)$, the baseline unpredictability due solely to the randomness of independent processes initiated at the same seed and spreading

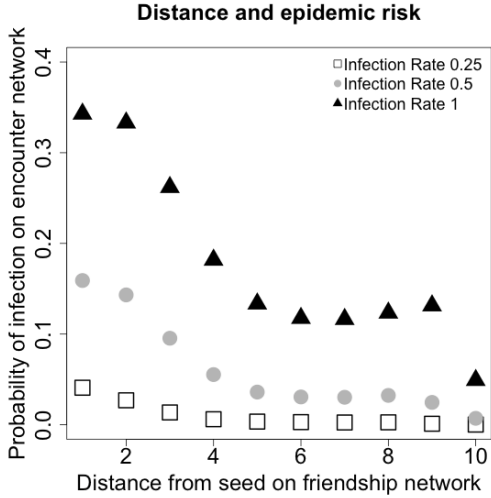


Fig. 1. Risk of infection on the encounter network versus distance from the seed on the friendship network. For each value of the infection rate β , 10,000 simulations on the encounter network initiated at random single seeds are run. The x -axis plots the distance d from the seed on the friendship network, the y -axis plots the empirical probability that nodes at distance d become infected on the encounter network.

on the encounter network. The middle panel shows the metrics $J_{E_1,S}(m; s)$, which includes the unpredictability due to the loss of temporal information in the static version of the encounter network. The right panel shows the metrics $J_{E_1,F}(m; s)$, which represents the unpredictability of using the friendship network to predict risk on the encounter network. For all values of the target m , two-sample paired t-tests support the hypotheses that $J_{E_1,E_2}(m; s)$ has larger average than both $J_{E_1,S}(m; s)$ and $J_{E_1,F}(m; s)$, and that $J_{E_1,S}(m; s)$ has larger average than $J_{E_1,F}(m; s)$ (SI Section 5-B). Notably, the intersections of the infected sets on the friendship and encounter network are substantially and significantly larger than the intersection of random sets (SI Section 5-C and Figure S8). A tight bound on $J_{E_1,F}(m; s)$ is not available for general m and s as individuals might be connected in one network and not in the other (there are $n_I = 76,933$ nodes in the intersection of the friendship and encounter network and $n_U = 225,028$ in their union), while $J_{E_1,E_2}(m; s)$ can be as large as one. Therefore, we also consider a rescaled version of $J_{E_1,F}(m; s)$ that highlights how the observed unpredictability results from structural differences of the networks (see SI Section 5-A). Together, the similarity measures $J_{\cdot,\cdot}(m; s)$ allow to characterize how the randomness of the infection process and the structural differences between the networks affect the unpredictability of epidemic risk. Our analyses show that, despite friendship networks produce similar epidemic dynamics at the macro level (SI Section 10 and 11), friendship provides a poor identification of the individuals at risk if the infection is driven by physical encounter. Despite the randomness of the infection determines unpredictability of the set of infected individuals (even between independent processes initiated at the same seed

on the encounter network), topological characteristics amplify such unpredictability when comparing two different networks. In addition, our results are not driven by the static nature of the friendship network opposed to the time-varying nature of the encounter network, as the static version of the encounter network provides more accurate prediction of risk than the friendship network.

The analyses reported thus far might erroneously lead to negative conclusions about the possibility of using the friendship network for prediction and containment of epidemic risk. From the point of view of risk prediction, we show that periodical and relatively infrequent monitoring of the real infection on the encounter network allows to correct the predicted infection on the friendship network, maintaining high levels of accuracy. This corresponds to a less extreme scenario in which the researcher has knowledge of the friendship network but, in addition, is able to monitor the infected population at given times. After each observation, the estimated set of infected individuals (on the friendship network) is updated to the real set of infected individuals (on the encounter network). Given a seed s present on both networks (and such that $t_0(s) \leq 900$), we consider two infections spreading on the encounter network and the friendship network for 500 time steps (denoted by F and E). Given an observation window W , every W time steps the predicted infected set $I_F(kW)$ on the friendship network is corrected to equal the real infected set $I_E(kW)$ on the encounter network. Figure 3 plots the Jaccard similarity of the sets of infected individuals on the two networks right before each correction, for window length $W \in \{10, 20, 50\}$ (6000 simulations for each W). A high level of prediction accuracy is established early in the process (after the first correction) and maintained over time (see SI Section 8 for detailed analyses and different measures). The accuracy decreases with larger window size, but even $W = 50$ guarantees good accuracy. Our results suggest that the ability to periodically monitor the infection on the encounter network is key to overcome the limits of the friendship network in predicting epidemic risk.

From the point of view of outbreak containment, we consider a scenario in which a fixed budget is available for immunization (e.g., limited amount of vaccine) and must be effectively allocated in order to contain an epidemic on the encounter network. In contrast to purely random immunization, we consider a strategy that selects random friends of randomly chosen individuals for immunization (friend immunization). Name-a-friend methods have already been proposed to predict the peak of an epidemic outbreak [16] and the spread of information online [28], and are motivated by the “friendship paradox” – the average friend of an individual is more connected than the average individual [24]. The method results in a more effective use of the immunization budget, substantially increasing the probability that an infection dies out in its early stage (Figure 4) and strongly reducing the final infection size (Figures 5 and S18-S20) with respect to random immunization. Moreover, it only requires a small additional cost to obtain the same effect as an ideal strategy that targets future encounters rather than friends (encounter immunization, see SI Section 9

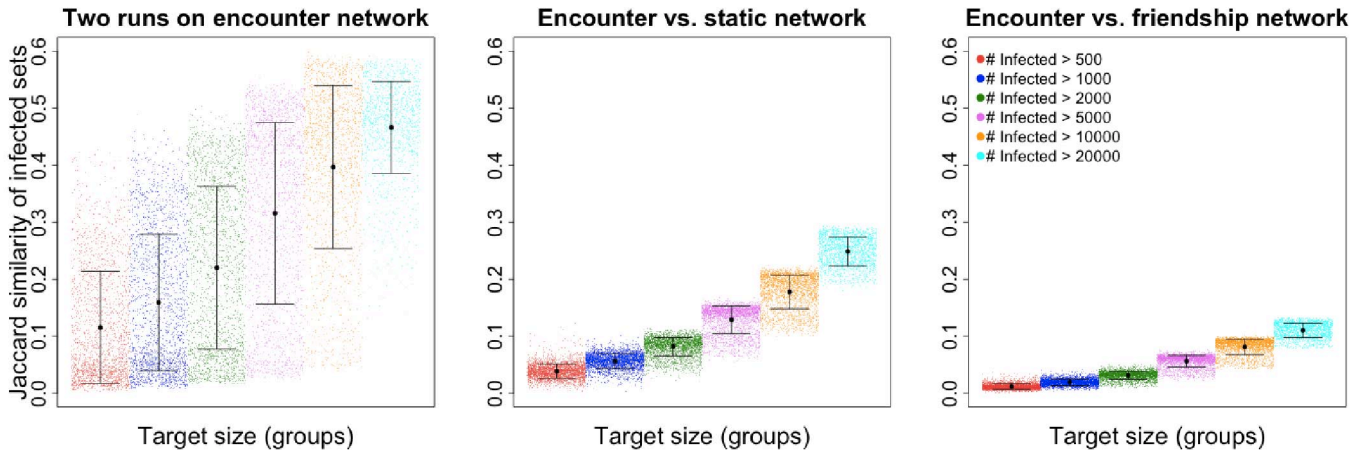


Fig. 2. Predictability of nodes' epidemic risk. For each of 5000 selections of a random single seed, two simulations on the encounter network, one on the static network and one on the friendship network are run independently. The similarities $J_{\cdot, \cdot}(m; s)$ of the infected sets are shown for different pairs of networks and different target infection size m . On the x -axis, observations for a given value of m form a block with a constant color (within the block, the x position is irrelevant). Black points represent the averages of the metrics over all observations such that the metrics are defined, and bars represent standard deviations.

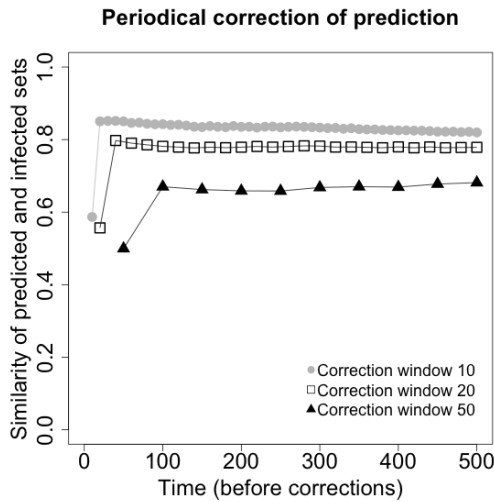


Fig. 3. Periodical correction of risk prediction using the friendship network. Shown here is the Jaccard similarity between the predicted infected set on the friendship network and the real infected set on the encounter network before each correction, for different values of the observation window W . For each $W \in \{10, 20, 50\}$, 6000 single seeds are selected at random, and for each seed one simulation on the encounter network and one (with correction) on the friendship network are run.

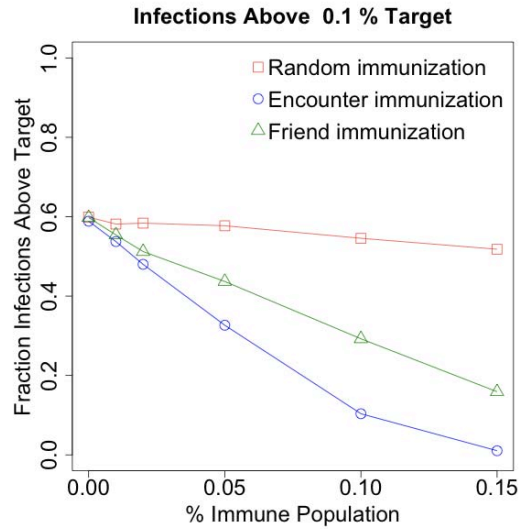


Fig. 4. Fraction of infections that do not die out in the early stage as a function of immunization budget b and immunization method. For each immunization type and $b \in \{1\%, 2\%, 5\%, 10\%, 15\%\}$, 5000 simulations on the encounter network initiated at random single seeds are run. The x -axis shows b , the y -axis shows the fraction of infections whose final size is above 0.1% of the entire population (taken as an indicator that the infection did not die out).

for details). Immunization budget is expressed as a fraction b of the entire population. For $b \in \{1\%, 2\%, 5\%, 10\%, 15\%\}$, Figure 4 shows the fraction of infections above 0.1% of the entire population as a function of b for all considered immunization methods (5000 simulations for each immunization method and b). The 0.1% target infection size is an indicator that the infection did not die out. The trend in Figure 4 is captured by a linear model with interactions between immunization type and immunization budget ($R^2 = 0.98$). Each 1% increase of the immunization budget determines: a 0.5%

decrease in the fraction of infections above the 0.1% target for random immunization (p-value= 0.0299); an additional 2.36% decrease for friend immunization (p-value= $2.77 \cdot 10^6$); and an additional 3.5% decrease for encounter immunization (p-value= $4.03 \cdot 10^8$). Regarding the size of the infected population, Figure 5 shows (for $b = 5\%$) the fraction of infections with final rate above given targets (5%, 10%, 15% of the entire population in the three panels) as a function of the infection start time $t_0(s)$ for all immunization methods. Friend immunization provides a large advantage with respect

to random immunization and, despite its simplicity, requires a relatively small additional cost to reach the same effectiveness as the benchmark of encounter immunization.

V. DISCUSSION

Since seminal work on the structure and growth of complex networks [6], [22], [87], interdisciplinary research has shown that biological networks, social networks and the Internet are governed by similar rules [1], [9], [44], [61] and share similar structure [29], [62], [67]. Very similar models have been proposed to characterize the spread of epidemics, information, behaviors, and cultural norms. Despite the macroscopic similarity between processes spreading on different networks, our work shows that the differences in local connectivity determined by the two definitions of edges result in striking differences between the dynamics at a microscopic level, that prevent the identification of the nodes at risk using a friendship network. However, this limitation can be overcome by periodical yet infrequent monitoring of the real infection on the encounter network. In addition, in the context of immunization with limited budget, simply asking individuals to name a friend enables the effective use of the available resources, increasing the probability that the infection dies out in its early stage and reducing the final size of the infected population.

We considered reviews as a proxy of physical encounter – an edge is active between two users on day t if they posted a review to the same business on day t . This constitutes an approximation to real physical encounter, that would require users to visit (rather than review) a business at about the same time. This approximation is justified as the time of a review is a proxy of the time of the visit to a business, and infections do not necessarily require direct physical contact. For example, in the case of airborne transmission, particles can remain suspended in the air for hours after an infected individual has occupied a room [10]. In the context of our dataset, after an infected user visits a business, the infection might spread to customers who visit the business later in the day. Also, the virus can infect customers which are not included in the dataset, and from them can infect another user who visits the business in a later moment.

When it is known who is infected or likely to become infected (e.g., individuals traveling to certain countries who might have come in contact with a pathogen), accurate prediction of the individuals at risk of contagion would allow targeted monitoring and immunization. Despite friendship and other social relationships might be informative about the encounters between individuals, our work suggests that they do not always give a complete picture of the paths a pathogen might take. Information to predict future encounters between individuals is likely to be unavailable, at least at a detailed level. However, a feasible approach could use past encounter as a proxy of future encounter. In fact, it is known that human mobility and encounter present high spatial and temporal regularity and predictability [11], [32], [79], [83]. From a practical perspective, networks based on social relationships (such as a friendship network) might be complemented by information

about past encounter. Our simulations are based on a large dataset that allowed us to build a static friendship network and a time-varying encounter network that is a candidate vehicle for the spread of a pathogen. The dataset considers a large number of individuals and spans several years of activity. In general, other datasets might be available and allow similar analyses. Friendship networks whose edges have a different semantic than that considered in the present work might lead to different observations.

REFERENCES

- [1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [2] Yaniv Altshuler, Nadav Aharoni, Yuval Elovici, Alex Pentland, and Manuel Cebrian. *Stealing reality: when criminals become data scientists (or vice versa)*. Springer, 2013.
- [3] Roy M Anderson and Robert McCredie May. *Infectious diseases of humans*, volume 1. Oxford university press Oxford, 1991.
- [4] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, pages 61–70. ACM, 2010.
- [5] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [7] Louise Barkhuus. The mismeasurement of privacy: using contextual integrity to reconsider privacy in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 367–376. ACM, 2012.
- [8] Louise Barkhuus and Anind K Dey. Location-based services for mobile telephony: a study of users’ privacy concerns. In *INTERACT*, volume 3, pages 702–712. Citeseer, 2003.
- [9] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.
- [10] Gabrielle Brankston, Leah Gitterman, Zahir Hirji, Camille Lemieux, and Michael Gardam. Transmission of influenza a in human beings. *The Lancet infectious diseases*, 7(4):257–265, 2007.
- [11] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [12] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PLOS ONE*, 5(7):e11596, 2010.
- [13] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.
- [14] Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z Sui. Exploring millions of footprints in location sharing services. *ICWSM*, 2011:81–88, 2011.
- [15] Eunjoon Cho, Seth A Myers, and Jure 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*, pages 1082–1090. ACM, 2011.
- [16] Nicholas A Christakis and James H Fowler. Social network sensors for early detection of contagious outbreaks. *PLOS ONE*, 5(9):e12948, 2010.
- [17] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC medicine*, 5(1):34, 2007.
- [18] Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleyesen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 2013.
- [19] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

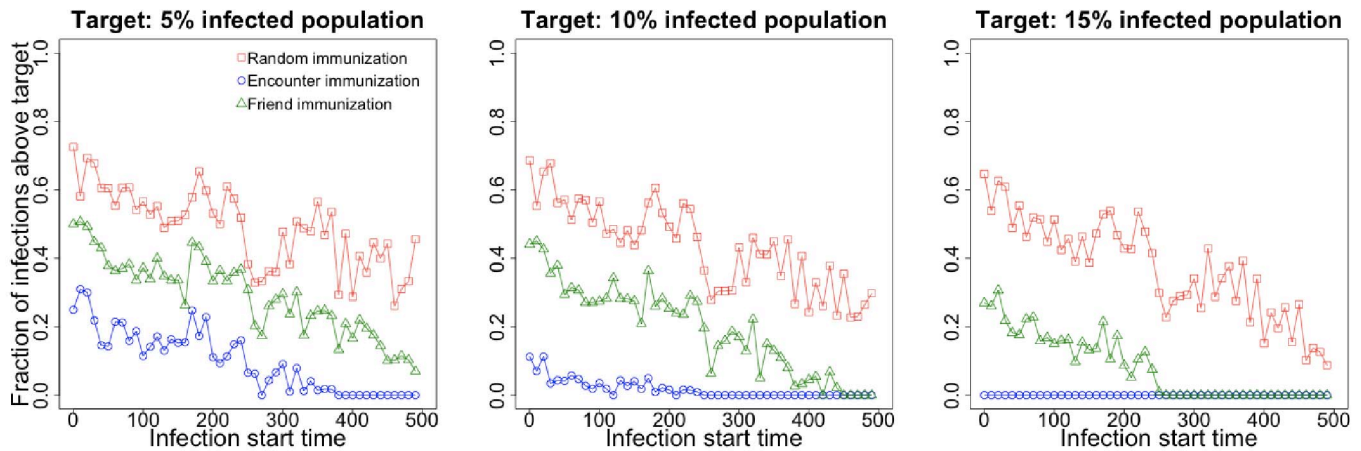


Fig. 5. Final infection size as a function of the immunization method and the infection start time. Given immunization budget $b = 5\%$ of the entire population, for each immunization type, 5000 simulations on the encounter network are initiated at random single seeds. Each panel considers a target infection size expressed as a percentage of the entire population. The x -axis shows the infection start time $t_0(s)$ of seed s , the y -axis shows the fraction of infections whose final size is above the given target.

- [20] W John Edmunds, CJ O'callaghan, and DJ Nokes. Who mixes with whom? a method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1384):949–957, 1997.
- [21] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.
- [22] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, pages 251–262. ACM, 1999.
- [23] Katayoun Farrahi, Remi Emonet, and Manuel Cebrian. Epidemic contact tracing via communication traces. *PLOS ONE*, 9(5):e95133, 2014.
- [24] Scott L Feld. Why your friends have more friends than you do. *American Journal of Sociology*, pages 1464–1477, 1991.
- [25] Neil M Ferguson, Derek AT Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437(7056):209–214, 2005.
- [26] Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.
- [27] Neil M Ferguson, Matt J Keeling, W John Edmunds, Raymond Gani, Bryan T Grenfell, Roy M Anderson, and Steve Leach. Planning for smallpox outbreaks. *Nature*, 425(6959):681–685, 2003.
- [28] Manuel Garcia-Herranz, Esteban Moro, Manuel Cebrian, Nicholas A Christakis, and James H Fowler. Using friends as sensors to detect global-scale contagious outbreaks. *PLOS ONE*, 9(4):e92413, 2014.
- [29] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
- [30] William Goffman et al. Generalization of epidemic theory. an application to the transmission of ideas. Technical report, DTIC Document, 1964.
- [31] Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M Elizabeth Halloran, and Alessandro Vespignani. Assessing the international spreading risk associated with the 2014 west african ebola outbreak. *PLOS Currents Outbreaks*, 1, 2014.
- [32] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [33] Thilo Gross and Bernd Blasius. Adaptive coevolutionary networks: a review. *Journal of The Royal Society Interface*, 5(20):259–271, 2008.
- [34] M Elizabeth Halloran, Ira M Longini, Azhar Nizam, and Yang Yang. Containing bioterrorist smallpox. *Science*, 298(5597):1428–1432, 2002.
- [35] M Elizabeth Halloran, Alessandro Vespignani, Nita Bharti, Leora R Feldstein, KA Alexander, Matthew Ferrari, Jeffrey Shaman, John M Drake, Travis Porco, JN Eisenberg, et al. Ebola: mobility data. *Science (New York, NY)*, 346(6208):433, 2014.
- [36] JAP Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons, 2000.
- [37] Tracy Heibeck and Alex Pentland. *Honest Signals: How They Shape Our World*. MIT Press, 2010.
- [38] Cesar A Hidalgo and C Rodríguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12):3017–3024, 2008.
- [39] Till Hoffmann, Mason A Porter, and Renaud Lambiotte. Generalized master equations for non-poisson dynamics on networks. *Physical Review E*, 86(4):046102, 2012.
- [40] Petter Holme. Network reachability of real-world contact sequences. *Physical Review E*, 71(4):046119, 2005.
- [41] Petter Holme. Information content of contact-pattern representations and predictability of epidemic outbreaks. *Scientific reports*, 5, 2015.
- [42] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, 2004.
- [43] Pan Hui, Augustin Chaintreau, James Scott, Richard Gass, Jon Crowcroft, and Christophe Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251. ACM, 2005.
- [44] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.
- [45] Hang-Hyun Jo, Juan I Perotti, Kimmo Kaski, and János Kertész. Analytically solvable model of spreading dynamics with non-poissonian processes. *Physical Review X*, 4(1):011041, 2014.
- [46] Márton Karsai, Mikko Kivela, Raj Kumar Pan, Kimmo Kaski, János Kertész, A-L Barabási, and Jari Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2):025102, 2011.
- [47] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.
- [48] Mikko Kivela, Raj Kumar Pan, Kimmo Kaski, János Kertész, Jari Saramäki, and Márton Karsai. Multiscale analysis of spreading in a large communication network. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(03):P03005, 2012.
- [49] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. Exploring privacy concerns about personal sensing. In *Pervasive Computing*, pages 176–183. Springer, 2009.

- [50] Jim Koopman. Modeling infection transmission. *Annu. Rev. Public Health*, 25:303–326, 2004.
- [51] Kai Kupferschmidt. Estimating the ebola epidemic. *Science*, 345(6201):1108–1108, 2014.
- [52] Scott Lederer, Jennifer Mankoff, and Anind K Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 724–725. ACM, 2003.
- [53] Suyu Liu, Nicola Perra, Márton Karsai, and Alessandro Vespignani. Controlling contagion processes in activity driven networks. *Physical review letters*, 112(11):118702, 2014.
- [54] Alun L Lloyd and Robert M May. How viruses spread among computers and people. *Science*, 292(5520):1316–1317, 2001.
- [55] Eric T Lofgren, M Elizabeth Halloran, Caitlin M Rivers, John M Drake, Travis C Porco, Bryan Lewis, Wan Yang, Alessandro Vespignani, Jeffrey Shaman, Joseph NS Eisenberg, et al. Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences*, 111(51):18095–18096, 2014.
- [56] Ira M Longini, Azhar Nizam, Shufu Xu, Kumnuan Ungchusak, Wanna Hanshaoworakul, Derek AT Cummings, and M Elizabeth Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.
- [57] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLOS ONE*, 10(9):e0136497, 2015.
- [58] Stefano Merler, Marco Ajelli, Laura Fumanelli, Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis L Chao, Ira M Longini, M Elizabeth Halloran, and Alessandro Vespignani. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases*, 2015.
- [59] RT Mikolajczyk, MK Akmatov, S Rastin, and Mirjam Kretzschmar. Social contacts of school children and the transmission of respiratory-spread pathogens. *Epidemiology and infection*, 136(06):813–822, 2008.
- [60] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 5(3):e74, 2008.
- [61] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [62] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [63] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PLOS ONE*, 7(5):e37027, 2012.
- [64] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. An empirical study of geographic user activity patterns in foursquare. *ICWSM*, 11:70–573, 2011.
- [65] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.
- [66] Eamonn O’Neill, Vassilis Kostakos, Tim Kindberg, Alan Penn, Danaë Stanton Fraser, Tim Jones, et al. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *UbiComp 2006: Ubiquitous Computing*, pages 315–332. Springer, 2006.
- [67] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
- [68] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *arXiv preprint arXiv:1408.2701*, 2014.
- [69] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.
- [70] Nicola Perra, Andrea Baronchelli, Delia Mocanu, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Random walks and search in time-varying networks. *Physical review letters*, 109(23):238701, 2012.
- [71] Chiara Poletto, MF Gomes, A Pastore y Piontti, Luca Rossi, L Bioglio, Denis L Chao, Ira M Longini, M Elizabeth Halloran, Vittoria Colizza, Alessandro Vespignani, et al. Assessing the impact of travel restrictions on international spread of the 2014 west african ebola epidemic. *Eurosurveillance*, 19(42), 2014.
- [72] Anatol Sapozort. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The bulletin of mathematical biophysics*, 15(4):523–533, 1953.
- [73] Jonathan M Read, Ken TD Eames, and W John Edmunds. Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface*, 5(26):1001–1007, 2008.
- [74] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput Biol*, 7(3):e1001109, 2011.
- [75] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.
- [76] Hiroki Sayama, Irene Pestov, Jeffrey Schoolmid, Benjamin James Bush, Chun Wong, Junichi Yamanoi, and Thilo Gross. Modeling complex systems with adaptive networks. *Computers & Mathematics with Applications*, 65(10):1645–1664, 2013.
- [77] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 247–262. IEEE, 2011.
- [78] Timo Smieszek, Victoria C Barclay, Indulaxmi Seeni, Jeanette J Rainey, Hongjiang Gao, Amra Uzicanin, and Marcel Salathé. How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC infectious diseases*, 14(1):136, 2014.
- [79] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- [80] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, et al. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9(1):87, 2011.
- [81] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. Measuring large-scale social networks with high resolution. *PLOS ONE*, 9(4):e95978, 2014.
- [82] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Manuel Cebrian. Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific reports*, 4, 2014.
- [83] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.
- [84] Thomas W Valente. *Network models of the diffusion of innovations*. Hampton Press Cresskill, NJ, 1995.
- [85] Alexei Vazquez, Balazs Racz, Andras Lukacs, and Albert-Laszlo Barabasi. Impact of non-poissonian activity patterns on spreading processes. *Physical review letters*, 98(15):158702, 2007.
- [86] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2012.
- [87] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.
- [88] Bin Zhou and Jian Pei. Preserving privacy in social networks against neighborhood attacks. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 506–515. IEEE, 2008.