Efficient detection of contagious outbreaks in massive metropolitan encounter networks

Lijun Sun¹,², Kay W. Axhausen¹,⁴, Der-Horng Lee² & Manuel Cebrian⁵

¹Future Cities Laboratory, Singapore-ETH Centre for Global Environmental Sustainability (SEC), 138602, Singapore, ²Department of Civil & Environmental Engineering, National University of Singapore, 117576, Singapore, ³Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA, ⁴Institute for Transport Planning and Systems (IVT), Swiss Federal Institute of Technology, Zürich, 8093, Switzerland, ⁵National Information and Communications Technology Australia, University of Melbourne, Victoria 3010, Australia.

Physical contact remains difficult to trace in large metropolitan networks, though it is a key vehicle for the transmission of contagious outbreaks. Co-presence encounters during daily transit use provide us with a city-scale time-resolved physical contact network, consisting of 1 billion contacts among 3 million transit users. Here, we study the advantage that knowledge of such co-presence structures may provide for early detection of contagious outbreaks. We first examine the “friend sensor” scheme - a simple, but universal strategy requiring only local information - and demonstrate that it provides significant early detection of simulated outbreaks. Taking advantage of the full network structure, we then identify advanced “global sensor sets”, obtaining substantial early warning times savings over the friends sensor scheme. Individuals with highest number of encounters are the most efficient sensors, with performance comparable to individuals with the highest travel frequency, exploratory behavior and structural centrality. An efficiency balance emerges when testing the dependency on sensor size and evaluating sensor reliability; we find that substantial and reliable lead-time could be attained by monitoring only 0.01% of the population with the highest degree.

Digital traces generated by citizens, during their commute across metropolitan transportation networks are helping answer long-standing questions on topics from individual mobility to collective interaction patterns. A series of landmark papers examining multiple large-scale digital traces has shifted the understanding of individual mobility patterns from random to highly structured and predictable¹–⁵. This has important implications in urban dynamics and epidemiology, particularly as the reproducible structure of metropolitan face-to-face encounters does significantly shape the spatial-temporal dynamics of disease spreading⁶–⁸. Therefore, advances in deciphering metropolitan encounter patterns play an important role in detection and mitigation of contagious outbreaks⁹–¹¹.

In detecting and containing contagious outbreaks, it is crucial to identify “super-spreaders”, as they may provide significant lead indicators for the early response of public health agencies¹²,¹³. To measure an individual’s importance in spreading processes, various centrality measures, such as degree, betweenness, closeness¹⁴, k-shell index¹⁵ and activity potential¹⁶ have been applied to theoretical diffusion models. Recent empirical works have confirmed the importance of these diverse measurements in real-world diffusion processes¹⁷–²⁰. To obtain such measurements, full knowledge about the contact network structure is usually required; however, other than simulating human interaction at this level of resolution²¹–²³, mapping such structure from real-world physical contact processes could be expensive to collect, computationally costly, laborious in the filtering of spurious connectivity, and privacy-sensitive²⁴. This has been particularly true for large metropolitan contact networks, where the availability of citywide datasets is still limited²⁵,²⁶.

Disease monitoring is extremely costly, privacy sensitive, and involves enormous technical difficulties. A low-cost contact network structure constructed from transit use may provide a way to design efficient monitoring strategies using a small fraction of the population. In this work, we examine the largest metropolitan encounter dataset collected to date - travel smart card data from all of Singapore’s bus users, covering approximately 3 million users during 1 week. Using one week’s tapping-in/tapping-out data collected from public transit services in Singapore, we built a large-scale high-resolution physical contact network. In a recent study based on this...
dataset, we demonstrated that physical encounters display a significant degree of temporal regularity and these rhythmic interactions form a large-scale spatial-temporal contact network, spanning all of Singapore for the whole week. The study emphasizes that encounters at this fine-grained scale are also very structured, and far from random. If the former study identified the global behavioral properties that generate this citywide co-presence network, our present study tries to identify the key individuals’ network properties that can be exploited to combat the spread of infectious disease.

As an alternative to constructing a global structure of contact networks, recent research exhibits an increasing interest in applying crowd-sourcing as a potential strategy to detect contagious outbreaks, from using declared “friends as sensors”, to aggregated search engine queries, to social media. Although these methods proposed are based on simple principles and require only small slices of information, they also show great advantages in providing early warning. Still, interesting questions remain in comparing the possible gains of using full knowledge vs. local methods in an epidemiological city-wide scenario. We perform such study in this high-resolution network, as a first evidence of its kind at a population and metropolitan level.

Results
To explore the dynamics of city-scale contagious outbreaks, we applied a general Susceptible-Exposed-Infected (SEI) model to simulate the spreading processes (see Methods and Supplementary Note 1). Briefly, a simulation run is initialized with ten infectious individuals (as index cases), who are selected randomly among all transit users on Saturday. In the temporal weighted physical encounter network (with each contact as an edge and its duration as weight), an infectious individual will transmit disease to neighbor with probability per 20 seconds (contagion rate ) is a universal parameter across the population and represents encounter duration; see Fig. 1A for example). Once a susceptible individual gets exposed, he/she becomes infectious after 2 hours, starting to spread the disease to other susceptible people. As almost all transit journeys are shorter than 2 hours, the introduction of this exposure stage prevents one from getting infected and then infects others directly during the same journey (which will significantly boost the spreading as instantaneous networks for a vehicle is always fully connected).

Note that is used in a high-resolution contact network in Ref7,20; we apply comparable values in our simulations. The full temporal resolution enables us to simulate the spreading processes during the whole week based on the proposed scheme for detecting contagious outbreaks, by registering infection time and transmission pathway on individual levels (Fig. 1B).

As mentioned above, a simple, but effective strategy for early detecting contagious outbreaks without mapping the detailed structure of a social network is to find friend sensors from the population. The inherent principle behind this method: a randomly selected “friend” (neighbor; in a friend group of one vertex (in a control group) has higher degree on average when the network has a heterogeneous degree distribution, implying that friend group is more central than the control group (or the population as a whole). This is commonly referred as the “friendship paradox”; your friends have more friends than you do. However, as social links initiated by physical encounters with strangers display a significant degree of heterogeneity, it remains unclear whether the friend sensor scheme obtained from a static network structure works in temporal spreading processes. Hence, to assess performance of the friend sensor scheme, we conducted multiple simulation experiments with different contagion rates . In each simulation, we first select 1% individuals from population randomly as a control set (the corresponding sensor group is composed of randomly selected neighbors of each individual in ( ) is a neighbor set of individual , respectively. Note that is a list
instead of a set since an individual might be selected repeatedly from different $N_i$. After obtaining results from 20 simulations, we measured the average infected ratio ($i/N_i$) of the sensor groups and ($i/N_P$) of the whole population temporally, finding that friend sensors have large lead-times (Fig. 1C). Given the heterogeneous individual participation and size of the time window\(^{35}\), spreading exhibits a linear increase - instead of a saturation process - after the explosive stage.

In Fig. 1D, we show the temporal change of infection incidence $i(t)$ from the same simulation as Fig. 1B. The sensor group $S$ is obtained by the same selection scheme; however, in this case, the control group $C$ is the whole population. Together with Fig. 1C, we found that spreading in $S$ not only happens earlier, but also faster than in the whole population, suggesting that the lead-time also varies with time (or infected ratio; see Supplementary Note 3 and Supplementary Fig. S1 and S2). Notably, although the temporal structure is not used in finding sensors, the friend sensor scheme is still efficient in early detecting outbreaks in our simulation experiments.

Considering that friend sensors are identified locally without using any centrality measures, they could be representative of a universally applicable strategy when it is costly or impossible to map the detailed network structure. To investigate the superiority of friend sensors in a comparable manner, we employed different centrality measures to quantify an individual's importance based on the both network structure and individual travel behavior employing the following centrality measures (see Supplementary Note 2): (1) Degree $k$, measuring total number of contacts of each individual during the week, (2) Travel frequency $f$, frequency of taking public transit services, ($f$ could also be interpreted as number of activities in temporal networks\(^{39}\)) (3) Shell index $k_s$, taken from $k$-shell decomposition\(^{13}\) on the static network and (4) Encounter entropy $S$, capturing temporal diversity of encounters:

$$S = - \sum_i p_i \ln p_i,$$

where $p_i$ is the probability of an individual's physical encounter in time $t$ (hourly). Using time-stamped encounter transactions, we can build the whole contact network and determine individual's centrality for both control and sensor sets (see Fig. 2).

Indeed, a sensor group is more central than the randomly selected control group in terms of degree $k$ (Fig. 2A); however, it is not yet known whether the friend paradox applies to other measures related to travel behavior (other than network structure). Before looking for additional sensors, we first measured other centrality distributions $P(f)$, $P(k_s)$ and $P(S)$ using both population and friend sensors. Although most people traveled less than 5 times during the week, we still found that $P(f)$ was characterized by a heavy tail across the population, indicating a significant degree of heterogeneity in individual transit use pattern (Fig. 2B). Moreover, we found that $P(f)$ of the sensor group clearly exhibited the friend paradox as well, indicating that the people you have encountered on buses traveled more often than you do. Using the same control and sensor groups, we then obtained the distributions $P(k_s)$ and $P(S)$. As Fig. 2C and D demonstrate, the friend paradox does exhibit in terms of shell index $k_s$ and encounter entropy $S$ as well, suggesting that friend sensors have higher $k$-shell indexes and show higher temporal encounter diversity than the population. Taken together, Fig. 2 suggests that the simple friend sensor scheme can universally identify more centrally located social sensors. Nevertheless, as the percentiles show (in all Fig. 2 panels), there are still significant differences between the most central individuals and friend sensors, further indicating that the efficiency of friend sensors might be limited. Taken together, as one might expect, the simple principle of friend sensor scheme also prevents itself from performing more efficiently, as better sensors could always be obtained by using more information on contact structure.

We next compare performance of the best sensors identified by each centrality measure against friend sensors by quantifying lead-time on a universal scale. When individual infected time cannot be obtained across the whole population, lead-time is estimated as difference between control and sensor samples in general\(^{22}\). However, since transit services are generally not operated 24 hours a day, the cumulative infection curve is not strictly monotonic increasing during the monitoring period in our case, resulting in significant difference when calculating lead-time from multiple runs; thus, using instantaneous lead-time is a biased measure of sensor performance (see Supplementary Note 3 and Supplementary Fig. S2). However, given that individual infection time can be traced from simulations, we can essentially quantify lead-time against the whole population instead of a small sample control group. For efficient early detection, we fixed the monitored infected ratio $\tilde{a} = [x_1, x_2] = [0.05, 0.25]$ and measured only the difference of infection time of people in $\tilde{a}$, obtaining infection time $t^a_F = \{t | x_1 \leq F_F(t) < x_2 \}$ from population and $t^a_s = \{t | x_1 \leq F_s(t) < x_2 \}$ from sensor group ($F$ represents the empirical distribution of exposed time). We re-define lead-time as the difference of average $t^a_s$ and $t^a_F$:

$$T = \langle t \rangle_F - \langle t \rangle_S$$
Next, we ordered individuals according to their centrality measure and divided the whole population into 100 percentiles. Using each percentile as a sensor group, we performed 20 simulation runs and measured the corresponding lead-times under different contagion rate $\beta$. As Fig. 3 shows, the top 1% slices from all these partitions are able to provide early detection; however, the less the average centrality is, the shorter the lead-time $T$ will be. For example, the sensor group provides no advance detection when $k \approx k_{0.4}$ and even falls behind the general population when $k > k_{0.4}$ ($k_{0.4}$ is the 40th percentile of degree). In this case, lead-time may reach infinity if the spreading cannot reach $s_2$ (25%) among sensor group. By comparing these centrality measures jointly, we found that they actually vary consistently on sensor composition; however, no one outperforms the others significantly (see Supplementary Fig. S3).

The efficiency of using such sensors to detect contagious outbreaks depends not only on centrality measures, but also sensor size $|S|$. On one hand, a small sample size induces large variation, providing poor reliability in potential applications. On the other hand, the average centrality measure might be more and more significant given the intrinsic heterogeneity of individual behavior, revealing that we may achieve longer lead-time with lower cost (if the cost is in proportion to sensor size). In Fig. 4A, we chose degree as primary centrality and measured lead-time for logistically spaced sampling rate $n_p = |S|/|P|$, spanning from 0.001% (only 27 people with highest degree) to 100% (the full population is used as sensors; lead-time is zero in this case). As the figure shows, smaller sample size indeed provides longer lead-time, but, with larger variation. In Fig. 4B, we show performance of friend sensors obtained from equally sized control groups. Given that the sensor group is always sampled from a deterministic population, we observed a constant average lead-time, independent of sampling rate $n$. However, the standard deviation of lead-time decreases as sample size gets larger in both Fig. 4A and B, corresponding to the law of large numbers when calculating lead-time in each simulation.

In practice, one should not just consider average lead-time and monitoring cost of such sensors; their reliability is equally important. To evaluate sensor reliability, we created a simulation result set with 500 runs and measured the lead-time distribution $P(T)$ for contagion rate $\beta = 0.001$. As Fig. 4C shows, average lead-time of different sensor sizes ($|S|$) when contagion rate $\beta = 0.001$, corresponding to panel (A). (D) The same plot as panel C, however, for friend sensors corresponding to panel (B).
Discussion

To summarize, we show the feasibility of a friend sensor scheme in providing early detection during a contagious outbreak in a metropolitan physical contact network. Indeed, the simple friend sensor scheme, which does not require a detailed network structure, works consistently well in finding sensors that are more central in the network. However, since all friend sensors are actually characterized by a deterministic neighbor population based on network structure, their performance is often limited by inherent characteristics of the neighbor population, providing constant early warning on average, independent of sample selection and sample size. Therefore, it is still crucial to show the value of full network structure, in particular for early detecting contagious outbreaks. Taking advantage of individual-based passive data collection techniques on city-scale (transit fare collection systems in this paper), we mapped detailed spatial and temporal structures as a whole and identified new sensors given diverse centrality measures, offering new insight into finding more efficient social sensors. Considering the weak, passive and indirect nature of social links enabled by these common daily physical encounters, $k$-shell index $k_s$ - a well-defined structural centrality - is less effective than the simple degree $k$ and frequency $f$ (number of activities) in contagious detection. Note that we did not use betweenness and closeness centralities as a measure in our study. On one hand, computing shortest path-based centralities is extremely time-consuming because of this network’s high density. On the other hand, considering the temporal nature of daily encounters, the role of static shortest path is not as significant as it is in social networks of personal relations. Based on the spreading settings examined in our study, a well-defined social sensor group based on degree may account for only 0.01% of the total population; however, it provides longer and more reliable lead-time - than the friend sensor scheme - allowing public health officials and governments to plan a quick and efficient response. In practice, those sensor individuals can be easily identified by transit agencies given their unique smart card IDs, and the remaining question is to monitor the health status of sensor groups. Although sensor groups are deterministic given our observations, it may still not be easy for public health agencies to monitor their status owing to privacy and technical issues, which are beyond the goal of our study. Nevertheless, one possibility is to use the emerging ICT in health monitoring - such as health-care applications in smartphones - and ask these special people to willingly provide anonymous information to public health agencies. Another possibility, assuming that health authorities could have access to the identities of these sensor group individuals, is to, instead of monitoring them directly, track their appearance as cases of selected contagious diseases reported by hospitals and local clinics. Were this sensor groups to appear with a certain statistical trend in clinical reports, we would have in our hands an early-warning signal for the future advancement of the contagious disease.

Having said that, influenza-like diseases are transmitted primarily by close contacts. Although the network used in our study is created across the whole metropolitan area, capturing all transit users’ contacts during a whole week, it still covers only a small slice of all potential contacts in our daily life, forming only a subnetwork of a network of contacts that would be important in the spread of actual epidemics. On the other hand, to simulate an outbreak, we fixed relatively unrealistic simulation settings, such as introducing only 2-hour exposed period and using an simplified SEI model instead of a full developed SIR or SEIR model, for the outbreak to travel at a speed where global and local methodologies could be tested. To what extent the simulation can match a real contagious outbreak and the relevance of the simulations findings to actual epidemics remain to be measured. Thus, it is important to note that the specific results in our study are embedded in the physical encounter network with a pre-defined spreading mechanism. Such encounters on transit vehicles occur more often between perfect strangers than among friends, colleagues or families, making the network incomplete for predicting epidemic spreading via all possible transmission pathways. Therefore, great caution should be exercised in interpreting the results. In reality, a full contact network for disease spreading consists of all of social links from diverse circumstances; it remains unclear to us which part should be given priority with respect to the characteristics of an unknown virus/disease. Nevertheless, with the rapid development of information and communication technologies, mapping the whole structure of close encounters from various data would be far less difficult and laborious today. Given the high individual and collective regularities rooted in human behaviors2-4, patterns of face-to-face encounters in various settings could be documented as well2-4, helping us build more comprehensive agent-based models to contain emerging epidemics5,6. Moreover, with our increasing knowledge about ourselves and various microorganisms around us, more efficient social sensors for different scenarios can be identified and applied in monitoring contagious spreading from day to day, providing early and accurate information to support better decision making. We believe that our work can serve as a base to help better combat the spread of disease on a citywide scale7,8 and better understand social contagion dynamics9-14.

Methods

Data sets. Trip records were collected from Singapore’s smart-card-based fare collection system, covering more than 90% of public transit trips. The system collects data for both bus and MRT (subway) modes. Smart card data is widely used in public transit network planning, service adjustments, providing ridership statistics, and indicating service performance. We employ bus, not MRT (Mass Rapid Transit, railway based) trip records in this study, since it is difficult to identify close proximity interactions on large MRT trains. For buses, once a smart card holder boards a vehicle (tapping-in), the system generates a temporary transaction record; after he/she leaves the vehicle (tapping-out), a complete record will be stored with detailed trip information. A full bus trip may contain more than one stages with transfers from one route/vehicle to another. The stage records are generated separately in the smart card system (with each tapping-in and tapping-off). Since our goal is to identify in-vehicle encounters and the people one may encounter in vehicles will differ from stage to stage, we use the term trip to represent stage in this document. After processing the raw data, we obtained the trip records used in this study. The fields and their contents are provided in Supplementary Tab. S1.

This study was performed on the trip records of one week in March, 2012. The dataset contains 22,455,159 bus trip transaction records from 2,969,320 individual smart card holders.

Simulation. To evaluate the performance of social sensors in the obtained interaction network, we use the SEI models to simulate contagious outbreaks among all transit users15, which are assumed to be in one of two states: susceptible (S) when they are prone to infection, exposed (E) between exposure and infectiousness, or infected (I) when they can transmit the disease to others. In studying the outbreak dynamics, we are more interested in the initial spreading processes and thus we do not consider the recovery stage in the simulation.

All simulations start on Saturday and end on the next Friday, spanning the whole week (given the dataset). In the spreading process, the duration of explosive stage (such that $0 < \beta < 1\%$) is highly determined by the number of index cases. Thus, a smaller index size induces larger in terms of temporal spreading processes; however, after this explosion, the spreading becomes steady and contagion rate $\beta$ determines the spreading speed of the rest spreading. Thus, to boost the initial spreading processes, we set ten index cases in our simulation, enabling us to observe outbreaks in one week. On the other hand, since people show great heterogeneity in their transit use behavior (such that $\tilde{\chi} = 5$ for almost 50% of the users during the week), a larger number of index cases also prevents the disease from dying out at initial stage. However, as sensor performance is monitored given infected ratio $5\% \leq F_2(t) \leq 25\%$ (after the explosive stage and during the steady spreading), lead-times and their variability are mainly determined by $\beta$ rather than number of index cases (see Supplementary Fig. S5). Thus, all our simulations start with ten infected people (ten infected cases), randomly selected across all transit users who were active on Saturday (who took buses on Saturday). After being infectious, individual will transmit disease to a susceptible individual $j$, who, individual $i$ encountered during his/her journey, with probability $p_0 = \beta \times d_{ij}$ (encounter duration). Here, $\beta$ is an important parameter determining the speed of contagious spreading. We chose a series of values from 0.001 to 0.005 per 20 seconds. On one hand, these values are similar to the value used in the other hand, simulating the speed of the different $\beta$, we can better evaluate the performance of different sensors for outbreaks with different $\beta$. Given any instantaneous network in a vehicle is a fully connected one, disease may spread very fast once one individual get infect. To avoid this, we intro-
duce the exposure (E) stage—which lasts for 2 hours—in our simulation: once an individual is exposed, he/she will not spread the disease immediately; however, he/she begins to spread the disease to other encounter people after 2 hours.

Considering that most transit trips take place in under 2 hours, one is unlikely to get infected and begin spreading disease to others during the same trip.

In the simulation, each time step represents 0.5 h, such as 7:00–7:30. In any step, we first identify all the neighbors he/she has encountered (the time they encountered each other should be within this time step) and then get them exposed with the defined probability $\beta$. The incubation time is selected as a constant given the time granularity (2 h), and thus, the exposed individuals become infectious in step $t + 4$. We also tested our results when setting the exposed period to be 6 h and 12 h, finding that sensors identified by degrees performs consistently better than others (Supplementary Fig. S6).

Based on these simulation settings, one can monitor the temporal spreading dynamics from a set of simulations with certain $\beta$ and random seeds as initial infected people. Meanwhile, individual infection time could be traced from each simulation. As Supplementary Fig. S1 shows, although contagion rate $\beta$ in each panel is the same, simulations still differ significantly from each other, in particular when $\beta$ is low. Thus, estimating lead-time universally is important to establish the difference.

1. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. Nature 439, 462–465 (2006).
2. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. Nature 453, 779–782 (2008).
3. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. Science 327, 1018–1021 (2010).
4. de Montjoye, Y. A., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. Unique in the Crowd: The privacy bounds of human mobility. Sci Rep 3, 1376 (2013).
5. Sun, L., Axhausen, K. W., Lee, D. H. & Huang, X. Understanding metropolitan patterns of daily encounters. Proc Natl Acad Sci USA 110, 13774–13779 (2013).
6. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci USA 106, 21484–21489 (2009).
7. Salathé, M. et al. A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci USA 107, 22020–22025 (2010).
8. Funk, S., Salathé, M. & Jansen, V. A. A. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 7, 1247–1256 (2010).
9. Huerta, R. & Tsimring, L. S. Contact tracing and epidemics control in social networks. Phys Rev E Stat Nonlin Soft Matter Phys 66, 056115 (2002).
10. Eubank, S. et al. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184 (2004).
11. Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding the spreading patterns of mobile phone viruses. Science 324, 1071–1076 (2009).
12. Galvani, A. P. & May, R. M. Epidemiology: dimensions of superspreading. Nature 438, 293–295 (2005).
13. Kitsak, M. et al. Identification of influential spreaders in complex networks. Nat Phys 6, 888–893 (2010).
14. Freeman, L. C. Centrality in social networks conceptual clarification. Soc Networks 1, 215–239 (1979).
15. Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of time varying networks. Sci Rep 2, 469 (2012).
16. Liu, L., Zhang, Y. C., Yeung, C. H. & Zhou, T. Leaders in Social Networks, the sociability and random seeds as initial infected people. forthcoming (2014).
17. Gilbert, M. T. et al. The emergence of HIV/AIDS in the americas and beyond. Proc Natl Acad Sci USA 104, 18566–18570 (2007).
18. Ginzberg, I. et al. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014 (2008).
19. Chan, E. H. et al. Global capacity for emerging infectious disease detection. Proc Natl Acad Sci USA 107, 21701–21706 (2010).
20. Garcia-Herranz, M. et al. Using friends as sensors to detect global-scale contagious outbreaks. arXiv preprint arXiv:1211.6512 (2012).
21. Shaman, J. & Karspeck, A. Forecasting seasonal outbreaks of influenza. Proc Natl Acad Sci USA 109, 20425–20430 (2012).
22. Hoda, N. O., Koot, F. & Lerman, K. Friendship paradox redux: Your friends are more interesting than you. arXiv preprint arXiv:1304.3480 (2013).
23. Salathé, M. et al. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 369(5), 401–404 (2013).
24. Anderson, R. M., May, R. M. & Anderson, B. Infectious diseases of humans: dynamics and control. Oxford University Press, New York, USA, (1992).
25. Feld, S. L. Why your friends have more friends than you do. Am J Sociol, 1464–1477 (1991).
26. Krings, G. et al. Effects of time window size and placement on the structure of an aggregated communication network. EPJ Data Sci 1, 1–16 (2012).
27. Smieszek, T. et al. Reconstructing the 2003/2004 H5N1 influenza epidemic in Switzerland with a spatially explicit, individual-based model. BMC Infect Dis 11, 115 (2011).
28. Acel Meyers, L., Newman, M. E. J., Martin, M. & Schrag, S. Applying network theory to epidemiics: control measures for mycoplasma pneumoniae outbreaks. Emerg Infect Dis 9, 204–210 (2003).
29. Ferguson, N. M. et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437, 209–214 (2005).
30. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. N Engl J Med 357, 370–379 (2007).
31. Bond, R. M. et al. A 61-million-person experiment in social influence and political mobilization. Nature 489, 295–298 (2012).
32. Christakis, N. A. & Fowler, J. H. Social contagion theory: examining dynamic social networks and human behavior. Stat Med 32, 556–577 (2013).

Acknowledgments
We thank Singapore’s Land Transport Authority for providing the smart card data. Special thanks to the developers ofigraph library. This work was supported by National Research Foundation of Singapore, which is the funding authority of Future Cities Laboratory. Manuel Cebran is funded by the Australian Government as represented by the Department of Broadband, Communications and Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

Author contributions
L.S., K.W.A. and M.C. wrote the paper. L.S., M.C. and K.W.A. designed the research. L.S. and M.C. performed the research. L.S. and M.C. wrote the paper.

Additional information
Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sun, L., Axhausen, K.W., Lee, D.-H. & Cebran, M. Efficient detection of contagious outbreaks in massive metropolitan encounter networks. Sci. Rep. 4, 5099, DOI:10.1038/srep05099 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images in this article are included in the article’s Creative Commons license, unless indicated otherwise in the image credit; if the image is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the image. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/