Predicting a Community's Flu Dynamics with Mobile Phone Data

Katayoun Farrahi, Rémi Emonet, Manuel Cebrian

Predicting a Community's Flu Dynamics with Mobile Phone Data

Katayoun Farrahi
Goldsmiths, University of London
Department of Computing
London, United Kingdom

Rémi Emonet
Jean Monnet University
Hubert Curien Laboratory
Saint Étienne, France

Manuel Cebrian
NICTA
University of Melbourne
Melbourne, Australia

ABSTRACT
Human interactions that are sensed ubiquitously by mobile phones can improve a significant number of public health problems, particularly helping to track the spread of disease. In this paper, we evaluate multiple avenues for the integration of high-resolution face to face Bluetooth-sensed interaction networks into standard epidemic models. Our goal is to evaluate the capacity of the different avenues of integration to track the spread of seasonal influenza on a real-world community of 72 individuals over a period of 17 weeks. The dataset considered contains real-time tracking of individual flu symptoms over the whole observation period, providing a concrete individualized source for evaluation. We obtain an error of less than 2 infected people on average for predicting the total number of individuals affected by the flu and precision of approximately 30% when predicting exactly which individual will become infected at a given time. To the best of our knowledge, this is the first study considering mobile phone Bluetooth-sensed interaction data for dynamic infectious disease simulation that is evaluated against real human influenza occurrence. Our remarkable results indicate that high-resolution mobile phone data can increase the predictive power of even the simplest of epidemic models.

Author Keywords
Health; Mobile sensing; Social interactions; Social computing; Epidemic models;

General Terms
Human Factors; Theory

INTRODUCTION
Traditionally, the field of epidemiology has faced two major challenges. The first regards the quality and granularity of data for epidemiological modeling (how much we know about an individual's health status, and how often we obtain updates of such health status). The second relates to the complexity of an epidemiological scenario (how epidemics are affected by seasonal effects, individual's health history, demographics, past and undergoing vaccination campaigns, and general cultural factors regarding hygiene and social interactions) [22]. Recent advances in face to face interaction sensing, for example sociometric badges [10], and wearable devices in general, are ground-breaking tools for epidemic research, and can significantly help address the challenges of data quality. In a number of studies [9, 11, 12, 20, 23], different communities were equipped with devices for fine-grained sensing of interaction dynamics for disease transmission modeling. However, these studies require individuals to wear such devices (normally around their necks) consistently during the study period. Given that the devices have no use other than helping the experimenters gather data, participants may forget to wear them, charge their battery, or activate them at the beginning of the day rendering them more suitable for small, experimental studies both in terms of the number of participants and the duration of time for which the face to face interactions are monitored.

Mobile phone Bluetooth sensors can complement such research by providing longitudinal dynamics of human interaction networks, with the major advantage of not requiring participants to use any additional sensors. Because most of the population carries mobile phones ubiquitously, this can potentially be a powerful tool for global research in health and disease transmission. While Bluetooth sensors can prove to be a much larger scale tool for understanding and modeling interactions, they also have some drawbacks. People may not turn their Bluetooth on; Bluetooth is available for devices other than phones (e.g. headsets, laptops), which can create spurious face to face interactions; Bluetooth signals may also be sensed across walls, creating even more spurious interactions; and of course, people may not always carry their phones with them, even though that seems to be a decreasing problem in recent times. However, given the value of Bluetooth as a global, cross-cultural interaction sensor, these drawbacks can be addressed potentially in the future by mobile phone's sheer volume of usage. This paper is a preliminary step towards evaluating Bluetooth sensors for simulating the spread of flu within a community. This is important as only limited number of previous studies have presented research in epidemiology simulated over mobile phone data at the individual level [5].

More generally, behavior inferred by mobile sensing has recently been shown to bear a relationship with changes in weather [19], political opinions [15, 4], and personality traits [3]. These promising advances in modeling human be-

Author-version of a paper submitted and accepted at *CSCW 2015*, March 14–18, 2015, Vancouver, BC, Canada.

havior reveal that mobile sensing may be a tool which could eventually address challenges currently facing epidemiology research, in term of both data quality as well as scenario complexity.

In this paper, we consider the interaction dynamics of a community of 72 participants obtained by mobile phone Bluetooth sensors, over a period of 17 weeks. The real mobile sensed interaction dynamics of the community are integrated and simulated in different epidemic models, in particular different Susceptible Infected Recovered (SIR) epidemic scenarios. SIR is the basic compartmental model used in epidemiology to model a group of individuals that can transmit a disease to others when they interact. Individuals in the SIR model can have one of three possible states: susceptible, infected, or recovered. The model details are covered in the section titled "The SIR Model". We consider two different classes of dynamic models simulated using the real interaction dynamics of a community obtained entirely by their mobile phone sensors. The interaction dynamics are considered to be aggregated on two different levels: on a weekly basis, and on a daily basis. We also consider a homogeneous interaction model, in which the amount of interaction between users (the weights of the edges in their interaction graph) are disregarded. We compare this to a heterogeneous scenario, where the pairwise interaction strengths between individuals are taken into account in the epidemic model. These multiple simulated scenarios are evaluated against participant real influenza occurrence, as reconstructed from survey questionnaires collected at the same time as the mobile phone data.

The general motivation of this paper is to determine whether mobile phones and the interaction information obtained from Bluetooth sensors, are suitable for modeling the spread of disease. In particular, we are interested in determining how much of the flu transmission in a community can be traced, and predicted using standard epidemic models. To the best of our knowledge, this is the first study considering mobile phone Bluetooth interaction data for dynamic infectious disease simulation that is evaluated against real human influenza occurrence – particularly over the long period of time we are able to obtain individual-level data data. Our longitudinal access to individual flu incidence of participants whose interactions are sensed makes the dataset and epidemiological scenario particularly unique and novel.

The data used in the present paper has previously been analyzed in studies involving public health [13]. However the focus in this previous study was to determine correlations between mobile phone usage patterns and individual well-being, not exploiting the structure of the interaction network for epidemiological studies. The focus of this work is to investigate the use of mobile phone interaction data for research in standard epidemiology. Our goal is to quantify the accuracy of incorporating mobile phone interaction data into dynamic epidemic models for infectious disease prediction.

Overall we find that Bluetooth sensors are a promising tool for obtaining large-scale continuous face to face interaction patterns in order to simulate disease spread. Furthermore, we find that it is possible to predict which particular individuals will become infected. The most accurate predictions result from the daily Bluetooth data, as the weekly data overestimates the infection rates. We also find a homogeneous model outperforms a heterogeneous one in predicting the infection rates and targeting which individuals become infected at a given time. This suggests that the important factor for epidemiological purposes is wether or not two individuals interacted, not necessarily the duration or intensity of their interaction. Taken together, these results provide robust support for the use of mobile phones as valuable tools for research in epidemiology, particularly since daily dynamics can be sensed for realistic epidemic simulations.

The contributions of this paper are as follows:

1. We propose two methods to incorporate real face to face interaction data into the SIR epidemic model. These methods are referred to as homogeneous and heterogeneous interaction models and encompass how the face to face interactions are incorporated (or weighted) into the SIR model. Note the methods proposed here are extensions of Stehlé et al. [23] which were presented over SEIR models (which is an extension of the SIR model to include a fourth state, exposed (E), that is not consider in the present work).

2. We simulate an epidemic over 17 weeks of real data-driven human interactions obtained by mobile phone data. This is the longest duration over which such a simulation has taken place. Previous studies presented results with real interaction data on the order of days.

3. We validate our results over these 17 weeks with actual participant infection ground truth, obtained simultaneously to the interaction data collection. Validation of the disease spread model to participant ground truth over a 17 week interval is novel, as are the evaluation measures obtained which can be used as a benchmark for future work.

4. This paper reports predictions of which particular individuals are infected over time, providing the first such performance results on a real dataset.

METHODOLOGY

Mobile Phone Data

The mobile phone data considered here has been collected at an undergraduate student dormitory at MIT [14], and has been made publicly available [18]. Please refer to Madan et al. [14, 18] for complete details about the dataset.

Our main dataset of interest is made of the physical proximity interactions obtained by mobile phone Bluetooth sensors. The data considered here are for 72 subjects over a 9 month period, between October 2008 to June 2009. The simulation results and results in the figures are presented based on weeks 15 to 31 of the study. This corresponds to the same period for which the influenza survey data was collected. This time period can be seen in figure 2. We consider each Bluetooth event sensed, not its duration, and we consider such Bluetooth dyadic interaction events as undirected. Whenever there is a Bluetooth interaction between individual A and individual B, we consider that both A interacted with B and B interacted

with A (interaction data is symmetrized). Bluetooth events are only considered between known devices (i.e. the 72 participants in the study). The distribution of the number of person to person interactions (averaged per day) is presented in figure 1. The top figure excludes $x = 0$ due to a large number of pairs of users never having interacted with anyone on a given day, on average. The bottom figure presented on a log-log scale includes $x = 0$. We can see that the majority of users have few daily interactions with each other, though several pairs of users interact regularly with a maximum of 55.3 average daily Bluetooth events sensed.

Though Bluetooth sensing by mobile phones contains many sources of noise, the distribution of the interactions over time presents a similar exponential distribution to that obtained by the RFID sensors by Stehlé et al. [23] (figure 1, bottom). In Stehlé et al.'s work, the distribution of the RFID interactions is presented over the duration in seconds, with a mean of 49 seconds. The average number of Bluetooth events sensed per day in this community is 1.1 and is presented in even counts.

Survey Data

Influenza-specific participant symptom data was collected using a daily survey designed by an experienced epidemiologist [13]. The survey consisted of 6 questions with Yes/No responses. In the present work we consider the following subset of four questions: (1) Do you have a sore throat or cough? (2) Do you have a runny nose, congestion or sneezing? (3) Do you have a fever? (4) Have you had any vomiting, nausea, or diarrhoea? Please refer to Madan et al. [13] for further details about the design of the questionnaire.

The number of participant self-reported symptoms, flu (vomiting, nausea, diarrhoea), fever, runny nose (including congestion or sneezing), and sore throat are plot in figure 2. For simulation results, we consider the number of unique participants who reported at least one symptom over time. This overall participant symptom data is presented over time in the results presented, and labeled as ground truth.

The SIR Model

The stochastic process based on the susceptible-infected-removed (SIR) epidemic model is used for cases where individuals recover after an infection of a particular strain of influenza. The SIR model is a compartmental model, meaning the population is divided into three different non-overlapping compartments: those individual who are susceptible to get the flu, $S(t)$; those infected with the flu, $I(t)$; and those removed (hospitalised) or recovered from the disease, $R(t)$. The standard model is given by the differential equations (1)-(3), describing the dynamics of the SIR process [1]:

$$\frac{dS}{dt} = -\beta SI \quad (1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

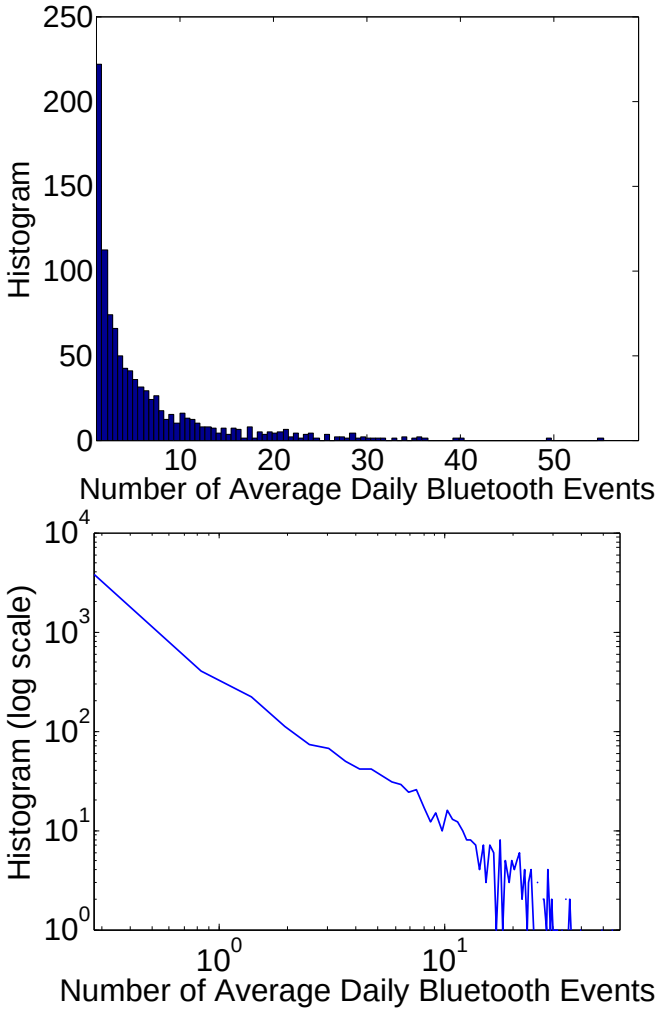

Figure 1. Distribution of the average daily individual Bluetooth interactions in the dataset. The top figure is for a number of events larger than 0, and the bottom figure — presented on a log log scale — includes the number of times individuals had 0 daily interactions. We can see a large number of users have no average daily interactions with any other individual, though several pairs of users have on average 10 or more (maximum 55.3) interactions per day.

where β is the rate of transmission of infection (the larger the rate, the higher the probability of the flu passing from an infected individual onto another upon interaction); γ is the recovery rate (the probability of an infected individual recovering spontaneously from the flu). In this paper, we assume there are no births or deaths as there are no new or lost users in the dataset, and thus $N = S + I + R$ is the total population size. We vary the values of β and γ in the results presented to evaluate a wide variety of epidemiological scenarios.

The SIR model, defined by equations (1)-(3), is not defined for network data. Therefore, here we adopt the methodology by Stehlé [22] to incorporate the interaction dynamics, and apply them them to the SIR process.

Stochastic Process

The stochastic process is simulated using a continuous time implementation of the SIR model. A continuous time ap-

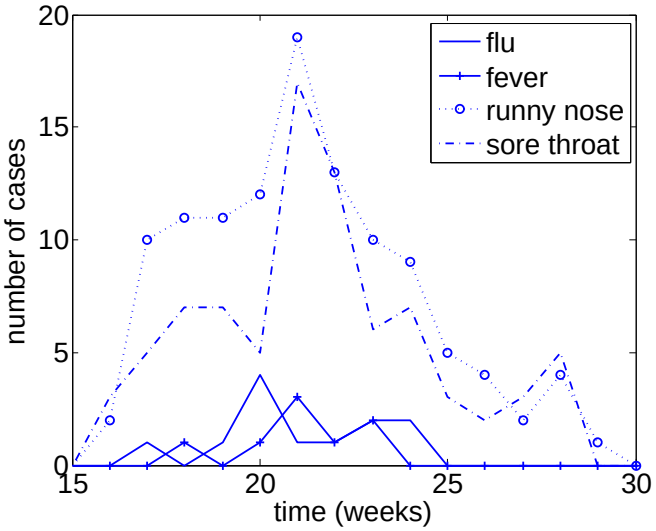

Figure 2. Distribution of participant self-reported symptoms, including flu (vomiting, nausea, diarrhoea), fever, runny nose (including congestion or sneezing), over time.

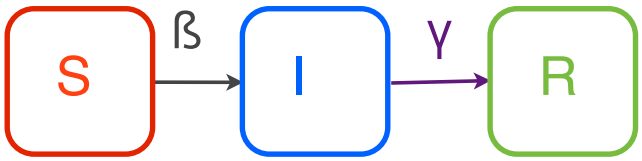

Figure 3. SIR model.

proach entails sampling the next time at which an individual will change his or her health state. We assume the probability of transitioning between health states is sampled from a geometric distribution with mean $\frac{1}{C}$ (from Susceptible to Infected) and $\frac{1}{D}$ (for Infected to Recovered) as follows.

$$p_{S\to I}(t) = Ce^{-Ct}, \quad C = k\beta \tag{4}$$

$$p_{I\to T}(t) = De^{-Dt}, \quad D = \gamma \tag{5}$$

At time t, for each susceptible individual n, we sample the next time of infection from equation (4). Note that k is the number of infected individuals in proximity to susceptible node n. We assume the time increment is $dt = 10 \times 10^{-6}$. Similarly, for each infected node, we sample the next time at which recovery will occur from equation (5). For each simulation result, we run $1,000$ random trials of the epidemic model.

For the initial infectious individuals, we consider two different scenarios. On the first experimental set, we assume one initial infectious individual, selected randomly from the whole population. On the second, for the results presented in figure 9, we use the actual initial infected individuals, as we are trying to predict which individuals are infected and not the overall number of infected individuals.

We consider dynamic models in which the interaction dynamics between the individuals simulated vary over time. This is in contrast to a static epidemiological model, which would

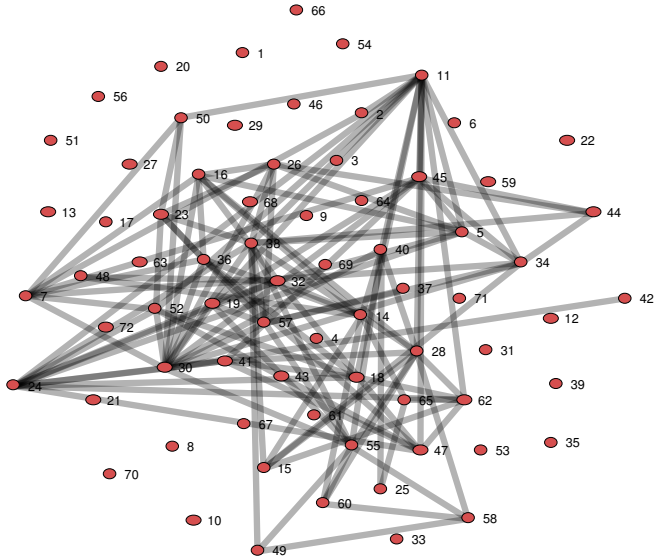

Figure 4. Homogeneous network model of participant interactions. This network is for day 10 of the dataset considered. The nodes are the participants and the edges represent interactions sensed by Bluetooth. Note that the edges are not weighed by the amount of interaction time for the homogeneous model.

assume that the interactions across individuals are constant over time (they either exist or they do not). Finally, we consider two dynamic network cases, a homogeneous model and a heterogeneous interaction model, which we explain next.

Dynamic Homogeneous Model

In the homogeneous case, the real interaction network is used to determine which pairs of individuals interact; however, the probability of disease transmission (the weight) is equivalent for each pair. This can be seen in figure 4, where the edges between nodes are all of equivalent weight and are simply present if at least one person-to-person interaction is sensed within the time interval x. The time interval x is either one week or one day depending on whether the simulation scenario is weekly or daily.

Dynamic Heterogeneous Model

The dynamic heterogeneous model extends the dynamic homogeneous model by considering network weights over time, as seen in figure 5. The greater the number of interactions between two individuals, the greater the weight and the more likely the infection will transmit between these two individuals. Again, this model considers the amount of contact between pairs of individuals in two scenarios of time-aggregation: a weekly or daily basis. The infection rate is weighed such that, $\beta W_{i,j}/\hat{W}$, where $W_{i,j}$ is the given week's (or day's) total number of interactions between nodes i and j, and $\hat{W}$ is the overall average interaction weight of the network.

RELATED WORK

Many applications which currently use individual interaction data sensed by social media can greatly benefit from the use of face to face interaction data sensed by mobile phone Bluetooth, particularly in the CSCW community. Applications

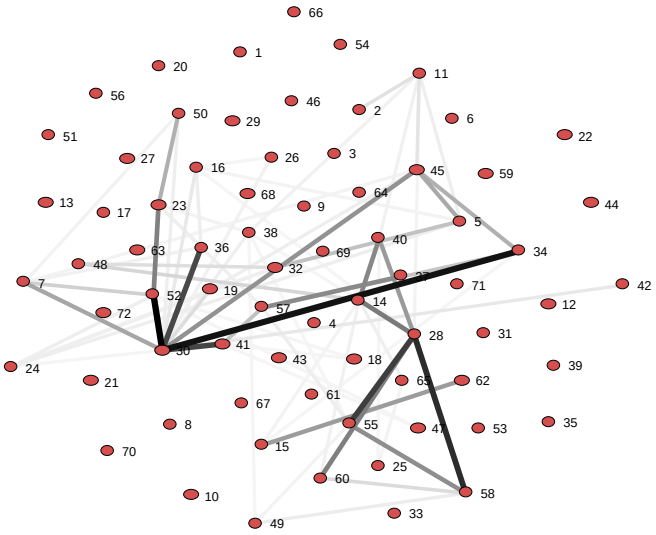

Figure 5. Heterogeneous network model of participant interactions. Network of interactions on day 10 of the dataset. The edges representing interactions are weighed in this network configuration, resulting in the appearance of far fewer interacting pairs in comparison to the homogeneous model of figure 4.

range from understanding information diffusion across social networks [21, 6], cross cultural studies [25], and collaboration [16], just to name a few. Some specific examples which have previously used face to face interaction data, beyond the health domain, include the inference of social relationships [17], mapping workplace behavior and understanding cultural differences [2].

As mentioned above, physical proximity interactions have recently been sensed using wearable badges for improved epidemic modeling. Stehlé et al. [23] consider a network of real-life human interaction patterns obtained by RFID devices. The authors simulate the spread of epidemics considering an SEIR model (SEIR is an extension of the SIR model which includes a fourth state, E for exposed). The 405 participants were attendees of a conference that lasted 2 days (3-4 June, 2009). The focus of the work is on varying timescales (20 second versus daily resolution) and network structures to determine the level of detail needed to correctly inform computational models for real epidemic management. Due to the lack of longitudinal data, the authors present different procedures to longitudinally extend their data.

Another closely related, motivating work is by Salathé et al. [20], where 788 participants' close proximity interactions (CPIs) are sensed on a school day to determine the high resolution dynamics of disease transmission considering an SEIR model. The authors' results suggest that contact network data is required for more effective immunization strategies. Isella et al. [8] use wearable RFID badges to monitor the dynamics of disease transmission in a hospital setting. In the study, 119 subjects are monitored over a one week period for monitoring the spread of respiratory infections; their goals are critical pattern discovery and tailored prevention strategies.

These recent studies reveal the importance of new data-driven approaches to epidemiology. In particular, high resolution human interaction network data have much to offer in the domain of public health (see [24] for a review). However, previous work has used real life human interaction data on the order of days. Bluetooth interaction data has much to offer in this respect, particularly due to its longitudinal nature (weeks to months to years). Even though Bluetooth interaction data may suffer from noise, it offers a *nearly high-resolution* human interaction network, comparable in quality to RFIDs, yet on orders of magnitude longer of duration and larger population samples.

RESULTS

Data Only

First we consider the likelihood of an infected individual being in proximity with another infected individual (figure 6). An infected individual is considered to have reported at least one symptom on a given day for the survey given. Figure 6 shows how often an infected individual was in proximity with another infected individual, considering a time window of up to 8 days (time presented on the x-axis). As the time window increases, there is a higher probability of an infected individual to have interacted with another infected individual. In this community, 30% of infected individuals did not have a Bluetooth interaction with another infected individual within a 8 day window (i.e. 7 days previous to the onset of infection reported). This result gives an indication of the best range of results we can expect to obtain from this experiment. Given a time window of 7 previous days, only 30% of infected individuals were not in physical proximity to another individual sensed by Bluetooth. This means that the 30% of infected individuals were likely either infected from an outside source or their interaction was not sensed by Bluetooth (if they were infected from someone within this population). Note that for the flu, we can never be certain who an individual is infected by based on interactions alone, but we can develop models which can make predictions about potential avenues for such infection.

Dynamic SIR Models

The simulation results of the homogeneous model are presented in figure 7. Our initial finding is that, considering weekly intervals for the interaction network, time aggregation tends to overestimate the actual number of infected cases. We observe that the daily time-aggregation gives more accurate counts (weekly is denoted by (W) and daily by (D) in figure 7). This is likely due to the network being much more dense when aggregated on a weekly basis, resulting in more predicted infection transmissions than really occur. This overestimation of weekly aggregates also holds for the heterogeneous model, presented in figure 8. We further evaluate the difference between the homogeneous and heterogeneous models below.

In figure 9 we consider the error over all combinations of parameters γ and β for all model cases. The error is the absolute difference in the real number of infectious cases and those predicted using the models run on the mobile phone data. The error is calculated as the absolute value of the average error determined for each week over the 17 week period.The results

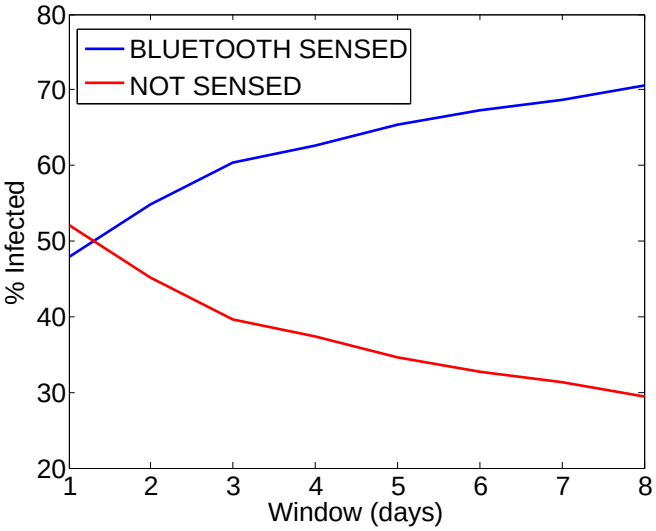

Figure 6. Percentage of infected individuals sensed (blue) / not sensed (red) in proximity to another infected individual with Bluetooth. The time window is given for the number of previous days up to infection (including the current day). Note, the "not sensed" cases are based on individuals with reported symptom data but without having been traced with Bluetooth as being in proximity with another infected individual.

re-confirm that the models perform better when we consider the data on a daily basis as opposed to a weekly basis. The minimum overall error occurs for the homogenous daily case, with an error of 1.89 (for $\beta = 0.2$ and $\gamma = 0.3$). The minimum error for the heterogenous daily experiments is 2.18 for $\beta = 0.1$ and $\gamma = 0.3$. The best performing model was able to predict with an overall average error of 1.89, which is the error in the number of infected individuals predicted. This means that, in theory, we could predict the number of infected individuals with an error of just under 2 people given this population of 72 people.

When optimizing the epidemic model parameters using only real data-driven interactions obtained by the mobile phones, we obtain the lowest errors for model parameters of β and γ which are in close agreement to parameters used in epidemiology research [7]. These results can be considered an indirect validation of our model and data.

We further consider the difference between the homogeneous and heterogeneous experiments on a daily scale for the optimal cases in figure 10, given that these parameter settings displayed the lowest errors. While the heterogenous model may be able to capture some of the 'jumps' in infectious cases over time, the homogeneous model provides better estimates for the number of infected cases over time. The reason for the better performance with the homogeneous model is likely the imbalance in weighting. With the current heterogenous model, if a few pair of individuals have a large number of interactions they will greatly outweigh the other interactions which may have occurred only once or twice (as can be seen by the difference between figures 4 and 5). However, the results indicate these interactions that may occur only a few times are important to weigh equivalently in order to better capture the spread of disease in this community.

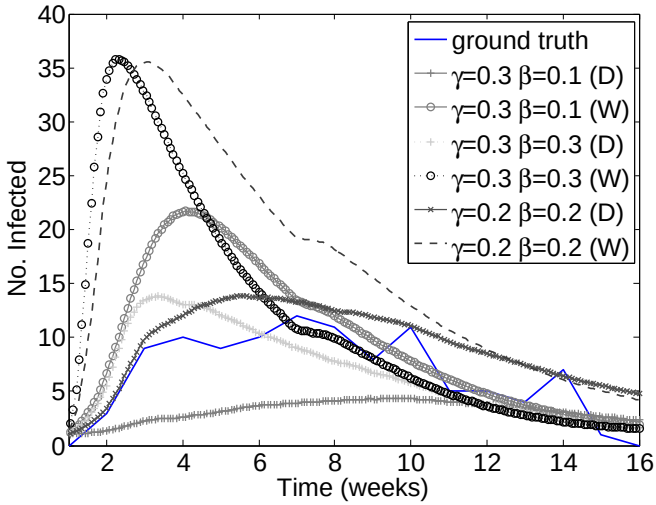

Figure 7. Simulation results for the dynamic homogeneous cases over varying model parameters. The real weekly (W) and daily (D) interactions are presented with the same colour for a given set of model parameters.

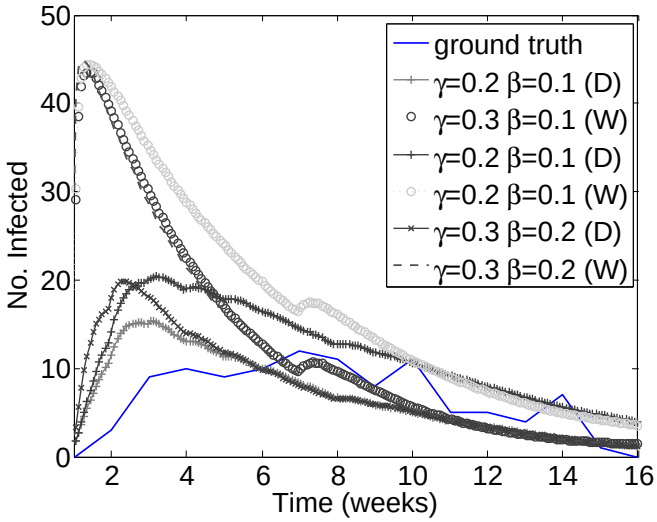

Figure 8. Simulation results for the dynamic heterogeneous cases over varying model parameters.

We further consider how well the homogeneous daily model can predict which particular individuals will become infected by looking at precision values next. In figure 11, the average precision is plot for a wide range of γ and β considering 1000 runs. Precision is the fraction of particular infected individuals correctly predicted by the model (note that this is not the aggregated number of infected individuals, but which particular individuals are correctly predicted to get the flu). In this case, the initial infections in the homogeneous daily model correspond to the actual infections in the dataset as opposed to initial infections chosen randomly as in previous experiments. Because increasing γ continuously improves the precision of the model, this reflects the model artefact that if takes longer and longer for infected individuals to recover, this artificially increases the likelihood of correctly identify-

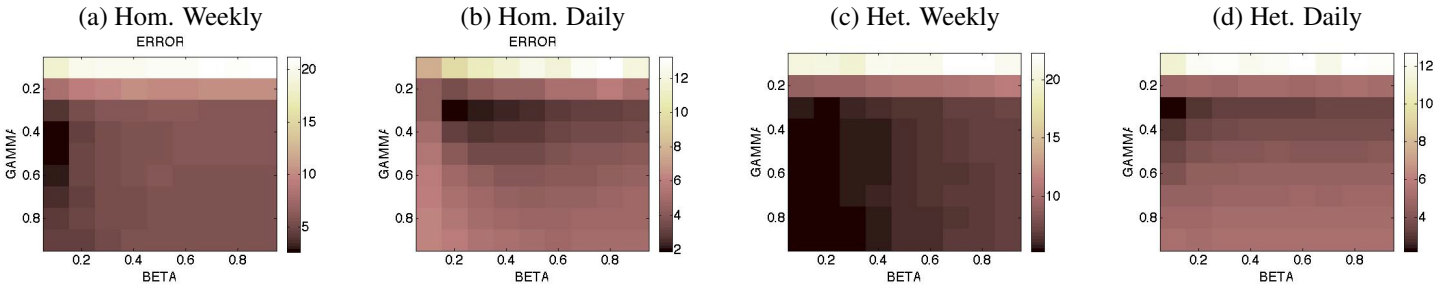

Figure 9. The overall average absolute error between the (a),(b) homogeneous and (c),(d) heterogenous models considering weekly and daily time frames.

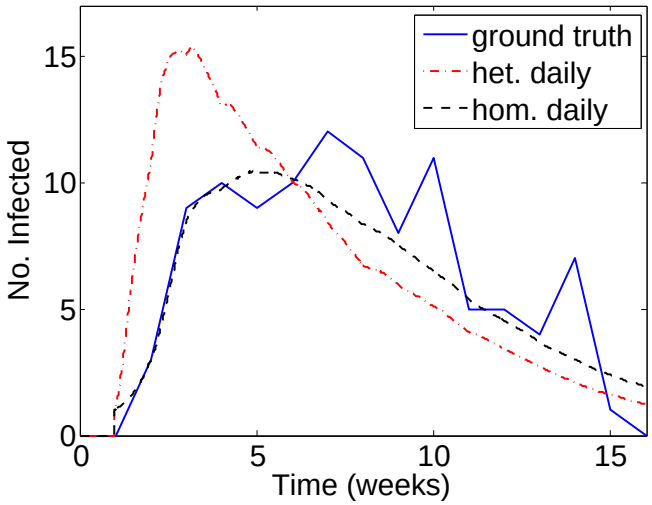

Figure 10. The best performing homogeneous (hom) and heterogeneous (het) daily models in comparison to the real data. For hom $\gamma = 0.3, \beta = 0.2$ and for het $\gamma = 0.3, \beta = 0.1$.

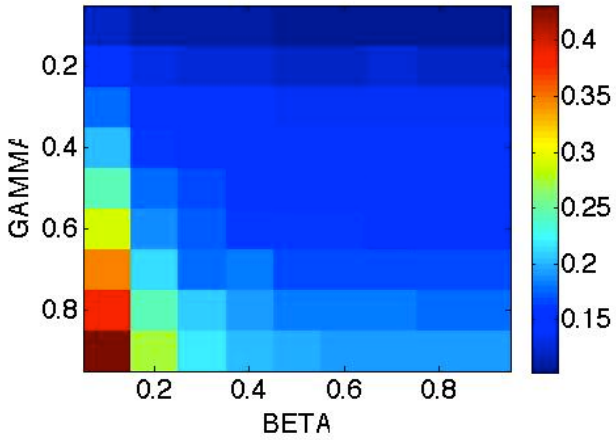

Figure 11. The average precision of the daily homogeneous model run over 1000 random trials. The epidemic was initialized with the actual infected cases. Precision here shows the fraction of infected individuals correctly predicted by the model.

ing them over time. This is an important insight to avoid mistakes when modeling with this type of data.

Looking at the more realistic parameter range of $\gamma = 0.1-0.5$ and $\beta = 0.1-0.5$, the homogeneous daily model can achieve a precision of nearly 30% over the entire 17 weeks simulated. Note, this implies that over the 17 weeks, 3 out 10 individuals were correctly pin-pointed as being infected. These results were averaged over 1000 random trials, and the results varied greatly over trials. However, in general, the precision is very high near the beginning of the experiment, and drops down to zero as time progresses, lowering the overall precision results presented here. This is expected since the errors in identifying infected individuals propagate given the nature of the models.

CONCLUSION

We incorporate an individual-level dataset obtained from mobile sensing (Bluetooth) into a classical epidemiological scenario and study how such new data can improve infectious disease prediction. By simulating weekly and daily Susceptible Infected Recovered (SIR) models for influenza transmission — i.e. models that incorporate mobile sensed interactions at a weekly and daily time scales — we find that daily sensed interactions provide improved prediction performance over weekly sensed interactions. Given a daily homogeneous model – using equal interaction weights for those pairs of individuals that interact at least once – we are able to achieve an error of less than 2 people out of 72 subjects when predicting the number of infected cases over a 17 week interval. Given the same homogeneous model, we can predict with 30% precision the actual individuals which will become infected within the 17 week period evaluated. This remarkable predictive power by such a standard epidemiological model indicates that high resolution mobile phone data can increase the predictive capacity of even the simplest epidemic models.

Extensions of this work include replicating the evaluation of our results with other Bluetooth interaction data collections in similar epidemiological scenarios where both the Bluetooth interactions and the health symptoms are obtained. Improved infection prediction based on mobile sensing could be further validated by considering other types of infectious diseases, as well as other existing epidemiological compartmental models.

ACKNOWLEDGEMENTS

The authors would like to thank Anmol Madan and Alex (Sandy) Pentland for collecting, processing and providing the empirical dataset. Thank you to Shirin Farrahi for proof reading the paper.

REFERENCES

1. L. J. S. Allen. An introduction to stochastic epidemic models mathematical epidemiology. volume 1945 of *Lecture Notes in Mathematics*, chapter 3, pages 81–130. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2008.
2. C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, and C. Mascolo. Tracking serendipitous interactions: how individual cultures shape the office. In *CSCW*, pages 1072–1081. ACM, 2014.
3. G. Chittaranjan, J. Blom, and D. Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal Ubiquitous Comput.*, 17(3):433–450, Mar. 2013.
4. K. Farrahi. *A Probabilistic Approach to Socio-Geographic Reality Mining*. PhD thesis, EPFL, Lausanne, 2011.
5. K. Farrahi, R. Emonet, and M. Cebrian. Epidemic contact tracing via communication traces. *PloS one*, 9(5):e95133, 2014.
6. M. Garcia-Herranz, E. Moro, M. Cebrian, N. A. Christakis, and J. H. Fowler. Using friends as sensors to detect global-scale contagious outbreaks. *PloS one*, 9(4):e92413, 2014.
7. R. Huerta and L. S. Tsimring. Contact tracing and epidemics control in social networks. *Physical Review E*, 66(5):056115, 2002.
8. L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. Van den Broeck, F. Gesualdo, E. Pandolfi, L. Rav, C. Rizzo, and A. E. Tozzi. Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLOS ONE*, 6(2):e17144, 02 2011.
9. L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J. Pinton, and W. Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–180, 2011.
10. T. Kim, A. Chang, L. Holland, and A. S. Pentland. Meeting mediator: enhancing group collaborationusing sociometric feedback. In *CSCW*, pages 457–466. ACM, 2008.
11. S. Lee, L. E. Rocha, F. Liljeros, and P. Holme. Exploiting temporal network structures of human interaction to effectively immunize populations. *PloS one*, 7(5):e36439, 2012.
12. A. Machens, F. Gesualdo, C. Rizzo, A. E. Tozzi, A. Barrat, and C. Cattuto. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC infectious diseases*, 13(1):185, 2013.
13. A. Madan, M. Cebrián, D. Lazer, and A. Pentland. Social sensing for epidemiological behavior change. In *Ubicomp*, pages 291–300, New York, NY, USA, 2010. ACM.
14. A. Madan, M. Cebrián, S. T. Moturu, K. Farrahi, and A. Pentland. Sensing the "health state" of a community. *IEEE Pervasive Computing*, 11(4):36–45, 2012.
15. A. Madan, K. Farrahi, D. Gatica-Perez, and A. Pentland. Pervasive sensing to model political opinions in face-to-face networks. In *Pervasive*, pages 214–231, Berlin, Heidelberg, 2011. Springer-Verlag.
16. R. C. Miller, H. Zhang, E. Gilbert, and E. Gerber. Pair research: matching people for collaboration, learning, and productivity. In *CSCW*, pages 1043–1048. ACM, 2014.
17. J.-K. Min, J. Wiese, J. I. Hong, and J. Zimmerman. Mining smartphone data to classify life-facets of social relationships. In *CSCW*, pages 285–294. ACM, 2013.
18. MIT Human Dynamics Lab. Social evolution dataset. http:realitycommons.media.mit.edu/socialevolution.html.
19. S. Phithakkitnukoon, T. W. Leong, Z. Smoreda, and P. Olivier. Weather effects on mobile social interactions: A case study of mobile phone users in lisbon, portugal. *PLoS ONE*, 7(10):e45745, 10 2012.
20. M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones. A high-resolution human contact network for infectious disease transmission. *PNAS*, 107(51):22020–22025, Dec. 2010.
21. K. Starbird and L. Palen. (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *CSCW*, pages 7–16. ACM, 2012.
22. J. Stehlé. *Human proximity networks: analysis, modeling and dynamical phenomena*. PhD thesis, Aix-Marseille Universit, 2012.
23. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Regis, J. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhems. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9(87), July 2011.
24. L. Sun, K. W. Axhausen, D.-H. Lee, and M. Cebrian. Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific Reports*, 4, 2014.
25. S. P. Wyche and R. E. Grinter. This is how we do it in my country: a study of computer-mediated family communication among kenyan migrants in the united states. In *CSCW*, pages 87–96. ACM, 2012.