Airborne Disease Propagation on Large Scale Social Contact Networks

Md Shahzamal
Macquarie University
Sydney, Australia
Md.Shahzamal@students.mq.edu.au

Raja Jurdak
Data61, CSIRO
Brisbane, Australia
Raja.Jurdak@data61.csiro.au

Reza Arablouei
Data61, CSIRO
Brisbane, Australia
Reza.Arablouei@data61.csiro.au

Minkyoung Kim
Data61, CSIRO
Brisbane, Australia
Minkyoung.Kim@data61.csiro.au

Kanchana Thilakarathna
Data61, CSIRO
Sydney, Australia
Kanchana.Thilakarathna@data61.csiro.au

Bernard Mans
Macquarie University
Sydney, Australia
Bernard.Mans@mq.edu.au

ABSTRACT

Social sensing has received growing interest in a broad range of applications from business to health care. The potential benefits of modeling infectious disease spread through geo-tagged social sensing data has recently been demonstrated, yet it has not considered contagion events that can occur even when co-located individuals are no longer in physical contact, such as for capturing the dynamics of airborne diseases. In this study, we exploit the location updates made by 0.6 million users of the Momo social networking application to characterize airborne disease dynamics. Airborne diseases can transmit through infectious particles exhaled by the infected individuals. We introduce the concept of same-place different-time (SPDT) transmission to capture the persistent effect of airborne particles in their likelihood to spread a disease. Because the survival duration of these infectious particles is dependent on environmental conditions, we investigate through large-scale simulations the effects of three parameters on SPDT-based disease diffusion: the air exchange rate in the proximity of infected individuals, the infectivity decay rates of pathogen particles, and the infection probability of inhaled particles. Our results confirm a complex interplay between the underlying contact network dynamics and these parameters, and highlight the predictive potential of social sensing for epidemic outbreaks.

CCS CONCEPTS

•**Computing methodologies** →*Modeling and simulation; Model development and analysis;*

KEYWORDS

Social sensing; contact network; infectious disease spreading; human movement

SocialSens 2017, Pittsburgh, PA USA

DOI: http://dx.doi.org/10.1145/3055601.3055604

ACM Reference format:
Md Shahzamal, Raja Jurdak, Reza Arablouei, Minkyoung Kim, Kanchana Thilakarathna, and Bernard Mans. 2017. Airborne Disease Propagation on Large Scale Social Contact Networks. In *Proceedings of The 2nd International Workshop on Social Sensing, Pittsburgh, PA USA, April 2017 (SocialSens 2017),* 6 pages.
DOI: http://dx.doi.org/10.1145/3055601.3055604

1 INTRODUCTION

Social sensing is a rapidly emerging research field where individuals play an important role in data collection. Social sensing includes data from social networking services, e.g, Facebook, Twitter and Google+, where user opinions, behaviors, and feelings can be collected through their online activities. It also includes the environmental and location information collected by sensors that are embedded in personal devices such as mobile phones. The potential of these data has been utilised in the fields of marketing, transportation, urban planning, and health care [1]. Social media data has been recently shown to help predict infectious disease spreading [3]. Geo-tagged data from social sensing have also high potential to facilitate modeling infectious disease spread based on human movement.

1.1 Related Work

A specific disease type that has not been extensively studied in the context of social sensing is airborne infectious diseases (AID). These diseases may be transmited from infected to susceptible individuals that are in close proximity. An AID-infected individual generates infectious pathogens as droplets through expiratory events like coughing and sneezing which scatter in their proximity. Thus, anybody within the reach of droplets may contract the disease by inhaling these pathogens [7]. The study of AID has shown that the generated droplets become light-weight particles after aerosalization that may suspend in the air for a long time [24]. Susceptible individuals may contract the disease if they visit an area where an infected individual has recently visited, even if the latter has left the location. In this case, there is no direct interaction (or contact) between the infected and susceptible individuals, but there is an indirect transmission path of pathogens (over time) through the visited location. The direct transmission path where both individuals are present in the proximity can be termed as the same-place

same-time (SPST) transmission. The indirect transmission path over time can be termed as the same-place different-time (SPDT) transmission. In SPDT, the probability of infection transmission decays with the elapsed time after the infected individual leaves the place as the generated pathogens are removed by air circulation and lose their infectivity over time [21]. While infectious disease spreading based on the SPST is widely studied [12], only few researchers have considered the indirect path in developing epidemic models [13, 17]. However, the influence of indirect transmission paths due to SPDT is not studied separately to understand its impact on epidemic spreading. Several studies have analysed the likelihood of infection for being within a small geographic region, i.e. locally [9, 20]. However, they did not analyse how this influences epidemic phenomena, i.e. globally. As AID are known to spread through indirect transmission paths, it is important to characterize the SPDT for an accurate modelling of AID spreading.

Several researchers [10, 16, 19] have applied empirical contact data to understand disease diffusion dynamics. These data are collected from small communities such as schools, hospitals, and conference venues. In these studies, the contact frequencies are high and the spreading dynamics are overestimated. They do not explain the disease spreading at larger spatial scales, such as where spreading occurs within neighbourhoods, cities, or countries. Contact information representing large populations can provide more realistic modelling. Some researchers [2, 22] used cell-phone call records to model human movements emulating large social networks. However, human movement can only be tracked at the base-station level (covering a wide area in many cases, on average within a 600 m error). Another approach of tracking the movements of people at global level is to use air passenger data as in [8]. This data provides a human movement model for a larger community but its spatial resolution of tracking individual locations is explicitly low and cannot capture individual interactions. In addition, researchers also investigated the potential of using geotagged Tweets for tracking the human movement as they provide finest locations information of individuals [11]. However, the update regularity varies significantly among users, leading to frequent data sparsity. In addition, users have to opt-in for the inclusion of geo-tags, limiting its ability to capture relevant contact for disease spreading. Furthermore, AID spread modeling requires information about co-location of the individuals that is not explicitly available in geotagged tweets.

1.2 Contributions and Structure

In this paper, we investigate the impact of SPDT transmission on disease spread through a data-driven study using social sensing. Data-driven studies are the first step to understand the dynamics of disease spreading on real contact networks. In previous data-driven studies, the interaction data have been limited to a community or a small number of individuals and thus cannot be used to analyze epidemic phenomena on a large-scale network. In our study, we exploit the location updates gathered by 0.6 million users of the Momo application in the entire city of Beijing. Momo is a location-based social networking application in which users locations are updated to the Momo server continuously as they move during the use of the app. The authors of [4] have collected about 356 million location updates for 6 million Momo users over 40 cities in the world over a 71 day period. We derive contact networks for the users in Beijing based on these location updates. For analysing AID spreading phenomena on this network, we first formulate a generic infection risk model to find the probability of infection by being in close proximity. The air exchange rate (AER), the rate of decrease in the infectiousness of airborne particles, and the infection probability for a pathogen are the crucial factors to determine the infection risk. We investigate how these parameters impact disease spread and identify the contribution of the SPDT transmission. Finally, we study the conditions for the disease to become a larger outbreak.

This paper is organised as follows. A generic formula for calculating infection risk based on likely exposure is introduced in Section II. We describe our methods for modeling disease dynamics in Section III. Simulation results and analysis are presented in the Section IV. In Section V, we discuss the outcomes and the limitations of our study while Section VI concludes the paper.

2 INFECTION RISK FORMULATION

To investigate airborne disease propagation via the SPDT propagation model, we first formulate the infection risk of a person that either has direct contact with an infected person or visits a location where an infected person has recently been present. Infection risk is modeled based on the availability of pathogen-containing particles in the considered proximity. Consider an infected individual coughs at a rate of f (cough/s) and each cough is of volume v (m^3). If the pathogen concentration in cough droplets is c (particle/m^3), the particle generation rate by an infected individual can be formulated as:

$$g = fvc \quad \text{(particle/s)} \tag{1}$$

The generated droplets evaporate quickly and their fate is determined by their sizes. Droplets below 5 μm in size dry down to droplet nuclei, which can be suspended in the air for a long time and can travel by air turbulence [14]. Droplets above 5 μm settled down to ground quickly. Around 20% of generated droplets can be in the range of 1 to 5 μm while droplets below 1 μm can be ignored as their pathogen concentration is almost negligible. Thus, the airborne particle generation rate by an infected individual is

$$\bar{g} = 0.2fvc \quad \text{(particle/s)} \tag{2}$$

In practice, these particles can be found 20 m away from the sources [7], and it has been shown to travel up to 100 m in the direction of airflow [9]. For simplicity, we consider that the particles are available homogeneously within a radius d m from the source of generation and up to a constant height h m. Thus, the generated particles will be distributed in the air of volume:

$$V = \pi h d^2 \quad (m^3) \tag{3}$$

The number of particles that are added to unit volume each second is:

$$n = \frac{\bar{g}}{V} = \frac{0.2fvc}{\pi h d^2} \quad \text{(particle/(sm}^3\text{))} \tag{4}$$

The particle concentration in the proximity of infected individuals continuously increases. However, in practice, there are environmental factors that reduce the particale concentration. As the air in a space is exchanged [6], particles are removed from the space and their concentration in the air decreases. The particle concentration

is also decreased due to the fact that the pathogen-carrying particles lose their ability to cause infection over time. The rate of decrease in particle infectivity depends on the humidity and temperature. We consider that at the initial stage of aerosolization, $(1-\rho)\%$ of particles lose their infectivity and further $\alpha\%$ of them lose their infectivity every second thereafter [21]. If the air exchange rate is r (1/s), the number of active particles in a unit volume after t seconds of their generation will be:

$$\begin{aligned} n_t &= \rho n (1-r)^t (1-\alpha)^t \\ &= \rho n b^t \end{aligned} \tag{5}$$

where we define $b = (1-r)(1-\alpha)$. If an infected individual stays t seconds in a location, the amount of active particles in a unit volume at time t is the accumulative sum of particles that survive until time t after their generation during time zero to time t. Thus, the number of active particles at time t exhaled by an infected individual can be computed as:

$$\begin{aligned} N_t &= \int_0^t \rho n b^{t-t'} dt' \\ &= \frac{\rho n}{\ln(b)} \left(1 - b^t\right) \end{aligned} \tag{6}$$

If a susceptible individual with a pulmonary rate of q stays within d meters of the infected individual from time t_0 to t_1, the exposure for direct contact E_d can be given by the following equation:

$$\begin{aligned} E_d &= q \int_{t_0}^{t_1} N_t dt \\ &= \frac{q\rho n}{\ln(b)} \int_{t_0}^{t_1} \left(1 - b^t\right) dt \\ &= \frac{q\rho n}{[\ln(b)]^2} \left[(t_1 - t_0)\ln(b) + b^{t_1} - b^{t_0}\right] \end{aligned} \tag{7}$$

If the infected individual leaves the location, the generated particles persist in the proximity which may infect susceptible individual. If the infected individual leaves the location at time t_1, the particle concentration decreases as follows:

$$N_t = N_{t_1} b^{t-t_1}$$

If a susceptible individual stays at the location from time t_2 to t_3, the exposure to the indirect contact E_i is :

$$\begin{aligned} E_i &= q \int_{t_2}^{t_3} N_{t_1} b^{t-t_1} dt \\ &= \frac{q N_{t_1}}{ln(b)} \left(b^{t_3-t_1} - b^{t_2-t_1}\right) \end{aligned} \tag{8}$$

The time t_2 is equal to time t_1 if the susceptible individual has arrived at the location before the infected individual left otherwise t_2 is greater than t_1. The total exposure of the susceptible individual due to both paths to the pathogen-carrying particles created by individual is

$$E = E_d + E_i \tag{9}$$

If the susceptible individual interacts with n infected individual in the same location, the total exposure is given by:

$$E_T = \sum_{k=0}^{n} E_k \tag{10}$$

where E_k is the total exposure due to the k^{th} infected individual. According to the Wells-Riley model [20], the probability of infection for the estimated exposure is given:

$$P = 1 - e^{-\gamma E_T} \tag{11}$$

where γ is the infectivity parameter for the particles.

3 METHODOLOGY

In this section, we present our methodology of evaluating the impact of SPDT transmission that leverages the derived probability of infection by (11), an epidemic dissemination model, and a contact network derived from a large-scale real world user mobility dataset.

3.1 Empirical Data

We utilize location information of the social discovery network *Momo*[1]. Momo enables the interaction of nearby users by sharing the user location with other users in the proximity. Every time a user launches the Momo app, a location update is forwarded to the Momo server. The server sends back the latest location updates of all users within the close proximity. We have previously collected these location updates via a set of network API provided by Momo servers in every 15 minutes for a period of 71 days (during May to October 2012) [23]. The dataset contains 356 million location updates from about 6 million Momo users around the world, but primarily in China. Each database entry includes coordinates of the location, time of update and user ID. In this study, we only consider the updates from Beijing, the city with the highest number of updates, for the period of 33 days from 17 September, 2012 to 19 October, 2012. This dataset has almost 56 million updates from 0.6 million users.

3.2 Epidemic Model

We adopt the susceptible-exposed-infected-recovered (SEIR) epidemic model to emulate airborne disease propagation over the dynamic contact network formed by the traces of the Momo users. Individuals are in one of four states any time and change the states as in a stochastic process through the four possible disease states, namely, susceptible (S), exposed (E), infectious (I) and recovered (R). If a susceptible individual in state S comes into contact with an infected individual, they will be exposed to infectious particles and may contract the disease. At the beginning of contraction, a susceptible individual enters the state E where they cannot infect others. The exposed individuals will be infectious (I) after a latency period of σ^{-1} days with a rate σ. The infected individual continues to produce infectious particles over its infection period of μ^{-1} days until they enter the recovered state R. Here, μ is the rate of recovering from the disease. It has been shown that σ^{-1} ranges between 1 to 2 days and μ^{-1} between 3 to 5 days for influenza-like diseases [5]. As the values can vary for each individual even for the same disease, we draw the parameters from a uniform distribution within the observed empirical ranges. We use the probability of infection (11) as the disease transmission probability.

[1] https://www.immomo.com

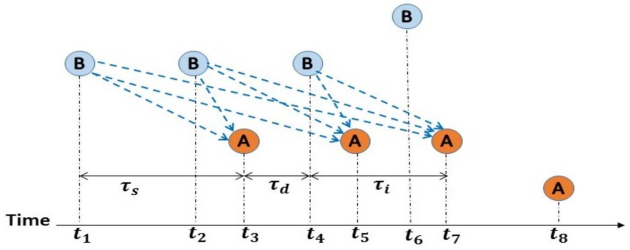


Figure 1: Link creation and its timing

3.3 Disease Transmission Links

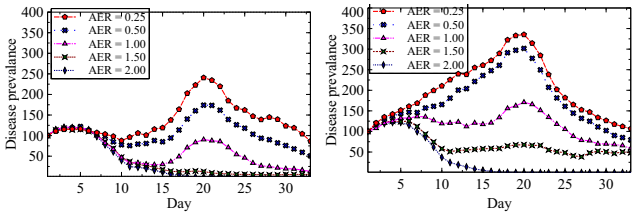
We utilize the location updates of Momo users to infer the potential infection dissemination links based on the proposed infection risk formulation and epidemic propagation model. We consider that there exists a disease transmission link e_{BA} between users A and B, when a susceptible user A makes a location update within δ seconds and distance d of a location update made by an infected user B. Here, δ is the maximum infectivity period of pathogen particles and d is the maximum distance infection particles travel.

Fig. 1 illustrates an example timing diagram where B circles represent the updates of the user B and the A circles present the updates of the user A. At time t_3, the user A is within d distance of the location B has updated at time t_1. Therefore, a link (dashed arrow from B to A) is established from time t_1. The subsequent updates of B are compared with the current update of A until the update time of B is greater than t_3 or the distance is greater than d. Thus, A has a link with B for the location updated at t_2. For the next update of A, the process repeats in checking all updates between t_1 and t_5 and the link is extended until the next update of B. For the update at t_7, A has a link with B at t_1, t_2 and t_4 but not at t_6 as their distance is beyond d for this update. Time t_4 is the departure time of B from the proximity of A. With the update at t_8, the link e_{BA} is broken and t_7 is the proximity departure time for A. The link lasts from t_1 to t_7 during which A inhales particles generated by B in the period of t_1 to t_4 through the direct path of duration (τ_d) and indirect path of duration (τ_i). The link creation process repeats for all updates of A. The user A can have several such links with the user B. Each link has three separate periods: initial pathogen store time τ_s when the infected individual exhales pathogens before the susceptible is around, the direct inhalation period τ_d when the susceptible and infected individuals are within proximity, and the indirect inhalation period τ_i when the infected has left the location but the susceptible continues inhaling the particles. Any period may have a zero value but τ_s and τ_i should be less than δ. We form two separate networks for our simulation 1) SPDT where links include all periods and 2) SPST where links exclude the indirect inhalation period τ_i.

4 SIMULATION AND ANALYSIS

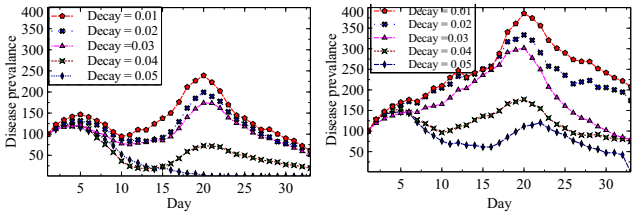
4.1 Simulation Setup

In this section, we use trace-driven simulations to explore the impact of the SPDT transmission on airborne disease spread across three parameters that underpin the diffusion process: air exchange rate (AER), infectivity decay rate, and infection probability by an inhaled pathogen. At the beginning of each simulation run, we



(a) SPST contact network **(b) SPDT contact network**

Figure 2: The influence of air exchange rate on disease spreading



(a) SPST contact network **(b) SPDT contact network**

Figure 3: The disease prevalence changes with various decay rates

randomly infect 100 users that have at least 5 contacts in the entire traces while other users are assumed to be susceptible. The simulation model of disease propagation determines which users are entering into the exposed state at the current day of simulation. To calculate the probability of a susceptible user becoming infected, we estimate the exposure of each user for all links created in the current day of simulation and accumulate over all the links. The accumulated exposure determines the probability of becoming infected, i.e, changing state to exposed state. We generate the incubation period from a uniform distribution in the range of 1 to 2 days and the infectious period from a uniform distribution in the range of 3 to 5 days. The next day's simulation starts with the exposed individuals and remaining infected individuals that have not yet recovered. We characterize the disease diffusion dynamics via the prevalence of disease I_p, which is the number of infected individuals in the current day, and the total infections I_t caused during the simulation period.

We set the distance d, within which a susceptible individual can inhale the infectious particles from infected individuals, to 20m [7]. We also assume that the particles can be scattered up to the ceiling height h of 3m. We set the other parameters as follows: cough frequency $f = 18$ (cough/hour) [15], total volume of the cough droplets $v = 6.7 \times 10^{-3}$ (ml) [25], pathogen concentration in the expiratory fluid $c = 3.7 \times 10^6$ (particles/m^3) [13], and pulmonary rate $q = 7.5$ (l/min) [25].

In our first experiment, we study the impact of the air exchange rate on the SPDT based diffusion dynamics. The AER is the proportion of air particles that are circulated out of an area per hour. If particles stay for longer, there is a higher chance of infecting others. The AER, r, in residential areas varies according to the type of the buildings and seasons. The authors of [18] reported that AER

Table 1: Result summary of all experiments

r	α	γ	$P_{mx}(di)$	$P_{mx}(d)$	I_{di}	I_d	C_i
0.25			335	240	1448	972	49%
0.50			301	174	1276	872	46%
1.00	0.03	0.05	181	120	819	422	94%
1.50			125	117	406	238	70%
2.00			122	117	194	169	14%
	0.01		385	226	2103	1070	96%
0.50	0.02	0.05	333	199	1641	950	72%
	0.04		176	123	833	476	75%
	0.05		146	121	287	218	31%
		0.09	531	343	2711	1489	82%
0.50	0.03	0.08	488	314	2499	1332	87%
		0.07	436	264	1934	1218	90%
		0.06	373	173	1695	1017	94%

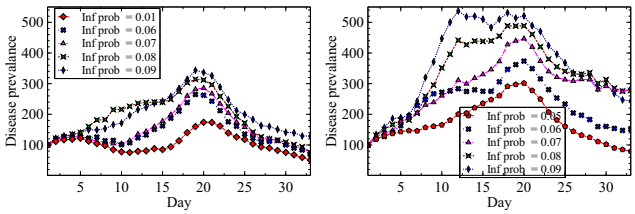
Table 2: Description of studied and model parameters

Symb	Description	Symb	Description
r (h^{-1})	Air exchange rate	I_{di}	Total infection in SPDT
α (min^{-1})	Infectivity decay rate	I_d	Total infection in SPST
γ	Infection prob by a pathogen	$P_{mx}(di)$	Maximum prevalence in SPDT
C_i (%)	Contribution of SPDT	$P_{mx}(d)$	Maximum prevalence in SPST

varies between 0.01 h^{-1} to 1.6 h^{-1}. The AER varies when opening doors, opening windows and using exhausts fans, or in open versus closed spaces. We assume the AER varies within the range [0.01, 4] h^{-1} with various medians. We run the simulations for medians 2, 1.5, 1, 0.5 and 0.25 h^{-1}, while setting the mean infectivity decay rate of droplet nuclei α to 0.03/min and it is within the range 0 to 0.05. Each simulation is repeated ten times and the results report the average of the simulation runs. In Table 1, we summarize the results for the maximum prevalence of disease and total infection occurred during the simulation period, and the increase in total number of infections for SPDT relative to SPST. Table 2 defines the notations used in the Table 1.

4.2 Results

Figure 2 plots the results comparing the SPST and SPDT spreading with varying AER. Overall, SPDT results in up to five times higher peak prevalence depending on the value of AER. For the high AER value of 2, where all the air circulates out of the area twice per hour, prevalence subsides once the initial infected individuals recover indicating that the high air circulation rates prevent any disease outbreaks. For smaller values of AER, SPDT leads to a significant increase in prevalence compared to SPST as the persistence of links in the contact network counteracts the effects of air circulation and the likelihood of individual encounters leading to infection increases. For AER=1.5, SPDT leads to a sustained prevalence of around 50 individuals, while SPST maintains fewer than 10 infected individuals. For higher AER, the rate of increase in prevalence with SPDT relative to SPST decreases. With smaller air circulation rates, the particles remain for longer in the air, but their infectivity also



(a) SPST contact network **(b) SPDT contact network**

Figure 4: The changes in disease dynamics with the infectiousness of pathogen

decreases over time, which limits the growth in the likelihood of infection.

In the above experiment, we have studied disease prevalence for a constant mean infectivity rate of 0.03 min^{-1}. To model the effects of temperature and humidity on the rate at which the particles lose their infectivity, we analyze the effect of infectivity decay rate with constant AER median of 0.5 h^{-1} that is calculated as the median for the residential areas in the Beijing [18]. We change the infectivity decay rate between 0.01 to 0.05 min^{-1}. Figure 3 shows the results for disease spreading on SPDT and SPST networks while varying IDR. SPDT leads to higher overall and peak prevalence (increase of 70-200%) over SPST for all decay rates. For the highest decay rate of 0.05, SPST fails to trigger any outbreak while SPDT causes an infection peak at around 21 days. For SPDT, slower decay rates in particle infectivity generally amplify this effect as links that survive longer due to indirect transmission paths are more relevant when the particles are infective for a longer period.

So far we assume that the probability of infection for an inhaled pathogen is 0.05 [13]. To model the impact of changing infection probability of a pathogen, we vary the infection probability between 0.05-0.09 with average conditions of AER 0.5 h^{-1} and infectivity decay rate 0.03 min^{-1}. The results are shown in Figure 4. As the infectivity of the pathogen increases, the outbreak peak grows in amplitude and width (duration). The peak prevalence is around 510 for SPDT with the infection probability 0.09 compared to around 330 for SPST. The number of total infected individuals for SPDT and SPST over the simulated period of 33 days are 2711 and 1489 respectively.

5 DISCUSSION

The main goal of our study has been to characterize the SPDT transmission process for AID. SPDT has strong dependency on individual movements and weather conditions. To understand the interplay of these parameters has required to study SPDT on real traces. In our study, we use social sensing traces from location-based social networking application Momo. Our simulations on a contact network of 0.6 million individuals with realistic disease parameters did not result in sustained epidemics. This is partly because the airborne path is one transmission route among other routes in AID and we consider only the infectious particle exhalation through coughs. The lack of sustained epidemics may also be due to the relatively small number of initially infected individuals (100 out of 0.6 million). However, the simulation results shows that SPDT transmission has a dominant role in AID spreading, signifcantly

increasing the total infections by up to 94% over SPST in the 33 day period. Specific scenarios modelling high air circulation or pathogen decay rates highlight the capacity of SPDT to lead to outbreaks when SPST fails to grow the number of infected individuals. While the SPST-based network shows slow progress of disease at moderate AER and infectivity decay rates, SPDT causes more infection and sensitive to the weather conditions due to its characteristics. There is also a complex interplay between the parameters we consider. For instance, smaller air circulation rates for SPDT starts to have diminishing effect at some point as pathogen infectivity decays over time.

Our approach has several limitations that we leave for future work. The Momo dataset allows us to construct a large-scale network for characterizing airborne disease spreading. However, some Momo users do not regularly use the application. Thus, a complete trajectory of an individual movement can not be tracked using this data as several locations of each user are not included in the updates. Secondly, the position estimates are often approximated by WiFi access points when users are indoors. In our infection risk model, we assume a homogeneous spatial distribution of infectious particles. An interesting direction for future work is to use a heterogeneous spatial-temporal distribution of particles.

The SPDT transmission model provides insights into the spreading dynamics of AID as well as other diseases whose transmission can be described by SPDT. The concept of SPDT can also be relevant to other diffusion networks such as modeling information propagation in ant colonies and content dissemination in online social networks by posting and reposting mechanisms.

6 CONCLUSION

In this study, we proposed the new concept of the same-place different-time to cover both direct and indirect disease transmission processes such as airborne infectious diseases. Based on the proposed SPDT framework, we modeled the spread of AID by incorporating realistic environmental factors such as air exchange rate and infectivity decay rate. Also, we conducted simulations using Momo data and showed that SPDT can cover indirect airborne disease transmission in all the experimental cases, in contrast to the same-place same-time. Future work will be to investigate the effects of network structures on disease spread with diverse experimental setups and extend the proposed model to study the countrywide disease outbreaks.

REFERENCES

[1] Charu C Aggarwal and Tarek Abdelzaher. 2013. Social sensing. In *Managing and mining sensor data*. Springer, 237–297.

[2] Linus Bengtsson, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet, and Renaud Piarroux. 2015. Using mobile phone data to predict the spatial spread of cholera. *Scientific reports* 5 (2015), 8923.

[3] Todd Bodnar and Marcel Salathé. 2013. Validating models for disease detection using twitter. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 699–702.

[4] Terence Chen, Mohamed Ali Kaafar, and Roksana Boreli. 2013. The Where and When of Finding New Friends: Analysis of a Location based Social Discovery Network.. In *ICWSM*.

[5] Gerardo Chowell, Santiago Echevarría-Zuno, Cecile Viboud, Lone Simonsen, James Tamerius, Mark A Miller, and Víctor H Borja-Aburto. 2011. Characterizing the epidemiology of the 2009 influenza A/H1N1 pandemic in Mexico. *PLoS Med* 8, 5 (2011), e1000436.

[6] Liuliu Du, Stuart Batterman, Christopher Godwin, Jo-Yu Chin, Edith Parker, Michael Breen, Wilma Brakefield, Thomas Robins, and Toby Lewis. 2012. Air change rates and interzonal flows in residences, and the need for multi-zone models for exposure and health analyses. *International journal of environmental research and public health* 9, 12 (2012), 4639–4661.

[7] Aaron Fernstrom and Michael Goldblatt. 2013. Aerobiology and its role in the transmission of infectious diseases. *Journal of pathogens* 2013 (2013).

[8] Lauren Gardner and Sahotra Sarkar. 2013. A global airport-based risk model for the spread of dengue infection via the air transport network. *PloS one* 8, 8 (2013), e72129.

[9] Zhuyang Han, Wenguo Weng, Quanyi Huang, and Shaobo Zhong. 2014. A Risk Estimation Method for Airborne Infectious Diseases Based on Aerosol Transmission in Indoor Environment. In *Proceedings of the World Congress on Engineering*, Vol. 2.

[10] Chunlin Huang, Xingwu Liu, Shiwei Sun, Shuai Cheng Li, Minghua Deng, Guangxue He, Haicang Zhang, Chao Wang, Yang Zhou, Yanlin Zhao, and others. 2016. Insights into the transmission of respiratory infectious diseases through empirical human contact networks. *Scientific Reports* 6 (2016).

[11] Raja Jurdak, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, and David Newth. 2015. Understanding human mobility from Twitter. *PloS one* 10, 7 (2015), e0131469.

[12] Matt J Keeling and Pejman Rohani. 2008. *Modeling infectious diseases in humans and animals*. Princeton University Press.

[13] Sheng Li, Joseph NS Eisenberg, Ian H Spicknall, and James S Koopman. 2009. Dynamics and control of infections transmitted from person to person through the environment. *American journal of epidemiology* 170, 2 (2009), 257–265.

[14] William G Lindsley, Francoise M Blachere, Robert E Thewlis, Abhishek Vishnu, Kristina A Davis, Gang Cao, Jan E Palmer, Karen E Clark, Melanie A Fisher, Rashida Khakoo, and others. 2010. Measurements of airborne influenza virus in aerosol particles from human coughs. *PloS one* 5, 11 (2010), e15100.

[15] Robert G Loudon and Linda C Brown. 1967. Cough Frequency in Patients with Respiratory Disease 1, 2. *American Review of Respiratory Disease* 96, 6 (1967), 1137–1143.

[16] Anna Machens, Francesco Gesualdo, Caterina Rizzo, Alberto E Tozzi, Alain Barrat, and Ciro Cattuto. 2013. 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 (2013), 185.

[17] Thomas O Richardson and Thomas E Gorochowski. 2015. Beyond contact-based transmission networks: the role of spatial coincidence. *Journal of The Royal Society Interface* 12, 111 (2015), 20150705.

[18] Shanshan Shi, Chen Chen, and Bin Zhao. 2015. Air infiltration rate distributions of residences in Beijing. *Building and Environment* 92 (2015), 528–537.

[19] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, and others. 2011. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC medicine* 9, 1 (2011), 87.

[20] GN Sze To and CYH Chao. 2010. Review and comparison between the Wells–Riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor Air* 20, 1 (2010), 2–16.

[21] GN Sze To, MP Wan, CYH Chao, F Wei, SCT Yu, and JKC Kwan. 2008. A methodology for estimating airborne virus exposures in indoor environments using the spatial distribution of expiratory aerosols and virus viability characteristics. *Indoor air* 18, 5 (2008), 425–438.

[22] Andrew J Tatem, Zhuojie Huang, Clothilde Narib, Udayan Kumar, Deepika Kandula, Deepa K Pindolia, David L Smith, Justin M Cohen, Bonita Graupe, Petrina Uusiku, and others. 2014. Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria journal* 13, 1 (2014), 52.

[23] Kanchana Thilakarathna, Suranga Seneviratne, Kamal Gupta, Mohamed Ali Kaafar, and Aruna Seneviratne. 2016. A deep dive into location-based communities in social discovery networks. *Computer Communications* (2016).

[24] Jianjian Wei and Yuguo Li. 2015. Enhanced spread of expiratory droplets by turbulence in a cough jet. *Building and Environment* 93 (2015), 86–96.

[25] S Yin, GN Sze-To, and Christopher YH Chao. 2012. Retrospective analysis of multi-drug resistant tuberculosis outbreak during a flight using computational fluid dynamics and infection risk assessment. *Building and environment* 47 (2012), 50–57.