

RESEARCH ARTICLE

Inferring a District-Based Hierarchical Structure of Social Contacts from Census Data

Zhiwen Yu^{1*}, Jiming Liu^{2*}, Xianjun Zhu¹

1 School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China, **2** Department of Computing, Hong Kong Baptist University, Hong Kong

* zhwyu@scut.edu.cn (ZY), jiming@comp.hkbu.edu.hk (JL)



OPEN ACCESS

Citation: Yu Z, Liu J, Zhu X (2015) Inferring a District-Based Hierarchical Structure of Social Contacts from Census Data. PLoS ONE 10(2): e0118085. doi:10.1371/journal.pone.0118085

Academic Editor: Frederic Amblard, University Toulouse 1 Capitole, FRANCE

Received: January 8, 2014

Accepted: January 4, 2015

Published: February 13, 2015

Copyright: © 2015 Yu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors are grateful for the constructive advice received from the anonymous reviewers of this paper. The work described in this paper was partially funded by a grant from the Hong Kong Scholars Program (Project No. XJ2012015) and supported by grants from the National Natural Science Foundation of China (NSFC) (Project Nos. 61273363, 61379033, 61472145), the NSFC-Guangdong Joint Fund (Project No. U1035004), a grant from the Fundamental Research Funds for the Central Universities (Project No. 2014G0007), a grant from China Postdoctoral Science Foundation (Project No. 2013M540655), a grant the Guangdong Natural Science Funds for Distinguished Young Scholars

Abstract

Researchers have recently paid attention to social contact patterns among individuals due to their useful applications in such areas as epidemic evaluation and control, public health decisions, chronic disease research and social network research. Although some studies have estimated social contact patterns from social networks and surveys, few have considered how to infer the hierarchical structure of social contacts directly from census data. In this paper, we focus on inferring an individual's social contact patterns from detailed census data, and generate various types of social contact patterns such as hierarchical-district-structure-based, cross-district and age-district-based patterns. We evaluate newly generated contact patterns derived from detailed 2011 Hong Kong census data by incorporating them into a model and simulation of the 2009 Hong Kong H1N1 epidemic. We then compare the newly generated social contact patterns with the mixing patterns that are often used in the literature, and draw the following conclusions. First, the generation of social contact patterns based on a hierarchical district structure allows for simulations at different district levels. Second, the newly generated social contact patterns reflect individuals social contacts. Third, the newly generated social contact patterns improve the accuracy of the SEIR-based epidemic model.

Introduction

Researchers have recently paid attention to capturing individuals social contacts and social network structures due to their useful applications in the social sciences, ecology, health care, communications, economic sociology [2] [1] [9] [5] [3] [10] and especially epidemic models for infectious disease transmission [17] [16] [15] [4]. Although many studies have examined the relationships among individuals and social network structures, few [20] have considered how to infer the social contact patterns of a hierarchical social structure at the population level directly from census data using a statistical approach.

Every kind of social demographical profile contains a hierarchical structure of social contacts and cross-district contact patterns. such as working, traffic and school patterns, that affect

(project No. S2013050014677), a grant from the key lab of cloud computing and big data in Guangzhou (Project No. SITGZ[2013]268-6), A grant from Key Enterprises and Innovation Organizations in Nanshan District in Shenzhen (Project No. KC2013ZDZJ0007A), and Hong Kong Baptist University (Project No. RGC/HKBU211212). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

the dynamics of an epidemic. For example, most individuals work in the Central and Western and Wan Chai districts in Hong Kong, but live in the Eastern and Southern districts. If an epidemic were to break out in the Central and Western district, the individuals in the Eastern and Southern districts would be affected immediately. In addition, age- and district-specific patterns have different effects on the spread of disease. Gaining a more complete understanding of epidemic modeling and control requires inferring the hierarchical structure of social contacts from census data.

There are many types of social contacts, such as social media, virtual world and face-to-face contacts. We pay specific attention to face-to-face social contact patterns, as the public health decision department often uses them to capture the dynamic of influenza transmission in the epidemic model. Face-to-face contact suggests physical contact between at least two persons. It includes a mixture of any types of physical contact, such as contact with relatives at home, on public transportation and in the workplace. The approaches applied to identify social contact patterns can be divided into three categories according to different data sources. The first approach involves mining social contact patterns from social networks. For example, Eubank et al. [18] modeled physical contact patterns using the social networks generated by urban traffic simulations built according to actual census, land-use and population-mobility data. Rohani et al. [39] studied the effect of age-structured contact patterns on an epidemic in a detailed contact network. Zelner et al. [47] investigated the effect of social relationships in social networks on the transmission of diarrheal disease. Volz et al. [38] explored the effects of heterogeneous and clustered contact patterns in social networks on the spreading of infectious disease. Lee et al. [30] exploited the structure of social contacts in temporal social networks based on empirical data. Ndeffo Mbah et al. [37] studied the effects of imitation behavior and contact heterogeneity in social contact networks on vaccination coverage. Huerta-Quintanilla et al. [25] modeled social contacts based on social networks in three elementary school case studies. Szell et al. [43] examined the social contact patterns in a large online social network. Son et al. [41] studied a massive online multiplayer social network. Rolls et al. [40] modeled social contacts and designed a contact network for people to explore the spread of hepatitis C. Ventresca et al. [44] constructed a social network based on individuals social contacts as determined by agent-based simulations, and used it to evaluate mitigation strategies. Del Valle et al. [11] proposed a new method of inferring the mixing contact patterns between age groups in social networks. In general, the approaches in the first category can be applied to determine individuals relationships and social contact structures based on their social networks. However, this kind of approach is limited in that it is difficult to construct a social network on a large scale, especially when face-to-face contacts are considered. In addition, the sparse properties of a high dimensional matrix that captures the social contacts from a social network complicates the process of social contact pattern mining.

The approaches in the second category adopt a survey method to collect social contact data from countries such as the United States, France, Belgium, Taiwan, Vietnam and countries in Europe. These data are used to identify social contact patterns. For example, Cauchemez et al. [7] explored the effect of school closure on social contact patterns in relation to influenza transmission based on survey data from France. They also investigated the effect of household contacts on household transmission based on confirmed cases of the 2009 H1N1 virus infection in the United States [8]. Wallinga et al. [45] estimated age-specific transmission parameters using self-reported social contact data. Eames et al. [12] explored the social contact patterns of schoolchildren using survey data from a self-completed questionnaire-based study. Kretzschmar et al. [27] conducted a representative survey involving 7,290 respondents in 8 European countries to capture social contact patterns. They also used social contact patterns to evaluate reproduction numbers based on serological data from five European countries [28]. Glasser

et al. [21] considered how to incorporate age-structured contact mixing patterns into the epidemic model using household and cross-sectional serological survey data. Fu et al. [19] investigated contact patterns based on diary-approach surveys conducted in Taiwan. Hens et al. [23] identified social contact patterns based on the data from a two-day population survey conducted in Belgium. Horby et al. [24] studied social contact patterns based on data from a household-based social contact diary in rural Vietnam. Kucharski et al. [29] studied the age patterns of immunity based on mixing data taken from social contact surveys. Melegaro et al. [33] analyzed types of contact based on contact survey data from five European countries. Willem et al. [46] collected data from a social contact survey in Belgium combined with local weather data to investigate the effect of weather conditions and social contact patterns on seasonal influenza transmission during epidemics. Eames et al. [13] conducted a prospective survey to investigate the social mixing patterns of schoolchildren. They also explored social contact patterns according to the results of an Internet-based social contact survey conducted for a cohort of participants over 9,000 times, and examined their simulation effects during the 2009 H1N1v influenza epidemic [14]. Stehle et al. [42] explored the face-to-face contact patterns of children and teachers based on a survey conducted in a French school. Bolton et al. [6] studied the effect of contact definitions on social contact patterns during disease transmission based on survey data from a convenience sample of 65 adults. Mikolajczyk et al. [35] conducted a series of surveys involving university students in Bielefeld from 2003 to 2006 to attain the patterns of direct contact between individuals. These second-category approaches are more suitable for mining social contact patterns on a small scale, such as a community or a small district. They are limited when the survey scale becomes too large, and this increases the human resource and financial costs involved in collecting data.

The approaches in the third category adopt a statistical method to infer social contact patterns directly from the census data published by governments or other authorities. For example, Iozzi et al. [26] identified contact patterns using a synthetic matrix inferred from Italian time-use and routine socio-demographic data. Fumanelli et al. [20] adopted a statistical approach to derive social contact structures from demographic data, which they used to analyze influenza epidemics. In general, the approaches in this category are suitable for mining social contact patterns on a very large scale. In addition, social contact patterns can be easily and correctly inferred from the census data provided by governments and other authorities.

In this paper, we propose a new third-category statistical approach to model a city's population; identify the social contact patterns of the hierarchical social structure at the population level, which are inferred from the detailed census data; and determine hierarchical-district-structure-based, cross-district and age-district-based social contact patterns. We also evaluate the hierarchical structure of social contacts in a simulation of the 2009 Hong Kong H1N1 epidemic to improve the accuracy of the epidemic model. Our work differs from Fumanelli's [20] in that we provide a more general approach taking into consideration not only the age groups, but also the districts with a hierarchical structure, during the formation of contact matrix.

In summary, the proposed approach has several properties. First, it considers social contacts in a hierarchical district structure. The social contact patterns of districts at different levels form a hierarchical structure that is useful for epidemic control. Second, it considers cross-district contact matrices that account for the contact patterns between individuals in different districts at the same level, which reflects the spatial information during the spread of infectious disease. Third, it considers contact matrices that correspond to both districts and age groups simultaneously. Fourth, it adopts the detailed 2011 Hong Kong census data (released in 2012) to infer the social contact structure. To our knowledge, this paper marks the first time that the detailed 2011 Hong Kong census data has been used to identify social contact patterns.

Methods

Census data

To estimate the contact matrices quantitatively, we collected the detailed social census data from Hong Kongs Census and Statistics Department. The most recent census data released by the government was collected in March 2011. The objective of the census is to determine characteristics and trends in the population based on the Hong Kong districts that help the government make decisions. During the 2011 census, 1 in 10 households completed the long questionnaire. Interviews were performed to guarantee the confidence of the provided information. Completed questionnaires were processed and converted into all kinds of statistical tables to ensure the privacy of the respondents information. The statistical tables in the detailed 2011 Hong Kong census data related to our work include the tables for domestic households by district council, household composition and size, domestic households by the sex and age group of the head of the household and by household composition, population by district council constituency area and age, population by sex and district council, age group and economic activity status, persons attending full-time courses by district council, place of study by age group, school type by age, working population by sex, place of work by age group and district council, work type by age and working population by occupation and industry, among others. All of the tables were processed to calculate the contact matrices among different sub-populations.

Preprocessing of census data

The process of generating synthetic populations in our work is more complicated than that in [20], as we account for the synthetic population based on the hierarchical district structure as shown in Fig. 1, cross-district synthetic population and synthetic population corresponding to both district and age. Fig. 2 illustrates the hierarchical district structure in Hong Kong. Hong Kong (level 0) contains three big districts (level 1), Hong Kong Island, the Kowloon Peninsula and the New Territories, and each of which contains several smaller districts (level 2). Using the detailed social census data, we calculate the number of households in each district, the household compositions, the age of each household member, the activity status of each member, the place of study for persons attending full-time courses, school type in relation to age, working population in relation to place of work, work type in relation to age and working population in relation to occupation. For example, we use the census data related to households by district, household composition and size, households by the sex and age group of the head of the household, and population by district and age to partition individuals into households in

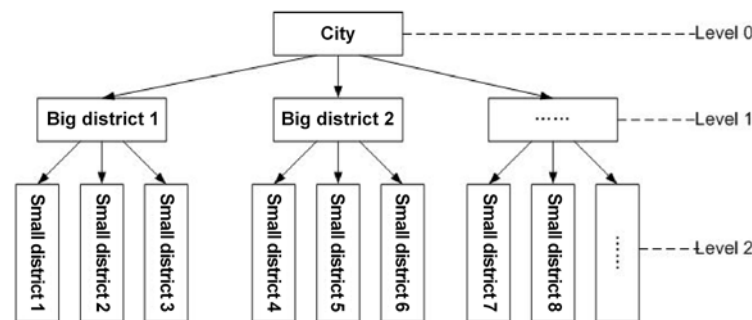


Fig 1. Hierarchical district structure in a city.

doi:10.1371/journal.pone.0118085.g001

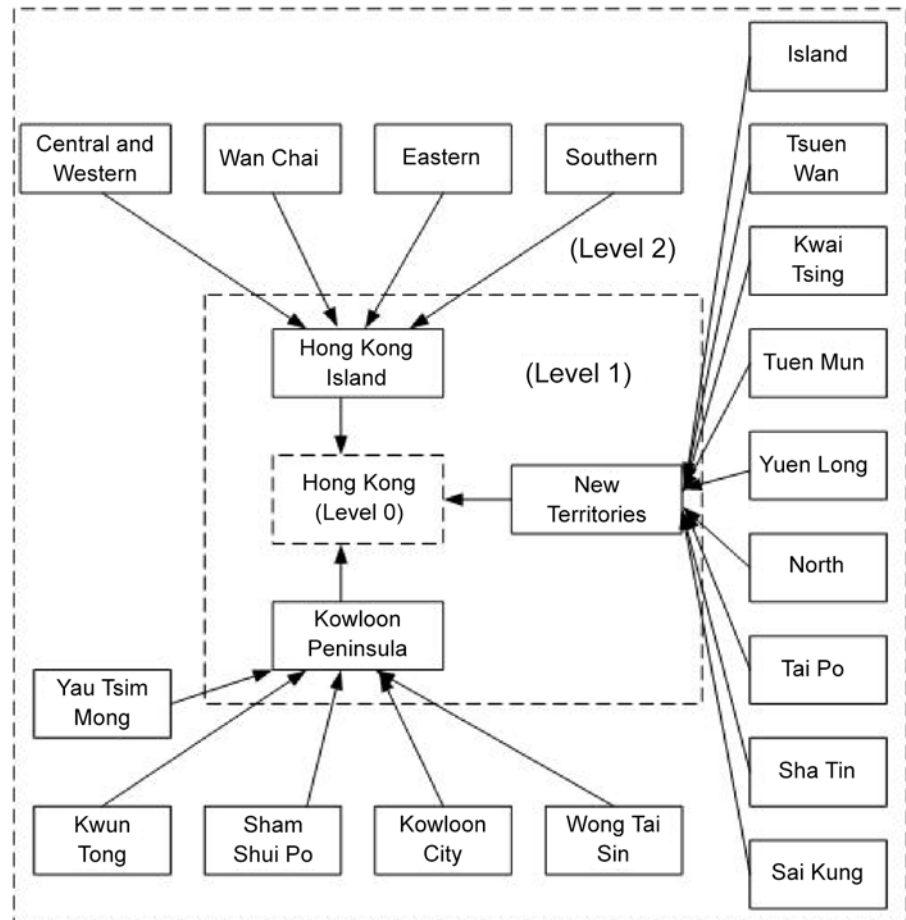


Fig 2. Hierarchical district structure in Hong Kong.

doi:10.1371/journal.pone.0118085.g002

different districts. The census data related to working population by district, place of work in relation to age group, persons attending full-time courses by district and places of study in relation to age group can be used to identify individuals who work or study in the same district or across different districts. Statistical data related to population by district, age group and economic activity status allow us to distribute individuals based on activities such as studying, working and staying at home. We use additional census data to generate synthetic populations, especially cross-district synthetic populations.

One of the challenges in the estimation process is how to calculate the cross-district synthetic populations at different levels of the hierarchical district structure. The proposed approach should calculate the average probability of contact between individuals in the i -th and j -th districts (where $i \neq j$).

Fig. 3 gives an overview of the simulation process for the individuals at different levels of the hierarchical district structure. The proposed approach first estimates the districts, their age structures, the households in each district and the members of each household. It then considers the activities of the household members, including staying at home, staying in the community, studying and working. If the individual stays at home or in the community, the district of the individuals activity is within the same district, which means that the individual belongs to the household or community contact type. If the individual goes to school, the individual and

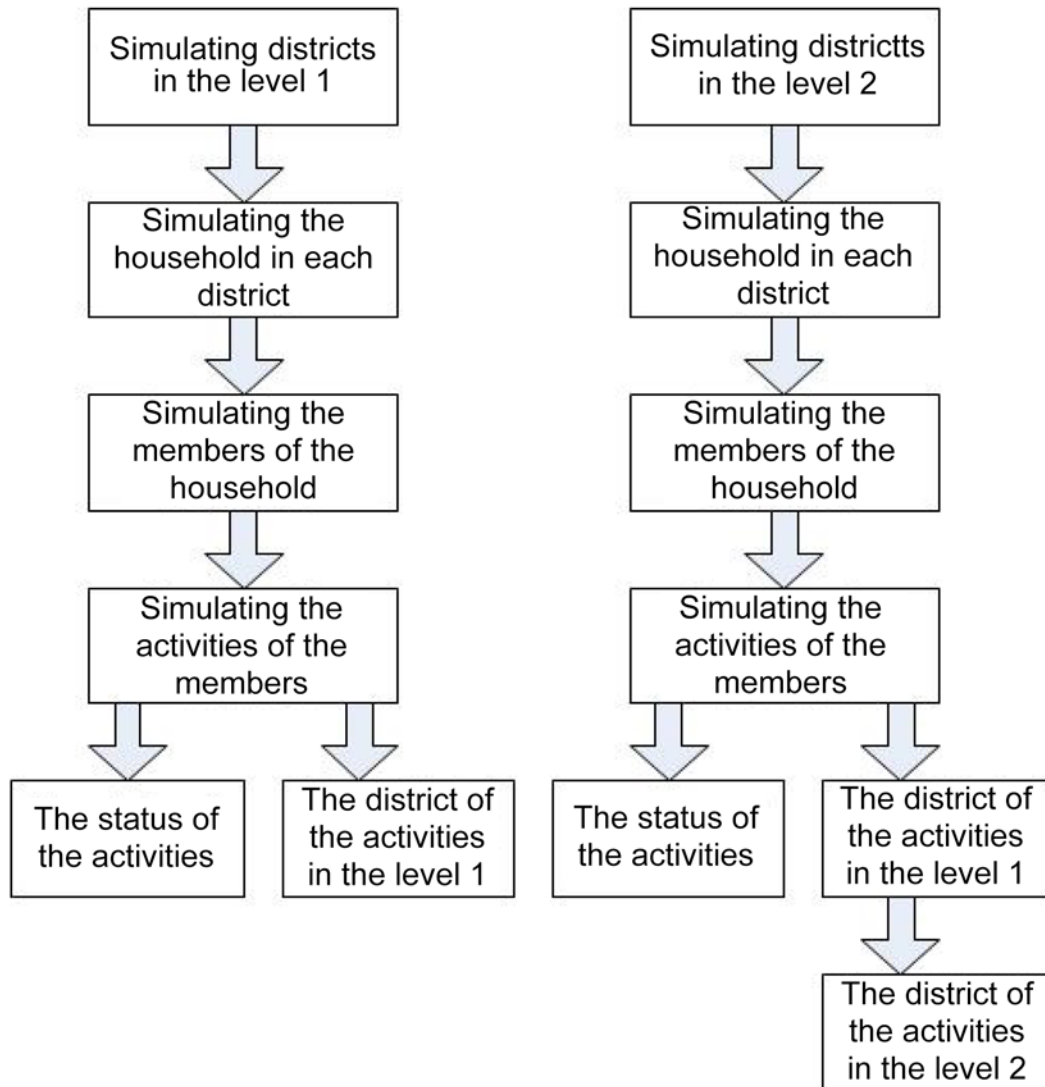


Fig 3. Simulation process in the hierarchical district structure.

doi:10.1371/journal.pone.0118085.g003

school belong to either the same district or different districts. The probability $p(I)$ of the individual I distributed to each case is calculated as follows:

$$p(I) = \frac{P_{ij}}{\sum_j P_{ij}} \tag{1}$$

where $i, j \in \{1, \dots, r\}$, r is the number of districts and $\sum_j P_{ij}$ denotes the total number of individuals in the i -th district who attend school in the j -th district. If $i = j$, P_{ij} denotes the number of individuals who attend school in the same district. Otherwise, P_{ij} denotes the number of individuals who attend school in different districts. Some of the social census data provide only the number of individuals who study in different districts at level 1. Three cases must be considered during the simulation to simulate an individual studying in different districts at level 2. In the first case, the individual and school belong to the same district at level 2. This problem can be

solved using the preceding approach. In the second case, the individual and school belong to different districts at level 2, but belong to the same district at level 1. In this case, the proposed approach first determines the level 2 districts that share the same district at level 1, and makes use of the radiation model [22] to distribute the individuals to the corresponding districts at level 2. The traditional radiation model unfortunately does not consider populations consisting of several age groups. The proposed modified radiation model related to age groups is defined as follows:

$$T_{ij}^h = T_i^h \frac{O_i^h O_j^h}{(O_i^h + S_{ij}^h)(O_j^h + O_j^h + S_{ij}^h)} \tag{2}$$

where O_i^h and O_j^h denote the study opportunities in relation to the h -th age group provided by the i -th and j -th districts, respectively; S_{ij}^h denotes the total study opportunities for the h -th age group in the circle of radius Γ_{ij} centered at the i -th district (excluding the source and destination study opportunities); T_i^h denotes the number of students in relation to the h -th age group who study outside the i -th district and T_{ij}^h denotes the number of students in the h -th age group who live in the i -th district but study in the j -th district, as predicted by the radiation model. In the third case, the individual and school belong to different districts at levels 1 and 2 simultaneously. The proposed approach first assigns the individual to the i -th district at level 1 as shown on the right side of Fig. 3, and then distributes the individual to the district at level 2 based on the radiation model [22].

If an individual is working, the proposed approach simulates two scenarios in which the individuals workplace is in the same district and a different district, respectively. The simulation process is similar to that used for individuals studying at school. In summary, the preceding simulation process captures the hierarchical-district-structure-based and cross-district contact patterns.

Calculation of contact matrix

Two important factors are closely related to the social structure of the population: district and age. District reflects the spatial properties of the population, and age captures its temporal properties. For example, workers in districts with more business centers tend to spend more time with their business partners at their companies, and teenagers prefer to spend more time with their classmates at school. This means that social contacts differ by district and age group. To capture the social and demographic structure of the population, heterogeneous contact patterns among individuals should be identified corresponding to district and age groups. We adopt contact matrices W to represent heterogeneous contact patterns, the entries $w_{(i-1)*a+h, (j-1)*a+k}$ (rewritten as w_{hk}^{ij}) of which denote the average contact probability between an individual belonging to the h -th age group in the i -th district and an individual belonging to the k -th age group in the j -th district (where $h, k \in \{1, \dots, a\}$, a is the number of age groups, $i, j \in \{1, \dots, r\}$ and r is the number of districts). Two individuals may share contact if they share the same physical environment [20], such as the household, school, workplace or community. We adopt the same assumption as [20], which states that the mixing in the finest unit, such as the single household or school, is homogeneous. Individuals have four types of social contact patterns: those that occur in the household, school, workplace and community. As a result, district and age group contact matrices are calculated based on four types of contact matrices W^e ($e \in \{1, 2, 3, 4\}$) related to the household W^1 , school W^2 , workplace W^3 and community W^4 , respectively. It is reasonable to assume that social contacts in the household and community occur within the same district, and that those related to the school and workplace occur

within the same district or across districts. Instead of considering the workplace, we divide working individuals by their occupation and industry, and view those individuals who are in the same occupation and industry as a homogeneous unit.

The proposed approach first calculates the contact probability $(U_{hk}^{ij})^{edb}$ per individual belonging to the h -th age group in the i -th district and the k -th age group in the j -th district (where the superscript $(\cdot)^{edb}$ denotes the social contacts appearing in the b -th unit in the d -th district in relation to e -th contact type; $b \in \{1, \dots, B\}$; B is the number of total units; $d \in \{1, \dots, r\}$; r is the number of districts in the corresponding level and e is the contact type) as follows:

$$(U_{hk}^{ij})^{edb} = \begin{cases} \frac{(P_k^j)^{edb} - \delta_{hk}^{ij}}{P^{edb} - 1} & \text{if } P^{edb} > 1, \\ 0 & \text{if otherwise,} \end{cases} \tag{3}$$

$$P^{edb} = \sum_{i=1}^r \sum_{h=1}^a (P_h^i)^{edb} \tag{4}$$

where $(P_h^i)^{edb}$ and $(P_k^j)^{edb}$ denote the total number of individuals belonging to the h -th age group in the i -th district and the k -th age group in the j -th district, respectively. a is the number of age groups. δ_{hk}^{ij} is a Kronecker delta formula and is defined as follows:

$$\delta_{hk}^{ij} = \begin{cases} 1 & \text{if } i = j \wedge h = k, \\ 0 & \text{if otherwise,} \end{cases} \tag{5}$$

This means that when the number of individuals is greater than 1, the preceding formula works. Otherwise, $(U_{hk}^{ij})^{edb} = 0$.

The total contact frequency matrix C^{ed} in the d -th district with the e -th contact type is calculated, the entries $(c_{hk}^{ij})^{ed}$ of which are as follows:

$$(c_{hk}^{ij})^{ed} = \sum_{b=1}^B (U_{hk}^{ij})^{edb} (P_h^i)^{edb} \tag{6}$$

We then calculate the average contact frequency matrix F^{ed} in the d -th district with the e -th contact type. Its entries $(f_{hk}^{ij})^{ed}$ are as follows:

$$(f_{hk}^{ij})^{ed} = \frac{(c_{hk}^{ij})^{ed}}{(N_h^i)^{ed} (N_k^j)^{ed}} \tag{7}$$

$$(N_h^i)^{ed} = \sum_{b=1}^B (P_h^i)^{edb} \quad \text{if } P^{edb} > 1 \tag{8}$$

$$(N_k^j)^{ed} = \sum_{b=1}^B (P_k^j)^{edb} \quad \text{if } P^{edb} > 1 \tag{9}$$

where $(N_h^i)^{ed}$ denotes the total number of individuals belonging to the h -th age group in the i -th district whose effective social contacts occur in the d -th district. $(N_k^j)^{ed}$ can be defined in a similar way.

We calculate the contact matrix W^{ed} in the d -th district with the e -th contact type, and its entries $(w_{hk}^{ij})^{ed}$ are as follows:

$$(w_{hk}^{ij})^{ed} = (\phi_{hk}^{ij})^{ed} (f_{hk}^{ij})^{ed} \tag{10}$$

$$(\phi_{hk}^{ij})^{ed} = \frac{(N_h^i)^{ed} (N_k^j)^{ed}}{P_h^d P_k^d} \tag{11}$$

where P_h^d is the total number of individuals belonging to the h -th age group who live in the d -th district and who participate in activities in the d -th district, which can be derived from the preceding simulation process. P_k^d can be defined in a similar way.

If we consider only one district, the contact matrix W^d in the d -th district is calculated. Its entries $(w_{hk}^{ij})^d$ are as follows:

$$(w_{hk}^{ij})^d = \sum_e \frac{\omega_e (w_{hk}^{ij})^{ed}}{\sum_e (w_{hk}^{ij})^{ed}} \tag{12}$$

where the values of the parameters ω_e are similar to [20].

If we account for all of the districts at the same level, the contact matrix W^e with the e -th contact type is calculated, and its matrices $(w_{hk}^{ij})^e$ are determined as follows:

$$(w_{hk}^{ij})^e = \sum_{d=1}^r \zeta_{hk}^d (w_{hk}^{ij})^{ed} \tag{13}$$

$$\zeta_{hk}^d = \frac{P_h^d P_k^d}{P_h P_k} \tag{14}$$

where P_h denotes the total number of individuals in the h -th age group and P_k denotes the total number of individuals in the k -th age group.

Finally, we calculate the contact matrix W in relation to all of the contact types. Its entries w_{hk}^{ij} are as follows:

$$w_{hk}^{ij} = \sum_e \frac{\omega_e (w_{hk}^{ij})^e}{\sum_e (w_{hk}^{ij})^e} \tag{15}$$

In summary, we use the proposed approach to calculate the average contact probability matrix W in terms of all of the contact types level by level in the hierarchical district structure, and form a set of hierarchical-district-structure-based contact matrices.

Contact matrix permutation

To determine the patterns in the contact matrix, we define the permutation operator π to re-sort the rows and columns. We consider the first l (where l is a parameter specified by the user) at the largest contact probability for each row or column, and define an objective function ζ_i for the i -th row or column, which is the sum of the first l for the i -th row or column. We then use the permutation operator π to re-sort the rows and columns in the contact matrix in

descending order according to the values of the objective function ζ as follows:

$$\begin{pmatrix} 1 & , & 2 & , & 3 & , & \dots & , & n \\ \pi(1) & , & \pi(2) & , & \pi(3) & , & \dots & , & \pi(n) \end{pmatrix} \tag{16}$$

The lines and columns may be rearranged because they are all independent, as all of the districts are independent. This means that there are no relationships between any two neighboring lines or columns. As a result, the rearrangement process based on the contact probabilities provides a chance to capture the patterns among the lines and columns.

Contact matrix validation

To evaluate the effectiveness of the newly generated contact matrix, we incorporate the contact matrix into the SEIR model [34], which is formulated as follows:

$$\frac{dS_h}{dt} = -\chi_h \cdot S_h \tag{17}$$

$$\frac{dE_h}{dt} = -\varphi_h \cdot E_h + \chi_h \cdot S_h \tag{18}$$

$$\frac{dI_h}{dt} = -\psi_h \cdot I_h + \varphi_h \cdot E_h \tag{19}$$

$$\frac{dR_h}{dt} = \psi_h \cdot I_h \tag{20}$$

where S_h, E_h, I_h and R_h denote the compartments of the SEIR model in the age-specified structure, including the susceptible, exposed, infectious and recovered compartments, respectively. $h \in \{1, \dots, a\}$ denotes the h -th age group. a is the number of age groups. φ_h is the incubation rate. ψ_h is the recovery rate. χ_h is the infection risk, which is defined as follows:

$$\chi_h = \frac{1}{a} \cdot \left(\sum_{k=1}^a w_{hk} \cdot \frac{I_k}{P_k} \right) \cdot \frac{S_h}{P_h} \cdot \beta_h \tag{21}$$

where w_{hk} denotes the value of the entry in the contact matrix W in relation to all of the contact types, and β_h denotes the infection risk. The infection risk χ_h consists of two parts: the risk of infectious contacts determined by w_{hk} and the generic infection vulnerability β_h .

Results

Simulation with the detailed 2011 Hong Kong census data

We evaluate the simulation process of the proposed approach using the detailed 2011 Hong Kong census data.

[Fig. 4](#) compares the real and simulated population data in relation to the age structures in the different level 1 districts. The numbers of simulated individuals in each age group in each district are close to the number of real individuals, indicating that the simulation process is successful.

[Fig. 5](#) compares the real and simulated social activity data in the different districts at level 1. In [Fig. 5](#), the number of simulated individuals who stay at home, study or work in correspondence with the number of householders, students and workers, respectively, is close to those of real individuals, indicating that the simulation of individuals social activities satisfies the

Comparison of age structures of real and simulated data in 3 districts in the level 1

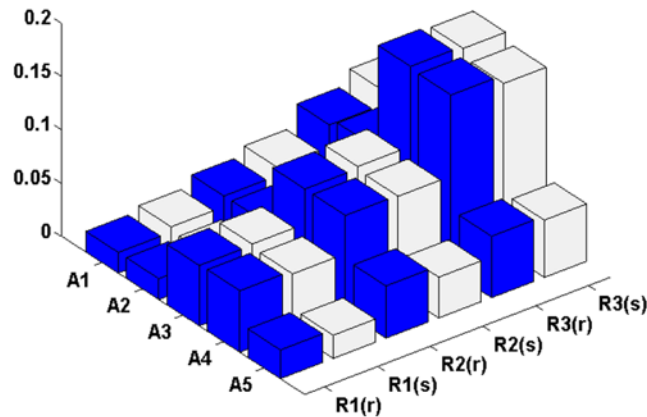


Fig 4. Comparison of the populations of simulated and real data in the districts at level 1, including Hong Kong Island (R1), Kowloon Peninsula (R2), and the New Territories (R3) (where R1(r), R2(r) and R3(r) denote the real data in the three districts, respectively, and R1(s), R2(s) and R3(s) denote the simulated data in the three districts, respectively).

doi:10.1371/journal.pone.0118085.g004

Comparison of social activities of real and simulated data in 3 districts in the level 1

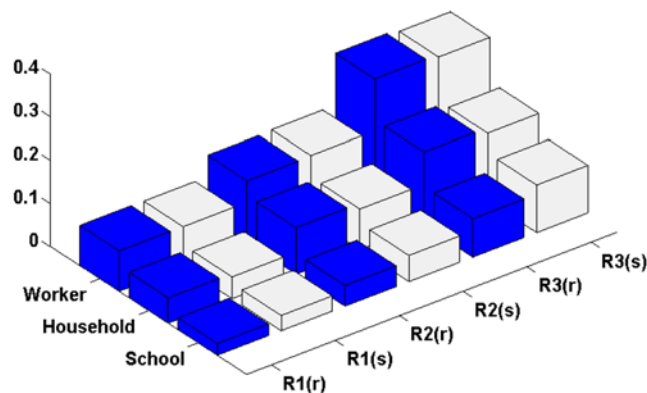


Fig 5. Comparison of the social activities of simulated and real data in the districts at level 1 (where R1(r), R2(r) and R3(r) denote the real data in the three districts, respectively, and R1(s), R2(s) and R3(s) denote the simulated data in the three districts, respectively).

doi:10.1371/journal.pone.0118085.g005

decision-making requirement. Figs. 6 and 7 compare the ratios of students and workers in relation to the simulated and real age group data in the hierarchical district structure, and show them to be similar. From Figs. 6 and 7, we can conclude that the proposed approach successfully simulates the students and workers in correspondence with the age groups and districts.

We also use a linear regression model with zero intercept between the real and simulated data in Figs. 4, 5, 6 and 7. The values of the coefficient of determination R^2 are 0.9039, 0.9677,

Comparison of the ratios of students w.r.t age groups in 3 districts in the level 1

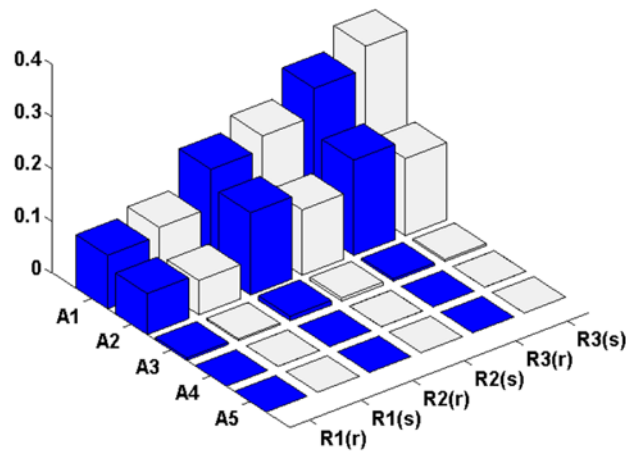


Fig 6. Comparison of the ratios of students in relation to age group for the simulated and real data in the level 1 districts.

doi:10.1371/journal.pone.0118085.g006

Comparison of the ratios of workers w.r.t age groups in 3 districts in the level 1

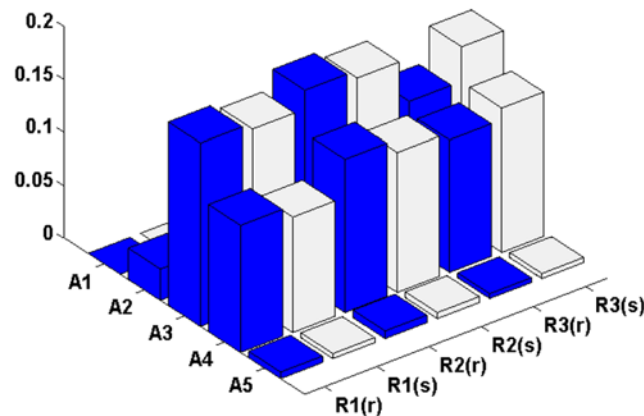


Fig 7. Comparison of the ratios of workers in relation to age group for the simulated and real data in the level 1 districts.

doi:10.1371/journal.pone.0118085.g007

0.9633 and 0.9781, respectively. This indicates that the difference between the estimated and real data is very small.

Hierarchical-district-structure-based contact matrices

Fig. 8 illustrates the hierarchical-district-structure-based contact matrices in Hong Kong, which include district and age-district matrices. Fig. 8 presents the following findings. First, as

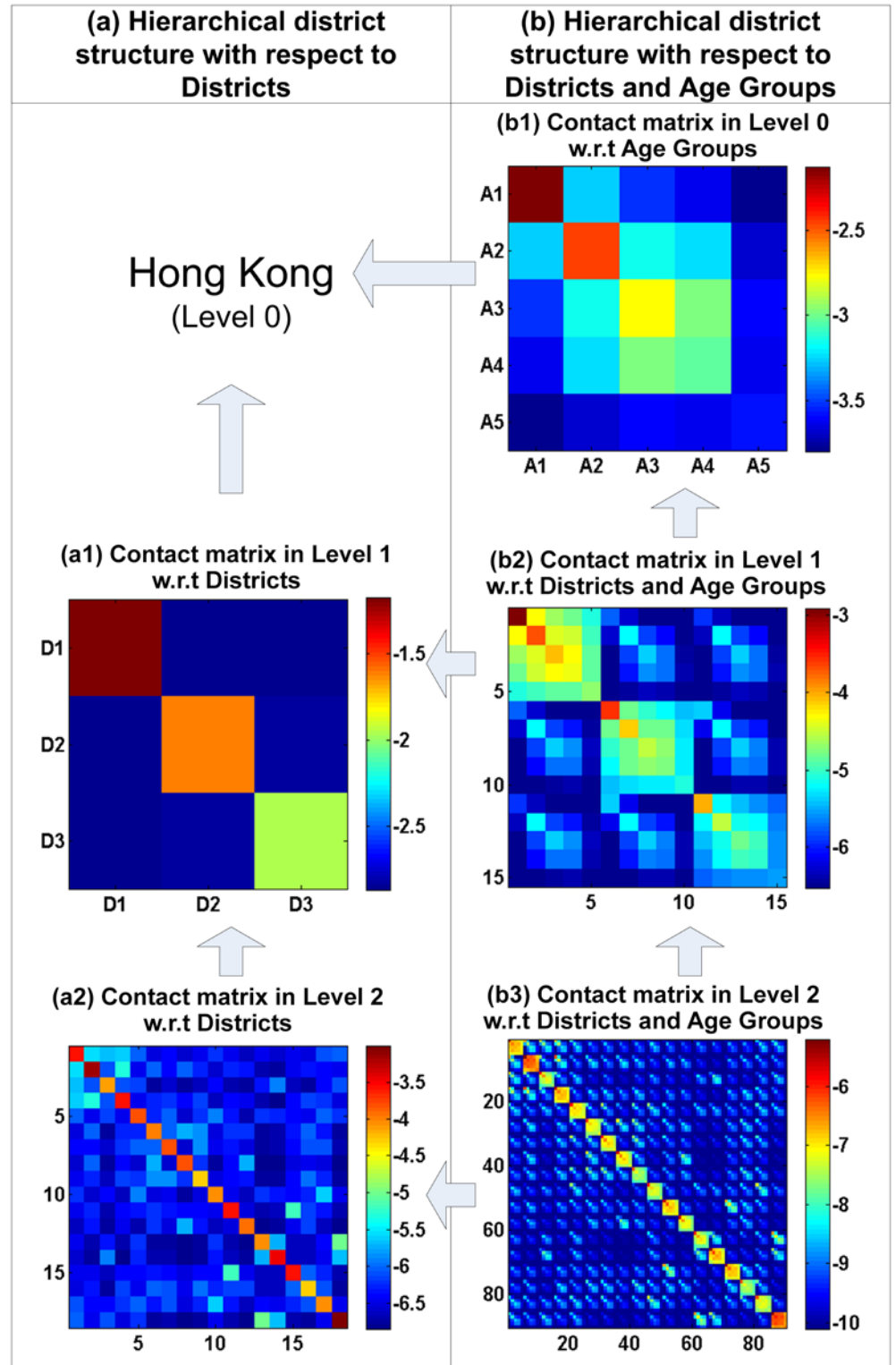


Fig 8. Hierarchical-district-structure-based contact matrices for Hong Kong. (All of the values are shown on a logarithmic scale. (b1) A1, A2, A3, A4 and A5 denote the five age groups (5–14, 15–24, 25–44, 45–64 and 65+, respectively. (b2) The numbers 1, 2, ..., 15 denote the combinations of 3 districts and 5 age groups in the level 1, including R1A1, R1A2, ..., R3A5, respectively. (b3) The numbers 1, 2, ..., 90 denote the combinations of the 18 districts (Central and Western, Wan Chai, Eastern, Southern, Yau Tsim Mong,

Sham Shui Po, Kowloon City, Wong Tai Sin, Kwun Tong, Kwai Tsing, Tsuen Wan, Tuen Mun, Yuen Long, North, Tai Po, Sha Tin, Sai Kung, Islands) and 5 age groups in the level 2, including R1A1, R1A2, . . . , R18A5, respectively.)

doi:10.1371/journal.pone.0118085.g008

illustrated in [Figs. 8\(a1\) and \(a2\)](#), the values of the entries on the diagonal of the district contact matrices are significantly higher than those not on the diagonal. Although the contact matrices in [Figs. 8\(a1\) and \(a2\)](#) belong to different levels in the hierarchical district structure, they possess the same pattern. This indicates that most of the social contacts appear in the same district, possibly because the social contacts made among individuals in the household and community often occur in the same district. Second, as shown in [Fig. 8\(b1\)](#), the values of the entries in the contact matrices among children and teenagers A1(5–14) and A2(15–24) are higher than those in the contact matrices among other age groups. The same contact patterns can be discovered in [Figs. 8\(b2\) and \(b3\)](#). Children and teenagers (A1(5–14) and A2(15–24)) in schools have a number of classmates, which may generate many social contacts compared with adults (A3 and A4) and the elderly (A5(65+)). Third, the contact matrix entries that correspond with the same age group in the same district have higher values than those related to different age groups. For example, the contact matrix entries that correspond with the A1, A2 and A3 age groups have the largest values in [Fig. 8\(b1\)](#), and the entries related to the A1, A2 and A3 age groups have larger values than those related to different age groups in the same district, as shown in [Figs. 8\(b2\) and \(b3\)](#). This means that individuals in social networks prefer to contact individuals within the same age group. For instance, infants make friends in kindergarten, schoolmates prefer to take part in activities with their classmates and workers prefer to be in contact with colleagues. Fourth, as illustrated in [Figs. 8\(b1\) and \(b2\)](#), the contact probabilities between the individuals in the A4 and A3 age groups are larger than those between the individuals in the A4 group and other groups. The adults in the A4 group have powerful social networks and participate in more social activities with the adults in the A3 group compared with the other age groups. Fifth, as shown in [Fig. 8\(b1\)](#), the values of the contact matrix entries that correspond with the A5 age group are small. Elderly individuals participate in few social activities and only make contact with their sons or daughters in the A3 and A4 groups. In summary, the contact matrices related to age groups and districts at different levels in the hierarchical district structure share similar patterns, reflecting Hong Kongs social and demographic situation.

Cross-district contact matrices

To determine the contact patterns from the cross-district contact matrices, we set the values of all of the entries in the same district to zero. [Fig. 9](#) shows the cross-district contact matrices in Hong Kong, including those that correspond with the district and age-district groups.

Several observations can be made. First, as shown in [Fig. 8\(a1\)](#), the contact probabilities between the individuals in the neighboring districts are large, such as between the individuals in R2 (Kowloon Peninsula) and R3 (the New Territories) and between the individuals in R1 (Hong Kong Island) and R2. There are several possible reasons for this. (i) Most of the individuals who work in the Kowloon Peninsula live in the New Territories, which has a lower cost of living. In addition, the Kowloon Peninsula has more job opportunities and good schools that attract students. (ii) Hong Kong Island is a business, cultural and government center. The individuals located there and in the Kowloon Peninsula have more contacts due to the higher prominence of business, cultural and political affairs.

Second, the values of the entries in the upper left corner of the cross-district contact matrix in [Fig. 8\(a2\)](#) are larger than those in the remaining area of the matrix. The upper left corner of the matrix corresponds with four districts at level 2 of the hierarchical district structure,

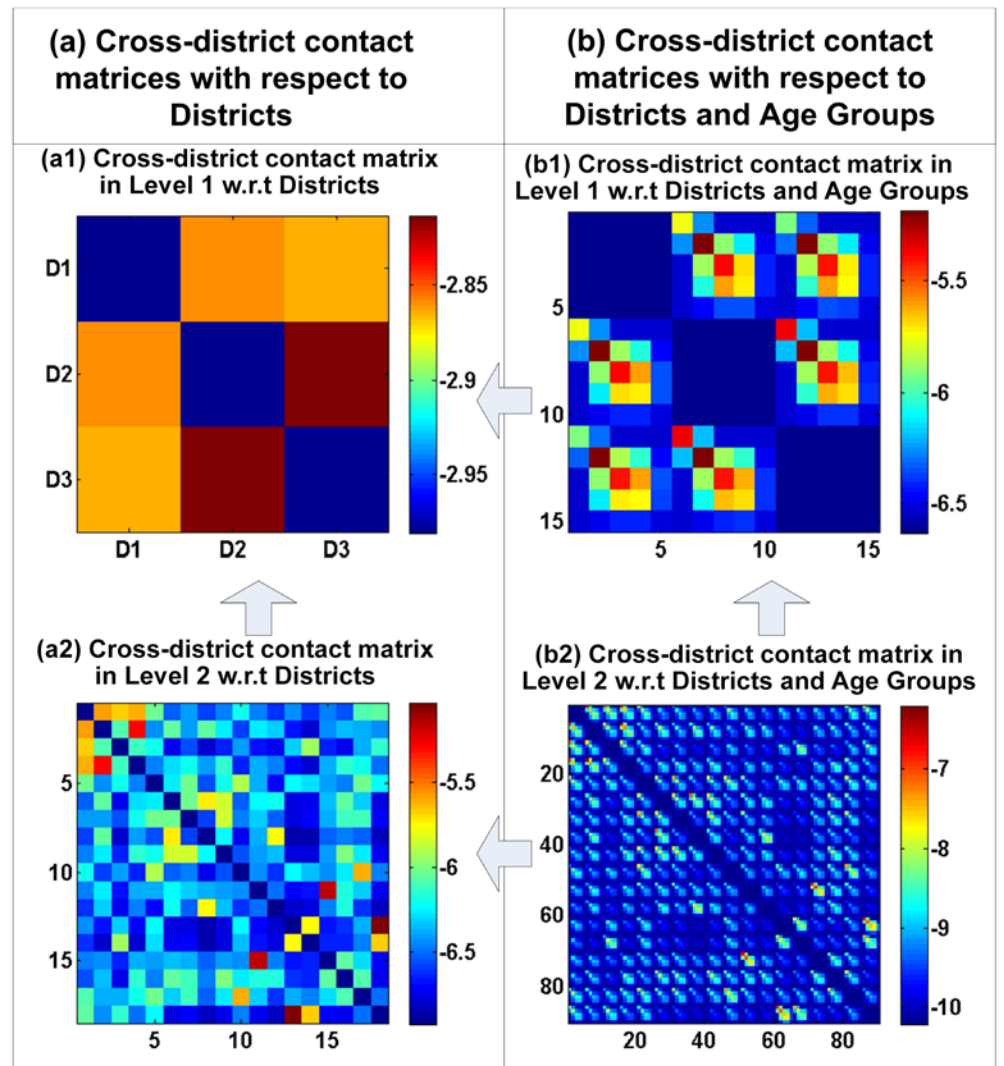


Fig 9. Cross-district contact matrices for Hong Kong. (All of the values are shown in logarithmic scale. (b1) The numbers 1, 2, . . . , 15 denote the combinations of 3 districts and 5 age groups in the level 1. (b2) The numbers 1, 2, . . . , 90 denote the combinations of the 18 districts and 5 age groups in the level 2.)

doi:10.1371/journal.pone.0118085.g009

including the Central and Western, Wan Chai, Eastern, and Southern districts. There are several possible reasons for the high contact probabilities among the individuals in these districts. (i) The Central and Western district is Hong Kong's central business and government district. It includes the headquarters of financial services corporations, government headquarters and the consulates of many countries. The Wan Chai district is the heart of Hong Kong. It includes government buildings, art centers, hotels, shopping malls and large exhibition and conference centers. It is reasonable to assume that there is a high probability of contact between the individuals in the Central and Western district and the Wan Chai district. The individuals in the companies or governments establish many social contacts to deal with business procedures between these two districts. (ii) The Eastern district has the second highest population among the 18 districts, including many public and large private housing estates. The Southern district has many residential areas. Most of the residents in the Eastern and Southern districts work in the Central

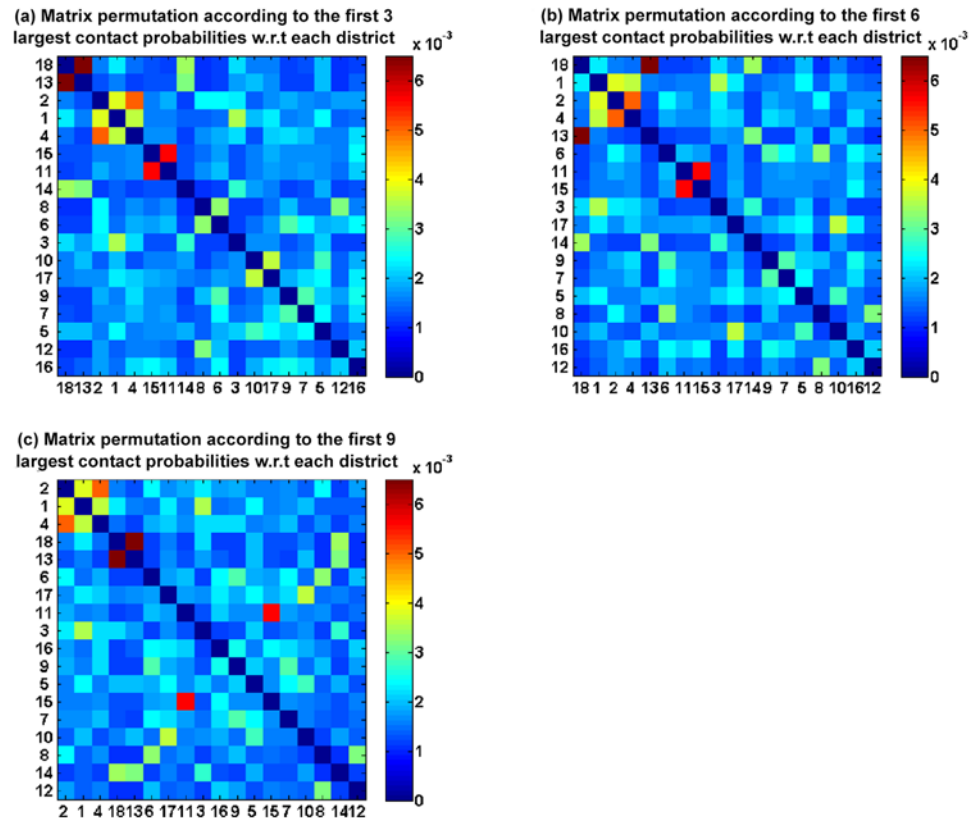


Fig 10. Cross-district contact matrix permutation for each district at level 2. (The numbers 1, 2, ..., 18 in (a), (b) and (c) denote the level 2 districts, including Central and Western, Wan Chai, Eastern, Southern, Yau Tsim Mong, Sham Shui Po, Kowloon City, Wong Tai Sin, Kwun Tong, Kwai Tsing, Tsuen Wan, Tuen Mun, Yuen Long, North, Tai Po, Sha Tin, Sai Kung and Hong Kong Island.)

doi:10.1371/journal.pone.0118085.g010

and Western district and Wan Chai district, leading to a high probability of contact among these individuals. The most interesting observation is that the contact probabilities between the individuals in the Eastern and Southern districts are lower than those among the individuals in other combinations of the four districts. Both the Eastern and Southern districts are residential areas, encouraging a low contact frequency between the individuals located there.

We perform a matrix permutation on the cross-district contact matrix at level 2 in relation to the first l largest contact probabilities for each district. Fig 10 illustrates the cross-district contact matrix permutation at level 2 according to the first 3, 6 and 9 largest contact probabilities for each district, respectively. The Yau Tsim Mong, Sham Shui Po, Kowloon City, Yuen Long and Sha Tin districts, which have index values of 5, 6, 7, 13 and 16, respectively, are always sorted in the first 10 rows as shown in Figs. 10(a), (b) and (c). There are several possible reasons for this. (i) The Yau Tsim Mong district is located across Victoria Harbour, which is the business center in the Knowloon City district at level 1. A lot of visitors shop in the Yau Tsim Mong district, which creates a lot of social contacts. (ii) The Sham Shui Po district is a commercial, industrial and transportation hub of the territory, and generates a number of social contacts. (iii) The Kowloon City district includes four universities, or half of the universities in Hong Kong. Individuals gather together in this district to study, which increases the probability of contact. (iv) The Yuen Long district has the youngest population in Hong Kong. The average ages for males and females in this district are 32 and 27, respectively, which means

that most of the individuals belong to the A3 age group. (v) The Sha Tin district is a living center in Hong Kong, and includes many residents and large residential areas. Most of the residents in the Sha Tin district work in other districts, which leads to a high contact probability.

We further interpret the implications of the cross-district contact matrices for epidemic modeling or control. Assuming an epidemic were to break out in the Central and Western district, the neighboring Wan Chai, Eastern and Southern districts would be affected immediately, as shown in [Fig. 10 \(a\)](#), due to their high contact probabilities with the Central and Western district. The Yau Tsim Mong district would also be affected, as it is closer to and has a higher probability of contact with the Hong Kong Island district compared with other districts. The Sham Shui Po and Kowloon City districts would then be affected in turn. The epidemic would eventually spread to all of the districts in Hong Kong. One way to prevent the spreading of disease in Hong Kong is to provide more vaccinations in the Yau Tsim Mong district, which would be useful in isolating the Hong Kong Island and Kowloon City districts.

In summary, the newly generated cross-district contact matrices not only reflect the situation of the commuter population in Hong Kong and the corresponding social factors, but also reflect the social, economic and cultural situations of the different districts.

Comparison with a traditional contact matrix for Hong Kong

The traditional contact matrices for Hong Kong consider only age groups. Hence, we compare the newly generated contact matrix, which corresponds with age groups, with the traditional contact matrix for Hong Kong [\[31\]](#) [\[32\]](#), which is derived according to the approach in [\[36\]](#). [Figs. 11\(a\) and \(b\)](#) compare the visualization results of the newly generated and traditional contact matrices, respectively. The two contact matrices share similar patterns. For example, the contact probability in the same age group is higher than those in different age groups in both contact matrices. The order of contact probability in terms of its value in the same age group is also similar between the two matrices. This indicates that the newly generated and traditional contact matrices are similar. We adopt a linear regression model for the traditional and newly generated contact matrices from our model with zero intercept, as illustrated in [Fig. 11\(c\)](#) (where X denotes the new contact matrix and Y denotes the traditional contact matrix). The coefficient of determination R^2 is 0.72. The two contact matrices are very similar, but exhibit a small difference that may be caused by the different data-driven models. Our newly generated contact matrix is inferred from the 2011 Hong Kong census data, and the traditional contact matrix is derived from survey data from eight European countries in 2008 [\[36\]](#). For example, our proposed data-driven model views individuals in the same occupation and industry as a homogeneous unit instead of accounting for the workplace, which lead to a change in contact probabilities among the individuals in the A3 and A4 age groups.

We use the SEIR model to simulate the spread of the infectious virus during the 2009 Hong Kong H1N1 swine flu epidemic. When the first infection case was confirmed by Hong Kongs Center for Health Protection (CHP) on May 2, 2009, the public health department adopted several intervention strategies, such as vaccination, the segregation of infection cases and school closures to control the dynamics of the epidemic. To reasonably validate the proposed epidemic model, we compare the infection case data accumulated between May 23 and July 27, 2009 with the simulation results. The SEIR model adopts the same parameter values adopted in [\[20\]](#) to simulate the dynamics of the H1N1 epidemic in Hong Kong in terms of the size of the accumulated infectious population. We consider five age groups during the simulation, including A1 (5–14), A2 (15–24), A3 (25–44), A4 (45–64) and A5 (65+).

The circle line in [Fig. 12](#) shows the number of infectious cases confirmed in the laboratory by the CHP in practice during the 2009 Hong Kong H1N1 swine flu epidemic from May 23 to

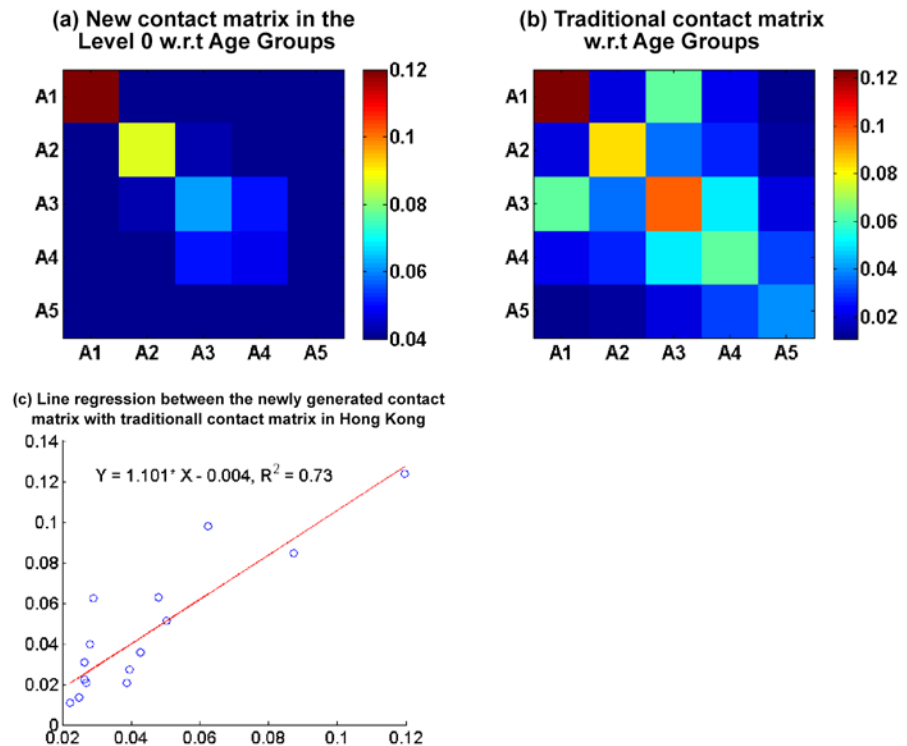


Fig 11. Linear regression between the newly generated and the traditional contact matrices for Hong Kong. (X denotes the values in the traditional contact matrix and Y denotes the values in the newly generated contact matrix.)

doi:10.1371/journal.pone.0118085.g011

July 27, 2009 without any intervention strategies taken. The red dotted line in Fig. 12 illustrates the simulation results obtained by the SEIR model in combination with the new contact matrix. The solid line in Fig. 12 shows the simulation results obtained by the SEIR model with the contact matrix obtained in [20].

Three of the curves in Fig. 12 are close to one another, which means that the simulation results obtained by the SEIR model in combination with the new contact matrix or the contact matrix used in [31] [32] are consistent with the observed dynamics of the infectious virus in practice, especially the dotted line in Fig. 12. The new contact matrix fully reflects the social contacts among the different age groups in Hong Kong, which further improves the simulation results of the SEIR model and gives a more accurate reflection of the spread of the infectious virus. In general, the simulation results of the SEIR model in combination with the new contact matrix qualify for epidemic dynamic simulation, and are suitable for exploring the effect of different vaccine distribution strategies.

We also compare the age- and district-specific disease attack rates to explore whether the generated hierarchical contact structure can reproduce the observed dynamics of a spreading disease. The age- and district-specific disease attack rates are defined as follows:

$$\Lambda_h = \frac{I_h}{\sum_k I_k} \tag{22}$$

$$\Lambda_i = \frac{I_i}{\sum_j I_j} \tag{23}$$

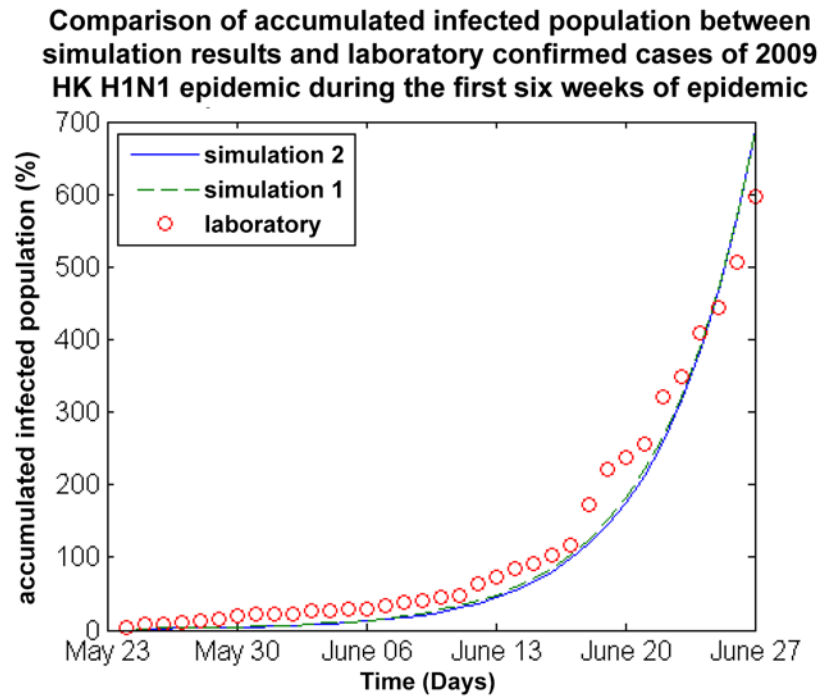


Fig 12. Comparison of the accumulated infected population based on simulation results and laboratory-confirmed infection cases during the first six weeks of the 2009 Hong Kong H1N1 epidemic. (Simulation 1 is based on the newly generated contact matrix and simulation 2 is based on the traditional Hong Kong contact matrix.)

doi:10.1371/journal.pone.0118085.g012

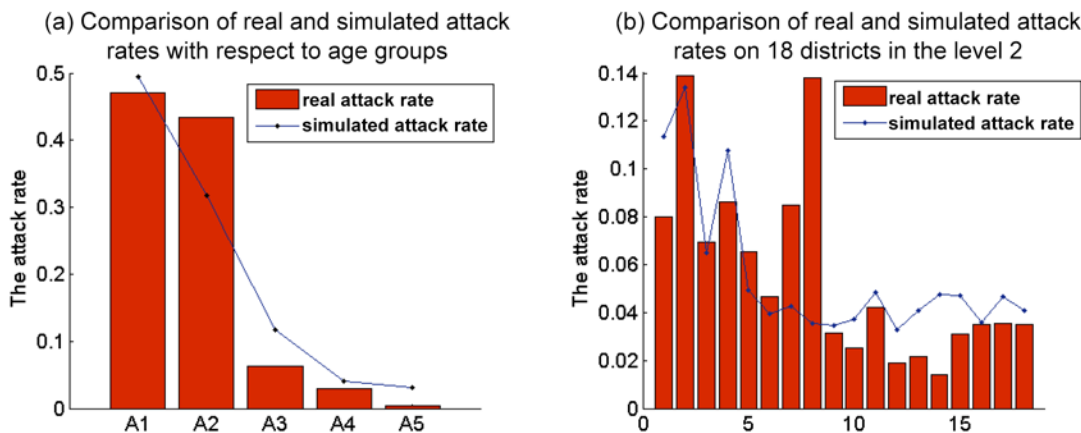


Fig 13. Comparison of the age- and district-specific disease attack rates in Hong Kong.

doi:10.1371/journal.pone.0118085.g013

where $h, k \in \{1, \dots, a\}$, a is the number of age groups, $i, j \in 1, \dots, r$ and r is the number of districts. Fig 13 compares the simulated and real results of the age- and district-specific disease attack rates in Hong Kong. The simulation results are close to the real results, as shown in Figs. 13 (a) and (b), especially the results in the A1 age group in Fig. 13 (a) and the results of the R1 district in Fig. 13 (b). The hierarchical contact structure can improve our understanding of epidemic dynamics.

Conclusion

In this paper, we investigate the problem of mining social contact patterns from census data. Our major contribution is a newly proposed approach to determining social contact patterns from detailed Hong Kong census data. Compared with previous research, our proposed approach not only captures the hierarchical relationships of social contact patterns in relation to districts at different levels, but also characterizes the social contact patterns among the districts according to age group, allowing social contact patterns to satisfy the requirements of the public health department at different levels. The newly generated contact matrices that use the proposed approach reflect the social contacts within Hong Kongs social, economic and demographic structures and other related factors. We also evaluate the newly generated contact matrix derived from the 2011 Hong Kong census data by conducting simulation-based model experiments and predicting the dynamics of the 2009 Hong Kong H1N1 epidemic. Our experimental results show that the derived contact matrix gives more accurate SEIR-model-based predictions of the spread of disease. In the future, we will include more sociological knowledge in our analyses.

Author Contributions

Conceived and designed the experiments: ZY JL XZ. Performed the experiments: ZY JL XZ. Analyzed the data: ZY JL XZ. Contributed reagents/materials/analysis tools: ZY JL XZ. Wrote the paper: ZY JL XZ.

References

1. Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. *Nature* 466(7307): 761–764. doi: [10.1038/nature09182](https://doi.org/10.1038/nature09182) PMID: [20562860](https://pubmed.ncbi.nlm.nih.gov/20562860/)
2. Apicella C., Marlowe FW, Fowler JH, Christakis NA (2012) Social networks and cooperation in hunter-gatherers. *Nature* 481(7382): 497–501. doi: [10.1038/nature10736](https://doi.org/10.1038/nature10736) PMID: [22281599](https://pubmed.ncbi.nlm.nih.gov/22281599/)
3. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092): 337–341. doi: [10.1126/science.1215842](https://doi.org/10.1126/science.1215842) PMID: [22722253](https://pubmed.ncbi.nlm.nih.gov/22722253/)
4. Ayaz H, David L, Brian L (2013) Predictive validation of an influenza spread model. *PLoS One* 8(6): e65459. doi: [10.1371/journal.pone.0065459](https://doi.org/10.1371/journal.pone.0065459)
5. Bastolla U, Fortuna M, Pascual-Garcia A, Ferrera A, Luque B, et al. (2009) The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* 458(7241): 1018–1020. doi: [10.1038/nature07950](https://doi.org/10.1038/nature07950) PMID: [19396144](https://pubmed.ncbi.nlm.nih.gov/19396144/)
6. Bolton KJ, McCaw JM, Forbes K, Nathan P, Robins G, et al. (2012) Influence of contact definitions in assessment of the relative importance of social settings in disease transmission risk. *PLoS One* 7(2): e30893. doi: [10.1371/journal.pone.0030893](https://doi.org/10.1371/journal.pone.0030893) PMID: [22359553](https://pubmed.ncbi.nlm.nih.gov/22359553/)
7. Cauchemez S, Valleron AJ, Bolle PY, Flahault A, Ferguson NM (2008) Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* 452(7188): 750–754. doi: [10.1038/nature06732](https://doi.org/10.1038/nature06732) PMID: [18401408](https://pubmed.ncbi.nlm.nih.gov/18401408/)
8. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, et al. (2009) Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine* 361: 2619–2627. doi: [10.1056/NEJMoa0905498](https://doi.org/10.1056/NEJMoa0905498) PMID: [20042753](https://pubmed.ncbi.nlm.nih.gov/20042753/)
9. Clauset A, Moore C, Newman ME (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453(7191): 98–101. doi: [10.1038/nature06830](https://doi.org/10.1038/nature06830) PMID: [18451861](https://pubmed.ncbi.nlm.nih.gov/18451861/)
10. Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, et al. (2012) Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* 337(6099): 1228–1231. doi: [10.1126/science.1219385](https://doi.org/10.1126/science.1219385) PMID: [22955834](https://pubmed.ncbi.nlm.nih.gov/22955834/)
11. Del Valle SY, Hyman JM, Hethcote HW, Eubank SG (2007) Mixing patterns between age groups in social networks. *Social Networks* 29(4): 539–554. doi: [10.1016/j.socnet.2007.04.005](https://doi.org/10.1016/j.socnet.2007.04.005)
12. Eames KT, Tilston NL, White PJ, Adams E, Edmunds WJ (2010) The impact of illness and the impact of school closure on social contact patterns. *Health Technology Assessment* 14(34): 267–312. doi: [10.3310/hta14340-04](https://doi.org/10.3310/hta14340-04) PMID: [20630125](https://pubmed.ncbi.nlm.nih.gov/20630125/)

13. Eames KT, Tilston NL, Edmunds WJ (2011) The impact of school holidays on the social mixing patterns of school children. *Epidemics* 3(2): 103–108. doi: [10.1016/j.epidem.2011.03.003](https://doi.org/10.1016/j.epidem.2011.03.003) PMID: [21624781](https://pubmed.ncbi.nlm.nih.gov/21624781/)
14. Eames KT, Tilston NL, Brooks-Pollock E, Edmunds WJ (2012) Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS Comput Biology* 8(3): e1002425. doi: [10.1371/journal.pcbi.1002425](https://doi.org/10.1371/journal.pcbi.1002425)
15. Elaine ON, Richard JB, Madhav VM (2012) Sensitivity analysis of an individual-based model for simulation of influenza epidemics. *PLoS One* 7(10), e45414. doi: [10.1371/journal.pone.0045414](https://doi.org/10.1371/journal.pone.0045414)
16. Chao DL, Halloran ME, Obenchain VJ, Longini IM (2010) FluTE, a Publicly Available Stochastic Inuenza Epidemic Simulation Model. *PLoS Computational Biology* 6: e1000656. doi: [10.1371/journal.pcbi.1000656](https://doi.org/10.1371/journal.pcbi.1000656) PMID: [20126529](https://pubmed.ncbi.nlm.nih.gov/20126529/)
17. Elaine ON, Scotland CL, Madhav VM (2014) A Dirichlet process model for classifying and forecasting epidemic curves. *BMC Infect Dis* 14: 12. doi: [10.1186/1471-2334-14-12](https://doi.org/10.1186/1471-2334-14-12)
18. Eubank S, Guclu H, Kumar VSA, Marathe MV, Srinivasan A, (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429: 180–184. doi: [10.1038/nature02541](https://doi.org/10.1038/nature02541) PMID: [15141212](https://pubmed.ncbi.nlm.nih.gov/15141212/)
19. Fu YC, Wang DW, Chuang JH (2012) Representative contact diaries for modeling the spread of infectious diseases in Taiwan. *PLoS One* 7(10): e45113. doi: [10.1371/journal.pone.0045113](https://doi.org/10.1371/journal.pone.0045113) PMID: [23056193](https://pubmed.ncbi.nlm.nih.gov/23056193/)
20. Fumanelli L, Ajelli M, Manfredi P, Vespignani A, Merler S (2012) Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput Biology* 8(9), e1002673. doi: [10.1371/journal.pcbi.1002673](https://doi.org/10.1371/journal.pcbi.1002673)
21. Glasser J, Feng Z, Moylan A, Del V.S., Castillo-Chavez C (2012) Mixing in age-structured population models of infectious diseases. *Mathematical Biosciences* 235(1): 1–7. doi: [10.1016/j.mbs.2011.10.001](https://doi.org/10.1016/j.mbs.2011.10.001) PMID: [22037144](https://pubmed.ncbi.nlm.nih.gov/22037144/)
22. Gonz lez FSMC, Maritan A, Barab si A (2012) A universal model for mobility and migration patterns. *Nature* 484: 96–100. doi: [10.1038/nature10856](https://doi.org/10.1038/nature10856)
23. Hens N, Goeyvaerts N, Aerts M, Shkedy Z, Van DP, Beutels P (2009) Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infectious Diseases* 9, 5. doi: [10.1186/1471-2334-9-5](https://doi.org/10.1186/1471-2334-9-5) PMID: [19154612](https://pubmed.ncbi.nlm.nih.gov/19154612/)
24. Horby P, Pham QT, Hens N, Nguyen TT, Le QM, et al. (2011) Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS One* 6(2): e16965. doi: [10.1371/journal.pone.0016965](https://doi.org/10.1371/journal.pone.0016965) PMID: [21347264](https://pubmed.ncbi.nlm.nih.gov/21347264/)
25. Huerta-Quintanilla R, Canto-Lugo E, Viga-de A.D (2013) Modeling social network topologies in elementary schools. *PLoS One* 8(2): e55371. doi: [10.1371/journal.pone.0055371](https://doi.org/10.1371/journal.pone.0055371) PMID: [23408976](https://pubmed.ncbi.nlm.nih.gov/23408976/)
26. Iozzi F, Trusiano F, Chinazzi M, Billari FC, Zagheni E, et al. (2010) Little Italy: an agent-based approach to the estimation of contact patterns-fitting predicted matrices to serological data. *PLoS Comput Biology* 6(12): e1001021. doi: [10.1371/journal.pcbi.1001021](https://doi.org/10.1371/journal.pcbi.1001021)
27. Kretzschmar M, Mikolajczyk RT (2009) Contact profiles in eight European countries and implications for modelling the spread of airborne infectious diseases. *PLoS One* 4(6): e5931. doi: [10.1371/journal.pone.0005931](https://doi.org/10.1371/journal.pone.0005931) PMID: [19536278](https://pubmed.ncbi.nlm.nih.gov/19536278/)
28. Kretzschmar M, Teunis PF, Pebody RG (2010) Incidence and reproduction numbers of pertussis: estimates from serological and social contact data in five European countries. *PLOS Medicine* 7(6): e1000291. doi: [10.1371/journal.pmed.1000291](https://doi.org/10.1371/journal.pmed.1000291) PMID: [20585374](https://pubmed.ncbi.nlm.nih.gov/20585374/)
29. Kucharski AJ, Gog JR (2012) The role of social contacts and original antigenic sin in shaping the age pattern of immunity to seasonal influenza. *PLoS Comput Biology* 8(10): e1002741. doi: [10.1371/journal.pcbi.1002741](https://doi.org/10.1371/journal.pcbi.1002741)
30. Lee S, Rocha LE, Liljeros F, Holme P (2012) Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS One* 7(5): e36439. doi: [10.1371/journal.pone.0036439](https://doi.org/10.1371/journal.pone.0036439) PMID: [22586472](https://pubmed.ncbi.nlm.nih.gov/22586472/)
31. Liu J, Xia S (2011) Toward effective vaccine deployment: a systematic study. *Journal of Medical Systems* 35(5): 1153–1164. doi: [10.1007/s10916-011-9734-x](https://doi.org/10.1007/s10916-011-9734-x) PMID: [21607707](https://pubmed.ncbi.nlm.nih.gov/21607707/)
32. Liu J, Xia S (2013) A computational approach to characterizing the impact of social influence on individuals' vaccination decision making. *PLoS One* 8(4): e60373. doi: [10.1371/journal.pone.0060373](https://doi.org/10.1371/journal.pone.0060373) PMID: [23585835](https://pubmed.ncbi.nlm.nih.gov/23585835/)
33. Melegaro A, Jit M, Gay N, Zagheni E, Edmunds WJ (2011) What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. *Epidemics* 3(3–4): 143–151. doi: [10.1016/j.epidem.2011.04.001](https://doi.org/10.1016/j.epidem.2011.04.001) PMID: [22094337](https://pubmed.ncbi.nlm.nih.gov/22094337/)
34. Kermack W, McKendrick A (1991) Contributions to the mathematical theory of epidemics?I. *Bulletin of Mathematical Biology* 53(1–2): 33–55. doi: [10.1007/BF02464423](https://doi.org/10.1007/BF02464423) PMID: [2059741](https://pubmed.ncbi.nlm.nih.gov/2059741/)

35. Mikolajczyk RT, Kretzschmar M (2008) Collecting social contact data in the context of disease transmission: prospective and retrospective study designs. *Social Networks* 30(2): 127–135. doi: [10.1016/j.socnet.2007.09.002](https://doi.org/10.1016/j.socnet.2007.09.002)
36. Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine* 5(3): e74. doi: [10.1371/journal.pmed.0050074](https://doi.org/10.1371/journal.pmed.0050074) PMID: [18366252](https://pubmed.ncbi.nlm.nih.gov/18366252/)
37. Ndeffo MML, Liu J, Bauch CT, Tekel YI, Medlock J, et al. (2012) The impact of imitation on vaccination behavior in social contact networks. *PLoS Comput Biology* 8(4): e1002469. doi: [10.1371/journal.pcbi.1002469](https://doi.org/10.1371/journal.pcbi.1002469)
38. Volz EM, Miller JC, Galvani A, Ancel ML (2011) Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput Biology* 7(6): e1002042. doi: [10.1371/journal.pcbi.1002042](https://doi.org/10.1371/journal.pcbi.1002042)
39. Rohani P, Zhong X, King AA (2010) Contact network structure explains the changing epidemiology of pertussis. *Science* 330: 982–985. doi: [10.1126/science.1194134](https://doi.org/10.1126/science.1194134) PMID: [21071671](https://pubmed.ncbi.nlm.nih.gov/21071671/)
40. Rolls DA, Wang P, Jenkinson R, Pattison PE, Robins GL, et al. (2013) Modelling a disease-relevant contact network of people who inject drugs. *Social Networks* 35(4): 699C710. doi: [10.1016/j.socnet.2013.06.003](https://doi.org/10.1016/j.socnet.2013.06.003)
41. Son S, Kang AR, Kim HC, Kwon T, Park J, et al. (2012) Analysis of context dependence in social interaction networks of a massively multiplayer online role-playing game. *PLoS One* 7(4): e33918. doi: [10.1371/journal.pone.0033918](https://doi.org/10.1371/journal.pone.0033918) PMID: [22496771](https://pubmed.ncbi.nlm.nih.gov/22496771/)
42. Stehle J, Voirin N, Barrat A, Cattuto C, Isella L, et al. (2011) High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One* 6(8): e23176. doi: [10.1371/journal.pone.0023176](https://doi.org/10.1371/journal.pone.0023176) PMID: [21858018](https://pubmed.ncbi.nlm.nih.gov/21858018/)
43. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America* 107(31): 13636–13641. doi: [10.1073/pnas.1004008107](https://doi.org/10.1073/pnas.1004008107) PMID: [20643965](https://pubmed.ncbi.nlm.nih.gov/20643965/)
44. Ventresca M, Aleman D (2013) Evaluation of strategies to mitigate contagion spread using social network characteristics. *Social Networks* 35(1): 75–88. doi: [10.1016/j.socnet.2013.01.002](https://doi.org/10.1016/j.socnet.2013.01.002)
45. Wallinga J, Teunis P, Kretzschmar M (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology* 164(10): 936–944. doi: [10.1093/aje/kwj317](https://doi.org/10.1093/aje/kwj317) PMID: [16968863](https://pubmed.ncbi.nlm.nih.gov/16968863/)
46. Willem L, Van KK, Chao DL, Hens N, Beutels P (2012) A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS One* 7(11): e48695. doi: [10.1371/journal.pone.0048695](https://doi.org/10.1371/journal.pone.0048695) PMID: [23155399](https://pubmed.ncbi.nlm.nih.gov/23155399/)
47. Zelner JL, Trostle J, Goldstick JE, Cevallos W, House JS, et al (2012) Social connectedness and disease transmission: social organization, cohesion, village context, and infection risk in rural Ecuador. *American Journal of Public Health* 102(12): 2233–2239. doi: [10.2105/AJPH.2012.300795](https://doi.org/10.2105/AJPH.2012.300795) PMID: [23078481](https://pubmed.ncbi.nlm.nih.gov/23078481/)