Estimating high-resolution daily encounter networks with activity-travel diaries

I. Saadi[a,*], E. Côme[a], L. Luong[b], S. Lassarre[a] and M. Zargayouna[a]

[a] Univ. Gustave Eiffel, F-77454 Marne-la-Vallée, France
{ismail.saadi,etienne.come,sylvain.lassarre,mahdi.zargayouna}@univ-eiffel.fr
[b] AP-HP Hôpital Cochin, F-75014 Paris, France
liem.luong@aphp.fr
[*] Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

April 12, 2022

Keywords: Activity-based model (ABM), daily contact networks, COVID-19

1 INTRODUCTION

SARS-CoV-2 has become one of the most devastating pandemics in human history and is still, two years after the pandemic onset, contributing to the breakdown of societies around the world. As a result, several countries are experiencing deep economic, social and political crises. Furthermore, multiple non pharmaceutical interventions (NPIs), i.e. social distancing, travel restrictions, partial/full lockdowns, have been put in place to mitigate human contacts. These interventions have important social, economic, public health, legal implications and they should be strongly justified. In general, decision making of such measures is based on epidemiological forecasts. Thus, understanding human contact patterns and their underlying social interactions is crucial to estimate reliable infectious disease transmission models. This is especially the case in large densely-crowded urban and transportation systems where complex social interactions are widespread. In this paper, we present a new estimation method of human contact patterns based on an activity-based model of travel demand. The latter are used in transportation research to model persons' travel behavior. Nonetheless, they are not explored enough in epidemiology. Therefore, we demonstrate the effectiveness of our modeling strategy by setting up an activity-based model based on publicly available data, such that the approach is easily reproducible. Afterwards, we derive large scale multi-layer contact networks. Finally, we estimate the related age-stratified contact matrices.

2 MODELING FRAMEWORK

The modeling framework consists of three parts. The first part is dedicated to the behavioral synthetic population of Île-de-France based on multiple open data sources. In the second part, we set up an efficient and fast algorithm to estimate the large scale multi-setting contact network based on the targeted behavioral synthetic population. Finally, at the end of this process, we compute the age-specific contact matrices that estimate contact frequencies of individuals clustered by age groups in two particular settings, i.e. household and overall layers.

2.1 Behavioral synthetic population

A behavioral synthetic population is a detailed representation of the true population based on multiple data sources. It includes several millions of individuals grouped into households. Socio-demographic attributes are inferred and daily activity-travel schedules are associated to the individuals of that population. The integrated modeling framework simulates the synthetic population of Île-de-France, a densely populated area of 12,011 km2 consisting of eight departments, including Paris the capital of France. As of January 2021, the population size of Île-de-France is around 12,213,447 inhabitants, resulting in a population density of 1,017 inhabitants/km2. The proposed behavioral synthetic population is based on multiple datasets which are publicly available, thus allowing full transparency and reproducibility of the present study. We use the generation process proposed by Hörl & Balac (2021). Data gathering and pipeline running are described at `https://github.com/eqasim-org/ile-de-france`.

2.2 Multi-layer contact network estimation

Infectious disease transmission dynamics are heavily reliant on the social interactions of the individuals constituting the population. Modeling the human mixing patterns is indeed crucial to understanding how epidemics spread in-between individuals. In this regard, the most appropriate approach consists of representing the contact patterns as a graph $G = (V, E)$ where V, the set of vertices, are the individuals and E, the set of edges, are the contacts.

The total number of contacts is computed based on the spatiotemporal co-presence. In other words, if a group of individuals share the same location at the same time, then we assume that they might have met. The individuals' spatiotemporal co-presence can be derived from the activity-travel diaries. In doing so, we obtain detailed information on the spatiotemporal meeting patterns, resulting in underlying exact upper-bound contact duration distributions. An activity travel diary contains the reference number of the person, the activity purpose, start/end times and location ID. For instance, we use activity-travel diaries (ATD) consisting of 20,716,218 records. We group the activity-travel diaries by unique location to apply the spatial constraint of the co-presence. Then we compute a systematic pairwise comparison of individuals sharing the same location to check for joint time windows.

In particular, we distinguish between the potential contact network $G_p = (V_p, E_p)$ and the effective contact network $G_e = (V_e, E_e)$, and the relationship is described by $|V_p| = |V_e|$ and $\pi|E_p| = |E_e|$. Indeed, we consider a Bernoulli model by assuming a single parameter π. It means that if a contact is detected, it is accepted with a probability π. We compute the contact duration and we assign its purpose, e.g. household (H), work (W), education (E), leisure (L), shop (S), other (O) and/or hybrid (X-X). Given that the daily average number of contacts μ_c is around 8 (5-14) (Béraud *et al.*, 2015), and taking into account the statistical properties of the potential contact network (before making the decision on preserving or not the contacts after detecting the spatiotemproal co-presence), we can estimate the optimal value of π. Indeed, we can state that $\mu_c = \frac{2(|E_p^H|+|E_p^O|\pi)}{N_{pop}}$ where μ_c is the observed overall average daily number of contacts (from the survey), $|E_p^H|$ is the total number of contacts in the household layer from the potential graph G_p, $|E_p^O|$ the total number of contacts in all the other layers from their corresponding potential graphs and $N_{pop}(= |V|)$ the number of vertices/individuals of the graph. Thus, $\pi = \frac{\frac{\mu_c N_{pop}}{2} - |E_p^H|}{|E_p^O|}$. Note that $|E_e| = |E_e^O| + |E_e^H| = |E_p^O|\pi + |E_p^H|$.

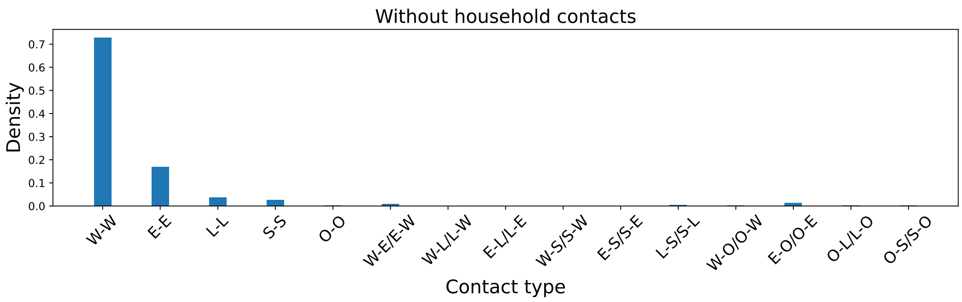
2.3 Numerical experiment

We build a representative population sample which represents 100% of the true population size. The sampling procedure consists of the generation of a behavioral synthetic population of around 12M persons grouped into 4M households. The basic statistical properties of the contact network are presented in Table 1.

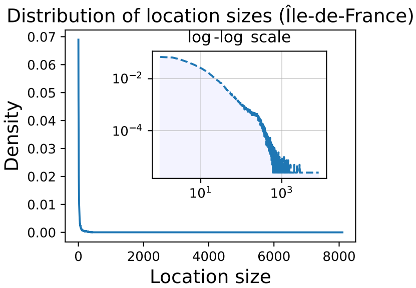
Parameter	Value
$\lvert E_e \rvert$	47,029,073
$\lvert V \rvert (= N_{pop})$	11,758,464
π	0.013

Table 1 – *Properties of the effective contact network*

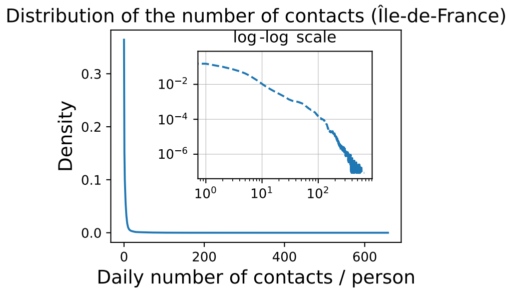
Most contacts are non-hybrid (Figure 1a). Furthermore, the modeling framework highlights two key aspects: the distributions of the location sizes (number of people sharing the same location) (Figure 1b) and the daily number of contacts per person (Figure 1c) look like power law distributions with cutoffs. The upper bound contact duration distribution is presented in Figure 1d. The waves are strongly influenced by the temporal patterns in workplaces and education-related locations, e.g. 8 hours/day of working time and/or morning/afternoon presence at workplaces. This is a mixture of layer-related contact duration distributions.



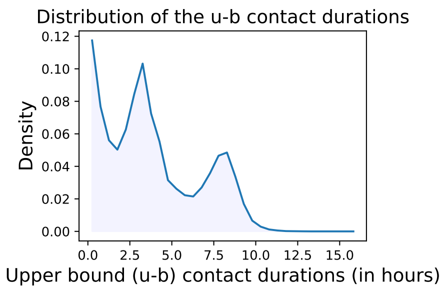
(a)



(b)



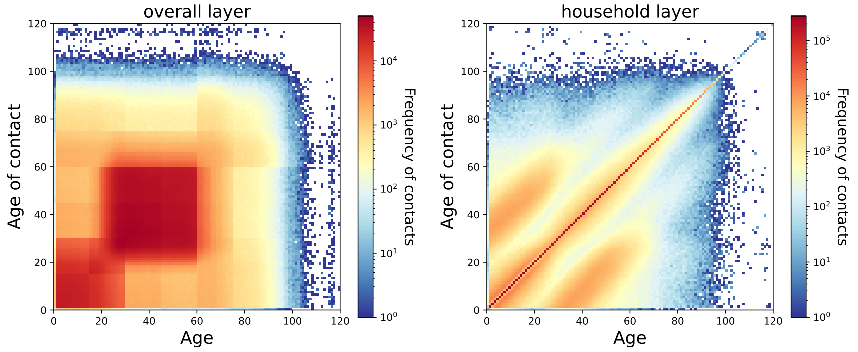
(c)



(d)

Figure 1 – *Key results*

In households (Figure 2b), the main diagonal corresponds to the contacts of people of similar age groups, i.e. children and parents. The secondary diagonals mainly correspond to the parent-children contacts. In work (W)/education (E)/leisure (L)/shop (S)/other (O) layers, we mainly find contacts in-between individuals of age 25-, 25 and 65, with some spread below



(a) (b)

Figure 2 – *Age-specific contact matrices*

and above these boundaries. Therefore, the corresponding patterns clearly emerge in the joint (non-household layers) age-specific contact matrix (Figure 2a).

3 CONCLUSIONS

The pipeline framework (Eqasim) (Hörl & Balac, 2021) was initially designed for simulating mobility of multi-agent systems. In this context, it adopts different modeling assumptions regarding the spatiotemporal activity sequencing. The socio-demographic description of the sythetic population is based on open data (INSEE) which are relatively reliable and highly detailed with the possibility of matching individuals with their corresponding households. Such data are rarely available in other countries. Nonetheless, the travel behavior description of the agents is based on the National Household Travel Survey of 2008 (ENTD) as no viable alternative is available, which may be a limitation. In fact, travel habits might have evolved. Although the initial purpose of Eqasim is about transport and mobility, we demonstrate in this paper that it can be an interesting option for modeling infectious disease transmission. From a contact pattern standpoint, we show that a conventional activity-based model of travel demand clearly captures the underlying human mixing patterns characterizing the different layers. We also demonstrated that it is possible to reproduce the visitation and contact patterns (super-spreaders, super-spreading events) previously described in the literature.

Further research is needed (a) to estimate the intra/inter-layer related parameters taking into account key statistical properties from the existing literature. At this stage, we only considered a parameter π. Incorporating variability into and/or in-between the layers is required to improve the statistical representativeness of the effective contact network G_e. (b) To incorporate dynamic features into the model. For example, a dynamic contact network can be used to take into account the effects of NPIs. (c) To validate the contact network with respect to key observed statistical indicators. (d) To run infectious disease transmission models on the estimated contact networks to assess the effects of various mobility-related NPIs on the confirmed cases/deaths.

References

Béraud, Guillaume, Kazmercziak, Sabine, Beutels, Philippe, Levy-Bruhl, Daniel, Lenne, Xavier, Mielcarek, Nathalie, Yazdanpanah, Yazdan, Boëlle, Pierre-Yves, Hens, Niel, & Dervaux, Benoit. 2015. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PloS one*, **10**(7), e0133203.

Hörl, Sebastian, & Balac, Milos. 2021. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, **130**, 103291.