

Mixing Matrices for Quasi-Spatial Population Stratifications: Age & Mobility Data

Jesse Knight

v.3.0

2021 April 29

1 Objective

Consider a population stratified by into $A = 11$ age groups (a) and $N = 510$ neighbourhoods (FSA, n). We are interested in defining the number of contacts formed between all individuals in group na and all individuals in group $n'a'$, which we denote $X_{nan'a'}$. Then, we will use a pre-determined assignment FSAs into $G = 10$ groups (deciles, g) to aggregate the FSA dimensions and obtain $X_{gag'a'}$.

2 Approach

2.1 Data

We know the sizes of populations in each FSA n , stratified by age groups a , which we denote P_{na} . Anonymous cell phone mobility data from Ontario has also been used to estimate the average proportion of individuals living in FSA n who travel to FSA n' per day for at least one hour, which we denote $B_{nn'}$. These proportions are aggregated into groups gg' and shown in Figure 4. Finally, the POLYMOD study estimates the average numbers of contacts per person per day between age groups a and a' , stratified by the following types (y) of contacts: *home, work, school, transport, leisure, other*; these data are denoted $C_{aa'y}$ and are illustrated in Figure 1.

2.2 Contact Rates

The total number of type y contacts per individual in age group a per day in the POLYMOD study can be obtained by summing across the “other” age groups a' :

$$C_{ay} = \sum_{a'} C_{aa'y} \quad (1)$$

We assume these contact rates apply to our population. Thus, the total number of type y contacts made available by age a individuals in FSA n is:

$$Q_{nay} = P_{na} C_{ay} \quad (2)$$

These contacts can be formed with three potential mixing pools: (a) others met while travelling to another FSA n' ; (b) others met who travelled to this FSA n ; and (c) others within this FSA n only. For a given contact type y , we assume that a proportion h_y of those contacts are be formed with pool (c). We further assume that the remaining contacts $(1 - h_y)$ are formed with pools (a,b) in proportion to population mobility. That is the proportion who are mobile form these contacts with (a), and the proportion who are not form these contacts with (b). Let $B_n = \sum_{B_{nn'}}$ be the total mobile proportion.¹ Then Q_{nay} are distributed amongst the mixing pools as:

$$Q_{nay} \sim \begin{cases} B_n(1 - h_y) & \text{(a) other traveller pools} \\ (1 - B_n)(1 - h_y) & \text{(b) local traveller pool} \\ h_y & \text{(c) local non-traveller pool} \end{cases} \quad (3)$$

Now consider the traveller pool in FSA n^* , comprising (a) contacts from other FSAs and (b) contacts from this FSA. The total number of type y contacts made available to the pool by age a individuals from FSA n is (where $1_{x=y}$ denotes 1 if $x = y$ else 0):

$$Q_{nay}^{n^*} = Q_{nay}(1 - h_y)(B_{nn^*} + 1_{n=n^*}(1 - B_n)) \quad (4)$$

Thus, the total number of contacts made available to the pool is:

$$T_{n^*y} = \sum_{na} Q_{nay}^{n^*} \quad (5)$$

We assume that contacts form randomly with respect to home FSA n , but the POLYMOD data suggest that contacts do not form randomly with respect to age. We cannot use the POLYMOD contact matrices directly, since these depend on the distribution of age groups from the POLYMOD study, and thus would not balance given the different age distributions between FSAs. However, we can assume a parametric form for age mixing, such as:

$$Q'_{aa'y} = Q_{ay}((1 - \epsilon_y)\pi_{a'y} + (\epsilon_y)1_{a=a'}), \quad \pi_{a'y} = \frac{Q_{a'y}}{\sum_{a'} Q_{a'y}} \quad (6)$$

where $\epsilon_y \in [0, 1]$ denotes the degree of assortative (like-with-like) mixing by age groups. The value of ϵ_y can be estimated based on POLYMOD data, by minimizing the difference between the empiric $Q_{aa'y}$ and the parametric approximation $Q'_{aa'y}$, using $Q_{ay} \approx C_{ay}$ (no population weighting). A good measure of “difference” is the Kullback-Leibler Divergence:²

$$D_{KL}(Q_{aa'y} \parallel Q'_{aa'y}) = \sum_{aa'} Q_{aa'y} \log \left(\frac{Q_{aa'y}}{Q'_{aa'y}} \right) \quad (7)$$

The estimated ϵ_y using this approach are given in Figure 1. Returning to the traveller pool in FSA n^* , the total number of type y contacts formed between people aged a from FSA n with people aged a' from FSA n' is then given by:

$$X_{nan'a'y}^{n^*} = (1 - \epsilon_y)\pi_{nan'a'y}^{n^*} + (\epsilon_y)1_{a=a'} \sum_{a'} \pi_{nan'a'y}^{n^*}, \quad \pi_{nan'a'y}^{n^*} = \frac{Q_{nay}^{n^*} Q_{n'a'y}^{n^*}}{T_{n^*y}} \quad (8)$$

¹ The current implementation in code introduces another parameter to control the total mobile population directly.

² The contact matrices $C_{aa'y}$ and $C'_{aa'y}$ are not strictly probability distributions because they sum to more than one. We could normalize $C_{aa'y}$ and $C'_{aa'y}$ by $\sum_a C_{ay}$ to compute a “proper” KL-Divergence. However, as can be seen in Eq (7), this constant will not affect the minimization.

The total number of contacts formed between FSA n and FSA n' is then the sum across all traveller pools, including those corresponding to $n = n^*$ and $n' = n^*$, plus the local contacts from case (c):

$$X_{nan'a'y} = \sum_{n^*} X_{nan'a'y}^{n^*} + 1_{n=n'} \left((1 - h_y) \sum_{n'} X_{nan'a'y}^{n^*} \right) \quad (9)$$

2.3 Caveats

- We introduced a 6 free parameters in h_y , representing the proportion of type y contacts that are formed with the local non-traveller pool. However, some values of h_y can likely be fixed, such as $h_1 = 1$: no *home* contacts formed with the traveller pool, etc.
- We assume that travelling individuals only form contacts in one destination. However, due to the large numbers involved, this assumption won't likely have any influence on the results.
- We have not considered differences in contact numbers by FSA n , for example, due to different types of employment or housing. However, this could be added to Q_{nay} in Eq. (2).

3 Results

Note: In Section 2 we described the methodology in relation to $N = 510$ FSAs. However, as noted in Section 1, we eventually intended to aggregate the mixing matrices to reflect $G = 10$ groupings of FSAs. Since the mobility data currently available are based on these groupings (g), the results below are for the groupings. When analyzing groups g instead of FSAs n , the methodology above can be repeated exactly using $n \rightarrow g$, $Q_{nay} \rightarrow Q_{gay}$, and $B_{nn'} \rightarrow B_{gg'}$, which can be obtained as suggested in v.1 of this report (where S_g is the set of n assigned to group g):

$$Q_{gay} = \sum_{n \in S_g} Q_{nay}, \quad B_{gg'} = \sum_{n \in S_g} \sum_{n' \in S_{g'}} B_{nn'} \quad (10)$$

The estimated values of ϵ_y corresponding to the degree of assortative age mixing by contact type were: *home*: 0.09, *work*: 0.03, *school*: 0.22, *transport*: 0.12, *leisure*: 0.13, *other*: 0.06. Figure 1 compares the original POLYMOD contact data with the values approximated by the parametric equation (6).

Figure 2 illustrates the total number of contacts (millions) predicted between deciles, stratified by contact type, using $h_y = [1, 0, 1, 0, \frac{1}{2}, \frac{1}{2}]$.³ We verified that $\sum_{g'a'} X_{gag'a'y} = Q_{gay}$ for all contact types y . When $h_y = 0$ (e.g. *home* contacts), the contact matrix has no off-diagonal elements. However, even for $h_y = 1$ (e.g. *work* contacts), the contact matrix is still dominated by the diagonal due to the tendency for individuals to move between FSA in the same decile (Figure 4), and the larger size of the non-mobile population (locals) in the local traveller pool. The proportions of all contacts formed within the same decile across the various contact types was: *home*: 1.00, *work*: 0.33, *school*: 1.00, *transport*: 0.33, *leisure*: 0.67, *other*: 0.67; and 0.72 overall. To improve readability, Figure 2 therefore also shows the number of contacts in the off-diagonal elements alone. Finally, Figure 3 similarly illustrates the total number of contacts (millions) predicted between deciles, after aggregating over contact types.

³ And also assuming 50% population mobility

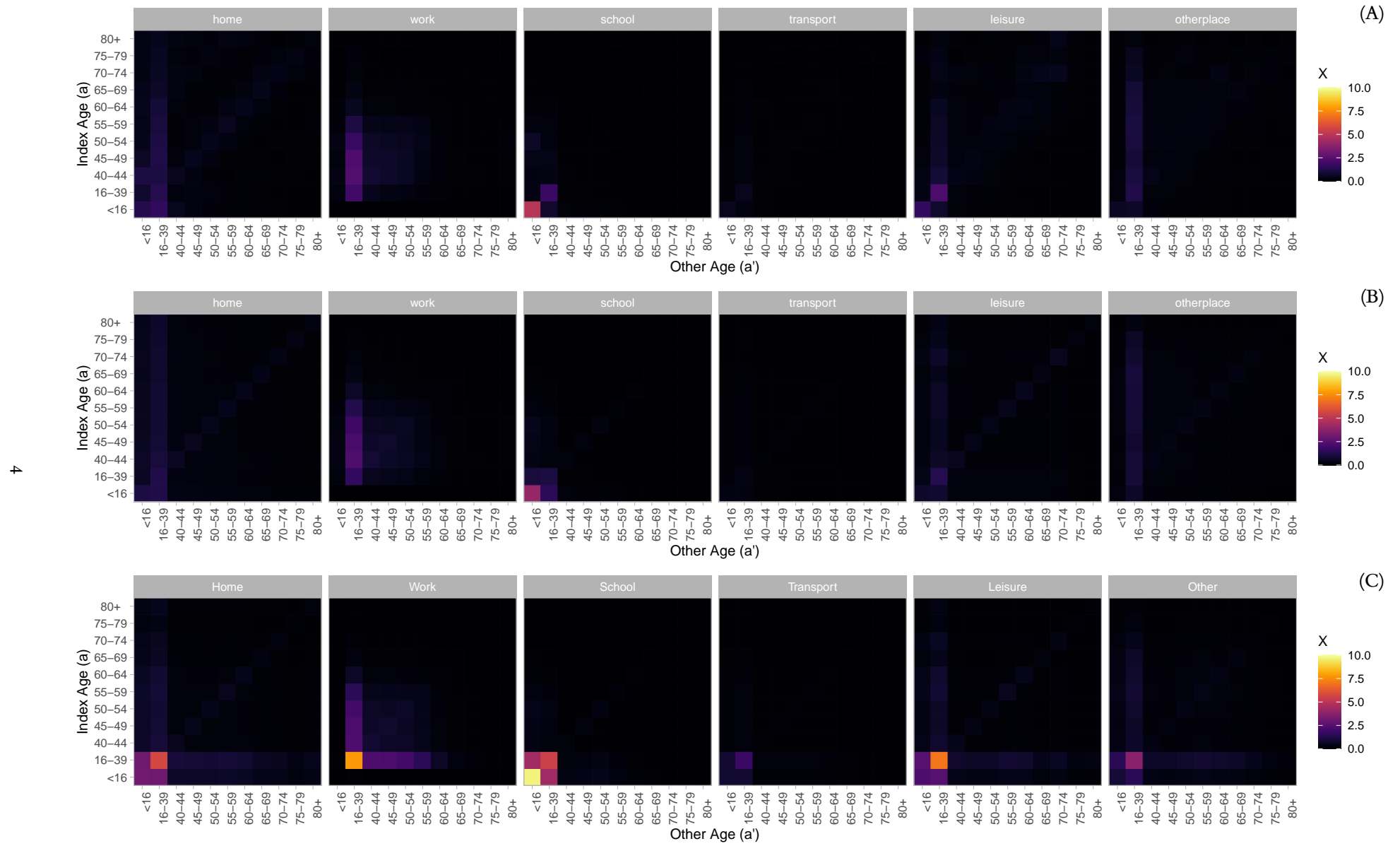


Figure 1: Average number and types of contacts between individuals of various age groups per day. A: original POLYMOD data; B: parametric approximation of POLYMOD data; C: total contacts modelled in Ontario in conjunction with mobility data.

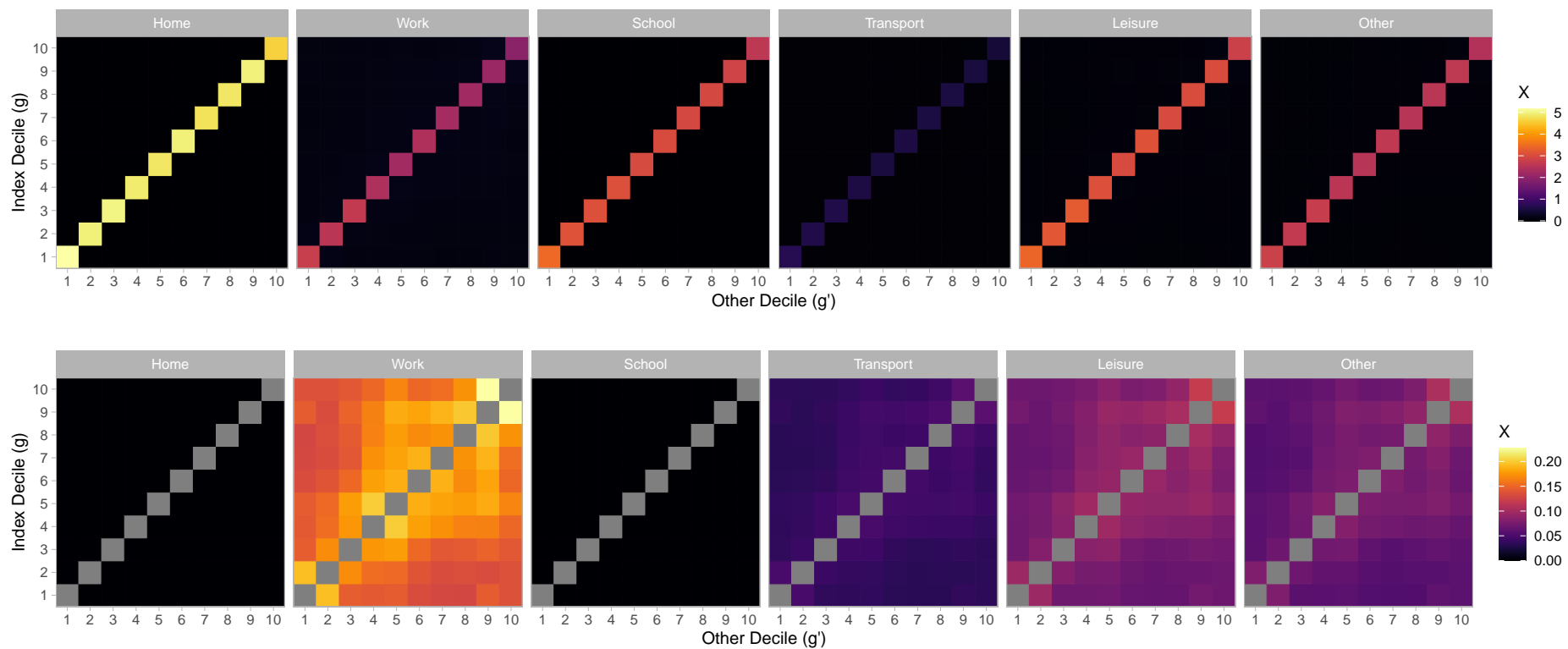


Figure 2: Total number and types of contacts formed between deciles (millions); raw values (top) and after removing the dominant diagonal (bottom)

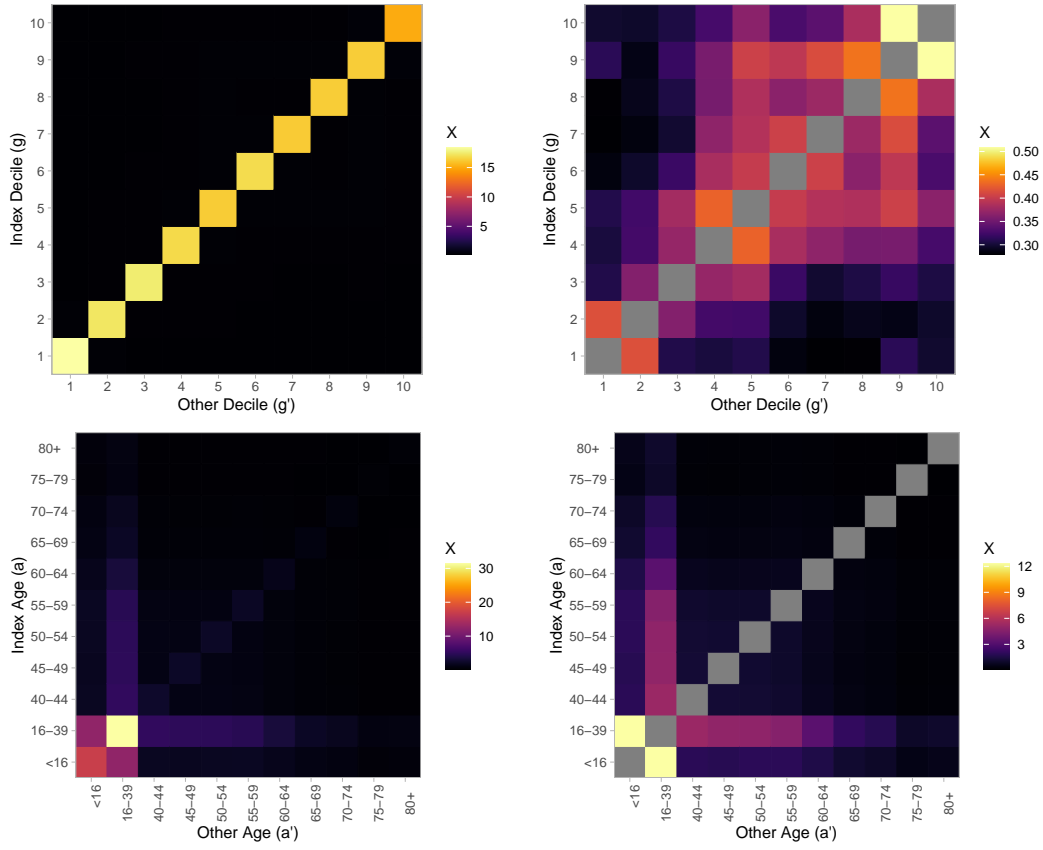


Figure 3: Total number of contacts (millions) formed between deciles (top) and age groups (bottom); complete values (left) and after removing the dominant diagonals (right)

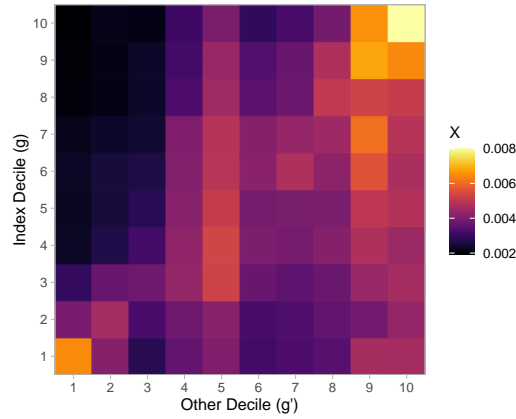


Figure 4: Cell phone mobility data: proportion of individuals residing in “index decile” who travelled to “other decile” for at least one hour per day