Mixing Matrices for Quasi–Spatial Population Stratification: Using Mobility Data

Jesse Knight

v.2.0

2021 April 26

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 an and all individuals in group $a'n'$, which we denote $X_{ana'n'}$. Then, we will use a pre–determined assignment FSAs into $G = 10$ groups (g) to aggregate the FSA dimensions and obtain $X_{aga'g'}$.

2 Approach

2.1 Data

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 A.1 We assume these rates of contacts are similar for our population (see Section 2.4 for further comment on these assumptions). The total number of type y contacts formed by people in age group a is $C_{ay} = \sum_{a'} C_{aa'y}$. We know the sizes of populations in each FSA n, stratified by age groups a, which we denote P_{na}. Thus, the total number of type y contacts made available by FSA n is:

$$Q_{ny} = P_{na} \times C_{ay} \tag{1}$$

Cell phone (TBC) mobility data from Ontario has been used to estimate the average proportion of individuals residing in FSA n who travel to FSA n' per day, denoted $B^*_{nn'}$. These proportions are aggregated into groups gg' and shown in Figure A.2. To use these data, we need to make several assumptions: 1) What proportion of the FSA population are captured in the data? 2) Is this proportion equally distributed across age groups and equal across FSAs? 3) With whom do travellers mix in the destination FSA: local residents and/or other travellers? 4) What types of contacts are formed where?

For now, we assume that: (1,2) the mobility data sample an equal proportion b_s across all FSA and age groups; and (3,4) a proportion of type y contacts b_y are formed either while travelling or with travellers to FSA n, and the remainder ($1 - b_y$) are formed only with other locals in FSA n. The proportion of people travelling between FSAs n and n' after correcting for under–sampling of mobility data is therefore $B_{nn'} = B^*_{nn'}/b_s \leq 1$, and the total proportion of people travelling away from FSA n per day is $B_n = \sum_n B_{nn'}$.

2.2 Numbers of Contacts

Consider Q_{ny}: the total number of type y contacts formed by individuals in FSA n. These contacts can be formed with three pools of potential "others": (a) others met while travelling to other FSAs; (b) others met who travelled to this FSA; and (c) others from this FSA (who may/not also travel at other times to form other types of contacts). The total contacts Q_{ny} are distributed between (a,b,c) as follows:

$$Q_{ny} \sim \begin{cases} B_n b_y & \text{(a) while travelling} \\ (1 - B_n) b_y & \text{(b) with local traveller pool} \\ (1 - b_y) & \text{(c) with locals only} \end{cases} \tag{2}$$

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

$$Q_{ny}^{n^*} = b_y Q_{ny} \big(B_{nn^*} + 1_{n=n^*} (1 - B_n) \big) \tag{3}$$

So, the total number of contacts made available to the pool from all FSAs is:

$$T_{n^* y} = \sum_n Q_{ny}^{n^*} \tag{4}$$

If random mixing is assumed within the traveller pool, then the total number of contacts formed between individuals from FSAs n and n' in the pool is:

$$X_{nn'y}^{n^*} = \frac{Q_{ny}^{n^*} Q_{n'y}^{n^*}}{T_{n^* y}} \tag{5}$$

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

$$X_{nn'y} = \sum_{n^*} X_{nn'y}^{n^*} + 1_{n=n'} \big((1 - b_y) Q_{ny} \big) \tag{6}$$

2.3 Free Parameters

The free parameters (count) introduced here are:

- b_s: (1) the proportion of individuals represented in the mobility data

- b_y: (6) the proportion of type y contacts that are formed with the traveller pool

However, some values of b_y can likely be fixed, such as $b_1 = 0$: no *home* contacts formed with the traveller pool; $b_2 = 1$: all *work* contacts formed with the traveller pool, etc.

2.4 Caveats

- 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_{ny} in Eq. (1).

- The proportion of contacts that are formed internally (within the same FSA) is sensitive to the parameter b_s because a large non–mobile population in FSA n^* (locals) will dominate the local traveller pool.

- Our current approach aggregates over age strata before calculating the contacts formed between FSA for simplicity, thus precluding age:FSA interactions in mixing. However, since contact types are disaggregated, age:contact–type interactions could be supported.

3 Results

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 $n \to g$, $Q_{ny} \to Q_{gy}$, and $B_{nn'} \to B_{gg'}$, which can be obtained as suggested in v.1 of this report (where S_g is the set of n assigned to group g):

$$Q_{gy} = \sum_{n \in S_g} Q_{ny}, \quad B_{gg'} = \sum_{n \in S_g} \sum_{n' \in S_{g'}} B_{nn'} \tag{7}$$

Figure 1 illustrates the total number of contacts (millions) predicted between deciles, stratified by contact type, using $b_s = .05$ and $b_y = [0, 1, 0, 1, \frac{1}{2}, \frac{1}{2}]$. We verified that $\sum_{g'} X_{gg'y} = Q_{gy}$ for all contact types y. When $b_y = 0$ (e.g. *home* contacts), the contact matrix has no off–diagonal elements. However, even for $b_y = 1$ (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 A.2), 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*: 0.15, *school*: 1.00, *transport*: 0.15, *leisure*: 0.58, *other*: 0.58; and 0.64 overall. To improve readability, Figure 1 therefore also shows the number of contacts in the off-diagonal elements alone. Finally, Figure 2 similarly illustrates the total number of contacts (millions) predicted between deciles, after aggregating over contact types.
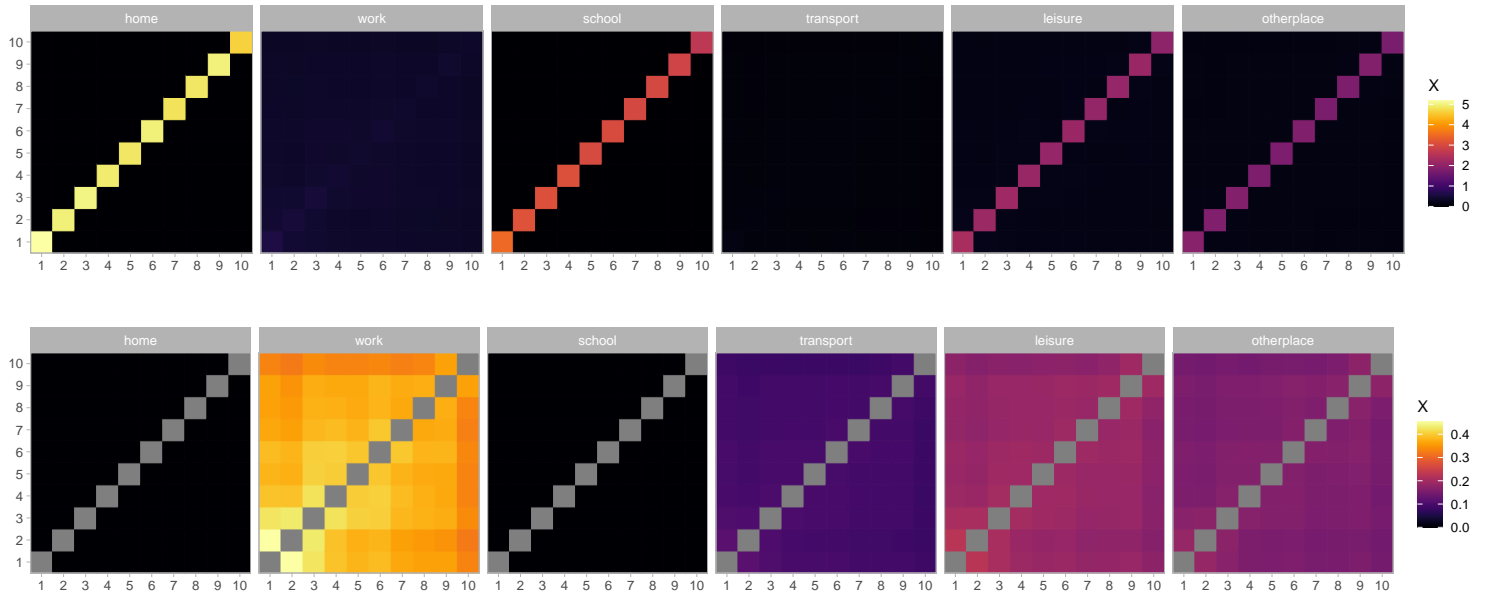
Figure 1: Total number and types of contacts formed between deciles (millions); raw values (top) and after removing the dominant diagonal (bottom)
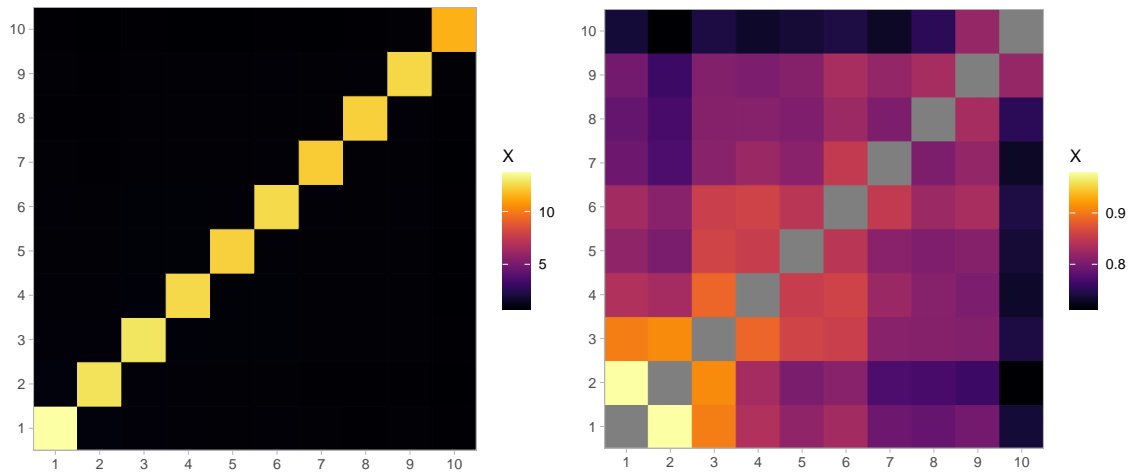

Figure 2: Total number of contacts formed between deciles (millions); raw values (left) and after removing the dominant diagonal (right)
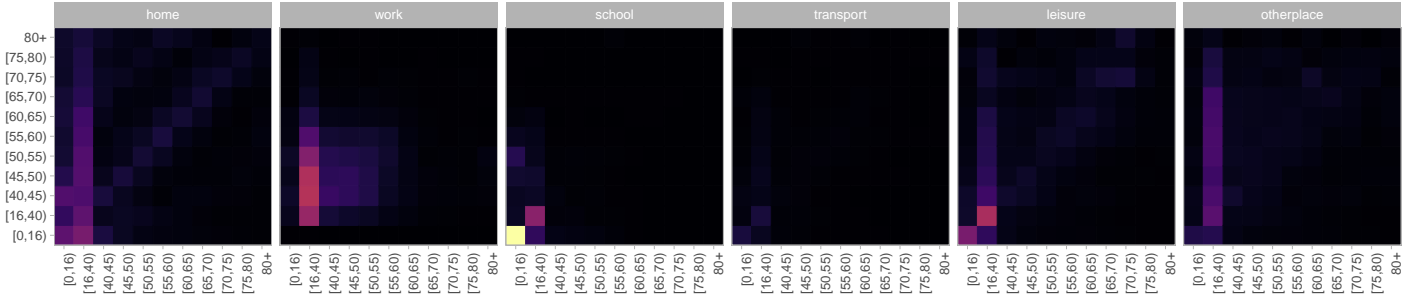
A   Supplemental Results

Figure A.1: Average number and types of contacts between various age groups per day (POLYMOD)
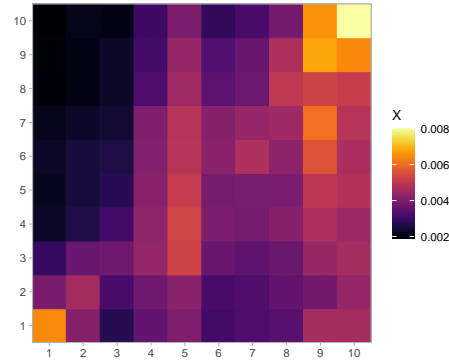

Figure A.2: Raw cell phone mobility data: proportion of individuals residing in "index decile" who travelled to "other decile" per day