Mixing Matrices for Quasi-Spatial Population Stratification: Adjacency Approach

Jesse Knight

21 April, 2021

1 Objective

The population is stratified into $N = 513$ "FSA" (neighbourhoods) indexed n. Each FSA has P_n population size, average number of contacts C_n, and is assigned to one of $G = 10$ groups indexed g; then S_g denotes the set of n assigned to group g. We are interested in constructing a mixing matrix $M_{gg'}$ describing the probability of contact formation between an average individual in group g with an average individual in group g'. The properties of $M_{gg'}$ include:

1. Bounded: $0 \le M_{gg'} \le 1$
2. Sum to One: $1 = \sum_{g'} M_{gg'}$
3. Balance: $P_g C_g M_{gg'} = P_{g'} C_{g'} M_{g'g}$

2 Toy Example: Adjacent Mixing

Consider the example network in Figure 1 with 5 FSAs: A, B, C, D, E. The FSA sizes are $P_n = [17, 10, 5, 5, 3]$, with $S_1 = \{A, D\}$ (red), and $S_2 = \{B, C, E\}$ (blue). For now, we assume $C_n = 1$. The adjacency matrix Λ is:

$$\Lambda_{nn'} = \begin{bmatrix} * & 1 & 1 & 1 & 0 \\ 1 & * & 1 & 0 & 0 \\ 1 & 1 & * & 1 & 0 \\ 1 & 0 & 1 & * & 1 \\ 0 & 0 & 0 & 1 & * \end{bmatrix} \tag{1}$$

2.1 Balancing FSA Contacts

Let $Q_n = P_n C_n$ be the total number of contacts made available by FSA n. We begin by exploring the absolute number of contacts formed between pairs of FSAs, denoted $X_{nn'}$. Consider X_{DE}. If we assumed

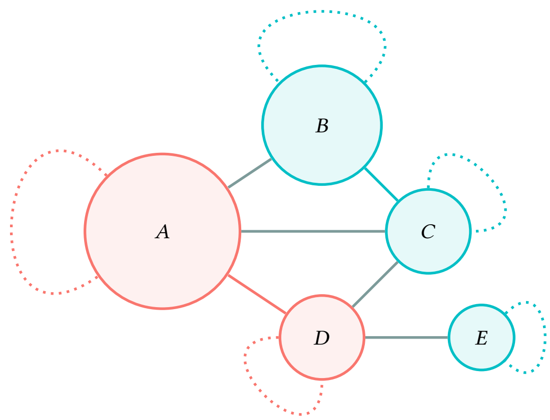

Figure 1: Diagram of toy FSA network: size (area), adjacency/external contacts (solid lines), internal contacts (dotted), and groupings (colour)

proportionate mixing between adjacent FSAs, then from D's perspective:

$$X_{DE} = Q_D \frac{Q_E}{Q_D + Q_A + Q_C + Q_E} \tag{2}$$

But from E's perspective:

$$X_{ED} = Q_E \frac{Q_D}{Q_D + Q_E} \tag{3}$$

Yet, we must have balanced contacts, $X_{DE} = X_{ED}$. In the approach above, "highly connected" FSAs (e.g. D) would expect fewer contacts with each adjacent FSA, whereas "minimally connected" FSAs (e.g. E) would expect more contacts with each adjacent FSA.[1] One way to assume a "compromise" would be to define a weighted average for the number of contacts formed with weights ω:

$$\begin{aligned} X^{\omega}_{DE} = X^{\omega}_{ED} &= (Q_D Q_E) \left[\frac{\omega_D}{Q_D + Q_A + Q_C + Q_E} + \frac{\omega_E}{Q_D + Q_E} \right] \\ &= (Q_D Q_E) \frac{\omega_D T_E + \omega_E T_D}{T_D T_E}, \quad \text{where} \begin{cases} T_D = Q_D + Q_A + Q_C + Q_E \\ T_E = Q_D + Q_E \end{cases} \end{aligned} \tag{4}$$

However, a danger of the compromise approach is that the adjacent FSAs may "request" more contacts than is possible for a highly connected FSA to "provide". A safer definition of X_{DE} could use the smaller number of expected contacts expected by either FSA:

$$\begin{aligned} X^{\mathrm{m}}_{DE} &= \min\left\{X_{DE}, X_{ED}\right\} \\ &= \frac{Q_D Q_E}{\max\left\{T_D, T_E\right\}} \end{aligned} \tag{5}$$

[1] To be precise, the expected number of contacts also depends on the relative sizes of adjacent FSAs.

More generally, and considering the adjacency matrix:

$$X^{\mathrm{m}}_{nn'} = \frac{Q_n Q_{n'}}{\max\{T_n, T_{n'}\}} \Lambda_{nn'} \tag{6}$$

The implication of this approach is that the more highly connected FSA will limit the number of contacts formed with other FSAs. We can further assume that the less connected FSAs would form the remaining contacts internally (within the same FSA) by adjusting the contact matrix:

$$X_{nn'} = X^{\mathrm{m}}_{nn'} + \mathrm{diag}\left(Q_n - \sum_{n'} X^{\mathrm{m}}_{nn'}\right) \tag{7}$$

2.2 Aggregating FSAs

The next step is to obtain $M_{gg'}$ from $X_{nn'}$. The total number of contacts formed between group g and group g' can be defined as:[2]

$$X_{gg'} = \sum_{n \in S_g} \sum_{n' \in S_{g'}} X_{nn'} \tag{8}$$

To resolve the mixing matrix $M_{gg'}$, we first define the total number of contacts made available by each group g as the sum of constitutive FSAs:

$$Q_g = \sum_{n \in S_g} Q_n \tag{9}$$

Then, since $X_{gg'}$ is defined as the product of Q_g with $M_{gg'}$ (constraint 3), we can define $M_{gg'}$ as:

$$M_{gg'} = \frac{X_{gg'}}{Q_g} \tag{10}$$

Following the approach above for the toy network, we obtain:

$$X_{nn'} = \begin{bmatrix} 7.81 & 4.59 & 2.30 & 2.30 & 0 \\ 4.59 & 4.05 & 1.35 & 0 & 0 \\ 2.30 & 1.35 & 0.68 & 0.68 & 0 \\ 2.30 & 0 & 0.68 & 1.53 & 0.50 \\ 0 & 0 & 0 & 0.50 & 2.50 \end{bmatrix} \tag{11a}$$

$$X_{gg'} = \begin{bmatrix} 13.93 & 8.07 \\ 8.07 & 9.93 \end{bmatrix} \tag{11b}$$

$$Q_g = \begin{bmatrix} 22 & 18 \end{bmatrix} \tag{11c}$$

$$M_{gg'} = \begin{bmatrix} .63 & .37 \\ .45 & .55 \end{bmatrix} \tag{11d}$$

[2] It may seem like we are double-counting off-diagonal elements of $X_{nn'}$, since these represent the same contacts as across the diagonal. In fact, the diagonal elements are already effectively double-counted due to forming contacts with themselves.

Also, if Z_{ng} is a $N \times G$ indicator matrix, then $X_{gg'}$ can also be obtained by matrix multiplication: $Z_{ng} X_{nn'} Z^{\top}_{n'g'}$.

3 More Types of Mixing

In Section 2, we assumed quasi-proportionate mixing amongst adjacent FSAs. We might also assume some global proportionate (r) mixing across all FSAs (such as for individuals travelling long distances), or some strictly assortative (i) mixing within FSAs (such as household contacts). The mixing matrices for these cases are:

$$M^r_{gg'} = \frac{Q_{g'}}{\sum_{g'} Q_{g'}}, \qquad M^i_{gg'} = \begin{cases} 1 & g = g' \\ 0 & g \neq g' \end{cases} \tag{12}$$

Assuming ϵ_r, ϵ_i, and ϵ_a, proportions of individuals form contacts through global proportionate, strictly assortative, and adjacent (a) mixing, we can define a final $M_{gg'}$ as:[3]

$$M_{gg'} = \epsilon_r M^r_{gg'} + \epsilon_i M^i_{gg'} + \epsilon_a M^a_{gg'} \tag{13}$$

[3] It can be shown that this approach can be implemented in terms of either FSAs (nn') or groups (gg') with equivalent results.