Contact Matrix Generation

Shamik Bose

July 27, 2020

1 Introduction

This document outlines the method employed in this project to generate contact matrices for a given population in multiple environments (schools and workplaces, in this case). A social contact matrix is defined as the mean number of contacts *per day reported between each age group*[2]. The rest of this work is based on works from [1] and [3]

with no distinction between mean different contacts, and mean repeated contacts, or some combination. Usually, the assumption is mean different contacts, but that is highly unlikely.

2 The Math and the Data

Our synthetic population is divided into 16 age groups, of which the first 15 are of size 5 and the last one includes all people over 75. This creates the following age groups:
$0 - 4, 5 - 9, 10 - 14, 15 - 19, 20 - 24, 25 - 29, 30 - 34, 35 - 39, 40 - 44, 45 - 49, 50 - 54, 55 - 59, 60 - 64, 65 - 69, 70 - 74, 75+$ *I think we could simplify by only using the groups 20-29, 30-39, 40-49, 50-59, 60-69, 70+*
As a starting point, we use the projected social contact matrices as created in [3] for the United States. We use three matrices, one each for schools, workplaces and overall contact. To project them to our synthetic population, we use the following equation from [1]

M_{i,j} is not clearly defined. It is the mean number of contacts between an individual i and an age-group j. This leads to the necessary symmetry constraint that must be satisfied: M_{i,j} N_i = M_{j,i} N_j

M_{ij} is the matrix obtained from [3]? It is not clear.

$$M'_{i,j} = M_{i,j}\, \frac{N N'_j}{N_j N'} \qquad (1)$$

*Does M'_{i,j} * N'_i == M'_{j,i} * N'_j? That is required to satisfy symmetry properties. Is M_{i,j} * N_i == M_{j,i} * N_j*

where the terms are defined as follows:

You mean M_{ij}?

- $M'_{i,j}$ is the entry for the original contact matrix M for contacts from group i to group j.

- N and N' are the total number of people in the original and the new demographic respectively

- N_j and N'_j are the number of people in age group j in the original and new demographic respectively

To this end, we use the values of N and n_j from the American Community Survey[1] data. The data is shown in Tables 1, 2 and 3. In Table 3, we can see that some of the data is for larger age groups than what we have (35-44 and 65-74). To align it with our data, we split the values in half to use those values as our ratios. At the time of writing this, I realized I should have scaled that according to the population ratios. That will be incorporated into the script soon. Using this data and Equation 1, we can build a contact matrix for the synthetic population when we start with the contact matrices provided for the United States in [3]. These values are given in the Supplemental Excel sheet Using the contact matrices, the population distribution for US, the population distribution for Leon County and the 1, we can project the contact matrices to Leon County. Since all age groups are not present in all schools and workplaces, the contact matrices are created for every school and workplace separately. To this end, we maintain the age distribution for every school in the synthetic data, aligning the age groups with those used in the contact matrices. Let us consider a school X with age groups k. The contact matrix for X is computed as follows:

$$CM_X[idx][j] = CM_{BaseSchool}[idx][j] * AP_x[idx] / AP_{USASchool}[idx] \forall j \in k \qquad (2)$$

use index i (why use idx?)

where

What is AP_x[idx]?

---

[1] https://data.census.gov/cedsci/table?q=dp&tid=ACSDP1Y2018.DP05

*Your formulas on the left as a capital "X"
The formulas on the right has little "x". Is that a mistype?*

Age Group	USA	Leon County, FL
0-4	19,646,315	15,704
5-9	19,805,900	16,472
10-14	21,392,922	15,348
15-19	21,445,493	19,532
20-24	21,717962	38,376
25-29	23,320,702	20,824
30-34	22,023,972	15,625
35-39	21,571,302	15,107
40-44	19,927,151	15,144
45-49	20,733,440	16.975
50-54	20,871,804	17,362
55-59	21,624,541	16,073
60-64	20,662,821	13,334
65-69	17,107,288	8,569
70-74	13,464,025	5,454
75+	21,851,801	10,643
Total population	327,167,439	260,542

In the paper, if we use this data, we should plot it.

Table 1: Age distribution in USA and Leon County

Age Group	USA	Leon County, FL
3 to 4 year olds in school	3,927,637	4,261
5 to 9 year olds in school	18,876,942	15,823
10 to 14 year olds in school	20,984,583	15,050
15 to 18 year olds in school	17,100,878	10,212
Total	60,890,040	45,346

Plots would be nice.

Table 2: Age distribution in schools in USA and Leon County

AP_{USA}? AP_{X} ?

Age Group	USA	Leon County, FL
16 to 19 years	17,309,142	5,043
20 to 24 years	21,717,962	23,321
25 to 29 years	23,320,702	17,008
30 to 34 years	22,023,972	11,968
35 to 44 years	41,498,453	24,337
45 to 54 years	41,605,244	27,891
55 to 59 years	21,624,541	11,889
60 to 64 years	20,662,821	7,700
65 to 74 years	30,571,313	4,250
75 years and over	21,851,801	778
Total	262,185,951	134,285

Plots would be nice, comparing the two columns, perhaps using percentages.

Table 3: Age distribution in workplaces in USA and Leon County

- idx is the index of every age present in the school in the list of following age groups
  $0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64,$
  $65-69, 70-74, 75+$

- CM_X is the contact matrix for school X,

- $CM_{BaseSchool}$ is the contact matrix provided for the school environment given in [3],

- AP_X is the age distribution for X

- $AP_{USASchool}$ is the age distribution for the US in schools

All contacts to and with absent age groups are set to zero. Contact matrices for workplaces in Leon County can also be calculated using Eq.2 with the following changes

- Replace $CM_{BaseSchool}$ with $CM_{BaseWorkplace}$ from [3]

- Replace $AP_{USASchool}$ with $AP_{USAWorkplace}$

The contact matrices generated for Leon County schools and workplaces are available on the repository under `ContactMatrices/Leon/` as both .csv and serialized pickle files. An example contact matrix for a school(ID: 450147248) in Leon County is given below:

	0-4	5-9	10-14	15-19
0-4	0	0	0	0
5-9	0	0.935826768	0.278592537	0.026150962
10-14	0	0.218607436	6.792703149	0.178885828
15-19	0	0.010554706	2.380692579	6.952417816

this means a child aged 5-9 only has contact on average with all kids in the same class, per day. Does that sound reasonable to you?

Table 4: Contact Matrix for school 450147248 in Leon County, FL

References

[1] Sergio Arregui, Alberto Aleta, Joaquín Sanz, and Yamir Moreno. Projecting social contact matrices to different demographic structures. *PLoS computational biology*, 14(12):e1006638, 2018.

[2] Ken Eames, Natasha Tilston-Lunel (Tilston), Ellen Brooks-Pollock, and W Edmunds. Measured dynamic social contact patterns explain the spread of h1n1v influenza. *PLoS computational biology*, 8:e1002425, 03 2012.

[3] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, 13(9):e1005697, 2017.