PyRossGeo: The geographical compartmental model for infective diseases

Lukas Kikuchi and Fernando Caballero

May 11, 2020

1 Model

We consider an extension of epidemiological compartmental models to geographical networks. For simplicity we consider an SIR model, but the model can be generalized easily to any of its variations. We consider a network of locations $i = 1, \ldots, M$, where typically $M \sim 10^3$, that represent geographical nodes in which we assume populations are well mixed, and therefore can be described using SIR dynamics. To consider commuting we expand the variables of the SIR model to tensors, $S \to S^{\alpha}_{ij}$, where i refers to where they live, and j where they are currently located, and where we allow for different age brackets through the superscript α. The i and j indices are required for book-keeping purposes, in order to ensure that all residents return to their homes at the end of a day.

The second main ingredient of our model is what we call the *commuterverse*, designed to account for infections incurred during traveling between different nodes of the network. The commuterverse is built as a secondary network that introduces a new node in each link between connected nodes of the geographical network. For instance, consider the scenario where there are susceptibles in age brackets $\alpha = 1, \ldots, K$, who are currently located at node j, who live at node i, and wants to travel to node k. Then these people will first have to spend the requisite amount of commuting time in a commuting node $S^{\alpha}_{i(j\to k)}$, before arriving at S^{α}_{ik}. The dynamics of the commuting node is slightly different from the normal SIR dynamics.

Let O mark a general class (S, I or R in this case), then the sub- and superscripts of the nodes represent

$$\begin{aligned} O^{\alpha}_{ij} &\Longrightarrow O^{\text{age bracket}}_{\text{home,current location}} \\ O^{\alpha}_{i(j\to k)} &\Longrightarrow O^{\text{age bracket}}_{\text{home,commuting from } j \text{ to } k} \end{aligned}.$$

In words, O^{α}_{ij} *is the number of people living at i and are located at j, are of the age bracket α and the class O. $O^{\alpha}_{i(j\to k)}$ is the number of people living at i and are travelling from j to k, are of the age bracket α and the class O.* In our nomenclature, all residents of i is a *community*, each j is a *geographical* *node* $gc(j,\alpha)$, the tuple $n(i,j,\alpha)$ is a *node*, and the tuple $cn(i,(j\to k),\alpha)$ is a *commuterverse node* or *cnode*. The total population of a community is thus

$$N_i = \sum_{O,j,\alpha} O^{\alpha}_{ij} + \sum_{O,j,k,\alpha} O^{\alpha}_{i(j\to k)}.$$

The movement between nodes is modelled using the transport matrix $T^{O,\alpha}_{i,j(j\to k)}$ which contains the rate at which people who live in i move from location j to location to the commuterverse $j\to k$, where O marks the class (S, I or R in this case). $T^{O,\alpha}_{i,(j\to k)k}$ gives the rate at which people move from the $j\to k$ commuterverse into node k.

An example of a commuter network is shown in Figure 1. The figure illustrates that there are separate commuterverses for people resident at different nodes, as the 4 and 1 nodes both have their own $(1\to 4)$ and $(4\to 1)$ commuterverses.

Strictly speaking, $T^{O,\alpha}_{i,ab}$ should be seen as a matrix with components $a,b\in \{1,2,\ldots,M\}\cup\{(j\to k)\ :\ j=1,2,\ldots,M,\ k=1,2,\ldots,M\}$, for each i. This is of course a very large matrix with around $M\times M^2\sim 10^9$ components. In reality $T^{O,\alpha}_{i,ab}$ will only have around ~ 10 non-zero elements for each i, and this sparsity will be taken advantage of in implementations.

Transport terms of the variables O^{α}_{ij} and $O^{\alpha}_{i(j\to k)}$ comes into their equations for each location as follows

$$\begin{aligned}\dot{O}^{\alpha}_{ij} =&\text{local terms} + \sum_k [T^{O,\alpha}_{i,(k\to j)j} - T^{O,\alpha}_{i,j(j\to k)}]\\ \dot{O}^{\alpha}_{i(j\to k)} =&\text{local terms} + T^{O,\alpha}_{i,j(j\to k)} - T^{O,\alpha}_{i,(j\to k)k}\end{aligned}$$

The local terms, in the case of the variable O^{r}_{ij} are the local terms of the SIR model, while the local terms of the variable $O^{\alpha}_{i(j\to k)}$ are only the linear terms of the SIR model and the nonlinearities responsible for infection.

Neglecting age-contact structure, the geographical compartmental model is

$$\begin{array}{ccccc}
\dot{S}_{ij} & = & -\beta\frac{I_j}{N_j}S_{ij} & + & \sum_k [T^{S}_{i,(k\to j)j} - T^{S}_{i,j(j\to k)}]\\
\dot{I}_{ij} & = & \beta\frac{I_j}{N_j}S_{ij} - \gamma I^{\alpha}_{ij} & + & \sum_k [T^{I}_{i,(k\to j)j} - T^{I}_{i,j(j\to k)}]\\
\dot{R}_{ij} & = & \gamma I^{\alpha}_{ij} & + & \sum_k [T^{R}_{i,(k\to j)j} - T^{R}_{i,j(j\to k)}]\\
\dot{S}_{i(j\to k)} & = & -\beta\frac{I_{(k)}}{N_{(k)}}S_{i(j\to k)} & + & T^{S}_{i,j(j\to k)} - T^{S}_{i,(j\to k)k}\\
\dot{I}_{i(j\to k)} & = & \beta\frac{I_{(k)}}{N_{(k)}}S_{i(j\to k)} - \gamma I_{i(j\to k)} & + & T^{I}_{i,j(j\to k)} - T^{I}_{i,(j\to k)k}\\
\dot{R}_{i(j\to k)} & = & \gamma I_{i(j\to k)} & + & T^{R}_{i,j(j\to k)} - T^{R}_{i,(j\to k)k}
\end{array}$$

where $I_j=\sum_r I_{rj}$, $N_j=\sum_r N_{rj}$, $I_{(k)}=\sum_{l,m} I_{l(m\to k)}$ and $N_{(k)}=\sum_{l,m} N_{l(m\to k)}$, and where β and γ are free parameters.

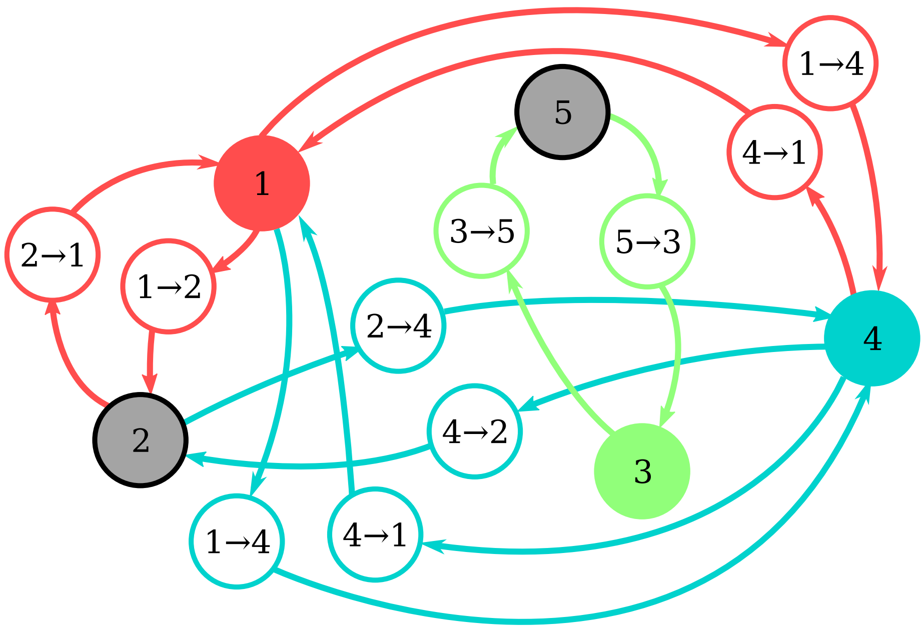

Figure 1: An example of a 4-node system. The graph shows the commuter networks of node 1 (red), 3 (green) and 4 (blue). The commuter networks of 2 and 5 are not displayed. Hollow nodes denote commuterverse nodes, and filled nodes denote geographic nodes. Note that there are both red and blue $(1 \to 4)$ and $(4 \to 1)$ commuterverses.

We can write the model in a compactified manner by letting $\boldsymbol{X}_{ij}^{\alpha} = (S_{ij}^{\alpha}, I_{ij}^{\alpha}, R_{ij}^{\alpha})^T$, where we have now included age-structure. We get

$$\begin{array}{rcccc} \dot{\boldsymbol{X}}_{ij}^{\alpha} & = & \boldsymbol{F}_{ij}^{\alpha}(t, \boldsymbol{X}_{ij}^{\alpha}, \boldsymbol{\theta}_{ij}) & + & \sum_k [\boldsymbol{T}_{i,(k\to j)j}^{\alpha} - \boldsymbol{T}_{i,j(j\to k)}^{\alpha}] \\ \dot{\boldsymbol{X}}_{i(j\to k)}^{\alpha} & = & \boldsymbol{G}_{i(j\to k)}^{\alpha}(t, \boldsymbol{X}_{i(j\to k)}^{\alpha}, \boldsymbol{\theta}_{i(j\to k)}) & + & \boldsymbol{T}_{i,j(j\to k)}^{\alpha} - \boldsymbol{T}_{i,(j\to k)k}^{\alpha} \end{array}$$

where

$$\boldsymbol{F}_{ij}^{\alpha}(t, \boldsymbol{X}_{ij}^{\alpha}, \boldsymbol{\theta}_{ij}) = \begin{pmatrix} -\lambda_j^{\alpha} S_{ij}^{\alpha} \\ \lambda_j^{\alpha} S_{ij}^{\alpha} - \gamma I_{ij}^{\alpha} \\ \gamma I_{ij}^{\alpha} \end{pmatrix}$$

$$\boldsymbol{G}_{i(j\to k)}^{\alpha}(t, \boldsymbol{X}_{i(j\to k)}^{\alpha}, \boldsymbol{\theta}_{i(j\to k)}) = \begin{pmatrix} -\tau_k^{\alpha} S_{i(j\to k)}^{\alpha} \\ \tau_k^{\alpha} S_{i(j\to k)}^{\alpha} - \gamma I_{i(j\to k)}^{\alpha} \\ \gamma I_{i(j\to k)}^{\alpha} \end{pmatrix}$$

and

$$\begin{aligned} \tau_k^{\alpha} &= \beta \frac{1}{\sum_{l,m} N_{l(m\to k)}} \sum_{\gamma=1}^{K} C^{\alpha\gamma} \sum_{l,m} I_{l(m\to k)}^{\gamma} \\ \lambda_j^{\alpha} &= \beta \sum_{\gamma=1}^{K} C^{\alpha\gamma} \frac{\sum_k I_{kj}^{\gamma}}{\sum_k N_{kj}^{\gamma}} \end{aligned} .$$

2 Coding architecture

The number of geographical locations is $M \sim 10^3$, and thus the number of possible nodes $\boldsymbol{X}_{ij}^{\alpha}$ are $M^2 \sim 10^6$. Additionally, the number of possible commuter nodes $\boldsymbol{X}_{i(j\to k)}^{\alpha}$ are of the order $M \times M^2 \sim 10^9$. In other words, the matrices are too large to be dealt with computationally. The matrices are also very sparse (the nodes $\boldsymbol{X}_{ij}^{\alpha}$ will on average only interact with very few other nodes ~ 10).

To set the stage for the optimisations to come, we will introduce some terminology. Let V_i^{α} be the set of nodes in the commuter network of the i-population in age-bracket α. In other words, the set of nodes that the i-population travel to. Formally, we define

$$V_i^{\alpha} = \{j \ : \ \phi_i^{\alpha}(i,j) = 1\}$$

where $\phi_i^{\alpha}(j,k) = 1$ if i and j are part of the same connected commuter-network, and $\phi_i^{\alpha}(j,k) = 0$ otherwise. We define the analogous set for the commuterverses

$$W_i^{\alpha} = \{(j \to k) \ : \ \psi_i^{\alpha}(j,k)\}$$

where $\psi_i^{\alpha}(j,k) = 1$ if $\exists t$ such that $T_{i,j(j\to k)} \neq 0$ or $T_{i,k(k\to j)} \neq 0$, and $\psi_i^{\alpha}(j,k) = 0$ otherwise. Finally, let $V_{\text{tot}}^{\alpha} = \{(i,j) \ : \ j \in V_i^{\alpha}, \ i = 1, \ldots, M\}$. and $W_{\text{tot}}^{\alpha} = \{(i, j \to k) \ : \ j \to k \in W_i^{\alpha}, \ i = 1, \ldots, M\}$.

The main sources of potential computational complexity are from the sums in λ_j, $\tau_{j\to k}$ and in $\dot{\boldsymbol{X}}^{\alpha}_{ij}$. Regard the the sums of the former two

$$\begin{aligned}\tau^{\alpha}_{j\to k} &= \left(\frac{1}{\sum_{l,m} N_{l,m\to k}} \sum_{\gamma=1}^{K} C_{\alpha\gamma} \sum_{l,m} I^{\gamma}_{l(m\to k)}\right) \beta_{j\to k} \\ \lambda^{\alpha}_{j} &= \beta \sum_{\gamma=1}^{M} C_{\alpha\gamma} \sum_{k} \frac{I^{\gamma}_{kj}}{N_j}\end{aligned} . \tag{1}$$

If done naively, $\sum_{l,m}$ and $\sum_k$ would run over the entirety of the $M \sim 10^3$ nodes. In reality there will only be on the order of ~ 10 non-zero summands, and the set of non-zero summands can also be determined during pre-processing. Let $v^{\alpha}_{j} = \{k \;:\; j \in V^{\alpha}_{k}\}$, in other words the set of k-populations who has j in their commuter-network. Let $w_{j\to k} = \{(l,m) \;:\; (m \to k) \in W^{\alpha}_{l}\}$, in other words the set of l-populations that has a commuter-verse linking to node k. We can replace the sums in Eq. 1 with

$$\begin{aligned}\tau^{\alpha}_{j\to k} &= \left(\frac{1}{\sum_{l,m} N_{l,m\to k}} \sum_{\gamma=1}^{K} C_{\alpha\gamma} \sum_{l,m\in w_{j\to k}} I^{\gamma}_{l(m\to k)}\right) \beta_{j\to k} \\ \lambda^{\alpha}_{j} &= \beta \sum_{\gamma=1}^{M} C_{\alpha\gamma} \sum_{k\in v^{\alpha}_{j}} \frac{I^{\gamma}_{kj}}{N_j}\end{aligned} .$$

Consider now the sum

$$\dot{\boldsymbol{X}}^{\alpha}_{ij} = \cdots + \sum_{k} [\boldsymbol{T}^{\alpha}_{i,(k\to j)j} - \boldsymbol{T}^{\alpha}_{i,j(j\to k)}]$$

this can be replaced by

$$\dot{\boldsymbol{X}}^{\alpha}_{ij} = \cdots + \sum_{k\in \overset{\rightarrow}{\chi}{}^{\alpha}_{ij}} \boldsymbol{T}^{\alpha}_{i,(k\to j)j} - \sum_{k\in \overset{\leftarrow}{\chi}{}^{\alpha}_{ij}} \boldsymbol{T}^{\alpha}_{i,j(j\to k)}]$$

where $\overset{\rightarrow}{\chi}{}^{\alpha}_{ij}$ is the set of commuterverses incoming into j, for population i, and $\overset{\leftarrow}{\chi}{}^{\alpha}_{ij}$ is the set of commuterverses leading out of j.

Now v^{α}_{j}, $w_{j\to k}$, $\overset{\rightarrow}{\chi}{}^{\alpha}_{ij}$ and $\overset{\leftarrow}{\chi}{}^{\alpha}_{ij}$ can all be computed during a pre-processing step, greatly reducing the complexity of the model. Furthermore, note that $\tau^{\alpha}_{j\to k}$ and λ^{α}_{j} do not have a home index i, so they only need to computed once every loop, and can then be used for every i at j.

2.1 Pseudocode

The core element of the implementation is the $\texttt{node}^{\alpha}_{ij}$ struct. It will contain pointers to all the necessary information to integrate its own local dynamics.

```
node^α_ij
 → X^α_ij
 → λ^α_j
 → List of incoming commuter nodes: {X^α_i(k→j)}_{k ∈ →χ^α_ij}
 → List of outgoing commuter nodes: {X^α_i(j→k)}_{k ∈ ←χ^α_ij}
```

The commuter nodes are

```
cnode^α_i(j→k)
 → X^α_i(j→k)
 → node^α_ij
 → node^α_ik
 → T^α_i,j(j→k)
 → T^α_i,(j→k)k
 → τ^α_j→k
```

The simulation is then

```
# Preprocessing

Compute v^α_j, w_j→k, →χ^α_ij and ←χ^α_ij.

# Simulation

Main loop
  Compute λ^α_j, for all j = 1,...,M, and α = 1,...,K
  Compute τ^α_j→k, for all j,k ∈ ∪_{i=1}^M W^α_i, and α = 1,...,K

  For i,j ∈ V^α_tot
    Update node^α_ij

  For i,j → k ∈ W^α_tot
    Update cnode^α_i(j→k)
```