# CONDENSATION OF SIR MODELS ON GRAPHS

Félicien Bourdin

# CONDENSATION OF SIR MODELS ON GRAPHS

FÉLICIEN BOURDIN, LMO - LABORATOIRE DE MATHÉMATIQUES D'ORSAY

Abstract. We develop deterministic SIR models on graphs where we have condensed different nodes of the network into communities of people. We analyze this condensation process within the framework of resistive networks and that of epidemiological models by comparing the condensed model to the underlying microscopic model. We identify qualitative properties that should satisfy the network for the difference between the predictions of those models to be small.

## 1. Introduction

**SIR models on graphs.** Due to the Covid-19 outbreak, there have been a recent better understanding of the so-called SIR models, introduced in the historical paper [8] to describe the dynamics of an epidemy in a population. In the most basic SIR model, under panmictic hypotheses on the population, one assume that the ratios of susceptible, infected, and recovered people, $S$, $I$ and $R$, evolve according the following differential equations:

$$\left\{\begin{array}{ll} \dfrac{dS}{dt} & = -\beta SI \\ \dfrac{dI}{dt} & = \beta SI - \gamma I \\ \dfrac{dR}{dt} & = \gamma I \end{array}\right. \tag{1}$$

where $\beta$ and $\gamma$ are constants of the model. At the other end of the range of granularity, probabilistic SIR models (see [9], [17], [11]) rely on an individual-based description of the outbreak, where the states $S$, $I$ or $R$ of the individuals are random variables and can change from $S$ to $I$ when a susceptible person has a contact with an infected one, and from $I$ to $R$ according to an exponential random variable. The random outbreak then spreads on the underlying graph that is assumed to be known a priori. The qualitative results are then obtained by Monte-Carlo arguments, by averaging the scenarii over a great number of realizations.

We deal here with deterministic SIR models on a graph $G$ where we define for each node $X$ epidemiologic quantities $S_X$, $I_X$, and $R_X$, like in [1], [2]. In addition to internal contaminations, nodes have interactions through the network that can lead to a global outbreak starting from a single node. These nodes might be singletons, geographical groups, groups of age, or more specific communities in structured situations, as classes in a school, or teams of coworkers in a company. In such a model, the equation on $I$ writes

$$I^{n+1} - I^n - S^n \mathcal{L} I^n = \beta S^n I^n - \gamma I^n, \tag{2}$$

2000 *Mathematics Subject Classification.* 92D30, 35R02, 94C15.

*Key words and phrases.* SIR models on graphs, resistive networks, Poisson problem, Laplace equation.

where $\beta, \gamma$ are epidemiological parameters and $\mathcal{L}$ being a propagation operator on $G$ associated to $\beta$. For a function $\phi$ defined on the nodes, it is defined as

$$\mathcal{L}\phi(X) = \sum_{Y \sim X} \beta_{XY}(\phi(Y) - \phi(X)), \tag{3}$$

where $Y \sim X$ means that $(X, Y)$ is an edge of $G$. From this formulation, $\mathcal{L}$ plays the role of a discrete elliptic operator (one can see for instance that in dimension 1 on a linear graph with constant $\beta_{XY}$, it is the discretization of the Laplace operator). One can thus read this equation on $I$ as the discretization of a generalized heat equation, with a diffusion coefficient $S_X^n$ that varies in space, and a source term that depends on the $S$ and $I$ ratios at each node.

**Propagation models.** The modeling of the propagation of quantities on a finite set of connected points leads to a general form of problem that can take two forms. Dynamical problems such as the propagation of heat on a graph or the establishment of the equilibrium on a network of capacitors involve a time derivative. They are mostly of the form

$$\partial_t \phi - \kappa \mathcal{L}\phi = f, \tag{4}$$

where $\kappa$ is a diffusion constant, $\mathcal{L}$ is a discrete Laplace operator defined on the graph, and $f$ a potentially time-varying source, see [10], [7]. In the case of a very fast diffusion, that is to say $\kappa \gg 1$, the heat equation (4) reduces to the discrete Poisson-type equation that does not involve anymore a time dependance:

$$-\mathcal{L}\phi = f. \tag{5}$$

This static model describes for instance the propagation of a current in a resistive circuit under the approximation of quasi-stationary regimes, see [15]. Solving this equation consists in inverting the discrete $-\mathcal{L}$ operator, that is to say in inverting a matrix. The diagonalization of this operator has been widely studied, see for instance [13] for a survey on the many results there are on the spectrum of the graph Laplacian. The discrete heat equation (4) can be seen as the description of the transitional regime, and the Poisson equation (5) as the long term behaviour.

**Model condensation on networks.** When modeling the spread of quantities on large populations, one may want to condense several nodes that tend to play a similar role in the network into one only node to have fewer parameters. In electrical circuits, those shall be the nodes linked by very high conductances. In epidemiological models, we expect a people that are in permanent contact to have the same probability to be infected. In other words, we expect groups of people that stay together to be described by same epidemiological quantities. Conversly, from a modeling standpoint, some situations involve networks that are too large to have a knowledge of every interaction. For instance, when dealing with a general model of propagation of an outbreak in a large population, one can not expect to build a model from a real description of the contacts between any pair of people. We must thus choose an a priori, for instance by drawing randomly an underlying graph of contacts. This kind of choice is arbirtary and one can not control the dependance of the model on this graph. One may rather want to define a mesoscopic model by grouping the nodes according to a feature (for instance age, or geography in epidemiologic models) and thereafter to model the interactions between those groups by defining mesoscopic interaction matrices.

Generically, we speak of model condensation on a microscopic graph $G = (V, E)$ when we have a law a evolution of a quantity $\phi_t$ that depends on quantities $K_{XY}$ attributed to the edges $(X, Y)$ of $G$. We then define a condensed graph $\bar{G} = (\bar{V}, \bar{E})$ by gathering some nodes of $G$, and a new

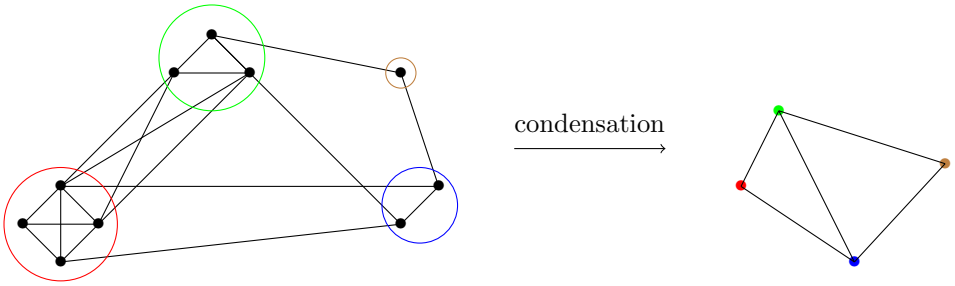


FIGURE 1. On the left handisde, a given graph, the length $l_{XY}$ of the edges being inversly proportional to their connectivity $K_{XY}$. The red circles delimits the communities we chose to condense into one only node. On the right handside, the corresponding condensed graph.

matrix $\bar{K}$ from $K$ and from a law of summation of the blocs of the previous matrix $K$ that depends on the problem (summation, mean, ...). Finally, we derive a new law of evolution of $\bar{\phi}_t$ defined on $\bar{G}$: we say that the evolution of $\bar{\phi}$ on $\bar{G}$ is the condensed model of the microscopic one, that is the evolution of $\phi$ on $G$. The example of the condensation of a graph is depicted on figure 1. As the length of the edges are inversly proportional to their conductivity, we chose to condense the dots in the four colored circles into only four nodes. For the summation law, we chose the addition of the conductances linking nodes of a group to another. We will study these two main questions raised by this condensation process:

- How to measure the error made by condensing, that is to say how to compare $\bar{\phi}$ and $\phi$ ?
- Given a microscopic graph and an associated matrix, what points should be gathered to have $\bar{\phi}$ close to $\phi$ ?

The rest of the paper is organized as follows. Section 2 is dedicated to an analysis of the condensation of resistive networks, for two distinct problems: a static Poisson equation where we inject current into internal nodes of the system, and a homogeneous Laplace problem where we impose a difference of potential at the extremities of the system. In section 3 is defined a non-linear SIR model on subcommunities based on probabilistic arguments. We then compare it to the native purely probabilistic model in the case where subcommunities are singletons, and proceed to the analysis of the condensation of this model in 4.

## 2. Condensation of resistive networks

In this section, we adress issues on the problematics of the condensation of a resistive network into subclasses of points.
A resistive network is composed of a symmetric connected graph $G = (V, E)$ and a symmetric matrix of conductances $K$ between nodes of $G$. We set $K_{xy} = 0$ if $(x, y)$ is not an edge of $G$, and $K_{xy} > 0$ when $(x, y) \in E$. We distinguish between a set $\Gamma \subset V$ of external nodes, and internal nodes in $V \setminus \Gamma$. We shall make later assumptions on $\Gamma$ so that the equations we write on $G$ are well posed.

### 2.1. Discrete Poisson equation on a resistive network.

In this subsection, we deal with the situation of a network where each internal node $x \in V \setminus \Gamma$ is plugged to a source $f_x$. We assume

that every exernal node $x \in \Gamma$ is connected to the earth and thus set to a potential equal to 0. We shall describe the behaviour of this system using the following two laws ruling the electrical circuits:

$$i = cu \quad \text{(Ohm's law)} \tag{6}$$

where $i$ is the intensity flowing through an edge, $c$ its conductance, and $u$ the difference of the potential at its extremities, and

$$\sum_{k=1}^{n} i_k^x = f_x \quad \text{(Kirchoff's circuits law)} \tag{7}$$

where the $(i_k^x)_{k=1,\dots,n}$ are the signed intensity meeting at any point $x$ in the circuit. Under these laws, finding the potential in the circuit results in solving the following discrete Poisson problem:

$$\left\{ \begin{array}{rcll} -\mathcal{L}_K u & = & f & \text{on } V \setminus \Gamma \\ u & = & 0 & \text{on } \Gamma \end{array} \right. \tag{8}$$

where $\mathcal{L}_K$ is the following discrete Laplace operator defined by

$$\mathcal{L}_K u(x) = \sum_{y \in G} K_{xy}(u_y - u_x) \quad \forall x \in G. \tag{9}$$

The first line of equation (8) expresses the conservation of the total intensity at every node (Kirchoff's law), whereas the boundary condition on $\Gamma$ is needed for uniqueness of the solution.

**Proposition 1.** *As soon as $G$ is connected,* $\mathrm{Ker}\mathcal{L}_K$ *is the set of constant functions on $G$ and thus has a dimension equal to 1.*

*Proof.* Let $u \in G^{\mathbb{R}}$ such that $\mathcal{L}_K u = 0$. One can write from (9) for every $x \in G$

$$u_x = \frac{\sum_{y \in G} K_{xy} u_y}{\sum_{y \in G} K_{xy}}. \tag{10}$$

$u_x$ is thus in the convex hull of the values of its neighbours. Let $x \in G$, and $x_0 \in G$ such that $u$ reachs its maximum in $x_0$. As $G$ is connected, let $x_0, \dots, x_k = x$ be a path in $G$ from $x_0$ to $x$. One shows by induction that for every $j = 0, \dots, k-1$, $u_{x_j} = u_{x_0}$. If $u_{x_j} = u_{x_0}$, as $u_{x_j}$ is in the convex hull of its neighbours and $(x_j, x_{j+1}) \in E$, we get $u_{x_j} = u_{x_{j+1}}$. Then $u_{x_0} = u_x$ and u is constant. □

We now define the process of condensation of the network according to a partition into given subcommunities: let $\mathcal{P}$ be a partition of the nodes of $G$ into classes, with $\Gamma \in \mathcal{P}$. We then identify every node in each class, by forcing them to have the same potential (that is to say we connect them by an infinite conductance). As in electrical circuits, where two conductances in parallel $c_1$ and $c_2$ between the same nodes are equivalent to a single conductance $c_{eq} = c_1 + c_2$, we set the equivalent conductance between the subcommunities $X$ and $Y$ as

$$\overline{K_{XY}} = \sum_{x \in X, y \in Y} K_{xy}. \tag{11}$$

Setting $f_X = \sum_{x \in X} f_x$, we now look for the solution of

$$\left\{ \begin{array}{rcll} -\mathcal{L}_{\bar{K}} w & = & f & \text{on } \mathcal{P} \setminus \{\Gamma\} \\ w_\Gamma & = & 0 & \end{array} \right. \tag{12}$$

where $w : \mathcal{P} \longrightarrow \mathbb{R}$. We then define $\bar{u} : V \longrightarrow \mathbb{R}$ the interpolation of $w$ on $G$ by $\bar{u}_x = w_X$ for every $X \in \mathcal{P}$ and $x \in X$.

**Remark 1.** *The process of computing the solution of the discrete version of a PDE by computations on a coarser network can be related to the so-called multigrid methods. Commonly used to approximate solutions of linear PDE's, they rely on several computations of the error of an iterative method on different meshes to get a fast and cheap convergence of all its frequencies, see* [16] *for a detailed introduction. The starting point of the multigrid idea is the observation that when using an iterative algorithm to solve a linear PDE, the error quickly becomes smooth, and not necessarly small. Therefore, one can deduce that on a given mesh only the high frequencies are efficiently computed, the notion of "high frequency" being a feature of the mesh more than an absolute concept depending on the operator only. The resulting idea is thus the computation of the low frequencies on a coarser mesh, provided satisfying operators of projection from a grid to another. Indeed, on a coarser grid, some of the low frequencies will become high for the new grid (the notion of "high" and "low" frequency depending on refinment of the mesh), and so become fast (and cheap) to compute. In the context of our condensation process, we may see the condensed graph as the coarser grid.*

Let us set a framework to compare the two potentials we introduced. Let

$$\mathcal{V} = \{v : V \longrightarrow \mathbb{R}\,,\ v = 0 \text{ on } \Gamma\}\,, \tag{13}$$

endowed with the norm $\|v\|_K^2 = \frac{1}{2}\sum_{x,y} K_{xy}(v_x - v_y)^2$.

**Remark 2.** *From the equivalence of norms in finite dimension, we get the existence of a constant $C > 0$ such that for every $v \in \mathcal{V}$,*

$$\|v\|_2 \leq C\|v\|_K. \tag{14}$$

*Equation* (14) *is a Poincaré-like equation, with a constant $C$ that degenerates as the graph grows, with the existence of long paths starting from the border $\Gamma$. Indeed, let us consider a path of length $n$ $(x_1, \ldots, x_n)$ with $x_1 \in \Gamma$, with local variations $(v_{x_i} - v_{x_i+1})$ of order 1. Then we may have $v_{x_n}$ of order $n$ (consider for instance a linear graph of size $n$ with $\Gamma$ being one of its extremities). Thus $C$ is going to infinity as $n$ grows.*

Like in the context of PDE's, let us then write the variational formulation of equation (8). By multiplying both sides by a field $\phi$, one has for any $x \in V \setminus \Gamma$

$$\sum_{y \in V} K_{xy}\phi_x(u_x - u_y) = f_x\phi_x. \tag{15}$$

By summation on $x$, one gets the following variational formulation of the Poisson problem:

$$\sum_{x,y \in V} K_{xy}(u_x - u_y)(\phi_x - \phi_y) = \sum_{x \in V} f_x\phi_x \tag{16}$$

One can see that equation (16) is an optimality condition for the minimization of the strictly convex functional

$$J : \left\{ \begin{array}{ccc} \mathcal{V} & \longrightarrow & \mathbb{R} \\ v & \longmapsto & \|v\|_K^2 - \sum\limits_{x \in V} f_x v_x. \end{array} \right. \tag{17}$$

We can thus define $u$ as the minimizer of $J$ on $\mathcal{V}$. Similarly, defining

$$\mathcal{V}_{\mathcal{P}} = \{v \in \mathcal{V}\,,\ v \text{ is constant on every } X \in \mathcal{P}\}\,, \tag{18}$$

we get that

$$\bar{u} = \operatorname*{argmin}_{v \in \mathcal{V}_{\mathcal{P}}} J(v). \tag{19}$$

The following proposition gives a first formal estimate of the distance between the two potentials for the $\|\cdot\|_K$ norm:

**Proposition 2.** *We have the following estimate:*

$$\|u - \bar{u}\|_K = \inf_{v \in \mathcal{V}_{\mathcal{P}}} \|u - v\|_K. \tag{20}$$

*Proof.* By Cea's lemma (see [3]), one has $\bar{u} = P_{\mathcal{V}_{\mathcal{P}}}(u)$, where $P_{\mathcal{V}_{\mathcal{P}}}$ is the orthogonal projection on $\mathcal{V}_{\mathcal{P}}$ for $\|\cdot\|_K$. □

We can thus compute $\bar{u}$ from $u$: for every $X \in \mathcal{P}$ and $x \in X$,

$$\bar{u}_x = \frac{\sum_{y \in X} K_y u_y}{\sum_{y \in X} K_y}, \tag{21}$$

where $K_y = \displaystyle\sum_{x \sim y} K_{xy}$. The condensed potential $\bar{u}$ is thus a weighted mean of $u$ on every set of the partition $\mathcal{P}$. Therefore, one gets

**Proposition 3.** *We can estimate*

$$\|u - \bar{u}\|_K = 2\left(\frac{1}{2}\sum_{X \in \mathcal{P}}\sum_{x \in X} K_x \left(u_x - \frac{\sum_{y \in X} K_y u_y}{\sum_{y \in X} K_y}\right)^2\right)^{\frac{1}{2}}. \tag{22}$$

We find the expected behaviour that $u$ and $\bar{u}$ should be close when $u$ does not vary much on each class of $\mathcal{P}$. The right handside of (22) can be interpreted as the sum of the variance within each class with respect to the natural norm, weighted by the conductivities $K_y$.

**A limit case with an explicit estimate of the distance to the condensed potential.** Let us detail a special case where we can compute an estimate of the distance between the potential and its condensation. As this distance goes to 0 in the limit we shall consider, it will enlight us about situations where the condensation is accurate, that is to say the potential and its condensed version are close.
For any matrix $K$ and $X \subset V$, such that $X \cap \Gamma = \emptyset$, let us consider the partition $\mathcal{P}_X$ where the only class with more than one point is $X$, and assume that for any $(x, y) \in X$, there exists a path from $x$ to $y$ in $X$ composed of edges of positive conductances. Let $\epsilon > 0$ and $K^\epsilon_{xy} = K_{xy} + \dfrac{\mathbf{1}_{x,y \in X}}{\epsilon} K_{xy}$: as $\epsilon$ goes to 0, the internal conductances in $X$ defined by $K^\epsilon$ blow up. Let $u_\epsilon$ the unconstrained Poisson problem defined by (8) for the conductance matrix $K^\epsilon$ and $\bar{u}_\epsilon$ the corresponding condensed problem, along the partition $\mathcal{P}_X$. One has the following

**Proposition 4.** *As $\epsilon$ goes to 0,*

$$\|\bar{u}_\epsilon - u_\epsilon\|_K = O(\epsilon). \tag{23}$$

*Proof.* Let $\bar{u}$ be the solution of the condensed Poisson equation (12) for the partition $\mathcal{P}_X$: $\bar{u}$ minimizes the functional $J$ over the set $\mathcal{V}_{\mathcal{X}} := \{v \in \mathcal{V}, v \text{ is constant on } X\}$. Defining the bilinear form $b(u,v) = \sum_{x,y\in X} K_{xy}(u_x - u_y)(v_x - v_y)$, one has $\mathcal{V}_{\mathcal{X}} = \mathrm{Ker}(b)$. As $b(u,v) = (Bu, Bv)_K$, where

$$Bu : \begin{cases} V & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \begin{cases} u(x) & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases} \end{cases} \tag{24}$$

and $(\ ,\ )_K$ is the scalar product associated to $K$, one can use the corollary 2.4 in [12] to get

$$\|\bar{u} - u_\epsilon\|_K = O(\epsilon). \tag{25}$$

$\bar{u}_\epsilon$ minimizes

$$J_\epsilon : \begin{cases} \mathcal{V} & \longrightarrow & \mathbb{R} \\ v & \longmapsto & \frac{1}{2}\sum\limits_{x,y\in V} K_{xy}(1 + \frac{\mathbf{1}_{x,y\in X}}{\epsilon})(v_x - v_y)^2 - \sum\limits_{x\in V} f_x v_x. \end{cases} \tag{26}$$

over $\mathcal{V}_{\mathcal{X}}$, but on the set of the constant fields on $X$ one has $J_\epsilon = J$. One thus has $\bar{u}_\epsilon = \bar{u}$. □

This proposition confirms the intuition that the higher are the internal conductance in a given group, the more accurate the condensation will be (for the $\|\cdot\|_K$ norm).

2.2. **Laplace equation on a resistive network.** In this subsection, we study the situation of a current flowing into the circuit when we impose a fixed difference of potential between two nodes: we force the potential to be 0 at a source node $s$, 1 at a target node $t$, and then compute the resulting potential is the circuit. This corresponds to plugging a 1-Volt battery at the two extremities $s, t$ of the system. This problem has been studied from a probabilistic standpoint in [5], and is known as the pressure drop problem in the fluid mechanics literature, see for instance [4]. Under this description, finding the potential $V$ results in solving the following discrete Laplace problem :

$$\begin{cases} -\mathcal{L}_K u &= 0 \quad \text{on } G \setminus \{s,t\} \\ u_s &= 0 \\ u_t &= 1 \end{cases} \tag{27}$$

We can solve the problem (27) by inverting a matrix: writing the nodes of $G$ as $\{1, \dots, n-2, s, t\}$, let us consider the matrix A defined by

$$\begin{array}{ll} A_{ij} = \frac{K_{i,j}}{K_i} & \text{for } i \le n-2, j = 1, \dots, n,\ j \neq i \\ A_{ii} = -1 & \text{for } i \le n-2 \\ A_{ij} = \delta_{ij} & \text{for } i > n-2, j = 1, \dots, n. \end{array} \tag{28}$$

where $K_i = \sum_{j\neq i} K_{ij}$. The equation (27) can be rewritten $Au = B$, where $B = (0, \dots, 0, 1)$.
As we did in subsection 2 for Poisson equation, we now define the condensed model of the previous equation according to a partition $\mathcal{P}$: we want to impose the potential to be constant in each class of $\mathcal{P}$. Let us consider that the classes of the source and the target are singletons. We still set $\overline{K_{XY}} = \sum\limits_{x\in X, y\in Y} K_{xy}$. The condensed equation for the potential is now

$$\begin{cases} -\mathcal{L}_{\bar{K}} w &= 0 \quad \text{on } \mathcal{P} \setminus \{s,t\} \\ w_s &= 0 \\ w_t &= 1. \end{cases} \tag{29}$$

We then define $\bar{u} : G \longrightarrow \mathbb{R}$ the interpolation of $w$ on $G$: for every $X \in \mathcal{P}$ and $x \in X$, $\bar{u}_x = w_X$. Let us define the matrix $C$ by

$$\begin{array}{lll} C_{ij} & = \frac{\bar{K}_{X_i,X_j}}{N_j \overline{K_{X_i}}} & \text{for } i \leq N-2, j = 1, \dots, N,\ X_i \neq X_j \\ C_{ij} & = -\delta_{ij} & \text{for } i \leq N-2, j \in X_i \\ C_{ij} & = \delta_{ij} & \text{for } i > N-2, j = 1, \dots, N, \end{array} \tag{30}$$

where $X_i$ is the class of $i$, $\overline{K_{X_i}} = \sum_{X \neq X_i} \bar{K}_{X,X_i}$, $N$ is the total number of classes, $N_j$ the cardinal of class $X_j$. The equation (29) writes $C\bar{u} = B$, with $B = (0, \dots, 0, 1)$. Once again, we want to define what could be a "good" matrix $K$ in ordre to have a decent approximation when proceeding to condensation. One can estimate for any matrix norm the error made in the condensation by

$$\begin{aligned} \|u - \bar{u}\| &= \|A^{-1}B - C^{-1}B\| \\ &\leq \|A^{-1}\|\|C - A\|\|C^{-1}B\| \end{aligned} \tag{31}$$

For the 2-norm, we get

$$\|A - C\|_2^2 \leq \|A^{-1}\|_2^2 \sum_{i=1}^{n-2} \left( \sum_{\substack{j=1 \\ j \notin X_i}}^{n-2} \left( \frac{K_{ij}}{K_i} - \frac{\bar{K}_{X_i X_j}}{\bar{K}_{X_i} N_j} \right)^2 + \sum_{\substack{j \neq i \\ j \in X_i}} \left( \frac{K_{ij}}{K_i} \right)^2 \right). \tag{32}$$

In the first term $\dfrac{K_{ij}}{K_i} - \dfrac{\bar{K}_{X_i X_j}}{\bar{K}_{X_i} N_j}$, $\dfrac{K_{ij}}{K_i}$ is the ratio of the conductances connected to $i$ involving $j$, while

$$\frac{\bar{K}_{X_i X_j}}{\bar{K}_{X_i} N_j} = \frac{\bar{K}_{X_i X_j}}{N_j} \frac{1}{\bar{K}_{X_i}} \tag{33}$$

is the ratio of the average conductance between the whole class $X_i$ and any $j \in X_j$ over the total conductance leading from $X_i$ to the outside of the class. This term is small if the relative importance for $i$ of every node $j \notin X_i$ does not vary with the representative $j \in X_j$ chosen.
The second term is a loss due to the rough estimation (31): it is lower when there is no internal connection in a group. One can nevertheless see that for fixed total internal conductances $\sum_{j \in X_i} K_{ij}$, it is lower in the isotropic situation where $K_{ij}$ is constant for $j \in X_i$.

**Remark 3.** *There are several ways to define the diagonal blocs of $C$, that are the $C_{ij}$ for $j \in X_i$. Indeed, one can add a constant $\beta_i$ to every $C_{ij}$ for $j \in X_i$ and $j \neq i$, and then substract $\beta_i(N_i - 1)$ to the diagonal term $C_{ii}$, without any modification of $C^{-1}B$ - the previous estimation is then the special case $\beta = 0$. The optimal domination in* (32) *would then be reached for*

$$\beta_i = \frac{\sum_{j \in X_i} \frac{K_{ij}}{K_i}}{(N_i - 1)^2 + (N_i - 1)}. \tag{34}$$

*However, in the ideal case $\dfrac{K_{ij}}{K_i} = c_i$ for every $j \in X_i$, we get $\beta_i = \dfrac{c_i}{N_i}$, and the second term in* (31) *becomes $\dfrac{(N_i - 1)^2 c_i}{N_i^3}$: one still have the unwanted effect that the second term only vanishes when there is no internal connection in a group.*

Despite the previous negative remark concerning the second term in (31), the behaviour of the first term illustrate the intuitive fact that points similarly connected to the rest of the network tend to have the same potential in the Laplace equation.

## 3. SIR model on subcommunities

3.1. **Elaboration of the model.** Following the literature in epidemiology (see [1], [2], [8]), we aim at defining formally a model of the propagation of a disease in a population. We assume a structure of the population into subcommunities: precisely, we consider a population $V = \{x_1, \dots, x_p\}$, and a partition $\mathcal{P} = \{X_1, \dots, X_n\}$ of $V$. A community $X$ is then composed of "epidemiologicaly equivalent" people , that are supposed to have the same behaviour in term of social interactions throughout the population. These subcommunities may be formed according to several features shared by the members: depending on the context, they may be groups of age, geographical groups, of more specific groups in very structured situations, as classes if the population is the public of a school, or teams of coworkers in a company.
We then assume that it's sufficient for each pair of subcommunities $X, Y$ to know $K_{XY}$, the daily total number of contacts between members of $X$ and members of $Y$ in order to describe the outbreak in the (structured) population: we consider that each individual $x \in X$ has $\overline{K_{XY}} = K_{XY}/N_X$ contacts with members of $Y$, where $N_X$ is the cardinal of $X$.

**Remark 4.** *From a modeling standpoint, one can consider that $\overline{K_{XY}}$ encodes the average number of interactions between people from $X$ and $Y$, and that we flatten the particularities by replacing for all $X, Y$-pairs the the number of contacts of any $x \in X$ by a single value $\overline{K_{XY}}$. This procedure might be particularly interesting when dealing with a large population one can only infer these average numbers of contacts between subcommunities and can not access to a precise description of every contact in the population.*

As it's usually done in the epidemiologic literature, see for instance the historical paper [8], we define for each group $X$ its proportion of susceptible, infected, and recovered people, $S_X^n$ , $I_X^n$ , $R_X^n \in [0, 1]$ at the unit of time $n$, and make the following assumptions:

- The epidemiologic quantities of the population evolve discretely in time.
- There is a fixed fraction $p$ of the contacts between infected and susceptibles people that leads to an infection.
- Infected people recover (or die, or are sent into quarantine ...) at rate $\gamma$.

Considering that a fraction $pI_Y$ of the $\overline{K_{XY}}$ contacts of each individual leads to an infection (indeed, there is a fraction $I_Y$ of their contacts that involves an infected person, and a fraction $p$ of these that leads to an infection), we now model the fraction of susceptible people infected by the $\overline{K_{XY}}$ contacts with $Y$ by:

$$1 - (1 - pI_Y)^{\overline{K_{XY}}} \tag{35}$$

The equation (35) leads to the following description of the epidemiologic dynamics:

**Definition 1.** *We define the condensed SIR model on the partition $V$ by the following dynamics:*

$$\begin{cases} S_X^{n+1} &= S_X^n \prod_{Y \in \mathcal{P}} (1 - I_Y^n p)^{\overline{K_{XY}}} \\ I_X^{n+1} &= (1-\gamma) I_X^n + S_X^n \left( 1 - \prod_{Y \in \mathcal{P}} (1 - I_Y^n p)^{\overline{K_{XY}}} \right) \\ R_X^{n+1} &= R_X^n + \gamma I_X^n, \end{cases} \tag{36}$$

for all $X \in \mathcal{P}$.

**Remark 5.** *The formulation of equation* (35) *stems from a probabilistic interpretation: the probability for each individual of not being infected by each $\overline{K_{XY}}$ contact with people being randomly infected with probability $I_Y$ is $1 - pI_Y$ for each contact. The probability of remaining disease-free after the sum of these contacts if $(1 - pI_Y)^{\overline{K_{XY}}}$. Although our model is built on this probabilistic intuition, one should care that is has no vocation to express the proper limit for some large numbers law of an underlying probabilistic model.*

**Remark 6.** *One expects the matrix $K$ to be symmetric in the previous formalism. Indeed, during a direct contact between two people there is no particular reason to favor a direction of transmission. Nevertheless, one may consider an asymmetric matrix $K$ when dealing with generalized contacts, for instance indirect contacts (succession at the same place of two groups of people in the case of an airborne infection).*

**Remark 7.** *Under the assumption $p \ll 1$ and when $V = \mathcal{P}$, one can see that the model* (36) *reduces to the classical homogeneous panmictic SIR model (see for instance* [14] *for a complete study of this kind of panmictic SIR model):*

$$\left\{ \begin{array}{rcl} S^{n+1} & = & S^n - \bar{p} S^n I^n \\ I^{n+1} & = & (1-\gamma) I^n + \bar{p} S^n I^n \\ R^{n+1} & = & R^n + \gamma I^n, \end{array} \right. \tag{37}$$

*where $\bar{p}$ is the product of $p$ by the daily average number of contact $K_{PP}$ in the population.*

3.2. **Microscopic deterministic model, comparison to the probabilistic model.** We now consider the extreme case where $\mathcal{P} = \{\{x_1\}, \dots, \{x_p\}\}$, i.e. where each group is composed of one only individual.

**Definition 2.** *In what follows, we will refer to the condensed model with $\mathcal{P} = \{\{x_1\}, \dots, \{x_p\}\}$ as the microscopic deterministic model:*

$$\left\{ \begin{array}{rcl} S_x^{n+1} & = & S_x^n \displaystyle\prod_{y \in V} \left(1 - I_y^n p\right)^{K_{xy}} \\ I_x^{n+1} & = & (1-\gamma) I_x^n + S_x^n \left( 1 - \displaystyle\prod_{y \in V} \left(1 - I_y^n p\right)^{K_{xy}} \right) \\ R_x^{n+1} & = & R_x^n + \gamma I_x^n. \end{array} \right. \tag{38}$$

In the following section, we will discuss some issues related to the comparison of the condensed SIR model and the deterministic microscopic model. Let us first end this section with a discussion on the links between this last model and probabilistic individual-based models built on Monte-Carlo approachs, as in [9], [11], [17].

We can define a probabilistic random model where $E_x^n$, the state of the individual $x$ at the day $n$, can either be $S$, $I$ or $R$, and has $K_{xy}$ independent chances to be infected by $y$ if they are in the state $I$. It is already clear that $S_x^{n+1}$ can not generically be the probability that $E_x^n = S$: to get an expression of the form (38), one should have

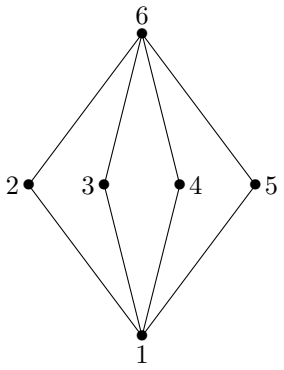


FIGURE 2. A configuration where the ways of infections for the node 6 are not independent when an initial random infection is introduced at node 1. When $p$ is high, either 1 has been infected and it is likely that many nodes in $\{2,3,4,5\}$ are infected at time 2, or 1 has not been infected and there is no outbreak at all.

$$\begin{aligned}\mathbf{P}\left(E_x^{n+1}=S\right) &= \mathbf{P}\left(E_x^n=S\right)\mathbf{P}\left(\bigcap_{y\in V}\{\text{y does not infect x}\}\right)\\ &= \mathbf{P}\left(E_x^n=S\right)\prod_{y\in V}\mathbf{P}\left(\text{y does not infect x}\right),\end{aligned} \tag{39}$$

which is true only if the possible ways of infections of $x$ are independent. This is clearly false on the example described on the figure 2, with an initial infection at the node 1 with probability $1/2$ at time $t=0$, and a high rate of infection $p$. At time 1, if one of the nodes $2,\dots,5$ is infected, so has been 1 and it is very likely that the other nodes have been infected as well: the ways of contamination of the node 6 are thus highly correlated.
Nevertheless, we observe that in common situations, the two models have a tendency to predict similar results on toy examples. We simulated the two models in the case of 3 people, with $K_{12}=K_{13}=1$ and $K_{23}=2$ and repeted the scheme for 5 units of time. Even if the predictions are not identical, the two scheme predict the same tendencies, see table 1. In particular, introducing the **risk** of an individual as the mean probability to be infected during the period of time of the experiment as the initial infection vary on the population, and their **dangerosity** as the mean number of people they will infect if they are the initial infected person, one can see on table 2 that the relative risks and dangerosity are the same for both models on this (very) simple situation. One notes that in the probabilistic model, the risk and the dangerosity are the same because of the underlying reversibility of the phenomenon of propagation, which is not the case in the deterministic model.

We expect the same common hierarchy in the prediction to occur in more complex situation, as it is still the case in the one represented by the graph on figure 3. We simulated both models associated to the contacts that label the edges of the graph. The dangerosity and the risk computed are presented on figure 4. Once again, the risk and the dangerosity take different values depending

| Model | S | I | R |
|---|---|---|---|
| Deterministic | 0.813 | 0.116 | 0.069 |
| Probabilistic | 0.787 | 0.131 | 0.081 |

TABLE 1. Comparison of the outputs of the deterministic and probabilistic models for $K_{12} = K_{13} = 1$ and $K_{23} = 2$. We infect initially the individual 1, and show the $S, I, R$ quantities of the person 3. We simulated 10000 times the experiment on 5 units of time, with a probability of transmission equal to 0.05 by contact and $\gamma = 0.2$.

| Model | Individual | Dangerosity | Risk |
|---|---|---|---|
| Deterministic | 1 | 0.458 | 0.455 |
| | 2 | 0.495 | 0.496 |
| | 3 | 0.495 | 0.496 |
| Probabilistic | 1 | 0.47 | 0.47 |
| | 2 | 0.52 | 0.52 |
| | 3 | 0.52 | 0.52 |

TABLE 2. Comparison of the dangerosities and risks of the deterministic and probabilistic models. We chose $K_{12} = K_{13} = 1$ and $K_{23} = 2$. We simulated 10000 times the experiment on 5 units of time, with a probability of transmission equal to 0.05 by contact and $\gamma = 0.2$.

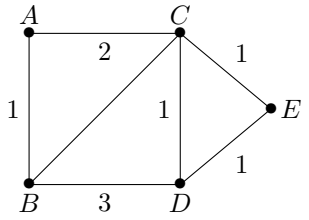


FIGURE 3. A graph of contacts between five people, labeled from $A$ to $E$. The number on the edges are the daily number of contacts between the vertices.

on the models, but one can see that the hierarchy of the risk/dangerosity is the same for both population: C-B-D-A-E.

## 4. Error of the condensation process.

Being given a contact matrix $K_{xy}$, one can define a microscopic SIR model with (38). From a partition $\mathcal{P}$, one can also condense the model by defining a SIR model on communities by (36). In this section, we want to estimate the error of the condensed model compared to the microscopic model, that is to say to define a criterion to measure the error we make by replacing the exact knowledge of all the contacts, encoded in the matrix $(K_{xy})_{x,y\in V}$ by the condensed contact matrix $(K_{XY})_{X,Y\in\mathcal{P}}$, with $K_{XY} = \sum_{x\in X, y\in Y} K_{xy}$.

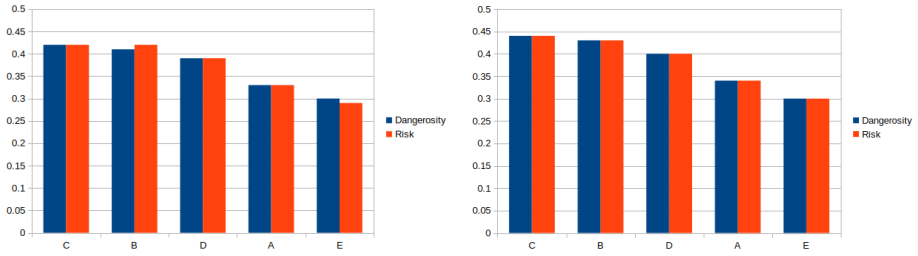


FIGURE 4. The risks and the dangerosities predicted by the two models (left: deterministic, right: probabilistic), for the contact data of the graph of the figure 3. We chose $p = 0.05$, $\gamma = 0.2$, and repeted the scheme for 5 time steps.

4.1. **An estimate on the quality of the condensation process.** Let us compare the models for a similar inital condition, that is that for a group $X$, the initial epidemiologic quantities $S_x^0, I_x^0, R_x^0$ don't depend on the representative $x \in X$ chosen. We can then define the condensation error at step $n$ by

$$\epsilon_n = \max\left( \max_{X\in\mathcal{P}\,,\ x\in X} |S_x^n - S_X^n|\ ,\ \max_{X\in\mathcal{P}\,,\ x\in X} |I_x^n - I_X^n| \right), \tag{40}$$

that is to say the worst error on the epidemiologic quantities made on the prediction at day $n$ by the condensed (so approximative) model compared to the microscopic (and thus more precise) one. We use the growth of this error to quantify the quality of a partition $\mathcal{P}$ of the population into classes.

The following proposition gives an estimate of the condensation error committed at time n as a function of K.

**Proposition 5.** *Defining $K_{xY} = \sum_{y\in Y} K_{xy}$ for $x \in V$ and $Y \subset V$, and $\overline{K_{XY}} = K_{XY}/N_X$ where $N_X$ is the cardinal of $X$, one has for every $n \geq 1$:*

$$\epsilon_n \leq \left( \left( 1 + 2p \max_{x\in V} K_{xV} \right)^n - 1 \right) \frac{\max_{X\in\mathcal{P}\,,\ x\in X} \sum_{Y\in\mathcal{P}} \left| K_{xY} - \overline{K_{XY}} \right|}{\max_{x\in V} K_{xV}}. \tag{41}$$

In order to prove proposition 5, we will use the following technical lemma:

**Lemma 1.** *Let $(a_i)_{i=1,\dots,n}$, $(b_i)_{i=1,\dots,n}$ real numbers in the range $[0\,,\ 1]$. One has*

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq \sum_{i=1}^n |a_i - b_i|\,. \tag{42}$$

*Proof of Lemma 1.* Straightforward by induction. □

*Proof of proposition 5.* We proceed by induction: let $n \in \mathbb{N}$.

$$\begin{aligned}
\left|S_x^{n+1} - S_X^{n+1}\right| &= \left| S_x^n \prod_{y\in V} \left(1 - pI_y^n\right)^{K_{xy}} - S_X^n \prod_{Y\in V} (1 - pI_Y^n)^{\overline{K_{XY}}} \right| \\
&\leq |S_x^n - S_X^n| \prod_{y\in V} \left(1 - pI_y^n\right)^{K_{xy}} + S_X^n \left| \prod_{y\in V} \left(1 - pI_y^n\right)^{K_{xy}} - \prod_{Y\in\mathcal{P}} (1 - pI_Y^n)^{\overline{K_{XY}}} \right| \\
&\leq \epsilon_n + \left| \prod_{Y\in\mathcal{P}} \prod_{y\in Y} \left(1 - pI_y^n\right)^{K_{xy}} - \prod_{Y\in\mathcal{P}} (1 - pI_Y^n)^{\overline{K_{XY}}} \right| \\
&\leq \epsilon_n + \left| \prod_{Y\in\mathcal{P}} \prod_{y\in Y} \left(1 - pI_y^n\right)^{K_{xy}} - \prod_{Y\in\mathcal{P}} \prod_{y\in Y} (1 - pI_Y^n)^{K_{xy}} \right| \\
&\quad + \left| \prod_{Y\in\mathcal{P}} (1 - pI_Y^n)^{K_{xY}} - \prod_{Y\in\mathcal{P}} (1 - pI_Y^n)^{\overline{K_{XY}}} \right|
\end{aligned}$$

We use three times the lemma 1 to get sucessively the following estimates:

$$\left| \left(1 - pI_y^n\right)^{K_{xy}} - (1 - pI_Y^n)^{K_{xy}} \right| \leq p\epsilon_n K_{xy} \tag{43}$$

$$\left| \prod_{y\in Y} \left(1 - pI_y^n\right)^{K_{xy}} - \prod_{y\in Y} (1 - pI_Y^n)^{K_{xy}} \right| \leq p\epsilon_n K_{xY} \tag{44}$$

$$\left| \prod_{Y\in\mathcal{P}} \prod_{y\in Y} \left(1 - pI_y^n\right)^{K_{xy}} - \prod_{Y\in\mathcal{P}} \prod_{y\in Y} (1 - pI_Y^n)^{K_{xy}} \right| \leq p\epsilon_n K_{xV}. \tag{45}$$

On the other hand, one has

$$\left| \prod_{Y\in\mathcal{P}} (1 - pI_Y^n)^{K_{xY}} - \prod_{Y\in\mathcal{P}} (1 - pI_Y^n)^{\overline{K_{XY}}} \right| \leq p \sum_{Y\in\mathcal{P}} \left|K_{xY} - \overline{K_{XY}}\right|. \tag{46}$$

We can then get the following bound: $\left|S_x^{n+1} - S_X^{n+1}\right| \leq \epsilon_n (1 + pK_{xV}) + p \sum_{Y\in\mathcal{P}} \left|K_{xY} - \overline{K_{XY}}\right|$. A similar calculus gives $\left|I_x^{n+1} - I_X^{n+1}\right| \leq \epsilon_n (1 - \gamma + 2pK_{xV}) + p \sum_{Y\in\mathcal{P}} \left|K_{xY} - \overline{K_{XY}}\right|$. The condensation error thus satisfies the following system:

$$\left\{ \begin{array}{rcl} \epsilon_{n+1} & \leq & \left(1 + p \max_{x\in V} K_{xV}\right) \epsilon_n + p \max_{X\in\mathcal{P},\, x\in X} \sum_{Y\in\mathcal{P}} \left|K_{xY} - \overline{K_{XY}}\right| \\ \epsilon_0 & = & 0 \end{array} \right. \tag{47}$$

One deduces, by comparison with an arithmetico-geometric sequence:

$$\epsilon_n \leq \left( \left(1 + 2p \max_{x\in V} K_{xV}\right)^n - 1 \right) \frac{\max_{X\in\mathcal{P},\, x\in X} \sum_{Y\in\mathcal{P}} \left|K_{xY} - \overline{K_{XY}}\right|}{\max_{x\in V} K_{xV}}. \tag{48}$$

□

In the limit $p \max_{x\in V} K_{xV} \ll 1$, the equation (41) reduces to

$$\epsilon_n \leq 2np \max_{X\in\mathcal{P},\, x\in X} \sum_{Y\in\mathcal{P}} \left|K_{xY} - \overline{K_{XY}}\right|. \tag{49}$$

One can thus expect the error to grow at most linearly in $n$, with a shape given by the worst approximation made by replacing $K_{xY}$, the microscopic interaction bewteen an individual $x \in V$ with some group $Y \in \mathcal{P}$ by the averaged version $\overline{K_{XY}}$.

4.2. **Exact condensation: an example.** The previous bound given by proposition 5 gives us a first example of a microscopic contact matrix $(K_{xy})_{x,y\in P}$ and a partition $\mathcal{P}$ that leads to an exact condensation:

**Corollary 1.** *Assume that for every groups $X, Y \in \mathcal{P}$, $K_{xY}$ does not depend on $x \in X$. Then the two models (the deterministic microscopic and the condensed one associated to the partition $\mathcal{P}$) give the same results, that is to say that the error $\epsilon_n$ defined in* (40) *is* 0 *for every $n \geq 0$.*

*Proof.* One has for every $X, Y \in \mathcal{P}$ and for every $x \in X$ $K_{xY} = \overline{K_{XY}}$, so the estimate (41) gives $\epsilon_n = 0$ for every $n$. □

**Remark 8.** *One can as well proceed by induction showing that the formula* (38) *does not depend on the representative of a group chosen.*

We now want to define a set of hypothesis on the population $V$ and a partition $\mathcal{P}$ in order to get a microscopic contact matrix $(K_{xy})_{x,y\in V}$ that leads to an exact condensation in a more general setting: we start from a microscopic matrix $K$, and allow the $K_{xy}$ contacts between $x$ and $y$ to have different probabilities of transmission $p_{xy}^1, \ldots, p_{xy}^{K_{xy}}$. The generalized microscopic model now writes

$$\begin{cases} S_x^{n+1} &= S_x^n \prod_{y\in V} \prod_{s=1}^{K_{xy}} \left(1 - I_y^n p_{xy}^s\right) \\ I_x^{n+1} &= (1-\gamma) I_x^n + S_x^n \left(1 - \prod_{y\in V} \prod_{s=1}^{K_{xy}} \left(1 - I_y^n p_{xy}^s\right)\right) \\ R_x^{n+1} &= R_x^n + \gamma I_x^n. \end{cases} \tag{50}$$

Let us describe the hypothesis needed in this context to define a (generalized) condensed model that coincide with this microscopic model.

**Remark 9.** *In a general context one may imagine a population where we have a full tracking of the social interactions, as in the study presented in* [6]. *In this study, people working in an office building were asked to wear a chest-sensor that detects every situation where two people face at short distance. One can then reconstruct the full contact matrix, with every contact associated to a time of contact. Being given $\lambda$ a rate of transmission per unit of time of contact, one can then associate to each contact of duration $t$ a probability $p = 1 - e^{-t\lambda}$.*

**Proposition 6.** *In the previous setting, we make the assumption that for every pair $X, Y \in \mathcal{P}$, $K_{xY}$ and the collection of the associated probabilities $(p_{xy}^s)_{y\in Y, s=1,\ldots,K_{xy}}$ do not depend on $x \in X$. Then, starting from initial quantities $S_x^0, I_x^0, R_x^0$ that do not depend on $x \in X$, the generalized microscopic*

*model predicts epidemiological quantities that do not depend on the representative of a group chosen. Defining the condensed macroscopic model as*

$$\left\{\begin{array}{rcl} S_X^{n+1} &=& S_X^n \prod\limits_{Y\in\mathcal{P}} \prod\limits_{s=1}^{K_{XY}} (1-I_Y^n p_{XY}^s) \\ I_X^{n+1} &=& (1-\gamma) I_X^n + S_X^n \left(1 - \prod\limits_{Y\in\mathcal{P}} \prod\limits_{s=1}^{K_{XY}} (1-I_Y^n p_{XY}^s)\right) \\ R_X^{n+1} &=& R_X^n + \gamma I_X^n, \end{array}\right. \tag{51}$$

*where* $(p_{XY}^s)_{s=1,\dots,K_{XY}} = (p_{xy}^i)_{y\in Y, i=1,\dots,K_{xy}}$ *for any* $x \in X$*, one has exact condensation between the two generalized models.*

*Proof.* By induction, one just has to group terms in the products in (50) and use that the set of probabilities $(p_{xy}^i)_{y\in Y,i=1,\dots,K_{xy}}$ does not depend on $x \in X$. □

**Remark 10.** *The assumption made in the proposition 6 might seem a bit over-restrictive. Nevertheless, let us consider a situation where groups of people that stay together move in a space (*$\mathbb{R}^2$*, or a graph ...). We consider that when there is a meeting event between two groups, every individual of each groupe has a contact of a fixed duration with a constant number of people of the other group (for instance, we say that every individual of group* $X$ *has* 3 *contacts of 1 minute with people of group* $Y$*). The situation we picture satisfies the previous hypothesis.*

Let us end this section with the analysis of the previous situation in the case where the probabilities $p_{XY}^s$ are very small. In this particular setting, one has

$$1 - \prod_{Y\in\mathcal{P}} \prod_{s=1}^{\overline{K_{XY}}} (1-I_Y^n p_{XY}^s) \approx \sum_{Y\in\mathcal{P}} I_Y^n \sum_{s=1}^{\overline{K_{XY}}} p_{XY}^s. \tag{52}$$

Following the remark 9, we set $p_{XY}^s = 1 - e^{-\lambda t_{XY}^s} \approx \lambda t_{XY}^s$. The fraction of people in $X$ infected at time unit $n$ is thus approximately

$$\lambda S_X^n \sum_{Y\in\mathcal{P}} I_Y^n \sum_{s=1}^{\overline{K_{XY}}} t_{XY}^s. \tag{53}$$

The condition for the exact condensation in this case is now that for every $X, Y \in \mathcal{P}$, the total time spent with some people in $Y$ by an individual $x \in X$, $\sum\limits_{y\in Y} \sum\limits_{s=1}^{K_{xy}} t_{xy}^s$, must not depend of the representative $x \in X$ chosen.

## Conclusion, perspectives

We defined and studied the process of condensation of a graph into subcommunities of nodes when studying the propagation of quantities on the graph. In the context of resistive networks in section 2, we estimated the distance between the potential and its condensed version and derived heuristics to understand in which case this distance is small. We defined in section 3 a condensed SIR model on a structured population, and compared it to the correspondant probabilistic SIR model. In section 4, we computed an estimation of the error commited when replacing the microscopic model by its condensed version. We eventually studied an example of exact condensation, that correspond to a situation where the people of a community stay together at every time and

have the same number of contacts when crossing another group.
A full implementation of this model under the previous hypotheses in the context of a school is currently in progress. It is designed as a web app at the disposition of school principals. One inputs a time-schedule of every class (that can be singletons in the case of teachers) in the establishment and a symbolic graph of the school. The data of contact are then computed from the paths, and an epidemiologic model of the type of (51) is computed. Some visualizations of the results are designed to help a school principal to build a schedule that minimizes the risk of an outbreak in their school and to identify some potentially dangerous subcommunities. We propose contact matrices, that encode the ratios of infected people in every community at the end of a certain period of detection of the outbreak as the initial infection is localized in a certain community. We provide as well the dangerosities and risks defined in section 3. The designing of this app is part of the project ModCov19[1].

**Acknowledgements**. We thank S. Faure and B. Maury for their fruitful suggestions.
The work of Félicien Bourdin is supported by the ERC grant NORIA.

## References

[1] D. Acemoglu, V. Chernozhukov, I. Werning, & M. D. Whinston, Optimal targeted lockdowns in a multi-group SIR model, National Bureau of Economic Research (Vol. 27102, 2020).
[2] B. Cantó, C. Coll, & E. Sánchez, Estimation of parameters in a structured SIR model, Advances in Difference Equations, 2017(1), 1-13.
[3] J. Cea, Approximation variationnelle des problèmes aux limites, Annales de l'Institut Fourier, Tome 14 (1964) no. 2, pp. 345-444.
[4] T.H. Chilton, & A. P. Colburn, II—pressure drop in packed tubes, Industrial & Engineering Chemistry, 23(8), 913-919 (1931).
[5] P.G. Doyle, & J. L. Snell, Random walks and electric networks (Vol. 22), American Mathematical Soc. (1984).
[6] M. Génois, C. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, & A. Barrat, Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers, Network Science, 3(3), 326-347 (2015).
[7] M. Keller, & D. Lenz, Unbounded Laplacians on graphs: basic spectral properties and the heat equation, Mathematical Modelling of Natural Phenomena, 5(4), 198-224 (2010).
[8] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics.", Proceedings of the royal society of London, Series A, Containing papers of a mathematical and physical character 115.772 (1927): 700-721.
[9] M. S. S. Khan , A computer virus propagation model using delay differential equations with probabilistic contagion and immunity, International Journal of Computer Networks & Communications (IJCNC) Vol.6, No.5, September 2014.
[10] R.I. Kondor & J. Lafferty, Diffusion kernels on graphs and other discrete structures, Proceedings of the 19th international conference on machine learning (Vol. 2002, pp. 315-322).
[11] S. Mauras, V. Cohen-Addad, G. Duboc, M. Dupré la Tour, P. Frasca ,C. Mathieu et al, Mitigating COVID-19 outbreaks in workplaces and schools by hybrid telecommuting, PLoS Comput Biol 17(8): e1009264 (2021).
[12] B. Maury, Numerical analysis of a finite element/volume penalty method, SIAM Journal on Numerical Analysis, 47(2), 1126-1148 (2009).
[13] B. Mohar, Y. Alavi, G. Chartrand, & O. R. Oellermann, The Laplacian spectrum of graphs, Graph theory, combinatorics, and applications, 2(871-898), 12 (1991).
[14] J.D. Murray, Mathematical biology: I. An introduction (Vol. 17), Springer Science & Business Media (2007).
[15] S. Seshu & M.B. Reed. Linear graphs and electrical networks, Addison-Wesley Publishing Company, 1961.
[16] U. Trottenberg,C. W. Oosterlee & A. Schuller, Multigrid, Elsevier (2000).
[17] K. Zhu and L. Ying, "Information Source Detection in the SIR Model: A Sample-Path-Based Approach," in IEEE/ACM Transactions on Networking, vol. 24, no. 1, pp. 408-421, Feb. 2016.

[1]See https://modcov19.math.cnrs.fr/