3

CAVITY METHOD

3.1 SPARSE AND TREE-LIKE GRAPHS

In the previous lecture we introduced the *naïve mean field* (NMF) approximation to study the epidemic threshold on a graph. We saw, however, that this approximation is appropriate only for dense graphs, while most of real world graphs are (luckily) sparse. We here investigate an alternative approach that is well suited for sparse random graphs and that builds an approximation based on the locally tree-like[1] structure of a sparse *Erdős-Rényi* (ER) graph.

3.1.1 LOCALLY TREE-LIKE GRAPHS

Let us first introduce the concept of *rooted* graph $\mathcal{G}_i(\mathcal{V}, \mathcal{E})$ that is a graph in which a particular node $i \in \mathcal{V}$ (the *root*) is specified. Denote with $\mathfrak{B}_i(t)$ the ball of radius t around the node i, *i.e.* the sub-graph made by the set of all nodes that can be reached from i in at most t steps and the corresponding edges. If the law of $\mathfrak{B}_i(t)$ under uniformly random sampling of the root admits a limit $\mathcal{L}$, then we call it *local weak limit*. In words, $\mathcal{L}$ is the asymptotic local distribution of $\mathcal{G}(\mathcal{V}, \mathcal{E})$ as seen from a random vertex. A relevant result concerning sparse ER graphs[2] is that they locally converge to a tree, as more formally stated in Property 3.1.

1 We recall that an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$is said to be a tree if it is connected and it does not contain any cycle.

2 But actually also other sparse random graphs.

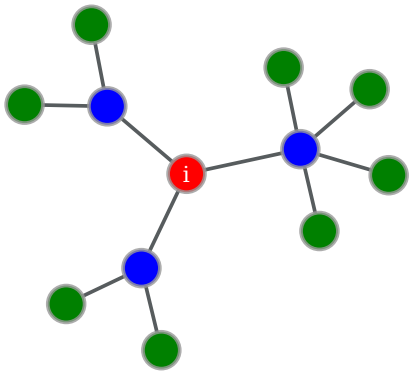

Figure 3.1: A toy example of a Poisson GW tree. In red the root i, in blue the first generation of nodes, in green the second.

Local convergence to a tree

Property 3.1 (Convergence to Poisson *Galton Watson tree* (GW) tree of ER)**.** *A* sparse *ER random graph with $n \to \infty$ rooted at i with average degree $d = O_n(1)$ converges locally to a Poisson GW tree so obtained: consider the node i as the root and generate d_i neighbours (called* sons*), where d_i is a Poisson random variable with parameter d and iteratively repeat the operation for each son.*

As a consequence of this property, a sparse ER graph locally *looks like* a tree and hence, with high probability, there are no cycles of finite size.[3] Figure 3.1 displays an example of a Poisson GW tree, rooted at i.

Let us now describe a fundamental property of probability distributions defined over the nodes of a tree, namely, *conditional independence.*

3.1.2 CONDITIONAL INDEPENDENCE

Consider three random variables x, y, z. We say that x and y are conditionally independent given z if

Conditional independence

$$\mathbb{P}[x, y \mid z] = \mathbb{P}[x \mid z]\mathbb{P}[y \mid z]. \tag{3.1}$$

Note that two random variables are independent if we can write $\mathbb{P}[xy] = \mathbb{P}[x]\mathbb{P}[y]$, that is what we did in the NMF approximation. Conditional independence holds only on the conditional probabilities and in general $\mathbb{E}[xy] \neq \mathbb{E}[x]\mathbb{E}[y]$. Now, the relation between conditional independence and trees is that all variables associated to the neighbors of a same node are conditionally independent given the value of their common neighbor, or, more formally

$$\forall j \neq k \in \partial i, \quad \mathbb{P}[x_j, x_k \mid x_i] = \mathbb{P}[x_j \mid x_i]\mathbb{P}[x_k \mid x_i].$$

Let us try to understand why. Suppose there is a piece of news that is propagating on the network through contacts. If a node knows it, then with some

3 Recall the local convergence definition is given in the asymptotic limit of $n \to \infty$. Finite cycles will exist, but their size will depend on n (for instance they may grow as $\log(n)$) and thus diverge in the large n limit.

probability it will talk about it to its neighbors that will also be aware of it from that moment on. Now, let us consider a node i (as the red one in Figure 3.1) and two of its neighbors j, k (in blue, same figure). If we let $x_i = 1$ if i knows the piece of information and $x_i = 0$ otherwise, then, clearly $\mathbb{P}[x_j, x_k] \neq \mathbb{P}[x_j]\mathbb{P}[x_k]$. If the variables were independent, the notion of x_j would not allow me to say anything about x_k, but it turns out that if I know x_j I can tell something about x_k. The two random variables bring information one of the other because there is a (short) path connecting them and the piece of information may flow from one node to the other. However, since we are considering a tree, there is only one such path, that is the one going through i. If we suppose to know x_i, then knowing also x_j does not add any information when trying to predict x_k, because the only influence j has on k is through i. This is the effect of conditional independence.

Notably, conditional independence implies the following relation that we will exploit later on.

$$\begin{aligned}\mathbb{P}(x_i, x_j, x_k) &= \mathbb{P}(x_j, x_k|x_i)\mathbb{P}(x_i) \\ &= \mathbb{P}(x_j|x_i)\mathbb{P}(x_k|x_i)\mathbb{P}(x_i) \\ &= \frac{\mathbb{P}(x_i, x_j)\mathbb{P}(x_i, x_k)}{P(x_i)}\end{aligned} \tag{3.2}$$

Knowing that a sparse ER random graph asymptotically "looks like" a tree, we can now simplify our analysis exploiting conditional independence and the graph structure to introduce the *cavity method*, or *belief propagation*.

3.2 FACTORIZING PROBABILITY DISTRIBUTIONS ON TREES

As we saw in Chapter 2, a difficulty of studying processes on a graph is to compute the edge marginal probability distributions that cannot be simply assumed to factorize as the product of the node marginals. The cavity method builds on the fact that the edge marginals can be exactly calculated on tree with a recursive formula, as stated in Lemma 3.1.

Lemma 3.1. *Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a tree and let $\mu(\boldsymbol{x})$ be a probability distribution defined on $\mathcal{G}(\mathcal{V}, \mathcal{E})$ that can be written as*

$$\mu(\boldsymbol{x}) = \prod_{(ij)\in\mathcal{E}} \phi_{ij}(x_i, x_j). \tag{3.3}$$

Then the edge marginal $\mu_{ij}(x_i, x_j) = \sum_{\boldsymbol{x}\setminus x_i, x_j} \mu(\boldsymbol{x})$ and the node marginal $\mu_i(x_i) = \sum_{\boldsymbol{x}\setminus x_i} \mu(\boldsymbol{x})$ can be written in the following form:

$$\mu_i(x_i) = \prod_{k\in\partial i} \eta_{ik}(x_i) \tag{3.4}$$

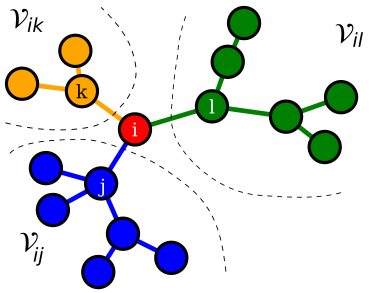

Figure 3.2: Sketch of a tree. The node i in red, while in green, blue and orange the edges and nodes $\mathcal{E}_{ix}, \mathcal{V}_{ix}$ with $x = j, k, l$, respectively. Note that i belongs to $\mathcal{V}_{ik}, \mathcal{V}_{ij}$ and $\mathcal{V}_{i\ell}$.

$$\mu_{ij}(x_i x_j) = \phi_{ij}(x_i, x_j) \prod_{k \in \partial i \setminus j} \eta_{ik}(x_i) \prod_{\ell \in \partial j \setminus i} \eta_{jl}(x_j). \tag{3.5}$$

The quantities $\eta_{ij}(x_i)$ are defined on the set of directed edges.

From Equations (3.4, 3.5), exploiting $\mu_i(x_i) = \sum_{x_j} \mu_{ij}(x_i, x_j)$, it is obtained that the *messages* have to satisfy the following fixed point equation

Message passing

$$\eta_{ij}(x_i) = \sum_{x_j} \phi_{ij}(x_i, x_j) \prod_{\ell \in \partial j \setminus i} \eta_{jl}(x_j). \tag{3.6}$$

Let us now sketch here the proof of Lemma 3.1 since it is very pedagogical and helpful to understand the essence of cavity method.

Proof of Lemma 3.1. Denote with $\mathcal{E}_d$ the set of directed edges of $\mathcal{G}(\mathcal{V}, \mathcal{E})$and consider $(ij) \in \mathcal{E}_{\mathrm{d}}$. We define $\mathcal{E}_{ij}$ as the set of all edges that can be reached from i only passing through j. As a consequence of the fact that on a tree there exists a unique path connecting any two nodes – since there are no cycles –, the two following properties are verified:

$$\forall\, i \in \mathcal{V}, \quad \mathcal{E} = \bigcup_{k \in \partial i} \mathcal{E}_{ik}\,; \tag{3.7}$$

$$\forall\, (ij) \in \mathcal{E}, \quad \mathcal{E} = \{(ij)\} \cup \underbrace{\bigcup_{k \in \partial i \setminus j} \mathcal{E}_{ik}}_{\text{reached from } (ji)} \cup \underbrace{\bigcup_{\ell \in \partial j \setminus i} \mathcal{E}_{jl}}_{\text{reached from } (ij)}. \tag{3.8}$$

Furthermore, note that $\forall\, j \neq k$, $\mathcal{E}_{ij} \cap \mathcal{E}_{ik} = \emptyset$. A pictorial representation of the definition of $\mathcal{E}_{ij}$ is given in Figure 3.2. Exploiting Equation (3.7), $\mu(\boldsymbol{x})$ can then be written as:

$$\mu(\boldsymbol{x}) = \prod_{(ab) \in \mathcal{E}} \phi_{ab}(x_a, x_b) = \prod_{k \in \partial i} \prod_{(ab) \in \mathcal{E}_{ik}} \phi_{ab}(x_a, x_b) := \prod_{k \in \partial i} \psi_{ik}(\boldsymbol{x}_{\mathcal{V}_{ik}}),$$

where $\mathcal{V}_{ik}$ is the set of nodes connected by edges in $\mathcal{E}_{ik}$ (i included) and $\boldsymbol{x}_{\mathcal{V}_{ik}}$ is the variable vector corresponding to those nodes. The node marginal can then be written in the following form

$$\mu_i(x_i) = \sum_{\boldsymbol{x}\setminus x_i} \mu(\boldsymbol{x}) = \sum_{\boldsymbol{x}\setminus x_i} \prod_{k\in\partial i} \psi_{ik}(\boldsymbol{x}_{\mathcal{V}_{ik}}) = \prod_{k\in\partial i} \sum_{\boldsymbol{x}_{\mathcal{V}_{ik}}\setminus x_i} \psi_{ik}(\boldsymbol{x}_{\mathcal{V}_{ik}}).$$

Denoting $\eta_{ik}(x_i) := \sum_{\boldsymbol{x}_{\mathcal{V}_{ik}}\setminus x_i} \psi_{ik}(\boldsymbol{x}_{\mathcal{V}_{ik}})$, we obtain the first equation of Lemma 3.1. Note that $\eta_{ki}(x_i)$ indeed only depends on x_i since the sum is run over all variables $\boldsymbol{x}_{\mathcal{V}_{ik}}$, except x_i. Proceeding in a similar way, the expression of the edge marginal is obtained from Equation (3.8).

$$\begin{aligned}
\mu_{ij}(x_i,x_j) &= \sum_{\boldsymbol{x}\setminus x_i x_j} \mu(\boldsymbol{x}) \\
&= \sum_{\boldsymbol{x}\setminus x_i x_j} \phi_{ij}(x_i,x_j) \prod_{k\in\partial i\setminus j} \prod_{(ab)\in\mathcal{E}_{ik}} \phi_{ab}(x_a,x_b) \prod_{\ell\in\partial j\setminus i} \prod_{(cd)\in\mathcal{E}_{k\ell}} \phi_{cd}(x_c,x_d) \\
&= \sum_{\boldsymbol{x}\setminus x_i x_j} \phi_{ij}(x_i,x_j) \prod_{k\in\partial i\setminus j} \psi_{ik}(\boldsymbol{x}_{\mathcal{V}_{ik}}) \prod_{\ell\in\partial j\setminus i} \psi_{j\ell}(\boldsymbol{x}_{\mathcal{V}_{j\ell}}) \\
&= \phi_{ij}(x_i,x_j) \left(\prod_{k\in\partial i\setminus j} \sum_{\boldsymbol{x}_{\mathcal{V}_{ik}}\setminus i} \psi_{ik}(\boldsymbol{x}_{\mathcal{V}_{ik}}) \right) \cdot \left(\prod_{\ell\in\partial j\setminus i} \sum_{\boldsymbol{x}_{\mathcal{V}_{j\ell}}\setminus j} \psi_{j\ell}(\boldsymbol{x}_{\mathcal{V}_{j\ell}}) \right) \\
&= \phi_{ij}(x_i,x_j) \prod_{k\in\partial i\setminus j} \eta_{ik}(x_i) \cdot \prod_{\ell\in\partial j\setminus i} \eta_{j\ell}(x_j).
\end{aligned}$$

□

The essence of the proof of Lemma 3.1 relies on the conditional independence of the node variables on trees. More specifically, to obtain Equation (3.7), one could imagine to remove the node i, obtaining d_i (the degree of i) disconnected sub-graphs in which variables are independent and hence factorize. Similarly Equation (3.8) is obtained removing the nodes i and j from the graph. We now show that on a tree, Equation (3.2) is verified.

Lemma 3.2. *Let $\mathcal{G}(\mathcal{V},\mathcal{E})$ be a tree and let $\mu(\boldsymbol{x})$ be a probability distribution defined on $\mathcal{G}(\mathcal{V},\mathcal{E})$ that can be written as per Equation (3.3). Then, taking $j,k\in\partial i$ with $j\neq k$ we can write*

$$\mathbb{P}(x_i,x_j,x_k) = \frac{\mathbb{P}(x_i,x_j)\mathbb{P}(x_i,x_k)}{\mathbb{P}(x_i)}$$

Proof. For simplicity, we will drop the dependence on the variables $\boldsymbol{x}$. Following the same procedure we used to prove Lemma 3.1, we can easily show that

$$\mathbb{P}(x_i,x_j,x_k) = \phi_{ij}\phi_{ik} \prod_{p\in\partial j\setminus i} \eta_{jp} \prod_{q\in\partial k\setminus i} \eta_{kq} \prod_{r\in\partial i\setminus\{j,k\}} \eta_{ir}.$$

This can be rewritten in the following form

$$\mathbb{P}(x_i, x_j, x_k) = \frac{\left(\phi_{ij} \prod_{p\in\partial j\setminus i} \eta_{jp} \prod_{r\in\partial i\setminus j} \eta_{ir}\right)\left(\phi_{ik} \prod_{q\in\partial k\setminus i} \eta_{kq} \prod_{r\in\partial i\setminus k} \eta_{ir}\right)}{\eta_{ik} \prod_{k\in\partial i\setminus k} \eta_{ir}}$$
$$= \frac{\mathbb{P}_{ij}(x_i, x_j)\mathbb{P}_{ik}(x_i, x_k)}{\mathbb{P}_i(x_i)},$$

where in the last step we used the relations shown in Lemma 3.1. □

Given these results on trees, let us now move to sparse graphs.

3.2.1 CAVITY METHOD ON TREE-LIKE GRAPHS

When we consider a graph that is not a tree, the proof we gave above does not generalize because $\mathcal{E}_{ij} \cup \mathcal{E}_{ik} \neq \emptyset$, due to the presence of cycles. In a tree-like graph, however, we know that cycles do not have a short length. When considering two nodes j, k in the neighborhood of a same node, there will be a short path of length 2 connecting them and other very long paths that pass through other nodes. The main intuition we have is that all those long paths are unimportant and the main channel of of relation is the short path connecting them. For this reason, we simply use conditional independence as an *ansatz* that is asymptotically verified on sparse graphs. We can then rewrite the cavity equations on a graph (with cycles) a follows.

Factorizing probabilities on sparse graphs

The cavity fixed point equations

$$\eta_{ji}(x_i) = \frac{Z_i}{Z_{ji}} \sum_{x_j} \phi_{ij}(x_i, x_j) \prod_{\ell\in\partial j\setminus i} \eta_{j\ell}(x_j).$$
$$\mu_i(x_i) \approx \frac{1}{Z_i} \prod_{k\in\partial i} \eta_{ik}(x_i)$$
$$\mu_{ij}(x_i, x_j) \approx \frac{1}{Z_{ij}} \phi_{ij}(x_i, x_j) \prod_{k\in\partial i\setminus j} \eta_{ik}(x_i) \prod_{\ell\in\partial j\setminus i} \eta_{j\ell}(x_j).$$

Given these equations, we now introduce the *non-backtracking matrix* or *Hashimoto operator* that naturally comes into play from the cavity method.

3.3 THE NON-BACKTRACKING MATRIX

Let us consider the fixed point cavity equation, letting $r_{ij}(x_i) = \log(\eta_{ij}(x_i))$ and $C_{ji} = \log(Z_i) - \log(Z_{ji})$.

$$r_{ji}(x_i) = C_{ji} + \log\left(\sum_{x_j} \exp\left\{\log \phi_{ij}(x_i, x_j) + \sum_{\ell \in \partial j \setminus i} r_{j\ell}(x_j).\right\}\right).$$

Focusing on the sum $\sum_{\ell \in \partial j \setminus i}$, we introduce the non-backtracking matrix.

The non-backtracking matrix

Let $\mathcal{E}_{\mathrm{d}}$ be the set of *directed* edges of a graph $\mathcal{G}$. We define the non-backtracking matrix $B \in [0,1]^{|\mathcal{E}_{\mathrm{d}}| \times |\mathcal{E}_{\mathrm{d}}|}$ as

$$B_{(ij),(k\ell)} = \delta_{jk}(1 - \delta_{i\ell}), \tag{3.9}$$

for all $(ij), (k\ell) \in \mathcal{E}_{\mathrm{d}}$. Then, given a vector $\boldsymbol{g} \in \mathbb{R}^{|\mathcal{E}_{\mathrm{d}}|}$, we have

$$(B\boldsymbol{g})_{(ij)} = \sum_{\ell \in \partial j \setminus i} g_{j\ell}.$$

The non-backtracking matrix is naturally related to the cavity method

In simple words, we can say that the non-backtracking matrix B is the linear operator associated to the cavity approximation. To interpret its definition, in essence we can see B as the adjacency matrix of graph in which each node is a directed edge of $\mathcal{G}(\mathcal{V}, \mathcal{E})$ and two nodes are neighboring if they are successive and the second one is not the reversed of the first. Unlike the adjacency matrix, the spectral radius of the non-backtracking matrix is "well behaved" in the sparse regime, as stated by the following theorem.

Theorem 3.1. *Consider a symmetric matrix $A \in [0,1]^{n \times n}$ in which the entries are set to 1 independently (up to symmetry) with probability $\mathbb{P}(A_{ij} = 1) = P_{ij}$ and $P_{ij} = O_n(n^{-1})$ for all i, j. Then, for all large n with high probability, the spectral radius of the non-backtracking matrix B associated with the adjacency matrix A is*

$$\rho(B) = \rho(P) + o_n(1).$$

This theorem is very general and allows us to consider easily both the ER and the configuration model. For the ER, we have $P = \frac{\langle d \rangle}{n} \mathbf{1}_n \mathbf{1}_n^T$ and $\rho(P) = \langle d \rangle$, the expected average degree. For the configuration model, instead, we can write $P = \frac{1}{2|\mathcal{E}|} \boldsymbol{d}\boldsymbol{d}^T$ and thus $\rho(B) = \frac{\langle d^2 \rangle}{\langle d \rangle}$.

Let us now see how the matrix B enters into play when studying the epidemic threshold on a sparse graph with the cavity method.

3.4 AN APPLICATION TO EPIDEMICS

Let us now use the cavity method to find the epidemic threshold on a sparse graph, in which the NMF is not appropriate. Let us consider Equation (2.2) that describes the dynamics of the infected state in a *Susceptible-Infected-Recovered model* (SIR) model. We can write

$$\partial_t \mathbb{P}(x_i(t) = I) = \beta \sum_{j \in \mathcal{V}} A_{ij} \mathbb{P}\left(x_i(t) = S, x_j(t) = I\right) - \mu \mathbb{P}(x_i(t) = I).$$

The whole point of going beyond NMF is to realize that $x_i(t), x_j(t)$ are not independent if $A_{ij} = 1$. To move forward, let us lighten a bit the notation. We define $\boldsymbol{p} \in \mathbb{R}^n$ the vector with entries $p_i(t) = \mathbb{P}(x_i(t) = I)$ and with $\boldsymbol{\chi}(t) \in \mathbb{R}^{2|\mathcal{E}|}$ the vector with entries $\chi_{ij}(t) = \mathbb{P}(x_i(t) = S, x_j(t) = I)$. Note that the vector $\boldsymbol{\chi}$ is defined over the set of *directed* edges of the graph and that $\chi_{ij}(t) \neq \chi_{ji}(t)$, in general. With this notation, we can rewrite the state evolution as

$$\partial_t \boldsymbol{p}(t) = \beta T \boldsymbol{\chi}(t) - \mu \boldsymbol{p}(t), \tag{3.10}$$

where we introduced the matrix $T \in \mathbb{R}^{n \times 2|\mathcal{E}|}$, defined as $T_{i,(ab)} = \delta_{ia} A_{ab}$. Now, to proceed, we need to write a state evolution equation for the vector $\boldsymbol{\chi}(t)$ as well. The probability that i and j are respectively susceptible and infected at a given time step implies that they were both susceptible and j got infected (but certainly not from i), while i did not or that j was already infected, it did not recover and i did not get infected.

$$\chi_{ij}(t + dt) =$$
$$\mathbb{E}\left[\delta[x_i(t) = S]\delta[x_j(t) = S] \underbrace{\left(1 - \beta dt \sum_{k \in \partial i \setminus j} \delta[x_k(t) = I]\right)}_{i \text{ does not get infected}} \underbrace{\beta dt \left(\sum_{\ell \in \partial j \setminus i} \delta[x_\ell(t) = I]\right)}_{j \text{ gets infected}}\right]$$
$$+ \mathbb{E}\left[\delta[x_i(t) = S]\delta[x_j(t) = I] \underbrace{\left(1 - \beta dt \sum_{k \in \partial i \setminus j} \delta[x_k(t) = I] - \beta dt\right)}_{i \text{ does not get infected}} \underbrace{(1 - \mu dt)}_{j \text{ recovers}}\right].$$

Now, there are two approximations we can perform to simplify the analysis. The first one is just to take the limit for $dt \to 0$ and remove the higher order terms. The second one is done exploiting conditional independence. We first remove the higher order terms in dt

$$\chi_{ij}(t + dt) \overset{dt \to 0}{=} \beta dt \left(\sum_{\ell \in \partial j \setminus i} \mathbb{E}\left[\delta[x_i(t) = S]\delta[x_j(t) = S]\delta[x_\ell(t) = I]\right]\right)$$

$$+\chi_{ij}(t)(1-\beta dt-\mu dt)-\beta dt\sum_{k\in\partial i\setminus j}\mathbb{E}\left[\delta[x_i(t)=S]\delta[x_j(t)=I]\delta[x_k(t)=I]\right].$$

We now exploit conditional independence. We denote with $\Omega_{ij}=\mathbb{P}(x_i(t)=S,x_j(t)=S)$ and $s_i=\mathbb{P}(x_i(t)=S)$ and write

$$\mathbb{E}\left[\delta[x_i(t)=S]\delta[x_j(t)=S]\delta[x_\ell(t)=I]\right]=\frac{\Omega_{ij}(t)\cdot\chi_{j\ell}(t)}{s_j(t)}$$
$$\mathbb{E}\left[\delta[x_i(t)=S]\delta[x_j(t)=I]\delta[x_k(t)=I]\right]=\frac{\chi_{ik}(t)\cdot\chi_{ij}(t)}{s_i(t)},$$

thus turning the state evolution equation into

$$\chi_{ij}(t+dt)\overset{dt\to0}{=}\beta dt\frac{\Omega_{ij}(t)}{s_j(t)}\sum_{\ell\in\partial j\setminus i}\chi_{j\ell}(t)+\chi_{ij}(t)(1-\beta dt-\mu dt)-\beta dt\frac{\chi_{ij}(t)}{s_i(t)}\sum_{k\in\partial i\setminus j}\chi_{ik}(t).$$

To get the epidemic threshold we now want to linearize around the epidemic-free fixed point that is $\Omega_{ij},s_i,s_j\to1$ and $\chi_{ij}\to0$ and get

$$\chi_{ij}(t+dt)\overset{dt\to0}{=}\beta dt\sum_{\ell\in\partial j\setminus i}\chi_{j\ell}(t)+\chi_{ij}(t)(1-\beta dt-\mu dt),$$

that can be written as

$$\partial_t\boldsymbol{\chi}(t)=(\beta B-(\beta+\mu)I_{2|\mathcal{E}|})\boldsymbol{\chi}(t).$$

Injecting this result in Equation (3.10)

$$\begin{pmatrix}\partial_t\boldsymbol{p}(t)\\ \partial_t\boldsymbol{\chi}(t)\end{pmatrix}=\begin{pmatrix}-\mu I_n & \beta T\\ \beta B-(\beta+\mu)I_{2|\mathcal{E}|} & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{p}(t)\\ \boldsymbol{\chi}(t)\end{pmatrix}.$$

From a simple calculation one sees that the stability condition is now obtained on heterogeneous random graphs as

The reproductive number according to the cavity method

$$R_0=\frac{\beta(\rho(B)-1)}{\mu}=\frac{\beta}{\mu}\left(\frac{\langle d^2\rangle}{\langle d\rangle}-1\right).$$

If we compare this result with the one obtained from NMF, the main difference is the appearance of the term "-1" that accounts the fact that when there is a susceptible-infected pair (ij), certainly i did not infect j. This comes "for free", in the sense that we did not have to add this correction manually and it naturally came from the equations. Other than that the results seem quite similar, but we must not forget that $\rho(A)$ is very close to $\rho(B)$ on dense networks but not on sparse ones. In other words, $\frac{\langle d^2\rangle}{\langle d\rangle}$ still is the good quantity to look at, but it is actually not the spectral radius of A in the sparse regime. Using the method explained in Chapter 2, we can derive the R_0 of the SIR model in the presences of vaccination. Note that, while with NMF this

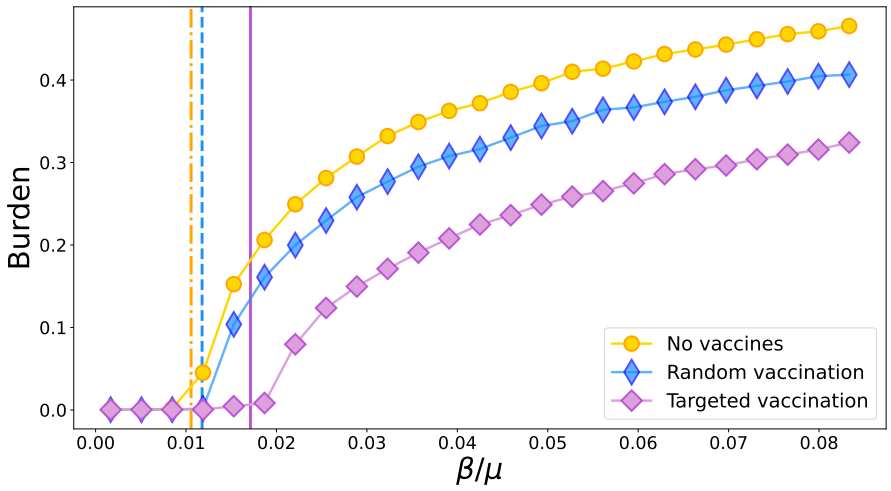

Figure 3.3: **Epidemic burden as a function of β/μ for different vaccination strategies**. The yellow dots are the burden (fraction of non-susceptible people at the end of the simulation) in absence of vaccination; the blue diamonds correspond to the random vaccination; the purple squares are the targeted vaccination in which each person is vaccinated with a probability proportional to the degree. The vertical lines (color coded) correspond to the position of the transition as predicted by the cavity method for the three different scenarios.

result was not rigorous because the effect of vaccination is to sparsify the network, with the cavity method we obtain e precise bound. The simulation shows the goodness of the cavity method in this setting as well as the efficacy of targeted vaccination strategies, as shown in Figure 3.3, we introduce $Q_{i,(ab)} = \delta_{ib}A_{ab}$ and $M_{(ab),(cd)} = A_{ab}A_{cd}\delta_{bc}\delta_{ad}$.

3.5 EFFICIENTLY COMPUTING THE SPECTRAL RADIUS OF B

We have seen that the non-backtracking matrix naturally appears when adopting the cavity approximation. The B matrix defined in Equation (3.9) however is large (its size scales with the number of edges, not of nodes) and it might not be so straightforward to build. However, we now show that there is a matrix B_{p} of size $2n \times 2n$ whose eigenvalues are also eigenvalues of B and it is much more easily built. Similarly to the matrix T introduced earlier $T_{i,(ab)} = \delta_{ia}A_{ab}$, we introduce $Q \in \mathbb{R}^{n\times 2|\mathcal{E}|}$ and $M \in \mathbb{R}^{2|\mathcal{E}|\times 2|\mathcal{E}|}$

$$\begin{aligned} Q_{i,(ab)} &= \delta_{ib}A_{ab} \\ M_{(ab),(cd)} &= \delta_{bc}\delta_{ad}A_{ab}A_{cd}. \end{aligned}$$

Now, the following relations[4] are satisfied:

4 You may try to obtain these relations as an exercise.

$$
\begin{aligned}
Q^T T - M &= B \\
TQ^T &= A \\
QQ^T &= D \\
QM &= T \\
TM &= Q
\end{aligned}
$$

With these relations at hand, suppose $\boldsymbol{g}$ is the leading eigenvector of B with eigenvalue ρ, then

$$\rho T\boldsymbol{g} = TB\boldsymbol{g} = T(Q^T T - M)\boldsymbol{g} = AT\boldsymbol{g} - Q\boldsymbol{g},$$

and

$$\rho Q\boldsymbol{g} = QB\boldsymbol{g} = Q(Q^T T - M)\boldsymbol{g} = DT\boldsymbol{g} - T\boldsymbol{g}.$$

Denoting $P\boldsymbol{g} = \boldsymbol{x}$ and $Q\boldsymbol{g} = \boldsymbol{y}$ for simplicity, we obtain,

$$\underbrace{\begin{pmatrix} A & -I_n \\ D - I_n & 0 \end{pmatrix}}_{B_{\mathrm{P}}} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \rho \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}, \tag{3.11}$$

where we introduced the smaller matrix $B_{\mathrm{P}} \in \mathbb{R}^{2n \times 2n}$. This matrix has the same eigenvalues as B (except those equal to ± 1 that have a different degeneracy). This matrix can thus be used to efficiently compute $\rho(B)$.

3.6 CONCLUSION

In this section we introduced the cavity method and the closely related non-backtracking matrix. Unlike the NMF approximation, this method is well suited for sparse graphs, being asymptotically exact on sparse random graphs, such as the ER. Consequently, this "second order" approximation (in which we assume independence at the edge, rather than the node level) is more accurate, but we must recall that real-world networks are often sparse but not locally tree-like. Sparse random graphs, in fact, tend to have a much smaller clustering coefficient than a real network with the same average degree. The low clustering coefficient, however, implies the absence of short loops, making the cavity method exact on random, but not on real world graphs, in general. This approximation is a improvement over NMF approach that is appropriate to deal with some of its limitations.

3.7 REFERENCES

- Wainwright, Jordan: *Graphical Models, Exponential Families, and Variational Inference*

This review is a milestone for physics methods on graphs and in Chapter 4 it treats the cavity method.

- Mezard, Montanari: *Information, Physics and Computation.*
  This is another relevant reference. The cavity method is discussed in Chapter 14.