

CONTAGIONS IN RANDOM NETWORKS WITH OVERLAPPING COMMUNITIES

EMILIE COUPECHOUX,* *Université Nice Sophia Antipolis and INRIA-ENS*

MARC LELARGE,** *INRIA-ENS*

Abstract

We consider a threshold epidemic model on a clustered random graph model obtained from local transformations in an alternating branching process that approximates a bipartite graph. In other words, our epidemic model is such that an individual becomes infected as soon as the proportion of his/her infected neighbors exceeds the threshold q of the epidemic. In our random graph model, each individual can belong to several communities. The distributions for the community sizes and the number of communities an individual belongs to are arbitrary. We consider the case where the epidemic starts from a single individual, and we prove a phase transition (when the parameter q of the model varies) for the appearance of a cascade, i.e. when the epidemic can be propagated to an infinite part of the population. More precisely, we show that our epidemic is entirely described by a multi-type (and alternating) branching process, and then we apply Sevastyanov's theorem about the phase transition of multi-type Galton–Watson branching processes. In addition, we compute the entries of the mean progeny matrix corresponding to the epidemic. The phase transition for the contagion is given in terms of the largest eigenvalue of this matrix.

Keywords: Random graphs; threshold epidemic model; branching processes; clustering

2010 Mathematics Subject Classification: Primary 60C05

Secondary 05C80; 91D30

1. Introduction

The spread of diseases or email viruses is well modeled by classical (SI, SIR, SIS) epidemics, whose study on complex networks has attracted a lot of attention in recent years; see Newman [24] for a review on complex networks and an introduction to epidemiological processes on such networks. The simplest epidemic model was first formulated by Reed and Frost for a population in which any two individuals can be in contact; see [28] for more details. It can be easily generalized on networks, and in that case is equivalent to a bond percolation on the network [21], which is widely studied on random graphs (see [15] and [19] for bond percolation on the configuration model).

In such epidemic models, each node can be independently influenced by each of his/her neighbors. For the diffusion of an innovation, an individual's adoption behavior is highly correlated with the behavior of his/her neighbors [27], and threshold epidemic models are more appropriate to model such diffusions. In this paper we consider the game-theoretic contagion model proposed by Blume [4] and Morris [20], and described below.

Consider an undirected graph G in which the nodes are the individuals in the population and there is an edge (i, j) if i and j can interact with each other. Each node has a choice between

Received 6 February 2014; revision received 27 November 2014.

* Postal address: Laboratoire I3S, Université Nice Sophia Antipolis, CS 40121, 06903 Sophia Antipolis Cedex, France.
Email address: emilie.coupechoux@gmail.com

** Postal address: INRIA-ENS, 23, avenue d'Italie, CS 81321, 75214 Paris Cedex 13, France.

two possible actions labeled A and B . On each edge (i, j) , there is an incentive for i and j to have their actions match, which is modeled as the following coordination game parametrized by a real number $q \in (0, 1)$. If i and j choose A (respectively B), they each receive a payoff of q (respectively $(1 - q)$); if they choose opposite actions, then they receive a payoff of 0. Then the total payoff of a player is the sum of the payoffs with each of his/her neighbors. If the degree of node i is d_i and S_i^B is the number of his/her neighbors playing B , then the payoff to i from choosing A is $q(d_i - S_i^B)$ while the payoff from choosing B is $(1 - q)S_i^B$. Hence, in a best-response dynamic, i should adopt B if $S_i^B > qd_i$ and A if $S_i^B \leq qd_i$. A number of qualitative insights can be derived from such a model even at this level of simplicity [16], [27]. Specifically, consider a network where all nodes initially play A . If a small number of nodes are forced to adopt strategy B (the seed) and we apply best-response updates to other nodes in the network, then these nodes will be repeatedly applying the following rule: switch to B if enough of your neighbors have already adopted B . There can be a cascading sequence of nodes switching to B such that a network-wide equilibrium is reached in the limit.

In [20], the contagion threshold of a connected infinite network was defined as the maximum threshold q_c at which a finite set of initial adopters can cause a complete cascade, i.e. the resulting cascade of adoptions of B eventually causes every node to switch from A to B . There are two possible models to consider depending on whether the initial adopters changing from A to B apply or not a best-response update. It was shown in [20] that the same contagion threshold arises in both models. Hence, without loss of generality, we restrict ourselves to the model where the initial adopters are forced to play B forever. In this case, the diffusion is monotone and the number of nodes playing B is nondecreasing. When the graph G is deterministic, in [20] Morris computed (and gave general bounds on) the corresponding contagion threshold.

The graph G that we consider here will have most of the properties observed in real-world networks. One of the most striking features shared by real-world networks is the scale-free property [2]: their degree distribution follows a power law. Random graphs with an arbitrary degree distribution [3] cover this property. The contagion model on such graphs was studied by heuristic means by Watts [29], and a generalization of it was studied rigorously by Lelarge [17]. In particular, it was shown in [17] that the notion of contagion threshold for a sequence of random graphs should be weakened in order to make sense (indeed random graphs are typically not connected) and the notion of pivotal players was introduced. Informally, a pivotal player is a node such that if this node is forced to play B forever and it is the only node playing B , then a global cascade occurs. Hence, in this paper we consider the case where a node playing B is forced to play B forever and where the seed consists of only one vertex. The graph G will be infinite and random, and we are interested in the cascade phenomenon, i.e. when an infinite subset of the population will eventually adopt B . In this setting, we will show a phase transition for this phenomenon, depending on the value of the parameter q of the model.

Another feature of real-world networks is that they all have a high clustering coefficient [30]; see also [23] for several examples. The clustering coefficient of a graph is, by definition, the probability that two given nodes are connected, knowing that they have a common neighbor. Since the asymptotic clustering coefficient of random graphs with an arbitrary degree distribution is 0 (locally, they look like trees), this random graph model fails to cover the clustering property of real-world networks. Recently, the contagion on clustered random graph models was studied by heuristic means in [11], and rigorously in [7] and [8] (in which a generalization of the contagion model is considered). The random graph models considered have a tunable clustering coefficient and an arbitrary degree distribution, which in particular allows the study of the clustering impact on the contagion model. However, these random graph models do

not cover the following property: in real-world networks, a node often belongs to several communities. A community is a set of nodes which are densely connected internally and only sparsely connected with other nodes of the network; in the clustered random graph models mentioned above, communities are represented by cliques, and a node can only belong to at most one clique. On the contrary, and as explained in [12] and [25], the structure of many real-world networks is close to the one-mode projection of a bipartite graph, in which each node belongs (possibly) to several cliques (communities).

The one-mode projection of a bipartite graph is defined as follows. Let $\Gamma \subset V \times E$ be a bipartite graph, i.e. a graph with two types of nodes: V -nodes and E -nodes, and in which there are no edges between nodes of the same type. For instance, each E -node of Γ represents a ‘community’ (a movie), and the V -nodes linked to a common E -node are members of the same community (actors in the same movie). The one-mode projection of Γ on the V -nodes (actors) is a unipartite graph: the nodes are the V -nodes of Γ , and there is an edge between two V -nodes if they belong to at least one common community (if these actors played together in at least one movie). One can construct a random graph by considering the one-mode projection of a random bipartite graph. To the best of the authors’ knowledge, neither rigorous proofs nor heuristics have been carried out for the contagion on such a model. There are several random bipartite graph models, and even the literature on the classical SIR epidemic on the one-mode projection of such models is incomplete. When the random bipartite graph has arbitrary degree distribution for each one of both types of nodes, heuristics have been derived by Newman [22] for the classical SIR epidemic on the one-mode projection of it. Rigorous results (for the classical SIR epidemic) were obtained by Britton *et al.* [6] and by Bollobás *et al.* [5], but the random graphs considered are such that their asymptotic degree distributions are respectively Poisson and mixed Poisson. Recently, Hackett *et al.* [13] studied by heuristic means the contagion model on random graphs with overlapping communities, i.e. with nodes that can belong to several cliques. They also derived results about the clustering effect on the contagion spread (for their random graph model). However, the communities in that random graph model are only of size three (that model does not come from the one-mode projection of a random bipartite graph).

Our random graph model is inspired from the one-mode projection of a random bipartite graph with arbitrary degree distributions. More precisely, we consider the one-mode projection of an alternating branching process that approximates locally this random bipartite graph (see [9, Section 7.2] for this approximation), and study rigorously the contagion on this random graph model. Our goal is twofold:

- (i) we study *rigorously* the contagion on random graphs with overlapping communities;
- (ii) our study provides heuristics for the contagion on the one-mode projection of random bipartite graphs with arbitrary degree distributions.

This paper is organized as follows. In Section 2 we define our random graph model and recall its degree distribution and clustering coefficient. In Section 3 we state our theorem about the phase transition for the contagion spread on our random graph model with overlapping communities. In Section 4 we compute the entries of the matrix involved in the phase transition. Finally, in Section 5 we prove this phase transition, applying Sevastyanov’s theorem about the phase transition of multi-type Galton–Watson branching processes.

2. Random graph model and its basic properties

In this paper we consider the one-mode projection of an alternating branching process. We define this branching process in Section 2.1, and our random graph model in Section 2.2. The degree distribution and clustering coefficient of our random graph model are the same as those computed in [25, Section IV], and are recalled in Section 2.3.

Let $\mathbf{p} = (p_d)_d$ and $\mathbf{q} = (q_w)_w$ be two probability distributions with positive finite means $\lambda := \sum_d dp_d \in (0, \infty)$ and $\mu := \sum_w wq_w \in (0, \infty)$, respectively.

2.1. Alternating branching process

Our random graph model is constructed from the following alternating branching process $\Gamma(\mathbf{p}, \mathbf{q})$, whose definition is given in Section 2.1.1 and phase transition (for $\Gamma(\mathbf{p}, \mathbf{q})$ to be finite/infinite) recalled in Section 2.1.2.

2.1.1. *Definition.* The branching process $\Gamma(\mathbf{p}, \mathbf{q})$ is an alternating one: each node is either of type V or E , and a generation of V -nodes gives birth to a generation of E -nodes, and conversely.

Let \tilde{D} and \tilde{W} be independent random variables with the following distributions (for any $d \geq 1, w \geq 1$):

$$\mathbb{P}(\tilde{D} = d - 1) = \frac{dp_d}{\lambda}, \quad \mathbb{P}(\tilde{W} = w - 1) = \frac{wq_w}{\mu}.$$

The variable \tilde{D} (respectively \tilde{W}) represents the offspring number of a V -node (respectively E -node), except for the root (whose offspring distribution is \mathbf{p}). The reasons why we choose this particular distribution are the following.

- This distribution gives a unimodular tree (the distribution is invariant by rerooting).
- The branching process $\Gamma(\mathbf{p}, \mathbf{q})$ is a local approximation for the random bipartite graph $\mathcal{B} = \mathcal{B}(\mathbf{p}, \mathbf{q})$ with arbitrary degree distributions \mathbf{p} and \mathbf{q} [9, Section 7.2]. Informally, the root of the branching process Γ represents a ‘typical’ vertex in \mathcal{B} . Hence, our random graph model G is a local approximation for the one-mode projection $G_{\mathcal{B}}(\mathbf{p}, \mathbf{q})$ of \mathcal{B} , and the root of G represents a ‘typical’ vertex in $G_{\mathcal{B}}(\mathbf{p}, \mathbf{q})$.

We can formally define the branching process $\Gamma = \Gamma(\mathbf{p}, \mathbf{q})$ as follows (only the definition of the i th V - and E -generations are required in the remaining sections). The root has type V and offspring distribution \mathbf{p} : its number D_0 of children satisfies $\mathbb{P}(D_0 = d) = p_d$. Let $(\tilde{W}_j^{(i)})_{i,j \geq 1}$ (respectively $(\tilde{D}_k^{(i)})_{i,k \geq 1}$) be random variables distributed as \tilde{W} (respectively \tilde{D}), all variables being independent and independent from D_0 . Then each child j ($1 \leq j \leq D_0$) of the root is an E -node that gives birth to $\tilde{W}_j^{(1)}$ V -nodes, so that the root has $\xi_V^{(1)} = \sum_{j=1}^{D_0} \tilde{W}_j^{(1)}$ grandchildren. Each such node k , $1 \leq k \leq \xi_V^{(1)}$, is a V -node that gives birth to $\tilde{D}_k^{(1)}$ E -nodes. Set $\xi_E^{(1)} = \sum_{k=1}^{\xi_V^{(1)}} \tilde{D}_k^{(1)}$. Notation is summarized in Figure 1.

The root corresponds to V -generation numbered 0, its children to E -generation 0, and so on (until now, we constructed V and E -generations 0 and 1). Assume that generations 0 to $i - 1$ are constructed ($i \geq 2$), with $\xi_E^{(i-1)}$ for the number of nodes in the $(i - 1)$ th E -generation. Then each node j , $1 \leq j \leq \xi_E^{(i-1)}$, gives birth to $\tilde{W}_j^{(i)}$ V -nodes, and we set $\xi_V^{(i)} = \sum_{j=1}^{\xi_E^{(i-1)}} \tilde{W}_j^{(i)}$ for the total number of nodes in the i th V -generation. Each such node k , $1 \leq k \leq \xi_V^{(i)}$,

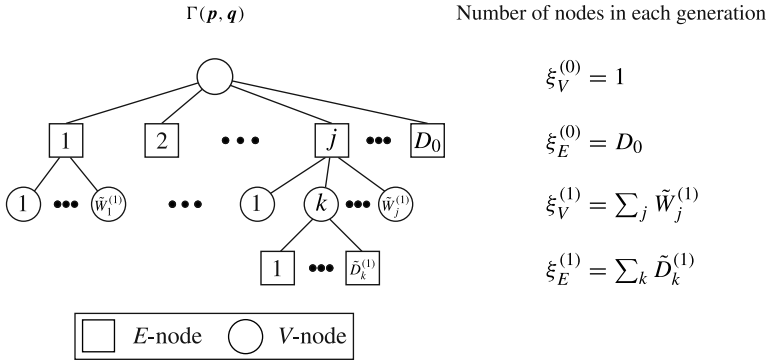


FIGURE 1: Notation for the alternating branching process $\Gamma(\mathbf{p}, \mathbf{q})$.

gives birth to $\tilde{D}_k^{(i)}$ E -nodes, and we set $\xi_E^{(i)} = \sum_{j=1}^{\xi_V^{(i)}} \tilde{D}_j^{(i)}$ for the total number of nodes in the i th E -generation.

2.1.2. *Extinction versus survival.* Let D (respectively W) be a random variable with distribution \mathbf{p} (respectively \mathbf{q}). We define the following generating functions for $x \in [0, 1]$:

$$F(x) = \sum_{d \geq 0} p_d x^d, \quad G(x) = \sum_{w \geq 1} \frac{w q_w}{\mu} x^{w-1}, \quad H(x) = \sum_{d \geq 1} \frac{d p_d}{\lambda} x^{d-1}.$$

The phase transition for $\Gamma(\mathbf{p}, \mathbf{q})$ to be finite/infinite is given by the next proposition, which is a direct consequence of [1, Theorem 1, Section I.A.5] (heuristics can also be found in [25, Section IV.A]).

Proposition 1. *Let p_{ext} be the probability that the branching process $\Gamma(\mathbf{p}, \mathbf{q})$ is finite.*

- *If $\mathbb{E}[W(W - 1)]\mathbb{E}[D(D - 1)] \leq \mathbb{E}[W]\mathbb{E}[D]$ then $p_{\text{ext}} = 1$.*
- *If $\mathbb{E}[W(W - 1)]\mathbb{E}[D(D - 1)] > \mathbb{E}[W]\mathbb{E}[D]$ then $p_{\text{ext}} < 1$. More precisely, we have*

$$p_{\text{ext}} = (F \circ G)(\eta),$$

where $\eta := \inf\{x \in [0, 1]: (H \circ G)(x) = x\}$.

2.2. Random graph model

We construct the (rooted and unipartite) random graph $G = G_\Gamma(\mathbf{p}, \mathbf{q})$ as follows. The root of G is the root of $\Gamma = \Gamma(\mathbf{p}, \mathbf{q})$, and the parent and the children of each E -node e in Γ are connected into a clique before e is removed, as illustrated in Figure 2. In other words, the random graph $G_\Gamma(\mathbf{p}, \mathbf{q})$ is the one-mode projection of Γ .

Even if the graph G is not a tree, its particular construction allows us to use the same terminology as for a tree (see Figure 2). More precisely, let v be a node in G , and let d be its distance from the root. By construction, there is exactly one node u , among the neighbors of v , which is at distance $d - 1$ from the root: it is called the *parent* of v . In addition, the neighbors of v that are at distance d from the root are called the *brothers* of v , and those at distance $d + 1$ the *children* of v . If we consider a given clique of G , the *parent of the clique* is the node whose distance from the root is minimal.

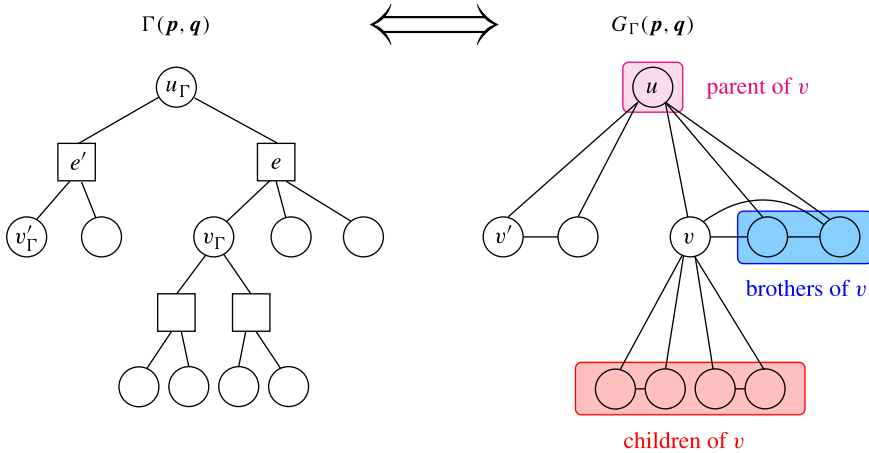


FIGURE 2: Random graph model $G_\Gamma(\mathbf{p}, \mathbf{q})$, constructed from the branching process $\Gamma(\mathbf{p}, \mathbf{q})$.

Note that the random graph G is infinite if and only if Γ is, so that the phase transition for G to be finite/infinite is given by Proposition 1 above.

2.3. Degree distribution and clustering coefficient

This section contributes to the description of our random graph model, but is not needed in the sequel. The next proposition can be found in Newman [22, Section III].

Proposition 2. (Newman [22].) *We consider the random graph $G_\Gamma(\mathbf{p}, \mathbf{q})$. Then the degree distribution D'_0 of the root is given by*

$$\mathbb{P}(D'_0 = k) = \sum_{d=1}^{\infty} p_d \sum_{w_1 + \dots + w_d = k+d} \frac{\prod_{j=1}^d w_j q_{w_j}}{\mu^d}.$$

Let T_0 be the number of triangles the root belongs to, and let P_0 be the number of connected triples the root belongs to. Then the local clustering coefficient C of the root is given by

$$\begin{aligned} C &:= \frac{\mathbb{E}(T_0)}{\mathbb{E}(P_0)} \\ &= \frac{1}{\mathbb{E}[W]} \mathbb{E}[W(W-1)(W-2)] \\ &\quad \times \left(\frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} \left(\frac{\mathbb{E}[W(W-1)]}{\mathbb{E}[W]} \right)^2 + \frac{1}{\mathbb{E}[W]} \mathbb{E}[W(W-1)(W-2)] \right)^{-1}, \end{aligned}$$

where D (respectively W) is a random variable with distribution \mathbf{p} (respectively \mathbf{q}).

3. Phase transition for the contagion: statement of our result

In Section 3.1 we define the contagion model described in the introduction, on the random graph $G = G_\Gamma(\mathbf{p}, \mathbf{q})$. In Section 3.2 we state our main theorem (whose proof is given in Section 5).

3.1. Epidemic model

Let $q \in (0, 1)$. We consider the contagion model described in the introduction, on the random graph $G = G_\Gamma(\mathbf{p}, \mathbf{q})$, with parameter q and the seed consisting in the root only. For simplicity, players B are called active vertices, and players A inactive vertices. The progressive dynamics of the contagion on the graph G operates as follows: the root starts out being active; all other vertices are inactive. Time operates in discrete steps $t = 1, 2, 3, \dots$. At a given time t , any inactive vertex becomes active if its proportion of active neighbors is strictly greater than q . Once active, a vertex stays active. Hence, the set of active vertices increases with time, and we define $G_{\text{act}}^{(q)}$ as the graph induced by the vertices in G that are active in the limit as time tends to ∞ . We say that a cascade occurs if the graph $G_{\text{act}}^{(q)}$ is infinite. We will show a phase transition for the cascade phenomenon, i.e. for the graph $G_{\text{act}}^{(q)}$ to be finite/infinite.

3.2. Phase transition

In the following, we make the additional assumption that the degrees of nodes in $\Gamma(\mathbf{p}, \mathbf{q})$ are bounded: there exist $\bar{d}, \bar{w} \geq 1$ such that $p_d = 0$ for $d > \bar{d}$ and $q_w = 0$ for $w > \bar{w}$. It ensures that the number of types for the multi-type branching process defined in Section 5.1 is finite, which is required in the proof of Theorem 1. In addition, we assume that $p_0 = q_0 = q_1 = 0$.

We consider the random graph $G = G_\Gamma(\mathbf{p}, \mathbf{q})$ defined in Section 2.2. We define the *type* of a vertex in G as its number of children in G . For all $x_0, x \in \{0, 1, \dots, \bar{d}\bar{w}\}$, we set $m_{x_0,x}$ for the mean number of *active* children of type x of an active vertex of type x_0 in G . In other words, we consider an active vertex u , different from the root, with x_0 children (if the probability that there exists such a vertex is 0, then we set $m_{x_0,x} = 0$). Once the contagion has spread among all its children, we count the number of such children that are active *and* have exactly x children. The mean of this quantity (among all possible realizations of G for the children and grandchildren of u) is called $m_{x_0,x}$. The matrix $M = (m_{x_0,x})_{0 \leq x_0, x \leq \bar{d}\bar{w}}$ will be computed in Section 4.

We have the following phase transition for the graph $G_{\text{act}}^{(q)}$ to be finite/infinite, and is proved in Section 5.

Theorem 1. *Let p_f be the probability that the random graph $G_{\text{act}}^{(q)}$ is finite, and let ρ be the largest eigenvalue of M . Then, we have*

- (i) if $q \geq \frac{1}{2}$ then $p_f = 1$;
- (ii) if $q < \frac{1}{2}$ then
 - if $p_2 = q_2 = 1$ then the random graph $G_{\text{act}}^{(q)}$ is infinite with probability 1;
 - otherwise $p_f = 1$ if and only if $\rho \leq 1$.

The idea developed in Sections 4 and 5 is the following. We first describe the epidemic (i.e. the graph $G_{\text{act}}^{(q)}$) by a branching process. The construction of the random graph G implies independence and recursion properties that left us with the study of the contagion spread inside one clique. Yet this propagation depends on the degree of the nodes (cf. Lemma 1 in Section 4.1). In other words, the number of active vertices at a given generation depends on the offspring number of these vertices. Our idea is to use a multi-type (and alternating) branching process to encode this information (cf. Section 5.1). The definition of this branching process requires not only the computation of the number of active vertices inside a clique, but rather the computation of the joint distribution for this number and the types of the vertices (cf. Lemma 2 in Section 4.1).

Once the epidemic has been described by a multi-type branching process, we use a phase transition theorem for multi-type branching processes. As developed in Sections 5.2 and 5.3,

classical results about multi-type branching processes do not apply directly in our case, since the matrix M defined above is not positively regular. We thus use Sevastyanov’s theorem. To complete the proof of Theorem 1, we compute the final classes (defined in Section 5.2) of our multi-type branching process.

We divided our work into two parts: although Lemmas 1 and 2 are used in both the proof of Theorem 1 and the computation of the matrix M , we chose to gather the computational part of our work in the next section.

4. Computation of the matrix M

Before computing the entries of the matrix M defined in Section 3.2, we state two lemmas about the number of active vertices inside a given clique. These lemmas will also be used in Section 5.1 (to define the multi-type branching process that describes the contagion spread in $G_{\Gamma}(p, q)$).

4.1. Lemmas about the number of active vertices inside a given clique

We consider a clique of size w whose parent u becomes active at time t . The other nodes inside the clique can be activated until time $t + w - 1$. More precisely, the first nodes (different from u) that are possibly activated inside the clique are those with fewer children. If such activations occur at time $t + 1$, then it can turn other nodes of the clique into active ones at time $t + 2$, and so on.

The following lemma computes the number of active vertices inside a clique, when the contagion has spread inside the clique.

Lemma 1. *We consider the contagion spread inside a clique of size w , whose parent u is initially active. Let L be the final number of active children (active vertices different from u) inside the clique. We denote by $\{1, \dots, w - 1\}$ the set of children inside the clique, and set X_i for the number of children of vertex i , $1 \leq i \leq w - 1$ (so that $X_{(i)} + w - 1$ is the degree of vertex i). Then L satisfies the following equation:*

$$L = \min\{i \in \{1, \dots, w - 1\} \mid \lfloor q(X_{(i)} + w - 1) \rfloor + 1 > i\} - 1, \tag{1}$$

where $X_{(1)} = \min_i X_i \leq X_{(2)} \leq \dots \leq X_{(w-1)} = \max_i X_i$ is the order statistics of $(X_i)_{1 \leq i \leq w-1}$, and $L = w - 1$ if this set is empty (i.e. if $\lfloor q(X_{(i)} + w - 1) \rfloor + 1 \leq i$ for all $1 \leq i \leq w - 1$).

Proof. By definition, a node i , $1 \leq i \leq w - 1$, becomes active if and only if the proportion of its active neighbors is strictly greater than q , i.e. if and only if its number of active neighbors is at least

$$A_i := \lfloor q(X_i + w - 1) \rfloor + 1. \tag{2}$$

We use the order statistics of $(A_i)_{1 \leq i \leq w-1}$ (or equivalently the one of $(X_i)_{1 \leq i \leq w-1}$): nodes with fewer children need fewer active neighbors to become active, and the first node(s) to become possibly active is (are) the one(s) with $X_{(1)}$ children. More precisely, if $A_{(1)} > 1$, no node different from u inside the clique can be activated, and $L = 0$. If $A_{(1)} \leq 1$, then at least one node (different from u) is activated. Then a second one is also activated if and only if $A_{(2)} \leq 2$, and the lemma follows by a simple induction.

The next lemma provides the joint distribution of L and the order statistics $(X_{(i)})_{1 \leq i \leq L}$, given the size w of the clique. The random variables $(X_i)_{1 \leq i \leq w-1}$ are independent and identically distributed (i.i.d.) and distributed as the following random variable X , which represents the

number of children of a vertex different from the root in $G_\Gamma(\mathbf{p}, \mathbf{q})$ (this equation is similar to the one of Proposition 2):

$$\mathbb{P}(X = k) = \sum_{d=1}^{\bar{d}} \frac{dp_d}{\lambda} \sum_{(w_1-1)+\dots+(w_{d-1}-1)=k} \frac{\prod_{j=1}^{d-1} w_j q_{w_j}}{\mu^{d-1}}. \tag{3}$$

For any sequence $x_1 \leq x_2 \leq \dots \leq x_\ell$ and any $i \in \{1, \dots, \ell - 1\}$, we set (omitting the dependency on (x_1, \dots, x_ℓ))

$$s_i := \begin{cases} \max\{j \geq 0 \mid x_i = x_{i+j}\} + 1 & \text{if } x_{i-1} < x_i \text{ or } i = 1, \\ 1 & \text{if } x_{i-1} = x_i. \end{cases} \tag{4}$$

When $x_{i-1} < x_i$ (or $i = 1$), s_i is the number of $i' \geq i$ such that $x_{i'} = x_i$. For instance, if $\ell = 6$ and $x_1 < x_2 = x_3 = x_4 < x_5 = x_6$, then $s_1 = 1, s_2 = 3, s_3 = s_4 = 1$, and $s_5 = 2$, so that $\prod_{i=1}^5 s_i! = 3! 2!$.

Lemma 2. *We consider a clique of size w whose parent u is initially active, and use the same notation as in Lemma 1. Then we have, for any $1 \leq \ell \leq w - 1$ and $0 \leq x_1 \leq x_2 \leq \dots \leq x_\ell$,*

$$\begin{aligned} p^{(E)}(\ell, x_1, \dots, x_\ell \mid w) &:= \mathbb{P}(L = \ell, X_{(1)} = x_1, \dots, X_{(\ell)} = x_\ell) \\ &= \left(\prod_{i=1}^{\ell} \mathbf{1}_{\{\lfloor q(x_i+w-1) \rfloor + 1 \leq i\}} \right) \frac{(w-1)!}{(w-1-\ell)! \prod_{i=1}^{\ell-1} s_i!} \\ &\quad \times \left(\prod_{i=1}^{\ell} \mathbb{P}(X = x_i) \right) (\mathbb{P}(\lfloor q(X+w-1) \rfloor > \ell))^{w-1-\ell}, \\ p^{(E)}(0 \mid w) &:= \mathbb{P}(L = 0) = (\mathbb{P}(\lfloor q(X+w-1) \rfloor > 0))^{w-1}. \end{aligned}$$

Proof. The joint distribution of the order statistics for a sequence of i.i.d. discrete random variables $(Y_i)_{1 \leq i \leq n}$, distributed as Y , is given by the following equation, for any $y_1 \leq y_2 \leq \dots \leq y_n$ (see [10, Equation (2.3)]),

$$\mathbb{P}(Y_{(1)} = y_1, \dots, Y_{(n)} = y_n) = \frac{n!}{\prod_{i=1}^{n-1} s_i!} \prod_{i=1}^n \mathbb{P}(Y = y_i) =: p_Y(y_1, \dots, y_n), \tag{5}$$

where $(s_i)_i$ (defined in (4)) correspond to the sequence (y_1, \dots, y_n) .

If $\lfloor q(x_i + w - 1) \rfloor + 1 > i$ for some $1 \leq i \leq \ell$, then $L < \ell$ (due to (1)), so that $p(\ell, x_1, \dots, x_\ell \mid w) = 0$. We assume that $\lfloor q(x_i + w - 1) \rfloor + 1 \leq i$ for all $1 \leq i \leq \ell$. Then, using (1) for the random variable X defined in (3), we have

$$\begin{aligned} p^{(E)}(\ell, x_1, \dots, x_\ell \mid w) &= \sum_{x_\ell < x_{\ell+1} \leq \dots \leq x_{w-1}} p_X(x_1, \dots, x_{w-1}) \mathbf{1}_{\{\lfloor q(x_{\ell+1}+w-1) \rfloor > \ell\}} \\ &= \sum_{x_{\ell+1} \leq \dots \leq x_{w-1}} p_X(x_1, \dots, x_{w-1}) \mathbf{1}_{\{\lfloor q(x_{\ell+1}+w-1) \rfloor > \ell\}}. \end{aligned} \tag{6}$$

The last equality comes from the fact that $\lfloor q(x_\ell + w - 1) \rfloor + 1 \leq \ell$. We set

$$g_X(\ell, w) := \sum_{x_\ell < x_{\ell+1} \leq \dots \leq x_{w-1}} \frac{(w-1-\ell)!}{\prod_{i=\ell+1}^{w-2} s_i!} \prod_{i=\ell+1}^{w-1} \mathbb{P}(X = x_i) \mathbf{1}_{\{\lfloor q(x_{\ell+1}+w-1) \rfloor > \ell\}}.$$

Then, using (6) and replacing $p_X(x_1, \dots, x_{w-1})$ by its expression in (5), we have

$$p^{(E)}(\ell, x_1, \dots, x_\ell \mid w) = \left(\frac{(w-1)!}{\prod_{i=1}^{\ell-1} s_i!} \prod_{i=1}^{\ell} \mathbb{P}(X = x_i) \right) \frac{g_X(\ell, w)}{(w-1-\ell)!}.$$

We now compute $g_X(\ell, w)$. Let $Z_1, \dots, Z_{w-1-\ell}$ be i.i.d. random variables distributed as X , and set $Z_{(1)} := \min_{1 \leq k \leq w-1-\ell} Z_k$. Then, using (5), we have

$$\begin{aligned} g_X(\ell, w) &= \sum_{x_{\ell+1} \leq \dots \leq x_{w-1}} p_X(x_{\ell+1}, \dots, x_{w-1}) \mathbf{1}_{\lfloor q(x_{\ell+1}+w-1) \rfloor > \ell} \\ &= \mathbb{P}(\lfloor q(Z_{(1)} + w - 1) \rfloor > \ell) \\ &= (\mathbb{P}(\lfloor q(X + w - 1) \rfloor > \ell))^{w-1-\ell}. \end{aligned}$$

This proves the first equation, and the second one follows from

$$\begin{aligned} \mathbb{P}(L = 0) &= \mathbb{P}(\lfloor q(X_i + w - 1) \rfloor + 1 > 1 \text{ for all } i \leq w - 1) \\ &= (\mathbb{P}(\lfloor q(X + w - 1) \rfloor > 0))^{w-1}, \end{aligned}$$

which completes the proof.

4.2. The number of active vertices of a given type

The next proposition gives the computation for the matrix M (whose largest eigenvalue is used in Theorem 1).

Proposition 3. *We define the following quantities, for $0 \leq x \leq \bar{d}\bar{w}$, $1 \leq w \leq \bar{w}$, $1 \leq k \leq w - 1$, $k + \lfloor q(x + w - 1) \rfloor \leq \ell \leq w - 1$, and $\lfloor q(x + w - 1) \rfloor \leq i \leq \ell - k$ (with $p^{(E)}(\ell, x_1, \dots, x_\ell \mid w)$ defined in Lemma 2),*

$$\begin{aligned} \mathcal{S}_{(x,k,\ell,i)} &:= \{(x_1, \dots, x_\ell) \mid x_1 \leq \dots \leq x_i < x_{i+1} = \dots = x_{i+k} = x < x_{i+k+1} \leq x_\ell\}, \\ P_{x \mid w}(k, \ell, i) &:= \sum_{(x_1, \dots, x_\ell) \in \mathcal{S}_{(x,k,\ell,i)}} p^{(E)}(\ell, x_1, \dots, x_\ell \mid w) \\ m_{x \mid w} &:= \sum_{1 \leq k \leq w-1} k \sum_{k + \lfloor q(x+w-1) \rfloor \leq \ell \leq w-1} \left(\sum_{\lfloor q(x+w-1) \rfloor \leq i \leq \ell-k} P_{x \mid w}(k, \ell, i) \right). \end{aligned}$$

Let v be an active vertex in $G = G_\Gamma(\mathbf{p}, \mathbf{q})$ with type x_0 , $0 \leq x_0 \leq \bar{d}\bar{w}$. Then the mean number $m_{x_0,x}$ of its active children having type x , $0 \leq x \leq \bar{d}\bar{w}$, is $m_{0,x} = 0$ if $x_0 = 0$ and otherwise is given by

$$m_{x_0,x} = \sum_{d=1}^{\bar{d}} \frac{dp_d}{\lambda} \sum_{(w_1-1)+\dots+(w_{d-1}-1)=x_0} \frac{\prod_{j=1}^{d-1} w_j q_{w_j}}{\mu^{d-1}} \sum_{i=1}^{d-1} m_{x \mid w_i}.$$

Proof. Using the definition of $p^{(E)}(\ell, x_1, \dots, x_\ell \mid w)$ in Lemma 2, we can easily see that $P_{x \mid w}(k, \ell, i)$ is the probability that a clique of size w , whose parent is initially active, has exactly ℓ active children, i of which having type strictly less than x and exactly k of which having type x .

Hence, the mean number, for a clique of size w , of its active children having type x is given by

$$N_{x|w} := \sum_{1 \leq k \leq w-1} k \sum_{0 \leq \ell \leq w-1} \left(\sum_{0 \leq i \leq \ell-k} P_{x|w}(k, \ell, i) \right).$$

In addition, for the children of type x to be activated inside the clique, there must be at least $\lfloor q(x + w - 1) \rfloor + 1$ active children of type strictly less than x (due to (2)). In other words, for $P_{x|w}(k, \ell, i)$ to be positive, we should have $i + 1 \geq \lfloor q(x + w - 1) \rfloor + 1$. Moreover, we have $\ell \geq k + i \geq k + \lfloor q(x + w - 1) \rfloor$, so that $N_{x|w} = m_{x|w}$.

The end of the proof follows easily, since we can study independently two different cliques having the same parent: with the notation of Figure 2, if the parent of v is active, the fact that v becomes active or not is independent of the activation/nonactivation of v' , when v' is not a brother of v .

5. Proof of Theorem 1

This section is organized as follows. First we define a multi-type (and alternating) branching process that completely describes the contagion spread in $G = G_\Gamma(\mathbf{p}, \mathbf{q})$ (using Lemma 2 of Section 4.1). Second we recall Sevastyanov’s theorem about the phase transition for multi-type Galton–Watson branching processes [26]. We then use this theorem to prove Theorem 1 (stated in Section 3.2).

5.1. Description of the contagion model by a multi-type (and alternating) branching process

We now define an alternating branching process Γ' in which the number of V -nodes in the i th V -generation is distributed as the number of vertices in the i th generation of $G_{\text{act}}^{(q)}$ (for any $i \geq 0$). As explained in Lemma 1, the number L of vertices at the end of the contagion spread inside a clique depends on:

- the clique size w ;
- the number of children in $G_\Gamma(\mathbf{p}, \mathbf{q})$ of each child i , $1 \leq i \leq w - 1$, inside the clique.

These dependencies lead to considering a multi-type branching process for Γ' (for an introduction on such processes; see, for example, Harris [14, Chapter 2] or Mode [18, Chapters 1 and 2]).

We need the following notation. For all $1 \leq d \leq \bar{d}$ and $0 \leq w_1, w_2, \dots, w_{d-1} \leq \bar{w}$, we set

$$p^{(V)}(d - 1, w_1, \dots, w_{d-1} \mid x) := \mathbf{1}_{\{\sum_{i=1}^{d-1} (w_i - 1) = x\}} \frac{dp_d}{\lambda} \left(\prod_{i=1}^{d-1} \frac{w_i q w_i}{\mu} \right) \frac{x!}{(H \circ G)^{(x)}(0)}, \quad (7)$$

where the generating functions G and H are defined in Section 2.1.2, and $(H \circ G)^{(x)}(0)$ is the value at point 0 of the x th derivative of $H \circ G$.

Proposition 4. *We define an alternating and multi-type branching process Γ' as follows. The root is a V -node that gives birth to d ($\leq \bar{d}$) E -nodes with probability p_d . Each of these E -nodes has type w with probability wq_w/μ .*

An E -node of type w gives birth to $\ell \in \{0, \dots, w - 1\}$ V -nodes of types $0 \leq x_1 \leq x_2 \leq \dots \leq x_\ell \leq \bar{d}\bar{w}$ with probability $p^{(E)}(\ell, x_1, \dots, x_\ell \mid w)$ defined in Lemma 2.

Finally, a V -node of type $x \geq 0$ gives birth to $d - 1$ E -nodes of types w_1, \dots, w_{d-1} with probability $p^{(V)}(d - 1, w_1, \dots, w_{d-1} \mid x)$ defined above.

As before, we set $\Gamma = \Gamma(\mathbf{p}, \mathbf{q})$ for the alternating branching process defined in Section 2.1, $G = G_\Gamma(\mathbf{p}, \mathbf{q})$ for the corresponding random graph model, and $G_{\text{act}}^{(q)}$ for the random graph of active vertices in G (as defined in Section 3.1). Then there is a coupling between Γ' and $G_{\text{act}}^{(q)}$ such that

- the number of V -nodes in the i th V -generation of Γ' is distributed as the number of vertices in the i th generation of $G_{\text{act}}^{(q)}$ (for any $i \geq 0$);
- the type w attached to each E -node in Γ' corresponds to its 'size' in Γ (i.e. its offspring number in Γ is $w - 1$);
- the type x attached to each V -node in Γ' corresponds to its number of children in G (i.e. its number of grandchildren in Γ is x).

Proof. Using Lemma 2 and the fact that the contagion spreads independently in two different cliques, we are left to prove that, when $\sum_{i=1}^{d-1} (w_i - 1) = x$, the probability

$$P := \mathbb{P}\left(\tilde{D} = d - 1, \tilde{W}_1 = w_1 - 1, \dots, \tilde{W}_{\tilde{D}} = w_{d-1} - 1 \mid \sum_{i=1}^{\tilde{D}-1} (\tilde{W}_i - 1) = x\right)$$

is given by (7), where \tilde{D} is the offspring number of a V -node in $G_\Gamma(\mathbf{p}, \mathbf{q})$ (with generating function H), and $(\tilde{W}_i)_i$ the offspring numbers of E -nodes in $G_\Gamma(\mathbf{p}, \mathbf{q})$ (with generating function G).

We have

$$P = \frac{\mathbb{P}(\tilde{D} = d - 1, \tilde{W}_1 = w_1 - 1, \dots, \tilde{W}_{\tilde{D}} = w_{d-1})}{\mathbb{P}(\sum_{i=1}^{\tilde{D}} \tilde{W}_i = x)}$$

The numerator is easy to compute:

$$\mathbb{P}(\tilde{D} = d - 1, \tilde{W}_1 = w_1 - 1, \dots, \tilde{W}_{\tilde{D}} = w_{d-1}) = \frac{dp_d}{\lambda} \prod_{i=1}^{d-1} \frac{w_i q_{w_i}}{\mu}$$

We use generating functions to compute the denominator. We have, for all $y \in [0, 1]$,

$$\begin{aligned} \mathbb{E}[y^{\sum_{j=1}^{\tilde{D}} \tilde{W}_j}] &= \sum_{d=1}^{\tilde{d}} \mathbb{P}(\tilde{D} = d - 1) \mathbb{E}[y^{\sum_{j=1}^{d-1} \tilde{W}_j}] \\ &= \sum_d \frac{dp_d}{\lambda} (\mathbb{E}[y^{\tilde{W}}])^{d-1} \\ &= (H \circ G)(y) \\ &= \sum_x \frac{1}{x!} (H \circ G)^{(x)}(0) y^x, \end{aligned}$$

which leads to

$$\mathbb{P}\left(\sum_{j=1}^{\tilde{D}} \tilde{W}_j = x\right) = \frac{1}{x!} (H \circ G)^{(x)}(0),$$

and completes the proof.

By definition, there is a cascade in $G_\Gamma(\mathbf{p}, \mathbf{q})$ if $G_{\text{act}}^{(q)}$ is infinite, which occurs if and only if the branching process Γ' is infinite.

5.2. Phase transition for multi-type Galton–Watson branching processes

We refer the reader to Harris [14, Chapter 2] for further information on multi-type Galton–Watson branching processes. Let Z be a multi-type Galton–Watson branching process, by k different types, starting from only one individual of a given type i_0 (results are the same when the process starts from a finite number of individuals, but we are interested only in this case here). Let $M = (m_{ij})_{1 \leq i, j \leq k}$ be the first-moment matrix of Z , i.e. m_{ij} is the mean number of children of type j that is created by a single individual of type i .

We next recall Sevastyanov’s theorem about the phase transition for multi-type Galton–Watson branching processes, that we will use in the next section. We define final classes as in [14, Chapter 2]. Let $m_{ij}^{(n)}$ be, for $n \geq 1$, the element in the i th row and j th column of M^n . The types i and j communicate if $m_{ij}^{(n)}, m_{ji}^{(n')} > 0$ for some $n, n' \geq 1$. A type that communicates neither with itself nor with any other type is called *singular*; a *class* is a set of types, each pair of which communicate, that is not contained in any other set having this property. The types fall uniquely into singular types and mutually exclusive classes. A *final class* C is a class in which any individual of type $i \in C$ has probability 1 to give birth to exactly one individual with type in C (other individuals whose type is not in C may also be produced). Let ρ be the largest eigenvalue of M . Sevastyanov’s theorem can be stated as follows.

Theorem 2. ([26].) *The probability of extinction of the branching process Z is 1 if and only if*

- (a) $\rho \leq 1$ and
- (b) *there are no final classes.*

The proof of this theorem originally appeared in Sevastyanov [26]. The statement used here corresponds to [14, Chapter 2, Theorem 10.1].

Remark 1. There is a classical version of Theorem 2 (with a simpler proof) for positively regular processes. The process Z is said to be *positively regular* if there exists $n \geq 1$ such that M^n is positive, i.e. all entries of M^n are (strictly) positive. The particular case of Theorem 2 for positively regular processes can be found in [14, Chapter 2, Theorem 7.1], or in any book dealing with multi-type branching processes (see also [18, Chapter 1] or [1, Chapter 5]). Some examples of nonpositively regular processes are also considered in [18, Chapter 2]. The hypothesis that the branching process has no final class is reduced to the hypothesis of nonsingularity in the positively regular case. We will see in the proof of Theorem 1 that the process we consider is not positively regular as soon as $p_1 > 0$, for instance, so that we need the stronger result of Sevastyanov.

5.3. Proof of Theorem 1

We construct a nonalternating branching process Γ'_V from Γ' by erasing E -generations, so that a V -generation directly gives birth to the next V -generation, and we apply Theorem 2 to the branching process Γ'_V .

Note that we have the following fact, as mentioned in the remark following Theorem 2. If $p_1 > 0$, then the process Γ'_V is not positively regular, since $m_{0x} = 0$ for all x . In addition, we cannot assume without loss of generality that $p_1 = 0$. Indeed the V -nodes of degree 1 in the original branching process $\Gamma = \Gamma(\mathbf{p}, \mathbf{q})$ play a role in the contagion inside a clique, since the

activation of the nodes with high degree depends on the activation of the nodes with low degree (cf. Lemma 1 in Section 4.1).

The proof of Theorem 1(i) is obvious. We assume that $q < \frac{1}{2}$. We will show that there is no final classes in Γ'_V if and only if either $p_2 < 1$ or $q_2 < 1$. Then applying Theorem 2 completes the proof (the $p_2 = q_2 = 1$ case is direct). We first show that, if there exists a final class C in Γ'_V , then necessarily $C = \{1\}$. We start with a definition and a lemma.

Definition 1. Let $0 < x \leq \bar{d}\bar{w}$. A configuration starting from x is an element σ_x of the form

$$\sigma_x = ((w_i, x_{i1}, \dots, x_{i\ell_i}))_{1 \leq i \leq d-1},$$

where $1 \leq d \leq \bar{d}$, $2 \leq w_i \leq \bar{w}$, $0 \leq \ell_i \leq w_i - 1$, and $0 \leq x_{i1} \leq \dots \leq x_{i\ell_i} \leq \bar{d}\bar{w}$ for all $1 \leq i \leq d - 1$. We say that a configuration σ_x has positive probability to occur if

- $p_d > 0$;
- $\sum_{i=1}^{d-1} (w_i - 1) = x$;
- $q_{w_i} > 0$ and $p^{(E)}(\ell_i, x_{i1}, \dots, x_{i\ell_i} \mid w_i) > 0$ for all $1 \leq i \leq d - 1$.

In other words, a configuration σ_x of positive probability is a possible realization for the next two generations starting from a V -node v of type x in Γ' (it contains all the information about the children and grandchildren of v in the sense that the w_i s represent the types of the children of v in Γ' , and x_{ij} the types of its grandchildren).

The next lemma provides a simple way to construct new configurations σ'_x of positive probability, if one knows a given configuration σ_x of positive probability (the proof is obvious, using the expression of $p^{(E)}(\ell_i, x_{i1}, \dots, x_{i\ell_i} \mid w_i)$ given in Lemma 2).

Lemma 3. Let $\sigma_x = ((w_i, x_{i1}, \dots, x_{i\ell_i}))_{1 \leq i \leq d-1}$ be a configuration that occurs with positive probability. Let $(y_{ij})_{i,j}$ be such that $y_{ij} \leq x_{ij}$ and $\mathbb{P}(X = y_{ij}) > 0$ for all $1 \leq i \leq d - 1$, $1 \leq j \leq \ell_i$, where the distribution of X is given by (3). Then the new configuration $\sigma'_x = ((w_i, y_{i1}, \dots, y_{i\ell_i}))_{1 \leq i \leq d-1}$ still occurs with positive probability.

We assume that there is at least one final class C , and first prove that necessarily $C = \{1\}$. Let $x \in C$ be such that $\mathbb{P}(X = x) > 0$, and let v be a V -node in Γ' with type x . We assume by contradiction that $x \neq 1$. Since 0 is singular, necessarily $x > 1$. By the definition of a final class, the number N of children of v whose type is in C is 1 almost surely. Hence, there exists a configuration $\sigma_x = ((w_i, x_{i1}, \dots, x_{i\ell_i}))_{1 \leq i \leq d-1}$ with a positive probability to occur and in which there exists a unique couple (i_0, j_0) such that $x_{i_0 j_0} \in C$. We distinguish several cases.

- If there exists (k, ℓ) such that $x_{k\ell} > x_{i_0 j_0}$, we consider the new configuration σ'_x with $y_{ij} = x_{ij}$ if $(i, j) \neq (k, \ell)$ and $y_{k\ell} = x_{i_0 j_0}$. Then $N = 2$ in σ'_x , and σ'_x has positive probability to occur by Lemma 3, which is a contradiction.
- If there exists (k, ℓ) such that $x_{k\ell} < x_{i_0 j_0}$, we consider the new configuration σ'_x with $y_{ij} = x_{ij}$ if $(i, j) \neq (i_0, j_0)$ and $y_{i_0 j_0} = x_{k\ell}$. Then $N = 0$ in σ'_x , and σ'_x has positive probability to occur by Lemma 3, which is a contradiction.
- Otherwise, v has only one grandchild (of type $x_{i_0 j_0}$) in Γ' . We first consider the case where $w_{i_0} = 2$. Since $\sum_{i=1}^{d-1} (w_i - 1) = x > 1$, there exists $k \neq i_0$ such that $w_k \geq 2$. We construct σ'_x by replacing w_k by $w_k - 1$ E -nodes of type 2, each of which having a

V -node of type $x_{i_0 j_0}$ as a child. Hence, $N = w_k \geq 2$ in σ'_x , and it is easy to see that σ'_x has positive probability to occur, which is a contradiction.

- The remaining case is when v has only one grandchild (of type $x_{i_0 j_0}$) in Γ' , and $w_{i_0} > 2$. We construct σ'_x by replacing $(w_{i_0}, x_{i_0 j_0})$ by $(w_{i_0}, x_{i_0 j_0}, x_{i_0 j_0})$. Hence, $N = 2$ in σ'_x , and it is easy to see that σ'_x has positive probability to occur, which is a contradiction.

Hence, $C = \{1\}$, which means that v has only one grandchild in the original branching process $\Gamma = \Gamma(\mathbf{p}, \mathbf{q})$. It implies in particular that $q_2 > 0$, and we are left to compute the following quantity:

$$p^{(E)}(\ell = 1, x = 1 \mid w = 2) = \mathbf{1}_{\lfloor 2q \rfloor \leq 0} \mathbb{P}(X = 1) = \frac{2p_2}{\lambda} \frac{2q_2}{\mu}.$$

By the definition of a final class, $p^{(E)}(\ell = 1, x = 1 \mid w = 2) = 1$, which implies that $p_2 = q_2 = 1$, and completes the proof.

6. Conclusion and perspectives

We studied rigorously the contagion [4], [20] on a clustered random graph model with overlapping communities. Our random graph model allows an arbitrary distribution for the community sizes, while the heuristic study of Hackett *et al.* [13] was performed on a random graph model with communities of size 3. To the best of the authors' knowledge, these are the only two studies of this epidemic model on random graphs with overlapping communities. There are several dependencies that made this study challenging:

- in the epidemic model itself, since the behavior of an individual depends on the behavior of all his/her neighbors;
- in the random graph model considered, that allows an arbitrary distribution for both the community size and the number of communities an individual belongs to.

In addition, our study provides heuristics for the contagion on the one-mode projection of a random bipartite graph with arbitrary degree distributions, which is well suited for modeling real-world networks [12], [25].

We showed that our epidemic is completely described by a multi-type and alternating branching process, and use a nonclassical theorem on phase transitions for multi-type branching processes, referred to as Sevastyanov's theorem, to prove a phase transition for our process. This opens the way to the study of the clustering effect on the cascade phenomenon in this case, as studied in [13] or [8].

Acknowledgement

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-11-JS02-005-01 (GAP project).

References

- [1] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, New York.
- [2] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512.
- [3] BENDER, E. A. AND CANFIELD, E. R. (1978). The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory Ser. A* **24**, 296–307.
- [4] BLUME, L. E. (1995). The statistical mechanics of best-response strategy revision. *Games Econom. Behav.* **11**, 111–145.

- [5] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2011). Sparse random graphs with clustering. *Random Structures Algorithms* **38**, 269–323.
- [6] BRITTON, T., DEIJFEN, M., LAGERÅS, A. N. AND LINDHOLM, M. (2008). Epidemics on random graphs with tunable clustering. *J. Appl. Prob.* **45**, 743–756.
- [7] COUPECHOUX, E. AND LELARGE, M. (2011). Impact of clustering on diffusions and contagions in random networks. In *Network Games, Control and Optimization* (NetGCoop, 2011), IEEE, New York, pp. 112–118.
- [8] COUPECHOUX, E. AND LELARGE, M. (2012). How clustering affects epidemics in random networks. Preprint. Available at <http://arxiv.org/abs/1202.4974>.
- [9] DARLING, R. W. R. AND NORRIS, J. R. (2008). Differential equation approximations for Markov chains. *Prob. Surveys* **5**, 37–79.
- [10] GAN, G. AND BAIN, L. J. (1995). Distribution of order statistics for discrete parents with applications to censored sampling. *J. Statist. Planning Infer.* **44**, 37–46.
- [11] GLEESON, J. P. (2008). Cascades on correlated and modular random networks. *Phys. Rev. E* **77**, 046117.
- [12] GUILLAUME, J.-L. AND LATAPY, M. (2006). Bipartite graphs as models of complex networks. *Physica A* **371**, 795–813.
- [13] HACKETT, A., MELNIK, S. AND GLEESON, J. P. (2011). Cascades on a class of clustered random networks. *Phys. Rev. E* **83**, 056107.
- [14] HARRIS, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.
- [15] JANSON, S. (2009). On percolation in random graphs with given vertex degrees. *Electron. J. Probab.* **14**, 87–118.
- [16] KLEINBERG, J. (2007). Cascading behavior in networks: algorithmic and economic issues. In *Algorithmic Game Theory*, Cambridge University Press, pp. 613–632.
- [17] LELARGE, M. (2012). Diffusion and cascading behavior in random networks. *Games Econom. Behav.* **75**, 752–775.
- [18] MODE, C. J. (1971). *Multitype Branching Processes – Theory and Applications*. American Elsevier, New York.
- [19] MOLLOY, M. AND REED, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6**, 161–179.
- [20] MORRIS, S. (2000). Contagion. *Rev. Econom. Stud.* **67**, 57–78.
- [21] NEWMAN, M. E. J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128.
- [22] NEWMAN, M. E. J. (2003). Properties of highly clustered networks. *Phys. Rev. E* **68**, 026121.
- [23] NEWMAN, M. E. J. (2003). Random graphs as models of networks. In *Handbook of Graphs and Networks*, Wiley-VCH, Weinheim, pp. 35–68.
- [24] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45**, 167–256.
- [25] NEWMAN, M. E. J., STROGATZ, S. H. AND WATTS, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118.
- [26] SEVAST'YANOV, B. A. (1951). The theory of branching random processes. *Uspehi Matem. Nauk (N.S.)* **6**, 47–99.
- [27] VEGA-REDONDO, F. (2007). *Complex Social Networks* (Econometric Soc. Monogr. **44**). Cambridge University Press.
- [28] VON BAHR, B. AND MARTIN-LÖF, A. (1980). Threshold limit theorems for some epidemic processes. *Adv. Appl. Probab.* **12**, 319–349.
- [29] WATTS, D. J. (2002). A simple model of global cascades on random networks. *Proc. Nat. Acad. Sci. USA* **99**, 5766–5771.
- [30] WATTS, D. J. AND STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442.