Cluster variation approximations for a contact process living on a graph

Nathalie Peyrard, Biometrie INRA (Avignon), Nathalie.Peyrard@avignon.inra.fr
Alain Franc, UMR BioGeCo INRA (Bordeaux), alain.franc@pierroton.inra.fr

Corresponding author : Nathalie Peyrard
Unite de Biometrie — tel : +33 (0)4 32 72 21 53
INRA Domaine St Paul - Site Agroparc — fax : +33 (0)4 32 72 21 82
84914 AVIGNON cedex 9, FRANCE — nathalie.peyrard@avignon.inra.fr

Abstract: A model classically used for modelling the spread of an infectious diseases in a network is the time continuous contact process, which is one simple example of interacting particles system. It displays a non-equilibrium phase transition, related to the burst of an epidemics within a population in case of an accidental introduction. Several studies have recently emphasized the role of some geometrical properties of the graph on which the contact process lives, like the degree distribution, for quantities of interest like the singlet density at equilibrium or the critical value of the infectivity parameter for the emergence of the epidemics, but this role is not yet fully understood. As the contact process on a graph still cannot be solved analytically (even in a 1D lattice), some approximations are needed. The more naive, but well studied approximation is the mean field approximation. We explore in this paper the potentiality of a finer approximation: the pair approximation used in ecology. We give an analytical formulation on a graph of the site occupancy probability at equilibrium, depending on the site degree, under pair approximation and another dependence structure approximation. We point out improvements brought in the case of realistic graph structures, far from the well-mixed assumption. We also identify the limits of the pair approximation to answer the question of the effects of the graph characteristics. We show how to improve the method using a more appropriate order cluster variation method, the Bethe approximation.

PACS code : 02.50.Ey; 11.80.Fv

Keywords: graph, contact process, phase transition, pair approximation, Bethe approximation.

1 Introduction

The contact process is a well studied toy model exhibiting a nonequilibrium phase transition ([1] or [2, chapter 6]). It has been defined on a lattice, where the cell state can be either 0 (empty) or 1 (alive). The time is continuous. A living cell becomes empty with rate dt. An empty cell becomes alive with rate $\beta a dt$ where a is the number of living neighbor cells. The transition is between an absorbing state, for $\beta < \beta_c$ where all cells are in the empty state,

and an excited state, for $\beta \geq \beta_c$ where the process is living forever on the lattice (rigorously, if the lattice size is infinite).

It has been designed for, and widely used, to study the spread of diseases within a population. The cells represent hosts, and the states are 'healthy' (0) or 'infected (1). The phase transition corresponds to the persistence of the disease, once an infected host has been introduced within the population. It has been used as well for modelling phenomena involving excitable media, like fire models [3], where the states are 'unburned' (0) and 'burning' (1), and the transition concerns the persistence of the fire. Following the works on networks of SIS models, (see [4]) a contact process living on a graph has been studied (as opposed to a regular grid), where the nodes model the hosts and the edges the paths for infectivity ([5]). Different types of diseases lead to different types of graphs. For example, some infectious disease in humans propagate through social networks, known to be scale free graphs [6]. Diseases in plants propagated by fungi lead to neighborhoods defined by distances ([7, 8]). A graph-based modeling approach has also been shown relevant to represent individuals movements within a network of cities in order to extend classical non-spatial SIR models for disease dynamics within each city ([9]).

One challenge is to understand how the structure of a graph can influence its function (here, propagation of a disease living on it). The structure of a graph can be described by the degree distribution (a random graph has a poisson degree distribution, whereas a scale-free graph exhibits a power law degree distribution, see [4]). More sophisticated indices are the clustering coefficient, correlation between degrees, etc. ... [4]. The difficulty is due to the model complexity: no analytical expression of the site occupancy probability is available and approximation are needed. One solution would be to explore the graph influence through simulations. We adopt here approximations from statistical mecanics, with the advantage of a generic analytical solution. A first step is to implement classical and tractable approximations, such as mean field approximation ([10]). When applied to the contact process, it leads to the family of SIR models often used in epidemiology. Although the mean field approximation assumes that the individuals are well mixed and behave like indepedent – the structure of the graph is not taken into account – it is possible to derive a mean field approximation taking into account the degree distribution [11, 12, 4, 5]. However, it is known to behave badly close to the phase transition. One challenge is to find a reliable approximation close to the phase transition, taking into account the structure of the graph. Therefore, we present here some cluster variation approximations for the contact process on a graph, which allows to take into account pair correlation (they are order 2 moment closure solutions). The cluster variation method has been developped in the 1930's in solid state physics, and has been successfully implemented on toy models like Ising or Pott model (see [13]). The approximations we consider here are the pair approximation and the Bethe approximation, whose generalization is the Kikuchi approximation in statistical physics (see [14, 15, 16, 17]).

To our knowledge, neither pair approximation nor the Bethe approximation have been developed for a contact process on a graph. Recently, [18] (see references therein) have developped a pair approximation for a spatialized SIRS model on an hypercubic lattice. Here, we develop for the CP on a graph (i) nested exact equations for the dynamics of singlets and

pairs (*ii*) a closure with pair approximation, which does not take into account the triangles (*iii*) finally a Bethe approximation, as a cluster variation method which introduces a correcting term in the PA, taking the existence of triangles into account. We identify then, through simulations, the types of graphs which can favor the mean field approximation and the ones where this method reaches its limits and can be improved by order 2 approximations.

The paper is organized as follows. The next section is dedicated to the description of the time continuous contact process model on a graph and to the exact (but not closed) expressions for transients of singlets and pairs. Then, in Section 3, we propose one way to implement pair approximation. The resulting analytic solution at equilibrium is established in Section 4 and compared to simulation results and mean field results on graphs with different distribution of the number of neighbors. The Bethe approximation with a correction term taking the density of triangles into account is presented in section 5. A conclusion ends the paper.

2 The model

2.1 Contact process on a graph

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined by a set $\mathcal{V}$ of vertices and a set $\mathcal{E}$ of edges. We will denote $s \sim r$ the event $(s, r) \in \mathcal{E}$. A vertex s is of degree k if s belongs to the subset $\mathcal{G}_k$ of vertices linked to exactly k other vertices by edges. We note $d_s \in \mathbb{N}$ the degree of vertex s. The degree distribution of a graph is given by the probability distribution $(p_k)_k$ where p_k is the probability for a vertex to be of degree k.
Each vertex s is in a state $z_s \in \{0, 1\}$ which varies with time, according to the dynamics of a time continuous contact process. The rules are as follow ([1] or [2, chapter 6]): if s is occupied (state 1) it becomes empty (state 0) with rate μ and we can always set $\mu = 1$. Concerning the contamination probability, if s is empty and has a_s occupied neighbors, it becomes occupied with rate βa_s, with β being the per site contamination rate. This parameter can also be expressed as $\beta = \lambda/\bar{d}$, with $\bar{d}$ the mean of the degree distribution. In order to specify the asymptotic behavior, it is standard to study the evolution of some quantity of interest, like: $\rho_k = P(z_s = 1 | d_s = k)$ and $\xi_{kl} = P(z_s = 0, z_r = 1 | s \sim r, d_s = k, d_r = l)$ The other densities for pairs are derived directly from these two.

2.2 Exact expression of the transients for singlets

The evolution of singlets is given by

$$\frac{d\rho_k}{dt} = (1 - \rho_k) P_k^{0\to 1} - \rho_k \tag{1}$$

where $P_k^{0\to 1} dt$ is the probability that a site of degree k and in state 0 moves to state 1 during the period dt. By definition

$$\begin{aligned} P_k^{0\to 1} dt &= P(z_s^{t+dt} = 1 | z_s^t = 0, d_s = k) \\ &= \sum_{\mathbf{z}} P(z_s^{t+dt} = 1 | z_s^t = 0, d_s = k, \mathbf{z}) \times P(\mathbf{z} | z_s^t = 0, d_s = k) \end{aligned}$$

where $\mathbf{z}$ is a configuration of the neighborhood of site s. Thus

$$P_k^{0\to 1} = \sum_{\mathbf{z}} \beta a(\mathbf{z}) P(\mathbf{z}|z_s^t = 0, d_s = k) \tag{2}$$

where $a(\mathbf{z})$ is the number of occupied vertices in configuration $\mathbf{z}$ of the neighborhood. Let us define $\mathcal{Z}_k$ as $\{0,1\}^k$. This is the set of all the possible configurations for $\mathbf{z}$ and we can consider the probability law Q on $\mathcal{Z}_k$ defined as

$$Q(\mathbf{z}) = P(\mathbf{z}|z_s^t = 0, d_s = k) \tag{3}$$

Equation (2) can then be read as follows:

$$P_k^{0\to 1} = \mathbb{E}_Q[\beta a(\mathbf{z})] \tag{4}$$

We use now the fact that $a(\mathbf{z})$ can we expressed as a sum of indicatrix variables

$$a(\mathbf{z}) = \sum_{i=1}^{k} \mathbf{1}_{\{z_{s(i)}=1\}} \tag{5}$$

where $s(1), \ldots, s(k)$ are the k vertices corresponding to the k neighbors of s (i.e. $\mathbf{z} = \{z_{s(i)}, i = 1, \ldots k\}$). Then

$$\begin{aligned} \mathbb{E}_Q[a(\mathbf{z})] &= \sum_{i=1}^{k} \mathbb{E}_Q\left[\mathbf{1}_{\{z_{s(i)=1}\}}\right] \\ &= \sum_{i=1}^{k} P(z_{s(i)} = 1 | z_s = 0, d_s = k) \end{aligned} \tag{6}$$

(We have dropped the reference to time in the notation since now all considered variables are at the same date). This yields

$$P_k^{0\to 1} = \beta \sum_{i=1}^{k} \frac{P(z_{s(i)} = 1, z_s = 0 | d_s = k)}{P(z_s = 0 | d_s = k)}$$

The probability $P(z_{s(i)} = 1, z_s = 0 | d_s = k)$ can be seen as a marginal probability, or a mean quantity:

$$\begin{aligned} P(z_{s(i)} = 1, z_s = 0 | d_s = k) &= \sum_{l} P(z_{s(i)} = 1, z_s = 0 | d_s = k, d_{s(i)} = l) p(d_{s(i)} = l | d_s = k) \\ &= \sum_{l} \xi_{kl} P(d_{s(i)} = l | d_s = k) \end{aligned}$$

We will denote $\xi_{k.}$ this expression. Then,

$$P_k^{0\to 1} = \beta \frac{k \xi_{k.}}{1 - \rho_k} \tag{7}$$

Hence

$$\begin{aligned}\frac{d\rho_k}{dt} &= (1-\rho_k)P_k^{0\to 1} - \rho_k \\ &= k\beta\xi_{k.} - \rho_k \end{aligned} \tag{8}$$

We have obtained an exact evolution equation for ρ_k involving the term $\xi_{k.}$ related to the pairs dynamics. $\xi_{k.}$ is unknown and depends itself on triplets dynamics. Hence, equation (8) is not closed, and is the first equation from a hierarchy involving probabilities of higher orders. However, we can already deduce from (8) a property at equilibrium:

$$\rho_k = k\beta\xi_{k.} \tag{9}$$

2.3 Exact expression of the transients for pairs

In the process to close (9) we consider now the evolution of pairs in the system.
When considering a pair evolution, four states are possible: (00), (01), (10) and (11). The probability to observe these states depends on the degree of the two sites of the pair. Let us study the evolution of pairs (00) among the pairs in $\mathcal{G}_{kl}$. During a period dt, pairs (00) can be created from pairs (01) or (10), and pairs (00) can disappear through creation of pairs of type (01) or (10). This can be expressed as

$$\frac{d(1-\rho_k-\xi_{kl})}{dt} = \xi_{kl} + \xi_{lk} - (1-\rho_k-\xi_{kl})(P_{kl}^{00\to 10} + P_{lk}^{00\to 10}) \tag{10}$$

and at equilibrium,

$$0 = \xi_{kl} + \xi_{lk} - (1-\rho_k-\xi_{kl})(P_{kl}^{00\to 10} + P_{lk}^{00\to 10}) \tag{11}$$

with $P_{kl}^{00\to 10}dt$ being the probability that a pair (00) in $\mathcal{G}_{kl}$ changes into state (10) during a period dt. First, we establish an exact equation for $P_{kl}^{00\to 10}$:

$$\begin{aligned}P_{kl}^{00\to 10}dt = \sum_{\tilde{\mathbf{z}}} & P(z_s^{t+dt}=1 | z_s^t=0, d_s=k, z_r^t=0, d_r=l, \tilde{\mathbf{z}}) \\ & \times P(\tilde{\mathbf{z}} | z_s^t=0, d_s=k, z_r^t=0, d_r=l)\end{aligned} \tag{12}$$

where $\tilde{\mathbf{z}}$ is the state, at time t, of the neighborhood of the site s except the known site r (note that this restricted neighborhood can be empty). By definition

$$P(z_s^{t+dt}=1 | z_s^t=0, d_s=k, z_r^t=0, d_r=l, \tilde{\mathbf{z}}) = \beta a(\tilde{\mathbf{z}})dt \tag{13}$$

where $a(\tilde{\mathbf{z}})$ is the number of occupied sites in $\tilde{\mathbf{z}}$ (we remind that the last neighbor, r, is empty). Thus

$$P_{kl}^{00\to 10} = \sum_{\tilde{\mathbf{z}}} \beta a(\tilde{\mathbf{z}}) P(\tilde{\mathbf{z}} | z_s^t=0, d_s=k, z_r^t=0, d_r=l) \tag{14}$$

$$\tag{15}$$

and in the same way that for singlets, this yields

$$\begin{aligned} P_{kl}^{00\to 10} &= \beta \sum_{i=1}^{k-1} P(z_{s(i)}=1 | z_s=0, z_r=0, d_s=k, d_r=l) \\ &= \beta \sum_{i=1}^{k-1} \frac{P(z_{s(i)}=1, z_s=0, z_r=0 | d_s=k, d_r=l)}{P(z_s=0, z_r=0 | d_s=k, d_r=l)} \end{aligned} \tag{16}$$

At this stage, the expression of $P_{kl}^{00\to 10}$ is an exact expression but is intractable. It involves probabilities on triplets. Approximation on the interactions are necessary in order to break this hierarchy, as presented in the following section.

3 Structure approximation

The system at equilibrium is described by the system of exact equations (9), (11) and (16). The unknown quantities in these equations are ρ_k, ξ_{kl} and $\xi_{k.}$. The system must be closed by choosing a way to approximate the triplets probability only in terms of the ρ_k, ξ_{kl} and $\xi_{k.}$. The simpler approximation is the mean field approximation of a contact process on a graph, which assumes that the sites are independent (see [11, 12, 5]). We present here the approach we have adopted, based on a simplification on the dependence structure induced by the degrees, and on a pair approximation, to take into account pair correlation between states. Such a technique has been implemented several times on regular lattices where k is constant over vertices and is the coordination number z (see [19, chapters 13 & 18]) or on graphs with constant degree (see [19, chapter 19]). Here, we extend it to more general graphs.

3.1 Degrees dependence structure

The main difference between a regular lattice and a graph is the introduction of an extra random variable: the degree of each site. This induces a new dependence structure, among the degrees, in addition of the dependence structure among the states. The consequence is that in general the degrees and the states are not independent. Let us consider for instance the evaluation of $P(z_{\mathcal{A}}, z_{\mathcal{B}} | d_{\mathcal{A}}, d_{\mathcal{C}})$ for disjoint subsets $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ of $\mathcal{V}$. Since the variables corresponding to the degrees of the sites are not independent and since the probability of the state of a site is not independent of its degree, the probability distribution actually depends on $d_{\mathcal{C}}$ and this conditional probability is not easily accessible. To circumvent the problem of degree dependence, we will assume (H) that for all disjoint subsets $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ of $\mathcal{V}$,

$$H: \quad P(z_{\mathcal{A}}, z_{\mathcal{B}} | d_{\mathcal{A}}, d_{\mathcal{C}}) = P(z_{\mathcal{A}}, z_{\mathcal{B}} | d_{\mathcal{A}})$$

In addition, the contact process is characterized by spatial dependencies between the states of the different sites of the graph and the computation of joint laws is of high complexity. In the next section we will address this second problem using the pair approximation.

3.2 Pair approximation on the states

The pair approximation is a method often used in ecology to simplify systems complexity (see [19, chapters 13,18,19]). The approximation is as follows, for any set of three sites a, b, c, with a linked to b and b to c:

$$p(z_a|z_b, z_c) \approx p(z_a|z_b) \tag{17}$$

Note that the approximation will be the same if c is linked to a or not. This implies the following approximation of the joint triplet distribution:

$$p(z_a, z_b, z_c) \approx \frac{p(z_a, z_b)p(z_b, z_c)}{p(z_b)} \tag{18}$$

Let us apply this principle to expression (16). We obtain, under (H):

$$P_{kl}^{00\to 10} = \beta\frac{(k-1)\xi_{k.}}{1-\rho_k} \tag{19}$$

It can be seen that $P_{kl}^{00\to 01}$ actually does not depend on l (as a consequence of H). One can note the similarity between this expression and the (exact) expression of the probability that a site of degree k moves from 0 to 1 during the period dt (see equation 7). The probability that a site s with k neighbors, and one of them at least in state 0, becomes occupied is approximated here by the probability that a site s with $k-1$ "effective" neighbors becomes occupied. This is quite natural, since the empty neighbor plays no role in a potential contamination of s.

4 Results

At this stage, under pair approximation, in the system of equations defining the equilibrium ((9), (11) and (19)), all the quantities have been expressed only in terms of singlet or pair densities. The quantities of interest are the ρ_ks and the ξ_{kl}s but the system still depends on the quantities $\xi_{k.}$s. We have established that under our approximation, $\xi_{kl} = \xi_{k.}$ at equilibrium and is equal to (see the Appendix for computational details)

$$\xi_{k.} = \beta(1-\rho_k-\xi_{k.})\frac{\xi_{k.}(k-1)}{1-\rho_k} \tag{20}$$

Consequently the system is closed and we can then derive the solution at equilibrium.

4.1 Solution at equilibrium

The equilibrium is solution of the system of equations (9) and (20). If (notation) $\alpha_k = k\beta - \frac{k}{k-1}$, this yields the solution for the singlet densities for $k > 1$ (the analytical solution $\rho_1 = 0$ is not considered here as it will be shown that the approximation involved in the calculation is relevant for large k only):

$$\rho_k = 1 - \frac{1}{1+\alpha_k}, \quad \forall k > 1 \tag{21}$$

The solution at equilibrium for the pair densities is:

$$\xi_{k\cdot} = \frac{1}{k\beta}\left(1 - \frac{1}{1+\alpha_k}\right), \quad \forall k > 1 \tag{22}$$

We can see, and this is coherent with intuition, that ρ_k is an increasing function of k and of β. In addition, ρ_k tends to 1 when one of these quantities tends to infinity. The right-hand side of equation (21) is positive if and only if $\beta > \frac{1}{k-1}$. We can also note that pair approximation, as mean field (see [5]), predicts that only the mean degree (through $\beta = \lambda/\bar{d}$) affects the value of ρ_k. This is probably not the only graph characteristic which can affect the process equilibrium.
If the graph is a lattice with coordination number (constant degree) z, the solution (21) becomes

$$\rho_z = 1 - \frac{z-1}{\lambda(z-1)-1} \tag{23}$$

In the case of the n-dimensional cubic lattice $\mathbb{Z}^n$ ($z = 2n$), we recover the critical value for λ under pair approximation: $\lambda_c = \frac{2n}{2n-1}$ (see [2], pp. 180).

4.2 Simulation results

We present here comparisons between the formula derived from pair approximation and simulations of a time continuous contact process. We compare also with the mean field prediction (see [11, 12, 5]), in order to understand in which situations pair approximation should be preferred. We have considered two sorts of graphs for simulations. A random (or Poisson) graph has a degree distribution which can be approximated by a Poisson distribution. The second family of graphs considered is graphs with power law degree distribution. This second example is known to be closer to networks in real world ([4]).
For both types of graphs we observe a better estimation of ρ_k for high degrees, or more precisely for degrees far from the critical value (see Figure 1). This critical value is given by $\beta = \frac{1}{k-1}$, or $k = \frac{\bar{d}}{\lambda} + 1$. We have also observed, from simulations for different values of λ that the difference between simulated an pair approximation values increases when λ decreases and becomes closer to the critical value. One of the reasons for the difficulty to capture the system behavior close to critical values with our pair approximation is in particular that, for small λ, simulations are far from assumption (H) (results not shown here). Working without this assumption is probably one of the paths to explore to improve our pair approximation.
We have also compared the pair and the mean field approximation results on a random graph and a power law graph. We observe that the mean field assumption leads to satisfying approximation of the proportions of occupied sites at equilibrium for a random graph (see Figures 1 (a) to (d), in (b) and (d) the mean field solution is undistinguishable of simulations). For such type of graphs where the variation in the degree is not important between sites and the mean degree is important (20 or 400 in our simulations), we are very close to the assumption of a well-mixed system. It is not surprising that the mean field approximation performs well in such conditions. It can even surpass the pair approximation. Our interpretation of this difference of behavior is that it lies in the approximation of the quantity ξ_{kl}. Even if the mean field is more naive in terms of states dependence, it is more

faithful in terms of degree dependence. In the case of a random graph, this will favor this approximation. But, if we compare the mean field and the pair approximations of the mean proportion of occupied sites at equilibrium ($\bar{\rho} = \sum p_k \rho_k$) with simulated values, we observe that for both methods the prediction is very precise for random graphs (we do not present the corresponding figure since the values are undistinguishable). And, when considering more real graphs, far for the assumption of well-mixed systems, such as the power law graph, we can see that it becomes more important to take into account the correlation between states and we observe the superiority of the pair approximation (see Figures 1 (e) and (f)).
We have finally compared the value of $\bar{\rho}$ at equilibrium for different values of β and for a power law graph and a random graph with same mean degree. On Figure 2 (a), we can see that for values of the parameter far from the phase transition, the two graphs lead to almost the same curve, and the mean field and pair approximation provide quite reliable approximations. Even the mean field approximation ignoring the graph structure, $\rho = 1 - 1/(\beta \bar{d})$, is satisfying (black curve). But if we look at the behaviour close to the phase transition (Figure 2 (b)), it is clear that the critical value of β and the sloap of the curve above this value are different for a random graph and a power law graph. If the mean field provides a rather good approximation for a random graph (coherent with the results mentionned above), neither mean field, nor pair approximation are satisfying to describe the behaviour of the power law graph. None of the approximations lies in the confidence interval obtained from simulations (blue dashed lines). This simple illustration shows the limits of the pair approximation close to the phase transition. This is the region the more difficult to approximate but this is also the more crucial in terms of propagation control.

5 How to improve the pair approximation?

In this section, we relate the pair approximation used along this paper and arising from complexity simplification in ecology to the Bethe approximation ([15]) from statistical mechanics. The latter is also known as Kirkwood approximation ([19, chapter 21]). The two methods we are discussing are different solutions for moment closure at order two when studying the evolution and equilibrium of a system of interacting particles. We present first the similarities and differences between the two approximation methods and explain why Bethe should be more powerful. We develop then the Bethe approximation of the contamination probability $P_{kl}^{00\to 10}$ and compare with the pair approximation (19).

5.1 Pair versus Bethe approximation

The pair approximation has been defined in Section 3.2. The idea for the Bethe approximation is to replace the probability to observe in a given configuration a set of sites linked by edges, by the product of the probabilities on the existing pairs, divided by the product of the probabilities of the overcounted singlets. When considering the subgraph formed by a vertex s of degree k and its neighbours, the existing edges are the k edges from s towards its neighbours, plus maybe edges between the neighbours themselves. Let us consider for instance the following subgraph: a site s and its three neighbours, i, j and k, and an edge

between i and j. If we apply the Bethe approximation to this example we obtain

$$P(z_s, z_i, z_j, z_k) \approx \frac{P(z_i, z_j) \prod_{l=i}^{k} P(z_s, z_l)}{P(z_s)^2 P(z_i) P(z_j)} \tag{24}$$

The first difference with the pair approximation is that Bethe is defined for any n-uplet of sites and not only for triplets. Let us then consider three sites a, b, c, with a linked to b and b to c. If there is no link between c and a, the Bethe approximation is:

$$p(z_a, z_b, z_c) \approx \frac{p(z_a, z_b) p(z_b, z_c)}{p(z_b)} \tag{25}$$

which is equal to the pair approximation (see (18)). But if c is linked to a, the Bethe approximation becomes

$$p(z_a, z_b, z_c) \approx \frac{p(z_a, z_b) p(z_b, z_c) p(z_a, z_c)}{p(z_a) p(z_b) p(z_c)} \tag{26}$$

So the pair and Bethe approximation are equivalent for a graph without triangle (a tree). But for a general graph, the Bethe approximation allows to take into account the presence or absence of triangles. Note that if in (26), the extra term $\frac{p(ac)}{p(a)p(c)}$ due to the presence of the edge between a and c is replaced by its mean field approximation $\frac{p(a)p(c)}{p(a)p(c)} = 1$, we recover (25) or (18) and the pair approximation. The pair approximation can thus be seen as a variant of the Bethe approximation on triplets, were triangle are handled with the mean field approximation.
We thus believe that the Bethe approximation should be a powerful tool, worth to investigate for understanding the particularities of the contact process on a graph. The application of the Bethe principle (26) leads to more complex computations but it should enhance the influence of other characteristics of the graph structure than just the mean degree, in particular the probability that a site and two of its neighbours form a triangle (related to the clustering coefficient of a graph). The main difference with the pair approximation solution will lie in the approximation of the transition probabilities, as developped in the next section.

5.2 Bethe approximation of the pair transition probability

Let us recall the exact value (16) of the pair transition rate $P_{kl}^{00\to 10}$:

$$P_{kl}^{00\to 10} = \beta \sum_{i=1}^{k-1} \frac{P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, r \sim s, s(i) \sim s)}{P(z_s = 0, z_r = 0 \mid d_s = k, d_r = l, r \sim s)} \tag{27}$$

where s is the site in the pair moving from 0 to 1 ($d_s = k$), r is the site in the pair remaining in state 0 ($d_r = l$), and $s(i)$ is one of the $k-1$th other neighbour of s ($d_{s(i)}$ is unknown). For a fixed neighbour $s(i)$ let us consider

$$A = P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, r \sim s, s(i) \sim s)$$

We can decompose first A, by distinguishing the case where the triplet $(s, r, s(i))$ form a triangle (labelled by $\triangle$) and the case where the edge $(r, s(i))$ does not exist (labelled by $\wedge$):

$$\begin{aligned} A &= q_{kl}\, P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, \triangle) \\ &+ (1 - q_{kl}) P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, \wedge) \end{aligned} \tag{28}$$

where $q_{kl} = P(s(i) \sim r \mid s \sim r, s \sim s(i), d_s = k, d_r = l)$ (this probability is assumed independent of the triplet $(s, r, s(i))$). The quantity q_{kl} measures a clustering property of the graph. Since $d_{s(i)}$ is unknown, then we can decompose A over the possible values for $d_{s(i)}$ yielding:

$$\begin{aligned} A &= q_{kl} \sum_d P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, d_{s(i)} = d, \triangle) \times \\ &\quad P(d_{s(i)} = d \mid d_s = k, d_r = l, \triangle) \\ &+ (1 - q_{kl}) \sum_d P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, d_{s(i)} = d, \wedge) \times \\ &\quad P(d_{s(i)} = d \mid d_s = k, d_r = l, \wedge) \\ &= q_{kl} \sum_d B_{kld} \times\ P(d_{s(i)} = d \mid d_s = k, d_r = l, \triangle) \\ &+ (1 - q_{kl})\, C_{kld} \times\ P(d_{s(i)} = d \mid d_s = k, d_r = l, \wedge) \end{aligned} \tag{29}$$

with

$$B_{kld} = P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, d_{s(i)} = d, \triangle) \tag{30}$$

$$C_{kld} = P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, d_{s(i)} = d, \wedge) \tag{31}$$

We have etablished the Bethe approximation of B_{kld} (see Appendix for computational details)

$$B_{kld} = \frac{\xi^{\triangle}_{kd} \,.\, \xi^{\triangle}_{ld} \,.\, (1 - \rho^{\triangle}_k - \xi^{\triangle}_{kl})}{\rho^{\triangle}_d \,.\, (1 - \rho^{\triangle}_l) \,.\, (1 - \rho^{\triangle}_k)} \tag{32}$$

where

$$P(z_i = 1 | d_i = d, \triangle) = \rho^{\triangle}_d, \quad P(z_i = 0, z_j = 1 | d_i = d, d_j = m, \triangle) = \xi^{\triangle}_{dm}$$
$$P(d_i = d, d_j = m | \triangle) = p^{\triangle}_{dm}$$

We define also

$$P(d_s = k, d_r = l, d_{s(i)} = d) = p^{\triangle}_{kld}$$

We obtain with the same approach the Bethe approximation of C_{kld}

$$C_{kld} = \frac{\xi^{\wedge}_{kd} \,.\, (1 - \rho^{\wedge}_k - \xi^{\wedge}_{kl})}{1 - \rho^{\wedge}_k} \tag{33}$$

(the quantities $\rho^{\wedge}_k$ and $\xi^{\wedge}_{kl}$ are defined as the corresponding $\rho^{\triangle}_k$ and $\xi^{\triangle}_{kl}$ except that the conditionning is now equal to $\wedge$.)
Using now (32) and (33), expression (29) of A becomes under Bethe approximation

$$A = q_{kl} \sum_d \frac{\xi^{\triangle}_{kd} \,.\, \xi^{\triangle}_{ld} \,.\, (1 - \rho^{\triangle}_k - \xi^{\triangle}_{kl})}{\rho^{\triangle}_d \,.\, (1 - \rho^{\triangle}_l) \,.\, (1 - \rho^{\triangle}_k)} \times \frac{p^{\triangle}_{kld}}{p^{\triangle}_{kl}} + (1 - q_{kl}) \sum_d \frac{\xi^{\wedge}_{kd} \,.\, (1 - \rho^{\wedge}_k - \xi^{\wedge}_{kl})}{1 - \rho^{\wedge}_k} \times \frac{p^{\wedge}_{kld}}{p^{\wedge}_{kl}}$$

At this stage, in order to get an expression of $P_{kl}^{00\to 10}$ with easy interpretation and comparison with its pair approximation, we will assume that

$$\rho_a^{\wedge} = \rho_a^{\triangle} = \rho_a, \qquad \xi_{ab}^{\wedge} = \xi_{ab}^{\triangle} = \xi_{ab} \tag{34}$$

for any $a, b \in \mathcal{V}$. In this case, if we denote

$$\alpha_{kl} = \sum_d \frac{\xi_{kd} \cdot \xi_{ld}}{\rho_d} \times \frac{p_{kld}^{\triangle}}{p_{kl}^{\triangle}} \quad \text{and} \quad \gamma_{kl} = \sum_d \xi_{kd} \frac{p_{kld}^{\wedge}}{p_{kl}^{\wedge}} \tag{35}$$

Then

$$\begin{aligned} P_{kl}^{00\to 10} &= \beta \sum_{i=1}^{k-1} \frac{A}{1-\rho_k-\xi_k} \\ &= \beta\,(k-1)\frac{1}{1-\rho_k}\left(\frac{q_{kl}}{1-\rho_l}\,\alpha_{kl} + (1-q_{kl})\gamma_{kl}\right) \end{aligned} \tag{36}$$

It can be shown that under the Bethe approximation, $\gamma_{kl} = \xi_{k.}^{\wedge} = \xi_{k.}$, and finally we obtain the approximation

$$P_{kl}^{00\to 10} = \beta\,\frac{k-1}{1-\rho_k}\left(\frac{q_{kl}}{1-\rho_l}\,\alpha_{kl} + (1-q_{kl})\xi_{k.}\right) \tag{37}$$

This approximation must be compared with the pair approximation of $P_{kl}^{00\to 10}$ established in Section (3.2)

$$P_{kl}^{00\to 10} = \beta\frac{(k-1)\xi_{k.}}{1-\rho_k} \tag{38}$$

We recover the relation mentioned previously: the pair approximation is a Bethe approximation assuming $q_{kl} = 0$, i.e. no triangle in the graph. With pair approximation we neglect the effect of the triangles in the process propagation. With the Bethe approximation, they are taken into account through the proportion q_{kl} and the correction term $\beta\,\frac{(k-1)\alpha_{kl}}{(1-\rho_k)(1-\rho_l)}$

6 Conclusion as a summary

The contact process is a powerful tool to represent and understand the basic mechanism of an epidemic's evolution, which is related in this model to the evolution of a single parameter: the transmission or contamination parameter. This model is also useful to study the evolution of a metapopulation. The evolution of such systems (epidemic or metapopulation) results of the combination of two effects: the intrinsic properties of the disesase/population and the properties of the network of relationship between the hosts or patches. The knowledge of global characteristics of the graph can thus help establishing strategies for targered disease control. This has been shown for instance in the case of urban social networks and vaccination, using dynamics bypartite graphs to model and simulate a realistic network ([20]). More generaly, in real cases, the interaction network can be regular (like it could be for orchard) but can also be irregular (e.g. forest trees, social or transporation networks).

We have focused on the second situation which is challenging since the characterization of a graph is quite complex (but well studied, [4]) and it is not yet well understood how these characteristics (degree distribution, clustering property, ...) can affect the evolution of the process living of this graph (critical point value, epidemy size, propagation speed). Most of the works on epidemiology on irregular networks are based on the mean field approximation, assuming no spatial dependencies between the states of each vertex of the graph (e.g. [11]). This assumption can be valid in the case of well-mixed systems (the environment of each individual is homogeneous) and this is not true for a large set of graphs. We investigated one step further here, using two approximations taking into account pair correlation: the pair and the Bethe approximations. The former is simpler to derive but do not take into account the presence of triangles in the graph. It is well known in ecology and we extend it to the case of irregular graphs. The Bethe approximation leads to more complex computations but allows to go deeper in the characterisation of the graph.

Comparisons between simulations of a time continuous contact process and mean field and pair approximations predictions allows us to distinguish in which situations one method should be preferred to the other. In the case of random graphs with sufficient parameter p, we are close to the representation of a well-mixed population: then the mean field approximation is sufficient and leads to satisfying prediction of the state of the system at equilibrium. On the contrary, in the case of graphs with a power law distribution of the degrees, we are far from the well-mixed assumptions: the mean field equations become less powerful when the model parameter decreases towards to the critical value. In that case, the pair approximation provides better predictions.

Concerning the understanding of the importance of the different graph's characteristics in the process contact equilibrium, the analytic solution derived from our pair approximation allows to enhance the effect of the degree distribution. Even if this method can improve the mean field approximation, our results show that the pair approximation can still be far from the exact process equilibrium. Indeed, it seems obvious that two graphs with same degree distribution can present different disease evolutions, in particular if they have different degrees correlations, or clustering coefficient(see [21] in the case of graphs with constant degree, and [11] in the case of power law graphs). Thus, we have extended our study by using the Bethe approximation, taking into account the existence of triangles. This leads to more elaborate developments but still is tractable. Next work will be to obtain more precise numerical estimations from closed forms in particular close to the critical point. (*Peyrard and Franc, in prep.*)

7 Appendix

Independence of ξ_{kl} on l

We establish here that (20) is the only solution of (11). To this order, we consider now the evolution of pairs (10) among the pairs of in $\mathcal{G}_{kl}$. During a period dt, pairs (10) can be created from pairs (00) or (11), and pairs (10) can disappear through creations of pairs of type (00) or (11). This can be expressed as

$$\frac{d\xi_{lk}}{dt} = \rho_l - \xi_{kl} + (1 - \rho_k - \xi_{kl})P_{kl}^{00\to 10} - \xi_{lk}(1 + P_{kl}^{10\to 11}) \tag{39}$$

with $P_{kl}^{10\to 11}dt$ being the probability that a pair (10) in $\mathcal{G}_{kl}$ changes into state (11) during a period dt. We use the same notations than in Section 2.3 and we assume that s is the vertex changing state.

$$\begin{aligned} P_{lk}^{10\to 11} &= \sum_{\tilde{\mathbf{z}}} \beta(a(\tilde{\mathbf{z}}) + 1)P(\tilde{\mathbf{z}}|z_s^t = 0, z_r^t = 1, d_s = k, d_r = l) \\ &= \beta\mathbb{E}_{\tilde{Q}}[a(\tilde{\mathbf{z}})] + \beta \\ &= P_{k.}^{00\to 10} + \beta \end{aligned} \tag{40}$$

Consequently, under (H) and pair approximation, $P_{lk}^{10\to 11}$ is independent of l.
At equilibrium, the derivate (39) is equal to zero so that

$$0 = \rho_l - \xi_{kl} + (1 - \rho_k - \xi_{kl})P_{k.}^{00\to 10} - \xi_{lk}\left(1 + P_{l.}^{00\to 10} + \beta\right) \tag{41}$$

The sum (11) + (41) leads to

$$0 = (1 - \rho_l)P_{l.}^{00\to 10} - \rho_l + \beta\xi_{lk} \tag{42}$$

A consequence of this equality is that at equilibrium ξ_{lk} is independant of k, and is thus equal to its mean $\xi_{l.}$. Thus (11) becomes

$$\xi_{k.} + \xi_{l.} = (1 - \rho_k - \xi_{k.})P_{k.}^{00\to 10} + (1 - \rho_l - \xi_{l.})P_{l.}^{00\to 10} \tag{43}$$

And by setting $k = l$, we recover (20).

Bethe approximation

We establish here the Bethe approximation of expression B_{kld}. The developments are similar for expression C_{kld}. To this order, we define a variable u_a for any $a \in \mathcal{V}$ as $u_a = (z_a, d_a)$.With this notation

$$\begin{aligned} B_{kld} &= P(z_{s(i)} = 1, z_s = 0, z_r = 0 \mid d_s = k, d_r = l, d_{s(i)} = d, \triangle) \\ &= \frac{P(u_{s(i)} = (1, d), u_s = (0, k), u_r = (0, l) \mid \triangle)}{P(d_{s(i)} = d, d_s = k, d_r = l \mid \triangle)} \end{aligned} \tag{44}$$

We apply now the Bethe approximation, as defined in the previous section, on the dependence structure related to variables u_a. This structure corresponds to states correlation, but also to degrees correlation. We obtain the following approximation of the numerator of expression (44)

$$\begin{aligned} & P(u_{s(i)} = (1,d), u_s = (0,k), u_r = (0,l) \mid \triangle) \\ = \; & \frac{P(u_{s(i)} = (1,d), u_s = (0,k)|\triangle) \times P(u_{s(i)} = (1,d), u_r = (0,l)|\triangle)}{P(u_{s(i)} = (1,d)|\triangle) \times P(u_s = (0,k)|\triangle)} \times \\ & \frac{P(u_s = (0,k), u_r = (0,l)|\triangle)}{P(u_r = (0,l)|\triangle)} \end{aligned} \tag{45}$$

Let us introduce some notations for lisibility

$$P(z_i = 1|d_i = d, \triangle) = \rho_d^{\triangle}, \quad P(z_i = 0, z_j = 1|d_i = d, d_j = m, \triangle) = \xi_{dm}^{\triangle}$$

and

$$P(d_i = d, d_j = m|\triangle) = p_{dm}^{\triangle}$$

Then, for instance

$$P(u_s = (0,k), u_r = (0,l)|\triangle) \;\; = \;\; (1 - \rho_k^{\triangle} - \xi_{kl}^{\triangle})\, p_{kl}^{\triangle} \tag{46}$$

Let us note as well

$$P(d_s = k, d_r = l, d_{s(i)} = d) = p_{kld}^{\triangle}$$

Equation (44) can be written as

$$B_{kld} \;\; = \;\; \frac{\xi_{kd}^{\triangle} \cdot p_{kd}^{\triangle} \cdot \xi_{ld}^{\triangle} \cdot p_{ld}^{\triangle} \cdot (1 - \rho_k^{\triangle} - \xi_{kl}^{\triangle}) \cdot p_{kl}^{\triangle}}{\rho_d^{\triangle} \cdot p_d^{\triangle} \cdot (1 - \rho_l^{\triangle}) \cdot p_l^{\triangle} \cdot (1 - \rho_k^{\triangle}) \cdot p_k^{\triangle} \cdot p_{kld}^{\triangle}} \tag{47}$$

In (47), the term $K = \frac{p_{kd}^{\triangle} \cdot p_{ld}^{\triangle} \cdot p_{kl}^{\triangle}}{p_d^{\triangle} \cdot p_l^{\triangle} \cdot p_k^{\triangle} \cdot p_{kld}^{\triangle}}$ multiplied by $p_{kld}^{\triangle}$ is exactly the Bethe approximation of $p_{kld}^{\triangle}$ (this is now the Bethe approximation only on degree dependence). Thus, under Bethe, $K = 1$, so that B_{kld} can be simplifyed:

$$B_{kld} = \frac{\xi_{kd}^{\triangle} \cdot \xi_{ld}^{\triangle} \cdot (1 - \rho_k^{\triangle} - \xi_{kl}^{\triangle})}{\rho_d^{\triangle} \cdot (1 - \rho_l^{\triangle}) \cdot (1 - \rho_k^{\triangle})} \tag{48}$$

References

[1] T. E. Harris. Contact interactions on a lattice. *Annals of probability*, 2:969–988, 1974.

[2] J. Marro and R. Dickman. *Nonequilibrium Phase Transition in Lattice Models*. Monographs and texts in statistical physics, Collection Alea Saclay. Cambridge University Press, Cambridge, UK, 1999.

[3] B. Drossel and F. Schwabl. Self organizing critical forest fire models. *Phys. Rev. Let.*, 69:1629–1632, 1992.

[4] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[5] A. Franc. Metapopulation dynamics as a contact process on a graph. *Ecological Complexity*, 1:49–63, 2004.

[6] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graphs models of social networks. *PNAS*, 99:2566–2572, 2002.

[7] B. Lung-Escarmant and Guyon D. Temporal and spatial dynamics of primary and secondary infection by Armillaria ostoyae in a Pinus pinaster plantation. *Ecology and Epidemiology*, 94(2):125–131, 2003.

[8] W. Otten, D. Bailey, and A. Gilligan. Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. *New Phytologist*, (163):125–132, 2004.

[9] I. Longini. A mathematical model for predicting the geographic spread of new infectious agents. *Math Biosci*, 90:367–383, 1988.

[10] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.

[11] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.

[12] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev.*, E63:066117, 2001.

[13] D.A. Lavis and G.M. Bell. *Statistical mechanics of lattice systems: I closed forms and exact solutions*. Springler, 1999.

[14] R. Kikuchi. A theory of cooperative phenomena. *Physical Review*, 81(6):988–1003, 1951.

[15] J. Yedidia, W. Freeman, and Y. Weiss. Bethe free energy, Kikuchi approximations, and belief propagation algorithms, 2000. MERL T.R. ; url: citeseer.ist.psu.edu/yedidia00bethe.html.

[16] J. S. Yedidia. *An Indiosyncratic Journey Beyond Mean Field Theory*, chapter 3 in *Advanced Mean Field Methods - Theory and Practice*, Opper, M. and Saad, D. (Ed.), pages 21–35. The MIT Press, Cambridge, Mas., USA, 2001.

[17] J. A. N. Filipe and G. J. Gibson. Comparing approximations to spatio-temporal models for epidemics with local spread. *Bulletin of Mathematical Biology*, 63:603–624, 2001.

[18] Joo J. and J. L. Lebowitz. Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercubic lattice. *arXiv:q-bio.PE/0404035 v1*, page 26 Apr 2004, 2004.

[19] U. Dieckmann, R. Law, and J. A. J. Metz, editors. *The geometry of ecological interactions - Simplifying spatial complexity*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, UK, 2000.

[20] S. Eubank, H. Guclu, V. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 149, 2004.

[21] M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. London B*, 266:859–867, 1999.

Figures

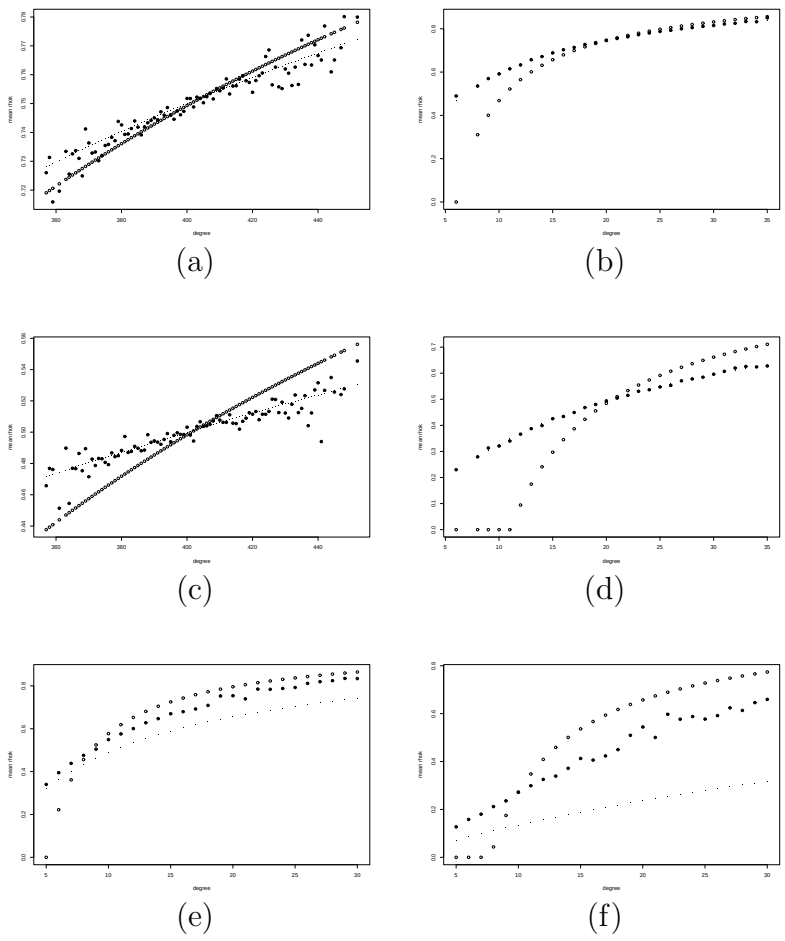
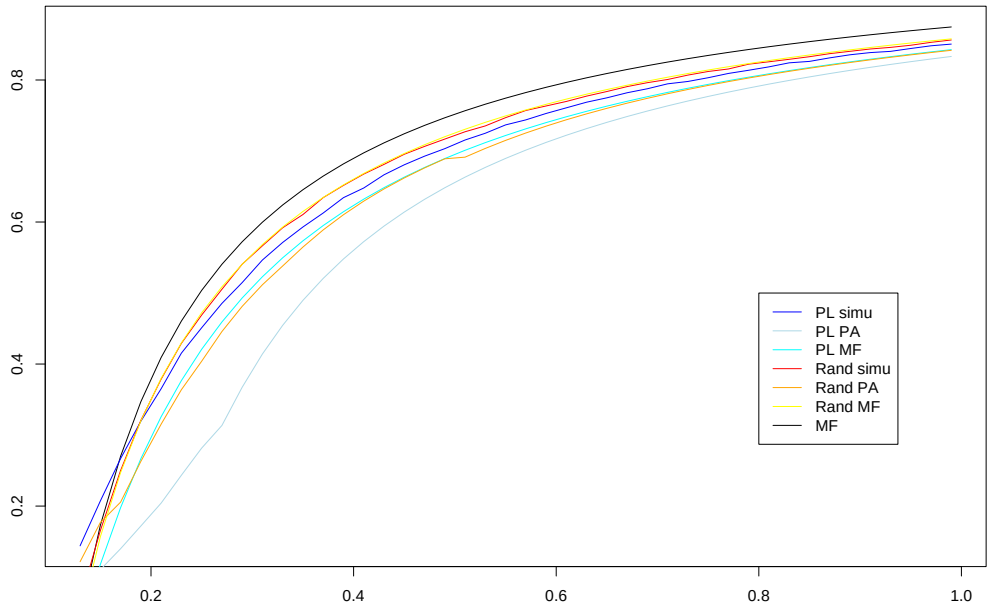
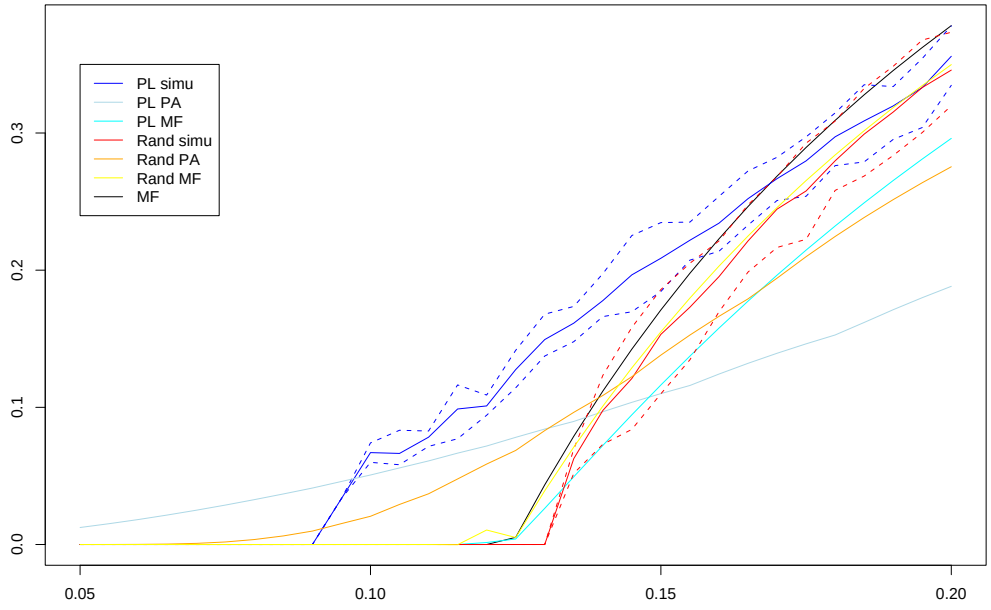


Figure 1: Comparison of simulated values (•), pair approximation (∘) and mean field approximation (.) of ρ_k at equilibrium (1000 vertices): (a) random graph, edge creation probability $p = 0.4$ and $\lambda = 4$, (b) random graph, $p = 0.02$ and $\lambda = 4$, (c) random graph, $p = 0.4$ and $\lambda = 2$, (d) random graph, $p = 0.02$ and $\lambda = 2$, (e) power law graph, power $= 3$, $\lambda = 2$, (f) power law graph, power $= 3$, $\lambda = 1.2$



(a)



(b)

Figure 2: Comparison of the occupancy rate $\bar{\rho}$ at equilibrium versus β for a power law graph and a random graph with same mean degree, and corresponding mean field and pair approximations (a) for β ranging from 0.13 to 1, (b) zooming of $\beta \in [0.05, 0.2]$ (dashed lines show the confidence intervals).