LONG RANGE CORRELATIONS IMPROVE UNDERSTANDING OF THE INFLUENCE OF NETWORK STRUCTURE ON EPIDEMIC DYNAMICS

BY

Nathalie Peyrard
Ulf Dieckmann
Alain Franc

Research Report No. 18
April 2007

Unité de Biométrie
Institut National de la Recherche Agronomique
Avignon, France
http://www.avignon.inra.fr/biometrie

Long range correlations improve understanding of the influence of network structure on epidemic dynamics

N. Peyrard[a], U. Dieckmann[b], A. Franc[c]

[a] Unité de Biométrie et Intelligence Artificielle, UR 875 , INRA, Chemin de Borde-Rouge, BP 52627, 31326 Castanet-Tolosan cedex, France (nathalie.peyrard@toulouse.inra.fr)

[b] Evolution and Ecology Program, International Institute for Applied Systems Analysis, A-2361, Laxenburg, Austria (dieckmann@iiasa.ac.at)

[c] INRA, UMR 1202 Biodiversité, Gènes et Communautés, 69 route d'Arcachon, Pierroton, 33612 Cestas cedex, France (alain.franc@pierroton.inra.fr)

Abstract

Models of infectious disease that do not confer permanent immunity are characterized by a phase transition: rapid extinction or persistence. A current challenge is to understand how the geometry of the interaction network associated with the disease can influence prevalence close to the critical point. Here, we bring new insights on how to introduce explicitly the network structure in the model to study this influence. To cope with the model complexity, we consider moment closures. Moment closures do not provide satisfying

estimation close to the critical value. Therefore, we propose a method to introduce longer range correlations in the closures. This method is technically simple, remains computationally reasonable and significantly improves approximation performance. Extended closures provide a tool to quantify the influence of the clustering coefficient and of new descriptors of a graph organization, the square clustering coefficient. We also compare the relative performance of different forms of closure from the literature, with or without extension. The normalized version of the Bethe approximation, extended within our frameworks, appears as a good candidate to study the influence of the graph features. Numerical results illustrate the role of the clustering and square clustering coefficients for low and median values of the transmission rate of the disease and the importance of path redundancy on prevalence.

Keywords: Contact process; Interaction network structure; Long range correlation; Moment closure; Phase transition.

1 Introduction

Infectious diseases where permanent immunity is not possible are classically represented as a SIS system which can be implemented on a lattice, or more generally on a graph. This model is characterized by the existence of two possible dynamics: either a rapid extinction of the disease or, at the opposite, its persistence over a long time (metastability on a finite lattice). This change of behaviour can be defined as a phase transition (rigorously, a phase transition occurs only on infinite size lattices or graphs), linked with the disease properties (e.g. transmission rate). The value of the parameter for which such a change occurs is a critical value. Critical values have received considerable attention in the study of systems in statistical physics, because (i) they enable to define ranges of parameters values where the qualitative behaviour of the system is unchanged, as sets bound by critical values (ii) classical approximations such as mean field approximation are known to yield tractable and reasonable estimates for variables of interest far from critical values, whereas approximation quality may break down close to them. For infectious diseases, studying the critical value is relevant for understanding both outbreak and prevalence close to transition. A better knowledge of the factors influencing theses values is important in order to improve control strategies to prevent disease outbreak.

Infectious diseases are usually characterized by propagation by contact between nearest neighbours. The geometric pattern of the connections between hosts, a contact enabling transmission, can efficiently be described by a graph. The associated interaction network plays an important role in the disease dynamics (Keeling 1999, Eguiluz and Klemm 2002, Keeling 2005,

Colizza et al. 2006). In particular, its structure can influence the value of the critical point of the phase transition. However, while the existence of this influence is well known, it is still not fully understood which are the important characteristics of the network controlling it, and how they affect the disease, favoring or limiting the propagation. For a long time, classical graphs have been restricted to random graphs (mean field model when complete, or island model) or connections to nearest neighbours on a lattice (stepping stone model), but more realistic families of graphs have been proposed recently in literature, each characterized by common values of degree distribution, clustering coefficient, etc... (Albert and Barabási 2002 or Newman 2003). In this paper, we propose geometric characteristics of the graphs which can shed new light on understanding how the structure of a graph may influence its function, here disease outbreak.

Individual based spatially explicit stochastic systems are called "Interacting Particle Systems" (Ligget 1985 or Durret and Levin 1994). IPS are well adapted to capture the propagation of infectious disease (e.g. Keeling 1999). They are largely used in ecological dynamics when individual contacts are through nearest neighbours too (Van Baalen and Rand 1998). However, very few exact results are computable on these models, which can be studied mostly through simulations. IPS however can be simplified by ODE on state variables, such as mean density of susceptible vertices. Critical values are often approximated by bifurcation points of these ODE. However, examples of failure of such mean-field models close to the critical values, assuming independence of the individuals, are numerous (Durret and Levin 1994, Dieckmann et al. 2000, Part B, Filipe and Gibson 1998, Keeling 2005). Therefore, considerable attention has been paid in literature to design ap-

proximations which behave better than mean field approximations, closer to the critical values of the parameters. Such methods are often referred to as moment closure methods (Bolker and Pacala 1997), such as *Pair approximation* (Matsuda et al. 1992) or *cluster variation methods* (Kikuchi 1951 or Lavis and Bell 1999) like *Bethe approximation*, etc. ... These methods are well-known and largely used in theoretical epidemiology or ecology (e.g. Dieckmann et al. 2000). However, all these classical methods take into account the graph structure in a limited way only and cannot be used as such to study the interaction network influence.

Analytically tractable models such as the Ising model teach us that approximations breakdown because long range correlations, ignored in the simplifications, do not fade away close to the critical point (Marro and Dickman 1999, Snyder and Nisbet 2000, Dickman and Martins de Oliveira 2005). Hence, one way to improve upon classical approximations is to take into account, when possible, long range correlations, in addition to correlations between nearest neighbours, as in classical Pair and Bethe approximations on graphs. Here, we start from several published approximations relying on correlations between nearest neighbours only, and propose for each of them new types of approximation by taking into account correlations at distances larger than one. We show that, in some cases, this leads to better estimation of the critical value of the parameter, as well as of the prevalence of the disease. In addition, these new approximations enable to explicitly take into account some features of the graph and thus to study their influence on the prevalence. Let us note that this is one of several ways to improve classical approximations: Petermann and De Los Rios (2004a) have suggested improvements through careful examinations of local patterns of size larger than

pairs, but with correlation between nearest neighbours only. If the method improves the description of the dynamics for a given interaction network, comparison of the same dynamics for varying values of some features of the graph is not straightforward.

As a toy model, we consider the well-studied contact process (Harris 1974, Marro and Dickman 1999), spatially explicit version of the SIS model, on a graph. It has been used for modelling phenomena involving excitable media, like fire models (Drossel and Schwabl 1992), or for modelling metapopulations (Franc 2004). It has also been widely used to study the spread of diseases within a population (Filipe and Gibson 1998, Pastor-Satorras and Vespignani 2001, Eguiluz and Klemm 2002).

This paper is organized as follows: in Section 2, we define the contact process and derive the open ODE system up to pairs dynamics. After introducing two classical closure methods, mean field and Bethe approximations in Section 3, we illustrate on Bethe the methodological contribution of this work: how to extend approximations based on nearest neighbours pairs to longer range correlations (Section 4). From this, we propose new graph characteristics with potential influence of the contact process dynamics. Section 5 presents other classical closures on which we apply the same extension. Comparison of the different approximations, with and without long range correlations, as well as the influence of graph features on prevalence are illustrated on two homogeneous graphs (Section 6). We conclude with a discussion including possible improvements on the method presented herein.

2 The Contact Process

2.1 Definition

Let us consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, defined by a set $\mathcal{V}$ of n vertices and a set $\mathcal{E}$ of edges. Vertices represent individuals (plants, human beings, ...) and edges represent possible paths for the disease propagation. On each vertex i of the graph a random variable z_i (the state of the vertex) is expressed, taking values in a discrete space $\Omega = \{0, 1\}$. The set Ω could be interpreted as {succeptible, infected}, as in epidemiology, or {empty, occupied}, as in metapopulation dynamics. We adopt here the first point of view. The states of the graph's vertices evolve according to epidemiological dynamics (infection and recovering).

We consider here a time-continuous contact process to model dynamics on the graph. The rules for evolution are as follows (Harris 1974, Marro and Dickman 1999, chapter 6): if $i \in \mathcal{V}$ is infected ($z_i = 1$) it becomes susceptible ($z_i = 0$) with rate μ. As time units are arbitrary, we set $\mu = 1$ in the following, meaning that the expected time between a site infection and its recovery is one unit of time. For the transmission probability, if i is susceptible and has a_i infected neighbours, it becomes infected with rate βa_i, with β being the per site transmission rate:

$$\begin{aligned} P(z_i^{t+dt} = 0 | z_i^t = 1) &= dt \\ P(z_i^{t+dt} = 1 | z_i^t = 0, a_i = a) &= a\beta dt \end{aligned} \tag{1}$$

The rate β is in fact dependent of the time unit. Indeed, if π is the probability that a transmission event occurs during period $[t, t + \triangle t]$,with $a_i = a$, then $1 - \pi = \exp -a\beta \triangle t$ and $\beta = -(\ln(1 - \pi))/(a \triangle t)$. Thus, β and $\triangle t$ must be chosen such that $\pi \leq 1$.

This model exhibits a non-equilibrium phase transition (on an infinite graph): below a value β_c (the critical value) of the transmission rate β, no establishment is possible and the only final global state is the one with all vertices susceptible. Establishment is possible above β_c.

2.2 Open system for dynamics equations

We will consider in this study homogeneous graphs, i.e. each vertex has the same degree h, for sake of simplicity. The first quantity of interest to study dynamics and equilibrium is ρ, i.e. the probability that a vertex is in state 1. As contact process transmission can only be done at short distance, between nearest neighbours, the evolution of ρ depends on the evolution of pairs of vertices linked by an edge, It is described by Filipe and Gibson (1998) or Peyrard and Franc (2005)

$$\frac{d\rho}{dt} = h\beta P^{(1)}(01) - \rho \tag{2}$$

with $P^{(1)}(01)$ being the probability that a pair of vertices linked by an edge is in state 01. Here and in the following, the exponent (1) recalls that the two sites of interest are linked by an edge (distance 1 pair).
Thus, we consider the evolution of $P^{(1)}(01)$. The probabilities for the other configurations of a pair at distance 1 can be derived from ρ and $P^{(1)}(01)$. The exact expression for the transients for $P^{(1)}(01)$ is given by

$$\begin{aligned} \frac{dP^{(1)}(01)}{dt} &= P^{(1)}(11) + \beta(h-1)P^{(1,1)}(100) \\ &\quad -P^{(1)}(01)\left[1+\beta+\beta(h-1)\frac{P^{(1,1)}(101)}{P^{(1)}(01)}\right] \end{aligned} \tag{3}$$

where $P^{(1,1)}(100) = P(z_i = 1, z_j = 0, z_k = 0 \mid (i,j),(j,k) \in \mathcal{E})$ is the probability to observe a vertex j in state 0 and two of its neighbours in states

0 (vertex k) and 1 (vertex i). It may seem complicated at first sight, but is nothing more than a straightforward consequence of careful counting of infection and recovery events starting from, or leading to, a pair 01 (see Appendix). Note that the three vertices do not play the same role here and the ordering of the states in $P^{(1,1)}(100)$ is important. The triplet of sites can either be closed (triangle) or open.

The system formed by equations (2) and (3) is not closed since the evolution of the triplet probabilities is unknown. This requires the knowledge of the evolutions of quadruplets and so on. Exact solving of the contact process (transient and equilibrium) is not possible, because of this hierarchy of dynamics. We present in the following section a classical method to simplify the model complexity leading to a system with a reduced number of equations.

3 Closure methods

To close a system like (2) and (3), the idea is to truncate the hierarchy at a certain level (order) and to approximate all the higher joint probabilities as functions of the lower ones. This technique is referred to as a closure method and finds its origin in the cluster variation method, developed in the 1930's in solid state physics (see Kikuchi 1951, Lavis and Bell 1999). The intuitive idea behind cluster variational method is to approximate the free energy related to a complex distribution by dealing exactly with subsets of nodes of reasonable size (the clusters) and to approximate correlations between nodes from different clusters (see Yedidia et al. 2005) for details on this and others region-based approximations).

The simplest closure is the closure at order one, namely the mean field approximation. It consists of assuming well-mixing among the vertices, i.e. each individual can have contact with all the others. This is equivalent to assume that individuals are independent ($P^{(1)}(01) = \rho(1-\rho)$). The corresponding approximated system reduces to a closed equation in ρ (Marro and Dickman 1999), if we set $h\beta = \lambda$:

$$\frac{d\rho}{dt} = \lambda\rho(1-\rho) - \rho = r\rho\left(1 - \frac{\rho}{K}\right) \tag{4}$$

which is the classical logistic equation with the rate of maximum population growth, r, equal to λ and the carrying capacity, K, equal to $1 + \lambda^{-1}$. If probabilities are replaced by proportions, one can recover the SIS model often used in non-spatial epidemiology. This equation is simple to solve and study, but except in particular cases (graphs with a high value of h, graphs in high dimensions, or high prevalence) the mean field assumption of well-mixing is too strong and the approximation is quantitatively poor (Durret and Levin 1994, Filipe and Gibson 1998, Dieckmann et al. 2000, Keeling 2005).

We are interested here in order 2 approximations, meaning that pair correlations are taken into account and that triplet probabilities are approximated in terms of singletons and pairs probabilities. Order 2 closures offer a trade-off between precision and complexity. First, let us recall that one of the quantities to approximate is $P^{(1,1)}(100)$. The approximation of the triplet probability should take into account that some triplets of sites are closed (triangle) and some are not. If θ is the clustering coefficient of the graph, i.e. the probability that two neigbours of a same site are neighbours to each others, then $P^{(1,1)}(100)$ can be decomposed as

$$P^{(1,1)}(100) = \theta P^{1,1,1}(100) + (1-\theta)P^{1,1,2}(100) \tag{5}$$

$P^{1,1,1}$ and $P^{1,1,2}$ being respective probabilities for closed ($\triangle$) and open ($\wedge$) triplet configurations. The three indices give distances between sites. Parameter θ is one of the characteristics of the graph well known in graph theory (Albert and Barabási 2002, Newman 2003).

Several forms have been proposed for the closure (see section 5 and Discussion). We present only one in this section, the *Bethe* approximation, as an example to illustrate the method that we propose to improve classical order 2 closures. In the case of a triangle, the Bethe approximation (BA) (Morita 1994, Yedidia et al. 2000) or *Kirkwood approximation* (Singer 2004) is as follows:

$$\widehat{P}_{BA}^{1,1,1}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)P^{(1)}(z_i, z_k)}{P(z_i)P(z_j)P(z_k)} \tag{6}$$

This definition can be interpreted as the product of the probabilities for the three distance 1 pairs of the triangle, divided by the overcounted singleton probabilities (since a vertex is present in two pairs).
In the case of an open triplet, the Bethe approximation is the Pair approximation, and reads

$$\widehat{P}_{BA}^{1,1,2}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)}{P(z_j)}$$

As Bethe, all classical closures at order 2 for models on graph are distance 1 closures at order 2: pairs of nodes separated by more than one edge are assumed to be independent. (As we deal with graphs, reference to a distance will always be reference to the geodesic distance). However, if middle range correlations can be neglected far from the critical point, this same choice leads to inaccurate approximations closer to the critical value β_c. As a step for improvement of classical order 2 closures, we present in the next section a method to take into account the correlation between two nodes at distance

larger than 1. The method will be implemented in the approximation of the triplet probabilities $P^{(1,1)}(100)$ and $P^{(1,1)}(101)$, and to close the system formed by (2) and (3).

4 Extension to longer range correlation

The method presented below can be applied to extend order 2 closures by taking into account correlations for any given distance. However, as for the choice of the order of the closure, there is a trade-off between complexity and precision. We detail here a method for adding distance 2 and distance 3 correlations (illustrated only on the Bethe closure). For longer distances, the method is still theoretically valid, but may not be tractable in practice.

4.1 Adding distance 2 correlations

A generalisation of the closure is proposed in order to take into account correlations at distance 2. If i, j, k are three vertices of the graph with associated distances d_{ij}, d_{jk} and d_{ik}, then the extended definition of the Bethe closure is

$$\widehat{P}^{(d_{ij},d_{jk},d_{ik})}(z_i, z_j, z_k) = \frac{P^{(d_{ij})}(z_i, z_j)P^{(d_{jk})}(z_j, z_k)P^{(d_{ik})}(z_i, z_k)}{P(z_i)P(z_j)P(z_k)} \tag{7}$$

In the above expression, we approximate a pair probability, for instance $P^{(d_{ik})}(z_i, z_k)$, with $P(z_i)P(z_k)$ if $d_{ik} > 2$ (instead of $d_{ik} > 1$ in the classical closures).

Let us start from equation (5). The approximation $\widehat{P}^{1,1,1}(z_i, z_j, z_k)$ for a closed triangle $\triangle$ is not modified with the extended closure because it involves only pairs at distance 1. Thus, we still use (6). For $\widehat{P}^{1,1,2}(z_i, z_j, z_k)$, i.e. for an open triplet $\wedge$, even if there is no direct path (an edge) between

sites i and k, many other paths of different lengths exist. One of them is the path through the middle point j of the triplet, which is a path of length 2, the shortest length. Taking into account this information (which is only part of the existing long-range correlations), the distance 2 Bethe closure of $P^{1,1,2}(z_i, z_j, z_k)$ is

$$\widehat{P}^{(1,1,2)}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)P^{(2)}(z_i, z_k)}{P(z_i)P(z_j)P(z_k)} \tag{8}$$

The correction term, when comparing with the classical pair approximation, is $\frac{P^{(2)}(z_i,z_k)}{P(z_i)P(z_k)}$, which arises from the fact that the variables z_i and z_k are not independent.

4.2 New approximated dynamics equations

We can now exploit the distance 2 order 2 closure to approach the system formed by equations (2) and (3). We recall that

$$\begin{cases} \dfrac{d\rho}{dt} = h\beta P^{(1)}(01) - \rho \\ \\ \dfrac{dP^{(1)}(01)}{dt} = P^{(1)}(11) + \beta(h-1)P^{(1,1)}(100) \\ \qquad\qquad - P^{(1)}(01)\left[1 + \beta + \beta(h-1)\dfrac{P^{(1,1)}(101)}{P^{(1)}(01)}\right] \end{cases}$$

The closure will be reached through an approximation $\widehat{P}^{(1,1)}(100)$ (and $\widehat{P}^{(1,1)}(101)$) which involves not only $P^{(1)}(01)$ but also $P^{(2)}(01)$, since from (5), (6) and

(8)

$$\begin{aligned}\widehat{P}^{(1,1)}(100) &= \theta\widehat{P}^{1,1,1}(100) + (1-\theta)\widehat{P}^{1,1,2}(100) \\ &= \theta\,\frac{P^{(1)}(01)^2P^{(1)}(00)}{\rho(1-\rho)^2} + (1-\theta)\,\frac{P^{(1)}(01)P^{(2)}(01)P^{(1)}(00)}{\rho(1-\rho)^2} \\ &= \frac{P^{(1)}(01)P^{(1)}(00)}{\rho(1-\rho)^2}\;\left(\theta\,P^{(1)}(01) + (1-\theta)\,P^{(2)}(01)\right)\end{aligned} \tag{9}$$

Then, an additional equation is required for the temporal evolution of $P^{(2)}(01)$ Let us define $t^{(2)}_{ab\to cd}$, the transition rate from states (a,b) to states (c,d) when the shortest path between the two sites of interest is of length 2. Counting the events starting from, and leading to, the state (01), we have:

$$\begin{aligned}\frac{dP^{(2)}(01)}{dt} &= -P^{(2)}(01)(t^{(2)}_{01\to 11} + t^{(2)}_{01\to 00}) \\ &\quad + P^{(2)}(00)t^{(2)}_{00\to 01} + P^{(2)}(11)t^{(2)}_{11\to 01}\end{aligned} \tag{10}$$

with the transition rates at distance 2 approximated by

$$\begin{aligned}\hat{t}^{(2)}_{01\to 11} &= \beta(h-1)\frac{\widehat{P}^{(1,2)}(101)}{P^{(2)}(01)} + \beta\frac{\widehat{P}^{(1,1,2)}(011)}{P^{(2)}(01)} \\ \hat{t}^{(2)}_{01\to 00} &= 1 \\ \hat{t}^{(2)}_{00\to 01} &= \beta(h-1)\frac{\widehat{P}^{(1,2)}(100)}{P^{(2)}(00)} + \beta\frac{\widehat{P}^{(1,1,2)}(010)}{P^{(2)}(00)} \\ \hat{t}^{(2)}_{11\to 01} &= 1\end{aligned} \tag{11}$$

and

$$\begin{aligned}\widehat{P}^{(1,2)}(z_i,z_j,z_k) &= \alpha_1\,\frac{P^{(1)}(z_i,z_j)P^{(2)}(z_j,z_k)P^{(1)}(z_i,z_k)}{P(z_i)P(z_j)P(z_k)} \\ &\quad + \alpha_2\,\frac{P^{(1)}(z_i,z_j)P^{(2)}(z_j,z_k)P^{(2)}(z_i,z_k)}{P(z_i)P(z_j)P(z_k)} \\ &\quad + \alpha_3\,\frac{P^{(1)}(z_i,z_j)P^{(2)}(z_j,z_k)}{P(z_j)}\end{aligned} \tag{12}$$

(see the appendix for details on derivation of the transition rates.)
In $\widehat{P}^{(1,2)}$ only two distances are known. There are three possibilities for the value of the third distance: equal 1, equal 2 or higher than 2. The coefficients α_1 to α_3 represent the proportions of the three corresponding patterns of triplets of sites respectively (see Figure 1 (a) and (b)). The extended closure therefore enables us to point out new characteristics of the graph which can affect the evolution of the process. We refer to these weights as the *square clustering coefficients*. Their influence on the evolution of the process will be easy to explore in the framework of distance 2 order 2 closures.
Finally, the system formed by (2) and (3) and (10), with approximations (9), (11) and (12) is now closed since pairs probabilities for states 00 and 11 can be recovered from ρ and pair probabilities for state 01.

4.3 Adding distance 3 correlations

The same method could be repeated in turn to refine the approximation of the correlations when two sites are separated by more than 2 edges, more than 3, ... and so on. We have implemented here one step further only, dealing with pairs of vertices at distance 3 without approximations. This modifies only the term with weight α_3 in (12) because this corresponds to the only situation where a distance higher than 2 exists between the three sites involved. Applying the same logic as for distance 2 correlations, we derive the extension by applying the following steps:
(1) we add a fourth equation in the system, for the temporal evolution of $P^{(3)}(01)$.
(2) we express transition rates $t^{(3)}_{01\to 11}$ and $t^{(3)}_{00\to 01}$ in terms of probabilities of the form $P^{(1,3)}(z_i, z_j, z_k)$: the probability to find three sites (i, j, k) in configuration (z_i, z_j, z_k), given that $(i, j) \in \mathcal{E}$, that the shortest path between j

and k is of length 3 and that d_{ik} is unknown (i is not on the path between j and k).
(3) we decompose $P^{(1,3)}(z_i, z_j, z_k)$ depending on the third distance in the triplet. There are three possibilities for the shortest path between i and k: 2 or 3 edges (see Figure 1 (c) and (d)) or 4 edges. Only one edge is not possible since it would mean that there is a path of length 2 between j and k while the distance between these two vertices is 3. We compute the weights $(\gamma_2, \gamma_3, \gamma_4)$ of the three patterns of triplets for the studied graph.
(4) we apply formula (7) for $d_{ik} = 2, 3$ or 4. In the case of a distance 4, we replace $P^{(4)}(z_i, z_k)$ by the product of the singleton probabilities.

Note that going one step further in the approximation enables us to introduce three new graph parameters $(\gamma_2, \gamma_3, \gamma_4)$.

5 Comparison with closures from the literature

To study the potential gain associated with the proposed extension, comparison with different classical order 2 closures from literature will be implemented.
The first closure presented here is Pair Approximation (PA). It is now standard to simplify complexity in spatio-temporal models in epidemiology and ecology when space is discrete (see Dieckmann et al. 2000, chapters 13,18,19). It reads as follows, for any set of three sites all linked by an edge:

$$\widehat{P}_{PA}^{1,1,1}(z_i, z_j, z_k) = \widehat{P}_{PA}^{1,1,2}(z_i, z_j, z_k) = \frac{P^{(1)}(z_i, z_j)P^{(1)}(z_j, z_k)}{P(z_j)} \qquad (13)$$

Note that the approximation is the same whether the triplet is closed or not. It does not account for the effect of the clustering coefficient of the graph, and distance 2 or distance 3 correlation cannot be introduced. Several improvement and variations around PA have been proposed (see Discussion). We do not introduce them here, because, as it is the case for PA, they cannot be extended to correlations at distances larger than 1.

As we have shown, the Bethe approximation (BA) enables such extension. However, the Bethe approximation of triplet probabilities is not normalized. So we will also consider here its normalized version (NBA):

$$\widehat{P}^{1,1,1}_{NBA}(z_i, z_j, z_k) = \frac{\widehat{P}^{1,1,1}_{BA}(z_i, z_j, z_k)}{\sum_{z_i,z_j,z_k} \widehat{P}^{1,1,1}_{BA}(z_i, z_j, z_k)} \tag{14}$$

Moment closure is also widely used in continuous space in population dynamics. The PA and BA have also been applied in this context. Recently more sophisticated order 2 closures have been proposed (Dieckmann et al. 2000, chapter 21, Murrel et al. 2004). They are referred to as respectively *Power 1* ($Pw1$) and *Power 2* ($Pw2$) closure according to the number of pairs multiplied together in an elementary building block of the closure. The Power 1 closure is defined as

$$\begin{aligned}\widehat{P}^{1,1,1}_{Pw1}(z_i, z_j, z_k) = & P(z_i)P^{(1)}(z_j, z_k) + P(z_j)P^{(1)}(z_i, z_k) + P(z_k)P^{(1)}(z_i, z_j) \\ & - 2P(z_i)P(z_j)P(z_k)\end{aligned} \tag{15}$$

Here, the building block is one pair probability among the three in the triangle, multiplied by the probability of the vertex opposite to the pair. For $Pw2$, the building block is the product of two pairs probabilities divided by the probability of the vertex common to the two pairs. This ensures, as for the other closures presented above, that the closure has the correct dimension.

Note that one of the building block is exactly PA,

$$\widehat{P}^{1,1,1}_{Pw2}(z_i,z_j,z_k) = \frac{1}{2}\left[\frac{P^{(1)}(z_i,z_j)P^{(1)}(z_j,z_k)}{P(z_j)} + \frac{P^{(1)}(z_i,z_j)P^{(1)}(z_i,z_k)}{P(z_i)} + \frac{P^{(1)}(z_i,z_k)P^{(1)}(z_j,z_k)}{P(z_k)} - P(z_i)P(z_j)P(z_k)\right] \tag{16}$$

Note also that a drawback of $Pw1$ and $Pw2$ is that they may lead to negative values, with low values of β (and thus of ρ), because of the strong correlation between vertices' states.

For each of these four closures, we have derived the extended versions, taking into account distance 2 and distance 3 correlations, following the same methodology as in Section 4.

6 Results

We compared the performance of each closure that we selected from literature (see section 3 and 5: Mean Field (MF), Pair approximation (PA), Bethe (BA), normalized Bethe (NBA), Power 1 ($Pw1$) and Power 2 ($Pw2$)) without and, when relevant, with the above extension to correlations at distance 2 (section 4.1) and 3 (section 4.3), regarding the transients and value at equilibrium of ρ and estimation of the critical value β_c. Closures were compared with simulations of the contact process.

6.1 Method

We used the Runge Kuta method from package `odesolve` of `R` to solve the ODE systems. Estimations of ρ at equilibrium from simulations of the continuous time contact process were obtained by averaging over 100 different

realisations, each of them run during 40000 iterations. Simulations of transients were obtained by performing realisations of 40000 iterations. One iteration is defined as the modification of the state of a single node in the graph (according to rates derived from equation (1)). The time scale between two iterations is incremented by the corresponding exponential random variable. The size of the graphs are about 1000 vertices.

Two different graphs with homogeneous degree are considered, with respectively $h = 6$, $h = 4$ edges by vertex (see Figure 2). The first one is often referred to as the triangular grid. These examples of homogeneous graphs have the same clustering coefficient, but different coefficients regarding distance 2 and distance 3 correlations. All the graph's statistics necessary to implement the closures approximations are easily obtain by combinatorial computation. Their values are reported in Table 1. We ran simulations on the triangular grid and the degree 4 graph with respectively 900 and 980 vertices. As the results are qualitatively similar, we illustrate them here only on the graph of degree 4.

6.2 Comparisons between different forms of closures

Estimation of the critical value β_c of the parameter β and of the density ρ (prevalence) at equilibrium close to this value is more challenging than estimation of densities far from the critical value. Therefore, the numerical simulations and comparisons with the simplified systems have been implemented close to the critical value of β, known from simulations.

h	θ	α_1	α_2	α_2	γ_2	γ_3	γ_4
6	2/5	2/15	6/15	7/15	2/15	6/15	7/15
4	2/5	4/45	4/15	29/45	2/45	12/45	31/45

Table 1: Statistics of the two homogeneous graphs studied: h is the degree of the graph, θ is the clustering coefficient, α_1 to α_3 are the square clustering coefficients for closures with correlations at range higher than 1, γ_2 to γ_4 are coefficients for closures with correlations at range higher than 2.

Estimation of β_c

Table 2 gives the estimates of β_c from simulation and for each closure, on the graph of degree 4. Note that an analytical expression of β_c is available for MF and for PA, which is equal respectively to $1/h$ and $1/(h+1)$. The results for the graph of degree 6 are similar (i.e. similar ordering of the quality of the approximations) and, as the degree is higher, the approximations behave slightly better. This shows that (i) estimations without the extension to correlation at distance 2 or larger are poor whatever the method, with the exception of $Pw1$, with the ordering $MF < Pw2 < PA < BA \leq NBA$, where $M1 < M2$ means that method $M1$ is less accurate than method $M2$ (ii) the estimation with method $Pw1$ without extension is accurate (iii) taking into account the correlations at distance 2 or 3 improves the estimation by NBA, whereas there is an improvement with distance 2 for BA, and a breakdown for distance 3 (iv) these extensions destabilise the numerical scheme for $Pw1$, probably because negative values for the probabilities are produced. Figure 3 illustrates points (ii) and (iii) around the value of β_c es-

timated from simulations. In the following, NBA with extension to distance 3 correlations, as well as $Pw1$ closure without extension, are kept for testing the estimation of ρ during transients.

Simulation	0.45		
Closure	Classical	Distance 2	Distance 3
Mean Field	0.25		
Pair approximation	0.33		
Bethe approximation	0.37	0.40	0.36
Normalized BA	0.37	0.41	0.42
Power 1	0.44	problem	problem
Power 2	0.30	0.32	0.33

Table 2: Estimated values of the critical value β_c from simulation and by solving the ODE systems corresponding to different closures without (Classical) and with extension to distance 2 and 3 correlations.

Estimation of ρ during transients and at equilibrium

Figure 4 shows simulated and approximated trajectories trajectories, for the closures that we selected because they yield the best fit for β_c. Every closure, without and with extension when relevant, has been implemented (figure not shown), and none is better than NBA with the extension or $Pw1$. Three behaviours can be distinguished:

(i) for $\beta < \beta_c$ (Figure 4, a), the transients towards extinction are well estimated by NBA with extension or by $Pw1$, and with less accuracy by other methods

(*ii*) for $\beta \simeq \beta_c$, with $\beta > \beta_c$ (figure 4, b), all methods overestimate the prevalence both at equilibrium (ρ_{eq}) and during the transients; however, this overestimation is slighter for NBA with the extensions or for $Pw1$

(*iii*) for $\beta \gg \beta_c$ (figure 4, c), all methods estimate correctly the prevalence at equilibrium, and more poorly the transients.

Empirical correction

The numerical results show that when $\beta > \beta_c$ ($\rho_{eq} > 0$) the discrepancy between simulations and approximations increases when $\beta - \beta_c$ decreases. The discrepancy is always an overestimation of ρ_{eq}. Therefore, a smallest but artificial value $\widetilde{\beta}$ of β could yield the correct prevalence at equilibrium. A priori, this value is unknown. As the value of ρ_{eq} is an increasing function of β in simulations and in the approximated system, as soon as $\beta > \beta_c$, there exists a value $\widetilde{\beta} = f(\beta)$ such that the approximated system with $\widetilde{\beta}$ yields the same value for ρ_{eq} as the one obtained by simulation. The question is: what does the function f look like? By numerical simulations, we observed a linear relationship for f (see Figure 5). A linear fitting $\widetilde{\beta} = a\,\beta + b + \epsilon$ yields $R^2 = 0.989$ with a $p-$value $p = 4.905 \times 10^{-5}$, with 13 degrees of freedom. It is worth noting that this linear fitting does not depend on the graph, as it works equally well, with the same coefficients, for the degree 4 graph as for the degree 6 graph. Even if we have not tested it thoroughly, we conjecture that the linear relationship between β and $\tilde{\beta}$ is universal, in the sense that it does not depend on the fine geometry of the graph. This allows a very good empirical prediction of the prevalence at equilibrium through the following process: (*i*) compute $\widetilde{\beta}$ as a linear function of β (*ii*) solve numerically the equilibrium for NBA with $\tilde{\beta}$. This procedure is an extension of the modified

mean field model, as proposed in Pascual et al. (2001), where an exponent q in the mean field equation is modified such that the system with the modified exponent yields a correct estimation of the density at equilibrium, for Wator model. In our estimation, the modified parameter $\widetilde{\beta}$ is a linear function of the correct one, whereas in Pascual et al. (2001), the modified parameter depends on the dimension of the embedding space only. This relationship leads to a good fit at equilibrium while progress remains to be made for transients. Nevertheless, this provides an ad-hoc tool to improve the closures at intermediate and high values of β.

6.3 Influence of the network structure

The extended closures we proposed can now be exploited to characterize the influence of features of the graph they depend on. These features can be good candidates for being simple characteristic of the graph which control the value of the critical point and the prevalence in the vicinity of it, both at equilibrium and during the transients.

We tested the influence of the clustering coefficient θ and of the square clustering coefficients $(\alpha_1, \alpha_2, \alpha_3)$ on ρ_{eq}, using the NBA approximation with distance 3 correlations. This closure and the classical $Pw1$ provide the best approximation but the latter does not enable us to take into account the square coefficients. To test the influence of a given element of the geometry we computed the solution of the closed ODE system for a range of values of the corresponding parameters (either θ, or $(\alpha_1, \alpha_2, \alpha_3)$), the other parameters being unchanged.

Figure 6 shows the influence of the clustering coefficient θ for different

values of β. The role of θ is highly dependent on the value of the transmission rate β: for high values (corresponding to a high density of infected nodes at equilibrium) there is almost no effect. This is consistent with the relevance of mean field approximation, which ignores the geometry of the graph, for large β. The effect of θ increases as β decreases towards its critical value. We recover the fact that a high clustering coefficient impedes disease propagation (see Keeling (1999) in the case of a spatially explicit SIR model, or Proulx et al. (2005)). Indeed, infected individuals in a common highly clustered area of the interaction network share many nearest neighbours. These links are wasted because some of the paths for disease spread becomes redundant.

Figure 7 shows that the repartition of the weights between the α_i's can significantly modify the density ρ_{eq} when the transmission rate is low: when α_1 is dominant, the prevalence is lower whereas when α_3 is dominant, the prevalence is higher. The interpretation is the same as for θ: when α_1 is dominant, there is a redundancy of paths of length 2 between pairs of vertices, and infected individuals have more overlapping contacts, which are wasted. Here again, the effect of these graph features decreases when β increases and can be ignored for high value of this parameter.

7 Discussion

Stochastic models on graphs are powerful tools to represent the role of space and connection networks in infectious diseases emergence and spread. Simulations may require intensive computations in order to understand the role of parameters, and in general no analytical result is known. Standard treatment is to derive approximations of exact results, namely order 2 moment closure approximations. Here we proposed a method to improve the classical

closures in order to explore easily the importance of the graph features.

The method relies on the introduction of longer range correlations in the classical expression of the closures. In closure for models of population dynamics in continuous space, this information has more naturally been taken into account (Bolker and Pacala 1997, Murrel et al. 2004) since the models are defined through diffusion kernels rather than nearest-neighbours interaction. However, to our knowledge, the importance of long range correlations in the closures quality has been little studied in discrete space. In Snyder and Nisbet (2000), the authors have introduced an empirically-based approximation of the spatial correlations as a function of the recovery rate of the contact process at the critical point. In this article we propose a alternative method with easier interpretation and avoiding simulations for approaching the critical point, which is technically simple and remains computationally reasonable.

Regarding the closures (extended or not) leading to the best fit with simulations, a first result is the characterization of three behaviours depending on the value of the transmission rate β: (i) for $\beta < \beta_c$ (the critical value), the transients towards extinction are well estimated, (ii), for $\beta \simeq \beta_c$, with $\beta > \beta_c$ all methods overestimate the prevalence both at equilibrium and during the transients, (iii) for $\beta \gg \beta_c$ all methods estimate correctly the prevalence at equilibrium, and more poorly the transients. Such a variability in the closure performance according to the region of the parameter space (sub-critical, critical and metastable region) has already been pointed out (e.g. Krishnarajah et al. 2005). Some questions remain open for IPS models: which are the model features which remain poorly captured by the order

2 closures above the critical point? Which are the graph features which can enhance the importance of these quantities? It seems that the methods overestimate the effective number of possible contacts between infected and susceptible individuals. Indeed, in the first stages of an epidemic, infected individuals are clustered in a small area of the interaction network and a large part of the population can not be reached by direct contact with these infected. Such concerns have already being considered. In the framework of a lattice-based epidemic model, Sato et al. (1994) have proposed an Improved Pair Approximation using a discounted term for the overestimated probability. In Filipe and Gibson (2001), the overestimation of the density of infected is compensated through a mixture with an approximation which underestimates the same quantity. A more meaningful solution would be to relate this behaviour to some parameters of the interaction network (like the diameter, linked to the notion of short cut) and to introduce them into the closures.

Another contribution of this work is the analysis of the effect of the choice of the closure on the quality of the results. We compared different forms of closure from the literature, two of them (Power 1 and Power 2) introduced from continuous space, and extended here to discrete space. Several other variations from the original pair approximation have been developed. We could not consider all of them since either, as for the pair approximation, their form does not enable us to introduce longer range correlation (Sato et al. 1994, Filipe and Gibson 2001, Bauch 2005), or they are similar to the ones studied (e.g. Van Baalen 2000) has proposed a form close to the Bethe approximation, taking into account the clustering coefficient). For all the closures we have considered, we derived the corresponding versions with

longer range correlation and we compared the performance with, and without, this extension. Our conclusions are (*i*) in most of the cases, introducing longer range correlations improves the quality of the approximation, (*ii*) the two closures leading to the best fit for prevalence, at equilibrium or during transients are Power 1 without extension and normalised Bethe approximation with distance 3 correlations. Even if these two closures do not always identify accurately the phase transition, they provide reasonable estimators of the critical value. Since our objective is the exploration of the influence of the graph charasteristics rather than a precise estimation, the normalised Bethe approximation with distance 3 correlations appears as a good candidate for this goal.

Through the methodology for extended closures presented in this work, we propose some characteristics of the graph as relevant features for a simplification of the process dynamics and a study of spatial determinants of disease spread. Among these features is the well known clustering coefficient. We identify as well new coefficients, the square clustering coefficients, which enrich the characterisation of spatial organisation in a network. In Caldarelli et al. (2004), the authors have pointed our that the clustering coefficient may be not sufficient and have considered as well a description of ordering in a networks via more complex patterns of nodes. Here we link such descriptions with the dynamics of a process spreading on the graph. Exploration of the state space of clustering and square clustering coefficients is easy through the extended normalised Bethe approximation, and yields (*i*) the influence of these elements of the graph geometry is strong for low to median values of the transmission rate, (*ii*) the clustering and square clustering coefficients are both indicators of the redundancy of paths in a graph. For graphs with ho-

mogeneous degree, computations of these parameters are easy. Furthermore, when considering a real and more complex interaction network, algorithms are available to evaluate them (see Schreiber and Schwöbbermeyer 2005 and references therein). In Petermann and De Los Rios (2004b), the authors proposed a different approach to integrate elements of the graph topology in an approximation of the spatial SIS model, based on the analysis of two consecutive steps of the process. Comparing the answer of the two approximations in term of quantification of the graph's role in a disease spread could be worth being done.

Finally, we presented the method in the case of graphs with constant degree, but we are aware that graphs with non-constant degree are far more realistic: interaction networks, except in some particular cases like in orchards, are highly complex (e.g. social contact networks, airline routes networks, see Newman 2003). This is also true for ecological networks (Proulx et al. 2005). The methodology presented here can be extended without conceptual problem to this situation and should enable us to take into account parameters specific to non-homogeneous graphs (mean degree, degree distribution). The limit will only be the available computational time and this cost remains to be evaluated. Pair approximation and Bethe approximations have already been derived for a graph with a general degree distribution (Peyrard and Franc 2005). Combining the influence of degree distribution and long-range correlations should provide quite realistic closures of the real epidemic dynamics close to the critical value.

Acknowledgements

The authors wish to thank the anonymous referees for their valuable suggestions and their helpful comments for improving a previous version of this manuscript.

Appendix

Derivation of the exact expression for evolution of $P^{(1)}(01)$

Pairs of nodes at distance 1 in state 01 can result from the infection of one node in a pair 00 or from the evolution toward susceptible of a node in a pair 11. Pairs 01 can also disappear, either by infection of the susceptible node, or by recovery of the infected one. If $t^{(1)}_{ab\to cd}$ is the transition rate of a pair of nodes at distance 1 from state ab to state cd, then

$$\frac{dP^{(1)}(01)}{dt} = P^{(1)}(00)\, t^{(1)}_{00\to 01} + P^{(1)}(11)\, t^{(1)}_{11\to 01} - P^{(1)}(01)\, t^{(1)}_{01\to 11} - P^{(1)}(01)\, t^{(1)}_{01\to 00}$$

Transition rates $t^{(1)}_{11\to 01}$ and $t^{(1)}_{01\to 00}$ are equal to 1. Transition rates $t^{(1)}_{00\to 01}$ and $t^{(1)}_{01\to 11}$ depend on the state of the neighbourhood of the node moving from state 0 to state 1.

Let us consider $t^{(1)}_{01\to 11}$. First we need to introduce some notations. We will call respectively i and j the node moving from 0 to 1 and the node staying in state 1. If $N(i)$ represents the set of the h nodes in $\mathcal{V}$ linked with i by an edge, and z^t_A denotes the state of the nodes in a subset A of $\mathcal{V}$ at time t, then the number of neighbours of node i which are infected at time t are $a_i(z^t_{N(i)}) = \sum_{l\in N(i)} \delta_1(z^t_l)$ (with $\delta_1(0) = 0$ and $\delta_1(1) = 1$). Since the state of node j is known we can rewrite this expression as $1 + \sum_{l\in N(i)\setminus\{j\}} \delta_1(z^t_l)$. So,

using properties of the mathematical expectation E,

$$\begin{aligned} t^{(1)}_{01\to 11} &= E[\beta\, a_i(z^t_{N(i)}) \mid z^t_i = 0, z^t_j = 1] \\ &= \beta\left(1 + \sum_{l\in N(i)\setminus\{j\}} E[\delta_1(z^t_l) \mid z^t_i = 0, z^t_j = 1]\right) \\ &= \beta\left(1 + \sum_{l\in N(i)\setminus\{j\}} P(z^t_l = 1 \mid z^t_i = 0, z^t_j = 1]\right) \\ &= \beta\left(1 + (h-1)\frac{P^{(1,1)}(101)}{P^{(1)}(01)}\right) \end{aligned}$$

The same logic leads to

$$t^{(1)}_{00\to 01} = \beta(h-1)\frac{P^{(1,1)}(100)}{P^{(1)}(00)}$$

Finally,

$$\begin{aligned} \frac{dP^{(1)}(01)}{dt} &= P^{(1)}(11) + \beta(h-1)P^{(1,1)}(100) \\ &\quad -P^{(1)}(01)\left[1 + \beta + \beta(h-1)\frac{P^{(1,1)}(101)}{P^{(1)}(01)}\right] \end{aligned}$$

Derivation of the distance 2 pair transition rates

Let us consider the transition rate $t^{(2)}_{01\to 11}$, and let say that the site moving from 0 to 1 is site j and the other is site k. To compute this rate, we have to compute the probability of each neighbour of j to be infected. One of these neighbours is exactly the site in between j and k (since we know that the shortest path between j and k is of length 2). The term $\frac{\widehat{P}^{(1,1,2)}(011)}{P^{(2)}(01)}$ corresponds to this particular neighbour. The term $(h-1)\frac{P^{(1,2)}(101)}{P^{(2)}(01)}$ corresponds to the $h-1$ other neighbours of j, so that

$$\hat{t}^{(2)}_{01\to 11} = \beta(h-1)\frac{\widehat{P}^{(1,2)}(101)}{P^{(2)}(01)} + \beta\frac{\widehat{P}^{(1,1,2)}(011)}{P^{(2)}(01)}$$

As for the temporal evolution of $P^{(1)}(01)$, these transitions involve triplet probabilities. In the same way that we have decomposed $\widehat{P}^{(1,1)}(z_i, z_j, z_k)$ using the clustering coefficient θ, $\widehat{P}^{(1,2)}(z_i, z_j, z_k)$ is decomposed according to the distance between node i and node k. More precisely, $P^{(1,2)}(z_i, z_j, z_k)$ is the probability to find three sites (i, j, k) in configuration (z_i, z_j, z_k), given that $(i, j) \in \mathcal{E}$, given that the shortest path between j and k is of length 2 and given that i is not the middle site between j and k. Either i and k are linked by an edge (Figure 1 (a)), or the two extreme vertices are at distance 2 (Figure 1 (b)), or they are at distance 3. Let us denote $\alpha_1, \alpha_2, \alpha_3$ the respective proportions of these three organizations. We can now decompose

$$\widehat{P}^{(1,2)}(z_i, z_j, z_k) = \alpha_1 \widehat{P}^{(1,2,1)}(z_i, z_j, z_k) + \alpha_2 \widehat{P}^{(1,2,2)}(z_i, z_j, z_k) + \alpha_3 \widehat{P}^{(1,2,3)}(z_i, z_j, z_k)$$

We apply the distance 2 Bethe closure to define each term. In the two first terms $\widehat{P}^{(1,2,1)}(z_i, z_j, z_k)$ and $\widehat{P}^{(1,2,2)}(z_i, z_j, z_k)$ only $P^{(1)}$ and $P^{(2)}$ are involved. In the case of $\widehat{P}^{(1,2,3)}(z_i, z_j, z_k)$, we assume that the states of the pair of sites at distance 3 are independent. So for the distance 2 Bethe approximation we have:

$$\begin{aligned} \widehat{P}^{(1,2)}(z_i, z_j, z_k) &= \alpha_1 \frac{P^{(1)}(z_i, z_j) P^{(2)}(z_j, z_k) P^{(1)}(z_i, z_k)}{P(z_i) P(z_j) P(z_k)} \\ &\quad + \alpha_2 \frac{P^{(1)}(z_i, z_j) P^{(2)}(z_j, z_k) P^{(2)}(z_i, z_k)}{P(z_i) P(z_j) P(z_k)} + \alpha_3 \frac{P^{(1)}(z_i, z_j) P^{(2)}(z_j, z_k)}{P(z_j)} \end{aligned}$$

References

Albert, R. and A.-L. Barabási (2002). Statistical mechanics of complex networks. *Review of Modern Physics 74*, 41–97.

Bauch, C. (2005). The spread of infectious diseases in spatially structured populations: an invasory pair approximation. *Mathematical Bio-*

sciences 198, 217–237.

Bolker, B. M. and S. W. Pacala (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology 52*(3), 179–197.

Caldarelli, G., R. Pastor-Satorras, and A. Vespignani (2004). Structure of cycles and local ordering in complex networks. *The European Physical Journal B - Condensed Matter 38*(2), 183–186.

Colizza, V., A. Barrat, M. Barthélemy, and A. Vespignani (2006). The modelling of global epidemics: stochastic dynamics and predictability. *Bulletin of Mathematical Biology 68*, 1893– 1921.

Dickman, R. and M. Martins de Oliveira (2005). Quasi-stationary simulation of the contact process. *Physica A 357*(iss. 1 [special issue]), 134–141.

Dieckmann, U., R. Law, and J. A. J. Metz (Eds.) (2000). *The geometry of ecological interactions - Simplifying spatial complexity*. Cambridge Studies in Adaptive Dynamics. UK: Cambridge University Press.

Drossel, B. and F. Schwabl (1992). Self organizing critical forest fire models. *Physical Review Letters 69*, 1629–1632.

Durret, R. and S. Levin (1994). The importance of being discrete (and spatial). *Theoretical Population Biology 46*, 363–394.

Eguiluz, V. and K. Klemm (2002). Epidemic threshold in structured scale-free networks. *Physical Review Letters 89*, 108701.

Filipe, J. A. N. and G. J. Gibson (1998). Studying and approximating spatio-temporal models for epidemic spread and control. *Philosophical Transactions of the Royal Society of London 353*, 2153–2162.

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

Franc, A. (2004). Metapopulation dynamics as a contact process on a graph. *Ecological Complexity 1*, 49–63.

Harris, T. E. (1974). Contact interactions on a lattice. *Annals of Probability 2*, 969–988.

Keeling, M. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London B 266*, 859–867.

Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theoretical Population Biology 67*, 1–8.

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

Krishnarajah, I., A. Cook, G. Marion, and G. Gibson (2005). Novel moment closure approximations in stochastic epdemics. *Mathematical Biology 67*, 855–873.

Lavis, D. and G. Bell (1999). *Statistical mechanics of lattice systems: I closed forms and exact solutions*. Springler.

Ligget, T. M. (1985). *Interacting Particle Systems*, Volume 276 of *Grundlehren der Matematischen Wissenschaften*. New-York: Springer verlag.

Marro, J. and R. Dickman (1999). *Nonequilibrium Phase Transition in Lattice Models*. Monographs and Texts in Statistical Physics, Collection Alea Saclay. Cambridge, UK: Cambridge University Press.

Matsuda, H., A. Ogita, N. Sasaki, and K. Sato (1992). Statistical mechanics of population - the lattice Lotka-Volterra model. *Progress in Theoretical Physics 88*(6), 1035–1049.

Morita, T. (1994). Formal structure of the cluster variation method. *Progress of Theoretical Physics Supplement 115*, 27–39.

Murrel, D., U. Dieckmann, and R. Law (2004). On moment closures for population dynamics in continuous space. *Journal of Theoretical Biology 229*, 421–432.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review 45*, 167–256.

Pascual, M., P. Mazzega, and S. A. Levin (2001). Oscillatory dynamics and spatial scale: the role of noise and unresolved pattern. *Ecology 82*, 2357–2369.

Pastor-Satorras, R. and A. Vespignani (2001). Epidemic spreading in scale-free networks. *Physical Review Letters 86*, 3200–3203.

Petermann, T. and P. De Los Rios (2004a). Cluster approximations for epidemic processes: a systematic description of correlations beyond the pair level. *Journal of Theoretical Biology 229*(1), 1–11.

Petermann, T. and P. De Los Rios (2004b). The role of clustering and gridlike ordering in epidemic spreading. *Physical Review E 69*, 066116.

Peyrard, N. and A. Franc (2005). Cluster variational approximations for a contact process living on a graph. *Physica A 358*, 575–592.

Proulx, S., D. Promislow, and P. Phillips (2005). Network thinking in ecology and evolution. *Trends in Ecology and Evolution 20*(6), 345–353.

Sato, K., H. Matsuda, and A. Sasaki (1994). Pathogen invasion and host extinction in lattice structured population. *Journal of Mathematical Biology 32*, 251–268.

Schreiber, F. and H. Schwöbbermeyer (2005). Frequency concepts and pattern detection for the analysis of motifs in networks. *Transactions on Computational Systems Biology 3 (LNBI 3737)*, 89–104.

Singer, A. (2004). Maximum entropy formulation of the Kirkwood superposition approximation. *Journal of Chemical Physics 121*(8), 3657–3666.

Snyder, R. and R. Nisbet (2000). Spatial structure and fluctuations in the contact process and related models. *Bulletin of Mathematical Biology 62*, 959–975.

Van Baalen, M. (2000). *Pair approximations for different spatial geometries*, Chapter 19 in The geometry of ecological interactions, Dieckmann, U., Law, R. and Metz, J.A.J. (Ed.), pp. 359–387. Cambridge University Press.

Van Baalen, M. and D. Rand (1998). The unit of selection in viscous populations and the evolution of altruism. *Journal of Theoretical Biology 193*, 631–648.

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

Yedidia, J., W. Freeman, and Y. Weiss (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory 51*(7), 2282–2312.

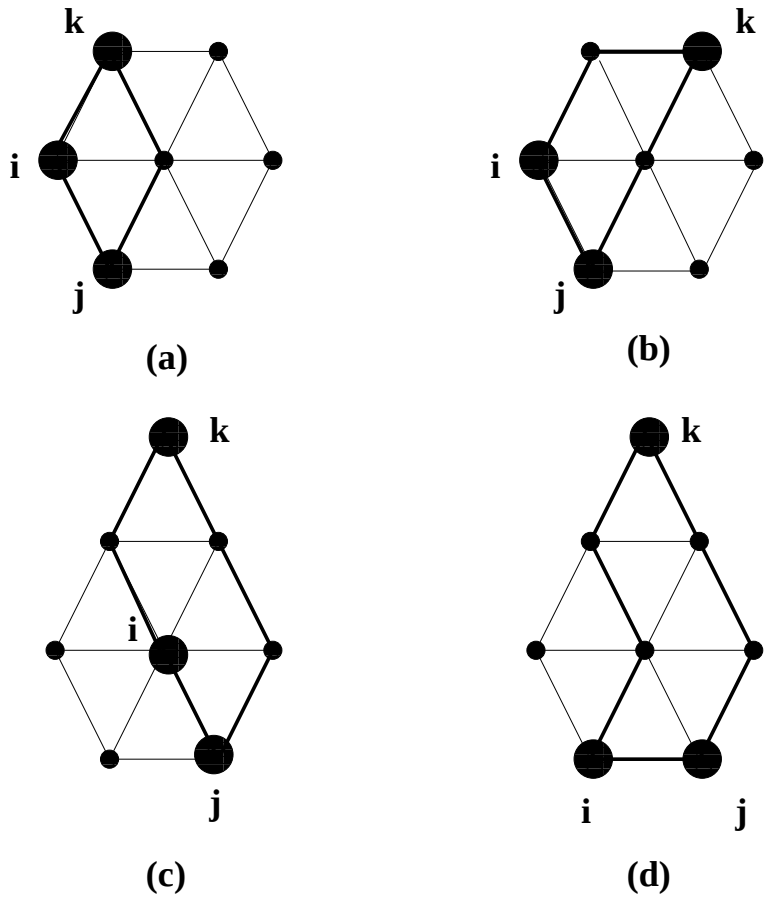

Figure 1: Illustration, on the triangle grid, of patterns of triplets of nodes i, j, k corresponding to weight (a) α_1, (b) α_2, (c) γ_2, (d) γ_3. Big vertices are the three nodes of the triplet.

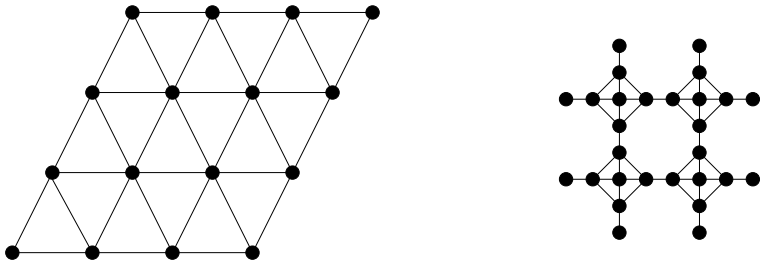

Figure 2: Examples of graphs with constant degree: left, triangular grid with $h = 6$, right, $h = 4$.

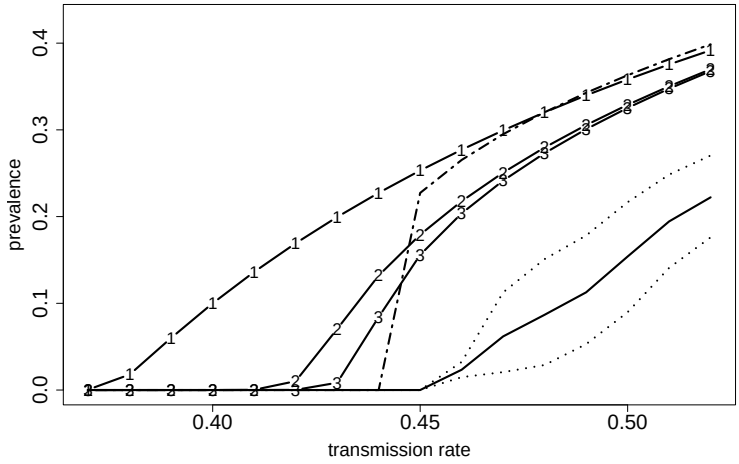

Figure 3: Estimations of the prevalence at equilibrium close to β_c by NBA, $Pw1$ and simulations, for the homogeneous graph with $h = 4$. Full line and dotted lines: mean and 90 % confidence interval from simulations, full line with symbols 1 to 3: NBA respectively without and with distance 2 and distance 3 correlations, dash-dotted line: $Pw1$ without extension to longer range correlations.

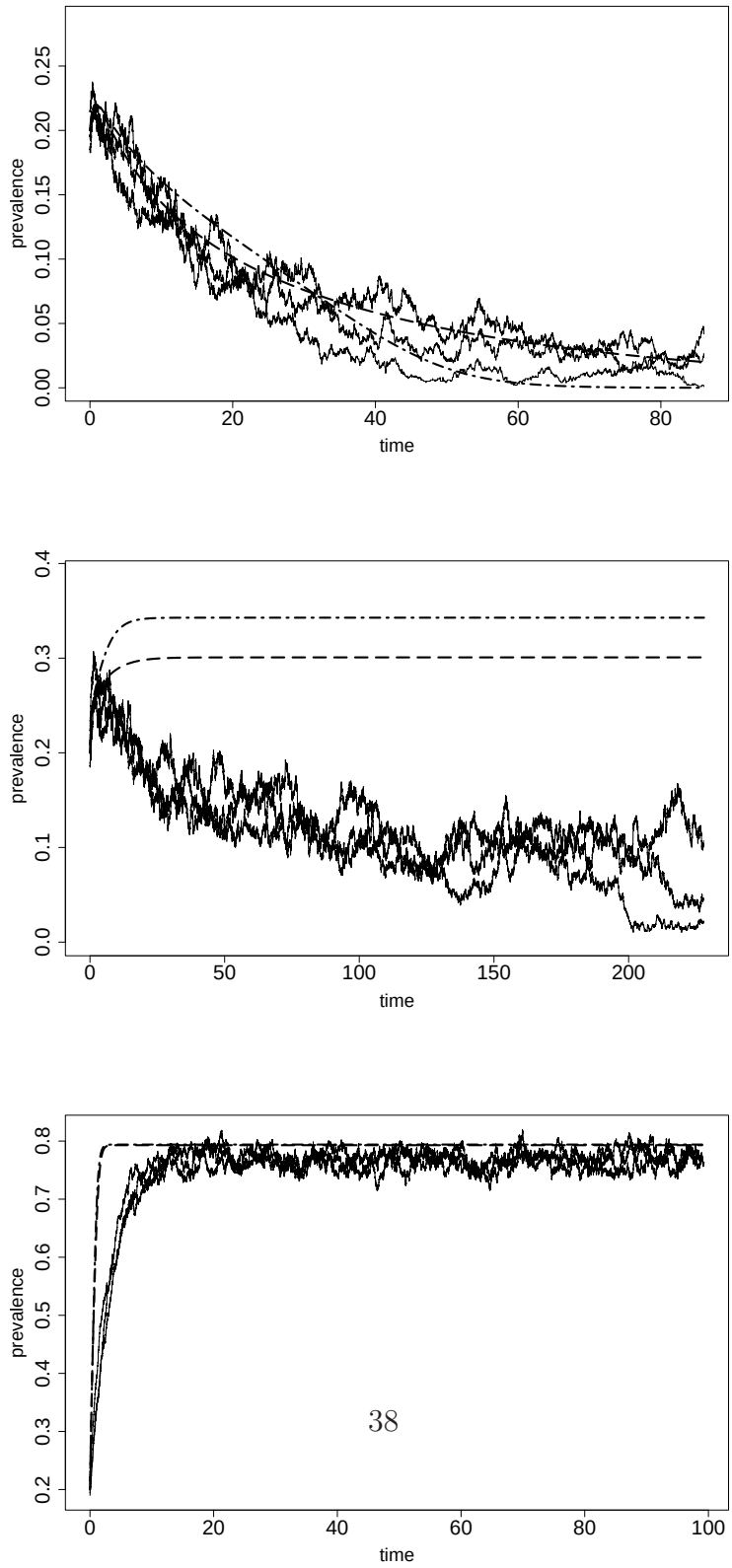

Figure 4: Estimations of the prevalence temporal evolution for the homogeneous graph with $h = 4$. x-axis: time, y-axis: prevalence, full lines: simulations of the contact process, dashed line: NBA with distance 3 correla-

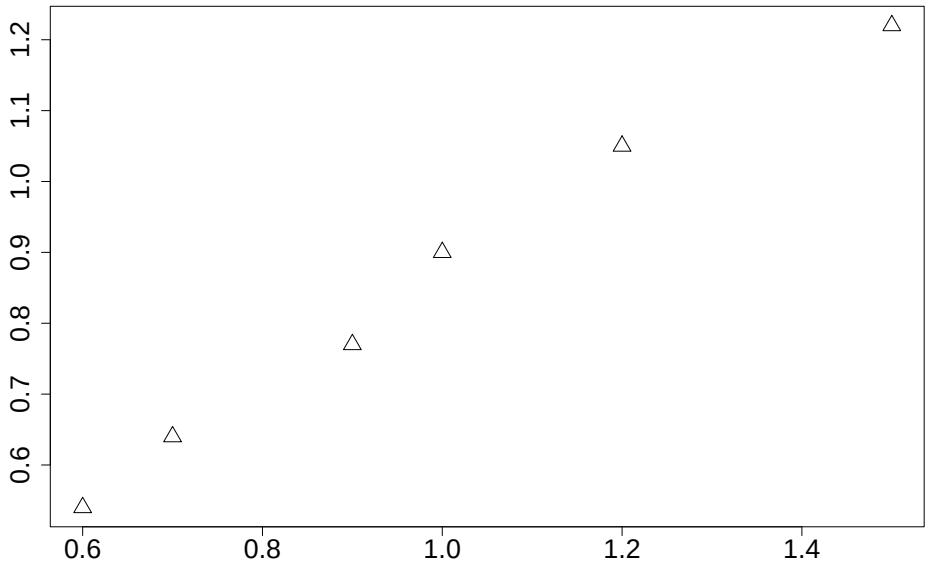

Figure 5: Linear relationship between the real value of β (x-axis) and the value leading to the same equilibrium with NBA extended to distance 3 correlations (y-axis) for the homogeneous graph with $h = 4$.

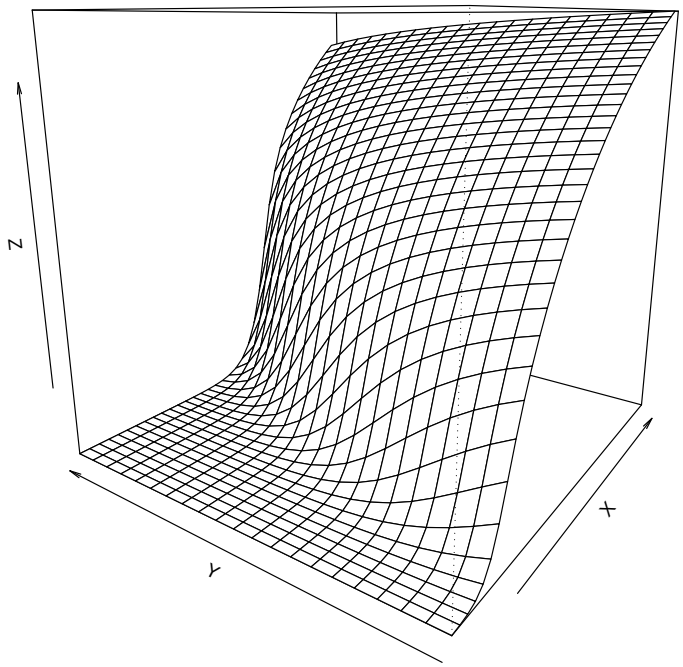

Figure 6: Influence of the value of the clustering coefficient θ on the prevalence at equilibrium for the homogeneous graph with $h = 4$. x-axis: β ranging from 0.33 to 0.69, y-axis: θ ranging from 0 to 1, z-axis: ρ_{eq} ranging from 0 to 0.58 (arrows indicate increasing values).

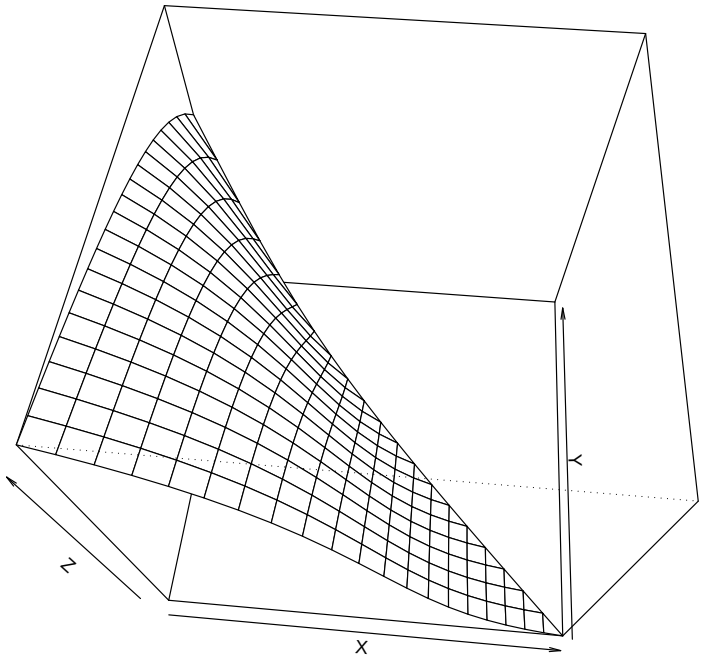

Figure 7: Influence of the square clustering coefficients $(\alpha_2, \alpha_2, \alpha_3)$ on the prevalence at equilibrium for the homogeneous graph with $h = 4$. x-axis: α_1 ranging from 0 to 1, y-axis: α_2 ranging from 0 to 1 , z-axis: ρ_{eq} ranging from 0.01 to 0.26 (arrows indicate increasing values). The y-axis can equivalently be interpreted as α_3 ranging from 0 to 1.