PHYSICAL REVIEW E **88**, 052801 (2013)

Epidemic thresholds for bipartite networks

D. G. Hernández
Centro Atómico Bariloche and Instituto Balseiro, 8400 S. C. de Bariloche, Argentina
and Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

S. Risau-Gusman*
Centro Atómico Bariloche, 8400 S. C. de Bariloche, Argentina
and Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

(Received 13 June 2013; published 4 November 2013)

It is well known that sexually transmitted diseases (STD) spread across a network of human sexual contacts. This network is most often bipartite, as most STD are transmitted between men and women. Even though network models in epidemiology have quite a long history now, there are few general results about bipartite networks. One of them is the simple dependence, predicted using the mean field approximation, between the epidemic threshold and the average and variance of the degree distribution of the network. Here we show that going beyond this approximation can lead to qualitatively different results that are supported by numerical simulations. One of the new features, that can be relevant for applications, is the existence of a critical value for the infectivity of each population, below which no epidemics can arise, regardless of the value of the infectivity of the other population.

DOI: 10.1103/PhysRevE.88.052801 PACS number(s): 89.75.Hc, 87.10.−e, 87.23.Ge, 87.19.X−

I. INTRODUCTION

Mathematical models have been used to study the spread of infectious diseases at least since the work of Bernoulli in 1760. But it was only after the pioneering work of Ross, and Kermack and McKendrick that the field of mathematical epidemiology began to be considered as a serious alternative for the prediction and control of infectious diseases [1,2]. The first models (still widely used today) were deterministic models in which the population is divided into compartments and the evolution of the number of individuals is given by a set of differential equations. It is assumed that when an infected individual interacts with a susceptible individual there is a fixed probability that the disease is transmitted. The models also need an assumption about the way in which members of the different compartments can interact. The most popular and simple assumption is the Law of Mass Action [3] which, drawing an analogy with the movement of particles in a gas, postulates that the probability that two individuals meet is simply proportional to the product of the relative populations of their respective compartments. This is also known as homogeneous mixing. For diseases which confer immunity (such as measles), or for which no cure is presently known (such as AIDS), the simplest model is called a SIR model because it has only three compartments: susceptibles (S), infecteds (I), and removeds (R), corresponding either to immune or dead individuals. If the disease does not confer immunity, or if this immunity is rapidly lost (as is the case for influenza or gonorrhea), the simplest possibility is the SIS model, in which infected individuals become susceptible after a certain amount of time (on average), dependent on the disease at hand. The most remarkable feature of these models is the existence of a threshold value for the infectivity, such that the epidemic spread of the disease depends on whether its infectivity is below or above this threshold. Even though adding compartments makes a model more realistic, SIR and SIS models are still widely used.

As mentioned above, homogeneous mixing is achieved by selecting at random from all the compartments the individuals that are to interact. In general, however, real individuals can interact only with a very limited set of the population. Thus, one way to go beyond the homogeneous mixing assumption is to consider that the members of the population form a social network. The structure of such a network depends on the disease being considered [4]. For example, for sexually transmitted diseases the relevant structure of the population is a bipartite graph representing the interactions between two different groups, men and women (if homosexual interactions are neglected) [5]. It has been shown that the SIR model on networks is very closely related to the problem of bond percolation [6,7] and can thus be solved using generating functions, at least for networks without loops. On the other hand, the SIS models on networks is equivalent to the contact process on graphs [8] which, because of the possibility of reinfection, is a much more difficult problem that has been solved analytically only in a few cases. Thus, in this case approximations can provide some useful insights. Probably the simplest and best known is the *mean field* (MF) approximation that leads to some interesting results. One example is the simple relationship that exists between the network and the epidemic threshold of the effective transmission rate, λ_c [1,9], for SIS models:

$$\lambda_c = \frac{\overline{k}}{\overline{k^2}}, \tag{1}$$

where $\overline{k}$ is the average of the degree distribution of the network, and $\overline{k^2}$ the average of the squared degree. Interestingly, a similar result holds for SIR models in networks [7]:

$$\lambda_c = \frac{\overline{k}}{\overline{k^2} - 2\overline{k}}. \tag{2}$$

For sexually transmitted diseases it is well known that in general the rate of infection is different for men and women [1].

*srisau@cab.cnea.gov.ar

Using a mean field approach an equation can be found for the critical value of both effective rates for SIS models in bipartite networks [10]:

$$\lambda_M \lambda_F = \frac{\overline{k_M}\,\overline{k_F}}{\overline{k_M^2}\,\overline{k_F^2}}, \tag{3}$$

where the subindices M and F denote that the corresponding quantities are calculated within the male and female populations. For SIR models the equation for the critical values is [7]

$$\frac{\lambda_M \lambda_F}{(\lambda_M+1)(\lambda_F+1)} = \frac{\overline{k_M}\,\overline{k_F}}{\left(\overline{k_M^2}-\overline{k_M}\right)\left(\overline{k_F^2}-\overline{k_F}\right)} \equiv \lambda_{\mathrm{SIR}}^2. \tag{4}$$

Equations (3) and (4) define rectangular hyperbolas in (λ_M,λ_F) space. Both are symmetric, with asymptotes at $\lambda_M = \lambda^*$ and $\lambda_F = \lambda^*$. The main difference between the models lies in the position of the asymptotes: whereas for the SIS model $\lambda^*_{\mathrm{SIS}} = 0$, for the SIR model we have $\lambda^*_{\mathrm{SIR}} = \lambda^2_{\mathrm{SIR}}/(1-\lambda^2_{\mathrm{SIR}})$. An important consequence of this is that in a SIR model it is possible to avoid the epidemic spread of a disease by acting on only one population because, if the infectivity of that population is lowered below λ^*, an epidemic becomes impossible, regardless of the infectivity of the other population. According to Eq. (3) this would not be possible for diseases of SIS type. However, it is well known that in some cases the results obtained using the mean field approximation can be qualitatively wrong. For instance, whereas Eq. (1) gives a positive epidemic threshold for a contact process on a network with a power law degree distribution if $\alpha > 2$, it has been shown rigorously that in fact the epidemic threshold vanishes for all values of α [11,12].

In the next sections we show that when the thresholds for bipartite networks are calculated with an approximation that goes beyond mean field the result is qualitatively different to Eq. (3). Our results suggest that for SIS models λ^*_{SIS} is always a positive number, and that, unlike the result for SIR models, it can have different values for men and women. Even though these results are not exact, numerical simulations seem to confirm their validity. In Sec. II we present the model and the approximation used, and we analyze the case of a regular bipartite network. In Sec. III we extend these results to a network with an arbitrary degree distribution and no assortativity. In the last section we draw some conclusions. As in the rest of this paper only the SIS model is analyzed, in the following the subindices that specify the model (SIS or SIR) will be dropped.

II. PAIR APPROXIMATION FOR REGULAR BIPARTITE NETWORKS

We consider a population of two types of agents, M (males) and F (females), placed on the vertices of a bipartite network, so that members of one group have only members of the other group as neighbors. Infected M and F individuals can transmit the disease to its neighbors at a rate β_M and β_F, respectively, and they recover from the disease with rates γ_M and γ_F. The probability $P_t(I_M^x)$ that, at time t, the male at site x is infected satisfies the equation:

$$\dot{P}_t\left(I_M^x\right) = -\gamma_M P_t\left(I_M^x\right) + \beta_F \sum_{y\in n_x} P_t\left(I_F^y S_M^x\right), \tag{5}$$

where $P_t(I_M^y S_F^x)$ is the probability that the female at site x and the male at site y are infected and susceptible, respectively. The sum runs over all the sites that are neighbors of site x. To obtain the corresponding equation for $P_t(I_F^x)$, M and F must be swapped in Eq. (5). Because S and I are the only two possible states, we have $P_t(I_M^x) + P_t(S_M^x) = P_t(I_F^x) + P_t(S_F^x) = 1$. The single site and pair probabilities are related by equations like $P_t(I_M^y S_F^x) + P_t(I_M^y I_F^x) = P_t(I_M^y)$, with the corresponding permutations of indices and states. For the pair probabilities the evolution equations are:

$$\begin{aligned}
\dot{P}_t\left(I_M^y S_F^x\right) &= \gamma_F P_t\left(I_M^y I_F^x\right) - \gamma_M P_t\left(I_M^y S_F^x\right) - \beta_M P_t\left(I_M^y S_F^x\right) \\
&\quad + \beta_F \sum_{z\neq y} P_t\left(S_F^x S_M^y I_F^z\right) - \beta_M \sum_{z\neq y} P_t\left(I_M^y S_F^x I_M^z\right) \\
\dot{P}_t\left(S_M^y S_F^x\right) &= \gamma_F P_t\left(S_M^y I_F^x\right) + \gamma_M P_t\left(I_M^y S_F^x\right) \\
&\quad - \beta_M \sum_{z\neq y} P_t\left(I_M^z S_F^x S_M^y\right) - \beta_F \sum_{z\neq y} P_t\left(I_F^z S_M^y S_F^x\right),
\end{aligned} \tag{6}$$

where we now have triplet probabilities such as $P_t(I_M^y S_F^x S_M^z)$ that represents the probability that the female at site x and the males at sites y and z (both connected to x) are susceptible, infected and susceptible, respectively. The equation for $P_t(S_M^x I_F^y)$ is obtained from Eqs. (6) by swapping the indices M and F, and $P_t(I_M^x I_F^y)$ is obtained from the normalization condition: $P_t(S_M^x I_F^y) + P_t(S_M^x S_F^y) + P_t(I_M^x S_F^y) + P_t(I_M^x I_F^y) = 1$.

We begin our analysis by considering the case of a regular bipartite network, where all males have k_M female neighbors (i.e., partners) and all females have k_F male neighbors (which implies that N_F and N_M must satisfy $k_M N_M = k_F N_F$). If we assume that the system is homogeneous, in the sense that the probabilities do not depend on the sites involved, we can drop the site indices and the equations become:

$$\begin{aligned}
\dot{P}_t(I_M) &= -\gamma_M P_t(I_M) + \beta_F k_M P_t(I_M S_F) \\
\dot{P}_t(I_M S_F) &= \gamma_F P_t(I_M I_F) - (\gamma_M + \beta_M) P_t(I_M S_F) \\
&\quad + \beta_F (k_M - 1) P_t(S_F S_M I_F) \\
&\quad - \beta_M (k_F - 1) P_t(I_M S_F I_M) \\
\dot{P}_t(S_M S_F) &= \gamma_F P_t(S_M I_F) + \gamma_M P_t(I_M S_F) \\
&\quad - \beta_M (k_F - 1) P_t(I_M S_F S_M) \\
&\quad - \beta_F (k_M - 1) P_t(I_F S_M S_F).
\end{aligned} \tag{7}$$

To avoid considering also the equations that correspond to the evolution of triplets, it is necessary to choose an ansatz for the relationship between pair and triplet probabilities. The most often used ansatz is

$$P_t(I_M S_F S_M) = \frac{P_t(I_M S_F) P_t(S_F S_M)}{P_t(S_F)}, \tag{8}$$

with analogous formulas for the other triplets. This is known as the pair approximation (PA) [13–15]. As we are here mostly interested in steady state properties, in the following

we will drop the temporal dependencies from all quantities. Using the PA, the system of four equations for the pair probabilities [Eqs. (7) with their left-hand sides set to 0] can be reduced to a system of two equations for the fraction of infected M and F individuals $n_F = (k_F\beta_M/\gamma_F)P(I_M S_F)$ and $n_M = (k_M\beta_F/\gamma_M)P(S_M I_F)$:

$$0 = -(\gamma_M(1+\lambda_M)(1-n_F) + \gamma_F n_F)\frac{(1-n_F)^2\gamma_F n_F}{\gamma_M\lambda_M} + \gamma_M n_M(1-n_F)(1-n_M)\left(k_F - \frac{\gamma_F n_F}{\gamma_M\lambda_M} - \frac{\gamma_M\beta n_M}{\gamma_F\lambda_F}\right)$$
$$+ \frac{\gamma_F(1-n_M)-\gamma_M n_M}{\gamma_M+\gamma_F}\left[\frac{(1-n_F)\gamma_M^2\beta n_M^2}{\gamma_F\lambda_F}\frac{\gamma_F^2(1-n_M)n_F^2}{\gamma_M\lambda_M} + (1-n_F)(1-n_M)(\gamma_F n_F + \beta\gamma_M n_M)\right], \quad (9)$$

where $\beta = k_F/k_M$ and λ_F and λ_M are the effective transmission rates, defined as $\lambda_F = \beta_F/\gamma_F$ and $\lambda_M = \beta_M/\gamma_M$. The remaining equation is obtained from Eq. (9) by swapping M and F and replacing β by $1/\beta$.

For unipartite networks it has been shown [15] that this approximation provides a significant improvement in the agreement between theory and numerical stochastic simulations. Figure 1 shows that this is also the case for bipartite networks. In particular, the figure shows that the approximation for the epidemic threshold is much better than the one provided by MF. To calculate explicitly the epidemic threshold we use the standard method of linearizing the steady state equations [Eqs. (7)]. The linearized equations can be compactly written in matrix form as

$$0 = \mathbf{M}P, \quad (10)$$

where $P = [P(I_M S_F), P(S_M I_F)]$ and

$$\mathbf{M} = \begin{pmatrix} -\gamma_F(1+\lambda_F\gamma) & \lambda_M(k_F-\gamma) \\ \lambda_F(k_M-1+\gamma) & -\gamma_M(1+\lambda_M(1-\gamma)) \end{pmatrix} \quad (11)$$

where $\gamma = \gamma_F/(\gamma_F+\gamma_M)$. The relationship between the critical values for the infectivities is obtained by imposing the condition that the determinant of $\mathbf{M}$ vanishes. This gives

$$\left(\frac{1}{\lambda_F}+\gamma\right)\left(\frac{1}{\lambda_M}+1-\gamma\right) = (k_F-\gamma)(k_M+\gamma-1). \quad (12)$$

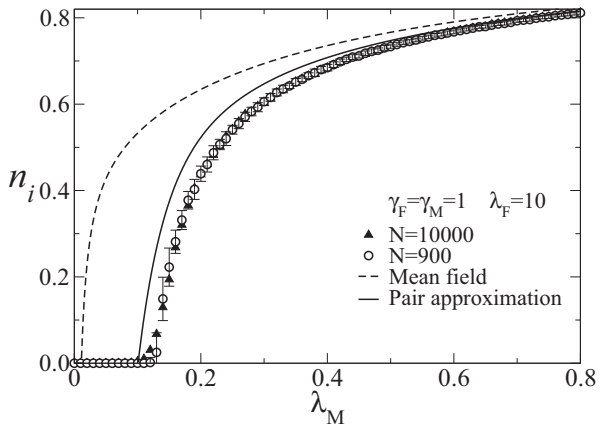

FIG. 1. Fraction of the population that is infected, in the steady state, for a regular bipartite network with $k_F = k_M = 3$. The curves give the predictions of the mean field (dashed) and the pair approximation (full). The points represent the average fraction of infecteds obtained from 100 simulations of an epidemic process in networks with 20 000 (triangles) and 1800 (circles) individuals.

Figure 2 shows that the PA provides a much better approximation to the simulation results than the MF. For the sake of comparison we also show the results of numerical simulations of a SIR epidemic in the same network and with the same infectivities, together with the theoretical curve obtained within the generating functional formalism [7], which is generally held to be in good agreement with simulations. The figure shows that the PA curve is above the MF curve in the whole range. From Eq. (12) it can be demonstrated that this is also the case for all values of k_F, k_M, and γ. There are also some important qualitative differences between the predictions of MF and PA. One of them is that the dependence on the recovery rates is not restricted to a rescaling of the infectivities, because γ appears explicitly in Eq. (12).

Another significant difference is the presence of nonvanishing horizontal and vertical asymptotes. This has the important consequence that an epidemic can be avoided by acting on only one population. Note that in the MF scenario this is not the case: If the infectivity of one population is lowered, the system can enter an epidemic-free region but only if the infectivity of the other group is small enough. In the PA scenario there are critical values for both infectivities, below which no epidemic is possible, no matter how large the infectivity of the other group. These values, defined by the asymptotes to

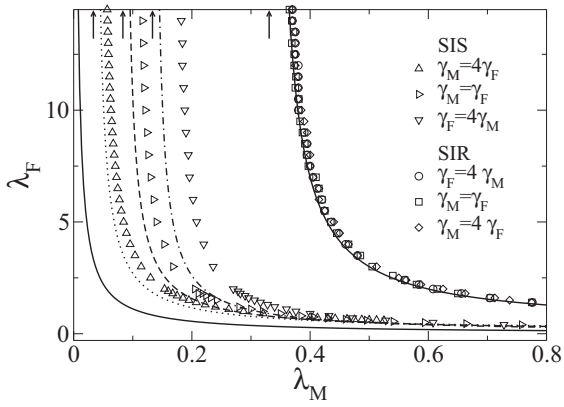

FIG. 2. Epidemic thresholds for a regular bipartite network with $k_f = k_m = 3$. Full curves correspond to mean field predictions for SIS (left) and SIR (right) epidemics. The other curves correspond, from left to right, to the prediction of the PA for $\gamma = 1/5$, $\gamma = 1/2$, $\gamma = 4/5$. The position of the asymptotes is shown by a vertical arrow to the left of each curve. The points represent the thresholds obtained from simulations of an epidemic process in a network with 20 000 males and 20 000 females. They are given by the infectivities for which the average number of infecteds is $\approx 1/\sqrt{N}$ ($N = 40\,000$).

the rectangular hyperbola of Eq. (12) are

$$\lambda_F^* = \frac{1-\gamma}{k_M k_F - \gamma k_M - (1-\gamma)k_F}$$
$$\lambda_M^* = \frac{\gamma}{k_M k_F - \gamma k_M - (1-\gamma)k_F}. \quad (13)$$

Note that λ_M^* is an increasing function of γ, whereas λ_F^* is a decreasing function of γ.

So far the distinctive features of the PA approximation have only been shown for a very simple regular network. In the next section we show that, even though the calculations are less straightforward, the same qualitative features can be shown to appear when the approximation is applied to bipartite networks with an arbitrary distribution of contacts.

III. GENERALIZATION TO ARBITRARY DEGREE DISTRIBUTIONS

In this section we consider a connected bipartite network where the agents in each group can belong to different subgroups with different connectivity properties. For example, one subgroup could be the husbands (or wives), i.e., people having only one partner that in turn has no other partner, which would be different to the subgroup of people connected with one sex worker (i.e., individuals having a very large number of partners). Note that networks with the same degree distributions might have different divisions into subgroups. The relationship between subgroups is characterized by the numbers N_{ij}^F (N_{ij}^M), which give for a female (male) of group i, the number of partners belonging to subgroup j. For the sake of clarity, female subgroups are indicated by even numbers and male subgroups by odd numbers.

Even though the generalization of Eqs. (5), (6), and (7) for this case is straightforward, it is in general not possible to obtain analytical results for the evolution of the fraction of infected individuals in the population. However it is possible to study the epidemic threshold and to obtain some general results. For this we turn to the linearization of the full set of equations [i.e., the generalization of Eq. (9)] using the PA, obtaining the same equation as in the previous section [Eq. (10)], but with a different matrix. In the general case, the matrix $\mathbf{M}$ can be written as

$$\mathbf{M} = \begin{pmatrix} -\gamma_F(1+\lambda_F\gamma))\,\mathbf{I} & \gamma_M\lambda_M\mathbf{A}_\gamma^F \\ \gamma_F\lambda_F\mathbf{A}_\gamma^M & -\gamma_M(1+(1-\gamma)\lambda_M)\,\mathbf{I} \end{pmatrix}. \quad (14)$$

The matrices $\mathbf{A}_\gamma^F$ and $\mathbf{A}_\gamma^M$ have dimension $c_F c_M \times c_F c_M$, where c_F (c_M) is the number of different connectivities of the degree distribution of women (men), and are defined as $\mathbf{A}_\gamma^F = \mathbf{A}^F - \gamma\mathbf{I}$ and $\mathbf{A}_\gamma^M = \mathbf{A}^M - (1-\gamma)\mathbf{I}$. $\mathbf{A}^F$ and $\mathbf{A}^M$ only depend on the network, and are defined as:

$$\mathbf{A}^F = \mathbf{C}\big[\mathbf{D}\big(N_{21}^F, N_{41}^F, \dots\big), \mathbf{D}\big(N_{23}^F, N_{43}^F, \dots\big), \dots\big]$$
$$\mathbf{A}^M = \mathbf{D}\big[\mathbf{C}\big(N_{12}^M, N_{14}^M, \dots\big), \mathbf{C}\big(N_{32}^M, N_{34}^M, \dots\big), \cdots\big], \quad (15)$$

where $\mathbf{C}(\mathbf{X_1}, \mathbf{X_2}, \dots)$ is a singular block matrix where $\mathbf{X_1}$ is repeated in all the blocks of the first column, $\mathbf{X_2}$ in all the blocks of the second column, etc., and $\mathbf{D}(\mathbf{X_1}, \mathbf{X_2}, \dots)$ is a diagonal block matrix whose diagonal blocks are $\mathbf{X_1}$, $\mathbf{X_2}$, etc. As in the previous section, the equation for the critical values of λ_F and λ_M is obtained by setting the determinant of $\mathbf{M}$ to 0. Using Schur complements [16], it is straightforward to check that this is equivalent to solving the equation

$$0 = \det\big[\lambda_M\lambda_F\mathbf{A}_\gamma^M\mathbf{A}_\gamma^F - (1+\lambda_F\gamma)(1+(1-\gamma)\lambda_M)\mathbf{I}\big], \quad (16)$$

which, in turn, is equivalent to solving an eigenvalue equation. Thus, the critical equation for the infectivities can be written as:

$$\left(\frac{1}{\lambda_F} + \gamma\right)\left(\frac{1}{\lambda_M} + 1 - \gamma\right) = \mu, \quad (17)$$

where μ is the largest eigenvalue of the matrix $\mathbf{A}_\gamma^M\mathbf{A}_\gamma^F$. This describes a hyperbola whose asymptotes are

$$\lambda_F^* = \frac{1-\gamma}{\mu - \gamma(1-\gamma)}$$
$$\lambda_M^* = \frac{\gamma}{\mu - \gamma(1-\gamma)}. \quad (18)$$

The fact that all N_{ij}^F and N_{ij}^M are natural numbers implies that, for all values of γ, both $\mathbf{A}_\gamma^F$ and $\mathbf{A}_\gamma^M$ are nonnegative matrices, whose product is a primitive matrix. Furthermore, it can be shown that, for all networks, $\mu \geqslant \gamma(1-\gamma)$, which implies that there always exists a pair of critical values for the infectivities.

Unfortunately, in most empirical studies, in the form of surveys on sexual behavior, detailed information for the connectivity between subgroups is not available. Furthermore, in general only the degree distribution is available. In this case each subgroup i of men and women has a different connectivity k_i^M or k_i^F, and it is necessary to propose an ansatz for the quantities N_{ij}^F and N_{ij}^M. Note that any ansatz must satisfy the identities $N_{ij}^M k_i^M = N_{ji}^F k_j^F$. The most natural ansatz is to assume that, even though it has a prescribed degree distribution, the network is otherwise random. In this case, we have:

$$N_{ij}^F = \frac{k_i^F k_j^M}{\overline{k_F}} n_i^F$$
$$N_{ij}^M = \frac{k_i^M k_j^F}{\overline{k_M}} n_i^M, \quad (19)$$

where $\overline{k_F}$ and $\overline{k_M}$ are the average number of contacts for females and males, respectively, and n_i^F and n_i^M are the fraction of females and males in group i, respectively. In this case, the matrices $\mathbf{A}^F$ and $\mathbf{A}^M$ have simpler expressions:

$$\mathbf{A}^F = (\overline{k^M}\,\overline{k^F})^{-1}\mathbf{C}\big[k_1^M n_1^M, k_3^M n_3^M, \dots\big] \otimes \mathbf{D}\big[k_2^F, k_4^F, \dots\big]$$
$$\mathbf{A}^M = (\overline{k^M}\,\overline{k^F})^{-1}\mathbf{D}\big[k_1^M, k_3^M, \dots\big] \otimes \mathbf{C}\big[n_2^M k_2^F, n_4^F k_4^F, \dots\big], \quad (20)$$

where $\otimes$ denotes the Kronecker matrix product. It is straightforward to show that the largest eigenvalue of the positive matrix $\mathbf{A}^M\mathbf{A}^F$ is $\frac{\langle k_M^2\rangle\langle k_F^2\rangle}{\langle k_M\rangle\langle k_F\rangle}$. On the other hand the matrix $\mathbf{A}_\gamma^M\mathbf{A}_\gamma^F$ is componentwise smaller than $\mathbf{A}^M\mathbf{A}^F$, which implies [16] that $\mu < \frac{\langle k_M^2\rangle\langle k_F^2\rangle}{\langle k_M\rangle\langle k_F\rangle}$. In turn, this implies that the curve given defined by Eq. (17) is always above the curve defined by Eq. (3).

It is important to note that in the ansatz used, Eq. (19), the quantities obtained are in general not natural numbers. This does not change any of the features mentioned, unless the numbers obtained are smaller than 1. In this case the

matrix may not be positive for all values of γ and the largest eigenvalue could even be negative. In order to avoid this, the formalism should be modified in some way. We have chosen the most obvious one, which is simply to set to 0 all negative components of the matrices $\mathbf{A}_\gamma^F$ and $\mathbf{A}_\gamma^M$. The product of the resulting matrices is again componentwise smaller than $\mathbf{A}^M\mathbf{A}^F$, and thus the curves obtained for the critical infectivities are always above the corresponding curve for the MF.

To test this approximation we apply it to a real network, obtained from data from the National Survey of Sexual Attitudes (NATSAL 2000) [17,18], conducted in Britain in 2000. Note that given the ansatz of random connectivity, we have to "extract" the connected part of the network, which usually has a different degree distribution. For example, in general there is an important fraction of the population that reports only one partner, which implies the existence of many couples, which, by definition, are not connected to the main component of the network. Thus, these individuals (and some others) must be removed from the network in order to obtain the connected component. An additional problem that arises in the analysis of data obtained in sexual behavior surveys is the inconsistency between the reported number of sexual partners of men and women. More specifically, the number of female partners reported by men usually [19] is larger than the number of male partners reported by women. Many different hypothesis have been advanced to account for this effect. If these data are to be used for a model, they have to be corrected to make both degree distributions consistent, which implies choosing one of the many explanations that exist [20]. As our purpose here is only to test the accuracy of the pair approximation, we have chosen the hypothesis of women underreporting, which provides one of the simplest corrections to the data. Thus, we modify the partner distribution of women by randomly augmenting the number of contacts of females until their total number of contacts is the same as the total reported by men. Afterwards, to build a plausible degree distribution for the connected part of this network, we have used only the individuals with degrees larger than 1, rescaling the distribution accordingly. This ensures that the resulting degree distribution corresponds to a connected random bipartite network [21]. The cumulative distributions of contacts obtained is shown in the inset of Fig. 3.

Once the degree distributions of the connected part of the network have been obtained, μ can be calculated, and the curves for the epidemic thresholds can be drawn. In Fig. 3 we compare the results from a numerical simulation of the spread of an STD in a network as the one mentioned above, with the theoretical prediction obtained by using the mean field and pair approximations. We show the case of identical values of the recovery rates, and also the case where the recovery rate for men is four times larger than the rate for women, which is close to the relationship between the values used in models of gonorrhea transmission [22,23]. The figure also displays the curve that results when this relationship is inverted, in order to show the asymmetry between men and women. We see that, as happens in the case of a regular network, the PA provides a more accurate prediction than MF. It is important to note that in this case the dependence of the thresholds on γ is even stronger than predicted by the PA.

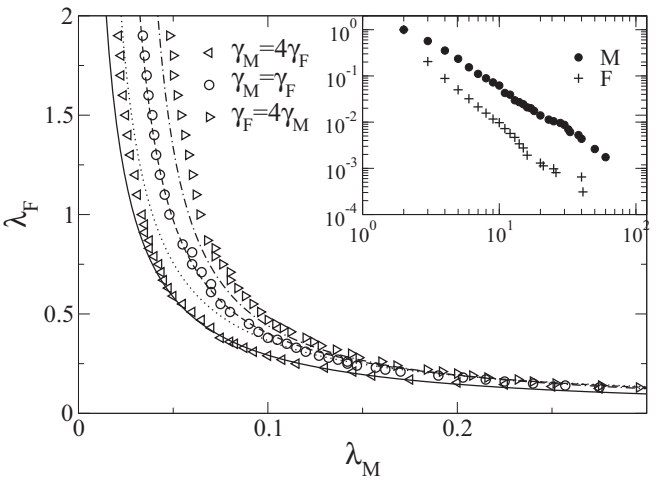

FIG. 3. Epidemic thresholds for a bipartite network obtained from data from NATSAL. The cumulative degree distributions for males (M) and females (F) are shown in the inset. The full curve in the main figure corresponds to the MF prediction. The other curves correspond, from left to right, to the prediction of the PA for $\gamma = 1/5$, $\gamma = 1/2$, $\gamma = 4/5$. The points represent the thresholds obtained from simulations of an epidemic process in a network with 13 000 males and 20 000 females. They are given by the infectivities for which the average number of infecteds is $\approx 1/\sqrt{N}$ ($N = 33\,000$).

IV. CONCLUSIONS

The importance of understanding the spread of sexually transmitted diseases in sexual networks can hardly be overestimated. Furthermore, perhaps the most important type of network in this regard are bipartite networks, because they are necessary to model those diseases that are transmitted between men and women. However, very few general results are known for the spread of STD on such networks, and they are in general obtained using the mean field approximation. Thus, it is important to see whether such results can be improved (at least in terms of accuracy) using a different approach. One of the possible ways to go beyond the mean field approach is to use the pair approximation.

In the previous sections we have shown that the pair approximation provides better predictions not only in a quantitative but also in a qualitative sense. In (λ_M,λ_F) space, MF gives a critical curve (i.e., a curve that separates the parts of the space where the endemic state is stable or not) which is a rectangular hyperbola with asymptotes at $\lambda_M^* = 0$ and $\lambda_F^* = 0$. On the other hand, the PA also gives a hyperbola but with asymptotes at nonvanishing values of the infectivities, $\lambda_M^* > 0$ and $\lambda_F^* > 0$. If this was true for real systems, it would have the important consequence that, by lowering the infectivity of only one population below a critical value (λ_M^* or λ_F^*), it would be possible to arrive at a disease free state, independently of the infectivity of the other population. This situation is not possible in the MF approach. Numerical simulations seem to confirm that such critical values do exist.

The other qualitatively different prediction concerns the effects of the duration of the disease on the spread of it. MF predicts that the epidemic threshold depends only on the quotient of the infectivity and the recovery rate. In other words,

a shorter duration of the disease in any population should be equivalent to a proportionally smaller infectivity. In the PA approximation this does not hold, and consequently, for different recovery rates the critical hyperbolas are different even when the infectivities are rescaled by the recovery rates. This is also seen in numerical simulations, in which the difference seems to be even larger than predicted by the PA.

The differences mentioned also imply that the role that the network plays is more important in the PA. To see this, note that in the MF approach the critical values do not change if men and women are "swapped" in the network. On the other hand, in the PA swapping is not enough, and the network must be rescaled for the situation to be epidemiologically equivalent.

All these features justify the use of the PA to make better and more accurate predictions. Here we have assumed that only the degree distribution of the sexual network is known. However, the PA is best suited for the case when more information is available about the network. Thus, it provides a useful tool to study the spread of epidemics in bipartite assortative networks, which are generally more realistic.

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, Oxford, UK, 1991).

[2] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases* (Wiley, Chichester, 2000).

[3] H. Heesterbeek, in *Ecological Paradigms Lost: Routes of Theory Change*, edited by K. Cuddington and B. Beisner (Elsevier, Amsterdam, 2005).

[4] M. J. Keeling and K. T. D. Eames, J. R. Soc. Interface **2**, 295 (2005).

[5] M. Kretzschmar, Sexually Transmitted Diseases **27**, 627 (2000).

[6] E. Kenah and J. M. Robins, Phys. Rev. E **76**, 036113 (2007).

[7] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).

[8] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes* (Springer, Berlin, 1999).

[9] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[10] J. Gómez-Gardeñes, V. Latora, Y. Moreno, and E. Profumo, Proc. Natl. Acad. Sci. USA **105**, 1399 (2008).

[11] S. Chatterjee and R. Durrett, Ann. Probab. **37**, 2332 (2009).

[12] R. Durrett, Proc. Natl. Acad. Sci. USA **107**, 4491 (2010).

[13] H. Matsuda, N. Tamachi, A. Sasaki, and N. Ogita, in *Mathematical Topics in Biology, Morphogenesis and Neurosciences*, edited by E. Teramoto and M. Yamaguti, Springer Lecture Notes in Biomathematics Vol. 71 (Springer, Berlin, 1987), pp. 154–161.

[14] H. Matsuda, N. Ogita, A. Sasaki, and K. Sato, Prog. Theor. Phys. **88**, 1035 (1992).

[15] S. A. Levin and R. Durrett, Philos. Trans. R. Soc. London B **351**, 1615 (1996).

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1990).

[17] National Centre for Social Research *et al*., National Survey of Sexual Attitudes and Lifestyles II, 2000-2001 [computer file]. Colchester, Essex: UK Data Archive [distributor], August 2005.

[18] A. M. Johnson, C. H. Mercer, B. Erens, A. J. Copas, S. McManus, K. Wellings, K. A. Fenton, C. Korovessis, W. Macdowall, K. Nanchahal, S. Purdon, and J. Field, Lancet **358**, 1835 (2001).

[19] T. W. Smith, Soc. Biol. **39**, 203 (1992).

[20] M. Morris, Nature (London) **365**, 437 (1993).

[21] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).

[22] H. W. Hethcote and J. A. Yorke, *Gonorrhoea: Transmission Dynamics and Control*, Lecture Notes in Biomathematics, Vol. 56 (Springer-Verlag, New York, 1984).

[23] M. Kretzschmar, M. Y. T. H. P. Van Duynhoven, and A. J. Severijnen, Am. J. Epidemiol. **144**, 306 (1997).