Comment on "A BINOMIAL MOMENT APPROXIMATION SCHEME FOR EPIDEMIC SPREADING IN NETWORKS" in U.P.B. Sci. Bull., Series A, Vol. 76, Iss. 2, 23-30, 2014

Istvan Z. Kiss[1,*] & Prapanporn Rattana[1]

July 2, 2014

[1] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK

Abstract

In this short comment we report on our test of the generalisation proposed by Shang in [8]. Shang in [8] claims to generalise previous results developed by Kiss and Simon in [3] and Nagy, Kiss and Simon in [6]. However, our tests show that the proposed generalisation performs poorly for all networks proposed by Shang, except for heterogenous networks with relatively high average degree. While the binomial closure gives good results, in that the solution of the full Kolmogorov equations, with the newly proposed infectious rates, agrees well with the closed system, the agreement with simulation is extremely poor. This disagreement invalidates Shang's generalisation and shows that the newly proposed infectious rates do not reflect the true stochastic process unfolding on the network. We emphasise that our simulations are run on networks which are constructed a priori followed by the simulation of the spreading process on these using a Gillespie-type approach [1, 2], where inter event times are chosen from an exponential distribution with a rate given by the sum of the rates of all possible events, followed by the choice of an event at random but proportionally to its rate. We conclude that the generalisation proposed by Shang [8] is incorrect and that Shang's simulation method and the excellent agreement with the ODE models is based on flawed or incorrectly implemented simulations. To support this statement, we also validate our simulation results by using the well-known pairwise and effective degree models.

Keywords: SIS epidemic; networks; Kolmogorov equations, pairwise model; effective-degree model

∗ corresponding author
email: i.z.kiss@sussex.ac.uk

1 Introduction

Kiss and Simon in [3] considered the susceptible-infected-suceptible (SIS)dynamics on a fully connected network with N nodes. The model was formulated in terms of the master equation given by

$$p_k(t) = a_{k-1}p_{k-1}(t) - (a_k + c_k)p_k(t) + c_{k+1}p_{k+1}(t), \tag{KE}$$

where $p_k(t)$ is the probability that there are k infectious nodes at time $t \geq 0$, with $k = 0, 1, 2, \ldots, N$. Furthermore, the rates of infection, a_k, and rates or recoveries, c_k, are given by

$$a_k = \tau k(N-k), \quad c_k = \gamma k \quad \text{for} \quad k = 0, 1, \ldots, N \quad \text{with} \quad a_{-1} = c_{N+1} = 0.$$

All infection and recovery processes are modelled as independent Poisson processes. The infection rates encode all the information about the network, and the rate of recovery is simply a rate corresponding to pooled Poisson processes. Kiss and Simon in [3] show that rather than solving this full system, it is possible to derive a low-dimensional ODE based on the assumption that the number of infectious nodes is binomially distributed. Namely, it is assumed that $p_k(t)$ is distributed binomially, i.e. $\mathcal{B}(n, p)$, where n and p depend on time.

More precisely, the low-dimensional ODE is formulated for the first moment of the distribution, and this will also involve the second moment and the third. However, due to the assumption that $p_k(t)$ is binomially distributed, it is possible to express the third moment in terms of the first and second. This then yields an ODE system with 2 equations only. We briefly focus on deriving equations for the moments. Namely, for

$$y_j(t) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^j p_k(t) \quad \text{or} \quad Y_j(t) = \sum_{k=0}^{N} k^j p_k(t), \tag{1}$$

where $N^j y_j = Y_j$ with $j = 1, 2, \ldots$. Deriving evolution equations for these is straightforward. For example, the derivative of the first moment, and in a similar way for all other moments, can be given in function of higher-order moments upon using the Kolmogorov equations, Eq. (KE). The derivation for the first moment is outlined below,

$$\begin{aligned} \dot{Y}_1(t) &= \sum_{k=0}^{N} k\dot{p}_k = \sum_{k=0}^{N} k(a_{k-1}p_{k-1} - (a_k + c_k)p_k + c_{k+1}p_{k+1}) \\ &= \sum_{k=0}^{N} (ka_{k-1}p_{k-1} - ka_kp_k + kc_kp_k + kc_{k+1}p_{k+1}). \end{aligned}$$

By changing the indices of the summation, plugging in the corresponding expressions for the transition rates a_k and c_k, and taking into account that $a_{-1} = c_{N+1} = 0$ the following expression holds,

$$\dot{Y}_1(t) = \sum_{k=0}^{N} (\tau(k + k^2)(N-k) - \tau k^2(n-k) - k^2\gamma + (k^2 - k)\gamma)p_k.$$

Based on our notations, see Eq. (1), the equation above reduces to

$$\dot{Y}_1(t) = \tau N Y_1 - \tau Y_2 - \gamma Y_1. \tag{2}$$

We emphasise that this was possible due to the special form of the a_k coefficients, namely that these are quadratic polynomials in k. Using a similar procedure, the equation for the second moment Y_2 can be easily computed and is given by

$$\dot{Y_2} = 2(\tau N - \gamma)Y_2 - 2\tau Y_3 + (\tau N + \gamma)Y_1 - \tau Y_2. \tag{3}$$

Equations (2) & (3) can be recast in terms of the density dependent moments y_js to give

$$\begin{aligned} \dot{y}_1 &= (\tau N - \gamma)y_1 - y_2, & (4)\\ \dot{y}_2 &= 2(\tau N - \gamma)y_2 - 2\tau y_3 + \frac{1}{N}\left((\tau N + \gamma)y_1 - \tau N y_2\right). & (5) \end{aligned}$$

The above equations are not closed or self-contained since the second moment depends on the third and an equation for this is also needed. It is easy to see that this dependence of the moments on higher moments leads to an infinite but countable number of equations. Hence, a closure is needed and below we show that it is possible to express Y_3 as a function of Y_1 and Y_2. The first three moments of the binomial distribution can be specified easily in terms of the two parameters and are as follows,

$$\begin{aligned} Y_1 &= np & (6)\\ Y_2 &= np + n(n-1)p^2 & (7)\\ Y_3 &= np + 3n(n-1)p^2 + n(n-1)(n-2)p^3. & (8) \end{aligned}$$

Using Eqs. (6) & (7), n and p can be expressed in term of Y_1 and Y_2 as follows,

$$p = 1 + Y_1 - \frac{Y_2}{Y_1}, \qquad n = \frac{Y_1^2}{Y_1 + Y_1^2 - Y_2}. \tag{9}$$

Plugging the expressions for p and n, Eq. (9), into Eq. (8), the closure for the third moment is found to be

$$Y_3 = \frac{2Y_2^2}{Y_1} - Y_2 - Y_1(Y_2 - Y_1).$$

This relation defines the new closure, and in terms of the density dependent moments this is equivalent to

$$y_3 = \frac{2y_2^2}{y_1} - y_1 y_2 + \frac{1}{N}(y_1^2 - y_2).$$

Using the equation for the first moment, Eq. (4), the closure at the level of the second moment yields the following approximate equation

$$\dot{x}_1 = (\tau N - \gamma)x_1 - \tau N x_1^2.$$

Using the equations for the first two moments, Eqs. (4) & (5), and the closure at the level of the third moment yields

$$\begin{aligned} \dot{x_1} &= (\tau N - \gamma)x_1 - \tau N x_2, \\ \dot{x_2} &= 2(\tau N - \gamma)x_2 - 2\tau N x_3 + \left(\left(\tau + \frac{\gamma}{N}\right)x_1 - \tau x_2\right), \end{aligned}$$

where

$$x_3 = \frac{2x_2^2}{x_1} - x_1 x_2 + \frac{1}{N}(x_1^2 - x_2).$$

Hence, we have derived two approximate systems, with the first and second closed at the level of the second and third moment, respectively. It is in general true that the higher the moment at which the closure the more likely that the resulting approximate model performs well. We note that we used x instead of y to highlight that the closed systems, defined in term of x, are only an approximation to the exact system given in terms of y.

The major challenge is generalising this to arbitrary networks is in finding a correct functional form for the infection rates a_k for any network in general. Kiss and Simon [3] have shown that for homogenous random networks and based on the random mixing argument a_k can be written as

$$a_k = \tau(N-k)\langle k\rangle \frac{k}{N-1},$$

where it is assumed that infectious nodes are distributed at random around susceptible nodes. Our numerical experiments also show that such a formula also performs well for Erdős-Rényi random networks. For other graphs no such immediate or intuitive formula exists.

Shang in [8] proposed that a_k in general could be written as

$$a_k = \frac{\tau k(N-k)\langle k^2\rangle}{\langle k\rangle(N-1)}, \quad c_k = \gamma k \quad \text{for} \quad k = 0, 1, \ldots, N \quad \text{with} \quad a_{-1} = c_{N+1} = 0, \tag{10}$$

where the network is given in terms of a degree distribution with $P(k)$ denoting the probability that a randomly chosen node has degree k, with $k = 0, 1, 2, \ldots, N-1$ for a network of size N. Moreover $\langle k\rangle = \sum kP(k)$ and $\langle k^2\rangle = \sum k^2 P(k)$. While there is not explicit explanation for this, we can heuristically explain how such a formula could be arrived at. A newly infected node, under the assumption of random mixing will have degree l with probability $lP(l)/\langle k\rangle$. Hence, such a node has l onward connections and one such links leads to a susceptible node with probability $(N-k)/(N-1)$. Putting this together for a single node and averaging across all degrees gives

$$\sum_l \frac{lP(l)}{\langle k\rangle} \times l \times \frac{N-k}{N-1},$$

and upon multiplying this with k, the number of infectious nodes, yields

$$a_k = \frac{\tau k(N-k)\langle k^2\rangle}{\langle k\rangle(N-1)}.$$

Shang then used the same procedure as above to derive a set of 2 ODEs for these potentially more general infection terms. His closed system yields

$$\dot{x}_1(t) = \left(\frac{\tau\langle k^2\rangle N}{\langle k\rangle(N-1)} - \gamma\right) x_1 - \frac{\tau\langle k^2\rangle N}{\langle k\rangle(N-1)} x_2, \tag{11}$$

$$\begin{aligned}\dot{x}_2(t) &= \left(\frac{\tau\langle k^2\rangle(2N-1)}{\langle k\rangle(N-1)} - 2\gamma\right) x_2 - \frac{2\tau\langle k^2\rangle N}{\langle k\rangle(N-1)} x_3 \\ &+ \left(\frac{\tau\langle k^2\rangle}{\langle k\rangle(N-1)} + \frac{\gamma}{N}\right) x_1,\end{aligned} \tag{12}$$

where the same closure applies, namely

$$x_3 = \frac{2x_2^2}{x_1} - x_1 x_2 + \frac{1}{N}(x_1^2 - x_2).$$

2 Testing Shang's generalisation

To carry out our tests we used the same networks and parameters as give in Shang's paper [8]. We note that some of these choices are not natural, as the proposed network have a very low average degree, which in general makes it very difficult to obtain good mean-field like approximation for stochastic processes unfolding on sparse networks.

Table 1: Network models with degrees in the range $1 \leq k \leq 20$ for the truncated power laws and $k \in \{0, 1, 2, \ldots\}$ for the networks with Poisson degree distributions.

Network	Degree distribution	$\langle k\rangle$	$\langle k^2\rangle$
Homogenous/regular	$P(4) = 1$	4	16
Bimodal	$P(2) = P(4) = 0.5$	3	10
Poisson	$P(k) = \langle k\rangle^k \frac{e^{-\langle k\rangle}}{k!}$	10	110
Truncated power law (a)	$P(k) = 0.673k^{-2}e^{-k/30}$	2.0406	9.6613
Truncated power law (b)	$P(20-k) = 0.673k^{-2}e^{-k/30}$	17.9635	328.1197

2.1 Full versus reduced/closed ODEs

Here we show that solving the master equations, Eq. (KE), directly with the more general infection terms, Eq. (10), gives good agreement with the solution of the closed/reduced system, Eqs. (11-12). In Fig. 1, we show that for a range of parameter values the agreement is excellent, and in line with what Shang found in [8], which simply means that the assumption of a binomial distribution for the number of infected individuals at a given time is a valid approximation. However, it does neither confirm nor invalidates the appropriateness of the choice of the new infection rates a_k, as proposed by Shang in [8]. Their appropriateness is tested via comparing the output from the master and / or reduced equations to the average of stochastic simulations and this is what we test next.

2.2 Comparison of Shang's generalisation to simulation

We first generate networks with the prescribed degree distribution by using the configuration method. This is followed by implementing the epidemic as a continuous-time Markov Chain on these networks. This is done by using a Gillespie-type approach [1, 2]. In this case, inter event times are chosen from an exponential distribution with a rate given by the sum of the rates of all possible events, followed by the choice of an event at random but proportionally to its rate.

We now move on to the crucial comparison of output based on the closed system to results from explicit stochastic network simulations. First, we validate our own simulations for the range of networks suggested by Shang in [8], see Table 1 for a summary. We use the pairwise [7], see Appendix 5.1, and effective degree models [4], see Appendix 5.2, and as shown in Figs. 2 and 3, the agreement with our simulations is excellent. As pointed out before, the small disagreements are due to the very small average degree of the networks used in [8]. A small average degree is well-known to make the approximation with mean-field type models difficult. The same figures show that the agreement improves as the average degree increases, see the case of networks with homogenous and heterogeneous degree distributions with $\langle k \rangle = k = 4$ and $\langle k \rangle = 10$, respectively.

In Figs. 4 and 5, we plot the prevalence based on Shang's closed model, Eqs. (11-12), versus that from simulations. These plots show clearly that the agreement is poor, except for heterogenous networks with relatively large average degree and for networks with the inverted truncated power law distribution with very high degree as shown in Fig. 5. Our tests significantly differ from Shang's results and we infer that Shang's simulation method, which is not described in [8], is flawed or incorrectly implemented. We point out that the results concerning the full master equation and its reduction are correct and we were able to reproduce these. However, this alone neither leads to nor guarantees agreement with results based on simulations. In all our tests, and in line with Shang's work, we also attempted to time shift the prevalence, see the right panel in Fig. 4, but this did not lead to better agreement. Moreover, a close visual inspection shows clearly that there are fundamental differences between Shang's closed model and simulation results and that no amount of time shifting will lead to a better agreement. For example, the equilibrium prevalence is very different and this again is in stark disagreement with Shang's results.

3 Discussion

It is our view that identifying general infectious terms a_k remains a major challenge as this is highly dependent on the structure of the network, parameters of the disease dynamics, and more importantly on the correlations that build up during the spreading process. It is unfortunate that this generalisation does not work and, as we shown in [6], it is possible to try to derive semi-analytical or numerical approximations for the infection rates. We conclude that Shang's simulation method is flawed and that Shang's generalisation is not valid. We look forward to any clarifications.

4 Acknowledgements

P. Rattana acknowledges funding for her Ph.D. studies from the Ministry of Science and Technology, Thailand.

5 Appendix

5.1 Appendix A: Compact pairwise model

House and Keeling [7] have successfully extended the general pairwise model of Eames and Kelling [5] to heterogeneous networks and for both SIR and SIS models. Here, we focus on the SIS model. The original pairwise equations of Eames and Kelling [5] are,

$$
\begin{aligned}
&[\dot{S_k}] = \gamma[I_k] - \tau[S_k I], \\
&[\dot{I_k}] = \tau[S_k I] - \gamma[I_k], \\
&[\dot{S_k S_l}] = -\tau \textstyle\sum_m ([S_k S_l I_m] + [I_m S_k S_l]) + \gamma([S_k I_l] + [I_l S_k]), \\
&[\dot{S_k I_l}] = \tau \textstyle\sum_m ([S_k S_l I_m] - [I_m S_k I_l]) - \tau[S_k I_l] - \gamma[S_k I_l] + \gamma[I_k I_l], \\
&[\dot{I_k I_l}] = \tau \textstyle\sum_m ([I_m S_k I_l] + [I_m S_l I_k]) + \tau([S_k I_l] + [S_l I_k]) - 2\gamma[I_k I_l],
\end{aligned}
\tag{13}
$$

where $[A_k]$ stands for the expected number of nodes of degree k across the whole network in state A, $[A_k B_l]$ represents the number of links of type $A - B$ when A has degree k and B has degree l, $[A_k B] = \sum_l [A_k B_l]$, τ is the transmission rate and γ is the recovery rate.

Then, they used the following more compact closure

$$
[A_k B] \approx [AB] \frac{k[A_k]}{\sum_l l[A_l]}.
$$

Using this in the standard pairwise SIS model, Eq. (13), the reduced/compact pairwise models is given by:

$$
\begin{aligned}
&[\dot{S_k}] = \gamma([k] - [S_k]) - \tau[SI] \frac{k[S_k]}{\sum_l l[S_l]}, \\
&[\dot{SI}] = \tau[SI]\Big(\sum_k k[S_k] - 2[SI]\Big) \frac{\sum_l l(l-1)[S_l]}{(\sum_m m[S_m])^2} - (\tau + \gamma)[SI] \\
&\qquad\qquad + \gamma\Big(\sum_k k([k] - [S_k]) - [SI]\Big),
\end{aligned}
$$

where $[k]$ is the number of nodes of degree k.

5.2 Appendix B: Effective degree model

Lindquist et al. [4] formulated the SIS mean-field model base on the effective degree approach. This model is based on keeping track of the expected number of susceptible and infected nodes with all possible neighbourhood combinations, S_{si}

and I_{si}, respectively. S_{si} represents the expected number of susceptible nodes that have s connections to other susceptible nodes and i connections to infected nodes, with similar argument for I_{si}.

Accounting for all possible transitions, the equations as formulated by Lindquist et al. [4] are:

$$\acute{S_{si}} = -\tau i S_{si} + \gamma I_{si} + \gamma\Big[(i+1)S_{s-1,i+1} - iS_{si}\Big]$$

$$+\frac{\sum_{k=1}^{M}\sum_{j+l=k}\tau jlS_{jl}}{\sum_{k=1}^{M}\sum_{j+l=k}jS_{jl}}\Big[(s+1)S_{s+1,i-1} - sS_{si}\Big],$$

$$\acute{I_{si}} = \tau i S_{si} - \gamma I_{si} + \gamma\Big[(i+1)I_{s-1,i+1} - iI_{si}\Big]$$

$$+\frac{\sum_{k=1}^{M}\sum_{j+l=k}\tau l^2 S_{jl}}{\sum_{k=1}^{M}\sum_{j+l=k}jI_{jl}}\Big[(s+1)I_{s+1,i-1} - sI_{si}\Big],$$

for $\{(s,i) : s \geq 0, i \geq 0, s+i \leq M\}$, where M is the maximum node degree in the network.

References

[1] D. T. Gillespie (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403-434.

[2] D. T. Gillespie (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340-2361.

[3] I.Z. Kiss and P.L. Simon (2012) New moment closures based on a priori distributions with applications to epidemic dynamics. *Bull. Math. Biol.* **74**, 1501-1515.

[4] J. Lindquist, J. Ma, P. van den Driessche and F. H. Willeboordse (2011) Effective degree network disease models. *J. Math. Biol.* **62**, 143-164.

[5] K.T.D. Eames and M.J. Keeling. (2002) Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci. USA.* **99**, 13330–13335.

[6] N. Nagy, I.Z. Kiss and P.L. Simon (2014) Approximate master equations for dynamical processes on graphs. *Math. Model. Nat. Phenom.* **9**, 43-57.

[7] T. House and M. J. Keeling (2011) Insights from unifying modern approximations to infections on networks. *J. Roy. Soc. Interface* **8**, 67-73.

[8] Y. Shang (2014) A binomial moment approximation scheme for epidemic spreading in networks. *U.P.B. Sci. Bull. Series A* **76**, 23-30.

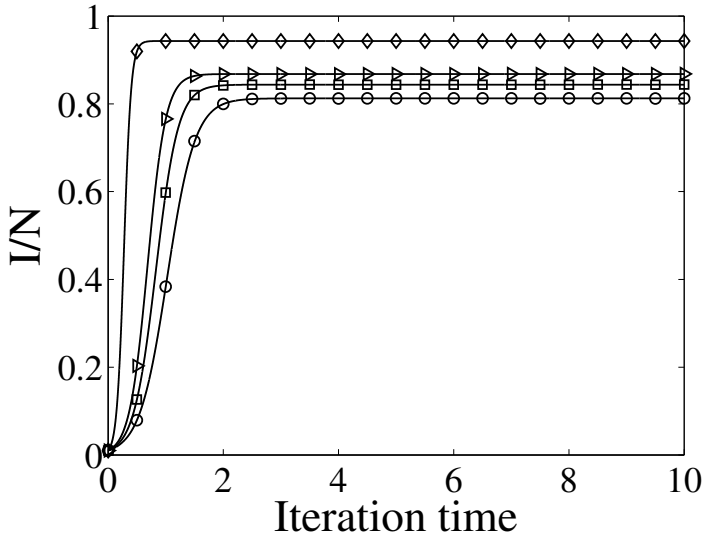

Figure 1: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. Continuous lines represent the solution of the full equations, see Eq. (KE), while the solution of reduced model is given by Eqs. (11-12) for ($\square$) - homogeneous distribution $P(4) = 1$, ($\circ$) - bimodal distribution $P(2) = P(4) = 0.5$, ($\diamond$) - Poisson distribution with $\langle k \rangle = 10$, and ($\triangleright$) - truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$. For all cases there is excellent agreement between the full and reduced equations.

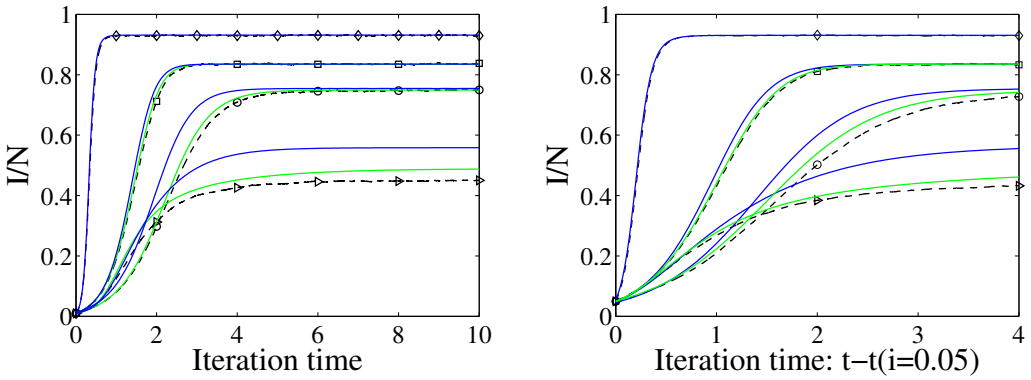

Figure 2: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these: homogeneous distribution $P(4) = 1$ ($\square$), bimodal distribution $P(2) = P(4) = 0.5$ ($\circ$), Poisson distribution with $\langle k \rangle = 10$ ($\diamond$) and truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$), (simulation: black dashed line, effective degree model: green line, compact pairwise model: blue line). We note that the effective degree model has not been implemented for networks with Poisson distribution due to the degrees being theoretically unbounded.

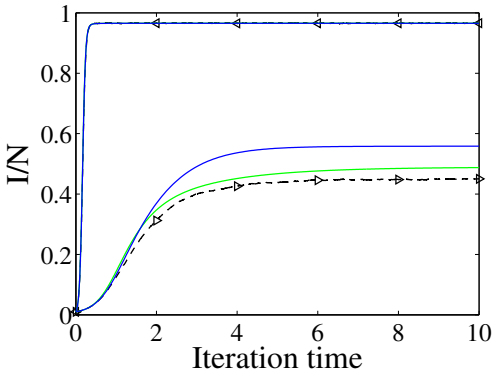

Figure 3: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. The networks have truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$) and degree inverted distribution ($\triangleleft$), i.e. $P(20-k) = 0.673k^{-2}\exp(-k/30)$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these, (simulation: black dashed line, effective degree model: gree line, compact pairwise model: blue line).

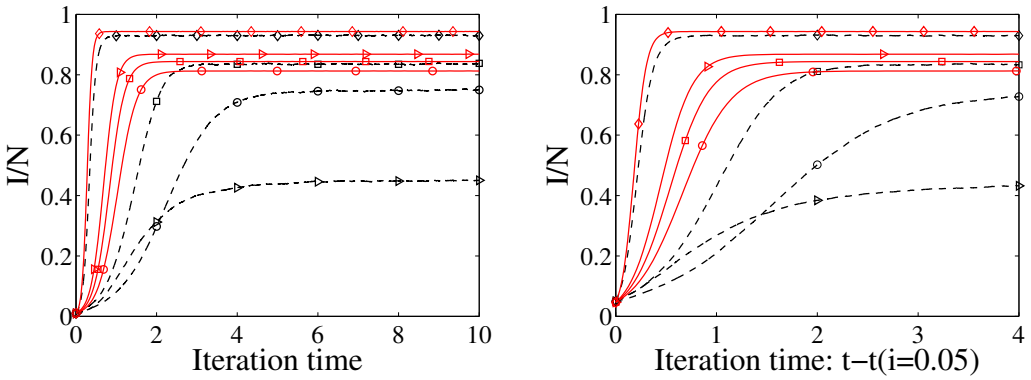

Figure 4: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these: homogeneous distribution $P(4) = 1$ ($\square$), bimodal distribution $P(2) = P(4) = 0.5$ ($\circ$), Poisson distribution with $\langle k \rangle = 10$ ($\diamond$) and truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$). Simulations are black dashed lines and results based on Shang's model, see Eqs. (11-12), are given by the red lines.

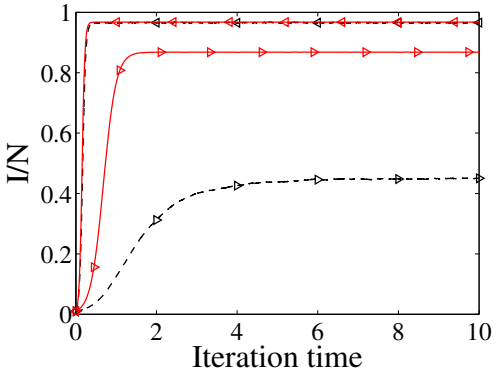

Figure 5: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. The networks have truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$) and degree inverted distribution ($\triangleleft$), i.e. $P(20-k) = 0.673k^{-2}\exp(-k/30)$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these. Simulations are black dashed lines and results based on Shang's model, see Eqs. (11-12), are given by red lines.