

SCIENTIFIC REPORTS



OPEN

The Stochastic Transport Dynamics of a Conserved Quantity on a Complex Network

Pablo Medina^{1,2,3}, Jaime Clark^{1,3}, Miguel Kiwi^{1,3}, Felipe Torres^{1,3}, José Rogan^{1,3} & Juan Alejandro Valdivia^{1,3}

The stochastic dynamics of conserved quantities is an emergent phenomena observed in many complex systems, ranging from social and to biological networks. Using an extension of the Ehrenfest urn model on a complex network, over which a conserved quantity is transported in a random fashion, we study the dynamics of many elementary packets transported through the network by means of a master equation approach and compare with the mean field approximation and stochastic simulations. By use of the mean field theory, it is possible to compute an approximation to the ensemble average evolution of the number of packets in each node which, in the thermodynamic limit, agrees quite well with the results of the master equation. However, the master equation gives a more complete description of the stochastic system and provides a probabilistic view of the occupation number at each node. Of particular relevance is the standard deviation of the occupation number at each node, which is not uniform for a complex network. We analyze and compare different network topologies (*small world*, *scale free*, *Erdos-Renyi*, among others). Given the computational complexity of directly evaluating the asymptotic, or equilibrium, occupation number probability distribution, we propose a scaling relation with the number of packets in the network, that allows to construct the asymptotic probability distributions from the network with one packet. The approximation, which relies on the same matrix found in the mean field approach, becomes increasingly more accurate for a large number of packets.

One of the most interesting and fundamental problem in physics is related to the understanding of how the reversible microphysics gives rise to irreversible thermodynamics. An important model that has contributed to this comprehension is the “Ehrenfest urn”¹, that was proposed in 1907 and solved exactly^{2,3} in 1947. In this model, we have N marbles, or packets, that move randomly and in a conserved manner between two urns, so that at each time step a packet is selected at random and changed from the original to the other urn.

Here we elaborate on a generalization of the Ehrenfest model by Clark *et al.*⁴, in which a number of urns are interconnected as a complex network, and the marbles or packets can only jump to urns to which the first one has a directed connection in a conservative manner.

In this respect, there has been a large amount of published research about complex networks in areas as diverse as physics, biology, and social sciences^{5–8}. Originally, the interest was on the topology of the networks, such as their characterization in terms of their connectivity distribution $P(k)$. For example, the study of *scale-free networks*, which follow a power law $P(k) \sim k^{-\alpha}$ for large values of k , became very popular^{9–18}. Models based on preferential attachment seem to explain a diversity of power law exponents^{5,19–21}. Another characterization of such networks involves the distribution of the shortest distance between nodes, and a noteworthy example are *small-world networks*²² that have short average distances. More recently, researches have began to study the topological evolution of these networks^{7,8}, or as dynamical systems over which a certain quantity is transported^{9,23–37}.

The generalization of the “Ehrenfest urn” to a complex network, as proposed by Clark *et al.*⁴, is an example of the nontrivial transport that can occur on a complex network^{38–41}, and should have similar properties to traffic in cities, electric networks, etc.

¹Departamento de Física, Facultad de Ciencias, Universidad de Chile, Santiago de Chile, Chile. ²Facultad de Física, Pontificia Universidad Católica de Chile, Santiago de Chile, Chile. ³Centro para el Desarrollo de la Nanociencia y Nanotecnología, Santiago, Chile. Pablo Medina, Jaime Clark, Miguel Kiwi, Felipe Torres, Jose Rogan and Juan Alejandro Valdivia contributed equally. Correspondence and requests for materials should be addressed to P.M. (email: pab-medi@uniandes.edu.co)

In this manuscript we construct the master equation that describes the evolution of the occupation number probability at each urn, and then compare the results with a stochastic simulation of the network of urns and the mean field approach proposed by⁴. The mean field theory approximation to the ensemble average evolution of the number of packets in each node agrees quite well with the results of the master equation, particularly in the thermodynamic limit. However, the master equation gives a more complete description of the stochastic system, and provides a probabilistic view of the occupation number on each node. Of particular relevance is the standard deviation of the occupation number at each node, which is not uniform for a complex network, and therefore provides an intriguing result from a statistical mechanics point of view.

Given the computational complexity of directly evaluating the asymptotic, or equilibrium, occupation number probability distribution; we propose a scaling relation with the number of packets in the network that allows to construct the asymptotic probability distributions from the network with one packet. Interestingly enough, the scaling approach requires the same matrix that is constructed for the mean field approach. We will notice that the approximation becomes increasingly more accurate as the number of packets becomes large.

Results

The model. The “Ehrenfest urn” over a complex network, as generalized by Clark *et al.*⁴, describes the transport of N packets between the M nodes of a directed network. At a given time t a random packet, which is at a node i , is chosen to move to one of the k_i nodes to which node i is connected, i.e., in its *outgoing set*. Similarly, the *incoming set* of node i corresponds to the nodes that connect to node i . Hence, the dynamics of the packets is conserved, so that at a given time we have $x_i(t)$ packets at the i^{th} node, with the restriction

$$N = \sum_{i=1}^M x_i(t). \quad (1)$$

In Fig. 1(a) we show a 3-node network that is represented by the adjacency matrix A

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (2)$$

Here $A_{ij} = 1$ if there is a directed connection from node i to node j , and 0 otherwise. The size of the *outgoing set* of the i^{th} node is then $k_i = \sum_j A_{ij}$. A stochastic simulation of the packet transport for $N = 100$ is plotted in Fig. 1(b), which shows that the system relaxes to an asymptotic state that is not uniform, e.g., the middle node has on average twice as many packets than the other 2 nodes. This results may have implications in many fields such as traffic in cities, lines in supermarkets, etc. Of course, we also see fluctuations around the mean asymptotic solution. This is much clearer in Fig. 1c where we display 10 simulations for the same network. Below, we find a way to describe these fluctuations, and in fact the whole probability distribution, with the help of a master equation.

Following the mean field approach proposed by Clark *et al.*⁴ one assumes that evaluating an ensemble average evolution $\langle m_i(t) \rangle$ of $x_i(t)$, in the thermodynamic limit, is equivalent to assume that all the N packets move to a new node in a time N , so that the evolution equation for the ensemble average is

$$\langle m_i(t+1) \rangle - \langle m_i(t) \rangle \approx \frac{1}{N} \left(-\langle m_i(t) \rangle + \sum_{j \neq i}^N \frac{A_{ji}}{k_j} \langle m_j(t) \rangle \right). \quad (3)$$

The first term on the right represents the transport of the $\langle m_i \rangle$ packets to the *outgoing set* of the i^{th} node. The second term represents the packets that get transported to the i^{th} node from all the nodes that have the i^{th} node in their *outgoing sets*. Of course, this is properly normalized by the size of each of the *outgoing sets*.

For large N we can approximate this expression by a time derivative, so that we can write it in vectorial form as

$$\frac{d}{dt} \langle \vec{m}(t) \rangle = \frac{1}{N} \mathbf{B} \langle \vec{m}(t) \rangle, \quad (4)$$

where $\langle \vec{m}(t) \rangle \rightarrow \{\langle m_1(t) \rangle, \langle m_2(t) \rangle, \dots, \langle m_M(t) \rangle\}$, and \mathbf{B} is the dynamical matrix whose elements are

$$B_{ij} = -\delta_{ij} + \frac{A_{ji}}{k_j}, \quad (5)$$

where δ_{ij} is the Kronecker delta. Hence, given an initial condition, we can evaluate $\langle \vec{m}(t) \rangle$ by integrating the above equation, to obtain

$$\langle \vec{m}(t) \rangle = e^{\mathbf{B}t/N} \langle \vec{m}(0) \rangle \quad (6)$$

which can be obtained by diagonalizing the dynamical matrix \mathbf{B} , such that

$$\mathbf{B} = \mathbf{V}^{-1} \mathbf{\Lambda} \mathbf{V}, \quad (7)$$

and writing

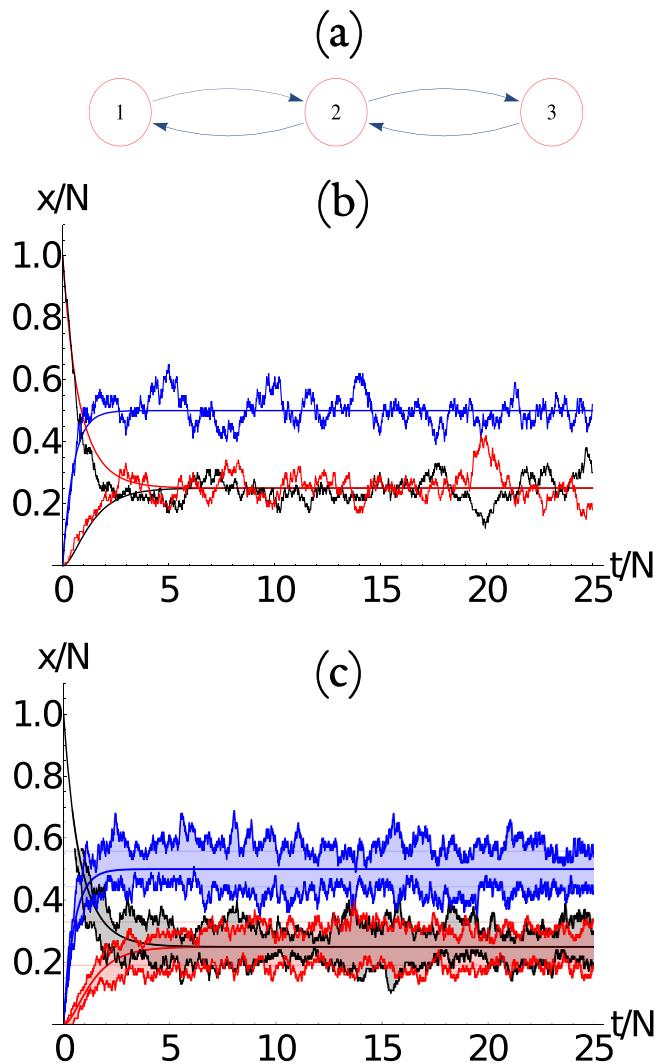


Figure 1. (a) Three-node network, in which the arrows determine the directed connectivity of the network. (b) Time evolution of the number of packets at each node for a stochastic simulation using $N = 100$. We show the evolution for node 1 (red), 2 (blue), and 3 (black). The initial condition is $m_1(0) = N$, $m_2(0) = m_3(0) = 0$. (c) The time evolution of the number of packets at each node for 10 stochastic simulations (similar to b), showing that the system evolves to an asymptotic state that presents fluctuations around a mean value (thick color lines) that can be obtained from the $\lambda_0 = 0$ eigenvector of the B matrix as described in the text. The standard deviation around the mean (thin color lines) for each node is constructed analytically from the master equation as described in the text. The thick color lines correspond to the maximum and minimum value of the 10 simulations for each node at each time.

$$\langle \vec{m}(t) \rangle = \mathbf{V}^{-1} e^{\Lambda t/N} \mathbf{V} \langle \vec{m}(0) \rangle, \tag{8}$$

where Λ is the diagonal matrix of the eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$ of \mathbf{B} , and \mathbf{V} is the matrix of column eigenvectors. Therefore, the time evolution of the particle number is dominated by the smallest nonzero eigenvalue $\tilde{\lambda}$, namely

$$\langle m_i(t) \rangle = e^{-|\Re[\tilde{\lambda}]|t/N} \langle m_i(0) \rangle, \tag{9}$$

where $|\Re[\tilde{\lambda}]|$ is the absolute value of the real part of the $\tilde{\lambda}$. The eigenvalues of the matrix \mathbf{B} give the timescales of the system, as studied in detail in ref.⁴ for different types of complex networks. Also from the $\lambda = 0$ eigenvector ($t \rightarrow \infty$), we can evaluate the asymptotic ensemble average occupation number. Such eigenstate exists because

$$\sum_{i=1}^M \langle m_i(t) \rangle = N.$$

The eigenvalues for matrix \mathbf{B} of the network of Fig. 1(a) are $\lambda_1 = -2$, $\lambda_2 = -1$, and $\lambda_0 = 0$, hence the relaxation time of the system is $\tau = N$. The evolution of the dynamics is shown as the continuous lines of Fig. 1(b). The asymptotic state can be recovered from the $\lambda_0 = 0$ eigenvector $\tilde{v} = \{1, 2, 1\}/5$, so that the asymptotic state is $\langle m_1 \rangle = \langle m_3 \rangle = N/4$ and $\langle m_2 \rangle = N/2$. The existence of such nonuniform asymptotic states is an interesting results in view of what we know about equilibrium statistical mechanics.

It is interesting to note that the $M = 2$ solution corresponds to the original ‘‘Ehrenfest urn’’ solution constructed by Marc Kac^{2,3}. The mean field approximation improves as we increase N , however we cannot evaluate the fluctuations within this approximation, and we need to resort to a master equation for the probability of occupation.

Let us notice that the mean field evolution equation can be cast into a rate equation for the particle number variation, given by

$$\frac{d}{dt} \langle m_i(t) \rangle = \sum_j \left(p_{j,i} \langle m_j(t) \rangle - p_{i,j} \langle m_i(t) \rangle \right), \tag{10}$$

where $p_{i,j}$ is the transition probability of one packet from the i -th node to the j -th node. Notice that the conservation of the particle number can be obtained directly from Eq. (10); indeed

$$\frac{d}{dt} \sum_i \langle m_i(t) \rangle = \sum_{i,j} \left(p_{j,i} \langle m_j(t) \rangle - p_{i,j} \langle m_i(t) \rangle \right) = 0. \tag{11}$$

Replacing $p_{i,j} = A_{i,j}/(Nk_i)$ into the Eq. (10), we obtain

$$\begin{aligned} \frac{d}{dt} \langle m_i(t) \rangle &= \frac{1}{N} \left(\sum_j \frac{A_{j,i}}{k_j} \langle m_j(t) \rangle - \sum_j \frac{A_{i,j}}{k_i} \langle m_i(t) \rangle \right), \\ &= \frac{1}{N} \left(\sum_j \frac{A_{j,i}}{k_j} \langle m_j(t) \rangle - \langle m_i(t) \rangle \right) \end{aligned} \tag{12}$$

which is equivalent to the mean field equation, since $\sum_j A_{i,j} = k_i$ in the last term.

Close to the steady state condition, Eq. (12) can be written as

$$\langle m_i(\tau) \rangle = \sum_j \frac{A_{j,i}}{k_j} \langle m_j(\tau) \rangle. \tag{13}$$

Therefore, for an undirected network (*outgoing set* is equal to *incoming set* of each node), a solution can be written as $\langle m_i \rangle = Ck_i$ with $C = N/\sum_j k_j$ so that we satisfy $\sum_i \langle m_i(\tau) \rangle = N$. Topologically, the connectivity distribution determines the asymptotic state for the mean occupation number at each node $\langle m_i(t \rightarrow \infty) \rangle$. Therefore, for a *scale free* network we obtain a power law distribution for the asymptotic mean occupation number. However, for a directed network the analysis is not that trivial, and there seems to be no simple connection to a topological property of the network. We plan to analyze this in detail in a future manuscript.

The Master Equation. We now construct the Master equation, that describes the evolution of the probability of occupation at each of the nodes. We start by defining the vector $\vec{n} = [n_1, n_2, \dots, n_M]$ that represents a given occupation of the nodes of the system, which has probability $P(\vec{n}, t)$ to occur at the iteration t . The convention is that $0 \leq n_i \leq N$ for all $i = 1, \dots, M$, and satisfy the restriction

$$N = \sum_{i=1}^M n_i.$$

We also consider all vectors of the type $\vec{\Omega}_{i,j} = [0, 0, \dots, n_i = 1, \dots, n_j = -1, \dots]$, whose components are equal to 0, except for $n_i = +1$ and $n_j = -1$. Using this definition, we can write the evolution of the probabilities as

$$P(\vec{n}, t + 1) = \sum_{\vec{\Omega}_{i,j}} \frac{(n_i + 1)A_{i,j}}{Nk_i} P(\vec{n} + \vec{\Omega}_{i,j}, t). \tag{14}$$

It is easy to show that the evolution equations satisfy probability conservation

$$\sum_{\vec{n}} P(\vec{n}, t + 1) = \sum_{\vec{n}} P(\vec{n}, t). \tag{15}$$

Once we know $P(\vec{n}, t)$, we can compute the expectation value

$$\langle n_i(t) \rangle = \sum_{\vec{n}} n_i P(\vec{n}, t), \tag{16}$$

the covariance matrix

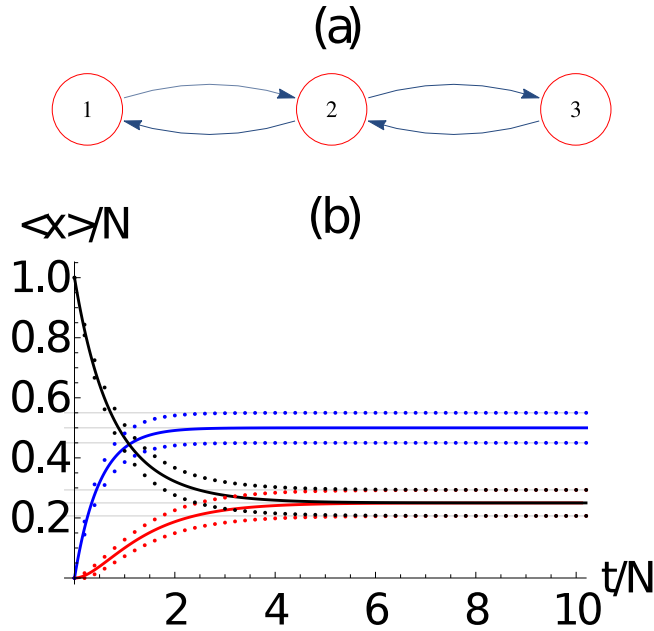


Figure 2. (a) Three-node network, in which the arrows determine the directed connectivity of the network. (b) Evaluation of the ensemble average number of packets at each node $\langle n_i(t) \rangle$ (continuous line) and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ (dotted curves), as a function of time. The horizontal dotted lines are the corresponding values obtained from the asymptotic solution. Here we use $N = 100$ packets, which corresponds to $N_{\text{eq}} = 1326$ coupled equations. We also observe the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field approach. There is an excellent agreement between the master equation and the mean field approach. Here $i = 1$ (red), 2 (blue), 3 (black).

$$\langle \sigma_{ij}^2(t) \rangle = \sum_{\vec{n}} (n_i - \langle n_i(t) \rangle)(n_j - \langle n_j(t) \rangle) P(\vec{n}, t) \tag{17}$$

and the occupation probability of a given node

$$P(n_i, t) = \sum_{[n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_M]} P(\vec{n}, t). \tag{18}$$

The steady state asymptotic solution can be defined as $P_e(\vec{n}) = P(\vec{n}, t + 1) = P(\vec{n}, t)$ for all \vec{n} . The number of equations for a given value of N and M is

$$N_{\text{eq}} = \frac{1}{M!} \prod_{i=1}^M (N + i), \tag{19}$$

so that in general the number of equations required to find the asymptotic solution grows very quickly as N^M , making it increasingly difficult to evolve large systems.

We show in Fig. 2(b) the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$, as a function of time, calculated at each time step from the evolution of the master equation. We have considered the network of Fig. 2(a) with $N = 50$ packets, which corresponds to $N_{\text{eq}} = 1326$ coupled equations. We compare the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field approach. We observe that the equivalent values obtained from the asymptotic solution of the master equation, as we will see below, can also be obtained exactly as the $\lambda_0 = 0$ eigenvector of the dynamical matrix \mathbf{B} .

We now study the network of Fig. 3(a), which displays an interesting dynamics. Figure 3(b) shows the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ as a function of time from the evolution of the master equation. We have considered the network of Fig. 3(a) with $N = 100$ packets, which corresponds to $N_{\text{eq}} = 10626$ coupled equations. We also compare the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field method. Again, both methods, the mean field and the master equations, provide very similar results, i.e., $\langle n_i(t) \rangle \approx \langle m_i(t) \rangle$.

We now turn our attention to the dynamics over *small-world* networks. To construct the *small-world* networks of Watts and Strogatz²² we start with a ring network of $M = 8$ nodes, as shown in Fig. 4(a). The evolution of the average number of packets in the network, along with its standard deviation, is shown in Fig. 4(b), which shows an excellent agreement with the mean field approach. As expected the system converges to $\langle n_i \rangle \rightarrow N/M$ for a ring.

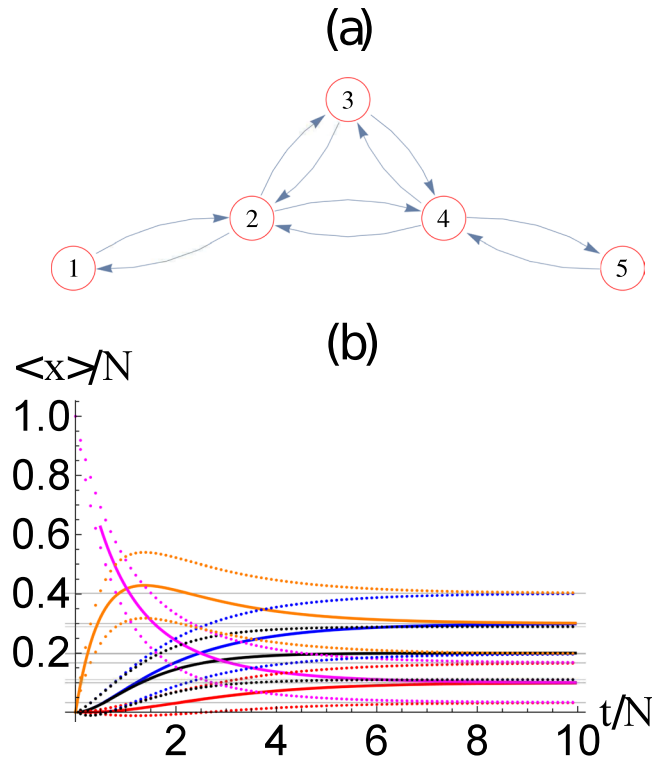


Figure 3. (a) Five-node network, in which the arrows determine the directed connectivity of the network. (b) Evaluation of the ensemble average number of packets at each node $\langle n_i(t) \rangle$ (continuous line) and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ (dotted curves), as a function of time. The horizontal dotted lines are the corresponding values obtained from the asymptotic solution. Here we use $N = 20$ packets. The evolution of the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field shows an excellent agreement. Here $i = 1$ (red), 2 (blue), 3 (black), 4 (orange), 5 (magenta).

For a *small-world* networks of Watts and Strogatz²², we start with the ring and then connect $M \times p$ distinct pairs of nodes, as shown in Fig. 5(a) for $p = 1$. These networks are called *small-world* networks because the average distance $\langle D \rangle / M$ between nodes decreases with p . The distance between nodes i and j is defined as the minimum number of steps required to reach node j from node i along the network, and considering the directed nature of the network. The evolution of the average number of packets in the network, along with its standard deviation, is shown in Fig. 5(b), which shows an excellent agreement with the mean field approach.

We have studied the package evolution in two other types of networks, namely the *Erdos-Renyi* (Fig. 6) network (a complete random network) and the scale-free network (Fig. 7). Figures 6(a) and 7(a) shows the graph representation of the networks we used in our simulations. Figures 6(b) and 7(b) shows the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ of its respective networks. As before, there is an excellent agreement between the results obtained from the mean field approach and the master equation. It is interesting to notice that the undirected network of Fig. 5 has a larger number of different asymptotic states than the undirected networks of Figs 6 or 7. Although highly dependent on the particular connectivity distribution, it is expected that in general the breaking of the undirected symmetry of a network should produce more different asymptotic states.

It is worth noticing the overshoot phenomena that appears in Figs 3–7. We have checked that different initial conditions (i.e., varying the position in which all packages are placed at the beginning of the simulation) may modify the first part of the dynamics, producing overshoot or damping at different nodes. Hence, the overshoot that occurs at a particular node depends on the distance to the initial node, but also on the connectivity of the neighboring nodes which control how the packages are taken from each node. Of course, we checked that the asymptotic behavior is the same in all cases.

Asymptotic equilibrium state. The equilibrium state obtained from the asymptotic solution of the master equation for the network of Fig. 2(a) is $\langle \vec{n} \rangle = [12.5, 25, 12.5]$, which is the same as the one obtained from the mean field approach. The asymptotic solution for the average occupation number at each node, compared with their dynamics produced by the mean field and master equation approach is shown in Fig. 2(c), showing excellent agreements. It is interesting to notice that the corresponding covariance matrix is

$$\sigma^2 = \begin{bmatrix} 9.37 & -6.26 & -3.12 \\ -6.25 & 12.50 & -6.25 \\ -3.12 & -6.25 & 9.37 \end{bmatrix},$$

so that the standard deviations $\sigma_i = \sqrt{\sigma_{ii}^2}$ (diagonal terms) are not all equal in the asymptotic state.

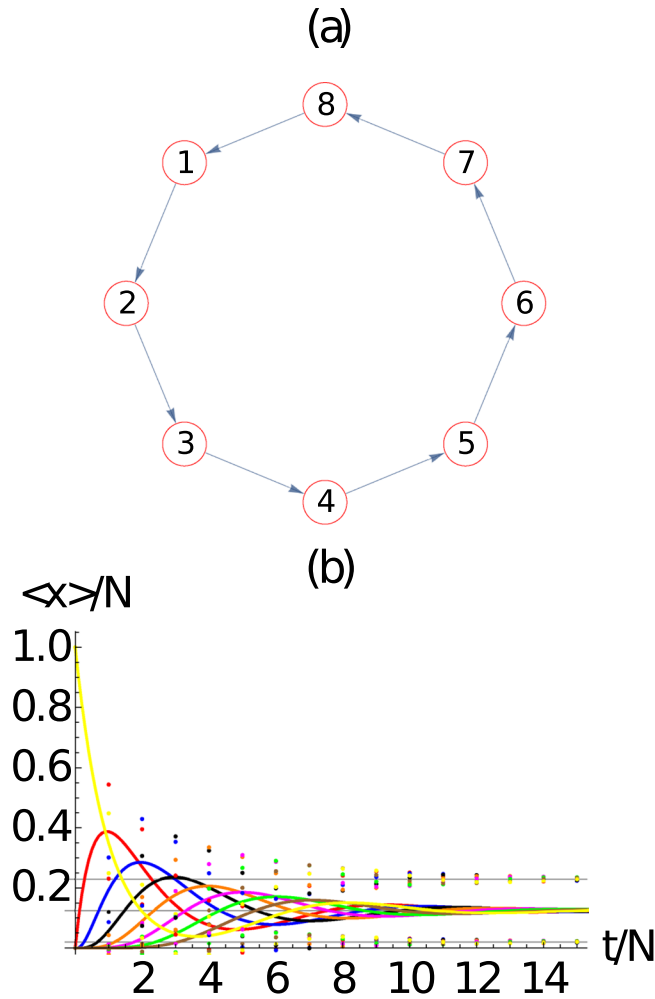


Figure 4. (a) Ring network with $M=8$ nodes in which the arrows determine the directed connectivity of the network. (b) Evaluation of the ensemble average number of packets at each node $\langle n_i(t) \rangle$ (continuous line) and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ (dotted curves), as a function of time. Here we use $N=10$ packets. We observe the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field approach presents an excellent agreement. Here $i=1$ (red), 2 (blue), 3 (black), 4 (orange), 5 (magenta), 6 (green), 7 (brown), 8 (yellow).

The asymptotic state obtained for the five-node network from the asymptotic solution of the master equation is $\langle \vec{n} \rangle = [2, 6, 4, 6, 2]$, which is the same as the one obtained from the mean field approach. The agreement between the mean field and master equation results is excellent, as displayed in Fig. 3(c). It is interesting to notice that the corresponding covariance matrix is

$$\sigma^2 = \begin{bmatrix} 1.8 & -0.6 & -0.4 & -0.6 & -0.2 \\ -0.6 & 4.2 & -1.2 & -1.8 & -0.6 \\ -0.4 & -1.2 & 3.2 & -1.2 & -0.4 \\ -0.6 & -1.8 & -1.2 & 4.2 & -0.6 \\ -0.2 & -0.6 & -0.4 & -0.6 & 1.8 \end{bmatrix},$$

so that the standard deviation σ_i (diagonal terms) are not all equal in the asymptotic state.

In Fig. 8(a,b), we show the asymptotic occupation number distribution $P(n_i)$ for the (a) three and (b) five node networks from Figs 2 and 3. Similarly, in Fig. 9 we show the asymptotic occupation number distribution for the (a) *small world*, (b) *Erdos-Renyi*, and (c) *scale free* networks with $M=8$ nodes and $N=10$ packages. Hence for small networks and packages, given the computational restrictions imposed by Eq. 19, it is reasonable to solve the master equation directly. Furthermore, for small M we notice that as N increase, the occupation number distributions at the i^{th} node approaches a normal distribution

$$P(n_i) \approx C_i \exp \left[-\frac{(n_i - \langle n_i \rangle)^2}{2\sigma_{ii}^2} \right], \tag{20}$$

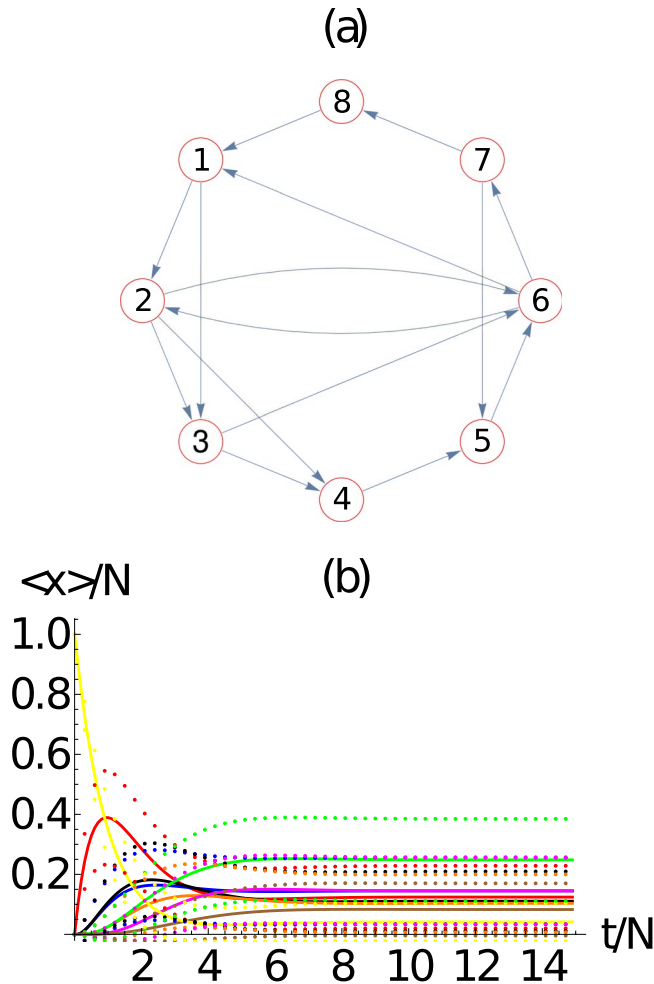


Figure 5. Small-world network with $M = 8$ nodes and $p = 1$, in which the arrows determine the directed connectivity of the network. (b) Evaluation of the average number of packets at each node $\langle n_i(t) \rangle$ (continuous line) and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ (dotted curves), as a function of time. Here we use $N = 10$ packets. We observe the excellent agreement between the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field approach. Here $i = 1$ (red), 2 (blue), 3 (black), 4 (orange), 5 (magenta), 6 (green), 7 (brown), 8 (yellow).

centered at $\langle n_i \rangle$ with a standard deviation given by σ_i . The normalization constant C_i is such that $\sum_{n_i=0}^N P(n_i) = 1$. The expected form of Eq. (20), for large N , is in agreement with Eq. (31), that is obtained from the continuous time scale description of the master equation (see Methods section for a derivation).

As N and M increase, the asymptotic state becomes increasingly more complicated to calculate, specially if we are interested in calculating the probability distribution and the standard deviation of n_i . However, let us note that the average value of $\langle n_i \rangle$ for the asymptotic state can be computed from the mean field approach. Hence, we can extrapolate $\langle n_i \rangle(N)$ as a function of N from the $N = 1$ case, namely

$$\langle n_i \rangle(N) = N \langle n_i \rangle(N = 1). \tag{21}$$

Similarly, in Fig. 10(a,b), we show the standard deviations $\sigma_i(N)$ as a function of N for the (a) three and (b) five node networks for Figs 2 and 3, respectively. The continuous lines, in each case, corresponds to the scaling

$$\sigma_{i,i}(N) = \sqrt{N} \sigma_{i,i}(N = 1), \tag{22}$$

which clearly gives an excellent approximation, even for relatively small values of N . Notice that this is expected from a stochastic system in which

$$\frac{\sigma_{ii}}{\langle n_i \rangle} \sim \frac{1}{\sqrt{N}}. \tag{23}$$

Therefore, we see that we can compute the asymptotic state of the master equation from the $N = 1$ case and then re-scale the distribution to larger values of N using the scaling properties just discussed. In fact, the

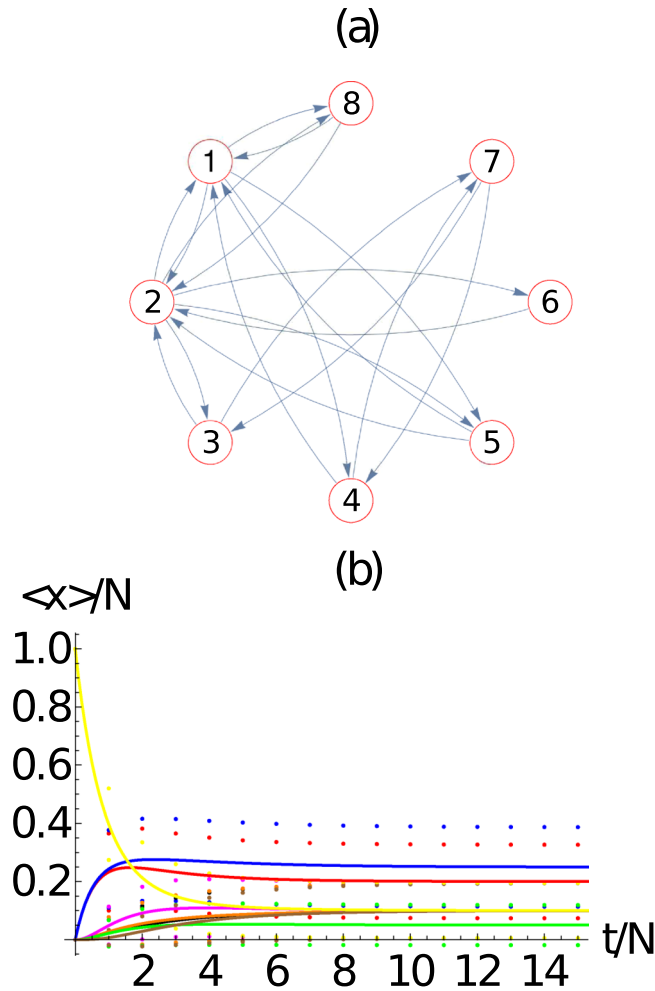


Figure 6. (a) *Erdos-Renyi* network with $M=8$ nodes and 10 bidirectional nodes which corresponds to a density of 35% of the total possible connections (28 in total). (b) Evaluation of the ensemble average number of packets at each node $\langle n_i(t) \rangle$ (continuous line) and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ (dotted curves), as a function of time. Here we use $N=10$ packets. We observe that the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field approach exhibit an excellent agreement. Here $i=1$ (red), 2 (blue), 3 (black), 4 (orange), 5 (magenta), 6 (green), 7 (brown), 8 (yellow).

distribution functions displayed in Fig. 8(a,b), are constructed in this manner, showing that it is an excellent approximation, specially as N increases (thermodynamic limit).

The same analysis has been done for the ring and *small network* of Figs 4(c) and 5(c), respectively, which is in close agreement with Eq. 20 and the scaling from the $N=1$ case.

Hence, if we are interested in estimating the syntactic probability distribution of the occupation number for N packets, it becomes of interest to solve the master equation for $N=1$ which has M possible states, namely

$$\vec{n}_i = [0, 0, \dots, 0, n_i = 1, 0, \dots, 0]. \tag{24}$$

The evolution equation for them (using $p_i(t) = P(\vec{n}_i, t)$) is

$$p_i(t+1) = \sum_{j \neq i} \frac{A_{j,i}}{k_j} p_j(t), \tag{25}$$

From this equation, it becomes clear that finding the steady state $p_i(t+1) = p_i(t) = p_i$ for $N=1$ is completely equivalent to finding the asymptotic mean field $\lambda_0=0$ eigenvector of the matrix \mathbf{B} given by Eq. 5, however in this case is easier to solve directly the M equations

$$\sum_{i=1}^M B_{i,j} p_j = 0,$$

with the restriction $\sum_{i=1}^M p_i = 1$, than solving the complete eigensystem.

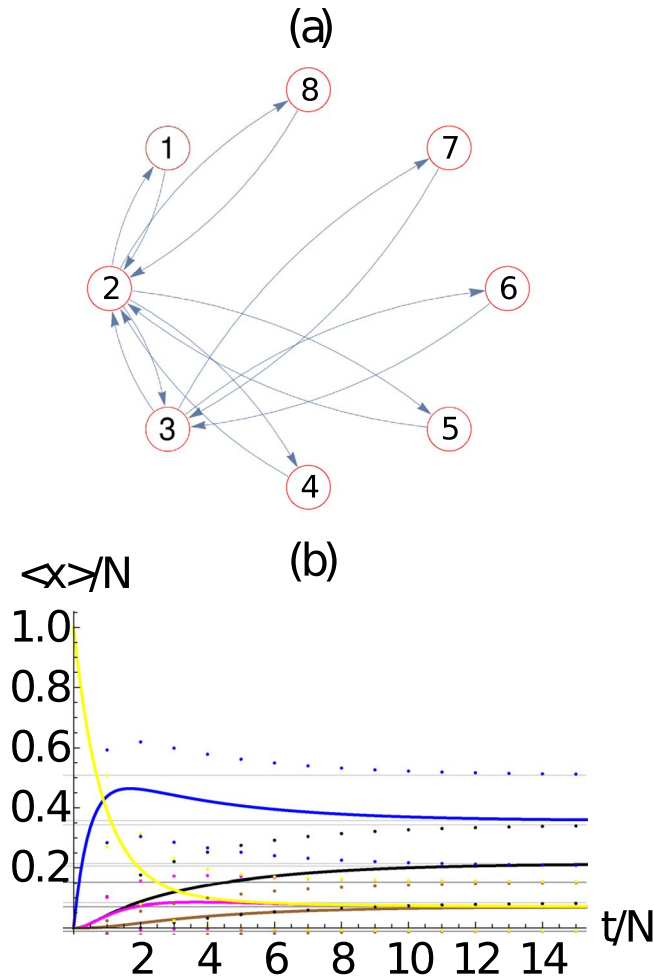


Figure 7. (a) Scale-free network with $M=8$ nodes and 1 bidirectional node added at each step of the free scale network algorithm of construction with probability of attachment proportional to the vertex degree. (b) Evaluation of the ensemble average number of packets at each node $\langle n_i(t) \rangle$ (continuous line) and their standard deviation $\langle n_i(t) \rangle \pm \sigma_{ii}$ (dotted curves), as a function of time. Here we use $N = 10$ packets. We observe the excellent agreement of the evolution of the average number of packets at each node $\langle n_i(t) \rangle$ obtained from the master equation and $\langle m_i(t) \rangle$ obtained from the mean field approach. Here $i = 1$ (red), 2 (blue), 3 (black), 4 (orange), 5 (magenta), 6 (green), 7 (brown), 8 (yellow).

We see that there is a clear connection between the mean field approach and the master equations, which are described in the previous text. Once, we find the asymptotic states given by p_i , we observe that $n_i p_j \rightarrow \delta_{ij} p_j$, so that the expected occupation value is

$$\langle n_i \rangle (N = 1) = p_i,$$

and the standard deviation can be found from

$$\sigma_{i,j}^2 (N = 1) = p_i \delta_{i,j} - p_i p_j,$$

which explain the negative off-diagonal values obtained above for the three and five node networks. Using these expressions, we can scale to any N and find $\langle n_i \rangle (N)$, σ^2 , and $P(n_i, N)$. We have to use these equations to construct the analytic approximation to the occupation probability at each node for any network, as was done for the cases of Fig. 8.

We can use this strategy, which is much less computational intensive, to re-construct the analytic approximation to the occupation probability at each node for any network, as was done in Fig. 8 for the 2 and 3 node example; and in Fig. 9 for the (a) *small world*, (b) *Erdos-Renyi*, and (c) *scale free* networks. The results show very good agreement with the master equation result, which becomes increasingly more accurate as N is incremented as can be observed in Fig. 8.

We verified our results in large scale complex networks instances of the I. *small-world*, II. *Erdos-Renyi*, and III. *scale-free* scenarios. Figure 11. summarizes these results. Panels I(a), II(a), III(a) present the frequency F of the

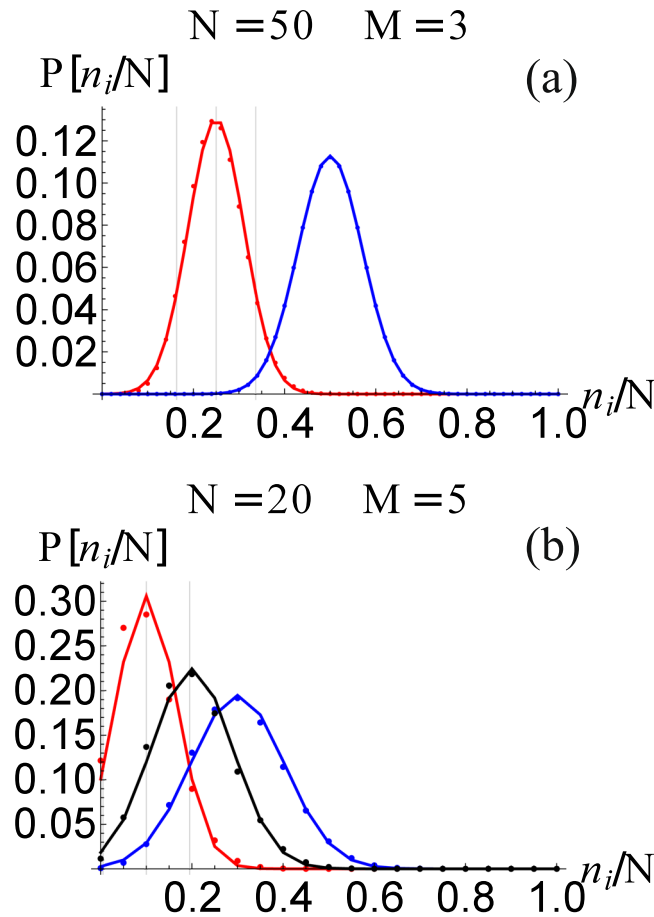


Figure 8. The occupation number distribution $P(n_i)$ for the (a) three and (b) five node networks for Figs 2 and 3, respectively. The dots are the values obtained from the master equation, while the continuous lines corresponds to Eq. 20 with the extrapolation from the $N = 1$ case discussed in the text. We also show $\langle n_i \rangle$ and $\langle n_i \rangle \pm \sigma_{ii}$ as vertical lines.

average number of packets in the asymptotic states for stochastic simulation, and the mean field solution of its respective networks (*small-world*, *Erdos-Renyi*, and *scale-free*). Panels (b) show the correlation between the average occupation of the master equation $\langle n_i^{ME} \rangle$ and the standard deviation of the stochastic simulation $\langle n_i^{SS} \rangle$ for its respective networks. Panels (c) exhibit the correlation between the standard deviation of the master equation σ_i^{ME} and the standard deviation of the stochastic simulation σ_i^{SS} , respectively. Each network contains 1000 nodes ($N = 1000$) and 10^4 packets ($M = 10^4$). In the *small-world* network, it was created considering a re-linking probability of 0.5; for the *scale-free* case, the network was built adding 2 bidirectional node at each step of the free scale network algorithm of construction with probability of attachment proportional to the vertex degree. The *Erdos-Renyi* network was built using a edge probability equal to 0.5. Sub-figures II and III were obtained studying the behaviors of the mean and the standard deviation of the master equation and the stochastic simulation of a particular node on the respective network. From figs. I(a), II(a), and III(a), despite small differences, the shape of the distributions are quite similar between stochastic simulation cases and mean field approach. Solid lines in sub-panels II and III are added to indicate the correlation equal to 1. To sum up, instances presented in Fig. 11 shows excellent agreement among stochastic simulation, mean field approach and master equation, allowing to observe notorious differences among different type of complex networks in large scales. It is worth noticing that the Erdos-Renyi case is simulated longer than small-world and scale-free cases: the Erdos-Renyi case was simulated with 10^5 simulation steps, while the other two cases were simulated with 5×10^4 simulation steps. The relaxation time scale $\tau = N|\Re[\tilde{\lambda}]|$ of the Erdos-Renyi network, calculated as the inverse value of the smallest nonzero eigenvalue of the B matrix, turns out to be twice as large as the time scale of the other two cases, explaining why we need to integrate longer for the ensemble averaged number of packets and its deviation to converge to the asymptotic value. This difference in time scales has to do with the fact that the degree distribution of Erdos-Renyi does not have large tails, so that the connectivity is spanned over a narrow range of values, which in turn leads to a continuous distribution of values for the ensemble averaged number of packets and its deviation. In contrast the small-world or scale-free networks has a hierarchical structure of the connectivity which is clearly evidenced by a discrete distribution for the ensemble averaged number of packets and its deviation.

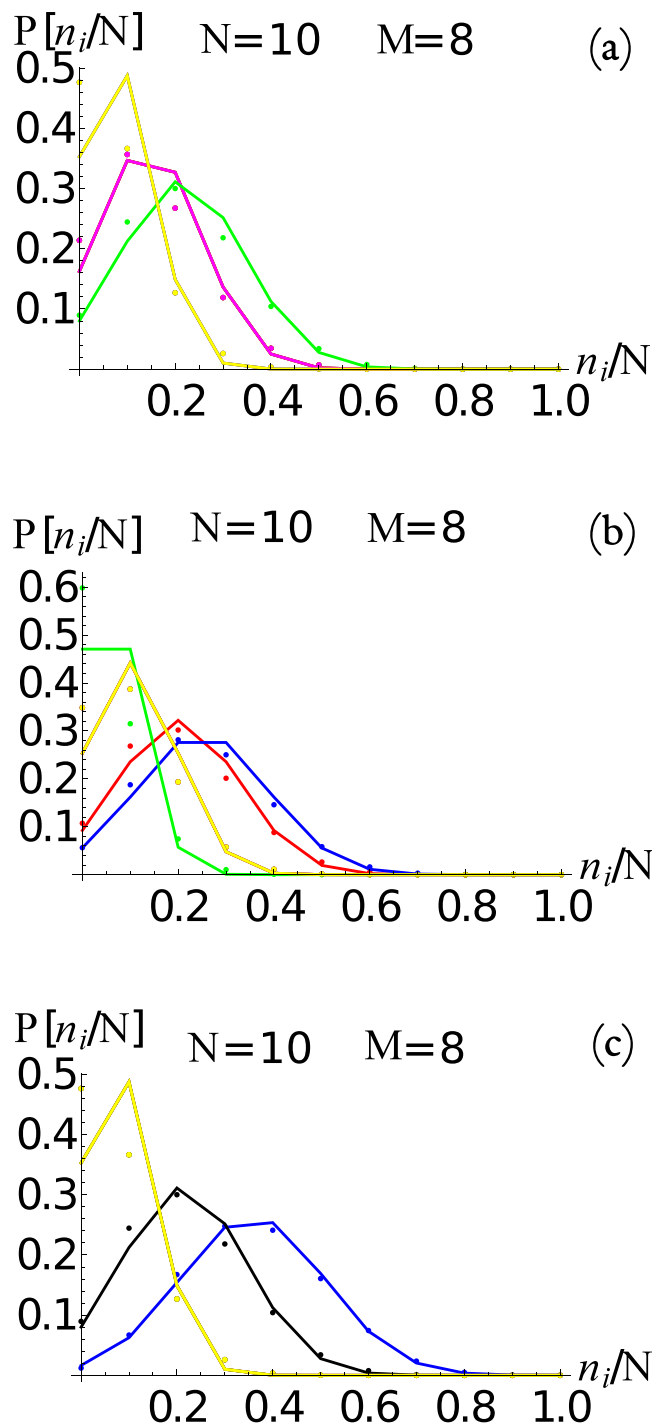


Figure 9. The occupation number distribution $P(n_i)$ for the (a) *small-world* (see Fig. 5), (b) *Erdos-Renyi* (see Fig. 6), and (c) *scale-free* (see Fig. 7) networks respectively, with eight nodes. The dots are the values obtained from the master equation, while the continuous lines corresponds to Eq. 20 with the extrapolation from the $N=1$ case discussed in the text.

Discussion

We have generalized the Ehrenfest urn model to a complex network of urns, in which the packets or marbles move from node to node following the network connections. We have constructed the master equation for the evolution of the probability of occupation of each of the nodes in the network. The calculated occupation number at each node $\langle n_i \rangle$ compares quite closely with analytic solution for the ensemble average evolution $\langle m_i \rangle$ of the number of packets at each node, obtained from a mean field approach in the thermodynamic limit (namely $N \gg 1$).

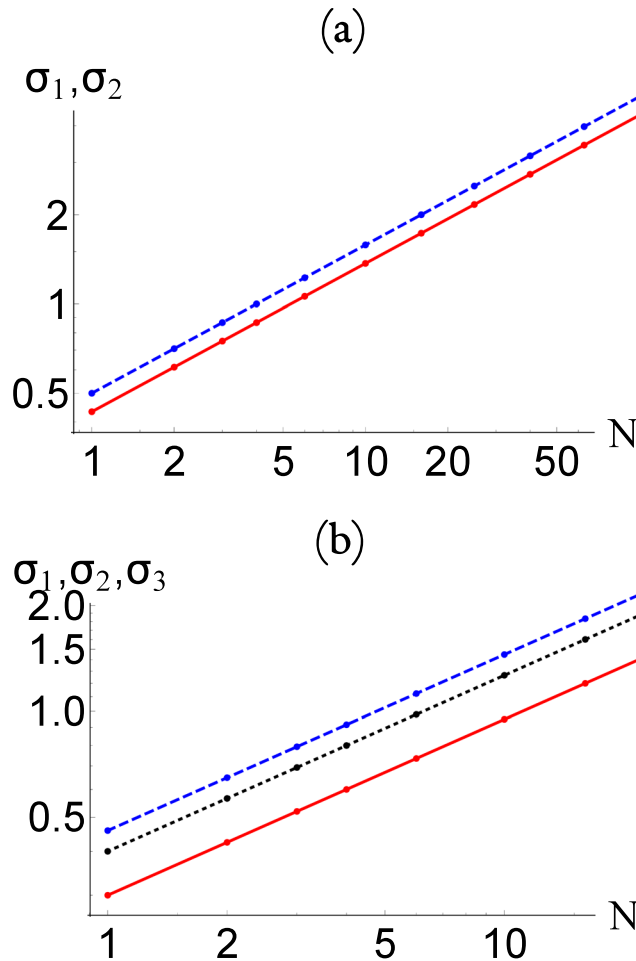


Figure 10. The standard deviation $\sigma_i(N)$ as a function of N for the (a) three and (b) five node networks of Figs 2 and 3, respectively. The dots corresponds to data obtained directly from the asymptotic state calculated from the master equation, and the straight line corresponds to the re-scaling of the distribution function as discussed in the text.

We clearly observe that mean field theory provides a good approximation for the evolution of the ensemble averaged number of packets, as compared to the the evolution of the more complete master equation. However, the master equation provides a more complete description, allowing to calculate all the statistical properties of the system at any time t .

We also notice that the asymptotic state provided by the master equation is quite useful to find the equilibrium distribution of the occupation at each node, providing a complete statistical description at equilibrium. Furthermore, we can find scaling laws to approximate the asymptotic solution to the occupation number probability at each node from the $N=1$ case, which involves the matrix \mathbf{B} used in the mean field approach. One of the main conclusions of the manuscript is that for small networks with a small number of packets, it is necessary to find the asymptotic solution, including the correlation matrix, directly from the master equation, which is in general computationally expensive. While for large values of N it is possible to estimate the asymptotic state, including the correlation matrix, from the $\lambda_0=0$ eigenvector of the matrix \mathbf{B} with $N=1$, as the whole distribution becomes normal as N increases with the distribution parameters satisfying the scaling relations given by Eqs 21 and 22. This approximation improves as we increase the number of packets, i.e., in the thermodynamic limit. Hence, the mean field matrix \mathbf{B} can be use to estimate not only the average occupation number, but also the occupation probability distribution, and in particular the standard deviation of the average occupation number.

By comparing the mean field evolution of the network of Fig. 2 with the networks of Figs 3, 4 and 5, we observe that there is an overshoot phenomena that occurs before the system reaches the asymptotic dynamics. However, the initial condition can, and the distance from the node that contains all packets initially, also determine if there is overshoot or not. For example, if we take the network of Fig. 3, there are 3 non-equivalent nodes in which we can initially place all the packages, namely, nodes 1, 2, and 3. In this sense initially placing all the packages in node 4 is equivalent to placing them in node 2. Similarly, node 5 is equivalent to node 1. Hence, initially placing all the packages in node 1, 2, 4, and 5 produce an overshoot phenomena in nodes 2, 1, 5, and 4, respectively. However, initially placing all the packages in node 3 does not produce the overshoot phenomena. The reason for the overshoot is the following: for example, when we initially place all packages in node 1, the overshoot phenomenon

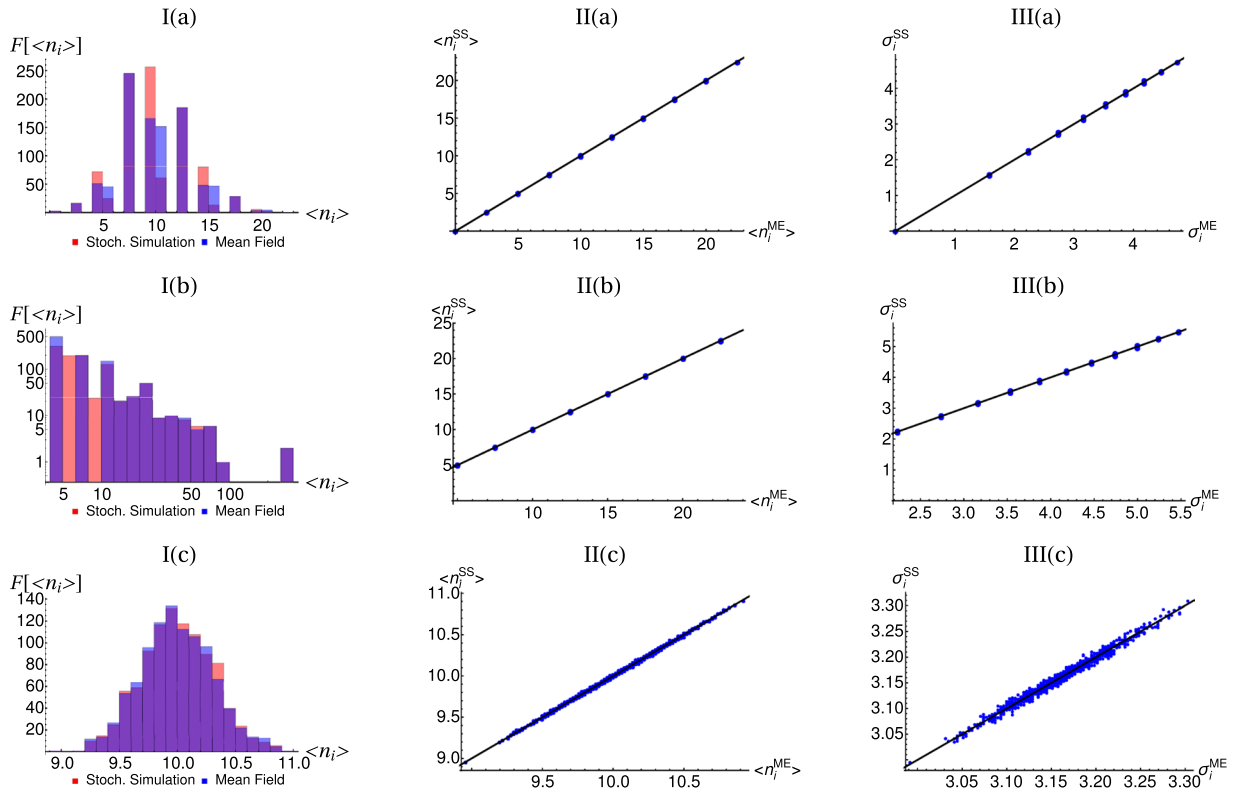


Figure 11. Simulations on large scale networks instances. I(a) (*small-world*), II(a) (*Erdos-Renyi*), III(a) (*scale-free*) present the probability distributions of the mean of the asymptotic states for stochastic simulation, the master equation approach, and the mean field solution. I(b), II(b), and III(b) shows the correlation between the average occupation of the master equation $\langle n_i^{ME} \rangle$ and the standard deviation of the stochastic simulation $\langle n_i^{SS} \rangle$ for its respective networks. I(c), II(c), and III(c) exhibit the correlation between the standard deviation of the master equation σ_i^{ME} and the standard deviation of the stochastic simulation σ_i^{SS} , respectively. Simulations were run on networks of 1000 nodes ($N = 1000$) and 10^4 packets ($M = 10^4$). For the *small-world* case, the network was built with a re-linking probability of 0.5; for the *scale-free* case, the network was built adding 2 bidirectional node at each step of the free scale network algorithm of construction with probability of attachment proportional to the vertex degree; and the Erdos-Renyi network was built using a edge probability equal to 0.5. Sub-figures II and III were obtained studying the evolution of a particular node in network type. The *Erdos-Renyi* case was simulated with 10^5 simulation steps, while the other two cases were simulated with 5×10^4 simulation steps.

occurs in node 2 as all the packages need to pass through node 2 before they can get distributed to the rest of the network. The opposite argument applies for the nonexistence of the overshoot phenomena when initially placing all packets in node 3. Hence, the topology of the network and the initial condition control the existence of this overshoot phenomenon, however, the asymptotic behavior is robust in all cases.

Finally, it is interesting to mention that the fact that the standard deviation of the occupation number at each node is not uniform in the asymptotic states, proves that the *equal a priori probabilities* proposed by Boltzmann does not apply for the transport in these networks, unless there is an underlying symmetry. The complex topology of the network provides a way to equilibrate fluctuations that become non-uniform throughout the system. This observation may have relevant implications in the understanding of the statistical mechanics of transportation networks. For example, as car change lanes in a 3 lane street, we would expect not only that the central lane will be more occupied on average than the other two lanes, but also that its fluctuations will also be larger.

Methods

Let us assume that temporal evolution of the probability $P(\vec{n}, t)$ given by Eq. (14) can be written in a continuous timescale as a transition equation

$$\frac{d}{dt}P(\vec{n}, t) = \sum_{\vec{n}'} (W_{\vec{n}, \vec{n}'}P(\vec{n}', t) - W_{\vec{n}', \vec{n}}P(\vec{n}, t)), \tag{26}$$

where $W_{\vec{n}, \vec{n}'}(W_{\vec{n}', \vec{n}})$ is the transition probability of the process $\vec{n} \rightarrow \vec{n}'(\vec{n}' \rightarrow \vec{n})$. Thus from

$$\frac{d}{dt} \sum_{\vec{n}} P(\vec{n}, t) = \sum_{\vec{n}, \vec{n}'} (W_{\vec{n}, \vec{n}'} P(\vec{n}', t) - W_{\vec{n}', \vec{n}} P(\vec{n}, t)) = 0, \quad (27)$$

we see that the total probability is conserved. From Eqs (16) and (26), we can write

$$\begin{aligned} \frac{d}{dt} \langle m_i(t) \rangle &= \sum_{\vec{n}} n_i \frac{dP(\vec{n}, t)}{dt}, \\ &= \sum_{\vec{n}, \vec{n}'} n_i (W_{\vec{n}, \vec{n}'} P(\vec{n}', t) - W_{\vec{n}', \vec{n}} P(\vec{n}, t)). \end{aligned} \quad (28)$$

Now let us suppose that close to the steady state the transition probabilities $W_{\vec{n}, \vec{n}'} \rightarrow 0$ and $(W_{\vec{n}', \vec{n}}) \rightarrow \omega$ (according to the ergodic theorem), thus the Eq. (28) takes the form

$$\frac{d}{dt} \langle m_i(t) \rangle = -\omega \sum_{\vec{n}'} n_i P(\vec{n}', t) = -\omega \langle m_i(t) \rangle,$$

so that the solution is $\langle m_i(t) \rangle = e^{-\omega t} \langle m_i(0) \rangle$. Comparing with Eq. (9), we note that the transition probability $\omega = |\Re[\tilde{\lambda}]|/N$. Furthermore, around the steady state \vec{n} , the master equation for the probability can be approximated as

$$\frac{d}{dt} P(\vec{n}, t) = \omega \sum_{\vec{\rho}} (P(\vec{n} + \vec{\rho}, t) - P(\vec{n}, t)), \quad (29)$$

$$\begin{aligned} &\approx \omega \left((\vec{\rho} \cdot \nabla) P(\vec{n}, t) + \frac{1}{2} (\vec{\rho} \cdot \nabla)^2 P(\vec{n}, t) \right) \\ &= \frac{\omega}{2} (\vec{\rho} \cdot \nabla)^2 P(\vec{n}, t). \end{aligned} \quad (30)$$

for small $\vec{\rho}$. For the steady state, $(\vec{\rho} \cdot \nabla) P(\vec{n}, t) = 0$, since at equilibrium the steady state is the most probable configuration. Hence, a direct calculation shows that the solution of the above equation is given by

$$P(\vec{n}, t) = \frac{P_0}{\sqrt{t}} e^{-(\vec{n} - \langle \vec{n} \rangle)^2 / 2\omega t}, \quad (31)$$

which implies that the expected distribution at the steady state configuration should be a Gaussian in the thermodynamic limit.

References

- Ehrenfest, P. & Ehrenfest, T. Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. *Physikalische Zeitschrift* **8**, 311 (1907).
- Kac, M. On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math. Soc.* **54**, 369 (1959).
- Kac, M. Random walk and the theory of Brownian motion. *Amer. Math. Mon.* **54**, 369 (1959).
- Clark, J., Kiwi, M., Torres, F., Rogan, J. & Valdivia, J. A. Generalization of the Ehrenfest urn model to a complex network. *Phys. Rev. E* **92**, 012103 (2015).
- Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47 (2002).
- Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167 (2003).
- Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Adv. Phys.* **51**, 1079 (2002).
- Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
- Redner, S. How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B* **4**, 131 (1998).
- Ferrer i Cancho, R., Janssen, C. & Solé, R. V. Topology of technology graphs: small world patterns in electronic circuits. *Phys. Rev. E* **64**, 046119 (2001).
- Aiello, W., Chung, F. & Lu, L. *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, 171–180 (2000).
- Aiello, W., Chung, F. & Lu, L. *Handbook of Massive Data Sets* (Kluwer, Netherlands, 2000).
- Broder, A. et al. Graph structure in the web. *Comput. Networks* **33**, 309 (2000).
- Jeong, H., Tombor, B., Albert, R., Oltvali, Z. N. & Barabási, A. L. The large-scale organization of metabolic networks. *Nat.* **407**, 651 (2000).
- Amaral, L. A. N., Scala, A. & Barthélemy, H. E. M. Stanley. Classes of small-world networks. *Proc. Nat. Acad. Sci.* **97**, 11149 (2000).
- Jeong, H., Mason, S. P., Barabási, A. L. & Oltvali, Z. N. Lethality and centrality in protein networks. *Nat.* **411**, 41 (2001).
- Faloutsos, M., Faloutsos, P. & Faloutsos, C. *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication* (1999).
- Ferrer i Cancho, R. & Solé, R. V. The small world of human language. *Proc. Royal Soc. B* **268**, 2261 (2001).
- Albert, R. & Barabási, A. L. Emergence of scaling random networks. *Science* **286**, 509 (1999).
- Barabási, A. L., Nitsch, L. & Dorobantu, I. A. Super-tracks and nth order windows in the chaotic regime. *Phys. Lett. A* **139**, 53 (1989).
- Lipowski, A., Lipowska, D. & Ferreira, A. L. Emergence of social structures via preferential selection. *Phys. Rev. E* **90**, 032817 (2014).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of small-world networks. *Nat.* **393**, 440 (1998).
- Liljeros, F., Edling, C. R. & Amaral, L. A. N. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes Infect.* **5**, 189 (2003).
- Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, E. H. & Aberg, Y. The web of human sexual contacts. *Nat.* **411**, 907 (2001).
- Tadic, B., Rodjers, G. J. & Thurner, S. Transport on complex networks: Flow, jamming and optimization. *Int. J. Bifurc. Chaos* **17**, 2363 (2007).

26. Farina, A., Graziano, A., Marianini, F., Recchioni, M. C. & Zirilli, F. Homogeneous and heterogeneous traffic of data packets on complex networks: The traffic congestion phenomenon. *Commun. Netw.* **4**, 147 (2012).
27. Barthélemy, M. & Flammini, A. Optimal traffic networks. *J. Stat. Mech.* **L07002** (2006).
28. Lammer, S., Gehlsen, B. & Helbing, D. Decentralised control of material or traffic flows in networks using phase-synchronisation. *Phys. A* **363**, 89 (2006).
29. Varas, A. D. *et al.* Resonance, criticality, and emergence in city traffic investigated in cellular automaton models. *Phys. Rev. E* **80**, 056108 (2009).
30. Castillo, F. *et al.* City traffic jam relief by stochastic resonance. *Phys. A* **403**, 65 (2013).
31. Nagatani, T. Fluctuation of riding passengers induced by chaotic motions of shuttle buses. *Phys. Rev. E* **68**, 036107 (2003).
32. Villalobos, J. *et al.* Regular transport dynamics produce chaotic travel times. *Phys. Rev. E* **89**, 062922 (2014).
33. Li Yang, S., Hsu, N. S., Louie, P. W. F. & Yeh, W. W. G. Water distribution network reliability: Connectivity analysis. *J. Infrastructure Syst.* **2**, 54 (1996).
34. Sen, P. *et al.* Small-world properties of the Indian railway network. *Phys. Rev. E* **67**, 036106 – Published 14 March 2003.
35. Goh, K., Lee, B., Kahng, D. & Kim, D. Sandpile on scale-free networks. *Phys. Rev. E* **91**, 148701 (2003).
36. Lee, D., Goh, K. I., Kahng, B. & Kim, D. Sandpile avalanche dynamics on scale-free networks. *Phys. A* **84**, 338 (2004).
37. Pasten, D., Munoz, V., Cisternas, A., Rogan, J. & Valdivia, J. A. Monofractal and multifractal analysis of the spatial distribution of earthquakes in the central zone of Chile. *Phys. Rev. E* **84**, 066123 (2011).
38. Kao, Y. & Luan, P. Poincaré cycle of a multibox Ehrenfest urn model with directed transport. *Phys. Rev. E* **67**, 031101 (2003).
39. Luan, P. G. & Kao, Y. Drifting diffusion on a circle as continuous limit of a multiurn Ehrenfest model. *Phys. Rev. E* **69**, 022102 (2004).
40. Luan, P. & Kao, Y. Poincaré cycle of an Ehrenfest multurn model in a one-dimensional ring. *Phys. Rev. E* **69**, 027103 (2004).
41. Lipowski, A. Criticality in a model with absorbing states. *Phys. Rev. E* **63**, 026105 (2001).

Acknowledgements

This work was supported by the Fondo Nacional de Investigaciones Científicas y Tecnológicas (FONDECYT, Chile) under grants #1120399 and 1130272 (MK and JR), #1150718 (JAV). We thank the support of Financiamiento Basal para Centros Científicos y Tecnológicos de Excelencia under grants FB0807 (MK, FT, JR, PM, and JAV), to Conicyt Anillos Científicos under grant ACT-1405 (JR and JAV), and US AFOSR under Grant FA9550-16-1-0384 (PM, FT, JR, MK, JAV). PM thanks the support of Fondecyt Posdoctoral project 3180315.

Author Contributions

J.C., P.M., M.K., F.T., J.R. and J.A.V. equally contributed to this work, developing simulations, calculations and contributing to the discussion. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018