View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repository of the University of Namur

Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche

researchportal.unamur.be

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Reactive explorers to unravel network topology

Adam, Ihusan; Fanelli, Duccio; Carletti, Timoteo; Innocenti, Giacomo

Published in:
European Physical Journal B

DOI:
https://doi.org/10.1140/epjb/e2019-90700-3

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (HARVARD):
Adam, I, Fanelli, D, Carletti, T & Innocenti, G 2019, 'Reactive explorers to unravel network topology' European Physical Journal B, vol. 92, no. 99. https://doi.org/10.1140/epjb/e2019-90700-3

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 21. May. 2019

Eur. Phys. J. B (2019) 92: 99
https://doi.org/10.1140/epjb/e2019-90700-3

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Reactive explorers to unravel network topology

Ihusan Adam[1,2], Duccio Fanelli[3,4,a], Timoteo Carletti[5], and Giacomo Innocenti[1]

[1] Dipartimento di Ingegneria dell'Informazione, Università di Firenze, Via S. Marta 3, 50139 Florence, Italy
[2] Università degli Studi di Firenze, Dipartimento di Fisica e Astronomia, CSDC and INFN, via G. Sansone 1, 50019 Sesto Fiorentino, Italy
[3] Dipartimento di Fisica e Astronomia and CSDC, Università degli Studi di Firenze, via G. Sansone 1, 50019 Sesto Fiorentino, Italy
[4] INFN Sezione di Firenze, via G. Sansone 1, 50019 Sesto Fiorentino, Italy
[5] naXys, Namur Institute for Complex Systems, University of Namur, Namur, Belgium

Received 29 November 2018 / Received in final form 21 March 2019
Published online 8 May 2019
© EDP Sciences / Società Italiana di Fisica / Springer-Verlag GmbH Germany, part of Springer Nature, 2019

Abstract. A procedure is developed and tested to recover the distribution of connectivity of an a priori unknown network, by sampling the dynamics of an ensemble made of reactive walkers. The relative weight between reaction and relocation is gauged by a scalar control parameter, which can be adjusted at will. Different equilibria are attained by the system, following the externally imposed modulation, and reflecting the interplay between reaction and diffusion terms. The information gathered on the observation node is used to predict the stationary density as displayed by the system, via a direct implementation of the celebrated Heterogeneous Mean Field (HMF) approximation. This knowledge translates into a linear problem which can be solved to return the entries of the sought distribution. A variant of the model is then considered which consists in assuming a localized source where the reactive constituents are injected, at a rate that can be adjusted as a stepwise function of time. The linear problem obtained when operating in this setting allows one to recover a fair estimate of the underlying system size. Numerical experiments are carried so as to challenge the predictive ability of the theory.

1 Introduction

Networks are abstract mathematical structures, often invoked in modeling the dynamics of complex interacting units [1–5]. The brain, Internet and the cyberworld, foodwebs and social contacts are examples, drawn from distinct fields of investigation, which can be ideally grouped under the unifying umbrella of network science. Nodes (vertices) can point to individual actors of the inspected dynamics (e.g. material units, bits of information, or, on a different scale, extended populations), while edges (links) stand for existing bilateral ties. Alternatively, nodes can tag spatial or functional niches, bounded regions of an embedding landscape, mutually connected by physical or virtual paths, as epitomized by the links [4,6–10]. In several cases of interest, punctual entities, also termed agents, may jump from one node to any of its adjacent neighbors, following the intricate network's architecture. Agents relocating across the network via multiple successive jumps are said to execute a random walk: their asymptotic distribution convey important information on the inherent organization of the underlying network. If walkers do not interact with each other, their steady state distribution reflects in fact the degree of connectivity of the nodes, a direct measure of the number of links possessed by any given node of the collection. Photographing the asymptotic nodes' occupancy, enables hence to reconstruct the distribution of connectivities, a topological quantity of paramount importance when aiming at classifying the characteristics of the underlying graph. Indeed, the structure of the network is often unknown and several methods have been devised in the literature to recover it, from functions back to structure, a non trivial task that hides formidable challenges [11–19]. Efficient schemes should gather the necessary information from a limited number of nodes, as monitoring the population on each vertex becomes virtually impracticable, for large network sizes. In a recent paper [20] a variant of the random walk problem was introduced which accounts for the mutual interference between agents, as stemming from the competition for the available space in crowded operating condition [21–29]. Nodes are assigned a finite carrying capacity, a sensible constraint which makes walkers dynamically intertwingled, through dedicated nonlinear terms. The asymptotic density distribution of walkers in the presence of crowding differs significantly from that obtained under diluted conditions. In crowded conditions, the equilibrium concentration saturates for large enough values of the

[a] e-mail: duccio.fanelli@unifi.it

connectivity. This observation opens up the perspective of recovering the unknown distribution of connectivities from repeated single-node measurements of the asymptotic dynamics, at increasing crowding. The nonlinearities that originate from the interference among microscopic agents competing for space is the key of success to the proposed approach. Building on this achievement, we here generalize the method to the setting where agents perform standard, hence linear, diffusion, but the nonlinearity comes from a local reaction term. In the first part of this paper, the relative strength of the reaction and diffusion contributions is weighted by a scalar control parameter. Different equilibria are attained by the system, by modulating the latter parameter, and reflecting the interplay between reaction and diffusion terms. The equilibrium distribution is sampled by punctual measurements performed on just one node. The information gathered on the observation node is used to predict the stationary density as displayed by the system, via a direct implementation of the celebrated Heterogeneous Mean Field (HMF) approximation [4,30,31]. The entries of the sought distribution link the solution, as obtained within the HMF working ansatz, to the average density sampled on the reference node. Solving the ensuing linear problem with standard optimization tools, returns a rather accurate estimate of the distribution of connectivity, as we shall prove for a selected gallery of test network models. In the second part of the paper, we consider a variant of the model by accounting for the presence of a source where the reactive constituents are injected, at a rate that we assume to be modulated as a stepwise function of time. This allows for the fixed point to be successively tweaked, as required by the reconstruction scheme here developed. In our application, the reaction model is assumed of the logistic type and it is therefore tempting to ideally interpret the reactive explorers, as living entities crawling on the unknown network support.

2 The mathematical framework

Label with x_i the concentration of the reactive species on node i. The dynamics of the system that we shall examine is governed by the following set of ordinary differential equations:

$$\dot{x}_i = \alpha f(x_i) + (1-\alpha)\sum_{j=1}^{N} L_{ij} x_j \tag{1}$$

where $L_{ij} = \frac{A_{ij}}{k_j} - \delta_{ij}$ are the entries of the random walk Laplacian operator $\mathbf{L}$; $\mathbf{A}$ is the adjacency matrix of the (undirected) network, while $k_i = \sum_j A_{ij}$ denotes the connectivity (or degree) of node i. The scalar parameter $\alpha \in [0,1]$ gauges the relative weight of the two terms, appearing on the right hand side of the above equation. In the following, we will operate under an idealized setting and assume that α can be freely tuned within the interval of pertinence. This choice has pedagogical value, and builds on the analysis in [32]: when $\alpha = 0$ the reaction term is silenced and the agents behave as standard linear walkers. By making α progressively larger, nonlinearities gain in relevance. The linear problem that is obtained when $\alpha = 0$ can be solved analytically. By direct inspection, it is immediate to conclude that, at equilibrium ($\dot{x}_i = 0$), $x_i = k_i / \sum_j (k_j)$. When nonlinearities come into play ($\alpha \neq 0$), the complexity of the problem rises considerably and no closed form solutions exist in general. Approximate techniques can be however put forward, to access information on the asymptotic fate of the system. In particular, for relatively small values of α it can be reasonably hypothesized that the displayed concentration is still arranged in classes of connectivities, as it happens in the reference setting of a pure random walk ($\alpha = 0$). This working ansatz motivates recasting the problem at hand in the form:

$$\dot{x}_k = \alpha f(x_k) + (1-\alpha)\left[k \sum_{k'} P(k' \mid k) \frac{x_{k'}}{k'} - x_k\right] \tag{2}$$

where x_k stands for the density displayed by the nodes that share the connectivity k. The discrete index k runs from 1 to $k_{\max}$, where $k_{\max}$ stands for the largest connectivity, as exhibited by the network being analyzed. $P(k' \mid k)$ is the conditional probability that a link exists from a given class k to a class k'. In (2) we have implicitly assumed that the nonlinear contribution $f(x)$, can be also organized in classes $f(x_k)$, as reflecting the degree of connectivity associated to individual nodes. While this is not true in general, it can be reasonably postulated as long as α is forced small, i.e. when the system under scrutiny is a perturbation to the linear random walk problem. Neglecting correlation among node degrees, one can break the probability as $P(k' \mid k) = \frac{k' P(k')}{\langle k \rangle}$, where $\langle k \rangle = \sum_k k P(k)$ and $P(k')$ is the connectivity distribution. This latter condition constitutes the core of the celebrated Heterogeneous Mean Field (HMF) approximation, to which we shall make extensive reference in the following. A straightforward manipulation yields

$$\dot{x}_k = \alpha f(x_k) + (1-\alpha)\left[\frac{k}{\langle k \rangle} \sum_{k'} P(k') x_{k'} - x_k\right]. \tag{3}$$

Introduce now the quantity $\Theta = \frac{1}{\langle k \rangle} \sum_{k'} P(k') x_{k'}$ which enables one to recast the previous equation in the compact form:

$$\dot{x}_k = \alpha f(x_k) + (1-\alpha)[k\Theta - x_k]. \tag{4}$$

Θ is a collective mean-field variable, which allows to formally decouple the dynamics, as seen on different nodes, grouped in classes of homologous connectivity. Stated differently, the knowledge of Θ is sufficient, under the range of validity of the HMF approximation, to solve for the densities at any time and for all degree classes k. In the following, we will focus on the equilibrium solution, which in turn amounts to setting $\dot{x}_k = 0\ \forall k$. We will then label with $\bar{x}_k$ the fixed points as displayed by the system and, consequently, $\bar{\Theta} = \frac{1}{\langle k \rangle} \sum_{k'} P(k') \bar{x}_{k'}$. Further, we will assume

the nonlinear function $f(\cdot)$ to be of the logistic type, and thus set $f(\bar{x}_k) = \bar{x}_k(1 - \bar{x}_k)$. This is not a mandatory step for the forthcoming analysis, as any generic nonlinear function would serve equally well the scope.[1] The advantage of using a logistic equation resides in that it allows for explicit analytical progress to be made.

From equation (4), at the fixed point, one gets:

$$\bar{x}_k = \frac{(2\alpha - 1) \pm \sqrt{(2\alpha - 1)^2 + 4\alpha[(1-\alpha)k\bar{\Theta}]}}{2\alpha}. \quad (5)$$

To elaborate on the fundamental interest of equation (5), we consider a numerical implementation of system (1), assuming a random network made of $N = 200$ nodes (see caption of Fig. 1) as the backbone support. Starting out of equilibrium, the system evolves towards a fixed point, as it can be appreciated by visual inspection of Figure 1. Trajectories stemming from nodes sharing the same connectivity cluster together, thus confirming a posteriori the validity of the HMF ansatz. The asymptotic attractor as attained by the system in its late time evolution can be effectively estimated by resorting to relation (5). More specifically, we select a randomly chosen node of the pool, with degree k^*, and measure the density therein displayed, $\bar{x}_{k^*}$. By inversion of (5) one gets an estimate for the mean-field variable $\bar{\Theta}$ as:

$$\bar{\Theta} = \frac{\alpha\bar{x}_{k^*}^2 - (2\alpha - 1)\bar{x}_{k^*}}{(1-\alpha)k^*}. \quad (6)$$

This latter is then inserted in equation (5) to predict the equilibrium solution $\bar{x}_k$ for all choices of the class index k. The predicted values are depicted in Figure 1 with a symbol (crosses) and match the equilibrium solution as obtained by direct integration of the dynamics. This observation forms the basis of the scheme of inversion that we shall outline in the following. We remind that the inverse scheme is ultimately targeted to reconstructing the distribution of connectivity of a network, a priori unknown, that happens to host the inspected dynamics. Moreover, the number of necessary information are to be gathered on just one node.

3 The inverse protocol

The procedure that we shall illustrate builds on the following recipe. Imagine to perform a series of experiments by tuning progressively the parameter α, in discrete, ascending steps. The sequence of the experiments is indexed by r, which ranges from 1 to at least $k_{\max}$. In each experiment the system is let to equilibrate, and the corresponding density $\bar{x}_{k^*}^{(r)}$ is recorded on a node of class k^* where the inspection is performed. From the knowledge of $\bar{x}_{k^*}^{(r)}$ one can infer an estimate of $\bar{\Theta}_r$, which can be used to access an approximate measure of $\bar{x}_k^{(r)}$, for $k \neq k^*$, by means of equation (5). Combining these information together,

[1] The only request is the existence of a stable fixed point and the boundedness of orbits to avoid escaping solutions.

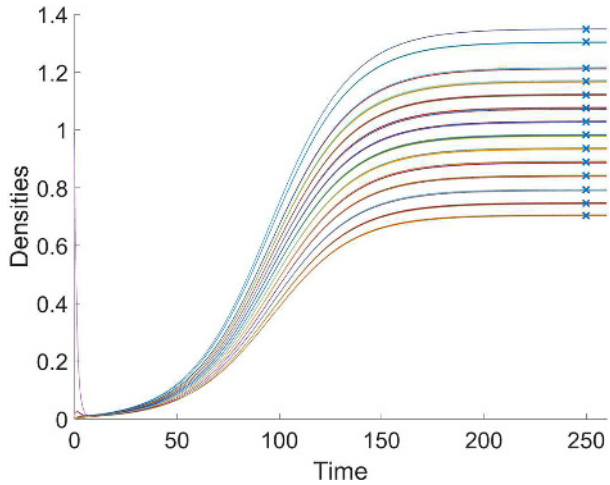

Fig. 1. Densities x_i are plotted against time, starting out-of-equilibrium and assuming a random Watts-Strogatz network with relocation probability $\beta = 0.99$ and $N = 200$ nodes. Solid lines refer to a direct integration of the governing equation (1), while symbols (crosses) stand for the HMF-based prediction, obtained following the procedure described in the main text. All 200 trajectories are displayed and cluster in families of homologous connectivity k, Each solid curve in the graph corresponds exactly to a given connectivity class. Here, $\alpha = 0.05$.

and recalling the definition of Θ, results in a linear problem for the unknown entries of the q-component vector $\vec{P} = (P(1) \dots P(q))$. In formulae:

$$\langle k \rangle \begin{bmatrix} \bar{\Theta}_1 \\ \vdots \\ \bar{\Theta}_q \end{bmatrix} = \underbrace{\begin{bmatrix} \bar{x}_1^{(1)} & \cdots & \bar{x}_q^{(1)} \\ \vdots & \ddots & \vdots \\ \bar{x}_1^{(q)} & \cdots & \bar{x}_q^{(q)} \end{bmatrix}}_{\doteq \mathbf{\Gamma}} \begin{bmatrix} P(1) \\ \vdots \\ P(q) \end{bmatrix}. \quad (7)$$

Solving the above problem for $(P(1) \dots P(q))$ implies inverting the matrix $\mathbf{\Gamma}$, a task that proved numerically cumbersome, being $\mathbf{\Gamma}$ poorly conditioned. To overcome this limitation we resorted to an optimization approach, which enforced the minimization of the norm $\|\vec{\Theta}\langle k \rangle - \mathbf{\Gamma}\vec{P}\|$, while imposing the entries of $\vec{P}$ to be positive defined. Here, $\vec{\Theta} = (\Theta_1 \dots \Theta_q)$. The average connectivity $\langle k \rangle$ is a priori unknown and it is therefore assumed, as a free control parameter in the optimization scheme. More specifically, we set $\langle k \rangle$ to a nominal value and run consequently the optimization protocol, recording as an output the quantity $\sum_{k'} P(k')$. An implicit requirement of the analysis that leads to (5) is the normalization of the distribution of connectivity, $\sum_{k'} P(k') = 1$. Among the solutions that are found by solving the problem in norm for different $\langle k \rangle$, we select the one that minimizes the positive residue $(1 - \sum_k P(k))^2$. By invoking this closure of the scheme, we also get an estimate for the average connectivity $\langle k \rangle$. This latter could be challenged against the true values in synthetic network model,

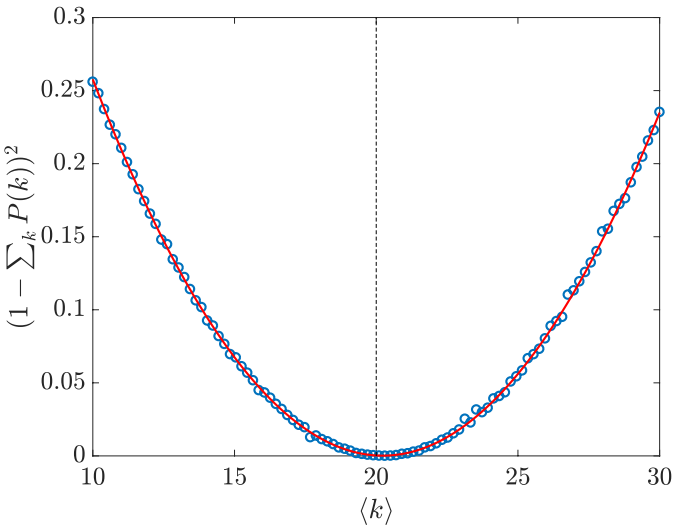

Fig. 2. The normalization error $\left(1-\sum_k P(k)\right)^2$ is plotted against the value of $\langle k \rangle$ imposed when running the inverse scheme. A clear minimum is displayed for $\langle k \rangle \simeq 20$, which is very close to the correct value of the average connectivity ($\langle k \rangle = 20$). The reconstruction procedure is hence able to single out the correct average connectivity as possessed by the network being analyzed. The solid line is obtained by fitting a parabola to the collected data.

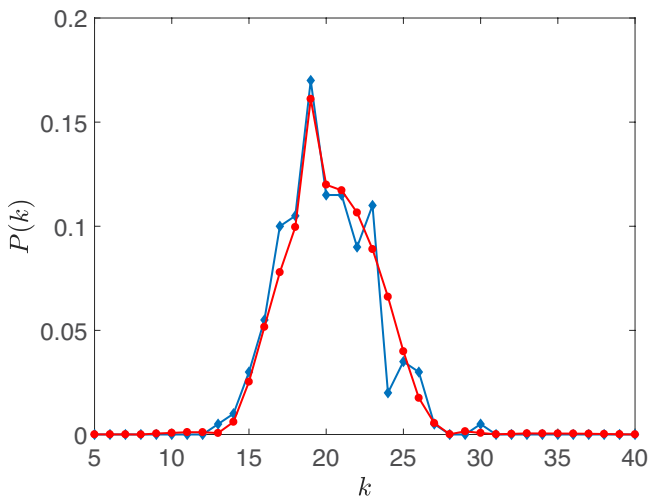

Fig. 3. The reconstructed distribution of connectivities: the blue line (with the diamond markers) represent the true degree distribution. The red line (with dot markers) stands for the distribution reconstructed via the procedure described in the main text. Here, 75 independent experiments are employed and the dynamics is sampled from just one node of the collection. The network is generated according to the Watts-Strogatz recipe with relocation probability $\beta = 0.99$. Here, $N = 200$.

in the aim of testing the adequacy of the proposed procedure.[2]

To this end we begin by considering the model (1) defined on a random network made of $N = 200$ nodes. The network is generated with the Watts-Strogatz recipe [33], for a large relocation probability, which makes the network almost random. The average connectivity is $\langle k \rangle = 20$. We performed $q = 75$ different measurements, sampling the dynamics on the very same node and letting α to change in uniform steps in the interval $[0.005, 0.4]$. In Figure 2 the normalization error $\left(1-\sum_k P(k)\right)^2$ is depicted against the imposed average connectivity. A clear minimum is displayed, for approximately the correct value of $\langle k \rangle$.

Setting $\langle k \rangle$ to the value that minimizes the normalization error returns the distribution of connectivity depicted in Figure 3. The blue line (with diamonds) stands for the true distribution, while the red curve (with dot markers) refer to the reconstructed profile. Changing the node from which the dynamics is sampled yields different estimates of the average connectivity $\langle k \rangle$ (and of the distribution that is consequently recovered). To provide a qualitative illustration of the degree of variability that stems from an arbitrary choice of the reference node, we plot in Figure 4a the histogram of $\langle k \rangle$, as obtained for all possible selections of the observation site. The distribution of predicted average connectivity is peaked around the correct solution. To further challenge the reconstruction scheme we also monitored the root mean square deviation between the reconstructed ($P(k)^{rc}$) and the exact ($P(k)^{ex}$) profiles. This latter quantity is labelled σ and formally defined as $\sigma = \sqrt{\sum_k \left(P(k)^{ex} - P(k)^{rc}\right)^2 / q}$. The histogram of the values of σ obtained when changing the observation nodes is depicted in Figure 4b and points to the adequacy of the proposed scheme. To improve on the accuracy of the method one can repeat the measurements on different sites and combine together the acquired information. This significantly improve on the ability of the HMF approximation to adhere on the exact asymptotic solution, as seen in direct simulations. In Figure 5 the reconstruction procedure is tested for a Watts Strogatz network with a smaller relocation probability and the quality of the reconstruction is still satisfying.

To further elaborate on the potential of the described procedure we considered a scale free network made of $N = 500$ nodes. More specifically $P(k) \propto k^{-\gamma}$ with $\gamma = 3$. The $P(k)$ is correctly recovered following the method, as displayed in Figure 6. The distribution of $\langle k \rangle$ as obtained by sampling the dynamics on all possible observation sites is shown in Figure 7a. At variance to the case of the random network analyzed above, changing the reference node impacts more significantly on the recorded $P(k)$, as seen from the histogram of σ in Figure 7b (while $\langle k \rangle$ is correctly estimated). However, we could not identify any significant correlation between the error displayed by the reconstruction and the topological characteristics of the observation node (as e.g. its degree). Averaging independent profiles obtained by sampling the dynamics from distinct nodes contributed to enhance the fidelity of the reconstruction.

Summing up to this point, we have introduced and successfully applied a procedure to reconstruct the distribution of connectivity of an unknown network by

[2] In principle, one could absorb $\langle k \rangle$ in the definition of P_k, compute the rescaled entries $\tilde{P}_k = P_k/\langle k \rangle$ via the linear problem and enforce a posteriori the normalization. This scheme proved however less stable that the one that we have illustrated in the main body of the paper.

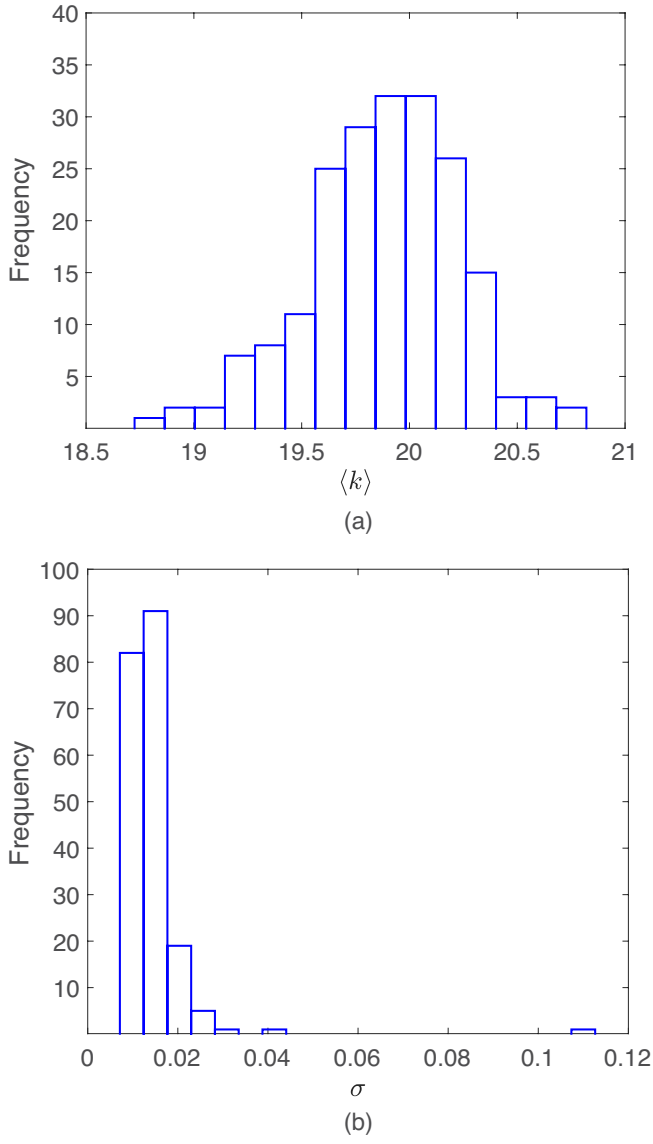

Fig. 4. (a) The histogram of the averaged connectivity $\langle k \rangle$ is displayed. Each value of $\langle k \rangle$ refers to a different selection of the node from which the dynamics is sampled. (b) The histogram of the root mean square error in reconstruction σ is plotted. Each value of σ refers to different observation nodes.

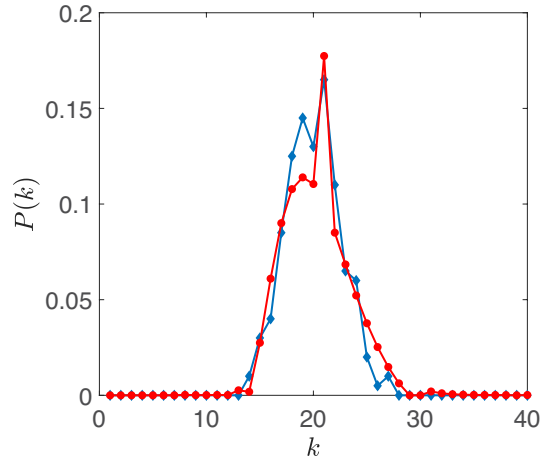

Fig. 5. The reconstructed distribution of connectivities, for a Watts-Strogatz network with relocation probability $\beta = 0.5$ and $N = 200$. Symbols are chosen as explained in the caption of Figure 3.

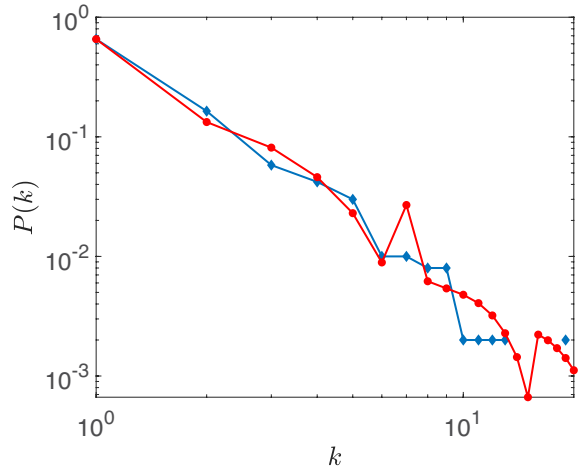

Fig. 6. The reconstructed distribution, assuming a scale free network. The network consists of 500 nodes. Here, $P(k) \propto k^{-\gamma}$ with $\gamma = 3$. Symbols follows the convention introduced in Figure 3.

performing local measurements on just on node. The inspected system combines nonlinear reactions to relocation. The method implemented requires adjusting a control parameter which sets the relative strength of the two aforementioned contributions, a condition that might prove difficult to meet in real experiments. To overcome this limitation and formulate a reconstruction scheme which could be viably implemented, we will hereafter consider a slightly modified version of the examined dynamical process. More precisely, we will accommodate for the presence of a set of sources, where the interacting elements can be injected at a given rate to be externally tuned.

4 Introducing the sources and modulating their strengths

The updated scheme builds on the following steps. First, we freeze α to a constant amount. Then, we modify the dynamics of a given class of nodes, say those characterized by a connectivity $\hat{k}$, by introducing an ensemble made of identical sources, characterized by a constant injection strength η. Mathematically, this corresponds to inserting on the right hand side of equation (1) a constant factor η, for the entries i which identify the selected pool of nodes. For any given choice of η, the asymptotic fate of the system can be analytically investigated by proceeding with the HMF approximation, in analogy with the results reported above. Formally, one ends up with an expression for the fixed point which trivially extends that displayed

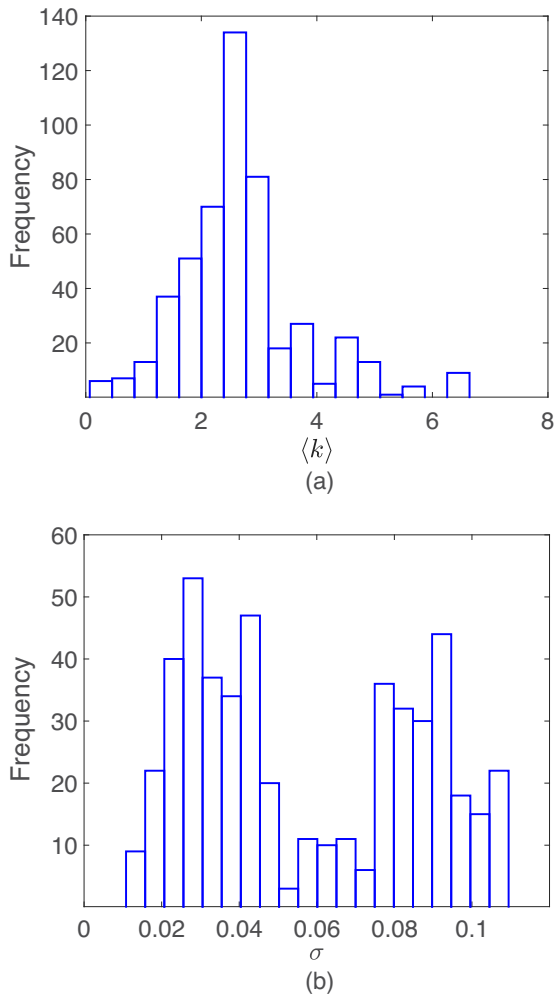

Fig. 7. (a) The histogram of the reconstructed averaged connectivity $\langle k \rangle$ is plotted, for a scale free network. Each value of $\langle k \rangle$ refers to a different selection of the node from which the dynamics is sampled. (b) The histogram of the root mean square error in reconstruction σ is represented. Each value of σ refers to different observation nodes.

in equation (5):

$$\bar{x}_k = \frac{(2\alpha - 1) \pm \sqrt{(2\alpha - 1)^2 + 4\alpha[(1-\alpha)k\bar{\Theta} + \eta\delta_{k\hat{k}}]}}{2\alpha} \tag{8}$$

where $\delta_{k\hat{k}}$ stands for the Kronecker delta.

The reconstruction scheme can be hence modified as follows: (i) change η within a given interval; (ii) for each choice of η, measure the asymptotic state attained by the system on a given node with degree $k = k^*$; (iii) use this knowledge to estimate the mean-field variable $\bar{\Theta}$, by inversion of equation (8) and, finally, (iv) predict the equilibrium solution $\bar{x}_k$, $\forall k \neq k^*$. Repeating this procedure for a sufficiently large set of distinct values of η, yields a linear problem of the type (7) which can be solved in norm to compute $\vec{P} = (P(1) \dots P(q))$. Also in this case the average connectivity is estimated by minimizing the residual error. As a demonstrative example, we compare in Figure 8 the reconstructed $P(k)$ to its exact homologue. Here, the underlying network is generated according to the Watts-Strogatz recipe and the agreement between the two depicted curves is satisfying.

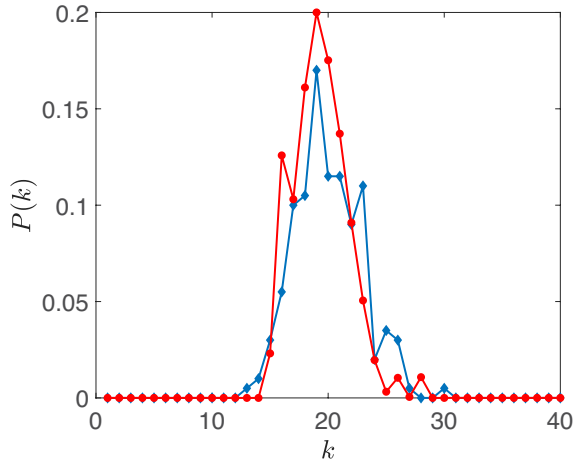

Fig. 8. The reconstructed distribution of connectivities: the blue line (with the diamond markers) stands for the true degree distribution. The red line (with dot markers) identifies the distribution reconstructed via the procedure described in the main text. Here, $\alpha = 0.005$ and η is changed in the interval $[0.005, 5.5]$ with uniform increments of 0.2. The sources are placed on the nodes sharing degree $\hat{k} = 15$ (the results do not depend on this specific choice). This class contains 7 different nodes, for the network realization here considered. The network employed is that of Figure 3.

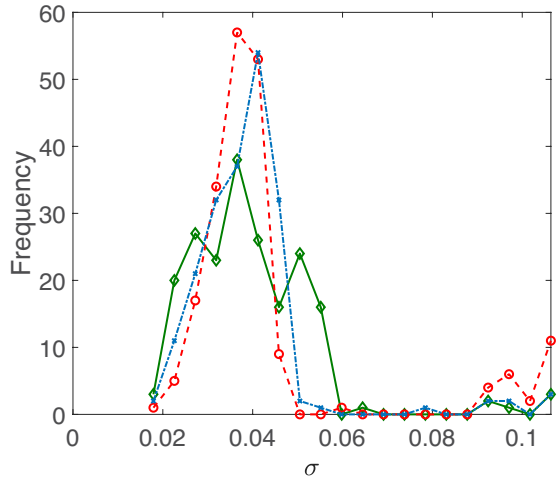

Fig. 9. The histogram of σ is plotted. Each curve refers to a specific choice of the source nodes. More specifically, the green line with diamond markers is obtained by setting the sources on the nodes that share connectivity $\hat{k} = 15$. The red dashed line with circle markers refers to $\hat{k} = 20$ and the dash dotted blue line with crosses stands for $\hat{k} = 25$. For each sources selection, we perform the reconstruction by sampling the dynamics on different observation nodes and quantify the associated error σ.

In Figure 9 we report the histogram of σ, the root mean square error of the reconstructed profile against

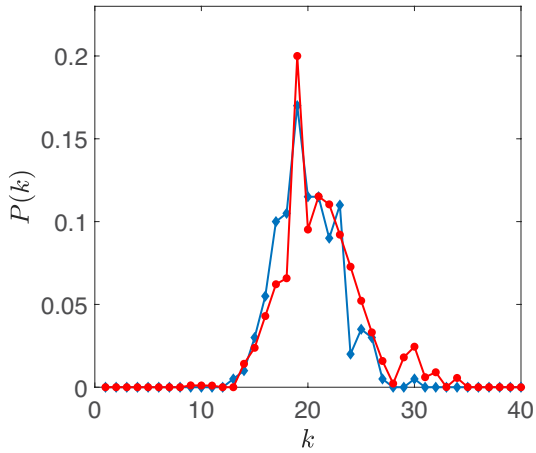

Fig. 10. The reconstructed distribution of connectivities: the blue line (with the diamond markers) stands for the true degree distribution. The red line (with dot markers) identifies the distribution reconstructed via the procedure described in the main text. Here, $\alpha = 0.005$ and η is changed in the interval $[0.005, 0.35]$, with 80 successive uniform increments. The source is placed on one of the nodes sharing degree $\hat{k} = 15$. The network employed is that of Figure 3.

the exact one. Each curve refers to a different choice of the nodes that act as sources. The corresponding ensemble of σ is obtained by changing the observation node, for each source selection. The results are quite consistent and point to the overall validity of the proposed method. The connectivity of the selected sources is, respectively, $\hat{k} = 15, 20, 25$, and it does not influence significantly the performance of the reconstruction algorithm.

As a final point, we will relax the assumption of dealing with a full class of nodes which behave as identical sources. More precisely, we will break the symmetry and assume that just one node of a given class $\hat{k}$ acts as a source. To handle this generalized setting, we revisit the definition of the mean field variable. Recall that $P(\hat{k})$ measures the number of nodes sharing connectivity $\hat{k}$ normalized to the system size N. Then one can define the following collective variable:

$$\Pi = \frac{1}{\langle k \rangle} \left[\sum_{k'} P(k') x_{k'} + \frac{y - x_{\hat{k}}}{N} \right] \tag{9}$$

where y denotes the value of the density at the source location. One can apply the HMF machinery[3] to yield:

$$\bar{x}_k = \frac{(2\alpha - 1) \pm \sqrt{(2\alpha - 1)^2 + 4\alpha(1-\alpha)k\Pi}}{2\alpha} \tag{10}$$

for the fixed point concentration as predicted on all nodes, but the source. This latter is characterized by an

[3] In doing so we postulate that the inserted source does not disrupt the organization in classes. The validity of this working ansatz is confirmed a posteriori by the quality of the obtained reconstruction.

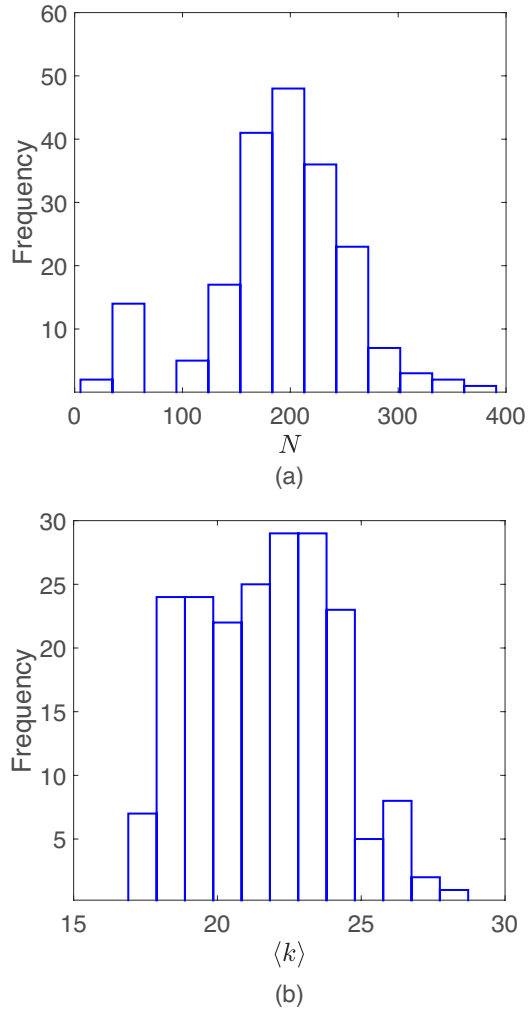

Fig. 11. (a) The histogram of N, the reconstructed network size. Each value of N refers to distinct observation node. (b) The histogram of the reconstructed averaged connectivity $\langle k \rangle$ is plotted. Each value of $\langle k \rangle$ refers to a different selection of the node from which the dynamics is sampled.

asymptotic density given by:

$$\bar{y} = \frac{(2\alpha - 1) \pm \sqrt{(2\alpha - 1)^2 + 4\alpha[(1-\alpha)k\Pi + \eta]}}{2\alpha}. \tag{11}$$

Building on the above, one can put forward a straightforward generalization of the reconstruction algorithm. For any given choice of the source injection rate η, one can measure the equilibrium density as displayed on one individual node, belonging to class k^*. By manipulating equation (10), one can then estimate Π and use this latter to predict the expected density on each of the classes ($\bar{x}_k$, $k \neq k^*$) and on the source ($\bar{y}$). To this end one makes explicit use of, respectively, equations (10) and (11). Repeating the analysis for a sufficiently large set of η returns a linear problem of the type discussed above, with the sole difference that now the size of the network N qualifies as one of the unknowns to be eventually recovered. In formulae, equation (9) can be cast in

the form:

$$\langle k\rangle \Pi = \sum_{k'} P(k') x_{k'} + \frac{z}{N}$$

where $z = y - x_{\hat{k}}$. We hence get:

$$\langle k\rangle \begin{bmatrix} \bar{\Pi}_1 \\ \vdots \\ \bar{\Pi}_q \end{bmatrix} = \begin{bmatrix} \bar{x}_{k_1}^{(1)} & \cdots & \bar{x}_q^{(1)} & \bar{z}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \bar{x}_1^{(q)} & \cdots & \bar{x}_q^{(q)} & \bar{z}_q \end{bmatrix} \begin{bmatrix} P(1) \\ \vdots \\ P(q) \\ \frac{1}{N} \end{bmatrix} \tag{12}$$

where q ($>k_{\max}$) stands for the number of repeated measurements performed at different choices of η. The reconstruction obtained following this updated strategy is displayed in Figure 10, for the same realization of the Watts-Strogatz network as considered above. The quality of the reconstruction is still adequate and the estimated value of N is in agreement with the true one. The performance of the algorithm may change depending on the node of observation but the distribution of N obtained when covering the full set of possible choices is peaked in correspondence of the correct value as shown in Figure 11a. Similarly, in Figure 11b the distribution of $\langle k\rangle$, the average value of the reconstructed connectivity, extends over a finite domain which contains the correct value.

5 Conclusion

In conclusion, we have here introduced and tested a procedure to access structural information on network topology. The method samples the dynamics of reactive walkers, microscopic entities which are made to explore the embedding network, while subject to nonlinear (and local) reaction terms. The imposed non-linearity makes it possible to recover the network's distribution of connectivity, from sequences of measurements performed on just one node of the collection. To reach this goal, we exploit the organization in classes of the ensuing dynamical equilibrium and make explicit use of the celebrated Heterogeneous Mean Field approximation. A variant of the method which consists in introducing localized sources of modulable strength, enables in turn to estimate the size of the scrutinized network. Future investigations will be aimed to improving on the optimization scheme and, consequently, on testing the predictive ability of the proposed techniques versus more challenging network architecture, including multiplex. Moreover, it would be also interesting to elaborate on possible generalization of the method so as to account for corrections beyond the simplified Heterogeneous Mean Field approximation.

Author contribution statement

I.A., D.F., T.C. and G.I. conceptualized the work and performed the analytical calculations. I.A. carried out the numerical study. All authors contributed to the discussion of the results and to the writing of the paper.

References

1. M.E.J. Newman, *Networks: An Introduction* (OUP, Oxford, 2010)
2. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Phys. Rep. **424**, 175 (2006)
3. R. Albert, A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002)
4. A. Barrat, M. Barthélemy, A. Vespignani, in *Dynamical Processes on Complex Networks*, 1st edn. (CUP, Cambridge, 2008)
5. G. Caldarelli, in *Scale Free Networks: Complex Webs in Nature and Technology*, 1st edn. (Oxford Finance Series, 2007)
6. J.D. Noh, H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004)
7. R. Sinatra, J. Gómez-Gardeñes, R. Lambiotte, V. Nicosia, V. Latora, Phys. Rev. E **83**, 030103(R) (2011)
8. V. Nicosia, F. Bagnoli, V. Latora, Europhys. Lett. **94**, 68009 (2011)
9. N. Masuda, M.A. Porter, R. Lambiotte, Phys. Rep. **716–717**, 1 (2017)
10. S. Manfredi, E. Di Tucci, V. Latora, arxiv.1801.01332 (2017)
11. E. Schneidman, M.J. Berry, R. Segev, W. Bialek, Nature **440**, 04701 (2006)
12. S. Coccoa, S. Leiblerb, R. Monasson, PNAS **106**, 14058 (2009)
13. R. Burioni, M. Casartelli, M. di Volo, R. Livi, A. Vezzani, Sci. Rep. **4**, 4336 (2014)
14. M. di Volo, R. Burioni, M. Casartelli, R. Livi, A. Vezzani, Phys. Rev. E **90**, 022811 (2014)
15. M. di Volo, R. Burioni, M. Casartelli, R. Livi, A. Vezzani, Phys. Rev. E **93**, 012305 (2016)
16. L. Michiels van Kessenich, L. de Arcangelis, H.J. Herrmann, Sci. Rep. **6**, 32071 (2016)
17. R. Russo, H.J. Herrmann, L. de Arcangelis, Sci. Rep. **4**, 4312 (2014)
18. S.G. Shandilya, M. Timme, New J. Phys. **13**, 013004 (2011)
19. I. Malvestio, T. Kreuz, R.G. Andrzejak, Phys. Rev. E **96**, 022203 (2017)
20. M. Asllani, T. Carletti, F. Di Patti, D. Fanelli, F. Piazza, Phys. Rev. Lett. **120**, 158301 (2018)
21. T.M. Ligget, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, 1st edn. (Springer-Verlag, Berlin, 1999)
22. E. Almaas, R.V. Kulkarni, D. Stroud, Phys. Rev. E **68**, 056105 (2003)
23. S. Kwon, Y. Kim, Phys. Rev. E **84**, 041103 (2011)
24. D. Fanelli, A.J. Mckane, Phys. Rev. E **82**, 021113 (2010)
25. M. Galanti, D. Fanelli, F. Piazza, Front. Phys. **4**, 33 (2016)
26. A.E. Fernando, K.A. Landman, M.J. Simpson, Phys. Rev. E **81**, 011903 (2010)
27. K.A. Landman, A.E. Fernando, Physica A **390**, 3742 (2011)
28. M. Galanti, D. Fanelli, A. Maritan, F. Piazza, Europhys. Lett. **107**, 20006 (2014)
29. M. Galanti, S. Traytak, D. Fanelli, F. Piazza, Phys. Chem. Chem. Phys. **18**, 15950 (2016)
30. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008)
31. F. Pittorino, M. Ibanez Berganza, M. di Volo, A. Vezzani, R. Burioni, Phys. Rev. Lett. **118**, 098102 (2017)
32. G. Cencetti, F. Battiston, D. Fanelli, V. Latora, Phys. Rev. E **98**, 052302 (2018)
33. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)