

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### Multigraph reconstruction via nonlinear random walk

de Kemmeter, Jean-François; Carletti, Timoteo

*Published in:*

Journal of Complex Networks

*DOI:*

[10.1093/comnet/cnae038](https://doi.org/10.1093/comnet/cnae038)

*Publication date:*

2024

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (HARVARD):*

de Kemmeter, J-F & Carletti, T 2024, 'Multigraph reconstruction via nonlinear random walk', *Journal of Complex Networks*, vol. 12, no. 5, cnae038. <https://doi.org/10.1093/comnet/cnae038>

#### 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.

# Multigraph reconstruction via nonlinear random walk

Jean-François de Kemmeter<sup>1,\*</sup> and Timoteo Carletti<sup>1,+</sup>

<sup>1</sup>Department of Mathematics and Namur Institute for Complex Systems, naXys, University of Namur, rue Grafé 2, B 5000 Namur, Belgium

\* Email: jean-francois.dekemmeter@unamur.be

+ Email: timoteo.carletti@unamur.be

## Abstract

Over the last few years, network science has proved to be useful in modeling a variety of complex systems, composed of a large number of interconnected units. The intricate pattern of interactions often allows the system to achieve complex tasks, such as synchronization or collective motions. In this regard, the interplay between network structure and dynamics has long been recognized as a cornerstone of network science. Among dynamical processes, random walks are undoubtedly among the most studied stochastic processes. While traditionally, the random walkers are assumed to be independent, this assumption breaks down if nodes are endowed with a finite carrying capacity, a feature shared by many real-life systems. Recently, a class of nonlinear diffusion processes accounting for the finite carrying capacities of the nodes was introduced. The stationary nodes densities were shown to be nonlinearly correlated with the nodes degrees, allowing to uncover the network structure by performing a few measurements of the stationary density at the level of a single arbitrary node and by solving an inverse problem. In this work, we extend this class of nonlinear diffusion processes to the case of multigraphs, in which links between nodes carry distinct attributes. Assuming the knowledge of the pattern of interactions associated with one type of links, we show how the degree distribution of the whole multigraph can be reconstructed. The effectiveness of the reconstruction algorithm is demonstrated through simulations on various multigraph topologies.

## 1 Introduction

We are surrounded by networks [1,2]. Physics [4], economy [5], biology [6] and sociology [7], are few but relevant research domains that can be studied in the

vast realm of network science. Notwithstanding the differences and the peculiarities of each domain, scholars have been able to provide a transversal and unified description of those networked systems in which individual units (e.g., chemicals, bits of information, nutrients) belonging to a certain population may move across network nodes by using the intricate web of available links, the latter defining the structure of a complex network. Those processes eventually give rise to the emergence of self-organized complex patterns [3,8–10], the latter depending on the network architecture. Hence, the bridge between network structure and dynamics can be explored by studying the evolution of simple processes; a stereotypical example is provided by a random walk, i.e., a simple model of diffusion, where basic units jump from node to node according to some microscopic rules and using available links.

Random walks play a central role in different fields of science [11–13]; once dealing with networks, a central research theme is devoted to studying the relation between patterns of diffusion and network structure [14], with relevant applications to centrality measures related to walkers density [17,21], e.g., Page rank, or community detection methods based on time of visit, e.g., Markov stability [18–20]. In those applications, walkers are assumed to be independent from each other, a condition that can be safely assumed to hold true once no competition for the available space is at play. To overcome this limitation, a novel class of nonlinear diffusion processes on networks has been recently studied, accounting for the finite carrying capacities of nodes [22,23]. Such processes incorporate the limited available space within the nodes, a feature of most real-life networks. Take for instance the case of an ecological network in which nodes would model local habitat patches. The latter certainly cannot accommodate for an infinite number of individuals due to competition for limited resources. In other words, setting a constraint on the maximal number of agents allowed to sit on a node, naturally induces correlations between the random walkers. As a consequence, the stationary node densities are no longer simply proportional to the node degree [24]. In some cases though, an explicit expression for the stationary density can still be worked out and thus, by taking advantage of the mass conservation, the degree distribution of a graph can be inferred from a few measurements performed at a *single* arbitrarily selected node [23]. An other approach based on an ensemble of reactive walkers performing linear diffusion was considered in [25]. The latter work has some similarities with the present one. Indeed, both rely on the use of nonlinear random walks to infer the structure of the underlying network, by performing measurements on a single node (or family of nodes). They however also show relevant differences. First of all, the random walk in [25] is driven by the standard random walk Laplace operator, defined on a simple network, while in the present case, a new nonlinear operator is used on a multigraph. Finally, in this work we develop a solution at the nodes level while in [25] authors used the heterogeneous mean-field assumption [26–28].

The purpose of this work is to extend the approach introduced in [23] to the case of multigraph networks, where two nodes can be connected by several links, each one associated with a different attribute. Examples include air transport networks in which air routes (links) between airports (nodes) carry as attribute

the name of the operating company, or social networks in which individuals (nodes) interact through several platforms (links). Assuming a finite set of distinct attributes, the corresponding non-simple networks can be represented as edge-coloured multigraphs - multigraphs for short -, each color mapped bijectively to a given value of the attribute.

In several relevant applications, the network structure is only partially known. This might be the case of social networks in which only a fraction of relationships is available or protein-protein interaction networks in which the identification of interactions is costly [29]. Methods able to uncover the network structure from partial knowledge result thus important. With this regard, link prediction has attracted increasing attention these last years. The purpose of link prediction is to compute the likelihood of a link between two nodes, based on observed links and nodes' attributes. The latter might be structural properties such as their degree or external information, such as age or hobby if nodes refer to individuals. Various algorithms have been designed for performing link prediction, among which similarity-based algorithms and probabilistic models [15,16]. Some other models take advantage of the interplay between network structure and dynamics. In particular, random walks allow for the identification of central nodes and the detection of communities, groups of nodes more tightly connected among them than with nodes outside the community [20].

In this work, we restrict our attention to multigraphs with two distinct attributes, i.e., two colors. We first extend the theory of non-linear diffusion process considered in [22,23] to the case of multigraphs, by assuming nodes of the latter to be endowed with a limiting carrying capacity. We explicitly determined the stationary node density and by using the fact that the latter at the level of a single node depends on the degree distribution of the whole multigraph, we are able to extend the reconstruction scheme proposed in [23] to the case of multigraphs. We investigate various multigraph topologies, assessing in each case, the accuracy of the reconstruction procedure. More precisely we consider multigraphs where the known and unknown layers are selected among an Erdős-Rényi graph, a Watts-Strogatz network or a scale-free one; then we reconstruct the first and second degree momenta of the unknown layer and we compare them with the ground truth as a function of the main model parameter, i.e., the probability to have a link among two nodes in the Erdős-Rényi graph, the rewiring probability in the case of Watts-Strogatz and the power law exponent for the scale-free case. As a general conclusion we can assert that with local and partial information we are able to reconstruct quite well the degree probability distribution and the degree momenta, the first momentum being generally closer to the right value than the second momentum. We can also claim that our results do not suggest any strong impact of the known layer on the reconstruction of the second, let us however stress that this can be a consequence of the made assumption of absence of correlations among node degrees. In the case the unknown network is an Erdős-Rényi graph we observe a degradation of the results as it becomes more sparse and in particular nodes with degree zero appear. This observation is also confirmed in the study of the real multigraphs we carried out, where one network possesses nodes with zero degree, the whole multigraph being still connected thanks to the

second network.

Moreover, still starting from the exact formula for the stationary node density we proposed a second method to reconstruct node degree momenta based on the assumption of diluted walkers density. The method requires less information than the previous one and still, it is able to reconstruct the first two momenta with a quite good precision.

In conclusion in this work we have thus strengthened the link existing between network topology and dynamical systems defined on top of networks, by proposing a model capable to infer global structural information from a multigraph with a limited amount of information, e.g., knowing the degree distribution for a single layer and measuring the random walk frequency of transit on a single node, designing in this way further studies of links prediction.

The paper is organized as follows. In Section 2, we introduce the non-linear diffusion process and derive the analytical formula for nodes' stationary densities. In Section 3, we explain how the multigraph structure can be recovered as the solution of an inverse problem based on the knowledge of the stationary node density at a single arbitrary node. In Section 4, we test the algorithm on various multigraph topologies and assess the robustness of the reconstruction. We then conclude and give some perspectives in Section 5.

## 2 The model

In this section, we introduce the nonlinear diffusion process evolving on a multigraph. The derivation is obtained from a microscopic formulation of the dynamics and builds on previous works [22, 23]. We start by considering a set  $V$  of  $\Omega$  nodes and a set  $E$  of edges (see Fig. 1 for a cartoon example). Links between nodes are distinguished by some attribute  $\mathcal{A}$  which we assume to take only two distinct values, say  $\mathcal{A} \in \{1, 2\}$ . We will denote by  $E_1$  (respectively  $E_2$ ) the subset of edges associated with the attribute 1 (respectively 2). The multigraph  $G = (V, E)$  is then obtained as the union of the two subgraphs  $G_1 = (V, E_1)$  and  $G_2 = (V, E_2)$ . The structure of each of the two subgraphs is encoded by two adjacency matrices,  $A^{(1)}$  and  $A^{(2)}$ , with  $A_{ij}^{(\ell)} = 1$  if nodes  $i$  and  $j$  are connected through a link with attribute  $\mathcal{A} = \ell$ .

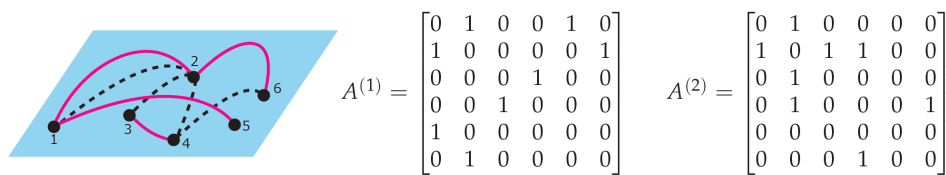


Figure 1: Multigraph with two types of links, distinguished in plain magenta and dashed black. The adjacency matrices  $A^{(1)}$  and  $A^{(2)}$  of the subgraphs  $G_1$  and  $G_2$  are indicated.

Given some  $\beta \in (0, 1)$ , we then assume a set of  $\lfloor \beta N \Omega \rfloor$  agents, also named walkers, to diffuse across the network under the constraint that at most  $N$  of them are found simultaneously in any given node. The parameter  $\beta$  thus measures the density of walkers within the network. The state of the system at time  $t$  may be recorded as a vector  $\mathbf{n}(t) := (n_1(t), n_2(t), \dots, n_\Omega(t))^\top$  with  $0 \leq n_i(t) \leq N$  being the number of walkers found in node  $i$  at time  $t$ . Denoting by  $\mathbb{P}(\mathbf{n}, t)$  the probability that the system is in state  $\mathbf{n}$  at time  $t$ , the latter obeys the *M-equation*:

$$\frac{d\mathbb{P}(\mathbf{n}, t)}{dt} = \sum_{\mathbf{n}' \neq \mathbf{n}} \{T(\mathbf{n}|\mathbf{n}')\mathbb{P}(\mathbf{n}', t) - T(\mathbf{n}'|\mathbf{n})\mathbb{P}(\mathbf{n}, t)\}, \quad (1)$$

being  $T(\mathbf{n}'|\mathbf{n})$  the transition probability from state  $\mathbf{n}$  to state  $\mathbf{n}'$ . As any node may host from 0 to  $N$  agents at a given time, the terms  $T(\mathbf{n}'|\mathbf{n})$  may be seen as the entries of a huge matrix of size  $(N+1)^\Omega$ . The matrix is essentially sparse as during the infinitesimal time interval  $[t, t+dt]$ , at most one random walker will have moved from a node to one of its neighbors. Assuming the departure node, say  $i$ , to contain  $n_i$  walkers and the arrival node, say  $j$ , to contain  $n_j$  ( $\leq N-1$ ) walkers, the corresponding transition probability is written as  $T(n_i-1, n_j+1|n_i, n_j)$ . To write down the latter, let us first notice that the probability for the walker sitting at node  $i$  to choose, among all the available connections incident to that node, a link leading to node  $j$ , is given by  $\frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_i}$ , with  $k_i = \sum_j \{A_{ij}^{(1)} + A_{ij}^{(2)}\}$  the number of connections incident to node  $i$ , regardless of their corresponding attributes. The probability for the walker to settle onto the arrival node is then modulated by the available space within node  $j$ : the more crowded the node  $j$ , the less likely the transition. Hence, the overall transition probability reads

$$T(n_i-1, n_j+1|n_i, n_j) = \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_i} \frac{n_j}{N} g\left(\frac{n_j}{N}\right), \quad (2)$$

where the function  $g$  takes into account the available space within the arrival node. Let us recall that the case of uncorrelated walkers is recovered with  $g(x) = 1$ . In contrast, we here impose  $g(1) = 0$  to forbid a walker to hop onto a fully occupied node. A common choice is to take  $g(x) = 1 - x$ . To make further progress in the analytical treatment, we now assume the nodes carrying capacities  $N$  to be large enough and introduce the continuous variable, also referred to as “node density”,  $\rho_i(t) := \lim_{N \rightarrow +\infty} \frac{\langle n_i(t) \rangle}{N}$ , where  $\langle \cdot \rangle$  denotes the average over stochastic realizations of the system. As shown in Appendix A, the time evolution of nodes densities is governed by the following system of ordinary differential equations.

$$\begin{aligned} \frac{d\rho_i}{dt} &= \sum_j \left\{ \rho_j \frac{A_{ji}^{(1)} + A_{ji}^{(2)}}{k_j} g(\rho_i) - \rho_i \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_i} g(\rho_j) \right\} \\ &= \sum_j \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_j} \left\{ \rho_j g(\rho_i) - \frac{k_j}{k_i} \rho_i g(\rho_j) \right\}, \end{aligned} \quad (3)$$

where the second equality has been obtained by exploiting the fact that the multigraph is undirected. Introducing the random walk Laplacian matrix  $L$  with elements

$$L_{ij} = \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_j} - \delta_{ij} \quad (4)$$

allows to rewrite the system of ODEs in a compact form, namely

$$\frac{d\rho_i}{dt} = \sum_j L_{ij} \left\{ \rho_j g(\rho_j) - \frac{k_j}{k_i} \rho_i g(\rho_j) \right\}. \quad (5)$$

The reconstruction scheme builds on the knowledge of the stationary node densities  $\rho_i^*$  [23]. The latter are solution of:

$$0 = \sum_j L_{ij} f_j(i) \quad \forall i = 1, \dots, \Omega, \quad (6)$$

where  $f_j(i) = \rho_j g(\rho_j) - \frac{k_j}{k_i} \rho_i g(\rho_j)$ . It thus follows that the vector  $(f_1(i), \dots, f_\Omega(i))$  should be proportional to the right eigenvector associated to the null eigenvalue<sup>1</sup> of the random walk Laplacian  $L$ . This eigenvector is proportional to the degree vector  $(k_1, \dots, k_\Omega)^\top$ . Hence, we must have:

$$f_j(i) = C_i k_j \quad \forall i, j = 1, \dots, \Omega, \quad (7)$$

for some constant  $C_i$ . Considering the particular case  $j = i$  for which  $f_i(i) = 0$ , we conclude that  $C_i = 0 \quad \forall i$ . As a consequence, the stationary node densities  $\rho_i^*$  are solution of:

$$\rho_j^* g(\rho_j^*) - \frac{k_j}{k_i} \rho_i^* g(\rho_j^*) = 0 \quad \forall i, j, \quad (8)$$

Let us note that, apart from the trivial case  $\beta = 0$  for which there is no walker diffusing across the nodes, the stationary node densities are all positive, i.e.,  $\rho_i^* \neq 0 \quad \forall i$ . Indeed if node  $i$  had a zero stationary node density ( $\rho_i^* = 0$ ), then all the other nodes should also have a zero stationary node density, which would imply  $\beta = 0$ . For  $\beta > 0$ , one may thus rewrite Eq. (8) as follows

$$\frac{k_i g(\rho_i^*)}{\rho_i^*} = \frac{k_j g(\rho_j^*)}{\rho_j^*} \quad \forall i, j, \quad (9)$$

from which the stationary node densities are found to satisfy

$$k_i g(\rho_i^*) = c \rho_i^*, \quad (10)$$

for some constant  $c \equiv c(\beta)$  that is uniquely fixed by mass conservation, i.e.,  $\sum_i \rho_i^* = \beta \Omega$ .

---

<sup>1</sup>Without loss of generality, we here assume the (undirected) multigraph to be connected. In that case, the random walk Laplacian  $L$  has a single null eigenvalue. If the multigraph had multiple disconnected components, there would be as many null eigenvalues as disconnected components and the analysis should be applied separately to any of these components.

### 3 Uncovering the network structure

Previous work has shown how the above presented nonlinear diffusion process allows to uncover the degree distribution of the corresponding (unknown) graph structure [23]. We now review and extend their approach by considering, for the sake of pedagogy, the case of multigraphs with two types of links; let us however observe that the method is general enough to deal with more than two attributes for the links. More specifically, we will assume the subgraph associated with the attribute  $\mathcal{A} = 1$  to be known and will aim at uncovering structural properties of the subgraph associated with the other attribute  $\mathcal{A} = 2$ , under the assumption that the degree distributions of both subgraphs are uncorrelated. The approach relies on an explicit expression of the stationary densities. From now on, we assume  $g(x) = 1 - x$ . For such choice, the stationary node densities read

$$\rho_i^* = \frac{c(\beta)k_i}{1 + c(\beta)k_i}. \quad (11)$$

The mass conservation then implies

$$\sum_i \frac{c(\beta)k_i}{1 + c(\beta)k_i} = \beta\Omega, \quad (12)$$

or, equivalently,

$$\sum_{k^{(1)}, k^{(2)}} p(k^{(1)}, k^{(2)}) \frac{c(\beta)[k^{(1)} + k^{(2)}]}{1 + c(\beta)[k^{(1)} + k^{(2)}]} = \beta, \quad (13)$$

where  $p(k^{(1)}, k^{(2)})$  is the probability that a node will have  $k^{(1)}$  links with attribute  $\mathcal{A} = 1$  and  $k^{(2)}$  links with attribute  $\mathcal{A} = 2$ . The degree distributions of the two subgraphs being uncorrelated, the joint probability distribution function factorizes, i.e.,  $p(k^{(1)}, k^{(2)}) = p_1(k^{(1)})p_2(k^{(2)})$ , leading to:

$$\sum_{k^{(2)}} \left( \sum_{k^{(1)}} p_1(k^{(1)}) \frac{c(\beta)[k^{(1)} + k^{(2)}]}{1 + c(\beta)[k^{(1)} + k^{(2)}]} \right) p_2(k^{(2)}) = \beta. \quad (14)$$

The latter expression holds for any arbitrary fixed value of  $\beta$ . Consider thus to perform  $M$  experiments (for some integer  $M$ ) each one involving a different number of agents,  $\lfloor \beta_i N \Omega \rfloor$ , walking on the multigraph and assume to measure for each experiment the asymptotic average number of walkers sitting on a randomly chosen node (fixed once for all <sup>2</sup>) and eventually obtain  $c(\beta)$  by using

---

<sup>2</sup>Let us observe that for each value of  $\beta_i$ , one could in principle choose an arbitrary node to measure  $\rho_i^*$ , since from Eq. (11) it is clear that the constant  $c(\beta_i)$  can be recovered if we know  $\rho_i^*$  and  $k_i$ . However this would assume a larger knowledge of the network with respect to the case hereby assumed where we only have access to the stationary node density of a given node, yet for distinct values of  $\beta$ .

Eq. (11). Recalling that the architecture of the network with attribute  $\mathcal{A} = 1$  is known, we can define the matrix with known entries:

$$F_{ij} = \sum_{k^{(1)}} p_1(k^{(1)}) \frac{c(\beta_i)[k^{(1)} + j]}{1 + c(\beta_i)[k^{(1)} + j]}. \quad (15)$$

Finally by introducing the vector  $\boldsymbol{\beta} := (\beta_1, \beta_2, \dots, \beta_M)^\top$  and writing  $\boldsymbol{p}_2 := (p_2(1), p_2(2), \dots, p_2(M))^\top$  for the unknown degree distribution of the network with attribute  $\mathcal{A} = 2$ , we can rewrite (14) as

$$F\boldsymbol{p}_2 = \boldsymbol{\beta}, \quad (16)$$

The probability distribution  $\boldsymbol{p}_2$  can thus be obtained by solving the linear system (16). Observe that inverting the matrix  $F$  is not recommended, since the matrix  $F$  is close to singular. Rather, the solution will be obtained by a least squares method, i.e., by minimizing the 2-norm  $\|F\boldsymbol{p}_2 - \boldsymbol{\beta}\|_2$ .

Before proceeding with the reconstruction of  $\boldsymbol{p}_2$ , let us for a moment come back to the expression given in Eq. (14). At low densities, i.e., for  $\beta \rightarrow 0$ , the stationary node densities,  $\rho_i^*$ , are close to 0. It follows, from Eq. (11), that the constant  $c(\beta)$  admits the following expansion  $c(\beta) = c_1\beta + c_2\beta^2 + c_3\beta^3 + O(\beta^4)$  as  $\beta \rightarrow 0$ . Inserting the latter expression into (14) and gathering terms of the same power in  $\beta$ , we deduce:

$$\begin{aligned} c_1 &= \frac{1}{\langle k_1 \rangle + \langle k_2 \rangle}, \\ c_2 &= \frac{\langle k_1^2 \rangle + \langle k_2^2 \rangle + 2\langle k_1 \rangle \langle k_2 \rangle}{(\langle k_1 \rangle + \langle k_2 \rangle)^3}, \\ c_3 &= 2 \frac{(\langle k_1^2 \rangle + \langle k_2^2 \rangle + 2\langle k_1 \rangle \langle k_2 \rangle)^2}{(\langle k_1 \rangle + \langle k_2 \rangle)^5} - \frac{\langle k_1^3 \rangle + \langle k_2^3 \rangle + 3\langle k_1^2 \rangle \langle k_2 \rangle + 3\langle k_1 \rangle \langle k_2^2 \rangle}{(\langle k_1 \rangle + \langle k_2 \rangle)^4}. \end{aligned} \quad (17)$$

The latter expressions provide a direct way of obtaining the degree momenta  $\langle k_2^m \rangle = \sum_{k^{(2)}} (k^{(2)})^m p_2(k^{(2)})$  of the unknown subgraph, given again the ones of the known network. Indeed, for a given  $\beta$ , the knowledge of the stationary density of a single (arbitrary) node allows to deduce the corresponding value of the constant  $c(\beta)$ . By varying  $\beta$ , we obtain a set of measurement points  $(\beta_i, c(\beta_i))$ , from which the constants  $c_1, c_2, \dots$  can be deduced by interpolation. The latter constants then allow to obtain the unknown momenta  $\langle k_2^m \rangle = \sum_{k^{(2)}} (k^{(2)})^m p_2(k^{(2)})$ , being the momenta  $\langle k_1^m \rangle = \sum_{k^{(1)}} (k^{(1)})^m p_1(k^{(1)})$  explicitly known.

We illustrate the procedure by considering a multigraph made of 100 nodes, the known subgraph of which is an Erdős-Rényi graph,  $G_1$ , with a probability  $q_1 = 0.15$  of link connection and whose unknown subgraph is also an Erdős-Rényi subgraph, this time with a probability  $q_2 = 0.3$  of link connection, of course this

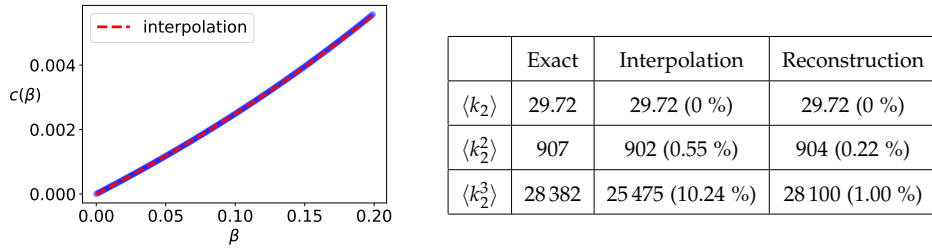


Figure 2: Reconstruction of the degree momenta of an Erdős-Rényi subgraph  $G_2$  made of  $n = 100$  nodes with a probability  $q_2 = 0.3$  of link connection. The known subgraph is an Erdős-Rényi graph made of 100 nodes with probability  $q_1 = 0.15$  of link connection. (Left) Data points  $(\beta_i, c(\beta_i))$  along with the cubic polynomial interpolation. (Right) Comparison between the exact and reconstructed degree momenta  $\langle k_2^m \rangle$  of  $G_2$  ( $m = 1, 2, 3$ ). The second column gives the exact values of the degree momenta. The third column shows the values obtained from the interpolation procedure, see (17) and the fourth column the results obtained by solving the linear system (16) to obtain the degree distribution  $p_2$  and the degree momenta. The relative errors are reported in parenthesis.

information will not be used in the reconstruction strategy. Fig. 2 (left) shows (in blue) the data points  $(\beta_i, c(\beta_i))$  obtained from Eq. (11) for  $0 \leq \beta_i \leq \beta_{\max} = 0.2$ , along with the cubic polynomial interpolation  $c_1\beta + c_2\beta^2 + c_3\beta^3$  (in dashed red). From these coefficients, one can obtain the first moment degrees, as shown in the Table provided in Fig. 2. These momenta are in close agreement with the exact degree momenta provided in the second column (Exact). The following two columns list the degree momenta obtained by solving the linear system (16) (Interpolation) and the set of equations (17) (Reconstruction); the values in parenthesis are the relative errors. Let us conclude by observing that the coefficients of the interpolating polynomial depend on the choice of  $\beta_{\max}$  and so do the reconstructed degree momenta (see Fig. 7 in Appendix B).

## 4 Impact of the network structure

We mentioned in the previous section that the sought degree distribution  $p_2(k^{(2)})$  can be found as the solution of a linear problem and that it could be obtained by minimizing the 2-norm  $\|Fp_2 - \beta\|_2$ . The purpose of this section is to show the effectiveness of the proposed strategy and to investigate how the accuracy of the reconstruction procedure is influenced by the topologies of the known and unknown subgraphs. To this aim, we will consider three distinct subgraph topologies, namely, Erdős-Rényi, scale-free and Watts-Strogatz. We further report in Appendix C the results obtained for a bimodal degree distribution. In each case, we will compute the degree distribution by means of a mean-square method. The accuracy of the reconstruction will be assessed by the (relative) error made on

the first two momenta of the degree distribution, i.e.,  $\langle k_2 \rangle$  and  $\langle k_2^2 \rangle$ . This choice is motivated by the fact that the first two momenta of the degree distribution already provide crucial information on the interplay between network structure and dynamics. For instance, the epidemic threshold  $\lambda_c$  of the SIS model occurs, in the mean-field limit, at  $\lambda_c = \langle k \rangle / \langle k^2 \rangle$ .

#### 4.1 Reconstruction of an Erdős-Rényi subgraph

In this subsection, we assume the unknown subgraph  $G_2$  to be given by an Erdős-Rényi graph of  $\Omega = 250$  nodes and probability  $q_2$  of connection among nodes, and we are interested in investigating the impact of the known topology of subgraph  $G_1$  on the quality of the reconstruction. We first consider the case where  $G_1$  is also an Erdős-Rényi graph of  $\Omega = 250$  nodes and probability  $q_1$  of links connection. The top panel of the first column of Fig. 3 provides the relative error (all the relative errors will be expressed in percent and in  $\log_{10}$  scale for better visualization) made on the estimated average degree  $\langle k_2 \rangle$  of the unknown subgraph  $G_2$ . Similarly, the bottom panel of the same column provides the relative error for the second moment,  $\langle k_2^2 \rangle$ . Those errors are computed for a large interval of values of the probabilities  $q_1$  and  $q_2$  used to build the Erdős-Rényi graphs. In practice, for each pair of parameters  $(q_1, q_2)$ , we sampled 10 multigraphs and computed for each of them the relative error made on the reconstructed degree momenta. The relative errors, obtained for each replica, were then averaged<sup>3</sup>. In the middle panel, we show, for a specific couple  $(q_1, q_2)$  of links connection probabilities, the reconstructed degree distribution (blue symbols) and the comparison with the exact one (red symbols).

We then repeat the same experiment in the case where  $G_1$  is a Watts-Strogatz graph (second column) and eventually in the case it has a scale-free degree distribution with exponent  $\gamma_1$  (third column). We here briefly comment on the procedure used to generate these subgraphs. The Watts-Strogatz graph is obtained by starting with a regular ring lattice of  $N$  nodes, with each node connected to its  $k$  nearest neighbors. With probability  $\tau_1$ , each edge  $(u, v)$  of the initially regular graph is replaced by a new edge  $(u, w)$  where the node  $w$  is chosen at random among all the nodes (distinct) from  $v$  [33]. In all the analyses shown here, the parameters  $N = 250$  and  $k = 32$  are fixed while we vary the parameter  $\tau_1$  to investigate the impact of the link rewiring on the accuracy of the reconstruction scheme. The scale-free graphs were generated according to the Simon model [30] which allows to build connected graphs with a power-law degree distribution  $p(k) \sim 1/k^{\gamma_1}$  with  $\gamma_1 \geq 2$ .

The numerical results reported in Fig. 3 suggest that the reconstruction of the first moment is better than for the second one and the goodness of the reconstruction increases with  $q_2$ , i.e., the larger the connectivity of the unknown graph, the better are the results. When  $G_1$  is a Watts-Strogatz graph (second column), the

---

<sup>3</sup>Let us note that one could in principle first average the reconstructed degree momenta and then compute the relative error. However as we are interested in quantifying the quality of a single reconstruction, we decided to first compute the relative error and then average the latter over the set of replicas.

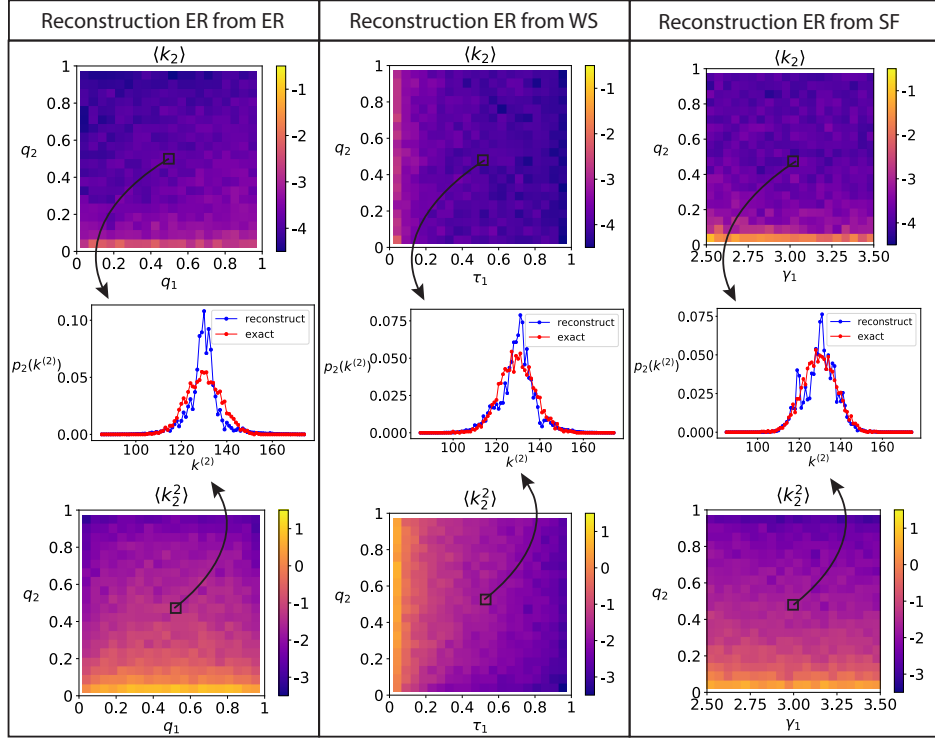


Figure 3: Reconstruction of an Erdős-Rényi (ER) subgraph  $G_2$  with  $\Omega = 250$  nodes and probability  $q_2$  of links connection. (First column) The known subgraph  $G_1$  is an Erdős-Rényi graph with  $\Omega = 250$  nodes and probability  $q_1$  of links connection. The relative error (in  $\log_{10}$ ) made on the average degree  $\langle k_2 \rangle$  and second moment  $\langle k_2^2 \rangle$  are reported on the first and third rows. For each panel the data were averaged over 10 independent replicas. The central panel shows, for some specific values of the parameters  $q_1$  and  $q_2$  the reconstructed distribution compared with the exact one. (Second column) Reconstruction of  $G_2$  by using a Watts-Strogatz (WS) graph  $G_1$  with probability  $\tau_1$  of link rewiring, starting from a ring topology in which each node is connected to its 32 nearest neighbors. (Third column) Reconstruction of  $G_2$  by using a graph  $G_1$  with a scale-free (SF) degree distribution with exponent  $\gamma_1$ , i.e.,  $p_1(k^{(1)}) \sim 1/[k^{(1)}]^\gamma$ .

reconstruction works better when the randomness of the network is increased, i.e., when the links rewiring parameter  $\tau_1$  is increased. In all the three cases, the results suggest that the impact of the known graph  $G_1$ , as measured by  $q_1$ ,  $\mu_1$  or  $\gamma_1$ , are quite limited. The middle panels confirm such claim, indeed the reconstructed probability distribution reproduces the main feature of the exact distribution, in particular it is centered around the right value, i.e., the average degree, but the widths are a bit narrower than the true ones.

## 4.2 Reconstruction of a scale-free graph

We now turn our attention to the reconstruction of a scale-free graph with  $\Omega = 250$  nodes and exponent  $\gamma_2$ . Results are reported in Fig. 4. The reconstruction of a scale-free graph from the knowledge of an Erdős-Rényi graph or a Watts-Strogatz graph is quite independent of the involved parameters  $q_1$ ,  $\gamma_2$  and  $\tau_1$ , as measured by the relative errors made on  $\langle k_2 \rangle$  and  $\langle k_2^2 \rangle$  (see first and second columns). The reconstruction of a scale-free graph from a scale-free graph (third column) is best achieved for large values of  $\gamma_1$  and low values of  $\gamma_2$ . Overall, the results suggest that the reconstruction works better when the known graph  $G_1$  is an Erdős-Rényi graph or a Watts-Strogatz graph.

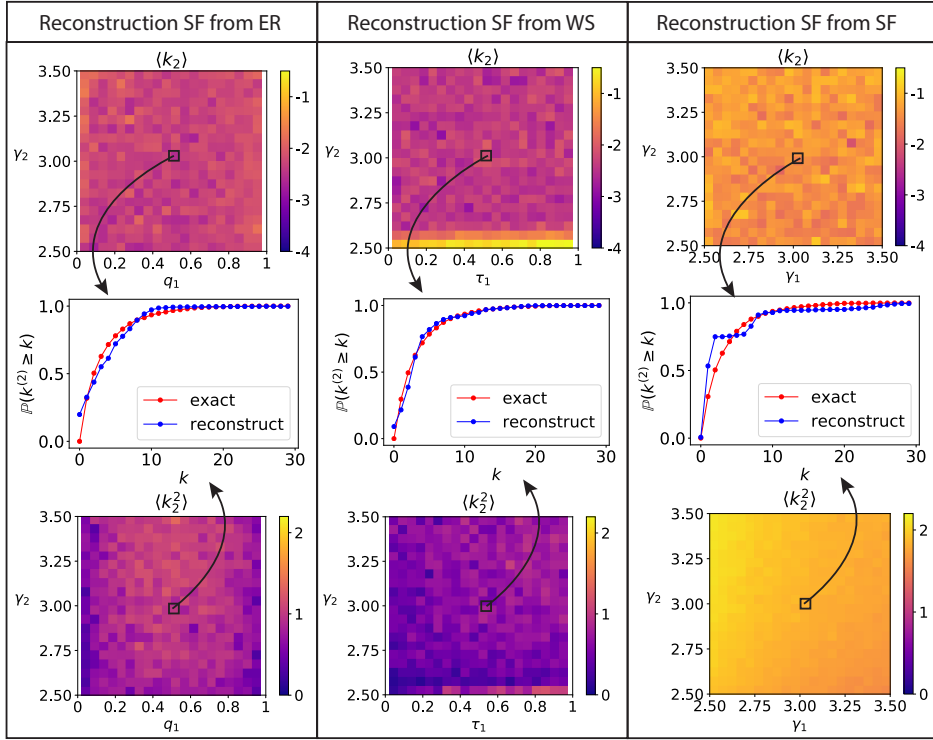


Figure 4: Reconstruction of a subgraph  $G_2$  with scale-free degree distribution with exponent  $\gamma_2$  and  $\Omega = 250$  nodes. (First column) The known subgraph  $G_1$  is an Erdős-Rényi graph with  $\Omega = 250$  nodes and probability  $q_1$  of link connection. The relative error (in  $\log_{10}$ ) made on the average degree  $\langle k_2 \rangle$  and second moment  $\langle k_2^2 \rangle$  are reported on the first and third rows (averaged over 10 independent configurations). The central panel shows the reconstructed distribution, for specific values of the parameters  $\gamma_2$  and  $q_1$ , averaged over several configurations, and the exact one. (Second column) Same with  $G_1$  a Watts-Strogatz graph. (Third column) Same with  $G_1$  a scale-free network with exponent  $\gamma_1$ , i.e.,  $p_1(k^{(1)}) \sim 1/[k^{(1)}]^{\gamma_1}$ .

### 4.3 Reconstruction of a Watts-Strogatz graph

We finally consider the reconstruction of a Watts-Strogatz graph  $G_2$ . As mentioned previously, the graph is obtained by starting with a regular ring lattice of  $N = 250$  nodes, in which each node is connected to its 32 nearest neighbors and by rewiring the edges with some probability  $\tau_2$ . As in the previous sections, we consider the known subgraph  $G_1$  to be chosen among an Erdős-Rényi graph, a Watts-Strogatz graph and a scale-free graph. The accuracy of the reconstruction is relatively independent from the known graph, see Fig. 5. When  $G_1$  is a Watts-Strogatz graph (second column), the accuracy decreases for low values of the links rewiring parameters, i.e., when the underlying subgraphs are almost regular.

### 4.4 Application to the reconstruction of a real multigraph

In this subsection, we test the reconstruction algorithm on a real multigraph describing the relationships between the members of a corporate law partnership [34,35]. Two additional examples are provided in App. E. The dataset used in the following contains 71 nodes (i.e., members) and 2223 links. There are three kinds of links [34]:

- *Co-work* links between two members indicate that they worked together.
- *Advice* links connect two members if one of the two consulted the other for some professional advice.
- *Friendship* links between two members indicate that the two of them have socialized outside work.

Links in the multigraph carry an orientation. We here discard this information and assume the links to be undirected. As the reconstruction algorithm introduced previously considers two subgraphs, i.e., the known subgraph  $G_1$  and the unknown one  $G_2$ , we grouped the co-work and advice links (which can be viewed as *professional* links). We then proceed with the reconstruction of the degree distribution of the professional links, assuming the subgraph of friendship relationships to be known. The reconstructed degree distribution is shown in Fig. 6 (Left) and is found to be in close agreement with the exact one. We then reverse the roles and apply the algorithm to the reconstruction of the degree distribution of the friendship links, under the knowledge of the subgraph of the professional links, see Fig. 6 (Right). The reconstruction is less accurate, due to the low density of the subgraph we aimed at reconstructing. In particular, a non-negligible fraction of the nodes have no friendship connections (such nodes have degree  $k^{(2)} = 0$ , but the network is still connected thanks to the links of the other kind). The algorithm is able to predict the presence of such nodes although the estimate ( $\approx 0.3$ ) overestimates the exact value ( $\approx 0.2$ ). We further investigate in App. D the impact of the density of the unknown subgraph on the accuracy of the reconstruction. Observe that the method hereby used relies on the assumption of absence of correlations among the degree distributions on each subgraph; in

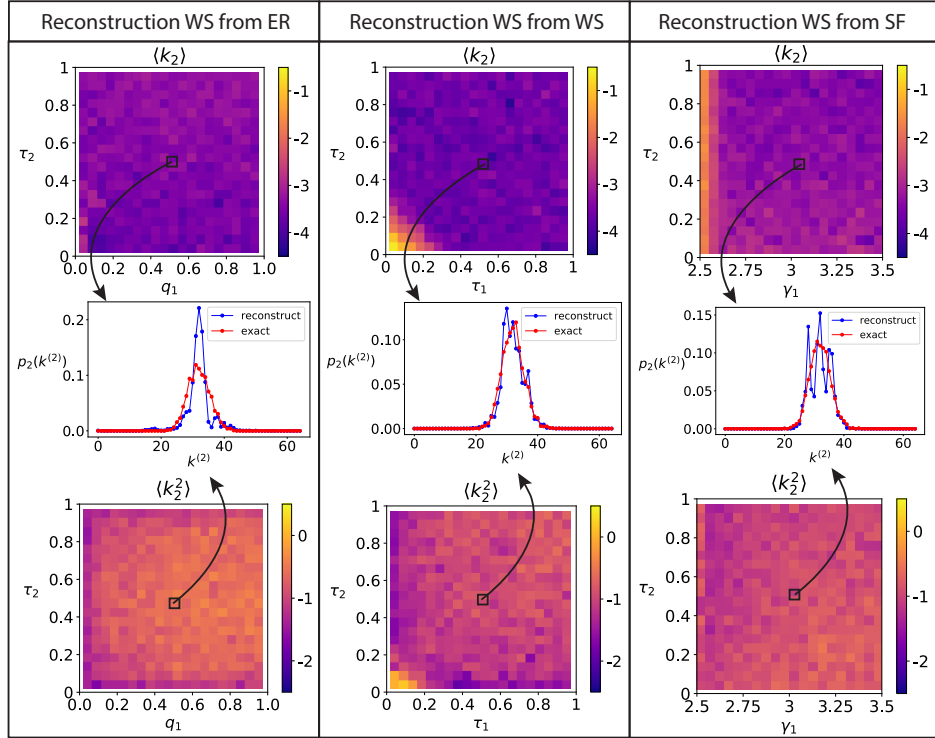


Figure 5: Reconstruction of a Watts-Strogatz graph  $G_2$  with  $\Omega = 250$  nodes, average connectivity  $k = 32$  and probability  $\tau_2$  of links rewiring. (First column) The known subgraph  $G_1$  is an Erdős-Rényi graph with  $\Omega = 250$  nodes and probability  $q_1$  of link connection. The relative error (in  $\log_{10}$ ) made on the average degree  $\langle k_2 \rangle$  and second moment  $\langle k_2^2 \rangle$  are reported on the first and third rows (averaged over 10 independent configurations). The central panel shows, for specific values of the parameters  $\tau_2$  and  $q_1$ , the reconstructed distribution, averaged over several configurations, and compared with the exact one. (Second column) Same with  $G_1$  a Watts-Strogatz graph, with average connectivity  $k = 32$  and probability  $\tau_1$  of links rewiring. (Third column) Same with  $G_1$  a scale-free network with exponent  $\gamma_1$ , i.e.,  $p_1(k^{(1)}) \sim 1/[k^{(1)}]^{\gamma_1}$ .

App. F we studied the impact of correlations by using a synthetic network and we observed a strong impact especially in the case the degree distributions are anti-correlated, i.e., the degree distribution is disassortative. Let us note that the degree momenta of the unknown subgraph can also be obtained by measuring the constant  $c(\beta)$  for small  $\beta$ , and by using Eq. (17). For the reconstruction of the degree distribution of the advice and co-work links, the procedure yields<sup>4</sup>

<sup>4</sup>We used an interpolation of degree 5, i.e.,  $c(\beta) = \sum_{j=1}^5 c_j \beta^j$ , with the constants  $c_j$  estimated from the measurement points  $(\beta_i, c(\beta_i))$  with  $\beta_i$  sampled on the interval  $[0, 0.02]$  by increment of 0.001.

the estimate  $\langle k_2 \rangle \approx 25.8$  and  $\langle k_2^2 \rangle \approx 733$ , to be compared with the exact values  $\langle k_2 \rangle = 25.6$  and  $\langle k_2^2 \rangle = 728$ . For the reconstruction of the degree distribution of the friendship links, we obtain  $\langle k_2 \rangle \approx 2.82$  and  $\langle k_2^2 \rangle \approx 20.5$ , while the exact values are given by  $\langle k_2 \rangle = 2.82$  and  $\langle k_2^2 \rangle = 15.3$ .

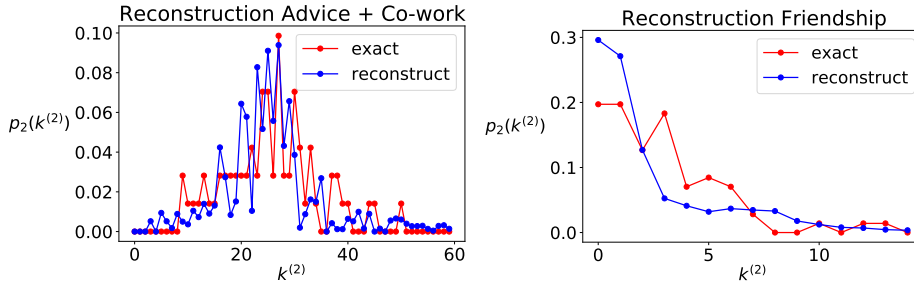


Figure 6: Application to a real corporate law partnership multigraph. Reconstruction of the degree distribution of professional links (co-work and advice links) from the knowledge of the friendship links (Left) and the other way around (Right).

## 5 Conclusion

We here studied a non-linear diffusion process on top of multigraphs, where pairs of nodes can be connected by different types of links. The process was derived from a microscopic description modelling the dispersal of a set of agents under limiting node carrying capacities. As for the case of simple graphs, the stationary node densities show a nonlinear dependence on the node degrees. From the explicit knowledge of the stationary node densities, we extended an existing algorithm to reconstruct the degree distribution associated to a given type of links, by assuming the pattern of the other types of interactions to be known. We first applied the algorithm to the case of synthetic multigraphs, assessing in each case the robustness of the reconstruction by measuring the relative errors on the first two degree momenta. We then applied the algorithm to three real multigraphs; this allowed us to benchmark the method on networks where the assumption of uncorrelated nodes degree is no longer true and thus to study the impact of assortativity on the results. We also considered a second method based on a diluted nonlinear random walk, i.e., once a smaller and smaller number of walkers is allowed to travel across the multigraph links, in this way we have been able to directly measure the first and second degree momenta without the need to compute the degree distribution. This algorithm has been successfully tested both on synthetic than empirical networks. Future work could include an extension of the reconstruction algorithm to the case of multilayer networks in which there are both interlayer and intralayer links.

## 6 Funding

This work was supported by the Fonds de la Recherche Scientifique—FNRS (Grant FC 38477 to J.-F. d K). Part of the results were obtained using the computational resources provided by the “Consortium des Equipements de Calcul Intensif” (CECI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

## Appendices

### A Derivation of the ODE from the M-equation

The purpose of this appendix is to derive the system of ordinary differential equations given in Eq. (3) and reproduced below:

$$\begin{aligned} \frac{d\rho_i}{dt} &= \sum_j \left\{ \rho_j \frac{A_{ji}^{(1)} + A_{ji}^{(2)}}{k_j} g(\rho_i) - \rho_i \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_i} g(\rho_j) \right\} \\ &= \sum_j \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_j} \left\{ \rho_j g(\rho_i) - \frac{k_j}{k_i} \rho_i g(\rho_j) \right\}, \end{aligned} \quad (18)$$

from the M-equation

$$\frac{d\mathbb{P}(\mathbf{n}, t)}{dt} = \sum_{\mathbf{n}' \neq \mathbf{n}} \{ T(\mathbf{n}|\mathbf{n}') \mathbb{P}(\mathbf{n}', t) - T(\mathbf{n}'|\mathbf{n}) \mathbb{P}(\mathbf{n}, t) \}, \quad (19)$$

with transition rates

$$T(n_i - 1, n_j + 1 | n_i, n_j) = \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_i} \frac{n_i}{N} g\left(\frac{n_j}{N}\right). \quad (20)$$

We refer the reader to the main text for more details about the notations used. The procedure used is quite standard, see e.g. [22]. We first recall that the average number of individuals found in node  $i$  at time  $t$  is given by  $\langle n_i(t) \rangle = \sum_{\mathbf{n}} n_i(t) \mathbb{P}(\mathbf{n}, t)$ . As during an infinitesimal time interval, at most one random walker will have moved, the states  $\mathbf{n}$  and  $\mathbf{n}'$  will have identical components, except for two of them, corresponding to the nodes involved in the transition. More precisely, if the state before transition has components  $\mathbf{n} = (n_1, \dots, n_i, \dots, n_j, \dots, n_\Omega)$ , then after the transition, the new state will have components  $\mathbf{n}' = (n_1, \dots, n_i \pm 1, \dots, n_j \mp 1, \dots, n_\Omega)$ . We thus obtain:

$$\begin{aligned} \frac{d\langle n_i(t) \rangle}{dt} &= \sum_{n_i} \sum_{n_j} n_i \left[ -T(n_i - 1, n_j + 1 | n_i, n_j) \mathbb{P}(n_i, n_j, t) \right. \\ &\quad \left. + T(n_i, n_j | n_i + 1, n_j - 1) \mathbb{P}(n_i + 1, n_j - 1, t) \right] \\ &\quad + \sum_{n_i} \sum_{n_j} n_i \left[ -T(n_i + 1, n_j - 1 | n_i, n_j) \mathbb{P}(n_i, n_j, t) \right. \\ &\quad \left. + T(n_i, n_j | n_i - 1, n_j + 1) \mathbb{P}(n_i - 1, n_j + 1, t) \right]. \end{aligned} \quad (21)$$

In the above expression,  $T(n_i - 1, n_j + 1 | n_i, n_j) \mathbb{P}(n_i, n_j, t)$  denotes the transition rate associated to the jump of an agent from node  $i$ , containing initially  $n_i$  agents,

to node  $j$ , containing initially  $n_j$  agents. The three other terms carry a similar interpretation. Playing with the summation indices in Eq. (21) leads to the following expression:

$$\begin{aligned} \frac{d\langle n_i(t) \rangle}{dt} &= \sum_j \left( \langle T(n_i + 1, n_j - 1 | n_i, n_j) \rangle - \langle T(n_i - 1, n_j + 1 | n_i, n_j) \rangle \right) \\ &= \sum_j \left( \frac{A_{ji}^{(1)} + A_{ji}^{(2)}}{k_j} \langle \frac{n_j}{N} g(\frac{n_i}{N}) \rangle - \frac{A_{ji}^{(1)} + A_{ji}^{(2)}}{k_i} \langle \frac{n_i}{N} g(\frac{n_j}{N}) \rangle \right) \end{aligned} \quad (22)$$

Rescaling time  $t \rightarrow tN$  and taking the limit  $N \rightarrow +\infty$  allows to neglect the correlations and leads to:

$$\frac{d\rho_i}{dt} = \sum_j \left\{ \rho_j \frac{A_{ji}^{(1)} + A_{ji}^{(2)}}{k_j} g(\rho_i) - \rho_i \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{k_i} g(\rho_j) \right\}, \quad (23)$$

as stated in Eq. (3).

## B Dependence of the degree momenta on the mass parameter $\beta$

In this appendix, we comment on the impact that the mass parameter  $\beta$  has on the reconstruction of the degree momenta, see Eq. (17). We recall that the latter are obtained by first computing the constant  $c(\beta)$  for distinct values of  $\beta$  in the interval  $[0, \beta_{\max}]$  and then by extrapolating  $c(\beta)$  by a polynomial whose coefficients allow to obtain the desired degree momenta. In Fig. 7, we compare, as a function of  $\beta_{\max}$ , the first and second degree momenta, as reconstructed from Eq. (17), with their exact values. The agreement is very satisfactory except for very low values of  $\beta_{\max}$ .

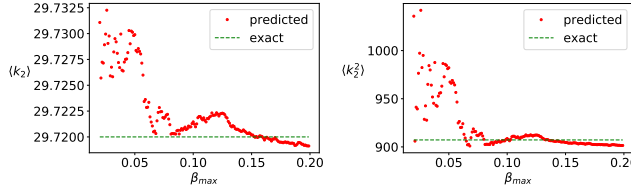


Figure 7: Reconstruction of the degree momenta of an Erdős-Rényi subgraph  $G_2$  made of  $n = 100$  nodes with a probability  $q_2 = 0.3$  of link connection. The known subgraph is an Erdős-Rényi graph made of 100 nodes with probability  $q_1 = 0.15$  of link connection. (Left panel) Comparison between the predicted and exact first moment  $\langle k_2 \rangle$ , as obtained from the interpolation on the interval  $0 \leq \beta_i \leq \beta_{\max}$ . (Right panel) Same for the second moment  $\langle k_2^2 \rangle$ .

## C Reconstruction in the case of bimodal degree distribution

In this appendix, we investigate the accuracy of the reconstruction procedure described in the main text when the subgraph  $G_1$  and/or the subgraph  $G_2$  has (have) a bimodal degree distribution. To construct a subgraph  $G_1$  with bimodal degree distribution, we proceed as follows. We first generate the sequence of nodes' degrees, with half of the nodes' degrees sampled from a Gaussian distribution<sup>5</sup> with mean  $\mu = 35$  and standard deviation  $\sigma = 10$ , while the remaining nodes degrees are sampled from a Gaussian distribution centered at  $\mu_1$  with standard deviation  $\sigma = 10$ , being  $\mu_1$  a varying parameter. Once the degree sequence has been generated, the subgraph  $G_1$  is built by using the configuration model with the previously generated sequence of nodes' degrees. We delete self-loops and multiple edges so that the adjacency matrix  $A^{(1)}$  associated to the known subgraph has binary entries<sup>6</sup>.

In Fig. 8, we show the reconstruction of a graph  $G_2$  with bimodal degree distribution, for four distinct topologies of the known subgraph  $G_1$ , namely, Erdős-Rényi (ER), Watts-Strogatz (WS), scale-free (SF) and bimodal (Bim). The reconstruction strategy works better when the known subgraph is an Erdős-Rényi, Watts-Strogatz or a scale-free graph. In contrast, the exact and reconstructed degree distributions deviate significantly when  $G_1$  has a bimodal degree distribution (see fourth column) although the bimodality is still predicted.

Fig. 9 shows the accuracy of the reconstruction of a graph when the known subgraph has bimodal degree distribution. The reconstruction works better if the unknown subgraph is an Erdős-Rényi (first column) or a Watts-Strogatz (third column).

---

<sup>5</sup>We here assume nodes' degrees to be positive integers between 1 and  $\Omega - 1$ , with  $\Omega$  the number of nodes. The numbers randomly sampled from the Gaussian distribution are thus rounded to their nearest integer value and values outside the interval  $(1, \Omega)$ , if any, are rejected during the sampling procedure.

<sup>6</sup>As a consequence of the suppression of self loops and multiple edges, the degree connectivity of the resulting graph is slightly reduced and, in particular, the two peaks of the observed degree distribution  $p_2(k^{(2)})$  will be slightly shifted towards lower values of  $k^{(2)}$ . Let us however note that one could still use the reconstruction algorithm presented in the main text in presence of multiple edges.

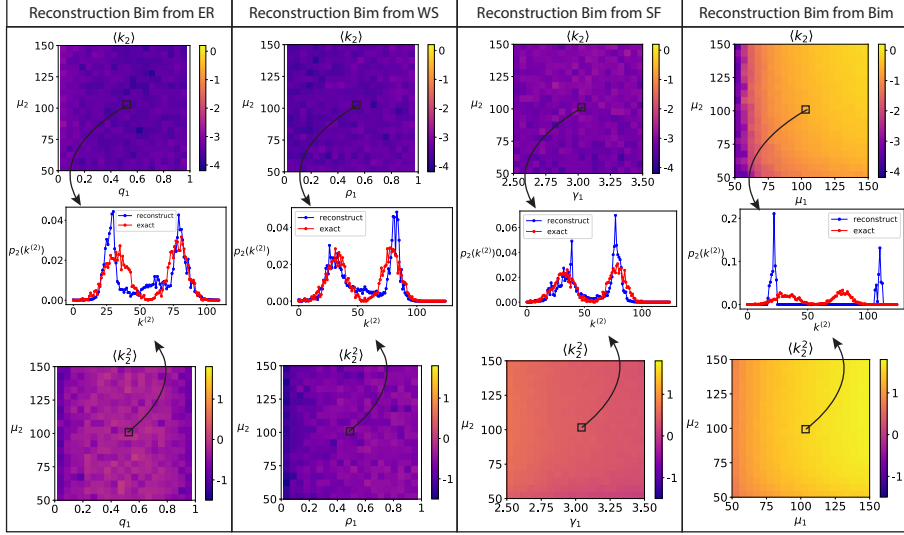


Figure 8: Reconstruction of a subgraph  $G_2$  with bimodal degree distribution (Bim) centred at 35 and  $\mu_2$  with variance 10 and  $\Omega = 250$  nodes. (First column) The known subgraph  $G_1$  is an Erdős-Rényi graph with  $\Omega = 250$  nodes and probability  $q_1$  of link connection. The relative error (in  $\log_{10}$ ) made on the average degree  $\langle k_2 \rangle$  and second moment  $\langle k_2^2 \rangle$  are reported on the first and third rows (averaged over 10 independent configurations). The central panel shows, for specific values of the parameters  $\mu_2$  and  $q_1$ , the reconstructed distribution, averaged over several configurations, and compared with the exact one. (Second column) Same with  $G_1$  a Watts-Strogatz graph with probability  $\tau_1$  of links rewiring. (Third column) Same with  $G_1$  a scale-free network with exponent  $\gamma_1$ , i.e.,  $p_1(k^{(1)}) \sim 1/[k^{(1)}]^{\gamma_1}$ . (Fourth column) Same with  $G_1$  having a bimodal degree distribution centred at 35 and  $\mu_1$ , with variance 10.

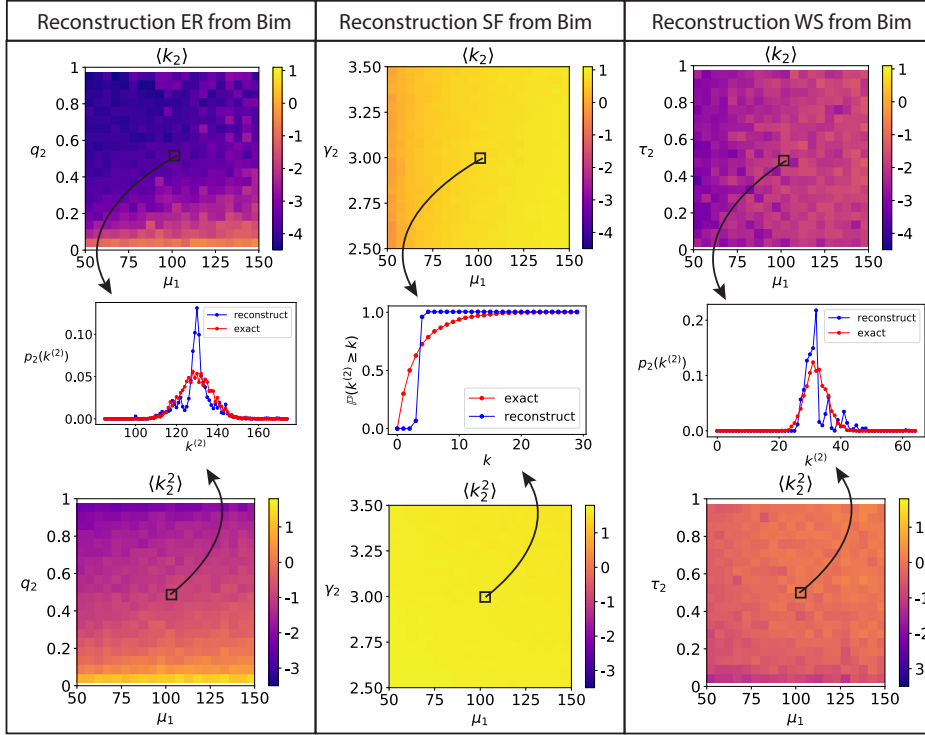


Figure 9: Reconstruction of a subgraph  $G_2$  when the known subgraph  $G_1$  has bimodal degree distribution (Bim) centred at 35 and  $\mu_2$  with variance 10 and  $\Omega = 250$  nodes. (First column) The unknown subgraph  $G_2$  is an Erdős-Rényi graph with  $\Omega = 250$  nodes and probability  $q_2$  of link connection. The relative error (in  $\log_{10}$ ) made on the average degree  $\langle k_2 \rangle$  and second moment  $\langle k_2^2 \rangle$  are reported on the first and third rows (averaged over 10 independent configurations). The central panel shows, for specific values of the parameters  $\mu_1$  and  $q_2$ , the reconstructed distribution, averaged over several configurations, and compared with the exact one. (Second column) Same with  $G_2$  a scale-free network with exponent  $\gamma_2$ , i.e.,  $p_2(k^{(2)}) \sim 1/[k^{(2)}]^{\gamma_2}$ . (Third column) Watts-Strogatz graph with probability  $\tau_2$  of links rewiring.

## D Reconstruction of low density networks

In this appendix, we investigate the accuracy of the reconstruction of low density connected subgraphs. To this goal, we construct synthetic multigraphs with  $G_1$  (known subgraph) and  $G_2$  (unknown subgraph) two Erdős-Rényi graphs with probabilities  $q_1 = 1 \gg 1/\Omega$  and  $q_2 \sim 1/\Omega$  of link connections. In this way the known subgraph is densely connected while the unknown subgraph is sparse, i.e., the node degrees  $k^{(2)}$  are low and possibly zero. In Fig. 10, we show the relative errors on the average reconstructed degree momenta  $\langle k_2 \rangle$  and  $\langle k_2^2 \rangle$  as a function of  $q_2$  with  $\Omega = 100$  nodes. Vertical bars indicate the standard deviations deduced from 100 experiments. As can be seen from this Figure, the averaged relative error and the standard deviations increase as the parameter  $q_2$  is reduced, implying thus some limitation of the reconstruction method in presence of zero degree nodes and low density of connections.

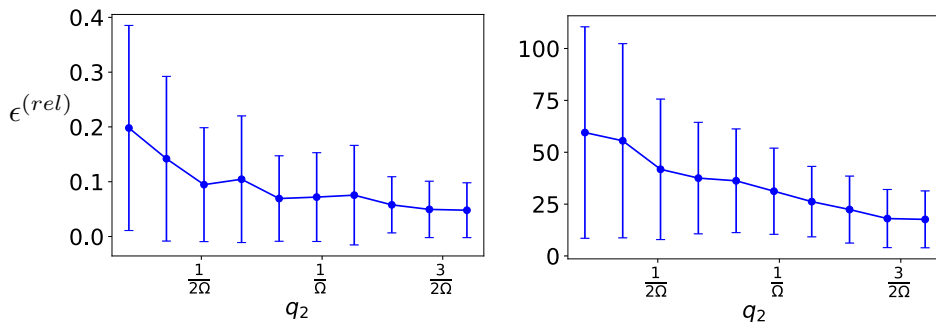


Figure 10: Reconstruction of an Erdős-Rényi subgraph  $G_2$  of  $\Omega = 100$  nodes with link connection  $q_2$  from the knowledge of an Erdős-Rényi subgraph  $G_1$  with link connection  $q_1 = 0.5$ . In the left panel we show the relative error  $\epsilon^{(rel)}$  on  $\langle k_2 \rangle$  while in the right panel the one for  $\langle k_2^2 \rangle$ . The data were averaged over 100 independent replicas. Vertical bars represent the standard deviations.

## E Reconstruction of real multigraphs

We further investigate in this Appendix the reconstruction of real multigraphs. The first instance is a cooperation multigraph among students in a course [31]. Nodes correspond to students and links identify cooperation between them. Three types of cooperation links were identified, whose attributes are denoted by *Time*, *Computer* and *Partners*. Partners links connect pairs of students that submitted their assignments together, computer links connect pairs of students that used the same computer for their (online) assignments and Time links connect pairs of students that had accessed the course's website almost at the same time. We refer the reader to [31] for more details about the procedure used to collect the data and construct the multigraph. As there are fewer *Computer* links, we discard them

and only focus on the Time and Partners links, dealing thus with a multigraph with two attributes. Moreover, we restrict our attention to the largest connected component<sup>7</sup> of the resulting multigraph, shown in Fig. 11 (Left). The latter contains 134 nodes and 242 links, of which 156 are Partners links and 86 are Time links. We then proceed with the reconstruction of the degree distribution corresponding to the Partners links, assuming the degree distribution of the *Time* links to be known, see top right panel of Fig. 11. We then repeat the experiment for the reconstruction of the degree distribution of the Time links, see bottom right panel of Fig. 11. In the latter case, the reconstruction is less satisfactory. The discrepancy observed might be the consequence of the correlation between the degree sequences of both subgraphs. The Pearson correlation coefficient measuring the correlation between both degree sequences is given by  $r \approx -0.5$ , which indicates a diassortative behavior: nodes with few Partners links tend to have more Time links and vice versa. We further investigate in Appendix F the impact of degree-degree correlations on the reconstruction accuracy. The degree momenta of the unknown subgraph can also be inferred by measuring the constant  $c(\beta)$  for small  $\beta$ , and by using Eq. (17). For the reconstruction of the degree distribution of the Partners links, the procedure yields the estimate  $\langle k_2 \rangle \approx 2.33$  and  $\langle k_2^2 \rangle \approx 5.17$ , to be compared with the exact values  $\langle k_2 \rangle = 2.33$  and  $\langle k_2^2 \rangle = 6.19$ . For the reconstruction of the degree distribution of the Time links, we obtain  $\langle k_2 \rangle \approx 1.28$  and  $\langle k_2^2 \rangle \approx 2.56$  while the exact values are given by  $\langle k_2 \rangle = 1.28$  and  $\langle k_2^2 \rangle = 3.58$ .

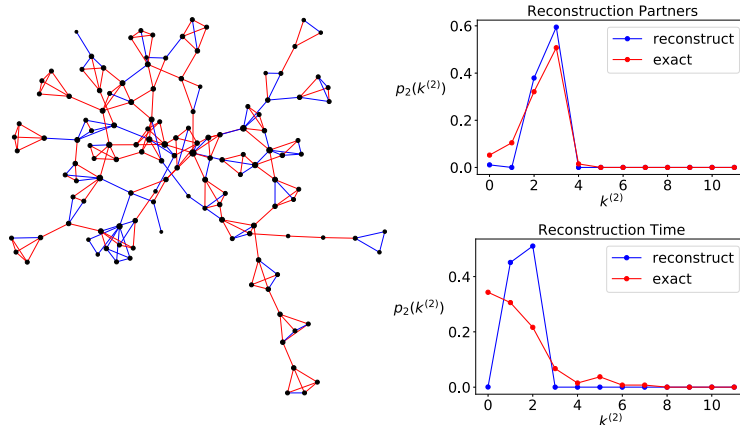


Figure 11: (Left) Cooperation multigraph. Blue links correspond to *Time* links while red links identify *Partners* links. (Right) The top panel (resp. bottom) gives the reconstruction of the degree distribution corresponding to the *Partners* (resp. *Time*) links from the knowledge of the degree distribution of the *Time* (resp. *Partners*) links.

The second instance is the European air transportation multigraph where links' attributes are commercial airlines [32]. We here restrict our attention to two

<sup>7</sup>There must exist a path joining any pair of nodes in the connected component.

operating companies, namely Lufthansa and Ryanair. The multigraph is shown in Fig. 12 (Left). It contains 198 nodes connected by 1690 links, of which 488 are from the Lufthansa operating company and the others from Ryanair. As before, we restrict our attention to the largest connected component. The top right panel of Fig. 12 shows the reconstructed degree distribution corresponding to Ryanair's links, assuming the degree distribution for Lufthansa's links to be known while the bottom right panel shows the other way around. The discrepancy observed between the exact and reconstructed degree distributions might in part be due to the presence of low-degree nodes and the degree-degree correlations, see Appendices D and F. Again, the degree momenta of the unknown subgraph can be inferred by measuring the constant  $c(\beta)$  for small  $\beta$ , and by using Eq. (17). For the reconstruction of the degree distribution of Ryanair's links, the procedure yields the estimate  $\langle k_2 \rangle \approx 6.07$  and  $\langle k_2^2 \rangle \approx 120$ , to be compared with the exact values  $\langle k_2 \rangle = 6.07$  and  $\langle k_2^2 \rangle = 143$ . For the reconstruction of the degree distribution of the Lufthansa links, we obtain  $\langle k_2 \rangle \approx 2.46$  and  $\langle k_2^2 \rangle \approx 54.8$  while the exact values are given by  $\langle k_2 \rangle = 2.46$  and  $\langle k_2^2 \rangle = 77.8$ .

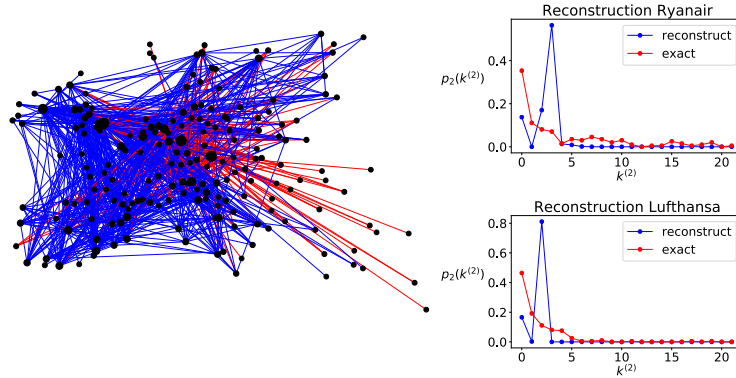


Figure 12: (Left) European air transportation multigraph. Links in blue (resp. red) indicate flight routes operated by Lufthansa (resp. Ryanair) (Right) The top panel (resp. bottom) gives the reconstruction of the degree distribution corresponding to Ryanair's (resp. Lufthansa's) links from the knowledge of the degree distribution of Lufthansa's (resp. Ryanair's) links.

## F Impact of the degree-degree correlations on the reconstruction

The reconstruction procedure described in the main text assumes the degree distributions of the known and unknown subgraphs  $G_1$  and  $G_2$  to be uncorrelated. This assumption allows to factorize the joint probability degree distribution function  $p(k^{(1)}, k^{(2)}) = p_1(k^{(1)})p_2(k^{(2)})$ , from which the unknown degree distribution

$p_2(k^{(2)})$  is recovered as the solution of a linear system, see Eq. (16). When the nodes degrees of the known subgraph correlate with that of the unknown subgraph, factorizing the joint probability distribution is no longer exact; it is an approximation that may induce some errors on the reconstruction outcomes. The purpose of this Appendix is precisely to investigate that aspect. To quantify the level of degree-degree correlations, we will use the Pearson correlation coefficient  $r$ . For the case at hand, the latter coefficient is given by:

$$r = \frac{\langle k_1 k_2 \rangle - \langle k_1 \rangle \langle k_2 \rangle}{\sigma_1 \sigma_2}, \quad (24)$$

where  $\langle k_1 k_2 \rangle = \sum_{k^{(1)}, k^{(2)}} k^{(1)} k^{(2)} p(k^{(1)}, k^{(2)})$ ,  $\langle k_1 \rangle = \sum_{k^{(1)}} k^{(1)} p(k^{(1)})$  (and similarly for  $\langle k_2 \rangle$ ) and  $\sigma_1, \sigma_2$  are the standard deviations associated to the nodes degrees of  $G_1$  and  $G_2$ .

To investigate the impact of degree-degree correlations on the reconstruction accuracy, we construct correlated multigraphs with a prescribed Pearson correlation coefficient  $r$ . The algorithm we used is similar to the one described in Section VII. A. in [36]. We start by constructing two synthetic graphs  $G_1$  and  $G_2$  (from some graph ensembles), each of which has  $\Omega$  nodes, labelled  $1, 2, \dots, \Omega$ . Keeping the graph structure of  $G_1$  fixed, the idea is to modify progressively the structure of  $G_2$  so as to get closer and closer to the desired Pearson correlation coefficient. In practice, the algorithm will stop when the observed Pearson correlation and the target one differ by a quantity smaller than some tolerance threshold  $\epsilon \ll 1$  set by the user.

We now describe in more details the algorithm. Suppose thus that we want to generate a multigraph with a Pearson correlation coefficient  $r$  given above. Starting with two random graphs  $G_1$  and  $G_2$ , we compute the difference  $\Delta r$ , in absolute value, between the observed Pearson correlation coefficient and the one we would like to obtain. If  $\Delta r$  is less than the tolerance threshold  $\epsilon$ , the algorithm outputs the multigraph. Otherwise, we select a pair of nodes of  $G_2$  at random, say  $i$  and  $j$ , and measure the new Pearson correlation coefficient we would obtain by swapping the labels of both nodes, i.e., by permuting rows  $i$  and  $j$  and columns  $i$  and  $j$  of the adjacency matrix  $A^{(2)}$ . If the operation would result in a smaller value  $\Delta_r^{\text{new}}$  compared to  $\Delta r$ , it is accepted with probability 1. If not, namely, if  $\Delta_r^{\text{new}} > \Delta r$  there is still a small probability  $e^{-(\Delta_r^{\text{new}} - \Delta r)/\lambda}$  to accept the modification, with  $\lambda > 0$  some parameter. We repeat the process until  $\Delta r$  is smaller than  $\epsilon$ . Let us note that the graph structure of  $G_1$  is preserved throughout the process.

The above algorithm generates correlated multigraphs. In particular, positive values of  $r$  indicate that high-degree nodes in  $G_1$  tend to be also high-degree nodes in  $G_2$ , while, in contrast, negative values of  $r$  imply the opposite phenomenon, namely, high degree nodes in  $G_1$  have a greater probability to be low degree nodes in  $G_2$ .

We now apply the above algorithm to generate correlated multigraphs and measure the relative errors made on the reconstructed degree momenta  $\langle k_2 \rangle$  and  $\langle k_2^2 \rangle$ . We first assume the graphs  $G_1$  and  $G_2$  to be Barabási-Albert graphs of  $\Omega = 100$  nodes. Figure 13 gives the relative errors made on the average degree

$\langle k_2 \rangle$  (left panel) and second degree moment  $\langle k_2^2 \rangle$  (right panel), as a function of the Pearson correlation coefficient  $r$ . As can be seen from the left panel, the relative error made on  $\langle k_2 \rangle$  is almost zero for positive values of the Pearson correlation coefficient. For negative values of  $r$ , the smaller  $r$  the larger the relative error. The right panel shows that the larger  $|r|$  and the larger the relative error. A similar behavior is observed if the known and unknown subgraphs are Erdős-Rényi graphs, see Fig. 14. Overall, these results suggest that degree-degree correlations reduce the performance of the reconstruction process. Our intuition about those results is the following. In the case of negatively correlated nodes degree it can happen that if  $k^{(1)}$  is small then  $k^{(2)}$  is large and as well the opposite case, but then the multigraph will have, on average, constant nodes degree, being each degree the sum of a large and a small number, hence the matrix elements  $F_{ij}$  given by Eq. (15) will be quite similar each other rendering difficult to solve the least-square problem. On the other hand if  $G_1$  and  $G_2$  have positive nodes degree correlations, then the multigraph will have nodes with very high degree, associated to large  $k^{(1)}$  and  $k^{(2)}$ , together with nodes with very low degree, associated to small  $k^{(1)}$  and  $k^{(2)}$ , the matrix elements  $F_{ij}$  will thus be quite localized and the least-square problem easier to be solved. This could thus explain the dependence of the relative error on  $\langle k_2 \rangle$ ; on the other hand we already observed that  $\langle k_2^2 \rangle$  is more sensitive to the reconstruction process and thus deviations from the assumption of independence among the distributions of  $k^{(1)}$  and  $k^{(2)}$  manifest into a large relative error for any non zero  $r$ .

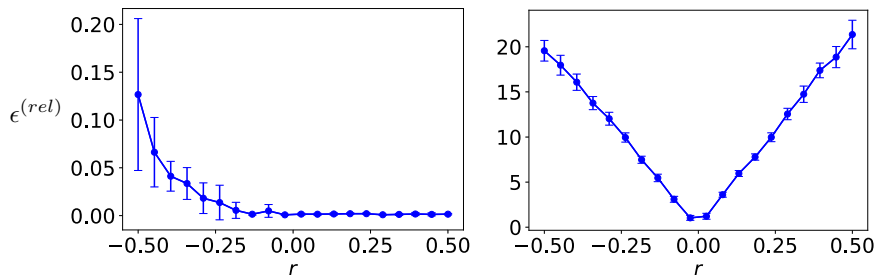


Figure 13: Relative errors  $\epsilon^{(rel)}$  made on  $\langle k_2 \rangle$  (left panel) and  $\langle k_2^2 \rangle$  (right panel) with  $G_1$  (known graph) and  $G_2$  (unknown graph) two Barabási-Albert subgraphs of  $\Omega = 100$  nodes, constructed by preferential attachment: new nodes are added to the graph and connect to 10 existing nodes. The degree sequences of both subgraphs are correlated, with a Pearson correlation coefficient  $r$  (x-axis). Each point was averaged over 10 independent replicas, with the standard deviations reported by the vertical bars. The parameter  $\lambda$  was set to 0.0001 (see the text for its meaning).

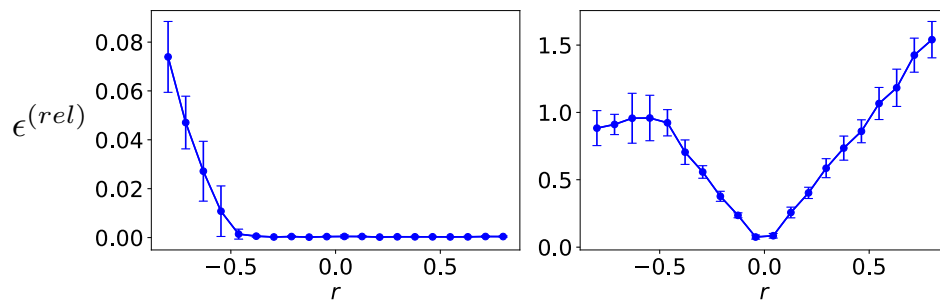


Figure 14: Relative errors  $\epsilon^{(rel)}$  made on  $\langle k_2 \rangle$  (left panel) and  $\langle k_2^2 \rangle$  (right panel) with  $G_1$  (known graph) and  $G_2$  (unknown graph) two Erdős-Rényi subgraphs of  $\Omega = 100$  nodes and probability  $q_1 = 1/2$  and  $q_2 = 1/2$  of link connection. The degree sequences of both subgraphs are correlated, with a Pearson correlation coefficient  $r$  (x-axis). Each point was averaged over 10 independent replicas, with the standard deviations reported by the vertical bars. The parameter  $\lambda$  was set to 0.0001 (see the text for its meaning).

## References

- [1] Newman, M.E.J., (2010) *Networks: An introduction*. Oxford: Oxford University Press.
- [2] Barabási, A.-L., (2016) *Network science*. Cambridge/England/UK: Cambridge University Press.
- [3] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. (2006) *Complex networks: Structure and dynamics*. Phys. Rep. **424**, 175.
- [4] Albert, R. & Barabasi, A.-L. (2002) *Statistical mechanics of complex networks*. Rev. Mod. Phys. **74**, 47.
- [5] Jackson, M. O. (2008) *Social and economic networks*. Oxford/England/UK: Princeton University Press.
- [6] Barabási, A.-L., Gulbahce, N. & Loscalzo, J. (2011) *Network medicine: a network-based approach to human disease*. Nat. Rev. Genet. **12**, 56.
- [7] Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge/England/UK: Cambridge University Press.
- [8] Pastor-Satorras, R. & Vespignani, A. (2010) *Complex networks: patterns of complexity*. Nat. Phys., **6**, 480.
- [9] Nicolis, G. & Prigogine I. (1977) *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York/US, London/UK, Sydney/Australia, Toronto/Canada: J. Wiley and Sons.
- [10] Barrat, A., Barthélemy, M. & Vespignani, A. (2008) *Dynamical Processes on Complex Networks*. 1st edn. Cambridge, England: Cambridge University Press.
- [11] Balescu, R. (1997) *Statistical dynamics: matter out of equilibrium*. London/UK: Imperial Coll.
- [12] ben Avraham, D. & Havlin, S. (2000) *Diffusion and reactions in fractals and disordered systems*. Cambridge, UK: Cambridge University Press.
- [13] Klafter, J. & Sokolov, I. M. (2011) *First steps in random walks: from tools to applications*. Oxford/UK: Oxford University Press.
- [14] Masuda, N., Porter, M. A., & R. Lambiotte (2017) *Random walks and diffusion on networks*. Phys. Rep., **716 – 717**, 1.
- [15] Martinez, V., Berzal, F. & Cubero, J.-C. (2016) *A survey of link prediction in complex networks*. ACM Comput. Surv. (CSUR), **49**, 4.
- [16] Daud, N. N., Ab Hamid, S. H., Saadoon, M., Sahran, F., & Anuar, N. B. (2020). *Applications of link prediction in social networks: A review*. J. Netw. Comput. Appl., **166**, 102716.

- [17] Brin, S. & Page, L. (1998) *The anatomy of a large-scale hypertextual Web search engine*. Computer Networks and ISDN Systems **30**, 107-117.
- [18] Rosvall, M. & Bergstrom, C. T. (2008) *Maps of random walks on complex networks reveal community structure*. Proc. Natl. Acad. Sci. USA, **105**, 1118.
- [19] Delvenne, J.-C., Yaliraki, S. N. & Barahona, M. (2010) *Stability of graph communities across time scales*. Proc. Natl. Acad. Sci. USA, **107**, 12755.
- [20] Lambiotte, R., Delvenne, J.-C. & Barahona, M. (2014) *Random walks, Markov processes and the multiscale modular organization of complex networks*. IEEE Trans. Netw. Sci. Eng., **1**, 76.
- [21] Lambiotte, R., Rosvall, M. (2012) *Ranking and clustering of nodes in networks with smart teleportation*. Phys. Rev. E **85**, 056107.
- [22] Carletti, T., Asllani, M., Fanelli, D. & Latora, V. (2020) *Nonlinear walkers and efficient exploration of congested networks*. Phys. Rev. Res., **2**, 033012.
- [23] Asllani, M., Carletti, T., Di Patti, F., Fanelli, D. & Piazza, F. (2018) *Hopping in the crowd to unveil network topology*. Phys. Rev. Lett., **120**, 158301.
- [24] de Kemmeter, J.-F., Carletti, T., & Asllani, M. (2022) *Self-segregation in heterogeneous metapopulation landscapes*. J. Theor. Biol., **554**, 111271.
- [25] Adam, I., Fanelli, D., Carletti, T. & Innocenti, G. (2019) *Reactive explorers to unravel network topology*, Eur. Phys. J. B, **92**, 1-8.
- [26] Colizza, V., Pastor-Satorras, R. & Vespignani, A. (2007) *Reaction-diffusion processes and metapopulation models in heterogeneous networks*, Nat. Phys., **3**, 276-282.
- [27] Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. (2008) *Critical phenomena in complex networks*, Rev. Mod. Phys., **80**, 1275-1335.
- [28] Pittorino, F., Ibáñez-Berganza, M., di Volo, M., Vezzani, A., & Burioni, R. (2017) *Chaos and correlated avalanches in excitatory neural networks with synaptic plasticity*, Phys. Rev. Lett., **118**, 098102.
- [29] Lei, C. & Ruan, J. (2013) *A novel link prediction algorithm for reconstructing protein-protein interaction networks by topological similarity*. Bioinformatics, **29**, 355-364.
- [30] Bornholdt, S. & Ebel, H. (2001) *World Wide Web scaling exponent from Simon's 1955 model*, Phys. Rev. E, **64**, 035104.
- [31] Fire, M., Katz, G., Elovici, Y., Shapira B. & Rokach, L. (2012) *Predicting student exam's scores by analyzing social network data*. In Huang, R. et al (Eds.), ATM2012, LNCS 7669, International Conference on ActiveMedia Technology, Springer-Verlang Berlin Heidelberg Germany. pp. 584–595.

- [32] Cardillo, A., Gómez-Gardenes, J., Zanin, M., Romance, M., Papo, D., Pozo, F. D. & Boccaletti, S. (2013) *Emergence of network features from multiplexity*. Sci. Rep., **3**, 1344.
- [33] Watts, D.J. & Strogatz, S.H. (1998) *Collective dynamics of small-world networks*, Nature, **393**, 440-442.
- [34] Lazega, E. (2001) *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford/UK: Oxford University Press.
- [35] Snijders, T.A.B., Pattison, P.E., Robins, G.L. & Handcock, M.S. (2006) *New specifications for exponential random graph models*. Sociol. Methodol. **393**, 99-153.
- [36] Nicosia, V. & Latora, V. (2015) *Measuring and modeling correlations in multiplex networks*, Phys. Rev. E, **92**, 032805.