Papers in Physics, vol. 17, art. 170002 (2025)

www.papersinphysics.org

Received: 28 March 2025, Accepted: 02 August 2025
Edited by: C. Anteneodo, Pontifícia Universidade Católica do Rio de Janeiro, Brazil

DOI: https://doi.org/10.4279/PIP.170002

ISSN 1852-4249

Random walks on random networks of cliques: Inferring the network structure

A Nannini[1]*, D H Zanette[1,2]†

We study the properties of discrete-time random walks on networks formed by randomly interconnected cliques, namely, random networks of cliques. Our purpose is to derive the parameters that define the network structure—specifically, the distribution of clique size and the abundance of inter-clique links—from the observation of selected statistical features along the random walk. To this end, we apply a Bayesian approach based on recording the times spent by the walker inside successively visited cliques. The procedure is illustrated with some numerical examples of diverse complexity, where the relevant structural parameters are successfully recovered.

I Introduction

Formally proposed more than two decades ago [1], network exploration is by now a well-established procedure for disclosing the structure of relational patterns in a wide variety of complex systems, covering biological and social community detection and identification [2, 3], survey of economic (commercial, industrial and financial) complexes [4], and big-data mining [5], among many other real-life applications. Suitably defined stochastic processes—concretely, random walks endowed with the capacity of recording selected features of their trajectories [6, 7]—have been pinpointed as a tool to efficiently explore such intricate structures.

In this contribution, we consider random walks evolving on a specific class of networks, namely, *random networks of cliques*. These networks are formed by small groups of fully interconnected nodes—the cliques—with more sparse, random connections between different groups. They have been introduced as a stylized model of interaction patterns with highly clustered architecture, for which many structural properties can be exactly computed [8]. Our main aim is to relate the statistical properties of the random walk to the structural features of the underlying network, both analytically and numerically, with the ultimate purpose of deriving the latter from the former.

In the next section, we recall the construction of random networks of cliques and review some of their structural features, which are necessary for our subsequent analysis. In Section III, random-walk statistical properties are analytically derived from those structural features. In Section IV, we study the inverse problem of deriving the network structure from the observation of the random walk, using a Bayesian inference approach. Three illustrative numerical examples are presented. Finally, we draw our conclusions in Section V.

* albano.nannini@ib.edu.ar
† damian.zanette@ib.edu.ar

[1] Centro Atómico Bariloche and Instituto Balseiro, Comisión Nacional de Energía Atómica and Universidad Nacional de Cuyo, 8400 San Carlos de Bariloche, Río Negro, Argentina.

[2] Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

II Random networks of cliques (RNoCs)

A random network of cliques (RNoC) is built by first constructing Q cliques. We recall that a clique is a (typically small) group of nodes fully connected to each other. The size n of each clique—namely, its number of nodes—is drawn at random from a prescribed probability distribution f_n ($n = 1, 2, \dots$). The expected total number of nodes in the network is thus $N = Q\langle n\rangle$, with $\langle n\rangle$ the mean value of n over the distribution f_n. Then, M *inter-clique* links are established between randomly chosen nodes belonging to different cliques, with the condition that at most one inter-clique link reaches any given node. The number M is chosen in such a way that the resulting fraction of nodes connected to inter-clique links has a prescribed value γ. This is achieved by taking $M = N\gamma/2$. Figure 1 shows an example of this construction for a small number of cliques. For future reference, we quote the probability $F_{n,m}$ of having a clique of size n with m inter-clique links ($0 \le m \le n$):

$$F_{n,m} = \binom{n}{m}\gamma^m(1-\gamma)^{n-m}f_n. \tag{1}$$

As shown in previous work [8], the simple rules used to build RNoCs enable a straightforward calculation of their structural properties, such as degree distribution, clustering, assortativity, and diameter. In many respects, the RNoC structure is similar to that of small-world networks [9, 10]. To our present purposes, a crucial question is whether a giant connected component exists or not. In fact, for a random walk to sample a statistically significant part of the network, it must take place on the giant component. We recall that, strictly speaking, a giant component is well defined for an infinitely large network only as a connected component that contains a finite fraction of the whole network [11]. In finite, but large, networks, it is identified as the largest connected component.

Since cliques are internally fully connected, whether a giant component is present in the RNoC is determined by the random pattern of inter-clique links. Making use of the generating function for the degree distribution, the theory of random networks [12] provides the tools to assess the existence and the size of the giant component. Concretely, it

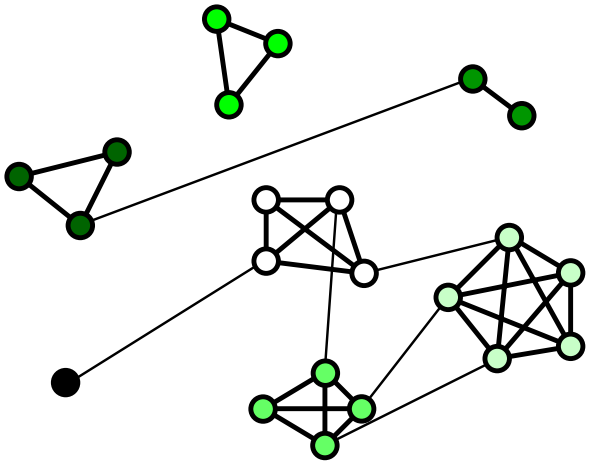

Figure 1: A random network of cliques with $Q = 7$ cliques, $N = 22$ nodes and $M = 6$ inter-clique links. The fraction of nodes with inter-clique links is $\gamma = 6/11 \approx 0.55$. In the plot, inter-clique links are thinner than links inside cliques, and nodes in different cliques are colored with different shades.

can be shown that a giant component comprising a fraction

$$s = 1 - G_0(u) \tag{2}$$

of all cliques exists in a RNoC when the equation $u = G_1(u)$ has a solution $0 < u < 1$. Here,

$$G_0(x) = \sum_{n,m} x^m F_{n,m} = \langle(1-\gamma+\gamma x)^n\rangle \tag{3}$$

is the generating function for the probability distribution of the number of inter-clique links, and

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)} = \frac{\langle n(1-\gamma+\gamma x)^{n-1}\rangle}{\langle n\rangle}, \tag{4}$$

where primes denote differentiation with respect to x. As above, $\langle\cdot\rangle$ stands for the average over the clique size distribution f_n. The solution for u effectively exists when the network is sufficiently connected by inter-clique links, i.e. when γ is larger than a certain critical value determined by the distribution f_n [8].

To relate the statistical features of a random walk on the giant component to the structural properties of the network, it is important to take into account that the distribution of clique sizes and inter-clique links is not the same inside the giant component

as over the entire network. Using standard results on the "microscopic" structure of the giant component [13], we find that the joint distribution of clique sizes and inter-clique connections inside it reads

$$F^G_{n,m} = \frac{1-u^m}{s} F_{n,m}, \tag{5}$$

where u and s are given by Eqs. (2) to (4), and $F_{n,m}$ is the corresponding distribution all over the RNoC, Eq. (1).

III Statistics of random walks on RNoCs

Now, we consider a random walker moving on the giant component of a RNoC in discrete time. At each time step, the walker jumps from the node it occupies to one of the neighbor nodes, chosen at random with equal probability. The new node can belong to the same clique or, if the original node had an inter-clique link, to a different one.

Having in mind the aim of inferring the structural features of the RNoC from the random walk statistics, our main interest is to find the frequency with which the walker stays for a certain time in a clique before jumping to another one. To this end, we first compute the probability $\Pi_{n,m}(T)$ that the walker spends exactly T steps inside a clique of size n with m inter-clique links. Along a single visit of the random walker to any given clique, the evolution can be conceived as a Markov process with three different states:

1. the walker is on a node *without* inter-clique link;
2. the walker is on a node *with* inter-clique link;
3. the walker has left the clique.

The time-dependent probabilities of each state can be arranged in a vector $\mathbf{p}(t) = [p_1(t), p_2(t), p_3(t)]$, which evolves in time as $\mathbf{p}(t+1) = \mathcal{M}\mathbf{p}(t)$. For a clique with n nodes and m inter-clique links ($1 \leq m \leq n$), the transference matrix reads

$$\mathcal{M} = \begin{pmatrix} (n-m-1)/(n-1) & (n-m)/n & 0 \\ m/(n-1) & (m-1)/n & 0 \\ 0 & 1/n & 1 \end{pmatrix}. \tag{6}$$

The probabilities at time t, given by

$$\mathbf{p}(t) = \mathcal{M}^{t-1}\mathbf{p}(1), \tag{7}$$

have to be evaluated using the initial condition $\mathbf{p}(1) = (0, 1, 0)$. The elements of $\mathcal{M}$ in Eq. (6) result from straightforward counting of the events that lead from one state to another.

The probability $\Pi_{n,m}(T)$ is given by the product of the probability that the walker is in state 2 at time T, times the probability that it leaves the clique in the next jump, $1/n$. Namely,

$$\Pi_{n,m}(T) = \frac{1}{n} p_2(T) = \frac{1}{n} \mathbf{p}(1) \cdot \mathcal{M}^{T-1}\mathbf{p}(1), \tag{8}$$

where $\cdot$ indicates scalar product. An explicit form for $\Pi_{n,m}(T)$ can be obtained by diagonalizing $\mathcal{M}$, whose eigenvalues turn out to be $\lambda_1 = 1$ and

$$\lambda_{2,3} = \frac{(n-1)^2 - m \mp R}{2n(n-1)}. \tag{9}$$

Here, $R = \sqrt{(n^2-1)^2 - 2m(n-1)^2 + m^2}$, and the upper and lower signs correspond to λ_2 and λ_3, respectively. Operating with the diagonalized version of $\mathcal{M}$ in Eq. (8), we find

$$\begin{aligned} \Pi_{n,m}(T) =& \lambda_2^{T-1} \frac{R + n^2 - 2nm + m - 1}{2nR} \\ &+ \lambda_3^{T-1} \frac{R - n^2 + 2nm - m + 1}{2nR}, \end{aligned} \tag{10}$$

which is correctly normalized for all n and m:

$$\sum_{T=1}^{\infty} \Pi_{n,m}(T) = 1. \tag{11}$$

For $n > 1$ and $m \leq n$, the eigenvalues satisfy $\lambda_2 < 0 < \lambda_3$ and $|\lambda_2| < \lambda_3 < 1$. These inequalities imply that, for large T, $\Pi_{n,m}(T)$ is dominated by the second term in Eq. (10), and therefore decreases as $\Pi_{n,m}(T) \sim \lambda_3^T$. Moreover, since λ_2 is negative, an oscillatory dependence is expected for small T due to the alternating sign of the first term.

Once $\Pi_{n,m}(T)$ has been evaluated, we can compute the probability $P(T)$ that a random walker on the giant component of a given RNoC stays exactly T steps inside any of its cliques. This probability results from a sum of the contributions of cliques of each size n and each number of inter-clique links m weighted by the frequency of such cliques, which is

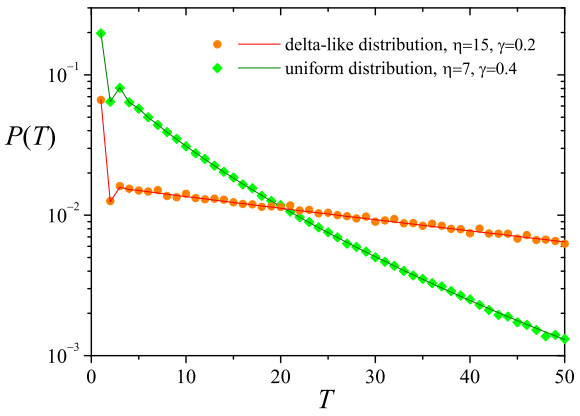

Figure 2: Probability $P(T)$ that the random walker spends T steps inside a clique, Eq. (12), for two distributions of clique sizes f_n, Eqs. (13) and (14), and two values of the fraction of nodes with inter-clique links, γ. Symbols correspond to simulation results, while lines join the values of $P(T)$ computed using Eq. (12).

given by the probability distribution $F^G_{n,m}$ obtained in the preceding section, Eq. (5). Moreover, each contribution is weighted by the probability of entering a given clique, which is proportional to the number of its inter-clique links, m. The result is

$$P(T) = \frac{\sum_n \sum_{m=1}^{n} m F^G_{n,m} \Pi_{n,m}(T)}{\sum_n \sum_{m=1}^{n} m F^G_{n,m}}, \qquad (12)$$

with the first sum running over the relevant values of n, which depend on the distribution of clique sizes. The normalization of $P(T)$ is ensured by that of $\Pi_{n,m}(T)$, Eq. (11). Generally, $P(T)$ cannot be given an explicit analytical form but can be straightforwardly computed by numerical means once $F^G_{n,m}$ has been specified.

Direct application of the result for $P(T)$, Eq. (12), to the statistics of the time spent inside each clique along an actual random walk on a RNoC requires neglecting correlations between successive visits to different cliques. In fact, after abandoning a given clique to enter a neighbor, there is a relatively large probability that the walker returns to the previous clique. This kind of correlation, however, should become less important as the random walk progresses, covering increasingly large portions of the network.

To confirm this conjecture, we have simulated random walks on the giant component of RNoCs with two distributions of clique sizes, namely, a delta-like distribution,

$$f_n = \delta_{n,\eta}, \qquad (13)$$

where all cliques have size η, and a uniform distribution,

$$f_n = \begin{cases} 1/(\eta - 2) & \text{for } 3 \le n \le \eta, \\ 0 & \text{otherwise,} \end{cases} \qquad (14)$$

where all the sizes between 3 and η have the same probability. Depending on the value of η, the fraction of nodes with inter-clique links, γ, has been chosen to ensure the presence of a well-developed giant component. Each random walk was conducted on a RNoC consisting of $Q = 10^5$ cliques, along 10^7 time steps. Figure 2 shows the results in two illustrative cases: the delta-like distribution, Eq. (13) with $\eta = 15$, for $\gamma = 0.2$, and the uniform distribution, Eq. (14) with $\eta = 7$, for $\gamma = 0.4$. Symbols stand for the results of numerical realizations of the random walk, and lines join the values of $P(T)$ obtained from Eq. (12) as a function of T. Up to small random fluctuations, the excellent agreement between numerical and analytical results strongly supports the above conjecture. Moreover, the plots illustrate the oscillations expected in $P(T)$ for small values of T due to the opposite signs of λ_2 and λ_3 in Eq. (10), as well as the exponential decrease for large T.

IV Bayesian inference of the RNoC structure

To apply the above results to the inference of the RNoC structure, we assume that, as the result of observing a random walk on the network, a vector $\mathbf{T} = (T_1, T_2, \ldots, T_K)$ has been recorded, whose components T_k are the times spent by the walker inside successively visited cliques. Our aim is to compute the conditional probability $P(\mathbf{\Phi}|\mathbf{T})$ that the network has a certain structure given the observed vector $\mathbf{T}$. Here, $\mathbf{\Phi}$ denotes a vector whose components are the parameters that specify the RNoC structure. These encompass the fraction of nodes with inter-clique links, γ, and the parameters needed to determine the distribution of clique sizes f_n. Generally, among the components of $\mathbf{\Phi}$, some

parameters may be known with certainty, while others are to be determined from probabilistic inference.

By virtue of Bayes' theorem [14], we have

$$P(\mathbf{\Phi}|\mathbf{T}) = \frac{P(\mathbf{T}|\mathbf{\Phi})}{P(\mathbf{T})} P(\mathbf{\Phi}), \tag{15}$$

where $P(\mathbf{T}|\mathbf{\Phi})$ is the probability of observing the vector $\mathbf{T}$ conditioned on the parameter set $\mathbf{\Phi}$, and $P(\mathbf{T}) = \sum_{\mathbf{\Phi}} P(\mathbf{T}|\mathbf{\Phi})P(\mathbf{\Phi})$. In turn, $P(\mathbf{\Phi})$ is the *a priori* probability of the set $\mathbf{\Phi}$, to be assigned on the basis of a suitable hypothesis, as explained later. In our case, Bayesian inference consists in estimating the most likely value of $\mathbf{\Phi}$ from the probability $P(\mathbf{\Phi}|\mathbf{T})$ calculated as in Eq. (15), using the information obtained from the measurements of $\mathbf{T}$. This requires applying Eq. (15) to each possible parameter set $\mathbf{\Phi}$ in order to evaluate the corresponding $P(\mathbf{\Phi}|\mathbf{T})$, and then selecting the set $\mathbf{\Phi}$ for which the probability is maximal. In practice, $P(\mathbf{\Phi}|\mathbf{T})$ is numerically computed for a discrete, but large, ensemble of parameter sets $\mathbf{\Phi}$, and its maximum over this ensemble is then detected.

For each parameter set $\mathbf{\Phi}$, the conditional probability $P(\mathbf{T}|\mathbf{\Phi})$ can be computed using the results obtained in Section III, as

$$P(\mathbf{T}|\mathbf{\Phi}) = \prod_{k=1}^{K} P(T_k|\mathbf{\Phi}), \tag{16}$$

where $P(T_k|\mathbf{\Phi})$ is the probability $P(T)$ given by Eq. (12), calculated for an RNoC with parameters $\mathbf{\Phi}$ and evaluated on time T_k. In Eq. (16), we have assumed that the probabilities of the components of $\mathbf{T}$ are mutually independent, as supported by the numerical results presented at the end of the preceding section. Thus, $P(\mathbf{T}|\mathbf{\Phi})$ is given by the product of the probabilities of all the individual components of $\mathbf{T}$.

In order to minimize the information provided to evaluate $P(\mathbf{\Phi}|\mathbf{T})$, we assign the same *a priori* probability to all the sets $\mathbf{\Phi}$, say $P(\mathbf{\Phi}) = P_0$ for all $\mathbf{\Phi}$. With this choice, Eq. (15) can be rewritten as

$$P(\mathbf{\Phi}|\mathbf{T}) = \frac{P(\mathbf{T}|\mathbf{\Phi})}{P(\mathbf{T})} P_0 \equiv \frac{P(\mathbf{T}|\mathbf{\Phi})}{Z(\mathbf{T})}, \tag{17}$$

with $Z(\mathbf{T}) = P(\mathbf{T})/P_0 = \sum_{\mathbf{\Phi}} P(\mathbf{T}|\mathbf{\Phi})$. Using Eq. (17), we compute $P(\mathbf{\Phi}|\mathbf{T})$ by first evaluating $P(\mathbf{T}|\mathbf{\Phi})$ from Eq. (16), and then normalizing over the relevant values of $\mathbf{\Phi}$ to find $Z(\mathbf{T})$. Once $P(\mathbf{\Phi}|\mathbf{T})$ has been obtained for each $\mathbf{\Phi}$, the unknown network parameters inferred from the vector $\mathbf{T}$ are those with the highest probability, namely those that maximize $P(\mathbf{\Phi}|\mathbf{T})$.

To assess the performance of this procedure, we operate as follows. First, we generate a RNoC with a prescribed set of parameters $\mathbf{\Phi}_0$. Next, a random walk is carried out on its giant component, recording the times T_k spent by the walker in the successive cliques it visits. This provides the vector $\mathbf{T} = (T_1, T_2, \ldots, T_K)$. Then, using Eqs. (16) and (17), we calculate the conditional probability $P(\mathbf{\Phi}|\mathbf{T})$ for a large ensemble of parameter sets $\mathbf{\Phi}$ and detect the set $\mathbf{\Phi}_{\max}$ for which this probability reaches a maximum. Finally, $\mathbf{\Phi}_{\max}$ is compared with $\mathbf{\Phi}_0$. The larger the coincidence, the better the performance.

i Inferring γ

As a first example, we assume that we know the distribution of clique sizes in the RNoC, f_n, but that the fraction of nodes with inter-clique links γ is unknown and has to be inferred as explained above. To test the procedure in this situation, we have taken f_n as the delta-like distribution of Eq. (13), with $\eta = 7$, and have built a RNoC with $Q = 10^5$

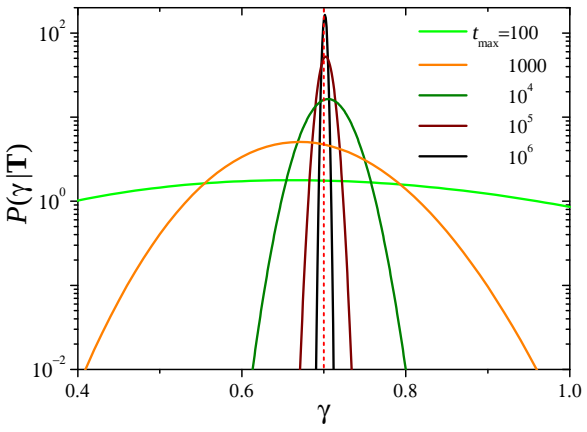

Figure 3: The probability distribution $P(\gamma|\mathbf{T})$ as a function of γ, computed from six random walks with different length $t_{\max}$ on the giant component of a RNoC with cliques of identical sizes $n = 7$. The dashed vertical line indicates the fraction of nodes with inter-clique links used to build the network, $\gamma_0 = 0.7$.

cliques and $\gamma_0 = 0.7$. On this network, we performed random walks of different lengths $t_{\max}$ and, for each random walk, we determined the vector $\mathbf{T}$ containing the times spent by the walker inside successively visited cliques.

Figure 3 shows the probability distributions $P(\gamma|\mathbf{T})$ calculated as prescribed by Eqs. (16) and (17), for six random walks with various values of $t_{\max}$. For each walk, $P(\gamma|\mathbf{T})$ has been computed as a function of γ in the interval $(0, 1)$ with discretization $\delta\gamma = 10^{-4}$, i.e. for an ensemble of some 10^4 values of γ. The results clearly show that $P(\gamma|\mathbf{T})$ becomes increasingly concentrated around a well-defined value of γ as $t_{\max}$ grows and the random walk samples the network more thoroughly. For $t_{\max} = 10^6$, $P(\gamma|\mathbf{T})$ attains its maximum at $\gamma_{\max} = 0.7013(1)$, which constitutes our best estimate for γ from this specific realization. The small difference between $\gamma_{\max}$ and γ_0, just below 0.2%, can be ascribed to the fact that inter-clique links in the RNoC are established at random as it is built, so that the fraction of nodes with such links in a given realization of the network can slightly differ from γ_0.

To evaluate the dispersion of the results obtained from different realizations of the random walk, we performed sets of 100 random walks for each value of $t_{\max}$. For each random walk, we have computed $P(\gamma|\mathbf{T})$ as a function of γ—now, with discretization $\delta\gamma = 10^{-2}$—and numerically evaluated the mean value

$$\langle\gamma\rangle = \int_0^1 \gamma P(\gamma|\mathbf{T}) d\gamma, \tag{18}$$

and the corresponding standard deviation σ_γ. For the peaked profiles of $P(\gamma|\mathbf{T})$, as in Fig. 3, the mean value $\langle\gamma\rangle$ yields an excellent estimate of the position of the maximum, $\gamma_{\max}$. Results are shown in the main panel of Fig. 4, where each dot corresponds to a single realization. We see that, as $t_{\max}$ grows, not only does σ_γ decrease, as already shown in Fig. 3, but the values of $\langle\gamma\rangle$ become more consistent between realizations. The inset shows the standard deviation averaged over realizations, $\bar{\sigma}_\gamma$, as a function of $t_{\max}$. The straight segment has a slope of $-1/2$, indicating that the average standard deviation decreases as $\bar{\sigma}_\gamma \sim t_{\max}^{-1/2}$ along the considered interval.

We mention that we have also tested the complementary task of inferring η for delta-like and uniform distributions, Eqs. (13) and (14), now assuming that γ is known. In this case, successful inference requires shorter random walks, because η is a discrete (integer) parameter. In fact, the probability $P(\eta|\mathbf{T})$ already has a maximum at the correct value of η for random walks of length between 10^3 and 10^4. For brevity, the results for these cases are not presented here.

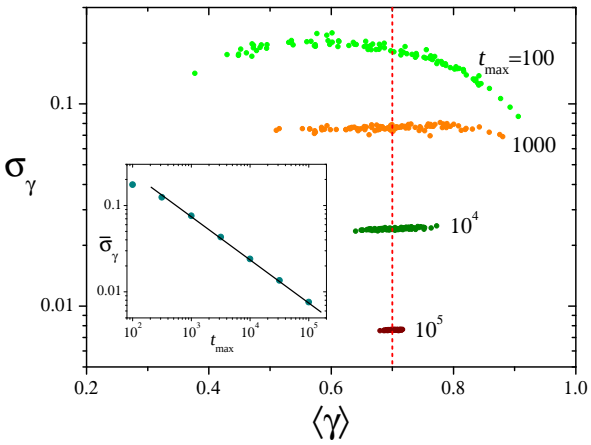

Figure 4: Main panel: Mean value $\langle\gamma\rangle$ and standard deviation σ_γ for the probability distribution $P(\gamma|\mathbf{T})$ obtained from random walks of length $t_{\max}$ on a RNoC with cliques of identical sizes $n = 7$. Each dot corresponds to a single random walk, among 100 for each value of $t_{\max}$. The dashed vertical line indicates the fraction of nodes with inter-clique links used to build the network, $\gamma_0 = 0.7$. Inset: The standard deviation averaged over realizations, $\bar{\sigma}_\gamma$, as a function of $t_{\max}$. The slope of the straight segment equals $-1/2$.

ii Simultaneous inference of γ and η

In our second example, we use the above procedure to infer two parameters, namely, γ and η, assuming that the clique sizes are uniformly distributed according to Eq. (14). To test the method, we constructed a RNoC with 10^5 cliques using $\gamma_0 = 0.75$ and $\eta_0 = 10$. The results presented here were obtained from three realizations of the random walk on the giant component of the network, with $t_{\max} = 10^4$, 10^5, and 10^6 steps, respectively.

Figure 5 depicts, in a color scale, the probability distribution $P(\gamma, \eta|\mathbf{T})$ on the plane (γ, η), as obtained from Eqs. (16) and (17) for an ensemble of some 5000 pairs (γ, η), using the vectors $\mathbf{T}$ recorded

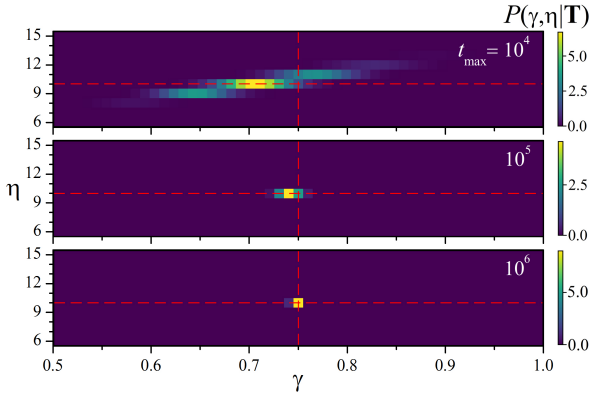

Figure 5: The probability $P(\gamma, \eta|\mathbf{T})$, in a color scale, on the plane spanned by the parameters γ and η, as estimated from three random walks of different length $t_{\max}$ on the giant component of a RNoC with $\gamma_0 = 0.75$ and a uniform distribution of clique sizes, Eq. (14), with $\eta_0 = 10$. The discretization step in γ is $\delta\gamma = 10^{-2}$. Straight dashed lines stand for the values of γ_0 and η_0.

along the three random walks. Dashed lines stand for the values of γ_0 and η_0. The plots show clearly that, much as observed in the previous example, the probability becomes more concentrated and better centered around the expected values of the parameters as the random walk increases in length.

Table 1 presents the mean values and standard deviations of γ and η derived from the distribution $P(\gamma, \eta|\mathbf{T})$ for each random walk. Note that the estimate of the discrete parameter η converges faster and more accurately to its expected value than that of γ. In any case, for $t_{\max} = 10^6$, both γ_0 and η_0 lie inside the intervals defined by the mean value and the standard deviation of each parameter.

$t_{\max}$	$\langle\gamma\rangle$	σ_γ	$\langle\eta\rangle$	σ_η
10^4	0.71	0.06	10.1	0.9
10^5	0.741	0.009	10.00	0.04
10^6	0.749	0.003	10.00	$< 10^{-5}$

Table 1: Mean value and standard deviation of the parameters γ and η deriving from the estimate of the probability $P(\gamma, \eta|\mathbf{T})$, as obtained from the three random walks described in the text.

iii Parameter inference by gradient ascent

When the network parameters to be inferred from a random walk on a RNoC are continuous, and/or when their number increases, the calculation of the probability $P(\mathbf{\Phi}|\mathbf{T})$ all over the multidimensional space spanned by $\mathbf{\Phi}$ may become computationally impractical. In such cases, a better strategy may be needed to efficiently locate the point where $P(\mathbf{\Phi}|\mathbf{T})$ attains its maximum. An alternative procedure consists in implementing a biased random walk on parameter space (not to be confused with the random walk on the RNoC, used to obtain $\mathbf{T}$), whose steps only occur in directions along which the probability grows. In this way, $P(\mathbf{\Phi}|\mathbf{T})$ needs to be calculated only along the random walk path and its immediate neighborhood. This is nothing but a *gradient ascent* procedure [15], which ensures that, for sufficiently long times, a local maximum of the probability is eventually reached.

We have implemented this method to infer two continuous parameters of a RNoC whose cliques have two possible sizes ($n = 3$ or 4) distributed with different frequencies, namely,

$$f_n = \begin{cases} \xi & \text{for } n = 3, \\ 1 - \xi & \text{for } n = 4, \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

with $0 < \xi < 1$. The parameters to be estimated are γ and ξ. In the specific example presented below, we have built the RNoC of 10^5 cliques using $\gamma_0 = 0.75$ and $\xi_0 = 0.5$. Then, we have run a single random walk of 10^7 steps on its giant component, in order to obtain the vector $\mathbf{T}$.

The gradient-ascent random walk on the parameter space (γ, ξ)—which, for computational representation, is tessellated with discretization $\delta\gamma$ and $\delta\xi$ in each direction—is carried out as follows. An initial position for the walker, say (γ_i, ξ_i), is chosen at random in the relevant parameter interval and the probability $P(\gamma_i, \xi_i|\mathbf{T})$ is computed. Then, a jump to a randomly chosen nearest-neighbor cell, say (γ_j, ξ_j), is attempted. If $P(\gamma_i, \xi_i|\mathbf{T}) < P(\gamma_j, \xi_j|\mathbf{T})$ the jump is accepted and the walker moves. Otherwise, another potential new position (γ_j, ξ_j) is selected. If the jump is rejected a prescribed number of times τ, the process ends. This procedure is expected to lead to the "trapping" of the random walker in a cell where the probability attains a local maximum.

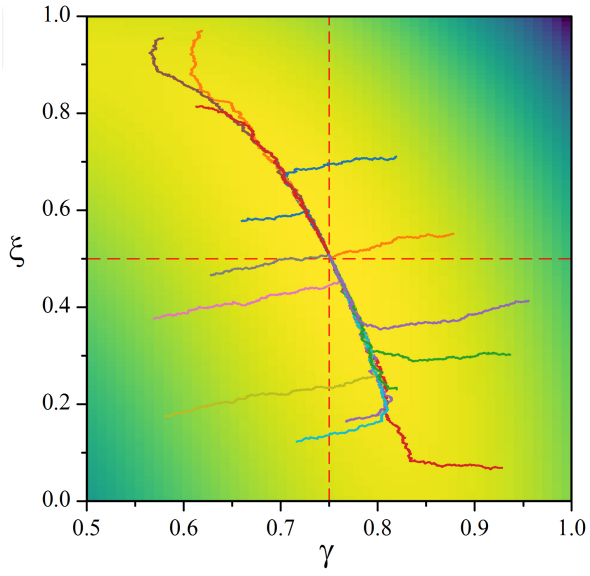

Figure 6: Fifteen trajectories of a gradient-ascent random walk on the parameter plane (γ, ξ), used to find the maximum of the probability $P(\gamma, \xi|\mathbf{T})$, as explained in the text. Dashed lines stand for the values γ_0 and ξ_0 of the RNoC whose parameters are inferred. The background represents $P(\gamma, \xi|\mathbf{T})$ in a color scale, with lighter shades corresponding to higher values.

Figure 6 shows 15 trajectories of the gradient-ascent random walker on the plane (γ, ξ), starting from different initial positions, with $\delta\gamma = \delta\xi = 10^{-3}$ and $\tau = 10$. It is apparent that, in all cases, the walker converges toward the expected point (γ_0, ξ_0), at the intersection of the dashed lines. The final positions of all the trajectories yield $\gamma = 0.750(1)$ and $\xi = 0.50(1)$ for the inferred parameters. Note also that all trajectories exhibit a first stage where the walker approaches a curved manifold and then proceeds along it. The slope of the manifold near (γ_0, ξ_0) indicates that the convergence is considerably faster along γ than along ξ.

V Conclusions

In this paper, we have tested a procedure to derive information on network structure from the observation of random walks evolving on the network, or, more precisely, on its giant component. Concretely, we have implemented the method for random networks of cliques (RNoCs), for which we derived the parameters defining the distribution of clique sizes and the density of inter-clique connections from the sequence of times spent by the random walker inside successively visited cliques. Parameter estimation was based on Bayesian inference, under the assumption of uniform *a priori* probability. Within this formulation, Bayes' theorem allows for the calculation of a probability distribution over parameter space, whose maximum is associated with the best estimate. For two of the examples presented here, we have shown how the results improve in precision as the random walk becomes longer. In all cases, we have obtained excellent estimates.

As we have demonstrated through specific examples, the procedure can be implemented in two alternative ways. When the number of parameters to be inferred is small, the probability distribution can be computed all over the relevant zone of parameter space, and its maximum can be detected with satisfactory precision. As the number of parameters grows, however, the calculation on a multi-dimensional space may become computationally very expensive. In this case, the position of the maximum can be found more efficiently using other methods. Here, we have used a random walk driven by gradient ascent but, if a complicated probability landscape is foreseen, a Monte Carlo algorithm could provide a more exhaustive exploration of parameter space.

Network exploration using random walks requires that, besides the rules that define its dynamics, the walker (or its observer) is endowed with the capability of recording some selected network features as the process progresses. In our case, we have assumed that it was possible to detect when the walker jumps between cliques, for instance, discerning between intra- and inter-clique connections. This assumption was essential to provide the series of times spent inside cliques, which fed the Bayesian approach. Inference based on recording other features along the random walk may be the subject of future work.

Acknowledgements - A. Nannini is supported by a scholarship of Comisión Nacional de Energía Atómica, Argentina, at Instituto Balseiro.

[1] S H Strogatz, *Exploring complex networks*, Nature **410**, 268 (2001).

[2] X Huang, D Chen, T Ren, D Wang, *A survey of community detection methods in multilayer networks*, Data Min. Knowl. Disc. **35**, 1 (2021).

[3] M Fukuda, K Nakajima, K Shudo, *Estimating the bot population on Twitter via random walk based sampling*, IEEE Access **10**, 17201 (2022).

[4] A R Faroque, S C Morrish, O Kuivalainen, S Sundqvist, L Torkkeli, *Microfoundations of network exploration and exploitation capabilities in international opportunity recognition*, Int. Business Rev. **30**, 101767 (2021).

[5] A Baptista, A González, A Baudot, *Universal multilayer network exploration by random walk with restart*, Commun. Physics **5**, 170 (2022).

[6] L Dall'Asta, I Alvarez Hamelin, A Barrat, A Vázquez, A Vespignani, *Statistical theory of Internet exploration*, Phys. Rev. E **71**, 036135 (2005).

[7] N Masuda, M A Porter, R Lambiotte, *Random walks and diffusion on networks*, Phys. Rep. **716-717**, 1 (2017).

[8] L A Sobehart, S Martínez Alcalá, A Chacoma, D H Zanette, *Structural properties of random networks of cliques*, Physica A **624**, 128998 (2023).

[9] L A Sobehart, D H Zanette, *Critical behavior of rumor propagation on random networks of cliques*, Pap. Phys. **16**, 160003 (2024).

[10] D J Watts, S H Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393**, 440 (1998).

[11] M Newman, *Networks. An Introduction*, Oxford University Press, Oxford (2018).

[12] M E J Newman, S H Strogatz, D J Watts, *Random graphs with arbitrary degree distribution and their applications*, Phys. Rev. E **64**, 026118 (2001).

[13] I Tishby, O Biham, E Katzav, R Kühn, *Revealing the microstructure of the giant component in random graph ensembles*, Phys. Rev. E **97**, 042318 (2018).

[14] W von der Linden, V Dose, U von Toussaint, *Bayesian Probability Theory: Applications in the Physical Sciences*, Cambridge University Press, Cambridge, UK (2014).

[15] S Boyd, L Vanderberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK (2004).