Generalized Markov stability of network communities

Aurelio Patelli,1,2 Andrea Gabrielli,2,3 and Giulio Cimini2,3

1Service de Physique de l’Etat Condensé, UMR 3680 CEA-CNRS, Université Paris-Saclay, CEA-Saclay, 91191 Gif-sur-Yvette (France)
2Istituto dei Sistemi Complessi (CNR) UoS Sapienza, Dipartimento di Fisica, “Sapienza” Università di Roma, P.le A. Moro 2, 00185 Rome (Italy)
3IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca (Italy)

We address the problem of community detection in networks by introducing a general definition of Markov stability, based on the difference between the probability fluxes of a Markov chain on the network at different time scales. The specific implementation of the quality function and the resulting optimal community structure thus become dependent both on the type of Markov process and on the specific Markov times considered. For instance, if we use a natural Markov chain dynamics and discount its stationary distribution – that is, we take as reference process the dynamics at infinite time – we obtain the standard formulation of the Markov stability. Notably, the possibility to use finite-time transition probabilities to define the reference process naturally allows detecting communities at different resolutions, without the need to consider a continuous-time Markov chain in the small time limit. The main advantage of our general formulation of Markov stability based on dynamical flows is that we work with lumped Markov chains on network partitions, having the same stationary distribution of the original process. In this way the form of the quality function becomes invariant under partitioning, leading to a self-consistent definition of community structures at different aggregation scales.

I. INTRODUCTION

Networks are systems made up of entities (nodes) embedded in a complex pattern of interconnections (links), which occur in a large variety of contexts – ranging from socio-economic systems and infrastructures to biological processes and ecosystems [1–4]. Networks observed in nature have a recurrent set of characteristics, such as fat-tail behavior of the degree distribution, small-world topology, and community structure – the latter referring to the internal organization of nodes into densely connected groups. Identifying the communities of a network means uncovering its mesoscopic structure, and is still an outstanding challenge for network science [5,7].

The first method proposed in the literature to partition a network in communities is based on the maximization of a quality function, the modularity, which compares the actual number of links in the network falling inside each community to the expectation of such number under a null network model [8]. The modularity function has been then generalized to various setups, like directed, weighted or bipartite networks (see e.g. [9–10]), and still nowadays represents the benchmark method for community detection [6]. However, by relying on a global null model, the modularity suffers from a resolution limit, that is, it cannot find communities smaller than a minimum size – which depends on the scale of the whole system [11]. Multi-resolution versions of the modularity address this issue using a tunable resolution parameter [12,13], whereas, the modularity-density functional employs a penalty function for splitting partitions [14]. Another popular approach to community detection consists in fitting the network to a stochastic blockmodel, namely a random graph with built-in communities [15], yet this approach was recently shown to be equivalent to modularity maximization [16]. Other well known community detection methods use clique percolation [17], spectral graph properties [18], spin glass models [13,19,20] or combinatorial arguments – notably this latter method, Surprise [21,22], is nearly unaffected by the resolution limit, but has the opposite drawback of overestimating the number of communities [23].

Another popular branch of community detection methods is based on random walks [24]. The idea is that communities correspond to network regions where the walker’s dynamics spends a relatively long time, because of the high density of links within communities and the sparse connections across communities. This phenomenon leads to the definition of a quality function known as Markov stability [25]. Notably, Markov stability allows interpolating between modularity and spectral clustering by simply varying the time scale of the dynamics [26]. Indeed, such a time scale effectively acts as a resolution parameter, with short scales leading to many small communities and long scales to a few large communities [25,27]. Using continuous-time random walks in the small time limit can even overcome the resolution limit of the modularity [25]. Among related methods, the Walktrap algorithm has been one of the first to use random walks for inferring similarities between nodes whence the network community structure [25]. The popular Infomap algorithm instead puts the community detection problem in information-theoretical terms [29,30]: the functional to be optimized with respect to the network partition is the description length for the moves of a random walker on the network. Hence the codebook and the codewords are based on the transition probabilities and stationary distribution of the random walk. Related to this, methods based on Boltzmann minimum description length have recently been proposed [31]. Random
walks have also been used to partition the links (rather than the nodes) of the network, and thus to uncover community structures using the concept of the line graph.

The plethora of community detection methods give similar but not identical results, and indeed no algorithm seems to be optimal for all possible community detection tasks. This happens because community detection is an ill-defined problem: there is no universal definition of communities, and thus no clear guidelines on how to build and assess a community detection method. For instance, the approaches based on the network topology (modularity and blockmodel) or on link combinatorics (surprise) use a null network model to assess the statistical significance of a network partition, and the freedom in choosing the null model introduces a degeneracy in the definition of the community structure. Physics-inspired methods suffer from the same pathology, since changing the definition of the interaction between nodes and the strength of the noise give different phases, whereas, methods based on random walks find different partitions depending on the particular dynamics implemented on the network.

Given that the quest for the “best method” to detect the ground truth communities of any network is possibly vain, here we follow up on the complementary viewpoint of random walks methods that any given dynamical process on the network induces a different community structure. We thus consider a general Markov diffusion process on the network, and derive a general quality function for the optimization problem using the transition probability fluxes of the dynamics at different time scales. In this way we generalize previous definitions of the Markov stability, which compare the Markov dynamics at finite times to a reference process given by its stationary distribution (i.e., the dynamics at infinite time). Indeed by varying the time scales of the Markov dynamics and of the reference process we can detect communities at both higher and lower resolutions. Remarkably, our approach is grounded on the definition of lumped Markov chains on network partitions, whose stationary distributions follow the same aggregating rules of the dynamics. Thanks to this property the form of the quality function becomes invariant under network partitioning, leading to a self-consistent definition of hierarchical structures at different scales. This leads not only to an elegant theoretical formulation of the problem but also to a convenient recursive algorithm for the optimization of the quality function.

II. MARKOV CHAIN ON NETWORKS

We start by recalling basic definitions and properties of Markov chains on networks. We then introduce lumped Markov chains on network partitions, and illustrate these concepts in the simple case of the natural Markov chain (i.e., the random walk).

A network is a set \( \mathcal{N} \) of \( N \) nodes, whose pattern of interconnections is described by the adjacency matrix – with generic element \( A_{ij} \) giving the weight of the link from node \( i \) to node \( j \) (in the case of binary networks, \( A_{ij} = 1 \) if the link \( i \to j \) exists and 0 otherwise). A Markov chain on a network is a discrete-time stochastic process that defines a temporal sequence of nodes (the possible states of the chain), and that satisfies the Markov property: the probability to be in any state at a given time step depends only on the state attained at the previous step. The process is thus described by the set of probabilities \( \{ p_{ij} \}_{j \in \mathcal{N}} \) of jumping from node \( i \) to node \( j \) at a given time step.

A Markov chain is ergodic if it is non-periodic and in the long time regime it visits each node of the network with a non-zero frequency, which converges to a stationary distribution \( \{ \pi_j \}_{j \in \mathcal{N}} \) satisfying the eigenvalue relation \( \pi_j = \sum_{i \in \mathcal{N}} \pi_i p_{ij} \). Transition probabilities for a finite number \( n \) of jumps, \( \{ p_{ij}^n \}_{j \in \mathcal{N}} \), are obtained from the \( n \)-th power of the single jump transition probability matrix. Thus the expected proportion of times that a chain starting from node \( i \) visits node \( j \) in the first \( n \) jumps is \( q_{ij}^n = n^{-1} \sum_{m=1}^{n} p_{ij}^m \). Because of ergodicity, in the long-time limit the latter quantity converges to the stationary frequency of visiting node \( j \):

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} p_{ij}^m \equiv \pi_j, \quad (1)
\]

which is independent on the initial node \( i \). Here \( \{ \pi_j \}_{j \in \mathcal{N}} \) is the set of infinite time transition probabilities. Finally, the stationary probability flux from node \( i \) to node \( j \) is the probability that the chain actually jumps from \( i \) to \( j \), and thus is given by the asymptotic joint probability of being in \( i \) and successively jump to \( j \):

\[
F_p(i \to j) = \pi_i p_{ij}. \quad (2)
\]

Lumped Markov chain on network partitions

A partition of the network nodes into a set of communities \( \{ C \} \) induces an aggregated dynamical process, described by transition probabilities \( \{ \tilde{p}_{CC'} \} \) between communities, that is a function of the original Markov chain. Such aggregated process is not necessarily Markovian, since the new transition probabilities could in principle depend on the whole sequence of visited nodes. However, if the original Markov chain is ergodic, it is possible to define a lumped Markov process by preserving the probability fluxes between communities:

\[
F_{\tilde{p}}(C \to C') = \sum_{i \in C} \sum_{j \in C'} F_p(i \to j). \quad (3)
\]

This property is called weak lumpability, and translates into a transition probability from \( C \) to \( C' \) of the form

\[
\tilde{p}_{CC'} = \sum_{i \in C} \sum_{j \in C'} \frac{\pi_i p_{ij}}{\pi_C}, \quad (4)
\]
where $\tilde{\pi}_C = \sum_{i \in C} \pi_i$. The generic diagonal term of this matrix, $\tilde{\pi}_{CC}$, is the persistence probability of community $C$ \cite{37,39}. Analogously, we can build the finite and infinite time transition probabilities of the lumped process as $\tilde{q}_{CC}^n = \tilde{\pi}_C^{-1} \sum_{i \in C} \sum_{j \in C'} \pi_i q_{ij}^n$ and $\tilde{\omega}_{CC'} = \tilde{\pi}_C^{-1} \sum_{i \in C} \sum_{j \in C'} \pi_i \omega_{ij}$ \cite{40}.

### The natural Markov chain

The natural Markov chain gives the simplest instance of transition probabilities between nodes in a network: the probability $p_{ij}$ to jump from node $i$ to one of its neighbors $j$ is uniform across the neighbors, and zero for unconnected nodes. If the network is weighted, $p_{ij}$ simply becomes proportional to $A_{ij}$. Hence in general $p_{ij} = A_{ij}/d_i$ where $d_i = \sum_{j \in N} A_{ij}$ denotes the total weight of outgoing connections for node $i$.

In the case of undirected networks ($A_{ij} = A_{ji}$ for each $i,j$) with a single connected component, the natural Markov chain is ergodic and the stationary distribution has the analytic form $\pi_i = d_i/(2L)$, where $2L = \sum_{ij \in E} d_{ij}$. Besides, the chain is reversible since it satisfies the detailed balance: the probability fluxes between any two nodes are equal, $F_p(i \rightarrow j) \equiv \pi_i p_{ij} = \pi_j p_{ji} \equiv F_p(j \rightarrow i)$. This relation holds simply because fluxes are proportional to the elements of the adjacency matrix, $F_p(i \rightarrow j) \sim A_{ij}$.

Due to this property, the lumped process of a natural Markov chain on a network partition can be mapped to a new weighted adjacency matrix, whose terms are given by the sum of elements of the original adjacency matrix corresponding to nodes in the considered partitions:

$$F_p(C \rightarrow C') \sim A_{CC'} = \sum_{i \in C} \sum_{j \in C'} A_{ij}. \quad (5)$$

### III. Community Detection with lumped Markov chains

We now use the general dynamical framework of lumped Markov chain introduced above to define a quality function for community detection tasks. We start from two key assumptions on which we base our definition of communities.

Firstly, as stated above, different Markov dynamics induce different partitions of the network. According to the principle behind the Markov stability, communities are regions of the network where the Markov chain remains confined for relatively long time – where “relatively long” has to be assessed using a reference process. The most natural choice is to use a reference that brings zero information both on the details of the network topology and on the initial state of the dynamics. The infinite time transition probabilities of eq. (3) satisfy this requirement, but this is just a possible choice – we can as well use as reference the Markov dynamics at any finite time.

Secondly, we require that any community has to be resilient to changes occurring locally elsewhere in the network, or equivalently that a community is a community almost independently on the topological details of the rest of the network. This assumption allows simplifying the assessment of an individual community using a lumped Markov chain with two states: the community itself and the rest of the network. Notably such a two-states Markov process can be described using only the stationary distribution $\tilde{\pi}_C$ and persistence probability $\tilde{\pi}_{CC}$ of the community concerned, since the conservation of probability fluxes implies that the flux from $C$ to anywhere else is equal to the flux from anywhere else to $C$ \cite{11}.

### Generalized Markov stability (GMS)

To find a good network partition, we aim at “maximizing the difference” between the Markov dynamics and the reference process. We can thus build a quality function based on the probability flux difference between these two processes. For simplicity we start by considering the single jump Markov dynamics, using as reference process its asymptotic behavior given by the infinite time transition probabilities. For any two nodes $i,j$ in the original Markov chain we define

$$D_{ij} = F_p(i \rightarrow j) - F_{\omega}(i \rightarrow j) = \pi_i (p_{ij} - \omega_{ij}), \quad (6)$$

while for two communities $C,C'$ in the lumped chain

$$D_{CC'} = F_p(C \rightarrow C') - F_{\omega}(C \rightarrow C') = \tilde{\pi}_C (\tilde{\pi}_{CC'} - \tilde{\omega}_{CC'}) \quad \text{ (7)}$$

meaning that $D_{CC'} \equiv \sum_{i \in C} \sum_{j \in C'} D_{ij}$. From this definition we see that the flux difference internal to community $C$, $D_{CC}$, satisfies the requirement of depending only on quantities related to $C$ itself – with respect to the rest of the network. As global quality function to assess the quality of a network partition we can thus take the trace

$$\mathcal{M}^{[1,\infty]}(\{C\}) = \sum_c \tilde{\pi}_C (\tilde{\pi}_{CC} - \tilde{\omega}_{CC}), \quad (8)$$

representing the probability flux that a random walker remains in a community within one time step, discounting the stationary distribution of the process. More generally, if we consider transition probabilities of $n$ jumps against visiting frequencies within $m$ jumps (with $n < m$) we can define

$$\mathcal{M}^{[n,m]}(\{C\}) = \sum_c \tilde{\pi}_C (\tilde{\pi}_{CC}^n - \tilde{\omega}_{CC}^m). \quad (9)$$

This quality function is a generalized Markov stability (GMS). Indeed for $m \rightarrow \infty$ the reference process is given by the infinite time transition probability as in eq. (8), which converges to the stationary distribution of the dynamics, and in this case $\mathcal{M}^{[n,\infty]}(\{C\})$ coincides with the traditional definition of Markov stability \cite{25,29}. Also,
for a natural Markov chain on an undirected network and 

\[ n = 1, \mathcal{M}^{[1,\infty]} \{ \{ C \} \} \]  

This equivalence holds because the modularity relies on a null network model (known as the Chung-Lu configuration model) that constrains node degrees [9], and the expectations of link probabilities under this null model coincide with the infinite time transition probabilities of the natural Markov chain.

The Markov stability and its generalized version have conceptual and practical advantages with respect to modularity. First of all, the modularity is based on simple link counts, as well as on a null model for the network topology. Typically, null model implementations are limited to the simple Erdős-Rényi random graph, the configuration model and the (possibly degree-corrected) stochastic blockmodel [16] – the few cases that have an analytic formulation. The Markov stability is instead based on a generic Markov process on the network: besides the natural Markov chain one is free to consider other dynamics, e.g., PageRank [42] or maximal entropic random walks [33], as well as higher-order Markov models [14]. Moreover, GMS is automatically defined in the case of directed networks, at stake with modularity [20, 39].

The peculiar advantage of our generalization of Markov stability is instead the possibility of choosing the reference process, in particular by setting its time horizon. This last aspect in particular relates to the resolution limit of the modularity. According to [5], “the resolution limit comes from the very definition of modularity, in particular from its random model. The weak point of the random model is the implicit assumption that each vertex can interact with every other vertex, which implies that each part of the network knows about everything else [...] It is certainly more reasonable to assume that each vertex has a limited horizon within the network”. In terms of Markov stability, this is implemented as in eq. (9) by using a finite time horizon for the reference process. In this way, the dynamics is compared not to its stationary distribution (which is achieved after the walker has explored the whole network), but to the finite-time frequency of visiting nodes (i.e., what the walker is able to explore in a finite number of steps). Thus, it is possible to find smaller communities than with modularity. Note also that previous attempts [25, 26] to overcome the resolution limit with standard Markov stability are based on a continuous-time process in the small time limit, rather than on a different reference process as in eq. (9).

Finally, the definition of the quality function using lumped Markov chains is invariant under hierarchical partitioning of the network. We use this feature when implementing the numerical search of communities (see pseudocode 1) using a variant of the Louvain algorithm [45] and a coarse-graining procedure.

**Algorithm 1 Louvain-based algorithm**

```plaintext
procedure GMS MAXIMIZATION
    input:
    \{p\} ← transition probabilities between nodes
    \{C\}: initial partition of single-node communities
    list: community membership of each node
    repeat
        \{C\} ← Moves(\{C\}, \{p\})
        if \mathcal{M}(\{C\}, \{p\}) > \mathcal{M}(\{\tilde{C}\}, \{p\}) then
            update list according to \{\tilde{C}\}
            \{p\} ← lumped transition probabilities – eq. (4)
            \{C\} ← coarse-grained \{\tilde{C}\} (one node per community)
        end if
    until \mathcal{M} reaches a maximum
    output list
    final step:
    \{p\} ← transition probabilities between nodes
    \{C\}: partition corresponding to list
    \{C\} ← Moves(\{C\}, \{p\})
end procedure

function Moves(\{C\}, \{p\})
(repeat a few times)
for all communities \(C \in \{C\}\) do
    for all nodes \(i \in C\) do
        find \(C' \neq C\) such that
        moving \(i\) from \(C\) to \(C'\) maximally increases \(M\)
        if such \(C'\) exists then move \(i\) from \(C\) to \(C'\)
    end if
end for
end for
end function
```

**Numerical optimization**

We first work at the finest level of nodes. We start with a configuration where each node is considered as a different community, giving the corresponding initial value for \(\mathcal{M}^{[n,m]}\). The moves we consider are successive changes of community for individual nodes. Each move is accepted if the induced change to \(\mathcal{M}^{[n,m]}\) is positive (such variation is computed locally because we consider only moves of single nodes and not of node groups). These moves are repeated until no further increase of \(\mathcal{M}^{[n,m]}\) can be achieved.

The communities found through this first procedure are then taken as the meta-nodes of a coarse-grained network, while the Markov process for this new network is defined using the lumpability condition of eq. (4). The local moves described above are then repeated again for this network until a new maximum of \(\mathcal{M}^{[n,m]}\) is reached. The corresponding partition is then used to build a more coarse-grained network, and the whole process is repeated until no further moves nor coarse-graining steps can increase \(\mathcal{M}^{[n,m]}\).

As a final step, we restart the method from the node level but imposing the community structure just found – that is, we check whether the move of a single node can refine the optimal partition. For instance in the case of the Karate Club network [46] (see below), this final step
causes a single node to switch community and the GMS value to rise from 0.4188 to 0.4198.

IV. RESULTS

Resolution of generalized Markov stability

We first explore the resolution of $\mathcal{M}^{[n,m]}$ with respect to different choices of Markov times $n$ and $m$. To this end, we consider a natural Markov chain process on the toy undirected network used in [11] – a ring-like configuration with 30 cliques of 5 nodes, each clique being connected to only two other cliques (Figure 1). For this graph, standard modularity optimization returns a maximum value of 0.8879 and a structure of 15 communities (each composed by a pair of cliques). Standard Markov stability $\mathcal{M}^{[n,\infty]}$ for $n > 1$ instead finds a community structure that is coarser than what is found by modularity [25]. In particular, since communities are defined as regions where the walker remains confined within $n$ jumps, by increasing the time horizon $n$ of the dynamics communities become less in number and bigger in size – as shown in the upper panel of Figure 1. Notably, if we use a reference process at finite time, we automatically obtain finer communities than with modularity. The lower panel of Figure 1 shows the case of the simple function $\mathcal{M}^{[1,m]}$: there is a sharp transition for the number of communities at a critical value $m^*$, below which the true structure of 30 communities (one for each clique) emerges.

The cliques-graphs considered above is a very simple example, especially because the cliques have the same size. We thus consider an heterogeneous graph of 30 cliques whose size is exponentially distributed. As shown in Figure 2 standard modularity maximization on a realization of this graph returns 22 communities, since it tends to group together the small nearest cliques (while being able to identify the large cliques). Instead $\mathcal{M}^{[1,m]}$ with $m = 3$ finds the true community structure.

Generalized Markov stability of different random processes

$\mathcal{M}^{[n,m]}$ of eq. (8) is defined for a generic Markov process on the network – the only requirement being the existence of the stationary distribution and its finite-time version. The induced community structure can thus strongly depend on which process is implemented. In the section above we used the natural Markov chain. Here we considered two other processes.

The first one is PageRank [42] (see also [39, 47]), which complements the natural Markov chain with a teleportation term allowing for jumps between any two nodes: $p_{ij} = (1 - \mu)A_{ij}/d_i + \mu/N$. Notably, the teleportation probability $\mu$ makes the chain ergodic even if the network has disconnected components or if it is directed and has transient parts (that the walker cannot access after leaving them). Figure 3 shows the number of communities found by $\mathcal{M}^{[n,\infty]}$ with PageRank for two undirected social networks: Dolphins [48] and Karate club [46]. Dolphins is the network of “swimming together” relations among a group of dolphins, whereas, Karate Club describes the social interactions between the members of a karate club before the its splits into two separated groups. Both these networks are representative small graphs with communities, often used to benchmark community detection methods. We see that, for both networks, the number of communities decreases with the teleportation rate $\mu$ due to a larger probability to jump outside communities.
FIG. 2. Visual representation of communities found by standard modularity $M_{[1,\infty]}$ (a) and generalized Markov stability $M_{[1,3]}$ (b), on a ring-like configuration with 30 cliques of varying (exponentially distributed) size, each clique being connected to only two other cliques. Each community is represented by a different color.

Hence teleportation has a similar effect of considering a longer time horizon for the natural Markov chain – see the upper panel of of Figure 1.

The second Markov process we consider is the maximal entropy random walk (MERW) \cite{43,49}, also known as the Ruelle-Bowen process in discrete time \cite{26}. MERW transition probabilities are such that all trajectories of given length and given endpoints are equiprobable, and take the form $p_{ij} = (A_{ij}/\lambda)/\psi_j/\psi_i$ – where $\lambda$ is the largest eigenvalue of the adjacency matrix and $\psi_i$ is the $i$-th component of the normalized eigenvector associated to $\lambda$. MERW has strong localization property, imprisoning the walkers in entropic wells \cite{43}. When applied to the networks of Figure 3, $M_{[1,\infty]}$ with MERW dynamics finds 3 communities in the Dolphins network and 2 in the Karate club network.

To visually grasp the effect of using a particular Markov dynamics, as an illustrative example we show in Figure 4 the communities detected by $M_{[1,\infty]}$ on the Dolphins network using the natural Markov chain, PageRank and MERW. We see that indeed the community structure varies using a different Markov processes, yet notably...
some communities seem more resilient to the different
dynamics (in this case, the top left part of the network).

Directed networks

As a final remark, we stress that the definition of gen-
eralized Markov stability does not depend on the spe-
cific network features. Therefore, GMS can be directly
implemented on directed networks, provided the consid-
ered Markov chain is ergodic (an easy solution for this
is the teleportation term of PageRank). Indeed, the case
\( n = 1 \) and \( m = \infty \) for simple random walks on directed
networks has been studied in [39] as a generalization of
standard modularity.

CONCLUSIONS

In this work we reformulated the use of ergodic Markov
chains applied to the problem of community detection in
networks. Specifically, we defined a lumped Markov pro-
cess between communities, whose transition probability
fluxes are built by aggregating the probability fluxes at
the level of nodes. This aggregated process is then used
to define a quality function to evaluate a network partition,
by requiring the probability fluxes internal to com-

communities to be maximally larger than those of a reference
case. This results in a generalized version of the Markov
stability. When the reference process corresponds to a
dynamics in which all the information on initial condi-
tions and nodes correlations is lost, as in the case of the
infinite time transition probability, we obtain the stan-
dard Markov stability. However, considering a reference
process with a finite time horizon allows finding commu-
nities of varying size – thus overcoming in a natural way
the resolution limit typical of the modularity and other
approaches.

The framework we developed in this work is general
and can possibly be applied to other kinds of networks
(e.g., bipartite graphs) or to detect overlapping communities.
Another interesting research direction would be to
compare, within the quality function, Markov processes
of different nature.

[1] R. Albert and A.-L. Barabási, Reviews of Modern Physics 74, 47 (2002)
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and
D.-U. Hwang, Physics Reports 424, 175 (2006).
[3] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes,
Reviews of Modern Physics 80, 1275 (2008).
[4] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli,
A. Gabrielli, and G. Caldarelli, Nature Reviews Physics 1, 58 (2019).
[5] S. Fortunato, Physics Reports 486, 75 (2010).
[6] S. Fortunato and D. Hric, Physics Reports 659, 1 (2016).
[7] M. T. Schaub, J.-C. Delvenne, M. Rosvall, and R. Lambiotte,
Applying Network Science 2, 4 (2017).
[8] M. Girvan and M. E. J. Newman, Proceedings of the National Academy of Sciences 99, 7821 (2002).
[9] M. E. J. Newman and M. Girvan, Physical Review E 69, 026113 (2003).
[10] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato,
PLoS ONE 6, 1 (2011).
[11] S. Fortunato and M. Barthélemy, Proceedings of the National Academy of Sciences 104, 36 (2007).
[12] A. Arenas, A. Fernández, and S. Gómez, New Journal of Physics 10, 053039 (2008).
[13] J. Reichardt and S. Bornholdt, Physical Review E 74, 016110 (2006).
[14] M. Chen, T. Nguyen, and B. K. Szymanski, "A new metric for quality of network community structure,”
http://arxiv.org/abs/1507.04308 (2015).
[15] B. Karrer and M. E. J. Newman, Physical Review E 83, 016107 (2011).
[16] M. E. J. Newman, Physical Review E 94, 052315 (2016).
[17] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, Nature 435, 814 (2005).
[18] M. E. J. Newman, Physical Review E 74, 036104 (2006).
[19] V. A. Traag, P. Van Dooren, and Y. Nesterov, Physical Review E 84, 016114 (2011).
[20] V. A. Traag, G. Krings, and P. Van Dooren, Scientific Reports 3, 2930 (2013).
[21] R. Aldecoa and I. Marin, PLoS ONE 6, 1 (2011).
[22] R. Aldecoa and I. Marin, Scientific Reports 3, 1060 (2013).
[23] V. A. Traag, R. Aldecoa, and J.-C. Delvenne, Physical Review E 92, 022816 (2015).
[24] N. Masuda, M. A. Porter, and R. Lambiotte, Physics Reports 716-717, 1 (2017).
[25] J. C. Delvenne, S. N. Yaliraki, and M. Barahona, Proceedings of the National Academy of Sciences 107, 12755 (2010).
[26] R. Lambiotte, J. Delvenne, and M. Barahona, IEEE Transactions on Network Science and Engineering 1, 76 (2014).
[27] M. Kheirkhahzadeh, A. Lancichinetti, and M. Rosvall, Physical Review E 93, 032309 (2016).
[28] P. Pons and M. Latapy, Journal of Graph Algorithms and Applications 10, 191 (2006).
[29] M. Rosvall and C. T. Bergstrom, Proceedings of the National Academy of Sciences 105, 1118 (2008).
[30] M. Rosvall, D. Axelsson, and C. T. Bergstrom, The European Physical Journal Special Topics 178, 13 (2009).
[31] J. I. Perotti, C. J. Tessone, A. Clauset, and G. Caldarelli, "Thermodynamics of the Minimum Description Length on community detection,”
https://arxiv.org/abs/1806.07005 (2018).
[32] T. S. Evans and R. Lambiotte, Physical Review E 80, 016105 (2009).
[33] A. Gabrielli, R. Mastrandrea, G. Caldarelli, and G. Cimini, Physical Review E 99, 030301 (2019).
[34] R. Aldecoa and I. Marin, Scientific Reports 3, 2216.
[35] L. Peel, D. B. Larremore, and A. Clauset, Science Advances 3, e1602548 (2017).
[36] V. Zlatic, A. Gabrielli, and G. Caldarelli, Physical Review E 82, 066109 (2010).
[37] C. Picardi, PLoS ONE 6, e27028 (2011).

We limit our analysis to time-homogeneous Markov chains, for which the transition probabilities do not depend on the current time step.

[38] Y. Kim, S. W. Son, and H. Jeong, Physical Review E 81, 016103 (2010).

This is generally different from taking the limit \( n \to \infty \) of the \( n \) jump transition probabilities for the lumped process.

[39] V. Zlatic, A. Gabrielli, and G. Caldarelli, Physical Review E 82, 066109 (2010).

An ergodic Markov chain with two possible states (labeled 1 and 2) can be described using only two parameters. Here we take \( \pi_1 \), the stationary distribution of state 1, and \( p_{12} \), the transition probability from state 1 to 2. Firstly, because of normalization we have \( \pi_2 = 1 - \pi_1 \) and \( p_{11} = 1 - p_{12} \). Then, since the chain is reversible we have \( F(1 \to 2) = F(2 \to 1) \) and thus \( p_{21} = p_{12} \pi_1/(1 - \pi_1) \) and \( p_{22} = 1 - p_{21} \).

[40] L. Page, S. Brin, R. Motwani, and T. Winograd, World Wide Web Internet and Web Information Systems 54, 1 (1998).
[41] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, Physical Review Letters 102, 160602 (2009).

This is generally different from taking the limit \( n \to \infty \) of the \( n \) jump transition probabilities for the lumped process.

[42] V. Salnikov, M. T. Schaub, and R. Lambiotte, Scientific Reports 6, 23194 (2016).
[43] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Journal of Statistical Mechanics: Theory and Experiment 10, P10008 (2008).

An ergodic Markov chain with two possible states (labeled 1 and 2) can be described using only two parameters. Here we take \( \pi_1 \), the stationary distribution of state 1, and \( p_{12} \), the transition probability from state 1 to 2. Firstly, because of normalization we have \( \pi_2 = 1 - \pi_1 \) and \( p_{11} = 1 - p_{12} \). Then, since the chain is reversible we have \( F(1 \to 2) = F(2 \to 1) \) and thus \( p_{21} = p_{12} \pi_1/(1 - \pi_1) \) and \( p_{22} = 1 - p_{21} \).

[44] W. Zachary, Journal of Anthropological Research 33, 452 (1977).
[45] R. Lambiotte and M. Rosvall, Physical Review E 85, 056107 (2012).

An ergodic Markov chain with two possible states (labeled 1 and 2) can be described using only two parameters. Here we take \( \pi_1 \), the stationary distribution of state 1, and \( p_{12} \), the transition probability from state 1 to 2. Firstly, because of normalization we have \( \pi_2 = 1 - \pi_1 \) and \( p_{11} = 1 - p_{12} \). Then, since the chain is reversible we have \( F(1 \to 2) = F(2 \to 1) \) and thus \( p_{21} = p_{12} \pi_1/(1 - \pi_1) \) and \( p_{22} = 1 - p_{21} \).

[46] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behavioral Ecology and Sociobiology 54, 396 (2003).
[47] J. Ochab and Z. Burda, The European Physical Journal Special Topics 216, 73 (2013).