Stability of graph communities across time scales

J.-C. Delvenne[a], S. N. Yaliraki[a,b], and M. Barahona[a,c,1]

[a]Institute for Mathematical Sciences, [b]Department of Chemistry, and [c]Department of Bioengineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

Edited by Mark Newman, University of Michigan, Ann Arbor, MI, and accepted by the Editorial Board May 21, 2010 (received for review March 23, 2009)

The complexity of biological, social, and engineering networks makes it desirable to find natural partitions into clusters (or communities) that can provide insight into the structure of the overall system and even act as simplified functional descriptions. Although methods for community detection abound, there is a lack of consensus on how to quantify and rank the quality of partitions. We introduce here the stability of a partition, a measure of its quality as a community structure based on the clustered autocovariance of a dynamic Markov process taking place on the network. Because the stability has an intrinsic dependence on time scales of the graph, it allows us to compare and rank partitions at each time and also to establish the time spans over which partitions are optimal. Hence the Markov time acts effectively as an intrinsic resolution parameter that establishes a hierarchy of increasingly coarser communities. Our dynamical definition provides a unifying framework for several standard partitioning measures: modularity and normalized cut size can be interpreted as one-step time measures, whereas Fiedler's spectral clustering emerges at long times. We apply our method to characterize the relevance of partitions over time for constructive and real networks, including hierarchical graphs and social networks, and use it to obtain reduced descriptions for atomic-level protein structures over different time scales.

community structure | Markov chain | modularity | multiscale modelling | networks

In recent years, there has been a surge of interest in the analysis of networks as models of complex systems. The literature is extensive, spanning areas as diverse as gene regulation, protein interactions and metabolic pathways, neural science, social networks or engineering systems such as transportation networks and the internet, to name but a few (1, 2). The tools for network analysis are firmly grounded on results in graph theory, with an influx of concepts from statistical physics, dynamical systems, and stochastic processes (3). Due to the large-scale, complex nature of many systems under study, an appealing idea is to obtain relevant partitions of the network (also called clusterings or communities) that can reveal the underlying structure of the system and hence provide insight into its function. These partitions could potentially lead to reduced, more manageable representations of the original system (4, 5).

The topic of community detection in graphs has a long history and multiple methods and heuristics have been proposed to partition graphs into communities or clusters. (See for instance ref. 6 and references therein for a recent survey.) However, the extensive list of partitioning methods comes with a parallel lack of theory or consensus on measures to quantify the goodness of a community structure. For instance, consider the simplest such measure: the *cut size*, i.e., the sum of the weights of edges that lie at the boundaries of different communities. As a general rule, good community structures should have small cut size implying that the communities are weakly interconnected. Unfortunately, this intuitive notion has negligible applicability since the partition with minimum cut size is often trivial. Therefore, a variety of generalized, compound measures have been proposed to induce more balanced partitions. These include, without claim of exhaustivity, normalized cut (7), (α,ϵ)-clustering (8), modularity (9, 10) and variants and extensions of modularity (11, 12). These methods, which are based on different heuristics, have distinct features and have been shown to produce reasonable community structures for a variety of examples. In particular, modularity does not require that the number of communities be specified in advance, unlike most of the other partitioning methods. However, it has been recently shown that optimizing modularity can overpartition or underpartition the network, failing to find the most natural community structure (13). To compensate for this, recent methods (12, 14, 15), have included an ad hoc resolution parameter that can be tuned to bias towards small or large communities. The introduction of such resolution parameters highlights the fact that one would expect that any given graph could be described by different natural community structures (finer or coarser) in different regimes.

Here we introduce a quality measure that has the intrinsic flexibility to find which clusterings are relevant at different time scales. This goal is achieved by establishing a link between the quality of the partition and a stochastic process taking place on the clustered graph. To establish this connection, we use the well known relationship between graphs and Markov chains: any graph has an associated random walk in which the probability of leaving a vertex is distributed among the outgoing edges according to their weight. Conversely, a Markov chain is represented by a graph with edges weighted by probabilities. This Markov viewpoint provides a dynamical interpretation of community detection: natural communities at a given time scale correspond to persistent dynamical basins, that is, to sets of states from which escape is unlikely within the given time scale. This correspondence can be established quantitatively through the autocovariance of the clustered Markov process, a measure that defines the persistence of a cluster in time. In essence, one can think of the time scale as an intrinsic *resolution parameter* for the clustering: over short time scales, many small clusters should be coherent; on the other hand, the expectation is that, as time evolves, there will be fewer, larger clusters that are persistent under the dynamics of the Markov chain.

A satisfying feature of our dynamical approach is that it provides a framework that unifies seemingly disparate clustering heuristics in the literature, which turn out to have a natural Markov probabilistic interpretation. In particular, we show below that modularity and normalized cut are related to the clustered autocovariance on paths of length one (i.e., at time $t = 1$), while Fiedler's spectral method is favored in the asymptotic limit of long paths (i.e, as time $t \to \infty$). Moreover, we also show that recent methods that include ad hoc resolution parameters (12, 14) are obtained as a linearization of the continuous Markov process at small times (i.e., as time $t \to 0$). In contrast, our measure incorporates paths of *all lengths* and provides an evaluation of

Author contributions: J.-C.D., S.N.Y., and M.B. designed research; J.-C.D. and M.B. performed research; J.-C.D. and M.B. analyzed data; and J.-C.D., S.N.Y., and M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.N. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: m.barahona@imperial.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0903215107/-/DCSupplemental.

the quality of a clustering at all times. The resulting sequence of partitions with maximum stability as a function of time includes typically clusterings that are both coarser (for $t > 1$) and finer (for $0 < t < 1$) than those obtained by modularity optimization. We now introduce the definition of stability of a partition, including the connections with other clustering heuristics, and exemplify the applications of the method with networks drawn from different fields to showcase the generality of the approach.

Methods

Stability of Graph Communities Based on Their Markov Autocovariance. Consider an undirected, connected graph with N vertices (or nodes) and m edges and assume that the graph is nonbipartite. For simplicity, we will assume that the graph is unweighted, although all our results apply equally to weighted graphs. The topology of the graph is described by the $N \times N$ adjacency matrix A, a symmetric 0–1 matrix with a "1" if two vertices are connected and a "0" otherwise. The number of edges of each vertex or degree, d_i, can be compiled into the vector $\mathbf{d} = A\mathbf{1}$, where $\mathbf{1}$ is the $N \times 1$ vector of ones. We will also use the diagonal matrix of degrees: $D = \mathrm{diag}(\mathbf{d})$.

A random walk on any such graph defines an associated Markov chain in which the probability of leaving a vertex is split uniformly among the outgoing edges with a transition probability $1/d_i$ for each edge:

$$\mathbf{p}_{t+1} = \mathbf{p}_t\, D^{-1}A \equiv \mathbf{p}_t M, \qquad [1]$$

where $\mathbf{p}_t$ is the $1 \times N$ (normalized) probability vector and M is the transition matrix. Under these assumptions, we have an ergodic and reversible Markov chain with stationary distribution $\pi = \mathbf{d}^T/\sum_i d_i = \mathbf{d}^T/2m$, given by $\pi = \pi M$. We will also use the corresponding diagonal matrix $\Pi = \mathrm{diag}(\pi)$.

Consider now a partition of the graph into c nonoverlapping communities. This (hard) clustering can be encoded in an $N \times c$ indicator matrix H, a 0–1 matrix that records which vertex belongs to which community. Each row of H is all zeros except for a one indicating the cluster to which the vertex belongs. Let us now observe the Markov process [**1**] under the prism of a given partition. If we assign a different real value α_i to the vertices of each of the c clusters, the observed signal is then a stationary, not necessarily Markovian, random variable $(X_t)_{t\in N}$ which consists of a sequence of α_i. If the partition of the graph involves good communities *over a given time scale*, we expect that the state is more likely to remain within the starting cluster for such a time span, as compared with that event occurring at random. This phenomenon can be quantified through the autocovariance of the observable $\mathrm{cov}[X_t, X_{t+\tau}] = E[X_t X_{t+\tau}] - E[X_t]^2$, where E denotes expectation. If the intercommunity connections are weak, the values of X_t and $X_{t+\tau}$ will be correlated for longer times. Indeed, how fast the autocovariance decays as a function of the lag τ is therefore an indicator of the quality of the clustering over the corresponding Markov time scale. The connection between autocovariance and clustering is the main idea underpinning our measure.

The covariance of X_t can be rewritten as $\mathrm{cov}[X_t, X_{t+\tau}] = \alpha^T R_\tau \alpha$, where α is the vector of labels of the c communities and the matrix R_t is the *clustered autocovariance matrix* of the graph:

$$R_t = H^T(\Pi M^t - \pi^T\pi)H, \qquad [2]$$

a matrix that depends only on the topology of the graph and on the given clustering. R_t describes the t-step dependence of the transfer probabilities between clusters: each element $(R_t)_{ij}$ corresponds to the probability of starting in a cluster i and being in another cluster j after t steps minus the probability that two independent random walkers are in i and j, evaluated at stationarity.

As stated above, a good partition over a given time scale should imply a high likelihood of remaining within the starting community. In terms of the clustered autocovariance matrix, the diagonal elements $(R_t)_{ii}$, which measure the probability of a random path of length t to start and end in the same community, should be larger than the off-diagonal ones. This observation leads to our definition of the *stability of the clustering*:

$$r(t;H) = \min_{0\le s\le t} \sum_{i=1}^{c} (R_s)_{ii} = \min_{0\le s\le t} \mathrm{trace}[R_s]. \qquad [3]$$

A good clustering over time t will have large stability, with a large trace of R_t over such a time span. Note that our definition involves the minimum value of the trace up to a time t, i.e., the stability is large only if its value is large *for all times* up to t. In this way, we assign low stability to partitions where there is a high probability of leaving the community and coming back to it later, as in the case of almost bipartite graphs.

The stability [**3**] is the fundamental tool we propose to assess the quality of different clusterings over time. For each candidate clustering, we can compute the stability at all times and rank the possible partitions. Clearly, certain partitions might only be optimal in particular time windows and different partitions will be optimal at different times. For each Markov time t, we seek the partition with the largest stability to obtain the *stability curve of the graph*:

$$r(t) = \max_H r(t;H). \qquad [4]$$

The stability curve establishes a time hierarchy of partitions, from finer to coarser as time grows, as exemplified in Fig. 1 for a social network. This curve encapsulates the idea that partitions are better or worse depending on the time of interest and it makes explicit the concept of the Markov time as an intrinsic resolution parameter that establishes when a partition is good.

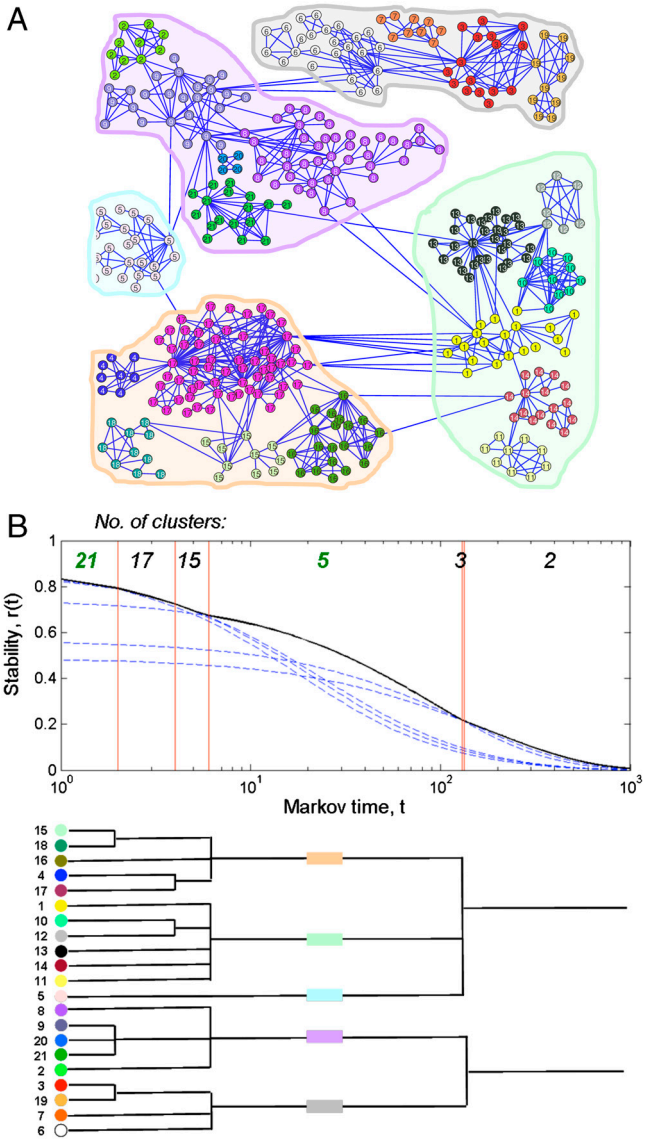

Fig. 1. (*A*) Largest connected component of a graph of scientific collaborations in network science (16). The vertices corresponding to $N = 379$ researchers are color coded according to the 21-way partition obtained by maximizing the stability [**3**] at $t = 1$ (or equivalently, modularity). The list of researcher names and groupings is available in the *SI Appendix*. (*B*) Stability curve [**4**] of this graph obtained with the divisive KVV algorithm and the corresponding dendrogram of the hierarchy of partitions. Note the simplicity of the dendrogram, which is not a binary tree, as compared with the many branching points obtained by standard binary partition methods. Only two clusterings are long-lived: the two-way split and the five-way partition represented by areas shaded in different colors in (*A*).

In this sense, one will consider that the most relevant partitions will be those that are optimal over long time windows, because they serve as good representations over extended time scales of the system. The stability curve can be estimated numerically using a variety of existing algorithms, as shown in Fig. 2.

Relationship of Stability with Diversity Index, Modularity, Cut, Normalized Cut, and Spectral Partitioning. An important feature of the stability [3] is that it encompasses several of the criteria for clustering that have been proposed in the literature and allows us to interpret those heuristics in terms of the relevant Markov time scales of the graph. To explore the unifying power of the framework, we study the autocovariance R_t and the stability $r(t)$ in different limits. An extended explanation of the following results and other generalizations is given in more detail in the *SI Appendix*.

First, consider short times. At time $t = 0$, the partition with the largest stability is the finest possible clustering, a fact that follows from elementary inequalities which show that $r(0) = 1 - \|\pi H\|_2^2$ becomes maximal when each vertex is in its own cluster. Interestingly, the quantity $r(0)$ is the so-called Simpson's diversity index, often used by biologists to measure how equally a population is divided among different species. The quantity $r(0)$ is also equivalent to other diversity measures such as the Hirschman-Herfindahl index, used to quantify monopolies in economics, or the Rényi entropy of order 2, used in information theory.

At time $t = 1$, we recover modularity, a popular measure for community detection (9): modularity is equal to the trace of R_1, the autocovariance matrix at $t = 1$, as follows from [3]. Therefore, maximizing $r(1)$ is equivalent to modularity optimization. (See also ref. 16 for an alternative, nondynamical take on this issue.) The stability at $t = 1$ is also related to other measures in the literature. Consider the cut size (Cut), defined as the number of intercommunity edges divided by the total number of edges of the graph. It is easy to see that $\text{Cut} = r(0) - r(1)$, from which it follows that Modularity = Diversity Index – Cut. This equality is the reason why modularity optimization tends to produce balanced partitions: minimizing cut favors few clusters, possibly of very unequal sizes, while maximizing the Diversity Index tends to favor many clusters of equal size. An alternative measure to modularity is the so-called normalized cut size (NCut) (7). For the case of two communities, NCut is the number of intercommunity edges multiplied by the sum of the inverse of the number of edges in each community, which equals $\text{NCut} = \frac{1}{2}(\rho(0) - \rho(1))$, where $\rho(t)$ is given by the same expression as the stability $r(t)$ replacing covariances by correlations.

The discussion above shows that modularity, Cut, and NCut are based on the one-step behavior of the Markov process. However, stability provides a measure at all times. In fact, the behavior of $r(t)$ in the long time limit $t \to \infty$ establishes a link with spectral clustering methods, the other standard toolbox for graph partitioning. Spectral clustering is generally based on the eigenvectors of the Laplacian matrix $L = D - A$. Classic methods proceed by partitioning the graph recursively into subgraphs according to the sign of the components of the Fiedler eigenvector (17, 18), i.e., the eigenvector associated with the second smallest eigenvalue of L. More recently, it has been proposed (19) that graph partitioning could be based on the Fiedler vector of the *normalized* Laplacian $\mathscr{L} = D^{-1/2}LD^{-1/2}$, which has been shown to be a heuristic for the optimal NCut two-way clustering (7).

The analysis of the definition of stability shows that spectral clustering is not just a heuristic but an exact method to find the most stable partitions at long time scales. This fact follows from the spectral decomposition of the normalized Laplacian $\mathscr{L}$, which is trivially related to that of $\mathscr{M} = D^{1/2}MD^{-1/2} = \sum_{i=1}^{N} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$. Here the eigenvalues λ_i are ranked in decreasing order and the corresponding eigenvectors $\mathbf{u}_i$ are orthonormal. In particular, $\lambda_1 = 1$ and $\mathbf{u}_1 = (1/\sqrt{2m})D^{1/2}\mathbf{1}$. If, as in most networks of interest, λ_2 is nondegenerate and dominates all eigenvalues except λ_1, then we have the asymptotic behavior

$$\text{trace}[R_t] = \sum_{i=2}^{N} \frac{\lambda_i^t}{2m} \|H^T D^{1/2}\mathbf{u}_i\|^2 \overset{t\to\infty}{\longrightarrow} \frac{\lambda_2^t}{2m} \|H^T D^{1/2}\mathbf{u}_2\|^2, \qquad [5]$$

which is dominated by $\mathbf{u}_2$, the normalized Fiedler eigenvector. In this case, the clustering with maximal stability at long times corresponds to the two-way partition according to the signs of the components of $\mathbf{u}_2$. The optimality of this partition follows from the fact that clustering vertices i and j together induce a variation in [5] given by $(\lambda_2^t/m)\sqrt{d_i d_j}u_{2,i}u_{2,j}$, which is only positive if the components of $\mathbf{u}_2$ for nodes i and j have the same sign.

This asymptotic convergence is typical but there are some nongeneric graphs for which the asymptotic limit is different. In those cases, stability also provides a consistent interpretation. When λ_2 is degenerate, the asymptotics are dominated by the subspace spanned by the degenerate eigenvectors. This degeneracy occurs in graphs with natural partitions into $k > 2$ groups for which the stability will indicate that the k-partition is optimal (Fig. 3). When $|\lambda_N| > \lambda_2$, as is the case for (almost) bipartite graphs, the stability at large times becomes negative for all clusterings except for the partition with all nodes in one community, for which the stability is trivially zero at all times. In this case, the dominant partition at long times is the one-way partition, the expected outcome for bipartite graphs, which do not have a natural two-way cut.

The overall picture that emerges from our analysis is that the sequence of partitions with highest stability evolves generically from the finest possible (each vertex by itself) at $t = 0$, through the partition with optimal modularity at $t = 1$, onto a sequence of generally coarser partitions which contain fewer and fewer clusters as the Markov time grows, typically towards the two-way spectral clustering as $t \to \infty$. It is important to remark that this asymptotic convergence does not imply that the normalized Fiedler two-way partition will be identified as a relevant community, since it can become asymptotically dominant at values of stability that are negligibly small. An example of this situation is shown in Fig. 3, which has an intrinsic fivefold symmetry in its community structure. We have also checked that stability does not impose the emergence of an artificial community structure when there is none, as in the case of random graphs (see *SI Appendix*). It is also easy to see within this framework that the modularity-optimal clustering might be too fine or too coarse for particular examples, since it might correspond to transient partitions (as in Fig. 3) or to time scales that are not relevant for the system.

Applications and Examples

We now show the applicability of the method by analyzing three examples drawn from social interactions, hierarchical scale-free graphs, and protein structural networks. Rather than being exhaustive, our goal is to highlight through each example some of the wider features of our approach.

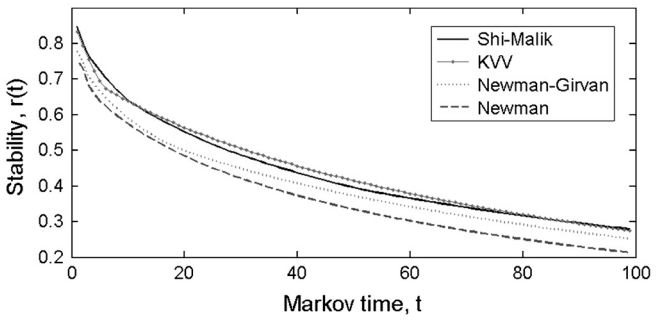

Fig. 2. A comparison of the stability curves of the network of scientific collaborations in Fig. 1*B* obtained through four divisive algorithms: Shi-Malik (7), KVV (8), Newman (without Kernighan-Lin) (16), and Newman-Girvan (9).

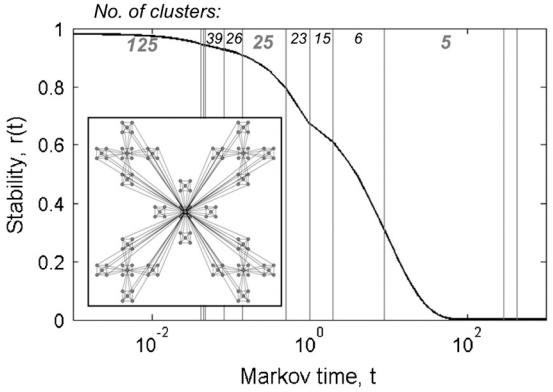

Fig. 3. Stability curve of a hierarchical, scale-free graph with $N = 125$ vertices proposed in (24) (shown in the *inset*) calculated for times smaller and larger than one. Note that the natural partitions into 25 and 5 communities persist over long time scales, while the modularity-optimal clustering (at $t = 1$) can be seen as a transient. As expected from the symmetry of the graph, the five-way partition dominates for large times and the two-way partition is not relevant.

Example 1—Time Hierarchy of Partitions, Optimization of Stability, and Comparison of Clustering Algorithms. Our first example deals with the graph of collaborations between researchers in network science shown in Fig. 1*A* (16). Community structures are relevant for social networks, where the identification of groups of people with strong ties can help unravel underlying patterns of interdependence (3). In Fig. 1*B* we show the time hierarchy of partitions associated with the stability curve of the network. Our measure [**3**] is used to rank partitions efficiently, since the stability of a given clustering $r(t;H)$ is directly computable in $\mathcal{O}(cmt)$, or estimated in $\mathcal{O}(Kt)$ with accuracy $\mathcal{O}(c/\sqrt{K})$ through K random walks of length t. In order to obtain the stability curve, one needs to maximize the stability over all partitions. Given that modularity optimization is provably NP-hard (20), it is likely that no efficient algorithm exists for the optimization of stability for arbitrary graphs. However, for all practical applications, we can still obtain sequences of partitions through the use of a number of heuristic strategies, such as aggregative (i.e., unifying clusters from the finest clustering) or divisive (i.e., splitting clusters from the coarsest clustering). Fig. 1*B* is the result of the application of Kannan, Vempala and Vetta's (KVV) conductance spectral divisive algorithm (8) to produce a sequence of partitions, which are then ranked according to their stability to estimate the stability curve $r(t)$. This curve is then translated into a nonbinary dendrogram representing the sequence of community structures with maximal stability as a function of time. The dendrogram has the advantage of being relatively simple, with fewer branching points compared with the binary trees produced by most hierarchical community detection algorithms. In this case, the time hierarchy of partitions indicates that the modularity-optimal clustering into 21 communities is short lived whereas a coarser partition into five communities persists over a long time window. This persistence suggests the relevance of this coarser meta-community structure as indicative of the likelihood of information to flow within the five subgroups of researchers. We use here the persistence of the partitions across long time scales as a crude measure of their relevance, although more sophisticated, but computationally costly, tools could be used (21, 22).

Stability can also be used to rank the sequences of partitions obtained by different algorithmic strategies. Fig. 2 presents the estimated stability curves from four algorithms chosen for their simplicity and popularity and because they represent different overall methodologies (see *SI Appendix*): the KVV method (8), Shi-Malik's recursive spectral method (7), Newman's spectral modularity optimization without the Kernighan-Lin step (16), and the Newman-Girvan betweenness algorithm (9). In all cases, we use a divisive strategy to produce a sequence of increasingly finer partitions and we obtain an estimate of the stability curve $r(t)$ by choosing the best partition at each time. In this case, Shi-Malik and KVV produce the partitions with highest stability at all shown times (alternatively better in different time windows), followed closely by the Newman-Girvan algorithm and Newman's spectral algorithm. At higher times (up to $t = 1{,}000$ at least), the KVV method slightly dominates Shi-Malik and Newman-Girvan algorithms, while Newman's clustering algorithm is worse by a factor of two. These observations are no evidence of overall superiority of one method over another, but an example of how to compare and use the different partitioning algorithms on a given network. Conversely, one can use a combination of optimization methods to find the best possible partitions that maximize stability at different time scales. Fig. 2 also shows that all algorithms produce stability curves of similar magnitude with comparable time dependence, a fact we have confirmed through the use of the Louvain algorithm, based on a distinct agglomerative heuristic (see *SI Appendix*) (23). Therefore, the NP-hard optimal stability can be estimated robustly through heuristic algorithms. This point is discussed further in the *SI Appendix*.

Example 2—Beyond the Resolution Limit of Modularity: the Small Time Limit of Stability. Recently, it has been shown that modularity optimization cannot produce partitions smaller than a certain relative size. This effect, termed the resolution limit of modularity, leads to partitions coarser then the expected "natural" community structure (13). So far, our analysis has shown that the modularity-optimal community structure corresponds to stability at time $t = 1$ while for $t > 1$, the most stable community structures are *coarser* than those found by modularity optimization. Our framework also leads naturally to *finer* communities than modularity (i.e., beyond the resolution limit) by considering stability at times between zero and one. This regime can be studied through the extension to the continuous-time version of [**2**] obtained by substituting M^t by the matrix exponential $\exp[(M - I)t]$, where I is the identity matrix (23). The linearized stability for small (continuous) times then gives:

$$r_c(t) \simeq (1 - t)r(0) + tr(1), \qquad 0 \leq t \leq 1. \qquad [6]$$

Note that this linear interpolation recovers modularity at $t = 1$ and the diversity index at time $t = 0$. Eq. **6** also provides an interpretation in terms of Markov time of the resolution parameter proposed by Reichardt and Bornholdt (12) and is related to a heuristic proposed by Arenas et al. (14) consisting of the addition of weighted self loops to the graph. These connections are discussed further in the *SI Appendix*.

As an example, Fig. 3 shows the stability curve extending to $t < 1$ of a 125-vertex hierarchical scale-free graph recently proposed by Ravász and Barabási (24). In this simple model, the natural clustering is not found through modularity. Our method, on the other hand, finds that the natural partitions into 25 and 5 clusters have long windows of stability while the partition obtained by modularity at $t = 1$ is a transient with no extended significance.

Example 3: Structural Graphs, Model Reduction, and Time Scales. Our final example shows an application of our framework to graphs derived from atomic-level protein structures and its relevance to model reduction of biophysical systems. Recently, new methods based on graphs of constraints have been proposed to simplify the complex dynamics of large biomolecules such as proteins. The idea is to obtain lower-dimensional descriptions of protein dynamics in terms of a few relatively rigid parts connected by flexible elements (4, 5, 25–28). Because coherent subunits are likely to result from a tightly-knit network of chemical bonds and chemical constraints, we expect that a reasonable approximation to the constrained flexibility of the protein will be revealed by the multiscale partitions of the structural graph of the protein, in which the atoms are vertices and edges correspond to bonds and chemical constraints (25).

Fig. 4*A* shows the stability curve and the time hierarchy of partitions of a full-atom ($N = 2{,}085$) structural graph of the protein Adenylate Kinase (AK) in its open configuration obtained using the Shi-Malik divisive algorithm. In this example, biophysical considerations indicate that optimizing modularity overpartitions the graph—the 31 communities obtained at $t = 1$ split several structural motifs such as β-sheets and α-helices. Some of the optimal partitions at longer times (notably those into 18 and 4 communities) prevail over relatively long time scales and contain significant biophysical features. To make this statement more precise, we evaluate the relative variation of the intracommunity positions of C_α carbons between two functional conformations of AK (open vs. closed) for all partitions. Fig. 4*B* shows the intracommunity stretching for all partitions obtained as follows: calculate all pair distances between atoms within each community in both configurations of the protein and obtain Δ, the average square variation of those distances over all communities. If the communities are completely rigid, the pair distances within com-

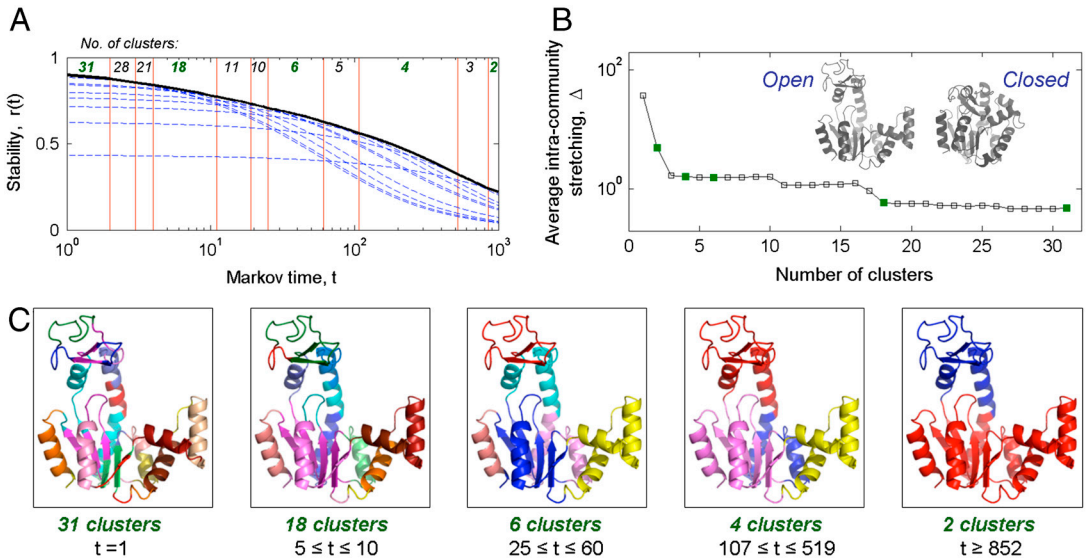

Fig. 4. Analysis of the atomic-level structural graph of the open conformation of the protein AK with $N = 2{,}085$ vertices (see the *SI Appendix* for a detailed explanation of how this graph is obtained). (*A*) The optimal stability curve at each Markov time (*solid curve*) is estimated from partitions obtained by the divisive Shi-Malik algorithm. The 31-way clustering with optimal modularity among the computed clusterings overpartitions the structure: it breaks β-sheets and α-helices, which should belong to the same cluster. The four-way and 18-way partitions have relatively long windows of stability with a good balance between over- and underpartitioning. (*B*) Evaluation of the validity of the partitions through a comparison of two experimental conformations of AK: open and closed. Partitions obtained from the graph of the open configuration are evaluated in terms of the error of the experimental distortion when comparing to the closed conformation by assuming rigidity of the predicted communities. The two plateaux in the error (from 4 to 10 and from 18 to 31 clusters) indicate that the four-way and 18-way partitions, which also show persistence over long time windows in (*A*), represent a parsimonious compromise between predicted rigidity and a small number of clusters. (*C*) Some of the partitions in the hierarchy that appear at different Markov time scales with structural communities represented by adjacent regions of the same color.

munities will not change and $\Delta = 0$. The maximum value $\Delta = 37$ Å^2 is the average square variation for all atoms in the protein (i.e, when we consider all of them in one community). As the number of communities grows, one expects that Δ will decrease, since the number of pair distances decreases. The key is to find when the addition of a community does not result in a significant decrease of Δ, thus implying that the new communities added are not significantly rigid. This situation is observed in the plateaux in Δ that follow the four-way and 18-way community structures and is consistent with the extended time scales of prevalence for both partitions in the stability curve. We deduce that the four-way and 18-way community structures are a reasonable compromise between simplicity and predictive power for rigidity. We remark for the application to protein structures that the "Markov time" is defined as an abstract entity, not to be assigned an immediate link with a physical quantity. The rigorous connection between the Markov time and the biophysical time of protein motions is the subject of current research.

Discussion and Future Work

In this work, we have introduced the stability [**3**] as a quality measure of a graph partition based on dynamical concepts. The stability of a partition, which is defined in terms of the clustered autocovariance of a Markov process taking place on the graph, is explicitly dependent on the Markov time, an intrinsic time scale of the network. We can therefore rank partitions and establish their relevance over time scales. Although Markov chains (29–31) and oscillator dynamics (27, 32, 33) have been used in relation to community detection, previous methods have not defined a quality measure, nor have they considered paths of all lengths to evaluate the quality of partitions across time scales.

Our measure can be used to obtain a sequence of partitions with maximum stability as a function of time, from finer to coarser as the Markov time grows. This time hierarchy can be used to establish the most relevant partitions over significant time scales. Hence, our method does not provide a unique, "optimal" partition for the graph. Rather, we propose that obtaining distinct partitions that are valid over different time windows and selecting those partitions that are relevant over extended time scales may be better suited for many applications. In particular, if a network is defined by an underlying dynamical process with well defined time scales, our analysis can suggest reduced representations valid over time windows of interest. On the other hand, if the network under study does not have an obvious temporal interpretation, the Markov time acts effectively as an *intrinsic* resolution parameter for graph partitions.

An important feature of the stability is that it gives a unified interpretation in terms of time scales of community detection methodologies that have been hitherto considered separately. We have shown that modularity, cut, and normalized cut can be understood in relation to the stability at $t = 1$, while spectral clustering based on the normalized Fiedler vector is linked to stability as $t \to \infty$. Additionally, the continuous version of stability (23) can go beyond the resolution limit of modularity and its linearization for small times can be linked to previously proposed ad hoc multiresolution methods (12, 14). Stability can also be connected with the concept of "anticlustering" and k-colorings (34, 35) based on the existence of recurrence patterns in the time-dependence of the trace of R_t, which become important in (almost) bipartite graphs.

Complex systems, from protein dynamics to metabolic and social interactions to the internet, are often described as networks. The methodology presented here, which extends seamlessly to both weighted and directed graphs, uses the intimate connection between structure and dynamics to identify communities that can be revealing of the network structure. In some cases, the original networks are static and our dynamical approach is a convenient construct to reveal the intrinsic resolution scales of the problem. If, on the other hand, the network has a dynamic origin, or indeed it can be related to a Markov process (4, 28), the stability of the graph provides information about the hierarchy of time scales of the underlying landscape of the system. From this dynamic viewpoint, the presence of communities relevant over particular time scales hints at a first step towards re-

duced representations in which the communities can be lumped into aggregate variables.

ACKNOWLEDGMENTS. We thank Renaud Lambiotte, João Costa, and Vincent Blondel for helpful discussions and João Costa for providing the graph and figures for the AK protein. J.-C. D. and M.B. acknowledge support from the Mathematics and Life Science Interface panels of the United Kingdom Engineering and Physical Sciences Research Council (EPSRC). J.-C. D. is supported by the Belgian Programme of Interuniversity Attraction Poles and an Action de Recherche Concertée (ARC) of the French Community of Belgium.

1. Strogatz SH (2001) Exploring complex networks. *Nature* 410:268–276.
2. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256.
3. da Costa L, Rodrigues FA, Travieso G, Villas Boas PR (2007) Characterization of complex networks: a survey of measurements. *Adv Phys* 56:167–242.
4. Gfeller D, De Los Rios P, Caflisch A, Rao F (2007) Complex network analysis of free-energy landscapes. *Proc Natl Acad Sci USA* 104:1817–1822.
5. Yaliraki SN, Barahona M (2007) Chemistry across scales: from molecules to cells. *Philos T Roy Soc A* 365:2921–2934.
6. Fortunato S, Castellano C (2008) Community structure in graphs. *Encyclopedia of Complexity and System Science* (Springer, Berlin).
7. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE T Pattern Anal* 22:888–905.
8. Kannan R, Vempala S, Vetta A (2000) On clusterings: good, bad, and spectral. *Proceedings of the 41st Annual Symposium on Foundations of Computer Science* 367–377.
9. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.
10. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.
11. Muff S, Rao F, Caflisch A (2005) Local modularity measure for network clusterizations. *Phys Rev E* 72:056107.
12. Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. *Phys Rev E* 74:016110.
13. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci USA* 104:36–41.
14. Arenas A, Fernández A, Gómez S (2008) Multiple resolution of the modular structure of complex networks. *New J Phys* 10:05039.
15. Lancichinetti A, Fortunato S, Kertész J (2009) Detecting the overlapping and hierarchical community structure of complex networks. *New J Phys* 11:033015.
16. Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E* 74:036104.
17. Fiedler M (1973) Algebraic connectivity of graphs. *Czech Math J* 23:298–305.
18. Fiedler M (1975) A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czech Math J* 25:619–633.
19. van Driessche R, Roose D (1995) An improved spectral bisection algorithm and its application to dynamic load balancing. *Parallel Comput* 21:29–48.
20. Brandes U, et al. (2008) On modularity clustering. *IEEE T Knowl Data En* 20:172–188.
21. Guimerà R, Sales-Pardo M, Amaral LAN (2004) Modularity from fluctuations in random graphs and complex networks. *Phys Rev E* 70:025101.
22. Ronhovde P, Nussinov Z (2009) Multiresolution community detection for megascale networks by information-based replica correlations. *Phys Rev E* 80:016109.
23. Lambiotte R, Delvenne J-C, Barahona M (2008) Laplacian dynamics and multiscale modular structure in networks. arXiv:0812.1770.
24. Ravász E, Barabási A-L (2003) Hierarchical organization in complex networks. *Phys Rev E* 67:026112.
25. Gohlke H, Thorpe MF (2006) A natural coarse graining for simulating large biomolecular motion. *Biophys J* 91:2115–2120.
26. Hemberg M, Yaliraki SN, Barahona M (2006) Stochastic kinetics of viral capsid assembly based on detailed protein structures. *Biophys J* 90:3029–3042.
27. Gfeller D, De Los Rios P (2008) Spectral coarse graining and synchronization in oscillator networks. *Phys Rev Lett* 100:174104.
28. Reichardt J, Bornholdt S (2004) Detecting fuzzy community structures in complex networks with a Potts model. *Phys Rev Lett* 93:218701.
29. van Dongen S (2000) A cluster algorithm for graphs. *Technical Report INS-R0010* (National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, Amsterdam).
30. Fouss F, Pirotte A, Renders J-M, Saerens M (2006) Random-walk computation of similarities between nodes of a graph, with application to collaborative recommendation. *IEEE T Knowl Data En* 19:355–369.
31. Latapy M, Pons P (2006) Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications* 10:191–218.
32. Arenas A, Díaz-Guilera A, Pérez-Vicente CJ (2006) Synchronization reveals topological scales in complex networks. *Phys Rev Lett* 96:114102.
33. Arenas A, Díaz-Guilera A (2007) Synchronization and modularity in complex networks. *Eur Phys J-Spec Top* 143:19–25.
34. Alon N (1998) Spectral techniques in graph algorithms. *LATIN '98: Proceedings of the Third Latin American Symposium on Theoretical Informatics* (Springer-Verlag, London), pp 206–215.
35. Arenas A, Fernández A, Fortunato S, Gómez S (2008) Motif-based communities in complex networks. *J Phys A-Math Theor* 41:224001.