MULTISCALE MODEL. SIMUL.
Vol. 14, No. 1, pp. 1–41
© 2016 Society for Industrial and Applied Mathematics

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

COMMUNITY DETECTION IN TEMPORAL MULTILAYER NETWORKS, WITH AN APPLICATION TO CORRELATION NETWORKS*

MARYA BAZZI†, MASON A. PORTER‡, STACY WILLIAMS§, MARK MCDONALD§, DANIEL J. FENN§, AND SAM D. HOWISON¶

Abstract. Networks are a convenient way to represent complex systems of interacting entities. Many networks contain "communities" of nodes that are more densely connected to each other than to nodes in the rest of the network. In this paper, we investigate the detection of communities in temporal networks represented as multilayer networks. As a focal example, we study time-dependent financial-asset correlation networks. We first argue that the use of the "modularity" quality function—which is defined by comparing edge weights in an observed network to expected edge weights in a "null network"—is application-dependent. We differentiate between "null networks" and "null models" in our discussion of modularity maximization, and we highlight that the same null network can correspond to different null models. We then investigate a multilayer modularity-maximization problem to identify communities in temporal networks. Our multilayer analysis depends only on the form of the maximization problem and not on the specific quality function that one chooses. We introduce a diagnostic to measure *persistence* of community structure in a multilayer network partition. We prove several results that describe how the multilayer maximization problem measures a trade-off between static community structure within layers and larger values of persistence across layers. We also discuss some computational issues that the popular "Louvain" heuristic faces with temporal multilayer networks and suggest ways to mitigate them.

Key words. community structure, multilayer networks, temporal networks, modularity maximization, financial correlation networks

AMS subject classifications. 62H30, 91C20, 94C15, 90C35

DOI. 10.1137/15M1009615

1. Introduction.

In its simplest form, a network is simply a graph: it consists of a set of *nodes* that represent entities and a set of *edges* between pairs of nodes that represent interactions between those entities. One can consider *weighted graphs* (in which each edge has an associated edge weight that quantifies the interaction of interest) or *unweighted graphs* (weighted graphs with binary edge weights). Networks provide useful representations of complex systems across many disciplines [54]. Common types include social networks (which arise via offline and/or online interactions), information networks (e.g., hyperlinks between webpages on the World Wide Web),

*Received by the editors February 23, 2015; accepted for publication (in revised form) September 14, 2015; published electronically January 6, 2016. This work was supported by HSBC Bank. The first and second authors were supported by FET-Proactive project PLEXMATH (FP7-ICT-2011-8, grant 317614) funded by the European Commission, and also received support from the James S. McDonnell Foundation (220020177).

http://www.siam.org/journals/mms/14-1/M100961.html

†Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, Oxford OX2 6GG, UK (bazzi@maths.ox.ac.uk). This author acknowledges a CASE studentship award from the EPSRC (BK/10/41).

‡Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, Oxford OX2 6GG, UK, and CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK (porterm@maths.ox.ac.uk). This author was supported by the EPSRC (EP/J001759/1).

§Global Research, HSBC Bank, London E14 5HQ, UK (stacy.williams@hsbcgroup.com, mark.mcdonald@hsbcib.com, dan.fenn@hsbcib.com).

¶Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, Oxford OX2 6GG, UK, and Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford OX2 6ED, UK (howison@maths.ox.ac.uk).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

infrastructure networks (e.g., transportation routes between cities), and biological networks (e.g., metabolic interactions between cells or proteins, food webs, etc.).

Given a network representation of a system, it can be useful to apply a coarse-graining technique in order to investigate features that lie between those at the "microscale" (e.g., nodes and pairwise interactions) and those at the "macroscale" (e.g., total edge weight and degree distribution[1]) [55, 61]. One thereby studies "mesoscale" features such as core-periphery structure and (especially) community structure. Loosely speaking, a *community* (or *cluster*) in a network is a set of nodes that are "more densely" connected to each other than they are to nodes in the rest of the network [24, 61]. Giving a precise definition of "densely connected" is, of course, necessary to have a method for community detection. It is important to recognize at the outset that this definition is subjective and in particular may depend on the application in question. Correspondingly, community detection methods may need to be tailored. We restrict ourselves to *hard partitions*, in which each node is assigned to exactly one community, and we use the term "partition" to mean "hard partition." It is also important, but beyond the scope of this paper, to consider "soft partitions," in which communities can overlap [24, 36, 58, 61].

Analysis of community structure has been very useful in a wide range of applications, many of which are described in [24, 26, 55, 61]. In social networks, communities can reveal groups of people with common interests, places of residence, or other similarities [56, 73]. In biological systems, communities can reveal functional groups that are responsible for synthesizing or regulating an important chemical product [31, 45]. In the present paper, we use financial-asset correlation networks as examples [7, 14]. Despite the diversity of markets, financial products, and geographical locations, financial assets can exhibit strong time-dependent correlations, both within and between asset classes. It is a primary concern for market practitioners (e.g., for portfolio diversification) to estimate the strengths of these correlations and to identify sets of assets that are highly correlated [48, 74].

Most methods for detecting communities are designed for static networks. However, in many applications, entities and/or interactions between entities evolve in time. In such applications, one can use the formalism of *temporal networks*, where nodes and/or their edge weights vary in time [33, 34]. This is important for numerous applications, including person-to-person communication [75], one-to-many information dissemination (e.g., Twitter networks [28] and Facebook networks [77]), cell biology [34], neuroscience [6], ecology [34], finance [21, 22, 23, 57], and more.

Two main approaches have been adopted to detect communities in time-dependent networks. The first entails constructing a static network by aggregating snapshots of the evolving network at different points in time into a single network (e.g., by taking the mean or total edge weight for each edge across all time points, which can be problematic if the set of nodes varies in time, and which also makes restrictive assumptions on the interaction dynamics between entities [32]). One can then use standard network techniques. The second approach entails using static community-detection techniques on each element of a time-ordered sequence of networks at different times or on each element of a time-ordered sequence of network aggregations[2] (computed as above)

[1] A node's "degree" is the number of edges attached to it; degree is the special case of "strength" (2.1) for unweighted networks.

[2] One needs to distinguish between this kind of aggregation and the averaging of a set of time series over a moving window to construct a correlation matrix, which one can then interpret as a fixed-time snapshot of a time-evolving network. Although both involve averaging over a time window, the former situation entails averaging a network, and the latter situation entails averaging over a collection of time series (one for each node) with no directly observable edge weights.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

over different time intervals (which can be either overlapping or nonoverlapping) and then tracking the communities across the sequence [2, 21, 22, 35, 47, 58].

A third approach consists of embedding a time-ordered sequence of networks in a larger network [50, 60] (and related ideas are also relevant in other contexts [18, 70]). Each element of the sequence is a network *layer*, and nodes at different time points are joined by *interlayer* edges. This approach was introduced in [50], and the resulting network is a type of *multilayer network* [10, 38]. The main difference between this approach and the previous approach is that the presence of nonzero interlayer edges introduces a dependence between communities identified in one layer and connectivity patterns in other layers. Thus far, most computations that have used a multilayer representation of temporal networks have assumed that interlayer connections are "diagonal" (i.e., they exist only between copies of the same node) and "ordinal" (i.e., they exist only between consecutive layers) [38]. Diagonal is a natural model of the persistence of node identity in time, while ordinal preserves the time ordering.

The authors of [50] derived a generalization of *modularity maximization*, a popular clustering method for static networks, to multilayer networks. Modularity is a function that measures the "quality" of a network partition into disjoint sets of nodes by computing the difference between the total edge weight in sets in the observed network and the total expected edge weight in the same sets in a "null network" generated from some "null model" [24, 61]. Modularity maximization consists of maximizing the modularity quality function over the space of network partitions. (In practice, given the combinatorial complexity of this maximization problem, one uses some computational heuristic and finds a local maximum [29].) Intuitively, the null model controls for connectivity patterns that one anticipates finding in a network, and one uses modularity maximization to identify connectivity patterns in an observed network that are stronger than anticipated.

In this paper, we address two main issues: (1) the choice of null network, and (2) the role of interlayer edges in multilayer modularity maximization. We discuss the first issue in section 4 and the second issue in section 5. Most of our conclusions from sections 4 and 5 are applicable to an arbitrary choice of single-layer networks within layers, and we use financial correlation networks as illustrative examples. In sections 2 and 3, we give an overview of existing results.

We give a precise definition of the modularity function for single-layer networks in section 2, where (importantly) we distinguish between a "null network" and a "null model" in modularity maximization. In section 4, we discuss the choice of null network for a given application. In section 3, we describe the generalization of single-layer modularity to multilayer networks proposed in [50]. To date, almost no theory has explained how a multilayer partition obtained with zero interlayer coupling (which reduces to single-layer modularity maximization on each layer independently) differs from a multilayer partition obtained with nonzero interlayer coupling. In section 5, we prove several theoretical properties of an optimal solution for the multilayer maximization problem to better understand how such partitions differ and how one can exploit this difference in practice. We also describe two computational issues that arise when using the popular Louvain heuristic [9] to solve the multilayer maximization problem, and we suggest ways to mitigate them. The results of section 5 depend only on the form of the maximization problem and still hold if one uses a quality function other than the modularity quality function, provided that it has the same form (e.g., they are valid for "stability" [17, 41, 42]). We conclude in section 6.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

2. Single-layer modularity maximization.

2.1. The modularity function. Consider an N-node network $\mathcal{G}$ and let the edge weights between pairs of nodes be $\{A_{ij}|i,j \in \{1,\ldots,N\}\}$, so that $\boldsymbol{A} = (A_{ij}) \in \mathbb{R}^{N\times N}$ is the *adjacency matrix* of $\mathcal{G}$. In this paper, we consider only symmetric adjacency matrices (and hence undirected networks), so $A_{ij} = A_{ji}$ for all i and j. The *strength* of a node i is

$$k_i = \sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{N} A_{ji}\,, \tag{2.1}$$

and it is given by the ith row (or column) sum of $\boldsymbol{A}$.

When studying the structure of a network, it is useful to compare what is observed with what is anticipated. We define a *null model* to be a probability distribution on the set of adjacency matrices, and a *null network* to be the expected adjacency matrix under a specified null model. In a loose sense, null models play the role of *prior models*, as they control for features that one anticipates finding in the system under investigation. One can thereby take into account known (or suspected) connectivity patterns that might obscure unknown connectivity patterns that one hopes to discover via processes like community detection. For example, in social networks, one often takes the strength of a node in a null network to be its observed strength k_i [53, 55, 61]. We discuss the use of this null network for financial-asset correlation networks in section 4. In spatial networks that represent the spread of a disease or information between different locations, some authors have used null networks in which edge weights between two locations scale inversely with the distance between them [20, 66].

As we discussed in section 1, one uses modularity maximization to partition a network into sets of nodes called "communities" that have a larger total internal edge weight than the expected total internal edge weight in the same sets in a null network, generated from some null model [24, 55, 56, 61]. *Modularity maximization* consists of finding a partition that maximizes this difference [24, 61]. (As we mentioned earlier, in practice, one uses some computational heuristic and finds a local maximum [29].) In the present paper, we do not restrict ourselves to the usual choice of null network (i.e., the "Newman–Girvan" null network in subsection 2.4.1) for the modularity quality function, and we ignore any normalization constant that depends on the choice of null network but does not affect the solution of the modularity-maximization problem for a given null network. Modularity thus acts as a "quality function" $Q : \mathcal{C} \to \mathbb{R}$, where the set $\mathcal{C}$ is the set of all possible N-node network partitions.

Suppose that we have a partition C of a network into K disjoint sets of nodes $\{C_1,\ldots,C_K\}$. We can then define a map $c(\cdot)$ from the set of nodes $\{1,\ldots,N\}$ to the set of integers $\{1,\ldots,K\}$ such that $c(i) = c(j) = k$ if and only if nodes i and j lie in C_k. We use the term *global maximum* to refer to a solution of the modularity-maximization problem, and the term *local maximum* to refer to a solution that one obtains with a computational heuristic. We call $c(i)$ the *set assignment* (or *community assignment* when C is a global or local maximum) of node i in partition C. The value of modularity for a given partition C is then

$$Q(C|\boldsymbol{A};\boldsymbol{P}) := \sum_{i,j=1}^{N} (A_{ij} - P_{ij})\delta(c_i, c_j)\,, \tag{2.2}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

where $\boldsymbol{P} = (P_{ij}) \in \mathbb{R}^{N\times N}$ is the adjacency matrix of the null network, c_i is shorthand notation for $c(i)$, and $\delta(c_i, c_j)$ is the Kronecker delta function. We state the modularity-maximization problem as follows:

$$\max_{C\in\mathcal{C}} \sum_{i,j=1}^{N} (A_{ij} - P_{ij})\delta(c_i, c_j)\,, \tag{2.3}$$

which we can also write as $\max_{C\in\mathcal{C}} Q(C|\boldsymbol{B})$ or $\max_{C\in\mathcal{C}} \sum_{i,j}^{N} B_{ij}\delta(c_i, c_j)$, where $\boldsymbol{B} = \boldsymbol{A} - \boldsymbol{P}$ is the so-called *modularity matrix* [53]. The number K of sets in a partition is free in the optimization problem (2.3). (In other words, one maximizes over the set of *all* N-node partitions.) We only consider a fixed value of K when we consider a particular partition $C \in \mathcal{C}$, with $|C| = K$. We allow self-edges in all numerical experiments with financial data of sections 4 and 5. (In particular, $\boldsymbol{A}$ is a Pearson correlation matrix with $A_{ii} = 1$.) We assume in the rest of the paper that each of the partitions in the set $\mathcal{C}$ contains sets that do not have multiple connected components in the graph with adjacency matrix $\boldsymbol{B}$. It is clear from (2.3) that pairwise contributions to modularity are only counted when two nodes are assigned to the same set. These contributions are positive (respectively, negative) when the observed edge weight A_{ij} between nodes i and j is larger (respectively, smaller) than the expected edge weight P_{ij} between them. If $A_{ij} < P_{ij}$ for all i and j, then the optimal solution is N singleton communities. Conversely, if $A_{ij} > P_{ij}$ for all i and j, then the optimal solution is a single N-node community. To obtain a partition of a network with a high value of modularity, one hopes to have many edges within sets that satisfy $A_{ij} > P_{ij}$ and few edges within sets that satisfy $A_{ij} < P_{ij}$. As is evident from (2.3), what one regards as "densely connected" in this setting depends fundamentally on the choice of null network.

It can be useful to write the modularity-maximization problem using the trace of matrices [53]. As before, we consider a partition C of a network into K sets of nodes $\{C_1, \ldots, C_K\}$. We define the *partition matrix* $\boldsymbol{S} \in \{0,1\}^{N\times K}$ as

$$S_{ij} = \delta(c_i, j)\,, \tag{2.4}$$

where $j \in \{1, \ldots, K\}$ and $c_i = j$ means that node i lies in C_j. The columns of $\boldsymbol{S}$ are orthogonal, and the jth column sum of $\boldsymbol{S}$ gives the number of nodes in C_j. This yields

$$\sum_{i,j=1}^{N} B_{ij}\delta(c_i, c_j) = \sum_{i,j=1}^{N}\sum_{k=1}^{K} S_{ik}B_{ij}S_{jk} = \mathrm{Tr}(\boldsymbol{S}^T\boldsymbol{B}\boldsymbol{S})\,,$$

where the (i,i)th term of $\boldsymbol{S}^T\boldsymbol{B}\boldsymbol{S}$ is twice the sum of modularity-matrix entries between pairs of nodes in C_i. (The (i,j)th off-diagonal term is the sum of modularity-matrix entries between pairs of nodes with one node in C_i and one node in C_j.) It follows that one can restate the modularity-maximization problem in (2.3) as

$$\max_{\boldsymbol{S}\in\mathcal{S}} \mathrm{Tr}(\boldsymbol{S}^T\boldsymbol{B}\boldsymbol{S})\,, \tag{2.5}$$

where $\mathcal{S}$ is the set of all partition matrices in $\{0,1\}^{N\times K}$ (with $K \leq N$).

Modularity maximization is one of myriad community-detection methods [24], and it has many limitations (e.g., a resolution limit on the size of communities [25]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

and a huge number of nearly degenerate local maxima [29]). Nevertheless, it is a popular method (which has been used successfully in numerous applications [24, 61]), and the ability to specify explicitly what one anticipates is a useful (and under-exploited) feature for users working on different applications. In section 4, we make some observations on one's choice of null network when using the modularity quality function.

2.2. The Louvain computational heuristic. For a given modularity matrix $\boldsymbol{B}$, a solution to the modularity-maximization problem is guaranteed to exist in any network with a finite number of nodes. However, the number of possible partitions in an N-node network, given by the Bell number [8], grows at least exponentially with N, so an exhaustive search of the space of partitions is infeasible. Modularity maximization was proved in [13] to be an NP-hard problem (at least for the null networks which we consider in this paper), so solving it requires the use of computational heuristics. In the present paper, we focus on the Louvain heuristic, which is a locally greedy modularity-increasing sampling process over the set of partitions [9].

The Louvain heuristic consists of two phases, which are repeated iteratively. Initially, each node in the network constitutes a set, which gives an initial partition that consists of N singletons. During phase 1, one considers the nodes one by one (in some order), and one places each node in a set (possibly where it already is) that results in the largest increase of modularity. This phase is repeated until one reaches a local maximum (i.e., until one obtains a partition in which the move of a single node cannot increase modularity). Phase 2 consists of constructing a reduced network $\mathcal{G}'$ from the sets of nodes in $\mathcal{G}$ that one obtains after the convergence of phase 1. We denote the sets in $\mathcal{G}$ at the end of phase 1 by $\{\hat{C}_1, \ldots, \hat{C}_{\hat{N}}\}$ (where $\hat{N} \leq N$) and the set assignment of node i in this partition by $\hat{c}_i$. Each set $\hat{C}_k$ in $\mathcal{G}$ constitutes a node k in $\mathcal{G}'$, and the reduced modularity matrix of $\mathcal{G}'$ is

$$\boldsymbol{B}' = \hat{\boldsymbol{S}}^{\boldsymbol{T}} \boldsymbol{B} \hat{\boldsymbol{S}}\,,$$

where $\hat{\boldsymbol{S}}$ is the partition matrix of $\{\hat{C}_1, \ldots, \hat{C}_{\hat{N}}\}$. This ensures that the all-singleton partition in $\mathcal{G}'$ has the same value of modularity as the partition of $\mathcal{G}$ that we identified at the end of phase 1. One then repeats phase 1 on the reduced network and continues iterating until the heuristic converges (i.e., until phase 2 induces no further changes).

Because we use a nondeterministic implementation of the Louvain heuristic—in particular, the node order is randomized at the start of each iteration of phase 1—the network partitions that we obtain for a fixed modularity matrix can differ across runs.[3] To account for this, one can compute the frequency of co-classification of nodes into communities for a given modularity matrix $\boldsymbol{B}$ across multiple runs of the heuristic instead of using the output partition of a single run. (See [44] for an application of such an approach to "consensus clustering" and [65] for an application of such an approach to hierarchical clustering.) We use the term *association matrix* for a matrix that stores the mean number of times that two nodes are placed in the same community across multiple runs of a heuristic, and we use the term *co-classification index* of nodes i and j to designate the (i, j)th entry of an association matrix.

[3]The implementation [37] of the heuristic that we use in this paper is a generalized version of the implementation in [9]. It is independent of the null network—it takes the modularity matrix as an input to allow an arbitrary choice of null network—and it randomizes the node order at the start of each iteration of phase 1 of the heuristic to increase the search space of the heuristic. When one chooses the same null network that was assumed in [9] and uses a node order fixed to $\{1, \ldots, N\}$ at each iteration of phase 1 (the value of N can change after each iteration of the heuristic's phase 2), then the implementation in [37] and the implementation described in [9] return the same output.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

There are many other heuristics that one can employ to maximize modularity [24, 54, 61], but the Louvain heuristic is a popular choice in practice [43]. It is very fast [24, 43], which is an important consideration in multilayer networks, for which the total number of nodes is the number of nodes in each layer multiplied by the number of layers. In section 5, we point out two issues that the Louvain heuristic (independently of how it is implemented) faces with temporal multilayer networks.

2.3. Multiscale community structure. Many networks include community structure at multiple scales [24, 61], and some systems even have a hierarchical community structure of "parts-within-parts" [68]. In such a situation, although there are dense interactions within communities of some size (e.g., friendship ties between students in the same school), there are even denser interactions in subsets of nodes that lie inside of these communities (e.g., friendship ties between students in the same school and in the same class year). Some variants of the modularity function have been proposed to detect communities at different scales. A popular choice is to scale the null network using a *resolution parameter* $\gamma \geq 0$ to yield a *multiscale modularity-maximization* problem [64]:

$$\max_{C \in \mathcal{C}} \sum_{i,j=1}^{N} (A_{ij} - \gamma P_{ij}) \delta(c_i, c_j) \,. \tag{2.6}$$

In some sense, the value of the parameter γ determines the importance that one assigns to the null network relative to the observed network. The corresponding modularity matrix and modularity function evaluated at a partition C are $\boldsymbol{B} = \boldsymbol{A} - \gamma \boldsymbol{P}$ and $Q(C|\boldsymbol{A}; \boldsymbol{P}; \gamma) = \sum_{i,j=1}^{N} (A_{ij} - \gamma P_{ij}) \delta(c_i, c_j)$. The special case $\gamma = 1$ yields the modularity matrix and modularity function in the modularity-maximization problem (2.3). This formulation of multiscale modularity has a dynamical interpretation [40, 41, 42] that we will discuss in the next subsection.

In most applications of community detection, the adjacency matrices of the observed and null networks have nonnegative entries. In these cases, the solution to (2.6) when $0 \leq \gamma \leq \gamma^- = \min_{i \neq j, P_{ij} \neq 0} (A_{ij}/P_{ij})$ is a single community regardless of any structure, however clear, in the observed network, because then

$$B_{ij} = A_{ij} - \gamma P_{ij} \geq 0 \quad \text{for all} \quad i, j \in \{1, \ldots, N\} \,.$$

(We exclude diagonal terms because a node is always in its own community.) However, the solution to (2.6) when $\gamma > \gamma^+ = \max_{i \neq j, P_{ij} \neq 0} (A_{ij}/P_{ij})$ is N singleton communities because

$$B_{ij} = A_{ij} - \gamma P_{ij} \leq 0 \quad \text{for all} \quad i, j \in \{1, \ldots, N\} \,.$$

(In the above inequality, we assume that $P_{ij} = 0$ can occur only if $A_{ij} = 0$.) Partitions at these boundary values of γ correspond to the coarsest and finest possible partitions of a network, and varying the resolution parameter between these bounds makes it possible to examine a network's community structure at intermediate scales.

For an observed and/or null network with signed edge weights, the intuition behind the effect of varying γ in (2.6) on an optimal solution is not straightforward. A single community and N singleton communities do not need to be optimal partitions for any value of $\gamma \geq 0$. In particular, B_{ij} has the same sign as A_{ij} for sufficiently small values of γ, and B_{ij} has the sign opposite that of P_{ij} for sufficiently large values of γ. For further discussion, see section 4, where we explore the effect of varying the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

resolution parameter on an optimal partition for an observed and null network with signed edge weights. We vary the resolution parameter in the interval $[0, \gamma^+]$ instead of $[\gamma^-, \gamma^+]$ in numerical experiments with signed networks because globally optimal partitions can be different when $\gamma \in [0, \gamma^-]$.[4]

It is important to differentiate between a "resolution limit" on the smallest community size that is imposed by a community-detection method [25] and inherently multiscale community structure in a network [24, 61, 68]. For the formulation of multiscale modularity in (2.6), the resolution limit described in [25] applies to any fixed value of γ. By varying γ, one can identify communities that are smaller than the limit for any particular γ value. In this sense, multiscale formulations of modularity help "mitigate" the resolution limit, though there remain issues [1, 29, 40]. In this paper, we do not address the issue of how to identify communities at different scales, though we note in passing that the literature includes variants of multiscale modularity (e.g., see [1]). We make observations on null networks in section 4, and we illustrate how our observations can occur in practice using the formulation of multiscale modularity in (2.6). (Our observations hold independently of the formulation of multiscale modularity that one adopts, but the precise manifestation can be different for different variants of multiscale modularity.)

We use the term *multiscale community structure* to refer to a set $\mathcal{C}_{\text{local}}(\boldsymbol{\gamma})$ of local optima that we obtain with a computational heuristic for a set of (not necessarily all distinct) resolution-parameter values $\boldsymbol{\gamma} = \{\gamma_1, \ldots, \gamma_l\}$, where $\gamma^- = \gamma_1 \leq \cdots \leq \gamma_l = \gamma^+$. We use the term *multiscale association matrix* for an association matrix $\hat{\boldsymbol{A}} \in [0,1]^{N \times N}$ that stores the co-classification index of all pairs of nodes for partitions in this set:

$$\hat{A}_{ij} = \frac{\sum_{C \in \mathcal{C}_{\text{local}}(\boldsymbol{\gamma})} \delta(c_i, c_j)}{|\mathcal{C}_{\text{local}}(\boldsymbol{\gamma})|} \,. \tag{2.7}$$

For each partition $C \in \mathcal{C}_{\text{local}}(\boldsymbol{\gamma})$, nodes i and j either are in the same community (i.e., $\delta(c_i, c_j) = 1$) or are not in the same community (i.e., $\delta(c_i, c_j) = 0$). It follows that $\hat{A}_{ij}$ is the mean number of partitions in $\mathcal{C}_{\text{local}}(\boldsymbol{\gamma})$ for which nodes i and j are in the same community. The value of $|\mathcal{C}_{\text{local}}(\boldsymbol{\gamma})|$ is the number of distinct values of γ that we consider multiplied by the number of runs of the Louvain algorithm performed for each value of γ. In section 4, we use a discretization of $[\gamma^-, \gamma^+]$ (respectively, $[0, \gamma^+]$) for unsigned (respectively, signed) networks and perform a single run for each resolution-parameter value. The number of partitions in $\mathcal{C}_{\text{local}}(\boldsymbol{\gamma})$ is then precisely the number of distinct values of γ that we consider (i.e., one partition per value of γ). We use the matrix $\hat{\boldsymbol{A}}$ repeatedly in our computational experiments in section 4.

2.4. Null models and null networks. In this section, we describe three null networks. We make several observations on the interpretation of communities that we obtain from Pearson correlation matrices using each of these null networks in the computational experiments of section 4.

2.4.1. Newman–Girvan null network. A popular choice of null network for networks with positive edge weights is the Newman–Girvan (NG) null network, whose adjacency-matrix entries are $P_{ij} = k_i k_j / (2m)$, where k_i are the observed node strengths

[4]For $\gamma \geq \gamma^+$, one can show that all modularity contributions no longer change signs. They are nonpositive (respectively, nonnegative) between pairs of nodes with $P_{ij} \geq 0$ (respectively, $P_{ij} \leq 0$).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

[52, 56]. This yields the equivalent maximization problems

$$\max_{C\in\mathcal{C}} \sum_{i,j=1}^{N} \left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j) \Leftrightarrow \max_{\boldsymbol{S}\in\mathcal{S}} \operatorname{Tr}\left[\boldsymbol{S}^T\left(\boldsymbol{A} - \frac{\boldsymbol{k}\boldsymbol{k}^T}{2m}\right)\boldsymbol{S}\right], \tag{2.8}$$

where $\boldsymbol{k} = \boldsymbol{A}\mathbf{1}$ is the $N \times 1$ vector of node strengths and $2m = \mathbf{1}^T\boldsymbol{A}\mathbf{1}$ is the total edge weight of the observed network. This null network can be derived from a variety of null models. One way to generate an unweighted network with expected adjacency matrix $\boldsymbol{k}\boldsymbol{k}^T/(2m)$ is to generate each of its edges and self-edges with probability $k_i k_j/(2m)$ (provided $k_i k_j \leq 2m$ for all i, j). That is, the presence and absence of edges and self-edges is a Bernoulli random variable with probability $k_i k_j/(2m)$ [11, 12]. More generally, any probability distribution on the set of adjacency matrices that satisfies $\mathbb{E}\big(\sum_{j=1}^{N} W_{ij}\big) = k_i$ (i.e., the expected strength equals the observed strength; see, e.g., [15]) and $\mathbb{E}(W_{ij}) = f(k_i)f(k_j)$ for some real-valued function f has an expected adjacency matrix of $\mathbb{E}(\boldsymbol{W}) = \boldsymbol{k}\boldsymbol{k}^T/(2m)$.[5] The adjacency matrix of the NG null network is symmetric and positive semidefinite.

We briefly describe a way of deriving the NG null network from a model on time-series data (in contrast to a model on a network). The *partial correlation* $\operatorname{corr}(a, b\,|\,c)$ between a and b given c is the Pearson correlation between the residuals that result from the linear regression of a with c and b with c, and it is given by

$$\operatorname{corr}(a, b\,|\,c) = \frac{\operatorname{corr}(a,b) - \operatorname{corr}(a,c)\operatorname{corr}(b,c)}{\sqrt{1-\operatorname{corr}^2(a,c)}\sqrt{1-\operatorname{corr}^2(b,c)}}\,. \tag{2.9}$$

Suppose that the data used to construct the observed network is a set of time series $\{z_i | i \in \{1, \ldots, N\}\}$, where $z_i = \{z_i(t) | t \in T\}$ and T is a discrete set of time points. The authors of [46] pointed out that $A_{ij} = \operatorname{corr}(z_i, z_j)$ implies that $k_i = \operatorname{cov}(\hat{z}_i, \hat{z}_{\text{tot}})$ and thus that

$$\frac{k_i k_j}{2m} = \operatorname{corr}(\hat{z}_i, \hat{z}_{\text{tot}})\operatorname{corr}(\hat{z}_j, \hat{z}_{\text{tot}})\,, \tag{2.10}$$

where $\hat{z}_i(t) = (z_i(t) - \langle z_i \rangle)/\sigma(z_i)$ is a standardized time series and $\hat{z}_{\text{tot}}(t) = \sum_{i=1}^{N} \hat{z}_i(t)$ is the sum of the standardized time series.[6] Taking $a = \hat{z}_i$, $b = \hat{z}_j$, and $c = \hat{z}_{\text{tot}}$, (2.9) implies that if $\operatorname{corr}(\hat{z}_i, \hat{z}_j\,|\,\hat{z}_{\text{tot}}) = 0$, then $\operatorname{corr}(\hat{z}_i, \hat{z}_j) = k_i k_j/(2m)$. That is, Pearson correlation coefficients between pairs of time series that satisfy $\operatorname{corr}(\hat{z}_i, \hat{z}_j\,|\,\hat{z}_{\text{tot}}) = 0$ are precisely the adjacency-matrix entries of the NG null network. One way of generating a set of time series in which pairs of distinct time series satisfy this condition is to assume that each standardized time series depends linearly on the mean time series and that residuals are mutually uncorrelated (i.e., $\hat{z}_i = \alpha_i \hat{z}_{\text{tot}}/N + \beta_i + \epsilon_i$ for some $\alpha_i, \beta_i \in \mathbb{R}$ and $\operatorname{corr}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$).

The multiscale modularity-maximization problem in (2.6) was initially introduced in [64] using an ad hoc approach. Interestingly, one can derive this formulation of

[5] The linearity of the expectation and the assumptions $\mathbb{E}(\sum_{j=1}^{N} W_{ij}) = k_i$ and $\mathbb{E}(W_{ij}) = f(k_i)f(k_j)$ imply that $f(k_i) = k_i/\sum_{j=1}^{N} f(k_j)$ and $\sum_{j=1}^{N} f(k_j) = \sqrt{2m}$. Combining these equations gives the desired result.

[6] The equality (2.10) holds for networks constructed using Pearson correlations between time series [46]. Such networks are examples of signed networks with edge weights in $[-1, 1]$. The strength of a node i is given by the ith (signed) column or row sum of the correlation matrix.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

the maximization problem for sufficiently large values of γ by considering a quality function based on a continuous-time Markov process $X(t)$ on an observed network [40, 41, 42]. The probability density of a continuous-time Markov process with exponentially distributed waiting times at each node parametrized by $\lambda(i)$ satisfies

$$\dot{\boldsymbol{p}} = \boldsymbol{p}\boldsymbol{\Lambda}\boldsymbol{M} - \boldsymbol{p}\boldsymbol{\Lambda}\,, \tag{2.11}$$

where the vector $\boldsymbol{p}(t) \in [0,1]^{1\times N}$ is the probability density of a random walker at each node [i.e., $p_i(t) := \mathbb{P}(X(t) = i)$ for each i], $\boldsymbol{\Lambda}$ is a diagonal matrix with the rate $\lambda(i)$ on its ith diagonal entry, and $\boldsymbol{M}$ is the transition matrix of a random walker (i.e., $M_{ij} := A_{ij}/k_i$). The solution to (2.11) is $\boldsymbol{p}(t) = \boldsymbol{p}_0 e^{\boldsymbol{\Lambda}(\boldsymbol{M}-\boldsymbol{I})t}$, and its stationary distribution is unique (provided the network is connected) and is given by $\boldsymbol{\pi} = c\boldsymbol{k}^T\boldsymbol{\Lambda}^{-1}/(2m)$, where the normalization constant c ensures that $\sum_{i=1}^N \pi_i = 1$. The *stability* of a partition is a quality function defined by [17, 40, 41, 42]

$$r(\boldsymbol{S}, t) = \mathrm{Tr}\left[\boldsymbol{S}^T\left(\boldsymbol{\Pi}e^{\boldsymbol{\Lambda}(\boldsymbol{M}-\boldsymbol{I})t} - \boldsymbol{\pi}^T\boldsymbol{\pi}\right)\boldsymbol{S}\right],$$

where $\Pi_{ij} = \delta(i,j)\pi_i$. Equivalently, the stability is

$$r(C,t) = \sum_{i,j=1}^{N}\left[\pi_i\left(e^{\boldsymbol{\Lambda}(\boldsymbol{M}-\boldsymbol{I})t}\right)_{ij} - \pi_i\pi_j\right]\delta(c_i,c_j)\,. \tag{2.12}$$

Taking $\boldsymbol{p}_0 = \boldsymbol{\pi}$, the first term in the square brackets on the right-hand side of (2.12) is $\mathbb{P}\big[(X(0) = i) \cap (X(t) = j)\big]$, and the second term in the square brackets is $\lim_{t\to\infty}\mathbb{P}\big[(X(0) = i) \cap (X(t) = j)\big]$ (provided the system is ergodic). The intuition behind the stability quality function is that a good partition at a given time before reaching stationarity corresponds to one in which the time that a random walker spends within communities is large compared with the time that it spends transiting between communities. In other words, a random walker that starts out at a community ends up there again in the early stages of the random walk, long before stationarity. The resulting maximization problem is $\max_{\boldsymbol{S}\in\mathcal{S}} r(\boldsymbol{S},t)$ or, equivalently, $\max_{C\in\mathcal{C}} r(C,t)$. By linearizing $e^{\boldsymbol{\Lambda}(\boldsymbol{M}-\boldsymbol{I})t}$ at $t = 0$ and taking $\boldsymbol{\Lambda} = \boldsymbol{I}$, one obtains the multiscale modularity-maximization problem in (2.6) at short time scales with $\gamma = 1/t$ and $P_{ij} = k_ik_j/(2m)$. This approach provides a dynamical interpretation of the resolution parameter γ as the inverse (after linearization) of the time used to explore a network by a random walker.

2.4.2. Generalization of Newman–Girvan null network to signed networks. In [27], Gómez, Jensen, and Arenas proposed a generalization of the NG null network to signed networks. We refer to this null network as the NGS null network. They separated $\boldsymbol{A}$ into its positive and negative edge weights:

$$\boldsymbol{A} = \boldsymbol{A}^+ - \boldsymbol{A}^-\,,$$

where $\boldsymbol{A}^+$ denotes the positive part of $\boldsymbol{A}$ and $-\boldsymbol{A}^-$ denotes its negative part. The adjacency-matrix entries of the NGS null network are given by $P_{ij} = k_i^+k_j^+/(2m^+) - k_i^-k_j^-/(2m^-)$, where k_i^+ and $2m^+$ (respectively, k_i^- and $2m^-$) are the strengths and total edge weight in $\boldsymbol{A}^+$ (respectively, $\boldsymbol{A}^-$). This yields the maximization problem

$$\max_{C\in\mathcal{C}}\sum_{i,j=1}^{N}\left[\left(A_{ij}^+ - \frac{k_i^+k_j^+}{2m^+}\right) - \left(A_{ij}^- - \frac{k_i^-k_j^-}{2m^-}\right)\right]\delta(c_i,c_j)\,. \tag{2.13}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

The intuition behind this generalization is to use an NG null network on both unsigned matrices $\boldsymbol{A}^+$ and $\boldsymbol{A}^-$ but to count contributions to modularity from negative edge weights (i.e., the second group of terms in (2.13)) in an opposite way to those from positive edge weights (i.e., the first group of terms in (2.13)). Negative edge weights that exceed their expected edge weight are penalized (i.e., they decrease modularity), and those that do not are rewarded (i.e., they increase modularity). One can generate a network with edge weights 0, 1, or -1 and expected edge weights $k_i^+k_j^+/(2m^+) - k_i^-k_j^-/(2m^-)$ by generating one network with expected edge weights $W_{ij}^+ = k_i^+k_j^+/(2m^+)$ and a second network with expected edge weights $W_{ij}^- = k_i^-k_j^-/(2m^-)$ using the procedure described for the NG null network in section 2.4.1, and then defining a network whose edge weights are given by the difference between the edge weights of these two networks. More generally, by footnote 5 and linearity of expectation, any probability distribution on the set $\{\boldsymbol{W} \in \mathbb{R}^{N\times N}\}$ of signed adjacency matrices that satisfies the two conditions

$$\mathbb{E}\left(\sum_{j=1}^{N} W_{ij}^+\right) = k_i^+ \text{ and } \mathbb{E}(W_{ij}^+) = f(k_i^+)f(k_j^+)\,, \tag{2.14}$$

$$\mathbb{E}\left(\sum_{j=1}^{N} W_{ij}^-\right) = k_i^- \text{ and } \mathbb{E}(W_{ij}^-) = g(k_i^-)g(k_j^-)\,, \tag{2.15}$$

where f and g are real-valued functions and $W_{ij} = W_{ij}^+ - W_{ij}^-$, has an expected edge weight of $\mathbb{E}(W_{ij}) = k_i^+k_j^+/(2m^+) - k_i^-k_j^-/(2m^-)$ for all $i, j \in \{1, \ldots, N\}$.

The authors of [50] derived a variant of the multiscale formulation of modularity in (2.6) for the NGS null network at short time scales by building on the random-walk approach used to derive the NG null network.[7] They considered the function

$$\hat{r}(C,t) = \sum_{i,j=1}^{N} \Big(\pi_i\big[\delta_{ij} + t\Lambda_{ii}(M_{ij} - \delta_{ij})\big] - \pi_i\rho_{i|j}\Big)\delta(c_i, c_j)\,, \tag{2.16}$$

where the term in square brackets on the right-hand side of (2.16) is a linearization of the exponential term in (2.12), $\boldsymbol{M}$ and π_i are as defined in (2.12) on a network with adjacency matrix $|\boldsymbol{A}| := \boldsymbol{A}^+ + \boldsymbol{A}^-$, and $\rho_{i|j}$ is the probability of jumping from node i to node j at stationarity in one step conditional on the network structure [50]. If the network is nonbipartite, unsigned, and undirected, then $\rho_{i|j}$ reduces to the stationary probability π_j.

2.4.3. Uniform null network. A third null network that we consider is a uniform (U) null network, with adjacency-matrix entries $P_{ij} = \langle k\rangle^2/(2m)$, where $\langle k\rangle := \big(\sum_{i=1}^N k_i\big)/N$ denotes the mean strength in a network. We thereby obtain the equivalent maximization problems

$$\max_{C\in\mathcal{C}} \sum_{i,j=1}^{N} \left(A_{ij} - \frac{\langle k\rangle^2}{2m}\right)\delta(c_i, c_j) \Leftrightarrow \max_{\boldsymbol{S}\in\mathcal{S}} \operatorname{Tr}\left[\boldsymbol{S}^T\left(\boldsymbol{A} - \frac{\langle k\rangle^2}{2m}\mathbf{1}_N\right)\boldsymbol{S}\right], \tag{2.17}$$

[7] In particular, they derived the multiscale formulation of modularity obtained using a Potts-model approach in [71]. This multiscale formulation results in one resolution parameter γ_1 for the term $(k_i^+k_j^+)/(2m^+)$ and a second resolution parameter γ_2 for the term $(k_i^-k_j^-)/(2m^-)$ in (2.13) (see [47] for an application of this multiscale formulation to the United Nations General Assembly voting networks). Without an application-driven justification for how to choose these parameters, this increases the parameter space substantially, so we consider only the case $\gamma_1 = \gamma_2$ in this paper.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

where $\boldsymbol{A}$ is an unsigned adjacency matrix and $\mathbf{1}_N$ is an $N \times N$ matrix in which every entry is 1.[8] The expected edge weight in (2.17) is constant and satisfies

$$\frac{\langle k \rangle^2}{2m} = \frac{\left(\sum_{i=1}^N k_i/N\right)^2}{\sum_{i=1}^N k_i} = \frac{2m}{N^2} = \langle \boldsymbol{A} \rangle \,,$$

where $\langle \boldsymbol{A} \rangle$ denotes the mean value of the adjacency matrix.[9] One way to generate an unweighted network with adjacency matrix $\langle \boldsymbol{A} \rangle \mathbf{1}_N$ is to generate each edge with probability $\langle \boldsymbol{A} \rangle$ (provided $\langle \boldsymbol{A} \rangle \leq 1$). That is, the presence and absence of an edge (including self-edges) are independent and identically distributed Bernoulli random variables with probability $\langle \boldsymbol{A} \rangle$. More generally, any probability distribution on the set of adjacency matrices that satisfies $\mathbb{E}\big(\sum_{i,j=1}^N W_{ij}\big) = 2m$ and $\mathbb{E}(W_{ij}) = \mathbb{E}(W_{i'j'})$ for all i, j, i', j' has an expected adjacency matrix $\mathbb{E}(\boldsymbol{W}) = \langle \boldsymbol{A} \rangle \mathbf{1}_N$. The adjacency matrix of the U null network is symmetric and positive semidefinite. One can derive the multiscale formulation in (2.6) for the U null network from the stability quality function in precisely the same way as it is derived for the NG null network, except that one needs to consider exponentially distributed waiting times at each node with rates proportional to node strength (i.e., $\Lambda_{ij} = \delta(i,j) k_i / \langle k \rangle$) [41, 42].

3. Multilayer modularity maximization.

3.1. Multilayer representation of temporal networks. We restrict our attention to temporal networks in which only edges vary in time. (Thus, each node is present in all layers.) We use the notation $\boldsymbol{A}_s$ for a *layer* in a sequence of adjacency matrices $\mathcal{T} = \{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{|\mathcal{T}|}\}$, and we denote node i in layer s by i_s. We use the term *multilayer network* for a network defined on the set of nodes $\{1_1, \ldots, N_1; 1_2, \ldots, N_2; \ldots; 1_{|\mathcal{T}|}, \ldots, N_{|\mathcal{T}|}\}$ [38].

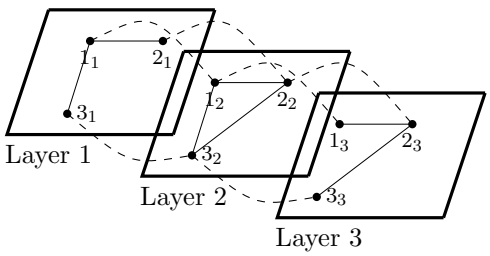
Thus far, computations that have used a multilayer framework for temporal networks have almost always assumed (1) that interlayer connections exist only between nodes that correspond to the same entity (i.e., between nodes i_s and i_r for some i and $s \neq r$), and (2) that the network layers are "ordinal" (i.e., interlayer edges exist only between consecutive layers) [6, 38, 49, 50, 63]. It is also typically assumed that (3) interlayer connections are uniform (i.e., interlayer edges have the same weight). In a recent review article on multilayer networks [38], condition (1) was called "diagonal" coupling, and condition (3) implies that a network is "layer-coupled." We refer to the type of coupling defined by (1), (2), and (3) as "diagonal, ordinal, and uniform" interlayer coupling, and we denote the value of the interlayer edge weight by $\omega \in \mathbb{R}$. We show a simple illustration of a multilayer network with diagonal, ordinal, and uniform interlayer coupling in Figure 1. One can consider more general interlayer connections (e.g., nonuniform ones). Although we restrict our attention to uniform coupling in our theoretical and computational discussions, we give an example of a nonuniform choice of interlayer coupling in section 5. Results similar to those of subsection 5.2 also apply in this more general case.

3.2. The multilayer modularity function. The authors of [50] generalized the single-layer multiscale modularity-maximization problem in (2.6) to a multilayer

[8] For a network in which all nodes have the same strength, the U and NG null networks are equivalent because $k_i = k_j$ for all $i, j \Leftrightarrow k_i = 2m/N = \langle k \rangle$ for all i. This was pointed out for an application to foreign exchange markets in [21].

[9] Although we use the U null network on unsigned adjacency matrices in this paper, the expected edge weight in the U null network is always nonnegative for correlation matrices, as positive semidefiniteness guarantees that $\langle \boldsymbol{A} \rangle = \mathbf{1}^T \boldsymbol{A} \mathbf{1}/(N^2) \geq 0$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



$$\longleftrightarrow \left[\begin{array}{ccc|ccc|ccc} 0 & 1 & 1 & \omega & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \omega & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & \omega & 0 & 0 & 0 \\ \hline \omega & 0 & 0 & 0 & 1 & 1 & \omega & 0 & 0 \\ 0 & \omega & 0 & 1 & 0 & 1 & 0 & \omega & 0 \\ 0 & 0 & \omega & 1 & 1 & 0 & 0 & 0 & \omega \\ \hline 0 & 0 & 0 & \omega & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \omega & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \omega & 0 & 1 & 0 \end{array}\right]$$

FIG. 1. *Example of a multilayer network (left) with unweighted intralayer connections (solid lines) and uniformly weighted interlayer connections (dashed curves) and its corresponding adjacency matrix (right). (The adjacency matrix that corresponds to a multilayer network is sometimes called a "supra-adjacency matrix" in the network-science literature* [38]*.)*

network using an approach similar to that used to derive the NGS null network from a stochastic Markov process on the observed network. For simplicity, we express intralayer and interlayer connections in an $N|\mathcal{T}|$-node multilayer network using a single $N|\mathcal{T}| \times N|\mathcal{T}|$ matrix. Each node i_s in layer s has the unique index $i' := i + (s-1)N$, and we use $\boldsymbol{\mathcal{A}}$ to denote the multilayer adjacency matrix, which has entries $\mathcal{A}_{i'j'} = A_{ijs}\delta(s,r) + \omega\delta(|s-r|,1)$ when the interlayer coupling is diagonal, ordinal, and uniform. (As discussed in [38], one can use either an adjacency tensor or an adjacency matrix to represent a multilayer network.) The generalization in [50] consists of applying the function in (2.16) to the $N|\mathcal{T}|$-node multilayer network:

$$\hat{r}(C,t) = \sum_{i,j=1}^{N|\mathcal{T}|} \Big(\pi_i\big[\delta_{ij} + t\Lambda_{ii}(\mathcal{M}_{ij} - \delta_{ij})\big] - \pi_i\rho_{i|j}\Big)\delta(c_i,c_j)\,, \tag{3.1}$$

where C is now a multilayer partition (i.e., a partition of an $N|\mathcal{T}|$-node multilayer network), $\boldsymbol{\Lambda}$ is the $N|\mathcal{T}| \times N|\mathcal{T}|$ diagonal matrix with the rates of the exponentially distributed waiting times at each node of each layer on its diagonal, $\boldsymbol{\mathcal{M}}$ (with entries $\mathcal{M}_{ij} := \mathcal{A}_{ij}/\sum_k \mathcal{A}_{ik}$) is the $N|\mathcal{T}| \times N|\mathcal{T}|$ transition matrix for the $N|\mathcal{T}|$-node multilayer network with adjacency matrix $\boldsymbol{\mathcal{A}}$, π_i is the corresponding stationary distribution (with the strength of a node and the total edge weight now computed from the multilayer adjacency matrix $\boldsymbol{\mathcal{A}}$), and $\rho_{i|j}$ is the probability of jumping from node i to node j at stationarity in one step conditional on the structure of the network within and between layers. The authors' choice of $\rho_{i|j}$, which accounts for the "sparsity pattern"[10] of interlayer edges in the multilayer adjacency matrix, motivates the *multilayer modularity-maximization problem*

$$\max_{C\in\mathcal{C}} \sum_{i,j=1}^{N|\mathcal{T}|} \mathcal{B}_{ij}\delta(c_i,c_j)\,, \tag{3.2}$$

[10]The *sparsity pattern* of a matrix $\boldsymbol{X}$ is a matrix $\boldsymbol{Y}$ with entries $Y_{ij} = 1$ when $X_{ij} \neq 0$, and $Y_{ij} = 0$ when $X_{ij} = 0$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

which we can also write as $\max_{C\in\mathcal{C}} Q(C|\boldsymbol{\mathcal{B}})$, where $\boldsymbol{\mathcal{B}}$ is the *multilayer modularity matrix*

$$\boldsymbol{\mathcal{B}} = \begin{bmatrix} \boldsymbol{B}_1 & \omega\boldsymbol{I} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \omega\boldsymbol{I} & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \omega\boldsymbol{I} \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \omega\boldsymbol{I} & \boldsymbol{B}_{|\mathcal{T}|} \end{bmatrix}, \tag{3.3}$$

and $\boldsymbol{B}_s$ is a single-layer modularity matrix computed on layer s. (For example, $\boldsymbol{B}_s = \boldsymbol{A}_s - \langle \boldsymbol{A}_s \rangle \boldsymbol{1}_N$ if one uses the U null network and sets $\gamma = 1$.) We rewrite the multilayer modularity-maximization problem in [50] as

$$\max_{C\in\mathcal{C}} \left[\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs}\delta(c_{i_s}, c_{j_s}) + 2\omega \sum_{s=1}^{|\mathcal{T}|-1} \sum_{i=1}^{N} \delta(c_{i_s}, c_{i_{s+1}}) \right], \tag{3.4}$$

where B_{ijs} denotes the (i,j)th entry of $\boldsymbol{B}_s$. Equation (3.4) clearly separates intralayer contributions (left term) from interlayer contributions (right term) to the multilayer quality function.

In practice, one can solve this multilayer modularity-maximization problem with the Louvain heuristic in subsection 2.2 by using the multilayer modularity matrix $\boldsymbol{\mathcal{B}}$ instead of the single-layer modularity matrix $\boldsymbol{B}$ as an input (the number of nodes in the first iteration of phase 1 becomes $N|\mathcal{T}|$ instead of N). In this case, the initial partition consists of $N|\mathcal{T}|$ singletons. One first places each of the $N|\mathcal{T}|$ nodes into a set (possibly where it already is) that results in the largest increase of multilayer modularity. We then iterate this procedure on a reduced network (as defined in subsection 2.2) until the heuristic converges. It is clear from (3.4) that placing nodes from different layers into the same set, which we call an interlayer *merge*, decreases the value of the multilayer quality function when $\omega < 0$, so we consider only $\omega \geq 0$. As with single-layer networks, we assume that each of the partitions in the set $\mathcal{C}$ of $N|\mathcal{T}|$-node partitions contains sets that do not have multiple connected components in the graph with adjacency matrix $\boldsymbol{\mathcal{B}}$.

In section 5, we try to gain some insight into how to interpret a globally optimal multilayer partition by proving several properties that it satisfies. The results that we show hold for *any* choice of matrices $\boldsymbol{B}_1, \dots, \boldsymbol{B}_{|\mathcal{T}|}$, so (for example) they still apply when one uses the stability quality function in (2.12) on each layer instead of the modularity quality function. For ease of writing (and because modularity is the quality function that we use in our computational experiments of section 5), we will continue to refer to the maximization problem (3.4) as a multilayer modularity-maximization problem.

4. Interpretation of community structure in correlation networks with different null networks. It is clear from the structure of $\boldsymbol{\mathcal{B}}$ in (3.3) that the choice of quality function within layers (i.e., diagonal blocks in the multilayer modularity matrix) and the choice of coupling between layers (i.e., off-diagonal blocks) for a given quality function affect the solution of the maximization problem in (3.4). In this section, we make some observations on the choice of null network for correlation networks when using the modularity quality function. To do this, we consider the multilayer modularity-maximization problem (3.4) with zero interlayer coupling (i.e.,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$\omega = 0$), which is equivalent to performing single-layer modularity maximization on each layer independently.

4.1. Toy examples. We describe two simple toy networks to illustrate some features of the NG (2.8) and NGS (2.13) null networks that can be misleading for asset correlation networks.

4.1.1. NG null network. Assume that the nodes in a network are divided into K nonoverlapping categories (e.g., asset classes) such that all intracategory edge weights have a constant value $a > 0$, and all intercategory edge weights have a constant value b, with $0 \leq b < a$. Let κ_i denote the category of node i, and rewrite the strength of node i as

$$k_i = |\kappa_i|a + (N - |\kappa_i|)b = |\kappa_i|(a - b) + Nb\,.$$

The strength of a node in this network scales affinely with the number of nodes in its category. Suppose that we have two categories κ_1, κ_2 that do not contain the same number of nodes. Taking $|\kappa_1| > |\kappa_2|$ without loss of generality, it follows that

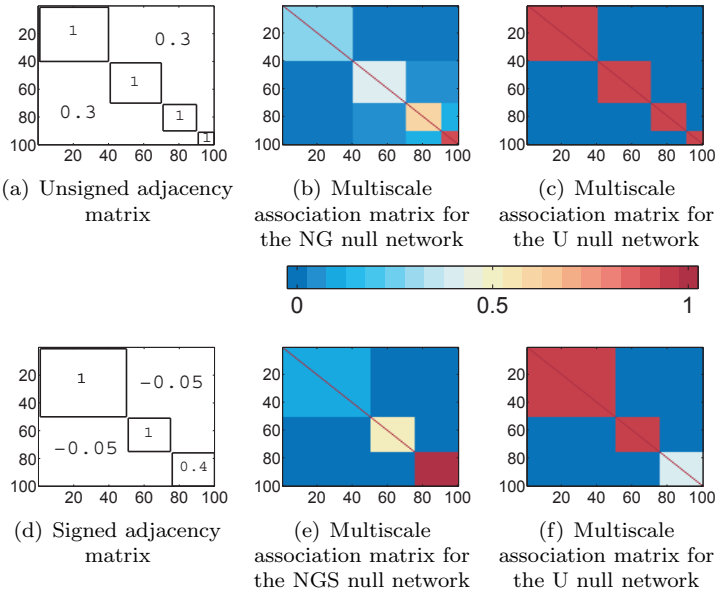
$$P_{i,j\in\kappa_1} = \frac{1}{2m}\Big[|\kappa_1|(a-b) + Nb\Big]^2 > \frac{1}{2m}\Big[|\kappa_2|(a-b) + Nb\Big]^2 = P_{i,j\in\kappa_2}\,, \tag{4.1}$$

where $P_{i,j\in\kappa_i}$ is the expected edge weight between pairs of nodes in κ_i in the NG null network. That is, pairs of nodes in an NG null network that belong to larger categories have a larger expected edge weight than pairs of nodes that belong to smaller categories.

To see how (4.1) can lead to misleading results, we perform a simple experiment. Consider the toy network in Figure 2(a) that contains 100 nodes divided into four categories of sizes 40, 30, 20, and 10. We set intracategory edge weights to 1 and intercategory edge weights to 0.3 (i.e., $a = 1$ and $b = 0.3$ in (4.1)). In Figure 2(b) (respectively, Figure 2(c)), we show the multiscale association matrix defined in (2.7) using an NG null network (respectively, a U null network). Colors scale with the frequency of co-classification of pairs of nodes into the same community across resolution-parameter values. Because the nodes are ordered by category, diagonal blocks in Figures 2(b) and (c) indicate the co-classification index of nodes in the same category, and off-diagonal blocks indicate the co-classification index of nodes in different categories. We observe in Figure 2(b) that larger categories are identified as a community across a smaller range of resolution-parameter values than smaller categories when using an NG null network. In particular, category κ is identified as a single community when $\gamma < a/P_{i,j\in\kappa}$ (with $a/P_{i,j\in\kappa_1} < a/P_{i,j\in\kappa_2}$ when $|\kappa_1| > |\kappa_2|$ by (4.1)). When $\gamma \geq a/P_{i,j\in\kappa}$, category κ is identified as $|\kappa|$ singleton communities. However, we observe in Figure 2(c) that all four categories are identified as a single community across the same range of resolution-parameter values when using the U null network. In particular, category κ is identified as a single community when $\gamma < a/\langle \boldsymbol{A}\rangle$ and as $|\kappa|$ singleton communities when $\gamma \geq a/\langle \boldsymbol{A}\rangle$.

The standard interpretation of multiscale modularity maximization is that the communities that one obtains for larger values of γ reveal "smaller" and "more densely" connected nodes in the observed network [40, 64]. Although all diagonal blocks in Figure 2(a) have the same internal connectivity, different ones are identified as communities for different values of γ when using the NG null network—as γ increases, nodes in the largest category split into singletons first, followed by those in the second largest category, etc. This example illustrates that one needs to be cautious

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) Unsigned adjacency matrix (b) Multiscale association matrix for the NG null network (c) Multiscale association matrix for the U null network

(d) Signed adjacency matrix (e) Multiscale association matrix for the NGS null network (f) Multiscale association matrix for the U null network

FIG. 2. (a) *Toy unsigned block matrix with constant diagonal and off-diagonal blocks that take the value indicated in the block.* (b) *Multiscale association matrix of* (a) *that gives the frequency of co-classification of nodes across resolution-parameter values using an NG null network.* (c) *Multiscale association matrix of* (a) *that uses a U null network.* (d) *Toy signed block matrix with constant diagonal and off-diagonal blocks that take the value indicated in the block.* (e) *Multiscale association matrix of (d) that uses an NGS null network.* (f) *Multiscale association matrix of* (d) *that uses a U null network. For the NG and U (respectively, NGS) null networks, our sample of resolution-parameter values is the set* $\{\gamma^-, \ldots, \gamma^+\}$ *(respectively,* $\{0, \ldots, \gamma^+\}$*) with a discretization step of* 10^{-3} *between each pair of consecutive values.*

when using multiscale community structure to estimate the strength of connectivity patterns in an observed network.

4.1.2. NGS null network. A key difference between an NG null network (2.8) and an NGS null network (2.13) is that the expected edge weight between two nodes must be positive in the former but can be negative in the latter. Consider a signed variant of the example in section 4.1.1 in which intracategory edge weights equal a constant $a > 0$, and intercategory edge weights equal a constant $b < 0$. The strengths of node i in the κth category are

$$k_i^+ = |\kappa| a \quad \text{and} \quad k_i^- = (N - |\kappa|) b \,.$$

We consider two categories κ_1, κ_2 with different numbers of nodes. Taking $|\kappa_1| > |\kappa_2|$ without loss of generality, it follows that

$$\begin{aligned} P_{i,j\in\kappa_1} &= \frac{1}{2m^+}\Big(|\kappa_1| a\Big)^2 - \frac{1}{2m^-}\Big[(N - |\kappa_1|) b\Big]^2 \\ &> \frac{1}{2m^+}\Big(|\kappa_2| a\Big)^2 - \frac{1}{2m^-}\Big[(N - |\kappa_2|) b\Big]^2 = P_{i,j\in\kappa_2} \,, \end{aligned}$$

where $P_{i,j\in\kappa_i}$ is the expected edge weight between pairs of nodes in κ_i in the NGS null network. As was the case for an NG null network, pairs of nodes in an NGS

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

null network that belong to larger categories have a larger expected edge weight than pairs of nodes that belong to smaller categories.

However, the fact that the expected edge weight can be negative can further complicate interpretations of multiscale community structure. A category κ for which $P_{i,j\in\kappa} < 0$ and $P_{i\in\kappa,j\notin\kappa} \geq 0$ is identified as a community when $-A_{ij} < -\gamma P_{ij}$ for all $i, j \in \kappa$ (this inequality must hold for sufficiently large γ because $P_{i,j\in\kappa} < 0$) and does not split further for larger values of γ. This poses a particular problem in the interpretation of multiscale community structure obtained with the NGS null network because nodes with negative expected edge weights do not need to be "densely connected" in the observed network to contribute positively to modularity. In fact, if one relaxes the assumption of uniform edge weights across categories, one can ensure that nodes in the category with *lowest* intracategory edge weight will *never* split. This is counterintuitive to standard interpretations of multiscale community structure [40].

In Figures 2(d) and (e), we illustrate the above feature of the NGS null network using a simple example. The toy network in Figure 2(d) contains 100 nodes divided into three categories: one of size 50 and two of size 25. The category of size 50 and one category of size 25 have an intracategory edge weight of 1 between each pair of nodes. The other category of size 25 has an intracategory edge weight of 0.4 between each pair of nodes. All intercategory edges have weights of -0.05. (We choose these values so that the intracategory expected edge weight is negative for the third category but positive for the first two, and so that intercategory expected edge weights are positive.) We observe in Figure 2(e) that the first and second categories split into singletons for sufficiently large γ, that the smaller of the two categories splits into singletons for a larger value of the resolution parameter, and that the third category never splits. We repeat the same experiment with the U null network in Figure 2(f) (after a linear shift of the adjacency matrix to the interval $[0, 1]$ using $A_{ij} \mapsto \frac{1}{2}(A_{ij} + 1)$ for all i and j), and we observe that the co-classification index of nodes reflects the value of the edge weight between them. It is largest for pairs of nodes in the first and second categories, and it is smallest for pairs of nodes in the third category.

4.2. Data sets. We illustrate how the features that we discussed in section 4.1 can occur in real data. We use two data sets of financial time series for our computational experiments.

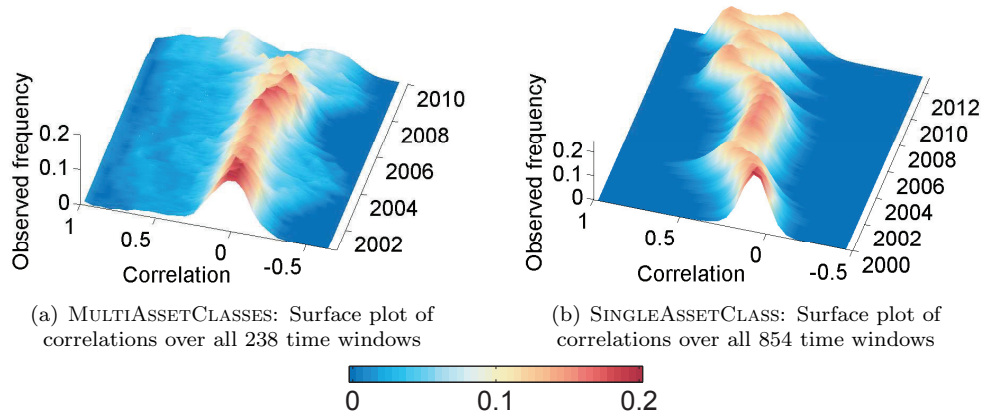
The first data set, which we call MultiAssetClasses, has multiple types of assets and consists of weekly price time series for $N = 98$ financial assets during the time period 01 Jan 99–01 Jan 10 (resulting in 574 prices for each asset). The assets are divided into seven asset classes: 20 government bond indices (Gov.), 4 corporate bond indices (Corp.), 28 equity indices (Equ.), 15 currencies (Cur.), 9 metals (Met.), 4 fuel commodities (Fue.), and 18 commodities (Com.). This data set was studied in [23] using principal component analysis, and a detailed description of the financial assets can be found in that paper.

The second data set, which we call SingleAssetClass, consists of daily price time series for $N = 859$ financial assets from the Standard & Poor's (S&P) 1500 index during the time period 01 Jan 99–01 Jan 13 (resulting in 3673 prices for each asset).[11] The financial assets are all equities and are divided into 10 sectors: 64 materials, 141 industrials, 150 financials, 142 information technology, 55 utilities, 47 consumer staples, 138 consumer discretionary, 48 energy, 68 health care, and 6 telecommunication services.

[11] We consider fewer than 1500 nodes because we include only nodes for which data is available at all time points to avoid issues associated with choices of data-cleaning techniques.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy



(a) MultiAssetClasses: Surface plot of correlations over all 238 time windows

(b) SingleAssetClass: Surface plot of correlations over all 854 time windows

Fig. 3. *Surface plots of the correlations over all time windows for* (a) MultiAssetClasses *data set and* (b) SingleAssetClass *data set. The colors in each panel scale with the value of the observed frequency (color available online).*

The precise way in which one chooses to compute a measure of similarity between pairs of time series and the subsequent choices that one makes (e.g., uniform or nonuniform window length, and overlap or no overlap if one uses a *rolling time window*) affect the values of the similarity measure. There are myriad ways to define similarity measures—the best choices depend on facets such as application domain, time-series resolution, and so on—and this is an active and contentious area of research [62, 67, 69, 76]. Constructing a similarity matrix from a set of time series and investigating community structure in a given similarity matrix are separate problems, and we are concerned with the latter in the present paper. Accordingly, in all of our experiments, we use Pearson correlation coefficients for our measure of similarity. We compute them using a rolling time window with a uniform window length and uniform amount of overlap.

We adopt the same network representation for both data sets. We use the term *time window* for a set of discrete time points and divide each time series into overlapping time windows that we denote by $\mathcal{T} = \{T_s\}$. The length of each time window $|T|$ and the amount of overlap between consecutive time windows $|T| - \delta t$ are uniform. The amount of overlap determines the number of data points that one adds and removes from each time window. It thus determines the number of data points that can alter the connectivity patterns in each subsequent correlation matrix (i.e., each subsequent layer). We fix $(|T|, \delta t) = (100, 2)$ for the MultiAssetClasses data set (which amounts to roughly two years of data in each time window) and $(|T|, \delta t) = (260, 4)$ for the SingleAssetClass data set (which amounts to roughly one year of data in each time window). Every network layer with adjacency matrix $\boldsymbol{A}_s$ is a Pearson correlation matrix between the time series of logarithmic returns during the time window T_s. We take correlations between logarithmic returns because it is standard practice [16], but one can also examine correlations between other quantities (such as arithmetic returns [30]). For each data set, we study the sequence of matrices

$$\left\{\boldsymbol{A}_s \in [-1,1]^{N\times N} | s \in \{1,\ldots,|\mathcal{T}|\}\right\} .$$

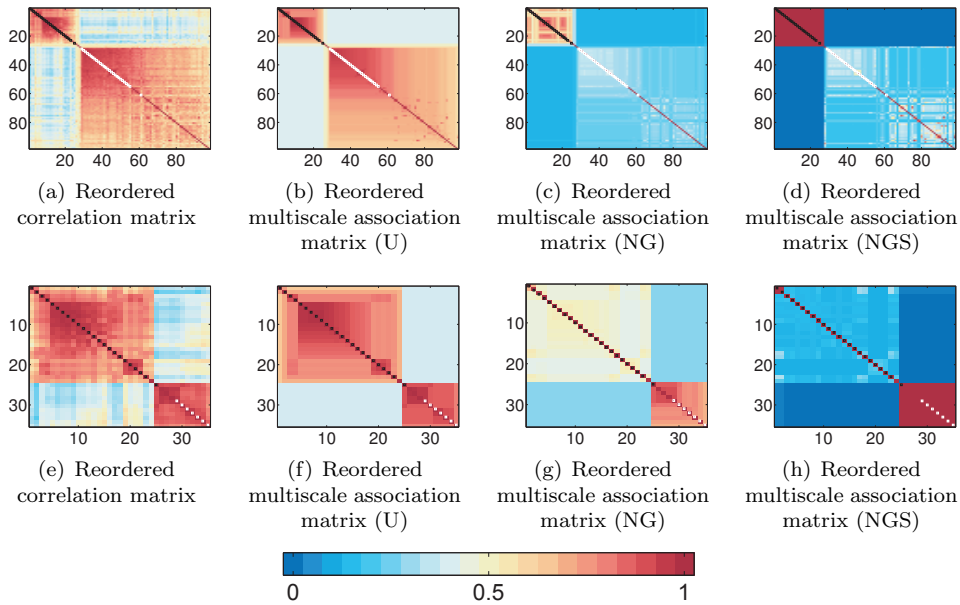
We show a surface plot of the observed frequency of correlations in each layer for each data set in Figure 3.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

4.3. Multiscale community structure in asset correlation networks. We perform the same experiments as in Figure 2 on the correlation matrices of both data sets. Our resolution-parameter sample is the set $\{\gamma^-, \ldots, \gamma^+\}$ (respectively, $\{0, \ldots, \gamma^+\}$) for the U and NG (respectively, NGS) null networks with a discretization step of the order of 10^{-3}. We store the co-classification index of pairs of nodes averaged over all resolution-parameter values in the sample. We use the U and NG null networks for a correlation matrix that is linearly shifted to the interval $[0, 1]$. For each null network, we thereby produce $|\mathcal{T}|$ multiscale association matrices with entries between 0 and 1 that indicate how often pairs of nodes are in the same community across resolution-parameter values.

We show the multiscale association matrices for a specific layer of MultiAsset-Classes in Figure 4. The matrix in Figure 4(a) corresponds to the correlation matrix during the interval 08 Feb 08–01 Dec 10. In accord with the results in [23], this matrix reflects the increase in correlation between financial assets that took place after the Lehman bankruptcy in 2008 compared to correlation matrices that we compute from earlier time periods. (One can also see this feature in the surface plot of Figure 3(a).) The matrices in Figures 4(b), (c), and (d) correspond, respectively, to the multiscale association matrix for the U, NG, and NGS null networks. We reorder all matrices (identically) using a node ordering based on the partitions that we obtain with the U null network that emphasizes block-diagonal structure in the correlation matrix.



(a) Reordered correlation matrix
(b) Reordered multiscale association matrix (U)
(c) Reordered multiscale association matrix (NG)
(d) Reordered multiscale association matrix (NGS)
(e) Reordered correlation matrix
(f) Reordered multiscale association matrix (U)
(g) Reordered multiscale association matrix (NG)
(h) Reordered multiscale association matrix (NGS)

FIG. 4. *Multiscale association matrix for the U, NG, and NGS null networks for the entire correlation matrix and a subset of the correlation matrix in the last layer of the* MultiAssetClasses *data set. In panel* (a)*, we show the entire matrix; in panels* (b), (c), (d)*, we show the multiscale association matrix that we obtain from this matrix using each of the three null networks. In panel* (e)*, we show the first* 35×35 *block of the correlation matrix from panel* (a)*; in panels* (f), (g), (h)*, we show the multiscale association matrix that we obtain from this subset of the correlation matrix using each of the three null networks. The colors scale with the entries of the multiscale association and the entries of the correlation matrix. Black squares on the diagonals correspond to government and corporate bond assets, and white squares correspond to equity assets. (Color available online.)*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

We observe that the co-classification indices in the multiscale association matrix of Figure 4(b) are a better reflection of the strength of correlation between assets in Figure 4(a) than the multiscale association matrices in Figures 4(c) and (d). As indicated by the darker shades of red in the upper left corner in Figures 4(c) and (d), we also observe that the government and corporate bond assets (which we represent with black squares on the diagonal) are in the same community for a larger range of resolution-parameter values than the range for which equity assets (which we represent with white squares on the diagonal) are in the same community. In fact, when we use an NGS null network, the expected weight between two government or corporate bonds is negative (it is roughly -0.1), and these assets are in the same community for arbitrarily large values of the resolution parameter. (In other words, they do not split into smaller communities for large γ.) One needs to be cautious when using the multiscale association matrices in Figures 4(c) and (d) to gain insight about the strength of connectivity between assets in Figure 4(a).

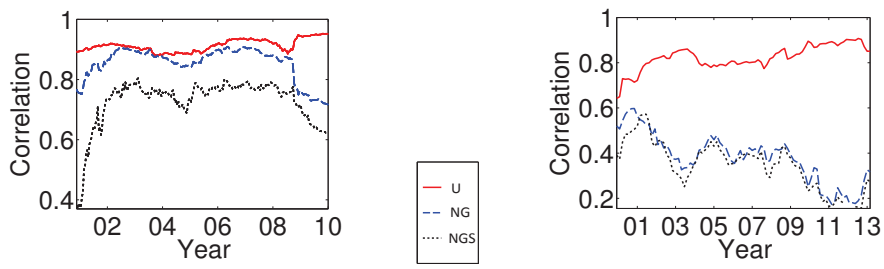
When studying correlation matrices of multiasset data sets, one may wish to vary the size of the asset classes included in the data (e.g., by varying the ratio of equity and bond assets). We show how doing this can lead to further misleading conclusions. By repeating the same experiment using only a subset of the correlation matrix (the first 35 nodes), we consider an example where we have inverted the relative sizes of the bond asset class and the equity asset class. As indicated by the darker shades of red in the lower right corner in Figures 4(g) and (h), equity assets now have a larger co-classification index than government and corporate bond assets when using the NG or NGS null networks. If one uses the co-classification index in the multiscale association matrices of Figures 4(c) and (d) (respectively, Figures 4(g) and (h)) to estimate the values of observed correlation between equity and bond assets in Figure 4(a) (respectively, Figure 4(e)), one can draw different conclusions despite the fact that these values have not changed. However, the multiscale association matrix with a U null network in Figure 4(f) reflects the observed correlation between equity and bond assets in Figure 4(e).[12]

To quantify the sense in which a multiscale association matrix of one null network "reflects" the values in the correlation matrix, we compute the Pearson correlation between the upper triangular part of each multiscale association matrix and its corresponding adjacency matrix across all time layers of both data sets for the U, NG, and NGS null networks. We show these correlation plots in Figure 5. Observe that the correlation between the adjacency and multiscale association matrix in Figures 5(a) and (b) is strongest in each layer for the U null network and weakest in (almost) each layer for the NGS null network.

The above observation can be explained as follows. Recall from (2.5) that we can write the modularity-maximization problem as $\max_{\boldsymbol{S}\in\mathcal{S}} \mathrm{Tr}(\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{B}\boldsymbol{S})$, where $\mathcal{S}$ is the set of partition matrices. When one uses a U null network, the entries of the modularity matrix are the entries of the adjacency matrix shifted by a constant $\gamma\langle\boldsymbol{A}\rangle$,

[12]The authors of [72] showed that a globally optimal partition for a null network called the "constant Potts model" (CPM), in which the edge weights are given by a constant that is *independent* of the network, is "sample-independent." Their result can be generalized as follows for the U null network (in which expected edge weights are constant but are not independent of the observed network). Suppose that $C_{\max}$ is a partition that maximizes $Q(C|\boldsymbol{A};\boldsymbol{P};\gamma_1)$, and consider the subgraph induced by the network on a set of communities $C_1,\ldots,C_l \in C_{\max}$. It then follows that $\{C_1 \cup C_2 \ldots \cup C_l\}$ maximizes $Q(C|\tilde{\boldsymbol{A}};\boldsymbol{P};\gamma_2)$, where $\tilde{\boldsymbol{A}}$ is the adjacency matrix of the induced subgraph and $\gamma_2 = \gamma_1\langle\boldsymbol{A}\rangle/\langle\tilde{\boldsymbol{A}}\rangle$. For the CPM null network, the same result holds with $\gamma_1 = \gamma_2$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) Correlation between multiscale association matrix and adjacency matrix for the MultiAssetClasses data set

(b) Correlation between multiscale association matrix and adjacency matrix for the SingleAssetClass data set

Fig. 5. *Correlation between the adjacency matrix and the multiscale association matrix for the U (solid curve), NG (dashed curve), and NGS (dotted curve) null networks over all time layers for* (a) *the* MultiAssetClasses *data set and* (b) *the* SingleAssetClass *data set. We compute the Pearson correlation coefficients between entries in the upper diagonal blocks in each matrix (to avoid double counting, as the matrices are symmetric), and we exclude diagonal entries (which, by construction, are equal to* 1 *in both matrices).*

and the quality function reduces to

$$\max_{\boldsymbol{S}\in\mathcal{S}} \left[\mathrm{Tr}(\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{S}) - \gamma\langle\boldsymbol{A}\rangle||c(\boldsymbol{S})||_2\right] , \tag{4.2}$$

where $||c(\boldsymbol{S})||_2 = ||\mathrm{Tr}(\boldsymbol{S}^{\boldsymbol{T}}\mathbf{1}_N\boldsymbol{S})||_2$ is the 2-norm of the vector of set sizes in $\boldsymbol{S}$ (i.e., $c(\boldsymbol{S})$ is the vector whose kth entry is $\sum_{i=1}^{N} S_{ik}$). It follows that modularity maximization with a U null network is equivalent to a block-diagonalization of the adjacency matrix $\boldsymbol{A}$ (the first term in (4.2)) with a penalty on the size of communities (the second term). As one increases the resolution parameter, one favors smaller sets of nodes with stronger internal connectivity. Note that one could also apply (4.2) on adjusted adjacency matrices $\boldsymbol{A'} = \boldsymbol{A} - \tilde{\boldsymbol{A}}$. For example, one can let $\tilde{\boldsymbol{A}}$ be a matrix that controls for random fluctuations in a correlation matrix $\boldsymbol{A}$ (e.g., the "random component" $\boldsymbol{C}^r$ in [46]).

For a general null network, (4.2) takes the form

$$\max_{\boldsymbol{S}\in\mathcal{S}} \left[\mathrm{Tr}(\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{S}) - \mathrm{Tr}(\boldsymbol{S}^{\boldsymbol{T}}(\gamma\boldsymbol{P})\boldsymbol{S})\right] ,$$

where $\boldsymbol{P}$ is the adjacency matrix of the null network. That is, modularity maximization finds block-diagonal structure in $\boldsymbol{A}$ (first term) that is not in $\gamma\boldsymbol{P}$ (second term). It is common to avoid using the U null network in applications because "it is not a good representation of most real-world networks" [53]. The extent to which one wants a null network to be a good representation of an observed network depends on the features that one wants to take as given. We argue that whether an NG null network is more appropriate than a U null network for a given situation depends at least in part on one's interpretation of node strength for that application. As we discussed in section 2.4, the strength of a node in correlation matrices is given by the covariance between its standardized time series and the mean time series. When using the NG null network, it thus follows that pairwise differences $B_{ij} - B_{i'j'}$ in the modularity quality function depend on $\mathrm{corr}(\hat{z}_i, \hat{z}_j), \mathrm{corr}(\hat{z}_{i'}, \hat{z}_{j'})$, and $\mathrm{corr}(\hat{z}_k, \hat{z}_{\mathrm{tot}})$, where $k \in \{i, j, i', j'\}$, the quantity $\hat{z}_i$ is the standardized time series of asset i defined in subsection 2.4.1,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

and $\hat{z}_{\text{tot}} = \sum_{i=1}^{N} \hat{z}_i$. When using the U null network, pairwise differences in the modularity quality function depend only on the observed edge weights $\text{corr}(\hat{z}_i, \hat{z}_j)$ and $\text{corr}(\hat{z}_{i'}, \hat{z}_{j'})$. The term $\text{corr}(\hat{z}_k, \hat{z}_{\text{tot}})$ introduces a dependence between the communities that one finds using the NG null network and the extent to which nodes in those communities are representative of the mean time series for the sample (as measured by $\text{corr}(\hat{z}_k, \hat{z}_{\text{tot}})$). In situations where one may wish to vary one's node sample (e.g., by changing the size of asset classes), one needs to bear such dependencies in mind when interpreting the communities that one obtains.

5. Effect of interlayer coupling on a multilayer partition. In section 4, we set the interlayer connection weights to 0 in the multilayer network. The solution to the multilayer modularity-maximization problem (3.4) then depends solely on the values in the modularity matrix of each time layer, and the multilayer modularity-maximization problem reduces to performing single-layer modularity maximization on each layer independently.

Recall the multilayer modularity-maximization problem

$$\max_{C \in \mathcal{C}} \left[\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs} \delta(c_{i_s}, c_{j_s}) + 2\omega \sum_{s=1}^{|\mathcal{T}|-1} \sum_{i=1}^{N} \delta(c_{i_s}, c_{i_{s+1}}) \right].$$

A solution to this problem is a partition of an $N|\mathcal{T}|$-node multilayer network. Its communities can contain nodes from the same layer and nodes from different layers. Nodes from different layers can be the same node at different times ((i_s, i_r) with $s \neq r$) or different nodes at different times ((i_s, j_r) with $i \neq j$ and $s \neq r$). We say that a node i_s remains in the same community (respectively, changes communities) between consecutive layers s and $s+1$ if $\delta(c_{i_s}, c_{i_{s+1}}) = 1$ (respectively, $\delta(c_{i_s}, c_{i_{s+1}}) = 0$).

Positive diagonal, ordinal, and uniform interlayer connections favor nodes remaining in the same community between consecutive layers. Every time a node does not change communities between two consecutive layers (i.e., $\delta(c_{i_s}, c_{i_{s+1}}) = 1$), a positive contribution of 2ω is added to the multilayer quality function. One thereby favors communities that do not to change in time because community assignments are transitive: if $\delta(c_{i_s}, c_{j_s}) = 1$ and $\delta(c_{i_s}, c_{i_{s+1}}) = \delta(c_{j_s}, c_{j_{s+1}}) = 1$, then $\delta(c_{i_{s+1}}, c_{j_{s+1}}) = 1$. We define the *persistence* of a multilayer partition to be the total number of nodes that do not change communities between layers:

$$\text{Pers}(C) := \sum_{s=1}^{|\mathcal{T}|-1} \sum_{i=1}^{N} \delta(c_{i_s}, c_{i_{s+1}}) \in \{0, \ldots, N(|\mathcal{T}|-1)\}. \tag{5.1}$$

As indicated in (5.1), $\text{Pers}(C)$ is an integer between 0, which occurs when no node ever remains in the same community across layers, and $N(|\mathcal{T}|-1)$, which occurs when every node always remains in the same community. (See [6] for a closely related measure called "flexibility" that has been applied to functional brain networks.) Let $\text{Pers}(C)|_s$ denote the number of nodes that remain in the same community between two consecutive layers s and $s+1$:

$$\text{Pers}(C)|_s := \sum_{i=1}^{N} \delta(c_{i_s}, c_{i_{s+1}}) \in \{0, \ldots, N\}, \tag{5.2}$$

so that $\text{Pers}(C) = \sum_{s=1}^{|\mathcal{T}|-1} \text{Pers}(C)|_s$. Persistence provides an insightful way of rewrit-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

ing the multilayer modularity-maximization problem:

$$\max_{C\in\mathcal{C}}\left[\sum_{s=1}^{|\mathcal{T}|}\sum_{i,j=1}^{N}B_{ijs}\delta(c_{i_s},c_{j_s})+2\omega\mathrm{Pers}(C)\right]. \tag{5.3}$$

The multilayer maximization problem thus measures a trade-off between static community structure within layers (the first term in (5.3)) and temporal persistence across layers (the second term in (5.3)).

To better understand how the partitions that one obtains with nonzero interlayer coupling (i.e., using $\omega > 0$) can differ from the partitions that one obtains without interlayer coupling (i.e., using $\omega = 0$), we define notation that helps to compare a multilayer partition to single-layer partitions. Let $\mathcal{N}_s := \{1_s, \ldots, N_s\}$ denote the set of nodes in layer s. The restriction of a set $C_l \subseteq \{1_1, \ldots, N_1; 1_2, \ldots, N_2; \ldots; 1_{|\mathcal{T}|}, \ldots, N_{|\mathcal{T}|}\}$ of nodes to a layer s is $C_l|_s := C_l \cap \mathcal{N}_s$, and we define the *partition induced by a multilayer partition* $C \in \mathcal{C}$ on layer s by

$$C|_s := \{C_l|_s, C_l \in C\}\,.$$

We refer to a "globally optimal partition" as an "optimal partition" in this section for ease of writing, and we refer to $\sum_{s=1}^{|\mathcal{T}|}\sum_{i,j=1}^{N}B_{ijs}\delta(c_{i_s},c_{j_s})$ (i.e., the first term in (5.3)) as *intralayer modularity*. In the next two subsections, we illustrate how the set of partitions induced by a multilayer partition with $\omega > 0$ on individual layers can differ from intralayer partitions obtained with $\omega = 0$.

5.1. Toy examples.

5.1.1. Changes in connectivity patterns. This toy example illustrates how interlayer coupling can enable us to detect and differentiate between changes in connectivity patterns across layers. In Figure 6, we show an unweighted multilayer network with $|\mathcal{T}| = 10$ layers and $N = 8$ nodes in each layer. Every layer except for layers 3 and 6 contains two 4-node cliques. In layer 3, node 5_3 is connected to nodes $\{1_3, 2_3\}$ instead of nodes $\{6_3, 7_3, 8_3\}$. In layer 6, node 5_6 is connected to nodes $\{1_6, 2_6, 3_6, 4_6\}$ instead of nodes $\{6_6, 7_6, 8_6\}$. We show the layers of the multilayer network in Figures 6(a)–(c). We examine its communities using a U null network with a resolution parameter $\gamma = 1$. Layer s then has the following single-layer modularity matrix:

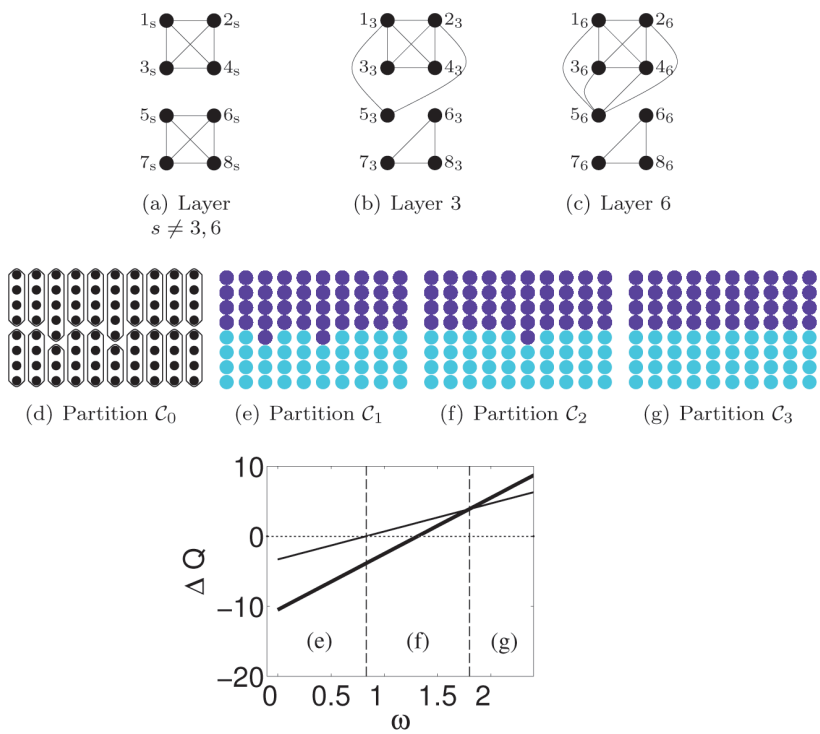
$$B_{ijs} = \begin{cases} 1-\langle \boldsymbol{A}_s\rangle & \text{if } i \text{ is connected to } j,\\ -\langle \boldsymbol{A}_s\rangle & \text{otherwise.}\end{cases}$$

The optimal partition in each layer is unique and is $C_s = \{\{1_s, 2_s, 3_s, 4_s\}, \{5_s, 6_s, 7_s, 8_s\}\}$ in layer s for $s \notin \{3, 6\}$ and $C_s = \{\{1_s, 2_s, 3_s, 4_s, 5_s\}, \{6_s, 7_s, 8_s\}\}$ in layers 3 and 6. When the value of interlayer coupling is 0, the optimal multilayer partition is the union of $|\mathcal{T}|$ disconnected optimal single-layer partitions. The resulting multilayer partition $\mathcal{C}_0 = \bigcup_{i=1}^{10} C_s$, which we show in Figure 6(d), has a persistence of $\mathrm{Pers}(\mathcal{C}_0) = 0$. For any $\omega > 0$, any partition with the same intralayer partitions as $\mathcal{C}_0$ and a nonzero value of persistence yields a higher value of multilayer modularity than $\mathcal{C}_0$. This follows immediately from the expression of the multilayer quality function:

$$Q(C|\boldsymbol{\mathcal{B}}) = \sum_{s=1}^{|\mathcal{T}|}\sum_{i,j=1}^{N}B_{ijs}\delta(c_{i_s},c_{j_s})+2\omega\mathrm{Pers}(C)\,.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy



(a) Layer $s \neq 3, 6$ (b) Layer 3 (c) Layer 6

(d) Partition $\mathcal{C}_0$ (e) Partition $\mathcal{C}_1$ (f) Partition $\mathcal{C}_2$ (g) Partition $\mathcal{C}_3$

(h) Change in multilayer modularity value with respect to partition $\mathcal{C}_1$

FIG. 6. *Toy example illustrating the use of diagonal, ordinal, and uniform interlayer coupling for detecting changes in community structure across layers. We consider* 10 *layers* ($|\mathcal{T}| = 10$) *with eight nodes* ($N = 8$) *in each layer. We show the network structures in* (a) *layers* $s \notin \{3, 6\}$, (b) *layer* 3, *and* (c) *layer* 6. *Panels* (d)–(g) *illustrate four different multilayer partitions. In each panel, the* s*th column of circles represents the nodes in the* s*th layer, which we order from* 1 *to* 8. *We show sets of nodes in the same community using solid curves in panel* (d) *(to avoid having to use* 20 *distinct colors) and using colors in panels* (e)–(g) *(color available online). In panel* (h), *we show the difference between the multilayer modularity value between the partition in panels* (f) *(thin line) and* (g) *(thick line) and the partition in panel* (e) *for different values of* ω. *We include the horizontal dotted line to show the point at which the thin line intercepts the horizontal axis. The panel labels in the regions defined by the area between two consecutive vertical lines in panel* (h) *indicate which of the multilayer partitions in panels* (e), (f), *and* (g) *has a higher value of modularity.*

Increasing persistence without changing intralayer partitions increases the last term of $Q(C|\boldsymbol{\mathcal{B}})$ without changing the other terms. (In section 5.2, we will prove that $\omega > 0$ is both necessary and sufficient for an optimal partition to have a positive value of persistence.) To obtain the multilayer partition in panel (e), we combine all of the sets in panel (d) that contain 1_s into one set, and all of the sets that contain N_s into another set. This partition has a persistence equal to $N(|\mathcal{T}| - 1) - 4$, and any other way of combining the sets in $\mathcal{C}_0$ yields a lower value of persistence.

We now examine Figure 6 further. We consider the multilayer partitions in panels (e)–(g). The example in panel (e) shows the structural changes from both layer 3 (see panel (b)) and layer 6 (see panel (c)), the example in panel (f) shows only the change from layer 6, and panel (g) does not show either change. As we soon quantify in terms of modularity cost, the change in layer 6 is the "stronger" of the two changes. We let

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$\mathcal{C}_1$ denote the multilayer partition in panel (e), $\mathcal{C}_2$ denote the multilayer partition in panel (f), and $\mathcal{C}_3$ denote the multilayer partition in panel (g). Additionally, note that $\text{Pers}(\mathcal{C}_1) < \text{Pers}(\mathcal{C}_2) < \text{Pers}(\mathcal{C}_3)$. The value ω of interlayer coupling determines which partition of these three has the largest value of multilayer modularity. To see this, we compute the modularity cost of changing static community structure within layers in partition $\mathcal{C}_1$ in favor of persistence. (Such a computation is a multilayer version of the calculations for static networks in [29].) The intralayer modularity cost in $\mathcal{C}_1$ of moving node 5_s from community $\{1_s, 2_s, 3_s, 4_s, 5_s\}$ to community $\{6_s, 7_s, 8_s\}$ in layers $s \in \{3, 6\}$ is

$$\begin{aligned}\Delta Q(s) &= 2\left(\sum_{j\in\{6,7,8\}} B_{5js} - \sum_{j\in\{1,2,3,4\}} B_{5js}\right)\\ &= \begin{cases} -4 + 2\langle \boldsymbol{A}_3\rangle \approx -3.3 & \text{if } s = 3,\\ -8 + 2\langle \boldsymbol{A}_6\rangle \approx -7.2 & \text{if } s = 6.\end{cases}\end{aligned}$$

The interlayer modularity cost from this move is $+4\omega$ in both cases; the first $+2\omega$ contribution of the $+4\omega$ follows by symmetry of $\boldsymbol{\mathcal{B}}$, and the second $+2\omega$ contribution of the $+4\omega$ follows from the fact that either move increases persistence by $+2$. Consequently, for $0 < 4\omega < |\Delta Q(3)|$, the partition in panel (e) yields a larger value of multilayer modularity than the partitions in (f) and (g). When $|\Delta Q(3)| < 4\omega < |\Delta Q(6)|$, the multilayer modularity value of the partition in (f) is larger than that of (e) or (g). Finally, when $4\omega > |\Delta Q(6)|$, the partition in panel (g) has the largest value of multilayer modularity of the three. When $4\omega = |\Delta Q(3)|$ (respectively, $4\omega = |\Delta Q(6)|$), the multilayer partition in panels (e) and (f) (respectively, (f) and (g)) have the same value of multilayer modularity. We illustrate these results in Figure 6(h) by plotting $Q(\mathcal{C}_2|\boldsymbol{\mathcal{B}}) - Q(\mathcal{C}_1|\boldsymbol{\mathcal{B}})$ and $Q(\mathcal{C}_3|\boldsymbol{\mathcal{B}}) - Q(\mathcal{C}_1|\boldsymbol{\mathcal{B}})$ against ω. This example is a simple illustration of how interlayer connections can help distinguish between changes in connectivity patterns: stronger changes (in terms of modularity cost) persist across larger values of interlayer coupling (see [3, 59] for other approaches to "change point detection" in temporal networks).

5.1.2. Shared connectivity patterns. In the previous toy example, the intralayer partitions induced on each layer by the multilayer partitions in Figures 6(e), (f), and (g) are optimal for at least one layer when $\omega = 0$ (see Figure 6(d)). This second example illustrates how interlayer coupling can identify intralayer partitions that are not optimal for any individual layer when $\omega = 0$ but which reflect connectivity patterns that are shared across layers.

In Figure 7, we consider an unweighted multilayer network with $|\mathcal{T}| = 3$ layers and $N = 13$ nodes in each layer. Every sth layer contains four 3-node cliques and a node that is connected to each of the three nodes in the sth clique, and to nodes 10_s and 12_s in the 4th clique. We show the layers of the multilayer network in panels (a)–(c). We examine its communities using a U null network with a resolution-parameter value of $\gamma = 1$. The optimal partition in each layer is unique and is $\{\{1_1, 2_1, 3_1, 13_1\}, \{4_1, 5_1, 6_1\}, \{7_1, 8_1, 9_1\}, \{10_1, 11_1, 12_1\}\}$ for layer 1, $\{\{1_2, 2_2, 3_2\}, \{4_2, 5_2, 6_2, 13_2\}, \{7_2, 8_2, 9_2\}, \{10_2, 11_2, 12_2\}\}$ for layer 2, and $\{\{1_3, 2_3, 3_3\}, \{4_3, 5_3, 6_3\}, \{7_3, 8_3, 9_3, 13_3\}, \{10_3, 11_3, 12_3\}\}$ for layer 3. We obtain the multilayer partition $\mathcal{C}_1$ in panel (d) by combining these sets such that induced intralayer partitions are optimal for each layer when $\omega = 0$ and persistence is maximized between layers. The multilayer partition $\mathcal{C}_2$ in panel (e) reflects connectivity patterns that are shared by all layers (i.e., node 13_s is with the fourth 3-node clique instead of the sth 3-node

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

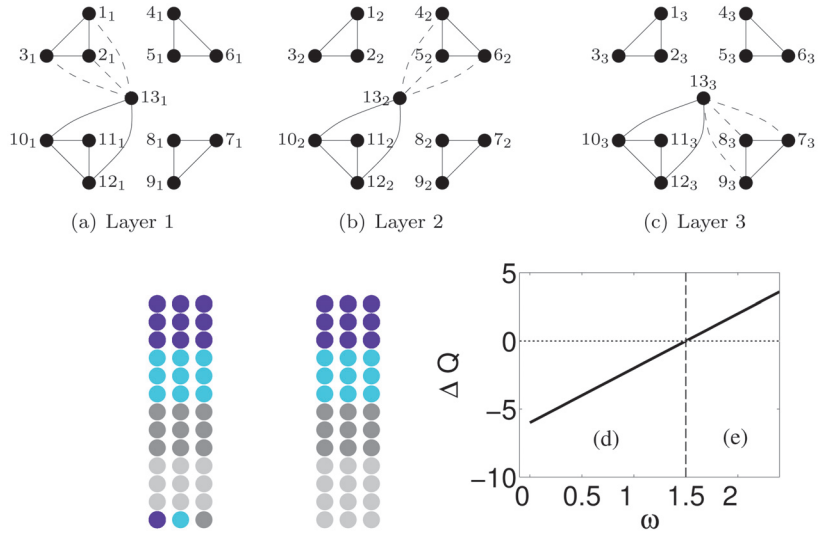


FIG. 7. *Toy example illustrating the use of diagonal, ordinal, and uniform interlayer coupling for detecting shared connectivity patterns across layers. We consider three layers ($|\mathcal{T}| = 3$) with* 13 *nodes ($N = 13$) in each layer. We show the network structures in* (a) *layer* 1, (b) *layer* 2, *and* (c) *layer* 3. *Solid lines represent edges present in all three layers, and dashed lines represent edges that are present only in one of the layers. Panels* (d) *and* (e) *illustrate two different multilayer partitions. In each panel, the sth column of circles represents the nodes in the sth layer, which we order* 1 *to* 13. *We show sets of nodes in the same community using colors in panels* (d) *and* (e) *(color available online). In panel* (f), *we show the difference between the multilayer modularity value between the partition in panel* (e) *and the partition in panel* (d) *for different values of ω. We include the horizontal dotted line to show the point at which the line intercepts the horizontal axis. The panel labels in the regions defined by the area between two consecutive vertical lines in panel* (f) *indicate which of the multilayer partitions in panels* (d) *and* (e) *has a larger value of multilayer modularity.*

clique), but its intralayer partitions are not optimal for any layer when $\omega = 0$. By carrying out calculations similar to those in the previous toy example, one can show that when $\omega > 3/2$,[13] the multilayer partition in panel (e) yields a larger modularity value than the multilayer partition in panel (d). We illustrate this result in Figure 7(f) by plotting $Q(\mathcal{C}_2|\boldsymbol{\mathcal{B}}) - Q(\mathcal{C}_1|\boldsymbol{\mathcal{B}})$ against ω. This example is a simple illustration of how interlayer connections can help identify connectivity patterns that are shared across layers.

5.2. Some properties of multilayer partitions. We now ask how introducing positive diagonal, ordinal, and uniform coupling (i.e., $\omega > 0$) alters the set of maximum-modularity partitions of static networks (i.e., the case $\omega = 0$). To clearly differentiate between intralayer and interlayer modularity contributions, we denote the quality function by

$$Q(C|\boldsymbol{B_1}, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) := \sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs}\delta(c_{i_s}, c_{j_s}) + 2\omega \mathrm{Pers}(C)$$

[13]That is, when $4\omega + 6[2(1 - \langle \boldsymbol{A}_s \rangle) - \langle \boldsymbol{A}_s \rangle - 3(1 - \langle \boldsymbol{A}_s \rangle)] > 0$, with $\langle \boldsymbol{A}_1 \rangle = \langle \boldsymbol{A}_2 \rangle = \langle \boldsymbol{A}_3 \rangle$ by construction in this example.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

instead of $Q(C|\boldsymbol{\mathcal{B}})$, where $C \in \mathcal{C}$ is a multilayer partition. We assume throughout this section that $|\mathcal{T}| \geq 2$. Let $\mathcal{C}_{\max}(\omega)$ denote the set of optimal partitions for the multilayer modularity-maximization problem (5.3), and let $C^{\omega}_{\max}$ be an arbitrary partition in $\mathcal{C}_{\max}(\omega)$. In the discussion that follows, it will be helpful to recall our assumption that each of the partitions in the set $\mathcal{C}$ contains sets that do not have multiple connected components in the weighted graph with adjacency matrix $\boldsymbol{\mathcal{B}}$. In particular, this applies to partitions $C^{\omega}_{\max} \in \mathcal{C}_{\max}(\omega)$. We prove several propositions that hold for an arbitrary choice of the matrices $\boldsymbol{B}_s$ (for example, if one uses the modularity quality function with a U null network and a resolution parameter value of 1, then $\boldsymbol{B}_s = \boldsymbol{A}_s - \langle \boldsymbol{A}_s \rangle \boldsymbol{1}_N$).

PROPOSITION 5.1. $\text{Pers}(C^{\omega}_{\max}) > 0 \Leftrightarrow \omega > 0$.

Proposition 5.1 ensures that as soon as (and only when) the value of ω is strictly positive, the value of persistence of an optimal solution is also positive. To prove this, it suffices to observe that if one rearranges sets in a multilayer partition by combining some of the sets into the same set *without* changing the partitions induced on individual layers, then one only changes the value of persistence in the expression of multilayer modularity. For example, this phenomenon occurs in Figure 6 when going from the partition in panel (d) to the partition in panel (e).

Proof.

$\Rightarrow$: We prove the contrapositive. Assume that $\omega = 0$, and consider a multilayer partition C such that $\text{Pers}(C) > 0$. The partition C contains at least one set with multiple connected components (because $\text{Pers}(C) > 0$ and nodes in different layers are not connected), and C is not optimal by our assumption that global optima do not contain sets with multiple connected components.

$\Leftarrow$: Assume that $\omega > 0$, and consider a multilayer partition C with $\text{Pers}(C) = 0$. We will show that C is not optimal. Let i_r be an arbitrary node in $\{1_1, \ldots, N_1; \ldots; 1_{|\mathcal{T}|}, \ldots, N_{|\mathcal{T}|}\}$, and let C_{i_r} denote the set in C that contains i_r. Let C' be the partition obtained from C by combining all sets that contain i_s, for some s, into one set:

$$C' = \left(C \setminus \bigcup_{s=1}^{|\mathcal{T}|} \{C_{i_s}\} \right) \cup \left\{ \bigcup_{s=1}^{|\mathcal{T}|} C_{i_s} \right\},$$

where C_{i_s} denotes the set in C that contains i_s. Consequently,

$$Q(C'|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) \geq Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) + 2\omega(|\mathcal{T}| - 1),$$

so C is not optimal. (Note that $2\omega(|\mathcal{T}| - 1)$ is strictly positive for $\omega > 0$ because we have assumed that $|\mathcal{T}| \geq 2$.) $\square$

PROPOSITION 5.2. *Let $C^{\omega}_{\max} \in \mathcal{C}_{\max}(\omega)$ and $C_l \in C^{\omega}_{\max}$. The set*

$$\{s \in \{1, \ldots, |\mathcal{T}|\} : C_l|_s \neq \varnothing\}$$

of layers consists of contiguous layers.

Proposition 5.2 ensures that if a community becomes empty in a given layer, then it remains empty in all subsequent layers. We omit the proof, as this result follows directly from the sparsity pattern of $\boldsymbol{\mathcal{B}}$ and our assumption that optimal partitions do not contain sets with multiple connected components in the graph with adjacency matrix $\boldsymbol{\mathcal{B}}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

PROPOSITION 5.3. $C^{\omega}_{\max}|_s = C^{\omega}_{\max}|_{s+1} \Leftrightarrow \mathrm{Pers}(C^{\omega}_{\max})|_s = N$.

Proposition 5.3 connects the notion of persistence between a pair of layers to the notion of change in community structure within layers. Various numerical experiments that have been performed with diagonal, ordinal, and uniform interlayer coupling consist of varying the value of ω and using information about when nodes change communities between layers as an indication of change in community structure within these layers [5, 6, 50]. The equivalence relation in Proposition 5.3 motivates the use of $\mathrm{Pers}(C)|_s$ (or a variant thereof) as an indication of intralayer change in community structure.

Proof.

$\Leftarrow$: The proof of this implication follows straightforwardly by transitivity of community assignments: if $\delta(c_{j_s}, c_{j_{s+1}}) = \delta(c_{i_s}, c_{i_{s+1}}) = 1$ for all i, j, then $\delta(c_{i_s}, c_{j_s}) = 1$ if and only if $\delta(c_{i_{s+1}}, c_{j_{s+1}}) = 1$ for all i, j. (This direction holds for *any* multilayer partition; it need not be optimal.)

$\Rightarrow$: Let $C \in \mathcal{C}$ be a multilayer partition such that $C|_s = C|_{s+1}$ and $\mathrm{Pers}(C)|_s < N$ for some $s \in \{1, \ldots, |\mathcal{T}|\}$. We show that C is not optimal.[14] Consider a set $C_l \in C$ of nodes such that $C_l|_s \neq \varnothing$. If $\delta(c_{i_s}, c_{i_{s+1}}) = 1$ (respectively, $\delta(c_{i_s}, c_{i_{s+1}}) = 0$) for some $i_s \in C_l|_s$, then $\delta(c_{j_s}, c_{j_{s+1}}) = 1$ (respectively, $\delta(c_{j_s}, c_{j_{s+1}}) = 0$) for all $j_s \in C_l|_s$ by transitivity of community assignments and because $C|_s = C|_{s+1}$ by hypothesis. Because $\mathrm{Pers}(C)|_s < N$ by hypothesis, there exists at least one set $C_k|_s$ of nodes (where $C_k \in C$) such that $\delta(c_{i_s}, c_{i_{s+1}}) = 0$ for all $i_s \in C_k|_s$. Let $C_m|_{s+1}$, with $C_m \in C$, denote the set of nodes in layer $s+1$ that contains i_{s+1} for all $i_s \in C_k|_s$. Consider the set $\cup_{r \leq s} C_k|_r$ of nodes in C_k that are in layers $\{1, \ldots, s\}$ and the set $\cup_{r > s} C_m|_r$ of nodes in C_m that are in layers $\{s+1, \ldots, |\mathcal{T}|\}$. Because $\delta(c_{i_s}, c_{i_{s+1}}) = 0$ for all $i_s \in C_k|_s$, it follows by Proposition 5.2 that $C_k = \cup_{r \leq s} C_k|_r$ and $C_m = \cup_{r > s} C_m|_r$. Define the partition C' by

$$C' = \Big(C \setminus \big(\{C_k\} \cup \{C_m\}\big)\Big) \bigcup \Big(\{C_k \cup C_m\}\Big).$$

This partition satisfies $C'|_r = C|_r$ for all $r \in \{1, \ldots, |\mathcal{T}|\}$, $\mathrm{Pers}(C')|_r = \mathrm{Pers}(C)|_r$ for all $r \neq s$, and $\mathrm{Pers}(C')|_s > \mathrm{Pers}(C)|_s$. It follows that $Q(C'|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) > Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega)$ and C is not optimal. ∎

Propositions 5.1, 5.2, and 5.3 apply to an optimal partition obtained with any positive value of ω. The next two propositions concern the existence of "boundary" values for ω.

PROPOSITION 5.4. *There exists $\omega_0 > 0$ such that*

$$\text{if } \omega < \omega_0\,, \text{ then } \bigcup_{s=1}^{|\mathcal{T}|} C^{\omega}_{\max}|_s \in \mathcal{C}_{\max}(0)\,.$$

Moreover, we show that $\omega_0 = \Delta Q / \big[2N(|\mathcal{T}| - 1)\big]$, where ΔQ is the difference between $Q(C^0_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0)$ and the second largest value of $Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0)$ among the partitions of $\mathcal{C}$.

Proposition 5.4 reinforces the idea of thinking of ω as the cost of breaking static community structure within layers in favor of larger values of persistence across layers. It demonstrates that there is a positive value of interlayer coupling such that for any smaller coupling, multilayer modularity maximization only gives more information

[14] Imposing $\mathrm{Pers}(C)|_s = N$ by setting $c_{i_{s+1}} = c_{i_s}$ is not sufficient because changing $\mathrm{Pers}(C)|_s$ locally can change $\mathrm{Pers}(C)|_{s+1}$ or $\mathrm{Pers}(C)|_{s-1}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

than single-layer modularity maximization in that it identifies the set of partitions in $\mathcal{C}_{\max}(0)$ with largest persistence. The proof of this property relies on the fact that the set of possible modularity values for a given modularity matrix is finite.

Proof. Let C be an arbitrary partition such that $\cup_{s=1}^{|\mathcal{T}|} C|_s \notin \mathcal{C}_{\max}(0)$. We will show that there exists a value ω_0 of the interlayer coupling parameter ω such that C is never optimal for any interlayer coupling less than ω_0. Given a sequence of single-layer modularity matrices $\{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}\}$, the set of possible multilayer modularity values for a fixed value of $\omega > 0$ is finite and is given by

$$\mathcal{Q}_\omega = \left\{ Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) | C \in \mathcal{C} \right\} ,$$

where C is a multilayer partition. Let $Q_0^1 = \max \mathcal{Q}_0$, $Q_0^2 = \max \mathcal{Q}_0 \setminus \{Q_0^1\}$, and $\Delta Q = Q_0^1 - Q_0^2 > 0$. By hypothesis,

$$Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) < Q(C_{\max}^0 | \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) ,$$

where $C_{\max}^0 \in \mathcal{C}_{\max}(0)$. Furthermore, by definition of persistence, it follows that

$$Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) \leq Q_0^2 + 2\omega N(|\mathcal{T}| - 1) \tag{5.4}$$

for all values of ω. By choosing $\omega < \omega_0$, with $\omega_0 = \Delta Q / \left[2N(|\mathcal{T}| - 1)\right]$, we obtain

$$Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) \leq Q_0^2 + 2\omega N(|\mathcal{T}| - 1) < Q_0^2 + \Delta Q = Q_1^0 ,$$

so C is not optimal for any interlayer coupling below ω_0. ∎

Clearly, $\omega_0 = \Delta Q / \left[2N(|\mathcal{T}| - 1)\right]$ is not an upper bound for the set $\left\{\omega \in \mathbb{R}^+ : \bigcup_{s=1}^{|\mathcal{T}|} C_{\max}^\omega|_s \in \mathcal{C}_{\max}(0)\right\}$,[15] but our main concern is that the smallest upper bound of this set is *not* 0. (In fact, we have shown that it must be at least as large as $\Delta Q / \left[2N(|\mathcal{T}| - 1)\right] > 0$.)

PROPOSITION 5.5. *There exists $\omega_\infty > 0$ such that*

$$\textit{if } \omega > \omega_\infty , \textit{ then } \mathrm{Pers}(C_{\max}^\omega)|_s = N \textit{ for all } s \in \{1, \ldots, |\mathcal{T}| - 1\} .$$

Moreover, we show that $\omega_\infty = |\mathcal{T}| N^2 \left[\max(\boldsymbol{\mathcal{B}}_{\mathrm{diag}}) - \min(\boldsymbol{\mathcal{B}}_{\mathrm{diag}})\right]/2$, where $\max(\boldsymbol{\mathcal{B}}_{\mathrm{diag}}) = \max_{ijs} B_{ijs}$ and $\min(\boldsymbol{\mathcal{B}}_{\mathrm{diag}}) = \min_{ijs} B_{ijs}$.

Proposition 5.5 implies that a sufficiently large value of interlayer coupling ω guarantees that $C_{\max}^\omega|_s$ remains the same across layers (by Proposition 5.3). The proof of this proposition is similar to the proof of Proposition 5.4.

Proof. Let C be an arbitrary partition of a multilayer network with $\mathrm{Pers}(C)|_s < N$ for some $s \in \{1, \ldots, |\mathcal{T}| - 1\}$. We show that there exists a value $\omega_\infty > 0$ of the interlayer coupling parameter ω such that C is never optimal for $\omega > \omega_\infty$. We first rewrite the quality function as

$$Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) = \beta_1 + 2\omega[N(|\mathcal{T}| - 1) - A] ,$$

where $\beta_1 = \sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs} \delta(c_{i_s}, c_{j_s})$ and $A \geq 1$ because $\mathrm{Pers}(C) < N(|\mathcal{T}| - 1)$ by assumption. Now consider the set of values

$$\boldsymbol{\mathcal{B}}_{\mathrm{diag}} = \left\{ B_{ijs} | i, j \in \{1, \ldots, N\}, s \in \{1, \ldots, |\mathcal{T}|\} \right\} \tag{5.5}$$

[15] For example, one could replace $N(|\mathcal{T}| - 1)$ in (5.4) by $N(|\mathcal{T}| - 1) - \mathrm{Pers}(\mathcal{C}_{\max}(0))$, where $\mathrm{Pers}(\mathcal{C}_{\max}(0))$ denotes the maximum value of persistence that one can obtain by combining sets in each partition of $\mathcal{C}_{\max}(0)$ without changing the partitions induced on individual layers. Proposition 5.4 still holds if one takes $\omega_0 = \Delta Q / [2(N(|\mathcal{T}| - 1) - \mathrm{Pers}(\mathcal{C}_{\max}(0)))]$, where $\Delta Q / [2(N(|\mathcal{T}| - 1) - \mathrm{Pers}(\mathcal{C}_{\max}(0)))] > \Delta Q / 2N(|\mathcal{T}| - 1)$, because $\mathrm{Pers}(\mathcal{C}_{\max}(0)) \geq |\mathcal{T}| - 1$ in any multilayer network.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

on the diagonal blocks of the multilayer modularity matrix $\boldsymbol{\mathcal{B}}$, and let $\max(\boldsymbol{\mathcal{B}}_{\text{diag}})$ and $\min(\boldsymbol{\mathcal{B}}_{\text{diag}})$, respectively, denote the maximum and minimum values of the set $\boldsymbol{\mathcal{B}}_{\text{diag}}$. Without loss of generality,[16] we assume that $\min(\boldsymbol{\mathcal{B}}_{\text{diag}}) < 0$ and $\max(\boldsymbol{\mathcal{B}}_{\text{diag}}) > 0$. Let C' be any multilayer partition with a maximal value of persistence. It follows that

$$Q(C'|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega) = \beta_2 + 2\omega N(|\mathcal{T}| - 1)$$

for some $\beta_2 \in \mathbb{R}$. Because $A \geq 1$, choosing

$$2\omega > |\mathcal{T}| N^2 \big[\max(\boldsymbol{\mathcal{B}}_{\text{diag}}) - \min(\boldsymbol{\mathcal{B}}_{\text{diag}})\big] \geq \beta_1 - \beta_2$$

ensures that C' yields a larger value of multilayer modularity than C for any β_1 and for all $A \in \{1, \ldots, N(|\mathcal{T}| - 1)\}$. ∎

The following proposition follows directly from Proposition 5.5.

PROPOSITION 5.6. *There exists $\omega_\infty > 0$ such that*

$$\textit{for all } r \in \{1, \ldots, |\mathcal{T}|\},\ C^{\omega}_{\max}|_r \textit{ is a solution of } \max_{C \in \mathcal{C}} Q\left(C|\sum_{s=1}^{|\mathcal{T}|} \boldsymbol{B}_s\right)$$

for all $\omega > \omega_\infty$.

Propositions 5.5 and 5.6 imply the existence of a "boundary value" for ω above which single-layer partitions induced by optimal multilayer partitions (1) are the same on all layers, and (2) are optimal solutions for the single-layer modularity-maximization problem defined on the mean modularity matrix.

Proof. Suppose that $\omega > \omega_\infty$, where ω_∞ is as defined in Proposition 5.5, and let $C^{\omega}_{\max} \in \mathcal{C}_{\max}(\omega)$. By Proposition 5.5, $\text{Pers}(C^{\omega}_{\max}) = N(|\mathcal{T}| - 1)$, and community assignments in $C^{\omega}_{\max}$ are the same across layers. Consequently, for $\omega > \omega_\infty$

$$\begin{aligned}
C^* &= \operatorname*{argmax}_{C \in \mathcal{C}} \left(\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs} \delta(c_{i_s}, c_{j_s}) + 2\omega \text{Pers}(C) \right), \\
\Leftrightarrow C^* &= \operatorname*{argmax}_{C \in \mathcal{C}} \left(\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs} \delta(c_i, c_j) + 2\omega N(|\mathcal{T}| - 1) \right), \\
\Leftrightarrow C^* &= \operatorname*{argmax}_{C \in \mathcal{C}} \left(\sum_{i,j=1}^{N} \left(\sum_{s=1}^{|\mathcal{T}|} B_{ijs} \right) \delta(c_i, c_j) \right),
\end{aligned}$$

where c_i denotes the community assignment of node i in all layers. ∎

The next two propositions formalize the intuition that an optimal multilayer partition measures a trade-off between static community structure within layers (i.e., intralayer modularity) and persistence of community structure across layers.

PROPOSITION 5.7. *Let $\omega_1 > \omega_2 > 0$. For all $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_2)$, one of the following two conditions must hold:*

(1) $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_1)$,

(2) $\text{Pers}(C^{\omega_2}_{\max}) < \text{Pers}(C^{\omega_1}_{\max})$ *for all* $C^{\omega_1}_{\max} \in \mathcal{C}_{\max}(\omega_1)$.

[16] If $\min(\boldsymbol{\mathcal{B}}_{\text{diag}})$ and $\max(\boldsymbol{\mathcal{B}}_{\text{diag}})$ have the same sign or if either is equal to 0, then every diagonal block of $\boldsymbol{\mathcal{B}}$ has either all nonnegative entries or all nonpositive entries. In both cases, an optimal partition $C^{\omega}_{\max}$ has maximal persistence for any value of $\omega > 0$, because $C^{\omega}_{\max}|_s$ is given either by a single community or by N singleton communities for all s. Consequently, by Proposition 5.3, $\text{Pers}(C^{\omega}_{\max})|_s = N$ for all s.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Proof. Let $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_2)$. If $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_1)$, then condition (1) is satisfied. Suppose that $C^{\omega_2}_{\max} \notin \mathcal{C}_{\max}(\omega_1)$, and assume that $\mathrm{Pers}(C^{\omega_2}_{\max}) \geq \mathrm{Pers}(C^{\omega_1}_{\max})$ for some $C^{\omega_1}_{\max} \in \mathcal{C}_{\max}(\omega_1)$. By definition of optimality, $C^{\omega_2}_{\max} \notin \mathcal{C}_{\max}(\omega_1)$ implies that

$$Q(C^{\omega_2}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_1) < Q(C^{\omega_1}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_1)\,, \tag{5.6}$$

where $\omega_1 > \omega_2$ by hypothesis. By writing

$$Q(C^{\omega_k}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_{k'}) = \sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs}\delta(c^{\omega_k}_{i_s}, c^{\omega_k}_{j_s}) + 2\omega_{k'}\mathrm{Pers}(C^{\omega_k}_{\max})\,,$$

where $c^{\omega_k}_{i_s}$ is the community assignment of node i_s in $C^{\omega_k}_{\max}$ and $k, k' \in \{1, 2\}$, and by substituting ω_1 by $\omega_2 + \Delta$ for some $\Delta > 0$, one can show that the inequality (5.6) implies

$$Q(C^{\omega_2}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_2) < Q(C^{\omega_1}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_2)\,,$$

which contradicts the optimality of $C^{\omega_2}_{\max}$. ∎

One can similarly prove the following proposition.

PROPOSITION 5.8. *Let $\omega_1 > \omega_2 > 0$. For all $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_2)$, one of the following two conditions must hold:*

(1) $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_1)$,

(2) $Q(C^{\omega_2}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) > Q(C^{\omega_1}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0)$ *for all* $C^{\omega_1}_{\max} \in \mathcal{C}_{\max}(\omega_1)$.

Proof. Let $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_2)$. If $C^{\omega_2}_{\max} \in \mathcal{C}_{\max}(\omega_1)$, then condition (1) is satisfied. Suppose that $C^{\omega_2}_{\max} \notin \mathcal{C}_{\max}(\omega_1)$, and assume that

$$Q(C^{\omega_2}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) \leq Q(C^{\omega_1}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) \tag{5.7}$$

for some $C^{\omega_1}_{\max} \in \mathcal{C}_{\max}(\omega_1)$. By definition of optimality, $C^{\omega_2}_{\max} \notin \mathcal{C}_{\max}(\omega_1)$ implies that

$$Q(C^{\omega_2}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_1) < Q(C^{\omega_1}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_1)\,, \tag{5.8}$$

where $\omega_1 > \omega_2$ by hypothesis. By writing

$$Q(C^{\omega_k}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_1) = Q(C^{\omega_k}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) + 2\omega_1\mathrm{Pers}(C^{\omega_k}_{\max}),$$

where $k \in \{1, 2\}$, and by using (5.7), one can show that (5.8) implies that

$$Q(C^{\omega_2}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_2) < Q(C^{\omega_1}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega_2)$$

for all $\omega_2 < \omega_1$. This contradicts the optimality of $C^{\omega_2}_{\max}$. ∎

To develop intuition for Propositions 5.7 and 5.8, it is helpful to think of a multilayer quality function $Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; \omega)$ for a given partition C as a linear function of ω that crosses the vertical axis at $Q(C|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0)$ with a slope $\mathrm{Pers}(C)$ (see, for example, the last panels of Figures 6 and 7).

The next three corollaries follow straightforwardly from Propositions 5.7 and 5.8. The first states that the largest achievable value of persistence for an optimal partition obtained with a given value of interlayer coupling is a nondecreasing function in ω. The second states that the largest achievable value of intralayer modularity for an optimal partition obtained with a given value of interlayer coupling is a nonincreasing

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

function in ω. The third property states that if two distinct values of ω have the same set of optimal partitions, then this set is also optimal for all intermediate values.

COROLLARY 5.9. *Let $\omega_1 > \omega_2 > 0$. It follows that*

$$\mathrm{Pers}(\mathcal{C}_{\max}(\omega_1)) \geq \mathrm{Pers}(\mathcal{C}_{\max}(\omega_2)),$$

where $\mathrm{Pers}(\mathcal{C}_{\max}(\omega)) := \max\{\mathrm{Pers}(C^{\omega}_{\max}), C^{\omega}_{\max} \in \mathcal{C}_{\max}(\omega)\}$.

COROLLARY 5.10. *Let $\omega_1 > \omega_2 > 0$. It follows that*

$$Q\left(\mathcal{C}_{\max}(\omega_1)|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0\right) \leq Q\left(\mathcal{C}_{\max}(\omega_2)|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0\right),$$

where $Q(\mathcal{C}_{\max}(\omega)|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0) := \max\{Q(C^{\omega}_{\max}|\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{|\mathcal{T}|}; 0), C^{\omega}_{\max} \in \mathcal{C}_{\max}(\omega)\}$.

COROLLARY 5.11. *Assume that $\mathcal{C}_{\max}(\omega_1) = \mathcal{C}_{\max}(\omega_2)$ for $\omega_1 > \omega_2 > 0$. It follows that*

$$\mathcal{C}_{\max}(\omega_1) = \mathcal{C}_{\max}(\omega) = \mathcal{C}_{\max}(\omega_2) \text{ for all } \omega \in (\omega_2, \omega_1).$$

One can extend the proofs of Propositions 5.1–5.8 so that they apply for interlayer coupling that is uniform between each pair of contiguous layers but may differ from pair to pair. In other words, one can obtain similar results for the maximization problem

$$\max_{C \in \mathcal{C}} \left[\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs} \delta(c_{i_s}, c_{j_s}) + 2 \sum_{s=1}^{|\mathcal{T}|-1} \omega_s \mathrm{Pers}(C)|_s \right].$$

Propositions 5.1–5.6 trivially extend to this case, and one can extend Propositions 5.7–5.8 if (for example) one assumes that $\omega_s^{(1)} > \omega_s^{(2)} > 0$ for all $s \in \{1, \ldots, |\mathcal{T}| - 1\}$, where $\boldsymbol{\omega}^{(1)}$ and $\boldsymbol{\omega}^{(2)}$ are $(|\mathcal{T}| - 1)$-dimensional vectors. (For example, one could set $\boldsymbol{\omega}^{(2)} = \omega \boldsymbol{\omega}^{(1)}$ and vary $\omega > 0$.)

5.3. Computational issues. We now examine issues that can arise when using the Louvain heuristic (see section 2.2) to maximize multilayer modularity (3.4).

5.3.1. Underemphasis of persistence. Consider the example network in Figure 8, which is a 3-layer network that has 5 nodes in each layer. Suppose that all nodes are strongly connected to each other in layers 1 and 3, and that the edge weight between node 1_2 and nodes $\{2_2, 3_2, 4_2, 5_2\}$ is smaller in layer 2 than the edge weight between node 1_s and nodes $\{2_s, 3_s, 4_s, 5_s\}$ when $s = 1, 3$. We use the uniform null network with $\gamma = 0.5$ and set $\omega = 0.1$. This produces a multilayer modularity matrix in which all the single-layer modularity entries B_{ijs} except those of node 1_2 are positive and exceed the value of interlayer coupling. Suppose that one loops over the nodes ordered from 1 to $N|\mathcal{T}|$ in phase 1 of the Louvain heuristic. The initial partition consists of $N|\mathcal{T}|$ singletons, and each node is then moved to the set that maximally increases modularity. The partition at the end of phase 1 is $\{\{1_1, 2_1, 3_1, 4_1, 5_1, 1_2\}, \{2_2, 3_2, 4_2, 5_2\}, \{1_3, 2_3, 3_3, 4_3, 5_3\}\}$. In phase 2, the second and third sets merge to form a single set,[17] and the Louvain heuristic gets trapped in a local optimum in which the smaller set of nodes (i.e., $\{1_1\}$) remains in the same community across layers 1 and 2, and the larger set of nodes (i.e., $\{2_1, 3_1, 4_1, 5_1\}$)

[17]Note that combining the first and second sets into a single set decreases modularity because the value of interlayer coupling is too small to compensate for the decrease in intralayer contributions to modularity.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

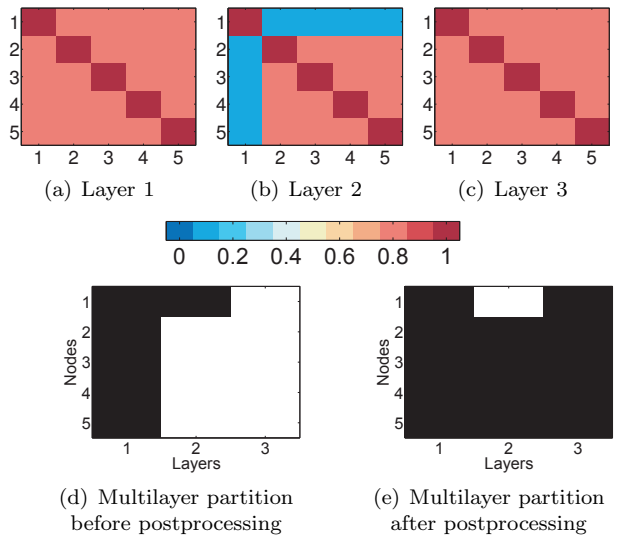


FIG. 8. *Toy example illustrating the effect of postprocessing on a multilayer partition by increasing multilayer modularity via community-assignment swaps that increase the value of persistence but do not change intralayer partitions. The colors in panels* (a)–(c) *scale with the entries of the adjacency matrix (color available online). Panel* (d) *(respectively, panel* (e)*) represents the output multilayer partition obtained with Louvain before (respectively, after) postprocessing. The horizontal axis represents the layers, and the vertical axis represents the nodes. The shading in panels* (d) *and* (e) *represents the community assignments of nodes in each layer.*

changes community. We show this multilayer partition in Figure 8(d). Repeating this experiment 1000 times using a randomized node order at the start of each iteration of phase 1 of the Louvain heuristic yields the same multilayer partition. One can modify this multilayer partition to obtain a new partition with a larger value of multilayer modularity by increasing the value of persistence across layers without changing intralayer partitions (we use this idea in the proof of Proposition 5.1). We show an example of this situation in Figure 8(e).

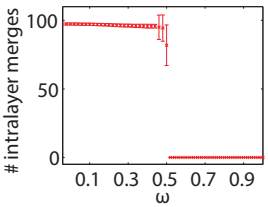
In Figure 8(d), we illustrate the above issue visually via abrupt changes in colors between layers. (These are more noticeable in larger networks.) Such changes are misleading because they imply a strong decrease in persistence that might not be accompanied by a significant change in intralayer partitions. In Figure 8(d), for example, the intralayer partitions differ in the community assignment of only a single node. To mitigate this problem, we apply a postprocessing function to all output partitions that maximizes persistence between layers without changing the partitions that are induced on each layer. We thereby produce a partition with a larger value of multilayer modularity. In our postprocessing, we relabel the community assignments of nodes in each layer such that (1) the number of nodes that remain in the same community between consecutive layers is maximized, and (2) the partition induced on each layer by the original multilayer partition is unchanged. By reformulating the problem of maximizing persistence for a given set of intralayer partitions as a weighted bipartite matching problem between consecutive layers, one can implement our postprocessing procedure using the *Hungarian algorithm* [19, 39, 51].

5.3.2. Abrupt drop in the number of intralayer merges. The Louvain heuristic faces a second problem in multilayer networks. When the value of inter-

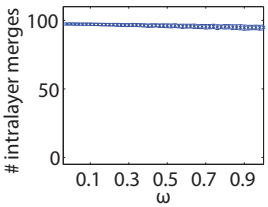
Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

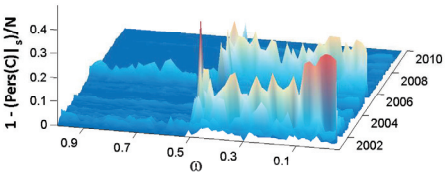
Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy



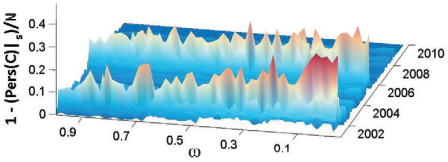
(a) Mean number of intralayer merges after the first completion of phase 1 of Louvain



(b) Mean number of intralayer merges after the first completion of phase 1 of LouvainRand



(c) Mean number of nodes that change community assignment between layers in output partition of Louvain



(d) Mean number of nodes that change community assignment between layers in output partition of LouvainRand

FIG. 9. *Comparison between the Louvain and LouvainRand algorithms. The sample of interlayer coupling values is the set* $\{0, 0.02, \ldots, 0.98, 1\}$ *with a discretization step of* 0.02 *between each pair of consecutive values. Top row: The number of nodes* n_{intra} *that have been merged with at least one node from the same layer after the first completion of phase* 1 *of* (a) *the Louvain heuristic and* (b) *the LouvainRand heuristic. For each heuristic, we average* n_{intra} *over* $|\mathcal{T}| = 238$ *layers and* 100 *iterations. The error bars in panels* (a) *and* (b) *indicate standard deviations. Bottom row: The value of* $1 - Pers(C)|_s/N$ *averaged over* 100 *runs of* (c) *the Louvain heuristic and* (d) *the LouvainRand heuristic after convergence of the algorithms to a local optimum.*

layer coupling satisfies

$$\omega > \max(\boldsymbol{\mathcal{B}}_{\text{diag}}) \,, \tag{5.9}$$

where $\boldsymbol{\mathcal{B}}_{\text{diag}}$ is the set of values on the diagonal blocks of $\boldsymbol{\mathcal{B}}$ defined in (5.5), the interlayer contributions to multilayer modularity are larger than the intralayer contributions for all pairs of nodes. Consequently, only interlayer merges occur during the first completion of phase 1 of the Louvain heuristic. In Figure 9(a), we illustrate this phenomenon using the MultiAssetClasses data set. The mean number of intralayer merges drops from roughly $N = 98$ (almost every node contains at least one other node from the same layer in its community) to 0. For ω values larger than $\max(\boldsymbol{\mathcal{B}}_{\text{diag}})$, every set at the end of the first completion of phase 1 only contains copies of each node in different layers and, in particular, does not contain nodes from the same layer. This can yield abrupt changes in the partitions induced on individual layers of the output multilayer partition.

In Figure 9(c), we show an example, using the MultiAssetClasses data set, of how the above issue can lead to an abrupt change in a quantitative measure computed from a multilayer output partition obtained with the Louvain heuristic. We note that the mean size of sets (averaged over 100 runs) after the first completion of phase 1 of the Louvain algorithm for $\omega > \max(\boldsymbol{\mathcal{B}}_{\text{diag}})$ is relatively small for MultiAsset-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Classes. (The mean is 3 nodes per set, and the maximum possible number of nodes per set is $|\mathcal{T}| = 238$, because each of these sets only contains copies of the same node when $\omega > \max(\boldsymbol{\mathcal{B}}_{\text{diag}})$.) Nevertheless, as ω increases past $\omega = \max(\boldsymbol{\mathcal{B}}_{\text{diag}})$, there is a sudden drop in the value of $(1-\text{Pers}(C)|_s/N)$ between consecutive layers in the output partition at $\omega = \max(\boldsymbol{\mathcal{B}}_{\text{diag}})$ (see Figure 9(c)). Nonzero values of $(1 - \text{Pers}(C)|_s/N)$ indicate that community assignments have changed between layers s and $s+1$ (by Proposition 5.3). Roughly speaking, Figure 9(c) suggests that when $\omega > \max(\boldsymbol{\mathcal{B}}_{\text{diag}})$, one abruptly moves away from the situation $\omega \approx \omega_0$ to a scenario that is closer to $\omega \approx \omega_\infty$.[18]

The above phenomenon occurs when the values of interlayer coupling are large relative to the entries of $\boldsymbol{\mathcal{B}}_{\text{diag}}$. In the correlation multilayer networks that we consider (or in unweighted multilayer networks), entries of the adjacency matrix satisfy $|A_{ijs}| \leq 1$. Assuming that one uses the modularity quality function on each layer and that $P_{ijs} \geq 0$ (e.g., $P_{ijs} = \langle \boldsymbol{A}_s \rangle$), this implies that

$$\max(\boldsymbol{\mathcal{B}}_{\text{diag}}) \leq 1 \text{ for all } \gamma \in [\gamma^-, \gamma^+]\,.$$

For networks in which the modularity cost of changing intralayer partitions in favor of persistence is large in comparison to the values of $\boldsymbol{\mathcal{B}}_{\text{diag}}$, it may be desirable to use $\omega > 1$ to gain insight into multilayer community structure. (For example, this occurs in both toy examples of section 5.1.)

To mitigate this problem, we change the condition for merging nodes in the Louvain heuristic. Instead of moving a node to a community that maximally increases modularity, we move a node to a community chosen uniformly at random from those that increase modularity. We call this heuristic *LouvainRand* [37], and we illustrate the results of using it in Figures 9(b) and (d). Although LouvainRand can increase the output variability (by increasing the search space of the optimization process), it seems to mitigate the problem for multilayer networks with diagonal, ordinal, and uniform coupling.

5.4. Multilayer community structure in asset correlation networks. In this section, we show the results of computational experiments in which we fix the value of the resolution parameter γ and vary the value of interlayer coupling ω. We use the uniform null network (i.e., $P_{ijs} = \langle \boldsymbol{A}_s \rangle$) and set $\gamma = 1$. We use the LouvainRand heuristic to identify multilayer partitions and apply our postprocessing procedure that increases persistence without changing partitions induced on individual layers to all output multilayer partitions. We showed in Proposition 5.5 that for $2\omega > 2\omega_\infty = |\mathcal{T}|N^2\big[\max(\boldsymbol{\mathcal{B}}_{\text{diag}}) - \min(\boldsymbol{\mathcal{B}}_{\text{diag}})\big]$, the set $\mathcal{C}_{\max}(\omega)$ of global optima no longer changes, and every optimal partition in this set has maximal persistence.[19] In our example, $|\mathcal{T}|N^2\big[\max(\boldsymbol{\mathcal{B}}_{\text{diag}}) - \min(\boldsymbol{\mathcal{B}}_{\text{diag}})\big] \leq 2|\mathcal{T}|N^2$, with $N = 98$ and $|\mathcal{T}| = 238$. However, for the purposes of the present paper, we take the set $\{0, 0.1, \ldots, 49.9, 50\}$ with a discretization step of 0.1 between consecutive values (giving 501 values in total) as our sample of ω values.

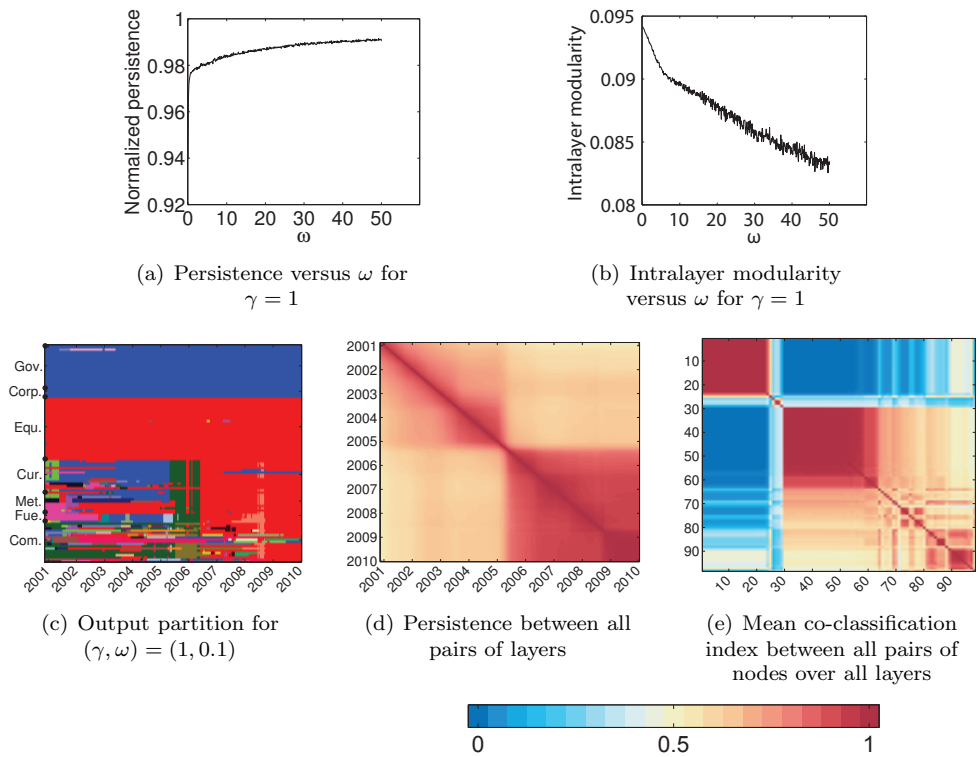
In agreement with the properties derived in Propositions 5.7 and 5.8, we observe in Figure 10(a) that normalized persistence (given by $\text{Pers}(C)/[N(|\mathcal{T}|-1)]$) tends to be larger for larger values of interlayer coupling, and in Figure 10(b) that intralayer

[18] For another instance in which an abrupt transition occurs as one increases ω, see the discussion in [63] on the change in eigenvalues of a multilayer Laplacian as a function of ω.

[19] Note that there can also be smaller values of ω_∞ for which this is true; in other words, we did not show that ω_∞ is the smallest lower bound of the set $\{w : \text{Pers}(C^\omega_{\max}) = N(|\mathcal{T}|-1) \text{ for all } C^\omega_{\max} \in \mathcal{C}_{\max}(\omega)\}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy



(a) Persistence versus ω for $\gamma = 1$

(b) Intralayer modularity versus ω for $\gamma = 1$

(c) Output partition for $(\gamma, \omega) = (1, 0.1)$

(d) Persistence between all pairs of layers

(e) Mean co-classification index between all pairs of nodes over all layers

FIG. 10. *Numerical experiments with* MULTIASSETCLASSES. *We sample the set of interlayer edge weights uniformly from the interval* $[0, 50]$ *with a discretization step of* 0.1 *(so there are* 501 *values of* ω *in total), and we use the uniform null network (i.e.,* $P_{ijs} = \langle \boldsymbol{A}_s \rangle$*) with* $\gamma = 1$. (a) *The persistence normalized by* $N(|\mathcal{T}| - 1)$ *for each value of* ω *averaged over* 20 *runs of Louvain-Rand.* (b) *The intralayer modularity* $\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} B_{ijs} \delta(c_{i_s}, c_{j_s})$ *normalized by* $\sum_{s=1}^{|\mathcal{T}|} \sum_{i,j=1}^{N} A_{ijs}$ *for each value of* ω *averaged over* 20 *runs of LouvainRand.* (c) *Sample output multilayer partition. Each point on the horizontal axis represents a single time window, and each position on the vertical axis is an asset. We order the assets by asset class, and the colors represent communities (color available online).* (d) *Association matrix of normalized persistence values between all pairs of layers averaged over all values of* $\omega \in [0, 50]$ *in our sample and* 20 *runs for each value. The normalized persistence between a pair of layers* $\{s, r\}$ *is* $\sum_{i=1}^{N} \delta(c_{i_s}, c_{i_r})/N$. (e) *Association matrix indicating the co-classification of nodes averaged over the set of partitions induced on each layer for each value of* ω *and* 20 *runs of LouvainRand.*

modularity (which we normalize by $\sum_{s=1}^{|\mathcal{T}|} (\mathbf{1}^T \boldsymbol{A}_s \mathbf{1})$) tends to be smaller for larger values of interlayer coupling. The increase of persistence and the decrease of intralayer modularity need not be monotonic, because we are finding a set of local optima for each value of ω rather than the set of global optima.

In Figure 10(c), we show a sample output of the multilayer partition (which contains 35 communities). (See section 4.2 for our definitions of the asset-class abbreviations.) Some of the changes in community structure correspond to known events (e.g., the Lehman bankruptcy in September 2008 (marked by an increase in the size of the equity asset class)). Observe that the two largest communities are the ones that contain the government bond assets and the equity assets. In particular, the community that contains equities becomes noticeably larger between 2006 and 2007, and again toward the end of 2008 (after the pink streak between 2008 and 2009 in Figure 10(c)). For larger values of the resolution parameter γ, this community instead

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

becomes noticeably larger only in 2008. (By inspecting the correlation matrices, one can check that the increase in correlation between equities and other assets is greater in 2008 than in 2006.)

In Figure 10(d), we show the matrix of mean values of persistence between all pairs of layers. The (s,r)th entry is the term $\sum_{i=1}^{N} \delta(c_{i_s}, c_{i_r})/N$ (where $s, r \in \{1, \ldots, |\mathcal{T}|\}$ need not be from consecutive layers), averaged over all values of $\omega \in [0, 50]$ in our sample and over multiple runs for each value of ω. Instead of only plotting $\mathrm{Pers}(C)|_s$ for consecutive layers, Figure 10(d) gives some indication as to whether nodes change communities between layers s and $s+1$ to join a community that contains a copy of some of these nodes from another time layer (i.e., $\sum_{i=1}^{N} \delta(c_{i_{s+1}}, c_{i_r}) \neq 0$ for some r) or to join a community that does not contain a copy of these nodes in any other time layer (i.e., $\sum_{i=1}^{N} \delta(c_{i_{s+1}}, c_{i_r}) = 0$ for all r). Figure 10(d) also gives some insight into whether there are sets of consecutive layers across which persistence values remain relatively large. This may shed light on when connectivity patterns change in a multilayer network. As indicated by the values on the color scale, the values of persistence in Figure 10(d) remain relatively high (which can be explained in part by the fact that equities and bonds remain in the same community across almost all layers, and these constitute roughly 50% of the node sample). The most noticeable separation into diagonal blocks in the middle of Figure 10(d) corresponds to the change in Figure 10(c) between 2005 and 2006, at which various currencies, metals, and fuels separate from the bond community (blue) to form a dark green community. The smaller diagonal block at the bottom right of Figure 10(d) corresponds to the change in Figure 10(c) after the Lehman bankruptcy between 2008 and 2009 (after the pink streak in Figure 10(c)).

In Figure 10(e), we show the mean co-classification index of nodes in partitions induced on individual layers. To calculate the mean, we average the co-classification index over layers, over all values of $\omega \in [0, 50]$ in our sample, and over multiple runs for each value of ω. (We reorder the nodes to emphasize diagonal blocks in the association matrix.) This figure yields insight into what sets of nodes belong to the same community across layers for increasing values of ω. This may help illuminate which connectivity patterns are shared across layers. Unsurprisingly, the first diagonal block corresponds primarily to bond assets, and the second diagonal block corresponds primarily to equity assets. Figures 10(d) and (e) complement each other: at a given γ resolution, the former gives an idea about *when* community structure has changed, and the latter gives an idea about *how* it has changed.

6. Conclusions. Modularity maximization in temporal multilayer networks is a clustering technique that produces a time-evolving partition of nodes. We have investigated two questions that arise when using this method: (1) the role of null networks in modularity maximization, and (2) the effect of interlayer edges on the multilayer modularity-maximization problem. We demonstrated that one must be cautious in interpreting communities obtained with a null network in which the distribution of expected edge weights is sample dependent. Furthermore, we showed that an optimal partition in multilayer modularity maximization reflects a trade-off between static community structure within layers and persistence of community structure across layers. One can try to exploit this in practice to detect changes in connectivity patterns and shared connectivity in a time-dependent network. Our results and observations hold for any choice of single-layer networks within layers; we discussed correlation networks in detail to use them as illustrative examples.

At the heart of modularity maximization is a comparison between what one antic-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

ipates and what one observes. The ability to specify what is anticipated is a desirable (albeit underexploited) feature of modularity maximization, because one can explicitly adapt it for different applications [4, 5, 20, 66]. By defining a null model as a probability distribution over the space of adjacency matrices and a null network as the expected adjacency matrix under the specified distribution, we highlight the important point that the *same* null network can correspond to *different* null models; this is not something that has been properly appreciated in the literature. Moreover, one needs to be very careful with one's choice of null network because it determines what one regards as densely connected in a network: different choices in general yield different communities. As we illustrated in section 4 for financial correlation networks, this choice can have a large impact on results and can lead to misleading conclusions.

In section 5, we proved several properties that describe the effect of diagonal, ordinal, and uniform interlayer coupling on multilayer modularity maximization or, more generally, on any maximization problem that can be cast in the form (3.4). Although our theoretical results do not necessarily apply to the local optima that one attains in practice, they do provide useful guidelines for how to interpret the outcome of a computational heuristic for maximizing modularity: if a multilayer partition is inconsistent with one of the proven properties, then it must be an artifact of the heuristic and not a feature of the quality function.

To further examine multilayer modularity maximization, we defined a measure that we called *persistence* to quantify how much community assignments change in time in a multilayer partition. For zero interlayer coupling, the value of persistence is 0, and it achieves a maximum finite value for sufficiently large interlayer coupling. We showed that the highest achievable value of persistence for an optimal partition obtained with a given value of interlayer coupling ω is a nondecreasing function in ω. Similarly, the highest achievable value of intralayer contributions to the quality function for an optimal partition obtained with a given value of interlayer coupling ω is a nonincreasing function in ω. The notion of persistence makes it possible to measure this trade-off between static community structure within layers and temporal persistence across layers. We illustrated this trade-off in our numerical experiments.

Finally, we showed that the Louvain heuristic can pose two issues when applied to multilayer networks with diagonal, ordinal, and uniform coupling. These can produce misleading values of persistence (or other quantitative measures of a multilayer partition) and can cause one to draw false conclusions about temporal changes in community structure in a network. We proposed ways to mitigate these problems and showed several numerical experiments on real data as illustrations. To further interpret these results, one needs to investigate more closely how the increase in persistence and the decrease in intralayer contributions to the quality function are exhibited in a multilayer partition between the "boundary cases" ω_0 and ω_∞ (see Propositions 5.4 and 5.5). This may help identify an interval of ω values in which the trade-off between persistence and intralayer modularity yields the most insights.

Acknowledgments. We thank HSBC Bank for providing the time-series data of financial assets. We thank Alex Arenas, Mihai Cucuringu, Sergio Gómez, Raphael Hauser, Lucas Jeub, and three anonymous referees for helpful comments.

REFERENCES

[1] A. Arenas, A. Fernández, and S. Gómez, *Analysis of the structure of complex networks at different resolution levels*, New J. Phys., 10 (2008), 053039.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

[2] T. Aynaud and J.-L. Guillaume, *Static community detection algorithms for evolving networks*, in Proceedings of WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Avignon, France, 2010, pp. 508–514.

[3] I. Barnett and J.-P. Onnela, *Change Point Detection in Correlation Networks*, preprint, arXiv:1410.0761v2 [stat.ME], 2015.

[4] D. S. Bassett, E. T. Owens, M. A. Porter, M. L. Manning, and K. E. Daniels, *Extraction of force-chain network architecture in granular materials using community detection*, Soft Matter, 11 (2015), pp. 2731–2744.

[5] D. S. Bassett, M. A. Porter, N. F. Wymbs, S. T. Grafton, J. M. Carlson, and P. J. Mucha, *Robust detection of dynamic community structure in networks*, Chaos, 23 (2013), 013142.

[6] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton, *Dynamic reconfiguration of human brain networks during learning*, Proc. Natl. Acad. Sci. USA, 108 (2011), pp. 7641–7646.

[7] S. Battiston, J. B. Glattfelder, D. Garlaschelli, F. Lillo, and G. Caldarelli, *The structure of financial networks*, in Network Science, Springer, London, 2010, pp. 131–163.

[8] E. T. Bell, *Exponential numbers*, Amer. Math. Monthly, 41 (1934), pp. 411–419.

[9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast unfolding of communities in large networks*, J. Stat. Mech. Theory Exp., no. 10 (2008), P10008.

[10] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, *The structure and dynamics of multilayer networks*, Phys. Rep., 544 (2014), pp. 1–122.

[11] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures Algorithms, 31 (2007), pp. 3–122.

[12] B. Bollobás and O. Riordan, *Random graphs and branching processes*, in Handbook of Large-Scale Random Networks, Bolyai Soc. Math. Stud. 18, Springer, Berlin, 2009, pp. 15–115.

[13] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner, *On modularity clustering*, IEEE Trans. Knowl. Data Eng., 20 (2008), pp. 172–188.

[14] M. Catanzaro and M. Buchanan, *Network opportunity*, Nature Phys., 9 (2013), pp. 121–123.

[15] F. Chung and L. Lu, *The average distance in a random graph with given expected degrees*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 15879–15882.

[16] R. Cont, *Empirical properties of asset returns: Stylized facts and statistical issues*, Quant. Finance, 1 (2001), pp. 223–236.

[17] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, *Stability of graph communities across time scales*, Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 12755–12760.

[18] M. De Domenico, A . Lancichinetti, A. Arenas, and M. Rosvall, *Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems*, Phys. Rev. X, 5 (2015), 011027.

[19] J. Edmonds and R. M. Karp, *Theoretical improvements in algorithmic efficiency for network flow problems*, J. ACM, 19 (1972), pp. 248–264.

[20] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, *Uncovering space-independent communities in spatial networks*, Proc. Natl. Acad. Sci. USA, 108 (2011), pp. 7663–7668.

[21] D. J. Fenn, M. A. Porter, M. McDonald, S. Williams, N. F. Johnson, and N. S. Jones, *Dynamic communities in multichannel data: An application to the foreign exchange market during the* 2007–2008 *credit crisis*, Chaos, 19 (2009), 033119.

[22] D. J. Fenn, M. A. Porter, P. J. Mucha, M. McDonald, S. Williams, N. F. Johnson, and N. S. Jones, *Dynamical clustering of exchange rates*, Quant. Finance, 12 (2012), pp. 1493–1520.

[23] D. J. Fenn, M. A. Porter, S. Williams, M. McDonald, N. F. Johnson, and N. S. Jones, *Temporal evolution of financial market correlations*, Phys. Rev. E, 84 (2011), 026109.

[24] S. Fortunato, *Community detection in graphs*, Phys. Rep., 486 (2010), pp. 75–174.

[25] S. Fortunato and M. Barthelémy, *Resolution limit in community detection*, Proc. Natl. Acad. Sci. USA, 104 (2006), pp. 36–41.

[26] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 7821–7826.

[27] S. Gómez, P. Jensen, and A. Arenas, *Analysis of community structure in networks of correlated data*, Phys. Rev. E, 80 (2009), 016114.

[28] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, *The dynamics of protest recruitment through an online network*, Sci. Rep., 1 (2011), 197.

[29] B. H. Good, Y.-A. de Montjoye, and A. Clauset, *Performance of modularity maximization in practical contexts*, Phys. Rev. E, 81 (2010), 046106.

[30] A. A. Groppelli and E. Nikbakht, *Barron's Finance*, 4th ed., Barron's Educational Series, Inc., Hauppauge, NY, 2000.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

[31] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, *Functional cartography of complex metabolic networks*, Nature, 443 (2005), pp. 895–900.

[32] T. Hoffmann, M. A. Porter, and R. Lambiotte, *Generalized master equations for non-Poisson dynamics on networks*, Phys. Rev. E, 86 (2012), 046102.

[33] P. Holme, *Modern temporal network theory: A colloquium*, Eur. Phys. J. B, 88 (2015), 234.

[34] P. Holme and J. Saramäki, *Temporal networks*, Phys. Rep., 519 (2012), pp. 97–125.

[35] J. Hopcroft, O. Khan, B. Kulli, and B. Selman, *Tracking evolving communities in large linked networks*, Proc. Natl. Acad. Sci. USA, 101 (2004), pp. 5250–5253.

[36] L. G. S. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, and M. W. Mahoney, *Think locally, act locally: Detection of small, medium-sized, and large communities in large networks*, Phys. Rev. E, 91 (2015), 012821.

[37] I. S. Jutla, L. G. S. Jeub, and P. J. Mucha, *A Generalized Louvain Method for Community Detection Implemented in* MATLAB*, Version* 2.0, http://netwiki.amath.unc.edu/GenLouvain/GenLouvain (2011–2014).

[38] M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *Multilayer networks*, J. Complex Networks, 2 (2014), pp. 203–271.

[39] H. W. Kuhn, *The Hungarian method for the assignment problem*, Naval Res. Logist. Quart., 2 (1955), pp. 83–97.

[40] R. Lambiotte, *Multi-scale modularity in complex networks*, in Proceedings of WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Avignon, France, 2010, pp. 546–553.

[41] R. Lambiotte, J.-C. Delvenne, and M. Barahona, *Laplacian Dynamics and Multiscale Modular Structure in Networks*, preprint, arXiv:0812.1770 [physics.soc-ph], 2008.

[42] R. Lambiotte, J.-C. Delvenne, and M. Barahona, *Random walks, Markov processes and the multiscale modular organization of complex networks*, IEEE Trans. Netw. Sci. Eng., 1 (2014), pp. 76–90.

[43] A. Lancichinetti and S. Fortunato, *Community detection algorithms: A comparative analysis*, Phys. Rev. E, 80 (2009), 056117.

[44] A. Lancichinetti and S. Fortunato, *Consensus clustering in complex networks*, Sci. Rep., 2 (2012), 336.

[45] A. C. F. Lewis, N. S. Jones, M. A. Porter, and C. M. Deane, *The function of communities in protein interaction networks at multiple scales*, BMC Syst. Biol., 4 (2010), 100.

[46] M. MacMahon and D. Garlaschelli, *Community detection for correlation matrices*, Phys. Rev. X, 5 (2015), 021006.

[47] K. T. Macon, P. J. Mucha, and M. A. Porter, *Community structure in the United Nations General Assembly*, Phys. A, 391 (2012), pp. 343–361.

[48] R. N. Mantegna, *Hierarchical structure in financial markets*, Eur. Phys. J. B, 11 (1999), pp. 193–197.

[49] P. J. Mucha and M. A. Porter, *Communities in multislice voting networks*, Chaos, 20 (2010), 041108.

[50] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community structure in time-dependent, multiscale and multiplex networks*, Science, 328 (2010), pp. 876–878.

[51] J. Munkres, *Algorithms for the assignment and transportation problems*, J. Soc. Indust. Appl. Math., 5 (1957), pp. 32–38.

[52] M. E. J. Newman, *Analysis of weighted network*, Phys. Rev. E, 70 (2004), 066146.

[53] M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices*, Phys. Rev. E, 74 (2006), 036104.

[54] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.

[55] M. E. J. Newman, *Communities, modules and large-scale structure in networks*, Nature Phys., 8 (2012), pp. 25–31.

[56] M. E. J. Newman and M. Girvan, *Finding and evaluating community structure in networks*, Phys. Rev. E, 69 (2004), 026113.

[57] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Dynamics of market correlations: Taxonomy and portfolio analysis*, Phys. Rev. E, 68 (2003), 056110.

[58] G. Palla, A.-L. Barabási, and T. Vicsek, *Quantifying social group evolution*, Nature, 446 (2007), pp. 664–667.

[59] L. Peel and A. Clauset, *Detecting change points in the large-scale structure of evolving networks*, in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Association for the Advancement of Artificial Intelligence, Palo Alto, CA, 2015, pp. 2914–2920.

[60] G. Pietri and P. Expert, *Temporal stability of network partitions*, Phys. Rev. E, 90 (2014), 022813.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[61] M. A. Porter, J.-P. Onnela, and P. J. Mucha, *Communities in networks*, Notices Amer. Math. Soc., 56 (2009), pp. 1082–1097, 1164–1166.

[62] M. Potters, J.-P. Bouchaud, and L. Laloux, *Financial applications of random matrix theory: Old laces and new pieces*, Acta Phys. Polon. B, 36 (2005), pp. 2767–2784.

[63] F. Radicchi and A. Arenas, *Abrupt transition in the structural formation of interconnected networks*, Nature Phys., 9 (2013), pp. 717–720.

[64] J. Reichardt and S. Bornholdt, *Statistical mechanics of community detection*, Phys. Rev. E, 74 (2006), 016110.

[65] M. Sales-Pardo, R. Guimerà, A. A. Moreira, and L. A. N. Amaral, *Extracting the hierarchical organization of complex systems*, Proc. Natl. Acad. Sci. USA, 39 (2007), pp. 15224–15229.

[66] M. Sarzynska, E. L. Leicht, G. Chowell, and M. A. Porter, *Null models for community detection in spatially embedded, temporal networks*, J. Complex Networks (2015), doi:10.1093/comnet/cnv027.

[67] R. Schafer and T. Guhr, *Local normalization: Uncovering correlations in non-stationary financial time series*, Phys. A, 389 (2010), pp. 3856–3865.

[68] H. A. Simon, *The architecture of complexity*, Proc. Amer. Phil. Soc., 106 (1962), pp. 467–482.

[69] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, *Network modelling methods for FMRI*, NeuroImage, 54 (2011), pp. 875–891.

[70] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, *Clustering Network Layers with the Strata Multilayer Stochastic Block Model*, preprint, arXiv:1507.01826v2 [cs.SI], 2015.

[71] V. A. Traag and J. Bruggeman, *Community detection in networks with positive and negative links*, Phys. Rev. E, 80 (2009), 036115.

[72] V. A. Traag and P. Van Dooren, *Narrow scope for resolution-limit-free community detection*, Phys. Rev. E, 84 (2011), 016114.

[73] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *Comparing community structure to characteristics in online collegiate social networks*, SIAM Rev., 53 (2011), pp. 526–543.

[74] S. Williams, *Risk trades will test investors through* 2011, Financial Times, January 5, 2011.

[75] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber, *Evidence for a bimodal distribution in human communication*, Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 18803–18808.

[76] A. Zalesky, A. Fornito, and E. Bullmore, *On the use of correlation as a measure of network connectivity*, NeuroImage, 60 (2012), pp. 2096–2106.

[77] Q. Zhao, Y. Tian, Q. He, N. Oliver, R. Jin, and W.-C. Lee, *Communication motifs: A tool to characterize social communications*, in Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010, pp. 1645–1648.

Downloaded 12/14/22 to 131.179.156.5 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.