CHAPTER

MULTILAYER SOCIAL NETWORKS

25

Brandon Oselio*, **Sijia Liu***, **Alfred Hero***

*EECS Department, University of Michigan, Ann Arbor, MI, United States**

25.1 INTRODUCTION

Social media's prevalence in daily life has led to a massive increase in data for modeling human behaviors. Often, this data is structured in a way that can be thought of as a connection of links between agents, i.e., a network of social media users. This precipitates a necessity for specialized algorithms that can handle multilayer structured network data for inference, evaluation, and prediction.

We often find heterogeneous structure in social media data—there may exist more than one type of relationship between agents, and these relationships may impose different topological characteristics. For instance, people may be connected by more than one social platform. Alternatively, we may observe explicit links between agents but also infer implicit affinities based on agent features.

Another example of this heterogeneous structure arises when relationships between agents appear and disappear over time; agents begin talking to each other at one time and end at another time, possibly signifying a change in relation. Both of the above examples can be explained by a multilayer network framework.

A multilayer network is a network where a set of elementary units is connected by intralayer and interlayer relationships ("edges"). This structure is a generalization of single-layer networks where there are only intralayer relationships. These layers represent heterogeneity in the structure or labeling of the data; a layer might correspond to a type of connection or a discrete timestep. The interlayer structure represents ties among nodes in the different layers; this structure may be observed, assumed, or estimated depending on the application. The interlayer structure in a social network often corresponds to the labels of each node, so that each node in a single layer is connected to its counterparts in the other layers. If the layers represent timesteps, each entity might be connected to its counterpart in layers before and after the present layer, which represents the localization of that layer's characteristics in time.

As the multilayer structure is more complicated than its single-layer counterpart, methods for single-layer analysis must be modified to accommodate accordingly, and new methods are developed specifically for the multilayer case. This chapter will review some of the approaches for modeling multilayer networks and some of the methods that are specific to this structure.

Cooperative and Graph Signal Processing. https://doi.org/10.1016/B978-0-12-813677-5.00025-0

Table 25.1 List of Notations

Symbol	Description
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	a graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$
$\mathcal{M}$, $\mathcal{G}_M$, $\mathbf{M}$	a multilayer network $\mathcal{M}$ with supragraph $\mathcal{G}_M$ and tensor form $\mathbf{M}$
$\mathbf{A}_M$, $\mathbf{L}_M$	supra-adjacency matrix $\mathbf{A}_M$ and supra-Laplacian matrix $\mathbf{L}_M$
$\mathbf{A}^{(\alpha)}$, $\mathbf{L}^{(\alpha)}$	adjacency matrix and Laplacian matrix for network at layer α
$[L]$	an integer set $\{1, 2, \ldots, L\}$
$\circ$, $\otimes$	outer/tensor product, and Kronecker product
$\mathbf{X}^{(t)}$, $\mathbf{Y}^{(t)}$	features from time $1, \ldots, t$

The rest of this chapter will proceed as follows: Section 25.2 will discuss the mathematical formulation of multilayer networks. Section 25.3 will cover some examples of multilayer node centralities. Section 25.4 will review some types of multilayer community detection methods. Section 25.5 will pivot and discuss the estimation of multilayer interaction networks for social media data. Section 25.6 will utilize some of the techniques discussed in the chapter on two application datasets. Finally, Section 25.7 will provide some concluding remarks. We list our notations used in this chapter in Table 25.1.

25.2 MATHEMATICAL FORMULATION OF MULTILAYER NETWORKS

In this section, we focus on the mathematical formulation of multilayer networks. Different from single-layer networks, they allow multiple types of interactions between each pair of nodes. In what follows, we introduce two network representations: supra-adjacency representation and tensor representation, each of which generalizes the notation of a single-layer network. We next show some real-life examples that involve the multilayer network structure.

25.2.1 MODELING AND REPRESENTATION

A single-layer network (also called a monoplex network) can be represented by a graph [1]. A graph is a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges that connects pairs of nodes. A multilayer network generalizes the notion of a single-layer network by incorporating the interlayer connections; see Fig. 25.1A for an illustrative example. More formally, a multilayer network is a pair $\mathcal{M} = (\mathcal{T}, \mathcal{C})$ [2], where $\mathcal{T} = \{\mathcal{G}_\alpha, \alpha \in [L]\}$ is a family of graphs $\mathcal{G}_\alpha = (\mathcal{V}_\alpha, \mathcal{E}_\alpha)$ with $\mathcal{V}_\alpha \subseteq \mathcal{V}$, $[L] := \{1, 2, \ldots, L\}$, and $\mathcal{C} = \{\mathcal{E}_{\alpha\beta} \subseteq \mathcal{V}_\alpha \times \mathcal{V}_\beta, \alpha, \beta \in [L]\}$ denotes the set of interlayer connections ($\alpha \neq \beta$). Here α is the layer index, and by convention, $\mathcal{E}_{\alpha\alpha} = \mathcal{E}_\alpha$. When $L = 1$, the multilayer network $\mathcal{M}$ simplifies to a single-layer network. In the rest of the chapter, unless specified otherwise, we assume that each layer contains the same set of nodes with $|\mathcal{V}_\alpha| = |\mathcal{V}| = N$ for $\alpha \in [L]$, where $|\mathcal{V}|$ denotes the cardinality of the node set $\mathcal{V}$.

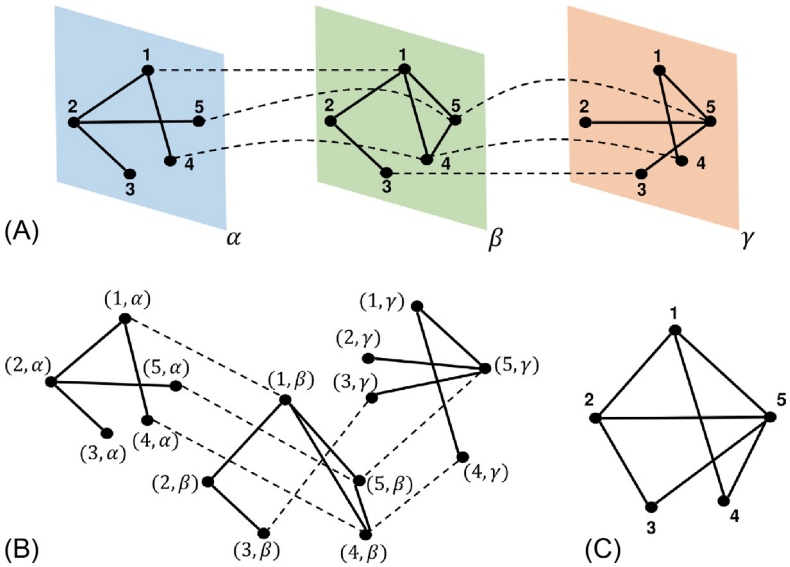

FIG. 25.1

Schematic illustration of multilayer network. Example with five nodes and three layers labeled α, β, and γ. (A) Multilayer network where *solid line* represents intralayer connection, and *dashed line* represents interlayer connection. (B) Supragraph representation. (C) Aggregated network.

Supragraph representation

Let $\mathcal{V}_M \subseteq \mathcal{V} \times [L]$ denote a set of node-layer combinations corresponding to $\mathcal{M}$, where $(v, \alpha) \in \mathcal{V}_M$ signifies that the node $v \in \mathcal{V}$ is present in layer $\alpha \in [L]$. Let $\mathcal{E}_M \subseteq \mathcal{V}_M \times \mathcal{V}_M$ be the set of edges between node-layer tuples. The multilayer network $\mathcal{M}$ can then be described by a graph $\mathcal{G}_M = (\mathcal{V}_M, \mathcal{E}_M)$, known as a supragraph, leading to a supra-adjacency matrix $\mathbf{A}_M$ and/or a supra-Laplacian matrix $\mathbf{L}_M$ [3]. Fig. 25.1B shows the supragraph representation of the multilayer network in Fig. 25.1A. Based on such a representation, many methods for single-layer networks, e.g., centrality-based network diagnostics and community detection methods, can be extended to multilayer networks [4,5].

In contrast to the supragraph, network aggregation provides the simplest representation for a multilayer network, where connections between nodes are aggregated in all layers to a single layer. The resulting graph is given by $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a)$, where $\mathcal{V}_a = \cup_{\alpha=1}^{L} \mathcal{V}_\alpha$ and $\mathcal{E}_a = \cup_{\alpha=1}^{L} \mathcal{E}_\alpha$. Often the aggregated network can be cast as a convex combination (e.g., linear combination) of graph adjacency matrices across all layers [6,7]. Although such an aggregation may cause loss of information about the interlayer network structure [3], it becomes useful when modeling across networks that have very similar interlayer connectivity. Fig. 25.1C shows the aggregated representation of the multilayer network in Fig. 25.1A.

Tensor representation

A multilayer network can be represented in a tensor form [2,8,9]. Let $\mathbf{M} \in \mathbb{R}^{N \times L \times N \times L}$ denote the fourth-order adjacency tensor of the L-layer network $\mathcal{M}$. Each element of $\mathbf{M}$ is defined by

$$M_{i\alpha j\beta} = \begin{cases} w_{i\alpha j\beta} & \text{if } (v_i^{(\alpha)}, v_j^{(\beta)}) \in \mathcal{E}_{\alpha\beta} \\ 0 & \text{otherwise,} \end{cases} \tag{25.1}$$

for $i, j \in [N]$ and $\alpha, \beta \in [L]$, where $v_i^{(\alpha)} \in \mathcal{V}_\alpha$ denotes node i at layer α, and $w_{i\alpha j\beta}$ is the weight corresponding to the edge $(v_i^{(\alpha)}, v_j^{(\beta)})$. We refer readers to [9] for a detailed background on tensors.

We can express the multilayer adjacency tensor (25.1) as a linear combination of tensors in the canonical basis

$$\mathbf{M} = \sum_{i=1}^{N}\sum_{\alpha=1}^{L}\sum_{j=1}^{N}\sum_{\beta=1}^{L} w_{i\alpha j\beta}(\mathbf{e}_i \circ \mathbf{e}_\alpha \circ \mathbf{e}_j \circ \mathbf{e}_\beta), \tag{25.2}$$

where $\circ$ represents the vector outer product (tensor product),[1] $\mathbf{e}_i$ is a basis vector in $\mathbb{R}^N$ with 1 at the ith coordinate and 0s elsewhere, and $\mathbf{e}_\alpha$ is a basis vector in $\mathbb{R}^L$. The tensor representation (25.2) can be viewed as a generalization of the graph adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ for the single-layer network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

$$\mathbf{A} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\mathbf{e}_i\mathbf{e}_j^T = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(\mathbf{e}_i \circ \mathbf{e}_j), \tag{25.3}$$

where w_{ij} is the weight associated with edge $(v_i, v_j) \in \mathcal{E}$.

In addition to the fourth-order tensor representation (25.2), a multilayer network is also modeled by a third-order tensor in [10], where each slice corresponds to the network at one layer, e.g., a dynamic network at one snapshot. In contrast with the third-order tensor, the fourth-order tensor encodes detailed information on the interlayer connection between any two nodes at different layers, namely, Eq. (25.1). Also, the fourth-order tensor $\mathbf{M}$ can be flattened out to the supra-adjacency matrix $\mathbf{A}_M$ of dimension $NL \times NL$. Therefore, the fourth-order tensor is a natural representation of multilayer networks, and many techniques on tensor algebra [8] can be used for network analysis.

25.2.2 EXAMPLES OF MULTILAYER NETWORKS

We next introduce three important classes of multilayer networks: node-colored networks, edge-colored networks, and temporal networks [3].

Node-colored networks are graphs in which each node is labeled by one color. Considering each color as designating a layer, node-colored graphs can be represented as multilayer networks. They are often used to model heterogeneous networks that contain nodes of different types.

Example 1. Bibliographic information networks contain information about researchers (authors) and publications they produce (documents). Links exist between papers and/or authors by the authorship, colleagueship, published venues, or topics [11].

[1] If $\mathbf{X} = \mathbf{a}_1 \circ \mathbf{a}_2 \circ \cdots \circ \mathbf{a}_n$, then each element of the tensor $\mathbf{X}$ is given by $X_{i_1 i_2 \ldots i_n} = [\mathbf{a}_1]_{i_1}[\mathbf{a}_2]_{i_2}\cdots[\mathbf{a}_n]_{i_n}$, where $[\mathbf{x}]_i$ denotes the ith entry of $\mathbf{x}$.

Example 2. Internet of Things (IoT) denotes the internetworking of smart phones, computers, vehicles, buildings, and other devices embedded with electronics, sensors, and actuators [12]. IoT allows autonomous exchange of useful information between "heterogeneous nodes."
Edge-colored networks are graphs with multiple types of edges, where similar to node-colored networks, color distinguishes between layers. Edge-colored graphs can be represented by multilayer networks, where nodes in each layer are fixed and linked by edges with a unique color. They can be used to model multirelational networks where nodes have relations of different types [13].
Example 3. Public social networks link social entities by several types of relationships, including friendship, vicinity, kinship, and membership in the same cultural society [14].
Example 4. Urban transportation networks describe the urban ecosystem, where nodes represent spatial locations (e.g., restaurants, shopping malls, schools, parks, and other places of interest), and edges represent vehicles of different types, e.g., taxis, buses, and subways, that are used to travel between two locations [15].

A temporal network is given by an ordered sequence of graphs. It can be interpreted as a special case of an edge-colored multigraph, where the set of time instants provides the set of edge colors, and the interlayer edges are between nodes and their counterparts across all time steps. The chromatin contact map over a time course of cell growth/development is an example of a temporal network in biology [16].

25.3 DIAGNOSTICS FOR MULTILAYER NETWORKS: CENTRALITY ANALYSIS

The study of centrality, i.e., evaluating the degree of nodal importance to the network structure, is often used to identify and rank essential nodes in complex networks. A number of centrality measures are commonly used, such as degree, eigenvector, clustering coefficient, closeness, betweenness, hubness, and authority, differing in what type of influence is to be emphasized [17]. For example, degree centrality measures the total number of connections a node has while eigenvector centrality measures the importance of a node by the importance of its neighbors [18]. Most centrality methods are only directly applicable to single-layer networks. Here we generalize some important single-layer centrality methods to multilayer networks.

25.3.1 OVERLAPPING DEGREE AND MULTIPLEX PARTICIPATION COEFFICIENT

Nodal degree is the simplest feature in network diagnostics. There exist several ways to define multilayered degree centrality. The simplest way is to use network aggregation, where two nodes are considered to be adjacent if and only if the number of edges that connect them in a multilayer network is larger than a threshold [19,20]. However, this measure does not fully consider the interlayer effect.

In a multilayer network, it is essential to study how the nodal degree is distributed across different layers. We recall from Section 25.2 that $\mathcal{M}$ denotes a multilayer network with N nodes and L layers, the degree of node i on layer α becomes

$$k_i^{(\alpha)} = \sum_{j=1}^{N} A_{ij}^{(\alpha)}, \tag{25.4}$$

where $A_{ij}^{(\alpha)}$ is the (i,j)th entry of the adjacency matrix associated with graph $\mathcal{G}_\alpha$ on layer α. The degree of node i in a multilayer network is a vector quantity

$$\mathbf{k}_i = \left[k_i^{(1)}, k_i^{(2)}, \ldots, k_i^{(L)}\right], \quad i \in [N]. \tag{25.5}$$

The overlapping degree of node i across all layers is defined as [6]

$$o_i = \sum_{\alpha=1}^{L} k_i^{(\alpha)} = \mathbf{1}^T \mathbf{k}_i, \quad i \in [N], \tag{25.6}$$

where $\mathbf{1}$ is the $L \times 1$ vector of all ones. The overlapping degree (25.6) can be used to identify hubs, nodes with high degree in the network. However, a node that is a hub in one layer may only have a few connections in another layer. Thus a more suitable multilayer hub definition is the multiplex participation coefficient [6,21],

$$P_i = \frac{L}{L-1}\left[1 - \sum_{\alpha=1}^{L}\left(\frac{k_i^{(\alpha)}}{o_i}\right)^2\right]. \tag{25.7}$$

Here P_i takes values in $[0,1]$ and measures the degree to which the degree of node i is uniformly distributed among the L layers. If $P_i = 1$, then node i has exactly the same number of edges on each layer, namely, $k_i^{(\alpha)} = o_i/L$. If $P_i = 0$, all the edges of node i are concentrated in just one layer. The multiplex participation coefficient thus captures the heterogeneity of nodal degrees across layers in multilayer networks.

25.3.2 EIGENVECTOR CENTRALITY IN SUPRAGRAPH

Eigenvector centrality describes the impact of a node on the network's global structure, and is defined by the dominant eigenvector of the graph adjacency matrix. Eigenvector centrality is widely used in many applications. For example, it is closely related to hubness and authority centrality used in the hyperlink-induced topic search (HITS) algorithm [22]. Because computing the dominant eigenvalue and eigenvector can be computed in a distributed setting, eigenvector centrality is often preferable to other types of global centralities such as betweenness [23,24].

The simplest way to generalize the concept of eigenvector centrality for multilayer networks is to use network aggregation and apply single-layer based methods [25]. However, as shown in Fig. 25.1, network aggregation oversimplifies the multilayer network. Therefore, we consider the supragraph representation $\mathcal{G}_M$ of a multilayer network with L layers and N nodes. The supra-adjacency matrix

$\mathbf{A}_M \in \mathbb{R}^{NL\times NL}$ of $\mathcal{G}_M$ can be separated into two parts: the intralayer component $\mathbf{A}_M^L$ and the interlayer component $\mathbf{A}_M^I$. That is,

$$\mathbf{A}_M = \mathbf{A}_M^L + \mathbf{A}_M^I, \quad \mathbf{A}_M^L = \text{diag}(\{\mathbf{A}^{(\alpha)}\}_{\alpha=1}^L), \tag{25.8}$$

where $\text{diag}(\{\mathbf{A}^{(\alpha)}\}_{\alpha=1}^L)$ denotes a block-diagonal matrix with diagonal elements $\mathbf{A}^{(\alpha)}$ for $\alpha \in [L]$, and recall that $\mathbf{A}^{(\alpha)}$ is the graph adjacency matrix on layer α. The interlayer supra-adjacency matrix $\mathbf{A}_M^I$ defines the interlayer connectivity between every two layers. If the interlayer connectivity is identical for all nodes [5], then $\mathbf{A}_M^I = \mathbf{A}^I \otimes \mathbf{I}_N$, where $\mathbf{A}^I \in \mathbb{R}^{L\times L}$ is an interlayer adjacency matrix whose elements represent the strength of the connection between every pair of layers. For example, in a temporal network, if layers are connected at consecutive time steps, then the interlayer supra-adjacency matrix becomes

$$\mathbf{A}_M^I = \mathbf{A}^I \otimes \mathbf{I}_N, \quad \mathbf{A}^I = \begin{bmatrix} 0 & 1 & 0 & \cdots \\ 1 & 0 & 1 & \ddots \\ 0 & 1 & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{25.9}$$

Here $\mathbf{A}^I$ models an undirected chain network in which each node is adjacent to its nearest neighbors. It is worth mentioning that the decomposition of the supra-adjacency matrix in Eq. (25.8) facilitates exploring the spectral properties of multilayer networks [5].

The eigenvector centrality $\mathbf{v}_M \in \mathbb{R}^{NL}$ of the supra-adjacency matrix $\mathbf{A}_M$ can then be defined as the solution to the following eigenvalue problem

$$\mathbf{A}_M\mathbf{v}_M = \lambda_{\max}\mathbf{v}_M, \tag{25.10}$$

where $\lambda_{\max}$ denotes the largest positive eigenvalue of $\mathbf{A}_M$. The entries of $\mathbf{v}_M$ give the centralities of each node-layer pair. It is convenient to map the eigenvector centralities $\mathbf{v}_M$ to an $N \times L$ matrix

$$V_{i\alpha} = v_{N(\alpha-1)+i}, \quad i \in [N],\ \alpha \in [L], \tag{25.11}$$

where $V_{i\alpha}$ corresponds to the joint centrality of the node-layer pair (i, α). Based on Eq. (25.11), we introduce the marginal node centrality $\hat{v}_i$ and the marginal layer centrality $\tilde{v}_\alpha$ [24]

$$\hat{v}_i = \sum_{\alpha=1}^{L} V_{i\alpha}, \quad i \in [N], \quad \tilde{v}_\alpha = \sum_{i=1}^{N} V_{i\alpha}, \ \alpha \in [L]. \tag{25.12}$$

Similar to the supra-adjacency matrix in Eq. (25.8), we can define the decomposition of the supra-Laplacian matrix, $\mathbf{L}_M = \mathbf{L}_M^L + \mathbf{L}_M^I$, where $\mathbf{L}_M^L = \text{diag}(\mathbf{A}_M^L\mathbf{1}) - \mathbf{A}_M^L$, and $\mathbf{L}_M^I = \text{diag}(\mathbf{A}_M^I\mathbf{1}) - \mathbf{A}_M^I$. The decomposition of the supra-Laplacian matrix corresponds to a diffusion process over nodes of the network. Specifically, the nodal dynamics follows the differential equation [5]

$$\dot{x}_{i\alpha} = \sum_{j=1}^{N} w_{ij}^{(\alpha)}(x_{j\alpha} - x_{i\alpha}) + \sum_{\beta=1}^{L} u_{\alpha\beta}(x_{i\beta} - x_{i\alpha}) \tag{25.13}$$

for any $i \in [N]$ and $\alpha \in [L]$, where $x_{i\alpha}$ denotes the state of node i at layer α, $w_{ij}^{(\alpha)}$ is the (i,j)th entry of the graph adjacency matrix $\mathbf{A}^{(\alpha)}$ at layer α, and $u_{\alpha\beta}$ is the interlayer coupling constant, namely, the (α,β)th entry of the interlayer adjacency matrix $\mathbf{A}^I$. The discretized matrix form of the diffusion equation (25.13) yields

$$\dot{\mathbf{x}} = -(\mathbf{L}_M^L + \mathbf{L}_M^I)\mathbf{x} = -\mathbf{L}_M\mathbf{x}. \tag{25.14}$$

Here the second smallest eigenvalue of $\mathbf{L}_M$ (also known as algebraic connectivity [1]) governs the convergence properties of the diffusion process.

25.3.3 NODAL CENTRALITY VIA TENSOR DECOMPOSITION

A fourth-order tensor was introduced in Section 25.2 to represent a multilayer network. Tensor decomposition is an effective tool for multiarray data analysis, and mono-layer centrality measures can be extended in order to identify key nodes in multilayer networks. It has been shown in [2] that the principal singular vectors obtained from the CANDECOMP/PARAFAC tensor decomposition [9] can provide hub and authority scores of all nodes in a multilayer network.

The fourth-order adjacency tensor $\mathbf{M} \in \mathbb{R}^{N\times L\times N\times L}$ of a multilayer network can be decomposed into a sum of rank-one tensors [9],

$$\mathbf{M} = \sum_{i=1}^{R} \sigma_i \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i \circ \mathbf{d}_i, \tag{25.15}$$

where $\{\sigma_i\}_{i=1}^R$ are singular values of $\mathbf{M}$ sorted in a descending order, $\mathbf{a}_i \in \mathbb{R}^N$, $\mathbf{b}_i \in \mathbb{R}^L$, $\mathbf{c}_i \in \mathbb{R}^N$, and $\mathbf{d}_i \in \mathbb{R}^L$ are singular vectors corresponding to the singular value σ_i, and R is the rank of $\mathbf{M}$. Considering the principal quadruplet $\{\mathbf{a}_1, \mathbf{b}_1, \mathbf{c}_1, \mathbf{d}_1\}$ in Eq. (25.15), the entries of $\mathbf{a}_1$ and $\mathbf{c}_1$ correspond to hub and authority scores of all nodes while the entries of $\mathbf{b}_1$ and $\mathbf{d}_1$ give hub and authority scores of all layers. Note that if $L = 1$, then the four-order adjacency tensor reduces to the second-order adjacency matrix, and the entries of $\mathbf{a}_1$ and $\mathbf{c}_1$ give the conventional hub and authority scores of nodes in a single-layer network. Given the hub and authority scores [2], one can further generalize HITS of multilayer networks [26].

Based on Eq. (25.15), the importance of a node-layer pair (i,α) can be evaluated as

$$H_{i,\alpha} = |a_{1,i}b_{2,\alpha}| + |c_{1,i}d_{1,\alpha}|, \tag{25.16}$$

where $a_{1,i}$ denotes the ith entry of the vector $\mathbf{a}_1$. The nodal importance measure defined in Eq. (25.16) is called EDCPTD (Essential nodes Determining based on CP Tensor Decomposition) centrality [2]. Given the joint centrality of the node-layer pair in Eq. (25.16), we can then define the marginal node centrality and the marginal layer centrality following Eq. (25.12).

In addition to the hubness and authority centrality, other generalized centrality measures such as clustering coefficient, modularity, and random walk centrality can also be defined using the tensor representation; see [8] for details.

25.4 CLUSTERING AND COMMUNITY DETECTION IN MULTILAYER NETWORKS

Discovering mesoscale structures such as communities in complex networks is a wide field of study [27]. These communities are generally described as a subset of nodes in the graph that are more densely connected than other nodes in the network. This is sometimes referred to in sociology as homophily [28]. Detecting these communities in a single-layer network has and continues to be an active research field. Furthermore, research in community detection for the more general multilayer case has become increasingly prevalent in the past decade.

Community detection in social networks facilitates the interpretation of the overall structure of the network. Generally, we expect to see communities in social networks that strongly relate different agents to one another, such as common activities, interests, or memberships to organizations. For instance, students that attend the same university, play the same sport, or like the same music are more likely to be connected in a particular link type. The concept of communities becomes more complex when multiple layers are introduced (see Fig. 25.2); communities that develop in one type of interaction may not be present in another or may be subsumed by a larger, more prevalent supercommunity. It is also possible that the community structure in each layer exhibits different homophilic clusters that

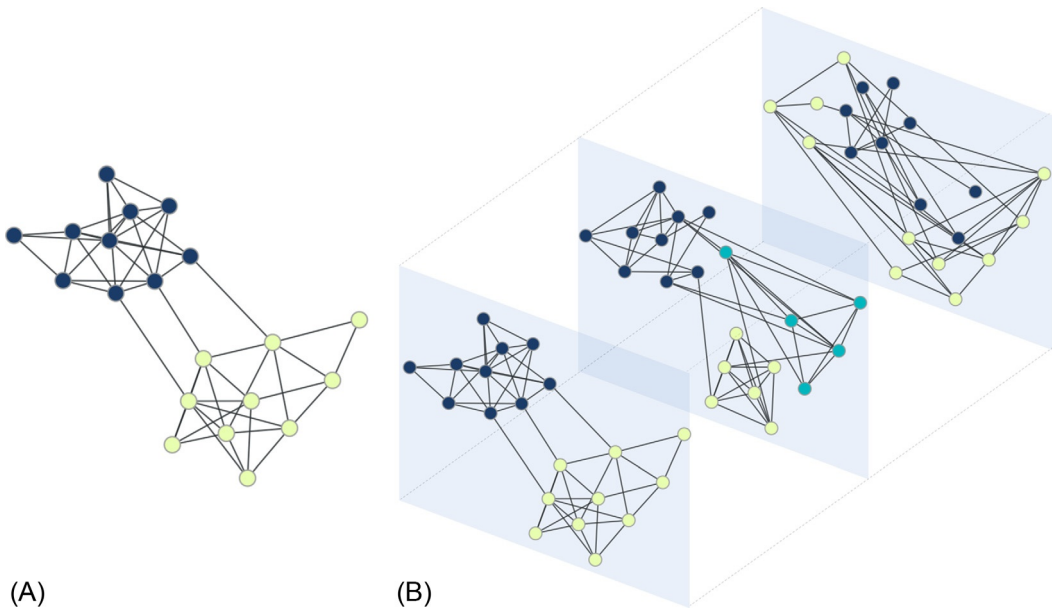

FIG. 25.2

Illustration of community detection. Examples of community detection with 20 nodes. (A) Single-layer community detection, where the community structure captures homophily among the nodes. (B) Community detection in a multilayer setting, where more complex situations can occur. The middle layer has a subcommunity in one of the larger communities displayed in the front layer while the back layer has a different latent structure altogether.

do not correlate across layers. Depending on the application, the main goal of analysis might be to utilize multiple layers to find communities that may not have been obvious in a single-layer slice of the network. In other applications, we may be interested in the similarities and dissimilarities of the community structure for each layer, which necessitates different approaches. Community detection in temporal networks deserves its own special treatment, as we often make temporal locality assumptions that allow for a more focused analysis.

We will briefly cover three types of methods for multilayer community detection: score-based methods, model-based methods, and aggregation methods. This list is by no means exhaustive nor are the types of methods meant to be canonical. Rather, we find them to be useful descriptors that tend to have reasonably well understood advantages and disadvantages in the multilayer setting. The main goal of aggregation methods is to find shared community structure by combining each slice of the multilayer graph into a single-layer network. Score-based methods rely on maximizing fitness functions based on an appropriate null model in order to detect communities. Finally, model-based methods rely on statistical models and formal inference to discover latent structure. These three types of methods are not necessarily disjointed from one another; for instance, many in the statistics community study models that are very similar to the null models that are used in score-based methods.

25.4.1 SCORE-BASED METHODS

In the single-layer setting, score-based methods operate by optimizing a fitness function. Perhaps the most popular method in this category is modularity maximization. Modularity for a single-layer network is defined as follows:

$$Q = \frac{1}{2m} \sum_{i,j \in \mathcal{E}} \left(\mathbf{A}_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j), \tag{25.17}$$

where m is the number of edges in the network, $\mathbf{A}$ is the adjacency matrix, k_i is the degree of node i, and c_i is the community label of node i. Modularity is qualitatively a comparison with the structure of the network to a random null model in which every edge between every node is equally likely [29]. Extensions of modularity, such as multiresolution variants [30], have been proposed in the literature.

In order to find community structure, we perform a maximization of Q over the community assignments c_i. Modularity maximization is typically performed using the Louvain algorithm or its appropriate variants. Modifying these fitness functions for a multilayer setting can be done by appropriately defining a null model [31], which takes into account intralayer connections and interlayer connections accordingly, in which case we have the following multilayer modularity:

$$\frac{1}{2\mu} \sum_{i,j \in \mathcal{E}_M, \alpha, \beta \in [L]} \left[\left(\mathbf{A}_{ij}^{(\alpha)} - \gamma_\alpha \frac{k_i^{(\alpha)} k_j^{(\alpha)}}{2m_\alpha} \right) \delta(\alpha, \beta) + \delta(i,j) C_{j\alpha\beta} \right] \delta(c_{i\alpha}, c_{j\beta}). \tag{25.18}$$

This model only takes into account interlayer connections between the same nodes, and their strengths are represented by $C_{j\alpha\beta}$. Further, each node in each layer has a different community label $c_{i\alpha}$, and μ is an appropriate normalization term; see [31] for details.

Another score-based method that allows for extensions to any single-layer fitness function involves Pareto optimality [32]. In this case, we assume that each node has one community label for every layer, so that $c_i = c_{i\alpha} = c_{i\beta}$. In this method, we define a fitness function for each layer, $f_1(\mathbf{c}), f_2(\mathbf{c}), \ldots, f_L(\mathbf{c})$, that we wish to jointly minimize. We could, for instance, choose (negative) modularity on each layer for our cost function. Alternatively, we could choose a similar cost function that arises when attempting to reduce the intercommunity connections—this is called spectral clustering [33]. Once we define these functions, we attempt to solve the multiobjective optimization problem:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} [f_1(\mathbf{c}), f_2(\mathbf{c}), \ldots, f_L(\mathbf{c})]. \tag{25.19}$$

The objective is to find the Pareto optimal solution or solutions. A non-Pareto optimal solution $\mathbf{c}$ is a solution such that there exists at least one other solution $\mathbf{d}$ such that, for all $\alpha \in [L]$, $f_\alpha(\mathbf{d}) \leq f_\alpha(\mathbf{c})$, and $f_\beta(\mathbf{d}) < f_\beta(\mathbf{c})$ for at least one layer $\beta \in [L]$. The set of Pareto points is the set of solutions for which the above is not true. The special case of using the spectral clustering score function has been explored in [32]. Other methods for finding approximate Pareto optimal points include evolutionary algorithms, and Pareto methods have been used in anomaly detection [34] and image retrieval [35].

25.4.2 MODEL-BASED METHODS

Model-based methods assume a specified statistical model for the network, and then use statistical methods for inference in order to discover the latent community structure. These models are often variants of a ubiquitous single-layer model called the stochastic block model (SBM) [36]. This model assumes that given the community structure, each edge is drawn independently as a Bernoulli random variable according to a parameter p_{ij}, where i and j are the communities of the nodes for the edge that is being drawn. Reference [37] generalize the SBM to have discrete types of layers and communities in each type. References [38–40] explore different extensions of the single-layer SBM. The inference for these models can be quite difficult from both a computational and statistical perspective. Work to find provably computationally and statistically efficient algorithms in various model cases continues to be an active field of research in the multilayer setting.

25.4.3 AGGREGATION-BASED METHODS

Aggregation-based methods attempt to find a single-layer network that holds information about the communities in the multilayer network, and then utilize single-layer community detection methods. Examples include [41–43]. A recent paper [7] utilizes spectral clustering and convex layer aggregation to perform community detection. Specifically, given a layer weight vector $w \in \mathcal{W}_L$, where $\mathcal{W}_L = \{w : w_\alpha \geq 0, \sum_{\alpha=1}^{L} w_\alpha = 1\}$, and a supra-adjacency matrix as defined in Section 25.2, we define the weighted adjacency layer matrix and associated Laplacian as:

$$\mathbf{A}_w = \sum_{\alpha=1}^{L} w_\alpha \mathbf{A}^{(\alpha)}, \quad \mathbf{L}_w = \sum_{\alpha=1}^{L} w_\alpha \mathbf{L}^{(\alpha)}. \tag{25.20}$$

The authors in [7] discuss theoretical guarantees and limits of this method under different models, and also provide a framework for model selection.

25.5 ESTIMATION OF DYNAMIC SOCIAL INTERACTION NETWORKS

With social media data, we are often faced with the situation where given some features over time, we would like to infer an interaction graph among the agents. The network structure we consider is a dynamic graph, which as discussed above can be thought of as a multilayer graph, with each layer corresponding to a discrete time point or a series of discrete time points.

In the stationary case, under Gaussian assumptions, this problem is related to structural estimation in Gaussian graphical models (GGM) [44–46]. The extension of GGM models to the time-varying setting is an active research area [47,48]. It can be especially valuable for understanding the interaction of multiple agents over time.

In order to capture these time-varying interactions on the network, we turn to using information theoretic measures. Specifically, directed information (DI) will be used as a measure of influence among these nodes and edges. DI was originally introduced as an extension of mutual information for a channel that exhibits feedback. Consider $\mathbf{X}^{(T)} = [X_1, X_2, \ldots, X_T]$ and $\mathbf{Y}^{(T)} = [Y_1, Y_2, \ldots, Y_T]$ to be features at two respective nodes at time periods $1, \ldots, T$. Then, the DI between $\mathbf{X}^{(T)}$ and $\mathbf{Y}^{(T)}$ is defined as:

$$I(\mathbf{X}^{(T)} \to \mathbf{Y}^{(T)}) = \sum_{i=1}^{T} I(\mathbf{X}^{(i)}; Y_i | \mathbf{Y}^{(i-1)}) \tag{25.21}$$

$$= H(\mathbf{Y}^{(T)}) - H(\mathbf{Y}^{(T)} || \mathbf{X}^{(T)}), \tag{25.22}$$

where $H(\mathbf{Y}^{(T)}||\mathbf{X}^{(T)})$ is defined as the causally conditioned entropy:

$$H(\mathbf{Y}^{(T)}||\mathbf{X}^{(T)}) = \sum_{i=1}^{T} = H(Y_i|\mathbf{X}^{(i)}, \mathbf{Y}^{(i-1)}). \tag{25.23}$$

This measure has been utilized extensively in information theory [49] as well as fMRI and EEG interaction studies [50–52], financial time-series analysis [53], and video indexing and retrieval [54]. It has proven robust in detecting nonlinear relationships among objects of interest over time.

The disadvantage of DI, as with most information theoretic measures, is the lack of ability to track changes over time. Most of these estimators assume stationary processes, and estimate (either parametrically or nonparametrically) the distributions and subsequently the functionals of the distributions. With this in mind, we introduce adaptive measures for directed information that allow for changes in the distributions of interactions over time. The adaptivity is a two-stage process; we first must allow for the parameters to be adaptable, and then also allow for the functional to adapt over time. In [55], we use an exponentially weighted filter applied to the DI:

$$(ADI_{X^t \to Y^t}) = \sum_{i=1}^{t} g(t,i) I(\mathbf{X}^{(i)}; Y_i | \mathbf{Y}^{(i-1)}), \tag{25.24}$$

where $g(t,i) = e^{(t-i)\lambda} c_t$. Under (time-varying) Gaussian assumptions, we can simplify ADI in terms of the varying covariance matrices:

$$(ADI_{X^t \to Y^t}) = \frac{1}{2} \sum_{i=1}^{t} g(t,i) \log \frac{\left|\Sigma_{Y_i|Y_{i-1}}\right|}{\left|\Sigma_{Y_i|\mathbf{Y}^{(i-1)},\mathbf{X}^{(i-1)}}\right|}. \tag{25.25}$$

In this case, we must estimate these covariances, for instance using a dynamic covariance model [56]. In other cases, such as the discrete case, other methods can be used to parametrically estimate the appropriate distributions. After the estimation of ADI, we can define our dynamic interaction graph at each time step with a directed weighted edge between each two nodes with nonzero ADI; functional p-value transformations for thresholding can also be used [55,57].

25.6 APPLICATIONS

In the following sections, we will demonstrate the utility of multilayer network methods on real data. First, we will examine a biological multilayer network to uncover topological roles in gene contact networks. We will also describe a Twitter dataset, and use the dynamic interaction graph estimation technique discussed in Section 25.5 to uncover novel interactions between US senators.

25.6.1 IDENTIFYING GENES ENCODING ALLELIC DIFFERENCES IN GENE CONTACT NETWORKS

Allelic differences between two homologous chromosomes (corresponding to paternal and maternal alleles) can affect the propensity of inheritance in humans [58]. Therefore, it is important to discriminate the contribution of the paternal (Pat) and maternal (Mat) genomes to the functional diploid human nucleome. In what follows, we perform multilayer network analysis to understand allelic differences at the gene level.

Genome technologies such as genome-wide chromosome conformation capture (Hi-C) can be used to measure the genomic structure [59–61]. Here, Hi-C evaluates long-range interactions between pairs of segments delimited by specific cutting sites using spatially constrained ligation [59]. As a result, we obtain a fragment read table, each row of which indicates a ligated pair of fragments from the genome with the coordinates of both fragments. Based on that, we can construct two-dimensional Hi-C contact maps at gene resolution [16,62]. We refer the reader to [62] for more details on data generation and preprocessing. From the network point of view, this leads to a sequence of intergene interaction networks over time (namely, cell cycle phases G1, S, and G2/M) under both Pat and Mat alleles. That is, we obtain an allele-specific multilayer network where each cell cycle stage corresponds to a layer. Our goal is to identify genes that yield significant contact differences between the Pat and Mat alleles.

We adopt the overlapping degree centrality and the multiplex participation coefficient to distinguish Pat allele from Mat allele. We recall from Section 25.3.1 that the overlapping degree centrality allows us to identify hubs from a network, and the multiplex participation coefficient can quantify the participation of a gene to different cell cycle phases. In Fig. 25.3B, we present z-scores of genes' overlapping degrees versus genes' participation coefficients. As we can see, due to allelic differences, there exist genes that play different topological roles on Pat and Mat alleles. Let z_i denote the z-score of the overlapping degree for gene i, and P_i denote its multiplex participation coefficient. We distinguish hubs (interacting with many genes) from regular nodes if $z_i \geq 2$. Motivated by [6], we call genes focused if contacts associated with them were concentrated on a single cell cycle phase, corresponding to $P_1 < 1/3$, and multiplex if their connected edges were homogeneously distributed across different cell cycle phases, corresponding to $P_1 > 2/3$. In the considered experiment, genes LEPREL1 and CTSS are hubs at Pat allele while they become regular nodes at Mat allele. And gene KBTBD2 is a

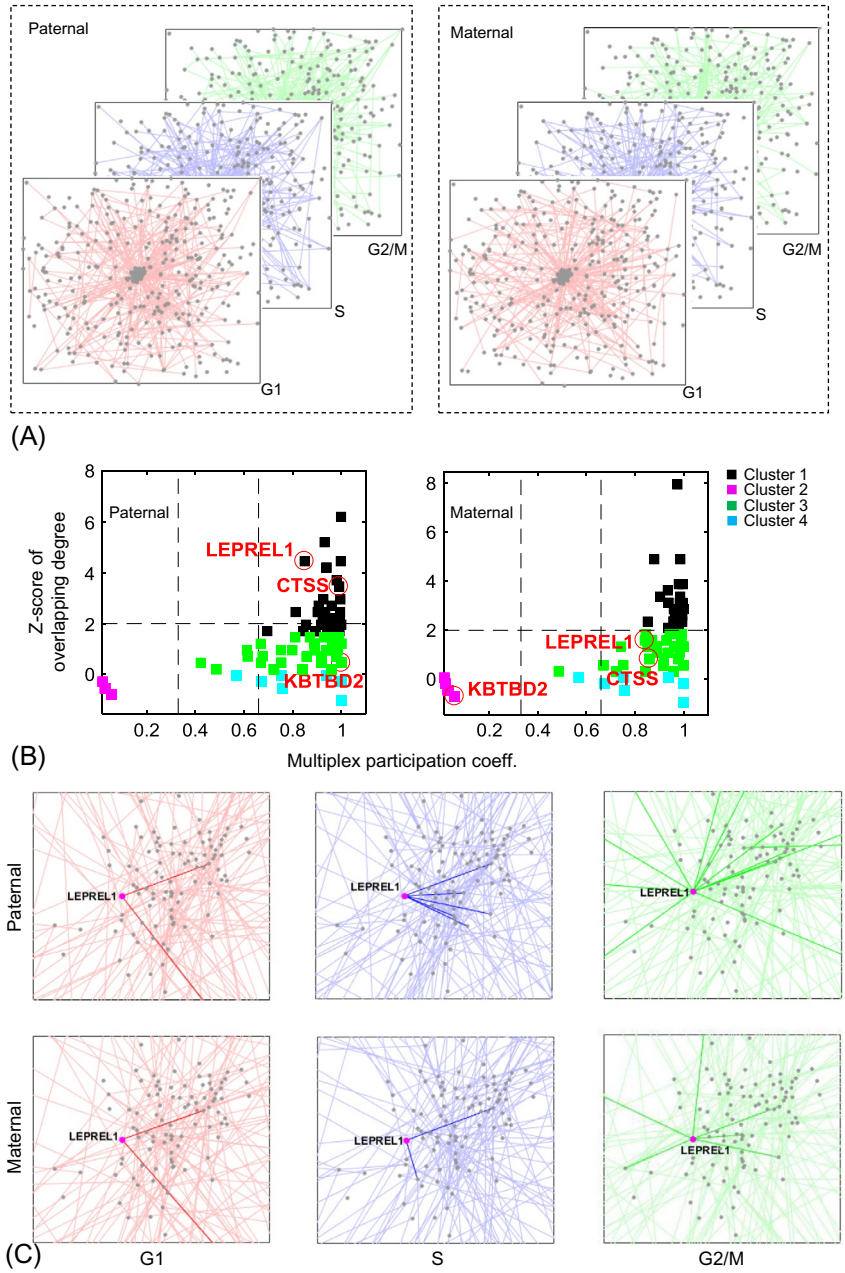

FIG. 25.3

Allele-specific intergene contact networks over different cell cycle phases. (A) Temporal network with implicit interlayer connections between genes at one cell cycle phase and their counterparts at other cell cycle phases. (B) Overlapping degree versus multiplex participation coefficient: genes are divided into 4 clusters via K-means. (C) Representative gene LEPREL1 with allelic differences in the topological structure.

multiplex node at Pat allele, but it becomes a focused node at Mat allele. We show the allelic differences in terms of contact differences of genes, e.g., LEPREL1 in Fig. 25.3C.

25.6.2 INTERACTION NETWORKS IN PRESIDENTIAL AND SENATORIAL DATASETS

In what follows, we demonstrate the method of interaction graph estimation on Twitter datasets, using the measures described in Section 25.5. The dataset is of the members of the US Senate, from October 1, 2015 to January 13, 2016. In total, the dataset consists of 96,090 tweets. In [55], adaptive directed information (ADI) was used to show time-varying interaction structure between the Twitter accounts. This is an extension of that work. Fig. 25.4 displays updated versions of ADI graphs at consecutive timesteps. Some senators are not displayed as they have no significant edges. We notice that there are nodes of high activity such as RB (Rob Bishop) and MK (Marcy Kaptur). Further, we see significant evolution in the network, with nodes adapting their behavior; this shows the method's ability to estimate changes in influence.

In Fig. 25.5 we plot the total degrees over time of ADI for a subset of senators. Total degree for a particular node is defined as the out-degree (sum of outgoing ADI) minus the in-degree (sum of incoming ADI). These senators were chosen to show examples of nodes that have high average influence (large positive total degree), senators that receive influence approximately equal to the amount they influence (small total degree), and senators that are recipients of influence on average

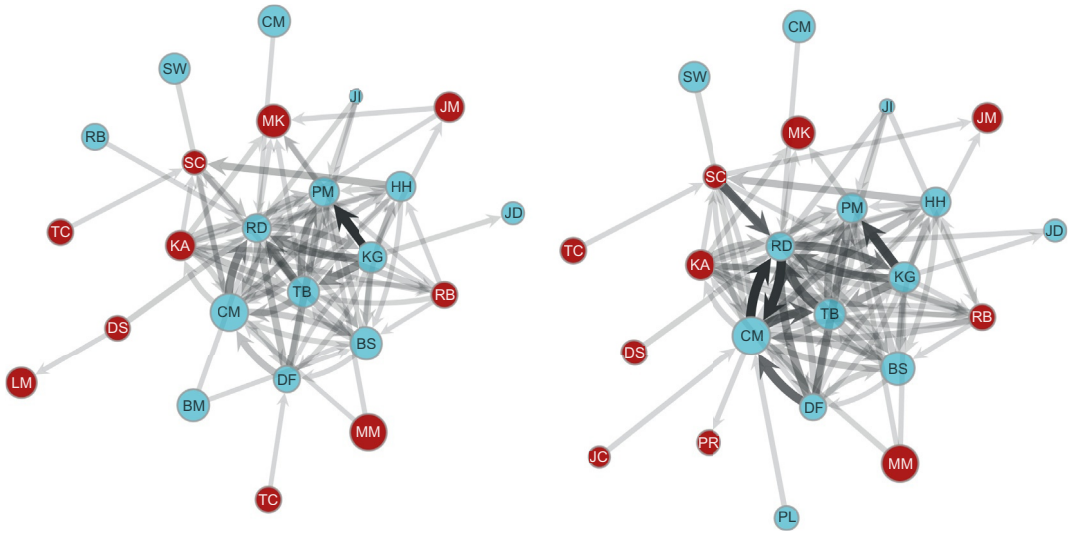

FIG. 25.4

ADI network of US Senators over two consecutive time periods from left to right. The width of the directed edge as well as the shade is related to the magnitude of the DI, and the size of each node represents the volume of tweets. We see a large connected component exhibiting mutual interaction, and significant evolution in the network, with nodes adapting their behavior.

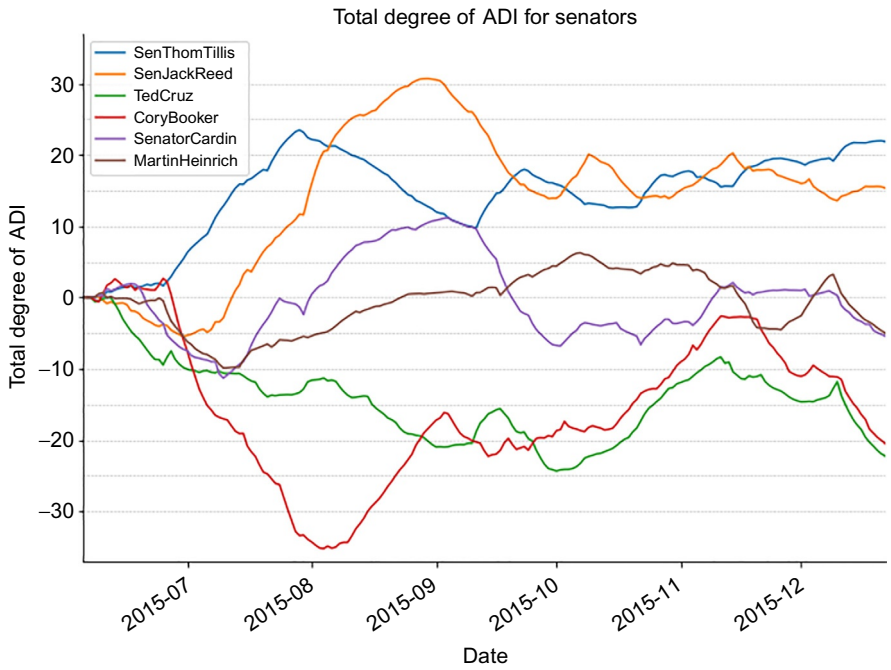

FIG. 25.5

Total degree of ADI network of a subset of US Senators. These senators were chosen as representative of senators that are high influencers (SenThomTillis, SenJackReed), average senators (SenatorCardin, MartinHeinrich), and senators that are high receivers (TedCruz, CoryBooker). We note that there is large variation over time of the total degree for each of these senators.

(large negative total degree). We note that in all cases ADI captures variation in degree over time. This is compared to total degree computed using DI, which is not sensitive to these temporal effects.

25.7 CONCLUSIONS

Social network datasets are ubiquitous in today's data landscape. We have discussed in this chapter some methods for dealing with multilayer social networks, and some of the difference when analyzing a single-layer network. After defining some formulations and representations of multilayer networks and some common examples that one might encounter, we covered some measures of multilayer network centrality. We also discussed a few types of methods for multilayer community detection, including briefly discussing some benefits and drawbacks of each type. We finally covered the problem of multilayer interaction graph estimation, with special focus on dynamic graphs. We then applied a few of the techniques to two datasets, a biological dataset and finally a social network dataset.

As the field of multilayer networks continue to grow, we expect that the methods that we have summarized here will continue to evolve and improve, and that the framework of multilayer graphs will become even more useful to the field of social network analysis in the years to come.

ACKNOWLEDGMENTS

This work was partially supported by the following grants: USAF FA8650-15-D-1845 and by ARO W911NF-15-1-0479. We also would like to thank Prof. Indika Rajapakse at Department of Computational Medicine and Bioinformatics, University of Michigan, who provided data on allele-specific Hi-C contact maps.

REFERENCES

[1] Chung FRK. Spectral graph theory, No. 92. American Mathematical Society; 1997.
[2] Wang D, Wang H, Zou X. Identifying key nodes in multilayer networks based on tensor decomposition. Chaos: Interdisciplinary J Nonlinear Sci 2017;27(6):063108.
[3] Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. J Complex Netw 2014;2(3):203–71.
[4] Cozzo E, Kivelä M, De Domenico M, Solé-Ribalta A, Arenas A, Gómez S, et al. Structure of triadic relations in multiplex networks. New J Phys 2015;17(7):073029.
[5] Sole-Ribalta A, De Domenico M, Kouvaris NE, Díaz-Guilera A, Gómez S, Arenas A. Spectral properties of the Laplacian of multiplex networks. Phys Rev E 2013;88(3):032807.
[6] Battiston F, Nicosia V, Latora V. Structural measures for multiplex networks. Phys Rev E 2014;89(3):032804.
[7] Chen PY, Hero AO. Multilayer spectral graph clustering via convex layer aggregation: theory and algorithms. IEEE Trans Signal Inf Process Netw 2017;3(3):553–67.
[8] De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y, Porter MA, et al. Mathematical formulation of multilayer networks. Phys Rev X 2013;3(4):041022.
[9] Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Rev 2009;51(3):455–500.
[10] Gauvin L, Panisson A, Cattuto C. Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach. PLOS One 2014;9(1):e86028.
[11] Sun Y, Yu Y, Han J. Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2009. p. 797–806.
[12] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of things (IoT): a vision, architectural elements, and future directions. Futur Gener Comput Syst 2013;29(7):1645–60.
[13] Cai D, Shao Z, He X, Yan X, Han J. Community mining from multi-relational networks. In: European conference on principles of data mining and knowledge discovery. Springer; 2005. p. 445–452.
[14] Yuan Z, Zhao C, Wang WX, Di Z, Lai YC. Exact controllability of multiplex networks. New J Phys 2014;16(10):103036.
[15] Aleta A, Meloni S, Moreno Y. A multilayer perspective for the analysis of urban transportation systems. Sci Rep 2017;7:44359.
[16] Liu S, Chen H, Ronquist S, Seaman L, Ceglia N, Meixner W, et al. Genome architecture leads a bifurcation in cell identity. bioRxiv; 2017. p. 151555.
[17] Newman M. Networks: an introduction. Oxford University Press; 2010.
[18] Bertrand A, Moonen M. Seeing the bigger picture: how nodes can learn their place within a complex ad hoc network topology. IEEE Signal Process Mag 2013;30(3):71–82.

[19] Bródka P, Kazienko P, Musiał K, Skibicki K. Analysis of neighbourhoods in multi-layered dynamic social networks. Int J Comput Intell Syst 2012;5(3):582–96.
[20] Bródka P, Skibicki K, Kazienko P, Musiał K. A degree centrality in multi-layered social network. In: International conference on computational aspects of social networks (CASoN). IEEE; 2011. p. 237–42.
[21] Guimera R, Amaral LAN. Functional cartography of complex metabolic networks. Nature 2005;433(7028): 895.
[22] Kleinberg JM. Authoritative sources in a hyperlinked environment. J ACM (JACM) 1999;46(5):604–32.
[23] Gleich DF. Pagerank beyond the web. SIAM Rev 2015;57(3):321–63.
[24] Taylor D, Myers SA, Clauset A, Porter MA, Mucha PJ. Eigenvector-based centrality measures for temporal networks. Multiscale Model Simul 2017;15(1):537–74.
[25] Solá L, Romance M, Criado R, Flores J, García del Amo A, Boccaletti S. Eigenvector centrality of nodes in multiplex networks. Chaos: Interdisciplinary J Nonlinear Sci 2013;23(3):033131.
[26] Kolda TG, Bader BW, Kenny JP. Higher-order web link analysis using multilinear algebra. In: Fifth IEEE international conference on data mining; 2005. p. 8.
[27] Fortunato S, Hric D. Community detection in networks: a user guide. Phys Rep 2016;659:1–44. https://doi.org/10.1016/j.physrep.2016.09.002.
[28] Mcpherson M, Smith-Lovin L, Cook J. Birds of a feather: homophily in social networks. Ann Rev Sociol 2001;27:415–44.
[29] Newman MEJ. Modularity and community structure in networks Proc Natl Acad Sci U S A 2006;103(23): 8577–82. https://doi.org/10.1073/pnas.0601602103.
[30] Xiang J, Hu XG, Zhang XY, Fan JF, Zeng XL, Fu GY, et al. Multi-resolution modularity methods and their limitations in community detection. Eur Phys J B 2012;85(10). https://doi.org/10.1140/epjb/e2012-30301-2.
[31] Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. Community structure in time-dependent, multi-scale, and multiplex networks. Science 2010;328(5980):876–8. https://doi.org/10.1126/science.1184819.
[32] Oselio B, Kulesza A, Hero AO. Multi-layer graph analysis for dynamic social networks. IEEE J Sel Top Signal Process 2014;8(4):514–23.
[33] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 2000;22(8): 888–905. https://doi.org/10.1109/34.868688.
[34] Hsiao KJ, Xu KS, Calder J, Hero AO. Multicriteria similarity-based anomaly detection using Pareto depth analysis. IEEE Trans Neural Netw Learn Syst 2016;27(6):1307–21. https://doi.org/10.1109/TNNLS.2015.2466686.
[35] Hsiao KJ, Calder J, Hero AO. Pareto-depth for multiple-query image retrieval. IEEE Trans Image Process 2015;24(2):583–94. https://doi.org/10.1109/TIP.2014.2378057.
[36] Holland P, Laskey K, Leinhardt S. Stochastic blockmodels: first steps. Soc Networks 1983;5:109–37.
[37] Stanley N, Shai S, Taylor D, Mucha PJ. Clustering network layers with the strata multilayer stochastic block model. IEEE Trans Network Sci Eng 2016;3(2):95–105. https://doi.org/10.1109/TNSE.2016.2537545.
[38] Peixoto TP. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. Phys Rev E 2015;92:042807. https://doi.org/10.1103/PhysRevE.92.042807.
[39] De Bacco C, Power EA, Larremore DB, Moore C. Community detection, link prediction, and layer interdependence in multilayer networks. Phys Rev E 2017;95:042317. https://doi.org/10.1103/PhysRevE.95.042317.
[40] Han Q, Xu KS, Airoldi EM. Consistent estimation of dynamic and multi-layer block models. In: Proceedings of the 32Nd international conference on international conference on machine learning. ICML'15, vol. 37. JMLR.org; 2015. p. 1511–20. http://dl.acm.org/citation.cfm?id=3045118.3045279.
[41] De Domenico M, Nicosia V, Arenas A, Latora V. Structural reducibility of multilayer networks Nat Commun 2015;6:6864.
[42] Taylor D, Shai S, Stanley N, Mucha PJ. Enhanced detectability of community structure in multilayer networks through layer aggregation. Phys Rev Lett 2016; 116:228301. https://doi.org/10.1103/PhysRevLett.116.228301.

[43] Berlingerio M, Coscia M, Giannotti F. Finding redundant and complementary communities in multidimensional networks. In: Proceedings of the 20th ACM international conference on information and knowledge management. CIKM '11. New York, NY: ACM; 2011. p. 2181–4. https://doi.org/10.1145/2063576.2063921.
[44] Hero A, Rajaratnam B. Hub discovery in partial correlation graphs. IEEE Trans Inf Theory 2012;58(9):6064–78. https://doi.org/10.1109/TIT.2012.2200825.
[45] Meng Z, Eriksson B, Hero A. Learning latent variable Gaussian graphical models. In: Proceedings of the 31st international conference on machine learning (ICML-14); 2014. p. 1269–77.
[46] Zhou S, Rütimann P, Xu M, Bühlmann P. High-dimensional covariance estimation based on Gaussian graphical models. J Mach Learn Res 2011;12:2975–3026. http://dl.acm.org/citation.cfm?id=1953048.2078201.
[47] Kolar M, Song L, Ahmed A, Xing EP. Estimating time-varying networks. Ann Appl Stat 2010:94–123.
[48] Ahmed A, Xing EP. Recovering time-varying networks of dependencies in social and biological studies. Proc Natl Acad Sci U S A 2009;106(29):11878–83.
[49] Jiao J, Permuter HH, Zhao L, Kim YH, Weissman T. Universal estimation of directed information. IEEE Trans Inf Theory 2013;59(10):6220–42. https://doi.org/10.1109/TIT.2013.2267934.
[50] Mehta K, Kliewer J. Directed information measures for assessing perceived audio quality using EEG. In: Conference record—Asilomar conference on signals, systems and computers; 2016. p. 123–7. https://doi.org/10.1109/ACSSC.2015.7421096.
[51] Chen X, Syed Z, Hero A. EEG spatial decoding with shrinkage optimized directed information assessment. In: ICASSP 2012 proceedings; 2012. p. 577–80. https://doi.org/10.1109/ICASSP.2012.6287945.
[52] Quinn C, Coleman TP, Kiyavash N, Hatsopoulos NG. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. J Comput Neurosci 2011;30(1):17–44. https://doi.org/10.1007/s10827-010-0247-2.
[53] Permuter HH, Kim YH, Weissman T. Interpretations of directed information in portfolio theory, data compression, and hypothesis testing. IEEE Trans Inf Theory 2011;57(6):3248–59.
[54] Chen X, Hero AO, Savarese S. Multimodal video indexing and retrieval using directed information. IEEE Trans Multimedia 2012;14(1):3–16. https://doi.org/10.1109/Tmm.2011.2167223.
[55] Oselio B, Hero A. Dynamic reconstruction of influence graphs with adaptive directed information. In: Proceedings of IEEE international conference on acoustics, speech and signal processing. ICASSP; 2017. p. 5935–9.
[56] Chen Z, Leng C. Dynamic covariance models. J Am Stat Assoc 2016;111(515):1196–207. https://doi.org/10.1080/01621459.2015.1077712.
[57] Rao A, Hero AO, States DJ, Engel JD. Using directed information to build biologically relevant influence networks. J Bioinform Comput Biol 2008;06(03):493–519. https://doi.org/10.1142/S0219720008003515.
[58] Leung D, Jung I, Rajagopal N, Schmitt A, Selvaraj S, Lee AY, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature 2015;518(7539):350.
[59] Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326(5950):289–93. https://doi.org/10.1126/science.1181369.
[60] Chen H, Chen J, Muir LA, Ronquist S, Meixner W, Ljungman M, et al. Functional organization of the human 4D Nucleome. Proc Natl Acad Sci U S A 2015;112(26):8002–7.
[61] Boulos RE, Tremblay N, Arneodo A, Borgnat P, Audit B. Multi-scale structural community organisation of the human genome. BMC Bioinformatics 2017;18(1):209.
[62] Chen H, Liu S, Seaman L, Najarian C, Wu W, Ljungman M, et al. Parental allele-specific genome architecture and transcription during the cell cycle. bioRxiv; 2017. p. 201715.