Abstract

We propose a method for simultaneously detecting shared and unshared communities in heterogeneous multilayer weighted and undirected networks. The multilayer network is assumed to follow a generative probabilistic model that takes into account the similarities and dissimilarities between the communities. We make use of a variational Bayes approach for jointly inferring the shared and unshared hidden communities from multilayer network observations. We show the robustness of our approach compared to state-of-the-art algorithms in detecting disparate (shared and private) communities on synthetic data as well as on real genome-wide fibroblast proliferation dataset.

1 Introduction

Community detection in networks is an ubiquitous area of active research as it enables exploration of network structural properties in real-world scenarios including, but not limited to, social and biological sciences [20]. There is a vast literature on single-layer network community detection using modularity optimization [32], spectral clustering [33] and statistical inference [22]. However, single-layer networks are not well suited to real-world networks, such as the Internet of Things (IoT), transportation, social, and biological networks where multi-relational, multidimensional, multiplex or multilayer network structure exists [3]. In social networks, for example, large company employees may be related to each other by similarity of their activities (functional relations) within the organization, and also by sharing the same office location (spatial relations) [30]. As another example,
in genomics, genes might be related either by transcriptional interactions (functional relations), e.g., measured by RNA-seq profile similarity, or by chromatin interactions (spatial relations), e.g., measured by chromatin conformation capture (Hi-C) of promoter-enhancer ligations \[7, 17, 19\]. In each of these examples, there may be a community structure that is common to each layer and a structure that is distinct between layers. Such multilayer networks create a need for new community detection methods \[9, 26, 28, 46, 50\]. Recovering the communities from each network independently is suboptimal as this strategy does not exploit the shared information across the network layers. Thus, current research efforts aim at developing joint inference methods by multilayer aggregation\[14, 15, 34\]. The simplest aggregation approach is to collapse the multilayer network to a single-layer network on which classical community detection methods can be applied \[2, 13, 16, 27, 45, 47, 54\]. Alternatively, some researchers have suggested performing community detection separately in each layer followed by consensus aggregation of the communities across layers \[2, 31, 35, 38, 51\]. Another approach is to extend single-layer Stochastic Block Models \[24\] to multilayer networks \[3, 21, 37, 39, 41, 42, 43, 48\].

In many applications, some communities might be shared between the different layers, while others might not be (see Figure 2a). However, few methods in the literature explicitly consider this general scenario. The Multilayer Extraction algorithm proposed in \[49\] allows identification of heterogeneous multilayer communities where the communities might be shared between a subset of layers. Wilson et al. \[49\] minimize a cost function and take into account similarities and dissimilarities between the layers’ communities. While the model used in \[49\] is realistic when considering multilayer graph connections, the method is only limited to unweighted graphs. The approach proposed in \[6\] extend \[52\] to weighted multilayer graphs and allows the extraction of coherent dense subgraphs (cliques) shared by subsets of layers. However, those methods are limited to identification of dense communities and might fail when the communities are connected but with only a few edges. We propose a new model-based method to simultaneously detect shared and unshared communities between heterogeneous weighted networks. We define joint weighted stochastic block models (WSBM) that share “a part” of their community structures. We develop a mean field variational Bayes approach to infer the latent shared and private communities from the proposed multilayer WSBM. This extends the works in \[1, 53\] devised for WSBM in single-layer graphs.

Our contributions are

- We derive a variational Bayes algorithm for automatically inferring shared and unshared communities from multilayer weighted graphs.
- We establish that the proposed algorithm is more accurate and robust than previous approaches to community detection in multi-layer networks in extracting both shared and unshared communities from weighted graph benchmarks.
- We illustrate a real-world use of our method in multinomic molecular biology where it enables discovery of heterogeneous multilayer communities of gene-gene interactions in human fibroblast proliferation.

Notations: Vectors are denoted with boldface lowercase letters and matrices by boldface uppercase letters. \(I_n \in \mathbb{R}^{n \times n}\) is the identity matrix and \(1_n\) is the column vector full of ones.

2 Joint Weighted Stochastic Block Models

To start with, let us recall the definition of a single-layer Stochastic Block Model generated by a weighted distribution \(D\) with sufficient statistic function \(T\) and natural parameter function \(\eta\). Given latent community label \(g_i \in \{1, \ldots, K\}\) (with \(K\) denoting the number of communities) of each vertex \(i\) and a community-wise connectivity matrix \(\theta \in \mathbb{R}^{K \times K}\), an edge is placed between two vertices \(i\) and \(j\) with an adjacency weight \(A_{ij}\) such that

\[
\mathbb{P}(A_{ij}|g_i, g_j, \theta_{g_i, g_j}) \propto \exp \left\{ T(A_{ij}) \eta(\theta_{g_i, g_j}) \right\}.
\]

Following a Bayesian approach, prior distributions are attributed to the labels \(g_i\) and the community-wise connectivity matrix \(\theta\).

We denote a multilayer graph, \(\mathcal{G}\), defining \(L\) as the number of layers and \(n\) as the number of vertices. The graph in the \(l\)-th layer is an undirected (possibly weighted) graph \(\mathcal{G}^{(l)} = (\mathcal{V}, \mathcal{E}^{(l)})\) with \(\mathcal{V}\).
We propose the following generative heterogeneous community structure of the multilayer graph $A$.

The overall prior distribution can thus be written as

1. We assume that each layer is subdivided into $K^{(l)}$ non-overlapping communities among which the first $K \leq \min_l K^{(l)}$ are shared between the layers as described below.
2. We first generate the label $g_i^{(1)}$ of each vertex $i$ in the first layer as $g_i^{(1)} \sim \text{Multinomial}(\mu_0^{(1)})$ where $\mu_0^{(1)} \in \mathbb{R}^{K^{(1)}}$ contains prior probabilities that the vertices belong to one of the $K^{(1)}$ communities.
3. For each vertex $i$, if $g_i^{(1)} \in \{1, \ldots, K\}$ then set $g_i^{(l)} = g_i^{(1)}$ for each layer $l$. Otherwise, generate for each layer $l$, $g_i^{(l)} \sim \text{Multinomial}(\mu_0^{(l)})$.
4. Given latent community labels $g_i^{(l)}$ (generated in steps 2 and 3) of each vertex $i$ and community-wise connectivity matrices $\theta^{(l)} \in \mathbb{R}^{K^{(l)} \times K^{(l)}}$ (the generation of which will be defined later), an edge is placed between two vertices $i$ and $j$ and it is assigned an adjacency weight $A_{ij}^{(l)}$ drawn according to

$$P(A_{ij}^{(l)}|g_i^{(l)}, g_j^{(l)}, \theta^{(l)}) \propto \exp \left\{ T^{(l)}(A_{ij}^{(l)}; \eta^{(l)}(g_i^{(l)}, g_j^{(l)})) \right\} \tag{1}$$

where $T^{(l)}$ is the sufficient statistic function and $\eta^{(l)}$ is the natural parameter function of the weights distribution.
5. The community-wise connectivity matrices $\theta^{(l)}$ are generated according to conjugate priors associated with the distribution characterized by $(T^{(l)}, \eta^{(l)})$ i.e.,

$$p^*(\theta^{(l)}_{ab}) = \frac{1}{Z^{(l)}(\tau_0^{(l)})} \exp(\tau_0^{(l)} \eta^{(l)}(\theta^{(l)}_{ab}))$$

with $\tau_0^{(l)}$ denoting the associated hyperparameters and $Z^{(l)}(\tau_0^{(l)})$ the normalization constants.

For illustration, we specialize the presentation to two communities, for which each of the matrices $\theta^{(l)}$ are decomposed into four blocks corresponding, respectively, to the shared-shared, shared-private, private-shared, private-private interconnections (see Figure 2). As in [1], we consider each sub-matrix $\theta_1^{(l)}, \theta_2^{(l)}, \theta_3^{(l)}, \theta_4^{(l)}$ as one-dimensional vectors where the elements are stacked. Let us denote by $r_1^{(l)}, r_2^{(l)}, r_3^{(l)}, r_4^{(l)}$ the indexing variables into each of the obtained vectors i.e., $r_1^{(l)} = 1, \ldots, K^2; r_2^{(l)} = 1, \ldots, K(K^{(l)} - K); r_3^{(l)} = 1, \ldots, K(K^{(l)} - K); r_4^{(l)} = 1, \ldots, (K^{(l)} - K)^2$. The overall prior distribution can thus be written as

![Figure 1: Generative graphical model. Circles and rectangles represent random and deterministic (parameters) variables respectively. Observed variables are shaded.](image)
where the correlations constraints on $g$ are $\text{N-P hard due to two main difficulties: the maximization is over all possible configurations of } g$, the calculation of the posterior distribution $P(g^{(l)})$, the goal is to infer the community labels $g^{(l)}$ for each node $i$ in each layer $l$ i.e., to find the most probable clustering $g^{(l)}$ of the vertices in each layer in the set of all different possible partitioning
\[
[(g^{(1)})^*, \cdots, (g^{(L)})^*] = \arg\max_{g^{(l)}, l=1, \ldots, L} P(g^{(l)}|A^{(l)}, \theta^{(l)}, l = 1, \ldots, L) \tag{3}
\]
with the correlations constraints on $g^{(l)}$ defined in Point 3 of Section 2. The optimization problem (3) is N-P hard due to two main difficulties: the maximization is over all possible configurations of $g^{(l)}$, the calculation of the posterior distribution $P(g^{(l)}|A^{(l)}, \theta^{(l)})$, which is intractable due to its high dimensional integral form.

Our approach to the optimization (3) is the mean field variational Bayes approximation [4,25] that uses a factorisable distribution as an approximation to the joint posterior $p(g^{(l)}, \theta^{(l)}) = P(g^{(l)}, \theta^{(l)}|A^{(l)})$.

3 Variational inference

3.1 Mean field variational Bayes inference

Denote by $q(g^{(l)}, \theta^{(l)})$ an approximating (factorisable) distribution that depends on tunable shaping parameters $\mu^{(l)}$ and $\tau^{(l)}$. The variational Bayes algorithm fits the distribution $q$ to the joint distribution by minimizing the KL-divergence, i.e., $q = \arg\min_{q} D_{KL}(r||p)$.

Here the distribution $q$ is taken to have the same parametric form as the prior $p^*$
\[
q(g^{(l)}, \theta^{(l)}, l = 1, \ldots, L) = \prod_{l=1}^{L} \prod_{i} \mu_{l,g^{(l)}}^{(l)} \prod_{r^{(l)}} \frac{1}{Z(l)(\tau_{r^{(l)})} \exp(\tau_{r^{(l)} \eta^{(l)}(\theta_{r^{(l)})}) \tag{4}
\]
where $\tau_{r^{(l)}}, \mu^{(l)} \in \mathbb{R}^{n \times K^{(l)}}$ are variational parameters corresponding to the random variables $\theta_{r^{(l)}}, g^{(l)}$ respectively. We can rewrite the original problem (3) as follows
\[
[(g^{(1)})^*, \cdots, (g^{(L)})^*] = \arg\max_{g^{(l)}} \int P(g^{(l)}, \theta^{(l)} | A^{(l)}) d\theta^{(l)}
\approx \arg\max_{g^{(l)}} \int q(g^{(l)}, \theta^{(l)}) d\theta^{(l)}
= \arg\max_{g^{(l)}} \prod_{l} \prod_{i} q(g^{(l)}_{i}) q(\theta^{(l)}) d\theta^{(l)}
= \arg\max_{g^{(l)}} \prod_{l} \prod_{i} q(g^{(l)}_{i}).
\]
Since \( q^{(l)} \) is a categorical distribution with parameter \( \mu^{(l)} \), the original problem (3) is equivalent to
\[
(g^{(l)}_i)^* = \arg\max_k \mu^{(l)}_{ik}
\]
for each node \( i \) and layer \( l \), and thus, the multilayer community detection boils down to a Maximum A Posteriori (MAP) estimator on each individual nodal variational parameter \( \mu^{(l)}_i \) for each layer \( l \).

### 3.2 Learning

As per [1], the constant model likelihood can be written as \( \log \mathbb{P}(A^{(1)}) = \mathcal{G}(q) + D_{KL}(q||p) \) with
\[
\mathcal{G}(q) = \mathbb{E}_q \log \mathbb{P}(A^{(1)}, A^{(2)}|g^{(1)}, g^{(2)}, \theta^{(1)}, \theta^{(2)}) + \mathbb{E}_q \frac{p^*}{q}
\]
where \( p^* \) is the prior distribution assigned to the parameters \( g^{(l)}, \theta^{(l)} \). Since the likelihood is constant, minimizing \( D_{KL}(q||p) \) (and thus making the approximation \( q \) to be the closest to the sought posterior \( p \)) is equivalent to maximizing \( \mathcal{G}(q) \) over the variational parameters. In the sequel, we devise a procedure to learn the parameters for which \( \mathcal{G}(q) \) is maximized.

We next address how to find the variational parameters \( \tau^{(l)}, \mu^{(l)} \) for which \( \mathcal{G}(q) \) is maximized. To this end, let us first compute \( \mathcal{G}(q) \) with the forms of the prior \( p^* \) and the approximation \( q \) defined in the previous section. For illustration, we specialize to \( L = 2 \) but the same principle applies to any number of layers. We have
\[
\mathcal{G}(q) = \mathbb{E}_q \log \mathbb{P}(A^{(1)}, A^{(2)}|g^{(1)}, g^{(2)}, \theta^{(1)}, \theta^{(2)}) + \mathbb{E}_q \frac{p^*}{q} \\
= \mathbb{E}_q \log \mathbb{P}(A^{(1)}, A^{(2)}|g^{(1)}, g^{(2)}, \theta^{(1)}, \theta^{(2)}) + \mathbb{E}_q \log \mathbb{P}(A^{(2)}|g^{(1)}, g^{(2)}, \theta^{(2)}) + \mathbb{E}_q \frac{p^*}{q} \tag{7}
\]
where in the last line, we use the chain rule along with condition of conditional independence between \( A^{(1)} \) and \( A^{(2)} \) given \( g^{(1)}, g^{(2)}, \theta^{(1)}, \theta^{(2)} \). The structure of the heterogeneous Joint Stochastic Block Model (Section 2) couples the random variables \( g^{(1)} \) and \( g^{(2)} \) in a simple manner that can be decomposed into the following cases, which we call the dependency cases:

- For a vertex pair \((i, j)\) belonging to a block with \( \theta^{(l)}_1, g^{(1)}_i = g^{(2)}_i \) and \( g^{(1)}_j = g^{(2)}_j \).
- For a vertex pair \((i, j)\) belonging to a block with \( \theta^{(l)}_2, g^{(1)}_i = g^{(2)}_i \) and \( g^{(1)}_j \neq g^{(2)}_j \).
- For a vertex pair \((i, j)\) belonging to a block with \( \theta^{(l)}_3, g^{(1)}_i \neq g^{(2)}_i \) and \( g^{(1)}_j = g^{(2)}_j \).
- For a vertex pair \((i, j)\) belonging to a block with \( \theta^{(l)}_4, g^{(1)}_i \neq g^{(2)}_i \) and \( g^{(1)}_j \neq g^{(2)}_j \).

Using these dependency cases with (7), we obtain an expression for \( \mathcal{G}(q) \) as per (7). After differentiation with respect to the sought variational parameters, we obtain updates for \( \tau^{(l)}, \mu^{(l)} \), which are stationary points of \( \mathcal{G}(q) \) and correspond to local maxima. The precision of the local maxima depends on the initial values for \( \mu^{(l)} \). A single run of a single-layer clustering algorithm shall lead to the optimal solution basin of attraction. Due to the dependency cases, the community memberships variational parameters \( \mu^{(l)}_{ik} \) depends on \( g^{(l)}_i \) either belonging to the set of shared communities \( \{1, \ldots, K\} \), or to the set of unshared communities \( \{K + 1, \ldots, K^{(l)}\} \). The derivation details are provided in the supplementary material.

Algorithm 1 provides the necessary equations for the updates of the variational parameters \( \tau^{(l)} \) and \( \mu^{(l)} \). Due to Equation (5), a max decision rule can then be used on \( \mu^{(l)} \) to assign labels to each node, namely \( \arg\max_k \mu^{(l)}_{ik} \) gives the label assigned to node \( i \) in graph \( G^{(l)} \). The label of node \( i \) is shared between different graphs \( \{G^{(l)}\} \) when \( \arg\max_k \mu^{(l)}_{ik} \in \{1, \ldots, K\} \) and the label is unshared otherwise.

Algorithm 1 is an extension of the variational Bayes algorithm for inferring hidden communities from single-layer graphs [2] to the inference of hidden shared and unshared communities from multilayer graphs. In Algorithm 1, the updates for the parameters \( \tau^{(l)} \) are done independently for each graph as in [2]. As for the community membership variational parameters \( \mu^{(l)} \in \mathbb{R}^{n \times K^{(l)}} \), the
Algorithm 1 Mean field inference of heterogeneous communities in multilayer graphs.

**Inputs:** For $l = 1, \ldots, L$, layers adjacencies $A^{(l)}$, layer distributions $D^{(l)} = [T^{(l)}, \eta^{(l)}, Z^{(l)}]$, number of shared communities $K$, total number $K_l$ of communities.

**Output:** $\mu_{(1)}, \ldots, \mu_{(L)}$.

For $l = 1, \ldots, L$, initialize $\mu_{(l)}$ and choose hyperparameters $\tau_{0(l)}$.

repeat
  for $l = 1$ to $L$ do
    for $r = 1$ to $(K_l)^2$ do
      $\tau_{r(l)}(\tau) = \tau_{0(l)} + \sum_{ij} \sum_{(g_i^{(l)}, g_j^{(l)}) = r(l)} T^{(l)}(A^{(l)}_{ij}) \mu^{(l)}_{i,g_i} \mu^{(l)}_{j,g_j}$
    end for
  end for
  for $i = 1$ to $n$ do
    for $k = 1$ to $K_l$ do
      $\mu_{ik}^{(l)} = \exp \left\{ \frac{1}{L} \sum_{l=1}^{L} \left[ \sum_{r_{(l)}^{(l)} \neq i} T^{(l)}(A^{(l)}_{ij}) \mu^{(l)}_{i,g_i} \tilde{\eta}^{(l)}_{r_{(l)}}(r) + \sum_{(l) \neq i} T^{(l)}(A^{(l)}_{ij}) \mu^{(l)}_{i,g_i} \tilde{\eta}^{(l)}_{r_{(l)}}(r) \right] \right\}$
    end for
  end for
  end for

until convergence

updates of the first $K$ columns of $\mu_{(l)}$ are identical and are computed by adding the contributions of each graph. The last $K_l - K$ columns of $\mu_{(l)}$ are updated independently using only the information about each graph. This is quite intuitive since the first $K$ columns of $\mu_{(l)}$ correspond to the shared community evidences and thus they should be updated using the contributions of the graphs altogether, while the last columns correspond to unshared communities and thus the updates should be done independently for each graph.

4 Experiments

4.1 Synthetic graphs

We first consider two Bernoulli SBM graphs $G^{(1)}$ and $G^{(2)}$ with the same intra-community probabilities and different inter-community probabilities in such a way that one graph is noisier than the other. Blindly identifying the community labels from each of the graphs would yield poor performances since we do not know in advance which graph has a clearer community structure than the other.
We next consider the same graph settings as before but with disparate distributions $A_{ij}^{(1)} \sim \text{Bernoulli}(\theta^{(1)})$ and $A_{ij}^{(2)} \sim \text{Poisson}(\theta^{(2)})$. Here the M-E algorithm is not exploitable since the latter is designed only for binary graphs (Bernoulli). Our joint variational algorithm is thus compared with single-graph clustering algorithms. The results are reported in the right figure of Figure 3 where our joint algorithm outperforms single-layer clustering algorithms (spectral clustering and mean field variational Bayes).

Although our algorithm is designed for detecting shared and unshared communities from multilayer graphs, here we compare our method with the state-of-the-art algorithms designed to find only shared communities between multilayer networks which is a particular case of our model with $K^{(1)} = K^{(2)} = K$. We use the synthetic dataset $\text{ml-128}$ designed by [8] which is an extension of the LFR benchmark [29] to multilayer networks. The parameter $\mu$ characterizes the variation in the vertex degrees among layers (the higher $\mu$ the more variations in the layers vertices’ degrees). In Figure 4 by varying the number of layers $\ell$, we compare in terms of Normalized Mutual Information (NMI) for increasing $\mu$, the output of our joint mean algorithm with some state-of-the-art methods for the identification of shared communities in multilayer networks. Although our method is designed to seek for shared and unshared communities at the same time, it competes well with the PMM [44] and the MLMAOP [40] methods, only optimized for shared community detection in multi-layer networks. In addition, as shown above, our method is able to also recover unshared communities between different layers of the network.

4.2 Real world graphs

In this section, we make use of our novel approach to understand the interplay between genome structure (form) and transcription (function) based on a human fibroblast proliferation dataset [10]. This dataset consists of Hi-C contact maps [30] that capture chromatin architectures and RNA-seq...
data that provide gene expression levels over 8 time points. We first build a correlation matrix between the RNA-seq values, where thresholding is applied to obtain a binary adjacency matrix $A^{(1)}$ representing functional correspondence between different genes. The threshold was determined using the asymptotic expression in [23] for the minimal RNA-seq correlation necessary to maintain functional interaction between genes for a given number of samples $n$ (here the number of time points), the number of variables $p$ (here the number of genes in one chromosome) and a given level of significance. We then construct an average (over the 8 time points) Hi-C matrix $A^{(2)}$ and round each entry of the average matrix to the closest integer value. For the application of the variational Bayes algorithm, the entries of $A^{(1)}$ are considered to be Bernoulli distributed while those of $A^{(2)}$ are considered to be Poisson distributed. More sophisticated models for the sample correlation graph, e.g., Wishart distributions, could also be considered but this is left for future work.

It is shown in [10, 11, 18] that the genome structure per chromosome can be divided into different topologically associating domains (TADs), each of which may contain differently expressed genes. Although [10] found that some genes in a single TAD can maintain similar expression levels, it is unclear how to effectively find such a mapping between TADs and gene expression considering the fact that there are more than 22000 genes in the human genome. In Figure 5, we focus on chromosome 4 as an example in order to show how our proposed method provides an elegant way to gain insights on genomic form-function relationship. Figure 5-(a) shows Hi-C contacts and gene expressions corresponding to a subset of genes in one of the shared clusters that we found. We observe from Figures 5-(a) that the genes in a shared cluster aggregate in one TAD, which indicates their frequent interactions. In this shared cluster, the same group of genes have very similar expressions. Our results confirm the biological findings in [10] that co-expressed genes exist in a single TAD. Moreover, as shown in Figure 5-(b), our analysis establishes the form-function relationship for genes.
in a RNA-seq private cluster. We observe that, as compared to Figure 5-(a), fewer genes belong to the same TAD even though they are more strongly co-expressed. Finally, Figure 5-(c) shows that for a Hi-C private cluster, the genes are possibly aggregated in a small region of the chromosome but they have significantly different expression profiles. To sum up, in contrast to a single-layer community detection algorithm, our method allows to differentiate groups of genes

i) loosely co-expressed but highly interconnected,
ii) loosely interconnected but highly co-expressed
iii) highly co-expressed and highly interconnected.

5 Conclusion

Our proposed joint variational algorithm is capable of extracting shared communities across all graph layers as well as identifying communities unique to each layer. The method is applicable to any multilayer network (with or without edge weights) and can provide important insights in the analysis of real-world systems as demonstrated for the human fibroblast dataset. An interesting direction of future investigation would be to consider extensions to the case that only a subset of layers share communities.

References

[1] Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, page cnu026, 2014.

[2] Alessia Amelio and Clara Pizzuti. Community detection in multidimensional networks. In *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on*, pages 352–359. IEEE, 2014.

[3] Pierre Barbillon, Sophie Donnet, Emmanuel Lazega, and Avner Bar-Hen. Stochastic block models for multiplex networks: an application to a multilevel network of researchers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1):295–314, 2017.

[4] David M Blei, Michael I Jordan, et al. Variational inference for dirichlet process mixtures. *Bayesian analysis*, 1(1):121–143, 2006.

[5] Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesús Gómez-Gardenes, Miguel Romance, Irene Sendina-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

[6] Brigitte Boden, Stephan Günnemann, Holger Hoffmann, and Thomas Seidl. Mining coherent subgraphs in multi-layer graphs with edge labels. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1258–1266. ACM, 2012.

[7] Rasha E Boulos, Nicolas Tremblay, Alain Arneodo, Pierre Borgnat, and Benjamin Audit. Multiscale structural community organisation of the human genome. *BMC bioinformatics*, 18(1):209, 2017.

[8] P Brodka and T Grecki. *mLFR Benchmark: Testing Community Detection Algorithms in Multilayer, Multiplex and Multiple Social Networks*, 2012.

[9] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han. Community mining from multi-relational networks. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 445–452. Springer, 2005.

[10] H. Chen, J. Chen, et al. Functional organization of the human 4d nucleome. *Proceedings of the National Academy of Sciences*, 112(26):8002–8007, 2015.

[11] Jie Chen, Alfred O Hero, and Indika Rajapakse. Spectral identification of topological domains. *Bioinformatics*, page btw221, 2016.

[12] Pin-Yu Chen and Alfred O Hero. Multilayer spectral graph clustering via convex layer aggregation: Theory and algorithms. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):553–567, 2017.
[13] Prakash Mandayam Comar, Pang-Ning Tan, and Anil K Jain. A framework for joint community detection across multiple related networks. *Neurocomputing*, 76(1):93–104, 2012.

[14] Caterina De Bacco, Eleanor A Power, Daniel B Larremore, and Cristopher Moore. Community detection, link prediction, and layer interdependence in multilayer networks. *Physical Review E*, 95(4):042317, 2017.

[15] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, 5(1):011027, 2015.

[16] Manlio De Domenico, Vincenzo Nicosia, Alexandre Arenas, and Vito Latora. Structural reducibility of multilayer networks. *Nature communications*, 6:6864, 2015.

[17] Job Dekker, Andrew S Belmont, Mitchell Guttman, Victor O Leshyk, John T Lis, Stavros Lomvardas, Leonid A Mirny, Clodagh C O’shea, Peter J Park, Bing Ren, et al. The 4d nucleome project. *Nature*, 549(7671):219, 2017.

[18] J. R. Dixon, S. Selvaraj, F. Yue, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[19] Jesse R Dixon, Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539):331, 2015.

[20] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[21] Qiuyi Han, Kevin Xu, and Edoardo Airoldi. Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning*, pages 1511–1520, 2015.

[22] Matthew B Hastings. Community detection as an inference problem. *Physical Review E*, 74(3):035102, 2006.

[23] Alfred Hero and Bala Rajaratnam. Large-scale correlation screening. *Journal of the American Statistical Association*, 106(496):1540–1552, 2011.

[24] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[25] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[26] Jungeun Kim and Jae-Gil Lee. Community detection in multi-layer graphs: A survey. *ACM SIGMOD Record*, 44(3):37–48, 2015.

[27] Jungeun Kim, Jae-Gil Lee, and Sungsu Lim. Differential flattening: A novel framework for community detection in multi-layer graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(2):27, 2017.

[28] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[29] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.

[30] Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009. doi: 10.1126/science.1181369.

[31] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.
[32] Mark EJ Newman. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23):8577–8582, 2006.

[33] Mark EJ Newman. Spectral methods for community detection and graph partitioning. Physical Review E, 88(4):042822, 2013.

[34] Vincenzo Nicosia and Vito Latora. Measuring and modeling correlations in multiplex networks. Physical Review E, 92(3):032805, 2015.

[35] Brandon Oselio, Alex Kulesza, and Alfred O Hero. Multi-layer graph analysis for dynamic social networks. IEEE Journal of Selected Topics in Signal Processing, 8(4):514–523, 2014.

[36] Brandon Oselio, Alex Kulesza, and Alfred Hero. Information extraction from large multi-layer social networks. In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, pages 5451–5455. IEEE, 2015.

[37] Subhadeep Paul and Yuguo Chen. Community detection in multi-relational data with restricted multi-layer stochastic blockmodel. arXiv preprint arXiv:1506.02699, 2015.

[38] Subhadeep Paul and Yuguo Chen. Null models and modularity based community detection in multi-layer networks. arXiv preprint arXiv:1608.00623, 2016.

[39] Tiago P Peixoto. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. Physical Review E, 92(4):042807, 2015.

[40] Clara Pizzuti and Annalisa Socievole. Many-objective optimization for community detection in multi-layer networks. In Evolutionary Computation (CEC), 2017 IEEE Congress on, pages 411–418. IEEE, 2017.

[41] Perla Reyes and Abel Rodriguez. Stochastic blockmodels for exchangeable collections of networks. arXiv preprint arXiv:1606.05277, 2016.

[42] Natalie Stanley, Saray Shai, dane Taylor, and Peter J Mucha. Clustering network layers with the strata multilayer stochastic block model. IEEE transactions on network science and engineering, 3(2):95–105, 2016.

[43] Tracy M Sweet, Andrew C Thomas, and Brian W Junker. Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions. Handbook on mixed membership models and their applications, pages 463–488, 2014.

[44] Lei Tang, Xufei Wang, and Huan Liu. Uncovering groups via heterogeneous interaction analysis. In Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on, pages 503–512. IEEE, 2009.

[45] Lei Tang, Xufei Wang, and Huan Liu. Community detection via heterogeneous interaction analysis. Data mining and knowledge discovery, 25(1):1–33, 2012.

[46] Wei Tang, Zhengdong Lu, and Inderjit S Dhillon. Clustering with multiple graphs. In Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on, pages 1016–1021. IEEE, 2009.

[47] Dane Taylor, Saray Shai, Natalie Stanley, and Peter J Mucha. Enhanced detectability of community structure in multilayer networks through layer aggregation. Physical review letters, 116(22):228301, 2016.

[48] Toni Valles-Catala, Francesco A Massucci, Roger Guimera, and Marta Sales-Pardo. Multilayer stochastic block models reveal the multilayer structure of complex networks. Physical Review X, 6(1):011036, 2016.

[49] James D Wilson, John Palowitch, Shankar Bhamidi, and Andrew B Nobel. Community extraction in multilayer networks with heterogeneous community structure. The Journal of Machine Learning Research, 18(1):5458–5506, 2017.

[50] Zhiang Wu, Zhan Bu, Jie Cao, and Yi Zhuang. Discovering communities in multi-relational networks. In User Community Discovery, pages 75–95. Springer, 2015.
[51] Ju Xiang, Xin-Guang Hu, Xiao-Yu Zhang, Jun-Feng Fan, Xian-Lin Zeng, Gen-Yi Fu, Ke Deng, and Ke Hu. Multi-resolution modularity methods and their limitations in community detection. The European Physical Journal B, 85(10):352, 2012.

[52] Zhiping Zeng, Jianyong Wang, Lizhu Zhou, and George Karypis. Coherent closed quasi-clique discovery from large dense graph databases. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 797–802. ACM, 2006.

[53] Anderson Y Zhang and Harrison H Zhou. Theoretical and computational guarantees of mean field variational inference for community detection. arXiv preprint arXiv:1710.11268, 2017.

[54] Zhongfeng Zhang, Qiudan Li, Daniel Zeng, and Heng Gao. User community discovery from multi-relational networks. Decision Support Systems, 54(2):870–879, 2013.