

Discovering Models from Structural and Behavioral Brain Imaging Data

Zilong Bai

University of California, Davis
Computer Science
zlbai@ucdavis.edu

Buyue Qian

Xi'an Jiaotong University
Department of Computer Science
qianbuyue@xjtu.edu.cn

Ian Davidson

University of California, Davis
Computer Science
davidson@cs.ucdavis.edu

ABSTRACT

Block models of graphs are used in a wide variety of domains as they find not only clusters (the blocks) but also interaction within and between the blocks. However, existing approaches primarily focus on *either* structural graphs (i.e. for MRI scans) *or* behavioral graphs (i.e. for fMRI scans). In both cases the block model's interaction or mixing matrix can be useful for understanding *potential* interaction (for structural graphs) and *actual* interaction (for behavioral graphs) between the blocks. In this paper we explore finding block models where there is both a structural network **and** multiple behavioral graphs. This provides significant modeling challenges, consider if there is strong behavioral connectivity but no structural connectivity between two nodes. We show why existing multi-graph settings such as multi-view learning are insufficient and instead propose a novel model to address the problem. Our method not only learns structurally *and* behaviorally cohesive blocks of nodes but also finds structurally *and* behaviorally feasible block interactions. We show in numerical evaluations on synthetic data that our method outperforms baseline approaches in recovering the ground-truth factor matrices in increasingly complex situations. We further apply our method to real-world datasets from two different domains (1) brain imaging data (a multi-cohort fMRI study) and to show its versatility (2) Twitter (following network and retweet behavior) and gain insights into the information flow and underlying generating mechanisms of these complex data.

CCS CONCEPTS

• **Information systems** → *Clustering*;

KEYWORDS

Structural-behavioral networks; brain imaging data; model discovery; block modeling

ACM Reference Format:

Zilong Bai, Buyue Qian, and Ian Davidson. 2018. Discovering Models from Structural and Behavioral Brain Imaging Data. In *KDD 2018: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3219819.3220080>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD 2018, August 19–23, 2018, London, United Kingdom

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5552-0/18/08...\$15.00

<https://doi.org/10.1145/3219819.3220080>

1 INTRODUCTION

Understanding community structure in graphs and most importantly the interaction between these communities is an important task. Common application domains include social networks [20], satellite imagery [2] and the main focus of this paper brain imaging [4, 15]. However, the graphs studied in these applications can be generalized into two categories: structural graphs (i.e. from MRI imaging) [7] which capture static relationships or behavioral graphs (i.e. from fMRI imaging) which capture behavioral relationships [18]. Thus the notion of interaction between communities differs for these two settings. For structural graphs it means the *potential* for interaction (or information exchange) and for behavioral graphs it means the *actual* interaction (or information exchange).

From the methodological perspective, block modeling [12, 21] has benefits over other graph clustering approaches [10, 13] as it simultaneously clusters the network nodes into blocks/communities and learns the inter-/intra-block interactions [4, 6]. This is valuable for brain imaging as we are interested in information flow between the blocks (i.e. coactivated regions). The traditional block model takes a single graph denoted by the adjacency matrix X and simplifies it as follows where F is an $n \times k$ indicator matrix showing the block memberships stacked column-wise whilst M is an $k \times k$ and shows the interaction between the blocks.

$$\operatorname{argmin}_{F, M} \|X - FMF'\| \quad (1)$$

The block model can be considered as finding structural equivalence in the graph: that is, it puts nodes in the same block if they have connectivity to the same nodes. Consider our earlier block model of fMRI data [4]: this work placed voxels in the same block if they had similar voxels they co-activated with and most importantly the M matrix shows information flow between the blocks.

Motivating Example and Aim. As mentioned above, regular block models are ideal for simplifying a single graph whether it be a structural graph or a behavioral graph. However, in many settings we have both structural and behavioral behavior data and it makes sense to analyze them simultaneously otherwise incorrect and incomplete results may be obtained. Consider brain imaging data. The structural graph (from MRI scans [14]) contains the known pathways between the voxels, whilst the behavioral graph (from fMRI scans) the correlations of functional activity between voxels. Previous work (including our own) focuses on block modeling of the later graph so that the blocks indicating regions of co-activating voxels and the mixing matrix the interaction (information flow) between different regions. However, by ignoring the structural graph these methods can find blocks which are **biologically not possible** as they place two voxels in the same block even though there may not be a structural pathway between them. Similarly

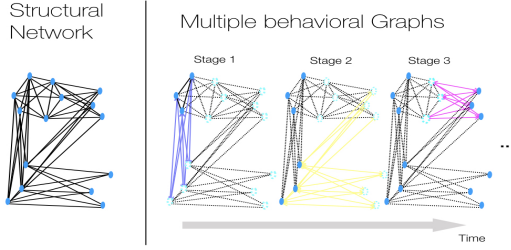


Figure 1: Structural network and multiple behavioral graphs with physically plausible correlations. In the multiple behavioral graphs, the solid lines in different colors indicate the behavioral interactions that are active at different stages along structural connections.

they may find information flow between two blocks even though there is little or no structural connectivity between them.

The aim of our work is to apply block modeling simultaneously to structural and behavioral graphs, this allows us to find new types of insights as we no longer find correlations between activities but (within the bounds of the accuracy of the structural graph) interactions which are *physically* plausible. This setting applies commonly in areas other than brain imaging analysis such as traffic data (the structural graph is the road network and the behavioral graph is start and endings of trips) and social network data (the structural graph could be the follower’s network in twitter and the behavioral graph the trajectory of the tweets through the space). To illustrate the use of our method beyond fMRI data we perform simple experiments on Twitter data.

Simple Solutions and Our Proposal. Perhaps the simplest way to model the situation with structural and behavioral graphs is as a multi-view graph setting. However, this threatens both types of graphs as being of the same kind and in our setting the structural graph plays a different role than the behavioral graph. Another simple solution is to turn the behavioral data into node labels L for the structural graph. However, extensions to block models to factor in vertex labels such as $\argmin_{F, M, G} \|X - FMF'\| + \|L - FG\|$ do not prevent the core problem of behaviorally strong but structurally unconnected nodes being in the same block due to additive nature of the formulation. Instead we propose a more powerful and elaborate setting shown in Figure 1. In addition to the structural graph we have multiple behavioral/functional graphs. These multiple behavioral graphs could be from multiple patients (where each graph represents mean behavior over time) or could be from one patient (where each graph represents the patients behavior over time). This allows us to model a range of situations but this produces several challenges. Most importantly we must directly allow the structural graph’s properties to prohibit some nodes being placed in the same block despite their behavior being similar.

We formulate the *Block Modeling from Structural and Behavioral Data* (BMSB) problem as a Linked Matrix-Tensor tri-Factorization

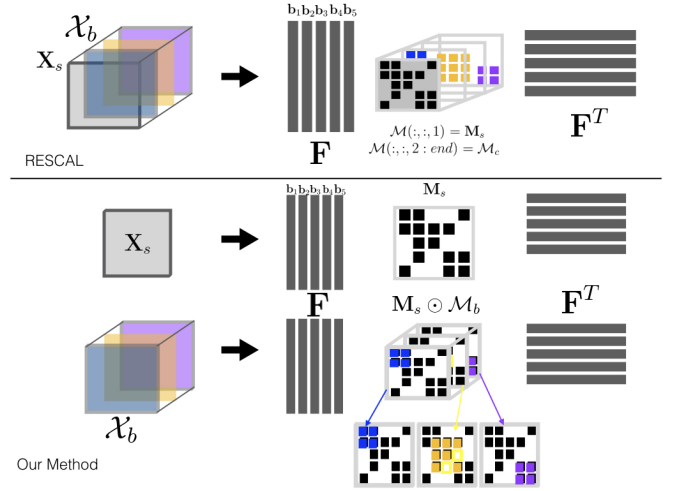


Figure 2: Our method compared to standard block models and decomposition of multiple graphs using RESCAL [11].

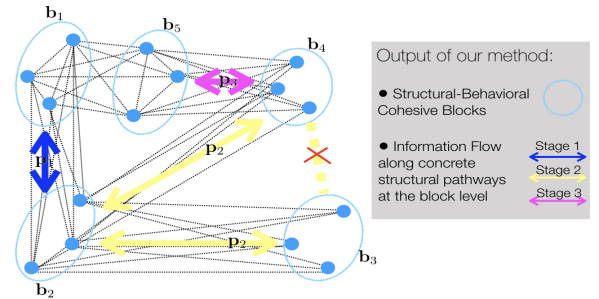


Figure 3: The expected output for the situation in Figure 1. Specifically, the blue circles indicate the blocks, each consists of nodes that are both structurally and behaviorally cohesive. The colored bi-directional arrows indicate the behavioral associations along structural pathways at different stages. Note the yellow dashed line between blocks b3 and b4 indicate their behavioral correlation as both blocks are respectively correlated to block b2. However, since there is no structural pathway between blocks b3 and b4, the behavioral correlation between b3 and b4 is excluded in our method.

(see equation 3) which not only couples the two categories of networks on their block structure F , but also explicitly requires interaction between blocks to be consistent with both the models. This is enforced by having a Hadamard product between structural pathways between the blocks M_s and the behavioral pathways M_b as shown in Figure 2 bottom). Compare this to using multiple interaction matrices such as RESCAL [11] (M_c Figure 2 top). This fundamentally changes the type of equivalence relation found by block modeling. Rather than finding structural equivalence based on one graph, two nodes are placed in the same block *if and only*

if they are structurally equivalent in **both** the structural network **and** the behavioral graphs.

The major contributions of this paper are:

- (1) Whereas previous block modeling work attempted to find blocks in a single graph we define a novel problem of discovering a block model from behavioral and structural data. This can be interpreted as a new type of equivalence relation between nodes showing realistic information flow or a specialized method for data fusion.
- (2) We design an effective solver for BMSB with novel update rules to assert the *asymmetric* dependency between the structural and behavioral mixing matrices.
- (3) We show in various use cases of synthetic data that our method outperforms existing approaches and their extensions (i.e., Orthogonal Nonnegative RESCAL with both structural and behavioral networks).
- (4) We show how our method can be used in a variety of brain imaging settings where the multiple behavioral graphs represent: i) different patients and ii) the same patient over time. This allows us to contrast information flow between cohorts.
- (5) Importantly, we show the benefit of fusing behavioral and structural data together. Our method is able to find spatially continuous blocks **without** the need for spatial regularization as was required in earlier work [4].
- (6) To illustrate the versatility of our method we explore twitter social networks.

The rest of the paper starts with discussions on related work. We formulate BMSB in section 3 and develop its solver in the following section. Section 5 presents the experimental results of our method on the aforementioned datasets. We conclude in section 6.

2 RELATED WORK

There are several fields which have similar input (multiple graphs) which we now discuss but our aims are fundamentally different.

Multi-view graph Clustering. Work including spectral clustering [22] takes as input multiple graph and attempts to find a clustering which is good for all graphs. Pareto optimization formulations [19] exist but all methods still treat each graph as being of the same type as all others. We show baseline comparisons against these methods and other community discovery methods for our Twitter experiments. These experiments are limited in size as Spectral clustering does not scale to large graphs. Whilst this provides a good baseline it treats both graphs as being of the same type where as we explicitly want to view one graph as structural and the other behavioral. Furthermore, these methods do not directly find the interaction between the blocks.

Learning Kernels from Vertex Labeled Graphs. Here the setting is a graph (i.e. our structural graph) and labels on the vertices (i.e. behavioral data) with the end result being pairwise distances which could be used for any clustering method [8]. Instead our work attempts to find block models based on both structural and behavioral behavior as shown in Figure 3.

Block Models. NMF-based approaches for block modeling and community detection have been nascent in analyzing complex networks as non-negativity enhances the models' interpretability with,

for instance, probabilistic and part-based interpretations of the discovered factor matrices. Seminal work on ONMF [6] learns complete block model (blocks structure and inter/intra-block connectivity) from single networks. Several methods exist to lift block models to address multiple graphs. RESCAL [11] and Linked Matrix Factorization (LMF) [16] are extensions to learn consistent block structure and varying block-level connectivity from multiple graphs. However, simply stacking structural network and behavioral graphs into a multi-graph fails to capture pivotal differences and relationship between them, in particular the dependency of behavioral connectivity over the structural pathways at the block level.

Our Previous Work. Two recent papers of ours explored block modeling of multiple graphs. Our recent KDD paper explored block modeling of behavioral behavior modeled as a graph combined with spatial regularization modeled as a second graph [4]. However, this method cannot model structural pathways at the block level; moreover, it encourages blocks to comprise vertices that are spatially close, while the structurally cohesive blocks can generate other connectivity patterns, such as disassortative, core-periphery, etc [1]. This formulation is shown below.

$$\begin{aligned} & \underset{\mathbf{F} \geq 0, \mathbf{M} \geq 0}{\text{Minimize}} \quad \|\mathbf{X} - \mathbf{F}\mathbf{M}\mathbf{F}'\|_F^2 + \text{Tr}(\mathbf{F}'\mathbf{F}) \\ & \text{s.t.} \quad \mathbf{F}'\mathbf{F} = \mathbf{I} \end{aligned} \quad (2)$$

Our recent paper at SDM 2018 [3] looks at simplifying multiple graphs into mixtures of block models. This work again explored only using behavioral graphs and does not use structural graphs.

3 FORMULATION

We formulate the problem of BMSB as a Nonnegative Matrix-Tensor tri-Factorization linked by blocks *and* structural pathways in equation 3. Our setting has one structural graph and multiple behavioral graphs as shown in Figure 1.

$$\begin{aligned} & \underset{\{\mathbf{F}, \mathbf{M}_s^l, \mathbf{M}_b^l\} \geq 0}{\text{Minimize}} \quad \alpha \|\mathbf{X}_s - \mathbf{F}\mathbf{M}_s\mathbf{F}'\|_F^2 + \sum_l \|\mathbf{X}_b^l - \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s)\mathbf{F}'\|_F^2 \\ & \quad + \gamma \|\mathbf{M}_s\|_1 \\ & \text{s.t.} \quad \mathbf{F}'\mathbf{F} = \mathbf{I} \end{aligned} \quad (3)$$

where

- \mathbf{X}_s : $n \times n$ adjacency matrix of the network of static structural connectivity - i.e., the structural network.
- \mathbf{X}_b^l : the l -th $n \times n$ adjacency matrix of the behavioral connectivity between the vertices of the structural network.
- \mathbf{F} : $n \times k$ cluster indicator matrix assigning the n vertices into k blocks, with each column indicating the membership of a vertex belonging to one block. Note there is just one block structure. We enforce orthogonality constraints $\mathbf{F}'\mathbf{F} = \mathbf{I}$ to discover non-overlapping blocks.
- \mathbf{M}_s : $k \times k$ mixing matrix for the structural pathways between blocks.
- \mathbf{M}_b^l : the $k \times k$ mixing matrix for the behavioral interactions over structural pathways in behavioral graph \mathbf{X}_b^l .

In the second term of the objective function of equation 3, we encode the *asymmetric* dependency between the physically plausible behavioral interactions and the structural connections. This

is achieved through the Hadamard product ($\mathbf{M}_b^l \odot \mathbf{M}_s$) in factorizing the l -th behavioral graph \mathbf{X}_b^l . That is, if $\mathbf{M}_s(i, j) = 0$, then $\mathbf{M}_b^l(i, j) * \mathbf{M}_s(i, j) = 0$, meaning no physically plausible behavioral association between blocks i and j . We incorporate L_1 norm sparsity regularization $\gamma \|\mathbf{M}_s\|_1$ to explore succinct block-level structural connectivity. This can withstand the potentially strong behavioral associations without structural connections from the perspective of numerical optimization.

4 SOLVER

The challenge with formulation (equation 3) is to incorporate \mathbf{M}_s into the discovery of both the block structure (\mathbf{F}) and the block level interactions from the behavioral multigraph through the Hadamard product between \mathbf{M}_s and \mathbf{M}_b^l . This is different from multi-view paradigms where the block level interactions from different views are discovered separately, posing challenge to learning \mathbf{M}_b^l . In this section we develop novel update rule for \mathbf{M}_b^l , $1 \leq l \leq L$ that are nonnegative and involves the \mathbf{M}_s in each iteration. We extend existing algorithms to solve for \mathbf{F} that links structural network and behavioral multigraph and \mathbf{M}_s with sparsity regularization. We develop an alternative update framework for solving equation 3. The algorithm is summarized in algorithm 1. We now present the key deductions in the following.¹

Algorithm 1 BMSB Discovery.

Require: i. k : Rank of factorization

ii. α : Weight of static structural network

iii. γ : Weight of sparsity regularization on \mathbf{M}_s

Input : One structural network adjacency matrix $\mathbf{X}_s^{n \times n}$
Multiple behavioral graphs \mathbf{X}_b^l , $1 \leq l \leq L$

Output: \mathbf{F} , \mathbf{M}_s , \mathbf{M}_b^l , $1 \leq l \leq L$

- 1: **Initialization:** Randomly initialize nonnegative matrices \mathbf{F} , $n \times k$, \mathbf{M}_s , $k \times k$, and \mathbf{M}_b^l , $1 \leq l \leq L$, $k \times k$;
 - 2: Solve equation 3 with multiplicative update rules;
 - 3: **while** Termination condition not satisfied **do**
 - 4: Update \mathbf{M}_s with equation (7);
 - 5: Update \mathbf{F} with equation (6);
 - 6: Record *temporarily rescaled* $\tilde{\mathbf{M}}_s$ with equation (12) and $\tilde{\mathbf{F}}$ with equation (13)
 - 7: Update \mathbf{M}_b^l with equation 11 using $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{M}}_s$
 - 8: **end while**
 - 9: Rescale \mathbf{M}_s with Equation (12);
 - 10: Rescale \mathbf{F} with Equation (13);
 - 11: **return:** \mathbf{F} , \mathbf{M}_s , and \mathbf{M}_b^l , $1 \leq l \leq L$
-

¹The vectorized matrices are denoted with boldface lower case letters - i.e., $\mathbf{m}_b^l = \text{vec}(\mathbf{M}_b^l)$, $\mathbf{m}_s = \text{vec}(\mathbf{M}_s)$, $\mathbf{x}_b^l = \text{vec}(\mathbf{X}_b^l)$, $\mathbf{x}_s = \text{vec}(\mathbf{X}_s)$.

4.1 Update F

By fixing \mathbf{M}_s and \mathbf{M}_b while updating \mathbf{F} , we have the Lagrangian function of \mathbf{F} in equation 4. λ is the matrix of Lagrangian multipliers.

$$\begin{aligned} \mathcal{L}(\mathbf{F}, \lambda) = & \text{Tr}(\lambda(\mathbf{F}^T \mathbf{F} - \mathbf{I})) + \frac{\alpha}{2} \|\mathbf{X}_s - \mathbf{F} \mathbf{M}_s \mathbf{F}^T\|_F^2 \\ & + \frac{1}{2} \sum_l \|\mathbf{X}_b^l - \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s) \mathbf{F}^T\|_F^2 \end{aligned} \quad (4)$$

$$\begin{aligned} \frac{\partial \mathcal{L}(\mathbf{F}, \lambda)}{\partial \mathbf{F}} = & 2\mathbf{F}\lambda + 2\alpha(-\mathbf{X}_s^T \mathbf{F} \mathbf{M}_s + \mathbf{F} \mathbf{M}_s \mathbf{F}^T \mathbf{F} \mathbf{M}_s) \\ & + 2\sum_l (-\mathbf{X}_b^{lT} \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s) + \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s) \mathbf{F}^T \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s)) \end{aligned} \quad (5)$$

By applying techniques of [4, 6], we have the closed form update rule without Lagrangian multipliers:

Multiplicative Update Rule of F:

$$\mathbf{F} \leftarrow \mathbf{F} \odot \sqrt{\frac{[\alpha \mathbf{X}_s^T \mathbf{F} \mathbf{M}_s + \sum_l \mathbf{X}_b^{lT} \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s)]}{\mathbf{F} \mathbf{F}^T [\alpha \mathbf{X}_s^T \mathbf{F} \mathbf{M}_s + \sum_l \mathbf{X}_b^{lT} \mathbf{F}(\mathbf{M}_b^l \odot \mathbf{M}_s)]}} \quad (6)$$

4.2 Update \mathbf{M}_s

Note the *asymmetric* dependency of $\mathbf{M}_b^l \odot \mathbf{M}_s$ over \mathbf{M}_s can be affected by updating \mathbf{M}_s with the entire objective function: the entries of \mathbf{M}_s can be forced by the minor values of \mathbf{M}_b^l into positive values so that the $\mathbf{M}_s \odot \mathbf{M}_b^l$ overfits the noise, i.e., the purely statistical correlations in \mathbf{X}_b . Therefore, we update \mathbf{M}_s solely based on the factorization of \mathbf{X}_s . On the other hand, the update rules of \mathbf{F} and \mathbf{M}_b^l assure the significant impact of the zero elements of \mathbf{M}_s in forming the block structure.

Multiplicative Update Rule of \mathbf{M}_s :

$$\mathbf{M}_s \leftarrow \mathbf{M}_s \odot \frac{\mathbf{F}^T \mathbf{X}_s \mathbf{F}}{\mathbf{F}^T \mathbf{F} \mathbf{M}_s \mathbf{F}^T \mathbf{F} + \gamma \mathbf{I}} \quad (7)$$

4.3 Update \mathbf{M}_b

We update \mathbf{m}_b^l , which is the vectorization of the l -th frontal slice of \mathbf{M}_b based on the partial differentiation of the objective function over \mathbf{m}_b^l and the general form for updating nonnegative variables in minimization [9]. For reading convenience, let $g^l(\mathbf{u}^l) = \frac{1}{2} \|\mathbf{x}_b^l - (\mathbf{F} \otimes \mathbf{F}) \mathbf{u}^l\|_2^2$ and $\mathbf{u}^l = \mathbf{m}_s \odot \mathbf{m}_b^l$. Then,

$$\begin{aligned} \frac{\partial \mathbf{u}^l}{\partial \mathbf{m}_b^l(\text{index})} = & \mathbf{m}_s(\text{index}) \odot \mathbf{e}_{\text{index}} \\ \text{where } \mathbf{e}_{\text{index}}(\text{index}) = & 1, \mathbf{e}_{\text{index}}(t \neq \text{index}) = 0. \end{aligned} \quad (8)$$

Based on the chain rule of partial differentiation:

$$\frac{\partial g^l(\mathbf{u}^l)}{\partial \mathbf{m}_b^l(t)} = \text{Tr}\left\{\left(\frac{\partial g^l(\mathbf{u}^l)}{\partial \mathbf{u}^l}\right)^T \frac{\partial \mathbf{u}^l}{\partial \mathbf{m}_b^l(t)}\right\} = \left(\frac{\partial g^l(\mathbf{u}^l)}{\partial \mathbf{u}^l}\right)_t \mathbf{m}_s(t) \quad (9)$$

$$\begin{aligned} \frac{\partial g^l(\mathbf{u}^l)}{\partial \mathbf{m}_b^l} = & \{(\mathbf{F} \otimes \mathbf{F})^T [(\mathbf{F} \otimes \mathbf{F})(\mathbf{m}_s \odot \mathbf{m}_b^l) - \mathbf{x}_b^l]\} \odot \mathbf{m}_s \\ = & [((\mathbf{F}^T \mathbf{F}) \otimes (\mathbf{F}^T \mathbf{F}))(\mathbf{m}_s \odot \mathbf{m}_b^l)] \odot \mathbf{m}_s - [(\mathbf{F} \otimes \mathbf{F})^T \mathbf{x}_b^l] \odot \mathbf{m}_s \end{aligned} \quad (10)$$

based on [9],

The Multiplicative Updating Rule of M_b :

$$\mathbf{m}_b^l \leftarrow \mathbf{m}_b^l \odot \left(\frac{[(\mathbf{F} \otimes \mathbf{F})^T \mathbf{x}_b^l] \odot \mathbf{m}_s}{[(\mathbf{F}^T \mathbf{F}) \otimes (\mathbf{F}^T \mathbf{F})](\mathbf{m}_s \odot \mathbf{m}_b^l)} \odot (\mathbf{m}_s + \epsilon) \right), \forall 1 \leq l \leq L \quad (11)$$

Then M_b^l can be attained from \mathbf{m}_b^l through the reverse operation of vectorization.

4.4 Rescaling M_s and F

Within each iteration, we *temporarily* rescale M_s and F before updating M_b^l :

$$M_s(i, j) \leftarrow M_s(i, j) * \max(F(:, i)) * \max(F(:, j)), \forall i, j \quad (12)$$

$$F(:, j) \leftarrow \frac{F(:, j)}{\max(F(:, j))}, \forall j \quad (13)$$

This is essential to our algorithm as we introduce “ $\odot \mathbf{m}_s$ ” in the numerator and “ $\odot (\mathbf{m}_s + \epsilon)$ ” in the denominator of equation 11 to **force** correlation without structural bases to zero. This is because M_s during the iterations absorbs sizes of the blocks, as the *orthogonality* between columns of F are implemented with *orthonormality* $F^T F = I$. The potentially large entries in M_s limit their effects on regularizing M_b^l , even when they are indeed ignorable according to the *relative* volumes within M_s .

5 EXPERIMENTS

In this section we first evaluate our method on synthetic data to show its potential and limitations; we then apply it to different real-world datasets to explore its capability in gaining insights. For synthetic data we specifically compare the performance of our method against baseline approaches in recovering the ground-truth factors, i.e., the underlying block structures and the mixing matrices that are used to generate the synthetic structural network and the simulated multiple behavioral graphs. Evaluating our method on two real-world datasets we explore: (1) brain imaging data² and to show its versatility (2) Twitter data³. For the former we explore two settings one where the behavioral data is for many patients and the other when the behavioral data is for the same person but over time. In each specific setting the static structural network is provided with its $n \times n$ adjacency matrix X_s , and the behavioral networks are presented as a $n \times n \times L$ tensor X_b with its l -th frontal slice being the adjacency matrix $X_b^l = X_b(:, :, l)$. We show our method’s benefit on the fMRI data by visually showing its benefits over baseline method. Importantly, our method was able to discover spatially continuous regions **without** the need for spatial regularization. All the matlab codes for reproducing our results will be made available on `github`⁴.

5.1 Synthetic Data

Synthetic Use Cases and Networks Generation. We generate the synthetic networks X_s and X_b with ground-truth factors F^g , M_s^g , and M_b^g based on equation 14 (please note that M_b^g is a tensor).

$$X_s^g = F^g M_s^g F^{gT}, X_b^g = F^g (M_b^g \odot M_s^g + \mathcal{E}) F^{gT} \quad (14)$$

²<http://adni.loni.usc.edu/study-design/collaborative-studies/dod-adni/>

³<https://snap.stanford.edu/data/higgs-twitter.html>

⁴<https://github.com/ZilongBai/KDD2018>

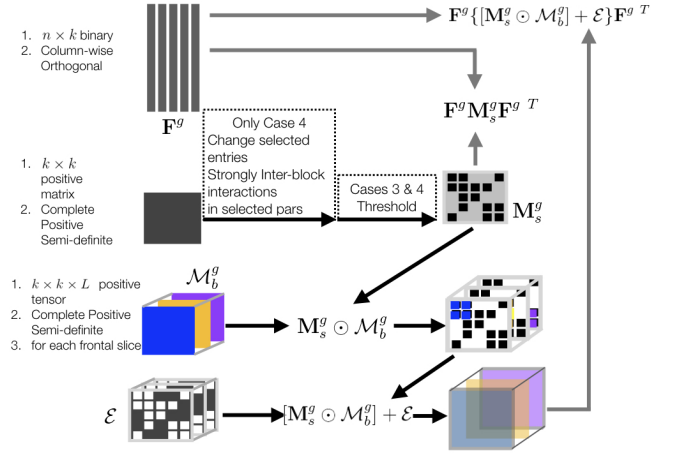


Figure 4: Pipeline for creating synthetic networks.

The pipeline shared by different use cases is shown in Figure 4. We progressively increase the complexity of the underlying structure from case (1) to (4). In all cases the ground truth block structure stays the same, but the comparability between the structural and behavioral ground truth interaction changes. There is complete agreement between the behavioral and structural graphs interaction matrices in the first simple case, but **not** in other three cases.

- (1) *Behavioral and structural interactions at the block level are compatible.* We configure all ground truth mixing matrix recording information density flow weights to be the same as the mixing matrix of potential pathways (i.e., $M_b^g \equiv M_s^g$).
- (2) *Behavioral and structural interactions at the block level are not compatible.* M_s^g is randomly generated. Each frontal slice in M_b^g is randomly and individually generated, independent from M_s^g . All elements in M_s^g and M_b^g are **non-zero**.
- (3) *Behavioral and structural interactions at the block level are not compatible.* Structural mixing matrix M_s^g is randomly generated, conditioned to be sparse. Each frontal slice in M_b^g is randomly and individually generated, independent from M_s^g . Simulated behavioral correlations **without** structural pathways (\mathcal{E}) increments.
- (4) *Behavioral and structural interactions at the block level are not compatible.* Structural mixing matrix M_s^g is randomly generated and conditioned to be sparse. Selected pairs of blocks (off-diagonal elements in M_s^g) have strong inter-block structural connectivity. Each frontal slice in M_b^g is randomly and individually generated, independent from M_s^g . Simulated behavioral correlations **without** structural pathways (\mathcal{E}) increments.

Numerical Measurements of Performance. We propose the following measurements to evaluate the performance of our method in recovering the ground-truth setting that generated the synthetic structural and behavioral networks. Each measures a different aspect of performance. We use notations with superscript g to denote the ground-truth factors. We use the term effective correlation (EC) in these measures. EC’s are generated by $M_s^g \odot M_b^g$ and is

a measure of the **physically plausible correlations** at the block level on behavioral graphs. In this paper, we define the Hadamard product between a $k \times k$ matrix \mathbf{M}_s and a $k \times k \times L$ tensor \mathcal{M}_b as $\mathbf{M}_b^l \odot \mathbf{M}_s, \forall 1 \leq l \leq L$, where $\mathcal{M}_b(:, :, l) \equiv \mathbf{M}_b^l$

- (1) **BDP**: the Discovery Precision of the ground-truth Block Structure (BS) \mathbf{F}^g . Equation 15 measures how accurate the discovered blocks match the the ground truth block based on the cosine similarity between block indicator vectors.

$$BDP(\mathbf{F}^g) = \frac{1}{k} \sum_j IBDP(\mathbf{F}^g(:, j)) \quad (15)$$

$$IBDP(\mathbf{F}^g(:, j)) = \max_i \left(\frac{\mathbf{F}^g(:, j)^T \mathbf{F}(:, i)}{\|\mathbf{F}^g(:, j)\|_2 \|\mathbf{F}(:, i)\|_2} \right)$$

- (2) **SDP**: the Discovery Precision of the ground-truth mixing matrix \mathbf{M}_s^g . Equation 16 measures how close the ground truth and discovered mixing matrices are for the structural graph with cosine similarity between vectorization of the matrices.

$$SDP(\mathbf{M}_s^g) = \frac{\text{vec}(\mathbf{M}_s)^T \text{vec}(\mathbf{M}_s^g)}{\|\text{vec}(\mathbf{M}_s)\|_2 \|\text{vec}(\mathbf{M}_s^g)\|_2} \quad (16)$$

- (3) **MECDP**: Mean Discovery Precision (DP) of the ground-truth Effective Correlations (EC) $\mathbf{M}_s^g \odot \mathcal{M}_b^g$. Equation 17 measures how close the discovered mixing matrices and the ground-truth *effective* interaction matrices are for the behavioral multigraphs with cosine similarity between the vectorized matrices.

$$MECDP(\mathbf{M}_s^g \odot \mathcal{M}_b^g) = \frac{1}{L} \sum_{1 \leq l \leq L} \frac{\text{vec}(\mathbf{M}_s^g \odot \mathcal{M}_b^g(:, :, l))^T \text{vec}(\mathbf{M}_s \odot \mathcal{M}_b(:, :, l))}{\|\text{vec}(\mathbf{M}_s^g \odot \mathcal{M}_b^g(:, :, l))\|_2 \|\text{vec}(\mathbf{M}_s \odot \mathcal{M}_b(:, :, l))\|_2} \quad (17)$$

- (4) **ECME**: Mean Errors (ME) in Effective Correlations. Equation 18 measures the absolute values of falsely detected statistical correlations *without structural connectivity*.

$$I = \text{find}(\mathbf{M}_s^g \equiv 0) \quad (18)$$

$$ECME = \frac{1}{|I| L} \sum_{i \in I, 1 \leq l \leq L} \|[\mathbf{M}_s \odot \mathcal{M}_b]([i], l)\|_1$$

Note the columns and rows of the discovered \mathbf{M}_s and $\mathbf{M}_b^l, \forall l$ must be reordered according to IBDP results to counter graph isomorphism. **Numerical Evaluation Results.** Figure 5 shows the performance of our method in discovering ground-truth factors in cases (1) and (2). In both cases our method recovers \mathbf{M}_s^g and $\mathbf{M}_s^g \odot \mathcal{M}_b^g$ with higher precisions (above 0.85 and 0.95 respectively in Figures 5a and 5b). Note when the mixing matrices of structural network and behavioral graph are compatible, the block discovery precision drops as the number of underlying blocks increments as compatible networks are not necessarily more informative than one single network. An unexpected drop (yet still at about 0.8) in the block structure discovery precision at $k^g = 4$ in Figure 5b can be caused by closeness in the *randomly* generated structural or behavioral inter-block correlations.

Figure 6 compares our method against three baseline approaches: (i) ONMtF on structural network, (ii) Orthogonal Nonnegative (ON-) RESCAL on behavioral multigraphs, and (iii) ON-RESCAL on multigraph with structural network stacked as another frontal slice to the behavioral multigraphs (referred to as extended ON-RESCAL)

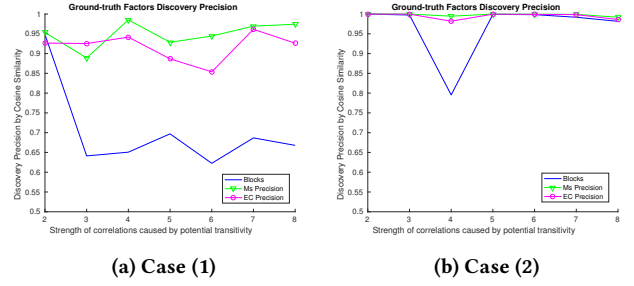


Figure 5: Discovery precisions for the ground-truth underlying factors in the two simple testing use cases (1) and (2). The higher the better. $\alpha = 1, \gamma = 0.1$. The x-axis marks the number of the ground-truth underlying blocks k^g .

in use cases (3) on the first row and (4) on the second row. With behavioral information incorporated, our method and extended ON-RESCAL demonstrate significant advantage over ONMtF in recovering the underlying structural mixing matrix (Figure 6a). By incorporating structural connectivity, our method *always* outperforms both ON-RESCAL and extended ON-RESCAL in discovering *effective* correlations and stand against simulated correlations *without* structural pathways (Figures 6b, 6f, 6c, and 6g). Incorporating behavioral information significantly helps to improve the block structure discovery precision in use case (3) (Figure 6d), while only with behavioral information is not sufficient to stand against the increasingly strong simulated correlations without structural pathways in use case (4) (Figure 6h). Note that our method shows advantage over the extended ON-RESCAL in both cases when the simulated correlations without structural pathways are exceptionally strong (at 0.9 and 1.0 in Figures 6d and 6h).

5.2 Brain Imaging Data

We apply our method to brain networks constructed with both MRI and fMRI scans. Specifically, the vertices of the networks are the 1730 *voxels* within the brain region on the 36-th slice of brain images. This slice is where most differences occur between healthy and unhealthy individuals for Alzheimers. The structural network \mathbf{X}_s can be created in a multitude of ways. Here we exploit the well known AAL atlas [17] which was constructed from MRI structural scans. This atlas divides the brain into 116 structural regions (see http://neuro.imm.dtu.dk/wiki/Automated_Anatomical_Labeling) and we create the structural graphs using the following rules: (1) all the voxels within the same structural/anatomical region are mutually connected, (2) all pairs of voxels that belong to spatially contiguous anatomical regions are connected, and (3) there are no connections otherwise.

Experiments with Cohorts of Individuals. Here we analyze three different cohorts of individuals (young healthy, elderly normal and Alzheimer's affected) using the same structural scan. This allows us to clearly understand the differences in biologically plausible information flow. For each cohort we created as many behavioral graphs as there were individuals in the cohort. The young healthy cohort contains 40 scans, elderly normal 19, and demented

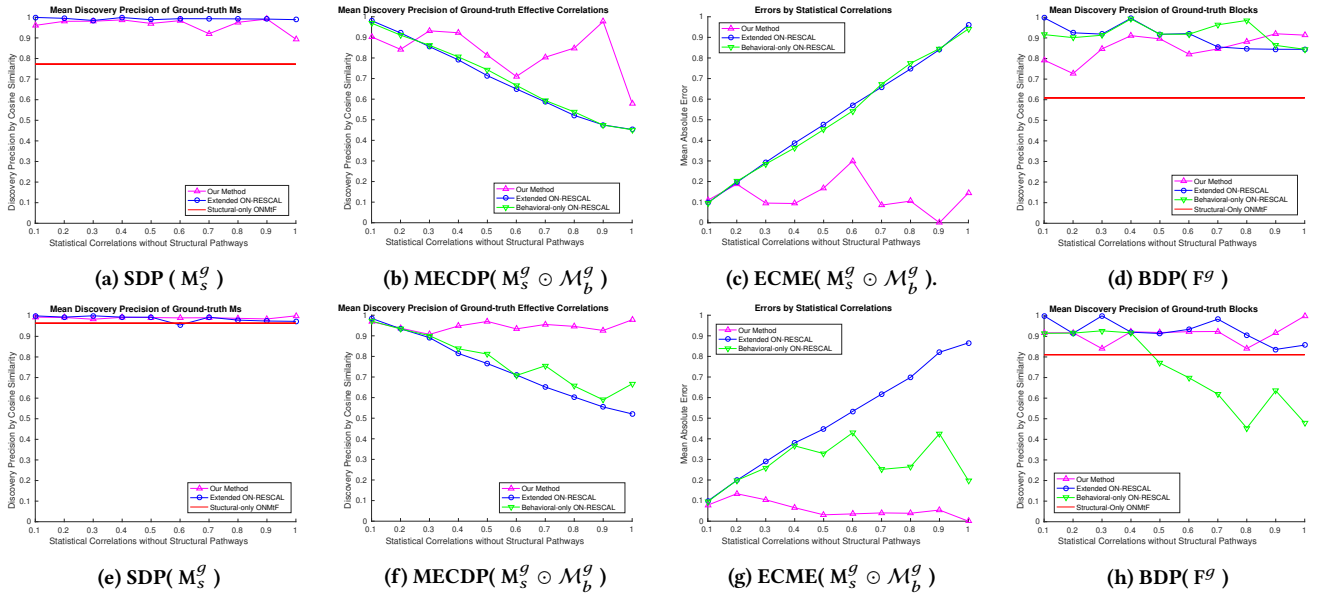


Figure 6: Performance in discovering ground-truth factors of our method with $k = 8$, $\alpha = 20$, $\gamma = 0.1$ compared against baseline approaches: ONMtF, ON-RESCAL, and extended ON-RESCAL of case (3) on the first row and case (4) on the second row. For discovery precisions (SDP, MECDP, and BDP), the higher the better; for discovery errors (ECME), the lower the better. The x-axis marks the strength of simulated behavioral correlations *without* structural pathways incrementing from 0.1 to 1, stepsize 0.1.

(Alzheimer’s affected) 21. In all cases the patients are at the resting state and the simple default mode network (DMN) is active.

Figure 7 shows the block structure found for each cohort and Figure 8 the interaction matrices. Though the former shows that parts of the DMN are active in all three cohorts (the DMN never disappears in demented individuals rather it deteriorates) the results for the mixing matrices are interesting. The bottom row of plots show a much simpler interaction for younger and normal patients **Experiments With Single Patients**. Here we analyze one typical scan from each of the three different cohorts using the same structural scan. This allows us to explore the biologically plausible information flow over time for individuals. For each instance we create 10 graphs from the fMRI scan by dividing the activation patterns into 10 slots and measuring the *absolute Pearson correlation* between the temporal patterns of each pair of voxels within each time slot.

Figure 9 shows the discovered block structures and Figure 10 shows their corresponding interaction matrices. Again, it is significant that our method discovers blocks that each are spatially continuous without enforcing spatial continuity regularization. This is not achievable with previous methods [3, 4]. Figures 9d and 9f discovered similar block structure from the typical young and demented scans, but their interactions between blocks in Figures 10h and 10j are different⁵. The results of baseline approaches are also presented to facilitate visual comparison. Note the extended ON-RESCAL on multigraph with both structural and behavioral data still fails to separate spatially non-contiguous blocks given strong behavioral correlations (Figure 9c).

⁵The observations are made after reordering to counter graph isomorphism.

5.3 Twitter Data

We use the twitter Higgs-Boson data set which is readily available⁶. The raw data consists of a network of 456,622 nodes. Amongst these nodes 328,132 retweet events occur (person k retweets person j ’s retweet of person i ’s tweet and so forth) over a seven day period. However, some of our baseline methods (i.e. multi-view spectral clustering) cannot handle such big graphs and we limit ourselves to a random sample of 2000 nodes. All tweets are on the Higgs Boson discovery but there are many different tweets on the topic from different sources.

Here we aim to answer a simple question. Does the block structure we find using structural and behavioral data differ from community discovery methods found using just the structural graph or just the behavioral graph?

Baselines. We performed simple community detection with the **Louvain** community detection method [5]. This method is very stable and finds 8 communities based on the follower relationship which is our structural graph. For the behavioral graph the method finds 11 communities. The Louvain method attempts to find a good approximation to the min-cut due its focus on modularity. For methods that can use both the behavioral and structural graphs we apply multi-view spectral clustering [22] and Pareto optimization multi-view clustering [19] and required them to find 8 clusters. The RAND index between the block models of these methods and our method is shown in Table 1. We find that the Louvain methods and the multi-view spectral clustering methods find similar community structure to each other but our proposed method finds a very different block structure to them. This is also

⁶Twitter dataset at <https://snap.stanford.edu/data/higgs-twitter.html>.

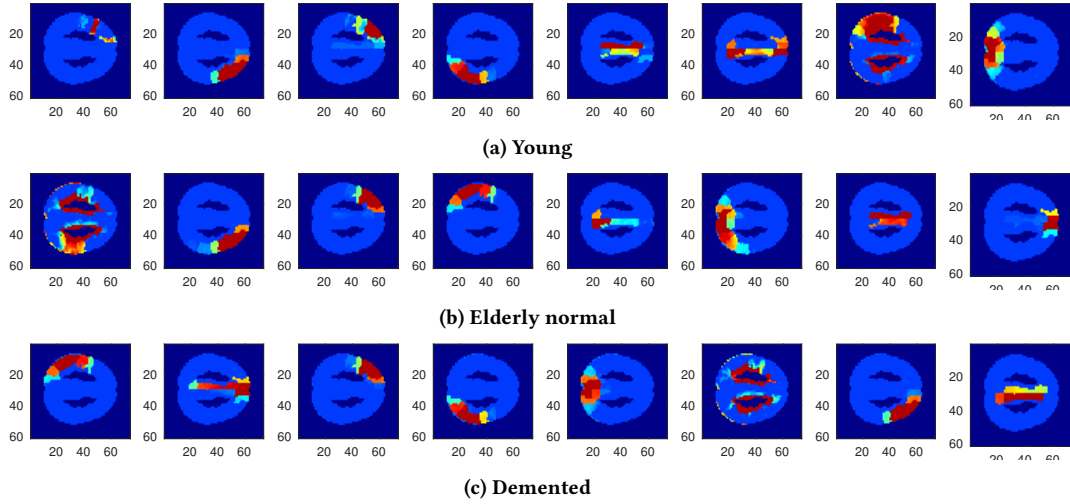


Figure 7: Blocks discovered by our method ($\alpha = 40, \gamma = 1$) from AAL-based structural network and behavioral multigraphs constructed with instances from three cohorts: young, elderly normal, and Alzheimer’s affected.

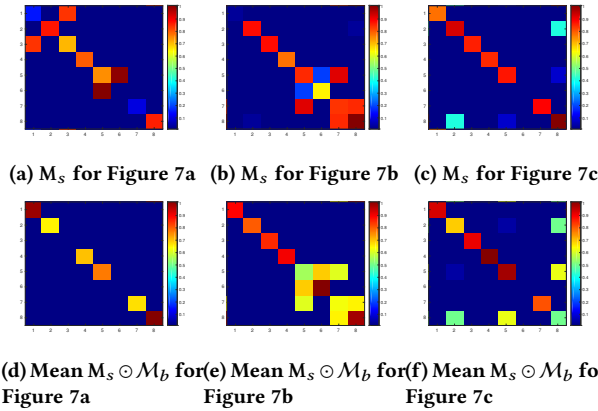


Figure 8: Discovered block level structural and mean behavioral interaction matrices over instances within the same cohort (corresponded to Figure 7).

the case for the Pareto optimization multi-view spectral clustering approach but those two methods (ours and Pareto) find different block structures to each other.

6 CONCLUSION

Existing work focuses on simplifying either a structural graph (i.e. MRI data) or a behavioral/functional graph (i.e. fMRI data) of brain activity. Often these graphs are simplified into block models which cluster the voxels into regions of interests (the blocks) and simultaneously discover the interaction between those blocks. However, not analyzing these two types of graphs together could lead to results that are biologically not plausible. For example, studying just behavioral graphs could yield a block model that shows interactions between voxels that are not structurally connected. Our paper attempts to analyze both types of graphs together for the

	Ours	Lou.-S	Lou.-F	MV:Spec.	Pareto
Ours	1.00	0.32	0.23	0.33	0.21
Louvain-Struct.	0.32	1.00	0.67	0.68	0.34
Louvain-Behav.	0.23	0.67	1.00	0.65	0.35
MV Spectral	0.33	0.68	0.65	1.00	0.34
MV Pareto	0.21	0.34	0.35	0.34	1.00

Table 1: The Rand index between our method and a variety of baselines (the higher the more similar). There are 10 rows as our method ingests 10 behavioral graphs and one structural graph whilst some methods can only ingest one graph at a time.

first time. This is achieved by formulating a coupled block models but the coupling is not simply through one matrix for the blocks, but also using a Hadamard product for the block level interactions. This means the block structure **and** block interaction must consider both behavioral and structural data. This is akin to a very specific form of data fusion and most significantly allows us to find spatially continuous regions despite there being no spatial regularization which was not possible before [4]. We experimentally verify our work in three settings: i) synthetic data and show our method outperforms other baseline multi-graph settings such as RESCAL, ii) real world fMRI and MRI data for Alzheimer’s data and iii) Twitter data to show the broadness of our work.

ACKNOWLEDGMENTS

This research is supported by NSF Grant IIS-1422218: Functional Network Discovery for Brain Connectivity.

REFERENCES

- [1] Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. 2014. Learning latent block structure in weighted networks. *Journal of Complex Networks* 3, 2 (2014), 221–248.

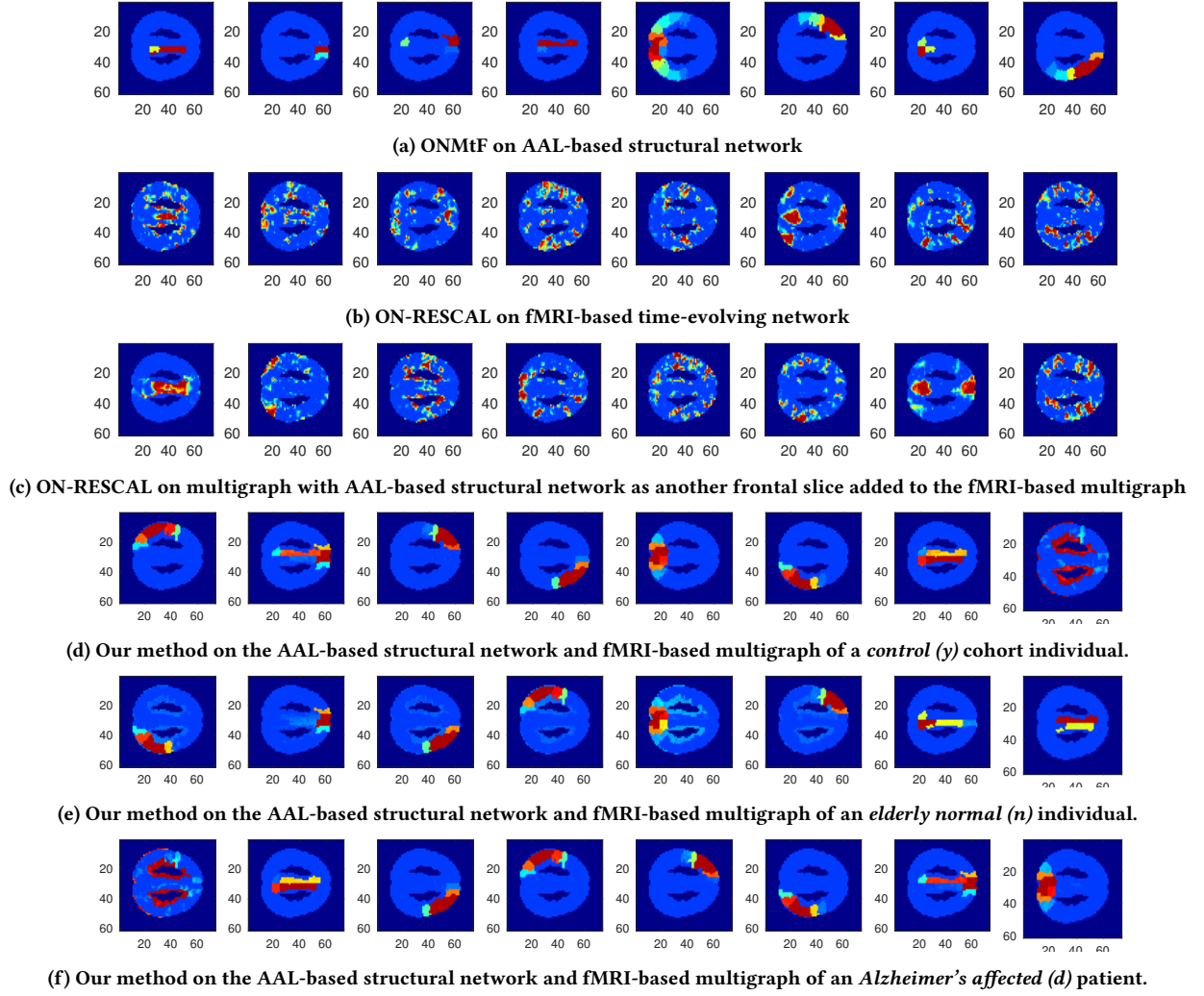


Figure 9: Blocks discovered by our method ($\alpha = 40, \gamma = 1$) and baseline approaches from AAL-based structural network and behavioral multigraphs, each being a time-evolving network constructed with a typical scan from one of the three distinct cohorts (young, elderly normal, or Alzheimer's affected).

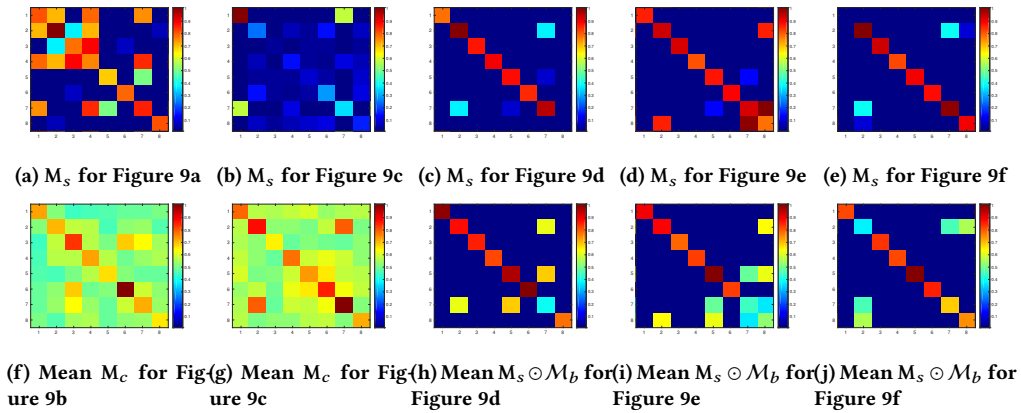


Figure 10: Discovered block level interaction matrices corresponded to the block structures in Figure 9.

- [2] RB Arango, AM Campos, Elías F Combarro, ER Canas, and Irene Díaz. 2016. Mapping cultivable land from satellite imagery with clustering algorithms. *International Journal of Applied Earth Observation and Geoinformation* 49 (2016), 99–106.
- [3] Zilong Bai, Peter Walker, and Ian Davidson. 2018. Mixtures of Block Models for Brain Networks. In *Proceedings of the 2018 SIAM International Conference on Data Mining*. SIAM, 46–54.
- [4] Zilong Bai, Peter Walker, Anna Tschiffely, Fei Wang, and Ian Davidson. 2017. Unsupervised Network Discovery for Brain Imaging Data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 55–64. <https://doi.org/10.1145/3097983.3098023>
- [5] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.
- [6] Chris Ding, Tao Li, Wei Peng, and Haesun Park. 2006. Orthogonal Nonnegative Matrix T-factorizations for Clustering. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 126–135. <https://doi.org/10.1145/1150402.1150420>
- [7] Patric Hagmann, Maciej Kurant, Xavier Gigandet, Patrick Thiran, Van J Wedeen, Reto Meuli, and Jean-Philippe Thiran. 2007. Mapping human whole-brain structural networks with diffusion MRI. *PLoS one* 2, 7 (2007), e597.
- [8] Hisashi Kashima, Koji Tsuda, and Akihiro Inokuchi. 2003. Marginalized kernels between labeled graphs. In *Proceedings of the 20th international conference on machine learning (ICML-03)*. 321–328.
- [9] Daniel D. Lee and H. Sebastian Seung. 2001. Algorithms for Non-negative Matrix Factorization. In *Advances in Neural Information Processing Systems 13*, T. K. Leen, T. G. Dietterich, and V. Tresp (Eds.). MIT Press, 556–562. <http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization.pdf>
- [10] Andrew Y Ng, Michael I Jordan, and Yair Weiss. 2002. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*. 849–856.
- [11] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A Three-way Model for Collective Learning on Multi-relational Data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11)*. Omnipress, USA, 809–816. <http://dl.acm.org/citation.cfm?id=3104482.3104584>
- [12] S. Paul and Y. Chen. 2016. Orthogonal symmetric non-negative matrix factorization under the stochastic block model. *ArXiv e-prints* (May 2016). [arXiv:stat.ML/1605.05349](https://arxiv.org/abs/1605.05349)
- [13] J. Qiu, J. Peng, and Y. Zhai. 2014. Network community detection based on spectral clustering. In *2014 International Conference on Machine Learning and Cybernetics*, Vol. 2. 648–652. <https://doi.org/10.1109/ICMLC.2014.7009685>
- [14] Yonggang Shi and Arthur W Toga. 2017. Connectome imaging for mapping human brain pathways. *Molecular psychiatry* 22, 9 (2017), 1230.
- [15] Olaf Sporns. 2013. Structure and function of complex brain networks. *Dialogues in clinical neuroscience* 15, 3 (2013), 247.
- [16] W. Tang, Z. Lu, and I. S. Dhillon. 2009. Clustering with Multiple Graphs. In *2009 Ninth IEEE International Conference on Data Mining*. 1016–1021. <https://doi.org/10.1109/ICDM.2009.125>
- [17] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 1 (2002), 273–289.
- [18] Kun Wang, Meng Liang, Liang Wang, Lixia Tian, Xinqing Zhang, Kuncheng Li, and Tianzi Jiang. 2007. Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. *Human brain mapping* 28, 10 (2007), 967–978.
- [19] Xiang Wang, Buyue Qian, Jieping Ye, and Ian Davidson. 2013. Multi-objective multi-view spectral clustering via pareto optimization. In *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 234–242.
- [20] J. Yang, J. McAuley, and J. Leskovec. 2013. Community Detection in Networks with Node Attributes. In *2013 IEEE 13th International Conference on Data Mining*. 1151–1156. <https://doi.org/10.1109/ICDM.2013.167>
- [21] Zhong-Yuan Zhang, Yujie Gai, Yu-Fei Wang, Hui-Min Cheng, and Xin Liu. 2018. On equivalence of likelihood maximization of stochastic block model and constrained nonnegative matrix factorization. *Physica A: Statistical Mechanics and its Applications* 503 (2018), 687–697.
- [22] Dengyong Zhou and Christopher J. C. Burges. 2007. Spectral Clustering and Transductive Learning with Multiple Views. In *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*. ACM, New York, NY, USA, 1159–1166. <https://doi.org/10.1145/1273496.1273642>