

Identifying Sub-Populations via Unsupervised Cluster Analysis on Multi-Edge Similarity Graphs

Madhura Ingalhalikar¹, Alex R. Smith¹, Luke Bloy¹, Ruben Gur²,
Timothy P.L. Roberts^{3,*}, and Ragini Verma¹

¹ Section of Biomedical Image Analysis,
University of Pennsylvania, Philadelphia, PA, USA
{Madhura.Ingalhalikar,Ragini.Verma}@uphs.upenn.edu

² Brain Behavior Laboratory, University of Pennsylvania, Philadelphia, PA, USA

³ Lurie Family Foundation's MEG Imaging Center, Department of Radiology,
Children's Hospital of Philadelphia, Philadelphia, PA, USA

Abstract. Pathologies like autism and schizophrenia are a broad set of disorders with multiple etiologies in the same diagnostic category. This paper presents a method for unsupervised cluster analysis using multi-edge similarity graphs that combine information from different modalities. The method alleviates the issues with traditional supervised classification methods that use diagnostic labels and are therefore unable to exploit or elucidate the underlying heterogeneity of the dataset under analysis. The framework introduced in this paper has the ability to employ diverse features that define different aspects of pathology obtained from different modalities to create a multi-edged graph on which clustering is performed. The weights on the multiple edges are optimized using a novel concept of 'holding power' that describes the certainty with which a subject belongs to a cluster. We apply the technique to two separate clinical populations of autism spectrum disorder (ASD) and schizophrenia (SCZ), where the multi-edged graph for each population is created by combining information from structural networks and cognitive scores. For the ASD-control population the method clusters the data into two classes and the SCZ-control population is clustered into four. The two classes in ASD agree with underlying diagnostic labels with 92% accuracy and the SCZ clustering agrees with 78% accuracy, indicating a greater heterogeneity in the SCZ population.

1 Introduction

Classifying subjects based on their underlying pathology, brain structure, behavior and cognition is an important step towards creating biomarkers. However, pathologies like ASD and other neuropsychiatric disorders are defined over a spectrum and the severity of the disease may vary within a population thus

* The authors would like to acknowledge support from the NIH grants: MH092862, MH079938 and DC008871.

making the data highly heterogeneous. Different modalities, like imaging, neurocognitive scores etc., may characterize different aspects of this heterogeneity to different degrees. This paper presents a method for unsupervised cluster analysis of populations using multi-edge similarity graphs that combine information of population heterogeneity from different modalities, producing classes that are more representative of population variability.

Traditional supervised classification methods, utilize predefined diagnostic labels for the subjects for training [1], [2], and hence new subjects can only be classified into one of these diagnostic categories, thereby overlooking the underlying heterogeneity of the pathology. These also require a large sample size to capture all the variability.

Unsupervised classification or clustering are powerful techniques for self-organized categorization of the underlying data [3] without the use of diagnostic labels. Earlier studies have used various clustering algorithms in classifying tissue types or segmenting lesions in brain images [4]. However, with diseases now being grouped into a variety of classes, population analysis using such unsupervised methods is gaining interest in the neuroimaging community [5]. Recently, one study by Filipovych et al. performed semi-supervised clustering on datasets [6]. The method was limited to use information from single imaging modality and thus overlooked other components of the pathology.

Ideally, for precisely grouping a subject into a certain category, information from diverse imaging modalities, psychological scores, demographics as well as genetic information can be combined together and clustered without utilizing clinical diagnostic information. Such a technique will thus provide a comprehensive grouping and aid in understanding the underlying patterns of pathology.

With such an aim, in this paper we present a novel method that employs unsupervised clustering on multiple features to better understand the underlying data structure and to identify coherent subpopulations if any. We define each subject by its structural networks computed from DTI data and a battery of cognitive scores. The nodes of the structural network are clustered and similarity between subjects is computed using the variance of information metric between these clusters. For the cognitive battery, the similarity is computed using Euclidean distance. Thus we get a dual-edged similarity graph, in which subjects represent the nodes and the similarities represent the edges. We then perform unsupervised spectral clustering on the linear combination of these similarity graphs. The optimal linear combination is defined via the concept of ‘holding power’ that provides a basis of certainty with which a subject belongs to a particular cluster. The weights on the individual similarity graph quantify the participation of that feature in the clustering process. We apply this method to two datasets, with autism and schizophrenia pathologies, respectively, to determine the ability of our method in identifying homogeneous subpopulations.

2 Methods

Here we describe the method of clustering on multi-edge graphs created on the population with the edges defining inter-subject similarity based on different

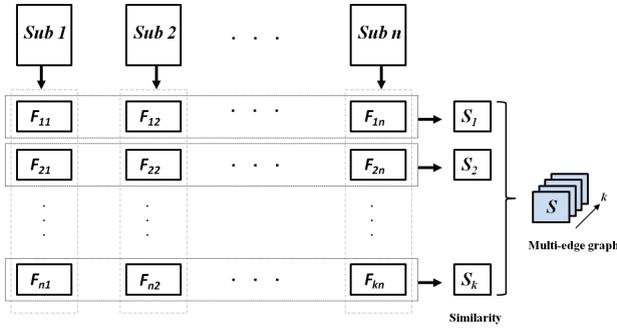


Fig. 1. This diagram depicts the technique used to create a multi-edge graph. Diverse features (F_1, \dots, F_k) are extracted for each subject that may include information from connectivity matrices, image intensity values, genomic data or cognitive scores. The similarity matrices (S_1, \dots, S_k) are then computed over all the subjects for each feature. Together these matrices form the multi-edge graph which defines various facets of similarity (edges) between the nodes (subjects).

modalities such as imaging and cognitive scores and define the concept of holding power of each node in the multi-edge graph, which represents the power with which the subject (represented by a node) belongs to a cluster.

2.1 Unsupervised Clustering on Multi-edge Graphs

Consider a similarity graph $S = (V, E)$, created over a population where the subjects are represented by the nodes V and the similarities based on the modalities are defined by the edges E . Each edge e in the graph represents the connection (similarity in our case) between two nodes with a weight w ($w \in R$). For a multi-edge graph, with k number of linkages (each representing a modality) the edges between any two nodes can be described by $w \in R^k$.

The goal of unsupervised clustering on such a graph is to partition the graph, utilizing information from all possible linkages/modalities, such that nodes with tighter connections (with high similarity) cluster together while nodes with loose connections (low similarity) are placed in different clusters.

Clustering a multi-edge graph is a challenging problem as each edge type conveys different information and thus can cause the subjects to cluster differently. We address this challenge by flattening the graph into a single edged graph by employing a linear function where $f = \sum \alpha_i S_i$ where $i = (1, 2, \dots, k)$ and α_i is the weight on all the edges of S_i . Our focus here is to obtain good clustering by using information from all the available features. Therefore the problem translates to finding values of α that optimize the clustering quality. To quantify the quality of clustering, we use the concept of pull and holding power as proposed by Rocklin et al. [7].

We begin with flattening the multi-edge graph where the α_i are chosen randomly under the condition $\sum \alpha_i = 1$. Unsupervised clustering is then applied

to this linear combination $(\alpha_1 S_1 + \alpha_2 S_2 + \dots + \alpha_k S_k)$. Any type of unsupervised clustering method can be applied (e.g. affinity clustering, graph clustering etc). Since our goal here is to demonstrate the importance of multi-edge graph, we use a standard spectral clustering algorithm [8] that clusters the nodes in M clusters. For each node $v \in V$, the ‘pull’ to each cluster C_m in $C = (C_1, C_2, \dots, C_M)$ is then defined as the average weights of edges between node v and the nodes categorized in cluster C_m . Therefore, for a given set of coefficients α , the pull on node v is defined by equation 1 where x is the number of nodes categorized in cluster C_m .

$$P_\alpha(v, C_m) = \frac{1}{x} \sum_{w=(u,v) \in E, u \in C_m} w(\alpha) \tag{1}$$

The holding power $H_\alpha(v)$ for each node, is then defined as the pull of the cluster to which the node belongs minus the next largest pull among other clusters. Thus, for a node v in cluster C_m , the holding power is defined by equation 2.

$$H_\alpha(v) = P(v, C_m) - \max_{C_i \in C, C_i \neq C_m} P(v, C_i) \tag{2}$$

If the holding power is positive, then we can say that the node is held in the right cluster. Thus, to achieve superior clustering that justifies the position of each node in that cluster, we can maximize the holding power as well as the number of nodes with positive holding power. For easier implementation of holding power in optimization routines, the function can be smoothed by using $\text{atan}(H_\alpha(v))$.

2.2 Multi-edge Graphs from Structural Networks and Cognitive Scores

In our study the two edge weights that we use are computed from the full brain structural connectivity networks the cognitive scores of each subject.

The structural network is a $n \times n$ connectivity matrix where n regions of interest (ROI’s) in gray matter are defined and the edge computation between two nodes is based on the density of white matter fibers between these nodes. Here, we explain the similarity between the structural networks of two subjects via the similarity between the community structures of the structural networks for the two subjects. Obtaining communities in the structural network is essentially equivalent to performing clustering on the nodes of the structural network. Here, we use a standard unsupervised spectral clustering algorithm for obtaining the community structure for each subject [8]. We then compute the distance between two subjects or two clusterings by using a variation of information (VI) metric [9] which is based on the mutual information between two clustering’s [9].

Consider $C_i = (C_i^1, C_i^2, \dots, C_i^K)$ to be a clustering on subject i and $C_j = (C_j^1, C_j^2, \dots, C_j^K)$ to be the clustering on subject j and let n be the total number of nodes. Then $P(C, k) = \frac{|C^k|}{n}$ is the probability that the node is in cluster C^k in a clustering C_i and in clustering C_j is $P(C_i, C_j, k, l) = \frac{|C_i^k \cap C_j^l|}{n}$. The entropy of

information in C_i is defined by equation 3, while the mutual information shared by C_i and C_j is given by equation 4.

$$H(C_i) = - \sum_{k=1}^K P(C_i, k) \log P(C_i, k) \quad (3)$$

$$I(C_i, C_j) = \sum_{k=1}^K \sum_{l=1}^{K'} P(C_i, C_j, k, l) \log P(C_i, C_j, k, l) \quad (4)$$

The variation of information metric that describes the distance between two clusterings is then defined by equation 5.

$$d_{VI}(C_i, C_j) = H(C_i) + H(C_j) - 2I(C_i, C_j) \quad (5)$$

The similarity between the battery of cognitive scores creates the second edge of the multi-edge graph. The edge ij of the similarity matrix is the Euclidean distance between the cognitive scores of subject i and subject j . For a vector of length r the Euclidean distance is computed by equation 6.

$$d(\text{cog}_i, \text{cog}_j) = \sqrt{\sum_{l=1}^r (\text{cog}_{il} - \text{cog}_{jl})^2} \quad (6)$$

To convert the variation of information distance matrix as well as the distance matrix of cognitive test scores, to similarity matrices, we use the negative exponential of the distance as proposed by Shepard [10]. Thus, we now have a multi-edge similarity graph on which unsupervised spectral clustering can be performed. For optimizing the holding power, optimizers such as gradient descent can be applied. However, in our case, since it's only a dual-edged graph, a simple grid search performs reasonably. We then apply the weights computed at the maximum holding power to obtain the final clustering on the population.

The clustering is then validated against the ground truth diagnosis. The validation enables us to understand the performance of the technique as well as provides insight into the heterogeneity of the dataset.

3 Results

3.1 Simulated Data

We consider a population of 8 subjects, described by 3 modalities. We simulated three 8×8 similarity matrices (S_1, S_2, S_3) each representing the similarities described by a specific modality. As can be seen from Fig 2, S_1 and S_3 were designed to impose a clustering (with modality S_3 characterizing the first 5 subjects better and modality S_1 characterizing the last 3 subjects better), while S_2 was diffuse. We then combined these matrices to create a 3 edged graph (as described in section 2.1). The weights (α) on these similarity matrices were optimized using a grid search method, to maximize the total holding power and

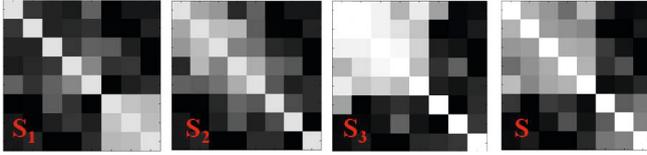


Fig. 2. Figure shows the simulated similarity matrices (S_1 , S_2 and S_3) which form the multi-edge graph. S is the linear combination of S_1 , S_2 and S_3 with maximum holding power and the 2 clusters in S are evident.

minimize the number of nodes (subjects) with negative holding power, that is, nodes that are mis-clustered. The maximum holding power was achieved at 0.4 for S_1 and 0.6 for S_3 , while the weight on S_2 was zero since it did not add anything to maximize the holding power, thereby establishing the feasibility of the holding power concept. The optimization implicitly puts more weight on the matrices with stronger connections, thus choosing S_1 and S_3 and removing S_2 . Spectral clustering on the final similarity matrix S caused it to cluster into 2 classes with the first five nodes in one cluster and last three in other cluster which matches the underlying clustering of the data.

3.2 Real Data

Two separate datasets were used in the unsupervised clustering:

- The SCZ dataset consisted of 29 female controls (CNT) and 23 female age matched patients with schizophrenia. The DWI images were acquired on Siemens 3T scanner with $b=1000$ s/mm^2 and 64 gradient directions. Neurocognitive testing was carried out on all the subjects and the speed and accuracy of memory, emotion, reasoning, and executive functioning were recorded.
- The ASD dataset consisted of 33 participants with ASD and 21 age matched typically developing controls (TD's). The DWI images were acquired on Siemens 3T scanner with $b=1000$ s/mm^2 and 30 gradient directions. The cognitive and psychological tests included verbal IQ, Social Responsiveness Scale (SRS), Social communication questionnaire (SCQ), Clinical evaluation of language fundamentals (CELF), Full scale IQ and Autism diagnostic observation schedule (ADOS) and perceptual reasoning index (PRI).

Computing the Structural Network. Cortical parcellation and sub-cortical segmentation of all the subjects was obtained using Freesurfer [11] on structural T1 images, and a total of 78 ROI's were extracted to represent the nodes of the structural network. These labels were then transferred to the diffusion space via intrasubject affine transformation. Probabilistic fiber tracking [12] was employed to determine the percentage of streamlines that exit ROI i and enter ROI j . The conditional probability is given by $p_{ij} = \frac{S_{i \rightarrow j}}{S_i}$, where $S_{i \rightarrow j}$ denotes the number of fibers reaching j , and S_i is the number of streamlines seeded in i .

We normalize p_{ij} by the active surface area R_i of the ROI i to get connectivity measure P_{ij} , which accounts for different sizes of the active seed region. This measure quantifies connectivity such that $P_{ij} \approx P_{ji}$ which upon averaging, gives an undirected weighted connectivity matrix.

We then perform clustering on the computed structural networks. We apply spectral clustering to each subject and then compute the VI matrix (Section 2.2, equation 5) for the two datasets. The similarity matrix from the cognitive scores is computed using the Euclidean distance as described in Section 2.2. Figure 3 shows the similarity matrices for the two datasets. The red lines divide the matrix to reveal patients and controls. The cognitive score of the ASD produces a ‘visual clustering’ in the matrix (d), however the psychological testing scores for SCZ (b), do not produce such a distinctive difference as is evident in the color coding of the matrix. The structural connectivity similarity matrices, do not impose such acute clustering.

Unsupervised Clustering on Multi-edge Graph. We performed spectral clustering on the multi-edge graphs for both the datasets, such that the holding power is maximized with minimum number of subjects having negative holding power. The weights at maximum holding power for SCZ data were 0.55 for structural network and 0.45 for the psychological scores, suggesting that the combination of information aided in the clustering process. The entire dataset was clustered into 4 groups with two clusters representing the SCZ patients and 2 clusters representing the controls with 78% accuracy.

For ASD data, the maximum holding power was achieved with a weight of 0.2 for structural and 0.8 for psychological scores. The spectral clustering on this linear combination split the data into two clusters: one with ASD and other with TD with 92% accuracy. This suggested that although DTI did not add much, the combination of information was important to maximize the hold of each subject in that cluster. When only psychological scores were used, the data was split into 3 clusters, where the third cluster consisted 4 subjects (2 ASD and 2 TD).

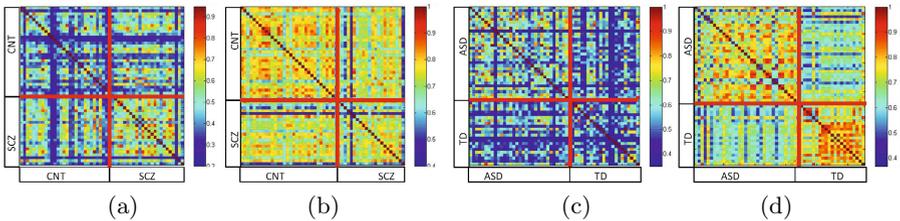


Fig. 3. (a) Similarity matrix computed from structural connectivity network for SCZ (b) similarity matrix computed from cognitive scores in SCZ (c) Similarity matrix computed from structural networks in ASD (d) Similarity matrix computed from cognitive scores in ASD. The horizontal and vertical red lines show the control-patient division in the group.

4 Conclusion

In this paper, we have created a novel technique for unsupervised clustering that creates population based multi-edged graphs using different modalities and features. Spectral clustering on these graphs in conjunction with maximizing the holding power of each subject in a cluster was used to identify population subgroups. The method was validated on simulated data and then applied to datasets with ASD and schizophrenia. We found two inherent clusters in the ASD data while schizophrenia data was more heterogeneous with four inherent clusters. A direct interpretation of such clusters is non-trivial, but it is our working hypothesis that relevant features found using this methodology will map onto the clinical space of cognitive scores. In future, we plan on expanding this idea for defining the heterogeneity index over a patient population in a spectrum disorder like ASD.

References

1. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: Compare: classification of morphological patterns using adaptive regional elements. *IEEE Trans. Med. Imaging* 26(1), 93–105 (2007)
2. Ingalhalikar, M., Parker, D., Bloy, L., Roberts, T.P.L., Verma, R.: Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of asd. *Neuroimage* 57(3), 918–927 (2011)
3. Duda, R., Hart, P., Stork, D.: *Pattern Classification*. Wiley, New York (2001)
4. Xu, R., Wunsch, D.C.: Clustering algorithms in biomedical research: a review. *IEEE Rev. Biomed. Eng.* 3, 120–154 (2010)
5. Sabuncu, M., Balci, S., Golland, P.: Discovering Modes of an Image Population through Mixture Modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II*. LNCS, vol. 5242, pp. 381–389. Springer, Heidelberg (2008)
6. Filipovych, R., Resnick, S.M., Davatzikos, C.: Semi-supervised cluster analysis of imaging data. *Neuroimage* 54(3), 2185–2197 (2011)
7. Rocklin, M., Pinar, A.: Latent Clustering on Graphs with Multiple Edge Types. In: Frieze, A., Horn, P., Prałat, P. (eds.) *WAW 2011*. LNCS, vol. 6732, pp. 38–49. Springer, Heidelberg (2011)
8. Newman, M.E.J.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U S A* 103(23), 8577–8582 (2006)
9. Meilă, M.: Comparing Clusterings by the Variation of Information. In: Schölkopf, B., Warmuth, M.K. (eds.) *COLT/Kernel 2003*. LNCS (LNAI), vol. 2777, pp. 173–187. Springer, Heidelberg (2003)
10. Shepard, R.N.: Toward a universal law of generalization for psychological science. *Science* 237(4820), 1317–1323 (1987)
11. Fischl, B., et al.: Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14(1), 11–22 (2004)
12. Behrens, T.E.J., et al.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* 6(7), 750–757 (2003)