



Explainable Classification of Brain Networks via Contrast Subgraphs

Tommaso Lanciano
Sapienza University, Rome, Italy
tommaso.lanciano@uniroma1.it

Francesco Bonchi
ISI Foundation, Turin, Italy
francesco.bonchi@isi.it

Aristides Gionis
KTH, Sweden
argioni@kth.se

ABSTRACT

Mining human-brain networks to discover patterns that can be used to discriminate between healthy individuals and patients affected by some neurological disorder, is a fundamental task in neuroscience. Learning simple and interpretable models is as important as mere classification accuracy. In this paper we introduce a novel approach for classifying brain networks based on extracting *contrast subgraphs*, i.e., a set of vertices whose induced subgraphs are dense in one class of graphs and sparse in the other. We formally define the problem and present an algorithmic solution for extracting contrast subgraphs. We then apply our method to a brain-network dataset consisting of children affected by Autism Spectrum Disorder and children Typically Developed. Our analysis confirms the interestingness of the discovered patterns, which match background knowledge in the neuroscience literature. Further analysis on other classification tasks confirm the simplicity, soundness, and high explainability of our proposal, which also exhibits superior classification accuracy, to more complex state-of-the-art methods.

ACM Reference Format:

Tommaso Lanciano, Francesco Bonchi, and Aristides Gionis. 2020. Explainable Classification of Brain Networks via Contrast Subgraphs. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '20)*, August 23–27, 2020, Virtual Event, CA, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3394486.3403383>

1 INTRODUCTION

The development of magnetic resonance imaging (MRI) techniques has paved the way to *connectomics* [32], i.e., modeling the brain as a network, allowing to tackle interesting neuroscience research questions as graph-analysis problems [6].

A connectome is a map describing neural connections between brain regions of interest (ROIs), either by observing anatomic fiber density (structural connectome), or by computing pairwise correlations between time series of activity associated to ROIs (functional connectome). The latter approach, known as functional magnetic resonance imaging (fMRI), exploits the link between neural activity and blood flow and oxygenation, to associate a time series to each ROI. A brain network can then be defined by creating a link between two ROIs that exhibit co-activation, i.e., strong correlation in

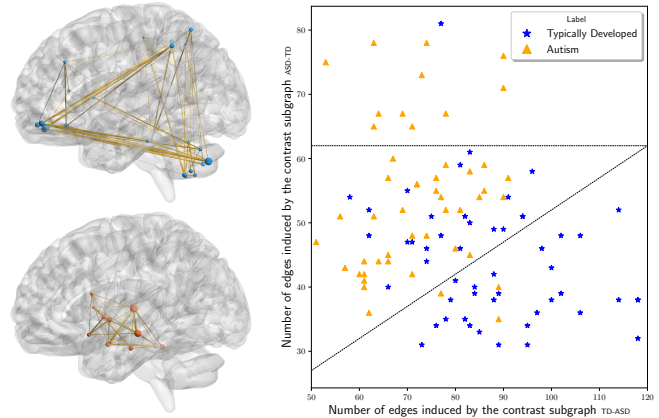


Figure 1: Example of real-world *contrast subgraphs* extracted from a brain-network dataset consisting of 49 children affected by Autism Spectrum Disorder (class ASD) and 52 Typically Developed (class TD) children. Top-left: the contrast subgraph TD-ASD. Bottom-left: the contrast subgraph ASD-TD. On the right, scatter plot showing for each individual the number of edges present in the subgraph induced by the contrast subgraph TD-ASD (x -axis) and by the contrast subgraph ASD-TD (y -axis). Details are given in Section 1.2 and Section 5.

their time series. An important problem in this domain is to identify connection patterns that might be associated to specific cognitive phenotypes or mental dysfunctions [10]. Given fMRI scans of patients affected by a mental disorder and scans of healthy individuals, the goal is to discover patterns in the corresponding connectomes that explain differences in the brain mechanism of the two groups. Identifying such patterns might provide important insight of the disorder and hint strategies to improve the condition of patients.

This task can be seen as a *graph classification* problem [25, 38, 42]. We are given two groups of individuals, a *condition group* \mathcal{A} and a *control group* \mathcal{B} . Each individual is represented by a brain network, i.e., a graph $G_i = (V, E_i)$, where each graph is defined over the same set V of vertices (corresponding to the brain ROIs). The set of edges E_i represents the connections, either structural or functional, between the ROIs observed G_i . The goal is to infer a model that, given an unseen graph $G_n = (V, E_n)$, predicts whether it belongs to class \mathcal{A} or \mathcal{B} .

Literature on graph classification is mostly based on kernel methods [31], graph embeddings [1, 16, 27], and deep learning [22, 39, 43]. The bulk of these methods for graph classification (with few exceptions discussed in Section 2), however, is not well-suited for the task of classification in brain networks, for the following reasons.

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 the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to 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 '20, August 23–27, 2020, Virtual Event, CA, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7998-4/20/08...\$15.00

<https://doi.org/10.1145/3394486.3403383>

Node-identity awareness: existing approaches are designed to find structural similarities among the input graphs, without taking into consideration the *correspondence* of nodes among the different networks, i.e., the fact that a specific vertex id corresponds to the same brain ROI in all the input networks. Not in every application domain it is possible to identify corresponding nodes among different graphs. If such a property holds, however, it is crucial to take it into account – just ignoring it represents a fatal loss of information.

Black-box effect: the inferred models are complex and difficult to understand and to explain why a certain prediction is made. *In this specific application domain, the simplicity and explainability of the models are of uttermost importance.* In fact, the neuroscientist needs to understand which are the ROIs and their interactions that best discriminate between patients and healthy individuals.

High number of parameters: a third important limitation is the very high number of parameters that need to be fine tuned, making the existing approaches too complex to be adopted by non-experts, especially in application domains where the number of examples is inherently small¹ and over-fitting can be hard to avoid.

1.1 Our proposal and contributions

We propose a simple and elegant, yet effective model, based on extracting *contrast subgraph*, i.e., a set of vertices whose induced subgraph is dense in the graphs of class \mathcal{A} and sparse in the graphs of class \mathcal{B} . Our model is extremely simple, it has only one parameter (governing the complexity of the model), and it has excellent interpretability. Although it is not the main assessment criteria, our model also exhibits very good classification accuracy in several different brain-classification tasks, outperforming more complex models which have several parameters and long training time.

Our main contributions can be summarized as follows:

- We introduce a novel problem aimed at extracting contrast subgraphs from two classes of graphs, defined on the same set of vertices (Section 3). We study the complexity of the problem and propose algorithmic solutions (Section 4).
- We apply our method to a brain-network dataset consisting of children affected by Autism Spectrum Disorder and children Typically Developed (Section 5). The analysis of this dataset confirms the interestingness, simplicity and high explainability of the patterns extracted by our methods, which match domain knowledge in the neuroscience literature.
- We further assess the classification performance of our method by comparing it against state-of-the-art methods, on several brain-classification tasks. The results show that our method, despite its simplicity, outperforms those more complex methods.

We next provide a preview of the type of patterns that our approach can extract from a real-world dataset.

¹Each data point requires an individual undergoing through a scan, which is an expensive operation.

1.2 A preview of the results

A preview of the type of structures and classification capabilities of our method is presented in Figure 1. The data are obtained from the Autism Brain Imagine Data Exchange (ABIDE) project [9]; more details about data are provided in Section 5. The specific dataset contains 101 brain networks: 49 patients affected by *Autism Spectrum Disorder* (class ASD) and 52 *Typically Developed* (class TD) individuals. Each individual is represented by an undirected unweighted graph defined over 116 vertices, each corresponding to a ROI.

On the top-left of Figure 1 we illustrate the contrast subgraph TD-ASD (the subgraph that maximizes the difference between the number of edges in the class TD with respect to the same measure for class ASD); on the bottom-left we illustrate the contrast subgraph ASD-TD. Vertex size represents the importance of the vertex in discriminating the two classes. Additionally, the figure depicts the edges whose discrimination capability is above a given threshold.²

By inspecting the two subgraphs, we can observe an evident complementarity: TD-ASD involves mainly connections between cerebellum, prefrontal cortex, posterior parietal cortex, inferior and middle temporal gyri, while ASD-TD only exhibits connections between striatum and limbic cortex. These first-sight findings are consistent with recent literature [19, 24]. Furthermore, the scatter plot on the right-hand side of Figure 1 reports for each individual (or graph) in the dataset, the number of edges present in the contrast subgraph TD-ASD (x -axis) and in the contrast subgraph ASD-TD (y -axis). We can see that these very simple and explainable measures, already provide a good separability of the two classes. For instance, the horizontal dotted line represents the following rule:

If an individual exhibits more than 62 edges among the 15 vertices of the contrast subgraph ASD-TD, then there are high chances that the individual is affected by ASD.

The diagonal dotted line corresponds to another simple rule:

If the number of edges induced by the contrast subgraph ASD-TD is smaller than half of the number of edges induced by the contrast subgraph TD-ASD, then there are high chances that the individual is not affected by ASD.

Simple rules like these are self-explainable and easily communicable to the neuroscientists for further investigation.

2 RELATED WORK

Graph classification. Graph classification, i.e., the task of building a model able to predict the target class for unseen graphs accurately, is receiving increasing attention as witnessed by the many approaches proposed in the last few years in the literature.

Shervashidze et al. [31], propose a graph kernel based on the Weisfeiler-Lehman test of graph isomorphism. Narayanan et al. [27] propose *graph2vec*, a method that considers rooted subgraphs as the components that define a specific graph and performs an embedding under this assumption. Adhikari et al. [1] propose *sub2vec*, whose aim is to learn subgraph embeddings, preserving structural and neighborhood properties. In our experimental comparison in

²As we explain in the next section, a contrast subgraph is defined exclusively by a set of vertices. However, in these illustrations we also highlight the most relevant edges.

Section 5 we compare the classification accuracy obtained by our method against state-of-the-art baselines [1, 27, 31]. As previously discussed, these methods are not node-identity aware: they do not take into consideration the correspondence of nodes in the different networks. Not in every application domain it is possible to identify corresponding nodes among different graphs, but in application domains (such as brain network classification) in which this is possible, not taking it in consideration is an important limitation. To the best of our knowledge, the only unsupervised embedding for networks with node identity awareness was proposed by Gutierrez et al. [16]: we also compare with this method in our experiments.

If the task of graph classification is something that arose recently, application of it to the domain of brain networks is a growing line of research (see [10] for a comprehensive survey). Recently several researchers [25, 26, 38] have proposed the use of deep learning architectures for the same problem we tackle in this paper, i.e., brain networks classification. However, these methods suffer from the black-box curse as they have very little interpretability. Moreover, they have very high number of parameters that need to be fine tuned and long training time. A recent work by Yan et al. [42], still based on neural networks, also focuses on explainability. Their proposed architecture has a node-grouping layer before the convolutional layer: it is claimed that such node-grouping layer might provide insight in the most predictive brain subnetworks w.r.t. the classification task at hand. Direct comparison with [42] is not doable as their pipeline starts directly from the timeseries associated to the ROIs, and early on splits the computation in two tracks: the positive and the negative correlation track. Instead, the input to the problem studied in this paper, is simply two groups of graphs.

Contrast and dense subgraphs. Contrast, sometimes called discriminative, subgraphs have been approached in the graph mining literature mostly as a *frequent subgraph mining* task [18, 33, 34, 41]: in this context the discriminative structure is a typically small motif which is defined over node-labelled graphs (e.g., a triangle h-h-o), that appears frequently in one set of graphs but infrequently in another set of graphs. Frequent subgraph discovery usually requires to enumerate all the subgraphs satisfying the search criteria, and it is more appropriate for applications in bioinformatics or cheminformatics dealing with large databases of proteins, molecules or chemical compounds. Counting frequency requires to solve many subgraph isomorphism instances, and as such it is computationally challenging. This part of the literature is rather far from the problem studied in this paper, as our structure of interest is a set of vertices whose induced subgraph is dense in a class of graphs and sparse in another: we do not have any notion of frequency, nor we require node-labels, nor need to solve subgraph isomorphism.

Closer to our work is, instead, the literature on dense-subgraph discovery [14]. Among the many notions of *density* studied in the literature, the most popular one is the *average-degree density*, as the problem of extracting the subgraph maximizing such density (commonly referred to as the *densest subgraph*) is solvable in polynomial time [15] and admits a linear time $\frac{1}{2}$ -approximation algorithm [2]. As discussed in Section 4 an alternative notion of dense subgraph is that of optimal quasi-clique introduced by Tsourakakis et al. [36], which has been shown to return denser subgraphs and with smaller diameter than the so-called densest subgraph.

Several recent works have dealt with the problem of extracting dense subgraphs from a set of multiple graphs sharing the same vertex set [5, 12, 17, 29, 40], however none of these deal with extracting contrast or discriminative subgraphs. The work probably most related to ours is due to Yang et al. [44], which study the density contrast subgraph, i.e., the set of vertices whose maximize the difference between the density of its induced subgraphs in two input graphs G^A and G^B . The work of Yang et al. [44] differs from ours as they adopt, as measure of density, the *average degree* [15] and *graph affinity* [8]. They show that their problem is equivalent to the densest subgraph problem with negative edges, which is shown to be NP-hard and hard to approximate. In our work, instead of adopting the densest subgraph definition — as done by Yang et al. [44] — which normalizes by the size of the subgraph S , we follow Tsourakakis et al. [36] and we balance the total edge weight with the term $\alpha \binom{|S|}{2}$. The parameter α also allows us to control the size of the extracted contrast subgraphs. Finally, Yang et al. [44] do not consider the case in which the input is constituted by many graphs in two classes and its potential application in graph classification.

3 PROBLEM STATEMENT

Notation. We consider a dataset \mathcal{D} of observations, where the i -th observation corresponds to a graph $G_i = (V, E_i)$. Without loss of generality we assume that all observation graphs are defined over the same set of vertices V ,³ i.e., brain regions of interest, while the edge set E_i represents connections between vertices in the observation graph G_i . For sake of simplicity of presentation we consider each G_i to be unweighted, even if the model can straightforwardly deal with weighted input graphs. The dataset \mathcal{D} is divided in two groups: the *condition group* $\mathcal{A} = \{G_1^A, \dots, G_{r_A}^A\}$ and the *control group* $\mathcal{B} = \{G_1^B, \dots, G_{r_B}^B\}$.

We aggregate the information in the groups \mathcal{A} and \mathcal{B} in two *summary graphs* $G^{\mathcal{A}} = (V, w^{\mathcal{A}})$ and $G^{\mathcal{B}} = (V, w^{\mathcal{B}})$, respectively. The summary graphs $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$ are undirected and weighted graphs, defined over the same set of vertices V as the original observation graphs, while $w^{\mathcal{A}}, w^{\mathcal{B}} : V \times V \rightarrow \mathbb{R}_+$ are weight functions assigning a value to each pair of vertices and summarizing the edges of the observation graphs in the groups \mathcal{A} and \mathcal{B} .

In one particular instantiation, given two vertices u and v in V , we define $w^{\mathcal{A}}(u, v)$ to be the *fraction* of graphs $G_i^A \in \mathcal{A}$ in which u is incident to v , that is,

$$w^{\mathcal{A}}(u, v) = \frac{1}{r_A} \left| G_i^A \in \mathcal{A} \text{ s.t. } (u, v) \in E_i^A \right|, \quad (1)$$

and similarly for $w^{\mathcal{B}}$. Note that according to this weighting function $w^{\mathcal{A}}(u, v) \in [0, 1]$ with $w^{\mathcal{A}}(u, v) = 0$ denoting the case in which there is no relationship (i.e., no edge) between u and v in $G^{\mathcal{A}}$. Alternative definitions of the weight functions $w^{\mathcal{A}}$ and $w^{\mathcal{B}}$ are possible. For instance, if the input graphs are all weighted, with $w_i^A(u, v)$ denoting the weight of the edge (u, v) in G_i^A , then we can define:

$$w^{\mathcal{A}}(u, v) = \frac{1}{r_A} \sum_{G_i^A \in \mathcal{A}} w_i^A(u, v).$$

³When the observation graphs have different set of vertices, we can consider that they are defined on the union of all vertex sets.

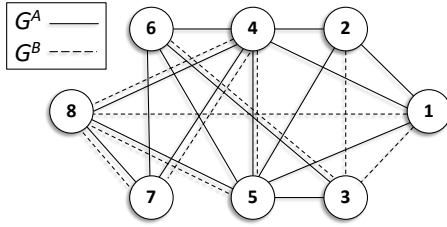


Figure 2: Run-through example of two input graphs $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$. Please refer to Example 1, 2, 3, and 4 for different definitions of contrast subgraphs.

As another example, one could require the summary graph itself to be binary, i.e., $w^{\mathcal{A}}(u, v) \in \{0, 1\}$, for instance by considering if the number of edges present in the observation graphs is larger than a given threshold.

Given a subset of vertices $S \subseteq V$, we consider the subgraph induced by S in the summary graphs $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$. We define

$$e^{\mathcal{A}}(S) = \sum_{u, v \in S} w^{\mathcal{A}}(u, v)$$

the sum of edge weights in the subgraph of $G^{\mathcal{A}}$ induced by the vertex set S . In the case of summary graphs with binary weights, the quantity $e^{\mathcal{A}}(S)$ corresponds to the number of edges in the subgraph induced by S in the summary graph $G^{\mathcal{A}}$. Analogous definitions apply to $G^{\mathcal{B}}$.

Problem definition. The basic definition of the contrast subgraph problem requires to find a subset of vertices whose induced subgraph is dense in a summary graph $G^{\mathcal{A}}$ and sparse in summary graph $G^{\mathcal{B}}$.

PROBLEM 1 (CONTRAST SUBGRAPH). *Given two sets of observation graphs, i.e., the condition group $\mathcal{A} = \{G_1^{\mathcal{A}}, \dots, G_{r_{\mathcal{A}}}^{\mathcal{A}}\}$ and the control group $\mathcal{B} = \{G_1^{\mathcal{B}}, \dots, G_{r_{\mathcal{B}}}^{\mathcal{B}}\}$, and corresponding summary graphs $G^{\mathcal{A}} = (V, w^{\mathcal{A}})$ and $G^{\mathcal{B}} = (V, w^{\mathcal{B}})$, we seek to find a subset of vertices $S^* \subseteq V$ that maximizes the contrast-subgraph objective*

$$\begin{aligned} \delta(S) &= e^{\mathcal{A}}(S) - e^{\mathcal{B}}(S) - \alpha \binom{|S|}{2} \\ &= \sum_{u, v \in S} (w^{\mathcal{A}}(u, v) - w^{\mathcal{B}}(u, v) - \alpha), \end{aligned}$$

where $\alpha \in \mathbb{R}_+$ is a user-defined parameter.

The last term of the objective, i.e., $-\alpha \binom{|S|}{2}$, is a regularization term penalizing solutions of large size: the larger the value of α , the smaller is the optimal contrast subgraph. Note that, to avoid the naïve solution, $S^* = \emptyset$, we have to ensure that

$$0 < \alpha < \max_{u, v \in V} (w^{\mathcal{A}}(u, v) - w^{\mathcal{B}}(u, v)),$$

otherwise we would encounter the case in which every pair of vertices is detrimental for the objective function.

EXAMPLE 1. *Figure 2 provides an example of two summary graphs $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$. In this example we assume that the summary graphs are unweighted, i.e., the edge weight function is binary. For $\alpha = 0.8$ the contrast subgraph distinguishing $G^{\mathcal{A}}$ from $G^{\mathcal{B}}$ is given by the set of vertices $S_3 = \{1, 2, 4\}$. In particular, S_3 forms a clique in $G^{\mathcal{A}}$ and an*

independent set in $G^{\mathcal{B}}$. The contribution of each edge to the objective function is $1 - 0 - 0.8 = 0.2$, giving a total value for the objective $\delta(S_3) = 0.6$. Note that the set $S_4 = \{1, 2, 4, 5\}$ is a larger clique in $G^{\mathcal{A}}$, but it induces an edge in $G^{\mathcal{B}}$. In this case, the contribution to the solution S_4 is 0.2 for all 5 edges appearing only in $G^{\mathcal{A}}$, and -0.8 for the edge appearing in both $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$, giving a total value for the objective $\delta(S_4) = 0.2$.

EXAMPLE 2. *In the example of Figure 2, with $\alpha = 0.5$, the contrast subgraph distinguishing $G^{\mathcal{A}}$ from $G^{\mathcal{B}}$ is given by the set of vertices $S_5 = \{1, 2, 4, 5, 6\}$. In this case, the contribution in the objective of each edge appearing only to $G^{\mathcal{A}}$ (there are 7) is $1 - 0 - 0.5 = 0.5$, while an edge appearing in both graphs (there is only one edge) contributes -0.5 , and the same for an edge that does not appear in none graph (there are 2 edges). Finally, an edge appearing only in $G^{\mathcal{B}}$ (there are no such edge in the subgraph induced by S_5) would contribute -1.5 . Thus, the total value of the objective is $\delta(S_5) = 2$. Note that $S_4 = \{1, 2, 4, 5\}$ achieves the same score.*

The last two examples show the function of the parameter α , which is that of governing the complexity of the extracted patterns. As any other parameter governing the complexity of a model, the “best” value of α might depend on the specific input dataset, as well as on trading-off different requirements of the analyst: for instance, when explainability is important, a smaller contrast subgraph (larger α) might be desirable, but if it is too small might end up being not interesting.

So far we have asked to find the contrast subgraph that distinguishes $G^{\mathcal{A}}$ from $G^{\mathcal{B}}$. Note, however, that our objective function is not symmetric with respect to $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$. Thus, one can consider the same problem for finding the contrast subgraph that distinguishes $G^{\mathcal{B}}$ from $G^{\mathcal{A}}$. Indeed, in our experiments in Section 5 we always consider both contrast subgraphs.

EXAMPLE 3. *In the example of Figure 2, with $\alpha = 0.8$, the optimal contrast subgraph that distinguishes $G^{\mathcal{B}}$ from $G^{\mathcal{A}}$ is a single edge appearing in $G^{\mathcal{B}}$ and not in $G^{\mathcal{A}}$. Thus, any of the following sets is an optimal solution: $\{1, 8\}$, $\{1, 3\}$ or $\{2, 3\}$.*

Symmetric variant. In some cases the analyst may want to find a subgraph having the largest absolute difference, in terms of edge weights, between $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$, disregarding on whether the larger weights are on one side or the other. To address this case, we consider a symmetric variant of Problem 1.

PROBLEM 2 (SYMMETRIC CONTRAST SUBGRAPH). *Given two sets of observation graphs, i.e., the condition group $\mathcal{A} = \{G_1^{\mathcal{A}}, \dots, G_{r_{\mathcal{A}}}^{\mathcal{A}}\}$ and the control group $\mathcal{B} = \{G_1^{\mathcal{B}}, \dots, G_{r_{\mathcal{B}}}^{\mathcal{B}}\}$, and corresponding summary graphs $G^{\mathcal{A}} = (V, w^{\mathcal{A}})$ and $G^{\mathcal{B}} = (V, w^{\mathcal{B}})$, we seek to find a subset of vertices $S^* \subseteq V$ that maximizes the contrast-subgraph objective*

$$\sigma(S) = \sum_{u, v \in S} \left(\left| w^{\mathcal{A}}(u, v) - w^{\mathcal{B}}(u, v) \right| - \alpha \right),$$

where $\alpha \in \mathbb{R}_+$ is a user-defined parameter.

EXAMPLE 4. *In Figure 2 the optimal symmetric contrast subgraph (Problem 2), for $\alpha = 0.5$, is $\{1, 2, 3, 5\}$. This vertex set induces a clique in the union of the two summary graphs, such that each edge belongs only to either $G^{\mathcal{A}}$ or $G^{\mathcal{B}}$, so that every pair of vertices produces a positive contribution to the objective function of Problem 2.*

4 COMPLEXITY AND ALGORITHMS

To tackle the problem of extracting a dense subgraph from a given unweighted graph, Tsourakakis et al. [36] introduce the notion of *edge-surplus* $f_\alpha(S)$ for a subgraph induced by a set of vertices $S \subseteq V$, which they defined as follows

$$f_\alpha(S) = \begin{cases} 0, & S = \emptyset; \\ g(e[S]) - \alpha h(|S|), & \text{otherwise,} \end{cases} \quad (2)$$

where $e[S]$ denotes the number of edges in the induced subgraph, $\alpha > 0$ is a fixed penalty parameter and g and h are two strictly-increasing functions. The rationale behind the definition of edge surplus is to counterbalance two opposite forces: the term $g(e[S])$ favors subgraphs with many edges, whereas the term $\alpha h(|S|)$ penalizes large subgraphs. They propose the problem of extracting the subgraph with maximum edge surplus as a general class of dense-subgraphs problems.

Tsourakakis et al. [36] then focus on a specific instance of the general problem, i.e., $g(x) = x$ and $h(x) = \binom{x}{2}$, which they dub as **OPTIMAL QUASI-CLIQUE (OQC)** problem. They propose an approximation algorithm for the OQC problem, which was later shown to be NP-hard in [35].

More recently, Cadena et al. [7] extend the OQC setting, and consider extracting dense subgraphs in *streaming signed networks for event detection*. They generalize OQC to weighted graphs and edge-dependent penalty parameter α :

PROBLEM 3 (GENERALIZED OPTIMAL QUASI-CLIQUE (GOQC) [7]). Given a graph $G = (V, E)$, and functions $w(u, v)$ and $\alpha(u, v)$, for each pair of vertices $u, v \in V$, find a subset of vertices $S \subseteq V$ that maximizes

$$f(S) = \sum_{u, v \in S} w(u, v) - \alpha(u, v).$$

Cadena et al. prove that Problem 3 is NP-complete and NP-hard to approximate within a factor $O(|V|^{1/2-\epsilon})$ [7, Theorem 1]. Then they develop an algorithm using a semidefinite-programming (SDP) based rounding to produce a solution, which is then refined by the *local-search* procedure of Tsourakakis et al. [36]. Their algorithm provides a $O(\log n)$ approximation guarantee, although in practice the approximation is shown to be much better.

We next show that, although defined in a totally different setting (event detection in a single, streaming, signed network), we can fruitfully make use of the algorithm developed by Cadena et al. to solve the contrast subgraph problems we defined in the previous section. Next proposition provides the mapping between our problems and the GOQC problem.

PROPOSITION 1. *Problems 1 and 2 can be mapped to Problem 3.*

PROOF. The mapping is given by setting $\alpha(u, v) = \alpha$ and

$$w(u, v) = \begin{cases} w^{\mathcal{A}}(u, v) - w^{\mathcal{B}}(u, v) & \text{(Problem 1)} \\ |w^{\mathcal{A}}(u, v) - w^{\mathcal{B}}(u, v)|, & \text{(Problem 2)} \end{cases}$$

for each pair of vertices $u, v \in V$. \square

Based on this proposition, the algorithms that we use for our problems are reported in details in Appendix A.1.

Table 1: Datasets used in the experiments.

Dataset	Description	TD	ASD
Children	Age ≤ 9	52	49
Adolescents	Age in [15, 20]	121	116
EyesClosed	Eyes closed during scanning	158	136
Male	Male individuals	418	420

5 EXPERIMENTS

The aim of our experimental evaluation is to show how contrast subgraphs can be exploited profitably for finding discriminative patterns between two groups of brain networks. In particular, our goal is to answer the following questions:

- (1) Can the contrast-subgraph approach be used to identify structural differences between two groups of brain-networks, which are not easy to detect through standard analysis?
- (2) Are the discovered contrast subgraphs interpretable?
- (3) Are the results obtained with the contrast-subgraph approach consistent with the neuroscience literature?
- (4) Does the approach achieve good performance for the classification tasks considered?

fMRI and the brain network. We start by providing some background on fMRI and brain networks. The analysis reported in this section are based on *resting-state functional MRI (rs-fMRI)* data. The rs-fMRI is a technique whose aim is to measure brain activity in the absence of any underlying controlled experimental paradigm, exploiting the link between neural activity and blood flow and oxygenation. When a neuron activates, the blood flow in the corresponding regions increases, and as a result, oxygen-rich blood displaces oxygen-depleted blood. Such a variation is measured as a signal, dubbed BOLD (blood-oxygen-level dependent), and it is what really influences the final magnetic resonance signal. Biswal et al. [4] were the first to demonstrate temporal coherence between several blood fluctuation in the primary sensor motor cortices and, along with several corroborating results, prompted researchers to think of the brain as a network, where nodes represent regions and edges represent functional connectivity measured by correlation. Since the quantity of signals detected by any MRI is huge, there is usually a process of aggregating voxels, so as to reduce the data dimensionality. Furthermore, since such signals are heavily subject to noise caused by different confounding factors, a plethora of pre-processing strategies, either from the signal-processing side (e.g., filtering, signal correction, etc.) or from the network-analysis side (e.g., thresholding) have been proposed. We refer the interested reader to the comprehensive survey of Lang et al. [21].

Data source. For our experimental evaluation we use a publicly-available dataset⁴ released by the Autism Brain Imagine Data Exchange (ABIDE) project [9]. The dataset contains neuroimaging data of 1112 different patients, 573 Typically Developed (TD) and 539 suffering from Autism Spectrum Disorder (ASD).

To extract the data used in this work, we have followed the pre-processing strategy denoted as DPARSF,⁵ followed by Band-Pass Filtering and Global Signal Regression. To parcellate the brain we

⁴<http://preprocessed-connectomes-project.org/abide/index.html>

⁵<http://preprocessed-connectomes-project.org/abide/dparsf.html>

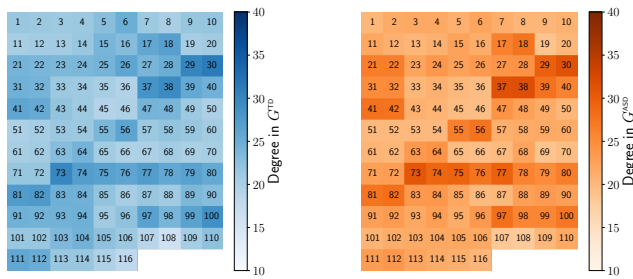


Figure 3: Weighted degree of the 116 nodes (ROIs) in G^{TD} (left) and G^{ASD} (right). The weighted degree of each vertex is almost identical in the two summary graphs, not revealing any structural difference between the two classes.

adopt the AAL atlas [37], which divides the brain into 116 ROIs. The final result of our pre-processing, for a single patient, is a set of 116 time series (each one associated to a ROI) of length 145. In order to obtain the brain connectome we compute the pairwise Pearson correlation between the time series of each pair of ROIs, producing a 116×116 correlation matrix. Finally, we use a threshold t , whose value is set equal to the 80-th percentile of the distribution of correlation values, and we draw an edge for each pair (u, v) having correlation larger than t ; these choices are typical in the literature, see for example the works of Lord et al. [23] and Rubinov et al. [30].

The end result of our data processing is an *undirected unweighted graph* for each patient.

It should be noted that fMRI data, in addition to capturing the status of an individual, are influenced by other phenotypic information, which for our purposes would result to confounding factors with respect to our target variable, i.e., the status TD and ASD. As an effort to mitigate the intrinsic variance and in accordance with neuroscience literature, we create four different datasets by selecting individuals who share some common characteristics, such as age, gender, or the condition during the scan, e.g., eyes closed.

More information about the four selected datasets, and the sizes of the two target classes, is reported in Table 1. We remind that in all the datasets the classes of the observations are TD and ASD, the name we use to identify each dataset is the common phenotypic features shared by the observations. Children contains only individuals whose age is at most 9 years. Adolescents contains individuals whose age is between 15 and 20 years. EyesClosed contains individuals who performed their fMRI with eyes closed. Finally, Male contains only male individuals. For all our datasets we consider only individuals for whom there are no missing observations in the time series, and we apply the pre-processing described above. As a result, each individual is represented by an undirected unweighted graph with $|V| = 116$ vertices.

5.1 Characterization

In this section we characterize the type of information produced by the approach of contrast subgraphs, focusing on the Children dataset. A preview of a contrast subgraph for this dataset was already given in Figure 1 (Section 1.2).

The first natural question we consider is whether the contrast subgraphs capture some simple, first-level information, which

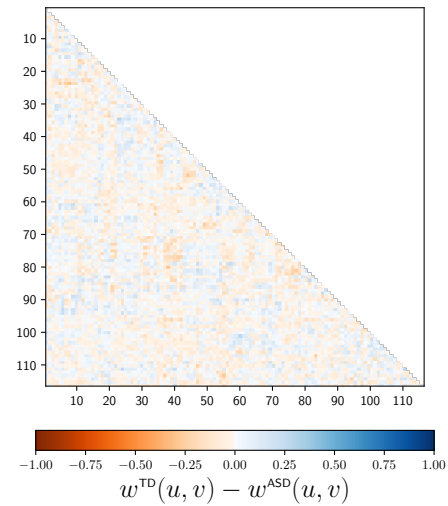


Figure 4: Adjacency matrix of $G^{\text{TD-ASD}}$. Elements of the matrix are colored according $w^{\text{TD}} - w^{\text{ASD}}$. Nodes are ordered by their id in the AAL atlas.

might characterize the two different classes. To answer this question, we check the two summary graphs G^{TD} and G^{ASD} , produced according to Equation (1). These are two weighted graphs, where edge weights take values in the interval $[0, 1]$, representing the fraction of individuals in the class who exhibit the edge.

In Figure 3 we report, for each of the 116 ROIs, their (weighted) degree in G^{TD} and G^{ASD} . The degrees of the nodes in a network is a key feature. If there was some evident structural difference between G^{TD} and G^{ASD} , we would expect to see different with respect to the degrees of their nodes in the two graphs. Instead, in Figure 3 we can observe *very similar* degrees for all ROIs in both summary graphs. In other words, we are not able to identify a few nodes that are “important” in G^{TD} and not in G^{ASD} , or vice versa.

We thus move our attention to the edge weights in the *difference graph* $G^{\text{TD-ASD}} = (V, w^{\text{TD}} - w^{\text{ASD}})$, looking for patterns that are distinct in one network or the other. The adjacency matrix of $G^{\text{TD-ASD}}$ is reported in Figure 4.⁶ It is particularly interesting to check whether in this adjacency matrix there are regions with a predominant color. Such regions could be either single cells or rows/columns, and they can be either dark red or dark blue. In the case of single cells, such dark regions would represent a single edge showing a large weight difference between the two classes. In the case of rows/columns, such dark regions would represent ROIs with a tendency to be hyper/hypo-connected in a specific class. However, we observe that none such pattern appears. Instead, the whole matrix is represented with light-colored cells (i.e., values are very close to 0), exhibiting very weak structural difference between the two classes.

Contrast subgraph (Problem 1). Despite the apparent lack of structural differences between the two classes, we have seen already in Section 1.2 (Figure 1) an example of a contrast subgraph, which provides simple discrimination rules between the two classes,

⁶We do not show adjacency matrix of $G^{\text{ASD-TD}}$ since it would be equal to the one in Figure 4 with opposite signs.

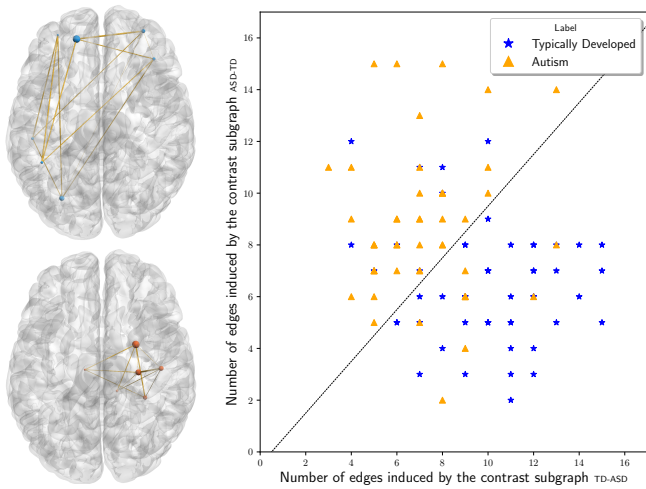


Figure 5: Contrast subgraphs from the Children dataset (obtained with a larger value of α w.r.t Figure 1). Top left: TD-ASD contrast subgraph. Bottom left: ASD-TD contrast subgraph. Size of nodes and thickness of edges are respectively proportional to their degree and weight in the *difference graph*. Edges shown are the ones whose weight is greater than 0.1. On the right, scatter plot showing for each individual the number of edges present in the subgraph induced by the contrast subgraph TD-ASD (x -axis) and by the contrast subgraph ASD-TD (y -axis).

TD and ASD. Another example, obtained using a larger value of α and thus yielding a smaller solution, is reported in Figure 5. In this setting the contrast subgraph TD-ASD (top left) contains only 7 nodes, while the contrast subgraph ASD-TD (bottom left) contains only 6 nodes. As in Figure 1, the size of a node in the contrast subgraph TD-ASD (respectively, ASD-TD) is proportional to its weighted degree in the *difference graph* $G^{\text{TD-ASD}}$ (respectively, $G^{\text{ASD-TD}}$). An induced edge (u, v) is depicted in the figure of the contrast subgraph TD-ASD only if $w^{\text{TD}} - w^{\text{ASD}}(u, v) \geq 0.1$ (and similarly for ASD-TD). Inspection of the data points in Figure 5 motivates us to derive a simple rule:

If the number of edges induced by the contrast subgraph ASD-TD is smaller than the number of edges induced by the contrast subgraph TD-ASD, then there are high chances that the individual is affected by ASD.

Khan et al. [19] have discovered high connectivity between cerebellum and prefrontal cortex, posterior parietal cortex, inferior and middle temporal gyri, as a fingerprint distinguishing TD children from children with ASD. These findings are consistent with the contrast subgraph TD-ASD in Figure 1. Di Martino et al. [24] have discovered hyper-connectivity in ASD children between the regions of striatum and the limbic cortex, which is consistent with the contrast subgraph ASD-TD in Figure 1.

Furthermore, by inspecting Figure 5 we are able to discover another insight: the ubiquitous presence of ROIs belonging to the left hemisphere in the contrast subgraph TD-ASD. This observation, again, is coherent with neuroscience findings: the effect of some

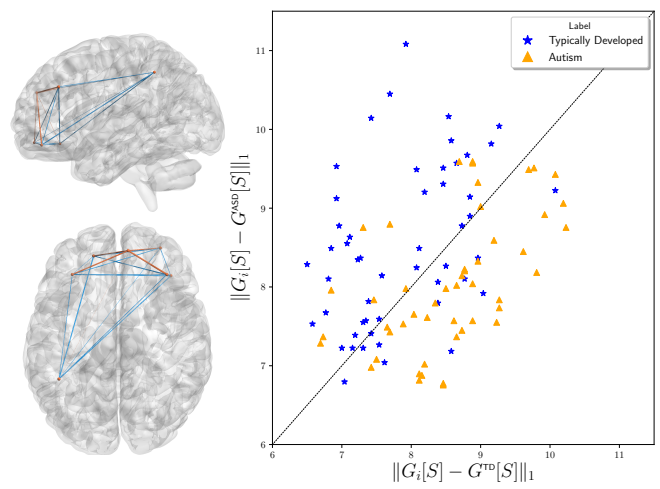


Figure 6: Example of contrast subgraphs from the Children dataset according to Problem 2. Top left: sagittal view. Bottom left: axial view. Thickness of edges is proportional to the absolute value of their weight in the *difference graph*. Blue edges represent the ones that have positive weight in $G^{\text{TD-ASD}}$, orange ones in $G^{\text{ASD-TD}}$. Edges shown are the ones whose weight in absolute value is greater than 0.1. Such formulation of the problem highlights again that prefrontal cortex and posterior parietal cortex are regions of hyper-connectivity in TD patients, but also that at the same time such regions shows hyper-connectivity with medial prefrontal cortex in ASD patients, and hypo-connectivity in TD ones. Such contemporaneity is caught by the maximization of specularity that Problem 2 operates, instead of taking into account difference.

neural function being specialized in only one side of the brain is called left/right lateralization of the brain. For example, when it comes to the task of speech production, it has been shown that most humans exhibit a left lateralization of the brain, however, it has been reported that ASD patients show significantly reduced left lateralization in speech-production-related ROIs [11, 20, 28], such as inferior frontal gyrus, inferior parietal lobule, and superior temporal gyrus, which mostly appear also in the solution we provide.

Symmetric contrast subgraph (Problem 2). In the case of the symmetric problem definition, we extract a single contrast subgraph for the two classes. An example is shown in Figure 6: blue edges show edges with positive weight in $G^{\text{TD-ASD}}$, and orange edges show edges with positive weight in $G^{\text{ASD-TD}}$. There are seven ROIs involved in this solution: five of those already appeared in the contrast subgraphs TD-ASD, previously shown in Figure 1 and Figure 5. All these five ROIs belong either to prefrontal cortex or to posterior parietal cortex, which confirms that such ROIs are discriminative for the two classes. The solution is completed by right middle frontal gyrus and superior frontal gyrus medial. This finding is coherent with the results of Gilbert et al. [13] who provide evidence that the medial prefrontal cortex has an anomalous behavior in ASD patients during several tasks.

Given that in this formulation of the problem there is a single contrast subgraph S for the two classes, to produce features to classify a patient G_i we compute the L_1 norm between the subgraph induced by S in G_i and the subgraph induced by S in G^{TD} and G^{ASD} , respectively. The results are shown in the scatter plot of Figure 6 (right), where we show each graph G_i with value $\|G_i[S] - G^{\text{TD}}[S]\|_1$ on the x -axis and value $\|G_i[S] - G^{\text{ASD}}[S]\|_1$ on the y -axis. Here, $G[S]$ denotes the subgraph induced by S in G . Also in this case, we can see a very good separability defined by the diagonal.

5.2 Classification

We evaluate the effectiveness of contrast subgraphs in classifying brain networks according to the target classes TD and ASD. We compare with four baselines: graph2vec [27], sub2vec [1], WL-Kern [31], and Embs [16]. All baselines provide a feature map, which we use for the classification task. Details about the implementation and parameter tuning of these methods are reported in Appendix A.2.

For our method, we use the same features as in Figure 5 (we refer to this approach as CS-P1) and Figure 6 (we refer to this approach as CS-P2): under both settings our method uses only two features.

We employ the SVM-based classification process of Narayanan et al. [27]. We randomly split the data into 80/20 training/test subsets. Using the training set and 5-fold cross validation we select the best hyperparameters for the classifier. We then apply the best classifier to the test set. We repeat this process 5 times for each method.

We report the average accuracy for all the classification experiments in Table 2. We can observe the overall good performance of our method in all these tasks. Moreover, a number of observations can be drawn:

Only two features: used as a proof of concept, our method was tested employing only two features derived by contrast subgraphs. In more complex classification tasks, however, one can consider additional features from different contrast subgraphs so as to improve the classification accuracy.

A single parameter: our method has only one parameter to tune, i.e., α , while the other methods (with the exception of WL-Kern) require tuning of many parameters.

Run-time efficiency: extracting the contrast subgraphs from this type of brain connectome datasets always take less than 30 seconds, while the other methods (with the exception of WL-Kern) have longer running times.

Explainability: in the pipeline we presented, our method uses only two simple features, instead of embedded features or convoluted kernels. These features enable simple description in form of rule, as well as visualization, of the decision boundary discriminating between the two classes.

Figures 7 and 8 reports the decision boundaries for CS-P1 and CS-P2 (respectively).

6 CONCLUSIONS AND FUTURE WORK

Learning models that are able to discriminate brains of patients affected by a mental disorder from those of healthy individuals, is attracting a lot of interest. However, often the accuracy of the models is a predominant goal over its interpretability, which is instead a key requirement in neuroscience.

Table 2: Results of the experiments performed over the datasets described in Table 1. Each value represents the average accuracy, along with its relative standard deviation.

	Children	Adolescents	EyesClosed	Male
CS-P1	0.86 \pm 0.07	0.72 \pm 0.07	0.71 \pm 0.03	0.63 \pm 0.01
CS-P2	0.86 \pm 0.04	0.71 \pm 0.04	0.72 \pm 0.08	0.65 \pm 0.03
graph2vec	0.72 \pm 0.13	0.65 \pm 0.05	0.60 \pm 0.04	0.56 \pm 0.02
sub2vec	0.66 \pm 0.01	0.59 \pm 0.04	0.60 \pm 0.01	0.57 \pm 0.01
WL-Kern	0.52 \pm 0	0.52 \pm 0	0.54 \pm 0	0.50 \pm 0
Embs	0.70 \pm 0.12	0.58 \pm 0.03	0.59 \pm 0.04	0.57 \pm 0.03

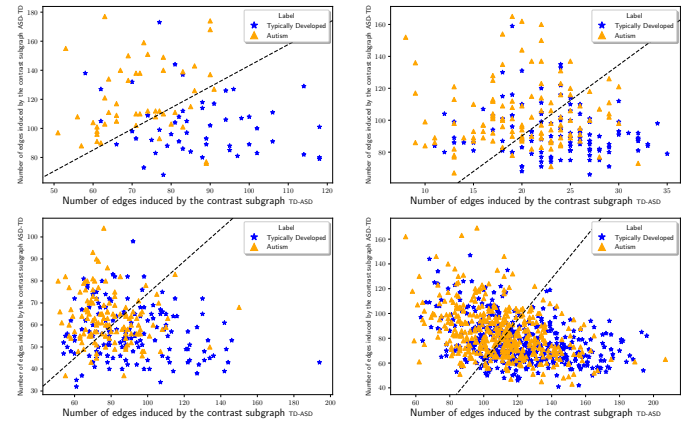


Figure 7: Decision boundaries for CS-P1. Top Children (left), Adolescents (right). Bottom EyesClosed (left), Male (right).

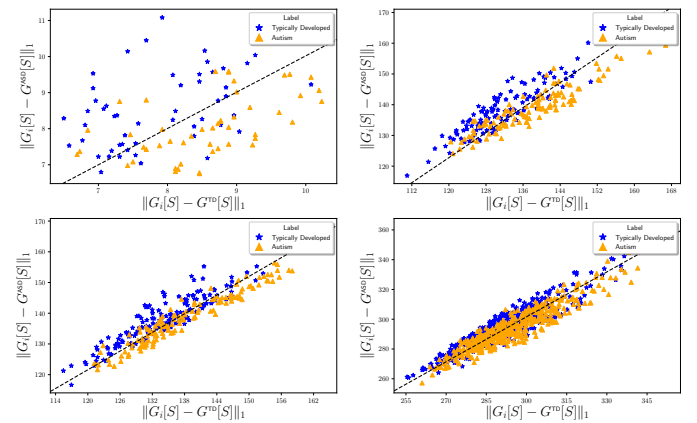


Figure 8: Decision boundaries for CS-P2. Top Children (left), Adolescents (right). Bottom EyesClosed (left), Male (right).

In this paper we approach the task of brain-network classification with a two-fold goal: to achieve good accuracy, but most importantly, to identify discriminant brain patterns that lead a model to classify an individual. The framework we propose, based on the notion of *contrast subgraph*, satisfies both of these conditions. It outperforms several state-of-the-art competitors and returns some very intuitive explanations, which can be shared with experts from the neuroscience field. Moreover, contrast subgraphs are exceptionally easy to compute, both in terms of runtime and memory.

Future work. In this first work we focus on Autism Spectrum Disorder mostly because of the wide availability of public datasets. However, the proposed approach can be applied in several other contexts and different types of mental disorders. To push forward this research direction, we have created an ongoing collaboration with neuroscientists. We are currently working with the domain experts to create datasets with patients affected by Bipolar Disorder and by Schizophrenia.

On the algorithmic side, we plan to investigate the extraction of the top- k contrast subgraphs, following an approach similar to that of Balalau et al. [3] for densest subgraphs. Exploiting the information given by multiple contrast subgraphs, could help both to improve the classification performances, and to detect further interesting patterns that may not appear in the single solution.

Acknowledgements. The authors wish to thank Edoardo Galimberti for many meaningful discussions in the early stage of this project. Francesco Bonchi acknowledges support from Intesa Sanpaolo Innovation Center. Aristides Gionis is supported by three Academy of Finland projects (286211, 313927, 317085), the ERC Advanced Grant REBOUND (834862), the EC H2020 RIA project “SoBigData++” (871042), and the Wallenberg AI, Autonomous Systems and Software Program (WASP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- [1] B. Adhikari, Y. Zhang, N. Ramakrishnan, and B. A. Prakash. Sub2vec: Feature learning for subgraphs. In *PAKDD 2018*.
- [2] Y. Asahiro, K. Iwama, H. Tamaki, and T. Tokuyama. Greedily finding a dense subgraph. *Journal of Algorithms*, 34(2):203–221, 2000.
- [3] O. D. Balalau, F. Bonchi, T. H. Chan, F. Gullo, and M. Sozio. Finding subgraphs with maximum total density and limited overlap. In *WSDM 2015*.
- [4] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34(4):537–541, oct 1995.
- [5] B. Boden, S. Gunnemann, H. Hoffmann, and T. Seidl. Mining coherent subgraphs in multi-layer graphs with edge labels. In *KDD 2012*.
- [6] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.
- [7] J. Cadena, A. K. Vullikanti, and C. C. Aggarwal. On dense subgraphs in signed network streams. In *ICDM 2016*.
- [8] L. Chu, S. Wang, S. Liu, Q. Huang, and J. Pei. ALID: scalable dominant cluster detection. *PVLDB*, 8(8):826–837, 2015.
- [9] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, C. Yan, and P. Bellec. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, (41), 2013.
- [10] Y. Du, Z. Fu, and V. D. Calhoun. Classification and prediction of brain disorders using functional connectivity: Promising but challenging. *Frontiers in Neuroscience*, 12:525, 2018.
- [11] L. T. Eylar, K. Pierce, and E. Courchesne. A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, 135(3):949–960, Feb. 2012.
- [12] E. Galimberti, F. Bonchi, and F. Gullo. Core decomposition and densest subgraph in multilayer networks. In *CIKM 2017*.
- [13] S. J. Gilbert, G. Bird, R. Brindley, C. D. Frith, and P. W. Burgess. Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: An fMRI study of two executive function tasks. *Neuropsychologia*, 46(9):2281–2291, July 2008.
- [14] A. Gionis and C. E. Tsourakakis. Dense subgraph discovery: Kdd 2015 tutorial. In *KDD 2015*.
- [15] A. V. Goldberg. Finding a maximum density subgraph. Technical report, University of California at Berkeley, 1984.
- [16] L. Gutiérrez-Gómez and J.-C. Delvenne. Unsupervised network embeddings with node identity awareness. *Applied Network Science*, 4(1), oct 2019.
- [17] D. Jiang and J. Pei. Mining frequent cross-graph quasi-cliques. *TKDD*, 2(4):16, 2009.
- [18] N. Jin and W. Wang. Lts: Discriminative subgraph mining by learning from search history. In *ICDE 2011*.
- [19] A. J. Khan, A. Nair, C. L. Keown, M. C. Datko, A. J. Lincoln, and R.-A. Müller. Cerebro-cerebellar resting-state functional connectivity in children and adolescents with autism spectrum disorder. *Biological Psychiatry*, 78(9):625 – 634, 2015. Adolescence, Cortical Functional Connectivity, and Psychopathology.
- [20] N. M. Kleinhans, R.-A. Müller, D. N. Cohen, and E. Courchesne. Atypical functional lateralization of language in autism spectrum disorders. *Brain Research*, 1221:115 – 125, 2008.
- [21] E. W. Lang, A. M. Tomé, I. R. Keck, J. M. Górriz-Sáez, and C. G. Puntonet. Brain connectivity analysis: A short survey. *Computational Intelligence and Neuroscience*, 2012:1–21, 2012.
- [22] J. B. Lee, R. Rossi, and X. Kong. Graph classification using structural attention. In *KDD 2018*.
- [23] A. Lord, D. Horn, M. Breakspear, and M. Walter. Changes in community structure of resting state functional connectivity in unipolar depression. *PLOS ONE*, 7(8):1–15, 08 2012.
- [24] A. D. Martino, C. Kelly, R. Grzadzinski, X.-N. Zuo, M. Mennes, M. A. Mairena, C. Lord, F. X. Castellanos, and M. P. Milham. Aberrant striatal functional connectivity in children with autism. *Biological Psychiatry*, 69(9):847 – 856, 2011. Genes, Autism, and Associated Phenotypes.
- [25] L. Meng and J. Xiang. Brain network analysis and classification based on convolutional neural network. *Frontiers in Computational Neuroscience*, 12, dec 2018.
- [26] M. F. Misman, A. A. Samah, F. A. Ezudin, H. A. Majid, Z. A. Shah, H. Hashim, and M. F. Harun. Classification of adults with autism spectrum disorder using deep neural network. In *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pages 29–34, Sep. 2019.
- [27] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, and S. Jaiswal. graph2vec: Learning distributed representations of graphs, 2017.
- [28] J. A. Nielsen, B. A. Zielinski, P. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson. Abnormal lateralization of functional connectivity between language and default mode regions in autism. *Molecular Autism*, 5(1):8, 2014.
- [29] J. Pei, D. Jiang, and A. Zhang. On mining cross-graph quasi-cliques. In *KDD 2005*.
- [30] M. Rubinov, S. A. Knock, C. J. Stam, S. Micheloyannis, A. W. Harris, L. M. Williams, and M. Breakspear. Small-world properties of nonlinear brain activity in schizophrenia. *Human Brain Mapping*, 30(2):403–416, 2009.
- [31] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, and K. M. Borgwardt. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 12(Sep):2539–2561, 2011.
- [32] O. Sporns, G. Tononi, and R. Kötter. The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 1(4), 2005.
- [33] M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A. Smola, L. Song, P. S. Yu, X. Yan, and K. M. Borgwardt. Discriminative frequent subgraph mining with optimality guarantees. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 3(5):302–318, 2010.
- [34] R. M. H. Ting and J. Bailey. Mining minimal contrast subgraph patterns. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 639–643, 2006.
- [35] C. Tsourakakis. Streaming Graph Partitioning in the Planted Partition Model. In *COSN 2015*.
- [36] C. Tsourakakis, F. Bonchi, A. Gionis, F. Gullo, and M. Tsiarli. Denser than the densest subgraph: extracting optimal quasi-cliques with quality guarantees. In *KDD 2013*.
- [37] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, 15(1):273 – 289, 2002.
- [38] S. Wang, L. He, B. Cao, C.-T. Lu, P. S. Yu, and A. B. Ragin. Structural deep brain network mining. In *KDD 2017*.
- [39] J. Wu, J. He, and J. Xu. Demo-net: Degree-specific graph neural networks for node and graph classification. In *KDD 2019*.
- [40] Y. Wu, X. Zhu, L. Li, W. Fan, R. Jin, and X. Zhang. Mining dual networks: Models, algorithms, and applications. *TKDD*, 10(4):40:1–40:37, 2016.
- [41] X. Yan, H. Cheng, J. Han, and P. S. Yu. Mining significant graph patterns by leap search. In *SIGMOD 2008*.
- [42] Y. Yan, J. Zhu, M. Duda, E. Solarz, C. Sripada, and D. Koutra. Groupinn: Grouping-interpretable neural network for classification of limited, noisy brain data. In *KDD 2019*.
- [43] P. Yanardag and S. Vishwanathan. Deep graph kernels. In *KDD 2015*.
- [44] Y. Yang, L. Chu, Y. Zhang, Z. Wang, J. Pei, and E. Chen. Mining density contrast subgraphs. In *ICDE 2018*.

A REPRODUCIBILITY

A.1 Pseudocode

We provide pseudocodes of the algorithms used for computing contrast subgraphs, according to Section 4. In particular, Algorithm 1 solves Problem 1, Algorithm 2 solves Problem 2, Algorithm 3 displays the algorithm DENS DP by Cadena et al. [7] based on the semi-definite programming (SDP) routine, while Algorithm 4 describes the local-search routine by Tsourakakis et al. [36], which refines the solution produced by the SDP method.

All our code and datasets are made available at:
<https://github.com/tlancian/contrast-subgraph>

Algorithms 1 and 2 simply apply Proposition 1 in order to map Problem 1 and 2 (respectively) to Problem 3, and then invoke Algorithm 3 over the difference graph $G^{\mathcal{A}-\mathcal{B}}$ and with the $\bar{\alpha}$ function taking constantly the α parameter of our problems.

The subroutine dubbed SDP in Algorithm 3, corresponds to the following semi-definite programming:

$$\begin{aligned} \max_{(u,z) \in E} \quad & \sum_{(u,z) \in E} w(u,z) \left(\frac{1 + \mathbf{v}_u \mathbf{v}_0 + \mathbf{v}_z \mathbf{v}_0 + \mathbf{v}_u \mathbf{v}_z}{4} \right) - \\ & \sum_{u,z \in V, u \neq z} \bar{\alpha}(u,z) \left(\frac{1 + \mathbf{v}_u \mathbf{v}_0 + \mathbf{v}_z \mathbf{v}_0 + \mathbf{v}_u \mathbf{v}_z}{4} \right) \\ \text{s.t.} \quad & \\ & \mathbf{v}_u^T \cdot \mathbf{v}_u = 1 \\ & \mathbf{v}_0, \mathbf{v}_u \in \mathbb{R}^{n+1}. \end{aligned}$$

The function $f_\alpha(\cdot)$ in Algorithm 4 corresponds to the notion of *edge-surplus*, mentioned in Section 4, Equation (2), by Tsourakakis et al. [36].

Algorithm 1 Contrast Subgraph (Problem 1)

Input: $\mathcal{A} = \{G_1^A, \dots, G_{r_A}^A\}$; $\mathcal{B} = \{G_1^B, \dots, G_{r_B}^B\}$; $\alpha \in \mathbb{R}_+$

Output: $S^* \subseteq V$, i.e., the solution to Problem 1

- 1: **for all** $(u, v) \in V \times V$ **do**
- 2: $w^{\mathcal{A}}(u, v) \leftarrow \frac{1}{r_A} |G_i^A \in \mathcal{A} \text{ s.t. } (u, v) \in E_i^A|$ {Eq. (1)}
- 3: $w^{\mathcal{B}}(u, v) \leftarrow \frac{1}{r_B} |G_i^B \in \mathcal{B} \text{ s.t. } (u, v) \in E_i^B|$
- 4: $G^{\mathcal{A}-\mathcal{B}} \leftarrow (V, w^{\mathcal{A}} - w^{\mathcal{B}})$ {Prop. 1}
- 5: $S^* \leftarrow \text{DENS DP}(G^{\mathcal{A}-\mathcal{B}}, \bar{\alpha}(\cdot) = \alpha)$ {Alg. 3}

Algorithm 2 Symmetric Contrast Subgraph (Problem 2)

Input: $\mathcal{A} = \{G_1^A, \dots, G_{r_A}^A\}$; $\mathcal{B} = \{G_1^B, \dots, G_{r_B}^B\}$; $\alpha \in \mathbb{R}_+$

Output: $S^* \subseteq V$, i.e., the solution to Problem 2

- 1: **for all** $(u, v) \in V \times V$ **do**
- 2: $w^{\mathcal{A}}(u, v) \leftarrow \frac{1}{r_A} |G_i^A \in \mathcal{A} \text{ s.t. } (u, v) \in E_i^A|$ {Eq. (1)}
- 3: $w^{\mathcal{B}}(u, v) \leftarrow \frac{1}{r_B} |G_i^B \in \mathcal{B} \text{ s.t. } (u, v) \in E_i^B|$
- 4: $G^{\mathcal{A}-\mathcal{B}} \leftarrow (V, |w^{\mathcal{A}} - w^{\mathcal{B}}|)$ {Prop. 1}
- 5: $S^* \leftarrow \text{DENS DP}(G^{\mathcal{A}-\mathcal{B}}, \bar{\alpha}(\cdot) = \alpha)$ {Alg. 3}

Algorithm 3 DENS DP [7]

Input: Weighted graph $G = (V, w)$; function $\bar{\alpha} : V \times V \rightarrow \mathbb{R}_+$

Output: $S^* \subseteq V$, i.e., the solution to Problem 3

- 1: $\mathcal{V} \leftarrow \text{SDP}(G, \bar{\alpha})$, where $\mathcal{V} = \{\mathbf{v}_0\} \cup \{\mathbf{v}_u \mid u \in V\}$
- 2: Sample \mathbf{r} from $\mathcal{N} \sim (\mathbf{0}_{(|V|+1)}, I_{(|V|+1) \times (|V|+1)})$
- 3: $T \leftarrow \sqrt{4 \log |V|}$
- 4: **for** $u \in V$ **do**
- 5: $z_u \leftarrow (\mathbf{v}_u \cdot \mathbf{r})/T$
- 6: **if** $|z_u| > 1$ **then** $y_u \leftarrow z_u/|z_u|$ **else** $y_u \leftarrow z_u$
- 7: $x_u \leftarrow \begin{cases} 1, & \text{with probability } \frac{1+y_u}{2} \\ -1 & \text{with probability } \frac{1-y_u}{2} \end{cases}$
- 8: $S' \leftarrow \{u \mid x_u = 1\}$
- 9: $S \leftarrow \text{LocalSearch}(G' = (S', w), \alpha = \bar{\alpha}(\cdot))$ {Alg. (4)}

Algorithm 4 LocalSearch [36]

Input: Weighted graph $G = (V, w)$, $\alpha \in \mathbb{R}_+$, maximum number of iterations T_{MAX}

Output: $S^* \subseteq V$

- 1: $S \leftarrow \{v\}$, where v is chosen uniformly at random
- 2: $b_1 \leftarrow \text{TRUE}$, $t \leftarrow 1$
- 3: **while** b_1 **and** $t \leq T_{MAX}$ **do**
- 4: $b_2 \leftarrow \text{TRUE}$
- 5: **while** b_2 **do**
- 6: **if** $\exists u \in V \setminus S$ s.t. $f_\alpha(S \cup \{u\}) \geq f_\alpha(S)$ **then**
- 7: $S \leftarrow S \cup \{u\}$
- 8: **else**
- 9: $b_2 \leftarrow \text{FALSE}$
- 10: **if** $\exists u \in S$ s.t. $f_\alpha(S \setminus \{u\}) \geq f_\alpha(S)$ **then**
- 11: $S \leftarrow S \setminus \{u\}$
- 12: **else**
- 13: $b_1 \leftarrow \text{FALSE}$
- 14: $t \leftarrow t + 1$
- 15: $S^* \leftarrow \arg \max_{\hat{S} \in \{S, V \setminus S\}} f_\alpha(\hat{S})$

A.2 Competitors and parameters tuning

In this section we provide further details of the experimental setup we described in Section 5.2.

Embeddings. In order to obtain embeddings for the baseline methods, we followed different strategies. `graph2vec` [27] and `Embs` [16] provide directly embeddings for a given set of networks. `sub2vec` [1] provides an embedding for a set of subgraphs in a single network, therefore we employed the strategy of Adhikari et al [1], and consider any brain network to be a specific subgraph of a single network made by the union of all these subgraphs. Regarding `WL-Kern` [31], we employed the feature map it produces for computing subsequently the kernel matrix.

Code source and parameters. We next provide links to source code of all the algorithms compared, together with the parameter settings used in the experiments. The ones not mentioned in this section, have been left to their default value in their respective sources. We have approached to the parameter tuning, taking values around the default value set by the respective authors. For every algorithms listed here, we have performed the experiment for any possible combination of this parameters, retaining only the one

that led to the best results, in term of average accuracy over 5 experiments.

CS-P1

Source: <https://github.com/tlancian/contrast-subgraph>

Parameters tuned:

- α for TD-ASD : {70, 75, 80, 85, 90, 93, 95, 97}
- α for ASD-TD : {70, 75, 80, 85, 90, 93, 95, 97}

Best Parameters:

	Children	Adolescents	EyesClosed	Male
α for TD-ASD	70	95	75	75
α for ASD-TD	80	70	75	70

In addition, we report that the α values used to produce Figure 1 are the same that resulted to be the best results for Children. The ones for Figure 5 are respectively 90 and 93.

CS-P2

Source: <https://github.com/tlancian/contrast-subgraph>

Parameters tuned:

- α : {70, 75, 80, 85, 90, 93, 95, 97}

Best Parameters:

	Children	Adolescents	EyesClosed	Male
α	70	75	70	70

In addition, we report that the α used to produce Figure 6 is the same that resulted to be the best results for Children.

graph2vec [27]

Source: https://github.com/MLDroid/graph2vec_tf

Parameters tuned:

- Batch Size: {64, 128, 256, 512}
- Epochs: {100}
- Embedding Size: {256, 512, 1024, 2048}
- Negative Samples: {10, 20, 30, 40}
- Height WL Kernel: {3, 4, 5}

Best Parameters:

	Children	Adolescents	EyesClosed	Male
Batch Size	256	256	64	128
Embedding Size	256	2048	512	1024
Negative Samples	30	30	30	20
Height WL Kernel	3	3	3	3

sub2vec [1]

Source: <https://goo.gl/Ef4q8g>

Parameters tuned:

- Property: {Neighborhood, Structural}
- Walk Length: {100000, 150000, 200000}
- Embedding Size: {128, 256, 512}
- Model: {DBON, DM}
- Training Iterations: {10, 20}

Best Parameters:

	Children	Adolescents	EyesClosed	Male
Property	S	S	S	S
Walk Length	100000	150000	150000	150000
Embedding Size	512	128	128	128
Model	DM	DM	DBON	DBON
Training Iterations	10	20	20	20

Embs [16]

Source: <https://github.com/leoguti85/GraphEmbs>

Parameters tuned:

- Epochs: {100, 300, 500}
- Batch Size: {64, 128, 256}
- Embedding Size: {128, 256, 512, 1024}
- Noise: {0.3, 0.5}

Best Parameters:

	Children	Adolescents	EyesClosed	Male
Epochs	500	500	300	100
Batch Size	64	128	256	256
Embedding Size	1024	1024	1024	1024
Noise	0.5	0.3	0.5	0.3

WL-Kern [31]

Source: <http://mlcb.is.tuebingen.mpg.de/Mitarbeiter/Nino/WL>

Parameters tuned:

- h : {1, 2, 3, 4, 5}

Best Parameters:

	Children	Adolescents	EyesClosed	Male
h	3	3	3	3