Differentially Describing Groups of Graphs

Corinna Coupette,[*1] **Sebastian Dalleiger,**[*2] **Jilles Vreeken**[2]

[1] Max Planck Institute for Informatics
[2] CISPA Helmholtz Center for Information Security

Abstract

How does neural connectivity in autistic children differ from neural connectivity in healthy children or autistic youths? What patterns in global trade networks are shared across classes of goods, and how do these patterns change over time? Answering questions like these requires us to differentially describe groups of graphs: Given a set of graphs and a partition of these graphs into groups, discover what graphs in one group have in common, how they systematically differ from graphs in other groups, and how multiple groups of graphs are related. We refer to this task as *graph group analysis*, which seeks to describe similarities and differences between graph groups by means of statistically significant subgraphs. To perform graph group analysis, we introduce GRAGRA, which uses maximum entropy modeling to identify a non-redundant set of subgraphs with statistically significant associations to one or more graph groups. Through an extensive set of experiments on a wide range of synthetic and real-world graph groups, we confirm that GRAGRA works well in practice.

Introduction

Differentially describing groups of graphs lies at the heart of many scientific and societal challenges. Neuroscientists, for example, might want to characterize brain activity in healthy subjects, elucidate how it differs from brain activity in subjects diagnosed with certain disorders or diseases (e.g., autism or Alzheimer's), and investigate whether their findings are the same across different groups of subjects (e.g., children, adolescents, and adults; or men and women). Policymakers, security experts, and epidemiologists alike could seek to understand patterns of human mobility, be it to improve the resilience of traffic infrastructure to random failures and targeted attacks, or to curb the spread of infectious diseases. And international economists might want to investigate patterns of world trade, e.g., imports and exports between countries, and ask how these vary across different years and product classes.

We refer to the common task underlying these scenarios as *graph group analysis*: Given a set of graphs and a partition of this set into *graph groups*, succinctly summarize the commonalities and differences between graphs in the same
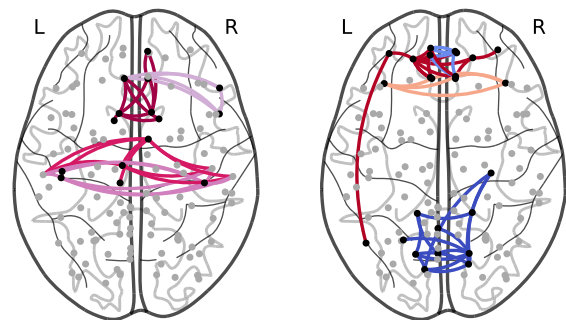

Figure 1: GRAGRA discovers common and contrastive graph patterns in noisy, heterogeneous groups of graphs, capturing, e.g., systematic similarities (left) and differences (right) between the functional brain networks of adolescents with and without autism spectrum disorder. Here, nodes represent centers of mass for brain regions from the AAL Atlas, and edge color classes correspond to significant subgraphs shared between (left) or specific to (right) groups, with individual edges signaling strong connectivity between regions.

group, between graphs in different groups, and between the relationships connecting the groups. In this paper, we formalize graph group analysis as a maximum likelihood modeling problem, using *significant subgraphs* as graph patterns to factorize our probability distribution. We introduce GRAGRA (*Gra*ph *gr*oup *a*nalysis) as an algorithm to solve this problem, which jointly discovers a set of graph patterns and an assignment of these patterns to graph groups.

As a real-world example of graph group analysis, consider Fig. 1. Here, we show the top shared (left) and specific (right) patterns identified in resting-state functional brain networks of adolescents with and without autism spectrum disorder, where nodes in the graphs correspond to brain regions, and edges signal strong connectivity between regions. On the right, patterns with red edges are characteristic of autistic adolescents, and patterns with blue edges are characteristic of non-autistic adolescents. They indicate over- and underconnectivity, respectively, in the brains of autistic adolescents when compared to typically developed controls. Although there is no consensus regarding the relationships between autism and neural connectivity (Hull et al. 2017), our

---

method identifies graph patterns that permit neuroscientific interpretation: For example, the dark blue pattern in Fig. 1 indicates underconnectivity between the visual cortex, responsible for processing visual information, and the lingual gyrus, involved in vision and word processing.

Graph group analysis is related to *graph classification* (e.g., Lee, Rossi, and Kong 2018), but we are interested not only in what is *different* but also in what is *similar* among our graph groups. Our task further shares some of its motivation with *significant subgraph mining* (e.g., Sugiyama et al. 2015), *graph summarization* (e.g., Liu et al. 2018), and *data clustering* with graphs as data points (e.g., Mukherjee, Sarkar, and Lin 2017). However, we focus on a *complete* characterization of a *set* of graphs under a *given* partition—a cornerstone of scientific discovery involving graph data.

The remainder of the paper is structured as follows. After settling our basic notation, we describe the theoretical foundations of our method and introduce our algorithm. Having covered related work, through experiments on synthetic and real-world data, we demonstrate that GRAGRA works well in practice, before rounding up with discussion and conclusions. All our data, code, and results are publicly available.[1]

Preliminaries

We consider a set $\mathcal{G} = \{G_1, \ldots, G_{|\mathcal{G}|}\}$ of $|\mathcal{G}|$ *node-aligned* graphs $G_i = (V, E_i)$ with $n = |V|$ nodes and $m_i = |E_i|$ edges, omitting the subscripts when clear from context. A partition $\Pi = \{\mathcal{G}_1, \ldots, \mathcal{G}_k\}$ is a set of k non-empty subsets of $\mathcal{G}_i \subseteq \mathcal{G}$, called *graph groups*, of cardinalities $c_i = |\mathcal{G}_i|$, whose disjoint union is $\mathcal{G}$. Our graphs can be undirected or directed, loopy or non-loopy, and unweighted, edge-labeled, or integer weighted, where for the purposes of our model, we treat distinct edge labels or edge weights as a set W of categories, and edges $e \in E_i$ are drawn from the set $\mathcal{E} = V \times V \times W$ of all possible weighted edges.

The *empirical frequency* of edge set $X \subseteq \mathcal{E}$ in group $\mathcal{G}_i$ is $q_i(X) = |\{(V, E) \in \mathcal{G}_i \mid X \subseteq E\}| / c_i$, and we denote by V_X the set of nodes incident with at least one edge in X.

We base our probabilistic model on the maximum entropy principle, by which the distribution that best reflects a given set of constraints without introducing additional assumptions is the distribution with maximum Shannon entropy (Jaynes 1982). Thus, the expected frequency of X in $\mathcal{G}_i$ under a given set of edge sets $S \subseteq 2^{\mathcal{E}}$ is

$$p_i(X \mid S) = \mathbb{E}_f[X] = \sum_{Y \in 2^{\mathcal{E}}, \, X \subseteq Y} f(Y \mid S) \,,$$

where $2^{\mathcal{E}}$ is the power set of $\mathcal{E}$, and f is the distribution maximizing $\arg\max_f \{-\sum f_X \log f_X\}$, subject to linear constraints $\mathbb{E}_f[X] = q_i(X)$ for all elements in S (Csiszár 1975). That is,

$$f(X \mid S) = \theta_0 \prod_{Y_i \in S, \, Y_i \subseteq X} \theta_i \,,$$

where θ_0 and all θ_i are real-valued model parameters. Finding the distribution f is a convex problem that involves computing the expected frequency $p_i(X \mid S)$ over exponentially

many elements. This is intractable if done naïvely, but there exist practical approaches that factorize p_i into a product of independent distributions (Mampaey, Vreeken, and Tatti 2012; Dalleiger and Vreeken 2020a,b).

Theory

We now lay the theoretical foundations of our method, introducing our probabilistic model, our objective function, and our statistical test. At a high level, our goal in graph group analysis is to discover a set S of *graph patterns*, i.e., edge sets of *connected subgraphs*, and an *association matrix* A assigning graph patterns to graph groups, such that S and A together reveal the similarities and differences between graphs in the same group, between graphs in different groups, and between the relationships connecting the groups. A pattern is *specific* if we assign it to only one graph group, and it is *shared* if we assign it to several graph groups. We choose which patterns to include in our model based on the information we gain from them, testing whether this gain is statistically significant to rule out spurious results.

To avoid redundancy, we assign a pattern X to a group $\mathcal{G}_i$ iff X is *informative* for $\mathcal{G}_i$, given what we already know about all groups. More precisely, using X as a column index of A in a slight abuse of notation, we set $A_{iX} = 1$ iff X is *informative* for $\mathcal{G}_i$ under our current model (S, A). We assess this by comparing the *empirical frequency* of X in group $\mathcal{G}_i$, $q_i(X)$, to its *expected frequency* in that group under our current model, $p_i(X \mid S_i)$, where $S_i = \{X \in S \mid A_{iX} = 1\}$, and p_i is obtained from a practical approximation of the maximum entropy distribution with constraint set S_i. X is *informative* for $\mathcal{G}_i$ iff $q_i(X)$ is significantly different from $p_i(X \mid S_i)$, as judged by a statistical test, and we add X to S (and column X to A) if X is informative for *some* $\mathcal{G}_i \in \Pi$.

To identify a suitable set of graph patterns S and an adequate association matrix A, we exploit the interplay between two steps. First, we discover the best pattern X to add to S, given the current (S, A), and second, we identify the best assignment of X to graph groups to update A, given the current (S, A) *and* a new pattern X. We now describe each step in more detail.

Identifying Informative Graph Patterns

To measure the likelihood of a set $S \subseteq 2^{\mathcal{E}}$ of graph patterns, we use the *Bayesian Information Criterion* (BIC),

$$\mathrm{BIC}(S) = \ell(S) + (k \cdot |S|)/2 \log |\mathcal{G}| \quad \text{(Schwarz 1978)},$$

where $k \cdot |S|$ is the number of coefficients in our model, and

$$\ell(S) = \sum_i \ell_i(S) = -\sum_i \sum_{G \in \mathcal{G}_i} \log p_i(G \mid S_i) \,,$$

is the log likelihood of S (with $S_i \subseteq S$ derived from A), assuming that the graphs in a group are independent and identically distributed. This allows us to identify a good set of graph patterns by minimizing the BIC score, i.e.,

$$\arg\min_{S \subseteq 2^{\mathcal{E}}} \{\mathrm{BIC}(S)\} \,.$$

Solving this problem exactly poses significant challenges in practice due to its combinatorial nature and the explosion

in the number of solution candidates. Therefore, we employ a greedy search strategy, iteratively selecting the graph pattern $X \subseteq \mathcal{E}$ that best improves our current model. That is, for a given (S, A), we select the graph pattern X that maximizes our likelihood, or equivalently, maximizes the difference $\mathrm{BIC}(S) - \mathrm{BIC}(S \cup \{X\})$, which we write as

$$\Delta(X) = \ell(S) - \ell(S \cup \{X\}) - k/2 \log |\mathcal{G}| .$$

In a nutshell, the core of our approach is the procedure

$$S \leftarrow S \cup \left\{ \underset{X \subseteq \mathcal{E},\, \Delta(X) > 0}{\arg\max} \{\Delta(X)\} \right\} , \qquad (1)$$

by which we iteratively and greedily insert into S the pattern $X \subseteq \mathcal{E}$ that locally maximizes our information gain.

Using a model selection criterion alone, however, we cannot tell if our information gain is due to random fluctuations or due to signal, especially if we only have a limited number of samples. Thus, to avoid modeling noise, we add X to S only if its information gain $\Delta(X)$ is *statistically significant*. Therefore, we test whether we can reject the null hypothesis

$$H_0 : \ \mathrm{BIC}(S) = \mathrm{BIC}(S \cup \{X\}) .$$

To this end, we use Vuong's closeness test (Vuong 1989), a likelihood ratio test designed for model selection problems under BIC. Vuong's test statistic is defined as $2\Delta(X)$, which is asymptotically χ^2-distributed with $\mathrm{df}_\Delta(X) = \mathrm{df}\, p_i(\,\cdot\mid S \cup \{X\}) - \mathrm{df}\, p_i(\,\cdot\mid S)$ degrees of freedom. To calculate $\mathrm{df}_\Delta(X)$, we count the coefficients θ that must be changed in every distribution if we insert X into S. As we add one coefficient for X, and update at least $|X|$ edge coefficients per group, we arrive at $|X| + 1$ additional degrees of freedom.

Discovering Differential Pattern Associations

Once we have selected a new pattern $X \subseteq \mathcal{E}$ to add to S, given the current S and A, we identify a good assignment of X to graph groups $\mathcal{G}_i \in \Pi$ to update A. Here, the significance of $\Delta(X)$, which is used to accept X into S, signals that X is informative for *some* $\mathcal{G}_i \in \Pi$, but it does not tell us for *which* $\mathcal{G}_i$. To assign X to a graph group $\mathcal{G}_i$, we hence rely on the *partial* information gain of X for $\mathcal{G}_i$,

$$\Delta_i(X) = \ell_i(S_i) - \ell_i(S_i \cup \{X\}) - k/2 \log |\mathcal{G}| .$$

Again, we use Vuong's closeness test to decide whether $\Delta_i(X)$ is significant; and if it is, we set $A_{iX} = 1$.

Algorithm

Having stated its theoretical foundations, we now introduce GRAGRA as an algorithm to differentially describe groups of graphs using sets of significant subgraphs. GRAGRA, whose pseudocode is given as Alg. 1, revolves around the procedure stated in Eq. (1), a greedy process that iteratively selects the graph pattern candidate that best enhances our model. This *could* involve myriad searches through the exponentially-sized space of all possible graphs with nodes from V, which is not only computationally infeasible in most cases but also unnecessary, as most candidates will be eliminated as uninformative by Vuong's test. Hence, rather than exhaustively searching for the best graph patterns, we propose to grow graphs by systematically adding edges to candidates.

Algorithm 1: GRAGRA

Input: groups of graphs $\mathcal{G}_1, \ldots, \mathcal{G}_k$
Output: set of graph patterns S, association matrix A
1   $S \leftarrow \mathcal{E}$
2   $A \leftarrow$ empty binary matrix with k rows and 0 columns
3   $C \leftarrow \{\{x, y\} \mid x, y \in \mathcal{E}, x \neq y, V_{\{x\}} \cap V_{\{y\}} \neq \emptyset\}$
4   **while** $C \neq \emptyset$
5     $\hat{X}, C \leftarrow \text{GROW}(C)$
6   **if** $\exists i \in [k]$ s.t. $h_i(\hat{X})$ is significant
7   resize A
8       $A_{i\hat{X}} = 1 \iff h_i(\hat{X})$ is significant $\forall i \in [k]$
9       $S \leftarrow S \cup \{\hat{X}\}$
10   estimate $p_i(\,\cdot\mid S_i) \ \forall i \in [k]$ s.t. $A_{i\hat{X}} = 1$
11   **return** $S \setminus \mathcal{E}, A$
12
13   **Fn.** $\text{GROW}(C)$
14     $X \leftarrow \underset{X \in C}{\arg\max} \{h(X) \text{ s.t. } h(X) \text{ is significant}\}$
15     $C \leftarrow C \cup (((V_X \times V \times W) \cup (V \times V_X \times W)) \setminus X)$
16     $C \leftarrow \{X \in C \mid h(X) \text{ is significant }\}$
17     $\hat{X} \leftarrow \underset{X \in C}{\arg\max} \{h(X)\}$
18   **if** $h(\hat{X}) > h(X)$
19   **return** $\text{GROW}(C)$
20   **else**
21   **return** $\hat{X}, C \setminus \{\hat{X}\}$

To enable our model to infer all possible graphs, we initialize it with the set $\mathcal{E}$ of all possible edges. As our initial graph to grow, we then select the most promising graph pattern from our initial candidates, i.e., the connected triples

$$C = \{\{x, y\} \mid x, y \in \mathcal{E}, x \neq y, V_{\{x\}} \cap V_{\{y\}} \neq \emptyset\} .$$

Starting with a graph pattern X, we explore all its expansions, $((V_X \times V \times W) \cup (V \times V_X \times W)) \setminus X$, from which we select the best candidate pattern to grow further, as long as we gain information and $\Delta(X)$ is significant. We summarize these steps in the function GROW of Alg. 1 (l. 13–21).

GROW requires many inferences of Δ, which involve inferring many more expected frequencies p_i, rendering exact computation impractical. We thus design a practical, pessimistic heuristic that only considers the information gain from graphs $G \in \mathcal{G}$ in which X is fully present:

$$h(X) = \sum_i c_i \cdot q_i(X) \log \frac{q_i(X)}{p_i(X)} - k/2 \log |\mathcal{G}| . \qquad (2)$$

We use $h(X)$ instead of $\Delta(X)$ because it involves inferring only *one* expected frequency per graph group. A derivation of this heuristic can be found in the Online Appendix.

To summarize, GRAGRA proceeds as follows. Starting with an initial set of candidates C (l. 3), we select (l. 14) and grow (l. 15) the best candidate, and retain all significant expansions (l. 16), until we have grown X to its fullest potential (l. 18–19). Afterwards, we test if the information gain provided by X is significant, and if so, we keep track of its graph group associations (l. 8), and insert X into S (l. 9).

The computational complexity of GRAGRA depends on the number of candidates, which can grow to at most $|2^{\mathcal{E}}|$. In practice, GRAGRA's complexity depends on the number of times we grow graph patterns, which is data-dependent and bounded by the size γ of the largest connected component observed in an input graph, as growing beyond that reduces the information gain. Multiplying γ by the initial set of candidates, GRAGRA achieves a complexity of $\mathcal{O}\left(\binom{n}{3}|W|\gamma\right)$ for all practial purposes, where we assume that the complexity of inferring the expected frequency is bounded.

Related Work

To the best of our knowledge, we are the first to differentially describe groups of graphs through sets of significant subgraphs. Our method is inspired by advances in graph similarity description (MOMO, Coupette and Vreeken 2021) and explainable pattern set mining using maximum-entropy modeling (DISC, Dalleiger and Vreeken 2020a,b). However, MOMO focuses on pairs and unpartitioned sets of graphs; DISC is designed for itemset data, ignores graph structure, and does not scale on graphs; and neither method uses a statistical test to select patterns. Further related work broadly falls into two categories: statistical inference on network populations, and graph mining for groups of graphs.

Statistical Inference on Network Populations. In the statistics literature, the task of analyzing multiple graphs simultaneously is typically framed as an inference problem for network-valued random variables (Durante, Dunson, and Vogelstein 2017; Lovato et al. 2020; Lunagómez, Olhede, and Wolfe 2020). Here, Ghoshdastidar et al. (2020) establish limits for distinguishing two population distributions given small sample sizes, and Lunagómez, Olhede, and Wolfe (2020) propose notions of mean and dispersion for a single population of networks, where the population mean is itself a network. Maugis et al. (2020) use subgraph counts to test if all graphs in a sample are drawn from the same distribution, and Signorelli and Wit (2020) propose a model-based clustering approach to describe subpopulations within a population of networks. Finally, Durante, Dunson, and Vogelstein (2017) extend latent space approaches designed for single graphs to capture the probabilistic mechanism that generates multiple graphs from a single population distribution. Their model has been used to characterize and test for differences between groups of brain networks (Durante, Dunson et al. 2018)—an actively studied application for which numerous statistical methods, mostly focusing on *testing* for differences, have been developed (Ginestet et al. 2017; Wang et al. 2019; Lukemire et al. 2020; Kundu et al. 2021; Lovato et al. 2021; Lehmann et al. 2021).

Prior work in the statistics literature has focused on describing *one* network population or distinguishing *two* populations. In contrast, with GRAGRA, we aim to construct a *differential description* of *any* number of populations. Furthermore, we ask not only *if* these populations are different, but also *how* they are different and how they are *similar*.

Graph Mining for Groups of Graphs. In the graph mining literature, groups of graphs have been studied in contexts as diverse as significant subgraph mining (Llinares-López

et al. 2015; Sugiyama et al. 2015), graph classification (Vogelstein et al. 2013; Yan et al. 2019; Lanciano, Bonchi, and Gionis 2020), graph clustering with graphs as data points (Mukherjee, Sarkar, and Lin 2017), anomalous graph detection (Gomes, Rao, and Neville 2018), and graph summarization for time series of graphs (Shah et al. 2015). Significant subgraph mining commonly considers small, node-labeled graphs with unaligned node sets, and hence, does not target our problem. However, our setup—i.e., medium-sized graphs with aligned node sets—, has received heightened attention in the graph classification community, again inspired by challenges from neuroscience (Vogelstein et al. 2013; Yan et al. 2019; Lanciano, Bonchi, and Gionis 2020).

The methods that are closest to our work are *contrast subgraphs* (Lanciano, Bonchi, and Gionis 2020) and *signal subgraphs* (Vogelstein et al. 2013), both designed for two groups of node-aligned graphs. *Contrast subgraphs* discover the densest subgraph in the difference of the summary graphs of the input groups (obtained by adding the graphs in each group separately and then subtracting the results), where the size of this subgraph depends on a user-specified regularization parameter α. *Signal subgraphs* assume edge independence as a prior to rank edges by the p-values of an edge-wise statistical test for distributional difference (e.g., Fisher's exact test). Like *signal subgraphs*, GRAGRA combines ideas from structural and statistical pattern mining to produce interpretable results that—unlike *contrast subgraphs*—are based on a statistical foundation. GRAGRA is more exploratory and more flexible than both competitors, however, because it treats graph group description as an end in itself and can handle any number of graph groups.

Experiments

We now present an extensive evaluation of our algorithm. To this end, we implement GRAGRA in C++ and expose a Python interface to facilitate experimentation. We run our experiments on Intel E5-2643 CPUs with 128 or 256 GB RAM, testing at a conservative significance level of 1×10^{-7} (or 1×10^{-5} when operating with less than 50 samples), and make all data, code, and results publicly available.[2] Our experiments revolve around two questions:

1. Can GRAGRA reliably recover the ground truth from groups of synthetic graphs?
2. Does GRAGRA discover meaningful patterns in groups of real graphs?

Recovering Ground Truth from Synthetic Graphs

To assess the reliability of GRAGRA, we run it on groups of synthetic graphs with planted patterns. We consider three scenarios, namely,

1. summarizing *one* group of graphs,
2. differentially describing *two* groups of graphs, and
3. differentially describing *four* groups of graphs.

In all three scenarios, each graph group consists of 100 graphs with 100 nodes, and our configurations differ in their
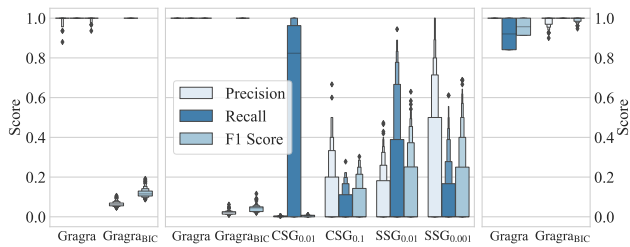
---

[2]https://doi.org/10.5281/zenodo.5758560

Figure 2: GRAGRA reliably recovers the ground truth from synthetic data. We show precision, recall, and F1 score distributions for GRAGRA, GRAGRA$_{BIC}$, *contrast subgraphs* (CSG), and *signal subgraphs* (SSG), separately for all experiments in our three different settings: one-group setting (left), two-group setting (middle), and four-group setting (right). Subscripts of CSG labels correspond to different choices of their regularization parameter α, and subscripts of SSG labels indicate different requirements for the p-values obtained from their edge-wise distributional difference test.

planted patterns (type, prevalence, and position) and noise levels. A detailed overview of our synthetic data configurations can be found in the Online Appendix.

For each scenario, we report the distribution of precision, recall, and F1 score, computed separately for each group of graphs, for the edges of the planted patterns across 100 graph group datasets sampled with different seeds. In all scenarios, we compare GRAGRA, which uses BIC with Vuong's closeness test for pattern selection, with a variant using only BIC and no statistical test to select patterns (GRAGRA$_{BIC}$). For configurations in the second scenario, we also compare our results with those from *contrast subgraphs* (CSG) and *signal subgraphs* (SSG), described in the previous section.

As shown in Fig. 2, GRAGRA$_{BIC}$ delivers good results in the four-group scenario but generally has poor precision, treating noise as signal. CSG and SSG identify only constrastive patterns, and fail even for contrastive patterns if the individual edges in planted patterns have similar occurrence probabilities across groups. GRAGRA, however, reliably recovers the ground truth across scenarios and configurations, which allows us to hope that it will also work in practice.

Discovering Meaningful Patterns in Real Graphs

To determine whether GRAGRA discovers meaningful patterns in groups of real graphs, we run 29 experiments on graph group data of various graph types from three domains: functional brain networks (undirected, unweighted), air transportation networks (directed, weighted), and international trade networks (directed, weighted). We compile basic statistics of these networks in the Technical Appendix, and present a quantitative overview of our results in Fig. 3. We observe that, in line with expectations derived from theory, *more graphs* or *graphs with more potential edges*, partitioned into *fewer groups*, generally yield *more patterns*.

Functional Brain Networks *Network neuroscience* has emerged as a promising approach to understanding neurological disorders and diseases (Bullmore and Sporns 2009;

Bassett and Sporns 2017; Fornito, Zalesky, and Breakspear 2015). One of its fundamental questions is whether certain disorders are systematically associated with structural or functional connectivity alterations in the brain (van den Heuvel and Sporns 2019). In particular, there is considerable uncertainty surrounding the neurological footprint of autism (and the delineation of its subtypes), and small sample sizes as well as covariates make many published findings hard to replicate (He, Byrge, and Kennedy 2020; King et al. 2019). This calls for methods that can detect signal in the presence of considerable noise and heterogeneity, identifying connectivity patterns that are statistically significantly associated with one or more groups of brain networks.

Motivated by this application, we obtain graphs from preprocessed functional connectomes provided by the Autism Brain Imaging Data Exchange (Craddock et al. 2013). In these graphs, each node corresponds to one of the 116 *regions of interest* from the automated anatomical labeling atlas (AAL, Rolls et al. 2020), and each edge indicates relatively strong connectivity between two regions, as measured by their blood-oxygen-level dependent signal correlation during resting-state functional magnetic resonance imaging. To facilitate comparisons, the data is processed and grouped as described by Lanciano, Bonchi, and Gionis (2020), but we remove the self-loops (corresponding to perfect self-correlations) that are present in their data.

We experiment with GRAGRA in four two-group settings (individuals with autism spectrum disorder [ASD] and typically developed controls [TD] in the categories *adolescents*, *children*, *eyes closed during scan*, and *males*), four one-group settings (autistic individuals in each category only), and one four-group setting (autistic and non-autistic children and adolescents), operating on graphs with $m \in [1\,320, 1\,348]$ edges and graph groups $\mathcal{G}_i$ with $c_i \in [49, 420]$ graphs. Our four-group experiment identifies significant overconnectivity across multiple brain regions as characteristic of ASD children versus all other groups, paralleling the neuroscience literature (Supekar et al. 2013; Nomi and Uddin 2015). However, as shown in Fig. 4, most of the patterns we identify in the two-group setting yield similar information gains across both groups (left), and there is significant structure to be exploited even *within* individual groups (right). This indicates that the differences between autistic and non-autistic brains in the settings under study are rather subtle, and that there is considerable heterogeneity also in the one-group data. To explore this heterogeneity and delineate neurosubtypes of autism (cf. Hong et al. 2020), our results could be used as inputs to multivariate subgroup discovery or clustering algorithms, where GRAGRA would effectively serve as a dimensionality reduction technique.

Air Transportation Networks We obtain data on passenger flows between domestic airports in the United States for each month over the sixteen years from January 2005 to December 2020 from the website of the Bureau of Transportation Statistics (Bureau of Transportation Statistics 2021). Restricting our analysis to United States mainland airports and carriers classified as national (100 million to 1 billion USD revenue in the previous year) or major (over 1 billion
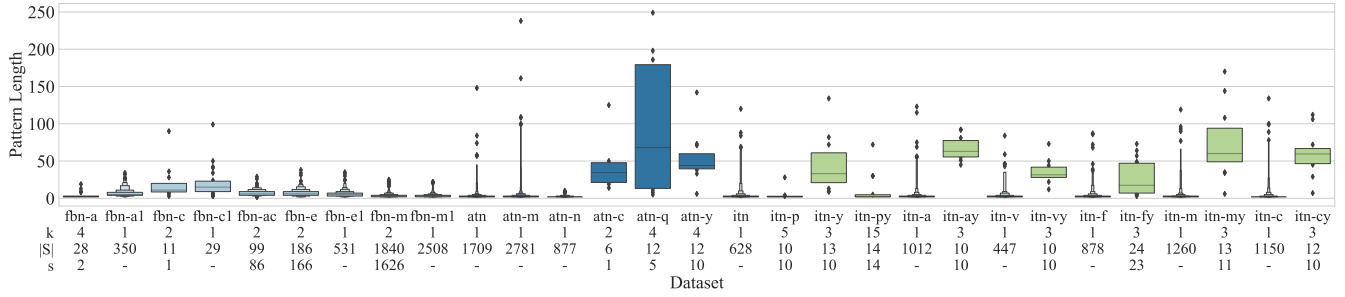
Figure 3: GRAGRA discovers long graph patterns in datasets with different numbers of graph groups. Here, we show the length distribution of the patterns identified in each of our experiments on real-world data, where each box corresponds to a dataset. The first number below a dataset identifier states the number of graph groups k in the dataset, the second number states the total number of patterns $|S|$, and the third number states the number of patterns s shared between at least two graph groups.
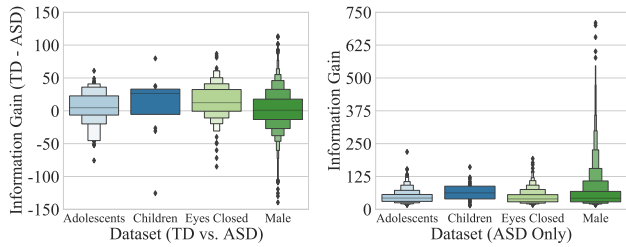

Figure 4: GRAGRA unveils shared and contrastive patterns in noisy and heterogeneous data. Here, we display the distribution of information gain differences per pattern in the two-group setting (left), and the distribution of information gains per pattern in the one-group setting (right), for our experiments on functional brain networks.

USD revenue in the previous year), we create one air transportation network per year, month, and carrier class. To this end, for each year and month, we aggregate the passenger flows between two airports by carrier class and filter edges corresponding to fewer than $3\,000$ passengers, which leaves edges between $n = 300$ airports (identified by three-letter IATA codes). Excluding graphs with fewer than $n-1 = 299$ edges, we arrive at 374 graphs, whose edges we discretize into ten weight categories using equal-width binning.

We are interested in discovering patterns that are shared across all graphs, identifying structures of connected routes that are specific to individual carrier classes, and unveiling both seasonal and temporal trends. Therefore, we run GRAGRA in six different settings: on all graphs as one group, on the graphs corresponding to each carrier class separately, on all graphs with carrier classes as groups, on all graphs with quarters as groups (starting from December to capture the winter holiday season), and on all graphs with consecutive four-year intervals as groups. Thus, our setup contains graphs with $m \in [335, 3\,533]$ edges and graph groups $\mathcal{G}_i$ with $c_i \in [86, 374]$ graphs. In Fig. 5, we depict a subset of our results from the experiments involving the distinction between carrier classes. GRAGRA reveals an air transportation backbone jointly serviced by both carrier classes

(middle), and it uncovers routes that are characteristically served by national or major carriers (left and right). Overall, we find that patterns corresponding to national carrier routes are often smaller and cover shorter distances than those corresponding to major carrier routes, mirroring the relatively smaller role of national carriers in the air traffic market.

International Trade Networks We obtain data on international trade flows from the website of the World Integrated Trade Solution (World Bank 2021), for the thirty years from 1989 to 2018 (inclusive). The raw data correspond to exports of goods between (mostly) countries, classified using the Harmonized System at the four-digit level (HS-4), whose trade values we aggregate per (source, destination, HS-4 code) triple. For each year and HS-4 code, we construct one directed, weighted graph with (roughly) countries as nodes and exports as edges, discretizing the edge weights into ten categories using equal-width binning. We eliminate all trade entities above the country level but retain trade entities below the country level (and countries that do not exist anymore) if they have an ISO3 code. Restricting our attention to the WITS product groups *Animals*, *Vegetables*, *Food Products*, *Minerals*, and *Chemicals*, we arrive at $3\,976$ graphs with $n = 250$ nodes and at least $n - 1 = 249$ edges.

Leveraging the richness of our data, we ask not only what graph patterns are characteristic of international trade as a whole, but also what structures emerge when we group trade networks by product class, ten-year interval, or product class *and* ten-year interval. As GRAGRA allows us to inspect our data at different scales, we further investigate the trade patterns it unveils when considering each product class separately, either treating all graphs from one product class as one group or splitting them by ten-year interval. Thus, we run our experiments on graphs with $m \in [256, 11\,415]$ edges and graph groups $\mathcal{G}_i$ with $c_i \in [70, 3\,976]$ graphs. In Fig. 6, we illustrate five patterns discovered in the experiments that explore all graphs together, grouped by product class and ten-year interval. Although the input consists of fifteen classes, GRAGRA discovers not only meaningful patterns but meaningful patterns *with meaningful assignments* to graph groups which, as highlighted by the pattern labels in Fig. 6, can be summarized succinctly. Across all experi-
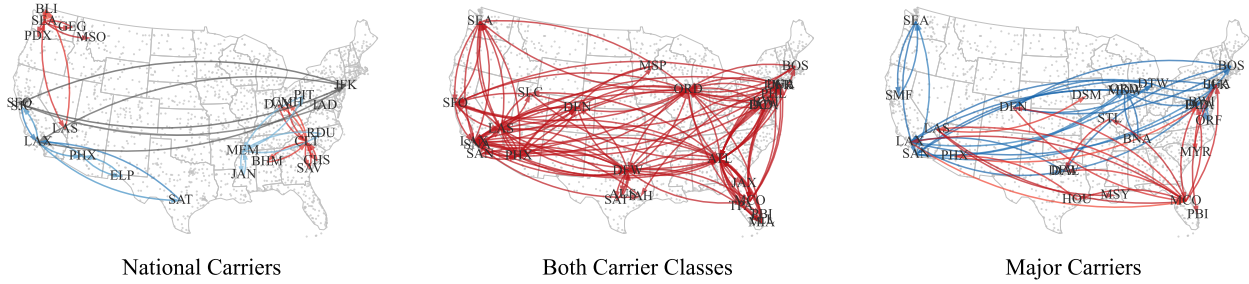
Figure 5: GRAGRA discovers large, meaningful graph patterns. Here, we depict some of the patterns discovered in the air transportation networks of national carriers (left, five patterns shown), major carriers (right, two patterns shown), and both carrier classes (middle, one pattern shown). Gray nodes represent airports, and node labels identify airports contained in at least one of the displayed patterns by their three-letter IATA codes. Directed edges represent flight segments, and edge colors are proportional to their weight bins, following different color maps (reds, blues, or grays) where necessary to make them visually distinguishable. All drawn patterns are among the top fifteen in terms of information gain for their respective experiment, and the pattern in the middle is the top shared pattern, corresponding to the United States air transportation backbone.
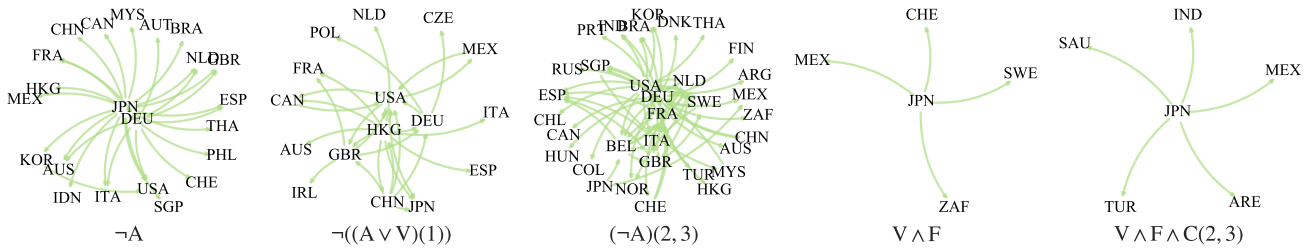

Figure 6: GRAGRA mines differential descriptions even when many graph groups are given as input. Here, we show the top five graph patterns identified in the international trade networks when split by product class and decade (fifteen graph groups in total). Nodes correspond to countries, which are represented by their ISO3 country codes. Directed edges correspond to trade flows between the countries, where the edge weights in all displayed patterns fall into the top weight bin. The patterns are labeled by rules identifying the graph groups in which they occur, with letters corresponding to the first letter of a product group, and numbers corresponding to the position of a ten-year interval. For example, the third pattern, labeled $(\neg A)(2, 3)$, occurs in all product classes except for *Animals*, in the second and the third ten-year interval, i.e., in $[99, 19]$.

ments, we observe that the patterns yielding the largest information gains are often composed entirely of edges in the top two weight bins. This suggests that the ranking of exporter-importer pairs is most stable on the upper end of the trade-value spectrum, which aligns with interdisciplinary research findings that international trade is highly stratified (Sacks, Ventresca, and Uzzi 2001; Lloyd, Mahutga, and De Leeuw 2009; Fagiolo, Reyes, and Schiavo 2010).

Conclusion

We study the *graph group analysis* problem: Given a set of graphs and a partition of this set into *graph groups*, succinctly summarize the commonalities and differences between graphs in the same group, between graphs in different groups, and between the relationships connecting the groups. We introduce GRAGRA as an algorithm to solve the problem, which uses maximum likelihood modeling, paired with a model selection criterion and a statistical test, to jointly discover a set of significant subgraphs, called

graph patterns, and an assignment of these patterns to graph groups. In our experiments, we demonstrate that GRAGRA differentially describes synthetic and real-world graph groups, even when faced with heterogeneity, noise, or large group numbers. As a byproduct, we introduce two novel datasets of node-aligned graphs, which might be of independent interest to the graph mining community.

However, our work also has limitations. First of all, we model edge weights as categories, which works well for binned edge weights in practice but is theoretically dissatisfying. Therefore, a natural enhancement of GRAGRA would be able to handle real edge weights, possibly using a maximum entropy model on its edge weight distribution. Second, we currently test all our graph patterns at the same alpha level. While this is theoretically defensible, given that we combine our statistical test with a model selection criterion, dynamically adjusting our alpha level might be an option worth exploring. Finally, GRAGRA is currently limited to groups of node-aligned graphs, and extending it to other graph types constitutes an open opportunity for future work.

References

Bassett, D. S.; and Sporns, O. 2017. Network neuroscience. *Nature Neuroscience*, 20(3): 353–364.

Bullmore, E.; and Sporns, O. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3): 186–198.

Bureau of Transportation Statistics. 2021. Data Bank 28DS - T-100 Domestic Segment Data (World Area Code). https://www.bts.gov/browse-statistical-products-and-data/bts-publications/data-bank-28ds-t-100-domestic-segment-data.

Coupette, C.; and Vreeken, J. 2021. Graph Similarity Description: How Are These Graphs Similar? In *KDD*, 185–195.

Craddock, C.; Benhajali, Y.; Chu, C.; Chouinard, F.; Evans, A.; Jakab, A.; Khundrakpam, B. S.; Lewis, J. D.; Li, Q.; Milham, M.; et al. 2013. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7.

Csiszár, I. 1975. I-divergence geometry of probability distributions and minimization problems. *Annals Prob.*, 146–158.

Dalleiger, S.; and Vreeken, J. 2020a. Explainable data decompositions. In *AAAI*, volume 34, 3709–3716.

Dalleiger, S.; and Vreeken, J. 2020b. The Relaxed Maximum Entropy Distribution and its Application to Pattern Discovery. In *ICDM*, 978–983.

Durante, D.; Dunson, D. B.; and Vogelstein, J. T. 2017. Nonparametric Bayes modeling of populations of networks. *JASA*, 112(520): 1516–1530.

Durante, D.; Dunson, D. B.; et al. 2018. Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis*, 13(1): 29–58.

Fagiolo, G.; Reyes, J.; and Schiavo, S. 2010. The evolution of the world trade web: a weighted-network analysis. *Journal of Evolutionary Economics*, 20(4): 479–514.

Fornito, A.; Zalesky, A.; and Breakspear, M. 2015. The connectomics of brain disorders. *Nature Reviews Neuroscience*, 16(3): 159–172.

Ghoshdastidar, D.; Gutzeit, M.; Carpentier, A.; and Von Luxburg, U. 2020. Two-sample hypothesis testing for inhomogeneous random graphs. *Annals Stat.*, 48(4): 2208–2229.

Ginestet, C. E.; Li, J.; Balachandran, P.; Rosenberg, S.; Kolaczyk, E. D.; et al. 2017. Hypothesis testing for network data in functional neuroimaging. *Annals Appl. Stat.*, 11(2): 725–750.

Gomes, G.; Rao, V.; and Neville, J. 2018. Multi-level Hypothesis Testing for Populations of Heterogeneous Networks. In *ICDM*, 977–982.

He, Y.; Byrge, L.; and Kennedy, D. P. 2020. Nonreplication of functional connectivity differences in autism spectrum disorder across multiple sites and denoising strategies. *Human Brain Mapping*, 41(5): 1334–1350.

Hong, S.-J.; Vogelstein, J. T.; Gozzi, A.; Bernhardt, B. C.; Yeo, B. T.; Milham, M. P.; and Di Martino, A. 2020. Toward neurosubtypes in autism. *Biological Psychiatry*, 88(1): 111–128.

Hull, J. V.; Dokovna, L. B.; Jacokes, Z. J.; Torgerson, C. M.; Irimia, A.; and Van Horn, J. D. 2017. Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in Psychiatry*, 7: 205.

Jaynes, E. 1982. On the rationale of maximum-entropy methods. *Proc. IEEE*, 70(9): 939–952.

King, J. B.; Prigge, M. B.; King, C. K.; Morgan, J.; Weathersby, F.; Fox, J. C.; Dean, D. C.; Freeman, A.; Villaruz, J. A. M.; Kane, K. L.; et al. 2019. Generalizability and reproducibility of functional connectivity in autism. *Molecular Autism*, 10(1): 1–23.

Kundu, S.; Ming, J.; Nocera, J.; and McGregor, K. M. 2021. Integrative learning for population of dynamic networks with covariates. *NeuroImage*, 118181.

Lanciano, T.; Bonchi, F.; and Gionis, A. 2020. Explainable Classification of Brain Networks via Contrast Subgraphs. In *KDD*, 3308–3318.

Lee, J. B.; Rossi, R.; and Kong, X. 2018. Graph classification using structural attention. 1666–1674.

Lehmann, B.; Henson, R.; Geerligs, L.; White, S.; et al. 2021. Characterising group-level brain connectivity: a framework using Bayesian exponential random graph models. *NeuroImage*, 225: 117480.

Liu, Y.; Safavi, T.; Dighe, A.; and Koutra, D. 2018. Graph summarization methods and applications: A survey. *ACM CSUR*, 51(3): 1–34.

Llinares-López, F.; Sugiyama, M.; Papaxanthos, L.; and Borgwardt, K. 2015. Fast and memory-efficient significant pattern mining via permutation testing. In *KDD*, 725–734.

Lloyd, P.; Mahutga, M. C.; and De Leeuw, J. 2009. Looking back and forging ahead: Thirty years of social network research on the world-system. *Journal of World-Systems Research*, 48–85.

Lovato, I.; Pini, A.; Stamm, A.; Taquet, M.; and Vantini, S. 2021. Multiscale null hypothesis testing for network-valued data: Analysis of brain networks of patients with autism. *Appl. Statist.*, 70(2): 372–397.

Lovato, I.; Pini, A.; Stamm, A.; and Vantini, S. 2020. Model-free two-sample test for network-valued data. *Computational Statistics and Data Analysis*, 144: 106896.

Lukemire, J.; Kundu, S.; Pagnoni, G.; and Guo, Y. 2020. Bayesian joint modeling of multiple brain functional networks. *JASA*, 1–13.

Lunagómez, S.; Olhede, S. C.; and Wolfe, P. J. 2020. Modeling network populations via graph distances. *JASA*, 1–18.

Mampaey, M.; Vreeken, J.; and Tatti, N. 2012. Summarizing Data Succinctly with the Most Informative Itemsets. *ACM TKDD*, 6: 1–44.

Maugis, P.-A.; Olhede, S.; Priebe, C.; and Wolfe, P. 2020. Testing for equivalence of network distribution using subgraph counts. *Journal of Computational and Graphical Statistics*, 29(3): 455–465.

Mukherjee, S. S.; Sarkar, P.; and Lin, L. 2017. On clustering network-valued data. In *NIPS*, 7074–7084.

Nomi, J. S.; and Uddin, L. Q. 2015. Developmental changes in large-scale network connectivity in autism. *NeuroImage: Clinical*, 7: 732–741.

Rolls, E. T.; Huang, C.-C.; Lin, C.-P.; Feng, J.; and Joliot, M. 2020. Automated anatomical labelling atlas 3. *Neuroimage*, 206: 116189.

Sacks, M. A.; Ventresca, M. J.; and Uzzi, B. 2001. Global institutions and networks: Contingent change in the structure of world trade advantage, 1965-1980. *American Behavioral Scientist*, 44(10): 1579–1601.

Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals Stat.*, 6(2): 461–464.

Shah, N.; Koutra, D.; Zou, T.; Gallagher, B.; and Faloutsos, C. 2015. Timecrunch: Interpretable dynamic graph summarization. In *KDD*, 1055–1064.

Signorelli, M.; and Wit, E. C. 2020. Model-based clustering for populations of networks. *Statistical Modelling*, 20(1): 9–29.

Sugiyama, M.; López, F. L.; Kasenburg, N.; and Borgwardt, K. M. 2015. Significant subgraph mining with multiple testing correction. In *SDM*, 37–45. SIAM.

Supekar, K.; Uddin, L. Q.; Khouzam, A.; Phillips, J.; Gaillard, W. D.; Kenworthy, L. E.; Yerys, B. E.; Vaidya, C. J.; and Menon, V. 2013. Brain hyperconnectivity in children with autism and its links to social deficits. *Cell Reports*, 5(3): 738–747.

van den Heuvel, M. P.; and Sporns, O. 2019. A cross-disorder connectome landscape of brain dysconnectivity. *Nature Reviews Neuroscience*, 20(7): 435–446.

Vogelstein, J. T.; Roncal, W. G.; Vogelstein, R. J.; and Priebe, C. E. 2013. Graph classification using signal-subgraphs: Applications in statistical connectomics. *IEEE TPAMI*, 35(7): 1539–1551.

Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307–333.

Wang, L.; Zhang, Z.; Dunson, D.; et al. 2019. Common and individual structure of brain networks. *Annals Appl. Stat.*, 13(1): 85–112.

World Bank. 2021. World Integrated Trade Solution. https://wits.worldbank.org/Default.aspx?lang=en.

Yan, Y.; Zhu, J.; Duda, M.; Solarz, E.; Sripada, C.; and Koutra, D. 2019. Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data. In *KDD*, 772–782.