View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Caltech Authors - Main

arXiv:1904.08934v1 [cs.DS] 18 Apr 2019

Convex Graph Invariant Relaxations for Graph Edit Distance

Utkan Onur Candogan[†] and Venkat Chandrasekaran[†,‡] *

[†] Department of Electrical Engineering
[‡] Department of Computing and Mathematical Sciences
California Institute of Technology
Pasadena, CA 91125

April 22, 2019

Abstract

The edit distance between two graphs is a widely used measure of similarity that evaluates the smallest number of vertex and edge deletions/insertions required to transform one graph to another. It is NP-hard to compute in general, and a large number of heuristics have been proposed for approximating this quantity. With few exceptions, these methods generally provide upper bounds on the edit distance between two graphs. In this paper, we propose a new family of computationally tractable convex relaxations for obtaining lower bounds on graph edit distance. These relaxations can be tailored to the structural properties of the particular graphs via *convex graph invariants*. Specific examples that we highlight in this paper include constraints on the graph spectrum as well as (tractable approximations of) the stability number and the maximum-cut values of graphs. We prove under suitable conditions that our relaxations are tight (i.e., exactly compute the graph edit distance) when one of the graphs consists of few eigenvalues. We also validate the utility of our framework on synthetic problems as well as real applications involving molecular structure comparison problems in chemistry.

Keywords: convex optimization; majorization; maximum cut; semidefinite programming; stability number; strongly regular graphs.

1 Introduction

Graphs are widely used to represent the structure underlying a collection of interacting entities. A common computational question arising in many contexts is that of measuring the similarity between two graphs. For example, the unknown functions of biological structures such as proteins, RNAs and genes are often deduced from structures which have similar sequences with known functions [18, 20, 26, 35, 36]. Evaluating graph similarity also plays a central role in various pattern recognition applications [9, 30], specifically in areas such as handwriting recognition [12, 25], fingerprint classification [19, 29] and face recognition [38].

The notion of similarity that is the most commonly considered is the *graph edit distance* [33]. The edit distance $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$ between two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ is the smallest number of operations required to transform $\mathcal{G}_1$ into $\mathcal{G}_2$ by a sequence of edits or changes applied to the vertices and

*Email: utkan@caltech.edu, venkatc@caltech.edu. The authors were supported in part by NSF grants CCF-1350590 and CCF-1637598, by AFOSR grant FA9550-16-1-0210, and by a Sloan research fellowship.

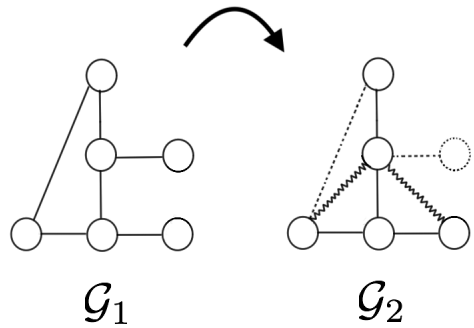


Figure 1: An instance of a graph edit distance problem in which we wish to calculate minimum number of edit operations required for transforming graph $\mathcal{G}_1$ to graph $\mathcal{G}_2$. The edit operations are encoded by line style: The dashed graph elements are to be removed from $\mathcal{G}_1$ and the zigzagged graph elements are to be added to $\mathcal{G}_1$ for transforming $\mathcal{G}_1$ to $\mathcal{G}_2$. Assuming a cost of 1 for every edit operation, we conclude that the graph edit distance between $\mathcal{G}_1$ and $\mathcal{G}_2$ is 5.

edges of $\mathcal{G}_1$. A particular sequence of edit operations transforming $\mathcal{G}_1$ into $\mathcal{G}_2$ is usually referred to as an edit path. For unlabeled graphs, the permissible set of edit operations are usually insertions/deletions of vertices/edges. For labeled graphs, the set of permissible edit operations can also include vertex/edge relabelings. In some situations, certain types of edits are considered more 'severe' than others and different edits have different costs associated to them; in such cases, the edit distance is the smallest cost over all edit paths that transform one graph to another, where the cost of an edit path is the sum of the costs of the edits that compose the path. See Figure 1 for an illustration of a simple graph edit distance problem.

The problem of computing the graph edit distance is NP-hard in general [13], and in practice exact calculation of the edit distance is only feasible for small-sized graphs. Thus, significant efforts have been directed towards developing computationally tractable heuristics for approximating the edit distance [1, 3, 10, 22, 23, 24, 31, 32, 39] or for exactly computing the edit distance for graphs from specific families such as planar graphs [28]. These methods are largely combinatorial in nature, and most of them aim at identifying an edit path that transforms one graph to the other. Consequently, much of the prior literature on this topic provides techniques that lead to upper bounds on the edit distance between two graphs. In contrast, far fewer approaches have been proposed for obtaining lower bounds on the edit distance. We are aware of three notable examples; in [22] and [32] the authors propose tractable linear programming relaxations of intractable combinatorial optimization formulations of the edit distance based on 0/1 optimization and the quadratic assignment problem, respectively, while in [39] the authors propose a combinatorial algorithm based on a multiset representation of the graph that enables the efficient computation of upper and lower bounds of the graph edit distance.

In this paper, we develop a computationally efficient framework for obtaining lower bounds on the graph edit distance. Our contributions differ qualitatively from the prior literature in two aspects. First, our approach can be tailored to the structural properties of the specific graphs at hand based on the notion of a *convex graph invariant*. These lead to useful lower bounds on the edit distance if one of the graphs is 'suitably structured'. Second, we give a theoretical analysis of conditions on pairs of graphs under which a tractable semidefinite relaxation based on the spectral properties of a graph provably computes the edit distance, i.e., our lower bound is tight.

1.1 Our Framework

Much of the focus of our development and our analysis is on the edit distance between two unlabeled graphs on the same number of vertices, with edge deletions and insertions being the allowed edit operations. We discuss in Section 5.1 an extension of our framework to settings in which the number of vertices in the two graphs may be different and in which vertex deletions and insertions are also allowed. Let $A_1 \in \mathbb{S}^n$ and $A_2 \in \mathbb{S}^n$ represent the adjacency matrices of two unweighted, unlabeled, simple and loopless graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ on n vertices. (Here $\mathbb{S}^n$ denotes the space of $n \times n$ real symmetric matrices.) The following optimization problem gives a combinatorial formulation of the computation of the edit distance between $\mathcal{G}_1$ and $\mathcal{G}_2$:

$$\begin{aligned} \mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2) = \min_{X,E \in \mathbb{S}^n} \quad & \sum_{1 \le i < j \le n} \mathbb{1}_{E_{ij} \neq 0} \\ s.t. \quad & X + E = A_2 \\ & X \in \{\Pi A_1 \Pi^T \ : \ \Pi \text{ is an } n \times n \text{ permutation matrix.}\} \\ & E_{ij} \ \in \ \{-1, 0, 1\} \ \forall \, i, j \in \{1, \ldots, n\}. \end{aligned} \tag{1}$$

The function $\mathbb{1}_{\{\cdot\}}$ denotes the usual indicator function, the decision variable X is an adjacency matrix representing $\mathcal{G}_1$, and the decision variable E specifies the edge deletions and insertions made to $\mathcal{G}_1$ to obtain $\mathcal{G}_2$. One aspect of the problem (1) is that its formulation is not symmetric with respect to $\mathcal{G}_1$ and $\mathcal{G}_2$, although the optimal value remains unchanged if A_1 and A_2 are swapped in the problem description, i.e., $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2) = \mathrm{GED}(\mathcal{G}_2, \mathcal{G}_1)$; we revisit this point in the sequel. Unsurprisingly, solving (1) is intractable in general as calculating the graph edit distance is an NP-hard problem. Our approach in this paper is to obtain tractable convex relaxations of (1). Relaxing the objective to a convex function is a straightforward matter; specifically, as the absolute value function constitutes a lower bound on the indicator function in the range $[-1, 1]$; we replace the objective of (1) with the convex function $\frac{1}{2}\|E\|_{\ell_1}$, where $\|\cdot\|_{\ell_1}$ denotes the (entrywise) ℓ_1 norm that sums the absolute values of the entries of a matrix.

The main remaining source of difficulty with obtaining a convex relaxation of (1) is to identify a tractable convex approximation of a set of the form $\{\Pi A \Pi^T \ : \ \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ for a given matrix $A \in \mathbb{S}^n$. When A specifies an adjacency matrix of a graph, this set consists of all the adjacency matrices that describe the graph, thus highlighting the *structural* attributes of the graph that remain invariant to vertex relabeling. Consequently, we seek a convex approximation that similarly remains invariant to vertex relabeling. We describe next a notion from [6] that aims to address this challenge:

Definition 1. *[6] A set $\mathcal{C} \subset \mathbb{S}^n$ is an* invariant convex set *if it is convex and if $M \in \mathcal{C}$ implies that $\Pi M \Pi^T \in \mathcal{C}$ for all $n \times n$ permutation matrices Π.*

Invariant convex sets provide a useful convex modeling framework to constrain graph properties that remain invariant to vertex relabeling. In particular, suppose that $\mathcal{C}_{\mathcal{G}_1} \subset \mathbb{S}^n$ is an invariant convex set that contains $\{\Pi A_1 \Pi^T \ : \ \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ and has an efficient description. Then, the following *convex program* provides a lower bound on $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$:

$$\begin{aligned} \mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1}) = \min_{X,E \in \mathbb{S}^n} \quad & \tfrac{1}{2} \|E\|_1 \\ s.t. \quad & X + E = A_2 \\ & X \in \mathcal{C}_{\mathcal{G}_1}, \end{aligned} \tag{P}$$

It is evident that this problem provides a lower bound on $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$ as the objective function here is a lower bound of the objective of (1) over the constraint set of (1), and further the constraint set of (P) is an outer approximation of the constraint set of (1). Unlike the optimal value $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$ of (1), the optimal value $\mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1})$ of (P) is not symmetric; specifically, if $\mathcal{C}_{\mathcal{G}_2}$ is some invariant convex set containing $\{\Pi A_2 \Pi^T : \Pi \text{ is an } n \times n \text{ permutation matrix}\}$, then in general $\mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1}) \neq \mathrm{GED}_{LB}(\mathcal{G}_2, \mathcal{G}_1; \mathcal{C}_{\mathcal{G}_2})$. Therefore, in practice we propose computing both $\mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1})$ and $\mathrm{GED}_{LB}(\mathcal{G}_2, \mathcal{G}_1; \mathcal{C}_{\mathcal{G}_2})$ for some invariant convex sets $\mathcal{C}_{\mathcal{G}_1}$ and $\mathcal{C}_{\mathcal{G}_2}$ corresponding to $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively, and taking the larger of these quantities as both constitute lower bounds on $\mathrm{GED}(\mathcal{G}_1, \mathcal{G}_2)$.

This discussion leads naturally to the following question – which invariant convex set $\mathcal{C}_{\mathcal{G}}$ best captures the structural properties of a graph $\mathcal{G}$? Employing such a set in the relaxation (P) would provide better lower bounds on the edit distance. Letting $A \in \mathbb{S}^n$ be an adjacency matrix of $\mathcal{G}$, the 'tightest' invariant convex set that contains the collection $\{\Pi A \Pi^T : \Pi \text{ is an } n \times n \text{ permutation matrix}\}$ is simply the convex hull of this collection. However, this convex hull is intractable to describe for general graphs $\mathcal{G}$ (unless P=NP). As a result, it is of interest to obtain computationally tractable invariant convex relaxations that reflect the structure in $\mathcal{G}$. In the next subsection, we give a list of invariant convex sets that are tractable to compute and that can be 'tuned' to the structure of $\mathcal{G}$. These invariants can either be used individually or combined (at increased computational expense), thus yielding a flexible and powerful framework for obtaining bounds on the graph edit distance. The focus of the rest of the paper is on investigating the utility of these invariants theoretically as well as via numerical experiments; we give a summary of our main contributions in Section 1.3.

1.2 Convexity and Graph Invariants

We list here a few examples of invariant convex sets that play a prominent role in this paper; we refer the interested reader to [6] for a more exhaustive list as well as additional properties of invariant convex sets.

Loopless and edge weight constraints Looplessness and edge weight bounds are not especially powerful constraints, but they nonetheless serve as simple examples of invariant convex sets. Looplessness corresponds to the constraint set $\{M \in \mathbb{S}^n \mid M_{ii} = 0 \text{ for } i = 1, \ldots, n\}$, and bounds on the edge weights for unweighted graphs (for example) can be specified via the set $\{M \in \mathbb{S}^n \mid 0 \leq M_{ij} \leq 1 \text{ for } i, j = 1, \ldots, n\}$.

Spectral invariants Let $\mathcal{G}$ be a graph represented by an adjacency matrix $A \in \mathbb{S}^n$ with eigenvalues $\lambda(A) \in \mathbb{R}^n$. The smallest convex set containing all graphs that are isospectral to $\mathcal{G}$ is given by the following *Schur-Horn orbitope* associated to A [34]:

$$\mathcal{C}_{\mathcal{SH}(\mathcal{G})} = \mathrm{conv}\{M \in \mathbb{S}^n \mid \lambda(M) = \lambda(A)\}.$$

This set consists precisely of those matrices whose spectra are majorized by $\lambda(A)$. One can replace the list of eigenvalues in this example with the degree sequence of a graph, and in a similar vein, consider the convex hull of all adjacency matrices representing graphs with the same degree sequence; see [6] for more details.

A prominent way in which invariant convex sets can be constructed is via sublevel sets of convex graph invariants:

Definition 2. *[6] A function $f : \mathbb{S}^n \to \mathbb{R}$ is a* convex graph invariant *if it is convex and if $f(M) = f(\Pi M \Pi^T)$ for all $M \in \mathbb{S}^n$ and all $n \times n$ permutation matrices Π.*

As with invariant convex sets, convex graph invariants characterize structural properties of a graph that are invariant to vertex relabeling. The following convex graph invariants play a prominent role in our work:

Inverse of the stability number A stable set (also known as independent set) of a graph $\mathcal{G}$ is a subset of vertices of $\mathcal{G}$ such that no two vertices in the subset are connected. The stability number of a graph $\mathcal{G}$ is a graph invariant that is equal to the size of the largest stable set of $\mathcal{G}$. It was shown by Motzkin and Straus [27] that the inverse of the stability number admits the following variational description, where $A \in \mathbb{S}^n$ is an adjacency matrix representing $\mathcal{G}$:

$$\begin{aligned} \text{inv-stab-number}(A) = \min_{x \in \mathbb{R}^n} \; & x'(I+A)x \\ s.t. \quad & \sum_i x_i = 1, \; x_i \geq 0 \text{ for } i = 1, \ldots, n. \end{aligned}$$

As the stability number of a graph is NP-hard to compute for general graphs (the above program may be reformulated as a conic program with respect to the completely positive cone), the following tractable relaxation based on doubly nonnegative matrices is widely employed:

$$\begin{aligned} f(A) = \min_{X \in \mathbb{S}^n} \; & \operatorname{Tr}(X(I+A)) \\ s.t. \quad & X \succeq 0, \; \mathbf{1}'X\mathbf{1} = 1, X_{ij} \geq 0 \text{ for } i,j = 1, \ldots, n. \end{aligned} \tag{2}$$

One can check that both inv-stab-number(A) and $f(A)$ are concave graph invariants.

Maximum cut The maximum cut value of a graph is the maximum over all partitions of the vertices of the sum of the weights of the edges between the partitions. For a graph $\mathcal{G}$ specified by adjacency matrix $A \in \mathbb{S}^n$, the maximum cut value is given as:

$$\text{max-cut}(A) = \max_{y \in \{-1,1\}^n} \tfrac{1}{4} \sum_{i,j} A_{i,j}(1 - y_i y_j).$$

As this value is NP-hard to compute for general graphs, the following celebrated efficiently-computable relaxation is commonly used [14]:

$$\begin{aligned} g(A) = \max_{X \in \mathbb{S}^n} \; & \tfrac{1}{4} \operatorname{Tr}(A\,(\mathbf{1}\mathbf{1}^T - X)) \\ s.t. \quad & X \succeq 0, \; X_{ii} = 1 \text{ for } i = 1, \ldots, n. \end{aligned} \tag{3}$$

Both max-cut(A) and $g(A)$ are convex graph invariants as they are each invariant under conjugation of the argument by a permutation matrix and they are each expressed as a pointwise maximum of affine functions.

1.3 Our Contributions

The invariant convex sets listed in the previous section when used in the context of the optimization problem (P) all lead to valid lower bounds on the edit distance between two graphs. The question then is whether certain invariants are more naturally suited to particular structural properties of graphs. The main focus of this paper is on identifying attributes of graphs for which the invariants described above are well-suited, and evaluating in these contexts the quality of the bounds obtained via (P) both theoretically and through numerical experiments. Specifically, we say that an invariant convex constraint set $\mathcal{C}_{\mathcal{G}}$ is well-suited to the structure of a graph $\mathcal{G}$ if $\text{GED}_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{G}})$ provides a tight (or high-quality) lower bound of $\text{GED}(\mathcal{G}, \mathcal{G}')$ for all graphs $\mathcal{G}'$ that are obtained via a small

number of edge deletions and insertions applied to $\mathcal{G}$ (here 'small' is interpreted relative to the total number of edges in $\mathcal{G}$).

In Section 2, we investigate theoretically the effectiveness of the Schur-Horn orbitope as an invariant convex set in providing lower bounds on the graph edit distance via (P). We consider a stylized setting in which a graph $\mathcal{G}$ on n vertices is modified to a graph $\mathcal{G}'$ by adding or removing at most d edges incident to each vertex of $\mathcal{G}$. We prove in Theorem 3 (see Section 2.1) that the optimal value of the convex program (P) with a Schur-Horn orbitope constraint set equals the graph edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $\mathrm{GED}_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{SH(\mathcal{G})}) = \mathrm{GED}(\mathcal{G}, \mathcal{G}')$ provided: 1) d is sufficiently small, 2) $\mathcal{G}$ has eigenspaces with the property that there exists a linear combination of the associated projection operators onto these spaces so that the largest entry in magnitude is suitably bounded, and 3) any matrix supported on entries corresponding to edits of $\mathcal{G}$ has only a small amount of its energy on each of the eigenspaces of $\mathcal{G}$; see Theorem 3 for precise details. Conditions similar to the third requirement appear in the authors' earlier work on employing the Schur-Horn orbitope in the context of the planted subgraph problem [5]. However, the second condition is novel and is motivated by the context of the present paper on graph edit distance. Under the additional assumption that $\mathcal{G}$ is vertex-transitive, Corollary 4 provides a simple formula on the maximum allowable number d of edge additions/deletions per vertex of $\mathcal{G}$; we illustrate the utility of this formula by computing bounds on d for many graph families such as Johnson graphs, Kneser graphs, Hamming graphs and other strongly regular graphs. Indeed, for some of these families, our results are 'order-optimal' in the sense that our bounds on d are on the order of the degree of $\mathcal{G}$. The proofs of the main results of Section 2 are given in Section 3.

In Section 4, we conduct a detailed numerical evaluation of the power and limitations of convex invariant relaxations based on the inverse stability number (via the tractable approximation (2)) and the maximum cut value (via the tractable approximation (3)). We do not provide precise theoretical guarantees due to a lack of a detailed characterization of the facial structure of the associated convex sets. Nonetheless, we identify classes of graph edit distance problems for which these constraints produce high-quality lower bounds. Specifically, we observe that a convex relaxation based on the Motzkin-Straus approximation of the inverse of the stability number provides useful lower bounds on graph edit distance if one of the graphs has the property that the removal of any edge increases the graph's stability number; graphs with such a property have been studied in the extremal graph theory literature [21] and we refer the reader to Section 4.1 for further details. Similarly, in Section 4.2, we observe that a convex relaxation based on the Goemans-Williamson approximation of the maximum cut value produces effective lower bounds on the graph edit distance if the addition of any edge to one of the graphs increases that graph's maximum cut value; windmill graphs are a prominent example that possess such a property. In both Sections 4.1 and 4.2, we present empirical results that corroborate our observations.

In Section 5 we demonstrate the utility of our framework in providing lower bounds on the average pairwise graph edit distance in two chemistry datasets consisting of a collection of molecules known as Alkanes and Polycyclic Aromatic Hydrocarbons (PAH). The PAH dataset in particular consists of large structures for which exact computation of graph edit distance is prohibitively expensive. The best-known upper bound on the average graph edit distance over all pairs of graphs in this dataset is 29.8, and to the best of our knowledge, the exact value of this quantity is not known [3]. Indeed, much of the literature featuring the PAH dataset aims at providing an upper bound on the average pairwise graph edit distance. Our framework provides a lower bound of 21.6 on the average pairwise graph edit distance of PAH, which appears to be the best available bound to date. In obtaining these results, we combine invariant convex sets based on the Schur-Horn orbitope, the Motzkin-Straus approximation of the inverse stability number, and the Goemans-Williamson approximation of the maximum-cut value.

Notation We denote the normal cone at a point $x \in \mathcal{C}$ of a closed convex set $\mathcal{C}$ by $\mathcal{N}_{\mathcal{C}}(x)$. The projection map onto a subspace $\mathcal{E} \subset \mathbb{R}^n$ is denoted by $P_{\mathcal{E}} : \mathbb{R}^n \to \mathbb{R}^n$. For a collection of subspaces $\mathcal{E}_i \subset \mathbb{R}^n,\ i \in \{1, \ldots, m\}$, the operator $\mathcal{P}_{ij} : \mathbb{S}^n \to \mathbb{S}^n$ is defined as $\mathcal{P}_{ij} := P_{\mathcal{E}_j} \otimes P_{\mathcal{E}_i}$, i.e., $\mathcal{P}_{ij}(A) = P_{\mathcal{E}_i} A P_{\mathcal{E}_j}$. The restriction of a (usually self-adjoint) linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ to an invariant subspace $\mathcal{E}$ of f is denoted by $f|_{\mathcal{E}} : \mathcal{E} \to \mathcal{E}$.

2 Theoretical Guarantees for the Schur-Horn Orbitope Constraint

In this section, we give theoretical guarantees that describe conditions under which employing the Schur-Horn orbitope as an invariant convex constraint set in (P) leads to the associated lower bound on the graph edit distance being *tight*, i.e., the optimal value of (P) equals the graph edit distance. Concretely, we consider conditions on a graph $\mathcal{G}$ and the structure of the edits that transform $\mathcal{G}$ to another graph $\mathcal{G}'$ so that $\mathrm{GED}(\mathcal{G}, \mathcal{G}') = \mathrm{GED}_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$. We begin with a description of our main theoretical results in Section 2.1, some consequences of these results for specific graph families in Section 2.2, and finally an experimental demonstration on the utility of the Schur-Horn orbitope on stylized problems in Section 2.3. The proofs of the results of this section are deferred to Section 3.

As the normal cones at extreme points of the Schur-Horn orbitope play a prominent role in the optimality conditions of (P), we state the relevant result here:

Lemma 1. *[5] Let $\mathcal{G}$ be any unweighted graph with m eigenvalues. A matrix W is an extreme point of $\mathcal{SH}(\mathcal{G})$ if and only if it has the same eigenvalues as $\mathcal{G}$ (counting multiplicity). Further, the relative interior of the normal cone relint($\mathcal{N}_{\mathcal{SH}(\mathcal{G})}(W)$) at an extreme point W consists of those matrices Q that satisfy the following conditions:*

1. *Q is simultaneously diagonalizable with W,*
2. *$\lambda_{\min}(Q|_{\mathcal{E}_i}) > \lambda_{\max}(Q|_{\mathcal{E}_{i+1}})\ \forall\, i \in \{1, \ldots, m-1\}$,*

where $\mathcal{E}_i$ for $i \in \{1, \ldots, m\}$ are eigenspaces of W ordered such that the corresponding eigenvalues are sorted in a decreasing order.

From Lemma 1 we observe that the relative interior of the normal cones of the Schur-Horn orbitope at extreme points are 'larger' if the underlying graph $\mathcal{G}$ consists of few distinct eigenvalues. This observation along with various properties of the eigenspaces of $\mathcal{G}$ play a prominent role in the analysis in this section.

2.1 Main Results

We present here the statements of our main theoretical results concerning the performance of the Schur-Horn orbitope as a constraint set in (P). In addition to various structural properties of $\mathcal{G}$, our results are described in terms of a parameter d that denotes the maximum number of deletions/additions of edges that are incident to any vertex of $\mathcal{G}$. Informally, we should expect the Schur-Horn orbitope constraint to be effective in exactly computing the graph edit distance if a matrix representing the edits from $\mathcal{G}$ to $\mathcal{G}'$ has only a small amount of its energy on each of the eigenspaces of $\mathcal{G}$. The reason behind this observation is that if the edits were largely concentrated in the eigenspaces of $\mathcal{G}$, then the eigenspaces of $\mathcal{G}'$ would be close to those of $\mathcal{G}$. This would result in an identifiability problem from the perspective of the Schur-Horn orbitope, which is based purely on the spectral properties of $\mathcal{G}$. To formalize this notion, we present the following definition which plays a key role in our analysis:

Definition 3. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues. Let $P_i,\ i = 1, \ldots, m$ represent projection maps onto the eigenspaces of $\mathcal{G}$ indexed by decreasing order of the corresponding eigenvalues and let $\mathcal{P}_{ii} = P_i \otimes P_i$. Fix a positive integer d and $\alpha \in [0,1]^m$. Define the parameter $\xi(\alpha, d, \mathcal{G})$ to be the smallest value*

$$\left\| [I - \sum_{i=1}^{m} \alpha_i \mathcal{P}_{ii}](W) \right\|_\infty \leq \xi(\alpha, d, \mathcal{G}) \left\| W \right\|_\infty .$$

for all $W \in \mathbb{S}^n$ with at most d nonzero entries per row/column.

Remark 1. *The maps P_i represent projections onto eigenspaces of an adjacency matrix representing $\mathcal{G}$, but we simply refer to these as eigenspaces of $\mathcal{G}$ with an abuse of terminology. The reason is that the quantity $\xi(\alpha, d, \mathcal{G})$ is a graph parameter (for each fixed α, d) and does not depend on a specific labeling of the vertices of $\mathcal{G}$.*

Remark 2. *The parameter $\xi(\alpha, d, \mathcal{G})$ is a restricted version of the induced (entrywise) infinity norm $\|I - \sum_i^m \alpha_i \mathcal{P}_{ii}\|_{\infty\to\infty}$, with the key difference being that $\xi(\alpha, d, \mathcal{G})$ computes the induced gain of the operator $I - \sum_i^m \alpha_i \mathcal{P}_{ii}$ restricted to inputs that have at most d nonzeros per row/column.*

The quantity $\xi(\alpha, d, \mathcal{G})$ helps quantify the idea described previously about the energy of the edits not being confined excessively to the eigenspaces of $\mathcal{G}$. As the specific edit pattern is not known in advance, this quantity is agnostic to the particular edits and is parametrized only in terms of the maximum number of edge deletions/additions that are incident to any vertex. In our main results described next, larger values of ξ make it harder to satisfy our sufficient conditions on tightness of our lower bounds. As the value of ξ depends on the selection of the parameter α, our main results allow for flexibility in the choice of this parameter, and we describe in Section 2.2 how specific choices lead to concrete consequences on the exactness of the relaxation (P) with the Schur-Horn orbitope constraint for various graph families. We present next a result that establishes basic optimality conditions of the convex program (P):

Lemma 2. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues, and let $\mathcal{G}'$ be a graph that is obtained from $\mathcal{G}$ via edge deletions/additions such that each vertex is incident to at most d edits. Let $A, A + E^* \in \mathbb{S}^n$ represent the graphs $\mathcal{G}$ and $\mathcal{G}'$, respectively; that is, E^* consists of at most d nonzeros per row/column. Let $\Omega \subset \mathbb{S}^n$ denote the subspace consisting of all matrices with nonzeros contained within the support of E^*. Suppose a vector $\alpha \in [0,1]^m$ and a matrix $Q \in \mathbb{S}^n$ satisfy the following conditions:*

1. $\mathcal{P}_\Omega(Q) = \text{sign}(E^*)$,
2. $||\mathcal{P}_{\Omega^c}(Q)||_\infty < 1$,
3. $Q \in \textit{relint}(\mathcal{N}_{\mathcal{SH}(\mathcal{G})}(A))$,
4. $\xi(\alpha, d, \mathcal{G}) < 1$.

Then we have that the convex relaxation (P) with the Schur-Horn orbitope constraint exactly computes the edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $GED(\mathcal{G}, \mathcal{G}') = GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$, with the optimal solution being unique and achieved at a matrix that specifies an optimal set of edits.

Proof. From the first three conditions, one can conclude that the pair (A, E^*) is an optimal solution of (P) by a direct application of the KKT conditions. Uniqueness can be established by standard arguments regarding transverse intersections of the subspace Ω and the invariant spaces of $\mathcal{G}$; see [6, Proposition 2] and [8, Lemma 6]. □

Conditions 1, 2, and 3 of this lemma essentially require that the subdifferential at a matrix specifying the edits with respect to the ℓ_1 norm has a nonempty intersection with the relative interior of the normal cone at an adjacency matrix representing $\mathcal{G}$ with respect to the Schur-Horn orbitope. In papers on the topic of low-rank matrix completion and matrix decomposition [4, 7], a convenient approach to ensuring that such types of conic intersection conditions can be satisfied is based on requiring that nullspace (the eigenspace corresponding to an eigenvalue of zero) of the low-rank matrix is suitably 'incoherent', i.e., that there are no elements of this nullspace with energy concentrated in a single location. In our context, all of the eigenspaces of $\mathcal{G}$ play a role rather than just a single distinguished eigenspace, and accordingly we describe next a weighted form of an incoherence-type condition:

Definition 4. *Let $\mathcal{G}$ be a graph on n nodes with m distinct eigenvalues and let $P_1, \ldots, P_m \in \mathbb{S}^n$ denote the projection matrices onto the associated eigenspaces indexed by decreasing order of the corresponding eigenvalues. Fix any $\gamma \in \mathbb{R}^m$. We define the parameter $\rho(\gamma, \mathcal{G})$ as follows:*

$$\rho(\gamma, \mathcal{G}) := \left\| \sum_{i=1}^{m} \gamma_i P_i \right\|_\infty .$$

Here the matrix $\|\cdot\|_\infty$ norm is the largest entry of the argument in magnitude. In the literature on inverse problems involving low-rank matrices, one typically considers the infinity norm of the projection map onto the nullspace as well as variants of this quantity. Thus, in this sense the parameter $\rho(\gamma, \mathcal{G})$ is a weighted generalization that is more suited to our setup. We state next our main theorem in terms of sufficient conditions involving the two parameters we have introduced in this section:

Theorem 3. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues, and let $\mathcal{G}'$ be a graph that is obtained from $\mathcal{G}$ via edge deletions/additions such that each vertex is incident to at most d edits. Suppose the following two conditions are satisfied for some $\gamma \in \mathbb{R}^m$ and $\alpha \in [0,1]^m$:*

1. $2\,\xi(\alpha, d, \mathcal{G}) + \rho(\gamma, \mathcal{G}) < 1$,
2. $\frac{(\alpha_i + \alpha_{i+1})\,(1+\rho(\gamma,\mathcal{G}))\,d}{1-\xi(\alpha,d,\mathcal{G})} < \gamma_{i+1} - \gamma_i, \ \forall\, i \in \{1, \ldots, m-1\}$.

Then the convex relaxation (P) with the Schur-Horn orbitope constraint exactly computes the edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $GED(\mathcal{G}, \mathcal{G}') = GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$, with the optimal solution being unique and achieved at a matrix that specifies an optimal set of edits.

This theorem states that the relaxation (P) with the Schur-Horn orbitope constraint set succeeds in calculating the graph edit distance exactly if 1) d is small enough, 2) there exists a vector α with small entries such that $\xi(\alpha, d, \mathcal{G})$ is also suitably small, and 3) there exists an ordered vector γ with well-separated entries that yields a small value of $\rho(\gamma, \mathcal{G})$. As discussed in the next subsection, graphs with a small number of well-separated eigenvalues offer an ideal candidate. Specifically, for such graph families, we give concrete consequences in terms of bounds on the maximum number d of edits per vertex via particular choices for α and γ in Theorem 3.

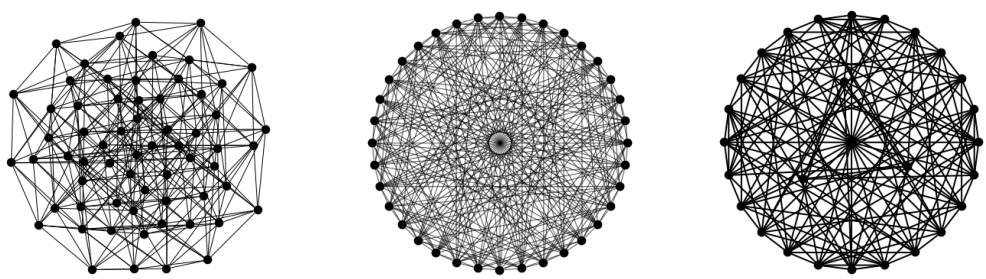
2.2 Consequences for Graph Families with Few Eigenvalues

Theorem 3 constitutes our most general result on the tightness of the Schur-Horn orbitope constraint in computing the graph edit distance when employed as a constraint set in the context of (P). The generality of the result stems from the wide range of flexibility provided by the vectors γ and α. In Corollary 4, we consider specific choices of these parameters to obtain concrete bounds in terms of graph parameters that can be computed easily:

Corollary 4. *Let $\mathcal{G}$ be a vertex-transitive graph on n vertices consisting of m distinct eigenvalues, and let κ denote the multiplicity of the eigenvalue with the second-highest multiplicity. Suppose $\mathcal{G}'$ is a graph on n vertices that can be obtained from $\mathcal{G}$ with the addition or removal of at most d edges incident to each vertex of $\mathcal{G}$. Then there exists a constant c depending only on m so that the optimal value $GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ of (P) equals $GED(\mathcal{G}, \mathcal{G}')$ provided*

$$d \leq c\frac{n}{\kappa}.$$

The particular dependence on the multiplicity of the eigenvalue with second-largest multiplicity is due to the choices of α and γ in Theorem 3 that we have employed in our proof; see Section 3 for more details. In the sequel we give consequences of this result for specific graph families in which the number of distinct eigenvalues is small (for example, three or four). In the context of such graphs, the relaxation (P) is tight even when the number of edits per vertex is large so long as the value of κ is suitably small. Indeed, for several graph families we observe that Corollary 4 produces 'order-optimal' bounds as the largest value of d that is allowed is on the same order as the degree of the underlying graphs.



(a) (b) (c)

Figure 2: From left to right: Hamming graph H(3,4), 9-Triangular graph, generalized quadrangle-(2,4) graph.

Johnson Graphs A Johnson graph $J(k, \ell)$ with $\ell > 0$ is a graph on $n = \binom{k}{\ell}$ vertices that correspond to the ℓ-element subsets of a set of k elements. Two vertices of a Johnson graph are connected if the corresponding subsets of these vertices contain $\ell - 1$ common elements. The Johnson graph $J(k, \ell)$ is vertex-transitive and contains $\ell + 1$ distinct eigenvalues. For $k \geq 2\ell$ and $j \in \{0, \ldots, \ell\}$, the multiplicity of its j'th eigenvalue is $\binom{k}{j}$ -$\binom{k}{j-1}$ for $j > 0$ and one for $j = 0$. For small values of ℓ, the multiplicity of the second most repeated eigenvalue is about $k^{\ell-1}$. As a result, for small fixed values of ℓ, Corollary 4 states that the convex relaxation (P) is tight provided

$$d \lesssim n^{\frac{1}{\ell}}.$$

Kneser Graphs A Kneser graph $K(k, \ell)$ with $\ell > 0$ shares certain aspects with Johnson graphs. Specifically, the vertices of $K(k, \ell)$ coincide with the ℓ-element subsets of a set of k elements, as with Johnson graphs. However, two vertices of a Kneser graph are connected if the subsets corresponding to these vertices are disjoint. Kneser graphs are vertex-transitive, and their eigenvalues exhibit the same multiplicities as those of the Johnson graphs $J(k, \ell)$. As a result, for small fixed values of ℓ, Corollary 4 implies that the relaxation (P) is tight provided:

$$d \lesssim n^{\frac{1}{\ell}}.$$

Hamming Graphs A Hamming graph $H(\ell, q)$ consists of q^ℓ vertices (see Figure 2a for a depiction of $H(3,4)$). Each vertex of $H(\ell, q)$ corresponds to a sequence of length ℓ from a set with q distinct elements. Two vertices are connected if their associated sequences differ in exactly one coordinate, i.e., their Hamming distance is equal to 1. Hamming graphs are vertex-transitive, and the spectrum of $H(\ell, q)$ consists of $\ell+1$ distinct eigenvalues with multiplicities $\binom{\ell}{i}(q-1)^i$, $i \in \{0, \ldots, \ell\}$. Therefore, for a small fixed value of ℓ, Corollary 4 states that the relaxation (P) is tight provided:

$$d \lesssim n^{\frac{1}{\ell}}.$$

Vertex-Transitive Strongly Regular Graphs A strongly regular graph on n vertices with degree r is defined by the property that every pair of adjacent vertices has d_a common neighbors and every pair of nonadjacent vertices has d_{na} common neighbors. Such graphs are generally denoted $srg(n, r, d_a, d_{na})$. Due to their rich algebraic structure, strongly regular graphs have only three distinct eigenvalues with multiplicities equal to one and $\frac{1}{2}\left[(n-1) \pm \frac{2r+(n-1)(d_a-d_{na})}{\sqrt{(d_a-d_{na})^2+4(r-d_{na})}}\right]$. Furthermore, many strongly regular graphs are also vertex-transitive and as a result, our Corollary 4 is applicable. We highlight two prominent examples:

- A k-Triangular graph T_k on $n = \binom{k}{2}$ vertices is a vertex-transitive strongly regular graph with parameters $srg(k(k-1)/2, 2(k-2), k-2, 4)$ (in fact, T_k is also isomorphic to the Johnson graph $J(k, 2)$); see Figure 2b for the 9-Triangular graph. Corollary 4 states that the convex relaxation (P) is tight provided:

  $$d \lesssim n^{\frac{1}{2}}.$$

  Incidentally, the degree of T_k also scales as $n^{\frac{1}{2}}$; as a result, Corollary 4 is tight for this family up to constant factors.

- A generalized quadrangle is an incidence relation satisfying certain geometric axioms on points and lines. A generalized quadrangle of order (s, t) gives rise to a strongly regular graph with parameters $srg((s+1)(st+1), s(t+1), s-1, t+1)$ denoted by $GQ(s, t)$ on $n = (s+1)(st+1)$ vertices – see Figure 2c for an illustration of the vertex-transitive graph $GQ(2, 4)$. Considering generalized quadrangle graphs $GQ(s, s^2)$ when they are vertex-transitive, Corollary 4 implies that the relaxation (P) is tight provided

  $$d \lesssim n^{\frac{1}{4}}.$$

In Section 2.3 we demonstrate the utility of our framework via numerical experiments on edit distance problems involving the graphs T_9 and $GQ(2, 4)$.

2.3 Numerical Experiments

We demonstrate the utility of the Schur-Horn orbitope as a constraint set in (P) in obtaining bounds on the graph edit distance between graphs $\mathcal{G}$ and $\mathcal{G}'$. In our experiments, we fix $\mathcal{G}$ to be either the 9-triangular graph T_9 (Figure 2b) or the generalized quadrangle-(2,4) graph $GQ(2, 4)$ (Figure 2c) introduced previously. The graph T_9 consists of 36 vertices and 252 edges and the graph GQ(2,4) consists of 27 vertices and 135 edges. Both of these are strongly regular graphs. In each case, the corresponding graph $\mathcal{G}'$ is obtained by adding/deleting edges randomly (both addition and deletion occur with equal probability) to achieve a desired number of edits. When $\mathcal{G}$ is T_9 we

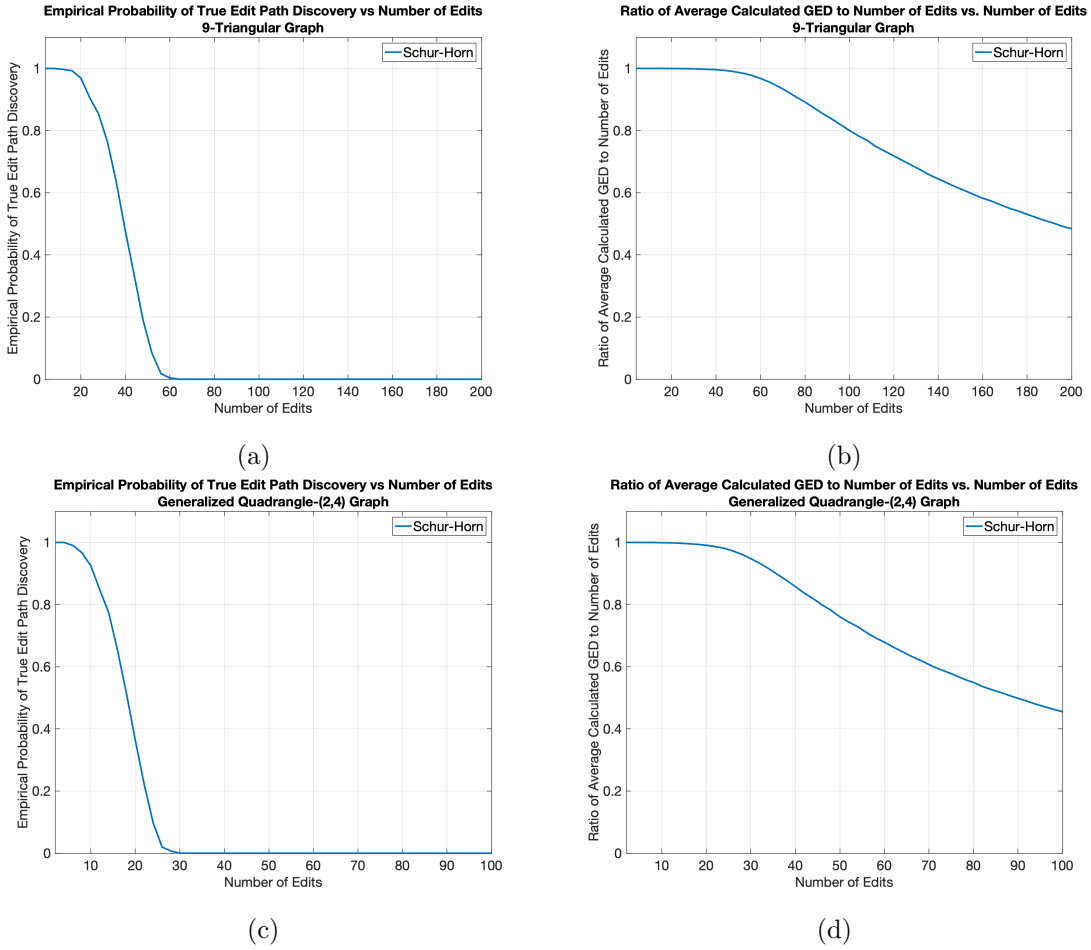


Figure 3: Performance of our framework (P) with the Schur-Horn constraint. Left, empirical probability of discovering the true graph edit path. Right, ratio of average calculated graph edit distance to number of edit operations. First row corresponds to the 9-triangular graph and the second row to the generalized quadrangle-(2,4) graph.

vary the number of edits from four to 200 in increments of four, and when $\mathcal{G}$ is $GQ(2,4)$ we vary the number of edits from two to 100 in increments of two. For each number of edits, we consider 1000 random trials and we report the probability that $\mathrm{GED}(\mathcal{G},\mathcal{G}') = \mathrm{GED}_{LB}(\mathcal{G},\mathcal{G}';\mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ and the ratio of the average computed lower bound $\mathrm{GED}_{LB}(\mathcal{G},\mathcal{G}';\mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ to the number of edits. In particular, we declare that $\mathrm{GED}(\mathcal{G},\mathcal{G}') = \mathrm{GED}_{LB}(\mathcal{G},\mathcal{G}';\mathcal{C}_{\mathcal{SH}(\mathcal{G})})$ if the infinity norm (maximum entrywise magnitude) of the difference between the optimal solution $\hat{E}$ and the true edit matrix E^* is less than 0.01. The results are shown in Figure 3 and they were obtained using the CVX parser [15, 16] and the SDPT3 solver [37]. As these plots demonstrate, the convex relaxation (P) with a Schur-Horn orbitope as an invariant convex constraint set is tight when the number of edits is small and leads to effective lower bounds when the number of edits is large.

3 Proofs of Results from Section 2

3.1 Constructing a Dual Certificate

We describe here a method for constructing a suitable dual certificate satisfying the conditions of Lemma 2, and we prove that this construction is valid whenever certain conditions involving the parameters ξ and ρ from Section 2 are satisfied. Our proofs are presented in the context of two intermediary lemmas, which are then used to prove Theorem 3. Specifically, our approach to constructing $Q \in \mathbb{S}^n$ that satisfies the requirements of Lemma 2 is to express Q as follows:

$$Q = R + \Delta.$$

Here $R \in \mathbb{S}^n$ plays the role of a 'reference' matrix that depends purely on the underlying graph $\mathcal{G}$, while $\Delta \in \mathbb{S}^n$ is a perturbation that additionally depends on the specific edits that transform $\mathcal{G}$ to $\mathcal{G}'$. We begin by stating an easily-proved result that serves as the basis for our subsequent development:

Lemma 5. *Let $\mathcal{G}$ be a graph on n vertices with m distinct eigenvalues, and let $\mathcal{G}'$ be a graph that is obtained from $\mathcal{G}$ via edge deletions/additions such that each vertex is incident to at most d edits. Let $A, A + E^* \in \mathbb{S}^n$ represent the graphs $\mathcal{G}$ and $\mathcal{G}'$, respectively; that is, E^* consists of at most d nonzeros per row/column. Let $\Omega \subset \mathbb{S}^n$ denote the subspace consisting of all matrices with nonzeros contained within the support of E^*. Let $P_i \in \mathbb{S}^n$, $i = 1, \ldots, m$ denote projection maps onto the eigenspaces of A indexed by decreasing order of the corresponding eigenvalues. Suppose a vector $\alpha \in [0,1]^m$, a vector $\gamma \in \mathbb{R}^m$, and a matrix $\Delta \in \mathbb{S}^n$ satisfy the following conditions with $R = \sum_i \gamma_i P_i$:*

1. $\mathcal{P}_\Omega(\Delta) + \mathcal{P}_\Omega(R) = \text{sign}(E^*)$,
2. $||\mathcal{P}_{\Omega^c}(\Delta)||_\infty + ||\mathcal{P}_{\Omega^c}(R)||_\infty < 1$,
3. $\mathcal{P}_{ij}(\Delta) = 0$, $\forall i, j \in \{1, \ldots, m\}$, $i \neq j$,
4. $||\mathcal{P}_{ii}(\Delta)||_2 + ||\mathcal{P}_{i+1,i+1}(\Delta)||_2 < \gamma_{i+1} - \gamma_i$, $\forall i \in \{1, \ldots, m-1\}$,
5. $\xi(\alpha, d, \mathcal{G}) < 1$.

Then the convex relaxation (P) with the Schur-Horn orbitope constraint exactly computes the edit distance between $\mathcal{G}$ and $\mathcal{G}'$, i.e., $GED(\mathcal{G}, \mathcal{G}') = GED_{LB}(\mathcal{G}, \mathcal{G}'; \mathcal{C}_{\mathcal{SH}(\mathcal{G})})$, with the optimal solution being unique and achieved at a matrix that specifies an optimal set of edits.

Proof. One can check that the conditions of Lemma 2 are satisfied by setting $Q = R + \Delta$. □

This lemma highlights the role of the parameter γ, in particular demonstrating that larger separation among the values of γ makes condition 4 easier to satisfy but may also increase the value of $||\mathcal{P}_{\Omega^c}(R)||_\infty$, thus making condition 2 potentially harder to satisfy.

We now move on to the perturbation term Δ. As this matrix must satisfy several of the constraints discussed in Lemma 5, its construction is somewhat delicate. We build on the ideas developed in [8] in the context of low-rank matrix recovery, but with certain adaptations that are crucial to our setting. We construct Δ as an element in the range of an operator $\mathcal{L}^\alpha : \mathbb{S}^n \to \mathbb{S}^n$ that is parametrized by $\alpha \in [0,1]^m$:

$$\mathcal{L}^\alpha := \left(\sum_{i=1}^m \alpha_i \mathcal{P}_{ii}\right) \mathcal{P}_\Omega \left[I - \left(I - \sum_{i=1}^m \alpha_i \mathcal{P}_{ii}\right) \mathcal{P}_\Omega\right]^{-1}.$$

All of the operators here are as defined before. A point of the departure in the description of this operator relative to the ideas in [8] is that our version allows for 'fractional' contractions (as well as integral ones) based on the choice of α. When it is well-defined (i.e., the term involving the inverse is indeed invertible), the operator $\mathcal{L}^\alpha$ possesses a number of properties that lead to a convenient approach for constructing a suitable dual variable:

(P1) $\mathcal{P}_{ij}\mathcal{L}^\alpha = 0 \ \forall i, j \in \{1, \ldots, m\}$, $i \neq j$

(P2) $\mathcal{P}_\Omega \mathcal{L}^\alpha = \mathcal{P}_\Omega$.

In the context of Lemma 5, property (P1) ensures that that Δ is completely contained in a desired subspace, as stipulated by condition 3 of Lemma 5. Further, property (P2) implies that condition 1 of Lemma 5 is satisfied – in particular, we make use of this property to ensure that Δ takes on a desired value when restricted to Ω. Conditions 2 and 4 of Lemma 5 require that the quantities $\|\mathcal{P}_{ii}(\Delta)\|_2$ and $\|\mathcal{P}_{\Omega^c}(\Delta)\|_\infty$ to be sufficiently small – these conditions are satisfied by the operator $\mathcal{L}^\alpha$ as well, as documented next:

Lemma 6. *Consider the same setup as in Lemma 5. Fix any $\alpha \in [0,1]^m$ such that $\xi(\alpha, d, \mathcal{G}) < 1$. Then the operator $\mathcal{L}^\alpha : \mathbb{S}^n \to \mathbb{S}^n$ is well-defined (i.e., the term containing the inverse is indeed invertible) and the following inequalities hold:*

1. $\|[\mathcal{P}_{\Omega^c}\mathcal{L}^\alpha](X)\|_\infty \leq \frac{\xi(\alpha,d,\mathcal{G})\|X\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}$,
2. $\|[\mathcal{P}_{ii}\mathcal{L}^\alpha](X)\|_2 \leq \frac{\alpha_i d\|X\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}$.

In addition to providing upper bounds that serve as a foundation for the proof of our main theorem, Lemma 6 conveys the significance of the parameter $\xi(\alpha, d, \mathcal{G})$. Specifically, a suitably small value of $\xi(\alpha, d, \mathcal{G})$ guarantees that $\mathcal{L}^\alpha$ is well-defined, along with the conclusion that elements in the range of $\mathcal{L}^\alpha$ have small infinity norm (when restricted to Ω) and small operator norm (restricted to eigenspaces of $\mathcal{G}$). The lemma also suggests that the operator norm of the restriction of $\mathcal{L}^\alpha(X)$ to any eigenspace of $\mathcal{G}$ scales with the corresponding entry of α. Consequently, one can adjust α and γ to ensure that every inequality in Lemma 5 condition 4 is satisfied. Identifying the best values of α and γ to achieve this may be accomplished in special cases based on underlying structure in $\mathcal{G}$, as demonstrated by Corollary 4 and the many concrete consequences that are described in Section 2.2. In particular, in the proof of Corollary 4, we choose γ such that the separation between consecutive γ_i's is proportional to the sum of consecutive α_i's and we demonstrate that this approach yields easily-computable bounds based on properties of the underlying graph $\mathcal{G}$ on the highest number d of tolerable edits per vertex.

3.2 Proofs

3.2.1 Proof of Lemma 6

Our proof is analogous to that of [8, Lemma 8]. In order to avoid notational clutter, we denote $(I - \sum_{i=1}^m \alpha_i \mathcal{P}_{ii})$ as $\mathcal{P}_T$. Then from the definition of $\xi(\alpha, d, \mathcal{G})$, we have:

$$\|[\mathcal{P}_T\mathcal{P}_\Omega](X)\|_\infty \leq \xi(\alpha, d, \mathcal{G}) \|X\|_\infty . \tag{4}$$

Due to the assumption that $\xi(\alpha, d, \mathcal{G}) < 1$, we have that the series $I + \mathcal{P}_T\mathcal{P}_\Omega + \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega + \ldots$ converges geometrically with rate $\frac{1}{1-\xi(\alpha,d,\mathcal{G})}$ and equals $(I - \mathcal{P}_T\mathcal{P}_\Omega)^{-1}$. As such, the operator $\mathcal{L}^\alpha$ is well-defined.

Next, we proceed to the upper bounds. First we have that:

$$\begin{aligned}\|[\mathcal{P}_{\Omega^c}\mathcal{L}^\alpha](X)\|_\infty &= \left\|[\mathcal{P}_{\Omega^c}\mathcal{P}_T\mathcal{P}_\Omega(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_\infty \\ &\leq \left\|[\mathcal{P}_T\mathcal{P}_\Omega(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_\infty \\ &\leq \xi(\alpha,d,\mathcal{G})\left\|(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}(X)\right\|_\infty \\ &\leq \frac{\xi(\alpha,d,\mathcal{G})\|X\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}.\end{aligned}$$

One can check that the first equality holds based on a term-by-term comparison. The first inequality follows by dropping the projection $\mathcal{P}_{\Omega^c}$. Bounding the resulting quantity from above using $\xi(\alpha,d,\mathcal{G})$ yields the second inequality. The last inequality follows from the geometric convergence of $(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}$.

Next we bound the quantity involving the operator norm:

$$\begin{aligned}\|[\mathcal{P}_{ii}\mathcal{L}^\alpha](X)\|_2 &= \left\|[\alpha_i\mathcal{P}_{ii}\mathcal{P}_\Omega(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_2 \\ &\leq \alpha_i\left\|[\mathcal{P}_\Omega(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_2 \\ &\leq \alpha_i d\left\|[\mathcal{P}_\Omega(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_\infty \\ &\leq \alpha_i d\left\|[(I-\mathcal{P}_T\mathcal{P}_\Omega)^{-1}](X)\right\|_\infty \\ &\leq \frac{\alpha_i d\|X\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}.\end{aligned}$$

The first inequality holds by dropping the projection $\mathcal{P}_{ii}$. The second inequality holds due to the fact that the operator norm of a matrix with at most d entries per row/column can be bounded above by d times the maximum element in magnitude of the matrix. The third inequality holds by dropping the operator $\mathcal{P}_\Omega$. The final inequality follows from geometric convergence, as before.

3.2.2 Proof of Theorem 3

We prove that under the assumptions of this theorem the sufficient conditions of Lemma 5 are satisfied. Set $R=\sum_{i=1}^m \gamma_i P_i$ where $P_i\in\mathbb{S}^n$ is the projection matrix corresponding to the i'th eigenspace of $\mathcal{G}$. Denote the edits by a matrix $E^*\in\mathbb{S}^n$, and let the subspace of matrices with nonzero entries contained inside the support of E^* be denoted Ω. Set $M=\text{sign}(E^*)-\mathcal{P}_\Omega(R)$ and note that $M\in\Omega$. Condition 5 of Lemma 5 is satisfied based on assumption 1 of Theorem 3. As a result, the operator $\mathcal{L}^\alpha$ is well-defined by Lemma 6. Set $\Delta=\mathcal{L}^\alpha(M)$. We prove that $Q=R+\Delta$ satisfies the requirements of Lemma 5.

Condition 1 of Lemma 5: One can check that:

$$\mathcal{P}_\Omega(\Delta)+\mathcal{P}_\Omega(R)=\mathcal{P}_\Omega(\mathcal{L}^\alpha(M))+\mathcal{P}_\Omega(R)=\mathcal{P}_\Omega(M)+\mathcal{P}_\Omega(R)=\text{sign}(E^*).$$

Here the second equality holds due to property (P1) of the operator $\mathcal{L}^\alpha$.

Condition 2 of Lemma 5: We have that:

$$\|\mathcal{P}_{\Omega^c}(\Delta)\|_\infty+\|\mathcal{P}_{\Omega^c}(R)\|_\infty\leq\frac{\xi(\alpha,d,\mathcal{G})\|M\|_\infty}{1-\xi(\alpha,d,\mathcal{G})}+\|R\|_\infty\leq\frac{\xi(\alpha,d,\mathcal{G})+\rho(\gamma,\mathcal{G})}{1-\xi(\alpha,d,\mathcal{G})}<1.$$

The first inequality employed assertion 1 of Lemma 6, the second inequality follows from the triangle inequality and the definition of M, and the last inequality holds by assumption 1 of the theorem.

Condition 3 of Lemma 5: Follows from property (P2) of operator $\mathcal{L}^\alpha$.

Condition 4 of Lemma 5: One can check that:

$$\|\mathcal{P}_{ii}(\Delta)\|_2 + \|\mathcal{P}_{i+1,i+1}(\Delta)\|_2 = \|[\mathcal{P}_{ii}\mathcal{L}^\alpha](M)\|_2 + \|[\mathcal{P}_{i+1,i+1}\mathcal{L}^\alpha](M)\|_2$$
$$\leq \frac{(\alpha_i + \alpha_{i+1})(1 + \rho(\gamma, \mathcal{G}))\, d}{1 - \xi(\alpha, d, \mathcal{G})} < \gamma_{i+1} - \gamma_i, \ \forall\, i \in \{1, \ldots, m-1\}.$$

Here the first inequality follows from assertion 2 of Lemma 6 and the triangle inequality, and the second inequality follows from the assumption of the theorem.

3.2.3 Proof of Corollary 4

For this proof we require the notion of incoherence of a subspace, which measures how well the subspace is aligned with the standard basis vectors. This notion appears prominently in results on sparse signal recovery via convex optimization [11].

Definition 5. *Let $S \subseteq \mathbb{R}^n$ be a subspace and let P_S be the corresponding projection onto S. The* incoherence *of S is denoted $\mu(S)$ and is defined as*

$$\mu(S) := \max_i \|P_S e_i\|_2.$$

Here e_i is the i'th standard basis vector.

For any projection matrix P_S, one can check that the inequality $\|P_S\|_\infty \leq \mu(S)^2$ is satisfied.

Remark 3. *For vertex-transitive graphs, the diagonal entries of a projection matrix associated to any eigenspace of the graph are identical. As a result, the incoherence of an eigenspace $\mathcal{E}$ of a vertex-transitive graph on n vertices is equal to*

$$\mu(\mathcal{E}) = \sqrt{\frac{\dim(\mathcal{E})}{n}}.$$

We now proceed to the proof of the corollary. Denote the eigenspaces of $\mathcal{G}$ by $\mathcal{E}_i$ for $i \in \{1, \ldots, m\}$ ordered by decreasing eigenvalue order. Remark 3 implies that $\mu(\mathcal{E}_i) = \sqrt{\frac{\dim(\mathcal{E}_i)}{n}}$. Denote the second largest coherence of the eigenspaces of $\mathcal{G}$ by $\bar{\mu} = \sqrt{\frac{\kappa}{n}}$, and denote the index of the eigenspace with the highest incoherence by ℓ. Set $\alpha_\ell = 1$ and the remaining entries of α to 0. Furthermore, choose γ such that:

$$\gamma_{i+1} - \gamma_i = c_1 \frac{\alpha_i + \alpha_{i+1}}{\bar{\mu}^2} + \epsilon, \quad \forall\, i \in \{1, \ldots, m-1\}, \text{ for some } c_1 > 0,\, \epsilon > 0.$$

Here c_1 and ϵ are positive constants that can be as small as desired. To establish condition 2 of Theorem 3, we prove that the inequality below holds for all $i \in \{1, \ldots, m-1\}$:

$$\frac{(\alpha_i + \alpha_{i+1})\,(1 + \rho(\gamma, \mathcal{G}))\, d}{1 - \xi(\alpha, d, \mathcal{G})} \leq \gamma_{i+1} - \gamma_i - \epsilon = c_1 \frac{\alpha_i + \alpha_{i+1}}{\bar{\mu}^2}, \ \forall\, i \in \{1, \ldots, m-1\}.$$

Clearly, if $\alpha_i + \alpha_{i+1} = 0$ for some i, then the corresponding inequality is satisfied. On the other hand, all the remaining inequalities can be collapsed to a single one by dividing both sides of all such inequalities by $\alpha_i + \alpha_{i+1}$:

$$\frac{(1 + \rho(\gamma, \mathcal{G}))d}{1 - \xi(\alpha, d, \mathcal{G})} \leq c_1 \frac{n}{\kappa},$$

a sufficient condition for which is:

$$\frac{(1+\rho(\gamma,\mathcal{G}))c}{1-\xi(\alpha,d,\mathcal{G})} \leq c_1. \tag{5}$$

In the remainder of the proof, we show that our particular choice of γ and α satisfy inequality (5) and Theorem 3 condition 1.

We bound $\rho(\gamma,\mathcal{G})$ from above via a change of variable. Setting $\tilde{\gamma} = \gamma_1 \frac{\bar{\mu}^2}{c_1}$, we have that:

$$\begin{aligned}
\rho(\gamma,\mathcal{G}) = \left\| \sum_{i=1}^{m} \gamma_i P_i \right\|_\infty &= \left\| \sum_{i=1}^{m} \Big[\gamma_1 + \sum_{j=1}^{i-1} \big(\frac{c_1(\alpha_j+\alpha_{j+1})}{\bar{\mu}^2} + \epsilon \big) \Big] P_i \right\|_\infty && (6)\\
&\leq \left\| \sum_{i=1}^{m} \Big[\gamma_1 + \sum_{j=1}^{i-1} \frac{c_1(\alpha_j+\alpha_{j+1})}{\bar{\mu}^2} \Big] P_i \right\|_\infty + \sum_{i=1}^{m}\sum_{j=1}^{i-1} \epsilon \left\| P_i \right\|_\infty && (7)\\
&\leq \frac{c_1}{\bar{\mu}^2} \left\| \sum_{i=1}^{m} \Big[\tilde{\gamma} + \sum_{j=1}^{i-1} (\alpha_j+\alpha_{j+1}) \Big] P_i \right\|_\infty + \epsilon c_3 && (8)\\
&\leq \frac{c_1}{\bar{\mu}^2} \sum_{i=1}^{m} \Big[|\tilde{\gamma} + \sum_{j=1}^{i-1} (\alpha_j+\alpha_{j+1})| \left\| P_i \right\|_\infty \Big] + \epsilon c_3 && (9)\\
&\leq c_1 c_2 + \epsilon c_3. && (10)
\end{aligned}$$

Here (7) follows by grouping all terms with ϵ and using the triangle inequality, (8) follows by the change of variables described above and bounding all the terms in the right summand from above by one, and (9) follows from the triangle inequality. We choose the remaining degree of freedom $\tilde{\gamma}$ to eliminate the contribution of the subspace with the highest incoherence parameter in the left summand. Consequently, (10) follows by bounding the infinity norms of the remaining projection matrices from above by $\bar{\mu}^2$. Crucially, the fact that $\alpha \in [0,1]^m$ and m are viewed as fixed enables us to bound the sum from above with positive constants c_2 and c_3 that depend only on m.

Next, we use our particular choice for α to bound $\xi(\alpha,d,\mathcal{G})$ from above. In particular, for any $W \in \mathbb{S}^n$ we have:

$$\begin{aligned}
&\left\| \big[(I - \sum_{i=1}^{m} \alpha_i \mathcal{P}_{ii}) \mathcal{P}_\Omega \big](W) \right\|_\infty \\
&= \left\| \sum_{\substack{i=1\\ i\neq \ell}}^{m} \Big(P_i \mathcal{P}_\Omega(W) + \mathcal{P}_\Omega(W) P_i - P_i \mathcal{P}_\Omega(W) P_i \Big) - \sum_{\substack{i=1\\ i\neq \ell}}^{m} \sum_{\substack{j=1\\ j\neq i,\ell}}^{m} P_i \mathcal{P}_\Omega(W) P_j \right\|_\infty \\
&\leq \sum_{\substack{i=1\\ i\neq \ell}}^{m} \Big(\left\| P_i \mathcal{P}_\Omega(W) \right\|_\infty + \left\| \mathcal{P}_\Omega(W) P_i \right\|_\infty + \left\| P_i \mathcal{P}_\Omega(W) P_i \right\|_\infty \Big) + \sum_{\substack{i=1\\ i\neq \ell}}^{m} \sum_{\substack{j=1\\ j\neq i,\ell}}^{m} \left\| P_i \mathcal{P}_\Omega(W) P_j \right\|_\infty \\
&\leq \left[\sum_{\substack{i=1\\ i\neq \ell}}^{m} 2\mu(\mathcal{E}_i)\sqrt{d} + \left(\sum_{\substack{i=1\\ i\neq \ell}}^{m} \mu(\mathcal{E}_i)^2 + \sum_{\substack{i=1\\ i\neq \ell}}^{m} \sum_{\substack{j=1\\ j\neq i,\ell}}^{m} \mu(\mathcal{E}_i)\mu(\mathcal{E}_j) \right) d \right] \left\| W \right\|_\infty \\
&\leq (c_4 \bar{\mu}\sqrt{d} + c_5 \bar{\mu}^2 d) \left\| W \right\|_\infty
\end{aligned}$$

$$
\begin{aligned}
&= (c_4\sqrt{\frac{\kappa d}{n}} + c_5\frac{\kappa d}{n})\,\|W\|_\infty \\
&\leq (c_4\sqrt{c} + c_5 c)\,\|W\|_\infty\,, \qquad (11)
\end{aligned}
$$

for some positive real numbers c_4 and c_5 depending only on m. Here the first equality is obtained by rearranging the sum in terms of the projection matrices P_i, the first inequality is due to the triangle inequality, and the second inequality is a consequence of the following inequalities:

$$
\begin{aligned}
\|P_i\mathcal{P}_\Omega(W)\|_\infty &\leq \mu(\mathcal{E}_i)\sqrt{d}\,\|W\|_\infty\,, \\
\|\mathcal{P}_\Omega(W)P_i\|_\infty &\leq \mu(\mathcal{E}_i)\sqrt{d}\,\|W\|_\infty\,, \\
\|P_i\mathcal{P}_\Omega(W)P_j\|_\infty &\leq \mu(\mathcal{E}_i)\mu(\mathcal{E}_j)d\,\|W\|_\infty\,;
\end{aligned}
$$

which hold for all $i, j \in \{1, \ldots, m\}$.

Equations (10) and (11) assert that $\rho(\gamma, \mathcal{G})$ and $\xi(\alpha, d, \mathcal{G})$ can lowered as desired by reducing the constants c_1, ϵ and c. Consequently, one can check that both condition 1 of Theorem 3 and equation (5) (which implies condition 2 of Theorem 3) can be satisfied by first choosing a sufficiently small c_1 and ϵ (both depending on m) to bound $\rho(\gamma, \mathcal{G})$ from above, and then suitably choosing a sufficiently small c depending on m, c_1 and ϵ.

4 Numerical Illustrations with Invariants based on Stable Sets and Cuts

In this section we evaluate the utility of two invariant convex sets based on (tractable relaxations of) the inverse of the stability number and the maximum cut value, both of which are described in Section 1.2. Our investigation is via numerical experiments rather than theoretical bounds as in Section 2. The primary reason for this choice is that we do not have a detailed understanding of the face structure of the invariant convex sets considered in this section; in contrast, we have a precise (and convenient for the purposes of analysis) characterization of the geometry of the Schur-Horn orbitope, which played a crucial role in the theoretical results of the previous section. Nonetheless, we pursue a systematic approach in the present section by identifying classes of graphs that are 'brittle' in the sense that deleting / adding a small number of edges results in large changes in their stability number / maximum cut value. Such graph families present excellent examples for which invariant convex sets based on the inverse of stability number and the maximum cut value are particularly well-suited to obtaining useful bounds on the graph edit distance. More broadly, our discussion in this section highlights the larger point that our framework (P) can be tailored to the particular structural properties of the underlying graphs to yield useful lower bounds on the edit distance.

4.1 Constraining the Inverse of the Stability Number

The function $f(A)$ described in Section 1.2 is an efficiently computable lower bound on the inverse of the stability number of a graph, and further it is a concave graph invariant. Consequently, super-level sets of this function provide tractable invariant convex sets that may be employed in our framework (P). Given a graph $\mathcal{G}$, we denote the associated set by $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$:

$$
\mathcal{C}_{\mathcal{IS}(\mathcal{G})} := \{M \in \mathbb{S}^n \mid f(M) \geq f(A)\} = \{M \in \mathbb{S}^n \mid \exists \mu \in \mathbb{S}^n,\ \mu \geq 0,\ I + M - \mu - f(A)11^T \succeq 0\}. \qquad (12)
$$

Here A is any adjacency matrix representing $\mathcal{G}$. From this description, it is immediately clear that for any edit to $\mathcal{G}$ that corresponds to an increase in the value of the function f, the constraint $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ is inactive. Adding edges to a graph can only reduce the stability number, and hence can potentially only increase the inverse of the stability number. Although the function f is only a lower bound on the inverse of the stability number, it satisfies a similar monotonocity property in that the value of f is non-decreasing with the addition of edges to a graph. The following lemma formalizes matters by describing the tangent cone at an adjacency matrix of a graph $\mathcal{G}$ with respect to the set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$:

Lemma 7. *For any graph $\mathcal{G}$ on n vertices and associated adjacency matrix $A \in \mathbb{S}^n$, let $\alpha^* = f(A)$, i.e., the value corresponding to the Motzkin-Straus relaxation of the inverse of the stability number. Then we have that:*

$$\mathcal{T}_{\mathcal{C}_{\mathcal{IS}(\mathcal{G})}}(A) = \{T \in \mathbb{S}^n \,|\, \exists\, \mu,\, \Lambda \in \mathbb{S}^n,\, \mu \geq 0,\, \Lambda \succeq 0,\; T + I + A - \alpha^* 11^T = \mu + \Lambda\}.$$

Proof. The proof of this lemma follows from a direct application of convex duality. □

The description of the tangent cone in Lemma 7 is based on the dual of the cone of doubly nonnegative matrices; see [2] for more details on this connection. In particular, this lemma implies that entrywise nonnegative matrices belong to the tangent cone at an adjacency matrix A representing a graph $\mathcal{G}$ with respect to the set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$; consequently, edits to $\mathcal{G}$ consisting purely of addition of edges are feasible directions with respect to the set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ and for such edits this set does not provide useful lower bounds on the edit distance. Thus, we investigate the utility of the constraint $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ in settings in which the edits consist mainly of edge deletions. Such problems arise in the context of *graph completion* in which the objective is to add edges to a given graph so that the resulting graph satisfies some desired property.

Building on this discussion, the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ is most likely to be useful for graphs $\mathcal{G}$ in settings in which the deletion of even a small number of edges of $\mathcal{G}$ results in an increase in the stability number. Graphs that have a large number of stable sets with cardinality equal to the stability number offer a natural prospect for further exploration. Fortunately, such graphs have been studied in extremal graph theory literature, from which we quote the following result [21]:

Theorem 8. *[21] For $s, n \in \mathbb{N}$ with $n \geq 6$, let*

$$h(n) = \begin{cases} 2 \times 3^{s-1} + 2^{s-1}, & \text{if } n = 3s, \\ 3^s + 2^{s-1}, & \text{if } n = 3s+1, \\ 4 \times 3^{s-1} + 3 \times 2^{s-2}, & \text{if } n = 3s+2. \end{cases}$$

Let $\mathcal{G}$ be any connected graph on n vertices, and denote the cardinality of the set of all maximum independent sets of $\mathcal{G}$ by $\phi(\mathcal{G})$. Then $\phi(\mathcal{G}) \leq h(n)$ with equality if and only if $\mathcal{G}$ is isomorphic to one of the graphs shown in Figure 4.

This theorem states that the graphs $E(n)$ shown in Figure 4 are precisely the connected graphs that have the largest number of distinct maximum independent sets. As such, they present a natural test case to investigate the utility of the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ in providing bounds on the graph edit distance, at least in settings in which the edits are composed predominantly of edge deletions. We illustrate here the results of numerical experiments conducted on the graph $E(30)$, which is a sparse graph with 39 edges and 396 nonedges. The setup of this experiment is the same as that described in Section 2.3 with one notable exception: in the present experiment, we assume asymmetric edits rather than symmetric edits so that 80% of the edits are edge deletions while

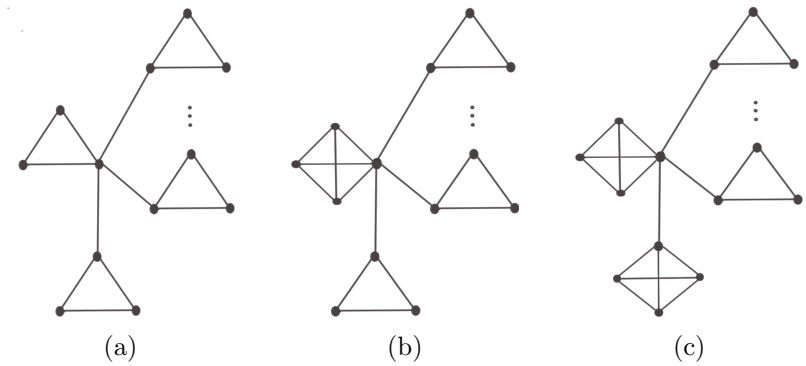


Figure 4: Left to right, $E(n)$ for $n = 3s$, $n = 3s + 1$, $n = 3s + 2$. For $n = 3s + r$, these graphs are formed by connecting $(s - r)\, K_3$'s and $r\, K_4$'s through edges connecting to a specific vertex.

20% are edge additions. We range the total number of edits from 5 to 45 with increments of 5, and for each number of edits we repeat our experiment 1000 times. In each iteration, we obtain a bound on the graph edit distance between $E(30)$ and the modified graph using our framework (P) with three different constraint sets: the Schur-Horn orbitope, the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$, and an invariant convex set based on the Goemans-Williamson relaxation of the maximum cut value (which is discussed in greater detail in the next subsection). Figure 5 reports the ratio of the average computed lower bound on the graph edit distance to the number of edit operations for each constraint set. (The number of edits is an upper bound on the true graph edit distance.) As one might expect, the relaxation based on the constraint $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ yields the best lower bounds of the three approaches. Specifically, even when a majority of the edges of $E(30)$ are removed, the constraint set $\mathcal{C}_{\mathcal{IS}(\mathcal{G})}$ continues to provide lower bounds that are at least 40% of the number of edit operations. In contrast, the bounds provided by the Schur-Horn orbitope constraint are much weaker, and those obtained using the Goemans-Williamson relaxation of the maximum cut value are ineffective.

4.2 Constraining the Maximum Cut Value

In analogy with the inverse of the stability number, the function $g(A)$ due to Goemans and Williamson [14] that is described in Section 1.2 provides an efficiently computable upper bound on the maximum cut value of a graph. As this function is invariant to conjugation of its argument by permutation matrices, its sublevel sets are invariant convex sets. For a graph $\mathcal{G}$, we denote the associated set by $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$:

$$\begin{aligned}\mathcal{C}_{\mathcal{MC}(\mathcal{G})} :=& \{M \in \mathbb{S}^n \mid g(M) \leq g(A)\} \\ =& \{M \in \mathbb{S}^n \mid \exists D \in \mathbb{S}^n \text{ diagonal}, M - D \succeq 0, \frac{1}{4}\operatorname{Tr}(M11' - D) \leq g(A)\},\end{aligned} \tag{13}$$

where A is an adjacency matrix representing $\mathcal{G}$. Reasoning in a similar manner as in the previous subsection, we observe that edits corresponding to a decrease in the value of the function g represent feasible directions with respect to the set $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$, and for such edits the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ is inactive. Deleting edges from a graph reduces its maximum cut value, and one can check that directions represented by entrywise nonpositive matrices belong to the tangent cone at an adjacency matrix A representing $\mathcal{G}$ with respect to $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$. Consequently, we should only expect the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ to potentially provide useful lower bounds on the graph edit distance in settings in which most of the edits to a graph $\mathcal{G}$ correspond to edge additions. In some sense, this type of a graph

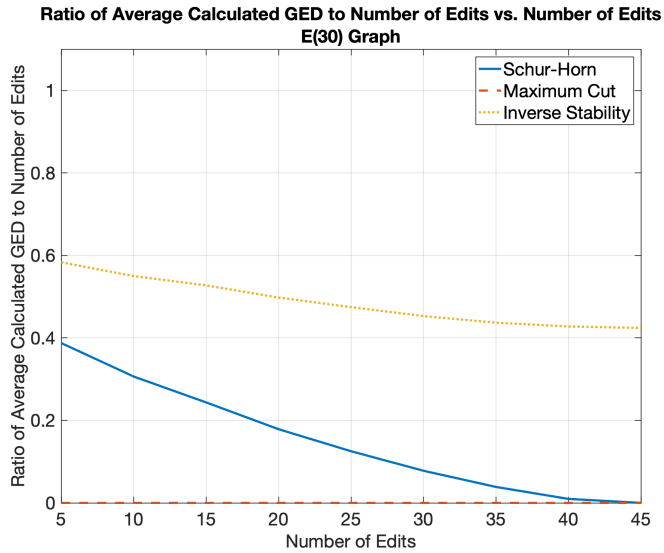


Figure 5: Ratio of average computed lower bound on graph edit distance to number of edit operations. Experiment conducted on $E(30)$ graph. The edit operations are 80% edge deletions and 20% edge additions.

(a) Windmill graph $D(4,7)$

(b) Windmill graph $D(5,5)$

Figure 6: Two sample Windmill graphs

inverse problem – removing the smallest number of edges from a graph so that it satisfies a desired property – is a complement of the graph completion problem discussed in the previous subsection.

Building further on the preceding discussion, we remark that the constraint set $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ is most likely to be effective if adding even a small number of edges to $\mathcal{G}$ increases the value of the function g. A prominent example of such graphs are the so-called Windmill graphs shown in Figure 6. The Windmill graph $D(m,n)$ is constructed by taking m copies of the complete graph K_n and intersecting them at a single vertex. Due to the ample amount of symmetry in these graphs, there are many partitions of the vertices into two sets that achieve the maximum cut value – the number of such partitions is $\binom{n-1}{n/2}^m$ for even n and $\binom{n}{(n-1)/2}^m$ for odd n. Thus, Windmill graphs present a natural test family to evaluate the power of the constraint set $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ when the graph edits consist primarily of the addition of edges. We present the results of numerical experiments on the Windmill graph $D(4,7)$ in a setting that closely mirrors the one in the previous subsection. The Windmill graph $D(4,7)$ is a graph on 25 nodes with 84 edges and 216 non-edges. The edits made to this

graph consist mostly of edge additions – 80% are edge additions and the remaining 20% are edge deletions. We vary the number of edits from 10 to 200 with increments of 5 and consider 1000 random instances of perturbations for each number of edits. For each problem instance, we obtain a lower bound on the edit distance by utilizing our framework (P) with the Schur-Horn orbitope constraint, the Motzkin-Straus relaxation from the previous subsection, and the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$. We report the average ratio of the computed lower bound on the graph edit distance to the number of edit operations in Figure 7. (As before the number of edits is an upper bound on the graph edit distance.)

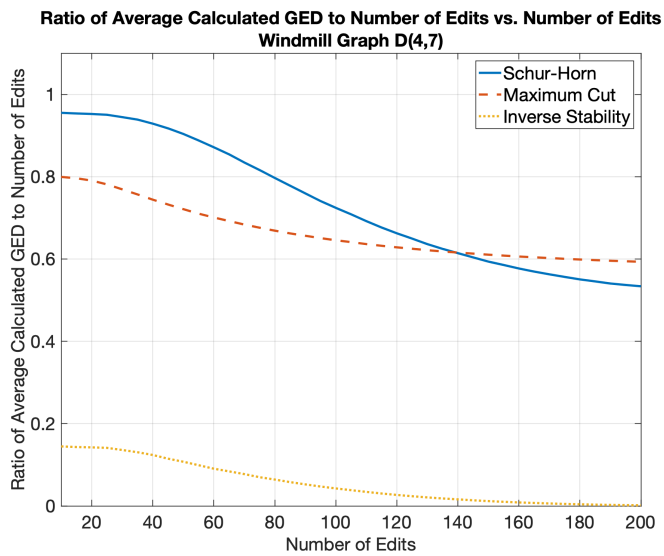


Figure 7: Ratio of average computed lower bound on the graph edit distance to number of edit operations. Experiment conducted on Windmill graph $D(4, 7)$. The edit operations are 80% edge additions and 20% edge deletions.

From Figure 7 we see that the Schur-Horn orbitope constraint produces the best lower bounds for the graph edit distance when the number of edits is small, whereas the constraint $\mathcal{C}_{\mathcal{MC}(\mathcal{G})}$ produces the best lower bounds when the number of edits is large. On average, both of these constraints provide bounds that are consistently better than 50% of the total number of edits. As the edits consist mainly of edge additions, the constraint based on the Motzkin-Straus relaxation of the inverse of the stability number performs poorly.

5 Experiments with Real Data

In this section, we present experimental results that demonstrate the utility of our framework on real data. We begin by introducing an extension of our framework to allow for edits that include vertex additions and deletions. We then describe the bounds obtained on two widely studied datasets consisting of molecular structures.

5.1 Enabling vertex additions and deletions

In many situations, one wishes to obtain bounds on the edit distance between two graphs consisting of different numbers of vertices. In such cases one allows vertex insertions and deletions in addition

to the usual operations of edge insertions and deletions that we've considered thus far. To extend our framework to this setting, we allow an adjacency matrix to take on nonzero values on the diagonal to denote the presence or absence of a vertex. Specifically, we consider a "vertex-indexed adjacency matrix" $A \in \mathbb{S}^n$ with entries equal to either zero or one and in which $A_{ij} = 1,\ j \neq i$ implies that $A_{ii} = 1$. In words, a value of one on the i'th diagonal entry implies that a vertex corresponding to that index is 'present' in the graph, and an edge being incident on a vertex implies that the vertex must be present in the graph. (Note that a value of one on a diagonal entry does not represent a "vertex weight" but instead the presence of a vertex in a graph.) With this notation in hand, we are now in a position to describe a generalization of our framework that allows for vertex deletions and insertions. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two unweighted and unlabeled graphs on n_1 and n_2 vertices, respectively, and let $n := \max\{n_1, n_2\}$. Letting $A_2 \in \mathbb{S}^n$ specify a vertex-indexed adjacency matrix for $\mathcal{G}_2$ with zeros on the diagonal corresponding to those indices that do not correspond to a vertex (when $n_2 < n$), consider the following convex optimization problem:

$$\begin{aligned} \mathrm{GED}_{LB}(\mathcal{G}_1, \mathcal{G}_2; \mathcal{C}_{\mathcal{G}_1}) = \min_{X,E \in \mathbb{S}^n} & \sum_{1 \leq i \leq j \leq n} |E_{ij}| \\ s.t. \quad & X + E = A_2 \\ & X \in \mathcal{C}_{\mathcal{G}_1} \\ & X_{ij} \leq X_{ii},\ X_{ij} \leq X_{jj}\ \forall i,j \in \{1,\ldots,n\}. \end{aligned} \tag{P_{ext}}$$

Here the set $\mathcal{C}_{\mathcal{G}_1}$ is an invariant convex set associated to $\mathcal{G}_1$, and the matrices X, A_1 and A_2 are to be interpreted as vertex-indexed adjacency matrices. There are two main differences between the convex program (P_{ext}) and the convex problem (P). The first is in the objective function in which we only sum the upper triangular elements of the matrix E in the program (P_{ext}), as we do not wish to double-count the edge edits relative to vertex edits. The second modification arises in the constraint in the last line of (P_{ext}) based on the observation that if edges are incident to a vertex, then this vertex must be 'present'. Using a line of reasoning similar to that following the presentation of (P), one can conclude that the optimal value of the convex program (P_{ext}) provides a lower bound on the graph edit distance between $\mathcal{G}_1$ and $\mathcal{G}_2$ with the permissible edit operations being vertex additions/deletions and edge additions/deletions. Finally, we note that our framework can also accommodate situations in which the cost of a vertex edit operation is different from that of an edge edit operation.

5.2 Experimental Results on Chemistry Datasets

We employ the convex program (P_{ext}) to obtain lower bounds on graph edit distance problems arising in chemistry. Specifically, we conduct experiments on two datasets known as the Polycyclic Aromatic Hydrocarbons (PAH) dataset and the Alkane dataset.[1] Both of these datasets consist of unlabeled, unweighted graphs representing chemicals, with the vertices of the graphs corresponding to carbon atoms in a molecule and edges specifying bonds between two carbons. These datasets have been used as benchmarks for evaluating the performances of graph edit distance algorithms; for example, see [1] and [10] for comparisons of the performance of various algorithms on these datasets. For each dataset, we compare upper bounds on the average edit distance taken over all pairs of graphs (obtained using other procedures) with lower bounds on the average obtained using our method.

The Alkane dataset consists of 150 unlabeled, acyclic graphs representing alkanes, with the number of vertices ranging from 1 to 10 vertices (the average is 8.9) and an average degree of

[1] Available online at https://brunl01.users.greyc.fr/CHEMISTRY/

1.8. As these graphs are relatively small in size, the average pairwise graph edit distance for this dataset can be calculated exactly using combinatorial algorithms such as the A^* procedure [17]. The PAH dataset consists of 94 graphs representing polycyclic aromatic hydrocarbons. As with the Alkane dataset, the vertices of the graphs in this dataset denote carbon atoms, and two vertices are connected if there exists a bond between the corresponding carbons. Unlike the Alkane dataset, the chemicals in the PAH dataset represent large compounds: the smallest graph in PAH has 10 vertices, the largest graph has 28 vertices, and the average number of vertices is 20.7. The average degree of the graphs in the PAH dataset is 2.4. Due to this larger size, calculating the exact average pairwise graph edit distance of the PAH dataset is prohibitively expensive. In fact, to the best of our knowledge, the exact average pairwise graph edit distance of the PAH dataset is unknown to this date [3]. Consequently, obtaining guaranteed lower bounds on the average graph edit distance of the PAH dataset is especially useful as a way to compare to known average upper bounds.

For each pair of graphs, we employ the convex program (P_{ext}) twice by switching the roles of $\mathcal{G}_1$ and $\mathcal{G}_2$, and take the larger optimal value as our lower bound. In each case we utilize four different types of invariant convex set constraints: the Schur-Horn orbitope ($\mathcal{C}_{\mathcal{SH}}$), the Motzkin-Straus bound on the inverse of the stability number ($\mathcal{C}_{\mathcal{IS}}$), the Goemans-Williamson bound on the maximum cut value ($\mathcal{C}_{\mathcal{MC}}$), and finally the intersection of all these three constraints ($\mathcal{C}_{\mathcal{SH}} \cap \mathcal{C}_{\mathcal{IS}} \cap \mathcal{C}_{\mathcal{MC}}$). In our experiments, we follow the convention adopted in the graph edit distance literature with these two datasets, namely that the cost of an edit operation is equal to three.[2] The average pairwise lower bounds obtained using our convex program (P_{ext}) on the Alkane and PAH datasets are given in Table 1.

Dataset Name	Best known upper bound on the average GED	Lower bounds on the average pairwise GED via (P_{ext})			
		$\mathcal{C}_{\mathcal{MC}}$	$\mathcal{C}_{\mathcal{IS}}$	$\mathcal{C}_{\mathcal{SH}}$	$\mathcal{C}_{\mathcal{MC}} \cap \mathcal{C}_{\mathcal{IS}} \cap \mathcal{C}_{\mathcal{SH}}$
Alkane	15.3 (exact) [10]	4.66	6.12	9.58	10.72
PAH	29.8 [10]	12.01	14.52	20.29	21.60

Table 1: Average pairwise graph edit distances of the Alkane and PAH datasets. Edit operations are limited to edge and vertex additions and removals. Every edit operation incurs a cost of 3.

There are a number of interesting aspects to these results. For both datasets, a constraint based only on the Goemans-Williamson relaxation seems to produce the worst lower bounds (4.66 and 12.01), while the Schur-Horn orbitope constraint produces the best lower bounds (9.58 and 20.29) when only a single type of invariant convex constraint is employed. As expected, the combination of all three individual constraint sets produces the best overall lower bounds (10.72 and 21.60). More broadly, these results demonstrate the effectiveness of our approach in producing useful lower bounds for graph edit distance problems arising from real data in a computationally tractable manner. Specifically, for the Alkane dataset the average lower bound 10.72 is obtained using our convex programming framework and the exact value of the average graph edit distance is 15.3 (which is obtained via combinatorial approaches). Our results have more interesting implications for the PAH dataset as it is prohibitively expensive to compute the exact average graph edit distance for this dataset due to the large size of its constituents. In particular, the best-known upper bound on the average graph edit distance of PAH is 29.8 [10]. Our convex relaxation framework produces a lower bound of 21.6 on the average graph edit distance over all pairs of graphs in PAH, which

[2]The reason for this choice in that community is that vertex/edge deletions/insertions are considered more significant edit operations than vertex/edge label substitutions which have a lower cost of one associated to them (in this paper, we do not consider such edits based on substitutions.)

provides a floor on the possible improvement that one should expect to obtain via better algorithms for computing graph edit distances.

6 Discussion

In this paper we introduce a framework based on convex graph invariants for obtaining lower bounds on the edit distance between two graphs. Much of the literature on this topic provides methods for computing upper bounds on the edit distance between two graphs by identifying a feasible sequence of edits to transform one graph to the other. Our approach is qualitatively different in that it is based on convex relaxation and it leads to guaranteed lower bounds on the edit distance. Further, our approach can be adapted to the structure underlying the two graphs. We provide both theoretical and empirical support for our method.

There are a number of potential directions for further investigation arising from our paper. First, our analysis of the performance of the Schur-Horn relaxation could potentially be tightened in order to obtain sharper conditions for the success of our algorithm. For example, Corollary 4 only utilizes information about the second most-repeated eigenvalue, and while this provides order-optimal scaling results for families such as triangular graphs, it may be possible to improve our analysis to obtain order-optimal bounds for other families as well. More broadly, a key step in carrying out a precise theoretical analysis of the power of an invariant convex constraint set is to obtain a full understanding of the facial structure of the set, and it would be of interest to develop such a characterization for a larger suite of invariant convex sets than those presented in this paper. Finally, a commonly encountered question in many applications is to test whether a given graph is a minor of another graph, and it would be useful to extend the framework described in our paper to address this problem.

References

[1] Zeina Abu-Aisheh et al. Graph edit distance contest: Results and future challenges. *Pattern Recognition Letters*, 100:96–103, 2017.

[2] Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*, volume 9. SIAM, 1994.

[3] Sébastien Bougleux et al. Graph edit distance as a quadratic assignment problem. *Pattern Recognition Letters*, 87:38–46, 2017.

[4] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, 2009.

[5] Utkan O. Candogan and Venkat Chandrasekaran. Finding planted subgraphs with few eigenvalues using the Schur–Horn relaxation. *SIAM Journal on Optimization*, 28(1):735–759, 2018.

[6] Venkat Chandrasekaran, Pablo A. Parrilo, and Alan S. Willsky. Convex graph invariants. *SIAM Review*, 54(3):513–541, 2012.

[7] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[8] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Constantine Caramanis. Low-rank matrix recovery from errors and erasures. *IEEE Trans. Inf. Theory*, 59(7):4324–4337, 2013.

[9] Donatello Conte, Pasquale Foggia, Carlo Sansone, and Mario Vento. Thirty years of graph matching in pattern recognition. *Int. J. Pattern Recognit. Artif. Intell*, 18(03):265–298, 2004.

[10] Évariste Daller, Sébastien Bougleux, Benoit Gaüzère, and Luc Brun. Approximate graph edit distance by several local searches in parallel. In *7th International Conference on Pattern Recognition Applications and Methods*, 2018.

[11] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[12] Andreas Fischer, Ching Y. Suen, Volkmar Frinken, Kaspar Riesen, and Horst Bunke. A fast matching algorithm for graph-based handwriting recognition. In *International Workshop on Graph-Based Representations in Pattern Recognition*, pages 194–203. Springer, 2013.

[13] Michael R. Garey and David S. Johnson. *Computers and intractability*, volume 29. W.H. Freeman New York, 2002.

[14] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[15] Michael C. Grant and Stephen P. Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx, September 2013.

[16] Michael C. Grant and Stephen P. Boyd. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, pages 95–110. Springer, 2008.

[17] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[18] Rashid Ibragimov, Maximilian Malek, Jiong Guo, and Jan Baumbach. GEDEVO: an evolutionary graph edit distance algorithm for biological network alignment. In *OASIcs-OpenAccess Series in Informatics*, volume 34. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2013.

[19] D. K. Isenor and Safwat G. Zaky. Fingerprint identification using graph matching. *Pattern Recognition*, 19(2):113–122, 1986.

[20] Tao Jiang, Guohui Lin, Bin Ma, and Kaizhong Zhang. A general edit distance between RNA structures. *Journal of Computational Biology*, 9(2):371–388, 2002.

[21] Min-Jen Jou and Gerard J. Chang. The number of maximum independent sets in graphs. *Taiwanese Journal of Mathematics*, 4(4):685–695, 2000.

[22] Derek Justice and Alfred Hero. A binary linear programming formulation of the graph edit distance. *IEEE Trans. Pattern Anal. Mach. Intell*, 28(8):1200–1214, 2006.

[23] Marius Leordeanu, Martial Hebert, and Rahul Sukthankar. An integer projected fixed point method for graph matching and MAP inference. In *Advances in Neural Information Processing Systems*, pages 1114–1122, 2009.

[24] Zhi-Yong Liu and Hong Qiao. GNCCP- graduated nonconvexity and concavity procedure. *IEEE Trans. Pattern Anal. Mach. Intell*, 36(6):1258–1267, 2014.

[25] Si Wei Lu, Ying Ren, and Ching Y. Suen. Hierarchical attributed graph representation and recognition of handwritten Chinese characters. *Pattern Recognition*, 24(7):617–632, 1991.

[26] Vesna Memišević and Nataša Pržulj. C-GRAAL: Common-neighbors-based global GRAph ALignment of biological networks. *Integrative Biology*, 4(7):734–743, 2012.

[27] Theodore S. Motzkin and Ernst G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canad. J. Math*, 17(4):533–540, 1965.

[28] Michel Neuhaus and Horst Bunke. An error-tolerant approximate matching algorithm for attributed planar graphs and its application to fingerprint classification. In *SSPR/SPR*, pages 180–189. Springer, 2004.

[29] Michel Neuhaus and Horst Bunke. A graph matching based approach to fingerprint classification using directional variance. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 191–200. Springer, 2005.

[30] Michel Neuhaus and Horst Bunke. Edit distance-based kernel functions for structural pattern classification. *Pattern Recognition*, 39(10):1852–1863, 2006.

[31] Kaspar Riesen and Horst Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision Computing*, 27(7):950–959, 2009.

[32] Kaspar Riesen, Andreas Fischer, and Horst Bunke. Computing upper and lower bounds of graph edit distance in cubic time. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pages 129–140. Springer, 2014.

[33] Alberto Sanfeliu and King-Sun Fu. A distance measure between attributed relational graphs for pattern recognition. *IEEE Trans. Syst., Man, Cybern.*, SMC-13(3):353–362, 1983.

[34] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. Orbitopes. *Mathematika*, 57(2):275–314, 2011.

[35] Roded Sharan et al. Conserved patterns of protein interaction in multiple species. *Proceedings of the National Academy of Sciences*, 102(6):1974–1979, 2005.

[36] Roded Sharan and Trey Ideker. Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, 24(4):427, 2006.

[37] Kim-Chuan Toh, Michael J. Todd, and Reha H. Tütüncü. SDPT3 - a MATLAB software package for semidefinite programming, version 1.3. *Optimization Methods and Software*, 11(1-4):545–581, 1999.

[38] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph Von Der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell*, 19(7):775–779, 1997.

[39] Zhiping Zeng, Anthony K.H. Tung, Jianyong Wang, Jianhua Feng, and Lizhu Zhou. Comparing stars: On approximating graph edit distance. *Proc. of the VLDB Endowment*, 2(1):25–36, 2009.