Trust your data or not - StQP remains StQP: Community Detection via Robust Standard Quadratic Optimization*

Immanuel M. Bomze[1], Michael Kahr[1] and Markus Leitner[1,2]

[1]University of Vienna, Department of Statistics and Operations Research, Austria.
{immanuel.bomze|m.kahr|markus.leitner}@univie.ac.at
[2]Department of Supply Chain Analytics, Vrije Universiteit Amsterdam, The Netherlands.
m.leitner@vu.nl

September 9, 2019

Abstract

We consider the Robust Standard Quadratic Optimization Problem (RStQP), in which an uncertain (possibly indefinite) quadratic form is optimized over the standard simplex. Following most approaches, we model the uncertainty sets by balls, polyhedra, or spectrahedra, more generally, by ellipsoids or order intervals intersected with sub-cones of the copositive matrix cone. We show that the copositive relaxation gap of the RStQP equals the minimax gap under some mild assumptions on the curvature of the aforementioned uncertainty sets, and present conditions under which the RStQP reduces to the Standard Quadratic Optimization Problem. These conditions also ensure that the copositive relaxation of an RStQP is exact. The theoretical findings are accompanied by the results of computational experiments for a specific application from the domain of graph clustering, more precisely, community detection in (social) networks. The results indicate that the cardinality of communities tend to increase for ellipsoidal uncertainty sets and to decrease for spectrahedral uncertainty sets.

1 Introduction

Robust optimization allows to guarantee feasibility of solutions when uncertainty affects problem parameters. This approach has gained remarkable attention during the last decades, most likely triggered by the influential papers of Ben-Tal and Nemirovski [3, 4, 5]. Another reason for the attractivity is that robust approaches (see, e.g., [9, 16, 20, 22] for comprehensive surveys) do not presume knowledge about the probability distribution of the uncertain data realizations, but only about bounds on them that define so-called *uncertainty sets*. However, there is typically a

*Supported by the Vienna Science and Technology Fund (project ICT15-014), and the Austrian Science Fund (project P 26755-N26).

trade-off between the chosen uncertainty sets and the computational tractability of optimization problems in which uncertain data realizations are considered [3]. For instance, in linear optimization computational tractability translates from the uncertainty set to the robust problem. By contrast, this is not true for conic programs in which even simple uncertainty sets typically increase the computational effort needed to solve robust optimization problems significantly [6, p.152ff].

It is frequently assumed that uncertainty only affects constraints [6, 22], sometimes exploiting the epigraph formulation for the objective function, but there are, however, important cases in which one can safely trust the constraints. In particular this is the case for the *Standard Quadratic Optimization Problem (StQP)* [10] in which a (possibly indefinite) quadratic form is optimized over the standard simplex

$$\Delta^n \coloneqq \{\mathbf{x} \in \mathbb{R}^n_+ : \mathbf{e}^\mathsf{T}\mathbf{x} = 1\},$$

in the n-dimensional Euclidean space $\mathbb{R}^n$. Indeed, thinking of probability distributions over $[1:n]$, the integer interval $\{1,\ldots,n\}$, it is widely accepted that probabilities are nonnegative and sum up to one. Here we already used some notation: $\mathbb{R}^n_+ \coloneqq \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i \in [1:n]\}$ denotes the nonnegative orthant in $\mathbb{R}^n$ and the vector of all-ones is denoted by $\mathbf{e} \coloneqq [1,\ldots,1]^\mathsf{T} \in \mathbb{R}^n_+$. Moreover, we will use $\mathbf{e}_i$ to denote the i-th column vector of the identity matrix I. Notice that throughout this article we reserve small bold-faced letters for column vectors (e.g., the all-zeros vector $\mathbf{o}$) with x_i referring to the i-th coordinate of a vector $\mathbf{x}$; and bold-faced, capital letters for matrices (e.g., the all-zeros matrix O). The set of *symmetric matrices* of order n is denoted by $\mathcal{S}^n \coloneqq \{\mathsf{A} \in \mathbb{R}^{n\times n} : \mathsf{A} = \mathsf{A}^\mathsf{T}\}$. For some $\mathsf{Q} \in \mathcal{S}^n$, the StQP is defined by

$$\max_{\mathbf{x}\in\Delta^n} \ \mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x}\,. \tag{1}$$

Despite its simplicity, the StQP is quite versatile and has numerous applications from different domains [12], e.g., finance (Markowitz portfolio selection), economics (evolutionary game-theoretic algorithms), ecology (replicator dynamics), machine learning (image analysis), and graph theory (graph clustering). In all these fields, uncertain input data may arise quite naturally. Robust problem variants are studied, for instance, in the context of Markowitz portfolio selection in [17, 18, 24, 37], where the objective functions are convex over Δ^n. Other applications, now from the domain of graph clustering and with possibly indefinite quadratic forms as objective functions, include the *(Robust) Maximum (Vertex-)Weighted Clique Problem*. Although it deviates a bit from the community detection scope of the current paper, this particular problem class serves as a good illustration of the potential of the approach we propose. We will therefore shortly discuss this graph optimization problem in Section 4 before turning to the main application in Section 5. Both graph clustering problems involve non-convexity in the objective function.

Mittal et al. [25] study robust quadratic programs with an uncertain *convex* objective function. They showed that copositive reformulations of such programs always exist when assuming a bounded mixed-integer polyhedral uncertainty set and developed an approximation method based on semidefinite relaxations which is demonstrated on instances of well-known problems such as the least squares problem. While their approach is therefore constrained to convex optimization as opposed to the general nonconvex case (with graph clustering applications, for instance), the method proposed here could also be applied to convex minimization, but of course multiplicity of local solutions is impossible in this context; which brings us to the main motivation for this paper:

Based upon this seemingly novel approach, Section 5 discusses a method for social network community detection based on the similarity among community members, which can only be roughly quantified and which therefore is naturally subject to uncertainty. The proposed method (i) allows to extract communities without partitioning the whole network, and (ii) allows for multiple community memberships, i.e., individuals can belong to several different communities which is naturally the case in social networks; cf. [29]. We remark that the desired multiplicity of communities is only possible if the problem is non-convex, which rules out the application of most of the procedures from existing literature mentioned above.

Already the deterministic (classical) version of StQP is NP-hard [23] and one would expect that problem variants with uncertain input data are even more difficult to solve in practice. In this article we show how to (possibly) overcome these computational issues and identify easy to check conditions under which an StQP with uncertain objective function parameters (see Definition 1) can be solved as a single deterministic StQP. The above-mentioned conditions are met by many, if not most, of the uncertainty set families that are popular in the current literature.

Definition 1. (Robust Standard Quadratic Optimization Problem (RStQP)). Let the nominal problem data be given by $\mathsf{Q} \in \mathcal{S}^n$, which is possibly affected by uncertain additive perturbations U ranging within an *uncertainty set* $\mathcal{U} \subset \mathcal{S}^n$. Then the RStQP is defined by

$$\max_{\mathbf{x}\in\Delta^n} \min_{\mathsf{U}\in\mathcal{U}} \ \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x}\,. \tag{2}$$

Remark 1. While we will discuss the RStQP in its max-min form in this article, all results can be easily transformed to the analogous min-max form. Alternatively, each StQP instance in maximization form can be solved in minimization form by appropriate manipulations of Q, see, e.g., [14]. Indeed this transformation also applies for the RStQP by manipulating set $\mathcal{U}$ accordingly. Moreover, we will desist from discussing general quadratic problems of the form $\mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x}$ over Δ^n, since they can always be homogenized by replacing Q in (1) with the rank-two update $\mathsf{Q}' = \mathsf{Q} + \mathbf{e}\mathbf{c}^\mathsf{T} + \mathbf{c}\mathbf{e}^\mathsf{T}$ in which case the objective values of the aforementioned general quadratic problem and the one of the StQP coincide [13]. We can also solve general robust quadratic problems over Δ^n with the RStQP by manipulating set $\mathcal{U}$ accordingly in (2) such that the uncertainty set for Q' is compatible with the uncertainty model for $(\mathsf{Q}, \mathbf{c})$. However, as in our main application, i.e., community detection in (social) networks, we deal with the homogeneous case in the first place, we will not pursue this issue further in this paper.

Contribution and outline. In the remainder of this section, we introduce further notation and terminology, as well as the completely positive (CP) relaxation of the RStQP. In Section 2 we show that the gap between the RStQP and its corresponding CP relaxation is equal to the minimax gap of von Neumann's theorem [36], and show that this gap is closed under some reasonable assumptions on the uncertainty sets. In Section 3 we show that the RStQP reduces to an StQP for several (commonly used) uncertainty sets, while in Section 4 we derive a relaxation of the RStQP that avoids the minimax problem structure for the case of polyhedral uncertainty sets. Moreover, in a first application to the *Robust Maximum (Vertex-)Weighted Clique Problem* we show how in particularly structured problems we again can close the gap by our methods. In Section 5 we turn to the main motivation of this study and apply our findings to the *Dominant Set Clustering Problem* [28, 30] which aims to identify homogeneous clusters in networks such as communities formed by similar individuals in social networks. We also provide new insights for the nominal and robust problem variants. The results of our computational experiments that accompany our theoretical findings are discussed in Section 6.

1.1 Further notation and terminology; lifting

We denote the *Frobenius norm* of a square matrix $\mathsf{A} \in \mathbb{R}^{n\times n}$ by $\|\mathsf{A}\|_{\mathrm{F}} \coloneqq (\sum_i \sum_j a_{ij}^2)^{1/2}$, where a_{ij} is the entry in the i-th row and j-th column of A.

We proceed with a lifting for problem (2) into a higher dimensional space in which the objective function is linear, and which goes back to Shor [33]. The aim hereby is to obtain the CP relaxation of (2) as will be seen later. Let the *trace* of matrix $\mathsf{A} \in \mathbb{R}^{n\times n}$ be denoted by $\mathrm{Tr}(\mathsf{A}) \coloneqq \sum_i a_{ii}$. Moreover, $\langle \mathsf{A}, \mathsf{B}\rangle \coloneqq \mathrm{Tr}(\mathsf{A}^\mathsf{T}\mathsf{B}) = \sum_i \sum_j a_{ij} b_{ij}$ denotes the *Frobenius inner product* of matrices A and B. Then observe that the objective funtion of (2) can be rearranged as

$$\mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x} = \mathrm{Tr}(\mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x}) = \mathrm{Tr}(\mathsf{Q}\mathbf{x}\mathbf{x}^\mathsf{T}) = \langle \mathsf{Q}, \mathbf{x}\mathbf{x}^\mathsf{T}\rangle = \langle \mathsf{Q}, \mathsf{X}\rangle\,, \tag{3}$$

where we define the matrix variable $\mathsf{X} \coloneqq \mathbf{x}\mathbf{x}^\mathsf{T}$ and consider it as a point in $\mathbb{R}^d$ with $d = \binom{n+1}{2}$.

For an arbitrary cone $\mathcal{K} \subseteq \mathbb{R}^d$ we use '$\preceq_\mathcal{K}$' to denote the *Löwner ordering with respect to* $\mathcal{K}$, meaning that $\mathsf{L} \preceq_\mathcal{K} \mathsf{R}$ if and only if $(\mathsf{R} - \mathsf{L}) \in \mathcal{K}$. Moreover, we denote a set which is induced by such a conic ordering by

$$[\mathsf{L}, \mathsf{R}]_\mathcal{K} \coloneqq \left\{\mathsf{U} \in \mathbb{R}^d : \mathsf{L} \preceq_\mathcal{K} \mathsf{U} \preceq_\mathcal{K} \mathsf{R}\right\}.$$

The dual cone of $\mathcal{K}$ is defined by

$$\mathcal{K}^* \coloneqq \left\{\mathbf{y} : \langle \mathbf{x}, \mathbf{y}\rangle \geq 0 \;\text{ for all }\; \mathbf{x} \in \mathcal{K}\right\}.$$

To introduce the most prominent example of a pair of dual cones in this paper, let $\mathrm{conv}\{\cdot\}$ denote the convex hull of a set, i.e., the smallest convex set containing that set. Then the cone of *completely positive matrices* (see, e.g, [8]) is defined by

$$\mathcal{CP}_n \coloneqq \mathrm{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} : \mathbf{x} \in \mathbb{R}^n_+\right\} = \left\{\mathsf{X} \in \mathcal{S}^n : \mathsf{X} = \sum_{i=1}^k \mathbf{y}_i\mathbf{y}_i^\mathsf{T} \;\text{ for some }\; \mathbf{y}_i \in \mathbb{R}^n_+\right\}$$

and its dual cone is

$$\mathcal{COP}_n \coloneqq \left\{\mathsf{Q} \in \mathcal{S}^n : \langle \mathsf{Q}, \mathsf{X}\rangle \geq 0 \;\text{ for all }\; \mathsf{X} \in \mathcal{CP}_n\right\} = \left\{\mathsf{Q} \in \mathcal{S}^n : \mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x} \geq 0 \;\text{ for all }\; \mathbf{x} \in \mathbb{R}^n_+\right\},$$

i.e., the cone of *copositive matrices.* Another familiar example, this time of a self-dual cone, is the cone of *positive semidefinite matrices* obtained by dropping the nonnegativity requirement in the definition of $\mathcal{CP}_n$, that is,

$$\mathcal{P}_n \coloneqq \mathrm{conv}\left\{\mathbf{x}\mathbf{x}^\mathsf{T} : \mathbf{x} \in \mathbb{R}^n\right\} = \left\{\mathsf{Q} \in \mathcal{S}^n : \mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x} \geq 0 \;\text{ for all }\; \mathbf{x} \in \mathbb{R}^n\right\}.$$

A further self-dual cone is the set of all symmetric matrices with no negative entries, $\mathcal{N}_n \coloneqq \mathcal{S}^n \cap \mathbb{R}^{n\times n}_+$. Combinations of the latter two self-dual cones will give approximations of $\mathcal{CP}_n$ and $\mathcal{COP}_n$ respectively which are dual to each other: the *doubly nonnegative (DNN) cone* $\mathcal{P}_n \cap \mathcal{N}_n \supset \mathcal{CP}_n$ and the *nonnegative decomposable (NND) cone* $\mathcal{P}_n + \mathcal{N}_n \subset \mathcal{COP}_n$.

By denoting the matrix of all-ones $\mathsf{E} \coloneqq \mathbf{e}\mathbf{e}^\mathsf{T} \in \mathcal{S}^n$ we define the *lifted standard simplex* (slightly abusing notation) by

$$\boldsymbol{\Delta}^{n\times n} \coloneqq \left\{\mathsf{X} \in \mathcal{CP}_n : \langle \mathsf{E}, \mathsf{X}\rangle = 1\right\},$$

which is a compact, convex set with infinitely many extremal points $\{\mathbf{x}\mathbf{x}^\mathsf{T} : \mathbf{x} \in \Delta^n\}$. Finally, using Equation (3), we can formulate the completely positive relaxation of the RStQP (in the lifted space) as a bilinear minimax problem over the set $\mathcal{U} \times \mathbf{\Delta}^{n\times n}$ by

$$\max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathsf{X}\rangle \,. \tag{4}$$

Notice, however, that although the objective function of (4) is bilinear, the problem difficulty did not vanish but is "hidden" in the constraint $\mathsf{X} \in \mathbf{\Delta}^{n\times n} \subset \mathcal{CP}_n$. The nominal case (which can be embedded by putting $\mathcal{U} = \{\mathsf{O}\}$ in the current setting) satisfies

$$\max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \langle \mathsf{Q}, \mathsf{X}\rangle = \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x}\,. \tag{5}$$

Thus, the deterministic counterpart of (4) is a conic reformulation of (1), a completely positive optimization problem. The conic dual is a copositive (COP) optimization problem of the form

$$\min\left\{y\in\mathbb{R} : y\mathsf{E}-\mathsf{S}=\mathsf{Q}\,,\, \mathsf{S}\in\mathcal{COP}_n\right\}, \tag{6}$$

which corresponds to identifying the smallest y such that $y\mathsf{E}-\mathsf{Q}$ is copositive. For this problem class, the conic duality gap and the relaxation gap are both zero, as shown by Bomze et al. [13]. By contrast, in the general case there is a relaxation gap which exactly corresponds to the gap in the minimax inequality, even if the conic duality gap is zero (e.g., because of a Slater condition for the conic problems). We will deal with this and further related results in Section 2.

2 CP relaxation and minimax gap

First we establish the fact that the *completely positive relaxation* (4) gives an upper bound on (2) and therefore deserves its name. We will refer to the difference between (4) and (2) as the *CP relaxation gap*. Furthermore, we will, under mild assumptions on the uncertainty set $\mathcal{U}$, establish an alternative interpretation of the relaxation value, and specify two conditions ensuring a zero gap.

We will also often use von Neumann's minimax theorem [36] which states that

$$\max_{\mathbf{x}\in\mathcal{X}} \min_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x},\mathbf{y}) = \min_{\mathbf{y}\in\mathcal{Y}} \max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x},\mathbf{y}). \tag{7}$$

if $\mathcal{X}\subset\mathbb{R}^n$, $\mathcal{Y}\subset\mathbb{R}^m$ are compact convex sets, and $f:\mathcal{X}\times\mathcal{Y}\mapsto\mathbb{R}$ is a continuous function such that $f(\cdot,\mathbf{y}):\mathcal{X}\mapsto\mathbb{R}$ is concave while $f(\mathbf{x},\cdot):\mathcal{Y}\mapsto\mathbb{R}$ is convex.

2.1 CP relaxation gap is minimax gap

Theorem 1. ***(CP relaxation gap is minimax gap).***

(a) For general $\mathcal{U}$, we have

$$\max_{\mathbf{x}\in\Delta^n} \min_{\mathsf{U}\in\mathcal{U}} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} \;\le\; \max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathsf{X}\rangle.$$

(b) Suppose $\mathcal{U}$ is closed and convex. Then

$$\max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathsf{X}\rangle \;=\; \min_{\mathsf{U}\in\mathcal{U}} \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x},$$

so that the CP relaxation gap is exactly the gap in the minimax inequality.

Proof. (a) follows directly by lifting $\mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} = \langle \mathsf{Q}+\mathsf{U}, \mathbf{x}\mathbf{x}^\mathsf{T}\rangle$ since $\{\mathbf{x}\mathbf{x}^\mathsf{T} : \mathbf{x} \in \Delta^n\} \subset \mathbf{\Delta}^{n\times n}$. To establish (b), we apply the already mentioned fact that for nominal StQP, the CP relaxation is exact, using the minimax theorem (7) for the lifted (linear) formulation:

$$\min_{\mathsf{U}\in\mathcal{U}} \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} = \min_{\mathsf{U}\in\mathcal{U}} \max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \langle \mathsf{Q}+\mathsf{U}, \mathsf{X}\rangle = \max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathsf{X}\rangle.$$

Hence, the result follows. □

By now, the possible non-concavity of the objective function $\mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x}$ in $\mathbf{x}$ suggests that relaxation (4) is in general not an exact reformulation of the RStQP (2), contrasting the case of the traditional StQP (with $\mathcal{U} = \{\mathsf{O}\}$). Somehow surprisingly, it turns out that some simple conditions on $\mathcal{U}$ in fact ensure that the CP relaxation of the RStQP is exact and, moreover, that the minimax theorem holds, despite the possible non-concavity.

2.2 Linear dependence closes gap

In this subsection we discuss the RStQP for arbitrary uncertainty sets, however, assuming that the worst-case data realizations and the nominal data are linearly dependent. Notice that such uncertainty sets occur naturally when the uncertain data realizations are estimated to lie within a *relative* tolerance with respect to the nominal data, e.g., within a tolerance of ±5%. Notation $\mathsf{L} = \lambda\mathsf{Q}$ is used in Theorem 2 to be consistent with theorems discussed later.

Theorem 2. ***(Linear dependence closes gap).*** *Let $\mathcal{U}$ be an arbitrary uncertainty set such that for all $\mathsf{X} \in \mathbf{\Delta}^{n\times n}$ we have*

$$\min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathsf{X}\rangle = \langle \mathsf{Q}+\mathsf{L}, \mathsf{X}\rangle \quad \text{for all } \mathsf{X} \in \mathbf{\Delta}^{n\times n} \tag{8}$$

where $\mathsf{L} = \lambda\mathsf{Q}$ for some suitable $\lambda \in \mathbb{R}$. Then the RStQP (2) reduces to an StQP equivalent to the nominal one (with data Q). Hence, its CP relaxation is exact and the minimax theorem holds.

Proof. Applying (8) to $\mathsf{X} = \mathbf{x}\mathbf{x}^\mathsf{T} \in \mathbf{\Delta}^{n\times n}$ for arbitrary $\mathbf{x} \in \Delta^n$, we get

$$\max_{\mathbf{x}\in\Delta^n} \min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathbf{x}\mathbf{x}^\mathsf{T}\rangle = \max_{\mathbf{x}\in\Delta^n} \langle \mathsf{Q}+\lambda\mathsf{Q}, \mathbf{x}\mathbf{x}^\mathsf{T}\rangle = (1+\lambda) \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x}\,,$$

a single StQP which has the same solutions as the nominal instance $\max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}\mathsf{Q}\mathbf{x}$. Using the general minimax inequality, we arrive at

$$\begin{aligned}
\min_{\mathsf{U}\in\mathcal{U}} \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} &\geq \max_{\mathbf{x}\in\Delta^n} \min_{\mathsf{U}\in\mathcal{U}} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} \\
&= \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\lambda\mathsf{Q})\mathbf{x} \\
&\geq \min_{\mathsf{U}\in\mathcal{U}} \max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x}\,.
\end{aligned}$$

So all inequalities above are in fact equalities, and the proof is complete. □

Remark 2. Note that we impose no curvature assumptions on $\mathsf{U}-\mathsf{L}$ for the minimax theorem to hold. By contrast, we will now assume that for some cone $\mathcal{K}$, the relation $(\mathsf{U}-\mathsf{L}) \in \mathcal{K}$ holds for all $\mathsf{U} \in \mathcal{U}$, which relaxes concavity in $\mathbf{x}$ and still ensures zero minimax gap.

2.3 COP-minimality closes gap.

Lemma 3. *Let $\mathcal{K} \subseteq \mathcal{COP}_n$ be a sub-cone of the cone of copositive matrices. Let $\{\mathsf{L}, \mathsf{U}\} \subset \mathcal{S}^n$ be given matrices such that $\mathsf{L} \preceq_{\mathcal{K}} \mathsf{U}$. Then $\langle \mathsf{L}, \mathsf{X} \rangle \leq \langle \mathsf{U}, \mathsf{X} \rangle$ for all $\mathsf{X} \in \mathcal{CP}_n$.*

Proof. Indeed $\mathsf{L} \preceq_{\mathcal{K}} \mathsf{U}$ is equivalent to $(\mathsf{U} - \mathsf{L}) \in \mathcal{K} \subseteq \mathcal{COP}_n$ and this implies $\langle \mathsf{U} - \mathsf{L}, \mathsf{X} \rangle \geq 0$ for all $\mathsf{X} \in \mathcal{CP}_n$. □

This elementary observation has the following consequences [35]:

Theorem 4. ***(COP-minimality closes gap).*** *Let $\mathcal{K} \subseteq \mathcal{COP}_n$ be a sub-cone of the cone of copositive matrices. Suppose that $\mathcal{U}$ contains a matrix $\mathsf{L} \in \mathcal{U}$ such that $\mathsf{L} \preceq_{\mathcal{K}} \mathsf{U}$ for all $\mathsf{U} \in \mathcal{U}$. Then the minimax theorem holds for the RStQP (2) and its CP relaxation (4) is exact.*

Proof. Observe that given $\mathsf{X} \in \mathbf{\Delta}^{n \times n}$, the inner minimization problem of (4) only depends on the realizations U. Using Lemma 3 it follows by $\mathbf{\Delta}^{n \times n} \subset \mathcal{CP}_n$ that

$$\min_{\mathsf{U} \in \mathcal{U}} \langle \mathsf{Q} + \mathsf{U}, \mathsf{X} \rangle = \langle \mathsf{Q} + \mathsf{L}, \mathsf{X} \rangle .$$

Repeating the argument of Theorem 2 completes the proof. □

Remark 3. Uncertainty sets constructed according to Theorem 4 also ensure that the copositive relaxation of the RStQP is exact, since strong duality holds for the StQP [11] (and its CP/COP relaxations). Hence,

$$\max_{\mathsf{X} \in \mathbf{\Delta}^{n \times n}} \min_{\mathsf{U} \in \mathcal{U}} \langle \mathsf{Q} + \mathsf{U}, \mathsf{X} \rangle = \min \left\{ y \in \mathbb{R} : y\mathsf{E} - \mathsf{S} = \mathsf{Q} + \mathsf{L}, \mathsf{S} \in \mathcal{COP}_n \right\} .$$

3 Trust your data or not – StQP remains StQP

In this section we will use the results from Section 2 to derive conditions under which the RStQP reduces to a single (nominal) StQP. In particular we discuss uncertainty sets $\mathcal{U}$ that allow to solve the inner minimization problem of (2) in closed form and, hence, do not increase the problem's difficulty compared to an StQP.

3.1 Conic uncertainty

We proceed with discussing uncertainty sets composed by the intersection of cones, which include the frequently investigated *box* and *spectrahedral* uncertainty sets (see, e.g., [7]), respectively, as special cases.

Corollary 1. ***(Conic uncertainty with zero CP relaxation gap).*** *Let $\mathcal{K} \subseteq \mathcal{COP}_n$ be a sub-cone of the cone of copositive matrices and $\mathsf{L}, \mathsf{R} \in \mathcal{S}^n$ be given matrices. Assume that $\mathcal{U} = [\mathsf{L}, \mathsf{R}]_{\mathcal{K}}$. Then the CP relaxation (4) is an exact reformulation of (2) and the RStQP reduces to an StQP (1) with problem data $\mathsf{Q} + \mathsf{L}$.*

Proof. Follows directly from Theorem 4. □

The results from above allow to construct (infinitely) many conic uncertainty sets while ensuring that the resulting RStQP can be solved as a nominal StQP of the same dimension. These include the most important cones in conic optimization, and the aforementioned, frequently considered uncertainty sets (i.e., *box-uncertainty* or *spectrahedral uncertainty*) as special cases. To only list a few of such conic uncertainty sets with the key property $\mathsf{L} \in \mathcal{U}$, take one of the following options for $\mathcal{U}$, along with their self-explanatory names:

(i) $[\mathsf{L},\mathsf{R}]_{\mathcal{CP}_n}$ *CP-uncertainty,*

(ii) $[\mathsf{L},\mathsf{R}]_{\mathcal{N}_n}$ *box-uncertainty,*

(iii) $[\mathsf{L},\mathsf{R}]_{\mathcal{P}_n}$ *spectrahedral uncertainty,*

(iv) $[\mathsf{L},\mathsf{R}]_{\mathcal{N}_n \cap \mathcal{P}_n}$ *DNN-uncertainty,*

(v) $[\mathsf{L},\mathsf{R}]_{\mathcal{N}_n + \mathcal{P}_n}$ *NND-uncertainty,* and

(vi) $[\mathsf{L},\mathsf{R}]_{\mathcal{COP}_n}$ *COP-uncertainty.*

Remark 4. Despite the fact that there is no efficient separation oracle for uncertainty sets (i) and (vi), we nonetheless end up with a robust problem that reduces to a nominal StQP of the same size. We further remark that the set $\mathcal{U} = [\mathsf{L},\mathsf{R}]_{\mathcal{K}}$ need not even be convex if $\mathcal{K}$ is not convex.

3.2 Ellipsoidal uncertainty

We will now to discuss *ellipsoidal uncertainty* sets which have also received much attention in robust optimization.

Theorem 5. ***(Ellipsoidal uncertainty).*** *Let* C *be an* $n \times k$ *matrix of full row rank (thus* $k \geq n$ *and* $(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}$ *exists) and define, for some scalar* $\rho > 0$*, the uncertainty set* $\mathcal{U} = \{\mathsf{U} \in \mathcal{S}^n : \|\mathsf{C}^\mathsf{T}\mathsf{U}\mathsf{C}\|_{\mathrm{F}} \leq \rho\}$*. Then*

$$\max_{\mathbf{x}\in\Delta^n} \min_{\mathsf{U}\in\mathcal{U}} \ \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} = \max_{\mathsf{X}\in\boldsymbol{\Delta}^{n\times n}} \ \langle \mathsf{Q} - \rho(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}, \mathsf{X}\rangle \, , \tag{9}$$

i.e., the RStQP reduces to an instance of an StQP, cf. (5)*.*

Proof. For any $\mathbf{x} \in \Delta^n$ let $\mathbf{y} := \mathsf{C}^\mathsf{T}(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}\mathbf{x} \in \mathbb{R}^k \setminus \{\mathbf{o}\}$ so that $\mathsf{C}\mathbf{y} = \mathbf{x}$. Then for any $\mathsf{U} \in \mathcal{U}$ we have via the Cauchy-Schwarz inequality

$$\mathbf{x}^\mathsf{T}\mathsf{U}\mathbf{x} = \mathbf{y}^\mathsf{T}(\mathsf{C}^\mathsf{T}\mathsf{U}\mathsf{C})\mathbf{y} = \langle \mathsf{C}^\mathsf{T}\mathsf{U}\mathsf{C}, \mathbf{y}\mathbf{y}^\mathsf{T}\rangle \geq -\|\mathsf{C}^\mathsf{T}\mathsf{U}\mathsf{C}\|_{\mathrm{F}} \|\mathbf{y}\mathbf{y}^\mathsf{T}\|_{\mathrm{F}} \geq -\rho \|\mathbf{y}\mathbf{y}^\mathsf{T}\|_{\mathrm{F}} = -\rho \|\mathbf{y}\|^2 \, . \tag{10}$$

Note that the Cauchy-Schwarz inequality above holds with equality if and only if $\mathsf{C}^\mathsf{T}\mathsf{U}\mathsf{C} = \lambda\mathbf{y}\mathbf{y}^\mathsf{T}$ for some scalar $\lambda \leq 0$. So choosing $\mathsf{U} = \mathsf{U}_\mathbf{x} := \lambda\mathbf{u}\mathbf{u}^\mathsf{T}$ for $\mathbf{u} := (\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}\mathbf{x}$ and $\lambda := -\rho/\|\mathbf{y}\|^2 < 0$ renders equality in (10) and $\mathsf{U}_\mathbf{x} \in \mathcal{U}$ since $\|\mathsf{C}^\mathsf{T}\mathsf{U}_\mathbf{x}\mathsf{C}\|_{\mathrm{F}} = \rho$ by elementary algebraic manipulations. The common value of all expressions in (10) equals

$$\mathbf{x}^\mathsf{T}\mathsf{U}_\mathbf{x}\mathbf{x} = -\rho\|\mathbf{y}\|^2 = -\rho \ \mathbf{x}^\mathsf{T}(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}\mathbf{x} \, .$$

Hence for all $\mathbf{x} \in \Delta^n$, we have $\min_{\mathsf{U}\in\mathcal{U}} \mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{U})\mathbf{x} = \mathbf{x}^\mathsf{T}(\mathsf{Q} - \rho(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1})\mathbf{x}$, and the result follows. □

Remark 5. By defining $\mathsf{L} := -\rho(\mathsf{C}\mathsf{C}^\intercal)^{-1}$, the statement above is equivalent to the result derived in Theorem 4, i.e.,

$$\max_{\mathbf{x}\in\Delta^n}\min_{\mathsf{U}\in\mathcal{U}} \ \mathbf{x}^\intercal(\mathsf{Q}+\mathsf{U})\mathbf{x} = \max_{\mathbf{x}\in\Delta^n} \ \mathbf{x}^\intercal(\mathsf{Q}+\mathsf{L})\mathbf{x}\,.$$

However, for $n > 1$ in case of ellipsoidal uncertainty, this L *never* belongs to $\mathcal{U}$, in contrast to the setting of Theorem 4. Indeed we have $\|\mathsf{C}^\intercal\mathsf{L}\mathsf{C}\|_{\mathrm{F}} = \rho\sqrt{n} > \rho$. As a special case, if C is an orthogonal square matrix, that is, $\mathsf{C}\mathsf{C}^\intercal = \mathsf{I}$, the ellipsoidal uncertainty set corresponds to a *Frobenius ball*, the simplest ellipsoid.

4 Polyhedral uncertainty

As discussed in Section 2, the gap of the CP relaxation (4) is equal to the minimax gap. This limits the applicability of conic reformulations of the RStQP for arbitrary uncertainty sets. Thus, the computational effort needed to solve instances of the RStQP may be significantly higher than the one of a single StQP.

However, as will be demonstrated below, one can obtain an approximation that avoids the minimax structure of the RStQP when considering polyhedral uncertainty sets. The cost incurred by tractable approximations comes from two gaps, the minimax and the approximation gap. The latter originates when replacing $\mathcal{CP}_n$ with a tractable cone, e.g. the DNN cone $\mathcal{P}_n \cap \mathcal{N}_n$.

4.1 General polyhedral case: approximations.

First we reformulate the RStQP as a quadratic optimization problem under nonconvex, quadratic and linear constraints:

Theorem 6. *Let* $\mathcal{U} = \{\mathsf{U} \in \mathcal{S}^n : \langle \mathsf{A}_i, \mathsf{U}\rangle \le b_i, \ \mathsf{A}_i \in \mathcal{S}^n, \ \text{for all} \ i \in [1:m]\}$ *be a nonempty polyhedral uncertainty set. Then*

$$\max_{\mathbf{x}\in\Delta^n}\min_{\mathsf{U}\in\mathcal{U}} \mathbf{x}^\intercal(\mathsf{Q}+\mathsf{U})\mathbf{x} = \max_{(\mathbf{x},\mathbf{y})\in\Delta^n\times\mathbb{R}^m_+} \left\{\mathbf{x}^\intercal\mathsf{Q}\mathbf{x} - \mathbf{b}^\intercal\mathbf{y} : -\sum_{i=1}^m y_i\mathsf{A}_i = \mathbf{x}\mathbf{x}^\intercal\right\}. \tag{11}$$

Proof. By assumption, the inner minimization problem of the RStQP is a linear program, hence strong duality holds. If $\mathbf{y} \in \mathbb{R}^m_+$ are dual variables associated with the constraints in the definition of $\mathcal{U}$, dualizing yields (11). □

Theorem 6 does not only enable the application of all methods for treating (nonconvex) all-quadratic problems, but also opens an avenue of getting tractable bounds:

Theorem 7. *Let* $\mathcal{U}$ *be as in Theorem 6. Then we have the following upper bounds for the RStQP in terms of linear conic optimization problems:*

$$\begin{aligned}\max_{\mathbf{x}\in\Delta^n}\min_{\mathsf{U}\in\mathcal{U}} \mathbf{x}^\intercal(\mathsf{Q}+\mathsf{U})\mathbf{x} &= \max_{\mathbf{x}\in\Delta^n}\min_{\mathsf{U}\in\mathcal{U}} \langle \mathsf{Q}+\mathsf{U}, \mathbf{x}\mathbf{x}^\intercal\rangle \\ \le \quad & \max_{(\mathsf{X},\mathbf{y})\in\boldsymbol{\Delta}^{n\times n}\times\mathbb{R}^m_+} \left\{\langle \mathsf{Q},\mathsf{X}\rangle - \mathbf{b}^\intercal\mathbf{y} : -\sum_{i=1}^m y_i\mathsf{A}_i = \mathsf{X}\right\} \end{aligned} \tag{12}$$

$$\leq \max_{(\mathsf{X},\mathbf{y})\in\mathcal{N}_n\cap\mathcal{P}_n\times\mathbb{R}^m_+} \left\{\langle \mathsf{Q},\mathsf{X}\rangle - \mathbf{b}^\mathsf{T}\mathbf{y} : -\sum_{i=1}^m y_i \mathsf{A}_i = \mathsf{X}\,,\ \langle \mathsf{E},\mathsf{X}\rangle = 1\right\}. \tag{13}$$

Proof. The first inequality is obvious by (11) via Shor lifting, and the second by the inclusion $\mathcal{CP}_n \subseteq \mathcal{P}_n \cap \mathcal{N}_n$. □

Inequality (12) corresponds exactly to the CP relaxation gap. Indeed, note that dualizing the lifted formulation in (12) by associating a dual variable $\mathsf{U} \in \mathcal{S}^n$ to the constraint yields

$$\max_{\mathsf{X}\in\mathbf{\Delta}^{n\times n}} \min_{\mathsf{U}\in\mathcal{S}^n} \ \{\langle \mathsf{Q}+\mathsf{U},\mathsf{X}\rangle : \langle \mathsf{A}_i,\mathsf{U}\rangle \leq b_i,\ \text{ for all }\ i\in[1:m]\}\,,$$

and a similar statement holds for the tractable approximation (13).

4.2 An application with tight approximation: the robust maximum weighted clique problem.

In some relevant application areas, the structure of problem uncertainty allows to obtain the same results as in Section 2.3 in the sense that all gaps can be closed with our approach. The so-called *Maximum (Vertex-)Weighted Clique Problem (MWCP)* considers simple undirected graphs $G = (V, E)$ with vertex set V of finite order $|V| = n$ and edge-set $E \subseteq V \times V$. A *clique* S is a vertex-set $S \subseteq V$ such that its induced subgraph G_S is complete. A clique S is *maximal* if there exists no other clique S' such that $S \subset S'$ and a *maximum* clique has the largest cardinality among all (maximal) cliques. The well-known Motzkin/Straus theorem [26] relates the search for the maximum cliques (a combinatorial optimization problem) to a suitable StQP (a continuous optimization problem).

In the MWCP we are given *vertex weights* $\mathbf{w} = (w_i)_{i\in V} \in \mathbb{R}^n_+$, and search for the clique $S \subseteq V$ with the largest total weight $W(S) := \sum_{i\in S} w_i$. This combinatorial optimization problem can again be recast as an StQP. A predecessor of this generalization of the Motzkin/Straus theorem has been introduced in [21]. This approach suffers from the possible existence of spurious StQP solutions which cannot be mapped into heavy cliques in the graph. Therefore we employ the improved formulation from [10, Thm.7]. For simplicity of exposition and without much loss of generality assume that $w_i \geq 1$ holds for all $i \in V$ (positive integer weights can, e.g., be always achieved by scaling with the common denominator of any set of positive rational weights, without changing the outcome). For a vertex-weighted instance $(G, \mathbf{w})$, we consider the matrix $\mathsf{A}(G, \mathbf{w})$ given by

$$a_{ij}(G,\mathbf{w}) := \left\{\begin{array}{ll} 1-\frac{1}{2w_i}, & \text{if } i=j\in V\,, \\ 1, & \text{if } \{i,j\}\in E \text{ and} \\ 1-\frac{1}{2w_i}-\frac{1}{2w_j}, & \text{otherwise;} \end{array}\right\} \tag{14}$$

derived from transforming the minimizing problem from [10, (5.16),(5.18)] into a maximization problem as treated here. Then all local maximizers $\mathbf{x}_S$ to the StQP $\max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}\mathsf{A}(G,\mathbf{w})\mathbf{x}$ are strict, and their supports S form maximal cliques of total weight $W(S) = [2(1-\mathbf{x}_S^\mathsf{T}\mathsf{A}(G,\mathbf{w})\mathbf{x}_S)]^{-1}$. Hence a global maximizer $\mathbf{x}^*$ of that StQP yields a maximum vertex weight clique $S^* = \{i : x_i^* > 0\}$. Therefore the quality of the local solutions in terms of objective are directly correlated to the weight of the corresponding cliques, and, speaking from the robust perspective, we must hedge against the worst case corresponding to an instance with small objective values.

We follow the robustness approach in the literature on robust MWCP, see, e.g. [34],[1] in that

[1] This paper deals with independent sets which are cliques in the complement graph.

the uncertainty is captured by variation in the weights. As we want to apply our methods, we therefore have to consider additive uncertainty with some structure. This can be done as follows: suppose we only know some bounds ℓ_i and r_i for w_i, i.e., $1 \le \ell_i \le w_i \le r_i$ for all $i \in V$ and take any value $\bar{w}_i$ inside this interval as the nominal one, giving rise to a nominal instance $\mathsf{A}(G, \bar{\mathbf{w}})$. The additive uncertainty now must be given by a matrix U such that the weights $\mathbf{w}$ corresponding to the uncertain instance $\mathsf{A}(G, \bar{\mathbf{w}}) + \mathsf{U}$ satisfy $\ell_i \le w_i \le r_i$ for all $i \in V$, and moreover respect the restrictions in (14) for all $w_i \in [\ell_i, r_i]$, giving rise to the uncertainty set

$$\mathcal{U}(G) := \left\{ \mathsf{U} \in \mathcal{S}^n : \begin{array}{lcll} u_{ii} & \in & [\frac{1}{2\bar{w}_i} - \frac{1}{2\ell_i}, \frac{1}{2\bar{w}_i} - \frac{1}{2r_i}] & \text{for all } i \in V\,, \\ u_{ij} & = & 0 & \text{for all } \{i,j\} \in E\,, \\ u_{ij} & = & u_{ii} + u_{jj} & \text{for all } \{i,j\} \notin E \text{ with } i \neq j\,. \end{array} \right\} \tag{15}$$

This set $\mathcal{U}(G)$ is clearly a polyhedron, namely the intersection of a linear subspace with a box uncertainty interval $[\mathsf{L}(G), \mathsf{R}(G)]_{\mathcal{N}^n}$ with $\mathsf{L}(G) := \mathsf{A}(G, \boldsymbol{\ell}) - \mathsf{A}(G, \bar{\mathbf{w}})$ and $\mathsf{R}(G) := \mathsf{A}(G, \mathbf{r}) - \mathsf{A}(G, \bar{\mathbf{w}})$, and it contains a copositive-minimal element $\mathsf{L}(G) \in \mathcal{U}(G)$.

By consequence, the robust MWCP can be recast as an StQP as shown in Section 2.3. Since a detailed empirical investigation of this approach goes beyond the scope of this paper which is focusing on community detection, we leave it as a future research topic. The above discussion only serves as an illustration that polyhedral uncertainty sets sometimes can have, by virtue of the application model, a beneficial structure allowing for efficient reformulation of the robust counterpart. In light of above discussion, even a (good) local, non-global solution to the resulting StQP $\max\limits_{\mathbf{x} \in \Delta^n} \mathbf{x}^{\mathsf{T}} \mathsf{A}(G, \boldsymbol{\ell}) \mathbf{x}$ would return a clique with a high total weight in an adverse environment.

5 Robust dominant set clustering

In this section we address the *Robust Dominant Set Clustering Problem* which has been proposed in its nominal form by Pavan [27] who established relations between *dominant sets* and optima of an StQP, arriving at a yet different generalization of the Motzkin/Straus theorem. Contrasting with the setting of Section 4.2 above, we here are given *edge weights* rather than vertex weights in a simple undirected graph $G = (V, E)$, i.e., a weight function $w : E \mapsto \mathbb{R}_+$.

5.1 Description and definition of the nominal problem

The *Dominant Set Clustering Problem* aims to cluster *similar* vertices of an undirected, weighted graph $(G, w) = (V, E, w)$ with a given weight function $w : E \mapsto \mathbb{R}_+$. An edge weight $w(i,j)$ reflects the similarity between vertices i and j for each $\{i,j\} \in E$. These edge weights are also used to define the weighted adjacency matrix $\mathsf{A} \in \mathcal{N}_{|V|}$ in which positive values are assigned to all pairs of adjacent vertices and zero otherwise, i.e., $a_{ij} = w(i,j)$ for all $\{i,j\} \in E$, $a_{ij} = 0$ for all $\{i,j\} \notin E$, and $a_{ii} = 0$ for all $i \in V$.

Dominant sets can informally be defined as sets containing entities that are *internally alike* but *externally different.* Thus, elements of a dominant set are dissimilar to elements outside that set. The following formal definition of dominant sets follows [27, 28]. For a non-empty subset $S \subseteq V$ let the *average weighted degree* of vertex $i \in S$ with respect to S be defined by $\mathrm{awdeg}_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij}$. For all $j \notin S$ let $\phi_S(i,j) = a_{ij} - \mathrm{awdeg}_S(i)$. Intuitively, $\phi_S(i,j)$ measures the similarity of i and j

with respect to the average similarity of i to its neighbors within S. The main idea behind dominant set clustering is to associate weights (*similarity values with respect to sets*) recursively defined via

$$w_S(i) = \begin{cases} 1 & \text{if } |S| = 1, \\ \sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) \cdot w_{S \setminus \{i\}}(j) & \text{otherwise,} \end{cases}$$

to vertices $i \in S \subseteq V$. Contrasting with Section 4.2, the total weight of a set S is now given by $W(S) = \sum_{i \in S} w_S(i)$ and intuitively, dominant sets are composed of vertices with similar weights; see Definition 2 for a formal definition.

Definition 2. (Dominant sets).
A non-empty set of vertices $S \subseteq V$ is a dominant set if

(i) $W(T) > 0$ for all $T \subseteq S$, $T \neq \emptyset$,

(ii) $w_S(i) > 0$ for all $i \in S$, and

(iii) $w_{S \cup \{i\}}(i) < 0$ for all $i \notin S$.

By analogy to the theorem of Motzkin and Straus [26], Pavan [27] showed that dominant sets correspond to strict local maximizers $\mathbf{x}'$ of $\max_{\mathbf{x} \in \Delta^{|V|}} \ \mathbf{x}^\mathsf{T} \mathsf{A} \mathbf{x}$. More precisely, if $\mathbf{x}'$ is a strict local maximizer, then $S = \sigma(\mathbf{x}') \coloneqq \{i \in V : x'_i > 0\}$ is a dominant set, if $w_{S \cup \{i\}}(i) \neq 0$ for all $i \in V \setminus S$. Conversely, any dominant set S coincides with the support $\sigma(\mathbf{x}')$ of a strict local maximizer $\mathbf{x}'$. We refer to [30] for a recent, comprehensive review on dominant set clustering, solution algorithms, and applications from several domains including bioinformatics, computer vision, climatology, and medical data analysis.

Some intuition (based on [30]) is provided in Figure 1 in which the concept of dominant sets is sketched for instances based on a complete graph. In Figure 1a, a dominant set is formed by vertices $\{i, j, l\}$, since they are equally similar to each other, while vertex k is quite different due to its weak connection to all other vertices. Conversely, in Figure 1b the comparably strong similarity between k and its neighbors induces a dominant set of cardinality four despite the dissimilarity of i, j, and l. Figure 1c shows that dominant sets need not be disjoint since vertex j belongs to dominant sets $\{i, j, k\}$ and $\{j, l, m\}$. Finally, Figure 1d depicts a similar situation as Figure 1a. Notice that although set $\{i, l, m\}$ is "more similar" (i.e., stronger connected) than set $\{i, j, l\}$ in Figure 1a it is not dominant. This also illustrates that similarity is a *relative* measure that is *immune against scaling* but *sensitive to translation*, i.e., additive perturbations on the instance matrix A.

We also observe that the dominant sets in Figure 1 form cliques. Indeed, we will show in Proposition 9 that this always holds true. Cliques corresponding to dominant sets need, however, not be maximal as this would be the case in the MWCP and in the *Maximum Edge-Weighted Clique Problem* [1] which aims to identify cliques of maximum edge-weight. Since maximal cliques are not necessarily internally alike, they need not form dominant sets.

The following results relate the notion of a dominant set to the structure of the underlying graph.

Proposition 8. ***(Connectedness).*** *Let* $\mathbf{x}'$ *be a locally optimal solution to* $\max_{\mathbf{x} \in \Delta^{|V|}} \ \mathbf{x}^\mathsf{T} \mathsf{A} \mathbf{x}$ *with* $\mathsf{A} \in \mathcal{N}_n \setminus \{\mathsf{O}\}$*, and* $\sigma(\mathbf{x}') = \{i \in V : x'_i > 0\}$*. Then the graph induced by node set* $\sigma(\mathbf{x}')$ *is connected.*

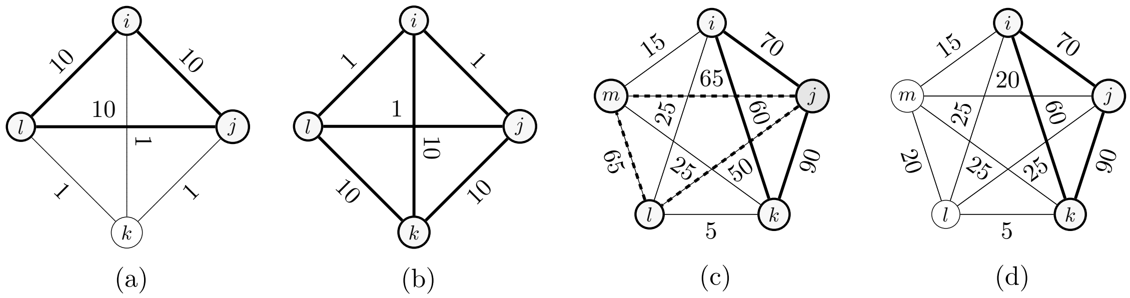

Figure 1: Dominant sets in a complete graph. Vertices connected with bold edges (of the line style) belong to a dominant set.

Proof. Without loss of generality suppose $\sigma(\mathbf{x}') = [1\!:\!q+p]$ and $\mathbf{x}' = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{o}]^\top$ such that $\mathbf{x}_1 \in \mathbb{R}^q_+$ and $\mathbf{x}_2 \in \mathbb{R}^p_+$ are both strictly positive vectors. From $\mathbf{e}^\top\mathbf{x}' = 1$ we gather that $\mathbf{x}'$ is a convex combination

$$\mathbf{x}' = (1-\lambda)\mathbf{u} + \lambda\mathbf{v} \quad \text{with} \qquad \{\mathbf{u}, \mathbf{v}\} = \{ \begin{bmatrix} \frac{\mathbf{x}_1}{\mathbf{e}^\top\mathbf{x}_1} \\ \mathbf{o} \\ \mathbf{o} \end{bmatrix}, \begin{bmatrix} \mathbf{o} \\ \frac{\mathbf{x}_2}{\mathbf{e}^\top\mathbf{x}_2} \\ \mathbf{o} \end{bmatrix} \} \subset \Delta^n \quad \text{with} \quad \lambda = \mathbf{e}^\top\mathbf{x}_2 \in (0,1)\,.$$

Since $\mathbf{x}'$ is a local maximizer of $f(\mathbf{x}) = \mathbf{x}^\top\mathsf{A}\mathbf{x}$ over Δ^n, also the univariate quadratic function $g(t) = f((1-t)\mathbf{u} + t\mathbf{v})$ on $t \in [0,1]$ is locally maximized at $t = \lambda$. Therefore we must have $\dot{g}(\lambda) = 0$ and $\ddot{g}(t) = \ddot{g}(\lambda) = (\mathbf{v}-\mathbf{u})^\top\mathsf{A}(\mathbf{v}-\mathbf{u}) \le 0$ for all t. Now, arguing by contradiction, assume

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_1 & \mathsf{O} & * \\ \mathsf{O} & \mathsf{A}_2 & * \\ * & * & * \end{bmatrix} \in \mathcal{N}_n \setminus \{\mathsf{O}\}$$

in which $\mathsf{A}_1 \in \mathcal{N}_q$, $\mathsf{A}_2 \in \mathcal{N}_p$, and asterisks * denote arbitrary block matrices of appropriate dimensions. The assumed structure of A, $\mathbf{u}$ and $\mathbf{v}$ yields

$$(\mathbf{v}-\mathbf{u})^\top\mathsf{A}(\mathbf{v}-\mathbf{u}) = \frac{1}{(1-\lambda)^2}\,\mathbf{x}_1^\top\mathsf{A}_1\mathbf{x}_1 + \frac{1}{\lambda^2}\,\mathbf{x}_2^\top\mathsf{A}_2\mathbf{x}_2 \ge 0\,, \tag{16}$$

hence, $\ddot{g}(t) = 0$ for all t so that g must be affine-linear in t. But since we have equality to zero in (16) as argued just before, we obtain $\mathbf{x}_i^\top\mathsf{A}_i\mathbf{x}_i = 0$ for both $i \in \{1,2\}$, hence $g(0) = f(\mathbf{u}) = 0 = f(\mathbf{v}) = g(1)$, hence also $f(\mathbf{x}') = g(\lambda) = 0$. Local optimality of $\mathbf{x}'$ over Δ^n implies $0 \le f(\mathbf{y}) \le f(\mathbf{x}') = 0$ for all $\mathbf{y} \in \Delta^n$ close to $\mathbf{x}'$ and then finally for all $\mathbf{y} \in \Delta^n$, since f is quadratic. This is absurd if $\mathsf{A} \ne \mathsf{O}$. □

Proposition 9. ***(Clique condition).*** *Every dominant set of a graph G forms a clique.*

Proof. Consider the weighted adjacency matrix A of G and let $\mathbf{x}'$ be a *strict local maximizer* of $\mathbf{x}^\top\mathsf{A}\mathbf{x}$ over Δ^n. Then $\{i,j\} \subseteq \sigma(\mathbf{x}')$ implies $x'_i x'_j > 0$. Let $t \ne 0$ with $|t|$ small enough and consider a slight perturbation of the maximizer $\mathbf{x}'$ of the form $\mathbf{x}(t) = \mathbf{x}' + t(\mathbf{e}_j - \mathbf{e}_i)$. Arguing as in the proof of Proposition 8, strict local optimality of $\mathbf{x}'$ implies the second-order condition $(\mathbf{e}_i - \mathbf{e}_j)^\top\mathsf{A}(\mathbf{e}_i - \mathbf{e}_j) = \ddot{g}(0) < 0$ for the quadratic function $g(t) = f(\mathbf{x}(t))$, i.e., $a_{ii} + a_{jj} - 2a_{ij} < 0$. Thus, $a_{ij} > 0$ holds (even if $a_{ii} \ge 0$ and $a_{jj} \ge 0$) and, hence, G contains an edge $\{i,j\} \in E$. □

5.2 Robust dominant set clustering

Adapting the notation of the RStQP slightly we define the *Robust Dominant Set Clustering Problem* by

$$\max_{\mathbf{x}\in\Delta^{|V|}} \min_{\mathsf{U}\in\mathcal{U}} \mathbf{x}^\mathsf{T}(\mathsf{A}+\mathsf{U})\mathbf{x}, \tag{17}$$

in which $\mathsf{A} \in \mathcal{N}_{|V|}$ is the weighted adjacency matrix of graph $G = (V, E, w)$, and edge weights are uncertain. Such uncertainty may arise if the appropriateness of a certain similarity measure is fuzzy. This is, for instance, the case when vertex similarities are used in social network analysis to identify communities. The choice of a particular similarity measure may be a further source of uncertainty. To this end, observe that there exists a large body of literature dedicated to the investigation of such measures; see, e.g., [19, 32].

Now, if the minimax theorem holds (e.g., if the uncertainty sets are constructed as described in Section 3) and if the maximizer $\mathbf{x}'$ of $\mathbf{x}^\mathsf{T}(\mathsf{Q}+\mathsf{L})\mathbf{x}$ is strict, then $S = \sigma(\mathbf{x}')$ is a clique for the graph with adjacency matrix $\mathsf{A} = \mathsf{Q}+\mathsf{L}$. However, it need not be a clique for the graph underlying the nominal instance Q. Whenever $a_{ij} > 0$ implies $q_{ij} > 0$ we can use Proposition 8 to show connectedness under Q. This will be the case if uncertainty sets $\mathcal{U}$ are chosen such that they respect the graph structure of the nominal instance. One such possibility is discussed in Section 2.2.

6 Computational experiments

This section discusses the results of our computational study. The results of the previous sections allow us to avoid using (tractable) approximations of the copositive reformulation of RStQP. Instead, we use an algorithm called *Infection Immunization Dynamics* [31] which quickly identifies a local solution for any starting point of the iteration; moreover, it is guaranteed that this procedure finds a dominant set after finitely many steps, see Section 6.2 for details. This algorithm has been implemented in MATLAB 2016b and each experiment has been performed on a single core of an Intel Xeon E5-2670v2 machine with 2.5 GHz and 8 GB RAM.

Besides reporting usual quantitative results such as numbers of identified dominant sets, their sizes, and corresponding runtimes, we also investigate the influence of different uncertainty sets and their sizes on (the structure of) the obtained solutions. All results will be compared to the nominal case, whose results will be provided in Section 6.3 after describing our benchmark instances and the used solution algorithm in Sections 6.1 and 6.2, respectively. The uncertain case is discussed in Section 6.4 in which the results for ellipsoidal uncertainty sets are given relative to box uncertainty sets $\mathcal{U} = [\mathsf{L}, \mathsf{B}]_{\mathcal{N}_n}$ for the sake of comparability.

6.1 Instance description

We use benchmark instances from the 10$^{\text{th}}$ *DIMACS Implementation Challenge* (on graph partitioning and graph clustering) [2] and similarity measures (which are detailed later), i.e., weight-functions $w^{(k)} : E \mapsto [0,1]^{|E|}$, to construct K similarity graphs $G^{(k)} = (V, E, w^{(k)})$ and weighted adjacency matrices $\mathsf{A}^{(k)}$ such that $a_{ij}^{(k)} := w_{ij}^{(k)}$ for all $k \in [1 : K]$ and $\{i, j\} \in E$. We assume that none of these similarity measures reflects the "true" similarity between two vertices $i, j \in V$ but use them to construct conic uncertainty sets as follows. From all similarity measures, i.e., weight-functions

$w^{(1)}, \ldots, w^{(K)}$, we construct weighted adjacency matrices $\mathsf{A}^{(1)}, \ldots, \mathsf{A}^{(K)}$ and derive *worst-case* data realizations $\mathsf{A}^{\min}$ by

$$[\mathsf{A}^{\min}]_{ij} = \min_{k \in [1:K]} [\mathsf{A}^{(k)}]_{ij}, \qquad \text{for all } \ i, j \in V, \tag{18}$$

i.e, by taking the entry-wise minimum data realizations over all similarity measures K. Moreover, we construct the nominal instance data by

$$\mathsf{A} = \frac{1}{K} \sum_{k=1}^{K} \mathsf{A}^{(k)}, \tag{19}$$

i.e., by averaging weight-functions $w^{(k)}$ across all $k \in [1:K]$. The conic uncertainty sets $\mathcal{U} = [\mathsf{L}, \mathsf{B}]_{\mathcal{N}_n}$ which we use as reference for ellipsoidal uncertainty sets are constructed such that $\mathsf{L} := \mathsf{A}^{\min} - \mathsf{A}$. Therefore, $\mathsf{A}^{\min} = \mathsf{A} + \mathsf{L}$ and $\mathsf{A}^{\min} \leq \mathsf{A} + \mathsf{U}$ holds entrywise for all $\mathsf{U} \in \mathcal{U}$. Notice that we could define matrix B similarly to (18) by taking the entry-wise maximum. This is, however, not necessary for the considered uncertainty sets which we detail in Section 6.4.

In particular we use three well-known (vertex-)similarity measures, i.e., $K = 3$, based on the common neighbors of vertices $i, j \in V$ defined by sets $|N(i) \cap N(j)|$. For each $i \in V$, $N(i) = \{j \in V \mid \{i, j\} \in E\}$ is its set of adjacent nodes, and the vertex similarities are defined as

$$w_{ij}^{(k)} = \frac{|N(i) \cap N(j)|}{d^{(k)}}, \ \text{ for all } \ i, j \in V, \ \ k \in [1:3].$$

Denominators $d^{(k)}$ are defined by $d^{(1)} = |N(i) \cup N(j)|$ for the *Jaccard-index*, $d^{(2)} = \sqrt{|N(i)||N(j)|}$ for the *Salton-index*, and $d^{(3)} = \min\{|N(i)|, |N(j)|\}$ for the *hub-promoted-index*; see, e.g., [38]. All similarity values have been scaled by a factor of 100.

6.2 Description of the solution algorithm

Infection Immunization Dynamics (InImDyn) [31] which is based on evolutionary game theory is used for all computations. The derived solutions satisfy the following conditions (20) and (21).

$$\begin{aligned} [\mathsf{A}\mathbf{x}]_i &= \mathbf{x}^\mathsf{T} \mathsf{A}\mathbf{x}, \ \text{ for all } \ i \in \sigma(\mathbf{x}), \\ [\mathsf{A}\mathbf{x}]_i &\leq \mathbf{x}^\mathsf{T} \mathsf{A}\mathbf{x}, \ \text{ for all } \ i \notin \sigma(\mathbf{x}). \end{aligned} \tag{20}$$

Conditions (20) refer (in a game-theoretic sense) to a *Nash strategy* $\mathbf{x} \in \Delta^n$ for a given payoff matrix A. To guarantee that Nash strategy $\mathbf{x} \in \Delta^n$ realizes a strict (local) maximum of the RStQP it has to be *evolutionary stable*, that is, it satisfies the condition

$$\mathbf{y}^\mathsf{T} \mathsf{A}\mathbf{x} = \mathbf{x}^\mathsf{T} \mathsf{A}\mathbf{x} \text{ and } \mathbf{y} \in \Delta^n \quad \implies \quad \mathbf{y}^\mathsf{T} \mathsf{A}\mathbf{y} < \mathbf{x}^\mathsf{T} \mathsf{A}\mathbf{y}. \tag{21}$$

InImDyn is a first-order method of the Frank-Wolfe-type with away steps, for which convergence with correct support separation (manifold identification) even in finitely many iterations has been shown; see [15] for a recent generalization. We will apply it here to the settings of Theorem 4 with the worst-case scenario $\mathsf{A}' = \mathsf{A} + \mathsf{L}$ and Theorem 5 with the choice $\mathsf{A}' = \mathsf{A} - \rho(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}$ for ellipsoidal uncertainty. We treat both as a particular instance of an StQP (note that we can offset the latter choice $\mathsf{A}' = \mathsf{A} - \rho(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}$ by a suitably large positive constant to ensure $\mathsf{A}' \in \mathcal{N}_n$ without changing the solutions). For the readers' convenience we specify an algorithmic scheme in Algorithm 1.

Data: Data matrix $\mathsf{A}' \in \mathcal{N}_n$, starting point $\mathbf{x} \in \Delta^n$, stopping tolerance $\tau = 10^{-9}$.
Result: (Local) optimum $\mathbf{x}$ of the problem $\max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}\mathsf{A}'\mathbf{x}$.
while $\epsilon(\mathbf{x}) > \tau$ **do**
  $\mathbf{y} \leftarrow \mathcal{S}_{\text{pure}}(\mathbf{x})$ `// selected point at boundary of` Δ^n
  $\alpha \leftarrow 1$
  $\pi \leftarrow (\mathbf{y}-\mathbf{x})^\mathsf{T}\mathsf{A}'\mathbf{y} - (\mathbf{y}-\mathbf{x})^\mathsf{T}\mathsf{A}'\mathbf{x}$
  if $\pi < 0$ `// negative curvature direction` $\mathbf{y}-\mathbf{x}$
  then
    $\alpha \leftarrow \min\left[\frac{(\mathbf{x}-\mathbf{y})^\mathsf{T}\mathsf{A}'\mathbf{x}}{\pi}, 1\right]$ `// optimal stepsize`
  $\mathbf{x} \leftarrow \alpha(\mathbf{y}-\mathbf{x}) + \mathbf{x}$ `// optimal convex combination of x and y`
  $\epsilon(\mathbf{x}) \leftarrow \sum_i \min\{x_i, \mathbf{x}^\mathsf{T}\mathsf{A}'\mathbf{x} - [\mathsf{A}'\mathbf{x}]_i\}^2$ `// Nash error - violation of equil.condition`

Algorithm 1: InImDyn Algorithm

6.3 Results for the nominal instances

Computational results for the benchmark instances of the nominal case, i.e., solving $\max_{\mathbf{x}\in\Delta^n} \mathbf{x}^\mathsf{T}\mathsf{A}\mathbf{x}$ are given in Table 1. We used $2|V|$ different starting points (indicated by superscript (0)) for each multi-start of InImDyn for each instance, i.e., $\mathbf{x}_m^{(0)} \in \Delta^n$ for all $m \in [1:2|V|]$, such that

$$\mathbf{x}_m^{(0)} = \begin{cases} \mathbf{e}_m & \text{if } 1 \le m \le |V|, \\ \lambda\mathbf{e}_u + (1-\lambda)\mathbf{e}_v & \text{otherwise,} \end{cases}$$

where $\lambda \in [0,1]$ is chosen uniformly at random. Indices u and v are chosen in non-decreasing order of the edge-weights a_{uv}, $\{u,v\} \in E$.

The results in Table 1 are the reference points for all computations considering uncertain data realizations.

6.4 Results for the robust instances

This section considers conic, ellipsoidal, and ball uncertainty and compares the obtained results for different sizes of these uncertainty sets. In particular we relate the size of an uncertainty set directly to the worst-case perturbations L which they impose on the nominal data matrix A. These perturbations are constructed as described in Section 6.1, and 21 different sizes are considered for each uncertainty set. For the conic uncertainty set these are constructed by scaling matrix L using scalar $\delta = \ell/20$ for all $\ell \in [0:20]$.

For an ellipsoidal uncertainty set we have already derived the worst-case perturbations $-\rho(\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}$ in Section 3.2, so that scalar $\rho > 0$ controls the size of that uncertainty set. Thus, for a meaningful comparison we need to relate ρ and δ. We consider the sum of all perturbations they impose on the data matrix as an indicator of the size of an uncertainty set. The conic perturbations is used as reference, i.e., $\delta\,\|\mathsf{L}\|_1$. Let $\text{Diag}(\mathbf{r}) := (\mathsf{C}\mathsf{C}^\mathsf{T})^{-1}$ where $\mathbf{r} = \mathbf{e}$ for the ball uncertainty set, and where $r_i \in [1/2, 3/2]$ is chosen uniformly at random across all $i \in V$ for the ellipsoidal uncertainty. Note that $\text{Diag}(\mathbf{r})$ is nonsingular by construction, hence, the diagonal matrices

Table 1: Results for the nominal problem instances described by graph order $|V|$ and graph size $|E|$ based on $2|V|$ multi-starts of InImDyn per instance. Numbers of identified different solutions ($\#\ \sigma(\mathbf{x})$) and the corresponding average cardinalities of dominant sets ($\overline{|\sigma(\mathbf{x})|}$) as well as average objective function values $\overline{f(\mathbf{x})}$ and average runtimes in seconds ($\bar{t}[s]$) of each multi-start iteration are reported.

Instance name	$\|V\|$	$\|E\|$	$\#\ \sigma(\mathbf{x})$	$\overline{\|\sigma(\mathbf{x})\|}$	$\overline{f(\mathbf{x})}$	$\bar{t}[s]$
`AG-Monien_3elt`	4720	13722	5346	3.32	20.86	0.10
`Arenas_celegans_metabolic`	453	2007	111	3.58	38.48	0.02
`Arenas_email`	1133	4229	149	3.79	33.34	0.06
`Arenas_jazz`	198	2734	17	6.76	69.69	0.04
`DIMACS10_chesapeake`	39	163	10	3.20	36.10	0.01
`DIMACS10_data`	2851	15081	1459	3.95	37.71	0.14
`DIMACS10_delaunay_n10`	1024	3056	482	3.30	23.99	0.03
`DIMACS10_delaunay_n11`	2048	6127	926	3.25	24.08	0.05
`DIMACS10_delaunay_n12`	4096	12264	1955	3.29	23.95	0.22
`Hamm_add20`	2395	7096	802	3.24	37.92	0.16
`Hamm_add32`	4960	8782	1664	3.13	24.92	0.21
`Newman_adjnoun`	112	308	6	3.83	27.70	0.01
`Newman_celegansneural`	297	2029	59	3.54	32.20	0.02
`Newman_dolphins`	62	121	7	3.71	36.15	0.00
`Newman_football`	115	517	18	6.17	45.34	0.01
`Newman_karate`	34	67	5	3.40	33.98	0.00
`Newman_lesmis`	77	232	9	5.00	64.45	0.01
`Newman_netscience`	1589	2521	289	3.94	54.56	0.06
`Newman_polblogs`	1490	16029	40	3.98	50.56	0.24
`Newman_polbooks`	105	423	22	4.45	38.35	0.01

$\mathsf{C}\mathsf{C}^\mathsf{T}$ and also C exist with entries $[\mathsf{C}]_{ii} = \sqrt{1/r_i}$ for all $i \in V$. Scalar ρ is then computed by

$$\rho = \frac{\delta \left\|\mathsf{L}\right\|_1}{(\max_{i\in V} r_i) \left\|\mathsf{E}\right\|_1 - \sum_{j\in V} r_j}\,, \tag{22}$$

which denominator is explained as follows. The ellipsoidal uncertainty set perturbs the data matrix A negatively on its diagonal elements. Hence, we consider the rank-one update $\mathsf{A}' = \mathsf{A} - \rho\,\mathrm{Diag}(\mathbf{r}) + \rho(\max_{i\in V} r_i)\mathsf{E}$ to obtain $\mathsf{A}' \in \mathcal{N}_n$ as input data for InImDyn. In other words, we include the rank-one update as a discount for computing ρ although only the diagonal elements are perturbed by $\sum_{j\in V} r_j$. Otherwise, ρ would grow rapidly compared to δ which exacerbates a comparison, since similarity measures are sensitive to translation (cf. Section 5.1).

Figure 2a depicts relative averages over the mean objective function values of all different solutions that have been identified by all multi-starts of InImDyn. The mean objective value in the nominal case (cf. Table 1) is used as reference value and therefore corresponds to point $(0, 1)$. Relative averages of numbers of different solutions are shown in Figure 2b while Figure 2c shows average cardinalities of dominant sets. Similar to Figure 2a the corresponding numbers of the nominal case are used as reference values. We observe that the worst-case results for ellipsoidal uncertainty sets results on average in larger objective function values than for conic and ball uncertainty sets. From Figure 2b we observe that the number of different solutions found typically decreases with increasing size of an uncertainty set when considering ellipsoidal uncertainty. The opposite trend can be observed from Figure 2c with respect to the average cardinality of found dominant sets. This observation may stem from the fact that the rank-one update used for the ellipsoidal uncertainty sets contributes to increase the similarity of vertices (until the whole graph is one dominant set). We also observe, however, that the opposite is true for conic uncertainty sets for which increasing the size of the uncertainty set results in a larger number of identified dominant sets of lower cardinality.

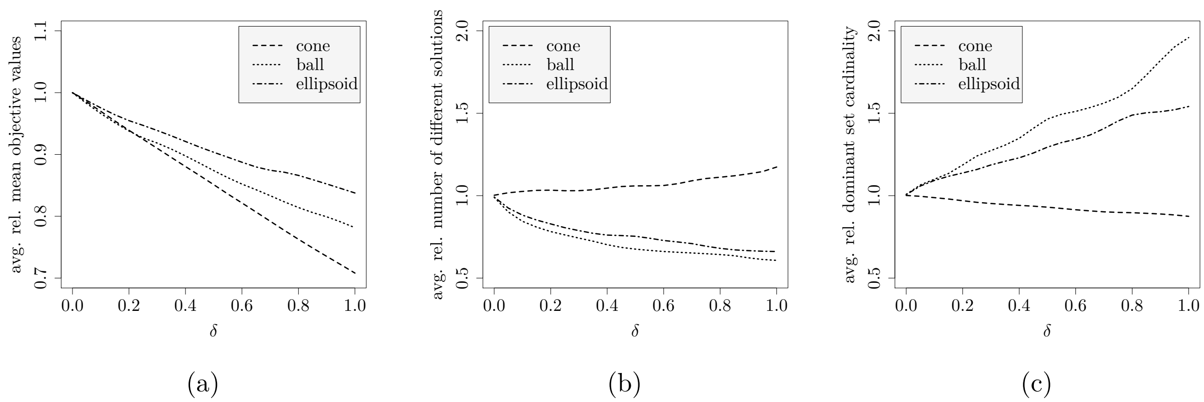

Figure 2: (a) Mean objective function values; (b) number of different solutions found; (c) corresponding dominant set cardinalities, relative to the nominal csse solutions for each instance with $2|V|$ multi-starts of InImDyn. The results are averaged across all instances.

6.5 Community detection in social networks.

This section demonstrates how (robust) dominant set clustering can be used to identify communities in social networks based on similarity among their members. In particular we discuss the instance `Newman_lesmis` in which vertices correspond to the characters of Victor Hugo's novel Les Misérables, while each edge weight corresponds to the number of simultaneous appearances of two characters in a chapter. The modifications described in Section 6.1 have been applied to the original instance. The computational results which are visualized in Figure 3 confirm that dominant sets correspond to communities formed by similar characters. The community (1) of Figures 3a-3d corresponds, for example, to the rebels in the novel. We also observe that the cardinalities of dominants sets decrease for conic uncertainty sets and increase for ball and ellipsoidal uncertainty when compared to the nominal case. The reverse effect can be observed regarding the number of identified different dominant sets ($\#\sigma(\mathbf{x})$). Finally, observe that all dominant sets form cliques (cf. Proposition 9) in the original graph except one in Figure 3d, which is, however, connected (cf. Proposition 8) as discussed in Section 5.2 and forms a clique in the perturbed instance.

7 Conclusions

In this article, we have introduced the Robust Standard Quadratic Optimization Problem and have shown that its CP relaxation gap is equal to the minimax gap. For several, frequently considered

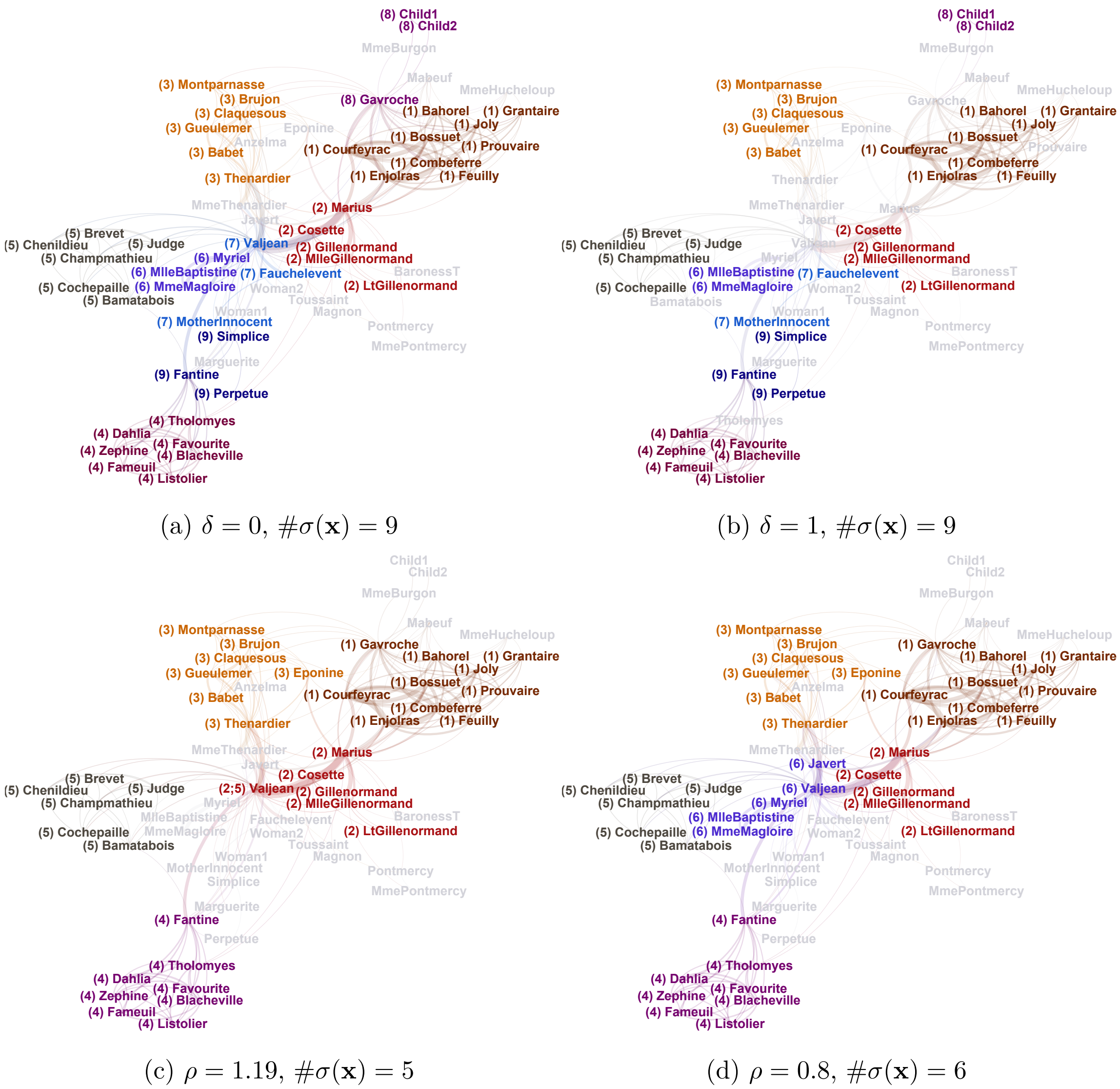

Figure 3: (Robust) dominant sets (indicated by $(\cdot)$ before the character names) of the instance `Newman_lesmis`: (a) nominal case, (b) conic uncertainty, (c) ball uncertainty, and (d) ellipsoidal uncertainty.

uncertainty sets we also have shown that this CP relaxation is an exact reformulation and that the minimax theorem holds without the usual convexity assumptions. A direct consequence of these results is that the RStQP reduces to an StQP in such cases which therefore retains the computational effort needed to solve the nominal problem variant. We have also investigated the robust variant of the dominant set clustering problem. For its nominal case, we have shown that dominant sets form cliques in the similarity graph. The results of our computational study indicate that considering ellipsoidal uncertainty sets tend to identify a smaller number of dominant sets but with higher cardinality and objective function values compared to the results of considering conic uncertainty sets. Promising research directions for future work include the consideration of less conservative robust optimization approaches such as, e.g., budgeted uncertainty or distributionally robust optimization.

Acknowledgments.

The authors thank Juan Vera Lizcano for pointing out a generalization of our initial argument in Theorem 4.

References

[1] B. Alidaee, F. Glover, G. Kochenberger, and H. Wang. Solving the maximum edge weight clique problem via unconstrained quadratic programming. *European Journal of Operational Research*, 181(2):592–597, 2007.

[2] D. A. Bader, H. Meyerhenke, P. Sanders, and D. Wagner. *Graph partitioning and graph clustering - 10th DIMACS Implementation Challenge Workshop*, volume 588 of *Contemporary Mathematics*. American Mathematical Society, Providence RI, 2012.

[3] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semidefinite programming. *SIAM Journal on Optimization*, 7(4):991–1016, 1997.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[5] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[6] A. Ben-Tal, L. El Ghaoui, and A. S. Nemirovski. *Robust optimization.* Princeton series in applied mathematics. Princeton University Press, Princeton NJ, 2009.

[7] A. Ben-Tal, D. den Hertog, and J.-P. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.

[8] A. Berman and N. Shaked-Monderer. *Completely positive matrices.* World Scientific, Singapore, 2003.

[9] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[10] I. M. Bomze. On standard quadratic optimization problems. *Journal of Global Optimization*, 13(4):369–387, 1998.

[11] I. M. Bomze. Copositive optimization – Recent developments and applications. *European Journal of Operational Research*, 216(3):509–520, Feb. 2012.

[12] I. M. Bomze and E. De Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24(2):163–185, 2002.

[13] I. M. Bomze, M. Dür, E. de Klerk, C. Roos, A. J. Quist, and T. Terlaky. On copositive programming and standard quadratic optimization problems. *Journal of Global Optimization*, 18(4):301–320, 2000.

[14] I. M. Bomze, M. Pelillo, and V. Stix. Approximating the maximum weight clique using replicator dynamics. *IEEE Transactions on Neural Networks*, 11(6):1228–1241, 2000.

[15] I. M. Bomze, F. Rinaldi, and S. Rota Bulò. First-order methods for the impatient: support identification in finite time with convergent Frank-Wolfe variants. *SIAM J. Optimization*, , to appear, 2019.

[16] M. Doumpos, C. Zopounidis, and E. Grigoroudis, editors. *Robustness Analysis in Decision Aiding, Optimization, and Analytics*, volume 241 of *International Series in Operations Research and Management Science*. Springer International, Switzerland, 2016.

[17] F. J. Fabozzi, D. Huang, and G. Zhou. Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176(1):191–220, 2010.

[18] F. J. Fabozzi, P. N. Kolm, D. A. Pachamanova, and S. M. Focardi. *Robust Modeling of Uncertain Parameters in Classical Mean-Variance Portfolio Optimization*, pages 361–394. John Wiley and Sons, Inc., 2012.

[19] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, Feb. 2010.

[20] V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483, 2014.

[21] L. E. Gibbons, D. W. Hearn, M. V. Ramana, and P. Pardalos. Continuous characterizations of the maximum clique problem. *Mathematics of Operations Research*, 22(3):754–768, 1997.

[22] B. L. Gorissen, I. Yanıkoglu, and D. den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, June 2015.

[23] R. Horst, P. Pardalos, and N. Van Thoai. *Introduction to Global Optimization*. Nonconvex Optimization and Its Applications. Springer, New York NY, 2000.

[24] A. Marandi, A. Ben-Tal, D. den Hertog, and B. Melenberg. Extending the scope of robust quadratic optimization. *Available on Optimization Online*, 2017.

[25] A. Mittal, C. Gokalp, and G. A. Hanasusanto. Robust quadratically constrained quadratic programming with mixed-integer uncertainty. *INFORMS Journal on Computing*, accepted, 2019.

[26] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics*, 17(4):533–540, 1965.

[27] M. Pavan. *A new graph-theoretic approach to clustering and segmentation*. PhD dissertation, Università Ca' Foscari di Venezia, 2004.

[28] M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):167–172, 2007.

[29] M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Notices of the AMS*, 56(9):1082–1097, 2009.

[30] S. Rota Bulò and M. Pelillo. Dominant-set clustering: A review. *European Journal of Operational Research*, 262(1):1–13, Oct. 2017.

[31] S. Rota Bulò, M. Pelillo, and I. M. Bomze. Graph-based quadratic optimization: A fast evolutionary approach. *Computer Vision and Image Understanding*, 115(7):984–995, 2011.

[32] S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, Aug. 2007.

[33] N. Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25(6):1–11, 1987.

[34] F. Talla Nobibon and R. Leus. Robust maximum weighted independent-set problems on interval graphs. *Optimization Letters*, 8(1):227–235, Jan. 2014.

[35] J. Vera Lizcano. Personal communication, Oberwolfach Workshop on Copositivity and Complete Positivity, 29 October-4 November, 2017.

[36] J. von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, Dec 1928.

[37] K. Ye, P. Parpas, and B. Rustem. Robust portfolio optimization: a conic programming approach. *Computational Optimization and Applications*, 52(2):463–481, Jun 2012.

[38] T. Zhou, L. Lü, and Y.-C. Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71(4):623–630, Oct 2009.