

BREAKING THE ODD CYCLE: REGULARIZATION AND FIRST-ORDER OPTIMIZATION FOR THE MAXIMUM S -PLEX PROBLEM

ELEONORA BELLESSO*, IMMANUEL M. BOMZE†, AND FRANCESCO RINALDI*

Abstract. Building on a new regularized continuous formulation which avoids spurious solutions of the maximum s -plex problem, we develop a tailored block-coordinate first-order method that combines away-step Frank-Wolfe updates on the standard simplex with linear optimization over the fractional b -matching polytope. We establish finite stabilization of the auxiliary variables, convergence to stationary points, and finite support identification under suitable conditions. Numerical experiments on standard benchmark graphs show that, also in practice, regularization avoids the support-identification failures observed with the original cubic formulation and enables the proposed method to quickly find high-quality s -plexes.

Key words. clique relaxations, maximum s -plex, Motzkin-Straus, Frank-Wolfe

AMS subject classifications. 05C35, 05C50, 65K05, 90C06, 90C30, 90C35

1. Introduction. The Motzkin-Straus (MS) formulation [14] is one of the most influential examples of how a combinatorial graph problem can be approached through continuous optimization. By maximizing a quadratic function over the standard simplex, it recovers the clique number of a graph from the resulting optimal value. Besides its theoretical significance, this connection has inspired several continuous approaches to the maximum clique problem (MCP); see, for example, [1, 9]. The correspondence, however, is incomplete at the level of solutions: maximizers of the MS formulation need not have clique support. These so-called spurious solutions are particularly problematic for algorithms that recover a clique from the support of a continuous solution and have motivated the development of suitable regularizations of the original formulation [16].

The influence of the MS formulation, however, extends beyond cliques. In many applications, particularly in community detection, requiring every pair of selected vertices to be adjacent is overly restrictive, since meaningful communities are generally dense but not complete. This observation has led to several clique-relaxation models, which describe cohesive subgraphs while allowing controlled deviations from complete connectivity; see [8, 15] and the references therein. In this paper, we focus on the s -plex model, introduced by Seidman and Foster [19] in the context of social network analysis. A vertex set is an s -plex if each of its vertices has at most $s - 1$ non-neighbors within the set. Thus, 1-plexes are precisely cliques, whereas values of $s > 1$ permit progressively weaker connectivity requirements. This model should not be confused with the s -defective clique relaxation, introduced by Yu et al. [22] in the context of protein-protein interaction networks, which allows up to s missing edges in the vertex set. These two models have often been studied together due to shared properties, such as heredity, despite the fact that the internal structures they can generate are fundamentally different and that, in general, no process exists that allows one structure to be transformed into the other. In [20], two cubic MS-type formulations for the maximum s -plex problem (MsPP) and the maximum s -defective clique problem

*Dipartimento di Matematica “Tullio Levi-Civita”, Università di Padova, Padova PD, Italy (eleonora.bellesso@studenti.unipd.it, rinaldi@math.unipd.it).

†Fakultät für Mathematik & ds:UniVie, Universität Wien, Wien, Austria (immanuel.bomze@univie.ac.at).

(MsDCP) were proposed separately. For both cases, in addition to a vertex vector belonging to the standard simplex, the formulations employ variables indexed by the graph's missing edges. These variables encode feasible sets of missing edges, subject to the constraint that at most $s - 1$ of them can be incident to any single vertex for the s -plex case, whereas at most s missing edges can be selected for the s -defective clique model. In both formulations, the global optimal value of the cubic problem allows one to derive the maximum cardinality of the sought relaxed structure. However, as is the case with the original MS formulation, this characterization of the optimal value does not automatically extend to the structure of the maximizers: the cubic formulation can indeed admit spurious solutions whose vertex supports are not clique relaxations.

For the s -plex formulation, eliminating these solutions is technically more involved than in both the clique and the s -defective clique cases. For the MCP, there are no missing-edge variables involved, so the regularization acts only on the vertex variables over the standard simplex; see [9]. The s -defective clique formulation does introduce missing-edge variables, and its continuous relaxation contains fractional feasible points. Nevertheless, its feasible region is an integral polytope for integer s , meaning that all its extreme points are binary. Consequently, strict convexity with respect to the missing-edge variables is sufficient to ensure that every local maximizer of the corresponding fixed-vertex problem is binary [3]. This argument does not extend directly to the s -plex formulation, where the separate degree constraints imposed at the vertices define a fractional b -matching polytope. Its extreme points may themselves be fractional; more precisely, they are half-integral, with their one-half components supported on vertex-disjoint odd cycles. Strict convexity therefore guarantees only that a local maximizer is an extreme point, not that its missing-edge component is binary. The purpose of the y -regularization is therefore to break the odd cycles supporting fractional extreme points, in the sense of making every such half-integral configuration locally improvable while leaving binary missing-edge selections unaffected. Achieving this requires additional conditions on the regularization and constitutes the main technical distinction of the s -plex setting.

It is hence easy to understand why a framework for the s -plex problem cannot be obtained through a direct adaptation of the regularizations developed for the clique and s -defective clique problems. Such a regularizer must simultaneously control the geometry of the objective over the simplex and eliminate half-integral configurations in the missing-edge block, while leaving binary missing-edge selections unpenalized. As we will see, by deriving regularization conditions that achieve both goals through an appropriate exploitation of the problem's distinctive polyhedral structure, we eliminate the spurious maximizers of the cubic formulation and ensure that every local maximizer encodes an s -plex and every global maximizer encodes a maximum s -plex. The main contributions of the paper can be summarized as follows:

- We introduce a general class of regularized continuous formulations for the MsPP and derive conditions on the regularization terms that exclude local maximizers with fractional missing-edge components and vertex supports that do not correspond to an s -plex. We first focus on the binary complementary structure. Taking into account the half-integrality of the fractional b -matching polytope, we show that the regularization rules out local maximizers with fractional missing-edge components. Since the one-half components of every non-binary extreme point are supported on vertex-disjoint odd cycles, this result excludes all such fractional odd-cycle configurations. We then characterize the local and global maximizers of the regularized problem. In particular, every local maximizer has a binary missing-edge component and a

vertex component given by the normalized characteristic vector of an s -plex, whereas global maximizers correspond to maximum s -plexes.

- We provide and analyze several families of regularization functions satisfying the proposed conditions.
- We develop a tailored block-coordinate first-order method that combines away-step Frank-Wolfe updates on the standard simplex with linear optimization over the fractional b -matching polytope. We establish finite stabilization of the missing-edge variables, convergence to stationary points, and finite support identification under suitable conditions.
- We evaluate the proposed method on a standard family of benchmark graphs. The reported numerical results show that regularization avoids the support-identification failures observed with the original cubic formulation and enables the proposed method to quickly find high-quality s -plexes.

Frank-Wolfe methods [7] are particularly attractive in this setting because they avoid projections and instead rely on linear optimization over the feasible region. This projection-free feature is crucial when dealing with geometrically complex polytopes, such as the fractional b -matching polytope considered here. Indeed, its Euclidean projection has no simple closed-form expression and would require solving a non-trivial constrained quadratic problem at every iteration, making projection-based methods computationally impractical. By contrast, the Frank-Wolfe linear optimization oracle can directly exploit the underlying b -matching structure. Moreover, such projection-free algorithms are able to generate sparse iterates and identify the support of a limiting solution. This aligns well with the combinatorial interpretation of the simplex variables. The away-step mechanism is especially useful for removing incorrectly selected vertices, while the block-coordinate structure makes it possible to treat the simplex and fractional b -matching components using procedures tailored to their respective geometries.

The remainder of the paper is organized as follows. In section 2, we introduce notation, review the cubic formulation, and recall some preliminary results on the fractional b -matching polytope. Section 3 develops the general regularization framework and characterizes the local and global maximizers of the resulting formulation. Section 4 presents and analyzes several families of regularization functions satisfying the proposed conditions. In section 5, we present the block-coordinate first-order method and establish its finite-time stabilization, convergence, and support-identification properties. Section 6 reports the computational study, including the comparison between the original and regularized formulations. Section 7 concludes the paper.

The implementation of the proposed method, together with the benchmark instances used in the computational study, is publicly available at the following repository: https://github.com/eleonora-bellesso/s-plex_FW.

2. Notation and preliminaries. Let $G = (V, E)$ be a simple undirected graph with vertex set $V = \{1, \dots, n\}$ and edge set $E \subseteq \binom{V}{2}$ consisting of $m = |E|$ unordered pairs of adjacent vertices. Similarly, denote by $\bar{G} = (V, \bar{E})$ the complementary graph of G , where $\bar{E} = \binom{V}{2} \setminus E$, and $\bar{m} = |\bar{E}|$ is the number of missing edges in G . For any subset of vertices $S \subseteq V$, we denote by $G[S] = (S, E(S))$ the subgraph induced by S in G , where $E(S) = E \cap \binom{S}{2}$. The notation adopted throughout this article largely follows that introduced in [20].

DEFINITION 2.1. *Given a positive integer s , a subset of vertices $S \subseteq V$ is an s -plex if every vertex in S is adjacent to at least $|S| - s$ other vertices of S . Equivalently,*

every vertex in S has at most $s - 1$ non-neighbors in S .

The maximum s -plex problem aims to find an s -plex of maximum cardinality in G . This cardinality is denoted by $w_s^p(G)$ and is called the s -plex number of G .

In [20], the authors showed that the quadratic MS formulation for the MCP can be extended to a continuous cubic formulation for the MsPP by introducing a vector y of auxiliary variables indexed by the missing edges of G . These variables represent the “fake edges” that may be added to the graph in order to recover a clique structure from the s -plex relaxation.

In the following, let $A_G = [a_{ij}]_{i,j \in V}$ be the $n \times n$ adjacency matrix of G such that $a_{ij} = 1$ if $\{i, j\} \in E$ and 0 otherwise, and $N(v)$ the neighborhood of any vertex $v \in V$ which indeed corresponds to the set of adjacent vertices to v in G . We instead indicate with $B_{\bar{G}} = [b_{vc}]_{v \in V, c \in \bar{E}}$ the $n \times \bar{m}$ incidence matrix of the complementary graph $\bar{G}$, such that $b_{vc} = 1$ if v is an endpoint of the missing edge c , so $v \in c$; and 0 otherwise.

Let $\Delta_n = \{x \in \mathbb{R}^n \mid x \geq 0, e^\top x = 1\}$ be the standard simplex in $\mathbb{R}^n$. Here e is always a vector with all components equal to 1 and the dimension clear from the context. Similarly, we denote by $\mathbb{I}_n$ the $n \times n$ identity matrix and by $e_i \in \mathbb{R}^n$ its i -th column. For any vector z whose components are indexed by a finite set I , we denote its support by $\text{supp}(z) := \{i \in I : z_i \neq 0\}$. For instance, for any nonempty $S \subseteq V$, we can define the (normalized) characteristic vector of S as

$$x(S) := \frac{1}{|S|} \sum_{i \in S} e_i \in \Delta_n$$

which is the barycenter of the face $\Delta(S) := \{x \in \Delta_n \mid \text{supp}(x) \subseteq S\}$ whose support is $\text{supp}(x(S)) = S$.

For each missing edge $\{i, j\} \in \bar{E}$, we formally introduce a binary variable y_{ij} . For any fixed positive integer $s \geq 1$, we define

$$(2.1) \quad P_s(G) = \{y \in \{0, 1\}^{\bar{m}} \mid B_{\bar{G}} y \leq (s - 1)e\}$$

as the set of incidence vectors of admissible missing-edge sets that may be added to the original graph G in order to complete an s -plex to a clique, and its continuous relaxation as

$$(2.2) \quad P'_s(G) = \{y \in [0, 1]^{\bar{m}} \mid B_{\bar{G}} y \leq (s - 1)e\}.$$

For any $y \in P'_s(G)$ we define $A_{\bar{G}}(y)$ as the matrix with entry in the i -th row and the j -th column given by $y_{ij} \in [0, 1]$ if the edge $\{i, j\} \in \bar{E}$ and 0 otherwise. That is, $A_{\bar{G}}(y)$ is obtained from the adjacency matrix $A_{\bar{G}}$ of the complementary graph $\bar{G}$ by replacing each non-zero entry with the corresponding y_{ij} value. Using the notation introduced above, we can now recall the main continuous characterization of the s -plex number established in [20]. The result shows that the integrality constraints on the missing-edge variables can be relaxed without changing the global optimal value, thereby extending the classical MS formulation from cliques to s -plexes.

THEOREM 2.2 ([20, Theorem 2]). *The s -plex number $w_s^p(G)$ of a graph G can be expressed in terms of the optimal value of the following continuous cubic problem:*

$$(2.3) \quad 1 - \frac{1}{w_s^p(G)} = \max_{(x,y) \in \Delta_n \times P'_s(G)} f_G(x, y)$$

where the objective function is given by

$$(2.4) \quad f_G(x, y) = x^\top (A_G + A_{\bar{G}}(y))x .$$

We observe that the objective function (2.4) is a polynomial of total degree three in the variables (x, y) . More precisely, it is quadratic with respect to the vertex variables x and affine with respect to the missing-edge variables y . Moreover, the projection of the feasible set onto the x variables coincides with the standard simplex Δ_n as in the original MS formulation. Indeed, for the specific value of $s = 1$, the feasible region for the y variable coincides with the single zero vector, $P'_1(G) = \{0\}$, and also the induced adjacency matrix is simply given by the zero matrix, $A_{\bar{G}}(y) = 0$. This finally means that no missing edges can be included in the computation, and so the above problem (2.3) trivially coincides with the original MS formulation for the MCP, as expected.

It is also important to note that establishing the correctness of formulation (2.3) is rather non-trivial. In particular, the authors in [20] exploited a classical result concerning the half-integrality property of the fractional b -matching polytope. Since this property will play a central role in our subsequent analysis, we decided to briefly review some fundamental concepts and results concerning the canonical b -matching problem. We begin by formally defining this structure.

DEFINITION 2.3. *Given an undirected graph $G = (V, E)$ and a non-negative integer vector $b \in \mathbb{Z}_n^+$, a subset of edges $M \subseteq E$ is said to be a b -matching in G if no vertex $i \in V$ is incident to more than b_i edges in M .*

Moreover, given non-negative weights w_{ij} associated with each edge $\{i, j\} \in E$, the maximum weight b -matching problem (MWbMP) aims to find a b -matching in G with the largest total possible weight. In particular, an integer linear program can be formulated for this combinatorial optimization problem, where for any edge $\{i, j\} \in E$ a binary variable z_{ij} is considered, indicating whether the corresponding edge is included or not in the b -matching structure.

$$(2.5) \quad \begin{aligned} \max \quad & \sum_{\{i,j\} \in E} w_{ij} z_{ij} \\ \text{s.t.} \quad & \sum_{j \in N(i)} z_{ij} \leq b_i, \quad \forall i \in V \\ & z_{ij} \in \{0, 1\}, \quad \forall \{i, j\} \in E. \end{aligned}$$

The linear relaxation of problem (2.5) reads

$$(2.6) \quad \begin{aligned} \max \quad & \sum_{\{i,j\} \in E} w_{ij} z_{ij} \\ \text{s.t.} \quad & \sum_{j \in N(i)} z_{ij} \leq b_i, \quad \forall i \in V \\ & z_{ij} \in [0, 1], \quad \forall \{i, j\} \in E. \end{aligned}$$

Clearly, problem (2.6) may have a fractional optimal solution and, in general, a higher optimal value. Consider, for example, the case where $b = e$, the unit vector, and $w_{ij} = 1$ for every edge $\{i, j\} \in E$, so the standard maximum matching problem, over the undirected graph $G = C_5$, so the simple cycle of 5 vertices. Then the optimal solution to problem (2.6) is obtained by setting $z_{ij} = \frac{1}{2}$ for every edge $\{i, j\} \in E(C_5)$.

This observation finally leads to the following well-known result on the fractional b -matching polytope [11, pp. 44-46].

LEMMA 2.4. *Let P be the feasible region of problem (2.6). Any extreme point $\hat{z}$ of P is half-integral, i.e., $\hat{z}_{ij} \in \{0, \frac{1}{2}, 1\}$, $\forall \{i, j\} \in E$. Moreover, the edge-induced subgraph $G\langle F \rangle$, where $F = \{\{i, j\} \in E \mid \hat{z}_{ij} = \frac{1}{2}\}$, consists of disjoint odd cycles.*

Here, the edge-induced subgraph $G\langle F \rangle = (V_F, F)$ has F as its edge set, while its vertex set $V_F \subseteq V$ includes all vertices incident to the edges of F .

We now return to the MS formulation (2.3) for the MsPP. As shown in the proof of Theorem 2.2, fixing the x -component at some $x^* \in \Delta_n$ yields the following maximization problem in the missing-edge variables:

$$(2.7) \quad \begin{aligned} \max \quad & 2 \sum_{\{i,j\} \in E} x_i^* x_j^* + 2 \sum_{\{i,j\} \in \bar{E}} x_i^* y_{ij} x_j^* \\ \text{s.t.} \quad & B_{\bar{G}} y \leq (s-1)e, \quad y \in [0, 1]^{\bar{m}}. \end{aligned}$$

Since the first term of the objective does not depend on y , problem (2.7) coincides, up to an additive constant, with the linear relaxation of the MWbMP (2.5) on the complement graph $\bar{G}$, with $b = (s-1)e$ and weights $w_{ij}^* = 2x_i^* x_j^*$ for every $\{i, j\} \in \bar{E}$.

The proof of Theorem 2.2 exploits this connection to show that, starting from an arbitrary global maximizer of formulation (2.3), one can construct another global maximizer (x^*, y^*) with the same objective value such that y^* is binary and $\text{supp}(x^*)$ is a maximum s -plex in G . Moreover, y^* selects precisely the missing edges needed to complete the subgraph induced by $\text{supp}(x^*)$ into a clique. This is an existence result and does not imply that every global maximizer has the same combinatorial structure. Indeed, as already pointed out in [20], formulation (2.3) may admit spurious global maximizers, namely optimal pairs (x^*, y^*) for which $\text{supp}(x^*)$ is not an s -plex in G . The following example provides a simple, yet representative illustration of this undesirable behavior. Consider the graph $G = (V, E)$ shown in Figure 1. Clearly, the 2-plex number of G is $w_2^p(G) = 3$ since both sets $C = \{1, 2, 3\}$ and $C' = \{2, 3, 4\}$ are 2-plexes of maximum cardinality in G . However, there exists a globally optimal solution $(x^*, y^*) \in \Delta_n \times P'_2(G)$ of (2.3) whose support in the x -component $S^* = \text{supp}(x^*) = \{1, 2, 3, 4\}$ is not a 2-plex in G . Indeed, the solution represented in Figure 1 satisfies $f_G(x^*, y^*) = \frac{2}{3} = 1 - \frac{1}{w_2^p(G)}$, and is therefore globally optimal by Theorem 2.2. This phenomenon is consistent with the fact that the cubic

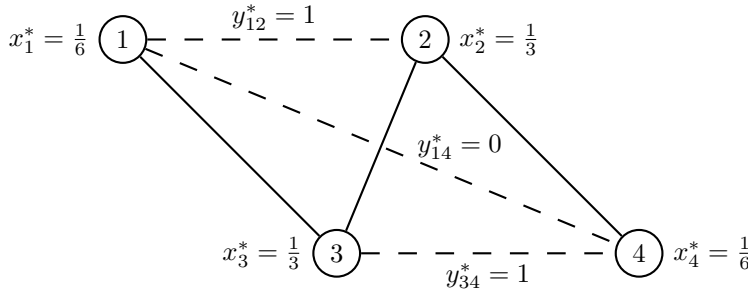


Fig. 1: Example of (x^*, y^*) spurious solution for G . The dashed lines indicate the missing edges.

formulation (2.3) for the MsPP was derived as a direct extension of the original MS

formulation. In particular, it inherits an analogous limitation at the level of solutions: although the formulation provides the exact optimal value, not every global maximizer has a vertex support that gives the desired combinatorial structure.

Accordingly, the primary aim of our study is to develop a regularized variant of formulation (2.3) that guarantees the vertex support of each local maximizer forms an s -plex in G .

3. General regularized formulation. In this section, we address the issue of spurious solutions for the cubic s -plex formulation (2.3) by introducing suitable regularization terms in the objective function, which ensure equivalence between the derived regularized version and the initial MsPP in both a local and global sense. In particular, no post-processing algorithms are needed to recover an s -plex structure from the optimal solution returned by the program.

Let us start by recalling a well-known property of convex maximization problems, which helps to locate optimal solutions as extreme points of the feasible region; see [18] for a more detailed discussion.

LEMMA 3.1. *Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty polytope, and let $f \in C^2(U)$ be a strictly convex function defined on an open set $U \supseteq \Omega$. Then every local maximizer, and hence every global maximizer, of $\max_{x \in \Omega} f(x)$ is an extreme point of Ω . Moreover, every local maximizer is strict.*

We continue the discussion by recalling some results from [9] on the regularized version of the quadratic MS formulation for the MCP, which will indeed play a fundamental role in the development of our subsequent analysis. Given the graph $G = (V, E)$, consider the following optimization problem

$$(3.1) \quad \begin{aligned} \max \quad & [x^\top A_G x + \Phi(x)] \\ \text{s.t.} \quad & x \in \Delta_n, \end{aligned}$$

where $\Phi : X \rightarrow \mathbb{R}$ is a twice continuously differentiable function defined in some open set $X \supseteq \Delta_n$ that satisfies the following conditions for every $x \in \Delta_n$:

- (C1) $\nabla^2 \Phi(x) > 0$, so Φ is strictly convex over the simplex;
- (C2) $\|\nabla^2 \Phi(x)\|_2 < 2$, or equivalently $\nabla^2 \Phi(x) - 2\mathbb{I}_n < 0$;
- (C3) Φ is constant on the set $\mathcal{P}(x) := \{\tilde{x} \in \Delta_n \mid \exists \sigma \in \mathcal{S}_n \text{ s.t. } \tilde{x}_i = x_{\sigma(i)} \forall i \in \{1, \dots, n\}\}$, where $\mathcal{S}_n$ is the set of permutations of $\{1, \dots, n\}$.

For simplicity, we now denote by

- $k_G(x) = x^\top A_G x + \Phi(x)$ the objective function of (3.1);
- $T(x) = \{d \in \mathbb{R}^n \mid e^\top d = 0, \text{ and } d_i \geq 0 \text{ for all } i \notin \text{supp}(x)\}$ the set of feasible directions for (3.1) at $x \in \Delta_n$.

Under these assumptions, the following results hold. Recall that for any $\emptyset \neq C \subseteq V$, the point $x(C) \in \Delta_n$ is the (normalized) characteristic vector of C which satisfies $x(C)_i = \frac{1}{|C|}$ if $i \in C$ and $x(C)_i = 0$ otherwise.

LEMMA 3.2 ([9, Proposition 3]). *If C^1 and C^2 are cliques, then*

$$|C^1| < |C^2| \iff k_G(x(C^1)) < k_G(x(C^2)).$$

LEMMA 3.3 ([9, Lemma 3]). *If C is a maximal clique in G , then*

$$x(C)^\top A_G d \leq 0, \quad \forall d \in T(x(C)).$$

LEMMA 3.4 ([9, Lemma 4]). *Let $S \subseteq V$ be non-empty. Then*

$$\nabla \Phi(x(S))^\top d \leq 0, \quad \forall d \in T(x(S)).$$

Moreover, the inequality is strict whenever $\text{supp}(d) \not\subseteq S$.

Finally, through the use of second-order sufficient optimality conditions, the authors in [9] were able to show that there exists a one-to-one correspondence between local (global) maximizers of (3.1) and maximal (maximum) cliques in G .

We can now turn to our results. Recall that we have denoted by A_G the adjacency matrix of the graph G and instead by $A_{\bar{G}}(y)$ the induced adjacency matrix of the complementary graph $\bar{G}$ for every $y \in P'_s(G)$. In particular, note that for any binary vector $y^* \in P_s(G)$, we can define the auxiliary graph G^* with adjacency matrix given by $A_{G^*} = A_G + A_{\bar{G}}(y^*)$ as the graph obtained from G by simply adding the missing edges selected by the binary y^* variable. Moreover, let us also indicate with $N(i) = \{j \in V \mid \{i, j\} \in E\}$ and $N^*(i) = \{j \in V \mid \{i, j\} \in E \cup \text{supp}(y^*)\}$ the neighborhoods of a vertex $i \in V$ for the graphs G and G^* , respectively. Finally recall that $\bar{E}(C)$ represents the set of missing edges with both end points in C .

Consider the following optimization problem:

$$(3.2) \quad \begin{aligned} \max \quad & h_G(x, y) \\ \text{s.t.} \quad & (x, y) \in \Delta_n \times P'_s(G), \end{aligned}$$

where the objective function is given by

$$h_G(x, y) = f_G(x, y) + \Phi(x) + \Psi(y).$$

In particular, the first term corresponds to the objective function (2.4) of the MS formulation for the s -plex relaxation, while the remaining terms are general regularizers for the variables x and y , separately. The choice of using separate regularization terms is motivated by the different roles and feasible-set geometries of the two variable blocks: x represents the vertices, whereas y represents the missing edges of the graph. Specifically, we consider $\Phi(x)$ as a general regularized function for the MS of the MCP, as introduced in [9] and reported in the discussion above, while, for the y -component, we introduce the separable regularization function

$$(3.3) \quad \Psi(y) := \sum_{\{i, j\} \in \bar{E}} \psi(y_{ij}),$$

where $\psi : Y \rightarrow \mathbb{R}$ is a twice continuously differentiable function, defined on some open set $Y \supseteq [0, 1]$, that satisfies the following requirements:

$$(R1) \quad \psi(0) = \psi(1) = 0$$

$$(R2) \quad \psi'(\tfrac{1}{2}) < -\tfrac{1}{2}$$

$$(R3) \quad \psi''(t) > 0 \text{ for all } t \in [0, 1], \text{ so } \psi \text{ is strictly convex over the unit interval.}$$

As will be formally shown below, these conditions ensure that the regularization term vanishes at binary vectors while ruling out half-integral extreme points of $P'_s(G)$ as local maximizers. Consequently, the y -component of every local maximizer of the regularized formulation is binary.

In the special case $s = 1$, the y -variables are constrained to be zero and, by condition (R1), $\Psi(y)$ vanishes. Consequently, the objective function h_G reduces exactly to that of the regularized MS formulation for the MCP considered in [9]. This consistency with the clique case shows that formulation (3.2) can be viewed as a natural extension of the regularized MS framework from cliques to s -plexes.

In the following theorem, we provide a complete characterization of the local maximizers of formulation (3.2). In particular, we show that every local maximizer is strict, has a binary y component, and has an x component equal to the characteristic

vector of an s -plex. In order to show this, we apply second-order sufficient optimality conditions following an approach similar to those in [3, 9]. For this reason, given any feasible point $(x, y) \in \Delta_n \times P'_s(G)$ we define

$$T(x, y) = \{(\bar{x}, \bar{y}) - (x, y) \mid (\bar{x}, \bar{y}) \in \Delta_n \times P'_s(G)\} \subseteq \mathbb{R}^{n+\bar{m}}$$

as the set of feasible directions at (x, y) , while for any vector $r \in \mathbb{R}^{n+\bar{m}}$ we have

$$T^0((x, y), r) = \{d \in T(x, y) \mid d^\top r = 0\}$$

as the intersection between the set $T(x, y)$ and the hyperplane orthogonal to r . So, in this setting, a feasible point $(x^*, y^*) \in \Delta_n \times P'_s(G)$ is a strict local maximizer of (3.2) if the following conditions hold:

$$(3.4) \quad \nabla h_G(x^*, y^*)^\top d \leq 0, \quad \forall d \in T(x^*, y^*),$$

$$(3.5) \quad d^\top \nabla^2 h_G(x^*, y^*) d < 0, \quad \forall d \in T^0((x^*, y^*), \nabla h_G(x^*, y^*)) \setminus \{0\}.$$

THEOREM 3.5. *The following statements are equivalent:*

- i. $(x^*, y^*) \in \Delta_n \times P'_s(G)$ is a local maximizer of (3.2);
- ii. $(x^*, y^*) \in \Delta_n \times P'_s(G)$ is a strict local maximizer of (3.2);
- iii. y^* is binary and $x^* = x(C)$ for C an s -plex in G which corresponds to a maximal clique in the auxiliary graph G^* induced by y^* . Moreover $y_{ij}^* = 1$ for every $\{i, j\} \in \bar{E}(C)$.

In either of these equivalent cases we have that $h_G(x^*, y^*) = 1 - \frac{1}{|C|} + \Phi(x(C))$.

Proof. (i) $\implies$ (iii) Let $(x^*, y^*) \in \Delta_n \times P'_s(G)$ be a local maximizer of (3.2).

First, we want to show that $y^* \in P_s(G)$. In order to do this, let us fix the x -component of the problem, so $x \equiv x^*$, thus deriving the following optimization problem in the single y variable:

$$(3.6) \quad \begin{aligned} \max \quad & g_G(y) \\ \text{s.t.} \quad & y \in P'_s(G), \end{aligned}$$

where the objective function is given by

$$g_G(y) = h_G(x^*, y) = (x^*)^\top [A_G + A_{\bar{G}}(y)] x^* + \Phi(x^*) + \sum_{\{i, j\} \in \bar{E}} \psi(y_{ij}).$$

Since (x^*, y^*) is a local maximizer of (3.2) by assumption, y^* must also be a local maximizer of (3.6). Moreover, the objective $g_G(y)$ has a positive-definite Hessian matrix over the polytope $P'_s(G)$, since for any $\{i, j\} \in \bar{E}$ and $y \in P'_s(G)$ we have $\frac{\partial^2}{\partial y_{ij}^2} g_G(y) = \psi''(y_{ij}) > 0$, while for any pair of distinct missing edges $\{i, j\}, \{k, l\} \in \bar{E}$ the corresponding entry is $\frac{\partial^2}{\partial y_{ij} \partial y_{kl}} g_G(y) = 0$. Hence, by Lemma 3.1 we can conclude that all local maximizers of (3.6) are extreme points of the feasible region. This in particular implies that the local maximizer y^* is a vertex of the polytope $P'_s(G)$.

On the other hand, we have already pointed out how the feasible region $P'_s(G)$ coincides with a fractional b -matching polytope in the complementary graph $\bar{G}$ with $b = (s-1)e$, see problem (2.7) above. Thus, by Lemma 2.4 we know that the vertices of $P'_s(G)$ are half-integral, i.e. $y_{ij}^* \in \{0, \frac{1}{2}, 1\}$ for every $\{i, j\} \in \bar{E}$. Let us assume, by

contradiction, that there exists an edge $\{i, j\} \in \bar{E}$ such that $y_{ij}^* = \frac{1}{2}$ is fractional. By Lemma 2.4, the fractional support of y^* contains an odd cycle, and every edge of this cycle has value $1/2$. Fix one such edge $\{i, j\} \in \bar{E}$ and let

$$\hat{y} := y^* - \varepsilon e_{ij}, \quad 0 < \varepsilon \leq \frac{1}{2},$$

where $e_{ij} \in \mathbb{R}^{\bar{m}}$ denotes the unit vector associated with the missing edge $\{i, j\}$. Since the constraints defining $P'_s(G)$ are upper-bound constraints with non-negative coefficients, $\hat{y} \in P'_s(G)$. By the convexity of ψ ,

$$\psi\left(\frac{1}{2} - \varepsilon\right) - \psi\left(\frac{1}{2}\right) \geq -\varepsilon \psi'\left(\frac{1}{2}\right).$$

Consequently,

$$\begin{aligned} g_G(\hat{y}) - g_G(y^*) &= -2\varepsilon x_i^* x_j^* + \psi\left(\frac{1}{2} - \varepsilon\right) - \psi\left(\frac{1}{2}\right) \\ &\geq -\varepsilon \left(2x_i^* x_j^* + \psi'\left(\frac{1}{2}\right)\right) > 0, \end{aligned}$$

where the last inequality follows from $2x_i^* x_j^* \leq 1/2$ and condition (R2). This contradicts the local optimality of y^* and finally shows that $y_{ij}^* \in \{0, 1\}$ for every $\{i, j\} \in \bar{E}$, so y^* is binary.

After showing this first property, let us fix the y -component of (3.2) instead, so $y \equiv y^*$. Then x^* must be a local maximizer of the function

$$h_G(x, y^*) = x^\top (A_G + A_{\bar{G}}(y^*))x + \Phi(x) = x^\top A_{G^*}x + \Phi(x)$$

over the standard simplex, where G^* is the auxiliary graph induced by $y^* \in P_s(G)$. As we can see, the considered optimization problem coincides with the regularized version of the maximum clique formulation (3.1) over G^* given in [9], thus it must hold that $x^* = x(C)$ for C a maximal clique in G^* . Since the auxiliary graph G^* is defined by adding the missing edges in G corresponding to the $\text{supp}(y^*)$, then C must also be an s -plex in G and $y_{ij}^* = 1$ for all $\{i, j\} \in \bar{E}(C)$.

(iii) $\implies$ (ii) Let $(x^*, y^*) \in \Delta_n \times P'_s(G)$ satisfy the properties outlined in the statement (iii) above. To show that it is a strict local maximizer of (3.2), we want to verify that the second-order sufficient optimality conditions (3.4) and (3.5) hold in this case.

For any missing edge $\{i, j\} \in \bar{E}$ we can compute

$$\frac{\partial}{\partial y_{ij}} h_G(x^*, y^*) = 2x_i^* x_j^* + \psi'(y_{ij}^*).$$

In the following we denote by $S := \text{supp}(y^*)$, and $\bar{S} := \bar{E} \setminus S$. Since by assumption $y_{ij}^* = 1$ for any $\{i, j\} \in \bar{E}(C)$, clearly $2x_l^* x_k^* = 0$ for any $\{l, k\} \in \bar{S}$. Moreover, by conditions (R1) and (R3) we necessarily have $\psi'(0) < 0$ and $\psi'(1) > 0$. Thus, for $\{l, k\} \in \bar{S}$ we get

$$(3.7) \quad \frac{\partial}{\partial y_{lk}} h_G(x^*, y^*) = \psi'(0) < 0,$$

while for $\{i, j\} \in S$

$$(3.8) \quad \frac{\partial}{\partial y_{ij}} h_G(x^*, y^*) = 2x_i^* x_j^* + \psi'(1) > 0,$$

where we use the fact that $y^* \in P_s(G)$ is binary.

Now, let d be a feasible direction at (x^*, y^*) , such that $d = (\bar{x}, \bar{y}) - (x^*, y^*)$ for some $(\bar{x}, \bar{y}) \in \Delta_n \times P'_s(G)$. We can separately consider $d = (d_x, d_y) = (\bar{x} - x^*, \bar{y} - y^*)$, where d_x and d_y are feasible directions for the x and y components of the optimization problem at x^* and y^* , respectively. Furthermore, for any $U \subseteq V$, we denote by $d_U := (d_x)_U$ the restriction of the x -component of d to the coordinates indexed by U .

Then we can compute

$$(3.9) \quad \begin{aligned} \nabla_x h_G(x^*, y^*)^\top d_x &= (2A_G x^* + 2A_{\bar{G}}(y^*) x^* + \nabla \Phi(x^*))^\top d_x \\ &= 2x(C)^\top A_{G^*} d_x + \nabla \Phi(x(C))^\top d_x \leq 0, \end{aligned}$$

where the inequality follows by Lemma 3.3 and Lemma 3.4 with respect to the MCP over the auxiliary graph G^* induced by y^* .

For the y -component we can instead compute

$$\begin{aligned} (\bar{y} - y^*)^\top \nabla_y h_G(x^*, y^*) &= \sum_{\{i,j\} \in \bar{E}} (\bar{y}_{ij} - y_{ij}^*) \frac{\partial}{\partial y_{ij}} h_G(x^*, y^*) \\ &= \sum_{\{i,j\} \in S} (\bar{y}_{ij} - 1) \frac{\partial}{\partial y_{ij}} h_G(x^*, y^*) + \sum_{\{i,j\} \in \bar{S}} \bar{y}_{ij} \frac{\partial}{\partial y_{ij}} h_G(x^*, y^*) \end{aligned}$$

thus

$$(3.10) \quad \nabla_y h_G(x^*, y^*)^\top d_y \leq 0$$

by inequalities (3.7), (3.8), and the fact that $\bar{y}_{ij} \leq 1$.

To summarize, combining (3.9) and (3.10) we can conclude that

$$\nabla h_G(x^*, y^*)^\top d = \nabla_x h_G(x^*, y^*)^\top d_x + \nabla_y h_G(x^*, y^*)^\top d_y \leq 0.$$

This finally proves that first-order condition (3.4) is satisfied.

Now, in order to show that the second condition (3.5) is also satisfied, notice that both terms in

$$\nabla h_G(x^*, y^*)^\top d = \nabla_x h_G(x^*, y^*)^\top d_x + \nabla_y h_G(x^*, y^*)^\top d_y$$

are nonpositive by (3.9) and (3.10). Therefore, equality can hold only if both terms vanish. By the strict part of Lemma 3.4 and the strict inequalities (3.7) and (3.8), this requires $(\bar{x}, \bar{y})$ to satisfy

- $\text{supp}(\bar{x}) \subseteq C$;
- $\bar{y}_{ij} = 1$ for every $\{i, j\} \in S$, while $\bar{y}_{lk} = 0$ for every $\{l, k\} \in \bar{S}$.

In particular, the second condition is equivalent to $\bar{y} = y^*$. We now show that these conditions are also sufficient for equality. Indeed, if $\text{supp}(\bar{x}) \subseteq C$, then $d_x = \bar{x} - x^*$ is supported on C and

$$e^\top d_C = e^\top \bar{x}_C - e^\top x_C^* = 1 - 1 = 0.$$

Since C is a clique in the auxiliary graph G^* , the components of the vector $A_{G^*} x(C)$ indexed by C are all equal to $\frac{|C|-1}{|C|}$. Moreover, condition (C3) implies that the

components of $\nabla\Phi(x(C))$ indexed by C are equal, since $x(C)$ is unchanged by every permutation of the vertices within C . Therefore, there exists $\lambda \in \mathbb{R}$ such that $\nabla_C\Phi(x(C)) = \lambda e$, and hence

$$\begin{aligned}\nabla_x h_G(x^*, y^*)^\top d_x &= 2x(C)^\top A_{G^*} d_x + \nabla\Phi(x(C))^\top d_x \\ &= \left(2\frac{|C|-1}{|C|} + \lambda\right) e^\top d_C = 0.\end{aligned}$$

Furthermore, $\bar{y} = y^*$ implies $d_y = \underline{0}$ and therefore $\nabla_y h_G(x^*, y^*)^\top d_y = 0$.

Thus, the two conditions above are also sufficient for equality, and we obtain

$$\begin{aligned}T^0((x^*, y^*), \nabla h_G(x^*, y^*)) &= \{d \in T(x^*, y^*) \mid d = (\bar{x}, \bar{y}) - (x^*, y^*), \bar{x}_{\bar{C}} = \underline{0}, \bar{y} = y^*\} \\ &= \{d = (d_x, d_y) \in T(x^*, y^*) \mid d_{\bar{C}} = \underline{0}, d_y = \underline{0}\}.\end{aligned}$$

For any $U \subseteq V$, we denote by $\nabla_{UU}^2 h_G(x^*, y^*)$ the principal submatrix of $\nabla^2 h_G(x^*, y^*)$ corresponding to the x -variables indexed by U . In particular, for any distinct pair of vertices $i, j \in C$, we have

$$\nabla_{x_i x_j}^2 h_G(x^*, y^*) = 2 + \nabla_{x_i x_j}^2 \Phi(x^*)$$

since C is a clique in the auxiliary graph G^* , while

$$\nabla_{x_i x_i}^2 h_G(x^*, y^*) = \nabla_{x_i x_i}^2 \Phi(x^*)$$

for every $i \in C$. Therefore,

$$\nabla_{CC}^2 h_G(x^*, y^*) = 2ee^\top - 2\mathbb{I}_{|C|} + \nabla_{CC}^2 \Phi(x^*).$$

Finally, let

$$d \in T^0((x^*, y^*), \nabla h_G(x^*, y^*)) \setminus \{0\}.$$

By the preceding characterization, $d_y = \underline{0}$ and $d_{\bar{C}} = \underline{0}$. Hence, only the components of d_x indexed by C contribute to the quadratic form, and we have $e^\top d_C = 0$ because d_x is a feasible direction for the standard simplex. Consequently,

$$\begin{aligned}d^\top \nabla^2 h_G(x^*, y^*) d &= d_C^\top \nabla_{CC}^2 h_G(x^*, y^*) d_C \\ &= d_C^\top (2ee^\top - 2\mathbb{I}_{|C|} + \nabla_{CC}^2 \Phi(x^*)) d_C \\ &= d_C^\top (\nabla_{CC}^2 \Phi(x^*) - 2\mathbb{I}_{|C|}) d_C < 0.\end{aligned}$$

The strict inequality follows from condition (C2), since every principal submatrix of a negative definite matrix is negative definite. This proves that the second condition (3.5) is satisfied.

To summarize, we have proved that the given point (x^*, y^*) satisfies the second-order sufficient optimality conditions (3.4) and (3.5) for the considered regularized formulation (3.2) and hence it is a strict local maximizer.

$(ii) \implies (i)$ is trivial. □

Theorem 3.5 highlights a key advantage of the regularized version (3.2), namely that, unlike the original formulation (2.3), it does not admit spurious local solutions, since every local maximizer corresponds to an s -plex in the given graph. Moreover, the x -component of the solution coincides with the characteristic vector of the detected s -plex structure.

We conclude this section with an extension of the previous result to the global equivalence case.

COROLLARY 3.6. *A feasible point $(x^*, y^*) \in \Delta_n \times P'_s(G)$ is a global maximizer of (3.2) if and only if y^* is binary and $x^* = x(C)$ for a maximum s -plex C in G ; moreover $y_{ij}^* = 1$ for every $\{i, j\} \in \bar{E}(C)$.*

Proof. Let (x^*, y^*) be a global maximizer of (3.2). By Theorem 3.5 we have that y^* is binary, $x^* = x(C)$ for an s -plex C in G and $y_{ij}^* = 1$ for every $\{i, j\} \in \bar{E}(C)$. Moreover, the corresponding value of the objective function is

$$h_G(x^*, y^*) = 1 - \frac{1}{|C|} + \Phi(x(C)).$$

Assume by contradiction that there exists an s -plex $C^1 \subseteq V$ in G such that $|C^1| > |C|$. By the heredity property of the s -plex model (see [15] and [21] for further details) there exists an s -plex $C^2 \subset C^1$ such that $|C| = |C^2|$. Therefore, let (x^1, y^1) be the solution corresponding to the s -plex C^1 , where $x^1 = x(C^1)$ and $y_{ij}^1 = 1$ if $\{i, j\} \in \bar{E}(C^1)$ and 0 otherwise, while $(x^2, y^2) = (x(C^2), y^1)$ is a feasible solution of (3.2) corresponding to C^2 . With these choices, we can compute

$$h_G(x^*, y^*) = 1 - \frac{1}{|C|} + \Phi(x(C)) = 1 - \frac{1}{|C^2|} + \Phi(x(C^2)) = h_G(x^2, y^2)$$

since Φ satisfies (C3) and $x(C^2) \in \mathcal{P}(x(C))$. Moreover, if we now fix the y component of problem (3.2) as $y \equiv y^1$, then (x^1, y^1) and (x^2, y^2) are both feasible solutions for the regularized version of the MS for the MCP over the auxiliary graph G^1 induced by y^1 , so

$$\begin{aligned} \max \quad & k_{G^1}(x) \\ \text{s.t.} \quad & x \in \Delta_n \end{aligned}$$

where the objective function is given by $k_{G^1}(x) := x^\top A_{G^1} x + \Phi(x)$. Finally, by Lemma 3.2, since $|C^2| < |C^1|$ by construction, we get

$$\begin{aligned} k_{G^1}(x^2) < k_{G^1}(x^1) &\iff h_G(x^2, y^1) < h_G(x^1, y^1) \\ &\iff h_G(x^2, y^2) < h_G(x^1, y^1) \\ &\iff h_G(x^*, y^*) < h_G(x^1, y^1) \end{aligned}$$

leading to a contradiction with the global optimality of (x^*, y^*) .

The converse follows immediately from the fact that any vector $(x^*, y^*) \in \Delta_n \times P'_s(G)$ satisfying the properties outlined in the above statement has the same value of the objective function h_G as any global maximizer of (3.2). $\square$

4. Regularization functions. We present three classes of regularization terms satisfying conditions (C1)-(C3):

- $\Phi_1(x) := \frac{\alpha}{2} \|x\|_2^2$, with $\alpha \in (0, 2)$;
- $\Phi_2(x) := \alpha_1 \|x + \varepsilon e\|_p^p$ with $\varepsilon > 0$, $p > 2$ and $0 < \alpha_1 < \frac{2}{p(p-1)(1+\varepsilon)^{p-2}}$;
- $\Phi_3(x) := \alpha_2 \sum_{i=1}^n (\exp(-\beta x_i) - 1)$ with $\beta > 0$ and $0 < \alpha_2 < \frac{2}{\beta^2}$.

Here, Φ_1 for $\alpha = 1$ is the two-norm regularization function introduced by Bomze [1]. The validity of these classes, with respect to conditions (C1)-(C3), was established in [9], to which we refer the reader for the corresponding proofs.

On the other hand, for the y -regularization term, we can define the following three families of regularization functions $\psi_\ell : Y_\ell \rightarrow \mathbb{R}$, where Y_ℓ is an open interval containing $[0, 1]$, with the indicated choices of parameters:

1. $\psi_1(t) := b_\nu t(t-1)(t^{\nu-2} + a_\nu)$, where:
 - for $\nu = 3$, $a_3 > 1$ and $b_3 > 2$;
 - for any integer $\nu \geq 4$ with $a_\nu > \frac{\nu-1}{2} \left(\frac{\nu-3}{\nu}\right)^{\nu-3}$ and $b_\nu > \frac{2^{\nu-2}}{\nu-2}$;
2. $\psi_2(t) := B_k (\exp(kt) - 1 - (\exp(k) - 1)t)$, with $k > 0$ and $B_k > \frac{1}{2}(\exp(k) - k \exp(k/2) - 1)^{-1}$;
3. $\psi_3(t) := D_q(\log(1-q)t - \log(1-tq))$, with $0 < q < 1$ and $D_q > -\frac{1}{2} \left(\log(1-q) + \frac{2q}{2-q}\right)^{-1}$.

In the following, we show that each class satisfies conditions (R1)-(R3).

First, note that all three terms satisfy (R1) trivially. We now address the remaining conditions separately for function ψ_1 . After some manipulations, we get

$$\psi_1'(t) = b_\nu(\nu t^{\nu-1} - (\nu-1)t^{\nu-2} + 2a_\nu t - a_\nu),$$

where substituting $t = \frac{1}{2}$ leads to

$$\psi_1'\left(\frac{1}{2}\right) = b_\nu(2-\nu) \left(\frac{1}{2}\right)^{\nu-1} < 0,$$

if we choose $b_\nu > 0$, for any integer $\nu > 2$ and independently of the parameter a_ν . Thus, condition (R2) is satisfied if

$$b_\nu(2-\nu) \left(\frac{1}{2}\right)^{\nu-1} < -\frac{1}{2} \iff b_\nu > \frac{2^{\nu-2}}{\nu-2}.$$

We now verify condition (R3). The case $\nu = 3$ needs to be treated separately. Here,

$$\psi_1''(t) = 2b_3(3t + a_3 - 1),$$

whose minimum over $[0, 1]$ is attained at $t = 0$. Therefore, $\psi_1''(t) > 0$ for every $t \in [0, 1]$ whenever $a_3 > 1$.

Suppose now that $\nu \geq 4$. We have

$$\psi_1''(t) = b_\nu ((\nu-1)t^{\nu-3}(\nu t - \nu + 2) + 2a_\nu).$$

Now, verifying the correctness of condition (R3) is rather non-trivial since we also want to determine the best lower bound for the parameter a_ν in order to ensure the strict convexity of ψ_1 over the unit interval. To ensure that $\psi_1''(t) > 0$ for all $t \in [0, 1]$, we first determine the minimum value of ψ_1'' on the unit interval. For this purpose, we separately consider the term of $\psi_1''(t)$ that depends explicitly on t :

$$g(t) := t^{\nu-3}(\nu t - \nu + 2).$$

Since we want to determine the minimum value of the function g , good candidate points are on the boundary of the unit interval, where we have $g(0) = 0$ and $g(1) = 2$, or on the set of stationary points. So, let us compute the derivative

$$g'(t) = (\nu-3)t^{\nu-4}(\nu t - \nu + 2) + \nu t^{\nu-3}.$$

One can easily check that

$$g'(t^*) = 0 \iff t^* = \frac{\nu-3}{\nu}$$

and in this case the function value is given by

$$g(t^*) = -\left(\frac{\nu-3}{\nu}\right)^{\nu-3} < 0$$

showing that t^* is the minimizer in the interval $[0, 1]$ for the function $g(t)$. Thus, we finally get

$$\begin{aligned} \psi_1''(t) > 0 &\iff \psi_1''(t^*) > 0 \\ &\iff -(\nu-1)\left(\frac{\nu-3}{\nu}\right)^{\nu-3} + 2a_\nu > 0 \\ &\iff a_\nu > \frac{\nu-1}{2}\left(\frac{\nu-3}{\nu}\right)^{\nu-3} \end{aligned}$$

proving the correctness of condition (R3) with the chosen parameter a_ν . Although the lower-bound sequence

$$L_\nu := \frac{\nu-1}{2}\left(\frac{\nu-3}{\nu}\right)^{\nu-3}$$

is not monotone for all ν , it diverges to $+\infty$ as $\nu \rightarrow +\infty$. Consequently, any sequence of admissible parameters satisfying $a_\nu > L_\nu$ also diverges.

In the remainder, let us instead verify the correctness of conditions (R2) and (R3) for functions ψ_2 and ψ_3 simultaneously. First, note that the function ψ_3 is well defined only for $q < 1$. Then, by computing the derivatives in the middle of the unit interval, we have

$$\begin{aligned} \psi_2'\left(\frac{1}{2}\right) &= B_k(k \exp(k/2) - \exp(k) + 1) \\ \psi_3'\left(\frac{1}{2}\right) &= D_q\left(\log(1-q) + \frac{2q}{2-q}\right) \end{aligned}$$

Since $k > 0$ and $0 < q < 1$, both quantities in parentheses are negative. Therefore, condition (R2) is satisfied provided that

$$\begin{aligned} B_k &> \frac{1}{2(\exp(k) - k \exp(k/2) - 1)} \\ D_q &> -\frac{1}{2}\left(\log(1-q) + \frac{2q}{2-q}\right)^{-1}. \end{aligned}$$

Finally, condition (R3) is always satisfied for all $t \in [0, 1]$ with the chosen parameters, because the second derivatives satisfy

$$\begin{aligned} \psi_2''(t) &= B_k k^2 \exp(kt) > 0 \\ \psi_3''(t) &= D_q \frac{q^2}{(1-qt)^2} > 0. \end{aligned}$$

Note that if the parameters B_k and D_q are non positive, then condition (R3) can never be satisfied.

5. FW variant for the MsPP. In the analysis that follows, we focus on the objective function below, which we adopt as a representative example of the class being studied:

$$(5.1) \quad h_G(x, y) = x^\top (A_G + A_{\bar{G}}(y))x + \frac{\alpha}{2} \|x\|_2^2 + \beta \sum_{\{i,j\} \in \bar{E}} y_{ij}(y_{ij} - 1)(y_{ij} + \gamma)$$

with $\alpha \in (0, 2)$, $\beta > 2$ and $\gamma > 1$. This objective is obtained by combining Φ_1 with the cubic regularization ψ_1 corresponding to $\nu = 3$, $a_3 = \gamma$ and $b_3 = \beta$. Accordingly, the conditions $\alpha \in (0, 2)$, $\beta > 2$, and $\gamma > 1$ ensure that (C1)-(C3) and (R1)-(R3) are satisfied. We note that the analysis below can also be extended to the other families of regularization functions introduced in Section 4, although the corresponding bounds depend on the chosen family and must be derived separately.

We first recall the notion of stationarity used throughout this section.

DEFINITION 5.1 (First-order stationary point). *Let $\Omega \subseteq \mathbb{R}^p$ be a nonempty compact convex set, and let F be continuously differentiable on an open set containing Ω . A point $z \in \Omega$ is called stationary for the maximization of F over Ω if*

$$\nabla F(z)^\top (u - z) \leq 0 \quad \text{for every } u \in \Omega.$$

We now describe the away-step Frank-Wolfe (AFW) method over the standard simplex, as introduced in [2] and summarized in Algorithm 5.1 below, since it plays a key role in the development of a tailored FW algorithm to solve the MsPP.

In a general setting, we consider the optimization problem

$$\begin{aligned} \max \quad & f(w) \\ \text{s.t.} \quad & w \in \Delta_n \end{aligned}$$

where we assume f to be a continuously differentiable function on the standard simplex Δ_n .

Algorithm 5.1 AFW on the standard simplex

- 1: Initialize $w^{(0)} \in \Delta_n$, $k := 0$.
 - 2: Set $s^{(k)} := e_{\hat{i}}$, with $\hat{i} \in \arg \max_{i=1,\dots,n} \nabla_i f(w^{(k)})$, and $d_{FW}^{(k)} := s^{(k)} - w^{(k)}$.
 - 3: **If** $w^{(k)}$ is stationary, **STOP**.
 - 4: Set $v^{(k)} := e_{\hat{j}}$, with $\hat{j} \in \arg \min_{j \in S_k} \nabla_j f(w^{(k)})$, where $S_k := \{j : w_j^{(k)} > 0\}$, and $d_A^{(k)} := w^{(k)} - v^{(k)}$.
 - 5: **If** $\nabla f(w^{(k)})^\top d_{FW}^{(k)} \geq \nabla f(w^{(k)})^\top d_A^{(k)}$ **then** $d^{(k)} := d_{FW}^{(k)}$ and $\alpha_k^{\max} := 1$; **else**

$$d^{(k)} := d_A^{(k)} \text{ and } \alpha_k^{\max} := \frac{w_{\hat{j}}^{(k)}}{1 - w_{\hat{j}}^{(k)}}.$$
 - 6: Choose $\alpha_k \in (0, \alpha_k^{\max}]$ using a suitable criterion.
 - 7: Update $w^{(k+1)} := w^{(k)} + \alpha_k d^{(k)}$.
 - 8: Set $k := k + 1$ and go to step 2.
-

As we can see from Algorithm 5.1, the AFW method at each iteration chooses between the standard FW direction $d_{FW}^{(k)}$ computed in step 2 and the away-step direction $d_A^{(k)}$ calculated in step 4, with the criterion in step 5. In particular, this second direction shifts the weight away from the worst vertex to the other vertices used to

represent the current iterate $w^{(k)}$. For the standard simplex case, this vertex coincides with the canonical basis vector that selects the lowest component of the gradient, or equivalently, the vertex that minimizes the first-order approximation of f given by $z \rightarrow f(w^{(k)}) + \nabla f(w^{(k)})^\top (z - w^{(k)})$ over the support of the current iterate $w^{(k)}$. The upper bound on the step-size α_k^{max} is then the largest feasible step-size for the away direction given the boundary conditions, so, when the algorithm performs an away-step, we have that either the support of the current iterate stays the same or decreases by one. In the latter case, $\alpha_k = \alpha_k^{max}$ and we remove the component whose index is associated with the away direction, generating a new point on the boundary of the current minimal face. This drop in dimension is what allows the AFW method to quickly identify low-dimensional faces containing $w^{(k)}$, and therefore the correct support of an optimal local solution. On the other hand, when the algorithm performs a standard FW step, it selects the vertex that maximizes the first-order approximation of f , which corresponds to the highest component of the gradient, and adds it to the support of the current iterate $w^{(k)}$. These properties are fundamental for the active-set identification of the AFW method.

Moreover, we often require the following lower bound on the step-sizes:

$$(5.2) \quad \alpha_k \geq \bar{\alpha}_k := \min \left(\alpha_k^{max}, c \frac{\nabla f(w^{(k)})^\top d^{(k)}}{\|d^{(k)}\|_2^2} \right)$$

for some $c > 0$ and, for convergence results, we need the following sufficient increase condition for some constant $\rho > 0$:

$$(5.3) \quad f(w^{(k)} + \alpha_k d^{(k)}) - f(w^{(k)}) \geq \rho \bar{\alpha}_k \nabla f(w^{(k)})^\top d^{(k)}$$

which generalizes properties of exact and Armijo line search. We refer the reader to [2] for further details and convergence properties of the AFW algorithm.

At this point, we can introduce our tailored Frank-Wolfe variant for the maximum s -plex formulation (3.2), namely FWsP, as presented in Algorithm 5.2, which exploits the product domain structure of the problem at hand by employing separate updating rules for the two blocks of variables x and y .

Algorithm 5.2 FWsP – FW variant for the s -plex

- 1: Initialize $x^{(0)} \in \Delta_n$ and $y^{(0)} \in P'_s(G)$, $k := 0$.
 - 2: **If** $(x^{(k)}, y^{(k)})$ is stationary, **then STOP**.
 - 3: Compute $x^{(k+1)}$ by applying one iteration of Algorithm 5.1 with $w^{(0)} := x^{(k)}$ and $f(w) := h_G(w, y^{(k)})$.
 - 4: Let $y^{(k+1)} \in \arg \max_{y \in P'_s(G)} \nabla_y h_G(x^{(k+1)}, y^{(k)})^\top y$.
 - 5: Set $k := k + 1$ and go to step 2.
-

In particular, at each iteration, the method first performs one AFW iteration on the standard simplex for the x variable, step 3, and then a full FW step on the y variable, step 4, which basically means solving a linear optimization problem. Such a problem is the linear relaxation of a maximum-weight b -matching problem on the complement graph $\bar{G}$, with capacity vector $b = (s - 1)e$ and edge-weight vector $\nabla_y h_G(x^{(k+1)}, y^{(k)})$. Throughout the following analysis, we assume that the linear optimization oracle used in step 4 always returns an extreme-point optimal solution. This entails no loss of generality, since a linear function over the nonempty compact polytope $P'_s(G)$ always admits an optimal solution at an extreme point.

Consequently, $y^{(k)}$ is a vertex of $P'_s(G)$ for every $k \geq 1$, and, by Lemma 2.4, we have that $y_{ij}^{(k)} \in \{0, \frac{1}{2}, 1\}$ for every $\{i, j\} \in \bar{E}$. Notice that the order in which these updates are performed is essential, since it allows us to compute the $y^{(k+1)}$ component of the new iterate as a function of the newly computed $x^{(k+1)}$ component, as highlighted in step 4 of Algorithm 5.2. Furthermore, as stated in the next proposition, the sequence $\{y^{(k)}\}_k$ is ultimately constant. This finally allows us to obtain convergence results by applying the general properties of the AFW method on the standard simplex for the x component, as proved in [2].

PROPOSITION 5.2. *In Algorithm 5.2, if $h_G(x^{(k+1)}, y^{(k)}) \geq h_G(x^{(k)}, y^{(k)})$ for every $k \geq 0$, then the sequence $\{y^{(k)}\}_k$ converges to a binary vertex of $P'_s(G)$ and can change at most $\frac{2}{\beta(\gamma-1)} \left(1 - \frac{2-\alpha}{2|C^*|} - \frac{\alpha}{2n} + \frac{\beta\bar{m}}{4}(\gamma+1)\right)$ times, with an s -plex C^* in G of maximum cardinality.*

Proof. First, we show that the function $y \rightarrow h_G(x, y)$ is strongly convex on $[0, 1]^{\bar{m}} \supset P'_s(G)$ with a convexity coefficient $\mu = 2\beta(\gamma - 1) > 0$, since $\gamma > 1$. Since this function is separable with respect to y , in the one-dimensional case we can consider $f(t) := w(x)t + \beta t(t - 1)(t + \gamma) + c(x)$, $\forall t \in [0, 1]$, for fixed x . Although $w(x)$ and $c(x)$ may depend on x , they are respectively a coefficient of a linear term and a constant term with respect to t ; therefore they do not contribute to the second derivative. Indeed, we get $f''(t) = 2\beta(3t + \gamma - 1) \geq 2\beta(\gamma - 1)$, since f'' reaches its minimum in $[0, 1]$ for $t = 0$. This shows that f is strongly convex in $[0, 1]$ with the strong-convexity coefficient $\mu := 2\beta(\gamma - 1)$. Because the Hessian with respect to y of $h_G(x, y)$ is diagonal with entries $f''(y_{ij})$, the same bound applies in $\bar{m}$ dimensions.

Now assume that $y^{(k)}$ and $y^{(k+1)}$ are two distinct vertices of $P'_s(G)$. Let $z^{(k)} = (x^{(k+1)}, y^{(k)})$ and $z^{(k+1)} = (x^{(k+1)}, y^{(k+1)})$. By the strong convexity just established for $y \rightarrow h_G(x, y)$, we have

$$\begin{aligned} h_G(z^{(k+1)}) - h_G(z^{(k)}) &\geq \nabla h_G(z^{(k)})^\top (z^{(k+1)} - z^{(k)}) + \frac{\mu}{2} \|z^{(k+1)} - z^{(k)}\|^2 \\ &= \nabla_y h_G(z^{(k)})^\top (y^{(k+1)} - y^{(k)}) + \frac{\mu}{2} \|y^{(k+1)} - y^{(k)}\|^2 \\ &\geq \frac{\mu}{2} \|y^{(k+1)} - y^{(k)}\|^2 \geq \frac{\mu}{2} \cdot \frac{1}{2} = \frac{\mu}{4}. \end{aligned}$$

The first-order term is non-negative because $y^{(k+1)}$ is an optimal solution of the linear problem in step 4 of Algorithm 5.2 and $y^{(k)}$ is feasible. Moreover, the squared distance between two distinct vertices of $P'_s(G)$ is at least $1/2$. Indeed, the vertices are half-integral, and a single difference of size $1/2$ would contradict the odd-cycle characterization of their fractional support. The previous inequality accounts only for the y update. To obtain an increase between two consecutive iterates, we use the assumption on the x update. As a consequence, whenever $y^{(k+1)} \neq y^{(k)}$, we have $h_G(x^{(k+1)}, y^{(k+1)}) - h_G(x^{(k)}, y^{(k)}) \geq \frac{\mu}{4}$. Hence, these increases can be telescoped over

all iterations in which the y -component changes. Therefore, $y^{(k)}$ can change at most

$$\begin{aligned}
& \max_{(x,y) \in \Delta_n \times P'_s(G)} \frac{h_G(x,y) - h_G(x^{(0)}, y^{(0)})}{\frac{\mu}{4}} \\
& \leq \frac{4}{\mu} \left(\max_{(x,y) \in \Delta_n \times P'_s(G)} h_G(x,y) - \min_{(x,y) \in \Delta_n \times P'_s(G)} h_G(x,y) \right) \\
& \leq \frac{4}{\mu} \left(1 - \frac{2-\alpha}{2|C^*|} - \frac{\alpha}{2n} + \frac{\beta \bar{m}}{4}(\gamma+1) \right) \\
& = \frac{2}{\beta(\gamma-1)} \left(1 - \frac{2-\alpha}{2|C^*|} - \frac{\alpha}{2n} + \frac{\beta \bar{m}}{4}(\gamma+1) \right)
\end{aligned}$$

times, where we used Corollary 3.6 with an s -plex C^* of maximum cardinality in G , and the fact that

$$\begin{aligned}
\min_{(x,y) \in \Delta_n \times P'_s(G)} h_G(x,y) & \geq \frac{\alpha}{2n} + \min_{y \in P'_s(G)} \beta \sum_{\{i,j\} \in \bar{E}} y_{ij}(y_{ij}-1)(y_{ij}+\gamma) \\
& \geq \frac{\alpha}{2n} + \beta \bar{m} \min_{t \in [0,1]} t(t-1)(t+\gamma) \\
& = \frac{\alpha}{2n} - \beta \bar{m} \max_{t \in [0,1]} t(1-t)(t+\gamma) \\
& \geq \frac{\alpha}{2n} - \beta \bar{m} \max_{t \in [0,1]} t(1-t) \cdot \max_{t \in [0,1]} (t+\gamma) \\
& = \frac{\alpha}{2n} - \frac{\beta \bar{m}}{4}(\gamma+1).
\end{aligned}$$

This is obtained by considering that the adjacency contribution to h_G is non-negative and $\|x\|_2^2 \geq 1/n$ for every $x \in \Delta_n$. Finally, if after this number of iterations the y -component of the solution corresponds to a fractional vertex of $P'_s(G)$, then the algorithm simply sets all fractional components to zero in the subsequent iteration. This occurs because the weights given by the gradient and associated with the fractional components of the solution are strictly negative in this case, since $\beta > 2$. $\square$

Remark 5.3. For $\beta > 0$ and $\gamma > 1$, the strong-convexity argument still implies finite stabilization of the y -sequence, using only the boundedness of the objective on the compact feasible set. The stronger condition $\beta > 2$ is needed to ensure that the stabilized vertex is binary and to invoke the equivalence results of Section 3. We also observe that the monotonicity assumption in Proposition 5.2 is satisfied under the hypotheses of the convergence result below. Indeed, condition (5.3), together with the fact that the direction selected by the AFW method is an ascent direction, implies $h_G(x^{(k+1)}, y^{(k)}) \geq h_G(x^{(k)}, y^{(k)})$. Consequently, Proposition 5.2 applies under the assumptions of Theorem 5.4.

We finally show the convergence of the newly proposed method.

THEOREM 5.4. *Let $\{(x^{(k)}, y^{(k)})\}_k$ be the sequence of points generated by Algorithm 5.2, with regularization coefficient $\alpha = 1$. If conditions (5.2) and (5.3) hold on the step-sizes, then the sequence converges to a first-order stationary point in the sense of Definition 5.1 and identifies the support of its x -component in finite time.*

Proof. By Proposition 5.2, after finitely many iterations, the y -variable generated by Algorithm 5.2 stabilizes at a binary vertex $y^* \in P'_s(G)$. Thereafter, the algorithm reduces to the AFW method applied over the standard simplex to maximize the

quadratic function $h_G(x, y^*) = x^\top A_{G^*} x + \frac{\alpha}{2} \|x\|_2^2$. As already noticed, the considered optimization problem corresponds to a regularized version of the MS for the MCP over the auxiliary graph G^* induced by the binary vector y^* .

For $\alpha = 1$, the fixed- y objective can be written as $h_G(x, y^*) = x^\top Q_{G^*} x$, with $Q_{G^*} := A_{G^*} + \frac{1}{2} \mathbb{I}_n$. By the argument used in the proof of Theorem 12 in [4], every principal submatrix of Q_{G^*} is nonsingular. Consequently, by Theorem 8 therein, there can be at most one stationary point in the relative interior of each face of the standard simplex when y^* is fixed. Furthermore, Theorem 2.5 in [13] implies that there exist finitely many stationary points of the fixed- y problem, all satisfying strict complementarity and having distinct supports. Therefore, all the assumptions of Theorem 4.5 in [2] are satisfied, and the sequence $\{x^{(k)}\}_k$ converges to a stationary point x^* of the fixed- y problem and identifies its support in finite time.

It remains to verify stationarity with respect to the y -component. Since $y^{(k+1)} = y^{(k)} = y^*$ for every sufficiently large k , step 4 of Algorithm 5.2 gives

$$\nabla_y h_G(x^{(k+1)}, y^*)^\top (\bar{y} - y^*) \leq 0 \quad \text{for every } \bar{y} \in P'_s(G).$$

Passing to the limit, we obtain

$$\nabla_y h_G(x^*, y^*)^\top (\bar{y} - y^*) \leq 0 \quad \text{for every } \bar{y} \in P'_s(G).$$

Combining this inequality with the stationarity of x^* for the fixed- y problem shows that $\nabla h_G(x^*, y^*)^\top ((\bar{x}, \bar{y}) - (x^*, y^*)) \leq 0$ for every $(\bar{x}, \bar{y}) \in \Delta_n \times P'_s(G)$. Hence, (x^*, y^*) is a first-order stationary point in the sense of Definition 5.1. Finally, since $y^{(k)}$ becomes constant and the AFW method identifies $\text{supp}(x^*)$ in finite time, the complete sequence converges to (x^*, y^*) and identifies the support of its x -component in finite time. $\square$

Although all local maximizers of formulation (3.2) correspond to s -plexes by Theorem 3.5, Theorem 5.4 does not rule out convergence of Algorithm 5.2 to saddle points. This possibility, however, does not appear to be problematic in practice. First, in our numerical tests, the method always converged to a local maximizer, in line with studies showing that many first-order methods avoid saddle points with high probability; see, e.g., [4, 12]. Second, while local maximizers may be locally attractive, saddle points are typically unstable and are unlikely to attract the iterates of a strictly monotone ascent method under standard regularity conditions. Lastly, for our specific problem, there are inexpensive strategies to escape saddle points even when the starting point is “unlucky,” for example, when it coincides with a saddle point.

We now describe one such strategy for Algorithm 5.2, to be applied after the y -component has stabilized at a binary vertex y^* , and if the $\text{supp}(x^{(k)})$ is not yet a clique in the auxiliary graph G^* . The first step is to select a pair of nonadjacent vertices $\{i, j\} \in \binom{\text{supp}(x^{(k)})}{2} \setminus (E \cup \text{supp}(y^*))$, an operation that requires checking at most $\binom{|\text{supp}(x^{(k)})|}{2}$ entries of the adjacency matrix of G^* . Assuming without loss of generality that $\frac{\partial}{\partial x_i} h_G(x^{(k)}, y^*) \leq \frac{\partial}{\partial x_j} h_G(x^{(k)}, y^*)$, the second step is to replace $x_i^{(k)}$ and $x_j^{(k)}$ with $(1 - \varepsilon)x_i^{(k)}$ and $x_j^{(k)} + \varepsilon x_i^{(k)}$, respectively, for some fixed $\varepsilon \in (0, 1]$. Denoting the resulting point by $\tilde{x}$ and setting $\delta := \varepsilon x_i^{(k)} > 0$, the quadratic structure of the objective gives

$$h_G(\tilde{x}, y^*) - h_G(x^{(k)}, y^*) = \delta \left(\frac{\partial h_G}{\partial x_j}(x^{(k)}, y^*) - \frac{\partial h_G}{\partial x_i}(x^{(k)}, y^*) \right) + \alpha \delta^2 > 0,$$

where the last term follows from the fact that $\{i, j\}$ is a missing edge of G^* . The point $(\tilde{x}, y^*)$ can therefore be used as a new starting point for Algorithm 5.2. In particular, if $(x^{(k)}, y^*)$ is a saddle point, a strictly monotone restart cannot converge back to the same saddle point.

Let y^* be a fixed binary vertex of $P'_s(G)$ found by Algorithm 5.2 after a finite number of iterations, and denote as usual by G^* the auxiliary graph induced by y^* . For a clique C of G^* we define $m(C, G^*) = \min_{v \in V \setminus C} |C| - |N^*(v) \cap C|$, that is, the minimum number of missing edges between an outside vertex v and C that would have to be added to make $C \cup \{v\}$ a clique.

We now give an explicit bound on how close the sequence $\{x^{(k)}\}_k$ generated by Algorithm 5.2 must be to the characteristic vector $x(C)$ for identification to occur.

PROPOSITION 5.5. *Let $\{(x^{(k)}, y^{(k)})\}_k$ be a sequence generated by Algorithm 5.2, y^* a binary vertex of $P'_s(G)$ and C a clique in the auxiliary graph G^* . Let δ_{max} be the largest eigenvalue of the the adjacency matrix related to the auxiliary graph $A_{G^*} := A_G + A_{\bar{G}}(y^*)$. Let k^* be a fixed index in $\mathbb{N}_0$ and I^c be the components of $\text{supp}(x^{(k^*)})$ with index not in C , and let $L := 2\delta_{max} + \alpha$. Assume that $y^{(k^*+j)} = y^*$ is constant for $0 \leq j \leq |I^c|$, that (5.2) holds for $c = \frac{1}{L}$, and that*

$$(5.4) \quad \|x^{(k^*)} - x(C)\|_1 \leq \frac{m_\alpha(C, G^*)}{m_\alpha(C, G^*) + 2|C|\delta_{max} + |C|\alpha}$$

for $m_\alpha(C, G^*) = m(C, G^*) - 1 + \frac{\alpha}{2}$. Then, the $\text{supp}(x^{(k^*+|I^c|)}) = C$.

Proof. Since after a certain number of iterations the y component of the sequence is fixed, the proof of the above result is identical to the proof reported in [3, Proposition 5.4, Appendix]. $\square$

Remark 5.6. It is well-known that for any graph the largest eigenvalue δ_{max} of the adjacency matrix is less than or equal to d_{max} as the maximum vertex degree of G^* (see, e.g., [6]). Then condition (5.4) can be replaced by

$$\|x^{(k^*)} - x(C)\|_1 \leq \frac{m_\alpha(C, G^*)}{m_\alpha(C, G^*) + 2|C|d_{max} + |C|\alpha}.$$

6. Numerical results. This section presents the results of the computational experiments conducted with the newly proposed FWsP method, Algorithm 5.2. All experiments use the specific objective function introduced in (5.1). We also provide a numerical comparison between the original formulation (2.3) and its regularized counterpart (3.2). We remark that, although the method finds local maxima whose x supports are s -plexes rather than necessarily maximum s -plexes in practice, it can be used as a heuristic to obtain lower bounds on the s -plex number, either independently or within a global optimization scheme. The experiments were conducted with the following objectives:

- to show that the proposed FW variant performs better than the general-purpose solver CONOPT when applied to the regularized problems, both in terms of CPU time and the quality of the solutions found;
- to show that our approach, based on solving the regularized formulation (3.2) with the FWsP algorithm, finds solutions of comparable quality to those obtained by applying the general-purpose solver CONOPT directly to the original cubic MS formulation (2.3), while requiring shorter running times and avoiding failures in the identification of maximal s -plexes.

In the experiments, the regularization parameters were set to $\alpha = 1$, $\gamma = 1 + \frac{1}{n}$ and $\beta = \frac{1}{10 \cdot \omega(G)^2}$, where $\omega(G)$ is the best-known clique number¹ for the given graph instance G . An intuitive motivation for this empirical choice of β can be obtained by revisiting the argument used in the proof of Theorem 3.5. For a fractional component $y_{ij}^* = 1/2$, that argument excludes local optimality whenever $2x_i^*x_j^* + \psi'(\frac{1}{2}) < 0$. When considering the regularization function in (5.1), we have $\psi'(1/2) = -\beta/4$, and hence the above inequality is equivalent to $\beta > 8x_i^*x_j^*$. The condition $\beta > 2$ follows from the uniform bound $x_i^*x_j^* \leq 1/4$ over the entire standard simplex. In practice, however, we expect the iterates to approach the normalized characteristic vector of a maximum s -plex C^* . At such a point, the nonzero products $x_i^*x_j^*$ are equal to $1/|C^*|^2$. Moreover, since every clique is also an s -plex, $|C^*| \geq \omega(G)$ and therefore $\frac{1}{|C^*|^2} \leq \frac{1}{\omega(G)^2}$. This observation motivates, at a heuristic level, scaling β as $\frac{1}{10 \cdot \omega(G)^2}$ in our experiments, thereby avoiding an excessively strong regularization effect. Although this argument does not provide the uniform guarantee ensured by $\beta > 2$, it offers an intuitive justification for the parameter scaling adopted in the computational experiments.

We used the step-size $\alpha_k = \bar{\alpha}_k$ with $\bar{\alpha}_k$ given by (5.2) for $c = 1$, corresponding to an estimate for the Lipschitz constant L of $\nabla_x h_G$. The Short Step Chain (SSC) procedure was adopted in the algorithm, as a gradient recycling scheme to use first-order information more efficiently (see [17] for details). The code was written in MATLAB, and the tests were performed on an Intel(R) Core(TM) Ultra 7 155H (1.40 GHz) 32 GB. The 50 graph instances that we used in our experiments are taken from the second DIMACS Implementation Challenge [10]. These graphs are a common benchmark to assess the performance of algorithms for maximum relaxed clique problems (see [3] and [20] and the references therein), and the particular instances we selected coincide with the ones employed in [20] to ensure a fair comparison of the performance. Following the scheme adopted in [20], for each pair (G, s) , with G a graph from the 50 selected instances and s an integer in $\{2, 3, 4, 5\}$, we imposed a cumulative time limit of 600 seconds. Within this time budget, the method was repeatedly run from independently generated random starting points until either 100 restarts had been completed or the time limit had been reached, whichever occurred first. In particular, the x -component of the starting point was generated with MATLAB function `rand()` and then normalized by dividing it by its sum. A slightly different procedure was adopted for the choice of the starting y -component: random values, again generated by the `rand()` function of MATLAB, were assigned only to the entries of y corresponding to a maximum fractional $(s-1)e$ -matching in the complementary graph of the instance. This initialization ensures that the resulting y -component is both feasible and randomly generated. To measure approximate stationarity in the numerical experiments, we use the FW gap associated with Definition 5.1. Let F denote the objective function of the formulation under consideration and let $\Omega := \Delta_n \times P'_s(G)$. For any $z = (x, y) \in \Omega$, we define

$$\begin{aligned} g_{\text{FW}}^F(z) &:= \max_{(u,v) \in \Omega} \nabla F(x, y)^\top ((u, v) - (x, y)), \\ &= \max_{u \in \Delta_n} \nabla_x F(x, y)^\top (u - x) + \max_{v \in P'_s(G)} \nabla_y F(x, y)^\top (v - y). \end{aligned}$$

The gap is non-negative, and $g_{\text{FW}}^F(z) = 0$ if and only if z is stationary in the sense of

¹If the clique number is not known, this scaling would require replacing $\omega(G)$ with an available estimate.

Definition 5.1. Finally, following [3], termination of the FWsP algorithm required two conditions: the FW gap defined above had to be less than or equal to the prescribed tolerance $\varepsilon := 2 \cdot 10^{-3}$, and the current support of the x -component had to induce an s -plex. With this stopping rule, the FWsP method satisfied both conditions in all the tested runs, and no support-identification failures were observed. These numerical results are consistent with the local equivalence properties established in this paper.

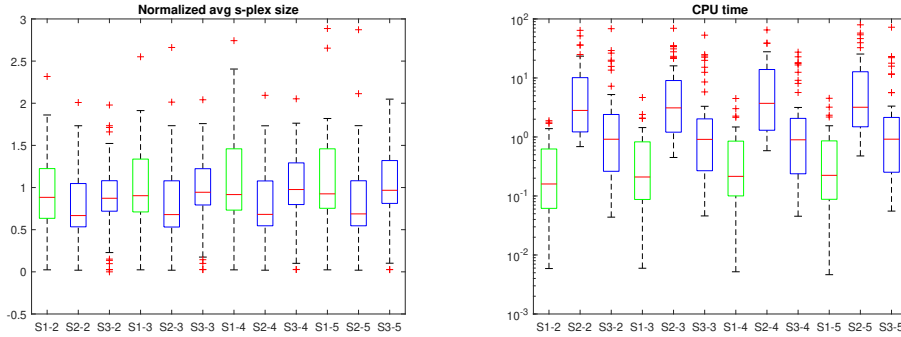
For the CONOPT solver the default random starting values and stopping criterion were used. We note that no postprocessing or repair procedure was applied to the solutions returned by either CONOPT or FWsP. This choice was deliberate, since the purpose of the comparison is to assess the intrinsic ability of the original and regularized formulations to produce solutions whose supports directly identify s -plexes. Applying a repair procedure could improve the final combinatorial solutions, but would conceal the support-identification failures that the proposed regularization is designed to eliminate.

In the box plots presented below, each series consists of 50 values corresponding to the aggregated data for the runs performed on the 50 instances. Here we list the strategies considered in our experiments:

- Strategy 1 (abbreviated as $S1$) is our proposed approach, in which the regularized formulation (3.2) is solved using the FWsP algorithm (Algorithm 5.2) with the parameters specified above;
- Strategy 2 (abbreviated as $S2$) consists of solving the regularized formulation (3.2) directly with the general-purpose solver CONOPT;
- Strategy 3 (abbreviated as $S3$) consists of solving the original cubic MS formulation (2.3) directly with CONOPT, in line with the approach in [20].

The red lines represent the median of the values in each series, and the boxes extend from the 25th percentile of the observed data to the 75th percentile. The whiskers extend to the most extreme observations within the prescribed interquartile range, while values outside this range are displayed as outliers. In Figure 2a, the box plot $Si-j$ represents the distribution of the average cardinality of the s -plex found by Strategy Si with parameter $s = j$, divided by the best-known clique number of the instance. To distinguish the algorithms visually, the box plots corresponding to the FWsP method ($S1$) are shown in green, whereas those corresponding to CONOPT ($S2$ and $S3$) are shown in blue. Notice that some normalized values exceed 1. This is expected because, for $s > 1$, the s -plex model is a relaxation of the clique model, and an s -plex may therefore have cardinality larger than the clique number. The solutions obtained by applying the FWsP algorithm to the regularized formulation (3.2) ($S1$) are generally larger than those returned by CONOPT for the same formulation ($S2$). Moreover, the median solution sizes obtained by $S1$ and by CONOPT applied to the original cubic MS formulation (2.3) ($S3$) are comparable.

In Figure 2b, $Si-j$ represents the distribution of average CPU times in seconds (on a logarithmic scale, explaining the asymmetry of the box plots) of Strategy Si for parameter $s = j$. As in the previous figure, green denotes the FWsP method, whereas blue denotes the CONOPT solver. Strategy $S1$ outperforms $S2$ and $S3$ in terms of computational efficiency, exhibiting a substantially lower median execution time. Moreover, $S1$ completes all 100 runs within the prescribed time limit of 600 seconds for every tested instance, whereas $S2$ and $S3$ fail to reach 100 runs within this time limit for some instances. Summarizing, the reported results indicate that the proposed Strategy $S1$, based on the regularized version (3.2) and a tailored FW variant, is a viable alternative when searching for maximal s -plexes in a graph. We refer the reader to the GitHub repository for detailed numerical results.



(a) The box plot $Si-j$ reports the average s -plex size found by strategy Si for $s = j$, divided by the best-known clique number of the instance.

(b) The box plot $Si-j$ reports the average running time (in seconds) of strategy Si for $s = j$.

Fig. 2: Comparison of the average s -plex size and running time for the considered strategies.

7. Conclusion. We introduced a general continuous regularization framework for the maximum s -plex problem, designed to eliminate the spurious maximizers admitted by the cubic Motzkin-Straus extension [20]. By exploiting the two-block structure of the formulation (i.e., the vertex variables and the missing-edge variables) together with the half-integral, odd-cycle structure of the fractional b -matching polytope, we derived sufficient conditions under which every local maximizer corresponds to an s -plex and every global maximizer yields a maximum s -plex. We also presented several families of regularization functions satisfying these conditions.

To solve the resulting formulation, we developed a tailored block-coordinate first-order algorithm that combines away-step Frank-Wolfe updates on the standard simplex with linear optimization over the fractional b -matching polytope. Under the stated assumptions, we established finite stabilization of the missing-edge variables, convergence to first-order stationary points, and finite identification of the support of the vertex variables. Experiments on standard benchmark graphs indicate that the tested regularized formulation avoids the support-identification failures observed for the original cubic model and returns high-quality s -plexes at a low computational cost.

Several directions merit further investigation. A first question is whether the theoretical guarantees can be retained under weaker or data-dependent regularization conditions, leading to adaptive parameter-selection rules. It would also be valuable to extend the framework to other clique-relaxation and network-clustering models. From a global-optimization perspective, the proposed method could be embedded in multistart, decomposition, or branch-and-bound schemes, where its fast local solutions could provide strong lower bounds. Finally, recent exact continuous formulations for broader classes of graph problems [5] suggest opportunities for combining structure-aware regularization with scalable and parallel first-order algorithms.

REFERENCES

- [1] I. M. BOMZE, *Evolution towards the maximum clique*, J. Global Optim., 10 (1997), pp. 143–164, <https://doi.org/10.1023/A:1008230200610>.
- [2] I. M. BOMZE, F. RINALDI, AND D. ZEFFIRO, *Active set complexity of the away-step Frank–Wolfe algorithm*, SIAM J. Optim., 30 (2020), pp. 2470–2500, <https://doi.org/10.1137/19M1309419>.
- [3] I. M. BOMZE, F. RINALDI, AND D. ZEFFIRO, *Fast cluster detection in networks by first-order optimization*, SIAM J. Math. Data Sci., 4 (2022), pp. 285–305, <https://doi.org/10.1137/21M1408658>.
- [4] I. M. BOMZE AND V. STIX, *Genetic engineering via negative fitness: Evolutionary dynamics for global optimization*, Ann. Oper. Res., 89 (1999), pp. 297–318, <https://doi.org/10.1023/A:1018935925761>.
- [5] S. BUTENKO, M. PARDALOS, AND A. RAVINDRAN, *Continuous formulations for combinatorial optimization problems based on the probabilistic method*, J. Global Optim., 95 (2026), pp. 65–109, <https://doi.org/10.1007/s10898-026-01620-x>.
- [6] D. CVETKOVIĆ AND P. ROWLINSON, *The largest eigenvalue of a graph: A survey*, Linear Multilinear Algebra, 28 (1990), pp. 3–33, <https://doi.org/10.1080/03081089008818026>.
- [7] M. FRANK AND P. WOLFE, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., 3 (1956), pp. 95–110, <https://doi.org/10.1002/nav.3800030109>.
- [8] T. GSCHWIND, S. IRNICH, F. FURINI, AND R. WOLFLER CALVO, *Social network analysis and community detection by decomposing a graph into relaxed cliques*, Tech. Report LM-2015-07, Gutenberg School of Management and Economics, Johannes Gutenberg University Mainz, Mainz, Germany, 2015.
- [9] J. T. HUNGERFORD AND F. RINALDI, *A general regularized continuous formulation for the maximum clique problem*, Math. Oper. Res., 44 (2019), pp. 1161–1173, <https://doi.org/10.1287/moor.2018.0954>.
- [10] *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge, October 11–13, 1993*, vol. 26 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, Providence, RI, 1996.
- [11] E. L. JOHNSON, *Programming in networks and graphs*, Tech. Report ORC 65-1, Operations Research Center, University of California, Berkeley, CA, 1965.
- [12] J. D. LEE, I. PANAGEAS, G. PILIOURAS, M. SIMCHOWITZ, M. I. JORDAN, AND B. RECHT, *First-order methods almost always avoid strict saddle points*, Math. Program., 176 (2019), pp. 311–337, <https://doi.org/10.1007/s10107-019-01374-3>.
- [13] A. MASSARO, M. PELILLO, AND I. M. BOMZE, *A complementary pivoting approach to the maximum weight clique problem*, SIAM J. Optim., 12 (2002), pp. 928–948, <https://doi.org/10.1137/S1052623400381413>.
- [14] T. S. MOTZKIN AND E. G. STRAUS, *Maxima for graphs and a new proof of a theorem of Turán*, Canad. J. Math., 17 (1965), pp. 533–540, <https://doi.org/10.4153/CJM-1965-053-6>.
- [15] J. PATILLO, N. YOUSSEF, AND S. BUTENKO, *On clique relaxation models in network analysis*, Eur. J. Oper. Res., 226 (2013), pp. 9–18, <https://doi.org/10.1016/j.ejor.2012.10.021>.
- [16] M. PELILLO AND A. JAGOTA, *Feasible and infeasible maxima in a quadratic program for maximum clique*, J. Artif. Neural Netw., 2 (1995), pp. 411–420.
- [17] F. RINALDI AND D. ZEFFIRO, *Avoiding bad steps in Frank–Wolfe variants*, Comput. Optim. Appl., 84 (2023), pp. 225–264, <https://doi.org/10.1007/s10589-022-00434-3>.
- [18] R. T. ROCKAFELLAR, *Convex Analysis*, vol. 28 of Princeton Mathematical Series, Princeton University Press, Princeton, NJ, 1970.
- [19] S. B. SEIDMAN AND B. L. FOSTER, *A graph-theoretic generalization of the clique concept*, J. Math. Sociol., 6 (1978), pp. 139–154, <https://doi.org/10.1080/0022250X.1978.9989883>.
- [20] V. STOZHOKOV, A. BUCHANAN, S. BUTENKO, AND V. BOGINSKI, *Continuous cubic formulations for cluster detection problems in networks*, Math. Program., 196 (2022), pp. 279–307, <https://doi.org/10.1007/s10107-020-01572-4>.
- [21] M. YANNAKAKIS, *Node- and edge-deletion NP-complete problems*, in Proceedings of the Tenth Annual ACM Symposium on Theory of Computing, New York, NY, 1978, Association for Computing Machinery, pp. 253–264, <https://doi.org/10.1145/800133.804355>.
- [22] H. YU, A. PACCANARO, V. TRIFONOV, AND M. GERSTEIN, *Predicting interactions in protein networks by completing defective cliques*, Bioinformatics, 22 (2006), pp. 823–829, <https://doi.org/10.1093/bioinformatics/btl014>.