Fast cluster detection in networks by first-order optimization

Immanuel M. Bomze* Francesco Rinaldi† Damiano Zeffiro‡

March 29, 2021

Abstract

Cluster detection plays a fundamental role in the analysis of data. In this paper, we focus on the use of s-defective clique models for network-based cluster detection and propose a nonlinear optimization approach that efficiently handles those models in practice. In particular, we introduce an equivalent continuous formulation for the problem under analysis, and we analyze some tailored variants of the Frank-Wolfe algorithm that enable us to quickly find maximal s-defective cliques. The good practical behavior of those algorithmic tools, which is closely connected to their support identification properties, makes them very appealing in practical applications. The reported numerical results clearly show the effectiveness of the proposed approach.

Keywords: Clique relaxations, maximum s-defective clique problem, support identification, Frank-Wolfe method.

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

1 Introduction

In the context of network analysis the clique model, dating back at least to the work of Luce and Perry [24] about social networks, refers to subsets with every two elements in a direct relationship. The problem of finding maximal cliques has numerous applications in domains including telecommunication networks, biochemistry, financial networks, and scheduling ([7], [30]). From an optimization perspective, this problem has been the subject of extensive studies stimulating new research directions in both continuous and discrete optimization (see, e.g., [5], [7], [8], [28]). The Motzkin-Straus quadratic formulation [25] in particular has motivated several algorithmic approaches (see [4], [18] and references therein) to the maximum clique problem, beside being of independent interest for its connection with Turán's theorem [1].

Since the strict requirement that every two elements have a direct connection is often not satisfied in practice, many relaxations of the clique model have been proposed (see, e.g., [26] for a survey). In this work we are interested in s-defective cliques ([11], [29], [31]), allowing up to s links to be missing, and introduced in [31] for the analysis of protein interaction networks obtained with large scale techniques subject to experimental errors.

In this paper, we first define a regularized version of a cubic continuous formulation for the maximum s-defective clique problem proposed in [28]. We then apply variants of the classic Frank-Wolfe (FW) method [14] to this formulation.

FW variants are a class of first order optimization methods widely used in the optimization and machine learning communities (see, e.g., [12], [19], [23], [21] and references therein) thanks to their sparse approximation properties, their weaker requirement of a linear minimization oracle instead of a projection oracle with respect to proximal gradient like methods, and their ability to quickly identify the support of a solution. These identification properties, first proved qualitatively for the Frank Wolfe method with in face directions (FDFW, [2], [15], [17]) in the strongly convex setting [17], were recently revisited for the away-step Frank-Wolfe (AFW) with quantitative bounds ([10], [16]) and extended to non convex objectives ([9], [10]).

*ISOR, VCOR & ds:UniVie, Universität Wien, Austria (immanuel.bomze@univie.ac.at)

†Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, Italy (rinaldi@math.unipd.it)

‡Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, Italy (damiano.zeffiro@math.unipd.it)

As we will see in the paper, the support identification properties of FW variants are especially suited for our maximal s-defective clique formulation, since in this case the optimization process can stop as soon as the support of a solution is identified.
Our contributions can be summarized as follows:

- We solve the spurious solution problem for the maximum s-defective clique formulation proposed in [28] by introducing a regularized version, for which we prove equivalence between local maximizers and maximal s-defective cliques. In particular, no postprocessing algorithms are needed to derive the desired structure from a local solution. Our work develops along the lines of analogous results proved for regularized versions of the Motzkin - Straus quadratic formulation ([7], [18]).
- We prove that the FDFW applied to our formulation identifies the support of a maximal s-defective clique in a finite number of iterations.
- We propose a tailored Frank-Wolfe variant for the s-defective clique formulation at hand exploiting its product domain structure. This method retains the identification properties of the FDFW while significantly outperforming it in numerical tests.

The paper is organized as follows: after giving some basic notation and preliminaries in Section 2, we study the regularized maximum s-defective clique formulation in Section 3. We then analyze, in Section 4, the FDFW algorithm and prove that it identifies a maximal s-defective clique in finite time. In Section 5, we describe our tailored FW variant and prove that it shares similar identification properties as FDFW. Finally, in Section 6, we report some numerical results showing the practical effectiveness of the proposed approach. In order to improve readability, some technical details and numerical result tables are deferred to a small appendix.

2 Notation and preliminaries

For any integers a, b we denote by $[a\!:\!b]$ the set of all integers k satisfying $a \leq k \leq b$.

For a vector $r \in \mathbb{R}^d$, the d-dimensional Euclidean space, and a set $A \subset [1:d]$, we denote with r_A the components of r with indexes in A; e is always a vector with all components equal to 1, and dimension clear from the context. Similarly, we denote by e_i the i-th column of an appropriately sized identity matrix.

Let $\mathcal{G} = (V, E)$ be a graph with vertices V and and edges E, $n = |V|$, $A_{\mathcal{G}}$ the adjacency matrix of $\mathcal{G}$, and let $\bar{\mathcal{G}} = (V, \bar{E})$ the complementary graph. The notation we use largely overlaps with the one introduced in [28]. For $s \in \mathbb{N}$ with $s \leq |\bar{E}|$ we define

$$D_s(\mathcal{G}) = \{y \in \{0,1\}^{\bar{E}} \mid e^\intercal y \leq s\} \,,$$

representing the set of "fake edges" to be added to the graph in order to complete an s-defective clique, and its continuous relaxation as

$$D_s'(\mathcal{G}) = \{y \in [0,1]^{\bar{E}} \mid e^\intercal y \leq s\} \,.$$

For $y \in D_s'(\mathcal{G})$ we define the induced adjacency matrix $A(y) \in \mathbb{R}^{n \times n}$ as

$$A(y)_{ij} = \begin{cases} y_{ij} & \text{if } \{i,j\} \in \bar{E} \,, \\ 0 & \text{if } \{i,j\} \notin E \,. \end{cases}$$

For $y \in D_s(\mathcal{G})$ in particular we define $\mathcal{G}(y)$ as the graph with adjacency matrix $A_{\mathcal{G}} + A(y)$, that is the graph where we add to $\mathcal{G}$ the edge $\{i,j\}$ whenever $y_{ij} = 1$. We also define $E(i)$ and $E^y(i)$ as the neighbors of i in $\mathcal{G}$ and $\mathcal{G}(y)$ respectively.
Let $\mathcal{P}_s = \Delta_{n-1} \times D_s'(\mathcal{G})$, with Δ_{n-1} the $(n-1)$-dimensional simplex. The objective of the s-defective clique relaxation defined in [28] is

$$f_{\mathcal{G}}(z) = f_{\mathcal{G}}(x, y) := x^\intercal [A_{\mathcal{G}} + A(y)] x \,, \quad z = (x, y) \in \mathcal{P}_s \tag{2.1}$$

so that when $A(y) = 0$ one retrieves Motzkin-Straus quadratic formulation.

3 A regularized maximum s-defective clique formulation

Here we consider the problem

$$\max\{h_{\mathcal{G}}(z) \mid z \in \mathcal{P}_s\}\,, \tag{P}$$

where $h_{\mathcal{G}} : \mathcal{P}_s \to \mathbb{R}_{>0}$ is a regularized version of $f_{\mathcal{G}}$:

$$h_{\mathcal{G}}(z) = h_{\mathcal{G}}(x, y) := f_{\mathcal{G}}(x, y) + \frac{\alpha}{2}\|x\|^2 + \frac{\beta}{2}\|y\|^2$$

for some $\alpha \in (0, 2)$ and $\beta > 0$. In particular, when $y = 0$ the objective $h_{\mathcal{G}}$ corresponds to the quadratic regularized maximal clique formulation introduced in [4].
For non-empty $C \subseteq V$ let $x^{(C)} = \frac{1}{|C|}\sum_{i\in C} e_i$ be the characteristic vector in Δ_{n-1} of the clique C, and

$$\Delta^{(C)} = \{x \in \Delta_{n-1} \mid x_i = 0 \text{ for all } i \in V \setminus C\}$$

be the minimal face of Δ_{n-1} containing $x^{(C)}$ in its relative interior.
For $p \in \mathcal{P}_s$ we define as $T_{\mathcal{P}_s}(p) = \{v - p : v \in \mathcal{P}_s\}$ as the cone of feasible directions at p in $\mathcal{P}_s$, while for $r \in \mathbb{R}^{|V|+|\bar{E}|}$ we define $T^0_{\mathcal{P}_s}(p, r)$ as the intersection between $T_{\mathcal{P}_s}(p)$ and the plane orthogonal to r:

$$T^0_{\mathcal{P}_s}(p, r) = \{d \in T_{\mathcal{P}_s}(p) \mid d^\intercal r = 0\}\,.$$

We now prove that there is a one to one correspondence between (strict) local maxima of $h_{\mathcal{G}}$ and s-defective cliques coupled together with s fake edges including the one missing on the clique.
Recall that in our polytope-constrained setting, (second order) sufficient conditions for the local maximality of $p \in \mathcal{P}_s$ are (see, e.g., [3])

$$\nabla h_{\mathcal{G}}(p)^\intercal d \le 0 \text{ for all } d \in T_{\mathcal{P}_s}(p) \tag{3.1}$$

and

$$d^\intercal D^2 h_{\mathcal{G}}(p) d < 0 \text{ for all } d \in T^0_{\mathcal{P}_s}(p, \nabla h_{\mathcal{G}}(p))\,. \tag{3.2}$$

In the rest of the article we use $\mathcal{M}_s(\mathcal{G})$ to denote the set of strict local maximizers of $h_{\mathcal{G}}$.

Proposition 3.1 (characterization of local maxima for $h_{\mathcal{G}}$)**.** *The following are equivalent:*

(i) $p \in \mathcal{P}_s$ *is a local maximizer for* $h_{\mathcal{G}}(x, y)$*;*

(ii) $p \in \mathcal{M}_s(\mathcal{G})$*;*

(iii) $p = (x^{(C)}, y^{(p)})$ *where* $s = e^\intercal y^{(p)} \in \mathbb{N}$*, with* C *an* s*-defective clique in* $\mathcal{G}$ *which is also a maximal clique in* $\mathcal{G}(y^{(p)})$*, and* $y^{(p)} \in D_s(\mathcal{G})$ *such that* $y^{(p)}_{ij} = 1$ *for every* $\{i, j\} \in \binom{C}{2} \cap \bar{E}$*.*

In either of these equivalent cases, we have

$$h_{\mathcal{G}}(p) = 1 - \frac{2-\alpha}{2|C|} + s\,\frac{\beta}{2}\,. \tag{3.3}$$

Proof. Let $p = (x^{(p)}, y^{(p)}) \in \mathcal{P}_s$, $g = \nabla h_{\mathcal{G}}(p), H = D^2 h_{\mathcal{G}}(p)$.
(ii) $\Rightarrow$ (i) is trivial.
(i) $\Rightarrow$ (iii). If $s := e^\intercal y^{(p)}$ were fractional, then for some $\{i, j\} \in \bar{E}$ we would have $y^{(p)}_{ij} < 1$. Furthermore

$$\frac{\partial h_{\mathcal{G}}(p)}{\partial y_{ij}} = 2x^{(p)}_i x^{(p)}_j + \beta y^{(p)}_{ij} \ge 0, \quad \frac{\partial h_{\mathcal{G}}(p)}{\partial^2 y_{ij}} = \beta > 0\,. \tag{3.4}$$

Thus for $\varepsilon > 0$ small enough we have $h_{\mathcal{G}}(p + \varepsilon e_{ij}) > h_{\mathcal{G}}(p)$ with $p + \varepsilon e_{ij} \in \mathcal{P}_s$, which means that p is not a local maximizer. Hence $s \in \mathbb{N}$ and obviously $s \le |\bar{E}|$ as well as $y^{(p)} \in D'_s(\mathcal{G})$.
Assume now by contradiction that p is a local maximizer but $y^{(p)} \notin D_s(\mathcal{G})$. Then for two distinct edges $\{i, j\}$,

$\{l, m\} \in \bar{E}$ we must have $y_{ij}^{(p)}, y_{lm}^{(p)} \in (0,1)$. Let $d = (0, e_{ij} - e_{lm})$. Since $\pm d$ are both feasible directions and p is a local maximizer, necessarily $g^\intercal d = 0$. But we also have

$$d^\intercal H d = \frac{\partial h_{\mathcal{G}}(p)}{\partial^2 y_{ij}} + \frac{\partial h_{\mathcal{G}}(p)}{\partial^2 y_{lm}} - 2\,\frac{\partial h_{\mathcal{G}}(p)}{\partial y_{ij} \partial y_{lm}} = 2\beta > 0\,. \tag{3.5}$$

so that again for $\varepsilon > 0$ small enough $h_{\mathcal{G}}(p + \varepsilon d) > h_{\mathcal{G}}(p)$ with $p + \varepsilon d \in \mathcal{P}_s$, a contradiction.
We proved that if p is a local maximizer, then $s = e^\intercal y^{(p)} \in \mathbb{N}$ and $y^{(p)} \in D_s(\mathcal{G})$. But $x^{(p)}$ must be a local maximizer for the function $x \mapsto h_{\mathcal{G}}(x, y^{(p)})$, which is (up to a constant) a regularized maximal clique relaxation for the augmented graph $\mathcal{G}(y^{(p)})$. Thus by well known results (see, e.g., [18], [4]) we must have $x = x^{(C)}$ with C a maximal clique in $\mathcal{G}(y^{(p)})$. In particular, since $\mathcal{G}(y^{(p)})$ is defined by adding s edges to $\mathcal{G}$, C must be an s-defective clique in $\mathcal{G}$.
(iii) $\Rightarrow$ (ii). For a fixed $p = (x^{(C)}, y^{(p)})$ with $C, y^{(p)}$ satisfying the conditions of point (iii) let $\bar{C} = V \setminus C$, $S = \text{supp}(y^{(p)})$ and $\bar{S} = \bar{E} \setminus S$. We abbreviate $E^{(p)}(i) = E^y(i)$ with $y = y^{(p)}$. For every $i \in V$ we have

$$g_i = \alpha x_i^{(C)} + \sum_{j \in E^{(p)}(i)} 2x_j^{(C)} \tag{3.6}$$

In particular for $i \in C$

$$g_i = \frac{\alpha}{|C|} + \sum_{j \in C \setminus \{i\}} 2x_j^{(C)} = \frac{1}{|C|}(\alpha + 2|C| - 2) \tag{3.7}$$

and for every $i \in \bar{C}$

$$g_i = \sum_{j \in E^{(p)}(i) \cap C} 2x_j^{(C)} \leq \frac{2|C| - 2}{|C|} \tag{3.8}$$

where we used $x_j^{(C)} = 1/|C|$ for every $j \in C$, $x_j^{(C)} = 0$ otherwise.
For $\{i, j\} \in \bar{E}$ we have

$$g_{ij} = \beta y_{ij}^{(p)} + 2x_i^{(C)} x_j^{(C)} \tag{3.9}$$

and in particular $g_{ij} = 0$ for $\{i, j\} \in \bar{S}$, while for $\{i, j\} \in S$

$$g_{ij} = \beta + 2x_i^{(C)} x_j^{(C)} \geq \beta > 0\,, \tag{3.10}$$

where we used $y_{ij}^{(p)} = 1$ for $\{i, j\} \in S$, 0 otherwise, and $x_i^{(C)} x_j^{(C)} = 0$ for $\{i, j\} \in \bar{S} \subseteq \bar{E}$.
Let d be a feasible direction from p, so that $d = v - p$ with $v \in \mathcal{P}_s$. Let $\sigma_S = \sum_{\{i,j\} \in S} g_{ij}$, $\sigma_C = \sum_{i \in C} v_i$ $= 1 - \sum_{i \in \bar{C}} v_i \in [0, 1]$, $m_{\bar{C}} = \max_{i \in \bar{C}} g_i$, so that by (3.8) we have $m_{\bar{C}} \leq \frac{2|C|-2}{|C|}$. Then

$$g^\intercal p = \sum_{i \in \bar{C}} x_i^{(C)} g_i + \sum_{i \in C} x_i^{(C)} g_i + \sum_{(i,j) \in S} y_{ij}^{(p)} g_{ij} = \frac{1}{|C|} \sum_{i \in C} g_i + \sum_{\{i,j\} \in S} g_{ij} = \frac{1}{|C|}(\alpha + 2|C| - 2) + \sigma_S \tag{3.11}$$

where we used (3.7) in the last equality. We also have

$$g_V^\intercal v_V = g_C^\intercal v_C + g_{\bar{C}}^\intercal v_{\bar{C}} \leq \frac{\alpha + 2|C| - 2}{|C|} \sigma_C + (1 - \sigma_C) m_{\bar{C}} \leq \frac{\alpha + 2|C| - 2}{|C|} \tag{3.12}$$

where we used (3.7) together with the Hölder inequality in the first inequality, $m_{\bar{C}} \leq \frac{2|C|-2}{|C|}$ in the second inequality and $\sigma_C \leq 1$. Finally,

$$g_{\bar{E}}^\intercal v_{\bar{E}} = g_S^\intercal v_S + g_{\bar{S}}^\intercal v_{\bar{S}} = g_S^\intercal v_S \leq \sigma_S \tag{3.13}$$

where we used $g_{\bar{S}} = 0$ in the second equality, and $v_i \leq 1$ for every $i \in \bar{E}$ in the inequality. We can conclude

$$g^\intercal d = g_V^\intercal v_V + g_{\bar{E}}^\intercal v_{\bar{E}} - g^\intercal p \leq 0 \tag{3.14}$$

where we used (3.13), (3.11) and (3.12) in the inequality. We have equality iff there is equality both in (3.12) and (3.13), and thus iff $v = (x^{(v)}, y^{(v)})$ with $\mathrm{supp}(x^{(v)}) = C$ and $y^{(v)} = y^{(p)}$. In particular p is a first order stationary point with

$$T^0_{\mathcal{P}_s}(p, g) = \{d \in T_{\mathcal{P}_s}(p) \mid d = v - p, v_{\bar{C}} = 0, v_{\bar{E}} = p_{\bar{E}}\} = \{d \in T_{\mathcal{P}_s}(p) \mid d_{\bar{C}} = d_{\bar{E}} = 0\}\,. \tag{3.15}$$

Let H_C be the submatrix of H with indices in C. We have, for $(i, j) \in C^2$ with $i \neq j$, $H_{ij} = 1$ since C is a clique in the augmented graph $\mathcal{G}(y_p)$, while $H_{ii} = \alpha$ for every $i \in V$ and in particular for every $i \in C$. This proves

$$H_C = 2ee^\intercal + (\alpha - 2)\mathbb{I}\,. \tag{3.16}$$

Now if $T^0_{\mathcal{P}_s}(p, g) \ni d \neq 0$ we have

$$d^\intercal H d = d_C^\intercal H_C d_C = d_C^\intercal(2ee^\intercal + (\alpha - 2)\mathbb{I})d_C = (\alpha - 2)\|d_C\|^2 < 0 \tag{3.17}$$

where we used $d_{\bar{C}} = d_{\bar{E}} = 0$ in the first equality, $e^\intercal d_C = e^\intercal(v_V - p_V) = 1 - 1 = 0$ in the third one. This proves the claim, since we have sufficient conditions for local optimality thanks to (3.14) and (3.17). ☐

As a corollary, the global optimum of $h_{\mathcal{G}}$ is achieved on maximum s-defective cliques.

Corollary 3.1. *The global maximizers of $h_{\mathcal{G}}(z)$ are all the points p of the form $p = (x^{C^*}, y^{(p)})$ where C^* is an s-defective clique of maximum cardinality, and $y^{(p)} \in D_s(\mathcal{G})$ such that $e^\intercal y^{(p)} = s$.*

Proof. Let $p = (x^{(C)}, y^{(p)})$ a local maximizer for $h_{\mathcal{G}}(z)$. Then its objective value is, by (3.3), $h_{\mathcal{G}}(p) = 1 - \frac{2-\alpha}{2|C|} + s\,\frac{\beta}{2}$, which is (globally) maximized when $|C|$ is as large as possible, because $2 - \alpha > 0$ by assumption. ☐

Thanks to Proposition 3.1, for every $p \in \mathcal{M}_s(\mathcal{G})$ we can define $y^{(p)} \in D_s(\mathcal{G})$ and a maximal clique C of $\mathcal{G}(y^{(p)})$ such that $p = (x^{(C)}, y^{(p)})$. We now recall that the face of a polytope $\mathcal{Q}$ exposed by a gradient $g \in \mathbb{R}^n$ is defined as

$$\mathcal{F}_e(g) = \underset{w \in Q}{\arg\max}\, g^\intercal w. \tag{3.18}$$

With this notation, we prove that the face of $\mathcal{P}_s$ exposed by the gradient in $p \in \mathcal{M}_s(\mathcal{G})$ is simply the product between $\Delta^{(C)}$ and the singleton $\{y^{(p)}\}$. This property, sometimes referred to as strict complementarity, is of key importance to prove identification results for Frank-Wolfe variants (see [9], [10], [16], and the discussion of external regularity in [6, Section 5.3]).

Lemma 3.1. *Let $p = (x^{(C)}, y^{(p)}) \in \mathcal{M}_s(\mathcal{G})$. Then the face exposed by $\nabla h_{\mathcal{G}}(p)$ coincides with the minimal face $\mathcal{F}(p)$ of $\mathcal{P}_s$ containing p:*

$$\mathcal{F}_e(\nabla h_{\mathcal{G}}(p)) = \mathcal{F}(p) = \Delta^{(C)}_{n-1} \times \{y^{(p)}\}\,. \tag{3.19}$$

Proof. To start with, the second equality follows from the fact that $y^{(p)}$ is a vertex of $D'_s(\mathcal{G})$ and that $\Delta^{(C)}_{n-1}$ is the minimal face of Δ_{n-1} containing $x^{(C)}$. The first equality is then equivalent to proving that for every vertex $a = (a_x, a_y)$ of $\mathcal{P}_s$ with $a \in \mathcal{P}_s \setminus \mathcal{F}(p)$ we have $\lambda_a(p) < 0$. Given that stationarity conditions must hold in Δ_{n-1} and $D'_s(\mathcal{G})$ separately, $\lambda_a(p) < 0$ if and only if

$$\lambda^x_a(p) := \nabla_x h_{\mathcal{G}}(p)^\intercal(a_x - x^{(C)}) \leq 0\,, \tag{3.20a}$$

$$\lambda^y_a(p) := \nabla_y h_{\mathcal{G}}(p)^\intercal(a_y - y^{(p)}) \leq 0\,, \tag{3.20b}$$

and at least one of these relations must be strict. Since a is a vertex of $\mathcal{P}_s$, $a_x = e_l$ with $l \in [1:n]$ and $a_y \in D_s(\mathcal{G})$, while $a \notin \mathcal{F}(p)$ implies $l \notin C$ or $a_y \neq y^{(p)}$. If $l \in C$ then $\lambda^x_a(p) = 0$ by stationarity conditions, otherwise

$$\nabla_x h_{\mathcal{G}}(p)^\intercal x^{(C)} = 2(x^{(C)})^\intercal[A + A(y^{(p)})]x^{(C)} + \alpha\|x^{(C)}\|^2 = 2 - \frac{2-\alpha}{|C|} \tag{3.21}$$

and

$$\nabla_x h_{\mathcal{G}}(p)^\intercal a_x = \frac{\partial}{\partial x_l} h_{\mathcal{G}}(p) = \alpha x_l + \sum_{j \in C \cap E^{(p)}(l)} 2x_j = 2\frac{|C \cap E^{(p)}(l)|}{|C|} \leq 2 - \frac{2}{|C|}\,, \tag{3.22}$$

where we used $a_x = e_l$ in the first equality, $l \notin C$ together with $x_j = 1/|C|$ for every $j \in C$ in the third equality, and the maximality of the clique C in the augmented graph $\mathcal{G}(y^{(p)})$ in the inequality. Combining (3.21) and (3.22), we obtain

$$\nabla_x h_{\mathcal{G}}(p)^\intercal (a_x - x^{(C)}) \leq -\frac{\alpha}{|C|} < 0\,, \tag{3.23}$$

which proves that (3.20a) holds with strict inequality if $l \notin C$, or else with equality if $l \in C$.

In a similar vain we proceed with (3.20b). If $a_y = y^{(p)}$ then (3.20b) holds with equality but then $l \in V \setminus C$ and we are done. So suppose $a_y \neq y^{(p)}$, and consider the supports $S_y = \{\{i,j\} \in \bar{E} \mid (a_y)_{ij} = 1\}$ and $S_p = \{\{i,j\} \in \bar{E} \mid y^{(p)}_{ij} = 1\}$. Since $a_y \in D_s(\mathcal{G})$, we have $|S_y| \leq s$ and on the other hand, by Proposition 3.1(iii), $|S_p| = s$. As S_y and S_p must be distinct, we conclude $S_y \setminus S_p \neq \emptyset$. Furthermore, by (3.4) for every $\{i,j\}$ in A_p we have

$$\frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) \geq \beta y^{(p)}_{ij} = \beta > 0\,, \tag{3.24}$$

while for every $\{i,j\}$ in $A_y \setminus A_p$ we have

$$\frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) = 0 \tag{3.25}$$

because $y^{(p)}_{ij} = 0$ by definition of A_p and $x^{(C)}_i x^{(C)}_j = 0$ since, again invoking Proposition 3.1(iii), $\{i,j\} \in \bar{E} \setminus \binom{C}{2}$. So we can finally prove (3.20b) by observing

$$\begin{aligned}\nabla_y h_{\mathcal{G}}(p)^\intercal (a_y - y^{(p)}) &= \sum_{\{i,j\}\in A_y} \frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) - \sum_{\{i,j\}\in A_p} \frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) \\ = \sum_{\{i,j\}\in A_y\setminus A_p} \frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) &- \sum_{\{i,j\}\in A_p\setminus A_y} \frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) = - \sum_{\{i,j\}\in A_p\setminus A_y} \frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(p) < 0\end{aligned} \tag{3.26}$$

where we used (3.25) in the third equality and (3.24) together with $A_p \setminus A_y \neq \emptyset$ in the inequality. □

4 Frank-Wolfe method with in face directions

Let $\mathcal{Q} = \text{conv}(A) \subset \mathbb{R}^n$ with $|A| < +\infty$. In this section, we consider the FDFW for the solution of the smooth constrained optimization problem

$$\max\{f(w) \mid w \in \mathcal{Q}\}\,.$$

In particular, $\{w_k\}$ is always a sequence generated by the FDFW applied to the polytope $\mathcal{Q}$ with objective f. For $w \in \mathcal{Q}$ we denote with $\mathcal{F}(w)$ the minimal face of $\mathcal{Q}$ containing w. The FDFW at every iteration chooses between the classic FW direction $d_k^{\mathcal{FW}}$ calculated at Step 2 and the in face direction $d_k^{\mathcal{FD}}$ calculated at Step 10 with the criterion in Step 12. The classic FW direction points toward the vertex maximizing the scalar product with the current gradient, or equivalently the vertex maximizing the first order approximation $w \mapsto f(w_k) + \nabla f(w_k)^\intercal w$ of the objective f. The in face direction $d_k^{\mathcal{FD}}$ is always a feasible direction in $\mathcal{F}(w_k)$ from w_k, and it points away from the vertex of the face minimizing the first order approximation of the objective. When the algorithm performs an in face step, we have that the minimal face containing the current iterate either stays the same or its dimension drops by one. The latter case occurs when the method performs a maximal feasible in face step (i.e., a step with $\alpha_k = \alpha_k^{\max}$ and $d_k = d_k^{\mathcal{FD}}$), generating a point on the boundary of the current minimal face. As we will see formally in Proposition 4.1, this drop in dimension is what allows the method to quickly identify low dimensional faces containing solutions.

We often require the following lower bound on the stepsizes:

$$\alpha_k \geq \bar{\alpha}_k := \min(\alpha_k^{\max}, c\frac{\nabla f(w_k)^\intercal d_k}{\|d_k\|^2}) \tag{S1}$$

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

$$f(w_k + \alpha_k d_k) - f(w_k) \geq \rho\bar{\alpha}_k\, \nabla f(w_k)^\intercal d_k\,. \tag{S2}$$

Algorithm 1 Frank-Wolfe method with in face directions (FDFW) on a polytope

1: **Initialize** $w_0 \in \mathcal{Q}$, $k := 0$
2: Let $s_k \in \arg\max_{y \in \mathcal{Q}} \nabla f(w_k)^\intercal y$ and $d_k^{\mathcal{FW}} := s_k - w_k$.
3: **if** w_k is stationary **then**
4:     STOP
5: **end if**
6: Let $v_k \in \arg\min_{y \in \mathcal{F}(w_k)} \nabla f(w_k)^\intercal y$ and $d_k^{\mathcal{FD}} := w_k - v_k$.
7: **if** $\nabla f(w_k)^\intercal d_k^{\mathcal{FW}} \geq \nabla f(w_k)^\intercal d_k^{\mathcal{FD}}$ **then**
8:     $d_k := d_k^{\mathcal{FW}}$
9: **else**
10:     $d_k := d_k^{\mathcal{FD}}$
11: **end if**
12: Choose the stepsize $\alpha_k \in (0, \alpha_k^{\max}]$ with a suitable criterion
13: Update: $w_{k+1} := w_k + \alpha_k d_k$
14: Set $k := k + 1$. Go to step 2.

As explained in the Appendix (see Lemma 7.1), these conditions generalize properties of exact and Armijo line search.

We also define the multiplier functions λ_a for $a \in A$, $w \in \mathbb{R}^n$ as

$$\lambda_a(w) = \nabla f(w)^\intercal (a - w)\,. \tag{4.1}$$

We adapt the well known FW gap ([10], [22]) to the maximization case, thus obtaining the following measure of stationarity

$$G(w) := \max_{y \in \mathcal{Q}} \nabla f(w)^\intercal (w - y) = \max_{a \in A} \nabla f(w)^\intercal (w - a) = \max_{a \in A} -\lambda_a(w)\,, \tag{4.2}$$

as well as an "in face" gap

$$G_{\mathcal{F}}(w) = \max(G(w), \max_{b \in \mathcal{F}(w) \cap A} \lambda_b(w))\,. \tag{4.3}$$

Using these definitions, we have

$$\begin{aligned} \nabla f(w_k)^\intercal d_k &= \max(\nabla f(w_k)^\intercal d_k^{\mathcal{FW}}, \nabla f(w_k)^\intercal d_k^{\mathcal{FD}}) \\ &= \max(G(w_k), \max_{y \in \mathcal{F}(w_k)} \nabla f(w_k)^\intercal (w_k - y)) = G_{\mathcal{F}}(w_k)\,, \end{aligned} \tag{4.4}$$

where in the second equality we used

$$\nabla f(w_k)^\intercal d_k^{\mathcal{FW}} = \max_{y \in Q} \nabla f(w_k)^\intercal (y - w_k) \tag{4.5}$$

and in the third equality

$$\nabla f(w_k)^\intercal d_k^{\mathcal{FD}} = \max_{b \in \mathcal{F}(w_k)} \nabla f(w_k)^\intercal (w_k - b) = \max_{b \in \mathcal{F}(w_k) \cap A} -\lambda_b(w_k)\,. \tag{4.6}$$

From the definitions it also immediately follows

$$G_{\mathcal{F}}(w) \geq G(w) \geq 0 \tag{4.7}$$

with equality iff w is a stationary point.
In order to obtain a local identification result, we need to prove that under certain conditions the method does consecutive maximal in face steps, thus identifying a low dimensional face containing a minimizer. First, in the following lemma we give an upper bound for the maximal feasible stepsize.

Lemma 4.1. *If w_k is not stationary, then $\alpha_k \leq G(w_k)/G_{\mathcal{F}}(w_k)$.*

Proof. Notice that since w_k is not stationary we have $G(w_k) > 0$ and therefore also $G_{\mathcal{F}}(w_k) > 0$. Now

$$\nabla f(w_k)^\intercal (w_k + \alpha_k d_k) \le \max_{y \in \mathcal{Q}} \nabla f(w_k)^\intercal y = \nabla f(w_k)^\intercal (w_k + d_k^{\mathcal{FW}}) = \nabla f(w_k)^\intercal w_k + G(w_k)\,,$$

where in the inequality we used $w_k + \alpha_k d_k \in \mathcal{Q}$. Subtracting $\nabla f(w_k)^\intercal w_k$ on both sides we obtain

$$\alpha_k \nabla f(w_k)^\intercal d_k \le G(w_k)\,. \tag{4.8}$$

and the thesis follows by applying (4.4) to the LHS. □

We can now prove a local identification result.

Proposition 4.1 (FDFW local identification)**.** *Let p be a stationary point for f defined on $\mathcal{Q}$ and assume that* (S1) *holds. We have the following properties:*

(a) there exists $r^*(p) > 0$ such that if $w_k \in B_{r^*(p)}(p) \cap \mathcal{F}_e(\nabla f(p))$ then $w_{k+1} \in \mathcal{F}_e(\nabla f(p))$;

(b) for any $\delta > 0$ there exists $r(\delta, p) \le \delta$ such that if $w_k \in B_{r(\delta,p)}(p)$ then $w_{k+j} \in \mathcal{F}_e(\nabla f(p)) \cap B_\delta(p)$ for some $j \le \dim(\mathcal{F}(w_k))$.

Proof. (a) Notice that by definition of exposed face and stationarity conditions

$$\lambda_a(p) \le 0 \tag{4.9}$$

for every $a \in A$, with equality iff $a \in \mathcal{F}_e(\nabla f(p))$. Then by continuity we can take $r^*(p)$ small enough so that $\lambda_a(w) < 0$ for every $a \in A \setminus (A \cap \mathcal{F}_e(\nabla f(p)))$. Under this condition, if $w_k \in B_{r^*(p)}(p)$ then the method cannot select a FW direction pointing toward an atom outside the exposed face $\mathcal{F}_e(\nabla f(p))$, because all the atoms maximizing the RHS of (4.2) must necessarily be in $\mathcal{F}_e(\nabla f(p))$. In particular if $w_k \in B_{r^*(p)}(p) \cap \mathcal{F}_e(\nabla f(p))$ then the method selects either an in face direction or a FW direction pointing toward a vertex in $\mathcal{F}_e(\nabla f(p))$. In both cases, $w_{k+1} \in \mathcal{F}_e(\nabla f(p))$.
(b) Let D be the diameter of $\mathcal{Q}$. We now consider $r^{(0)}(\delta, p) \le \min(\delta, r^*(p))$ such that, for every $w \in B_{r^{(0)}(\delta,p)}(p)$

$$\max_{a \in A} \lambda_a(w) < \min_{b \in A \setminus \mathcal{F}_e(\nabla f(p))} \min(-\lambda_b(w), \frac{c}{D^2} \lambda_b(w)^2)\,. \tag{4.10}$$

As we will see in the rest of the proof this upper bound together with Lemma 4.1 ensures in particular that the FDFW performs maximal in face steps in $B_{r^{(0)}(\delta,p)}(p) \setminus \mathcal{F}_e(\nabla f(p))$. Furthermore, (4.10) can always be satisfied thanks to (4.9) and by the continuity of multipliers. We then define recursively for $1 \le l \le n$ a sequence $r^{(l)}(\delta, p) \le r^{(l-1)}(\delta, p)$ of radii small enough so that, for

$$M_l = \sup_{w \in B_{(l)}(p) \setminus \mathcal{F}_e(\nabla f(p))} G(w)/G_{\mathcal{F}}(w)\,, \tag{4.11}$$

with $B_{(l)}(p) := B_{r^{(l)}(\delta,p)}(p)$ we have

$$r^{(l)}(\delta, p) + D M_l < r^{(l-1)}(\delta, p)\,. \tag{4.12}$$

Again this sequence can always be defined thanks to the continuity of multipliers. Finally, we define $r(\delta, p) = r^{(n)}(\delta, p)$.
Given these definitions, when $w_k \in B_{(l)}(p) \subset B_{(0)}(p)$ and w_k is not in $\mathcal{F}_e(\nabla f(p))$ an in face direction is selected, because

$$\nabla f(w_k)^\intercal d_k^{\mathcal{FW}} = \max_{a \in A} \lambda_a(w) < \min_{b \in A \setminus \mathcal{F}_e(\nabla f(p))} -\lambda_b(w) \le \max_{b \in \mathcal{F}(w_k) \cap A} -\lambda_b(w_k) = \nabla f(x_k)^\intercal d_k^{\mathcal{FD}}\,, \tag{4.13}$$

where we used (4.10) in the first inequality, $w_k \notin \mathcal{F}_e(p)$ in the second, and (4.6) in the second equality. We now want to prove that in this case α_k is maximal reasoning by contradiction. On the one hand, we have

$$\alpha_k \ge c \frac{\nabla f(x_k)^\intercal d_k}{\|d_k\|^2} \ge \frac{c}{D^2} \nabla f(x_k)^\intercal d_k = \frac{c}{D^2} G_{\mathcal{F}}(w_k) \tag{4.14}$$

where we used the assumption (S1) in the first inequality, $\|d_k\| \leq D$ in the second and $G_{\mathcal{F}}(w_k) = \nabla f(x_k)^\intercal d_k^{\mathcal{FD}}$ together with $d_k = d_k^{\mathcal{FD}}$ in the last one.
On the other hand,

$$\begin{aligned} G(w_k) = \max_{a \in A} \lambda_a(w_k) <& \frac{c}{D^2} \min_{b \in A \setminus \mathcal{F}_e(\nabla f(p))} \lambda_b(w)^2 \leq \frac{c}{D^2} \max_{b \in \mathcal{F}(w_k)} \lambda_b(w)^2 \\ =& \frac{c}{D^2} (\nabla f(w_k)^\intercal d_k)^2 = \frac{c}{D^2} G_{\mathcal{F}}(w_k)^2 \end{aligned} \tag{4.15}$$

where we used (4.10) in the first inequality, $w_k \notin \mathcal{F}_e(\nabla f(p))$ in the second, (4.6) together with $d_k = d_k^{\mathcal{FD}}$ in the second equality, and (4.4) in the third equality.
The inequality (4.15) leads us to a contradiction with the lower bound on α_k given by (4.14), since it implies

$$\alpha_k \leq \frac{G(w_k)}{G_{\mathcal{F}}(w_k)} < \frac{c}{D^2} G_{\mathcal{F}}(w_k)\,, \tag{4.16}$$

where we applied Lemma 4.1 in the first inequality and (4.15) in the second.
Assume now $w_k \in B_{(n)}(p)$. We prove by induction that, for every $j \in [-1 : \dim(\mathcal{F}(w_k)) - 1]$, if $\{w_{k+i}\}_{0 \leq i \leq j} \cap \mathcal{F}_e(\nabla f(p)) = \emptyset$ then $w_{k+j+1} \in B_{(n-j-1)}(p)$. For $j = -1$ we have $w_k \in B_{(n)}(p)$ by assumption. Now if $\{w_{k+i}\}_{0 \leq i \leq j} \cap \mathcal{F}_e(\nabla f(p)) = \emptyset$ we have

$$\begin{aligned} &\|w_{k+j+1} - p\| \leq \|w_{k+j} - p\| + \|w_{k+j+1} - w_{k+j}\| < r^{(n-j)}(\delta, p) + \|w_{k+j+1} - w_{k+j}\| \\ =& r^{(n-j)}(\delta, p) + \alpha_k \|d_k\| \leq r^{(n-j)}(\delta, p) + D\frac{G(w_k)}{G_{\mathcal{F}}(w_k)} \leq r^{(n-j)}(\delta, p) + DM_{n-j} < r^{(n-j-1)}(\delta, p)\,, \end{aligned} \tag{4.17}$$

where we used the inductive hypothesis $w_{k+j} \in B_{(n-j)}(p)$ in the second inequality, Lemma 4.1 in the third inequality, (4.11) in the fourth inequality and the assumption (4.12) in the last one. In particular $w_{k+j+1} \in B_{(n-j-1)}(p)$ and the induction is completed.
Since $B_{(n-j)}(p) \subset B_{(0)}(p)$, if $w_{k+j} \in (B_{(n-j)}(p) \setminus \mathcal{F}_e(\nabla f(p))$ then α_{k+j} must be maximal and therefore $\dim(\mathcal{F}(w_{k+j+1})) < \dim(\mathcal{F}(w_{k+j}))$. But starting from the index k the dimension of the current face can decrease at most $\dim(\mathcal{F}(w_k)) < n$ times in consecutive steps, so there must exists $j \in [0, \dim(\mathcal{F}(w_k))]$ such that $w_{k+j} \in \mathcal{F}_e(\nabla f(p))$. Taking the minimum j satisfying this condition we also obtain $w_{k+j} \in B_{(0)}(p) \subset B_\delta(p)$. □

A straightforward adaptation of results from [10] implies convergence to the set of stationary points for the FDFW.

Proposition 4.2. *If* (S1) *and* (S2) *hold, then all the limit points of the FDFW are contained in the set of stationary points of* f.

Proof. The proof presented in the special case of the simplex in [10], where the FDFW coincides with the away-step Frank-Wolfe, extends to generic polytopes in a straightforward way. □

In the next lemma we improve the FDFW local identification result given in Proposition 4.1 under an additional strong concavity assumption for the face containing the solution, satisfied in particular by $h_{\mathcal{G}}$.

Lemma 4.2. *Let* p *be a stationary point for* f *restricted to* $\mathcal{Q}$. *Assume that* (S1) *holds and that* f *is strongly concave*[1] *in* $\mathcal{F}_e(\nabla f(p))$. *Then, for a neighborhood* $U(p)$ *of* p, *if* $w_0 \in U(p)$:

(a) if $\{f(w_k)\}$ *is increasing, there exists* $k \in [0\colon \mathit{dim}(\mathcal{F}(w_0))]$ *such that* $w_{k+i} \in \mathcal{F}_e(\nabla f(p))$ *for every* $i \geq 0$;

(b) if in addition also (S2) *holds, then* $\{w_{k+i}\}_{i \geq 0}$ *converges to* p.

Proof. (a) Let μ be the strong concavity constant of f restricted to $\mathcal{F}_e(\nabla f(p))$, so that

$$f(w) \leq f(p) - \frac{\mu}{2}\|w - p\|^2 \tag{4.18}$$

[1] in fact, we only need strict concavity of f here.

for every $w \in \mathcal{F}_e(\nabla f(p))$. For $\varepsilon = \frac{\mu r^*(p)^2}{2}$, let $\mathcal{L}_\varepsilon$ be the superlevel of f for $f(p) - \varepsilon$:

$$\mathcal{L}_\varepsilon = \{y \in \mathcal{Q} \mid f(y) > f(p) - \varepsilon\}. \tag{4.19}$$

Let now $\bar{r} = r(\delta, p)$ defined as in Proposition 4.1, with $\delta = r^*(p)$. By (4.18) it follows $\mathcal{L}_\varepsilon \cap \mathcal{F}_e(\nabla f(p)) \subset B_{r^*(p)}(p)$. Assume now $w_0 \in U(p)$ with $U(p) = B_{\bar{r}}(p) \cap \mathcal{L}_\varepsilon$. By applying Proposition 4.1 we obtain that there exists $k \in [0 : \dim(\mathcal{F}(w_0))]$ such that w_k is in $\mathcal{F}_e(\nabla f(p)) \cap B_{r^*(p)}(p)$. But since $f(w_k) \geq f(w_0) > f(p) - \varepsilon$ we have the stronger condition $w_k \in \mathcal{L}_\varepsilon \cap \mathcal{F}_e(\nabla f(p))$.To conclude, notice that the sequence cannot escape from this set, because for $i \geq 0$ $w_{k+i} \in \mathcal{L}_\varepsilon$ implies that also w_{k+i+1} is in $\mathcal{L}_\varepsilon$, and $w_{k+i} \in \mathcal{L}_\varepsilon \cap \mathcal{F}_e(\nabla f(p)) \subset B_{r^*(p)}(p) \cap \mathcal{F}_e(\nabla f(p))$ implies that also w_{k+i+1} is in $\mathcal{F}_e(\nabla f(p))$.
(b) By point (a) $\{w_{k+i}\}_{i\geq 0}$ is contained in $\mathcal{F}_e(\nabla f(p))$. But by assumption f is strongly concave in $\mathcal{F}_e(\nabla f(p))$ with p global maximum and the only stationary point. To conclude it suffices to apply Proposition 4.2. ☐

Corollary 4.1. *Let $\{w_k\}$ be a sequence generated by the FDFW and assume that at least one limit point p is stationary and such that f is strongly concave in $\mathcal{F}_e(\nabla f(p))$. Then under the conditions* (S1) *and* (S2) *on the stepsize we have $w_k \to p$ with $w_k \in \mathcal{F}_e(\nabla f(p))$ for k large enough.*

Proof. Follows from Lemma 4.2 by observing that the sequence must be ultimately contained in $U(p)$. ☐

We can now prove local convergence and identification for the FDFW applied to the $s-$defective maximal clique formulation (P).

Proposition 4.3 (FDFW local convergence)**.** *Let $p = (x^{(C)}, y^{(p)}) \in \mathcal{M}_s(\mathcal{G})$, let $\{z_k\}$ be a sequence generated by the FDFW. Then under* (S1) *there exists a neighborhood $U(p)$ of p such that if $\bar{k} := \min\{k \in \mathbb{N}_0 \mid z_k \in U(p)\}$ we have the following properties:*

(a) if $h_\mathcal{G}(z_k)$ is monotonically increasing, then supp$(z_k) = C$ and $y_k = y^{(p)}$ for every $k \geq \bar{k} + \dim \mathcal{F}(w_{\bar{k}})$;

(b) if (S2) *also holds then $z_k \to p$.*

Proof. Let $A(p) = A_\mathcal{G} + A(y^{(p)})$. Then for $x \in \Delta^{(C)}$

$$\begin{aligned} x^\intercal A(p)x = &\sum_{(i,j)\in V^2} x_i A(p)_{ij} x_j = \sum_{i\in C} x_i (\sum_{j\in C} A(p)_{ij} x_j) = \sum_{i\in C} x_i (\sum_{j\in C\setminus\{i\}} x_j) \\ = &\sum_{i\in C}(x_i \sum_{j\in C} x_j - x_i^2) = (\sum_{i\in C} x_i)^2 - \sum_{i\in C} x_i^2\,, \end{aligned} \tag{4.20}$$

where in the first equality we used supp$(x) = C$, in the second that C is a clique n $G(y^{(p)})$, and $\sum_{i\in C} x_i = \sum_{i\in V} x_i = 1$.
Observe now that the function $x \mapsto h_\mathcal{G}(x, y^{(p)})$ is strongly concave in $\Delta^{(C)}$. Indeed for $x \in \Delta^{(C)}$

$$\begin{aligned} h_\mathcal{G}(x, y^{(p)}) = x^\intercal A(p)x + \frac{\alpha}{2}\|x\|^2 + \frac{\beta}{2}\|y^{(p)}\|^2 = &(\sum_{i\in C} x_i)^2 - \sum_{i\in C} x_i^2 + \frac{\alpha}{2}\|x\|^2 + \frac{\beta}{2}\|y^{(p)}\|^2 \\ = &1 - (1 - \frac{\alpha}{2})\sum_{i\in C} x_i^2 + \frac{\beta}{2}\|y^{(p)}\|^2\,, \end{aligned} \tag{4.21}$$

where in the second equality we used (4.20). The RHS of (4.21) is strongly concave in x since $\alpha \in (0, 2)$ so that $-(1 - \alpha/2) \in (-1, 0)$. This together with Lemma 3.1 gives us the necessary assumptions to apply Lemma 4.2. ☐

As a corollary, we have the following global convergence result under the mild assumption that the set of limit points contains one local minimizer.

Corollary 4.2 (FDFW global convergence)**.** *Let $\{z_k\}$ be a sequence generated by the FDFW and assume that at least one limit point $p = (x^{(C)}, y^{(p)})$ of $\{z_k\}$ is in $\mathcal{M}_s(\mathcal{G})$. Then under the conditions* (S1) *and* (S2) *on the stepsize we have $z_k \to p$ with supp$(x_k) \subset C$ and $y_k = y_p$ for k large enough.*

Proof. Follows from Corollary 4.1 where all the necessary assumptions are satisfied as for Proposition 4.3. ☐

Remark 4.1. *When the sequence converges to a first order stationary point $z^* = (x^*, y^*)$ which is not a local maximizer, one can use the procedure described in [28] to obtain an s-defective clique C and $y \in D_s(\mathcal{G})$ with $h_{\mathcal{G}}((x^{(C)}, y)) > h(z^*)$. The cost of the procedure is $O(|supp(x^*)|^2)$.*

5 A Frank-Wolfe variant for s-defective clique

As can be seen from numerical results, one drawback of the standard FDFW applied to the s-defective clique formulation (P) is the slow convergence of the high dimensional y component. Since this component is "tied" to the x component, it is not possible to speed up the convergence by changing the regularization term without compromising the quality of the solution. Motivated by this challenge, we introduce a tailored Frank-Wolfe variant, namely FWdc, for the maximum s-defective clique formulation (P), which exploits the product domain structure of the problem at hand by employing separate updating rules for the two blocks.
In particular, at every iteration the method alternates a FDFW step on the x variables (Step 5) with a full FW

Algorithm 2 Frank-Wolfe variant for s-defective clique

1: **Initialize** $z_0 := (x_0, y_0) \in \mathcal{P}_s$, $k := 0$
2: **if** z_k is stationary **then**
3: STOP
4: **end if**
5: Compute x_{k+1} applying one iterate of Algorithm 1 with $w_0 = x_k$ and $f(w) = h_{\mathcal{G}}(w, y_k)$.
6: Let $y_{k+1} \in \arg\max_{y \in D'_s(\mathcal{G})} \nabla_y h_{\mathcal{G}}(x_k, y_k)^\intercal y$.
7: Set $k := k + 1$. Go to step 2.

step on the y variable (Step 6), so that y_k is always chosen in the set of vertices $D_s(\mathcal{G})$ of $D'_s(\mathcal{G})$. Furthermore, as we prove in the next proposition, $\{y_k\}$ is ultimately constant. This allows us to obtain convergence results by applying the general properties of the FDFW proved in the previous section to the x component.

Proposition 5.1. *In the FWdc if $h_{\mathcal{G}}(z_k)$ is increasing then $\{y_k\}$ can change at most $\frac{2}{\beta} - \frac{2-\alpha}{\beta|C^*|} + s$ times, with C^* s-defective clique of maximal cardinality.*

Proof. Assume that y_k and y_{k+1} are distinct vertices of $D'_s(\mathcal{G})$. Then

$$\begin{aligned} h_{\mathcal{G}}(z_{k+1}) - h_{\mathcal{G}}(z_k) &\geq \nabla h_{\mathcal{G}}(z_k)^\intercal (z_{k+1} - z_k) + \frac{\beta}{2} \|z_{k+1} - z_k\|^2 \\ =& \nabla_y h_{\mathcal{G}}(z_k)^\intercal (y_{k+1} - y_k) + \frac{\beta}{2} \|y_{k+1} - y_k\|^2 \geq \frac{\beta}{2} > 0 \end{aligned} \tag{5.1}$$

where we used the $\beta-$strong convexity of $y \mapsto h_{\mathcal{G}}(x, y)$ in the first inequality, $x_k = x_{k+1}$ in the equality, $y_{k+1} \in \arg\max_{y \in \mathcal{P}_s} \nabla_y h_{\mathcal{G}}(z_k)^\intercal y$ and the fact that the distance between vertices of $D'_s(\mathcal{G})$ is at least 1 in the second inequality.
Therefore y_k can change at most

$$\max_{z \in \mathcal{P}_s} \frac{2(h_{\mathcal{G}}(z) - h_{\mathcal{G}}(z_0))}{\beta} \leq \max_{z \in \mathcal{P}_s} \frac{2h_{\mathcal{G}}(z)}{\beta} = \frac{1 - 1/|C^*| + \alpha/2|C^*| + s\beta/2}{\beta/2} = \frac{2}{\beta} + \frac{\alpha - 2}{\beta|C^*|} + s$$

times, where we used $h_{\mathcal{G}} \geq 0$ in the first inequality, and Corollary 3.1 in the second inequality. ☐

Corollary 5.1. *Let $\{z_k\}$ be a sequence generated by Algorithm 2.*

1. *If conditions* (S1) *and* (S2) *hold on the stepsizes, then $\{z_k\}$ converges to the set of stationary points.*
2. *If the stepsize is given by exact line search or Armijo line search and the set of limit points of $\{z_k\}$ is finite, then $z_k \to p$ with p stationary.*

Proof. As a corollary of Proposition 5.1, an application of Algorithm 2 reduces, after a finite number of changes for the variable y, to an application of the FDFW on the simplex for the optimization of a quadratic objective. After noticing that on the simplex the FDFW coincides with the AFW, point 1 follows directly from Proposition 4.2, and point 2 follows from [10, Theorem 4.5]. □

For a clique C of $\mathcal{G}(y)$ different from $\mathcal{G}$ we define $m(C, \mathcal{G}(y))$ as

$$\min_{v \in V \setminus C} |C| - |E^y(v) \cap C| \,, \tag{5.2}$$

that is the minimum number of edges needed to increase by 1 the size of the clique.
We now give an explicit bound on how close the sequence $\{x_k\}$ generated by Algorithm 2 must be to $x^{(C)}$ for the identification to happen.

Proposition 5.2. *Let $\{z_k\}$ be a sequence generated by Algorithm 2, $\bar{y} \in D^s(\mathcal{G})$, C be a clique in $\mathcal{G}(\bar{y})$, let $\delta_{\max}$ the maximum eigenvalue of the adjacency matrix $\bar{A} := A_{\mathcal{G}} + A(\bar{y})$. Let $\bar{k}$ be a fixed index in $\mathbb{N}_0$, I^c the components of supp$(x_{\bar{k}})$ with index not in C and let $L := 2\delta_{\max} + \alpha$. Assume that $y_{\bar{k}+j} = \bar{y}$ is constant for $0 \leq j \leq |I^c|$, that* (S1) *holds for $c = 1/L$, and that*

$$\|x_{\bar{k}} - x^{(C)}\|_1 \leq \frac{m_\alpha(C, \mathcal{G}(y_{\bar{k}}))}{m_\alpha(C, \mathcal{G}(y_{\bar{k}})) + 2|C|\delta_{\max} + |C|\alpha} \tag{5.3}$$

for $m_\alpha(C, \mathcal{G}(y_{\bar{k}})) = m(C, \mathcal{G}(y_{\bar{k}})) - 1 + \alpha/2$. Then supp$(x_{\bar{k}+|I^c|}) = C$.

Proof. Since y_k does not change for $k \in [\bar{k} : \bar{k} + |I^c|]$, Algorithm 2 corresponds to an application of the AFW to the simplex Δ_{n-1} on the variable x. For $1 \leq i \leq n$ let $\lambda_i(x) = \frac{\partial}{\partial x_i} h_{\mathcal{G}}(x, y_{\bar{k}})$ be the multiplier functions associated to the vertices of the simplex, and let

$$\lambda_{\min} = \min_{i \in V \setminus C} -\lambda_i(x^{(C)}) \,, \tag{5.4}$$

be the smallest negative multiplier with corresponding index not in C. Let L' be a Lipschitz constant for $\nabla_x h_{\mathcal{G}}(x, y)$ with respect to the variable x. By [10, Theorem 3.3] if

$$\|x_{\bar{k}} - x^{(C)}\|_1 < \frac{\lambda_{\min}}{\lambda_{\min} + 2L'} \tag{5.5}$$

we have the desired identification result.
We now prove that we can take L' equal to L in the following way:

$$\|\nabla_x h_{\mathcal{G}}(x', y_{\bar{k}}) - \nabla_x h_{\mathcal{G}}(x, y_{\bar{k}})\| = \|2\bar{A}(x' - x) + \alpha(x' - x)\| \leq (2\delta_{\max} + \alpha)\|x' - x\| \,, \tag{5.6}$$

where we used $\nabla_x h_{\mathcal{G}}(x, y) = 2\bar{A}x + \alpha x$ in the equality. As for the multipliers, for $i \in V \setminus C$ we have the lower bound

$$\begin{aligned} -\lambda_i(x^{(C)}) = \nabla_x h_{\mathcal{G}}(x^{(C)}, y_{\bar{k}})^\intercal (x^{(C)} - e_i) =& \frac{-2|C \cap E^{y_{\bar{k}}}(i)| + 2|C| - 2 + \alpha}{|C|} \\ \geq& \frac{2m_\alpha(C, \mathcal{G}(y_{\bar{k}}))}{|C|} \end{aligned} \tag{5.7}$$

by combining (3.21) and (3.22) in the second equation. We can now bound $\lambda_{\min}$ from below:

$$\lambda_{\min} = \min_{i \in V \setminus C} -\lambda_i(x^{(C)}) \geq \min_{i \in V \setminus C} \frac{2|C| - 2|C \cap E^y_{\bar{k}}(i)| - 2 + \alpha}{|C|} \geq \frac{2m_\alpha(C, \mathcal{G}(y_{\bar{k}}))}{|C|} \,, \tag{5.8}$$

where we applied (5.7) in the inequality. Finally, we have

$$\frac{\lambda_{\min}}{\lambda_{\min} + 2L} \leq \frac{m_\alpha(C, \mathcal{G}(y_{\bar{k}}))}{m_\alpha(C, \mathcal{G}(y_{\bar{k}})) + 2|C|\delta_{\max} + |C|\alpha} \tag{5.9}$$

where we applied (5.7) together with (5.8) in the inequality. The thesis follows applying (5.9) to the RHS of (5.5). □

Remark 5.1. *It is a well known result that for any graph the maximal eigenvalue* $\delta_{\max}$ *of the adjacency matrix is less than or equal to* $d_{\max}$*, the maximum degree of a node (see, e.g., [13]). Then condition* (5.3) *can be replaced by*

$$\|x_{\bar{k}} - x^{(C)}\|_1 \leq \frac{m_\alpha(C, \mathcal{G}(y_{\bar{k}}))}{m_\alpha(C, \mathcal{G}(y_{\bar{k}})) + 2|C|d_{\max} + |C|\alpha} \,. \tag{5.10}$$

6 Numerical results

In this section we report on a numerical comparison of the methods. We remark that, even though these methods only find maximal s-defective cliques, they can still be applied as a heuristic to derive lower bounds on the maximum s-defective clique within a global optimization scheme. With our tests, we aim to achieve the followings:

- empirically verify the active set identification property of the proposed methods;
- prove that the proposed FW variant is faster than the FDFW on these problems, while mantaining the same solution quality;
- make a preliminary comparison between the FW methods and the CONOPT solver used in [28].

In the tests, the regularization parameters were set to $\alpha = 1$ and $\beta = 2/n^2$. An intuitive motivation for this choice of β can be given by imposing that the missing edges for an identified s-defective clique are always included in the support of the FW vertex. Formally, if $x_k = x^{(C)}$ with C an s-defective clique and $(y_k)_{ij} = 0$ with $\{i, j\} \in \binom{C}{2}$ we want to ensure that the FW vertex $s_k = (x^{(s_k)}, y^{(s_k)})$ is such that $y^{(s_k)}_{ij} = 1$. Now for $\{l, m\} \notin \binom{C}{2}$ and assuming $|C| < n$ (otherwise $C = V$ and the problem is trivial) we have

$$\frac{\partial}{\partial y_{ij}} h_{\mathcal{G}}(x_k, y_k) = \frac{2}{|C|^2} > \frac{2}{n^2} = \beta \geq \frac{\partial}{\partial y_{lm}} h_{\mathcal{G}}(x_k, y_k) \tag{6.1}$$

where the first equality and the last inequality easily follow from (3.4). From (6.1) it is then immediate to conclude that $\{i, j\}$ must be in the support of $y^{(s_k)}$.
We used the stepsize $\alpha_k = \bar{\alpha}_k$ with $\bar{\alpha}_k$ given by (S1) for $c = 2$, corresponding to an estimate of 0.5 for the Lipschitz constant L of $\nabla h_{\mathcal{G}}$. A gradient recycling scheme was adopted to use first order information more efficiently (see [27] for details). The code was written in MATLAB and the tests were performed on an Intel Core i7-10750H CPU 2.60GHz, 16GB RAM.
The 50 graph instances we used in the tests are taken from the Second DIMACS Implementation Challenge [20]. These graphs are a common benchmark to assess the performance of algorithms for maximum (defective) clique problems (see references in [28]), and the particular instances we selected coincide with the ones employed in [28] in order to ensure a fair comparison at least for the quality of the solutions. Following the rule adopted in [28], for each triple $(\mathcal{G}, s, \mathcal{A})$ with $\mathcal{G}$ a graph from the 50 instances considered, $s \in [1:4]$, $\mathcal{A}$ the FDFW or the FWdc, we set a global time limit of 600 seconds and employed a simple restarting scheme with up to 100 random starting points. The algorithms always completed 100 runs within the time limit, with the exception of 2 instances (see Table 2). For both algorithms the x component of the starting point was generated with MATLAB's function rand and then normalized dividing it by its sum. An analogous rule was applied to generate the y component for the starting point of the FDFW, while for the FWdc the y component was simply initialized to 0. For the stopping criterion, two conditions are required: the current support of the x components coincides with an s-defective clique, and the FW gap is less than or equal to $\varepsilon := 10^{-3}$. In the experiments, both algorithms always terminated having identified an s-defective clique, thus providing an empirical verification of the results we proved in this paper.

In the boxplots, each series consists of 50 values corresponding to aggregate data for the runs performed on the 50 instances. The data for the CONOPT solver are taken from [28]. The red lines represent the median of the values in each series, and the boxes extend from the 25th percentile q_1 of the observed data to the 75th percentile q_3. The whiskers cover all the other values in a range of $[q_1 - w(q_3 - q_1), q_3 + w(q_3 - q_1)]$, with the

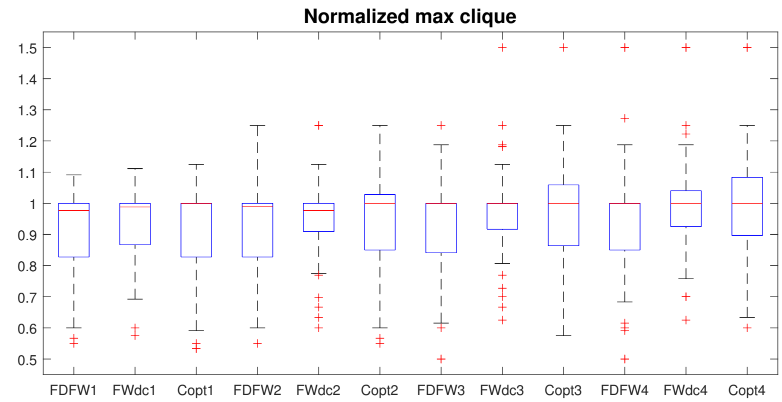

Figure 1: $\mathcal{A}i$ is the box plot of the maximum clique found within the 600 seconds/ 100 runs limit for each instance by the method $\mathcal{A}$ for $s = i$ divided by the maximum cardinality clique of the instance.

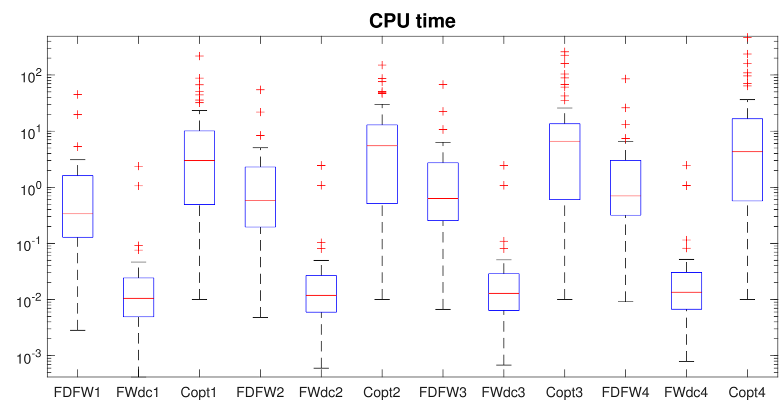

Figure 2: $\mathcal{A}i$ is the box plot of the average running time for each instance for the method $\mathcal{A}$ and $s = i$.

coefficient w equal to 2.7 times the standard deviation of the values.
In Fig. 1, the bar $\mathcal{A}i$ represents the distribution of the maximum cardinality of the s-defective clique found by method $\mathcal{A}$ with $s = i$, divided by the maximum clique cardinality of the instance. Notice that some data points are greater than 1, as expected since for $s > 0$ the cardinality of an s-defective clique can exceed the maximum clique cardinality. While the variance is higher for the max cliques found by the CONOPT solver, no significant differences can be seen for the median size of max cliques.
In Fig. 2, $\mathcal{A}i$ represents the distribution of average running times in seconds (on a logarithmic scale, explainng the asymmetry of the box plots) of method $\mathcal{A}$ for $s = i$. Here we can see that FWdc outperforms FDFW by about 1.5 orders of magnitude, which in turns outperforms the CONOPT solver by about 1 order of magnitude. This is even more impressive if we take into account the fact that the CONOPT solver is written in C++, while FW variants are written in MATLAB.

7 Appendix

7.1 Line searches

Here we briefly report some relevant results about line searches proved in [10] showing a connection between well known line searches and conditions (S1), (S2).

Recall that the exact line search stepsize is given by

$$\alpha_k \in \operatorname*{arg\,max}_{\alpha \in [0, \alpha_k^{\max}]} f(w_k + \alpha d_k)\,. \tag{7.1}$$

The stronger condition $\alpha_k = \max\{\arg\max_{\alpha \in [0, \alpha_k^{\max}]} f(w_k + \alpha d_k)\}$ is required for [9, Theorem 4.3], used to prove Corollary 5.1,
The stepsize α_k given by the Armijo line search always satisfies the condition

$$f(w_k + \alpha_k d_k) - f(w_k) \geq c_1 \alpha_k \nabla f(w_k)^\intercal d_k\,, \tag{7.2}$$

for some constant $c_1 \in (0, 1)$. This stepsize is produced by considering a sequence $\{\beta_k^{(j)}\}_{j \in \mathbb{N}_0}$ of tentative stepsizes given by $\beta_k^{(0)} = \alpha_k^{\max}$, $\beta_k^{(j+1)} = \gamma \beta_k^{(j)}$, with $\gamma \in (0, 1)$, and taking the largest tentative stepsize satisfying (7.2).
We report here for completeness results from [10] proving that Armijo and exact line search satisfy conditions (S1) and (S2).

Lemma 7.1. *Consider a sequence $\{w_k\}$ in $\mathcal{Q}$ such that $w_{k+1} = w_k + \alpha_k d_k$ with $\alpha_k \in \mathbb{R}_{\geq 0}$, $d_k \in \mathbb{R}^n$. Assume that d_k is a proper ascent direction in w_k, i.e. $\nabla f(w_k)^\intercal d_k > 0$.*

1. *If α_k is given by exact line search, then* (S1) *and* (S2) *are satisfied with $c = \frac{1}{L}$ and $\rho = \frac{1}{2}$.*
2. *If α_k is given by the Armijo line search described above, then* (S1) *and* (S2) *are satisfied with $c = \frac{2\gamma(1-c_1)}{L}$ and $\rho = c_1 \min\{1, 2\gamma(1 - c_1)\} < 1$.*

Proof. Point 1 follows from [10, Lemma B.1] and point 2 follows from [10, Lemma B.3]. □

7.2 Detailed numerical results

We report in this section detailed numerical results for each of the 50 graphs we used in our tests.

Table 1: Clique sizes for the FDFW

Graph	$s = 1$			$s = 2$			$s = 3$			$s = 4$		
	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std
brock200_1	19	15.6	1.3	21	17.0	1.4	22	17.9	1.2	21	18.6	1.3
brock200_2	10	8.0	0.9	11	9.0	0.9	11	9.5	0.8	12	9.5	1.0
brock200_3	13	10.2	1.1	14	11.5	1.0	14	12.2	0.9	14	12.5	0.9
brock200_4	15	11.9	1.2	16	13.2	1.1	17	14.0	1.1	17	14.3	1.3
brock400_1	21	17.5	1.4	22	18.8	1.3	23	20.0	1.3	24	20.9	1.2
brock400_2	21	17.8	1.4	22	18.9	1.2	23	20.0	1.4	24	20.9	1.1
brock400_3	21	17.8	1.5	22	19.0	1.3	23	20.2	1.2	24	20.9	1.2
brock400_4	21	17.7	1.5	22	18.7	1.3	23	20.0	1.3	24	21.1	1.3
c-fat200-1	12	11.5	1.1	12	11.6	0.9	12	11.6	0.9	12	11.7	0.8
c-fat200-2	24	20.8	3.1	24	21.9	2.0	24	22.2	1.2	24	22.3	0.9
c-fat200-5	58	45.7	11.7	58	54.4	5.0	58	56.2	3.0	58	56.7	2.3
c-fat500-1	14	13.6	1.1	14	13.7	0.7	14	13.7	0.7	14	13.7	0.7
c-fat500-2	26	25.1	2.0	26	25.4	1.4	26	25.6	1.1	26	25.6	0.9
c-fat500-5	64	55.2	10.1	64	59.7	5.2	64	62.2	2.6	64	62.4	3.1
c-fat500-10	126	99.8	20.8	126	114.8	11.8	126	123.7	4.1	126	125.2	1.0
hamming6-2	32	22.6	4.6	32	23.3	4.8	32	16.2	5.5	32	13.7	5.6
hamming6-4	4	3.7	0.5	5	3.7	0.5	4	3.2	0.7	4	2.7	0.9
hamming8-2	121	83.3	14.5	122	86.1	15.1	123	87.1	15.3	123	88.1	15.2
hamming8-4	14	10.5	1.2	15	11.9	1.1	15	12.4	1.0	15	12.7	1.0
hamming10-2 [2]	454	309.6	42.1	456	313.9	43.2	468	315.5	44.4	468	317.5	45.3
hamming10-4	31	27.8	1.3	32	29.1	1.3	33	30.3	1.2	34	31.4	1.3
johnson8-2-4	4	2.4	1.1	4	2.2	1.2	4	2.1	1.2	4	2.0	1.2
johnson8-4-4	14	9.5	1.6	14	9.9	1.6	14	9.9	1.8	14	7.6	2.3
johnson16-2-4	8	7.8	0.4	9	8.5	0.6	9	8.4	0.6	9	8.3	0.6
johnson32-2-4	16	14.9	0.7	17	15.9	0.8	17	16.4	0.7	18	17.3	0.8
keller4	11	8.3	0.9	12	9.5	0.7	12	10.0	1.0	13	9.9	0.9
keller5	20	17.2	1.1	20	18.3	1.0	21	19.4	0.9	23	20.4	1.0
MANN_a9	16	14.6	1.0	17	11.1	2.9	17	9.4	3.1	17	8.0	2.9
MANN_a27	118	117.6	0.5	119	118.6	0.6	120	119.5	0.7	121	120.4	0.7
MANN_a45 [3]	331	330.5	0.5	332	331.5	0.7	333	332.4	0.8	334	333.3	0.6
p_hat300-1	8	6.4	0.8	9	7.3	0.7	9	7.7	0.7	9	7.9	0.8
p_hat300-2	25	20.8	1.6	24	21.6	1.4	26	22.3	1.6	27	22.6	1.8
p_hat300-3	33	28.9	1.7	34	30.1	1.9	36	30.9	1.9	36	32.0	1.9
p_hat500-1	9	7.2	0.9	10	8.1	0.8	10	8.7	0.8	11	9.2	0.8
p_hat500-2	35	29.7	2.1	35	30.6	2.1	37	31.5	2.2	37	32.1	2.2
p_hat500-3	46	41.9	2.1	48	42.7	2.5	49	43.6	2.4	51	44.8	2.5
p_hat700-1	9	7.4	0.8	11	8.4	0.8	12	8.9	0.8	11	9.6	0.7
p_hat700-2	41	36.5	2.1	43	37.2	2.3	44	38.2	2.4	44	39.1	2.1
p_hat700-3	57	52.1	2.5	60	53.0	2.5	61	54.0	2.3	61	55.3	2.7
san200_0.7_1	17	15.6	0.5	18	16.8	0.4	18	17.2	0.7	19	17.3	0.8
san200_0.7_2	13	12.9	0.4	14	14.0	0.2	15	14.5	0.6	16	15.0	0.6
san200_0.9_1	46	45.4	0.5	47	46.6	0.6	48	47.6	0.7	49	48.5	0.7
san200_0.9_2	39	35.3	2.2	43	36.4	2.4	41	37.2	2.4	42	38.0	2.6
san200_0.9_3	32	28.2	2.0	33	28.8	2.2	35	29.7	2.6	35	30.0	2.3
san400_0.5_1	8	6.5	0.9	9	9.0	0.0	10	9.9	0.3	11	10.5	0.6
san400_0.7_1	22	20.6	0.7	22	21.9	0.3	23	23.0	0.2	24	23.9	0.2
san400_0.7_2	16	15.3	0.7	17	17.0	0.2	18	18.0	0.0	19	18.8	0.4
san400_0.7_3	13	12.1	1.0	14	13.9	0.4	15	14.9	0.2	16	15.6	0.5
sanr200_0.7	16	13.1	1.2	17	14.4	1.1	18	15.5	1.1	18	15.9	1.2
sanr200_0.9	37	32.4	2.0	39	33.7	2.1	40	34.7	2.2	40	35.7	2.0

[2] The time limit was reached in 29 runs
[3] The time limit was reached in 13 runs

Table 2: Running times for the FDFW

Graph	$s = 1$		$s = 2$		$s = 3$		$s = 4$	
	Time	Std	Time	Std	Time	Std	Time	Std
brock200_1	0.053	0.0013	0.179	0.0271	0.227	0.0468	0.261	0.0341
brock200_2	0.167	0.0297	0.218	0.0336	0.263	0.0727	0.327	0.1844
brock200_3	0.156	0.0216	0.196	0.0350	0.254	0.0697	0.318	0.1616
brock200_4	0.151	0.0178	0.189	0.0321	0.231	0.0483	0.278	0.0777
brock400_1	0.677	0.0989	0.968	0.0246	1.250	0.1431	1.495	0.0679
brock400_2	0.591	0.0197	0.983	0.0939	1.258	0.1185	1.493	0.0812
brock400_3	0.595	0.0201	0.972	0.0316	1.242	0.0518	1.493	0.1350
brock400_4	0.597	0.0190	0.961	0.0352	1.258	0.1030	1.519	0.1454
c-fat200-1	0.181	0.0057	0.318	0.0080	0.348	0.0082	0.376	0.0119
c-fat200-2	0.180	0.0079	0.314	0.0123	0.345	0.0169	0.379	0.0208
c-fat200-5	0.148	0.0071	0.264	0.0174	0.317	0.0247	0.356	0.0251
c-fat500-1	1.488	0.0249	2.441	0.0446	2.728	0.0458	3.016	0.0515
c-fat500-2	1.519	0.0257	2.579	0.1139	2.891	0.1459	3.256	0.1961
c-fat500-5	1.600	0.0641	2.408	0.1294	2.868	0.2600	3.311	0.4613
c-fat500-10	1.664	0.1425	2.436	0.3427	2.719	0.5139	3.073	0.5811
hamming6-2	0.008	0.0011	0.015	0.0036	0.020	0.0037	0.026	0.0107
hamming6-4	0.013	0.0005	0.024	0.0009	0.025	0.0018	0.027	0.0013
hamming8-2	0.172	0.0350	0.267	0.0515	0.346	0.0604	0.428	0.0645
hamming8-4	0.172	0.0296	0.312	0.0283	0.368	0.0487	0.409	0.0330
hamming10-2	19.686	8.3243	21.857	10.8597	22.468	9.8968	25.991	11.2051
hamming10-4	5.295	0.4853	8.365	0.5668	10.680	0.7274	13.257	0.7791
johnson8-2-4	0.003	0.0005	0.005	0.0005	0.007	0.0030	0.009	0.0049
johnson8-4-4	0.011	0.0009	0.020	0.0044	0.024	0.0033	0.030	0.0067
johnson16-2-4	0.037	0.0048	0.065	0.0048	0.080	0.0030	0.097	0.0041
johnson32-2-4	1.235	0.3833	1.547	0.2246	1.900	0.1825	2.325	0.2487
keller4	0.081	0.0163	0.164	0.0733	0.209	0.1159	0.253	0.1667
keller5	2.354	0.3020	3.819	0.6645	5.140	1.3110	6.438	2.0281
MANN_a9	0.005	0.0013	0.012	0.0081	0.018	0.0132	0.027	0.0260
MANN_a27	1.757	0.3220	2.504	0.5624	3.362	0.7169	4.371	0.8807
MANN_a45	44.945	5.8122	54.466	8.6046	67.499	9.7882	85.228	11.6795
p_hat300-1	0.348	0.0133	0.600	0.0408	0.634	0.0592	0.683	0.0746
p_hat300-2	0.323	0.0105	0.548	0.0177	0.631	0.0246	0.717	0.0754
p_hat300-3	0.262	0.0182	0.401	0.0197	0.466	0.0237	0.534	0.0332
p_hat500-1	1.209	0.0944	1.996	0.2460	2.293	0.8532	2.449	0.5990
p_hat500-2	1.353	0.0713	2.295	0.0999	2.836	0.1464	3.261	0.1831
p_hat500-3	1.369	0.1112	2.161	0.1798	2.535	0.1734	2.803	0.1708
p_hat700-1	2.549	0.1642	4.080	0.3388	4.468	0.8064	4.943	1.0512
p_hat700-2	2.765	0.1417	5.040	0.2496	6.340	0.2623	7.396	0.3281
p_hat700-3	3.084	0.2889	4.929	0.5101	5.871	0.5302	6.607	0.5871
san200_0.7_1	0.159	0.0454	0.277	0.0429	0.358	0.0659	0.481	0.0947
san200_0.7_2	0.271	0.0732	0.402	0.1320	0.404	0.1214	0.452	0.0791
san200_0.9_1	0.128	0.0218	0.220	0.0302	0.296	0.0282	0.373	0.0332
san200_0.9_2	0.129	0.0171	0.207	0.0225	0.260	0.0278	0.319	0.0295
san200_0.9_3	0.105	0.0134	0.183	0.0186	0.237	0.0392	0.319	0.0786
san400_0.5_1	1.605	0.1003	2.178	0.5510	2.011	0.6413	2.860	0.8604
san400_0.7_1	1.706	0.6290	2.266	0.6993	2.469	0.1244	2.956	0.1424
san400_0.7_2	1.684	0.5364	2.069	0.7020	2.133	0.3488	2.497	0.1083
san400_0.7_3	1.713	0.3159	2.131	0.7214	2.020	0.6281	2.244	0.4235
sanr200_0.7	0.103	0.0045	0.186	0.0248	0.227	0.0476	0.297	0.0961
sanr200_0.9	0.108	0.0110	0.173	0.0175	0.220	0.0178	0.283	0.0223

Table 3: Clique sizes for the FWdc

Graph	$s=1$			$s=2$			$s=3$			$s=4$		
	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std
brock200_1	21	18.2	1.01	21	18.2	0.90	21	18.5	1.02	22	18.7	1.05
brock200_2	10	8.6	0.91	11	8.9	0.81	11	9.3	0.91	12	9.5	0.86
brock200_3	13	11.4	0.87	14	11.5	0.88	15	11.9	1.02	14	12.0	0.97
brock200_4	16	13.3	1.01	16	13.7	0.93	16	14.0	1.01	17	14.1	1.14
brock400_1	24	20.7	1.10	25	21.4	1.30	24	21.3	1.25	25	21.7	1.33
brock400_2	24	20.9	1.12	25	21.3	1.26	26	21.4	1.32	25	21.8	1.16
brock400_3	24	20.7	0.97	24	21.0	1.18	25	21.2	1.09	24	21.4	1.06
brock400_4	23	20.6	1.16	23	21.1	1.07	24	21.5	1.07	25	21.6	1.38
c-fat200-1	12	11.4	1.69	12	10.8	2.64	12	8.7	4.31	12	7.9	4.49
c-fat200-2	24	21.2	4.76	24	20.3	6.25	24	18.7	7.63	24	17.2	8.45
c-fat200-5	58	55.1	7.08	58	53.3	9.88	58	52.0	12.88	58	53.3	10.66
c-fat500-1	14	13.5	1.13	14	12.7	2.69	14	10.8	4.57	14	9.5	5.24
c-fat500-2	26	25.5	2.47	26	24.5	5.23	26	22.8	7.65	26	21.6	8.73
c-fat500-5	64	60.8	10.85	64	61.7	8.62	64	58.8	14.73	64	56.2	18.38
c-fat500-10	126	122.6	12.02	126	119.8	17.20	126	118.0	24.57	126	115.7	29.14
hamming6-2	32	28.6	4.58	32	27.9	4.63	32	27.5	4.40	32	27.4	4.09
hamming6-4	4	3.7	0.46	5	4.2	0.61	6	4.4	0.66	6	4.8	0.63
hamming8-2	128	121.1	9.18	128	120.2	9.07	128	118.8	10.29	128	116.9	11.86
hamming8-4	16	12.7	2.68	16	12.6	2.31	16	12.5	2.24	17	12.8	2.28
hamming10-2	512	498.9	14.52	512	497.0	15.59	512	495.5	17.16	512	493.9	18.06
hamming10-4	36	31.6	2.93	36	32.2	2.67	37	32.1	2.80	37	32.9	2.32
johnson8-2-4	4	4.0	0.00	5	4.9	0.27	5	5.0	0.10	6	5.3	0.47
johnson8-4-4	14	11.9	1.73	14	11.7	1.46	14	11.8	1.35	15	11.9	1.18
johnson16-2-4	8	8.0	0.00	9	9.0	0.00	9	9.0	0.00	10	9.8	0.39
johnson32-2-4	16	16.0	0.00	17	17.0	0.00	17	17.0	0.00	18	17.8	0.41
keller4	12	9.3	1.15	12	9.7	0.81	13	10.1	0.80	13	10.6	0.82
keller5	27	20.7	1.77	26	21.1	1.72	26	21.5	1.50	27	21.5	1.55
MANN_a9	17	16.3	0.67	18	16.8	0.71	19	17.4	0.72	19	17.6	0.82
MANN_a27	120	118.2	0.43	120	119.2	0.37	121	120.1	0.43	122	121.1	0.38
MANN_a45	332	331.0	0.17	333	332.0	0.17	334	333.0	0.20	335	334.0	0.22
p_hat300-1	8	6.9	0.68	9	7.1	0.79	9	7.4	0.75	9	7.5	0.85
p_hat300-2	26	21.9	1.25	25	22.0	1.20	25	22.2	1.18	26	22.2	1.27
p_hat300-3	35	31.4	1.35	34	31.8	1.32	35	31.8	1.30	36	32.2	1.19
p_hat500-1	10	7.9	0.81	10	8.1	0.83	10	8.1	0.90	11	8.4	0.96
p_hat500-2	35	31.6	1.84	35	31.9	1.73	36	31.8	1.79	35	31.8	1.57
p_hat500-3	48	44.8	1.58	49	44.8	1.76	49	45.1	1.73	49	45.4	1.67
p_hat700-1	9	8.0	0.70	10	8.2	0.78	10	8.4	0.78	10	8.6	0.77
p_hat700-2	44	39.9	1.77	43	40.0	1.77	44	40.2	1.81	44	40.0	2.00
p_hat700-3	62	57.0	1.94	60	57.6	1.73	61	57.5	1.85	62	57.9	1.94
san200_0.7_1	18	16.6	0.84	19	17.1	1.23	20	18.0	1.17	21	18.7	1.44
san200_0.7_2	15	13.1	0.26	15	14.0	0.14	16	14.7	0.49	16	15.2	0.42
san200_0.9_1	65	48.9	4.23	65	49.4	3.94	68	50.1	4.11	70	50.6	3.85
san200_0.9_2	52	39.5	2.74	52	39.9	2.53	55	40.8	3.32	55	41.5	3.35
san200_0.9_3	36	33.4	1.24	36	33.7	1.30	38	34.1	1.43	39	34.4	1.51
san400_0.5_1	9	8.1	0.30	10	9.0	0.32	10	9.6	0.50	12	10.1	0.57
san400_0.7_1	23	21.8	0.77	24	22.5	0.86	25	22.9	1.30	25	23.5	1.73
san400_0.7_2	23	17.4	1.16	20	17.9	0.78	21	18.5	0.93	21	19.0	0.75
san400_0.7_3	17	15.1	0.93	18	15.6	0.85	18	16.2	0.92	19	16.6	0.88
sanr200_0.7	17	14.9	0.86	18	15.2	1.09	17	15.6	0.89	19	15.8	0.99
sanr200_0.9	41	37.5	1.79	41	37.5	1.67	42	38.1	1.75	43	38.3	1.73

Table 4: Running times for the FWdc

Graph	$s = 1$		$s = 2$		$s = 3$		$s = 4$	
	Time	Std	Time	Std	Time	Std	Time	Std
brock200_1	0.0060	0.005 60	0.0060	0.000 64	0.0064	0.000 70	0.0069	0.000 83
brock200_2	0.0045	0.000 53	0.0051	0.000 41	0.0051	0.000 50	0.0053	0.000 42
brock200_3	0.0044	0.000 45	0.0050	0.000 56	0.0052	0.000 47	0.0052	0.000 38
brock200_4	0.0045	0.000 39	0.0053	0.000 55	0.0055	0.000 54	0.0057	0.000 52
brock400_1	0.0122	0.000 88	0.0142	0.001 57	0.0141	0.000 82	0.0144	0.000 80
brock400_2	0.0129	0.000 66	0.0133	0.001 18	0.0144	0.000 95	0.0147	0.000 72
brock400_3	0.0129	0.000 60	0.0138	0.000 98	0.0139	0.000 57	0.0146	0.000 96
brock400_4	0.0130	0.000 65	0.0136	0.001 00	0.0142	0.000 69	0.0146	0.000 67
c-fat200-1	0.0065	0.000 61	0.0070	0.000 80	0.0068	0.000 52	0.0071	0.000 74
c-fat200-2	0.0060	0.001 00	0.0067	0.001 09	0.0067	0.000 98	0.0067	0.000 91
c-fat200-5	0.0071	0.002 16	0.0081	0.002 37	0.0084	0.002 61	0.0089	0.002 59
c-fat500-1	0.0351	0.002 54	0.0370	0.001 88	0.0366	0.002 10	0.0364	0.002 43
c-fat500-2	0.0311	0.003 60	0.0331	0.003 74	0.0334	0.003 55	0.0334	0.003 14
c-fat500-5	0.0296	0.006 21	0.0330	0.007 30	0.0343	0.005 99	0.0347	0.005 92
c-fat500-10	0.0390	0.013 53	0.0400	0.016 17	0.0422	0.016 85	0.0433	0.017 48
hamming6-2	0.0014	0.000 12	0.0021	0.000 17	0.0023	0.000 27	0.0026	0.000 29
hamming6-4	0.0009	0.000 15	0.0011	0.000 14	0.0012	0.000 14	0.0013	0.000 15
hamming8-2	0.0204	0.001 62	0.0245	0.001 95	0.0287	0.002 72	0.0302	0.003 08
hamming8-4	0.0060	0.000 55	0.0065	0.000 59	0.0068	0.000 75	0.0069	0.000 75
hamming10-2	1.0562	0.064 27	1.0825	0.067 12	1.0816	0.075 77	1.0717	0.078 35
hamming10-4	0.0762	0.002 68	0.0805	0.004 00	0.0805	0.003 79	0.0823	0.003 42
johnson8-2-4	0.0004	0.000 06	0.0006	0.000 09	0.0007	0.000 13	0.0008	0.000 13
johnson8-4-4	0.0011	0.000 09	0.0015	0.000 17	0.0018	0.000 21	0.0020	0.000 27
johnson16-2-4	0.0023	0.000 28	0.0029	0.000 36	0.0034	0.000 41	0.0037	0.000 55
johnson32-2-4	0.0218	0.002 18	0.0240	0.002 29	0.0248	0.002 32	0.0259	0.002 82
keller4	0.0034	0.000 34	0.0044	0.000 53	0.0044	0.000 65	0.0051	0.000 62
keller5	0.0428	0.002 07	0.0446	0.002 06	0.0455	0.002 39	0.0460	0.002 36
MANN_a9	0.0010	0.000 13	0.0014	0.000 17	0.0016	0.000 24	0.0018	0.000 31
MANN_a27	0.0910	0.006 10	0.1029	0.007 23	0.1096	0.007 23	0.1150	0.006 99
MANN_a45	2.3774	0.075 29	2.4419	0.101 45	2.4544	0.083 25	2.4712	0.091 95
p_hat300-1	0.0071	0.000 29	0.0078	0.000 68	0.0080	0.000 46	0.0083	0.000 40
p_hat300-2	0.0085	0.000 31	0.0096	0.000 61	0.0103	0.000 80	0.0104	0.000 62
p_hat300-3	0.0095	0.000 46	0.0110	0.000 70	0.0119	0.000 72	0.0126	0.000 92
p_hat500-1	0.0163	0.000 41	0.0169	0.000 39	0.0170	0.000 43	0.0171	0.000 57
p_hat500-2	0.0208	0.001 29	0.0224	0.001 33	0.0228	0.001 33	0.0235	0.001 31
p_hat500-3	0.0243	0.000 85	0.0266	0.001 00	0.0278	0.001 25	0.0281	0.001 00
p_hat700-1	0.0298	0.000 70	0.0307	0.000 91	0.0309	0.001 09	0.0311	0.000 79
p_hat700-2	0.0394	0.001 08	0.0421	0.002 11	0.0427	0.001 43	0.0431	0.001 61
p_hat700-3	0.0466	0.001 51	0.0498	0.001 81	0.0508	0.001 77	0.0520	0.001 44
san200_0.7_1	0.0049	0.000 59	0.0056	0.000 71	0.0061	0.000 66	0.0064	0.000 68
san200_0.7_2	0.0048	0.000 68	0.0063	0.000 93	0.0073	0.001 34	0.0081	0.001 48
san200_0.9_1	0.0086	0.000 76	0.0108	0.001 04	0.0118	0.001 07	0.0129	0.001 49
san200_0.9_2	0.0075	0.000 53	0.0092	0.000 75	0.0101	0.000 79	0.0113	0.000 73
san200_0.9_3	0.0070	0.000 58	0.0084	0.000 67	0.0092	0.000 69	0.0098	0.000 84
san400_0.5_1	0.0116	0.000 56	0.0128	0.000 94	0.0141	0.001 72	0.0142	0.001 36
san400_0.7_1	0.0130	0.000 56	0.0143	0.000 77	0.0152	0.001 01	0.0158	0.001 10
san400_0.7_2	0.0128	0.000 64	0.0139	0.000 68	0.0147	0.000 82	0.0157	0.001 38
san400_0.7_3	0.0125	0.000 43	0.0136	0.000 59	0.0143	0.000 72	0.0148	0.000 84
sanr200_0.7	0.0049	0.000 61	0.0057	0.000 71	0.0058	0.000 64	0.0062	0.000 78
sanr200_0.9	0.0070	0.000 53	0.0087	0.000 72	0.0095	0.000 78	0.0119	0.000 90

References

[1] Martin Aigner, Günter M Ziegler, Karl H Hofmann, and Paul Erdos. *Proofs from the Book*, volume 274. Springer, 2010.

[2] Mohammad Ali Bashiri and Xinhua Zhang. Decomposition-invariant conditional gradient for general polytopes with line search. In *Advances in Neural Information Processing Systems*, pages 2690–2700, 2017.

[3] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.

[4] Immanuel M Bomze. Evolution towards the maximum clique. *Journal of Global Optimization*, 10(2):143–164, 1997.

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

[6] Immanuel M. Bomze. Regularity versus degeneracy in dynamics, games, and optimization: a unified approach to different aspects. *SIAM Rev.*, 44(3):394–414, 2002.

[7] Immanuel M Bomze, Marco Budinich, Panos M Pardalos, and Marcello Pelillo. The maximum clique problem. In *Handbook of Combinatorial Optimization*, pages 1–74. Springer, 1999.

[8] Immanuel M Bomze, Mirjam Dür, Etienne De Klerk, Cornelis Roos, Arie J Quist, and Tamás Terlaky. On copositive programming and standard quadratic optimization problems. *Journal of Global Optimization*, 18(4):301–320, 2000.

[9] Immanuel M Bomze, Francesco Rinaldi, and Samuel Rota Bulo. First-order methods for the impatient: Support identification in finite time with convergent Frank-Wolfe variants. *SIAM Journal on Optimization*, 29(3):2211–2226, 2019.

[10] Immanuel M Bomze, Francesco Rinaldi, and Damiano Zeffiro. Active set complexity of the away-step Frank–Wolfe algorithm. *SIAM Journal on Optimization*, 30(3):2470–2500, 2020.

[11] Xiaoyu Chen, Yi Zhou, Jin-Kao Hao, and Mingyu Xiao. Computing maximum k-defective cliques in massive graphs. *Computers & Operations Research*, 127:105131, 2021.

[12] Kenneth L Clarkson. Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. *ACM Transactions on Algorithms (TALG)*, 6(4):1–30, 2010.

[13] Dragiša Cvetković and Peter Rowlinson. The largest eigenvalue of a graph: A survey. *Linear and multilinear algebra*, 28(1-2):3–33, 1990.

[14] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

[15] Robert M Freund, Paul Grigas, and Rahul Mazumder. An extended Frank–Wolfe method with in-face directions, and its application to low-rank matrix completion. *SIAM Journal on Optimization*, 27(1):319–346, 2017.

[16] Dan Garber. Revisiting Frank-Wolfe for polytopes: Strict complementary and sparsity. *arXiv preprint arXiv:2006.00558*, 2020.

[17] Jacques Guelat and Patrice Marcotte. Some comments on Wolfe's away step. *Mathematical Programming*, 35(1):110–119, 1986.

[18] James T Hungerford and Francesco Rinaldi. A general regularized continuous formulation for the maximum clique problem. *Mathematics of Operations Research*, 44(4):1161–1173, 2019.

[19] Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *ICML (1)*, pages 427–435, 2013.

[20] David S Johnson. Cliques, coloring, and satisfiability: second dimacs implementation challenge. *DIMACS series in discrete mathematics and theoretical computer science*, 26:11–13, 1993.

[21] Thomas Kerdreux, Lewis Liu, Simon Lacoste-Julien, and Damien Scieur. Affine invariant analysis of Frank-Wolfe on strongly convex sets. *arXiv preprint arXiv:2011.03351*, 2020.

[22] Simon Lacoste-Julien. Convergence rate of Frank-Wolfe for non-convex objectives. *arXiv preprint arXiv:1607.00345*, 2016.

[23] Simon Lacoste-Julien and Martin Jaggi. On the global linear convergence of Frank-Wolfe optimization variants. In *Advances in Neural Information Processing Systems*, pages 496–504, 2015.

[24] R Duncan Luce and Albert D Perry. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, 1949.

[25] Theodore S Motzkin and Ernst G Straus. Maxima for graphs and a new proof of a theorem of turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

[26] Jeffrey Pattillo, Nataly Youssef, and Sergiy Butenko. On clique relaxation models in network analysis. *European Journal of Operational Research*, 226(1):9–18, 2013.

[27] Francesco Rinaldi and Damiano Zeffiro. Avoiding bad steps in Frank Wolfe variants. *arXiv preprint arXiv:2012.12737*, 2020.

[28] Vladimir Stozhkov, Austin Buchanan, Sergiy Butenko, and Vladimir Boginski. Continuous cubic formulations for cluster detection problems in networks. *Mathematical Programming*, online, 2020.

[29] Svyatoslav Trukhanov, Chitra Balasubramaniam, Balabhaskar Balasundaram, and Sergiy Butenko. Algorithms for detecting optimal hereditary structures in graphs, with application to clique relaxations. *Computational Optimization and Applications*, 56(1):113–130, 2013.

[30] Qinghua Wu and Jin-Kao Hao. A review on algorithms for maximum clique problems. *European Journal of Operational Research*, 242(3):693–709, 2015.

[31] Haiyuan Yu, Alberto Paccanaro, Valery Trifonov, and Mark Gerstein. Predicting interactions in protein networks by completing defective cliques. *Bioinformatics*, 22(7):823–829, 2006.