Generalization to hypergraphs

September 18, 2023

1 Introduction

In recent years, the rapid growth of data availability has posed new challenges and opportunities in various fields, from image analysis and bioinformatics to marketing and social network analysis. The ability to extract meaningful insights and patterns from large and complex datasets has become a crucial endeavor. Data clustering, a fundamental technique in unsupervised learning, plays a pivotal role in uncovering hidden structures within data by grouping similar data points together. The clique model, dating back at least to [11] about social networks, refers to subsets where every elements is in a direct relation to all of the others.

The problem of finding maximal cliques has a huge field of applications in domains including telecommunication networks, biochemistry, financial networks, and scheduling (see [19] and [12]). The definition of a clique has been extended to hypergraphs, which can model relations between more than two elements, and numerous heuristics has been proposed and tested on real world data for the finding of maximum cliques in hypergraphs (see e.g. [7], [15], [18], [21]).

Since the strict requirement that every two elements have a direct relation is often not satisfied in practice (for example due to experimental errors when working on proteins, see [20] where the concept of s-defective clique was introduced), many relaxations of the clique model have been proposed (see, e.g., [14] for a survey). In this work we are interested in s-defective cliques (see also for example [17] and [13]), where s links can be missing.

In [5], following a long history of continuous formulations of this kind of problems, the authors defined a regularized version of a cubic continuous formulation for the Maximum s-defective Clique Problem proposed in [16], and then applied variants of the classic Frank–Wolfe (FW) method (see [8]) to this formulation. The goal of this work is to generalize this continuous regularized formulation to k-uniform hypergraphs, where an edge is no longer a link between a pair of nodes but between $k > 2$ different nodes.

This work is organized under the same structure as [5] : after giving some basic notations in section 2, we summarize previous works on the topic before studying the regularized maximum s-defective clique formulation for hypergraphs in section 3, which we easily extends to s-plexes in section 5. Finally, in section 6, we generalize the results obtained for the FDFW algorithm and the FW variant tailored to the Maximum s-defective Clique Problem that can be found in [5].

2 Common definitions and notations

A *k-graph* is a pair $G := (V, E)$, where $V := [1, \dots, n]$ is a finite set of *vertices* and $E \subseteq \binom{V}{k}$ is a set of k-subsets of V, each of which is called a *hyperedge*. 2-graphs are typically called *graphs*. We denote by $\overline{G} := (V, \overline{E} := \binom{V}{k} \setminus E)$ its complementary. We denote by $G(y) := (V, E(y))$ the weighted graph which has the vertices of G with weights equal to 1 and vertices of $\mathrm{supp}(y)$ with weights equal to y_e, where y_e is the coefficient in y coding for edge e.

The lagrangian of graph G L_G is defined such that $L_G(x) := \sum_{e \in E} \prod_{i \in e} x_i$. The lagrangian of weighted graph $G(y)$ $L_{G(y)}$ is defined such that $L_{G(y)}(x) := \sum_{e \in E} y_e \prod_{i \in e} x_i$.

We denote $\Delta := \{x \in \mathbb{R}^n | x \geq 0 \text{ and } \mathbf{1}^T x = 1\}$ the n-dimensional simplex. The search spaces in the following problems will be defined later and will always be denoted $\mathcal{P}_s$.

We denote $\mathbf{1}_C$ the characteristic function with logical expression C i.e. $\mathbf{1}_C = 0$ when C is false and $\mathbf{1}_C = 1$ when C is true, and $\mathbf{e}_i$ the i-th column of the identity matrix i.e. a vector with a 1 at index i and 0 everywhere else.

A characteristic vector of a subset A is defined by $x^{(A)} := (\frac{\mathbf{1}_{i\in A}}{|A|})_i \in \Delta$.

For any vector x and real p, we denote $x^{[p]} := (x_i^p)_i$ the element-wise power of x.

For $p \in \mathcal{P}_s$ we define as $T_{\mathcal{P}_s}(p) := \{v - p | v \in \mathcal{P}_s\}$ the cone of feasible directions at p in $\mathcal{P}_s$, while for $r \in \mathbb{R}^{|V|+|\overline{E}|}$ we define $T^0_{\mathcal{P}_s}(p, r)$ as the intersection between $T_{\mathcal{P}_s}(p)$ and the plane orthogonal to r :

$$T^0_{\mathcal{P}_s}(p, r) := \left\{d \in T_{\mathcal{P}_s}(p) | d^T r = 0\right\}$$

3 A continuous characterization of maximal s-defective cliques in k-graphs

3.1 Specific definitions and notations

A clique is a fully connected subset of vertices. A clique is said to be *maximal* if it is not contained in any other clique, while it is called *maximum* if it has maximum cardinality. An s-defective clique is a relaxation of this definition allowing to have up to s missing edges in the clique. Finding the biggest s-defective clique in a graph is known as the Maximum s-defective Clique Problem, which is a combinatorial problem.

Let $\mathcal{D}_s(G) := \{y \in [0,1]^{\overline{E}} | \mathbf{1}^T y \leq s\}$ and $\mathcal{P}_s := \Delta \times \mathcal{D}_s(G)$. The rest of the section is dedicated to the derivation of a continuous formulation of the Maximum s-defective Clique Problem.

4 Previous work

Turàn's theorem states that :

$$\max_{\text{s.t. } x\in\Delta} x^T A x = 1 - \frac{1}{\omega(G)} \tag{1}$$

where $\Delta = \{x \in \mathbb{R}^n | x \geq 0 \text{ and } 1^T x = 1\}$, A is the adjacency matrix of the graph, and $\omega(G)$ is the size of the maximum clique of the graph. Characteristic vectors of maximum cliques are global maxima of the objective function, but some global maxima are not linked to maximum cliques (these maximizers are called spurious solutions). To address it, in [1] a regularization was introduced to suppress those spurious maxima :

$$\max_{\text{s.t. } x\in\Delta} x^T A x + \alpha \|x\|_2^2 \tag{2}$$

with $\alpha \in (0, 1)$. The global and local maximizers of this problem are then the characteristic vectors of respectively maximum and maximal clique.

The continuous unregularized formulation for the Maximum Clique Problem was generalized in [16] to the Maximum s-defective Clique Problem :

$$\max_{\text{s.t. } (x,y)\in\Delta\times\mathcal{D}_s(G)} x^T (A + A(y)) x \tag{3}$$

where $\Delta = \{x \in \mathbb{R}^n | x \geq 0 \text{ and } 1^T x = 1\}$ and $\mathcal{D}_s(G) = \{y \in \{0,1\}^{\overline{E}} | 1^T y \leq s\}$ (which is the set of all the possible choices of s additional "fake" edges to complete the graph). Here, $A(y)$ is the adjacency matrix of the graph made of the s selected fake edges. In the end, the problem can be formulated as such : finding the s best edges to add to the original graph to have the largest clique in the augmented graph.

$y \in \{0,1\}^{\overline{E}}$ in the definition of $\mathcal{D}_s(G)$ can be relaxed to $y \in [0,1]^{\overline{E}}$ in order to obtain a fully continuous formulation.

Some maximizers of the previous problem are spurious, and like for the Maximum Clique Problem it is possible to add a regularization to suppress those maxima, which was done in [5] :

$$\max_{\text{s.t. } (x,y)\in\Delta\times\mathcal{D}_s(G)} x^T(A+A(y))x + \frac{\alpha}{2}\|x\|^2 + \frac{\beta}{2}\|y\|^2 \tag{4}$$

where $\alpha \in (0,2)$ and $\beta > 0$. This formulation ensures that the local and global maximizers of the above problems are strictly characteristic vectors of respectively maximal and maximum s-defective cliques of the graph. The authors also introduced a variant of the Frank-Wolfe algorithm adapted to this formulation and proved various convergence results.

Finally, the regularized maximum clique formulation was generalized to k-uniform hypergraphs in [6] :

$$\min_{\text{s.t. } x\in\Delta} \sum_{e\in\overline{E}} \prod_{i\in e} x_i + \tau\|x\|_k^k \tag{5}$$

where $0 < \tau \leq \frac{1}{k(k-1)}$ (with strict inequality when $k=2$). Beware that contrary to the previous formulations that searched for a maximum on the graph, this one searches for a minimum on the **complement** of the graph (the sum is indexed on $\overline{E}$ and not E).

4.1 Formulation considered

Let $G := (V,E)$ be a k-graph with vertices V and edges E. Let $0 < \alpha \leq \frac{1}{k(k-1)}$ (with strict inequality for $k=2$) and $\beta > 0$ and consider the following problem :

$$\min_{\text{s.t. } (x,y)\in\mathcal{P}_s} L_{\overline{G}}(x) - L_{G(y)}(x) + \alpha\|x\|_k^k - \beta\|y\|_2^2 \tag{6}$$

In the following, we will denote $h(x,y) := L_{\overline{G}}(x) - L_{G(y)}(x) + \alpha\|x\|_k^k - \beta\|y\|_2^2$. We claim that minimizers of this problem are attained at $p = (x^{(C)}, y^{(p)})$, where $s \geq l = \mathbf{1}^T y^{(p)} \in \mathbb{N}$, with C an l-defective clique in G which is also a maximal clique in $G \cup G(y^{(p)})$, and $y^{(p)} \in \{0,1\}^{\overline{E}}$ such that $y_e^{(p)} = 1$ for every $e \in \binom{C}{k} \cap \overline{E}$ and with $\mathrm{supp}(y^{(p)})$ of maximum cardinality under these constraints.

Remark 4.1. *To explain briefly the motivation behind this formulation, it comes pretty straightforwardly from the combination of the ideas developed in [6] (namely the objective function in (5), and that it is necessary to search for a minimum on the complement of the graph instead of a maximum on the graph itself) and in [5] (namely that the Maximum s-defective Clique Problem can be seen as two nested problems : first finding the best s fake edges to add to the graph and then reducing the problem to a standard Maximum Clique Problem on the augmented graph with an additional regularization which is visible in (4)).*

These ideas suggest the following formulation :

$$\min_{s.t.\ (x,y)\in\mathcal{P}_s} L_{\overline{G\cup G(y)}}(x) + \alpha\|x\|_k^k \pm \beta\|y\|_l^l \tag{7}$$

where $\pm$ and l are to be chosen later.

Then it seems natural to develop $L_{\overline{G\cup G(y)}}(x) = L_{\overline{G}\setminus\overline{G(y)}}(x) = L_{\overline{G}}(x) - L_{\overline{G(y)}}(x)$. At this point $L_{\overline{G(y)}}(x)$ only has a meaning when $y \in \{0,1\}^{\overline{E}}$, and it seems logical to define it as $L_{\overline{G(y)}}(x) := \sum_{e\in\overline{E}} y_e \prod_{i\in e} x_i$ in order to obtain something behaving similarly to $x^T A(y) x$ in [5]. This is even more obvious if we notice that the lagrangian can be expressed as a tensor product by $L_G(x) = \mathcal{A}x^k$ where $\mathcal{A}$ is the adjacency tensor of G. Then $\pm$ and l can be chosen to make the reasoning in [5] works.

Recall that in our polytope-constrained setting, (second order) sufficient conditions for the local minimality of $p \in \mathcal{P}_s$ are

$$\nabla h(p)^T d \geq 0 \text{ for all } d \in T_{\mathcal{P}_s}(p) \tag{8}$$

and

$$d^T \nabla^2 h(p) d > 0 \text{ for all } d \in T^0_{\mathcal{P}_s}(p, \nabla h(p)) \tag{9}$$

Recall also that we have the following result (given by [6]) :

Theorem *Let G be a k-graph and $0 < \alpha \leq \frac{1}{k(k-1)}$ (with strict inequality for $k = 2$). A vector $x \in \Delta$ is a local (global) minimizer of (50) if and only if it is the characteristic vector of a maximal (maximum) clique of G.*

4.2 Useful formulas

$$L_{\overline{G}}(x) - L_{G(y)}(x) = \sum_{e \in \overline{E}} \prod_{i \in e} x_i - \sum_{e \in E(y)} y_e \prod_{i \in e} x_i \tag{10}$$

When $y \in \{0,1\}^{\overline{E}}$, we can rewrite

$$L_{\overline{G}}(x) - L_{G(y)}(x) = L_{\overline{G} \setminus G(y)}(x) = L_{\overline{G \cup G(y)}}(x) \tag{11}$$

The first and second order derivatives of h are the following :

$$\frac{\partial h}{\partial x_j}(x,y) = \sum_{e \in \overline{E}} \mathbf{1}_{j \in e} \prod_{i \in e \setminus \{j\}} x_i - \sum_{e \in E(y)} y_e \mathbf{1}_{j \in e} \prod_{i \in e \setminus \{j\}} x_i + \alpha k x_j^{k-1} \tag{12}$$

$$\frac{\partial^2 h}{\partial x_i \partial x_j}(x,y) = \mathbf{1}_{i \neq j} \left[\sum_{e \in \overline{E}} \mathbf{1}_{i,j \in e} \prod_{l \in e \setminus \{i,j\}} x_i - \sum_{e \in E(y)} y_e \mathbf{1}_{i,j \in e} \prod_{l \in e \setminus \{i,j\}} x_i \right] + \mathbf{1}_{i=j} \alpha k(k-1) x_i^{k-2} \tag{13}$$

$$\frac{\partial h}{\partial y_e}(x,y) = -\prod_{i \in e} x_i - 2\beta y_e \tag{14}$$

$$\frac{\partial^2 h}{\partial y_e \partial y_{e'}}(x,y) = -2\beta \mathbf{1}_{e=e'} \tag{15}$$

For readers that are not familiar with tensors, we advise reading [10], but knowledge of tensors are necessary only for convergence results at the end of the paper. Denote $\mathcal{A}(G)$ the adjacency tensor of graph $G = (V,E)$, defined such that

$$\forall e \in \binom{|V|}{k}, \; (\mathcal{A}(G))_e := \mathbf{1}_{e \in E} \tag{16}$$

which is a symmetric tensor. Denoting $e = (i_1, \ldots, i_k)$, we can express the Lagrangian of graph G and its derivatives at $x \in \Delta$ with means of tensor theory :

$$L_G(x) = \frac{1}{k!} \langle \mathcal{A}(G), \underbrace{x \circ \cdots \circ x}_{\text{k times}} \rangle = \frac{1}{k!} \sum_{i_1,\ldots,i_k=1}^{|V|} \mathcal{A}(G)_e \prod_{j \in e} x_j = \frac{1}{k!} \mathcal{A}(G) x^k \tag{17}$$

$$\nabla L_G(x) = \frac{1}{(k-1)!} \left(\sum_{i_1,\ldots,i_{k-1}=1}^{|V|} \mathcal{A}(G)_e \prod_{i \in e \setminus \{i_k\}} x_i \right)_{1 \leq i_k \leq |V|} = \frac{1}{(k-1)!} \mathcal{A}(G) x^{k-1} \tag{18}$$

$$\nabla^2 L_G(x) = \frac{1}{(k-2)!} \left(\sum_{i_1,\ldots,i_{k-2}=1}^{|V|} \mathcal{A}(G)_e \prod_{i \in e \setminus \{i_k, i_{k-1}\}} x_i \right)_{1 \leq i_k, i_{k-1} \leq |V|} = \frac{1}{(k-2)!} \mathcal{A}(G) x^{k-2} \tag{19}$$

where $\circ$ is the outer product.

4.3 Equivalence with a previously derived formulation for the case $k = 2$

Here we provide a short proof of the equivalence between the new formulation in the case $k = 2$ and formulation (4) derived in [5].

For $x \in \Delta$,

$$1 = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j = 2 \sum_{i<j} x_i x_j + \sum_{i=1}^{n} x_i^2 \tag{20}$$

$$= 2 \sum_{[i,j]\in E} x_i x_j + 2 \sum_{[i,j]\in \overline{E}} x_i x_j + \sum_{i=1}^{n} x_i^2 \tag{21}$$

$$= 2L_G(x) + 2L_{\overline{G}}(x) + \|x\|_2^2 \tag{22}$$

From where we obtain

$$L_{\overline{G}}(x) = \frac{1}{2} - L_G(x) - \frac{1}{2}\|x\|_2^2 \tag{23}$$

Thus

$$\min_{(x,y)\in\mathcal{P}_s} \left[L_{\overline{G}}(x) - L_{G(y)}(x) + \alpha\|x\|_2^2 - \beta\|y\|_2^2\right] \tag{24}$$

$$= \min_{(x,y)\in\mathcal{P}_s} \left[\frac{1}{2} - L_G(x) - L_{G(y)}(x) - \frac{1}{2}\|x\|_2^2 + \alpha\|x\|_2^2 - \beta\|y\|_2^2\right] \tag{25}$$

$$= \frac{1}{2} - \max_{(x,y)\in\mathcal{P}_s} \left[L_{G\cup G(y)}(x) + \frac{1}{2}\|x\|_2^2 - \alpha\|x\|_2^2 + \beta\|y\|_2^2\right] \tag{26}$$

So for $\alpha \in (0, \frac{1}{2})$ and $\beta > 0$, we recover the initial formulation.

4.4 Correctness of the formulation in general

Proposition 4.2 (Characterization of local maxima for h)**.** *The following are equivalent :*

(i) $p = (x, y) \in \mathcal{P}_s$ is a local minimizer for $h(x, y)$

(ii) p is a strict local minimizer

(iii) $p = (x^{(C)}, y^{(p)})$, where $s \geq l = \mathbf{1}^T y^{(p)} \in \mathbb{N}$, with C an l-defective clique in G which is also a maximal clique in $G \cup G(y^{(p)})$, and $y^{(p)} \in \{0, 1\}^{\overline{E}}$ such that $y_e^{(p)} = 1$ for every $e \in \binom{C}{k} \cap \overline{E}$ and with supp$(y^{(p)})$ of maximum cardinality under these constraints.

In either of these equivalent cases, we have

$$h(p) = \alpha|C|^{1-k} - \beta l \tag{27}$$

Proof. Let $p := (x^{(p)}, y^{(p)}) \in \mathcal{P}_s$.

$(ii) \Rightarrow (i)$: if p is a strict minimum, it is of course a minimum.

$(i) \Rightarrow (iii)$: If p is a minimizer of h, then in particular $y^{(p)}$ must be a minimizer of $h(x^{(p)}, \cdot)$ which is strictly concave (indeed $\frac{\partial^2 h}{\partial y_e \partial y_{e'}}(x, y) = -2\beta\mathbf{1}_{e=e'}$ so its hessian is definite negative), thus y must be a vertex of $[0, 1]^{\overline{E}}$ i.e. $y \in \{0, 1\}^{\overline{E}}$.

Then $x^{(p)}$ must also be a local minimizer for $h(\cdot, y^{(p)})$, which is (up to a constant) a regularized maximal clique relaxation for the augmented graph $G \cup G(y^{(p)})$.

The theorem recalled in preamble then gives us that the local minimizers of this function are $x = x^{(C)}$ with C a maximal clique in $G \cup G(y^{(p)})$. Since $G \cup G(y^{(p)})$ is defined by adding s fake edges to G, C is necessarily an s-defective clique in G.

$(iii) \Rightarrow (ii)$: For a fixed $p = (x^{(C)}, y^{(p)})$ with C, $y^{(p)}$ satisfying the conditions of point (iii).

Case 1 : $C = V$ i.e. the clique is the whole graph (possible if the graph misses less than s edges and the budget is enough to fully complete it). Thus necessarily $y^{(p)} = \mathbf{1}$ i.e. the vector with 1 everywhere and $x^{(C)} = x^{(V)}$ i.e. the characteristic vector for the whole graph. Denote $l = \mathbf{1}^T y^{(p)}$. Then

$$h(x^{(C)}, y^{(p)}) = \alpha \left(\frac{1}{|V|}\right)^{k-1} - l\beta \tag{28}$$

Assume there exists a minimizer $z' = (x', y') \neq (x, y)$. By (i) $\Rightarrow$ (iii), z' satisfies the conditions of (iii), and there is at least $a \geq 1$ missing edges in $G \cup G(y')$ and $b \geq 1$ missing nodes in $\mathrm{supp}(x')$ which is a clique. So

$$h(x', y') = \alpha \left(\frac{1}{|V| - b}\right)^{k-1} - (l - a)\beta \tag{29}$$

$$> \alpha \left(\frac{1}{|V|}\right)^{k-1} - l\beta \tag{30}$$

$$= h(x^{(C)}, y^{(p)}) \tag{31}$$

This is true for every minimizer, so in particular it is true for every element of $\mathcal{P}_s$, i.e. $p = (x^{(C)}, y^{(p)})$ is a strict minimizer.

Case 2 : $C \neq V$ i.e. at least one node is not in the clique. Let $\overline{C} := V \setminus C$, $S := \mathrm{supp}(y^{(p)})$, $\overline{S} := \overline{E} \setminus S$, $g := \nabla h(p)$ and $H := \nabla^2 h(p)$.

For every $i \in V$ we have

$$\begin{aligned} g_i = \frac{\partial h}{\partial x_i}(x, y) &= \sum_{e \in \overline{E}} \mathbf{1}_{i \in e} \prod_{j \in e \setminus \{i\}} x_j^{(C)} - \sum_{e \in E(y)} \mathbf{1}_{i \in e} y_e \prod_{j \in e \setminus \{i\}} x_j^{(C)} + \alpha k (x_i^{(C)})^{k-1} \\ &= \sum_{e \in \overline{E} \setminus E(y)} \mathbf{1}_{i \in e} \prod_{j \in e \setminus \{i\}} x_j^{(C)} + \alpha k (x_i^{(C)})^{k-1} \end{aligned} \tag{32}$$

In particular, for $i \in C$

$$g_i = \alpha k \left(\frac{1}{|C|}\right)^{k-1} \tag{33}$$

because the product is non-null only if for all x_j in the edge except x_i, $x_j = \frac{1}{|C|}$, i.e. only if all $j \in C$, so because $i \in C$ then the edge would be in the clique so not in $\overline{E} \setminus E(y)$.

For every $i \in \overline{C}$

$$g_i = \sum_{e \in \overline{E} \setminus E(y)} \mathbf{1}_{i \in e} \prod_{j \in e \setminus \{i\}} x_j^{(C)} \geq \left(\frac{1}{|C|}\right)^{k-1} > \alpha k \left(\frac{1}{|C|}\right)^{k-1} \tag{34}$$

because first by hypothesis there is at least one missing edge between $k - 1$ nodes of the clique and one missing node (or else the node could be added to the clique which would not be maximal), then $\alpha \leq \frac{1}{k(k-1)}$ with strict inequality for $k = 2$ so $\alpha k < 1$.

For $e \in \overline{E}$ we have

$$g_e = \frac{\partial h}{\partial y_e}(x, y) = -\prod_{i \in e} x_i^{(C)} - \beta k (y_e^{(p)})^{k-1} \tag{35}$$

and in particular $g_e = 0$ for $e \in \overline{S}$ (because then $y_e = 0$ and at least one $x_i = 0$ or else the edge would be in the clique and so in $E \cup S$), while for $e \in S$

$$g_e = -\left(\frac{1}{|C|}\right)^k - \beta k < 0 \tag{36}$$

Let d be a feasible direction from p so that $d := v - p$ with $v \in \mathcal{P}_s$. Let $\sigma_S := \sum_{e \in S} g_e$, $\sigma_C := \sum_{i \in C} v_i = 1 - \sum_{i \in \overline{C}} v_i \in [0,1]$, and $m_{\overline{C}} := \min_{i \in \overline{C}} g_i$.

$$\begin{aligned} g^T p &= \sum_{i \in C} x_i^{(C)} g_i + \sum_{e \in \overline{E}} y_e^{(p)} g_e + \sum_{i \in \overline{C}} x_i^{(C)} g_i && (37)\\ &= \frac{1}{|C|} \sum_{i \in C} g_i + \sum_{e \in S} g_e && (38)\\ &= \alpha k \left(\frac{1}{|C|}\right)^{k-1} + \sigma_S && (39) \end{aligned}$$

Denoting by g_V the part of the gradient dedicated to the vertices, we also have :

$$g_V^T v_V = g_C^T v_C + g_{\overline{C}}^T v_{\overline{C}} \tag{40}$$

First,

$$g_C^T v_C = \alpha k \left(\frac{1}{|C|}\right)^{k-1} \sum_{i \in C} v_i = \alpha k \left(\frac{1}{|C|}\right)^{k-1} \sigma_C = \alpha k \left(\frac{1}{|C|}\right)^{k-1} (1 - \sigma_{\overline{C}}) \tag{41}$$

Then,

$$g_{\overline{C}}^T v_{\overline{C}} \geq m_{\overline{C}} \sigma_{\overline{C}} \geq \alpha k \left(\frac{1}{|C|}\right)^{k-1} \sigma_{\overline{C}} \tag{42}$$

Thus,

$$g_V^T v_V \geq \alpha k \left(\frac{1}{|C|}\right)^{k-1} \tag{43}$$

We also have

$$g_{\overline{E}}^T v_{\overline{E}} = g_S^T v_S + g_{\overline{S}}^T v_{\overline{S}} = g_S^T v_S \geq \sigma_S \tag{44}$$

because $g_e < 0$ and $v_e \leq 1$ for every $e \in S$.

In the end,

$$g^T d = g_V^T v_V + g_{\overline{E}}^T v_{\overline{E}} - g^T p \geq 0 \tag{45}$$

We have equality iff there is equality in (42) and (44), thus iff $v = (x^{(v)}, y^{(v)})$ with $\operatorname{supp}(x^{(v)}) \subset C$ (as $\alpha k < 1$ it is impossible to have an equality if $\sigma_{\overline{C}} \neq 0$) and $y^{(v)} = y^{(p)}$ (we need $v_e = 1$ for each $e \in S$ and then all the budget has been used if $s \leq \overline{E}$ or there is simply nowhere to spend it in the contrary case). In particular p is a first order stationary point with

$$T^0_{\mathcal{P}_s}(p, g) = \left\{ d \in T_{\mathcal{P}_s}(p) | d = v - p, v_{\overline{C}} = 0, v_{\overline{E}} = p_{\overline{E}} \right\} = \left\{ d \in T_{\mathcal{P}_s}(p) | d_{\overline{C}} = d_{\overline{E}} = 0 \right\} \tag{46}$$

Now denote by H_C the submatrix of the hessian with indices in C. The objective function here is the same (up to a constant) as for [6], which gives us

$$H_C = \alpha k (k-1) \left(\frac{1}{|C|}\right)^{k-2} \mathbf{I} > 0 \tag{47}$$

This proves the claim since we have sufficient conditions for local minimality. $\square$

As a corollary, the global optimum of h is achieved on maximum s-defective cliques.

Corollary 4.3. *The global minimizers of h are all the points p of the form $p = (x^{(C^*)}, y^{(p)})$ where C^* is an s-defective clique of maximum cardinality, and $y^{(p)} \in \{0,1\}^{\overline{E}}$ such that $\mathbf{1}^T y^{(p)} = \min(s, |\overline{E}|)$.*

Proof. Let $p = (x^{(C^*)}, y^{(p)})$ be a local minimizer for h and denote $l = \mathbf{1}^T y^{(p)}$. Then its objective value is, by (27), $h(p) = \alpha|C^*|^{1-k} - \beta l$, which is globally minimized when $|C^*|$ and l are as large as possible under the constraints $l \leq s$ and $l \leq |\overline{E}|$ because $y^{(p)} \in \{0,1\}^{\overline{E}}$. □

5 A continuous characterization of maximal s-plexes in k-graphs

5.1 Specific definitions and notations

An s-plex is a subset S of vertices such that each vertex in S is adjacent to at least $|S| - s$ other vertices from S. An s-plex is said to be *maximal* if it is not contained in any other s-plex, while it is called *maximum* if it has maximum cardinality. Finding the biggest s-plex in a graph is known as the Maximum s-plex Problem, which is a combinatorial problem.

For a given k-graph $G = (V, E)$, denote by $A(\overline{G}) \in \{0,1\}^{|V|\times|\overline{E}|}$ the incidence matrix of its complement graph, defined by :

$$\forall e \in \overline{E},\ \forall i \in V,\ \ \left(A(\overline{G})\right)_{ie} = \begin{cases} 1 \text{ if } i \in e \\ 0 \text{ else} \end{cases} \tag{48}$$

i.e. each row corresponds to a vertex $i \in V$, and each column corresponds to an edge $e \in \overline{E}$, such that, for a given $y \in \{0,1\}^{|\overline{E}|}$ coding for the addition of fake edges, doing the product $A(\overline{G})y$ yields a vector of size $|V|$ giving us the number of fake edges each vertex belongs to.

Let $\mathcal{D}_s(G) := \{y \in [0,1]^{\overline{E}} | A(\overline{G})y \leq (s-1)\mathbf{1}\}$ and $\mathcal{P}_s := \Delta \times \mathcal{D}_s(G)$.

The rest of the section is dedicated to the derivation of a continuous formulation of the Maximum s-plex Problem.

5.2 Previous work

In [16], the authors provide an adaptation of their continuous unregularized formulation for the Maximum s-defective Clique Problem, in order to make it work for the Maximum s-plex Problem, which is simply a change in the search space :

$$\max_{\text{s.t. } (x,y)\in\Delta\times\mathcal{D}_s(G)} x^T(A + A(y))x \tag{49}$$

where $\Delta = \{x \in \mathbb{R}^n | x \geq 0 \text{ and } 1^T x = 1\}$ and $\mathcal{D}_s(G) = \{y \in \{0,1\}^{\overline{E}} | By \leq (s-1)\mathbf{1}\}$ where B is the incidence matrix of graph G (instead of $\mathcal{D}_s(G) = \{y \in \{0,1\}^{\overline{E}} | 1^T y \leq s\}$ in the s-defective clique case). Again, $y \in \{0,1\}^{\overline{E}}$ in the definition of $\mathcal{D}_s(G)$ can be relaxed to $y \in [0,1]^{\overline{E}}$ in order to obtain a fully continuous formulation.

5.3 Formulation considered

The Maximum s-plex Problem admits the same continuous formulation as the Maximum s-defective Clique Problem, with just a difference in the search space.

Let $G := (V, E)$ be a k-graph with vertices V and edges E. Let $0 < \alpha \leq \frac{1}{k(k-1)}$ (with strict inequality for $k = 2$) and $\beta > 0$ and consider the following problem :

$$\min_{\text{s.t. } (x,y)\in\mathcal{P}_s} L_{\overline{G}}(x) - L_{G(y)}(x) + \alpha\|x\|_k^k - \beta\|y\|_2^2 \tag{50}$$

where $\mathcal{P}_s = \Delta \times \mathcal{D}_s(G)$ with $\mathcal{D}_s(G) = \{y \in [0,1]^{\overline{E}} | A(\overline{G})y \leq (s-1)\mathbf{1}\}$ (which still is a polytope) instead of $\{y \in [0,1]^{\overline{E}} | \mathbf{1}^T y \leq s\}$.

We claim that minimizers of this problem are attained at $p = (x^{(C)}, y^{(p)})$, where $s - 1 \geq l = \max(A(\overline{G})y) \in \mathbb{N}$, with C an l-plex in G which is also a maximal clique in $G \cup G(y^{(p)})$, and $y^{(p)} \in \{0,1\}^{\overline{E}}$ such that $y^{(p)}_e = 1$ for every $e \in \binom{C}{k} \cap \overline{E}$ and with $\mathrm{supp}(y^{(p)})$ of maximum cardinality under these constraints.

5.4 Correctness of the formulation

The reasoning for the Maximum s-defective Clique Problem can be almost exactly applied to this new problem with the exception of the application of the theorem given by [6] (Maximum Clique Problem in k-graphs) at the end of the proof of (i) $\Rightarrow$ (iii) in 4.2. As the result is applied to an augmented graph where $y^{(p)}$ cannot add more than $s-1$ edges to each node, the resulting maximal clique in this setting becomes a maximal l-plex instead of a maximal l-defective clique.

6 Results for variants of the Frank-Wolfe algorithm

6.1 Specific notations and definitions

In the following, we will denote $\tilde{h} := -h$ to stay in accordance with the usual conventions, which work on a maximization problem instead of a minimization problem like here.

Let $\Delta^{(C)} := \{x \in \Delta | x_i = 0 \text{ for all } i \in V \setminus C\}$ be the minimal face of Δ containing $x^{(C)}$ in its relative interior.

We define the face of a polytope $\mathcal{Q}$ exposed by a gradient $g \in \mathbb{R}^n$ as

$$\mathcal{F}_e(g) := \arg\max_{w \in \mathcal{Q}} g^T w \tag{51}$$

For every $x \in \Delta$, $i \in [1, \ldots, n]$, the multiplier function $\lambda_i : \Delta \to \mathbb{R}$ for a function f is defined as

$$\lambda_i(x) := \nabla f(x)^T (\mathbf{e}_i - x) \tag{52}$$

or in vector form

$$\lambda(x) := \nabla f(x) - x^T \nabla f(x) \mathbf{1} \tag{53}$$

We recall the FDFW algorithm, applied on a polytope $\mathcal{Q} := \mathrm{conv}(A) \subset \mathbb{R}^n$ with $|A| < +\infty$ with objective function f :

Algorithm 1 FDFW on a polytope

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

The FDFW at every iteration chooses between the classic FW direction $d_k^{\mathcal{FW}}$ and the in face direction $d_k^{\mathcal{FD}}$. 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 $w \to f(w_k) + \nabla f(w_k)^T w$ 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.

Remark 6.1. *For a more illustrated explanation, take $\mathcal{Q} = \Delta$. The problem can be seen as the attribution of a budget, and we have two possibilities to spend it better :*

- *We take some budget from the other vertices and give it to a profitable one (which corresponds to the classic FW direction $d_k^{\mathcal{FW}}$)*
- *Or we take the budget of a poor vertex and give it to the others (which corresponds to the in face direction $d_k^{\mathcal{FD}}$)*

When $f = \tilde{h}$ and $\mathcal{Q} = \mathcal{P}_s$, it is not difficult to see that the main cost to compute v_k is finding the smallest s components of a vector with size at most $|\overline{E}|$. After 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 proved in **Proposition 7.2** of [5], 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 :

$$\alpha_k \geq \overline{\alpha}_k := \min\left(\alpha_k^{\max}, c\frac{\nabla f(w_k)^T d_k}{\|d_k\|_2^2}\right) \tag{54}$$

Furthermore, for some convergence results we need the following sufficient increase condition (used in the Armijo / backtracking line search) for some constant $\rho > 0$:

$$f(w_k + \alpha_k d_k) - f(w_k) \geq \rho \overline{\alpha}_k \nabla f(w_k)^T d_k \tag{55}$$

As showed for the case $k = 2$ of the Maximum s-defective Clique Problem in [5], the convergence of the previous optimization scheme applied to our objective can be slow and inefficient. Since y is tied to x, it is not possible to efficiently change the regularization parameters to speed up convergence (either the algorithm ignores x if the penalty coefficient on the y variable is large or it ignores y if this coefficient is small). This motivated the authors to introduce a tailored FW variant, named FWdc, exploiting the cross product structure of the search space by splitting the update rules, which is recalled hereafter :

Algorithm 2 FWdc

Initialize : $z_0 = (x_0, y_0) \in \Delta \times \mathcal{D}_s(G), k = 0$
if z_k is stationary **then**
  STOP
end if
Compute x_{k+1} applying one step of the previous algorithm with $w_0 = x_k$ and $f(w) = \tilde{h}(w, y_k)$.
Let $y_{k+1} \in \arg\max_{y \in \mathcal{D}_s(G)} \nabla_y h_G(x_k, y_k)^T y$.
Set k = k+1. Go to step 2.

At every iteration the method alternates an FDFW step on the x variable with a full FW step on the y variable so that y_k always belongs to $\{0,1\}^{\overline{E}}$.

The rest of the section is dedicated to the generalization of the identification results for the FDFW and FWdc variants applied to the objective $\tilde{h}$ that were found in [5].

6.2 Results

For any of the two considered problems, thanks to the previous parts we can express any maximizer p of $\tilde{h}$ as $p = (x^{(C)}, y^{(p)})$ where $y^{(p)} \in \{0,1\}^{\overline{E}}$ such that $\mathbf{1}^T y^{(p)} \le s$ and C is a maximal clique of $G \cup G(y^{(p)})$.

We now prove that the face of $\mathcal{P}_s$ exposed by the gradient in p a maximizer is simply the product between Δ and the singleton $\{y^{(p)}\}$. This property, sometimes called strict complementarity, is of key importance to prove identification results for variants of the Frank-Wolfe algorithm (see [4], [3] and [9]), and the discussion of external regularity in **Section 5.3** of [2].

Lemma 6.2. *Let $p := (x^{(C)}, y^{(p)})$ a strict minimizer. Then the face exposed by $\nabla\tilde{h}(p)$ coincides with the minimal face $\mathcal{F}(p)$ of $\mathcal{P}_s$ containing p :*

$$\mathcal{F}_e(\nabla\tilde{h}(p)) = \mathcal{F}(p) = \Delta^{(C)} \times \left\{y^{(p)}\right\} \tag{56}$$

Proof. To start with, the second equality follows from the fact that $y^{(p)}$ is a vertex of $\mathcal{D}_s(G)$ and that $\Delta^{(C)}$ is the minimal face of Δ 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 and $\mathcal{D}_s(G)$ separately, $\lambda_a(p) < 0$ iff

$$\lambda_a^x(p) = \nabla_x\tilde{h}(p)^T(a_x - x^{(C)}) \le 0 \tag{57}$$

$$\lambda_a^y(p) = \nabla_y\tilde{h}(p)^T(a_y - y^{(p)}) \le 0 \tag{58}$$

and at least one of these relations must be strict. Since a is a vertex of $\mathcal{P}_s$, $a_x = \mathbf{e}_l$ with $l \in [1, \dots, n]$ and $\{0,1\}^{\overline{E}}$, while $a \notin \mathcal{F}(p)$ implies $l \notin C$ or $a_y \neq y^{(p)}$. We have

$$\nabla_x\tilde{h}(p)^T x^{(C)} = \sum_{j=1}^{n} x_j^{(C)} \frac{\partial\tilde{h}}{\partial x_j}(p) = -\sum_{j \in C} x_j^{(C)} \left[\sum_{e \in \overline{E} \setminus E(y)} \mathbf{1}_{j \in e} \prod_{i \in e \setminus \{j\}} x_i^{(C)} + \alpha k \left(x_j^{(C)}\right)^{k-1}\right] \tag{59}$$

$$= -\sum_{j \in C} \left[\sum_{e \in \overline{E} \setminus E(y)} \mathbf{1}_{j \in e} \prod_{i \in e} x_i^{(C)} + \alpha k \left(x_j^{(C)}\right)^{k}\right] \tag{60}$$

$$= -\alpha k \|x^{(C)}\|_k^k \tag{61}$$

and

$$\nabla_x\tilde{h}(p)^T a_x = \frac{\partial\tilde{h}}{\partial x_l}(p) = -\sum_{e \in \overline{E} \setminus E(y)} \mathbf{1}_{l \in e} \prod_{i \in e \setminus \{l\}} x_i^{(C)} - \alpha k \left(x_l^{(C)}\right)^{k-1} = -\alpha k \left(x_l^{(C)}\right)^{k-1} \tag{62}$$

because the clique is maximal. Combining the two, we obtain

$$\lambda_a^x(p) = \nabla_x \tilde{h}(p)^T (a_x - x^{(C)}) = \alpha k \left(x_l^{(C)} \right)^{k-1} - \alpha k \|x^{(C)}\|_k^k \tag{63}$$

which is equal to 0 if $l \in C$ and is strictly negative otherwise. This proves that (57) holds with strict inequality if $l \notin C$ or else with equality if $l \in C$.

In a similar vein we proceed with (58). If $a_y = y^{(p)}$, then (58) holds with equality, but then $l \in V \setminus C$, and we are done. So assume $a_y \neq y^{(p)}$, and consider the supports $S_a = \left\{ e \in \overline{E} | (a_y)_e = 1 \right\}$ and $S_y = \left\{ e \in \overline{E} | y_e^{(p)} = 1 \right\}$.

Necessarily, $S_y \setminus S_a \neq \emptyset$. Indeed, whether we are in the case of the Maximum s-defective Clique Problem or in the case of the Maximum s-plex Problem, both a_y and $y^{(p)}$ are elements of the same polytope i.e. they satisfy the same constraints, and, because p is a minimizer, $y^{(p)}$ has a support of maximal cardinality under these constraints, so if $S_y \subset S_a$ then $|S_y| = |S_a|$, but because $a_y \neq y^{(p)}$ then $S_y \not\subset S_a$ i.e. $S_y \setminus S_a \neq \emptyset$.

Then, for every e in S_y we have

$$\frac{\partial \tilde{h}}{\partial y_e}(p) = \prod_{i \in e} x_i^{(C)} + 2\beta y_e^{(p)} \geq 2\beta > 0 \tag{64}$$

while for every e in $S_a \setminus S_y$ we have

$$\frac{\partial \tilde{h}}{\partial y_e}(p) = 0 \tag{65}$$

because $y_e^{(p)} = 0$ by definition of S_y and $e \in \overline{E} \setminus E(y^{(p)})$. Then

$$\begin{aligned}
\lambda_a^y(p) = \nabla_y \tilde{h}(p)^T (a_y - y^{(p)}) &= \sum_{e \in S_a} \frac{\partial \tilde{h}}{\partial y_e}(p) - \sum_{e \in S_y} \frac{\partial \tilde{h}}{\partial y_e}(p) && (66) \\
&= \sum_{e \in S_a \setminus S_y} \frac{\partial \tilde{h}}{\partial y_e}(p) - \sum_{e \in S_y \setminus S_a} \frac{\partial \tilde{h}}{\partial y_e}(p) && (67) \\
&= - \sum_{e \in S_y \setminus S_a} \frac{\partial \tilde{h}}{\partial y_e}(p) < 0 && (68)
\end{aligned}$$

which proves that for any minimizer p, (57) and (58) both holds with one of the two strictly. $\square$

This result allows us to prove the following local convergence and identification result for the FDFW applied to our maximal s-defective clique formulation.

Theorem 6.3 (FDFW local identification and convergence)**.** *Let $p := (x^{(C)}, y^{(p)})$ be a strict minimizer, let z_k be a sequence generated by the FDFW. Then under (54) there exists a neighborhood $U(p)$ of p such that if $K := \min\{k \in \mathbb{N} | z_k \in U(p)\}$ we have the following properties :*

(a) if $\tilde{h}(z_k)$ is monotonically increasing, then supp$(z_k) = C$ and $y_k = y^{(p)}$ for every $k \geq K + dim\ \mathcal{F}(w_k)$

(b) if (55) also holds, then $z_k \to p$.

Proof. $\tilde{h}(\cdot, y^{(p)})$ is strongly concave in $\Delta^{(C)}$ (see (47) bearing in mind $\tilde{h} = -h$), and by 6.2, we have that $\mathcal{F}_e(\nabla \tilde{h}(p)) = \mathcal{F}(p) = \Delta^{(C)} \times \left\{ y^{(p)} \right\}$. We have all the necessary assumptions to apply **Lemma 7.4** of [5], which we recall hereafter :

> ***Lemma :*** *Let p be a local maximizer for f restricted to $\mathcal{Q}$. Assume that (54) holds and that f is strongly concave 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 $k \in [0, \ldots, dim(\mathcal{F}(w_0))]$ such that $w_{k+i} \in \mathcal{F}_e(\nabla f(p))$ for every $i \geq 0$*
>
> *(b) if in addition (55) holds, then $\{w_{k+i}\}_{i \geq 0}$ converges to p.*

□

As a corollary, we have the following global convergence result under the mild hypothesis that the set of limit points contains no saddle points.

Corollary 6.4 (FDFW global convergence)**.** *Let $\{z_k\}$ be a sequence generated by the FDFW, and assume that there are no saddle points in the limit set of $\{z_k\}$. Then under the conditions (54) and (55) on the step size we have $z_k \to p := (x^{(C)}, y^{(p)})$ with p a strict minimizer such that $supp(x_k) \subset C$ and $y_k = y^{(p)}$ for k large enough.*

Proof. As for for 6.3, we have all the necessary assumptions to apply **Corollary 7.5** of [5], which is recalled hereafter :

> ***Corollary :*** *Let $\{w_k\}$ be a sequence generated by the FDFW algorithm. Assume that there are no saddle points in the limit set of $\{w_k\}$ and that for every local maximizer p the objective f is strongly concave in $\mathcal{F}_e(\nabla f(p))$. Then under the conditions (54) and (55) on the stepsize, we have $w_k \to p$ with p a local maximizer satisfying $w_k \in \mathcal{F}_e(\nabla f(p))$ for k large enought.*

□

As recalled in the introduction of this section, the FDFW is not adapted to the structure of problem as the two variables are tied one to the other, which motivated the authors of [5] to design a tailored variant of the Frank-Wolfe algorithm where the updates rules are separated.

The next proposition gives us that, in the FWdc variant, $\{y_k\}$ is ultimately constant, which allows us to obtain convergence results by applying the general properties of the FDFW proved before to the x component.

Proposition 6.5. *In the FWdc variant, if $\tilde{h}(z_k)$ is increasing at each separate update of y_k and x_k, then $\{y_k\}$ can change at most $l + \frac{|\overline{E}|+\alpha\left(1-|C^*|^{1-k}\right)}{\beta}$ times, with C^* a maximum l-defective clique if we consider the Maximum s-defective Clique Problem and a maximum l-plex if we consider the Maximum s-plex Problem.*

Proof. Assume that for this step of the algorithm, y_k and y_{k+1} are distinct vertices of $\mathcal{D}_s$. Then

$$\begin{aligned}
\tilde{h}(z_{k+1}) - \tilde{h}(z_k) &= \tilde{h}(x_{k+1}, y_{k+1}) - \tilde{h}(x_k, y_k) && (69)\\
&= \tilde{h}(x_{k+1}, y_{k+1}) - \tilde{h}(x_{k+1}, y_k) + \tilde{h}(x_{k+1}, y_k) - \tilde{h}(x_k, y_k) && (70)\\
&\geq \tilde{h}(x_{k+1}, y_{k+1}) - \tilde{h}(x_{k+1}, y_k) && (71)\\
&\geq \nabla_y \tilde{h}(z_k)^T (y_{k+1} - y_k) + \beta \|y_{k+1} - y_k\|_2^2 && (72)\\
&\geq \beta > 0 && (73)
\end{aligned}$$

where we used the hypothesis that $\tilde{h}(z_k)$ is increasing at each separate update of x_k and y_k in (71), the 2β-strong convexity of $\tilde{h}(x, \cdot)$ in (72), and $y_{k+1} \in \arg\max_{y \in \mathcal{D}'_s} \nabla_y \tilde{h}(z_k)^T y$ and the fact that the distance between vertices of $\mathcal{D}_s$ is at least 1 in the last line.

By summing inequality (73), we get

$$\tilde{h}(z_N) - \tilde{h}(z_0) = \sum_{i=0}^{N-1} \left[\tilde{h}(z_{k+1}) - \tilde{h}(z_k)\right] \geq changes(N)\beta \tag{74}$$

where $changes(N)$ is the number of times y_k changes between index 0 and index N. But using $-\tilde{h}(z_0) = h(z_0) \leq |\overline{E}| + \alpha$ and 4.3, we have that for all $N \in \mathbb{N}$

$$\tilde{h}(z_N) - \tilde{h}(z_0) \leq \max_{z \in \mathcal{P}_s} \tilde{h}(z) - \tilde{h}(z_0) \leq \max_{z \in \mathcal{P}_s} \tilde{h}(z) + |\overline{E}| + \alpha \leq \beta l + |\overline{E}| + \alpha \left(1 - |C^*|^{1-k}\right) \tag{75}$$

Thus $\{changes(N)\}_N$ is bounded, so because it is an increasing integer sequence it necessary becomes stationary after a certain rank. By the previous inequality, we can now bound the number of times y_k can change by

$$changes(+\infty) \leq l + \frac{|\overline{E}| + \alpha\left(1 - |C^*|^{1-k}\right)}{\beta} \tag{76}$$

□

The following theorem allows to explicitly bound how close the sequence $\{x_k\}$ generated by the FWdc must be to $x^{(C)}$ for the identification to happen.

Proposition 6.6. *Let $\{z_k\}$ be a sequence generated by the FWdc, $\overline{y} \in \{0,1\}^{\overline{E}} \cap \mathcal{D}_s$ (where $\mathcal{D}_s$ is defined in the sense of one of the two considered problems), and C be a clique in $\overline{G \cup G(\overline{y})}$; let $\delta_{\max}$ be the maximum eigenvalue of the adjacency tensor $\mathcal{A}(\overline{G \cup G(\overline{y})})$.*

For a node $j \in V \setminus C$, we define

$$E^C(j) := \left|\left\{e \in \overline{E} \setminus E(\overline{y}) | j \in e,\, e \in \{j\} \times \binom{C}{k-1}\right\}\right| \tag{77}$$

the number of missing edges containing node j and $k-1$ nodes of the clique.

We also define

$$m(C, G \cup G(\overline{y})) := \min_{v \in V \setminus C} E^C(v) \tag{78}$$

the minimum number of edges needed to increase by 1 the size of the clique, and

$$m_\alpha(C, G \cup G(\overline{y})) := m(C, G \cup G(\overline{y})) - \alpha k \tag{79}$$

Let K be a fixed index in $\mathbb{N}$ and I^c be the components of $supp(x_K)$ with index not in C, and let $L := \frac{1}{(k-2)!}\delta_{\max} + k(k-1)\alpha$. Assume that $y_{K+j} = \overline{y}$ is constant for $0 \leq j \leq |I^c|$, that (54) hold for $c := \frac{1}{L}$, and that

$$\|x_K - x^{(C)}\|_1 \leq \frac{m_\alpha(C, G \cup G(\overline{y}))}{m_\alpha(C, G \cup G(\overline{y})) + 2|C|^{k-1}\left(\frac{1}{(k-2)!}\delta_{\max} + k(k-1)\alpha\right)} \tag{80}$$

Then $supp(x_{K+|I^c|}) = C$.

Proof. Since y_k does not change for $k \in [K, K + |I^c|]$, the FWdc corresponds to an application of the AFW to the simplex Δ on the variable x. Let

$$\lambda_{\min} := \min_{i \in V \setminus C} -\lambda_i(x^{(C)}) \tag{81}$$

be the smallest negative multiplier with corresponding index not in C. Let L' be a Lipschitz constant for $\nabla_x h(x, y)$ with respect to the variable x. By Theorem 3.3 of [4], if

$$\|x_k - x^{(C)}\|_1 < \frac{\lambda_{\min}}{\lambda_{\min} + 2L'} \tag{82}$$

we have the desired identification result.

We now have to prove that we can take $L' = L$. In the following, we will abbreviate $L_{\overline{G \cup G(\overline{y})}} := L$ and $\mathcal{A}(\overline{G \cup G(\overline{y})}) := \mathcal{A}$.

$$\|\nabla_x \tilde{h}(x', \overline{y}) - \nabla_x \tilde{h}(x, \overline{y})\|_2 = \|\nabla L(x') - \nabla L(x) + k\alpha(x'^{[k-1]} - x^{[k-1]})\|_2 \tag{83}$$

$$\leq \|\nabla L(x') - \nabla L(x)\|_2 + k\alpha\|(x'^{[k-1]} - x^{[k-1]}\|_2 \tag{84}$$

Then by the mean value theorem we have that $\|x'^{[k-1]} - x^{[k-1]}\|_2 \leq (k-1)\|x' - x\|_2$.

About the first term, denoting $\lambda_{\max}(u)$ the maximum eigenvalue of the Hessian matrix at a point u, we have :

$$\begin{aligned}
\lambda_{\max}(u) &= \max_{\|v\|_2=1} |v^T \nabla^2 L(u) v| && (85)\\
&= \frac{1}{(k-2)!} \max_{\|v\|_2=1} |\mathcal{A} u^{k-2} v^2| && (86)\\
&\leq \frac{1}{(k-2)!} \max_{\|v\|_2=1} \frac{|\mathcal{A} u^{k-2} v^2|}{\|u\|_2^{k-2}} \text{ because } \|u\|_2 \leq \|u\|_1 \leq 1 && (87)\\
&\leq \frac{1}{(k-2)!} \max_{\|v_i\|_2=1,\ 1\leq i\leq k} |\mathcal{A} v_1 \dots v_k| && (88)\\
&= \frac{1}{(k-2)!} \max_{\|v\|_2=1} |\mathcal{A} v^k| = \frac{1}{(k-2)!}\delta_{\max} && (89)
\end{aligned}$$

where we used **(1.2)** and **(1.3)** of [10] for the final equality.

This means that all eigenvalues of the Hessian are majored by $\frac{1}{(k-2)!}\delta_{\max}$, thus

$$\|\nabla L(x') - \nabla L(x)\|_2 \leq \frac{1}{(k-2)!}\delta_{\max}\|x' - x\|_2 \tag{90}$$

Combining the two Lipschitz constants found, we get

$$\|\nabla_x \tilde{h}(x', \overline{y}) - \nabla_x \tilde{h}(x, \overline{y})\|_2 \leq \left(\frac{1}{(k-2)!}\delta_{\max} + k(k-1)\alpha\right)\|x' - x\|_2 \tag{91}$$

About the multipliers, for $j \in V \setminus C$ we have the equality

$$\begin{aligned}
-\lambda_j(x^{(C)}) &= \nabla_x \tilde{h}(x^{(C)}, \overline{y})^T (x^{(C)} - \mathbf{e}_j) && (92)\\
&= \sum_{e \in \overline{E} \setminus E(\overline{y})} \mathbf{1}_{j\in e} \prod_{i \in e\setminus\{j\}} x_i^{(C)} - \frac{1}{|C|}\sum_{l\in C}\left[\sum_{e \in \overline{E} \setminus E(\overline{y})} \mathbf{1}_{l\in e} \prod_{i \in e\setminus\{l\}} x_i^{(C)} + \alpha k \left(\frac{1}{|C|}\right)^{k-1}\right] && (93)\\
&= \frac{|E^C(j)| - \alpha k}{|C|^{k-1}} && (94)
\end{aligned}$$

We can now find $\lambda_{\min}$

$$\begin{aligned}
\lambda_{\min} = \min_{\in V\setminus C} -\lambda_i(x^{(C)}) = \min_{i\in V\setminus C} \frac{|E^C(i)| - \alpha k}{|C|^{k-1}} && (95)\\
= \frac{m(C, G \cup G(\overline{y})) - \alpha k}{|C|^{k-1}} && (96)\\
= \frac{m_\alpha(C, G \cup G(\overline{y}))}{|C|^{k-1}} && (97)
\end{aligned}$$

Finally we have

$$\frac{\lambda_{\min}}{\lambda_{\min} + 2L} = \frac{m_\alpha(C, G \cup G(\overline{y}))}{m_\alpha(C, G \cup G(\overline{y})) + 2|C|^{k-1}\left(\frac{1}{(k-2)!}\delta_{\max} + k(k-1)\alpha\right)} \tag{98}$$

□

Remark 6.7. *This result is very similar to what has been found in [5] for the case $k = 2$: we have $L := \frac{1}{(k-2)!}\delta_{\max} + k(k-1)\alpha$ where $0 < \alpha \leq \frac{1}{k(k-1)}$ with strict inequality for $k = 2$ instead of $L := 2\delta + \alpha$ where $0 < \alpha < 2$ and*

$$\frac{m_\alpha(C, G \cup G(\overline{y}))}{m_\alpha(C, G \cup G(\overline{y})) + 2|C|^{k-1}\left(\frac{1}{(k-2)!}\delta_{\max} + k(k-1)\alpha\right)} \tag{99}$$

with $m_\alpha(C, G \cup G(\overline{y})) := m(C, G \cup G(\overline{y})) - \alpha k$ instead of

$$\frac{m_\alpha(C, G \cup G(\overline{y}))}{m_\alpha(C, G \cup G(\overline{y})) + 2|C|\delta_{\max} + |C|\alpha} \tag{100}$$

with $m_\alpha(C, G \cup G(\overline{y})) := m(C, G \cup G(\overline{y})) - 1 + \frac{\alpha}{2}$.

References

[1] Immanuel M. Bomze. "Evolution towards the Maximum Clique". In: *Journal of Global Optimization* 10 (1997), pp. 143–164. URL: https://api.semanticscholar.org/CorpusID:33853863.

[2] Immanuel M. Bomze. "Regularity versus Degeneracy in Dynamics, Games, and Optimization: A Unified Approach to Different Aspects". In: *SIAM Review* 44.3 (2002), pp. 394–414. ISSN: 00361445. URL: http://www.jstor.org/stable/4148380 (visited on 08/23/2023).

[3] Immanuel M. Bomze, Francesco Rinaldi, and Samuel Rota Bulò. "First-order Methods for the Impatient: Support Identification in Finite Time with Convergent Frank–Wolfe Variants". In: *SIAM Journal on Optimization* 29.3 (2019), pp. 2211–2226. DOI: 10.1137/18M1206953. eprint: https://doi.org/10.1137/18M1206953. URL: https://doi.org/10.1137/18M1206953.

[4] Immanuel M. Bomze, Francesco Rinaldi, and Damiano Zeffiro. *Active set complexity of the Away-step Frank-Wolfe Algorithm*. 2019. arXiv: 1912.11492 [math.OC].

[5] Immanuel M. Bomze, Francesco Rinaldi, and Damiano Zeffiro. *Fast cluster detection in networks by first-order optimization*. 2021. arXiv: 2103.15907 [math.OC].

[6] Samuel Rota Bulò and Marcello Pelillo. "A generalization of the Motzkin–Straus theorem to hypergraphs". In: *Optimization Letters* 3 (2009), pp. 287–295. DOI: 10.1007/s11590-008-0108-3.

[7] Philip S. Chodrow, Nate Veldt, and Austin R. Benson. *Generative hypergraph clustering: from blockmodels to modularity*. 2021. arXiv: 2101.09611 [cs.SI].

[8] Marguerite Frank and Philip Wolfe. "An algorithm for quadratic programming". In: *Naval Research Logistics Quarterly* 3 (1956), pp. 95–110. URL: https://api.semanticscholar.org/CorpusID:122654717.

[9] Dan Garber. *Revisiting Frank-Wolfe for Polytopes: Strict Complementarity and Sparsity*. 2020. arXiv: 2006.00558 [math.OC].

[10] Chang Liang and Yuning Yang. "A note on Banach's results concerning homogeneous polynomials associated with nonnegative tensors". In: 15 (Mar. 2021), pp. 419–429. DOI: 10.1007/s11590-020-01602.

[11] Duncan Luce and Albert D. Perry. "A method of matrix analysis of group structure". In: *Psychometrika* 14 (1949), pp. 95–116. URL: https://api.semanticscholar.org/CorpusID:16186758.

[12] Panos Pardalos and Jue Xue. "The Maximum Clique Problem". In: *Journal of Global Optimization* 4 (Apr. 1994), pp. 301–328. DOI: 10.1007/BF01098364.

[13] Bharath Pattabiraman et al. "Fast Algorithms for the Maximum Clique Problem on Massive Graphs with Applications to Overlapping Community Detection". In: *Internet Mathematics* 11.4-5 (May 2015), pp. 421–448. DOI: 10.1080/15427951.2014.986778. URL: https://doi.org/10.1080%2F15427951.2014.986778.

[14] Jeffrey Pattillo, Nataly Youssef, and Sergiy Butenko. "On clique relaxation models in network analysis". In: *European Journal of Operational Research* 226.1 (2013), pp. 9–18. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2012.10.021. URL: https://www.sciencedirect.com/science/article/pii/S0377221712007679.

[15] Nicolò Ruggeri et al. *Community Detection in Large Hypergraphs*. 2023. arXiv: 2301.11226 [cs.SI].

[16] Vladimir Stozhkov et al. *Continuous cubic formulations for cluster detection problems in networks*. 2022. DOI: 10.1007/s10107-020-01572-4.

[17] Svyatoslav Trukhanov et al. "Algorithms for Detecting Optimal Hereditary Structures in Graphs, with Application to Clique Relaxations". In: *Comput. Optim. Appl.* 56.1 (Sept. 2013), pp. 113–130. ISSN: 0926-6003. DOI: 10.1007/s10589-013-9548-5. URL: https://doi.org/10.1007/s10589-013-9548-5.

[18] Sara Venturini et al. *Laplacian-based Semi-Supervised Learning in Multilayer Hypergraphs by Coordinate Descent*. 2023. arXiv: 2301.12184 [cs.LG].

[19] Qinghua Wu and Jin-Kao Hao. "A review on algorithms for maximum clique problems". In: *European Journal of Operational Research* 242 (May 2015), pp. 693–709. DOI: 10.1016/j.ejor.2014.09.064.

[20] Haiyuan Yu et al. "Predicting interactions in protein networks by completing defective cliques". In: *Bioinformatics* 22.7 (Feb. 2006), pp. 823–829. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btl014. eprint: https://academic.oup.com/bioinformatics/article-pdf/22/7/823/48841534/bioinformatics_22_7_823.pdf. URL: https://doi.org/10.1093/bioinformatics/btl014.

[21] Yaoming Zhen and Junhui Wang. *Community Detection in General Hypergraph via Graph Embedding*. 2021. arXiv: 2103.15035 [stat.ML].