A Continuous Characterization of Maximal Cliques in k-uniform Hypergraphs

Samuel Rota Bulò, and Marcello Pelillo

Dipartimento di Informatica.
Università Ca' Foscari di Venezia.
{srotabul,pelillo}@dsi.unive.it

Abstract. In 1965 Motzkin and Straus established a remarkable connection between the local/global maximizers of the Lagrangian of a graph G over the standard simplex Δ and the maximal/maximum cliques of G. In this work we generalize the Motzkin-Straus theorem to k-uniform hypergraphs, establishing an isomorphism between local/global minimizers of a particular function over Δ and the maximal/maximum cliques of a k-uniform hypergraph. This theoretical result opens the door to a wide range of further both practical and theoretical applications, concerning continuous-based heuristics for the maximum clique problem on hypergraphs, as well as the discover of new bounds on the clique number of hypergraphs. Moreover we show how the continuous optimization task related to our theorem, can be easily locally solved by mean of a dynamical system.

1 Introduction

Many problems of practical interest are inherently intractable, in the sense that it is not possible to find fast (i.e., polynomial time) algorithms to solve them exactly, unless the classes P and NP coincide. The Maximum Clique Problem (MCP) is one of the most famous intractable combinatorial optimization problems, that asks for the largest complete subgraph of a given graph. This problem is even hard to approximate within a factor of $n/2^{(\log n)^{1-\epsilon}}$ for any $\epsilon > 0$ where n is the number of nodes in the graph [15]. Although this pessimistic state of affairs and because of its important applications in different fields such as computer vision, experimental design, information retrieval and fault tolerance, much attention has gone into developing efficient heuristics for the MCP, even if no formal guarantee of performance may be provided, but are nevertheless useful in practical applications. Moreover many important problems can be easily reduced to maximum clique problem e.g. boolean satisfiability problem, subgraph isomorphism problem, vertex cover problem etc.

Plenty of heuristics have been proposed over the last 50 years and we refer to [7] for a complete survey about complexity issues and applications of the MCP. In this introduction, we will focus our attention in particular on the continuous-based class of heuristics, since they are strongly related to the topics addressed in this paper. The heuristics of this class are mostly based on a result of Motzkin and Straus [18] that establishes a remarkable connection between the maximum clique problem and the extrema of the Lagrangian of a graph (1). In Section 2 we will see in deeper details the Motzkin-Straus theorem, but briefly it states, especially in its regularized version, an isomorphism between the set of maximal/maximum cliques of an undirected graph G and the set of local/global maximizers of the Lagrangian of G. This continuous formulation of the MCP suggests a fundamental new way of solving this problem, by allowing a shift from the discrete domain to the continuous one in an elegant manner. As pointed out in [21] continuous formulations of discrete problems are particularly attractive, because they not only allow us to exploit the full arsenal of continuous optimization techniques, thereby leading to the development of new algorithms, but may also reveal unexpected theoretical properties.

From an applicative point of view the Motzkin-Straus result led to the development of several MCP heuristics [6, 8, 13, 22, 24], but very interesting are also its theoretical implications. This result in fact was originally achieved by Motzkin and Straus to support an alternative proof of a slightly weaker version of the fundamental Turán theorem [27], moreover it was successfully used to achieve several bounds for the clique number of graphs [9, 28, 29]. The Motzkin-Straus theorem was also successfully generalized to vertex- weighted graphs [14] and edge-weighted graphs [23].

Recently the interest of researchers in many fields is focusing on hypergraphs, i.e. generalizations of graphs where edges are subsets of vertices, because of their greater expressiveness in representing higher-order relations. Just as graphs naturally represent many kinds of information in mathematical and computer science problems, hypergraphs also arise naturally in important practical problems [10, 20, 30]. Moreover, many theorems involving graphs, as for example the Ramsey's theorem or the Szemerédi lemma, also hold for hypergraphs, in particular for k-uniform hypergraphs (or more simply k-graphs), i.e. hypergraphs whose edges have all cardinality k. Nevertheless, all known intractable problems on graphs can be reformulated on hypergraphs and in particular the maximum clique problem.

Even if clique problems on hypergraphs are gaining increasing popularity in several scientific communities, a bridge from these discrete structures to the continuous domain is still missing. With our work we will fill up this gap, in the same way as the Motzkin-Straus theorem filled it up in the context of graphs. Hence the contribution of this paper is purely theoretical and basically consists in a generalization of the Motzkin-Straus theorem to k-uniform hypergraphs. However, as happened for the Motzkin-Straus theorem, our hope is to open the door to a wide range of further both practical and theoretical applications. First of all, we furnish a continuous characterization of maximal cliques in k-graphs, allowing the development of continuous-based heuristics for the maximum clique problem over hypergraphs based on it. Thereby, in Section 5 we provide a discrete dynamical system to elegantly find maximal cliques in k-graphs, that turns out to include the heuristic for MCP developed by Pelillo [24] on graphs as a special case (in fact graphs are 2-uniform hypergraphs). Moreover our theorem can be used to achieve new bounds for the clique number on k-graphs, a very popular problem in the extremal graph theory field, however we leave this topic as a future development of this work.

2 The Motzkin-Straus theorem

Let $G = (V, E)$ be an (undirected) graph, where $V = \{1, \ldots, n\}$ is the vertex set and $E \subseteq \binom{V}{2}$ is the edge set, with $\binom{V}{k}$ denoting the set of all k-element subsets of V. A *clique* of G is a subset of mutually adjacent vertices in V. A clique is called *maximal* if it is not contained in any other clique. A clique is called *maximum* if it has maximum cardinality. The maximum size of a clique in G is called the *clique number* of G and is denoted by $\omega(G)$.

Consider the following function $L_G : \Delta \mapsto \mathbb{R}$, sometimes called the *Lagrangian* of graph G

$$L_G(\boldsymbol{x}) = \sum_{\{i,j\} \in E} x_i x_j \tag{1}$$

where

$$\Delta = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq 0, \ \sum_{i=1}^{n} x_i = 1 \right\}$$

is the *standard simplex.*

In 1965, Motzkin and Straus [18] established a remarkable connection between the maxima of the Lagrangian of a graph and its clique number.

Theorem 1 (Motzkin-Straus). *Let G be a graph with clique number $\omega(G)$, and $\boldsymbol{x}^*$ a maximizer of L_G then*

$$L_G(\boldsymbol{x}^*) = \frac{1}{2}\left[1 - \frac{1}{\omega(G)}\right]$$

Additionally Motzkin and Straus proved that a subset of vertices S is a maximum clique of G if and only if its *characteristic vector* $\boldsymbol{x}^S$ is a global maximizer of L_G.[1] The characteristic vector of a set S is the vector in Δ defined as:

$$\boldsymbol{x}_i^S = \frac{1_{i \in S}}{|S|}$$

where $|S|$ indicates the cardinality of the set S and 1_P is an indicator function giving 1 if property P is satisfied and 0 otherwise. With $\sigma(\boldsymbol{x})$ we will denote the *support* of a vector $\boldsymbol{x} \in \Delta$, i.e. the set of positive components in $\boldsymbol{x}$. For example, the support of the characteristic vector of a set S is S.

[1] Actually, Motzkin and Straus provided just the “only if” part of this theorem, even if the converse direction is a direct consequence of their results [25].

Gibbons, Hearn, Pardalos and Ramana [14], and Pelillo and Jagota [25], extended the theorem of Motzkin and Straus, providing a characterization of maximal cliques in terms of local maximizers of L_G, however not all local maximizers were in the form of a characteristic vector. Finally Bomze et al. [6] introduced a regularizing term in the graph Lagrangian obtaining $L_G^\tau : \Delta \mapsto \mathbb{R}$ defined as

$$L_G^\tau(\boldsymbol{x}) = L_G(\boldsymbol{x}) + \tau \sum_{i \in V} x_i^2$$

and proved that all local maximizers of L_G^τ are strict and in one-to-one relation with the characteristic vector of the maximal cliques of G, provided that $0 < \tau < \frac{1}{2}$.

Theorem 2 (Bomze). *Let G be a graph and $0 < \tau < \frac{1}{2}$. A vector $\boldsymbol{x} \in \Delta$ is a local (global) maximizer of L_G^τ over Δ if and only if it is the characteristic vector of a maximal (maximum) clique of G.*

The Motzkin-Straus theorem was successfully extended also to vertex-weighted graphs by Gibbons et al. [14] and edge-weighted graphs by Pavan and Pelillo [23].

In this paper we provide a further generalization of the Motzkin-Straus as well as the Bomze theorems to k-uniform hypergraphs, but firstly, we will introduce hypergraphs and review another generalization of the Motzkin-Straus theorem due to Sós and Straus [26].

3 *k*-uniform hypergraphs

A *k-uniform hypergraph*, or simply a *k-graph*, is a pair $G = (V, E)$, where $V = \{1, \ldots, n\}$ 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 also called *graphs*. The *complement* of a k-graph G is given by $\bar{G} = (V, \bar{E})$ where $\bar{E} = \binom{V}{k} \setminus E$. A subset of vertices $C \subseteq V$ is called a *clique* if $\binom{C}{k} \subseteq E$. 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. The *clique number* of a k-graph G, denoted by $\omega(G)$, is defined as the cardinality of a maximum clique.

Given a k-graph G with n vertices, the *Lagrangian* of G is the following homogeneous multilinear polynomial in n variables:

$$L_G(\boldsymbol{x}) = \sum_{e \in E} \prod_{i \in e} x_i \ . \tag{2}$$

In this paper, we are interested in studying a generalization of the Motzkin-Straus Theorem to hypergraphs. As it turns out, L_G cannot be directly used to extend the Motkzin-Straus theorem to k-graphs. Frankl and Rödl in 1984 [12] proved that by taking a maximizer $\boldsymbol{x}^*$ of L_G with support as small as possible, the subhypergraph induced by S is a 2-cover, i.e. a hypergraph such that every pair of vertices is contained in some hyperedge. Since 2-covers in graphs are basically cliques, we could expect a possible generalization of the Motkzin-Straus theorem where the clique number is replaced by the size l of the maximum 2-cover in the hypergraph. However $\boldsymbol{x}^*$ is not necessarily in the form of a characteristic vector, and it is not in general possible to express l as a function of $L_G(\boldsymbol{x}^*)$. Nevertheless, this result was used by Mubay [19] to achieve a bound for $L_G(\boldsymbol{x}^*)$ in terms of l on k-graphs, obtaining

$$L_G(\boldsymbol{x}^*) \leq \binom{l}{k} l^{-k}$$

and he used it to provide an hypergraph extension of the Turán's theorem.

A further attempt to generalize the Motzkin-Straus theorem to hypergraphs is due to Sós and Straus [26]. They attach a nonnegative weight $x(H_l)$ to every complete l-subgraph H_l of a graph G, normalized by the condition

$$\sum_{H_l \subseteq G} x(H_l)^l = 1$$

and to every complete $(l+1)$-subgraph H_{l+1} of G they attach the weight

$$x(H_{l+1}) = \prod_{H_l \subset H_{l+1}} x(H_l)$$

and they define

$$f_G(x) = \sum_{H_{l+1} \subseteq G} x(H_{l+1}).$$

Then they get the following.

Theorem 3. $\max_x f_G(x) = \binom{k}{l+1} / \binom{k}{l}^{(l+1)/l}$*, where* k *is the order of a maximum clique* K *of* G*. This maximum is attained by attaching weights* $\binom{k}{l}^{-1/l}$ *to the* l*-subgraphs of* K *and weight* 0 *to all other complete* l*-subgraphs.*

Note that, the case $l = 1$ is exactly the Motzkin-Straus theorem.

Even if this result does not explicitly apply to hypergraphs, actually this theorem could be extended to k-graphs by attaching weights to subsets of hyperedges. However in order this theorem to work, the k-graph should satisfy a strong property that it to be a complete-subgraph graph of an ordinary graph (also said to be *conformal* [4]). This restricts the applicability of this theorem to a class of hypergraphs isomorphic to a subclass of 2-graphs having cliques of cardinality $\geq k$. We will see in the subsequent sections that our generalization applies to all k-uniform hypergraphs.

4 Characterization of maximal cliques on k-graphs

Given a k-graph G, consider the following non-linear program.

$$\begin{aligned} \text{minimize} \quad & h_{\bar{G}}(\boldsymbol{x}) = L_{\bar{G}}(\boldsymbol{x}) + \tau \sum_{i=1}^{n} x_i^k, \\ \text{subject to} \quad & \boldsymbol{x} \in \Delta \end{aligned} \qquad (3)$$

where $\tau \in \mathbb{R}$ and $L_{\bar{G}}$ is the Lagrangian of the complement of G. In order to simplify the notation we write h instead of $h_{\bar{G}}$ where the context is non ambiguous.

A *local solution* of problem (3) is a vector $\boldsymbol{x} \in \Delta$ for which there exists a local neighborhood $N_{\boldsymbol{x}}$ such that $h(\boldsymbol{y}) \geq h(\boldsymbol{x})$ for all $\boldsymbol{y} \in N_{\boldsymbol{x}}$, while a *global solution* is a vector $\boldsymbol{x} \in \Delta$ such that $h(\boldsymbol{y}) \geq h(\boldsymbol{x})$, for every $\boldsymbol{y} \in \Delta$. We say that $\boldsymbol{x}$ is a *strict* global/local solution if the inequalities are strict where $\boldsymbol{y} \neq \boldsymbol{x}$.

A necessary point for a vector $\boldsymbol{x}$ to be a local solution of our program is to satisfy the Karush-Kuhn-Tucker (KKT) conditions [16] for (3), i.e. there should exists $\lambda \in \mathbb{R}$ such that for all $j \in V$,

$$h_j(\boldsymbol{x}) \begin{cases} = \lambda & \text{if } j \in \sigma(\boldsymbol{x}) \\ \geq \lambda & \text{if } j \notin \sigma(\boldsymbol{x}) \end{cases}. \tag{4}$$

Here, $h_j(\boldsymbol{x})$ denotes the partial derivative of h with respect to x_j, i.e.

$$h_j(\boldsymbol{x}) = \sum_{e \in \bar{E}} 1_{j \in e} \prod_{i \in e \setminus \{j\}} x_i + \tau k x_j^{k-1},$$

and similarity $h_{j\ell}(\boldsymbol{x})$ will denote the partial derivative with respect to x_j and x_ℓ, i.e.

$$h_{j\ell}(\boldsymbol{x}) = 1_{j \neq \ell} \sum_{e \in \bar{E}} 1_{j,\ell \in e} \prod_{i \in e \setminus \{j,\ell\}} x_i + 1_{j=\ell} \tau k(k-1) x_j^{k-2}.$$

A sufficient condition for $\boldsymbol{x}$ to be a local solution of program (3) is to be a KKT point and to have the Hessian matrix of h in $\boldsymbol{x}$ positive definite on the subspace $M(\boldsymbol{x})$ defined as

$$M(\boldsymbol{x}) = \{\boldsymbol{\varepsilon} \in \mathbb{R}^n : \sum_{i=1}^{n} \varepsilon_i = 0, \varepsilon_j = 0 \text{ for all } j \text{ such that } h_j(\boldsymbol{x}) > \lambda\},$$

where the Hessian matrix of h in $\boldsymbol{x}$ is defined as

$$H(\boldsymbol{x}) = [h_{j\ell}(\boldsymbol{x})]_{j,\ell \in V}$$

In other words if $\boldsymbol{x}$ is a KKT point and for all $\boldsymbol{\varepsilon} \in M(\boldsymbol{x})$, $\boldsymbol{\varepsilon}' H(\boldsymbol{x}) \boldsymbol{\varepsilon} > 0$, then $\boldsymbol{x}$ is a local solution of (3).

Lemma 1. *Let G be a k-graph and let $\boldsymbol{x}$ be a local (global) solution of (3) with $\tau > 0$. If $C = \sigma(\boldsymbol{x})$ is a clique of G then it is a maximal (maximum) clique and $\boldsymbol{x}$ is the characteristic vector of C.*

Proof. Since $\boldsymbol{x}$ is a local solution of (3), it satisfies the KKT conditions (4). Therefore for all $j \in C$ we have that $\lambda = \tau k x_j^{k-1}$ and it follows that $\boldsymbol{x}$ is the characteristic vector of C. Moreover if there exists a larger clique D that contains C, then there exists a vertex $j \in D \setminus C$ such that $h_j(\boldsymbol{x}) = 0 < \lambda$. This contradicts conditions (4). Hence, C is a maximal clique of G.

Finally, $h(\boldsymbol{x}) = \tau |\sigma(\boldsymbol{x})|^{1-k}$ attains its global minimum when $\boldsymbol{x}$ is the characteristic vector of a maximum clique. ◻

Lemma 2. *Let G be a k-graph and $\boldsymbol{x}$ a local (global) solution of (3). If one of the following conditions holds*

1. $0 < \tau < \frac{1}{k(k-1)}$,
2. $\tau = \frac{1}{2}$, $k = 2$ *and* $\sigma(\boldsymbol{x})$ *is as small as possible,*
3. $\tau = \frac{1}{k(k-1)}$ *and* $k > 2$,

then $\boldsymbol{x}$ is the characteristic vector of a maximal (maximum) clique of G.

Proof. We claim that the support of $\boldsymbol{x}$ is a clique of G. Otherwise suppose that an edge $\tilde{e} \subseteq \sigma(\boldsymbol{x})$ is missing. Let $j, \ell \in \tilde{e}$ such that $x_j \leq x_\ell \leq \min_{i \in \tilde{e} \setminus \{j,\ell\}} x_i$ and take $\boldsymbol{y} = \boldsymbol{x} + \varepsilon(\boldsymbol{e}^j - \boldsymbol{e}^\ell) \in \Delta$, where $\boldsymbol{e}^j$ denotes a zero vector except for the j-th element set to 1 and where $0 < \varepsilon \leq x_\ell$.

We study the sign of $h(\boldsymbol{y}) - h(\boldsymbol{x})$ in a local neighborhood of $\boldsymbol{x}$ as $\varepsilon \to 0$ by means of the Taylor expansion of $h(\boldsymbol{x})$ truncated at the second-order term, where the first-order term cancels out since $\boldsymbol{x}$ satisfies (4), as it is a local solution of (3), and thereby $h_j(\boldsymbol{x}) = h_\ell(\boldsymbol{x})$:

$$
\begin{aligned}
&h(\boldsymbol{y}) - h(\boldsymbol{x}) = \\
&= \frac{\varepsilon^2}{2}\left[h_{jj}(\boldsymbol{x}) + h_{\ell\ell}(\boldsymbol{x}) - 2h_{j\ell}(\boldsymbol{x})\right] + \cdots = \\
&= \frac{\varepsilon^2}{2}\left[\tau k(k-1)\left(x_j^{k-2} + x_\ell^{k-2}\right) - 2\sum_{e \in \bar{E}} 1_{j,\ell \in e} \prod_{i \in e \setminus \{j,\ell\}} x_i\right] + \ldots
\end{aligned}
$$

Let $\mu = 2\sum_{e \in \bar{E}} 1_{j,\ell \in e} \prod_{i \in e \setminus \{j,\ell\}} x_i - (x_j^{k-2} + x_\ell^{k-2})$. Clearly $\mu \geq 0$. Then we can write

$$
\begin{aligned}
&h(\boldsymbol{y}) - h(\boldsymbol{x}) = \\
&= \frac{\varepsilon^2}{2}\left[\tau k(k-1)(x_j^{k-2} + x_\ell^{k-2}) - \mu - (x_j^{k-2} + x_\ell^{k-2})\right] + \cdots = \\
&= \frac{\varepsilon^2}{2}\left\{\left(x_j^{k-2} + x_\ell^{k-2}\right)\left[\tau k(k-1) - 1\right] - \mu\right\} + \ldots
\end{aligned}
\tag{5}
$$

Note from (5) that the second-order term is nonpositive and becomes zero only if $\tau = \frac{1}{k(k-1)}$ and $\mu = 0$. We proceed now by distinguishing 3 cases, each of which yields a contradiction, thereby proving that $\sigma(\boldsymbol{x})$ is a clique of G. This in conjunction with Lemma 1 concludes the proof.

Case 1: $0 < \tau < \frac{1}{k(k-1)}$ **or** $\mu > 0$**.**

In this case, $h(\boldsymbol{y}) - h(\boldsymbol{x})$ is strictly negative for sufficiently small values of ε, contradicting the local minimality of $\boldsymbol{x}$.

Case 2: $\tau = \frac{1}{k(k-1)}$**,** $k = 2$ **and** $\mu = 0$**.**

Here, from the hypothesis, we have that $\sigma(\boldsymbol{x})$ is as small as possible. For $k = 2$ we have trivially that $h(\boldsymbol{y}) - h(\boldsymbol{x}) = 0$, contradicting the minimality of the support size of $\boldsymbol{x}$.

Case 3: $\tau = \frac{1}{k(k-1)}$**,** $k > 2$ **and** $\mu = 0$**.**

Note that if $\mu = 0$, then $\tilde{e}$ is the only edge in $\bar{E}$ with vertices in $\sigma(\boldsymbol{x})$ that contains both j and ℓ. Moreover x_i is constant for all $i \in \tilde{e}$. It follows that we could arbitrarily have chosen j, ℓ in $\tilde{e}$ for the construction of $\boldsymbol{y}$. Hence, for every pair of vertices in $\tilde{e}$ there exists only one edge in $\bar{E}$ with vertices in $\sigma(\boldsymbol{x})$ containing them, namely $\tilde{e}$.

Let $m \in \arg\min_{i\in\tilde{e}\setminus\{j,\ell\}} x_i$ and take $\boldsymbol{z} = \boldsymbol{x} + \varepsilon\left[(\boldsymbol{e}^j + \boldsymbol{e}^\ell)/2 - \boldsymbol{e}^m\right] \in \Delta$ where $0 < \varepsilon \leq x_m$. We study the sign of $h(\boldsymbol{z}) - h(\boldsymbol{x})$ in a local neighborhood of $\boldsymbol{x}$ as $\varepsilon \to 0$ by means of the Taylor expansion of $h(\boldsymbol{x})$ truncated at the third-order term. Here again, the first-order term cancels out since $\boldsymbol{x}$ satisfies (4) and therefore we yield

$$\begin{aligned} &h(\boldsymbol{z}) - h(\boldsymbol{x}) = \\ &= \frac{\varepsilon^2}{2}\left[\frac{h_{jj}(\boldsymbol{x}) + h_{\ell\ell}(\boldsymbol{x})}{4} + h_{mm}(\boldsymbol{x}) - h_{jm}(\boldsymbol{x}) - h_{\ell m}(\boldsymbol{x}) + \frac{h_{j\ell}(\boldsymbol{x})}{2}\right] + \\ &+ \frac{\varepsilon^3}{6}\left[\frac{h_{jjj}(\boldsymbol{x}) + h_{\ell\ell\ell}(\boldsymbol{x})}{8} - h_{mmm}(\boldsymbol{x}) - \frac{3}{2}h_{j\ell m}(\boldsymbol{x})\right] + \dots \end{aligned} \tag{6}$$

where h_{uvw} denotes the partial derivative of h with respect to x_u, x_v and x_w, i.e.

$$h_{uvw}(\boldsymbol{x}) = 1_{u\neq v}1_{u\neq w}1_{v\neq w}\sum_{e\in\bar{E}} 1_{u,v,w\in e}\prod_{i\in e\setminus\{u,v,w\}} x_i + 1_{u=v=w}(k-2)x_u^{k-3}.$$

By the observation made at the beginning of this case and by setting $\xi = x_j$, it follows that $\forall u, v \in \tilde{e}.h_{uv}(\boldsymbol{x}) = \xi^{k-2}$, and $\forall u \in \tilde{e}.h_{uuu}(\boldsymbol{x}) = (k-2)\xi^{k-3}$ and finally $h_{j\ell m}(\boldsymbol{x}) = \xi^{k-3}$. Hence, the sign of $h(\boldsymbol{z}) - h(\boldsymbol{x})$ for sufficiently small values of ε is given by the sign of $-\frac{\varepsilon^3}{8}k\xi^{k-3}$ which is clearly negative and this contradicts the local minimality of $\boldsymbol{x}$. □

An interesting observation deriving in a straightforward manner from Lemma 2 is that all minimizers of (3) are strict provided that conditions (1) and (3) hold. The only case where "spurious" solutions could arise is when considering $k = 2$ and $\tau = \frac{1}{2}$ that is equivalent to maximize the Lagrangian of graphs over Δ.

The following theorem provides a generalization of the Motzkin-Straus Theorem (1) to k-graphs.

Theorem 4. *Let G be a k-graph with clique number $\omega(G)$. Then h attains its minimum over Δ at $\tau\,\omega(G)^{1-k}$ provided that $0 < \tau \leq \frac{1}{k(k-1)}$.*

Proof. Let $\boldsymbol{x}$ be a global solution of (3) with support as small as possible. Then by Lemma 2 we have that $\boldsymbol{x}$ is the characteristic vector of a maximum clique of G. It follows that $h(\boldsymbol{x}) = \tau|\sigma(\boldsymbol{x})|^{1-k} = \tau\omega(G)^{1-k}$. □

Note that this result is equivalent to the original Motzkin-Straus Theorem (1) for graphs, if we take $k = 2$ and $\tau = \frac{1}{2}$. In fact, in this case we obtain

$$L_G(\boldsymbol{x}) = \sum_{\{i,j\}\in E} x_i x_j = \frac{1}{2} - \sum_{\{i,j\}\in\bar{E}} x_i x_j - \frac{1}{2}\sum_{i=1}^n x_i^2 = \frac{1}{2} - h(\boldsymbol{x})$$

and it follows that

$$\max_{\boldsymbol{x}\in\Delta} L_G(\boldsymbol{x}) = \frac{1}{2} - \min_{\boldsymbol{x}\in\Delta} h(\boldsymbol{x}) = \frac{1}{2} - \frac{1}{2\omega(G)} = \frac{1}{2}\left[1 - \frac{1}{\omega(G)}\right].$$

The next lemma is instrumental to prove a generalization of the Bomze Theorem (2) to k-graphs.

Lemma 3. *Let G be a k-graph and $\boldsymbol{x}^C$ the characteristic vector of a maximal (maximum) clique C of G. Then $\boldsymbol{x}^C$ is a strict local (global) solution of (3) provided that $0 < \tau < \frac{1}{k}$.*

Proof. For simplicity let $\boldsymbol{x} = \boldsymbol{x}^C$. We will show that $\boldsymbol{x}$ is a strict local solution of (3) by proving that it satisfies the sufficient conditions introduced at the beginning of this section. First we prove that $\boldsymbol{x}$ satisfies (4) and then we show that $H(\boldsymbol{x})$ is positive definite on the subspace $M(\boldsymbol{x})$.

For all $j \in \sigma(\boldsymbol{x})$ we have $h_j(\boldsymbol{x}) = \tau k|C|^{1-k} = \lambda$, while for all $\ell \notin \sigma(\boldsymbol{x})$ we have $h_\ell(\boldsymbol{x}) \geq |C|^{1-k} > \lambda$, since $\sigma(\boldsymbol{x})$ is a maximal clique and therefore at least one edge joining ℓ and $k-1$ vertices in C is missing. Hence, $\boldsymbol{x}$ is a KKT point.

Moreover all eigenvalues of $H(\boldsymbol{x})|_{\sigma(\boldsymbol{x})}$, i.e. the Hessian in $\boldsymbol{x}$ restricted to the support of $\boldsymbol{x}$, are positive. In fact, $H(\boldsymbol{x})|_{\sigma(\boldsymbol{x})}$ is a diagonal matrix with positive diagonal entries

$$H(\boldsymbol{x})|_{\sigma(\boldsymbol{x})} = \tau k(k-1)|C|^{2-k} I$$

where I is the identity matrix. This implies that $H(\boldsymbol{x})$ is positive definite on the subspace $M(\boldsymbol{x})$.

Finally, $h(\boldsymbol{x}^C) = \tau|C|^{1-k}$ attains its global minimum where C is as large as possible, i.e. a maximum clique. □

Theorem 5. *Let G be a k-graph and $0 < \tau \leq \frac{1}{k(k-1)}$ (with strict inequality for $k = 2$). A vector $\boldsymbol{x} \in \Delta$ is a local (global) solution of (3) if and only if it is the characteristic vector of a maximal (maximum) clique of G.*

Proof. It follows from Lemmas 2 and 3. □

Note that if we take $k = 2$ and $0 < \tau < \frac{1}{2}$ then local (global) minimizers of h correspond to local (global) maximizers of $L_G^{\frac{1}{2}-\tau}$. In fact

$$h(\boldsymbol{x}) = \sum_{\{i,j\}\in\bar{E}} x_i x_j + \tau \sum_{i=1}^{n} x_i^2 = \frac{1}{2} - \sum_{\{i,j\}\in E} x_i x_j + \left(\tau - \frac{1}{2}\right) \sum_{i=1}^{n} x_i^2 =$$

$$= \frac{1}{2} - \left[\sum_{\{i,j\}\in E} x_i x_j + \left(\frac{1}{2} - \tau\right) \sum_{i=1}^{n} x_i^2\right] = \frac{1}{2} - L_G^{\frac{1}{2}-\tau}(\boldsymbol{x}).$$

Since $0 < \frac{1}{2} - \tau < \frac{1}{2}$, what we obtain is an equivalent formulation of the Bomze Theorem on graphs in terms of a minimization task.

5 Finding maximal cliques of k-graphs

Summarizing our results, we propose a generalization of a well-known theorem in the extremal graph theory field to k-graphs that turns out to provide a continuous characterization of a purely discrete problem, i.e. finding maximal cliques in k-graphs. More precisely, we implicitly provide an isomorphism between the set of maximal/maximum cliques of a k-graph G and the set of local/global minimizers of a particular function $h_{\bar{G}}$ over Δ, that permits to perform local optimization on $h_{\bar{G}}$ in order to extract, through the isomorphism, a maximal clique of the k-graph G. In this section we will see that the optimization of $h_{\bar{G}}$ may be easily carried out thanks to a theorem due to Baum and Eagon [1].

In the late 1960s, Baum and Eagon [1] introduced a class of nonlinear transformations in probability domain and proved a fundamental result which turns out to be very useful for the optimization task at hand. Their result generalizes an earlier one by Blakley [5] who discovered similar properties for certain homogeneous quadratic transformations. The next theorem introduces what is known as the Baum-Eagon inequality.

Theorem 6 (Baum-Eagon). *Let $P(\boldsymbol{x})$ be a homogeneous polynomial in the variables x_i with nonnegative coefficients, and let $\boldsymbol{x} \in \Delta$. Define the mapping $\boldsymbol{z} = \mathcal{M}(\boldsymbol{x})$ as follows:*

$$z_i = x_i \frac{\partial P(\boldsymbol{x})}{\partial x_i} \Big/ \sum_{j=1}^{n} x_j \frac{\partial P(\boldsymbol{x})}{\partial x_j}, \qquad i = 1, \ldots, n. \tag{7}$$

Then $P(\mathcal{M}(\boldsymbol{x})) > P(\boldsymbol{x})$, unless $\mathcal{M}(\boldsymbol{x}) = \boldsymbol{x}$. In other words $\mathcal{M}$ is a growth transformation for the polynomial P.

This result applies to homogeneous polynomials, however in a subsequent paper, Baum and Sell [3] proved that Theorem 6 still holds in the case of arbitrary polynomials with nonnegative coefficients, and further extended the result by proving that $\mathcal{M}$ increases P homotopically, which means that

$$P(\eta\mathcal{M}(\boldsymbol{x}) + (1-\eta)\boldsymbol{x}) \geq P(\boldsymbol{x}), \qquad 0 \leq \eta \leq 1$$

with equality if and only if $\mathcal{M}(\boldsymbol{x}) = \boldsymbol{x}$.

The Baum-Eagon inequality provides an effective iterative means for maximizing polynomial functions in probability domains, and in fact it has served as the basis for various statistical estimation techniques developed within the theory of probabilistic functions of Markov chains [2]. As noted in [3], the mapping $\mathcal{M}$ defined in Theorem 6 makes use of the first derivative only and yet is able to take finite steps while increasing P. This contrasts sharply with classical gradient methods, for which an increase in the objective function is guaranteed only when infinitesimal steps are taken, and determining the optimal step size entails computing higher-order derivatives. Additionally, performing gradient ascent in Δ requires some projection operator to ensure that the constraints not be violated, and this causes some problems for points lying on the boundary [11, 17]. In (7), instead, a computationally simple vector normalization is required.

It is worth noting that not all stationary points of the mapping $\mathcal{M}$ correspond to local maxima of the polynomial P; consider the vertices of Δ as an example. However all local maxima are the only stationary states that are stable, or even asymptotically stable if they are strict. Therefore if the dynamics gets trapped in non optimal stationary states, it suffices a small perturbation to get rid of the problem. We will see an example of this fact in Section 6.

Moving a step back to our function $h_{\bar{G}}$, it satisfies the hypothesis of Theorem 6 since it is a homogeneous polynomial of degree k with nonnegative coefficients in the variables x_i with $\boldsymbol{x} \in \Delta$. However our targets are not local maxima but local minima. Fortunately, it turns out that we can transform our minimization problem into an equivalent maximization one, by keeping the conditions of Theorem 6 still satisfied.

Note that all coefficients of $h_{\bar{G}}(\boldsymbol{x})$ are positive and upper bounded by 1. Furthermore, let $\xi = \max\left[\tau, \frac{1}{k!}\right]$ and note that ξ can be expressed as a complete homogeneous polynomial $\pi(\boldsymbol{x})$ of degree k in the variables x_i as follows

$$\xi = \xi\left(\sum_{i=1}^{n} x_i\right)^k = \pi(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \Delta.$$

It is trivial to verify that the polynomial $\pi(\boldsymbol{x}) - h_{\bar{G}}(\boldsymbol{x})$ is a homogeneous polynomial of degree k with nonnegative coefficients. Moreover

$$\arg\min_{\boldsymbol{x}\in\Delta} h_{\bar{G}}(\boldsymbol{x}) = \arg\max_{\boldsymbol{x}\in\Delta}\left[\xi - h_{\bar{G}}(\boldsymbol{x})\right] = \arg\max_{\boldsymbol{x}\in\Delta}\left[\pi(\boldsymbol{x}) - h_{\bar{G}}(\boldsymbol{x})\right].$$

Therefore, in order to minimize $h_{\bar{G}}$ we can apply Theorem 6 considering $P(\boldsymbol{x}) = \pi(\boldsymbol{x}) - h_{\bar{G}}(\boldsymbol{x})$, and since

$$\frac{\partial \pi(\boldsymbol{x})}{\partial x_i} = k\xi\left(\sum_{i=1}^{n} x_i\right)^{k-1} = k\xi$$

we end up with the following dynamics for the minimization of $h_{\bar{G}}$ over Δ

$$x_i^{(t+1)} = \frac{x_i^{(t)}\left[k\xi - h_{\bar{G}}^{i}(\boldsymbol{x}^{(t)})\right]}{k\xi - \sum_{j=1}^{n} x_j^{(t)} h_{\bar{G}}^{j}(\boldsymbol{x}^{(t)})}, \tag{8}$$

that will converge to a local minima of $h_{\bar{G}}$ starting from any state $\boldsymbol{x}$ in the interior of Δ, which corresponds by Theorem 5 to a maximal clique of the k-graph G.

6 A toy example

This section is not intended to provide experimental evidence that the dynamics (8) works, since we have a proof that guarantees it. Indeed, we provide a very simple toy example.

Figure 1 represents a 3-graph T, and the two sets that seem to be 4-edges are actually complete 3-subgraphs on the respective vertex sets. Hence T contains all possible 3-edges on the 5 vertices except for $\{0, 3, 4\}$ and $\{1, 3, 4\}$. T is a small example of a non conformal graph, i.e. it is not a complete-subgraphs graph of an ordinary graph. In other words, there exists no ordinary graph that has the same maximal cliques and therefore the generalization of the Motzkin-Straus theorem due to Sós and Straus [26] does not hold on this 3-graph. The set of maximal cliques of T is $\{\{0, 1, 2, 3\}, \{0, 1, 2, 4\}, \{2, 3, 4\}\}$.

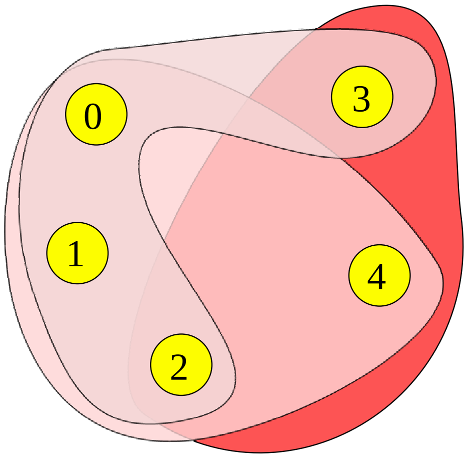


Fig. 1. A non conformal 3-graph T. Note that the sets $\{0, 1, 2, 3\}$ and $\{0, 1, 2, 4\}$ should be interpreted as complete 3-subgraphs on the respective vertex set. We draw them as 4-edges only for graphical clarity. In other words T contains all possible 3-edges on the 5 vertices except for $\{0, 3, 4\}$ and $\{1, 3, 4\}$.

We illustrate the behaviour of the dynamics (8) when applied to our toy example. Our parameters choice in this test is $\tau = \frac{1}{12}$, but no matter what is chosen as long as $0 < \tau < \frac{1}{6}$ as stated in Theorem 5, and therefore in the dynamics we have $k\xi = \frac{1}{2}$. The initial state encodes the hypothesis we make about the likelihood of a vertex to be part of a maximal clique, in fact if we set for example the i-th component of the initial state vector to zero then the i-th vertex will never be considered in a solution. Figure 2 presents three plots of the evolution of the state vector of the dynamics (8) for the 3-graph T over time. The initial states are respectively set to the simplex barycenter in the first two plots in order to have full uncertainty, and to $\boldsymbol{x}^{(0)} = (0.1, 0.1, 0.1, 0.35, 0.35)'$ in the last one in order to provide an initial stronger preference on the vertices 3 and 4.

Analysing our toy graph, we see that vertex 2 belongs to every maximal clique of T, while vertices 0 and 1 are shared between the two maximal 4-cliques and finally vertices 3 and 4 belong individually to a different maximal 4-clique, but together to the maximal 3-clique. Considering the first 114 iterations of the first two plots in Figure 2, we see that without advancing preferences of vertices, i.e. we start from the barycenter of the simplex, the dynamics converges to a stationary state, that is not optimal and hence not stable, but very informative. In fact, vertex 2 that certainly belongs to a maximal clique, has the highest likelihood, followed by vertices 0 and 1, that are shared between the two biggest maximal cliques in T and finally we find vertices 3 and 4. By inducing a small perturbation at that point, we introduce some random preference on vertices that leads the dynamics to a certain solution; in the first case we end up with the maximal 4-clique $\{0, 1, 2, 3\}$, while in the second one we end up with the maximal 4-clique $\{0, 1, 2, 4\}$. Let us move now our attention on the last plot in Figure 2. In order to extract the smallest maximal clique, that has a smaller basin of attraction we have to put stronger preferences on some vertices; for example we put stronger hypothesis on the vertices 3 and 4, since the only maximal clique they share is the smallest one. As we can see, the dynamics is able to extract also the maximal clique $\{2, 3, 4\}$. The solutions we found for this small example are the only stable ones for the dynamics (8) when applied to T. Hence randomly choosing the initial state we will certainly end up with a maximal clique, but clearly the maximal cliques with a larger basin of attraction are more likely to be extracted.

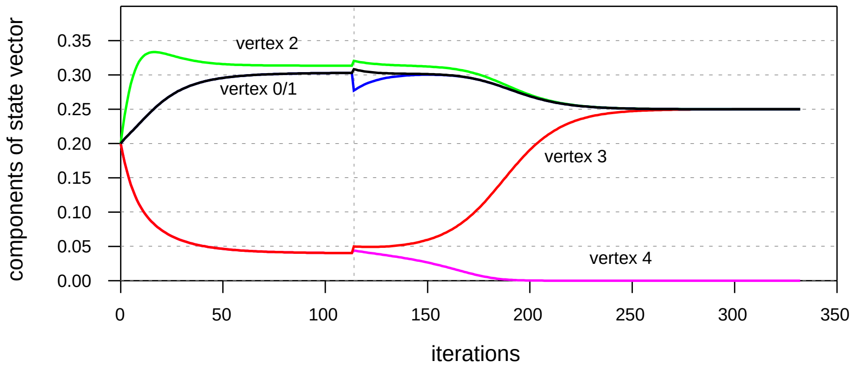
7 Conclusions and future work

In this paper we provide a generalization of a well-known extremal graph theory result, i.e. the Motzkin-Straus theorem, to k-uniform hypergraphs, and through it, we are able to provide a bridge between the purely discrete problem of finding maximal cliques in k-graphs and a minimization task of a continuous function. More precisely we introduce an isomorphism from the set of maximal/maximum cliques of a k-graph G and the set of local/global minima of the function $h_{\bar{G}}$. In this way we can focus our attention on minimizing $h_{\bar{G}}$ in order to find maximal cliques in G. Nevertheless, in the last section we provide also a dynamical system, derived from a result due to Baum and Eagon, to easily solve the optimization problem at hand. This basically furnishes an heuristic for the maximum clique problem on k-graphs.

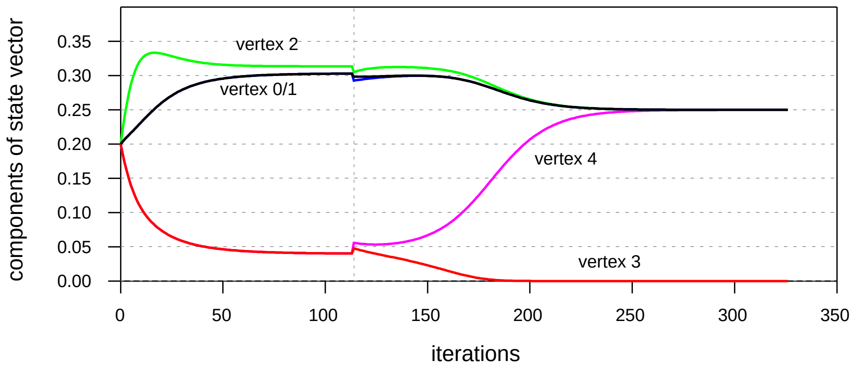
This result opens a wide range of possible future works. First of all, we may conduct experiments on the effectiveness of our heuristic for the maximum clique problem on k-graphs, however we expect in general performances on hypergraphs comparable with those obtained by Pelillo [24] on simple graphs. Even more interesting could be the theoretical applications carrying on with our work, such as finding new bounds on the clique number of k-uniform hypergraphs, or further generalizing the Motzkin-Straus theorem to vertex-weighted and edge-weighted hypergraphs.

References

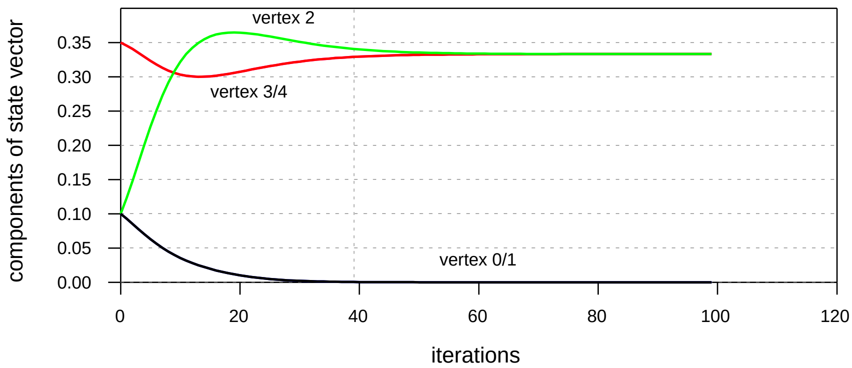
1. L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology.



(a) Extraction of the maximal clique $\{0, 1, 2, 3\}$



(b) Extraction of the maximal clique $\{0, 1, 2, 4\}$



(c) Extraction of the maximal clique $\{2, 3, 4\}$

Fig. 2. Evolution of the components of the state vector $\boldsymbol{x}^{(t)}$ for the k-graph in Figure 1, using the dynamics (8)

Bull. Amer. Math. Soc., 73:360–363, 1967.
2. L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *Ann. Math. Statistics*, 41:164–171, 1970.
3. L. E. Baum and G. R. Sell. Growth transformations for functions on manifolds. *Pacific J. Math.*, 27:221–227, 1968.
4. C. Berge. *Hypergraphs: combinatorics of finite sets.* Elsevier, 1989.
5. G. R. Blakley. Homogeneous nonnegative symmetric quadratic transformations. *Bull. Amer. Math. Soc.*, 70:712–715, 1964.
6. I. Bomze, M. Pelillo, and R. Giacomini. Evolutionary approach to the maximum clique problem: empirical evidence on a larger scale. In *Developments in Global Optimization*, pages 95–108. I. Bomze, T. Csendes, R. Horst and P. M. Pardalos, 1997.
7. I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo. The maximum clique problem. In *Handbook of Combinatorial Optimization (Supplement Volume A)*, pages 1–74. 1999.
8. I. M. Bomze, M. Budinich, M. Pelillo, and C. Rossi. Annealed replication: a new heuristic for the maximum clique problem. *Discr. Appl. Math.*, 121(1–3):27–49, 2002.
9. M. Budinich. Exact bounds on the order of the maximum clique of a graph. *Discr. Appl. Math.*, 127:535–543, 2003.
10. H. Bunke, P. J. Dickinson, and M. Kraetzl. Theoretical and algorithmic framework for hypergraph matching. In *Int. Conf. on Image Analysis and Processing*, volume 3617, pages 463–470, 2005.
11. O. D. Faugeras and M. Berthod. Improving consistency and reducing ambiguity in stochastic labeling: an optimization approach. *IEEE Trans. Pattern Anal. Machine Intell.*, 3:412–424, 1981.
12. P. Frankl and V. Rödl. Hypergraphs do not jump. *J. Combinatorica*, 4:149–159, 1984.
13. L. E. Gibbons, D. W. Hearn, and P. M. Pardalos. A continuous based heuristic for the maximum clique problem. In *Cliques, Coloring and Satisfiability: 2nd DIMACS Impl. Chall.*, volume 26, pages 103–124. 1996.
14. L. E. Gibbons, D. W. Hearn, P. M. Pardalos, and M. V. Ramana. Continuous characterizations of the maximum clique problem. *Math. Oper. Res.*, 22:754–768, 1997.
15. S. Khot. Improved inapproximability results for maxclique, chromatic number and approximate graph coloring. In *IEEE Symp. on Found. of Comp. Science*, pages 600–609, 2001.
16. D. G. Luenberger. *Linear and nonlinear programming.* Addison Wesley, 1984.
17. J. L. Mohammed, R. A. Hummel, and S. W. Zucker. A gradient projection algorithm for relaxation labeling methods. *IEEE Trans. Pattern Anal. Machine Intell.*, 5:330–332, 1983.
18. T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canad. J. Math.*, 17:533–540, 1965.
19. D. Mubay. A hypergraph extension of Turán's theorem. *J. Combin. Theory Series B*, 96:122–134, 2006.
20. D. A. Papa and I. Markov. Hypergraph partitioning and clustering. In *Approximation Algorithms and Metaheuristics*, pages 61.1–61.19. 2007.
21. P. M. Pardalos. *Continuous approaches to discrete optimization problems*, pages 313–328. 1996.

22. P. M. Pardalos and A. T. Phillips. A global optimization approach for solving the maximum clique proble. *Int. J. Comput. Math.*, 33:209–216, 1990.
23. M. Pavan and M. Pelillo. Generalizing the motzkin-straus theorem to edge-weighted graphs, with applications to image segmentation. In *Energy Minim. Methods in Computer Vision and Patt. Recogn.*, volume 2683 of *Lecture Notes in Computer Science*, pages 485–500. Springer, 2003.
24. M. Pelillo. Relaxation labeling networks for the maximum clique problem. *J. Artif. Neural Networks*, 2:313–328, 1995.
25. M. Pelillo and A. Jagota. Feasible and infeasible maxima in a quadratic program for maximum clique. *J. Artif. Neural Networks*, 2:411–420, 1995.
26. V. Sós and E. G. Straus. Extremal of functions on graphs with applications to graphs and hypergraphs. *J. Combin. Theory Series B*, 63:189–207, 1982.
27. P. Turán. On an extremal problem in graph theory (in hungarian). *Mat. ès Fiz. Lapok*, 48:436–452, 1941.
28. H. S. Wilf. The eigenvalues of a graph and its chromatic number. *J. London Math. Soc.*, 42:330–332, 1967.
29. H. S. Wilf. Spectral bounds for the clique and independence numbers of graphs. *J. Combin. Theory Series B*, 40:113–117, 1986.
30. D. Zhou, J. Huang, and B. Schölkopf. Learning with hypergraphs: clustering, classification, embedding. In *Adv. in Neur. Inf. Processing Systems*, volume 19, pages 1601–1608, 2006.