Chapter 5
Orthogonal representations I: stability number and coloring

Nachdiplomvorlesung by László Lovász
ETH Zürich, Spring 2014

1 Definition

Let $G = (V, E)$ be a simple graph. We will denote by $\overline{G} = (V, \overline{E})$ its complement.

An *orthogonal representation* of a graph $G = (V, E)$ in $\mathbb{R}^d$ assigns to each $i \in V$ a vector $\mathbf{u}_i \in \mathbb{R}^d$ such that $\mathbf{u}_i^\mathsf{T} \mathbf{u}_j = 0$ whenever $ij \in \overline{E}$. An *orthonormal representation* is an orthogonal representation in which all the representing vectors have unit length. Clearly we can always scale the nonzero vectors in an orthogonal representation this way, and usually this does not change any substantial feature of the problem.

Note that we did not insist that different nodes are represented by different vectors, nor that adjacent nodes are mapped on non-orthogonal vectors. If these conditions also hold, we call the orthogonal representation *faithful*.

Example 1.1 Every graph has a trivial orthonormal representation in $\mathbb{R}^V$, in which node i is represented by the standard basis vector $\mathbf{e}_i$. This representation is not faithful unless the graph is has no edges.

Of course, we are interested in "nontrivial" orthogonal representations, which are more "economical" than the trivial one.

Example 1.2 Figure 1 below shows that for the graph obtained by adding a diagonal to the pentagon a simple orthogonal representation in 2 dimensions can be constructed.

This example can be generalized as follows.

Example 1.3 Let $k = \overline{\chi}(G)$, and let $\{B_1, \dots, B_k\}$ be a family of disjoint complete subgraphs covering all the nodes. Let $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ be the standard basis of $\mathbb{R}^k$. Then mapping every node of B_i to $\mathbf{e}_i$ is an orthonormal representation.

A more "geometric" orthogonal representation is described by the following example.

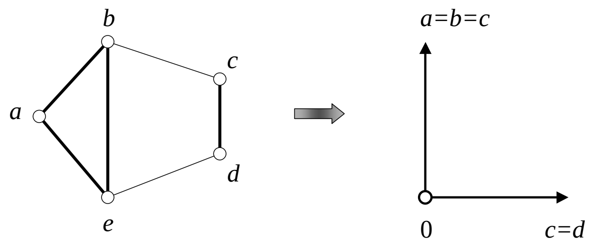

Figure 1: An (almost) trivial orthogonal representation

Example 1.4 Since $\overline{\chi}(C_5) = 3$, the previous example gives a representation of C_5 in 3-space (Figure 2, left). To get a less trivial representation, consider an "umbrella" in $\mathbb{R}^3$ with 5 ribs of unit length (Figure 2). Open it up to the point when non-consecutive ribs are orthogonal. This way we get 5 unit vectors $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$, assigned to the nodes of C_5 so that each $\mathbf{u}_i$ forms the same angle with the "handle" and any two non-adjacent nodes are labeled with orthogonal vectors. These vectors give an orthogonal representation of C_5 in 3-space.

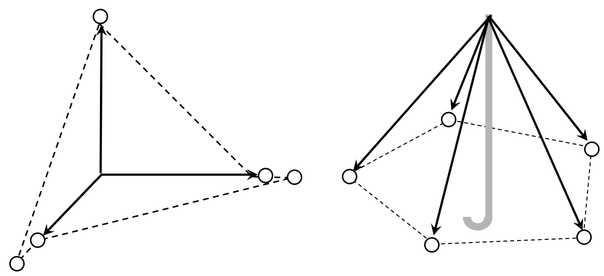

Figure 2: Two orthogonal representations of C_5.

2 Smallest cone and the theta function

When looking for "economic" orthogonal representations, we can define "economic" in several ways. For example, we may want to find an orthogonal representation in a dimension as low as possible (even though this particular way of phrasing the question does not seem to be the most fruitful; we will return to more interesting versions of the minimum dimension problem in the next section). We start with an application that provided the original motivation for considering orthogonal representations.

2.1 Shannon capacity

We start with the problem from information theory that motivated the introduction of orthogonal representations [23] and several of the results to be discussed in this chapter.

Consider a noisy channel through which we are sending messages over a finite alphabet V. The noise may blur some letters so that certain pairs can be confused. We want to select as many words of length k as possible so that no two can possibly be confused. As we shall see, the number of words we can select grows as Θ^k for some $\Theta \geq 1$, which is called the *Shannon zero-error capacity* of the channel.

In terms of graphs, we can model the problem as follows. We consider V as the set of nodes of a graph, and connect two of them by an edge if they can be confused. This way we obtain a graph $G = (V, E)$, which we call the *confusion graph* of the alphabet. We denote by $\alpha(G)$ the maximum number of independent points (the maximum size of a stable set) in the graph G. If $k = 1$, then the maximum number of non-confusable messages is $\alpha(G)$.

To describe longer messages, we use the notion of *strong product* of two graphs (see the Appendix). In terms of the confusion graph, $\alpha(G^{\boxtimes k})$ is the maximum number of non-confusable words of length k: words composed of elements of V, so that for every two words there is at least one i $(1 \leq i \leq k)$ such that the i-th letters are different and non-adjacent in G, i.e., non-confusable. It is easy to see that

$$\alpha(G \boxtimes H) \geq \alpha(G)\alpha(H). \tag{1}$$

This implies that

$$\alpha(G^{\boxtimes(k+l)}) \geq \alpha(G^{\boxtimes k})\alpha(G^{\boxtimes l}), \tag{2}$$

and

$$\alpha(G^{\boxtimes k}) \geq \alpha(G)^k. \tag{3}$$

The *Shannon capacity* of a graph G is the value

$$\Theta(G) = \lim_{k\to\infty} \alpha(G^{\boxtimes k})^{1/k}. \tag{4}$$

Inequality (2) implies that the limit exists, and (3) implies that

$$\Theta(G) \geq \alpha(G). \tag{5}$$

Rather little is known about this graph parameter. For example, it is not known whether $\Theta(G)$ can be computed for all graphs by any algorithm (polynomial or not), although there are several special classes of graphs for which this is not hard. The behavior of $\Theta(G)$ and the convergence in (4) are rather erratic; see Alon [2] and Alon and Lubetzky [4].

Example 2.1 Let C_4 denote a 4-cycle with nodes (a, b, c, d). By (5), we have $\Theta(G) \geq 2$. On the other hand, if we use a word, then all the 2^k words obtained from it by replacing a and b by each other, as well as c and d by each other, are excluded. Hence $\alpha(C_4^{\boxtimes k}) \leq 4^k/2^k = 2^k$, which implies that $\Theta(C_4) = 2$.

The argument for bounding Θ from above in this last example can be generalized as follows. Let $\overline{\chi}(G)$ denote the minimum number of complete subgraphs covering the nodes of G (this is the same as the chromatic number of the complementary graph.) Trivially

$$\alpha(G) \le \overline{\chi}(G). \tag{6}$$

Any covering of G by $\overline{\chi}(G)$ cliques and of H by $\overline{\chi}(H)$ cliques gives a "product covering" of $G \boxtimes H$ by $\overline{\chi}(G)\overline{\chi}(H)$ cliques, and so

$$\overline{\chi}(G \boxtimes H) \le \overline{\chi}(G)\overline{\chi}(H) \tag{7}$$

Hence

$$\alpha(G^{\boxtimes k}) \le \overline{\chi}(G^{\boxtimes k}) \le \overline{\chi}(G)^k,$$

and thus

$$\Theta(G) \le \overline{\chi}(G). \tag{8}$$

It follows that if $\alpha(G) = \overline{\chi}(G)$, then $\Theta(G) = \alpha(G)$; for such graphs, nothing better can be done than reducing the alphabet to the largest mutually non-confusable subset.

Example 2.2 The smallest graph for which $\Theta(G)$ cannot be computed by these means is the pentagon C_5. If we set $V(C_5) = \{0,1,2,3,4\}$ with $E(C_5) = \{01, 12, 23, 34, 40\}$, then $C_5^{\boxtimes 2}$ contains the stable set $\{(0,0),(1,2),(2,4),(3,1),(4,3)\}$. So $\alpha(C_5)^{\boxtimes 2k} = \alpha(C_5^{\boxtimes 2})^k \ge 5^k$, and hence $\Theta(C_5) \ge \sqrt{5}$.

We will show that equality holds here [23], but first we need some tools.

2.2 Definition and basic properties of the theta-function

Comparing Example 1.3 and the bound 6 suggests that using orthogonal representations of G other than those obtained from clique coverings might give better bounds on the Shannon capacity. To this end, we have to figure out what should replace the number of different cliques in this more general setting. it turns out that the smallest half-angle ϕ of a rotational cone (in arbitrary dimension) which contains all vectors in an orthogonal representation of the graph does the job [23]. We will work with a transformed version of this quantity, namely

$$\vartheta(G) = \frac{1}{(\cos\phi)^2} = \min_{(\mathbf{u}_i),\mathbf{c}} \max_{i \in V} \frac{1}{(\mathbf{c}^{\mathsf{T}}\mathbf{u}_i)^2},$$

where the minimum is taken over all orthonormal representations $(\mathbf{u}_i : i \in V)$ of G and all unit vectors $\mathbf{c}$. (We call $\mathbf{c}$ the "handle" of the representation; for the origin of the name, see Example 2.2. Of course, we could fix $\mathbf{c}$, but this is not always convenient.)

From the trivial orthogonal representation (Example 1.1) we get that $\vartheta(G) \le |V|$. Tighter inequalities can be proved:

Theorem 2.3 *For every graph G,*

$$\alpha(G) \le \vartheta(G) \le \overline{\chi}(G).$$

Proof. First, let $S \subseteq V$ be a maximum independent set of nodes in G. Then in every orthonormal representation $(\mathbf{u}_i)$, the vectors $\{\mathbf{u}_i : i \in S\}$ are mutually orthogonal unit vectors. Hence

$$1 = \mathbf{c}^\mathsf{T}\mathbf{c} \ge \sum_{i\in S}(\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2 \ge |S| \min_i (\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2,$$

and so

$$\max_{i\in V} \frac{1}{(\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2} \ge |S| = \alpha(G).$$

This implies the first inequality.

The second inequality follows from Example 1.3, using $\mathbf{c} = \frac{1}{\sqrt{k}}(\mathbf{e}_1 + \cdots + \mathbf{e}_k)$ as the handle. □

From Example 1.4 we get, using elementary trigonometry that

$$\vartheta(C_5) \le \sqrt{5}. \tag{9}$$

We'll see that equality holds here.

Lemma 2.4 *For any two graphs G and H, we have $\vartheta(G \boxtimes H) \le \vartheta(G)\vartheta(H)$.*

We will prove later (Corollary 2.10) that equality holds here.

Proof. The *tensor product* of two vectors $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$ and $\mathbf{v} = (v_1, \dots, v_m) \in \mathbb{R}^m$ is the vector

$$\mathbf{u} \circ \mathbf{v} = (u_1v_1, \dots, u_1v_m, u_2v_1, \dots, u_2v_m, \dots, u_nv_1, \dots, u_nv_m) \in \mathbb{R}^{nm}.$$

The inner product of two tensor products can be expressed easily: if $\mathbf{u}, \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{v}, \mathbf{y} \in \mathbb{R}^m$, then

$$(\mathbf{u} \circ \mathbf{v})^\mathsf{T}(\mathbf{x} \circ \mathbf{y}) = (\mathbf{u}^\mathsf{T}\mathbf{x})(\mathbf{v}^\mathsf{T}\mathbf{y}). \tag{10}$$

Now let $(\mathbf{u}_i : i \in V)$ be an optimal orthogonal representation of G with handle $\mathbf{c}$ ($\mathbf{u}_i, \mathbf{c} \in \mathbb{R}^n$), and let $(\mathbf{v}_j : j \in V(H))$ be an optimal orthogonal representation of H with handle $\mathbf{d}$ ($\mathbf{v}_j, \mathbf{d} \in \mathbb{R}^m$). It is easy to check, using (10), that the vectors $\mathbf{u}_i \circ \mathbf{v}_j$ ($(i, j) \in V(G) \times V(H)$) form an orthogonal representation of $G \boxtimes H$. Furthermore, taking $\mathbf{c} \circ \mathbf{d}$ as its handle, we have by (10) again that

$$\left((\mathbf{c} \circ \mathbf{d})^\mathsf{T}(\mathbf{u}_i \circ \mathbf{v}_j)\right)^2 = (\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2(\mathbf{d} \circ \mathbf{v}_j)^2 \ge \frac{1}{\vartheta(G)} \cdot \frac{1}{\vartheta(H)},$$

and hence

$$\vartheta(G\boxtimes H) \le \max_{i,j} \frac{1}{\big((\mathbf{c}\circ\mathbf{d})^{\mathsf T}(\mathbf{u}_i\circ\mathbf{v}_j)\big)^2} \le \vartheta(G)\vartheta(H).$$

□

This inequality has some important corollaries. We have

$$\alpha(G^{\boxtimes k}) \le \vartheta(G^{\boxtimes k}) \le \vartheta(G)^k,$$

which implies

Corollary 2.5 *For every graph G, we have $\Theta(G) \le \vartheta(G)$.*

In particular, combining with (9) and Example 2.2, we get a solution of Shannon's problem:

$$\sqrt{5} \le \Theta(C_5) \le \vartheta(C_5) \le \sqrt{5},$$

and equality must hold throughout.

We can generalize this argument. The product graph $G\boxtimes\overline{G}$ has independent points in the diagonal, implying that

$$\vartheta(G\boxtimes\overline{G}) \ge \alpha(G\boxtimes\overline{G}) \ge |V|.$$

Together with Lemma 2.4, this yields:

Corollary 2.6 *For every graph $G=(V,E)$, we have $\vartheta(G)\vartheta(\overline{G}) \ge |V|$.*

Since $\overline{C_5} \cong C_5$, Corollary 2.6 implies that $\vartheta(C_5) \ge \sqrt{5}$, and together with (9), we get another proof of the fact that $\vartheta(C_5)=\sqrt{5}$. Equality does not hold in general in Corollary 2.6, but it does when G has a node-transitive automorphism group. We postpone the proof of this fact until some further formulas for ϑ will be developed (Corollary 2.12).

2.3 More expressions for ϑ

We prove a number of formulas for $\vartheta(G)$, which together have a lot of implications. The first bunch of these will be proved together. We define a number of parameters, which will all turn out to be equal to $\vartheta(G)$.

We start with a geometric definition, discovered independently by Karger, Motwani and Sudan [18]. In terms of the complementary graph, this value is sometimes called the "vector chromatic number". Let

$$\vartheta_1 = \min\Big\{t \ge 2 : \exists \mathbf{w}_i \in \mathbb{R}^d \text{ such that } |\mathbf{w}_i| = 1,\ \mathbf{w}_i^{\mathsf T}\mathbf{w}_j = -\frac{1}{t-1}\ (\forall ij \in \overline{E})\Big\}. \qquad (11)$$

Next, we give a couple of formulas for ϑ in terms of semidefinite optimization. Consider the following two semidefinite programs:

$$\begin{aligned} &\text{minimize} && t && \text{maximize} && \sum_{i,j\in V} Z_{ij} \\ &\text{subject to} && Y \succeq 0 && \text{subject to} && Z \succeq 0 \\ &&& Y_{ij} = -1 \quad (\forall\, ij \in \overline{E}) &&&& Z_{ij} = 0 \quad (\forall\, ij \in E) \\ &&& Y_{ii} = t-1 &&&& \mathrm{tr}(Z) = 1 \end{aligned} \tag{12}$$

It is not hard to check that these are dual programs. The first program has a strictly feasible solution (just choose t large enough), so by the Duality Theorem of semidefinite programming, the two programs have the same objective value; we call this number ϑ_2.

Finally, we use orthonormal representations of the complementary graph: define

$$\vartheta_3 = \max \sum_{i\in V} (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2, \tag{13}$$

where the maximum extends over all orthonormal representations $(\mathbf{v}_i :\ i \in V)$ of the complementary graph $\overline{G}$ and all unit vectors $\mathbf{d}$.

The main theorem of this section asserts that all these definitions lead to the same value.

Theorem 2.7 *For every graph G, $\vartheta(G) = \vartheta_1 = \vartheta_2 = \vartheta_3$.*

Proof. We prove that

$$\vartheta(G) \le \vartheta_1 \le \vartheta_2 \le \vartheta_3 \le \vartheta(G). \tag{14}$$

To prove the first inequality, let $(\mathbf{w}_i :\ i \in V)$ be the representation that achieves the minimum in (11). Let $\mathbf{c}$ be a vector orthogonal to all the $\mathbf{w}_i$ (we increase the dimension of the space if necessary). Let

$$\mathbf{u}_i = \frac{1}{\sqrt{\vartheta_1}}(\mathbf{c}+\mathbf{w}_i).$$

Then $|\mathbf{u}_i| = 1$ and $\mathbf{u}_i^\mathsf{T}\mathbf{u}_j = 0$ for $ij \in \overline{E}$, so $(\mathbf{u}_i)$ is an orthonormal representation of G. Furthermore, with handle $\mathbf{c}$ we have $\mathbf{c}^\mathsf{T}\mathbf{u}_i = 1/\sqrt{\vartheta_1}$, which implies that $\vartheta(G) \le \vartheta_1$.

Second, let (Y, t) be an optimal solution of (12). Then $\vartheta_2 = t$, and the matrix $1/(t-1)Y$ is positive semidefinite, and so it can be written as a Gram matrix, i.e., there are vectors $\mathbf{w}_i \in \mathbb{R}^n$ such that

$$\mathbf{w}_i^\mathsf{T}\mathbf{w}_j = \frac{1}{\sqrt{t-1}} Y_{ij} = \begin{cases} -\dfrac{-1}{t-1}, & \text{if } ij \in \overline{E}, \\ 1, & \text{if } i = j. \end{cases}$$

(no condition if $ij \in E$). So these vectors satisfy the conditions in (11) for the given t. Since ϑ_1 is the smallest t for which this happens, this proves that $\vartheta_1 \le t = \vartheta_2$.

To prove the third inequality in (14), let Z be an optimum solution of the dual program in (12) with objective function value ϑ_2. We can write Z is a Gram matrix: $Z_{ij} = \mathbf{z}_i^\mathsf{T}\mathbf{z}_j$ where $\mathbf{z}_i \in \mathbb{R}^k$ for some $k \geq 1$. Let us rescale the vectors $\mathbf{z}_i$ to get the unit vectors $\mathbf{v}_i = \mathbf{z}_i^0$ (if $\mathbf{z}_i = 0$ then we take a unit vector orthogonal to everything else as $\mathbf{v}_i$). Define $\mathbf{d} = (\sum_i \mathbf{z}_i)^0$.

By the properties of Z, the vectors $\mathbf{v}_i$ form an orthonormal representation of $\overline{G}$, and hence

$$\vartheta_3 \geq \sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2.$$

To estimate the right side, we use the equation

$$\sum_i |\mathbf{z}_i|^2 = \sum_i \mathbf{z}_i^\mathsf{T}\mathbf{z}_i 1 = \mathrm{tr}(Z) = 1$$

and Cauchy Schwarz:

$$\sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2 = \Big(\sum_i |\mathbf{z}_i|^2\Big)\Big(\sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2\Big) \geq \Big(\sum_i |\mathbf{z}_i| \mathbf{d}^\mathsf{T}\mathbf{v}_i\Big)^2$$
$$= \Big(\sum_i \mathbf{d}^\mathsf{T}\mathbf{z}_i\Big)^2 = \Big(\mathbf{d}^\mathsf{T}\sum_i \mathbf{z}_i\Big)^2 = \Big|\sum_i \mathbf{z}_i\Big|^2 = \sum_{i,j} \mathbf{z}_i^\mathsf{T}\mathbf{z}_j = \vartheta_2.$$

This proves that $\vartheta_3 \geq \vartheta_2$.

Finally, to prove the last inequality in (14), it suffices to prove that if $(\mathbf{u}_i : i \in V)$ is an orthonormal representation of G in $\mathbb{R}^n$ with handle $\mathbf{c}$, and $(\mathbf{v}_i : i \in V)$ is an orthonormal representation of $\overline{G}$ in $\mathbb{R}^m$ with handle $\mathbf{d}$, then

$$\sum_{i \in V} (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2 \leq \max_{i \in V} \frac{1}{(\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2}. \tag{15}$$

By a similar computation as in the proof of Lemma 2.4, we get that the vectors $\mathbf{u}_i \circ \mathbf{v}_i$ $(i \in V)$ are mutually orthogonal unit vectors, and hence

$$\sum_i (\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2(\mathbf{d}^\mathsf{T}\mathbf{v}_j)^2 = \sum_i \big((\mathbf{c}\circ\mathbf{d})^\mathsf{T}(\mathbf{u}_i \circ \mathbf{v}_i)\big)^2 \leq 1. \tag{16}$$

On the other hand,

$$\sum_i (\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2(\mathbf{d}^\mathsf{T}\mathbf{v}_j)^2 \geq \min_i (\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2 \sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_j)^2,$$

which implies that

$$\sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_j)^2 \geq \frac{1}{\min_i (\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2} = \max_i \frac{1}{(\mathbf{c}^\mathsf{T}\mathbf{u}_i)^2}.$$

This proves (15) and completes the proof of Theorem 2.7. □

These formulations have many versions, one of which are stated below without proof, which is left to the reader as an exercise.

Proposition 2.8 *$\vartheta(G)$ is the minimum of the largest eigenvalues of all symmetric matrices $M \in \mathbb{R}^{V\times V}$ such that $M_{ij} = 1$ whenever $ij \in \overline{E}$ or $i = j$.*

2.4 Consequences of the formulas

From the fact that equality holds in (14), it follows that equality holds in all the arguments above. Let us formulate some consequences. Considering the optimal orthogonal representation constructed in the first step of the proof, we get:

Corollary 2.9 *Every graph G has an orthonormal representation $(\mathbf{u}_i)$ with handle $\mathbf{c}$ such that for every node i,*

$$\mathbf{c}^\mathsf{T}\mathbf{u}_i = \frac{1}{\sqrt{\vartheta(G)}}.$$

As a further application of the duality established in Section 2.3, we prove that equality holds in Lemma 2.4:

Proposition 2.10 *For any two graphs G and H,*

$$\vartheta(G \boxtimes H) = \vartheta(G)\vartheta(H).$$

Proof. Let $(\mathbf{v}_i, \mathbf{d})$ be an orthonormal representation of $\overline{G}$ which is optimal in the sense that $\sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2 = \vartheta(G)$, and let $(\mathbf{w}_j, \mathbf{e})$ be an orthonormal representation of $\overline{H}$ such that $\sum_i (\mathbf{e}^\mathsf{T}\mathbf{w}_i)^2 = \vartheta(H)$. It is easy to check that the vectors $\mathbf{v}_i \circ \mathbf{w}_j$ form an orthonormal representation of $\overline{G \boxtimes H}$, and so using handle $\mathbf{d} \circ \mathbf{e}$ we get

$$\vartheta(G \boxtimes H) \geq \sum_{i,j} \big((\mathbf{d} \circ \mathbf{e})^\mathsf{T}(\mathbf{v}_i \circ \mathbf{w}_j)\big)^2 = \sum_{i,j} (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2(\mathbf{e}^\mathsf{T}\mathbf{w}_j)^2 = \vartheta(G)\vartheta(H).$$

We already know the reverse inequality, which completes the proof. □

The matrix descriptions of ϑ imply an important property of optimal orthogonal representations, namely automorphism invariance.

Proposition 2.11 *Every graph G has an orthonormal representation $(\mathbf{u}_i, \mathbf{c})$ in $\mathbb{R}^n$ with $\mathbf{c}^\mathsf{T}\mathbf{u}_i = 1/\sqrt{\vartheta(G)}$ for all nodes i, and its complement has an orthonormal representation $(\mathbf{v}_i, \mathbf{d})$ in $\mathbb{R}^n$ with $\sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2 = \vartheta(G)$, such that every automorphism of G can be lifted to an orthogonal transformation of $\mathbb{R}^n$ that leaves both representations invariant.*

Proof. We give the proof for the orthonormal representation of the complement. The set of optimum solutions of the dual semidefinite program in (12) form a bounded convex set, which is invariant under the transformations $Z \mapsto PZP$, where P is the permutation matrix defined by an automorphism of G. The center of gravity of this convex set is a matrix Z which is fixed by these transformations, i.e., it satisfies $PZP = Z$ for all automorphisms P. The construction of an orthonormal representations of $\overline{G}$ in the proof of Theorem 2.7 can be done in a canonical way (e.g., choosing the rows of $Z^{1/2}$ as the vectors $\mathbf{z}_i$), and so the obtained optimal orthonormal representation will be invariant under the automorphism group of G (acting as permutation matrices). □

Corollary 2.12 *If G has a node-transitive automorphism group, then*

$$\vartheta(G)\vartheta(\overline{G}) = |V|.$$

Proof. It follows from Proposition 2.11 that $\overline{G}$ has an orthonormal representation $(\mathbf{v}_i, \mathbf{d})$ in $\mathbb{R}^n$ such that $\sum_i (\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2 = \vartheta(G)$, and $\mathbf{d}^\mathsf{T}\mathbf{v}_i$ is independent of i. So $(\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2 = \vartheta(G)/|V|$ for all nodes i, and hence

$$\vartheta(\overline{G}) \le \max_i \frac{1}{(\mathbf{d}^\mathsf{T}\mathbf{v}_i)^2} = \frac{|V|}{\vartheta(G)}.$$

Since we already know the reverse inequality (2.6), this proves the Corollary. □

Corollary 2.13 *If G is a self-complementary graph with a node-transitive automorphism group, then $\Theta(G) = \vartheta(G) = \sqrt{|V|}$.*

Proof. The diagonal in $G \boxtimes \overline{G}$ is independent, so $\alpha(G \boxtimes G) = \alpha(G \boxtimes \overline{G}) \ge |V|$, and hence $\Theta(G) \ge \sqrt{|V|}$. On the other hand, $\vartheta(G) \le \sqrt{|V|}$ follows by Corollary 2.12. □

Example 2.14 (Paley graphs) The *Paley graph* Pal_p is for a prime $p \equiv 1 \pmod 4$. We take the $\{0, 1, \dots, p-1\}$ as nodes, and connect two of them if their difference is a quadratic residue. It is clear that these graphs have a node-transitive automorphism group, and it is easy to see that they are self-complementary. So Corollary 2.13 applies, and gives that $\Theta(\text{Pal}_p) = \vartheta(\text{Pal}_p) = \sqrt{p}$. To determine the stability number of Paley graphs is a difficult unsolved number-theoretic problem, but it is conjectured that $\alpha(\text{Pal}_p) = O((\log p)^2)$. Supposing this conjecture is true, we get an infinite family for which the Shannon capacity is non-trivial (i.e., $\Theta > \alpha$), and can be determined exactly.

2.5 Random graphs

The Paley graphs are quite similar to random graphs, and indeed, for random graphs ϑ behaves similarly as for the Paley graphs, namely it is of the order $\sqrt{n}$ (Juhász [17]). (It is not known, however, how large the Shannon capacity of a random graph is.) As usual, $G(n, p)$ denotes a random graph on n nodes with edge density p. Here we assume that p is a constant and $n \to \infty$. The analysis extends to the case when $(\ln n)1/6/n \le p \le 1-(\ln n)1/6/n$ (Coja-Oghlan and Taraz [8]). See Coja-Oghlan [7] for more results about the concentration of this value.

Theorem 2.15 *With high probability, $\sqrt{(1-p)n}/3 < \vartheta(G(n,p)) < 3\sqrt{n/p}$.*

2.6 The TSTAB body

The basic technique of applying linear programming in discrete optimization is polyhedral combinatorics. Instead of surveying this broad topic, we illustrate it by recalling some results on the stable set polytope. A detailed account can be found e.g. in [14].

Let $G = (V, E)$ be a graph; it is convenient to assume that it has no isolated nodes. The *Stable Set Problem* is the problem of finding $\alpha(G)$. This problem is NP-hard.

The basic idea in applying linear programming to study the stable set problem is the following. For every subset $S \subseteq V$, let $\mathbb{1}^S \in \mathbb{R}^V$ denote its indicator vector. The *stable set polytope* $\text{STAB}(G)$ of G is the convex hull of incidence vectors of all stable sets.

There is a system of linear inequalities whose solution set is exactly the polytope $\text{STAB}(G)$, and if we can find this system, then we can find $\alpha(G)$ by optimizing the linear objective function $\sum_i x_i$. Unfortunately, this system is in general exponentially large and very complicated. But if we can find at least some linear inequalities valid for the stable set polytope, then using these we get an upper bound on $\alpha(G)$, and for special graphs, we get the exact value.

Let us survey some classes of known constraints.

NON-NEGATIVITY CONSTRAINTS:

$$x_i \geq 0 \quad (i \in V). \tag{17}$$

EDGE CONSTRAINTS:

$$x_i + x_j \leq 1 \quad (ij \in E). \tag{18}$$

These inequalities define a polytope $\text{FSTAB}(G)$. The integral points in $\text{FSTAB}(G)$ are exactly the incidence vectors of stable sets, but $\text{FSTAB}(G)$ may have other non-integral vertices, and is in general larger than $\text{STAB}(G)$ (see Figure 3).

Proposition 2.16 (a) $\text{STAB}(G) = \text{FSTAB}(G)$ *iff G is bipartite.*

(b) *The vertices of* $\text{FSTAB}(G)$ *are half-integral.*

CLIQUE CONSTRAINTS: For every clique (complete subgraph) B, we write up an inequality

$$\sum_{i \in B} x_i \leq 1. \tag{19}$$

Inequalities (17) and (19) define a polytope $\text{QSTAB}(G)$, which is contained in $\text{FSTAB}(G)$, but is in general larger than $\text{STAB}(G)$.

Recall that for a graph G, we denote by $\omega(G)$ the size of its largest clique, and by $\chi(G)$, its chromatic number. A graph G is called *perfect*, if for every induced subgraph G' of G, we have $\omega(G') = \chi(G')$. To be perfect is a rather strong structural property; nevertheless, many interesting classes of graphs are perfect (bipartite graphs, their complements and their

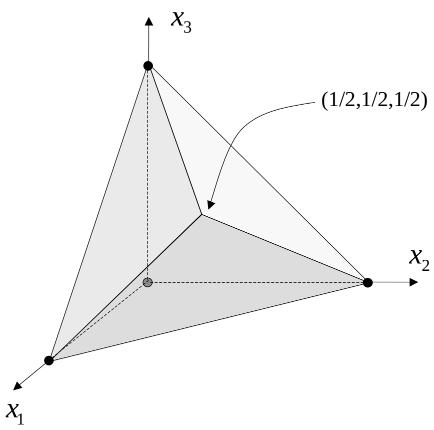

Figure 3: The fractional stable set polytope of the triangle. The black dots are incidence vectors of stable sets; the vertex $(1/2, 1/2, 1/2)$ (closest to us) is not a vertex of $\mathrm{STAB}(K_3)$.

linegraphs; interval graphs; comparability and incomparability graphs of posets; chordal graphs). We do not discuss perfect graphs, only to the extent needed to show their connection with orthogonal representations.

The following deep characterization perfect graphs was conjectured by Berge in 1961 and proved by Chudnovsky, Robertson, Seymour and Thomas [6].

Theorem 2.17 (The Strong Perfect Graph Theorem [6]) *A graph is perfect if and only if neither the graph nor its complement contains a chordless odd cycle longer than* 3.

As a corollary we can formulate the "The Weak Perfect Graph Theorem" proved much earlier [22]:

Theorem 2.18 *The complement of a perfect graph is perfect.*

From this it follows that in the definition of perfect graphs we could replace the equation $\omega(G') = \chi(G')$ by $\alpha(G') = \chi(\overline{G'})$. In particular, if G is a perfect graph, then $\alpha(G) = \chi(\overline{G})$, and so by Theorem 2.3,

Corollary 2.19 *For every perfect graph G, $\Theta(G) = \vartheta(G) = \alpha(G) = \chi(\overline{G})$.*

Orthogonality constraints: For every orthonormal representation $(\mathbf{v}_i, \mathbf{c})$ of $\overline{G}$, we consider the linear constraint

$$\sum_{i \in V} (\mathbf{c}^\mathsf{T} \mathbf{v}_i)^2 x_i \le 1. \tag{20}$$

It is easy to see that these inequalities are valid for $\mathrm{STAB}(G)$; we call them *orthogonality constraints*. The solution set of non-negativity and orthogonality constraints is denoted by

TSTAB(G). This construction was introduced by Grötschel, Lovász and Schrijver [13], who also established the properties below.

It is clear that TSTAB is a closed, convex set. The incidence vector of any stable set S satisfies (20) (we have seen this in the proof of Theorem 2.3). Furthermore, every clique constraint is an orthogonality constraint. Indeed, for every clique B, the constraint $\sum_{i\in B} x_i \le 1$ is obtained from the orthogonal representation

$$i \mapsto \begin{cases} \mathbf{e}_1, & i \in A, \\ \mathbf{e}_i, & \text{otherwise}, \end{cases} \qquad \mathbf{c} = \mathbf{e}_1.$$

Hence $\mathrm{STAB}(G) \subseteq \mathrm{TSTAB}(G) \subseteq \mathrm{QSTAB}(G)$ for every graph G.

There is a dual characterization of TSTAB. For every orthonormal representation $\mathbf{u} = (\mathbf{u}_i :\ i \in V)$, consider the vector $\mathbf{x}(\mathbf{u}) = ((\mathbf{c}^{\mathsf{T}}\mathbf{u}_i)^2 :\ i \in V)$. Then

$$\mathrm{TSTAB}(G) = \{\mathbf{x}(\mathbf{u}) :\ (\mathbf{u}, \mathbf{c}) \text{ is an orthonormal representation of } G\}. \tag{21}$$

This can be derived from semidefinite duality.

Not every orthogonality constraint follows from the clique constraints; in fact, the number of essential orthogonality constraints is infinite in general:

Theorem 2.20 TSTAB(G) *is polyhedral if and only if the graph is perfect. In this case* TSTAB = STAB = QSTAB.

While TSTAB is a rather complicated set, in many respects it behaves much better than, say, STAB. For example, it has a very nice connection with graph complementation:

Theorem 2.21 TSTAB($\overline{G}$) *is the antiblocker of* TSTAB(G).

2.7 Algorithmic applications

Perhaps the most important consequence of the formulas proved in Section 2.3 is that the value of $\vartheta(G)$ is polynomial time computable [11]. More precisely,

Theorem 2.22 *There is a polynomial time algorithm that computes, for every graph G and every $\varepsilon > 0$, a real number t such that*

$$|\vartheta(G) - t| < \varepsilon.$$

Algorithms proving this theorem can be based on almost any of our formulas for ϑ. The simplest is to refer to 12, and the polynomial time solvability of semidefinite programs (see the Background Material).

The significance of this fact is underlined if we combine it with Theorem 2.3: The two important graph parameters $\alpha(G)$ and $\chi(\overline{G})$ are both NP-hard, but they have a polynomial time computable quantity sandwiched between them. This fact is particularly useful for perfect graphs.

Corollary 2.23 *The independence number and the chromatic number of a perfect graph are polynomial time computable.*

Theorem 2.22 extends to the weighted version of the theta function. Maximizing a linear function over STAB(G) or QSTAB(G) is NP-hard; but, surprisingly, TSTAB behaves much better: *Every linear objective function can be maximized over* TSTAB(G) *(with arbitrarily small error) in polynomial time.* The maximum of $\sum_i x_i$ over TSTAB(G) is the familiar function $\vartheta(G)$. See [14, 21] for more detail.

We mention another important application to a coloring problem. Suppose that somebody gives a graph and guarantees that the graph is 3-colorable, without telling us its 3-coloring. Can we find this 3-coloring? (This may sound artificial, but this kind of situation does arise in cryptography and other data security applications; one can think of the hidden 3-coloring as a "watermark" that can be verified if we know where to look.)

It is easy to argue that knowing that the graph is 3-colorable does not help: it is still NP-hard to find the 3-coloration. But suppose that we would be satisfied with finding a 4-coloration, or 5-coloration, or $(\log n)$-coloration; is this easier? It is known that to find a 4-coloration is still NP-hard, but little is known above this. Improving earlier results, Karger, Motwani and Sudan [18] gave a polynomial time algorithm that, given a 3-colorable graph, computes a coloring with $O^*(n^{1/4})$ colors. More recently, this was improved by Blum and Karger [5] to $O^*(n^{3/14})$. The theorem behind the algorithm of Karger, Motwani and Sudan is the following.

Theorem 2.24 *For every 3-colorable graph G, we can construct in randomized polynomial time a stable set of size $n^{3/4}/(3\sqrt{\ln n})$.*

Proof. Let $G = (V, E)$. We may assume that G contains a triangle (just add one hanging from an edge). Then $\omega(G) = \chi(\overline{G}) = \vartheta(\overline{G}) = 3$. If G contains a node i with degree $n^{3/4}$, then $G[N(i)]$ is bipartite, and so it contains a stable set with $n^{3/4}/2$ nodes (and we can find in it easily). So we may suppose that all degrees are less than $n^{3/4}$, and hence $|E| < n^{7/4}/4$.

By Theorem 2.7, there are unit vectors $\mathbf{u}_i \in \mathbb{R}^n$ such that $\mathbf{u}_i^\mathsf{T}\mathbf{u}_j = -1/2$ whenever $ij \in E$. In other words, the angle between $\mathbf{u}_i$ and $\mathbf{u}_j$ is 120°.

For given $0 < s < 1$ and $\mathbf{v} \in S^{n-1}$, let $C_{\mathbf{v},s}$ denote a cap on S^{n-1} cut off by a halfspace $\mathbf{v}^\mathsf{T}\mathbf{x} \geq s$. This cup has center $\mathbf{v}$. Let its surface area be $\mathrm{vol}_{n-1}(C_{\mathbf{v},s}) = a(s)\mathrm{vol}_{n-1}(S^{n-1})$, and let $S = \{i \in V : \ \mathbf{u}_i \in C_{\mathbf{v},s}\}$. Choosing the center $\mathbf{v}$ uniformly at random on S^{n-1}, the expected number of points $\mathbf{u}_i$ in $C_\mathbf{v}$ is $a(s)n$.

We want to bound the expected number of edges spanned by S. Let $ij \in E$, then the probability that $C_\mathbf{v}$ contains both $\mathbf{u}_i$ and $\mathbf{u}_j$ is

$$b(s) = \frac{\mathrm{vol}_{n-1}(C_{\mathbf{u}_i} \cap C_{\mathbf{u}_j})}{\mathrm{vol}_{n-1}(S^{n-1})}$$

So the expected number of edges induced by S is $b(s)|E| < b(s)n^{7/4}$.

Deleting one endpoint of every edge induced by S, we get a stable set of nodes. Hence $\alpha(G) \geq |S| - |E(S)|$, and taking expectation, we get

$$\alpha(G) \geq a(s)n - b(s)n^{7/4}. \tag{22}$$

So it suffices to estimate $a(s)$ and $b(s)$ as functions of s, and then choose the best s. This is elementary computation in geometry, but it is spherical geometry in n-space, and the computations are a bit tedious.

Any point $\mathbf{x} \in C_{\mathbf{u}_i} \cap C_{\mathbf{u}_j}$ satisfies $\mathbf{u}_i^\mathsf{T}\mathbf{x} \geq s$ and also $\mathbf{u}_j^\mathsf{T}\mathbf{x} \geq s$, hence it satisfies

$$(\mathbf{u}_i + \mathbf{u}_j)^\mathsf{T}\mathbf{x} \geq 2s. \tag{23}$$

Since $\mathbf{u}_i + \mathbf{u}_j$ is a unit vector, this implies that $C_{\mathbf{u}_i,s} \cap C_{\mathbf{u}_j,s} \subseteq C_{\mathbf{u}_i+\mathbf{u}_j,2s}$, and so $b(s) \leq a(2s)$. Thus

$$\alpha(G) \geq a(s)n - a(2s)n^{7/4}. \tag{24}$$

It is known from geometry that

$$\frac{1}{10s\sqrt{n}}(1-s^2)^{(n-1)/2} < a(s) < \frac{1}{s\sqrt{n}}(1-s^2)^{(n-1)/2}.$$

Thus

$$\alpha(G) \geq a(s)n - a(2s)n^{7/4} \geq \frac{\sqrt{n}}{10s}(1-s^2)^{(n-1)/2} - \frac{n^{5/4}}{4s}(1-4s^2)^{(n-1)/2}.$$

We want to choose s so that it maximizes the right side. By elementary computation we get that $s = \sqrt{(\ln n)/(2n)}$ is an approximately optimal choice, which gives the estimate in the theorem. $\square$

The algorithm of Karger, Motwani and Sudan starts with computing an independent set of size $O^*(n^{3/4})$. Using the previous theorem, they find a stable set of size $\Omega(n^{3/4}/\sqrt{\ln n})$. Deleting this set from G and iterating, they get a coloring of G with $O^*(n^{1/4})$ colors.

Exercise 2.25 If $\vartheta(\overline{G}) = 2$, then G is bipartite.

Exercise 2.26 (a) If H is an induced subgraph of G, then $\vartheta(H) \leq \vartheta(G)$. (b) If H is a spanning subgraph of G, then $\vartheta(H) \geq \vartheta(G)$.

Exercise 2.27 Let G be a graph and $v \in V$. (a) $\vartheta(G-v) \geq \vartheta(G) - 1$. (b) If v is an isolated node, then equality holds. (c) If v is adjacent to all other nodes, then $\vartheta(G-v) = \vartheta(G)$.

Exercise 2.28 Let $G = (V,E)$ be a graph and let $V = S_1 \cup \cdots \cup S_k$ be a partition of V. (a) Then $\vartheta(G) \leq \sum_i \vartheta(G[S_i])$. (b) If no edge connects nodes in different sets S_i, then equality holds. (c) Suppose that any two nodes in different sets S_i are adjacent. How can $\vartheta(G)$ be expressed in terms of the $\vartheta(G[S_i])$?

Exercise 2.29 Let $G = (V, E)$ be a graph, and for $i \in V$, let $\overline{N}(i) = V \setminus (N(i) \cup \{i\}$ and $G_i = G[\overline{N}(i)]$. Then

$$\vartheta(G)(\vartheta(G)-1)^2 \le \sum_{i \in V} \vartheta(G_i)^2.$$

Furthermore, every graph $G = (V, E)$ has a node j such that

$$(\vartheta(G)-1)^2 \le \sum_{i \in V} \vartheta(G_i)^2.$$

Exercise 2.30 For a graph G, let $t(G)$ denote that radius of the smallest sphere (in any dimension) on which a given graph G can be drawn so that the euclidean distance between adjacent nodes is 1. Prove that $t(G)^2 = \frac{1}{2}\big(1 - 1/\vartheta(\overline{G})\big)$.

Exercise 2.31 (a) Show that any stable set S provides a feasible solution of dual program in (12). (b) Show that any k-coloring of $\overline{G}$ provides a feasible solution of the primal program in (12). (c) Give a new proof of the Sandwich Theorem 2.3 based on (a) and (b).

Exercise 2.32 Prove that $\vartheta(G)$ is the maximum of the largest eigenvalues of matrices $(\mathbf{v}_i^\mathsf{T}\mathbf{v}_j)$, taken over all orthonormal representations $(\mathbf{v}_i)$ of $\overline{G}$.

Exercise 2.33 Let $G = (V, E)$ be a graph and let $S, T \subseteq V$. Then $\vartheta(G[S \cup T]) \le \vartheta(G[S]) + \vartheta(G[T])$.

Exercise 2.34 (Alon and Kahale) Let G be a graph, $v \in V$, and let H be obtained from G by removing v and all its neighbors. Prove that

$$\vartheta(G) \le 1 + \sqrt{|V(H)|\vartheta(H)}.$$

Exercise 2.35 The *fractional chromatic number* $\chi^*(G)$ is defined as the least t for which there exists a family $(A_j : \; j = 1, \dots, p)$ of stable sets in G, and nonnegative weights $(\tau_j : \; j = 1, \dots, p)$ such that $\sum_j \tau_j = t$ and $\sum_j \tau_j \mathbb{1}_{A_j} \ge \mathbb{1}_V$.

(a) Prove that $\chi^*(G)$ is equal to the largest s for which there exist nonnegative weights $(\sigma_i : \; i \in V)$ such that $\sum_i \sigma_i = s$ and $\sum_{i \in A} \sigma_i \le 1$ for every stable set A.

(b) Prove that $\vartheta(\overline{G}) \le \chi^*(G) \le \chi(G)$.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi: A note on Ramsey numbers, *J. Combin. Theory A* **29** (1980), 354–360.

[2] N. Alon: The Shannon capacity of a union, *Combinatorica* **18** (1998), 301–310.

[3] N. Alon and N. Kahale: Approximating the independence number via the ϑ-function, *Math. Programming* **80** (1998), Ser. A, 253–264.

[4] N. Alon and E. Lubetzky: The Shannon capacity of a graph and the independence numbers of its powers, *IEEE Trans. Inform. Theory* **52** (2006), 2172–2176.

[5] A. Blum and D. Karger: An $\widetilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs, *J. Inform. Proc. Lett.* **61** (1997), 49–53.

[6] M. Chudnovsky, N. Robertson, P. Seymour and R. Thomas: The strong perfect graph theorem, *Annals of Math.*, **164** (2006), 51–229.

[7] A. Coja-Oghlan: The Lovász number of random graphs, in: *Approximation, randomization, and combinatorial optimization*, Lecture Notes in Comput. Sci. **2764**, Springer, Berlin (2003), 228–239.

[8] A. Coja-Oghlan and A. Taraz: Exact and approximative algorithms for coloring $G(n,p)$, *Random Structures and Algorithms* **24** (2004), 259–278.

[9] U. Feige: Randomized graph products, chromatic numbers, and the Lovász θ function, *Combinatorica* **17** (1997), 79–90.

[10] Z. Fredi and J. Komls: The eigenvalues of random symmetric matrices, *Combinatorica* **1** (1981), 233–241.

[11] M. Grötschel, L. Lovász and A. Schrijver: The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1** (1981), 169-197.

[12] M. Grötschel, L. Lovász and A. Schrijver: Polynomial algorithms for perfect graphs, *Annals of Discrete Math.* **21** (1984), 325-256.

[13] M. Grötschel, L. Lovász, A. Schrijver: Relaxations of vertex packing, *J. Combin. Theory B* **40** (1986), 330-343.

[14] M. Grötschel, L. Lovász, A. Schrijver: Geometric Algorithms and Combinatorial Optimization, Springer (1988).

[15] W. Haemers: On some problems of Lovász concerning the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **25** (1979), 231–232.

[16] J. Håstad: Clique is hard to approximate within a factor of $n^{1-\varepsilon}$, *Acta Math.* **182** (1999), 105–142.

[17] F. Juhász: The asymptotic behaviour of Lovász' ϑ function for random graphs, *Combinatorica* **2** (1982) 153–155.

[18] D. Karger, R. Motwani, M. Sudan: Approximate graph coloring by semidefinite programming, *Proc. 35th FOCS* (1994), 2–13.

[19] B. S. Kashin and S. V. Konyagin: On systems of vectors in Hilbert spaces, *Trudy Mat. Inst. V.A.Steklova* **157** (1981), 64–67; English translation: *Proc. of the Steklov Inst. of Math.* (AMS 1983), 67–70.

[20] V. S. Konyagin, Systems of vectors in Euclidean space and an extremal problem for polynomials, *Mat. Zametky* **29** (1981), 63–74. English translation: *Math. Notes of the Academy USSR* **29** (1981), 33–39.

[21] D. E. Knuth: The sandwich theorem, *The Electronic Journal of Combinatorics* **1** (1994), 48 pp.

[22] L. Lovász: Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972), 253-267.

[23] L. Lovász: On the Shannon capacity of graphs, *IEEE Trans. Inform. Theory* **25** (1979), 1–7.

[24] M. Szegedy: A note on the θ number of Lovász and the generalized Delsarte bound, *Proc. 35th FOCS* (1994), 36–39.