

THE SUBGAUSSIAN CONSTANT AND CONCENTRATION INEQUALITIES

S.G. Bobkov*, C. Houdré†, P. Tetali‡

Abstract

We study concentration inequalities for Lipschitz functions on graphs by estimating the optimal constant in exponential moments of subgaussian type. This is illustrated on various graphs and related to the spread constant, introduced by Alon, Boppana, and Spencer. We also settle, in the affirmative, a question of Talagrand on a deviation inequality for the discrete cube.

1 Introduction

Let $G = (V, \mathcal{E})$ be a finite, connected, undirected graph. We are interested in finding or estimating the optimal value of the constant $\sigma^2 = \sigma^2(G)$ satisfying the inequality

$$\mathbb{E}e^{t(f - \mathbb{E}f)} \leq e^{\sigma^2 t^2 / 2}, \quad \text{for all } t \in \mathbf{R}, \quad (1.1)$$

where f is an arbitrary Lipschitz function on V , and where the expectations are taken with respect to the normalized counting measure π on V . The Lipschitz property is taken with respect to a metric d associated with the graph, typically, d is the standard *graph distance*, given by the length of the shortest path between vertices. In this case, by f Lipschitz we mean that $|f(x) - f(y)| \leq 1$, whenever $\{x, y\} \in \mathcal{E}$. The quantity $\sigma^2(G)$ in (??), which we call the *subgaussian constant* of the graph, is akin to the so-called *spread constant*

$$c^2(G) = \sup_{f \in \mathcal{F}(G)} \text{Var } f,$$

*School of Mathematics, University of Minnesota, Minneapolis, MN 55455; research supported in part by EPSRC Visiting Fellowship.

†School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160; research supported in part by NSF Grant No. DMS-9803239, DMS-0100289.

‡School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160; research supported in part by NSF Grant No. DMS-9800351.

⁰Key words: isoperimetric inequalities, concentration, graph products, subgaussian constants, spread constants, log-Sobolev constants.

studied in [?] (cf. also [?]). Above, the supremum is taken over the family $\mathcal{F}(G)$ of all Lipschitz functions f on V , and Var is the variance of f with respect to π . Both constants quantify the deviation of a Lipschitz function f from its mean $\mathbb{E}f$. The advantage of the subgaussian constant is however the fact that it is responsible for the subgaussian tails of Lipschitz functions: it follows from (??) that, for all $h > 0$,

$$\pi\{f - \mathbb{E}f \geq h\} \leq e^{-h^2/(2\sigma^2)}. \quad (1.2)$$

The subgaussian constant is also the optimal value in a transport inequality on (V, d) ,

$$W_1^2(\pi, \nu) \leq 2\sigma^2 D(\nu||\pi) = 2\sigma^2 \sum_{x \in V} \nu(x) \log(\text{card}(V) \nu(x)),$$

relating the Kantorovich-Rubinstein (or Wasserstein) distance $W_1(\pi, \nu)$, the minimal “cost” needed in order to transport π to an arbitrary probability measure ν on V , to the informational divergence $D(\nu||\pi)$ also known as the relative entropy or Kullback-Liebler “distance” of ν with respect to π (cf. [?]).

The generic inequality, $\mathbb{E}|f - m(f)| \leq \sqrt{\text{Var } f}$, where $m(f)$ is a median of $f \in \mathcal{F}(G)$, implies together with (1.2) that

$$\pi\{f - m(f) \geq h\} \leq e^{-(h-c)^2/(2\sigma^2)}, \quad h \geq c = c(G). \quad (1.3)$$

These inequalities may further be connected to the isoperimetric problem on the graph where one minimizes the measure of $A^h = \{x \in V : d(x, A) \leq h\}$, the h -neighborhood of A for the metric d , given that the set A has a prescribed size. In particular, applying (1.3) to the Lipschitz function $f(x) = d(x, A)$ with an arbitrary $A \subset V$ such that $\pi(A) \geq 1/2$, and noting that, for such an f , $\mathbb{E}f \leq c(G) \leq \sigma(G)$, we arrive at a concentration type inequality

$$\pi\{x : d(x, A) \geq h\} \leq \exp\left\{-\frac{(h - \sigma)^2}{2\sigma^2}\right\}, \quad h \geq \sigma.$$

For h integer, the above is just

$$1 - \pi(A^h) \leq \exp\left\{-\frac{(h + 1 - \sigma)^2}{2\sigma^2}\right\}, \quad h + 1 \geq \sigma, \quad h = 0, 1, 2, \dots \quad (1.4)$$

It is in this way that we study concentration inequalities by trying to compute or bound from above the subgaussian constants.

A first motivation for such an approach is a question raised by Talagrand as part of his general investigations on isoperimetry in product probability spaces. (See the remarks following the proof of Corollary 2.2.3 in [?].) Is it indeed the case that for every $A \subset \Omega^n$ with $\mu(A) \geq 1/2$,

$$1 - \mu(A^h) \leq K e^{-2h^2/n}, \quad (1.5)$$

where $(\Omega^n = \Omega_1 \times \cdots \times \Omega_n, \mu = \mu_1 \times \cdots \times \mu_n)$ is an arbitrary product probability space, A^h is the enlargement of A with respect to the Hamming distance d on Ω^n , and K is a universal constant? Recall that the Hamming distance between $x \in \Omega^n$ and $A \subset \Omega^n$ is given by

$$d(x, A) = \min\{k : \exists y \in A; \text{card}\{i \leq n; x_i \neq y_i\} \leq k\}.$$

Talagrand also remarked that using certain more or less standard arguments (such as those used in [?]) it suffices to restrict to down-sets (also called hereditary sets) A in $\Omega^n = \{0, 1\}^n$ equipped with the product measure μ , where $\mu_i = \pi_p, i = 1, \dots, n$ is the Bernoulli measure with success probability p for $0 < p < 1$. Indeed, Steps 1 through 4 of the proof of Theorem 7 in [?] carry out such a reduction. (Recall also that $A \subset \{0, 1\}^n$ is a down-set if $x \in A$ and $y \leq x$ imply that $y \in A$. Here $y \leq x$, if for every i , $y_i \leq x_i$.) In the following (see the end of Section ??) we settle Talagrand's question in the affirmative. In fact, we show that the case of the discrete cube with down-sets follows easily from the results of Jogdeo-Samuels [?] and Bollobás-Leader [?].

A second crucial motivation is the simple observation that, for G^n the Cartesian product graph (equipped with an ℓ^1 -type metric),

$$\sigma^2(G^n) = n\sigma^2(G). \tag{1.6}$$

Therefore one may say that (in contrast to (??)) the property (??) tensorizes. Combining (??) and (??), we obtain concentration for the product graph in the form of an asymptotic isoperimetric inequality:

Proposition 1.1 *For all $A \subset V^n$ with $\pi^n(A) \geq 1/2$,*

$$1 - \pi^n(A^h) \leq \exp \left\{ -\frac{(h + 1 - \sigma\sqrt{n})^2}{2n\sigma^2} \right\}, \quad h + 1 \geq \sigma\sqrt{n}, \quad h = 0, 1, 2, \dots \tag{1.7}$$

Inequalities such as (??) and (??) are well-known in several situations. In the present work, we obtain inequalities of this type by computing or estimating the sub-gaussian constant for the following graphs:

1. The weighted two point space and the weighted discrete cube (section ??);
2. P_v : a path on v -vertices (v -path) and some generalizations (section ?? and appendix);
3. C_v : a cycle on v -vertices (v -cycle) (section ??);
4. (S_v, d_v) : the symmetric group with Hamming distance, and (S_v, ρ_v) : the symmetric group under transpositions (section ??);
5. K_v : the complete graph on v -vertices (v -clique) (section ??);
6. All of the above (section ??).

In particular, we find the exact constants $\sigma^2(P_v)$ and $\sigma^2(K_v)$, and our estimate for σ^2 of (S_v, ρ_v) is tight up to a multiplicative factor of 4, and that of (S_v, d_v) is tight up to a multiplicative factor of 16. Note that in general finding the extremal sets which minimize the size of A^h for $h = 1, 2, \dots$ is an extremely nontrivial problem; in particular, for the examples 4 above, this is still an open problem; while for K_v^n (rather than K_v) this has only been very recently solved by Harper (see [?]). Bollobás and Leader found the extremal sets for the n -dimensional grid graph P_v^n (see [?]), while for the n -dimensional discrete torus C_v^n , v even, this is due to Karakhanyan [?], Bollobás and Leader [?] and Riordan [?]. (To the best of our knowledge, for v odd this isoperimetric problem for C_v^n is still open.) One of the purposes of the present paper is to illustrate the fact that using a more functional analytic approach, it is possible to provide essentially best possible concentration inequalities without knowing the extremal sets. This is also compared to concentration inequalities obtained via log-Sobolev inequalities.

Since for every f , $\text{Var } f = \lim_{t \rightarrow 0} \text{E} \frac{e^{t(f - \text{E}f)} - 1}{t^2/2}$, $c^2(G) \leq \sigma^2(G)$. In general, this inequality is strict, it would however be worthwhile to describe those graphs for which $c^2(G)$ and $\sigma^2(G)$ coincide. We will show cases of equality for some of the above examples. To start with, introduce the functions

$$L_G(t, f) = \text{E} e^{t(f - \text{E}f)}, \quad L_G(t) = \sup_{f \in \mathcal{F}(G)} \text{E} e^{t(f - \text{E}f)}, \quad t \in \mathbf{R}.$$

Clearly, $L_G(-t) = L_G(t)$. More precise information is contained in L_G than in $\sigma^2(G)$, so it is reasonable to first try to find L_G and then to consider the analytical problem of computing the subgaussian constant via the relation

$$\sigma^2(G) = \sup_{t > 0} \frac{\log L_G(t)}{t^2/2}.$$

Note also that $L_{G^n}(t) = L_G(t)^n$, for all $t \in \mathbf{R}$, (see [?] for a proof) and thus $\sigma^2(G^n) = n\sigma^2(G)$.

2 Two point and Hamming spaces

The simplest graph of interest is the two point space $V = \{0, 1\}$ with uniform measure $\pi = \frac{\delta_0 + \delta_1}{2}$ which is a particular case of all of the above examples 2–4 when $v = 2$. It is well known, and due to Hoeffding, that for this graph $\sigma^2 = 1/4$ (cf. e.g., [?]). Moreover, since the seminal work of Harper [?], the solution to the isoperimetric problem is also known for $\{0, 1\}^n$ with the uniform measure. In the sequel, we will however need V to be equipped with an arbitrary probability measure $\mu = \mu_p$ assigning the mass $p \in (0, 1)$ to the point 1 and the mass $q = 1 - p$ to the point 0. The subgaussian constant for (V, μ) has been computed in [?], and for the sake of completeness it is also given here.

Proposition 2.1 *Given a function f on $\{0, 1\}$, the optimal value of σ^2 in the inequality*

$$\mathbb{E}e^{t(f-\mathbb{E}f)} \leq e^{\sigma^2 t^2/2}, \quad (2.1)$$

where $t \in \mathbf{R}$ is arbitrary, is given by

$$2\sigma^2 = \frac{p-q}{\log p - \log q} (f(1) - f(0))^2. \quad (2.2)$$

Here and throughout this section, the expectations and the other integral quantities, like the variance, on $\{0, 1\}$ are understood with respect to the Bernoulli measure $\mu_p = p\delta_1 + q\delta_0$. The discrete cube $\{0, 1\}^n$ is itself equipped with the product probability measure μ_p^n .

For $p = q = 1/2$, the value of $2\sigma^2$ is defined, as the limit as $p \rightarrow 1/2$, to be $(f(1) - f(0))^2/2$. This value maximizes the right hand side of (??) over all p 's. In particular,

$$\mathbb{E}e^{t(f-\mathbb{E}f)} \leq e^{t^2/8},$$

for all $t \in \mathbf{R}$, as soon as $0 \leq f \leq 1$.

Introducing the entropy functional

$$\text{Ent } g = \mathbb{E}g \log g - \mathbb{E}g \log \mathbb{E}g, \quad g \geq 0,$$

the proof of Proposition ?? relies on:

Lemma 2.2 *The optimal constant $c = c_p$ in the inequality*

$$c \text{Var } g \leq \mathbb{E}g \text{Ent } g, \quad (2.3)$$

where g is an arbitrary nonnegative function on $\{0, 1\}$, is given by

$$c_p = pq \frac{\log p - \log q}{p - q}.$$

When $p = q = 1/2$, the value of c_p becomes $\lim_{p \rightarrow 1/2} c_p = 1/2$. As we will also see, (??) becomes equality for the function g such that $g(0) = p/q$, $g(1) = q/p$.

The inequality (??) can be viewed as a converse to the general inequality $\text{Var } g \geq \mathbb{E}g \text{Ent } g$ which holds on an arbitrary probability space. As for the constant, it is the same as the the optimal c in

$$\text{Ent } g^2 \leq c \mathbb{E}|\nabla g|^2,$$

where $|\nabla g| = |g(1) - g(0)|$ (cf. [?], [?]); however, the relationship with (??) is not that transparent.

Proof Set $g(1) = a \geq 0$, $g(0) = b \geq 0$, and without loss of generality assume that in (??)

$$\mathbb{E}g = pa + qb = 1, \quad (2.4)$$

so that (??) takes the form

$$cpq(b-a)^2 \leq pa \log a + qb \log b.$$

By symmetry, we also assume that $p \geq 1/2$, $a \leq b$, so that $a \leq 1 \leq b \leq 1/q$. Let

$$\varphi(b) = \frac{pa \log a + qb \log b}{(b-a)^2}, \quad 1 \leq b \leq 1/q,$$

where a depends on b according to (??). At $b = 1$ (which is the only case where $a = b$) p is extended by continuity to be $pq/2$. We prove below that φ has only one point of minimum $b = p/q$ (and thus $a = q/p$). To this end, it will suffice to show that this point is the only solution of the equation $\varphi'(b) = 0$ in the interval $1 < b < 1/q$ when $p > 1/2$. Differentiating in b and recalling that $a'(b) = -q/p$, we find:

$$\varphi'(b) = \frac{pq(b-a)(\log b - \log a) - 2(pa \log a + qb \log b)}{p(b-a)^3}.$$

It is easy to verify that $\varphi'(p/q) = 0$. Now let

$$\psi(b) = pq(b-a)(\log b - \log a) - 2(pa \log a + qb \log b), \quad 1 \leq b \leq 1/q,$$

so that $\varphi'(b) = 0 \iff \psi(b) = 0$, for $b > 1$. Two more differentiations give:

$$\psi'(b) = q \left[\left(\frac{1}{a} + \log a \right) - \left(\frac{1}{b} + \log b \right) \right], \quad \psi''(b) = \frac{q}{(pab)^2} (b-1)(2qb-1).$$

Consequently, ψ is strictly concave in $[1, 1/(2q)]$ and strictly convex in $[1/(2q), 1/q]$. In addition, $\psi(1) = \psi'(1) = 0$. Therefore, for some $b_0 \in (1, 1/q)$, this function is strictly decreasing on the interval $[1, b_0]$ and strictly increasing on $[b_0, 1/q]$. In particular, the equation $\psi(b) = 0$ has at most one solution on $(1, 1/q)$ and, as we know, this solution exists and is given by $b = p/q$. When $p = q = 1/2$, the interval of concavity of ψ degenerates to the point $b = 1 = p/q$. Lemma ?? is proved.

Proof of Proposition ?? We may assume that $Ef = 0$ and that $f(0) \neq f(1)$ so that a priori $\sigma > 0$. The entropy functional has the following well-known general representation:

$$\text{Ent } g = \sup\{Eug : Ee^u \leq 1\}.$$

Hence, the inequality $Ee^u \leq 1$ is equivalent to $Eug \leq \text{Ent } g$, where $g \geq 0$ is arbitrary. Applying this to $u = tf - \sigma^2 t^2/2$, we see that (??) is equivalent to

$$E(tf - \sigma^2 t^2/2)g \leq \text{Ent } g, \quad g \geq 0, \quad t \in \mathbf{R}. \quad (2.5)$$

For a fixed g with $Eg > 0$, the above left hand side is maximized for

$$t = \frac{Efg}{\sigma^2 Eg}, \quad (2.6)$$

and (??) becomes

$$(Efg)^2 \leq 2\sigma^2 Eg \text{Ent } g, \quad g \geq 0.$$

But, on the two point space, since $Ef = 0$, we have $(Efg)^2 = pq(f(1) - f(0))^2 \text{Var } g$. Hence, the above inequality can be rewritten as

$$\frac{pq(f(1) - f(0))^2}{2\sigma^2} \text{Var } g \leq Eg \text{Ent } g,$$

where g is an arbitrary nonnegative function on $\{0, 1\}$. It remains to apply Lemma ?? to obtain Proposition ??.

We may now summarize:

Proposition 2.3 *Let $0 < p < 1$. For the discrete cube $V = \{0, 1\}^n$ equipped with the product measure μ_p^n , the subgaussian constant is given by*

$$\sigma^2 = \frac{n(p - q)}{2(\log p - \log q)}.$$

Applying Proposition ?? to the Lipschitz function $f(x) = x_1 + \dots + x_n$ on V , implies via (??) a subgaussian deviation inequality for the number S_n of successes in n independent Bernoulli trials with success probability p :

$$\Pr \left\{ \frac{S_n - np}{\sqrt{n}} \geq h \right\} \leq \exp \left\{ -\frac{\log p - \log q}{p - q} h^2 \right\} \leq \exp\{-2h^2\}, \quad h \geq 0. \quad (2.7)$$

We now return to Talagrand's question mentioned in the introductory section and further study concentration on the Hamming space. Before removing the question mark after (??), and for the sake of completeness we first state the results of Jogdeo-Samuels and Bollobás-Leader we need. Theorem 3.2 and Corollary 3.1 of [?] assert the following: Consider n independent Bernoulli trials, the i th trial with success probability $0 < p_i < 1$. Then

- (i) if the mean number of successes is an integer k then the median is also k ,
- (ii) if the mean number of successes is between the integers k and $k + 1$ then the median is either k or $k + 1$.

On the other hand, one of the main results (Corollary 5) of [?] asserts that for any down-set A in the discrete cube $\{0, 1\}^n$, if $\mu_p^n(A) \geq \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}$, then

$$\mu_p^n(A^h) \geq \sum_{i=0}^{k+h} \binom{n}{i} p^i (1-p)^{n-i}. \quad (2.8)$$

Proposition 2.4 For $n \geq 1$, let $\mu = \mu_p^n$, with $\mu_p(\{1\}) = p$ and $\mu_p(\{0\}) = 1 - p$, for $0 < p < 1$. Then for every $A \subset \{0, 1\}^n$, with $\mu(A) \geq 1/2$, and for every integer $h \geq 1$, we have

$$1 - \mu(A^h) \leq K e^{-2h^2/n}, \quad (2.9)$$

where $K > 0$ is an absolute constant, and $A^h = \{x \in \{0, 1\}^n : d(x, A) \leq h\}$, d being the Hamming distance.

Proof. As indicated before, we can take A to be a down-set with $\mu(A) \geq 1/2$. In light of the results of [?], observe that

$$\mu(A) \geq 1/2 \geq \sum_{i=0}^{\lfloor np \rfloor - 1} \binom{n}{i} p^i (1-p)^{n-i},$$

also assuming that $\lfloor np \rfloor > 0$. (The second inequality follows from the fact that a median is either $\lfloor np \rfloor$ or $\lfloor np \rfloor + 1$.) Even if $\lfloor np \rfloor = 0$, it is still true that $\mu(A) \geq (1-p)^n$, since every down-set contains $(0, \dots, 0)$. Now using (??) and also the fact that only the values $h \leq n - 1$ need to be considered, we conclude that

$$\begin{aligned} 1 - \mu(A^h) &\leq \sum_{i=\lfloor np \rfloor + h}^n \binom{n}{i} p^i (1-p)^{n-i} \\ &= \Pr\{S_n \geq \lfloor np \rfloor + h\} \leq K e^{-2h^2/n}, \end{aligned}$$

where the last inequality is standard and also easily follows from the second subgaussian inequality in (??). This proves the result with $K = e^4$. Actually, using the first inequality in (??) and for $p \neq \frac{1}{2}$, the exponent -2 can be improved to $-\frac{\log p - \log q}{p - q}$.

Since for $\{0, 1\}^n$, with $p = 1/2$, we have $\sigma^2 = n/4$, (??) can be rewritten as

$$1 - \mu(A^h) \leq K e^{-h^2/(2\sigma^2)}, \quad (2.10)$$

with $K = e^4$. Therefore, one may wonder whether or not this last inequality remains valid for an arbitrary (weighted) graph, or equivalently whether or not the spread constant $c(G)$ in the deviation inequality (??), can be removed at the expense of an absolute multiplicative constant K . A positive answer to this query would also easily settle Talagrand's question since the subgaussian constant of any Hamming space is at most $n/4$. However, the weighted discrete cube (of increasing dimension n) provides a counterexample to such an intriguing question (and although the particular case $n = 1$ is affirmatively solved by Proposition ??).

Indeed, assume $(1-p)^n = 1/2$ so that p is of order $\frac{\log 2}{n} + O\left(\frac{1}{n^2}\right)$, as $n \rightarrow \infty$. By Proposition ??, the inequality (??), for which we are seeking a counterexample,

becomes

$$1 - \mu(A^h) \leq K \exp \left\{ -\frac{\log p - \log q}{n(p - q)} h^2 \right\}.$$

Now, the only set of μ -measure $1/2$ is the one point set $A = \{(0, \dots, 0)\}$. For $h = n - 1$, we thus have $A^h = \{0, 1\}^n \setminus \{(1, \dots, 1)\}$ and the last inequality simplifies to

$$\frac{\log p - \log q}{p - q} \left(1 - \frac{1}{n}\right)^2 - \log \frac{1}{p} \leq \frac{\log K}{n}. \quad (2.11)$$

But, a simple Taylor expansion shows that the main term on the left of (??) is $2p \log \frac{1}{p}$ which is of order $\frac{2 \log 2 \log n}{n}$, and so cannot be bounded by the right hand side of (??).

It is standard that (??), for $p = \frac{1}{2}$, implies

$$\mu\{f - m(f) \geq h\} \leq K e^{-h^2/2n}, \quad (2.12)$$

for any f , Lipschitz (with constant 1) with respect to the Hamming distance on $\{0, 1\}^n$ and again $K = e^4$. It is thus also natural to wonder if this last inequality can be sharpened by replacing $1/2n$ with $2/n$, possibly worsening the absolute constant K . This would then match the deviation inequality from the mean, up to the universal constant K . On $\{0, 1\}$ this is indeed immediately true from (??) (and with $K = e^{11/2}$), and on $\{0, 1\}^n$ for p small ($p \leq 1/(e^8 + 1)$ will do) or p close to 1 this is also true by Proposition ???. Now, in general, $e^{-2h^2/n} \leq e^5 e^{-2(h+1)^2/n}$ for all $h \leq n$, $n \geq 2$. Thus (??) admits a small improvement:

$$1 - \mu(A^{h-1}) \leq e^9 e^{-2h^2/n},$$

for all $h \geq 1$, integer and all $A \subset \{0, 1\}^n$ with $\mu(A) \geq 1/2$ where μ is the weighted product probability measure as in Proposition ??. Given a Lipschitz function f on V , applying the above inequality to $A = \{x : f(x) \leq m(f)\}$, and since A^{h-1} is contained in $\{f \leq m(f) + (h - 1)\}$, gives

$$\mu\{f - m(f) \geq h\} \leq e^9 e^{-2h^2/n}. \quad (2.13)$$

3 v -path

v -path is the graph $G = P_v$ with $V = \{1, 2, \dots, v\}$ where $\{x, x + 1\}$ ($x = 1, \dots, v - 1$) are the only pairs of connected vertices. Let us find the function L_G in this case, assuming that $v \geq 2$. An element of $\mathcal{F}(G)$ is an arbitrary function f on V such that

$$|f(x + 1) - f(x)| \leq 1, \quad \text{for all } x = 1, \dots, v - 1.$$

We want to show that, whenever $t \in \mathbf{R}$, in the class $\mathcal{F}(G)$ the value of $\mathbb{E}e^{t(f-\mathbb{E}f)}$ is maximized for the identity function

$$f^*(x) = x.$$

In fact, there is a general principle involving this statement:

Proposition 3.1 *Let μ be a Borel probability measure on \mathbf{R} such that the half-axes $(-\infty, x]$ are extremal in the isoperimetric problem for μ , i.e., for all $p \in (0, 1)$ and $h > 0$, the infimum*

$$\inf\{\mu(A + (-h, h)) : A \text{ Borel}, \mu(A) \geq p\}$$

is attained at the half-axis $A = (-\infty, x]$, for some $x \in \mathbf{R}$. Then, for any Lipschitz function f on \mathbf{R} (with Lipschitz constant at most 1), and for all $t \in \mathbf{R}$,

$$\mathbb{E}_\mu e^{t(f-\mathbb{E}_\mu f)} \leq \mathbb{E}_\mu e^{t(f^*-\mathbb{E}_\mu f^*)}. \quad (3.1)$$

The proof of this proposition and of the following one are given at the end of this section. It should be clear that the uniform measure $\mu = \pi$ on V satisfies the condition of Proposition ???. Therefore, $L_G(t) = L_G(t, f^*)$, and since $\mathbb{E}_\pi f^* = (v+1)/2$, it follows that

$$\begin{aligned} L_G(t, f^*) &= \frac{1}{v} \sum_{x=1}^v e^{tx} e^{-t(v+1)/2} = e^t \frac{e^{vt} - 1}{v(e^t - 1)} e^{-t(v+1)/2} \\ &= \frac{e^{vt/2} - e^{-vt/2}}{v(e^{t/2} - e^{-t/2})} = \frac{\text{sh}(vt/2)}{v \text{sh}(t/2)}. \end{aligned}$$

Thus:

Proposition 3.2 *For any Lipschitz function f on the v -path P_v and for all $t \in \mathbf{R}$,*

$$\mathbb{E}e^{t(f-\mathbb{E}f)} \leq \frac{\text{sh}(vt/2)}{v \text{sh}(t/2)},$$

where the expectations are with respect to the uniform measure. In particular,

$$\sigma^2(P_v) = \frac{v^2 - 1}{12}.$$

Remark 3.3 For every probability measure μ on \mathbf{R} and every Lipschitz function f on \mathbf{R} ,

$$\text{Var}_\mu(f) = \frac{1}{2} \iint_{\mathbf{R}^2} (f(x) - f(y))^2 d\mu(x)d\mu(y) \leq \frac{1}{2} \iint_{\mathbf{R}^2} (x - y)^2 d\mu(x)d\mu(y) = \text{Var}_\mu(f^*).$$

Therefore, the spread constant of the v -path is

$$c^2(P_v) = \text{Var}_\pi(f^*) = \frac{1}{v} \sum_{x=1}^v x^2 - \left(\frac{v+1}{2}\right)^2 = \frac{v^2 - 1}{12}.$$

Thus, $c^2(P_v) = \sigma^2(P_v)$. According to (??)-(??), we obtain:

Proposition 3.4 *Let P_v^n be the n -th power of the v -path P_v with the uniform measure π^n on the set of vertices V^n . For all $A \subset V^n$, such that $\pi^n(A) \geq 1/2$,*

$$1 - \pi^n(A^h) \leq \exp \left\{ -\frac{6 \left(h + 1 - \sqrt{\frac{v^2-1}{12} n} \right)^2}{n(v^2-1)} \right\}, \quad h + 1 \geq \sqrt{\frac{v^2-1}{12} n}, \quad h = 0, 1, 2, \dots \quad (3.2)$$

As shown by Bollobás and Leader, a result stronger than Proposition ?? is true for the n -th power of *any* graph G . (Their result has an h in place of $h + 1 - \sigma\sqrt{n}$ in the right hand side of (??).) Using Corollary 14 of [?], we can also write Proposition ?? for arbitrary G^n . For completeness, let us state their Corollary 14 here. (In plain words, it states that G^n has the worst isoperimetry, or minimum vertex boundary, when G is a path.) Let $V^n = \{0, 1, \dots, v-1\}^n$ denote the set of vertices of the n -th power of a v -path, let

$$B_v^{(n)}(r) = \left\{ x \in V^n : \sum_i x_i \leq r \right\}, \quad r = 0, 1, \dots,$$

and let $b_v^{(n)}(r) = |B_v^{(n)}(r)|$ (here and below, $|\cdot|$ denotes cardinality). Then

Lemma 3.5 ([?]) *Let G_1, \dots, G_n be arbitrary connected graphs, each on v vertices, and let \square denote Cartesian product. Then for all $A \subset G_1 \square G_2 \cdots \square G_n$, with $|A| \geq b_v^{(n)}(r)$,*

$$|A^h| \geq b_v^{(n)}(r + h), \quad \text{for all } h.$$

Together Proposition ?? and Lemma ?? imply as claimed:

Proposition 3.6 *Let $G^n = G_1 \square \cdots \square G_n$. Then for all $A \subset G^n$ such that $\pi^n(A) \geq 1/2$,*

$$1 - \pi^n(A^h) \leq \exp \left\{ -\frac{6 \left(h + 1 - \sqrt{\frac{v^2-1}{12} n} \right)^2}{n(v^2-1)} \right\}, \quad h + 1 \geq \sqrt{\frac{v^2-1}{12} n}, \quad h = 0, 1, 2, \dots$$

Perhaps it is to be remarked that Bollobás and Leader obtain their stronger version of Proposition ?? by actually finding the extremal sets minimizing $\pi(A^h)$ for all $h \geq 1$, whereas the proof (of our weaker result) is simpler since we derive Proposition ?? directly, with a functional-analytic approach, without having to find the extremal sets.

Finally, we note that Proposition ?? guarantees that the v -path is extremal for the subgaussian constant (and hence the spread constant) among all graphs on v vertices:

Corollary 3.7 *For any connected undirected graph G on v vertices, $\sigma^2(G) \leq \frac{v^2-1}{12}$.*

Proof The point is that any function which is Lipschitz with respect to an arbitrary G on v vertices is also Lipschitz with respect to P_v . To be formal, let $t \in \mathbf{R}$ be fixed, and let g be Lipschitz with respect to $G = (V, E)$ with $|V| = v$ vertices. Without loss of generality let $\min_{x \in V} g(x) = 1$. Then clearly we may order g as follows:

$$1 = g(\tau_1) \leq g(\tau_2) \leq \cdots \leq g(\tau_v),$$

where τ is an appropriate permutation of V , and furthermore, $g(\tau_i) \leq i$. Thus for each $t \in \mathbf{R}$, we may indeed view g as a Lipschitz function on the path on $\{1, 2, \dots, v\}$. Now the result follows from Proposition ??, which establishes the extremality of the identity function f^* on P_v (see in particular (??)).

Note that it also follows now that the v -path is extremal for $c^2(G)$ among all graphs on v -vertices – one can either use the above argument or simply appeal to the fact that $c^2(G) \leq \sigma^2(G)$ for all G .

Remark 3.8 Noga Alon pointed out to us that the above argument extends to implying that the maximum over all 2-connected graphs¹ G of $\sigma^2(G)$ is achieved by C_v , the v -cycle.

We can now pass to the proofs of Propositions ?? and ??. As shown in [?], the extremal property of the half-axes in the isoperimetric problem for μ on the real line implies that μ has a finite exponential moment. In particular, f^* and thus all Lipschitz functions on \mathbf{R} are μ -integrable. Hence, both sides of the inequality (??) are well-defined. First we establish this inequality for monotone Lipschitz functions.

Lemma 3.9 *Let μ be a probability measure on \mathbf{R} with finite first moment, i.e., $E_\mu |f^*| = \int_{\mathbf{R}} |x| d\mu(x) < +\infty$. Then, (??) holds true for any non-decreasing Lipschitz function f on \mathbf{R} and for all $t \geq 0$. If the measure μ is symmetric about a point, then (??) holds for all $t \in \mathbf{R}$.*

In general (??) is not true for all Lipschitz functions f on \mathbf{R} , even if μ is symmetric. A simple counterexample to (??) is given by the function $f(x) = |x|$ with respect to measures of the form $\mu = p\delta_x + p\delta_{-x} + q\delta_0$ with sufficiently small p and large x . Thus, in order to obtain an extremal property for the function f^* in the class of all Lipschitz functions, an extra condition on μ is required. Such an extra condition, based on the extremal property of the half-axes as stated in Proposition ??, will be used.

¹Recall that a graph is 2-connected if the removal of any single vertex still leaves the graph connected.

Proof of Lemma ?? We use the following version of the well-known functional representation of the entropy: for every measurable function h defined on some probability space,

$$\log \mathbb{E}e^h = \sup_{\mathbb{E}g=1} [\mathbb{E}gh - \mathbb{E}g \log g], \quad (3.3)$$

where the sup is taken over all measurable non-negative (and for simplicity bounded) functions g with $\mathbb{E}g = 1$. Clearly, this supremum is attained at $g = e^h/\mathbb{E}e^h$. Thus, when the measure μ and the functions g, h are considered on the real line \mathbf{R} , and h is non-decreasing on \mathbf{R} , the extremal g is non-decreasing, as well. Hence in this case, it suffices to restrict ourselves to non-decreasing functions g in (??). In particular, applying (??) to the nondecreasing function $h = t(f - \mathbb{E}_\mu f)$, we have a representation in terms of the covariances $\text{cov}_\mu(f, g) = \mathbb{E}fg - \mathbb{E}f\mathbb{E}g$:

$$\log \mathbb{E}_\mu e^{t(f - \mathbb{E}_\mu f)} = \sup_{\mathbb{E}g=1} [t \text{cov}(f, g) - \mathbb{E}g \log g], \quad g \text{ nondecreasing.} \quad (3.4)$$

Now use

$$\text{cov}(f, g) = \frac{1}{2} \iint_{\mathbf{R}^2} (f(x) - f(y))(g(x) - g(y)) d\mu(x)d\mu(y).$$

Moreover, if both f and g are non-decreasing and if f is Lipschitz, then for all $x, y \in \mathbf{R}$,

$$(f(x) - f(y))(g(x) - g(y)) \leq (x - y)(g(x) - g(y)) = (f^*(x) - f^*(y))(g(x) - g(y)).$$

Hence, $\text{cov}_\mu(f, g) \leq \text{cov}_\mu(f^*, g)$, and thus, by (??), $\log \mathbb{E}_\mu e^{t(f - \mathbb{E}_\mu f)} \leq \log \mathbb{E}_\mu e^{t(f^* - \mathbb{E}_\mu f^*)}$. This proves Lemma ?? (the second statement is trivial since then $f^* - \mathbb{E}_\mu f^*$ and $\mathbb{E}_\mu f^* - f^*$ are identically distributed).

Clearly, Lemma ?? can equivalently be formulated as follows: Let ξ and η be integrable random variables on some probability space (Ω, μ) . If, for some non-decreasing Lipschitz function f from \mathbf{R} to \mathbf{R} , η and $f(\xi)$ are identically distributed, then, for all $t \geq 0$,

$$\mathbb{E}_\mu e^{t(\eta - \mathbb{E}_\mu \eta)} \leq \mathbb{E}_\mu e^{t(\xi - \mathbb{E}_\mu \xi)}. \quad (3.5)$$

In order to check the assumptions of this statement and thus get (??), one may use the following characterization proved in [?] (Proposition 2.6 therein):

Lemma 3.10 *Given two random variables ξ and η on (Ω, μ) , the existence of a non-decreasing Lipschitz map f from \mathbf{R} to \mathbf{R} such that η and $f(\xi)$ are identically distributed is equivalent to the inequality*

$$\mu\{\eta \leq m_p(\eta) + h\} \geq \mu\{\xi \leq m_p(\xi) + h\}, \quad (3.6)$$

holding for all $p \in (0, 1)$ and $h > 0$, and where m_p denotes the minimal quantile (of order p) of a random variable.

Proof of Proposition ?? The extremal property of the half-axes implies that

- 1) μ is symmetric about some point;
- 2) for every Lipschitz function on \mathbf{R} , there exists a non-decreasing Lipschitz function on \mathbf{R} with the same distribution (with respect to μ).

The property 2) is stronger than 1) which in turn follows from 2) applied to the function $f = -f^*$. In order to derive 2) from the extremality of the half-axes, we use Lemma ???. Indeed, given a Lipschitz function f on \mathbf{R} , let

$$A_p = \{x \in \mathbf{R} : f(x) \leq m_p(f)\}, \quad 0 < p < 1,$$

where m_p is the minimal quantile of f with respect to μ . By assumption, there exists $x \in \mathbf{R}$ such that $\mu((-\infty, x]) \geq p$ and $\mu(A_p + (-h, h)) \geq \mu((-\infty, x + h))$, for all $h > 0$. The minimal value of x with the property $\mu((-\infty, x]) \geq p$ is $x = m_p(f^*)$ in which case $\mu((-\infty, x + h)) = \mu\{f^* - m_p(f^*) < h\}$. Thus,

$$\mu(A_p + (-h, h)) \geq \mu\{f^* - m_p(f^*) < h\}.$$

But since f is Lipschitz, $A_p + (-h, h) \subset \{x \in \mathbf{R} : f(x) < m_p(f) + h\}$. Therefore,

$$\mu\{f - m_p(f) < h\} \geq \mu\{f^* - m_p(f^*) < h\}, \quad h > 0, \quad 0 < p < 1.$$

So, (??) and thus (??) hold for $\eta = f$ and $\xi = f^*$. Proposition ??? is proved.

Proof of Proposition ??? It remains to show that the optimal value of $\sigma^2 = \sigma^2(P_v)$ in

$$\frac{\text{sh}(vt/2)}{v \text{sh}(t/2)} \leq e^{\sigma^2 t^2/2}, \quad t \in \mathbf{R}, \quad (3.7)$$

is $\sigma^2(P_v) = \frac{v^2 - 1}{12}$. Taking logarithm of both sides in (??) and setting $s = t/2$, we need to find the optimal constant σ^2 satisfying the inequality

$$\varphi(s) = \log(\text{sh}(vs)) - \log(\text{sh}(s)) - 2\sigma^2 s^2 \leq \log v, \quad s \geq 0, \quad (3.8)$$

where by continuity, $\varphi(0) = \log v$. Next, for $s > 0$,

$$\begin{aligned} \varphi'(s) &= v \frac{\text{ch}(vs)}{\text{sh}(vs)} - \frac{\text{ch}(s)}{\text{sh}(s)} - 4\sigma^2 s, \\ \varphi''(s) &= -\frac{v^2}{\text{sh}^2(vs)} + \frac{1}{\text{sh}^2(s)} - 4\sigma^2, \\ \frac{1}{2} \varphi'''(s) &= \frac{v^3 \text{ch}(vs)}{\text{sh}^3(vs)} - \frac{\text{ch}(s)}{\text{sh}^3(s)}. \end{aligned}$$

Using Taylor's expansion for the hyperbolic functions, we easily find that

$$\varphi'(0^+) = \lim_{s \rightarrow 0^+} \varphi'(s) = 0, \quad \varphi''(0^+) = \frac{v^2 - 1}{3} - 4\sigma^2.$$

Therefore, (??) implies $\varphi''(0^+) \leq 0$, i.e., for the optimal σ^2 in (??), we have $\sigma^2 \geq \frac{v^2 - 1}{12}$. It remains to show that $\sigma^2 = \frac{v^2 - 1}{12}$ satisfies (??). To do so, it is enough to show that φ is concave, i.e., $\varphi''(s) \leq 0$, for all $s > 0$. Since $\varphi''(0^+) = 0$, it is sufficient to show that, for all $s > 0$, $\varphi'''(s) \leq 0$, i.e.,

$$\frac{(vs)^3 \text{ch}(vs)}{\text{sh}^3(vs)} \leq \frac{s^3 \text{ch}(s)}{\text{sh}^3(s)}.$$

Clearly, this inequality follows if the function $\psi(s) = \frac{s^3 \text{ch}(s)}{\text{sh}^3(s)}$ is non-increasing in $s > 0$.

That is, if the function

$$\theta(s) = \log \psi(s) = 3 \log s + \log \text{ch}(s) - 3 \log \text{sh}(s),$$

is non-increasing in $s > 0$. So, let us verify that

$$\theta'(s) = \frac{3}{s} + \frac{\text{sh}(s)}{\text{ch}(s)} - \frac{3 \text{ch}(s)}{\text{sh}(s)} = \frac{3}{s} - \frac{1 + \text{ch}^2(s)}{\text{sh}(s)\text{ch}(s)} \leq 0,$$

which can be rewritten as

$$u(s) = 3 \text{sh}(s)\text{ch}(s) - s(1 + \text{ch}^2(s)) \leq 0, \quad s \geq 0. \quad (3.9)$$

Since $u(0) = 0$ and $u'(s) = 4 \text{sh}(s) (\text{sh}(s) - s \text{ch}(s)) \leq 0$, u is non-increasing in $s \geq 0$. This proves (??) and thus Proposition ??.

4 v -cycle and discrete torus

The v -cycle $G = C_v$ can be viewed as the subset V of the complex plane \mathbf{C} given by

$$V = \{x_k = e^{2\pi i k/v} : k = 0, 1, \dots, v-1\}$$

where $\{x_k, x_{k+1}\}$ ($k = 0, 1, \dots, v-1$) are the only pairs of connected points (with the agreement that $x_v = x_0$). For example, a 2-cycle is also a 2-path, but for $v \geq 3$, a v -cycle is not a v -path. The graph distance on the v -cycle is up to a constant the geodesic distance on V considered as a subset of the unit circle $S^1 \subset \mathbf{R}^2$. Thus, an element of $\mathcal{F}(G)$ is an arbitrary function f on V such that

$$|f(x_k) - f(x_{k-1})| \leq 1, \quad (4.1)$$

for all $k = 1, \dots, v$. In analogy with the v -path, one can suggest the following.

Conjecture 4.1 *In the class $\mathcal{F}(G)$ of all Lipschitz function f on the v -cycle $G = C_v$, for all $t \in \mathbf{R}$, the value of*

$$L_G(t, f) = \mathbb{E} e^{t(f - \mathbb{E}f)}$$

is maximized for the function $f(x) = d(x, x_0)$ or $f(x) = -d(x, x_0)$, $x \in V$ (the expectations are with respect to the normalized counting measure π on V).

We verified that the conjecture is true for $v \leq 4$; we present below a proof for the case $v = 3$, and compute the corresponding constant $\sigma^2(C_3)$. Moreover, see Remark ?? below which observes that the conjecture is true when v is even. As a result, this will allow us to compute the corresponding constants $\sigma^2(C_v)$.

First, let us simplify the problem of maximization of the functional $f \rightarrow L_G(t, f)$, $f \in \mathcal{F}(G)$. This functional is translation invariant, $L_G(t, f + c) = L_G(t, f)$, so we can always assume that $f(x_0) = 0$. Denote the set of such Lipschitz functions by $\mathcal{F}_0(G)$. Next, this functional is clearly convex. Therefore, since $\mathcal{F}_0(G)$ is a convex and compact set, $L_G(t, f)$ is maximized for some extremal function f of $\mathcal{F}_0(G)$. In order to describe the extremal functions, we associate every function $f \in \mathcal{F}_0(G)$, according to (??), with the vector

$$y = (y_1, \dots, y_v) \in [-1, 1]^v, \quad y_k = f(x_k) - f(x_{k-1}), \quad k = 1, \dots, v,$$

such that

$$y_1 + \dots + y_v = 0. \tag{4.2}$$

Thus, the map $T : y \rightarrow f$ allows us to identify $\mathcal{F}_0(G)$ with the intersection M_0 of the cube $[-1, 1]^v$ with the hyperplane defined by (??). But, as easily seen (and proved), when $v = 2n$ is even, the extremal points of M_0 are the sequences

$$y = (\pm 1, \dots, \pm 1) \tag{4.3}$$

with the number of pluses equal to the number of minuses ($= n$). When $v = 2n + 1$ is odd, the extremal points of M_0 are the sequences of the form

$$y = (\pm 1, \dots, \pm 1, 0, \pm 1, \dots, \pm 1) \tag{4.4}$$

also with the number of pluses equal to the number of minuses ($= n$), but with 0 at some place. Thus, in order to maximize $L_G(t, f)$ in the class $\mathcal{F}(G)$, it suffices to consider the functions f on V with the property that

$$f(x_k) - f(x_{k-1}) = \pm 1, \tag{4.5}$$

for all $k = 1, \dots, v$ in case $v = 2n$, and in case $v = 2n + 1$, with the property that (??) holds for all $k = 1, \dots, v$ except one value of k for which $f(x_k) - f(x_{k-1}) = 0$, with in addition in both cases the number of +1 equal to n .

Now, denote by S the shift operator on \mathbf{R}^v : $(Sy)_k = y_{k+1}$ with the agreement that $y_{v+1} = y_1$. Then, $T(Sy) = T(y) + c$, so $L_G(t, T(Sy)) = L_G(t, T(y))$. Therefore, maximizing $L_G(t, T(y))$ among all $y \in M_0$, we can restrict ourselves to extremal sequences $y \in \text{ex}(M_0)$ as in (??)–(??) whose transforms by shift operation form the whole set $\text{ex}(M_0)$.

For example, when $v = 3$, up to the shift transformation there exist only two extremal sequences

$$(0, 1, -1) \quad \text{and} \quad (0, -1, 1),$$

and the rest of the sequences $(1, -1, 0)$, $(-1, 1, 0)$, $(1, 0, -1)$, $(-1, 0, 1)$ can be obtained from the first two sequences using the shift operator possibly applied twice. The sequence $y = (0, 1, -1)$ corresponds to the function $f = T(y)$ such that $f(x_0) = 0$, $f(x_1) = 1$, $f(x_2) = 0$, that is $f(x) = 1 - d(x, x_1)$ which, up to an additive constant, has the same distribution as the function $-d(x, x_0)$. The second sequence $y = (0, -1, -1)$ corresponds to the function $f = T(y)$ such that $f(x_0) = 0$, $f(x_1) = -1$, $f(x_2) = 0$, that is, $f(x) = d(x, x_1) - 1$ which up to a summand has the same distribution as the function $d(x, x_0)$. This proves the conjecture in the case $v = 3$:

$$L_G(t) = \max\{L_G(t, f), L_G(-t, f)\}, \quad \text{where } f(x) = d(x, x_0), \quad x \in V. \quad (4.6)$$

$$\text{Now, } \mathbb{E}f = \frac{f(x_0) + f(x_1) + f(x_2)}{3} = \frac{2}{3},$$

$$L_G(t, f) = \mathbb{E}e^{tf} e^{-2t/3} = \frac{1 + 2e^t}{3} e^{-2t/3} = \frac{e^{-2t/3} + 2e^{t/3}}{3}.$$

As is easily seen, $L_G(t, f) \leq L_G(-t, f)$, for all $t \geq 0$, in which case the function $f(x) = -d(x, x_0)$ maximizes $L_G(t, f)$, while for $t \leq 0$, the function $f(x) = d(x, x_0)$ maximizes $L_G(t, f)$. We can now summarize:

Proposition 4.2 *For the 3-cycle $G = C_3$, we have*

$$L_G(t) = \frac{e^{2|t|/3} + 2e^{-|t|/3}}{3}, \quad t \in \mathbf{R}.$$

In particular,

$$\sigma^2(C_3) = \frac{1}{6 \log 2}.$$

Proof It only remains to find the optimal constant σ^2 satisfying the inequality

$$L_G(t) = \frac{1}{3} e^{2t/3} + \frac{2}{3} e^{-t/3} \leq e^{\sigma^2 t^2/2}, \quad t \geq 0.$$

Note that $L_G(t) = \mathbb{E}e^{t(\xi - \mathbb{E}\xi)}$ where ξ is a Bernoulli random variable taking the values 1 and 0 with probabilities $p = 1/3$ and $q = 2/3$, respectively. But by Proposition ??,

$$2\sigma^2 = \frac{p - q}{\log p - \log q}. \quad (4.7)$$

When $p = 1/3$, $2\sigma^2 = \frac{1}{3 \log 2}$. Proposition ?? is proved.

Remark 4.3 The functional $f \rightarrow \text{Var } f$ is convex, hence it attains its maximum in $\mathcal{F}(G)$ at the function $f(x) = d(x, x_0)$ (since $\text{Var}(-f) = \text{Var } f$ and since the functions f and $-f$ are the only functions to be considered as explained above). Thus,

$$c^2(C_3) = \frac{2}{9} < \sigma^2(C_3).$$

Also, recall that, for the 3-path P_3 , $\sigma^2(P_3) = \frac{v^2 - 1}{12} = \frac{2}{3}$, which is greater than the corresponding constant for the 3-cycle.

Let us now pass to the (more symmetric) case of the 4-cycle C_4 . When $v = 4$, up to the shift operation the number of extremal sequences in $ex(M_0)$ is larger than in the case $v = 3$, but since the reflection $y \rightarrow -y$ in M_0 corresponds to the map $f \rightarrow -f$ and since $L_G(t, -f) = L_G(-t, f)$, we need only consider extremal sequences up to reflections. Up to these two types of transformations, there exist only three sequences (recall (??)):

$$(1, 1, -1, -1), (1, -1, 1, -1), (1, -1, -1, 1).$$

The first sequence corresponds to the function f such that $f(x_0) = 0$, $f(x_1) = 1$, $f(x_2) = 2$, $f(x_3) = 1$, that is to the function $f(x) = d(x, x_0)$. In this case, $Ef = 1$ and

$$L_G(t, f) = \frac{1 + 2e^t + e^{2t}}{4} e^{-t} = \frac{1 + \text{ch}(t)}{2}.$$

The third sequence corresponds to the function f such that $f(x_0) = 0$, $f(x_1) = 1$, $f(x_2) = 0$, $f(x_3) = -1$ in which case $Ef = 0$ and therefore $f(x) = 1 + d(x, x_3)$ (and thus again $L_G(t, f) = \frac{1 + \text{ch}(t)}{2}$). The second sequence corresponds to the function f such that $f(x_0) = 0$, $f(x_1) = 1$, $f(x_2) = 0$, $f(x_3) = 1$ in which case $Ef = 1/2$ and

$$L_G(t, f) = \frac{1 + e^t}{2} e^{-t/2} = \text{ch}\left(\frac{t}{2}\right).$$

Since $\text{ch}\left(\frac{t}{2}\right) \leq \frac{1 + \text{ch}(t)}{2}$ for all $t \in \mathbf{R}$, this last function is not a ‘‘point’’ of maximum of the functional $L_G(t, \cdot)$. Thus, we verified Conjecture ?? in the case $v = 4$.

Proposition 4.4 *For the 4-cycle C_4 , we have*

$$L_G(t) = \frac{1 + \text{ch}(t)}{2}, \quad t \in \mathbf{R}.$$

In particular,

$$c^2(C_4) = \sigma^2(C_4) = \frac{1}{2}.$$

Proof It only remains to find an optimal constant σ^2 satisfying the inequality

$$\varphi(t) = \log \frac{1 + \text{ch}(t)}{2} - \frac{\sigma^2 t^2}{2} \leq 0, \quad t \geq 0.$$

We have $\varphi(0) = 0$,

$$\varphi'(t) = \frac{\text{sh}(t)}{1 + \text{ch}(t)} - \sigma^2 t, \quad \varphi''(t) = \frac{1}{1 + \text{ch}(t)} - \sigma^2.$$

Since $\varphi(0) = \varphi'(0) = 0$, the condition $\varphi(t) \leq 0$ for all $t \geq 0$ implies $\varphi''(0) \leq 0$, that is, $\sigma^2 \geq 1/2$. On the other hand, with the value $\sigma^2 = 1/2$, $\varphi''(t) \leq 0$, for all $t \in \mathbf{R}$. Hence φ is concave and thus non-positive. It is also clear that, for the function $f(x) = d(x, x_0)$, $\text{Var } f = 1/2$. Thus $c^2(C_4) = 1/2$. Proposition ?? is proved.

Remark 4.5 While the above is self-contained, we could have deduced using the tensoring property that $\sigma^2(C_4) = 2\sigma^2(P_2) = 1/2$, since $C_4 = P_2 \square P_2$, the Cartesian product graph of P_2 with itself.

Remark 4.6 Colin McDiarmid pointed out to us that in the case of even v , the above conjecture is true and follows from an elementary argument based on some of the above observations. Suppose that v is even and that the conjecture is not true. Fix $t \in \mathbf{R}$ and consider an arbitrary Lipschitz f . Further suppose (without loss of generality, see e.g. the discussion leading to (??) above) that we restrict ourselves to Lipschitz f whose range lies in $R = \{0, 1, \dots, v/2\}$. Then either every value in R has precisely two pre-images or there exists an $i \in R$ with at least three pre-images. In either case, it is easy to see that we can define (by simply permuting the values of f) a Lipschitz g so that $Ee^{t(g-Eg)} = Ee^{t(f-Ef)}$, and that $g(x_k) - g(x_{k-1}) = 0$, for some $1 \leq k \leq v$; but now (??) shows that such a g can not be extremal!

Unfortunately, the above argument does not seem to extend to the *odd* case. It is rather interesting that it is also the odd case in which the characterization of extremal sets is still unsolved for the discrete isoperimetric problem on the n -dimensional discrete torus. However, see Remark ?? below.

Finally to conclude this section, we state an analog of Conjecture ?? when maximizing the variance, and postpone the proof to the Appendix.

Proposition 4.7 *In the class of all Lipschitz functions f on the v -cycle $G = C_v$, $\text{Var } f$ is maximized for the function $f(x) = d(x, x_0)$, $x \in V$ (the variance is with respect to the normalized counting measure π on V). In particular,*

$$c^2(C_v) = \begin{cases} \frac{v^2 + 8}{48}, & \text{if } v \text{ is even,} \\ \frac{(v^2 - 1)(v^2 + 3)}{48v^2}, & \text{if } v \text{ is odd.} \end{cases} \quad (4.8)$$

Remark 4.8 Very recently the last author and Marcus Sammer showed [?], using the transport formulation of the subgaussian (mentioned in the introduction) that for $v \geq 3$,

$$\sigma^2(C_v) = (1 + o(1))\sigma^2(P_{\lceil v/2 \rceil}) = (1 + o(1))\frac{v^2}{48},$$

where the $o(1)$ goes to zero as v goes to infinity.

5 Symmetric group

In the following we consider two natural metrics on S_v , the symmetric group of all permutations of elements in the sequence $(1, \dots, v)$. Any element x of S_v may be viewed as a bijection of the set $I_v = \{1, \dots, v\}$ onto itself. For $x, y \in S_v$, the product xy is the bijection such that $(xy)(i) = y(x(i))$, for all $i = 1, \dots, v$. As usual, we also write x_i instead of $x(i)$. The canonical metric d_v on S_v (cf. [?]) is induced from the Hamming space I_v^v (of which S_v is a subspace):

$$d_v(x, y) = \text{card}\{i \leq v : x_i \neq y_i\}, \quad x, y \in S_v. \quad (5.1)$$

Proposition 5.1 *For all $v \geq 2$,*

$$\sigma^2(S_v, d_v) \leq v - 1.$$

In other words, for every d_v -Lipschitz function f on S_v and all $t \in \mathbf{R}$,

$$\mathbb{E}e^{t(f - \mathbb{E}f)} \leq e^{(v-1)t^2/2}, \quad (5.2)$$

where the expectations are with respect to the normalized counting measure π_v on S_v . In particular, for all $A \subset S_v$ with $\pi_v(A) \geq 1/2$, and all integer $h \geq \sqrt{v-1}$,

$$\pi_v\{x \in S_v : d_v(A, x) \geq h\} \leq \exp\left\{-\frac{(h - \sqrt{v-1})^2}{2(v-1)}\right\}. \quad (5.3)$$

A concentration inequality for (S_v, d_v) was first obtained by Maurey in [?] who proved that for all $A \subset S_v$ with $\pi_v(A) \geq p$,

$$\pi_v\{x \in S_v : d_v(x, A) \geq h\} \leq \exp\left\{-\frac{\left(\frac{h}{2} - 2\sqrt{v \log(1/p)}\right)^2}{4v}\right\}. \quad (5.4)$$

When $p = 1/2$, (??) slightly improves upon (??).

Denote by $s_{i,j}$ the transposition of $i, j \in I_v$: $(s_{i,j})_i = j$, $(s_{i,j})_j = i$, and $(s_{i,j})_k = k$, for $k \neq i, j$. In particular, $s_{i,i}$ is the identity permutation, and we also have $s_{i,j}^{-1} = s_{i,j} = s_{j,i}$. Given $x \in S_v$, the permutations $y = s_{i,j}x$ ($i < j$) are at the least d_v -distance from x and could be considered as its neighbors. Note however that $d_v(x, s_{i,j}x) = 2$ (and that the metric cannot take the value 1). In this sense, the metric d_v is not the usual graph metric, since the distance between any two neighbors in the graph is 1. The inner graph metric $\rho_v(x, y)$, if we consider $\{s_{i,j}x : i \neq j\}$ as the set of all neighbors of x , should be defined as the least number of transpositions z_1, \dots, z_k such that $z_1 \cdots z_k x = y$. In particular, a function f on S_v is ρ_v -Lipschitz if and only if, for all $x \in S_v$ and all $1 \leq i < j \leq v$,

$$|f(s_{i,j}x) - f(x)| \leq 1. \quad (5.5)$$

We are thus faced with another problem: To get a concentration inequality for (S_v, ρ_v) . In general,

$$\rho_v(x, y) \leq d_v(x, y) - 1, \quad x \neq y,$$

from which it follows that $\rho_v(A, x) \leq d_v(A, x) - 1$, $x \notin A$. Hence, for $h \geq 1$, we have

$$\{x \in S_v : \rho_v(x, A) \geq h\} \subset \{x \in S_v : d_v(x, A) \geq h + 1\}$$

so that, by (??),

$$\pi_v\{x \in S_v : \rho_v(x, A) \geq h\} \leq \exp\left\{-\frac{(h+1-\sqrt{v-1})^2}{2(v-1)}\right\}, \quad h = 1, 2, \dots$$

Thus, concentration for (S_v, d_v) implies concentration for (S_v, ρ_v) . For the converse, the Hamming distance d_v is at most twice the transposition distance ρ_v (trivially, every transposition displaces at most two elements, increasing the Hamming distance by at most two). Just as $\rho_v \leq d_v - 1$ can be tight, $d_v \leq 2\rho_v$ can also be tight: consider $(123456\dots(v-1)v)$ and $(214365\dots v(v-1))$, for v even. Then $d_v = v$ and $\rho_v = v/2$. Thus (??) implies (??). Indeed, the sharper concentration inequality for (S_v, ρ_v) can be obtained as shown next (this is similar to a corresponding property of the discrete cube $\{0, 1\}^{v-1}$):

Proposition 5.2 *For all $v \geq 2$,*

$$\frac{v-1}{16} \leq c^2(S_v, \rho_v) \leq \sigma^2(S_v, \rho_v) \leq \frac{v-1}{4}.$$

In other words, the upper bound shows that for every ρ_v -Lipschitz function f on S_v and all $t \in \mathbf{R}$,

$$\mathbf{E}e^{t(f-\mathbf{E}f)} \leq e^{(v-1)t^2/8}, \quad (5.6)$$

where the expectations are with respect to the normalized counting measure π_v on S_v . In particular, for all $A \subset S_v$ with $\pi_v(A) \geq 1/2$, and all integer $h \geq \frac{1}{2}\sqrt{v-1}$,

$$\pi_v\{x \in S_v : \rho_v(A, x) \geq h\} \leq \exp\left\{-\frac{(2h-\sqrt{v-1})^2}{2(v-1)}\right\}. \quad (5.7)$$

The inequality (??) is not new. It appears in a paper of McDiarmid [?] inside a martingale-based proof of a version of the concentration inequality (??) (see the proof of Theorem 6.7 and Example 7.1 there). The inequality (??) can also be proved using the method of bounded differences as in [?]. The method goes back to the work of Maurey [?] and has been used since then by many authors, cf. e.g., [?], [?], [?]. To date, the best estimate obtained by this method seems to be given by $\sigma^2(S_v, d_v) \leq 4(v-1)$. Versions of Maurey's concentration result have also been recovered (using a different method) by Talagrand (see Section 5 in [?]). However, we would like to present below a related but direct inductive proof of (??) without having to introduce the notion of martingale.

We also remark that (S_v, ρ_v) was considered in the context of card-shuffling, and in particular, in the estimation of log-Sobolev constants of Markov kernels by [?]. (See Section ?? for a more detailed analysis of the interplay between the log-Sobolev and subgaussian constants in the context of graphs.)

Remark 5.3 The following argument due to Schechtman and communicated to us by Kwapien also provides an estimate of the subgaussian constant of (S_v, d_v) equipped with the normalized counting measure π_v . Let $I_v = \{1, 2, \dots, v\}$ and let λ_v be the normalized counting measure on I_v . Let $J_v = I_v \times I_{v-1} \times \dots \times I_1$ and let $\nu_v = \lambda_v \times \lambda_{v-1} \times \dots \times \lambda_1$. Then ν_v is the normalized counting measure on J_v which is equipped with the Hamming distance. Then, there exists a Lipschitz map $T : J_v \rightarrow S_v$ which is one-to-one and onto and of Lipschitz constant 3. As a result, for any Lipschitz (with constant 1) function $f : (S_v, d_v) \rightarrow \mathbf{R}$,

$$\pi_v\{f - \mathbb{E}_{\pi_v} f \geq h\} = \nu_v\{T^{-1}f - \mathbb{E}_{\nu_v}(T^{-1}f) \geq h\}.$$

It follows that $\sigma^2(S_v, d_v) \leq 9\sigma^2(J_v, d_v)$, and any estimate on $\sigma^2(J_v, d_v)$ gives an estimate on $\sigma^2(S_v, d_v)$. The rough bound, $\sigma^2(J_v, d_v) \leq v$, is easily verified and in turn, it implies that $\sigma^2(S_v, d_v) \leq 9v$. This is of the same order in v as known estimates, e.g., $4(v-1)$, but also less precise than (??). Let us now describe (recursively) T . If $x \in J_v$, then $y = T(x)$ is defined by $y(k) = x(k)$ if $x(j) \neq x(k)$ for all $j < k$. Otherwise, take $j_1 = \max\{j : j < k, x(j) = x(k)\}$, $x_1 = n+1 - j_1$ and put $y(k) = x_1$ if $x(j) \neq x_1$ for all $j < j_1$. Otherwise, take $j_2 = \max\{j : j < j_1, x(j) = x_1\}$, $x_2 = n+1 - j_1$, then set $y(k) = x_2$ if $x(j) \neq x_2$ for all $j < j_2$. Otherwise let $j_3 = \max\{j < j_2 : x(j) = x_2\}$, $x_3 = n+1 - x_2$, and put $y(k) = x_3$ if $x(j) \neq x_3$ for $j < j_3$. Continue this way until we find x_s such that $x_s \neq x_j$ for $j < j_s$ and then set $y(k) = x_s$.

Proof of Proposition ?? The case $v = 2$ is trivial. Let $v \geq 3$ and assume that (??) holds for S_{v-1} . Set $H = \{x \in S_v : x_v = v\}$. Thus, H realizes a natural embedding of S_{v-1} in S_v , and we may think of H as a copy of S_{v-1} . In particular, one may apply the induction hypothesis to d_{v-1} -Lipschitz functions on H . Set $e_i = s_{v,i}$ and introduce

$$H_i = He_i = \{xe_i : x \in H\} = \{x \in S_v : x_v = i\}, \quad i = 1, \dots, v.$$

In particular, $H_v = H$. The family $(H_i)_{1 \leq i \leq v}$ forms a partition of S_v : $H_i \cap H_j = \emptyset$, for all $1 \leq i < j \leq v$ and $\cup_{i=1}^v H_i = S_v$.

We note the following property of the metric d_v :

$$1) \quad d_v(x^{-1}, y^{-1}) = d_v(x, y), \text{ for all } x, y \in S_v.$$

Indeed,

$$\{i \leq v : x^{-1}(i) = y^{-1}(i)\} = \{i \leq v : i = x(y^{-1}(i))\} = \{y(j) \leq v : y(j) = x(j)\}.$$

Hence, $\text{card}\{i \leq v : x_i^{-1} = y_i^{-1}\} = \text{card}\{j \leq v : x_j = y_j\}$ which proves 1). It follows from this property that:

$$2) \quad d_v(xe_i, ye_i) = d_v(e_ix, e_iy) = d_v(x, y), \text{ for all } x, y \in H.$$

There is nothing to prove when $i = v$. If $i < v$, to see the second equality, let us write the elements of x, y, e_ix and e_iy , recalling that $x_v = y_v = v$:

$$\begin{aligned} x &= (x_1, \dots, x_{v-1}, v) & e_ix &= (x_1, \dots, x_{i-1}, v, x_{i+1}, \dots, x_{v-1}, x_i) \\ y &= (y_1, \dots, y_{v-1}, v) & e_iy &= (y_1, \dots, y_{i-1}, v, y_{i+1}, \dots, y_{v-1}, y_i). \end{aligned}$$

According to (??), we obtain that $d_v(e_ix, e_iy) = d_v(x, y)$. To prove the first equality in 2), we may combine 1) and the second equality in 2):

$$d_v(xe_i, ye_i) = d_v((xe_i)^{-1}, (ye_i)^{-1}) = d_v(e_ix^{-1}, e_iy^{-1}) = d_v(x^{-1}, y^{-1}) = d_v(x, y).$$

We also note the following analogue of 2) for the metric ρ_v (needed for the proof of Proposition ??):

$$3) \quad \rho_v(xe_i, ye_i) = \rho_v(x, y), \text{ for all } x, y \in H.$$

It suffices to check this property assuming that the elements $x, y \in H$ are neighbors. In this case, $y = s_{\alpha, \beta}x$, for some $1 \leq \alpha < \beta \leq v - 1$. Hence, $ye_i = s_{\alpha, \beta}(xe_i)$, so, xe_i and ye_i are neighbors. In fact, conversely, given $x, y \in H$, if xe_i and ye_i are neighbors in S_v , then, $ye_i = s_{\alpha, \beta}(xe_i)$, for some $1 \leq \alpha < \beta \leq v$. Hence, $y = s_{\alpha, \beta}x$. In particular, $\beta < v$ since otherwise $y_\alpha = x_v = v$ contradicting the condition $y \in H$. Thus, x and y are neighbours in H as in S_{v-1} .

Finally, we need:

$$4) \quad \text{For all } 1 \leq i < j \leq v, \text{ there is a bijection } \varphi_{i,j} : H_i \rightarrow H_j \text{ such that}$$

$$\rho_v(x, \varphi_{i,j}(x)) = 1, \quad \text{that is,} \quad d_v(x, \varphi_{i,j}(x)) = 2.$$

Indeed, one may take $\varphi_{i,j}(x) = s_{i, x^{-1}(j)}x$ which replaces only two coordinates of x .

We are now ready to perform the induction step to prove (??). Given a Lipschitz function f on S_v , introduce the functions $f_i(x) = f(xe_i)$ defined on H ($1 \leq i \leq v$). From property 4), these functions are d_{v-1} -Lipschitz on H . Hence, by the induction hypothesis, for all $t \in \mathbf{R}$,

$$\mathbb{E}_{\pi_{v-1}} e^{tf_i} \leq e^{(v-2)t^2/2} e^{t\mathbb{E}_{\pi_{v-1}} f_i}. \quad (5.8)$$

Denote by μ_i the normalized counting measure on H_i . In particular, μ_v is the normalized counting measure on H which was denoted in (??) by π_{v-1} . Clearly,

$$\pi_v = \frac{1}{v} \sum_{i=1}^v \mu_i. \quad (5.9)$$

Since $E_{\pi_{v-1}} e^{tf_i} = E_{\mu_i} e^{tf}$, it follows from (??) and (??) that

$$E_{\pi_v} e^{tf} = \frac{1}{v} \sum_{i=1}^v E_{\mu_i} e^{tf} \leq e^{(v-2)t^2/2} \frac{1}{v} \sum_{i=1}^v e^{tE_{\pi_{v-1}} f_i}. \quad (5.10)$$

In terms of the function

$$g(i) = E_{\pi_{v-1}} f_i = E_{\mu_i} f = \int_{H_i} f(x) d\mu_i(x), \quad i \in I_v,$$

the relation (??) has the form

$$E_{\pi_v} e^{tf} \leq e^{(v-2)t^2/2} \int_{I_v} e^{tg(i)} \lambda(i), \quad (5.11)$$

where λ is the normalized counting measure on I_v . Now, we use the property 4) for d_v . As a bijection, $\varphi_{i,j}$ transforms μ_i into μ_j , so,

$$g(j) = \int_{H_i} f(\varphi_{i,j}(x)) d\mu_i(x).$$

Hence, since f is d_v -Lipschitz,

$$|g(i) - g(j)| = \left| \int_{H_i} (f(x) - f(\varphi_{i,j}(x))) d\mu_i(x) \right| \leq \int_{H_i} d_v(x, \varphi_{i,j}(x)) d\mu_i(x) \leq 2.$$

Therefore, since the function g has Lipschitz seminorm at most 2 with respect to the metric $1_{\{i \neq j\}}$ on I_v , for all $t \in \mathbf{R}$,

$$E_{\lambda} e^{t(g - E_{\lambda} g)} \leq e^{(2t)^2/8}. \quad (5.12)$$

Combining (??) and (??), we obtain that

$$E_{\pi_v} e^{tf} \leq e^{(v-2)t^2/2} e^{4t^2/8} e^{tE_{\lambda} g} = e^{(v-1)t^2/2} e^{tE_{\lambda} g}. \quad (5.13)$$

It remains to note that, by (??),

$$E_{\lambda} g = \frac{1}{v} \sum_{i=1}^v E_{\mu_i} f = E_{\pi_v} f.$$

Thus, (??) is exactly (??). Proposition ?? is proved.

Proof of Proposition ??

(i) upper bound. The proof of (??) uses the same induction line: introduce the functions f_i , g and the probability measures μ_i , λ as above. From the property 3)

and since f is ρ_v -Lipschitz, the functions f_i are ρ_{v-1} -Lipschitz on H . Hence, by the induction hypothesis, we have an analogue of (??) and (??):

$$\mathbb{E}_{\pi_v} e^{tf} \leq e^{(v-2)t^2/8} \int_{I_v} e^{tg(i)} \lambda(i). \quad (5.14)$$

Also, by the property 4) for ρ_v ,

$$|g(i) - g(j)| = \left| \int_{H_i} (f(x) - f(\varphi_{i,j}(x))) d\mu_i(x) \right| \leq \int_{H_i} \rho_v(x, \varphi_{i,j}(x)) d\mu_i(x) \leq 1.$$

That is, the function g has Lipschitz seminorm at most 1 with respect to the metric $1_{\{i \neq j\}}$ on I_v . Hence, for all $t \in \mathbf{R}$,

$$\mathbb{E}_\lambda e^{t(g - \mathbb{E}_\lambda g)} \leq e^{t^2/8}. \quad (5.15)$$

Therefore, combining (??) and (??) and recalling that $\mathbb{E}_\lambda g = \mathbb{E}_{\pi_v} f$, we obtain that

$$\mathbb{E}_{\pi_v} e^{tf} \leq e^{(v-2)t^2/8} e^{t^2/8} e^{t\mathbb{E}_\lambda g} = e^{(v-1)t^2/8} e^{t\mathbb{E}_{\pi_v} f}. \quad (5.16)$$

Thus, (??) is exactly (??).

(ii) lower bound. It suffices to give a specific Lipschitz function f with $\text{Var } f \geq \frac{v-1}{16}$. The cases $v = 1$ and $v = 2$ are obvious since $c_1(S_1, \rho_1) = 0$, $c_2(S_2, \rho_2) = 1/4$. Assume $v \geq 3$. For $x \in S_v$, consider a function $f(x) = \text{card}\{i \leq a : x_i \leq b\}$, where the integers $a, b \in [2, v]$ will be chosen later. First note that f is ρ_v -Lipschitz with constant 1. Then simple calculations yield the following, with respect to the normalized counting measure: $\mathbb{E}f = \frac{ab}{v}$, and since $f^2(x) = \sum_{i,j \leq a} \mathbf{1}_{x_i \leq b} \mathbf{1}_{x_j \leq b}$, we also have $\mathbb{E}f^2 = \frac{ab}{v} + \frac{ab(a-1)(b-1)}{v(v-1)}$. The above clearly yields

$$c^2(S_v, \rho_v) \geq \text{Var } f = \frac{ab(ab - (a+b)v + v^2)}{v^2(v-1)}.$$

If $v = 3$, take $a = b = 2$, so that $\text{Var } f = \frac{2}{9} \geq \frac{1}{8}$. If $v \geq 4$ is even, take $a = b = v/2$, and if $v \geq 4$ is odd, take $a = (v-1)/2$, $b = (v+1)/2$. In both cases $a+b = v$, $ab \geq (v-1)(v+1)/4$. Hence,

$$\text{Var } f = \frac{(ab)^2}{v^2(v-1)} \geq \left(\frac{v+1}{v}\right)^2 \frac{v-1}{16}.$$

Proposition ?? is proved.

Remark 5.4 McDiarmid [?] considers another graph denoted by $S_{v,k}$, which is a generalization of the above permutation graph (S_v, ρ_v) . The k -tuple graph $S_{v,k}$ (Example 7.3 of [?]) has vertices the k -tuples of distinct members of $\{1, 2, \dots, v\}$, and two vertices are adjacent if and only if either they differ in exactly one coordinate or they differ in exactly two coordinates. The induction argument of the upper bound of Proposition ?? easily extends to yield $\sigma^2(S_{v,k}) \leq k/4$. This graph is of interest since it is closely related to the k -subset graph $\hat{S}_{v,k}$, a graph of special interest to combinatorialists. In particular, as observed in the proof of Proposition 7.13 of [?], they have the same concentration phenomenon. The vertices of the k -subset graph are the k -subsets of an v -set, and two vertices are adjacent if and only if their intersection is exactly $k - 1$, i.e. they are obtainable from each other by removing one element and adding another element.

6 Complete graph

Let $K_v = (V, \mathcal{E})$ be the complete graph: V is a non-empty finite set of cardinality v , and

$$\mathcal{E} = \{\{x, y\} : x, y \in V, x \neq y\}.$$

The graph metric here is $d(x, y) = 1_{\{x \neq y\}}$. Thus, a function f on V is Lipschitz if and only if, for all $x, y \in V$,

$$|f(x) - f(y)| \leq 1. \tag{6.1}$$

Assume V is equipped with a probability measure μ and define

$$p(\mu) = \inf\{\mu(A) : A \subset V, \mu(A) \geq 1/2\}.$$

Proposition 6.1

$$\sigma^2(K_v, \mu) = \frac{p - q}{2(\log p - \log q)}, \tag{6.2}$$

where $p = p(\mu)$ and $q = 1 - p$.

If $p = 1/2$, the above expression is defined to be $1/4$ by continuity. In the particular case where $\mu = \pi$, the normalized counting measure, when $v = \text{card } V = 2r$, we have $p(\pi) = 1/2$, and when $v = 2r + 1$, we have $p(\pi) = (r + 1)/(2r + 1)$. Therefore:

Corollary 6.2 For the completely connected graph V of cardinality v equipped with the normalized counting measure, $\sigma^2(K_v) = 1/4$, if $v = 2r$, and

$$\sigma^2(K_v) = \frac{1}{2(2r + 1) \log \frac{r+1}{r}}, \quad \text{if } v = 2r + 1.$$

Thus, if K_v^n (on the set of vertices V^n) is the n -th power of K_v , with $v = 2r + 1$, with the Hamming distance d and the normalized counting measure π^n , for every set $A \subset V^n$ of measure $\pi^n(A) \geq 1/2$, and for $h \geq \sqrt{\frac{n}{2(2r+1) \log \frac{r+1}{r}}}$,

$$\pi^n \{x \in V^n : d(x, A) \geq h\} \leq \left(\frac{r}{r+1}\right)^{\frac{(2r+1)}{n}} \left(h - \sqrt{\frac{n}{2(2r+1) \log \frac{r+1}{r}}}\right)^2. \quad (6.3)$$

When $v = 3$, we thus recover (the second part of) Proposition ??, since the 3-cycle is completely connected. If $v = 2r$, the inequality (??) should be replaced by the slightly weaker inequality

$$\pi^n \{x \in V^n : d(x, A) \geq h\} \leq \exp \left\{ -\frac{(2h - \sqrt{n})^2}{2n} \right\}, \quad 2h \geq \sqrt{n}. \quad (6.4)$$

Note that since the complete graph is a Hamming space, (??) is slightly better than (??). Note also that $c^2(K_v, \mu) = p(\mu)(1 - p(\mu))$. Hence, with respect to the normalized counting measure on K_v , $c^2(K_v) = \sigma^2(K_v) = 1/4$, if v is even, and $c^2(K_v) < \sigma^2(K_v)$, otherwise. Finally note that the general bound, $\sigma^2 \leq D^2/4$ (which follows from Corollary 3.3 in [?]), where D is the diameter of the graph G , can be tight as the computation of σ^2 of the complete graph shows.

Proof of Proposition ?? The functional $f \rightarrow L(t, f) = \mathbf{E}_\mu e^{t(f - \mathbf{E}_\mu f)}$ is translation invariant, so maximizing this functional in the class of functions satisfying (??), we can restrict ourselves to $0 \leq f \leq 1$. The class \mathcal{F}_0 of such functions is compact and convex, and the functional $f \rightarrow L(t, f)$ is convex. Hence, it attains its maximum on \mathcal{F}_0 at some extremal ‘‘point’’ of \mathcal{F}_0 . But the extremal functions in \mathcal{F}_0 are just indicator functions $f = 1_A$, $A \subset V$. Thus,

$$\sup_{f \in \mathcal{F}(G)} L(t, f) = \sup_{A \subset V} L(t, 1_A).$$

Therefore, the optimal constant $\sigma^2 = \sigma^2(K_v, \mu)$ in the inequality

$$L(t, f) \leq e^{\sigma^2 t^2 / 2}, \quad f \in \mathcal{F}(G), \quad t \in \mathbf{R},$$

satisfies $\sigma^2 = \sup_{A \subset V} \sigma_A^2$, where σ_A^2 is the optimal constant in the inequality

$$L(t, 1_A) \leq e^{\sigma_A^2 t^2 / 2}, \quad t \in \mathbf{R}.$$

Any function $f = 1_A$ is a Bernoulli random variable on the probability space (V, μ) , taking the values 1 and 0 with probabilities $\mu(A)$ and $\mu(B)$, respectively, where $B = V \setminus A = A^c$. Hence by Proposition ??,

$$\sigma_A^2 = \frac{\mu(A) - \mu(B)}{2(\log \mu(A) - \log \mu(B))}.$$

Since $\sigma_A^2 = \sigma_B^2$, we may restrict ourselves to the cases $\mu(A) \geq \mu(B)$, i.e, $\mu(A) \geq 1/2$. Thus,

$$\sigma^2 = \sup_{\mu(A)=p \geq 1/2} \frac{p - q}{2(\log p - \log q)},$$

where $q = 1 - p$, $p = \mu(A)$, and the sup is taken over all $A \subset V$ with $\mu(A) \geq 1/2$. To prove (??), it remains to show that the function $u(p) = \frac{p - q}{\log p - \log q}$ is decreasing in $(1/2, 1)$. Let

$$v(p) = \frac{\log p - \log q}{p - q}, \quad 1/2 < p < 1.$$

We have: $v'(p) > 0$ iff $\frac{p - q}{pq} \geq 2(\log p - \log q)$ which can be rewritten as

$$w(p) = 2pq(\log p - \log q) - (p - q) < 0, \quad 1/2 < p < 1.$$

Since $w' = -2(p - q)(\log p - \log q) < 0$, the function w is decreasing. But $w(1/2) = 0$, so w is negative on $(1/2, 1)$. Thus, v is increasing, that is, u is decreasing. Proposition ?? is proved.

7 Log-Sobolev and subgaussian constants

As shown by Aida, Masuda and Shigekawa ([?]) and by Ledoux ([?]) (with an argument going back to Herbst), it is also possible to derive concentration inequalities, whenever a log-Sobolev inequality holds. Based on these works and on ([?]), such derivations under log-Sobolev as well as Poincaré inequalities were given for products of Markov kernels and graph products in ([?]). It is thus quite appropriate to try to compare the log-Sobolev and subgaussian approaches. Since studying graphs and Lipschitz functions on them, a natural notion of discrete gradient of f at the point $x \in V$ is

$$\nabla_{\infty}^+ f(x) = \sup_{y: \{x, y\} \in \mathcal{E}} (f(x) - f(y))^+,$$

with a matching definition for ∇_{∞}^- , i.e., $\nabla_{\infty}^- f(x) = \sup_{y: \{x, y\} \in \mathcal{E}} (f(x) - f(y))^-$. Then, the corresponding log-Sobolev inequality (with the optimal constant $\rho_{\infty}^+ > 0$), is

$$\rho_{\infty}^+ \left[\mathbb{E} f^2 \log f^2 - \mathbb{E} f^2 \log \mathbb{E} f^2 \right] \leq \mathbb{E} (\nabla_{\infty}^+ f)^2,$$

where expectation is with respect to any measure π on V .

Proposition 7.1 *Let f be a function on G , then*

$$\mathbb{E} e^{t(f - \mathbb{E}f)} \leq e^{\frac{t^2 \|\nabla_{\infty}^+ f\|_{\infty}^2}{4\rho_{\infty}^+}}, \quad t \geq 0. \quad (7.1)$$

In particular,

$$\sigma^2(G) \leq \frac{1}{2\rho_{\infty}^+(G)}. \quad (7.2)$$

Proof Note that if $g = e^{tf/2}$, $t \geq 0$, then

$$(g(x) - g(y))^{+2} \leq \frac{t^2}{4}(f(x) - f(y))^{+2}e^{tf(x)}.$$

Hence,

$$\rho_{\infty}^+ \left[\mathbb{E}e^{tf} \log e^{tf} - \mathbb{E}e^{tf} \log \mathbb{E}e^{tf} \right] \leq \frac{t^2}{4} \mathbb{E}(\nabla_{\infty}^+ f)^2 e^{tf}. \quad (7.3)$$

Now applying Herbst's method, i.e., differentiating $H(t) = t^{-1} \log \mathbb{E}e^{tf}$, $H(0) = \mathbb{E}f$, (see [?]), lead to (??), where $\|\nabla_{\infty}^+ f\|_{\infty} = \sup_{x \in V} \nabla_{\infty}^+ f(x)$. For (??), just note that if $t \leq 0$, then (??) hence (??) continue to hold with ∇_{∞}^+ replaced by ∇_{∞}^- and that moreover if f is Lipschitz with respect to the graph distance, then

$$\sup_{y: \{x,y\} \in \mathcal{E}} (f(x) - f(y))^{\pm} \leq \sup_{y: \{x,y\} \in \mathcal{E}} |f(x) - f(y)| \leq 1.$$

Again, (??) implies deviation from the mean, i.e., for any Lipschitz function f and $h > 0$,

$$\pi\{f \geq \mathbb{E}f + h\} \leq \exp \left\{ -\frac{\rho_{\infty}^+ h^2}{\|\nabla_{\infty}^+ f\|_{\infty}^2} \right\} \leq \exp \left\{ -\rho_{\infty}^+ h^2 \right\}. \quad (7.4)$$

Next, since the graph distance is also the natural distance, (e.g., see [?])

$$d(x, y) = \sup_{|\nabla_{\infty} f| \leq 1} |f(x) - f(y)|, \quad x, y \in V,$$

where $|\nabla_{\infty} f(x)| = \sup_{y: \{x,y\} \in \mathcal{E}} |f(x) - f(y)|$, it easily follows from (??) that for all $A \subset V$,

$$1 - \pi(A^h) \leq e^{-\rho_{\infty}^+ (\pi(A))^2 (h+1)^2}, \quad h = 0, 1, 2, \dots \quad (7.5)$$

We now return to the examples analyzed in the previous sections and show that for the normalized counting measure, (??) often provides a good estimate of σ^2 . So, for the rest of the section, π is the normalized counting measure.

- $G = P_v$. Modifying the computations in Example 3.6 in [?] and since the maximal degree of P_v is 2, it is easy to see that:

$$\frac{1}{2v^2} \leq \rho_{\infty}^+(P_v) \leq \frac{\pi^2}{2v^2}.$$

- $G = C_v$. Using Example 4.2 in [?], and since the maximal degree of C_v is also 2, we get

$$\frac{8\pi^2}{25v^2} \leq \rho_{\infty}^+(C_v) \leq \frac{4\pi^2}{v^2}.$$

- $G = (S_v, \rho_v)$. Below, the upper bound follows from the computations in Example 4.3 in [?] while the lower bound is easily obtained via generic lower bounds given in [?], [?].

$$\frac{1}{v-1} \leq \rho_\infty^+(S_v) \leq \frac{2}{v-1}.$$

- $G = K_v$. It is easy to show that ρ_∞^+ of the complete graph can be bounded above and below independently of v . In fact, using bounds obtained in [?], [?], it also easily follows that:

$$\frac{1}{4 + (2 + 2\sqrt{2}) \log 2} \leq \rho_\infty^+(K_v) \leq 2,$$

if v is even and

$$\frac{r+1}{4(r+1) + (2 + 2\sqrt{2})r \log \frac{2r+1}{r}} \leq \rho_\infty^+(K_v) \leq \frac{(2r+1)^2}{2r(r+1)},$$

if $v = 2r + 1$ is odd.

Although in all the above examples, σ^2 and $1/\rho_\infty^+$ are of the same order in v , for non uniform measures, they can be quite different. A case at hand is the weighted two point space with probability of success p , then $\rho_\infty^+ = \min\left(\frac{1}{p}, \frac{1}{q}\right) \left(\frac{p-q}{\log p - \log q}\right)$, which, for p small, is of the same order $1/(-\log p)$ as σ^2 given in (?). Hence, in that case, the concentration inequality obtained via the subgaussian constant is much stronger than the one obtained via the log-Sobolev constant.

Appendix

Proof of Proposition ?? Let $V = \{x_0, \dots, x_{v-1}\}$ denote the vertices of the v -cycle. Let f be an extremal function, achieving the spread constant c^2 of the v -cycle. We make crucial use of Theorem 2.1 of [?], which we state here (as Fact 1) for the sake of completeness. This theorem is about the extremal Lipschitz functions defined on any undirected graph which maximize the variance, thus achieving the so-called spread constant of the corresponding graph. (Although the following theorem is stated (in [?]) in the context of the normalized counting measure, it holds true under an arbitrary probability measure on the set of vertices of the undirected graph.)

Fact 1 ([?]). *There is an integer-valued optimal f , and moreover there is a subset of vertices U , and an assignment of a sign $s(C) \in \{-1, 1\}$ to every connected component C of $V \setminus U$ so that for every vertex $z \in U$, we have $f(z) = 0$, and for every vertex x in a component C as above, $f(x)$ is the product of $s(C)$ and the (graph) distance between x and U .*

For the rest of the proof we specialize to the case of a v -cycle, and all the expectations and the variances are with respect to the normalized counting measure. In view of Fact 1, we may assume that an optimal f is integer-valued, and also that $Ef \geq 0$, since $\text{Var } f = \text{Var}(-f)$.

Let U be the set of vertices, as in Fact 1, such that $f(z) = 0$, for $z \in U$. The proof of Proposition ?? consists in showing that if $|U| = 1$ or 2 , then f can be translated to become $d(x, x_0)$ or $-d(x, x_0)$, and that $|U| \geq 3$ is not possible.

Case 1. Note that if $|U| = 1$, then we may assume by symmetry that $U = \{x_0\}$, and then the only connected component is a path on x_1, \dots, x_{v-1} , and we are done in view of Fact 1.

Case 2. Let $|U| = 2$. Without loss of generality, let $U = \{x_0, x_k\}$, where $k \leq (v-1)/2$.

(a) Suppose v is odd. If $k = 1$, then we still have only one connected component as in Case 1, and by using Fact 1, f is $(1, 2, \dots, (v-1)/2, \dots, 2, 1)$ (it cannot be $(-1, -2, \dots, -(v-1)/2, \dots, -2, -1)$, since we assumed that $Ef \geq 0$), on the component. Replacing f by $f - (v-1)/2$, we get $f(x) = -d(x, x_{(v-1)/2})$, and we are done by the symmetry of the v -cycle, in particular, by treating $x_{(v-1)/2}$ as x_0 . If $k > 1$, then we have two connected components, and since we assumed $Ef \geq 0$, we know that f is positive on the (larger) component consisting of x_{k+1} through x_{v-1} . Moreover, f is negative on the other component – otherwise define $g = -f$, on this component, and $g = f$ everywhere else, then it is easy to see that $\text{Var } g > \text{Var } f$. Also note that since v is odd and $|U| = 2$, exactly one of the connected components has an odd number of vertices – depending on whether k is odd or even. This leaves two possible sequences for f on $(x_0, x_1, \dots, x_k, \dots, x_{v-1})$: either

$$(0, -1, \dots, -(k-1)/2, -(k-1)/2, \dots, -1, 0, 1, \dots, (v-k)/2, \dots, 1),$$

or

$$(0, -1, \dots, -k/2, \dots, -1, 0, 1, \dots, (v-k-1)/2, (v-k-1)/2, \dots, 1).$$

Replacing f by $f - (v-k)/2$ in the first case, and by $f + k/2$ in the second case, and by using the symmetry of the cycle, we are done.

(b) Suppose v is even. Then we claim that there has to be an odd number of vertices between x_0 and x_k . For otherwise, there would be an even number of vertices in the connected components (one of them might have size zero). If $k = 1$, then one of them has size zero, and in view of Fact 1, the other one has f -values, $(1, 2, \dots, (v-2)/2, (v-2)/2, \dots, 2, 1)$; an easy calculation shows that this is not extremal, since the function g with the sequence of values $(1, 2, \dots, (v-2)/2, v/2, (v-2)/2, \dots, 2, 1)$ has $\text{Var } g = \text{Var } f + 1/4$. If $k > 1$, then the two connected components must have f -values,

$$(-1, -2, \dots, -(k-1)/2, -(k-1)/2, \dots, -2, -1),$$

and

$$(1, 2, \dots, (v-k-1)/2, (v-k-1)/2, \dots, 2, 1).$$

Replacing f by $f + (k-1)/2$, gives a sequence with two 0's next to each other, and reduces to the case of $k=1$, which we handled already.

Thus if there are two 0's then they have to be separated by an odd number of vertices. This means that the f -values of the two connected components are

$$(-1, -2, \dots, -k/2 + 1, -k/2, -k/2 + 1, \dots, -2, -1),$$

and

$$(1, 2, \dots, (v-k)/2 - 1, (v-k)/2, (v-k)/2 - 1, \dots, 2, 1).$$

Replacing f by $f + k/2$, results in $f(x) = d(x, x_{k/2})$, and we are done by symmetry.

Case 3. Suppose $|U| \geq 3$. First note that there cannot be three consecutive 0's in the sequence of f -values. Indeed, if f assigns three consecutive 0's, and assuming as before $Ef > 0$, define g to be equal to -1 in place of the middle zero, and otherwise to be identical to f . Then $Eg = Ef - 1/v$, and $Eg^2 = Ef^2 + 1/v$. Then $\text{Var } g = \text{Var } f + 1/v - 1/v^2 + 2(Ef)/v > \text{Var } f$, since $Ef > 0$ and $v \geq 2$.

To finish the proof, we know there must be at least two connected components, once we delete the vertices in U , and we may also assume there is a vertex in U which sees negative f -values on one side and positive f -values on the other side. This is because, if f were non-negative everywhere, then changing the sign of every value on the smallest connected component will then yield a function with a smaller mean (since $Ef > 0$ to begin with), and larger variance.

Thus, without loss of generality, let $f(x_0) = 0$, and let x_k and x_l with $0 < k < l$, be the closest vertices (with respect to the graph distance) to x_0 such that $f(x_k) = f(x_l) = 0$. We can also assume that the connected component C^- between x_l and x_0 has negative f -values, $(-1, -2, \dots, -l_1, -l_1 + 1, \dots, -2, -1)$, and the component C^+ between x_0 and x_k has positive f -values, $(1, 2, \dots, k_1, k_1 - 1, \dots, 2, 1)$, where l_1 and k_1 are easily computable constants depending on l and k . Note that $|C^-| > 0$, and furthermore, we can easily dispense with the case that $|C^+| = 0$: suppose that on the vertices x_{v-1}, x_0, x_1 , f takes the values $-1, 0, 0$, respectively, then define g to be $-1, -1, 0$ on the same vertices, and otherwise to be identical to f . Then it is easy to see that $\text{Var } g > \text{Var } f$.

Case 3(a). First assume $f(x_{k+1}) \in \{0, 1\}$. Then define g on the sequence of vertices, $(x_{v-1}, x_0, x_1, \dots, x_{k-1}, x_k)$, to take the values, $(-1, -1, 0, 1, \dots, 2, 1)$, and to be identical to f on the rest of the vertices. (Note that we have essentially introduced an extra -1 , and "adjusted" the rest of the sequence to the right by one position to maintain the Lipschitz property. In the process we moved the zero at x_0 to x_1 and we lost the zero at x_k .) Now note that $Eg^2 = Ef^2 + (-1)^2/v$, and $Eg = Ef + (-1)/v$. Thus,

$$\text{Var } g = \text{Var } f + \frac{1}{v} - \frac{1}{v^2} + \frac{2}{v} Ef > \text{Var } f,$$

since, $v \geq 2$, and $Ef \geq 0$.

Case 3(b). To finish Case 3, we need to consider the possibility that $f(x_{k+1}) = -1$. Note that at this stage we must have at least four components, alternating positive and negative valued, since two negative components cannot be next to each other. Indeed, otherwise the sequence $\dots, -1, 0, -1, \dots$ can be replaced by $\dots -1, -1, -1, \dots$, thus introducing an extra -1 , and it can be checked that this increases the variance, since $Ef > 0$. Thus, starting from x_{v-1} , f takes the sequence of values,

$$(-1, 0, \underline{1, 2, \dots, k_1, k_1 - 1, \dots, 2, 1, 0, -1, \dots, -k_2, -k_2 + 1, \dots, -1, 0, 1, \dots}).$$

Define g to be f , except with the above underlined portion reflected over as follows:

$$(-1, 0, -1, \dots, -k_2 + 1, -k_2, \dots, -1, 0, 1, 2, \dots, k_1 - 1, k_1, \dots, 2, 1, 0, 1, \dots).$$

Clearly, $\text{Var } g = \text{Var } f$, but crucially, we introduced two -1 's surrounding a zero. This cannot be extremal since, as mentioned before, replacing the zero by a -1 increases the variance. \square

Acknowledgement. We thank David Wilson for the idea regarding the lower bound in Proposition ???. We also thank Colin McDiarmid and Noga Alon for various insightful comments.

References

- [A-B-S] Alon, N., Boppana, R., Spencer, J. An asymptotic isoperimetric inequality. *Geom. and Funct. Anal.* **8** (1998), 411–436.
- [A-M-S] Aida, S., Masuda, T., Shigekawa, I. Logarithmic Sobolev inequalities and exponential integrability. *J. Funct. Anal.*, **126** (1994), 83–101.
- [A-M] Amir, D., Milman, V.D. Unconditional and symmetric sets in n -dimensional normed spaces. *Israel J. of Math.* **37** (1980), 3–20.
- [B] Bobkov, S.G. On deviations from medians. Manuscript (1998).
- [B-G] Bobkov, S.G., Götze, F. Exponential integrability and transportation cost related to logarithmic transportation inequalities. *J. Funct. Anal.* **163** (1999), 1–28.
- [B-H1] Bobkov, S.G., Houdré, C. Variance of Lipschitz functions and an isoperimetric problem for a class of product measures, *Bernoulli* **2** (1996), 249–255.

- [B-H2] Bobkov, S.G., Houdré, C. A characterization of Gaussian measures via the isoperimetric properties of half-spaces. *Zap. Nauchn. Semin. S.-Petersburg. Otdel. Mat. Inst. im. V.A.Steklova RAN.* **228** (1996), 31–38 (in Russian). English translation: *J. Math. Sciences*.
- [B-L] Bobkov, S.G., Ledoux, M. Poincaré’s inequalities and Talagrand’s concentration phenomenon for the exponential distribution. *Probab. Theory and Related Fields* **107** (1997), 383–400.
- [B-L1] Bollobás, B., Leader, I. An isoperimetric inequality on the discrete torus. *SIAM J. Discrete Math.* **3** (1990), 32–37.
- [B-L2] Bollobás B., Leader, I. Compressions and Isoperimetric Inequalities, *J. Comb. Theory (Series A)* **56** (1991), 47–62.
- [B-L3] Bollobás, B., Leader, I. Isoperimetric inequalities and fractional set systems, *J. Comb. Theory (Series A)* **56** (1991), 63–74.
- [D-SC] Diaconis, P., Saloff-Coste, L. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. of Appl. Probab.* **6** (1996), 695–750.
- [Ha1] Harper, L.H. Optimal numberings and isoperimetric problems on graphs, *J. Combinatorial Theory* **1** (1966), 385–393.
- [Ha2] Harper, L.H. On an isoperimetric problem for Hamming graphs. Proceedings of the Conference on Optimal Discrete Structures and Algorithms—ODSA ’97 (Rostock). *Discrete Appl. Math.* 95 (1999), no. 1–3, 285–309.
- [Ho] Houdré, C. Mixed and isoperimetric estimates on the Log–Sobolev constants of graphs and Markov chains. *Combinatorica* **21** (2001).
- [H-S] Houdré, C., Stoyanov, S. Improved estimates on the log–Sobolev constants of graphs. Preprint (2001).
- [H-T] Houdré, C., Tetali, P. Concentration of measure for products of Markov kernels via functional inequalities. *Combin. Probab. Comput.* **10** (2001), 1–28.
- [J-S] Jogdeo, K., Samuels, S.M. Monotone convergence of binomial probabilities and a generalization of Ramanujan’s equation, *Ann. of Math. Stat.* **39** (1968), 1191–1195.
- [K] Karakhanyan, V. M. A discrete isoperimetric problem on a multidimensional torus. (Russian) *Akad. Nauk Armyan. SSR Dokl.* **74** (1982), 61–65.

- [L] Ledoux, M. Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités, XXXIII, 120–216, *Lecture Notes in Math.*, **1709**, Springer, Berlin, 1999.
- [M] Maurey, B. Construction de suites symétriques. *C. R. Acad. Sc. Paris. Ser. A-B* **288** (1979), A679–681.
- [McD] McDiarmid, C. On the method of bounded differences. Surveys in Combinatorics (1989). *London Math. Society Lecture Notes*, vol. 141, Cambridge Univ. Press, 148–188.
- [M-S] Milman, V.D., Schechtman, G. Asymptotic theory of finite dimensional normed spaces. *Lecture Notes in Math.* **1200** (1986), Springer, Berlin.
- [P] Pisier, G. Probabilistic methods in the geometry of Banach spaces. Probability and Analysis, Varenna (Italy) 1985. *Lecture Notes Math.* **1206** (1986), 167–241.
- [R] Riordan, O. An ordering on the even discrete torus. *SIAM J. Discrete Math.* **11** (1998), no. 1, 110–127
- [SC] Saloff-Coste, L. Lectures on finite Markov chains. *Ecole d'Été de Probabilités de St-Flour* (1996); 301–413, *Lecture Notes in Math.*, **1665**, Springer, Berlin, 1997.
- [S] Schechtman, G. Lévy type inequalities for a class of metric spaces. *Martingale Theory in Harmonic Analysis and Banach spaces*, Lecture Notes in Math., **939**, Springer-Verlag, 1981, 211–215.
- [S-T] Sammer, M., Tetali, P. Transportation and Concentration, preprint (February, 2004).
- [Ta1] Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* No. **81** (1995), 73–205.
- [Ta2] Talagrand, M. Isoperimetry and integrability of the sum of independent Banach space valued random variables, *Ann. Probab.* **17** (1989), 1546–1570.