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 various graph constants. 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

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

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 (1.1), which we call the *subgaussian constant* of the graph, is related to the so-called *spread constant*

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

[*]School of Mathematics, University of Minnesota, Minneapolis, MN 55455;research supported in part by NSF Grant No. DMS–0405587 and by EPSRC Visiting Fellowship.

[†]Laboratoire d'Analyse et de Mathematiques Appliquées, CNRS, UMR 8050, Université Paris XII, 94010 Créteil Cedex, France, and 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 and College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0160;research supported in part by NSF Grant No. DMS–0401239.

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

studied in [A-B-S] (cf. also [B-H1]). 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 $\mathrm{E}f$. The advantage of the subgaussian constant is however the fact that it is responsible for the subgaussian tails of Lipschitz functions: it follows from (1.1) that, for all $h > 0$,

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

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

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

relating (cf. [B-G]) 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 $D(\nu\|\pi)$ (or the relative entropy or Kullback-Leibler "distance" of ν with respect to π).

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

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

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, $\mathrm{E}f \leq c(G) \leq \sigma(G)$, 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

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

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 [Ta1].) Is it indeed the case that for every $A \subset \Omega^n$ with $\mu(A) \geq 1/2$,

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

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

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

Talagrand also remarked that using certain more or less standard arguments (such as those used in [Ta2]) it suffices to restrict the problem to down-sets (also called hereditary sets) $A \subset \Omega^n = \{0,1\}^n$ equipped with the product measure μ, where $\mu_i = \pi_p, i = 1, \ldots, n$ is the Bernoulli measure with success probability p for $0 < p < 1$. Indeed, Steps 1 through 4 of the proof of Theorem 7 in [Ta2] 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 2) 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 and Samuels [J-S] as well as Bollobás and Leader [B-L3].

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 (1.4)) the property (1.1) tensorizes. Combining (1.4) and (1.6), 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 $\pi^n(A) \geq 1/2$,*

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

Inequalities such as (1.4) and (1.7) are well-known in several situations. In the present work, we obtain inequalities of this type by computing or estimating the subgaussian constant for the following graphs:

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

6. All of the above (section 7).

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, while 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 $h = 1, 2, \ldots$ is an extremely nontrivial problem; in particular, for C_v 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 [Ha2]). Bollobás and Leader found the extremal sets for the n-dimensional grid graph P_v^n (see [B-L2]), while for the n-dimensional discrete torus C_v^n, v even, this is due to Karakhanyan [K], Bollobás and Leader [B-L1] and Riordan [R]. (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, $\operatorname{Var} f = \lim_{t\to 0} \mathrm{E}\, \frac{e^{t(f-\mathrm{E}f)}-1}{t^2/2}$, $c^2(G) \le \sigma^2(G)$. We will show cases of equality for some of the above examples. In general, this inequality is strict, the weighted two-point space, and the unweighted three-point space (i.e., the complete graph on three vertices with uniform probability over the vertices) provide examples of strict inequality. Furthermore, it has recently been established (see [S-T]) that the family of bounded-degree *expander graphs* – bounded-degree graphs with spectral gap bounded uniformly (independent of the size of the graph) away from zero – provides a class of examples for which the spread constant c^2 is bounded from above by an absolute constant, while the subgaussian σ^2 could grow at least as large as $\log v$ (up to a universal multiplicative constant), where v is the number of vertices in the expander graph. Since a randomly chosen bounded-degree graph is an expander graph, asymptotically almost surely, we may conclude that the inequality is in fact typically strict.

To start with, introduce the functions

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

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 [A-B-S] 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 $\pi = \dfrac{\delta_0 + \delta_1}{2}$ which is a particular case of the examples 2–4 described above when $v = 2$. We denote this graph by $K_2(p)$. It is well known, and due to Hoeffding, that $\sigma^2(K_2(1/2)) = 1/4$ (cf. e.g., [McD]). Moreover, since the seminal work of Harper [Ha1], 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.

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

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

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

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

The above constant (2.2) was first computed in an unpublished work [B]. 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 $2\sigma^2$ in (2.2) is defined, as the limit as $p \to 1/2$, to be $(f(1) - f(0))^2/2$. This value maximizes the right hand side of (2.2) over all p's. In particular,

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

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

Introducing the entropy functional

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

the proof of Proposition 2.1 relies on:

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

$$c \,\mathrm{Var}\, g \leq \mathrm{E} g \;\; \mathrm{Ent}\, g, \tag{2.3}$$

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 $\lim_{p\to 1/2} c_p = 1/2$. As we will also see, (2.3) becomes equality for the function g such that $g(0) = p/q$, $g(1) = q/p$.

The inequality (2.3) can be viewed as a converse to the general inequality $\mathrm{Var}\, g \geq \mathrm{E}g\ \mathrm{Ent}\, g$ which holds on an arbitrary probability space. As for the constant, it is the same as the optimal c in

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

where $|\nabla g| = |g(1) - g(0)|$ (cf. [D-SC], [S]); however, the relationship with (2.3) is not that transparent.

Proof Set $g(1) = a \geq 0$, $g(0) = b \geq 0$, and without loss of generality assume that in (2.3)

$$\mathrm{E}g = pa + qb = 1, \tag{2.4}$$

so that (2.3) takes the form

$$c\, pq(b-a)^2 \leq pa \log a + qb \log b.$$

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 (2.4). 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 $\varphi'(b) = 0 \Longleftrightarrow \psi(b) = 0$, 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 2.2 is proved.

Proof of Proposition 2.1 We may assume that $\mathrm{E}f = 0$ and that $f(0) \neq f(1)$ so that a priori $\sigma > 0$. The entropy functional has the following well-known representation:

$$\mathrm{Ent}\, g = \sup\{\mathrm{E}\, ug : \mathrm{E}e^u \leq 1\}.$$

Hence, the inequality $\mathrm{E}e^u \leq 1$ is equivalent to $\mathrm{E}ug \leq \mathrm{Ent}\, g$, where $g \geq 0$ is arbitrary. Applying this to $u = tf - \sigma^2 t^2/2$, we see that (2.1) is equivalent to

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

For a fixed g with $\mathrm{E}g > 0$, the above left hand side is maximized for

$$t = \frac{\mathrm{E}fg}{\sigma^2\, \mathrm{E}g}\,, \tag{2.6}$$

and (2.5) becomes

$$(\mathrm{E}fg)^2 \leq 2\sigma^2\, \mathrm{E}g\ \mathrm{Ent}\, g, \quad g \geq 0.$$

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

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

where g is an arbitrary nonnegative function on $\{0, 1\}$. It remains to apply Lemma 2.2 to obtain Proposition 2.1.

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 2.3 to the Lipschitz function $f(x) = x_1 + \ldots + x_n$ on V, implies via (1.2) a subgaussian deviation inequality for the number S_n of successes in n independent Bernoulli trials with success probability p:

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

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 (1.5), and for the sake of completeness we first state the results of Jogdeo and Samuels and of Bollobás and Leader we need. Theorem 3.2 and Corollary 3.1 of [J-S] assert the following: Consider n independent Bernoulli trials, the ith 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 [B-L3] 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

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

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

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

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 [J-S], 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 $(0, \ldots, 0)$. Now using (2.8) 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 (2.7). This proves the result with $K = e^4$. Actually, using the first inequality in (2.7) and for $p \neq 1/2$, the exponent -2 can be improved to $-\dfrac{\log p - \log q}{p - q}$.

Since for $\{0,1\}^n$, with $p = 1/2$, we have $\sigma^2 = n/4$, (2.9) can be rewritten as

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

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 (1.3), 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 2.1).

Indeed, assume $(1-p)^n = 1/2$ so that p is of order $\dfrac{\log 2}{n} + O\left(\dfrac{1}{n^2}\right)$, as $n \to \infty$. By Proposition 2.3, the inequality (2.10), 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 $A = \{(0,\ldots,0)\}$. For $h = n-1$, we thus have $A^h = \{0,1\}^n \setminus \{(1,\ldots,1)\}$ and the last inequality simplifies to

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

But, a simple Taylor expansion shows that the main term on the left of (2.11) is $-2p\log p$ which is of order $\dfrac{2\log 2\,\log n}{n}$, and so cannot be bounded by the right hand side of (2.11).

It is standard that (2.10), for $p = 1/2$, implies

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

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 (2.12) (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 2.3. 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 (2.9) 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 2.4. 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

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

3 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 $x, y \in V$,

$$|f(x) - f(y)| \leq 1. \tag{3.1}$$

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 3.1

$$\sigma^2(K_v, \mu) = \frac{p - q}{2(\log p - \log q)}, \tag{3.2}$$

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 $v = \operatorname{card} V = 2r$, we have $p(\pi) = 1/2$, and when $v = 2r + 1$, we have $p(\pi) = (r+1)/(2r+1)$. Therefore:

Corollary 3.2 *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*

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

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 $h \geq \sqrt{\frac{n}{2(2r+1)\log\frac{r+1}{r}}}$,

$$\pi^n\{x \in V^n : d(x, A) \geq h\} \leq \left(\frac{r}{r+1}\right)^{\frac{(2r+1)}{n}\left(h - \sqrt{\frac{n}{2(2r+1)\log\frac{r+1}{r}}}\right)^2}. \tag{3.3}$$

When $v = 3$, this yields the second part of Proposition 5.2 below, since $K_3 = C_3$. If $v = 2r$, the inequality (3.3) should be replaced by the slightly weaker inequality

$$\pi^n\{x \in V^n : d(x, A) \geq h\} \leq \exp\left\{-\frac{(2h - \sqrt{n})^2}{2n}\right\}, \qquad 2h \geq \sqrt{n}. \tag{3.4}$$

Note that since the complete graph is a Hamming space, (2.9) is slightly better than (3.4). 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 [A-B-S]), 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 3.1 The functional $f \to L(t, f) = \mathrm{E}_\mu e^{t(f - \mathrm{E}_\mu f)}$ is translation invariant, so maximizing this functional in the class of functions satisfying (3.1), 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 $f \to L(t, f)$ 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

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

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

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

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 2.1,

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

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 (3.2), it remains to show that the function $u(p) = \dfrac{p - q}{\log p - \log q}$ is decreasing in $(1/2, 1)$. Let

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

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

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

Since $w'(p) = -2(p-q)(\log p - \log q) < 0$, 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 3.1 is proved.

4 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 $\mathrm{E}e^{t(f - \mathrm{E}f)}$ is maximized for the identity function

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

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

Proposition 4.1 *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 $t \in \mathbf{R}$,

$$\mathrm{E}_\mu e^{t(f - \mathrm{E}_\mu f)} \leq \mathrm{E}_\mu e^{t(f^* - \mathrm{E}_\mu f^*)}. \tag{4.1}$$

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 4.1. Therefore, $L_G(t) = L_G(t, f^*)$, and since $\mathrm{E}_\pi f^* = (v+1)/2$, it follows that

$$L_G(t, f^*) \;=\; \frac{1}{v} \sum_{x=1}^{v} e^{tx} \, e^{-t(v+1)/2} = e^t \, \frac{e^{vt} - 1}{v(e^t - 1)} \, e^{-t(v+1)/2} \;=\; \frac{\mathrm{sh}(vt/2)}{v \, \mathrm{sh}(t/2)}.$$

Thus:

Proposition 4.2 *For any Lipschitz function f on the v-path P_v and for all $t \in \mathbf{R}$,*

$$\mathrm{E}e^{t(f-\mathrm{E}f)} \le \frac{\mathrm{sh}(vt/2)}{v\ \mathrm{sh}(t/2)},$$

the expectations being with respect to the uniform measure. In turn, $\sigma^2(P_v) = (v^2 - 1)/12$.

Remark 4.3 For every probability measure μ on $\mathbf{R}$ and every Lipschitz function f on $\mathbf{R}$,

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

Therefore, the spread constant of the v-path is

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

Thus, $c^2(P_v) = \sigma^2(P_v)$. According to (1.4)–(1.7), we obtain:

Proposition 4.4 *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 $\pi^n(A) \ge 1/2$,*

$$1 - \pi^n(A^h) \le \exp\left\{-\frac{6\left(h+1-\sqrt{\frac{v^2-1}{12}\,n}\right)^2}{n\,(v^2-1)}\right\}, \quad h+1 \ge \sqrt{\frac{v^2-1}{12}n},\ h = 0,1,2\ldots. \tag{4.2}$$

As shown by Bollobás and Leader, a result stronger than Proposition 4.4 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 (4.2).) Using Corollary 14 of [B-L2], we can also write Proposition 4.4 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 $V^n = \{0, 1, \ldots, v-1\}^n$ denote the set of vertices of the n-th power of a v-path, let

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

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

Lemma 4.5 ([B-L2]) *Let $G_1, \ldots, G_n$ 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 $|A| \ge b_v^{(n)}(r)$,*

$$|A^h| \ge b_v^{(n)}(r+h), \qquad \textit{for all } h.$$

Together Proposition 4.4 and Lemma 4.5 imply as claimed:

Proposition 4.6 *Let $G^n = G_1 \Box \cdots \Box G_n$. Then for all $A \subset G^n$ such that $\pi^n(A) \geq 1/2$,*

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

Perhaps it is to be remarked that Bollobás and Leader obtain their stronger version of Proposition 4.6 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 4.4 directly, with a functional-analytic approach, without having to find the extremal sets.

The following corollary, whose proof makes use of Proposition 4.2, guarantees that the v-path is extremal for the subgaussian constant (and hence the spread constant) among all graphs on v vertices:

Corollary 4.7 *For any connected undirected graph G on v vertices, $\sigma^2(G) \leq \dfrac{v^2-1}{12}$.*

Proof For a graph G, by a Lipschitz function *with respect to* G, we mean a function which is Lipschtiz with respect to the usual graph metric given by G. The proof follows from the observation that any function which is Lipschitz with respect to an arbitrary G, with v vertices, is also Lipschitz with respect to the v-path 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 P_v. Now the result follows from Proposition 4.2, which establishes the extremality of the identity function f^* on P_v (see in particular (4.1)).

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.

We can now pass to the proofs of Propositions 4.1 and 4.2. As shown in [B-H2], 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 (4.1) are well-defined. First we establish this inequality for monotone Lipschitz functions.

Lemma 4.8 *Let μ be a probability measure on $\mathbf{R}$ with finite first moment, i.e.,* $\mathrm{E}_\mu|f^*| = \int_{\mathbf{R}} |x| d\mu(x) < +\infty$. *Then,* (4.1) *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* (4.1) *holds for all $t \in \mathbf{R}$.*

In general (4.1) is not true for all Lipschitz functions f on $\mathbf{R}$, even if μ is symmetric. A simple counterexample to (4.1) 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 4.1, will be used.

Proof of Lemma 4.8 We use the following well-known functional representation of the entropy: for every measurable function h defined on some probability space,

$$\log \mathrm{E}e^h = \sup_{\mathrm{E}g=1} \left[\mathrm{E}gh - \mathrm{E}g \log g\right], \tag{4.3}$$

where the sup is taken over all measurable non-negative (and for simplicity bounded) functions g with $\mathrm{E}g = 1$. Clearly, this supremum is attained at $g = e^h/\mathrm{E}e^h$. 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 (4.3). In particular, applying (4.3) to the nondecreasing function $h = t(f - \mathrm{E}_\mu f)$, we have a representation in terms of the covariances $\mathrm{cov}_\mu(f, g) = \mathrm{E}fg - \mathrm{E}f\mathrm{E}g$:

$$\log \mathrm{E}_\mu e^{t(f - \mathrm{E}f)} = \sup_{\mathrm{E}g=1} \left[t \,\mathrm{cov}(f, g) - \mathrm{E}g \log g\right], \quad g \ \text{ nondecreasing.} \tag{4.4}$$

Now use

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

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, $\mathrm{cov}_\mu(f, g) \leq \mathrm{cov}_\mu(f^*, g)$, and thus, by (4.4), $\log \mathrm{E}_\mu e^{t(f - \mathrm{E}f)} \leq \log \mathrm{E}_\mu e^{t(f^* - \mathrm{E}f^*)}$. This proves Lemma 4.8 (the second statement is trivial since then $f^* - \mathrm{E}_\mu f^*$ and $\mathrm{E}_\mu f^* - f^*$ are identically distributed).

Clearly, Lemma 4.8 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 $t \geq 0$,

$$\mathrm{E}_\mu e^{t(\eta - \mathrm{E}_\mu \eta)} \leq \mathrm{E}_\mu e^{t(\xi - \mathrm{E}_\mu \xi)}. \tag{4.5}$$

In order to check the assumptions of this statement and thus get (4.5), one may use the following characterization proved in [B-H2] (Proposition 2.6 therein):

Lemma 4.9 *Given two random variables ξ and η on (Ω, μ), the existence of a non-decreasing Lipschitz function f from $\mathbf{R}$ to $\mathbf{R}$ such that η and $f(\xi)$ are identically distributed is equivalent to the inequality*

$$\mu\{\eta \le m_p(\eta) + h\} \ge \mu\{\xi \le m_p(\xi) + h\}, \tag{4.6}$$

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 4.1 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), since 2) applied to the function $f = -f^*$ implies 1). In order to derive 2) from the extremality of the half-axes, we use Lemma 4.9. Indeed, given a Lipschitz function f on $\mathbf{R}$, let

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

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

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

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

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

So, (4.6) and thus (4.5) hold for $\eta = f$ and $\xi = f^*$. Proposition 4.1 is proved.

Proof of Proposition 4.2 It remains to show that the optimal value of $\sigma^2 = \sigma^2(P_v)$ in

$$\frac{\mathrm{sh}(vt/2)}{v\ \mathrm{sh}(t/2)} \le e^{\sigma^2 t^2/2}, \quad t \in \mathbf{R}, \tag{4.7}$$

is $(v^2 - 1)/12$. Taking logarithm of both sides in (4.7) and setting $s = t/2$, we need to find the optimal constant σ^2 satisfying the inequality

$$\varphi(s) = \log(\mathrm{sh}(vs)) - \log(\mathrm{sh}(s)) - 2\sigma^2 s^2 \le \log v, \quad s \ge 0, \tag{4.8}$$

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

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

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

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

Therefore, (4.8) implies $\varphi''(0^+) \le 0$, i.e., for the optimal σ^2 in (4.8), we have $\sigma^2 \ge \frac{v^2-1}{12}$. It remains to show that $\sigma^2 = \frac{v^2-1}{12}$ satisfies (4.8). To do so, it is enough to show that φ is concave, i.e., $\varphi''(s) \le 0$, for all $s > 0$. Since $\varphi''(0^+) = 0$, it is sufficient to show that, for all $s > 0$, $\varphi'''(s) \le 0$, i.e.,

$$\frac{(vs)^3\mathrm{ch}(vs)}{\mathrm{sh}^3(vs)} \le \frac{s^3\mathrm{ch}(s)}{\mathrm{sh}^3(s)}.$$

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

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

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

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

which can equivalently be rewritten as

$$u(s) = 3\,\mathrm{sh}(s)\mathrm{ch}(s) - s(1+\mathrm{ch}^2(s)) \le 0, \quad s \ge 0. \tag{4.9}$$

Since $u(0) = 0$ and $u'(s) = 4\,\mathrm{sh}(s)\,(\mathrm{sh}(s) - s\,\mathrm{ch}(s)) \le 0$, u is non-increasing in $s \ge 0$. This proves (4.9) and thus Proposition 4.2.

5 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

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

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

$$|f(x_k) - f(x_{k-1})| \leq 1, \tag{5.1}$$

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

Conjecture 5.1 *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*

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

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 the conjecture to be true for $v \leq 4$. We present below the proof for the case $v = 3$, in part to illustrate the fact that the corresponding constant $\sigma^2(C_3)$ is distinct from the spread constant $c^2(C_3)$. More importanly, the general observations made as part of this proof lead to verifying the conjecture for v even (see Remark 5.5 below).

First, let us simplify the problem of maximization of the functional $f \to L_G(t, f)$, $f \in \mathcal{F}(G)$. 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 (5.1), with the vector

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

such that

$$y_1 + \cdots + y_v = 0. \tag{5.2}$$

Thus, the map $T : y \to f$ allows us to identify $\mathcal{F}_0(G)$ with the intersection M_0 of the cube $[-1, 1]^v$ with the hyperplane defined by (5.2). But, as easily seen (and proved), when $v = 2n$ is even, the extreme points of M_0 are the sequences

$$y = (\pm 1, \cdots, \pm 1) \tag{5.3}$$

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

$$y = (\pm 1, \cdots, \pm 1, 0, \pm 1, \cdots, \pm 1) \tag{5.4}$$

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

$$f(x_k) - f(x_{k-1}) = \pm 1, \tag{5.5}$$

for all $k = 1, \dots, v$ in case $v = 2n$, and in case $v = 2n+1$, with the property that (5.5) holds for all $k = 1, \dots, v$ except for 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 $y \in ex(M_0)$ as in (5.3)–(5.4) whose transforms by shift operation form the whole set $ex(M_0)$.

For example, when $v = 3$, up to the shift transformation there exist only two extremal sequences $(0,1,-1)$ and $(0,-1,1)$, 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 $v = 3$:

$$L_G(t) = \max\{L_G(t,f), L_G(-t,f)\}, \quad \text{where} \quad f(x) = d(x,x_0),\ x \in V. \tag{5.6}$$

Now, $\mathrm{E}f = \dfrac{f(x_0)+f(x_1)+f(x_1)}{3} = \dfrac{2}{3}$, $\ L_G(t,f) = \mathrm{E}e^{tf}\ e^{-2t/3} = \dfrac{e^{-2t/3}+2e^{t/3}}{3}$.

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

Proposition 5.2 *For the 3-cycle $G = C_3$, we have*

$$L_G(t) = \frac{e^{2|t|/3} + 2e^{-|t|/3}}{3}, \quad t \in \mathbf{R}. \quad \textit{In particular,} \quad \sigma^2(C_3) = \frac{1}{6\log 2}.$$

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

$$L_G(t) = \frac{1}{3}\,e^{2t/3} + \frac{2}{3}\,e^{-t/3} \le e^{\sigma^2 t^2/2}, \qquad t \ge 0.$$

Note that $L_G(t) = \mathrm{E}e^{t(\xi - \mathrm{E}\xi)}$ 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 2.1,

$$2\sigma^2 = \frac{p-q}{\log p - \log q}. \tag{5.7}$$

When $p = 1/3$, $2\sigma^2 = \dfrac{1}{3\log 2}$. Proposition 5.2 is proved.

Remark 5.3 The functional $f \to \operatorname{Var} f$ is convex, hence it attains its maximum in $\mathcal{F}(G)$ at the function $f(x) = d(x, x_0)$ (since $\operatorname{Var}(-f) = \operatorname{Var} f$ and since the functions f and $-f$ are the only functions to be considered as explained above). Thus,

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

We omit the proof of the conjecture for $\sigma^2(C_4)$ since it follows, using the tensoring property, that $\sigma^2(C_4) = 2\sigma^2(P_2) = 1/2$ – observe that $C_4 = P_2 \square P_2$, the Cartesian product graph of P_2 with itself. In fact, we have:

Proposition 5.4 *For the* 4*-cycle* C_4, $\quad L_G(t) = \dfrac{1+\mathrm{ch}(t)}{2}, \quad t \in \mathbf{R}.$

In particular, $\quad c^2(C_4) = \sigma^2(C_4) = \dfrac{1}{2}.$

Remark 5.5 C. McDiarmid pointed out to us that in the case of v even, 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 (5.5) above) that we restrict ourselves to Lipschitz f whose range lies in $R = \{0, 1, \ldots, v/2\}$. Then either every value in $R \setminus \{0, v/2\}$ 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 $1 \le k \le v$; but now (5.5) shows that such a g can not be extremal!

Unfortunately, the above argument does not seem to extend to the *odd* case. However, recently Marcus Sammer and the last author showed [S-T], using the transport formulation (mentioned in the introduction) of the subgaussian, that for $v \ge 3$,

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

To conclude this section, we state a result which corresponds to Conjecture 5.1 when maximizing the variance. The proof requires some tedious case analysis and can be found in [BHT].

Proposition 5.6 *In the class of all Lipschitz functions f on the v-cycle $G = C_v$,* 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,*

$$c^2(C_v) = \begin{cases} \dfrac{v^2+8}{48}, & \text{if } v \text{ is even,} \\ \\ \dfrac{(v^2-1)(v^2+3)}{48v^2}, & \text{if } v \text{ is odd.} \end{cases} \tag{5.8}$$

6 Symmetric group

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

$$d_v(x, y) = \text{card}\{i \le v : x_i \ne y_i\}, \qquad x, y \in S_v. \tag{6.1}$$

Proposition 6.1 *For all $v \ge 2$, $\sigma^2(S_v, d_v) \le v - 1$. In other words, for every d_v-Lipschitz function f on S_v and all $t \in \mathbf{R}$,*

$$\mathrm{E} e^{t(f - \mathrm{E}f)} \le e^{(v-1)\, t^2/2}, \tag{6.2}$$

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 $\pi_v(A) \ge 1/2$, and all integer $h \ge \sqrt{v-1}$,

$$\pi_v\{x \in S_v : d_v(A, x) \ge h\} \le \exp\left\{-\frac{(h - \sqrt{v-1})^2}{2(v-1)}\right\}. \tag{6.3}$$

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

$$\pi_v\{x \in S_v : d_v(x, A) \ge h\} \le \exp\left\{-\frac{\left(\frac{h}{2} - 2\sqrt{v \log(1/p)}\right)^2}{4v}\right\}. \tag{6.4}$$

When $p = 1/2$, (6.3) slightly improves upon (6.4).

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 $k \ne i, j$. 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 that $d_v(x, s_{i,j}x) = 2$ and that the metric cannot take the value 1. There is another natural metric to consider on the permutations: The usual 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 $z_1, \cdots, z_k$ 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 $1 \leq i < j \leq v$,

$$|f(s_{i,j}x) - f(x)| \leq 1. \tag{6.5}$$

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

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

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 (6.3),

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

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 $(123456..(v-1)v)$ and $(214365...v(v-1))$, for v even. Then $d_v = v$ and $\rho_v = v/2$. Thus (6.7) implies (6.3). 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 6.2 *For all* $v \geq 2$, $\dfrac{v-1}{16} \leq c^2(S_v, \rho_v) \leq \sigma^2(S_v, \rho_v) \leq \dfrac{v-1}{4}$.
In other words, the upper bound shows that for every ρ_v-Lipschitz function f on S_v and all $t \in \mathbf{R}$,

$$\mathrm{E}e^{t(f - \mathrm{E}f)} \leq e^{(v-1)\, t^2/8}, \tag{6.6}$$

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 $h \geq \frac{1}{2}\sqrt{v-1}$,

$$\pi_v\{x \in S_v : \rho_v(A, x) \geq h\} \leq \exp\left\{-\frac{(2h - \sqrt{v-1})^2}{2(v-1)}\right\}. \tag{6.7}$$

The inequality (6.6) appears in a paper of McDiarmid [McD] inside a martingale-based proof of a version of the concentration inequality (6.7) (see the proof of Theorem 6.7 and Example 7.1 there). The inequality (6.2) can also be proved using Hoeffding's inequality (see Page 18 of [Hoef]) for bounded martingale differences. The method has been popularised by the work of Maurey [M] and has been used since then by many authors, cf. e.g., [Sc], [P], [M-S]. To date, the best estimate obtained by this method seems to be given by $\sigma^2(S_v, d_v) \le 4(v-1)$. Versions of Maurey's concentration result have also been recovered (using a different method) by Talagrand (see Section 5 in [Ta1]). However our proofs of (6.2) and (6.6) are elementary, simply using induction. Since the results are essentially known, we omit the proofs and refer the interested reader to [BHT].

Proof of the lower bound in Proposition 6.2.

It suffices to give a specific Lipschitz function f with $\mathrm{Var}\, f \ge \frac{v-1}{16}$. The cases $v=1$ and $v=2$ are obvious since $c^2(S_1,\rho_1)=0$, $c^2(S_2,\rho_2)=1/4$. Assume $v\ge 3$. For $x\in S_v$, consider a function $f(x)=\mathrm{card}\{i\le a : x_i \le b\}$, 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: $\mathrm{E}f = \frac{ab}{v}$, and since $f^2(x)=\sum_{i,j\le a} \mathbf{1}_{x_i\le b}\mathbf{1}_{x_j\le b}$, we also have $\mathrm{E}f^2 = \frac{ab}{v} + \frac{ab(a-1)(b-1)}{v(v-1)}$, where the expectation is with respect to the normalized counting measure. The above clearly yields

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

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

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

Proposition 6.2 is proved.

7 Log-Sobolev and subgaussian constants

As shown by Aida, Masuda and Shigekawa ([A-M-S]) and by Ledoux ([L]) (with an argument going back to Herbst), it is also possible to derive subgaussian concentration inequalities, whenever a log-Sobolev inequality holds. Based on these works and on ([B-L]), such derivations under log-Sobolev as well as Poincaré inequalities were given for products of Markov kernels and graph products in ([H-T]). It is thus quite appropriate to try to compare the log-Sobolev and subgaussian approaches. For 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

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

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

We state the following without a proof, since the derivation is straight forward using the so-called Herbst argument. Interested reader may find the proof in [BHT].

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

$$\mathrm{E}e^{t(f-\mathrm{E}f)} \le e^{\frac{t^2 \|\nabla_\infty^+ f\|_\infty^2}{4\rho_\infty^+}}, \quad t \in \mathbf{R}. \tag{7.1}$$

In particular,

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

We may return to the examples analyzed in the previous sections and observe that for the normalized counting measure, (7.2) provides estimates of σ^2 which are tight up to universal multiplicative constants. For example, for $G = (S_v, \rho_v)$, with π being the normalized counting measure, we have:

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

the upper bound follows from the computations in Example 4.3 in [D-SC] while the lower bound is easily obtained via generic lower bounds given in [Ho], [St]. Although in the examples of the previous sections, 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 (2.2). Hence, in that case, the concentration inequality obtained via the subgaussian constant is much stronger than the one obtained via the log–Sobolev constant.

Acknowledgement. We thank David Wilson for the idea regarding the lower bound in Proposition 6.2. The last author would also like to 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. Func. 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. Characterization of Gaussian measures in terms of 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. Sci.* (New York) **93** (1999), 270–275.

[BHT] Bobkov, S.G., Houdré, C., Tetali, P. The subguassian constant and concentration inequalities. Expanded version available at http://www.math.gatech.edu/∼tetali/PUBLIS/BHT_SUB_full.tex

[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.

[Hoef] Hoeffding, W. Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.* **58** (1963), 13–30.

[Ho] Houdré, C. Mixed and isoperimetric estimates on the Log–Sobolev constants of graphs and Markov chains. *Combinatorica* **21** (2001), 489–513.

[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

[S] Saloff-Coste, L. Lectures on finite Markov chains. *Ecole d'Eté de Probabilités de St-Flour* (1996); 301–413, *Lecture Notes in Math.,* **1665**, Springer, Berlin, 1997.

[S-T] Sammer, M., Tetali, P. Concentration on the Discrete Torus using Transportation, submitted (2005).

[Sc] 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.

[St] Stoyanov, S. Isoperimetric and Related Constants for Graphs and Markov chains. Ph.D. thesis (2001), Georgia Institute of Technology.

[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.