Bennett-Bernstein inequality and the spectral gap of random regular graphs

Pierre Youssef

Laboratoire de Probabilités et de Modèles aléatoires
Université Paris Diderot
E-mail: youssef@math.univ-paris-diderot.fr
http://www.lpma-paris.fr/pageperso/youssef/

Contents

These notes are based on a mini-course given by the author at the Lebanese University-Faculty of Science II- Fanar. They are meant to introduce a general audience to some concentration inequalities and the problem of estimating the spectral gap of a random graph. This short course (9 hours) was prepared with the intention to be accessible to first year master students with very basic knowledge in probability theory and without any prerequisite in graph theory. It will however get more technical as the course advances.

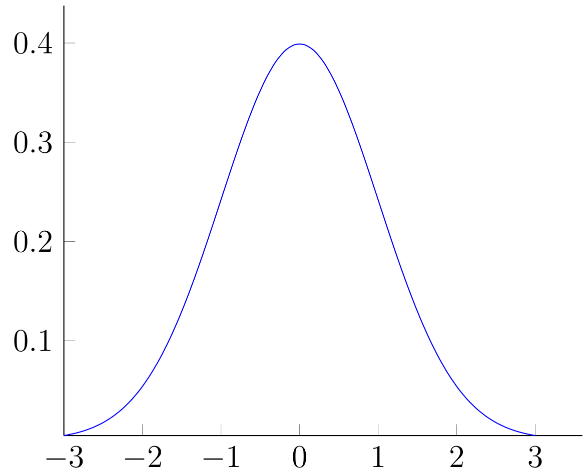
I would like to thank Pascal Lefèvre for arranging this opportunity for me to share some of my research interests at the Lebanese University. I also thank Ihab Alam and Georges Habib for their hospitality and for organizing everything on a very short notice. In preparing these notes, I have used for the graph theoretical part [5] (thanks to Justin Salez for suggesting and landing me the book) and [9] (we refer to references in [9] for a more complete exposition on the topic). I would also like to thank Marwa Banna for many valuable comments which helped improve these notes.

These notes are organized as follows. In the first part, we establish some concentration inequalities for sums of independent random variables. We focus on Bennett and Bernstein's inequalities which control the deviation from the mean in terms of the variance. We then give two applications of these inequalities: in the first, we establish structural properties of random Bernoulli matrices which will be used later; in the second, we apply Bernstein's inequality in order to prove the Johnson-Lindenstrauss flattening lemma, which says that n points in a Hilbert space can be embedded into a $\ln n$-dimensional space while keeping the distances between these points almost the same. We then introduce basic notions on d-regular graphs and their adjacency matrices. We consider the edge expansion constant of a d-regular graph and establish its relation to the spectral properties of the graph. We present Alon-Milman's result which asserts that a large gap between the largest and the second largest eigenvalue implies that the graph has good expansion properties, and vice versa. Finally, we study the Erdös-Renyi random graph and show that in the dense regime, it has a large spectral gap with high probability.

Everywhere in the text, we assume that n is a sufficiently large natural number. For a finite set I, by $|I|$ we denote its cardinality. For any positive integer m, the set $\{1, 2, \dots, m\}$ will be denoted by $[m]$. We denote by I^c the complement of I in the corresponding set. For a real number a, $\lfloor a \rfloor$ is the largest integer not exceeding a. For a vector $y \in \mathbb{R}^n$, we denote by $\operatorname{supp} y$ the set of indices of its non-zero coordinates i.e. $\operatorname{supp} y = \{i \in [n] : y_i \neq 0\}$. By $\langle \cdot, \cdot \rangle$ we denote the standard inner product in $\mathbb{R}^n$, by $\| \cdot \|_2$ — the standard Euclidean norm in $\mathbb{R}^n$. The notation $a \lesssim b$ (resp. $a \gtrsim b$) means that $a \leq c\, b$ (resp. $a \geq c\, b$) for some numerical constant c. Finally, $a \sim b$ means that $a \gtrsim b$ and $a \lesssim b$.

1 Concentration inequalities and applications

The concentration of measure phenomenon is a vast topic which has many applications in different fields. In these notes, we will investigate this phenomenon for a sum of independent random variables. Let us first illustrate the basic idea of concentration starting with the normal distribution. Given $g \sim \mathcal{N}(0, 1)$ a standard Gaussian random variable, its density function is given by $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$:



We can see that most of the mass is concentrated around the mean while the probability of being far from the mean decays very fast.

In general, there is no reason for the mass to be concentrated around the mean. Consider the experience of tossing a coin where one gets head or tail with equal probability 1/2. In this case, no outcome is privileged. Denote ξ the corresponding standard Bernoulli random variable i.e. $\mathbb{P}\{\xi = 1\} = \mathbb{P}\{\xi = 0\} = \frac{1}{2}$. Repeating this experience many times, one expects to get an equal number of head and tail. More precisely, if $\{\xi = 1\}$ denotes the event that "we got head" then the number of occurence of head in n experiences is equal to $\sum_{i \le n} \xi_i$ where $\xi_1, \ldots, \xi_n$ are independent standard Bernoulli random variables. From the law of large numbers, we know that

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i \underset{n \to \infty}{\longrightarrow} \mathbb{E}\, \xi = \frac{1}{2} \quad a.s.$$

Therefore, the more we repeat this experience, the closer the average number of head occurence will be to 1/2 and the phenomenon of concentration will be more apparent. One of the goals of concentration inequalities is to quantify the above convergence by finding the correct rate at which the average is approaching the mean. Once this is done, one can deduce the minimal number of experiences needed in order to get with high probability the same number of heads and tails.

1.1 Bennett-Bernstein's inequalities

The moment generating function of a random variable ξ is given by

$$M_\xi(\lambda) = \mathbb{E}\, e^{\lambda \xi}, \quad \lambda \in \mathbb{R},$$

whenever this expectation exists. We start with the Chernoff bound which is of similar flavor to Markov and Tchebysheff's inequalities.

Lemma 1.1 (Chernoff bound)**.** *Let S be a random variable. Then for any $t \in \mathbb{R}$*

$$\mathbb{P}\Big\{S \geq t\Big\} \leq \inf_{\lambda>0}\{e^{-\lambda t} M_S(\lambda)\}.$$

Proof. Fix $\lambda > 0$ and write

$$\mathbb{P}\big\{S \geq t\big\} = \mathbb{P}\big\{\lambda S \geq \lambda t\big\} = \mathbb{P}\Big\{e^{\lambda S} \geq e^{\lambda t}\Big\},$$

where we used that the exponential function is increasing. Now, notice that

$$\mathbb{P}\big\{e^{\lambda S} \geq e^{\lambda t}\big\} = \mathbb{E}\,\mathbf{1}_{\{e^{\lambda S} \geq e^{\lambda t}\}} = e^{-\lambda t}\,\mathbb{E}\,\mathbf{1}_{\{e^{\lambda S} \geq e^{\lambda t}\}}\, e^{\lambda t} \leq e^{-\lambda t}\, M_S(\lambda).$$

Since this is the true for any $\lambda > 0$, then the lemma follows. □

The previous lemma suggests that a good bound on the moment generating function of a random variable allows to control its tails. In the next lemma, we estimate the moment generating function of bounded random variables and those having nice moments growth.

Lemma 1.2. *Let $b \geq 0$ and let ξ be a centered random variable with $\mathbb{E}\,\xi^2 \leq \sigma^2$ for some positive number σ. Then we have*

1. *If $|\xi| \leq b$ a.s. then $M_\xi(\lambda) \leq \exp\left(\frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right)$ for any $\lambda > 0$.*

2. *If $\mathbb{E}\,\xi^k \leq \frac{1}{2}k!\,\sigma^2 b^{k-2}$ for any $k \geq 3$, then $M_\xi(\lambda) \leq \exp\left(\frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right)$ for any $\lambda \in (0, 1/b)$.*

Proof. Let $\lambda > 0$. From the power series expansion of the exponential function, we can write

$$M_\xi(\lambda) = 1 + \lambda\mathbb{E}\,\xi + \frac{1}{2}\lambda^2\mathbb{E}\,\xi^2 + \sum_{k\geq 3}\frac{\lambda^k\mathbb{E}\,\xi^k}{k!}.$$

Using that $\mathbb{E}\,\xi = 0$ and $\mathbb{E}\,\xi^2 \leq \sigma^2$, we get

$$M_\xi(\lambda) \leq 1 + \frac{1}{2}\lambda^2\sigma^2 + \sum_{k\geq 3}\frac{\lambda^k\mathbb{E}\,\xi^k}{k!}. \tag{1}$$

Now if $|\xi| \leq b$ a.s. then

$$\sum_{k\geq 3}\frac{\lambda^k\mathbb{E}\,\xi^k}{k!} \leq \sum_{k\geq 3}\frac{\lambda^k b^{k-2}\mathbb{E}\,\xi^2}{k!} \leq \frac{\sigma^2}{b^2}\sum_{k\geq 3}\frac{(\lambda b)^k}{k!} = \frac{\sigma^2}{b^2}(e^{\lambda b} - \frac{\lambda^2 b^2}{2} - \lambda b - 1),$$

which together with (1) implies that

$$M_\xi(\lambda) \leq 1 + \frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1) \leq \exp\left(\frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right),$$

and proves the first part of the lemma.

Now suppose that $\mathbb{E}\,\xi^k \le \frac{1}{2}k!\,\sigma^2 b^{k-2}$ for any $k \ge 3$. Injecting this assumption in (1), we deduce

$$M_\xi(\lambda) \le 1 + \frac{1}{2}\lambda^2\sigma^2 + \frac{1}{2}\lambda^2\sigma^2 \sum_{k\ge 3}(\lambda b)^{k-2}.$$

When $\lambda < 1/b$, the last term of the previous inequality is the sum of a convergent geometric series. Thus, we get for any $0 < \lambda < 1/b$

$$M_\xi(\lambda) \le 1 + \frac{1}{2}\lambda^2\sigma^2 + \frac{1}{2}\lambda^2\sigma^2 \frac{\lambda b}{1-\lambda b} = 1 + \frac{\lambda^2\sigma^2}{2(1-\lambda b)} \le \exp\left(\frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right),$$

which shows the second part of the lemma. □

Now that we estimated the MGF of bounded random variables, we will use this to establish the concentration of a sum of independent bounded random variables around their mean. Define a function $H(t)$ on the positive semi-axis as

$$H(t) := (1+t)\ln(1+t) - t. \tag{2}$$

Note that $H(\cdot)$ is increasing. Moreover, it is easy to check that

$$H(t) \ge \frac{t^2}{2(1+t/3)} \quad \text{for any } t \ge 0. \tag{3}$$

Theorem 1.3 (Bennett's inequality)**.** *Let $n \in \mathbb{N}$ and $\xi_1, \dots, \xi_n$ be independent centered random variables satisfying $|\xi_i| \le b$ and $\mathbb{E}\,\xi_i^2 \le \sigma_i^2$ for some positive numbers b and σ_i, $i = 1, \dots, n$. Then for any $t \ge 0$, we have*

$$\mathbb{P}\Big\{\big|\sum_{i\le n}\xi_i\big| \ge t\Big\} \le 2\exp\left(-\frac{\sigma^2}{b^2}H\Big(\frac{tb}{\sigma^2}\Big)\right).$$

Proof. Let $\lambda > 0$ and $i \le n$. Applying the first part of Lemma 1.2, we have

$$M_{\xi_i}(\lambda) \le \exp\left(\frac{\sigma_i^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right). \tag{4}$$

Now denote $S = \sum_{i\le n}\xi_i$ and note that

$$M_S(\lambda) = \mathbb{E}\,e^{\lambda S} = \mathbb{E}\prod_{i=1}^n e^{\lambda\xi_i} = \prod_{i=1}^n \mathbb{E}\,e^{\lambda\xi_i},$$

where we used the independence of the ξ_i's in the last equality. This, together with (4), implies that

$$M_S(\lambda) \le \prod_{i=1}^n \exp\left(\frac{\sigma_i^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right) = \exp\left(\frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right),$$

where $\sigma^2 = \sum_{i\le n}\sigma_i^2$. Using the above estimate together with Lemma 1.1, we get for any $t>0$

$$\mathbb{P}\Big\{\sum_{i\le n}\xi_i \ge t\Big\} \le \inf_{\lambda>0}\exp\left(-\lambda t + \frac{\sigma^2}{b^2}(e^{\lambda b}-\lambda b-1)\right).$$

The above expression is minimized for $\lambda = \frac{1}{b}\ln\big(1+bt/\sigma^2\big)$. Replacing this value, we deduce that

$$\mathbb{P}\Big\{\sum_{i\le n}\xi_i \ge t\Big\} \le \exp\left(-\frac{\sigma^2}{b^2}H\Big(\frac{tb}{\sigma^2}\Big)\right),$$

where H is the function defined in (2). Now reapplying the above with the random variables $-\xi_1,\dots,-\xi_n$, we get for any $t>0$

$$\mathbb{P}\Big\{\sum_{i\le n}\xi_i \le -t\Big\} \le \exp\left(-\frac{\sigma^2}{b^2}H\Big(\frac{tb}{\sigma^2}\Big)\right).$$

We finish the proof by using the two previous inequalities together with the fact that for any $t>0$

$$\mathbb{P}\Big\{\big|\sum_{i\le n}\xi_i\big| \ge t\Big\} = \mathbb{P}\Big\{\sum_{i\le n}\xi_i \ge t\Big\} + \mathbb{P}\Big\{\sum_{i\le n}\xi_i \le -t\Big\}.$$

□

Remark 1.4.

- Let us note that if the ξ_i's are not centered, one just needs to apply the previous statement with $\xi_i - \mathbb{E}\,\xi_i$ to obtain the same phenomenon of concentration of the sum of the ξ_i's around their expectation.
- Since $\big|\sum_{i\le n}\xi_i\big| \le nb$ with probability 1, then the concentration inequality makes sense whenever $t\le nb$. On the other hand, it has a non-trivial consequence whenever $2\exp\Big(-\frac{\sigma^2}{b^2}H\Big(\frac{tb}{\sigma^2}\Big)\Big) \le 1$ meaning for $t \ge \frac{\sigma^2}{b}H^{-1}\big(\frac{b^2\ln 2}{\sigma^2}\big)$. This holds when t is larger than a proportion of b.
- From (3), we get that the concentration inequality exhibits two behaviors depending on the regime we are interested in. When $t\le \sigma^2/b$, then the probability of deviating from the mean is bounded by $2\exp\big(-t^2/(4\sigma^2)\big)$ while when $t\ge\sigma^2/b$, this probability is bounded by $2\exp\big(-t/(4b)\big)$. Therefore, we have a subgaussian behavior (the tail inequality is similar to the Gaussian case) in the range where $t\le\sigma^2/b$ while in the other range we have a subexponential behavior.
- Going back to the experience of tossing a coin. Fix $\varepsilon\in(0,1)$ and let $n\in\mathbb{N}$ be the number of experiences one makes. If one wants to increase chances that the number of tails and heads differ by at most εn, then one should have

$$\mathbb{P}\Big\{\big|\sum_{i=1}^n\xi_i - \sum_{i=1}^n(1-\xi_i)\big| \le \varepsilon n\Big\} \ge \frac{1}{2}.$$

This translates to having

$$\mathbb{P}\Big\{\big|\sum_{i=1}^{n}(\xi_i - \mathbb{E}\,\xi_i)\big| \geq \frac{\varepsilon n}{2}\Big\} \leq \frac{1}{2},$$

which can be ensured if

$$2\exp\left(-\frac{n}{2}H(\varepsilon)\right) \leq \frac{1}{2}.$$

Thus one should repeat the experience at least $\ln 8/H(\varepsilon)$ times.

In the previous theorem, we only dealt with bounded random variables. We will see that this assumption can be dropped and replaced by a control on the moments.

Theorem 1.5 (Bernstein's inequality)**.** *Let $\xi_1, \ldots, \xi_n$ be independent centered random variables satisfying*

$$\mathbb{E}\xi_i^2 \leq \sigma_i^2 \quad \text{and} \quad \mathbb{E}|\xi_i|^k \leq \frac{1}{2}k!\,\sigma_i^2 b^{k-2} \text{ for any } k \geq 3,$$

for some positive numbers b and σ_i, $i = 1, \ldots, n$. Then for any $t > 0$

$$\mathbb{P}\Big\{\big|\sum_{i=1}^{n}\xi_i\big| \geq t\Big\} \leq 2\exp\left(-\frac{t^2}{2(\sigma^2 + tb)}\right),$$

where $\sigma^2 = \sum_{i=1}^{n}\sigma_i^2$.

Proof. Let $\lambda > 0$ and $i \leq n$. Applying the second part of Lemma 1.2, we have

$$M_{\xi_i}(\lambda) \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right) \tag{5}$$

Denote $S = \sum_{i\leq n}\xi_i$. Using the independence of the ξ_i's together with (5), we get

$$M_S(\lambda) \leq \exp\left(\frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right),$$

where $\sigma^2 = \sum_{i\leq n}\sigma_i^2$. This estimate, together with Lemma 1.1, implies that for any $t > 0$

$$\mathbb{P}\Big\{\sum_{i\leq n}\xi_i \geq t\Big\} \leq \inf_{\lambda>0}\exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right).$$

Taking $\lambda := \frac{t}{\sigma^2+bt}$ we deduce that

$$\mathbb{P}\Big\{\sum_{i\leq n}\xi_i \geq t\Big\} \leq \exp\left(-\frac{t^2}{2(\sigma^2 + tb)}\right).$$

To finish the proof, one should reapply the above for $-\xi_1, \ldots, -\xi_n$ and conclude in a similar manner to what is done in Theorem 1.3. $\square$

Let us finish this subsection by a natural question: "Can we see Bennett and Bernstein's inequalities as tensorization of single inequalities involving the cumulative distribution function of each variable?"

More presicely, given positive numbers b, $(\sigma_i)_{i\leq n}$ and $\xi_1,\dots,\xi_n$ centered random variables satisfying $\mathbb{E}\,\xi_i^2 \leq \sigma_i^2$ and

$$\mathbb{P}\Big\{|\xi_i| \geq t\Big\} \leq 2\exp\left(-\frac{\sigma_i^2}{b^2}H\big(\frac{tb}{\sigma_i^2}\big)\right),$$

for any $t\geq 0$. Is is true that for any $t\geq 0$

$$\mathbb{P}\Big\{|\sum_{i=1}^n \xi_i| \geq t\Big\} \leq 2\exp\left(-\frac{\sigma^2}{b^2}H\big(\frac{tb}{\sigma^2}\big)\right),$$

where $\sigma^2 = \sum_{i\leq n}\sigma_i^2$?

The same question can be asked for Bernstein's inequality. Eventhought this formulation is quite natural, it is not common in the literature. Indeed, we usually work with the moment generating function, take advantage of the fact that the mgf of a sum of independent random variables is the product of the individual mgf's, then use Lemma 1.1 to turn a bound on the mgf to a control on the cumulative distribution function. It would be interesting to know if the above question has an affirmative answer.

1.2 Structure of a symmetric Bernoulli matrix

Let $d,n\in\mathbb{N}$ and ξ be a random variable following the Bernoulli distribution with parameter d/n i.e. $\mathbb{P}\{\xi=1\} = d/n = 1-\mathbb{P}\{\xi=0\}$. Let B be an $n\times n$ symmetric random matrix with zero diagonal whose entries above the diagonal are independent copies of ξ i.e. entries of B above the diagonal are independent random variables following the Bernoulli distribution with parameter d/n. Our goal in this subsection is to understand the structure of such matrix. More precisely, we will see that B will inherit some of the structural properties satisfied by its expectation given by $\mathbb{E}\,B = \frac{d}{n}C$, where C denotes the $n\times n$ matrix with zero diagonal and all other entries equal to 1. One of the nice structural properties of $\mathbb{E}\,B$ is that it is an almost "d- double stochastic" matrix meaning that the entries in each row and column almost sum up to d (they sum up to $d(1-1/n)$). The next proposition shows how the concentration inequalities we established imply that B is "almost d-double stochastic" with high probability. Given $\varepsilon\in(0,1)$, we will say that an $n\times n$ matrix A with nonnegative entries is ε-almost d-double stochastic matrix if for any $1\leq i,j\leq n$

$$\frac{1}{d}\Big|\sum_{k=1}^n a_{kj} - d\Big| \underset{n\to\infty}{\longrightarrow} \varepsilon \quad\text{and}\quad \frac{1}{d}\Big|\sum_{k=1}^n a_{ik} - d\Big| \underset{n\to\infty}{\longrightarrow} \varepsilon.$$

Proposition 1.6. *For any $\varepsilon\in(0,1)$ and $d > 2\ln(2n)/H(\varepsilon)$, B is ε-almost d-double stochastic with probability going to one as n tends to ∞.*

Proof. Let us denote by $(\varepsilon_{ij})_{1\leq i,j\leq n}$ the entries of B and $\xi_{ij} = \varepsilon_{ij} - \frac{d}{n}$ for any $1 \leq i \neq j \leq n$. Clearly for any given $i \in [n]$, the random variables $(\xi_{ij})_{j\neq i}$ are independent and centered. Moreover, we have $|\xi_{ij}| \leq 1$ and $\mathbb{E}\,\xi_{ij}^2 \leq \frac{d}{n}$. Given $i \in \{1,\ldots,n\}$, let us denote by $\mathcal{E}_i$ the following event

$$\mathcal{E}_i = \Big\{(1-\varepsilon-1/n)d \leq \sum_{k=1}^{n} \varepsilon_{ik} \leq (1+\varepsilon-1/n)d\Big\} = \Big\{\big|\sum_{k\neq i} \xi_{ik}\big| \leq \varepsilon d\Big\}.$$

Fix for a moment $i \in \{1,\ldots,n\}$ and apply Theorem 1.3 for the random variables $(\xi_{ik})_{k\neq i}$ (with $b=1$, $\sigma_k^2 = d/n$ and $t = \varepsilon d$) to deduce that

$$\mathbb{P}\Big\{\mathcal{E}_i\Big\} \geq 1 - 2\exp\left(-dH(\varepsilon)\right).$$

Using this, together with the union bound, we get

$$\begin{aligned}\mathbb{P}\Big\{\bigcap_{1\leq i\leq n} \mathcal{E}_i\Big\} &= 1 - \mathbb{P}\Big\{\bigcup_{1\leq i\leq n} \mathcal{E}_i^c\Big\}\\ &\geq 1 - 2n\exp\left(-dH(\varepsilon)\right)\\ &\geq 1 - \frac{1}{2n},\end{aligned}$$

where in the last inequality we used the assumption on d. □

Remark 1.7. Note that when ε is small, we have $H(\varepsilon) \sim \varepsilon^2$. This dependence on ε will appear again in the next section when dealing with the Johnson-Lindenstrauss lemma. Let us note also that the use of concentration inequality in the above proposition is artificial and not needed since the sum of independent bernoulli variables has the binomial distribution which would allow us to compute explicitly the desired probabilities.

Another structural property of the matrix B is that the number of non-zero entries in a given block is comparable to a fixed proportion of the size of this block. More presicely, given $I, J \subset \{1,\ldots,n\}$, let us define

$$e(I,J) = \Big|\Big\{(i,j) \in I\times J :\, \varepsilon_{ij} \neq 0\Big\}\Big| = \sum_{(i,j)\in I\times J} \varepsilon_{ij}.$$

Note that $\mathbb{E}\,e(I,J) = \frac{d}{n}|I|\,|J|$. The next proposition establishes a structural property of B which will be useful for us later.

Proposition 1.8. *Let $K \geq 1$ and let $\mathcal{E}_{1.8}(K)$ be the event that for all subsets $S, T \subset [n]$ at least one of the following is true:*

$$e(S,T) \leq 3\,\frac{d}{n}\,|S|\,|T|, \tag{6}$$

or

$$e(S,T)\,\ln\left(\frac{e(S,T)}{\frac{d}{n}\,|S|\,|T|}\right) \leq 3(K+5)\max(|S|,|T|)\,\ln\left(\frac{e\,n}{\max(|S|,|T|)}\right). \tag{7}$$

Then we have

$$\mathbb{P}\Big\{\mathcal{E}_{1.8}(K)\Big\} \geq 1 - \frac{1}{n^K}.$$

Proof. We will drop in the proof the symmetry assumption on B, and therefore suppose that all entries of B are independent (it will be sufficient to imply the statement for the symmetric case since the proposition is concerned with large deviation of $e(S,T)$). Let $S,T\subset[n]$. For every $(i,j)\in S\times T$, define $\xi_{ij}=\varepsilon_{ij}-\frac{d}{n}$ so that

$$e(S,T)=\sum_{(i,j)\in S\times T}\xi_{ij}+\frac{d}{n}|S|\,|T|.$$

Note that the ξ_{ij}'s are independent and centered. Moreover, we have $|\xi_{ij}|\le 1$ and $\mathbb{E}\,\xi_{ij}^2\le\frac{d}{n}$. Let $r>0$. Applying Theorem 1.3 with $t=r\frac{d}{n}|S|\,|T|$, we get

$$\mathbb{P}\Big\{\Big|\sum_{(i,j)\in S\times T}\xi_{ij}\Big|\ge r\frac{d}{n}|S|\,|T|\Big\}\le 2\exp\left(-\frac{d}{n}|S|\,|T|H(r)\right).$$

In particular, we deduce that for any $S,T\subset[n]$ and any $r>0$

$$\mathbb{P}\Big\{e(S,T)\ge(1+r)\frac{d}{n}|S|\,|T|\Big\}\le 2\exp\left(-\frac{d}{n}|S|\,|T|H(r)\right). \tag{8}$$

Define

$$r_1=H^{-1}\left[\frac{(K+5)\max(|S|,|T|)}{\frac{d}{n}\,|S|\,|T|}\,\ln\left(\frac{e\,n}{\max(|S|,|T|)}\right)\right],$$

and

$$r_0:=\max(2,r_1).$$

Note that if $e(S,T)\le(1+r_0)\,\frac{d}{n}|S|\,|T|$ then either (6) or (7) holds. Indeed, if $r_0=2$ the first property clearly holds. Otherwise, if $r_0=r_1\ge 2$ then

$$\ln\left(\frac{e(S,T)}{\frac{d}{n}\,|S|\,|T|}\right)\le\ln(1+r_1).$$

Since $\ln(1+r_1)>0$, this implies that

$$e(S,T)\,\ln\left(\frac{(S,T)}{\frac{d}{n}\,|S|\,|T|}\right)\le\frac{d}{n}|S|\,|T|\,(1+r_1)\ln(1+r_1).$$

Using that $(1+r)\ln(1+r)\le 3\,H(r)$ for any $r\ge 2$, we deduce that

$$e(S,T)\,\ln\left(\frac{e(S,T)}{\frac{d}{n}\,|S|\,|T|}\right)\le 3\,\frac{d}{n}|S|\,|T|\,H(r_1)=3(K+5)\max(|S|,|T|)\,\ln\left(\frac{e\,n}{\max(|S|,|T|)}\right),$$

which means that (7) holds. From the above, we deduce that

$$\mathbb{P}\{\mathcal{E}_{1.8}^c\}\le\mathbb{P}\Big\{\exists S,T\subset[n]:\,e(S,T)>(1+r_0)\frac{d}{n}|S|\,|T|\Big\},$$

which after using (8) implies

$$\begin{aligned}
\mathbb{P}\{\mathcal{E}_{1.8}^c\} &\le 2\sum_{k,\ell=1}^{n}\binom{n}{k}\binom{n}{\ell}\exp\left(-\frac{d\,k\,\ell}{n}\,H(r_0)\right)\\
&\le 2\sum_{k,\ell=1}^{n}\exp\left(k\ln\big(\frac{en}{k}\big)+\ell\ln\big(\frac{en}{\ell}\big)-\frac{d\,k\,\ell}{n}\,H(r_1)\right)\\
&\le 2\sum_{k,\ell=1}^{n}\exp\left[-(K+3)\max(k,\ell)\,\ln\left(\frac{e\,n}{\max(k,\ell)}\right)\right]\\
&\le \frac{2}{n^{K+1}}\le\frac{1}{n^K},
\end{aligned}$$

where before the last inequality we used that $\max(k,\ell)\,\ln\left(\frac{e\,n}{\max(k,\ell)}\right)\ge \ln n$. □

Remark 1.9. Let us note that (6) is destined to large sets S and T while (7) takes care of the small ones. Indeed, we know that $e(S,T)$ is positive with probability equal to $1-(d/n)^{|S|\,|T|}$. Therefore, if $|S|\,|T|<n/(3d)$ then (6) fails with probability $1-(d/n)^{|S|\,|T|}$.

1.3 Johnson-Lindenstrauss flattening lemma

Another nice application of the concentration inequalities established above, is the Johnson-Lindenstrauss lemma. Given m points in a Hilbert space, one can naturally embed these points in $\mathbb{R}^m$ while preserving the Euclidean distances separating them. Our goal is to embed these points into a smaller dimensional space. It turns out that random projections are good embeddings, Gaussian operators as well. Using the concentration inequality established before, we will use Rademacher matrices and show that they allow to reduce the dimension of the space containing our original set of points.

Lemma 1.10. *Let $\varepsilon_1,\dots,\varepsilon_n$ be independent Rademacher random variables i.e. $\mathbb{P}\{\varepsilon_i=-1\}=\mathbb{P}\{\varepsilon_i=1\}=\frac{1}{2}$. Let $X:=(\varepsilon_1,\dots,\varepsilon_n)$ be the n-dimensional random vector whose coordinates are given by the ε_i's. Then for any $x\in S^{n-1}$, the random variable $\eta=\langle X,x\rangle^2$ satisfies*

$$\mathbb{E}\,\eta=1 \quad \text{and} \quad \mathbb{E}\,\eta^k\le e^{4k}k! \quad \text{for any } k\ge 2.$$

Proof. Let $x\in S^{n-1}$. Since the ε_j's are independent and centered, we have

$$\mathbb{E}\,\eta=\mathbb{E}\left(\sum_{j=1}^{n}\varepsilon_j x_j\right)^2=\sum_{j=1}^{n}x_j^2+\mathbb{E}\sum_{1\le j\ne k\le n}\varepsilon_j\varepsilon_k x_j x_k=1$$

Now fix $k\ge 2$ and note that

$$\eta^k=\left(\sum_{i=1}^{n}\varepsilon_i x_i\right)^{2k}=\sum_{p_1+\ldots+p_n=2k}\binom{2k}{p_1,\dots,p_n}\varepsilon_1^{p_1}\dots\varepsilon_n^{p_n}\,x_1^{p_1}\dots x_n^{p_n},$$

where the multinomial coefficients are defined by

$$\binom{2k}{p_1, \dots, p_n} = \frac{(2k)!}{p_1! \dots p_n!}.$$

Now since the ε_i's are independent, centered and $\varepsilon_i^p = \varepsilon_i$ if p is odd and $\varepsilon_i^p = 1$ if p is even then

$$\mathbb{E}\, \eta^k = \sum_{k_1+\ldots+k_n=k} \binom{2k}{2k_1, \dots, 2k_n} x_1^{2k_1} \dots x_n^{2k_n}, \tag{9}$$

where we only kept the even powers compared to the first equality. Using that $(p/e)^p \le p! \le p^p$ for any integer p, we have

$$\binom{2k}{2k_1, \dots, 2k_n} \le e^{4k} \binom{k}{k_1, \dots, k_n}^2 \le e^{4k} k! \binom{k}{k_1, \dots, k_n}.$$

Plugging this in (9), we deduce that

$$\mathbb{E}\, \eta^k \le e^{4k} k! \sum_{k_1+\ldots+k_n=k} \binom{k}{k_1, \dots, k_n} (x_1^2)^{k_1} \dots (x_n^2)^{k_n} = e^{4k} k! \left(\sum_{i=1}^n x_i^2 \right)^k = e^{4k} k!,$$

where the last equality follows follows from the fact that $x \in S^{n-1}$. □

We are now ready to state and prove the main result of this subsection.

Theorem 1.11 (Johnson-Lindenstrauss's flattening lemma). *Let $m, n \in \mathbb{N}$ and $\varepsilon \in (0,1)$. There exists $N = N(m, \varepsilon)$ satisfying $N \le 1 + \lfloor e^9 \varepsilon^{-2} \ln(2m^2) \rfloor$ such that the following holds. There exists a linear map A from $\mathbb{R}^n$ into $\mathbb{R}^N$ such that for any set T of m points in $\mathbb{R}^n$ and any $x, y \in T$, we have*

$$(1-\varepsilon)\|x-y\|_2 \le \|Ax - Ay\|_2 \le (1+\varepsilon)\|x-y\|_2.$$

Proof. Let $N \in \mathbb{N}$ to be specified later. Let B be the $N \times n$ random matrix whose entries are independent Rademacher random variables. Let $u \in S^{n-1}$, we write

$$\|Bu\|_2^2 = \sum_{i=1}^N \langle \mathrm{row}_i(B), u \rangle^2,$$

where $\mathrm{row}_i(B)$ denotes the i^{th}-row of B. For any $i \le N$, denote $\eta_i = \langle \mathrm{row}_i(B), u\rangle^2$, $\xi_i = \eta_i - 1$ and note that the ξ_i's are independent and $|\xi_i| \le \max(1, \eta_i)$ for any $i \le N$. It follows from Lemma 1.10 that for any $i \le N$

$$\mathbb{E}\, \eta_i = 1 \quad \text{ and } \quad \mathbb{E}\, \eta_i^k \le e^{4k} k! \quad \text{ for any } k \ge 2.$$

Therefore, the random variables $(\xi_i)_{i \le N}$ are centered and satisfy

$$\mathbb{E}\, |\xi_i|^k \le 1 + \mathbb{E}\, \eta_i^k \le 1 + e^{4k} k!$$

Thus, denoting $\sigma_i^2 := e^9$ and $b := e^4$, a short calculation shows that for any $i \le N$

$$\mathbb{E}\,\xi_i^2 \le \sigma_i^2 \quad \text{and} \quad \mathbb{E}\,|\xi_i|^k \le \frac{1}{2}k!\sigma_i^2 b^{k-2} \quad \text{for any } k \ge 3.$$

Therefore, applying Theorem 1.5 with the random variables ξ_i's and noting that $\sum_{i\le N}\xi_i = \|Bu\|_2^2 - N$, we deduce that for any $u \in S^{n-1}$ and any $t > 0$

$$\mathbb{P}\Big\{\big|\|Bu\|_2^2 - N\big| \ge t\Big\} \le 2\exp\left(-\frac{t^2}{2(Ne^9 + te^4)}\right).$$

Let $x, y \in T$ and $\varepsilon \in (0,1)$ and apply the above with $u = (x-y)/\|x-y\|_2 \in S^{n-1}$ and $t = 2\varepsilon N$ to get

$$\mathbb{P}\Big\{\Big|\|Bx - By\|_2^2 - N\|x-y\|_2^2\Big| \ge 2\varepsilon N\|x-y\|_2^2\Big\} \le 2\exp\left(-\frac{\varepsilon^2 N}{e^9}\right),$$

where we used that $2\varepsilon e^4 \le e^9$. For any $x, y \in T$, if we denote $\mathcal{E}_{x,y}$ the event

$$\mathcal{E}_{x,y} = \left\{(1-\varepsilon)\|x-y\|_2 \le \frac{1}{\sqrt{N}}\|Bx - By\|_2 \le (1+\varepsilon)\|x-y\|_2\right\}$$

then the above implies that $\mathbb{P}\{\mathcal{E}_{x,y}^c\} \le 2\exp(-\varepsilon^2 N e^{-9})$. Therefore

$$\mathbb{P}\Big\{\bigcap_{(x,y)\in T\times T}\mathcal{E}_{x,y}\Big\} = 1 - \mathbb{P}\Big\{\bigcup_{(x,y)\in T\times T}\mathcal{E}_{x,y}^c\Big\} \ge 1 - 2|T|^2\exp\left(-\varepsilon^2 N e^{-9}\right).$$

The above probability is positive whenever $N > e^9\varepsilon^{-2}\ln(2m^2)$ in which case there exists $\omega \in \bigcap_{(x,y)\in T\times T}\mathcal{E}_{x,y}$ and a corresponding linear map B_ω for which the conclusion of the Theorem holds. $\square$

Remark 1.12. Rademacher matrices are a special case of subgaussian random matrices (matrices whose entries are iid subgaussian); these matrices are also good embeddings for the Johnson-Lindenstrauss lemma. To avoid defining subgaussian random variables and their properties (since this is not the subject of this course), we took a different route and used Bernstein's inequality eventhought it is not the ideal way to attack the above problem in this case. Our goal was to give another application of Bernstein's inequality.

2 Regular graphs

2.1 Introduction and definitions

A graph G is a pair of sets (V, E), where V denotes the set of vertices and E the set of edges, formed by pairs of vertices. In these notes, we will only consider *simple* graphs i.e. graphs in which there is at most one edge connecting two vertices (the general case is refered to as a *multi-graph*). Moreover, we will only consider undirected graphs i.e. the set of edges

E is a set of unordered couples. Loops are not allowed meaning that we cannot have an edge connecting a vertex to itself. Therefore, in what follows, **all graphs are supposed to be undirected, simple and without loops**. We will always label vertices from 1 to n and take $V = \{1, \ldots, n\}$. Therefore, two vertices i and j are connected by an edge if $(i, j) \in E$.

Given $S \subset [n]$, we define its (edge) *boundary* ∂S as the set of edges connecting S to $[n] \setminus S$ i.e.

$$\partial S = \{(i, j) \in E : i \in S \text{ and } j \notin S\}.$$

The *edge expanding constant*, or *edge isoperimetric constant* of G, is defined by

$$h(G) = \inf \left\{ \frac{|\partial S|}{|S|} : S \subset [n], |S| \leq n/2 \right\}.$$

The *edge isoperimetric constant* measures the "quality" of G as a network: if $h(G)$ is large, then many edges connect any subset of vertices to its complementary meaning that the information transmitted from each vertex propagates well through the network.

To illustrate this, let us first consider the complete graph G_1 on n vertices i.e. the graph where each vertex is connected to all others. Then for any $S \subset [n]$, we have $|\partial S| = |S|(n - |S|)$ which implies that $h(G_1) = n - \lfloor n/2 \rfloor \sim n/2$. Now consider the cycle G_2 on n vertices i.e. the graph where each vertex i is connected to the $(i+1)^{th}$ vertex for any $1 \leq i \leq n-1$ and the n^{th} vertex connected to the first one. Then taking $S = [n/2]$, we have $|\partial S| = 2$ which implies that $h(G_2) \leq \frac{2}{\lfloor n/2 \rfloor} \sim \frac{4}{n}$. Therefore, we see that $h(\cdot)$ captures the connectivity of the graph. This leads us to the definition of *expanders*, a family of graphs where an information propagates quickly through the graph.

Definition 2.1. *A family of graphs $G_n = ([n], E_n)$ is a family of expanders if there exists $\delta > 0$ such that $h(G) \geq \delta$ for every $n \in \mathbb{N}$.*

Following this definition, we see that a sequence of complete graphs is a family of expanders while a sequence of cycles is not. In the construction of a "good" network, one should try to achieve a good connectivity while keeping the number of edges as low as possible. To this aim, we will unify the number of edges at any particular vertex and define the following class of graphs.

Definition 2.2. *Given $d, n \in \mathbb{N}$ with $d \leq n$. A graph $G = ([n], E)$ is d-regular if for every vertex $i \in [n]$, the set $\{j \in [n] : (i, j) \in E\}$ is of cardinality d.*

Informally speaking, an undirected graph is k regular if each vertex has exactly k neighbors (counting possibly the vertex itself in case of a loop). With this definition, the complete graph on n vertices is $(n-1)$-regular and the cycle is 2-regular.

To every graph G on n vertices, we can naturally associate an $n \times n$ matrix A, called *Adjacency matrix* of G, defined by

$$a_{ij} = 1 \quad \text{if } (i, j) \in E \qquad \text{and} \qquad 0 \text{ otherwise.}$$

Note that A completely determines the graph G. Moreover, since the graph G is undirected and without loops, then A is symmetric with zero diagonal. For example, the adjacency matrix of the complete graph is the $n \times n$ matrix whose all entries are equal to one except those on the diagonal which are equal to zero. Note that G is d regular if its adjacency matrix is *d-double stochastic*.

The spectrum of a graph G is the spectrum of its adjacency matrix. We will list these eigenvalues in the non-increasing order

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n.$$

We will see how the expansion properties of the graph are related to some of its spectral properties. Let us first start with a basic property of the spectrum of a d-regular graph.

Proposition 2.3. *Let G be a d-regular graph on n vertices. Then*

$$\lambda_1 = d \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -d.$$

Proof. Let A be the adjacency matrix of G. First we show that $|\lambda| \leq d$ for any eigenvalue λ. Let x be an eigenvector associated with the eigenvalue λ i.e. $Ax = \lambda x$. Let $i_0 \in [n]$ be such that $|x_{i_0}| = \|x\|_\infty$. Then, we can write

$$|\lambda x_{i_0}| = \big|\sum_{j=1}^{n} A_{i_0 j} x_j\big| \leq \max_{1\leq j\leq n} |x_j| \, \big|\sum_{j=1}^{n} A_{i_0 j}\big| = d\,|x_{i_0}|,$$

which implies that $|\lambda| \leq d$.

Now note that $A\mathbf{1} = d\mathbf{1}$ where $\mathbf{1}$ is the n-dimensional vector with all coordinates equal to 1. Therefore, $\mathbf{1}$ is an eigenvector of A associated with the eigenvalue d. This, together with the above, implies that $\lambda_1 = d$. □

2.2 Edge expansion and the spectral gap

The isoperimetric problem started with *Elissar* (also known as *Dido*), the Phoenician queen of Carthage. When she arrived in 814BC on the coast of Tunisia, she asked for a piece of land. Her request was satisfied provided that the land could be encompassed by an ox-hide. She sliced the hide into very thin strips, tied them together, and was able to enclose a sizable area which became the city of Carthage.

The problem was to find, among all figures with the same perimeter, the one which maximizes the area. More generally, it is about establishing inequalities between the area and the perimeter, or the volume and the surface. In the graph setting, we consider a subset of vertices. One can see the number of vertices a set S contains as its volume. Then, one could define the boundary of S as being the vertices which are connected to its complement. Instead, we will look at the "edge boundary" which is the set of edges connecting S to its complement. Then finding a relation between the volume of S and its boundary reduces to estimating the edge expansion constant we introduced above. In this subsection, we will give upper and lower bounds on $h(G)$ in terms of the spectrum of G. These inequalities are sometimes referred to as isoperimetric inequalities.

Lemma 2.4. *Let $G = ([n], E)$ be a d-regular graph on n vertices. Then for any $S \subset [n]$ with $|S| \leq n/2$, we have*

$$\frac{|\partial S|}{|S|} \geq (d - \lambda_2)(1 - \frac{|S|}{n}).$$

Proof. Let $S \subset [n]$ be such that $|S| \leq n/2$. Denote $\mathbf{1}_S$ the n-dimensional vectors whose coordinates indexed by S are equal to 1 and the remaining are zero. With this notation, we have $\mathbf{1}_{S^c} = \mathbf{1} - \mathbf{1}_S$, where $\mathbf{1}$ is the n-dimensional vector with all coordinates equal to one. Now notice that

$$|\partial S| = |\mathbf{E}_G(S, S^c)| = \sum_{i \in S, j \in S^c} a_{ij},$$

where $(a_{ij})_{i,j \leq n}$ denote the entries of the adjacency matrix A of G. Therefore, we can write

$$|\partial S| = \langle A\mathbf{1}_{S^c}, \mathbf{1}_S \rangle = \langle A\mathbf{1}, \mathbf{1}_S \rangle - \langle A\mathbf{1}_S, \mathbf{1}_S \rangle = d|S| - \langle A\mathbf{1}_S, \mathbf{1}_S \rangle, \tag{10}$$

where in the last equality we used that $\mathbf{1}$ is an eigenvector associated with the eigenvalue d and that $\langle \mathbf{1}, \mathbf{1}_S \rangle = |S|$. It follows from the spectral theorem (and that $\lambda_1 = d$ established in Proposition 2.3) that $A = \frac{d}{n}\,\mathbf{1}\mathbf{1}^t + \lambda_2\, v_2 v_2^t + \ldots + \lambda_n\, v_n v_n^t$, where $(v_i)_{2 \leq i \leq n}$ denote the normalized eigenvectors associated with the eigenvalues $(\lambda_i)_{2 \leq i \leq n}$. Plugging this back in (10), we get

$$|\partial S| = d|S| - \frac{d}{n}|S|^2 - \sum_{i=2}^{n} \lambda_i \langle \mathbf{1}_S, v_i \rangle^2 \geq d|S| - \frac{d}{n}|S|^2 - \lambda_2 \sum_{i=2}^{n} \langle \mathbf{1}_S, v_i \rangle^2.$$

Since $\frac{1}{\sqrt{n}}\mathbf{1}, v_2, \ldots, v_n$ form an orthonormal basis of $\mathbb{R}^n$, then

$$|S| = \|\mathbf{1}_S\|_2^2 = \frac{1}{n}\langle \mathbf{1}, \mathbf{1}_S \rangle^2 + \sum_{i=2}^{n} \langle \mathbf{1}_S, v_i \rangle^2 = \frac{|S|^2}{n} + \sum_{i=2}^{n} \langle \mathbf{1}_S, v_i \rangle^2.$$

Therefore, we deduce that

$$|\partial S| \geq d|S| - \frac{d}{n}|S|^2 - \lambda_2 |S| + \frac{\lambda_2}{n}|S|^2.$$

This implies that

$$\frac{|\partial S|}{|S|} \geq (d - \lambda_2) - \frac{d - \lambda_2}{n}|S|,$$

and finishes the proof. □

The next theorem shows the relation between the expansion properties of a d-regular graph (captured by $h(\cdot)$) and its spectrum. It shows that whenever the gap between the largest eigenvalue (which is equal to d) and the second largest eigenvalue is big, then the edge expansion constant is big which means the graph tends to be a good expander. Inversely, whenever this gap gets smaller, the edge expansion constant decreases. For instance, the adjacency matrix of the complete graph has one eigenvalue equal to $n - 1$ and all others are equal to -1. This means that the gap between the two largest eigenvalues is big, and we already saw that its edge expansion constant is large as well.

Theorem 2.5 (Alon-Milman)**.** *Let $G = ([n], E)$ be a d-regular graph on n vertices. Then*

$$\frac{d-\lambda_2}{2} \leq h(G) \leq \sqrt{2d(d-\lambda_2)}.$$

Proof. The first inequality follows easily from Lemma 2.4. To prove the second inequality, we will show the existence of a set S with $|S| \leq n/2$ such that

$$\frac{|\partial S|}{|S|} \leq \sqrt{2d(d-\lambda_2)}. \tag{11}$$

Let v_2 be the eigenvector associated with the eigenvalue λ_2. Let u be the n-dimensional vector whose positive coordinates coincide with those of v_2 and the remaining are zero. Without loss of generality, we may assume that $\|u\|_\infty = 1$ and $|\operatorname{supp} u| \leq n/2$ (otherwise we renormalize v_2 for the first assumption, and replace v_2 by $-v_2$ for the second assumption).

Let γ be a random variable uniformly distributed on $[0,1]$. We consider the random set

$$S := \{i \in \operatorname{supp} u : \ u_i^2 \geq \gamma\}.$$

We will show that there exists a realization of S satisfying (11). First note that $|S| \leq n/2$ by construction.

$$|\partial S| = \sum_{(i,j)\in S\times S^c} a_{ij} = \sum_{i,j\in \operatorname{supp} u} a_{ij}\chi_{\{i\in S, j\notin S\}},$$

where χ denotes the indicator function. Therefore,

$$\mathbb{E}|\partial S| = \sum_{i,j\in \operatorname{supp} u} a_{ij}\,\mathbb{P}\{i\in S \text{ and } j\notin S\} = \sum_{\substack{i,j\in \operatorname{supp} u\\ u_j\leq u_i}} a_{ij}\,\mathbb{P}\{u_j^2 \leq \gamma \leq u_i^2\}.$$

Now, using that $\mathbb{P}\{u_j^2 \leq \gamma \leq u_i^2\} = u_i^2 - u_j^2$, we get

$$\mathbb{E}|\partial S| = \sum_{\substack{i,j\in \operatorname{supp} u\\ u_j\leq u_i}} a_{ij}\,(u_i^2 - u_j^2) = \frac{1}{2}\sum_{\substack{i,j\in \operatorname{supp} u\\ u_j\leq u_i}} a_{ij}\,(u_i^2 - u_j^2) + \frac{1}{2}\sum_{\substack{i,j\in \operatorname{supp} u\\ u_i\leq u_j}} a_{ji}\,(u_j^2 - u_i^2).$$

Since A is symmetric, we deduce from the above that

$$\mathbb{E}|\partial S| = \frac{1}{2}\sum_{\substack{i,j\in \operatorname{supp} u\\ u_j\leq u_i}} a_{ij}\,|u_i^2 - u_j^2| + \frac{1}{2}\sum_{\substack{i,j\in \operatorname{supp} u\\ u_i\leq u_j}} a_{ij}\,|u_j^2 - u_i^2| = \frac{1}{2}\sum_{i,j\in \operatorname{supp} u} a_{ij}\,|u_i^2 - u_j^2|.$$

Writing that $|u_i^2 - u_j^2| = |u_i - u_j|\,|u_i + u_j|$ then using Cauchy-Schwarz inequality and that $a_{ij}^2 = a_{ij}$, we can write

$$\mathbb{E}|\partial S| \leq \frac{1}{2}\left(\sum_{i,j\in \operatorname{supp} u} a_{ij}\,(u_i - u_j)^2\right)^{\frac{1}{2}}\left(\sum_{i,j\in \operatorname{supp} u} a_{ij}\,(u_i + u_j)^2\right)^{\frac{1}{2}}. \tag{12}$$

Using that A is d-double stochastic, we have

$$\sum_{i,j\in \operatorname{supp} u} a_{ij}\, u_i^2 \le d\|u\|_2^2. \tag{13}$$

Moreover,

$$\langle Au, u\rangle = \langle A(u-v_2), u\rangle + \langle Av_2, u\rangle = \langle A(u-v_2), u\rangle + \lambda_2\langle v_2, u\rangle = \langle A(u-v_2), u\rangle + \lambda_2\|u\|_2^2.$$

Since $\langle A(u-v_2), u\rangle \ge 0$ (which follows from the fact that all coordinates of $u-v_2$ and u are non negative), the above implies that

$$\sum_{i,j\in \operatorname{supp} u} a_{ij} u_i u_j = \langle Au, u\rangle \ge \lambda_2\|u\|_2^2. \tag{14}$$

Putting together (12), (13), (14) and using that $\langle Au, u\rangle \le d\|u\|_2^2$, we deduce that

$$\mathbb{E}|\partial S| \le \sqrt{2d(d-\lambda_2)}\, \|u\|_2^2.$$

Now notice that

$$\mathbb{E}\,|S| = \sum_{i\in \operatorname{supp} u} \mathbb{P}\{i\in S\} = \sum_{i\in \operatorname{supp} u} \mathbb{P}\{\gamma \le u_i^2\} = \sum_{i\in \operatorname{supp} u} u_i^2 = \|u\|_2^2.$$

Therefore, we have

$$\mathbb{E}|\partial S| \le \sqrt{2d(d-\lambda_2)}\, \mathbb{E}\,|S|,$$

which implies the existence of a set S of size at most $n/2$ satisfying (11). □

It follows from the previous statement that the graph is connected if and only if $\lambda_2 < d$. Indeed, G is disconnected if and only if there exists $S\subset [n]$ such that $\mathbf{E}_G(S,S^c)=\emptyset$ which means that $h(G)=0$ and thus equivalent to $\lambda_2 = d$ by Theorem 2.5. The previous theorem suggests that if one want a graph to have good expansion, it needs to have a large spectral gap. The next theorem shows how big can a spectral gap be. Let us define

$$\lambda := \lambda(G) = \max(|\lambda_2|, |\lambda_n|).$$

Theorem 2.6 (Alon-Boppana)**.** *There exists a positive constant C such that the following holds. For any d-regular graph on n vertices, we have*

$$\lambda \ge 2\sqrt{d-1}\Big(1-\frac{C\ln^2 d}{\ln^2 n}\Big).$$

We will not give a proof of the above theorem here (see [9]). Instead, we state and prove a weaker version.

Proposition 2.7. *For any d-regular graph on n vertices, we have*

$$\lambda \ge \sqrt{d\Big(1-\frac{d-1}{n-1}\Big)}.$$

Proof. Using that A is d-double stochastic, it is easy to see that

$$\mathrm{Tr}(A^2) = \sum_{1\le i,j\le n} a_{ij}^2 = \sum_{1\le i,j\le n} a_{ij} = nd.$$

On the other hand, since $(\lambda_i^2)_{i\le n}$ are the eigenvalues of A^2, we can write

$$\mathrm{Tr}(A^2) = \sum_{i=1}^{n} \lambda_i^2 = d^2 + \sum_{i=2}^{n} \lambda_i^2 \le d^2 + (n-1)\lambda^2.$$

Therefore, we deduce that

$$\lambda^2 \ge \frac{d(n-d)}{n-1},$$

and finish the proof. □

We saw in Theorem 2.5 that a large spectral gap implies good expansion properties for the corrsponding graph. On the other hand, Theorem 2.6 indicates how big can the spectral gap be, meaning how good of an expander a graph can be. This motivates the following definition.

Definition 2.8 (Ramanujan graphs)**.** *A d-regular graph on n vertices is called Ramanujan, if*

$$\lambda \le 2\sqrt{d-1}.$$

A family of Ramanujan graphs is of course a family of expanders. By the remark above, these are the best expanders. An important line of research is to construct such families. However, only recently the existence of Ramanujan graphs of all degrees was established [8]. This will not be the subject of this course, instead we will investigate random graphs.

2.3 Random regular graphs

As we mentioned above, constructing Ramanujan graphs is not an easy task. We turn our attention to random graphs and see if they are suitable candidates for being good expanders. There are several ways to introduce randomness in a graph. The model which will be considered in this course is the Erdös-Renyi graph $\mathcal{G}(n,p)$ where we decide to put an edge between two different vertices independently with the same probability p. We will write $p = d/n$ for some $d \le n$. The corresponding adjacency matrix to $\mathcal{G}(n,p)$ is an $n\times n$ symmetric matrix with zero diagonal whose entries above the diagonal are independent Bernoulli variables with parameter d/n.

We saw in Proposition 1.6 that such a symmetric Bernoulli matrix is almost d-double stochastic with high probability, provided that $d \gtrsim \ln n$. Therefore, whenever $d \gtrsim \ln n$, the Erdös-Renyi graph is almost d-regular with high probability. We may therefore investigate if it is almost Ramanujan by studying its spectral gap i.e. the gap between the largest and the second largest (in absolute value) eigenvalue. This will be the target of this subsection, but first let us note that when $d \lesssim \ln n$ the graph cannot be a good expander as it contains isolated vertices i.e. vertices with no edges on them.

Proposition 2.9. *Let $\varepsilon \in (0,1)$ and suppose that $d \lesssim \ln(\varepsilon n)$. Then*

$$\mathbb{P}\Big\{\exists \text{ an isolated vertex in } \mathcal{G}(n,d/n)\Big\} \geq 1-\varepsilon.$$

Proof. For any $i \leq n$, define $\mathcal{E}_i = \{\text{vertex } i \text{ is isolated}\}$. Clearly, we have

$$\mathbb{P}\{\mathcal{E}_i\} = \Big(1-\frac{d}{n}\Big)^{n-1}.$$

Moreover, for any $i,j \in [n]$, it is not difficult to see that

$$\mathbb{P}\Big\{\mathcal{E}_i \cap \mathcal{E}_j\Big\} = \Big(1-\frac{d}{n}\Big)^{2n-3}.$$

Let η denote the number of isolated vertices. We have

$$\eta = \sum_{i\leq n} \chi_i,$$

where χ_i denotes the indicator function of the event $\mathcal{E}_i$. From the above, we have

$$\mathbb{E}\,\eta = n\Big(1-\frac{d}{n}\Big)^{n-1} \quad \text{and} \quad \mathbb{E}\,\eta^2 = n\Big(1-\frac{d}{n}\Big)^{n-1} + n(n-1)\Big(1-\frac{d}{n}\Big)^{2n-3}. \tag{15}$$

Now the Paley-Zygmund inequality implies that for any $\theta \in (0,1)$, we have

$$\mathbb{P}\Big\{\eta \geq \theta \mathbb{E}\,\eta\Big\} \geq (1-\theta)^2 \frac{(\mathbb{E}\,\eta)^2}{\mathbb{E}\,\eta^2}.$$

This, together with (15) implies

$$\mathbb{P}\Big\{\eta \geq \theta n\Big(1-\frac{d}{n}\Big)^{n-1}\Big\} \geq (1-\theta)^2 \frac{n\Big(1-\frac{d}{n}\Big)^{n-1}}{1+(n-1)\Big(1-\frac{d}{n}\Big)^{n-2}}.$$

Now if $\varepsilon \in (0,1)$ and $d \lesssim \ln(\varepsilon n)$, one can check that

$$\mathbb{P}\Big\{\eta \gtrsim \theta/\varepsilon\Big\} \geq (1-\theta)^2(1-\varepsilon).$$

Choosing θ of the correct order finishes the proof. □

In the remaining, we will consider the case where $d \gtrsim \ln n$ and show that in this setting $\mathcal{G}(n,d/n)$ has a large spectral gap. We denote by B its adjacency matrix which is a symmetric $n \times n$ matrix having independent Bernoulli entries with parameter d/n on and above the diagonal. The main theorem in this subsection is the following.

Theorem 2.10. *For any $K > 0$, there exists $C = C(K)$ such that the following holds. Let $d,n \in \mathbb{N}$. Denote λ the second largest (in absolute value) eigenvalue of $\mathcal{G}(n,d/n)$. Then*

$$\mathbb{P}\big\{\lambda \geq C\sqrt{d}\big\} \leq \frac{1}{n^K}.$$

Let us start first by characterizing the second largest (in absolute value) eigenvalue. The following is a special case of the courant-Fisher formula.

Proposition 2.11. *Let A be an $n \times n$ symmetric matrix and let λ denotes its second largest (in absolute value) eigenvalue. Then*

$$\lambda = \min_{z \in S^{n-1}} \max_{x \in S^{n-1} \cap z^\perp} \|Ax\|_2 = \min_{z \in S^{n-1}} \max_{\substack{x,y \in S^{n-1} \\ x \perp z}} \langle Ax, y \rangle.$$

Proof. Let $(\lambda_i)_{i \leq n}$ be the eigenvalues of A and suppose that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. From the spectral theorem, we can write $A = \sum_{i=1}^n \lambda_i v_i v_i^t$ where v_i are the normalized eigenvectors corresponding to the eigenvalues $(\lambda_i)_{i \leq n}$. Let $z \in S^{n-1}$ and $x \in z^\perp \cap \operatorname{span}\{v_1, v_2\} \cap S^{n-1}$ (the latter is non-empty since $z^\perp$ is of dimension $n-1$ and $\operatorname{span}\{v_1, v_2\}$ is of dimension 2). Then using that x is orthogonal to all but v_1 and v_2, we get

$$\|Ax\|_2^2 = \|\sum_{i \leq n} \lambda_i \langle x, v_i \rangle v_i\|_2^2 = \lambda_1^2 \langle x, v_1 \rangle^2 + \lambda_2^2 \langle x, v_2 \rangle^2 \geq \lambda_2^2.$$

Since shows that

$$|\lambda_2| \leq \min_{z \in S^{n-1}} \max_{x \in S^{n-1} \cap z^\perp} \|Ax\|_2.$$

On the other hand, taking $z = v_1$ one gets the reverse inequality. □

Having in mind that Ramanujan graphs have their second largest eigenvalue of order $\sqrt{d}$, we will try to show that with high probability the same holds for $\mathcal{G}(n, d/n)$. Using Proposition 2.11, and choosing $z = \mathbf{1}$ the n-dimensional vector whose all coordinates are equal to one, it is sufficient to show that

$$\mathbb{P}\Big\{ \max_{\substack{x,y \in S^{n-1} \\ x \perp \mathbf{1}}} \langle Bx, y \rangle \gtrsim \sqrt{d} \Big\} \underset{n \to \infty}{\longrightarrow} 0.$$

Let us denote

$$S_0^{n-1} = \mathbf{1}^\perp \cap S^{n-1} = \{x \in S^{n-1} : \sum_{i=1}^n x_i = 0\}$$

and for any $(x, y) \in S_0^{n-1} \times S^{n-1}$ the event

$$\mathcal{E}_{x,y} = \Big\{ \langle Bx, y \rangle \gtrsim \sqrt{d} \Big\}.$$

Then our task is to show that

$$\mathbb{P}\Big\{ \bigcup_{(x,y) \in S_0^{n-1} \times S^{n-1}} \mathcal{E}_{x,y} \Big\} \underset{n \to \infty}{\longrightarrow} 0.$$

Similarly to previous arguments in these notes, we would like to estimate $\mathbb{P}\{\mathcal{E}_{x,y}\}$ for a fixed (x, y) then use a union bound argument to bound the above quantity. This motivates us to try to discretize the sphere and show that controlling $\langle Bx, y \rangle$, for x, y in some discrete set, is sufficient in order to bound it for all other vectors of the sphere.

Definition 2.12 (ε-Net)**.** *Given $\varepsilon > 0$ and $S \subset S^{n-1}$, we say that $\mathcal{N} \subset S$ is an ε-net of S if for any $x \in S$, there exists $y \in \mathcal{N}$ such that $\|x - y\|_2 \le \varepsilon$.*

In the next lemma, we will show that there always exists an ε-net of suitable cardinality.

Lemma 2.13 (Net construction)**.** *Given $\varepsilon > 0$ and $S \subset S^{n-1}$, there exists an ε-net $\mathcal{N}$ of S satisfying*

$$|\mathcal{N}| \le (1 + 2/\varepsilon)^n.$$

Proof. Let $\mathcal{N} = \{y_1, \dots, y_s\} \subset S$ be an ε-separated set of maximal size i.e.

$$\|y_i - y_j\|_2 > \varepsilon \qquad \forall i \neq j,$$

and the above property is violated if one adds an element to $\mathcal{N}$. Therefore, for any $x \in S$, the set $S \cup \{x\}$ is not ε-separated. Thus, there exists $i \le s$ such that $\|x - y_i\|_2 \le \varepsilon$. This implies that $\mathcal{N}$ is an ε-net of S.

Let us now estimate its cardinality. Note that since the y_i's are ε-separated, then the Euclidean balls $\mathcal{B}(y_i, \varepsilon/2)$ of center y_i and radius $\varepsilon/2$ are disjoint. Moreover, since the y_i's belong to the sphere then

$$\bigcup_{i=1}^{s} \mathcal{B}(y_i, \varepsilon/2) \subseteq \mathcal{B}(0, 1 + \varepsilon/2).$$

Therefore, using that $\mathcal{B}(y_i, \varepsilon/2)$ $(i = 1, \dots, s)$ are disjoint, we get

$$\mathrm{Vol}\Big(\bigcup_{i=1}^{s} \mathcal{B}(y_i, \varepsilon/2)\Big) = \sum_{i=1}^{s} \mathrm{Vol}\Big(\mathcal{B}(y_i, \varepsilon/2)\Big) \le \mathrm{Vol}\Big(\mathcal{B}(0, 1 + \varepsilon/2)\Big).$$

This implies that

$$s\Big(\frac{\varepsilon}{2}\Big)^n \le \Big(1 + \frac{\varepsilon}{2}\Big)^n,$$

which after rearrangement finishes the proof. □

Now that we saw how to construct a net, we need to check that controlling the quantity $\langle Bx, y\rangle$ for vectors from the net implies a control for all other vectors. This is done in the next lemma.

Lemma 2.14. *Let $\varepsilon \in (0, 1/2)$, $\mathcal{N}_\varepsilon$ an ε-net of S^{n-1}, $\mathcal{N}_\varepsilon^0$ an ε-net of S_0^{n-1}, and A be an $n \times n$ matrix. If $|\langle Ax, y\rangle| \le \beta$ for all $(x, y) \in \mathcal{N}_\varepsilon \times \mathcal{N}_\varepsilon^0$, then $|\langle Ax, y\rangle| \le \beta/(1 - 2\varepsilon)$ for all $(x, y) \in S^{n-1} \times S_0^{n-1}$.*

Proof. Let $(x_0, y_0) \in S^{n-1} \times S_0^{n-1}$ be such that $a := \sup_{(x,y)\in S^{n-1}\times S_0^{n-1}} \langle Ax, y\rangle = \langle Ax_0, y_0\rangle$. By the definition of $\mathcal{N}_\varepsilon$ and $\mathcal{N}_\varepsilon^0$, there exist $(x_0', y_0') \in \mathcal{N}_\varepsilon \times \mathcal{N}_\varepsilon^0$ such that $\|x_0 - x_0'\|_2 \le \varepsilon$ and $\|y_0 - y_0'\| \le \varepsilon$. Using this, together with the fact that the normalized difference of two elements in S_0^{n-1} remains in S_0^{n-1}, we get

$$\begin{aligned}\langle Ax_0, y_0\rangle &= \langle A(x_0 - x_0'), y_0\rangle + \langle Ax_0', y_0 - y_0'\rangle + \langle Ax_0', y_0'\rangle \\ &\le a\|x_0 - x_0'\|_2 + a\|y_0 - y_0'\|_2 + \sup_{(x,y)\in\mathcal{N}_\varepsilon\times\mathcal{N}_\varepsilon^0} |\langle Ax, y\rangle|.\end{aligned}$$

This, together with the hypothesis of the lemma, implies that

$$a \le 2\varepsilon\, a + \beta,$$

which gives that $a \le \beta/(1-2\varepsilon)$.

Now let $(x,y) \in S^{n-1} \times S_0^{n-1}$. By the definition of $\mathcal{N}_\varepsilon$ and $\mathcal{N}_\varepsilon^0$, there exist $(x',y') \in \mathcal{N}_\varepsilon \times \mathcal{N}_\varepsilon^0$ such that $\|x-x'\|_2 \le \varepsilon$ and $\|y-y'\| \le \varepsilon$. One can easily check that

$$\langle Ax, y\rangle = \langle A(x-x'), y\rangle + \langle Ax', y-y'\rangle + \langle Ax', y'\rangle \le 2\varepsilon a + \beta \le \frac{\beta}{1-2\varepsilon},$$

and finish the proof. □

Once the net is constructed, we can focus on estimating the probability of $\mathcal{E}_{x,y}$ for fixed $(x,y) \in S_0^{n-1} \times S^{n-1}$. Let us try describing the strategy of the proof. First note that

$$\langle Bx, y\rangle = \sum_{1\le i<j\le n} \varepsilon_{ij} x_j y_i + \sum_{1\le j<i\le n} \varepsilon_{ij} x_j y_i,$$

which implies that

$$\mathbb{P}\Big\{\langle Bx, y\rangle \ge 2C\sqrt{d}\Big\} \le \mathbb{P}\Big\{\sum_{1\le i<j\le n} \varepsilon_{ij} x_j y_i \ge C\sqrt{d}\Big\} + \mathbb{P}\Big\{\sum_{1\le j<i\le n} \varepsilon_{ij} x_j y_i \ge C\sqrt{d}\Big\},$$

where C is a universal constant. We may then focus on estimating one of the terms above as the second could be treated in a similar manner. The main advantage is that when $i<j$ the random variables ε_{ij} are independent and we may therefore use the concentration inequality established at the beginning of these notes in order to control the large deviation of a sum of independent random variables. Let us denote $\alpha_{ij} = \varepsilon_{ij} x_j y_i$ for any $i,j \in [n]$. Then $|\alpha_{ij}| \le \|x\|_\infty \|y\|_\infty$ and

$$\mathbb{E}\,\alpha_{ij} = \frac{d}{n} x_j y_i \quad \text{ and } \quad \mathbb{E}\,\alpha_{ij}^2 = \frac{d}{n} x_j^2 y_i^2.$$

We can therefore use Theorem 1.3 (with $b = \|x\|_\infty \|y\|_\infty$, $\sigma_{ij}^2 = \frac{d}{n} x_j^2 y_i^2$ and $t \sim \sqrt{d}$) to deduce that

$$\mathbb{P}\Big\{\Big|\sum_{1\le i<j\le n} \big(\varepsilon_{ij} - \frac{d}{n}\big) x_j y_i\Big| \gtrsim \sqrt{d}\Big\} \le 2\exp\left(-\frac{d}{n\|x\|_\infty^2 \|y\|_\infty^2} H\Big(\frac{n}{\sqrt{d}} \|x\|_\infty \|y\|_\infty\Big)\right).$$

It is easy to check that when $\|x\|_\infty \|y\|_\infty \gtrsim \sqrt{d}/n$ then the probability estimate above is always smaller than $\exp(-n)$. Having in mind the union bound argument over the constructed net of size $\sim \exp(n)$, this suggests that the strategy will fail when $\|x\|_\infty \|y\|_\infty \gtrsim \sqrt{d}/n$. This motivates the following splitting of the vectors x and y depending on the order of magnitude of their coordinates. Given $(x,y) \in S_0^{n-1} \times S^{n-1}$, let us define

$$\mathcal{L}(x,y) := \big\{(i,j) \in [n]^2 :\, |x_j y_i| \le \sqrt{d}/n\big\} \quad \text{ and } \quad \mathcal{H}(x,y) := \big\{(i,j) \in [n]^2 :\, |x_j y_i| > \sqrt{d}/n\big\}.$$

The notation $\mathcal{L}(x,y)$ stands for *light couples* while $\mathcal{H}(x,y)$ refers to *heavy couples*. From what is described above, we will be able to deal with the *light* part using Bennett's inequality.

Lemma 2.15. *Let $(x,y) \in S_0^{n-1} \times S^{n-1}$. Then for any $t > 0$ we have*

$$\mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \varepsilon_{ij}\Big| \geq (1+2t)\sqrt{d}\Big\} \leq 4\exp\big(-n\,H(t)\big).$$

Proof. Let $(x,y) \in S_0^{n-1} \times S^{n-1}$. First notice that for any $t > 0$

$$\mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \big(\varepsilon_{ij} - \mathbb{E}\,\varepsilon_{ij}\big)\Big| \geq 2t\sqrt{d}\Big\} \leq \mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y),\, i<j} x_j y_i \big(\varepsilon_{ij} - \mathbb{E}\,\varepsilon_{ij}\big)\Big| \geq t\sqrt{d}\Big\}$$
$$+\, \mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y),\, i>j} x_j y_i \big(\varepsilon_{ij} - \mathbb{E}\,\varepsilon_{ij}\big)\Big| \geq t\sqrt{d}\Big\}.$$

Setting $\alpha_{ij} = x_j y_i\big(\varepsilon_{ij} - \mathbb{E}\,\varepsilon_{ij}\big)$ for any $(i,j) \in \mathcal{L}(x,y)$, we have that $|\alpha_{ij}| \leq |x_j y_i| \leq \sqrt{d}/n$ and $\mathbb{E}\,\alpha_{ij}^2 \leq d/n$. Applying Theorem 1.3 to the independent centered random variables $(\alpha_{ij})_{i<j}$, we get

$$\mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y),\, i<j} x_j y_i \big(\varepsilon_{ij} - \mathbb{E}\,\varepsilon_{ij}\big)\Big| \geq t\sqrt{d}\Big\} \leq 2\exp\left(-nH(t)\right),$$

where we used that $\sum_{i<j} \mathbb{E}\,\alpha_{ij}^2 \leq d/n$ since $x, y \in S^{n-1}$. Obviously, the same inequality holds for $(\alpha_{ij})_{i>j}$. Therefore, we deduce that

$$\mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \big(\varepsilon_{ij} - \mathbb{E}\,\varepsilon_{ij}\big)\Big| \geq 2t\sqrt{d}\Big\} \leq 4\exp\left(-nH(t)\right). \tag{16}$$

Since the coordinates of x sum up to zero, we have for any $i \leq n$:

$$\Big| \sum_{j\in\{k:\,(i,k)\in\mathcal{L}(x,y)\}} x_j y_i\Big| = \Big| \sum_{j\in\{k:\,(i,k)\in\mathcal{H}(x,y)\}} x_j y_i\Big| \leq \sum_{j\in\{k:\,(i,k)\in\mathcal{H}(x,y)\}} \frac{(x_j y_i)^2}{\sqrt{d}/n},$$

where in the last inequality we used that $|x_j y_i| \geq \sqrt{d}/n$ for $(i,j) \in \mathcal{H}(x,y)$. Summing the previous inequality over all rows and using the condition $\|x\| = \|y\| = 1$, we get

$$\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \mathbb{E}\,\varepsilon_{ij}\Big| = \frac{d}{n}\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i\Big| \leq \sqrt{d}$$

This, together with (16), finishes the proof. □

For the *heavy couples*, we will assume the following result of Kahn-Szeméredi [7] who show that for any matrix having the nice structural properties as in Proposition 1.8, its action on the *heavy couples* can be controlled.

Lemma 2.16 (Kahn-Szemerédi). *For any $K > 0$, there exist β depending only on K such that the following holds. If $A \in \mathcal{E}_{1.8}(K)$ then for any $x, y \in S^{n-1}$, we have*

$$\Big| \sum_{(i,j)\in\mathcal{H}(x,y)} x_j y_i A_{i,j}\Big| \leq \beta\sqrt{d}.$$

We have now all ingredients in place to prove the main Theorem.

Proof of Theorem 2.10. Let $K \geq 1$ and let β be the constant coming from Lemma 2.16. Moreover let $r := H^{-1}(1 + \ln 81)$ and denote

$$\mathcal{E} := \{\lambda \geq 2(\beta + r)\sqrt{d}\}.$$

From Proposition 2.11, we have

$$\mathbb{P}\{\mathcal{E} \mid \mathcal{E}_{1.8}\} \leq \mathbb{P}\Big\{ \exists (x,y) \in S_0^{n-1} \times S^{n-1} \text{ such that } |\langle Bx, y\rangle| \geq 2(1+\beta+2r)\sqrt{d} \mid \mathcal{E}_{1.8}\Big\}.$$

Let $\mathcal{N}$ be a 1/4-net of S^{n-1} and $\mathcal{N}_0$ a 1/4-net of S_0^{n-1}. Using Lemma 2.13 we may take $\mathcal{N}$ and $\mathcal{N}_0$ such that $\max(|\mathcal{N}|, |\mathcal{N}_0|) \leq 9^n$. Applying Lemma 2.14, we get

$$\begin{aligned}\mathbb{P}\{\mathcal{E} \mid \mathcal{E}_{1.8}\} &\leq \mathbb{P}\Big\{\exists (x,y) \in \mathcal{N}_0 \times \mathcal{N} \text{ such that } |\langle Bx, y\rangle| \geq (1+\beta+2r)\sqrt{d} \mid \mathcal{E}_{1.8}\Big\} \\ &\leq (81)^n \max_{(x,y)\in S_0^{n-1}\times S^{n-1}} \mathbb{P}\Big\{|\langle Bx, y\rangle| \geq (1+\beta+2r)\sqrt{d} \mid \mathcal{E}_{1.8}\Big\}. \qquad (17)\end{aligned}$$

Now, given $(x,y) \in S_0^{n-1} \times S^{n-1}$, note that

$$|\langle Bx, y\rangle| \leq \Big|\sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \varepsilon_{i,j}\Big| + \Big|\sum_{(i,j)\in\mathcal{H}(x,y)} x_j y_i \varepsilon_{i,j}\Big|.$$

From Lemma 2.16, we have $\Big|\sum_{(i,j)\in\mathcal{H}(x,y)} x_j y_i \varepsilon_{i,j}\Big| \leq \beta\sqrt{d}$ whenever $B \in \mathcal{E}_{1.8}$. Therefore, from the above and (17), we get

$$\begin{aligned}\mathbb{P}\{\mathcal{E} \mid \mathcal{E}_{1.8}\} &\leq (81)^n \max_{(x,y)\in S_0^{n-1}\times S^{n-1}} \mathbb{P}\Big\{\Big|\sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \varepsilon_{i,j}\Big| \geq (1+2r)\sqrt{d} \mid \mathcal{E}_{1.8}\Big\} \\ &\leq \frac{(81)^n}{\mathbb{P}\{\mathcal{E}_{1.8}\}} \max_{(x,y)\in S_0^{n-1}\times S^{n-1}} \mathbb{P}\Big\{\Big|\sum_{(i,j)\in\mathcal{L}(x,y)} x_j y_i \varepsilon_{i,j}\Big| \geq (1+2r)\sqrt{d}\Big\},\end{aligned}$$

which after applying Lemma 2.15 (with $t = r$) implies that

$$\mathbb{P}\{\mathcal{E} \mid \mathcal{E}_{1.8}\} \leq \frac{(81)^n}{\mathbb{P}\{\mathcal{E}_{1.8}\}} 4\exp(-n\,H(r)) \leq \frac{4e^{-n}}{\mathbb{P}\{\mathcal{E}_{1.8}\}},$$

by the choice of r. To finish the proof, note that

$$\mathbb{P}\{\mathcal{E}\} \leq \mathbb{P}\{\mathcal{E} \mid \mathcal{E}_{1.8}\}\,\mathbb{P}\{\mathcal{E}_{1.8}\} + \mathbb{P}\{\mathcal{E}_{1.8}^c\}$$

and use the estimate established above together with the one obtained in Proposition 1.8. □

Remark 2.17. In the above subsection, we investigated the homogeneous Erdös-Renyi graph $\mathcal{G}(n, d/n)$ and established the spectral gap in the dense regime. This model is not very convenient in applications for several reasons. The first one is that people's connections tend to be dependent while in this model we supposed their independence. Another reason is that it is more realist to define an inhomogeneous model where some weight is put on each edge favoring for example connections among communities. One can therefore define the model $\mathcal{G}(n, P)$, where $P = (p_{ij})_{1\leq i,j\leq n}$ is a symmetric double stochastic matrix, where vertices i and j are connected with probability p_{ij}. One can study the expansion properties of such model in terms of the matrix P.

Another model of interest is by dropping the independence assumption. This can be done by considering the following model. Let $\mathcal{G}_{n,d}$ the set of all undirected graphs on n vertices. Then consider G a random graph uniformly distributed on $\mathcal{G}_{n,d}$. The adjacency matrix is a symmetric $n \times n$ random matrix which is d-double stochastic. The main difficulty here is the dependence between the entries of this matrix. It was conjectured by Alon [1] and proved by Friedman [6] (and later by Bordenave [2]) that for d fixed, this model has a largest possible spectral gap with high probability i.e. it is almost Ramanujan with high probability. It was later conjectured by Vu [11] that the same phenomenon holds for dense graphs, more precisely for any $d \leq n/2$. Broder, Frieze, Suen and Upfal [3] verified this for $d \leq \sqrt{n}$ and Cook, Goldstein and Johnson [4] for $d \leq n^{2/3}$. Very recently, with Konstantin Tikhomirov [10], we answered this question by bounding the spectral gap for any $\sqrt{n} \leq d \leq n/2$. The proof follows similar analysis to the proof described in this subsection. The only difference is that Bennett's inequality is not usable because of the lack of independence. The whole difficulty is therefore to establish a similar inequality which deals with dependent random variables, where the dependence is dictated by the uniform model on $\mathcal{G}_{n,d}$.

References

[1] N. Alon, Eigenvalues and expanders, Combinatorica **6** (1986), no. 2, 83–96. MR0875835

[2] Ch. Bordenave, A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts, arXiv:1502.04482.

[3] A. Z. Broder, A. M. Frieze, S. Suen, E. Upfal, Optimal construction of edge-disjoint paths in random graphs, SIAM J. Comput. **28** (1999), no. 2, 541–573 (electronic). MR1634360

[4] N. Cook, L. Goldstein, T. Johnson, Size biased couplings and the spectral gap for random regular graphs, arXiv:1510.06013.

[5] G. Davidoff, P. Sarnak, and A. Valette, Elementary number theory, group theory, and Ramanujan graphs, In: London Mathematical Society Student Texts, vol. 55, Cambridge University Press, Cambridge, 2003, pp. x + 144.

[6] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100 pp. MR2437174

[7] J. Friedman, J. Kahn, E. Szemerédi, On the second eigenvalue of random regular graphs, Proceedings of the twenty-first annual ACM symposium on Theory of computing (1989), 587–598.

[8] A. Marcus, D. A. Spielman, and N. Srivastava, Interlacing families I: Bipartite Ramanujan graphs of all degrees, Ann. of Math. 182 (2015), 307-325.

[9] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their application, Bull Am Math Soc (N.S.) 43(4) (2006), 439-561.

[10] K. Tikhomirov, P. Youssef, The uniform model for d-regular graphs: concentration inequalities for linear forms and the spectral gap, Preprint.

[11] V. Vu, Random discrete matrices, in *Horizons of combinatorics*, 257–280, Bolyai Soc. Math. Stud., 17, Springer, Berlin. MR2432537

Pierre Youssef,
Laboratoire de Probabilités et de Modèles aléatoires, Université Paris Diderot,
E-mail: youssef@math.univ-paris-diderot.fr