Discrete Applied Mathematics 243 (2018) 286–289

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Note

Discrepancy minimizing spectral clustering

Marianna Bolla *, Ahmed Elbanna

Institute of Mathematics, Technical University of Budapest, Hungary

ARTICLE INFO

Article history:
Received 24 August 2017
Received in revised form 4 January 2018
Accepted 19 February 2018
Available online 22 March 2018

Keywords:
Multiway discrepancy
Normalized adjacency matrix
Spectral subspaces
k-means clustering

ABSTRACT

In this short note we strengthen a former result of Bolla (2011), where in a multipartition (clustering) of a graph's vertices we estimated the pairwise discrepancies of the clusters with the normalized adjacency spectra. There we used the definition of Alon et al. (2010) for the volume-regularity of the cluster pairs. Since then, in Bolla (2016) we defined the so-called k-way discrepancy of a k-clustering and estimated the kth largest (in absolute value) normalized adjacency eigenvalue with an increasing function of it. In the present paper, we estimate the new discrepancy measure with this eigenvalue. Putting these together, we are able to establish a relation between the large spectral gap (as for the $(k-1)$th and kth non-trivial normalized adjacency eigenvalues) and the sudden decrease between the $k-1$ and k-way discrepancies. It makes rise to a new paradigm of spectral clustering, which minimizes the multiway discrepancy.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Our purpose is to relate discrepancy and spectra in a multiway classification of a graph's vertices. To do so, we develop an estimate for the multiway discrepancy by spectral tools. To simplify notation, we use unweighted graphs, though the result can as well be extended to edge-weighted ones.

Let $G = (V, \boldsymbol{A})$ be an undirected, unweighted graph on the n-element vertex-set V with the $n \times n$ adjacency matrix $\boldsymbol{A} = (a_{ij})$. We will use the normalized adjacency matrix $\boldsymbol{A_D} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ of G, where $\boldsymbol{D} = \mathrm{diag}(d_1, \dots, d_n)$ is the diagonal *degree-matrix* and d_v is the degree of vertex v. Assume that G is *connected*, i.e., $\boldsymbol{A}$ is *irreducible*. Then the eigenvalues of $\boldsymbol{A_D}$, enumerated in decreasing absolute values are $1 = \mu_0 \geq |\mu_1| \geq |\mu_2| \geq \cdots \geq |\mu_{n-1}|$, and μ_0 is a single eigenvalue. (Note that $\mu_1 = -1$ can be if G is bipartite.)

Let $1 < k < n$ be a fixed integer. We look for the proper k-partition (clustering) $V_1, \dots, V_k$ of the vertices such that the within- and between-cluster discrepancies are minimized. Let $a(X, Y) = \sum_{u\in X}\sum_{v\in Y} a_{ij}$ be the edge-cut between $X, Y \subset V$, and $\mathrm{Vol}(X) = \sum_{v\in X} d_v$ be the volume of the vertex-subset X. Further, let $\rho(X, Y) := \frac{a(X,Y)}{\mathrm{Vol}(X)\mathrm{Vol}(Y)}$ be the volume–density between X and Y. The multiway discrepancy [6] of $G = (V, \boldsymbol{A})$ in the clustering $V_1, \dots, V_k$ of its vertices is

$$\mathrm{md}(G; V_1, \dots, V_k) = \max_{1\leq i\leq j\leq k} \max_{X\subset V_i,\, Y\subset V_j} \frac{|a(X, Y) - \rho(V_i, V_j)\mathrm{Vol}(X)\mathrm{Vol}(Y)|}{\sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}}.$$

The *minimum k-way discrepancy* of G is $\mathrm{md}_k(G) = \min_{(V_1,\dots,V_k)\in\mathcal{P}_k} \mathrm{md}(G; V_1, \dots, V_k)$, where $\mathcal{P}_k$ denotes the set of proper k-partitions of V. This generalizes the notion of the volume-regular cluster pairs of [2] in the following way. If $\alpha = \mathrm{md}(G; V_1, \dots, V_k)$, then α is the smallest positive constant such that for every V_i, V_j pair and for every $X \subset V_i$, $Y \subset V_j$,

$$|a(X, Y) - \rho(V_i, V_j)\mathrm{Vol}(X)\mathrm{Vol}(Y)| \leq \alpha\sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}$$

* Corresponding author.
E-mail addresses: marib@math.bme.hu (M. Bolla), ahmed@math.bme.hu (A. Elbanna).

https://doi.org/10.1016/j.dam.2018.02.016
0166-218X/© 2018 Elsevier B.V. All rights reserved.

holds. In [2], the α volume-regularity of the cluster pairs requires only

$$|a(X,Y)-\rho(V_i,V_j)\mathrm{Vol}(X)\mathrm{Vol}(Y)| \le \alpha\sqrt{\mathrm{Vol}(V_i)\mathrm{Vol}(V_j)}, \tag{1}$$

a condition on which the estimate of [4] is based.

In Section 2, we give a strengthened discrepancy estimate by means of spectral tools. We use the following measure for the 'goodness' of a k-clustering. Assume that $|\mu_{k-1}| > |\mu_k|$, and denote by $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$ the unit-norm, pairwise orthogonal eigenvectors, corresponding to the leading (non-trivial) eigenvalues $\mu_1, \dots, \mu_{k-1}$ of $\boldsymbol{A_D}$. Let $\mathbf{r}_1, \dots, \mathbf{r}_n \in \mathbb{R}^{k-1}$ be the row vectors of the $n\times(k-1)$ matrix of column vectors $\boldsymbol{D}^{-1/2}\mathbf{u}_1, \dots, \boldsymbol{D}^{-1/2}\mathbf{u}_{k-1}$; they are called $(k-1)$-dimensional representatives of the vertices. The *weighted k-variance* of these representatives is defined as

$$\tilde{S}_k^2 = \min_{(V_1,\dots,V_k)\in\mathcal{P}_k} \sum_{i=1}^k \sum_{v\in V_i} d_v \|\mathbf{r}_v - \mathbf{c}_i\|^2,$$

where $\mathbf{c}_i = \frac{1}{\mathrm{Vol}(V_i)}\sum_{v\in U_i} d_v \mathbf{r}_v$ is the weighted center of the cluster V_i. Actually, $\tilde{S}_k^2$ is the minimal distance between the subspace spanned by the vectors $\mathbf{1}, \boldsymbol{D}^{-1/2}\mathbf{u}_1, \dots, \boldsymbol{D}^{-1/2}\mathbf{u}_{k-1}$ and that spanned by step-vectors over the k-partitions in $\mathcal{P}_k$, see [5]. It is the *weighted k-means algorithm* that provides this minimum. This is the generalization of the k-means algorithm, for which there are polynomial time approximating schemes (PTAS), see [12]. The so-called spectral relaxation (base of the spectral clustering) means that we can approximately find discrepancy minimizing clustering via applying the weighted k-means algorithm to the $(k-1)$-dimensional vertex representatives.

In Theorem 1, for a general term of a graph sequence having 'no dominant vertices', we estimate $\mathrm{md}_k(G)$ with $|\mu_k|$, $\tilde{S}_k$, and exact constants by strengthening the estimate of Theorem 2.1 of [4] in that there instead of exact constants only big-O, and instead of $\mathrm{md}_k(G)$ the weaker discrepancy measure (1) for the volume-regularity of cluster pairs was used.

Conversely, in [6] we estimated $|\mu_k|$ with $\mathrm{md}_k(G)$ as

$$|\mu_k| \le 9\,\mathrm{md}_k(G)(k+2-9k\ln \mathrm{md}_k(G)). \tag{2}$$

Putting these back and forth statements together, in Section 3 we establish an important relation between the gaps in the spectral and in the discrepancy view for expanding graph sequences, see Corollary 2. Hence, we justify for the discrepancy minimizing spectral clustering, and contribute to the characterization of the generalized quasirandom graphs (see [11]) by spectra and spectral subspaces.

So we are able to prove back and forth relations between k-clusterings of the vertices and the inner spectral gap (at the kth eigenvalue) of the normalized adjacency spectrum, with the same k, thus also indicating the choice of the optimal k. Relations between the overall discrepancy of a graph and the spectral gap at the edge of the spectrum have long been investigated together with quasirandomness, e.g., Thomason [13], Chung–Graham–Wilson [10], Bollobás–Nikiforov [8], Bilu–Linial [3], Butler [9]. However, their results apply to the $k=1$ case. Similar results for bipartite graphs, e.g., Alon [1] and Thomason [14], correspond to the $k=2$ case, but their findings do not extend trivially to a general k. For simulation results and applications see [7].

2. The strengthened discrepancy estimate

Theorem 1. *Let G_n be the general term of a connected simple graph sequence, G_n has n vertices. (We do not denote the dependence of the vertex-set V and adjacency matrix $\boldsymbol{A}$ of G_n on n.) Assume that there are constants $0 < c < C < 1$ such that except $o(n)$ vertices, the degrees satisfy $cn \le d_v \le Cn$, $v = 1, \dots, n$. Let the eigenvalues of the normalized adjacency matrix of G_n, enumerated in decreasing absolute values, be*

$$1 = \mu_0 \ge |\mu_1| \ge \dots \ge |\mu_{k-1}| > \varepsilon \ge |\mu_k| \ge \dots \ge |\mu_{n-1}| = 0.$$

The partition $(V_1, \dots, V_k)$ of V is defined so that it minimizes the weighted k-variance $s^2 = \tilde{S}_k^2$ of the optimal $(k-1)$-dimensional vertex representatives of G_n. Assume that this k-partition $(V_1, \dots, V_k)$ satisfies the balancing condition: $\frac{|V_i|}{n} \to r_i$ for $i = 1, \dots, k$ as $n \to \infty$ with some positive reals $r_1, \dots, r_k$. Then

$$\mathrm{md}_k(G_n) \le \mathrm{md}(G_n; V_1, \dots, V_k) \le 2\left(\frac{C}{c} + o(1)\right)(\sqrt{2k}s + \varepsilon).$$

Proof. At the beginning, we follow the proof of [4], the idea of which is summarized briefly. Denote by $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{k-1}$ the unit-norm, pairwise orthogonal eigenvectors of $\boldsymbol{A}_D$ corresponding to the k largest (in absolute value) eigenvalues $\mu_0, \mu_1, \dots, \mu_{k-1}$ of $\boldsymbol{A_D}$. The $(k-1)$-dimensional representatives of the vertices of G_n are row vectors of the matrix with column vectors $\mathbf{x}_i = \boldsymbol{D}^{-1/2}\mathbf{u}_i$ $(i = 1, \dots, k-1)$. The representatives can as well be regarded as k-dimensional ones, as by inserting the vector $\mathbf{x}_0 = \boldsymbol{D}^{-1/2}\mathbf{u}_0$ will not change the weighted k-variance $s^2 = \tilde{S}_k^2$, because it is the vector of all 1 coordinates. Suppose that the minimum weighted k-variance is attained at the k-partition $(V_1, \dots, V_k)$ of the vertices. By an easy analysis of variance argument (see [5]) it follows that $s^2 = \sum_{i=0}^{k-1}\mathrm{dist}^2(\mathbf{u}_i, F)$, where $F = \mathrm{Span}\{\boldsymbol{D}^{1/2}\mathbf{z}_1, \dots, \boldsymbol{D}^{1/2}\mathbf{z}_k\}$ with the so-called normalized partition vectors $\mathbf{z}_1, \dots, \mathbf{z}_k$ of coordinates $z_{ji} = \frac{1}{\sqrt{\mathrm{Vol}(V_i)}}$ if $j \in V_i$ and 0, otherwise $(i = 1, \dots, k)$.

Note that the vectors $\boldsymbol{D}^{1/2}\mathbf{z}_1, \dots, \boldsymbol{D}^{1/2}\mathbf{z}_k$ form an orthonormal system. By [5], we can find another orthonormal system $\mathbf{v}_0, \dots, \mathbf{v}_{k-1} \in F$ such that $s^2 \le \sum_{i=0}^{k-1} \|\mathbf{u}_i - \mathbf{v}_i\|^2 \le 2s^2$ (note that $\mathbf{v}_0 = \mathbf{u}_0$). Then we approximate the matrix $\boldsymbol{A_D} = \sum_{i=0}^{n-1} \mu_i \mathbf{u}_i \mathbf{u}_i^T$ by the rank k matrix $\sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T$ with the following accuracy (in spectral norm):

$$\left\| \sum_{i=0}^{n-1} \mu_i \mathbf{u}_i \mathbf{u}_i^T - \sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T \right\| \le \sqrt{2k}s + \varepsilon, \tag{3}$$

see [4]. Based on these considerations and relation between the cut norm and the spectral norm, we rewrite our estimates in terms of stepwise constant vectors in the following way. The vectors $\mathbf{y}_i := \boldsymbol{D}^{-1/2}\mathbf{v}_i$ are stepwise constants on the partition $(V_1, \dots, V_k)$, $i = 0, \dots, k-1$. The matrix $\sum_{i=0}^{k-1} \mu_i \mathbf{y}_i \mathbf{y}_i^T$ is therefore a symmetric block-matrix on $k \times k$ blocks belonging to the above k-partition of the vertices. Let $\hat{a}_{ij}$ denote its entries in the (i, j) block $(i, j = 1, \dots, k)$. Using (3), the rank k approximation of the matrix $\boldsymbol{A}$ is performed with the following accuracy of the perturbation $\boldsymbol{E} = (\eta_{uv})$:

$$\|\boldsymbol{E}\| = \left\| \boldsymbol{A} - \boldsymbol{D}(\sum_{i=0}^{k-1} \mu_i \mathbf{y}_i \mathbf{y}_i^T)\boldsymbol{D} \right\| = \left\| \boldsymbol{D}^{1/2}(\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2} - \sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T)\boldsymbol{D}^{1/2} \right\|.$$

Therefore, the entries of $\boldsymbol{A}$ – for $u \in V_i$, $v \in V_j$ – can be decomposed as $a_{uv} = d_u d_v \hat{a}_{ij} + \eta_{uv}$.

From here, we develop the strengthened estimate. The cut-norm of the $n \times n$ symmetric error matrix $\boldsymbol{E}$, restricted to $X \times Y$ (otherwise it contains all zero entries) and denoted by $\boldsymbol{E}_{XY}$, is estimated as follows:

$$\begin{aligned}
\|\boldsymbol{E}_{XY}\|_\square &\le \sqrt{|X||Y|}\|\boldsymbol{E}_{XY}\| \le \sqrt{|X||Y|}\|\boldsymbol{D}_X^{1/2}\|(\sqrt{2k}s + \varepsilon)\|\boldsymbol{D}_Y^{1/2}\| \\
&\le \sqrt{|X||Y|}\sqrt{\frac{C}{c}\frac{\mathrm{Vol}(X)}{|X|}}\sqrt{\frac{C}{c}\frac{\mathrm{Vol}(Y)}{|Y|}}(\sqrt{2k}s + \varepsilon) \\
&= \frac{C}{c}\sqrt{\mathrm{Vol}(X)}\sqrt{\mathrm{Vol}(Y)}(\sqrt{2k}s + \varepsilon).
\end{aligned}$$

Here the diagonal matrix $\boldsymbol{D}_X$ contains the diagonal part of $\boldsymbol{D}$ restricted to X, otherwise zeros. But by the degree conditions,

$$\|\boldsymbol{D}_X\| = \max_{v \in X} d_v \le \frac{C}{c} \min_{v \in X \setminus V_0} d_v \le \left(\frac{C}{c} + o(1)\right)\frac{\mathrm{Vol}(X)}{|X|},$$

where V_0 is the exceptional class of size $o(n)$ (where the degree conditions do not hold) and the constants c and C do not depend on n. We also used that $\mathrm{Vol}(X) = \sum_{v \in X \setminus V_0} d_i + o(n)$. Likewise,

$$\|\boldsymbol{E}_{V_iV_j}\|_\square \le \frac{C}{c}\sqrt{\mathrm{Vol}(V_i)}\sqrt{\mathrm{Vol}(V_j)}(\sqrt{2k}s + \varepsilon)$$

and, from the balancing condition, $|V_i| = r_i n + o(n)$.

Consequently, for $i, j = 1, \dots, k$ and $X \subseteq V_i$, $Y \subseteq V_j$:

$$\begin{aligned}
&\left| a(X, Y) - \rho(V_i, V_j)\mathrm{Vol}(X)\mathrm{Vol}(Y) \right| \\
&= \left| \sum_{v \in X}\sum_{u \in Y}(d_v d_u \hat{a}_{ij} + \eta_{uv}) - \frac{\mathrm{Vol}(X)\mathrm{Vol}(Y)}{\mathrm{Vol}(V_i)\mathrm{Vol}(V_j)}\sum_{v \in V_i}\sum_{u \in V_j}(d_u d_v \hat{a}_{ij} + \eta_{uv}) \right| \\
&= \left| \sum_{v \in X}\sum_{u \in Y}\eta_{uv} - \frac{\mathrm{Vol}(X)\mathrm{Vol}(Y)}{\mathrm{Vol}(V_i)\mathrm{Vol}(V_j)}\sum_{v \in V_i}\sum_{u \in V_j}\eta_{uv} \right| \\
&\le \left(\frac{C}{c} + o(1)\right)\sqrt{\mathrm{Vol}(X)}\sqrt{\mathrm{Vol}(Y)}(\sqrt{2k}s + \varepsilon) \\
&+ \frac{\mathrm{Vol}(X)\mathrm{Vol}(Y)}{\mathrm{Vol}(V_i)\mathrm{Vol}(V_j)}\left(\frac{C}{c} + o(1)\right)\sqrt{\mathrm{Vol}(V_i)}\sqrt{\mathrm{Vol}(V_j)}(\sqrt{2k}s + \varepsilon) \\
&= \left(\frac{C}{c} + o(1)\right)\sqrt{\mathrm{Vol}(X)}\sqrt{\mathrm{Vol}(Y)}(\sqrt{2k}s + \varepsilon) \\
&+ \frac{\sqrt{\mathrm{Vol}(X)}\sqrt{\mathrm{Vol}(Y)}}{\sqrt{\mathrm{Vol}(V_i)}\sqrt{\mathrm{Vol}(V_j)}}\sqrt{\mathrm{Vol}(X)}\sqrt{\mathrm{Vol}(Y)}\left(\frac{C}{c} + o(1)\right)(\sqrt{2k}s + \varepsilon) \\
&\le 2\left(\frac{C}{c} + o(1)\right)(\sqrt{2k}s + \varepsilon)\sqrt{\mathrm{Vol}(X)}\sqrt{\mathrm{Vol}(Y)}
\end{aligned}$$

that gives the required estimate. □

3. Conclusion

Corollary 2. *Let G_n be the general term of a sequence of simple graphs; G_n has n vertices and $n \times n$ normalized adjacency matrix $\boldsymbol{A}_{\boldsymbol{D},n}$; further, satisfies the degree conditions of Theorem 1. Let k be a fixed positive integer, whereas $n \to \infty$. Consider the following two properties:*

(a) *There exists a constant $0 < \delta < 1$ (independent of n) such that $\boldsymbol{A}_{\boldsymbol{D},n}$ has $k-1$ structural eigenvalues (except the trivial 1) that are greater than δ (in absolute value), while the remaining eigenvalues are $o(1)$. The weighted k-variance $\tilde{S}^2_{k,n}$ of the $(k-1)$-dimensional vertex representatives, based on the transformed eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{A}_{\boldsymbol{D},n}$, is $o(1)$. The k-partition $(V_{1,n}, \dots, V_{k,n})$ minimizing this k-variance satisfies the following: there are positive reals $r_1, \dots, r_k$ such that*

$$\frac{|V_{i,n}|}{n} \to r_i \quad \text{for} \quad i = 1, \dots, k \quad \text{as} \quad n \to \infty. \tag{4}$$

(b) *There are vertex-classes $V_{1,n}, \dots, V_{k,n}$ obeying the balancing condition (4), and there is a constant $0 < \theta < 1$ (independent of n) such that $\mathrm{md}_1(G_n) > \theta, \dots, \mathrm{md}_{k-1}(G_n) > \theta$, and $\mathrm{md}(G_n; V_{1,n}, \dots, V_{k,n}) = o(1)$.*

Then property (a) *implies property* (b).

Proof. Assume that there is a constant $0 < \delta < 1$ such that $\boldsymbol{A}_{\boldsymbol{D},n}$ has $k-1$ eigenvalues (except the trivial 1) that are greater than δ in absolute value, while the remaining eigenvalues are $o(1)$; further, the squareroot of weighted k-variance $\tilde{S}^2_{k,n}$ is also $o(1)$. Using that there are no dominant vertices, we apply Theorem 1. According to this, $\mathrm{md}(G_n; V_{1,n}, \dots, V_{k,n}) = o(1)$, where $V_{1,n}, \dots, V_{k,n}$ are the spectral clusters.

Now indirectly assume that there is no absolute constant $0 < \theta < 1$ such that $\mathrm{md}_1(G_n) > \theta, \dots, \mathrm{md}_{k-1}(G_n) > \theta$. Then there is an $1 \le i \le k-1$ with $\mathrm{md}_i(G_n) \le \varepsilon$, for any $0 < \varepsilon < 1$. But Inequality (2) estimates $|\mu_{n,i}|$ with a (near zero) strictly increasing function of $\mathrm{md}_i(G_n)$. In view of this, there should be an $0 < \varepsilon' < 1$ so that $|\mu_{n,i}| \le \varepsilon'$, where ε' can be any small positive number (depending on ε). This contradicts to the $|\mu_{n,i}| > \delta$ assumption. □

The message of the above statement is that a sudden gap in the spectrum and cluster variances is an indication of a sudden gap in the multiway discrepancies. The (a) → (b) implication may be converted with some additional conditions when it is included in the chain of implications between so-called generalized quasirandom properties. If the chain is closed through other properties and with some additional conditions, then (b) → (a) also holds, and our graph sequence is generalized quasirandom in the sense of [11]. Simulation results supporting this idea are shown in [7].

References

[1] N. Alon, Eigenvalues and expanders, Combinatorica 6 (1986) 83–96.
[2] N. Alon, A. Coja-Oghlan, H. Han, M. Kang, V. Rödl, M. Schacht, Quasi-randomness and algorithmic regularity for graphs with general degree distributions, SIAM J. Comput. 39 (6) (2010) 2336–2362.
[3] Y. Bilu, N. Linial, Lifts, discrepancy and nearly optimal spectral gap, Combinatorica 26 (2006) 495–519.
[4] M. Bolla, Spectra and structure of weighted graphs, Electron. Notes Discrete Math. 38 (2011) 149–154.
[5] M. Bolla, Spectral Clustering and Biclustering, Wiley, 2013.
[6] M. Bolla, Relating multiway discrepancy and singular values of nonnegative rectangular matrices, Discrete Appl. Math. 203 (2016) 26–34.
[7] M. Bolla, A. Elbanna, Matrix and discrepancy view of generalized random and quasirandom graphs, Spec. Matrices 4 (1) (2016) 31–45.
[8] B. Bollobás, V. Nikiforov, Hermitian matrices and graphs: singular values and discrepancy, Discret. Math. 285 (2004) 17–32.
[9] S. Butler, Using discrepancy to control singular values for nonnegative matrices, Linear Algebra Appl. 419 (2006) 486–493.
[10] F. Chung, R. Graham, R.K. Wilson, Quasi-random graphs, Combinatorica 9 (1989) 345–362.
[11] L. Lovász, V. T.-Sós, Generalized quasirandom graphs, J. Combin. Theory Ser. B 98 (2008) 146–163.
[12] R. Ostrovsky, Y. Rabani, L.J. Schulman, C. Swamy, The effectiveness of lloyd-type methods for the k-means problem, J. ACM 28 (2012) 59.6.
[13] A. Thomason, Pseudo-random graphs, in: M. Karónski (Ed.), Proc. RandOm Graphs, Poznán, 1985, Ann. Discrete Math. 33 (1987) 307–331.
[14] A. Thomason, Dense expanders and pseudo-random bipartite graphs, Discrete Math. 75 (1989) 381–386.