Notes and errata to the book

Bolla, M., Spectral Clustering and Biclustering.
Learning Large Graphs and Contingency Tables, Wiley (2013)

Notes

- In the Introduction we promise to connect graph theory and statistics. Under this we mean that we apply the techniques of multivariate statistical methods to find a hidden (low dimensional) structure in the graph, based on its spectrum and spectral subspaces. We indeed make assessments on the dimension of the low-dimensional vertex representatives, and relate spectra to multiway cuts and discrepancies, a new paradigm of spectral clustering.

  In the last section we also consider parametric probabilistic models, and use the iteration of the EM algorithm for parameter estimation, where the starting clustering is obtained by spectral clustering tools. However, we do not deal with graphical models at all in the book. Graphical models provide a framework for describing statistical dependencies in a collection of random variables. They rely on probability and information theory, and though their techniques use graph decompositions, those have nothing to do with spectral clustering.

- In Section 2.4, we note that the complete and complete multipartite graphs have the zero as the largest modularity eigenvalue. Dragan Stevanovic conjectured (posed as an open problem in July 2012, at the Conference on Applications of Graph Spectra in Computer Science, Barcelona) that these are the only graphs with a negative semidefinite modularity matrix (of which zero is always an eigenvalue). In 2013, I was leading an undergraduate research course on spectral clustering at the Budapest Semester of Mathematics for visiting undergraduate students from the US. I posed there this problem, and just when the proofreading of this book was over, we managed to give an affirmative answer to this question and proved that *the modularity matrix of a simple connected graph is negative semidefinite if and only if it is complete multipartite.* (Note that a complete graph is also considered as complete multipartite with singleton classes.)

  The same is true for the normalized modularity matrix. The backward statement can be extended to edge-weighted graphs showing that if certain patterns appear in the edge-weight matrix, then the modularity matrix should be indefinite. The above fact has important consequences for the isoperimetric inequality, the symmetric maximal correlation, and the Newman–Girvan modularity, see

  Bolla, M., Bullins, B., Chaturapruek, S., Chen, S., Friedl, K., Spectral properties of modularity matrices, *Linear Algebra and Its Applications* **73** (2015), 359-376.
  http://dx.doi.org/10.1016/j.laa.2014.10.039

- In Section 3.3 (see Theorems 3.3.3 and 3.3.8) we managed to estimate the discrepancy of the cluster pairs (in the graph and rectangular array setup) in a k-clustering by means of the gap between the k structural and the other singular values and the k-variance of the vertex representatives (based on the eigenvectors corresponding to the structural eigenvalues). Since then, we have managed to prove the converse: estimated the kth largest singular value by a (near 0) strictly increasing function of the *k-way discrepancy*, defined for this purpose. The back and forth results are generalizations of the expander mixing lemma and and its converse for the k-cluster case, see

Bolla, M., Relating multiway discrepancy and singular values of nonnegative rectangular matrices, *Discrete Applied Mathematics* **203** (2016), 26-34.
http://dx.doi.org/10.1016/j.dam.2015.09.013

Errata

- In the Acknowledgements: Katalin Friedl, not Katalin Friedll.
- p.26: in Definition 1.4.1, after the first sentence we should include: The joint distribution of X_i and Y_i is $\mathbb{W}$ $(i = 1, \ldots, k)$.
- p.27, l.12: in the sentence 'Observe that in the case of an irreducible contingency table', instead of irreducible the correct wording is *non-decomposable* (irreducibility is defined for quadratic matrices). See Definition A.3.28, where the notion of a *decomposable* rectangular matrix is defined. Note that sometimes it is called *degenerate*.
- p.50, l.6: correctly, row vectors of $\mathbf{Z}_k$.
- p.71: in the main formula of Proposition 2.3.3, in the second line under the min, correctly $\psi, \psi' \in H$ *i.d.* should stand.
- p.90: in the first line of the top formula, correctly $\frac{\sigma_{bb}}{\sqrt{\mathrm{Vol}(C_b)}}$ should stand.
- p.115: Table 3.1. correctly is as follows (there are some minor changes).

Graph $G = (V, \boldsymbol{B})$	Adjacency matrix $\boldsymbol{B}$	Laplacian matrix $\boldsymbol{D} - \boldsymbol{B}$	Normalized Laplacian $\boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{B}\boldsymbol{D}^{-1/2}$	Normalized modularity $\boldsymbol{D}^{-1/2}\boldsymbol{B}\boldsymbol{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$
$\boldsymbol{B} = \oplus_{i=1}^{k} \boldsymbol{B}_i$, where the $n_i \times n_i$ $\boldsymbol{B}_i$ has diagonal ν_i and off-diagonal μ_i $V = (V_1, \dots, V_k)$ $\lvert V_i \rvert = n_i$ $(i = 1, \dots, k)$	$\lambda_i = (n_i - 1)\mu_i + \nu_i$ $(i = 1, \dots, k)$ with piecewise constant eigenvectors over V_i's; $\nu_i - \mu_i$ with multiplicity $n_i - 1$, and eigenvectors with 0-sum coordinates over V_i $(i = 1, \dots, k)$	0 with multiplicity k and piecewise constant eigenvectors over V_i's; $n_i\mu_i$ with multiplicity $n_i - 1$, and eigenvectors with 0-sum coordinates over V_i $(i = 1, \dots, k)$	0 with multiplicity k and stepwise constant eigenvectors over V_i's; $\frac{n_i\mu_i}{\nu_i + (n_i - 1)\mu_i}$ with multiplicity $n_i - 1$, and eigenvectors with 0-sum coordinates over V_i $(i = 1, \dots, k)$	1 with multiplicity $k - 1$ and stepwise constant eigenvectors over V_i's; $1 - \frac{n_i\mu_i}{\nu_i + (n_i - 1)\mu_i}$ with multiplicity $n_i - 1$, and eigenvectors with 0-sum coordinates over V_i $(i = 1, \dots, k)$
$G = K_{n_1, \dots, n_k}$ with independent sets V_i's, $\lvert V_i \rvert = n_i$ $(i = 1, \dots, k)$. w.l.g. assume that $n_1 \le \dots \le n_k$ $(n = \sum_{i=1}^{k} n_i)$	0 with multiplicity $n - k$ with eigenvectors of 0-sum coordinates over V_i's; the other k eigenvalues are in $[-n_k, -n_1] \cup [n - n_k, n - n_1]$ with piecewise constant eigenvectors	0 single; n with multiplicity $k - 1$ and piecewise constant eigenvectors over V_i's; $n - n_i$ with multiplicity $n_i - 1$ and eigenvectors with 0-sum coordinates over V_i's	0 single; 1 with multiplicity $n - k$; $k - 1$ eigenvalues in $[1 + \delta, 2]$, where δ does not depend on n under $\frac{n_i}{n} \ge c$ $(i = 1, \dots, k)$	0 with multiplicity $n - k$; $k - 1$ eigenvalues in $[-1, -\delta]$, where δ does not depend on n under $\frac{n_i}{n} \ge c$ $(i = 1, \dots, k)$
$\boldsymbol{B}$ is the blown-up matrix of $\boldsymbol{P} = (p_{ij})$ $(i, j = 1, \dots, k)$, with blow-up sizes $n_1, \dots, n_k$ and clusters $V_1, \dots, V_k$; $\lvert V_i \rvert = n_i$ $(n = \sum_{i=1}^{k} n_i)$ $\mathrm{rank}(\boldsymbol{P}) = k$, $\frac{n_i}{n} \ge c$	0 with multiplicity $n - k$ with eigenvectors of 0-sum coordinates over V_i's; k non-zero eigenvalues $\lambda_1, \dots, \lambda_k = \Theta(n)$ with piecewise constant eigenvectors over $V_1, \dots, V_k$	0 single; $\lambda_1, \dots, \lambda_{k-1} = \Theta(n)$ with piecewise constant eigenvectors over V_i's; $\gamma_i = \sum_{j \ne i} n_j p_{ij}$ with multiplicity $n_i - 1$ and zero-sum coordinates over V_i $(i = 1, \dots, k)$; $\sum_{i=1}^{k-1} \lambda_i = \sum_{i=1}^{k} \gamma_i$	$\exists 0 < \delta < 1$ s.t. there are k eigenvalues (including the 0) in $[0, 1 - \delta] \cup [1 + \delta, 2]$ with piecewise constant eigenvectors over V_i's, and the 1 with multiplicity $n - k$	$\exists 0 < \delta < 1$ s.t. there are $k - 1$ eigenvalues (excluding the 1) in $[-1, -\delta] \cup [\delta, 1)$ with piecewise constant eigenvectors over V_i's, and the 0 with multiplicity $n - k$

Table 1: Spectra and spectral subspaces of some special block- and blown-up matrices

- p.117, l.1: the title of Lemma 3.1.16 correctly is Berstein inequality, not Chernoff inequality (this form of a large deviation theorem is attributed to Bernstein, originally).

- p.118, l.4: $\boldsymbol{A}_n = \boldsymbol{B}_n + \boldsymbol{E}_n$ is the correct formula.

- p.119, l.-2: Lemma 3.1.18, not Theorem 3.1.8.

- p.125: in the last paragraph of this page, there are problems with the indexing. This paragraph correctly is as follows.

  'With an appropriate Wigner-noise we can guarantee that the noisy table $\boldsymbol{A}_{m\times n}$ contains 1's in the (u,v)-th block with probability p_{uv}, and 0's otherwise. Indeed, for indices $1 \le u \le a$, $1 \le v \le b$, and $i \in R_u$, $j \in C_v$ let
  $$w_{ij} := \begin{cases} 1 - p_{uv} & \text{with probability} \quad p_{uv} \\ -p_{uv} & \text{with probability} \quad 1 - p_{uv} \end{cases} \tag{1}$$
  be independent random variables. This $\boldsymbol{W}_{m\times n}$ satisfies the conditions of Definition 3.2.1 with uniformly bounded entries of zero expectation. Let $R_1, \dots, R_a$ and $C_1, \dots, C_b$ denote the row- and column clusters induced by the blow-up. In the random 0-1 contingency table $\boldsymbol{A}_{m\times n}$, the row and column categories of R_u and C_v are in interaction with probability p_{uv}.'

- p.143: the long formula starting in l.11 is as follows, correctly.
$$\begin{aligned} |w(X,Y) - \mathrm{Vol}(X)\mathrm{Vol}(Y)| &= \left| \sum_{i=1}^{n-1} \mu_i a_i b_i \right| \le \max_{i\ge 1} |\mu_i| \sum_{i=1}^{n-1} |a_i||b_i| \\ &\le \|\boldsymbol{M}_D\| \cdot \sqrt{\sum_{i=1}^{n-1} a_i^2 \sum_{i=1}^{n-1} b_i^2} \\ &= \|\boldsymbol{M}_D\| \cdot \sqrt{\mathrm{Vol}(X)(1-\mathrm{Vol}(X))\mathrm{Vol}(Y)(1-\mathrm{Vol}(Y))} \\ &\le \|\boldsymbol{M}_D\| \cdot \sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}, \end{aligned}$$

- p.151: the long formula starting in l.10 is as follows, correctly.
$$\begin{aligned} |c(R,C) - \mathrm{Vol}(R)\mathrm{Vol}(C)| &= \left| \sum_{i=1}^{r-1} s_i a_i b_i \right| \le s_1 \sum_{i=1}^{r-1} |a_i||b_i| \\ &\le s_1 \sqrt{\sum_{i=1}^{r-1} a_i^2 \sum_{i=1}^{r-1} b_i^2} \\ &= s_1 \sqrt{\mathrm{Vol}(R)(1-\mathrm{Vol}(R))\mathrm{Vol}(C)(1-\mathrm{Vol}(C))} \\ &\le s_1 \sqrt{\mathrm{Vol}(R)\mathrm{Vol}(C)}, \end{aligned}$$

- p.152, last line: correctly $\mathbb{R}^m$, not $\mathbb{R}^n$.

- p.153, first line: correctly $\mathbb{R}^n$, not $\mathbb{R}^m$.

- p.175, l.8. of Section 4.5: in the parenthesis, 'note that $G_{\boldsymbol{B}_n}$' stands correctly.

- p.180, in Theorem 4.6.3, the following modifications (for the better understanding) should be made. As the first sentence, include 'Let $(G_n) \to W$'. In the 5th line of this theorem, correctly write $\boldsymbol{D}_n^{-1/2}\mathbf{u}_{n,1}, \ldots, \boldsymbol{D}_n^{-1/2}\mathbf{u}_{n,k-1}$.
- p.206. The citation correctly:
  Holland P, Laskey KB and Leinhardt S 1983 Stochastic blockmodels: some first steps. *Social Networks* **5**, 109–137.
- p.222: in Equation (A.11), $\boldsymbol{A}^T$ should be written, not $\boldsymbol{A}^*$.
- p.231, l.4: Note that in Definition A.3.28, the notion of a *decomposable* rectangular matrix is equivalent to the notion of a *degenerate* one.
- p.237, l.-3: 'conitional expectations' should stand correctly, not 'conditional distributions'.
- p.239, l.-8: Proposition B.1.5 should stand, not Theorem B.1.5.
- p.251: in the second Equation, under the max, in the 3rd line $\mathrm{Cov}_{\mathbb{P}}\psi\psi_i = 0$, $i = 1, \ldots, k-1$, and in the 4th line $\mathrm{Cov}_{\mathbb{Q}}\phi\phi_i = 0$, $i = 1, \ldots, k-1$ should stand correctly.
- p.254, l.5: 'structure matrix' instead of 'design matrix' is a better wording.
- p.268, in the Index: under Szemerédi's regularity lemma, weak, 163, 166, 186 are the correct numbers of the pages.

Marianna Bolla

Last modified on August 9, 2016.