Spectral Partitioning of Random Graphs with Given Expected Degrees

Amin Coja-Oghlan[1], Andreas Goerdt[2], and André Lanka[2]

[1] Humboldt Universität zu Berlin, Institut für Informatik
Unter den Linden 6, 10099 Berlin, Germany
`coja@informatik.hu-berlin.de`
[2] Fakultät für Informatik, Technische Universität Chemnitz
Straße der Nationen 62, 09107 Chemnitz, Germany
`{goerdt, lanka}@informatik.tu-chemnitz.de`

Abstract. It is a well established fact, that – in the case of classical random graphs like (variants of) $G_{n,p}$ or random regular graphs – spectral methods yield efficient algorithms for clustering (e. g. colouring or bisection) problems. The theory of large networks emerging recently provides convincing evidence that such networks, albeit looking random in some sense, cannot sensibly be described by classical random graphs. A variety of new types of random graphs have been introduced. One of these types is characterized by the fact that we have a fixed expected degree sequence, that is for each vertex its expected degree is given.
Recent theoretical work confirms that spectral methods can be successfully applied to clustering problems for such random graphs, too – provided that the expected degrees are not too small, in fact $\geq \log^6 n$. In this case however the degree of *each* vertex is concentrated about its expectation. We show how to remove this restriction and apply spectral methods when the expected degrees are bounded below just by a suitable constant. Our results rely on the observation that techniques developed for the classical sparse $G_{n,p}$ random graph (that is $p = c/n$) can be transferred to the present situation, when we consider a suitably normalized adjacency matrix: We divide each entry of the adjacency matrix by the product of the expected degrees of the incident vertices. Given the host of spectral techniques developed for $G_{n,p}$ this observation should be of independent interest.

1 Introduction

For definiteness we specify the model of random graphs to be considered first. This model is very similar to that considered and convincingly motivated in [9]. (In particular, we refer to Subsection 1.3 of that paper where the model is defined.)

1.1 The model

Our random graphs with planted partition and given expected degree sequence are generated as follows. Let $V = \{1, \ldots, n\}$ be the set of nodes. Partition V into k disjoint subsets $V_1, \ldots, V_k$, where k is fixed. We assume that the size of each set $|V_j| \geq \delta n$ for some arbitrarily small but constant $\delta > 0$. For $i \in V$ we let $\psi(i)$ denote the number of the subset i belongs to, that is $i \in V_{\psi(i)}$. We fix some symmetric $k \times k$-matrix $D = (d_{ij})$ with non-negative constants as entries. Moreover, we assign some weight w_i to each node $i \in V$. We let $W = \sum w_i$ and $\overline{w} = W/n$ be the arithmetic mean of the w_i's. We construct the random graph $G = (V, E)$ by inserting each edge $\{i, j\}$ independently with probability $w_i \cdot w_j \cdot d_{\psi(i),\psi(j)}/W$. Of course the parameters should be chosen such that each probability is bounded above by 1. (It has some mild technical advantages to allow for loops as we do.) Depending on the matrix D, we can model a variety of random instances of clustering problems. For example we can generate 3-colourable graphs, then the V_j are the colour classes, or graphs having a small bisection, in which case the V_j are the two sides of the bisection, or graphs with subsets of vertices which are very dense or sparse... The algorithmic problem is to efficiently reconstruct the V_j (or large parts thereof) given such a random G. Note that the model from [9] allows for directed edges where the minimum expected in- and out-degree of each vertex is $\log^6 n$. We restrict our attention to undirected graphs. We denote the expected degree of vertex i by w_i', then

$$w_i' = \frac{w_i}{W} \cdot \sum_{j \in V} w_j \cdot d_{\psi(i),\psi(j)}.$$

In order for our algorithm to work properly we impose the following restrictions on the model's parameters:

1. The matrix D has full rank.
2. We have $w_i \geq \varepsilon \cdot \overline{w}$ for all i, where ε is some arbitrarily small constant.
3. $\overline{w} \geq d$, where $d = d(\varepsilon, D, \delta)$ is a sufficiently large constant.

Our asymptotics is such that n gets large, while $D, k, \varepsilon, \delta, d$ are fixed. On the other hand the weights w_i can be picked arbitrarily subject to our restrictions (in particular depending on n) and the subsets V_j with $|V_j| \geq \delta n$ are arbitrary, too.

Our restrictions 2. and 3. imply that $l \cdot w_i \leq w_i' \leq u \cdot w_i$ for constants $l = l(\varepsilon, D, \delta)$ and $u = u(\varepsilon, D, \delta)$ that is $w_i' = \Theta(w_i)$. This shows the extent to which we consider graphs with given expected degree sequence. Note that depending on the weight w_i 2. and 3. allow w_i' among others to be constant, independent of n.

1.2 Motivation and related literature

The analysis of large real life networks, like the internet graph, social or bibliographical networks is one of the current topics not only of Computer Science.

Clearly it is important to obtain efficient algorithms adapted to the characteristics of these networks. One particular problem of interest is the problem of detecting some kind of clusters, that is subsets of vertices having extraordinarily many or few edges. Such clusters are supposed to mirror some kind of relationship among its members (= vertices of the network). Heuristics based on the eigenvalues and eigenvectors of the adjacency matrix of the network provide one of the most flexible approaches to clustering problems applied in practice. See for example [15] or the review [19] or [18]. Note that the eigenvalues and eigenvectors of *symmetric* real valued matrices, first are real valued and second can be approximated efficiently to arbitrary precision.

The relationship between spectral properties of the adjacency matrix of a graph on the one hand and clustering properties of the graph itself on the other hand is well established. Usually this relationship is based on some separation between the (absolute) values of the largest eigenvalues and the remaining eigenvalues. It has a long tradition of being exploited in practice, among others for numerical calculations. However, it is in general not easy to obtain convincing proofs certifying the quality of spectral methods in these cases, see [23] for a notable exception.

Theoretically convincing analyses of this phenomenon have been conducted in the area of random graphs. This leads to provably efficient algorithms for clustering problems in situations where purely combinatorial algorithms do not seem to work, just to cite some examples [2], [3], or [4], or the recent [20] and subsequent work such as [14]. In particular [3] has lead to further results [10], [11]. The reason for this may be that [3] is based on a rather flexible approach to obtain spectral information about random graphs [12]: Spectral information directly follows from clustering properties known to be typically present in a random graph by (inefficient) counting arguments. We apply this technique here, too.

In order to explain the success of spectral algorithms to detect clustering properties of large real life networks the preceding results do not seem to be readily applicable. As opposed to classical random graphs such networks are well known to have many vertices whose degree deviates considerably from the average degree, that is the degree distribution has a "heavy tail", or it seems to follow a "power law" , see for example [1] . And in fact in [21] it is shown that the largest eigenvalues of a random graph with power law degree distribution are proportional to the square root of the largest degrees, and thus do not reveal any non-local information about the graph. This result looks somehow related to the fact that the largest eigenvalue of a sparse random graph $G_{n,p}$ where $p = c/n$ is always the square root of the largest degree of the graph and that there is an unbounded number of eigenvalues of this size, see [16]. However, in the case of classical random graphs it helps to delete the vertices of highest degree as observed by [3] leaving the clustering properties of the graph essentially unchanged. However, in the case of a degree distribution with a heavy tail this trick is not useful, because significant parts of the graph may

just be ignored in this way. Thus, the adjacency matrix itself does not seem appropriate to represent graphs with heavy-tailed degree distributions.

To come to terms with varying degrees the Laplacian matrix is considered, see [5] for a nice exposition of the relationship of the Laplacian spectrum to clustering properties of general graphs. It is also used in practical applications, cf. [22]. However, for randomly generated graphs it is more difficult to handle theoretically than the adjacency matrix. As far as classical random graphs are concerned it is already a major difficulty to get insight into the Laplacian spectrum, at least in the interesting sparse case. The difficulty stems from the fact that in this case the graph is not asymptotically regular. See however [6] for very recent progress in this direction. Clustering problems in the denser case can be treated with the help of the Laplacian even for random graphs modelling real networks as our model does (which allows for arbitrary, in particular heavily tailed degree distributions): In [9] it is shown that the Laplacian eigenvalues allow to find the partition in the model considered here, too (provided that the average degree is $\gg \ln^6 n$). Laplacian eigenvalues of random graphs with given expected degree sequence are also investigated in [8]. Both papers rely on [13] and in part on [17] to obtain information about the spectrum. This makes it inevitable that the degree is $\gg \log n$, in fact $\geq \log^6 n$ in the case of [9]. The case of small expected degrees as considered here is interesting because the actual degree of a vertex is not any more concentrated at the expected degree. It is also mentioned in the concluding section of [9]. Independently of its applications to graph partitioning problems, we have also investigated the Laplacian eigenvalues of sparse graphs with given expected degrees in [7].

1.3 Techniques and result

We consider the following algorithm to reconstruct the V_j for random graphs as generated by our model. Only for technical simplicity we restrict our attention to $k = 2$. It poses no substantial difficulties to extend the algorithm to arbitrary, yet constant k: Instead of the two eigenvectors s_2, s_3 we use k eigenvectors $s_2, \ldots, s_{k+1}$. The sufficiently large constants C_1, C_2, C_3 depend on the actual partioning problem. The values can be calculated with the knowledge of D, ε and δ.

Algorithm 1.
> Input: The adjacency matrix A of some graph $G = (V, E)$ generated
>   in the above model and the expected degree sequence $w'_1, \ldots, w'_n$.
> Output: A partition V'_1, V'_2 of V.

1. Calculate the expected average degree, $\overline{w}' = \sum_{i=1}^{n} w'_i / n$.
2. Construct $R = (r_{ij})$ with $r_{ij} = \overline{w}'^2 \cdot a_{ij} / (w'_i \cdot w'_j)$.
3. Let $s_1 = R \cdot \mathbf{1}$ where $\mathbf{1}$ is the all one's vector.
4. Let $U = \{i \in V : \sum_{j=1}^{n} r_{ij} \leq C_1 \cdot \overline{w}'\}$ for some sufficiently large constant C_1.
5. Construct R^* from R by setting all entries r_{ij} with $i \notin U$ or $j \notin U$ to 0.

6. Calculate the eigenvectors of R^*.
7. Let s_2, s_3 be two eigenvectors of R^* belonging to different occurrences of eigenvalues which are $\geq C_2 \cdot \overline{w}'$ in absolute value.
8. At least one of the s_1, s_2, s_3 turns out to have the property that all but $C_3 \cdot (n/\overline{w}')$ entries are close to two sufficiently different values c_1, c_2. Let V_i' be all the entries close to c_i for $i = 1, 2$. Distribute the remaining entries arbitrarily among the V_i'.

Some remarks are in order. First observe that the algorithm besides the graph needs the expected degree sequence as additional information. Note that the algorithm of [9] even gets the w_i themselves. The main idea is to use the normalized adjacency matrix R, where we divide each entry of the adjacency matrix by the expected degrees of the incident vertices (the additional factor of $\overline{w}'^2$ is only for technical convenience.) It is this choice of the matrix which makes our analysis possible.

Of course, a natural idea is to divide the entries by the *actual* degrees rather than the expected degrees, in order to remove the requirement that $w_1', \ldots w_n'$ are given at the input. In fact, it turns out that this approach can be carried out successfully, i.e., the resulting matrix is suitable to recover the planted partition as well. Nonetheless, since the analysis is technically significantly more involved, we omit the details from the present extended abstract (the complete analysis will be given in the full paper version of this work).

In fact using R we get a situation formally rather similar to the case: classical sparse random graph with a planted partition and adjacency matrix, the situation as considered in [3] or [20]. Note that all entries r_{ij} with the same $(\psi(i), \psi(j))$ have the same expected value which makes the analogy possible. In particular we can apply [12]. The vector s_1 is necessary in order to recognize partitions which can be readily recognized just from the row sums of R. Step 5. has the analogous effect on the spectrum of R as has the deletion of high degree vertices in the case of sparse random graphs on the spectrum of the adjacency matrix. Being eigenvectors of different occurrences of eigenvalues, s_2 and s_3 are orthogonal to each other. Notions "vague" up to now, like "close" or the C_i, c_i in the algorithm are made precise through the subsequent proof of

Theorem 2. *Let D, ε, δ as defined above. There exists constants C_1, C_2, C_3 with $C_i = C_i(D, \varepsilon, \delta)$ such that the following property holds:*

Let G be some graph generated in the above model. With probability $1 - o(1)$ with respect to G Algorithm 1 produces a partition which differs from the original partition V_1, V_2 only in $O(n/\overline{w}')$ vertices.

Note that the number of vertices not classified correctly is $O(n/\overline{w}') = O(n/\overline{w})$ and thus decreases linearly in $\overline{w}$.

We present the proof of Theorem 2 in the following two sections. The proof in section 3 is based on some notions and lemmas used throughout. These are presented in section 2.

2 Notation and basic facts

We use the following notation.

1. $\|\cdot\|$ denotes the l_2-norm of a vector or matrix.
2. The transpose of a matrix or vector M is written as M^t.
3. For $U \subseteq \mathbb{N}$ and a vector v we construct the vector $v_{|U}$ by setting the ith component of $v_{|U}$ to v_i if $i \in U$ and to 0 if $i \notin U$. If U is clear from the context, we write simply v^*. For a matrix M we obtain M^* by setting all entries $m_{ij} := 0$ if $i \notin U$ or $j \notin U$. For a set of vectors S we define $S^* = \{v^* : v \in S\}$.
4. We abbreviate $(1, \ldots, 1)^t$ by $\mathbf{1}$.
5. For a matrix $M = (m_{ij})$ we define

$$s_M(X, Y) = \sum_{\substack{x \in X \\ y \in Y}} m_{xy}.$$

The Courant-Fischer characterization of eigenvalues reads

Fact 3. *Let $A \in \mathbb{R}^{n \times n}$ be some symmetric matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$. Then*

$$\lambda_{j+1} = \min_{\substack{U \\ \dim U = j}} \max_{\substack{x \in U^\perp \\ \|x\| = 1}} x^t A x$$

$$\lambda_{n-j} = \max_{\substack{U \\ \dim U = j}} \min_{\substack{x \in U^\perp \\ \|x\| = 1}} x^t A x$$

where $U^\perp$ denotes the orthogonal complement to U.

The next two lemmas are slight generalizations of two lemmas from [3]. Lemma 1 is proved as Lemma 3.4 in that paper for $0 - 1$ random variables. Our generalization can be derived analogously.

Lemma 1. *Let $x_1, \ldots, x_n$ independent random variables each having exactly two possible values from the interval $[0, b]$ and the same expectation μ, such that for all i*

$$\mathbf{Pr}\,[x_i = 0] = 1 - p_i \quad and \quad \mathbf{Pr}\,[x_i \neq 0] = \mathbf{Pr}\,[x_i = \mu/p_i] = p_i.$$

Let $a_1, \ldots, a_n$ real numbers from $[-a, a]$ and $Z = \sum_{i=1}^n a_i \cdot x_i$. If for S, D and some constant $c > 0$

$$\sum_{i=1}^n a_i^2 \leq D \qquad and \qquad S \leq c \cdot \mathrm{e}^c \cdot D \cdot \mu/a$$

hold, then

$$\mathbf{Pr}\,[|Z - \mathbf{E}\,[Z]| \geq S] \leq 2\mathrm{e}^{\frac{-S^2}{2\mu \cdot \mathrm{e}^c \cdot D \cdot b}}.$$

Let R be some $n \times n$-matrix with random entries r_{ij} and let $V = \{1, \ldots, n\}$ be the set of indices. We assume either that all r_{ij} are independent or that the only dependence is due to symmetry. We assume that the collection of the r_{ij}'s otherwise has the same properties as the x_i's in Lemma 1, in particular $\mathbf{E}[r_{ij}] = \mu$. The subsequent Lemma 2 is as Lemma 3.6 in [3]. Its proof is analogous. A similar lemma occurs as Lemma 2.5 in [12].

Lemma 2. *With probability* $1 - o(1)$ *for any pair* (A, B) *of sets* $A, B \subseteq V$ *the following holds:*
If $m = \max\{|A|, |B|\} \leq n/2$ *then*

1. $s_R(A, B) = O(\mathbf{E}[s_R(A, B)])$ *or*
2. $s_R(A, B) \cdot \ln \frac{s_R(A,B)}{\mathbf{E}[s_R(A,B)]} = O(m \cdot \ln \frac{n}{m})$.

Let R be a random matrix as above and $B > 1$ be some constant. For symmetric R let $U \subseteq V$ be given by

$$u \in U \ \text{ if and only if } s_R(V, \{u\}) = s_R(\{u\}, V) \leq B \cdot \mu \cdot n.$$

For non-symmetric R we define

$$U = \{u \in V : \max(s_R(\{u\}, V), s_R(V, \{u\})) \leq B \cdot \mu \cdot n\}.$$

The following lemma is at the heart of our results. It is a transfer of Lemma 3.3 in [3] and Theorem 2.2 in [12]. In contrast to [3] and [12] we require that only the vector y is perpendicular to $\mathbf{1}$. The proof is similar to [3] and [12]. In particular recall item 3. of the notation as introduced above.

Lemma 3. *For R and U as above with probability* $1 - o(1)$ *we have for all unit vectors* $x, y \in (\mathbb{R}^n)^*$ *with* $y \perp \mathbf{1}$ *that* $|x^t R y| = O(\sqrt{\mu \cdot n})$.

3 The analysis of the algorithm

Let $G = (V, E)$, D, V_1, V_2 and $w_1, \ldots, w_n$ as in Subsection 1.1. Let d_i be the actual degree of i in G. For $W \subseteq V$ we define $\Phi(W) = \sum_{i \in W} w_i$ and abbreviate $\Phi_i := \Phi(V_i)/\Phi(V)$. Since all $w_i \geq \varepsilon \cdot \overline{w} = \Omega(\overline{w})$ and $|V_i| = \Omega(n)$ we have

$$\Phi_i = \frac{\Phi(V_i)}{\Phi(V)} \geq \frac{\sum_{j \in V_i} w'_j}{\overline{w} n} \geq \frac{\sum_{j \in V_i} \varepsilon \overline{w}}{\overline{w} n} \geq \frac{\sum_{j \in V_i} \varepsilon}{n} = \Omega(\varepsilon)$$

and each Φ_i are bounded away from 0 by some constant. For $i \in V_1$ we have

$$\mathbf{E}[d_i] = w'_i = \sum_{j \in V_1} d_{11} \cdot \frac{w_i \cdot w_j}{\overline{w} \cdot n} + \sum_{j \in V_2} d_{12} \cdot \frac{w_i \cdot w_j}{\overline{w} \cdot n} = w_i \cdot (d_{11}\Phi_1 + d_{12}\Phi_2)$$

and for $i \in V_2$ we get

$$w'_i = w_i \cdot (d_{12}\Phi_1 + d_{22}\Phi_2).$$

Since D is of full rank, we have no row containing only 0. So, each w_i' is $\neq 0$ and $w_i' = \Theta(w_i)$. The expected average degree $\overline{w}'$ in G is

$$\overline{w}' = \sum_{i=1}^{n} \frac{w_i'}{n} = \sum_{i \in V_1} \frac{w_i'}{n} + \sum_{i \in V_2} \frac{w_i'}{n}$$

$$= \frac{d_{11}\Phi_1\Phi(V_1) + 2 \cdot d_{12}\Phi_1\Phi(V_2) + d_{22}\Phi_2\Phi(V_2)}{n}$$

$$= \overline{w} \cdot (d_{11}\Phi_2^2 + 2 \cdot d_{12}\Phi_1\Phi_2 + d_{22}\Phi_2^2) = \Theta(\overline{w}).$$

Let A be the adjacency matrix of G. We construct R by multiplying each entry a_{ij} with $\overline{w}'^2/(w_i' \cdot w_j') = \Theta(\overline{w}^2/(w_i \cdot w_j)) = O(1/\varepsilon^2)$. So each entry in R is bounded by some constant.

We have for $i, j \in V_1$

$$\mathbf{E}\left[r_{ij}\right] = d_{11} \cdot \frac{w_i \cdot w_j}{\overline{w} \cdot n} \cdot \frac{\overline{w}'^2}{w_i' \cdot w_j'}, = d_{11} \cdot \frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot \frac{1}{(d_{11}\Phi_1 + d_{12}\Phi_2)^2},$$

for $i \in V_1$, $j \in V_2$ or the other way round

$$\mathbf{E}\left[r_{ij}\right] = d_{12} \cdot \frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot \frac{1}{(d_{11}\Phi_1 + d_{12}\Phi_2) \cdot (d_{12}\Phi_1 + d_{22}\Phi_2)},$$

and finally for $i, j \in V_2$

$$\mathbf{E}\left[r_{ij}\right] = d_{22} \cdot \frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot \frac{1}{(d_{12}\Phi_1 + d_{22}\Phi_2)^2}.$$

We obtain a symmetric 2×2-matrix $M = (m_{ij})$ of expectations such that $\mathbf{E}\left[r_{ij}\right] = m_{\psi(i),\psi(j)}$. With

$$X = \begin{pmatrix} (d_{11}\Phi_1 + d_{12}\Phi_2)^{-1} & 0 \\ 0 & (d_{12}\Phi_1 + d_{22}\Phi_2)^{-1} \end{pmatrix}$$

we get

$$M = \frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot X \cdot \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \cdot X = \frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot X \cdot D \cdot X$$

If $e = (e_1 \; e_2)^t$ is some eigenvector of D, then $(e_1/x_{11} \; e_2/x_{22})^t$ is an eigenvector of $X \cdot D \cdot X$ with the same eigenvalue. So, the eigenvalues of $X \cdot D \cdot X$ are determined only by D, and are $\neq 0$.

We divided the entries of e by the x_{ii}. This makes the entries larger, but at most by some constant factor independent of $\overline{w}'$ or n. So, the normalized eigenvectors of $X \cdot D \cdot X$ have entries, that are bounded away from 0 by some constant. We need this fact later.

We summarize, M has 2 eigenvalues, whose absolute value is

$$\Omega(\overline{w}'^2/(\overline{w}n)) = \Omega(\overline{w}'/n)$$

and all the entries of the normalized eigenvectors are $\Omega(1)$.

The expected row-sum $s_R(\{i\}, V)$ for some $i \in V_1$ is

$$\frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot \left(\frac{d_{11}|V_1|}{x_{11}^2} + \frac{d_{12}|V_2|}{x_{11} \cdot x_{22}} \right) = \Theta(\overline{w}') \tag{1}$$

and for $i \in V_2$

$$\frac{\overline{w}'^2}{\overline{w} \cdot n} \cdot \left(\frac{d_{12}|V_1|}{x_{11} \cdot x_{22}} + \frac{d_{22}|V_2|}{x_{22}^2} \right) = \Theta(\overline{w}'). \tag{2}$$

The number of rows with $s_R(\{i\}, V) \geq 5 \cdot \mathbf{E}\left[s_R(\{i\}, V)\right]$ is with high probability $\mathrm{e}^{-\Omega(\overline{w}')} \cdot n$. This can be shown as follows: Use Lemma 1 to calculate the probability that a fixed i is such a row. This probability is $\mathrm{e}^{-\Omega(\overline{w}')}$. So, we have an expected number of such rows bounded by $\mathrm{e}^{-\Omega(\overline{w}')} \cdot n$. Since the dependence between any two rows is small, we have a relatively small variance and Chebycheff's inequaltity gives the result.

If (1) and (2) differ by a factor of at least 25, we can simply detect large parts of V_1 and V_2 by partitioning the rows by the value of $s_R(\{i\}, V)$. This is the reason for s_1 in the algorithm. If (1) and (2) are closer, then both are relatively near to the average row-sum, which is $\Theta(\overline{w}')$. Now, let U be the set of all i, with $s_R(\{i\}, V) \leq C \cdot \overline{w}'$. The exact value of C depends on D, ε and the lower bound δ on $|V_i|/n$. A similar calculation as above shows, that $|U| \geq (1 - \mathrm{e}^{-\Omega(\overline{w}')}) \cdot n$.

Lemma 4. *With high probability for any set $X \subseteq V$ with $|X| = e^{-\Omega(\overline{w})} \cdot n$ we have $s_R(X, V) = \mathrm{e}^{-\Omega(\overline{w}')} \cdot n$.*

Proof. Let $X_i = X \cap V_i$. We have that

$$s_R(X, V) = \sum_{i,j=1}^{2} s_R(X_i, V_j).$$

If we can show, that with high probability for each summand the bound $e^{-\Omega(\overline{w}')} \cdot n$ holds, then the assertion follows. We give the proof for $s_R(X_1, V_1)$ explicitly. The remaining cases follow analogously.

Fix some set $X_1 \subseteq V_1$ with $|X_1| = \delta n = e^{-c_1 \cdot \overline{w}'} \cdot n$, where c_1 is some arbitrarily small constant. Then $\mathbf{E}\left[s_R(X_1, V_1)\right] = \Theta(m_{11} \cdot |X_1| \cdot |V_1|) = \Theta(\overline{w}' \cdot |X_1|) = \mathrm{e}^{-\Omega(\overline{w}')} \cdot n$.

Let $t = |X_1| \cdot |V_1|$. We use Lemma 1. For $\{u, v\} \subseteq X_1$ we set x_i in the lemma to r_{uv} with $u < v$ and a_i to 2, because such entries are counted twice in the sum. For the other terms in $s_R(X_1, V_1)$, namely r_{uv} with $u \in X_1$ and $v \notin X_1$ we let $x_i = r_{uv}$ and $a_i = 1$. This gives for the lemma, that $a = 2$, $D \leq 2t$ and $\mu = m_{11}$. We choose $S = c \cdot e^c \cdot m_{11} \cdot t = c \cdot e^c \cdot \Theta(\overline{w}' \cdot \delta n) = \mathrm{e}^{-\Omega(\overline{w}')} \cdot n$ for some constant c determined later. Then

$$\mathbf{Pr}\left[|s_R(X_1, V_1) - m_{11} \cdot t| \geq S\right] \leq 2 \cdot \mathrm{e}^{-\Omega(S^2/(m_{11} \cdot e^c \cdot t))}$$
$$= 2 \cdot e^{-\Omega(c^2 \cdot e^c \cdot m_{11} \cdot t)}$$
$$= 2 \cdot e^{-\Omega(c^2 \cdot e^c \cdot \overline{w}' \cdot \delta n)}$$

The number of sets X_1 possible is bounded by

$$\binom{|V_1|}{\delta n} \le \binom{n}{\delta n} \le \left(\frac{e}{\delta}\right)^{\delta n} = e^{\delta n - \delta \ln \delta \cdot n} = e^{\delta n + O(\delta n \cdot \overline{w}')}.$$

A union bound gives that the probability for the existence of a set X_1 contradicting the claim is

$$2 \cdot e^{\delta n + O(\delta n \cdot \overline{w}') - \Omega(c^2 \cdot e^c \cdot \overline{w}' \cdot \delta n)} = o(1),$$

if c is large enough (but still constant). For sets X_1 with cardinality $\le \delta n$ the same bounds for $s_R(X_1, V_1)$ and the probability hold, since we can fill them up until they contain exactly δn elements without decreasing $s_R(X_1, V_1)$. $\square$

By the above lemma we see that the sum of the entries we loose by building R^* is bounded $e^{-\Omega(\overline{w}')} \cdot n$. Thus, we have that $\|R - R^*\| \le e^{-\Omega(\overline{w}')} \cdot n$. And for all unit vectors f, g we have $\max_{f,g} |f^t (R - R^*) g| \le \|R - R^*\| = e^{-\Omega(\overline{w}')} \cdot n$.

Let $e = (e_1 \; e_2)^t$ be some normalized eigenvector of M and χ_1, χ_2 be the characteristic vectors of V_1, V_2 ($\chi_i(j) = 1$ if $j \in V_i$ and 0 otherwise) and $\alpha = |V_1|/n$, $\beta = |V_2|/n$. Let $g = e_1 \cdot \beta \cdot \chi_1 + e_2 \cdot \alpha \cdot \chi_2$. Then $\|g\| = \sqrt{e_1^2 \alpha \beta^2 n + e_2^2 \alpha^2 \beta n} = \Theta(\sqrt{n})$.

We have with probability $1 - o(1)$ that asymptotically

$$\begin{aligned}
g^t R g &= e_1^2 \cdot \beta^2 s_R(V_1, V_1) + 2e_1 e_2 \cdot \alpha \beta s_R(V_1, V_2) + e_2^2 s_R(V_2, V_2) \\
&= e_1^2 \cdot \alpha^2 \beta^2 \cdot n^2 \cdot m_{11} + 2e_1 e_2 \cdot \alpha^2 \beta^2 \cdot n^2 \cdot m_{12} + e_2^2 \cdot \alpha^2 \beta^2 \cdot n^2 \cdot m_{22} \\
&= \alpha^2 \beta^2 \cdot n^2 \cdot (e_1^2 \cdot m_{11} + 2e_1 e_2 \cdot m_{12} + e_2^2 m_{22}) \\
&= \alpha^2 \beta^2 \cdot n^2 \cdot (e_1 \; e_2) \cdot M \cdot \binom{e_1}{e_2}.
\end{aligned}$$

Since all eigenvalues of M are in absolute value $\Omega(\overline{w}'/n)$ we get

$$|g^t R^* g| \ge |g^t R g| - e^{-\Omega(\overline{w}')} \cdot n = \alpha^2 \beta^2 \cdot n^2 \cdot \Omega(\overline{w}'/n) - e^{-\Omega(\overline{w}')} \cdot \Theta(n) = \Omega(\overline{w}' \cdot n),$$

by using the triangle inequality. Thus, using the 2 eigenvectors of M, we can construct 2 orthogonal vectors g and h for R^* such that

$$\left| \frac{g^t}{\|g\|} \cdot R^* \cdot \frac{g}{\|g\|} \right| = \Omega(\overline{w}') \qquad \text{and} \qquad \left| \frac{h^t}{\|h\|} \cdot R^* \cdot \frac{h}{\|h\|} \right| = \Omega(\overline{w}').$$

By Fact 3 we obtain, that at least two eigenvalues of R^* are $\Omega(\overline{w}')$ in absolute value.

It is important that all the other eigenvalues of R^* are bounded by $O(\sqrt{\overline{w}'})$ in absolute value. Let u and v some unit-vectors with u perpendicular to g and h. Because both g and h are linear combinations of χ_1 and χ_2, u is also perpendicular to χ_1 and χ_2.

We partition u into u_1, u_2 as V is partitioned into V_1, V_2. By the same principle we construct v_i, $R_{i,j}$ and $R^*_{i,j}$. Then

$$\max_{\substack{u,v \\ u \perp g,h}} \left| v^t R^* u \right| = \max_{\substack{u,v \\ u \perp g,h}} \left| \sum_{i,j=1}^{2} v_i^t R^*_{i,j} u_j \right| \leq \max_{\substack{u,v \\ u \perp g,h}} \sum_{i,j=1}^{2} \left| v_i^t R^*_{i,j} u_j \right|. \qquad (3)$$

If u and v maximize the above terms, we can assume that $u = u^*$ and $v = v^*$.

Then the $u_j = u_j{}^*$ are perpendicular to $\mathbf{1}$. In addition we have $v_i^t \cdot R^*_{i,j} \cdot u_j = v_i^{*t} \cdot R_{i,j} \cdot u_j{}^*$. By the construction of R we have for all $R_{i,j}$ that the entries are bounded by some constant and the expectation of each entry is the same, namely $\Theta(d_{ij} \cdot \overline{w}'/n)$. Lemma 3 allows us to bound each term in the above sum by $O(\sqrt{\overline{w}'})$. Fact 3 can be used to bound the remaining eigenvalues of R^* by $O(\sqrt{\overline{w}'})$.

Finally we show that it is possible to obtain V_1 and V_2 by investigating the eigenvectors of R^*.

For this let v_1, v_2 be two orthonormal eigenvectors of R^* with eigenvalue $\Omega(\overline{w}')$ (in absolute value). Then v_i can be written as $v_i = c_i \cdot m_i + d_i \cdot u_i$ with $\|m_i\| = \|u_i\| = 1$ and $c_i^2 + d_i^2 = 1$. m_i comes from the space spanned by g and h, and u_i comes from the orthogonal complement. Then by the bound for (3)

$$|v^t R^* u_i| = \Omega(\overline{w}') \cdot |v_i^t \cdot u_i| = \Omega(\overline{w}') \cdot |d_i| = O(\sqrt{\overline{w}'}),$$

and $|d_i|$ must be $O(1/\sqrt{\overline{w}'})$. As $|c_i| + |d_i| \geq c_i^2 + d_i^2 = 1$, we have $|c_i| = 1 - O(1/\sqrt{\overline{w}'})$.

Since

$$0 = v_1^t v_2 = c_1 c_2 m_1^t m_2 + c_1 d_2 m_1^t u_2 + c_2 d_1 u_1^t m_2 + d_1 d_2 u_1^t u_2$$

we have

$$\begin{aligned}
|c_1 c_2 m_1^t m_2| &= |c_1 d_2 m_1^t u_2 + c_2 d_1 u_1^t m_2 + d_1 d_2 u_1^t u_2| \\
&\leq |c_1 d_2 m_1^t u_2| + |c_2 d_1 u_1^t m_2| + |d_1 d_2 u_1^t u_2| \\
&= |d_1 d_2 u_1^t u_2| \leq |d_1 d_2| \\
&= O(1/\overline{w}').
\end{aligned}$$

Together with $c_i = 1 - O(1/\sqrt{\overline{w}'})$ we can follow that m_1 and m_2 must be almost perpendicular. We write $m_i = \gamma_i \cdot \chi_1/\sqrt{n} + \delta_i \cdot \chi_2/\sqrt{n}$. For at least one i we have $|\gamma_i - \delta_i| > \varepsilon$ for some small constant ε, otherwise m_1 and m_2 could not be almost perpendicular. Taking this m_i, we have that the entries belonging to V_1 differ from the other entries by at least $\varepsilon/\sqrt{n}$. This gives us the chance to identify the V_1, V_2 by the entries of m_i.

Unfortunaly, we have only v_i and not m_i. But we can assume, that in $c_i \cdot m_i$ the distance of $\varepsilon/(2\sqrt{n})$ still holds, because $c_i \geq (1 - O(1/\overline{w}')) > 1/2$. It is possible, that some entries j in u change the value of $c_i \cdot m_i(j)$, such that we put j into the wrong partition. This may happen, if the value is changed by at least $\varepsilon/(4\sqrt{n})$. But such entries are relatively rare. The entry in u_i must have an absolute value of $\Omega(\sqrt{\overline{w}'}) \cdot \varepsilon/(4\sqrt{n})$, because $|d_i| = O(1/\sqrt{\overline{w}'})$ is small. The number of such entries is bounded by $O(n/\overline{w}')$ since u has length 1. We obtain, that we are able to partition at least $(1 - O(1/\overline{w}')) \cdot n$ vertices correctly by visiting the eigenvector v_i of R^*. This finishes our proof of Theorem 2.

References

1. Aiello, W, Chung, F., Lu, L.: A random graph model for massive graphs. Proc. 33rd. SToC (2001), 171–180.
2. Alon, N. Spectral techniques in graph algorithms. Proc. LATIN (1998), LNCS 1380, Springer, 206–215.
3. Alon, N., Kahale, N.: A spectral technique for coloring random 3-colorable graphs. SIAM J. Comput. **26** (1997) 1733–1748.
4. Boppana, R.B.: Eigenvalues and graph bisection: An average case analysis. Proc. 28th FoCS (1987), 280–285.
5. Chung, F.K.R.: Spectral Graph Theory. American Mathematical Society (1997).
6. Coja-Oghlan, A.: On the Laplacian eigenvalues of $G_{n,p}$. Preprint (2005) http://www.informatik.hu-berlin.de/∼coja/de/publikation.php.
7. Coja-Oghlan, A., Lanka, A.: The Spectral Gap of Random Graphs with Given Expected Degrees. Preprint (2006).
8. Chung, F.K.R., Lu, L., Vu, V.: The Spectra of Random Graphs with Given Expected Degrees. Internet Mathematics **1** (2003) 257–275.
9. Dasgupta, A., Hopcroft, J.E., McSherry, F.: Spectral Analysis of Random Graphs with Skewed Degree Distributions. Proc. 45th FOCS (2004) 602–610.
10. Feige, U., Ofek, E.: Spectral Techniques Applied to Sparse Random Graphs. Random Structures and Algorithms, **27(2)** (2005), 251–275.
11. Flaxman, A.: A spectral technique for random satisfiable 3CNF formulas. Proc. 14th SODA (2003) 357–363.
12. Friedman, J., Kahn, J., Szemerédi, E.: On the Second Eigenvalue in Random Regular Graphs. Proc. 21th STOC (1989) 587–598.
13. Füredi, Z., Komloś, J.: The eigenvalues of random symmetric matrices. Combinatorica **1** (1981) 233–241.
14. Giesen, J., Mitsche, D.: Reconstructing Many Partitions Using Spectral Techniques. Proc. 15th FCT (2005) 433–444.
15. Husbands, P., Simon, H., and Ding, C.: On the use of the singular value decomposition for text retrieval. In 1st SIAM Computational Information Retrieval Workshop (2000), Raleigh, NC.
16. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. Combinatorics, Probability and Computing **12** (2003) 61–72.
17. Krivelevich, M., Vu, V.H.: On the concentration of eigenvalues of random symmetric matrices. Microsoft Technical Report 60 (2000).
18. Lempel, R., Moran, S. Rank-stability and rank-similarity of link-based web ranking algorithms in authority-connected graphs. Information retrieval, special issue on Advances in Mathematics/Formal methods in Information Retrieval (2004) Kluwer.
19. Meila, M., Varna D.: A comparison of spectral clustering algorithms. UW CSE Technical report 03-05-01.
20. McSherry, F.: Spectral Partitioning of Random Graphs. Proc. 42nd FoCS (2001) 529–537.
21. Mihail, M., Papadimitriou, C.H.: On the Eigenvalue Power Law. Proc. 6th RANDOM (2002) 254–262.
22. Pothen, A., Simon, H.D., Liou, K.-P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. **11** (1990) 430–452
23. Spielman, D.A., Teng, S.-H.: Spectral partitioning works: planar graphs and finite element meshes. Proc. 36th FOCS (1996) 96–105.