MAX CUT in Weighted Random Intersection Graphs and Discrepancy of Sparse Random Set Systems

Sotiris Nikoletseas ORCID
Computer Engineering & Informatics Department, University of Patras, Greece
Computer Technology Institute, Greece

Christoforos Raptopoulos[1] ORCID
Computer Engineering & Informatics Department, University of Patras, Greece

Paul Spirakis ORCID
Department of Computer Science, University of Liverpool, UK
Computer Engineering & Informatics Department, University of Patras, Greece
Computer Technology Institute, Greece

Abstract

Let V be a set of n vertices, $\mathcal{M}$ a set of m labels, and let $\mathbf{R}$ be an $m \times n$ matrix of independent Bernoulli random variables with probability of success p; columns of $\mathbf{R}$ are incidence vectors of label sets assigned to vertices. A random instance $G(V, E, \mathbf{R}^T\mathbf{R})$ of the weighted random intersection graph model is constructed by drawing an edge with weight equal to the number of common labels (namely $[\mathbf{R}^T\mathbf{R}]_{v,u}$) between any two vertices u, v for which this weight is strictly larger than 0. In this paper we study the average case analysis of Weighted Max Cut, assuming the input is a weighted random intersection graph, i.e. given $G(V, E, \mathbf{R}^T\mathbf{R})$ we wish to find a partition of V into two sets so that the total weight of the edges having exactly one endpoint in each set is maximized.

In particular, we initially prove that the weight of a maximum cut of $G(V, E, \mathbf{R}^T\mathbf{R})$ is concentrated around its expected value, and then show that, when the number of labels is much smaller than the number of vertices (in particular, $m = n^\alpha, \alpha < 1$), a random partition of the vertices achieves asymptotically optimal cut weight with high probability. Furthermore, in the case $n = m$ and constant average degree (i.e. $p = \frac{\Theta(1)}{n}$), we show that with high probability, a majority type randomized algorithm outputs a cut with weight that is larger than the weight of a random cut by a multiplicative constant strictly larger than 1. Then, we formally prove a connection between the computational problem of finding a (weighted) maximum cut in $G(V, E, \mathbf{R}^T\mathbf{R})$ and the problem of finding a 2-coloring that achieves minimum discrepancy for a set system Σ with incidence matrix $\mathbf{R}$ (i.e. minimum imbalance over all sets in Σ). We exploit this connection by proposing a (weak) bipartization algorithm for the case $m = n, p = \frac{\Theta(1)}{n}$ that, when it terminates, its output can be used to find a 2-coloring with minimum discrepancy in a set system with incidence matrix $\mathbf{R}$. In fact, with high probability, the latter 2-coloring corresponds to a bipartition with maximum cut-weight in $G(V, E, \mathbf{R}^T\mathbf{R})$. Finally, we prove that our (weak) bipartization algorithm terminates in polynomial time, with high probability, at least when $p = \frac{c}{n}, c < 1$.

2012 ACM Subject Classification Mathematics of computing → Random graphs

Keywords and phrases Random Intersection Graphs, Maximum Cut, Discrepancy

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

Given an undirected graph $G(V, E)$, the Max Cut problem asks for a partition of the vertices of G into two sets, such that the number of edges with exactly one endpoint in each set of the

[1] Corresponding author

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

partition is maximized. This problem can be naturally generalized for weighted (undirected) graphs. A weighted graph is denoted by $G(V, E, \mathbf{W})$, where V is the set of vertices, E is the set of edges and $\mathbf{W}$ is a weight matrix, which specifies a weight $\mathbf{W}_{i,j} = w_{i,j}$, for each pair of vertices i, j. In particular, we assume that $\mathbf{W}_{i,j} = 0$, for each edge $\{i, j\} \notin E$.

▶ **Definition 1** (Weighted Max Cut). *Given a weighted graph $G(V, E, \mathbf{W})$, find a partition of V into two (disjoint) subsets A, B, so as to maximize the cumulative weight of the edges of G having one endpoint in A and the other in B.*

Weighted Max Cut is fundamental in theoretical computer science and is relevant in various graph layout and embedding problems [10]. Furthermore, it also has many practical applications, including infrastructure cost and circuit layout optimization in network and VLSI design [19], minimizing the Hamiltonian of a spin glass model in statistical physics [3], and data clustering [18]. In the worst case Max Cut (and also Weighted Max Cut) is APX-hard, meaning that there is no polynomial-time approximation scheme that finds a solution that is arbitrarily close to the optimum, unless P = NP [17].

The average case analysis of Max Cut, namely the case where the input graph is chosen at random from a probabilistic space of graphs, is also of considerable interest and is further motivated by the desire to justify and understand why various graph partitioning heuristics work well in practical applications. In most research works the input graphs are drawn from the Erdős-Rényi random graphs model $\mathcal{G}_{n,m}$, i.e. random instances are drawn equiprobably from the set of simple undirected graphs on n vertices and m edges, where m is a linear function of n (see also [13, 7] for the average case analysis of Max Cut and its generalizations with respect to other random graph models). One of the earliest results in this area is that Max Cut undergoes a phase transition on $\mathcal{G}_{n,\gamma n}$ at $\gamma = \frac{1}{2}$ [8], in that the difference between the number of edges of the graph and the Max-Cut size is $O(1)$, for $\gamma < \frac{1}{2}$, while it is $\Omega(n)$, when $\gamma > \frac{1}{2}$. For large values of γ, it was proved in [4] that the maximum cut size of $G_{n,\gamma n}$ normalized by the number of vertices n reaches an absolute limit in probability as $n \to \infty$, but it was not until recently that the latter limit was established and expressed analytically in [9], using the interpolation method; in particular, it was shown to be asymptotically equal to $(\frac{\gamma}{2} + P_* \sqrt{\frac{\gamma}{2}})n$, where $P_* \approx 0.7632$. We note however that these results are existential, and thus do not lead to an efficient approximation scheme for finding a tight approximation of the maximum cut with large enough probability when the input graph is drawn from $\mathcal{G}_{n,\gamma n}$. An efficient approximation scheme in this case was designed in [8], and it was proved that, with high probability, this scheme constructs a cut with at least $\left(\frac{\gamma}{2} + 0.37613\sqrt{\gamma}\right) n = (1 + 0.75226 \frac{1}{\sqrt{\gamma}}) \frac{\gamma}{2} n$ edges, noting that $\frac{\gamma}{2} n$ is the size of a random cut (in which each vertex is placed independently and equiprobably in one of the two sets of the partition). Whether there exists an efficient approximation scheme that can close the gap between the approximation guarantee of [8] and the limit of [9] remains an open problem.

In this paper, we study the average case analysis of Weighted Max Cut when input graphs are drawn from the generalization of another well-established model of random graphs, namely the *weighted random intersection graphs model* (the unweighted version of the model was initially defined in [15]). In this model, edges are formed through the intersection of label sets assigned to each vertex and edge weights are equal to the number of common labels between edgepoints.

▶ **Definition 2** (Weighted random intersection graph). *Consider a universe $\mathcal{M} = \{1, 2, \dots, m\}$ of labels and a set of n vertices V. We define the $m \times n$ representation matrix $\mathbf{R}$ whose entries are independent Bernoulli random variables with probability of success p. For $\ell \in \mathcal{M}$ and $v \in V$, we say that vertex v has chosen label ℓ iff $\mathbf{R}_{\ell,v} = 1$. Furthermore, we draw* *an edge with weight* $[\mathbf{R}^T\mathbf{R}]_{v,u}$ *between any two vertices* u, v *for which this weight is strictly larger than 0.The weighted graph* $G = (V, E, \mathbf{R}^T\mathbf{R})$ *is then a random instance of the weighted random intersection graphs model* $\overline{\mathcal{G}}_{n,m,p}$.

Random intersection graphs are relevant to and capture quite nicely social networking; vertices are the individual actors and labels correspond to specific types of interdependency. Other applications include oblivious resource sharing in a (general) distributed setting, efficient and secure communication in sensor networks [20], interactions of mobile agents traversing the web etc. (see e.g. the survey papers [6, 16] for further motivation and recent research related to random intersection graphs). In all these settings, weighted random intersection graphs, in particular, also capture the strength of connections between actors (e.g. in a social network, individuals having several characteristics in common have more intimate relationships than those that share only a few common characteristics). One of the most celebrated results in this area is equivalence (measured in terms of total variation distance) of random intersection graphs and Erdős-Rényi random graphs when the number of labels satisfies $m = n^{\alpha}, \alpha > 6$ [12]. This bound on the number of labels was improved in [22], by showing equivalence of sharp threshold functions among the two models for $\alpha \geq 3$. Similarity of the two models has been proved even for smaller values of α (e.g. for any $\alpha > 1$) in the form of various translation results (see e.g. Theorem 1 in [21]), suggesting that some algorithmic ideas developed for Erdős-Rényi random graphs also work for random intersection graphs (and also weighted random intersection graphs).

In view of this, in the present paper we study the average case analysis of Weighted Max Cut under the weighted random intersection graphs model, for the range $m = n^{\alpha}, \alpha \leq 1$ for two main reasons: First, the average case analysis of Max Cut has not been considered in the literature so far when the input is a drawn from the random intersection graphs model, and thus the asymptotic behaviour of the maximum cut remains unknown especially for the range of values where random intersection graphs and Erdős-Rényi random graphs differ the most. Furthermore, studying a model where we can implicitly control its intersection number (indeed m is an obvious upper bound on the number of cliques that can cover all edges of the graph) may help understand algorithmic bottlenecks for finding maximum cuts in Erdős-Rényi random graphs.

Second, we note that the representation matrix $\mathbf{R}$ of a weighted random intersection graph can be used to define a random set system Σ consisting of m sets $\Sigma = \{L_1, \dots, L_m\}$, where L_ℓ is the set of vertices that have chosen label ℓ; we say that $\mathbf{R}$ is the *incidence matrix* of Σ. Therefore, there is a natural connection between Weighted Max Cut and the discrepancy of such random set systems, which we formalize in this paper. In particular, given a set system Σ with incidence matrix $\mathbf{R}$, its *discrepancy* is defined as $\text{disc}(\Sigma) = \min_{\mathbf{x} \in \{\pm 1\}^n} \max_{L \in \Sigma} \left|\sum_{v \in L} x_v\right| = \|\mathbf{R}\mathbf{x}\|_\infty$, i.e. it is the minimum imbalance of all sets in Σ over all 2-colorings $\mathbf{x}$. Recent work on the discrepancy of random rectangular matrices defined as above [1] has shown that, when the number of labels (sets) m satisfies $n \geq 0.73 m \log m$, the discrepancy of Σ is at most 1 with high probability. The proof of the main result in [1] is based on a conditional second moment method combined with Stein's method of exchangeable pairs, and improves upon a Fourier analytic result of [14], and also upon previous results in [11], [20]. The design of an efficient algorithm that can find a 2-coloring having discrepancy $O(1)$ in this range still remains an open problem. Approximation algorithms for a similar model for random set systems were designed and analyzed in [2]; however, the algorithmic ideas there do not apply in our case.

1.1 Our Contribution

In this paper, we introduce the model of weighted random intersection graphs and we study the average case analysis of Weighted Max Cut through the prism of Discrepancy of random set systems. We formalize the connection between these two combinatorial problems for the case of arbitrary weighted intersection graphs in Corollary 4. We prove that, given a weighted intersection graph $G = (V, E, \mathbf{R}^T\mathbf{R})$ with representation matrix $\mathbf{R}$, and a set system with incidence matrix $\mathbf{R}$, such that $\text{disc}(\Sigma) \leq 1$, a 2-coloring has maximum cut weight in G if and only if it achieves minimum discrepancy in Σ. In particular, Corollary 4 applies in the range of values considered in [1] (i.e. $n \geq 0.73m \log m$), and thus any algorithm that finds a maximum cut in $G(V, E, \mathbf{R}^T\mathbf{R})$ with large enough probability can also be used to find a 2-coloring with minimum discrepancy in a set system Σ with incidence matrix $\mathbf{R}$, with the same probability of success.

We then consider weighted random intersection graphs in the case $m = n^\alpha, \alpha \leq 1$, and we prove that the maximum cut weight of a random instance $G(V, E, \mathbf{R}^T\mathbf{R})$ of $\overline{\mathcal{G}}_{n,m,p}$ concentrates around its expected value (see Theorem 5). In particular, with high probability over the choices of $\mathbf{R}$, $\texttt{Max-Cut}(G) \sim \mathbb{E}_{\mathbf{R}}[\texttt{Max-Cut}(G)]$, where $\mathbb{E}_{\mathbf{R}}$ denotes expectation with respect to $\mathbf{R}$. The proof is based on the Efron-Stein inequality for upper bounding the variance of the maximum cut. As a consequence of our concentration result, we prove in Theorem 6 that, in the case $\alpha < 1$, a random 2-coloring (i.e. biparition) $\mathbf{x}^{(rand)}$ in which each vertex chooses its color independently and equiprobably, has cut weight asymptotically equal to $\texttt{Max-Cut}(G)$, with high probability over the choices of $\mathbf{x}^{(rand)}$ and $\mathbf{R}$.

The latter result on random cuts allows us to focus the analysis of our randomized algorithms of Section 4 on the case $m = n$ (i.e. $\alpha = 1$), and $p = \frac{c}{n}$, for some constant c (see also the discussion at the end of subsection 3.1), where the assumptions of Theorem 6 do not hold. It is worth noting that, in this range of values, the expected weight of a fixed edge in a weighted random intersection graph is equal to $mp^2 = \Theta(1/n)$, and thus we hope that our work here will serve as an intermediate step towards understanding when algorithmic bottlenecks for Max Cut appear in sparse random graphs (especially Erdős-Rényi random graphs) with respect to the intersection number. In particular, we analyze a Majority Cut Algorithm 1 that extends the algorithmic idea of [8] to weighted intersection graphs as follows: vertices are colored sequentially (each color $+1$ or -1 corresponding to a different set in the partition of the vertices), and the t-th vertex is colored opposite to the sign of $\sum_{i \in [t-1]} [\mathbf{R}^T\mathbf{R}]_{i,t} x_i$, namely the total available weight of its incident edges, taking into account colors of adjacent vertices. Our average case analysis of the Majority Cut Algorithm shows that, when $m = n$ and $p = \frac{c}{n}$, for large constant c, with high probability over the choices of $\mathbf{R}$, the expected weight of the constructed cut is at least $1 + \beta$ times larger than the expected weight of a random cut, for some constant $\beta = \beta(c) \geq \sqrt{\frac{16}{27\pi c^3}} - o(1)$. The fact that the lower bound on beta is inversely proportional to $c^{3/2}$ was to be expected, because, as p increases, the approximation of the maximum cut that we get from the weight of a random cut improves (see also the discussion at the end of subsection 3.1).

In subsection 4.2 we propose a framework for finding maximum cuts in weighted random intersection graphs for $m = n$ and $p = \frac{c}{n}$, for constant c, by exploiting the connection between Weighted Max Cut and the problem of discrepancy minimization in random set systems. In particular, we design a Weak Bipartization Algorithm 2, that takes as input an intersection graph with representation matrix $\mathbf{R}$ and outputs a subgraph that is "almost" bipartite. In fact, the input intersection graph is treated as a multigraph composed by overlapping cliques formed by the label sets $L_\ell = \{v : \mathbf{R}_{\ell,v} = 1\}, \ell \in \mathcal{M}$. The algorithm attempts to destroy all odd cycles of the input (except from odd cycles that are formed by labels with only two vertices) by replacing each clique induced by some label set L_ℓ by a random maximal matching. In Theorem 11 we prove that, with high probability over the choices of $\mathbf{R}$, if the Weak Bipartization Algorithm terminates, then its output can be used to construct a 2-coloring that has minimum discrepancy in a set system with incidence matrix $\mathbf{R}$, which also gives a maximum cut in $G(V, E, \mathbf{R}^T\mathbf{R})$. It is worth noting that this does not follow from Corollary 4, because a random set system with incidence matrix $\mathbf{R}$ has discrepancy larger than 1 with (at least) constant probability when $m = n$ and $p = \frac{c}{n}$. Our proof relies on a structural property of closed 0-strong vertex-label sequences (loosely defined as closed walks of edges formed by distinct labels) in the weighted random intersection graph $G(V, E, \mathbf{R}^T\mathbf{R})$ (Lemma 8). Finally, in Theorem 12, we prove that our Weak Bipartization Algorithm terminates in polynomial time, with high probability, if the constant c is strictly less than 1. Therefore, there is a polynomial time algorithm for finding weighted maximum cuts, with high probability, when the input is drawn from $\overline{\mathcal{G}}_{n,n,\frac{c}{n}}$, with $c < 1$. We believe that this part of our work may also be of interest regarding the design of efficient algorithms for finding minimum disrepancy colorings in random set systems.

Due to lack of space, some of the proofs are given in a clearly marked Appendix, to be read at the discretion of the program committee.

2 Notation and preliminary results

We denote weighted undirected graphs by $G(V, E, \mathbf{W})$; in particular, $V = V(G)$ (resp. $E = E(G)$) is the set of vertices (resp. set of edges) and $\mathbf{W} = \mathbf{W}(G)$ is the weight matrix, i.e. $\mathbf{W}_{i,j} = w_{i,j}$ is the weight of (undirected) edge $\{i, j\} \in E$. We allow $\mathbf{W}$ to have non-zero diagonal entries, as these do not affect cut weights. We also denote the number of vertices by n, and we use the notation $[n] = \{1, 2, \ldots, n\}$. We also use this notation to define parts of matrices, for example $\mathbf{W}_{[n],1}$ denotes the first column of the weight matrix.

A bipartition of the sets of vertices is a partition of V into two sets A, B such that $A \cap B = \emptyset$ and $A \cup B = V$. Bipartitions correspond to 2-colorings, which we denote by vectors $\mathbf{x}$ such that $x_i = +1$ if $i \in A$ and $x_i = -1$ if $i \in B$.

Given a weighted graph $G(V, E, \mathbf{W})$, we denote by $\texttt{Cut}(G, \mathbf{x})$ the weight of a cut defined by a bipartition $\mathbf{x}$, namely $\texttt{Cut}(G, \mathbf{x}) = \sum_{\{i,j\} \in E: i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{\{i,j\} \in E} w_{i,j}(x_i - x_j)^2$. The maximum cut of G is $\texttt{Max-Cut}(G) = \max_{\mathbf{x} \in \{-1,+1\}^n} \texttt{Cut}(G, \mathbf{x})$.

For a weighted random intersection graph $G(V, E, \mathbf{R}^T\mathbf{R})$ with representation matrix $\mathbf{R}$, we denote by S_v the set of labels chosen by vertex $v \in V$, i.e. $S_v = \{\ell : \mathbf{R}_{\ell,v} = 1\}$. Furthermore, we denote by L_ℓ the set of vertices having chosen label ℓ, i.e. $L_\ell = \{v : \mathbf{R}_{\ell,v} = 1\}$. Using this notation, the weight of an edge $\{v, u\} \in E$ is $|S_v \cup S_u|$; notice also that this is equal to 0 when $\{v, u\} \notin E$. We also note here that we may also think of a weighted random intersection graph as a simple weighted graph where, for any pair of vertices v, u, there are $|S_v \cap S_u|$ simple edges between them.

A set system Σ defined on a set V is a family of sets $\Sigma = \{L_1, L_2, \ldots, L_m\}$, where $L_\ell \subseteq V, \ell \in [m]$. The incidence matrix of Σ is an $m \times n$ matrix $\mathbf{R} = \mathbf{R}(\Sigma)$, where for any $\ell \in [m], v \in [n]$, $\mathbf{R}_{\ell,v} = 1$ if $v \in S_\ell$ and 0 otherwise. The discrerpancy of Σ with respect to a 2-coloring $\mathbf{x}$ of the vertices in V is $\text{disc}(\Sigma, \mathbf{x}) = \max_{\ell \in [m]} \left|\sum_{v \in V} \mathbf{R}_{\ell,v} x_v\right| = \|\mathbf{R}\mathbf{x}\|_\infty$. The discrepancy of Σ is $\text{disc}(\Sigma) = \min_{\mathbf{x} \in \{-1,+1\}^n} \text{disc}(\Sigma, \mathbf{x})$.

It is well-known that the cut size of a bipartition of the set of vertices of a graph $G(V, E)$ into sets A and B is given by $\frac{1}{4} \sum_{\{i,j\} \in E} (x_i - x_j)^2$, where $x_i = +1$ if $i \in A$ and $x_i = -1$ if $i \in B$. This can be naturally generalized for multigraphs and also for weighted graphs. In particular, the $\texttt{Max-Cut}$ size of a weighted graph $G(V, E, \mathbf{W})$ is given by

$$\texttt{Max-Cut}(G) = \max_{\mathbf{x} \in \{-1,+1\}^n} \frac{1}{4} \sum_{\{i,j\} \in E} \mathbf{W}_{i,j} (x_i - x_j)^2. \tag{1}$$

In particular, we get the following Corollary (refer to Section A of the Appendix for the proof):

▶ **Corollary 3.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a weighted intersection graph with representation matrix $\mathbf{R}$. Then, for any $\mathbf{x} \in \{-1, +1\}^n$,*

$$\textit{Cut}(G, \mathbf{x}) = \frac{1}{4} \left(\sum_{i,j \in [n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \|\mathbf{R}\mathbf{x}\|^2 \right) \tag{2}$$

and so

$$\textit{Max-Cut}(G) = \frac{1}{4} \left(\sum_{i,j \in [n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \min_{\mathbf{x} \in \{-1,+1\}^n} \|\mathbf{R}\mathbf{x}\|^2 \right), \tag{3}$$

where $\|\cdot\|$ denotes the 2-norm. In particular, the expectation of the size of a random cut, where each entry of $\mathbf{x}$ is independently and equiprobably either +1 or -1 is equal to $\mathbb{E}_{\mathbf{x}} \left[\textit{Cut}(G, \mathbf{x})\right] = \frac{1}{4} \sum_{i \neq j, i,j \in [n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j}$, where $\mathbb{E}_{\mathbf{x}}$ denotes expectation with respect to $\mathbf{x}$.

Since $\sum_{i,j \in [n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j}$ is fixed for any given representation matrix $\mathbf{R}$, the above Corollary implies that, to find a bipartition of the vertex set V that corresponds to a maximum cut, we need to find an n-dimensional vector in $\arg\min_{\mathbf{x} \in \{-1,+1\}^n} \|\mathbf{R}\mathbf{x}\|^2$. We thus get the following (refer to Section B of the Appendix for the proof):

▶ **Corollary 4.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a weighted intersection graph with representation matrix $\mathbf{R}$ and Σ a set system with incidence matrix $\mathbf{R}$. If $disc(\Sigma) \leq 1$, then $\mathbf{x}^* \in \arg\min_{\mathbf{x} \in \{-1,+1\}^n} \|\mathbf{R}\mathbf{x}\|^2$ if and only if $\mathbf{x}^* \in \arg\min_{\mathbf{x} \in \{-1,+1\}^n} disc(\Sigma, \mathbf{x})$. In particular, if the minimum discrepancy of Σ is at most 1, a bipartition corresponds to a maximum cut iff it achieves minimum discrepancy.*

Notice that above result is not necessarily true when $\text{disc}(\Sigma) > 1$, since the minimum of $\|\mathbf{R}\mathbf{x}\|$ could be achieved by 2-colorings with larger discrepancy than the optimal.

2.1 Range of values for p

Concerning the success probability p, we note that, when $p = o\left(\sqrt{\frac{1}{nm}}\right)$, direct application of the results of [5] suggest that $G(V, E, \mathbf{R}^T\mathbf{R})$ is chordal with high probability, but in fact the same proofs reveal that a stronger property holds, namely that there is no closed vertex-label sequence (refer to the precise definition in subsection 4.2) having distinct labels. Therefore, in this case, finding a bipartition with maximum cut weight is straightforward: indeed, one way to construct a maximum cut is to run our Weak Bipartization Algorithm 2 from subsection 4.2, and then to apply Theorem 11 (noting that Weak Bipartization termination condition trivially holds, since the set $\mathcal{C}_{odd}(G^{(b)})$ defined in subsection 4.2 is empty). Furthermore, even though we consider weighted graphs, we will also assume that $mp^2 = O(1)$, noting that, otherwise, $G(V, E, \mathbf{R}^T\mathbf{R})$ will be almost complete with high probability (indeed, the unconditional edge existence probability is $1-(1-p^2)^m$, which tends to 1 for $mp^2 = \omega(1)$). In particular, we will assume that $C_1\sqrt{\frac{1}{nm}} \leq p \leq C_2\frac{1}{\sqrt{m}}$, for arbitrary positive constants C_1, C_2; C_1 can be as small as possible, and C_2 can be as large as possible, provided $C_2\frac{1}{\sqrt{m}} \leq 1$. We note that, when p is asymptotically equal to the upper bound $C_2\frac{1}{\sqrt{m}}$, there is no constant weight upper bound that holds with high probability, whereas, when p is asymptotically equal to the lower bound $C_1\sqrt{\frac{1}{nm}}$, all weights in the graph are bounded by a small constant with high probability. Our results in Section 3 assume this range of values for p, and thus graph instances may contain edges with large (but constant) weights. On the other hand, in the analysis of our randomized algorithms in section 4, we assume $n = m$ and $p = \Theta\left(\frac{1}{n}\right)$; this range of values gives sparse graph instances (even though the distribution is different from sparse Erdős-Rényi random graphs).

3 Concentration of Max-Cut

In this section we prove that the size of the maximum cut in a weighted random intersection graph concentrates around its expected value. We note however, that the following Theorem does not provide an explicit formula for the expected value of the maximum cut.

▶ **Theorem 5.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model with $m = n^a, \alpha \leq 1$, and $C_1\sqrt{\frac{1}{nm}} \leq p \leq C_2\frac{1}{\sqrt{m}}$, for arbitrary positive constants C_1, C_2, and let $\mathbf{R}$ be its representation matrix. Then $\texttt{Max-Cut}(G) \sim \mathbb{E}_{\mathbf{R}}[\texttt{Max-Cut}(G)]$ with high probability, where $\mathbb{E}_{\mathbf{R}}$ denotes expectation with respect to $\mathbf{R}$, i.e. $\texttt{Max-Cut}(G)$ concentrates around its expected value.*

Proof. Let $G = G(V, E, \mathbf{R}^T\mathbf{R})$ be a weighted random intersection graph, and let $\mathbf{D}$ denote the (random) diagonal matrix containing all diagonal elements of $\mathbf{R}^T\mathbf{R}$. In particular, equation (3) of Corollary 3 can be written as

$$\texttt{Max-Cut}(G) = \frac{1}{4}\left(\sum_{i\neq j, i,j\in[n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \min_{\mathbf{x}\in\{-1,+1\}^n} \mathbf{x}^T\left(\mathbf{R}^T\mathbf{R} - \mathbf{D}\right)\mathbf{x}\right).$$

Furthermore, for any given $\mathbf{R}$, notice that, if we select each element of $\mathbf{x}$ independently and equiprobably from $\{-1, +1\}$, then $\mathbb{E}_{\mathbf{x}}[\mathbf{x}^T\left(\mathbf{R}^T\mathbf{R} - \mathbf{D}\right)\mathbf{x}] = 0$, where $\mathbb{E}_{\mathbf{x}}$ denotes expectation with respect to $\mathbf{x}$. By the probabilistic method, we thus have $\min_{\mathbf{x}\in\{-1,+1\}^n} \mathbf{x}^T\left(\mathbf{R}^T\mathbf{R} - \mathbf{D}\right)\mathbf{x} \leq 0$, implying the following bound:

$$\frac{1}{4}\sum_{i\neq j, i,j\in[n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} \leq \texttt{Max-Cut}(G_{n,m,p}) \leq \frac{1}{2}\sum_{i\neq j, i,j\in[n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j}, \tag{4}$$

where the second inequality follows trivially by observing that $\frac{1}{2}\sum_{i\neq j, i,j\in[n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j}$ equals the sum of the weights of all edges.

By linearity of expectation, we have $\mathbb{E}_{\mathbf{R}}\left[\sum_{i\neq j, i,j\in[n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j}\right] = \mathbb{E}_{\mathbf{R}}\left[\sum_{i\neq j, i,j\in[n]}\sum_{\ell\in[m]} \mathbf{R}_{\ell,i}\mathbf{R}_{\ell,j}\right] = n(n-1)mp^2 = \Theta(n^2mp^2)$, which goes to infinity as $n \to \infty$, because $np = \Omega\left(\sqrt{\frac{n}{m}}\right) = \Omega(1)$ in the range of parameters that we consider. In particular, by (4), we have

$$\mathbb{E}_{\mathbf{R}}[\texttt{Max-Cut}(G)] = \Theta(n^2mp^2). \tag{5}$$

By Chebyshev's inequality, for any $\epsilon > 0$, we have

$$\Pr\left(|\texttt{Max-Cut}(G) - \mathbb{E}_{\mathbf{R}}[\texttt{Max-Cut}(G)]| \geq \epsilon n^2mp^2\right) \leq \frac{\mathrm{Var}_{\mathbf{R}}(\texttt{Max-Cut}(G))}{\epsilon^2n^4m^2p^4}, \tag{6}$$

where $\mathrm{Var}_{\mathbf{R}}$ denotes variance with respect to $\mathbf{R}$. To bound the variance on the right hand side of the above inequality, we use the Efron-Stein inequality. In particular, we write $\texttt{Max-Cut}(G) := f(\mathbf{R})$, i.e. we view $\texttt{Max-Cut}(G)$ as a function of the label choices. For $\ell \in [m], i \in [n]$, we also write $\mathbf{R}^{(\ell,i)}$ for the matrix $\mathbf{R}$ where entry (ℓ, i) has been replaced by an independent, identically distributed (i.i.d.) copy of $\mathbf{R}_{\ell,i}$, which we denote by $\mathbf{R}'_{\ell,i}$. By the Efron-Stein inequality, we now have

$$\mathrm{Var}_{\mathbf{R}}(\texttt{Max-Cut}(G)) \leq \frac{1}{2} \sum_{\ell\in[m],i\in[n]} \mathbb{E}\left[\left(f(\mathbf{R}) - f\left(\mathbf{R}^{(\ell,i)}\right)\right)^2\right]. \tag{7}$$

Notice now that, given all entries of $\mathbf{R}$ except $\mathbf{R}_{\ell,i}$, the probability that $f(\mathbf{R})$ is different from $f\left(\mathbf{R}^{(\ell,i)}\right)$ with probability at most $\Pr(\mathbf{R}_{\ell,i} \neq \mathbf{R}'_{\ell,i}) = 2p(1-p)$. Furthermore, if $L_\ell \backslash \{i\}$ is the set of vertices different from i which have selected ℓ, we then have that $\left(f(\mathbf{R}) - f\left(\mathbf{R}^{(\ell,i)}\right)\right)^2 \leq |L_\ell \backslash \{i\}|^2$, because the intersection graph with representation matrix $\mathbf{R}$ differs by at most $|L_\ell\backslash\{i\}|$ edges from the intersection graph with representation matrix $\mathbf{R}^{(\ell,i)}$. Notice now that, by definition, $|L_\ell\backslash\{i\}|$ follows the Binomial distribution $\mathcal{B}(n-1,p)$. In particular, $\mathbb{E}\left[|L_\ell\backslash\{i\}|^2\right] = (n-1)p(np-2p+1)$, implying $\mathbb{E}\left[\left(f(\mathbf{R}) - f\left(\mathbf{R}^{(\ell,i)}\right)\right)^2\right] \leq 2p(1-p)(n-1)p(np-2p+1)$, for any fixed $\ell \in [m], i \in [n]$.

Putting this all together, (7) becomes

$$\begin{aligned}\mathrm{Var}_{\mathbf{R}}(\texttt{Max-Cut}(G)) &\leq \frac{1}{2} \sum_{\ell\in[m],i\in[n]} 2p(1-p)(n-1)p(np-2p+1) \\ &= nmp(1-p)(n-1)p(np-2p+1) = O(n^3mp^3), \end{aligned} \tag{8}$$

where the last equation comes from the fact that, in the range of values that we consider, we have $p = o(1)$ and $np = \Omega(1)$. Therefore, by (6), we get

$$\Pr\left(|\texttt{Max-Cut}(G) - \mathbb{E}_{\mathbf{R}}[\texttt{Max-Cut}(G)]| \geq \epsilon n^2mp^2\right) \leq \frac{O(n^3mp^3)}{\epsilon^2 n^4 m^2 p^4} = O\left(\frac{1}{\epsilon^2 nmp}\right),$$

which goes to 0 in the range of values that we consider. Together with (5), the above bound proves that $\texttt{Max-Cut}(G)$ is concentrated around its expected value, and the proof is completed. ◀

3.1 Max-Cut for small number of labels

Using Theorem 5, we can now show that, in the case $m = n^\alpha, \alpha < 1$, and $p = O\left(\frac{1}{\sqrt{m}}\right)$, a random cut has asymptotically the same weight as $\texttt{Max-Cut}(G)$, where $G = G(V, E, \mathbf{R}^T\mathbf{R})$ is a random instance of $\overline{\mathcal{G}}_{n,m,p}$. In particular, let $\mathbf{x}^{(rand)}$ be constructed as follows: for each $i \in [n]$, set $x_i^{(rand)} = -1$ independently with probability $\frac{1}{2}$, and $x_i^{(rand)} = +1$ otherwise.

The proof details of the following Theorem can be found in Section C of the Appendix. In view of equation (3), the main idea is to prove that, with high probability over random $\mathbf{x}$ and $\mathbf{R}$, $\|\mathbf{R}\mathbf{x}\|^2$ is asymptotically smaller than the expectation of the weight of the cut defined by $\mathbf{x}^{(rand)}$, in which case the theorem follows by concentration of $\texttt{Max-Cut}(G)$ around its expected value (Theorem 5), and straightforward bounds on $\texttt{Max-Cut}(G)$.

▶ **Theorem 6.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model with $m = n^a, \alpha < 1$, and $C_1\sqrt{\frac{1}{nm}} \leq p \leq C_2\frac{1}{\sqrt{m}}$, for arbitrary positive constants C_1, C_2, and let $\mathbf{R}$ be its representation matrix. Then the cut weight of the random 2-coloring $\mathbf{x}^{(rand)}$ satisfies $\texttt{Cut}(G, \mathbf{x}^{(rand)}) = (1 - o(1))\texttt{Max-Cut}(G)$ with high probability over the choices of $\mathbf{x}^{(rand)}$, $\mathbf{R}$.*

We note that the same analysis also holds when $n = m$ and p is sufficiently large (e.g. $p = \omega(\frac{\ln n}{n})$); more details can be found at the end of Section C of the Appendix. In view of this, in the following sections we will only assume $m = n$ (i.e. $\alpha = 1$) and also $p = \frac{c}{n}$, for some positive constant c. Besides avoiding complicated formulae for p, the reason behind this assumption is that, in this range of values, the expected weight of a fixed edge in $G(V, E, \mathbf{R}^T\mathbf{R})$ is equal to $mp^2 = \Theta(1/n)$, and thus we hope that our work will serve as an intermediate step towards understanding algorithmic bottlenecks for finding maximum cuts in Erdős-Rényi random graphs $G_{n,c/n}$ with respect to their intersection number.

4 Algorithmic results (randomized algorithms)

4.1 The Majority Cut Algorithm

In the following algorithm, the 2-coloring representing the bipartition of a cut is constructed as follows: initially, a small constant fraction ϵ of vertices are randomly placed in the two partitions, and then in each subsequent step, one of the remaining vertices is placed in the partition that maximizes the weight of incident edges with endpoints in the opposite partition.

Algorithm 1 Majority Cut

Input: $G(V, E, \mathbf{R}^T\mathbf{R})$ and its representation matrix $\mathbf{R} \in \{0, 1\}^{m \times n}$
Output: Large cut 2-coloring $\mathbf{x} \in \{-1, +1\}^n$

```
Let v_1, ..., v_n an arbitrary ordering of vertices;
for t = 1 to εn do
    Set x_t to either −1 or +1 independently with equal probability;
for t = εn + 1 to n do
    if Σ_{i∈[t−1]} [R^T R]_{i,t} x_i ≥ 0 then
        x_t = −1;
    else
        x_t = +1;
return x;
```

Clearly the Majority Algorithm runs in polynomial time in n, m. Furthermore, the following Theorem provides a lower bound on the expected weight of the cut constructed by the algorithm in the case $m = n$, $p = \frac{c}{n}$, for large constant c, and $\epsilon \to 0$. The full proof details can be found in Section D of the Appendix.

▶ **Theorem 7.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $m = n$, and $p = \frac{c}{n}$, for large positive constant c, and let $\mathbf{R}$ be its representation matrix. Then, with high probability over the choices of $\mathbf{R}$, the majority algorithm constructs a cut with expected weight at least $(1 + \beta)\frac{1}{4}\mathbb{E}\left[\sum_{i \neq j, i,j \in [n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j}\right]$, where $\beta = \beta(c) \geq \sqrt{\frac{16}{27\pi c^3}} - o(1)$ is a constant, i.e. at least $1 + \beta$ times larger than the expected weight of a random cut.*

Proof sketch. Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $m = n$, and $p = \frac{c}{n}$, for some large enough constant c. For $t \in [n]$, let M_t denote the constructed cut size just after the consideration of a vertex v_t, for some $t \geq \epsilon n + 1$. By equation (3) for $n = t$, and since the values $x_1, \ldots, x_{t-1}$ are already decided in previous steps, we have $M_t = \frac{1}{4}\left(\sum_{i,j \in [t]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \min_{x_t \in \{-1,+1\}} \left\|\mathbf{R}_{[m],[t]}\mathbf{x}_{[t]}\right\|^2\right)$, and after careful calculation we get the recurrence

$$M_t = M_{t-1} + \frac{1}{2} \sum_{i \in [t-1]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,t} + \frac{1}{2} |Z_t| ,$$

where $Z_t = Z_t(\mathbf{x}, \mathbf{R}) = \sum_{i\in[t-1]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,t} x_i = \sum_{\ell\in[m]} \mathbf{R}_{\ell,t} \sum_{i\in[t-1]} \mathbf{R}_{\ell,i}x_i$. Observe that, in the latter recursive equation, the term $\frac{1}{2} \sum_{i\in[t-1]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,t}$ corresponds to the expected increment of the constructed cut if the t-vertex chose its color uniformly at random. Therefore, lower bounding the expectation of $\frac{1}{2} |Z_t|$ will tell us how much better the Majority Algorithm does when considering the t-th vertex.

Towards this end, we note that, given $\mathbf{x}_{[t-1]} = \{x_i, i \in [t-1]\}$, and $\mathbf{R}_{[m],[t-1]} = \{\mathbf{R}_{\ell,i}, \ell \in [m], i \in [t-1]\}$, Z_t is the sum of m independent random variables, since the Bernoulli random variables $\mathbf{R}_{\ell,t}, \ell \in [m]$, are independent, for any given t (note that the conditioning is essential for independence, otherwise the inner sums in the definition of Z_t would also depend on the x_i's, which are not random when i is large). By using a domination argument, we can then prove that

$$\mathbb{E}[|Z_t| \big| \mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}] \geq \mathrm{MD}(Z_t^B),$$

where Z_t^B is a certain Binomial random variable (formally defined in the full proof), and $\mathrm{MD}(\cdot)$ is the mean absolute difference of (two independent copies of) Z_t^B, namely $\mathrm{MD}(Z_t^B) = \mathbb{E}[\big|Z_t^B - Z_t'^B\big|]$. Even though we are aware of no simple closed formula for $\mathrm{MD}(Z_t^B)$, we resort to Gaussian approximation of $Z_t^B - Z_t'^B$ through the Berry-Esseen Theorem, ultimately showing that $|Z_t^B - Z_t'^B|$ follows approximately the *folded normal distribution*. In particular, we show that $\mathrm{MD}(Z_t^B) \geq \sqrt{\frac{c(t-1)}{3\pi n}} - o(1)$, and since the right hand side is independent of $\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}$, we get the same lower bound on the expectation of $|Z_t|$, namely, $\mathbb{E}[|Z_t|] \geq \sqrt{\frac{c(t-1)}{3\pi n}} - o(1)$. Summing over all $t \geq \epsilon n + 1$, we get

$$\sum_{t \geq \epsilon n+1} \mathbb{E}\left[|Z_t|\right] \geq \sqrt{\frac{c}{3\pi}} \left(\frac{2}{3} - \epsilon^{3/2}\right) n - o(n),$$

and the result follows by noting that the expected weight of a random cut is equal to $\frac{1}{4}n(n-1)mp^2 = \frac{c^2}{4}n + o(n)$, and taking $\epsilon \to 0$.

◀

4.2 Intersection graph (weak) bipartization

Notice that we can view a weighted intersection graph $G(V, E, \mathbf{R}^T\mathbf{R})$ as a multigraph, composed by m (possibly) overlapping cliques corresponding to the sets of vertices having chosen a certain label, namely $L_\ell = \{v : \mathbf{R}_{\ell,v}\}, \ell \in [m]$. In particular, let $K^{(\ell)}$ denote the clique induced by label ℓ. Then $G = \cup^+_{\ell\in[m]} K^{(\ell)}$, where $\cup^+$ denotes union that keeps multiple edges. In this section, we present an algorithm that takes as input an intersection graph G given as a union of overlapping cliques and outputs a subgraph that is "almost" bipartite.

To facilitate the presentation of our algorithm, we first give some useful definitions. A *closed vertex-label sequence* is a sequence of alternating vertices and labels starting and ending at the same vertex, namely $\sigma := v_1, \ell_1, v_2, \ell_2, \cdots, v_k, \ell_k, v_{k+1} = v_1$, where the size of the sequence $k = |\sigma|$ is the number of its labels, $v_i \in V$, $\ell_i \in \mathcal{M}$, and $\{v_i, v_{i+1}\} \subseteq L_{\ell_i}$, for all $i \in [k]$ (i.e. v_i is connected to v_{i+1} in the intersection graph). We will also say that label ℓ is *strong* if $|L_\ell| \geq 3$, otherwise it is *weak*. For a given closed vertex-label sequence σ, and any integer $\lambda \in [|\sigma|]$, we will say that σ is λ-*strong* if $|L_{\ell_i}| \geq 3$, for λ indices $i \in [|\sigma|]$. The structural Lemma below is useful for our analysis (see Section E of the Appendix for the proof).[2]

▶ **Lemma 8.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $m = n$, and $p = \frac{c}{n}$, for some constant $c > 0$. With high probability over the choices of $\mathbf{R}$, 0-strong closed vertex-label sequences in G do not have labels in common.*

The following definition is essential for the presentation of our algorithm.

▶ **Definition 9.** *Given a weighted intersection graph $G = G(V, E, \mathbf{R}^T\mathbf{R})$ and a subgraph $G^{(b)} \subseteq G$, let $\mathcal{C}_{odd}(G^{(b)})$ be the set of odd length closed vertex-label sequences $\sigma := v_1, \ell_1, v_2, \ell_2, \cdots, v_k, \ell_k, v_{k+1} = v_1$ that additionally satisfy the following:*

- ***(a)*** *σ has distinct vertices (except the first and the last) and distinct labels.*
- ***(b)*** *v_i is connected to v_{i+1} in $G^{(b)}$, for all $i \in [|\sigma|]$.*
- ***(c)*** *σ is λ-strong, for some $\lambda > 0$.*

Algorithm 2 initially replaces each clique $K^{(\ell)}$ by a random maximal matching $M^{(\ell)}$, and thus gets a subgraph $G^{(b)} \subseteq G$. If $\mathcal{C}_{odd}(G^{(b)})$ is not empty, then the algorithm selects $\sigma \in \mathcal{C}_{odd}(G^{(b)})$ and a strong label $\ell \in \sigma$, and then replaces $M^{(\ell)}$ in $G^{(b)}$ by a new random matching of $K^{(\ell)}$. The algorithm repeats until all odd cycles are destroyed (or runs forever trying to do so).

Algorithm 2 Intersection Graph Weak Bipartization

```
Input: Weighted intersection graph G = ∪⁺_{ℓ∈[m]} K^(ℓ)
Output: A subgraph of G^(b) that has only 0-strong odd cycles
for each ℓ ∈ [m] do
    Let M^(ℓ) be a random maximal matching of K^(ℓ);
Set G^(b) = ∪⁺_{ℓ∈[m]} M^(ℓ) ;
while C_odd(G^(b)) ≠ ∅ do
    Let σ ∈ C_odd(G^(b)) and ℓ a label in σ with |L_ℓ| ≥ 3;
    Replace the part of G^(b) corresponding to ℓ by a new random maximal matching
      M^(ℓ);
return G^(b);
```

The following results are the main technical tools that justify the use of the Weak Bipartization Algorithm for Weighted Max Cut. The proof details for Lemma 10 and Theorem 11 can be found in Sections F and G of the Appendix respectively.

▶ **Lemma 10.** *If $\mathcal{C}_{odd}(G^{(b)})$ is empty, then $G^{(b)}$ may only have 0-strong odd cycles.*

▶ **Theorem 11.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $n = m$ and $p = \frac{c}{n}$, where $c > 0$ is a constant, and let $\mathbf{R}$ be its representation matrix. Let also Σ be a set system with incidence matrix $\mathbf{R}$. With high probability over the choices of $\mathbf{R}$, if Algorithm 2 for weak bipartization terminates on input G, its output can be used to construct a 2-coloring $\mathbf{x}^{(disc)} \in \arg\min_{\mathbf{x}\in\{\pm 1\}^n} disc(\Sigma, \mathbf{x})$, which also gives a maximum cut in G, i.e. $\mathbf{x}^{(disc)} \in \arg\max_{\mathbf{x}\in\{\pm 1\}^n} Cut(G, \mathbf{x})$.*

[2] We conjecture that the structural property of Lemma 8 also holds if we replace 0-strong with λ-strong, for any constant λ, but this stronger version is not necessary for our analysis.

The fact that Theorem 11 is not an immediate consequence of Corollary 4 follows from the observation that a random set system with incidence matrix $\mathbf{R}$ has discrepancy larger than 1 with (at least) constant probability when $m = n$ and $p = \frac{c}{n}$. Indeed, by a straightforward counting argument, we can see that the expected number of 0-strong odd cycles is at least constant. Furthermore, in any 2-coloring of the vertices at least one of the weak labels forming edges in a 0-strong odd cycle will be monochromatic. Therefore, with at least constant probability, for any $\mathbf{x} \in \{-1, +1\}^n$, there exists a weak label ℓ, such that $x_i x_j = 1$, for both $i, j \in L_\ell$, implying that $\text{disc}(L_\ell) = 2$.

We close this section by a result indicating that the conditional statement of Theorem 11 is not void, namely there is a range of values for c where the Weak Bipartization Algorithm terminates in polynomial time. The proof details can be found in Section H of the Appendix.

▶ **Theorem 12.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $n = m$ and $p = \frac{c}{n}$, where $0 < c < 1$ is a constant, and let $\mathbf{R}$ be its representation matrix. With high probability over the choices of $\mathbf{R}$, Algorithm 2 for weak bipartization terminates on input G in $O\left((n + \sum_{\ell \in [m]} |L_\ell|) \cdot \log n\right)$ polynomial time.*

5 Discussion and some open problems

In this paper, we introduced the model of weighted random intersection graphs and we studied the average case analysis of Weighted Max Cut through the prism of discrepancy of random set systems. In particular, in the first part of the paper, we proved concentration of the weight of a maximum cut of $G(V, E, \mathbf{R}^T\mathbf{R})$ around its expected value, and we used it to show that, with high probability, the weight of a random cut is asymptotically equal to the maximum cut weight of the input graph, when $m = n^\alpha, \alpha < 1$. On the other hand, in the case where the number of labels is equal to the number of vertices (i.e. $m = n$), we proved that a majority algorithm gives a cut with weight that is larger than the weight of a random cut by at least a constant factor, when $p = \frac{c}{n}$ and c is large.

In the second part of the paper, we highlighted a connection between Weighted Max Cut of sparse weighted random intersection graphs and Discrepancy of sparse random set systems, formalized through our Weak Bipartization Algorithm and its analysis. We demonstrated how our proposed framework can be used to find optimal solutions for these problems, with high probability, in special cases of sparse inputs ($m = n, p = \frac{c}{n}, c < 1$).

One of the main problems left open in our work concerns the termination of our Weak Bipartization Algorithm for large values of c. We conjecture the following:

▶ **Conjecture 13.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $m = n$, and $p = \frac{c}{n}$, for some constant $c \geq 1$. With high probability over the choices of $\mathbf{R}$, on input G, Algorithm 2 for weak bipartization terminates in polynomial time.*

We also leave the problem of determining whether Algorithm 2 terminates in polynomial time, in the case $m = n$ and $p = \omega(1/n)$, as an open question for future research.

Towards strengthening the connection between Weighted Max Cut under the $\overline{\mathcal{G}}_{n,m,p}$ model, and Discrepancy in random set systems, we conjecture the following:

▶ **Conjecture 14.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $m = n^\alpha, \alpha \leq 1$ and $mp^2 = O(1)$, and let $\mathbf{R}$ be its representation matrix. Let also Σ be a set system with incidence matrix $\mathbf{R}$. Then, with high probability over the choices of $\mathbf{R}$, there exists $\mathbf{x}^{disc} \in \arg\min_{\mathbf{x} \in \{-1,+1\}^n} disc(\Sigma, \mathbf{x})$, such that $\texttt{Cut}(G, \mathbf{x}^{disc})$ is asymptotically equal to $\texttt{Max-Cut}(G)$.*

References

1. D. Altschuler, and J. Niles-Weed. The Discrepancy of Random Rectangular Matrices. CoRR abs/2101.04036 (2021)
2. N. Bansal, and R. Meka. On the discrepancy of random low degree set systems. In Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms SODA 2019: 2557-2564.
3. F. Barahona, M. Gr otschel, M. J unger, and G. Reinelt. An Application of Combinatorial Optimization to Statistical Physics and Circuit Layout Design. Operations Research. 36 (3): 493–513, 1988.
4. M. Bayati, D. Gamarnik, and P. Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Ann. Probab. 41 (2013), 4080–4115.
5. M. Behrisch, A. Taraz, and M. Ueckerdt. Coloring Random Intersection Graphs and Complex Networks. SIAM J. Discret. Math. 23(1): 288-299 (2009).
6. M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas, and K. Rybarczyk. Recent Progress in Complex Network Analysis: Models of Random Intersection Graphs. Studies in Classification, Data Analysis, and Knowledge Organization, Springer 2015, pages 69-78.
7. A. Coja-Oghlan, C. Moore, and V. Sanwalani. MAX k-CUT and approximating the chromatic number of random graphs. Random Struct. Algorithms 28(3): 289-322 (2006).
8. D. Coppersmith, D. Gamarnik, M. Hajiaghayi, and G. Sorkin. Random maxsat, random maxcut, and their phase transitions. Rand. Struct. Alg. 24 (2004), no. 4, 502–545.
9. A. Dembo, A. Montanari, and S. Sen. Extremal Cuts of Sparse Random Graphs. The Annals of Probability, 2017, Vol. 45, No. 2, 1190–1217.
10. J. Díaz, J. Petit, and M. Serna. A survey on graph layout problems. ACM Comput. Surveys 34 (2002), 313–356.
11. E. Ezra, and S. Lovett. On the Beck-Fiala Conjecture for Random Set Systems. In Proceedings of Approximation, Randomization, and Combinatorial Optimization - Algorithms and Techniques (APPROX-RANDOM) 2016: 29:1-29:10.
12. J. Fill, E. Sheinerman, and K. Singer-Cohen. Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models. Random Struct. Algorithms 16(2), 156-176 (2000).
13. D. Gamarnik, and Q. Li. On the max-cut of sparse random graphs. Random Struct. Algorithms 52(2): 219-262 (2018).
14. R. Hoberg, and T. Rothvoss. A Fourier-Analytic Approach for the Discrepancy of Random Set Systems. In Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) 2019: 2547-2556.
15. M. Karoński, E. Scheinerman, and K. Singer-Cohen. On random intersection graphs: the subgraph problem. Combinatorics, Probability and Computing journal 8: 131-159 (1999).
16. S. Nikoletseas, C. Raptopoulos, and P. Spirakis. Efficient Approximation Algorithms in Random Intersection Graphs. Handbook of Approximation Algorithms and Metaheuristics (2), Chapman and Hall/CRC, 2018.
17. C. Papadimitriou, and M. Yannakakis. Optimization, approximation, and complexity classes. Journal of Computer and System Sciences, 43 (3): 425–440, 1991.
18. J. Poland, and T. Zeugmann. Clustering pairwise distances with missing data: Maximum cuts versus normalized cuts. In Lecture Notes in Comput. Sci., 4265 (2006), pp. 197-208.
19. S. Poljak, and Z. Tuza. Maximum cuts and largest bipartite subgraphs. DIMACS series in Discrete Mathematics and Theoretical Computer Science, vol. 20, pp. 181–244, American Mathematical Society, Providence, R.I., 1995.
20. A. Potukuchi. Discrepancy in random hypergraph models. CoRR abs/1811.01491 (2018)
21. C. Raptopoulos, and P. Spirakis. Simple and Efficient Greedy Algorithms for Hamilton Cycles in Random Intersection Graphs. In Proceedings of the 16th International Symposium on Algorithms and Computation (ISAAC) 2005: 493-504.
22. K. Rybarczyk. Equivalence of a random intersection graph and $G(n, p)$. Random Structures and Algorithms 38(1-2)): 205-234 (2011).

23 I. Shevtsova. On the absolute constants in the Berry Esseen type inequalities for identically distributed summands. arXiv:1111.6554 [math.PR].

A Proof of Corollary 3

We first prove the following Lemma, by straightforward calculation from equation (1):

▶ **Lemma 15.** *Let $G(V, E, \mathbf{W})$ be a weighted graph such that $\mathbf{W}$ is symmetric and $\mathbf{W}_{i,j} = 0$ if $\{i, j\} \notin E$. Then*

$$\texttt{Max-Cut}(G) = \frac{1}{4}\left(\sum_{i,j\in[n]^2} \mathbf{W}_{i,j} - \min_{\mathbf{x}\in\{-1,+1\}^n} \mathbf{x}^T\mathbf{W}\mathbf{x}\right). \tag{9}$$

Proof. For any $\mathbf{x} \in \{-1, +1\}^n$, we write

$$\begin{aligned}
\sum_{i,j\in[n]^2} \mathbf{W}_{i,j} - \mathbf{x}^T\mathbf{W}\mathbf{x} &= \sum_{i,j\in[n]^2} \mathbf{W}_{i,j} - \sum_{i,j\in[n]^2} \mathbf{W}_{i,j}x_ix_j \\
&= \frac{1}{2}\sum_{i,j\in[n]^2} \mathbf{W}_{i,j}\left(x_i^2 + x_j^2 - 2x_ix_j\right) \\
&= \frac{1}{2}\sum_{i,j\in[n]^2} \mathbf{W}_{i,j}\left(x_i - x_j\right)^2 \\
&= \sum_{\{i,j\}\in E} \mathbf{W}_{i,j}\left(x_i - x_j\right)^2.
\end{aligned}$$

By (1), this completes the proof. ◀

Proof of Corollary 3. Notice that diagonal entries of the weight matrix in (9) cancel out, and so, for any $\mathbf{x} \in \{-1, +1\}^n$, we have

$$\sum_{i,j\in[n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \|\mathbf{R}\mathbf{x}\|^2 = \sum_{i\neq j, i,j\in[n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \sum_{i\neq j, i,j\in[n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} x_ix_j.$$

Taking expectations with respect to $\mathbf{x}$, the contribution of the second sum in the above expression equals 0, which completes the proof. ◀

B Proof of Corollary 4

Proof. Since $\text{disc}(\Sigma, \mathbf{x}^*) \le 1$, then each component of $\mathbf{R}\mathbf{x}^*$ is either 0 or 1, for any $\mathbf{x}^* \in \{-1, +1\}^n$. In particular, for any $\ell \in [m]$, $[\mathbf{R}\mathbf{x}^*]_\ell$ is 0 if the number of ones in the ℓ-th row is even and it is equal to 1 otherwise. This is the best one can hope for, since sets with an odd number of elements cannot have discrepancy less than 1. Therefore, $\|\mathbf{R}\mathbf{x}^*\|$ is also the minimum possible. In particular, this implies that, in the case $\text{disc}(\Sigma, \mathbf{x}^*) \le 1$, any 2-coloring that achieves minimum discrepancy gives a bipartition that corresponds to a maximum cut and vice versa. ◀

C Proof of Theorem 6

Proof. Let $G = G(V, E, \mathbf{R}^T\mathbf{R})$ be a weighted random intersection graph. By equation (2) of Corollary 3, for any $\mathbf{x} \in \{-1, +1\}^n$, we have:

$$\texttt{Cut}(G, \mathbf{x}) = \frac{1}{4}\left(\sum_{i,j\in[n]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \|\mathbf{R}\mathbf{x}\|^2\right).$$

Taking expectations with respect to random $\mathbf{x}$ and $\mathbf{R}$, we get

$$\mathbb{E}_{\mathbf{x},\mathbf{R}}[\texttt{Cut}(G,\mathbf{x})] = \frac{1}{4}\cdot\mathbb{E}_{\mathbf{R}}\left[\sum_{i,j\in[n]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \sum_{i\in[n]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,i}\right]$$

$$= \frac{1}{4}\cdot\mathbb{E}_{\mathbf{R}}\left[\sum_{i\neq j,i,j\in[n]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j}\right] = \frac{1}{4}n(n-1)mp^2. \tag{10}$$

To prove the Theorem, we will show that, with high probability over random $\mathbf{x}$ and $\mathbf{R}$, we have $\|\mathbf{Rx}\|^2 = o\left(\mathbb{E}_{\mathbf{R}}\left[\frac{1}{4}\sum_{i\neq j,i,j\in[n]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j}\right]\right) = o(n^2mp^2)$, in which case the theorem follows by concentration of $\texttt{Max-Cut}(G)$ around its expected value (Theorem 5), and the fact that $\texttt{Max-Cut}(G) \geq \frac{1}{4}\sum_{i\neq j,i,j\in[n]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j}$.

To this end, fix $\ell \in [m]$ and consider the random variable counting the number of ones in the ℓ-th row of $\mathbf{R}$, namely $Y_\ell = \sum_{i\in[n]}\mathbf{R}_{\ell,i}$. By the multiplicative Chernoff bound, for any $\delta > 0$,

$$\Pr(Y_\ell > (1+\delta)np) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{np}.$$

Since $np \geq C_1\sqrt{\frac{n}{m}} = C_1 n^{\frac{1-\alpha}{2}}$, taking any $\delta \geq 2$, we get

$$\Pr(Y_\ell > 3np) \leq \left(\frac{e^2}{27}\right)^{np} = o\left(\frac{1}{m}\right). \tag{11}$$

Therefore, by the union bound,

$$\Pr(\exists \ell \in [m] : Y_\ell > 3np) = o(1), \tag{12}$$

implying that, all rows of $\mathbf{R}$ have at most $3np$ non-zero elements with high probability.

Fix now ℓ and consider the random variable corresponding to the ℓ-th entry of $\mathbf{Rx}$, namely $Z_\ell = \sum_{i\in[n]}\mathbf{R}_{\ell,i}x_i$. In particular, given Y_ℓ, notice that Z_ℓ is equal to the sum of Y_ℓ independent random variables $x_i \in \{-1,+1\}$, for i such that $\mathbf{R}_{\ell,i} = 1$. Therefore, since $\mathbb{E}_{\mathbf{x}}[Z_\ell] = \mathbb{E}_{\mathbf{x}}[Z_\ell|Y_\ell] = 0$, by Hoeffding's inequality, for any $\lambda \geq 0$,

$$\Pr(|Z_\ell| > \lambda|Y_\ell) \leq e^{-\frac{\lambda^2}{2Y_\ell}}.$$

Therefore, by the union bound, and taking $\lambda \geq \sqrt{6np\ln n}$,

$$\Pr(|Z_\ell| > \lambda) \leq \Pr(\exists \ell \in [m] : Y_\ell > 3np) + me^{-\frac{\lambda^2}{6np}} = o(1) + \frac{m}{n} = o(1), \tag{13}$$

implying that all entries of $\mathbf{Rx}$ have absolute value at most $\sqrt{6np\ln n}$ with high probability over the choices of $\mathbf{x}$ and $\mathbf{R}$. Consequently, with high probability over the choices of $\mathbf{x}$ and $\mathbf{R}$, we have $\|\mathbf{Rx}\|^2 = 6mnp\ln n$, which is $o(n^2mp^2)$, since $np = \omega(\ln n)$ in the range of parameters considered in this theorem. This completes the proof. ◀

We note that the same analysis also holds when $n = m$ and p is sufficiently large (e.g. $p = \omega(\frac{\ln n}{n})$). In particular, similar probability bounds hold in equations (11), (12) and (13), for the same choices of $\delta \geq 2$ and $\lambda \geq \sqrt{6np\ln n}$, implying that $\|\mathbf{Rx}\|^2 = 6mnp\ln n = o(n^2mp^2)$ with high probability.

D Proof of Theorem 7

Proof. Let $G(V, E, \mathbf{R}^T\mathbf{R})$ (i.e. the input to the Majority Cut Algorithm 1) be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model, with $m = n$, and $p = \frac{c}{n}$, for some large enough constant c. For $t \in [n]$, let M_t denote the constructed cut size just after the consideration of a vertex v_t, for some $t \geq \epsilon n + 1$. In particular, by equation (3) for $n = t$, and since the values $x_1, \dots, x_{t-1}$ are already decided in previous steps, we have

$$M_t = \frac{1}{4}\left(\sum_{i,j\in[t]^2}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \min_{x_t\in\{-1,+1\}}\left\|\mathbf{R}_{[m],[t]}\mathbf{x}_{[t]}\right\|^2\right) \tag{14}$$

The first of the above terms is

$$\frac{1}{4}\sum_{i,j\in[t]^2}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j} = \frac{1}{4}\left(\sum_{i,j\in[t-1]^2}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j} + 2\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t} + \left[\mathbf{R}^T\mathbf{R}\right]_{t,t}\right) \tag{15}$$

and the second term is

$$\begin{aligned}
&-\frac{1}{4}\min_{x_t\in\{-1,+1\}}\left\|\mathbf{R}_{[m],[t]}\mathbf{x}_{[t]}\right\|^2 \\
&= -\frac{1}{4}\min_{x_t\in\{-1,+1\}}\left\|\mathbf{R}_{[m],t}x_t + \sum_{i\in[t-1]}\mathbf{R}_{[m],i}x_i\right\|^2 \\
&= -\frac{1}{4}\min_{x_t\in\{-1,+1\}}\sum_{i,j\in[t]^2}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j}x_ix_j \\
&= -\frac{1}{4}\left(\sum_{i,j\in[t-1]^2}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j}x_ix_j + 2\min_{x_t\in\{-1,+1\}}\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t}x_ix_t + \left[\mathbf{R}^T\mathbf{R}\right]_{t,t}\right)
\end{aligned} \tag{16}$$

By (14), (15) and (16), we have

$$\begin{aligned}
M_t &= M_{t-1} + \frac{1}{2}\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t} - \frac{1}{2}\min_{x_t\in\{-1,+1\}}\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t}x_ix_t \\
&= M_{t-1} + \frac{1}{2}\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t} + \frac{1}{2}\left|\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t}x_i\right|
\end{aligned} \tag{17}$$

Define now the random variable

$$Z_t = Z_t(\mathbf{x}, \mathbf{R}) = \sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t}x_i = \sum_{\ell\in[m]}\mathbf{R}_{\ell,t}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i,$$

so that $M_t = M_{t-1} + \frac{1}{2}\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t} + \frac{1}{2}|Z_t|$. Observe that, in the latter recursive equation, the term $\frac{1}{2}\sum_{i\in[t-1]}\left[\mathbf{R}^T\mathbf{R}\right]_{i,t}$ corresponds to the expected increment of the constructed cut if the t-vertex chose its color uniformly at random. Therefore, lower bounding the expectation of $\frac{1}{2}|Z_t|$ will tell us how much better the Majority Algorithm does when considering the t-th vertex.

Towards this end, we first note that, given $\mathbf{x}_{[t-1]} = \{x_i, i \in [t-1]\}$, and $\mathbf{R}_{[m],[t-1]} = \{\mathbf{R}_{\ell,i}, \ell \in [m], i \in [t-1]\}$, Z_t is the sum of m independent random variables, since the Bernoulli random variables $\mathbf{R}_{\ell,t}, \ell \in [m]$, are independent, for any given t (note that the conditioning is essential for independence, otherwise the inner sums in the definition of Z_t would also depend on the x_i's, which are not random when i is large). Furthermore, $\mathbb{E}[Z_t|\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}] = p\sum_{\ell\in[m]}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i$ and $\text{Var}(Z_t|\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}) = p(1-p)\sum_{\ell\in[m]}\left(\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i\right)^2$. Given $\mathbf{x}_{[t-1]} = \{x_i, i \in [t-1]\}$, and $\mathbf{R}_{[m],[t-1]} = \{\mathbf{R}_{\ell,i}, \ell \in [m], i \in [t-1]\}$, define the sets $A_t^+ = \{\ell \in [m] : \sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i > 0\}$ and $A_t^- = \{\ell \in [m] : \sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i < 0\}$. In particular, given $\mathbf{x}_{[t-1]} = \{x_i, i \in [t-1]\}$, and $\mathbf{R}_{[m],[t-1]} = \{\mathbf{R}_{\ell,i}, \ell \in [m], i \in [t-1]\}$, Z_t can be written as

$$Z_t = \sum_{\ell\in A_t^+}\mathbf{R}_{\ell,t}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i - \sum_{\ell\in A_t^-}\mathbf{R}_{\ell,t}\left|\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i\right|, \tag{18}$$

where $\mathbf{R}_{\ell,t}, \ell \in A_t^+ \cup A_t^-$ are independent Bernoulli random variables with success probability p.

It is a matter of careful calculation to show that $\mathbb{E}\left[|Z_t|\middle|\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}\right]$ is smallest when the conditional expectation $\mathbb{E}\left[Z_t\middle|\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}\right]$ is 0, which happens when the sum of positive factors for the Bernoulli random variables in the definition of Z_t is equal to the sum of negative ones, namely $\sum_{\ell\in A_t^+}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i = \sum_{\ell\in A_t^-}\left|\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i\right|$. Furthermore, we note that $\mathbb{E}[|Z_t|\big|\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}]$ does not increase if we replace $\sum_{\ell\in A_t^+}\mathbf{R}_{\ell,t}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i$ and $\sum_{\ell\in A_t^-}\mathbf{R}_{\ell,t}\left|\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i\right|$ in the expression (18) for Z_t by independent binomial random variables $Z_t^+ \sim \mathcal{B}\left(\sum_{\ell\in A_t^+}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i, p\right)$ and $Z_t^- \sim \mathcal{B}\left(\sum_{\ell\in A_t^-}\left|\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i\right|, p\right)$, respectively.[3]

In view of the above, if Z_t^B is a random variable which, given $\mathbf{x}_{[t-1]} = \{x_i, i \in [t-1]\}$, and $\mathbf{R}_{[m],[t-1]} = \{\mathbf{R}_{\ell,i}, \ell \in [m], i \in [t-1]\}$, follows the Binomial distribution $\mathcal{B}(N_t, p)$, where

$$N_t \stackrel{\text{def}}{=} \max\left(\sum_{\ell\in A_t^+}\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i, \sum_{\ell\in A_t^-}\left|\sum_{i\in[t-1]}\mathbf{R}_{\ell,i}x_i\right|\right), \tag{19}$$

then

$$\mathbb{E}[|Z_t|\big|\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}] \geq \text{MD}(Z_t^B), \tag{20}$$

where $\text{MD}(\cdot)$ is the mean absolute difference of (two independent copies of) Z_t^B. In particular, $\text{MD}(Z_t^B) = \mathbb{E}[|Z_t^B - Z_t'^B|]$, where $Z_t^B, Z_t'^B$ are independent random variables following $\mathcal{B}(N_t, p)$. Unfortunately, we are aware of no simple closed formula for $\text{MD}(Z_t^B)$, and so we resort to Gaussian approximation through the Berry-Esseen Theorem:

▶ **Theorem** (Berry-Esseen Theorem [23]). *Let $X_1, X_2, \dots,$ be independent, identically distributed random variables, with $\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = \sigma^2 > 0$, and $\mathbb{E}[|X_i|^3] = \rho < \infty$. For $N > 0$, let $F_N(\cdot)$ be the cumulative distribution function of $\frac{X_1+\dots+X_N}{\sigma\sqrt{N}}$, and let $\Phi(\cdot)$ be the cumulative distribution function of the standard normal distribution. Then,* $\sup_{x\in\mathbb{R}}|F_N(x) - \Phi(x)| \leq \frac{0.4748\rho}{\sigma^3\sqrt{N}}$.

---

[3] This property follows inductively, by noting that, if $X = \sum_{i=1}^{k} a_iX_i - \sum_{i=k}^{N} a_iX_i$, and $X' = \sum_{i=1}^{k-1} a_iX_i + (a_k - 1)X_k + X_k' - \sum_{i=k}^{N} a_iX_i$, where $k, N, a_i \in \mathbb{N}^+, i \in [N]$, and $X_i, i \in [N], X_k'$ are independent, identically distributed Bernoulli random variables, then $\mathbb{E}[|X|] \geq \mathbb{E}[|X'|]$. Indeed, notice that, the independence of X_k, X_k' implies that these random variables work against each other (with respect to the absolute value) at least half of the time.

In our case, we write $Z_t^B = \sum_{i=1}^{N_t} Z_{t,i}^B$, $Z_t'^B = \sum_{i=1}^{N_t} Z_{t,i}'^B$, and set $X_i = Z_{t,i}^B - Z_{t,i}'^B$, where $Z_{t,i}^B, Z_{t,i}'^B$ are independent Bernoulli random variables with success probability p, for any $i \in [N_t]$. In particular, we have $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = \mathbb{E}[|X_i|^3] = 2p(1-p)$. Therefore, by the Berry-Esseen Theorem, given $\mathbf{x}_{[t-1]} = \{x_i, i \in [t-1]\}$, and $\mathbf{R}_{[m],[t-1]} = \{\mathbf{R}_{\ell,i}, \ell \in [m], i \in [t-1]\}$,the distribution of $Z_t^B - Z_t'^B$ is approximately Normal $\mathcal{N}(0, 2p(1-p)N_t)$, with approximation error $\frac{0.4748}{\sqrt{2p(1-p)N_t}}$.

Notice that the latter approximation error bound becomes $o(1)$ if $N_t = \Theta(n), p = \frac{c}{n}$ and $c \to \infty$. Therefore, we next show that, with high probability over the choices of $\mathbf{R}$, $N_t = \Theta(n)$, for any $t \geq \epsilon n + 1$, where ϵ is the constant used in the Majority Algorithm. In particular, even though we cannot control the variables $x_i \in \{-1, +1\}, i \in [t-1]$, in the definition of N_t, we will find a lower bound that holds whp by using the random variable

$$Y_t = Y_t(\mathbf{R}, \mathbf{x}) \stackrel{\text{def}}{=} \left| \ell \in [m] : \sum_{i \in [t-1]} \mathbf{R}_{\ell,i} \text{ is odd} \right|,$$

and employing the following inequality

$$N_t \geq \frac{Y_t}{2}. \tag{21}$$

Indeed, (21) holds because, for any $i \in [t-1]$, if $\sum_{i \in [t-1]} \mathbf{R}_{\ell,i}$ is odd, then $\left|\sum_{i \in [t-1]} \mathbf{R}_{\ell,i} x_i\right| \geq 1$, no matter what value the x_i's have. Therefore, $\sum_{i \in [t-1]} \mathbf{R}_{\ell,i} x_i$ will contribute at least 1 to one of the two terms in the maximum from the right side of (19), and thus (21) follows.

Notice now that, for any fixed i and $t \geq \epsilon n + 1$, we have $\Pr(\sum_{i \in [t-1]} \mathbf{R}_{\ell,i} \text{ is odd}) = \sum_{j \text{ odd}} \binom{t-1}{j} p^j (1-p)^{t-1-j} = \frac{1}{2}\left(1 - (1-2p)^{t-1}\right) \geq \frac{1}{2}\left(1 - e^{-2p(t-1)}\right) \geq \frac{1}{2}\left(1 - e^{-2c\epsilon}\right)$, where in the last inequality we set $p = \frac{c}{n}$. Taking $c \to \infty$, the latter bound becomes $\frac{1}{2} - o(1)$. Therefore, by independence of the entries of $\mathbf{R}$, Y_t stochastically dominates a binomial random variable $\mathcal{B}(t-1, \frac{1}{3})$. Furthermore, by the multiplicative Chernoff (upper) bound, for any $\delta > 0$,

$$\Pr\left(Y_t < (1-\delta)\frac{t-1}{3}\right) < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\frac{t-1}{3}}.$$

Taking $\delta = \frac{1}{2}$ and noting that $t \geq \epsilon n + 1$, we have

$$\Pr\left(Y_t < \frac{t-1}{6}\right) < \left(\frac{e}{2}\right)^{-\frac{\epsilon n}{6}},$$

which is $o(1/n)$, for any constant $\epsilon > 0$. By the union bound,

$$\Pr\left(\exists t : t \geq \epsilon n + 1, Y_t < \frac{t-1}{6}\right) = o(1).$$

By inequality (21), we thus have that, with high probability over the choices of $\mathbf{R}$, $N_t \geq \frac{t-1}{12} \geq \frac{\epsilon n}{12}$, for all $t \geq \epsilon n + 1$, as needed.

Combining the above, by the Berry-Esseen Theorem, given $\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}$, the distribution of $Z_t^B - Z_t'^B$ is approximately Normal $\mathcal{N}(0, 2p(1-p)N_t)$ with approximation error $o(1)$ as $c \to \infty$, with high probability over the choices of $\mathbf{R}$. In particular, given $\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}$, $|Z_t^B - Z_t'^B|$ follows approximately (i.e. with the same approximation error $o(1)$) the *folded normal distribution* with mean value (at least) $\sqrt{\frac{2}{\pi}\text{Var}(Z_t^B - Z_t'^B | \mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]})}$. Notice now that, by inequality (21), we have

$$\text{Var}(Z_t^B - Z_t'^B | \mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}) \geq p(1-p)Y_t.$$

Since $Y_t \geq \frac{t-1}{6} \geq \frac{\epsilon n}{6}$ with high probability, and also $p = \frac{c}{n}$, we get that $\mathrm{Var}(Z_t^B - Z_t'^B | \mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}) \geq \frac{c(t-1)}{6n} - o(1)$, with high probability, where the $o(1)$ comes from the approximation error given by the Berry-Esseen Theorem. Consequently, by inequality (20), with high probability over the choices of $\mathbf{R}$ (which is $1 - o(1)$),

$$\mathbb{E}\left[|Z_t|\right] = \mathbb{E}\left[\left|\sum_{i \in [t-1]} \left[\mathbf{R}^T \mathbf{R}\right]_{i,t} x_i\right|\right] \geq \sqrt{\frac{c(t-1)}{3\pi n}} - o(1).$$

Summing over all $t \geq \epsilon n + 1$, we get

$$\sum_{t \geq \epsilon n+1} \mathbb{E}\left[|Z_t|\right] \geq \sqrt{\frac{c}{3\pi n}} \sum_{t \geq \epsilon n} \sqrt{t} - o(n) = \sqrt{\frac{c}{3\pi n}} \left(\sum_{t \geq 1} \sqrt{t} - \epsilon n \sqrt{\epsilon n}\right) - o(n).$$

Using the fact that $\sum_{t \geq 1} \sqrt{t} = \frac{2}{3} n^{3/2} + o(n)$, we thus have that

$$\sum_{t \geq \epsilon n+1} \mathbb{E}\left[|Z_t|\right] \geq \sqrt{\frac{c}{3\pi}} \left(\frac{2}{3} - \epsilon^{3/2}\right) n - o(n).$$

On the other hand, we have that the expected weight of a random cut is equal to $\frac{1}{4} n(n-1) m p^2 = \frac{c^2}{4} n + o(n)$ (see e.g. equation (10)). The proof is completed by taking $\epsilon \to 0$. ◀

E Proof of Lemma 8

Proof. We will use the first moment method and so we need to prove that the expectation of the number of pairs of distinct 0-strong closed vertex-label sequences in G that have at least one label in common goes to 0. To this end, for $j \in [\min(k, k') - 1]$, let $A_j(k, k')$ denote the number of such sequences σ, σ', with $k = |\sigma|, k' = |\sigma'|$, that have j labels in common. In particular, for integers k, k', let $\sigma := v_1, \ell_1, v_2, \ell_2, \cdots, v_k, \ell_k, v_{k+1} = v_1$, and let $\sigma' := v_1', \ell_1', v_2', \ell_2', \cdots, v_{k'}', \ell_{k'}', v_{k'+1}' = v_1$. Notice that, any such fixed pair σ, σ' has the same probability to appear, namely $p^{2(k+k'-j)}(1-p)^{(n-2)(k+k'-j)}$; indeed, $p^{2k}(1-p)^{(n-2)k}$ is the probability that σ appears (recall that σ has k labels and it is 0-strong, i.e. each label is only selected by two vertices) and $p^{2(k'-j)}(1-p)^{(n-2)(k'-j)}$ is the probability that σ' appears given that σ has appeared. Furthermore, the number of such pairs of sequences is dominated by the number of sequences that overlap in j consecutive labels (e.g. the first j), which is at most $n^k m^k n^{k'-j-1} m^{k'-j}$ (notice that j common labels implies that there are at least $j' + 1$ common vertices). Overall, since $n = m$ and $p = \frac{c}{n}$, we have

$$\begin{aligned} \mathbb{E}[A_j(k, k')] &\leq (1 + o(1)) \frac{1}{n} (np)^{2(k+k'-j)} (1-p)^{(n-2)(k+k'-j)} \\ &= (1 + o(1)) \frac{1}{n} \left(c^2 (1-p)^{n-2}\right)^{k+k'-j}. \end{aligned}$$

Since $n \to \infty$ and $p = \frac{c}{n}$, by elementary calculus we have that $c^2(1-p)^{n-2}$ bounded by a constant (which depends only on c) strictly less than 1. Therefore, the above expectation is at most $e^{-\ln n - \Theta(1)(k+k'-j)}$. Therefore, summing over all choices of $k, k' \in [n]$ and $j \in [\min(k, k') - 1]$, we get that the expected number of pairs of distinct 0-strong closed vertex-label sequences that have at least one label in common is at most

$$\sum_{k,k' \in [n]} \sum_{j \in [\min(k,k')-1]} e^{-\ln n - \Theta(1)(k+k'-j)} = o(1),$$

and the proof is completed by Markov's inequality. ◀

F Proof of Lemma 10

Proof. For the sake of contradiction, assume $\mathcal{C}_{odd}(G^{(b)}) = \emptyset$, but $G^{(b)} = \cup^{+}_{\ell \in [m]} M^{(\ell)}$ has an odd cycle C_k that is not 0-strong and has minimum length. Notice that C_k corresponds to a closed vertex-label sequence, say $\sigma := v_1, \ell_1, v_2, \ell_2, \cdots, v_k, \ell_k, v_{k+1} = v_1$, where $\{v_i, v_{i+1}\} \in M^{(\ell_i)}$, for all $i \in [k]$. Furthermore, by assumption, conditions (b) and (c) of Definition 9 are satisfied by σ (indeed $\{v_i, v_{i+1}\} \in M^{(\ell_i)}$, for all $i \in [k]$, and σ is λ-strong, for some $\lambda > 0$). Therefore, the only reason for which σ does not belong to $\mathcal{C}_{odd}(G^{(b)})$ is that condition (a) of Definition 9 is not satisfied, i.e. there are distinct indices $i > i' \in [k]$ such that $\ell_i = \ell_{i'}$. Clearly, such indices are not consecutive (i.e. $i' \neq i+1$), because ℓ_i is strong and step 6 of our algorithm implies that $M^{(\ell_i)}$ is a matching of $K^{(\ell_i)}$. But then either the vertex-label sequence $v_1, \ldots, v_i, \ell_i, v_{i'+1}, \ell_{i'+1}, v_{i'+2}, \ldots, v_{k+1} = v_1$ or the vertex-label sequence $v_{i+1}, \ell_{i+1}, v_{i+2}, \ldots, v_{i'}, \ell_i, v_{i+1}$ corresponds to a shorter odd cycle, which is a contradiction on the minimality of C_k. ◀

G Proof of Theorem 11

Proof. By construction, the output of Algorithm 2, namely $G^{(b)}$, has only 0-strong odd cycles. Furthermore, by Lemma 8 these cycles correspond to vertex-label sequencies that are label-disjoint. Let H denote the subgraph of $G^{(b)}$ in which we have destroyed all 0-strong odd cycles by deleting a single (arbitrary) edge e_C from each 0-strong odd cycle C (keeping all other edges intact), and notice that e_C corresponds to a weak label. In particular, H is a bipartite multi-graph and thus its vertices can be partitioned into two independent sets A, B constructed as follows: In each connected component of H, start with an arbitrary vertex v and include in A (resp. in B) the set of vertices reachable from v that are at an even (resp. odd) distance from v. Since H is bipartite, it does not have odd cycles, and thus this construction is well-defined, i.e. no vertex can be placed in both A and B.

We now define $\mathbf{x}^{(disc)}$ by setting $x_i^{(disc)} = +1$ if $i \in A$ and $x_i^{(disc)} = +1$ if $i \in B$. Let $\mathcal{M}_0$ denote the set of weak labels corresponding to the edges removed from $G^{(b)}$ in the construction of H. We first note that, for each $\ell_C \in \mathcal{M}_0$ corresponding to the removal of an edge e_C, we have $\left|\sum_{i \in L_{\ell_C}} x_i^{(disc)}\right| = 2$. Indeed, since e_C belongs to an odd cycle in $G^{(b)}$, its endpoints are at even distance in H, which means that either they both belong to A or they both belong to B. Therefore, their corresponding entries of $\mathbf{x}^{(disc)}$ have the same sign, and so (taking into account that the endpoints of e_C are the only vertices in L_{ℓ_C}), we have $\left|\sum_{i \in L_{\ell_C}} x_i^{(disc)}\right| = 2$. Second, we show that, for all the other labels $\ell \in [m] \backslash \mathcal{M}_0$, $\left|\sum_{i \in L_\ell} x_i^{(disc)}\right|$ will be equal to 1 if $|L_\ell|$ is odd and 0 otherwise. For any label $\ell \in [m] \backslash \mathcal{M}_0$, let $M^{(\ell)}$ denote the part of $G^{(b)}$ corresponding to a maximal matching of $K^{(\ell)}$, and note that all edges of $M^{(\ell)}$ are contained in H. Since H is bipartite, no edge in $M^{(\ell)}$ can have both its endpoints in either A or B. Therefore, by construction, the contribution of entries of $\mathbf{x}^{(disc)}$ corresponding to endpoints of edges in $M^{(\ell)}$ to the sum $\sum_{i \in L_\ell} x_i^{(disc)}$ is 0. In particular, if $|L_\ell|$ is even, then $M^{(\ell)}$ is a perfect matching and $\left|\sum_{i \in L_\ell} x_i^{(disc)}\right| = 0$, otherwise (i.e. if $|L_\ell|$ is odd) there is a single vertex not matched in $M^{(\ell)}$ and $\left|\sum_{i \in L_\ell} x_i^{(disc)}\right| = 1$.

To complete the proof of the theorem, we need to show that $\mathrm{Cut}(G, \mathbf{x}^{(disc)})$ is maximum. By Corollary 3, this is equivalent to proving that $\|\mathbf{R}\mathbf{x}^{(disc)}\| \leq \|\mathbf{R}\mathbf{x}\|$ for all $\mathbf{x} \in \{-1, +1\}^n$.

Suppose that there is some $\mathbf{x}^{(min)} \in \{-1,+1\}^n$ such that $\|\mathbf{R}\mathbf{x}^{(disc)}\| > \|\mathbf{R}\mathbf{x}^{(min)}\|$. As mentioned above, for all $\ell \in [m]\backslash\mathcal{M}_0$, we have $[\mathbf{R}\mathbf{x}^{(disc)}]_\ell \leq 1$, and so $[\mathbf{R}\mathbf{x}^{(disc)}]_\ell \leq [\mathbf{R}\mathbf{x}^{(min)}]_\ell$. Therefore, the only labels where $\mathbf{x}^{(min)}$ could do better are those corresponding to edges e_C that are removed from $G^{(b)}$ in the construction of H, i.e. $\ell_C \in \mathcal{M}_0$, for which we have $[\mathbf{R}\mathbf{x}^{(disc)}]_{\ell_C} = 2$. However, any such edge e_C belongs to an odd cycle C, and thus any 2-coloring of the vertices of C will force at least one of the 0-strong labels corresponding to edges of C to be monochromatic. Taking into account the fact that, by Lemma 8, with high probability over the choices of $\mathbf{R}$, all 0-strong odd cycles correspond to vertex-label sequences that are label-disjoint, we conclude that $\|\mathbf{R}\mathbf{x}^{(disc)}\| \leq \|\mathbf{R}\mathbf{x}^{(min)}\|$, which completes the proof. ◀

H Proof of Theorem 12

We first prove the following structural Lemma regarding the expected number of closed vertex label sequences.

▶ **Lemma 16.** *Let $G(V, E, \mathbf{R}^T\mathbf{R})$ be a random instance of the $\overline{\mathcal{G}}_{n,m,p}$ model. Let also C_k denote the number of distinct closed vertex-label sequences of size k in G. Then*

$$\mathbb{E}[C_k] = \frac{1}{k}\frac{n!}{(n-k)!}\frac{m!}{(m-k)!}p^{2k}. \tag{22}$$

In particular, when $m = n \to \infty$, $p = \frac{c}{n}, c > 0$, and $k \geq 3$, we have $\mathbb{E}[C_k] \leq \frac{e}{2\pi}c^{2k}$.

Proof. Notice that there are $\frac{1}{k}\frac{n!}{(n-k)!}$ ways to arrange k out of n vertices in a cycle. Furthermore, in each such arrangement, there are $\frac{m!}{(m-k)!}$ ways to place k out of m labels so that there is exactly one label between each pair of vertices. Since labels in any given arrangement must be selected by both its adjacent vertices, (22) follows by linearity of expectation.

Setting $m = n$ and $p = \frac{c}{n}$, and using the inequalities $\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n} \leq n! \leq en^{n+\frac{1}{2}}e^{-n}$,

$$\begin{aligned}
\mathbb{E}[C_k] &= \frac{1}{k}\left(\frac{n!}{(n-k)!}\right)^2\left(\frac{c}{n}\right)^{2k} \\
&\leq \frac{1}{k}\frac{e^2n^{2n+1}e^{-2n}}{2\pi(n-k)^{2n-2k+1}e^{2k-2n}}\left(\frac{c}{n}\right)^{2k} = \frac{1}{k}\frac{e^2}{2\pi}\left(\frac{n}{n-k}\right)^{2n-2k+1}\left(\frac{c}{e}\right)^{2k} \\
&\leq \frac{e^2}{2\pi}\frac{n}{k(n-k)}e^{\frac{k}{n-k}(2n-2k)}\left(\frac{c}{e}\right)^{2k} = \frac{e^2}{2\pi}\frac{n}{k(n-k)}c^{2k}.
\end{aligned}$$

When n goes to ∞ and $k \geq 3$, then the above is at most $\frac{e}{2\pi}c^{2k}$ as needed. ◀

We are now ready for the proof of the Theorem.

Proof of Theorem 12. We will prove that, when $m = n \to \infty$, $p = \frac{c}{n}, c < 1$, and $k \geq 3$, with high probability, there are no closed vertex-label sequences that have labels in common. To this end, recalling Definition 9 for $\mathcal{C}_{odd}(G^{(b)})$, we provide upper bounds on the following events: $A \stackrel{\text{def}}{=} \{\exists k \geq \log n : C_k \geq 1\}$, $B \stackrel{\text{def}}{=} \{|\mathcal{C}_{odd}(G^{(b)})| \geq \log n\}$ and $C \stackrel{\text{def}}{=} \{\exists \sigma \neq \sigma' \in \mathcal{C}_{odd}(G^{(b)}) : \exists \ell \in \sigma, \ell \in \sigma'\}$.

By the union bound, Markov's inequality and Lemma 16, we get that, whp all closed vertex-label sequences have less than $\log n$ labels:

$$\Pr(A) \leq \sum_{k \geq \log n}\mathbb{E}[C_k] \leq \sum_{k \geq \log n}\frac{e}{2\pi}c^{2k} = \frac{e}{2\pi}\frac{c^{2\log n}}{1-c^2} = O\left(c^{2\log n}\right) = o(1), \tag{23}$$

where the last equality follows since $c < 1$ is a constant. Furthermore, by Markov's inequality and Lemma 16, and noting that any closed vertex-label sequence in $\mathcal{C}_{odd}(G^{(b)})$ must have at least $k \geq 3$ labels, we get that, whp there less than $\log n$ closed vertex-label sequences in $\mathcal{C}_{odd}(G^{(b)})$:

$$\Pr(B) \leq \frac{1}{\log n} \sum_{k \geq 3} \mathbb{E}[C_k] \leq \frac{1}{\log n} \sum_{k \geq 3} \frac{e}{2\pi} c^{2k} = \frac{1}{\log n} \frac{e}{2\pi} \frac{c^6}{1 - c^2} = O\left(\frac{1}{\log n}\right). \tag{24}$$

To bound $\Pr(C)$, fix a closed vertex-label sequence σ, and let $|\sigma| \geq 3$ be the number of its labels. Notice that, the probability that there is another closed vertex-label sequence that has labels in common with σ implies the existence of a vertex-label sequence $\breve{\sigma}$ that starts with either a vertex or a label from σ, ends with either a vertex or a label from σ, and has at least one label or at least one vertex that does not belong to σ. Let $|\breve{\sigma}|$ denote the number of labels of $\breve{\sigma}$ that do not belong to σ. Then the number of different vertex-label sequences $\breve{\sigma}$ that start and end in labels from σ is at most $|\sigma|^2 n^{|\breve{\sigma}|+1} m^{|\breve{\sigma}|}$; indeed $\breve{\sigma}$ in this case has $|\breve{\sigma}|$ labels and $|\breve{\sigma}| + 1$ vertices that do not belong to σ. Therefore, by independence, each such sequence $\breve{\sigma}$ has probability $p^{2|\breve{\sigma}|+2}$ to appear. Similarly, the number of different vertex-label sequences $\breve{\sigma}$ that start and end in vertices from σ is at most $|\sigma|^2 n^{|\breve{\sigma}|-1} m^{|\breve{\sigma}|}$ and each one has probability $p^{2|\breve{\sigma}|}$ to appear. Finally, the number of different vertex-label sequences $\breve{\sigma}$ that start in a vertex from σ and end in a label from σ (notice that this also covers the case where $\breve{\sigma}$ starts in a label from σ and ends in a vertex from σ) is at most $|\sigma|^2 n^{|\breve{\sigma}|} m^{|\breve{\sigma}|}$ and each one has probability $p^{2|\breve{\sigma}|+1}$ to appear. Overall, for a given sequence σ, the expected number of sequences $\breve{\sigma}$ described above that additionally satisfies $|\breve{\sigma}| < \log n$, is at most

$$\sum_{k=0}^{\log n - 1} |\sigma|^2 n^{k+1} m^k p^{2k+2} + \sum_{k=1}^{\log n - 1} |\sigma|^2 n^{k-1} m^k p^{2k} + \sum_{k=1}^{\log n - 1} |\sigma|^2 n^k m^k p^{2k+1} \leq c|\sigma|^2 \frac{\log n}{n}, \tag{25}$$

where in the last inequality we used the fact that $m = n, p = \frac{c}{n}$ and $c < 1$. Since the existence of a sequence $\breve{\sigma}$ for σ that additionally satisfies $|\breve{\sigma}| \geq \log n$ implies event A, and on other hand the existence of more than $\log n$ different sequences $\sigma \in |\mathcal{C}_{odd}(G^{(b)})|$ implies event B, by Markov's inequality and (25), we get

$$\Pr(C) \leq \Pr(A) + \Pr(B) + c\frac{(\log n)^4}{n} = O\left(c^{2 \log n}\right) + O\left(\frac{1}{\log n}\right) + O\left(\frac{(\log n)^4}{n}\right) = O\left(\frac{1}{\log n}\right).$$

We have thus proved that, with high probability over the choices of $\mathbf{R}$, closed vertex-label sequences in $\mathcal{C}_{odd}(G^{(b)})$ are label disjoint, as needed.

In view of this, the proof of the Theorem follows by noting that, since closed vertex label sequences in $\mathcal{C}_{odd}(G^{(b)})$ are label disjoint, steps 5 and 6 within the while loop of the Weak Bipartization Algorithm will be executed exactly once for each sequence in $\mathcal{C}_{odd}(G^{(b)})$, where $G^{(b)}$ is defined in step 3 of the algorithm; indeed, once a closed vertex label sequence $\sigma \in \mathcal{C}_{odd}(G^{(b)})$ is destroyed in step 6, no new closed vertex label sequence is created. In fact, once σ is destroyed we can remove the corresponding labels and edges from $G^{(b)}$, as these will no longer belong to other closed vertex label sequences. Furthermore, to find a closed vertex label sequences in $\mathcal{C}_{odd}(G^{(b)})$, it suffices to find an odd cycle in $G^{(b)}$, which can be done by running DFS, requiring $O(n + \sum_{\ell \in [m]} |L_\ell|)$ time, because $G^{(b)}$ has at most $\sum_{\ell \in [m]} |L_\ell|$ edges. Finally, by (24), we have $|\mathcal{C}_{odd}(G^{(b)})| < \log n$ with high probability, and so the running time of the Weak Bipartization Algorithm is $O((n + \sum_{\ell \in [m]} |L_\ell|) \log n)$, which concludes the proof of Theorem 12. ◀