Tolerant Bipartiteness Testing in Dense Graphs

Arijit Ghosh ✉ ⌂
Indian Statistical Institute, Kolkata, India

Gopinath Mishra ✉ ⌂
University of Warwick, Coventry, UK

Rahul Raychaudhury ✉
Duke University, Durham, NC, USA

Sayantan Sen ✉ ⌂
Indian Statistical Institute, Kolkata, India

─── **Abstract** ───

Bipartite testing has been a central problem in the area of property testing since its inception in the seminal work of Goldreich, Goldwasser, and Ron. Though the non-tolerant version of bipartite testing has been extensively studied in the literature, the tolerant variant is not well understood. In this paper, we consider the following version of tolerant bipartite testing problem: Given two parameters $\varepsilon, \delta \in (0, 1)$, with $\delta > \varepsilon$, and access to the adjacency matrix of a graph G, we have to decide whether G can be made bipartite by editing at most εn^2 entries of the adjacency matrix of G, or we have to edit at least δn^2 entries of the adjacency matrix to make G bipartite. In this paper, we prove that for $\delta = (2 + \Omega(1))\varepsilon$, tolerant bipartite testing can be decided by performing $\widetilde{\mathcal{O}}\left(1/\varepsilon^3\right)$ many adjacency queries and in $2^{\widetilde{\mathcal{O}}(1/\varepsilon)}$ time complexity. This improves upon the state-of-the-art query and time complexities of this problem of $\widetilde{\mathcal{O}}\left(1/\varepsilon^6\right)$ and $2^{\widetilde{\mathcal{O}}(1/\varepsilon^2)}$, respectively, due to Alon, Fernandez de la Vega, Kannan and Karpinski, where $\widetilde{\mathcal{O}}(\cdot)$ hides a factor polynomial in $\log{(1/\varepsilon)}$.

1 Introduction

The field of property testing refers to the model where the main goal is to design efficient algorithms that reads only "small" part of the input. Over the past few years, the field has had very rapid growth, and several interesting techniques and results have emerged. See, for example, the new property testing book by Goldreich [11] for an introduction to property testing.

The field of graph property testing was first introduced in the seminal work of Goldreich, Goldwasser, and Ron [12]. In that work, the authors studied various interesting and important problems in dense graphs and testing bipartiteness was one of them. In non-tolerant variant

of bipartiteness testing, we are given a dense graph G and a proximity parameter $\varepsilon \in (0,1)$ as the input, and the goal is to decide if G is bipartite, or do we need to modify at least εn^2 many entries of the adjacency matrix of G to make it bipartite, using as few queries to the adjacency matrix of G as possible.

Due to the fundamental nature of the problem, *bipartite testing* has been extensively studied over the past two decades [12]. Though there are several works on non-tolerant testing of various graph properties across all models in graph property testing [12, 13, 7], there are very few works related to their tolerant counterparts (See, for example, the property testing book by Goldreich [11] for an extensive list of various results). To the best of our knowledge, this is the first time tolerant bipartite testing has been *explicitly* studied in the literature.

Now we formally define the notion of *bipartite distance* and state our main result. Then we discuss our result vis-a-vis the related works.

▶ **Definition 1.1** (Bipartite distance). *A bipartition of a graph G is a function $f : V(G) \to \{L, R\}$, where $V(G)$ denotes the vertex set of G [1]. The bipartite distance of G with respect to the bipartition f is denoted and defined as*

$$d_{bip}(G, f) := \left[\sum_{v \in V : f(v) = L} |N(v) \cap f^{-1}(L)| + \sum_{v \in V : f(v) = R} |N(v) \cap f^{-1}(R)| \right].$$

Here $N(v)$ denotes the neighborhood of v in G. Informally, $d_{bip}(G, f)$ measures the distance of the graph G from being bipartite, with respect to the bipartition f. The *bipartite distance* of G is defined as the minimum bipartite distance of G over all possible bipartitions f of G, that is,

$$d_{bip}(G) := \min_f d_{bip}(G, f).$$

Now we are ready to formally state our result.

▶ **Theorem 1.2** (Main result). *Given query access to the adjacency matrix of a dense graph G with n vertices and a proximity parameter $\varepsilon \in (0,1)$, there exists an algorithm that, with probability at least $\frac{9}{10}$, decides whether $d_{bip}(G) \leq \varepsilon n^2$ or $d_{bip}(G) \geq (2 + \Omega(1))\varepsilon n^2$, by sampling $\mathcal{O}\left(\frac{1}{\varepsilon^3} \log \frac{1}{\varepsilon}\right)$ many vertices in $2^{\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)}$ time, and performs $\mathcal{O}\left(\frac{1}{\varepsilon^3} \log^2 \frac{1}{\varepsilon}\right)$ many queries.*

1.1 Our result in the context of literature

Recall that non-tolerant bipartite testing refers to the problem where we are given query access to the adjacency matrix of an unknown graph G and a proximity parameter $\varepsilon \in (0,1)$, and the objective is to decide whether $d_{bip}(G) = 0$ or $d_{bip}(G) \geq \varepsilon n^2$. The problem of non-tolerant bipartite testing in the dense graph model was first studied in the seminal work of Goldreich, Goldwasser, and Ron [12], and they showed that it admits an algorithm with query complexity $\widetilde{\mathcal{O}}\left(1/\varepsilon^3\right)$. Later, Alon and Krivelevich [4] improved the query complexity of the problem to $\widetilde{\mathcal{O}}\left(1/\varepsilon^2\right)$. They further studied the problem of testing c-colorability of dense graph. Note that bipartite testing is a special case of testing c-colorability, when $c = 2$. They proved that c-colorability can be tested by performing $\widetilde{\mathcal{O}}\left(1/\varepsilon^4\right)$ many queries, for $c \geq 3$. This bound was later improved to $\widetilde{\mathcal{O}}\left(1/\varepsilon^2\right)$ by Sohler [18]. On the other hand, for

---

[1] L and R denote left and right respectively.

non-tolerant bipartite testing, Bogdanov and Trevisan [6] proved that $\Omega(1/\varepsilon^2)$ and $\Omega(1/\varepsilon^{3/2})$ many adjacency queries are required by any non-adaptive and adaptive testers, respectively. Later, Gonen and Ron [14] further explored the power of adaptive queries for bipartiteness testing. Bogdanov and Li [5] showed that bipartiteness can be tested with one-sided error in $\mathcal{O}(1/\varepsilon^c)$ queries, for some constant $c < 2$, assuming a conjecture [2].

Though the non-tolerant variant of bipartite testing is well understood, the query complexity of the tolerant version (even for restricted cases like we consider in Theorem 1.2) has not yet been addressed in the literature. From the result of Alon, Vega, Kannan and Karpinski [1],for estimating MAXCUT [3] for any given ε ($0 < \varepsilon < 1$), it implies that that the bipartite distance of a (dense) graph G can be estimated up to an additive error of εn^2, by performing $\widetilde{\mathcal{O}}\left(1/\varepsilon^6\right)$ many queries (see Appendix A for details, and in particular, see Corollary A.2). Even for the tolerant version that we consider in Theorem 1.2, their algorithm does not give any bound better than $\widetilde{\mathcal{O}}\left(1/\varepsilon^6\right)$. Note that Alon, Vega, Kannan and Karpinski [1] improved the result of Goldreich, Goldwasser, and Ron [12], who were the first to prove that MAXCUT can be estimated with an additive error of εn^2 by performing $\widetilde{\mathcal{O}}(1/\varepsilon^7)$ many adjacency queries and with time complexity $2^{\widetilde{\mathcal{O}}(1/\varepsilon^3)}$. Though we improve the bound for tolerant bipartite testing (for the restricted case as stated in Theorem 1.2) substantially from the work of Alon et al. [1], we would like to note that this is the first work that studies tolerant bipartite testing explicitly.

1.2 Other related works

Apart from the dense graph model, this problem has also been studied in other models of property testing. Goldreich and Ron [13] studied the problem of bipartiteness testing for bounded degree graphs, where they gave an algorithm of $\widetilde{\mathcal{O}}(\sqrt{n})$ queries, where n denotes the number of vertices of the graph. Later, Kaufman, Krivelevich and Ron [15] studied the problem in the general graph model and gave an algorithm with query complexity $\widetilde{\mathcal{O}}\left(\min\left\{\sqrt{n}, n^2/m\right\}\right)$, where m denotes the number of edges of the graph. Few years back, Czumaj, Monemizadeh, Onak, and Sohler [7] studied the problem for planar graphs (more generally, for any minor-free graphs), where they employed a random walk based technique, and proved that constant number of queries are enough for bipartiteness testing. Apart from bipartite testing, there have been extensive works related to property testing in the dense graph model and its connection to the regularity lemma [3, 2, 9].

1.3 Organization

In Section 2, we present an overview of our algorithm along with a brief description of its analysis. In Section 3, we formally describe our algorithm, followed by its correctness analysis in Section 4. Finally, we conclude in Section 5. The proofs that are omitted in the main text are either presented in the appendix or in the full version of the paper [10].

1.4 Notations

All graphs considered here are undirected, unweighted, and have no self-loops or parallel edges. For a graph $G = (V(G), E(G))$, $V(G)$ and $E(G)$ denote the vertex and edge sets of G respectively. For a vertex $v \in V(G)$, $N_G(v)$ denotes the neighborhood of a vertex v in

---

[2] The conjecture is that if the graph G is ε-far from bipartite, then an induced subgraph on $\widetilde{\mathcal{O}}(1/\varepsilon)$ many random vertices would be $\widetilde{\Omega}(\varepsilon)$-far from being bipartite with probability at least 1/2.

[3] MAXCUT of a graph G denotes the size of the largest *cut* in G.

G, and we will write it as $N(v)$ when the graph G is clear from the context. Since we are only considering undirected graphs, we write an edge as $\{u, v\} \in E(G)$. For a set of pairs of vertices Z, we will denote the set of vertices present in at least one pair in Z by $V(Z)$. For a function $f : V(G) \to \{L, R\}$, $f^{-1}(L)$ and $f^{-1}(R)$ represent the set of vertices that are mapped to L and R by f respectively. We denote by $\binom{V(G)}{2}$ the set of unordered pairs of the vertices of G. Finally, $a = (1 \pm \varepsilon)b$ represents $(1 - \varepsilon)b \leq a \leq (1 + \varepsilon)b$.

2 Overview of the proof of Theorem 1.2

In this section, we give an overview of our algorithm. The detailed description of the algorithm is presented in Section 3, while its analysis is presented in Section 4. We will prove the following theorem, which is our main technical result.

▶ **Theorem 2.1.** *There exists an algorithm* TOL-BIP-DIST(G, ε) *that given adjacency query access to a dense graph G with n vertices and a parameter $\varepsilon \in (0, 1)$, decides with probability at least $9/10$, whether $d_{bip}(G) \leq \varepsilon n^2$ or $d_{bip}(G) \geq (2 + k)\varepsilon n^2$, by sampling $\mathcal{O}(\frac{1}{k^5 \varepsilon^2} \log \frac{1}{k\varepsilon})$ many vertices in $2^{\mathcal{O}(\frac{1}{k^3 \varepsilon} \log \frac{1}{k\varepsilon})}$ time, using $\mathcal{O}(\frac{1}{k^8 \varepsilon^3} \log^2 \frac{1}{k\varepsilon})$ many queries to the adjacency matrix of G.*

Note that Theorem 2.1 implies Theorem 1.2, by taking $k = \Omega(1)$.

2.1 Brief description of the algorithm

Assume C_1, C_2, C_3 are three suitably chosen large absolute constants. At the beginning of our algorithm, we generate t many subsets of vertices $X_1, \ldots, X_t$, each with $\lceil \frac{C_2}{k^3 \varepsilon} \log \frac{1}{k\varepsilon} \rceil$ many vertices chosen randomly, where $t = \lceil \log \frac{C_1}{k\varepsilon} \rceil$. Let $\mathcal{C} = X_1 \cup \ldots \cup X_t$. Apart from the X_i's, we also randomly select a set of pairs of vertices Z, with $|Z| = \lceil \frac{C_3}{k^5 \varepsilon^2} \log \frac{1}{k\varepsilon} \rceil$. We find the neighbors of each vertex of Z in $\mathcal{C}$. Then for each vertex pair in Z, we check whether it is an edge in the graph or not. Loosely speaking, the set of edges between $\mathcal{C}$ and $V(Z)$ [4] will help us generate partial bipartitions, restricted to $X_i \cup V(Z)$'s, for each $i \in [t]$, and the edges among the pairs of vertices of Z will help us in estimating the bipartite distance of some *specific kind* of bipartitions of G. Here we would like to note that no further query will be performed by the algorithm. The set of edges with one vertex in $\mathcal{C}$ and the other in $V(Z)$, and the set of edges among the vertex pairs in Z, when treated in a *specific* manner, will give us the desired result. Observe that the number of adjacency queries performed by our algorithm is $\mathcal{O}(\frac{1}{k^8 \varepsilon^3} \log^2 \frac{1}{k\varepsilon})$.

For each $i \in [t]$, we do the following. We consider all possible bipartitions $\mathcal{F}_i$ of X_i. For each bipartition f_{ij} (of X_i) in $\mathcal{F}_i$, we extend f_{ij} to a bipartition of $X_i \cup V(Z)$, say f'_{ij}, such that both f_{ij} and f'_{ij} are identical with respect to X_i. Moreover, we assign $f'_{ij}(z)$ (to either L or R), for each $z \in V(Z) \setminus X_i$, based on the neighbors of z in X_i. To design a rule of assigning $f'_{ij}(z)$, for each $z \in V(Z) \setminus X_i$ for our purpose, we define the notions of *heavy* and *balanced* vertices, with respect to a bipartition (see Definitions 4.1 and 4.2). Heavy and balanced vertices are defined in such a manner that when the bipartite distance of G is at most εn^2 (that is, G is ε-close), we can infer the following interesting connections. Let f be a bipartition of $V(G)$ such that $d_{bip}(G, f) \leq \varepsilon n^2$. We will prove that the total number of edges, with no endpoints in X_i and whose at least one endpoint is a balanced vertex with respect to f, is bounded (see Claim 4.12). Moreover, if we generate a bipartition f' such that

---

[4] Recall that $V(Z)$ denotes the set of vertices present in at least one pair in Z.

f and f' differ for *large* number of heavy vertices, then the bipartite distance with respect to f' cannot be small. To guarantee the correctness of our algorithm, we will prove that a heavy vertex v with respect to f, can be detected and $f(v)$ can be determined, with probability at least $1 - o(k\varepsilon)$. Note that the testing of being a heavy vertex will be performed only for the vertices in $V(Z)$. We will see shortly how this will help us to guarantee the completeness of our algorithm.

Finally, our algorithm computes ζ_{ij}, that is, the fraction of vertex pairs in Z that are monochromatic [5] edges with respect to f'_{ij}. If we find at least one i and j such that $\zeta_{ij} \leq \left(2 + \frac{k}{20}\right)\varepsilon$, the algorithm decides that $d_{bip}(G) \leq \varepsilon n^2$. Otherwise, it will report that $d_{bip}(G) \geq (2 + k)\varepsilon n^2$.

2.2   Completeness

Let us assume that the bipartite distance of G is at most εn^2, and let f be a bipartition of $V(G)$ that is optimal. Let us now focus on a particular $i \in [t]$, that is, an X_i. Since we are considering all possible bipartitions $\mathcal{F}_i$ of X_i, there exists a $f_{ij} \in \mathcal{F}_i$, such that f_{ij} and f are identical with respect to X_i. To complete our argument, we introduce (in Definition 4.3) the notion of SPECIAL bipartition $\mathrm{SPL}_i^f : V(G) \to \{L, R\}$, with respect to f by f_{ij} such that $f(v)$, $f_{ij}(v)$ and $\mathrm{SPL}_i^f(v)$ are identical for each $v \in X_i$, and at least $1 - o(k\varepsilon)$ fraction of heavy vertices, with respect to f, are mapped identically both by f and SPL_i^f. We shall prove that the bipartite distance of G with respect to SPL_i^f is at most $\left(2 + \frac{k}{50}\right)\varepsilon n^2$ (see Lemma 4.6). Now let us think of generating a bipartition f''_{ij} of $V(G)$ such that, for each $v \in V(G) \setminus X_i$, if we determine $f''_{ij}(v)$ by the same rule used by our algorithm to determine $f_{ij}(z)$, for each $z \in V(Z) \setminus X_i$. Note that our algorithm does not find f''_{ij} explicitly, it is used only for the analysis purpose. The number of heavy vertices, with respect to the bipartition f, that have different mappings by f and f''_{ij}, is at most $o(k\varepsilon n)$ with constant probability. So, with a constant probability, f''_{ij} is a SPECIAL bipartition with respect to f by f_{ij}. Note that, if we take $|Z| = \mathcal{O}(\frac{1}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon})$ many random vertex pairs and determine the fraction χ_{ij}^f of pairs that form monochromatic edges with respect to the SPECIAL bipartition f''_{ij}, we can show that $\chi_{ij}^f \leq (2 + \frac{k}{20})\varepsilon$, with probability at least $1 - 2^{-\Omega(\frac{1}{k^3\varepsilon} \log \frac{1}{k\varepsilon})} \geq \frac{9}{10}$. However, we are not finding either f''_{ij} or χ_{ij}^f explicitly. We just find ζ_{ij}, that is, the fraction of vertex pairs in Z that are monochromatic edges with respect to f'_{ij}. But the above argument still holds, since Z is chosen randomly and there exists a f''_{ij}, such that $f''_{ij}(z) = f'_{ij}(z)$, for each $z \in V(Z)$, and the probability distribution of ζ_{ij} is identical to that of χ_{ij}^f.

2.3   Soundness

Let us now consider the case when the bipartite distance of G is at least $(2+k)\varepsilon n^2$, and f be any bipartition of $V(G)$. To prove the soundness of our algorithm, we introduce the notion of DERIVED *bipartition* $\mathrm{DER}_i^f : V(G) \to \{L, R\}$ with respect to f by f_{ij} (see Definition 4.4), such that $f(v)$, $f_{ij}(v)$ and $\mathrm{DER}_i^f(v)$ are identical for each $v \in X_i$. Observe that the bipartite distance of G with respect to any DERIVED bipartition is at least $(2 + k)\varepsilon n^2$ as well. Similar to the discussion of the completeness, if we generate a bipartition f''_{ij} of $V(G)$, f''_{ij} will be a DERIVED bipartition, with respect to f by f_{ij}. If we take $|Z| = \mathcal{O}(\frac{1}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon})$ many random pairs of vertices and determine the fraction χ_{ij}^f of pairs that form monochromatic edges with respect to the DERIVED bipartition f''_{ij}, we can prove that $\chi_{ij}^f \leq \left(2 + \frac{k}{20}\right)\varepsilon$ holds, with

---

[5] An edge is said to be monochromatic with respect to f'_{ij} if both its endpoints have the same f'_{ij} values.

probability at most $2^{-\Omega(\frac{1}{k^3\varepsilon}\log\frac{1}{k\varepsilon})}$. We want to re-emphasize that we are not determining f''_{ij}, as well as χ^f_{ij} explicitly. The argument follows due to the facts that Z is chosen randomly and there exists an f''_{ij} such that $f'_{ij}(z) = f''_{ij}(z)$, for each $z \in V(Z)$, and the probability distribution of ζ_{ij} is identical to that of χ^f_{ij}. Using the union bound, we can say that the algorithm rejects with probability at least $\frac{9}{10}$.

3 Algorithm for Tolerant Bipartite Testing (Proof of Theorem 2.1)

In this section, we formalize the ideas discussed in Section 2, and prove Theorem 2.1.

Formal description of algorithm TOL-BIP-DIST(G, ε)

Step-1 Let C_1, C_2, C_3 be three suitably chosen large constants and $t := \lceil \log \frac{C_1}{k\varepsilon} \rceil$.
  (i) We start by generating t many subset of vertices $X_1, \ldots, X_t \subset V(G)$, each with $\lceil \frac{C_2}{k^3\varepsilon} \log \frac{1}{k\varepsilon} \rceil$ many vertices, sampled randomly without replacement [6].
  (ii) We sample $\lceil \frac{C_3}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon} \rceil$ many random pairs of vertices, with replacement, and denote those sampled pairs of vertices as Z. Note that $X_1, \ldots, X_t, Z$ are generated independent of each other.
  (iii) We find all the edges with one endpoint in $\mathcal{C} = X_1 \cup X_2 \cup \ldots X_t$ and the other endpoint in one of the vertices of $V(Z)$ [7], by performing $\mathcal{O}\left(\frac{1}{k^8\varepsilon^3} \log^2 \frac{1}{k\varepsilon}\right)$ many adjacency queries.
Step-2 (i) Let $\{a_1, b_1\}, \ldots \{a_\lambda, b_\lambda\}$ be the pairs of vertices of Z, where $\lambda = \lceil \frac{C_3}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon} \rceil$. Now we find the pairs of Z that are edges in G, by performing adjacency queries to all the pairs of vertices of Z (after this step, the algorithm does not make any query further).
  (ii) For each $i \in [t]$, we do the following:
    (a) Let $\mathcal{F}_i$ denote the set of all possible bipartitions of X_i, that is,
    $$\mathcal{F}_i = \left\{ f_{ij} : X_i \to \{L, R\} : j \in \left[2^{|X_i|-2}\right] \right\}.$$

    (b) For each bipartition f_{ij} (of X_i) in $\mathcal{F}_i$, we extend f_{ij} to $f'_{ij} : X_i \cup Z \to \{L, R\}$ to be a bipartition of $X_i \cup Z$, such that the mapping of each vertex of X_i are identical in f_{ij} and f'_{ij}, and is defined as follows:

    $$f'_{ij}(z) = \begin{cases} f_{ij}(z), & z \in X_i \\ L, & z \notin X_i \text{ and } \left|N(z) \cap f_{ij}^{-1}(R)\right| > \left|N(z) \cap f_{ij}^{-1}(L)\right| + \frac{k\varepsilon|X_i|}{225000} \\ R, & z \notin X_i \text{ and } \left|N(z) \cap f_{ij}^{-1}(L)\right| > \left|N(z) \cap f_{ij}^{-1}(R)\right| + \frac{k\varepsilon|X_i|}{225000} \\ L \text{ or } R \text{ arbitrarily}, & \text{otherwise} \end{cases}$$

    Note that this step can be performed from the adjacency information between the vertices of $\mathcal{C}$ and Z, which have already been computed before.
    (c) We now find the fraction of the vertex pairs of Z that are edges and have the same label with respect to f'_{ij}, that is,

    $$\zeta_{ij} = 2 \cdot \frac{\left|\{\{a_\ell, b_\ell\} : \ell \in [\lambda], \{a_\ell, b_\ell\} \in E(G) \text{ and } f'_{ij}(a_\ell) = f'_{ij}(b_\ell)\}\right|}{\lambda} \; [8].$$

    (d) If $\zeta_{ij} \leq \left(2 + \frac{k}{20}\right)\varepsilon$, we ACCEPT G as ε-close to being bipartite, and QUIT the algorithm.

---

[6] Since we are assuming n is sufficiently large with respect to $\frac{1}{\varepsilon}$, sampling with and without replacement are the same.
[7] Recall that $V(Z)$ denotes the set of vertices present in at least one pair in Z.

(iii) If we arrive at this step, then $\zeta_{ij} > \left(2 + \frac{k}{20}\right)\varepsilon$, for each $i \in [t]$ and $f_{ij} \in \mathcal{F}_i$ in Step-(ii). We REJECT and declare that G is $(2+k)\varepsilon$-far from being bipartite.

We split the analysis of algorithm TOL-BIP-DIST(G, ε) into five parts:

Completeness: If G is ε-close to being bipartite, then TOL-BIP-DIST(G, ε) reports the same, with probability at least $\frac{9}{10}$.

Soundness: If G is $(2+k)\varepsilon$-far from being bipartite, then TOL-BIP-DIST(G, ε) reports the same, with probability at least $\frac{9}{10}$.

Sample Complexity: The sample complexity of TOL-BIP-DIST(G, ε) is $\mathcal{O}(\frac{1}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon})$.

Query Complexity: The query complexity of TOL-BIP-DIST(G, ε) is $\mathcal{O}(\frac{1}{k^8\varepsilon^3} \log^2 \frac{1}{k\varepsilon})$.

Time Complexity: The time complexity of TOL-BIP-DIST(G, ε) is $2^{\mathcal{O}(\frac{1}{k^3\varepsilon} \log \frac{1}{k\varepsilon})}$.

The last three quantities can be computed from the description of TOL-BIP-DIST(G, ε). In **Step-1(i)**, we sample vertices of G to generate $t = \lceil \log \frac{C_1}{k\varepsilon} \rceil$ subsets, each with $\lceil \frac{C_2}{k^3\varepsilon} \log \frac{1}{k\varepsilon} \rceil$ many vertices. Thereafter in **Step-1(ii)** and **Step-1(iii)**, we randomly choose $\lceil \frac{C_3}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon} \rceil$ many pairs of vertices and perform adjacency queries for each vertex in any pair of Z to every X_i. Thus the sample complexity of TOL-BIP-DIST(G, ε) is $\mathcal{O}(\frac{1}{k^5\varepsilon^2} \log \frac{1}{k\varepsilon})$ and query complexity is $\mathcal{O}(\frac{1}{k^8\varepsilon^3} \log^2 \frac{1}{k\varepsilon})$. The time complexity of the algorithm is $2^{\mathcal{O}(\frac{1}{k^3\varepsilon} \log \frac{1}{k\varepsilon})}$, which follows from **Step-2(ii)**, that dominates the running time.

4 Proof of Correctness of TOL-BIP-DIST(G, ε)

Before proceeding to the proof, we introduce some definitions for classifying the vertices of the graph, with respect to any particular bipartition, into two categories:

  (i) **heavy** vertex, and
  (ii) **balanced** vertex.

These definitions will be mostly used in the proof of completeness. Informally speaking, a vertex v is said to be **heavy** with respect to a bipartition f, if it has *substantially* large number of neighbors in one side of the bipartition (either L or R), as compared to the other side.

▶ **Definition 4.1** (Heavy vertex). *A vertex $v \in V$ is said to be L-**heavy** with respect to a bipartition f, if it satisfies two conditions:*

  (i) $\left| N(v) \cap f^{-1}(L) \right| \geq \left| N(v) \cap f^{-1}(R) \right| + \frac{k\varepsilon n}{150}$;
  (ii) *If* $\left| N(v) \cap f^{-1}(R) \right| \geq \frac{1}{(1+\frac{k}{200})} \frac{k\varepsilon n}{150}$, *then*

$$\left| N(v) \cap f^{-1}(L) \right| \geq (1 + \tfrac{k}{200}) \left| N(v) \cap f^{-1}(R) \right|;$$

*We define R-**heavy** vertices analogously. The union of the set of L-**heavy** and R-**heavy** vertices, with respect to a bipartition f, is defined to be the set of heavy vertices (with respect to f), and is denoted by $\mathcal{H}_f$.*

*Similarly, a vertex v is said to be **balanced** if the number of neighbors of v are similar in both L and R, with respect to a bipartition f. We define it formally as follows:*

▶ **Definition 4.2** (Balanced vertex). *A vertex $v \in V$ is said to be **balanced** with respect to a bipartition f, if $v \notin \mathcal{H}_f$, that is, it satisfies at least one of the following conditions:*
*(i) **Type 1:** $\left| \left| N(v) \cap f^{-1}(R) \right| - \left| N(v) \cap f^{-1}(L) \right| \right| < \frac{k\varepsilon n}{150}$;*

(ii) Type 2: Either

$$\left| N(v) \cap f^{-1}(L) \right| \leq \left| N(v) \cap f^{-1}(R) \right| < (1 + \frac{k}{200}) \left| N(v) \cap f^{-1}(L) \right|,$$

or,

$$\left| N(v) \cap f^{-1}(R) \right| \leq \left| N(v) \cap f^{-1}(L) \right| < (1 + \frac{k}{200}) \left| N(v) \cap f^{-1}(R) \right|.$$

The set of balanced vertices of type 1 with respect to f is denoted as $\mathcal{B}_f^1$, and the set of balanced vertices of type 2 with respect to f is denoted as $\mathcal{B}_f^2$. The union of $\mathcal{B}_f^1$ and $\mathcal{B}_f^2$ is denoted by $\mathcal{B}_f$. Note that $\mathcal{B}_f^1$ and $\mathcal{B}_f^2$ may not be disjoint.

In order to prove the completeness (in Section 4.1), we also use a notion of SPECIAL *bipartition* to be defined below. The definition of SPECIAL bipartition is based on an optimal bipartition f of $V(G)$, and notions of heavy and balanced vertices. We would also like to note that, later in Lemma 4.6, we show that when $d_{bip}(G) \leq \varepsilon n^2$, the bipartite distance of G with respect to any SPECIAL bipartition is bounded by $(2 + \frac{k}{50})\varepsilon n^2$.

▶ **Definition 4.3** (SPECIAL bipartition). *Let $d_{bip}(G) \leq \varepsilon n^2$, and $f : V(G) \to \{L, R\}$ be an optimal bipartition of $V(G)$, that is, $d_{bip}(G, f) \leq \varepsilon n^2$, and there does not exist any bipartition g such that $d_{bip}(G, g) < d_{bip}(G, f)$. For an X_i selected in **Step-1(i)** of the algorithm, let $f_{ij} \in \mathcal{F}_i$ be the bipartition of X_i such that $f \mid_{X_i} = f_{ij}$. Then bipartition $\mathrm{SPL}_i^f : V(G) \to \{L, R\}$ is said to be a SPECIAL **bipartition** with respect to f by f_{ij} such that*
- *$\mathrm{SPL}_i^f \mid_{X_i} = f \mid_{X_i} = f_{ij}$;*
- *There exists a subset $\mathcal{H}_f' \subset \mathcal{H}_f$ such that $\left| \mathcal{H}_f' \right| \geq (1 - o(k\varepsilon)) \left| \mathcal{H}_f \right|$, and for each $v \in \mathcal{H}_f'$, $\mathrm{SPL}_i^f(v)$ is defined as follows:*

$$\mathrm{SPL}_i^f(v) = \begin{cases} R, & v \notin X_i \text{ and } v \text{ is } L - heavy \\ L, & v \notin X_i \text{ and } v \text{ is } R - heavy \end{cases}$$

- *For each $v \notin (\mathcal{H}_f' \cup X_i)$, $\mathrm{SPL}_i^f(v)$ is set to L or R arbitrarily.*

In our proof of the soundness theorem (in Section 4.2), we need the notion of DERIVED *bipartition*. Unlike the definition of SPECIAL bipartition, the definition of DERIVED bipartition is more general, in the sense that it is not defined based on either any optimal bipartition, or on heavy or balanced vertices.

▶ **Definition 4.4** (DERIVED bipartition). *Let $f : V(G) \to \{L, R\}$ be a bipartition of $V(G)$. For an X_i selected in **Step-1(i)** of the algorithm, let $f_{ij} \in \mathcal{F}_i$ be the bipartition of X_i such that $f \mid_{X_i} = f_{ij}$. A bipartition $\mathrm{DER}_i^f : V(G) \to \{L, R\}$ is said to be DERIVED bipartition with respect to f by f_{ij}, if $\mathrm{DER}_i^f \mid_{X_i} = f \mid_{X_i} = f_{ij}$.*

4.1 Proof of Completeness

In this section, we prove the following theorem:

▶ **Theorem 4.5.** *Let us assume G is ε-close to being bipartite. Then $\mathrm{TOL\text{-}BIP\text{-}DIST}(G, \varepsilon)$ reports the same, with probability at least $9/10$.*

The proof of Theorem 4.5 will cricially use the following lemma, which says that the bipartite distance of G with respect to any SPECIAL bipartition is bounded by a $\left(2 + \frac{k}{50} \right) \varepsilon n^2$.

▶ **Lemma 4.6** (SPECIAL bipartition lemma). *Let f be a bipartition such that $d_{bip}(G, f) \leq \varepsilon n^2$ and there does not exist any bipartition g such that $d_{bip}(G, g) < d_{bip}(G, f)$. For any SPECIAL bipartition SPL_i^f with respect to f, we have*

$$d_{bip}(G, \mathrm{SPL}_i^f) \leq \left(2 + \frac{k}{50}\right) \varepsilon n^2.$$

We will prove the above lemma later. For now, we want to establish (in Lemma 4.8) that there exists an $i \in [t]$ and a $f_{ij} \in \mathcal{F}_i$ which can be thought of as a *random restriction* of some SPECIAL bipartition with respect to f by f_{ij}. In other words, Lemma 4.8 basically states that if G is ϵ-close to being bipartite, then the extension according to the rule in **Step-2(ii)(b)** of the mapping obtained by restricting an optimal bipartition to a random X_i is likely to correspond to a SPECIAL bipartition, and therefore, the number of monochromatic edges (with respect to a SPECIAL bipartition) in the randomly picked Z is likely to be low with respect to that bipartition. Thus, ζ_{ij} must be low for some i, j with high probability.

To prove Lemma 4.8, we need the following lemma (Lemma 4.7) about heavy vertices. In Lemma 4.7, we basically prove that a heavy vertex with respect to a bipartition f will have significantly more neighbors in the part of X_i, that corresponds to the heavy side of that vertex (with respect to f). Basically, if a vertex v is L-heavy with respect to f, it has more neighbors in the subset of X_i on the L-side as compared to the subset of X_i on the R-side of f.

▶ **Lemma 4.7** (Heavy vertex lemma). *Let f be a bipartition of G. Consider a vertex $v \in V$.*
 (i) *For every L-heavy vertex v, $\left|N(v) \cap f^{-1}(L) \cap X_i\right| - \left|N(v) \cap f^{-1}(R) \cap X_i\right| \geq \frac{k^2 \varepsilon |X_i|}{225000}$ with probability at least $1 - o(k\varepsilon)$.*
 (ii) *For every R-heavy vertex v, $\left|N(v) \cap f^{-1}(L) \cap X_i\right| - \left|N(v) \cap f^{-1}(R) \cap X_i\right| \geq \frac{k^2 \varepsilon |X_i|}{225000}$ with probability at least $1 - o(k\varepsilon)$.*

We would like to note that Lemma 4.7 holds for any bipartition. However, we will use it only for completeness with resepct to an optimal bipartition f.

▶ **Lemma 4.8.** *If $d_{bip}(G) \leq \varepsilon n^2$, then there exists an $i \in [t]$ and $f_{ij} \in \mathcal{F}_i$ such that $\zeta_{ij} \leq \left(2 + \frac{k}{20}\right) \varepsilon$ holds, with probability at least $1 - o(k\varepsilon)$.*

Proof. Let f be an optimal bipartition such that $d_{bip}(G, f) \leq \varepsilon n^2$. First, consider a SPECIAL bipartition SPL_i^f, and consider a set of random vertex pairs Y such that $|Y| = |Z|$. Now consider the fraction of monochromatic edges of Y, with respect to the bipartition SPL_i^f, that is,

$$\chi_{ij}^f = 2 \cdot \frac{\left|\left\{\{a, b\} \in Y : \{a, b\} \in E(G) \text{ and } \mathrm{SPL}_i^f(a) = \mathrm{SPL}_i^f(b)\right\}\right|}{|Y|}.$$

▶ **Observation 4.9.** *$\chi_{ij}^f \leq \left(2 + \frac{k}{20}\right) \varepsilon$ holds, with probability at least $9/10$.*

Proof. By Lemma 4.6, we know that when $d_{bip}(G) \leq \varepsilon n^2$, $d_{bip}(G, \mathrm{SPL}_i^f) \leq \left(2 + \frac{k}{50}\right) \varepsilon n^2$. So, $\mathbb{E}[\chi_{ij}^f] \leq \left(2 + \frac{k}{50}\right) \varepsilon$. Using Chernoff bound (see Lemma B.1), we can say that

$$\mathbb{P}(\chi_{ij}^f \geq \left(2 + \tfrac{k}{20}\right) \varepsilon) \leq \frac{1}{2^{\Omega\left(\frac{1}{k^3 \varepsilon} \log \frac{1}{k\varepsilon}\right)}} \leq \frac{1}{10}. \qquad \blacktriangleleft$$

Now, we claim that bounding χ_{ij}^f is equivalent to bounding ζ_{ij}.

▷ **Claim 4.10.**   For any $i \in [t]$, there exists a bipartition $f_{ij} \in \mathcal{F}_i$ such that the probability distribution of ζ_{ij} is identical to that of χ_{ij}^f, for some SPECIAL bipartition f with respect to f_{ij}, with probability at least $1/2$.

As $t = \mathcal{O}(\log \frac{1}{k\varepsilon})$, the above claim implies that there exists an $i \in [t]$ and $f_{ij} \in \mathcal{F}_i$ such that the probability distribution of ζ_{ij} is identical to that of χ_{ij}^f, with probability at least $1 - o(k\varepsilon)$.

Now we prove Claim 4.10. Recall the procedure of determining ζ_{ij} as described in **Step 2** of algorithm TOL-BIP-DIST(G, ε) presented in Section 3.

Fact 1: For any vertex $v \in \mathcal{H}_f \cap Z$, $\text{SPL}_i^f(v) = f'_{ij}(v)$, with probability at least $1 - o(k\varepsilon)$, where $\mathcal{H}_f$ denotes the set of heavy vertices of X_i with respect to the bipartition f. This follows according to Claim 4.7, along with the definition of $f'_{ij}(z)$.

Fact 2: Consider a bipartition $f_{ij} \in \mathcal{F}_i$ of X_i, and its extension f'_{ij} to $X_i \cup Z$, as considered in the algorithm. Assume a bipartition f''_{ij} of $V(G)$, constructed by extending f'_{ij} according to the rule of **Step-2(ii)(b)** of the algorithm. From Heavy vertex lemma (Lemma 4.7), we know that the expected number of vertices in $\mathcal{H}_f$ such that $f''_{ij}(v) \neq f(v)$, is at most $o(k\varepsilon) |\mathcal{H}_f|$. Using Markov inequality, we can say that, with probability at least $\frac{1}{2}$, the number of vertices in $\mathcal{H}_f$ such that $f''_{ij}(v) \neq f(v)$, is at most $o(k\varepsilon) |\mathcal{H}_f|$. Thus, with probability at least $\frac{1}{2}$, there exists a set of vertices $\mathcal{H}'_f$ such that $f''_{ij}(v) = f(v)$ holds for at least $(1 - o(k\varepsilon)) |\mathcal{H}'_f|$ vertices. Note that the bipartition f''_{ij} is a SPECIAL bipartition f with respect to f_{ij}.

From **Fact 1** and **Fact 2**, we can deduce that, there exists a SPECIAL bipartition SPL_i^f such that $\text{SPL}_i^f(v) = f'_{ij}(v)$ for each $z \in Z$.

Since we choose Z uniformly at random, Lemma 4.8 follows. ◀

According to the description of algorithm TOL-BIP-DIST(G, ε), the algorithm reports that $d_{bip}(G) \leq \varepsilon n^2$, if there exists a ζ_{ij} such that $\zeta_{ij} \leq \left(2 + \frac{k}{20}\right)\varepsilon$, for some $i \in [t]$ and $j \in [2^{|X_i|-2}]$. Hence, by Lemma 4.8, we are done with the proof of the completeness theorem (Theorem 4.5).

Now we focus on proving SPECIAL bipartition lemma (Lemma 4.6) and Heavy vertex lemma (Lemma 4.7), starting with the proof of SPECIAL bipartition lemma.

Proof of SPECIAL bipartition lemma (Lemma 4.6)

The idea of the proof relies on decomposing the bipartite distance with respect to a SPECIAL bipartition into a sum of three terms and then carefully bounding the cost of each of those parts individually.

Let us first recall the definition of bipartite distance of G with respect to a special bipartition SPL_i^f.

$$d_{bip}(G, \text{SPL}_i^f) = \left| \left\{ (u, v) \in E(G) : \text{SPL}_i^f(u) = \text{SPL}_i^f(v) \right\} \right|. \tag{1}$$

By abuse of notation, here we are denoting $E(G)$ as the set of ordered edges.

We will upper bound $d_{bip}(G, \text{SPL}_i^f)$ as the sum of three terms defined below. Here $\mathcal{H}_f$ and $\mathcal{B}_f$ denote the set of heavy vertices and balanced vertices (with respect to f), as defined in Definition 4.1 and Definition 4.2, respectively. Also, $\mathcal{H}'_f \subseteq \mathcal{H}_f$ denotes the set of vertices of $\mathcal{H}_f$ that are mapped according to f, as defined in the definition of SPECIAL bipartition in Definition 4.3. The three terms that are used to upper bound $d_{bip}(G, \text{SPL}_i^f)$ are as follows:

(a) $D_{\mathcal{H}'_f \cup X_i, \mathcal{H}'_f \cup X_i} = \left| \left\{ (u,v) \in E(G) : u \in \mathcal{H}'_f \cup X_i \;\&\; v \in \mathcal{H}'_f \cup X_i, \mathrm{SPL}_i^f(u) = \mathrm{SPL}_i^f(v) \right\} \right|$;

(b) $D_{\mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i), V(G)} =$
$$\left| \left\{ (u,v) \in E(G) : u \in \mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i), \;\&\; v \in V(G), \mathrm{SPL}_i^f(u) = \mathrm{SPL}_i^f(v) \right\} \right|;$$

(c) $D_{\mathcal{B}_f \setminus X_i, V(G)} = \left| \left\{ (u,v) \in E(G) : u \in \mathcal{B}_f \setminus X_i \text{ and } v \in V(G), \mathrm{SPL}_i^f(u) = \mathrm{SPL}_i^f(v) \right\} \right|$.

Now from Equation 1 along with the above definitions, we can upper bound $d_{bip}(G, \mathrm{SPL}_i^f)$ as follows:

$$d_{bip}(G, \mathrm{SPL}_i^f) \leq D_{\mathcal{H}'_f \cup X_i, \mathcal{H}'_f \cup X_i} + D_{\mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i), V(G)} + D_{\mathcal{B}_f \setminus X_i, V(G)}. \tag{2}$$

We now upper bound $d_{bip}(G, \mathrm{SPL}_i^f)$ by bounding each term on the right hand side of the above expression separately, via the two following claims which we will prove later.

▷ **Claim 4.11.**

(i) $D_{\mathcal{H}'_f \cup X_i, \mathcal{H}'_f \cup X_i} \leq d_{bip}(G, f) - \Pi$, where

$$\Pi := \left[\sum_{v \in \mathcal{B}_f \setminus X_i : f(v) = L} |N(v) \cap f^{-1}(L)| + \sum_{v \in \mathcal{B}_f \setminus X_i : f(v) = R} |N(v) \cap f^{-1}(R)| \right];$$

(ii) $D_{\mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i), V(G)} \leq o(k\varepsilon)n^2$;

▷ **Claim 4.12.** $D_{\mathcal{B}_f \setminus X_i, V(G)} \leq 2 \left(1 + \frac{k}{400} \right) \Pi + \frac{k\varepsilon n^2}{150}$.

Assuming Claim 4.11 and Claim 4.12 hold, along with Equation 2, $d_{bip}(G, \mathrm{SPL}_i^f)$ can be upper bounded as follows:

$$
\begin{aligned}
d_{bip}(G, \mathrm{SPL}_i^f) &\leq d_{bip}(G, f) - \Pi + o(k\varepsilon)n^2 + 2\left(1 + \frac{k}{400}\right)\Pi + \frac{k\varepsilon n^2}{150} \\
&\leq d_{bip}(G, f) + \Pi + \frac{k}{200}\Pi + \frac{k\varepsilon n^2}{100}.
\end{aligned}
$$

Note that $\Pi \leq d_{bip}(G, f)$ and $d_{bip}(G, f) \leq \varepsilon n^2$. Hence, we can say the following:

$$d_{bip}(G, \mathrm{SPL}_i^f) \leq \left(2 + \frac{k}{50} \right) \varepsilon n^2.$$

So, we are done with the proof of the SPECIAL bipartition lemma. We are left with the proofs of Claim 4.11 and Claim 4.12.

Proof of Claim 4.11. (i) We use the following observation in our proof. The observation follows due to the fact that the bipartition f considered is an optimal bipartition.

▶ **Observation 4.13.** *Let v be a L-heavy vertex v with respect to f. Then $f(v) = R$. Similarly, for every R-heavy vertex v with respect to f, $f(v) = L$.*

Following the definition of SPECIAL bipartition, we know that there exists a set of vertices $\mathcal{H}'_f \subset \mathcal{H}_f$ such that $\left| \mathcal{H}'_f \right| \geq (1 - o(k\varepsilon)) |\mathcal{H}_f|$, and for each $v \in \mathcal{H}'_f$, the following holds:

$$\mathrm{SPL}_i^f(v) = \begin{cases} R, & v \notin X_i \text{ and } v \text{ is } L - \text{heavy} \\ L, & v \notin X_i \text{ and } v \text{ is } R - \text{heavy} \end{cases}$$

By Observation 4.13, we know that for every $v \in \mathcal{H}'_f$, $\mathrm{SPL}_i^f(v) = f(v)$. Moreover, for each $v \in X_i$, $\mathrm{SPL}_i^f(v) = f(v)$, following the definition of SPECIAL bipartition SPL_i^f. Thus, for every $v \in \mathcal{H}'_f \cup X_i$, we have $\mathrm{SPL}_i^f(v) = f(v)$. Hence,

$$D_{\mathcal{H}'_f \cup X_i, \mathcal{H}'_f \cup X_i}$$
$$= \left| \left\{ (u,v) \in E(G) : u \in \mathcal{H}'_f \cup X_i \text{ and } v \in \mathcal{H}'_f \cup X_i, \mathrm{SPL}_i^f(u) = \mathrm{SPL}_i^f(v) \right\} \right|$$
$$= \left| \left\{ (u,v) \in E(G) : u \in \mathcal{H}'_f \cup X_i, \text{ and } v \in \mathcal{H}'_f \cup X_i, f(u) = f(v) \right\} \right|$$
$$\left(\because \text{ for every } v \in \mathcal{H}'_f \cup X_i, \mathrm{SPL}_i^f(v) = f(v) \right)$$
$$= d_{bip}(G, f) -$$
$$\left[\sum_{v \in V \setminus (\mathcal{H}'_f \cup X_i) : f(v) = L} |N(v) \cap f^{-1}(L)| + \sum_{v \in V \setminus (\mathcal{H}'_f \cup X_i) : f(v) = R} |N(v) \cap f^{-1}(R)| \right]$$
$$\leq d_{bip}(G, f) - \left[\sum_{v \in \mathcal{B}_f \setminus X_i : f(v) = L} |N(v) \cap f^{-1}(L)| + \sum_{v \in \mathcal{B}_f \setminus X_i : f(v) = R} |N(v) \cap f^{-1}(R)| \right]$$
$$= d_{bip}(G, f) - \Pi.$$

(ii) By the definition of $\mathcal{H}'_f$, we know that $\left| \mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i) \right|$ is upper bounded by $o(k\varepsilon) |\mathcal{H}_f|$. Following the definition of $D_{\mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i), V(G)}$, we can say the following:

$$D_{\mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i), V(G)}$$
$$= \left| \left\{ (u,v) \in E(G) : u \in \mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i) \text{ and } v \in V(G), \mathrm{SPL}_i^f(u) = \mathrm{SPL}_i^f(v) \right\} \right|$$
$$\leq \left| \mathcal{H}_f \setminus (\mathcal{H}'_f \cup X_i) \right| \times |V(G)| = o(k\varepsilon) |\mathcal{H}_f| \times n \leq o(k\varepsilon) n^2.$$

The last inequality follows as $|\mathcal{H}_f|$ is at most n. $\qquad \triangleleft$

Proof of Claim 4.12. Observe that

$$D_{\mathcal{B}_f \setminus X_i, V(G)}$$
$$= \left| \left\{ (u,v) \in E(G) : u \in \mathcal{B}_f \setminus X_i \text{ and } v \in V(G), \mathrm{SPL}_i^f(u) = \mathrm{SPL}_i^f(v) \right\} \right|$$
$$\leq |\{ (u,v) \in E(G) : u \in \mathcal{B}_f \setminus X_i \text{ and } v \in V(G) \}| = \sum_{v \in \mathcal{B}_f \setminus X_i} |N(v)|$$

As $\mathcal{B}_f = \mathcal{B}_f^1 \cup \mathcal{B}_f^2$, we have

$$D_{\mathcal{B}_f \setminus X_i, V(G)} \leq \sum_{v \in \mathcal{B}_f^1 \setminus X_i} |N(v)| + \sum_{v \in \mathcal{B}_f^2 \setminus X_i} |N(v)|. \tag{3}$$

We will bound $D_{\mathcal{B}_f \setminus X_i, V(G)}$ by bounding $\sum_{v \in \mathcal{B}_f^1 \setminus X_i} |N(v)|$ and $\sum_{v \in \mathcal{B}_f^2 \setminus X_i} |N(v)|$ separately, which we prove in the following claim:

$\triangleright$ Claim 4.14. Consider T_1 and T_2 defined as follows:

$$T_1 = 2 \left(\sum_{v \in f^{-1}(L) \cap (\mathcal{B}_f^1 \setminus X_i)} \left| N(v) \cap f^{-1}(L) \right| + \sum_{v \in f^{-1}(R) \cap (\mathcal{B}_f^1 \setminus X_i)} \left| N(v) \cap f^{-1}(R) \right| \right) + \frac{k\varepsilon n^2}{150},$$

$$T_2 = \left(2 + \frac{k}{200}\right) \left(\sum_{v \in f^{-1}(L) \cap (\mathcal{B}_f^2 \setminus X_i)} \left|N(v) \cap f^{-1}(L)\right| + \sum_{v \in f^{-1}(R) \cap (\mathcal{B}_f^2 \setminus X_i)} \left|N(v) \cap f^{-1}(R)\right|\right).$$

Then

(i) For balanced vertices of **Type 1**, we have

$$\sum_{v \in \mathcal{B}_f^1 \setminus X_i} |N(v)| \le T_1$$

(ii) For balanced vertices of **Type 2**, we have

$$\sum_{v \in \mathcal{B}_f^2 \setminus X_i} |N(v)| \le T_2$$

See the full version of the paper [10] for the proof of the above claim. Using Claim 4.14 and Equation (3), we get

$$D_{\mathcal{B}_f \setminus X_i, V(G)}$$
$$= \sum_{v \in \mathcal{B}_f^1 \setminus X_i} |N(v)| + \sum_{v \in \mathcal{B}_f^2 \setminus X_i} |N(v)|$$
$$\le T_1 + T_2$$
$$\le 2\left(1 + \frac{k}{400}\right)\Pi + \frac{k\varepsilon n^2}{150}$$

The last inequality follows from the definitions of T_1, T_2 and Π. $\triangleleft$

Proof of Heavy vertex lemma (Lemma 4.7)

Before proceeding to prove the Heavy vertex lemma, we will first prove two intermediate claims that will be crucially used in the proof of the lemma. The first claim states that when we consider a bipartition f of G, if a vertex $v \in G$ has a *large* number of neighbors on one side of the partition defined by f, the proportion of its neighbors in X_i on the same side of f will be approximately preserved, where X_i is a set of vertices picked at random in **Step-1(i)** of the algorithm TOL-BIP-DIST(G, ε). The result is formally stated as follows:

▷ **Claim 4.15.** Let f be a bipartition of G. Consider a vertex $v \in V$.
 (i) Suppose $\left|N(v) \cap f^{-1}(L)\right| \ge \frac{k\varepsilon n}{150}$. Then, with probability at least $1 - o(k\varepsilon)$, we have

$$\left|N(v) \cap f^{-1}(L) \cap X_i\right| = \left(1 \pm \frac{k}{500}\right)\left|N(v) \cap f^{-1}(L)\right| \frac{|X_i|}{n}.$$

 (ii) Suppose $\left|N(v) \cap f^{-1}(R)\right| \ge \frac{k\varepsilon n}{150}$. Then, with probability at least $1 - o(k\varepsilon)$, we have

$$\left|N(v) \cap f^{-1}(R) \cap X_i\right| = \left(1 \pm \frac{k}{500}\right)\left|N(v) \cap f^{-1}(R)\right| \frac{|X_i|}{n}.$$

The next claim is in similar spirit as that of Claim 4.15. Instead of considering vertices with large number of neighbors, it considers the case when a vertex has *small* number of neighbors on one side of a bipartition f.

▷ **Claim 4.16.** Let f be a bipartition of G. Consider a vertex $v \in V$.
 (i) Suppose $\left|N(v) \cap f^{-1}(L)\right| \le \frac{1}{1 + \frac{k}{200}} \frac{k\varepsilon n}{150}$. Then, with probability at least $1 - o(k\varepsilon)$, we have

$$\left|N(v) \cap f^{-1}(L) \cap X_i\right| \le \frac{1}{1 + \frac{k}{300}} \frac{k\varepsilon |X_i|}{150}.$$

(ii) Suppose $\left|N(v) \cap f^{-1}(R)\right| \leq \frac{1}{1+\frac{k}{200}} \frac{k\varepsilon n}{150}$. Then, with probability at least $1 - o(k\varepsilon)$, we have

$$\left|N(v) \cap f^{-1}(R) \cap X_i\right| \leq \frac{1}{1+\frac{k}{300}} \frac{k\varepsilon |X_i|}{150}.$$

Claim 4.15 and Claim 4.16 can be proved by using large deviation inequalities (stated in Appendix B), and the proofs can be found in the full version of the paper [10].

Assuming Claim 4.15 and Claim 4.16 hold, we now prove the Heavy vertex lemma (Lemma 4.7).

Proof of Lemma 4.7. We will only prove (i) here, which concerns the L-heavy vertices. (ii) can be proved in similar fashion.

We first characterize L-heavy vertices into two categories:

(a) Both $\left|N(v) \cap f^{-1}(L)\right|$ and $\left|N(v) \cap f^{-1}(R)\right|$ are large, that is, $\left|N(v) \cap f^{-1}(L)\right| \geq \frac{k\varepsilon n}{150}$ and $\left|N(v) \cap f^{-1}(R)\right| \geq \frac{1}{1+\frac{k}{200}} \frac{k\varepsilon n}{150}$. Also, $\left|N(v) \cap f^{-1}(L)\right| \geq \left(1 + \frac{k}{200}\right) \left|N(v) \cap f^{-1}(R)\right|$.

(b) $\left|N(v) \cap f^{-1}(L)\right|$ is large and $\left|N(v) \cap f^{-1}(R)\right|$ is small, that is, $\left|N(v) \cap f^{-1}(L)\right| \geq \frac{k\varepsilon n}{150}$ and $\left|N(v) \cap f^{-1}(R)\right| \leq \frac{1}{1+\frac{k}{200}} \frac{k\varepsilon n}{150}$.

Case (a): Here $\left|N(v) \cap f^{-1}(L)\right| \geq \left(1 + \frac{k}{200}\right) \frac{k\varepsilon n}{150}$, and $\left|N(v) \cap f^{-1}(R)\right| \geq \frac{k\varepsilon n}{150}$. From Claim 4.15, the following hold, with probability at least $1 - o(k\varepsilon)$:

$$\left|N(v) \cap f^{-1}(L) \cap X_i\right| = \left(1 \pm \frac{k}{500}\right) \left|N(v) \cap f^{-1}(L)\right| \frac{|X_i|}{n}$$

and

$$\left|N(v) \cap f^{-1}(R) \cap X_i\right| = \left(1 \pm \frac{k}{500}\right) \left|N(v) \cap f^{-1}(R)\right| \frac{|X_i|}{n}.$$

So, with probability at least $1 - o(k\varepsilon)$, we have the following
$$\left|N(v) \cap f^{-1}(L) \cap X_i\right| - \left|N(v) \cap f^{-1}(R) \cap X_i\right|$$
$$\geq \left(1 - \frac{k}{500}\right) \left|N(v) \cap f^{-1}(L)\right| \frac{|X_i|}{n} - \left(1 + \frac{k}{500}\right) \left|N(v) \cap f^{-1}(R)\right| \frac{|X_i|}{n}$$
$$\geq \left(1 - \frac{k}{500} - \frac{1 + \frac{k}{500}}{1 + \frac{k}{200}}\right) \frac{\left|N(v) \cap f^{-1}(L)\right| |X_i|}{n}$$
$$\left(\because \left|N(v) \cap f^{-1}(L)\right| \geq (1 + \frac{k}{200}) \left|N(v) \cap f^{-1}(R)\right|\right)$$
$$\geq \frac{k}{1500} \times \frac{k\varepsilon |X_i|}{150}$$
$$\geq \frac{k^2 \varepsilon |X_i|}{225000} \qquad\qquad (\because \ k \leq 100)$$

Case (b): Here $\left|N(v) \cap f^{-1}(L)\right| \geq \frac{k\varepsilon n}{150}$ and $\left|N(v) \cap f^{-1}(R)\right| \leq \left(\frac{1}{1+\frac{k}{200}}\right) \frac{k\varepsilon n}{150}$. So, from Claim 4.15 and Claim 4.16, the following hold, with probability at least $1 - o(k\varepsilon)$:

$$\left|N(v) \cap f^{-1}(L) \cap X_i\right| = \left(1 \pm \frac{k}{500}\right) \left|N(v) \cap f^{-1}(L)\right| \frac{|X_i|}{n}$$

and

$$\left|N(v) \cap f^{-1}(R) \cap X_i\right| \leq \frac{1}{1 + \frac{k}{300}} \frac{k\varepsilon |X_i|}{150}.$$

Thus, with probability at least $1 - o(k\varepsilon)$, we have the following:

$$\left|N(v) \cap f^{-1}(L) \cap X_i\right| - \left|N(v) \cap f^{-1}(R) \cap X_i\right|$$

$$\geq (1 - \frac{k}{500})\left|N(v) \cap f^{-1}(L)\right|\frac{|X_i|}{n} - \frac{1}{1 + \frac{k}{300}}\frac{k\varepsilon|X_i|}{150}$$

$$= (1 - \frac{k}{500})\frac{k\varepsilon|X_i|}{150} - \frac{1}{1 + \frac{k}{300}}\frac{k\varepsilon|X_i|}{150}$$

$$\geq \frac{1}{1500}\left(2k - \frac{k^2}{100}\right)\frac{k\varepsilon|X_i|}{150}$$

$$\geq \frac{k^2\varepsilon|X_i|}{225000} \qquad\qquad\qquad (\because\ k \leq 100)$$

This completes the proof of part (i) of Lemma 4.7. ◀

4.2 Proof of Soundness

In this section, we prove the following theorem:

▶ **Theorem 4.17.** *Let us assume that G is $(2 + k)\varepsilon$-far from being bipartite. Then* $\textsc{Tol-Bip-Dist}(G, \varepsilon)$ *reports the same, with probability at least* $9/10$.

Assume f be a bipartition of $V(G)$. Now let us consider a DERIVED bipartition $\textsc{Der}_i^f$ with respect to f by f_{ij}, and choose a set of random vertex pairs Y such that $|Y| = |Z|$. Let χ_{ij}^f denote the fraction of vertex pairs of Y that are monochromatic with respect to the bipartition $\textsc{Der}_i^f$, that is,

$$\chi_{ij}^f = 2 \cdot \frac{\left|\left\{\{a, b\} \in Y : \{a, b\} \in E(G) \text{ and } \textsc{Der}_i^f(a) = \textsc{Der}_i^f(b)\right\}\right|}{|Y|}.$$

▶ **Observation 4.18.** $\chi_{ij}^f \leq \left(2 + \frac{k}{20}\right)\varepsilon$ *holds with probability at most* $\frac{1}{10N}$, *where* $N = 2^{\mathcal{O}(\frac{1}{k^3\varepsilon}\log\frac{1}{k\varepsilon})}$.

Proof. Since G is $(2 + k)\varepsilon$-far from being bipartite, the same holds for the bipartition $\textsc{Der}_i^f$ as well, that is, $d_{bip}(G, \textsc{Der}_i^f) \geq (2 + k)\varepsilon n^2$. So, $\mathbb{E}[\chi_{ij}^f] \geq (2 + k)\varepsilon$. Using Chernoff bound (see Lemma B.1), we can say that, $\mathbb{P}\left(\chi_{ij}^f \leq \left(2 + \frac{k}{20}\right)\varepsilon\right) \leq \frac{1}{10N}$. Since $|Z| = \mathcal{O}\left(\frac{1}{k^5\varepsilon^2}\log\frac{1}{k\varepsilon}\right)$, the result follows. ◀

We will be done with the proof by proving the following claim, that says that bounding χ_{ij}^f is equivalent to bounding ζ_{ij}.

▷ **Claim 4.19.** For any $i \in [t]$, and any $f_{ij} \in \mathcal{F}_i$, the probability distribution of ζ_{ij} is identical to that of χ_{ij}^f for some DERIVED bipartition with respect to f by f_{ij}.

Proof. Consider a bipartition $f_{ij} \in \mathcal{F}_i$ of X_i, and the bipartition f_{ij}' of $X_i \cup Z$, constructed by extending f_{ij}, as described in the algorithm. For the sake of the argument, let us construct a new bipartition f_{ij}'' of $V(G)$ by extending the bipartition f_{ij}', following the same rule of **Step-2 (ii) (b)** of the algorithm. Observe that $f_{ij}''(v) = f_{ij}(v)$, for each $v \in X_i$. Thus f_{ij}'' is a DERIVED bipartition with respect to some f by f_{ij}. Hence, the claim follows according to the way we generate ζ_{ij}, along with the fact that Z is chosen uniformly at random by the algorithm in **Step-1 (ii)**. ◁

Let us now define a pair (X_i, f_{ij}), with $i \in [t]$ and $f_{ij} \in \mathcal{F}_i$ as a **configuration**. Now we make the following observation which follows directly from the description of the algorithm.

▶ **Observation 4.20.** *Total number of possible configurations is $N = 2^{\mathcal{O}(\frac{1}{k^3 \varepsilon} \log \frac{1}{k\varepsilon})}$.*

Note that Claim 4.19 holds for a particular $f_{ij} \in \mathcal{F}_i$. Recall that in **Step-2(iii)**, our algorithm ToL-Bip-Dist(G, ε) reports that G is $(2+k)\varepsilon$-far if $\zeta_{ij} > \left(2 + \frac{k}{20}\right)\varepsilon$, for all $i \in [t]$ and $f_{ij} \in \mathcal{F}_i$. So, using the union bound, along with Observation 4.18, Claim 4.19 and Observation 4.20, we are done with the proof of Theorem 4.17.

5   Conclusion

We believe that our result will certainly improve the current understanding of (tolerant) bipartite testing in the dense graph model. However, one may wonder whether the analysis can be improved to show that the algorithm (presented in Section 3) can decide whether $d_{bip}(G) \leq \varepsilon n^2$ or $d_{bip}(G) \geq c\varepsilon n^2$ for any $c > 1$. There is a bottleneck in our technique as we are bounding error due to the balanced vertices by the sum of degrees of the balanced vertices (as done in Claim 4.12). Because of this reason, it is not obvious if our algorithm (and its analysis) can be used to get a result, like Theorem 2.1, for all $c > 1$ with the same query complexity.

On a different note, we can decide $d_{bip}(G) \leq \varepsilon n^2$ or $d_{bip}(G) \geq (1 + k)\varepsilon n^2$ by using $\widetilde{\mathcal{O}}\left((1/k\varepsilon)^6\right)$ queries, which can be derived from the work of Alon, Vega, Kannan and Karpinski [1] (see Corollary A.3 in Appendix A). Hence, any algorithm that solves the general bipartite distance problem with query complexity $o\left((1/k\varepsilon)^6\right)$, will be of huge interest.

―― **References** ――

1   Noga Alon, Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. Random Sampling and Approximation of MAX-CSPs. *Journal of Computer and System Sciences*, 67(2):212–243, 2003. `doi:10.1016/S0022-0000(03)00008-4`.

2   Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000. `doi:10.1007/s004930070001`.

3   Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009. `doi:10.1137/060667177`.

4   Noga Alon and Michael Krivelevich. Testing k-colorability. *SIAM Journal on Discrete Mathematics*, 15(2):211–227, 2002. `doi:10.1137/S0895480199358655`.

5   Andrej Bogdanov and Fan Li. A better tester for bipartiteness? *arXiv preprint*, 2010. `arXiv:1011.0531`.

6   Andrej Bogdanov and Luca Trevisan. Lower Bounds for Testing Bipartiteness in Dense Graphs. In *CCC*, pages 75–81, 2004. `doi:10.1109/CCC.2004.1313803`.

7   Artur Czumaj, Morteza Monemizadeh, Krzysztof Onak, and Christian Sohler. Planar graphs: Random walks and bipartiteness testing. *Random Structures & Algorithms*, 55(1):104–124, 2019. `doi:10.1002/rsa.20826`.

8   Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009. URL: `http://www.cambridge.org/gb/knowledge/isbn/item2327542/`.

9   Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM Journal on Computing*, 37(2):482–501, 2007. `doi:10.1137/060652324`.

10 Arijit Ghosh, Gopinath Mishra, Rahul Raychaudhury, and Sayantan Sen. Tolerant bipartiteness testing in dense graphs. *arXiv preprint*, 2022. `arXiv:2204.12397`.

11 Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. `doi:10.1017/9781108135252`.

12 Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property Testing and its Connection to Learning and Approximation. *Journal of the ACM*, 45(4):653–750, 1998. `doi:10.1145/285055.285060`.

13 Oded Goldreich and Dana Ron. A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999. `doi:10.1007/s004930050060`.

14 Mira Gonen and Dana Ron. On the benefits of adaptivity in property testing of dense graphs. In *APPROX-RANDOM*, pages 525–539, 2007. `doi:10.1007/978-3-540-74208-1_38`.

15 Tali Kaufman, Michael Krivelevich, and Dana Ron. Tight bounds for testing bipartiteness in general graphs. *SIAM Journal on computing*, 33(6):1441–1483, 2004. `doi:10.1137/S0097539703436424`.

16 Claire Mathieu and Warren Schudy. Yet Another Algorithm for Dense Max Cut: Go Greedy. In *SODA*, pages 176–182, 2008. URL: `http://dl.acm.org/citation.cfm?id=1347082.1347102`.

17 Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. `doi:10.1017/CBO9780511813603`.

18 Christian Sohler. Almost Optimal Canonical Property Testers for Satisfiability. In *FOCS*, pages 541–550, 2012. `doi:10.1109/FOCS.2012.59`.

A   Bipartite distance estimation with query complexity $\widetilde{\mathcal{O}}\left(1/\varepsilon^6\right)$

Formally, we state the following theorem.

▶ **Theorem A.1.** *Given an unknown graph G on n vertices and any approximation parameter $\varepsilon \in (0,1)$, there is an algorithm that performs $\widetilde{\mathcal{O}}(1/\varepsilon^6)$ adjacency queries, and outputs a number $\widehat{d}_{bip}(G)$ such that, with probability at least $9/10$, the following holds:*

$$d_{bip}(G) - \varepsilon n^2 \leq \widehat{d}_{bip}(G) \leq d_{bip}(G) + \varepsilon n^2,$$

where $d_{bip}(G)$ denotes the bipartite distance of G.

We have the following two corollaries of the above theorem.

▶ **Corollary A.2.** *There exists an algorithm that given adjacency query access to a graph G with n vertices and a parameter $\varepsilon \in (0,1)$ such that, with probability at least $9/10$, decides whether $d_{bip}(G) \leq \varepsilon n^2$ or $d_{bip}(G) \geq (2 + \Omega(1))\varepsilon n^2$ using $\widetilde{\mathcal{O}}\left(1/\varepsilon^6\right)$ many queries to the adjacency matrix of G.*

▶ **Corollary A.3.** *There exists an algorithm that given adjacency query access to a graph G with n vertices and a parameter $\varepsilon \in (0,1)$ such that, with probability at least $9/10$, decides whether $d_{bip}(G) \leq \varepsilon n^2$ or $d_{bip}(G) \geq (1 + k)\varepsilon n^2$ using $\widetilde{\mathcal{O}}\left((1/k\varepsilon)^6\right)$ many queries to the adjacency matrix of G.*

To prove Theorem A.1, we first discuss the connection between MaxCut and bipartite distance of a graph G. Then we use the result for MaxCut estimation by Alon, Vega, Kannan and Karpinski [1].

Connection between MaxCut and $d_{bip}(G)$

For a graph $G = (V, E)$ on the *vertex* set V and *edge* set E, let S be a subset of V. We define

$$\text{Cut}(S) := |\ \{\{u, v\} \in E\ \mid\ |\{u, v\} \cap S| = 1\}\ |$$

Maximum Cut (henceforth termed as MaxCut), denoted by $M(G)$, is a partition of the vertex set V of G into two parts such that the number of edges crossing the partition is maximized, that is,

$$M(G) := \max_{S \subseteq V} \text{Cut}(S).$$

The following equation connects MaxCut and the bipartite distance of a graph G:

$$d_{bip}(G) = |E(G)| - M(G). \tag{4}$$

So, $d_{bip}(G)$ can be estimated by estimating $|E(G)|$ and $M(G)$.

Result on edge estimation

Observe that estimating $|E(G)|$ with εn^2 additive error is equivalent to *parameter estimation problem* in probability theory, see Mitzenmacher and Upfal [17, Section 4.2.3].

▶ **Proposition A.4** (Folklore). *Given any graph G on n vertices and an input parameter $\varepsilon \in (0,1)$, the size of the edge set $E(G)$ can be estimated within an additive εn^2 error, with probability at least $9/10$, using $\mathcal{O}(1/\varepsilon^2)$ many adjacency queries to G.*

MaxCut **estimation by using** $\widetilde{\mathcal{O}}\left(1/\varepsilon^6\right)$ **queries**

Let $G = (V, E)$ be an n vertex graph. Both Alon, Vega, Kannan and Karpinski [1] and Mathieu and Schudy [16] showed that if S is a t-sized random subset of V, where $t = O\left(\frac{1}{\varepsilon^4} \log \frac{1}{\varepsilon}\right)$, then, with probability at least $\frac{9}{10}$, we have the following:

$$\left| \frac{M(G \mid_S)}{t^2} - \frac{M(G)}{n^2} \right| \leq \frac{\varepsilon}{2}$$

where $G \mid_S$ denotes the induced graph of G on the vertex set S. So, the above inequality tells us that if we can get an $\varepsilon t^2/2$ additive error to $M(G \mid_S)$, then we can get an εn^2 additive estimate for $M(G)$. Observation A.5 implies that using $O\left(t/\varepsilon^2\right) = O\left(\frac{1}{\varepsilon^6} \log \frac{1}{\varepsilon}\right)$ many adjacency queries to $G \mid_S$, we can get an $\varepsilon t^2/2$ additive estimate to $M(G \mid_S)$. Therefore, the query complexity of MaxCut algorithms of Alon, Vega, Kannan and Karpinski [1] and Mathieu and Schudy [16] is at most $O\left(\frac{1}{\varepsilon^6} \log \frac{1}{\varepsilon}\right)$.

Now we state and prove the following observation.

▶ **Observation A.5** (Folklore). *For a graph G with n vertices and an approximation parameter $\varepsilon \in (0,1)$, $\Theta\left(n/\varepsilon^2\right)$ many adjacency queries to G are sufficient to get an εn^2 additive approximation to MaxCut $M(G)$, with probability at least $9/10$.*

Proof. We sample t many pairs of vertices $\{a_1, b_1\}, \ldots, \{a_t, b_t\}$ uniformly at random and independent of each other, where $t = \Theta(n/\varepsilon^2)$. Thereafter, we perform t many adjacency queries to those sampled pairs of vertices. Now fix a subset $S \subset V(G)$ and let us denote $(S, \overline{S})$ to be the set of edges between S and $\overline{S}$.

Let us now define a set of random variables, one for each sampled pair of vertices as follows:

$$X_i = \begin{cases} 1, & \text{if } \{a_i, b_i\} \in (S, \overline{S}) \\ 0, & \text{Otherwise} \end{cases}$$

We will output $\max_{S \subset V(G)} \widehat{M_S}$ as our estimate of $M(G)$, where $\widehat{M_S} = \frac{\binom{n}{2}}{t} \sum_{i=1}^{t} X_i$.

Let us denote $X = \sum_{i=1}^{t} X_i$. Note that

$$\mathbb{E}\left[X_i\right] = \mathbb{P}\left(X_i = 1\right) = \frac{\left|(S, \overline{S})\right|}{\binom{n}{2}},$$

and hence

$$\mathbb{E}\left[\widehat{M_S}\right] = \frac{\binom{n}{2}}{t}\mathbb{E}\left[\sum_{i=1}^{t} X_i\right] = \left|(S, \overline{S})\right|.$$

Using Hoeffding's Inequality (see Lemma B.2), we can say that

$$\mathbb{P}\left(\left|\left|(S, \overline{S})\right| - \widehat{M_S}\right| \geq \frac{\varepsilon n^2}{10}\right) \leq \mathbb{P}\left(\left|X - \mathbb{E}[X]\right| \geq \frac{\varepsilon t}{10}\right) \leq 2e^{-\Theta(\varepsilon^2 t^2/t)} \leq 2e^{-\Theta(n)}.$$

Using union bound over all $S \subset V(G)$, we can show that with probability at least $3/4$, for each $S \subset V(G)$, $\widehat{M_S}$ approximates $\left|(S, \overline{S})\right|$ with εn^2 additive error. Therefore $\max_{S \subset V(G)} \widehat{M_S}$ estimates $M(G)$ with additive error εn^2, with probability at least $3/4$.   ◀

B   Large Deviation Inequalities

▶ **Lemma B.1** (Chernoff-Hoeffding bound, see [8]). *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \in [0, 1]$. For $X = \sum_{i=1}^{n} X_i$ and $\mu_l \leq \mathbb{E}[X] \leq \mu_h$, the followings hold for any $0 < \varepsilon < 1$:*
   (i) $\mathbb{P}\left(X \geq (1 + \varepsilon)\mu_h\right) \leq \exp\left(\frac{-\varepsilon^2 \mu_h}{3}\right).$
   (ii) $\mathbb{P}\left(X \leq (1 - \varepsilon)\mu_l\right) \leq \exp\left(\frac{-\varepsilon^2 \mu_l}{3}\right).$

▶ **Lemma B.2** (Hoeffding's Inequality, see [8]). *Let $X_1, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$ and $X = \sum_{i=1}^{n} X_i$. Then, for all $\delta > 0$,*

$$\mathbb{P}\left(\left|X - \mathbb{E}[X]\right| \geq \delta\right) \leq 2\exp\left(\frac{-2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$