MOSER-TARDOS ALGORITHM: BEYOND SHEARER'S BOUND

KUN HE, QIAN LI, AND XIAOMING SUN

Abstract. In a seminal paper (Moser and Tardos, JACM'10), Moser and Tardos developed a simple and powerful algorithm to find solutions to combinatorial problems in the *variable* Lovász Local Lemma (LLL) setting. Kolipaka and Szegedy (Kolipaka and Szegedy, STOC'11) proved that the Moser-Tardos algorithm is efficient up to the tight condition of the *abstract* Lovász Local Lemma, known as Shearer's bound. A fundamental problem around LLL is whether the efficient region of the Moser-Tardos algorithm can be further extended.

In this paper, we give a positive answer to this problem. We show that the efficient region of the Moser-Tardos algorithm goes beyond the Shearer's bound of the underlying dependency graph, if the graph is not chordal. Otherwise, the dependency graph is chordal, and it has been shown that Shearer's bound exactly characterizes the efficient region for such graphs (Kolipaka and Szegedy, STOC'11; He, Li, Liu, Wang and Xia, FOCS'17).

Moreover, we demonstrate that the efficient region can exceed Shearer's bound by a constant by explicitly calculating the gaps on several infinite lattices.

The core of our proof is a new criterion on the efficiency of the Moser-Tardos algorithm which takes the intersection between dependent events into consideration. Our criterion is strictly better than Shearer's bound whenever the intersection exists between dependent events. Meanwhile, if any two dependent events are mutually exclusive, our criterion becomes the Shearer's bound, which is known to be tight in this situation for the Moser-Tardos algorithm (Kolipaka and Szegedy, STOC'11; Guo, Jerrum and Liu, JACM'19).

1. Introduction

Suppose $\mathcal{A} = \{A_1, \cdots, A_m\}$ is a set of bad events. If the events are mutually independent, then we can avoid all of these events simultaneously whenever no event has probability 1. Lovász Local Lemma (LLL) [EL75], one of the most important probabilistic methods, allows for limited dependency among the events, but still concludes that all the events can be avoided simultaneously if each individual event has a bounded probability. In the most general setting (a.k.a. abstract LLL), the dependency among $\mathcal{A}$ is characterized by an undirected graph $G_D = ([m], E_D)$, called a *dependency graph* of $\mathcal{A}$, which satisfies that for any vertex i, A_i is independent of $\{A_j : j \notin \mathcal{N}_{G_D}(i) \cup \{i\}\}$. Here $\mathcal{N}_G(i)$ stands for the set of neighbors of vertex i in a given graph G.

We use $\mathcal{A} \sim (G_D, \boldsymbol{p})$ to denote that (i) G_D is a dependency graph of $\mathcal{A}$ and (ii) the probability vector of $\mathcal{A}$ is $\boldsymbol{p}$. Given a graph G_D, define the *abstract interior* $\mathcal{I}_a(G_D)$ to be the set consisting of all vectors $\boldsymbol{p}$ such that $\mathbb{P}\left(\cap_{A\in\mathcal{A}}\overline{A}\right) > 0$ for any $\mathcal{A} \sim (G_D, \boldsymbol{p})$. In this context, the most frequently used abstract LLL can be stated as follows:

Theorem 1.1 ([Spe77]). *Given any graph $G_D = ([m], E_D)$ and any probability vector $\boldsymbol{p} \in (0,1]^m$, if there exist real numbers $x_1, ..., x_m \in (0,1)$ such that $p_i \le x_i \prod_{j\in\mathcal{N}_{G_D}(i)}(1-x_j)$ for any $i \in [m]$, then $\boldsymbol{p} \in \mathcal{I}_a(G_D)$.*

Shearer [She85] obtained the strongest possible condition for abstract LLL. Let $\mathrm{Ind}(G_D)$ be the set of all independent sets of an undirected graph $G_D = ([m], E_D)$ and $\boldsymbol{p} = (p_1, \cdots, p_m) \in (0,1]^m$. For each $I \in \mathrm{Ind}(G_D)$, define the quantity

$$q_I(G_D, \boldsymbol{p}) = \sum_{J\in\mathrm{Ind}(G_D), I\subseteq J} (-1)^{|J|-|I|} \prod_{i\in J} p_i.$$

(Kun He, Qian Li, and Xiaoming Sun) Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. University of Chinese Academy of Sciences. Beijing, China. E-mail: hekun.threebody@foxmail.com, liqian@ict.ac.cn, and sunxiaoming@ict.ac.cn.

$\boldsymbol{p}$ is called *in Shearer's bound* of G_D if $q_I(G_D, \boldsymbol{p}) > 0$ for any $I \in \text{Ind}(G_D)$. Otherwise we say $\boldsymbol{p}$ is *beyond Shearer's bound* of G_D. Shearer's result can be stated as follows.

Theorem 1.2 ([She85])**.** *For any graph $G_D = ([m], E_D)$ and any probability vector $\boldsymbol{p} \in (0,1]^m$, $\boldsymbol{p} \in \mathcal{I}_a(G_D)$ if and only if $\boldsymbol{p}$ is in Shearer's bound of G_D.*

Variable Lovász Local Lemma. Variable Lovász Local Lemma (VLLL) is another quite general and common setting of LLL, which applies to variable-generated event systems. In this setting, there is a set of underlying mutually independent random variables $\{X_1, \cdots, X_n\}$, and each event A_i can be fully determined by some variables $\text{vbl}(A_i)$ of them. The dependency between events and variables can be naturally characterized by a bipartite graph $G_B = ([m], [n], E_B)$, known as the event-variable graph, such that edge $(i, j) \in [m] \times [n]$ exists if and only if $X_j \in \text{vbl}(A_i)$.

The variable setting is important, mainly because most applications of LLL have natural underlying independent variables, such as the satisfiability of CNF formulas[GMSW09, GST16, Moi19a, FGYZ20], hypergraph coloring [McD97, GLLZ19], and Ramsey numbers[Spe75, Spe77, Har16]. In particular, the groundbreaking result by Moser and Tardos [MT10] on constructive LLL applies in the variable setting.

There is a natural choice for the dependency graph of variable-generated systems, called the *canonical dependency graph*: two events are adjacent if they share some common variables. Formally, given a bipartite graph $G_B = (U, V, E_B)$, its *base graph* is defined as the graph $G_D(G_B) = (U, E_D)$ such that for any two vertices $u_i, u_j \in U$, $(u_i, u_j) \in E_D$ if and only if u_i and u_j share common neighbors in G_B. If G_B is the event-variable graph of a variable-generated system $\mathcal{A}$, then $G_D(G_B)$ is the canonical dependency graph of $\mathcal{A}$.

Given a graph G_D, define the variable interior $\mathcal{I}_v(G_D)$ to be the set consisting of all vectors $\boldsymbol{p}$ such that $\mathbb{P}\left(\cap_{A \in \mathcal{A}} \overline{A}\right) > 0$ for any *variable-generated* event system $\mathcal{A} \sim (G_D, \boldsymbol{p})$. Obviously, $\mathcal{I}_v(G_D) \supseteq \mathcal{I}_a(G_D)$ for any G_D. In contrast with the abstract LLL, the Shearer's bound (of the canonical dependency graph) turns out to be not tight for variable-generated systems [HLL+17]: the containment is proper if and only if G_D is not chordal[1].

Constructive (variable) Lovász Local Lemma and Moser-Tardos algorithm. The abstract LLL and the variable LLL mentioned above are not constructive in that they do not indicate how to efficiently find an object avoiding all the bad events. In a seminal paper [MT10], Moser and Tardos developed an amazingly simple efficient algorithm for variable-generated systems, depicted in Algorithm 1[2], and showed that this algorithm terminates quickly under the condition in Theorem 1.1. Following the Moser-Tardos algorithm (or MT algorithm for short), a large amount of effort devoted to constructive LLL, including the remarkable works which extend the MT techniques beyond the variable setting [HS14, AIV17, AIK19, AIS19, IS20, HV20]. The MT algorithm has been applied to many important problems, including k-SAT [GST16], hypergraph coloring [Har16], Hamiltonian cycle [Har16], and their counting and sampling [GJL19, Moi19a, FGYZ20, FHY21, JPV21, HSW21].

Algorithm 1: Moser-Tardos Algorithm

Assign random values to $X_1, \cdots, X_n$;
while $\exists i \in [m]$ *such that A_i holds* **do**
  Arbitrarily select one such i and resample all variables X_j in $\text{vbl}(A_i)$;
Return the current assignment;

Mainly because such a simple algorithm is so powerful and general-purpose, it is one of the most intriguing and fundamental problems on constructive LLL how powerful the MT algorithm is. Given a graph G_D, define the *Moser-Tardos interior* $\mathcal{I}_{MT}(G_D)$ to be the set consisting of all vectors $\boldsymbol{p}$ such that the MT algorithm is efficient for any *variable-generated* event system $\mathcal{A} \sim (G_D, \boldsymbol{p})$. Clearly, $\mathcal{I}_{MT}(G_D) \subseteq \mathcal{I}_v(G_D)$ for any G_D. A major line of follow-up works explores $\mathcal{I}_{MT}(G_D)$ [KSX12, Peg14, KS11, CCS+17].

[1]A graph is chordal if it has no induced cycle of length at least four.

[2]Throughout the paper, the Moser-Tardos algorithm is allowed to follow arbitrary selection rules.

The best known criterion is obtained by Kolipaka and Szegedy [KS11]. They extended the MT interior to the Shearer's bound. That is, they showed that $\mathcal{I}_{MT}(G_D) \supseteq \mathcal{I}_a(G_D)$. As mentioned above, if G_D is not chordal, $\mathcal{I}_a(G_D)$ is properly contained in $\mathcal{I}_v(G_D)$, so it is possible to further push $\mathcal{I}_{MT}(G_D)$ beyond Shearer's bound.

In this paper, we concentrate on the following open problem:

***Problem 1**: does $\mathcal{I}_{MT}(G_D)$* properly *contain $\mathcal{I}_a(G_D)$ for some G_D? If so, for what kind of graph G_D?*

Rather than potential applications, our main motivations are the following fundamental problems around LLL itself:

- *The limitation of the constructive LLL in the variable setting.* In the most fascinating problems around LLL, a mysterious conjecture says that there is an algorithm which is efficient for all variable-generated systems $\mathcal{A}$ if $\mathcal{A} \sim (G_D, \boldsymbol{p})$ for some G_D and $\boldsymbol{p} \in \mathcal{I}_v(G_D)$ [Sze13]. It would be a small miracle if the conjecture is true, since if so, one can always *construct* a solution efficiently in the variable setting if solutions are guaranteed to *exist* by the LLL condition. Towards this conjecture, a good start is to show that $\mathcal{I}_{MT}(G_D) \supsetneqq \mathcal{I}_a(G_D)$ for some G_D, as $\mathcal{I}_v(G_D) \supsetneqq \mathcal{I}_a(G_D)$ for G_D which is not chordal.
- *The limitation of the MT algorithm.* The MT algorithm is one of the most intriguing topics in modern algorithm researches, not only because it is very simple and with magic power, but also because it is closely related to the famous Walksat algorithm for random k-SAT. A mysterious problem about the MT algorithm is where is its true limitation [Sze13, CCS+17]. It is conjectured that $\mathcal{I}_{MT}(G_D) = \mathcal{I}_v(G_D)$ for any G_D [Sze13]. To prove this conjecture, the first step is to give a positive answer to Problem 1. Moreover, due to the connection between Shearer's bound and the Repulsive Lattice Gas model, it is conjectured that *essential connection exists between statistical mechanics and the MT algorithm* [Sze13]. Whether $\mathcal{I}_{MT}(G_D) = \mathcal{I}_a(G_D)$ for each G_D is critical to this conjecture.

Remark 1.3. *To explore the power of the MT algorithm in specific applications, one may employ special structures of the applications, such as the way the variables interact, to obtain sharp bounds rather than in terms of the canonical dependency graph only. Nevertheless, characterizing the power of the MT algorithm in terms of the canonical dependency graph is a very fundamental problem and also the focus of the major line of researches [MT10, Peg14, BFPS11, KS11]. Moreover, a major difficulty to strengthen the guarantees of the MT algorithm is that the analysis should be valid for all possible variable-generated event systems. It is not quite surprising to obtain better bounds if the event system has further restrictions. To substantially improve the guarantees of the MT algorithm and provide deep insight about its dynamics, we would rather focus on the general variable LLL setting than employ the special structures in the applications.*

We should emphasize that Problem 1 is still quite open! As mentioned before, it has been proved that the Shearer's bound is not tight for variable-generated systems [HLL+17]. However, this only says that there is some probability vector $\boldsymbol{p}$ beyond the Shearer's bound such that all variable-generated event systems $\mathcal{A} \sim (G_D, \boldsymbol{p})$ must have a satisfying assignment. It is unclear whether the MT algorithm can construct such an assignment efficiently.

It also has been proved that the MT algorithm can still be efficient even beyond the Shearer's bound *for some specific applications* [Har16]. Despite its novel contribution, this result does not provide an answer to Problem 1. The result in [Har16] focuses on the event systems with special structures. Thus, it only implies that there is a probability vector $\boldsymbol{p}$ beyond the Shearer's bound such that the MT algorithm is efficient for **some restricted** variable-generated event systems $\mathcal{A} \sim (G_D, \boldsymbol{p})$. However, to show $\mathcal{I}_{MT}(G_D) \supsetneqq \mathcal{I}_a(G_D)$, one must prove that the MT algorithm is efficient for **all possible** event systems, and this is one major difficulty to resolve Problem 1.

1.1. **Results and contributions.** We provide a complete answer to Problem 1 (Theorem 1.5): if G_D is not chordal, then $\mathcal{I}_{MT}(G_D) \supsetneqq \mathcal{I}_a(G_D)$, i.e., the efficient region of the MT algorithm goes beyond Shearer's bound. Otherwise, $\mathcal{I}_{MT}(G_D) = \mathcal{I}_a(G_D)$, because $\mathcal{I}_a(G_D) \subseteq \mathcal{I}_{MT}(G_D) \subseteq \mathcal{I}_v(G_D)$ and $\mathcal{I}_v(G_D) = \mathcal{I}_a(G_D)$ for chordal graphs G_D [HLL+17].

The core of the proof of Theorem 1.5 is a new convergence criterion for the MT algorithm (Theorem 1.6), which may be of independent interest. This new criterion takes the intersection between dependent events into consideration, and is strictly better than Shearer's bound when there exists a pair of dependent events which are not mutually exclusive.

1.1.1. *Moser-Tardos algorithm: beyond Shearer's bound.* Given a dependency graph $G_D = ([m], E_D)$ and a probability vector $\boldsymbol{p} = (p_1, p_2, \cdots, p_m) \in (0,1)^m$, we say that $\boldsymbol{p}$ is on the Shearer's boundary of G_D if $(1-\varepsilon)\boldsymbol{p}$ is in Shearer's bound and $(1+\varepsilon)\boldsymbol{p}$ is not for any $\varepsilon > 0$. A chordless cycle in a graph G_D is an induced cycle of length at least 4. A chordal graph is a graph without chordless cycles.

Given two vectors $\boldsymbol{p}$ and $\boldsymbol{q}$, we say $\boldsymbol{p} \leq \boldsymbol{q}$ if the inequality holds entry-wise. Additionally, if the inequality is strict on at least one entry, we say that $\boldsymbol{p} < \boldsymbol{q}$.

Definition 1.4 (Maximum L_1-gap to the Shearer's bound)**.** *Given a dependency graph G_D and a probability vector $\boldsymbol{p}$ beyond the Shearer's bound of G_D, define the maximum L_1-gap from $\boldsymbol{p}$ to the Shearer's bound of G_D as*

$$d(\boldsymbol{p}, G_D) \triangleq \arg \sup_{||\boldsymbol{q}||_1} \{\boldsymbol{p} - \boldsymbol{q} \notin \mathcal{I}_a(G_D) : \boldsymbol{q} \leq \boldsymbol{p}\}.$$

For convenience, we let $d(\boldsymbol{p}, G_D) = -1$ if $\boldsymbol{p}$ is in the Shearer's bound of G_D.

Intuitively, $d(\boldsymbol{p}, G_D)$ measures how far $\boldsymbol{p}$ is from the Shearer's bound of G_D. One can verify that $d(\boldsymbol{p}, G_D) < 0$ if $\boldsymbol{p}$ is in the Shearer's bound, $d(\boldsymbol{p}, G_D) = 0$ if $\boldsymbol{p}$ is on the Shearer's boundary, and $d(\boldsymbol{p}, G_D) > 0$ if $\boldsymbol{p}$ is beyond Shearer's bound but not on the Shearer's boundary. Now, we are ready to state our main result.

Theorem 1.5. *For any chordal graph G_D, $\mathcal{I}_{MT}(G_D) = \mathcal{I}_a(G_D)$, i.e., $\boldsymbol{p} \in \mathcal{I}_{MT}(G_D)$ iff $d(\boldsymbol{p}, G_D) < 0$.*

For any graph G_D which is not chordal, $\boldsymbol{p} \in \mathcal{I}_{MT}(G_D)$ if

$$d(\boldsymbol{p}, G_D) < \frac{1}{545} \cdot \sum_{i \leq \ell} |C_i| \big(\min_{j \in C_i} p_j\big)^4 \cdot \left(\max\left\{\frac{2\sum_{j \in C_i} \sqrt{p_j}}{|C_i|} - 1, 0\right\}\right)^2$$

for some disjoint chordless cycles $C_1, C_2, \cdots, C_\ell$ in G_D. In particular, there is a probability vector $\boldsymbol{p}$ with $d(\boldsymbol{p}, G_D) \geq 2^{-20}K^{-3}$ satisfying the above condition, where K is the length the shortest chordless cycle. This implies that $\mathcal{I}_{MT}(G_D)$ contains a probability vector $\boldsymbol{p}$ with $d(\boldsymbol{p}, G_D) \geq 2^{-20}K^{-3}$.

The intuition of Theorem 1.5 is as follows. The theorem characterizes the efficient region of the MT algorithm with $d(\boldsymbol{p}, G_D)$. It shows that if $d(\boldsymbol{p}, G_D)$ is upper bounded by a *non-negative* quantity related to the chordless cycles in G_D, then the MT algorithm is efficient. Since $\mathcal{I}_a(G_D)$ is the set of $\boldsymbol{p}$ where $d(\boldsymbol{p}, G_D) < 0$, our criterion is at least as good as Shearer's bound. Moreover, for each G_D which is not chordal, our criterion is strictly better: there exists some $\boldsymbol{p}$ with $d(\boldsymbol{p}, G_D) \geq 2^{-20}K^{-3}$ satisfying our criterion. Intuitively, Theorem 1.5 implies that *chordless cycles in G_D enhance the power of the MT algorithm.*

We emphasize that Theorem 1.5 provides a *complete* answer to Problem 1: $\mathcal{I}_{MT}(G_D)$ properly contains $\mathcal{I}_a(G_D)$ if and only if G_D is not chordal.

1.1.2. *A new constructive LLL for non-extremal instances.* Given a set $\mathcal{A}$ of events with dependency graph G_D, $\mathcal{A}$ is called *extremal* if all pairs of dependent events are mutually exclusive, and *non-extremal* otherwise. Kolipaka and Szegedy [KS11] showed that the MT algorithm is efficient up to the Shearer's bound. In particular, Shearer's bound is the tight convergence criterion for extremal instances [KS11, GJL19]. Here, we provide a new convergence criterion (Theorem 1.6) which is a strict improvement of Kolipaka and Szegedy's result: this criterion is strictly better than Shearer's bound when the instance is non-extremal, and becomes Shearer's bound when the instance is extremal. This criterion, named *intersection LLL*, is the core of our proof of Theorem 1.5.

Let $G_D = ([m], E_D)$ be a canonical dependency graph and $\boldsymbol{p} = (p_1, \cdots, p_m) \in (0,1)^m$ be a probability vector. Let $\mathcal{M} = \{(i_1, i_1'), (i_2, i_2'), \cdots\} \subseteq E_D$ be a matching of G_D, and $\boldsymbol{\delta} = (\delta_{i_1,i_1'}, \delta_{i_2,i_2'}, \cdots) \in (0,1)^{|\mathcal{M}|}$ be another probability vector. We say that an event set $\mathcal{A}$ is of the setting $(G_D, \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$, and write $\mathcal{A} \sim (G_D, \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$, if $\mathcal{A} \sim (G_D, \boldsymbol{p})$ and $\mathbb{P}(A_i \cap A_{i'}) \geq \delta_{i,i'}$ for each pair $(i, i') \in \mathcal{M}$. Given $(G_D, \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$, define $\boldsymbol{p}^- \in (0,1)^m$ as follows:

$$\forall i \in [m]: \quad p_i^- = \begin{cases} p_i - \frac{1}{17}\cdot\delta_{i,i'}^2, & \text{if } (i,i') \in \mathcal{M} \text{ for some } i'; \\ p_i, & \text{otherwise.} \end{cases}$$

Theorem 1.6 (intersection LLL (informal)). *For any $\mathcal{A} \sim (G_D, \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$, MT algorithm terminates quickly if $\boldsymbol{p}^-$ is in the Shearer's bound of G_D.*

The intuition of Theorem 1.6 is as follows. For any matching $\mathcal{M}$ in G_D, if the intersection of events on each edge (i,i') in $\mathcal{M}$ has a lower bound $\delta_{i,i'}$, then one can subtract $\frac{1}{17}\cdot\delta_{i,i'}^2$ from the probabilities of endpoints i and i', and the MT algorithm is guaranteed to be efficient whenever the reduced probability vector is in the Shearer's bound.

Remark 1.7. *In many applications of LLL [McD97, GST16, GMSW09, Moi19a, GKPT17], the dependent bad events naturally intersect with each other. For instance, in a CNF formula, if the common variables in two clauses are both either positive or negative, then the bad events corresponding to these two clauses are dependent and intersect with each other. Thus our intersection LLL may be capable of improving bounds for these applications. However, currently the improvement is weak because only the intersections between the matched events are considered in Theorem 1.6.*

Nevertheless, the primary motivation of this work is to explore the power of the MT algorithm in the general variable LLL setting. This basic problem is very important in itself, besides its potential applications.

1.1.3. *Application to lattices.* To illustrate the application of Theorem 1.5, we estimate the efficient region of the MT algorithm on some lattices explicitly. For simplicity, we focus on symmetric probabilities, i.e., $\boldsymbol{p} = (p, p, \cdots, p)$. Our lower bounds on the gaps between the efficient region of the MT algorithm and the Shearer's bound are summarized in Table 1. For example, when the canonical dependency graph is the square lattice, the vector $(0.1193, 0.1193, \cdots)$ is on the Shearer's boundary, and the MT algorithm is provably efficient whenever the probability of each event is at most $0.1193 + 1.858 \times 10^{-22}$.

TABLE 1. Summary of lower bounds on the gaps

Lattice	Shearer's bound	lower bound on the gaps
Square	0.1193 [GF65, Tod99]	1.858×10^{-22}
Hexagonal	0.1547 [Tod99]	2.597×10^{-25}
Simple Cubic	0.0744 [Gau67]	7.445×10^{-23}

1.2. **Technique overview.** As mentioned before, the Shearer's bound is the tight criterion for MT algorithm on extremal instances. Thus in order to show that MT algorithm goes beyond Shearer's bound, we need to take advantage of the intersection between dependent events. Specifically, Theorem 1.5 immediately follows from two results about non-extremal instances. One is the intersection LLL criterion (Theorem 1.6), which goes beyond Shearer's bound whenever there are intersections between dependent events. The other result is a lower bound on the amount of intersection between dependent events for general instances (Theorem 4.1).

1.2.1. *Proof overview of Theorem 1.6.* Let us first remember Kolipaka and Szegedy's argument [KS11], which shows that the MT algorithm is efficient up to the Shearer's bound. We assume that $\{A_i\}_{i=1}^m$ is a fixed set of events with dependency graph $G_D = ([m], E_D)$ and probabilities $\boldsymbol{p} = (p_1, \cdots, p_m)$. The notion of a witness DAG[3] (abbreviated wdag) is central to their argument. A wdag is a DAG whose each node v has a label $L(v)$ from $[m]$ and in which two nodes v and v' are connected by an arc if and only if $L(v) = L(v')$ or $(L(v), L(v')) \in E_D$. With a resampling sequence $\boldsymbol{s} = s_1, s_2, \cdots, s_T$ (i.e., MT algorithm picks the events $A_{s_1}, A_{s_2}, \cdots, A_{s_T}$ for resampling in this order), we associate a wdag $D_{\boldsymbol{s}}$ on node set $\{v_1, \cdots, v_T\}$ as follows: (a) $L(v_k) = s_k$ and (b) there is an arc from v_k to v_ℓ with $k < \ell$ if and only if either $s_k = s_\ell$ or $(s_k, s_\ell) \in E_D$ (see an example in Figure 1). We say that a wdag D occurs in the resampling

[3]In the paper [KS11], the role of witness DAGs was played by "stable set sequences", but the concepts are essentially the same: there is a natural bijection between stable set sequences and wdags.

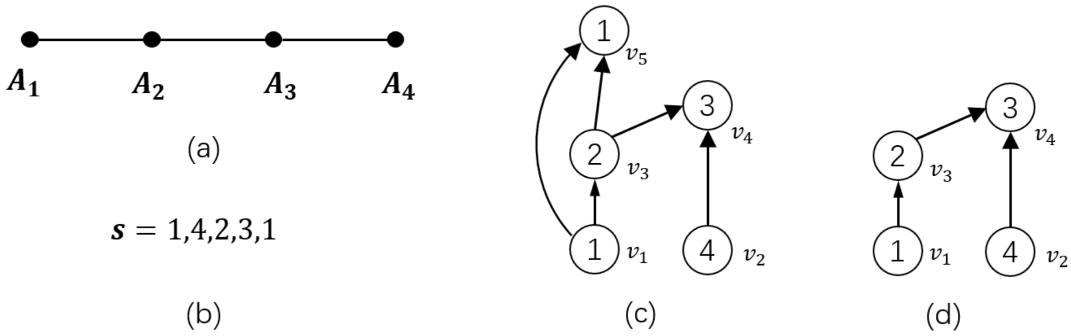

FIGURE 1. (a) a dependency graph G_D; (b) a resample sequence; (c) the D_s; (d) a wdag occurring in s.

sequence s if there is subset U of nodes in D_s such that D is a subgraph of D_s induced by the nodes that have a directed path to U (Figure 1 (d) is an example, where $U = \{v_4\}$). An useful observation is that $\mathbb{E}[T] = \sum_{D\in\mathcal{D}} \mathbb{P}_s[D \text{ occurs in } s]$. Here, $\mathcal{D}$ denotes the set of all single-sink wdags (a.k.a. proper wdags) of G_D.

We define the weight of a wdag D to be $\Pi_{v\in D} p_{L(v)}$. The crucial lemma in Kolipaka and Szegedy's argument (the idea is from Moser-Tardos analysis) is that the probability of occurrence of a certain wdag D is upper bounded by its weight. The idea is that we can assume (only for the analysis) that the MT algorithm has a preprocessing step where it prepares an infinite number of independent samples for each variable. These independent samples create a table X, called *the resampling table* (see Figure 2 in Section 3.1 for an example). When the MT algorithm decides to resample variable X_j, it picks a new sample of X_j from the resampling table. Suppose a certain wdag D occurs, then for each of its events we can determine a particular set of samples in the resampling table that must satisfy the event, where we say that D is consistent with the resampling table X and denote it by $D \sim X$. Hence, $\mathbb{P}_s[D \text{ occurs in } s] \le \mathbb{P}_X[D \sim X] = \Pi_{v\in D} p_{L(v)}$.

Finally, they solved beautifully the summation of weights of proper wdags, i.e., $\sum_{D\in\mathcal{D}} \Pi_{v\in D} p_{L(v)}$, which turns out to converge if and only if p is in the Shearer's bound of G_D.

Viewing Theorem 1.6 as an improvement of Kolipaka and Szegedy's result, we begin by providing a tighter upper bound on $\sum_{D\in\mathcal{D}} \mathbb{P}_s[D \text{ occurs in } s]$ when the instance is non-extremal (Theorem 3.7). First, note that for each wdag D, there exist selection rules to make $\mathbb{P}_s[D \text{ occurs in } s] = \Pi_{v\in D} p_{L(v)}$, so it is *impossible* to give a better upper bound on $\mathbb{P}_s[D \text{ occurs in } s]$ which holds for all selection rules. Our idea is to group proper wdags, and consider the sum of $\mathbb{P}_s[D \text{ occurs in } s]$ over a group. For example, suppose that A_1 and A_2 are dependent and $\mathbb{P}[A_1 \cap A_2] \ge \delta_{1,2}$. Let D_1 denote the proper wdag which consists of only one arc $A_1 \to A_2$, and D_2 denote the proper wdag consisting of only $A_2 \to A_1$. D_1 and D_2 cannot both occur, but they may be both consistent with a given resampling table. So the total weights of D_1 and D_2 is an overestimate of the probability that D_1 or D_2 occurs. Formally,

$$\begin{aligned}\mathbb{P}_s[D_1 \text{ occurs in } s] + \mathbb{P}_s[D_2 \text{ occurs in } s] =& \mathbb{P}_s[(D_1 \text{ occurs in } s) \vee (D_2 \text{ occurs in } s)]\\ \le& \mathbb{P}_X[(D_1 \sim X) \vee (D_2 \sim X)]\\ =& \mathbb{P}_X[D_1 \sim X] + \mathbb{P}_X[(D_2 \sim X) \wedge (D_1 \nsim X)]\\ \le& p_1p_2 + p_1p_2 - \delta_{1,2}^2,\end{aligned}$$

where the last inequality is according to the Cauchy–Schwarz inequality (see Proposition 3.3). Importantly, the upper bound holds for all selection rules.

It is crucial as well as the difficulty that our improvement over the weight of wdags should be "exponential": since the quantity $\sum_{D\in\mathcal{D}} \Pi_{v\in D} p^-_{L(v)}$ converges if and only if p^- is in the Shearer's bound, constant factor or even sub-exponential improvements over $\sum_{D\in\mathcal{D}} \Pi_{v\in D} p_{L(v)}$ do not help to show the desired convergence criterion. Our exponential improvement relies on a delicate grouping and a tricky random partition of the union of $D \sim X$ across wdags.

We first state how we group proper wdags: define $\mathcal{D}(i,r)$ to be the set of proper wdags whose unique sink node is labelled with i and in which there are exactly r nodes labelled with i. Noticing that at most one wdag in $\mathcal{D}(i,r)$ can occur, we have that

$$\sum_{D\in\mathcal{D}(i,r)} \mathbb{P}_s[D \text{ occurs in } s] = \mathbb{P}_X\left[\bigvee_{D\in\mathcal{D}(i,r)} (D \text{ occurs})\right] \le \mathbb{P}_X\left[\bigvee_{D\in\mathcal{D}(i,r)} (D\sim X)\right].$$

Now, we partition the space $\bigvee_{D\in\mathcal{D}(i,r)}(D\sim X)$ across wdags in $\mathcal{D}(i,r)$. The notions of *reversible arcs* (see Definition 2.4) and a *auxiliary table* (see Section 3.1) are two central concepts here. Specifically, an arc $u\to v$ in a wdag D is said reversible, if the directed graph obtained from D by reversing the direction of $u\to v$ is also a wdag. The auxiliary table is a table Y of independent fair coins corresponding to directions of reversible arcs. We say a wdag D is consistent with (X,Y), denoted by $D\sim(X,Y)$ if (i) $D\sim X$; and (ii) for each reversible arc whose direction is *not* consistent with Y, the wdag obtained by reversing the arc is *not* consistent with X. The crucial lemma (Lemma 3.1) shows that for any certain assignment y of the auxiliary table Y, $\bigvee_{D\in\mathcal{D}(i,r)}(D\sim X) = \bigvee_{D\in\mathcal{D}(i,r)}(D\sim(X,y))$. The point is that $(D\sim(X,y))$'s have much less overlap with each other so that they can be viewed as a "approximate" partition of the space. By applying union bound, we get

$$\begin{aligned}\mathbb{P}_X\left[\bigvee_{D\in\mathcal{D}(i,r)} (D\sim X)\right] &= \mathbb{E}_Y\mathbb{P}_X\left[\bigvee_{D\in\mathcal{D}(i,r)} (D\sim X)\right] = \mathbb{E}_Y\mathbb{P}_X\left[\bigvee_{D\in\mathcal{D}(i,r)} (D\sim (X,Y))\right]\\ &\le \mathbb{E}_Y \sum_{D\in\mathcal{D}(i,r)} \mathbb{P}_X\left[D\sim(X,Y)\right]\\ &= \sum_{D\in\mathcal{D}(i,r)} \mathbb{E}_Y\mathbb{P}_X\left[D\sim(X,Y)\right].\end{aligned}$$

Then we are able to provide an upper bound on $\mathbb{E}_Y\mathbb{P}_X[D\sim(X,Y)]$ which is "exponentially" smaller than $\Pi_{v\in D}p_{L(v)}$ (Lemma 3.4), and then complete the proof of Theorem 3.7.

The next step is to show that the tighter upper bound converges when p^- is in the Shearer's bound. For each vertex i in the matching $\mathcal{M}$, we "split" vertex i into two new connected vertices $i^\uparrow$ and $i^\downarrow$. Let $G^{\mathcal{M}}$ be the resulted dependency graph (see an example in Figure 3). Define $p^{\mathcal{M}}_{i\uparrow} = p'_i$ and $p^{\mathcal{M}}_{i\downarrow} = p^-_i - p'_i$ (see the definition of p'_i in Section 2.3). One can see that (G_D, p^-) and $(G^{\mathcal{M}}, p^{\mathcal{M}})$ are essentially the same: suppose $\mathcal{A}\sim(G_D,p^-)$, then for each $i\in\mathcal{M}$, we view A_i as the union of two mutually exclusive events $A_{i\uparrow}$ and $A_{i\downarrow}$ whose probabilities are p'_i and $p^- - p'_i$ respectively. Such a representation of $\mathcal{A}$ is of the setting $(G^{\mathcal{M}}, p^{\mathcal{M}})$. Thus, the sum of weights of proper wdags in the setting (G_D, p^-) is equal to that in the setting $(G^{\mathcal{M}}, p^{\mathcal{M}})$ (Proposition 3.9). So it suffices to show that our tighter upper bound is upper bounded by the sum of weights of proper wdags in the setting $(G^{\mathcal{M}}, p^{\mathcal{M}})$ (Theorem 3.13). Our idea is to construct a mapping which maps each $D\in\mathcal{D}(G_D)$ to a subset of $\mathcal{D}(G^{\mathcal{M}})$ and satisfies that:

(a) distinct proper wdags of G_D are mapped to disjoint subsets of $\mathcal{D}(G^{\mathcal{M}})$; and
(b) for each $D\in\mathcal{D}(G_D)$, the bound in Lemma 3.4 is upper bounded by the sum of weights of proper wdags over the subset that D is mapped to.

We present such a mapping in Definition 3.11. Conditions (a) and (b) are verified in Theorem 3.12 and Theorem 3.13 respectively.

The idea of constructing a mapping between wdags of two dependency graphs may be of independent interest, and may be applied elsewhere when we wish to show some properties about Shearer's bound.

1.2.2. *Proof overview of Theorem 4.1.* The proof of Theorem 4.1 mainly consists of two parts. First, we show that there is an elementary event set which approximately achieves the minimum amount of the intersection between dependent events (Lemma 4.2). Here, we call an event $A_i\in\mathcal{A}$ elementary, if there is a subset S^i_j of the domain of variable X_j for each variable in $\mathrm{vbl}(A)$ such that A happens if and only if $X_j\in S^i_j$ for all variables in $\mathrm{vbl}(A)$. We call a set $\mathcal{A}$ of events elementary if every $A_i\in\mathcal{A}$ is elementary. Then, for elementary event sets, by applying AM-GM inequality, we obtain a lower bound on the total amount of overlap on common variables, which further implies a lower bound on the amount of intersection between dependent events (Lemma 4.5).

1.3. **Related works.** Beck proposed the first constructive LLL, which provides efficient algorithms for finding the perfect object avoiding all "bad" events [Bec91]. His methods were refined and improved by a long line of research [Alo91, MR98, CS00, HSS11]. In a groundbreaking work, Moser and Tardos proposed a new algorithm, i.e., Algorithm 1, and proved that it finds such a perfect object under the condition in Theorem 1.1 in the variable setting [MT10]. Pegden [Peg14] proved that the MT algorithm efficiently converges even under the condition of the cluster expansion local lemma [BFPS11]. Kolipaka and Szegedy [KS11] pushed the efficient region to Shearer's bound. The phenomenon that the MT algorithm can still be efficient beyond Shearer's bound was known to exist *for sporadic and toy examples* [Har16]. However, such result employs the special structures in the examples and only applies to **some restricted** variable-generated event systems $\mathcal{A} \sim (G_D, \boldsymbol{p})$. By contrast, the results in this work applies to **all** variable-generated event systems.

Besides the line of research exploring the efficient region of the MT algorithm, there is a large amount of effort devoted to derandomizing or parallelizing the MT algorithm [MT10, CGH13, Har19, BFH+16, Gha16, CPS17, HH17, Har18] and to extending the Moser-Tardos techniques beyond the variable setting [HS14, AI16, HV20, AIV17, AIK19, Mol19, IS20, AIS19].

There is a line of works studying the gap between non-constructive VLLL and Shearer's bound [KS11, HLL+17, Gil19, HLSZ19]. Kolipaka and Szegedy [KS11] obtained the first example of gap existence where the canonical dependency graph is a cycle of length 4. The paper [HLL+17] showed that Shearer's bound is not tight for VLLL. More precisely, Shearer's bound is tight for non-constructive VLLL if and only if the canonical dependency graph is chordal. The first paper to study quantitatively the gaps systematically is [HLSZ19], which provides lower bounds on the gap when the canonical dependency graph containing many chordless cycles.

Erdös and Spencer [ES91] introduced the lopsided-LLL, which extends the results in [EL75] to lopsidependency graphs. Lopsided LLL has many interesting applications in combinatorics and theoretical computer science, such as the k-SAT [GST16], random permutations [LS07], Hamiltonian cycles [AFR95], and matchings on the complete graph [LS09]. Shearer's bound is also the tight condition for the lopsided LLL [She85].

LLL has a strong connection to sampling. Guo, Jerrum and Liu [GJL19] proved that the MT algorithm indeed uniformly samples a perfect object if the instance is extremal. For extremal instances, they developed an algorithm called "partial rejection sampling" which resamples in a parallel fashion, since the occurring bad events form an independent set in the dependency graph. Actually, a series of sampling algorithms for specific problems are the parallel resampling algorithm running in the extremal case [GJL19, GJ19, GH20, GJ18]. In a celebrated work, Moitra [Moi19b] introduced a novel approach that utilizes LLL to sample k-CNF solutions. This approach was then extended by several works [GLLZ19, GGGY20, FGYZ20, FHY20, JPV20, JPV21].

1.4. **Organization of the paper.** In Section 2, we recall and introduce some definitions and notations. In Section 3, we prove Theorem 1.6. Section 4 is about the proof of Theorem 4.1, which gives a lower bound on the amount of the intersection between dependent events. In Section 5, we prove Theorem 1.5. In Section 6, we provide a explicit lower bound for the gaps between the efficient region of MT algorithm and Shearer's bound on periodic Euclidean graphs.

2. Preliminaries

Let $\mathbb{N} = \{0, 1, 2, \cdots\}$ denote the set of non-negative integers. Let $\mathbb{N}^+ = \{1, 2, \cdots\}$ denote the set positive integers. For $m \in \mathbb{N}^+$, we define $[m] = \{1, \cdots, m\}$. Throughout this section, we fix a canonical dependency graph $G_D = ([m], E_D)$.

2.1. **Witness DAG.** If for a given run, MT algorithm picks the events $A_{s_1}, A_{s_2}, ..., A_{s_T}$ for resampling in this order, we say that $\boldsymbol{s} = s_1, s_2 ..., s_T$ is *a resample sequence*. If the algorithm never finishes, the resample sequence is infinite, and in this case we set $T = \infty$.

Definition 2.1 (Witness DAG). *We define a witness DAG (abbreviated wdag) of G_D to be a DAG D, in which each node v has a label $L(v)$ from $[m]$, and which satisfies the additional condition that for all distinct nodes $v, v' \in D$ there is an arc between v and v' (in either direction) if and only if $L(v) = L(v')$ or $\big(L(v), L(v')\big) \in E_D$.*

We say D is a proper *wdag (abbreviated pwdag) if D has only one sink node. Let $\mathcal{D}(G_D)$ denote the set of pwdags of G_D.*

Given a resampling sequence $\boldsymbol{s} = s_1, s_2, ..., s_T$, we associate a wdag $D_{\boldsymbol{s}}$ on the node set $\{v_1, ..., v_T\}$ such that (i) $L(v_k) = s_k$ and (ii) $v_k \to v_\ell$ with $k < \ell$ is as an arc of $D_{\boldsymbol{s}}$ if and only if either $s_k = s_\ell$ or $(s_k, s_\ell) \in E_D$. See Figure 1 for an example of $D_{\boldsymbol{s}}$.

Given a wdag D and a set U of nodes of D, we define $D(U)$ to be the induced subgraph on all nodes which has a directed path to some $u \in U$. Note that $D(U)$ is also a wdag. We say that H is a prefix of D, denoted by $H \trianglelefteq D$, if $H = D(U)$ for some node set U.

Definition 2.2. *We say a wdag D occurs in a resampling sequence $\boldsymbol{s}$ if $D \trianglelefteq D_{\boldsymbol{s}}$. Let χ_D be the indicator variable of the event that D occurs in $\boldsymbol{s}$.*

Similar to Lemma 12 in [KS11], we have that $T = \sum_{D \in \mathcal{D}(G_D)} \chi_D$. For $i \in [m]$ and $r \in \mathbb{N}^+$, define $\mathcal{D}(i, r)$ to be the set of pwdags whose unique sink node is labelled with i and in which there are exactly r nodes labelled with i. Let $\chi_{\mathcal{D}(i,r)}$ be the indicator variable of the event that there is a $D \in \mathcal{D}(i, r)$ occurring in $\boldsymbol{s}$. It is easy to see that only one pwdag in $\mathcal{D}(i, r)$ can occur in $\boldsymbol{s}$. Thus $\chi_{\mathcal{D}(i,r)} = \sum_{D \in \mathcal{D}(i,r)} \chi_D$, which further implies that

Fact 2.3. $T = \sum_{i \in [m]} \sum_{r \in \mathbb{N}^+} \chi_{\mathcal{D}(i,r)}$.

2.2. **Reversible arc.** In the rest of this section, we fix a matching $\mathcal{M} \subseteq E_D$ of G_D. Given $i \in [m]$, with a slight abuse of notation, we sometimes say $i \in \mathcal{M}$ if there is some $i' \in [m]$ such that $(i, i') \in \mathcal{M}$.

Definition 2.4 (Reversibility). *We say that an arc $u \to v$ is reversible in a wdag D if the directed graph obtained from D by reversing the direction of the arc is still a DAG.*

Furthermore, we say that $u \to v$ is $\mathcal{M}$-reversible in D if $u \to v$ is reversible in D and $(L(u), L(v)) \in \mathcal{M}$.

By definition, we have the following two observations.

Fact 2.5. *$u \to v$ is reversible in D if and only if it is the unique path from u to v in D.*

Fact 2.6. *If $u \to v$ is reversible in a wdag D of G_D, then the directed graph obtained from D by reversing the direction of $u \to v$ is also a wdag of G_D.*

Given a pwdag $D = (V, E, L)$, define

$$\mathcal{V}(D) \triangleq \{v : \exists u \in V \text{ such that } u \to v \text{ or } v \to u \text{ is } \mathcal{M}\text{-reversible in } D\}$$

to be the set of nodes participating in reversible arcs, and $\overline{\mathcal{V}}(D) \triangleq V \setminus \mathcal{V}(D)$. For $i \in [m]$, define $\mathcal{V}(D, i) \triangleq \mathcal{V}(D) \cap \{v : L(v) = i\}$.

2.3. **Other notations.** Let $\boldsymbol{p} = (p_1, \cdots, p_m) \in (0, 1]^m$ and $\boldsymbol{\delta} \in (0, 1)^{\mathcal{M}}$ be two probability vectors. Recall that $\boldsymbol{p}^- = (p_1^-, \cdots, p_m^-)$ is defined as

$$\forall i \in [m]: \quad p_i^- = \begin{cases} p_i - \frac{\delta_{i,i'}^2}{17} & \text{if } (i, i') \in \mathcal{M} \text{ for some } i', \\ p_i & \text{otherwise.} \end{cases} \tag{1}$$

For each $i \in [m]$ where $(i, i') \in \mathcal{M}$ for some $i' \in [m]$, define

$$c_i \triangleq \frac{\delta_{i,i'}^2}{8 p_i p_{i'}} \qquad \text{and} \qquad p_i' \triangleq p_i(1 - c_i) = p_i - \frac{\delta_{i,i'}^2}{8 p_{i'}}.$$

Fact 2.7. $p_i^- + p_{i'}^-(p_i^- - p_i') \geq p_i$ *for each* $(i, i') \in \mathcal{M}$.

3. Proof of Theorem 1.6

The proof of Theorem 1.6 consists of two parts. First, we provide a tighter upper bound on the complexity of MT algorithm (Section 3.1). Then, we show that the tighter upper bound converges if $\boldsymbol{p}^-$ is in the Shearer's bound of G_D (Section 3.2).

3.1. A tighter upper bound on the complexity of MT algorithm.

In this subsection, we prove Theorem 3.7, which follows from Lemma 3.1 and Lemma 3.4 immediately. We first recall and introduce some concepts and notations.

Resampling Table. One key analytical technique of Moser and Tardos [MT10] is to precompute the randomness in a resampling table $\boldsymbol{X}$. Specifically, we can assume (only for the analysis) that MT algorithm has a preprocessing step where it draws an infinite number of independent samples $X_j^1, X_j^2, \cdots$ for each variable X_j. These independent samples create a table $\boldsymbol{X} = (X_j^k)_{j\in[m],k\in\mathbb{N}^+}$, called the resampling table (see Figure 2). MT algorithm takes that first column as the initial assignments of $X_1, \cdots, X_n$. Then, when X_j is to be resampled, MT algorithm goes right in the row corresponding to X_j and picks the sample.

Consistency with the resampling table. For a wdag D, a node v, and a variable $X_j \in \text{vbl}(A_{L(v)})$, we define

$$\mathcal{L}(D, v, j) \triangleq |\{u : \text{ there is a directed path from } u \text{ to } v \text{ in } D \text{ and } X_j \in \text{vbl}(A_{L(u)})\}| + 1.$$

Moreover, let $\boldsymbol{X}_{D,v} \triangleq \{X_j^{\mathcal{L}(D,v,j)} : X_j \in A_{L(v)}\}$. We say that D is *consistent* with $\boldsymbol{X}$, denoted by $D \sim \boldsymbol{X}$, if for each node v in D, the event $A_{L(v)}$ holds on $\boldsymbol{X}_{D,v}$. Intuitively, suppose D occurs, then $\boldsymbol{X}_{D,v}$ are the assignments of $\text{vbl}(A_{L(v)})$ just before the time that the MT algorithm picks the event corresponding to v to resample, hence $A_{L(v)}$ must hold on $\boldsymbol{X}_{D,v}$. We sometimes use $\mathcal{L}(v, j)$ and $\boldsymbol{X}_v$ instead of $\mathcal{L}(D, v, j)$ and $\boldsymbol{X}_{D,v}$ respectively if D is clear from the context. Besides, we use $\mathcal{D}(i, r) \sim \boldsymbol{X}$ to denote that there is some $D \in \mathcal{D}(i, r)$ such that $D \sim \boldsymbol{X}$.

X_1^1	X_1^2	X_1^3	$\cdots$
X_2^1	X_2^2	X_2^3	$\cdots$
X_3^1	X_3^2	X_3^3	$\cdots$
X_4^1	X_4^2	X_4^3	$\cdots$

$Y_{1,2}^1$	$Y_{1,2}^2$	$Y_{1,2}^3$	$\cdots$
$Y_{3,4}^1$	$Y_{3,4}^2$	$Y_{3,4}^3$	$\cdots$
$Y_{5,6}^1$	$Y_{5,6}^2$	$Y_{5,6}^3$	$\cdots$
$Y_{7,8}^1$	$Y_{7,8}^2$	$Y_{7,8}^3$	$\cdots$

Figure 2. The left is a resampling table where there are four variables $X_1, \cdots, X_4$. The right is an auxiliary table where $\mathcal{M} = \{(1, 2), (3, 4), (5, 6), (7, 8)\}$.

Auxiliary Table. We introduce another central concept in the proof of Theorem 3.7, called the auxiliary table, which is a table of independent fair coins. Specifically, for each pair $(i, i') \in \mathcal{M}$, we draw an infinite number of independent fair coins $Y_{i,i'}^1, Y_{i,i'}^2, \cdots$, where $\mathbb{P}(Y_{i,i'}^k = i) = \mathbb{P}(Y_{i,i'}^k = i') = 1/2$. These independent coins form the auxiliary table $\boldsymbol{Y} = (Y_{i,i'}^k)_{(i,i')\in\mathcal{M},k\in\mathbb{N}^+}$ (see Figure 2). The auxiliary table is used to encode directions of $\mathcal{M}$-reversible arcs, according to which we partition the space $\bigvee_{D\in\mathcal{D}(i,r)}(D \sim \boldsymbol{X})$.

Consistency with the resampling table and the auxiliary table. We need some notations about reversible arcs. Suppose D has a unique sink node w and $u \to v$ is reversible in D. Let D' be the DAG obtained from D by reversing the direction of $u \to v$. We define $\varphi(D, u, v) \triangleq D'(\{w\})$. In other words, $\varphi(D, u, v)$ is the prefix of D' with a unique sink node w. Given $(i, i') \in \mathcal{M}$ and a pwdag D, let $\text{List}(D, i, i')$ denote the sequence listing all nodes in D with labels i or i' in a topological order of G_D[4]. Given a node v in D, if $(L(v), i) \in \mathcal{M}$[5], we define

$$\lambda(v, D) \triangleq |\{u : (u \to v \text{ is in } D) \wedge (L(u) \in \{i, L(v)\})\}| + 1$$

[4] It is easy to see that $\text{List}(D, i, i')$ is well defined. That is, all topological orderings of D induce the same $\text{List}(D, i, i')$.

[5] Because $\mathcal{M}$ is a matching, there is at most one such i.

to be the order of v in $\text{List}(D, L(v), i)$. For simplicity of notations, we will use $\lambda(v)$ instead of $\lambda(v, D)$ if D is clear from the context.

Given a wdag D, we say an $\mathcal{M}$-reversible arc $u \to v$ is inconsistent with the auxiliary table Y if $y^{\lambda(u)}_{L(u),L(v)} = L(v)$. We say D is consistent with $(\boldsymbol{X}, Y)$, denoted by $D \sim (\boldsymbol{X}, Y)$, if (i) $D \sim \boldsymbol{X}$ and (ii) for any $\mathcal{M}$-reversible arc $u \to v$ inconsistent with $\boldsymbol{y}$, $\varphi(D, u, v) \nsim \boldsymbol{X}$. We say $\mathcal{D}(i, r) \sim (\boldsymbol{X}, Y)$ if there is some $D \in \mathcal{D}(i, r)$ such that $D \sim (\boldsymbol{X}, Y)$.

The intuition of the notion "consistency" is as follows. Suppose $u \to v$ in a $\mathcal{M}$-reversible arc in D, and both D and $\varphi(D, u, v)$ are consistent with the resampling table. But D and $\varphi(D, u, v)$ cannot both occur. It is according to the auxiliary table to which one of D and $\varphi(D, u, v)$ we assign $(D \sim \boldsymbol{X}) \wedge (\varphi(D, u, v) \sim \boldsymbol{X})$.

Lemma 3.1. *For each $i \in [m]$ and $r \in \mathbb{N}^+$, $\mathbb{P}_{\boldsymbol{X}}[\mathcal{D}(i, r) \sim \boldsymbol{X}] = \mathbb{P}_{\boldsymbol{X},Y}[\mathcal{D}(i, r) \sim (\boldsymbol{X}, Y)]$.*

Proof. Fix an arbitrary assignment $\boldsymbol{x}$ of $\boldsymbol{X}$ and an arbitrary assignment $\boldsymbol{y}$ of Y. Suppose $\mathcal{D}(i, r) \sim \boldsymbol{x}$, i.e, $\exists D_0 \in \mathcal{D}(i, r)$ such that $D \sim \boldsymbol{x}$. We will show that there must exist some $D \in \mathcal{D}(i, r)$ such that $D \sim (\boldsymbol{x}, \boldsymbol{y})$. This will imply the conclusion immediately.

We apply the following procedure to find such a pwdag $D \in \mathcal{D}(i, r)$.

Initially, $k = 0$;
while $\exists$ *an $\mathcal{M}$-reversible arc $u_k \to v_k$ in D_k inconsistent with $\boldsymbol{y}$ such that $\varphi(D_k, u_k, v_k) \sim \boldsymbol{x}$* **do**
  let $D_{k+1} := \varphi(D_k, u_k, v_k)$ and $k := k + 1$;
Return D_k;

By induction on k, it is easy to check that $D_k \sim \boldsymbol{x}$ and $D_k \in \mathcal{D}(i, r)$ for each k. Furthermore, if the procedure terminates, then in the final wdag D, for every $\mathcal{M}$-reversible arc $u \to v$ inconsistent with $\boldsymbol{y}$, we have that $\varphi(D, u, v) \nsim \boldsymbol{x}$. So $D \sim (\boldsymbol{x}, \boldsymbol{y})$. In the following, we will show that the procedure always terminates, which finishes the proof.

Note that each D_k has no more nodes than D_0 and that there are finite number of wdags in $\mathcal{D}(i, r)$ with no more nodes than D_0, so it suffices to prove that each wdag appears at most once in the procedure.

By contradiction, assume $D_j = D_k$ for some $j \le k$. Recall that $u_j \to v_j$ is reversible in D_j and inconsistent with $\boldsymbol{y}$. So $y^{\lambda(v_j,D_j)-1}_{L(u_j),L(v_j)} = y^{\lambda(u_j,D_j)}_{L(u_j),L(v_j)} = L(v_j)$.

Let D_ℓ be the last wdag in $D_{j+1}, \cdots, D_k$ such that $\lambda(v_j, D_\ell) < \lambda(v_j, D_j)$. Observing that $\lambda(v_j, D_{j+1}) = \lambda(v_j, D_j) - 1$, we have such D_ℓ must exist. By $\lambda(v_j, D_k) = \lambda(v_j, D_j)$, we have $\lambda(v_j, D_\ell) = \lambda(v_j, D_j) - 1$, $\lambda(v_j, D_{\ell+1}) = \lambda(v_j, D_j)$. Therefore, $\lambda(v_j, D_{\ell+1}) = \lambda(v_j, D_\ell) + 1$. Combining with that $u_\ell \to v_\ell$ is the inconsistent arc in D_ℓ which is reversed in $D_{\ell+1}$, we have $u_\ell = v_j$, $(L(u_j), L(v_j)) = (L(u_\ell), L(v_\ell)) \in \mathcal{M}$ and $y^{\lambda(u_\ell,D_\ell)}_{L(u_\ell),L(v_\ell)} = L(v_\ell)$. Thus we have $L(v_\ell) = L(u_j)$ and $y^{\lambda(u_\ell,D_\ell)}_{L(u_\ell),L(v_\ell)} = L(u_j)$. Note that $\lambda(v_\ell, D_\ell) = 1 + \lambda(u_\ell, D_\ell) = 1 + \lambda(v_j, D_\ell)$. Combining with $\lambda(u_j, D_j) = \lambda(v_j, D_j) - 1$, we have $\lambda(u_\ell, D_\ell) = \lambda(u_j, D_j)$. Combining with $y^{\lambda(u_\ell,D_\ell)}_{L(u_\ell),L(v_\ell)} = L(u_j)$, we have $y^{\lambda(u_j,D_j)}_{L(u_\ell),L(v_\ell)} = L(u_j)$. This is contradicted with $y^{\lambda(u_j,D_j)}_{L(u_j),L(v_j)} = L(v_j)$.

□

The following two propositions will be used in the proof of Lemma 3.4. The first proposition is an easy observation, and the second one is a direct application of the Cauchy-Schwarz inequality. For the sake of completeness, we present their proof in the appendix.

Proposition 3.2. *Given any wdag D, there exists a set $\mathcal{P}$ of disjoint $\mathcal{M}$-reversible arcs*[6] *such that: for each $i \in \mathcal{M}$,*

$$|\{v : \exists u \text{ such that } u \to v \text{ or } v \to u \text{ is in } \mathcal{P}\} \cap \{v : L(v) = i\}| \ge \frac{1}{2} \cdot \mathcal{V}(D, i).$$

[6] We say two arc $u \to v$ and $u' \to v'$ are disjoint if their node sets are disjoint, i.e. $\{u, v\} \cap \{u', v'\} = \emptyset$.

Proposition 3.3. *Suppose X, Y and Z are three independent random variables, A is an event determined by $\{X, Y\}$, and A' is an event determined by $\{Y, Z\}$. Let X_1, Y_1, Y_2, Z_1 be four independent samples of X, Y, Y, Z, respectively. Then the following holds with probability at most $\mathbb{P}(A)\mathbb{P}(A') - \mathbb{P}(A \cap A')^2$:*

- *A is true on (X_1, Y_1), A' is true on (Y_2, Z_1), and*
- *either A is false on (X_1, Y_2) or A' is false on (Y_1, Z_1).*

Now, we are ready to show Lemma 3.4.

Lemma 3.4. *For each pwdag D,*

$$\mathbb{P}[D \sim (\boldsymbol{X}, \boldsymbol{Y})] \leq \left(\prod_{v \in \overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v \in \mathcal{V}(D)} p'_{L(v)}\right).$$

Proof. Let $\mathcal{P}$ be the set of disjoint $\mathcal{M}$-reversible arcs defined in Proposition 3.2. Let $V(\mathcal{P})$ denote the set of nodes which appears in $\mathcal{P}$, and $\overline{V(\mathcal{P})}$ consists of the other nodes. Proposition 3.2 says that for each $i \in \mathcal{M}$,

$$V(\mathcal{P}) \cap \{v : L(v) = i\}| \geq \frac{1}{2} \cdot \mathcal{V}(D, i).$$

For each $v \in \overline{V(\mathcal{P})}$, let B_v denote the event that $A_{L(v)}$ holds on $\boldsymbol{X}_v$. It is easy to see that $\mathbb{P}[B_v] = p_{L(v)}$. Besides,

Claim 3.5. *If $D \sim (\boldsymbol{X}, \boldsymbol{Y})$, then B_v holds for each $v \in \overline{V(\mathcal{P})}$.*

Proof. Note that $\boldsymbol{X}_v$ are the assignments of $\text{vbl}(A_{L(v)})$ just before the time that the MT algorithm picks the event corresponding to v to resample. MT algorithm decides to pick $A_{L(v)}$ only if $A_{L(v)}$ holds. Hence $A_{L(v)}$ must hold on $\boldsymbol{X}_v$. □

Let $u \longrightarrow v$ be an arc in $\mathcal{P}$, where $L(u) = i$ and $L(v) = i'$. Then by the definition of $\mathcal{P}$, we have $u \longrightarrow v$ is reversible in D. Let D' be the wdag obtained by reversing the direction of $u \longrightarrow v$ in D. Recalling the definition of $\boldsymbol{X}_{D',v}$, one can verify that

$$\boldsymbol{X}_{D',u} := \left\{X_{j,\mathcal{L}(v,j)} : X_j \in \text{vbl}\left(A_i\right) \cap \text{vbl}\left(A_{i'}\right)\right\} \cup \left\{X_{j,\mathcal{L}(u,j)} : X_j \in \text{vbl}\left(A_i\right) \setminus \text{vbl}\left(A_{i'}\right)\right\}$$

and

$$\boldsymbol{X}_{D',v} := \left\{X_{j,\mathcal{L}(u,j)} : X_j \in \text{vbl}\left(A_i\right) \cap \text{vbl}\left(A_{i'}\right)\right\} \cup \left\{X_{j,\mathcal{L}(v,j)} : X_j \in \text{vbl}\left(A_{i'}\right) \setminus \text{vbl}\left(A_i\right)\right\}.$$

For simplicity, let $\lambda := \lambda(u, D)$. We define $B_{u,v}$ to be the event that the following hold:

(a) A_i holds on X_u, and $A_{i'}$ holds on X_v;
(b) If $Y^{\lambda}_{i,i'} = i'$, then either A_i is false on $\boldsymbol{X}_{D',u}$ or A'_i is false on $\boldsymbol{X}_{D',v}$.

Conditioned on that $Y^{\lambda}_{i,i'} = i$, $B_{u,v}$ happens with probability $p_i p_{i'}$. Condition on that $Y^{\lambda}_{i,i'} = i'$, by using Proposition 3.3, $B_{u,v}$ happens with probability at most $p_i p_{i'} - \delta^2_{i,i'}$. Thus,

$$\begin{aligned}\mathbb{P}[B_{u,v}] &\leq \mathbb{P}[Y^{\lambda}_{i,i'} = i]p_i p_{i'} + \mathbb{P}[Y^{\lambda}_{i,i'} = i']\left(p_i p_{i'} - \delta^2_{i,i'}\right) \leq \frac{1}{2} \cdot p_i p_{i'} + \frac{1}{2} \cdot \left(p_i p_{i'} - \delta^2_{i,i'}\right) \\ &\leq p_i p_{i'}(1 - 2c_i)(1 - 2c_{i'}).\end{aligned}$$

Claim 3.6. *If $D \sim (\boldsymbol{X}, \boldsymbol{Y})$, then $B_{u,v}$ holds for each $u \longrightarrow v$ in $\mathcal{P}$.*

Proof. Suppose $D \sim (\boldsymbol{X}, \boldsymbol{Y})$. Similar to the argument in Claim 3.5, we can see that Item (a) holds. In the following, we show Item (b) holds.

By contradiction, assume $Y^{\lambda}_{i,i'} = i'$, A_i holds on $\boldsymbol{X}_{D',u}$, and $A_{i'}$ holds on $\boldsymbol{X}_{D',v}$. Then, we have $u \longrightarrow v$ in D is inconsistent with $\boldsymbol{Y}$ and $D' \sim \boldsymbol{X}$. Thus, $\varphi(D, u, v) \sim \boldsymbol{X}$ since $\varphi(D, u, v)$ is a prefix of D'. By definition, we have $D \nsim (\boldsymbol{X}, \boldsymbol{Y})$, a contradiction. □

Since the events $\{B_v : v \in \overline{V(\mathcal{P})}\}$ and $\{B_{u,v} : u \to v \text{ is in } \mathcal{P}\}$ depend on distinct entries of $\boldsymbol{X}$ and $\boldsymbol{Y}$, they are mutually independent. Therefore,

$$\begin{aligned}
\mathbb{P}\left[D \sim (\boldsymbol{X},\boldsymbol{Y})\right] &\le \mathbb{P}\left[\left(\bigcap_{w\in\overline{V(\mathcal{P})}} B_w\right)\bigcap\left(\bigcap_{u\to v \text{ is in } \mathcal{P}} B_{u,v}\right)\right] = \left(\prod_{w\in\overline{V(\mathcal{P})}} \mathbb{P}(B_w)\right)\left(\prod_{u\to v \text{ is in } \mathcal{P}} \mathbb{P}(B_{u,v})\right)\\
&\le \left(\prod_{w\in\overline{V(\mathcal{P})}} p_{L(w)}\right)\left(\prod_{u\to v \text{ is in } \mathcal{P}} p_{L(u)}p_{L(v)}\left(1-2c_{L(u)}\right)\left(1-2c_{L(u)}\right)\right)\\
&= \left(\prod_{v \text{ in } D} p_{L(v)}\right)\cdot\left(\prod_{i\in[m]} (1-2c_i)^{|\mathcal{P}\cap\{v:L(v)=i\}|}\right)\\
&\le \left(\prod_{v \text{ in } D} p_{L(v)}\right)\cdot\left(\prod_{i\in[m]} (1-2c_i)^{|\mathcal{V}(i)|/2}\right) \le \left(\prod_{v \text{ in } D} p_{L(v)}\right)\cdot\left(\prod_{i\in[m]} (1-c_i)^{|\mathcal{V}(i)|}\right)\\
&= \left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right).
\end{aligned}$$

□

Now we are ready to prove the main theorem of this subsection.

Theorem 3.7. $\mathbb{E}[T] \le \sum_{D\in\mathcal{D}(G_D)} \left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right)$.

Proof. First, according to Lemmas 3.1 and 3.4,

$$\begin{aligned}
\mathbb{P}[\chi_{\mathcal{D}(i,r)}] \le \mathbb{P}[\mathcal{D}(i,r) \sim \boldsymbol{X}] = \mathbb{P}[\mathcal{D}(i,r) \sim (\boldsymbol{X},\boldsymbol{Y})] &\le \sum_{D\in\mathcal{D}(i,r)} \mathbb{P}[D \sim (\boldsymbol{X},\boldsymbol{Y})]\\
&\le \sum_{D\in\mathcal{D}(i,r)}\left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right).
\end{aligned}$$

Then, by Fact 2.3 and the above inequality, we have

$$\begin{aligned}
\mathbb{E}[T] = \sum_{i\in[m]}\sum_{r\in\mathbb{N}^+}\mathbb{P}[\chi_{\mathcal{D}(i,r)}] &\le \sum_{i\in[m]}\sum_{r\in\mathbb{N}^+}\sum_{D\in\mathcal{D}(i,r)}\left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right)\\
&\le \sum_{D\in\mathcal{D}(G_D)}\left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right).
\end{aligned}$$

□

3.2. **Mapping between wdags.** In this section, we will prove Theorem 3.13, which provides a upper bound of $\mathbb{E}[T]$ in terms of $\boldsymbol{p}^-$.

Definition 3.8 (Homomorphic dependency graph)**.** *Given a dependency graph $G_D = ([m], E_D)$ and a matching $\mathcal{M}$ of G_D, we define a graph $G^{\mathcal{M}} = (V^{\mathcal{M}}, E^{\mathcal{M}})$ homomorphic to G_D respected to $\mathcal{M}$ as follows.*

- $V^{\mathcal{M}} = [m] \setminus \{i_0, i_1 : (i_0, i_1) \in \mathcal{M}\} \cup \{i_0^{\uparrow}, i_0^{\downarrow}, i_1^{\uparrow}, i_1^{\downarrow} : (i_0, i_1) \in \mathcal{M}\}$;
- $\forall (i_0, i_1) \in E_D$, *each pair of vertices in* $\{i_0, i_1, i_0^{\uparrow}, i_0^{\downarrow}, i_1^{\uparrow}, i_1^{\downarrow}\} \cap V^{\mathcal{M}}$ *are connected in* $G^{\mathcal{M}}$.

Besides, we associate a probability vector $\boldsymbol{p}^{\mathcal{M}}$ with $G^{\mathcal{M}}$ as follows:

$$\forall v \in V^{\mathcal{M}}: \quad p_v^{\mathcal{M}} = \begin{cases} p'_i & \text{if } v = i^{\uparrow} \text{ for some } i \in [m],\\ p_i^- - p'_i & \text{if } v = i^{\downarrow} \text{ for some } i \in [m],\\ p_i^- & \text{otherwise, } v = i \text{ for some } i \in [m].\end{cases}$$

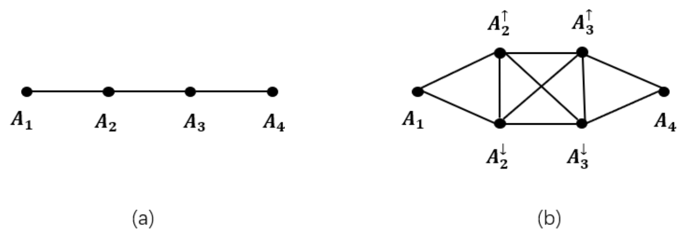

FIGURE 3. (a) a dependency graph G_D; (b) the $G^{\mathcal{M}}$ when $\mathcal{M} = \{(2,3)\}$.

In fact, $(G_D, \boldsymbol{p}^-)$ and $(G^{\mathcal{M}}, \boldsymbol{p}^{\mathcal{M}})$ are essentially the same: suppose $\mathcal{A} \sim (G_D, \boldsymbol{p}^-)$, then for each $i \in \mathcal{M}$, we view A_i as the union of two mutually exclusive events $A_{i\uparrow} \cup A_{i\downarrow}$ whose probabilities are p'_i and $p^- - p'_i$ respectively. Such a representation of $\mathcal{A}$ is of the setting $(G^{\mathcal{M}}, \boldsymbol{p}^{\mathcal{M}})$.

We have the following proposition, whose proof can be found in the appendix.

Proposition 3.9. $\sum_{D' \in \mathcal{D}(G^{\mathcal{M}})} \prod_{v' \text{ in } D'} p^{\mathcal{M}}_{L'(v')} = \sum_{D \in \mathcal{D}(G_D)} \prod_{v \text{ in } D} p^-_{L(v)}$.

Given a pwdag $D = (V, E, L)$, recall that $\mathcal{V}(D)$ is the set of nodes of $\mathcal{M}$-reversible arcs in D. Define $\mathscr{M}(D) \triangleq \{v : L(v) \in \mathcal{M}\}$ to be the set of nodes v in D where $L(v)$ is contained in an edge in $\mathcal{M}$. Obviously, $\mathcal{V}(D) \subseteq \mathscr{M}(D)$. For simplicity of notations, we will omit D from the notations if D is clear from the context.

Given a pwdag $D = (V, E, L)$, we use $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\}$ to represent a partition of $\mathscr{M}(D)$ where $\mathcal{V} \subseteq \mathcal{S}_1$ (some of these four sets are possibly empty). Let $\psi(D)$ denote the set consisting of all such partitions. The formal definition is as follows.

Definition 3.10 (Partition)**.** *Given a pwdag $D = (V, E, L)$ of G_D, define*

$$\psi(D) \triangleq \{\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\} : \mathcal{V} \subseteq \mathcal{S}_1 \text{ and } \mathscr{M} = \mathcal{S}_1 \sqcup \mathcal{S}_2 \sqcup \mathcal{S}_3 \sqcup \mathcal{S}_4\}.$$

Given a wdag D, there may be two or more topological ordering of D. We fix an arbitrary topological ordering, and denote it by π_D. In the following, we define an injection h from $\{(D, \mathcal{S}) : D \in \mathcal{D}(G_D), \mathcal{S} \in \psi(D)\}$ to $\mathcal{D}(G^{\mathcal{M}})$.

Definition 3.11. *Given a pwdag D and $\mathcal{S} \in \psi(D)$, define $h(D, \mathcal{S})$ to be a directed graph $D' = (V', E', L')$ constructed as follows.*

Constructing V'. $V' = V'_1 \sqcup V'_2$ where $|V'_1| = |V|$ and $|V'_2| = |\mathcal{S}_3 \cup \mathcal{S}_4|$. For convenience of presentation, we fix two bijections $f : V \to V'_1$ and $f^* : \mathcal{S}_3 \cup \mathcal{S}_4 \to V'_2$ to name nodes in V'. In order to distinguish between nodes in D and those in D', we will always use u, v, w to represent the nodes of D and u', v', w' to present the nodes of D'. Given $v' \in V'$, we use $g(v')$ to denote the unique node $v \in V$ such that $f(v) = v'$ (if $v' \in V'_1$) or $f^*(v) = v'$ (if $v' \in V'_2$).

Description of L'. For each node $v' \in V'_1$, where $v' = f(v)$,

$$L'(v') = \begin{cases} (L(v))^{\uparrow}, & \text{if } v \in \mathcal{S}_1, \\ (L(v))^{\downarrow}, & \text{if } v \in \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4, \\ L(v), & \text{otherwise, } v \notin \mathscr{M}. \end{cases} \tag{2}$$

For each node $v' \in V'_2$, assuming $v \in \mathcal{S}_3 \cup \mathcal{S}_4$ is the node such that $v' = f^*(v)$ and $i \in [m]$ is the node such that $((L(v), i) \in \mathcal{M}$,

$$L'(v') = \begin{cases} i^{\uparrow}, & \text{if } v \in \mathcal{S}_3, \\ i^{\downarrow}, & \text{otherwise, } v \in \mathcal{S}_4. \end{cases} \tag{3}$$

Constructing E'. $E' = E'_1 \sqcup E'_2$ where $E'_1 = \{f^*(v) \to f(v) : v \in \mathcal{S}_3 \cup \mathcal{S}_4\}$ and

$$E'_2 = \{u' \to v' : \big((L'(u') = L'(v')\big) \vee \big((L'(u'), L'(v')) \in E^{\mathcal{M}}\big)\big) \wedge (g(u') < g(v') \text{ in } \pi_D)\}.$$

Theorem 3.12. *$h(\cdot, \cdot)$ is an injection from $\{(D, \mathcal{S}) : D \in \mathcal{D}(G_D), \mathcal{S} \in \psi(D)\}$ to $\mathcal{D}(G^{\mathcal{M}})$.*

The proof of Theorem 3.12 is in the appendix. Now we can prove the main theorem of this subsection.

Theorem 3.13. $\sum_{D\in\mathcal{D}(G_D)}\left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right) \le \sum_{D\in\mathcal{D}(G_D)}\prod_{v \text{ in } D} p^-_{L(v)}$.

Proof. For each $i\in[m]$ where $(i,j)\in\mathcal{M}$, let

$$q_i^1 \triangleq p'_i,\quad q_i^2 \triangleq p_i^- - p'_i,\quad q_i^3 \triangleq (p_i^- - p'_i)p'_j,\quad \text{and} \quad q_i^4 \triangleq (p_i^- - p'_i)(p_j^- - p'_j).$$

According to Fact 2.7, $q_i^1+q_i^2+q_i^3+q_i^4 = p_i^- + p_j^-(p_i^- - p'_i) \ge p_i$.

Given $D=(V,E,L)\in\mathcal{D}(G_D)$ and $\mathcal{S}\in\psi(D)$, let $D'=h(D,\mathcal{S})$. For each v in D where $(L(v),j)\in\mathcal{M}$ for some $j\in[m]$, according to the definition of $\boldsymbol{p}^{\mathcal{M}}$, (2), and (3), we have that

- if $v\in\mathcal{S}_1$, then $p^{\mathcal{M}}_{L'(f(v))} = p'_{L(v)} = q^1_{L(v)}$;
- if $v\in\mathcal{S}_2$, then $p^{\mathcal{M}}_{L'(f(v))} = p^-_{L(v)} - p'_{L(v)} = q^2_{L(v)}$;
- if $v\in\mathcal{S}_3$, then $p^{\mathcal{M}}_{L'(f(v))}\cdot p^{\mathcal{M}}_{L'(f^*(v))} = (p^-_{L(v)} - p'_{L(v)})p'_j = q^3_{L(v)}$;
- if $v\in\mathcal{S}_4$, then $p^{\mathcal{M}}_{L'(f(v))}\cdot p^{\mathcal{M}}_{L'(f^*(v))} = (p^-_{L(v)} - p'_{L(v)})(p^-_j - p'_j) = q^4_{L(v)}$.

Moreover, for each $u\in V\setminus\mathscr{M}(D) = \overline{\mathcal{V}(D)}\setminus\mathscr{M}(D)$, we have $p^{\mathcal{M}}_{L'(f(v))} = p_{L(v)}$. Thus, for each $\mathcal{S}\in\psi(D)$,

$$\begin{aligned}\prod_{v' \text{ in } h(D,\mathcal{S})} p^{\mathcal{M}}_{L'(v')} &= \prod_{v\in\overline{\mathcal{V}}\setminus\mathscr{M}} p_{L(v)}\prod_{v\in\mathcal{S}_1} q^1_{L(v)}\prod_{v\in\mathcal{S}_2} q^2_{L(v)}\prod_{v\in\mathcal{S}_3} q^3_{L(v)}\prod_{v\in\mathcal{S}_4} q^4_{L(v)}\\ &= \prod_{v\in\overline{\mathcal{V}}\setminus\mathscr{M}} p_{L(v)}\prod_{v\in\mathcal{V}} p'_{L(v)}\prod_{v\in\mathcal{S}_1\setminus\mathcal{V}} q^1_{L(v)}\prod_{v\in\mathcal{S}_2} q^2_{L(v)}\prod_{v\in\mathcal{S}_3} q^3_{L(v)}\prod_{v\in\mathcal{S}_4} q^4_{L(v)}.\end{aligned}$$

So

$$\begin{aligned}\sum_{\mathcal{S}\in\psi(D)}\prod_{v' \text{ in } h(D,\mathcal{S})} p^{\mathcal{M}}_{L'(v')} &= \sum_{\mathcal{S}\in\psi(D)}\prod_{v\in\overline{\mathcal{V}}\setminus\mathscr{M}} p_{L(v)}\prod_{v\in\mathcal{V}} p'_{L(v)}\prod_{v\in\mathcal{S}_1\setminus\mathcal{V}} q^1_{L(v)}\prod_{v\in\mathcal{S}_2} q^2_{L(v)}\prod_{v\in\mathcal{S}_3} q^3_{L(v)}\prod_{v\in\mathcal{S}_4} q^4_{L(v)}\\ &= \prod_{v\in\overline{\mathcal{V}}\setminus\mathscr{M}} p_{L(v)}\prod_{v\in\mathcal{V}} p'_{L(v)}\sum_{\mathcal{S}\in\psi(D)}\prod_{v\in\mathcal{S}_1\setminus\mathcal{V}} q^1_{L(v)}\prod_{v\in\mathcal{S}_2} q^2_{L(v)}\prod_{v\in\mathcal{S}_3} q^3_{L(v)}\prod_{v\in\mathcal{S}_4} q^4_{L(v)}\\ &= \prod_{v\in\overline{\mathcal{V}}\setminus\mathscr{M}} p_{L(v)}\prod_{v\in\mathcal{V}} p'_{L(v)}\prod_{v\in\mathscr{M}\setminus\mathcal{V}}\left(q^1_{L(v)}+q^2_{L(v)}+q^3_{L(v)}+q^4_{L(v)}\right)\\ &\ge \prod_{v\in\overline{\mathcal{V}}\setminus\mathscr{M}} p_{L(v)}\prod_{v\in\mathcal{V}} p'_{L(v)}\prod_{v\in\mathscr{M}\setminus\mathcal{V}} p_{L(v)}\\ &= \prod_{v\in\overline{\mathcal{V}}} p_{L(v)}\prod_{v\in\mathcal{V}} p'_{L(v)},\end{aligned}$$

where the third equality is according to the definition of $\psi(D)$. Finally,

$$\begin{aligned}\sum_{D\in\mathcal{D}(G_D)}\left(\prod_{v\in\overline{\mathcal{V}}(D)} p_{L(v)}\right)\left(\prod_{v\in\mathcal{V}(D)} p'_{L(v)}\right) &\le \sum_{D\in\mathcal{D}(G_D)}\sum_{\mathcal{S}\in\psi(D)}\prod_{v' \text{ in } h(D,\mathcal{S})} p^{\mathcal{M}}_{L'(v')} \le \sum_{D'\in\mathcal{D}(G^{\mathcal{M}})}\prod_{v \text{ in } D'} p^{\mathcal{M}}_{L'(v')}\\ &= \sum_{D\in\mathcal{D}(G_D)}\prod_{v \text{ in } D} p^-_{L(v)},\end{aligned}$$

where the second inequality is due to Theorem 3.12 and the equality is by Proposition 3.9. □

3.3. **Putting all things together.**

The following lemma is implicitly proved in [KS11].

Lemma 3.14 ([KS11])**.** *For any undirected graph $G_D=([m],E_D)$ and probability vector $\boldsymbol{p}\in\mathcal{I}_a(G_D)/(1+\varepsilon)$, $\sum_{i\in[m]}\frac{q_{\{i\}}(G_D,\boldsymbol{p})}{q_\emptyset(G_D,\boldsymbol{p})}\le m/\varepsilon$.*

Theorem 1.6 (restated)**.** *For any $\mathcal{A}\sim(G_D,\boldsymbol{p},\mathcal{M},\boldsymbol{\delta})$, if $(1+\varepsilon)\cdot\boldsymbol{p}^-\in\mathcal{I}_a(G_D)$, then the expected number of resampling steps performed by MT algorithm is most m/ε, where m is the number of events in $\mathcal{A}$.*

Proof. Fix any such $\mathcal{A}$. We have that

$$\mathbb{E}[T] \le \sum_{D \in \mathcal{D}(G_D)} \prod_{v \text{ in } D} p^-_{L(v)} \le \frac{\sum_{i\in[m]} q_{\{i\}}(G_D, \boldsymbol{p}^-)}{q_{\emptyset}(G_D, \boldsymbol{p}^-)} \le \frac{m}{\varepsilon},$$

where the first inequality is by Theorems 3.7 and 3.13, the second inequality is due to Theorem 4 in [KS11], and the last inequality is according to Lemma 3.14. □

4. Lower bound on the amount of intersection

In order to explore how far beyond Shearer's bound MT algorithm is still efficient in general, we provide a lower bound on the amount of intersection between dependent events for general instances (Theorem 4.1).

We first introduce some notations. Given a bipartite graph $G_B = ([m], [n], E_B)$, we call the vertex $i \in [m]$ left vertex and the vertex $j \in [n]$ right vertex. We call G_B *linear*[7] if any two left vertices in $[m]$ share at most one common neighbor in $[n]$. Let $\Delta_D(G_B)$ denote the maximum degree of $G_D(G_B)$, and $\Delta_B(G_B)$ denote the maximum degree of the left vertices in G_B. If G_B is clear from the context, we may omit G_B from these notations. In addition, for a bipartite graph $G = (L \subset [m], R, E)$ and a probability vector $\boldsymbol{p} \in (0,1)^m$, we define[8]

$$F(G, \boldsymbol{p}) \triangleq \frac{\left(\min_{i\in L} p_i\right)^2 \cdot \left(-|\cup_{i\in L} \mathcal{N}_G(i)| + \sum_{i\in L} |\mathcal{N}_G(i)| \cdot p_i^{1/|\mathcal{N}_G(i)|}\right)}{\sqrt{|L|} \cdot \Delta_D(G) \cdot \Delta_B(G)^2}.$$

and $F^+(G, \boldsymbol{p}) \triangleq \max\{F(G, \boldsymbol{p}), 0\}$.

We use $\mathcal{A} \sim (G_B, \boldsymbol{p})$ to denote that (i) G_B is an event-variable graph of $\mathcal{A}$ and (ii) the probability vector of $\mathcal{A}$ is $\boldsymbol{p}$. Let $\mathcal{M} = \{(i_1, i_1'), (i_2, i_2'), \cdots\}$ be a matching of $G_D(G_B)$, and $\boldsymbol{\delta} = (\delta_{i_1,i_1'}, \delta_{i_2,i_2'}, \cdots) \in (0,1)^{|\mathcal{M}|}$ be another probability vector. We say that an event set $\mathcal{A}$ is of the setting $(G_B, \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$, and write $\mathcal{A} \sim (G_B, \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$, if $\mathcal{A} \sim (G_B, \boldsymbol{p})$ and $\mathbb{P}(A_i \cap A_{i'}) \ge \delta_{i,i'}$ for each pair $(i, i') \in \mathcal{M}$.

We call an event A *elementary*, if A can be written as $(X_{i_1} \in S_{i_1}) \wedge (X_{i_2} \in S_{i_2}) \wedge \cdots \wedge (X_{i_k} \in S_{i_k})$ where $S_{i_1}, \cdots, S_{i_k}$ are subsets of the domains of variables. We call an event set $\mathcal{A}$ *elementary* if all events in $\mathcal{A}$ are elementary.

Theorem 4.1. *Let $G_B = ([m], [n], E_B)$ be a bipartite graph, $\boldsymbol{p} \in (0,1]^m$ be a probability vector, and $L_1, L_2, \cdots, L_t$ be a collection of disjoint subsets of $[m]$. For each $k \in [t]$, let G_k denote the induced subgraph on $L_k \cup \left(\cup_{i\in L_k} \mathcal{N}_{G_B}(i)\right)$ and E_k denote the edge set of $G_D(G_k)$. If all G_k's are linear, then the following holds.*

If $\mathcal{A} \sim (G_B, \boldsymbol{p})$, then there is a matching $\mathcal{M}$ of $G_D(G_B)$ satisfying that $\sum_{(i,i')\in\mathcal{M}\cap E_k} \mathbb{P}(A_i \cap A_{i'})^2 \ge (F^+(G_k, \boldsymbol{p}))^2$ for any k.

The proof of Theorem 4.1 mainly consists of two parts. First, we show that there is an elementary event set which approximately achieves the minimum amount of intersection between dependent events (Lemma 4.2). Then, for elementary event sets, by applying AM-GM inequality, we obtain a lower bound on the total amount of overlap on common variables, which further implies a lower bound on the amount of intersection between dependent events (Lemma 4.5).

Lemma 4.2. *Let $G_B = ([m], [n], E_B)$ be a linear bipartite graph, E_D be the edge set of $G_D(G_B)$, and $\boldsymbol{p} \in (0,1]^m$ is a probability vector. Let γ denote the minimum $\sum_{(i_0,i_1)\in E_D} \mathbb{P}[A_{i_0} \cap A_{i_1}]$ among all event sets $\mathcal{A} = (A_1, \cdots, A_m) \sim (G_B, \boldsymbol{p})$. Then there is an elementary event set $\mathcal{A}'$ such that $\sum_{(i_0,i_1)\in E_D} \mathbb{P}[A'_{i_0} \cap A'_{i_1}] \le (\Delta_B(G_B))^2 \cdot \gamma$.*

Proof. For simplicity, we let $\Delta \triangleq \Delta_B(G_B)$. Without loss of generality, we assume that each random variable X_i is uniformly distributed over $[0,1]$. Let $\mathcal{A} \sim (G_B, \boldsymbol{p})$ be an event set where $\sum_{(i_0,i_1)\in E_D} \mathbb{P}[A_i \cap$

[7] The notion is not arbitrary. The bipartite graph G_B can be represented by a hypergraph in a natural way: each right vertex j is represented by a node v_j in the hypergraph, each left vertex i is represented by a hyperedge e_i, and v_j is in e_i if and only if $(i, j) \in E_B$. A hypergraph is called linear if any two hyperedges share at most one node.

[8] It is possible that $F(G, \boldsymbol{p}) < 0$.

$A_j] = \gamma$. We will replace A_i with an elementary A'_i one by one for each $i = 1, 2, \cdots, m$, so that the resulted event set $\mathcal{A}'$ satisfies $\sum_{(i_0,i_1)\in E_D} \mathbb{P}[A'_{i_0} \cap A'_{i_1}] \le \Delta^2 \cdot \sum_{(i_0,i_1)\in E_D} \mathbb{P}[A_{i_0} \cap A_{i_1}] = \Delta^2 \cdot \gamma$.

More precisely, fix $i \in [m]$ and suppose $A_1, \cdots, A_{i-1}$ have been replaced with elementary events $A'_1, \cdots, A'_{i-1}$ respectively. For simplicity of notations, for any pair $i_0 < i_1$, we abbreviate $\mathbb{P}[A_{i_0} \cap A_{i_1}]$, $\mathbb{P}[A'_{i_0} \cap A_{i_1}]$ and $\mathbb{P}[A'_{i_0} \cap A'_{i_1}]$ to $p_{i_0,i_1}, p'_{i_0,i_1}$ and p''_{i_0,i_1} respectively. Without loss of generality, we assume A_i depends on variables $X_1, X_2, \cdots, X_k$. For every $j \in [k]$, we define

$$P_j(x_j) := \sum_{i_0<i, i_0\in\mathcal{N}_{G_B}(j)} \frac{1}{\Delta} \cdot \mathbb{P}[A'_{i_0} \mid X_j = x_j] + \sum_{i_0>i, i_0\in\mathcal{N}_{G_B}(j)} \mathbb{P}[A_{i_0} \mid X_j = x_j].$$

for $x_j \in [0, 1]$. Without loss of generality, we assume $P_j(\cdot)$ is non-decreasing. Let $\mu : [0, 1]^k \to \{0, 1\}$ be the indicator of A_i, then

$$\int_{x_1,\cdots,x_k} \mu(x_1, \cdots, x_k)\mathrm{d}x_1 \cdots \mathrm{d}x_k = \mathbb{P}[A_i],$$

For each $j \in [k]$, let

$$\mu_j(x_j) := \mathbb{P}[A_i \mid X_j = x_j] = \int_{x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_k} \mu(x_1, \cdots, x_k)\mathrm{d}x_1 \cdots \mathrm{d}x_{j-1}\mathrm{d}x_{j+1} \cdots \mathrm{d}x_k.$$

Noticing that G_B is linear (i.e., any two events share at most one common variable), we have

$$\int_{x_j} P_j(x_j)\mu_j(x_j)\mathrm{d}x_j = \sum_{i_0<i, i_0\in\mathcal{N}_{G_B}(j)} \frac{p'_{i_0,i}}{\Delta} + \sum_{i_0>i, i_0\in\mathcal{N}_{G_B}(j)} p_{i_0,i}. \tag{4}$$

Let A'_i be an elementary event such that it happens if and only if $(x_1, \cdots, x_k) \in [0, q_1] \times \cdots \times [0, q_k]$. Here $q_1, \cdots, q_k$ is a set of positive real numbers satisfying that

(i) $\Pi_{j=1}^k q_j = \mathbb{P}[A_i]$. That is, $\mathbb{P}[A'_i] = \mathbb{P}[A_i]$;
(ii) $\int_{x_1\ge q_1} \mu_1(x_1)\mathrm{d}x_1 = \int_{x_2\ge q_2} \mu_2(x_2)\mathrm{d}x_2 \cdots = \int_{x_k\ge q_k} \mu_k(x_k)\mathrm{d}x_k$.

Claim 4.3. *Such $\{q_1, \cdots, q_k\}$ exists. Thus so does A'_i.*

Proof. We prove a generalized statement in which $\Pi_{j=1}^k q_j$ can be required to be an arbitrary number in $[0, 1]$. Our proof is by induction on k. The base case when $k = 1$ is trivial. Now we assume that for any preset $q' \in (0, 1]$, there exist $\{q_1, \cdots, q_{k-1}\}$ satisfying that

(i) $\Pi_{j=1}^{k-1} q_j = q'$ and
(ii) $\int_{x_1\ge q_1} \mu_1(x_1)\mathrm{d}x_1 = \cdots = \int_{x_{k-1}\ge q_{k-1}} \mu_{k-1}(x_{k-1})\mathrm{d}x_{k-1}$.

Let $f(q')$ denote the minimum $\int_{x_1\ge q_1} \mu_1(x_1)\mathrm{d}x_1$ among all such $\{q_1, \cdots, q_{k-1}\}$'s. It is easy to see that $f(1) = 0$ and f is continuous and non-increasing.

Fix an arbitrary $q \in [0, 1]$. We define $g(q'') := \int_{x_k\ge q/q'} \mu_k(x_k)\mathrm{d}x_k$ for $q'' \in [q, 1]$. Obviously, $g(q) = 0$ and g is continuous and non-decreasing. So there must exist a $q^* \in [q, 1]$ such that $g(q^*) = f(q^*)$. Then let $\{q^*_1, \cdots, q^*_{k-1}\}$ be a set of positive real numbers where

(i) $\Pi_{j=1}^{k-1} q^*_j = q^*$ and
(ii) $f(q^*) = \int_{x_1\ge q^*_1} \mu_1(x_1)\mathrm{d}x_1 = \cdots = \int_{x_{k-1}\ge q^*_{k-1}} \mu_{k-1}(x_{k-1})\mathrm{d}x_{k-1}$.

Let $q^*_k = q/q^*$. It is obvious that $\Pi_{j=1}^k q^*_j = q$ and $f(q^*) = g(q^*) = \int_{x_k\ge q^*_k} \mu_k(x_k)\mathrm{d}x_k$. This completes the induction step. □

Claim 4.4. *For every $j \in [k]$, we have*

$$\sum_{i_0<i, i_0\in\mathcal{N}_{G_B}(j)} \frac{p''_{i_0,i}}{\Delta} + \sum_{i_0>i, i_0\in\mathcal{N}_{G_B}(j)} p'_{i_0,i} \le \sum_{i_0<i, i_0\in\mathcal{N}_{G_B}(j)} p'_{i_0,i} + \Delta \cdot \sum_{i_0>i, i_0\in\mathcal{N}_{G_B}(j)} p_{i_0,i}.$$

Proof. Let $\mu_{i'}, \mu_{i\cap i'}$, and $\mu_{i'\setminus i}$ denote the indicator functions of the events A'_i, $A'_i \cap A_i$, and $A'_i \setminus A_i$ respectively. Since $\mathbb{P}[A'_i] = \mathbb{P}[A_i]$,

$$\int_{x_1 \geq q_1} \mu_1(x_1)\mathrm{d}x_1 + \cdots + \int_{x_k \geq q_k} \mu_k(x_k)\mathrm{d}x_k \geq \mathbb{P}[A_i \setminus A'_i] = \mathbb{P}[A'_i \setminus A_i] = \int_{x_1,\cdots,x_k} \mu_{i'\setminus i}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k.$$

Fix $j \in [k]$, then

$$\int_{x_j \geq q_j} \mu_j(x_j)\mathrm{d}x_j \geq \frac{1}{k} \cdot \int_{x_1,x_2,\cdots,x_k} \mu_{i'\setminus i}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k.$$

Since $P_j(x_j)$ is non-decreasing and $k \leq \Delta$, we have

$$\int_{x_j \geq q_j} P_j(x_j)\mu_j(x_j)\mathrm{d}x_j \geq \frac{1}{\Delta} \cdot \int_{x_1,x_2,\cdots,x_k} P_j(x_j)\mu_{i'\setminus i}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k.$$

According to Equation 4,

$$\begin{aligned}
&\sum_{i_0<i, i_0\in\mathcal{N}_{G_B}(j)} p'_{i_0,i} + \Delta \cdot \sum_{i_0>i, i_0\in\mathcal{N}_{G_B}(j)} p_{i_0,i} = \Delta \cdot \int_{x_j} P_j(x_j)\mu_j(x_j)\mathrm{d}x_j \\
=&\Delta \cdot \int_{x_j \geq q_j} P_j(x_j)\mu_j(x_j)\mathrm{d}x_j + \Delta \cdot \int_{x_j < q_j} P_j(x_j)\mu_j(x_j)\mathrm{d}x_j \\
\geq&\Delta \cdot \int_{x_j \geq q_j} P_j(x_j)\mu_j(x_j)\mathrm{d}x_j + \Delta \cdot \int_{x_1,\cdots,x_k} P_j(x_j)\mu_{i\cap i'}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k \\
\geq&\int_{x_1,\cdots,x_k} P_j(x_j)\mu_{i'\setminus i}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k + \int_{x_1,\cdots,x_k} P_j(x_j)\mu_{i\cap i'}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k \\
=&\int_{x_1,\cdots,x_k} P_j(x_j)\mu_{i'}(x_1,\cdots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k \\
=&\sum_{i_0<i, i_0\in\mathcal{N}_{G_B}(j)} \frac{\mathbb{P}[A'_{i_0} \cap A'_i]}{\Delta} + \sum_{i_0>i, i_0\in\mathcal{N}_{G_B}(j)} \mathbb{P}[A_{i_0} \cap A'_i].
\end{aligned}$$

This completes the proof. □

From Claim 4.4, we have

$$\sum_{i_0<i, i_0\in\mathcal{N}_{G_D}(i)} \frac{p''_{i_0,i}}{\Delta} + \sum_{i_0>i, i_0\in\mathcal{N}_{G_D}(i)} p'_{i_0,i} \leq \sum_{i_0<i, i_0\in\mathcal{N}_{G_D}(i)} p'_{i_0,i} + \Delta \cdot \sum_{i_0>i, i_0\in\mathcal{N}_{G_D}(i)} p_{i_0,i}, \tag{5}$$

By summation over all $i \in [m]$, we finish the proof:

$$\sum_{(i_0,i)\in E_D} \frac{p''_{i_0,i}}{\Delta} \leq \Delta \cdot \sum_{(i_0,i)\in E_D} p_{i_0,i}.$$

□

Lemma 4.5. *Let $G_B = ([m],[n],E_B)$ be a linear bipartite graph and $\boldsymbol{p}$ be a probability vector. Then for any elementary $\mathcal{A} = (A_1,\cdots,A_m) \sim (G_B,\boldsymbol{p})$,*

$$\sum_{(i_0,i_1)\in E_D} \mathbb{P}\left(A_{i_0} \cap A_{i_1}\right) \geq \sqrt{m} \cdot \Delta_D(G_B) \cdot \Delta_B(G_B)^2 \cdot F(G_B,\boldsymbol{p}),$$

where E_D is the edge set of $G_D(G_B)$

Proof. For simplicity of notation, we let $\mathcal{N}$ stand for $\mathcal{N}_{G_B}$. Without loss of generality, we assume that each variable X_i is uniformly distributed over $[0,1]$. As $\mathcal{A}$ is elementary, each A_i can be written as $\bigwedge_{j\in\mathcal{N}(i)}[X_j \in S_i^j]$ where $S_i^j \subset [0,1]$. Let μ be the Lebesgue measure.

On one hand, according to the AM–GM inequality,

$$\sum_{i\in[m]} \sum_{j\in\mathcal{N}(i)} \mu(S_i^j) \geq \sum_{i\in[m]} |\mathcal{N}(i)| \cdot \left(\Pi_{j\in\mathcal{N}(i)}\mu(S_i^j)\right)^{1/|\mathcal{N}(i)|} = \sum_{i\in[m]} |\mathcal{N}(i)| \cdot p_i^{1/|\mathcal{N}(i)|}. \tag{6}$$

On the other hand,

$$\sum_{i\in[m]}\sum_{j\in\mathcal{N}(i)}\mu(S_i^j)=\sum_{j\in[n]}\sum_{i\in\mathcal{N}(i)}\mu(S_i^j)\le n+\sum_{j\in[n]}\sum_{i_0\neq i_1\in\mathcal{N}(j)}\mu\left(S_{i_0}^j\cap S_{i_1}^j\right) \tag{7}$$

By Inequalities 6 and 7 and noticing G_B is linear, we have that

$$\sum_{(i_0,i_1)\in E_D}\sum_{j\in\mathcal{N}(i_0)\cap\mathcal{N}(i_1)}\mu\left(S_{i_0}^j\cap S_{i_1}^j\right)=\sum_{j\in[n]}\sum_{i_0\neq i_1\in\mathcal{N}(j)}\mu\left(S_{i_0}^j\cap S_{i_1}^j\right)\ge\left(\sum_{i\in[m]}|\mathcal{N}(i)|\cdot p_i^{1/|\mathcal{N}(i)|}\right)-n. \tag{8}$$

Moreover, given any $(i_0,i_1)\in E_D$, where $\{j\}=\mathcal{N}(i)\cap\mathcal{N}(i')$, we have that

$$\begin{aligned}\mathbb{P}(A_{i_0}\cap A_{i_1})&\ge\mu\left(S_{i_0}^j\cap S_{i_1}^j\right)\cdot\left(\prod_{k\in\mathcal{N}(i_0)\setminus\{j\}}\mu(S_{i_0}^k)\right)\cdot\left(\prod_{k'\in\mathcal{N}(i_1)\setminus\{j\}}\mu(S_{i_1}^{k'})\right)\\&\ge\mu\left(S_{i_0}^j\cap S_{i_1}^j\right)\cdot p_{i_0}\cdot p_{i_1}.\end{aligned} \tag{9}$$

Finally, combining (8) with (9), we concludes that

$$\begin{aligned}\sum_{(i_0,i_1)\in E_D}\mathbb{P}(A_{i_0}\cap A_{i_1})&\ge\sum_{(i_0,i_1)\in E_D}\sum_{j\in\mathcal{N}(i_0)\cap\mathcal{N}(i_1)}\mu\left(S_{i_0}^j\cap S_{i_1}^j\right)\cdot p_{i_0}\cdot p_{i_1}\\&\ge\left(\min_{i\in[m]}p_i\right)^2\left(\sum_i|\mathcal{N}(i)|\cdot p_i^{1/|\mathcal{N}(i)|}-n\right)\\&=\sqrt{m}\cdot\Delta_D(G_B)\cdot\Delta_B(G_B)^2\cdot F(G_B,\boldsymbol{p}).\end{aligned}$$

□

The following lemma is a special case of Theorem 4.1 where $t=1$ and $L_1=[m]$. In fact, Theorem 4.1 is proved by applying Lemma 4.6 to each G_k separately.

Lemma 4.6. *Let $G_B=([m],[n],E_B)$ be a linear bipartite graph and $\boldsymbol{p}$ be a probability vector. If $\mathcal{A}\sim(G_B,\boldsymbol{p})$, then $\mathcal{A}\sim(G_B,\boldsymbol{p},\mathcal{M},\boldsymbol{\delta})$ for some matching $\mathcal{M}$ of $G_D(G_B)$ and some $\boldsymbol{\delta}\in(0,1)^{|\mathcal{M}|}$ satisfying that $\sum_{(i,i')\in\mathcal{M}}\delta_{i,i'}^2\ge\left(F^+(G_B,\boldsymbol{p})\right)^2$.*

Proof. Given an instance $\mathcal{A}\sim(G_B,\boldsymbol{p})$, we construct such a $\mathcal{M}$ greedily as follows.

We maintain two sets E and $\mathcal{M}$, which are initialized as E_D and $\emptyset$ respectively. We do the following iteratively until E becomes empty: select a edge (i_0,i_1) with maximum $\mathbb{P}(A_{i_0}\cap A_{i_1})$ from E, add (i_0,i_1) to $\mathcal{M}$, and delete all edges connecting i_0 or i_1 from E (including (i_0,i_1)).

Let Δ_D and Δ_B denote $\Delta_D(G_B)$ and $\Delta_B(G_B)$ respectively. In each iteration, at most $2\Delta_D$ edges are deleted from E and for each deleted edge (i,i'), $\mathbb{P}(A_i\cap A_{i'})^2\le\mathbb{P}(A_{i_0}\cap A_{i_1})^2$. Based on this observation, it is easy to see that

$$\sum_{(i_0,i_1)\in\mathcal{M}}\mathbb{P}(A_{i_0}\cap A_{i_1})^2\ge\frac{1}{2\Delta_D}\sum_{(i,i')\in E_D}\mathbb{P}(A_i\cap A_{i'})^2. \tag{10}$$

Moreover, according to Lemma 4.2 and 4.5, it has that

$$\sum_{(i,i')\in E_D}\mathbb{P}(A_i\cap A_{i'})^2\ge\frac{1}{|E_D|}\cdot\left(\sum_{(i,i')\in E_D}\mathbb{P}(A_i\cap A_{i'})\right)^2\ge\frac{m\cdot\Delta_D^2\cdot(F^+(G_B,\boldsymbol{p}))^2}{|E_D|}, \tag{11}$$

By combining Inequality 10 and 11, setting $\delta_{i,i'}=\mathbb{P}(A_i\cap A_{i'})$, and noting $2|E_D|\le m\Delta_D$, we finish the proof. □

Proof of Theorem 4.1. For each $k\in[t]$, by applying Lemma 4.6 to G_k, we have that $\mathcal{A}\sim(G_B,\boldsymbol{p},\mathcal{M}_k,\boldsymbol{\delta}_k)$ for some matching $\mathcal{M}_k\subseteq E_k$ and some $\boldsymbol{\delta}_k$ where $\sum_{(i,i')\in\mathcal{M}_k}\delta_{i,i'}^2\ge\left(F^+(G_k,\boldsymbol{p})\right)^2$. Note that E_k's are disjoint with each other, so $\mathcal{M}_1\cup\mathcal{M}_2\cup\cdots\cup\mathcal{M}_t$ is still a matching. By letting $\mathcal{M}=\mathcal{M}_1\cup\mathcal{M}_2\cup\cdots\cup\mathcal{M}_t$ and $\boldsymbol{\delta}=(\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_t)$, we conclude the theorem. □

Remark 4.7. *Given a bipartite graph G, its* simplified graph *is defined to be obtained from G by deleting all the right nodes which only have one neighbor and combining all the right nodes with the same neighbor set. Notice that if G is linear, so is its simplified graph.*

Theorem 4.1 can be slightly generalized: it is sufficient that the simplified graph of G_k instead of G_k itself is linear.

5. The Moser-Tardos algorithm is beyond Shearer's bound

In this section, we prove Theorem 1.5. Given a dependency graph G_D, a vector $\boldsymbol{p}$ and a chordless cycle C in G_D, define

$$r(G_D, \boldsymbol{p}, C) \triangleq |C| \cdot \left(\min_{j \in C} p_j\right)^4 \cdot \left(\frac{2\sum_{j\in C}\sqrt{p_j}}{|C|} - 1\right)^2.$$

and

$$r^+(G_D, \boldsymbol{p}, C) \triangleq |C| \cdot \left(\min_{j \in C} p_j\right)^4 \cdot \left(\max\left\{\frac{2\sum_{j\in C}\sqrt{p_j}}{|C|} - 1, 0\right\}\right)^2.$$

Then Theorem 1.5 is obvious by Lemmas 5.1 and 5.2.

Lemma 5.1. *Given G_D, $\boldsymbol{p}$ and $\varepsilon > 0$, let $C_1, C_2, \cdots, C_\ell$ be any disjoint chordless cycles in G_D. If*

$$d((1+\varepsilon)\boldsymbol{p}, G_D) < \frac{1}{544}\sum_{i\le\ell} r^+(G_D, \boldsymbol{p}, C_i),$$

then for any variable-generated event system $\mathcal{A} \sim (G_D, \boldsymbol{p})$, the expected number of resampling steps performed by MT algorithm is most m/ε.

Proof. Fix such an instance $\mathcal{A}$. Define $\delta_{i,i'} := \mathbb{P}(A_i \cap A_{i'})$. Let G_B denote the event-variable graph of $\mathcal{A}$. Let G_k denote the induced subgraph of G_B on $C_k \cup \left(\cup_{i\in C_k}\mathcal{N}_{G_B}(i)\right)$. According to Remark 4.7, it is lossless to assume G_k is a cycle of length $2|C_k|$. Thus we have

$$F^+(G_k, \boldsymbol{p}) \ge \frac{\left(\min_{i\in C_k} p_i\right)^2 \cdot \left(-|C_k| + \sum_{i\in L} 2\sqrt{p_i}\right)}{8\sqrt{|C_k|}}. \tag{12}$$

According to Theorem 4.1, there is a matching $\mathcal{M}$ of G_D such that $\sum_{(i,i')\in\mathcal{M}} \delta_{i,i'}^2 \ge \sum_{k\le\ell}\left(F^+(G_k, \boldsymbol{p})\right)^2$. Define $\boldsymbol{p}^-$ as (1). We have $(1+\varepsilon)\boldsymbol{p}^- \le (1+\varepsilon)\boldsymbol{p}$ and

$$||(1+\varepsilon)\boldsymbol{p} - (1+\varepsilon)\boldsymbol{p}^-||_1 \ge ||\boldsymbol{p} - \boldsymbol{p}^-||_1 \ge \frac{2}{17}\sum_{(i,i')\in\mathcal{M}} \delta_{i,i'}^2 \ge \frac{2}{17}\sum_{k\le\ell}\left(F^+(G_k, \boldsymbol{p})\right)^2.$$

Combining with (12), we have

$$||(1+\varepsilon)\boldsymbol{p} - (1+\varepsilon)\boldsymbol{p}^-||_1 \ge \frac{1}{544}\sum_{i\le\ell} r^+(G_D, \boldsymbol{p}, C_i) > d((1+\varepsilon)\boldsymbol{p}, G_D),$$

where the last inequality is by the condition of the lemma. Thus by Definition 1.4, we have $(1+\varepsilon)\boldsymbol{p}^-$ is in the Shearer's bound of G_D. Combining with Theorem 1.6, we have the expected number of resampling steps performed by the Moser-Tardos algorithm is most m/ε. □

Lemma 5.2. *Given G_D and any chordless cycle C in G_D, there is some probability vector $\boldsymbol{p}$ beyond the Shearer's bound of G_D and with*

$$d(\boldsymbol{p}, G_D) \ge \frac{1}{545} \cdot r(G_D, \boldsymbol{p}, C) > 2^{-20}\ell^{-3}$$

such that for any variable-generated event system $\mathcal{A} \sim (G_D, \boldsymbol{p})$, the expected number of resampling steps performed by MT algorithm is most $2^{29} \cdot m^2 \cdot |C|^3$.

The following two lemmas will be used in the proof of Lemma 5.2.

Lemma 5.3. *[She85] $q_\emptyset(G_D, \boldsymbol{p}) = 1 - \mathbb{P}(\bigcup_{A\in\mathcal{A}} A)$ holds for any extremal instance $\mathcal{A} \sim (G_D, \boldsymbol{p})$.*

Lemma 5.4. *[She85] Suppose $\boldsymbol{p}$ is the Shearer's bound of $G_D = ([m], E_D)$. Then for $i \in [m]$,*

$$\frac{\partial q_{\emptyset}(G_D, \boldsymbol{p})}{\partial p_i} = -\mathbb{P}\left(\bigcap_{j \notin N_{G_D}(i) \cup \{i\}} \overline{A_j}\right)$$

holds for any $\mathcal{A} \sim (G_D, \boldsymbol{p})$ satisfying that $A_{i'} \cap A_{i''} = \emptyset$ for any $(i', i'') \in E_D$ where $i', i'' \neq i$.

Proof of Lemma 5.2. Let $\ell = |C|$ and $\boldsymbol{\lambda} = \left(\frac{1}{4}, \cdots, \frac{1}{4}, \frac{1}{4}\right)$. Let $\mathcal{A} \sim (C, \boldsymbol{\lambda})$ be an extremal instance defined as follows: $\mathcal{A} = (A_1, \cdots, A_\ell)$ is a variable-generated event system fully determined a set of underlying mutually independent random variables $\{X_1, \cdots, X_\ell\}$. Moreover, $A_i = [X_i < 1/2] \wedge [X_{i+1} \geq 1/2]$ for each $i \in [\ell - 1]$, and $A_\ell = [X_\ell < 1/2] \wedge [X_1 \geq 1/2]$. According to Lemma 5.3,

$$q_{\emptyset}(C, \boldsymbol{\lambda}) = \mathbb{P}\left(\bigcup_{i \in [\ell]} A_i\right) = \frac{1}{2^{\ell-1}}.$$

Besides, according to Lemma 5.4, for any $\boldsymbol{\lambda}' = \left(\frac{1}{4}, \cdots, \frac{1}{4}, \frac{1}{4} + \varepsilon\right)$ in the Shearer's bound of C,

$$\frac{\partial q_{\emptyset}(C, \boldsymbol{\lambda}')}{\partial \lambda'_\ell} = -\mathbb{P}\left(\bigcap_{i \in [2, \ell-2]} \overline{A_i}\right) = -\frac{\ell - 2}{2^{\ell-3}}.$$

Thus, for any $\boldsymbol{\lambda} \leq \boldsymbol{\lambda}' \leq \boldsymbol{\lambda}'' := \left(\frac{1}{4}, \cdots, \frac{1}{4}, \frac{1}{4} + \frac{1}{4(\ell-1)}\right)$, we have that

$$q_{\emptyset}(C, \boldsymbol{\lambda}'') = q_{\emptyset}(C, \boldsymbol{\lambda}) + \int_{\frac{1}{4}}^{\lambda''_\ell} \frac{\partial q_{\emptyset}(C, \boldsymbol{\lambda}')}{\partial \lambda'_\ell} d\lambda'_\ell > \frac{1}{2^{\ell-1}} - \frac{\ell - 2}{\ell - 1} \cdot \frac{1}{2^{\ell-1}} = \frac{1}{\ell - 1} \cdot \frac{1}{2^{\ell-1}}.$$

Hence $\boldsymbol{\lambda}''$ is in the Shearer's bound of C. Thus, there exists $q > 0$ such that $\boldsymbol{q}$ defined as follows is on the Shearer's boundary of G_D:

$$\forall i \in [m]: \quad q_i = \begin{cases} \frac{1}{4} & \text{if } i \in [\ell - 1], \\ \frac{1}{4} + \frac{1}{4(\ell-1)} & \text{if } i = \ell, \\ q & \text{otherwise.} \end{cases}$$

One can verify that

$$r^+(G_D, \boldsymbol{q}, C) = r(G_D, \boldsymbol{q}, C) > \ell \cdot \frac{1}{4^4} \cdot \left(\frac{1}{2\ell^2}\right)^2 > \frac{1}{2^{10} \cdot \ell^3}. \tag{13}$$

Define

$$f(\delta) = 545 \cdot d((1 + \delta)\boldsymbol{q}, G_D) - r^+(G_D, (1 + \delta)\boldsymbol{q}, C).$$

One can verify that $f(0) < 0$ because $d(\boldsymbol{q}, G_D) = 0$ and $r^+(G_D, \boldsymbol{q}, C) > 0$. Moreover, let δ' be large enough such that $(1 + \delta')\boldsymbol{q} \notin \mathcal{I}_v(G_D)$. One can verify that such δ' must exist. We have $f(\delta') \geq 0$. This is because otherwise $f(\delta') < 0$ and then

$$d((1 + \delta')\boldsymbol{q}, G_D) < \frac{1}{545} \cdot r^+(G_D, (1 + \delta')\boldsymbol{q}, C).$$

By following the proof of Lemma 5.1, we have the MT algorithm terminates at $(1 + \delta')\boldsymbol{q}$, which is contradictory with $(1 + \delta')\boldsymbol{q} \notin \mathcal{I}_v(G_D)$.

Moreover, $f(\delta)$ is a continuous function of δ, because $d((1 + \delta)\boldsymbol{q}, G_D)$ and $r^+(G_D, (1 + \delta)\boldsymbol{q}, C)$ are both continuous functions of δ. Combining with $f(0) < 0$ and $f(\delta') > 0$, we have there must be a $0 \leq \delta \leq \delta'$ such that $f(\delta) = 0$. Let $\boldsymbol{p} = (1 + \delta)\boldsymbol{q}$. By $f(\delta) = 0$, we have

$$d(\boldsymbol{p}, G_D) = \frac{1}{545} \cdot r^+(G_D, \boldsymbol{p}, C). \tag{14}$$

Combining with $r^+(G_D, \boldsymbol{p}, C) = r(G_D, \boldsymbol{p}, C) > r(G_D, \boldsymbol{q}, C)$ and (13), we have $d(\boldsymbol{p}, G_D) > 2^{-20}\ell^{-3}$.

Fix a variable-generated event system $\mathcal{A} \sim (G_D, \boldsymbol{p})$. Define $\delta_{i,i'} := \mathbb{P}(A_i \cap A_{i'})$. Let G_B denote the event-variable graph of $\mathcal{A}$. Let G denote the induced subgraph of G_B on $C \cup \left(\cup_{i\in C}\mathcal{N}_{G_B}(i)\right)$. According to Remark 4.7, it is lossless to assume that G is a cycle of length $2|C|$. Thus we have

$$F^+(G, \boldsymbol{p}) \geq \frac{\left(\min_{i\in C} p_i\right)^2 \cdot \left(-|C| + \sum_{i\in L} 2\sqrt{p_i}\right)}{8\sqrt{|C|}}. \tag{15}$$

According to Theorem 4.1, there is a matching $\mathcal{M}$ of G_D such that $\sum_{(i,i')\in\mathcal{M}} \delta^2_{i,i'} \geq (F^+(G,\boldsymbol{p}))^2$. Define $\boldsymbol{p}^-$ as (1). We have

$$||\boldsymbol{p} - \boldsymbol{p}^-||_1 \geq \frac{2}{17} \sum_{(i,i')\in\mathcal{M}} \delta^2_{i,i'} \geq \frac{2}{17} \sum_{k\leq \ell} \left(F^+(G_k, \boldsymbol{p})\right)^2.$$

Combining with (15), we have

$$||\boldsymbol{p} - \boldsymbol{p}^-||_1 \geq \frac{1}{544} \cdot r^+(G_D, \boldsymbol{p}, C).$$

Let

$$\varepsilon \triangleq \frac{1}{2^{29} \cdot \ell^3 \cdot m}.$$

By (13) we have

$$m\varepsilon \leq \frac{1}{545 \cdot 544 \cdot 2^{10} \cdot \ell^3} \leq \left(\frac{1}{544} - \frac{1}{545}\right) r^+(G_D, \boldsymbol{q}, C) \leq \left(\frac{1}{544} - \frac{1}{545}\right) r^+(G_D, \boldsymbol{p}, C).$$

Thus we have

$$||\boldsymbol{p} - (1+\varepsilon)\boldsymbol{p}^-||_1 > ||\boldsymbol{p} - \boldsymbol{p}^-||_1 - m\varepsilon \geq \frac{r^+(G_D, \boldsymbol{p}, C)}{544} - m\varepsilon \geq \frac{r^+(G_D, \boldsymbol{p}, C)}{545} \geq d(\boldsymbol{p}, G_D),$$

where the last inequality is by (14). Thus by Definition 1.4, we have $(1+\varepsilon)\boldsymbol{p}^-$ is in the Shearer's bound of G_D. Combining with Theorem 1.6, we have the expected number of resampling steps performed by the MT algorithm is most m/ε. □

6. Application to periodic Euclidean graphs

In this section, we explicitly calculate the gaps between our new criterion and Shearer's bound on periodic Euclidean graphs, including several lattices that have been studied extensively in physics. It turns out the efficient region of MT algorithm can exceed *significantly* beyond Shearer's bound.

A periodic Euclidean graph G_D is a graph that is embedded into a Euclidean space naturally and has a *translational unit* G_U in the sense that G_D can be viewed as the union of periodic translations of G_U. For example, a cycle of length 4 is a translational unit of the square lattice.

Given a dependency graph G_D, it naturally defines a bipartite graph $G_B(G_D)$ as follows. Regard each edge of G_D as a variable and each vertex as an event. An event A depends on a variable X if and only if the vertex corresponding to A is an endpoint of the edge corresponding to X.

For simplicity, we only focus on symmetric probabilities, where $\boldsymbol{p} = (p, p, \cdots, p)$. Given a dependency graph G_B and a vector $\boldsymbol{p}$, remember that $\boldsymbol{p}$ is on Shearer's boundary of G_D if $(1-\varepsilon)\boldsymbol{p}$ is in Shearer's bound and $(1+\varepsilon)\boldsymbol{p}$ is not for any $\varepsilon > 0$.

Given a dependency graph $G_D = ([m], E_D)$ and two vertices $i, i' \in [m]$, we use $\mathrm{dist}(i, i')$ to denote the distance between i and i' in G_D. The following Lemma will be used.

Lemma 6.1. *Suppose $\boldsymbol{p}_a = (p_a, p_a, \cdots, p_a)$ is on Shearer's boundary of $G_D = ([m], E_D)$. For any probability vector $\boldsymbol{p}$ other than $\boldsymbol{p}_a$, it is in the Shearer's bound if there exist $K, d \in \mathbb{N}^+$, $\mathcal{S} \subseteq 2^{[m]}$ where $\cup_{S\in\mathcal{S}} = [m]$, and $f : \mathcal{S} \to 2^{[m]}$ such that the following conditions hold:*

(a) for each $i \in [m]$, there are at most K subsets $S \in \mathcal{S}$ such that $f(S) \ni i$;
(b) if $f(S) = T$, then $\mathrm{dist}(i, i') \leq d$ for each $i \in S$ and $i' \in T$;
(c) if $f(S) = T$, then

$$\left(\frac{1-p_a}{p_a}\right)^{d-1} \cdot \frac{K}{p_a} \cdot \sum_{i\in S} \max\{p_i - p_a, 0\} \leq \sum_{i\in T} \max\{p_a - p_i, 0\}.$$

While Lemma 6.1 looks involved, the basic idea is simple: by contradiction, suppose there is such a vector $\boldsymbol{p}'$ beyond Shearer's bound; then we apply Lemma D.1 repeatedly to transfer probability from one event to another while keeping the probability vector still beyond Shearer's bound; finally, the vector $\boldsymbol{p}'$ will be changed to a vector strictly below $\boldsymbol{p}$, which makes a contradiction to the assumption that $\boldsymbol{p}$ is on the Shearer's boundary. The involved part is a transferring scheme which changes $\boldsymbol{p}'$ to another probability vector strictly below $\boldsymbol{p}$. We leave the proof to the appendix.

The main result of this section is as follows.

Theorem 6.2. *Let $G_D = (V_D, E_D)$ be a periodic Euclidean graph with maximum degree Δ, and $\boldsymbol{p}_a = (p_a, \cdots, p_a)$ be the probability vector on Shearer's boundary of G_D. Suppose $G_U = (V_U, E_U)$ is a translational unit of G_D with diameter D. Let*

$$q \triangleq \frac{p_a^{D+2}\left(F^+(G_B(G_U), \boldsymbol{p}_a)\right)^2}{17 \cdot (\Delta+1) \cdot |V|^2 \cdot (1-p_a)^{D+1}}.$$

Then for any $\mathcal{A} \sim (G_B(G_D), \boldsymbol{p})$ where $(1+\varepsilon)\boldsymbol{p} \leq (p_a + q, \cdots, p_a + q)$, the expected number of resampling steps performed by the MT algorithm is most $|V_D|/\varepsilon$.

Proof. Fix any $\mathcal{A} \sim (G_B(G_D), \boldsymbol{p})$ where $(1+\varepsilon)\boldsymbol{p} \leq (p_a + q, \cdots, p_a + q)$. Let δ_{v_0,v_1} denote $\mathbb{P}(A_{v_0} \cap A_{v_1})$ for $(v_0, v_1) \in E_D$. We construct a matching $\mathcal{M} \subset E_D$ greedily as follows: we maintain two sets E and $\mathcal{M}$, which are initialized as E_D and $\emptyset$ respectively. We do the following iteratively until E becomes empty: select a edge (v_0, v_1) with maximum δ_{v_0,v_1} from E, add (v_0, v_1) to $\mathcal{M}$, and delete all edges connecting v_0 or v_1 from E (including (v_0, v_1)). Let $\boldsymbol{\delta} = \left(\delta_{v_0,v_1} : (v_0, v_1) \in \mathcal{M}\right)$. Then $\mathcal{A} \sim (G_B(G_D), \boldsymbol{p}, \mathcal{M}, \boldsymbol{\delta})$.

Define $\boldsymbol{p}^-$ as (1). In the remaining part of the proof, we will show that $(1+\varepsilon)\boldsymbol{p}^-$ is in the Shearer's bound. This implies the conclusion immediately by Theorem 1.6.

In fact, it is a direct application of Lemma 6.1 to show that $(1+\varepsilon)\boldsymbol{p}^-$ is in the Shearer's bound. To provide more detail, we need some notations. We use $v, v', v_1, v_2, \cdots$ to represent vertices in G_D, and use $u, u', u_1, u_2, \cdots$ to represent vertices in G_U. Let $G_U^1, G_U^2, \cdots$ be the periodic translations of G_U in G_D. And we use a surjection[9] $h : \mathbb{N}^+ \times V_U \to V_D$ to represent how these periodic translations constitute G_D: $h(k, u) = v$ if the copy of $u \in V_U$ in k-th translation (i.e., G_U^k) is $v \in V_D$. In particular, the vertex set of G_U^k , denoted by V_U^k, is $\{h(k, u) : u \in V\}$, and the edge set of G_U^k , denoted by E_U^k, is $\{(h(k, u), h(k, u')) : (u, u') \in E_U\}$. Besides, let $\mathcal{N}^+(v) := \mathcal{N}_{G_D}(v) \cup \{v\}$ for $v \in V_D$. For $V \subset V_D$, let $\mathcal{N}^+(V) := \cup_{v \in V} \mathcal{N}^+(v)$. Let $T_k := \{(v_0, v_1) \in \mathcal{M} : v_0, v_1 \in \mathcal{N}^+(G_U^k)\}$ stand for the pairs in $\mathcal{M}$ adjacent to G_U^k. With some abuse of notation, we sometimes use $v \in T_k$ to denote that $(v, v') \in T_k$ for some $v' \in V_D$.

The following claim says that $\boldsymbol{p}^-$ is much smaller than $\boldsymbol{p}$ even projected on a single translation. Its proof uses a similar idea to Theorem 4.6 and can be found in the appendix.

Claim 6.3. $\sum_{(v_0,v_1) \in T_k} \delta_{v_0,v_1}^2 \geq \left(F^+(G_B(G_U), \boldsymbol{p})\right)^2$ *holds for any k.*

To apply Lemma 6.1, let $K := (\Delta+1)|V_U|$, $d := D + 2$, $\mathcal{S} := \{V_U^1, V_U^2, \cdots\}$, and $f(V_U^k) := T_k$. Based on Claim 6.3, one can check that all the three conditions in Lemma 6.1 hold (see the appendix for details). Thus, according to Lemma 6.1, $(1+\varepsilon)\boldsymbol{p}^-$ is in Shearer's bound. □

We apply Theorem 6.2 to three lattices: square lattice, Hexagonal lattice, and simple cubic lattice. For square lattice, we take the 5×5 square with 25 vertices as the translational unit. For Hexagonal lattice, we take a graph consisting of 19 hexagons as the translational unit, in which there are 3,4,5,4,3 hexagons in the five columns, respectively. For simple cubic lattice, we take the $3 \times 3 \times 3$ cube with 27 vertices as the translational unit. The explicit gaps are summarized in Table 1. Finally, the lower bounds for these three lattices in Table 1 hold for all bipartite graphs with the given canonical dependency graph, because all such bipartite graphs are essentially the same under the reduction rules defined in [HLL+17].

[9] h is possibly not a injection, as these translations are possibly overlapped with each other.

References

[AFR95] Michael Albert, Alan Frieze, and Bruce Reed. Multicoloured hamilton cycles. *the electronic journal of combinatorics*, 2(1):R10, 1995.

[AI16] Dimitris Achlioptas and Fotis Iliopoulos. Random walks that find perfect objects and the lovász local lemma. *Journal of ACM*, 63(3):22:1–22:29, 2016.

[AIK19] Dimitris Achlioptas, Fotis Iliopoulos, and Vladimir Kolmogorov. A local lemma for focused stochastic algorithms. *SIAM Journal on Computing*, 48(5):1583–1602, 2019.

[AIS19] Dimitris Achlioptas, Fotis Iliopoulos, and Alistair Sinclair. Beyond the lovász local lemma: Point to set correlations and their algorithmic applications. In *Proceedings of Symposium on Foundations of Computer Science (FOCS)*, pages 725–744, 2019.

[AIV17] Dimitris Achlioptas, Fotis Iliopoulos, and Nikos Vlassis. Stochastic control via entropy compression. In *Proceedings of International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 80, pages 83:1–83:13, 2017.

[Alo91] Noga Alon. A parallel algorithmic version of the local lemma. *Random Structures and Algorithms*, 2(4):367–378, 1991.

[Bec91] József Beck. An algorithmic approach to the Lovász local lemma. *Random Structures and Algorithms*, 2(4):343–365, 1991.

[BFH+16] Sebastian Brandt, Orr Fischer, Juho Hirvonen, Barbara Keller, Tuomo Lempiäinen, Joel Rybicki, Jukka Suomela, and Jara Uitto. A lower bound for the distributed lovász local lemma. In *Proceedings of Symposium on Theory of Computing (STOC)*, pages 479–488, 2016.

[BFPS11] Rodrigo Bissacot, Roberto Fernández, Aldo Procacci, and Benedetto Scoppola. An improvement of the Lovász local lemma via cluster expansion. *Combinatorics, Probability and Computing*, 20(05):709–719, 2011.

[CCS+17] Jan Dean Catarata, Scott Corbett, Harry Stern, Mario Szegedy, Tomas Vyskocil, and Zheng Zhang. The Moser-Tardos resample algorithm: Where is the limit?(an experimental inquiry). In *2017 Proceedings of the Ninteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 159–171, 2017.

[CGH13] Karthekeyan Chandrasekaran, Navin Goyal, and Bernhard Haeupler. Deterministic algorithms for the lovász local lemma. *SIAM Journal on Computing*, 42(6):2132–2155, 2013.

[CPS17] Kai-Min Chung, Seth Pettie, and Hsin-Hao Su. Distributed algorithms for the lovász local lemma and graph coloring. *Distributed Computing*, 30(4):261–280, 2017.

[CS00] Artur Czumaj and Christian Scheideler. Coloring nonuniform hypergraphs: A new algorithmic approach to the general Lovász local lemma. *Random Structures and Algorithms*, 17(3-4):213–237, 2000.

[EL75] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets*, 10(2):609–627, 1975.

[ES91] Paul Erdös and Joel Spencer. Lopsided lovász local lemma and latin transversals. *Discrete Applied Mathematics*, 30(2-3):151–154, 1991.

[FGYZ20] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Fast sampling and counting k-sat solutions in the local lemma regime. In *Proceedings of Symposium on Theory of Computing (STOC)*, pages 854–867, 2020.

[FHY20] Weiming Feng, Kun He, and Yitong Yin. Sampling constraint satisfaction solutions in the local lemma regime. *arXiv preprint arXiv:2011.03915*, 2020.

[FHY21] Weiming Feng, Kun He, and Yitong Yin. Sampling constraint satisfaction solutions in the local lemma regime. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1565–1578, 2021.

[Gau67] David S Gaunt. Hard-sphere lattice gases. ii. plane-triangular and three-dimensional lattices. *The Journal of Chemical Physics*, 46(8):3237–3259, 1967.

[GF65] David S Gaunt and Michael E Fisher. Hard-sphere lattice gases. i. plane-square lattice. *The Journal of Chemical Physics*, 43(8):2840–2863, 1965.

[GGGY20] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Kuan Yang. Counting solutions to random CNF formulas. In *ICALP*, volume 168 of *LIPIcs*, pages 53:1–53:14, 2020.

[GH20] Heng Guo and Kun He. Tight bounds for popping algorithms. *Random Structures & Algorithms*, 2020.

[Gha16] Mohsen Ghaffari. An improved distributed algorithm for maximal independent set. In Robert Krauthgamer, editor, *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 270–277, 2016.

[Gil19] András Gilyén. *Quantum singular value transformation & its algorithmic applications*. PhD thesis, University of Amsterdam, 2019.

[GJ18] Heng Guo and Mark Jerrum. Approximately counting bases of bicircular matroids. *arXiv preprint arXiv:1808.09548*, 2018.

[GJ19] Heng Guo and Mark Jerrum. A polynomial-time approximation algorithm for all-terminal network reliability. *SIAM Journal on Computing*, 48(3):964–978, 2019.

[GJL19] Heng Guo, Mark Jerrum, and Jingcheng Liu. Uniform sampling through the lovász local lemma. *Journal of the ACM (JACM)*, 66(3):18, 2019.

[GKPT17] Ioannis Giotis, Lefteris Kirousis, Kostas I Psaromiligkos, and Dimitrios M Thilikos. Acyclic edge coloring through the Lovász local lemma. *Theoretical Computer Science*, 665:40–50, 2017.

[GLLZ19] Heng Guo, Chao Liao, Pinyan Lu, and Chihao Zhang. Counting hypergraph colorings in the local lemma regime. *SIAM Journal on Computing*, 48(4):1397–1424, 2019.

[GMSW09] Heidi Gebauer, Robin A Moser, Dominik Scheder, and Emo Welzl. The lovász local lemma and satisfiability. In *Efficient Algorithms*, pages 30–54. Springer, 2009.

[GST16] Heidi Gebauer, Tibor Szabó, and Gábor Tardos. The local lemma is asymptotically tight for SAT. *Journal of ACM*, 63(5):43, 2016.

[Har16] David G Harris. Lopsidependency in the moser-tardos framework: beyond the lopsided Lovász local lemma. *ACM Transactions on Algorithms (TALG)*, 13(1):17, 2016.

[Har18] David G. Harris. Deterministic parallel algorithms for fooling polylogarithmic juntas and the lovász local lemma. *ACM Transaction on Algorithms*, 14(4):47:1–47:24, 2018.

[Har19] David G. Harris. Deterministic algorithms for the lovasz local lemma: simpler, more general, and more parallel. *CoRR*, abs/1909.08065, 2019.

[HH17] Bernhard Haeupler and David G. Harris. Parallel algorithms and concentration bounds for the lovász local lemma via witness dags. *ACM Transaction on Algorithms*, 13(4):53:1–53:25, 2017.

[HLL+17] Kun He, Liang Li, Xingwu Liu, Yuyi Wang, and Mingji Xia. Variable-version Lovász local lemma: Beyond shearer's bound. In *Proceedings of Symposium on Foundations of Computer Science (FOCS)*, pages 451–462, 2017.

[HLSZ19] Kun He, Qian Li, Xiaoming Sun, and Jiapeng Zhang. Quantum lovász local lemma: Shearer's bound is tight. In *Proceedings of Symposium on Theory of Computing (STOC)*, pages 461–472, 2019.

[HS14] David G. Harris and Aravind Srinivasan. A constructive algorithm for the lovász local lemma on permutations. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 907–925, 2014.

[HSS11] Bernhard Haeupler, Barna Saha, and Aravind Srinivasan. New constructive aspects of the lovász local lemma. *Journal of ACM*, 58(6):1–28, 2011.

[HSW21] Kun He, Xiaoming Sun, and Kewen Wu. Perfect sampling for (atomic) lov\'asz local lemma. *arXiv preprint arXiv:2107.03932*, 2021.

[HV20] Nicholas J. A. Harvey and Jan Vondrák. An algorithmic proof of the lovász local lemma via resampling oracles. *SIAM Journal on Computing*, 49(2):394–428, 2020.

[IS20] Fotis Iliopoulos and Alistair Sinclair. Efficiently list-edge coloring multigraphs asymptotically optimally. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2319–2336, 2020.

[JPV20] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. Towards the sampling lovász local lemma. *CoRR*, abs/2011.12196, 2020.

[JPV21] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. On the sampling lovász local lemma for atomic constraint satisfaction problems. *CoRR*, abs/2102.08342, 2021.

[KS11] Kashyap Babu Rao Kolipaka and Mario Szegedy. Moser and tardos meet lovász. In *Proceedings of Symposium on Theory of Computing (STOC)*, pages 235–244, 2011.

[KSX12] Kashyap Babu Rao Kolipaka, Mario Szegedy, and Yixin Xu. A sharper local lemma with improved applications. In *Proceedings of APPROX/RANDOM*, volume 7408 of *Lecture Notes in Computer Science*, pages 603–614, 2012.

[LS07] Linyuan Lu and László Székely. Using lovász local lemma in the space of random injections. *the electronic journal of combinatorics*, 14(1):R63, 2007.

[LS09] Linyuan Lu and Laszlo A Szekely. A new asymptotic enumeration technique: the lovász local lemma. *arXiv preprint arXiv:0905.3983*, 2009.

[McD97] Colin McDiarmid. Hypergraph colouring and the Lovász local lemma. *Discrete Mathematics*, 167:481–486, 1997.

[Moi19a] Ankur Moitra. Approximate counting, the lovasz local lemma, and inference in graphical models. *Journal of ACM*, 66(2):10, 2019.

[Moi19b] Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *J. ACM*, 66(2):10:1–10:25, 2019.

[Mol19] Michael Molloy. The list chromatic number of graphs with small clique number. *Journal of Combinatorial Theory, Series B*, 134:264–284, 2019.

[MR98] Michael Molloy and Bruce Reed. Further algorithmic aspects of the local lemma. In *Proceedings of Symposium on Theory of Computing (STOC)*, pages 524–529, 1998.

[MT10] Robin A Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *Journal of ACM*, 57(2):11, 2010.

[Peg14] Wesley Pegden. An extension of the moser–tardos algorithmic local lemma. *SIAM Journal on Discrete Mathematics*, 28(2):911–917, 2014.

[She85] James B. Shearer. On a problem of spencer. *Combinatorica*, 5(3):241–245, 1985.

[Spe75] Joel Spencer. Ramsey's theorem—a new lower bound. *Journal of Combinatorial Theory, Series A*, 18(1):108–115, 1975.

[Spe77] Joel Spencer. Asymptotic lower bounds for Ramsey functions. *Discrete Mathematics*, 20:69–76, 1977.

[Sze13] Mario Szegedy. The lovász local lemma–a survey. In *International Computer Science Symposium in Russia*, pages 1–11. Springer, 2013.

[Tod99] Synge Todo. Transfer-matrix study of negative-fugacity singularity of hard-core lattice gas. *International Journal of Modern Physics C*, 10(04):517–529, 1999.

Appendix A. Missing Proofs in Section 2

Proof of Proposition 3.2. The following simple greedy procedure will find such a $\mathcal{P}$.

Initially, $\mathcal{P} = \emptyset$;
for *each* $(i, i') \in \mathcal{M}$ **do**
  for *each* k *from* 1 *to* $|\mathrm{List}(D, i, i')| - 1$ **do**
    if *the* k*-th node and* $(k+1)$*-th node in* $\mathrm{List}(D, i, i')$ *form a reversible arc* **then**
      add this arc to $\mathcal{P}$, and $k := k + 2$;
    else
      $k := k + 1$;
Return $\mathcal{P}$;

Obviously, for each $(i, i') \in \mathcal{M}$, the procedure contains at least half of all reversible arcs $u \to v$ where $\{L(u), L(v)\} = \{i, i'\}$, hence at least half of nodes in $\mathcal{V}(D, i)$. □

Appendix B. Proof of Proposition 3.9

Given a pwdag $D = (V, E, L)$ of G_D and a Boolean string $\boldsymbol{R} \in \{0,1\}^{\mathscr{M}(D)}$, define $h(D, \boldsymbol{R})$ to be a directed graph $D' := (V', E', L')$ where $V' = V$, $E' = E$, and

$$\forall v \in V: \quad L'(v) = \begin{cases} (L(v))^{\uparrow}, & \text{if } v \in \mathscr{M} \text{ and } R_v = 0; \\ (L(v))^{\downarrow}, & \text{if } v \in \mathscr{M} \text{ and } R_v = 1; \\ L(v), & \text{otherwise, } v \notin \mathscr{M}. \end{cases}$$

It is easy to verify that $h(D, \boldsymbol{R})$ is a pwdag of $G^{\mathcal{M}}$. Moreover, given any $D' \in \mathcal{D}(G^{\mathcal{M}})$, there is one and only one $D \in \mathcal{D}(G_D)$ and $\boldsymbol{R} \in \{0,1\}^{\mathscr{M}(D)}$ such that $h(D, \boldsymbol{R}) = D'$. In other words, h is a bijection between $\{(D, \boldsymbol{R}) : D \in \mathcal{D}(G_D), \boldsymbol{R} \in \{0,1\}^{\mathscr{M}(D)}\}$ and $\mathcal{D}(G^{\mathcal{M}})$. So

$$\begin{aligned}
\sum_{D' \in \mathcal{D}(G^{\mathcal{M}})} \prod_{v' \text{ in } D'} p^{\mathcal{M}}_{L'(v')} &= \sum_{D \in \mathcal{D}(G_D)} \sum_{\boldsymbol{R} \in \{0,1\}^{\mathscr{M}(D)}} \prod_{v' \text{ in } h(D,\boldsymbol{R})} p^{\mathcal{M}}_{L'(v')} \\
&= \sum_{D \in \mathcal{D}(G_D)} \sum_{\boldsymbol{R} \in \{0,1\}^{\mathscr{M}(D)}} \prod_{v \text{ in } D} p^{\mathcal{M}}_{L'(v)} \\
&= \sum_{D \in \mathcal{D}(G_D)} \prod_{v \notin \mathscr{M}(D)} p^{\mathcal{M}}_{L'(v)} \left(\sum_{\boldsymbol{R} \in \{0,1\}^{\mathscr{M}(D)}} \prod_{v \in \mathscr{M}(D)} p^{\mathcal{M}}_{L'(v)} \right) \\
&= \sum_{D \in \mathcal{D}(G_D)} \prod_{v \notin \mathscr{M}(D)} p^{\mathcal{M}}_{L(v)} \prod_{v \in \mathscr{M}(D)} \left(p^{\mathcal{M}}_{L(v)^{\uparrow}} + p^{\mathcal{M}}_{L(v)^{\downarrow}} \right) \\
&= \sum_{D \in \mathcal{D}(G_D)} \prod_{v \notin \mathscr{M}(D)} p^{-}_{L(v)} \prod_{v \in \mathscr{M}(D)} \left(p'_{L(v)} + p^{-}_{L(v)} - p'_{L(v)} \right) \\
&= \sum_{D \in \mathcal{D}(G_D)} \prod_{v \text{ in } D} p^{-}_{L(v)},
\end{aligned}$$

where the second equality is by that $V = V'$, the forth equality is by the definition of L', and the fifth equality is by the definition of $\boldsymbol{p}^{\mathcal{M}}$.

Appendix C. Proof of Theorem 3.12

We first verify that the image of h is a subset of $\mathcal{D}(G^{\mathcal{M}})$.

Lemma C.1. *For any $D \in \mathcal{D}(G_D)$ and $\mathcal{S} \in \psi(D)$, $h(D, \mathcal{S}) \in \mathcal{D}(G^{\mathcal{M}})$.*

Proof. First, we prove that $h(D, \mathcal{S}) = (V', E', L')$ is a DAG. Define a total order π' over the set V' as follows: for any two distinct nodes $u', v' \in V'$,

- if $g(u') \neq g(v')$, then $u' \prec v'$ in π' if and only if $g(u') \prec g(v')$ in π_D;
- if $g(u') = g(v')$, then $u' \prec v'$ in π' if and only if $u' = f^*(g(u'))$ (and then $v' = f(g(u'))$).

One can verify that π' is a topological order of $h(D, \mathcal{S})$, which means that $h(D, \mathcal{S})$ is a DAG.

Secondly, we prove that $h(D, \mathcal{S})$ is a wdag of $G^{\mathcal{M}}$. As $h(D, \mathcal{S})$ has been shown to be a DAG, we only need to verify that: for any two distinct nodes u', v' in D', there is a arc between u' and v' (in either direction) if and only if either $L'(u') = L'(v')$ or $(L'(v'), L'(u')) \in E^{\mathcal{M}}$.

$\Longrightarrow$: By symmetry, suppose $(u' \to v') \in E'$. If $(u' \to v') \in E'_1$, then $u' = f^*(w)$ and $v' = f(w)$ for some vertex $w \in \mathcal{S}_3 \cup \mathcal{S}_4$. Thus, by (2) and (3) we have $L'(u') \in \{i^{\uparrow}, i^{\downarrow}\}$ and $L'(v') = L(w)^{\downarrow}$ where $(L(w), i) \in \mathcal{M}$. By $(L(w), i) \in \mathcal{M}$, any two vertices in $\{(L(w))^{\uparrow}, (L(w))^{\downarrow}, i^{\uparrow}, i^{\downarrow}\}$ are connected in $G^{\mathcal{M}}$. In particular, $(L'(v'), L'(u')) \in E^{\mathcal{M}}$. If $(u' \to v') \in E'_2$, we have $L'(u') = L'(v')$ or $(L'(u'), L'(v')) \in E^{\mathcal{M}}$ immediately.

$\Longleftarrow$: Suppose $u', v' \in V'$ are two distinct nodes where $L'(u') = L'(v')$ or $(L'(u'), L'(v')) \in E^{\mathcal{M}}$. If $g(u') \neq g(v')$, then either $g(u') \prec g(v')$ or $g(v') \prec g(u')$ in π_D, which implies that either $(u' \to v') \in E'_2$ or $(v' \to u') \in E'_2$. Otherwise, $g(u') = g(v')$. Let $v := g(u') = g(v')$. By (2) and (3), we have $v \in \mathcal{S}_3 \cup \mathcal{S}_4$ and $\{u', v'\} = \{f(v), f^*(v)\}$. Therefore either $u' \to v'$ or $v' \to u'$ is in E'_1.

Finally, one can check that $f(v)$ where v is the unique sink of D is the unique sink of D'. This completes the proof. $\square$

In the rest of this section, we show that h is injective. Given $D \in \mathcal{D}(G_D)$ and $(i, j) \in \mathcal{M}$, recall that $\mathrm{List}(D, i, j)$ is the sequence listing all nodes in D' labelled with i or j in the topological order. Similarly,

Definition C.2. *Given $D' = (V', E', L') \in \mathcal{D}(G^{\mathcal{M}})$ and $(i, j) \in \mathcal{M}$, we use $\mathrm{List}'(D', i, j)$ to denote the unique sequence listing all nodes in D' with label in $\{i^{\uparrow}, i^{\downarrow}, j^{\uparrow}, j^{\downarrow}\}$ in the topological order.*

Claims C.3 and C.5 are two properties about $\mathrm{List}'(D', i, j)$, which will be used to show the injectiveness of h.

Claim C.3. *Suppose $D' = h(D, \mathcal{S})$ for some $D \in \mathcal{D}(G_D)$ and $\mathcal{S} \in \psi(D)$. Let $(i, j) \in \mathcal{M}$. Then for any node v' in D',*

- *(a) $v' \in \mathrm{List}'(D', i, j)$ if and only if $g(v') \in \mathrm{List}(D, i, j)$;*
- *(b) for any other node u' in D', if $g(u')$ precedes $g(v')$ in $\mathrm{List}(D, i, j)$, then u' precedes v' in $\mathrm{List}'(D', i, j)$;*
- *(c) if $v \in \mathcal{S}_3 \cup \mathcal{S}_4$, then $f(v)$ is next to $f^*(v)$ in $\mathrm{List}'(D', i, j)$.*

Proof. Part (a) is immediate by Definition 3.11.

Now, we show Part (b). Suppose $g(u')$ precedes $g(v')$ in $\mathrm{List}(D, i, j)$. Then $g(u') \prec g(v')$ in π_D. Thus one can check that all the four arcs $f(g(u')) \to f(g(v'))$, $f^*(g(u')) \to f(g(v'))$, $f(g(u')) \to f^*(g(v'))$, and $f^*(g(u')) \to f^*(g(v'))$ are contained in E'_2. In particular, $(u' \to v') \in E'$ as $u' \in \{f(g(u')), f^*(g(u'))\}$ and $v' \in \{f(g(v')), f^*(g(v'))\}$. This implies that u' precedes v' in $\mathrm{List}'(D', i, j)$.

Finally, we prove Part (c). According to Part (b), $f(v)$ and $f^*(v)$ are adjacent in $\mathrm{List}'(D', i, j)$. Besides, as there is an arc $f^*(v) \to f(v)$ in E'_1, we conclude that $f(v)$ is next to $f^*(v)$ in $\mathrm{List}'(D', i, j)$. $\square$

Definition C.4. *For a reversible arc $u' \to v'$ in D', we call it $(*, \downarrow)$-reversible in D' if $L'(u') \in \{i^{\uparrow}, i^{\downarrow}\}$ and $L'(v') = j^{\downarrow}$ for some $(i, j) \in E_D$.*

Claim C.5. *Suppose $D' = h(D, \mathcal{S})$ for some $D \in \mathcal{D}(G_D)$ and $\mathcal{S} \in \psi(D)$. Let $(i, j) \in \mathcal{M}$. Let u', v' be two nodes in $\mathrm{List}'(D', i, j)$ where v' is next to u'. Then $u' \in V'_2$ if and only if $u' \to v'$ is $(*, \downarrow)$-reversible in D' and $v' \in V'_1$.*

Proof. $\Longrightarrow$: Let $u := g(u')$. Assume $u' \in V'_2$, i.e., $u' = f^*(u)$. By Definition 3.11, $u \in \mathcal{S}_3 \cup \mathcal{S}_4$. According to Part (c) of Claim C.3, as v' is next to u', we have $v' = f(u)$ and then $v' \in V'_1$.

Now we show that $u' \to v'$ is $(*, \downarrow)$-reversible. First, by Definition 3.11, either $L'(u') \in \{i^{\uparrow}, i^{\downarrow}\}$ and $L'(v') = j^{\downarrow}$, or $L'(u') \in \{j^{\uparrow}, j^{\downarrow}\}$ and $L'(v') = i^{\downarrow}$. What remains is to show $u' \to v'$ is reversible, by Fact 2.5 which is equivalent to show that $f^*(u) \to f(u)$ is the unique path from u' to v' in D'. By contradiction, assume that there is a path $f^*(u) \to w'_1 \to \cdots \to w'_k \to f(u)$ in D' where $w'_1 \neq f(u)$ and $w'_k \neq f^*(u)$. As $w'_1 \neq f(u)$, we have $(f^*(u) \to w'_1)$ is not in E'_1 and then should be in E'_2, which further implies that $u \prec g(w'_1)$ in π_D. Similarly, we have $g(w'_k) \prec u$ in π_D. So $g(w'_k) \prec u \prec g(w'_1)$. Meanwhile, for each $\ell < k$, if $(w'_\ell \to w'_{\ell+1}) \in E'_1$, then $g(w'_\ell) = g(w'_{\ell+1})$; if $(w'_\ell \to w'_{\ell+1}) \in E'_2$, then $g(w'_\ell) \prec g(w'_{\ell+1})$ in π_D. So, it always holds that $g(w'_\ell) \preccurlyeq g(w'_{\ell+1})$ in π_D for each $\ell < k$. In particular, $g(w'_1) \preccurlyeq g(w'_k)$. A contradiction.

$\Longleftarrow$: Let $u := g(u')$ and $v := g(v')$. Assume $u' \notin V'_2$ and $v' \in V'_1$, i.e., $u' = f(u)$ and $v' = f(v)$. Furthermore, assume $L'(v') = j^{\downarrow}$, then $v \notin \mathcal{S}_1$ and $L(v) = j$. We will show that $(f(u) \to f(v))$ is not reversible.

Note that $(f(u) \to f(v))$ should be in E'_2 and then $u \prec v$ in π_D. By $L'(u') \in \{i^{\uparrow}, i^{\downarrow}\}$, $u' = f(u)$, and (2), we have $L(u) = i$. Thus, $(L(u), L(v)) = (i, j) \in \mathcal{M} \subseteq E_D$. As D is a wdag and $u \prec v$ in π_D, the arc $(u \to v)$ exists in D. Since $v \notin \mathcal{S}_1$, $v \notin \mathcal{V}$, which means that $u \to v$ is not reversible in D. According to Fact 2.5, there is a path $u = w_1 \to w_2 \to \cdots \to w_k \to w_{k+1} = v$ from u to v in D other than the arc $u \to v$, where $w_\ell \prec w_{\ell+1}$ in π_D and $(L(w_\ell) = L(w_{\ell+1})) \vee ((L(w_\ell), L(w_{\ell+1})) \in E_D)$ for each $\ell \in [k]$.

According to the definition of $G^{\mathcal{M}}$ and (2), one can check that $(f(w_i) \to f(w_{i+1})) \in E'_2$. Therefore $u' = f(w_1) \to f(w_2) \to \cdots \to f(w_k) \to f(w_{k+1}) = v'$ is a path from u' to v' in D', which implies that $u' \to f(v)$ is not reversible in D' by Fact 2.5. $\square$

Having Claims C.3 and C.5, we are ready to show that h is injective.

Lemma C.6. *h is injective.*

Proof. Fix a $D = (V, E, L) \in \mathcal{D}(G_D)$ and a $\mathcal{S} \in \psi(D)$. Let $D' = (V', E', L')$ denote $h(D, \mathcal{S})$. We show $(D, \mathcal{S})$ can be recovered from D', which implies the injectiveness of h.

First, we recover the partition (V_1', V_2'). That is, given a node $u' \in V'$, we distinguish whether $u' \in V_1'$ or $u' \in V_2'$. If $L'(u') \in [m] \setminus \mathcal{M}$, then $u' \in V_1'$ according to (2). Otherwise, we have $L'(u') \in \{i^\uparrow, i^\downarrow\}$ for some $(i, j) \in \mathcal{M}$, hence u' is in $\text{List}'(D', i, j)$. Assume the nodes in $\text{List}'(D, i, j)$ are $v_1' v_2' v_3' \cdots v_k'$. According to Claim C.5, we can see that the following procedure distinguishes whether $v_\ell' \in V_1'$ or $v_k' \in V_2'$ for all $v_\ell' \in \text{List}'(D', i, j)$, including u'.

Initially, mark that $v_k' \in V_1'$, and let $\ell := k - 1$;
while $\ell \geq 1$ **do**
  if *the arc* $(v_\ell' \to v_{\ell+1}')$ *is* $(*, \downarrow)$*-reversible and* $v_{\ell+1}' \in V_1'$ **then**
    Mark that $v_\ell' \in V_2'$;
  else
    Mark that $v_\ell' \in V_1'$;
  $\ell := \ell - 1$;

Secondly, we can easily recover $D = (V, E, L)$ from D' and (V_1', V_2'). Ignoring labels, it is easy to see that D is exactly the induced subgraph of D' on V_1'. By the way, we also get the function $f : V \to V_1'$. For labels, we simply replace each label $i^\uparrow$ or $i^\downarrow$ with i.

Finally, we recover $\mathcal{S}$ from D', D and (V_1', V_2). That is, we distinguish which one of $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_3\}$ contains a given node $v \in \mathscr{M}(D)$. Assume $L(v) = i$ and $(i, j) \in \mathcal{M}$. Let u' be the node previous to $f(v)$ in $\text{List}'(D, i, j)$. According to Part (c) of Claim C.3, $u' \in V_2'$ if and only if $v \in \mathcal{S}_3 \cup \mathcal{S}_4$. When $v \in \mathcal{S}_3 \cup \mathcal{S}_4$, $v \in \mathcal{S}_3$ if $L'(u') = j^\uparrow$, and $v \in \mathcal{S}_4$ if $L'(u') = j^\downarrow$. When $v \notin \mathcal{S}_3 \cup \mathcal{S}_4$, $v \in \mathcal{S}_1$ if $L'(v') = i^\uparrow$, and $v \in \mathcal{S}_2$ if $L'(v') = i^\downarrow$.

□

Appendix D. Proof of Lemma 6.1

Let $\boldsymbol{e_i}$ denote the vector whose coordinates are all 0 except the i-th that equals 1. The following lemmas will be used in the proof.

Lemma D.1. *[HLSZ19] Let $G_D = ([m], E_D)$ be a dependency graph and $\boldsymbol{p}$ be a probability vector beyond the Shearer's bound. Suppose $i, i_1, i_2, \cdots, i_{k-1}, i'$ form a shortest path from i to i' in G_D. Then for any $q \leq p_{i'}$, $\boldsymbol{p} - q\boldsymbol{e}_{i'} + \left(\prod_{\ell \in [k-1]} \frac{1-p_{i_\ell}}{p_{i_\ell}}\right) \cdot \frac{1-p_i}{p_{i'}} \cdot q\boldsymbol{e_i}$ is also beyond the Shearer's bound.*

Without loss of generality, we assume that $p_i - p_a$ is rational for each $i \in [m]$. By contradiction, let $\boldsymbol{p}$ be such a vector which is beyond Shearer's bound. Let $S_+ := \{i \in [m] : p_i > p\}$ and $S_- := \{i \in [m] : p_i < p\}$. Let Δ_p be a real number such that the following hold:

- For each $i \in S_+$, $p_i - p_a = \gamma_i \cdot \Delta_p$ for some $\gamma_i \in \mathbb{N}^+$. Intuitively, we cut $p_i - p_a$ into γ_i pieces each of size Δ_p. Besides, we call such pieces *positive pieces*.
- For each $i \in S_-$,

$$p_a - p_i = \tau_i \cdot K \cdot \left(\frac{1 - p_a}{p_a}\right)^{d-1} \cdot \frac{\Delta_p}{p_a}$$

  for some $\tau_i \in \mathbb{N}^+$. Intuitively, we cut $p_a - p_i$ into $\tau_i \cdot K$ pieces each of size $\left(\frac{1-p_a}{p_a}\right)^{d-1} \cdot \frac{\Delta_p}{p_a}$. We call such pieces *negative pieces*.

We use $\mathcal{R} := \{(i, r) : i \in S_+, r \in [\gamma_i]\}$ and $\mathcal{T} \triangleq \{(i', t, k) : i' \in S_-, t \in [\tau_{i'}], k \in [K]\}$ to denote the set of positive pieces and negative pieces respectively.

For convenience, let $\gamma_i = 0$ if $i \notin S_+$, and $\tau_i = 0$ if $i \notin S_-$. Then Condition (c) can be restated as: for $f(S) = T$, the positive pieces in S are no more than the negative pieces in T, i.e.,

$$\sum_{i \in S} \gamma_i \le \sum_{i' \in T} \tau_{i'}. \tag{16}$$

The basic idea of Lemma 6.1 is relatively simple: for each $S \in \mathcal{S}$, we move positive pieces in S to $f(S)$ such that (i) all the positive pieces in S are absorbed by the negative pieces in $f(S)$ and (ii) the resulted probability vector is still beyond Shearer's bound. Finally, all positive pieces will be absorbed, and we will get a vector strictly smaller than $\boldsymbol{p}$. By Lemma D.1, this vector is beyond Shearer's bound, which makes a contradiction.

For $i' \in [m]$, remember Condition (a) which says that there are at most K subsets $S \subset \mathcal{S}$ such that $i' \in f(S)$, and we use $S^1_{i'}, S^2_{i'}, \cdots$ to represent these subsets. Let $g : \mathcal{R} \to \mathcal{T}$ be a injection mapping each $(i, r) \in \mathcal{R}$ to some $(i', t, k) \in \mathcal{T}$ satisfying that (i) $i \in S^k_{i'}$ and (ii)

$$\sum_{i_0 \in S^k_{i'}, i_0 < i} \gamma_{i_0} + r = \sum_{i_1 \in f(S^k_{i'}), i_1 < i'} \tau_{i_1} + t.$$

By (16), one can verify that such mapping g exists. In addition, according to Condition (b), if $g(i, r) = (i', t, k)$, then $\mathrm{dist}(i, i') \le d$.

In the following, we will apply Lemma D.1 repeatedly.

Let g_0 be g, S_0 be S_- and $\mathcal{R}_0$ be $\mathcal{R}$. Given an injection $g_\kappa : \mathcal{R} \to \mathcal{T}$, S_κ and $\mathcal{R}_\kappa$ where $\mathrm{dis}(i, j) \le d$ if $g_\kappa(i, r) = (j, t, k)$, we construct another injection $g_{\kappa+1} : \mathcal{R} \to \mathcal{T}$, $S_{\kappa+1}$ and $\mathcal{R}_{\kappa+1}$ as follows. There are two possible cases for g_κ, S_κ and $\mathcal{R}_\kappa$.

(1) there exists i, r, j, t, k such that $(i, r) \in \mathcal{R}_\kappa$, $g_\kappa(i, r) = (j, t, k)$ and there is a shortest path between i and j such that no vertex in S_κ is on the path;
(2) For each $g_\kappa(i, r) = (j, t, k)$ where $(i, r) \in \mathcal{R}_\kappa$ and each shortest path between i and j, there is a vertex in S_κ on the path.

For case (1), we let $g_{\kappa+1} = g_\kappa$, $\mathcal{R}_{\kappa+1} = \mathcal{R}_\kappa \setminus \{(i, r)\}$, and

$$S_{\kappa+1} = \{j \in S_- : \text{there exists } i, r, t, k \text{ where } (i, r) \in \mathcal{R}_{\kappa+1} \text{ such that } g_{\kappa+1}(i, r) = (j, t, k)\}.$$

For case (2), there must be $(i_1, r_1, j_1, t_1, k_1), \cdots, (i_n, r_n, j_n, t_n, k_n)$ for some $n \in \mathbb{N}^+$ such that

- $(i_\ell, r_\ell) \in \mathcal{R}_\kappa$, $j_\ell \in S_\kappa$, $g_\kappa(i_\ell, r_\ell) = (j_\ell, t_\ell, k_\ell)$ for each $\ell \in [n]$,
- $j_{\ell+1}$ is on a shortest path between i_ℓ and j_ℓ for each $\ell \in [n-1]$,
- j_1 is on a shortest path between i_n and j_n.

We define the injection $F(g_\kappa)$ as follows.

$$\begin{cases} F(g_\kappa)(i_n, r_n) & = (j_1, t_1, k_1), \\ F(g_\kappa)(i_\ell, r_\ell) & = (j_{\ell+1}, t_{\ell+1}, k_{\ell+1}) \text{ for each } \ell \in [n-1], \\ F(g_\kappa)(i, r) & = g_\kappa(i, r) \text{ for other } (i, r). \end{cases}$$

One can verify that $\mathrm{dis}(i, j) \le d$ if $F(g_\kappa)(i, r) = (j, t, k)$ and

$$N \triangleq \sum_{\substack{(i,r,j,t,k): \\ g_\kappa(i,r)=(j,t,k)}} \mathrm{dis}(i, j) \ge 1 + \sum_{\substack{(i,r,j,t,k): \\ F(g_\kappa)(i,r)=(j,t,k)}} \mathrm{dis}(i, j).$$

Since N is bounded, there must be a constant $\ell \le N$ and i, r, j, t, k such that $(i, r) \in \mathcal{R}_\kappa$, $F^\ell(g_\kappa)(i, r) = (j, t, k)$ and there is a shortest path between i and j such that no vertex in S_κ is on the path. Let $g_{\kappa+1} = F^\ell(g_\kappa)$, $\mathcal{R}_{\kappa+1} = \mathcal{R}_\kappa \setminus \{(i, r)\}$ and

$$S_{\kappa+1} = \{j \in S_- : \text{there exists } i, r, t, k \text{ where } (i, r) \in \mathcal{R}_{\kappa+1} \text{ such that } g_{\kappa+1}(i, r) = (j, t, k)\}.$$

One can verify that in both cases, $g_{\kappa+1}$ is an injection from $\mathcal{R}$ to $\mathcal{T}$ and $\mathrm{dis}(i, j) \le d$ if $g_{\kappa+1}(i, r) = (j, t, k)$.

Let g' be $g_{|\mathcal{R}|}$. For each $\ell \in [|\mathcal{R}|]$, let (i_ℓ, r_ℓ) be the unique element in $\mathcal{R}_{\ell-1} \setminus \mathcal{R}_\ell$. Let (j_ℓ, t_ℓ, k_ℓ) denote $g'(i_\ell, r_\ell)$. Thus, we have

- g' is an injection from $\mathcal{R}$ to $\mathcal{T}$,
- $\mathrm{dis}(i_\ell, j_\ell) \le d$ for each $\ell \in [|\mathcal{R}|]$,

- there is a shortest path between i_ℓ and j_ℓ such that $j_{\ell+1}, j_{\ell+2}, \cdots, j_{|\mathcal{R}|} \in S_\ell$ are not on the path.

For each $j \in S_-$, define

$$\eta_j = |\{(i,r) : g'(i,r) = (j,t,k) \text{ for some } t \in [\tau_j], k \in [K]\}|.$$

Because g' is an injection, we have $\eta_j \le \tau_j \cdot K$. Let

$$\boldsymbol{p}'' \triangleq \boldsymbol{p}' + \sum_{j \in S_-} (K \cdot \tau_j - \eta_j) \cdot \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{\Delta_p}{p} \cdot \boldsymbol{e}_j.$$

By $\boldsymbol{p}'$ is beyond Shearer's bound and $\eta_j \le K \cdot \tau_j$ for each $j \in S_-$, we have $\boldsymbol{p}''$ is also beyond Shearer's bound. For each $\ell \in [0, |\mathcal{R}|]$, let

$$\boldsymbol{p}_\ell \triangleq \boldsymbol{p}'' - \Delta_p \cdot \left(\sum_{\kappa \le \ell - 1} \left(\boldsymbol{e}_{i_\kappa} - \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{1}{p} \cdot \boldsymbol{e}_{j_\kappa} \right) + \boldsymbol{e}_{i_\ell} - \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{1}{p + \Delta_p} \cdot \boldsymbol{e}_{j_\ell} \right).$$

Then we have the following claim.

Claim D.2. *For $\ell \in [0, |\mathcal{R}|]$, $\boldsymbol{p}_\ell$ is beyond Shearer's bound.*

Proof. We prove this claim by induction. Obviously, $\boldsymbol{p}_0$ is beyond Shearer's bound. In the following, we prove that if $\boldsymbol{p}_{\ell-1}$ is beyond Shearer's bound, then $\boldsymbol{p}_\ell$ is also beyond Shearer's bound.

Let

$$\boldsymbol{q} \triangleq \boldsymbol{p}'' - \Delta_p \cdot \sum_{\kappa \le \ell - 1} \left(\boldsymbol{e}_{i_\kappa} - \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{1}{p} \cdot \boldsymbol{e}_{j_\kappa} \right).$$

Obviously, $\boldsymbol{q} \ge \boldsymbol{p}_{\ell-1}$. By $\boldsymbol{p}_{\ell-1}$ is beyond Shearer's bound, we have $\boldsymbol{q}$ is also beyond Shearer's bound. Note that there is a shortest path $i_\ell, k_1, k_2, \cdots, k_n, j_\ell$ between i_ℓ and j_ℓ such that $j_{\ell+1}, j_{\ell+2}, \cdots, j_{|\mathcal{R}|}$ are not on the path. Because $\boldsymbol{q}$ is beyond Shearer's bound, by Lemma D.1, we have

$$\boldsymbol{q}' \triangleq \boldsymbol{q} - \Delta_p \cdot \left(\boldsymbol{e}_{i_\ell} - \left(\prod_{t \in [n]} \frac{1 - q_{k_t}}{q_{k_t}} \right) \cdot \frac{1}{q_i} \cdot \boldsymbol{e}_{j_\ell} \right)$$

is also beyond Shearer's bound. Meanwhile, by $(i_\ell, r_\ell) \in \mathcal{R}$, we have

$$q_{i_\ell} = p_i' - \Delta_p \sum_{\kappa \in \ell - 1} \mathbb{1}(i_\kappa = i_\ell) \ge p_i' - (\gamma_i - 1)\Delta_p \ge p_i + \Delta_p.$$

For each $t \in [n]$, if $k_t \notin S_-$, we have $q_{k_t} \ge p$. Otherwise, $k_t \in S_-$, and $k_t \ne j_\kappa$ for each $\kappa \ge \ell$. Thus, we have $\sum_{\kappa \in \ell-1} \mathbb{1}(j_\kappa = k_t) = \eta_{k_t}$. Therefore,

$$\begin{aligned} q_{k_t} &= p_{k_t}' + (K \cdot \tau_{k_t} - \eta_{k_t}) \cdot \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{\Delta_p}{p} + \sum_{\kappa \in \ell - 1} \mathbb{1}(j_\kappa = k_t) \cdot \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{\Delta_p}{p} \\ &= p_{k_t}' + (K \cdot \tau_{k_t} - \eta_{k_t}) \cdot \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{\Delta_p}{p} + \eta_{k_t} \cdot \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{\Delta_p}{p} = p. \end{aligned}$$

By $\mathrm{dis}(i,j) \ge d$, $q_{i_\ell} \ge p + \Delta_p$ and $q_{k_t} \ge p$ for each $t \in [n]$, we have

$$\left(\prod_{t \in [n]} \frac{1 - q_{k_t}}{q_{k_t}} \right) \cdot \frac{1}{q_i} < \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{1}{p + \Delta_p}.$$

Thus, by $\boldsymbol{q}'$ is beyond Shearer's bound, we have

$$\boldsymbol{p}_\ell = \boldsymbol{q} - \Delta_p \cdot \left(\boldsymbol{e}_{i_\ell} - \left(\frac{1-p}{p}\right)^{d-1} \cdot \frac{1}{p + \Delta_p} \cdot \boldsymbol{e}_{j_\ell} \right)$$

is also beyond Shearer's bound. □

Thus, we have $\boldsymbol{p}_{|\mathcal{R}|}$ is beyond Shearer's bound. It is easy to verify that $\boldsymbol{p}_{|\mathcal{R}|} < \boldsymbol{p}$, which is contradictory with that $\boldsymbol{p}$ is on Shearer's boundary.

Appendix E. Missing part in the proof of Theorem 6.2

Proof of Claim 6.3. Observe that for each $(v,v') \in E_U^k$, if $(v,v') \notin \mathcal{M}$, then one of its neighboring edge (v_0,v_1) is in T_k and satisfies that $\delta_{v,v'} \le \delta_{v_0,v_1}$. Here, we say two edges neighboring if they share a common vertex. Besides, note that each edge has at most 2Δ neighboring edges. So

$$\sum_{(v_0,v_1)\in T_k} \delta^2_{v_0,v_1} \ge \frac{1}{2\Delta} \sum_{(v,v')\in E_U^k} \delta^2_{v,v'}. \tag{17}$$

Moreover, according to Lemma 4.2 and 4.5, it has that

$$\sum_{(v,v')\in E_U^k} \delta^2_{v,v'} \ge \frac{1}{|E_U^k|} \cdot \left(\sum_{(v,v')\in E_U^k} \delta_{v,v'} \right)^2 \ge \frac{|V_U^k| \cdot \Delta^2}{|E_U^k|} \cdot \left(F^+(G_B(G_D), \boldsymbol{p})\right)^2, \tag{18}$$

By combining Inequality 17, 18 and the fact that $2|E_U^k| \le |V_U^k|\Delta$, we finish the proof. □

Let $K := (\Delta+1)|V_U|$, $d := D+2$, $\mathcal{S} := \{V_U^1, V_U^2, \cdots\}$, and $f(V_U^k) := T_k$. In the following, we check that all the three conditions in Lemma 6.1 hold.

Condition (a). That is, we want to show $|\{k : T_k \ni v\}| \le (\Delta+1)|V_U|$ for each $v \in V_D$. Observe that if $v \in T_k$, then $v \in \mathcal{N}^+(V_U^k)$. So

$$|\{k : T_k \ni v\}| \le |\{k : \mathcal{N}^+(V_U^k) \ni v\}| \le |\{k : \mathcal{N}^+(v) \cap V_U^k \ne \emptyset\}| \le \sum_{v' \in \mathcal{N}^+(v)} |\{k : V_U^k \ni v'\}| \le (\Delta+1)\cdot|V_D|.$$

The last inequality uses the fact that $h(k',u) \ne h(k,u)$ if $k \ne k'$.

Condition (b). That is, we want to show $\mathrm{dist}(v,v') \le D+2$ for any $v \in V_U^k$ and $v' \in T_k$. This is obvious, because if $v' \in T_k$, then $v' \in \mathcal{N}^+(V_U^k)$.

Condition (c). We verify that

$$\left(\frac{1-p_a}{p_a}\right)^{D+1} \cdot \frac{K}{p_a} \cdot \sum_{i\in S} \max\{p_i - p_a, 0\} \le \sum_{i\in T} \max\{p_a - p_i, 0\}. \tag{19}$$

On one hand, noting that $\max\{(1+\varepsilon)p_v^- - p_a, 0\} \le \max\{(1+\varepsilon)p_v - p_a, 0\} \le q$, we have

$$\text{L.H.S of (19)} \le \left(\frac{1-p_a}{p_a}\right)^{D+1} \cdot \frac{(\Delta+1)|V_U|^2}{p_a} \cdot q. \tag{20}$$

On the other hand, observe that

$$\begin{aligned}\max\{p_a - (1+\varepsilon)p_v^-, 0\} &\ge p_a - (1+\varepsilon)p_v^- = (p_a + q - (1+\varepsilon)p_v^-) - q \ge (1+\varepsilon)(p_v - p_v^-) - q \\ &\ge (p_v - p_v^-) - q,\end{aligned}$$

where the last inequality is due to the assumption that $(1+\varepsilon)\boldsymbol{p} \le (p_a+q, \cdots, p_a+q)$. Then

$$\begin{aligned}\text{R.H.S of (19)} &\ge \left(\sum_{v\in V_U^k} (p_v - p_v^-)\right) - |\mathcal{N}^+(V_U^k)|q \ge \frac{2}{17}\left(\sum_{(v_0,v_1)\in T_k} \delta^2_{v_0,v_1}\right) - \Delta|V_U|q \\ &\ge \frac{2}{17}\left(F^+(G_B(G_D), \boldsymbol{p})\right)^2 - \Delta|V_U|q. \qquad (21)\end{aligned}$$

Putting Inequality 20 and 21 together and noting that $\frac{1-p_a}{p_a} \ge 1$.