AIMS Mathematics, 11(7): 19659–19715.
DOI: 10.3934/math.2026798
Received: 16 April 2026
Revised: 23 June 2026
Accepted: 25 June 2026
Published: 03 July 2026

https://www.aimspress.com/journal/Math

Review

The Lovász local lemma: foundations and applications

Igal Sason[1,2,*]

[1] The Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion-Israel Institute of Technology, Haifa 3200003, Israel

[2] Faculty of Mathematics, Technion–Israel Institute of Technology, Haifa 3200003, Israel

* **Correspondence:** Email: eeigal@technion.ac.il; Tel: +97248294699.

Abstract: The Lovász local lemma (LLL) is a central tool in probabilistic combinatorics, providing a sufficient condition under which a finite collection of undesirable events with limited dependencies can be simultaneously avoided with positive probability. This paper offers a self-contained expository treatment of the lemma and its strengthened versions, emphasizing mathematical foundations, conceptual clarity, and applications. We begin with a pedagogically motivated proof of the LLL based entirely on unconditional probability inequalities. Particular attention is given to the symmetric form of the lemma and several subsequent strengthenings. We also discuss a variety of classical applications of both the symmetric and asymmetric forms of the LLL in combinatorics and graph theory, including bounds for the edge-disjoint paths problem, satisfiability of Boolean formulas in conjunctive normal form, lower bounds on diagonal and off-diagonal Ramsey numbers, hypergraph coloring results, structural properties of directed graphs, and acyclic graph colorings. Additional observations and refinements are provided throughout. We also introduce the algorithmic framework of Moser and Tardos, highlighting its constructive counterpart to the LLL, together with an introduction to the entropy-compression principle. The lopsided LLL, a refinement of the LLL, is presented along with an application to the Latin transversal problem. We further discuss the cluster-expansion lemma and its relation to the LLL, and present an alternative treatment of the Latin transversal problem from the cluster-expansion perspective that yields an improved result. The exposition concludes with a high-level overview of the iterated LLL, also known as the semi-random method.

Keywords: Lovász local lemma; Moser–Tardos algorithm; probabilistic combinatorics
Mathematics Subject Classification: 05C35, 05C50, 05C69, 05C76, 94A15

1. Introduction

In many applications of the probabilistic method, one considers a finite collection of undesirable events $\{A_i\}_{i=1}^n$ and seeks to show that, with positive probability, none of them occurs. If the sum of their probabilities is strictly less than one, then the union bound guarantees this. If, instead, the events are independent and each occurs with probability strictly less than one, then the probability that none of the events occurs equals the product of the probabilities of their complements, which is positive.

In typical combinatorial settings, however, the events exhibit significant dependencies and their probabilities are often not sufficiently small for direct probabilistic estimates to be effective. Nevertheless, in many applications, each event depends on only a limited number of others, and the event probabilities are sufficiently small compared to the reciprocal of the maximum number of events on which any given event depends. This observation underlies a fundamental result, namely the Lovász local lemma (LLL), introduced by Erdős and Lovász [1]. The lemma provides a sufficient condition under which a finite collection of events with limited dependencies can be simultaneously avoided with positive probability.

The LLL and its variants have become indispensable tools in probabilistic combinatorics and theoretical computer science. Over the past several decades, they have found applications across a broad spectrum of areas, including probabilistic and extremal combinatorics. Background on the probabilistic method can be found in [2–4]; graph-theoretic aspects are also covered in [5–7]; extremal combinatorics is treated in [8, 9]; and algorithmic aspects are discussed in [10]. Representative combinatorial applications of the LLL and its variants to graph-theoretic problems include edge-disjoint cycles [11–13], Ramsey theory [14–16], directed graphs [17, 18], graph coloring [19–21], combinatorial arrays [22], and partially ordered sets [23, 24]. Further applications arise in satisfiability and computational complexity. Classical SAT and computational complexity are treated in [25–27]; algorithmic aspects of the LLL for satisfiability are developed in [28–30]; and applications to constraint satisfaction problems are discussed in [31, 32]. Quantum extensions of the LLL are considered in [33]. The LLL has also found numerous applications in information theory, coding theory, communication networks, and group testing. Interactive communication is treated in [34–36]; communication-network applications appear in [37, 38]; coding-theoretic applications include synchronization strings [39], permutation and insertion–deletion codes [40–42], spatially-coupled codes [43, 44], multimedia and IPP codes [45, 46], and separating codes [47–49]; group-testing applications are discussed in [50, 51]; and superimposed-code constructions and related combinatorial structures are studied in [52–54]. Connections between the LLL and statistical physics are explored in [55, 56]; cluster-expansion methods from statistical physics are used to strengthen the LLL in [57]. For surveys devoted to the LLL, see [58, 59]; for broader expository accounts of the probabilistic method and its applications, including the LLL, see [60–62].

In its classical form, the LLL is *nonconstructive*: It guarantees the existence of an assignment that avoids all the undesirable events, yet does not, in general, provide an explicit method for finding such an assignment. To describe the dependence structure among events underlying the LLL, we introduce the following notions. Throughout, $[n] := \{1, \dots, n\}$ for $n \in \mathbb{N}$, with $[0] := \varnothing$.

Definition 1.1. (Mutual independence) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $n \in \mathbb{N}$, and $A_1, \dots, A_n \in \mathcal{F}$. For $i \in [n]$, the event A_i is independent of the σ-algebra $\sigma(\{A_j : j \in [n] \setminus \{i\}\})$ if

$$\mathbb{P}(A_i \cap B) = \mathbb{P}(A_i)\,\mathbb{P}(B) \qquad \text{for all } B \in \sigma(A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_n). \tag{1.1}$$

If this holds for all $i \in [n]$, then the events $A_1, \ldots, A_n$ are said to be *mutually independent*.

Remark 1.2. For $i \in [n]$, $\sigma(\{A_j : j \in [n] \setminus \{i\}\})$ is called the *σ-algebra generated* by the events $A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n$, and it is defined as the smallest σ-algebra containing these events. Since the family $A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n$ is finite, the σ-algebra it generates consists precisely of all events that are obtained from these events by finitely many applications of unions, intersections, and complements.

Definition 1.3. (Dependency digraph) Let $\{A_i\}_{i=1}^n$ be events in a probability space. A directed graph (digraph) $D = ([n], E)$ is a *dependency digraph* for $\{A_i\}_{i=1}^n$ if, for all $i \in [n]$, the event A_i is independent of the σ-algebra generated by $\{A_j : j \neq i,\ (i, j) \notin E\}$, where (i, j) denotes an arc from i to j.

Remark 1.4. Definition 1.3 requires more than pairwise independence: for each $i \in [n]$, the event A_i must be independent of the σ-algebra generated by the events $\{A_j : j \neq i,\ (i, j) \notin E\}$.

Remark 1.5. (Nonuniqueness of the dependency digraph) In general, the dependency digraph for the events $\{A_i\}_{i=1}^n$ is not unique. Indeed, if $D = ([n], E)$ is a dependency digraph, then any digraph obtained by adding arcs to E is also valid, since enlarging E weakens the independence requirement.

Theorem 1.6. *(LLL)* Let $A_1, \ldots, A_n$ be events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $D = ([n], E)$ be a dependency digraph for these events. If there exist $x_1, \ldots, x_n \in [0, 1)$ such that

$$\mathbb{P}(A_i) \leq x_i \prod_{j:(i,j)\in E} (1 - x_j), \quad \forall\, i \in [n], \tag{1.2}$$

then,

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n} (1 - x_i). \tag{1.3}$$

In particular, the probability that none of the events $\{A_i\}_{i=1}^n$ occurs is positive.

To illustrate the power of the LLL beyond the settings covered by the union bound and independence, suppose there exists an absolute constant $c < 1/2$ such that, for all $i \in [n]$,

$$\mathbb{P}(A_i) \leq c, \qquad \sum_{j:(i,j)\in E} \mathbb{P}(A_j) \leq \tfrac{1}{4}.$$

Setting $x_i := 2\,\mathbb{P}(A_i)$ yields $x_i \in [0, 1)$, and

$$\prod_{j:(i,j)\in E} (1 - x_j) \geq 1 - \sum_{j:(i,j)\in E} x_j = 1 - 2 \sum_{j:(i,j)\in E} \mathbb{P}(A_j) \geq \tfrac{1}{2}.$$

Consequently,

$$\mathbb{P}(A_i) = \frac{x_i}{2} \leq x_i \prod_{j:(i,j)\in E} (1 - x_j), \quad \forall\, i \in [n].$$

By the LLL (Theorem 1.6), since $0 \leq x_i \leq 2c < 1$ for all $i \in [n]$, it follows that

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n} (1 - x_i) \geq (1 - 2c)^n,$$

yielding an explicit positive lower bound on the probability that no event in $\{A_i\}_{i=1}^n$ occurs, with the lower bound decaying exponentially in n. This extends a statement in [6, p. 12], restricted to the case where $c = \frac{1}{8}$, where the positivity of this probability is noted.

We briefly review several major algorithmic and structural developments related to the LLL. Over the years, a variety of important extensions and refinements of the lemma have been developed, many of which address its historically nonconstructive nature. A central line of research, initiated in 1991 by Beck [63] and Alon [64], concerns *algorithmic versions* of the lemma. In this context, the *variable version of the LLL*, where events are modeled as functions of independent random variables, plays a fundamental role. A landmark breakthrough in this direction is due to Moser and Tardos [65], building on the earlier breakthrough of Moser [66], who introduced an efficient randomized resampling algorithm for avoiding all bad events in this variable setting. Subsequently, Kolipaka and Szegedy [67] showed that the Moser–Tardos algorithm remains efficient throughout the full regime characterized by Shearer's bound [68].

This framework has also led to a refined understanding of dependency structures, including sharp characterizations of feasibility boundaries. Recent developments have further revealed gaps between the variable setting and the abstract dependency-graph formulation of the LLL [69, 70]. Moreover, [70] established improved efficiency guarantees in the variable setting, thereby further separating it from the abstract LLL framework, for which Shearer's bound is optimal under dependency-graph formulations.

The analysis of these algorithms is closely related to the *entropy-compression method*, which provides a unifying framework for bounding the probability of long resampling sequences and has evolved into a powerful and widely used technique in probabilistic combinatorics [71–73]. Subsequent works have further extended and refined this framework; see, e.g., the generalizations of the Moser–Tardos algorithm by Pegden [74] and by Harvey and Vondrák [75]. Deterministic algorithms for the LLL have also been developed [76, 77]. Another important direction concerns lopsided variants of the LLL, which replace the classical mutual independence assumptions by weaker asymmetric dependency conditions [78]. Several strengthenings of the LLL, including the lopsided LLL and the cluster-expansion lemma, have proven successful in applications where the standard LLL does not apply directly (see, e.g., [7, 57]).

This paper offers a self-contained expository treatment of the LLL and its strengthenings, emphasizing mathematical foundations and applications and providing observations and refinements throughout. Its main features are as follows:

(1) A reformulated proof of the LLL (Section 2), together with a pedagogical motivation in Remark 2.2.

(2) A treatment of the symmetric version of the LLL (Section 3), including strengthened guarantees on the probability of avoiding all bad events (Corollaries 3.4 and 3.7, followed by Remark 3.8).

(3) A presentation of classical combinatorial applications of the symmetric and asymmetric LLL (Section 4), together with several observations and refinements, including:

   (a) Tightened results in Theorems 4.1 and 4.2, concerning the problems of edge-disjoint paths and satisfiability of Boolean formulas, respectively.

   (b) A slight tightening of Spencer's lower bound on diagonal Ramsey numbers (Theorem 4.5), together with a refined asymptotic formulation (Proposition 4.6 and Remark 4.7). A tightened lower bound on off-diagonal Ramsey numbers, building on Spencer's application of the (asymmetric) LLL, is also presented (Theorem 4.9).

(c) Strengthened results on hypergraph colorings (Theorems 4.17, 4.18, 4.20, and Remark 4.21).

(d) Strengthened bounds on cycle lengths in digraphs (Corollaries 4.24, 4.25, and Remark 4.26).

(e) An application of the classical LLL from [19] leads to a derivation of a tightened upper bound on the acyclic chromatic number of a graph (Proposition 4.34).

(4) A presentation of the Moser–Tardos algorithm [65] for the constructive proof of the LLL in the variable setting (Section 5), and an introduction to the entropy-compression principle (Section 6).

(5) A presentation of the lopsided LLL, together with its application to the existence of Latin transversals in matrices (Section 7).

(6) A presentation of the cluster-expansion lemma, its implications, and an alternative treatment of the Latin transversal problem by that lemma, which yields an improved result (Section 8).

(7) A presentation of the iterated LLL (Section 9) and a brief outlook on open directions (Section 10).

2. A reformulated proof of the LLL

This section presents a reformulated proof of the LLL in Theorem 1.6, followed by its pedagogical motivation in Remark 2.2.

Proof. We present a proof based entirely on unconditional probability inequalities. In particular, no step requires assuming that a conditioning event has positive probability.

Step 1: An auxiliary lemma. For $S \subseteq [n]$, define

$$F(S) := \bigcap_{j \in S} \overline{A_j}. \tag{2.1}$$

Lemma 2.1. Assume that the conditions of Theorem 1.6 hold. Then, for every $S \subset [n]$ and $i \in [n] \setminus S$,

$$\mathbb{P}(A_i \cap F(S)) \leq x_i \, \mathbb{P}(F(S)). \tag{2.2}$$

Proof of Lemma 2.1. We prove (2.2) by induction on $|S|$.

Base case. If $S = \varnothing$, then $F(S) = \Omega$, and (2.2) reduces to $\mathbb{P}(A_i) \leq x_i$. Indeed, by (1.2),

$$\mathbb{P}(A_i) \leq x_i \prod_{j:(i,j) \in E} (1 - x_j) \leq x_i,$$

where the last inequality holds since each factor satisfies $1 - x_j \in [0, 1]$.

Induction hypothesis. Fix an integer $m \geq 1$ and assume that (2.2) holds for every $S' \subset [n]$ with $|S'| < m$ and every $i' \in [n] \setminus S'$.

Induction step. Let $S \subset [n]$ be an arbitrary set with $|S| = m$, and let $i \in [n] \setminus S$. Define

$$S_1 := \{j \in S : (i, j) \in E\}, \qquad S_2 := S \setminus S_1. \tag{2.3}$$

If $S_1 = \varnothing$, then $(i, j) \notin E$ and $j \neq i$ for all $j \in S$. By Definition 1.3, the event A_i is independent of the σ-algebra generated by $\{A_j : j \in S\}$, and in particular it is independent of

$$F(S) = \bigcap_{j \in S} \overline{A_j}.$$

Therefore,

$$\mathbb{P}(A_i \cap F(S)) = \mathbb{P}(A_i)\,\mathbb{P}(F(S)) \leq x_i\,\mathbb{P}(F(S)),$$

and (2.2) follows.

Assume now that $S_1 \neq \varnothing$. By (2.3), $S_2 \subseteq \{j : (i, j) \notin E\}$. Since $i \notin S$ and $S_2 \subseteq S$, we have $i \notin S_2$. Therefore, $S_2 \subseteq \{j : j \neq i,\ (i, j) \notin E\}$. By Definition 1.3, A_i is independent of the σ-algebra generated by $\{A_j : j \in S_2\}$; in particular, A_i is independent of $F(S_2)$. As $S_2 \subseteq S$, we have $F(S) \subseteq F(S_2)$, so

$$\begin{aligned} \mathbb{P}(A_i \cap F(S)) &\leq \mathbb{P}(A_i \cap F(S_2)) \\ &= \mathbb{P}(A_i)\,\mathbb{P}(F(S_2)) \\ &\leq x_i \prod_{j:(i,j)\in E} (1 - x_j)\,\mathbb{P}(F(S_2)), \end{aligned} \tag{2.4}$$

where the last inequality holds by (1.2). We next derive a lower bound on $\mathbb{P}(F(S))$ in terms of $\mathbb{P}(F(S_2))$. Let $S_1 = \{j_1, \ldots, j_r\}$ and, for $t \in [r]$, let

$$T_t := S_2 \cup \{j_1, \ldots, j_{t-1}\}.$$

Note that $|T_t| = |S_2| + t - 1 \leq |S| - 1$, hence, $|T_t| < |S|$. Define, for $t = 0, 1, \ldots, r$, the events

$$G_t := F(S_2) \cap \bigcap_{s=1}^{t} \overline{A_{j_s}},$$

so that $G_0 = F(S_2)$ and $G_r = F(S)$ (recall that $S = S_1 \dot\cup S_2$ is a disjoint union of S_1 and S_2). Then, $G_t = \overline{A_{j_t}} \cap G_{t-1}$, and therefore

$$\mathbb{P}(G_t) = \mathbb{P}(G_{t-1}) - \mathbb{P}(A_{j_t} \cap G_{t-1}), \qquad t \in [r]. \tag{2.5}$$

Since $G_{t-1} = F(T_t)$ and $j_t \notin T_t$, the induction hypothesis applied to the pair $(i', S') = (j_t, T_t)$ gives

$$\mathbb{P}(A_{j_t} \cap G_{t-1}) = \mathbb{P}(A_{j_t} \cap F(T_t)) \leq x_{j_t}\,\mathbb{P}(F(T_t)) = x_{j_t}\,\mathbb{P}(G_{t-1}).$$

Substituting into (2.5) yields

$$\mathbb{P}(G_t) \geq (1 - x_{j_t})\,\mathbb{P}(G_{t-1}), \qquad t \in [r]. \tag{2.6}$$

Iterating for $t = 1, 2, \ldots, r$, we obtain

$$\mathbb{P}(F(S)) = \mathbb{P}(G_r) \geq \prod_{t=1}^{r} (1 - x_{j_t})\,\mathbb{P}(G_0) = \prod_{j \in S_1} (1 - x_j)\,\mathbb{P}(F(S_2)). \tag{2.7}$$

Since $S_1 \subseteq \{j : (i,j) \in E\}$ and each factor $1 - x_j$ lies in $[0,1]$, it follows that

$$\prod_{j \in S_1} (1 - x_j) \geq \prod_{j:(i,j)\in E} (1 - x_j). \tag{2.8}$$

Combining (2.4), (2.7), and (2.8) gives

$$\begin{aligned} \mathbb{P}(A_i \cap F(S)) &\leq x_i \prod_{j:(i,j)\in E} (1 - x_j)\, \mathbb{P}(F(S_2)) \\ &\leq x_i \prod_{j \in S_1} (1 - x_j)\, \mathbb{P}(F(S_2)) \\ &\leq x_i\, \mathbb{P}(F(S)), \end{aligned}$$

which is (2.2). This completes the induction and proves Lemma 2.1. □

Step 2: Concluding the proof of the LLL. For $k = 0, 1, \ldots, n$, define

$$F_k := \bigcap_{j=1}^{k} \overline{A_j}, \quad \text{with } F_0 = \Omega. \tag{2.9}$$

Let $k \in [n]$ and $S = \{1, \ldots, k-1\}$ (if $k = 1$, then $S = \varnothing$). By (2.1) and (2.9), we have $F(S) = F_{k-1}$. Applying Lemma 2.1 with this choice of S and $i = k$ (noting that $i \in [n] \setminus S$), we obtain

$$\mathbb{P}(A_k \cap F_{k-1}) \leq x_k\, \mathbb{P}(F_{k-1}),$$

and therefore

$$\begin{aligned} \mathbb{P}(F_k) &= \mathbb{P}(F_{k-1}) - \mathbb{P}(A_k \cap F_{k-1}) \\ &\geq (1 - x_k)\, \mathbb{P}(F_{k-1}), \end{aligned} \tag{2.10}$$

with $\mathbb{P}(F_0) = 1$. Iterating (2.10) for $k = 1, 2, \ldots, n$ and using (2.9), we obtain

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) = \mathbb{P}(F_n) \geq \prod_{i=1}^{n} (1 - x_i),$$

which is (1.3). In particular, since $x_i \in [0, 1)$ for all $i \in [n]$, it follows that

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$

This completes the proof. □

Remark 2.2. (On avoiding conditioning assumptions) A common presentation of the LLL proves a variation of Lemma 2.1, namely,

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j \in S} \overline{A_j}\right) \leq x_i, \qquad \forall\, S \subset [n], \quad i \in [n] \setminus S, \tag{2.11}$$

by manipulating conditional probabilities via identities such as

$$\mathbb{P}(A \mid B \cap C) = \frac{\mathbb{P}(A \cap B \mid C)}{\mathbb{P}(B \mid C)}, \tag{2.12}$$

and expressing them in terms of conditional probabilities of the form

$$\mathbb{P}\left(A_i \cap \bigcap_{j \in S \setminus S'} \overline{A_j} \,\Big|\, \bigcap_{j \in S'} \overline{A_j}\right) \quad \text{and} \quad \mathbb{P}\left(A_i \,\Big|\, \bigcap_{j \in S'} \overline{A_j}\right), \qquad S' \subset S.$$

See, for example, [1, pp. 616,617], [2, pp. 70–72], [3, pp. 147–150], [4, pp. 100–103], as well as [5, pp. 21–23], [7, pp. 226–228], [8, pp. 30–31], [9, pp. 280–282], [10, p. 266], [14, 15], and [16, pp. 111–114]. In standard proofs, such conditional probabilities are manipulated within an inductive argument. However, their definition requires that the corresponding conditioning events have positive probability, a fact whose validity is established only within the proof itself. Hence, intermediate steps involve expressions whose validity depends on properties that are justified only a posteriori. The approach adopted here avoids introducing conditional probabilities altogether. Instead, we work with inequalities such as

$$\mathbb{P}\left(A_i \cap \bigcap_{j \in S} \overline{A_j}\right) \leq x_i \, \mathbb{P}\left(\bigcap_{j \in S} \overline{A_j}\right), \tag{2.13}$$

which remain well-defined regardless of whether $\mathbb{P}(\cap_{j \in S} \overline{A_j})$ is positive or zero. This leads to a self-contained argument in which the positivity of $\mathbb{P}(\overline{A_1} \cap \cdots \cap \overline{A_n})$ follows without requiring separate justification of the positivity of intermediate conditioning events.

An alternative proof of the (asymmetric) LLL that does not rely on conditional probabilities, brought to our attention by one of the anonymous referees, is presented in the lecture notes of Vondrák [79].

3. Symmetric LLL

In a wide range of applications, the events $\{A_i\}_{i=1}^n$ satisfy uniform bounds on their probabilities and dependencies, allowing the conditions in Theorem 1.6 to be simplified. This gives the next result.

Theorem 3.1. *(Symmetric version of the LLL)* Let $A_1, \ldots, A_n$ be events in an arbitrary probability space. Suppose that, for every $i \in [n]$, the event A_i is independent of the σ-algebra generated by all the remaining events, except for at most $d \geq 1$ of them, and assume that $\mathbb{P}(A_i) \leq p$ for all $i \in [n]$. If $ep(d+1) \leq 1$, then

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) \geq \left(\frac{d}{d+1}\right)^n > e^{-\frac{n}{d}}. \tag{3.1}$$

In particular, the probability that none of the events occurs is positive.

Remark 3.2. If $d = 0$, then the events $\{A_i\}_{i=1}^n$ are mutually independent, so

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) = \prod_{i=1}^n \mathbb{P}(\overline{A_i}) \geq (1-p)^n > 0.$$

Thus, the probability that none of the events $\{A_i\}_{i=1}^n$ occurs is positive whenever $p < 1$, in contrast to the stronger condition $p \leq \frac{1}{e}$ that would result from formally substituting $d = 0$ into the condition $ep(d+1) \leq 1$.

Proof. Let $d \geq 1$, and set $x_j := \frac{1}{d+1}$ for all $j \in [n]$. By the assumption of Theorem 3.1, there exists a dependency digraph $D = ([n], E)$ in which every vertex has outdegree at most d (see Definition 1.3). Fix such a dependency digraph D for $\{A_i\}_{i=1}^n$. Then, for all $i \in [n]$,

$$\mathbb{P}(A_i) \leq \frac{1}{(d+1)e} \tag{3.2a}$$

$$\leq \frac{1}{d+1}\left(1+\frac{1}{d}\right)^{-d} \tag{3.2b}$$

$$= \frac{1}{d+1}\left(1-\frac{1}{d+1}\right)^{d} \tag{3.2c}$$

$$\leq x_i \prod_{j:(i,j)\in E}(1-x_j), \tag{3.2d}$$

where (3.2a) holds by the assumption of the theorem; Eq (3.2b) follows from the fact that the sequence $\{(1+\frac{1}{k})^k\}_{k\in\mathbb{N}}$ is monotonically increasing and converges to e as $k \to \infty$; Eq (3.2c) follows by straightforward algebra; finally, Eq (3.2d) holds by the choice $x_i = \frac{1}{d+1}$ for all $i \in [n]$ and since the outdegrees of the considered dependency digraph are at most d. It then follows from Theorem 1.6 that

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) \geq \prod_{i=1}^n (1-x_i) = \left(\frac{d}{d+1}\right)^n > e^{-\frac{n}{d}},$$

where the last inequality holds since $\left(1+\frac{1}{d}\right)^d < e$. □

By skipping inequality (3.2a) and starting instead from inequality (3.2b), the next sharper symmetric criterion is obtained.

Corollary 3.3. *(Spencer's bound)* Let $A_1, \ldots, A_n$ be events in an arbitrary probability space. Suppose that, for every $i \in [n]$, the event A_i is independent of the σ-algebra generated by all the remaining events, except for at most $d \geq 1$ of them, and assume that $\mathbb{P}(A_i) \leq p$ for all $i \in [n]$. If

$$p \leq \frac{d^d}{(d+1)^{d+1}}, \tag{3.3}$$

then

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) \geq \left(\frac{d}{d+1}\right)^n > e^{-\frac{n}{d}}. \tag{3.4}$$

Corollary 3.4. If $ep(d+\frac{1}{2}) \leq 1$, then (3.4) holds.

Proof. The condition $ep(d+\frac{1}{2}) \leq 1$ yields (3.3), as justified in Appendix A. Hence, Eq (3.4) holds by Corollary 3.3. □

Remark 3.5. The original version of the LLL, introduced by Erdős and Lovász in [1], asserts that

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} \overline{A_i}\Big) > 0$$

provided that $4pd \leq 1$, where p is an upper bound on the probability of each event A_i, and d is an upper bound on the number of dependencies of each event. The standard sharpened formulation of the symmetric LLL, stated in Theorem 3.1, establishes the condition $ep(d+1) \leq 1$. A further refinement, due to Spencer [15, Theorem 1.4], shows that the same conclusion holds under the weaker condition Eq (3.3). Theorem 3.1 and Corollary 3.3 also provide the explicit lower bound $\left(\frac{d}{d+1}\right)^n$ on the probability that none of the events $\{A_i\}_{i=1}^n$ occurs, thereby strengthening the classical assertion that this probability is merely positive. By Corollary 3.4, the lower bound in (3.4) holds, in particular, if $ep\left(d+\frac{1}{2}\right) \leq 1$.

In [15], Spencer defined $f(d)$ as the supremum of the set of all $x \in [0,1)$ such that, whenever $\mathbb{P}(A_i) \leq x$ for all $i \in [n]$ and each event A_i is independent of the σ-algebra generated by all but at most $d \geq 1$ of the remaining events, then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$

He further asked whether the limit $\lim_{d\to\infty} d\, f(d)$ exists, and if so, what its value is. The following theorem of Shearer [68, Theorem 2] answers this question.

Theorem 3.6. *(Shearer)* For every integer $d \geq 1$,

$$f(d) = \begin{cases} \frac{1}{2}, & \text{if } d = 1, \\ \frac{(d-1)^{d-1}}{d^d}, & \text{if } d \geq 2. \end{cases} \tag{3.5}$$

Consequently,

$$\lim_{d\to\infty} d f(d) = \frac{1}{e}, \tag{3.6}$$

and the constant e in the condition $ep(d+1) \leq 1$ of Theorem 3.1 is asymptotically best possible.

Corollary 3.7. *([10, Problem 319])* Let $A_1, \ldots, A_n$ be events in an arbitrary probability space. Suppose that, for all $i \in [n]$, the event A_i is independent of the σ-algebra generated by all the remaining events, except for at most $d \geq 1$ of them, and $\mathbb{P}(A_i) \leq p$. If $epd \leq 1$, then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$

Proof. For every $d \in \mathbb{N}$, we have $\frac{1}{ed} < f(d)$. Indeed, if $d = 1$, then $\frac{1}{e} < \frac{1}{2} = f(1)$, and if $d \geq 2$, then

$$\frac{1}{ed} < \frac{1}{d}\left(1 + \frac{1}{d-1}\right)^{-(d-1)} = \frac{(d-1)^{d-1}}{d^d} = f(d).$$

Hence, if $epd \leq 1$, then $p < f(d)$, and the result follows from Theorem 3.6. □

Remark 3.8. Corollary 3.4 is not implied by Corollary 3.7. Indeed, under the condition $epd \leq 1$ of Corollary 3.7, the probability that none of the events $\{A_i\}_{i=1}^n$ occurs is only guaranteed to be positive for a dependency digraph with maximum degree at most d. By contrast, under the stronger condition $ep\left(d + \frac{1}{2}\right) \leq 1$ of Corollary 3.4, this probability is lower-bounded by $e^{-n/d}$ whenever $d \geq 1$. For an undirected dependency graph on n vertices, where the maximum degree and its upper bound d may exceed those of a corresponding dependency digraph, Vaccaro recently communicated to us an unpublished proof that if $epd \leq 1$, then the probability that none of the events occurs is at least $e^{-n/d}$ whenever $d \geq 1$ (personal communication, June 22, 2026).

4. Combinatorial applications of the LLL

This section presents several classical applications of the symmetric and asymmetric LLL in probabilistic combinatorics and graph theory. These applications include bounds for the problem of edge-disjoint paths (Section 4.1), the satisfiability problem for Boolean formulas in conjunctive normal form (Section 4.2), lower bounds on diagonal and off-diagonal Ramsey numbers (Sections 4.3 and 4.4, respectively), hypergraph coloring results (Section 4.5), the existence of directed cycles with prescribed modular length (Section 4.6), and an upper bound on the acyclic chromatic number of graphs (Section 4.7). These applications highlight the versatility of the LLL as a tool for probabilistic existence proofs, and strengthened results and additional observations are incorporated throughout this section.

4.1. Edge-disjoint paths

Assume that in a communication network, n pairs of users need to communicate via edge-disjoint paths, and that for each pair there are at least a given number of candidate paths. Using the symmetric version of the LLL, one can show that if, for every two distinct pairs of users, each candidate path of one pair intersects (in an edge) only a limited number of candidate paths of the other pair, then there exists a choice of edge-disjoint paths connecting all n pairs. The following result makes this statement precise and provides a strengthened version of [3, Theorem 6.12].

Theorem 4.1. *(Edge-disjoint paths)* Let $G = (V, E)$ be a graph, and let $\{\{x_i, y_i\}\}_{i=1}^n$ be different unordered pairs of distinct vertices. For each $i \in [n]$, let $\mathcal{Q}_i$ be a set of paths connecting x_i and y_i, where $|\mathcal{Q}_i| \geq m$ for some fixed $m \in \mathbb{N}$. Assume that for all distinct $i, j \in [n]$, each path in $\mathcal{Q}_i$ shares an edge with at most k paths in $\mathcal{Q}_j$. If

$$\begin{cases} k < m, & n = 2, \\ e \cdot \frac{k}{m} \cdot (2n - 4) \leq 1, & n \geq 3, \end{cases} \tag{4.1}$$

then there exist paths $P_i \in \mathcal{Q}_i$, for all $i \in [n]$, such that $P_1, \ldots, P_n$ are pairwise edge-disjoint.

Proof. For each $i \in [n]$, let the path P_i be chosen independently and uniformly at random from $\mathcal{Q}_i$. Consider the bad events $A_{i,j} = \{P_i \text{ and } P_j \text{ share an edge}\}$, for all pairs (i, j) such that $1 \leq i < j \leq n$.

Step 1: Bounding the probability of a bad event. Fix $i, j \in [n]$ with $i < j$. Once the path P_i is chosen, it shares an edge with at most k paths in $\mathcal{Q}_j$. Since P_j is chosen independently and uniformly at random from $\mathcal{Q}_j$, where $|\mathcal{Q}_j| \geq m$, it follows that

$$\mathbb{P}(A_{i,j}) \leq \frac{k}{m}. \tag{4.2}$$

For $n = 2$, there is only one bad event, namely that the two selected paths share an edge. Since $\mathbb{P}(A_{1,2}) \leq \frac{k}{m}$, the sharp sufficient condition for the existence of a choice of two edge-disjoint paths in this case is $k < m$. Conversely, if $k = m$, then the conclusion may fail; for example, this occurs if every path in $\mathcal{Q}_1$ intersects every path in $\mathcal{Q}_2$.

We next consider the case where $n \geq 3$.

Step 2: The dependency structure. Each event $A_{i,j}$ depends only on the random choices P_i and P_j. Since the random paths $P_1, \ldots, P_n$ are mutually independent, it follows that $A_{i,j}$ is independent of the σ-algebra generated by all the remaining bad events, except possibly for those events $A_{r,s}$ for which $\{r, s\} \cap \{i, j\} \neq \emptyset$. These excluded events are precisely those of the form $A_{i,\ell}$ or $A_{\ell,i}$, and $A_{j,\ell}$ or $A_{\ell,j}$, where $\ell \in [n] \setminus \{i, j\}$. Other than $A_{i,j}$, there are exactly $n - 2$ bad events involving the index i, and similarly $n - 2$ bad events involving the index j. Hence, every bad event $A_{i,j}$ is independent of the σ-algebra generated by all but at most $2n - 4$ of the remaining bad events.

Step 3: Applying the symmetric LLL. For $n \geq 3$, we apply Corollary 3.7 with

$$p = \frac{k}{m}, \qquad d = 2n - 4. \tag{4.3}$$

If $epd \leq 1$, then with positive probability no bad event $A_{i,j}$ (for $1 \leq i < j \leq n$) occurs. This condition coincides with (4.1), which holds by assumption. Therefore, there exist pairwise edge-disjoint paths $P_i \in \mathcal{Q}_i$ for all $i \in [n]$. □

4.2. Satisfibability problems

The k-satisfibability (k-SAT) problem is a fundamental object in probabilistic combinatorics, concerned with the satisfiability of Boolean formulas in conjunctive normal form, where a formula is expressed as an AND of clauses, each clause being an OR of literals (variables or their negations), and each clause contains exactly k literals. In this setting, each clause can be associated with a bad event, namely that the clause is not satisfied. By endowing the space of assignments with a suitable probability measure, one obtains a collection of (typically dependent) bad events, making k-SAT a natural framework for applications of the LLL. The lemma then provides sufficient conditions under limited dependency to guarantee the existence of an assignment that avoids all such events, and hence satisfies the formula. The following result strengthens a result appearing in [3, Theorem 6.13], [6, Theorem 3.1] and [9, Lemma 19.8].

Theorem 4.2. *(Satisfiability criterion for a k-SAT formula)* A k-SAT formula is satisfiable if no variable appears in more than $\frac{2^k}{ek}$ clauses. Furthermore, for such a k-SAT formula with n clauses, a uniformly random assignment satisfies the formula with probability at least $\exp(-en2^{-k})$.

Proof. Consider a k-SAT formula in which each variable appears in at most $\frac{2^k}{ek}$ clauses, and choose a random assignment by setting each variable independently to 0 or 1, each with probability $\frac{1}{2}$. For each clause C, let A_C be the bad event that C is not satisfied. Since C contains k literals, all of which must evaluate to 0 for A_C to occur, we have

$$p := \mathbb{P}(A_C) = 2^{-k}. \tag{4.4}$$

Each event A_C is independent of the σ-algebra generated by all the remaining bad events, except for those events whose clause shares a variable with C. Each of the k variables appearing in C appears in at

most $\frac{2^k}{ek}$ clauses. Hence, excluding C, the number of clauses sharing at least one variable with C is at most

$$d \le k \cdot \frac{2^k}{ek} - 1 = \frac{2^k}{e} - 1. \tag{4.5}$$

Therefore, A_C is independent of the σ-algebra generated by all but at most d of the remaining bad events. By Theorem 3.1, since $ep(d+1) \le 1$, it follows that with positive probability no bad event occurs. This implies that there exists an assignment that satisfies all clauses.

Finally, we apply the asymmetric LLL (Theorem 1.6) to derive the claimed lower bound on the probability that a uniformly random assignment satisfies a k-SAT formula with n clauses. Let $x := 1-e^{-a}$ with $a := ep = e2^{-k}$. Consider the dependency graph whose vertex set consists of all bad events, and where any two vertices are adjacent if and only if their corresponding clauses share at least one variable. Then, the degree of every vertex is at most $d \le \frac{2^k}{e} - 1 = \frac{1}{a} - 1$. Consequently,

$$p = 2^{-k} = \frac{a}{e} \le \frac{e^a - 1}{e} = (1 - e^{-a})\, e^{-1+a} = x(1-x)^{\frac{1}{a}-1} \le x(1-x)^d. \tag{4.6}$$

Thus the hypotheses of the LLL are satisfied, and therefore

$$\mathbb{P}\left(\bigcap_C \overline{A_C}\right) \ge (1-x)^n = \exp(-en2^{-k}). \tag{4.7}$$

This completes the proof. □

4.3. Lower bounds on the diagonal Ramsey numbers

A fundamental principle of combinatorics is that complete disorder cannot persist in sufficiently large systems. Ramsey theory provides a precise formulation of this phenomenon, and Ramsey numbers measure the threshold beyond which prescribed structures must inevitably appear (see, e.g., [16]).

Definition 4.3. (Ramsey numbers) Let $k, \ell \ge 2$ be integers. The *Ramsey number* $R(k,\ell)$ is the smallest integer $n \ge 2$ such that every 2-coloring of the edges of the complete graph K_n contains either a monochromatic copy of K_k in the first color or a monochromatic copy of K_ℓ in the second color.

- The *diagonal Ramsey numbers* are the numbers $R(k,k)$; equivalently, $R(k,k)$ is the smallest integer $n \ge 2$ such that every graph on n vertices contains either a clique or an independent set on k vertices.
- The *off-diagonal Ramsey numbers* are all the remaining numbers $R(k,\ell)$ with $k \ne \ell$.

This subsection revisits the lower bound on the diagonal Ramsey numbers derived in [14], relying on the symmetric version of the LLL (Theorem 3.1). Theorem 4.4 restates [14, Theorem 2]. A slight sharpening, based on the same argument, is presented in Theorem 4.5, and the identical asymptotic behavior of the two lower bounds is analyzed in Proposition 4.6, which refines [14, Corollary 1].

Theorem 4.4. Let $k \ge 2$ be an integer. If

$$e\binom{k}{2}\binom{n-2}{k-2}2^{1-\binom{k}{2}} \le 1, \tag{4.8}$$

for some integer $n \ge 2$, then $R(k,k) > n$.

Proof. Color the edges of the complete graph K_n using two colors, uniformly at random and independently. For a set S of k vertices, let A_S be the event that all edges with both endpoints in S are monochromatic. Then, $\mathbb{P}(A_S) = 2^{1-\binom{k}{2}}$. Moreover, A_S is independent of the σ-algebra generated by the events

$$\{A_T : T \subseteq [n],\ |T| = k,\ |T \cap S| \leq 1\},$$

since, whenever $|T \cap S| \leq 1$, the events A_S and A_T depend on disjoint sets of independent edge-color variables. Thus, the only events A_T, with $T \neq S$, that may depend on A_S are those for which $|T \cap S| \geq 2$. Their number is at most $\binom{k}{2}\binom{n-2}{k-2} - 1$. Indeed, for each pair $\{i, j\} \subseteq S$, let

$$F_{\{i,j\}} := \{T \subseteq [n] :\ |T| = k,\ \{i, j\} \subseteq T\}.$$

Then,

$$|F_{\{i,j\}}| = \binom{n-2}{k-2}.$$

Clearly,

$$\{T \subseteq [n],\ |T| = k,\ |T \cap S| \geq 2,\ T \neq S\} = \left(\bigcup_{\{i,j\} \subseteq S} F_{\{i,j\}}\right) \setminus \{S\},$$

so, by the subadditivity of cardinality,

$$\begin{aligned} d := \left|\{T \subseteq [n],\ |T| = k,\ |T \cap S| \geq 2,\ T \neq S\}\right| &\leq \sum_{\{i,j\} \subseteq S} |F_{\{i,j\}}| - 1 \\ &= \binom{k}{2}\binom{n-2}{k-2} - 1. \end{aligned}$$

Finally, by Theorem 3.1 with

$$p = 2^{1-\binom{k}{2}}, \quad d' = \binom{k}{2}\binom{n-2}{k-2} - 1, \tag{4.9}$$

and $d \leq d'$, it follows that if $ep(d' + 1) \leq 1$, or equivalently, if (4.8) holds, then

$$\mathbb{P}\left(\bigcap_{S \subseteq [n]:\ |S| = k} \overline{A_S}\right) > 0. \tag{4.10}$$

In other words, with positive probability, the random coloring of K_n does not contain any monochromatic copy of K_k. Hence, $R(k, k) > n$, which proves Theorem 4.4. □

A variation of Theorem 4.4, which gives a slightly better lower bound on the diagonal Ramsey numbers is given as follows.

Theorem 4.5. Let $k \geq 2$ be an integer. If, for some integer $n \geq 2$,

$$e\left[\binom{n}{k} - \binom{n-k}{k} - k\binom{n-k}{k-1}\right] 2^{1-\binom{k}{2}} \leq 1, \tag{4.11}$$

then $R(k, k) > n$.

Proof. The proof follows the same argument as that of Theorem 4.4, except that we evaluate the exact value of d instead of using the upper bound d' in (4.9).

For an arbitrary k-subset $S \subseteq [n]$, d is the (fixed) number of k-subsets $T \subseteq [n]$, distinct from S, that satisfy $|T \cap S| \geq 2$. Since there are $\binom{n}{k}$ k-subsets of $[n]$, of which $\binom{n-k}{k}$ are disjoint from S and $k\binom{n-k}{k-1}$ intersect S in exactly one element, it follows that

$$d = \binom{n}{k} - \binom{n-k}{k} - k\binom{n-k}{k-1} - 1. \tag{4.12}$$

Thus, applying Theorem 3.1 with $p = 2^{1-\binom{k}{2}}$ (see (4.9)) and d given by (4.12), we conclude that if $ep\,(d+1) \leq 1$, or equivalently, if (4.11) is satisfied, then (4.10) holds. Therefore, with positive probability, the random coloring of K_n contains no monochromatic copy of K_k, and so $R(k,k) > n$. □

Numerical experiments comparing the lower bounds on the diagonal Ramsey numbers $R(k,k)$ given in Theorems 4.4 and 4.5 suggest that the fractional improvement provided by Theorem 4.5 decreases with k, and tends to zero as $k \to \infty$.

The next result, which follows from Theorem 4.4, refines the asymptotic lower bound on the diagonal Ramsey numbers in [14, Corollary 1].

Proposition 4.6. For every $\varepsilon \in (0,1)$, there exists $k_0 = k_0(\varepsilon) \in \mathbb{N}$ such that

$$R(k,k) > (1-\varepsilon)\left(\frac{\sqrt{2}}{e}\right) k 2^{\frac{k}{2}}, \quad \forall\, k \geq k_0. \tag{4.13}$$

Remark 4.7. An explicit closed-form expression for a valid $k_0 = k_0(\varepsilon)$ is derived in the proof of Proposition 4.6 (see (4.17)), and its behavior is then analyzed as $\varepsilon \to 0^+$ (see (4.19a)).

Proof. Let $c_\varepsilon := (1-\varepsilon)\,\frac{\sqrt{2}}{e}$. We show that there exists $k_0 = k_0(\varepsilon) \in \mathbb{N}$ such that

$$R(k,k) > c_\varepsilon\, k\, 2^{k/2}, \qquad \forall\, k \geq k_0.$$

Fix $n := \lfloor c_\varepsilon\, k\, 2^{k/2} \rfloor$. By Theorem 4.4, it suffices to show that inequality (4.8) holds for all sufficiently large k. Since $n \leq c_\varepsilon k 2^{k/2}$ and $\binom{n-2}{k-2} \leq \frac{n^{k-2}}{(k-2)!}$, we get

$$e\binom{k}{2}\binom{n-2}{k-2} 2^{1-\binom{k}{2}} \leq e\binom{k}{2}\frac{(c_\varepsilon k)^{k-2}}{(k-2)!}\, 2^{1-k/2}, \tag{4.14}$$

By Stirling's inequality, asserting that $n! \geq \sqrt{2\pi n}\,(n/e)^n$ for all $n \in \mathbb{N}$, together with the inequality $\left(\frac{k}{k-2}\right)^{k-2} \leq e^2$ that holds for all $k \geq 3$, it follows from (4.14) that

$$e\binom{k}{2}\binom{n-2}{k-2} 2^{1-\binom{k}{2}} \leq \sqrt{\frac{3}{8\pi}}\, e^3\, k^{3/2}\,(1-\varepsilon)^{k-2}, \quad \forall\, k \geq 3. \tag{4.15}$$

Since $\varepsilon \in (0,1)$, the right-hand side of (4.15) tends to zero as $k \to \infty$. Consequently, there exists $k_0 = k_0(\varepsilon) \in \mathbb{N}$ such that (4.8) holds for all $k \geq k_0$. By Theorem 4.4, $R(k,k) \geq n+1$ for all $k \geq k_0$, and therefore

$$R(k,k) > c_\varepsilon\, k\, 2^{k/2}, \quad \forall\, k \geq k_0,$$

which proves (4.13) by the expression for c_ε. □

The proof of Proposition 4.6 enables one to get an explicit closed-form expression for a suitable choice of $k_0 = k_0(\varepsilon)$. Let $\varepsilon \in (0,1)$. From (4.15), together with the requirement to satisfy (4.8) for all $k \geq k_0$, it follows that k_0 can be selected to be the smallest integer $k \geq 3$ that satisfies the inequality

$$\sqrt{\frac{3}{8\pi}}\, e^3\, k^{3/2}\, (1-\varepsilon)^{k-2} \leq 1. \tag{4.16}$$

A suitable choice of $k_0 = k_0(\varepsilon)$ is obtained by solving inequality (4.16) subject to the constraint $k \geq 3$, as detailed in Appendix B. This gives

$$k_0 = \max\left\{3, \left\lceil \frac{3W_{-1}(\beta \ln(1-\varepsilon)\,(1-\varepsilon)^{4/3})}{2\ln(1-\varepsilon)} \right\rceil\right\}, \tag{4.17}$$

where $W_{-1}(\cdot)$ denotes the secondary branch of the Lambert W-function, and

$$\beta = \frac{4}{3e^2}\sqrt[3]{\frac{\pi}{3}} \approx 0.183242. \tag{4.18}$$

We have

$$\begin{aligned} &W_{-1}(x) = \ln(-x) - \ln(-\ln(-x)) + o(1) \quad \text{as } x \to 0^-, \\ &\ln(1-x) = -x + O(x^2) \quad \text{as } x \to 0, \end{aligned}$$

which, applied to the right-hand side of (4.17), yields

$$k_0(\varepsilon) \approx \frac{3}{2\varepsilon}\left[\ln\left(\frac{1}{\beta\varepsilon}\right) + \ln\ln\left(\frac{1}{\beta\varepsilon}\right)\right] \tag{4.19a}$$

$$= \Theta\left(\frac{1}{\varepsilon}\,\ln\frac{1}{\varepsilon}\right). \tag{4.19b}$$

We close this subsection by noting that the lower bounds on the diagonal Ramsey numbers $R(k,k)$ given in Theorems 4.4 and 4.5 are well below the state-of-the-art bounds for small values of k. For example, the resulting lower bounds on $R(10,10)$ are 99 and 105, respectively, whereas the current best lower bound is 798, due to [80] (see also [81, Table 1a]). The significance of the LLL-based analysis in this subsection lies instead in Proposition 4.6 and (4.17)–(4.19b), which provide a refinement of the LLL-based analysis underlying the asymptotically best known lower bound on $R(k,k)$ due to [14, Corollary 1].

4.4. Lower bounds on the off-diagonal Ramsey numbers

As a continuation of Section 4.3, we now apply the asymmetric LLL (Theorem 1.6) to derive lower bounds on off-diagonal Ramsey numbers. This extends the use of the symmetric LLL in the preceding subsection to derive lower bounds on diagonal Ramsey numbers. This presentation follows Spencer's analysis in [14, 15].

Theorem 4.8. *(Lower bound on off-diagonal Ramsey numbers)* Let $k, \ell \geq 2$ be distinct integers, and define $a := \binom{k}{2}$ and $b := \binom{\ell}{2}$. If, for some integer $n \geq 2$, there exist numbers $p, x, y \in (0,1)$ such that

$$p^a \leq x\,(1-x)^{\varepsilon}\,(1-y)^{\vartheta}, \tag{4.20}$$

$$(1-p)^b \le y\,(1-x)^{\varphi}\,(1-y)^{\rho}, \tag{4.21}$$

where

$$\varepsilon := \binom{k}{2}\binom{n-2}{k-2}, \quad \vartheta := \binom{k}{2}\binom{n-2}{\ell-2}, \quad \varphi := \binom{\ell}{2}\binom{n-2}{k-2}, \quad \rho := \binom{\ell}{2}\binom{n-2}{\ell-2}. \tag{4.22}$$

Then, $R(k,\ell) > n$.

Proof. Color the edges of the complete graph K_n independently at random, assigning each edge the color blue with probability p and the color red with probability $1-p$. For a set $S \subseteq [n]$ with $|S| = k$, let A_S be the event that all edges with both endpoints in S are colored blue; likewise, for a set $T \subseteq [n]$ with $|T| = \ell$, let B_T be the event that all edges with both endpoints in T are colored red. Then, $\mathbb{P}(A_S) = p^a$ and $\mathbb{P}(B_T) = (1-p)^b$. Any two of these events are independent whenever their corresponding sets share at most one element (i.e., their corresponding cliques share no edge).

Let $x, y \in (0,1)$, and set $x_{A_S} = x$ and $x_{B_T} = y$ for all such subsets $S, T \in [n]$ with $|S| = k$ and $|T| = \ell$. Consider an associated dependency graph, whose vertex set is the above family of events, and where any two of them are adjacent if and only if they intersect in at least two elements.

(1) The total number of edges linking A_S with any of the events in $\{A_{S'}\}$, where $S' \in [n]$, $|S'| = k$, and $|S \cap S'| \ge 2$, is at most ε. Likewise, the total number of edges linking A_S with any of the events in $\{B_{T'}\}$, where $T' \in [n]$, $|S \cap T'| \ge 2$, and $|T'| = \ell$, is at most ϑ.

(2) The total number of edges linking B_T with any of the events in $\{A_{S'}\}$, where $S' \in [n]$, $|S'| = k$, and $|T \cap S'| \ge 2$, is at most φ. Likewise, the total number of edges linking B_T with any of the events in $\{B_{T'}\}$, where $T' \in [n]$, $|T \cap T'| \ge 2$, and $|T'| = \ell$, is at most ρ.

By the LLL (Theorem 1.6), it follows that if the conditions in (4.20) and (4.21) hold, then

$$\mathbb{P}\left(\bigcap_{S\subseteq[n]:\,|S|=k} \overline{A_S} \;\cap\; \bigcap_{T\subseteq[n]:\,|T|=\ell} \overline{B_T}\right) > 0. \tag{4.23}$$

This implies that, with positive probability, the complete graph K_n contains neither a blue copy of K_k nor a red copy of K_ℓ, which implies that $R(k,\ell) > n$. □

In analogy with the tightened lower bound for diagonal Ramsey numbers in Theorem 4.5, we next present a tightened version of Theorem 4.8. As in the proof of Theorem 4.5, the improvement is obtained by replacing the upper bounds on the numbers of edges in each of the four cases in the same dependency graph with their exact values.

Theorem 4.9. *(Lower bound on off-diagonal Ramsey numbers)* Let $k, \ell \ge 2$ be distinct integers, and define $a := \binom{k}{2}$ and $b := \binom{\ell}{2}$. If, for some integer $n \ge 2$, there exist numbers $p, x, y \in (0,1)$ such that

$$p^a \le x\,(1-x)^{\varepsilon'}\,(1-y)^{\vartheta'}, \tag{4.24}$$

$$(1-p)^b \le y\,(1-x)^{\varphi'}\,(1-y)^{\rho'}, \tag{4.25}$$

where

$$\varepsilon' := \binom{n}{k} - \binom{n-k}{k} - k\binom{n-k}{k-1} - 1, \quad \vartheta' := \binom{n}{\ell} - \binom{n-k}{\ell} - k\binom{n-k}{\ell-1}, \tag{4.26}$$

$$\varphi' := \binom{n}{k} - \binom{n-\ell}{k} - \ell\binom{n-\ell}{k-1}, \qquad \rho' := \binom{n}{\ell} - \binom{n-\ell}{\ell} - \ell\binom{n-\ell}{\ell-1} - 1. \tag{4.27}$$

Then, $R(k,\ell) > n$.

Remark 4.10. Following the proof of Theorem 4.8, the subtraction of 1 in the definitions of ε' and ρ' in Eqs (4.26) and (4.27), respectively, accounts for the fact that, in the corresponding same-type counts, the event itself must be excluded. In contrast, no subtraction of 1 appears in the mixed counts corresponding to ϑ' and φ', since the events A_S and B_T are distinct for all sets S and T with $|S| = k$ and $|T| = \ell$ (even when $S = T$ and $k = \ell$).

The next asymptotic lower bound on off-diagonal Ramsey numbers is a consequence of Theorem 4.8 (see [15, Theorem 2.2] for a proof).

Theorem 4.11. *(Asymptotics)* For every fixed integer $k \geq 3$, there exists a constant $c_k > 0$ such that

$$R(k,\ell) \geq c_k \left(\frac{\ell}{\log \ell}\right)^{\frac{k+1}{2}} \tag{4.28}$$

for all sufficiently large ℓ.

We close this subsection by noting that several recent works (see [82–85]) derive improved lower bounds on off-diagonal Ramsey numbers using probabilistic techniques that do not rely on the LLL.

4.5. Coloring hypergraphs

Using the LLL, the seminal work [1] established the existence of hypergraphs with a prescribed chromatic number. Their approach highlights the power of probabilistic methods in demonstrating the existence of sparse hypergraphs with nontrivial coloring properties.

This subsection concerns the coloring of hypergraphs and is based in part on [1] and [2, Section 5.2], together with some reformulated proofs and strengthened results on hypergraph colorings (Theorems 4.17, 4.18, 4.20, and Remark 4.21).

Definition 4.12. A *hypergraph* $H = (V, E)$ consists of a set of vertices V and a collection E of subsets of V. Each element $h \in E$ is called a *hyperedge*. A hypergraph $H = (V, E)$ is *k-colorable* if there is a k-coloring of V such that no hyperedge is monochromatic.

Theorem 4.13. Let $H = (V, E)$ be a hypergraph in which every hyperedge has size at least r, and each hyperedge intersects at most d other hyperedges. If $ed \leq k^{r-1}$, with $k \geq 2$, then H is k-colorable.

Proof. Assign to each vertex v of H a random color chosen independently and uniformly from $[k]$. For each hyperedge $h \in E$, let A_h be the bad event that h is monochromatic. By assumption $|h| \geq r$, which implies that

$$\mathbb{P}(A_h) = k\left(\frac{1}{k}\right)^{|h|} \leq k^{1-r} \quad \text{for all } \ h \in E. \tag{4.29}$$

Moreover, due to the independent coloring of the vertices, A_h is independent of the σ-algebra generated by all bad events $A_{h'}$ such that $h \cap h' = \varnothing$. Since h intersects at most d other hyperedges, it follows

that A_h is independent of the σ-algebra generated by all other bad events except for at most d of them. Hence, there exists a dependency graph for the family of events $\{A_h\}_{h\in E}$ whose maximum degree is at most d. Applying Corollary 3.7 with $p = k^{1-r}$, it follows that if $edk^{1-r} \leq 1$, then

$$\mathbb{P}\left(\bigcap_{h\in E} \overline{A_h}\right) > 0. \tag{4.30}$$

This shows that if $ed \leq k^{r-1}$, then H admits a k-coloring in which no hyperedge is monochromatic. By definition, H is k-colorable. □

Definition 4.14. A hypergraph H is *r-uniform* if each hyperedge has size r, and it is *b-regular* if every vertex in V is contained in exactly b hyperedges of H.

Corollary 4.15. Let H be an r-uniform and b-regular hypergraph with $b, r \geq 2$, and let $k \geq 2$. If $er(b-1) \leq k^{r-1}$, then H is k-colorable.

Proof. Each hyperedge $h \in E$ contains exactly r vertices, and each of these vertices belongs to exactly b hyperedges. Therefore, for every $h \in E$, the number of hyperedges intersecting h is at most $r(b-1)$. Hence, Theorem 4.13 applies with $d = r(b-1)$, yielding the desired result. □

Corollary 4.16. Let H be an r-uniform and r-regular hypergraph. Then, H is 2-colorable if $r \geq 9$, and 3-colorable if $r \geq 5$, 4-colorable if $r \geq 4$, 5-colorable if $r \geq 3$, and 6-colorable if $r \geq 2$.

Proof. By Corollary 4.15, applied with $r = b$, it follows that if $er(r-1) \leq k^{r-1}$, then H is k-colorable. This inequality holds if and only if $r \geq 9, 5, 4, 3$, and 2 for $k = 2, 3, 4, 5$, and 6, respectively. □

The next two results restate [1, Theorems 2 and 3], respectively, and present reformulated proofs that yield slightly stronger conclusions, based on Corollary 3.7 and on replacing the constant 4 in [1] by e. Moreover, these proofs are somewhat simpler, as they do not rely on the line graph of H, but instead work directly with H.

Theorem 4.17. Let $k, r \geq 2$, and let H be an r-uniform hypergraph. If every hyperedge of H intersects at most $\frac{1}{e}\, k^{r-1}$ other hyperedges, then H is k-colorable. In particular, if every vertex of H has degree at most $\frac{k^{r-1}}{er}$, then H is k-colorable.

Proof. The first assertion is a direct restatement of Theorem 4.13. Since every hyperedge of H intersects at most $\frac{1}{e}\, k^{r-1}$ other hyperedges, it follows that H is k-colorable.

The second assertion follows immediately from the first. Indeed, every hyperedge contains r vertices, each of degree at most $\frac{k^{r-1}}{er}$. Therefore, the number of hyperedges intersecting a given hyperedge is at most $r \cdot \frac{k^{r-1}}{er} = \frac{1}{e}\, k^{r-1}$, where we rely on the subadditivity of cardinality. Applying the first assertion completes the proof. □

Theorem 4.18. Let $r \geq k \geq 2$, and let H be an r-uniform hypergraph. If every hyperedge of H intersects at most $\left\lfloor \frac{k^{r-1}}{e(k-1)^r} \right\rfloor$ other hyperedges, then H admits a k-coloring in which each hyperedge contains all colors.

Proof. As in the proof of Theorem 4.13, assign to each vertex of H a color chosen independently and uniformly at random from $[k]$. For each hyperedge $h \in E$, let A_h denote the bad event that h does not contain all k colors. Since the event A_h is the union of the k events that a particular color is missing from h, it follows by the union bound that

$$\mathbb{P}(A_h) \leq k\left(1 - \frac{1}{k}\right)^r := p \quad \text{for all} \ \ h \in E. \tag{4.31}$$

Since A_h depends only on the colors assigned to the vertices of h, and the vertices of H are colored independently, A_h is independent of the σ-algebra generated by all bad events $A_{h'}$ such that $h \cap h' = \emptyset$. Let

$$d := \left\lfloor \frac{k^{r-1}}{e(k-1)^r} \right\rfloor. \tag{4.32}$$

Since h intersects at most d other hyperedges, it follows that A_h is independent of the σ-algebra generated by all other bad events except for at most d of them. Hence, there exists a dependency graph for the family of events $\{A_h\}_{h \in E}$ with maximum degree at most d. This implies that

$$epd \leq ek\left(1 - \frac{1}{k}\right)^r \cdot \frac{k^{r-1}}{e(k-1)^r} = 1. \tag{4.33}$$

By Corollary 3.7, with positive probability, no bad event A_h ($h \in E$) occurs. Therefore, there exists a k-coloring of the vertices of H in which every hyperedge contains all k colors. □

In the following, we propose a stronger version of Theorem 4.18. To derive an exact closed-form expression for the probability $\mathbb{P}(A_h)$ in place of the upper bound on the right-hand side of (4.31), we recall the definition of the Stirling numbers of the second kind.

Definition 4.19. Let $n, k \in \mathbb{N}$. The *Stirling number of the second kind*, denoted by $S(n, k)$, is the number of ways to partition the set $[n]$ into k nonempty and pairwise disjoint subsets. If $k > n$, then $S(n, k) = 0$.

In the proof of Theorem 4.18, the r vertices of a hyperedge $h \in E$ are independently assigned colors chosen uniformly at random from $[k]$. Hence, all k^r colorings of h are equally likely. By Definition 4.19, the number of colorings of h that use all k colors is $k!\, S(r, k)$. Therefore, the probability of the bad event A_h, namely, that h does not contain all k colors, is

$$\mathbb{P}(A_h) = 1 - \frac{k!\, S(r, k)}{k^r}. \tag{4.34}$$

By [86, Eq (6.19)], which gives a closed-form expression for $S(r, k)$, it follows that if $r \geq k \geq 2$, then

$$\mathbb{P}(A_h) = 1 - \frac{1}{k^r} \sum_{j=0}^{k} \left\{ (-1)^{k-j} \binom{k}{j} j^r \right\}, \quad h \in E. \tag{4.35}$$

This allows one to sharpen the result of Theorem 4.18 through an application of Corollary 3.7, replacing the upper bound p on $\mathbb{P}(A_h)$ in (4.31) with the exact value given by (4.35). Consequently, the following result holds:

Theorem 4.20. Let $r \geq k \geq 2$, and let H be an r-uniform hypergraph. If every hyperedge of H intersects at most

$$\frac{1}{e\left(1 - \frac{1}{k^r}\sum_{j=0}^{k}\left\{(-1)^{k-j}\binom{k}{j}j^r\right\}\right)}$$

other hyperedges, then H admits a k-coloring in which each hyperedge contains all colors.

The following remark shows that the improvement afforded by the sharpened result in Theorem 4.20 over Theorem 4.18 allows one to increase the number of hyperedges intersecting any given hyperedge by at most 1.

Remark 4.21. Let $r \geq k \geq 2$, and define

$$D_{r,k} := k\left(1 - \frac{1}{k}\right)^r, \quad N_{r,k} := 1 - \frac{k!\,S(r,k)}{k^r}. \tag{4.36}$$

Furthermore, set

$$A_{r,k} := \left\lfloor \frac{1}{eD_{r,k}} \right\rfloor, \quad B_{r,k} := \left\lfloor \frac{1}{eN_{r,k}} \right\rfloor. \tag{4.37}$$

Then, $A_{r,k}$ and $B_{r,k}$ represent, respectively, the largest number of hyperedges that may intersect a given hyperedge in an r-uniform hypergraph such that Theorem 4.18 or Theorem 4.20 guarantees the existence of a k-coloring in which every hyperedge contains all k colors.

The quantity $D_{r,k}$ is the expected number of empty boxes when r balls are thrown independently and uniformly into k boxes, and $N_{r,k}$ is the probability that at least one box is empty. Indeed, $D_{r,k}$ is the sum of the probabilities that the individual boxes are empty, and therefore, by the union bound, $N_{r,k} \leq D_{r,k}$. Consequently, $B_{r,k} \geq A_{r,k}$, and we next show that their difference is at most one. Let X denote the number of empty boxes, i.e.,

$$X = \sum_{i=1}^{k} I_i, \quad I_i := \mathbf{1}_{\{\text{box } i \text{ is empty}\}} \quad \text{for all } \ i \in [k],$$

where $\mathbf{1}_E$ denotes the indicator function of an event E. Then,

$$\mathbb{E}[I_i] = \left(1 - \frac{1}{k}\right)^r, \quad \mathbb{E}[X] = D_{r,k}, \quad \mathbb{P}(X \geq 1) = N_{r,k},$$
$$\mathbb{E}[I_i I_j] = \left(1 - \frac{2}{k}\right)^r, \quad \text{for all } \ i \neq j,$$

so

$$\mathrm{Cov}(I_i, I_j) = \left(1 - \frac{2}{k}\right)^r - \left(1 - \frac{1}{k}\right)^{2r} < 0,$$

thus the indicators $\{I_i\}$ are negatively correlated. Therefore,

$$\mathrm{Var}(X) = \sum_{i=1}^{k} \mathrm{Var}(I_i) + \sum_{i \neq j} \mathrm{Cov}(I_i, I_j) \leq \sum_{i=1}^{k} \mathrm{Var}(I_i).$$

Since each I_i is a Bernoulli random variable with mean $p = \left(1 - \frac{1}{k}\right)^r$, it follows that

$$\mathrm{Var}(I_i) = p(1-p) \le p,$$
$$\mathrm{Var}(X) \le kp = D_{r,k}.$$

By the Cauchy–Schwarz inequality, and since X is a nonnegative integer-valued random variable,

$$\begin{aligned}(\mathbb{E}[X])^2 &= (\mathbb{E}[X\mathbf{1}_{\{X\ge 1\}}])^2 \\ &\le \mathbb{E}[X^2]\,\mathbb{E}[\mathbf{1}_{\{X\ge 1\}}] \\ &= \mathbb{E}[X^2]\,\mathbb{P}(X \ge 1),\end{aligned} \tag{4.38}$$

and therefore

$$N_{r,k} = \mathbb{P}(X \ge 1) \ge \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}. \tag{4.39}$$

Since

$$\mathbb{E}[X^2] = \mathrm{Var}(X) + (\mathbb{E}[X])^2 \le D_{r,k} + D_{r,k}^2,$$

we get

$$N_{r,k} \ge \frac{D_{r,k}^2}{D_{r,k} + D_{r,k}^2} = \frac{D_{r,k}}{1 + D_{r,k}}.$$

Consequently,

$$0 \le \frac{1}{eN_{r,k}} - \frac{1}{eD_{r,k}} \le \frac{1}{e} < 1,$$

and therefore, by (4.37),

$$B_{r,k} - A_{r,k} \in \{0, 1\}. \tag{4.40}$$

The equality $B_{r,k} = A_{r,k} + 1$ can hold for some pairs (r, k) with $r \ge k \ge 2$, and Table 1 provides several such numerical examples.

Table 1. Examples of pairs (r, k) with $r \ge k \ge 2$ for which the equality $B_{r,k} = A_{r,k} + 1$ holds.

(r, k)	$A_{r,k}$	$B_{r,k}$
(21, 5)	7	8
(22, 5)	9	10
(19, 6)	1	2
(28, 7)	3	4
(35, 8)	4	5
(41, 8)	10	11
(48, 8)	27	28

4.6. Length of cycles in directed graphs

The following result, due to Alon and Linial [17], provides an insightful application of the symmetric version of the LLL (Theorem 3.1). Unlike the three preceding applications in Sections 4.3–4.5, the next theorem is not directly concerned with graph coloring, although its proof relies on coloring arguments.

Theorem 4.22. *([17])* Let $D = (V, E)$ be a finite, simple, and directed graph with minimum out-degree $\delta \geq 1$ and maximum in-degree at most $\Delta \geq 1$. If, for some $k \in \mathbb{N}$,

$$e(\delta\Delta + 1)\left(1 - \frac{1}{k}\right)^{\delta} \leq 1, \tag{4.41}$$

then D contains a directed cycle whose length is divisible by k.

Remark 4.23. By Corollary 3.7 and the proof of Theorem 4.22, the condition in (4.41) can be relaxed to

$$e\delta\Delta\left(1 - \frac{1}{k}\right)^{\delta} \leq 1. \tag{4.42}$$

For completeness, we present a more detailed version of the proof appearing in [2, p. 88] and [17, Theorem 2.3].

Proof. We may assume that all out-degrees are equal to δ. Indeed, by deleting arcs if necessary, we obtain a spanning subgraph of D in which every vertex has out-degree δ, and the maximum in-degree remains at most Δ. Any directed cycle in this spanning subgraph is also present in D; in particular, so is any directed cycle whose length is divisible by k.

We use the LLL to show that there exists a vertex coloring $c\colon V \to [k]$ such that every vertex $v \in V$ has an out-neighbor u with $c(u) \equiv c(v) + 1 \pmod{k}$.

We first show that the existence of such a k-coloring implies the theorem. Let $c\colon V \to [k]$ be such a coloring. Since D is finite and every vertex has out-degree at least 1, we may construct a directed walk as follows. Start at some vertex $v_1 \in V$, and for each $i \geq 1$, choose an out-neighbor v_{i+1} of v_i with $c(v_{i+1}) \equiv c(v_i) + 1 \pmod{k}$. Since D is finite, some vertex must repeat. Let j be the smallest index such that $v_j = v_\ell$ for some $\ell < j$. Consider the directed cycle

$$v_\ell \to v_{\ell+1} \to \cdots \to v_j = v_\ell.$$

Then,

$$c(v_j) \equiv c(v_{j-1}) + 1 \equiv \cdots \equiv c(v_\ell) + (j - \ell) \pmod{k},$$

and therefore

$$(j - \ell) \equiv 0 \pmod{k}.$$

Hence, the constructed directed cycle in D has length $j - \ell$, which is divisible by k.

Next, we use the LLL to prove the existence of such a k-coloring. Let $N_+(v)$ and $N_-(v)$ denote the sets of out-neighbors and in-neighbors of a vertex v in a directed graph D, respectively. These are referred to as the *open out-neighborhood* and *open in-neighborhood* of v. Furthermore, define

$$N_+[v] := N_+(v) \cup \{v\}, \qquad N_-[v] := N_-(v) \cup \{v\},$$

which are the *closed out-neighborhood* and *closed in-neighborhood* of v, respectively.

We next show that there exists a coloring $c\colon V \to [k]$ such that every vertex $v \in V$ has an out-neighbor $u \in N_+(v)$ satisfying $c(u) \equiv c(v) + 1 \pmod{k}$. Consider a random coloring $c\colon V \to [k]$, where each vertex of D is assigned a color independently and uniformly at random (each color with probability $\frac{1}{k}$). For each $v \in V$, let A_v denote the event that $c(u) \not\equiv c(v) + 1 \pmod{k}$ for all $u \in N_+(v)$. We show, using Theorem 3.1, that

$$\mathbb{P}\left(\bigcap_{v\in V} \overline{A_v}\right) > 0, \tag{4.43}$$

which gives the desired result. For each $v \in V$, since the colors are assigned independently and uniformly, we have

$$\mathbb{P}(A_v) = \left(1 - \frac{1}{k}\right)^{\delta}, \tag{4.44}$$

where, by assumption, $\delta = |N_+(v)|$ for all $v \in V$. Moreover, for every $v \in V$, if the colors of all vertices outside $N_+(v)$ are fixed, then the conditional probability of A_v remains $\left(1 - \frac{1}{k}\right)^{\delta}$. It therefore follows that A_v is mutually independent of the collection

$$\{A_u : N_+(v) \cap N_+[u] = \varnothing\}. \tag{4.45}$$

The condition in (4.45) fails if and only if one of the following holds:

(1) $u \in N_+(v)$;

(2) $N_+(u) \cap N_+(v) \neq \varnothing$ and $u \neq v$.

The number of vertices u satisfying Condition 1 is δ. Also, for each $z \in N_+(v)$, there are at most $\Delta - 1$ vertices $u \neq v$ such that $u \in N_-(z)$. Since there are δ such vertices z, the number of vertices satisfying Condition 2 is at most $\delta(\Delta - 1)$. Therefore, the number of vertices $u \in V$ for which $N_+(v) \cap N_+[u] \neq \varnothing$ is at most

$$\delta + \delta(\Delta - 1) = \delta\Delta.$$

Consequently, by the symmetric version of the LLL (Theorem 3.1) with parameters

$$p = \left(1 - \frac{1}{k}\right)^{\delta}, \quad d = \delta\Delta,$$

it follows that (4.43) holds provided that (4.41) is satisfied, which completes the proof. □

A *d-regular digraph* is a directed graph in which the in-degree and out-degree of every vertex are equal to d. The next corollary is a general result that includes regular digraphs as a special case, whereas the subsequent corollary is stated exclusively for regular digraphs.

Corollary 4.24. Let $D = (V, E)$ be a finite, simple, and directed graph with minimum out-degree $\delta \geq 1$ and maximum in-degree at most $\Delta \geq 1$. If

$$\Delta \leq \frac{2^{\delta}}{e\delta}, \tag{4.46}$$

then the digraph D contains an even-length directed cycle. In particular, this holds for every d-regular digraph with $d \geq 8$.

Proof. The first part follows from Theorem 4.22 and Remark 4.23 with $k = 2$. The second part holds since, for a d-regular digraph, $\delta = \Delta = d$, and

$$\frac{2^x}{ex^2} \geq 1 \qquad \Longleftrightarrow \qquad x \geq 7.09719\ldots,$$

so the condition in (4.46) holds for every d-regular digraph if and only if $d \geq 8$. □

The next result is an improved version of [2, Corollary 2.5].

Corollary 4.25. For every finite, simple, and d-regular digraph D with $d \geq 1$, and for every integer k satisfying

$$k \leq \frac{\sqrt[d]{ed^2}}{\sqrt[d]{ed^2} - 1}, \tag{4.47}$$

there exists a directed cycle in D whose length is divisible by k.

Proof. For a d-regular digraph, we have $\Delta = \delta = d$. By Theorem 4.22 and Remark 4.23, there exists a directed cycle of length divisible by k if

$$ed^2\left(1 - \frac{1}{k}\right)^d \leq 1.$$

Taking the d-th root of both sides and rearranging terms gives the condition in (4.47). □

Remark 4.26. Since the digraph D is, by assumption, finite and every vertex in D has an out-degree at least 1, it follows that D contains a directed cycle. Corollary 4.25 is therefore nontrivial if and only if the right-hand side of (4.47) is at least 2; this occurs if and only if $d \geq 8$. Moreover, it strengthens [2, Corollary 2.5], which asserts the same conclusion under the stronger condition

$$k \leq \frac{d}{1 + \ln(d^2 + 1)}, \tag{4.48}$$

whose right-hand side is strictly smaller than that of (4.47). For comparison, the right-hand side of Eq (4.48) is at least 2 if and only if $d \geq 12$. It can be verified that the difference between the right-hand sides of (4.47) and (4.48) converges to $\frac{1}{2}$ as $d \to \infty$. Indeed, let $x_d := \frac{1+2\ln d}{d}$, then $\sqrt[d]{ed^2} = e^{x_d}$ and $x_d \to 0$ as we let $d \to \infty$. Hence, the right-hand side of (4.47) is

$$\frac{\sqrt[d]{ed^2}}{\sqrt[d]{ed^2} - 1} = \frac{e^{x_d}}{e^{x_d} - 1} = \frac{1}{1 - e^{-x_d}} = \frac{1}{x_d} + \frac{1}{2} + o(1),$$

and the right-hand side of (4.48) is

$$\frac{d}{1 + \ln(d^2 + 1)} = \frac{d}{1 + 2\ln d + \ln(1 + d^{-2})} = \frac{1}{x_d} + o(1).$$

Subtracting yields

$$\lim_{d\to\infty}\left(\frac{\sqrt[d]{ed^2}}{\sqrt[d]{ed^2} - 1} - \frac{d}{1 + \ln(d^2 + 1)}\right) = \frac{1}{2}. \tag{4.49}$$

As an illustration of the improvement in Corollary 4.25, if $d \in \{17, 18, 19, 20, 21\}$, then (4.47) allows $k \in \{2, 3\}$, whereas (4.48) permits only $k = 2$. Consequently, every finite, simple, and d-regular digraph with $17 \leq d \leq 21$ not only contains a directed cycle of even length, but also contains a directed cycle whose length is divisible by 3.

We conclude this section by quoting a result on directed cycles in regular digraphs from [11, Theorem 2], whose proof relies on the symmetric version of the LLL (Theorem 3.1).

Theorem 4.27. *([11])* Let G be a d-regular digraph without parallel edges. Then, G contains at least εd^2 pairwise edge-disjoint directed cycles for some absolute constant $\varepsilon > 0$. In particular, $\varepsilon = \frac{3}{2^{19}}$ suffices.

4.7. Acyclic coloring of graphs

This last subsection presents a sophisticated application of the LLL (Theorem 1.6), in the context of graph coloring, appearing in a paper by Alon, McDiarmid, and Reed [19].

Let $G = (V, E)$ be a finite, simple, and undirected graph. For $v \in V$, let $N(v) = \{u \in V : \{u, v\} \in E\}$ denote the neighborhood of v, let $d(v) = |N(v)|$ denote its degree, and let

$$\Delta = \Delta(G) := \max_{v \in V} d(v)$$

denote the maximum degree of G. For vertices $u, v \in V$, let $\lambda(u, v) := |N(u) \cap N(v)|$ denote the number of common neighbors of u and v in G. Recall that a k-coloring $f \colon V \to [k]$ is called *proper* if $\{u, v\} \in E$ implies that $f(u) \neq f(v)$ (i.e., no two adjacent vertices are assigned the same color). A coloring of G is called *acyclic* if it is proper and contains no bichromatic cycle (equivalently, every cycle in G receives at least three colors). The minimum number of colors required for an acyclic coloring, denoted by $A(G)$, is called the *acyclic chromatic number* of G. Clearly, $A(G) \geq \chi(G)$, where $\chi(G)$ denotes the chromatic number of G, i.e., the minimum number of colors required for a proper coloring. Equality holds, for example, if G is an odd cycle, a complete graph, or a forest.

Introduced by Grünbaum in the study of planar graphs [87], the acyclic chromatic number has become an important topic in graph coloring and structural graph theory. Borodin settled Grünbaum's conjecture by proving that $A(G) \leq 5$ for every planar graph G [88]. A major reason for the importance of acyclic colorings is their close connection to sparse graph classes. These connections have led to numerous applications in structural and algorithmic graph theory, including graph decompositions, graph homomorphisms, and efficient algorithms for sparse graphs. Acyclic colorings also play a role in the study of graph layouts, graph drawing, and graphs embedded on surfaces (see, e.g., [89]).

In [19], it is proved that there exists an absolute positive constant C such that every graph of maximum degree Δ admits an acyclic coloring with at most $C\Delta^{4/3}$ colors (in [19, Theorem 1.1], $C = 50$, though no attempt was made to optimize that constant). The main tool in proving this result in [19] is the LLL (Theorem 1.6), and a key innovation in [19] is the introduction of special pairs of vertices and a carefully chosen family of bad events that allow the LLL to yield an upper bound of order $\Delta^{4/3}$. This order of growth in Δ is shown to be nearly optimal. More concretely, letting $B(\Delta) := \sup\{A(G) : \Delta(G) = \Delta\}$, for $\Delta \in \mathbb{N}$, [19, Theorem 1.2] establishes the asymptotic lower bound

$$B(\Delta) = \Omega\left(\frac{\Delta^{\frac{4}{3}}}{\sqrt[3]{\log \Delta}}\right). \tag{4.50}$$

Following standard notation, let the complete bipartite graph with vertex classes of sizes a and b be denoted by $K_{a,b}$. The proof of (4.50) relies on a probabilistic construction that gives, almost surely, a graph G with maximum degree Δ, containing no copy of $K_{2,\gamma}$, where $\gamma = O(\Delta^{2/3}\sqrt[3]{\log\Delta})$, and whose acyclic chromatic number $A(G)$ scales as $\Delta^{4/3}(\log\Delta)^{-1/3}$. This contrasts with planar graphs, which may have arbitrarily large maximum degree, yet whose acyclic chromatic number is at most 5.

The next result is derived in [19] via an application of the LLL, and the presentation here follows the main ideas of the proof.

Theorem 4.28. *([19, Theorem 1.1])* For every graph finite, simple, and undirected graph G of maximum degree Δ, its acyclic chromatic number satisfies

$$A(G) \leq \lceil 50\,\Delta^{\frac{4}{3}} \rceil. \tag{4.51}$$

Proof. Let each vertex in G be assigned a color from $[k]$ independently and uniformly at random, where $k := \lceil 50\,\Delta^{4/3} \rceil$. The central idea is to distinguish pairs of nonadjacent vertices having many common neighbors. A pair of nonadjacent vertices, $\{u, v\} \notin E$, is called *a special pair* if $\lambda(u, v) > \Delta^{2/3}$.

Lemma 4.29. For every vertex $u \in V$,

$$\Big|\{v : \{u, v\} \text{ is a special pair in } G\}\Big| < \Delta^{\frac{4}{3}}. \tag{4.52}$$

Proof. The number of all length-2 paths in G is at most $\Delta(\Delta - 1) < \Delta^2$, and every special pair contributes more than $\Delta^{2/3}$ such paths. Let $u \in V$ and let $s = s(u)$ denote the number of special partners of u. Since each special partner of u contributes more than $\Delta^{2/3}$ paths of length 2 having u as an endpoint, we have $s\Delta^{2/3} < \Delta^2$, and therefore $s < \Delta^{4/3}$. □

To apply the LLL, four types of bad events are introduced for such a coloring $f : V \to [k]$.

(1) Monochromatic edges. Let $e = \{u, v\} \in E$. Define $A = A_e := \{f(u) = f(v)\}$, whose probability is $\mathbb{P}(A) = \frac{1}{k}$.

(2) Bichromatic induced paths of length four. Let $u_1 - u_2 - u_3 - u_4 - u_5$ be an induced 4-path. Define $B := \{f(u_1) = f(u_3) = f(u_5),\ f(u_2) = f(u_4)\}$, whose probability is $\mathbb{P}(B) = \frac{1}{k^3}$.

(3) Bichromatic induced cycles of length four. Let $u_1 - u_2 - u_3 - u_4 - u_1$ be an induced 4-cycle such that neither of the opposite pairs, $\{u_1, u_3\}$ or $\{u_2, u_4\}$, is a special pair. Define the event $C := \{f(u_1) = f(u_3),\ f(u_2) = f(u_4)\}$, whose probability is $\mathbb{P}(C) = \frac{1}{k^2}$.

(4) For every special pair $p = \{u, v\}$ (recall that $p \notin E$), define $D = D_p := \{f(u) = f(v)\}$. Then, $\mathbb{P}(D) = \frac{1}{k}$.

The next lemma shows that excluding these four types of bad events guarantees that the coloring is acyclic.

Lemma 4.30. If none of the bad events of types 1–4 occurs, then the coloring is acyclic.

Proof. Since no Type 1 event occurs, the vertex coloring in G is proper.

Suppose, for the sake of contradiction, that there exists a bichromatic cycle, and let C be a shortest such cycle. Since the coloring is proper, C has even length. Hence it suffices to consider the cases

$|C| = 4$ and $|C| \geq 6$. Furthermore, C is induced; otherwise, a chord of C together with a subpath of C would yield a shorter bichromatic cycle, contradicting the choice of C. We next consider the following possibilities and rule out each of them.

(a) Suppose that C is of length at least 6. Take five consecutive vertices $u_1 - u_2 - u_3 - u_4 - u_5$ on the cycle. Since the cycle is bichromatic, $f(u_1) = f(u_3) = f(u_5)$ and $f(u_2) = f(u_4)$. Since C is induced, these vertices form an induced path. Thus a Type-2 event occurs, which leads to a contradiction.

(b) Suppose that C is a cycle of length 4, denoted by $u_1 - u_2 - u_3 - u_4 - u_1$. Since C is bichromatic, we have $f(u_1) = f(u_3)$ and $f(u_2) = f(u_4)$. If one opposite pair is special, then a Type 4 event occurs. Otherwise, neither opposite pair is special and a Type 3 event occurs, leading again to a contradiction.

Hence, no bichromatic cycle exists, so the coloring is acyclic. □

The next lemma appears as [19, Lemma 2.4], and its proof is omitted here.

Lemma 4.31. *([19, Lemma 2.4])* A vertex in V belongs to at most Δ events of type 1, $3\Delta^4$ events of type 2, $\Delta^{8/3}$ events of type 3, and $\Delta^{4/3}$ events of type 4.

Consider a dependency graph whose vertices are the types 1–4 bad events, where two such events are adjacent whenever their corresponding subsets of vertices of G intersect. In light of Lemma 4.31, consider the asymmetric matrix

$$\mathbf{M} = \begin{pmatrix} 2\Delta & 6\Delta^4 & 2\Delta^{8/3} & 2\Delta^{4/3} \\ 5\Delta & 15\Delta^4 & 5\Delta^{8/3} & 5\Delta^{4/3} \\ 4\Delta & 12\Delta^4 & 4\Delta^{8/3} & 4\Delta^{4/3} \\ 2\Delta & 6\Delta^4 & 2\Delta^{8/3} & 2\Delta^{4/3} \end{pmatrix}, \tag{4.53}$$

whose (i, j) entry, with $1 \leq i, j \leq 4$, is an upper bound on the number of type j events that are adjacent to a fixed type i event in the dependency graph (see [19, Lemma 2.5]).

For the purpose of using the LLL, the following values are assigned to the four types of events as above:

$$x_A = x_D = \frac{2}{k}, \quad x_B = \frac{2}{k^3}, \quad x_C = \frac{2}{k^2}. \tag{4.54}$$

Recall that the probabilities of these bad events are given by $\mathbb{P}(A) = \frac{1}{k} = \mathbb{P}(D)$, $\mathbb{P}(B) = \frac{1}{k^3}$, and $\mathbb{P}(C) = \frac{1}{k^2}$. By replacing the exact dependency counts with the upper bounds given by the matrix $\mathbf{M}$ in (4.53), the conditions in (1.2) are implied by the following three inequalities:

$$\frac{1}{k} \leq \frac{2}{k}\left(1 - \frac{2}{k}\right)^{2\Delta + 2\Delta^{4/3}} \left(1 - \frac{2}{k^3}\right)^{6\Delta^4} \left(1 - \frac{2}{k^2}\right)^{2\Delta^{8/3}}, \tag{4.55}$$

$$\frac{1}{k^3} \leq \frac{2}{k^3}\left(1 - \frac{2}{k}\right)^{5\Delta + 5\Delta^{4/3}} \left(1 - \frac{2}{k^3}\right)^{15\Delta^4} \left(1 - \frac{2}{k^2}\right)^{5\Delta^{8/3}}, \tag{4.56}$$

$$\frac{1}{k^2} \leq \frac{2}{k^2}\left(1 - \frac{2}{k}\right)^{4\Delta + 4\Delta^{4/3}} \left(1 - \frac{2}{k^3}\right)^{12\Delta^4} \left(1 - \frac{2}{k^2}\right)^{4\Delta^{8/3}}. \tag{4.57}$$

These inequalities are verified in [19, p. 282] to hold for $k = \lceil 50\Delta^{4/3} \rceil$. By the LLL, with positive probability, none of the types 1–4 bad events occurs. Hence, by Lemma 4.30, there exists an acyclic coloring of G. □

Remark 4.32. (Key idea in the proof of Theorem 4.28) The heart of the argument is the introduction of the threshold $\Delta^{2/3}$ for the number of common neighbors for special pairs. By Lemma 4.29, the number of special pairs incident to a vertex is at most $\Delta^{4/3}$, and every nonspecial pair of nonadjacent vertices has at most $\Delta^{2/3}$ common neighbors by definition, and therefore participates in at most $\Delta^{2/3}$ relevant induced 4-cycles. The equality $\Delta^{2/3} \cdot \Delta^{4/3} = \Delta^2$ is exactly what makes the LLL inequalities compatible with a color set of size that scales as $\Delta^{4/3}$, producing the celebrated bound in (4.51).

The LLL (Theorem 1.6) has another noteworthy application in [19], leading to the following result.

Theorem 4.33. *([19, Theorem 1.3])* Let G be a finite, simple, and undirected graph with maximum degree $\Delta \geq 1$, and suppose that for some integer $\gamma \geq 1$, the graph G contains no copy of $K_{2,\gamma+1}$ in which the two vertices in the class of size 2 are nonadjacent. Then,

$$A(G) \leq \lceil 32\sqrt{\gamma}\,\Delta \rceil. \tag{4.58}$$

The reader is referred to [19, Section 3] for a proof of Theorem 4.33. In its setting, $A(G)$ scales at most linearly with Δ. Hence, if the girth (length of shortest cycle) of G is at least 5, then $A(G) = O(\Delta)$.

We close this subsection by presenting a tightened version of the upper bound on the acyclic chromatic number in Theorem 4.28.

Proposition 4.34. For every finite, simple, and undirected graph G of maximum degree Δ, its acyclic chromatic number satisfies

$$A(G) \leq \lceil 20.571\ \Delta(\sqrt[3]{\Delta} + 1) \rceil. \tag{4.59}$$

Proof. The proof of Theorem 4.28 is not affected by the value of k, up to the derivation of inequalities (4.55)–(4.57). We modify the last part of this proof to obtain a smaller value of k than $\lceil 50\Delta^{4/3} \rceil$. Let $k := \lceil C\Delta(\sqrt[3]{\Delta} + 1) \rceil$, for some constant $C > 0$. The validity of inequality (4.56) clearly implies that the other two inequalities in (4.55) and (4.57) hold. Inequality (4.56) is equivalent to

$$\left(1 - \frac{2}{k}\right)^{5\Delta + 5\Delta^{4/3}} \left(1 - \frac{2}{k^3}\right)^{15\Delta^4} \left(1 - \frac{2}{k^2}\right)^{5\Delta^{8/3}} \geq \frac{1}{2}. \tag{4.60}$$

Referring to the left-hand side of (4.60), we have

$$\begin{aligned}
&\left(1 - \frac{2}{k}\right)^{5\Delta + 5\Delta^{4/3}} \left(1 - \frac{2}{k^3}\right)^{15\Delta^4} \left(1 - \frac{2}{k^2}\right)^{5\Delta^{8/3}} \\
&\geq \left(1 - \frac{10(\Delta^{4/3} + \Delta)}{k}\right)\left(1 - \frac{30\Delta^4}{k^3}\right)\left(1 - \frac{10\Delta^{8/3}}{k^2}\right) \\
&> \left(1 - \frac{10}{C}\right)\left(1 - \frac{30}{C^3}\right)\left(1 - \frac{10}{C^2}\right),
\end{aligned} \tag{4.61}$$

where the first inequality follows from Bernoulli's inequality $(1-x)^n \geq 1-nx$ for $x \in [0,1]$ and $n \in \mathbb{N}$, and the last inequality follows from the definition $k = \lceil C(\Delta^{4/3} + \Delta)$, for some $C > 0$, which implies that $k \geq C(\Delta^{4/3} + \Delta)$, $k^2 > C^2\Delta^{8/3}$, and $k^3 > C^3\Delta^4$. It therefore suffices that the right-hand side of (4.61) be at least $\frac{1}{2}$ to imply that (4.60) holds. The smallest C for which the latter condition is satisfied is $C_{\min} = 20.5707\ldots$, which yields inequality (4.59). □

The tightened upper bound in (4.59) improves upon the bound in (4.51), and the improvement increases from 16% for $\Delta = 1$ to a limiting value of 58.9% as $\Delta \to \infty$.

5. The Moser–Tardos algorithm and the LLL in the variable setting

In this section, we present a constructive version of the LLL in the variable setting, due to Moser and Tardos [65]. In this framework, the bad events depend on a family of mutually independent random variables, and dependencies among events arise through shared variables.

While the LLL guarantees the existence of an assignment avoiding all bad events under suitable conditions, its standard form is nonconstructive. The Moser–Tardos algorithm provides an efficient randomized procedure for finding such an assignment by iteratively resampling the variables associated with violated events.

The analysis relies on the dependency structure among events, captured by proper witness trees, and yields an explicit upper bound on the expected number of resampling steps in the algorithm. This implies that the algorithm terminates almost surely and provides quantitative control over its expected running time.

5.1. The variable setting

Let $\{X_i\}_{i=1}^n$ be mutually independent discrete random variables, where each X_i takes values in a finite set. Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be a finite family of bad events, where each event $A \in \mathcal{A}$ depends only on a subset of $\{X_i\}_{i=1}^n$. Denote by $\mathrm{vbl}(A)$ the set of variables on which A depends. The dependency graph for $\mathcal{A}$ is the graph with vertex set $\mathcal{A}$, in which two distinct events $A, B \in \mathcal{A}$ are adjacent if and only if $\mathrm{vbl}(A) \cap \mathrm{vbl}(B) \neq \emptyset$. In that case, we write $B \sim A$.

5.2. The LLL in the variable setting

The following result specializes Theorem 1.6 to the variable setting, providing a sufficient condition under which none of the events in $\mathcal{A}$ occurs with positive probability.

Theorem 5.1. *(LLL: Variable setting)* Suppose there exist numbers $x(A) \in [0,1)$ for each $A \in \mathcal{A}$ such that

$$\mathbb{P}(A) \leq x(A) \prod_{B:\, B \sim A} (1 - x(B)) \qquad \forall\, A \in \mathcal{A}. \tag{5.1}$$

Then,

$$\mathbb{P}\left(\bigcap_{A \in \mathcal{A}} \overline{A}\right) > 0. \tag{5.2}$$

The importance of the variable setting of the LLL stems from the observation that many of its applications are naturally formulated in terms of underlying independent random variables. Notable examples include conjunctive normal form satisfiability (Section 4.2), Ramsey numbers (Sections 4.3 and 4.4), and hypergraph coloring (Section 4.5). The Moser–Tardos algorithm gives a constructive counterpart of the LLL in this setting.

5.3. The Moser–Tardos algorithm

The Moser–Tardos algorithm proceeds as follows [65]:

- Sample all variables independently.
- While there exists an event $A \in \mathcal{A}$ that occurs, select such an A and resample all the variables in $\mathrm{vbl}(A)$.

The procedure terminates when no event in $\mathcal{A}$ occurs.

5.4. Performance of the Moser–Tardos algorithm

Theorem 5.2. *([65])* Under the condition in (5.1), the Moser–Tardos algorithm terminates almost surely. Moreover, for each $A \in \mathcal{A}$, the expected number of resamplings of A during the execution of the algorithm is at most $\frac{x(A)}{1-x(A)}$. Consequently, the expected total number of resamplings until termination is at most

$$\sum_{A\in\mathcal{A}} \frac{x(A)}{1-x(A)},$$

and the assertion in (5.2) also holds.

Corollary 5.3. *(Symmetric Moser–Tardos criterion)* Suppose that $\mathbb{P}(A) \le p$ for all $A \in \mathcal{A}$, and that each event in $\mathcal{A}$ is adjacent to at most d other events in the dependency graph. If $ep(d+1) \le 1$, then the Moser–Tardos algorithm terminates almost surely. Moreover, if $d \ge 1$, then the expected total number of resamplings is at most $\frac{|\mathcal{A}|}{d}$.

Proof. We apply Theorem 5.2 with

$$x(A) = \frac{1}{d+1}, \quad A \in \mathcal{A}.$$

As in the proof of Theorem 3.1, the condition $ep(d+1) \le 1$ implies that

$$\mathbb{P}(A) \le x(A) \prod_{B\sim A} (1 - x(B)), \quad \forall\, A \in \mathcal{A}.$$

Hence, Theorem 5.2 applies, and for each $A \in \mathcal{A}$,

$$\mathbb{E}[N_A] \le \frac{x(A)}{1-x(A)} = \frac{1}{d},$$

where N_A denotes the number of times A is selected for resampling (equivalently, the number of times all the variables in $\mathrm{vbl}(A)$ are resampled due to the selection of A) during the execution of the algorithm. Some variables in $\mathrm{vbl}(A)$ may also be resampled when other events are selected (due to overlaps among

the variables of distinct events), and such resampling may cause A to occur again. Summing over all $A \in \mathcal{A}$, we obtain

$$\mathbb{E}\left[\text{total number of resamplings}\right] = \sum_{A \in \mathcal{A}} \mathbb{E}[N_A] \leq \frac{|\mathcal{A}|}{d}.$$

In particular, the Moser–Tardos algorithm terminates almost surely. □

5.5. Proper witness trees

We introduce the notion of a proper witness tree, which is used in the proof of Theorem 5.2. We consider rooted trees whose vertices are labeled by events in $\mathcal{A}$, writing $[u] \in \mathcal{A}$ for the label of vertex u.

Definition 5.4. A *proper witness tree* is a rooted tree T whose vertices are labeled by events in $\mathcal{A}$ with the following properties:

(1) For every edge (u, v), where v is a child of u, the labels of u and v share a common variable, i.e.,

$$\mathrm{vbl}([u]) \cap \mathrm{vbl}([v]) \neq \varnothing.$$

(2) For every vertex u, the children of u have pairwise distinct labels.

In Definition 5.4, the labels need not be globally unique: The same event may label multiple vertices in a proper witness tree, as distinctness is only required among the children of any vertex.

Remark 5.5. The first condition in Definition 5.4 reflects the dependency structure: A child corresponds to an earlier resampling of an event that shares a common variable with its parent, and is therefore relevant to the dependency chain leading to the root. Resamplings of events depending on disjoint sets of variables may occur during the execution of the algorithm, but they are included in the proper witness tree only if they are connected to the root through a chain of overlapping events. In particular, every vertex in the tree is connected to the root via a path of overlapping events. The second condition ensures that no event appears more than once among the children of the same vertex.

Let $\mathcal{T}_A$ be the set of finite, proper witness trees whose root is labeled by $A \in \mathcal{A}$, that is, all rooted trees that may arise from the backward construction in some execution, where each vertex is connected to the root via a path of overlapping events.

Fix a random execution of the Moser–Tardos algorithm, and let $A^{(1)}, A^{(2)}, A^{(3)}, \ldots$ be the sequence of resampled events. For each resampling step t, the proper witness tree W_t is uniquely determined by the prefix $A^{(1)}, \ldots, A^{(t)}$ via the following backward construction.

Starting from a root labeled $A^{(t)}$, the indices $t-1, t-2, \ldots, 1$ are processed in decreasing order. A vertex labeled $A^{(s)}$ is added to the tree if and only if $A^{(s)}$ shares a common variable with the label of a vertex already present in the tree; in that case, it is attached as a child of the deepest (i.e., farthest from the root) such vertex. If there is more than one such deepest vertex, then a fixed deterministic tie-breaking rule is used to select one of them.

Every edge in W_t connects two events that share a common variable, so every vertex in W_t is connected to the root by a path of overlapping events. Hence, W_t is a proper witness tree with root labeled $A^{(t)}$.

We next relate resampling steps to proper witness trees. In counting the number of resamplings of an event, each proper witness tree can appear at most once during a single execution; this is established in the following lemma.

Lemma 5.6. *(Distinct proper witness trees)* In a fixed execution, for every pair of distinct resampling steps $s < t$, we have $W_s \neq W_t$. In particular, each proper witness tree τ appears at most once in the execution.

Proof. Fix two distinct resampling steps $s < t$. The proper witness tree W_t is constructed by processing the sequence $A^{(1)}, \ldots, A^{(t)}$ in reverse order, and its root is labeled $A^{(t)}$. When the construction reaches step s, either $A^{(s)}$ is inserted into W_t, or it is skipped.

(1) If $A^{(s)}$ is inserted into W_t, then the construction creates a vertex labeled $A^{(s)}$ in W_t. Since $s < t$, this vertex is not the root of W_t. On the other hand, the root of W_s is labeled $A^{(s)}$. Therefore, W_s and W_t cannot coincide as rooted labeled trees, and hence $W_s \neq W_t$.

(2) If $A^{(s)}$ is not inserted into W_t, then $A^{(s)}$ does not overlap any label already present in the partial proper witness tree. In particular, it does not overlap the root label $A^{(t)}$. Hence $A^{(s)} \neq A^{(t)}$, so W_s and W_t have different root labels, and therefore are distinct.

Thus, in both cases $W_s \neq W_t$. □

By Lemma 5.6, distinct resamplings of A produce distinct proper witness trees in $\mathcal{T}_A$. Hence, the number of resamplings of A equals the number of proper witness trees in $\mathcal{T}_A$ that occur during the execution.

Let N_A denote the number of resamplings of A during this execution, i.e., the number of times A is selected for resampling. Then,

$$N_A = \sum_{\tau \in \mathcal{T}_A} \mathbf{1}_{\{\exists\, t \text{ such that } W_t = \tau\}},$$

and therefore, by linearity of expectation,

$$\mathbb{E}[N_A] = \sum_{\tau \in \mathcal{T}_A} \mathbb{P}\left(\exists\, t \text{ such that } W_t = \tau\right). \tag{5.3}$$

5.6. Witness tree lemma

A proper witness tree τ is said to appear if there exists a resampling step t such that $W_t = \tau$. We now derive an upper bound on the probability that a fixed proper witness tree appears during an execution of the Moser–Tardos algorithm. This bound is a key ingredient in the analysis, as it controls the expected number of resamplings via the summation over all proper witness trees on the right-hand side of (5.3).

Lemma 5.7. *(Witness tree lemma)* For every proper witness tree τ,

$$\mathbb{P}(\tau \text{ appears}) \leq \prod_{v \in V(\tau)} \mathbb{P}([v]), \tag{5.4}$$

where $[v]$ denotes the label of vertex v.

Proof. We use the standard *resampling-table* representation of the algorithm. For each variable X_i, let $X_i^{(0)}, X_i^{(1)}, X_i^{(2)}, \ldots$ be an infinite sequence of samples, each distributed as X_i, and assume that all these samples are mutually independent over all i and all indices. The Moser–Tardos algorithm may be viewed as a deterministic procedure on these tables: initially the value of X_i is $X_i^{(0)}$, and whenever an event A is resampled, the algorithm advances by one step in the table of each variable in $\mathrm{vbl}(A)$.

Fix a proper witness tree τ, and let $A_v := [v]$ denote the label of vertex $v \in V(\tau)$. We shall show that

$$\mathbb{P}(\tau \text{ appears}) \leq \prod_{v \in V(\tau)} \mathbb{P}(A_v). \tag{5.5}$$

Consider the following checking procedure for τ. Initially, for each variable X_i, the current table entry is $X_i^{(0)}$. Process the vertices of τ in any order $v_1, \ldots, v_m$ such that each vertex is processed before its children. Processing a vertex v_r means that we inspect the current table entries of the variables in $\mathrm{vbl}(A_{v_r})$ and check whether the event A_{v_r} occurs under these entries. If A_{v_r} occurs, advance by one step in the table of each variable in $\mathrm{vbl}(A_{v_r})$ and continue; otherwise, the checking procedure fails.

Let F_r be the event that the check at v_r succeeds. At the moment v_r is processed, conditioned on $F_1 \cap \cdots \cap F_{r-1}$, the table entries inspected for the variables in $\mathrm{vbl}(A_{v_r})$ are fresh, independent samples with the same joint distribution as the variables on which A_{v_r} depends. Hence,

$$\mathbb{P}(F_r \mid F_1 \cap \cdots \cap F_{r-1}) = \mathbb{P}(A_{v_r}), \tag{5.6}$$

where the conditioning on the left-hand side of (5.6) is omitted for $r = 1$. Therefore, by the chain rule,

$$\begin{aligned} \mathbb{P}(\text{the checking procedure for } \tau \text{ succeeds}) &= \mathbb{P}\left(\bigcap_{r=1}^{m} F_r\right) \\ &= \prod_{r=1}^{m} \mathbb{P}(F_r \mid F_1 \cap \cdots \cap F_{r-1}) \\ &= \prod_{v \in V(\tau)} \mathbb{P}(A_v), \end{aligned} \tag{5.7}$$

where the last equality holds by (5.6).

We next show that

$$\{\tau \text{ appears}\} \subseteq \{\text{the checking procedure for } \tau \text{ succeeds}\}. \tag{5.8}$$

Suppose that τ appears during an execution of the Moser–Tardos algorithm, say $\tau = W_t$. The backward construction of W_t associates each vertex of τ with a resampling step at which the corresponding event occurred. Moreover, when the sequence is processed in reverse order, if two resampled events share a common variable, then the earlier resampling is attached as a descendant of the later one. We now run the checking procedure on τ. Since the vertices are processed from the root towards the leaves, the advances of the table indices before processing a vertex v account exactly for the later resamplings involving the same variables, represented by the ancestors of v. Therefore, when v is processed, the table entries inspected are precisely those corresponding to the occurrence of A_v at its associated resampling step in the execution. Since the algorithm resamples an event only when it occurs, each such check succeeds. Hence, if τ appears, then the checking procedure for τ succeeds, which proves (5.8).

Finally, combining (5.7) and (5.8), it follows that

$$\mathbb{P}(\tau \text{ appears}) \leq \mathbb{P}(\text{the checking procedure for } \tau \text{ succeeds}) = \prod_{v \in V(\tau)} \mathbb{P}(A_v),$$

which proves inequality (5.5). □

5.7. Bounding the expected number of resamplings

Combining (5.3) and (5.4) implies that, for all $A \in \mathcal{A}$,

$$\mathbb{E}[N_A] \leq \sum_{\tau \in \mathcal{T}_A} \prod_{v \in V(\tau)} \mathbb{P}([v]). \tag{5.9}$$

We next bound this sum using a branching process.

5.8. Branching process

A Galton–Watson branching process is a stochastic process in which each vertex independently generates a random number of children according to a common offspring distribution, thereby generating a random rooted tree.

Here and throughout, $B \sim C$ denotes that the events B and C depend on a common variable, i.e., $\mathrm{vbl}(B) \cap \mathrm{vbl}(C) \neq \emptyset$. For an event $B \in \mathcal{A}$, let

$$\Gamma(B) := \{C \in \mathcal{A} \setminus \{B\} : C \sim B\}$$

be the set of events, excluding B itself, that share at least one variable with B; equivalently, $\Gamma(B)$ is the set of neighbors of B in the dependency graph. Furthermore, let $\Gamma^+(B) := \Gamma(B) \cup \{B\}$.

Fix $A \in \mathcal{A}$. We define a branching process that generates proper witness trees rooted at A.

- Start with a root labeled A.
- For each node labeled B and each event $C \in \Gamma^+(B)$, independently include a child labeled C with probability $x(C)$.

The process continues until it becomes extinct, i.e., until no new vertices are generated in some generation (depending on the probabilities, this may occur with probability strictly less than one). This randomized process generates labeled rooted trees in which every edge connects overlapping events and children have distinct labels. In particular, every tree produced by this process is a proper witness tree (though not every such tree necessarily arises from an execution of the algorithm). In this randomized process, each node labeled B generates a random subset $S \subseteq \Gamma^+(B)$ with probability

$$\prod_{C \in S} x(C) \prod_{C \in \Gamma^+(B) \setminus S} (1 - x(C)).$$

We now compute the probability that this process generates a given proper witness tree.

Lemma 5.8. Let τ be a proper witness tree rooted at A. Then, the probability that the branching process produces τ is

$$p_\tau = \frac{1 - x(A)}{x(A)} \prod_{v \in V(\tau)} \left(x([v]) \prod_{C \in \Gamma([v])} (1 - x(C)) \right). \tag{5.10}$$

Proof. Fix a proper witness tree τ rooted at A. We compute the probability that the branching process produces exactly τ.

In the branching process, each vertex independently selects its set of children. Therefore, the probability of generating the tree τ is equal to the product, over all vertices $v \in V(\tau)$, of the probabilities

that each vertex selects precisely its children as prescribed by τ. For a vertex $v \in V(\tau)$, let children(v) denote the set of labels of the children of v in the tree τ.

Let $v \in V(\tau)$ be a vertex with label $[v]$. The set of its children in τ is a subset of $\Gamma^+([v])$, i.e., children$(v) \subseteq \Gamma^+([v])$. By the construction of the branching process, each vertex $C \in \Gamma^+([v])$ is included as a child independently with probability $x(C)$. Hence, the probability that v selects exactly its children in τ is

$$\prod_{C \in \text{ children}(v)} x(C) \prod_{C \in \Gamma^+([v]) \setminus \text{ children}(v)} (1 - x(C)).$$

Multiplying these probabilities over all vertices $v \in V(\tau)$, we obtain

$$p_\tau = \prod_{v \in V(\tau)} \left(\prod_{C \in \text{ children}(v)} x(C) \prod_{C \in \Gamma^+([v]) \setminus \text{ children}(v)} (1 - x(C)) \right).$$

For each vertex v, we rewrite the second factor as

$$\prod_{C \in \Gamma^+([v]) \setminus \text{ children}(v)} (1 - x(C)) = \frac{\prod_{C \in \Gamma^+([v])} (1 - x(C))}{\prod_{C \in \text{ children}(v)} (1 - x(C))}.$$

Substituting this identity gives

$$p_\tau = \prod_{v \in V(\tau)} \left(\prod_{C \in \Gamma^+([v])} (1 - x(C)) \right) \cdot \prod_{v \in V(\tau)} \prod_{C \in \text{ children}(v)} \frac{x(C)}{1 - x(C)}.$$

Let r denote the root of τ. Then, $[r] = A$. Since each nonroot vertex appears exactly once as a child of its parent,

$$\prod_{v \in V(\tau)} \prod_{C \in \text{ children}(v)} \frac{x(C)}{1 - x(C)} = \prod_{v \in V(\tau) \setminus \{r\}} \frac{x([v])}{1 - x([v])}.$$

Rewriting this as a product over all vertices gives

$$\prod_{v \in V(\tau) \setminus \{r\}} \frac{x([v])}{1 - x([v])} = \frac{1 - x(A)}{x(A)} \prod_{v \in V(\tau)} \frac{x([v])}{1 - x([v])},$$

and therefore

$$p_\tau = \frac{1 - x(A)}{x(A)} \prod_{v \in V(\tau)} \left(\frac{x([v])}{1 - x([v])} \prod_{C \in \Gamma^+([v])} (1 - x(C)) \right).$$

Finally, since $\Gamma^+([v]) = \Gamma([v]) \cup \{[v]\}$ and the union is disjoint, equality (5.10) follows. □

Since the Galton–Watson branching process produces at most one tree (not necessarily from $\mathcal{T}_A$, as it can produce an infinite tree), we have

$$\sum_{\tau \in \mathcal{T}_A} p_\tau \leq 1.$$

Therefore, by Lemma 5.8 and since $x([v]) \in [0, 1)$,

$$\sum_{\tau \in \mathcal{T}_A} \prod_{v \in V(\tau)} \left(x([v]) \prod_{C \in \Gamma([v])} (1 - x(C)) \right) \leq \frac{x(A)}{1 - x(A)}. \tag{5.11}$$

5.9. Completion of the proof of Theorem 5.2

By condition (5.1), for every vertex $v \in V(\tau)$,

$$\mathbb{P}([v]) \le x([v]) \prod_{C \in \Gamma([v])} (1 - x(C)).$$

Hence, for every $\tau \in \mathcal{T}_A$ with $A \in \mathcal{A}$,

$$\prod_{v \in V(\tau)} \mathbb{P}([v]) \le \prod_{v \in V(\tau)} \left(x([v]) \prod_{C \sim [v]} (1 - x(C)) \right). \tag{5.12}$$

Combining (5.9), (5.11), and (5.12), we obtain

$$\mathbb{E}[N_A] \le \sum_{\tau \in \mathcal{T}_A} \prod_{v \in V(\tau)} \mathbb{P}([v]) \le \frac{x(A)}{1 - x(A)},$$

and summing over all events $A \in \mathcal{A}$ implies that

$$\mathbb{E}[\text{total resamplings}] \le \sum_{A \in \mathcal{A}} \frac{x(A)}{1 - x(A)} < \infty.$$

Consequently, because the expected total number of resamplings is finite, the algorithm terminates almost surely after finitely many resampling steps and produces an assignment in which none of the events in $\mathcal{A}$ occurs. Consequently, inequality (5.2) holds.

5.10. Shearer's bound and beyond for the Moser–Tardos algorithm

We close this section by noting that, in the variable setting considered here, recent works [67, 69, 70] have shown that the efficient region of the Moser–Tardos algorithm reaches, and in some cases exceeds, Shearer's bound (Theorem 3.6). This does not contradict the optimality of Shearer's criterion in the abstract LLL (Theorem 1.6), where only the dependency graph and the event probabilities are taken into account. In the variable model, however, one retains additional structural information beyond the dependency graph. Specifically, the variable model is described by a bipartite variable–event graph whose vertices are the variables and bad events, with an edge joining a variable to a bad event whenever the latter depends on the former. The corresponding dependency graph is then obtained by projecting this bipartite graph onto the event vertices, linking two events whenever they depend on a common variable.

The connection between Shearer's criterion and the Moser–Tardos algorithm was established by Kolipaka and Szegedy [67], who showed that, when only the dependency graph and event probabilities are taken into account, the efficient region of the algorithm coincides with the Shearer region.

Building on the additional structure available in the variable setting, the authors in [69] developed a refined analysis of the Moser–Tardos algorithm and showed that the efficient region extends beyond the Shearer region if the dependency graph is nonchordal, that is, if it contains an induced cycle of length at least four. By contrast, for chordal dependency graphs (i.e., graphs in which every cycle of length at least four has a chord), Shearer's bound still exactly characterizes the efficient region.

Subsequently, [70] clarified the role of nonchordality by studying the relationship between the abstract-LLL and variable-LLL boundaries. In particular, it showed that the two boundaries coincide for trees, whereas the presence of an induced cycle of length at least four gives rise to a genuine gap between them. Thus, while Shearer's criterion remains the optimal threshold in the abstract dependency-graph setting, the variable model retains structural information that is discarded when passing to the dependency graph and that can be exploited algorithmically to obtain convergence guarantees beyond the Shearer threshold.

6. The entropy-compression principle

The entropy-compression principle, commonly attributed to Moser [66] and whose name was coined by Tao [71], is closely related to the Moser–Tardos resampling framework for the LLL. The basic idea is to analyze a randomized correction algorithm through suitably defined execution logs that record the bad events encountered and the corresponding resampling steps. Suppose that every execution surviving at least t steps can be associated with a pair consisting of an execution log and the state of the algorithm after step t, and that this correspondence is injective. If, moreover, the total number of such log–state pairs grows at a strictly smaller exponential rate than the number of possible random input sequences, then the proportion of inputs generating long executions decays exponentially with t. Consequently, the algorithm terminates almost surely.

This viewpoint has proved particularly fruitful in probabilistic combinatorics [72, 73], where it often yields constructive existence proofs and quantitative bounds through direct counting arguments on the possible execution histories of randomized algorithms.

The following theorem abstracts the counting argument underlying the entropy-compression method introduced by Moser [66]; see also Tao [71].

Theorem 6.1. *(Entropy-compression principle)* Let Ω be a finite set of size q, and consider a randomized algorithm whose random choices are independent and uniformly distributed over Ω. Assume that once the sequence of random choices is fixed, the execution of the algorithm is uniquely determined. For each integer $t \geq 1$, let B_t be defined as

$$B_t := \Big\{ R \in \Omega^t : \text{the execution determined by } R \text{ reaches step } t \Big\}, \tag{6.1}$$

which is a deterministic subset of Ω^t. For every $R \in B_t$, let $L_t(R) \in \mathcal{L}_t$ be a record (or log) of the information collected during the first t steps of the execution, and let $S_t(R) \in \mathcal{S}_t$ be the state after step t. Assume that for every $t \geq 1$, the mapping

$$\Phi_t \colon B_t \to \mathcal{L}_t \times \mathcal{S}_t, \qquad \Phi_t(R) = (L_t(R), S_t(R)), \tag{6.2}$$

is injective, and that

$$|\mathcal{L}_t \times \mathcal{S}_t| \leq C\alpha^t, \tag{6.3}$$

for some constants $C > 0$ and $\alpha < q$, independent of t. Then,

$$\mathbb{P}(\text{the algorithm reaches step } t) \leq C \left(\frac{\alpha}{q}\right)^t. \tag{6.4}$$

Consequently, the algorithm terminates almost surely.

Proof. Since, by assumption, the mapping $\Phi_t\colon B_t \to \mathcal{L}_t \times \mathcal{S}_t$ is injective, it follows that

$$|B_t| \le |\mathcal{L}_t \times \mathcal{S}_t| \le C\alpha^t. \tag{6.5}$$

Therefore, for every integer $t \ge 1$,

$$\begin{aligned}\mathbb{P}(\text{the algorithm lasts at least } t \text{ steps}) &= \frac{|B_t|}{q^t} \\ &\le C\left(\frac{\alpha}{q}\right)^t. \end{aligned} \tag{6.6}$$

Since $\alpha < q$, the right-hand side of (6.6) tends to 0 as $t \to \infty$. Let T denote the execution time of the algorithm. Then,

$$\lim_{t\to\infty} \mathbb{P}(T > t) = 0. \tag{6.7}$$

Hence $\mathbb{P}(T < \infty) = 1$, and therefore the algorithm terminates almost surely. □

Remark 6.2. (Connection to the Shannon entropy) The terminology *entropy compression* stems from the following information-theoretic interpretation. Since the entries of the random vector $R = (r_1, \ldots, r_t)$ are independent and uniformly distributed over an alphabet Ω of size q, the Shannon entropy of R equals

$$\mathrm{H}(R) = t \log q. \tag{6.8}$$

Furthermore, since the map $R \mapsto (L_t, S_t)$ is injective, the pair (L_t, S_t) uniquely determines the random input vector R. This implies that the conditional entropy of R given (L_t, S_t) satisfies

$$\mathrm{H}(R \,|\, L_t, S_t) = 0. \tag{6.9}$$

Consequently, by the chain rule for the entropy (see [90, Theorem 2.5.1]), it follows that

$$\begin{aligned}\mathrm{H}(R) &\le \mathrm{H}(R, L_t, S_t) \\ &= \mathrm{H}(R \,|\, L_t, S_t) + \mathrm{H}(L_t, S_t) \\ &= \mathrm{H}(L_t, S_t). \end{aligned} \tag{6.10}$$

Since the Shannon entropy of a random vector is upper bounded by the logarithm of the number of its possible outcomes (see [90, Theorem 2.6.4]), it follows from (6.3) that

$$\begin{aligned}\mathrm{H}(L_t, S_t) &\le \log(C\alpha^t) \\ &= \log C + t \log \alpha. \end{aligned} \tag{6.11}$$

Combining (6.8)–(6.11) therefore gives

$$t \log q = \mathrm{H}(R) \le \mathrm{H}(L_t, S_t) \le \log C + t \log \alpha. \tag{6.12}$$

Since by assumption α and C are fixed positive constants with $\alpha < q$, inequality (6.12) is violated for sufficiently large t. Thus, excessively long executions would imply that the random input vector R can be represented by the pair (L_t, S_t), whose number of possible outcomes grows asymptotically like α^t with $\alpha < q$. In this sense, the information contained in the random source is effectively compressed, contradicting the entropy lower bound $\mathrm{H}(R) = t \log q$.

Remark 6.3. (Connection to the LLL) The entropy-compression principle is closely related to the Moser–Tardos algorithmic proof of the LLL. The entropy-compression framework applies naturally to the Moser–Tardos resampling algorithm. Indeed, the successive resampling decisions may be viewed as a sequence of independent random choices drawn from a finite alphabet, and the resulting execution history is a deterministic function of this random sequence. Thus, the Moser–Tardos algorithm is an instance of the abstract randomized process considered in the entropy-compression principle. In both settings, one considers a randomized correction procedure in which variables are sampled independently and local resampling steps are performed whenever a bad event occurs.

The two approaches differ primarily in how the execution process is analyzed. The Moser–Tardos framework studies the resampling algorithm through witness trees and branching-process estimates, whereas entropy-compression arguments analyze execution histories via combinatorial logs and injective reconstruction maps. More precisely, the entropy-compression approach shows that excessively long executions would yield an injective encoding of the random input into a family of descriptions whose cardinality is too small to accommodate the entropy of the underlying random source. Consequently, the number of possible long execution histories is exponentially smaller than the number of random inputs that would have to be encoded by them.

7. The lopsided LLL

The LLL has an important refinement in which the standard notion of a dependency graph is replaced by the weaker notion of a *lopsidependency graph*. This refinement is especially useful in problems involving random permutations, matchings, and transversals, where the relevant bad events are generally not independent but satisfy a suitable negative-dependence condition. In Section 7.1 we present the lopsided LLL, and in Section 7.2 we apply it to the problem of Latin transversals.

7.1. Formulation of the lopsided LLL

In this subsection, we present the lopsided LLL and discuss several related remarks.

Definition 7.1. (Lopsidependency graph) Let $\{A_i\}_{i\in I}$ be a finite family of events in a probability space. A graph G on the vertex set I is called a *lopsidependency graph* for $\{A_i\}_{i\in I}$ if, for every $i \in I$ and every set

$$S \subseteq I \setminus (\Gamma(i) \cup \{i\}), \tag{7.1}$$

where $\Gamma(i)$ denotes the set of neighbors of i in G, we have

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j\in S} \overline{A_j}\right) \leq \mathbb{P}(A_i). \tag{7.2}$$

Theorem 7.2. *(LLL)* Let $\{A_i\}_{i\in I}$ be a finite family of events in a probability space, and let G be a lopsidependency graph for $\{A_i\}_{i\in I}$. If there exist numbers $\{x_i\}_{i\in I} \subseteq [0, 1)$ such that

$$\mathbb{P}(A_i) \leq x_i \prod_{j\in\Gamma(i)} (1 - x_j), \qquad \forall\, i \in I, \tag{7.3}$$

then

$$\mathbb{P}\left(\bigcap_{i\in I} \overline{A_i}\right) \geq \prod_{i\in I} (1 - x_i) > 0. \tag{7.4}$$

Proof. This result follows from the proof of the LLL given in Section 2. Specifically, Lemma 2.1 remains valid under the weaker assumptions (7.2) and (7.3), with the obvious notational modification that the index set $[n]$ is replaced by the finite set I. Consequently, Step 2 of that proof remains valid under the same assumptions. □

A convenient symmetric form follows from Theorem 7.2; compared with Theorem 3.1, it requires a weaker hypothesis.

Corollary 7.3. *(Symmetric lopsided LLL)* Under the assumptions of Theorem 7.2 with $\mathbb{P}(A_i) \leq p$ for all $i \in I$, suppose that every vertex of G has degree at most $d \geq 1$, and $ep(d+1) \leq 1$. Then,

$$\mathbb{P}\left(\bigcap_{i\in I} \overline{A_i}\right) \geq \left(\frac{d}{1+d}\right)^{|I|}. \tag{7.5}$$

Proof. Apply Theorem 7.2 with $x_i = \frac{1}{1+d}$ for all $i \in I$. The rest of the proof proceeds as in the proof of Theorem 3.1. □

Remark 7.4. The lopsided LLL is a refinement of the LLL (Theorem 1.6) when the latter is formulated in terms of undirected dependency graphs. Indeed, in this formulation, each event A_i is independent of the σ-algebra generated by the events A_j with $j \notin \Gamma(i) \cup \{i\}$. This implies, in particular, that

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j\in S} \overline{A_j}\right) = \mathbb{P}(A_i), \tag{7.6}$$

for all $S \subseteq I \setminus (\Gamma(i) \cup \{i\})$, whenever the conditional probability is well-defined. Hence, every dependency graph is also a lopsidependency graph.

The converse is not true in general. The lopsided condition requires only that conditioning on the nonoccurrence of non-neighboring bad events does not increase the probability of a given bad event. Consequently, the lopsided LLL can be applicable even when the ordinary LLL is not. As illustrated in Section 7.2, this is particularly useful in permutation and matching problems, where the dependence structure often prevents a direct application of the ordinary LLL, while still satisfying the negative-dependence condition (7.2) required by the lopsided LLL.

Remark 7.5. The conditions in Definition 7.1 and Theorem 7.2 may be formulated slightly more generally using auxiliary numbers $\{p_i\}_{i\in I} \subseteq [0, 1)$, satisfying

$$p_i \geq \mathbb{P}(A_i). \tag{7.7}$$

Indeed, it suffices to assume that for every $i \in I$ and every subset S satisfying (7.1), we have

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j\in S} \overline{A_j}\right) \leq p_i, \tag{7.8}$$

and that there exist numbers $x_i \in [0, 1)$ such that

$$p_i \leq x_i \prod_{j\in\Gamma(i)} (1 - x_j), \qquad \forall\, i \in I. \tag{7.9}$$

Note that taking $S = \varnothing$ in (7.8) yields (7.7) for every $i \in I$. The standard proof of the lopsided LLL then applies verbatim, yielding (7.4).

7.2. Latin transversals via the lopsided LLL

Latin transversals are important combinatorial objects that arise naturally in matching and assignment problems, combinatorial design theory, graph coloring, and scheduling. Their study is closely connected to Latin squares, permutation problems, and the probabilistic method. In particular, deriving sufficient conditions for the existence of a Latin transversal in a given matrix is a classical application of the lopsided LLL [78].

Definition 7.6. (Latin transversal of a matrix) A *Latin transversal* of an $n \times n$ matrix $\mathbf{A} = (a_{i,j})$ is a selection of n entries, one from each row and each column, whose values are pairwise distinct. Equivalently, it is a set of positions

$$\{(i, \pi(i)) : 1 \le i \le n\},$$

where π is a permutation of $[n]$, such that the sequence $\{a_{i,\pi(i)}\}_{i=1}^{n}$ has pairwise distinct terms.

Theorem 7.7. *(Existence of Latin transversals)* Let $\mathbf{A} = (a_{i,j})$ be an $n \times n$ matrix in which every symbol occurs at most k times. If

$$k \le \frac{n-1}{4e} + 1 - \frac{1}{4n}, \tag{7.10}$$

then $\mathbf{A}$ contains a Latin transversal.

Proof. Let π be a uniformly random permutation of $[n]$. Denote by $\mathcal{T}$ the set of all ordered quadruples (i, j, i', j') satisfying $i < i'$, $j \ne j'$, and $a_{i,j} = a_{i',j'}$. Thus, (i, j) and (i', j') are positions containing the same symbol and lying in distinct rows and distinct columns. For every $(i, j, i', j') \in \mathcal{T}$, define the event

$$A_{i,j,i',j'} = \{(\pi(i), \pi(i')) = (j, j')\}. \tag{7.11}$$

We first observe that

$$\mathbf{A} \text{ contains a Latin transversal} \iff \mathbb{P}\left(\bigcap_{(i,j,i',j')\in\mathcal{T}} \overline{A_{i,j,i',j'}}\right) > 0. \tag{7.12}$$

We prove each implication separately.

- Suppose that $\mathbf{A}$ contains a Latin transversal. Then, there exists a permutation π_0 of $[n]$ such that $a_{i,\pi_0(i)} \ne a_{i',\pi_0(i')}$ for all $i < i'$. Hence, for every $(i, j, i', j') \in \mathcal{T}$, we have $(\pi_0(i), \pi_0(i')) \ne (j, j')$ since otherwise $a_{i,\pi_0(i)} = a_{i,j} = a_{i',j'} = a_{i',\pi_0(i')}$, contradicting the fact that the selected entries are distinct. Hence, none of the events $A_{i,j,i',j'}$ with $(i, j, i', j') \in \mathcal{T}$ occurs when $\pi = \pi_0$. Since π is uniformly distributed over the $n!$ permutations of $[n]$, we have $\mathbb{P}(\pi = \pi_0) = \frac{1}{n!} > 0$. Consequently,

$$\mathbb{P}\left(\bigcap_{(i,j,i',j')\in\mathcal{T}} \overline{A_{i,j,i',j'}}\right) > 0. \tag{7.13}$$

- Conversely, suppose that (7.13) holds. Then, there exists a permutation π_0 of $[n]$ for which no event $A_{i,j,i',j'}$ with $(i, j, i', j') \in \mathcal{T}$ occurs. We claim that

$$T_{\pi_0} := \{(i, \pi_0(i)) : i \in [n]\}$$

is a Latin transversal. Otherwise, there exist indices $i < i'$ such that $a_{i,\pi_0(i)} = a_{i',\pi_0(i')}$. Set $j := \pi_0(i)$ and $j' := \pi_0(i')$. Since π_0 is a permutation and $i \neq i'$, we have $j \neq j'$. Then, $(i, j, i', j') \in \mathcal{T}$ and $(\pi_0(i), \pi_0(i')) = (j, j')$. Therefore, the event $A_{i,j,i',j'}$ occurs when $\pi = \pi_0$, contradicting the choice of π_0. Hence the entries $\{a_{i,\pi_0(i)}\}_{i=1}^n$ are pairwise distinct, and therefore T_{π_0} is a Latin transversal.

Now fix $(i, j, i', j') \in \mathcal{T}$. Since π is a uniformly random permutation of $[n]$, and $i \neq i'$, $j \neq j'$,

$$\mathbb{P}(A_{i,j,i',j'}) = \mathbb{P}(\pi(i) = j,\ \pi(i') = j') = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} =: p.$$

Define a graph G on the vertex set $\mathcal{T}$ by joining two distinct vertices (i, j, i', j') and (ℓ, m, ℓ', m') whenever

$$\{i, i'\} \cap \{\ell, \ell'\} \neq \varnothing \quad \text{or} \quad \{j, j'\} \cap \{m, m'\} \neq \varnothing. \tag{7.14}$$

By a result of Erdős and Spencer [78, Section 2], the graph G is a lopsidependency graph for the family of events

$$\{A_{i,j,i',j'} : (i, j, i', j') \in \mathcal{T}\}.$$

By Definition 7.1, let $(i, j, i', j') \in \mathcal{T}$, and let $S \subseteq \mathcal{T}$ be a set of vertices of G that are nonadjacent to (i, j, i', j') in G. Then,

$$\mathbb{P}\left(A_{i,j,i',j'} \,\middle|\, \bigcap_{(\ell,m,\ell',m') \in S} \overline{A_{\ell,m,\ell',m'}}\right) \leq \mathbb{P}(A_{i,j,i',j'}),$$

which is the lopsidependency condition required by the lopsided LLL.

It remains to derive an upper bound on the maximum degree of G. Fix a vertex $(i, j, i', j') \in \mathcal{T}$. An adjacent vertex must contain a position (s, t) such that $s \in \{i, i'\}$ or $t \in \{j, j'\}$. There are at most $4n$ such positions (s, t). For each such position (s, t), since every symbol occurs at most k times in the matrix, there are at most $k - 1$ other positions (s', t') satisfying $a_{s,t} = a_{s',t'}$. Each admissible pair of positions (s, t) and (s', t') lying in distinct rows and distinct columns determines at most one vertex of $\mathcal{T}$. Consequently, every vertex of G has degree at most $4n(k-1)$, i.e.,

$$d \leq 4n(k-1). \tag{7.15}$$

By the symmetric version of the lopsided LLL (Corollary 7.3), it is enough that $ep(d+1) \leq 1$ to satisfy Eq (7.13). Using the bounds above, this is implied by

$$e \cdot \frac{1}{n(n-1)} \cdot (4n(k-1) + 1) \leq 1,$$

which is equivalent to (7.10). Finally, by (7.12) and (7.13), the matrix $\mathbf{A}$ contains a Latin transversal. □

Remark 7.8. The symmetric version of the LLL (Theorem 3.1) cannot be applied directly in the proof of Theorem 7.7. The reason is that, under the uniform distribution on permutations, events involving disjoint rows and columns are generally not independent. For instance, suppose that i, i', ℓ, ℓ' are distinct and that j, j', m, m' are distinct. Then, for $n \geq 4$,

$$\mathbb{P}(A_{i,j,i',j'} \cap A_{\ell,m,\ell',m'}) = \frac{1}{n(n-1)(n-2)(n-3)},$$

whereas

$$\mathbb{P}(A_{i,j,i',j'})\,\mathbb{P}(A_{\ell,m,\ell',m'}) = \frac{1}{n^2(n-1)^2}.$$

Thus, such events are not independent, even though their corresponding positions lie in disjoint rows and columns. In fact, the former quantity is larger than the latter, showing that these events are positively correlated. Consequently, the occurrence of one such event makes the occurrence of the other more likely, whereas the nonoccurrence of one event tends to make the occurrence of the other less likely. The lopsided LLL is therefore essential here. Rather than requiring independence, it suffices that conditioning on the nonoccurrence of non-neighboring bad events does not increase the probability of a given bad event. Erdős and Spencer showed that this property holds for the events arising from random permutations [78], which allows the lopsided LLL to be applied in the proof of Theorem 7.7.

Remark 7.9. Theorem 7.7 does not guarantee the existence of a Latin transversal in an $n \times n$ Latin square with $n \geq 2$, since in that case every symbol appears exactly n times. Thus $k = n$, and the inequality (7.10) is violated. Latin squares of even order need not have a Latin transversal; for example, the Latin square $\mathbf{A} = (a_{i,j})_{i,j=1}^{n}$, where $a_{i,j} := (i + j) \pmod{n}$, has no Latin transversal for even $n \geq 2$. In contrast, Ryser's conjecture states that every Latin square of odd order has a Latin transversal; see, for example, [91, Conjecture 3.2].

8. The cluster-expansion lemma

This section presents the cluster-expansion lemma, due to Bissacot, Fernández, Procacci, and Scoppola [57]. Their original proof relied on cluster-expansion techniques from statistical physics. Subsequently, Pegden [74] established an algorithmic counterpart within the Moser–Tardos framework, and later Harvey and Vondrák [75] gave a short purely combinatorial proof based on Shearer's work [68]. Since then, the lemma has yielded several improved bounds in combinatorics (see, e.g., [57, 92, 93]).

We begin by stating the cluster-expansion lemma. We then discuss its relationship to the LLL (Theorem 1.6), including its connection to the symmetric form of the LLL. Finally, we revisit the Latin transversal problem from Section 7.2 and show that the cluster-expansion lemma yields a stronger sufficient condition for the existence of a Latin transversal than that obtained there via the lopsided LLL.

Theorem 8.1. *(Cluster-expansion lemma [57])* Let $A_1, \ldots, A_n$ be a finite collection of events in a probability space, and let $G = ([n], E)$ be a dependency graph for $\{A_i\}_{i=1}^{n}$. For each $i \in [n]$, define

$$\Gamma_G(i) = \{j \in [n] \setminus \{i\} : \{i, j\} \in E\}, \qquad \Gamma_G^+(i) = \Gamma_G(i) \cup \{i\}.$$

For $S \subseteq [n]$, let $G[S]$ denote the subgraph of G induced by the vertex set S, and let $\mathrm{Ind}(G[S])$ denote the family of independent sets of $G[S]$. Suppose that there exist numbers $y_1, \ldots, y_n > 0$ such that

$$\mathbb{P}(A_i) \leq \frac{y_i}{\displaystyle\sum_{I \in \mathrm{Ind}(G[\Gamma_G^+(i)])} \prod_{j \in I} y_j}, \qquad \forall\, i \in [n]. \tag{8.1}$$

Then,

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geq \left(\sum_{I \in \mathrm{Ind}(G)} \prod_{i \in I} y_i\right)^{-1}. \tag{8.2}$$

8.1. Connection between the cluster-expansion lemma and the LLL

The cluster-expansion lemma can be viewed as a refinement of the LLL when the latter is formulated in terms of undirected dependency graphs. The following remarks show that the undirected-graph formulation of the LLL follows from a suitable relaxation of the cluster-expansion lemma.

Remark 8.2. (From digraph to graph) The LLL is formulated in Theorem 1.6 in terms of a dependency digraph $D = ([n], E)$. Given such a digraph, one may associate an undirected graph $G = ([n], E')$, defined by

$$\{i, j\} \in E' \quad \Longleftrightarrow \quad (i, j) \in E \text{ or } (j, i) \in E. \tag{8.3}$$

All notions of independence in Theorem 8.1 are with respect to this graph G. It follows from (8.3) that the neighborhood of a vertex in G consists of all vertices that are connected to it by either an incoming or an outgoing edge in D. Consequently, the neighborhood of a vertex in G may be strictly larger than the set of vertices reachable from it by outgoing edges in D. This distinction is relevant when comparing results stated for dependency digraphs, such as the LLL, with results formulated for undirected dependency graphs, such as the cluster-expansion lemma.

Remark 8.3. (Product-type lower bound and relation to the LLL) Under the assumptions of Theorem 8.1,

$$\sum_{I \in \mathrm{Ind}(G)} \prod_{i \in I} y_i \le \sum_{I \subseteq [n]} \prod_{i \in I} y_i = \prod_{i=1}^{n} (1 + y_i), \tag{8.4}$$

where the last inequality holds since the summation on the left is restricted to the independent sets of G, whereas the summation on the right ranges over all subsets of $[n]$. Combining this with (8.2) yields

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i} \right) \ge \prod_{i=1}^{n} \frac{1}{1 + y_i}. \tag{8.5}$$

Similarly, for all $i \in [n]$, since every independent set in the induced subgraph $G[\Gamma_G^+(i)]$ is, in particular, a subset of $\Gamma_G^+(i)$, we obtain

$$\begin{aligned} \sum_{I \in \mathrm{Ind}(G[\Gamma_G^+(i)])} \prod_{j \in I} y_j &\le \sum_{I \subseteq \Gamma_G^+(i)} \prod_{j \in I} y_j \\ &= \prod_{j \in \Gamma_G^+(i)} (1 + y_j) \\ &= (1 + y_i) \prod_{j:\, \{i,j\} \in E} (1 + y_j), \end{aligned} \tag{8.6}$$

and therefore (recall that $y_i > 0$),

$$\frac{y_i}{\displaystyle\sum_{I \in \mathrm{Ind}(G[\Gamma_G^+(i)])} \prod_{j \in I} y_j} \ge \frac{y_i}{1 + y_i} \prod_{j:\, \{i,j\} \in E} \frac{1}{1 + y_j}. \tag{8.7}$$

By (8.7), the condition in (8.1) can be replaced by the following stronger sufficient condition:

$$\mathbb{P}(A_i) \le \frac{y_i}{1+y_i} \prod_{j:\,\{i,j\}\in E} \frac{1}{1+y_j}, \qquad \forall\, i \in [n]. \tag{8.8}$$

Setting $y_i := \frac{x_i}{1-x_i}$, with $x_i \in (0,1)$, gives $\frac{1}{1+y_i} = 1 - x_i$ and $\frac{y_i}{1+y_i} = x_i$. Hence, Eqs (8.8) and (8.5) reduce to Eqs (1.2) and (1.3), respectively, recovering the standard form of the LLL. The derivation above relies on a crude product bound that ignores independence constraints, and is therefore not tight. In this sense, the lower bound provided by the cluster-expansion lemma reduces to that of the LLL after the above relaxation. The cluster-expansion lemma is formulated in terms of an undirected dependency graph, whereas the LLL (Theorem 1.6) is formulated in terms a dependency digraph. By Remark 8.2, passing from a dependency digraph to its associated undirected graph may strengthen the sufficient conditions. Therefore, the above argument does not yield a derivation of the LLL from the cluster-expansion lemma, but rather highlights the relationship between the two results.

8.2. Latin transversals via the cluster-expansion lemma

In Section 7.2, we proved Theorem 7.7 by applying the lopsided LLL to a family of bad events associated with a uniformly random permutation of $[n]$. The same probabilistic model admits a stronger analysis via the cluster-expansion lemma, which yields an improved sufficient condition for the existence of a Latin transversal.

Theorem 8.4. *([57, Proposition 4.2])* Let $\mathbf{A} = (a_{i,j})$ be an $n \times n$ matrix in which, for every symbol, the maximum number of its occurrences is at most k. If

$$k \le \frac{27}{256}\,(n-1) + 1, \tag{8.9}$$

then $\mathbf{A}$ contains a Latin transversal.

Proof. We retain the notation and probabilistic construction from the proof of Theorem 7.7. Thus, π is a uniformly random permutation of $[n]$, $\mathcal{T}$ is the set of all ordered quadruples (i, j, i', j') such that $i < i'$, $j \ne j'$, and $a_{i,j} = a_{i',j'}$, and

$$A_{i,j,i',j'} = \{(\pi(i), \pi(i')) = (j, j')\}, \quad (i, j, i', j') \in \mathcal{T}.$$

As shown earlier,

$$\mathbb{P}(A_{i,j,i',j'}) = \frac{1}{n(n-1)} =: p, \quad (i, j, i', j') \in \mathcal{T}. \tag{8.10}$$

Furthermore, the existence of a Latin transversal is equivalent to the validity of inequality (7.13), so it suffices to prove the latter.

Let G be the dependency graph on vertex set $\mathcal{T}$ introduced in the proof of Theorem 7.7, whose adjacency condition is given by (7.14)). Fix a vertex $x = (i, j, i', j') \in \mathcal{T}$. Following the argument in [57], the neighborhood of x can be covered by the four cliques C_i, $C_{i'}$, C_j, $C_{j'}$, where C_i and $C_{i'}$ consist of all vertices in $\mathcal{T}$ involving rows i and i', respectively, and C_j and $C_{j'}$ consist of all vertices in

$\mathcal{T}$ involving columns j and j', respectively. Indeed, if a vertex (ℓ, ℓ', m, m') is adjacent to x, then by the adjacency condition (7.14), it must share one of the rows i, i' or one of the columns j, j'.

As observed in the proof of Theorem 7.7, each of these four cliques contains at most $n(k-1)$ vertices. Indeed, after fixing a row or a column, there are at most n choices for one entry and at most $k-1$ choices for a second entry containing the same symbol in a distinct row and column.

Let $\mu > 0$, and set $y_x = \mu$ for all $x \in \mathcal{T}$. Since every independent set of the induced subgraph $G[\Gamma_G^+(x)]$ contains at most one vertex from each of the four cliques, we have

$$\sum_{I \in \mathrm{Ind}(G[\Gamma_G^+(x)])} \prod_{z \in I} y_z = \sum_{I \in \mathrm{Ind}(G[\Gamma_G^+(x)])} \mu^{|I|} \leq \big(1 + \mu n(k-1)\big)^4. \tag{8.11}$$

Therefore, condition (8.1) is satisfied whenever

$$p \leq \frac{\mu}{\big(1 + n(k-1)\mu\big)^4}. \tag{8.12}$$

Substituting $p = \frac{1}{n(n-1)}$ from (8.10) and setting

$$\mu := \frac{\alpha}{n(k-1)}, \quad \alpha > 0, \tag{8.13}$$

inequality (8.12) becomes

$$k - 1 \leq \frac{\alpha\,(n-1)}{(1+\alpha)^4}, \quad \alpha > 0. \tag{8.14}$$

Define $f(\alpha) = \frac{\alpha}{(1+\alpha)^4}$ for $\alpha > 0$. Elementary calculus shows that f attains its maximum at $\alpha = \frac{1}{3}$, where $f(\frac{1}{3}) = \frac{27}{256}$. Maximizing the right-hand side of (8.14) over $\alpha > 0$, we recover inequality (8.9). Consequently, by (8.10) and the choice $y_x = \mu$ for all $x \in \mathcal{T}$, condition (8.1) is satisfied for the family of events $\{A_{i,j,i',j'}\}_{(i,j,i',j') \in \mathcal{T}}$. By the cluster-expansion lemma, inequality (7.13) holds. Hence, by the equivalence established in (7.12), the matrix $\mathbf{A}$ contains a Latin transversal. □

Remark 8.5. Theorem 8.4 improves Theorem 7.7 by replacing the condition (7.10) with the weaker condition (8.9). Since $\frac{27}{256} \approx 0.10547$ and $\frac{1}{4e} \approx 0.09197$, the cluster-expansion refinement permits a strictly larger admissible range of values for k. This improvement was one of the motivating examples in [57], illustrating how the cluster-expansion lemma can exploit the clique structure of dependency neighborhoods more effectively than the classical LLL when the dependency graph is undirected. Harris and Srinivasan [94] gave the first randomized polynomial-time algorithm for finding a Latin transversal under the condition of Theorem 8.4.

9. The iterated LLL

In the preceding sections, the LLL is applied as a one-shot probabilistic tool: one defines a collection of bad events and proves that there is a positive probability that none of them occurs simultaneously. In many combinatorial problems, however, a direct application of the lemma is either impossible or yields

only weak results because the dependencies among the bad events are too strong. A powerful extension of this idea is provided by the *iterated LLL*, also known as the *semi-random method*, which is one of the cornerstones of probabilistic combinatorics.

The origins of this methodology can be traced to Rödl's introduction of the nibble method in his work on hypergraph packings and coverings [95]. Building on earlier probabilistic constructions and ideas, including those of Ajtai, Komlós, Pintz, Spencer, and Szemerédi [96], the nibble method and related semi-random techniques evolved into a powerful framework that underlies many of the strongest results in modern probabilistic combinatorics, particularly in graph and hypergraph coloring [7, 97, 98].

The basic philosophy is to construct the desired object gradually. Instead of making all random choices at once, one performs a partial random construction, removes the resulting conflicts, and then repeats the procedure on the remaining unresolved portion of the problem. The residual structure typically becomes simpler after each stage, eventually reducing the problem to a setting where a final application of the LLL or a deterministic argument completes the construction.

A simple example arises in graph coloring. Let G be a graph with maximum degree Δ, and suppose that one wishes to produce a proper coloring using relatively few colors. Rather than coloring all vertices at once, consider the following iterative procedure. During a given round, every uncolored vertex independently selects a tentative color from a fixed set of colors. If two adjacent vertices are assigned the same color, their tentative assignments are discarded. Vertices that are not involved in any conflict keep their colors permanently, and the procedure is repeated on the subgraph induced by the remaining uncolored vertices.

The key question is whether the residual graph becomes significantly simpler after one round. To analyze this, one studies local parameters such as the number of uncolored neighbors of a given vertex v after the conflict-resolution step. By linearity of expectation, one first shows that the expected residual degree of v is smaller than its original degree, often by a fixed multiplicative factor. However, expectation alone is insufficient; one must also show that most vertices behave close to this average. Appropriate concentration inequalities are therefore used to show that large deviations from the expected behavior are unlikely, occurring with exponentially small probability; see [99, 100] for treatments of the subject.

One then defines a bad event A_v to be the event that the residual degree of v exceeds a prescribed threshold. Although each bad event has small probability, the events are not independent because nearby vertices are influenced by many of the same random color choices. Nevertheless, each event depends only on random choices within a bounded neighborhood of v, and hence each event is dependent on only a limited number of others. The LLL can therefore be applied to show that, with positive probability, none of the events $\{A_v\}$ occurs. Consequently, there exists a round in which every vertex simultaneously experiences the prescribed reduction in residual degree.

This combination of concentration inequalities and the LLL is the hallmark of the semi-random method. Concentration inequalities provide local control by showing that undesirable deviations are individually unlikely, whereas the LLL converts these local estimates into a global statement asserting that all vertices satisfy the required property simultaneously. As a result, after one iteration, the maximum degree of the residual graph is significantly smaller than that of the original graph.

The procedure can then be repeated. If the maximum degree decreases by a constant factor in each round, then after $O(\log \Delta)$ iterations, the residual graph has bounded degree. At that stage, the remaining coloring problem can often be completed by elementary methods or by a final application of the LLL. This iterative reduction of complexity is the essence of the semi-random method.

The semi-random method should also be distinguished from the algorithmic framework of Moser and Tardos in Section 5. Recall that, in the latter approach, one begins with a complete random assignment and repeatedly resamples the variables involved in occurring bad events. By contrast, the semi-random method constructs the desired object incrementally, repeatedly simplifying the remaining instance through a sequence of random and deterministic steps. Although both approaches originate from the ideas underlying the LLL, they employ fundamentally different mechanisms and have led to distinct developments in probabilistic combinatorics.

10. Concluding remarks and outlook

Fifty years after its inception in [1], the LLL continues to pose fundamental conceptual and technical challenges. While Shearer's criterion [68] (Theorem 3.6) provides a sharp characterization of the existential regime, an equally complete understanding of the algorithmic landscape remains elusive.

The algorithmic aspects of the LLL are now well understood in the variable framework, most notably through the Moser–Tardos resampling algorithm [65] and its subsequent analyses (see [70] and references therein). Despite substantial progress, the situation is less clear in more general settings, particularly in the presence of arbitrary dependency structures. The diversity of modern formulations, ranging from lopsided and resampling-oracle variants to commutativity-based conditions [58], suggests that a unifying perspective is still lacking.

Moreover, important directions remain only partially understood. One direction concerns the complexity of distributed and local algorithms for constructing configurations guaranteed by the lemma, particularly near the threshold of its applicability. Another direction concerns approximate counting and sampling in parameter regimes where the LLL guarantees the existence of configurations avoiding all bad events. A further direction seeks to extend and refine the lemma in more general settings, including measurable probability spaces and noncommutative (quantum) frameworks. Although substantial progress has been made on measurable versions of the LLL and related measurable graph-coloring problems, notably through the work of Bernshteyn [101], a complete understanding of the scope, limitations, and optimal conditions of such extensions has yet to emerge. In each of these directions, many fundamental questions concerning sharp thresholds, constructive methods, and the precise boundaries of applicability remain open.

Use of Generative-AI tools declaration

While writing this expository paper, the author used ChatGPT (OpenAI) solely for language editing and stylistic refinement of the manuscript. The author assumes full responsibility for its content.

Acknowledgments

The author gratefully acknowledges the three anonymous referees for their timely reports and constructive comments, which helped improve the presentation of this paper, and Ugo Vaccaro for raising a question that led to Corollary 3.4.

Conflict of interest

The author declares no conflict of interest.

References

1. P. Erdős, L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, *Colloq. Math. Soc. János Bolyai*, North-Holland, 1975, 609–627.
2. N. Alon, J. H. Spencer, *The probabilistic method*, 4 Eds., Wiley, 2016.
3. M. Mitzenmacher, E. Upfal, *Probability and computing: randomization and probabilistic techniques in algorithms and data analysis*, 2 Eds., Cambridge University Press, 2017.
4. P. Brémaud, *Discrete probability models and methods: probability on graphs and trees, Markov chains and random fields, entropy and coding*, Springer, 2017. https://doi.org/10.1007/978-3-319-43476-6
5. B. Bollobás, *Random graphs*, 2 Eds., Cambridge University Press, 2001. https://doi.org/10.1017/CBO9780511814068
6. M. Molloy, The probabilistic method, In: M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed, *Probabilistic methods for algorithmic discrete mathematics*, Springer, 1998, 1–35. https://doi.org/10.1007/978-3-662-12788-9_1
7. M. Molloy, B. Reed, *Graph colouring and the probabilistic method*, Springer, 2002. https://doi.org/10.1007/978-3-642-04016-0
8. J. H. van Lint, R. M. Wilson, *A course in combinatorics*, 2 Eds., Cambridge University Press, 2001. https://doi.org/10.1017/CBO9780511987045
9. S. Jukna, *Extremal combinatorics with applications in computer science*, 2 Eds., Springer, 2011. https://doi.org/10.1007/978-3-642-17364-6
10. D. E. Knuth, *The art of computer programming, Volum 4B, the: combinatorial algorithms*, Addison-Wesley Professional, 2022.
11. N. Alon, C. McDiarmid, M. Molloy, Edge-disjoint cycles in regular directed graphs, *J. Graph Theory*, **22** (1996), 231–237.
12. A. Z. Broder, A. M. Frieze, E. Upfal, Existence and construction of edge-disjoint paths on expander graphs, *SIAM J. Comput.*, **23** (1994), 976–989. https://doi.org/10.1137/S0097539792232021
13. Y. J. Chang, Q. He, W. Li, S. Pettie, J. Uitto, Distributed edge coloring and a special case of the constructive Lovász local lemma, *ACM Trans. Algorithms*, **16** (2020), 8. https://doi.org/10.1145/3365004
14. J. Spencer, Ramsey's theorem–a new lower bound, *J. Combin. Theory Ser. A*, **18** (1975), 108–115. https://doi.org/10.1016/0097-3165(75)90071-0
15. J. Spencer, Asymptotic lower bounds for Ramsey functions, *Discrete Math.*, **20** (1977), 69–76. https://doi.org/10.1016/0012-365X(77)90044-9
16. Y. Li, Q. Lin, *Elementary methods of graph Ramsey theory*, Springer, 2022. https://doi.org/10.1007/978-3-031-12762-5

17. N. Alon, N. Linial, Cycles of length 0 (mod k) in directed graphs, *J. Combin. Theory Ser. B*, **47** (1989), 114–119. https://doi.org/10.1016/0095-8956(89)90071-3

18. N. Alon, Independent sets in regular graphs and sum-free subsets of finite groups, *Israel J. Math.*, **73** (1991), 247–256. https://doi.org/10.1007/BF02772952

19. N. Alon, C. McDiarmid, B. Reed, Acyclic coloring of graphs, *Random Structures Algorithms*, **2** (1991), 277–288. https://doi.org/10.1002/rsa.3240020303

20. A. Kırtışoğlu, L. Özkahya, Coloring of graphs avoiding bicolored paths of a fixed length, *Graphs Combin.*, **40** (2024), 11. https://doi.org/10.1007/s00373-023-02739-4

21. K. Kolipaka, M. Szegedy, Y. Xu, A sharper local lemma with improved applications, In: A. Gupta, K. Jansen, J. Rolim, R. Servedio, *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, Springer, 2012, 603–614, https://doi.org/10.1007/978-3-642-32512-0_51

22. D. Deng, D. R. Stinson, R. Wei, The Lovász local lemma and its applications to some combinatorial arrays, *Des. Codes Crypt.*, **32** (2004), 121–134, https://doi.org/10.1023/B:DESI.0000029217.97956.26

23. Z. Füredi, J. Kahn, On the dimensions of ordered sets of bounded degree, *Order*, **3** (1986), 15–20. https://doi.org/10.1007/BF00403406

24. W. T. Trotter, Applications of the probabilistic method to partially ordered sets, In: R. L. Graham, J. Nešetřil, S. Butler, *The mathematics of Paul Erdős II*, Springer, 2013, 313–329. https://doi.org/10.1007/978-1-4614-7254-4_20

25. J. Kratochvíl, P. Savický, Z. Tuza, One more occurrence of variables makes satisfiability jump from trivial to NP-complete, *SIAM J. Comput.*, **22** (1993), 203–210. https://doi.org/10.1137/0222015

26. H. Gebauer, Disproof of the neighborhood conjecture with implications to SAT, *Combinatorica*, **32** (2012), 573–587. https://doi.org/10.1007/s00493-012-2679-y

27. H. Gebauer, T. Szabó, G. Tardos, The local lemma is asymptotically tight for SAT, *J. ACM*, **63** (2016), 43. https://doi.org/10.1145/2975386

28. B. Haeupler, B. Saha, A. Srinivasan, New constructive aspects of the Lovász local lemma, *J. ACM*, **58** (2011), 28. https://doi.org/10.1145/2049697.2049702

29. H. Gebauer, R. A. Moser, D. Scheder, E. Welzl, The Lovász local lemma and satisfiability, In: S. Albers, H. Alt, S. Näher, *Efficient algorithms*, Springer, 2009, 30–54. https://doi.org/10.1007/978-3-642-03456-5_3

30. A. Moitra, Approximate counting, the Lovász local lemma, and inference in graphical models, *J. ACM*, **66** (2019), 10. https://doi.org/10.1145/3268930

31. J. Livieratos, *Constraint satisfaction problems: probabilistic approach and applications to social choice theory*, Ph.D. Thesis, Department of Mathematics, National and Kapodistrian University of Athens, 2020.

32. Z. Chen, A. Lonkar, C. Wang, K. Yang, Y. Yin, Counting random k-SAT near the satisfiability threshold, *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, 2025, 867–878. https://doi.org/10.1145/3717823.3718163

33. A. Ambainis, J. Kempe, O. Sattath, A quantum Lovász local lemma, *J. ACM*, **59** (2012), 24. https://doi.org/10.1145/2371656.2371659

34. L. J. Schulman, Deterministic coding for interactive communication, *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, 1993, 747–756. https://doi.org/10.1145/167088.167279

35. N. Alon, M. Braverman, K. Efremenko, R. Gelles, B. Haeupler, Reliable communication over highly connected noisy networks, *Distributed Comput.*, **32** (2016), 505–515. https://doi.org/10.1007/s00446-017-0303-5

36. R. Gelles, Coding for interactive communication: a survey, *Found. Trends Theor. Comput. Sci.*, **13** (2017), 1–157. https://doi.org/10.1561/0400000079

37. F. T. Leighton, B. M. Maggs, S. B. Rao, Packet routing and job-shop scheduling in O(congestion + dilation) steps, *Combinatorica*, **14** (1994), 167–186. https://doi.org/10.1007/BF01215349

38. F. T. Leighton, B. M. Maggs, A. Richa, Fast algorithms for finding O(congestion + dilation) packet routing schedules, *Combinatorica*, **19** (1999), 375–401. https://doi.org/10.1007/s004930050061

39. K. Cheng, B. Haeupler, X. Li, A. Shahrasbi, K. Wu, Synchronization strings: highly efficient deterministic constructions over small alphabets, *Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms*, 2019, 2185–2204. https://doi.org/10.1137/1.9781611975482.132

40. P. Keevash, C. Y. Ku, A random construction for permutation codes and the covering radius, *Des. Codes Cryptography*, **41** (2006), 79–86. https://doi.org/10.1007/s10623-006-0017-3

41. R. Con, A. Shpilka, I. Tamo, Reed Solomon codes against adversarial insertions and deletions, *IEEE Trans. Inf. Theory*, **69** (2023), 2991–3000. https://doi.org/10.1109/TIT.2023.3237711

42. R. Con, A. Shpilka, I. Tamo, Corrections to "Reed Solomon codes against adversarial insertions and deletions", *IEEE Trans. Inf. Theory*, **71** (2025), 3250–3251. https://doi.org/10.1109/TIT.2025.3538114

43. L. Huang, Bounds and constructions of high-memory spatially-coupled codes, *2025 IEEE Information Theory Workshop (ITW)*, 2025. https://doi.org/10.1109/ITW62417.2025.11240539

44. L. Huang, Counting and entropy bounds for structure-avoiding spatially-coupled LDPC constructions, *2026 IEEE International Symposium on Information Theory (ISIT)*, 2026. https://doi.org/10.48550/arXiv.2601.09674

45. C. Aranda, M. Fernández, Improved existence bounds on IPP codes using the Clique Lovász local lemma, *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, 3170–3174. https://doi.org/10.1109/ISIT.2017.8007112

46. M. Fernández, G. Kabatiansky, S. Martín, C. Tavernier, A constructive approach to multimedia codes with complete traceability resistant to δ-noise, *2023 IEEE Information Theory Workshop (ITW)*, 2023, 254–259. https://doi.org/10.1109/ITW55543.2023.10161685

47. M. Fernández, J. Livieratos, S. Martín, Bounds and constructions of parent identifying schemes via the algorithmic version of the Lovász local lemma, *IEEE Trans. Inf. Theory*, **69** (2023), 7049–7069. https://doi.org/10.1109/TIT.2023.3282452

48. M. Fernández, J. Livieratos, S. Martín, Combinatorial constructions of separating codes, *J. Complexity*, **86** (2025), 101906. https://doi.org/10.1016/j.jco.2024.101906
49. M. Fernández, F. J. Martínez-Zaldívar, V. M. García-Mollá, M. A. Simarro, J. Livieratos, A. González, GPU generation of binary 2-separating codes, *J. Supercomput.*, **82** (2026), 149, https://doi.org/10.1007/s11227-026-08280-4
50. H. G. Yeh, *d*-disjunct matrices: bounds and Lovász local lemma, *Discrete Math.*, **253** (2002), 97–107. https://doi.org/10.1016/S0012-365X(01)00452-6
51. M. Dalai, S. D. Fiore, A. A. Rescigno, U. Vaccaro, An efficient algorithm for group testing with runlength constraints, *Discrete Appl. Math.*, **360** (2025), 181–187. https://doi.org/10.1016/j.dam.2024.09.001
52. A. A. Rescigno, U. Vaccaro, Bounds and algorithms for generalized superimposed codes, *Inf. Process. Lett.*, **182** (2023), 106365. https://doi.org/10.1016/j.ipl.2023.106365
53. A. A. Rescigno, U. Vaccaro, Improved algorithms and bounds for list union-free families, *IEEE Trans. Inf. Theory*, **70** (2024), 2456–2463. https://doi.org/10.1109/TIT.2023.3316435
54. L. Gargano, A. A. Rescigno, U. Vaccaro, Low-weight superimposed codes and related combinatorial structures: bounds and applications, *Theor. Comput. Sci.*, **806** (2020), 655–672. https://doi.org/10.1016/j.tcs.2019.10.032
55. A. D. Scott, A. D. Sokal, The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma, *J. Stat. Phys.*, **118** (2005), 1151–1261. https://doi.org/10.1007/s10955-004-2055-4
56. A. D. Scott, A. D. Sokal, On dependency graphs and the lattice gas, *Combin. Probab. Comput.*, **15** (2006), 253–279. https://doi.org/10.1017/S0963548305007182
57. R. Bissacot, R. Fernández, A. Procacci, B. Scoppola, An improvement of the Lovász local lemma via cluster expansion, *Combin. Probab. Comput.*, **20** (2011), 709–719. https://doi.org/10.1017/S0963548311000253
58. M. Szegedy, The Lovász local lemma–a survey, In: A. A. Bulatov, A. M. Shur, *Computer science – theory and applications*, Springer, 2013. https://doi.org/10.1007/978-3-642-38536-0_1
59. A. Faragó, A meeting point of probability, graphs, and algorithms: the Lovász local lemma and related results–a survey, *Algorithms*, **14** (2021), 355. https://doi.org/10.3390/a14120355
60. M. Grötschel, J. Nešetřil, The mathematics of László Lovász, In: H. Holden, R. Piene, *The Abel prize 2018–2022*, Springer, 2024, 535–594. https://doi.org/10.1007/978-3-031-33973-8_19
61. J. Beck, Games, randomness, and algorithms, In: R. L. Graham, J. Nešetřil, S. Butler, *The mathematics of Paul Erdős I*, Springer, 2013, 311–342. https://doi.org/10.1007/978-1-4614-7258-2_21
62. J. Kahn, On some hypergraph problems of Paul Erdős and the asymptotics of matchings, Covers and Colorings, In: R. L. Graham, J. Nešetřil, S. Butler, *The mathematics of Paul Erdős I*, Springer, 2013, 343–369. https://doi.org/10.1007/978-1-4614-7258-2_22
63. J. Beck, An algorithmic approach to the Lovász local lemma. I, *Random Structures Algorithms*, **2** (1991), 343–365. https://doi.org/10.1002/rsa.3240020402
64. N. Alon, A parallel algorithmic version of the local lemma, *Random Structures Algorithms*, **2** (1991), 367–378. https://doi.org/10.1002/rsa.3240020403

65. R. A. Moser, G. Tardos, A constructive proof of the general Lovász local lemma, *J. ACM*, **57** (2010), 11. https://doi.org/10.1145/1667053.1667060

66. R. A. Moser, A constructive proof of the Lovász local lemma, *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, 343–350. https://doi.org/10.1145/1536414.1536462

67. K. B. R. Kolipaka, M. Szegedy, Moser and Tardos meet Lovász, *Proceedings of the forty-third annual ACM symposium on Theory of computing*, 2011, 235–244. https://doi.org/10.1145/1993636.1993669

68. J. B. Shearer, On a problem of Spencer, *Combinatorica*, **5** (1985), 241–245. https://doi.org/10.1007/BF02579368

69. K. He, Q. Li, X. Sun, Moser–Tardos algorithm: beyond Shearer's bound, *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2023. https://doi.org/10.1137/1.9781611977554.ch129

70. K. He, L. Li, X. Liu, Y. Wang, M. Xia, Variable version Lovász local lemma: a tale of two boundaries, *Inf. Comput.*, **308** (2026), 105386. https://doi.org/10.1016/j.ic.2025.105386

71. T. Tao, Moser's entropy compression argument, unpublished notes, 2009. Available from: `https://page.math.tu-berlin.de/~felsner/SemWS17-18/Tao-moser-entropy-compression.pdf`.

72. L. Esperet, A. Parreau, Acyclic edge-coloring using entropy compression, *Eur. J. Combin.*, **34** (2013), 1019–1027. https://doi.org/10.1016/j.ejc.2013.02.007

73. V. Dujmović, G. Joret, J. Kozik, D. R. Wood, Nonrepetitive colouring via entropy compression, *Combinatorica*, **36** (2016), 661–686. https://doi.org/10.1007/s00493-015-3070-6

74. W. Pegden, An extension of the Moser–Tardos algorithmic local lemma, *SIAM J. Discrete Math.*, **28** (2014), 911–917. https://doi.org/10.1137/110828290

75. N. J. A. Harvey, J. Vondrák, An algorithmic proof of the Lovász local lemma via resampling oracles, *SIAM J. Comput.*, **49** (2020), 394–428. https://doi.org/10.1137/18M1167176

76. K. Chandrasekaran, N. Goyal, B. Haeupler, Deterministic algorithms for the Lovász local lemma, *SIAM J. Comput.*, **42** (2013), 2132–2155. https://doi.org/10.1137/100799642

77. D. G. Harris, Deterministic algorithms for the Lovász local lemma: simpler, more general, and more parallel, *Random Structures Algorithms*, **63** (2023), 716–752. https://doi.org/10.1002/rsa.21152

78. P. Erdős, J. Spencer, Lopsided Lovász local lemma and Latin transversals, *Discrete Appl. Math.*, **30** (1991), 151–154. https://doi.org/10.1016/0166-218X(91)90040-4

79. J. Vondrák, *Math 233: nonconstructive methods in combinatorics*, Stanford University, 2016. Available from: `https://theory.stanford.edu/~jvondrak/MATH233-2016/Math233-lec02.pdf`.

80. R. Mathon, Lower bounds for Ramsey numbers and association schemes, *J. Combin. Theory Ser. B*, **42** (1987), 122–127. https://doi.org/10.1016/0095-8956(87)90067-0

81. S. P. Radziszowski, Small Ramsey numbers, *Electron. J. Combin.*, 2026. https://doi.org/10.37236/21

82. J. Ma, W. Shen, S. Xie, An exponential improvement for Ramsey lower bounds, *Invent. Math.*, 2026. https://doi.org/10.1007/s00222-026-01421-9
83. Z. Hunter, A. Milojević, B. Sudakov, Gaussian random graphs and Ramsey numbers, *ArXiv*, 2025. https://doi.org/10.48550/arXiv.2512.17718
84. Q. Lin, L. Niu, Sharper Ramsey lower bounds from refined Gaussian estimates, *ArXiv*, 2026. https://doi.org/10.48550/arXiv.2605.25843
85. D. Bradač, Off-diagonal Ramsey numbers, *ArXiv*, 2026. https://doi.org/10.48550/arXiv.2605.28793
86. R. L. Graham, D. E. Knuth, O. Patashnik, *Concrete mathematics: a foundation for computer science*, 2 Eds., Addison–Wesley, 1989. Available form: `https://www-cs-faculty.stanford.edu/~knuth/gkp.html`.
87. B. Grünbaum, Acyclic colorings of planar graphs, *Israel J. Math.*, **14** (1973), 390–408. https://doi.org/10.1007/BF02764716
88. O. V. Borodin, On acyclic colorings of planar graphs, *Discrete Math.*, **25** (1979), 211–236. https://doi.org/10.1016/0012-365X(79)90077-3
89. J. Nešetřil, P. O. de Mendez, *Sparsity: graphs, structures, and algorithms*, Springer, 2012. https://doi.org/10.1007/978-3-642-27875-4
90. T. M. Cover, J. A. Thomas, *Elements of information theory*, 2 Eds., John Wiley & Sons, Inc., 2006. https://doi.org/10.1002/047174882X
91. I. M. Wanless, Transversals in Latin squares: a survey, In: R. Chapman, *Surveys in combinatorics 2011*, Cambridge University Press, 2011, 403–437. https://doi.org/10.1017/CBO9781139004114.010
92. J. Böttcher, Y. Kohayakawa, A. Procacci, Properly coloured copies and rainbow copies of large graphs with small maximum degree, *Random Structures Algorithms*, **40** (2012), 425–436. https://doi.org/10.1002/rsa.20383
93. S. Ndreca, A. Procacci, B. Scoppola, Improved bounds on coloring of graphs, *Eur. J. Combin.*, **33** (2012), 592–609. https://doi.org/10.1016/j.ejc.2011.12.002
94. D. G. Harris, A. Srinivasan, A constructive algorithm for the Lovász local lemma on permutations, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2014)*, 2014, 907–925. https://doi.org/10.1137/1.9781611973402.68
95. V. Rödl, On a packing and covering problem, *Eur. J. Combin.*, **6** (1985), 69–78. https://doi.org/10.1016/S0195-6698(85)80023-8
96. M. Ajtai, J. Komlós, J. Pintz, J. Spencer, E. Szemerédi, Extremal uncrowded hypergraphs, *J. Combin. Theory Ser. A*, **32** (1982), 321–335. https://doi.org/10.1016/0097-3165(82)90049-8
97. A. Johansson, Asymptotic choice number for triangle-free graphs, *Technical Report*, 1996, 91–95.
98. M. Molloy, The list chromatic number of graphs with small clique number, *J. Combin. Theory Ser. B*, **134** (2019), 264–284. https://doi.org/10.1016/j.jctb.2018.06.007
99. S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities: a nonasymptotic theory of independence*, Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

100. M. Raginsky, I. Sason, Concentration of measure inequalities in information theory, communications, and coding, *Found. Trends Commun. Inf. Theory*, **10** (2013), 1–247. https://doi.org/10.1561/0100000064

101. A. Bernshteyn, Measurable versions of the Lovász local lemma and measurable graph colorings, *Adv. Math.*, **353** (2019), 153–223. https://doi.org/10.1016/j.aim.2019.06.031

Appendix

A. Completion of the proof of Corollary 3.4

Lemma A.1. For every $d \in \mathbb{N}$,

$$\frac{d^d}{(d+1)^{d+1}} > \frac{1}{e(d+\frac{1}{2})}. \tag{A.1}$$

Moreover, $\alpha = \frac{1}{2}$ is the smallest constant such that

$$\frac{d^d}{(d+1)^{d+1}} \geq \frac{1}{e(d+\alpha)}, \quad \forall\, d \in \mathbb{N} \tag{A.2}$$

holds.

Proof. Define $\phi(x) := x\ln x - (x+1)\ln(x+1) + 1 + \ln(x+\frac{1}{2})$ for $x > 0$. Then, for each $x > 0$, the inequality $\phi(x) > 0$ is equivalent to $\frac{x^x}{(x+1)^{x+1}} > \frac{1}{e\left(x+\frac{1}{2}\right)}$. Therefore, to prove (A.1), it suffices to show that $\phi(x) > 0$ for all $x > 0$. Differentiating and setting $t := \frac{1}{x} > 0$, we obtain

$$\phi'(x) = -\ln(1+t) + \frac{2t}{2+t}.$$

To show that $\phi'(x) < 0$, consider the auxiliary function

$$f(t) := \ln(1+t) - \frac{2t}{2+t}$$

for all $t \geq 0$. Since $f(0) = 0$ and

$$f'(t) = \frac{t^2}{(1+t)(2+t)^2} > 0$$

for all $t > 0$, the function f is strictly increasing on $[0, \infty)$. Hence $f(t) > 0$ for all $t > 0$, which proves that $\phi'(x) < 0$ for all $x > 0$. Therefore, ϕ is strictly decreasing on $(0, \infty)$, and

$$\lim_{x\to\infty} \phi(x) = \lim_{x\to\infty} \left\{ \ln\left(1 + \frac{1}{2x}\right) - (x+1)\ln\left(1+\frac{1}{x}\right) + 1 \right\} = 0,$$

where the second limit holds by the standard expansion $\ln(1+u) = u + O(u^2)$. Since ϕ is strictly decreasing and tends to 0 at infinity, it follows that $\phi(x) > 0$ for all $x > 0$. In particular, inequality (A.1) holds for all $d \in \mathbb{N}$.

It remains to prove that the constant $\alpha = \frac{1}{2}$ is the smallest value for which inequality (A.2) holds. Suppose that (A.2) holds for some $\alpha > 0$. Then,

$$\alpha \geq \frac{d+1}{e}\left(1+\frac{1}{d}\right)^d - d, \quad \forall\, d \in \mathbb{N}.$$

Using the asymptotic expansion as $d \to \infty$

$$\left(1+\frac{1}{d}\right)^d = e\left(1-\frac{1}{2d}+O\left(\frac{1}{d^2}\right)\right),$$

it follows that $\alpha \geq \frac{1}{2}$. Since (A.1) shows that $\alpha = \frac{1}{2}$ is admissible, it is the smallest admissible value. □

B. Derivation of (4.17)

Let

$$a := -\ln(1-\varepsilon) > 0, \qquad C := \sqrt{\frac{3}{8\pi}}\, e^3. \tag{B.1}$$

Then, inequality (4.16) can be rewritten as

$$-\frac{2a}{3}\, k\, e^{-\frac{2a}{3}k} \geq -\frac{2a}{3}\left(\frac{e^{-2a}}{C}\right)^{2/3}.$$

Applying the branch W_{-1} of the Lambert W function to both sides, we obtain

$$k \geq -\frac{3}{2a}\, W_{-1}\left(-\frac{2a}{3}\left(\frac{e^{-2a}}{C}\right)^{2/3}\right). \tag{B.2}$$

The branch W_{-1} is used rather than the principal branch W_0 since the argument of the Lambert W function is negative and tends to 0 as $\varepsilon \to 0$. The branch W_0 would yield a bounded solution for k, whereas the desired solution corresponds to the large-k regime; this is captured by the branch W_{-1}, for which $W_{-1}(x) \to -\infty$ as $x \to 0^-$. Substituting the values of a and C from (B.1) into (B.2) gives

$$k \geq \left(\frac{3}{2\ln(1-\varepsilon)}\right) W_{-1}\left(\frac{4}{3e^2}\sqrt[3]{\frac{\pi}{3}}\, \ln(1-\varepsilon)\,(1-\varepsilon)^{4/3}\right). \tag{B.3}$$

Combining inequality (B.3) with the requirement $k \geq 3$ in (4.15) finally gives, together with (4.18), the valid selection of the integer k_0 in (4.17).