Expected Clique Counts and the Probability of Monochromatic Cliques

Jennifer Lee, Shaunak Chittimalla, Malindu Talagala, Sahithya Kalisetti, Ayaan Kayastha

October 2026

Abstract

Every red–blue coloring of the edges of a complete graph on 18 vertices contains four vertices with all six edges between them the same color, called a monochromatic K_4. Yet some colorings on 17 vertices contain none. How much smaller can the graph be if a monochromatic K_4 is required in more than 90% of random colorings rather than in every coloring? This study estimates the analogous 90% thresholds for clique orders 3–7 and measures the number of monochromatic cliques for orders 3–5. The first tested sizes with observed proportions above 90% were 5, 9, 15, 24, and 36, respectively. Sample mean counts closely matched the exact expectations. The study also compares random search with a Boolean satisfiability (SAT) solver for finding colorings without a monochromatic K_4. At each size from 13 through 17, none of twenty random searches of five seconds each succeeded, while one SAT solve found a verified avoiding coloring in under one second. These results describe the tested implementation and budgets; they do not establish exact probability thresholds or new Ramsey bounds.

1 Introduction

The diagonal Ramsey number $R(k, k)$ is the smallest n such that every red–blue coloring of the edges of the complete graph K_n contains a monochromatic K_k: a set of k vertices whose connecting edges all have one color. Table 1 gives established values and bounds for the clique orders studied here [1]. The guarantee for every coloring can require a much larger graph than the appearance of a clique in

Table 1: Known diagonal Ramsey values and bounds from the April 24, 2026 survey [1]. Intervals are bounds, not exact values.

k	$R(k, k)$	Status
3	6	Exact
4	18	Exact
5	$43 \leq R(5, 5) \leq 46$	Exact value unknown
6	$102 \leq R(6, 6) \leq 160$	Exact value unknown
7	$205 \leq R(7, 7) \leq 492$	Exact value unknown

most random colorings. For example, although $R(4, 4) = 18$, more than 90% of random colorings may already contain a monochromatic K_4 for some $n < 18$. This study estimates the smallest such n. It also counts how many monochromatic K_3, K_4, and K_5 subgraphs occur, and tests whether SAT search finds avoiding colorings more effectively than repeated random sampling.

Three research questions guide the experiments:

1. For monochromatic K_3, K_4, and K_5 counts, how do the sampled means compare with the exact expectations as n increases? For K_3 and K_4, how do the count distributions and modes change?
2. For each $3 \leq k \leq 7$, what is the smallest n for which more than 90% of all red–blue edge colorings of K_n contain a monochromatic K_k? How do the empirical estimates compare with the known Ramsey values or bounds and a simple mathematical approximation?
3. To find a coloring of K_n with no monochromatic K_4, is repeated random sampling or a SAT solver more effective under a five-second allowance per run, for $10 \leq n \leq 17$?

The experiments estimate the second question using 3,000 independently sampled colorings per tested n. An observed proportion above 90% is evidence about the population probability, not a proof of its exact threshold. The contribution is the combined empirical dataset, reproducible workflow, and comparison under a stated computational budget. The expectation formula, Ramsey values, and SAT encoding are established benchmarks.

The experimental questions were motivated by Gasarch's Ramsey project notes [2].

2 Mathematical Background

The complete graph K_n has n vertices and an edge between every pair. In a random two-coloring, each of its $\binom{n}{2}$ edges is colored red or blue independently with equal probability. Let $X_{n,k}$ count the k-vertex subsets whose induced edges all have the same color. The diagonal Ramsey number $R(k,k)$ is the smallest n such that every red–blue coloring of K_n contains at least one monochromatic K_k. It is a worst-case guarantee, whereas $P(X_{n,k} > 0)$ describes a random coloring.

The identity $R(4,4) = 18$ means every red–blue coloring of K_{18} contains a monochromatic K_4. There can nevertheless be an $n < 18$ for which more than 90% of all colorings contain one. Write $p_k(n) = P(X_{n,k} > 0)$, and define $n_{0.9}(k)$ as the smallest n with $p_k(n) > 0.9$. The experiments estimate $n_{0.9}(k)$, rather than $R(k,k)$. Independent fair coloring of every edge makes all edge colorings equally likely.

Theorem 1. *For independent fair red–blue edge colorings and integers $n \geq k \geq 2$, $E[X_{n,k}] = \binom{n}{k} 2^{1-\binom{k}{2}}$*

Proof. For each k-vertex subset S, let I_S be 1 when all its edges have the same color and 0 otherwise. Exactly two of the $2^{\binom{k}{2}}$ equally likely assignments are monochromatic, so $E[I_S] = 2^{1-\binom{k}{2}}$. Summing over the $\binom{n}{k}$ subsets and using linearity of expectation proves the formula, without requiring independence between clique indicators. □

Corollary 1.1. *For fixed k, $E[X_{n,k}] \sim Cn^k/k!$ as $n \to \infty$, where $C = 2^{1-\binom{k}{2}}$.*

Proof. By Stirling's approximation, $\binom{n}{k} \sim n^k/k!$ as $n \to \infty$, so setting $C = 2^{1-\binom{k}{2}}$, $E[X_{n,k}] = \binom{n}{k} 2^{1-\binom{k}{2}} \sim Cn^k/k!$ as $n \to \infty$. □

Remark 1. Since $E[X_{n,k}] \sim Cn^k/k!$ as $n \to \infty$, we can approximate the mean as a degree-k polynomial in n, reflected as the leading terms in Table 2.

Table 2: Exact expected counts and asymptotic leading terms from Theorem 1.

Clique	Exact Expectation	Leading term
K_3	$\binom{n}{3}/4$	$n^3/24$
K_4	$\binom{n}{4}/32$	$n^4/768$
K_5	$\binom{n}{5}/512$	$n^5/61440$

The first-moment bound gives $P(X_{n,k} > 0) \leq \min(1, E[X_{n,k}])$. If $E[X_{n,k}] < 1$, an avoiding coloring exists and $R(k,k) > n$. This is the probabilistic argument behind Erdős's exponential lower bound $R(k,k) > 2^{(k/2)}$ for $k \geq 3$ [6]. When $E[X_{n,k}] \geq 1$, this bound does not locate the 90% threshold. For a simple comparison, suppose heuristically that $X_{n,k}$ were Poisson with mean $\mu = E[X_{n,k}]$. Its presence probability would be $1 - \exp(-\mu)$, exceeding 0.9 when $\mu > \ln(10)$. Define $n^p_{0.9}(k)$ as the first integer n satisfying this inequality. Table 4 compares this computed benchmark with the empirical estimates. Overlapping cliques create dependence, so this Poisson calculation is not an exact probability formula or a bound on the true threshold.

Goodman's theorem determines the exact minimum number of monochromatic triangles [3]: $n(n-2)(n-4)/24$ for even n; $n(n-1)(n-5)/24$ for $n \equiv 1 \pmod 4$; and $(n+1)(n-3)(n-4)/24$ for $n \equiv 3 \pmod 4$. Thus the proven minima at $n = 5$, 6, and 10 are 0, 2, and 20. These mathematical minima are distinct from minima observed in a finite sample; the same inference cannot be made for K_4 from these trials.

3 Methods and Reproductibility

The full profile generated 2,000 independent colorings at each n for the K_3 distribution $5 \le n \le 10$ and K_4 distribution $10 \le n \le 25$. It recorded the full count frequency table, all tied modes, sample minima and maxima, and means. Separate growth sweeps used 2,000 colorings per n, starting at $n = 5$, with a 600-second target per n, a maximum n of 100, and a 10-million-entry cap on the precomputed clique-edge index. Probability searches used 3,000 colorings per n and a 300-second target per clique order, stopping at the first sampled proportion at least 0.9 or a cap. Avoidance searches used 50,000 colorings per n for K_3 at $4 \le n \le 6$ and K_4 at $8 \le n \le 17$. Independent pseudorandom streams were derived from seed 20260928 and the experiment, k, and n; the code and per-experiment metadata specify the formulas. The implementation is available in the project repository [4].

For SAT, each edge e is represented by Boolean x_e. For each four-vertex subset, one six-literal clause excludes all-red and another excludes all-blue. The CNF has $\binom{n}{2}$ variables and $2 \cdot \binom{n}{4}$ clauses [5]. A Minisat22 solve ran once per $10 \le n \le 17$, while uniform random search ran twenty independently seeded repetitions per n. Each run received a five-second wall-clock allowance. SAT wall time includes process startup and CNF construction. All 57 saved SAT/random benchmark witnesses and six Monte Carlo witnesses were checked again with an independent clique detector; a saved coloring certifies existence at its n. The uploaded run used Python 3.13.2, NumPy 2.5.3, and PySAT 1.9.dev15; the separate K_5 supplement used Python 3.12.14 and NumPy 2.3.5. Machine specifications were not captured, so measured wall times are not portable performance estimates.

For T independent coloring trials with count standard deviation σ, the estimated standard error of the sample mean is $\sigma/\sqrt{T}$. The distribution frequency tables retain enough information to calculate this quantity, reported below for selected settings. The archived growth summaries do not retain second moments or individual counts, so standard errors for the large-n growth curves cannot be reconstructed from those summaries. Their observed ranges are not confidence intervals. Future growth runs should save trial counts or count sums and sums of squares.

4 Count Distributions and Growth

The mean-count experiments support the exact formula in Theorem 1 across the tested ranges. Figure 1 shows the predicted cubic, quartic, and quintic growth alongside the sampled means. The mode in Table 3 is the most frequent number of monochromatic cliques in one coloring, not the number of trials attaining that count.

Table 3: Selected count distributions; 2,000 independent colorings at each point. Modes can tie, especially when counts spread over many values. The complete distributions and their plots are in the run archive.

Clique	n	Mode(s)	Sample mean	$E[X]$	Sample min-max	Standard error
K_3	5	3	2.454	2.500	0-10	0.031
K_3	6	4	4.989	5.000	2-16	0.043
K_3	10	31	30.131	30.000	20-55	0.107

Clique	n	Mode(s)	Sample mean	$E[X]$	Sample min-max	Standard error
K_4	10	4	6.646	6.562	0-37	0.102
K_4	13	17	22.311	22.344	5-83	0.215
K_4	17	61, 66, 71	74.447	74.375	33-211	0.462
K_4	25	361	397.713	395.312	259-729	1.360

For the triangle rows, the observed minima 0, 2, and 20 coincide with Goodman's proven minima at $n = 5$, 6, and 10. This agreement checks the implementation against a theorem; it does not derive the theorem from sampling. At $n = 100$, the K_3 mean was $40,428.459$ versus exact expectation $40,425$ (ratio 1.00009). At $n = 60$, the K_4 mean was $15,241.343$ versus $15,238.594$ (ratio 1.00018). The uploaded K_5 sweep saved only $5 \le n \le 8$; its final batch reports $23,889.717$ seconds, far beyond the intended 600 seconds. Its cause cannot be established from the archived summary. To fill this gap without attributing new data to the uploaded run, a separate seeded rerun completed 2,000 colorings at each $5 \le n \le 30$.

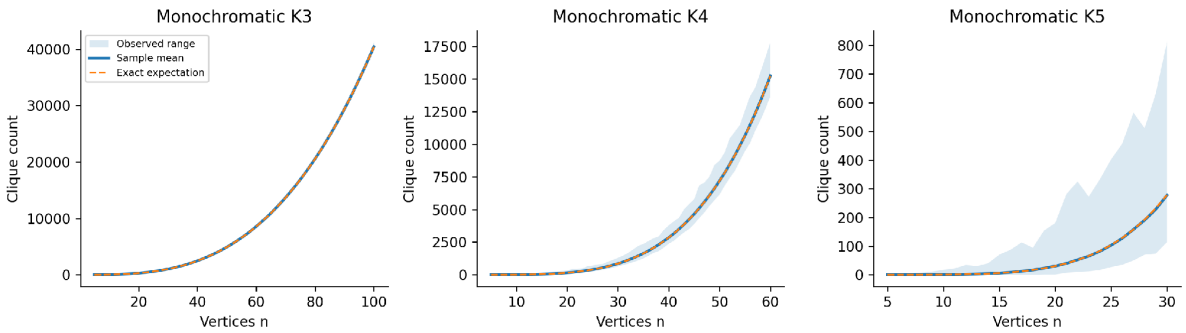


Figure 1: Sample means, observed ranges, and exact expectations for K_3 ($5 \le n \le 100$) and K_4 ($5 \le n \le 60$) from the uploaded run, plus K_5 ($5 \le n \le 30$) from a separate targeted rerun. Each point has 2,000 complete trials; the K_5 data used a different environment.

At $n = 30$ its mean was 277.281 versus exact expectation 278.332 (ratio 0.9962). Agreement with the known formula is an implementation check, not a newly established quintic coefficient.

The growth is consistent with the fixed-k asymptotic formula in Section 2. The reported mean-to-expectation ratios describe agreement but do not, by themselves, show whether each deviation is statistically distinguishable from sampling noise. That assessment requires the standard errors unavailable for these archived growth batches.

5 Sizes at Which More Than 90 Percent Contain a Clique

The estimated smallest sizes are $n = 5$, 9, 15, 24, and 36 for $k = 3$, 4, 5, 6, and 7, respectively. At the preceding tested size, each observed proportion was below 90%; at the reported size, it was above 90% (Table 4). These sizes are below $R(3, 3) = 6$ and $R(4, 4) = 18$, and below the known lower bounds 43, 102, and 205 for the remaining Ramsey numbers in Table 1.

The Poisson benchmark gives 5, 9, 13, 20, and 31. It matches the observed threshold estimates for $k = 3$ and 4 but gives smaller sizes by 2, 4, and 5 vertices for $k = 5$, 6, and 7. Thus the exact mean alone, combined with a Poisson assumption, does not accurately reproduce all observed 90% thresholds. The table compares an approximation with finite-sample estimates; neither column proves the exact population threshold. For K_3 at $n = 5$, the separate avoidance search found 520 zero-count colorings in

Table 4: Estimated 90% thresholds and the Poisson benchmark defined in Section 2. Each observed proportion uses 3,000 colorings. "Previous n" is the preceding tested size; "estimated n" is the first tested size with observed proportion above 0.9. These are empirical estimates of $n_{0.9}(k)$.

Clique	Previous n	Observed fraction	Estimated n	Observed fraction	Poisson n
K_3	4	2156/3000 (0.719)	5	2970/3000 (0.990)	5
K_4	8	2372/3000 (0.791)	9	2820/3000 (0.940)	9
K_5	140	2532/3000 (0.844)	15	2788/3000 (0.929)	13
K_6	23	2603/3000 (0.868)	24	2787/3000 (0.929)	20
K_7	35	2563/3000 (0.854)	36	2756/3000 (0.919)	31

50,000 trials (1.04%); exhaustive enumeration of the 2^{10} possible K_5 colorings gives $12/1024 = 1.171875\%$. For K_4, the avoidance search found 27/50,000 at $n = 11$ and none in 50,000 at each $12 \le n \le 17$. Zero observations do not imply nonexistence: if the true zero probability is q, the chance of missing it in T independent trials is $(1 - q)^T$. Indeed, verified SAT models below show K_4-avoiding colorings through $n = 17$, consistent with the established $R(4, 4) = 18$ [1].

6 Finding Avoiding Colorings with Random Search and SAT

A successful search returns one complete red–blue edge coloring with no monochromatic K_4. SAT means Boolean satisfiability: the solver searches for edge-color assignments satisfying clauses that forbid an all-

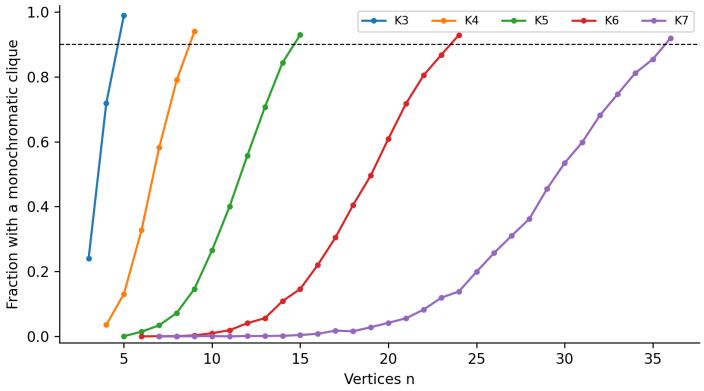


Figure 2: Sampled presence probability across n for $K_3, K_4, \ldots K_7$. The horizontal line marks 0.9; full trial counts and n values are in thresholds.csv.

red or all-blue K_4. Under the five-second allowance, SAT found a verified avoiding coloring at every tested size $10 \leq n \leq 17$. Random search succeeded in all twenty runs at $n = 10$, 11; in nine runs at $n = 12$; and in none at $13 \leq n \leq 17$. SAT was therefore more effective at finding an avoiding coloring in these tested runs, especially at $13 \leq n \leq 17$; the experiment does not measure a general solver success probability or establish a ranking across other instances. At $n = 12$, uniform sampling succeeded in

Table 5: Searches for a coloring with no monochromatic K_4. Each random-search entry is the number of successful runs out of twenty, with five seconds per run. Each SAT entry records one solve under the same allowance; every returned coloring was independently verified. SAT time includes setup. A random-search failure means no witness was found within the allowance, not that no avoiding coloring exists.

n	Random searches successful/20	SAT solve successful/1	SAT wall time (seconds)
10	20/20	1/1 verified	0.326
11	20/20	1/1 verified	0.337
12	9/20	1/1 verified	0.354
13	0/20	1/1 verified	0.345
14	0/20	1/1 verified	0.532
15	0/20	1/1 verified	0.487
16	0/20	1/1 verified	0.435
17	0/20	1/1 verified	0.326

9/20 runs; at $13 \leq n \leq 17$ it succeeded in 0/20 per n, despite hundreds of thousands of attempted colorings at each size. A 0/20 outcome has a Wilson 95% upper limit of approximately 0.161 for the chance that this specified five-second random-search procedure succeeds in one independent repetition. It says nothing directly about the fraction of all edge colorings that avoid K_4. At $n = 13$, 78 variables and 1,430 clauses were generated. Minisat22, Glucose3, and Glucose4 each returned a verified model in a single five-second survey run, with respective wall times 0.567, 0.352, and 0.337 seconds. One run per deterministic solver cannot rank their general performance [6].

7 Interpretation and Limits

The experiments give three answers. First, sampled mean clique counts closely follow the exact expectation in Theorem 1 over the completed growth ranges. Second, the estimated 90% thresholds for $3 \leq k \leq 7$ are 5, 9, 15, 24, and 36, respectively; the simple Poisson benchmark agrees for $k = 3, 4$ and underestimates the empirical sizes for $5 \leq k \leq 7$. Third, SAT found verified K_4-avoiding colorings at every $10 \leq n \leq 17$ under the five-second allowance, while random search failed in all twenty runs at each $13 \leq n \leq 17$. This supports using SAT for finding witnesses in the tested setting. Avoiding colorings through $n = 17$ and the SAT encoding are already known [1, 5]. Broader performance claims would require recorded hardware, repeated solver trials, multiple budgets, independent environments, and more search baselines. A timeout or random failure does not prove nonexistence; an unsatisfiability claim requires suitable independently checkable evidence.

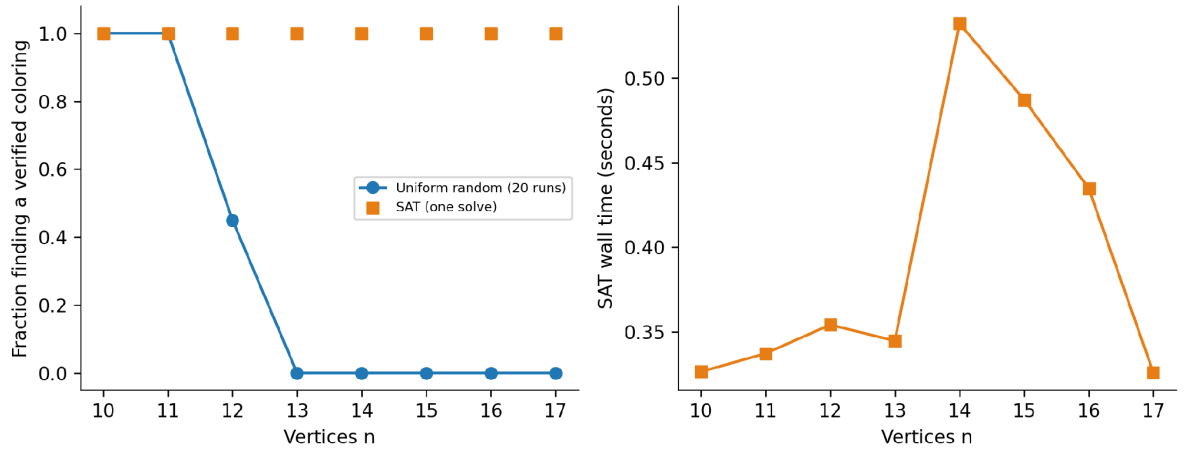


Figure 3: Left: fraction of twenty random searches that succeeded, alongside the successful outcome of one SAT solve at each n. Right: SAT elapsed time including setup. Every search seeks a coloring without a monochromatic K_4; SAT markers show individual outcomes, not estimated success probabilities.

The run archive contains CSV summaries, metadata, DIMACS CNF, figures, and verified witnesses. It does not include every individual growth-trial count; exact reruns rely on the public code, seed rules, and recorded software versions. The K_5 supplement fills the plotted range but came from a different Python/NumPy environment than the uploaded run; runtime comparisons across them would be inappropriate. Conclusions in this paper are limited to the completed experiments and their stated budgets.

For future timed benchmarks, record the processor model and core count, installed RAM, operating system, package versions, thread settings, and relevant concurrent workload. Hardware and load influence wall time and the number of attempts completed within a budget. Distribution estimates also retain sampling uncertainty even when the same code is used.

8 Code and Data Availability

The project repository is `https://github.com/jenniferlilli/ramsey-repo` [4]. The supplied repository package contains the source code, recorded full-run data, the separately identified K_5 supplement, and figure-generation scripts. Run commands from the repository root after installing the dependencies. On Windows:

```
py -m pip install -r requirements.txt
py run_all.py --profile quick --out fresh-quick
py run_all.py --profile full --out fresh-full
```

The quick profile checks the pipeline with fewer trials; the full profile regenerates the experiments from scratch and produces data and plots, subject to the documented time and size limits. Fresh outputs may differ from the manuscript, especially when time limits or software versions differ. To redraw the three manuscript figures from the archived data without rerunning trials:

```
py plot_full_report.py --input study-data/full-run --k5-csv
study-data/k5-supplement/growth_k5.csv --output report-figures-new
```

A fresh full run may stop the K_5 sweep before $n = 30$; the manuscript's K_5 curve comes from the separately archived supplement. New results should be accompanied by updated tables, captions, and numerical discussion.

Acknowledgements

We are deeply grateful to Professor William Gasarch for his invaluable mentorship, guidance, and feedback throughout this project.

References

[1] S. P. Radziszowski. "Small Ramsey Numbers". In: *Electronic Journal of Combinatorics* DS1 (2026). Dynamic Survey DS1, version April 24, 2026. URL: https://www.combinatorics.org/ojs/index.php/eljc/article/view/DS1.

[2] W. Gasarch. *HS PROJ 2026 New Ramsey*. University of Maryland. 2026. URL: https://www.cs.umd.edu/~gasarch/HSPROJ26/NRAMSEY/nramsey.html.

[3] J. Cummings et al. *Monochromatic triangles in three-coloured graphs*. Theorem 1 (Goodman's theorem). 2012. arXiv: 1206.1987. URL: https://www.math.cmu.edu/users/jcumming/papers/monotri.pdf.

[4] Project authors. *Random Ramsey Theory computational repository*. GitHub. URL: https://github.com/jenniferlilli/ramsey-repo.

[5] B. Molnár et al. "A continuous-time MaxSAT solver with high analog performance". In: *Nature Communications* 9 (2018), p. 4864. DOI: 10.1038/s41467-018-07327-2. URL: https://doi.org/10.1038/s41467-018-07327-2.

[6] A. Ignatiev, A. Morgado, and J. Marques-Silva. "PySAT: A Python Toolkit for Prototyping with SAT Oracles". In: *Theory and Applications of Satisfiability Testing – SAT 2018*. 2018. URL: https://pysathq.github.io/.