Maximizing the number of q-colorings

Po-Shen Loh, Oleg Pikhurko and Benny Sudakov

Abstract

Let $P_G(q)$ denote the number of proper q-colorings of a graph G. This function, called the *chromatic polynomial* of G, was introduced by Birkhoff in 1912, who sought to attack the famous four-color problem by minimizing $P_G(4)$ over all planar graphs G. Since then, motivated by a variety of applications, much research was done on minimizing or maximizing $P_G(q)$ over various families of graphs. In this paper, we study an old problem of Linial and Wilf, to find the graphs with n vertices and m edges which maximize the number of q-colorings. We provide the first approach that enables one to solve this problem for many nontrivial ranges of parameters. Using our machinery, we show that for each $q \geqslant 4$ and sufficiently large $m < \kappa_q n^2$, where $\kappa_q \approx 1/(q \log q)$, the extremal graphs are complete bipartite graphs minus the edges of a star, plus isolated vertices. Moreover, for $q = 3$, we establish the structure of optimal graphs for all large $m \leqslant n^2/4$, confirming (in a stronger form) a conjecture of Lazebnik from 1989.

1. *Introduction*

The fundamental combinatorial problem of graph coloring is as ancient as the cartographer's task of coloring a map without using the same color on neighboring regions. In the context of general graphs, we say that an assignment of a color to every vertex is a *proper coloring* if no two adjacent vertices receive the same color, and we say that a graph is *q-colorable* if it has a proper coloring using only at most q different colors.

The problem of counting the number $P_G(q)$ of q-colorings of a given graph G has been the focus of much research over the past century. Although it is already NP-hard even to determine whether this number is nonzero, the function $P_G(q)$ itself has very interesting properties. It was first introduced by Birkhoff [6], who proved that $P_G(q)$ is always a polynomial in q. It is now called the *chromatic polynomial* of G. Although $P_G(q)$ has been studied for its own sake (for example, Whitney [34] expressed its coefficients in terms of graph theoretic parameters), perhaps more interestingly there is a long history of diverse applications which has led researchers to minimize or maximize $P_G(q)$ over various families of graphs. In fact, Birkhoff's original motivation for investigating the chromatic polynomial was to use it to attack the famous four-color theorem. Indeed, one way to show that every planar graph is four-colorable is to minimize $P_G(4)$ over all planar G, and show that the minimum is nonzero. In this direction Birkhoff [7] proved the tight lower bound $P_G(q) \geqslant q(q-1)(q-2)(q-3)^{n-3}$ for all n-vertex planar graphs G when $q \geqslant 5$, later conjecturing with Lewis in [8] that it extended to $q = 4$ as well.

Linial [22] arrived at the problem of minimizing the chromatic polynomial from a completely different motivation. The worst-case computational complexity of determining whether a particular function $f : V(G) \to \mathbb{R}$ is a proper coloring (that is, satisfies $f(x) \neq f(y)$ for every

pair of adjacent vertices x and y) is closely related to the number of *acyclic orientations* of a graph, which equals $|P_G(-1)|$, obtained by substituting $q = -1$ into the formal polynomial expression of $P_G(q)$. Lower bounding the worst-case complexity, therefore, corresponds to minimizing $|P_G(-1)|$ over the family $\mathcal{F}_{n,m}$ of graphs with n vertices and m edges. Linial showed that surprisingly, for any n, m there is a graph which *simultaneously* minimizes each $|P_G(q)|$ over $\mathcal{F}_{n,m}$, for *every* integer q. This graph is simply a clique K_k with an additional vertex adjacent to l vertices of the K_k, plus $n - k - 1$ isolated vertices, where k, l are the unique integers satisfying $m = \binom{k}{2} + l$ with $k > l \geqslant 0$. At the end of his paper, Linial posed the problem of maximizing $P_G(q)$ over all graphs in $\mathcal{F}_{n,m}$.

Around the same time, Wilf arrived at exactly that maximization problem while analyzing the *backtrack* algorithm for finding a proper q-coloring of a graph (see [**5**, **35**]). Although this generated much interest in the problem, it was only solved in sporadic cases. The special case $q = 2$ was completely solved for all m, n, by Lazebnik [**18**]. For $q \geqslant 3$, the only nontrivial pairs m, n for which extremal graphs were known corresponded to the number of vertices and edges in the Turán graph $T_r(n)$, which is the complete r-partite graph on n vertices with all parts of size either $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$. In this vein, Lazebnik [**20**] proved that $T_r(n)$ is optimal for very large $q = \Omega(n^6)$, and proved with Pikhurko and Woldar [**21**] that $T_2(2k)$ is optimal when $q = 3$ and asymptotically optimal when $q = 4$.

Outside these isolated cases, very little was known for general m, n. Although many upper and lower bounds for $P_G(q)$ were proved by various researchers [**10**, **18**, **19**, **23**], these bounds were widely separated. Even the $q = 3$ case resisted solution: twenty years ago, Lazebnik [**18**] conjectured that when $m \leqslant n^2/4$, the n-vertex graphs with m edges which maximized the number of three-colorings were complete bipartite graphs minus the edges of a star, plus isolated vertices. Only very recently, Simonelli [**24**] managed to make some progress on this conjecture, verifying it under the additional very strong assumption that all optimal graphs are already bipartite.

Perhaps part of the difficulty for general m, n, q stems from the fact that the maximal graphs are substantially more complicated than the minimal graphs that Linial found. For number-theoretic reasons, it is essentially impossible to explicitly construct maximal graphs for general m, n. Furthermore, even their coarse structure depends on the density m/n^2. For example, when m/n^2 is small, the maximal graphs are roughly complete bipartite graphs, but after $m/n^2 > \frac{1}{4}$, the maximal graphs become tripartite. At the most extreme density, when m, n correspond to the Turán graph $T_q(n)$, the unique maximal graph is obviously the complete q-partite graph. Therefore, in order to tackle the general case of this problem, one must devise a unified approach that can handle all of the outcomes.

In this paper, we propose such an approach, developing the machinery that one might be able to use to determine the maximal graphs in many nontrivial ranges of m, n. Our methodology can be roughly outlined as follows. We show, via Szemerédi's regularity lemma, that the asymptotic solution to the problem reduces to a certain quadratically constrained linear program in $2^q - 1$ variables. For any given q, this task can, in principle, be automated by a computer code that symbolically solves the optimization problem, although a more sophisticated approach was required to solve this for all q. Our solutions to the optimization problem then give us the approximate structure of the maximal graphs. Finally, we use various local arguments, such as the so-called 'stability' approach introduced by Simonovits [**25**], to refine their structure into precise results.

We successfully applied our machinery to solve the Linial–Wilf problem for many nontrivial ranges of m, n, and $q \geqslant 3$. In particular, for $q = 3$, our results confirm a stronger form of Lazebnik's conjecture when m is large. In addition, for each $q \geqslant 4$ we show that for all densities m/n^2 up to approximately $1/(q \log q)$, the extremal graphs are also complete bipartite graphs minus a star. In order to state our results precisely, we need the following definition.

DEFINITION 1.1. Let $a \leqslant b$ be positive integers. We say that G is a *semi-complete subgraph of $K_{a,b}$* if the number of missing edges $E(K_{a,b}) \setminus E(G)$ is less than a, and they form a star (that is, they share a common endpoint v which we call the *center*). If v belongs to the larger side of $K_{a,b}$, then we also say that G is *correctly oriented*.

Define the constant

$$\kappa_q = \left(\sqrt{\frac{\log(q/(q-1))}{\log q}} + \sqrt{\frac{\log q}{\log(q/(q-1))}} \right)^{-2} \approx \frac{1}{q \log q}.$$

All logarithms here and in the rest of the paper are in base $e \approx 2.718$. In the following theorems, we write $o(1)$ to represent a quantity that tends to zero as $m, n \to \infty$.

THEOREM 1.2. *For every fixed integer $q \geqslant 3$, and any $\kappa < \kappa_q$, the following holds for all sufficiently large m with $m \leqslant \kappa n^2$. Every n-vertex graph with m edges which maximizes the number of q-colorings is a semi-complete subgraph (correctly oriented if $q \geqslant 4$) of some $K_{a,b}$, plus isolated vertices, where $a = (1 + o(1))\sqrt{m \cdot \log \frac{q}{q-1} / \log q}$ and $b = (1 + o(1))\sqrt{m \cdot \log q / \log \frac{q}{q-1}}$. The corresponding number of q-colorings is $q^n e^{(-c+o(1))\sqrt{m}}$, where $c = 2\sqrt{\log \frac{q}{q-1} \cdot \log q}$.*

REMARK. The part sizes of the maximal graphs above all have the ratio roughly $\log q / \log \frac{q}{q-1}$. The constant κ_q corresponds to the density m/n^2 at which the number of isolated vertices becomes $o(n)$ in the optimal construction.

For three colors, we can push our argument further, beyond the density κ_3. Now, due to the absence of isolated vertices, a rare exception occurs, which requires us to include an additional possibility. Here, a 'pendant edge' means that a new vertex is added, along with a single edge between it and any other vertex in the graph. Proposition B.1 shows that this outcome is in fact necessary.

THEOREM 1.3. *The following holds for all sufficiently large $m \leqslant n^2/4$. Every n-vertex graph with m edges and the maximum number of three-colorings is either* (i) *a semi-complete subgraph of some $K_{a,b}$, plus isolated vertices if necessary, or* (ii) *a complete bipartite graph $K_{a,b}$ plus a pendant edge. Furthermore,*

(i) *If $m \leqslant \kappa_3 n^2$, then $a = (1 + o(1))\sqrt{m \cdot \log \frac{3}{2} / \log 3}$ and $b = (1 + o(1)) \sqrt{m \cdot \log 3 / \log \frac{3}{2}}$. The corresponding number of colorings is $3^n e^{-(c+o(1))\sqrt{m}}$, where $c = 2\sqrt{\log \frac{3}{2} \cdot \log 3}$.*

(ii) *If $\kappa_3 n^2 \leqslant m \leqslant \frac{1}{4}n^2$, then $a = (\frac{1}{2} + o(1))(n - \sqrt{n^2 - 4m})$ and $b = (\frac{1}{2} + o(1))(n + \sqrt{n^2 - 4m})$. The corresponding number of colorings is $2^{b + o(n)}$.*

We also considered another conjecture of Lazebnik (see, for example, [21]) that the Turán graphs $T_r(n)$ are always extremal when $r \leqslant q$. Building upon the techniques in [21] that answered the $r = 2, q = 3$ case, we confirmed this conjecture for large n and $r = q - 1$.

THEOREM 1.4. *Fix an integer $q \geqslant 4$. For all sufficiently large n, the Turán graph $T_{q-1}(n)$ has more q-colorings than any other graph with the same number of vertices and edges.*

We close by mentioning some related work. Tomescu [**26**–**33**] and Dohmen [**11**, **12**] considered the problem of maximizing or minimizing the number of q-colorings of G given some other parameters, such as chromatic number, connectedness, planarity, and girth. Wright [**36**] asymptotically determined the total number of q-colored labeled n-vertex graphs with m edges, for the entire range of m; this immediately gives an asymptotic approximation for the *average* value of $P_G(q)$ over all labeled n-vertex graphs with m edges.

Graph coloring is also a special case of a homomorphism problem, and as we discuss in our concluding remarks, our approach easily extends to that more general setting. Recall that a graph homomorphism $\phi : G \to H$ is a map from the vertices of G to those of H, such that adjacent vertices in G are mapped to adjacent vertices in H. Thus, the number of q-colorings of G is precisely the number of homomorphisms from G to K_q. Another interesting target graph H is the two-vertex graph consisting of a single edge, plus a loop at one vertex. Then, the number of homomorphisms is precisely the number of independent sets in G, and the problem of estimating that number given some partial information about G is motivated by various questions in statistical physics and the theory of partially ordered sets. Alon [**1**] studied the maximum number of independent sets that a k-regular graph of order n can have, and Kahn [**16**, **17**] considered this problem under the additional assumption that the k-regular graph is bipartite. This was recently resolved for non-bipartite graphs by Zhao [**38**]. Galvin and Tetali [**15**] generalized the main result from [**16**] to arbitrary target graphs H.

Another direction of related research was initiated by the question of Erdős and Rothschild (see Erdős [**13**, **14**], Yuster [**37**], Alon, Balogh, Keevash, and Sudakov [**2**], Balogh [**4**], and others), about the maximum over all n-vertex graphs of the number of q-edge-colorings (not necessarily proper) that do not contain a monochromatic K_r-subgraph. Our method is somewhat similar to that in [**2**], and these two problems may be more deeply related than just a similarity in their formulations.

The rest of this paper is organized as follows. The next section contains some definitions, and a formulation of the Szemerédi regularity lemma. In Section 3, we prove Theorems 3.2 and 3.3, which (asymptotically) reduce the general case of the problem to a quadratically constrained linear program. Then, in the next section we solve the relevant instances of the optimization problem to give approximate versions of our main theorems. Sections 5 and 6 refine these into the precise forms of Theorems 1.2 and 1.3. We prove Theorem 1.4 in Section 7. The final section contains some concluding remarks and open problems.

2. *Preliminaries*

The following (standard) asymptotic notation will be utilized extensively. For two functions $f(n)$ and $g(n)$, we write $f(n) = o(g(n))$ if $\lim_{n \to \infty} f(n)/g(n) = 0$, and $f(n) = O(g(n))$ or $g(n) = \Omega(f(n))$ if there exists a constant M such that $|f(n)| \leqslant M|g(n)|$ for all sufficiently large n. We also write $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ are satisfied.

We use $[q]$ to denote the set $\{1, 2, \ldots, q\}$, and $2^{[q]}$ to denote the collection of all of its subsets. As mentioned in Section 1, the *Turán graph* $T_q(n)$ is the complete r-partite graph on n vertices with all parts of size either $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$.

Given two graphs with the same number of vertices, their *edit distance* is the minimum number of edges that need to be added or deleted from one graph to make it isomorphic to the other. We say that two graphs are *d-close* if their edit distance is at most d.

The rest of this section is devoted to formulating the celebrated Szemerédi regularity lemma. This theorem roughly states that every graph, no matter how large, can be approximated by an object of bounded complexity, which corresponds to a union of a bounded number of random-looking graphs. To measure the randomness of edge distribution, we use the following definition.

Let the edge density $d(A, B)$ be the fraction $e(A, B)/(|A||B|)$, where $e(A, B)$ is the number of edges between A and B.

DEFINITION 2.1. A pair (X, Y) of disjoint subsets of a graph is ϵ-*regular* if every pair of subsets $X' \subset X$ and $Y' \subset Y$ with $|X'| \geqslant \epsilon|X|$ and $|Y'| \geqslant \epsilon|Y|$ has $|d(X', Y') - d(X, Y)| < \epsilon$.

In this paper, we use the following convenient form of the regularity lemma, which is essentially Theorem IV.5.29$'$ in the textbook [**9**].

THEOREM 2.2. *For every $\epsilon > 0$, there is a natural number $M' = M'(\epsilon)$ such that every graph $G = (V, E)$ has a partition $V = \bigcup_{i=1}^{M} V_i$ with the following properties. The sizes of the vertex clusters V_i are as equal as possible (differing by at most 1), their number is between $1/\epsilon \leqslant M \leqslant M'$, and all but at most ϵM^2 of the pairs (V_i, V_j) are ϵ-regular.*

3. *Reduction to an optimization problem*

In this section, we show that the solution of the following quadratically constrained linear program answers our main problem asymptotically. (Observe that the logarithms are merely constant multipliers for the variables α_A.)

OPTIMIZATION PROBLEM 1. Fix an integer $q \geqslant 2$ and a real parameter γ. Consider the following objective and constraint functions:

$$\mathrm{OBJ}(\boldsymbol{\alpha}) := \sum_{A \neq \emptyset} \alpha_A \log |A|; \qquad \mathrm{V}(\boldsymbol{\alpha}) := \sum_{A \neq \emptyset} \alpha_A, \quad \mathrm{E}(\boldsymbol{\alpha}) := \sum_{A \cap B = \emptyset} \alpha_A \alpha_B.$$

The vector $\boldsymbol{\alpha}$ has $2^q - 1$ coordinates $\alpha_A \in \mathbb{R}$ indexed by the nonempty subsets $A \subset [q]$, and the sum in $\mathrm{E}(\boldsymbol{\alpha})$ runs over *unordered* pairs of disjoint nonempty sets $\{A, B\}$. Let $\mathrm{FEAS}(\gamma)$ be the *feasible set* of vectors defined by the constraints $\boldsymbol{\alpha} \geqslant 0$, $\mathrm{V}(\boldsymbol{\alpha}) = 1$, and $\mathrm{E}(\boldsymbol{\alpha}) \geqslant \gamma$. We seek to maximize $\mathrm{OBJ}(\boldsymbol{\alpha})$ over the set $\mathrm{FEAS}(\gamma)$, and we define $\mathrm{OPT}(\gamma)$ to be this maximum value, which exists by compactness. We will write that the vector $\boldsymbol{\alpha}$ *solves* $\mathrm{OPT}(\gamma)$ when both $\boldsymbol{\alpha} \in \mathrm{FEAS}(\gamma)$ and $\mathrm{OBJ}(\boldsymbol{\alpha}) = \mathrm{OPT}(\gamma)$.

NOTE. In the remainder of this paper, we will write $\sum_A$ instead of $\sum_{A \neq \emptyset}$ because it is clear from the definition of $\boldsymbol{\alpha}$ that the empty set is excluded.

CONSTRUCTION 1: $G_{\boldsymbol{\alpha}}(n)$. Let n and m be the desired numbers of vertices and edges, and let $\boldsymbol{\alpha} \in \mathrm{FEAS}(m/n^2)$ be a feasible vector. Consider the following n-vertex graph, which we call $G_{\boldsymbol{\alpha}}(n)$. Partition the vertices into (possibly empty) clusters V_A such that each $|V_A|$ differs from $n\alpha_A$ by less than 1. For every pair of clusters (V_A, V_B) which is indexed by disjoint subsets, place a complete bipartite graph between the clusters.

Observe that any coloring that for each cluster V_A uses only colors from A is a proper coloring. Therefore, if all $n\alpha_A$ happened to be integers, then $G_{\boldsymbol{\alpha}}(n)$ would have at least $\prod_A |A|^{n\alpha_A} = e^{\mathrm{OBJ}(\boldsymbol{\alpha})n}$ colorings, and also precisely $\mathrm{E}(\boldsymbol{\alpha})n^2$ edges. But we cannot simply apply Construction 1 to the $\boldsymbol{\alpha}$ that solves $\mathrm{OPT}(m/n^2)$, because it may happen that $G_{\boldsymbol{\alpha}}(n)$ has fewer than m edges if the entries of $\boldsymbol{\alpha}$ are not integer multiples of $1/n$. Fortunately, the shortfall cannot be substantial.

PROPOSITION 3.1.   *The number of edges in any $G_{\boldsymbol{\alpha}}(n)$ differs from $\mathrm{E}(\boldsymbol{\alpha})n^2$ by less than $2^q n$. Also, for any other vector $\boldsymbol{\nu}$, the edit distance between $G_{\boldsymbol{\alpha}}(n)$ and $G_{\boldsymbol{\nu}}(n)$ is at most $\|\boldsymbol{\alpha} - \boldsymbol{\nu}\|_1 n^2 + 2^{q+1} n$, where $\| \cdot \|_1$ is the L^1-norm.*

The proof is elementary and routine, so we will defer it to Section 3.4 so as not to interrupt this exposition. To recover from the $\mathrm{O}(n)$ edge deficit, we extend the construction in the following way.

CONSTRUCTION 2: $G'_{\boldsymbol{\alpha}}(n)$.   Let n and m be the desired number of vertices and edges, and let $\boldsymbol{\alpha} \in \mathrm{FEAS}(m/n^2)$ be a feasible vector. If $G_{\boldsymbol{\alpha}}(n)$ from Construction 1 already has at least m edges, then set $G'_{\boldsymbol{\alpha}}(n) = G_{\boldsymbol{\alpha}}(n)$.

Otherwise, $G_{\boldsymbol{\alpha}}(n)$ is short by, say, k edges, and $k = \mathrm{O}(n)$ by Proposition 3.1. Let V_A be its largest cluster whose index A is not a singleton. Suppose first that $|V_A| \geqslant 2\lceil\sqrt{k}\rceil$. So far V_A does not span any edges, so we can add k edges to $G_{\boldsymbol{\alpha}}(n)$ by selecting two disjoint subsets $U_1, U_2 \subset V_A$ of size $\lceil\sqrt{k}\rceil$, and setting a k-edge bipartite graph between them. Call the result $G'_{\boldsymbol{\alpha}}(n)$.

The last case is $|V_A| < 2\lceil\sqrt{k}\rceil$. We later show that this only arises when the maximum number of colorings is only $2^{\mathrm{o}(n)}$, and this is already achieved by the Turán graph $T_q(n)$. Therefore, to clean up the statements of our theorems, we just define $G'_{\boldsymbol{\alpha}}(n) = T_q(n)$ here.

3.1. *Structure of asymptotic argument*

We are now ready to state our theorem, which shows that solutions to Optimization Problem 1 produce graphs which asymptotically maximize the number of q-colorings.

THEOREM 3.2.   *For any $\epsilon > 0$, the following holds for any sufficiently large n, and any m less than or equal to the number of edges in the Turán graph $T_q(n)$.*
  (i) *Every n-vertex graph with m edges has fewer than $e^{(\mathrm{OPT}(m/n^2)+\epsilon)n}$ proper q-colorings.*
  (ii) *Any $\boldsymbol{\alpha}$ which solves $\mathrm{OPT}(m/n^2)$ yields a graph $G'_{\boldsymbol{\alpha}}(n)$ via Construction 2 which has at least m edges and more than $e^{(\mathrm{OPT}(m/n^2)-\epsilon)n}$ proper q-colorings.*

REMARK.   The number of colorings can only increase when edges are deleted, so one may take an arbitrary m-edge subgraph of $G'_{\boldsymbol{\alpha}}(n)$ if one requires a graph with exactly m edges.

The key ingredient in the proof of Theorem 3.2 is Szemerédi's regularity lemma. Part (ii) is routine, and full details are given in Section 3.4. On the other hand, the argument for part (i) is more involved, so we highlight its structure here so that the reader does not get lost in the details. The proof breaks into the following claims.

CLAIM 1.   *For any $\delta > 0$, there exists n_0 such that the following holds for any graph $G = (V, E)$ with $n > n_0$ vertices and m edges. The regularity lemma gives a special partition of the vertex set into sets $V_1, \ldots, V_M$ of almost equal size, where M is upper bounded by a constant depending only on δ. Then, we may delete at most δn^2 edges of G in such a way that the resulting graph G' has the following properties.*
  (i) *Each $G'[V_i]$ spans no edges.*
  (ii) *If G' has any edges at all between two parts V_i and V_j, then in fact it has an edge between every pair of subsets $U \subset V_i$, $W \subset V_j$ with $|U| \geqslant \delta|V_i|$ and $W \geqslant \delta|V_j|$.*
Note that since G' is a subgraph of G, the number of q-colorings can only increase.

CLAIM 2.   *Let $\mathcal{C}_1$ be the set of colorings of G'. Then, if we keep only those colorings $\mathcal{C}_2 \subset \mathcal{C}_1$ with the property that in each V_i, any color is used either zero times or at least $\delta|V_i|$ times, we will still have $|\mathcal{C}_2| \geqslant e^{-c_\delta n}|\mathcal{C}_1|$. Here, c_δ is a constant that tends to zero with δ. Now each coloring in $\mathcal{C}_2$ has the special property that whenever the same color appears on two parts V_i and V_j, then there cannot be any edges between those entire parts.*

CLAIM 3.   *By looking at which colors appear on each part V_i, we may associate each coloring with a map $[M] \to 2^{[q]}$. Let $\phi : [M] \to 2^{[q]}$ be a map which is associated with the maximum number of colorings in $\mathcal{C}_2$. Then, if we keep only those colorings $\mathcal{C}_3 \subset \mathcal{C}_2$ which give ϕ, we still have $|\mathcal{C}_3| \geqslant 2^{-qM}|\mathcal{C}_2|$.*

CLAIM 4.   *For every nonempty $A \subset [q]$, let V_A be the union of those parts V_i for which $\phi(i) = A$. (These are the parts that in all colorings in $\mathcal{C}_3$ are colored using exactly colors from A.) Define the vector $\boldsymbol{\alpha}$ by setting each $\alpha_A = |V_A|/n$. Then $G' \subset G_{\boldsymbol{\alpha}}(n)$, and since G' only differs from our original G by at most δn^2 edges, we also have $\boldsymbol{\alpha} \in \mathrm{FEAS}(m/n^2 - \delta)$. Thus*

$$|\mathcal{C}_3| \leqslant \prod_A |A|^{|V_A|} = e^{\mathrm{OBJ}(\boldsymbol{\alpha})n} \leqslant e^{\mathrm{OPT}(m/n^2 - \delta)n}.$$

CLAIM 5.   *The function OPT is uniformly continuous. Thus, for an appropriate (sufficiently small) choice of $\delta > 0$, we have for all sufficiently large n that*

$$P_G(q) \leqslant P_{G'}(q) \leqslant e^{c_\delta n} \cdot 2^{qM} \cdot e^{\mathrm{OPT}(m/n^2 - \delta)n} < e^{(\mathrm{OPT}(m/n^2) + \epsilon)n},$$

as desired. (Recall that $P_G(q)$ is the number of q-colorings of G.)

By combining these five claims with an elementary analysis argument, we also obtain a stability result, which roughly states that if a graph has 'close' to the optimal number of colorings, then it must resemble a graph from Construction 1. A stability result is very useful, because the approximate structure later allows us to apply combinatorial arguments to refine our asymptotic results into exact results. We quantify this in terms of the edit distance, which we defined in Section 2. Recall that we say that two graphs are d-close when their edit distance is at most d. We prove the following theorem in Section 3.5.

THEOREM 3.3.   *For any $\epsilon, \kappa > 0$, the following holds for all sufficiently large n. Let G be an n-vertex graph with $m \leqslant \kappa n^2$ edges, which maximizes the number of q-colorings. Then G is ϵn^2-close to some $G_{\boldsymbol{\alpha}}(n)$ from Construction 1, for an $\boldsymbol{\alpha}$ which solves $\mathrm{OPT}(\gamma)$ for some $|\gamma - m/n^2| \leqslant \epsilon$ with $\gamma \leqslant \kappa$.*

REMARK.   *This theorem is only useful if the resulting γ falls within the range of densities for which the solution of OPT is known. The technical parameter κ is used to keep γ within this range.*

3.2. Finer resolution in the sparse case

The regularity lemma is nontrivial only for graphs with positive edge density (that is, a quadratic number of edges). This typically presents a serious and often insurmountable obstacle when trying to extend regularity-based results to situations involving sparse graphs. Although much work has been done to develop sparse variants of the regularity lemma, the resulting analogs are weaker and much more difficult to apply.

Let us illustrate the issue by attempting to apply Theorem 3.2 when $m = o(n^2)$. Then, we find that the maximum number of q-colorings of any n-vertex graph with m edges is $e^{cn+o(n)}$, where $c = \text{OPT}(0) = \log q$ is a constant entirely determined by q. Note that the final asymptotic is independent of m, even if m grows extremely slowly compared with n^2. This is because the key parameter was the density m/n^2, which already vanished once $m = o(n^2)$. Thus, the interesting question in the sparse case is to distinguish between sparse graphs and very sparse graphs, by looking inside the $o(n)$ error term in the exponent.

We are able to circumvent these difficulties by making the following key observation which allows us to pass to a dense subgraph. As it turns out, every sparse graph which maximizes the number of q-colorings has a nice structure: most of the vertices are isolated, and *all* of the edges are contained in a subgraph which is dense, but not too dense. Section 3.6 contains the following lemma's short proof, which basically boils down to a comparison against the smallest Turán graph with at least m edges.

LEMMA 3.4.   *Fix an integer $q \geqslant 2$ and a threshold $\kappa > 0$. Given any positive integer m, there exists an $n_0 = \Theta(\sqrt{m})$ with $m/n_0^2 \leqslant \kappa$ such that the following holds for any $n \geqslant n_0$. In every n-vertex graph G with m edges, which maximizes the number of q-colorings, there is a set of n_0 vertices which spans all of the edges.*

The fact that our graph is sparse becomes a benefit rather than a drawback, because it allows us to limit the edge density from above by any fixed threshold. This is useful, because we can completely solve the optimization problem for all densities below

$$\kappa_q = \left(\sqrt{\frac{\log(q/(q-1))}{\log q}} + \sqrt{\frac{\log q}{\log(q/(q-1))}} \right)^{-2}.$$

We will prove the following proposition in Section 4.1.

PROPOSITION 3.5.   *Fix an integer $q \geqslant 3$. For any $0 \leqslant \gamma \leqslant \kappa_q$, the unique solution (up to a permutation of the ground set $[q]$) to $\text{OPT}(\gamma)$ has the following form.*

$$\alpha_{\{1\}} = \sqrt{\gamma \cdot \log \frac{q}{q-1} / \log q}, \quad \alpha_{\{2,\dots,q\}} = \frac{\gamma}{\alpha_{\{1\}}}, \quad \alpha_{[q]} = 1 - \alpha_{\{1\}} - \alpha_{\{2,\dots,q\}}, \tag{1}$$

with all other $\alpha_A = 0$. This gives $\text{OPT}(\gamma) = \log q - 2\sqrt{\gamma \cdot \log \frac{q}{q-1} \cdot \log q}$.

Since we have the complete solution of the relevant instance of the optimization problem, we can give explicit bounds when we transfer our asymptotic results from the previous section to the sparse case. We can also explicitly describe the graph that approximates any optimal graph, as follows. Let t_1 and t_2 be real numbers that satisfy $t_1/t_2 = \log \frac{q}{q-1} / \log q$ and $t_1 t_2 = m$. Take a complete bipartite graph between two vertex clusters V_1 and V_2 with sizes $|V_i| = \lceil t_i \rceil$, and add enough isolated vertices to make the total number of vertices exactly n. Call the result $G_{n,m}$.

PROPOSITION 3.6.   *Fix an integer $q \geqslant 3$. The following hold for all sufficiently large $m \leqslant \kappa_q n^2$.*

(i) *The maximum number of q-colorings of an n-vertex graph with m edges is $q^n e^{(-c+o(1))\sqrt{m}}$, where $c = 2\sqrt{\log \frac{q}{q-1} \cdot \log q}$. Here, the $o(1)$ term tends to zero as $m \to \infty$.*

(ii) *For any $\epsilon > 0$, as long as m is sufficiently large, every n-vertex graph G with m edges, which maximizes the number of q-colorings, is ϵm-close to the graph $G_{n,m}$ which we described above.*

We prove this proposition in Section 3.6. Note that part (i) is precisely the final claim of Theorem 1.2.

3.3. *Proof of Theorem 3.2(i)*

This section contains the proofs of the claims in Section 3.1, except for Claim 3, which is obvious. Together, these establish part (i) of Theorem 3.2, which gives the asymptotic upper bound for the number of q-colorings of an n-vertex graph with m edges.

Proof of Claim 1. Apply Szemerédi's regularity lemma (Theorem 2.2) with parameter $\epsilon = \delta/3$ to partition V into nearly equal parts $V_1, \ldots, V_M$. Then, all but ϵM^2 of the pairs (V_i, V_j) are ϵ-regular, and $M \geqslant 1/\epsilon$. Importantly, M is also upper bounded by a constant independent of n. We clean up the graph in a way typical of many applications of the regularity lemma. Delete all edges in each induced subgraph $G[V_i]$, all edges between pairs (V_i, V_j) which are not ϵ-regular, and all edges between pairs (V_i, V_j) whose edge density is at most ϵ. Since all $|V_i| = (1 + o(1))n/M$, the number of deleted edges is at most

$$(1 + o(1)) \left[M \binom{n/M}{2} + \epsilon M^2 \left(\frac{n}{M} \right)^2 + \epsilon \binom{n}{2} \right] \leqslant (1 + o(1)) \left[\frac{\epsilon n^2}{2} + \epsilon n^2 + \frac{\epsilon n^2}{2} \right],$$

which is indeed less than δn^2 when n is sufficiently large.

It remains to show property (ii). The only edges remaining in G' are those between ϵ-regular pairs (V_i, V_j) with edge density greater than ϵ. By the definition of ϵ-regularity (and since $\delta > \epsilon$), the edge density between every pair of sets $|U| \geqslant \delta|V_i|$, $|W| \geqslant \delta|V_j|$ must be positive. In particular, there must be at least one edge, which establishes property (ii). $\qquad\square$

Proof of Claim 2. We aim to establish $|\mathcal{C}_2| \geqslant e^{-c_\delta n}|\mathcal{C}_1|$, with $c_\delta = q\delta \log(e^2/\delta)$. It is a simple calculus exercise to verify that $c_\delta \to 0$ as $\delta \to 0$. Let us show that we can obtain any coloring $\psi \in \mathcal{C}_1$ by starting with an appropriate coloring $\psi' \in \mathcal{C}_2$, and changing only a few color choices. Since we may assume $\delta < 1/q$, every part V_i has some color c_i^* which appears on at least δ-fraction of its vertices. Now consider each V_i. For every color c which appears less than $\delta|V_i|$ times in V_i, use color c_i^* to re-color all vertices of V_i that had color c under ψ. Now all colors appear either 0 or at least $\delta|V_i|$ times, so once we verify that the coloring is still proper, we will have our desired $\psi' \in \mathcal{C}_2$. But the only way to make a monochromatic edge is to have two distinct parts V_i, V_j, with $c_i^* = c_j^*$, joined by at least one edge. Then part (ii) of Claim 1 implies that there is also some edge between the $\delta|V_i|$ vertices in V_i originally colored c_i^* under ψ, and the $\delta|V_j|$ vertices in V_j originally colored c_j^*. This contradicts the fact that ψ was a proper coloring.

Reversing the process, it is clear that ψ can be recovered by taking $\psi' \in \mathcal{C}_2$ and changing the colors of at most $\delta|V_i|$ vertices for every color $c \in [q]$ and every $1 \leqslant i \leqslant M$. Note that for each $c \in [q]$, we recolor a subset of G of total size at most $\sum_i \delta|V_i| = \delta n$. Using the bounds $\binom{n}{r} \leqslant (en/r)^r$ and $(1 + x) \leqslant e^x$, we see that the total number of distinct ways in which we can modify any given $\psi' \in \mathcal{C}_2$ is at most

$$\left[\sum_{r=0}^{\delta n} \binom{n}{r} \right]^q \leqslant \left[(1 + \delta n) \binom{n}{\delta n} \right]^q \leqslant \left[e^{\delta n} \left(\frac{en}{\delta n} \right)^{\delta n} \right]^q = e^{c_\delta n},$$

which provides the desired upper bound on $|\mathcal{C}_1|/|\mathcal{C}_2|$.

The final part of this claim is a simple consequence of property (ii) of Claim 1. Indeed, suppose that some coloring in $\mathcal{C}_2$ assigns the same color c to some vertices $U_i \subset V_i$ and $U_j \subset V_j$. Since this is a proper coloring, there cannot be any edges between U_i and U_j. Yet $|U_i| \geqslant \delta|V_i|$ and $|U_j| \geqslant \delta|V_j|$ by definition of $\mathcal{C}_2$. Therefore, by property (ii) of Claim 1, there are no edges at all between V_i and V_j, as claimed. $\square$

Proof of Claim 4.   Recall that $G_{\boldsymbol{\alpha}}(n)$ was obtained in Construction 1 by setting a complete bipartite graph between every pair (V_A, V_B) indexed by disjoint subsets. The last part of Claim 2 implies that G' has no edges at all between parts V_i and V_j which receive overlapping color sets under $\mathcal{C}_3$. Furthermore, each $G'[V_i]$ is empty by part (i) of Claim 1. Therefore, G' has no edges in each V_A, and also has no edges between any V_A and V_B that are indexed by overlapping sets. Hence G' is indeed a subgraph of $G_{\boldsymbol{\alpha}}(n)$.

Furthermore, $G_{\boldsymbol{\alpha}}(n)$ has at least $m - \delta n^2$ edges, because G' differs from G by at most δn^2 edges. Yet all $n\alpha_A$ are integers by construction, so $G_{\boldsymbol{\alpha}}(n)$ has precisely $\mathrm{E}(\boldsymbol{\alpha})n^2$ edges. Therefore, $\boldsymbol{\alpha} \in \mathrm{FEAS}(m/n^2 - \delta)$, as claimed. The final inequality in Claim 4 follows from the fact that $\mathcal{C}_3$ only uses colors from A to color each V_A, and the definitions of $\alpha_A = |V_A|/n$ and $\mathrm{OBJ}(\boldsymbol{\alpha}) = \sum_A \alpha_A \log|A|$. $\square$

Proof of Claim 5.   The only nontrivial part of this claim is the continuity of OPT on its domain, which is the set of γ for which $\mathrm{FEAS}(\gamma) \neq \emptyset$. This is easily recognized as the interval $(-\infty, \frac{q-1}{2q}]$, where the upper endpoint, which corresponds to the q-partite Turán graph, equals $\mathrm{E}(\boldsymbol{\alpha})$ for the vector $\boldsymbol{\alpha}$ with $\alpha_A = 1/q$ for all singletons A. Note that the constraint $\boldsymbol{\alpha} \geqslant 0$ already guarantees that $\mathrm{E}(\boldsymbol{\alpha}) \geqslant 0$, so OPT is constant on $(-\infty, 0]$.

Fix an $\epsilon > 0$. Since OPT is monotonically decreasing by definition, and constant on $(-\infty, 0]$, it suffices to show that any $0 \leqslant \gamma < \gamma' \leqslant (q-1)/(2q)$ with $|\gamma' - \gamma| < \epsilon^2$ has $\mathrm{OPT}(\gamma') > \mathrm{OPT}(\gamma) - 2^{q+1}\epsilon \log q$. Select any $\boldsymbol{\alpha}$ which solves $\mathrm{OPT}(\gamma)$. We will adjust $\boldsymbol{\alpha}$ to find an $\boldsymbol{\alpha}' \in \mathrm{FEAS}(\gamma')$ with $\mathrm{OBJ}(\boldsymbol{\alpha}') > \mathrm{OBJ}(\boldsymbol{\alpha}) - 2^{q+1}\epsilon \log q$, using essentially the same perturbation as in Construction 2.

If there is an $\alpha_A \geqslant 2\epsilon$ with $|A| \geqslant 2$, shift ϵ of α_A's value to each of $\alpha_{\{i\}}$ and $\alpha_{\{j\}}$ for some pair of distinct $i, j \in A$ (formally, α_A falls by 2ϵ, and each of $\alpha_{\{i\}}$ and $\alpha_{\{j\}}$ increase by ϵ). This clearly keeps $\mathrm{V}(\boldsymbol{\alpha})$ invariant, and it increases $\mathrm{E}(\boldsymbol{\alpha})$ by at least ϵ^2 because $\alpha_{\{i\}}\alpha_{\{j\}}$ is a summand of $\mathrm{E}(\boldsymbol{\alpha})$. Yet it only reduces $\mathrm{OBJ}(\boldsymbol{\alpha})$ by at most $2\epsilon \log|A| \leqslant 2\epsilon \log q$, so $\mathrm{OBJ}(\boldsymbol{\alpha}') \geqslant \mathrm{OBJ}(\boldsymbol{\alpha}) - 2\epsilon \log q$, finishing this case.

On the other hand, if all nonsingletons A have $\alpha_A < 2\epsilon$, then $\mathrm{OBJ}(\boldsymbol{\alpha})$ is already less than $2^q \cdot 2\epsilon \log q$. Since OPT is always nonnegative, we trivially have $\mathrm{OPT}(\gamma') \geqslant 0 > \mathrm{OPT}(\gamma) - 2^{q+1}\epsilon \log q$, as desired. $\square$

3.4. *Proof of Theorem 3.2(ii)*

In this section, we establish the asymptotic tightness of our upper bound, by showing that Construction 2 produces graphs that asymptotically maximize the number of q-colorings. We will need Proposition 3.1, so we prove it first.

Proof of Proposition 3.1.   Define the variables $n_A = n\alpha_A$ (not necessarily integers), and call the expressions $\sum_A n_A$ and $\sum_{A \cap B = \emptyset} n_A n_B$, the number of *fractional vertices* and *fractional edges*, respectively. Initially, there are exactly n fractional vertices and $\mathrm{E}(\boldsymbol{\alpha})n^2$ fractional edges.

Recall that the construction rounds each n_A either up or down to the next integer. Let us perform these individual roundings sequentially, finishing all of the downward roundings before the upward roundings. This ensures that the number of fractional vertices is kept at most n throughout the process. But each iteration changes the number of fractional edges by at most $\sum_A n_A \leqslant n$, and there are at most 2^q iterations, so our final number of edges is indeed within $2^q n$ of m.

The second part of the proposition is proved similarly. We can apply the same iterative process to change each part size from $\alpha_A n$ to $\nu_A n$, in such a way that all downward adjustments are performed first. When updating the coordinate indexed by $A \subset [q]$, we affect at most $(|\alpha_A n - \nu_A n| + 2)n$ edges, where the extra 2 comes from the fact that the part sizes were rounded off. Therefore, after at most 2^q total iterations, the total number of edges we edit is indeed at most $\|\boldsymbol{\alpha} - \boldsymbol{\nu}\|_1 n^2 + 2^{q+1}n$. $\qquad\square$

Proof of Theorem 3.2(ii). Let n and m be given, with m less than the number of edges in the Turán graph $T_q(n)$. Suppose that we have a vector $\boldsymbol{\alpha} \in \text{FEAS}(m/n^2)$ which achieves the maximum $\text{OBJ}(\boldsymbol{\alpha}) = \text{OPT}(m/n^2)$. Construction 2 produces a graph $G'_{\boldsymbol{\alpha}}(n)$ with n vertices and at least m edges, which we will show has more than $e^{(\text{OPT}(m/n^2)-\epsilon)n}$ proper q-colorings, as long as n is sufficiently large.

If $G_{\boldsymbol{\alpha}}(n)$ already has at least m edges, then we defined $G'_{\boldsymbol{\alpha}}(n) = G_{\boldsymbol{\alpha}}(n)$, which has at least $\prod_A |A|^{\lfloor n\alpha_A \rfloor} \geqslant \prod_A |A|^{n\alpha_A - 1} = e^{\text{OBJ}(\boldsymbol{\alpha})n} / \prod_A |A| = e^{\text{OBJ}(\boldsymbol{\alpha})n - O(1)}$ colorings, because all colorings that use only colors from A for each V_A are proper.

Otherwise, $G_{\boldsymbol{\alpha}}(n)$ is short by, say, k edges, which is at most $2^q n$ by Proposition 3.1. If the largest $|V_A|$ indexed by a nonsingleton is at least $2\lceil \sqrt{k} \rceil$, then our construction places a k-edge bipartite graph between $U_1, U_2 \subset V_A$. Let c_1 and c_2 be two distinct colors in A. Even if we force every vertex in each U_i to take the color c_i, we only lose at most a factor of $q^{2\lceil \sqrt{k} \rceil} = e^{o(n)}$ compared with the bound in the previous paragraph. This is because each of the $2\lceil \sqrt{k} \rceil$ vertices in $U_1 \cup U_2$ had its number of color choices reduced from $|A| \leqslant q$ to 1. Therefore, $G'_{\boldsymbol{\alpha}}(n)$ still has at least $e^{\text{OBJ}(\boldsymbol{\alpha})n - o(n)}$ colorings.

The final case is when all parts of V_A indexed by nonsingletons are smaller than $2\lceil \sqrt{k} \rceil$. Here, the construction simply defines $G'_{\boldsymbol{\alpha}}(n)$ to be the Turán graph $T_q(n)$. Since $\log|A| = 0$ for singletons A, the upper bound on $|V_A|$ implies that $\text{OBJ}(\boldsymbol{\alpha}) \leqslant 2^q \cdot 2\lceil \sqrt{k} \rceil / n \cdot \log q$. This is less than ϵ for sufficiently large n, because we had $k \leqslant 2^q n$. Then, $e^{(\text{OPT}(m/n^2)-\epsilon)n} < 1$, which is of course less than the number of q-colorings of the Turán graph $T_q(n)$. This completes our proof. $\qquad\square$

3.5. *Proof of Theorem 3.3*

In this section, we prove that any n-vertex graph with m edges, which maximizes the number of q-colorings, is in fact close (in edit-distance) to a graph $G_{\boldsymbol{\alpha}}(n)$ from Construction 1. In fact, we prove something slightly stronger: if a graph has 'close' to the maximum number of q-colorings, then it must be 'close' (in edit-distance) to an asymptotically optimal graph from Construction 1.

LEMMA 3.7. *For any $\epsilon, \kappa > 0$, there exists $\delta > 0$ such that the following holds for all sufficiently large n. Let G be an n-vertex graph with $m \leqslant \kappa n^2$ edges and at least $e^{(\text{OPT}(m/n^2)-\delta)n}$ proper q-colorings. Then G is ϵn^2-close to some $G_{\boldsymbol{\alpha}}(n)$ from Construction 1, for an $\boldsymbol{\alpha}$ which solves $\text{OPT}(\gamma)$ for some $|\gamma - m/n^2| \leqslant \epsilon$ with $\gamma \leqslant \kappa$.*

Note that this lemma immediately implies Theorem 3.3, because Theorem 3.2 established that the maximum number of colorings of an n-vertex graph with m edges was $e^{(\text{OPT}(m/n^2)+o(1))n}$. Its proof is an elementary analysis exercise in compactness, which only requires the continuity of OBJ, OPT, V, and E, the fact that $\boldsymbol{\alpha}$ and the edge densities m/n^2 reside in compact spaces, and the following consequence of Claims 1–4 of Section 3.1 (whose simple proof we omit):

COROLLARY 3.8. *For every $\delta > 0$, the following holds for all sufficiently large n. Every q-colorable, n-vertex graph G with m edges is δn^2-close to a subgraph of some $G_{\boldsymbol{\alpha}}(n)$ with $\boldsymbol{\alpha} \in \mathrm{FEAS}(m/n^2 - \delta)$. Also, G has at most $e^{(\mathrm{OBJ}(\boldsymbol{\alpha})+\delta)n}$ proper q-colorings.*

Proof of Lemma 3.7. We proceed by contradiction. Then, there is some fixed $\epsilon > 0$, a sequence $\delta_i \to 0$, and a sequence of graphs G_i with the following properties:
   (i) G_i has at least as many vertices as required to apply Corollary 3.8 with parameter δ_i;
   (ii) G_i has at least $e^{(\mathrm{OPT}(m_i/n_i^2)-\delta_i)n_i}$ colorings, where n_i and m_i are its number of vertices and edges, and $m_i \leqslant \kappa n_i^2$;
   (iii) G_i is at least ϵn_i^2-far from $G_{\boldsymbol{\alpha}}(n_i)$ for every $\boldsymbol{\alpha}$ that solves $\mathrm{OPT}(\gamma)$ with $|\gamma - m_i/n_i^2| \leqslant \epsilon$.
Applying Corollary 3.8 to each G_i with parameter δ_i, we find vectors $\boldsymbol{\alpha}_i \in \mathrm{FEAS}(m_i/n_i^2 - \delta_i)$ such that G_i is $\delta_i n_i^2$-close to some subgraph G_i' of $G_{\boldsymbol{\alpha}_i}(n_i)$, and each G_i has at most $e^{(\mathrm{OBJ}(\boldsymbol{\alpha}_i)+\delta_i)n_i}$ proper q-colorings. Combining this with property (ii) above, we find that each $\mathrm{OBJ}(\boldsymbol{\alpha}_i) \geqslant \mathrm{OPT}(m_i/n_i^2) - 2\delta_i$. The densities m_i/n_i^2 and the vectors $\boldsymbol{\alpha}_i$ live in bounded (hence compact) spaces. Therefore, by passing to a subsequence, we may assume that $m_i/n_i^2 \to \gamma \leqslant \kappa$ and $\boldsymbol{\alpha}_i \to \boldsymbol{\alpha}$ for some limit points γ and $\boldsymbol{\alpha}$.

Observe that by continuity, both $\boldsymbol{\alpha} \in \mathrm{FEAS}(\gamma)$ and $\mathrm{OBJ}(\boldsymbol{\alpha}) \geqslant \mathrm{OPT}(\gamma)$. Therefore $\boldsymbol{\alpha}$ solves $\mathrm{OPT}(\gamma)$, that is, $\mathrm{OBJ}(\boldsymbol{\alpha}) = \mathrm{OPT}(\gamma)$. Furthermore, although *a priori* we only knew that $\mathrm{E}(\boldsymbol{\alpha}) \geqslant \gamma$, maximality implies that in fact $\mathrm{E}(\boldsymbol{\alpha}) = \gamma$. Indeed, if not then one could shift more mass to $\alpha_{[q]}$ to increase $\mathrm{OBJ}(\boldsymbol{\alpha})$ while staying within the feasible set. This would contradict that $\mathrm{OBJ}(\boldsymbol{\alpha}) = \mathrm{OPT}(\gamma)$.

We finish by showing that eventually G_i is ϵn_i^2-close to $G_{\boldsymbol{\alpha}}(n_i)$, contradicting (iii). To do this, we show that all three of the edit-distances between $G_i \leftrightarrow G_i' \leftrightarrow G_{\boldsymbol{\alpha}_i}(n_i) \leftrightarrow G_{\boldsymbol{\alpha}}(n_i)$ are $o(n_i^2)$. The closeness of the first pair follows by construction since $\delta_i \to 0$, and the closeness of the last pair follows from Proposition 3.1 because $\boldsymbol{\alpha}_i \to \boldsymbol{\alpha}$.

For the central pair, recall that G_i' is actually contained in $G_{\boldsymbol{\alpha}_i}(n_i)$, so we only need to compare their number of edges. In fact, since we already established $o(n_i^2)$-closeness of the first and last pairs, it suffices to show that the difference between the number of edges in G_i and $G_{\boldsymbol{\alpha}}(n_i)$ is $o(n_i^2)$. Recall from above that $\mathrm{E}(\boldsymbol{\alpha}) = \gamma$, and therefore by Proposition 3.1, $G_{\boldsymbol{\alpha}}(n_i)$ has $\mathrm{E}(\boldsymbol{\alpha})n_i^2 + o(n_i^2) = (\gamma + o(1))n_i^2$ edges. Yet G_i also has $(\gamma + o(1))n_i^2$ edges, because $m_i/n_i^2 \to \gamma$. This completes the proof. $\qquad\square$

3.6. *Proofs for the sparse case*

In this section, we prove the statements which refine our results in the case when the graph is sparse, that is, when $m = o(n^2)$. We begin with the lemma which shows that every sparse graph with the maximum number of colorings has a dense core which spans all of the edges.

Proof of Lemma 3.4. Let n_1 be the number of nonisolated vertices in G, and let r be the number of connected components in the subgraph induced by the nonisolated vertices. Since all such vertices there have degree at least 1, we have $r \leqslant n_1/2$.

Any connected graph on t vertices has at most $q(q-1)^{t-1}$ proper q-colorings, because we may iteratively color the vertices along a depth-first-search tree rooted at an arbitrary vertex; when we visit any vertex other than the root, there will only be at most $q - 1$ colors left to choose from. Therefore, G has at most $q^{n-n_1} \cdot q^r \cdot (q-1)^{n_1-r}$ colorings, where the first factor comes from the fact that isolated vertices have a free choice over all q colors. Using $r \leqslant n_1/2$, this bound is at most $q^{n-n_1/2}(q-1)^{n_1/2}$.

But since G is optimal, it must have at least as many colorings as the Turán graph $T_q(n_2)$ plus $n - n_2$ isolated vertices, where $n_2 = \Theta(\sqrt{m})$ is the minimum number of vertices in a q-partite Turán graph with at least m edges. The isolated vertices already give the latter graph at least

q^{n-n_2} colorings, so we must have $q^{n-n_2} \leqslant q^{n-n_1/2}(q-1)^{n_1/2}$, which implies that

$$n_1 \leqslant n_2 \cdot (2\log q)/\left(\log \frac{q}{q-1}\right). \tag{2}$$

The expression on the right-hand side is $\Theta(n_2) = \Theta(\sqrt{m})$, so if we define the integer n_0 to be the maximum of right-hand side in (2) and $\sqrt{m/\kappa}$ (rounding up to the next integer if necessary) then we indeed have $n_1 \leqslant n_0 = \Theta(n_2) = \Theta(\sqrt{m})$. □

Next, we prove the first part of Proposition 3.6, which claims that the maximum number of q-colorings of an n-vertex graph with $m \leqslant \kappa_q n^2$ edges is asymptotically $q^n e^{(-c+o(1))\sqrt{m}}$, where

$$\kappa_q = \left(\sqrt{\frac{\log(q/(q-1))}{\log q}} + \sqrt{\frac{\log q}{\log(q/(q-1))}} \right)^{-2} \quad \text{and} \quad c = 2\sqrt{\log \frac{q}{q-1} \cdot \log q}.$$

Proof of Proposition 3.6(i). Let G be an n-vertex graph with m edges, which maximizes the number of q-colorings. Let n_0 be the integer obtained by applying Lemma 3.4 with threshold κ_q. If $n \geqslant n_0$, then the lemma gives a dense n_0-vertex subgraph $G' \subset G$ which contains all of the edges. Otherwise, set $G' = G$. In either case, we obtain a graph G' whose number of vertices n' is $\Theta(\sqrt{m})$, and $m/(n')^2 \leqslant \kappa_q$.

Since the vertices in $G \setminus G'$ (if any) are isolated, the number of q-colorings of G is precisely $q^{n-n'}$ times the number of q-colorings of G'. Therefore, G' must also have the maximum number of q-colorings over all n'-vertex graphs with m edges. Applying Theorem 3.2 to G', we find that G' has $e^{(\text{OPT}(m/(n')^2)+o(1))n'}$ colorings. Proposition 3.5 gives us the precise answer $\text{OPT}(m/(n')^2) = \log q - 2\sqrt{\frac{m}{(n')^2} \cdot \log \frac{q}{q-1} \cdot \log q}$, so substituting that gives us that the number of q-colorings of G is:

$$q^{n-n'} \cdot e^{(\text{OPT}(m/(n')^2)+o(1))n'} = q^{n-n'} \cdot q^{n'} e^{(-c+o(1))\sqrt{m}} = q^n e^{(-c+o(1))\sqrt{m}},$$

where c is indeed the same constant as claimed in the statement of this proposition. □

We finish this section by proving the stability result which shows that any optimal sparse graph is ϵm-close (in edit-distance) to the graph $G_{n,m}$ defined in Section 3.2.

Proof of Proposition 3.6(ii). Let G be an n-vertex graph with m edges, which maximizes the number of q-colorings. We will actually show the equivalent statement that G is $O((\epsilon + \sqrt{\epsilon})m)$-close to $G_{n,m}$.

As in the proof of part (i) above, we find a dense n'-vertex subgraph $G' \subset G$ that spans all of the edges, which itself must maximize the number of q-colorings. Using the same parameters as above, we have $n' = \Theta(\sqrt{m})$ and $m \leqslant \kappa_q(n')^2$. By Theorem 3.3, G' must be $\epsilon(n')^2$-close to a graph $G_{\boldsymbol{\alpha}}(n')$ from Construction 1, for some $\boldsymbol{\alpha}$ that solves $\text{OPT}(\gamma)$ with $\gamma \leqslant \kappa_q$. Since $n' = \Theta(\sqrt{m})$, the graphs are $O(\epsilon m)$-close. The γ is within the range in which Proposition 3.5 solved Optimization Problem 1, so $G_{\boldsymbol{\alpha}}(n')$ is a complete bipartite graph plus isolated vertices, which indeed resembles $G_{n,m}$.

Moreover, the ratio between the sizes of the sides of the complete bipartite graph in $G_{\boldsymbol{\alpha}}(n')$ is correct, because it tends to the constant $\log \frac{q}{q-1}/\log q$ regardless of the value of γ. Also, their product, which equals the number of edges in $G_{\boldsymbol{\alpha}}(n')$, is within $O(\epsilon m)$ of m because $G_{\boldsymbol{\alpha}}(n')$ is $O(\epsilon m)$-close to the m-edge graph G'. Therefore, each of the sides of the complete bipartite graph in $G_{\boldsymbol{\alpha}}(n')$ differs in size from its corresponding side in $G_{n,m}$ by at most $O(\sqrt{\epsilon m})$. Since each side of the bipartite graph in $G_{n,m}$ has size $\Theta(\sqrt{m})$, we can transform $G_{\boldsymbol{\alpha}}(n')$ into $G_{n,m}$ by adding isolated vertices and editing at most $O(\sqrt{\epsilon} \cdot m)$ edges. Yet by the construction of $\boldsymbol{\alpha}$, the graphs G' and $G_{\boldsymbol{\alpha}}(n')$ were $O(\epsilon m)$-close, modulo isolated vertices. Therefore, G and $G_{n,m}$ are indeed $O((\epsilon + \sqrt{\epsilon})m)$-close, as claimed. □

4. *Solving the optimization problem*

In this section, we solve the optimization problem for low densities, for all values of q. We also solve it for all densities in the case when $q = 3$.

4.1. *Sparse case*

The key observation is that when the edge density is low, we can reduce the optimization problem to one with no edge density parameter and no vertex constraint. This turns out to be substantially easier to solve.

OPTIMIZATION PROBLEM 2.   Fix an integer q, and consider the following objective and constraint functions:

$$\text{OBJ}^*(\boldsymbol{\alpha}) := \sum_A \alpha_A \log \frac{|A|}{q}; \quad \text{E}(\boldsymbol{\alpha}) := \sum_{A \cap B = \emptyset} \alpha_A \alpha_B.$$

The vector $\boldsymbol{\alpha}$ has $2^q - 2$ coordinates $\alpha_A \in \mathbb{R}$ indexed by the nonempty *proper* subsets $A \subset [q]$, and the sum in $\text{E}(\boldsymbol{\alpha})$ runs over unordered pairs of disjoint sets $\{A, B\}$. Let FEAS^* be the feasible set of vectors defined by the constraints $\boldsymbol{\alpha} \geqslant 0$ and $\text{E}(\boldsymbol{\alpha}) \geqslant 1$. We seek to maximize $\text{OBJ}^*(\boldsymbol{\alpha})$ over the set FEAS^*, and we define OPT^* to be this maximum value, which we will show to exist in Section 4.1.1. We write that the vector $\boldsymbol{\alpha}$ *solves* OPT^* when both $\boldsymbol{\alpha} \in \text{FEAS}^*$ and $\text{OBJ}^*(\boldsymbol{\alpha}) = \text{OPT}^*$.

PROPOSITION 4.1.   *For any given $q \geqslant 3$, the unique solution (up to a permutation of the base set $[q]$) to Optimization Problem 2 is the vector $\boldsymbol{\alpha}^*$ with*

$$\alpha_{\{1\}}^* = \sqrt{\log \frac{q}{q-1} \Big/ \log q}, \quad \alpha_{\{2,\dots q\}}^* = \frac{1}{\alpha_{\{1\}}^*}, \quad \text{and all other } \alpha_A^* = 0.$$

This gives $\text{OBJ}^*(\boldsymbol{\alpha}^*) = -2\sqrt{\log \frac{q}{q-1} \cdot \log q}$.

Let us show how Proposition 4.1 implies Proposition 3.5, which gave the solution to Optimization Problem 1 for sufficiently low edge densities γ.

Proof of Proposition 3.5.   Let $\boldsymbol{\alpha}^*$ be the unique maximizer for Optimization Problem 2, and consider any number $t \geqslant \text{V}(\boldsymbol{\alpha}^*)$. Then $\boldsymbol{\alpha}^*$ is still the unique maximizer of $\text{OBJ}^*(\boldsymbol{\alpha})$ when $\boldsymbol{\alpha}$ is required to satisfy the vacuous condition $\text{V}(\boldsymbol{\alpha}) \leqslant t$ as well. Let $\overline{\boldsymbol{\alpha}}$ be the vector obtained by dividing every entry of $\boldsymbol{\alpha}^*$ by t, and adding a new entry $\overline{\alpha}_{[q]}$ so that $\text{V}(\overline{\boldsymbol{\alpha}}) = 1$.

Then, $\overline{\boldsymbol{\alpha}}$ is the unique maximizer of $\text{OBJ}^*(\boldsymbol{\alpha})$ when $\boldsymbol{\alpha}$ is constrained by $\text{V}(\boldsymbol{\alpha}) = 1$ and $\text{E}(\boldsymbol{\alpha}) \geqslant t^{-2}$. But when $\text{V}(\boldsymbol{\alpha}) = 1$ is one of the constraints, then $\text{OBJ}^*(\boldsymbol{\alpha}) = \text{OBJ}(\boldsymbol{\alpha}) - \log q$, so this implies that $\overline{\boldsymbol{\alpha}}$ is the unique solution to $\text{OPT}(t^{-2})$. Using the substitution $\gamma = t^{-2}$, we see that $\overline{\boldsymbol{\alpha}}$ is precisely the vector described in (1). Since $t \geqslant \text{V}(\boldsymbol{\alpha}^*)$ was arbitrary, we conclude that this holds for all γ below

$$\text{V}(\boldsymbol{\alpha}^*)^{-2} = \left(\sqrt{\frac{\log(q/(q-1))}{\log q}} + \sqrt{\frac{\log q}{\log(q/(q-1))}} \right)^{-2} = \kappa_q. \qquad \square$$

4.1.1. *Observations for Optimization Problem 2.*

We begin by showing that OBJ^* attains its maximum on the feasible set FEAS^*. Since FEAS^* is clearly nonempty, there is some finite $c \in \mathbb{R}$ for which $\text{OPT}^* \geqslant c$. In the formula for OBJ^*, all coefficients $\log(|A|/q)$ of the α_A are

negative, so we only need to consider the compact region bounded by $0 \leqslant \alpha_A \leqslant c/(\log(|A|/q))$ for each A. Therefore, by compactness, OBJ^* indeed attains its maximum on FEAS^*.

Now that we know the maximum is attained, we can use perturbation arguments to determine its location. The following definition will be convenient for our analysis.

DEFINITION 4.2. Let the *support* of a vector $\boldsymbol{\alpha}$ be the collection of A for which $\alpha_A \neq 0$.

The following lemma will allow us to reduce to the case of considering optimal vectors whose supports are a partition of $[q]$.

LEMMA 4.3. *One of the vectors $\boldsymbol{\alpha}$ which solves OPT^* has support that is a partition of $[q]$ (a collection of disjoint sets whose union is $[q]$). Furthermore, if the only partitions that support optimal vectors consist of a singleton plus a $(q-1)$-set, then in fact every vector which solves OPT^* is supported by such a partition.*

Proof.   We begin with the first statement. Let $\boldsymbol{\alpha}$ be a vector which solves OPT^*, and suppose that its support contains two intersecting sets A and B. We will perturb α_A and α_B while keeping all the other α fixed. Since A and B intersect, the polynomial $\mathrm{E}(\boldsymbol{\alpha})$ has no products $\alpha_A \alpha_B$, that is, it is of the form $x\alpha_A + y\alpha_B + z$, for some constants $x, y, z \geqslant 0$.

Furthermore, $x \neq 0$, or else we could reduce α_A to zero without affecting $\mathrm{E}(\boldsymbol{\alpha})$, but this would strictly increase $\mathrm{OBJ}^*(\boldsymbol{\alpha})$ because all coefficients $\log|A|/q$ in OBJ^* are negative. Similarly, $y \neq 0$. Therefore, we may perturb α_A by $+ty$ and α_B by $-tx$, while keeping $\mathrm{E}(\boldsymbol{\alpha})$ fixed. Since we may use both positive and negative t and OBJ^* itself is linear in α_A and α_B, optimality implies that OBJ^* does not depend on t. Hence we may choose a t which drives one of α_A or α_B to zero (we are free to pick which one), and OBJ^* will remain unchanged.

Repeating this process, we eventually obtain a vector $\boldsymbol{\alpha}$ which is supported by disjoint sets. Their union must be the entire $[q]$, because otherwise we could simply grow one of the sets in the support by adding the unused elements of $[q]$. This would not affect $\mathrm{E}(\boldsymbol{\alpha})$, but it would strictly increase OBJ^*.

It remains to prove the second part of our lemma. Let $\boldsymbol{\alpha}$ be an optimal vector, and apply the above reduction process to simplify its support. At the end, we will have a vector supported by $|A| = 1$ and $|B| = q - 1$, by assumption. Each iteration of the reduction removes exactly one set from the support, so the second to last stage will have some $\boldsymbol{\alpha}'$ supported by three distinct sets, two of which are the final A and B, and the third which we call C.

In the reduction, when we consider two overlapping sets, we are free to select which one is removed. Therefore, we could choose to keep the third set C and remove one of A and B, and then continue reducing until the support is disjoint, while keeping OBJ^* unchanged. Yet no matter what C was, it is impossible for this alternative reduction route to terminate in a partition of $[q]$, contradicting the above observation that any reduction must terminate in a partition. $\square$

DEFINITION 4.4. Let $\boldsymbol{\alpha}$ be a fixed vector whose support is a partition of $[q]$. For each $A \subset [q]$, define the expressions:

$$I_A = \alpha_A \sum_{B \neq A} \alpha_B, \quad J_A = \frac{1}{\mathrm{OBJ}^*(\boldsymbol{\alpha})} \cdot \alpha_A \log \frac{|A|}{q}.$$

LEMMA 4.5.   *Let $\boldsymbol{\alpha}$ be a vector which solves* OPT*, *whose support is a partition of* $[q]$. *Then:*

(i) *For every $A \subset [q]$, we have $I_A = 2J_A$. In particular, for each A in the support, $I_A/\alpha_A = 2J_A/\alpha_A$.*

(ii) *Suppose that A and B are both in the support, and $|A| = |B|$. Then $\alpha_A = \alpha_B$ as well.*

Proof.   We begin with part (i). Fix any $A \subset [q]$. Consider the following operation for small $\epsilon > 0$. First, replace α_A by $(1 + \epsilon)\alpha_A$. Observe that $I_A = \alpha_A \sum_{B:B\cap A=\emptyset} \alpha_B$ because the support of $\boldsymbol{\alpha}$ is a partition of $[q]$. Therefore we increase $\mathrm{E}(\boldsymbol{\alpha}) = \sum_{A\cap B=\emptyset} \alpha_A\alpha_B$ by ϵI_A. Next, multiply all the α (including the one we just increased) by $(1 + \epsilon I_A)^{-1/2}$. Then $\mathrm{E}(\boldsymbol{\alpha})$ is still at least 1 and our perturbed vector is in FEAS*. Its new objective equals $\mathrm{OBJ}^*(\boldsymbol{\alpha}) \cdot (1 + \epsilon J_A)/\sqrt{1 + \epsilon I_A}$. Since $\boldsymbol{\alpha}$ maximized the objective (which is always negative), we must have $(1 + \epsilon J_A)/\sqrt{1 + \epsilon I_A} \geqslant 1$. Rearranging, this implies that $I_A \leqslant 2J_A + \epsilon J_A^2$. Sending $\epsilon \to 0$, we see that $I_A \leqslant 2J_A$. The opposite inequality follows from considering the replacement of α_A by $(1 - \epsilon)\alpha_A$, and then multiplying α all the by $(1 - \epsilon I_A)^{-1/2}$. This establishes part (i).

For part (ii), let $S = \sum_C \alpha_C$. Since the support of $\boldsymbol{\alpha}$ is a partition of $[q]$, $S - \alpha_A = I_A/\alpha_A$. By part (i), this equals $2J_A/\alpha_A = \log(|A|/q)/\mathrm{OBJ}^*(\boldsymbol{\alpha})$, which is determined by the cardinality of A. Therefore, $S - \alpha_A = S - \alpha_B$, which implies (ii).   □

4.1.2.   *Solution to Optimization Problem 2 for $q < 9$.*   In its original form, Optimization Problem 2 involves exponentially many variables, but Lemma 4.3 dramatically reduces their number by allowing us to consider only supports that are partitions of $[q]$. Therefore, we need to make one computation per partition of $[q]$, which can actually be done *symbolically* (hence exactly) by *Mathematica*. The running time of *Mathematica*'s symbolic maximization is double-exponential in the number of variables, so it was particularly helpful to reduce the number of variables.[†]

Let us illustrate this process by showing what needs to be done for the partition $7 = 2 + 2 + 3$. This corresponds to maximizing $\alpha_A \log \frac{2}{7} + \alpha_B \log \frac{2}{7} + \alpha_C \log \frac{3}{7}$ subject to the constraints $\alpha_A\alpha_B + \alpha_B\alpha_C + \alpha_C\alpha_A \geqslant 1$ and $\boldsymbol{\alpha} \geqslant 0$. By Lemma 4.5(ii), we may assume $\alpha_A = \alpha_B$, so it suffices to maximize $2x \log \frac{2}{7} + y \log \frac{3}{7}$ subject to $x^2 + 2xy \geqslant 1$ and $x, y \geqslant 0$. This is achieved by *Mathematica*'s `Maximize` function:

```
Maximize[{2 x Log[2/7] + y Log[3/7], x^2 + 2 x y >= 1 && x >= 0
          && y >= 0}, {x, y}]
```

Mathematica answers that the maximum value is $-\sqrt{-(\log \frac{7}{3})^2 + 4 \log \frac{7}{3} \log \frac{7}{2}} \approx -1.9$, which is indeed less than the claimed value $-2\sqrt{\log \frac{7}{7-1} \cdot \log 7} \approx -1.1$.

We performed one such computation per partition of each $q \in \{3, \ldots, 8\}$. In every case except for the partition $q = 1 + (q - 1)$, the maximum indeed fell short of the claimed value. That final partition is completely solved analytically (that is, including the uniqueness result) by Lemma 4.6 in the next section. This completes the analysis for all $q < 9$.

4.1.3.   *Solution to Optimization Problem 2 for $q \geqslant 9$.*   We begin by ruling out several extreme partitions that our general argument below will not handle. As one may expect, each of these special cases has a fairly pedestrian proof, so we postpone the proofs of the following two lemmas to the appendix.

[†]The entire computation for $q \in \{3, \ldots, 8\}$ took less than an hour. The complete *Mathematica* program and output accompany the arXiv version of this paper.

LEMMA 4.6. *Fix any integer $q \geqslant 3$, and let $\boldsymbol{\alpha}$ be a vector which solves* OPT*. *If the support of $\boldsymbol{\alpha}$ is a partition of $[q]$ into exactly two sets, then (up to permutation of the ground set $[q]$) $\boldsymbol{\alpha}$ must be equal to the claimed unique optimal vector $\boldsymbol{\alpha}^*$ in Proposition 4.1.*

LEMMA 4.7. *Fix any integer $q \geqslant 4$, and let $\boldsymbol{\alpha}$ be a vector which solves* OPT*, *whose support is a partition of $[q]$. Then that partition cannot have any of the following forms:*
  (i) *all singletons;*
  (ii) *all singletons, except for one 2-set;*
  (iii) *have a $(q-2)$-set as one of the parts.*

The heart of the solution to the optimization problem is the following general case, which we will prove momentarily.

LEMMA 4.8. *Fix any integer $q \geqslant 9$, and let $\boldsymbol{\alpha}$ be a vector which solves* OPT*, *whose support is a partition of $[q]$. Then that partition must have a set of size at least $q - 2$.*

These collected results show that OPT* has the unique solution that we claimed at the beginning of this section.

Proof of Proposition 4.1 for $q \geqslant 9$. Let $\boldsymbol{\alpha}$ be a vector which solves OPT*. By Lemma 4.3, we may assume that its support is a partition of $[q]$. It cannot be a single set (of cardinality q), because indices are proper subsets of $[q]$, and by Lemmas 4.7(iii) and 4.8, the support cannot contain a set of size $\leqslant q - 2$.

Thus, the support must contain a set of size $q - 1$, and since it is a partition, the only other set is a singleton. Then Lemma 4.6 gives us that $\boldsymbol{\alpha}$ equals the claimed unique optimal vector $\boldsymbol{\alpha}^*$, up to a permutation of the ground set $[q]$. This completes the proof. □

In the remainder of this section, we prove the general case (Lemma 4.8). The following definition and fact are convenient, but the proof is a routine calculus exercise, so we postpone it to the appendix.

LEMMA 4.9. *Define the function $F_q(x) = \log \frac{q}{q-x} \cdot \log \frac{q}{x}$.*
  (i) *For $q > 0$, $F_q(x)$ strictly increases on $0 < x < q/2$ and strictly decreases on $q/2 < x < q$.*
  (ii) *For $q \geqslant 9$, we have the inequality $F_q(3) > 2F_q(1) \cdot (q-3)/(q-2)$.*

Proof of Lemma 4.8. Assume for the sake of contradiction that all sets in the support of the optimal $\boldsymbol{\alpha}$ have size at most $q - 3$. In terms of the expressions I and J from Definition 4.4, we have the following equality, where the sums should be interpreted as only over sets in the support of $\boldsymbol{\alpha}$:

$$\frac{2\log(|A|/q)}{\text{OBJ}^*(\boldsymbol{\alpha})} = \frac{2J_A}{\alpha_A} = \frac{I_A}{\alpha_A} = \sum_{B \neq A} \alpha_B = \sum_{B \neq A} \frac{J_B \cdot \text{OBJ}^*(\boldsymbol{\alpha})}{\log(|B|/q)}.$$

(The second equality is Lemma 4.5(i), and the other three equalities come from the definitions of I and J.) Note that the above logarithms are always negative. It is cleaner to work with positive quantities, so we rewrite the above equality in the equivalent form:

$$\frac{2\log(q/|A|)}{\text{OBJ}^*(\boldsymbol{\alpha})} = \sum_{B \neq A} \frac{J_B \cdot \text{OBJ}^*(\boldsymbol{\alpha})}{\log(q/|B|)}.$$

Since every B in the above sum is disjoint from A and we assumed all sets in the support have size at most $q - 3$, we have that every B above has size $|B| \leqslant q - \max\{|A|, 3\}$. This gives the upper bound:

$$\frac{2 \log(q/|A|)}{\text{OBJ}^*(\boldsymbol{\alpha})} \leqslant \sum_{B \neq A} \frac{J_B \cdot \text{OBJ}^*(\boldsymbol{\alpha})}{\log(q/(q - \max\{|A|, 3\}))}$$

$$\frac{2 \cdot \log(q/|A|) \cdot \log(q/(q - \max\{|A|, 3\}))}{\text{OBJ}^*(\boldsymbol{\alpha})^2} \leqslant \sum_{B \neq A} J_B.$$

Since $|A| \leqslant \max\{|A|, 3\}$, the left-hand side is at least $2F_q(\max\{|A|, 3\})/\text{OBJ}^*(\boldsymbol{\alpha})^2$. Also, $F_q(x)$ is symmetric about $x = q/2$ and we assumed that $3 \leqslant q/2$ and $|A| \leqslant q - 3$, so Lemma 4.9(i) implies that this is in turn at least $2F_q(3)/\text{OBJ}^*(\boldsymbol{\alpha})^2$. Lemma 4.9(ii) bounds this in terms of $F_q(1)$, which ultimately gives us the following bound for $\sum_{B \neq A} J_B$:

$$\frac{q-3}{q-2} \leqslant \frac{\text{OBJ}^*(\boldsymbol{\alpha}^*)^2}{\text{OBJ}^*(\boldsymbol{\alpha})^2} \cdot \frac{q-3}{q-2} = \frac{4F_q(1)}{\text{OBJ}^*(\boldsymbol{\alpha})^2} \cdot \frac{q-3}{q-2} < \frac{2F_q(3)}{\text{OBJ}^*(\boldsymbol{\alpha})^2} \leqslant \sum_{B \neq A} J_B. \qquad (3)$$

Here, $\boldsymbol{\alpha}^*$ is the claimed optimal vector in Proposition 4.1, and we recognize $4F_q(1) = \text{OBJ}^*(\boldsymbol{\alpha}^*)^2$. The first inequality follows from the maximality of $\boldsymbol{\alpha}$, and its direction is reversed because OBJ^* is always negative.

Let t be the number of sets in the support of $\boldsymbol{\alpha}$. Summing (3) over all sets A in the support:

$$t \cdot \frac{q-3}{q-2} < \sum_A \sum_{B \neq A} J_B = \sum_B J_B(t-1).$$

Yet $\sum_B J_B = 1$ by definition, so this implies $t/(t-1) < (q-2)/(q-3)$, which forces $t > q - 2$. Then, the support must be all singletons, except possibly for a single 2-set. This contradicts Lemma 4.7, and completes our proof. $\qquad \square$

4.2. *Solving the optimization problem for three colors*

In this section, we provide the complete analytic solution to Optimization Problem 1, for the entire range of the edge density parameter γ when the number of colors q is exactly 3. To simplify notation, we will write α_{12} instead of $\alpha_{\{1,2\}}$, etc.

Recall that κ_q with $q = 3$ is

$$\kappa_3 = \left(\sqrt{\frac{\log(3/2)}{\log 3}} + \sqrt{\frac{\log 3}{\log(3/2)}} \right)^{-2} \approx 0.1969.$$

PROPOSITION 4.10.   *The unique solution (up to a permutation of the index set* $\{1, 2, 3\}$*) of Optimization Problem 1 with edge density parameter* γ *is the vector* $\boldsymbol{\alpha}$ *defined as follows. (All unspecified* α_A *below are zero.)*

   (i) *If* $0 \leqslant \gamma \leqslant \kappa_3$*, then* $\alpha_3 = \sqrt{\gamma \cdot \log \frac{3}{2} / \log 3}$*,* $\alpha_{12} = \gamma/\alpha_3$*, and* $\alpha_{123} = 1 - \alpha_{12} - \alpha_3$*. This gives* $\text{OPT}(\gamma) = \log 3 - 2\sqrt{\gamma \cdot \log 3 \cdot \log \frac{3}{2}}$*.*

   (ii) *If* $\kappa_3 \leqslant \gamma \leqslant \frac{1}{4}$*, then* $\alpha_{12} = \frac{1+\sqrt{1-4\gamma}}{2}$ *and* $\alpha_3 = 1 - \alpha_{12}$*, which gives* $\text{OPT}(\gamma) = (1 + \sqrt{1-4\gamma})/2 \cdot \log 2$*.*

   (iii) *If* $\frac{1}{4} \leqslant \gamma \leqslant \frac{1}{3}$*, then* $\alpha_{12} = \frac{1-\sqrt{12\gamma-3}}{2}$*,* $\alpha_1 = \alpha_2 = \frac{1-2\alpha_{12}}{3}$*, and* $\alpha_3 = \frac{1+\alpha_{12}}{3}$*, which gives* $\text{OPT}(\gamma) = \frac{1-\sqrt{12\gamma-3}}{2} \cdot \log 2$*.*

This covers the entire range of admissible γ, because $\gamma = 1/3$ corresponds to the density of the Turán graph $T_3(n)$, which is the densest three-colorable graph.

4.2.1. *Outline of solution.* The strategy of the solution is as follows. Suppose that we have some $\boldsymbol{\alpha}$ that solves $\mathrm{OPT}(\gamma)$. Since we may permute the index set, we may assume without loss of generality that $\alpha_1 \leqslant \alpha_2 \leqslant \alpha_3$. We then use perturbation arguments to pinpoint the location of $\boldsymbol{\alpha}$. Although the problem initially looks cumbersome (there are seven nontrivially related variables), the solution cleanly follows from six short steps.

Step 1. By *shifting mass* (adjusting the values of the α_A while conserving their sum $\sum_A \alpha_A$) between the α_A with $|A| = 2$, we deduce that both α_{23} and α_{13} are zero.

Step 2. By smoothing together α_1 and α_2, we deduce that $\alpha_1 = \alpha_2$.

Step 3. By shifting mass between the variables α_A with $|A| = 1$, we reduce to one of the following two situations. Either $\alpha_1 = \alpha_2 = 0$, or $0 < \alpha_1 = \alpha_2 = \alpha_3 - \alpha_{12}$.

Step 4. We solve the first case resulting from Step 3, which is vastly simpler than the original problem. We find that the solution corresponds to outcomes (i) and (ii) of Proposition 4.10.

Step 5. It remains to consider the second case resulting from Step 3. By taking mass away from both α_{123} and α_1, and giving it to α_{12}, we conclude that $\alpha_{123} = 0$.

Step 6. We are left with the situation where the only nonzero variables are α_1, α_2, α_3, and α_{12}, and they are related by the equation $\alpha_1 = \alpha_2 = \alpha_3 - \alpha_{12}$. Again, this is vastly simpler than the original problem, and we find that its solution corresponds to outcome (iii) of Proposition 4.10.

4.2.2. *Details of solution.* We begin by recording a simple result that we will use repeatedly in the solution.

LEMMA 4.11. *Let* $\boldsymbol{\alpha}$ *be a vector that solves* $\mathrm{OPT}(\gamma)$. *Then* $\mathrm{E}(\boldsymbol{\alpha}) = \gamma$. *Furthermore, if* $\boldsymbol{\alpha}'$ *is obtained from* $\boldsymbol{\alpha}$ *by shifting mass from some* α_A *to another* α_B *with* $|A| = |B|$, *then* $\mathrm{E}(\boldsymbol{\alpha}') \leqslant \mathrm{E}(\boldsymbol{\alpha})$.

Proof. Suppose for contradiction that $\mathrm{E}(\boldsymbol{\alpha}) > \gamma$. The slack in the edge constraint lets us shift some more mass to α_{123} while keeping $\mathrm{E}(\boldsymbol{\alpha}) \geqslant \gamma$. But in the definition of OBJ, the coefficient $(\log 3)$ of α_{123} is the largest, so this shift strictly increases OBJ, contradicting maximality of $\boldsymbol{\alpha}$.

For the second claim, observe that OBJ is invariant under the shift since $|A| = |B|$. Now suppose for contradiction that $\mathrm{E}(\boldsymbol{\alpha}') > \mathrm{E}(\boldsymbol{\alpha})$. Then, as above, we could shift more mass to α_{123}, which would strictly increase OBJ, again contradicting the maximality of $\boldsymbol{\alpha}$. $\qquad\square$

Step 1. Consider shifting mass among $\{\alpha_{12}, \alpha_{23}, \alpha_{13}\}$. If we hold all other α_A constant, then $\mathrm{E}(\boldsymbol{\alpha}) = \alpha_1 \alpha_{23} + \alpha_2 \alpha_{13} + \alpha_3 \alpha_{12} + \text{constant}$, which is linear in the three variables of interest.

Let us postpone the uniqueness claim for a moment. Since we ordered $\alpha_1 \leqslant \alpha_2 \leqslant \alpha_3$, shifting all of the mass from $\{\alpha_{13}, \alpha_{23}\}$ to α_{12} will either strictly grow $\mathrm{E}(\boldsymbol{\alpha})$ if $\alpha_2 < \alpha_3$, or keep $\mathrm{E}(\boldsymbol{\alpha})$ unchanged. Also, $\mathrm{OBJ}(\boldsymbol{\alpha})$ will be invariant. Therefore, if we are only looking for an upper bound for $\mathrm{OPT}(\gamma)$, then we may perform this shift, and reduce to the case when $\alpha_{13} = 0 = \alpha_{23}$ without loss of generality.

We return to the topic of uniqueness. The next five steps of this solution will deduce that, conditioned on $\alpha_{13} = 0 = \alpha_{23}$, the unique optimal $\boldsymbol{\alpha}$ always has either $\alpha_2 < \alpha_3$ or $\alpha_{12} = \alpha_{13} = \alpha_{23} = 0$. We claim that this implies that our initial shift of mass to α_{12} *never happened*. Indeed, in the case with $\alpha_2 < \alpha_3$, the previous paragraph shows that an initial shift would have strictly increased $\mathrm{E}(\boldsymbol{\alpha})$, violating Lemma 4.11. And in the case with $\alpha_{12} = \alpha_{13} = \alpha_{23} = 0$, there was not even any mass at all to shift. Therefore, this will imply the full uniqueness result.

Step 2. Consider shifting mass between α_1 and α_2 until they become equal. If we hold all other α_A constant, then $\mathrm{E}(\boldsymbol{\alpha}) = \alpha_1\alpha_2 + (\alpha_1 + \alpha_2)\alpha_3 +$ constant. This 'smoothing' operation strictly increases the first term, while keeping the other terms invariant. But Lemma 4.11 prohibits $\mathrm{E}(\boldsymbol{\alpha})$ from increasing, so we conclude that we must have had $\alpha_1 = \alpha_2$.

Step 3. Consider shifting mass among $\{\alpha_1, \alpha_2, \alpha_3\}$. That is, fix $S = \alpha_1 + \alpha_2 + \alpha_3$, and vary $t = \alpha_3$ in the range $0 \leqslant t \leqslant S$. By Step 2, $\alpha_1 = \alpha_2 = (S - t)/2$. Step 1 gave $\alpha_{13} = \alpha_{23} = 0$, so we have

$$\mathrm{E}(\boldsymbol{\alpha}) = \alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3 + \alpha_{12}\alpha_3 = \frac{(S-t)^2}{4} + 2 \cdot \frac{S-t}{2} \cdot t + \alpha_{12}t$$

$$= -\frac{3}{4}t^2 + \left(\frac{S}{2} + \alpha_{12}\right)t + \frac{S^2}{4}.$$

By Lemma 4.11, $\alpha_3 = t$ must maximize this downward-opening parabola in the range $0 \leqslant t \leqslant S$. Recall that quadratics $f(x) = ax^2 + bx + c$ reach their extreme value at $x = -b/2a$, which corresponds to $t = -(S/2 + \alpha_{12})/(2 \cdot (-3/4)) = (S + 2\alpha_{12})/3$ above. Thus, if $(S + 2\alpha_{12})/3 < S$, then we must have $\alpha_3 = (S + 2\alpha_{12})/3 = (\alpha_1 + \alpha_2 + \alpha_3 + 2\alpha_{12})/3$. Step 2 gave us $\alpha_1 = \alpha_2$, which forces $0 < \alpha_1 = \alpha_2 = \alpha_3 - \alpha_{12}$. This is the second claimed outcome of this step.

On the other hand, if $(S + 2\alpha_{12})/3 \geqslant S$, then the quadratic is strictly increasing on the interval $0 \leqslant t \leqslant S$. Therefore, we must have $\alpha_3 = S$, forcing $\alpha_1 = \alpha_2 = 0$. This is the first claimed outcome of this step.

Step 4. In this case, only α_3, α_{12}, and α_{123} are nonzero. Then the edge constraint is simply $\mathrm{E}(\boldsymbol{\alpha}) = \alpha_3\alpha_{12} = \gamma$ (Lemma 4.11 forces equality). Note that since $\alpha_3 + \alpha_{12} \leqslant \mathrm{V}(\boldsymbol{\alpha}) = 1$, their product $\alpha_3\alpha_{12}$ is always at most $1/4$, so we can only be in this case when $\gamma \leqslant 1/4$.

Now let $x = \alpha_3$ and $y = \alpha_{12}$. The vertex constraint forces $\alpha_{123} = 1 - x - y$, so we are left with the routine problem of maximizing $\mathrm{OBJ} = y\log 2 + (1 - x - y)\log 3 = \log 3 - x\log 3 - y\log\frac{3}{2}$ subject to the constraints

$$x, y \geqslant 0, \quad x + y \leqslant 1, \quad xy = \gamma.$$

These constraints specify a segment of a hyperbola (a convex function) in the first quadrant of the xy-plane, and the objective is linear in x and y. Therefore, by convexity, the maximum would be at the global maximum of OBJ on the entire first quadrant branch of the hyperbola, unless that fell outside the segment, in which case it must be at an endpoint, forcing $x + y = 1$.

The maximum over the entire branch of $xy = \gamma$ follows easily from the inequality of arithmetic and geometric means: $\mathrm{OBJ} \leqslant \log 3 - 2\sqrt{x\log 3 \cdot y\log\frac{3}{2}} = \log 3 - 2\sqrt{\gamma \cdot \log 3 \cdot \log\frac{3}{2}}$, with equality when $x\log 3 = y\log\frac{3}{2}$. Using $xy = \gamma$ to solve for x and y, we see that the unique global maximum is at $x = \sqrt{\gamma \cdot \log\frac{3}{2}/\log 3}$ and $y = \sqrt{\gamma \cdot \log 3/\log\frac{3}{2}}$. This lies on our segment (satisfies $x + y \leqslant 1$) precisely when γ is below $\kappa_3 \approx 0.1969$ and these values of $\alpha_3 = x$ and $\alpha_{12} = y$ indeed match those claimed in that regime.

On the other hand, when $\gamma > \kappa_3$, we are outside the segment, so by the above we must have $x + y = 1$, and we may substitute $x = 1 - y$. We are left with the single-variable maximization of $\mathrm{OBJ} = y\log 2$ subject to $0 \leqslant y \leqslant 1$ and $(1 - y)y = \gamma$. By the quadratic formula, this is at $\alpha_{12} = y = (1 + \sqrt{1 - 4\gamma})/2 \leqslant 1$, which produces $\alpha_3 = x = 1 - y = 1 - \alpha_{12}$. This indeed matches outcome (ii) of our proposition.

Step 5. The remaining case is $0 < \alpha_1 = \alpha_2 = \alpha_3 - \alpha_{12}$, and we will show that this forces $\alpha_{123} = 0$. Indeed, suppose for the sake of contradiction that $\alpha_{123} > 0$. Shift mass to α_{12} by taking ϵ from α_{123} and $\epsilon' = \epsilon\alpha_3/\alpha_2$ from α_1. Since many α_A are zero, $\mathrm{E}(\boldsymbol{\alpha}) = \alpha_1(\alpha_2 + \alpha_3) + \alpha_2\alpha_3 + \alpha_{12}\alpha_3$. Our perturbation decreases the first term by $\epsilon'(\alpha_2 + \alpha_3)$, increases the third

term by $(\epsilon + \epsilon')\alpha_3$, and does not change the second term, so our choice of ϵ' keeps $\mathrm{E}(\boldsymbol{\alpha})$ invariant.

On the other hand, OBJ increases by $(\epsilon + \epsilon') \log 2 - \epsilon \log 3$. Since we know $\alpha_2 = \alpha_3 - \alpha_{12}$, in particular we always have $\alpha_3 \geqslant \alpha_2$, which implies that $\epsilon' \geqslant \epsilon$ because we assume $\alpha_2, \alpha_3 > 0$. Hence the increase in OBJ is $(\epsilon + \epsilon') \log 2 - \epsilon \log 3 \geqslant (\epsilon + \epsilon) \log 2 - \epsilon \log 3 > 0$, contradicting the maximality of $\boldsymbol{\alpha}$. Therefore, we must have had $\alpha_{123} = 0$.

Step 6. Now only α_1, α_2, α_3, and α_{12} remain. Let $t = \alpha_3$ and $r = \alpha_{12}$. Step 3 gives $\alpha_1 = \alpha_2 = \alpha_3 - \alpha_{12} = t - r$. We use the vertex constraint to eliminate t: $1 = \mathrm{v}(\boldsymbol{\alpha}) = 2(t - r) + t + r$, so $t = (1 + r)/3$. Substituting this for t, we are left with $\alpha_1 = \alpha_2 = (1 - 2r)/3$ and $\alpha_3 = (1 + r)/3$. Since we need all $\alpha_A \geqslant 0$, the range for r is $0 \leqslant r \leqslant 1/2$.

The above expressions give

$$\mathrm{E}(\boldsymbol{\alpha}) = \left(\frac{1 - 2r}{3}\right)^2 + 2\left(\frac{1 - 2r}{3}\right)\left(\frac{1 + r}{3}\right) + \left(\frac{1 + r}{3}\right)r = \frac{r^2 - r + 1}{3},$$

and Lemma 4.11 forces $\mathrm{E}(\boldsymbol{\alpha}) = \gamma$.

The quadratic formula gives the roots $r = (1 \pm \sqrt{12\gamma - 3})/2$. These are only real when $12\gamma - 3 \geqslant 0$, so this case only occurs when $\gamma \geqslant 1/4$. Furthermore, the only root within the interval $0 \leqslant r \leqslant 1/2$ is $r = (1 - \sqrt{12\gamma - 3})/2$. Plugging this value of r into the expressions for the α_A, we indeed obtain outcome (iii) of Proposition 4.10.

Conclusion. The only steps which proposed possible maxima were Steps 4 and 6. Conveniently, Step 4 also required that $\gamma \leqslant 1/4$, while Step 6 required $\gamma \geqslant 1/4$, so we do not need to compare them except at $\gamma = 1/4$, which is trivial. Finally, note that all extremal outcomes indeed have $\alpha_2 < \alpha_3$, except at $\gamma = 1/3$, in which case $\alpha_{12} = \alpha_{13} = \alpha_{23} = 0$. This justifies the uniqueness argument that we used at the end of Step 1, and completes our proof of Proposition 4.10.

5. Exact result for sparse graphs

In this section, we determine the precise structure of the sparse graphs that maximize the number of colorings, completing the proof of Theorem 1.2. Proposition 3.6(ii) showed that in this regime, the optimal graphs were close, in edit distance, to complete bipartite graphs. As a warm-up for the arguments that will follow in this section, let us begin by showing that the semi-complete subgraphs of Definition 1.1 are optimal among bipartite graphs. We will use this in the final stage of our proof of the exact result.

LEMMA 5.1. *Let $q \geqslant 3$ and $r < a \leqslant b$ be positive integers. Among all subgraphs of $K_{a,b}$ with r missing edges, the ones which maximize the number of q-colorings are precisely:*
  (i) *both the correctly and incorrectly oriented semi-complete subgraphs, when $q = 3$, and*
  (ii) *the correctly oriented semi-complete subgraph, when $q \geqslant 4$ and $\frac{b}{a} \geqslant \log q / \log \frac{q-2}{q-3}$ and a is sufficiently large (that is, $a > N_q$, where N_q depends only on q).*

REMARK. The above result is not as clean when more than three colors are used, but is sufficient for our purposes. In the sparse case, we encounter only highly unbalanced bipartite graphs, all of which have part size ratio approximately $\log q / \log \frac{q}{q-1}$. Apparently out of sheer coincidence (and good fortune), this is just barely enough to satisfy the additional condition of the lemma. Nevertheless, it would be nice to remove that condition.

Proof Lemma 5.1(ii). Let $A \cup B$ be the vertex partition of $K_{a,b}$, with $|A| = a$ and $|B| = b$. Let F^* be the correctly oriented semi-complete subgraph of $K_{a,b}$ with exactly r missing edges. Let F be another non-isomorphic subgraph of $K_{a,b}$ with the same number of edges. We will show that F has fewer colorings. Since F and F^* are both bipartite, they share every coloring that uses disjoint sets of colors on the sides of the bipartition. Discrepancies arise when the same color appears on both sides. Note, however, that whenever this occurs, every edge between same-colored vertices must be missing from the graph. This set of forced missing edges (in this lemma, *missing edges* refer only to those missing from the bipartite $K_{a,b}$, not the entire K_{a+b}), which we call the coloring's *footprint*, is always a union of vertex-disjoint complete bipartite graphs, one per color that appears on both sides. For each subset H of the missing edges of F, let n_H be the number of colorings of F with footprint H. Then, $\sum n_H$ is exactly the number of colorings of F. To give each n_H a counterpart from F^*, fix an arbitrary bijection ϕ between the missing edges of F and F^*, and let n_H^* be the number of colorings of F^* with footprint $\phi(H)$. Since F^* has $\sum n_H^*$ colorings, it suffices to show that $n_H \leqslant n_H^*$ for all H, with strict inequality for at least one H.

Clearly, when H is empty, or a star centered in B, then $n_H = n_H^*$. We observed that all footprints are unions $\Gamma_1 \cup \ldots \cup \Gamma_k$ of vertex-disjoint complete bipartite graphs, so all H not of that form automatically have $n_H = 0 \leqslant n_H^*$. It remains to consider H that have this form, but are not stars centered in B. Colorings with this footprint are monochromatic on each Γ_i, and there are $\binom{q}{k}k!$ ways to choose a distinct color for each Γ_i. The remaining $q - k$ colors are partitioned into two sets, one for $A \setminus V(H)$ and one for $B \setminus V(H)$. Crucially, $|B \setminus V(H)| \leqslant b - 2$ because H is not a star centered in B. Thus,

$$n_H \leqslant \left[\binom{q}{k}k! \right] \cdot \sum_{i=1}^{q-k-1} \binom{q-k}{i} i^{|A \setminus V(H)|} (q-k-i)^{|B \setminus V(H)|}$$

$$\leqslant q^k \cdot \sum_{i=1}^{q-k-1} \binom{q-k}{i} i^a (q-k-i)^{b-2}.$$

To see that the sum is dominated by the $i = 1$ term, note that since we assumed that $\frac{b}{a} \geqslant \log q / \log \frac{q-2}{q-3}$, for sufficiently large a we have

$$\frac{b-2}{a} \geqslant \log(q-1)/\log\frac{q-2}{q-3} \geqslant \log(q-k)/\log\frac{q-k-1}{q-k-2},$$

so we may apply Inequality B.2(ii) from the Appendix. This gives $n_H \leqslant q^k \cdot 1.1(q-k)(q-k-1)^{b-2}$. Next, we claim that this bound is greatest when k is smallest. Indeed, when k increases by one, q^k increases by the factor q, but $(q-k-1)^{b-2}$ decreases by a factor of at least $((q-2)/(q-3))^{b-2} \gg q$ for large b. Hence we have $n_H \leqslant 1.1q(q-1)(q-2)^{b-2}$.

On the other hand, $\phi(H)$ is always a star centered in B, so we can easily construct $q(q-1)(q-2)^{b-1}$ colorings of F^*. Indeed, choose one color for the vertices of the graph $\phi(H)$, a different color for the remainder of $A \setminus \phi(H)$, and allow each vertex left in $B \setminus \phi(H)$ to take any of the other $q - 2$ colors. Since $\phi(H)$ intersects B in exactly one vertex, $n_H^* \geqslant q(q-1)(q-2)^{b-1}$, as claimed. But $q - 2 \geqslant 2$, so we have the desired strict inequality $n_H^* \geqslant 2q(q-1)(q-2)^{b-2} > n_H$ for all remaining H. $\qquad \square$

Part (i) is a consequence of the following more precise result, which we will also need later.

LEMMA 5.2. *Let F be a subgraph of the complete bipartite graph $K_{a,b}$ with vertex partition $A \cup B$, and $r < \max\{a, b\}$ missing edges. Suppose that F has $x \in A$ and $y \in B$ with x complete to B and y complete to A. Then its number of three-colorings is precisely $3 \cdot 2^a + 3 \cdot 2^b - 6 + 6s$, where s is the number of nonempty subsets of missing edges which form complete bipartite graphs. This is at most $3 \cdot 2^a + 3 \cdot 2^b + 6 \cdot (2^r - 2)$, with equality exactly when the missing edges form a star.*

Proof. As in the proof of Lemma 5.1(ii), let n_H be the number of three-colorings of F with footprint H. The key observation is that for every nonempty H, $n_H = 6$ when H is a complete bipartite graph, and $n_H = 0$ otherwise. Indeed, if H is not a complete bipartite graph, then it cannot be a footprint of a three-coloring, so $n_H = 0$. Otherwise, there are three ways to choose a color for the vertices of H, and then by definition of footprint, the remaining two colors must be split between $A \setminus H$ and $B \setminus H$. Both of these sets are nonempty, because $A \setminus H$ must contain the given vertex x and $B \setminus H$ must contain y, so the only way to split the two colors is to use one on all of $A \setminus H$ and the other on all of $B \setminus H$. There are two ways to decide how to do this. Therefore, $n_H = 3 \cdot 2 = 6$, as claimed, and this produces the $6s$ in the formula.

The rest of the formula follows from $n_\emptyset = 3 \cdot 2^a + 3 \cdot 2^b - 6$. Indeed, the terms correspond to the colorings that use a single color (for which there are three choices) on B and allow the other two on A, those that use one on A and allow the others on B, and those that use only one on each of A and B (hence were double-counted). The final claim in the statement comes from the fact that stars are the only r-edge graphs which have all $2^r - 1$ of their nonempty subgraphs complete bipartite. $\square$

Proof of Lemma 5.1(i). Since the number of missing edges r is less than both $|A|$ and $|B|$, the vertices x and y of Lemma 5.2 must exist. Therefore, its equality condition implies that the optimal subgraphs are indeed semi-complete. $\square$

5.1. *Structure of proof*

We will use several small constants with relative order of magnitude $\epsilon_1 \ll \epsilon_2 \ll \epsilon_3$, related by $\epsilon_1 = \epsilon_2^2 = \epsilon_3^3$. We do not send them to zero; rather, we show that there is an eventual choice of the ϵ_i, determined by q and κ, that makes our argument work. Therefore, to avoid confusion, the O, Θ, and o notation that we employ in this proof will only mask constants depending on q, κ alone. For example, we will write $X = O(\epsilon_2 Y)$ when there is a constant $C_{q,\kappa}$ such that $|X| \leqslant C_{q,\kappa} \epsilon_2 |Y|$ for sufficiently large m and n. Occasionally, we will use phrases like 'almost all colorings have property P' when $(1 - o(1))$-fraction of all colorings have that property.

Proof of Theorem 1.2. Let $G = (V, E)$ be an optimal graph with n vertices and $m \leqslant \kappa n^2$ edges. We begin with a convenient technical modification: if G has an isolated edge xy, replace it with an edge between x and another non-isolated vertex of minimal degree. Do this only once, even if G had multiple isolated edges. The number of colorings stays the same because both graphs share the same partial colorings of $V \setminus \{x\}$, and each of those has exactly $q - 1$ extensions (in each graph) to the degree-1 vertex x.

This adjustment will not compromise the uniqueness claim, because it cannot create one of the optimal graphs listed in Theorem 1.2. Indeed, if it did, then the degree-1 vertex x would now have to be the center of the missing star of the semi-complete subgraph $H \subset K_{a,b}$. But we made x adjacent to a vertex of minimal degree, so x must be on the smaller side of H's bipartition. Then the number of $K_{a,b}$-edges missing from the semi-complete H is precisely $b - d(x) = b - 1$. This exceeds a for all optimal graphs listed in Theorem 1.2, but our definition of semi-completeness required that the number of missing edges was strictly less than the size of the smaller part. This contradiction shows that we may assume without loss of generality that if G has an isolated edge uv, then it also contains a degree-1 vertex $x \notin \{u, v\}$.

Define $u_1 = \sqrt{m \cdot \log \frac{q}{q-1} / \log q}$ and $u_2 = \sqrt{m \cdot \log q / \log \frac{q}{q-1}}$, and note that $\frac{u_1}{u_2} = \log \frac{q}{q-1} / \log q$ and $u_1 u_2 = m$. Therefore, Proposition 3.6(ii) gives disjoint subsets $U_1, U_2 \subset V$ of size $|U_i| = \lceil u_i \rceil$, such that by editing at most $\epsilon_1 m$ edges, we can transform G into the complete bipartite graph between U_1 and U_2, with all other vertices isolated. Call that graph G^*.

Let (V_1, V_2) be a max-cut partition of the *nonisolated* vertices of G, such that V_1 contains at least as many vertices of U_1 as V_2 does. We would like to show that this partition is very close

to (U_1, U_2), so we keep track of the U_i by defining $U_i' = U_i \cap V_i$ and $U_i'' = U_i \cap V_{3-i}$ for each $i \in \{1, 2\}$. To help us recognize vertices that are 'mostly correct,' let $X_i \subset U_i'$ be the vertices that are adjacent to all but at most $\epsilon_2 \sqrt{m}$ vertices of U_{3-i}'.

The following series of claims will complete the proof of Theorem 1.2, since Proposition 3.6(i) already determined the asymptotic maximum number of colorings.

CLAIM 1. *For each i, $|U_i'|$ is within $O(\epsilon_1 \sqrt{m})$ of u_i, $|X_i|$ is within $O(\epsilon_2 \sqrt{m})$ of u_i, and $|U_i''| \leqslant O(\epsilon_1 \sqrt{m})$.*

CLAIM 2. *Almost all colorings of G are (X_1, X_2)-regular, which means that they only use one color on X_1, and avoid that color on X_2.*

CLAIM 3. *At most one non-isolated vertex v_0 has degree $\leqslant 2\epsilon_3 \sqrt{m}$. We use this to show that each $|V_i|$ is within $O(\epsilon_2 \sqrt{m})$ of u_i. Let $V_0 = \{v_0\}$ if it exists; otherwise, let $V_0 = \emptyset$. Let $V_i^* = V_i \setminus V_0$.*

CLAIM 4. *Almost all colorings are (V_1^*, V_2^*)-regular, that is, use one color for V_1^*, and avoid it on V_2^*.*

CLAIM 5. *Each V_i^* is an independent set, and v_0 (if it exists) has neighbors in only one of the V_i^*. Hence G is a bipartite graph plus isolated vertices.*

CLAIM 6. *G is a semi-complete subgraph of $K_{|V_1|,|V_2|}$ plus isolated vertices, correctly oriented if $q \geqslant 4$.*

$\square$

5.2. Details of proof

Proof of Claim 1. We know that by editing at most $\epsilon_1 m$ edges, G can be transformed into G^*, the complete bipartite graph between (U_1, U_2), plus isolated vertices. Since $|U_i| = \lceil u_i \rceil = \Theta(\sqrt{m})$, all vertices in the U_i have degree $\Theta(\sqrt{m})$ in G^*. Therefore, the number of U_i-vertices that are isolated in G is at most $\epsilon_1 m / \Theta(\sqrt{m}) = O(\epsilon_1 \sqrt{m})$, implying in particular that the number of U_1-vertices in $V_1 \cup V_2$ is at least $|U_1| - O(\epsilon_1 \sqrt{m}) \geqslant \frac{2}{3} u_1$. (Recall that (V_1, V_2) is a max-cut partition of the *nonisolated* vertices of G.) Since more U_1-vertices are in V_1 than in V_2, and $U_1' = U_1 \cap V_1$, we have $|U_1'| \geqslant \frac{1}{3} u_1 = \Theta(\sqrt{m})$.

Also, G^* has at least m edges crossing between (U_1, U_2), so G has at least $m - \epsilon_1 m$ edges crossing between (U_1, U_2), and at least that many between its max-cut (V_1, V_2). As G has only m edges, this shows that each $G[V_i]$ spans at most $\epsilon_1 m$ edges. But the sets $U_1', U_2'' \subset V_1$ are complete to each other in G^*, so among the $\leqslant \epsilon_1 m$ edges of $G[V_1]$, at least $|U_1'||U_2''| - \epsilon_1 m$ of them must go between U_1' and U_2''. Combining this with the above result that $|U_1'| \geqslant \Theta(\sqrt{m})$, we obtain the desired bound $|U_2''| \leqslant O(\epsilon_1 \sqrt{m})$.

Then U_2', the set of U_2-vertices in V_2, has size at least $u_2 - O(\epsilon_1 \sqrt{m}) \geqslant \Theta(\sqrt{m})$, because only $O(\epsilon_1 \sqrt{m})$ of the U_2-vertices are isolated and $|U_2''| \leqslant O(\epsilon_1 \sqrt{m})$ of them are in V_1. Repeating the previous paragraph's argument with respect to U_2' and U_1'', we find that $|U_1''| \leqslant O(\epsilon_1 \sqrt{m})$, which then implies that $|U_1'| \geqslant u_1 - O(\epsilon_1 \sqrt{m})$.

It remains to control X_i, which we recall to be the vertices of U_i' which had at most $\epsilon_2 \sqrt{m}$ nonneighbors in U_{3-i}'. The U_i' are complete to each other in G^*, so each vertex not in X_i

contributes at least $\epsilon_2\sqrt{m}$ to the total edit distance of $\leqslant \epsilon_1 m$. We set $\epsilon_2^2 = \epsilon_1$, so this implies that all but at most $\epsilon_2\sqrt{m}$ vertices of U_i' belong to X_i. Since $|U_i'|$ is within $O(\epsilon_1\sqrt{m})$ of u_i, this gives the desired result. $\qquad\square$

Proof of Claim 2. We bound the number of colorings that are not (X_1, X_2)-regular. For each partition $[q] = C_0 \cup C_1 \cup C_2 \cup C_3$, we count the colorings which use the colors C_1 in X_1 but not X_2, use C_2 in X_2 but not X_1, use C_3 in both X_1 and X_2, and do not use C_0 in either X_1 or X_2. Then we sum over all *irregular* partitions, which are all partitions with $|C_1| \geqslant 2$ or $|C_3| \geqslant 1$. It suffices to show that the result is of smaller order than the total number of colorings of G.

For any given partition with $|C_i| = c_i$, we claim that the corresponding number of colorings is at most $(|X_1||X_2|)^{c_3} \cdot c_1^{|X_1|-q\epsilon_2\sqrt{m}} \cdot c_2^{|X_2|-q\epsilon_2\sqrt{m}} \cdot q^{n-2c_3-(|X_1|-q\epsilon_2\sqrt{m})-(|X_2|-q\epsilon_2\sqrt{m})}$. The first factor comes from choosing c_3 pairs of vertices $x_i \in X_1$, $y_i \in X_2$ on which to use each color of C_3. Then, every vertex in the common neighborhood of $\{y_i\}$ must avoid C_3 in order to produce a proper coloring. By definition of X_2, the number of vertices of U_1' that are not in this common neighborhood is at most $|C_3|\epsilon_2\sqrt{m} \leqslant q\epsilon_2\sqrt{m}$. Thus all but at most $q\epsilon_2\sqrt{m}$ vertices of $X_1 \subset U_1'$ are adjacent to every $\{y_i\}$, and therefore restricted to colors in C_1. This produces the second factor in our bound, and the third factor is obtained analogously. Of course every vertex has at most q color choices, and we use that trivial bound for all remaining vertices, producing our final factor. Using that each $|X_i|$ is within $O(\epsilon_2\sqrt{m})$ of $u_i = \Theta(\sqrt{m})$, we find that the sum Σ_1 of this bound over all $\leqslant 4^q$ irregular partitions is

$$\Sigma_1 = \sum_{\text{irregular}} (|X_1||X_2|)^{c_3} \cdot c_1^{|X_1|-q\epsilon_2\sqrt{m}} \cdot c_2^{|X_2|-q\epsilon_2\sqrt{m}} \cdot q^{n-2c_3-(|X_1|-q\epsilon_2\sqrt{m})-(|X_2|-q\epsilon_2\sqrt{m})}$$

$$\leqslant e^{O(\epsilon_2\sqrt{m})} \sum_{\text{irregular}} (\Theta(\sqrt{m}) \cdot \Theta(\sqrt{m}))^{c_3} \cdot c_1^{u_1} \cdot c_2^{u_2} \cdot q^{n-u_1-u_2}$$

$$\leqslant e^{O(\epsilon_2\sqrt{m})} \cdot 4^q \cdot O(m^q) \cdot \max_{c_1 \geqslant 2 \text{ or } c_3 \geqslant 1} \{c_1^{u_1} c_2^{u_2}\} \cdot q^{n-u_1-u_2}.$$

For any irregular partition with $c_1 + c_2 < q$, it is clear that $c_1^{u_1} c_2^{u_2}$ increases when C_1 is replaced by $C_1 \cup C_0 \cup C_3$, and C_0 and C_3 are reduced to $\emptyset$. It is also clear that this procedure gives another irregular partition, but this time with $c_1 + c_2 = q$. Yet $\frac{u_2}{u_1} = \log q / \log \frac{q}{q-1} \geqslant \log q / \log \frac{q-1}{q-2}$, so we may apply Inequality B.2(i), which gives

$$\max_{c_1 \geqslant 2 \text{ or } c_3 \geqslant 1} c_1^{u_1} c_2^{u_2} = 2^{u_1}(q-2)^{u_2} \leqslant 1.5^{-u_1} \cdot 1^{u_1}(q-1)^{u_2} = e^{-\Theta(\sqrt{m})} \cdot (q-1)^{u_2}.$$

Thus for small ϵ_2, we have $\Sigma_1 \leqslant e^{-\Theta(\sqrt{m})} \cdot (q-1)^{u_2} \cdot q^{n-u_1-u_2}$.

On the other hand, Proposition 3.6(i) shows that the optimal graph has at least $\Sigma_0 := q^n e^{(-c-\epsilon_1)\sqrt{m}}$ colorings, where $c = 2\sqrt{\log\frac{q}{q-1} \cdot \log q}$. Since $u_1 = \sqrt{m \cdot \log\frac{q}{q-1}/\log q}$ and $u_2 = \sqrt{m \cdot \log q / \log\frac{q}{q-1}}$, routine algebra shows that Σ_0 is precisely $e^{-\epsilon_1\sqrt{m}}(q-1)^{u_2}q^{n-u_1-u_2}$. Therefore, for small ϵ_1 we have $\Sigma_1/\Sigma_0 \leqslant e^{-\Theta(\sqrt{m})} = o(1)$, that is, almost all colorings of G are (X_1, X_2)-regular. $\qquad\square$

Before proving the next claim, it is convenient to establish the following lemma, which should be understood in the context of Claim 3.

LEMMA 5.3. *Let x, y be a pair of nonisolated vertices of G, such that xy is not an isolated edge. Then $d(x) + d(y) \geqslant |X_1| - 1$.*

Proof. Suppose for contradiction that there is such a pair x, y with $d(x) + d(y) \leqslant |X_1| - 2$. Let G' be the graph obtained by deleting the $\leqslant |X_1| - 2$ edges incident to x or y, and adding back as many edges between x and $X_1 \setminus \{x, y\}$. Define an $(X_1 \setminus \{x, y\}, X_2 \setminus \{x, y\})$-regular partial coloring to be a proper coloring of the vertices $V \setminus \{x, y\}$, which uses only one color on $X_1 \setminus \{x, y\}$, and avoids that color on $X_2 \setminus \{x, y\}$. Every such partial coloring of G' has exactly $q - 1$ extensions to x since only one color appears on $N_{G'}(x) \subset X_1 \setminus \{x, y\}$, and then exactly q further extensions to the newly-isolated vertex y. On the other hand, in G, x and y both have degree at least 1 and do not form an isolated edge, so one of them, say x, has a neighbor in the rest of the graph. Therefore, in G the same partial coloring has at most $q - 1$ extensions to the vertex x, and then at most $q - 1$ further extensions to the nonisolated vertex y. Yet by Claim 2, almost all colorings of G arise in this way, so for sufficiently large m, G has fewer colorings than G', contradiction. $\qquad \square$

Proof of Claim 3. Recall that our initial technical adjustment allows us to assume that if G contains an isolated edge uv, then it also contains a degree-1 vertex $x \notin \{u, v\}$. This would give $d(x) + d(u) = 2 \ll |X_1| - 1$, contradicting Lemma 5.3 because xu cannot be an isolated edge. Hence G in fact has no isolated edges. But then the same lemma implies that at most one vertex v_0 has degree $\leqslant 2\epsilon_3\sqrt{m}$, since $|X_1| = \Theta(\sqrt{m})$ by Claim 1.

It remains to show that each $|V_i|$ is within $O(\epsilon_2\sqrt{m})$ of u_i. Recall that U_1' and U_2'' are the U_1- and U_2-vertices that are in V_1. All other vertices of V_1 are isolated in the graph G^* which is within edit distance $\epsilon_1 m$ of G. Therefore by the previous paragraph, each of them (except v_0 if it exists) has degree at least $2\epsilon_3\sqrt{m}$, and thus contributes at least $2\epsilon_3\sqrt{m}$ to the edit distance between G and G^*. Therefore, there are at most $1 + (\epsilon_1 m)/(2\epsilon_3\sqrt{m}) \ll \epsilon_2\sqrt{m}$ of them, where we used $\epsilon_3^3 = \epsilon_2^2 = \epsilon_1$. Claim 1 controls $|U_i'|$ and $|U_i''|$, so we indeed find that $|V_1|$ is within $O(\epsilon_2\sqrt{m})$ of u_1. The analogous result for V_2 follows by a similar argument. $\qquad \square$

Proof of Claim 4. Since almost all colorings are (X_1, X_2)-regular, it suffices to prove this claim only for those colorings. Therefore, we bound the (X_1, X_2)-regular colorings that (i) use a common color on both V_2^* and V_1^*, or (ii) use at most $q - 2$ colors on V_2^*. Note that every (X_1, X_2)-regular coloring which avoids both (i) and (ii) must use exactly $q - 1$ colors on V_2^* and only the remaining color on V_1^*, and so is automatically (V_1^*, V_2^*)-regular. It therefore suffices to show that these two types of colorings constitute $o(1)$-fraction of all colorings. The key observation is that every $v \in V_2^*$ has a neighbor in X_1. Indeed, (V_1, V_2) is a max-cut, so at least half of the $\geqslant 2\epsilon_3\sqrt{m}$ neighbors of v must be in V_1. These cannot all avoid X_1, because Claims 1 and 3 show that only $O(\epsilon_2\sqrt{m})$ vertices of V_1 are outside X_1, and $\epsilon_2 \ll \epsilon_3$.

To bound the number of colorings of type (i) above, first choose a color c_1 for all X_1. By the key observation, c_1 cannot appear on V_2^*, so the shared color c_2 must be different. Hence we have $q - 1$ choices for c_2, and must pick a pair of vertices $x \in V_1^* \setminus X_1$ and $y \in V_2^*$ to use it on. The $\geqslant \epsilon_3\sqrt{m}$ neighbors of x in V_2^* must avoid c_2 as well as c_1, so they each have at most $q - 2$ color choices. Every other vertex of V_2^* must still avoid c_1, so we use the bound of $\leqslant q - 1$ color choices there. Using the trivial bound $\leqslant q$ for all other vertices, and the fact that $|X_i|$ and $|V_i^*|$ are within $O(\epsilon_2\sqrt{m})$ of $u_i = \Theta(\sqrt{m})$, we find that the number of type-(i) colorings is at most

$$
\Sigma_2 := q \cdot (q - 1) \cdot |V_1^* \setminus X_1| |V_2^*| \cdot (q - 2)^{\epsilon_3\sqrt{m}} \cdot (q - 1)^{|V_2^*| - \epsilon_3\sqrt{m}} \cdot q^{n - |X_1| - |V_2^*| - 1}
$$

$$
\leqslant O(m) \cdot \left(\frac{q - 2}{q - 1} \right)^{\epsilon_3\sqrt{m}} \cdot (q - 1)^{|V_2^*|} \cdot q^{n - |X_1| - |V_2^*| - 1}
$$

$$
\leqslant e^{O(\epsilon_2\sqrt{m})} \cdot \left(\frac{q - 2}{q - 1} \right)^{\epsilon_3\sqrt{m}} \cdot (q - 1)^{u_2} \cdot q^{n - u_1 - u_2}.
$$

On the other hand, we showed at the end of the proof of Claim 2 that G had at least $\Sigma_0 = e^{-\epsilon_1\sqrt{m}}(q-1)^{u_2}q^{n-u_1-u_2}$ colorings. Since $\epsilon_1 \ll \epsilon_2 \ll \epsilon_3$, we have $\Sigma_2/\Sigma_0 \leqslant e^{-\Theta(\epsilon_3\sqrt{m})} = o(1)$, as desired.

The number of type-(ii) colorings is easily bounded by $\Sigma_3 := q \cdot (q-1) \cdot (q-2)^{|V_2^*|} \cdot q^{n-|X_1|-|V_2^*|}$. The four factors correspond to choosing a color for X_1, choosing another color to avoid on V_2^*, coloring V_2^*, and coloring all remaining vertices. Using that $|X_i|$ and $|V_i^*|$ are within $O(\epsilon_2\sqrt{m})$ of u_i, we obtain $\Sigma_3 \leqslant e^{O(\epsilon_2\sqrt{m})}(q-2)^{u_2}q^{n-u_1-u_2}$, so $\Sigma_3/\Sigma_0 \leqslant e^{O(\epsilon_2\sqrt{m})}\left((q-2)/(q-1)\right)^{u_2}$. Since $u_2 = \Theta(\sqrt{m})$, for small enough ϵ_2 we indeed have $\Sigma_3/\Sigma_0 \leqslant e^{-\Theta(\sqrt{m})} = o(1)$, as desired. $\qquad\square$

Proof of Claim 5. Almost all colorings are (V_1^*, V_2^*)-regular, so $G[V_1^*]$ spans no edges. We turn our attention to V_2^*, and start by showing that all degrees within $G[V_2^*]$ are at most $\epsilon_3\sqrt{m}$. Indeed, suppose for contradiction that some $x \in V_2^*$ has at least $\epsilon_3\sqrt{m}$ neighbors in V_2^*. Then the number of (V_1^*, V_2^*)-regular colorings is at most $\Sigma_4 := q \cdot (q-1) \cdot (q-2)^{\epsilon_3\sqrt{m}} \cdot (q-1)^{|V_2^*|-\epsilon_3\sqrt{m}} \cdot q^{n-|V_1^*|-|V_2^*|}$. Here, the factors correspond to choosing a color c_1 for $|V_1^*|$, choosing a color c_2 for x, coloring $V_2^* \cap N(x)$ without c_1 or c_2, coloring the rest of V_2^* without c_1, and coloring the remaining vertices. Using that each $|V_i^*|$ is within $O(\epsilon_2\sqrt{m})$ of u_i, we find that

$$\Sigma_4 \leqslant e^{O(\epsilon_2\sqrt{m})} \cdot q \cdot (q-1) \cdot (q-2)^{\epsilon_3\sqrt{m}} \cdot (q-1)^{u_2-\epsilon_3\sqrt{m}} \cdot q^{n-u_1-u_2}$$

$$\leqslant e^{O(\epsilon_2\sqrt{m})} \cdot \left(\frac{q-2}{q-1}\right)^{\epsilon_3\sqrt{m}} \cdot (q-1)^{u_2}q^{n-u_1-u_2}.$$

Yet we showed at the end of the proof of Claim 2 that G had at least $\Sigma_0 = e^{-\epsilon_1\sqrt{m}}(q-1)^{u_2}q^{n-u_1-u_2}$ colorings, so using $\epsilon_1 \ll \epsilon_2 \ll \epsilon_3$, we obtain $\Sigma_4/\Sigma_0 \leqslant e^{-\Theta(\epsilon_3\sqrt{m})}$. This contradicts the fact that Σ_4 includes almost all colorings. Therefore, all degrees within $G[V_2^*]$ are indeed at most $\epsilon_3\sqrt{m}$.

We now use this intermediate bound to show that all such degrees are in fact zero. Suppose for contradiction that some $x \in V_2^*$ has neighbors within V_2^*. Let G' be the graph obtained by deleting all edges between x and V_2^* and all edges incident to v_0 (if it exists), and adding back as many edges between V_1^* and some formerly isolated vertex z.[†] This is possible because $d(v_0) \leqslant 2\epsilon_3\sqrt{m}$ and x has at most $\epsilon_3\sqrt{m}$ neighbors within V_2^*, while $|V_1^*| = \Theta(\sqrt{m})$. Observe that any $(V_1^*, V_2^* \setminus \{x\})$-regular partial coloring of $V \setminus \{x, z, v_0\}$ has exactly $(q-1)^2q^{|V_0|}$ extensions to all of G', because x and z only need to avoid the single color which appears on V_1^*, and v_0 is now isolated, if it exists. On the other hand, we claim that the same partial coloring has at most $(q-2)q(q-1)^{|V_0|}$ extensions in G. Indeed, there are at most $q-2$ extensions to x because x must avoid the color of V_1^* as well as some (different) color which appears on its neighbor in V_2^*. Then, there are q ways to color the isolated vertex z, and finally at most $q-1$ further extensions to the nonisolated vertex v_0 if it exists. Yet by Claim 2, almost all colorings of G arise in this way, so for sufficiently large m, G has fewer colorings than G'. This is impossible, so V_2^* must indeed be an independent set.

It remains to show that v_0, if it exists, has neighbors in only one V_i^*. Suppose for contradiction that v_0 is adjacent to both V_1^* and V_2^*, and consider the graph G' obtained by deleting all edges incident to v_0, and replacing them with edges to V_1^* only. This is possible because $d(v_0) \leqslant 2\epsilon_3\sqrt{m}$ and $|V_1^*| = \Theta(\sqrt{m})$. Any partial (V_1^*, V_2^*)-regular coloring of $G \setminus \{v_0\}$ has at most $q-2$ extensions to v_0, because neighbors of v_0 in V_2^* are colored differently from its neighbors in V_1^*. Yet the same partial coloring has exactly $q-1$ extensions with respect to G',

---

[†]Isolated vertices exist because Claim 3 shows that each $|V_i|$ is within $O(\epsilon_2\sqrt{m})$ of u_i, so the number of nonisolated vertices is $|V_1 \cup V_2| \leqslant u_1 + u_2 + O(\epsilon_2\sqrt{m})$. This is strictly below n for small ϵ_2, because $u_1 + u_2 = \sqrt{m/\kappa_q}$, and we assumed that $m \leqslant \kappa n^2$ with $\kappa < \kappa_q$.

since it uses the same color on all neighbors of v_0 (now in V_1^*). Therefore, for sufficiently large m, G' has more colorings than G, giving the required contradiction. $\quad\square$

Proof of Claim 6. First, consider the case when V_0 is empty. Then all nonisolated vertices are already in the bipartite graph (V_1^*, V_2^*). If that subgraph is less than $|V_1^*|$ edges away from being complete bipartite, then Lemma 5.1 already implies[†] that $G[V_1^* \cup V_2^*]$ is semi-complete (and correctly oriented if $q \geqslant 4$), so we are done. On the other hand, if that subgraph has at least $|V_1^*|$ missing edges, then we can construct an n-vertex graph G' with at least m edges by taking $K_{|V_1^*|, |V_2^*|-1}$ and adding enough isolated vertices. Then, G' has at least $q(q-1)^{|V_2^*|-1}q^{n-|V_1^*|-|V_2^*|+1}$ colorings because there are q choices of a single color for the $|V_1^*|$-side, $q-1$ color choices for each vertex on the other side, and q choices for each remaining (isolated) vertex. However, the same counting shows that G has exactly $q(q-1)^{|V_2^*|}q^{n-|V_1^*|-|V_2^*|}$ colorings that are (V_1^*, V_2^*)-regular, which includes almost all colorings by Claim 4. Hence for sufficiently large m, G' has more colorings, and this contradiction completes the case when V_0 is empty.

Now suppose that the vertex v_0 with degree at most $2\epsilon_3\sqrt{m}$ exists. By counting (V_1^*, V_2^*)-regular colorings, we find that G has at most $\Sigma_5 := (1+o(1))q(q-1)^{|V_2^*|}(q-1)q^{n-|V_1^*|-|V_2^*|-1}$ colorings. Here, the factors correspond to choosing a color for V_1^*, coloring V_2^*, coloring the nonisolated vertex v_0 which must avoid a neighbor's color, and coloring the remaining vertices. Observe that if there were at least $d(v_0)$ edges missing between V_1^* and V_2^*, then we could isolate v_0 by deleting its edges and adding back as many between V_1^* and V_2^*. The resulting graph would have at least $q(q-1)^{|V_2^*|}q^{n-|V_1^*|-|V_2^*|}$ colorings, where the factors correspond to choosing a color for V_1^*, coloring V_2^*, and coloring the remaining (isolated) vertices. For sufficiently large m, this exceeds the number of colorings of G, which is impossible. Therefore, less than $d(v_0)$ edges are missing between (V_1^*, V_2^*).

By Claim 5, v_0 has neighbors in only one V_i^*. If it is V_1^*, we must have $V_1 = V_1^*$ and $V_2 = V_2^* \cup \{v_0\}$ because (V_1, V_2) is a max-cut. The previous paragraph then implies that less than $|V_1|$ edges are missing between (V_1, V_2), so Lemma 5.1 shows that G is indeed semi-complete on its nonisolated vertices (and correctly oriented if $q \geqslant 4$).

The only remaining case is when v_0 has neighbors only in V_2^*, which we will show is impossible. This time, the max-cut gives $V_1 = V_1^* \cup \{v_0\}$ and $V_2 = V_2^*$. Since $d(v_0) \leqslant 2\epsilon_3\sqrt{m}$, there are at least $|V_2| - 2\epsilon_3\sqrt{m}$ missing edges between (V_1, V_2). Therefore, if we let $t = \lfloor(|V_2| - 2\epsilon_3\sqrt{m})/|V_1|\rfloor = \lfloor u_2/u_1 - O(\epsilon_3)\rfloor = \lfloor \log q/\log \frac{q}{q-1} - O(\epsilon_3)\rfloor$, we can construct an n-vertex graph G' with at least m edges by taking $K_{|V_1|, |V_2|-t}$ and adding enough isolated vertices. This graph has at least $\Sigma_6 := q(q-1)^{|V_2|-t}q^{n-|V_1|-|V_2|+t}$ colorings, by the same counting as earlier in this proof. Let us compare this with the number of colorings Σ_5 of G, which we calculated above. Since $|V_1^*| = |V_1| - 1$ and $|V_2^*| = |V_2|$, we have $\Sigma_6/\Sigma_5 \geqslant (1-o(1))(q/(q-1))^t \cdot 1/(q-1)$.

Crucially, $\log q/\log \frac{q}{q-1}$ is always irrational, because any positive integral solution to $q^x = (q/(q-1))^y$ would require q and $q-1$ to have a nontrivial common factor. Therefore, by choosing our ϵ_i sufficiently small in advance (based only on q), we may ensure that $t \geqslant \log q/\log \frac{q}{q-1} - 1 + c_q$ for some small positive constant c_q.

Yet

$$\left(\frac{q}{q-1}\right)^{\log q/\log \frac{q}{q-1}-1} \cdot \frac{1}{q-1} = 1,$$

so this gives $\Sigma_6/\Sigma_5 \geqslant (1-o(1))(q/(q-1))^{c_q}$, which exceeds 1 for large m, leaving G' with more colorings than G. This contradiction finishes our last case, and our entire proof. $\quad\square$

[†] V_1^* is the smaller side of the bipartite graph (V_1^*, V_2^*) because Claim 3 shows that $|V_1^*|$ is within $O(\epsilon_2\sqrt{m})$ of $u_1 = \sqrt{m \cdot \log \frac{q}{q-1}}/\log q$ and $|V_2^*|$ is within $O(\epsilon_2\sqrt{m})$ of $u_2 = \sqrt{m \cdot \log q/\log \frac{q}{q-1}}$.

6. *Exact result for three colors*

Our arguments can be pushed further when only three colors are used. In this section, we complete the proof of Theorem 1.3, determining the precise structure of the graphs that maximize the number of 3-colorings, for edge densities up to $m \leqslant \frac{1}{4}n^2$ (the density of the complete bipartite graph). The structure of this proof closely resembles that of the previous section, so parts that are essentially the same are rewritten briefly.

We would, however, like to draw attention to a new piece of notation. Recall that, as defined in the previous section, a coloring is (X, Y)-regular if it uses only one color on X but does not use this color on Y. This time, we will also need a symmetric version of this concept, which we denote with square brackets. We will say that a coloring is $[X, Y]$-*regular* if one of X or Y is monochromatic, and the other avoids that color entirely. Note that this is equivalent to having no colors shared between X and Y, because there are only three colors altogether.

Proof of Theorem 1.3. Theorem 1.2 already established our result for densities up to $m \leqslant \kappa n^2$ for some constant κ, so we may assume that $m = \Theta(n^2)$. Routine algebra verifies that Proposition 4.10 and Theorem 3.2 establish the claimed numbers of colorings in this theorem. This leaves us to concentrate on the optimal graph structure. We use several constants $\epsilon_1 \ll \epsilon_2 \ll \epsilon_3$, related by $\epsilon_1 = \epsilon_2^2 = \epsilon_3^3$, and show that there is an eventual choice that makes our argument work. To avoid confusion, our O, Θ, and o notation will only mask absolute constants.

Let $G = (V, E)$ be an optimal graph whose density m/n^2 is between κ and $1/4$. Let $u_1 = \alpha_3 n$ and $u_2 = \alpha_{12} n$, where the α are determined by Proposition 4.10 with density parameter $\gamma = m/n^2$. Note that since $\kappa \leqslant \gamma \leqslant \frac{1}{4}$, each $u_i = \Theta(n)$. Theorem 3.3 gives disjoint subsets $U_1, U_2 \subset V$ with $|U_i| \in \{\lfloor u_i \rfloor, \lceil u_i \rceil\}$, such that by editing at most $\epsilon_1 n^2$ edges, we can transform G into the complete bipartite graph between U_1 and U_2, plus isolated vertices. Call that graph G^*.

Let (V_1, V_2) be a max-cut partition of the nonisolated vertices of G, such that V_1 contains at least as many vertices of U_1 as V_2 does. Define $U_i' = U_i \cap V_i$ and $U_i'' = U_i \cap V_{3-i}$, and let $X_i \subset U_i'$ be the vertices that are adjacent to all but at most $\epsilon_2 n$ vertices of U_{3-i}'. The following series of claims will complete the proof of Theorem 1.3.

CLAIM 1. *For each* i, $|U_i'|$ *is within* $O(\epsilon_1 n)$ *of* u_i, $|X_i|$ *is within* $O(\epsilon_2 n)$ *of* u_i, *and* $|U_i''| \leqslant O(\epsilon_1 n)$.

CLAIM 2. *Almost all colorings of* G *are* $[X_1, X_2]$-*regular, meaning that one* X_i *is monochromatic, and the other* X_{3-i} *avoids that color entirely.*

CLAIM 3. *All nonzero degrees are at least* $2\epsilon_3 n$, *except possibly for either* (i) *only one isolated edge* $w_1 w_2$, *or* (ii) *only one nonisolated vertex* v_0. *We use this to show that each* $|V_i|$ *is within* $O(\epsilon_2 n)$ *of* u_i. *Let* $V_0 = \{w_1, w_2\}$ *if exception* (i) *occurs, let* $V_0 = \{v_0\}$ *if* (ii) *occurs, and let* $V_0 = \emptyset$ *otherwise. Let* $V_i^* = V_i \setminus V_0$.

CLAIM 4. *Almost all colorings are* $[V_1^*, V_2^*]$-*regular.*

CLAIM 5. *Each* V_i^* *is an independent set, and* v_0 (*if it exists*) *has neighbors in only one of the* V_i^*. *Hence* G *is a bipartite graph plus isolated vertices.*

CLAIM 6. G *is either a semi-complete subgraph of* $K_{|V_1|, |V_2|}$ *plus isolated vertices, or a complete bipartite subgraph* $K_{|V_1^*|, |V_2^*|}$ *plus a pendant edge to* v_0.

$\square$

6.1. Supporting claims

Proof of Claim 1. The sets $|U_i| = \Theta(n)$ are complete to each other in G^*, so all U_i-vertices have degree $\Theta(n)$ in G^*. As G is at most $\epsilon_1 n^2$ edges away from G^*, the number of U_i-vertices that are isolated in G is at most $\epsilon_1 n^2 / \Theta(n) = O(\epsilon_1 n)$. Since V_1 received more nonisolated U_1-vertices than V_2 did, we must have $|U_1'| \geqslant \frac{1}{3} u_1 = \Theta(n)$. By Proposition 3.1, G^* has at least $m - O(n)$ edges, all of which cross between (U_1, U_2). Therefore, G has at least $m - O(n) - \epsilon_1 n^2$ edges there, and at least that many between its max-cut (V_1, V_2). As G has only m edges, this shows that each $G[V_i]$ spans $O(\epsilon_1 n^2)$ edges. But the sets $U_1', U_2'' \subset V_1$ are complete to each other in G^*, so $|U_1'||U_2''| - \epsilon_1 n^2 \leqslant e(G[V_i]) \leqslant O(\epsilon_1 n^2)$. Using $|U_1'| \geqslant \Theta(n)$, we indeed obtain $|U_2''| \leqslant O(\epsilon_1 n)$.

Then $|U_2'| \geqslant u_2 - O(\epsilon_1 n) \geqslant \Theta(n)$, because only $O(\epsilon_1 n)$ of the U_2-vertices are isolated and $|U_2''| \leqslant O(\epsilon_1 n)$ of them are in V_1. Therefore, repeating the above with respect to U_2' and U_1'' instead of U_1' and U_2'', we find that $|U_1''| \leqslant O(\epsilon_1 n)$, which then implies that $|U_1'| \geqslant u_1 - O(\epsilon_1 n)$.

To control X_i, observe that since the U_i' are complete to each other in G^*, each vertex not in X_i contributes at least $\epsilon_2 n$ to the total edit distance of at most $\epsilon_1 n^2$ between G and G^*. We set $\epsilon_2^2 = \epsilon_1$, so all but at most $\epsilon_2 n$ vertices of U_i' belong to X_i. Since $|U_i'|$ is within $O(\epsilon_1 n)$ of u_i, this gives the desired result. $\qquad\square$

Proof of Claim 2. For each partition $\{1, 2, 3\} = C_0 \cup C_1 \cup C_2 \cup C_3$, we count the colorings which use the colors C_1 in X_1 but not X_2, use C_2 in X_2 but not X_1, use C_3 in both X_1 and X_2, and do not use C_0 in either X_1 or X_2. Then we sum over all *irregular* partitions, which are all partitions with $|C_3| \geqslant 1$. Note that a coloring is $[X_1, X_2]$-regular if and only if it does not use any color on both X_i, so this sum will include all other colorings.

For any given partition with $|C_i| = c_i$, we have that the corresponding number of colorings is at most $(|X_1||X_2|)^{c_3} \cdot c_1^{|X_1| - 3\epsilon_2 n} \cdot c_2^{|X_2| - 3\epsilon_2 n} \cdot 3^{n - 2c_3 - (|X_1| - 3\epsilon_2 n) - (|X_2| - 3\epsilon_2 n)}$, by the calculation in Claim 2 of Section 5.2 with q replaced by 3 and $\sqrt{m}$ replaced by n. Using that each $|X_i|$ is within $O(\epsilon_2 n)$ of $u_i = \Theta(n)$ and all irregular colorings have $|C_3| \geqslant 1 \Rightarrow c_1 + c_2 \leqslant 2$, we find that the sum Σ_1 of this bound over all $\leqslant 4^3$ irregular partitions is

$$\Sigma_1 = \sum_{\text{irregular}} (|X_1||X_2|)^{c_3} \cdot c_1^{|X_1| - 3\epsilon_2 n} \cdot c_2^{|X_2| - 3\epsilon_2 n} \cdot 3^{n - 2c_3 - (|X_1| - 3\epsilon_2 n) - (|X_2| - 3\epsilon_2 n)}$$

$$\leqslant e^{O(\epsilon_2 n)} \sum_{\text{irregular}} (\Theta(n) \cdot \Theta(n))^{c_3} \cdot c_1^{u_1} \cdot c_2^{u_2} \cdot 3^{n - u_1 - u_2}$$

$$\leqslant e^{O(\epsilon_2 n)} \cdot 4^3 \cdot O(n^6) \cdot \max_{c_1 + c_2 \leqslant 2} \{c_1^{u_1} c_2^{u_2}\} \cdot 3^{n - u_1 - u_2} = e^{O(\epsilon_2 n)} \cdot 3^{n - u_1 - u_2}.$$

On the other hand, Proposition 4.10, Theorem 3.2, and routine algebra show that just as in the sparse case, the optimal graph has at least $\Sigma_0 := e^{-\epsilon_1 n} \cdot 2^{u_2} \cdot 3^{n - u_1 - u_2}$ colorings. Using $u_2 = \Theta(n)$, we find that $\Sigma_1 / \Sigma_0 \leqslant e^{-\Theta(n)} = o(1)$, that is, almost all colorings of G are $[X_1, X_2]$-regular. $\qquad\square$

Before proving the next claim, it is convenient to establish the following lemma, which should be understood in the context of Claim 3.

LEMMA 6.1. *Let x, y be a pair of nonisolated vertices of G, such that xy is not an isolated edge. Then $d(x) + d(y) \geqslant \min\{|X_1|, |X_2|\} - 1$.*

Proof. Suppose for contradiction that there is such a pair x, y with $d(x) + d(y) \leqslant \min\{|X_1|, |X_2|\} - 2$. Also suppose that among the $[X_1 \setminus \{x, y\}, X_2 \setminus \{x, y\}]$-regular partial colorings of $V \setminus \{x, y\}$, at least half of them have $X_1 \setminus \{x, y\}$ monochromatic. (The case when

at least half have $X_2 \setminus \{x, y\}$ monochromatic follows by a similar argument.) Let G' be the graph obtained by deleting the $\leqslant |X_1| - 2$ edges incident to x or y, and adding back as many edges between x and $X_1 \setminus \{x, y\}$.

Consider any $[X_1 \setminus \{x, y\}, X_2 \setminus \{x, y\}]$-regular partial coloring of $V \setminus \{x, y\}$. If it is monochromatic in X_1, which happens at least half the time by assumption, then in G' it has exactly two extensions to x, followed by three further extensions to the newly isolated vertex y. The rest of the time, the partial coloring is monochromatic in X_2 and uses at most two colors in X_1. Then, in G' it has at least one extension to x, followed by three further extensions to y.

On the other hand, in G, x and y both have degree at least 1 and do not form an isolated edge, so one of them, say x, has a neighbor in the rest of the graph. Therefore, in G the same partial coloring has at most two extensions to the vertex x, and then at most two further extensions to the nonisolated vertex y. Yet by Claim 2, almost all colorings of G arise in this way, so the ratio of G'-colorings to G-colorings is at least $\frac{1}{2}(\frac{2 \cdot 3}{2 \cdot 2} + \frac{1 \cdot 3}{2 \cdot 2}) - o(1) = \frac{9}{8} - o(1) > 1$, contradiction. $\qquad\square$

Proof of Claim 3. If there is an isolated edge $w_1 w_2$, then Lemma 6.1 implies that any other vertex x has $d(x) + 1 = d(x) + d(w_1) \geqslant \min\{|X_1|, |X_2|\} - 1 = \Theta(n)$, giving exception (i). Otherwise, the same lemma implies there is at most one vertex v_0 of degree at most $2\epsilon_3 n$, giving exception (ii). The rest of this claim, that each $|V_i|$ is within $O(\epsilon_2 n)$ of u_i, follows by the same argument as in Claim 3 of Section 5.2, but with $\sqrt{m}$ replaced by n throughout. $\qquad\square$

Proof of Claim 4. Note that a coloring is $[V_1^*, V_2^*]$-regular if and only if it does not use any color on both V_i^*. Therefore, we bound the colorings that share a color on both V_i^*, but (i) use only one color on X_1 and a subset of the other two on X_2, or (ii) one on X_2 and a subset of the other two on X_1. Since almost all colorings are $[X_1, X_2]$-regular, it suffices to show that these two types of colorings constitute $o(1)$-fraction of all colorings. The same calculation as in Claim 4 of Section 5.2, with q replaced by 3 and $\sqrt{m}$ replaced by n, shows that the number of type-(i) colorings is at most

$$\Sigma_2 := 3 \cdot 2 \cdot |V_1^* \setminus X_1||V_2^*| \cdot 1^{\epsilon_3 n} \cdot 2^{|V_2^*| - \epsilon_3 n} \cdot 3^{n - |X_1| - |V_2^*| - 1}$$
$$\leqslant e^{O(\epsilon_2 n)} \cdot O(n^2) \cdot 2^{-\epsilon_3 n} \cdot 2^{u_2} \cdot 3^{n - u_1 - u_2}.$$

On the other hand, we showed at the end of the proof of Claim 2 that G had at least $\Sigma_0 = e^{-\epsilon_1 n} \cdot 2^{u_2} \cdot 3^{n - u_1 - u_2}$ colorings. Since $\epsilon_1 \ll \epsilon_2 \ll \epsilon_3$, we have $\Sigma_2 / \Sigma_0 \leqslant e^{-\Theta(\epsilon_3 n)} = o(1)$, as desired. The analogous result for type-(ii) colorings follows by a similar argument. $\qquad\square$

Proof of Claim 5. We first show that v_0 cannot have neighbors in both V_i^*. Suppose for contradiction that this is not the case. Almost all colorings are $[V_1^*, V_2^*]$-regular by Claim 4, so there is $I \in \{1, 2\}$ such that V_I^* is monochromatic in at least $(\frac{1}{2} - o(1))$-fraction of all colorings. Let G' be obtained by deleting all edges incident to v_0 (at most $2\epsilon_3 n$), and replacing them with edges to $|V_I^*| = \Theta(n)$ only. Consider any partial $[V_1^*, V_2^*]$-regular coloring of $V \setminus \{v_0\}$. If it uses only one color on V_I^* (which happens at least half the time by assumption), in G' it has exactly two extensions to v_0. The rest of the time, it still uses at most two colors on V_I^*, so there is at least one extension. On the other hand, in G the same partial coloring always has at most one extension to v_0, because the neighbors of v_0 in V_1^* are colored differently from the neighbors in V_2^*. By Claim 2, almost all colorings of G arise in this way, so the ratio of number of colorings of G' to G is at least $\frac{1}{2} \cdot (\frac{2}{1} + \frac{1}{1}) - o(1) = \frac{3}{2} - o(1)$, contradiction. Therefore, v_0 cannot have neighbors in both V_i^*, as claimed.

It remains to show that both $G[V_i^*]$ are empty. Suppose for contradiction that some $x \in V_2^*$ has neighbors within V_2^*. (The analogous result for V_1^* follows by a similar argument.) Almost every coloring is $[V_1^*, V_2^*]$-regular, but V_2^* can never be monochromatic because it contains

edges. Therefore, almost all colorings are in fact (V_1^*, V_2^*)-regular. (Recall that round brackets denote 'ordered' regularity, where V_1^* is monochromatic, and V_2^* avoids that color.) Therefore, the same argument as in Claim 5 of Section 5.2, with q replaced by 3 and $\sqrt{m}$ replaced by n, shows that x has at most $\epsilon_3 n$ neighbors within V_2^*.

Case 1: there is some $z_0 \in V_0$. Let G' be obtained by deleting the $\leqslant \epsilon_3 n$ edges between x and V_2^* and the $\leqslant 2\epsilon_3 n$ edges incident to anything in V_0, and adding back as many edges between z_0 and $|V_1^*| = \Theta(n)$. Every $(V_1^*, V_2^* \setminus \{x\})$-regular partial coloring of $V \setminus (V_0 \cup \{x\})$ has exactly $2 \cdot 2 \cdot 3^{|V_0|-1}$ extensions to all of G', because x and z_0 only need to avoid the single color which appears on V_1^*, and the rest of V_0 (if any) is now isolated. On the other hand, in G the same partial coloring has at most one extension to x because x must avoid the color of V_1^* as well as some (different) color which appears on its neighbor in V_2^*. Then, it has at most $3^{|V_0|-1}$ further extensions to $V_0 \setminus \{z_0\}$ by the trivial bound, and at most two further extensions to the non-isolated vertex z_0. Note that all (V_1^*, V_2^*)-regular colorings of G arise in this way, which is almost all of the total by our remark before we split into cases. Hence for sufficiently large m, G has fewer colorings than G', contradiction.

Case 2: $V_0 = \emptyset$, but there is some isolated vertex z. Define G' by deleting the $\leqslant \epsilon_3 n$ edges between x and V_2^*, and adding back as many edges between z and $|V_1^*| = \Theta(n)$. By the same arguments as in Case 1, all $(V_1^*, V_2^* \setminus \{x\})$-regular partial colorings of $V \setminus \{x, z\}$ have exactly $2 \cdot 2$ extensions to G', but in G they have at most one extension to x, followed by three further extensions to the isolated z. This produces almost all colorings of G, so G' has more colorings for large m, contradiction.

Case 3: $V_1^* \cup V_2^* = V$. We observed that the edges in V_2^* force almost all colorings to use only one color for V_1^* and the other two on V_2^* (hence $G[V_2^*]$ is bipartite). There are three color choices for V_1^*, so the number of colorings of G is $(3 + o(1)) \cdot \#\{2\text{-colorings of } V_2^*\}$. Recall that the number of two-colorings of any bipartite graph F is precisely 2^r, where r is its number of connected components.

We claim that the bipartite $G[V_2^*]$ has at most $|V_2^*| - 2\sqrt{t} + 1$ components, where t is the number of edges in $G[V_2^*]$. Indeed, for fixed t, the optimal configuration is to have all isolated vertices except for a single nontrivial (bipartite) component C. The sizes a, b of the sides of that bipartite C should minimize $a + b$ subject to the constraint $ab \geqslant t$, so by the inequality of the arithmetic and geometric means, we have $a + b \geqslant 2\sqrt{t}$, as desired. Therefore, G has at most $(3 + o(1)) \cdot 2^{|V_2^*|-2\sqrt{t}+1}$ colorings.

Let G' be the complete bipartite graph with sides s and $n - s$, such that s is as large as possible subject to $s(n - s) \geqslant m$. Note that $|V_1^*| \cdot |V_2^*| \geqslant m - t$ because all but t of the m edges of G cross between the V_i^*, so Inequality B.3 routinely shows that $s \geqslant |V_2^*| - \lceil\sqrt{t}\rceil$. Since G' is complete bipartite, it has exactly $3 \cdot 2^s + 3 \cdot 2^{n-s} - 6$ colorings, and thus our bound on s implies that G' has strictly more than $3 \cdot 2^s \geqslant 3 \cdot 2^{|V_2^*|-\lceil\sqrt{t}\rceil}$ colorings. Yet for $t \geqslant 3$, one may check that $-\lceil\sqrt{t}\rceil \geqslant (-2\sqrt{t} + 1) + 0.4$, giving G' more colorings than G, which is impossible.

We are left with the cases $t \in \{1, 2\}$, but for these values there is always a vertex $y \in V_2^*$ with exactly one neighbor z in $G[V_2^*]$. This forces all edges to be present between the V_i^*, because otherwise we could increase the number of (V_1^*, V_2^*)-regular colorings by a factor of 2 by deleting the edge yz and adding one of the missing edges between the V_i^*. The presence of the complete bipartite graph forces *every* coloring of G to use exactly two colors on V_2^*, and the other on V_1^*. Together with the observation that the maximum number of connected components of $G[V_2^*]$ is $|V_2^*| - t$ when $t \in \{1, 2\}$, we find that G has *exactly* $3 \cdot 2^r \leqslant 3 \cdot 2^{|V_2^*|-t}$ colorings. On the other hand, we showed above that G' had more than $3 \cdot 2^{|V_2^*|-\lceil\sqrt{t}\rceil}$ colorings. Since $t = \lceil\sqrt{t}\rceil$ for $t \in \{1, 2\}$, G' has more colorings than G, contradiction. $\square$

Proof of Claim 6. Let $G_0 = G[V_1 \cup V_2]$ be the graph formed by the nonisolated vertices of G, and let $n_0 = |V_1 \cup V_2|$. Since the number of colorings of G is precisely 3^{n-n_0} times the number of colorings of G_0, the optimality of G implies that G_0 must also be optimal among n_0-vertex graphs with m edges. Furthermore, Claim 4 also implies that almost all colorings of G_0 are $[V_1^*, V_2^*]$-regular.

Case 1: V_0 *is empty.* Let $\{a, b\}$ be the sizes of the V_i^*, with $a \leqslant b$. If there are less than a missing edges between the V_i^*, then Lemma 5.1 shows that G_0 is semi-complete, so we are done. On the other hand, if there are at least a missing edges, then $K_{a,b-1}$ plus one isolated vertex has n_0 vertices and at least m edges, but also exactly $(3 \cdot 2^a + 3 \cdot 2^{b-1} - 6) \cdot 3$ colorings. Yet G_0 has no vertices outside $V_1^* \cup V_2^*$, and almost all colorings are $[V_1^*, V_2^*]$-regular, so G_0 has at most $(1 + o(1)) \cdot (3 \cdot 2^a + 3 \cdot 2^b)$ colorings, which is smaller, contradiction.

Case 2: V_0 *is the single edge* $w_1 w_2$. We show that this is impossible. Let $\{a, b\}$ be the sizes of the V_i^*, with $a \leqslant b$. Since there are always exactly six ways to color the endpoints $\{w_1, w_2\}$ of the isolated edge independently of the rest of V, and almost all colorings are $[V_1^*, V_2^*]$-regular, G_0 has $(6 + o(1)) \cdot (3 \cdot 2^a + 3 \cdot 2^b)$ colorings. Let G' be the complete bipartite graph $K_{a-1,b+3}$, and let G'' be the complete bipartite graph $K_{a-1,b+2}$ plus one isolated vertex. Both graphs have the same number of vertices as G_0, so it suffices to show that at least one of them has more edges and more colorings than G_0.

Claim 3 gives $a/b \geqslant u_1/u_2 - O(\epsilon_2)$, and Proposition 4.10 implies that $u_1/u_2 \geqslant \log \frac{3}{2} / \log 3 \approx 0.37$. Therefore, for small ϵ_2 and large n, we have that $ab + 3a - b - 3 > ab + 1$, hence G' has more edges than G_0. Also, G' has $3 \cdot 2^{b+3} = 24 \cdot 2^b$ colorings that use only one color on the $(a - 1)$-side and the other two on the $(b + 3)$-side. We claim that this already exceeds the number of colorings of G_0 whenever $b \geqslant a + 2$. Indeed, then $2^a \leqslant \frac{1}{4} \cdot 2^b$, so the number of colorings of G_0 is at most:

$$(6 + o(1)) \cdot (3 \cdot 2^a + 3 \cdot 2^b) \leqslant (6 + o(1)) \cdot \frac{5}{4} \cdot 3 \cdot 2^b = (22.5 + o(1)) \cdot 2^b,$$

which is indeed less than the number of colorings of G'.

It remains to consider $a \leqslant b \leqslant a + 1$. Here, G'' has $ab + 2a - b - 2 > ab + 1$ edges, and exactly $(3 \cdot 2^{a-1} + 3 \cdot 2^{b+2} - 6) \cdot 3$ colorings. Using $a \geqslant b - 1$, this is at least $(1 - o(1)) \cdot \frac{17}{16} \cdot 3 \cdot 2^{b+2} \cdot 3 = (38.25 - o(1)) \cdot 2^b$. On the other hand, using $a \leqslant b$, the number of colorings of G_0 is at most $(36 + o(1)) \cdot 2^b$, which is smaller. Therefore, G'' is superior on this range, and we are done.

Case 3: V_0 *is the single vertex* v_0. Let I be the index (unique by Claim 5) such that V_I^* contains neighbors of v_0. Let $J = 3 - I$ be the other index, and let $a = |V_I^*|$, $b = |V_J^*|$. Note that G_0 is bipartite with partition $(V_I^*, V_J^* \cup \{v_0\})$. If at least $d(v_0)$ edges are missing between V_I^* and V_J^*, then we can isolate v_0 while only adding edges between V_I^* and V_J^*. This increases the number of $[V_I^*, V_J^*]$-regular colorings by a factor of $\frac{3}{2} + o(1)$, which is impossible. Therefore, less than $d(v_0)$ edges are missing between V_I^* and V_J^*, which implies that less than a edges are missing between V_I^* and $V_J^* \cup \{v_0\}$. Hence G_0 is a subgraph of $K_{a,b+1}$ with less than a missing edges.

When $a \leqslant b + 1$, Lemma 5.1 shows that G_0 is semi-complete, as desired. It remains to consider $a > b + 1$. Some vertex of the set V_I^* of size a is complete to $V_J^* \cup \{v_0\}$, because less than a edges are missing between V_I^* and $V_J^* \cup \{v_0\}$. But we also showed that less than $d(v_0) \leqslant 2\epsilon_3 n \ll |V_J^*|$ edges are missing between V_I^* and V_J^*, so some vertex of V_J^* must be complete to V_I^*. Thus, Lemma 5.2 implies that since G_0 is an optimal graph, the missing edges $E(K_{a,b+1}) \setminus E(G_0)$ form a star, which must have center v_0 because $d(v_0) \leqslant 2\epsilon_3 n \ll \min\{a, b\}$. In particular, the number of missing edges is then exactly $a - d$, where $d = d(v_0)$, and then the same lemma shows that G_0 has exactly $3 \cdot 2^a + 3 \cdot 2^{b+1} + 6 \cdot (2^{a-d} - 2)$ colorings.

Consider the graph G' obtained by removing a $(b - d)$-edge star from the complete bipartite graph $K_{a+1,b}$. This has as many vertices and edges as G_0, and $3 \cdot 2^{a+1} + 3 \cdot 2^b + 6 \cdot (2^{b-d} - 2)$ colorings by Lemma 5.2. The difference between the numbers of colorings of G' and G_0 is

$$3 \cdot 2^a - 3 \cdot 2^b + 6 \cdot (2^{b-d} - 2^{a-d}) = \left(3 - \frac{6}{2^d}\right) \cdot (2^a - 2^b),$$

which exceeds zero for $d \geqslant 2$ because we are in the case $a > b + 1$. Optimality of G_0 thus forces $d(v_0) = 1$.

We showed there were less than $d(v_0)$ edges missing between the V_i^*, so now we know that the nonisolated vertices of G form a complete bipartite subgraph (V_1^*, V_2^*) plus a pendant edge to v_0. Finally, observe that G cannot have any isolated vertex z, or else we could replace the pendant edge with the (isolated) edge $v_0 z$, and this would not change the number of colorings because every partial coloring of $V \setminus \{v_0\}$ would still have exactly two extensions to the degree-1 vertex v_0. But the resulting graph is not optimal by the same argument as in Case 2 of this claim. Therefore, G is only a complete bipartite subgraph plus a pendant edge, with no isolated vertices. This completes the final case of our final claim, and our entire proof. $\qquad\square$

7. Exact result for Turán graphs

We now study the extremality of Turán graphs. As we mentioned in the introduction, Lazebnik conjectured that Turán graphs $T_r(n)$ were the unique graphs that maximized the number of q-colorings whenever $r \leqslant q$. Note that Theorem 1.3 implies this result for $q = 3$ and $r = 2$ when n is large, because it shows that all optimal graphs are bipartite, and no other bipartite graph has as many edges as $T_2(n)$. In this section, we prove Theorem 1.4, which confirms (for large n) Lazebnik's conjecture when $r = q - 1$, for all remaining q. Our proof relies on the following special case of a result of Simonovits [25]. Let $t_r(n)$ denote the number of edges of the r-partite Turán graph $T_r(n)$ with n vertices.

FACT 7.1. Let F be a graph with chromatic number $r + 1$. Suppose that there is an edge whose deletion makes F r-colorable. Then for all sufficiently large n, the Turán graph $T_r(n)$ is the unique n-vertex graph with at least $t_r(n)$ edges that does not contain a subgraph isomorphic to F.

We use this fact to prove the following lemma, which we will need later.

LEMMA 7.1. Let $q \geqslant 4$ be fixed. The following holds for all sufficiently large n. Let $G \neq T_{q-1}(n)$ have n vertices, and at least as many edges and q-colorings as $T_{q-1}(n)$. Let Δ be the difference between the number of edges of G and $T_{q-1}(n)$, and let $n' = n - (q - 1)$. Then there is an n'-vertex graph H with at least $\Delta + 1$ more edges than $T_{q-1}(n')$, and at least half as many q-colorings as G has.

Proof. We begin with a convenient technical adjustment. If G has $k \geqslant 2$ connectivity components C_i that are not isolated vertices, then choose vertices $v_i \in C_i$ and glue the components together by merging all of the v_i into a single vertex v. Add $k - 1$ isolated vertices $w_1, \ldots, w_{k-1}$ to restore the vertex count, and let G' be the resulting graph. Clearly, G' has as many edges as G, and it also is not $T_{q-1}(n)$ because G' has a vertex whose deletion increases the number of components while $T_{q-1}(n)$ does not. Furthermore, we claim that G and G' have the same number of colorings. Indeed, by symmetry, for an arbitrary color c,

the total number of colorings of G is precisely q^k times the number of colorings of G which use c for every v_i. The obvious correspondence gives a bijection between these colorings and partial colorings of $G' \setminus \{w_1, \ldots, w_{k-1}\}$ which use c on the merged vertex v. Yet the w_i are isolated, so each of these partial colorings has exactly q^{k-1} extensions to all of G'. Again by symmetry, the total number of colorings of G' is precisely q times the number that use c on v. Setting everything together, we find that G and G' indeed have the same number of colorings. Therefore, by replacing G with G', we may assume without loss of generality that G has only one nontrivial connectivity component.

Fact 7.1 implies that for large n, G has a subgraph F which is the complete $(q-1)$-partite graph on $V(F) = X_1 \cup \ldots \cup X_{q-1}$ with each part $X_i = \{u_i, w_i\}$ consisting of two vertices, plus an extra edge $u_1 w_1$. Let $U = \{u_1, \ldots, u_{q-1}\}$ and $W = \{w_1, \ldots, w_{q-1}\}$, and let $A = U \cup \{w_1\}$.

Let δ be the difference between the number of edges of $T_{q-1}(n)$ and $T_{q-1}(n')$. We claim that if there is a set Y of $q-1$ vertices of A such that the sum of their degrees is at most $\delta + \binom{q-1}{2} - 1$, then $H = G - Y$ satisfies the lemma's assertion. Clearly, H has the correct number of vertices, and it has the correct number of edges because $Y \subset A$ induces a complete graph K_{q-1}, so the number of deleted edges is at most $\delta - 1$. We now show that every q-coloring of H extends to at most two q-colorings of G.

If $Y = U$, since $\{u_1\} \cup W$ induces a K_q-subgraph in G, every coloring of $H \supset W$ has at most one extension to u_1. Then, every other u_i has at most one choice because $\{u_1, u_i\} \cup (W \setminus \{w_i\})$ induces a K_q-subgraph in which u_i is the only uncolored vertex. Thus when $Y = U$, every coloring of H colors W and hence has at most one extension to G. On the other hand, up to a symmetry of F, the only other case is when $Y = \{w_1\} \cup (U \setminus \{u_{q-1}\})$. As before, $\{u_1\} \cup W$ induces a K_q-subgraph in G, but this time H contains neither u_1 nor w_1 (although it contains the rest). Any partial coloring of $q-2$ vertices of K_q has only two completions, so there are at most two ways to extend any coloring of H to include u_1 and w_1. But then every other u_i has at most one choice because $\{u_1, u_i\} \cup (W \setminus \{w_i\})$ induces a K_q-subgraph in which u_i is the only uncolored vertex. Therefore, every coloring of H has at most two extensions to G, as claimed.

It remains to consider the case when every set of $q-1$ vertices of A has degrees summing to at least $\delta + \binom{q-1}{2}$. We will show that then G has fewer colorings than $T_{q-1}(n)$, which is impossible. Let $B = V(G) \setminus A$. By an averaging argument, the sum of degrees of A is at least $\frac{q}{q-1}[\delta + \binom{q-1}{2}]$. Since $|A| = q$, the number of edges between A and B is at least $\frac{q}{q-1}[\delta + \binom{q-1}{2}] - 2\binom{q}{2}$.

Let B_0 be the set of isolated vertices of G, and for $2 \leqslant i \leqslant q-1$, let B_i be the set of vertices of B that send i edges to A. Note that no vertex can send $q = |A|$ edges to A because that would create a K_{q+1}-subgraph, making G not q-colorable. Therefore, if we let $B_1 = B \setminus (B_0 \cup B_2 \cup \ldots \cup B_{q-1})$, then every vertex of B_1 either sends exactly one edge to A, or it is a nonisolated vertex that sends no edges to A. Let $b_i = |B_i|$. By counting the number of edges between A and B, we obtain

$$\sum_{i=1}^{q-1} i b_i \geqslant \frac{q}{q-1}\left[\delta + \binom{q-1}{2}\right] - 2\binom{q}{2}. \tag{4}$$

We now bound the number of q-colorings of G in terms of the b_i. There are exactly $q!$ ways to color A because it induces K_q. Then, there are exactly q^{b_0} ways to extend this partial coloring to B_0 because each isolated vertex has a free choice of the q colors. Next, for every $i \in \{2, \ldots, q-1\}$, each vertex in B_i has at most $q - i$ color choices left because it is adjacent to i vertices in A, all of which received different colors since $G[A] = K_q$. Finally, we color the vertices of B_1 by considering them in an order such that whenever we color a vertex, it always has a neighbor that we already colored. This is possible because our initial technical adjustment allows us to assume that G has only one nontrivial connectivity component. Hence each vertex in B_1 will have at most $q - 1$ choices. Setting this all together, we find that the number of

q-colorings of G is at most

$$q! \cdot \prod_{i=0}^{q-1}(q-i)^{b_i} \leqslant q! \cdot \prod_{i=0}^{q-1} 2^{(q-i-1)b_i} \leqslant q! \cdot 2^{(q-1)(n-q)} \cdot 2^{-\frac{q}{q-1}\left[\delta+\binom{q-1}{2}\right]+2\binom{q}{2}},$$

where we used the inequality $x + 1 \leqslant 2^x$ for $x \in \mathbb{Z}$, the identity $\sum b_i = n - q$ (since $\cup B_i = V(G) \setminus A$), and the bound for $\sum i b_i$ from inequality (4). Inequality B.5 routinely verifies that this final bound is always strictly less than the number of colorings of $T_{q-1}(n)$, contradicting our assumption that G had at least that many colorings.   □

Proof of Theorem 1.4.   Let $q \geqslant 4$ be fixed, and let N be the corresponding minimum number of vertices for which Lemma 7.1 holds (it is valid only for sufficiently large n). We will show that Theorem 1.4 holds for all $n \geqslant q\binom{N}{2}$. Therefore, suppose for contradiction that $G \neq T_{q-1}(n)$ is an n-vertex graph with at least as many edges and q-colorings as $T_{q-1}(n)$.

Define a sequence of graphs as follows. Start with $G_0 = G$. If G_i is the current graph, stop if G_i has fewer colorings than the $(q-1)$-partite Turán graph with $n - (q-1)i$ vertices. Otherwise, let G_{i+1} be the graph H obtained by applying Lemma 7.1 to G_i. We claim that this process terminates before the graph G_i has fewer than N vertices, so we will always be able to apply the lemma. Indeed, each G_i has exactly $n - (q-1)i$ vertices, so it will take more than $\binom{N}{2}$ iterations before G_i has fewer than N vertices. Yet if $\Delta \geqslant 0$ is the difference between the number of edges of G and $T_{q-1}(n)$, then each G_i has at least $\Delta + i$ more edges than the $(q-1)$-partite Turán graph with $n - (q-1)i$ vertices. Therefore, after $\binom{N}{2}$ iterations, G_i would certainly have more than the maximum number of edges of an N-vertex graph, and we indeed can never reach a graph with fewer than N vertices.

Therefore, we stop at some G_t, which has $n' = n - (q-1)t$ vertices and fewer colorings than $T_{q-1}(n')$, but at least 2^{-t} times as many colorings as G. Divide n by $q - 1$, so that $n = s(q-1) + r$ with $0 \leqslant r < q - 1$, and note that $n' = (s-t)(q-1) + r$. Lemma B.4 calculates that $T_{q-1}(n')$ has exactly $q! \cdot \left[(q-1+r)2^{s-t-1} - q + 2\right]$ colorings, so G has at most 2^t times that many, hence fewer than $q! \cdot \left[(q-1+r)2^{s-1} - q + 2\right]$. Yet by the same lemma, that final bound equals the number of colorings of $T_{q-1}(n)$. Thus G has fewer colorings than $T_{q-1}(n)$, a contradiction.   □

8.   *Concluding remarks*

• We have developed an approach that we hope future researchers can use to determine the graphs that maximize the number of q-colorings. Theorems 3.2 and 3.3 reduce any instance of this problem to a quadratically constrained linear program, which can be solved for any case of interest. Thus, thanks to modern computer algebra packages, these theorems imply that for any fixed q, approximately determining the extremal graphs amounts to a finite symbolic computation.

The remaining challenge is to find analytic arguments which solve the optimization problem for general q, and then refine the approximate structure into precise results. We accomplished this for low densities m/n^2, and the natural next step would be to extend the result to the range $\frac{m}{n^2} \leqslant \frac{1}{4}$. In this range, and for all q, we expect the solution to the optimization problem to correspond to a bipartite graph plus isolated vertices. This common form gives hope that perhaps one can find a solution which works across all q.

• For $q = 3$, we also know the approximate form of the extremal graphs when $m/n^2 > \frac{1}{4}$, since Proposition 4.10 solved the entire $q = 3$ case of the optimization problem. However, we did not pursue the precise structure of the optimal graphs because it appears that their description is substantially more involved, and this paper was already quite long.

• Our methods in Section 3 can easily be adapted to maximize the number of graph homomorphisms to an arbitrary H (not just K_q). The analogs of Theorems 3.2 and 3.3 show that for any fixed H, the asymptotic maximum number of homomorphisms from an n-vertex, m-edge graph to H can be determined by solving a certain quadratically constrained linear program. Although this can in principle be done, it appears that the computations become rather messy even for graphs H of small order.

However, in the interesting case when H is the two-vertex graph consisting of a single edge plus a loop, one can easily determine the extremal graphs via a direct argument. As we mentioned in the introduction, this corresponds to maximizing the number of independent sets. By considering the complement of the graph, this is equivalent to maximizing the number of cliques.

We claim that for any n, m, the same graph that Linial found to minimize the number of colorings also happens to maximize the number of cliques. This graph G^* was a clique K_k with an additional vertex adjacent to l vertices of the K_k, plus $n - k - 1$ isolated vertices, where k, l are the unique integers satisfying $m = \binom{k}{2} + l$ with $k > l \geqslant 0$. We will show that for any t, every n-vertex graph G with m edges has at most as many t-cliques as G^*. The only nontrivial values of t to check are $2 \leqslant t \leqslant k$.

If $l + 2 \leqslant t \leqslant k$, then G^* has exactly $\binom{k}{t}$ cliques of size t. Suppose for contradiction that G has more t-cliques. Construct a t-uniform hypergraph with at least $\binom{k}{t} + 1 = \binom{k}{t} + \binom{t-1}{t-1}$ hyperedges by defining a hyperedge for each t-clique. By the Kruskal-Katona theorem (see, for example, the book [3]), the number of 2-sets that are contained in some hyperedge is at least $\binom{k}{2} + \binom{t-1}{1} \geqslant \binom{k}{2} + (l + 1)$, which exceeds the number of edges of G. This contradicts the definition of the hyperedges, because each of these 2-sets must be an edge of G.

On the other hand, if $2 \leqslant t \leqslant l + 1$, G^* has exactly $\binom{k}{t} + \binom{l}{t-1}$ cliques of size t. A similar argument shows that if G has at least $\binom{k}{t} + \binom{l}{t-1} + 1 = \binom{k}{t} + \binom{l}{t-1} + \binom{t-2}{t-2}$ cliques of size t, then G must have at least $\binom{k}{2} + \binom{l}{1} + \binom{t-2}{0} \geqslant \binom{k}{2} + l + 1$ edges, a contradiction.

Therefore, G^* indeed maximizes the number of cliques. Furthermore, we can classify all extremal graphs, because our argument shows that any other graph G with as many cliques as G^* must also have exactly the same number of t-cliques for all integers t. In particular, using $t = k$, we see that G must also contain a K_k. If $l \neq 1$, we can use $t = l + 1$ to conclude that the remaining edges form a star with all endpoints in the K_k. Therefore, the maximizer is unique unless $l = 1$, in which case the extremal graphs are K_k plus an arbitrary edge (not necessarily incident to the K_k).

Appendix A. Routine verifications for Optimization Problem 2

In this section, we present the postponed proofs of the results stated in Section 4.1.3. We begin by disposing of Lemma 4.9, which states some analytical facts about the function $F_q(x) = \log \frac{q}{q-x} \cdot \log \frac{q}{x}$.

Proof of Lemma 4.9. For part (i), observe that if we reparametrize with $t = x/q$, then we need to show that the function $f(t) = \log \frac{1}{1-t} \log \frac{1}{t}$ is strictly increasing on $0 < t < 1/2$ and strictly decreasing on $1/2 < t < 1$. Instead of presenting a tedious analytic proof (which is routine and not very interesting), we refer the reader to *Mathematica*'s plot of $f(t)$ in Figure A1(i).

For part (ii), define the functions $g(x) = F_x(3) = \log \frac{x}{x-3} \cdot \log \frac{x}{3}$ and $h(x) = 2F_x(1) \cdot \frac{x-3}{x-2} = 2 \cdot \log \frac{x}{x-1} \log x \cdot \frac{x-3}{x-2}$. We need to show that $g(x) > h(x)$ for all $x \geqslant 9$. Direct substitution yields $g(9) \approx 0.4454$ and $h(9) \approx 0.4437$, so it is true at $x = 9$.
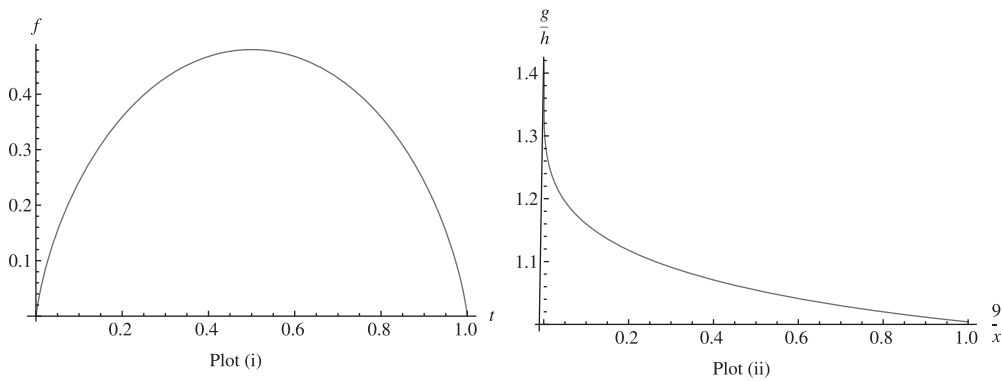
FIGURE A1. *Plot (i) displays the function* $f(t) = \log \frac{1}{1-t} \log \frac{1}{t}$. *Plot (ii) displays the ratio* $g(x)/h(x)$, *where g and h are as defined above, and the horizontal axis is parametrized by $9/x$.*

Also, a quick estimate shows that asymptotically, as $x \to \infty$, $g(x) = \log \left(1 + \frac{3}{x-3}\right) \cdot \log \frac{x}{3} = (1 + o(1))\frac{3}{x} \cdot \log x$ and $h(x) = 2 \cdot \log(1 + \frac{1}{x-1}) \cdot \log x \cdot \frac{x-3}{x-2} = (2 + o(1))\frac{1}{x} \cdot \log x$. Therefore, the ratio $g(x)/h(x)$ tends to 1.5, which is indeed greater than 1.

Again, instead of writing a routine analytic proof to fill in the gap between 9 and infinity, we refer the reader to Figure A1(ii), which shows that the ratio g/h steadily increases as x grows from 9. Thus, $g(x) > h(x)$ for all $x \geqslant 9$, as required.

The monotonicity of $F_q(x)$ on $0 < x < q/2$, which we just established, is useful for our next proof. This is Lemma 4.6, which stated that if $\boldsymbol{\alpha}$ solves OPT* and is supported by a partition of $[q]$ consisting of exactly two sets, then $\boldsymbol{\alpha}$ must have the same form as $\boldsymbol{\alpha}^*$, the claimed optimal vector in Proposition 4.1. $\qquad\square$

Proof of Lemma 4.6. Let A and B denote the two sets in the support, with $|A| \leqslant |B|$. Write $a = |A|$. Flipping the fractions to make the logarithms positive, we have OBJ*$(\boldsymbol{\alpha}) = -\alpha_A \log \frac{q}{a} - \alpha_B \log \frac{q}{q-a} \leqslant -2\sqrt{\alpha_A \log \frac{q}{a} \cdot \alpha_B \log \frac{q}{q-a}}$ by the inequality of arithmetic and geometric means. Yet $\alpha_A \alpha_B = \text{E}(\boldsymbol{\alpha}) \geqslant 1$ since $\boldsymbol{\alpha}$ is in the feasible set FEAS*, so OBJ*$(\boldsymbol{\alpha}) \leqslant -2\sqrt{\log \frac{q}{a} \cdot \log \frac{q}{q-a}} = -2\sqrt{F_q(a)}$. Here, F_q is the function which Lemma 4.9(i) claimed was strictly increasing between 0 and $q/2$. In particular, since $1 \leqslant a \leqslant q/2$, the final bound is at most $-2\sqrt{F_q(1)}$, which we recognize as OBJ*$(\boldsymbol{\alpha}^*)$, where $\boldsymbol{\alpha}^*$ is the claimed unique optimal vector in Proposition 4.1.

Since $\boldsymbol{\alpha}$ was assumed to be maximal, we must have equality in all of the above inequalities. Checking the equality conditions, we find that $\boldsymbol{\alpha}$ must indeed have the unique form claimed in Proposition 4.1. $\qquad\square$

The remaining lemma from Section 4.1.3 ruled out a handful of partitions as possible supports for optimal vectors. It turns out that each of those excluded partitions is a special case of the following result.

LEMMA A.1. *Fix any integer $q \geqslant 3$, and let $\boldsymbol{\alpha}$ be a vector which solves OPT*, whose support is a partition of $[q]$. Then that partition cannot be $\{1, \ldots, t\} \cup \{t+1\} \cup \{t+2\} \cup \ldots \cup \{q\}$, where $1 \leqslant t \leqslant q - 2$.*

Proof. Assume for the sake of contradiction that $\boldsymbol{\alpha}$ is supported by the above partition. Let $x = \alpha_{\{t+1\}} = \ldots = \alpha_{\{q\}}$, which are all equal by Lemma 4.5(ii). We assumed that $\boldsymbol{\alpha}$ was

maximal, so in particular $\text{OBJ}^*(\boldsymbol{\alpha}) \geqslant \text{OBJ}^*(\boldsymbol{\alpha}^*) = -2\sqrt{\log \frac{q}{q-1} \log q}$, where $\boldsymbol{\alpha}^*$ is the feasible vector constructed in Proposition 4.1. Therefore,

$$(q-t)x \log \frac{1}{q} > \alpha_{\{1,\dots,t\}} \log \frac{t}{q} + (q-t)x \log \frac{1}{q} = \text{OBJ}^*(\boldsymbol{\alpha}) \geqslant -2\sqrt{\log \frac{q}{q-1} \log q},$$

and we conclude that $(q-t)x < 2\sqrt{\log \frac{q}{q-1}/\log q}$. On the other hand, we also know by Lemma 4.5(i) for the set $A = \{1, \dots, t\}$ that $(q-t)x = I_A/\alpha_A = 2J_A/\alpha_A = (2\log \frac{t}{q})/\text{OBJ}^*(\boldsymbol{\alpha})$. Using the final bound for $(q-t)x$ above, this gives

$$\text{OBJ}^*(\boldsymbol{\alpha}) = \left(2\log \frac{t}{q}\right)/((q-t)x) < \log \frac{t}{q} \cdot \sqrt{(\log q)/\log \frac{q}{q-1}}.$$

(The inequality reversed because $\log(t/q)$ is negative.)

To get our contradiction, it remains to show that this is less than $\text{OBJ}^*(\boldsymbol{\alpha}^*) = -2\sqrt{\log \frac{q}{q-1} \log q}$. Cancelling the common factor of $\sqrt{\log q}$ and rearranging terms, this reduces to showing that $\log \frac{q}{t} > 2\log \frac{q}{q-1}$.

Since $t \leqslant q-2$ by definition, it suffices to show that $\log \frac{q}{q-2} > 2\log \frac{q}{q-1}$. Removing the logarithms reduces us to showing that $q/(q-2) > q^2/(q-1)^2$. This is equivalent to $(q-1)^2 > q(q-2)$, which is easily seen to be true by multiplying out each side. $\square$

Proof of Lemma 4.7. Part (i), the partition of all singletons, is precisely the case of the previous lemma when $t = 1$. Similarly, part (ii), the partition of all singletons except for a 2-set, corresponds to the $t = 2$ case. For part (iii), which concerns partitions that include a $(q-2)$-set, first note that if the partition is a $(q-2)$-set plus two singletons, then it is precisely the $t = q - 2$ case of the previous lemma. The only other possibility is that the partition is a $(q-2)$-set plus a 2-set, and this is excluded by Lemma 4.6. $\square$

Appendix B. Routine verifications for exact results

PROPOSITION B.1. *Let r be a sufficiently large positive integer. Then the complete bipartite graph $K_{r,2r}$ plus one pendant edge achieves the maximum number of colorings among all $(3r+1)$-vertex graphs with $2r^2 + 1$ edges.*

Proof. Every three-coloring of $K_{r,2r}$ has exactly two extensions to the pendant vertex, so Lemma 5.2 shows that the above graph has exactly $(3 \cdot 2^r + 3 \cdot 2^{2r} - 6) \cdot 2 = (1 + o(1)) \cdot 3 \cdot 2^{2r+1}$ colorings. Plugging $n = 3r + 1$ and $m = 2r^2 + 1$ into the dense case of Theorem 1.3, we see that the only other graphs we need to consider are semi-complete subgraphs of some $K_{a,b}$ with $a = (1 + o(1))r$ and $b = (2 + o(1))r$, plus isolated vertices. Note that we must have $a \geqslant r$, because when $a \leqslant r - 1$ and $a + b \leqslant 3r + 1$, convexity implies that $ab \leqslant (r-1)(2r+2) = 2r^2 - 2 < 2r^2 + 1$, and there would not be enough edges.

Let G' be one of the above graphs with $a = r + t$ for some $t \geqslant 0$. We must have $b \geqslant 2r - 2t + 1$, because $(r+t)(2r-2t) = 2r^2 - 2t^2 < 2r^2 + 1$, so any smaller b would not produce enough edges. This leaves $n - a - b \leqslant t$ isolated vertices. Observe that when $t = 0$, this forces G' to be a semi-complete subgraph of $K_{r,2r+1}$ with exactly $r - 1$ missing edges. Lemma 5.2 then shows that the number of colorings of G' is $3 \cdot 2^r + 3 \cdot 2^{2r+1} + 6 \cdot (2^{r-1} - 2)$, which is exactly the same as G.

It remains to consider $t > 0$. By definition, any semi-complete subgraph of $K_{a,b}$ is missing at most $a - 1$ edges, so Lemma 5.2 implies that the number of three-colorings of G' is at most $3^{n-a-b} \cdot (3 \cdot 2^a + 3 \cdot 2^b + 6 \cdot (2^{a-1} - 2))$. This expression is largest when b is as small as possible, so using $b \geqslant 2r - 2t + 1$ and $n = 3r + 1$, we find that G' has at most $3^t \cdot (3 \cdot 2^a + 3 \cdot$

$2^{2r-2t+1} + 6 \cdot (2^{a-1} - 2))$ colorings. Since $a = (1 + o(1))r$, this is at most $\left(\left(\frac{3}{4}\right)^t + o(1)\right) \cdot 3 \cdot 2^{2r+1}$, which is indeed less than the number of colorings of G when r is large. $\qquad\square$

REMARK. A similar argument shows that for any $c \in \{0, \pm 1, \pm 2\}$ and large r, $K_{r,2r+c}$ plus a pendant edge is optimal among graphs with $3r + c + 1$ vertices and $r(2r + c) + 1$ edges. Interestingly enough, it can also be shown that these values of n, m are the only ones which produce optimal graphs that are not semi-complete plus isolated vertices, when n, m are large.

INEQUALITY B.2. *Let a, b, t be positive integers, with $t \geqslant 3$ and $\frac{b}{a} \geqslant \log t / \log \frac{t-1}{t-2}$. Then*
(i) *The product $i^a(t - i)^b$ falls by a factor of at least 1.5^a when i increases by 1, for all $i \in \{1, \ldots, t - 2\}$.*
(ii) *If we further assume that a is sufficiently large (depending only on t), then $\sum_{i=1}^{t-1} \binom{t}{i} i^a (t - i)^b \leqslant 1.1 \cdot t(t - 1)^b$, that is, the first summand dominates.*

Proof. When $i \in \{1, \ldots, t - 2\}$ increases by 1, i grows by a factor of at most 2, but $t - i$ falls by at least $(t - 1)/(t - 2)$. Thus, the product $i^a(t - i)^b$ falls by a factor of at least

$$\left(\frac{1}{2}\right)^a \left(\frac{t-1}{t-2}\right)^b = \left(\frac{1}{2} \cdot \left(\frac{t-1}{t-2}\right)^{b/a}\right)^a \geqslant \left(\frac{1}{2} \cdot t\right)^a.$$

Since $t \geqslant 3$, this gives (i).

For part (ii), when i increases by 1, the term $\binom{t}{i}$ in the summand grows by a factor of at most t, but by (i) the rest of the summand falls by a factor of at least 1.5^a. Thus for sufficiently large a, each successive term of the sum falls by a factor of at least $1.4^a > 20$. The result follows by bounding the sum by a geometric series, since $1 + 1/20 + 1/20^2 + \ldots < 1.1$. $\qquad\square$

INEQUALITY B.3. *Let m, n, t, and v_1 be positive integers, with $m \leqslant n^2/4$ and $v_1(n - v_1) \geqslant m - t$. Let s be the largest integer that satisfies $s(n - s) \geqslant m$. Then $s \geqslant v_1 - \sqrt{t}$.*

Proof. The inequality for s rearranges to $s^2 - ns + m \leqslant 0$, so the quadratic formula implies that s is precisely $\lfloor (n + \sqrt{n^2 - 4m})/2 \rfloor$. Similarly, the inequality for v_1 rearranges to $v_1^2 - nv_1 + (m - t) \leqslant 0$, so the quadratic formula implies that $v_1 \leqslant \lfloor (n + \sqrt{n^2 - 4m + 4t})/2 \rfloor$. Therefore,

$$v_1 - s \leqslant \left\lfloor \frac{n + \sqrt{n^2 - 4m + 4t}}{2} \right\rfloor - \left\lfloor \frac{n + \sqrt{n^2 - 4m}}{2} \right\rfloor$$

$$\leqslant \left\lceil \frac{n + \sqrt{n^2 - 4m + 4t}}{2} - \frac{n + \sqrt{n^2 - 4m}}{2} \right\rceil = \left\lceil \frac{\sqrt{(n^2 - 4m) + 4t} - \sqrt{n^2 - 4m}}{2} \right\rceil.$$

Since the function $\sqrt{x}$ is concave and we assumed $n^2 - 4m \geqslant 0$, this final bound is largest when $n^2 - 4m = 0$. Therefore, $v_1 - s \leqslant \lceil \sqrt{t} \rceil$, which gives the claimed result. $\qquad\square$

LEMMA B.4. *The number of q-colorings of the Turán graph $T_{q-1}(n)$ is exactly $q! \cdot [(q - 1 + r)2^{s-1} - q + 2]$, where s and r are defined by $n = s(q - 1) + r$ with $0 \leqslant r < q - 1$.*

Proof. The complete $(q - 1)$-partite graph $T_{q-1}(n)$ has r parts of size $s + 1$ and $q - 1 - r$ parts of size s, and any q-coloring must use different colors on each part. The number of q-colorings that use exactly one color on each part is exactly $q \cdot (q - 1) \cdots 2 = q!$. All other colorings use two colors on one part, and one color on each of the other parts. There are $\binom{q}{2}$

ways to choose which two colors are paired. If the pair of colors is used on one of the r parts of size $s+1$, then there are $2^{s+1} - 2$ ways to color that part with exactly two colors, followed by $(q-2)!$ ways to choose which color goes to each of the remaining parts. Otherwise, if the pair of colors appears on one of the $q-1-r$ parts of size s, then there $(2^s - 2)(q-2)!$ colorings of this form. Therefore, the number of q-colorings of $T_{q-1}(n)$ is exactly

$$q! + \binom{q}{2} \cdot \left[r \cdot (2^{s+1} - 2)(q-2)! + (q-1-r) \cdot (2^s - 2)(q-2)! \right] = q! \cdot \left[(q-1+r)2^{s-1} - q + 2 \right],$$

as claimed. □

INEQUALITY B.5. *Fix any $q \geqslant 4$. For all sufficiently large n, the number of q-colorings of the Turán graph $T_{q-1}(n)$ is strictly greater than*

$$q! \cdot 2^{(q-1)(n-q)} \cdot 2^{-\frac{q}{q-1}\left[\delta + \binom{q-1}{2}\right] + 2\binom{q}{2}}, \tag{5}$$

where δ is the difference between the number of edges of $T_{q-1}(n)$ and $T_{q-1}(n-q+1)$.

Proof. Divide n by $q-1$, so that $n = s(q-1) + r$ with $0 \leqslant r < q-1$. Then $T_{q-1}(n)$ has exactly r parts of size $s+1$ and $q-1-r$ parts of size s, and $T_{q-1}(n-q+1)$ is obtained by deleting one vertex per part. Each deleted vertex in a part of size $s+1$ had degree $n-s-1$, while each deleted vertex in a part of size s had degree $n-s$. Thus, the number of deleted edges is $\delta = r(n-s-1) + (q-1-r)(n-s) - \binom{q-1}{2}$, where we had to subtract the double-counted edges of the K_{q-1} induced by the set of deleted vertices. Substituting this into (5) and using $n = s(q-1) + r$ to simplify the expression, we obtain:

$$q! \cdot 2^{(q-1)(n-q)} \cdot 2^{-\frac{q}{q-1}\left[\delta + \binom{q-1}{2}\right] + 2\binom{q}{2}} = q! \cdot 2^{(q-1)(n-q)} \cdot 2^{-\frac{q}{q-1}\left[r(n-s-1) + (q-1-r)(n-s)\right] + 2\binom{q}{2}}$$

$$= q! \cdot 2^s \cdot 2^{\frac{r}{q-1}}.$$

It remains to show that this is strictly less than the number of colorings of $T_{q-1}(n)$, which Lemma B.4 calculated to be $q! \cdot \left[(q-1+r)2^{s-1} - q + 2 \right] = (1 - o(1)) \cdot q! \cdot 2^s \cdot (q-1+r)/2$. Here, the $o(1)$ term tends to zero as n grows (and $s = \lfloor n/(q-1) \rfloor$ grows). Recall that $0 \leqslant r < q-1$, so when $r \geqslant 1$ and $q \geqslant 4$ we always have $2^{r/(q-1)} < 2^1 \leqslant (q-1+r)/2$, giving the desired result. On the other hand, when $r = 0$, the result follows from $2^{r/(q-1)} = 2^0 < \frac{3}{2} \leqslant (q-1+r)/2$. □

References

1. N. ALON, 'Independent sets in regular graphs and sum-free subsets of finite groups', *Israel J. Math.* 73 (1991) 247–256.
2. N. ALON, J. BALOGH, P. KEEVASH, and B. SUDAKOV, 'The number of edge colorings with no monochromatic cliques', *J. London Math. Soc.* 70 (2004) 273–288.
3. I. ANDERSON, *Combinatorics of finite sets* (Dover, Mineola, NY, 2002).
4. J. BALOGH, 'A remark on the number of edge colorings of graphs', *European. J. Combin.* 27 (2006) 565–573.
5. E. BENDER and H. WILF, 'A theoretical analysis of backtracking in the graph coloring problem', *J. Algorithms* 6 (1985) 275–282.
6. G. BIRKHOFF, 'A determinant formula for the number of ways of coloring a map', *Ann. Math.* 14 (1912) 42–46.
7. G. BIRKHOFF, 'On the number of ways of colouring a map', *Proc. Edinb. Math. Soc.* (2) 2 (1930) 83–91.
8. G. BIRKHOFF and D. LEWIS, 'Chromatic polynomials', *Trans. Am. Math. Soc.* 60 (1946) 355–451.
9. B. BOLLOBÁS, *Modern graph theory* (Springer, New York, NY, 1998).
10. O. BYER, 'Some new bounds for the maximum number of vertex colorings of a (v, e)-graph', *J. Graph Theory* 28 (1998) 115–128.
11. K. DOHMEN, 'Lower bounds and upper bounds for chromatic polynomials', *J. Graph Theory* 17 (1993) 75–80.
12. K. DOHMEN, 'Bounds to the chromatic polynomial of a graph', *Results Math.* 33 (1998) 87–88.

13. P. Erdős, 'Some new applications of probability methods to combinatorial analysis and graph theory', *Cong. Numer.* 10 (1974) 39–51.

14. P. Erdős, 'Some of my favourite problems in various branches of combinatorics', *Matematiche (Catania)* 47 (1992) 231–240.

15. D. Galvin and P. Tetali, 'On weighted graph homomorphisms', *Graphs, morphisms and statistical physics*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 63 (American Mathematical Society, Providence, RI, 2004) 97–104.

16. J. Kahn, 'An entropy approach to the hard-core model on bipartite graphs', *Combin. Prob. Comput.* 10 (2001) 219–237.

17. J. Kahn, 'Entropy, independent sets and antichains: a new approach to Dedekind's problem', *Proc. Am. Math. Soc.* 130 (2002) 371–378.

18. F. Lazebnik, 'On the greatest number of 2 and 3 colorings of a (V, E)-graph', *J. Graph Theory* **13** (1989), 203–214.

19. F. Lazebnik, 'New upper bounds for the greatest number of proper colorings of a (V, E)-graph', *J. Graph Theory* 14 (1990) 25–29.

20. F. Lazebnik, 'Some corollaries of a theorem of Whitney on the chromatic polynomial', *Discrete Math.* 87 (1991) 53–64.

21. F. Lazebnik, O. Pikhurko and A. Woldar, 'Maximum number of colorings of $(2k, k^2)$-graphs', *J. Graph Theory* 56 (2007) 135–148.

22. N. Linial, 'Legal coloring of graphs', *Combinatorica* 6 (1986) 49–54.

23. R. Liu, 'On the greatest number of proper 3-colorings of a graph', *Math. Appl.* 6 (1993) 88–91.

24. I. Simonelli, 'Optimal graphs for chromatic polynomials', *Discrete Math.* 308 (2008) 2228–2239.

25. M. Simonovits, 'A method for solving extremal problems in graph theory, stability problems', *Theory of graphs*, Proceedings of the Colloquium, Tihany, 1966, (eds P. Erdős and G. Katona, Academic Press, New York, 1968) 279–319.

26. I. Tomescu, 'Le nombre maximal de colorations d'un graphe', *C. R. Acad. Sci. Paris* 272 (1971) 1301–1303.

27. I. Tomescu, 'Le nombre maximal de 3-colorations d'un graphe connexe', *Discrete Math.* 1 (1972) 351–356.

28. I. Tomescu, 'Le nombre minimal de colorations d'un graphe', *C. R. Acad. Sci. Paris* 274 (1972) 539–542.

29. I. Tomescu, 'Problèmes extremaux concernant le nombre des colorations des sommets d'un graphe fini', *Combinatorial programming: methods and applications, Proceedings of NATO Advanced Study* Institute, Versailles, 1974), NATO Advanced Study Institute Series, Series C: Mathematical and Physical Science 19 (Reidel, Dordrecht, 1975) 327–336.

30. I. Tomescu, 'Le nombre maximal de colorations d'un graphe hamiltonien', *Discrete Math.* 16 (1976) 353–359.

31. I. Tomescu, 'Maximal chromatic polynomials of connected planar graphs', *J. Graph Theory* 14 (1990) 101–110.

32. I. Tomescu, 'Maximum chromatic polynomials of 2-connected graphs', *J. Graph Theory* 18 (1994) 329–336.

33. I. Tomescu, 'Maximum chromatic polynomial of 3-chromatic blocks', *Discrete Math.* 172 (1997) 131–139.

34. H. Whitney, 'A logical expansion in mathematics', *Bull. Am. Math. Soc.* 38 (1932) 572–579.

35. H. Wilf, 'Backtrack: an $O(1)$ expected time algorithm for the graph coloring problem', *Inf. Process. Lett.* 18 (1984) 119–121.

36. E. Wright, 'Counting coloured graphs III', *Can. J. Math.* 24 (1972) 82–89.

37. R. Yuster, 'The number of edge colorings with no monochromatic triangle', *J. Graph Theory* 21 (1996) 441–452.

38. Y. Zhao, 'The number of independent sets in a regular graph', *Combin. Prob. Comput.*, to appear.

Po-Shen Loh
Department of Mathematics
Princeton University
Princeton, NJ 08544
USA

ploh@math.princeton.edu

Oleg Pikhurko
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15123
USA

pikhurko@andrew.cmu.edu

Benny Sudakov
Department of Mathematics
UCLA
Los Angeles
CA 90095
USA

bsudakov@math.ucla.edu