Cycles and coloring in graphs and digraphs

by

Patrick Hompe

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2022

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis includes work from 5 coauthored papers, all of which I played a major role in producing. Many of the words in this thesis appear verbatim from these papers, which I wrote. They are listed below, in the order in which their contents appear.

- P. Hompe, P. Pelikanova, A. Pokorna, S. Spirkl, "On Aharoni's rainbow generalization of the Caccetta-Häggkvist conjecture," *Discrete Mathematics*, 344(5), 2021, https://doi.org/10.1016/j.disc.2021.112319.

- P. Hompe and S. Spirkl, "Further approximations for Aharoni's rainbow generalization of the Caccetta-Häggkvist conjecture," *Electronic Journal of Combinatorics*, 29(1), 2022, https://doi.org/10.37236/10418.

- P. Hompe, Z. Qu, and S. Spirkl, "Improved bounds for the triangle case of Aharoni's rainbow generalization of the Caccetta-Häggkvist conjecture," ArXiv:2206.10733, 2022.

- A. Carbonero, P. Hompe, B. Moore, and S. Spirkl, "A counterexample to a conjecture about triangle-free induced subgraphs of graphs with large chromatic number," ArXiv:2201.08204, 2022.

- A. Carbonero, P. Hompe, B. Moore, and S. Spirkl, "Digraphs with all induced directed cycles of the same length are not $\vec{\chi}$-bounded," ArXiv:2203.15575, 2022.

Abstract

We show results in areas related to extremal problems in directed graphs. The first concerns a rainbow generalization of the Caccetta-Häggkvist conjecture, made by Aharoni. The Caccetta-Häggkvist conjecture states that if G is a simple digraph on n vertices with minimum out-degree at least k, then there exists a directed cycle in G of length at most $\lceil n/k \rceil$. Aharoni proposed a generalization of this well-known conjecture, namely that if G is a simple edge-colored graph (not necessarily properly colored) on n vertices with n color classes each of size at least k, then there exists a rainbow cycle in G of length at most $\lceil n/k \rceil$.

In this thesis, we first prove that if G is an edge-colored graph on n vertices with n color classes each of size at least $\Omega(k \log k)$, then G has a rainbow cycle of length at most $\lceil n/k \rceil$. Then, we develop more techniques to prove the stronger result that if there are n color classes, each of size at least $\Omega(k)$, then there is a rainbow cycle of length at most $\lceil n/k \rceil$. Finally, we improve upon existing bounds for the triangle case, showing that if there are n color classes of size at least $0.3988n$, then there exists a rainbow triangle, and also if there are $1.1077n$ color classes of size at least $n/3$, then there is a rainbow triangle.

Let $\chi(G)$ denote the *chromatic number* of a graph G and let $\omega(G)$ denote the *clique number*. Similarly, let $\vec{\chi}(D)$ denote the *dichromatic number* of a digraph D and let $\omega(D)$ denote the clique number of the underlying undirected graph of D. In the second part of this thesis, we consider questions of χ-boundedness and $\vec{\chi}$-boundedness. In the undirected setting, the question of χ-boundedness concerns, for a class $\mathcal{C}$ of graphs, for what functions f (if any) is it true that $\chi(G) \leq f(\omega(G))$ for all graphs $G \in \mathcal{C}$. In a similar way, the notion of $\vec{\chi}$-boundedness refers to, given a class $\mathcal{C}$ of digraphs, for what functions f (if any) is it true that $\vec{\chi}(D) \leq f(\omega(D))$ for all digraphs $D \in \mathcal{C}$. It was a well-known conjecture, sometimes attributed to Esperet, that for all $k, r \in \mathbb{N}$ there exists n such that in every graph with G with $\chi(G) \geq n$ and $\omega(G) \leq k$, there exists an induced subgraph H of G with $\chi(H) \geq r$ and $\omega(H) = 2$. We disprove this conjecture. Then, we examine the class of k-chordal digraphs, which are digraphs such that all induced directed cycles have length equal to k. We show that for $k \geq 3$, the class of k-chordal digraphs is not $\vec{\chi}$-bounded, generalizing a result of Aboulker, Bousquet, and de Verclos in [1] for $k = 3$. Then we give a hardness result for determining whether a digraph is k-chordal, and finally we show a result in the positive direction, namely that the class of digraphs which are k-chordal and also do not contain an induced directed path on k vertices is $\vec{\chi}$-bounded.

We discuss the work of others stemming from and related to our results in both areas, and outline directions for further work.

Acknowledgements

I would like to thank all the people who made this thesis possible. First, my advisor Sophie, for devoting so much time to work with me to produce research, and being patient with me. Her support was invaluable. I would thank to thank my readers, Jim Geelen and Luke Postle, for taking the time to read my thesis and giving insightful feedback. I would also like to thank the department as a whole, both students and faculty, for being supportive and accommodating while teaching me many things.

I would also like to thank Paul Seymour, for encouraging me to apply to graduate school and helping me in the process. Finally, I thank all my friends and family who have always been there to support me along the way.

Table of Contents

1 Basic definitions and background **1**

1.1 General background and motivation . . . 1

1.2 Basic definitions . . . 3

2 Aharoni's rainbow conjecture **6**

2.1 Background, preliminaries, and main results . . . 6

2.1.1 Background and preliminaries . . . 6

2.1.2 Main results . . . 8

2.1.3 Other required results . . . 9

2.2 n colors each with $ck \log k$ edges . . . 11

2.3 $n + c_1 k$ colors each with $c_2 k$ edges . . . 14

2.4 n colors each with ck edges . . . 20

2.5 Triangle case: $(1 + \delta)n$ colors . . . 24

2.6 Triangle case: n colors of size at least tn . . . 30

2.7 Conclusions and further work . . . 33

3 χ-boundedness and $\vec{\chi}$-boundedness **35**

3.1 Background and main results . . . 35

3.2 Disproving Esperet's conjecture . . . 38

3.3 t-chordal digraphs are not $\vec{\chi}$-bounded . . . 41

3.4 Deciding if a graph is t-chordal is coNP-complete . . . 45

3.5 Conclusions and further work . . . 47

References **48**

Chapter 1

Basic definitions and background

1.1 General background and motivation

In this section, we provide some general background and motivation for the contents of this thesis. In the following section, we will give some basic definitions for, among other things, the terms used here. The thesis is split roughly into two topics, which are contained in Chapters 2 and 3, respectively. In each chapter, further background and relevant results are given in the first section, as well as an outline of the main results of the chapter. Then, in the final section of each chapter, areas for potential further work for that topic are detailed.

The motivation for the contents of this thesis originally arise from the field of *graph theory*, a rich area of mathematics which emerged in the 20^{th} century. Let a graph be *planar* if it can be drawn in the plane with no crossings. One notable result that helped motivate the development of the subfield of *structural graph theory* was the *Four Color Theorem*, namely that every planar graph has chromatic number at most 4. This result was shown in 1976 by Appel and Haken [6], and motivates studying how the structure of a graph (in this case, planarity) affects its chromatic number. Another area of graph theory that emerged is termed broadly as *extremal graph theory*. A fundamental example from this field is *Turán's Theorem* from 1941 [41], which characterizes the minimum number of edges on a graph on n vertices to guarantee the existence of a K_t subgraph, where K_t is the complete graph on $t \leq n$ vertices.

Generally speaking, the results of this paper are pertaining to digraphs, which are a natural generalization of graphs. Chapter 2 is motivated by a famous extremal problem in digraphs known as the *Caccetta-Häggkvist conjecture*, and Chapter 3 is motivated by the notion of χ-boundedness for graphs, and an attempt to extend this to the digraph setting. Here, we discuss these motivations at a broad level, giving some historical context.

As mentioned above, extremal graph theory studies, generally speaking, when the existence of certain local substructures can be guaranteed by global conditions. For example, one can ask in the undirected setting, what is the minimum number of edges needed to guarantee the existence of a cycle of length at most k? When $k = 3$, this is simply given by Turán's Theorem, but for $k \geq 4$ it is a different problem. This was asymptotically resolved by Erdős and Stone in [18], where they establish a more general result known as the *Erdős-Stone Theorem*.

In the directed setting, things are significantly more nuanced. A transitive tournament (an acyclic digraph with its underlying undirected graph being the complete graph) shows that having a large number of edges does not necessarily guarantee the existence of a short directed cycle. It turns out that a condition on the minimum out-degree (or, equivalently, the minimum in-degree) is what seems appropriate for this question in the digraph setting. The following tantalizing conjecture was made by Caccetta and Häggkvist in 1978 [11]:

Conjecture 1.1.1 (Caccetta-Häggkvist [11]) *Suppose n, k are positive integers, and let G be a simple digraph on n vertices with $\delta^+(v) \geq k$ for all $v \in V(G)$; then G contains a directed cycle of length at most $\lceil n/k \rceil$.*

We consider this conjecture to be quite elegant, and, while some partial results are known, it has proven to be notoriously difficult over the years. This conjecture serves as the basis for motivating the results given in Chapter 2, where we consider a generalization of this conjecture due to Aharoni [4]. Aharoni's conjecture is again related to finding certain short cycles in undirected graphs.

On the other hand, in structural graph theory, a natural question that arises is, broadly speaking, what causes a graph to have large chromatic number? Certainly, if a graph G contains K_t as a subgraph, then we have $\chi(G) \geq t$. However, what if G has, say, no triangles? Are there more subtle ways in which the chromatic number of a graph must be large?

It turns out there are, and this is the motivation for the study of what is known as the χ-*boundedness* problem, which was first studied systematically by Gyárfás in [23]. We let a class $\mathcal{C}$ of graphs be χ-*bounded* if there exists a function f such that $\chi(G) \leq f(\omega(G))$ for all $G \in \mathcal{C}$. There is a fundamental conjecture in this branch of structural graph theory known as the Gyárfás-Sumner conjecture:

Conjecture 1.1.2 (Gyárfás [22], Sumner [39]) *For any forest F, the class of graphs not containing F as an induced subgraph is χ-bounded.*

Similar to the Caccetta-Häggkvist conjecture, despite much study and some partial results (see the survey by Scott and Seymour [35]) the Gyárfás-Sumner conjecture remains largely open, and the field of χ-boundedness is still not that well-understood. In this thesis, we disprove the following well-known conjecture attributed to Esperet related to χ-boundedness:

Conjecture 1.1.3 (Esperet (see Scott, Seymour [35])) *For all $k, r \in \mathbb{N}$ there is an $n \in \mathbb{N}$ such that for every graph G with $\chi(G) \geq n$ and $\omega(G) \leq k$, there is an induced subgraph H of G with $\chi(H) \geq r$ and $\omega(H) = 2$.*

We also consider what happens when we translate these notions over to the context of digraphs. For this, we use the notion of a coloring of a digraph defined in the previous section, but we now give some justification for this perhaps unintuitive definition.

Indeed, at first glance it is not entirely clear how to define a coloring for a directed graph. The notion of a coloring for digraphs that we use in this thesis, namely a partitioning of the vertex set into induced acyclic subdigraphs, is motivated in part by the setting of tournaments. A *tournament* is a digraph D such that the underlying undirected graph of D is a complete graph. In tournaments, a transitive tournament (acyclic tournament) in many ways plays the role of a stable set, and defining a coloring of a tournament as a partitioning of the vertex set into transitive tournaments is therefore tempting. Furthermore, a beautiful characterization is given by Berger, Choromanski, Chudnovsky, Fox, Loebl, Scott, Seymour, and Thomassé in [7] of the set of heroes for tournaments, where a tournament H is a hero if the set of tournaments which do not contain H as a subtournament have bounded chromatic number. This provides further motivation for these definitions, as we are in some sense searching for a way to generalize that hero result to the non-tournament digraph setting.

Therefore, we consider the following notion. For a class $\mathcal{C}$ of digraphs, we say that $\mathcal{C}$ is *$\vec{\chi}$-bounded* if there exists a function f such that $\vec{\chi}(D) \leq f(\omega(D))$ for all $D \in \mathcal{C}$. Aboulker, Charbit, and Naserasr proposed the following intriguing digraph generalization of the Gyárfás-Sumner conjecture in [2]:

Conjecture 1.1.4 (Aboulker, Charbit, Naserasr [2]) *For any oriented forest F, the class of digraphs not containing F as an induced subdigraph is $\vec{\chi}$-bounded.*

Investigating this relatively new domain of $\vec{\chi}$-boundedness is the basis of motivation for Chapter 3.

1.2 Basic definitions

In this thesis, we explore areas related to extremal problems in directed graphs. This field of study is broadly contained in what is known as *graph theory*, which is the study of objects known as graphs. In this section, we introduce some standard notions in graph theory, see for example the text by Diestel [15]. A *graph* $G = (V, E)$ is specified by a *vertex set* $V = V(G)$ and an *edge set* $E = E(G)$, which is a collection of sets of pairs of vertices in $V(G)$. Often times, some basic assumptions are made on the graphs that we work with. A *simple* graph is a graph with no loops or parallel edges; that is, for every $e \in E$, the two vertices contained in e are distinct, and also for every two distinct edges $e, f \in E$, e and f do not contain the same pair of vertices. A fundamental concept in graph theory is that of a coloring, or more specifically a vertex-coloring. A *vertex-coloring* of a simple graph G with k colors is a map $f : V(G) \to [k] = \{1, 2, \cdots, k\}$ such that for all edges $e = (uv) \in E(G)$, $f(u) \neq f(v)$. In this thesis, a coloring of a graph G refers to a vertex-coloring. We let $\chi(G)$ denote the *chromatic number* of a graph G, which is the minimum positive integer k such that there exists a k-coloring of G.

Another fundamental concept in graph theory is that of a clique. A *clique* in a simple graph G is a subset $C \subseteq V(G)$ such that for every pair of vertices $u, v \in C$, we have $(uv) \in E(G)$. We let $\omega(G)$ denote the *clique number* of G, which is the size of the largest clique in G. It is immediate that $\chi(G) \geq \omega(G)$.

Next, we introduce the concept of a cycle. A *cycle of length t* in a simple graph G is a sequence of vertices $\{v_1, v_2, \cdots, v_t, v_{t+1} = v_1\}$ such that for all $1 \leq i \leq t$, we have $(v_i v_{i+1}) \in E(G)$.

Finally, we define the concepts of a subgraph and an induced subgraph. A *subgraph* of $G = (V, E)$ is a graph $G' = (V', E')$ such that $V' \subseteq V$ and $E' \subseteq E$. An *induced subgraph* of $G = (V, E)$ is a subgraph $G' = (V', E')$ of G such that for every pair of vertices $u, v \in V'$, if $(uv) \in E$ then $(uv) \in E'$. Informally, an induced subgraph is obtained by taking all possible edges within a subset of the vertices of G.

Now, we move on to basic definitions for digraphs, which are a natural generalization of graphs and are the main subject of this thesis. A digraph $D = (V, E)$ is a set $V = V(D)$ of vertices and a set $E = E(D)$ of edges which are ordered pairs of vertices in $V(G)$. Informally, a digraph arises from taking a graph and assigning each of its edges a direction. Given a digraph D, we define its *underlying undirected graph* G to be that graph with $V(G) = V(D)$ and in which $u, v \in V(G)$ are adjacent if D contains an edge from u to v or from v to u. Informally, the underlying undirected graph of D is obtained by ignoring the directions on the edges.

We let a digraph D be *simple* if the underlying undirected graph of D is simple. We let a *directed cycle of length t* in a digraph D be a sequence of vertices $\{v_1, v_2, \cdots, v_t, v_{t+1} = v_1\}$ such that for all $1 \leq i \leq t$ we have $(v_i v_{i+1}) \in E(D)$.

A *subdigraph* of a digraph $D = (V, E)$ is a digraph $D' = (V', E')$ such that $V' \subseteq V$ and $E' \subseteq E$. An *induced subdigraph* of $D = (V, E)$ is a subdigraph $D' = (V', E')$ of D such that for every pair of vertices $u, v \in V'$, if $(uv) \in E$ then $(uv) \in E'$. Informally, an induced subdigraph is obtained by taking all possible edges within a subset of the vertices of D.

Now, we define the notion of a vertex-coloring of a digraph D, which was introduced in [17, 32] by Erdős and Neumann-Lara, respectively. We say that a *k-coloring* of a digraph D is a map from $V(D)$ to $f : V(D) \to [k] = \{1, 2, \cdots, k\}$ such that for each $1 \leq i \leq k$, the subdigraph induced by the vertex set $f^{-1}(i)$ does not contain any directed cycles. Informally, a coloring of a digraph is an assignment of colors to vertices such that there is no monochromatic directed cycle. We let $\vec{\chi}(D)$ denote the minimum positive integer k such that there exists a k-coloring of D.

Lastly, the *clique number* of a digraph D is simply the clique number of the underlying undirected graph of D.

Chapter 2

Aharoni's rainbow conjecture

We note that many of the words in this chapter appear verbatim from the set of papers listed in the *Statement of Contributions* section at the beginning of this thesis. These words were written by the author of this thesis.

2.1 Background, preliminaries, and main results

2.1.1 Background and preliminaries

Now, a graph or digraph is *simple* if there are no loops or parallel edges. For a simple digraph G and a vertex $v \in V(G)$, let $\delta^+(v)$ denote the number of out-neighbors of v in G. A famous conjecture in graph theory is the following, due to Caccetta and Häggkvist [11]:

Conjecture 2.1.1 (Caccetta-Häggkvist, 1978) *Suppose n, k are positive integers, and let G be a simple digraph on n vertices with $\delta^+(v) \geq k$ for all $v \in V(G)$; then G contains a directed cycle of length at most $\lceil n/k \rceil$.*

Despite a substantial amount of work and results on the conjecture, it remains as a whole open and out of reach. Recently, Razborov's method of flag algebras [33] has been applied by Hladký, Král, and Norin to get the best known approximate results for the triangle case of the Caccetta-Häggkvist conjecture [24], and indeed the conjecture provided some of the motivation for the development of flag algebras. Showing the exact result for the triangle case, however, appears to still be quite difficult.

One issue that has presented itself is finding a natural generalization of the Caccetta-Häggkvist conjecture. One may hope to find a nice formulation for the case with non-uniform out-degrees, but this has proved difficult. There are existing results shown by

Aharoni, Berger, Chudnovsky, and Zerbib in [3] concerning the case of non-uniform out-degrees, but this particular formulation will not generalize the Caccetta-Häggkvist conjecture itself, as demonstrated by an example in the undergraduate thesis of this author [25]. In this chapter, we consider a rather nice conjectured generalization of the Caccetta-Häggkvist conjecture made by Aharoni, concerning short rainbow cycles in edge-colored graphs. The conjecture appears to be harder than the Caccetta-Häggkvist conjecture, but is nice because it involves arbitrary graphs (the Caccetta-Häggkvist conjecture concerns digraphs with large minimum out-degree), and in addition it introduces a new degree of freedom, namely in the number of color classes. It also is nice because it has a natural generalization to matroids, which is explored by DeVos, Drescher, Funk, de la Maza, Guo, Huynh, Mohar, and Montejano in [16].

In the remainder of this section, we will first describe Aharoni's conjecture and existing results on it. Then, we will outline the results that we will show in the ensuing sections. Finally, we list a number of results of others that we will use in the proofs of this chapter.

Now, for a graph G and a function $c : E(G) \to \{1, \dots, |V(G)|\}$, a *rainbow cycle (with respect to c)* is a cycle C in G such that for all $e, f \in E(C)$ with $e \neq f$, we have $c(e) \neq c(f)$. We will refer to c as a *coloring* of the edges of G.* We say that c has *color classes of size at least k* for $k \in \mathbb{N}$ if $|c^{-1}(i)| \geq k$ for all $i \in \{1, \dots, |V(G)|\}$.

In [4], Aharoni proposes a generalization of Conjecture 2.1.1:

Conjecture 2.1.2 (Aharoni [4]) *Let n, k be positive integers, and let G be a simple graph on n vertices. Let $c : E(G) \to \{1, \dots, n\}$ be a coloring of the edges of G with color classes of size at least k; then G has a rainbow cycle of length at most $\lceil n/k \rceil$.*

In the original paper [4], Aharoni, DeVos, and Holzman focus on the triangle case of the conjecture, and establish some results related to the conjecture. We let (α, β) be *triangular* if every simple edge-colored graph on n vertices with at least αn color classes, each with at least βn edges, contains a rainbow triangle. In [4], the following partial results in this direction are shown:

Theorem 2.1.3 (Aharoni, DeVos, Holzman [4]) $(9/8, 1/3)$ *is triangular;* $(1, 2/5)$ *is triangular.*

In a paper that followed, DeVos, Drescher, Funk, de la Maza, Guo, Huynh, Mohar, and Montejano [16] proved that Conjecture 2.1.2 is true for $k = 2$:

Theorem 2.1.4 (DeVos, Drescher, Funk, de la Maza, Guo et al. [16]) *Let G be a simple graph on n vertices, and let c be a coloring of the edges of G with color classes of size at least 2; then there exists a rainbow cycle of length at most $\lceil n/2 \rceil$.*

*Note that c is not required to be a proper edge-coloring.

This result is generalized in [3], where Aharoni, Berger, Chudnovsky, and Zerbib show the following:

Theorem 2.1.5 (Aharoni, Berger, Chudnovsky, Zerbib [3]) *Let $F_1, F_2, \cdots, F_n$ be sets of edges in a simple graph G each of size at most two; then there exists a rainbow cycle of length at most $\lceil \sum_{1 \le i \le n} \frac{1}{|F_i|} \rceil$.*

Unfortunately, this formulation is not true in general, due to the counterexample given by this author in their undergraduate thesis [25]. However, it may be true up to a constant factor, namely that:

Conjecture 2.1.6 *There is a constant c such that if $F_1, F_2, \cdots, F_n$ are sets of edges in a simple graph G on n vertices, then there is a rainbow cycle of length at most $c \cdot \lceil \sum_{1 \le i \le n} \frac{1}{|F_i|} \rceil$.*

This would be in a sense a common generalization of a result of Theorem 2.1.10 (see below) and the following result of [3]:

Theorem 2.1.7 (Aharoni, Berger, Chudnovsky, Zerbib [3]) *If D is a simple digraph, then there exists a directed cycle of length less than $2 \cdot \sum_{v \in V(D)} \frac{1}{\delta^+(v)}$.*

Finally, Aharoni and Guo [5] consider the case where each color class forms a matching, which is in some sense the opposite of the case where each color class forms a star (the Caccetta-Häggkvist conjecture is a case where all color classes form stars by a reduction due to Aharoni, DeVos, and Holzman [4]). Here, they are able to show much stronger results than are given by Aharoni's conjecture (which in this case would give $\lceil n/2 \rceil$):

Theorem 2.1.8 (Aharoni, Guo [5]) *Let $F_1, F_2, \cdots, F_n$ be sets of edges in a simple graph G on n vertices, such that each F_i forms a matching of size at least two. Then there exists a rainbow cycle of size at most $2 \log n$.*

2.1.2 Main results

Now, we outline our contributions to this developing body of literature. We begin by showing the following approximate result for Aharoni's conjecture.

Theorem 2.1.9 (Hompe, Pelikánová, Pokorná, Spirkl [26]) *Let $k > 1$ be an integer, and let G be a simple graph on n vertices. Suppose that we have a coloring of the edges of G with n color classes of size at least $301k \log k$. Then G has rainbow girth at most $\lceil n/k \rceil$.*

Then, we develop additional techniques to prove the following stronger result, which shows that Aharoni's conjecture is true up to a constant factor.

Theorem 2.1.10 (Hompe, Spirkl [28]) *Let $k \geq 1$ be an integer, and let G be a simple graph on n vertices. Suppose we have a coloring of the edges of G with n color classes of size at least ck, where $c = 10^{11}$. Then G has rainbow girth at most n/k.*

Finally, we consider the triangle case of Aharoni's conjecture. We improve upon the results of Aharoni, DeVos, and Holzman from [4], showing the following:

Theorem 2.1.11 (Hompe, Qu, Spirkl [27]) $(1.1077, 1/3)$ *is triangular;* $(1.3481, 1/4)$ *is triangular.*

Theorem 2.1.12 (Hompe, Qu, Spirkl [27]) $(1, 0.3988)$ *is triangular.*

An outline of the chapter is as follows. In Section 2.2, we show our initial result that Aharoni's conjecture is true if the color classes have size $\Omega(k \log k)$. In Section 2.3, we prove a number of results on the case where the number of colors is $n + ck$ for some constant c. Then, in Section 2.4, we build upon the techniques of section 1.1 to prove that Aharoni's conjecture is true up to a constant factor.

Afterwards, we move to considering the triangle case in particular. In Section 2.5, we consider the case of $(1+\delta)n$ colors for $\delta > 0$, showing Theorem 2.1.11, and in fact a more general result. Then, in Section 2.6, we consider the case of n color classes each of size at least tn, and show Theorem 2.1.12. In the final section of the chapter, we make some concluding remarks, and suggest areas of further work.

2.1.3 Other required results

We finish this section by listing a number of results of others that we will make use of in the proofs of this chapter. We make use of the following results due to Bollobás and Szemerédi [9] and Shen [38], respectively. The first deals with the girth of a simple graph, while the second is an approximate result for Conjecture 2.1.1. In this thesis, log denotes the logarithm with base 2.

Theorem 2.1.13 (Bollobás, Szemerédi [9]) *For all $n \geq 4$ and $k \geq 2$, if G is a simple graph on n vertices with $n + k$ edges, then G contains a cycle of length at most*

$$\frac{2(n+k)}{3k}(\log k + \log\log k + 4).$$

Theorem 2.1.14 (Shen [38]) *Let G be a simple digraph with $\delta^+(v) \geq k$ for all $v \in V(G)$. Then G contains a directed cycle of length at most $\lceil n/k \rceil + 73$.*

We will use the following immediate corollary of Theorem 2.1.13, as well:

Corollary 2.1.15 *For all $n \geq 4$ and $k \geq 2$, if G is a simple graph on n vertices with $n+k$ edges, then G contains a cycle of length at most*

$$\frac{14(n+k)\log k}{3k}.$$

Proof. By Theorem 2.1.13, we have that the girth is at most:

$$\frac{2(n+k)}{3k}(\log k + \log\log k + 4) \leq \frac{14(n+k)\log k}{3k}$$

since that is equivalent to:

$$\log\log k + 4 \leq 6\log k.$$

To see that this is true, let $f(k) = 6\log k - \log\log k - 4$. Then $f(2) = 6 - 4 = 2 \geq 0$, and for all $k \geq 2$ we have:

$$f'(k) = \frac{6}{k\ln 2} - \frac{1}{\log k \ln 2}\frac{1}{k\ln 2} \geq \frac{4}{k\ln 2} \geq 0.$$

It follows that $f(k) \geq 0$ for all $k \geq 2$, as desired. This proves Corollary 2.1.15. ∎

We also make use of a set of Chernoff bounds and Chebyshev's Inequality:

Theorem 2.1.16 (Mitzenmacher, Upfal [31]) *Let $\{X_i\}_{i=1}^m$ be independent indicator random variables, and let $X = \sum_{i=1}^m X_i$. Then for any $\epsilon > 0$, we have:*

$$\mathbb{P}(X \leq (1-\epsilon)\mathbb{E}[X]) \leq \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}[X]\right);$$
$$\mathbb{P}(X \geq (1+\epsilon)\mathbb{E}[X]) \leq \exp\left(-\frac{\epsilon^2}{2+\epsilon}\mathbb{E}[X]\right).$$

Theorem 2.1.17 (Chebyshev's Inequality) *Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k > 0$ we have:*

$$\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

When considering the triangle case in particular, we make use of the following result (which we will also modify to obtain a stronger bound under certain conditions), due to Goodman, which was used by Aharoni, DeVos, and Holzman in [4] for the proof of Theorem 2.1.3 as well.

Theorem 2.1.18 (Goodman [21]) *Suppose G is a simple graph with n vertices and m edges, and let $t(G)$ denote the number of triangles in G; then we have:*

$$t(G) \geq \frac{4m}{3n}\left(m - \frac{n^2}{4}\right)$$

Finally, we will use the following state-of-the-art approximate result on the Caccetta-Häggkvist conjecture:

Theorem 2.1.19 (Hladký, Král, Norin [24]) *Let D be a simple digraph on n vertices with minimum out-degree $0.3465n$; then D contains a directed triangle.*

We note that improvements to the approximation in Theorem 2.1.19 would result in improvements to Theorem 2.1.12.

2.2 n colors each with $ck \log k$ edges

Our main result in this section is the following. The proof proceeds by either reducing the problem to an approximate version of the Caccetta-Häggkvist conjecture, or otherwise finding an induced subgraph which has significantly more colors appearing than its number of vertices. This will allow us to use existing results on the girth of undirected graphs.

Theorem 2.2.1 *Let $k > 1$ be an integer, and let G be a graph. Let c be a coloring of the edges of G with color classes of size at least $301k \log k$. Then G contains a rainbow cycle of length at most $\lceil n/k \rceil$.*

In the proof, we will need the following definitions. For $v \in V(G)$ and $i \in \{1, \ldots, n\}$, we say that i *is dominant at* v if v is incident with at least $7f(k)$ edges e such that $c(e) = i$. We call a vertex $v \in V(G)$ *color-dominating* if there exists $i \in \{1, \ldots, n\}$ such that i is dominant at v. We call a color $i \in \{1, \ldots, n\}$ *vertex-dominating* if there exists a vertex $v \in V(G)$ such that i is dominant at v. Let us say that $H \subseteq V(G)$ is *nice* if

- for every vertex-dominating color $i \in \{1, \ldots, n\}$, there is a vertex $v \in V(G) \setminus H$ such that i is dominant at v; and
- there are at most $|H|$ colors $i \in \{1, \ldots, n\}$ such that i is not vertex-dominating and for all $e \in c^{-1}(i)$, at least one end of e is in H.

Proof. We proceed by induction on the number of vertices. Let $f(k) = 7k \log k$, and let G be a graph on n vertices. Let c be a coloring of the edges of G with color classes of size at least $43f(k)$. Suppose for a contradiction that there is no rainbow cycle of length at most $\lceil n/k \rceil$. Note that G has at least $43f(k)n$ edges, and therefore, $n > 43f(k)$.

Claim 2.2.2 *If there is a nice set $H \subseteq V(G)$ with $6f(k) \leq |H| < n$, then there is a nice set $H' \subseteq V(G)$ with $|H'| = \lceil 6f(k) \rceil$.*

Proof. We remove vertices from H one-by-one such that the remaining set is nice. Suppose that we have removed $j \geq 0$ vertices from H, leaving a nice set H_j with $|H_j| > \lceil 6f(k) \rceil$. Let C_j be the set of colors $i \in \{1, \ldots, n\}$ which are not vertex-dominating and also do not have an edge e with $c(e) = i$ such that both ends of e are in $V(G) \setminus H_j$. From the definition of a nice set, we know $|C_j| \leq |H_j|$. If $|C_j| < |H_j|$, then removing any vertex from H_j gives a smaller nice set. So, we may assume that $|C_j| = |H_j|$. If there is a color i in C_j and an edge $e = uv \in c^{-1}(i)$ with $v \in H_j$ and $u \in G \setminus H_j$, then $H_j \setminus \{v\}$ is nice. If there is no such $i \in C_j$, then for every color $i \in C_j$, all edges in $c^{-1}(i)$ have both their ends in H_j. Now applying induction to the subgraph of G with vertex set H_j and edge set $c^{-1}(C_j)$ gives a rainbow cycle of length at most $\lceil n/k \rceil$ in G, a contradiction. This proves Claim 2.2.2. ∎

Claim 2.2.3 *There is a nice set $H' \subseteq V(G)$ with $|H'| = \lceil 6f(k) \rceil$.*

Proof. For each vertex-dominating color i, we pick a vertex v_i such that i is dominant at v_i, and let S be the set of these vertices v_i. Let $H = V(G) \setminus S$. Note that H is nice; thus by Claim 2.2.2, we may assume that either $|H| < 6f(k)$ or $|H| = n$.

We first consider the case when $|H| = n$. Since $43f(k) \geq 2$, Theorem 2.1.4 guarantees the existence of a rainbow cycle K of length at most $n/2 + 1$ in G. Let $H' = V(G) \setminus V(K)$. Then H' is nice, and $n > |H'| \geq n/2 - 1 \geq 6f(k)$; so by Claim 2.2.2, G contains a nice set of size $\lceil 6f(k) \rceil$.

Now we may assume that $|H| < 6f(k)$. We construct a digraph G' with $V(G') = S$, and for all i, j with $v_i, v_j \in S$, there is an arc $v_i \to v_j$ if $v_i v_j \in E(G)$ and $c(v_i v_j) = i$. Every vertex v_i is incident with at least $7f(k)$ edges e with $c(e) = i$, and since $|H| < 6f(k)$, there are at least $f(k)$ edges $e = v_i u$ with $c(e) = i$ and $u \in S$. Therefore, $\delta^+(G') \geq f(k)$.

Now, we claim $n/f(k) + 74 \leq n/k$, which is equivalent to $74kf(k) \leq n(f(k) - k)$. Since $k \geq 2$, we have $\log(k) \geq 117/301$, and thus $74kf(k) \leq 43f(k)(f(k) - k) \leq n(f(k) - k)$, as claimed.

Then, by applying Theorem 2.1.14 to G' we obtain a directed cycle K of length at most $\lceil n/f(k) \rceil + 73 \leq \lceil n/k \rceil$ in G'. The edges of G that correspond to arcs of K form a rainbow cycle of length at most $\lceil n/k \rceil$ in G, a contradiction. This proves Claim 2.2.3. ∎

Claim 2.2.4 *Let $H \subseteq V(G)$ be a nice set with $|H| = \lceil 6f(k) \rceil$. Then there exists $H' \subseteq H$ such that $|H'| \geq \lceil 2f(k) \rceil$ and such that for at least $n - \lceil f(k) \rceil + 1$ colors i, at least one edge $e \in c^{-1}(i)$ has both ends in $V(G) \setminus H'$.*

Proof. Let C be the set of colors i which are not vertex-dominating and for which no edge of $c^{-1}(i)$ has both ends in $V(G) \setminus H$. Since H is nice, it follows that $|C| \leq |H| = \lceil 6f(k) \rceil$. Let $D \subseteq C$ be the set of colors $i \in C$ such that there is a vertex $v \in H$ which is incident with all edges in $c^{-1}(i)$ that have one end in H and the other in $V(G) \setminus H$. We claim that $|D| \leq \lceil f(k) \rceil - 1$. Indeed, for each color $i \in D$, there are at least $\lceil 36f(k) \rceil$ edges in $c^{-1}(i)$ with both ends in H since i is not vertex-dominating. If $|D| > \lceil f(k) \rceil - 1$, then we obtain more than $(f(k) - 1)(36f(k))$ edges with both ends in H. Now, since $k \geq 2$, we have $f(k) \geq 72/23$, and it follows that:

$$(f(k) - 1)(36f(k)) \geq \frac{49f(k)^2}{2} \geq \frac{(6f(k) + 1)^2}{2} \geq \frac{|H|^2}{2}$$

which gives a contradiction. Thus, $|D| \leq \lceil f(k) \rceil - 1$.

Next, we claim there exists $H' \subseteq H$ such that $|H'| = \lceil 2f(k) \rceil$ and such that for all $i \in \{1, \ldots, n\} \setminus D$, there is an edge $e \in c^{-1}(i)$ with both ends in $V(G) \setminus H'$. To see this, we construct a graph J with vertex set H and the following set of edges. For each $i \in C \setminus D$, we choose two vertices $v_1^i, v_2^i \in H$, each incident with an edge in $c^{-1}(i)$ whose other end is in $V(G) \setminus H$; we know from the definition of D that this is possible. Now, the graph J has $|H|$ vertices and at most $|H|$ edges, and so J has a stable set $H' \subseteq V(J)$ of size at least $|V(J)|/3 \geq 2f(k)$; and so $|H'| \geq \lceil 2f(k) \rceil$.

Now, for every color $i \in C \setminus D$, $V(G) \setminus H'$ contains at least one of v_1^i, v_2^i, and therefore, there is an edge in $c^{-1}(i)$ with both ends in $V(G) \setminus H'$. Moreover, for every $i \in \{1, \ldots, n\} \setminus C$, either i dominates a vertex v in $V(G) \setminus H \subseteq V(G) \setminus H'$ (and so, since $|H'| < 7f(k)$, there is an edge in $c^{-1}(i)$ incident with v whose other end is not in H'); or there is an edge in $c^{-1}(i)$ with both ends in $V(G) \setminus H \subseteq V(G) \setminus H'$. Thus, for at least $n - |D| \geq n - \lceil f(k) \rceil + 1$ colors i, at least one edge in $c^{-1}(i)$ has both ends in $V(G) \setminus H'$. This proves Claim 2.2.4. ∎

By combining Claim 2.2.3 and Claim 2.2.4, we conclude that there exists $H' \subseteq V(G)$ with $|H'| \geq \lceil 2f(k) \rceil$, and such that for at least $n - \lceil f(k) \rceil + 1$ colors i, at least one edge in $c^{-1}(i)$ has both ends in $V(G) \setminus H'$. Let H'' be a subgraph of G with vertex set $V(G) \setminus H'$, obtained by taking exactly one edge in $c^{-1}(i)$ with both ends in $V(G) \setminus H'$ for all $i \in \{1, \ldots, n\}$ which have such an edge. It follows that $|E(H'')| \geq |V(H'')| + \lceil f(k) \rceil$.

Now, we claim that $\frac{2(n+f(k))}{3(f(k))}(\log\log(f(k)) + \log(f(k)) + 4) \leq \frac{n}{k}$. Using $f(k) < n/43$, it suffices to show:

$$\frac{88(\log\log(f(k)) + \log(f(k)) + 4)}{129} \leq 7\log(k)$$

Let $g(k) = 7\log(k) - \frac{88}{129}(\log\log(f(k)) + \log(f(k)) + 4)$. We have that $g(2) > 0$, and for $k \geq 2$ we have:

$$f(k)g'(k)\ln(2) = 49\log(k) - \frac{88}{129}f'(k)\left(\frac{1}{\log(f(k))\ln(2)} + 1\right) > 0$$

since for $k \geq 2$ we have:

$$f'(k)\left(\frac{1}{\log(f(k))\ln(2)} + 1\right) < (7 + 7\log(k))(3) \leq 49\log(k)$$

So $g'(k) > 0$ for $k \geq 2$, and it follows that $g(k) \geq 0$ for $k \geq 2$, as desired.

Then, Theorem 2.1.13 gives a rainbow cycle of length at most $\frac{2(n+f(k))}{3(f(k))}(\log\log(f(k)) + \log(f(k)) + 4) \leq \lceil \frac{n}{k} \rceil$, a contradiction. This proves Theorem 2.2.1. ∎

We have an immediate corollary which gives us a result for the case of $\Omega(n\log n)$ color classes each of size k:

Corollary 2.2.5 *Let k be a positive integer and let G be a simple graph on n vertices. Let $c : E(G) \to \{1, \ldots, t\}$ with $t \geq 303n\log n$, and with $|c^{-1}(i)| \geq k$ for all $i \in \{1, \ldots, t\}$. Then G contains a rainbow cycle in G of length at most $\lceil n/k \rceil$.*

Proof. Note that $t \geq 303n\log n \geq 303n\log k$. Since $303n\log k \geq n\lceil 301\log k \rceil$, there exists a partition of $\{1, \ldots, t\}$ into n parts, each of size at least $\lceil 301\log k \rceil$; that is, there is a function $f : \{1, \ldots, t\} \to \{1, \ldots, n\}$ such that $|f^{-1}(i)| \geq \lceil 301\log k \rceil$ for all $i \in \{1, \ldots, n\}$. By Theorem 2.2.1, applied to G and $f \circ c$, we obtain a rainbow cycle of length at most $\lceil n/k \rceil$ in G with respect to $f \circ c$, which is also rainbow with respect to c. This proves Corollary 2.2.5. ∎

2.3 $n + c_1k$ colors each with c_2k edges

We now consider a relaxation of Conjecture 2.1.2 where we have $n + c_1k$ color classes each of size at least c_2k, for constants c_1, c_2 which we will specify. In this case, we obtain upper bounds for the rainbow girth that are stronger than $\lceil n/k \rceil$ to a surprising degree. For this reason, these results are interesting in their own right. They are also used in the proof in the next section.

Our first result is the following. The argument uses probabilistic methods to obtain a rainbow subgraph with a large number of edges which is the disjoint union of stars, and then uses an argument which contracts the stars to finish the proof.

Theorem 2.3.1 *Let $k > 1$ be an integer, and let G be a simple graph on n vertices. Suppose we have a coloring of the edges of G with $n + k$ color classes of size at least ck, where $c = 0.99 \cdot 10^9$. Then G has rainbow girth at most 6 or G has rainbow girth at most:*

$$\frac{n(\log k)^2}{10k^{3/2}} + 14 \log k.$$

Proof. Since the graph is simple, we have that $n^2 \geq |E(G)| \geq cnk$ and thus we may assume that $n \geq ck$. Now, we claim there exists a set of vertices S with $|S| \leq n \log k/(140\sqrt{k})$ such that every color class has at least one edge incident to a vertex in S. To see this, we let $s = \lfloor 2 \log k \rfloor$ and $t = \lfloor n/(560\sqrt{k}) \rfloor$. We will iteratively construct s sets of vertices $S_1, \cdots, S_s$, each of size at most t, as follows. Suppose we have constructed $S_1, \cdots, S_i$ so far. Let $T_i = \cup_{j=1}^{i} S_j$, and let C_i denote the set of colors whose color class has no edge incident to a vertex in T_i. Let H be a random set of t vertices chosen uniformly with repetition. For any color class a, note that the number of vertices which are incident to an edge of color a is at least $\sqrt{ck}$, since if there are at most $\sqrt{ck}$ vertices incident to edges of color a, the number of edges of color a will be at most $ck/2$. Also, we have that $t \geq n/(560\sqrt{k}) - 1 \geq n/(1120\sqrt{k})$ since $n \geq 1120\sqrt{k}$ which is implied by $n \geq ck$. Using these two observations, we have that the expected number of colors in C_i whose color class has no edges incident to the vertices of H is at most:

$$\left(1 - \frac{\sqrt{ck}}{n}\right)^t |C_i| \leq \left(1 - \frac{\sqrt{ck}}{n}\right)^{n/(1120\sqrt{k})} |C_i| \leq e^{-\sqrt{c}/1120}|C_i|.$$

Now, let S_{i+1} be such that $|C_{i+1}| \leq e^{-\sqrt{c}/1120}|C_i|$, and iterate. When we finish, we have a collection of sets $\{S_1, S_2, \cdots, S_s\}$ such that:

$$\begin{aligned} |C_s| \leq e^{-\sqrt{c}s/1120}(n+k) &\leq e^{-\sqrt{c}(2\log k - 1)/1120}(n+k) \\ &\leq 2ne^{-\sqrt{c}(\log k)/1120} \\ &\leq \frac{n \log k}{280\sqrt{k}} \end{aligned}$$

where that last inequality is true for $k \geq 2$ since:

$$2nk^{-\sqrt{c}/(1120 \ln 2)} \leq \frac{n \log k}{280\sqrt{k}} \iff 560 \leq k^{\sqrt{c}/(1120 \ln 2) - 1/2} \log k$$

which is true for $k = 2$ and thus for all $k \geq 2$. Now, we have that T_s is a set of vertices with $|T_s| \leq \frac{n \log k}{280\sqrt{k}}$ such that at most $\frac{n \log k}{280\sqrt{k}}$ colors a have no edge of their color class adjacent to any of the vertices in T_s. It follows that, by adding at most $\frac{n \log k}{280\sqrt{k}}$ vertices, there exists a set of vertices of size at most $\frac{n \log k}{140\sqrt{k}}$ which is incident to at least one edge of every color class, as desired.

Now, let S be a set of at most $\frac{n \log k}{140\sqrt{k}}$ vertices such that S is incident to at least one edge of every color. For each color a, choose one edge e_c of color a such that e_c is incident to at least one vertex in S. Let E be the set of these chosen edges e_c. Then $|E| = n+k$ and E contains exactly one edge of each color. Now, let H be the subgraph with $V(H) = \bigcup_{(uv)\in E}\{u,v\}$ and $E(H) = E$, and let $S = \{v_1, v_2, \cdots, v_p\}$, where $p = |S|$. Partition $V(H) \setminus S$ into $X_1, \cdots, X_p$ such that $X_i \subseteq N_H(v_i)$ for all $1 \le i \le p$. Now, contract each $H_i = X_i \cup \{v_i\}$ to a single vertex (by contracting each edge of H_i iteratively), and let the resulting graph be H'. We have that $|V(H')| = |S| \le \frac{n \log k}{140\sqrt{k}}$ and $|E(H')| = |S| + k$. Note that a rainbow cycle C in H' corresponds to a rainbow cycle in G with length at most $3|C|$, by replacing each contracted vertex by at most a two-edge path. We may assume that H' is simple, since otherwise we obtain a rainbow cycle of length at most 6 in G. Then applying Corollary 2.1.15 to H' gives a rainbow cycle in G of length at most:

$$\frac{14\left(\frac{n \log k}{140\sqrt{k}} + k\right)\log k}{k} = \frac{n(\log k)^2}{10k^{3/2}} + 14\log k$$

as desired. This proves Theorem 2.3.1. ∎

We immediately obtain the following interesting corollary:

Corollary 2.3.2 *Let $k > 1$ be an integer, and let G be a simple graph on n vertices. Suppose that we have a coloring of the edges of G with $n+k$ color classes of size at least ck, where $c = 0.99 \cdot 10^9$, and suppose also that $\frac{140k^{3/2}}{\log k} \le n$. Then G has rainbow girth at most $\frac{n(\log k)^2}{5k^{3/2}}$.*

Proof. The condition on the size of n is equivalent to:

$$14\log k \le \frac{n(\log k)^2}{10k^{3/2}}.$$

We have that G has rainbow girth at least 7 since $\frac{n(\log k)^2}{5k^{3/2}} \ge 7$ is implied by the condition. Then Corollary 2.3.2 gives that G has rainbow girth at most:

$$\frac{n(\log k)^2}{10k^{3/2}} + 14\log k \le \frac{n(\log k)^2}{5k^{3/2}}$$

as desired. This proves Corollary 2.3.2. ∎

If we would like rainbow girth to be at most roughly $n/k^{3/2}$ (as promised by Corollary 2.3.2), it is necessary that $k^{3/2} < n$, since a simple graph cannot have rainbow girth less than three. Corollary 2.3.2 can be interpreted as saying that, for the region where it makes sense (where $k^{3/2} < n$, roughly), when we relax the number of colors slightly from n to $n+k$, we obtain a much shorter rainbow cycle of length at most approximately $n/k^{3/2}$, in comparison to the tight bound of n/k for the case of n colors.

Next, we present a result of a similar flavor for the case where k is large relative to n. The argument uses a different method to obtain a large rainbow subgraph which is the disjoint union of stars, and then applies the same contraction method as for Theorem 2.3.1.

Theorem 2.3.3 *Let $k > 1$ be an integer, and let G be a simple graph on n vertices. Suppose we have a coloring of the edges of G with $n + k$ color classes of size at least ck, where $c = 0.99 \cdot 10^9$, and also that $140k^{10/9} \geq n$. Then G has rainbow girth at most 6.*

In the proof, we will need the following definitions. Let a *colorful star* be a subgraph H of G such that H is a star with at least $\frac{ck^2}{4n}$ edges such that no color appears more than $c^{2/3}k^{2/3}$ times in $E(H)$. Let a *collection of colorful stars* be a set $C = \{H_1, H_2, \cdots, H_m\}$ of colorful stars such that every color appears in at most one of the $E(H_i)$.

Proof. We may assume that $n \geq ck$ since otherwise we have at least $nck > n^2$ edges which is a contradiction since G is simple. For a collection C of colorful stars, for $1 \leq i \leq m$, let v_i be the center of the star H_i, and let $V(C) = \{v_1, \cdots, v_p\}$ and $E(C) = \cup_{i=1}^{p} E(H_i)$ be the set of all star centers and the set of all edges, respectively.

Now, let C be a collection of colorful stars in G, chosen to be maximal with respect to the number of stars. We first prove the following claim, which says that the number of colors appearing in $E(C)$ is large.

Claim 2.3.4 *At most $k/2$ colors do not appear in $E(C)$.*

Proof. We proceed by a proof by contradiction. Suppose that more than $k/2$ colors do not appear in $E(C)$. Let S be the set of colors which do not appear in any of the $E(v_i)$. Note that $|S| > k/2$. For each color $s \in S$, for $v \in V(G)$ let $d_s(v)$ be the number of edges incident to v of color s, and set $d'_s(v) = d_s(v)$ if $d_s(v) \leq c^{2/3}k^{2/3}$, and otherwise set $d'_s(v) = 0$. Now, let H be the set of vertices $v \in V(G)$ with $d_s(v) > c^{2/3}k^{2/3}$. Note that $|H| < 2ck/(c^{2/3}k^{2/3}) = 2c^{1/3}k^{1/3}$. Then the number of edges of color s with both ends in H is at most $4c^{2/3}k^{2/3}$, so it follows that:

$$\sum_{v \in V(G)} d'_s(v) \geq ck - 4c^{2/3}k^{2/3} \geq \frac{ck}{2}$$

since the last inequality is equivalent to $k \geq 8^3/c$ which is true for $k > 1$ since $c = 0.99 \cdot 10^9$. Then, on average, a vertex v has:

$$\sum_{s \in S} d'_s(v) > \frac{ck^2}{4n}.$$

Now, let v be a vertex for which $\sum_{s \in S} d'_s(v) > \frac{ck^2}{4n}$, and construct a colorful star with center v and $d'_s(v)$ edges of color s incident with v for all $s \in S$. Then we add v to C and obtain a larger collection of colorful stars, which contradicts the maximality of C. It follows that there are at most $k/2$ colors which do not appear in $E(C)$, as desired. This proves Claim 2.3.4. ∎

We now prove a second claim, which says the number of colorful stars in C is small.

Claim 2.3.5 $|C| < \frac{n^{1/5}}{12}$.

Proof. We proceed by a proof by contradiction. Suppose that $|C| \geq \frac{n^{1/5}}{12}$. It suffices to show a contradiction for the case where $t = |C| = \lceil \frac{n^{1/5}}{12} \rceil$, so that $\frac{n^{1/5}}{12} \leq t < \frac{n^{1/5}}{12} + 1 \leq \frac{n^{1/5}}{6}$ since $n \geq ck \geq 10^9 k \geq 12^5$. Now, for a colorful star H_i with center v_i, let $M(H_i) = V(H_i) \setminus \{v_i\}$. We claim that for any two colorful stars $H_i, H_j \in C$ with centers v_i and v_j, if $H_{ij} = M(H_i) \cap M(H_j)$, then either v_1 has all its edges in H_1 to H_{ij} in the same color class, or v_2 has all its edges in H_2 to H_{ij} in the same color class. Suppose not. Then without loss of generality there are two edges $e_1 = (v_i, w_1)$ and $e_2 = (v_i, w_2)$ for $w_1, w_2 \in H_{ij}$ such that e_1 and e_2 have colors a_1 and a_2 with $a_1 \neq a_2$. Let a_3 be the color of (v_j, w_1). Then clearly a_3 is also the color of (v_j, w_2), since otherwise we obtain a rainbow cycle of length 4. Now, consider an arbitrary edge (v_j, w_3) to a vertex $w_3 \in H_{ij}$ with $w_3 \notin \{w_1, w_2\}$. We claim that (v_j, w_3) has color a_3. Indeed, if (v_i, w_3) does not have color a_1 then the 4-cycle (v_i, w_1, v_j, w_3) implies that (v_j, w_3) has color a_3, and if (v_i, w_3) does not have color a_2 then the 4-cycle (v_i, w_2, v_j, w_3) implies that (v_j, w_3) has color a_3. Since (v_i, w_3) cannot have both color a_1 and color a_2 it follows that (v_j, w_3) has color a_3 for all $w_3 \in H_{ij}$, as desired.

This implies that for all $v_i, v_j \in V(C)$ we have $|M(v_i) \cap M(v_j)| \leq c^{2/3} k^{2/3}$. Then it follows that every colorful star H_i has at least:

$$\frac{ck^2}{4n} - tc^{2/3}k^{2/3}$$

vertices in $M(H_i)$ which are not in $\cup_{j \neq i} M(H_j)$. The condition $140k^{10/9} \geq n$ implies $k \geq \frac{n^{9/10}}{140^{9/10}}$, and we have that:

$$\frac{tck^2}{8n} \geq tc\frac{n^{9/5}}{8n140^{9/5}} \geq \frac{10^9 n}{96 \cdot 140^{9/5}} > n.$$

We also have that:

$$\frac{tck^2}{8n} > t^2 c^{2/3} k^{2/3}$$

since the inequality is equivalent to $c^{1/3}k^{4/3} > 8nt$, which is true since (using $k^{10/9} \geq n/140$ and $t \leq n^{1/5}/6$ from above):

$$c^{1/3}k^{4/3} > \frac{8 \cdot 140^{6/5} k^{4/3}}{6} \geq \frac{8n^{6/5}}{6} \geq 8nt.$$

Then we have that:

$$\begin{aligned}\left|\bigcup_{H_i \in C} V(H_i)\right| &\geq t\left(\frac{ck^2}{4n} - tc^{2/3}k^{2/3}\right) \\ &= \frac{tck^2}{8n} + \frac{tck^2}{8n} - t^2c^{2/3}k^{2/3} \\ &> n + t^2c^{2/3}k^{2/3} - t^2c^{2/3}k^{2/3} \\ &> n\end{aligned}$$

which gives a contradiction. This proves Claim 2.3.5. ∎

Now, for each color class with at least one edge in $E(C)$, we choose exactly one such edge. Let the resulting set of edges be F; from Claim 2.3.4, we know that $|F| \geq n + \frac{k}{2}$. Now, let H be the subgraph with $V(H) = \bigcup_{(uv)\in F}\{u, v\}$ and $E(H) = F$, and let $S = \{v_1, v_2, \cdots, v_p\}$, where $p = |S|$. Partition $V(H) \setminus S$ into $X_1, \cdots, X_p$ such that $X_i \subseteq N_H(v_i)$ for all $1 \leq i \leq p$. Now, contract each $H_i = X_i \cup \{v_i\}$ to a single vertex, and let the resulting graph be H'. By Claim 2.3.5, we have that $|V(H')| < \frac{n^{1/5}}{12}$, and, since $k^{10/9} \geq n/140$ and $n \geq c = 0.99 \cdot 10^9$, we obtain:

$$|E(H')| = |V(H')| + \frac{k}{2} \geq \frac{k}{2} \geq \frac{n^{9/10}}{2 \cdot 140^{9/10}} > \frac{n^{2/5}}{144} > |V(H')|^2.$$

Thus we obtain a rainbow cycle of length at most 2 in H', which gives a rainbow cycle of length at most 6 in H, as desired. This proves Theorem 2.3.3. ∎

We conclude this section with an immediate corollary of the above results which will be used in the proof of the next section:

Corollary 2.3.6 *Let $k > 1$ be an integer, and let G be a simple graph on n vertices. Suppose we have a coloring of the edges of G with $n + k$ color classes of size at least ck, where $c = 0.99 \cdot 10^9$. Then G has rainbow girth at most n/k.*

Proof. If $28k \log k \leq n$, then by Theorem 2.3.1 we have rainbow girth at most:

$$\frac{n(\log k)^2}{10k^{3/2}} + 14 \log k \leq \frac{n}{2k} + \frac{n}{2k} = \frac{n}{k}$$

since $(\log k)^2 \leq 5\sqrt{k}$ holds for $k \geq 2$. To see this, note that it is equivalent to $\log k \leq \sqrt{5}k^{1/4}$. Let $f(k) = \sqrt{5}k^{1/4} - \log k$. We compute:

$$f'(k) = \frac{\sqrt{5}}{4k^{3/4}} - \frac{1}{k \ln 2}$$

and it follows that $f(k)$ achieves its minimum for $k_0 \geq 2$, $k \in \mathbb{R}$ at the point $k_0 = \left(\frac{4}{\sqrt{5}\ln 2}\right)^4$. We verify that $f(k_0) \geq 0$, so it follows that $f(k) \geq 0$ for all $k \geq 2$, as desired.

If $28k \log k > n$, we claim that $140k^{10/9} \geq n$. Indeed, $140k^{10/9} > 28k \log k$ is equivalent to $5k^{1/9} > \log k$ which is true for $k \geq 2$. To see this, by taking derivatives as before it suffices to verify that the inequality is true for k_0 such that $k_0^{1/9} = 9/(5 \ln 2)$, which is true. Then, Theorem 2.3.3 gives that G has rainbow girth at most 6. Since $n^2 \geq |E(G)| \geq cnk$, we have that $n/k \geq c \geq 6$, so it follows that G has rainbow girth at most n/k, as desired. This proves Corollary 2.3.6. ∎

2.4 n colors each with ck edges

Now we are ready to prove Theorem 2.1.10, which we restate. The argument uses a similar structure to the proof of Theorem 2.2.1, but with some more refined probabilistic arguments. That is, we either apply existing approximate results for the Caccetta-Häggkist conjecture, or we reduce the problem to the case of Corollary 2.3.6.

Theorem 2.4.1 *Let $k \geq 1$ be an integer, and let G be a simple graph on n vertices. Suppose we have a coloring of the edges of G with n color classes of size at least ck, where $c = 10^{11}$. Then G has rainbow girth at most n/k.*

In the proof, we will use the following definition. We say that a color a *dominates* a vertex $v \in V(G)$ if there are at least $\frac{t}{100} + 8k$ edges incident to v with color a. Call a vertex v *color-dominated* if there exists a color a which dominates v, and call a color a *vertex-dominating* if there exists a vertex v which is dominated by a. The definition is motivated by a desire to reduce to the case of the Caccetta-Häggkvist conjecture where each color class is a star centered at a different vertex, as was done by Aharoni, DeVos, and Holzman in [4]. A color being vertex-dominating means that its edges form a large star, which will be useful in applying existing approximate results for the Caccetta-Häggkvist conjecture.

Proof. If $k = 1$, then taking one edge of each color gives a rainbow cycle of length at most n. So we may assume $k > 1$. Also, since G is simple, we have that the number of edges $|E(G)|$ satisfies $n^2 \geq |E(G)| \geq nck$, and thus we may assume that $n \geq ck$. Now, let $t = ck$. By removing edges if necessary, we may assume that every color class has exactly t edges. Now, for each vertex-dominating color a, pick one vertex v_a dominated by a (not necessarily unique), and let the resulting set of vertices be S. Let $H = V(G) \setminus S$.

Suppose first that $|H| \leq \frac{t}{100}$. Let b be the coloring of the edges. We construct a digraph G' with $V(G') = S$, and for all i, j with $v_i, v_j \in S$, there is an arc $v_i \to v_j$ if

$v_iv_j \in E(G)$ and $b(v_iv_j) = i$. Every vertex v_i is incident with at least $\frac{t}{100} + 8k$ edges e with $b(e) = i$, and since $|H| \leq \frac{t}{100}$, there are at least $8k$ edges $e = v_iu$ with $b(e) = i$ and $u \in S$. Therefore, $\delta^+(G') \geq 8k$.

Now, we claim $n/(8k) + 74 \leq n/k$, which is equivalent to $n \geq \frac{592k}{7}$ which is true since $n \geq ck = 10^{11}k$.

Then, by applying Theorem 2.1.14 to G' we obtain a directed cycle K of length at most $\lceil n/(8k) \rceil + 73 \leq n/k$ in G'. The edges of G that correspond to arcs of K form a rainbow cycle of length at most n/k in G.

So we may assume that $|H| > \frac{t}{100}$. Let $r = |H|$, so we have $\frac{t}{100} < r \leq n$. Let $T \subseteq H$ be a random set of vertices in H where each vertex in H is included in T independently with probability $\frac{4k}{r}$.

Now consider a color a which does not dominate a vertex in S (and thus does not dominate any vertex). Let $s = 0.99t = 0.99 \cdot 10^{11}k$. We will show that the probability that a has at least $s/100$ edges with both ends in $G \setminus T$ is at least $1 - \frac{k}{2r}$. We claim that we may assume that at least s of the edges of a have both ends in H. Indeed, if this is not the case, perform the following iterative process while there is still an edge e of color a not contained in H.

If e has both ends in $G \setminus H$, then remove e. Now, note that at most 200 vertices are incident to at least $t/100$ edges of color a. Since $|H| > \frac{t}{100} = 2^9k$, there exists a pair of vertices $v_1, v_2 \in H$ such that there is no edge of color a between v_1 and v_2 and v_1, v_2 are both incident to less than $t/100$ edges of color a. Then add an edge of color a between v_1 and v_2. If instead e has one end in H, say the vertex w, then remove e and if there is a vertex $v \in H$ such that there is not already an edge between v and w of color a and both v and w are incident to less than $t/100$ edges of color a, then add an edge of color a from w to any vertex $v \in H$. Repeat this process until we obtain a graph G' where all the edges of a have both ends in H.

We claim that when this process terminates at a graph G', we have that each color class has at least s edges in G'. Indeed, note that for each vertex $w \in H$ there are at most $8k + 200 \leq 208k$ edges incident with w that are removed from the graph without being replaced. Also, the number of such problematic vertices is at most:

$$\frac{2t}{\frac{t}{100} - 200} \leq \frac{2 \cdot 10^{11}}{10^9 - 200}$$

Then we have that the total number of edges of color a which are deleted without replacement is at most:

$$\frac{416 \cdot 10^{11}}{10^9 - 200}k < 0.01 \cdot 10^{11}k$$

since $416 < 0.01(10^9 - 200)$. It follows that each color has at least s edges in G'.

Then, if we show that for G' the probability that a has at least $s/100$ edges in $G' \setminus T$ is at least $1 - \frac{k}{2r}$, then it clearly follows that the probability that a has at least $s/100$ edges in $G \setminus T$ is also at least $1 - \frac{k}{2r}$. Thus, we may assume without loss of generality that at least s edges of a have both ends in H, as claimed.

Now, let the edges of a with both ends in H be $\{e_1, \cdots, e_s\}$. Let the random variable E_i have value 1 if $e_i \in G \setminus T$ and have value 0 otherwise. Let $E = \sum_{i=1}^{s} E_i$, and for a random variable R let $\mathrm{Var}(R)$ denote the variance of R. Since a is not vertex-dominating, we have that each edge e_i shares an end with at most $\frac{t}{50} + 16k$ edges of the same color. It follows that each E_i is dependent on at most $\frac{t}{50} + 16k$ of the variables $\{E_1, E_2, \cdots, E_s\}$. Let $x = \frac{r-4k}{r}$, and note that the probability that an edge e_i is in $G \setminus T$ is simply x^2, so for all $1 \leq i \leq t$ we have that E_i is a Bernoulli random variable with probability equal to x^2. Then it follows that $\mathrm{Var}(E_i) = x^2(1 - x^2)$ and furthermore, if e_i and e_j share an end, we obtain:

$$\begin{aligned}\mathrm{Cov}(E_i, E_j) &= \mathbb{E}(E_i E_j) - \mathbb{E}(E_i)\mathbb{E}(E_j) \\ &= x^3 - x^4\end{aligned}$$

and thus we have:

$$\begin{aligned}\mathrm{Var}(E) &= \sum_{i=1}^{s} \mathrm{Var}(E_i) + \sum_{1 \leq i \neq j \leq s} \mathrm{Cov}(E_i, E_j) \\ &\leq f(x) := sx^2(1 - x^2) + \left(\frac{s}{0.99 \cdot 50} + 16k\right) s(x^3 - x^4).\end{aligned}$$

Claim 2.4.2 *Let $\alpha = 1 - \frac{400}{c}$. For all $\alpha \leq y < 1$ we have:*

$$f(y) \leq s^2 \left(y^2 - \frac{1}{100}\right)^2 \frac{k}{2r}.$$

Proof. Since $s/100 < t/100 < r$, we have that $\frac{100}{c} > \frac{k}{r} > 0$ and thus $\alpha = 1 - \frac{400}{c} < y < 1$. Define $g(y)$ as follows:

$$g(y) = s^2 \left(y^2 - \frac{1}{100}\right)^2 \frac{1 - y}{8}.$$

We claim that $f(y) \leq g(y)$ for all $\alpha \leq y < 1$. To see this, let $h(y) = g(y) - f(y)$. Then $h(y) \geq 0$ is equivalent to:

$$h_1(y) = \frac{h(y)}{s(1 - y)} = \frac{s\left(y^2 - \frac{1}{100}\right)^2}{8} - y^2 - \left(\frac{s}{49.5} + 16k + 1\right) y^3 \geq 0.$$

We claim that $h_1(\alpha) \geq 0$ and $h_1'(y) \geq 0$ for all $\alpha \leq y < 1$. For the first claim, since $s = 0.99 \cdot 10^{11} k$ and $\alpha = 1 - \frac{400}{c}$, we have that:

$$\frac{s\left(\alpha^2 - \frac{1}{100}\right)^2}{8} \geq \frac{s}{32} \geq \frac{s}{49.5} + 16k + 2 \geq \alpha^2 + \left(\frac{s}{49.5} + 16k + 1\right) \alpha^3.$$

To show $h_1'(y) \geq 0$ for all $\alpha \leq y < 1$, we compute:

$$h_1'(y) = \frac{s}{2}\left(y^2 - \frac{1}{100}\right) y - 2y - 3\left(\frac{s}{49.5} + 16k + 1\right) y^2.$$

Since $y > 0$, $h_1'(y) \geq 0$ is equivalent to $h_2(y) \geq 0$, where:

$$h_2(y) = \frac{s}{2}\left(y^2 - \frac{1}{100}\right) - 2 - 3\left(\frac{s}{49.5} + 16k + 1\right) y.$$

Now, we claim that $h_2(\alpha) \geq 0$ and $h_2'(y) \geq 0$ for $\alpha \leq y < 1$. The first claim follows from the facts $s = 0.99 \cdot 10^{11} k$ and $\alpha = 1 - \frac{400}{c}$:

$$\frac{s\left(\alpha^2 - \frac{1}{100}\right)}{2} \geq \frac{s}{4} \geq 2 + 3\left(\frac{s}{49.5} + 16k + 1\right) \geq 2 + 3\left(\frac{s}{49.5} + 16k + 1\right)\alpha.$$

To show $h_2'(y) \geq 0$ for all $\alpha \leq y < 1$, we compute:

$$h_2'(y) = sy - 3\left(\frac{s}{49.5} + 16k + 1\right).$$

Now, $s\alpha - 3\left(\frac{s}{49.5} + 16k + 1\right) \geq 0$ for all $k \geq 1$, so it follows that $h_2'(y) \geq h_2'(\alpha) \geq 0$ for all $\alpha \leq y < 1$. This implies that $h_1(y) \geq 0$ for all $\alpha \leq y < 1$, which in turn gives $h(y) \geq 0$ for all $\alpha \leq y < 1$. Thus $f(y) \leq g(y)$ for all $\alpha \leq y < 1$, and we obtain:

$$f(y) \leq g(y) \leq s^2\left(y^2 - \frac{1}{100}\right)^2 \frac{k}{2r}.$$

as desired. This completes the proof of Claim 2.4.2. ∎

Now, Claim 2.4.2 gives:

$$\mathrm{Var}(E) \leq f(x) \leq s^2\left(x^2 - \frac{1}{100}\right)^2 \frac{k}{2r}.$$

Let $\lambda = s\left(x^2 - \frac{1}{100}\right)$. Then we have shown that $\mathrm{Var}(E) \leq \lambda^2 \frac{k}{2r}$. Let q be the probability that the color a has at least $s/100$ of its edges in $G \setminus T$. Note that $\mathbb{E}(E) = sx^2$, so:

$$1 - q = \mathbb{P}\left(E \leq s/100\right) = \mathbb{P}(\mathbb{E}(E) - E \geq \lambda).$$

Then by Theorem 2.1.17 (Chebyshev's Inequality), we have:

$$1 - q \leq \mathbb{P}(|E - \mathbb{E}(E)| \geq \lambda) \leq \frac{\mathrm{Var}(E)}{\lambda^2} \leq \frac{k}{2r}.$$

We say that a color a is *bad* if a is not vertex-dominating and a has less than $s/100$ of its edges in $G \setminus T$. Let B be the set of bad colors, and let $Y = |B|$. Since $1 - q \leq \frac{k}{2r}$, we have that $\mathbb{E}(Y) \leq k/2$. It follows from Markov's Inequality that $\mathbb{P}(Y \geq k) \leq 1/2$. Recall that T was formed by choosing each vertex in H independently with probability $4k/r$. Then $\mathbb{E}(|T|) = 4k$. Applying Theorem 2.1.16 yields that for all $k \geq 2$:

$$\mathbb{P}(|T| \geq 8k) + \mathbb{P}(|T| \leq 2k) \leq \exp(-4k/3) + \exp(-k/2) < 1/2.$$

Since $k > 1$ is an integer, it follows that with positive probability we have both $2k < |T| < 8k$ and $Y < k$, so there exists a set $T \subset G \setminus S$ with $|T| \leq 8k$ and such that $|T| - Y \geq k$. If $G' = G \setminus T$, then since $|T| \leq 8k$ it follows that for every vertex-dominating color class at least $s/100$ of its edges are in G'. Then we have that at least $|V(G')| + k$ colors a have at least $s/100$ edges in G'. Applying Corollary 2.3.6 to G' gives that G' has rainbow girth at most $|V(G')|/k$ and thus G has rainbow girth at most n/k, as desired. This completes the proof. ∎

2.5 Triangle case: $(1 + \delta)n$ colors

In this section, we consider the case where we have at least $(1+\delta)n$ color classes each of size at least tn and we want to find a rainbow triangle. We show the following general result. At a high-level, the proof either reduces to existing results on the Caccetta-Häggkvist conjecture, uses the fact that not that many color classes form a large star to improve a bound in a triangle counting argument made by Aharoni, DeVos, and Holzman in [4], or uses the existence of two color classes with large stars centered at the same vertex to find improvements to the same triangle counting argument.

Theorem 2.5.1 *Let t, δ be positive real numbers. Suppose G is a simple edge-colored graph with at least $(1+\delta)n$ color classes of size at least tn. Then there exists a rainbow triangle if there exists $0 < \epsilon < 1/2$ such that all of the following conditions hold:*

$$\frac{(1 + \delta - \epsilon\delta)t}{2} < \frac{4}{3}(1 + \delta)\left((t(1 + \delta) - \frac{1}{4}\right)$$

$$\frac{8}{3t}\left(t(1 + \delta) - \frac{1}{4}\right)(1 + \delta) + \frac{t}{12}\left(\sqrt{1 - 2\epsilon} + 1\right)\left(4(1 - \epsilon) - \left(1 - \sqrt{1 - 2\epsilon}\right)^2\right) > 1 + \delta$$

$$\frac{16}{3}(1 + \delta) > \frac{2}{3t} + 1$$

Proof. It suffices to show the claim when each color class has size equal to $\lceil tn \rceil$ and when the number of colors is equal to $\lceil (1+\delta)n \rceil$, so we assume both of those conditions. Let δ_1, t_1 be real numbers such that $(1+\delta_1)n = \lceil (1+\delta)n \rceil$ and $t_1 n = \lceil tn \rceil$. We proceed by a proof by contradiction. Suppose that there are no rainbow triangles. Now, we say that a color class c is *good* if there are at least $(1 - \epsilon)\frac{n^2}{2}t_1^2$ triangles in G with at least two of the three edges having color c. Now, we first show the following claim.

Claim 2.5.2 *More than n of the colors are good.*

Proof. We proceed by a proof by contradiction. Suppose that at most n of the colors are good. We have that at least $\delta_1 n$ of the colors are not good, namely that at least $\delta_1 n$ of the colors have at most $(1-\epsilon)\frac{n^2}{2}t_1^2$ triangles with at least two of the three edges having color c. For every triangle T in G, at least two of the edges in T have the same color. It follows that:

$$\begin{aligned} t(G) &\leq n(1+\delta_1)\binom{t_1 n}{2} - n\epsilon\delta_1 \frac{t_1^2 n^2}{2} \\ &\leq n(1+\delta_1)\frac{t_1^2 n^2}{2} - n\epsilon\delta_1 \frac{t_1^2 n^2}{2} \\ &= n^3(1+\delta_1-\epsilon\delta_1)\frac{t_1^2}{2} \end{aligned}$$

By Theorem 2.1.18, since $m = (1+\delta_1)t_1 n^2$, we have that:

$$t(G) \geq \frac{4m}{3n}\left(m - \frac{n^2}{4}\right) = n^3 \frac{4(1+\delta_1)t_1}{3}\left((1+\delta_1)t_1 - \frac{1}{4}\right)$$

We know by assumption that the following holds:

$$\frac{4}{3}(1+\delta)\left(t(1+\delta) - \frac{1}{4}\right) - \frac{(1+\delta-\epsilon\delta)t}{2} > 0$$

We claim that for all $t_1 \geq t$ and $\delta_1 \geq \delta$, we have:

$$\frac{4}{3}(1+\delta_1)\left(t_1(1+\delta_1) - \frac{1}{4}\right) - \frac{(1+\delta_1-\epsilon\delta_1)t_1}{2} > 0$$

To show this, it suffices to show that the derivative of the above expression is positive at all points (t_1, δ_1) with $t_1 \geq t$ and $\delta_1 \geq \delta$. Indeed, the derivative of the above expression with respect to t at the point (t_1, δ_1) is:

$$\frac{4(1+\delta_1)^2}{3} - \frac{1+\delta_1-\epsilon\delta_1}{2} > \frac{4}{3}(1+\delta_1)^2 - \frac{1}{2}(1+\delta_1) > 0$$

since $1+\delta_1 > \frac{3}{8}$.

Now, by assumption we have that:

$$\begin{aligned} \frac{16}{3}(1+\delta_1) &\geq \frac{16}{3}(1+\delta) \\ &> \frac{2}{3t} + 1 \\ &\geq \frac{2}{3t_1} + 1 \\ &> \frac{2}{3t_1} + (1-\epsilon) \end{aligned}$$

Taking the derivative with respect to δ at the point (t_1, δ_1) then gives:

$$\frac{8t_1(1+\delta_1)-1}{3} - \frac{(1-\epsilon)t_1}{2} > 0$$

and therefore we have:

$$\frac{(1+\delta_1-\epsilon\delta_1)t_1}{2} < \frac{4}{3}(1+\delta_1)\left(t_1(1+\delta_1)-\frac{1}{4}\right)$$

Thus, we obtain a contradiction. ∎

So, by Claim 2.5.2 we have that more than n colors are good. Next, we show the following.

Claim 2.5.3 *Suppose that color c is good. Then there exists a vertex in G incident to at least αn edges of c, where $\alpha = \frac{t_1}{2}(\sqrt{1-2\epsilon}+1)$.*

Proof. We proceed by a proof by contradiction. Suppose that color c is good and each vertex of G is incident to less than αn edges of c. Observe that the maximum number of two-edge paths which have both edges of color c will be obtained when the edges of color c form two stars, such that there is an edge of color c between the centers of the stars. It follows that the number of two-edge paths which have both edges of color c is at most:

$$\binom{\alpha n}{2} + \binom{(t_1-\alpha)n}{2} + (t_1-\alpha)n$$

Since c is good, we have that the number of two edge monochromatic paths of color c is at least $(1-\epsilon)\frac{n^2t_1^2}{2}$, so we get that:

$$\begin{aligned}
(1-\epsilon)\frac{n^2t_1^2}{2} &\le \binom{\alpha n}{2} + \binom{(t_1-\alpha)n}{2} + (t_1-\alpha)n \\
&= \frac{(\alpha n)^2-\alpha n}{2} + \frac{(t_1n-\alpha n)^2-(t_1n-\alpha n)}{2} + (t_1-\alpha)n \\
&= \frac{\alpha^2n^2}{2} + \frac{(t_1-\alpha)^2n^2}{2} + \left(\frac{t_1}{2}-\alpha\right)n \\
&< \frac{\alpha^2n^2}{2} + \frac{(t_1-\alpha)^2n^2}{2}.
\end{aligned}$$

since $\alpha = \frac{t_1}{2}(\sqrt{1-2\epsilon}+1) > \frac{t_1}{2}$. Cancelling a factor of $\frac{n^2}{2}$ immediately gives:

$$(1-\epsilon)t_1^2 < \alpha^2 + (t_1-\alpha)^2.$$

Now, $\alpha = \frac{t_1}{2}(\sqrt{1-2\epsilon}+1)$ implies that:

$$\begin{aligned}
\alpha^2 + (t_1-\alpha)^2 &= \left(\frac{t_1}{2}\left(1+\sqrt{1-2\epsilon}\right)\right)^2 + \left(\frac{t_1}{2}\left(1-\sqrt{1-2\epsilon}\right)\right)^2 \\
&= \frac{t_1^2}{4}\left(2+2(1-2\epsilon)\right) \\
&= (1-\epsilon)t_1^2 \\
&< \alpha^2 + (t_1-\alpha)^2
\end{aligned}$$

which is a contradiction, as desired. This completes the proof. ∎

Now, since more than n colors are good, Claim 2.5.3 implies that there exists a vertex $v \in V(G)$ such that there exist two colors c_1 and c_2 which both have at least αn edges of their color incident to v, where $\alpha = \frac{t_1}{2}(\sqrt{1-2\epsilon}+1)$. Let S_1 be the set of vertices which v has an edge of color c_1 to, so that $|S_1| \geq \lceil \alpha n \rceil$, and let S_2 be the vertices which v has an edge of color c_2 to, so that $|S_2| \geq \lceil \alpha n \rceil$. Then since c_1 is good, by counting good triangles, it follows that the number of edges with both ends in S_1 is at least:

$$\begin{aligned}
(1-\epsilon)\frac{n^2t_1^2}{2} - \binom{t_1n-|S_1|}{2} - (t_1n-|S_1|) &= (1-\epsilon)\frac{n^2t_1^2}{2} - \frac{(t_1n-|S_1|)^2}{2} - \frac{t_1n-|S_1|}{2} \\
&\geq (1-\epsilon)\frac{n^2t_1^2}{2} - \frac{(t_1n-\lceil\alpha n\rceil)^2}{2} - \frac{t_1n-\lceil\alpha n\rceil}{2},
\end{aligned}$$

and since c_2 is good, an identical argument gives that the same is true for S_2.

Now, if there is no rainbow triangle in G, then the only edges between S_1 and S_2 will have color c_1 or c_2, and it follows that there are at most $2t_1n - |S_1| - |S_2|$ edges between S_1 and S_2. Now, let H be the graph on three vertices with exactly one edge. For each edge with both ends in S_1, if neither of its ends have an edge to S_2 then we get at least $\lceil \alpha n \rceil$ induced copies of H in G (each containing that edge and a vertex from S_2), and likewise the same is true for each edge in S_2. Let $h(G)$ denote the number of induced copeis of H in G. Each edge with one end in S_1 and one end in S_2 will be incident to at most $|S_1| + |S_2|$ edges contained in either S_1 or S_2, and it follows that $h(G)$ is at least:

$$2\lceil\alpha n\rceil\left((1-\epsilon)\frac{t_1^2n^2}{2} - \frac{(t_1n-\lceil\alpha n\rceil)^2}{2} - \frac{t_1n-\lceil\alpha n\rceil}{2}\right) - (2t_1n-|S_1|-|S_2|)\left(|S_1|+|S_2|\right)$$

which, using the fact that $|S_1| \geq \alpha n > \frac{t_1n}{2}$ and likewise for $|S_2|$, is at least:

$$2\lceil\alpha n\rceil\left((1-\epsilon)\frac{t_1^2n^2}{2} - \frac{(t_1n-\lceil\alpha n\rceil)^2}{2}\right) - \lceil\alpha n\rceil(t_1n-\lceil\alpha n\rceil) - 2\lceil\alpha n\rceil\left(2t_1n-2\lceil\alpha n\rceil\right)$$

which equals:

$$2\lceil \alpha n\rceil \left((1-\epsilon)\frac{t_1^2 n^2}{2} - \frac{(t_1 n - \lceil \alpha n\rceil)^2}{2}\right) - 5\lceil \alpha n\rceil (t_1 n - \lceil \alpha n\rceil)$$

Observing again that $\alpha > t_1/2$, we obtain that:

$$h(G) \geq 2\alpha n \left((1-\epsilon)\frac{t_1^2 n^2}{2} - \frac{(t_1 n - \alpha n)^2}{2}\right) - 5(t_1 n - \alpha n)(\alpha n)$$

Now, we will use this fact about the number of induced copies of H to improve the lower bound on the number of triangles in Theorem 2.1.18, in the following:

Claim 2.5.4 *Let G be a simple graph on n vertices with m edges, and let $h(G)$ denote the total number of induced copies of H in G. Then the number of triangles in G is at least:*

$$\frac{h(G)}{3} + \frac{4m}{3n}\left(m - \frac{n^2}{4}\right)$$

Proof. Note that for each edge $e = (uv) \in E(G)$, we have that the number of triangles it is contained in is at least $\delta(u) + \delta(v) - n + h_e$, where h_e denotes the number of copies of H in G containing e. Let $t(G)$ denote the number of triangles in G. Summing this over all edges $e \in E(G)$, and noting that we count each triangle at most three times, we obtain:

$$\begin{aligned} t(G) &\geq \frac{1}{3}\sum_{e=(uv)\in E(G)} (\delta(u) + \delta(v) - n + h_e) \\ &= \frac{h(G)}{3} - \frac{mn}{3} + \frac{1}{3}\sum_{v\in V(G)} \delta(v)^2. \end{aligned}$$

and by the Cauchy-Schwarz inequality, we obtain that this is at least:

$$\frac{h(G)}{3} - \frac{mn}{3} + \frac{4m^2}{3n} = \frac{h(G)}{3} + \frac{4m}{3n}\left(m - \frac{n^2}{4}\right),$$

as desired. This completes the proof. ∎

So, in our case, since $m = (1+\delta)t_1 n^2$, Claim 2.5.4 gives that the number of triangles in G is at least:

$$\frac{1}{3}\left(2\alpha n \left((1-\epsilon)\frac{t_1^2 n^2}{2} - \frac{(t_1 n - \alpha n)^2}{2}\right)\right) - \frac{5}{3}(t_1 n - \alpha n)(\alpha n) + n^3 \frac{4(1+\delta_1)t_1}{3}\left((1+\delta_1)t_1 - \frac{1}{4}\right)$$

Now, if there is no rainbow triangle, then each triangle contains at least two edges of the same color, and it follows that the number of triangles is at most:

$$(1+\delta_1)n\binom{t_1 n}{2} = (1+\delta_1)n\frac{t_1 n(t_1 n - 1)}{2}$$

We claim that the following is true, which will give an immediate contradiction:

$$\frac{1}{3}\left(2\alpha n\left((1-\epsilon)\frac{t_1^2 n^2}{2} - \frac{(t_1-\alpha)^2 n^2}{2}\right)\right) - \frac{5}{3}\alpha(t_1-\alpha)n^2 + n^3\frac{4(1+\delta_1)t_1}{3}\left((1+\delta_1)t_1 - \frac{1}{4}\right)$$
$$> (1+\delta_1)n\frac{t_1 n(t_1 n-1)}{2} \qquad (2.1)$$

Note that we may assume that $5t_1\epsilon < 1$ because $\epsilon \leq 1/2$ and $t_1 < 2/5$ since otherwise we are done by Theorem 2.1.3. Then the assumption that $5t_1\epsilon < 1$ implies that:

$$\begin{aligned}
\frac{5}{3}\alpha(t_1-\alpha) &< 5\alpha(t_1-\alpha) \\
&= \frac{5t_1^2}{4}(1+\sqrt{1-2\epsilon})(1-\sqrt{1-2\epsilon}) \\
&= \frac{5t_1^2\epsilon}{2} \\
&< \frac{t_1}{2} \\
&< \frac{t_1(1+\delta_1)}{2}
\end{aligned}$$

since $5\epsilon t_1 < 1$. It follows that (2.1) is implied by:

$$\frac{1}{3}\left(2\alpha n\left((1-\epsilon)\frac{t_1^2 n^2}{2} - \frac{(t_1-\alpha)^2 n^2}{2}\right)\right) + n^3\frac{4(1+\delta_1)t_1}{3}\left((1+\delta_1)t_1 - \frac{1}{4}\right) > (1+\delta_1)n\frac{t_1^2 n^2}{2}$$

which, after plugging in $\alpha = \frac{t_1}{2}(1+\sqrt{1-2\epsilon})$ and cancelling $\frac{t_1 n^3}{2}$ from both sides, is equivalent to:

$$\frac{8}{3}\left(t_1(1+\delta_1) - \frac{1}{4}\right)(1+\delta_1) + \frac{t_1^2}{12}(\sqrt{1-2\epsilon}+1)\left(4(1-\epsilon) - \left(1-\sqrt{1-2\epsilon}\right)^2\right) - (1+\delta_1)t_1 > 0$$

Now, we claim that the derivative of the above expression with respect to t_1 and with respect to δ_1 is positive for all (t_1, δ_1) with $t_1 \geq t$ and $\delta_1 \geq \delta$. Indeed, the derivative with respect to t_1 is equal to:

$$\begin{aligned}
\frac{8}{3}(1+\delta_1)^2 + \frac{t_1}{6}\left(\sqrt{1-2\epsilon}+1\right)\left(4(1-\epsilon) - \left(1-\sqrt{1-2\epsilon}\right)^2\right) - (1+\delta_1) &> \frac{8}{3}(1+\delta_1)^2 - (1+\delta_1) \\
&> 0
\end{aligned}$$

since $\epsilon < 1/2$ and $1 + \delta_1 > \frac{3}{8}$. Now, the derivative with respect to δ_1 is equal to:

$$\frac{16}{3} t_1 (1 + \delta_1) - \frac{2}{3} - t_1 > 0$$

since by assumption we have:

$$\frac{16}{3}(1 + \delta_1) \geq \frac{16}{3}(1 + \delta) > \frac{2}{3t} + 1 \geq \frac{2}{3t_1} + 1$$

Therefore, we get that the condition with δ_1 and t_1 is implied by:

$$\frac{8}{3t}\left(t(1+\delta) - \frac{1}{4}\right)(1+\delta) + \frac{t}{12}(\sqrt{1-2\epsilon}+1)\left(4(1-\epsilon) - \left(1-\sqrt{1-2\epsilon}\right)^2\right) - (1+\delta) > 0$$

which was assumed to be true. This gives the desired contradiction, and completes the proof. ∎

This immediately gives Theorem 2.1.11, which we restate here:

Theorem 2.5.5 $(1.1077, 1/3)$ *is triangular;* $(1.3481, 1/4)$ *is triangular.*

Proof. One can verify that $t = 1/3$, $\delta = 0.1077$, and $\epsilon = 0.4746$ satisfy the above inequalities, and also that $t = 1/4$, $\delta = 0.3481$, and $\epsilon = 0.2774$ also satisfy the above inequalities. This completes the proof. ∎

In general, for a value of t we determine the minimum value of δ that this result gives by using the following tool, which we link here:
https://www.desmos.com/calculator/1x06qadpqp.

2.6 Triangle case: n colors of size at least tn

In this section, we show Theorem 2.1.12 by first showing the following more general statement. The proof at a high-level proceeds by either reducing to existing approximate results for the Caccetta-Häggkvist conjecture or using the fact that not that many color classes form large stars to improve the triangle counting argument made by Aharoni, DeVos, and Holzman in [4]. Since the case where there are two large stars centered at the same vertex is not a tight case here, it is actually simpler than the proof of the last section (and less effective).

Theorem 2.6.1 *For real* $0 < \epsilon < 1/2$, $0 < \delta \leq 1$, $t > 0$, *suppose that the following conditions hold:*

$$\frac{(1-\epsilon)t - \delta}{1 - \delta} \geq 0.3465$$

$$t > \frac{1}{2} - (1-\epsilon)^2 t^2$$

$$\frac{8}{3}\left(t - \frac{1}{4}\right) > (1 - \delta(2\epsilon - 2\epsilon^2))t$$

$$1 + 2\delta\epsilon^2 > 4\epsilon\delta$$

Then if G is a simple edge-colored graph with n colors classes of size at least tn, there exists a rainbow triangle in G.

Proof. Let t_1 be a real number such that $t_1 n = \lceil tn \rceil$. Now, let us call a color c *concentrated* if there exists a vertex which is incident to at least $(1-\epsilon)|C|$ edges of color c, where C is the set of edges of color c. We proceed by a proof by contradiction. Suppose that there is no rainbow triangle, and suppose first that at least δn colors are not concentrated. Since there is no rainbow triangle, then the number of triangles is at most the number of monochromatic paths of length 2. Observe that for a color c which is not concentrated, the maximum number of paths of two-edges of color c is obtained when $\lfloor (1-\epsilon)|C| \rfloor$ edges form a star and $\lceil \epsilon |C| \rceil$ edges form another star, and there is an edge of color c between the centers of the two stars, where C is the set of edges of color c. It follows that, since there are no rainbow triangles, the number of triangles is at most:

$$(1 - \delta(1 - \epsilon^2 - (1-\epsilon)^2))\frac{t_1 n^2 (t_1 n - 1)}{2} + \delta\epsilon t_1 n^2 \leq (1 - \delta(2\epsilon - 2\epsilon^2))\frac{t_1^2 n^3}{2}$$

since:

$$\frac{1 - \delta(1 - \epsilon^2 - (1-\epsilon)^2)}{2} > \delta\epsilon$$

is equivalent to:

$$1 + 2\delta\epsilon^2 > 4\epsilon\delta$$

which is true by assumption. Then by Theorem 2.1.18, we have the number of triangles is at least:

$$\frac{4m}{3n}\left(m - \frac{n^2}{4}\right) \geq \frac{4t_1 n}{3}\left(t_1 n^2 - \frac{n^2}{4}\right)$$

and since we have by assumption that:

$$\frac{8}{3}\left(t - \frac{1}{4}\right) > (1 - \delta(2\epsilon - 2\epsilon^2))t$$

this implies that:

$$\frac{8}{3}\left(t_1 - \frac{1}{4}\right) > (1 - \delta(2\epsilon - 2\epsilon^2))t_1$$

and we obtain a contradiction. Thus, we may assume that at least $(1-\delta)n$ of the colors are concentrated. Now, there are two cases. Suppose first that for each vertex

$v \in V(G)$, there is at most one color concentrated at v. Here, we proceed with a method inspired by the reduction of Conjecture 2.1.1 to Conjecture 2.1.2 made by Aharoni, DeVos, and Holzman in [4]. Let $T \subset V(G)$ be the set of vertices v for which there exists a color c_v concentrated at v. Form a digraph D with vertex set equal to T, and for each vertex v and edge $e = (vu)$ of color c_v which is incident to v, add an arc from v to u. It follows that D has minimum out-degree at least $(1-\epsilon)t_1 - (n - |T|)$, and therefore since $|T| \geq (1-\delta)n$, we have that:

$$\frac{\delta^+(D)}{|D|} \geq \frac{(1-\epsilon)t_1 n - (n - |T|)}{|T|} \geq \frac{(1-\epsilon)t_1 - \delta}{1-\delta}$$

By assumption we have that:

$$\frac{(1-\epsilon)t_1 - \delta}{1-\delta} \geq \frac{(1-\epsilon)t - \delta}{1-\delta} \geq 0.3465$$

so it follows from Theorem 2.1.19 that there exists a directed triangle T in D. Looking at the underlying edges in G corresponding to the directed edges of T, we obtain a rainbow triangle in G.

So, we may assume that there is some vertex v at which two colors c_1 and c_2 are concentrated. Let S_1 be the set of vertices u such that (vu) is an edge of color c_1 and let S_2 be the set of vertices u such that (vu) is an edge of color c_2. Then, since there are no rainbow triangles, it follows that the only edges with one end in S_1 and one end in S_2 have color either c_1 or c_2. Therefore, we get at least $(1-\epsilon)^2 t_1^2 n^2 - 2\epsilon t_1 n$ non-edges in G, and thus the number of edges is at most:

$$\frac{n(n-1)}{2} - (1-\epsilon)^2 t_1^2 n^2 + 2\epsilon t_1 n \leq \frac{n^2}{2} - (1-\epsilon)^2 t_1^2 n^2$$

since $2\epsilon t_1 < 1/2$ because $\epsilon < 1/2$ and $t_1 < 0.4$ (otherwise we are done by Theorem 2.1.3). However, we know that there are at least $t_1 n^2$ edges in G, and by assumption we have:

$$t > \frac{1}{2} - (1-\epsilon)^2 t^2$$

which implies that:

$$t_1 > \frac{1}{2} - (1-\epsilon)^2 t_1^2$$

and thus we obtain a contradiction. Therefore, G contains a rainbow triangle, as desired. This completes the proof. ∎

Now, we immediately obtain Theorem 2.1.12, which we restate here:

Theorem 2.6.2 $(1, 0.3988)$ *is triangular.*

Proof. We verify that $\epsilon = 0.03846$, $\delta = 0.0681$, and $t = 0.3988$ satisfy all the conditions of Theorem 2.6.1. Since the edges of G are colored with n colors each of size at least $tn = 0.3988n$, it follows that there exists a rainbow triangle, as desired. This completes the proof. ∎

2.7 Conclusions and further work

In this chapter, we showed a number of results concerning Aharoni's rainbow generalization of the Caccetta-Häggkvist conjecture. There are many directions in which further research can go; indeed, one can look at the Caccetta-Häggkvist conjecture literature for inspiration for some of these.

First, one could try to strengthen the constant c in Theorem 2.1.10. We made no attempt to optimize it, but it seems that new methods would have to be introduced to have it be significantly small. Another direction would be to prove Conjecture 2.1.6, which would generalize our result (up to a constant). This is interesting because, while some of our methods seem useful in the setting of non-uniform sizes of color classes, we think that some new ideas are needed.

Another interesting question is whether there exist extremal examples for Conjecture 2.1.2 which are not inherited from Conjecture 2.1.1, namely that do not have the property that for each vertex $v \in V(G)$ there exists a color a whose color class is the edge set of a star centered at v. Now, a related problem which we did not consider is a relaxation of Conjecture 2.1.2, which is the following:

Conjecture 2.7.1 (Aharoni, Devos, Holzman [4]) *Let n, k be positive integers, and let G be a simple graph on n vertices. Let b be a coloring of the edges of G with n color classes of size at least k; then G has a cycle C of length at most $\lceil n/k \rceil$ such that no two incident edges of C are the same color.*

This conjecture is interesting because it still implies Conjecture 2.1.1, but seems like it might be substantially easier than Conjecture 2.1.2, as it deals with a local condition rather than a global condition. However, we suspect it may require different methods than those used in this paper. Perhaps some of the existing methods from the Caccetta-Häggkvist conjecture literature would translate over better to Conjecture 2.7.1 than Conjecture 2.1.2.

We also note that nowhere in this chapter did we use induction, while a number of the results for Conjecture 2.1.1 utilize induction. Is there a way to use inductive arguments in this context?

On the subject of the triangle case, our methods proved far more effective for the case of $(1+\delta)n$ colors than for the case of n colors. Further work could improve upon either of the results, but we find the prospect of introducing a new method to make significant progress on the case of n colors to be interesting. For those familiar with flag algebras, it would be interesting to see whether they can be leveraged to make progress on the case of n colors.

Finally, we make the following conjecture concerning the bipartite case, which is inspired by the tight results for a bipartite version of the Caccetta-Häggkvist conjecture given by Seymour and Spirkl in [37].

Conjecture 2.7.2 *Let $G = (A, B)$ be a simple bipartite graph such that $|A| = |B| = n$, and suppose that $E(G)$ is colored with $2n$ color classes each with size more than $\frac{n}{k+1}$; then there exists a rainbow cycle of length at most $2k$.*

For the same reason that Aharoni's conjecture is a generalization of the Caccetta-Häggkvist conjecture, we have that this is a generalization of the following conjecture of Seymour and Spirkl [37]:

Conjecture 2.7.3 (Seymour, Spirkl [37]) *Let $D = (A, B)$ be a bipartite digraph such that $|A| = |B| = n$ and the minimum out-degree is more than $\frac{n}{k+1}$; then there exists a directed cycle of length at most $2k$.*

Since Seymour and Spirkl [37] are able to make substantially more progress on the bipartite version of the Caccetta-Häggkvist conjecture than has been made on the general Caccetta-Häggkvist conjecture, we wonder whether significant progress can be made towards Conjecture 2.7.2.

Chapter 3

χ-boundedness and $\vec{\chi}$-boundedness

We note that many of the words in this chapter appear verbatim from the set of papers listed in the *Statement of Contributions* section at the beginning of this thesis. These words were written by the author of this thesis.

3.1 Background and main results

Now, a fundamental question in graph theory is what substructures must a graph with large chromatic number contain? Certainly, containing a large clique implies that a graph has large chromatic number, but if a graph with bounded clique number has large chromatic number, what substructures must we find? A famous conjecture along these lines is the Gyárfás-Sumner conjecture:

Conjecture 3.1.1 (Gyárfás [22], Sumner [39]) *For any forest F, the class of graphs not containing F as an induced subgraph is χ-bounded.*

Now, in Section 2.2, we disprove the following fundamental conjecture, which would have implied that to show a class $\mathcal{C}$ is χ-bounded, one would only need to show this for the triangle-free graphs in the class (its origin appears somewhat unclear; it is attributed to Louis Esperet by Scott and Seymour in [35], while Thomassé, Trotignon, and Vušković state that "we could not find a reference" in [40]):

Conjecture 3.1.2 (Esperet (see Scott, Seymour [35])) *For all $k, r \in \mathbb{N}$ there is an $n \in \mathbb{N}$ such that for every graph G with $\chi(G) \geq n$ and $\omega(G) \leq k$, there is an induced subgraph H of G with $\chi(H) \geq r$ and $\omega(H) = 2$.*

Here, $\chi(G)$ denotes the *chromatic number* of a graph G and $\omega(G)$ denotes the *clique number*. This conjecture is the induced-subgraph analogue of the following theorem:

Theorem 3.1.3 (Rödl [34]) *For every $r \in \mathbb{N}$ there is an $n \in \mathbb{N}$ such that for every graph G with $\chi(G) \geq n$, there is a (not necessarily induced) subgraph H of G with $\chi(H) \geq r$ and $\omega(H) = 2$.*

We will show:

Theorem 3.1.4 (Carbonero, Hompe, Moore, Spirkl [12]) *For every $n \in \mathbb{N}$, there is a graph G with $\chi(G) \geq n$ and $\omega(G) \leq 3$ such that every induced subgraph H of G with $\omega(H) \leq 2$ satisfies $\chi(H) \leq 4$.*

This is almost best possible, because of the following:

Theorem 3.1.5 (Scott, Seymour [36]) *There is a function $f : \mathbb{N} \to \mathbb{N}$ such that for every graph G with no induced cycle of odd length at least 5, we have $\chi(G) \leq f(\omega(G))$.*

Letting f as in Theorem 3.1.5, it follows that every graph G with $\chi(G) > f(\omega(G))$ contains an induced cycle of odd length at least 5, and therefore contains an induced subgraph H with $\omega(H) = 2$ and $\chi(H) = 3$.

Our construction is based on a construction of Kierstead and Trotter [30] and produces a digraph with large dichromatic number, which we define below. Throughout this chapter, we only consider simple digraphs D, that is, for every two distinct vertices u and v, the digraph D contains either an edge from u to v, or an edge from v to u, or neither; but not both. For a digraph, we write uv for an edge from u to v. Given a digraph D, we define its *underlying undirected graph* G to be that graph with $V(G) = V(D)$ and in which $u, v \in V(G)$ are adjacent if D contains an edge from u to v or from v to u. The *clique number* $\omega(D)$ of a digraph D is defined as the clique number of the underlying undirected graph of D.

An analogue of chromatic number for directed graphs was introduced in [17, 32] by Erdős and Neumann-Lara, respectively. A digraph is *acyclic* if it contains no directed cycle. For $k \in \mathbb{N}$, a *k-dicoloring* of a digraph D is a function $f : V(D) \to \{1, \ldots, k\}$ such that for every $i \in \{1, \ldots, k\}$, the induced subdigraph of D with vertex set $\{v \in V(D) : f(v) = i\}$ is acyclic. The *dichromatic number* $\vec{\chi}(D)$ is the smallest integer k such that D has a k-dicoloring.

We show that the digraph analogue of Theorem 3.1.5 does not hold:

Theorem 3.1.6 (Carbonero, Hompe, Moore, Spirkl [12]) *For every n, there is a digraph D with $\vec{\chi}(D) \geq n$, $\omega(D) \leq 3$ and with no induced directed cycle of odd length at least 5.*

In the paper containing those results, since we were interested in the dichromatic number we posed various open problems. In particular, we asked if the class of digraphs where every induced directed cycle has length t, for a fixed integer $t \geq 3$, is $\vec{\chi}$-bounded. We will refer to a digraph in which all induced directed cycles have length t as *t-chordal*.

In [1], Aboulker, Bousquet, and de Verclos answered our question in the negative for $t = 3$. In Section 2.3, we extend this negative answer to all $t \geq 3$:

Theorem 3.1.7 (Carbonero, Hompe, Moore, Spirkl [13]) *For $t \geq 3$, and every $N \in \mathbb{N}$, there exists a t-chordal digraph D with $\omega(D) \leq 3$ and $\vec{\chi}(D) \geq N$, and if $t > 3$, then $\omega(D) \leq 2$.*

Then we give a positive result, which in some sense demonstrates that Theorem 3.1.7 is tight, by proving $\vec{\chi}$-boundedness for a subclass of t-chordal digraphs:

Theorem 3.1.8 (Carbonero, Hompe, Moore, Spirkl [13]) *For every integer t, there is a function $f_t : \mathbb{N} \to \mathbb{N}$ such that for every digraph D which is t-chordal and has no induced directed path with exactly t vertices, we have $\vec{\chi}(D) \leq f_t(\omega(D))$.*

Since cycles of length more than t contain a directed path on t vertices, Theorem 3.1.8 is equivalent to saying that for every t, digraphs with no induced t-vertex path and no induced cycle on strictly less than t vertices are $\vec{\chi}$-bounded.

This is a partial result towards the directed path case of the following digraph generalization of Conjecture 3.1.1:

Conjecture 3.1.9 (Aboulker, Charbit, Naserasr [2]) *For any oriented forest F, the class of digraphs not containing F as an induced subdigraph is $\vec{\chi}$-bounded.*

Despite not being $\vec{\chi}$-bounded, one could hope that t-chordal digraphs still have a "nice" structural description. We show that any clean structural description of t-chordal digraphs is unlikely, as deciding if a digraph is t-chordal is coNP-complete:

Theorem 3.1.10 (Carbonero, Hompe, Moore, Spirkl [13]) *Fix an integer $t \geq 2$. Deciding if a given digraph is t-chordal digraph is coNP-complete.*

An outline of the chapter is as follows. In Section 3.2, we present the construction which disproves Conjecture 3.1.2, and which also shows Theorem 3.1.6. Then, in Section 3.3, we present a construction showing that t-chordal digraphs are not $\vec{\chi}$-bounded, and show Theorem 3.1.8. In Section 3.4, we present the hardness result which is Theorem 3.1.10. Finally, in Section 3.5, we discuss future directions for work in this area.

3.2 Disproving Esperet's conjecture

At a high-level, our construction is inspired by the construction of Kierstead and Trotter [30], which in turn utilizes Zykov's construction from [42]. It differs in the way that additional edges are added to the digraph, and then the underlying undirected graph is what will actually give us the counterexample for Esperet's conjecture.

We construct a sequence of digraphs $\{D_n\}$ as follows. Let D_1 be the digraph with a single vertex. For $n \geq 2$, we take $n-1$ disjoint copies of the digraph D_{n-1} and call them $D_{n-1}^1, \ldots, D_{n-1}^{n-1}$. Let $\mathcal{T}$ be the set of all sequences $T = (x_1, \ldots, x_{n-1})$ with $x_i \in V(D_{n-1}^i)$ for all $i \in \{1, \ldots, n-1\}$. Now, for every $T = (x_1, \ldots, x_{n-1}) \in \mathcal{T}$ we create a vertex v_T and for every $i \in \{1, \ldots, n-1\}$, we add an edge from x_i to v_T. The resulting digraph with vertex set

$$V(D_{n-1}^1) \cup \cdots \cup V(D_{n-1}^{n-1}) \cup \{v_T : T \in \mathcal{T}\}$$

and edge set

$$E(D_{n-1}^1) \cup \cdots \cup E(D_{n-1}^{n-1}) \cup \{x_i v_T : i \in \{1, \ldots, n-1\}, T = (x_1, \ldots, x_{n-1}) \in \mathcal{T}\}$$

is called D_n.

We note that the graph D_n is the graph of red edges in the proof of Theorem 3 of [30] by Kierstead and Trotter, where the following was proved:

Lemma 3.2.1 (Kierstead and Trotter [30]) *For all $n \in \mathbb{N}$, we have:*

- *D_n is acyclic;*
- *for every two vertices $u, v \in V(D_n)$ there is at most one directed path from u to v in D_n.*

Proof. We include a proof for completeness. For $n \geq 1$, let us define a partition of $V(D_n)$ into sets $T_1^n, \ldots, T_n^n$ as follows: For $n = 1$, let $T_1^1 = V(D_1)$. For $n > 1$ and $i \in \{1, \ldots, n-1\}$, let T_i^n be the union of the sets T_i^{n-1} in $D_{n-1}^1, \ldots, D_{n-1}^{n-1}$, and let T_n^n be the set of remaining vertices (and thus T_n^n is the set of vertices v_T added when constructing D_n).

By construction we have that for all $i \in \{1, \ldots, n\}$, the set T_i^n is a stable set and the only edges between T_i^n and $T_1^n \cup \cdots \cup T_{i-1}^n$ are edges from $T_1^n \cup \cdots \cup T_{i-1}^n$ to T_i^n. It follows that D_n is acyclic, as desired.

For the second bullet, note that every edge is from T_i^n to T_j^n for some $i < j$. Now, suppose we have vertices u, v such that there exists a directed path P from u to v. Then it follows that $u \in T_i^n$ and $v \in T_j^n$ for $i < j$, and the vertex set of P is contained in $T_i^n \cup \cdots \cup T_j^n$. Let H be the copy of D_{j-1} that u is contained in from the construction of D_j. By construction, every edge of D_n with one end in H and one end x in $V(D_n) \setminus V(H)$ satisfies $x \in T_k^n$ for some $k \geq j$. Since v is the only vertex of P in $T_j^n \cup T_{j+1}^n \cup \cdots \cup T_n^n$, it follows that all vertices of $P \setminus v$ are contained in H. Note that v has exactly one in-neighbor in H; let that in-neighbor be w. It follows that any directed path from u to v must go through w. By induction on n (since $P \setminus v$ is contained in a copy of D_{j-1} with $j \leq n$), we have that there is at most one directed path from u to w, so it follows that there is at most one directed path from u to v, as desired. This completes the proof. ∎

We define the *length* of a (directed) path as its number of edges. Now, we construct a sequence of digraphs $\{D'_n\}$ as follows. We take a copy of D_n, and create a new graph D'_n with $V(D'_n) = V(D_n)$, and the following edges. For every two vertices u, v where there exists a directed path in D_n from u to v,

- we add an edge from u to v if that path has length equal to 1 modulo 3; and
- we add an edge from v to u if that path has length equal to 2 modulo 3.

From Lemma 3.2.1, it follows that D'_n is well-defined and a simple digraph. In our analysis, it will be useful to consider a partition of the edges of D'_n into two sets, positive and negative, which we call the *sign* of an edge. Let us call an edge *positive* if it was added as a result of the first bullet above, and *negative* if it was added as a result of the second bullet. Clearly, this is a partition of the edges of D'_n. Note that in particular, if $uv \in E(D_n)$, then the edge uv is added to D'_n according to the first bullet, and hence D_n is a (non-induced) subdigraph of the positive edges of D'_n.

Lemma 3.2.2 *Let $u, v, w \in V(D'_n)$. If uv and vw are edges of D'_n of the same sign, then wu is an edge of D'_n of the opposite sign.*

Proof. Suppose first that uv and vw are positive edges. Then by definition there exists a path P_1 from u to v in D_n with length equal to 1 modulo 3, and a path P_2 from v to w in D_n with length equal to 1 modulo 3. Then clearly $P_3 = P_2 \cup P_1$ is a directed walk from u to w, and since D_n is acyclic by Lemma 3.2.1, it follows that P_3 is the unique directed path from u to w. Then P_3 has length equal to 2 modulo 3, so it follows that wu is a negative edge, as desired.

Suppose instead that uv and vw are negative edges. Then there exists a path P_1 from v to u and a path P_2 from w to v such that P_1 and P_2 both have length equal to 2 modulo 3. Then clearly $P_3 = P_2 \cup P_1$ is a directed walk from w to u, and since D_n is acyclic by Lemma 3.2.1, it follows that P_3 is the unique path from w to u. Then P_3 has length equal to 1 modulo 3 and it follows that wu is a positive edge, as desired. This completes the proof. ∎

Lemma 3.2.3 *Let $u, v, w \in V(D'_n)$. Then not all of uv, vw, uw are edges of D'_n.*

Proof. We only consider the case when uw is positive; the case when uw is negative is analogous. It follows that there is a directed path P_1 from u to w of length congruent to 1 modulo 3. By Lemma 3.2.2, we may assume that uv and vw do not have the same sign. We consider two cases.

If uv is negative, then vw is positive. It follows that there is a directed path P_2 from v to u of length congruent to 2 modulo 3. Now $P_3 = P_2 \cup P_1$ is a directed walk and since D_n is acyclic by Lemma 3.2.1, a directed path, from v to w. But P_3 has length congruent to 0 modulo 3, and so from the construction of D'_n, it follows that v and w are not adjacent in either direction, a contradiction.

Now uv is positive, and vw is negative. It follows that there is a directed path P_2 from w to v of length congruent to 2 modulo 3. Now $P_3 = P_1 \cup P_2$ is a directed walk and since D_n is acyclic by Lemma 3.2.1, a directed path from u to v. But P_3 has length congruent to 0 modulo 3, and so from the construction of D'_n, it follows that v and u are not adjacent in either direction, a contradiction. ∎

Now, we present our main theorem.

Theorem 3.2.4 *For every n, there is a graph G with $\chi(G) \geq n$ and $\omega(G) \leq 3$ such that every induced subgraph H of G with $\omega(H) \leq 2$ satisfies $\chi(H) \leq 4$.*

Proof. Let $\{G_n\}$ be the sequence of graphs such that G_n is the underlying undirected graph of D'_n. Then we claim that taking $G = G_n$ will show the desired result.

Indeed, we first show that $\chi(G_n) \geq n$. Since D_n is a subgraph of D'_n, it suffices to show, by induction, that the underlying undirected graph H_n of D_n has chromatic number at least n (which was also shown in [30] by Kierstead and Trotter, and follows from the fact that the n-th Zykov graph [42] is a subgraph of H_n; here we give the short proof for completeness). The base case is trivial. By induction, we know that the underlying undirected graphs H_{n-1} of the $n-1$ copies of D_{n-1} that were used to build D_n all have chromatic number at least $n-1$. So, if we take a coloring of H_n with colors $\{1, \ldots, n-1\}$, it follows that for every $i \in \{1, \ldots, n-1\}$, there exists a vertex $x_i \in V(D^i_{n-1})$ which receives color i. Then, letting $T = (x_1, \ldots, x_{n-1})$, the corresponding vertex v_T must receive a color not in $\{1, \ldots, n-1\}$, and it follows that the coloring uses at least n colors. Thus, $\chi(H_n) \geq n$ for all $n \geq 1$, as claimed.

Next, we claim that $\omega(G_n) \leq 3$. Suppose not; then G_n contains a clique K of size 4. Let $u \in K$ with at least two outneighbors in the digraph induced by K in D'_n (which is possible, since the average outdegree in this four-vertex digraph is 1.5), and let v, w be two outneighbours of u in K. By symmetry, we may assume that vw is an edge of D'_n. But now uv, vw, uw are all edges of D'_n, contrary to Lemma 3.2.3.

Now, suppose that we have an induced subgraph H of G_n with $\omega(H) \leq 2$. If we look at the corresponding induced subdigraph H' of D'_n, it follows by Lemma 3.2.2 that H' does not contain a directed 2-edge path with both edges of the same sign as a subdigraph. Thus we can partition the vertices into two sets A, B such that every vertex in A is not the head of a positive edge and every vertex in B is not the tail of a positive edge. Then note that there can be no positive edges between any two vertices in A, and also there are no positive edges between any two vertices in B. Likewise, we can find a similar partition $V(H') = A' \cup B'$ for the negative edges. Now $(A \cap A', A \cap B', B \cap A', B \cap B')$ is a partition of the vertices of H into four stable sets, and thus $\chi(H) \leq 4$, as claimed. This completes the proof. ∎

The collection of digraphs $\{D'_n\}$ also gives the following result on $\vec{\chi}$-boundedness.

Theorem 3.2.5 *For every n, there is a digraph D with $\vec{\chi}(D) \geq n$ and $\omega(D) \leq 3$ and with no induced directed cycle of odd length at least 5.*

Proof. We claim that taking $D = D'_{4n}$ gives the desired result. Indeed, we know from the previous proof that $\omega(D) \leq 3$. Furthermore, suppose that D contains an induced odd directed cycle of length at least 5. Then it follows that there exist two consecutive edges in that cycle of the same sign; but now Lemma 3.2.2 gives a third edge which contradicts the fact that the cycle is induced.

It remains to show that $\vec{\chi}(D) = \vec{\chi}(D'_{4n}) \geq n$. Indeed, note that any acyclic induced subdigraph H' of D has no directed path on two edges of the same sign, because then H' would contain a directed triangle, by Lemma 3.2.2, contrary to the assumption that H' is acyclic. Now, let H be the underlying undirected graph of H'. Then the argument from the previous proof shows that $\chi(H) \leq 4$. Since $\chi(G_{4n}) \geq 4n$ it follows that if $V(D)$ is partitioned into t sets which induce acyclic subdigraphs, then $\chi(G_{4n}) \leq 4t$; therefore, we must have $t \geq n$. Thus, $\vec{\chi}(D) \geq n$, as claimed. This completes the proof. ∎

3.3 t-chordal digraphs are not $\vec{\chi}$-bounded

Our construction uses a key idea similar to and inspired by the construction of Aboulker, Bousquet, and de Verclos in [1], making sure that in every k-dicoloring, certain independent sets miss a color, and arranging this for one independent set at a time. This is accomplished through the following lemma, from which Theorem 3.1.7 will be derived.

Lemma 3.3.1 *For $t \geq 3$, suppose that D is a t-chordal digraph with $\omega(D) \leq 3$ and $\vec{\chi}(D) = k$. If $\mathcal{C} = \{I_1, \ldots, I_p\}$ is a collection of independent sets in D, then there exists a t-chordal digraph D' with the following properties:*

- *$\omega(D') \leq 3$, and if $\omega(D) \leq 2$ and $t > 3$, then $\omega(D') \leq 2$; and*
- *for every k-coloring of D', there exists a copy of D as an induced subgraph of D' such that for each $1 \leq i \leq p$, the copy of I_i contained in that copy of D is colored with at most $k-1$ colors.*

Proof. We will prove the lemma by induction on the chromatic number of an auxiliary graph that we define now. For a digraph D and a collection $\mathcal{C}$ of independent sets, let $G_{D,\mathcal{C}}$ be the graph with $V(G_{D,\mathcal{C}}) = \{1, 2, \ldots, p\}$, and where $ij \in E(G)$ if and only if $|I_i \cap I_j| \neq \emptyset$; in other words, $G_{D,\mathcal{C}}$ is the intersection graph of $\mathcal{C}$.

The base case is when $\chi(G_{D,\mathcal{C}}) = 0$, (and thus, $\mathcal{C} = \emptyset$) where the statement is trivially true. Now, for $s \geq 0$, suppose that the statement is true for all digraphs D paired with a collection of independent sets $\mathcal{C}$ such that the corresponding graph $G_{D,\mathcal{C}}$ has $\chi(G_{D,\mathcal{C}}) \leq s$, and suppose that $\chi(G_{D,\mathcal{C}}) = s+1$. Take an $(s+1)$-coloring of $G_{D,\mathcal{C}}$, say f. Let $X_0 = \{j_1, \ldots, j_q\}$ be a color class of f, and let $S_0 = \{I_{j_1}, \ldots, I_{j_q}\}$. It follows that $\chi(G_{D,\mathcal{C}\setminus S_0}) = s$, and that the sets $I_{j_1}, \ldots, I_{j_q}$ are pairwise disjoint.

We define a sequence $D_0, \ldots, D_q$ with the following properties:

- $D_0 = D$;
- for all $i \in \{0, \ldots, q\}$, D_i is a digraph with clique number at most 3 (at most 2 if $\omega(D) \leq 2$ and $t > 3$);
- D_i contains pairwise disjoint copies $G_i^1, \ldots, G_i^{t^i}$ of D;
- in every k-dicoloring of D_i, there is a $t^* \in \{1, \ldots, t^i\}$ such that for every $r \in \{1, \ldots, i\}$, the independent set corresponding to I_{j_r} in the copy $G_i^{t^*}$ of D is colored with at most $k-1$ colors; and
- for every $r \in \{i+1, \ldots, q\}$, the union of the independent sets corresponding to I_{j_r} in $G_i^1, \ldots, G_i^{t^i}$ is an independent set in D_i.

Clearly, D_0 satisfies the properties above for $i = 0$. Suppose that we have defined D_i for some $i \in \{0, \ldots, q-1\}$. Let I be the union of the sets corresponding to $I_{j_{i+1}}$ in $G_i^1, \ldots, G_i^{t^i}$. From the properties of D_i, it follows that I is an independent set. We create a new digraph D_{i+1} as follows. Let $D_i^1, \ldots, D_i^t$ be t copies of D_i, and let $V(D_{i+1}) = \bigcup_{j=1}^{t} V(D_i^j)$. In addition to the edges corresponding to each copy D_i^j of D_i, we add the following edges. For $j \in \{1, \ldots, t\}$, let I^j denote the copy of I in D_i^j. Then, for each vertex $v \in I^j$ and $u \in I^{j+1}$, we add the edge vu (where indices are taken modulo t, so I^{t+1} means I^1).

Observe that the clique number of D_{i+1} is at most 3, and at most 2 if $t > 3$ and $\omega(D) \leq 2$. This follows since a clique in D_{i+1} either has all its vertices in D_i^j for some j, or it has at most one vertex contained in each copy of D_i^j. Similarly, D_{i+1} is t-chordal, as every induced cycle either uses precisely one vertex from each D_i^j, or it is contained completely in a copy D_i^j for some j.

Since D_{i+1} contains t copies of D_i, each copy of D in D_i gives rise to t copies of D in D_{i+1}, for $t \cdot t^i = t^{i+1}$ copies overall. Let us label them as $G_{i+1}^1, \dots, G_{i+1}^{t^{i+1}}$ arbitrarily. It follows that the last bullet holds for D_{i+1}, since it holds for D_i, and since the only edges between copies of D_i in D_{i+1} are between vertices that are in a set corresponding to $I_{j_{i+1}}$ in a copy of D, and $I_{j_{i+2}}, \dots, I_{j_q}$ are disjoint from $I_{j_{i+1}}$ from the choice of S_0.

Now, let c be a k-dicoloring of D_{i+1} (if one exists). Then, not all k colors occur in each of $I^1, \dots, I^t$, for otherwise there would be a monochromatic t-vertex cycle. It follows that there is a $j \in \{1, \dots, t\}$ such that the copy of I in D_i^j is colored with at most $k-1$ colors. Consequently, for each of the copies of $G_i^1, \dots, G_i^{t^i}$ of D in D_i^j, the copy of I_{j_i} in this copy of D is colored with at most $k-1$ colors. Together with the fact that the third bullet holds for D_i, by applying it to D_i^j, it follows that it holds for D_{i+1}.

This completes the definition of $D_0, \dots, D_q$. Note that D_q has the property that in every k-dicoloring of D_q, there is a copy of D in D_q such that each of $I_{j_1}, \dots, I_{j_q}$ is colored with at most $k-1$ colors.

Now we construct a collection of independent sets $\mathcal{C}'$. For every independent set $I \in \mathcal{C} \setminus S_0$, and for every $t^* \in \{1, \dots, t^q\}$, we add the copy of I in $G_q^{t^*}$ to $\mathcal{C}'$. Note that we can s-color $G_{D_q,\mathcal{C}'}$ by fixing an s-coloring of $G_{D,\mathcal{C}\setminus S_0}$, and assigning to each independent set in a copy of D the color of the independent set in D it corresponds to. This, by construction, is an s-coloring, since two sets which are assigned the same color are either copies of disjoint sets, or disjoint copies of the same set.

It follows that we can apply the inductive hypothesis to D_q and $\mathcal{C}'$; and so there is a digraph D^* with $\omega(D^*) \leq 3$ (and if $\omega(D) \leq 2$ and $t > 3$ then $\omega(D^*) \leq 2$), and for every k-coloring of D^*, there is a copy D_q' of D_q in D^* such that for every $I \in \mathcal{C}'$, the copy of I in D_q' is colored with at most $k-1$ colors. It follows that in D_q', there are copies $G_q^1, \dots, G_q^{t^q}$ of D, and for every $I \in \mathcal{C} \setminus S_0$, every copy of I in every one of $G_q^1, \dots, G_q^{t^q}$ receives at most $k-1$ colors. But also, from the properties of D_q, there is a $t^* \in \{1, \dots, t^q\}$ such that for $G_q^{t^*}$ in D_q', for each $I \in S_0$, the copy of I in $G_q^{t^*}$ receives at most $k-1$ colors. This shows that in this copy $G_q^{t^*}$ in D_q' of D, the second bullet of the lemma holds. This concludes the proof. ∎

Now, we present our main result, which is proved in the same way as the main result of Aboulker, Bousquet, and de Verclos in [1] (in which the main theorem is derived from their Lemma 3 exactly like we derive ours from Lemma 3.3.1 above):

Theorem 3.3.2 *For all $t \geq 3$, there exists a sequence of t-chordal digraphs $\{D_n\}$ such that for all $n \geq 1$, we have $\omega(D_n) \leq 3$ (and $\omega(D_n) \leq 2$ if $t > 3$) and $\vec{\chi}(D_n) \geq n$.*

Proof. Let D_1 be the digraph with one vertex and no edges, and define the sequence of digraphs inductively as follows. For $k \geq 1$, take $\vec{\chi}(D_k)$ disjoint copies of D_k, forming a digraph D'_k, and construct a collection $\mathcal{C}$ of independent sets of D'_k as follows: for each set of $\vec{\chi}(D_k)$ vertices, one in each copy of D_k in D'_k, place this set in $\mathcal{C}$. Then by Lemma 3.3.1, we have that there exists a t-chordal digraph D_{k+1} with $\omega(D_{k+1}) \leq 3$ (and $\omega(D_n) \leq 2$ if $t > 3$) such that for any $\vec{\chi}(D_k)$-dicoloring of D_{k+1} there exists a copy of D'_k such that each independent set in $\mathcal{C}$ uses at most $\vec{\chi}(D_k) - 1$ colors. Now, for a $\vec{\chi}(D_k)$-coloring of D'_k, it follows that in each of the $\vec{\chi}(D_k)$ copies of D_k in D'_k there is a vertex of each color, so there exists an independent set in $\mathcal{C}$ with exactly one vertex of each color, which is a contradiction. It follows that there does not exist a $\vec{\chi}(D_k)$-coloring of D_{k+1} and thus $\vec{\chi}(D_{k+1}) \geq \vec{\chi}(D_k) + 1$, and therefore $\vec{\chi}(D_k) \geq k$ for all $k \geq 1$, as desired. This completes the inductive step and completes the proof. ∎

We now show the following positive result which shows that the above construction is in some sense tight:

Theorem 3.3.3 *For every l, there is a function $f_l : \mathbb{N} \to \mathbb{N}$ such that for every digraph D with no induced directed cycle of length less than l and no induced directed path with exactly l vertices, we have $\vec{\chi}(D) \leq f_l(\omega(D))$.*

Proof. Let us denote by $\mathcal{C}_l$ the class of digraphs with no induced directed cycle of length less than l and no induced directed path with exactly l vertices.

We will show that $f_l(\omega) = (l+1)^\omega$ gives the desired result. We proceed by induction on $k = \omega(D)$. For $k = 1$, the statement is true, since in that case $\vec{\chi}(D) = 1 \leq l + 1 = (l+1)^\omega$.

Now, suppose that the statement is true for $k = n$, namely that for all digraphs $D' \in \mathcal{C}_l$ with $\omega(D') = n$, we have $\vec{\chi}(D') \leq (l+1)^n$. Now, consider a digraph $D \in \mathcal{C}_l$ with $\omega(D) = n + 1$. We will show that $\vec{\chi}(D) \leq (l+1)^{n+1}$. Suppose for the sake of a contradiction that $\vec{\chi}(D) > (l+1)^{n+1}$. First, note that we may assume that D is strongly connected, since the dichromatic number of a digraph is equal to the maximum of the dichromatic numbers of its strongly connected components. Let $v_1 \in D$ be an arbitrary vertex, and let $C_1, \ldots, C_t$ be the strongly connected components of $D \setminus (\{v_1\} \cup N(v_1))$, where $N(v_1)$ denotes the set of vertices which have an in-edge or an out-edge to v_1. We claim that there exists a component C_i with $\vec{\chi}(C_i) > (l+1)^{n+1} - (l+1)^n$. Indeed, since $\omega(N(v_1)) \leq n$, we have by induction that $\vec{\chi}(N(v_1)) \leq (l+1)^n$. So if $\vec{\chi}(C_i) \leq (l+1)^{n+1} - (l+1)^n$ for all $1 \leq i \leq t$, we can color each component C_i with the same set of $(l+1)^{n+1} - (l+1)^n$ colors, color $N(v_1)$ with a disjoint set of $(l+1)^n$ colors, and reuse one of the colors from the C_i to color v_1. Clearly, this is a coloring of D with at most $(l+1)^{n+1}$ colors, which is a contradiction. Thus we may assume without loss of generality that $\vec{\chi}(C_1) > (l+1)^{n+1} - (l+1)^n$.

Now, since D is strongly connected, there exists a directed path from v_1 to a vertex in C_1. Let P' be the shortest such path, and let the second-to-last vertex in P' be v_2. Now, we let P be the portion of the path P' from v_1 to v_2. First, we note that P is forward-induced, meaning that there are no edges in the direction of the path other than those in P. Furthermore, since P' was the shortest directed path from v_1 to C_1, it follows that v_2 is the only vertex in P which has an out-edge to a vertex in C_1.

Let $D_1 = D$. We define a sequence of digraphs as follows. Let $D_2 = \{v_2\} \cup C_1$, and note that since C_1 is strongly connected and v_2 has an out-edge to C_1, it follows that there exists a directed path from v_2 to every vertex in D_2. Then we iterate this argument, letting C_2 be a strongly connected component of $D_2 \setminus (\{v_2\} \cup N(v_2))$ with $\vec{\chi}(C_1') > (l+1)^{n+1} - 2(l+1)^n$, taking a shortest directed path Q from v_2 to C_2, and adding to our path P the subpath of Q from v_2 to the second-to-last vertex of Q, which we call v_3. Since $(l+1)^{n+1} - (l+1)(l+1)^n \geq 0$, we can iterate this argument $l+1$ times and obtain a forward-induced path P with a subset of the vertices of P specified as $v_1, \ldots, v_{l+2}$. In particular, the l vertices $v_3, \ldots, v_{l+2}$ are not in $\{v_1\} \cup N(v_1)$, since all vertices of P after v_2 are in C_1, and thus not in $\{v_1\} \cup N(v_1)$.

Now, let w be the last vertex in P which is in $N(v_1)$. Then it follows from the properties of P that there exists a forward-induced path P' with l vertices starting at w such that w is the only vertex in P' which is in $\{v_1\} \cup N(v_1)$. Since P' is a forward-induced path on l vertices and there is no induced directed path on l vertices and no induced directed cycle of length less than l, it follows that P' is an induced directed cycle of length l. Let the vertices of P' be $(w_1, \ldots, w_l)$ such that $w_1 = w$. Now, if w_1 is an out-neighbor of v_1, then we have that $(v_1 w_1 w_2, \ldots, w_{l-1})$ is an induced directed path on l vertices, which is a contradiction. Suppose instead that w_1 is an in-neighbor of v_1. Then we have that $(w_3 w_4, \ldots, w_l w_1 v_1)$ is an induced directed path on l vertices, which is a contradiction as well. Hence, $\vec{\chi}(D) \leq (l+1)^n$. This finishes the proof by induction and shows that $\vec{\chi}(D) \leq (l+1)^{\omega(D)}$ for all digraphs $D \in \mathcal{C}_l$, as desired. ∎

3.4 Deciding if a graph is t-chordal is coNP-complete

In this section, we prove Theorem 3.1.10. The construction is very similar to constructions of Kawarabayashi and Kobayashi [29] and Bienstock [8], where it was shown that finding certain induced paths and cycles in graphs and digraphs is NP-hard. Note that Theorem 3.4.1 can be modified in a straightforward way to show that the problem of determining if a non-simple digraph (where both uv and vu are allowed to be present at the same time) is 2-chordal is coNP-complete as well, whereas we can decide in polynomial time if a simple digraph is 2-chordal by checking if it is acyclic.

Theorem 3.4.1 *For $t \geq 3$, the problem of determining whether a digraph D is t-chordal is coNP-complete.*

Proof. Observe the problem is in coNP. To see this, given any digraph D, a certificate that it is not t-chordal is an induced directed cycle C which has length not equal to t. Further, such a certificate can be checked in polynomial time.

For the rest of the proof, let $t \geq 2$ be a fixed integer. We will give a reduction from 3-SAT, a well-known NP-complete problem. Let ϕ be an instance of 3-SAT, with variables $x_1, \dots, x_n$ and clauses $C_1, \dots, C_m$. We will construct a digraph D such that D is t-chordal if and only if ϕ is a NO-instance of 3-SAT, which suffices to prove the result.

We build D in stages. First, for each variable x_i $i \in \{1, \dots, n\}$, we create a variable gadget, D_{x_i}, which is a digraph with two vertices v_1^i, v_2^i and two directed paths P_1^i, P_2^i, both directed from v_1^i to v_2^i, which are vertex-disjoint aside from the vertices v_j^i, $j \in \{1, 2\}$ and both contain exactly t edges. For each $j \in \{1, 2\}$, let z_j^i be the vertex in P_j^i adjacent from v_1^i and q_j^i the vertex adjacent to v_2^i in P_j^i.

Second, for each clause C_i, $i \in \{1, \dots, m\}$, we create a clause gadget, D_{C_i} which has two vertices u_1^i, u_2^i, and for each literal y contained in C_i, we add a directed path u_1^i, w^y, u_2^i.

Now we create a digraph D whose vertex set is

$$\{V(D_{x_i}) : i \in \{1, \dots, n\}\} \cup \{V(D_{C_i}) : i \in \{1, \dots, m\}\}.$$

For the edges, for $i \in \{1, \dots, n-1\}$, we add the edge $v_2^i v_1^{i+1}$, as well as the edge $v_2^n u_1^1$. For $j \in \{1, \dots, m-1\}$ we add the edges $u_2^i u_1^{i+1}$, as well as the edge $u_2^m v_1^1$. For a clause C_i, if the non-negated variable x_i appears in the clause, then we add edges $w^{x_i} z_2^i$ and $q_2^i w^{x_i}$. For a clause C_i if the negated variable x_i appears in the clause, then we add edges $w^{x_i} z_1^i$ and $q_2^i w^{x_i}$.

Now we claim that D is t-chordal if and only if the 3-SAT instance has no solution. Suppose there is an induced directed cycle C of length other than t. Suppose first that C uses an edge of the form $w^{x_i} z_j^i$ or $q_j^i w^{x_i}$ for some $j \in \{1, 2\}$. If C uses $w^{x_i} z_j^i$, then, since z_j^i has a unique out-neighbor, it follows that C contains the path $P_j^i \setminus v_2^i$. In particular, C contains q_j^i; and since C contains w^{x_i}, and C is induced, it follows that C contains $q_j^i w^{x_i}$, and so C has length t. If $q_j^i w^{x_i}$, then, since q_j^i has a unique in-neighbor, it follows similarly that C has length t. Now, by construction, it follows C has non empty-intersection with every variable gadget and every clause gadget.

For each variable x_i, if C takes the path P_1^i, we set x_i to true, and otherwise we set x_i to false. We claim this gives a satisfying assignment to the 3-SAT instance. If not, then some clause evaluates to false, say C_1 without loss of generality. We may assume up to symmetry that $C \cap D_{C_1}$ is the path associated to the variable x_i. If this variable is not negated in C then this implies that C must use the P_2^i path to set x_i to false, however in this case there are edges from vertices in P_2^i to w^{x_i} which are not apart of C, contradicting that C is induced. An analogous argument holds when x_i is negated in C_1.

Similarly for the converse, if an assignment to $x_1, \dots, x_m$ satisfies the formula, we construct the induced directed cycle by adding either P_1^i if x_i is true, or P_2^i if x_i is false, taking any path from each clause gadget which corresponds to a true literal in that clause, and adding the edges between clause gadgets and variable gadgets in the natural way. It is easy to check that this cycle is an induced directed cycle of length not equal to t, completing the proof. ∎

3.5 Conclusions and further work

Here we discuss some results that have been published relating to the work of this chapter, and areas of future work.

We let a hereditary class C of graphs be *polynomially χ-bounded* if there exists a polynomial f such that for all graphs $G \in C$ we have $\chi(G) \leq f(\omega(G))$. Shortly after our preprint showing Theorem 3.1.4 was published on the ArXiv, Briański, Davies, and Walczak [10] built upon the ideas of our paper to disprove the following conjecture of Esperet.

Conjecture 3.5.1 (Esperet [19]) *If a hereditary class C of graphs is χ-bounded then it is polynomially χ-bounded.*

In fact, they showed the stronger result that for any function f, there exists a hereditary class of graphs which is χ-bounded but does not have f as a χ-bounding function.

Shortly after that paper, Girão, Illingworth, Powierski, Savery, Scott, Tamitegama, and Tan [20] published a paper showing the following, generalizing both Theorem 3.1.4 and the results of Briański, Davies, and Walczak [10]:

Theorem 3.5.2 ([20]) *For every graph F with at least one edge, there is a constant c_F such that there are graphs of arbitrarily large chromatic number and the same clique number as F in which every F-free induced subgraph has chromatic number at most c_F.*

One direction of further work would be building upon these works (there are a number of open problems in the paper of Girão, Illingworth, Powierski, Savery, Scott, Tamitegama, and Tan [20]). In a somewhat different direction, Conjecture 3.1.9 remains very open, even for the case of directed paths. Showing this case, or disproving the conjecture, would both be quite interesting.

In general, it seems that the digraph analogues of χ-boundedness are much less true, as the development of counterexamples has shown some very fruitful results so far. So one can further explore constructions, such as the one we showed for t-chordal graphs, or one could instead make strides towards positive results in this setting, which likely require the development of new techniques which would be interesting.

References

[1] P. Aboulker, N. Bousquet, and R. de Verclos, "Chordal directed graphs are not χ-bounded," ArXiv:2202.01006, 2022.

[2] P. Aboulker, P. Charbit, and R. Naserasr, "Extension of Gyárfás-Sumner conjecture to digraphs," ArXiv:2009.13319, 2020.

[3] R. Aharoni, E. Berger, M. Chudnovsky, and S. Zerbib, "Two Remarks on the Caccetta-Häggkvist Conjecture", ArXiv:2110.11183, 2021.

[4] R. Aharoni, M. DeVos, and R. Holzman, "Rainbow triangles and the Caccetta-Häggkvist conjecture", *J. Graph Theory*, 92(4):347-360, 2019.

[5] R. Aharoni and H. Guo, "Rainbow cycles for families of matchings," ArXiv:2110.14332, 2021.

[6] K. Appel, W. Haken, "Every Planar Map is Four-Colorable," *Contemporary Mathematics*, vol. 98 (1989).

[7] E. Berger, K. Choromanski, M. Chudnovsky, J. Fox, M. Loebl, A. Scott, P. Seymour, and S. Thomassé, "Tournaments and colouring," *Journal of Combinatorial Theory, Series B*, 103(1):1-20, 2013.

[8] D. Bienstock, "On the complexity of testing for odd holes and induced odd paths," *Discrete Mathematics*, 90(1):85-92, 1991.

[9] B. Bollobás, E. Szemerédi, "Girth of sparse graphs", *J. Graph Theory*, 39(3):194-200, 2002.

[10] M. Briański, J. Davies, and B. Walczak, "Separating polynomial χ-boundedness from χ-boundedness," ArXiv:2201.08814, 2022.

[11] L. Caccetta and R. Häggkvist, "On minimal digraphs with given girth", *Congr. Numer.*, 21:181-187, 1978.

[12] A. Carbonero, P. Hompe, B. Moore, and S. Spirkl, "A counterexample to a conjecture about triangle-free induced subgraphs of graphs with large chromatic number," ArXiv:2201.08204, 2022.

[13] A. Carbonero, P. Hompe, B. Moore, and S. Spirkl, "Digraphs with all induced directed cycles of the same length are not $\vec{\chi}$-bounded," ArXiv:2203.15575, 2022.

[14] V. Chvátal and E. Szemerédi, "Short cycles in directed graphs," *Journal of Combinatorial Theory, Series B*, 35(3):323-327, 1983.

[15] R. Diestel, "Graph Theory (Graduate Texts in Mathematics)," *Springer*, 2005.

[16] M. DeVos, M. Drescher, D. Funk, S. de la Maza, K. Guo, T. Huynh, B. Mohar, and A. Montejano, "Short rainbow cycles in graphs and matroids," *J. Graph Theory*, 1–11, 2020.

[17] P. Erdős, "Problems and results in number theory and graph theory," *Proc. Ninth Manitoba Conference on Numerical Math. and Computing*, 3-21, 1979.

[18] P. Erdős, A.H. Stone, "On the structure of linear graphs," *Bulletin of the American Mathematical Society*, 52(12): 1087-1091, 1946.

[19] L. Esperet, "Graph colorings, flows and perfect matchings," Habilitation thesis, Universite Grenoble, Alpes, 2017.

[20] A. Girão, F. Illingworth, E. Powierski, M. Savery, A. Scott, Y. Tamitegama, and J. Tan, "Induced subgraphs of induced subgraphs of large chromatic number," ArXiv:2203.03612, 2022.

[21] A.W. Goodman, "On sets of acquaintances and strangers at any party," *Amer. Math. Monthly* 66:778-783, 1959.

[22] A. Gyárfás, "On Ramsey covering-numbers," *Infinite and Finite Sets*, 2:801-816, 1975.

[23] A. Gyárfás, "Problems from the world surrounding perfect graphs," *Applicationes Mathematicae*, 19(3-4):413-441, 1987.

[24] J. Hladký, D. Král, and S. Norin, "Counting flags in triangle-free digraphs," *Combinatorica*, 37:49-76, 2017.

[25] P. Hompe, Senior Thesis, Princeton University, 2019.

[26] P. Hompe, P. Pelikánová, A. Pokorná, and S. Spirkl, "On Aharoni's rainbow generalization of the Caccetta-Häggkvist conjecture," *Discrete Mathematics*, 344(5): 2021.

[27] P. Hompe, Z. Qu, and S. Spirkl, "Improved bounds for the triangle case of Aharoni's rainbow generalization of the Caccetta-Haggkvist conjecture," ArXiv.2206.10733, 2022.

[28] P. Hompe and S. Spirkl, "Further approximations for Aharoni's rainbow generalization of the Caccetta-Häggkvist conjecture," *Electronic Journal of Combinatorics*, 29(1), 2022.

[29] K. Kawarabayashi and Y. Kobayashi, "The induced disjoint paths problem," *International Conference on Integer Programming and Combinatorial Optimization*, 47-61, 2008.

[30] H.A. Kierstead, and W.T. Trotter, "Colorful induced subgraphs," *Discrete Mathematics*, 101(1-3):165-169, 1992.

[31] M. Mitzenmacher, E. Upfal, "Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis," *Cambridge University Press*, 2017.

[32] V. Neumann-Lara, "The dichromatic number of a digraph," *Journal of Combinatorial Theory, Series B*, 33(3):265-270, 1982.

[33] A.A. Razborov, "Flag algebras," *J. Symbolic Logic*, 72(4):1239-1282, 2007.

[34] V. Rödl, "On the chromatic number of subgraphs of a given graph," *Proceedings of the American Mathematical Society*, 64(2):370-371, 1977.

[35] A. Scott and P. Seymour, "A survey of χ-boundedness," *Journal of Graph Theory*, 95(3):473-504, 2020.

[36] A. Scott and P. Seymour, "Induced subgraphs of graphs with large chromatic number. I. Odd holes," *Journal of Combinatorial Theory, Series B*, 121:68-84, 2016.

[37] P. Seymour and S. Spirkl, "Short directed cycles in bipartite digraphs," *Combinatorica*, 40(4):575-599, 2020.

[38] J. Shen, "On the Caccetta-Häggkvist conjecture," *Graphs and Combinatorics*, 18(3): 645-654, 2002.

[39] D.P. Sumner, "Subtrees of a graph and chromatic number," in The Theory and Applications of Graphs, (G. Chartrand, ed.), John Wiley & Sons, New York, 557–576, 1981.

[40] S. Thomassé, N. Trotignon, and K. Vušković, "A polynomial Turing-kernel for weighted independent set in bull-free graphs," *Algorithmica*, 77(3):619-641, 2017.

[41] P. Turán, "On an extremal problem in graph theory," *Matematikai és Fizikai Lapok* 48: 436–452, 1941.

[42] A.A. Zykov, "On some properties of linear complexes," *Matematicheskii sbornik*, 66(2):163-188, 1949.