SOME PROBLEMS IN EXTREMAL GRAPH THEORY AVOIDING THE USE OF THE REGULARITY LEMMA

by

IAN MARC LEVITT

A dissertation submitted to

The Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Endre Szemerédi

and approved by

New Brunswick, New Jersey

May, 2009

ABSTRACT OF THE DISSERTATION

Some problems in Extremal Graph Theory avoiding the use of the Regularity Lemma

by Ian Marc Levitt

Dissertation Director: Endre Szemerédi

In this thesis we present two results in Extremal Graph Theory. The first result is a new proof of a conjecture of Bollobás on embedding trees of bounded degree. The second result is a new proof of the Pósa conjecture.

Let $G = (W, E)$ be a graph on n vertices having minimum degree $\delta(G) \geq n/2 + c \log n$, where c is a constant. Béla Bollobás conjectured that every tree on n vertices with bounded degree can be embedded into G. We show that this conjecture is true. In fact we show more, that unless G is very close to either the union of two complete graphs on $n/2$ vertices, or the complement, then a minimum degree of $n/2$ is sufficient to embed any tree of bounded degree.

The k^{th} *power* of C is the graph obtained from C by joining every pair of vertices at a distance at most k in C. In 1962 Pósa conjectured that any graph G of order n and minimum degree at least $\frac{2}{3}n$ contains the square of a Hamiltonian cycle. The conjecture was proven for $n > n_0$ by Komlós, Sárközy and Szemerédi in [17] using the Regularity Lemma and Blow-up Lemma. The new proof removes the use of the Regularity Lemma and establishes the Pósa conjecture using combinatorial arguments, thus vastly reducing n_0.

Acknowledgements

I would like first to acknowledge my advisor, Endre Szemerédi, without whose patience and gentle guidance this would not have been possible. The time that he has shared with me, given generously, has shown me what mathematics truly is. He has welcomed me into his home, and into his work, and for this I am grateful. I would also like to thank Endre's wife, Anna, whose graciousness and words of encouragement may be as responsible for the success of my studies as anything.

I thank my beautiful wife, Elizabeth, for standing by me and giving me the time and encouragement to perform this work. I would like to recognize the sacrifice that she has made, and I hope that I have the chance to repay it in kind. I would also like to thank my mother, Molly, whose capacity for her children and the ones that she loves has no bounds. Without my wife and my mother this thesis could not have been written.

I would like to extend my gratitude to Stuart Searight, who made sure that I had the freedom at work to complete my studies. I would also like to thank all of my colleagues who supported me in completing my studies.

I would like to recognize all of those people who made such a difference in my life and my family's lives during our time at Rutgers: Lolly Schenk, who is the village; Mary and Lillie and Tracy and all of the good people at Pine Grove who have helped so many little ones; Allesandro and Michela Silva who we miss dearly; Benji and Sayaka, Geta and Gehnadi, and all of the other parents and their children that gave us a reason to come outside.

Finally I would like to thank my children, Tallulah and Moses, for still shining brightly even when their daddy couldn't play.

Dedication

For Lui

Table of Contents

Abstract ii

Acknowledgements iii

Dedication iv

1. Introduction 1

1.1. A New Proof of the Bollóbas Conjecture on Embedding Trees 1

1.1.1. G is not Extremal 2

T has Many Leaves 2

T has Few Leaves 3

1.1.2. G is Extremal 3

G is Close to $K_{n/2} \cup K_{n/2}$ 3

G is Close to $K_{n/2,n/2}$ 4

1.2. A New Proof of the Pósa Conjecture 5

2. Proof of the Bollobás Conjecture on Embedding Trees 8

2.1. Introduction 8

2.2. Definitions and Notation 8

2.3. The non-extremal case 9

2.3.1. Proof sketch 9

2.3.2. Some tools for the proof 9

2.3.3. Preparations 13

2.3.4. Decomposition of T_* 15

2.3.5. Finishing the embedding of T 16

2.3.6. The Main Mapping Procedure . . . 16
2.3.7. The Second Mapping Method . . . 17
2.3.8. The Third Mapping Method . . . 19
2.3.9. Description of the Embedding Algorithm . . . 21
2.3.10. T has at most $\gamma^2 n$ leaves . . . 22
2.4. The Extremal Case . . . 26
2.4.1. G is close to $K_{\lceil n/2 \rceil} \cup K_{\lfloor n/2 \rfloor}$ . . . 28
2.4.2. G is close to $K_{n/2,n/2}$ . . . 33

3. How to avoid using the Regularity Lemma; Pósa's Conjecture revisited 35
3.1. Notations and definitions . . . 35
3.2. Outline of the proof . . . 35
3.3. Connecting . . . 37
3.4. Absorbing . . . 42
3.5. The reservoir . . . 44
3.6. The proof of Theorem 1 . . . 45
3.6.1. T is larger than $\alpha^8 n$ . . . 46
3.6.2. T is smaller than $\alpha^8 n$ . . . 49
3.7. Conclusion . . . 56
References . . . 57
Vita . . . 59

Chapter 1

Introduction

1.1 A New Proof of the Bollóbas Conjecture on Embedding Trees

An easy folklore result is that any graph G with minimum degree d contains every tree on d edges. This result is tight in the sense that K_{d+1} does not contain any tree on more edges. The well-known Erdős-Sós Conjecture weakens the statement: any graph G of *average* degree d contains all trees with d edges. This has proven to be far more difficult. In both these cases G is allowed to be arbitrarily large. The related problem, where G and the tree to be embedded have the same size, was conjectured by Bollóbas:

Conjecture 1 *For any $\delta > 0$ there exists a constant c such that if T is a tree of order n and maximum degree at most $cn/\log n$, and G is a graph of order n and minimum degree at least $(1/2 + \delta)n$, then T is a subgraph of G.*

In [22], Komlós, Sárközy and Szemerédi proved this conjecture for $n > n_0$ with the use of the Regularity Lemma. Here we prove the following

Theorem 1 *For any K there exists a c and a n_0 such that if G is a graph of order $n > n_0$ with minimum degree $n/2 + c\log n$ and T is a tree of order n with minimum degree K, then $T \subset G$.*

Our proof establishes embedding procedures that rely on elementary arguments, avoiding the use of the Regularity Lemma. The n_0 so obtained is therefore much smaller. In fact the proof presented in this thesis implies a stronger result, that unless G or its complement is very close to either the union of two complete graphs on $n/2$ vertices each then a minimum degree of $n/2$ is sufficient to embed every tree of bounded degree.

The proof of Theorem 1 is divided into cases. If G is not an extremal graph then we consider separately when T has many leaves and when T has few leaves. Then we consider the two extremal cases. In every case there is the general theme that we reserve a constant proportion of the vertices of T to be embedded last. This reduces the problem to embedding a smaller tree T' into G where the difference in their orders provides some freedom. In each case, before the embedding we identify random subsets in G that we use both for the embedding of T' and for embedding the remainder of T by a matching argument.

1.1.1 G is not Extremal

In both subcases, we need only that the minimum degree in G is $n/2$.

T has Many Leaves

In this case we first identify a small but constant-sized subtree T_0 which contains many leaves and identify a subset S of constant proportion of leaves of T_0. We then take 3 random subsets D_1, D_2, and M with $|D_1| = |D_2| = |S|$. The plan is to embed the vertices of S into D_2 at the end. We use D_2 for the parents of the vertices of S which are matched to D_1 in the final step. The random subset M plays an important role throughout the embedding process. We begin by embedding $T_0 - S$ greedily, but making sure that the parents of the vertices of S end up in D_2.

The next step is to decompose $T - T_0$ into subtrees that are exponentially large in K. We embed these subtrees breadth-first, establishing their inter-connections via the random set M. Let Q be the set of vertices of G that are not yet covered. When we connect through M, we use only polynomially in K many vertices of M, thus embedding into exponentially in K many vertices of Q. This ensures that the vertices of M are never exhausted. Here again is a common theme throughout the proof, of covering certain vertices of G at a proper rate.

If we get to a point where we can no longer greedily map any subtrees of the decomposition because Q is sparse, then we find a subset of Q which is densely connected to subtrees that have already been embedded. We rearrange the mapping in such a way that we return a dense subset to Q and can continue mapping the subtrees. Here again M plays an important role. When we rearrange the mapping of an already-embedded subtree, we have to ensure that it is properly connected to the rest of the embedding.

Finally, when Q is small enough, we greedily map the rest of $T - T_0$ into D_2. While the vertices of D_2 were reserved for S, this sacrifice is small enough so that by mapping a portion of S to Q we can still find the desired matching.

T has Few Leaves

In this case we have an easier time. We define a *stem* of a tree T to be a path P such that every vertex on P apart from possibly the endpoints has degree 2 in T. When T has only a few leaves it has many stems. We find many (a constant proportion of n) disjoint long stems in T and define a new tree T' obtained by contracting the middle edge of each stem. As T' is smaller than G it is easy to embed. We begin by embedding the contracted stems in such a way that most of their edges are randomly selected. At the end, the remaining $|G| - |T'|$ vertices of G are easily matched to the embedded stems, where a vertex is matched to a stem only if it contains an edge of the stem in its neighborhood in G. Thus we can insert the vertices into their assigned stems, which finishes the embedding. The randomness of the embedding of the contracted stems is used for the matching.

1.1.2 G is Extremal

G is Close to $K_{n/2} \cup K_{n/2}$

Note that here we need the full degree condition from the conjecture since $K_{n/2} \cup K_{n/2}$ is not even connected. An extremal G is one for which there is a bipartition of its vertex set into two classes, A and B, such that $|A| = |B|$ and the induced graph on each part has $n^2/4 - \gamma n$ edges. The basic plan in this case is to find a subset $S \subset T$ of size $O(\log n)$ so that the components $T - S$ can be grouped in such a way that the number of vertices in each group is $|A| - |S|$ and $|B|$. We will map S into A and the components of $T - S$ into their assigned parts. It is important that $S = O(\log n)$ as we may need as many as KS

edges going from A to B in order to connect the components. To acheive this we use the following folklore result

Lemma 2 *Let J be any tree on m vertices. Then J has a vertex $u \in V(J)$ such that it is possible to group the vertices of $J - u$ into two forests, J_1 and J_2 such that $m/3 \leq |J_1|, |J_2| \leq 2m/3$ and there is no edge connecting J_1 and J_2 in $J - u$.*

We first divide T into many subtrees of roughly equal size by splitting a constant number of times, then recursively split trees to reduce the error in packing the vertices by a constant factor.

Before splitting T, we first identify two large subtrees T_X and T_Y of size around $n/100$ each, and two smaller subtrees, $T_1 \subset T_X$ and $T_2 \subset T_Y$. We embed $T_X - T_1$ into A and $T_Y - T_2$ into B by a greedy algorithm which covers one minimum degree vertex of each class for every $O(K)$ vertices mapped. Thus, when we are finished with embedding these trees the minimum degree in its class of every uncovered vertex is very close to $n/2$. As we map T_1 and T_2 last, the uncovered vertices always induce a subgraph of large degree, and so mapping the components in the decomposition is always possible. Again, for the connections between components and for mapping T_1 and T_2 at the end we use random subsets that were set aside at the beginning of the procedure. We map T_1 and T_2 in the final step by a matching argument.

G is Close to $K_{n/2,n/2}$

This case is similar to the previous extremal case in that we begin by mapping a large subtree via a greedy algorithm that covers the vertices with lowest degree across the partition and we leave a smaller subtree to embed by a matching argument at the end. The decomposition of T into subtrees in this case is such that we can 2-color the components so that the union of the red color classes and the union of the blue color classes are the right size. We embed the edges connecting the colored components first, then embed the components, taking care of their connections through random subsets that were set aside at the beginning. We finish again by a matching argument to map the large subtree held in reserve.

1.2 A New Proof of the Pósa Conjecture

The k^{th} *power* of C is the graph obtained from C by joining every pair of vertices at a distance at most k in C. Let G be a graph on $n \geq 3$ vertices. A classical result of Dirac [4] asserts that if $\delta(G) \geq n/2$, then G contains a Hamiltonian cycle. As a natural generalization of Dirac's theorem, in 1962 Pósa (see Erdős [5]) conjectured the following:

Conjecture 2 (Pósa). *Let G be a graph on n vertices. If $\delta(G) \geq \frac{2}{3}n$, then G contains the square of a Hamiltonian cycle.*

Later in 1974 Seymour [25] generalized this conjecture:

Conjecture 3 (Seymour). *Let G be a graph on n vertices. If $\delta(G) \geq \frac{k}{k+1}n$, then G contains the k^{th} power of a Hamiltonian cycle.*

Seymour indicated the difficulty of the conjecture by observing that the truth of this conjecture would imply the remarkably difficult Hajnal-Szemerédi Theorem [14], namely that if $\Delta(G) < r$, then G is r colorable such that the sizes of the color classes are all $\lfloor \frac{n}{r} \rfloor$ or $\lceil \frac{n}{r} \rceil$.

These problems received significant attention. In the direction of Conjecture 2, first Jacobson (unpublished) showed that if $\delta(G) \geq 5n/6$, then the conclusion of the conjecture holds.

Faudree, Gould, Jacobson and Schelp [12] confirmed the conclusion with $\delta(G) \geq (3/4+\varepsilon)n+C(\varepsilon)$. Later the same authors improved this to $\delta(G) \geq 3n/4$. By using a result in [13], Häggkvist (unpublished) gave a very simple proof in the case $\delta(G) \geq 1 + 3n/4$ and $n \equiv 0 \pmod 4$. Fan and Häggkvist in [6] lowered the bound to $\delta(G) \geq 5n/7$. Fan and Kierstead improved this further to $\delta(G) \geq (17n + 9)/24$ in [7], and Faudree, Gould and Jacobson [11] to $\delta(G) \geq 7n/10$. Then Fan and Kierstead [8] improved the condition to the almost optimal $\delta(G) \geq \left(\frac{2}{3} + \varepsilon\right) n + C(\varepsilon)$. They also proved [9] that already $\delta(G) \geq (2n - 1)/3$ is sufficient for the existence of the square of a Hamiltonian *path*. Furthermore, they also proved [10] that if $\delta(G) \geq 2n/3$ and G contains the square of a cycle with length greater than $2n/3$, then G contains the square of a Hamiltonian cycle. Finally, Kierstead and

Quintana [16] proved that if $\delta(G) \geq 2n/3$ and G contains a maximal 4-clique, then G contains the square of a Hamiltonian cycle.

For Conjecture 3, in the above mentioned paper of Faudree *et al* in [12], it is proved that for any $\varepsilon > 0$ and positive integer k there is a C such that if an n-graph G satisfies

$$\delta(G) \geq \left(\frac{2k-1}{2k} + \varepsilon\right) n + C,$$

then G contains the k^{th} power of a Hamiltonian cycle.

Using the Regularity Lemma-Blow-up Lemma method first in [19], J. Komlós, G.N. Sárközy and E. Szemerédi, proved Conjecture 3 in asymptotic form, then in [17] and [20] they proved both conjectures for $n \geq n_0$. The proofs used the Regularity Lemma [26], the Blow-up Lemma [18], [21] and the Hajnal-Szemerédi Theorem [14]. Since the proofs used the Regularity Lemma the resulting n_0 is very large (it involves a tower function). The new result presented in this thesis is another proof for $k = 2$ that avoids the use of the Regularity Lemma.

Theorem 3. *There exists a natural number n_0 such that if a graph G has order n with $n \geq n_0$ and $\delta(G) \geq \frac{2}{3}n$, then G contains the square of a Hamiltonian cycle.*

This new proof employs two lemmas which, conceptually, could be used to remove the Regularity Lemma from other extremal graph theoretical proofs. The first lemma, more specific to the case of embedding k^{th} powers of cycles, is the Reservoir-Connecting Lemma. It states simply that any two ending edges of a square path can be connected through a random set (the reservoir) by a short square path. The second lemma, the Absorbing Lemma, states that there exists a square path P_A such that *any* sufficiently small set of vertices can be inserted into P_A without destroying the property of being a square path. The structure of the proof is to extend P_A to a maximal square path P, and then by combinatorial arguments show that we can rearrange P such that we either increase its length or we can find a long square path which we connect to P via the Reservoir-Connecting Lemma. The arguments hold until there is a small set of vertices not covered by P. At that point we connect the endpoints of P and absorb the remaining vertices by the Absorbing Lemma. It is this last step that provides quite a bit of power in removing the Regularity Lemma.

As in [20], the proof makes use of the assumption that G is not extremal. For the case of $k = 2$, the extremal graph G is one in which there is a set of roughly $\frac{1}{3}n$ vertices which induce a subgraph with very low density. It is easy to find a square path in the particular extremal graph comprised of an independent set S of size $\frac{1}{3}n$, implying that $G - S$ has minimum degree $|G - S|/2$. Dirac's theorem ensures that there is a Hamiltonian cycle C in $G - S$, and the vertices of S are easily inserted into C to extend it to a square cycle. The full extremal case is not far from this.

In fact the combinatorial methods developed can be extended and generalized to prove Conjecture 3 for $k \geq 3$, but the details become considerably more complicated. We have identified another way to remove the Regularity lemma from the proof of Conjecture 3 which applies for $k \geq 2$ and is similar to the arguments in [21]. We are able to cover most of the vertices of G with a collection of $K_k(t)$s by Kővári-Sós type arguments. We can easily convert the $K_k(t)$s to disjoint k^{th} powers of cycles and then connect them via the Connecting Lemma. Some elementary arguments are then used to insert the remaining vertices of G. In the case of $k = 2$ the Connecting-Absorbing method yields the smallest n_0 and it is this argument that we present in full detail.

We have also identified an alternate Lemma for connecting square paths to replace the costly Reservoir-Connecting Lemma. This new Connecting Lemma can also be generalized for $k \geq 3$. While the is work ongoing, the expectation is that this will allow us to push n_0 down far enough to develop a program to check the conjecture for the remaining n. In this, the combinatorial structure identified in the proof will prove useful as it lends itself to fast algorithms. It is our hope to establish the truth of Conjecture 2 for every n.

Chapter 2

Proof of the Bollobás Conjecture on Embedding Trees

2.1 Introduction

Let $G = (W, E(G))$ be a graph on n vertices. Assume that its minimum degree, $\delta(G) \geq n/2 + c\log n$. Let $T = (V, E(T))$ be a bounded degree tree on n vertices. We denote its maximum degree by K. Our goal is to prove that $T \subset G$ if $n \geq n_0$. We will show that there exists a bijective adjacency preserving mapping $I : V \to W$ by the help of a randomized embedding algorithm.

We divide the problem into two subproblems: we make distinctions depending on G being *extremal* or *non-extremal.* We call G η-extremal, if either G or its complement $\overline{G}$ contains a subgraph on $n/2$ vertices with at most ηn^2 induced edges for $\eta > 0$. It will be η-non-extremal otherwise. Let $\gamma = K^{-20}$. First we prove the statement for γ-non-extremal graphs, then for γ-extremal graphs.

2.2 Definitions and Notation

For a graph G, we denote by $v(G)$ the number of vertices of G and by $e(G)$ the number of edges in G. If $H \subset G$, we let $G - H$ denote the subgraph of G obtained by deleting the vertices of H from G. In the case of a single vertex v of G, we let $G - v$ denote the graph obtained by deleting v. For sets A and B, we let $A \setminus B = \{a : \ a \in A, \ a \notin B\}$. The minimum degree of G is denoted by $\delta(G)$. For $v \in G$, we denote by $N(v)$ the neighborhood of v. For a subset of vertices A, we will sometimes write $N_A(v)$ for $N(v) \cap A$.

Assume that we are given a rooted tree F with root ρ. Let x and y be any two vertices of F. We say that y is *below* x if the only path connecting y with ρ goes through x. $F(x)$

denotes the subtree rooted at x containing every vertex which is below x. For a tree F with root r, the *depth* of a vertex $x \in T$ is the length of the unique path from x to r.

2.3 The non-extremal case

The case of a non-extremal G is divided into two subcases depending on the structure of T. The first case is when T has at least $\gamma^2 n$ leaves, the second is when the number of its leaves is smaller than $\gamma^2 n$.

2.3.1 Proof sketch

Since we have many leaves, we are able to find a subtree $T_0 \subset T$ of size at most $\gamma^5 n$ which has about $\gamma^8 n$ distinguished leaves, each of distance at least 12 from each other. While we map most of T_0 in the beginning, the set of the distinguished leaves will be taken care of at the end of the embedding process, giving us some freedom.

Let $T_* = T - T_0$. We find a decomposition of T_* such that most of its vertices belong to certain subtrees of bounded size. We map the subtrees of the decomposition one at a time with the help of the Main Mapping Procedure. As this procedures requires that the vertices of G not yet covered contains a large subgraph with a non-negligible minimum degree, we may have to stop before all of the subtrees are mapped. If the set of uncovered vertices of G is too sparse, we rethink the way we mapped certain subtrees. We *rearrange* the mapping of some subtrees in such a way that we use many uncovered vertices and the new set of uncovered vertices contains a large dense bipartite graph. This dense bipartite subgraph will then help us to continue the embedding of T. When we arrive at the point that at most $\gamma^{10} n$ vertices of T are left to be mapped (plus the set of distinguished leaves of T_0), we can easily finish the embedding.

2.3.2 Some tools for the proof

We will need the following remark:

Remark 1 *If G is not γ-extremal then, since $\delta(G) > n/2$, there are at least γn^2 edges in the neighborhood of every vertex. In particular, every vertex is in at least γn^2 triangles.*

We make another simple statement which is also used throughout the proof:

Lemma 4 *Let H be a tree of size m. If J is a graph with $\delta(J) \geq m$, then there is an embedding of H into J. If $J' = J'(A, B)$ is a bipartite graph such that every $b \in B$ has at least $m/2$ neighbors in A and every $a \in A$ has at least m neighbors in B then $H \subset J'$.*

Proof: We leave the proof of the first part to the reader. For the second part we remark that H is 2-colorable, and one of its color classes has size at most $m/2$. □

The following standard lemma can be found in, e.g., [3].

Lemma 5 *Every graph H has a subgraph H' such that $\delta(H') \geq e(H)/v(H)$.*

Proof: Omitted. □

Let $F = F(A, B)$ be a bipartite graph satisfying the following requirements:

- $|A| = t$ and $|B| = t^2$,
- $\gamma^{-3} \ll t$,
- every $b \in B$ has at least $(1/2 + \gamma^3)t$ neighbors in A.

Then we have the following *Cleaning Lemma.*

Lemma 6 (Cleaning lemma - first version) *F has a subgraph $F' = F'(A', B')$ such that $A' \subset A$, $B' \subset B$, every $b \in B'$ has at least $(1/2 + \gamma^3/2)t$ neighbors in A' and every $a \in A'$ has at least t neighbors in B'.*

Proof: We may assume that every vertex in B has exactly $(1/2+\gamma^3)t$ neighbors in A—if necessary, we discard edges from those vertices of B which have larger degrees. Then we have $|E(F)| = (1/2 + \gamma^3)t^3$. Step by step we find the desired subgraph of F.

Let $A_1 = \{a \in A : \ |N(a) \cap B| < t\}$, and let $B_1 = \{b \in B : \ |N(b) \cap A_1| \geq \gamma^3 t/2\}$. We delete the vertices of A_1 from A and the vertices of B_1 from B. This is what we call the *cleaning step*—getting rid of vertices from both color classes which obviously violate the requirements of the lemma. In the cleaning step we lose at most $|A_1|t+2|A_1|t(1/2+\gamma^3)/\gamma^3$ edges from F.

Assume that we have performed k cleaning steps, identifying the sets $A_1, A_2, \dots A_k \subset A$ and $B_1, B_2, \dots B_k \subset B$. In the next step we set

$$A_{k+1} = \{a \in A \setminus \bigcup_{i=1}^{k} A_i : \ |N(a) \cap (B \setminus \bigcup_{i=1}^{k} B_1)| < t\}$$

and

$$B_{k+1} = \{b \in B \setminus \bigcup_{i=1}^{k} B_i : \ |N(b) \cap (\bigcup_{i=1}^{k} A_k)| \geq \gamma^3 t/2\}.$$

By deleting the vertices of A_{k+1} and B_{k+1} we lose at most $|A_{k+1}|t + 2|A_{k+1}|t(1/2+\gamma^3)/\gamma^3$ edges.

After k cleaning steps, the number of edges we lose is at most

$$\sum_{1 \leq i \leq k} |A_i|(t + 2t(1/2+\gamma^3))/\gamma^3 < t^2(3 + 1/\gamma^3) \ll t^3/2.$$

At some point the procedure will stop, say at step k_0. After the k_0^{th} cleaning step either every vertex satisfies the degree requirements of the lemma or there is no vertex left. As at most $t^3/2 < (1/2+\gamma^2)t^3$ edges have been lost, we must have the former case. Setting $A' = A \setminus \bigcup_{i=1}^{k_0} A_i$ and $B' = B \setminus \bigcup_{i=1}^{k_0} B_i$, the induced bipartite graph $F' = (A', B')$ is easily seen to satisfy the lemma. □

We will need another version. For that we assume that F satisfies the following:

- $|A| = t$ and $|B| = t^2$,
- $\gamma^{-4} \ll t$,
- every $b \in B$ has at least $(1/2 - \gamma^4)t$ neighbors in A.

Then we have the second version of the Cleaning Lemma:

Lemma 7 (Cleaning lemma - second version) *F has a subgraph $F' = F'(A', B')$ such that $A' \subset A$, $B' \subset B$, every $b \in B'$ has at least $(1/2 - 2\gamma^4)t$ neighbors in A' and every $a \in A'$ has at least t neighbors in B'.*

Proof: Very similar to the proof of the first version, we omit the details. □

We choose three disjoint subsets randomly from W. Let D_1, D_2 and M, be randomly chosen such that they are disjoint, $|D_1| = |D_2| = \gamma^8 n$, and $|M| = \gamma^{10} n$. The common size of the D_is will be denoted by s. Applying the Chernoff bound one can conclude that every $u \in W$ will have at least $s/2 - O(\sqrt{n \log n})$ neighbors in the D_is and $|M|/2 - O(\sqrt{n \log n})$ neighbors in M with high probability. If G is γ-non-extremal, then the subgraphs induced by D_1, D_2 and M will be $\gamma/3$-non-extremal with high probability. Moreover, the bipartite subgraph induced by D_1 and D_2 will be $\gamma/3$-non-extremal with high probability, i.e., if $A \subset D_1$, $B \subset D_2$, both having size $s/2$, then there are at least $\gamma s^2/3$ edges in between A and B with high probability. The proofs of these statements are straightforward. We sketch a proof of the non-extremality of D_i. We will show that the induced subgraph on D_1 cannot have a sparse subset of size $s/2$. The analogous statement, that the complement of the above induced subgraph cannot have a large subset with a few edges can be proved similarly.

It is a simple exercise that if G is not extremal, then any set U for which $\forall a, b \in U$, $|N(a) \cup N(b)| \leq (1/2 + \gamma)n$ cannot contain more than $(1/2 - \gamma)n$ vertices. Thus, with high probability, a set $A \subset D_1$ of size $s/2$ will have the property that for many pairs of vertices $a, b \in A$,

$$|(N(a) \cup N(b)) \cap D_1| \geq (1/2 + \gamma)s + O(\sqrt{n \log n}).$$

It follows that for many pairs of vertices, $|(N(a) \cup N(b)) \cap A| \geq \gamma s$. Thus A is dense.

We will need the following Lemma:

Lemma 8 *Let G be a non-extremal graph, and D_1, D_2 be as above. Then there is a perfect matching in the bipartite subgraph of G induced by D_1 and D_2.*

Proof: We will check the König-Hall conditions in order to prove the existence of a perfect matching. Since every $u \in W$ has at least $s/2 - o(n)$ neighbors in D_2, every $H \subset D_1$ such that $|H| \leq s/2 - \gamma^2 s$ will have a neighborhood of size at least $|H|$ in D_2. Let us assume, that $|H| > s/2 - \gamma^2 s$. If D_2 had a subset $\widehat{H}$ of size $s/2 - \gamma^2 s$ such that $e(H, \widehat{H}) = 0$, then the induced bipartite subgraph on D_1 and D_2 would be extremal – if we complete H and $\widehat{H}$ by adding at most $\gamma^2 s$ new vertices to both, the number of newly added edges is not more than $2\gamma^2 s^2 < \gamma s^2/3$. Hence, every $H \subset D_1$ having size $s/2 - \gamma^2 s$ has a neighborhood of size at least $s/2 + \gamma^2 s$ in D_2. We finish by noticing that every vertex in D_2 has $s/2 - o(n)$ neighbors, therefore, if $H \subset D_1$ is larger than $s/2 + o(n)$ then it is neighboring with the whole D_2. □

Corollary 9 *Let us modify the bipartite subgraph of the previous lemma: we discard at most $\gamma^2 s$ vertices from D_2, and substitute those with arbitrary vertices which all have at least $|D_1|/2 - o(n)$ neighbors in D_1. Then the resulting new graph has a perfect matching.*

Proof: The proof of Lemma 8 works with a minor change, we omit the details. □

Lemma 10 *Assume that G is non-extremal and $u, v \in W$. Then there are at least $\gamma n/5$ vertex disjoint paths of length 3 connecting u and v.*

Proof: Set $a = |N(u) \cap N(v)|$. If $a \geq n/2 - \gamma n/2$, then by the non-extremality of G there are at least $\gamma n^2/2$ edges in $N(u) \cap N(v)$. One can choose $\gamma n/2$ vertex disjoint edges from these, and every such edge with u and v forms a path of length 3 between u and v. The other extreme is when $a \leq \gamma n/2$. Then there are at least γn^2 edges in $N(u)$ induced by $\overline{G}$, since G is non-extremal. If v' is the endpoint of such an edge, then it has a neighbor in $N(v)$. This overall means at least γn^2 edges between $N(u)$ and $N(v)$, and as before, we can find the vertex disjoint path of length 3 between u and v.

Assume now that $\gamma n/2 \leq a \leq n/2 - \gamma n/2$. Then the number of edges connecting $N(u) \cap N(v)$ with $N(u) \cup N(v)$ is at least $a(n/2 - a) \geq \gamma n^2/5$. One can choose $\gamma n/5$ vertex disjoint edges from these, which give us the desired vertex disjoint paths of length 3. □

2.3.3 Preparations

Finding T_0 and the distinguished leaves

We need to prepare T for the embedding. We begin by choosing an arbitrary root ρ for T, which will be changed later. Our main goal is to determine a subtree T_0 with size at most $\gamma^5 n/4$ and number of leaves at least $\gamma^7 n/K^2$.

To find T_0 we employ a simple algorithm. First color every vertex of T red. Then pick an arbitrary red leaf $x \in T$, and consider a vertex y_1 with the property that $x \in T(y_1)$,

$v(T(y_1)) \leq \gamma^5 n/4$ and if z is any other vertex on the path connecting y_1 with ρ, then $v(T(z)) > \gamma^5 n/4$. Erase the color of the vertices of $T(y_1)$ and repeat the procedure with another red leaf, if it exists. The algorithms terminates when every leaf of T is contained in a subtree $T(y_i)$.

We need a good upper bound on the number of subtrees we have found this way. Consider a related collection of subtrees, whose roots are obtained by moving one vertex in the direction of ρ from every y_i. All of these new subtrees will have size at least $\gamma^5 n/4$, hence, the number of them is at most $4\gamma^{-5}$. Since every subtree in this new collection contains at most K subtrees output by the algorithm, we get an upper bound of $4K\gamma^{-5}$ on the number of subtrees $T(y_i)$.

There are at least $\gamma^2 n$ leaves in T and so there is a subtree $T(y_i)$ of size at most $\gamma^5 n/4$ having at least $\gamma^7 n/K^2$ leaves. We denote this subtree by T_0, and take its root r to be the root of T throughout.

By the properties of T_0, we can easily find $s = \gamma^8 n$ subtrees of T_0, denoted by $\{F_1, F_2, \ldots, F_s\}$, such that each F_i has depth 4, and for each i, j, the distance between F_i and F_j is at least 4. For each $1 \leq i \leq s$, distinguish a leaf of F_i which is at depth 4 and denote it by a_i. The immediate ancestor of a_i is denoted by $P(a_i)$ and the root of F_i is called $A(a_i)$. By construction, $P(a_i)$ is distance three from $A(a_i)$. Denote the set of distinguished leaves by $S = \{a_1, a_2, \ldots, a_s\}$. We will map $P(a_1), P(a_2), \ldots, P(a_s), a_1, a_2, \ldots, a_s$ in a special way. In particular, we will map the $P(a_i)$s to the vertices of D_1, and most of the a_is to D_2.

Embedding T_0

We start the embedding of T by that of T_0. Denote by $T_0' = T_0 - \cup(F_i - A(a_i)$ the subtree of T_0 obtained by removing each F_i but leaving its root. We map T_0' greedily into $W \setminus (M \cup D_1 \cup D_2)$. When we are done, all of the $A(a_i)$ will have been mapped. We then map the $P(a_i)$ into D_1 arbitrarily. By Lemma 10 we find vertex disjoint paths of length 3 connecting $I(A(a_i))$ with $I(P(a_i))$, such that these paths avoid D_2 and M, and contain vertices from D_1 only as an endpoint. We finish the embedding of the F_i greedily into $W \setminus M \cup D_1 \cup D_2)$.

We set $T_* = T - T_0$.

2.3.4 Decomposition of T_*

We find a decomposition of T_* into subtrees $T_1, T_2, \ldots, T_k$ such that for every $1 \leq j \leq k$, $e^c < |T_j| < Ke^c$ (c will be chosen later). Choose a maximal path P from r such that

(i) for every $x \in P$, $|T_*(x)| \geq Ke^c$

(ii) for $x, y \in P$ with x the parent of y, $T_*(y)$ is the largest subtree rooted at a child of x

If x is the last vertex on P, let x_1 be such that $T_*(x_1)$ is the largest subtree rooted at a child of x. We define $T_1 = T_*(x_1)$, easily seen to be of the desired size. Let $T_*^{(1)} = T_* - (T_1 - x_1)$ be the remainder of T_* after deleting all but the root of T_1 from T_*.

Assume that we have performed i iterations. Repeat the procedure on $T_*^{(i)}$, finding x_{i+1} and defining $T_{i+1} = T_*^{(i)}(x_{i+1})$ and $T_*^{(i+1)} = T_*^{(i)} - (T_{i+1} - x_{i+1})$. Let k be such that $|T_*^{(k)}| < Ke^c$, at which point the procedure stops. We set $T_{k+1} = T_*^{(k)}$ and note that T_{k+1} has r as its root. Throughout the proof, we let $t_i = |T_i|$

We remark that in this decomposition, T_i and T_j may have at most one vertex in common, and this common vertex is a leaf of one of the subtrees, say T_i, and the root of the other, T_j. If a leaf of a subtree T_i is the root of another, we call this a connecting leaf of T_i. For technical reasons, we have to avoid certain conflicts between connecting leaves of the same subtree. If two connecting leaves of T_i are such that their paths of length 3 towards the root intersect, we call these connecting leaves "conflicting". We adjust the decomposition so that there are no conflicting connecting leaves by adding a check at the end of each iteration. If at the completion of the i^{th} iteration the root x_i found is such that it is distance less than 3 from x_j for $j < i$ we choose the child x_i' of x_i such that $T_*(x_i')$ is the largest subtree rooted at a child of x. If x_i' is still in potential conflict with the root of another subtree, we repeat. After at most 3 corrections, we arrive at a T_i whose root will not conflict as a connecting leaf. Thus, the sizes of the subtrees in the decomposition may be as small as $\frac{e^c}{K^3}$, but this will not cause any difficulties.

2.3.5 Finishing the embedding of T

Let us assume that we have already mapped most of T_*, only the vertices of S and a small forest F are left out such that $|F| \leq \gamma^2 s$. Let $T'_* = T_* - (F \cup S)$. We embed F into D_2, connecting through M. This is easily done as D_2 is a randomly chosen subset of W and $|D_2| = s \gg \gamma^2 s$. We map all but $|F|$ vertices of S arbitrarily to $D_2 \setminus I(F)$. The remainder of S we map to $W \setminus (I(T'_*) \cup D_2)$. The embedding will be finished if we can find a perfect matching between D_1 and $I(S)$. As $I(S) = D_2 \setminus I(F) \cup (W \setminus (I(T'_*) \cup D_2))$, Corollary 9 applies and so such a perfect matching exists. Hence, if we can map most of T_* such that we avoid D_2, we can finish the embedding of T. The goal of the rest of this section is to achieve this mapping.

2.3.6 The Main Mapping Procedure

Here we define the method for mapping the subtrees of the decomposition. We start with T_{m+1}, which we map greedily into $W \setminus (I(T_0) \cup M \cup D_1 \cup D_2)$. This is the first step.

After the k^{th} step, denote the mapped portion of T_* by T'_* and set $Q = W \setminus (I(T'_*) \cup M \cup D_1 \cup D_2)$, the set of vertices of W that we will use for the embedding. Choose a connecting leaf y that is a leaf of T'_*, and let T_i be the subtree which has y as its root. Assume that there is a subset $Q' \subset Q$ with minimum degree $\gamma^{10} n$. Denote y's children by $y_1, y_2, \ldots, y_k$, and the children of the y_is by $r_1, r_2, \ldots, r_\ell$.

Assume that M has at least $(1-\gamma/20)|M|$ vacant vertices and denote the set of already covered vertices of M by X. Since M is a randomly chosen set, $|N(I(y)) \cap (M \setminus X)| \geq (1/2 - \gamma/19)|M|$. Define the set

$$U = \{u : u \in M \setminus X, |N(u) \cap Q'| \geq \sqrt{n}\}.$$

Since M is randomly chosen, every vertex of Q' has at least $(1/2 - \gamma/19)|M|$ neighbors in $M \setminus X$, which in turn implies that $|U| \geq (1/2 - \gamma/18)|M|$. Since the induced subgraph on M is $\gamma/3$-non-extremal, there are at least $\gamma/3|M|^2 - \gamma/9|M|^2$ edges in between U and $N(I(y)) \cap (M \setminus X)$. We map the y_is to such vertices in $N(I(y)) \cap (M \setminus X)$ which have at least K^2 neighbors in U. If $\{y_i, r_j\}$ is an edge in T, then $I(r_j)$ is chosen among the neighbors of $I(y_i) \cap U$. It is possible to embed the rest of T_i in Q' greedily since $|T_i|$ is constant and Q' is dense.

The general notion of connecting pieces of the embedding through M will be used throughout the proof. To this end, we make the following remark:

Remark 2 *In the procedure described above, we cover less than K^2 vertices from M while we embedded into at least $e^c - K^2$ vertices of Q. In general, when an embedding procedure is applied, if the ratio of the number of vertices of M to the number of vertices of Q covered in the procedure is always less than $\gamma^{12} \leq \frac{\gamma|M|}{20n}$, then we will be assured that less than $\gamma|M|/20$ vertices of M are covered at the end. We choose c so that this holds.*

2.3.7 The Second Mapping Method

Let us again denote by Q the set of those vertices of $W \setminus (M \cup D_1 \cup D_2)$ which have not yet been covered by a vertex of T_*. In the Main Mapping Procedure we used that there was a dense subset $Q' \subset Q$. This time we allow Q to be very sparse. Our goal now is to rearrange the embedding so that we return a dense subgraph to Q.

We call a subtree of the decomposition of Section 2.3.4 "good" if it contains less than γ^{-6} connecting leaves. As there are at most k connecting leaves overall, it follows that all but at most $\gamma^6 k$ subtrees are good, and so all but at most $\gamma^6 n$ vertices of T_* are in the union of the good subtrees. Without loss of generality, we assume that the good subtrees are $T_1, T_2, \ldots, T_m$.

Again denote by T'_* the portion of T_* that is already embedded. Assume that one of the subtrees of the decomposition $T_i \subset T'_*$ is such that Q has at least t_i^2 vertices all of which have at least $(1/2 + \gamma^3)t_i$ neighbors in T_i. If this condition is satisfied, we call T_i a *remappable* subtree. By the first version of the Cleaning Lemma we find a bipartite graph (A', B') such that $A' \subset I(T_i)$, $B' \subset Q$, every $b \in B'$ has at least $(1 + \gamma^3)t_i/2$ neighbors in A' and every $a \in A'$ has at least t_i neighbors in B'.

We embed T_i via the bipartite graph F' and Lemma 4, connecting to $I(T'_* - T_i)$ through M as in the Main Mapping Procedure with some minor modifications. We need to be careful with the embedding of the at most γ^{-6} connecting leaves of T_i, and so we precondition T_i before the remapping. Let $l_1, l_2, \ldots, l_d$ be the connecting leaves of T_i and for each connecting leaf l_j, let the path of length 3 towards the root beginning with l_j be (l_j, w_j, u_j, v_j). We will let T_i' be the result of trimming off the subtrees rooted at l_i, leaving the roots. More precisely, we set $T_i' = T_i - \bigcup_{j=1}^{d}(T_i(l_j) - l_j)$. We will embed T_i' first, and finish with the trimmed subtrees, ensuring that the newly embedded T_i is properly connected into the rest of the embedding.

First, using the notation from the previous section, we assume that the r_i, the vertices at distance 2 from the root of T_i, are in the smaller color class of T_i. If this assumption does not hold, then we embed the children of the r_i into $U \subset M$ rather than the r_i. This is easily done using the degree conditions and non-extremality of M. The second modification is in the definition of U. For the Second Mapping Method, we define

$$U = \{u : u \in M \setminus X, |N(u) \cap B'| \geq \sqrt{t_i}\}.$$

The rest of the procedure goes through: e.g. in the case that r_i is in the small color class, we map the y_i to $N(I(y)) \cap (M \setminus X)$, we map the r_i appropriately in U, and map the children of r_i into B'. At that point, Lemma 4 applies, and we greedily map the remainder of T_i' into F'. Finally we take care of the subtrees containing the connecting leaves l_j by first mapping the path (l_j, w_j, u_j, v_j) through M respecting the fact that l_j and v_j are already mapped. Then we map the remainder of T_i which lies below the path through U, into B', and then via F'. Note that we must be careful to respect the parity so that we use less than $t_i/2$ vertices of A' overall, mapping an extra generation in M if necessary. As the subtrees below the v_j are all disjoint by construction, this last step is possible.

As T_i may have γ^{-6} connecting leaves, we have used up a polynomial in K number of vertices from M without mapping any new subtrees. On the other hand, we have rearranged the mapping such that a dense bipartite graph $F''(A'', B'') \subset F(A', B')$ is moved to Q, where $A'' \subset A'$ and $B'' \subset B'$. As every vertex in B'' has degree to A'' at least $\frac{\gamma^3}{2}t_i$, there are at least $\frac{\gamma^3}{2}t|B''|$ edges in F''. The number of vertices in F'' is $|A''| + |B''| \leq 2|B''|$. Therefore F'' has a subgraph of minimum degree at least $\frac{\gamma^3}{4}t_i \geq \gamma^4 t_j$ for any unmapped tree T_j.

Assume that we are able to find γ^{-4} remappable subtrees. We rearrange their mappings, returning a dense subgraph as above to Q for each one. Then we take an unmapped tree T_j whose root is already mapped and decompose it into subtrees of size at most $K\gamma^4 t_j$ but at least $\gamma^4 t_j$ in a similar manner to section 2.3.4. Starting with the already mapped root, we connect through M and into one of the dense bipartite subgraphs in Q generated by the remapping. We map this piece greedily witin the dense bipartite graph and then repeat with another piece whose root is now mapped. As the decomposition of T_j is into at most γ^4 pieces, we use at most a polynomial in K number of vertices of M to connect the decomposed pieces of T_j.

We cover overall a polynomial in K number of vertices in M by this procedure, but map a new subtree which is of size at least $\frac{e^c}{K^3}$ so we do not use up the vertices of M too quickly. Again, c is chosen such that the ratio of newly-mapped vertices of M to newly-mapped vertices of Q is favorable.

2.3.8 The Third Mapping Method

As in the second mapping method we assume that Q is very sparse and $|Q| > \gamma^{10}n$. The sparsity of Q is made precise. As the Main Mapping Procedure can not be applied, there is no subgraph of Q with minimum degree at least $\gamma^{10}n$. Thus, the number of edges in Q is at most $\gamma^{10}nq$, where $q = |Q|$. We further assume that we cannot apply the Second Mapping Method. That is, for all but at most γ^{-4} good subtrees T_i, there are less than t_i^2 vertices having at least $(1/2 + \gamma^3)t_i$ neighbors in T_i. We remark that it follows easily that most of T_* has been already embedded.

As we cannot find enough good remappable subtrees to extend the embedding, we will look for a *weakly remappable subtree.* A weakly remappable subtree is a good subtree T_i for which there are at least t_i^2 vertices in Q each having at least $(1/2 - \gamma^2)t_i$ neighbors in T_i.

We claim that most of the embedded subtrees are weakly remappable if we cannot apply the Second Mapping Method. More precisely:

Lemma 11 *In the above setup if none of the good T_i subtrees are remappable then at least $(1-\gamma/2)n$ vertices are in weakly remappable good subtrees.*

Proof: If T_i is not a weakly remappable subtree, then as T_i is not remappable, there are at most t_i^2 vertices in Q which have more than $(1/2+\gamma^3)t_i$ neighbors in T_i. As it is not weakly remappable, the rest of the vertices of Q have less than $(1/2-\gamma^2)t_i$ neighbors. If T_i is weakly remappable then again, as it is not remappable, at most t_i^2 vertices of Q have t_i neighbors in T_i, and the remaining vertices of Q have at most $(1/2+\gamma^3)t_i$ neighbors in T_i.

On the other hand, $|T_0| \le \gamma^5 n/4$, there are at most $\gamma^{-4}Ke^c$ vertices in $I(T'_*)$ that are in remappable subtrees, there are at most $\gamma^6 n$ vertices in bad subtrees, and $e(Q) \le \sqrt{n}q$, and $|D_1 \cup D_2 \cup M| < 3\gamma^8 n$. It follows that the number of edges from Q to the good embedded subtrees which are not remappable is at least

$$(n/2 - \gamma^5 n/4 - \gamma^{-4}Ke^c - \gamma^6 n - 3\gamma^8 n - \gamma^{10} n)q \ge (1/2 - \gamma^5/3)nq.$$

Let $\mathcal{I}_1 = \{i : T_i \text{ is good, not remappable, and not weakly remappable}\}$ and $\mathcal{I}_2 = \{i : T_i$ is good and weakly remappable$\}$. We have the following inequality:

$$(\frac{1}{2} - \frac{\gamma^5}{3})nq \le \sum_{i \in \mathcal{I}_1} \left(t_i^3 + (\frac{1}{2} - \gamma^2)t_i q\right) + \sum_{i \in \mathcal{I}_2} \left(t_i^3 + (\frac{1}{2} + \gamma^3)t_i q\right) \le$$

$$\frac{n}{e^c/K^3}(Ke^c)^3 + (\frac{1}{2} - \gamma^2)dq + (\frac{1}{2} + \gamma^3)(n-d)q$$

where d denotes the number of vertices in subtrees which are not weakly remappable. A simple calculation shows that $d < \gamma n/2$. □

If T_i is weakly remappable then there is a subset of Q having size at least t_i^2 such that all of its vertices have at least $(1/2-\gamma^2)t_i$ neighbors in T_i. By the second version of the Cleaning Lemma we can remap most of T_i, leaving at most $K\gamma^2 t_i$ vertices unmapped. If this partial remapping of T_i is such that we gain a large dense subgraph in Q, then we can use that subgraph to build T_* further. Denote by R_i the part of $I(T_i)$ that is not used in the remapping. Let $r_i = |R_i|$. By the above $|\cup_i R_i| = \sum_i r_i \ge n(1/2 - 2\gamma/3)$. Since G

is non-extremal, $\cup_i R_i$ has many edges, at least $\gamma n^2/3$. We will show that there are many pairs (R_i, R_j) which have many edges between them.

Lemma 12 *There exist at least γ^{-2} disjoint pairs (R_i, R_j) such that $e(R_i, R_j) \geq \gamma r_i r_j$.*

Proof: Let us assume that, on the contrary, every pair has density less than γ. Then

$$e(\cup R_i) = 1/2 \sum_{i \neq j} e(R_i, R_j) + \sum_i e(R_i).$$

The sum $\sum_i e(R_i) = o(n^2)$, so we concentrate on the first part.

$$1/2 \sum_{i \neq j} e(R_i, R_j) \leq \gamma/2 \sum_{i \neq j} r_i r_j \leq \gamma/2 \sum_i r_i(n/2 - r_i).$$

This last expression is at most $\gamma(n/4 \sum_i r_i - \sum_i r_i^2) \leq \gamma n^2/4 - o(n^2)$, a contradiction. Thus for some positive (but small) constant α, we can find at least αn^2 dense pairs. In particular there are at least γ^{-2} disjoint dense pairs. □

If most of the subtrees are weakly remappable and less than γ^{-4} are remappable, then we do the following. Apply Lemma 12 to find γ^{-2} disjoint dense pairs. For one such pair (R_i, R_j) remap T_i and T_j. After the remapping at most $K\gamma^2 t_i$ and $K\gamma^2 t_j$ vertices are left unmapped from these subtrees. As (R_i, R_j) is a dense pair, there are at least $\gamma r_i r_j$ edges between them. By Lemma 5 we can find a subgraph of this dense pair in which the minimum degree is at least $\gamma \frac{r_i r_j}{r_i + r_j}$. Using that $e^c/K^3 \leq t_i \leq Ke^c$ and $r_i \geq t_i/2$, it can be shown that the minimum degree in this subgraph is enough to map the leftover portions of T_i and T_j of size $K\gamma^2 t_i$ and $K\gamma^2 t_j$ and still have minimum degree in this subgraph of at least $\gamma^2 K e^c$. Thus, if we remap such that at least γ^{-2} such subgraphs are returned to Q, then in a manner similar to the Second Mapping Method, we can map a new subtree by way of these subgraphs. We end with the remark that the we use at most a polynomial in K number of vertices from M while we map an exponential number of vertices from Q.

2.3.9 Description of the Embedding Algorithm

We are now ready to give the embedding method in case G is non-extremal and T has many leaves. After we are done with the mapping of the vertices of T_0 except the vertices of S, we decompose teh remainder of T, T_*, and apply the Main Mapping Procedure to map the subtrees of the decomposition starting from the root. It is possible to do so until Q drops below $n/2$, at which point we are not guaranteed to have the necessary minimum degree in Q. We look for at least γ^{-4} remappable subtrees that we can remap and use the resulting dense subgraphs to map a new tree from the decomposition. If there are not enough remappable subtrees then most of T_* has been mapped, and most of the embedded good subtrees are weakly remappable. It follows that the set comprised of those vertices of the weakly remappable subtrees which are not used in the potential (partial) remappings is nearly $n/2$, and so contains many edges. This in turn implies that, there are many disjoint pairs of weakly remappable subtrees for which remapping them would return a dense bipartite subgraph to Q. Again, we can use these dense subgraphs to map a new unmapped tree of the decomposition. We proceed in this manner until the size of Q drops below $\gamma^{10}n$. Then we apply the matching method of Section 2.3.5, and find the embedding of T into G.

2.3.10 T has at most $\gamma^2 n$ leaves

In this section we will discuss the case of embedding T with at most $\gamma^2 n$ leaves. We first give a sketch of the procedure.

A *stem* in a tree T is a path P such that every vertex on P apart from possibly the endpoints has degree 2 in T. In the case that T has only $\gamma^2 n$ leaves, it must contain many long stems. We find a set of stems $\{P_1, \ldots, P_{k'}\}$ where each P_i has $1/(4\gamma^2)$ vertices, and k' is around $\gamma^{14}n$. These stems are chosen from the good subtrees (recall the tree decomposition of the previous case). We remove one vertex from every P_i and glue the two new endpoints together. We denote the resulting new stems by $\widehat{P}_i$ and the resulting tree by $\widehat{T}$, which has roughly $(1-\gamma^{14})n$ vertices.

We prepare G for the embedding of $\widehat{T}$ by first embedding the stems $\widehat{P}_i$ by way of a randomized procedure. Then we apply the Main, Second, and Third embedding procedures of the previous section to embed the rest of $\widehat{T}$. Note that we will have no problem embedding $\widehat{T}$ as it is sufficiently smaller than G and the applicability conditions for at least one of the procedures will always apply. Then we handle the vertices which were removed from the stems by inserting one vertex into each $\widehat{P}_i$. This will be done by finding a perfect matching between the unmapped vertices of G and the $\widehat{P}_i$ such that a vertex v matched to a stem $\widehat{P}_i$ implies that v is adjacent to two consecutive vertices on $I(\widehat{P}_i)$.

We first show that if F is a tree having only a few leaves then we can find many long stems.

Lemma 13 *Let F be a tree on t vertices with dt leaves for some $0 < d \ll 1$. Then it is possible to find $dt/10$ vertex disjoint stems in F, each of size $1/(2d)$.*

Proof: Choose a root ρ. Substitute every maximal stem in F by one edge (we keep ρ). Then every vertex in the new tree F' will have degree 1 or at least 3 except possibly the root. We have at most $2dt$ vertices in F', and at least dt edges. We had $t-1$ edges in F, hence an average edge of F' corresponds to a stem of length a least $1/(2d)$. Cut out a part of this stem of length $1/(2d)$, and then glue together the resulting two endpints. Repeat this procedure: construct a new F' as before, and then find a long stem again by an averaging argument. We can continue this way, and find many stems of length $1/(2d)$, until the leftover number of edges in F is at least $t/2$. □

Apply the decomposition of Section 2.3.4, and find the good subtrees. Call a good subtree *long* if it has size t and has at most $2\gamma^2 t$ leaves. It is easy to see that almost half of the vertices are in long subtrees. Then we will find k stems of length $1/(4\gamma^2)$ by choosing less than $10\frac{K^4}{e^c}\gamma^{12}n$ long subtrees and finding $\gamma^2 t_i/10$ long stems in subtree T_i by the help of Lemma 13. Note that in Lemma 13 if $d \ll \gamma^2$ we find fewer stems of much greater length. In these cases, of course, we divide the stems into the appropriate length. Finally, for each stem that we obtain we trim two edges off of the end before choosing it as one of the stems that we embed. This is to ensure that they are all at distance at least three from each other. We denote by $P_1, P_2, \ldots, P_k$ these long chosen stems.

We prepare G for the embedding by picking a random subset $M \subset W(G)$ with size $\gamma^{10}n$. This will be the random set M that we need when performing the mapping procedures of the previous sections. We begin with the following Randomized Path Embedding procedure.

Step 1. Pick randomly, independently $\frac{k}{16\gamma^2}$ edges from $G - M$, with replacement.

Step 2. Form randomly subsets of these edges each having size $\frac{1}{16\gamma^2}$ and discard those subsets which contain repeated vertices. Connect the edges in every remaining subset into a path of length $\frac{1}{4\gamma^2} - 1$.

Step 3. Remove a vertex from each stem of the long subtrees that will be mapped onto the surviving random paths. Map the long subtrees such that the modified long stems map to the random paths.

Step 1 is obvious.

We analyze the edge sets that we found in Step 1. Let v be any vertex in G. The expected number of occurrences of v in the randomly chosen edges is proportional to γ^{12}. This number is a random variable which follows a Poisson distribution. Therefore, the probability that v has at least two occurrences in the random edge set is at most

$$\sum_{i \geq 2} \frac{\gamma^{12i} e^{-\gamma^{12}}}{i!} < \gamma^{20}.$$

A subset of $\frac{1}{16\gamma^2}$ edges will contain a vertex which occurs multiple times in the original draw with probability at most $\frac{\gamma^{20}}{8\gamma^2}$. Therefore, with high probability, $k' = (1 - \gamma^{18})k$ subsets will contain vertices not repeated in any other subset. To complete Step 2, for each subset of $\frac{1}{16\gamma^2}$ edges, we use Lemma 10 to connect the endpoints of the edges of each subset in series by paths of length 3. We need to extend the resulting paths by 2 edges to ensure that they are of length exactly $\frac{1}{4\gamma^2} - 1$.

Identify k' of the stems, say $P_1, P_2, \ldots, P_{k'}$, to be mapped onto the randomly constructed paths. For each $1 \leq i \leq k'$, take three consecutive vertices on P_i—x_i, y_i and z_i. We remove y_i from the path and add the edge $\{x_i, z_i\}$. We get a new path $\widehat{P}_i$ and a deleted vertex y_i. We denote the resulting tree on $(1 - k')n$ vertices by $\widehat{T}$.

Finally, for building the long subtrees we map greedily, connecting the endpoints of the P_i to their respective long subtrees through M. As we will need to perform less than $2\gamma^{12}n$ connections through M, we will not violate the requirement that at most $(\gamma/20)|M|$ vertices may be mapped for the mapping procedures that follow. Notice also that the condition that the stems are distance 3 from each other is used to be sure that the connections go through.

We say that we can insert a vertex u into $I(\widehat{P}_i)$ if u has two consecutive neighbors on the stem. Let us estimate the probability that u cannot be inserted to a given path $I(\widehat{P}_i)$. There are at least γn^2 edges in the neighborhood of u, since G is non-extremal. The probability that none of the $\frac{1}{16\gamma^2}$ randomly chosen edges is in the neighborhood of u is at most

$$(1-\gamma)^{\frac{1}{16\gamma^2}} \approx e^{-\frac{1}{16\gamma}}.$$

Since we choose the edges and then form the edge sets randomly, by Chernoff's inequality every vertex u can be inserted to at least $(1-e^{-\frac{1}{16\gamma}})k' - o(n)$ stems with probability tending to 1. Similarly, we get that the image of every mapped stem has many neighbors, i.e. vertices that can be inserted into it—at least $(1-e^{-\frac{1}{16\gamma}})n - o(n)$ with probability tending to 1. Let $A \subset W(G)$ be arbitrary, such that $|A| = k'$. Construct the following bipartite graph $F(A, B)$: the vertices of A are identified by the vertices removed from the embedded stems, and the vertices of B are identified by the randomly embedded stems. We connect $a \in A$ and $b \in B$ if the vertex corresponding to a can be inserted into the the image of the stem corresponding to b. As $|A| = |B| = k'$, the discussion above implies that the König-Hall conditions are satisfied with high probability. Therefore, we can find a perfect matching in F.

Now we are ready to discuss the embedding of T in case it has only a few leaves. First we choose $M \subset W(G)$ randomly of size $\gamma^{10}n$. We find paths of length $1/(4\gamma^2)-1$ randomly by the help of the Randomized Path Embedding procedure. We discard the bad paths, then correspond the remaining paths to the stems identified in the long subtrees. We cut out the middle vertex of these stems and embed the long subtrees containing them such that the stems map to the random paths.

Next we proceed to map the rest of the subtrees. All of the arguments go through for the Main Mapping Procedure, the Second Mapping Procedure, and the Third Mapping Procedure, with a single caveat. When mapping a new subtree into G, we may find that it contains a connecting leaf which is the root of one of the long subtrees that has already been embedded. Thus we will need to make an extra connection through M. This will not present a problem as there are at most $10\frac{K^4}{e^c}\gamma^{12}n$ extra connections to make throughout the embedding process.

The embedding goes smoothly as long as there are enough unmapped vertices left. As $v(\widehat{T}) = (1-k')n \le (1-\gamma^{12}/2)n$, there are *always* enough unmapped vertices, and so we map all of $\widehat{T}$. Finally, the remaining unmapped vertices are matched and then insterted into the modified stems, completing the embedding.

2.4 The Extremal Case

In this case we embed a tree T into an extremal graph G. An extremal graph G is either very close to being a complete bipartite graph on color classes having size $n/2$ with some edges in each part, or it is the union of two complete graphs on $n/2$ vertices each with some edges going between the parts. This two cases, while being very similar, will be handled separately.

We haven't used the full strength of the minimum degree of G in the non-extremal case. Even a minimum degree of $(1/2-\varepsilon)n$ for some small ε is sufficiently large if G is not extremal. On the other hand, if $G = K_{n/2} \cup K_{n/2}$, then it is not even connected. Likewise, if $G = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, then embedding T is only possible if there exists a two-coloring whose color classes are both of size $n/2$.

We assume that $\delta(G) \ge n/2 + K^4 C \log_{3/2} n$, where C is an absolute constant that does not depend on any other parameters.

In both extremal cases we will make use of a folklore result concerning trees.

Proposition 14 *Let J be any tree on m vertices. Then J has a vertex $u \in V(J)$ such that it is possible to group the vertices of $J - u$ into two forests, J_1 and J_2 such that $m/3 \le |J_1|, |J_2| \le 2m/3$ and there is no edge connecting J_1 and J_2 in $J - u$.*

We will call the vertex u a *split vertex.* We will repeatedly apply the above proposition, and get smaller and smaller subtrees.

The following embedding lemma will be applied in both cases:

Lemma 15 *Let F be a tree with maximum degree K and color classes Q_F and R_F. Let $H = (Q_H, R_H)$ be a bipartite graph such that*

1. *$q = |Q_F| = |Q_H|$ and $r = |R_F| = |R_H|$;*

2. *if $u \in Q_H$ then $deg(u) \geq (1 - \frac{1}{2K})r$;*

3. *if $v \in R_H$ then $deg(v) \geq (1 - \frac{1}{2K})q$.*

Then F is a spanning tree of H.

Proof: First we pick a random bijective mapping $\phi : Q_F \to Q_H$, then map every $x \in Q_F$ onto $\phi(x) \in Q_H$. We will extend ϕ to an embedding of F by a matching argument. Next we construct the auxiliary bipartite graph $L(R_F, R_H)$. Let $x \in R_F$ and $y_1, y_2, \ldots, y_k$ be the neighbors of x in F. We connect $u \in R_H$ with $x \in R_F$ if u is adjacent to $\phi(y_1), \phi(y_2), \ldots, \phi(y_k)$ in H. Obviously, if we can find a perfect matching in L then we have the claimed extension of ϕ to an embedding of F into H. For proving the existence of the perfect matching we will check the König-Hall conditions. In particular, we will show that every $u \in R_H$ is adjacent to at least $r/2$ vertices in R_F and that every $x \in R_F$ is adjacent to at least $r/2$ vertices in R_H with high probability.

Let $u \in R_H$ and $x \in R_F$ be arbitrary. As K is constant, the probability that $\{u, x\} \in E(L)$ is approximately

$$\left(1 - \frac{1}{2K}\right)^K > 0.56$$

By Azuma's inequality we get that every $u \in R_H$ is adjacent to more than $r/2$ vertices of R_F with very high probability.

We show that every $x \in R_F$ has at least $r/2$ neighbors in R_F. Let the neighbors of $x \in R_F$ be $y_1, y_2, \ldots, y_K \in Q_F$. By the minimum degree condition in H, every set of size K in Q_F has a common neighborhood of size at least $r/2$ in R_F, and so x is adjacent to at least $r/2$ vertices of R_F in L. $\square$

Remark 3 *Observe, that if there are $o(r)$ vertices of R_H such that these don't have many neighbors in R_F, but still connected to a constant proportion, we still can find the perfect matching in L, hence, find the embedding of F into H.*

2.4.1 G is close to $K_{\lceil n/2 \rceil} \cup K_{\lfloor n/2 \rfloor}$

In this case we find the vertex classes A and B such that $A, B \subset W(G)$, $A \cap B = \emptyset$ and $|A| = |B| = n/2$, and $e(G|_A), e(G|_B) \geq \binom{n/2}{2} - \gamma n^2$.

First we state a simple lemma on the degrees of the vertices of A and B :

Lemma 16 *Let $0 < \eta < 1/2$. Then the number of vertices of A having less than $n/2 - \eta n$ neighbors in A is at most $\frac{2\gamma n}{\eta}$. The analogue is true for B.*

Proof: Obvious. □

Preparations for the embedding – Step 1: Switching

We perform the following switching procedure: Let $A' \subset A$ contain those vertices which have at most $n/4$ neighbors in A, and let $B' \subset B$ contain those vertices which have at most $n/4$ neighbors in B. By Lemma 16 we have that $|A'|$, $|B'| \leq 8\gamma n$.

Redefine $A := (A \setminus A') \cup B'$ and $B := (B \setminus B') \cup A'$. Assume without loss of generality that after this switching $|B| \geq |A|$. We note that, despite rearranging edges, Lemma 16 still holds.

We will call a vertex $v \in A$ "heavy" if $\deg(v, A) \geq n/2 - 100\gamma n$. We define the heavy vertices in B similarly. We denote by H_A the heavy vertices of A, and H_B those of B. Note that the H_A and H_B comprise more than 97% of the vertices of A and B respectively by Lemma 16.

Preparations for the embedding – Step 2: Decomposition of T

Our next goal is to find a bipartition of the vertex set of T into two sets, X and Y such that $|X| = |A|, |Y| = |B|$ and $e(X, Y) = O(\log n)$. We begin by finding two subtrees of T, T_X and T_Y, such that each subtree contains every descendant of some vertex in T, and $n/81 \leq |T_X|, |T_Y| \leq n/27$. To find T_X and T_Y, first apply Proposition 14 yielding the two forests $T^{(1)}$ and and $T^{(2)}$, along with the split vertex u. To each *tree* $T^{(i)} \cup \{u\}$, apply Proposition 14 and keep the split vertex with the forest that does not include u, thus joining the forest into a rooted subtree of T. Repeat the procedure on the rooted subtree containing the last split vertex, keeping the new split vertex in the component that does not include the previous root. Stop when the subtree so obtained is of the prescribed size. Let the roots of T_X and T_Y be r_X and r_Y respectively.

Next we find two vertices $r_1 \in T_X$ and $r_2 \in T_Y$ such that the subtrees $T_1 := T_X(r_1)$ and $T_2 := T_Y(r_2)$ have sizes $K^{10}\gamma n \leq |T_1|, |T_2| \leq K^{11}\gamma n$. These subtrees can be found in a a manner similar to Section 2.3.3. For a leaf x of T_X, let y be the ancestor of x for which $|T_X(y)| \geq K^6\gamma n$ but if z is a child of y then $|T_X(z)| < K^6\gamma n$. Choose a leaf x for which such a y is at the maximum depth over all leaves. Then $|T_X(y)| \leq K^7\gamma n$. Set $r_1 = y$ and repeat on T_Y to find r_2.

We denote the rest of T by $T^* = T - T_X - T_Y$ Beginning with the set $S = \{r_X, r_Y\}$, recursively apply Proposition 14 to T^* a constant number of times, putting the split vertices into S such that $T^* - S$ is comprised of subtrees of size at most $\gamma^2 n$. For technical reasons, if when we add a split vertex to S there are two components of S which are at distance less than 9 from each other, we add the vertices of the path of length at most 8 to S.

We try to divide vertices of T into an (X, Y)-partition such that $|X| = |A|$, $|Y| = |B|$, and the number of edges between X and Y is $O(\log n)$. We begin by putting $V(T_X)$ and S into X and $V(T_Y)$ into Y. We then pack the components of $T^* - S$ into X and Y in such a way that that $||X| - |A||$ is minimized. This minimum is at most $\gamma^2 n$. If the error is not zero, then we split a component of $T^* - S$ by Proposition 14, updating S with the split vertex and any short paths connecting two components of S. We again divide the components of $T^* - S$ optimally, reducing the error by a constant factor. Because of the resizing of S at each step, when the error is less than, say, 10, we simply split by brute force. We can achieve an $(X, Y)-$bipartition with $|X| = |A|$ with at most $O(\log n)$ applications of Proposition 14, and so as the degree of T is bounded by a constant, the number of edges across the partition is $O(\log n)$. Set $s = |S|$.

For a component F of $T^* - S$, denote the "connection points" of F as those vertices which have a neighbor in S. As F can have at most one connection point to a single component of S, the condition that the components of S are distance at least 9 from each other ensures that the distance between any two connection points of F is at least 7.

Preparations for the embedding – Step 4: Choosing random subsets from A and B

Consider a proper two-coloring of T_1 and T_2. Denote the size of the color class containing r_1 by m_1, and the size of the color class containing r_2 by m_2. We choose $M_A \subset A$ of size m_1 randomly from H_A and M_B of size m_2 randomly from H_B. By the degree condition on T, a color class of T_i contains at least $K^9\gamma n$ vertices, so $|M_A|, |M_B| \geq K^9\gamma n$. We note that as $|H_A| > n/2 - n/50$ and $\delta(A) \geq n/4 - 8\gamma n$, any vertex $v \in A$ has at least $n/5$ heavy vertices in its neighborhood. We also note that for any $v \in A$ and any $u \in H_A$, $|N_A(v) \cap N_A(u)| \geq n/4 - 108\gamma n$. Similar remarks hold for B.

Observe that every vertex in A has at least $0.45|M_A|$ neighbors in M_A and every vertex in B has at least $0.45|M_B|$ neighbors in M_B with high probability. This follows from the fact that the minimum degree in A and B is around $|A|/2$ and $|B|/2$ respectively, and that M_A and M_B are chosen from the majority of the vertices from their respective super-sets.

Sketch of the embedding algorithm

First, we embed the forest induced on S into $A \setminus M_A$ This is easy to do greedily since S is very small. We follow by embedding $T_X - T_1$ and $T_Y - T_2$ in such a way that we cover the lowest-degree vertices. The sizes of T_X and T_Y are such that we have no problem with embedding greedily, but we have enough time to cover many low degree vertices. Then we continue with the other subtrees in arbitrary order, leaving T_1 and T_2 for last. This ensures that at each step the uncovered vertices induce a subgraph with high enough degree to continue the embedding.

At the end there will be two subtrees left unmapped, T_1 and T_2. Perhaps $K^2 s$ vertices will have be covered in M_B and a subset of size at least $K^9\gamma n$ in $A \setminus M_A$ and $B \setminus M_B$ will be left uncovered. For embedding T_1 and T_2 we will find a perfect matching in an auxiliary bipartite graph, where we use the randomness of M_A and M_B and the large degrees of the uncovered vertices.

Details of the embedding algorithm - mapping most of T

We begin by embedding S greedily into $A \setminus M_A$. Then for each vertex $x \in S$ we map the neighborhood of x into the neighborhood of $I(x)$, where vertices of $N_T(x)$ assigned to A map to $N(I(x)) \cap (H_A \setminus M_A)$ and vertices assigned to B map to $N(I(x)) \cap (B \setminus M_A)$. This is possible by the remarks on the connectivity of the heavy vertices, and as every vertex in A has $O(\log n)$ neighbors in B.

Next we embed $T_X - T_1$ into A and $T_Y - T_2$ into B. The embedding will be done almost greedily, however, we employ a technique to use the vertices with small degrees fast. First we discuss the case of $T_Y - T_2$

Lemma 17 *We can embed $T_Y - T_2$ into $B \setminus M_B$ in such a way that after the embedding there will be no uncovered vertex in B which has less than $n/2 - K^8\gamma n$ neighbors in B.*

Proof:

We provide an simple embedding algorithm that embeds at most $2K$ vertices at each iteration and covers an uncovered vertex of minimum degree. Suppose that we are in the process of embedding $T_Y - T_2$ into B and that $x \in V(T_Y - T_2)$ is a vertex which has been mapped, but for which there is a child z that is unmapped. Assume further that z is not a leaf and that x is mapped onto a heavy vertex, say $I(x) = u$. Let the children of z be $z_1, z_2, \ldots, z_k$.

Then we perform the following iteration. Let $w \in B$ be an uncovered vertex of minimum degree in B and let (u, v, w) be a path from $u = I(x)$ to w. We map $I(z) = v$ and $I(z_1) = w$. We map $z_2, z_3, \ldots, z_k$ to uncovered heavy vertices in $N_B(v)$, and we map the children of z_1 (if there are any) to uncovered heavy vertices in $N_B(w)$.

During the embedding procedure, we define a mapped vertex as “active” if at the end of an iteration it has children that are not yet mapped. Note that at the end of an iteration all active vertices are mapped onto heavy vertices. We have mapped at most $2K$ vertices and have covered a vertex of minimum degree.

Now we can describe the embedding of $T_Y - T_2$. We note that $r_Y \in S$ and has already been embedded in A and the children of r_Y, $x_1, x_2, \ldots, x_k$, are embedded into $B \setminus M_H$. Let $y_1, y_2, \ldots, y_ell$ be the children of the x_1. By the previous remark, we can map the y_i into $H_B \setminus M_B$. At this point, all active vertices are mapped onto heavy vertices. We iterate the embedding procedure breadth-first. Note that by the remarks on the connectivity of the heavy vertices and the size of $T_Y - T_2$, we can always perform the iteration as long as there is an active vertex whose children are not all leaves. When we can no longer iterate, we finish the procedure by mapping the unmapped leaves greedily. Assuming that none of the leaves are mapped onto low-degree vertices by the procedure, we are able to cover the $\frac{|T_y - T_2|/K}{2K} \geq \frac{n}{200K^2}$ lowest-degree vertices. By Lemma 16 every vertex left uncovered has degree at least $n/2 - K^8\gamma n$ for $K \geq 3$. □

As long as T_2 remains unmapped, there will be at least $K^9\gamma n$ uncovered vertices in B. By Lemma 17 the minimum degree in the graph induced on the uncovered vertices is at least $K^9\gamma n - K^8\gamma n \gg \gamma^2 n$

We embed $T_X - T_1$ just as we did $T_Y - T_2$ except that we have to avoid the vertices that have already been mapped by the vertices of S and their neighbors. This set of vertices is $O(\log n)$ and so does not cause a problem. We can assume similarly that after embedding $T_X - T_1$ but before embedding T_1, the graph induced by the uncovered vertices of A has minimum degree much bigger than $\gamma^2 n$

Then we are in a position to embed the components of $T^* - S$ into A and B according to the (X, Y)-bipartition. We describe embedding the components assigned to A, the procedure for B is identical. Note that for a component F so assigned, its connection points have already been mapped into A. Starting at an arbitrary connection point, we map greedily until we get to a vertex $x \in F$ that is at distance 2 from another connection point y. Similar to the Main Mapping Procedure, as the neighborhoods of x and y intersect in a constant percentage of M_A, we can connect x to y through M_A. Because $M_A \subset H_A$, every vertex of M_A is adjacent to at least $K^9\gamma n - 100\gamma n$ uncovered vertices. Hence, the children of the vertex in M_A used for the connection can be mapped greedily into the uncovered vertices and we use only one vertex of M_A for each connection. The distance conditions on the connection points ensures that we can continue greedily mapping F, taking care to properly establish the connections.

Finishing the embedding

Now we are in the position to finish the embedding. Only the vertices of T_1 and T_2 are left out. The embedding algorithms for these two subtrees are slightly different. We present the details of embedding T_1. Recall that the root of T_1 is denoted by r_1, and that one neighbor of r_1 is already embedded, say, onto the vertex v. Let A' be the unmapped vertices of $A \setminus M_A$.

It is easy to check that $H = (M_A, A')$ satisfies the conditions of Lemma 15 with the exception that the mapping of r_1 is restricted to $N(v) \cap M_A$. This restriction does not affect the proof of the lemma significantly, as $|N(v) \cap M_A|$ is large.

Embedding T_2 is similar, except that a set M' of $O(\log n)$ vertices of M_B have been covered already. Letting B' be the unmapped vertices of $B \setminus M_B$ and arbitrarily taking a subset $B'' \subset B'$ of size $|M'|$, by Remark 3 we can embed T_2 into the bipartite graph $H = ((M_B \setminus M') \cup B', B \setminus B')$.

2.4.2 G is close to $K_{n/2,n/2}$

This case is very similar to the previous one, so the emphasis will be on the differences. Let us sketch the embedding algorithm.

Sketch of the embedding algorithm

G has two vertex classes A and B of size $n/2$ such that the number of edges inside each class is at most γn^2. By the minimum degree condition, every vertex will have at least $K^4 C \log_{3/2} n$ neighbors in its vertex class, where C is an absolute constant.

First we prepare G for the embedding by repeatedly perform a switching procedure whenever it is possible. If we identify a vertex $v \in A$ and a vertex $u \in B$ such that $e(A, B) < e(A - v + u, B - u + v)$, then redefine $A := A - v + u$ and $B := B - u + v$. Since the number of edges in between A and B increases at every step of the switching, this is a finite process. When the process terminates, there is not a pair of vertices $v \in A, u \in B$ such that $|N(v) \cap B| \leq n/4$ and $|N(u) \cap A| \leq n/4$. Therefore, in one of the vertex classes, say A, every vertex has at least $n/4$ neighbors in B.

Next we find a decomposition of the tree T into a set of subtrees $T_1, T_2, \ldots, T_s$ such that (1) $s = O(\log n)$ and (2) there is a proper two-coloring of the subtrees such that the sum of the sizes of the red vertices is $|A|$ and the sum of the sizes of the blue vertices is $|B|$. We stress that the two-colorings are only proper within a subtree—the vertices of an edge connecting two subtrees may have the same color. We require that the largest subtree of this decomposition, T_1 has at least $n/(3K)$ vertices, but not larger than $n/9$, and that T_i has size less than $\gamma^2 n$ for $i \geq 2$. Finally, we find a subtree $T_0 \subset T_1$ of size about $K^{10}\gamma n$ such that the leaves of T_0 are that of T, and set it aside until the end of the embedding. This decomposition can be found in a manner similar to the previous case. Again, we need to make sure that the connecting points are at a sufficient distance from each other. During the decomposition process if we find two connecting points in the same component that are too close, we move one of them into the neighboring component. This may increase the number of components by at most a constant factor.

Given such a decomposition of T, we collect the edges connecting the trees $T_1, \ldots, T_s$. into a set S. The vertices of S are colored by the coloring of the decomposition. We embed S into $A \cup B$ such that whenever a vertex is red, we place it into A, and otherwise we place it into B. The size of S and the degree conditions on A and B ensure that this is easily accomplished. Every vertex in A has many neighbors in B, but it is possible that B has vertices with only a few neighbors in A. We set aside a subset $A' \subset A$ which is the union of disjoint neighborhoods of size K, one for each vertex of S in B.

Then we choose two random subsets as in the previous extremal case: $M_A \subset A - A'$ and $M_B \subset B$ among those which have at least $n/2 - 100\gamma n$ neighbors in the other vertex class. We choose $|M_A|$ to be the number of red vertices of T_0 and $|M_B|$ is the number of blue vertices of T_0.

We start the embedding process by first embedding $T_1 - T_0$, again eating up vertices with small degree into the opposite class. Afterwards, the uncovered vertices will all have large degree into the other class. Then we begin embedding $T_2, \ldots, T_s$, connecting to S through $M_A \cup M_B$. We may need at most K^2 vertices from $M_A \cup M_B$ when connecting a new subtree, overall using at most $O(\log n)$ vertices. Again, this will not cause problems for the final matching. Since a large subset of T is left to be embedded into $M_A \cup M_B$ at the end, there is always a sufficient minimum degree in the uncovered vertices to complete.

Finally, we embed T_0 into the remaining uncovered vertices by the help of a matching procedure (see Lemma 15) similar to the previous extremal case.

Chapter 3

How to avoid using the Regularity Lemma; Pósa's Conjecture revisited

3.1 Notations and definitions

$V(G)$ and $E(G)$ denote the vertex-set and the edge-set of the graph G. (A, B, E) denotes a bipartite graph $G = (V, E)$, where $V = A + B$, and $E \subset A \times B$. For a graph G and a subset U of its vertices, $G|_U$ is the restriction of G to U. $N(v)$ is the set of neighbors of $v \in V$, and $N_S(v)$ is the set of neghbors of $v \in V \cap S$. Hence the size of $N(v)$ is $|N(v)| = deg(v) = deg_G(v)$, the degree of v. $\delta(G)$ stands for the minimum, $\Delta(G)$ for the maximum and $\overline{d}(G) = \frac{1}{|V(G)|}\sum_{v \in V(G)} deg(v)$ for the average degree in G. $K_r(t)$ is the balanced complete r-partite graph with color classes of size t. We write $N(p_1, p_2, \ldots) = \cap_i N(p_i)$, the set of common neighbors. When A, B are subsets of $V(G)$, we denote by $e(A, B)$ the number of edges of G with one endpoint in A and the other in B. In particular, we write $deg(v, U) = e(\{v\}, U)$ for the number of edges from v to U. For non-empty A and B,

$$d(A, B) = \frac{e(A, B)}{|A||B|}$$

is the *density* of the graph between A and B. In particular, we write $d(A) = d(A, A) = 2|E(G|_A)|/|A|^2$.

3.2 Outline of the proof

We will follow the same rough outline (connecting-absorbing-reservoir) as in [24]; however, the main ingredient there, the Regularity Lemma, will be replaced with more elementary arguments here.

We will use the following main parameter

$$\alpha = \frac{1}{10}. \tag{3.1}$$

We assume throughout that n is sufficiently large.

Let us consider a graph G of order n with

$$\delta(G) \geq \frac{2}{3}n. \tag{3.2}$$

We must show that G contains the square of a Hamiltonian cycle.

In [17] the proof was divided into two main cases, the extremal case when G satisfies the following so-called extremal condition and the non-extremal case when this condition is not satisfied.

Extremal Condition (EC) with parameter α: *There exist (not necessarily disjoint) $A, B \subset V(G)$ such that*

- $\left(\frac{1}{3} - \alpha\right) n \leq |A|, |B| \leq \frac{1}{3}n$, *and*
- $d(A, B) < \alpha$.

In the extremal case in [17] the proof did not use the Regularity Lemma, thus we can use that part of the proof here again.

Lemma 18 (Lemma 12 in [17]). *There exists a natural number n_1 such that if a graph G has order n with $n \geq n_1$, $\delta(G) \geq \frac{2}{3}n$ and G satisfies the extremal condition EC with parameter α, then G contains the square of a Hamiltonian cycle.*

Hence we may assume that our graph G does not satisfy the extremal condition EC with parameter α. In this case our proof technique will follow the same outline (and notation) as in [24]. First in Section 3 we will prove the auxiliary Connecting Lemma that claims that any two disjoint ordered pairs of vertices can be connected by a short square-path. Then using the Connecting Lemma and the probabilistic method in Section 4 we will construct a "not too long" *absorbing* square-path P_A that will have the remarkable property that *every* "not too large" subset of vertices can be absorbed into this square-path. Thus if this P_A will be a part of a square-cycle C that contains "most" of the vertices already, then immediately P_A (and thus C) absorbs the leftover vertices and we have a Hamiltonian square-cycle. This is a significant simplification of the proof technique from [17], where the corresponding step in Section 6 was quite complicated.

Thus our goal is to construct a square-cycle C that contains the absorbing path P_A and most of the other vertices. For this purpose we will need another technical lemma in Section 5, the Reservoir Lemma, which allows us to use the Connecting Lemma (through the reservoir) even if some of the vertices are already occupied by the square-cycle we are building. Finally in the main part of the proof in Section 6 we will show that unless C contains most of the vertices already, we can extend it by using leftover vertices that are not from the reservoir or the absorbing path. This is where, in contrary to the proof in [24], we are able to achieve this goal without the use of the Regularity Lemma, but using more elementary arguments. Thus the main point of the chapter is that the proof method of [24] can be adapted into a method that avoids the use of the Regularity Lemma. We believe that this new approach (although some of the arguments are problem-specific) could be successful for other well-known extremal problems where the Regularity Lemma-Blow-up Lemma method has been used (e.g. Conjecture 3 for $k > 2$ or the main Dirac-type result of [24] itself).

3.3 Connecting

A k-square-path (or simply a k-path) in G is a sequence of vertices $\{v_1, v_2, \ldots, v_k\}$ such that $\{v_i, v_{i+1}\} \in E(G)$ for each $1 \leq i \leq k-1$ and $\{v_i, v_{i+2}\} \in E(G)$ for each $1 \leq i \leq k-2$. We say that P connects the ordered pairs (v_1, v_2) and (v_k, v_{k-1}) and these will be called the endpairs of P. Thus an endpair (a, b) is an ordered pair, a is the first (or the last) vertex on the path, and b is the second (or the second-to-last) vertex on the path. We will often call a square-path simply a path.

For two paths P and Q, let (a, b) be an endpair of P and (b, a) be an endpair of Q, and assume that $V(P) \cap V(Q) = \{a, b\}$. By $P \circ Q$ we denote the path obtained (in a unique way) as a concatenation of P and Q. We can extend this definition to more than two paths. The Connecting Lemma claims that two disjoint ordered pairs can be connected by a short path.

Lemma 19 (Connecting Lemma). *For every two disjoint ordered edges of G, (a,b) and (c,d), there is a k-path, $k \le \frac{10}{\alpha^4}$, which connects (a,b) and (c,d). Furthermore, this statement remains true even if at most $\alpha^9 n$ vertices are forbidden to be used on this connecting path.*

Proof: We will build a similar cascade structure as in the proof of the Connecting Lemma in [24]. We construct sets $A_0, A_1, A_2, \dots$ and bipartite graphs $G_1, G_2, \dots$, where $V(G_i) = A_{i-1} \cup A_i$, as follows. Let $A_0 = \{b\}$ and $A_1 = \{x \mid (a,x),(b,x) \in E(G)\}$ and let G_1 be the star with b as the center and A_1 as the set of its leaves. Note that $|A_1| \ge n/3$. Further, let

$$A_2' = \{y \mid \exists x \in A_1 \text{ such that } (b,y),(x,y) \in E(G)\}$$

and

$$G_2' = \{(x,y) \mid x \in A_1, y \in A_2', (b,y),(x,y) \in E(G)\}.$$

Then for every edge $(x,y) \in G_2'$ that is disjoint from (a,b) the vertices (a,b,x,y) form a 4-path in G. Furthermore, for each $x \in A_1$, we have $deg_{G_2'}(x) \ge n/3$. Let

$$A_2^0 = \{y \in A_2' \mid deg_{G_2'}(y) < \alpha^4 n\}, A_2 = A_2' \setminus A_2^0 \text{ and } G_2 = G_2'[A_1 \cup A_2].$$

Assume that we have constructed $A_0, A_1, \dots, A_j$ and $G_1, \dots, G_j, j \ge 2$ already. To construct A_{j+1} and G_{j+1} we do the following. First for every $y \in A_j$ we consider the auxiliary bipartite graph B_y^j between $N_{G_j}(y)$ and $V(G)$, where a pair $(x,z) \in E(B_y^j)$ for $x \in N_{G_j}(y)$, $z \in V(G)$ if $(x,z),(y,z) \in E(G)$. Define

$$A_{j+1}' = \{z \mid \exists y \in A_j \text{ such that } deg_{B_y^j}(z, N_{G_j}(y)) \ge \alpha^8 n\}$$

and

$$G_{j+1}' = \{(y,z) \mid y \in A_j, deg_{B_y^j}(z, N_{G_j}(y)) \ge \alpha^8 n\}.$$

Finally, let

$$A_{j+1}^0 = \{z \in A_{j+1}' \mid deg_{G_{j+1}'}(z) < \alpha^4 n\},$$

$$A_{j+1} = A_{j+1}' \setminus A_{j+1}^0 \text{ and } G_{j+1} = G_{j+1}'[A_j \cup A_{j+1}].$$

We call the entire structure $A_0, A_1, A_2, \ldots$ along with the bipartite graphs $G_1, G_2, \ldots$ an (a,b)-cascade. Notice that some of the sets A_j may intersect. For the sake of the construction we treat them as disjoint. Note also that we had to change the construction slightly for $j \geq 3$ and require $deg_{B_y^j}(z) \geq \alpha^8 n$ to make sure that we can always return from any edge of G_j back to (a,b) by a legitimate square-path on which all the vertices are distinct, even if at most $\alpha^9 n$ vertices are forbidden.

A vertex $y \in A_j$ is called *heavy* if $deg_{G_j}(y) \geq (1/3 + \alpha^4)n$.

Claim 20. *There exists an index $j \leq j_0 = \lceil \frac{4}{\alpha^4} \rceil + 2$ such that A_j contains at least $\alpha^4 n$ heavy vertices.*

For the proof of this claim, first we prove that for every $j \geq 2$ and for every $y \in A_j$ we have

$$deg_{G'_{j+1}}(y) \geq (1/3 - \alpha^4)n. \tag{3.3}$$

Indeed, let s be the number of vertices $z \in V(G)$ with $deg_{B_y^j}(z) < \alpha^8 n$. Then

$$s\alpha^8 n + (n-s)|N_{G_j}(y)| \geq |E(B_y^j)| \geq |N_{G_j}(y)|n/3.$$

From this using $|N_{G_j}(y)| = deg_{G_j}(y) \geq \alpha^4 n$ and $s \leq n$, we get

$$n - s \geq n/3 - \frac{s\alpha^8 n}{|N_{G_j}(y)|} \geq n/3 - \alpha^4 n,$$

proving (3.3). Note also that the total number of edges of G'_{j+1} incident to the exceptional vertices in A^0_{j+1} is smaller than $\alpha^4 n^2$.

Let us look at the sequence of sets $A_1, A_2, \ldots$, where we have $|A_1| \geq n/3$. Clearly we must have a $j \leq \lceil \frac{4}{\alpha^4} \rceil$ for which

$$|A_{j+1}|, |A_{j+2}| \leq (1+\alpha^4)|A_j|. \tag{3.4}$$

Indeed, if $j = 1$ does not satisfy (3.4), then either A_2 or A_3 (say A_3) has size at least $(1+\alpha^4)|A_1| \geq (1+\alpha^4)n/3$. If $j = 3$ does not satisfy (3.4), then either A_4 or A_5 (say A_5) has size at least $(1+\alpha^4)|A_3| \geq (1+\alpha^4)n/3$. Continuing in this fashion, in each step we add at least $\alpha^4 n/3$ new vertices to A_1, so in at most $\lceil \frac{2n/3}{\alpha^4 n/3} \rceil = \lceil \frac{2}{\alpha^4} \rceil$ steps we get a set A_j with more than n vertices, a contradiction.

Furthermore, we may assume that for this j in addition to (3.4) the following holds as well

$$|A_{j+1}| \geq (1-\alpha^3)|A_j|. \tag{3.5}$$

Otherwise A_{j+1} would contain at least $\alpha^4 n$ heavy vertices and we would be finished with the proof of Claim 20. Indeed, suppose not. On one hand from the above we have

$$|E(G_{j+1})| \geq |A_j|n/3 - 2\alpha^4 n^2, \tag{3.6}$$

but on the other hand using (3.5) we would get

$$|E(G_{j+1})| \leq \alpha^4 n|A_j| + |A_{j+1}|(1/3+\alpha^4)n \leq$$

$$\leq \alpha^4 n|A_j| + (1-\alpha^3)(1/3+\alpha^4)|A_j|n \leq |A_j|n/3 - \alpha^3|A_j|n/3 + 2\alpha^4|A_j|n,$$

a contradiction (using (3.1)).

Thus we may assume that there is a $j \leq \lceil \frac{4}{\alpha^4} \rceil$ for which both (3.4) and (3.5) hold. We fix this j. We will show that A_{j+2} contains at least $\alpha^4 n$ heavy vertices as desired in the claim. For this purpose first we show that A_{j+1} contains at least $\alpha n/2$ vertices z for which

$$deg_{G_{j+1}}(z) \geq (1/3-\alpha)n. \tag{3.7}$$

Otherwise, similarly as above using (3.4) we would get

$$|E(G_{j+1})| \leq \alpha n|A_j|/2 + |A_{j+1}|(1/3-\alpha)n \leq$$

$$\leq \alpha n|A_j|/2 + (1+\alpha^4)(1/3-\alpha)|A_j|n \leq |A_j|n/3 - \alpha|A_j|n/2 + \alpha^4|A_j|n/3,$$

a contradiction with (3.6) (using (3.1)).

Consider a vertex $z \in A_{j+1}$ satisfying (3.7). Next we show that

$$deg_{G'_{j+2}}(z) \geq (1/3+\alpha)n. \tag{3.8}$$

Indeed, otherwise let $A \subset N_{G_{j+1}}(z)$, $B \subset N_G(z) \setminus N_{G'_{j+2}}(z)$ be arbitrary subsets with sizes $(1/3-\alpha)n \leq |A|, |B| \leq n/3$ (this is possible as both of these sets have size at least $(1/3-\alpha)n$). Since G does not satisfy EC with parameter α, we have $d(A,B) \geq \alpha$. In particular, we can pick a vertex $w \in B$ with $deg_G(w, A) \geq \alpha|A| \gg \alpha^8 n$, a contradiction, since in this case w would belong to $N_{G'_{j+2}}(z)$ by definition.

Thus we get from (3.3) and (3.8)

$$|E(G_{j+2})| \geq |A_{j+1}|n/3 + \alpha^2 n^2/2 - 2\alpha^4 n^2. \tag{3.9}$$

However, this implies that in A_{j+2} we must have at least $\alpha^4 n$ heavy vertices, and thus proving the claim. Indeed, otherwise using (3.1), (3.4) and (3.5) we would get

$$|E(G_{j+2})| \leq \alpha^4 n|A_{j+1}| + |A_{j+2}|(1/3 + \alpha^4)n \leq$$

$$\leq \alpha^4 n|A_{j+1}| + (1 + 2\alpha^3)(1 + \alpha^4)(1/3 + \alpha^4)|A_{j+1}|n \leq |A_{j+1}|n/3 + \alpha^3|A_{j+1}|n,$$

a contradiction with (3.9).

Now to finish the proof of the Connecting Lemma, given two disjoint ordered edges of G, (a, b) and (c, d), we consider the (a, b)-cascade $(A_j^{(1)}, G_j^{(1)})$ and the (c, d)-cascade $(A_j^{(2)}, G_j^{(2)})$. For $i = 1, 2$, let $A_{j(i)}^{(i)}$ be the set that contains many ($\geq \alpha^4 n$) heavy vertices as guaranteed by Claim 20. An easy averaging argument shows that there must be many ($\geq \alpha n$) vertices $u \in V(G)$ such that u has many ($\geq \alpha^5 n$) heavy neighbors in both $A_{j(i)}^{(i)}$, $i = 1, 2$. Consider one such a u, a heavy neighbor $h^{(1)}$ of u in $A_{j(1)}^{(1)}$ and a heavy neighbor $h^{(2)}$ of u in $A_{j(2)}^{(2)}$. It is easy to see from the definition that we have

$$deg_{G'_{j(1)+1}}(h^{(1)}), deg_{G'_{j(2)+1}}(h^{(2)}) \geq 2n/3,$$

since $h^{(1)}$ and $h^{(2)}$ are heavy vertices.

Let $A \subset N_G(u) \cap N_{G'_{j(1)+1}}(h^{(1)})$, $B \subset N_G(u) \cap N_{G'_{j(2)+1}}(h^{(2)})$ be arbitrary subsets with sizes $\lfloor n/3 \rfloor$ (this is possible as both of these sets have size at least $n/3$). Since G does not satisfy EC with parameter α, we have $d(A, B) \geq \alpha$. In particular, we can pick an edge $(v^{(1)}, v^{(2)})$ with $v^{(1)} \in A$ and $v^{(2)} \in B$. By the definition of the (a, b)-cascade, there is a $(j(1) + 3)$-path $P^{(1)}$ connecting (a, b) and $(v^{(1)}, h^{(1)})$ and by the definition of the (c, d)-cascade, there is a $(j(2) + 3)$-path $P^{(2)}$ connecting (c, d) and $(v^{(2)}, h^{(2)})$. By putting together $P^{(1)}$ and $P^{(2)}$ and including u in the middle we get a k-path connecting (a, b) and (c, d) with

$$k = (j(1) + 3) + (j(2) + 3) + 2 \leq 2(j_0 + 4) \leq \frac{10}{\alpha^4}.$$

Furthermore, the condition $deg_{B_y^j}(z) \geq \alpha^8 n$ guarantees that the proof goes through (and we can find a connecting path) even if we have a set of at most $\alpha^9 n$ forbidden vertices. $\square$

3.4 Absorbing

Again we follow the method of [24], we just have to adapt the ideas to square-cycles in graphs instead of tight cycles in hypergraphs. For the sake of completeness we give the details here again. As we sketched above the Absorbing Lemma claims that we can construct a “not too long” *absorbing* path A that absorbs *every* “not too large” subset of vertices.

Lemma 21 (Absorbing Lemma). *There is an l-path P_A in G with $l \leq \alpha^9 n$, such that for every subset $U \subset V(G) \setminus V(P_A)$ of size at most $\alpha^{20} n$ there is a path P_{A_U} in G with $V(P_{A_U}) = V(P_A) \cup U$ and such that P_{A_U} has the same endpairs as P_A.*

Proof: Given a vertex $v \in V(G)$ we say that an ordered 5-tuple of vertices (x, a, b, c, d) absorbs v if these 5 vertices are all neighbors of v, the vertices $\{a, b, c, d\}$ are all neighbors of x and (a, b, c, d) is a (simple) path in G. Indeed, in this case the (square-)path (a, b, x, c, d) *may* absorb v to get the extended path (a, b, v, x, c, d). Note that both paths have the same endpairs. First we show that for every $v \in V(G)$ there are many 5-tuples absorbing v.

Claim 22. *For every $v \in V(G)$ there are at least $\frac{1}{2(6^4)}\alpha^4 n^5$ 5-tuples absorbing v.*

Indeed, let us consider an arbitrary $v \in V(G)$. We can choose x as an arbitrary neighbor of v. Since G satisfies (3.2) we can choose x in at least $\frac{2}{3}n$ different ways. Consider the common neighbors of v and x, $N(v, x)$. We know from (3.2) that $|N(v, x)| \geq \frac{n}{3}$. Keep a subset $N \subseteq N(v, x)$ with $|N| = \lfloor \frac{n}{3} \rfloor$. Since G does not satisfy the extremal condition EC with parameter α, we know that $d(N) = d(N, N) \geq \alpha$. Then we have $\overline{d}(G|_N) \geq \alpha|N|$, and so we can choose a subgraph H of $G|_N$ with $\delta(H) > \frac{\alpha}{2}|N|$. In particular, we also have $|V(H)| \geq \frac{\alpha}{2}|N|$. Let a be an arbitrary vertex of H (at least $\frac{\alpha}{2}|N|$ different choices), let b be an arbitrary neighbor of a in H (at least $\frac{\alpha}{2}|N|$ different choices), let c be an arbitrary neighbor of b in H that is different from a (at least $\frac{\alpha}{2}|N|-1$ different choices), and finally let d be an arbitrary neighbor of c in H that is different from a and b (at least $\frac{\alpha}{2}|N|-2$ different choices). Then (x,a,b,c,d) is a good 5-tuple that absorbs v. The number of ways we can select (x,a,b,c,d) from the above is at least

$$\frac{2}{3}n\frac{\alpha}{2}|N|\frac{\alpha}{2}|N|\left(\frac{\alpha}{2}|N|-1\right)\left(\frac{\alpha}{2}N-2\right)\geq\frac{1}{2(6^4)}\alpha^4n^5$$

(for sufficiently large n), finishing the proof of the claim.

For each $v\in V(G)$, let $\mathcal{A}_v$ be the family of all 5-tuples absorbing v. The next claim can be proved by an application of the probabilistic method.

Claim 23. *There exists a family $\mathcal{F}$ of at most $2\alpha^{14}n$ disjoint, absorbing 5-tuples of vertices of G such that for every $v\in V(G)$ we have $|\mathcal{A}_v\cap\mathcal{F}|>\alpha^{20}n$.*

For this purpose we first select a family $\mathcal{F}'$ of 5-tuples at random by including each of $n(n-1)(n-2)(n-3)(n-4)\sim n^5$ of them independently with probability $\alpha^{14}n^{-4}$ (some of the selected 5-tuples may not be absorbing at all). Using Chernoff's inequality (see, e.g. [15]) with probability $1-o(1)$, as $n\to\infty$, we have

- $|\mathcal{F}'|<2\alpha^{14}n$,
- for each $v\in V(G)$, $|\mathcal{A}_v\cap\mathcal{F}'|\geq\frac{1}{3(6^4)}\alpha^{18}n$.

Furthermore, the expected number of intersecting pairs of 5-tuples in $\mathcal{F}'$ is at most

$$n^5\times 5\times 5\times n^4\times(\alpha^{14}n^{-4})^2=25\alpha^{28}n,$$

and thus, by Markov's inequality, with probability at least 1/26,

- there are at most $26\alpha^{28}n$ pairs of intersecting 5-tuples in $\mathcal{F}'$.

Thus with positive probability, a random family $\mathcal{F}'$ satisfies all the three properties above. Thus there exists one such a family, for simplicity, we also denote this family by $\mathcal{F}'$. From $\mathcal{F}'$ we delete all 5-tuples that intersect other 5-tuples and all 5-tuples that are not absorbing at all. Let us denote by $\mathcal{F}$ the remaining subfamily. Then $\mathcal{F}$ consists of disjoint, absorbing 5-tuples such that for each $v\in V(G)$ we have using (3.1)

$$|\mathcal{A}_v\cap\mathcal{F}|\geq\frac{1}{3(6^4)}\alpha^{18}n-52\alpha^{28}n>\frac{1}{4(6^4)}\alpha^{18}n>\alpha^{20}n,$$

proving Claim 23.

Let $f = |\mathcal{F}|$, let $F_1, \ldots, F_f$ be the 5-tuples in $\mathcal{F}$ and let $F = \cup_{i=1}^{f} F_i$. Since for each $i = 1, \ldots, f$, F_i is absorbing for at least one vertex $v \in V(G)$, F_i spans a 5-path. Our next task is to connect all these 5-paths into one, not too long absorbing path P_A. For this purpose, we will apply the Connecting Lemma (Lemma 19) repeatedly, and for each $i = 1, \ldots, f-1$ we will connect the endpairs of F_i and F_{i+1} by a short path. Thus we get the following claim.

Claim 24. *There exists a path P_A in G of the form*

$$P_A = F_1 \circ C_1 \circ \ldots \circ F_{f-1} \circ C_{f-1} \circ F_f,$$

where the paths $C_1, \ldots, C_{f-1}$ each have at most $\frac{10}{\alpha^4}$ vertices.

Indeed, we apply Lemma 19 to connect F_1 and F_2, we apply Lemma 19 again to connect F_2 and F_3, etc. finally we apply Lemma 19 to connect F_{f-1} and F_f. Note that Lemma 19 can always be applied as the set of forbidden vertices (vertices on the part of P_A that is constructed already and vertices in F) has size at most

$$f\left(\frac{10}{\alpha^4} + 5\right) \leq 2\alpha^{14} \frac{11}{\alpha^4} n \leq \alpha^9 n.$$

Thus we connected all paths in $\mathcal{F}$ into one path of length at most $\alpha^9 n$. It remains to show that P_A has the absorbing property. Let $U \subset V \setminus V(P_A)$ of size at most $\alpha^{20} n$. Since for every $v \in U$ we have $|\mathcal{A}_v \cap \mathcal{F}| > \alpha^{20} n$, we can insert all vertices of U into P_A one by one, each time using a new absorbing 5-tuple.

3.5 The reservoir

In the Reservoir Lemma we will set aside some vertices that we can always use for connecting even if the other vertices are occupied already.

Lemma 25 (Reservoir Lemma). *For every subset $W \subset V(G)$, $|W| \leq \alpha^9 n$, there exists a subset $R \subset V(G) \setminus W$ (a reservoir) such that $|R| = \lfloor \alpha^{20} n/2 \rfloor$ and*

$$deg_G(x, R) \geq (2/3 - \alpha^{10})|R| \text{ for every } x \in V(G). \tag{3.10}$$

Proof: Set $r = \lfloor \alpha^{20} n/2 \rfloor$. We choose R randomly out of all $\binom{n-|W|}{r}$ possibilities and apply the probabilistic method again. By Chernoff's bound again, for sufficiently large n, (3.10) will be true for R with high probability. Then we can fix a choice of R for which (3.10) is true. □

Then indeed, we can connect through the reservoir.

Lemma 26 (Reservoir-Connecting Lemma). *For every two disjoint ordered edges of G, (a, b) and (c, d), there is a k-path in $R \cup \{a, b, c, d\}$, $k \leq \frac{10}{\alpha^4}$, which connects (a, b) and (c, d). Furthermore, this statement remains true even if at most $\alpha^9|R|$ vertices of R are forbidden to be used on this connecting path.*

Proof: Indeed, since by (3.10) inside R we have almost the same degree condition as in G, the proof of the Connecting Lemma goes through inside R, the slight loss in the minimum degree is not going to create any problems. Note also that we may assume that $G|_R$ does not satisfy the EC with parameter α as this is true with high probability. □

3.6 The proof of Theorem 1

We start with the outline of the proof.

Step 1: By applying the Absorbing Lemma (Lemma 21), we find an absorbing path P_A with $|P_A| \leq \alpha^9 n$.

Step 2: By applying the Reservoir Lemma (Lemma 25), we set aside a reservoir $R \subset V(G) \setminus V(P_A)$ with $|R| = \lfloor \alpha^{20} n/2 \rfloor$.

Step 3: We find a (square-)cycle C in G that contains P_A as a subpath, all but at most $\alpha^{20} n/2$ vertices of $V(G) \setminus (V(P_A) \cup R)$ (denote the set of these missing vertices by T) and some vertices of R (denote the set of remaining vertices in R by R'). Note that $|R' \cup T| \leq \alpha^{20} n$.

Step 4: Using the absorbing property of P_A, insert $R' \cup T$ into C, resulting in a Hamiltonian cycle of G.

It remains to explain Step 3 in the outline above. The rest of the chapter contains the construction of this C. We start with an arbitrary path P in G that starts with

P_A as a subpath. Then we will gradually extend this P (sometimes with the use of the Reservoir-Connecting Lemma, so using vertices from the reservoir R) until it contains all but at most $\alpha^{20}n/2$ vertices of $V(G) \setminus (V(A) \cup R)$. We connect the two endpairs of P through the reservoir by applying the Reservoir-Connecting Lemma one more time to get the cycle C that is desired in Step 3. Thus we only have to show how to extend P until it contains all but at most $\alpha^{20}n/2$ vertices of $V(G) \setminus (V(P_A) \cup R)$.

Denote by P' the square-path without the absorbing path, $V(P') = V(P) \setminus V(P_A)$, and set $m := |V(P')|$. Throughout the chapter, we will represent the neighborhood on P' of a vertex $a \in T$ by a bitstring of length $|P'|$, indexed by the vertices of P' in their order along the path. For $a \in T$, denote this bitstring by I_a, and write $I_a(S)$ for the substring on the vertices $S \subset P$, retaining the original order. In the case that $S = \{v\}$ we will write simply $I_a(v)$. For $v \in P'$, $I_a(v)$ is a one iff $v \in N_{P'}(a)$.

A general observation is that for any $a \in T$, there can be no run of ones longer than 3 in I_a, otherwise we could easily extend P by inserting a between the vertices of any run of length 4. Call a zero followed by a maximal run of ones a 3-, 2-, or 1-block, depending on the length of the maximal run. Call a zero that is followed by another zero a 0-block. Thus I_a is comprised of disjoint 3-, 2-, 1-, and 0-blocks. We note that only the 3-blocks have a density of ones that is greater than 2/3.

For any given $a \in T$, we will often make use of a partition of I_a into substrings (and thus, a partition of P' into subpaths) according to the 3-blocks. We denote the substrings, which we refer to as intervals, by $I_a^0, I_a^1, I_a^2 \ldots, I_a^l$. The interval I_a^j is defined to begin *after* the j^{th} 3-block and end with the $(j+1)^{th}$ 3-block. Of course, I_a^l may not end with a 3-block. Our first case is when T is large compared to the absorbing path.

3.6.1 T is larger than $\alpha^8 n$

An interval I_a^j comprised of a (possibly empty) run of 2-blocks followed by a 3-block is called a "heavy" interval. Note that only the heavy intervals have a density of ones greater than 2/3.

We begin by defining an operation, `HEAVY SWAP`, which exchanges vertices of T with vertices of P' in such a way as to extend P'. We will identify the conditions necessary for the operation to take place.

1. There exists a vertex $x_1 \in P'$ such that $H = \{a \in T \mid a$ has a heavy interval beginning at $x_1\}$ is nonempty

2. The minimum length of the heavy intervals beginning at x_1 is less than $3|H| - 2$.

With these conditions in place, we define the operation. Let $a \in H$ be a vertex whose heavy interval beginning at x_1 is of minimum length, say $3k + 1$. Define the subpath $Q \subset P'$ of length $3k + 4$ comprised of the 3 vertices preceeding x_1 and the vertices along the heavy interval in I_a,

$$Q = (o_1, o_2, o_3, x_1, o_4, o_5, x_2, o_6, o_7, \ldots, x_{k-1}, o_{2k}, o_{2k+1}, x_k, o_{2k+2}, o_{2k+3}, o_{2k+4}).$$

We have $o_i \in N_{P'}(a)$ and $x_i \notin N_{P'}(a)$, and the substring

$$I_a(Q) = (1, 1, 1, 0, 1, 1, 0, 1, 1, \ldots, 0, 1, 1, 0, 1, 1, 1).$$

In fact, by the minimality of a, for every $b \in H$,

$$I_b(Q \setminus \{o_{2k+4}\}) = I_a(Q \setminus \{o_{2k+4}\}).$$

The conditions ensure that $|H| \geq k + 1$, and so we can find k vertices $b_1, b_2, \ldots, b_k$ from $H \setminus \{a\}$. The path

$$Q' = (o_1, o_2, b_1, o_3, o_4, b_2, o_5, o_6, b_3, o_7, \ldots, o_{2k}, b_k, o_{2k+1}, o_{2k+2}, a, o_{2k+3}, o_{2k+4})$$

is a legitimate square path, with which we replace Q in P'. This defines the operation, which extends P' by one.

Claim 27. *If there exists a subset $H_0 \subset T$ of size $n^{3/4}$ such that for all $a \in H_0$, I_a contains at least $3n^{3/4}$ heavy intervals, then we can extend P.*

We will demonstrate that the conditions necessary for `HEAVY SWAP` are satisfied. Call a heavy interval "short" if it is of length less than $n^{1/4}$. Then for each $a \in H_0$, there are at least $2n^{3/4}$ *short* heavy intervals in I_a. Indeed, otherwise we get for the size of the union of the long heavy intervals strictly more than $n^{3/4}n^{1/4} = n$ vertices, a contradiction. By the pigeonhole principle, there is a vertex $x_1 \in P'$ where at least

$$\frac{2n^{3/4}n^{3/4}}{m} \geq 2n^{1/2}$$

short heavy intervals begin. Let $H \subset H_0$ be those vertices which all have a short heavy interval beginning at the vertex x_1. As $|H| \geq 2n^{1/2}$ and for every $a \in H$ the length of the heavy interval of I_a beginning at x_1 is less than $n^{1/4} \ll 2n^{1/2}$, we can perform the operation `HEAVY SWAP`. □

A simple calculation shows that if $deg(a, P') \geq \frac{2}{3}m + n^{3/4}$, then I_a contains at least $3n^{3/4}$ heavy intervals. Otherwise, with s the size of the union of the heavy intervals, recalling the observation on the density of the heavy intervials,

$$deg(a, P') < \frac{2}{3}(s - 3n^{3/4}) + 3n^{3/4} + \frac{2}{3}(m - s) = \frac{2}{3}m + n^{3/4},$$

a contradiction. Setting $T' = \{a \in T \mid deg(a, P') \geq \frac{2}{3}m + n^{3/4}\}$, assuming that the premise of Claim 27 fails, it follows that $|T'| \leq n^{3/4}$. Letting $|T| = t$, we have in this case that for every $a \in T \setminus T'$,

$$deg_T(a) \geq \frac{1}{2}t + \frac{1}{2}(\alpha t + n^{3/4}). \tag{3.11}$$

Indeed, for $a \in T \setminus T'$, using $deg(a, P') \leq \frac{2}{3}m + n^{3/4}$, $|R| \leq \alpha^{20}n$, $|P| \leq \alpha^9 n$,

$$\begin{aligned} deg_T(a) &\geq \frac{2}{3}n - (\frac{2}{3}m + n^{2/4}) - \alpha^9 n - \alpha^{20} n \geq \frac{2}{3}t - \frac{\alpha^9 n + \alpha^{20} n}{3} - n^{3/4} \\ &\geq \frac{1}{2}t + \frac{\alpha t + n^{3/4}}{2} \end{aligned}$$

if $t \geq \frac{1}{1-\alpha}(2\alpha^9 n + 2\alpha^{20} n + 7n^{3/4})$, which is true for large enough n when $t \geq \alpha^8 n$.

With degree condition (3.11), we are guaranteed to find a square-path in $T \setminus T'$ of length at least αt. Indeed, as any two vertices in $T \setminus T'$ have degree in T of $\frac{1}{2}t + \frac{1}{2}(\alpha t + n^{3/4})$, they have a common neighborhood in T of size at least $\alpha t + n^{3/4}$. Hence, the greedy algorithm is guaranteed to be able to extend any square path of length less than αt in T by a vertex not in T' and not on the path being extended. In this case, we can extend P by connecting a square path of length αt through the reservoir.

Thus, we may assume that T is close to the size of the absorbing path.

3.6.2 T is smaller than $\alpha^8 n$

The outline of the proof in this case is as follows: assuming that the premise of Claim 27 does not hold, we find a large matching in P' that can be moved out of the path by exchanging with vertices of T without disturbing P. Then we can either extend P or there is a large set of disjoint triangles in P' that can be moved out without disturbing P. If we still are unable to extend P then we can find a long square path in P' which we move out and then connect through Lemma 25.

We call an interval I_a^j "even" if it contains no 0-blocks and exactly one 1-block. Note that the even intervals have a density of ones exactly $2/3$.

As in the case of the heavy intervals, we will define the operation `EVEN SWAP` that exchanges vertices of T with vertices of P', taking advantage of a certain alignment of even intervals. In this case we will not be able to extend P', but rather we will identify vertices of P' and vertices of T which can be exchanged for the purpose of guaranteeing a certain structure in T. We first identify the conditions necessary to perform the operation `EVEN SWAP`:

1. There exists a vertex $x_1 \in P'$ such that $D = \{a \in T \mid a$ has an even interval whose 1-block begins at $x_1\}$ is non-empty
2. The minimum length of the even intervals whose 1-blocks begin at x_1 is less than $3|D|$.

With these conditions in place, we define the operation. Let $a \in D$ be a vertex whose even interval aligned with x_1 is of minimum length, say $3k$. Define the subpath $Q \subset P'$ of length $3k+2$ comprised of the 2 vertices preceding x_1 and the vertices of the even interval of I_a containing the position of x_1,

$$Q = (o_1, o_2, x_1, o_3, x_2, o_4, o_5, \ldots, x_{k-1}, o_{2k-2}, o_{2k-1}, x_k, o_{2k}, o_{2k+1}, o_{2k+2}).$$

We have $o_i \in N_{P'}(a)$ and $x_i \notin N_{P'}(a)$, and the substring

$$I_a(Q) = (1, 1, 0, 1, 0, 1, 1, \ldots, 0, 1, 1, 0, 1, 1, 1).$$

As $|D| \geq k$ we can find distinct vertices $b_1, b_2, \ldots, b_{k-1}$ from $D \setminus \{a\}$, and the path

$$Q' = (o_1, o_2, b_1, o_3, o_4, b_2, o_5, \ldots, o_{2k-2}, b_{k-1}, o_{2k-1}, o_{2k}, a, o_{2k+1}, o_{2k+2})$$

is a legitimate subpath of P' of length $3k+2$ with which Q can be replaced. This defines the operation.

For $a \in T$ and I_a^j an even interval for a, referring to the notation defined above, we consider x_1 and x_2 to be "swappable" with a via the operation `EVEN SWAP`. Unfortunately, in order to bring either x_1 or x_2 from the path into T, we are forced to swap every x_i. For this reason, we will only consider the zero vertices surrounding the 1-block of an even interval swappable if the interval is of length less than $\frac{1}{\alpha^2}$. We define a "short" even interval to be one whose length is less than $\frac{1}{\alpha^2}$.

There is one other class of vertex which is swappable with a. Let x be any vertex whose position is a zero in I_a. If the two positions preceding x and the two positions succeeding x are all ones in I_a, then a and x can be exchanged. More precisely, if the subpath

$$Q = (o_1, o_2, x, o_3, o_4)$$

is such that $o_i \in N_{P'}(a)$ and $x \notin N_{P'}(a)$, then

$$Q' = (o_1, o_2, a, o_3, o_4)$$

is a legitimate subpath of P'.

We are now prepared to define S_a, the set of swappable vertices for a. For $a \in T$, we define $v \in S_a$ iff $I_a(v) = 0$ and v is either

(a) preceded and succeeded by two ones in I_a, or

(b) the zero of a 1-block in a short even interval for I_a, or

(c) the zero following a 1-block in a short even interval for I_a.

We observe that in the case of small T, since $|R| \leq \alpha^{20}n$, $|P| \leq \alpha^9 n$, $|T| \leq \alpha^8 n$, for every $a \in T$,

$$deg(a, P') \geq \frac{2}{3}m - 2\alpha^8 n.$$

Again, define $H_0 = \{a \in T \mid I_a$ contains at least $3n^{3/4}$ heavy intervals$\}$. Claim 27 holds regardless of the size of T, and so we may assume $|H_0| \leq n^{3/4}$. It follows by the remark following Claim 27 that for $a \in T \setminus H_0$, $deg(a, P') \leq \frac{2}{3}m + n^{3/4}$.

For $a \in T \setminus H_0$, we provide a lower bound on the size of S_a, the main consequence of which is that S_a is α-dense for every $a \in T \setminus H_0$.

Claim 28. *For $a \in T \setminus H_0$, $|S_a| > (\frac{1}{3} - 3\alpha^2)n$*

By the last two remarks, as $a \in T \setminus H_0$,

$$\frac{2}{3}n - 2\alpha^8 n \leq deg(a, P') \leq \frac{2}{3}m + n^{3/4}. \tag{3.12}$$

Define a counting function on bitstrings as follows: to each zero assign a value of -2 and to each one a $+1$ and sum the values over the length of the bitstring. Thus a bitstring with a density of ones exactly 2/3 has a count of 0. By (3.12) the number of ones in I_a is at least $(\frac{2}{3} - 2\alpha^8)n$, and thus the number of zeros is at most $m - (\frac{2}{3} - 2\alpha^8)n$. Hence, the count for I_a is at least

$$\frac{2}{3}n - 2\alpha^8 n - 2(m - \frac{2}{3}n + 2\alpha^8 n) = 2n - 2m - 6\alpha^8 n \geq -6\alpha^8 n.$$

Consider the intervals I_a^j. At most $3n^{3/4}$ are heavy, each with a count of $+1$, and thus the total contribution to the count from the heavy intervals is at most $3n^{3/4}$. The rest of the I_a^j include at least a 1-block or a 0-block. For each I_a^j containing at least one 1-block, distinguish an arbitrary 1-block of I_a^j, and denote by b_1^i the number of undistinguished 1-blocks.. For any I_a^j, denote by b_0^i the number of 0-blocks in I_a^j, if any. If I_a^j contains at least one 1-block, the count on I_a^j is $-b_1^j - 2b_0^j$. If I_a^j contains only 0-blocks, the count is $-2b_0^j + 1 < b_0^j$. Let $b_0 = \sum_j b_0^j$ and $b_1 = \sum_j b_1^j$. Summing over the intervals, the count on I_a is at most $-b_1 - b_0 + 3n^{3/4}$. It follows that $b_0 + b_1 \leq 6\alpha^8 n + 3n^{3/4} \leq 7\alpha^8 n$.

As $m \geq n - 2\alpha^8 n$, using (3.12), the number of zeros in I_a is at least

$$m - (\frac{2}{3}n + n^{3/4}) \geq \frac{1}{3}n - 2\alpha^8 n - n^{3/4}.$$

For each 0-block or undistinguished 1-block in the interval I_a^j, the zero belonging to the block, the zero following the block, and the 2 zeros surrounding the distinguished 1-block (if it exists), are not in positions corresponding to vertices of S_a. There are at most $2\alpha^2 n$ zeros surrounding the 1-blocks of long even intervals. The rest of the zeros correspond to positions of vertices in S_a. Thus, the total number of vertices of P' in S_a is at least

$$\frac{1}{3}n - 2\alpha^8 n - n^{3/4} - 4(b_0 + b_1) - 2\alpha^2 n \geq \frac{1}{3}n - 3\alpha^2 n,$$

proving the claim. □

Let $|T \setminus H_0| = t$. From Claim 28, an easy calculation shows that there is a set $S \subset \bigcup_a S_a \subset P'$ such that $|S| \geq (\frac{1}{3} - 4\alpha^2)n$ and $\forall v \in S$, we have $v \in S_a$ for at least $\alpha^2 t$ vertices $a \in T \setminus H_0$.

Let $S' \subset S$ be such that

1. $|S'| \leq \alpha^4 t$, and

2. for any $u, v \in S'$, the distance from u to v along the path P' is at least $\frac{1}{\alpha^2}$.

Then we can move S' from the path to T by exchanging the vertices of S' with vertices of $T \setminus H_0$. To see this, let $S' \subset S$ be as described. For any $v \in S'$, and for each a for which $v \in S_a$, it is either of type **(a)**, **(b)**, or **(c)** by definition of S_a. By the pigeonhole principle there is a set of at least $\alpha^2 t/3$ vertices $a \in T \setminus H_0$ for which v is of the same type in S_a. If this is type **(a)** then v can be exchanged directly. If it is type **(b)** or **(c)**, then we have short even intervals for at least $\alpha^2 t/3 \gg 1/\alpha^2$ vertices $a \in T \setminus H_0$, all of whose whose 1-blocks are aligned. The conditions for `EVEN SWAP` are satisfied and we can exchange v with a vertex of $T \setminus H_0$, but we may have to perform as many as $1/3\alpha^2$ exchanges of other vertices in the short even interval. The distance condition on $u, v \in S'$ precludes the possibility that exchanging vertices within an interval of length $\frac{1}{\alpha^2}$ in order to move u from P' to T destroys the conditions necessary to move v out–it ensures that exchanging u does not diminish our ability to exchange v apart from simply using up vertices of T. From these observations, each exchange from S' to T may use up to $\frac{1}{3\alpha^2}$ vertices of T. When trying to exchange $u \in S'$, as long as there are $\frac{1}{\alpha^2}$ vertices $a \in T$ for which $u \in S_a$, we are able to perform the operation. With fewer than

$$\frac{\alpha^2 t/3}{1/3\alpha^2} = \alpha^4 t$$

exchanges, this condition is guaranteed.

By the extremal condition, as $|S| \geq (\frac{1}{3} - 4\alpha^2)n$, S has density α. We can easily find a matching M with $\alpha^4 t/2$ edges such that every two vertices of $V(M)$ are separated by constant distance on P'. By the previous remark, we can exchange these vertices of M without disturbing P'. Therefore, we assume that $V(M) \subset T$. We define H_0 as before, and let M' be the set of edges of M that are disjoint from H_0. Then we have $|M'| \geq \frac{\alpha^4 t}{2} - n^{3/4} \geq \frac{\alpha^4 t}{3}$.

For $(a, b) \in M'$, define the "overlap" of (a, b) to be $S_{a,b} = S_a \cap S_b$. We have the following claim:

Claim 29. *If $|S_{a,b}| > n^{1/2}$ for at least $3n^{1/2}$ edges $(a, b) \in M'$, then we can extend P.*

For any edge $(a, b) \in M$ and $x \in S_{a,b}$, we classify x as being of one of two types with respect to (a, b). If the two vertices following x on P' are both in $N_{P'}(a)$ and the two vertices preceding x are both in $N_{P'}(b)$, (or vice versa) we say that it is of type **(1)**. Otherwise it is of type **(2)**.

If the premise of the claim holds, then by the pigeonhole principle there is a vertex $x \in P'$ such that for at least $n^{1/2} 3n^{1/2}/m > 3$ edges (a, b), (c, d), and (e, f), x is in the overlap for all three edges. We study two cases. In Case 1, x is of type **(1)** with respect to one of these three edges. In Case 2, x is of type **(2)** for all three edges.

The vertices along P' around x are relevant. Specifically, we focus on the subpath of length 10 on P',

$$Q = (u_0, u_1, u_2, u_3, u_4, u_5, u_6, x, u_7, u_8).$$

When specifying a substring we will let a "$*$" denote that the value of the substring in that position is unrestricted. Of course, the value in position x is a zero for all strings.

Case 1: In this case, x is of type **(1)** with respect to one of the edges, say (a, b). Without loss of generality, the substrings on Q for a and b are

$$I_a(Q) = (*, *, *, *, *, 1, 1, 0, 1, *),$$

and

$$I_b(Q) = (*, *, *, *, *, *, 1, 0, 1, 1).$$

In this case we define the subpath

$$Q' = (u_0, u_1, u_2, u_3, u_4, u_5, u_6, a, b, u_7, u_8).$$

Replacing Q by Q' results in a legitimate square path, where we have replaced the subpath Q of length 10 by Q' of length 11, and thus extended P.

Case 2: In this case x of type **(2)** with respect to all three edges. By definition of S_a, whenever $x \in S_a$ either the two vertices which follow x on the path or the two vertices which precede x on the path are in $N_{P'}(a)$. It follows without loss of generality that there are two edges, say (a, b) and (c, d), for which the two vertices following x are in the neighborhoods of a, b, c, and d. Because x is not of type **(1)** for any edge, and because x is in an even interval for every vertex a, b, and c, their substrings on Q are all of the form

$$I_a(Q) = I_b(Q) = I_c(Q) = (*, *, *, 1, 1, 0, 1, 0, 1, 1).$$

If u_2 is not in the neighborhood of a or b, then the substrings on Q for a and b are both exactly

$$I_a(Q) = I_b(Q) = (1, 1, 0, 1, 1, 0, 1, 0, 1, 1),$$

in which case we are back to Case 1, with u_2 taking the place of x. Otherwise u_2 is in the neighborhood of, say, a, and so the substrings of a and b on Q are

$$I_a(Q) = (*, 0, 1, 1, 1, 0, 1, 0, 1, 1),$$

and

$$I_b(Q) = (*, *, *, 1, 1, 0, 1, 0, 1, 1).$$

In this case, we define the subpath

$$Q' = (u_0, u_1, u_2, u_3, a, b, u_4, u_6, c, u_7, u_8).$$

Replacing the subpath Q of length 10 by Q' of length 12 results in a legitimate square path, extending P' by 2. These cases are exhaustive, and thus the claim is established. $\square$

Thus we may assume that there exists a set $M'' \subset M'$ of at least $\alpha^4 t/3 - 3n^{1/2} \geq \alpha^4 t/4$ edges such that each edge of M'' has an overlap of less than $n^{1/2}$. From Claim 28 it follows that for each edge $(a, b) \in M''$, we have

$$|S_a \cup S_b| \geq 2(\frac{1}{3} - 4\alpha^2)n - n^{1/2} \geq (\frac{2}{3} - 9\alpha^2)n.$$

Again, we conclude that there is a set $S \subset \bigcup_a S_a \subset P$ such that $|S| \geq (\frac{2}{3} - 10\alpha^2)n$ and $\forall v \in S$ we have $v \in S_a \cup S_b$ for at least $\alpha^2|M''| \geq \alpha^6 t/4$ edges (a, b) of M''. The reader may check that any set $S' \subset S$ satisfying

1. $|S'| \leq \alpha^8 t/4$, and

2. $\forall u, v, \in S'$, the distance from u to v along P' is less than $\frac{1}{\alpha^2}$

can be exchanged for vertices from distinct edges of M''.

By (3.2), for any $v \in S$,

$$N_S(v) \geq \frac{2}{3}n - \Big(n - (\frac{2}{3} - 10\alpha^2)n\Big) = (\frac{1}{3} - 10\alpha^2)n.$$

As G is not extremal, the density of $N_S(v)$ is at least α. It follows that every vertex in S is contained in many triangles within S. Let $Z \subset S$ be a set of $\alpha^8 t/12$ vertex-disjoint triangles whose vertex set $V(Z)$ satisfies the above two conditions. By the observation we can exchange $V(Z)$ with vertices of M'' without disturbing P, and so we assume that there is a set of $\alpha^8 t/12$ vertex-disjoint triangles $Z \subset T$. As before, we let $Z' = Z \setminus H_0$ be the set of at least $\alpha^8 t/12 - n^{3/4} \geq \alpha^8 t/24$ triangles all of whose vertices satisfy (3.12).

As Claim 29 applies to any set of disjoint edges from Z', we can assume that there is a set Z'' containing at least $\alpha^8 t/24 - 3n^{1/2} \geq \alpha^8 t/48$ vertex-disjoint triangles such that any edge (a, b) of any triangle of Z'' has $|S_{a,b}| < n^{1/2}$. By this bound and Claim 28, it follows that for every triangle $(a, b, c) \in Z''$, $|S_a \cup S_b \cup S_c| \geq (1 - \alpha)n$. Again, we can find a set $S \subset \bigcup_{a \in V(Z'')} S_a$ such that $|S| \geq (1 - 2\alpha)n$ and for every $v \in S$ there are at least $\alpha|U''| \geq \alpha^9 t/48$ triangles $(a, b, c) \in Z''$ such that $v \in S_a \cup S_b \cup S_c$. For any set $S' \subset S$ of size $\alpha^{11} t/48$ all of whose vertices are distance $\frac{1}{\alpha^2}$ apart on P', we can exchange at once each vertex of S' with vertices of $V(Z'')$. With $deg_S(v) \geq (2/3 - 2\alpha)n$, we can easily find a square path of length $\alpha^{13}t < \alpha^{11}t/48$ satisfying the distance condition. Exchanging this path into T and connecting through the reservoir extends P.

In every case we either extend P directly or find a square path of length at least $\alpha^{13}t \geq \alpha^{24}n$ which we connect through the reservoir. The allowance for forbidden vertices in Lemma 25 ensures that we can continue to perform connections until T is small enough that $R \cup T$ can be absorbed by P_A. As in the program outlined at the beginning of the section, at this point we connect the endpairs of P through the reservoir to form a cycle containing P_A, and finally absorb $R \cup T$.

3.7 Conclusion

In order to solve the Pósa problem for every n in the case of $k = 2$, we plan to replace the costly Connecting Lemma with an alternative which we are developing. It is our hope that we will be able to push down n_0 to around 100 at which point we will be able to develop a computer program, taking advantage of much of the structure identified in this thesis to solve the conjecture for every n. This is a work in progress.

References

[1] B. Bollobás, *Extremal Graph Theory*, Academic Press, London (1978).

[2] B. Bollobás, P. Erdős, M. Simonovits, On the structure of edge graphs II, *J. London Math. Soc.*, 12 (2), 1976, pp. 219-224.

[3] R. Diestel, *Graph Theory*, third ed., Graduate Texts in Mathematics, vol. 173, Springer-Verlag, Berlin, 2005.

[4] G. A. Dirac, Some theorems on abstract graphs, *Proc. London Math. Soc.*, 2 (1952), pp. 68-81.

[5] P. Erdős, Problem 9, *Theory of graphs and its applications* (M. Fieldler ed.), Czech. Acad. Sci. Publ., Prague (1964), p. 159.

[6] G. Fan and R. Häggkvist, The square of a hamiltonian cycle, *SIAM J. Disc. Math.*, (1994), pp. 203-212.

[7] G. Fan and H.A. Kierstead, The square of paths and cycles, manuscript.

[8] G. Fan and H.A. Kierstead, The square of paths and cycles, *Journal of Combinatorial Theory, Ser. B*, 63 (1995), pp. 55-64.

[9] G. Fan and H.A. Kierstead, Hamiltonian square-paths, *Journal of Combinatorial Theory, Ser. B*, 67 (1996), pp. 167-182.

[10] G. Fan and H.A. Kierstead, Partitioning a graph into two square-cycles, *Journal of Graph Theory*, 23 (1996), pp. 241-256.

[11] R.J. Faudree, R.J. Gould and M. Jacobson, On a problem of Pósa and Seymour, manuscript.

[12] R.J. Faudree, R.J. Gould, M. S. Jacobson and R. H. Schelp, On a problem of Paul Seymour, *Recent Advances in Graph Theory* (V. R. Kulli ed.), Vishwa International Publication (1991), pp. 197-215.

[13] R. Häggkvist, On F-hamiltonian graphs, *Graph Theory and Related Topics* (J. A. Bondy and U. S. R. Murty eds.), Academic Press, New York (1979), pp. 219-231.

[14] A. Hajnal and E. Szemerédi, Proof of a conjecture of Erdős, Combinatorial Theory and its Applications vol. II (P. Erdős, A. Rényi and V.T. Sós eds.), Colloq. Math. Soc. J. Bolyai 4, North-Holland, Amsterdam (1970), pp. 601-623.

[15] S. Janson, T. Łuczak, A. Ruciński, *Random graphs*, 2000, Wiley, New York.

[16] H.A. Kierstead, J. Quintana, Square Hamiltonian cycles in graphs with maximal 4-cliques, *Discrete Mathematics*, 178 (1998), pp. 81-92.

[17] J. Komlós, G.N. Sárközy and E. Szemerédi, On the square of a Hamiltonian cycle in dense graphs, *Random Structures and Algorithms*, 9 (1996), pp. 193-211.

[18] J. Komlós, G.N. Sárközy and E. Szemerédi, Blow-up Lemma, *Combinatorica*, 17 (1) (1997), pp. 109-123.

[19] J. Komlós, G.N. Sárközy, E. Szemerédi, On The Pósa-Seymour conjecture, *Journal of Graph Theory*, 29 (1998), pp. 167-176.

[20] J. Komlós, G.N. Sárközy and E. Szemerédi, Proof of the Seymour conjecture for large graphs, *Annals of Combinatorics*, 2 (1998), pp. 43-60.

[21] J. Komlós, G.N. Sárközy and E. Szemerédi, An algorithmic version of the Blow-up Lemma, *Random Structures and Algorithms*, 12 (1998), pp. 297-312.

[22] J. Komlós, G.N. Sárközy and E. Szemerédi, Spanning Trees in Dense Graphs, *Combinatorics, Probability and Computing*, 10 (2001), pp. 397-416.

[23] P. Kővári, V.T. Sós, P. Turán, On a problem of Zarankiewicz, *Colloq. Math.*, 3 (1954), pp. 50-57.

[24] V. Rödl, A. Rucinski, E.Szemerédi, A Dirac-type theorem for 3-uniform hypergraphs, *Combinatorics, Probability and Computing*, 15 (2006), pp. 229-251.

[25] P. Seymour, Problem section, *Combinatorics: Proceedings of the British Combinatorial Conference 1973* (T. P.McDonough and V.C. Mavron eds.), Cambridge University Press (1974), pp. 201-202.

[26] E. Szemerédi, Regular partitions of graphs, Colloques Internationaux C.N.R.S. Nº 260 - Problèmes Combinatoires et Théorie des Graphes, Orsay (1976), pp. 399-401.

Vita

Ian M Levitt

Education:

2009 Ph. D. in Mathematics, Rutgers the State University of New Jersey, New Brunswick, NJ

1999 B. Sc. in Mathematics from Stevens Institute of Technology, Hoboken, NJ

Positions:

2007-2009 Mathematician, Research and Development, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City, NJ

2003-2006 Teaching assistant, Department of Mathematics, Rutgers the State University of New Jersey, Piscataway, NJ