1 Warmup

To warm up to the subject of the talk, I'll begin by presenting a few of my favorite theorems, and some of their extremely elegant proofs. For the first, recall that a *tournament* is a complete oriented graph, and that a *Hamiltonian (directed) path* is a (directed) path visiting all vertices exactly once. The following theorem is, as far as I know, the first result ever proved about tournaments.

Theorem 1.1 (Rédei, 1934)**.** *Every tournament contains a Hamiltonian directed path.*

There are many beautiful proofs known of this theorem, and in fact Rédei originally proved something stronger: every tournament contains an *odd* number of Hamiltonian paths. Let me show you my favorite proof, which is due (essentially) to Havet and Thomassé, although it was perhaps known earlier. It actually gives a bit more, namely an efficient algorithm to construct a Hamiltonian directed path in a given tournament.

Proof. Let T be an N-vertex tournament, and arbitrarily label the vertices as $v_1, \dots, v_N$. If $v_1 \to v_2 \to \cdots \to v_N$, then we have found a Hamiltonian directed path. If not, there is some i so that $v_i \leftarrow v_{i+1}$. Swap the labels of v_i and v_{i+1}, and repeat this process.

Clearly, if this process ever terminates, we have found a Hamiltonian directed path, so it suffices to prove that it does terminate. To do this, note that at every step of the process, we strictly increase the number of edges of the tournament that point forward, i.e. from a vertex of lower label to one of higher label. Indeed, at every swap, the edge $v_i v_{i+1}$ goes from being oriented backwards to being oriented forwards, whereas every other edge maintains its orientation. Since we strictly increase this finite quantity at every step, the algorithm must terminate (in fact, in at most $\binom{N}{2}$ steps). □

Our next warmup result looks completely unrelated (apart from the fact that it was also proved in Hungary in the mid-1930s). It is a classical result of Erdős and Szekeres, and is one of the most influential theorems in Ramsey theory.

Theorem 1.2 (Erdős–Szekeres, 1935)**.** *Every sequence of N real numbers contains a monotone subsequence of length at least $\sqrt{N}$.*

To make the connection to Rédei's theorem a little more apparent, let's actually prove the following stronger result, which is essentially equivalent to Theorem 1.2.

Theorem 1.3 (Erdős–Szekeres, 1935)**.** *Every 2-coloring of the edges of a complete ordered graph K_N contains a monochromatic monotone path of length at least $\sqrt{N}$.*

Here, by an *ordered complete graph* we just mean a complete graph K_N whose vertices are labeled by the integers $1, \dots, N$, and a *monotone* path is a path whose vertices are increasing with respect to this ordering. Morever, throughout this talk, the *length* of a path will always refer to the number of vertices it comprises.

To see that Theorem 1.3 implies Theorem 1.2, we can color our graph according to whether pairs of elements in the sequence are increasing or decreasing; it is then clear that a monochromatic monotone path in such a coloring precisely corresponds to a monotone subsequence. The prettiest proof I know of Theorem 1.3 (or Theorem 1.2) is the following proof, due to Seidenberg.

Proof of Theorem 1.3. Label each vertex of K_N by a pair of integers, recording the length of the longest red (resp. blue) monotone path ending at that vertex. Note that any two vertices must receive distinct labels: for example, if x precedes y and x and y are joined by a red edge, then the length of the longest monotone red path ending at y must be strictly greater than that of x, as we can extend any red path ending at x to one ending at y. This means that the total number of labels used must be at least N, and hence at least one label must feature an entry that is at least $\sqrt{N}$; by definition, this implies the existence of a monochromatic monotone path of length at least $\sqrt{N}$. □

To conclude the warmup, we turn to the following beautiful theorem, proven independently by Gallai, Hasse, Roy, and Vitaver in the 1960s.

Theorem 1.4. *Let G be a graph. Every orientation of G contains a directed path of length at least $\chi(G)$.*

Proof of Theorem 1.4. Let us try to mimic Seidenberg's proof of Theorem 1.3. Thus, let us denote by $\ell(v)$ the length of the longest directed path ending at a vertex v. Our goal is to prove that ℓ is a proper coloring of G, hence must use at least $\chi(G)$ different colors, and hence there is a vertex v for which $\ell(v) \geqslant \chi(G)$, as desired.

To prove this, we need to show that if u and v are adjacent, then $\ell(u) \neq \ell(v)$. And this is very simple: if, say, the edge uv is oriented as $u \to v$, then clearly $\ell(u) < \ell(v)$, since we can extend any directed path ending at u to a strictly longer one ending at v. QED.

Unfortunately, this "proof" is definitely wrong. The issue is that the longest directed path ending at u may have already passed through v, and hence we cannot just extend it to a longer path ending at v. To see how wrong this is, consider a directed cycle of length N: every vertex is the endpoint of a Hamiltonian directed path, hence $\ell(v) = N$ for all vertices v, so certainly ℓ is not a proper coloring.

To fix this, we do the following. Let H be a maximal acyclic subgraph of the given orientation of G. That is, H is acyclic, but if we add to it any edge of G not already present, we create a directed cycle. Let us now denote by $\ell'(v)$ the length of the longest directed path *in H* ending at v. We now claim that ℓ' is a proper coloring of G, which would complete the proof as above.

In order to prove this, let us fix some edge $uv \in E(G)$, and suppose that it is oriented as $u \to v$. If $uv \in E(H)$, then we can argue as above: any directed path ending at u cannot have already visited v, for otherwise we would have a directed cycle in H, and therefore it can be extended to a strictly longer path ending at v, implying that $\ell'(v) > \ell'(u)$. On the other hand, if $uv \notin E(H)$, then we know that adding uv to $E(H)$ would create a directed cycle. That is, there is an oriented path $v \to w_1 \to w_2 \to \cdots \to w_k \to u$, for some vertices $w_1, \ldots, w_k$, such that all these are edges of H. By the argument above, we must have $\ell'(v) < \ell'(w_1) < \cdots < \ell'(w_k) < \ell'(u)$, and in particular $\ell'(v) \neq \ell'(u)$. Thus, in both cases, we find that $\ell'(u) \neq \ell'(v)$, and hence ℓ' is a proper coloring of G, completing the proof. □

An immediate corollary of Theorem 1.4 is the following Ramsey-theoretic statement, observed independently by Chvátal and Gyárfás–Lehel.

Corollary 1.5 (Chvátal–Gyárfás–Lehel)**.** *In any q-edge-coloring of any N-vertex tournament, there is a monochromatic directed path of length at least $N^{1/q}$.*

Proof. Viewing each color class as an oriented graph, one of these graphs must have chromatic number at least $N^{1/q}$, since the chromatic number of a union of graphs is at most the product of their individual chromatic numbers. Applying Theorem 1.4 to this color class gives the desired result. □

Corollary 1.5 is a rather powerful result; in particular, two of its (very) special cases are the first two theorems we started with. Indeed, Theorem 1.1 is nothing more than the $q = 1$ case of this corollary, and Theorem 1.3 is precisely this statement applied to a *transitive* tournament, so that monotone paths and directed paths coincide.

We remark that the bound in Corollary 1.5 is tight (at least when N is a perfect power of q), even in the special case of transitive tournaments (which in particular implies that the bound in Theorems 1.2 and 1.3 is tight). To see this, consider a transitive tournament on N vertices, and split its vertex set into $N^{1/q}$ intervals, each of length $N^{1-1/q}$. Color all edges between these intervals in the first color. Note that in this first color, any directed path can visit each interval at most once, hence has length at most $N^{1/q}$. Inside each interval, again split into $N^{1/q}$ intervals, and color all edges between them with color 2. Then the graph of color-2 edges is disconnected (with each top-level interval being a connected component), hence any directed path in color 2 must again use at most $N^{1/q}$ vertices. Continuing in this fashion, we can color all edges with q colors and no directed path longer than $N^{1/q}$.

2 Color-avoiding paths

Corollary 1.5 is a classical type of Ramsey-theoretic result, stating that any colored object must contain a "large" monochromatic substructure of a specified type. However, another very influential line of work in Ramsey theory concerns *color-avoiding* problems. In such problems, we are given a q-colored object, and aim to find a substructure which *avoids* at least one of the colors. In case $q = 2$, the two problems coincide, but once $q \geqslant 3$ they are rather different. Nevertheless, in a wide range of contexts (e.g. in the multicolor Erdős–Hajnal conjecture), the color-avoiding setting turns out to be substantially more natural to work with than the monochromatic setting.

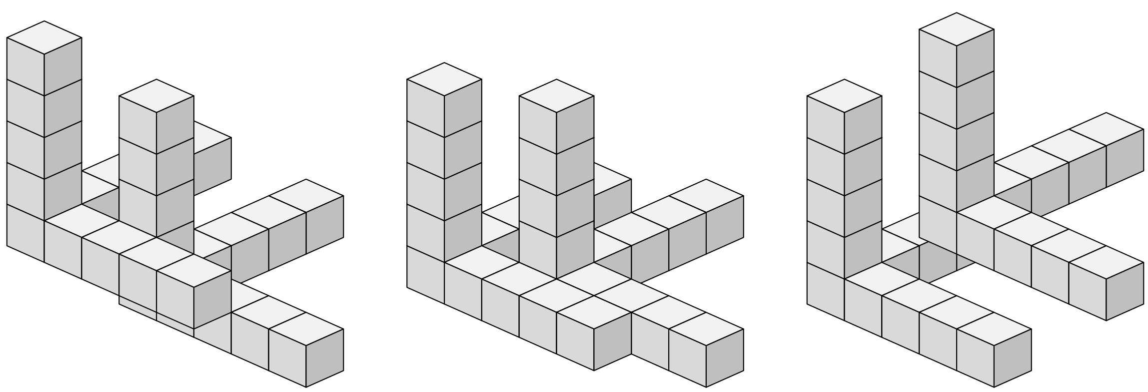
In this talk, we will study the following such question: given a q-edge-coloring of an N-vertex tournament, how long of a color-avoiding directed path can we guarantee? More precisely, let us denote by $g_q(N)$ the minimum, over all q-edge-colored N-vertex tournaments, of the length of the longest color-avoiding directed path. Thus, a lower bound on $g_q(N)$ is precisely a statement of the form "every q-edge-colored N-vertex tournament contains a color-avoiding path of length at least…"

This may seem to be a somewhat "out of nowhere" question, but it turns out to arise naturally in a wide array of different contexts. Let me briefly discuss a few of them.

1. Motivated by questions in both group theory and coding theory, in the 1960s Stein and his collaborators started studying the following packing question: what is the maximum density with which *tripods* can pack 3-dimensional space? A tripod is a shape consisting of 3 infinite rays of $1 \times 1 \times 1$ cubes meeting at a shared corner, called the *apex*. A tripod can be seen as a Euclidean analogue of a radius-1 Hamming ball, and hence the question of how densely they pack is very reminiscent of fundamental questions in coding theory.

   By a standard limiting argument, this is essentially the same as asking for the asymptotics in k of how many disjoint tripods we can pack such that their apices lie in a $k \times k \times k$ box. Moreover, we might as well assume that the bottom corner of each apex is an integer vector, i.e. lies in $[k]^3$.

   Let us say that a tripod T' *stabs* a tripod T if one leg of T' goes through a plane spanned by two of the legs of T. We also include in this definition the degenerate case where the apices lie in the same axis-aligned plane. In case neither T nor T' stabs the other, the only remaining option is that T *nestles* T' (or vice versa); the following pictures show two examples of stabbing (one degenerate), and an example of nestling.



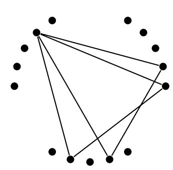
   Suppose we are able to pack N such tripods with apices in $[k]^3$. Define a 3-edge-colored N-vertex tournament as follows. The vertices of the tournament are identified with the tripods. Given two tripods T, T', suppose first that T' stabs T. We then orient the edge between them as $T \to T'$, and color it with the direction of the leg doing the stabbing (so in the pictures above, we would color this edge with color z, as the leg in the z-direction is stabbing). In case T nestles T', we again direct the edge as $T \to T'$, and color it with one of the three colors arbitrarily.

   Now, we claim that this N-vertex 3-edge-colored tournament does not contain a color-avoiding directed path of length longer than k. Indeed, consider (say) a directed path avoiding color z. Because no edge along this path was ever colored z, as we walk along the path, the z-coordinates of the apices must be strictly increasing; if they ever decrease, then we must have a z-stabbing pair, which is impossible. Since all z coordinates lie between 1 and k, we see that there is no color-avoiding path of length greater than k.

Summarizing, we have found that if we can pack N tripods with apices in $[k]^3$, then $g_3(N) \leqslant k$, since we have found some tournament and some 3-edge-coloring of it whose longest color-avoiding directed path has length at most k. In other words, lower bounds on $g_3(N)$ imply upper bounds on the packing density of tripods.

2. Brass studied the following class of extremal questions. Fix k points in convex position in the plane. How many triangles spanned by these points can one select while avoiding certain geometric configurations of triangles? For some configurations this problem is quite easy to answer, but others turn out to be rather more involved. A specific question raised by Brass concerned the following pair of configurations: he asked to avoid two triangles which share two nodes, as well as two triangles which share one node and are geometrically nested, i.e. in the form .

   Aronov, Dujmović, Morin, Ooms, and Schultz Xavier da Silveira found that this problem is also related to $g_3(N)$. To see this, suppose we have such a collection of N triangles. Let us assume, without too much loss of generality, that the k points of our convex polygon are divided into three intervals of $k/3$ points each, and each triangle from our collection transverses the three intervals. We define a 3-edge-colored N-vertex tournament as follows, viewing the collection of triangles as our vertices. Any pair of triangles shares at most one node by assumption. Moreover, whenever they share exactly one node, the other nodes must either both be clockwise or counterclockwise, otherwise we would get a nesting pair.



   Now, if T and T' share a node, we orient the edge so that the other nodes proceed clockwise along the directed edge, and we color the edge according to which of the three intervals contains the shared node. If T and T' share no nodes, we again orient the edge so that at least two proceed clockwise. If all three proceed clockwise, we color it arbitrarily; if only two do, we color it according to which of the three intervals contains the counterclockwise motion.

   Now, consider a color-avoiding directed path in this tournament. Since it avoids some color, in that interval we must only ever move clockwise along the directed path, hence the directed path has length at most $k/3$. That is, again, a lower bound for $g_3(N)$ yields an upper bound to Brass's question.

3. For an odd integer k, a *k-majority tournament* is obtained by taking k linear orderings on a vertex set, and then orienting each edge xy depending on whether $x < y$ or $x > y$ in the majority of the orderings. Such tournaments have a wide array of interesting properties, and have been extensively well-studied; one important question is how large of a transitive subtournament they must contain. Loh developed a new construction for

this question, but was unable to fully analyze it; however, he noticed that for $k = 3$, the size of the largest transitive subtournament in his construction is essentially controlled by the length of the longest color-avoiding path in an auxiliary coloring, where each edge is colored according to which of the three linear orderings is in the minority (and colored arbitrarily if all three agree). This led him independently to study $g_3(N)$, and he seems to have been the first to explicitly cast this as a question about color-avoiding paths.

4. A MathOverflow question asked by Hamilton concerned a certain property of sequences of correlated random variables, which he called “everywhere probably increasing”. Sawin and Tao were able to essentially completely resolve his question, using rather intricate Fourier-analytic arguments, notably an application of *Hilbert's inequality*. As observed by Pohoata, this problem (and its solution) can be recast as one about color-avoiding paths in *transitive* tournaments; in particular, the results of Sawin and Tao imply very strong bounds on the variant of $g_q(N)$ where we only restrict our attention to q-colorings of a transitive tournament.

With these motivations in mind, let us turn to what is actually known about the function $g_q(N)$. First, recall the coloring we used to show that Corollary 1.5 is tight, an iterated blowup of a monochromatic coloring on $N^{1/q}$ parts. Note that any color-avoiding path in this coloring must have length at most $N^{1-1/q}$, from which we deduce the bound $g_q(N) \leqslant N^{1-1/q}$. As far as I know, this remains the only upper bound on $g_q(N)$ known in general. On the other hand, the only known lower bound prior to our work is $g_q(N) \geqslant \sqrt{N}$, which follows from the following (extremely wasteful) argument: given any q-edge-colored N-vertex tournament, arbitrarily group the colors into two collections. Viewing each collection as a “super-color”, we obtain a 2-coloring of an N-vertex tournament, which thus contains a monochromatic directed path of length at least $\sqrt{N}$ by Corollary 1.5. In the original coloring, this directed path only uses one collection of the colors, and in particular avoids all the others, hence is color-avoiding; this shows $g_q(N) \geqslant \sqrt{N}$.

The only improvement on these trivial bounds is in the case $q = 3$, which, as discussed above, has been studied a great deal over the years in many different guises. In particular, Hamaker and Stein found a recursive construction for the tripod packing problem which, when converted to the language of tournaments, yields an upper bound of $g_3(N) \leqslant N^{0.663}$, which is in particular polynomially better than our earlier bound of $g_3(N) \leqslant N^{2/3}$. Their bound was improved several times over the years, culminating in important work of Gowers and Long, who (unaware of these earlier developments) introduced a continuous relaxation of this problem and used it to prove the upper bound $g_3(N) \leqslant N^{0.647}$, where the exponent is some irrational number arising as the optimum of a certain optimization problem. That is, they construct a 3-edge-colored tournament whose longest color-avoiding directed path has length at most $N^{0.647}$. Rather interestingly, their construction (as well as those coming before it) use highly non-transitive tournaments, and it is conjectured that the $N^{2/3}$ bound we saw above is tight when restricted to transitive tournaments.

Summarizing, we have the bounds

$$\sqrt{N} \leqslant g_q(N) \leqslant N^{1-1/q}$$

for all $q \geqslant 3$, with the upper bound being slightly improved in the case $q = 3$ (and only in this case). Note that the gap between the lower and upper bounds only widens as q increases, and it is very natural to wonder which is closer to the truth. Given that the lower bound argument does nothing more than the rather naive and wasteful color-merging approach, it is natural to expect that it is far from the truth, and indeed Gowers and Long raised as an open problem to prove $g_q(N) \geqslant N^{\frac{1}{2}+\delta}$ for even a single q and some fixed $\delta > 0$. Our main result, whose proof I'll discuss in the remainder of the talk, does this in a strong form.

Theorem 2.1 (Fox–Sudakov–W.). *For every $\varepsilon > 0$ and every sufficiently large q, we have $g_q(N) \geqslant N^{1-\varepsilon}$.*

More precisely, we have the bound $g_q(N) \geqslant N^{1-C/\sqrt{\log q}}$, for some absolute constant C.

That is, as $q \to \infty$, the true exponent of $g_q(N)$ tends to 1, and thus the upper bound of $N^{1-1/q}$ is in some sense not far from the truth. However, for any fixed q, we still have a rather large gap.

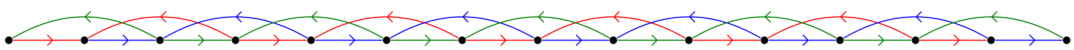
3 Proof outline

I now want to sketch some of the ingredients that go into the proof of Theorem 2.1. First, recall that the constructions of Hamaker–Stein and Gowers–Long, which prove stronger upper bounds on $g_3(N)$, use highly non-transitive tournaments. Our first technical statement, which is quite interesting in its own right, shows that this phenomenon is fundamentally restricted to the 3-color setting: for $q \geqslant 4$, every q-edge-colored tournament that is far from transitive contains color-avoiding directed paths of *linear* length. To make this precise, let us recall that a tournament is *δ-far from transitive* if at least a δ-fraction of its edges must be reversed to obtain a transitive tournament.

Lemma 3.1. *If a q-edge-colored N-vertex tournament is δ-far from transitive, then it contains a directed path of length $\Omega(\delta^2 N/q^3)$ which is colored by at most 3 colors. In particular, if $q \geqslant 4$, it contains a color-avoiding path of length $\Omega_{q,\delta}(N)$.*

Proof. The main input we need is the following nice result of Fox and Sudakov, which can be viewed as a directed version of the triangle removal lemma: if an N-vertex tournament T is δ-far from transitive, then it contains $\Omega(\delta^2 N^3)$ cyclic triangles. The proof of this lemma, in turn, is quite short and elementary, using nothing but averaging arguments and induction.

With this lemma in hand, we can apply the pigeonhole principle to find that there is a collection $\mathcal{T}$ of $\Omega(\delta^2 N^3/q^3)$ cyclic triangles that receive exactly the same triple of colors on their edges. By repeatedly deleting edges that lie on too few triangles in $\mathcal{T}$, we can pass to a collection of edges, each of which lies on $\Omega(\delta^2 N/q^3)$ triangles of $\mathcal{T}$. We now greedily use these to build a path: for each edge we reach, we pick a triangle in $\mathcal{T}$ containing it and which we have not already used, and continue on the next edge in this triangle. We clearly build a directed path during this process, and the only colors we will ever encounter are the 3 colors used on triangles in $\mathcal{T}$.



Moreover, we can keep this process going for $\Omega(\delta^2 N/q^3)$ steps, since each edge remaining lies on at least that many triangles in $\mathcal{T}$. □

Since our goal is only to find a color-avoiding path of length $N^{1-\varepsilon}$, Lemma 3.1 gives us far more than we need. In particular, we may assume throughout that our given q-edge-colored tournament is very close to transitive, as otherwise we are immediately done. To keep things simple, we will actually assume from now on that our tournament is *exactly* transitive. While this is a serious simplification, the transitive case already captures the majority of the ideas in our proof. As mentioned above, the Fourier-analytic arguments of Sawin and Tao can give much stronger bounds in this special case, but they seem impossible to adapt to the nearly transitive setting, whereas our combinatorial arguments can be modified without too much difficulty.

As our tournament is transitive, we can simply view it as an ordered complete graph on the vertex set $[N]$; additionally, a path in this complete graph is directed if and only if it is monotone. So our task is to prove that if ε is fixed and q is sufficiently large, then any q-edge-colored complete graph on $[N]$ contains a color-avoiding monotone path of length at least $N^{1-\varepsilon}$, and we prove this by induction on N. For a q-edge-colored transitive tournament T and a color $i \in [q]$, let $\ell_i(T)$ denote the length of the longest directed path in T avoiding color i. Thus, our goal is to prove by induction that $\ell_i(T) \geqslant N^{1-\varepsilon}$ for some i. We will actually strengthen the induction hypothesis, and prove that $\Pi(T) \geqslant N^{q-\varepsilon q}$, where we define $\Pi(T) = \prod_{i=1}^q \ell_i(T)$. The statement $\Pi(T) \geqslant N^{q-\varepsilon q}$ immediately implies that $\ell_i(T) \geqslant N^{1-\varepsilon}$ for some $i \in [q]$, hence this really is a stronger statement.

Our strategy is as follows. We divide $[N]$ into two subintervals $T_1 = [1, N/2]$ and $T_2 = (N/2, N]$, and apply the inductive hypothesis in each interval. Our dream scenario is to take a longest i-avoiding path P_1 in the first half and a longest i-avoiding path P_2 in the second half, and to glue them together to obtain an i-avoiding path of double the length. Of course, for this to work, we need the edge joining the end of P_1 and the start of P_2 to *not* be colored with color i; if we can ensure this, then their concatenation $P_1 \cup P_2$ truly is an i-avoiding path of length $|P_1| + |P_2|$. Thus, if we can do this for color i, we learn that

$$\ell_i(T) \geqslant \ell_i(T_1) + \ell_i(T_2) \overset{\text{AM-GM}}{\geqslant} 2\sqrt{\ell_i(T_1)\ell_i(T_2)}.$$

If, in turn, we can do this for *all* the colors, we conclude that

$$\Pi(T) = \prod_{i=1}^q \ell_i(T) \geqslant \prod_{i=1}^q 2\sqrt{\ell_i(T_1)\ell_i(T_2)} = 2^q\sqrt{\Pi(T_1)\Pi(T_2)}$$

Inductively, we know that $\min\{\Pi(T_1), \Pi(T_2)\} \geqslant (N/2)^{q-\varepsilon q}$, hence we learn that $\Pi(T) \geqslant 2^q(N/2)^{q-\varepsilon q} \geqslant N^{q-\varepsilon q}$. That is, in the dream scenario, we can prove the desired result.

Of course, there is no reason for the dream scenario to be remotely close to true—we shouldn't even expect to be able to glue together a longest i-avoiding path in each half even for *a single* color i, let alone for all of them. The proof now consists of showing that either we can glue together *almost* longest i-avoiding paths for *almost* all colors i, or else that we win for another reason; the losses inherent in these "almosts" are what contribute to the error term in the exponent.

We begin by explaining the first "almost". Because we are allowed to tolerate errors, it is not necessary for P_1 and P_2 above to be truly the longest i-avoiding paths in T_1, T_2, respectively. Instead, it suffices that their lengths are nearly as long as possible, so that $\ell_i(T) \geqslant (2 - o(1))\sqrt{\ell_i(T_1)\ell_i(T_2)}$. This gives us a lot more flexibility, since it now suffices to just find one such pair P_1, P_2, of nearly maximal length, such that the edge joining the end of P_1 and the start of P_2 is not colored with color i. To accomplish this, we will consider a large number of nearly maximal i-avoiding paths in each half, and attempt to glue together every pair.

Concretely, fix some parameter $1 \ll s \ll N$. We wish to pick s different i-avoiding paths in each half; moreover, since our goal is to try gluing every pair of paths from each half, we want these s paths to have different endpoints. To accomplish this, let us define $\ell_i({}^{\to}v)$ to be the length of the longest i-avoiding directed path ending at v, and similarly $\ell_i(w^{\to})$ to be the length of the longest i-avoiding directed path starting at w. We now define X_i to be the set of the s vertices in the first half with the highest value of $\ell_i({}^{\to}\bullet)$. Similarly, we define Y_i to be the set of the s vertices in the second half with the highest value of $\ell_i(\bullet^{\to})$. The paths P_1 that we consider will then be the longest i-avoiding paths ending at a vertex of X_i, and similarly the P_2 we consider are the longest i-avoiding paths starting at a vertex of Y_i. Heuristically, since $s \ll N$ and X_i consists of the s "best" endpoints for i-avoiding paths in T_1, every vertex of X_i is the endpoint of such a path of length almost $\ell_i(T_1)$.

We now call a color i *compressed* if the majority of the vertices of X_i and the majority of the vertices of Y_i are close to the midpoint $N/2$, i.e. in the interval $[\frac{N}{2} - 4s, \frac{N}{2} + 4s]$. The claim now is that we can glue together i-avoiding paths in T_1 and T_2 for nearly all compressed colors. Indeed, if i is a compressed color and we cannot perform such a gluing, that means that all edges from X_i to Y_i are of color i, and in particular we have found many edges of color i within the interval $[\frac{N}{2} - 4s, \frac{N}{2} + 4s]$. As there are only $O(s^2)$ edges within this interval, we must be able to glue for the vast majority of compressed colors: each color in which we cannot contributes many edges in this interval, and there are not so many such edges.

Thus, if nearly all colors are compressed, we are "almost" in the dream scenario and we can prove the claim inductively by absorbing all the losses in the $N^{-\varepsilon q}$ factor. It remains to understand what happens when we have a large number, say $2p$, of non-compressed colors; by symmetry, we may assume that at least p of them are *left-diffuse*, i.e. that the majority of X_i for these colors lies in the interval $[1, \frac{N}{2} - 4s]$. Let $U_i = X_i \cap [1, \frac{N}{2} - 4s]$, for each of these p left-diffuse colors i.

The key claim now is the following. Suppose that v, w are vertices, where v precedes w in the ordering of $[N/2]$, and assume that $v \in X_i, w \notin X_i$. Then the edge vw is necessarily colored with color i. Indeed, if it were not, then we can extend any i-avoiding path ending at v to a longer one ending at w, but then the fact that $v \in X_i, w \notin X_i$ contradicts the definition of X_i. As in the incorrect proof of the GHRV theorem, it is here that we crucially use the fact that T is transitive, and it is this step that takes the most care to adapt to the nearly-transitive setting.

The key claim above has two immediate corollaries. First, if i, j are distinct left-diffuse colors, then U_i and U_j are disjoint. Indeed, since there are few vertices of $X_i \cup X_j$ in $[\frac{N}{2} - 4s, \frac{N}{2}]$, we can find $w \notin X_i \cup X_j$ in this interval. If $v \in U_i \cap U_j$, then the observation above implies that vw must receive both color i and color j, which is impossible, hence $U_i \cap U_j = \varnothing$. The second corollary is that the color of every edge between U_i and U_j must be left-diffuse; indeed, by the key claim, every such edge must in fact be either of color i or of color j.

This immediately has the following remarkable consequence. For every color k which is *not* one of our p left-diffuse colors, we have

$$\ell_k(T) \geqslant \sum_{i \text{ left-diffuse}} \ell_k(T[U_i]) \overset{\text{AM-GM}}{\geqslant} p \left(\prod_{i \text{ left-diffuse}} \ell_k(T[U_i]) \right)^{1/p}.$$

Indeed, we get the desired k-avoiding path by using the vertices from the longest k-avoiding path in each U_i. This path is k-avoiding as every edge of the path is either between vertices from the same U_i or must be one of the p left-diffuse colors. In other words, we find that for $q - p$ choices of color k, we essentially gain a factor of p in the length of the longest k-avoiding path. This is an enormous gain, much larger than the factor of 2 we were winning in the dream scenario; moreover, by choosing $p = o(q)$, we can obtain such a gain for nearly all colors. The loss comes from the fact that we now need to apply the inductive hypothesis to the subtournament induced on each U_i, which is quite small; however, by picking p and s appropriately, we can ensure that the amount we win by multiplying the lengths by p dominates this loss.