

Turán’s Theorem Through Algorithmic Lens

Fedor V. Fomin¹, Petr A. Golovach¹, Danil Sagunov², and Kirill Simonov³

¹ Department of Informatics, University of Bergen, Norway.
{fomin, petr.golovach}@ii.uib.no

² St. Petersburg Department of V.A. Steklov Institute of Mathematics, Russia.
danilka.pro@gmail.com

³ Hasso Plattner Institute, University of Potsdam, Germany.
kirillsimonov@gmail.com

Abstract. The fundamental theorem of Turán from Extremal Graph Theory determines the exact bound on the number of edges $t_r(n)$ in an n -vertex graph that does not contain a clique of size $r + 1$. We establish an interesting link between Extremal Graph Theory and Algorithms by providing a simple compression algorithm that in linear time reduces the problem of finding a clique of size ℓ in an n -vertex graph G with $m \geq t_r(n) - k$ edges, where $\ell \leq r + 1$, to the problem of finding a maximum clique in a graph on at most $5k$ vertices. This also gives us an algorithm deciding in time $2.49^k \cdot (n + m)$ whether G has a clique of size ℓ . As a byproduct of the new compression algorithm, we give an algorithm that in time $2^{\mathcal{O}(td^2)} \cdot n^2$ decides whether a graph contains an independent set of size at least $n/(d+1) + t$. Here d is the average vertex degree of the graph G . The multivariate complexity analysis based on ETH indicates that the asymptotical dependence on several parameters in the running times of our algorithms is tight.

1 Introduction

In 1941, Pál Turán published a theorem that became one of the central results in extremal graph theory. The theorem bounds the number of edges in an undirected graph that does not contain a complete subgraph of a given size. For positive integers $r \leq n$, the *Turán’s graph* $T_r(n)$ is the unique complete r -partite n -vertex graph where each part consists of $\lfloor \frac{n}{r} \rfloor$ or $\lceil \frac{n}{r} \rceil$ vertices. In other words, $T_r(n)$ is isomorphic to K_{a_1, a_2, \dots, a_r} , where $a_i = \lceil \frac{n}{r} \rceil$ if i is less than or equal to n modulo r and $a_i = \lfloor \frac{n}{r} \rfloor$ otherwise. We use $t_r(n)$ to denote the number of edges in $T_r(n)$.

Theorem 1 (Turán’s Theorem [28]). *Let $r \leq n$. Then any K_{r+1} -free n -vertex graph has at most $t_r(n)$ edges. The only K_{r+1} -free n -vertex graph with exactly $t_r(n)$ edges is $T_r(n)$.*

The theorem yields a polynomial time algorithm that for a given n -vertex graph G with at least $t_r(n)$ edges decides whether G contains a clique K_{r+1} . Indeed, if a graph G is isomorphic to $T_r(n)$, which is easily checkable in polynomial time, then it has no clique of size $r + 1$. Otherwise, by Turán’s theorem,

G contains K_{r+1} . There are constructive proofs of Turán’s theorem that also allows to find a clique of size $r + 1$ in a graph with at least $t_r(n)$ edges.

The fascinating question is whether Turán’s theorem could help to find efficiently larger cliques in sparser graphs. There are two natural approaches to defining a “sparser” graph and a “larger” clique. These approaches bring us to the following questions; addressing these questions is the primary motivation of our work.

First, what happens when the input graph has a bit less edges than the Turán’s graph? More precisely,

Is there an efficient algorithm that for some $k \geq 1$, decides whether an n -vertex graph with at least $t_r(n) - k$ edges contains a clique of size $r + 1$?

Second, could Turán’s theorem be useful in finding a clique of size larger than $r + 1$ in an n -vertex graph with $t_r(n)$ edges? That is,

Is there an efficient algorithm that for some $\ell > r$ decides whether an n -vertex graph with at least $t_r(n)$ edges contains a clique of size ℓ ?

We provide answers to both questions, and more. We resolve the first question by showing a *simple* fixed-parameter tractable (FPT) algorithm where the parameter is k , the “distance” to the Turán’s graph. Our algorithm builds on the cute ideas used by Erdős in his proof of Turán’s theorem [11]. Viewing these ideas through algorithmic lens leads us to a simple preprocessing procedure, formally a linear-time polynomial compression. For the second question, unfortunately, the answer is negative.

Our contribution. To explain our results, it is convenient to state the above questions in terms of the computational complexity of the following problem.

TURÁN’S CLIQUE

Input: An n -vertex graph G , positive integers $r, \ell \leq n$, and k such that $|E(G)| \geq t_r(n) - k$.

Question: Is there a clique of size at least ℓ in G ?

Our first result is the following theorem (Theorem 2). Let G be an n -vertex graph with $m \geq t_r(n) - k$ edges. Then there is an algorithm that for any $\ell \leq r + 1$, in time $2.49^k \cdot (n + m)$ either finds a clique of size at least ℓ in G or correctly reports that G does not have a clique of size ℓ . Thus for $\ell \leq r + 1$, TURÁN’S CLIQUE is FPT parameterized by k . More generally, we prove that the problem admits a compression of size linear in k . That is, we provide a linear-time procedure that reduces an instance (G, r, ℓ, k) of TURÁN’S CLIQUE to an equivalent instance (G', p) of the CLIQUE problem with at most $5k$ vertices. The difference between CLIQUE and TURÁN’S CLIQUE is that we do not impose any bound on the number of edges in the input graph of CLIQUE. This is why we use the term compression rather than kernelization,⁴ and we argue that stating our reduction

⁴ A *kernel* is by definition a reduction to an instance of the same problem. See the book [13] for an introduction to kernelization.

in terms of compression is far more natural and helpful. Indeed, after reducing the instance to the size linear in the parameter k , the difference between CLIQUE and TURÁN'S CLIQUE vanes, as even the total number of edges in the instance is automatically bounded by a function of the parameter. On the other hand, CLIQUE is a more general and well-studied problem than TURÁN'S CLIQUE.

Pipelined with the fastest known exact algorithm for MAXIMUM INDEPENDENT SET of running time $\mathcal{O}(1.1996^n)$ [29], our reduction provides the FPT algorithm for TURÁN'S CLIQUE parameterized by k . This algorithm is single-exponential in k and linear in $n + m$, and we also show that the existence of an algorithm subexponential in k would contradict Exponential Time Hypothesis (Corollary 5). Thus the running time of our algorithm is essentially tight, up to the constant in the base of the exponent.

The condition $\ell \leq r + 1$ required by our algorithm is, unfortunately, unavoidable. We prove (Theorem 4) that for any fixed $p \geq 2$, the problem of deciding whether an n -vertex graph with at least $t_r(n)$ edges contains a clique of size $\ell = r + p$ is NP-complete. Thus for any $p \geq 2$, TURÁN'S CLIQUE parameterized by k is para-NP-hard. (We refer to the book of Cygan et al. [6] for an introduction to parameterized complexity.)

While our hardness result rules out finding cliques of size $\ell > r + 1$ in graphs with $t_r(n)$ edges in FPT time, an interesting situation arises when the ratio $\xi := \lfloor \frac{n}{r} \rfloor$ is small. In the extreme case, when $n = r$, the n -vertex graph G with $t_r(n) = n(n-1)/2$ is a complete graph. In this case the problem becomes trivial.

To capture how far the desired clique is from the Turán's bound, we introduce the parameter

$$\tau = \begin{cases} 0, & \text{if } \ell \leq r, \\ \ell - r, & \text{otherwise.} \end{cases}$$

The above-mentioned compression algorithm into CLIQUE with at most $5k$ vertices yields almost “for free” a compression of TURÁN'S CLIQUE into CLIQUE with $\mathcal{O}(\tau\xi^2 + k)$ vertices. Hence for any ℓ , one can decide whether an n -vertex graph with $m \geq t_r(n) - k$ edges contains a clique of size ℓ in time $2^{\mathcal{O}(\tau\xi^2 + k)} \cdot (n + m)$. Thus the problem is FPT parameterized by $\tau + \xi + k$. This result has an interesting interpretation when we look for a large independent set in the complement of a graph. Turán's theorem, when applied to the complement \overline{G} of a graph G , yields a bound

$$\alpha(G) \geq \frac{n}{d+1},$$

where $\alpha(G)$ is the size of the largest independent set in G (the independence number of G), and d is the average vertex degree of G . This motivates us to define the following problem.

TURÁN'S INDEPENDENT SET

Input: An n -vertex graph G with average degree d , a positive integer t .

Question: Is there an independent set of size at least $\frac{n}{d+1} + t$ in G ?

By Theorem 3, we have a simple algorithm (Corollary 3) that compresses an instance of TURÁN'S INDEPENDENT SET into an instance of INDEPENDENT

SET with $\mathcal{O}(td^2)$ vertices. Pipelined with an exact algorithm computing a maximum independent set, the compression results in the algorithm solving TURÁN'S INDEPENDENT SET in time $2^{\mathcal{O}(td^2)} \cdot n^2$.

As we already mentioned, TURÁN'S CLIQUE is NP-complete for any fixed $\tau \geq 2$ and $k = 0$. We prove that the problem remains intractable being parameterized by any pair of the parameters from the triple $\{\tau, \xi, k\}$. More precisely, TURÁN'S CLIQUE is also NP-complete for any fixed $\xi \geq 1$ and $\tau = 0$, as well as for any fixed $\xi \geq 1$ and $k = 0$. These lower bounds are given in Theorem 4.

Given the algorithm of running time $2^{\mathcal{O}(\tau\xi^2+k)} \cdot (n+m)$ and the lower bounds for parameterization by any pair of the parameters from $\{\tau, \xi, k\}$, a natural question is, what is the optimal dependence of a TURÁN'S CLIQUE algorithm on $\{\tau, \xi, k\}$? We use the Exponential Time Hypothesis (ETH) of Impagliazzo, Paturi, and Zane [20] to address this question. Assuming ETH, we rule out the existence of algorithms solving TURÁN'S CLIQUE in time $f(\xi, \tau)^{\mathcal{O}(k)} \cdot n^{f(\xi, \tau)}$, $f(\xi, k)^{\mathcal{O}(\tau)} \cdot n^{f(\xi, k)}$, and $f(k, \tau)^{\mathcal{O}(\sqrt{\xi})} \cdot n^{f(k, \tau)}$, for any function f of the respective parameters.

Related work. CLIQUE is a notoriously difficult computational problem. It is one of Karp's 21 NP-complete problems [22] and by the work of Håstad, it is hard to approximate CLIQUE within a factor of $n^{1-\epsilon}$ [19]. CLIQUE parameterized by the solution size is W[1]-complete [8]. The problem plays the fundamental role in the W-hierarchy of Downey and Fellows, and serves as the starting point in the majority of parameterized hardness reductions. From the viewpoint of structural parameterized kernelization, CLIQUE does not admit a polynomial kernel when parameterized by the size of the vertex cover [3]. A notable portion of works in parameterized algorithms and kernelization is devoted to solving INDEPENDENT SET (equivalent to CLIQUE on the graph's complement) on specific graph classes like planar, H -minor-free graphs and nowhere-dense graphs [7, 2, 10, 27].

Our algorithmic study of Turán's theorem fits into the paradigm of the “above guarantee” parameterization [25]. This approach was successfully applied to various problems, see e.g. [1, 5, 14, 15, 16, 17, 18, 24, 26, 21, 12].

Most relevant to our work is the work of Dvorak and Lidicky on independent set “above Brooks' theorem” [9]. By Brooks' theorem [4], every n -vertex graph of maximum degree at most $\Delta \geq 3$ and clique number at most Δ has an independent set of size at least n/Δ . Then the INDEPENDENT SET OVER BROOK'S BOUND problem is to decide whether an input graph G has an independent set of size at least $\frac{n}{\Delta} + p$. Dvorak and Lidicky [9, Corollary 3] proved that INDEPENDENT SET OVER BROOK'S BOUND admits a kernel with at most $114p\Delta^3$ vertices. This kernel also implies an algorithm of running time $2^{\mathcal{O}(p\Delta^3)} \cdot n^{\mathcal{O}(1)}$. When average degree d is at most $\Delta - 1$, by Corollary 3, we have that INDEPENDENT SET OVER BROOK'S BOUND admits a compression into an instance of INDEPENDENT SET with $\mathcal{O}(p\Delta^2)$ vertices. Similarly, by Corollary 4, for $d \leq \Delta - 1$, INDEPENDENT SET OVER BROOK'S BOUND is solvable in time $2^{\mathcal{O}(p\Delta^2)} \cdot n^{\mathcal{O}(1)}$. When $d > \Delta - 1$, for example, on regular graphs, the result of Dvorak and Lidicky is non-comparable with our results.

2 Algorithms

While in the literature it is common to present Turán's theorem under the implicit assumption that n is divisible by r , here we make no such assumption. For that, it is useful to recall the precise value of $t_r(n)$ in the general setting, as observed by Turán [28].

Proposition 1 (Turán [28]). *For positive integers $r \leq n$,*

$$t_r(n) = \left(1 - \frac{1}{r}\right) \cdot \frac{n^2}{2} - \frac{s}{2} \cdot \left(1 - \frac{s}{r}\right)$$

where $s = n - r \cdot \lfloor \frac{n}{r} \rfloor$ is the remainder in the division of n by r .

Note that [28] uses the expression $t_r(n) = \frac{r-1}{2r} \cdot (n^2 - s^2) + \binom{s}{2}$, however it can be easily seen to be equivalent to the above.

We start with our main problem, where we look for a K_{r+1} in a graph that has slightly less than $t_r(n)$ edges. Later in this section, we show how to derive our other algorithmic results from the compression routine developed next.

2.1 Compression algorithm for $\ell \leq r + 1$

First, we make a crucial observation on the structure of a TURÁN'S CLIQUE instance that will be the key part of our compression argument. Take a vertex v of maximum degree in G , partition $V(G)$ on $S = N_G(v)$ and $T = V(G) \setminus S$, and add all edges between S and T while removing all edges inside S . It can be argued that this operation does not decrease the number of edges in G while also preserving the property of being K_{r+1} -free. Performing this recursively yields that $T_r(n)$ has indeed the maximum number of edges for a K_{r+1} -free graph, and this is the gist of Erdős' proof of Turán's Theorem [11]. Now, we want to extend this argument to cover our above-guarantee case. Again, we start with the graph G and perform exactly the same recursive procedure to obtain the graph G' . While we cannot say that G' is equal to G , since the latter has slightly less than $t_r(n)$ edges, we can argue that every edge that gets changed from G to G' can be attributed to the "budget" k . Thus we arrive to the conclusion that G is different from G' at only $\mathcal{O}(k)$ places. The following lemma makes this intuition formal.

Lemma 1. *There is an $\mathcal{O}(m+k)$ -time algorithm that for non-negative integers $k \geq 1$, $r \geq 2$ and an n -vertex graph G with $m \geq t_r(n) - k$ edges, finds a partition V_1, V_2, \dots, V_p of $V(G)$ with the following properties*

- (i) $p \geq r - k$;
- (ii) For each $i \in \{1, \dots, p\}$, there is a vertex $v_i \in V_i$ with $N_G(v_i) \supset V_{i+1} \cup V_{i+2} \cup \dots \cup V_p$;
- (iii) If $p \leq r$, then for the complete p -partite graph G' with parts V_1, V_2, \dots, V_p , we have $|E(G')| \geq |E(G)|$ and $|E(G) \Delta E(G')| \leq 3k$. Moreover, all vertices covered by $E(G) \setminus E(G')$ are covered by $E(G') \setminus E(G)$ and $|E(G') \setminus E(G)| \leq 2k$.

Let us clarify this technical definition. The lemma basically states that if a graph G has at least $t_r(n) - k$ edges, then it either has a clique of size $r + 1$, or it has at most $3k$ edit distance to a complete multipartite graph G' consisting of $p \in [r - k, r]$ parts. Moreover, G has a clique of size p untouched by the edit, i.e. this clique is present in the complete p -partite graph G' as well.

We should also note that Lemma 1 is close to the concept of stability of Turán's theorem. This concept received much attention in extremal graph theory (see e.g. recent work of Korándi et al. [23]), and appeals the structural properties of graphs having number of edges close to the Turán's number $t_r(n)$. Lemma 1 can also be seen as a stability version of Turán's theorem, but from the algorithmic point of view. We move on to the proof of the lemma.

Proof (of Lemma 1). First, we state the algorithm, which follows from the Erdős' proof of Turán's Theorem from [11]. We start with an empty graph G' defined on the same vertex set as G , and set $G_1 = G$. Then we select the vertex $v_1 \in V(G_1)$ as an arbitrary maximum-degree vertex in G_1 , i.e. $\deg_{G_1}(v_1) = \max_{u \in V(G_1)} \deg_{G_1}(u)$. We put $V_1 = V(G_1) \setminus N_{G_1}(v_1)$ and add to G' all edges between V_1 and $V(G_1) \setminus V_1$.

We then put $G_2 := G_1 - V_1$ and, unless G_2 is empty, apply the same process to G_2 . That is, we select $v_2 \in V(G_2)$ with $\deg_{G_2}(v_2) = \max_{u \in V(G_2)} \deg_{G_2}(u)$ and put $V_2 = V(G_2) \setminus N_{G_2}(v_2)$ and add all edges between V_2 and $V(G_2) \setminus V_2$ to G' . We repeat this process with $G_{i+1} := G_i - V_i$ until G_{i+1} is empty. The process has to stop eventually as each V_i is not empty. In this way three sequences are produced: $G = G_1, G_2, \dots, G_p, G_{p+1}$, where G_1 is G and G_{p+1} is the empty graph; v_1, v_2, \dots, v_p , and V_1, V_2, \dots, V_p . Note that the sequences $\{v_i\}$ and $\{V_i\}$ satisfy property (ii) by construction. Observe that this procedure can be clearly performed in time $\mathcal{O}(n^2)$, and for any $r \geq 2$, $m + k = t_r(n) = \Theta(n^2)$, thus the algorithm takes time $\mathcal{O}(m + k)$.

Clearly, G' is a complete p -partite graph with parts V_1, V_2, \dots, V_p as in G' we added all edges between V_i and $V(G_i) \setminus V_i = (V_{i+1} \cup V_{i+2} \cup \dots \cup V_p)$ for each $i \in \{1, \dots, p\}$ and never added an edge between two vertices in the same V_i . Since a p -partite graph is always K_{p+1} -free, by Theorem 1 $|E(G')| \leq t_p(n)$.

Claim. $|E(G')| - |E(G)| \geq \sum_{i=1}^p |E(G[V_i])|$ and for each $u \in V(G)$, $\deg_G(u) \leq \deg_{G'}(u)$.

Proof (of Claim). For each $i \in \{1, \dots, p\}$, denote by E_i the edges of G' added in the i -th step of the construction. Formally, $E_i = V_i \times (V_{i+1} \cup V_{i+2} \cup \dots \cup V_p)$ for $i < p$ and $E_p = \emptyset$. We aim to show that $|E_i| - |E(G_i) \setminus E(G_{i+1})| \geq |E(G[V_i])|$. The first part of the claim will follow as $|E(G')| = \sum_{i=1}^p |E_i|$ and $|E(G)| = \sum_{i=1}^p |E(G_i) \setminus E(G_{i+1})|$.

Denote by d_i the degree of v_i in G_i . Since $N_{G_i}(v_i) = (V_{i+1} \cup V_{i+2} \cup \dots \cup V_p)$, $|E_i| = d_i |V_i|$. As v_i is a maximum-degree vertex in G_i , $d_i \geq \deg_{G_i}(u)$ for every

$u \in V_i$, so $|E_i| \geq \sum_{u \in V_i} \deg_{G_i}(u)$. Recall that $G_{i+1} = G_i - V_i$. Then

$$\begin{aligned} |E(G_i) \setminus E(G_{i+1})| &= \sum_{u \in V_i} \deg_{G_i}(u) - |E(G_i[V_i])| = \sum_{u \in V_i} \deg_{G_i}(u) - |E(G[V_i])| \\ &\leq |E_i| - |E(G[V_i])|, \end{aligned}$$

and the first part of the claim follows.

To show the second part, note that for a vertex $u \in V_i$, $\deg_G(u) \leq \sum_{j=1}^{i-1} |V_j| + \deg_{G_i}(u)$. On the other hand, u is adjacent to every vertex in $V_1 \cup V_2 \cup \dots \cup V_{i-1} \cup V_{i+1} \cup \dots \cup V_p$ in G' . We have already seen that $|V_{i+1} \cup \dots \cup V_p| \geq \deg_{G_i}(u)$. Thus, $\deg_G(u) \leq \deg_{G'}(u)$. Proof of the claim is complete. \square

The claim yields that $|E(G)| \leq t_p(n)$, so $t_p(n) \geq t_r(n) - k$. By Theorem 1, we have that $t_i(n) > t_{i-1}(n)$, as $T_{i-1}(n)$ is distinct from $T_i(n)$, so $t_i(n) \geq t_{i-1}(n) + 1$ for every $i \in [n]$. Hence if $r \geq p$ then $k \geq t_r(n) - t_p(n) \geq r - p$. It concludes the proof of (i).

It is left to prove (iii), i.e. that $|E(G) \Delta E(G')| \leq 3k$ under assumption $p \leq r$. First note that $E(G) \setminus E(G') = \sum E(G[V_i])$. Second, since $|E(G')| \leq t_p(n) \leq t_r(n)$ and $|E(G)| \geq t_r(n) - k$, $|E(G')| - |E(G)| \leq k$. By Claim, we have that $|E(G')| - |E(G)| \geq \sum |E(G[V_i])|$. Finally

$$\begin{aligned} |E(G) \Delta E(G')| &= |E(G')| - |E(G)| + 2|E(G) \setminus E(G')| \\ &= |E(G')| - |E(G)| + 2 \sum |E(G[V_i])| \leq 3k. \end{aligned}$$

By Claim, each vertex covered by $E(G) \setminus E(G')$ is covered by $E(G') \setminus E(G)$. The total size of these edge sets is at most $3k$, while $|E(G') \setminus E(G)| - |E(G) \setminus E(G')| = |E(G')| - |E(G)| \leq k$. Hence, the size of $|E(G) \setminus E(G')|$ is at most $2k$. This concludes the proof of (iii) and of the lemma. \square

We are ready to prove our main algorithmic result. Let us recall that we seek a clique of size ℓ in an n -vertex graph with $t_r(n) - k$ edges, and that $\tau = \max\{\ell - r, 0\}$.

Theorem 2. TURÁN'S CLIQUE with $\tau \in \{0, 1\}$ admits an $\mathcal{O}(n + m)$ -time compression into CLIQUE on at most $5k$ vertices.

Proof. Let (G, r, k, ℓ) be the input instance of TURÁN'S CLIQUE. If $r < 2$ or $n \leq 5k$, a trivial compression is returned. Apply the algorithm of Lemma 1 to (G, r, k, ℓ) and obtain the partition V_1, V_2, \dots, V_p . Observe that this takes time $\mathcal{O}(m + k) = \mathcal{O}(n + m)$ since $n > 5k$. By the second property of Lemma 1, v_1, v_2, \dots, v_p induce a clique in G , so if $p \geq \ell$ we conclude that (G, r, k, ℓ) is a yes-instance. Formally, the compression returns a trivial yes-instance of CLIQUE in this case.

We now have that $r - k \leq p \leq r$. Then the edit distance between G and the complete p -partite graph G' with parts V_1, V_2, \dots, V_p is at most $3k$. Denote by X the set of vertices covered by $E(G) \Delta E(G')$. Denote $R = E(G') \setminus E(G)$ and

$A = E(G) \setminus E(G')$. We know that $|R| + |A| \leq 3k$, $|R| \leq 2k$ and $|R| \geq |A|$. By Lemma 1, R covers all vertices in X , so $|X| \leq 2|R|$.

Clearly, (G, r, k, ℓ) as an instance of TURÁN'S CLIQUE is equivalent to an instance (G, ℓ) of CLIQUE. We now apply the following two reduction rules exhaustively to (G, ℓ) . Note that these rules are an adaption of the well-known two reduction rules for the general case of CLIQUE (see, e.g., [29]). Here the adapted rules employ the partition V_1, V_2, \dots, V_p explicitly.

Reduction rule 1 *If there is $i \in [p]$ such that $V_i \not\subseteq X$ and V_i is independent in G , remove V_i from G and reduce ℓ by one.*

Reduction rule 2 *For each $i \in [p]$ with $|V_i \setminus X| > 1$, remove all but one vertices in $V_i \setminus X$ from G .*

Since the reduction rules are applied independently to parts V_1, V_2, \dots, V_p , and each rule is applied to each part at most once, clearly this can be performed in linear time. We now argue that these reduction rules always produce an equivalent instance of CLIQUE. The proof of the following claim can be found in appendix.

Claim. Reduction rule 1 and Reduction rule 2 are safe.

It is left to upperbound the size of G after the exhaustive application of reduction rules. In this process, some parts among V_1, V_2, \dots, V_p are removed from G . W.l.o.g. assume that the remaining parts are V_1, V_2, \dots, V_t for some $t \leq p$. Note that parts that have no common vertex with X are eliminated by Reduction rule 1, so $t \leq |X|$. On the other hand, by Reduction rule 2, we have $|V_i \setminus X| \leq 1$ for each $i \in [t]$.

Consider $i \in [t]$ with $|V_i \setminus X| = 1$. By Reduction rule 1, $G[V_i]$ contains at least one edge. Since V_i is independent in G' , $E(G[V_i]) \subseteq A$. Hence, the number of $i \in [t]$ with $|V_i \setminus X| = 1$ is at most $|A|$. We obtain

$$\begin{aligned} |V(G)| &= \sum_{i=1}^t |V_i| = \sum_{i=1}^t |V_i \cap X| + \sum_{i=1}^t |V_i \setminus X| \\ &\leq |X| + |A| \leq 2|R| + |A| \leq |R| + (|R| + |A|) \leq 5k. \end{aligned}$$

We obtained an instance of CLIQUE that is equivalent to (G, r, k, ℓ) and contains at most $5k$ vertices. The proof is complete. \square

Combining the polynomial compression of Theorem 2 with the algorithm of Xiao and Nagamochi [29] for INDEPENDENT SET running in $\mathcal{O}(1.1996^n)$, we obtain the following.

Corollary 1. TURÁN'S CLIQUE with $\tau \leq 1$ is solvable in time $2.49^k \cdot (n + m)$.

Proof. Take a given instance of TURÁN'S CLIQUE and compress it into an equivalent instance (G, ℓ) of CLIQUE with $|V(G)| \leq 5k$. Clearly, $(G, |V(G)| - \ell)$ is an instance of INDEPENDENT SET equivalent to (G, ℓ) . Use the algorithm from [29] to solve this instance in $\mathcal{O}(1.1996^{|V(G)|})$ running time. Since $1.1996^5 < 2.49$, the running time of the whole algorithm is bounded by $2.49^k \cdot n^{\mathcal{O}(1)}$. \square

2.2 Looking for larger cliques

In this subsection we consider the situation when $\tau > 1$. As we will see in Theorem 4, an FPT algorithm is unlikely in this case, unless we take a stronger parameterization. Here we show that TURÁN'S CLIQUE is FPT parameterized by $\tau + \xi + k$. Recall that $\xi = \lfloor \frac{n}{r} \rfloor$. Theorem 4 argues that this particular choice of the parameter is necessary.

First, we show that the difference between $t_\ell(n)$ and $t_r(n)$ can be bounded in terms of τ and ξ . This will allow us to employ Theorem 2 for the new FPT algorithm by a simple change of the parameter. A careful counting proof of the following lemma is moved to appendix due to the space restrictions.

Lemma 2. *Let n, r, ℓ be three positive integers with $r < \ell \leq n$. Let $\xi = \lfloor \frac{n}{r} \rfloor$ and $\tau = \ell - r$. Then for $\tau = \mathcal{O}(r)$, $t_\ell(n) - t_r(n) = \Theta(\tau\xi^2)$.*

The following compression algorithm is a corollary of Lemma 2 and Theorem 2. It provides a compression of size linear in k and τ .

Theorem 3. *TURÁN'S CLIQUE admits a compression into CLIQUE on $\mathcal{O}(\tau\xi^2 + k)$ vertices.*

Proof. Let (G, k, r, ℓ) be the given instance of TURÁN'S CLIQUE. If $\ell \leq r+1$, then the proof follows from Theorem 2. Otherwise, reduce (G, k, r, ℓ) to an equivalent instance $(G, k + t_\ell(n) - t_r(n), \ell, \ell)$ of TURÁN'S CLIQUE just by modifying the parameters. This is a valid instance since $|E(G)| \geq t_r(n) - k \geq t_\ell(n) - (t_\ell(n) + t_r(n) + k)$. Denote $k' = k + (t_\ell(n) - t_r(n))$. By Lemma 2, $k' = k + \mathcal{O}(\tau\xi^2)$. Apply polynomial compression of Theorem 2 to (G, k', ℓ, ℓ) into CLIQUE with $\mathcal{O}(k')$, i.e. $\mathcal{O}(\tau\xi^2 + k)$, vertices. \square

Pipelined with a brute-force algorithm computing a maximum independent set in time $\mathcal{O}(2^n)$, Theorem 3 yields the following corollary.

Corollary 2. *TURÁN'S CLIQUE is solvable in time $2^{\mathcal{O}(\tau\xi^2+k)} \cdot (n+m)$.*

2.3 Independent set above Turán's bound

Another interesting application of Theorem 3 concerns computing INDEPENDENT SET in graphs of small average degree. Recall that Turán's theorem, when applied to the complement \bar{G} of a graph G , yields a bound

$$\alpha(G) \geq \frac{n}{d+1}.$$

Here $\alpha(G)$ is the size of the largest independent set in G (the independence number of G), and d is the average vertex degree of G . Then in TURÁN'S INDEPENDENT SET, the task is for an n -vertex graph G and positive integer t to decide whether there is an independent set of size at least $\frac{n}{d+1} + t$ in G .

Theorem 3 implies a compression of TURÁN'S INDEPENDENT SET into INDEPENDENT SET. In other words, we give a polynomial time algorithm that for

an instance (G, t) of TURÁN'S INDEPENDENT SET constructs an equivalent instance (G', p) of INDEPENDENT SET with at most $\mathcal{O}(td^2)$ vertices. That is, the graph G has an independent set of size at least $\frac{n}{d+1} + t$ if and only if G' has an independent set of size p .

Corollary 3. TURÁN'S INDEPENDENT SET *admits a compression into INDEPENDENT SET on $\mathcal{O}(td^2)$ vertices.*

Proof. For simplicity, let us assume that n is divisible by $d + 1$. (For arguments here this assumption does not make an essential difference.) We select $r = \frac{n}{d+1}$, $\tau = t$, and $k = 0$. Then $d = \frac{n}{r} - 1 = \xi - 1$. The graph \overline{G} has at most $nd/2$ edges, hence G has at least $\frac{n(n-1)}{2} - nd/2 = \frac{n(n-1)}{2} - n(\xi - 1)/2 \geq t_r(n)$ edges, see Proposition 1. An independent set of size $\frac{n}{d+1} + t$ in graph \overline{G} , corresponds in graph G to a clique of size $r + t$. Since Theorem 3 provides compression into a CLIQUE with $\mathcal{O}(\tau\xi^2 + k) = \mathcal{O}(\tau\xi^2)$ vertices, for independent set and graph \overline{G} this corresponds to a compression into an instance of INDEPENDENT SET with $\mathcal{O}(td^2)$ vertices. \square

By Corollary 3, we obtain the following corollary.

Corollary 4. TURÁN'S INDEPENDENT SET *is solvable in time $2^{\mathcal{O}(td^2)} \cdot n^2$.*

3 Lower bounds

In this section, we investigate how the algorithms above are complemented by hardness results. First, observe that k has to be restricted, otherwise the TURÁN'S CLIQUE problem is not any different from CLIQUE. In fact, reducing from INDEPENDENT SET on sparse graphs, one can show that there is no $2^{o(k)}$ -time algorithm for TURÁN'S CLIQUE even when $\tau \leq 1$. (The formal argument is presented in Theorem 5.) This implies that the $2^{\mathcal{O}(k)}$ -time algorithm given by Corollary 1 is essentially tight.

Also, the difference between r and ℓ has to be restricted, as it can be easily seen that TURÁN'S CLIQUE admits no $n^{o(\ell)}$ -time algorithm even when $k = 0$, assuming ETH. This is observed simply by considering the special case of TURÁN'S CLIQUE where $r = 1$, there the only restriction on G is that $|E(G)| \geq t_r(n) - k = 0$, meaning that the problem is as hard as CLIQUE. However, Theorem 4 shows that even for any fixed $\tau \geq 2$ and $k = 0$ TURÁN'S CLIQUE is NP-complete. This motivates Theorem 3, where the exponential part of the running time has shape $2^{\mathcal{O}(\tau\xi^2k)}$. In the rest of this section, we further motivate the running time of Theorem 3. First, in Theorem 4 we show that not only setting τ and k to constants is not sufficient to overcome NP-hardness, but also that the same holds for any choice of two parameters out of $\{\tau, \xi, k\}$.

Theorem 4. TURÁN'S CLIQUE *is NP-complete. Moreover, it remains NP-complete in each of the following cases*

- (i) *for any fixed $\xi \geq 1$ and $\tau = 0$;*

- (ii) for any fixed $\xi \geq 1$ and $k = 0$;
 (iii) for any fixed $\tau \geq 2$ and $k = 0$.

Proof. Towards proving (i) and (ii), we provide a reduction from CLIQUE. Let $\xi \geq 1$ be a fixed constant. Let (G, ℓ) be a given instance of CLIQUE and let $n = |V(G)|$. We assume that $\ell \geq \xi$, otherwise we can solve (G, ℓ) in polynomial time. Construct a graph G' from G as follows. Start from $G' = G$ and $\ell' = \ell$. Then add $\max\{\xi\ell - n, 0\}$ isolated vertices to G' . Note that (G, k) and (G', k') are equivalent and $|V(G')| \geq \xi\ell'$. If we have $\xi\ell' \leq |V(G')| < (\xi + 1)\ell'$, we are done with the construction of G' . Otherwise, repeatedly add a universal vertex to G' , increasing ℓ' by one, so $|V(G')| - (\xi + 1)\ell'$ decreases by ξ each time. We repeat this until $|V(G')|$ becomes less than $(\xi + 1)\ell'$. Since the gap between $\xi\ell'$ and $(\xi + 1)\ell'$ is at least ξ at any moment, we derive that $\xi\ell' \leq |V(G')| < (\xi + 1)\ell'$. The construction of G' is complete. Note that (G', ℓ') is an instance of CLIQUE equivalent to (G, ℓ) . We added at most $\max\{n, \xi\ell\}$ vertices to G' , hence this is a polynomial-time reduction.

By the above, $\lfloor |V(G')|/\ell' \rfloor = \xi$, so we can reduce (G', ℓ') to an equivalent instance $(G', \ell', \binom{|V(G')|}{2}, \ell')$ of TURÁN'S CLIQUE. Clearly, this instance have the required fixed value of ξ and $\tau = 0$. This proves (i). For (ii), we use the fact that $t_1(n) = 0$ for every $n > 0$ and reduce (G', ℓ') to $(G', 1, 0, \ell')$.

To show (iii), we need another reduction from CLIQUE. Let $\tau \geq 2$ be a fixed integer constant. Take an instance (G, ℓ) of CLIQUE with $\ell \geq 2\tau$. We denote $n = |V(G)|$. To construct G' from G , we start from a large complete $(\ell - 1)$ -partite graph with equal-sized parts. The size of each part equals x , so $|V(G')| = (\ell - 1)x$. We denote $N = |V(G')|$ and choose the value of x later, for now we only need that $N \geq n$. Clearly, $|E(G')| = t_{\ell-1}(N)$ at this point. To embed G into G' , we select arbitrary n vertices in G' and make them isolated. This removes at most $n(\ell - 2)x$ edges from G' . Then we identify these n isolated vertices with $V(G)$ and add edges of G between these vertices in G' correspondingly. This operation does not decrease $|E(G')|$. This completes the construction of G' . Since G' is isomorphic to a complete $(\ell - 1)$ -partite graph united disjointly with G , we have that (G, ℓ) and (G', ℓ) are equivalent instances of CLIQUE.

We now want to reduce (G', ℓ) to an instance $(G', \ell - \tau, 0, \ell)$ of TURÁN'S CLIQUE. To do so, we need $|E(G')| \geq t_{\ell-\tau}(N)$. By Lemma 2, $t_{\ell-1}(N) - t_{\ell-\tau}(N) \geq C \cdot (\tau - 1) \cdot \left(\frac{N}{\ell - \tau}\right)^2$ for some constant $C > 0$. Since $|E(G')| \geq t_{\ell-1}(N) - n(\ell - 2)x$, we want to choose x such that

$$n(\ell - 2)x \leq C \cdot (\tau - 1) \cdot \left(\frac{N}{\ell - \tau}\right)^2.$$

By substituting $N = (\ell - 1)x$, we derive that x should satisfy

$$\frac{n}{C} \cdot \frac{(\ell - 2)(\ell - \tau)}{(\ell - 1)^2} \cdot \frac{\ell - \tau}{\tau - 1} \leq x.$$

Now simply pick as x the smallest integer that satisfies the above. Then $(G', \ell - \tau, 0, \ell)$ is an instance of TURÁN'S CLIQUE that is equivalent to the instance (G, k) of CLIQUE and is constructed in polynomial time. \square

Now, recall that Theorem 3 gives an FPT-algorithm for TURÁN'S CLIQUE that is single-exponential in $\tau\xi^2 + k$. The previous theorem argues that all three of τ , ξ , k have to be in the exponential part of the running time. However, that result does not say anything about what can be the best possible dependency on these parameters. The next Theorem 5 aims to give more precise lower bounds based on ETH, in particular it turns out that the dependency on τ and k cannot be subexponential unless ETH fails. First, we need to show the relation between the parameter ξ and the average degree of \overline{G} . The proof of the following proposition is moved to appendix due to the space restrictions.

Proposition 2. *Let G be an n -vertex graph, $r \leq n$ be an integer, and denote $\xi = \lfloor \frac{n}{r} \rfloor$. Let \overline{G} denote the complement of G and \overline{d} denote the average degree of \overline{G} . Then $\overline{d} \leq \xi$ if $|E(G)| \geq t_r(n)$ and $|E(G)| \geq t_r(n)$ if $\overline{d} \leq \xi - 1$.*

We are ready to give lower bounds for algorithms solving TURÁN'S CLIQUE in terms of the parameters τ , ξ , and k .

Theorem 5. *Unless the Exponential Time Hypothesis fails, for any function f there is no $f(\xi, \tau)^{o(k)} \cdot n^{f(\xi, \tau)}$, $f(\xi, k)^{o(\tau)} \cdot n^{f(\xi, k)}$, or $f(k, \tau)^{o(\sqrt{\xi})} \cdot n^{f(k, \tau)}$ algorithm for TURÁN'S CLIQUE.*

The proof of this result is moved to appendix due to the space restrictions. It is based on the proof of Theorem 4, but is much more careful to details and contains some new ideas. Moreover, the proof of the first point of the theorem lets us observe that our $2.49^k \cdot (n + m)$ -time algorithm for TURÁN'S CLIQUE with $\tau \leq 1$ is essentially tight.

Corollary 5. *Assuming ETH, there is no $2^{o(k)} \cdot n^{O(1)}$ algorithm for TURÁN'S CLIQUE with $\ell \leq r + 1$.*

4 Conclusion

We conclude by summarizing natural questions left open by our work. Theorem 5 rules out (unless ETH fails) algorithms with running times subexponential in τ and k . However, when it comes to ξ , the dependency in the upper bound of Corollary 2 is $2^{\mathcal{O}(\tau\xi^2+k)} \cdot n^{O(1)}$, while Theorem 5 only rules out the running time of $f(k, \tau)^{o(\sqrt{\xi})} \cdot n^{f(k, \tau)}$ under ETH. Thus, whether the correct dependence in ξ is single-exponential or subexponential, is left open. Similarly, the question whether TURÁN'S CLIQUE admits a compression into CLIQUE whose size is linear in ξ , τ , and k , is open. A weaker variant of this question (for the case $k = 0$) for TURÁN'S INDEPENDENT SET, whether it admits a compression or kernel linear in d and in t , is also open.

References

1. N. ALON, G. GUTIN, E. J. KIM, S. SZEIDER, AND A. YEO, *Solving MAX-r-SAT above a tight lower bound*, *Algorithmica*, 61 (2011), pp. 638–655.
2. H. L. BODLAENDER, F. V. FOMIN, D. LOKSHTANOV, E. PENNINKX, S. SAURABH, AND D. M. THILIKOS, *(Meta) Kernelization*, *J. ACM*, 63 (2016), pp. 44:1–44:69.
3. H. L. BODLAENDER, B. M. P. JANSEN, AND S. KRATSCH, *Kernelization lower bounds by cross-composition*, *SIAM Journal on Discrete Mathematics*, 28 (2014), pp. 277–305.
4. L. R. BROOKS, *On colouring the nodes of a network*, *Proc. Cambridge Philos. Soc.*, 37 (1941), pp. 194–197.
5. R. CROWSTON, M. JONES, G. MUCIACCIA, G. PHILIP, A. RAI, AND S. SAURABH, *Polynomial kernels for lambda-extendible properties parameterized above the Poljak-Turzik bound*, in *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, vol. 24 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Germany, 2013, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 43–54.
6. M. CYGAN, F. V. FOMIN, L. KOWALIK, D. LOKSHTANOV, D. MARX, M. PILIPCZUK, M. PILIPCZUK, AND S. SAURABH, *Parameterized Algorithms*, Springer, 2015.
7. E. D. DEMAINE, F. V. FOMIN, M. HAJIAGHAYI, AND D. M. THILIKOS, *Subexponential parameterized algorithms on graphs of bounded genus and H-minor-free graphs*, *J. ACM*, 52 (2005), pp. 866–893.
8. R. G. DOWNEY AND M. R. FELLOWS, *Parameterized complexity*, Springer-Verlag, New York, 1999.
9. Z. DVORÁK AND B. LIDICKÝ, *Independent sets near the lower bound in bounded degree graphs*, in *Proceedings of the 34th International Symposium on Theoretical Aspects of Computer Science (STACS)*, vol. 66 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017, pp. 28:1–28:13.
10. Z. DVORÁK AND M. MNICH, *Large independent sets in triangle-free planar graphs*, *SIAM J. Discret. Math.*, 31 (2017), pp. 1355–1373.
11. P. ERDŐS, *On the graph theorem of Turán*, *Mat. Lapok*, 21 (1970), pp. 249–251.
12. F. V. FOMIN, P. A. GOLOVACH, D. LOKSHTANOV, F. PANOLAN, S. SAURABH, AND M. ZEHAVI, *Going far from degeneracy*, *SIAM J. Discrete Math.*, 34 (2020), pp. 1587–1601.
13. F. V. FOMIN, D. LOKSHTANOV, S. SAURABH, AND M. ZEHAVI, *Kernelization: theory of parameterized preprocessing*, Cambridge University Press, 2019.
14. S. GARG AND G. PHILIP, *Raising the bar for vertex cover: Fixed-parameter tractability above a higher guarantee*, in *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SIAM, 2016, pp. 1152–1166.
15. G. GUTIN, E. J. KIM, M. LAMPIS, AND V. MITSOU, *Vertex cover problem parameterized above and below tight bounds*, *Theory of Computing Systems*, 48 (2011), pp. 402–410.
16. G. GUTIN, L. VAN IERSEL, M. MNICH, AND A. YEO, *Every ternary permutation constraint satisfaction problem parameterized above average has a kernel with a quadratic number of variables*, *J. Computer and System Sciences*, 78 (2012), pp. 151–163.

17. G. Z. GUTIN AND V. PATEL, *Parameterized traveling salesman problem: Beating the average*, SIAM J. Discrete Math., 30 (2016), pp. 220–238.
18. G. Z. GUTIN, A. RAFIEY, S. SZEIDER, AND A. YEO, *The linear arrangement problem parameterized above guaranteed value*, Theory Comput. Syst., 41 (2007), pp. 521–538.
19. J. HÅSTAD, *Clique is hard to approximate within $n^{1-\epsilon}$* , Acta Math., 182 (1999), pp. 105–142.
20. R. IMPAGLIAZZO, R. PATURI, AND F. ZANE, *Which problems have strongly exponential complexity*, J. Computer and System Sciences, 63 (2001), pp. 512–530.
21. B. M. P. JANSEN, L. KOZMA, AND J. NEDERLOF, *Hamiltonicity below Dirac’s condition*, in Proceedings of the 45th International Workshop on Graph-Theoretic Concepts in Computer Science (WG), vol. 11789 of Lecture Notes in Computer Science, Springer, 2019, pp. 27–39.
22. R. M. KARP, *Reducibility among combinatorial problems*, in Complexity of computer computations, Plenum Press, New York, 1972, pp. 85–103.
23. D. KORÁNDI, A. ROBERTS, AND A. SCOTT, *Exact stability for turán’s theorem*, Advances in Combinatorics, (2021), p. 31079.
24. D. LOKSHTANOV, N. S. NARAYANASWAMY, V. RAMAN, M. S. RAMANUJAN, AND S. SAURABH, *Faster parameterized algorithms using linear programming*, ACM Trans. Algorithms, 11 (2014), pp. 15:1–15:31.
25. M. MAHAJAN AND V. RAMAN, *Parameterizing above guaranteed values: MaxSat and MaxCut*, J. Algorithms, 31 (1999), pp. 335–354.
26. M. MAHAJAN, V. RAMAN, AND S. SIKDAR, *Parameterizing above or below guaranteed values*, J. Computer and System Sciences, 75 (2009), pp. 137–153.
27. M. PILIPCZUK AND S. SIEBERTZ, *Kernelization and approximation of distance- r independent sets on nowhere dense graphs*, Eur. J. Comb., 94 (2021), p. 103309.
28. P. TURÁN, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok, 48 (1941), pp. 436–452.
29. M. XIAO AND H. NAGAMUCHI, *Exact algorithms for maximum independent set*, Inf. Comput., 255 (2017), pp. 126–146.

Appendix A Proof of Lemma 2

Proof. Throughout the proof, we assume $\xi = \frac{n}{r}$ since this does not influence the desired Θ estimation. Let s_r be the remainder in the division of n by r and s_ℓ be the remainder in the division of n by ℓ . By Proposition 1,

$$t_\ell(n) - t_r(n) = \frac{\tau n^2}{2r\ell} + \left(\frac{s_r}{2} \cdot \left(1 - \frac{s_r}{r} \right) - \frac{s_\ell}{2} \cdot \left(1 - \frac{s_\ell}{\ell} \right) \right). \quad (1)$$

The first summand in (1) is $\Theta(\xi^2\tau)$. Indeed, since $\tau = \mathcal{O}(r)$ we have

$$\frac{\tau n^2}{2r\ell} = \frac{\tau}{2} \cdot \frac{n}{r} \cdot \frac{n}{r + \tau} = \frac{\xi^2\tau}{2} \cdot \frac{r}{r + \tau} = \Theta(\xi^2\tau). \quad (2)$$

For the second summand,

$$\begin{aligned} \frac{s_r}{2} \cdot \left(1 - \frac{s_r}{r}\right) - \frac{s_\ell}{2} \cdot \left(1 - \frac{s_\ell}{\ell}\right) &= \frac{\ell s_r(r - s_r) - r s_\ell(\ell - s_\ell)}{2r\ell} = \frac{(rs_\ell^2 - \ell s_r^2) + r\ell(s_r - s_\ell)}{2r\ell} \\ &= \frac{(rs_\ell^2 - rs_r^2 - \tau s_r^2) + r\ell(s_r - s_\ell)}{2r\ell} \end{aligned} \quad (3)$$

$$\begin{aligned} &= \frac{r(s_\ell - s_r)(s_\ell + s_r) + r\ell(s_r - s_\ell)}{2r\ell} - \frac{\tau s_r^2}{2r\ell} \\ &= \frac{(s_r - s_\ell)(\ell - (s_\ell + s_r))}{2\ell} - \frac{\tau s_r^2}{2r\ell}. \end{aligned} \quad (4)$$

Since $n = \lfloor \frac{n}{\ell} \rfloor \cdot \ell + s_\ell$, we have that

$$s_r \equiv \left\lfloor \frac{n}{\ell} \right\rfloor \cdot \ell + s_\ell \pmod{r},$$

and

$$s_r \equiv \left\lfloor \frac{n}{\ell} \right\rfloor \cdot (r + \tau) + s_\ell \pmod{r}.$$

Hence,

$$s_r - s_\ell \equiv \left\lfloor \frac{n}{\ell} \right\rfloor \cdot \tau \pmod{r}.$$

By definition $s_r < r$, thus we get from the above that $s_r - s_\ell \leq \lfloor \frac{n}{\ell} \rfloor \cdot \tau \leq \xi\tau$.

Analogously,

$$s_\ell - s_r \equiv \left\lfloor \frac{n}{r} \right\rfloor \cdot (-\tau) \pmod{\ell}$$

Since $s_\ell - s_r > -r > -\ell$, we have that $s_\ell - s_r \geq \lfloor \frac{n}{r} \rfloor \cdot (-\tau) \geq -\xi\tau$. Therefore $|s_\ell - s_r| \leq \xi\tau$. It is easy to see that $|\ell - (s_\ell + s_r)| \leq \ell + (s_\ell + s_r) \leq 3\ell$. Finally, $\frac{\tau s_r^2}{2r\ell}$ is non-negative and is upper bounded by $\frac{\tau r^2}{2r\ell} \leq \frac{\tau}{2}$. Thus, the absolute value of (4), is at most

$$\frac{\xi\tau \cdot 3\ell}{2\ell} + \frac{\tau}{2} = \mathcal{O}(\xi\tau).$$

By putting together (2) and (4), we conclude that $t_\ell(n) - t_r(n) = \Theta(\xi^2\tau) + \mathcal{O}(\xi\tau) = \Theta(\xi^2\tau)$. \square

Appendix B Proof of Claim in Theorem 2

Proof. For Reduction rule 1, note that there is a vertex $v \in V_i \setminus X$ such that $N_G(v) = N_G(V_i) = V(G) \setminus V_i$. Since V_i is independent, for any vertex set C that induces a clique in G , we have $|C \cap V_i| \leq 1$. On the other hand, if $C \cap V_i = \emptyset$, $C \cup \{v\}$ also induces a clique in G as $C \subseteq N_G(v)$. Hence, any maximal clique in G contains exactly one vertex from V_i , so Reduction rule 1 is safe.

To see that Reduction rule 2 is safe, observe that $N_G(u) = N_G(v)$ for any two vertices $u, v \in V_i \setminus X$. Then no clique contains both u and v , and if $C \ni v$ induces a clique in G , $C \setminus \{v\} \cup \{u\}$ also induces a clique in G of the same size. Hence, v can be safely removed from G so Reduction rule 2 is safe. \square

Appendix C Proof of Proposition 2

Proof. Let $s < r$ be the remainder in the division of n by r . Then

$$\begin{aligned} \binom{n}{2} - t_r(n) &= \frac{n^2}{2} - \frac{n}{2} - \left(1 - \frac{1}{r}\right) \cdot \frac{n^2}{2} + \frac{s}{2} \cdot \left(1 - \frac{s}{r}\right) = \frac{n}{2} \cdot \frac{n}{r} - \frac{n}{2} + \frac{s}{2} \cdot \left(1 - \frac{s}{r}\right) \\ &= \frac{n}{2} \cdot \xi + \frac{n}{2} \cdot \frac{s}{r} - \frac{n}{2} + \frac{s}{2} - \frac{s}{2} \cdot \frac{s}{r} = \frac{n}{2} \cdot \xi - \frac{n-s}{2} \cdot \left(1 - \frac{s}{r}\right) = \\ &= \frac{n}{2} \cdot \xi - \frac{(n-s)(r-s)}{2r} = \frac{n}{2} \cdot \xi - \frac{r-s}{2} \cdot \xi. \end{aligned}$$

Since $|E(\overline{G})| = \binom{n}{2} - |E(G)|$, one direction is proved: assuming $|E(G)| \geq t_r(n)$, $|E(\overline{G})| \leq \binom{n}{2} - t_r(n) \leq \frac{\xi n}{2}$.

For the other direction, assume that $\bar{d} \leq \xi - 1$, i.e. $|E(\overline{G})| \leq \frac{n}{2} \cdot (\xi - 1)$. As $\binom{n}{2} - t_r(n) \geq \frac{n}{2} \cdot \xi - \frac{n}{2}$, we have $\binom{n}{2} - t_r(n) \geq |E(\overline{G})|$. Then $|E(G)| \geq t_r(n)$ follows and the proof is complete. \square

Appendix D Proof of Theorem 5

Proof. It is well-known that under ETH INDEPENDENT SET cannot be solved in $2^{o(n)}$ time on instances with linear number of edges [6]. This is a basis of our proof: we provide several reductions from INDEPENDENT SET with linear number of edges. Note that these reductions mostly repeat the reductions given in the proof of Theorem 4 but are different in terms of requirements for τ, ξ and k . In fact, we give three polynomial-time algorithms that reduce an instance (G, q) of INDEPENDENT SET, where $n = |V(G)|$ and $|E(G)| = \mathcal{O}(n)$, to an equivalent instance of TURÁN'S CLIQUE such that

- τ, ξ are constant but $k = \mathcal{O}(n)$;
- ξ, k are constant but $\tau = \mathcal{O}(q)$;
- τ, k are constant but $\xi = \mathcal{O}(nq)$.

Clearly, once we show these three reductions, the proof of the theorem is complete.

For an instance of INDEPENDENT SET (G, q) we denote $n = |V(G)|$ and $m = |E(G)|$. We always assume that the number of edges in G is linear, so $m = \mathcal{O}(n)$ and the average degree of G is not greater than some fixed constant D . We also assume that $n \geq 2(D+1)$ and $q > 2$.

The first reduction takes (G, q) and trivially reduces it to an instance (\overline{G}, n, m, q) of TURÁN'S CLIQUE. Note that $|E(\overline{G})| = \binom{n}{2} - m = t_n(n) - m$, so this is a valid instance of the problem. For this instance, $\xi = 1$ and $\tau = 0$ but $k = m = \mathcal{O}(n)$.

The second algorithm reduces (G, q) to an equivalent instance $(\overline{G}, r, 0, q)$, where $r = \lfloor \frac{n}{D+1} \rfloor$. As $\lfloor n/r \rfloor - 1$ upper bounds the average degree of G , by Proposition 2 we have that $|E(\overline{G})| \geq t_r(n)$, so $(\overline{G}, r, 0, q)$ is a valid instance. This instance has $k = 0$ and

$$\xi = \lfloor n/r \rfloor < n \cdot \left(\frac{n}{D+1} - 1 \right)^{-1} = (D+1) \cdot \frac{n}{n - (D+1)} \leq 2(D+1),$$

but $\tau \leq q = \mathcal{O}(q)$.

To show the last reduction, we reduce from the instance (G, ℓ) of CLIQUE instead of (G, q) of INDEPENDENT SET, since constraints on the number of edges in G are not necessary for it. Formally it means that we reduce from (G, q) of INDEPENDENT SET to (\overline{G}, q) of CLIQUE, and then apply reductions as required. Slightly abusing the notion we denote (\overline{G}, q) by (G, ℓ) .

We adjust the last reduction from CLIQUE from the proof of Theorem 4. Recall that in this reduction we reduce an instance (G, ℓ) of CLIQUE to an equivalent instance (G', ℓ) of CLIQUE with $|V(G')| = (\ell - 1)x$ for some chosen integer x . For $(G', \ell - \tau, 0, \ell)$ to be a valid equivalent instance of TURÁN'S CLIQUE, it is enough that x satisfies

$$x \geq \frac{n}{C} \cdot \frac{\ell - \tau}{\ell - 1} \cdot \frac{\ell - \tau}{\tau - 1},$$

where the fixed constant $C > 0$ comes from Lemma 2.

To show the third reduction, we pick $\tau := 2$. Then we choose $x := \lceil n\ell/C \rceil$, so $(G', \ell - \tau, 0, \ell)$ is a valid instance of TURÁN'S CLIQUE equivalent to (G, ℓ) . This instance has $k = 0$ and $\tau = 2$, but $\xi \leq |V(G')|/\ell < x = \mathcal{O}(n\ell)$. \square