Inapproximabilty of Densest κ-Subgraph from Average Case Hardness

Noga Alon [*] Sanjeev Arora[†] Rajsekar Manokaran[†] Dana Moshkovitz[‡]
Omri Weinstein[†]

December 8, 2011

Abstract

We establish two results about the inapproximability of the Densest κ-Subgraph (DκS) problem. Both results are of similar flavor: ruling out constant factor approximations in polynomial time for the DκS problem under an "average case" hardness assumption.

The first result asserts that if Random k-AND formulas are hard to distinguish from ones that are $2^{-c\sqrt{k}}$ satisfiable, then the Densest κ-Subgraph problem is hard to approximate to within any constant factor.

The second result, which is of a similar flavor, asserts that if the problem of finding a planted clique of size $n^{1/3}$ in the random graph $\mathcal{G}(n, 1/2)$ is hard, then so is approximating the Densest κ-Subgraph to within any constant factor, for a subgraph of size $\kappa = N^{1-\epsilon}$ for any $2/3 \geq \epsilon > 0$ in an N vertex graph. Depending on the hardness of the Hidden Clique problem, this result carries over to superconstant hardness factors for approximating DκS. Our result also implies the optimality (assuming appropriate hardness of the planted clique problem) of an existing algorithm by Feige and Seltser [FS97], for the problem of distinguishing between a graph containing a clique of size κ and one in which the densest subgraph of size κ is of density at most δ.

Both results are based on gap-amplification arguments: we believe that these arguments can be useful elsewhere as well.

[*] Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and Institute for Advanced Study, Princeton, New Jersey, 08540, USA. Email: nogaa@tau.ac.il. Research supported in part by an ERC Advanced grant, by a USA-Israeli BSF grant and by NSF grant No. DMS-0835373.

[†] Princeton University. Email: {arora,rajsekar,oweinste}@cs.princeton.edu

[‡] MIT. Email: dmoshkov@mit.edu

1 Introduction

In the *Densest κ-Subgraph Problem* (DκS) we are given a graph $G = (V, E)$ and a parameter κ, and need to find a subset $S \subseteq V$ of size κ that contains the most edges among all such subsets. This is a natural extension of the MAXCLIQUE problem.

This problem is NP-hard, as a simple consequence of the NP-hardness of MAXCLIQUE, see [FPK01], and therefore attention has focused on approximation algorithms. Since it is a bicriterion problem, approximation could be defined in multiple ways. Throughout this paper, by an α-factor approximation algorithm we mean one that outputs a set of size κ that contains at least $1/\alpha$ times as many edges as the optimum subset. Current algorithms only compute fairly weak approximations; the best algorithm (due to Bhaskara et al [BCC+10]) computes an $O(n^{1/4})$ approximation in quasi-polynomial time (and an $O(n^{1/4+\epsilon})$ approximation in polynomial time). These algorithms rely on state-of-the-art techniques including spectral methods and lift and project machinery, and nevertheless achieve rather poor guarantees. Thus researchers tend to believe that the problem is hard to approximate.

The hardness of the problem has also been found to have important consequences, especially hardness of the "planted" version whereby a large dense subgraph is planted in a random graph and the algorithm has to discover it. Note that a good approximation algorithm would find such a set since the random graph in question is extremely unlikely to have another subgraph of even remotely the same density. Assuming this planted problem is hard, Applebaum, Barak and Wigderson [ABW10] proposed a new public-key cryptosystem, and Arora, Barak, Brunnermeier, and Ge [ABBG10] showed that derivative pricing is hard on "real-life" distributions.

Unfortunately, the known inapproximability results for the problem are very weak. In his paper on the connection between average-case complexity and inapproximability, Feige showed that computing a $1 + \epsilon$-approximation is at least as hard as refuting random 3-SAT clauses [Fei02] (for some $\epsilon > 0$). In his paper on quasirandom PCPs, Khot attempted to prove hardness results similar to Feige's while relying on worst-case hardness assumptions, rather than on average-case ones. He was only able to show that a $1 + \epsilon$-approximation is hard assuming $NP \not\subseteq \cap_{\epsilon>0} DTIME(2^{n^\epsilon})$. Assuming what is called the Small Set Expansion Conjecture, Raghavendra et al [RST10] rule out all constant factor approximation for DκS.

In this paper, we provide additional evidence showing that the Densest κ-Subgraph problem is hard to approximate to within any constant factor in polynomial time. We actually present two independent pieces of evidence which both rely on the intractability of other, better known average-case problems: random k-AND, and hidden CLIQUE.

While incomparable (upto the current state of knowledge), the two conjectures have their merits that we would like to point out. The hidden CLIQUE problem has received lot of attention (refer below for a short exposition) in the past few decades and still resisted polynomial time algorithms making it a natural average case assumption to use.

However, the hidden CLIQUE problem is solvable in $n^{O(\log n)}$ time, and hence the inapproximability obtained using it can provide hardness of at most quasi-polynomial time. On the other hand, the random k-AND problem has no sub-exponential time algorithms and there is a good evidence suggesting why (refer [Tul09]). The reduction from k-AND thus suggests that DκS is hard to approximate well by even sub-exponential time algorithms.

Our reduction from the hidden CLIQUE problem provides much higher factors of inapproximability for DκS than the known ones: assuming no $n^{o(\log n)}$ algorithm solves the hidden CLIQUE problem, the DκS problem cannot be approximated even upto a $2^{(\log n)^{2/3}}$ factor in polynomial time.

Finally, both results involve a gap-amplification technique that seems of independent interest for future study of the densest subgraph problem and related ones. We now describe the two complexity assumptions and state our main results. The rest of the paper contains the proofs.

Complexity of Random k-AND Formulas. This complexity assumption was introduced by Feige (Hypothesis 3 in [Fei02]) in his study of random 3-SAT formulas. Our reduction to densest subgraph is analogous to Feige's reduction, combined with a gap-amplification technique using graph powering.

A k-AND formula is a collection of clauses, each containing k literals, where each literal is either a variable or its negation. A boolean assignment to the variables is said to satisfy a clause if every literal is set to true in the assignment.

A random k-AND formula with m clauses and n variables is picked by picking each of the literals in each of the clauses independently at random from the $2n$ literals.

For m large enough compared to n, at most $2^{-k}m(1+o(1))$ clauses can be satisfied simultaneously by any assignment to the variables. On the other hand, it is NP-hard to distinguish ($\sim 2^{-k}m$)-satisfiable instances from ($\sim 2^{-c\sqrt{k}}m$)-satisfiable instances for some $c > 0$ (see [ST00]). Feige conjectures that even random instances (which are $\sim 2^{-k}m$-satisfiable) are indistinguishable from ($\sim 2^{-c\sqrt{k}}m$)-satisfiable instances.

Hypothesis 1.1. *For some constant $c > 0$, for every k, for Δ a sufficiently large constant independent of n, there is no polynomial time algorithm that on most random k-AND formulas with n variables and $m = \Delta n$ clauses outputs* `typical`*, but never outputs* `typical` *on k-AND formulas with at least $m/2^{c\sqrt{k}}$ satisfiable clauses.*

Our reduction will exploit the structure (or lack thereof to be more precise) in random instances: for example, in such instances, *every* assignment would set roughly half the literals in most clauses to true. This fact is used to prove:

Theorem 1.2. *If Hypothesis 1.1 is true, then no polynomial time approximation algorithm achieves a constant factor approximation for the DκS problem.*

The Hidden Clique Problem. The problem of finding a hidden clique in a random graph has been open since the works of [Jer92], [Kuc95]. The input for this problem is a graph G obtained by planting a clique of size t in the random graph $\mathcal{G}(n, 1/2)$, where t is much bigger than the typical clique number of $\mathcal{G}(n, 1/2)$, which is roughly $2\log_2 n$. The objective is to find the clique, or, more modestly, to distinguish between the random graph $\mathcal{G}(n, 1/2)$ and the random graph with the t-clique planted in it. For $t = o(\sqrt{n})$, there is no known polynomial time algorithm that finds even a $(1+\epsilon)\log_2 n$ clique, for any constant $\epsilon > 0$. When $t = \Theta(\sqrt{n})$, the authors of [AKS98] describe a polynomial time algorithm that does find the hidden clique of size t. Subsequent algorithms with similar performance appear in [FK00] , [FR10], [DGGP10].

It is widely believed that there is no polynomial time algorithm that solves the hidden clique problem, even when t is as large as n^c for any fixed $c < 1/2$, see [AAK$^+$07, AKS98, HK11, Jer92, Kuc95, FK00, DGGP10]. Indeed, there are some known hardness results for various computational problems assuming the hidden clique problem is hard. In [AAK+07] it is shown that hardness of the hidden clique problem implies a hardness result for the problem of deciding whether a given distribution is close to being k-wise independent. In [HK11] it is shown that hardness of the hidden clique problem implies hardness of the problem of approximating best Nash Equilibrium.

In the present paper we prove that hardness of the hidden clique problem implies hardness of approximation of the densest κ-subgraph problem to within any constant factor.

Theorem 1.3. *If there is no polynomial time algorithm for solving the hidden clique problem for a planted clique of size $n^{1/3}$, then for any $2/3 \geq \epsilon > 0, \delta > 0$, there is no polynomial time algorithm that distinguishes between a graph G on N vertices containing a clique of size $\kappa = N^{1-\epsilon}$, and a graph G' on N vertices in which the densest subgraph on κ vertices has density at most δ.*

The hidden clique problem can be easily solved in time $n^{O(\log n)}$, by simply enumerating all subsets of size, say, $3 \log n$ of the given input graph, to check if there is a clique of size at least $3 \log n$ (which is, with high probability, a subset of the planted clique). Therefore, our proof of the theorem above, described in Section 3, only establishes a conditional quasi-polynomial hardness for the densest κ-subgraph problem. In fact, even assuming that the best running time of an algorithm for solving the hidden clique problem with $t = n^{1/3}$ is $n^{\Omega(\log n)}$ (which is the best known running time of an algorithm for the problem), our proof only provides an $N^{c(\epsilon,\delta)\log N}$ lower bound for the running time of any algorithm for the instances of the κ-densest subgraph problem described in the theorem. This is tight, as these instances deal with the problem of distinguishing between a graph containing a clique of size κ and one in which the densest subgraph of size κ is of density at most δ. For this problem, there is a simple elegant algorithm of Feige and Seltser [FS97] that solves the problem in time $N^{c(\delta)\log N}$ for any $\delta < 1$.

2 Reduction from random k-AND formulas

The reduction from k-AND formulas to DκS is along the lines of Feige's followed by a gap amplification technique. We will describe the two parts and how we put them together once we set up notation.

2.1 Notation and Setup

Graphs, Subsets and Powers For an integer m, let $[m]$ denotes the set $\{1, 2, \ldots, m\}$. For a set S and a natural number t, let S^t denotes the set of t-tuples of S: $\{(x_1, x_2, \ldots x_t) \mid x_i \in S\}$.

Given a bipartite graph, $G = (A \cup B, E)$ and a subset $S \cup T$, $S \subseteq A$, $T \subseteq B$, we will be interested in the number of edges in the subset as a function of the sizes of the sets S and T. We will represent S by its indicator function $f : A \to [0, 1]$ and similarly, T by g. The number of edges inside $S \cup T$ is therefore $|E| \cdot \mathbf{E}_{(a,b)}[f(a)g(b)]$, where the expectation is over a random edge $(a, b) \in E$. The size of the set S is represented by $\mu_G(f) \overset{\text{def}}{=} 1/|A| \sum_{a \in A} f(a)$, and similarly for T. We will drop the subscripts when the graph in question is clear from the context.

Given G as above, $G^{\otimes \ell}$ is the graph $(A^\ell \cup B^\ell, E_\ell)$ where vertices $(a_1, a_2, \ldots a_\ell)$ and $(b_1, b_2, \ldots b_\ell)$ are connected if for every i, (a_i, b_i) is in E. Subsets of these graphs are represented by functions $f_\ell : A^\ell \to [0, 1]$ and similarly g_ℓ.

Random k-degree Bipartite Graphs Given integers m, n, and k, $\mathcal{G}_{m,n}^k$ is the ensemble of bipartite graphs over $[m] \cup [n]$ where each vertex on the left, $a \in [m]$, is connected to k random vertices from $[n]$ picked at random with replacement. It is well-known that in almost all graphs in this family, the number of edges contained in subsets is strongly determined by the size of the subset. We will need an analog of this fact for "fractional subsets", which we make precise as follows.

Theorem 2.1. *There exists (non-negative) constants k_0 and c_0 such that for every $k \geq k_0$ there exists $\Delta_0 = \Delta_0(k)$ such that for every $m \geq \Delta_0 n$, a graph $G = (A \cup B, E)$ randomly picked from $\mathcal{G}_{m,n}^k$ satisfies the following property with probability at least $1 - exp(-c_0 n)$:*

For every $f : A \to [0,1]$ and $g : B \to [0,1]$ (to be thought of as "fractional subsets"),

$$\mathbf{E}_{(a,b)\in E}[f(a)g(b)] \leq \mu(f)\mu(g) + \frac{1}{k^{1/9}}\mu(f) + 2^{-k^{3/4}}$$

Proof Outline. The expected value of the left hand side is exactly $\mu(f)\mu(g)$. Further, for large enough k, and $\mu(f)$, $\mu(g)$ big enough (say, both $\Omega(1)$), the quantity on the left can be shown to be tightly concentrated around its mean. However, for sets of much smaller size ($\mu(f) \simeq \exp(-k)$), one needs the additive error term $2^{-k^{3/4}}$. The proof follows from a case analysis, based on $\mu(f)$. We defer the full proof to the appendix (refer Section C). □

2.2 Reduction to Bipartite Densest Subgraph Problem

Bipartite DκS. The bipartite variant of DκS is where we are given a bipartite graph $G = (A \cup B, E)$ and two parameters κ_1 and κ_2. We need to find a subsets $S \subseteq A$, $T \subseteq B$ such that S contains κ_1 elements, T contains κ_2 elements and the number of edges contained in $S \cup T$ is the maximum possible among such subsets.

Feige's Reduction. Given a k-AND formula $\mathcal{F}$ on n variables consisting of $m = \Delta n$ clauses, construct a bipartite graph $G = (A \cup B, E)$ with $A = [m]$ identified by the clauses, $B = [2n]$ identified by the literals. An edge (a, b) is in G exactly when the clause a contains the literal b. Now, set $\kappa_1 = 2^{-c\sqrt{k}}m$ (where c is from Hypothesis 1.1) and $\kappa_2 = n$.

If the k-AND formula has an assignment satisfying κ_1 clauses, pick T to be the set of true literals (one per variable; sums to n vertices in B). Vertices corresponding to satisfied clauses have all their k neighbors in T. Thus, G contains subsets of the required sizes (the satisfiable clauses and the true literals) that contain $\kappa_1 k$ edges.

On the other hand, in a typical random k-AND formula, a fixed assignment is expected to set roughly half the literals in a clause to true. For large enough Δ, this intuition can be proven. Observe that a random k-AND formula maps to a random element of $\mathcal{G}_{m,n}^k$. Fix subsets S and T of the prescribed sizes; let $f :$ and g denote their indicators. Then, $\mu(f) = \kappa_1/m$ and $\mu(G) = 1/2$. Applying Theorem 2.1, we get that

$$\mathbf{E}[f(a)g(b)] \leq \left(2^{-c\sqrt{k}}(1/2) + 2^{-c\sqrt{k}}k^{-1/9} + 2^{-k^{-3/4}}\right) \leq 2^{-c\sqrt{k}} \cdot (1/2 + o(1))$$

for large enough k and Δ. Thus, the densest subgraph of the prescribed sizes contains at most $\kappa_1 k(1/2 + o(1))$ edges. Now, Hypothesis 1.1 already gives a factor 2-inapproximability for the bipartite version.

2.3 Gap Amplification using Products

Amplification via graph products is a standard technique for increasing the gap in the value of the objective. For example, the clique number of $G^{\otimes l}$ is the clique number of G raised to power l. The analogous statement for densest subgraph is false since dense subsets of the product graph may not correspond to a single dense subset of the original graph. Luckily we can argue such a thing in the case when the graph we started with was random.

Let $\mathcal{F}$ be a k-AND formula as in the above section and let G be the graph produced from the above reduction. For a parameter ℓ, let $G^{\otimes \ell} = (A^\ell \cup B^\ell, E_\ell)$ denote ℓ-th power of G. The following lemma extends Theorem 2.1 to graph powers. Qualitatively, we show that the number of edges in a set of size κ^ℓ in $G^{\otimes \ell}$ is roughly at most the number of edges in a set of size κ in G raised to the ℓ-th power; the slack in the bound decays with ℓ.

Lemma 2.2. *There exist constants k_1, c_1 such that for every $k \geq k_1$, there is a $\Delta_1 = \Delta_1(k)$ such that, for every $m \geq \Delta_1 n$, a $G = (A \cup B, E)$ randomly picked from $\mathcal{G}^k_{m,n}$ satisfies the following property with probability at least $1 - exp(-c_1 n)$:*

For every non-negative integer ℓ, and every pair of functions $f_\ell : A^\ell \to [0,1]$, $g_\ell : B^\ell \to [0,1]$; the graph $G^{\otimes \ell} = (A^\ell \cup B^\ell, E_\ell)$ satisfies

$$\mathbf{E}_{a_1,a_2,\ldots,a_\ell,b_1,b_2,\ldots,b_\ell \in E_\ell}[f_\ell(a_1,a_2,\ldots,a_\ell)g_\ell(b_1,b_2,\ldots,b_\ell)] \leq \mu(f_\ell)\mu(g_\ell) + \ell\delta\mu(f_\ell) + \ell\epsilon$$

where $\delta = 1/k^{1/9}$ and $\epsilon = 2^{-k^{3/4}}$.

Proof. Given f_ℓ and g_ℓ, define

$$f_i(a_1,a_2,\ldots a_i) \stackrel{\text{def}}{=} \frac{1}{|A|^{\ell-i}} \sum_{(a_{i+1},\ldots,a_\ell)} f(a_1,\ldots a_\ell)$$

and g_i similarly for $1 \leq i < \ell$. Now, for a fixed $a_1,a_2,\ldots,a_{\ell-1},b_1,b_2,\ldots,b_{\ell-1}$, Theorem 2.1 implies that with probability $1 - exp(-c_0 n)$,

$$\mathbf{E}_{a_\ell,b_\ell}\left[f_\ell(a_1,a_2,\ldots,a_\ell)g_\ell(b_1,b_2,\ldots,b_\ell)\right] \leq f_{\ell-1}(\cdot)g_{\ell-1}(\cdot) + \delta f_{\ell-1}(\cdot) + \epsilon$$

where "$(\cdot)$" represents the first $\ell-1$ coordinates. Now, $\mathbf{E}[f_{\ell-1}] = \mu(f_\ell)$. Leaving only the first term. Applying this iteratively $\ell - 1$ more times gives the result (with $c_1 = c_0$). □

Parameters. Let δ, ϵ be as above. Set ℓ such that $\ell\delta \leq 2^{-\ell}$ and $\ell\epsilon \leq 2^{-k^{5/8}}$ ($\ell = \theta(\log k)$ for large enough k). Set k large enough and $\Delta = \Delta_1(k)$ from the above theorem. Finally, set $\kappa_1 = 2^{-c\sqrt{k}\ell}m^\ell$ and $\kappa_2 = 2^{-\ell}n^\ell$.

As before, suppose $\mathcal{F}$ had a $2^{-c\sqrt{k}}$ fraction of the clauses that could be simultaneously satisfied by an assignment. Let S denote the satisfied clauses and T denote the set of true literals in the assignment. In $G^{\otimes\ell}$, $S^\ell \cup T^\ell$ contains at least $2^{-c\sqrt{k}\ell}m^\ell k^\ell$ edges. Further S and T have κ_1 and κ_2 edges respectively. On the other hand, typical k-AND formulas output G such that $G^{\otimes\ell}$ has no such S and T containing more than $3 \cdot 2^{-\ell}2^{-c\sqrt{k}\ell}m^\ell k^\ell$ edges (by direct application of the above theorem with the set parameters), improving the gap to $2^{-\ell}/3 = k^{\Omega(1)}$ for large enough k.

2.4 Reduction to DκS

Keeping the setup and parameters as above, we are now in a shape to describe the complete reduction. Given a k-AND formula $\mathcal{F}$, construct G as in 2.2 and its power, $G^{\otimes \ell} = (A^\ell \cup B^\ell, E_\ell)$. Let κ_1, κ_2 be as immediately above.

To reduce to DκS, we use a Lagrangian relaxation style trick. Set $\lambda = \Delta^{2\ell}(\kappa_2 + \ell\delta n^\ell)/\kappa_1$ and $\kappa = \kappa_1\lambda + \kappa_2$. Construct graph H by taking λ disjoint copies of A^ℓ and connecting them in the same way to B^ℓ as in $G^{\otimes \ell}$. Output H and κ as the final DκS instance.

Running Time. For a fixed large enough k and Δ, the total reduction runs in $\Delta^{O(\log k)} m^{O(\log k)}$ steps which is polynomial in m for any fixed k. A closer look into Theorem 2.1 says that $\Delta = exp(k)$ suffices. Thus, for large enough k, the reduction runs in type $\exp(k) m^{O(\log k)}$.

Density of κ-subgraphs in H As before, if $\mathcal{F}$ is well satisfiable, we can find a subgraph of size κ containing $\lambda\kappa_1 k^\ell$ edges (since H is just a disjoint union of many copies of $G^{\otimes \ell}$).

Fix a subset S containing κ vertices in H. Let $f_1, f_2, \ldots f_\lambda$ be the indicator of S within each of the disjoint copies of A^ℓ. Let g be the indicator of $S \cap B^\ell$. Now $\kappa = \sum_i \mu(f_i) m^\ell + \mu(g) n^\ell$. The number of edges in S is, by Lemma 2.2, at most $\sum_i m^\ell \left(\mu(f_i)\mu(g) + \ell\delta\mu(f_i) + \ell\epsilon\right)$. For the purpose of bounding this expression, we can assume that $\mu(f_i) = \mu(f)$ (independent of i).

Our choice of λ was such that the function $f(x,y) = xy + \ell\delta x + \ell\epsilon$ under the constraint that $\lambda x + y = \lambda\kappa_1/m^\ell + \kappa_2/n^\ell$ is maximum when $x = \kappa_1/m^\ell$ and $y = \kappa_2/n^\ell$. In the above setting, this means that typical k-AND formulas produce H with at most $3\lambda\kappa_1\kappa_2 k^\ell/n^\ell$ edges in any subset induced on κ vertices (a $3 \cdot 2^{-\ell}$ factor lesser than when $\mathcal{F}$ is well satisfiable). We finish by stating this as a theorem; Theorem 1.2 is a corollary of the following.

Theorem 2.3. *There are constants k_2, c_2 and c_3 and an algorithm $\mathcal{A}$ that takes an AND formula and outputs a DκS instance (H, κ) such that for all $k \geq k_2$:*

- *Given a k-AND formula on n variables consisting of m clauses, $\mathcal{A}$ runs in $2^{O(k)} m^{O(\log k)}$ steps.*
- *For every k, there is a $\nu = \nu(k,m)$ such that on typical random k-AND formulas, $\mathcal{A}$ produces graphs with at most ν in any subgraph of size κ whilst always outputting graphs that contain a $k^{c_2}\nu$ subgraph of the same size when fed a formula that is $m/2^{c\sqrt{k}}$ satisfiable.*

3 Reduction from Hidden Clique

3.1 Notations and Setup

For notational ease, in this section, we measure the density of subsets of vertices in graphs as follows.

Graph density For any (undirected) graph $G = (V, E)$, and for any subset $X \subseteq V$, the *density* of X, denoted $d(X)$, is defined as $d(X) = \frac{|E(X)|}{\binom{|X|}{2}}$ where $E(X)$ is the number of edges in the induced subgraph $G|_X$. Similarly, for any two disjoint subsets $X, Y \subseteq V$, the density between X and Y is $d(X,Y) = \frac{|E(X,Y)|}{|X||Y|}$, where $E(X,Y)$ is the number of edges that have one endpoint in X and the other in Y in the induced subgraph $G|_{X \cup Y}$.

We denote by $\mathcal{G}(n, 1/2)$ the random graph on n vertices, where each edge is drawn independently at random with probability $1/2$.

The following definition, which is a slight variant of the power graph definition used in section 2 will be central to our analysis:

Definition 3.1 (Subset Power Graphs). *Let $G = (V, E)$ be an (undirected) graph on $|V| = n$ vertices. For any integer ℓ, the ℓ-th subset power graph* of G, denoted $G^\ell = (V^\ell, E^\ell)$, is the graph whose set of vertices V^ℓ consists of all $\binom{n}{\ell}$ subsets of cardinality ℓ of V, and for any pair $(A = (u_1, u_2, ..., u_\ell), B = (v_1, v_2, ..., v_\ell))$, $(A, B) \in E^\ell \iff A \cup B$ forms a clique in G (that is, for every two distinct $a, b \in A \cup B$, $ab \in E$).

3.2 The proof of Theorem 1.3

The reduction we present proceeds as follows. First, we take the subset-power graph of the given hidden clique instance, so that in the "Yes" case the clique is preserved whereas in the "No" case we show that the density of any subgraph (of the right size) decreases exponentially with the square of ℓ (the powering parameter). Theorem 3.3 together with Corollary 3.4 assert this fact. Then, we "blow up" the graph, replacing each vertex of the power graph with a clique, so that the relative size of the largest clique grows from $n^{1/3}$ to $N^{1-\epsilon}$, while this operation is shown to have negligible contribution to the density of the sparse graphs of the "No" instances. We wrap up the proof with Theorem 3.7.

We begin with the following lemma which will come handy in the proof of Theorem 3.3:

Lemma 3.2. *Let $G = \mathcal{G}(n, 1/2)$ and $\mathcal{G}^\ell = (V^\ell, E^\ell)$ be the ℓ-th subset-product graph of G for some constant $\ell \geq 1$. Then with probability $1 - o(1)$ the following condition holds:*

Condition (1): For every two disjoint subsets $\mathcal{A}_\ell, \mathcal{A}_{\ell'} \subseteq V^\ell$, $|\mathcal{A}_\ell| = |\mathcal{A}_{\ell'}| = c \log n$ where $c = c(\ell) = 2^{\ell^2}$ and $\forall\ A \in \mathcal{A}_\ell \cup \mathcal{A}_{\ell'} \quad |S_{\mathcal{A}_\ell, \mathcal{A}_{\ell'}}(A)| \geq 0.1\ell$ where $S_{\mathcal{A}_\ell, \mathcal{A}_{\ell'}}(A) \overset{\text{def}}{=} \{v \in A \ : \ \forall\ B \in \mathcal{A}_\ell \cup \mathcal{A}_{\ell'}, B \neq A \ \ v \notin B\}$ (that is, every set in the collection $\mathcal{A}_\ell \cup \mathcal{A}_{\ell'}$ has at least a constant fraction of elements which do not appear in any other set of the collection), the density $d(\mathcal{A}_\ell, \mathcal{A}_{\ell'})$ satisfies:

$$d(\mathcal{A}_\ell, \mathcal{A}_{\ell'}) \leq 2 \cdot 2^{-0.01\ell^2}$$

Proof sketch. The proof uses a standard Chernoff bound to upper bound the number of edges between two given sets $\mathcal{A}_\ell$ and $\mathcal{A}_{\ell'}$ which satisfy the premises of the condition. We defer the proof to the appendix.

Theorem 3.3. *Let $G = \mathcal{G}(n, 1/2)$ and $\mathcal{G}^\ell = (V^\ell, E^\ell)$ be the ℓ-th subset-product graph of G for some constant $\ell \geq 1$, and suppose that G satisfies condition* 1 *in Lemma 3.2. Then*

$$\forall\ \mathcal{B}_\ell \subset V^\ell, |\mathcal{B}_\ell| = \binom{n^{1/3}}{\ell} \qquad d(\mathcal{B}_\ell) \leq 3 \cdot 2^{-0.01\ell^2}$$

Notation. For convenience, we denote $\Delta = \binom{n^{1/3}}{\ell}$.

Proof. Given $\mathcal{B}_\ell$, define the set $Y = \{v \in V \mid v \text{ belongs to more than } \Delta \cdot n^{-1/20} \text{ members of } \mathcal{B}_\ell\}$. Note that the total number of vertices (from the original graph G) in all members of $\mathcal{B}_\ell$ (double counting appearances) is $\ell|\mathcal{B}_\ell| = \ell \cdot \Delta$, and so the average number of members in which $v \in V$ appears is $\frac{\ell \cdot \Delta}{n}$. Thus, the number of vertices $v \in V$ which appear in more than $\Delta \cdot n^{-1/20}$ members is at most $\frac{\ell\Delta}{\Delta \cdot n^{-1/20}} = \ell \cdot n^{1/20}$. Therefore, $|Y| \leq \ell \cdot n^{1/20}$.

Intuitively, we say a member $A \in \mathcal{B}_\ell$ is *bad* if it has "too many" vertices in Y. We claim that $\mathcal{B}_\ell$ has a small number of bad members. Indeed, let $\mathcal{B}_\ell^{bad} \overset{\text{def}}{=} \{B \in \mathcal{B}_\ell \mid |B \cap Y| \geq 0.9\ell\}$. Then $|\mathcal{B}_\ell^{bad}| \leq \binom{|Y|}{0.9\ell} n^{0.1\ell} < \Theta(n)^{\frac{0.9\ell}{20}} \cdot n^{0.1\ell} < n^{\frac{5\ell}{33}}$. Let $\mathcal{B}_\ell^{good} \overset{\text{def}}{=} \mathcal{B}_\ell - \mathcal{B}_\ell^{bad}$. Then for large enough n,

$$|\mathcal{B}_\ell^{good}| \geq \Delta - n^{\frac{5\ell}{33}} \tag{1}$$

For $A \in V^\ell$, let $T(A) \overset{\text{def}}{=} \{v \in A \mid v \notin Y\}$ (note that every $A \in \mathcal{B}_\ell^{good}$ has $T(A) \geq 0.1\ell$. W.l.o.g we can assume equality, otherwise take an arbitrary subset of exact size 0.1ℓ). Partition $\mathcal{B}_\ell^{good}$ arbitrarily into $q = \frac{|\mathcal{B}_\ell^{good}|}{c\log n} \leq \frac{\Delta}{c\log n}$ sets S_i of size $c\log n$ (we assume w.l.o.g that q is an integer. otherwise, throw the remainder of the vertices into $\mathcal{B}_\ell^{bad}$). Define an auxiliary graph on the sets S_i, where S_i is connected to S_j iff $i \neq j$ and $\exists A \in S_i, B \in S_j$ s.t $T(A) \cap T(B) \neq \emptyset$. Then each S_i ($1 \leq i \leq q$) is connected to at most $|S_i| \cdot 0.1\ell\Delta n^{-1/20}$ distinct S_j's. Thus, there are at most $\frac{q \cdot c\log(n) 0.1\ell\Delta n^{-1/20}}{2} \leq \Delta^2 n^{-1/30}$ pairs S_i, S_j connected. By Lemma 3.2, the number of edges of G^ℓ between disconnected S_i, S_j is at most $|S_i||S_j| 2 \cdot 2^{-0.01\ell^2}$ (since the pair S_i, S_j satisfies $(\star)$). Thus, the total number of edges in B_ℓ is at most

$$|B_\ell^{bad}||B_\ell| + q\binom{c\log n}{2} + \Delta^2 n^{-1/30}(c\log n)^2 + \binom{q}{2}(c\log n)^2 \cdot 2 \cdot 2^{-0.01\ell^2} \tag{2}$$

where the first term upper bounds the number of edges that have at least one endpoint in B_ℓ^{bad}, the second accounts for the edges inside each set S_i for all $1 \leq i \leq q$, the third term accounts for the number of edges between all connected pairs S_i, S_j (for which we assume a complete bipartite graph) and the fourth accounts for the edges between all of at most $\binom{q}{2}$ disconnected pairs. Plugging in the numbers in (2), we get

$$|E(B_\ell)| \leq \Delta n^{5\ell/33} + \frac{\Delta}{c\log n}(c\log n)^2 + \Delta^2 n^{-1/30}(c\log n)^2 + \binom{\Delta}{2} 2 \cdot 2^{-0.01\ell^2} \tag{3}$$

and dividing by $\binom{|B_\ell|}{2} = \binom{\Delta}{2}$ yields

$$d(B_\ell) = \frac{|E(B_\ell|}{\binom{|B_\ell|}{2}} \leq 4\frac{n^{5\ell/33}}{\Delta} + \frac{4c\log(n)}{\Delta} + 4n^{-1/30}(c\log n)^2 + 2 \cdot 2^{-0.01\ell^2} \tag{4}$$

where for the three first terms we used the fact that $\binom{\Delta}{2} \geq \frac{\Delta^2}{4}$. Note that for any fixed ℓ these three terms in (4) tend to 0 when $n \longrightarrow \infty$. In particular, there exists an $m = m(\ell)$ such that for all $n \geq m$ all three terms are at most $\frac{1}{3} \cdot 2^{-0.01\ell^2}$. Thus, if $n \geq m$, we have

$$d(B_\ell) \leq 2^{-0.01\ell^2} + 2 \cdot 2^{-0.01\ell^2} = 3 \cdot 2^{-0.01\ell^2} \tag{5}$$

as desired. □

Corollary 3.4. *Let $G = \mathcal{G}(n, 1/2)$ and $\mathcal{G}^\ell = (V^\ell, E^\ell)$ be the ℓ-th subset-product graph of G for some constant $\ell \geq 1$. Then with probability $1 - o(1)$*

$$\forall\, \mathcal{B}_\ell \subset V^\ell, |\mathcal{B}_\ell| = \Delta \qquad d(\mathcal{B}_\ell) \leq 3 \cdot 2^{-0.01\ell^2}$$

Proof. By Lemma 3.2, condition 1 holds w.p $1 - o(1)$, and so the claim follows from Theorem 3.3.□

Remark: Note that Theorem 3.3 is completely deterministic. The randomness of the underlying graph $\mathcal{G}(n, 1/2)$ affects only the probability of satisfying condition 1.

We proceed to the second step of our reduction.

Definition 3.5. *Let $G = (V, E)$ be an undirected graph. The k-clique blowup* of G is the graph $G_k = (V_k, E_k)$ whose vertices are $V \times [k]$ and (v, i) is connected to (u, j) whenever $(u, v) \in E$ or $u = v$. That is, each vertex $v \in G$ is mapped to a clique of size k, which we refer to as the *block* B_v of v, and two blocks are connected (form a complete bipartite graph) whenever their corresponding vertices are connected in G.

Lemma 3.6. *Let $G_k = (V_k, E_k)$ be a k-clique blowup of $G = (V, E)$, and let d_s denote the maximal density of a set of size s in G. Then the maximum density of a set of size $k \cdot s$ in G_k is at most $d_s + \frac{1}{s}$.*

Proof. We defer the proof to the appendix. □

We are now ready to prove Theorem 1.3 for which we consider the following counter-positive version:

Theorem 3.7. *If $\exists\, \epsilon > 0, \delta > 0$ for which the Densest κ-Subgraph$(1, \delta)$ with $\kappa = N^{1-\epsilon}$ can be solved in time $N^{O(1)}$, then the Hidden Clique for clique of size $n^{1/3}$ can be solved in time $n^{O(\frac{\sqrt{\log(\frac{1}{\delta})}}{\epsilon})}$.*

Proof. Let $G = (V, E)$, $|V| = n$ be the given instance of the Hidden Clique problem. Let G^ℓ be the ℓ-subset product graph of G, and G^ℓ_q be the q-clique blowup of G^ℓ, for $\ell = \sqrt{c \log(\frac{1}{\delta})}$ (we determine c shortly), and q such that $\binom{n^{1/3}}{\ell} q = (\binom{n}{\ell} q)^{1-\epsilon}$. It is easy to check that $q = n^{\Theta(\ell/\epsilon)}$. Let N denote the number of vertices of G^ℓ_q, so $N = \binom{n}{\ell} q$. Note that G^ℓ_q can be constructed in polynomial time. If G has a planted clique H of size $n^{1/3}$, then G^ℓ has a clique of size Δ (all $\binom{n^{1/3}}{\ell}$ subsets of H), and G^ℓ_q "blows" up this clique by a factor of q, so that G^ℓ_q has a clique of size $\kappa = \Delta \cdot q = N^{1-\epsilon}$ (by the choice of q). Thus in the "Yes" instance G^ℓ_q has a κ-subgraph of density 1. On the other hand, if G is a "No" instance, i.e $G = \mathcal{G}(n, 1/2)$, then by Corollary 3.4, with probability $1 - o(1)$ any induced subgraph H_ℓ of size Δ in G^ℓ satisfies $d(H_\ell) \leq 3 \cdot 2^{-0.01\ell^2}$. But G^ℓ_q is a q-clique blowup of G^ℓ, and thus by Lemma 3.6, the density of any induced subgraph of size $\Delta q = N^{1-\epsilon}$ is at most

$$\left(\max_{B_\ell \subseteq V^\ell, |B_\ell| = \Delta} d(B_\ell)\right) + \frac{1}{\Delta} \leq 3 \cdot 2^{-0.01\ell^2} + \frac{1}{\Delta} = 3 \cdot 2^{-0.01c\log(1/\delta)} + \frac{1}{\Delta} = 3 \cdot \delta^{0.01c} + \frac{1}{\binom{n^{1/3}}{\ell}} \tag{6}$$

and for an appropriate choice of c and large enough n, each of the terms in (6) is no larger than $\frac{\delta}{2}$, which yields

$$\left(\max_{H \subseteq V_q^\ell, |H| = N^{1-\epsilon}} d(H)\right) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta \tag{7}$$

completing the soundness side of the proof. The resulting graph is therefore an instance of DκS $(1, \delta)$, $\kappa = N^{1-\epsilon}$, which by the premises of the theorem is solvable in time $N^{O(1)} = n^{O(\frac{\sqrt{\log(\frac{1}{\delta})}}{\epsilon})}$ □

It is not hard to see that the proof of Theorem 1.3 carries over to sub-constant values of δ as well, depending on the hardness of the Hidden Clique problem. In particular,

Corollary 3.8. *If the Hidden Clique problem for a planted clique of size $t = n^{1/3}$ in $\mathcal{G}(n, 1/2)$ cannot be solved in time $n^{o(\log n)}$, then there is no algorithm that distinguishes between a graph on N vertices containing a clique of size $\kappa = N^{1/3}$ and one in which the densest subgraph on κ vertices has density at most $2^{-\Omega((\log N)^{2/3}))}$ in time $N^{o((\log N)^{1/3})}$.*

Proof. Set $\ell = \Theta(\sqrt{\log n})$ in the proof of Theorem 3.7 (Note that one cannot take a bigger ℓ, as 2^{ℓ^2} cannot exceed some fixed power of n for the proof to work). With this choice of ℓ, $N = n^{\Theta(\sqrt{\log n})}$ and $\delta = 2^{-\Theta(\ell^2)} = 2^{-\Theta(\log n)} = 2^{-\Theta((\log N)^{2/3})}$. The existence of an algorithm with running time $N^{o((\log N)^{1/3})} = n^{o(\log n)}$ for identifying a clique of size $N^{1/3}$ in the subset power graph G^ℓ would show that we can identify the hidden clique in the original graph $\mathcal{G}(n, 1/2)$ in time $n^{o(\log n)}$, contradicting the hardness assumption. □

Remark 3.9. *It is worth to note that an analogous extension of theorem 1.2 to super-constant density ratios also requires a stronger version of hypothesis 1.1 (namely, that random k-SAT formulas are hard to distinguish for super-constant values of k). For super-constant density ratios, the above reduction has the advantage that the density of the "No" instance produced by it decreases in a rate proportional to $2^{-\ell^2}$ (ℓ being the graph-powering parameter), whereas in the reduction described in the proof of theorem 1.2 it only deceases as fast as $2^{-\ell}$. Thus, for a super-constant value of ℓ, the above reduction produces hardness for a significantly higher approximation ratio.*

4 Open Problems

It seems plausible that the DκS problem is hard to approximate even up to an n^c factor in polynomial time for some fixed $c > 0$. It will be interesting to decide whether or not some version of our gap amplification technique can yield such an $n^{\Omega(1)}$ inapproximability result.

As mentioned in the introduction, the complexity of finding a hidden dense subgraph in a random graph has interesting applications, and it will be nice to establish hardness for this problem using some hardness assumptions like the ones considered here. Note that our reduction in Section 3 provides hardness of finding dense subgraphs in certain semi-random graphs (obtained by taking powers of a random graph), but not in usual random graphs.

Finally, proving a $1 + \epsilon$ NP-hardness of approximation for DκS for any fixed $\epsilon > 0$ is a long standing open problem.

References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *STOC*, pages 496–505. ACM, 2007.

[ABBG10] Sanjeev Arora, Boaz Barak, Markus Brunnermeier, and Rong Ge. Computational complexity and information asymmetry in financial products (extended abstract). In *ICS*, pages 49–65. Tsinghua University Press, 2010.

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *STOC*, pages 171–180. ACM, 2010.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *SODA*, pages 594–598, 1998.

[BCC+10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $n^{1/4}$ approximation for densest -subgraph. In *STOC*, pages 201–210. ACM, 2010.

[DGGP10] Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. *CoRR*, abs/1010.2997, 2010.

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *STOC*, pages 534–543. ACM Press, 2002.

[FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Struct. Algorithms*, 16(2):195–208, 2000.

[FPK01] Uriel Feige, David Peleg, and Guy Kortsarz. The dense -subgraph problem. *Algorithmica*, 29(3):410–421, 2001.

[FR10] Uriel Feige and Dorit Ron. Finding hidden cliques in linear time. In *AOFA*, 2010.

[FS97] Uriel Feige and Michael Seltser. Todo. *Algorithmica*, 29:2001, 1997.

[HK11] Elad Hazan and Robert Krauthgamer. How hard is it to approximate the best nash equilibrium? *SIAM J. Comput.*, 40(1):79–91, 2011.

[Jer92] Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3(4):347–360, 1992.

[Kuc95] Ludek Kucera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2-3):193–212, 1995.

[RST10] Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:172, 2010.

[ST00] Alex Samorodnitsky and Luca Trevisan. A pcp characterization of np with optimal amortized query complexity. In *STOC*, pages 191–199, 2000.

[Tul09] Madhur Tulsiani. Csp gaps and reductions in the lasserre hierarchy. In *STOC*, pages 303–312. ACM, 2009.

A Proof of Lemma 3.2

Proof. We say that two sets $\mathcal{A}_\ell, \mathcal{A}_{\ell'}$ satisfy $(\star)$ if they have the properties specified in condition (1). Let $\mathcal{A}_\ell, \mathcal{A}_{\ell'} \subset V^\ell$ be two subsets that satisfy $(\star)$, and let $|E(\mathcal{A}_\ell, \mathcal{A}_{\ell'})|$ denote the random variable whose value is the number of edges between the two sets in the induced subgraph $G^l|_{\mathcal{A}_\ell \cup \mathcal{A}_{\ell'}}$. Let $S \stackrel{\text{def}}{=} \sum_{A \in \mathcal{A}_\ell, A' \in \mathcal{A}_{\ell'}} I_{A,A'}$ where $I_{A,A'}$ is the indicator random variable whose value is 1 iff all the vertices in $S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A)$ are connected to all the vertices in $S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A')$ in G (i.e, the edges between $S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A)$ and $S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A')$ form a complete bipartite graph on these two sets). Note that this event contains the event that $(A, A') \in E^\ell$, and so $|E(\mathcal{A}_\ell, \mathcal{A}_{\ell'})| \leq S$. W.l.o.g, we can assume that $|S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A)| = 0.1\ell\ \forall A \in \mathcal{A}_\ell \cup \mathcal{A}_{\ell'}$ (otherwise, take an arbitrary subset $T \subset S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A)$ of size 0.1ℓ and use it in our following analysis).

Clearly, $\mathbf{E}[\mathrm{I}_{\mathrm{A,A'}}] = 2^{-(0.1\ell)^2}$ for any pair (A, A') (as the number of edges (from G) in the complete bipartite graph $K_{|S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A)|,|S_{\mathcal{A}_\ell,\mathcal{A}_{\ell'}}(A')|}$ is $(0.1\ell)^2$). By $(\star)$ and linearity of expectation, we therefore have

$$\mathbf{E}[|\mathrm{E}(\mathcal{A}_\ell, \mathcal{A}_{\ell'})|] \leq \mathbf{E}[\mathrm{S}] = \sum_{\mathrm{A} \in \mathcal{A}_\ell, \mathrm{A'} \in \mathcal{A}_{\ell'}} \mathbf{E}[\mathrm{I}_{\mathrm{A,A'}}] = |\mathcal{A}_\ell||\mathcal{A}_{\ell'}|2^{-(0.1\ell)^2} \tag{8}$$

where the expectation is over the edges in $\mathcal{G}(n, 1/2)$. By $(\star)$, the events $I_{A,A'}$ are independent (we never considered a single edge from G more than once), and so S is the sum of $|\mathcal{A}_\ell||\mathcal{A}_{\ell'}|$ i.i.d 0/1-random variables and has expectation $\mu = |\mathcal{A}_\ell||\mathcal{A}_{\ell'}| \cdot 2^{-0.01\ell^2}$. We can therefore apply Chernoff bound[1] to obtain

$$\Pr[S > 2 \cdot |\mathcal{A}_\ell||\mathcal{A}_{\ell'}| \cdot 2^{-(0.1\ell)^2}] = \Pr[S > 2\mu] \leq \Pr[|S - \mu| > \mu] < 2e^{-\mu/3} < 2^{1 - \frac{1}{3}(c^2 \log^2(n) 2^{-0.01\ell^2})} \tag{9}$$

where the first transition is by (8). Since $|E(\mathcal{A}_\ell, \mathcal{A}_{\ell'})| \leq S$, we get

$$Pr[d(\mathcal{A}_\ell, \mathcal{A}_{\ell'}) > 2 \cdot 2^{-(0.1\ell)^2}] \leq \Pr[S > |\mathcal{A}_\ell||\mathcal{A}_{\ell'}|2 \cdot 2^{-(0.1\ell)^2}] \leq 2^{1 - \frac{1}{3}(c^2 \log^2(n) 2^{-0.01\ell^2})} \tag{10}$$

Since the total number of pairs of disjoint subsets of size $c \log n$ in V^ℓ is at most $\binom{|V^\ell|}{c \log n}^2 < n^{2\ell c \log n} = 2^{2\ell c \log^2 n}$, a union bound on all pairs yields that the probability that there exist a pair

[1] See N.Alon and J. Spencer, "The Probabilistic Method" (Third Edition) ,Corollary A.1.14, p.312. Note that since $\epsilon = 1$ in our case, $c_\epsilon = \min(\ln 4 - 1, 1/2) > \frac{1}{3}$.

of subsets $\mathcal{A}_\ell, \mathcal{A}_{\ell'}$ with $d(\mathcal{A}_\ell, \mathcal{A}_{\ell'}) > 2 \cdot 2^{-(0.1\ell)^2}$ is at most

$$2^{1+c\log^2 n(2\ell-\frac{1}{3}c2^{-0.01\ell^2})}$$

which is $o(1)$ (with respect to n) for $c = 2^{\ell^2}$.

B Proof of Lemma 3.6

Given a set of vertices $T \subseteq V_k$ in G_k, let $g(T) = \sum_{v\in T} B(v)$, where $B(v)$ is the number from $[n]$ associated with u s.t $v \in B_u$. (I.e, g is charging each vertex its block number).

Let $T \subseteq V_k, |T| = ks$ be a set of maximal density in G_k. We first show that among all such sets W of maximal density, if T also minimizes $g(W)$, then T must be a union of exactly s *full* blocks (I.e, blocks B s.t $T \cap B = B$). Indeed, if this is not the case, then let $B_{v_1}, B_{v_2}, ..., B_{v_m}$ be the set of all non-full blocks which have nonempty intersection with T (i.e, $0 < |T \cap B_{v_i}| < k$) and assume w.l.o.g that B_{v_1} and B_{v_m} have the maximal and minimal degrees in T, d_{max} and d_{min} respectively (where a degree of a block is simply the degree of any of its vertices. This is well defined since by construction, all vertices in a block have the same degree). If $d_{max} > d_{min}$, then we claim that replacing a vertex from $T \cap B_{v_m}$ with a vertex from $B_{v_1} - (T \cap B_{v_1})$ increases the density of T. Indeed, removing a vertex from $T \cap B_{v_m}$ incurs a loss of d_{min} edges, while inserting a vertex from $B_{v_1} - (T \cap B_{v_1})$ contributes d_{max} edges in case B_{v_1} was connected to B_{v_m} [2], and $d_{max} + 1$ if the blocks were not connected. Since $d_{max} > d_{min}$, in both cases this operation increases the number of edges, resulting in a set T' of size ks with $d(T') > d(T)$, which contradicts the maximality of $d(T)$.

If $d_{max} = d_{min}$, let $B_1, B_2, ..., B_n$ be an ordering of all blocks according to $B(v)$ (i.e, B_j is the block of vertex j in G. Thus, if $v \in B_j, B(v) = j$), and let B_{i_o} be the first non-full block according to that order. Since the total number of vertices in T is a multiple of k, there must exist another non-full block B_j, with $i_0 < j$. If B_{i_0} and B_j are not connected, then by the above argument, replacing a vertex $w \in (B_j \cap T)$ with a new vertex $w' \in (B_{i_0} - (T \cap B_{i_0}))$ yields a set T' with higher density, in contradiction to the maximality of $d(T)$. Finally, if the blocks are connected, since $i_0 < j$, $g(T') - g(T) = B(w') - B(w) = i_0 - j < 0$, which contradicts the minimality of T with respect to g (note that the new set maintains the maximal density).

By the above, it is enough to analyze the maximal density of T in the case that it is a union of s full blocks (as we can always find such T with equivalent density). Let $T = B_{v_1} \cup B_{v_2} \cup .... \cup B_{v_s}$, and denote $S = \{v_1, v_2, ..., v_s\}$. Note that since $S \subseteq V$ is a set of size s in the original graph G, it satisfies $d(S) \leq d_s$. The density of T can be therefore written as

$$\begin{aligned} d(T) &= \frac{|\text{edges between blocks}| + |\text{edges inside blocks}|}{\binom{ks}{2}} = \frac{k^2|E(S)| + \binom{k}{2}s}{\binom{ks}{2}} \\ &\leq \frac{k^2 s(s-1)d_s + k(k-1)s}{ks(ks-1)} \leq d_s + \frac{1}{s} \end{aligned}$$

□

[2] since the degrees of vertices inside the same block are equal, but may have decreased by one if the blocks of the removed and the added vertices were connected. The degree of the added vertex inside $B_{v_1} \cap T$ is higher by 1 than it was before the replacement.

C Proof of Theorem 2.1

Theorem C.1 (Theorem 2.1, restated). *There exists (non-negative) constants k_0 and c_0 such that for every $k \geq k_0$ there exists $\Delta_0 = \Delta_0(k)$ such that for every $m \geq \Delta_0 n$, a $G = (A \cup B, E)$ randomly picked from $\mathcal{G}^k_{m,n}$ satisfies the following property with probability at least $1 - exp(-c_0 n)$:*

For every $f : A \to [0,1]$ and $g : B \to [0,1]$,

$$\mathbf{E}[f(a)g(b)] \leq \mu(f)\mu(g) + \frac{1}{k^{1/8}}\mu(f) + 2^{-k^{3/4}}$$

Proof. Note that the inequality is multilinear in f and g; hence, f and g can be assumed to map onto $\{0,1\}$. Fix one such f and g. Let $S = \{a \in A | f(a) = 1\}$ and let $\{X^i_a\}$ be a collection of independent random-variables taking 1 with probability $\mu(g)$ and 0 otherwise. From the definition of $\mathcal{G}^k_{m,n}$,

$$\mathbf{E}[f(a)g(b)] = \frac{1}{mk}\sum_{i,a} X^i_a.$$

Note that the expected value of the above quantity is exactly $\mu(f)\mu(g)$. The rest of the argument follows in three cases, based on $\mu(f)$.

Case 1: $\mu(f) \leq \theta = 2^{-k^{3/4}}/k^5$: Using Talagrand's inequality,

$$\begin{aligned}\Pr\left[\mathbf{E}[fg] \geq \mu(f)\mu(g) + \epsilon\right] &= \Pr[\sum_{a,i} X^i_a \geq \mu(f)\mu(g)mk + \left(\epsilon\sqrt{mk/\mu(f)}\right)\sqrt{\mu(f)mk}] \\ &\leq \exp(-c\epsilon^2 mk/\mu(f)) \leq \exp(-c\epsilon^2 mk/\theta)\end{aligned}$$

There are at most $\exp(H(\theta)m)$ different choices for f and 2^n different choices for g. Setting $\epsilon = exp(-\Omega(k^{3/4}))$, choosing large enough Δ and k, and a simple union bound argument gives that with probability at least $1 - \exp(-\Omega(n))$, $\mathbf{E}[fg] \leq \mu(f)\mu(g) + \epsilon$ for all f, g such that $\mu(f) < \theta$.

Case 2: $\theta < \mu(f) < 1/\sqrt{k}$: Again, using Talagrand's inequality,

$$\begin{aligned}\Pr\left[\mathbf{E}[fg] \geq \mu(f)\mu(g) + \delta\mu(f)\right] &= \Pr[\sum_{a,i} X^i_a \geq \mu(f)\mu(g)mk + \delta\mu(f)mk] \\ &= \Pr[\sum_{a,i} X^i_a \geq \mu(f)\mu(g)mk + \delta\sqrt{\mu(f)mk}\sqrt{\mu(f)mk}] \\ &\leq \exp(-c\delta^2\mu(f)mk)\end{aligned}$$

Again, there are at most $\exp(H(\mu(f))m)$ different choices for f, 2^n different choices for g and m different choices for $\mu(f)$. $H(\mu(f)) \leq \mu(f)\log(1/\mu(f)) \leq \mu(f)k^{3/4}$. Thus, setting $\delta = \Omega\left(k^{-1/8}\right)$ gives that $\mathbf{E}[fg] \leq \mu(f)\mu(g) + \Omega\left(k^{-1/8}\right)\mu(f)$ with probability at least $1 - \exp(-\Omega(n))$.

Case 3: $\mu(f) \geq 1/\sqrt{k}$: As in the previous case, set $\delta = k^{-1/8}$ and using $H(\mu(f)) < 1$ gives the required bound, with probability exponentially close to 1.

Putting together all the cases, we have that:

$$\mathbf{E}[f(a)g(b)] \leq \mu(f)\mu(g) + \frac{1}{k^{1/8}}\mu(f) + 2^{-k^{3/4}}$$

with probability at least $1 - \exp(-cn)$. □