

Technical Report

TR-2003-008

Embedding graphs with bounded degree in sparse pseudorandom graphs

by

Y. Kohayakawa, V. Rodl, P. Sissokho

MATHEMATICS AND COMPUTER SCIENCE

EMORY UNIVERSITY

Embedding graphs with bounded degree in sparse pseudorandom graphs*

Y. Kohayakawa[†] V. Rödl[‡] P. Sissokho[§]

May 31, 2002

Abstract

In this paper, we show the equivalence of some *quasi-random* properties for sparse graphs, that is, graphs G with edge density $p = |E(G)|/\binom{n}{2} = o(1)$, where $o(1) \rightarrow 0$ as $n = |V(G)| \rightarrow \infty$. In particular, we prove the following embedding result. For a graph J , write $N_J(x)$ for the neighborhood of the vertex x in J , and let $\delta(J)$ and $\Delta(J)$ be the minimum and the maximum degree in J . Let H be a *triangle-free* graph and set $d_H = \max\{\delta(J) : J \subset H\}$. Moreover, put $D_H = \min\{2d_H, \Delta(H)\}$. Let $C > 1$ be a fixed constant and suppose $p = p(n) \gg n^{-1/D_H}$. We show that if G is such that

- (i) $\deg_G(x) \leq Cpn$ for all $x \in V(G)$,
- (ii) for all $2 \leq r \leq D_H$ and for all distinct vertices $x_1, \dots, x_r \in V(G)$,

$$|N_G(x_1) \cap \dots \cap N_G(x_r)| \leq Cnp^r,$$

- (iii) for all but at most $o(n^2)$ pairs $\{x_1, x_2\} \subset V(G)$,

$$||N_G(x_1) \cap N_G(x_2)| - np^2| = o(np^2),$$

then G contains H as a subgraph. We discuss a setting under which an arbitrary graph H (not necessarily triangle-free) can be embedded in G . We also present an embedding result for directed graphs.

Contents

1	Introduction	2
1.1	The constant density case	3
1.2	The vanishing density case	6

*Research supported by a CNPq/NSF cooperative grant.

[†]Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508-090 São Paulo, SP, Brazil. *E-mail*: yoshi@ime.usp.br. Partially supported by MCT/CNPq through ProNEx Programme (Proc. CNPq 664107/1997-4) and by CNPq (Proc. 300334/93-1 and 468516/2000-0)

[‡]Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA. *E-mail*: rodl@mathcs.emory.edu. Partially supported by NSF Grant 0071261.

[§]Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA. *E-mail*: psissok@emory.edu. Supported by NSF grant CCR-9820931

2	Statements of the main results	8
2.1	The undirected case	8
2.2	The directed case	11
3	Proofs of the main results	12
3.1	Preliminaries	12
3.2	Proof of Theorems 16 and 18	13
3.2.1	Degenerate orderings	16
3.2.2	The extension lemma and a corollary	17
3.2.3	Proof of Lemma 30	20
3.2.4	Proof of Theorem 16	24
3.3	Proof of Theorem 23	24
3.4	Remarks about Theorem 25 (the directed case)	29
4	Auxiliary facts and related work	32
4.1	General facts	32
4.2	Proof of Proposition 6	33
4.2.1	Construction of Alon's graph	33
4.3	Connection to random graphs	37
4.3.1	Theorem 1 for subgraphs of random graphs	37
4.3.2	Turán's problem for subgraphs of random graphs	38
5	Concluding remarks	39

1 Introduction

Let H be a fixed graph with k vertices and e edges. In what follows, $o(1)$ terms denote functions of n such that $o(1) \rightarrow 0$ as $n \rightarrow \infty$. It is well known that, for any constant p , asymptotically almost surely the random graph $\mathcal{G}(n, p)$ contains $(1 + o(1))n^k p^e$ labeled (not necessarily induced) copies of H . Throughout this paper, we think of a *labeled copy* of a graph H in a graph G as an injective function from $V(H)$ to $V(G)$ that preserves edges.

Let $k \geq 4$ be a fixed integer and $p \in (0, 1)$. Suppose we have a sequence of graphs $\{G_n\}_{n=1}^\infty$, where G_n has n vertices and $(1 + o(1))p\binom{n}{2}$ edges. We say that $\{G_n\}_{n=1}^\infty$ is (k, p) -*quasi-random*, or simply *quasi-random* for short, if G_n contains $(1 + o(1))n^k p^e$ labeled (not necessarily induced) copies of H for any graph H with k vertices, where e is the number of edges in H . It turns out that, for constant p , this notion of quasi-randomness can be equivalently described in terms of some other properties involving parameters other than the number of subgraphs (see [8, 20] and also [5, Chapter 9]). When $p = o(1)$ (i.e., $p \rightarrow 0$ as $n \rightarrow \infty$), some of these properties fail to describe quasi-randomness in the above sense.

In this paper, we investigate quasi-random sparse graphs. We consider both directed and undirected graphs. In Section 1, we outline some well-known results about quasi-random graphs, when p is constant, as well as a few new results

when $p = o(1)$. In Sections 2 and 3, we state and prove our main results. We present the proofs of our results for undirected graphs in Sections 3.2 and 3.3 and sketch the proof of our result for directed graphs in Section 3.4. In Section 4, we present some auxiliary facts and related work. In the last section, Section 5, we summarize a few open questions.

Terminology and notation. Our terminology and notation are fairly standard. Our $o(1)$ terms refer to functions that tend to 0 as $n \rightarrow \infty$. More generally, $o(f(n))$ denotes a function $g(n)$ such that $g(n)/f(n) \rightarrow 0$ as $n \rightarrow \infty$. We also write $g(n) \ll f(n)$ if $g(n) = o(f(n))$. Moreover, for A , B , and $\delta > 0$, we write $A \sim_\delta B$ to mean $(1 - \delta)B < A < (1 + \delta)B$. Similarly, we write $A \not\sim_\delta B$ if $A \leq (1 - \delta)B$, or $A \geq (1 + \delta)B$. Also, if $f(n)$ and $g(n)$ are functions, we write $f(n) \sim g(n)$ if $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$.

For any integer n , let $[n] = \{1, \dots, n\}$. For any set X , we denote the set of all r -elements subsets of X by $[X]^r$ and we denote the set of all ordered r -tuples of X by $X^r = X \times \dots \times X$. The cardinality of X will be denoted by $|X|$. We use the following non-standard notation. If $U = (u_1, \dots, u_k)$ is an ordered k -tuple, we let $U^{\text{set}} = \{u_1, \dots, u_k\}$ be the set of the elements occurring in the vector U .

Let $G = (V, E)$ be a graph with vertex set $V = V(G)$ and edge set $E = E(G)$. We write $N(x) = N_G(x)$ for the neighborhood of a vertex x in G , and if $X \subset V$, we let $N(X) = N_G(X)$ be the *joint*, or *common*, neighborhood

$$\bigcap_{x \in X} N(x)$$

of the vertices in X . We denote the degree of $x \in V$ by $\deg(x) = \deg_G(x) = |N(x)|$. We denote the number of edges in G by $e(G)$. If $X \subset V$, we sometimes write $e(X) = e_G(X)$ for the number of edges induced by X in G . The maximum and the minimum degree in G are denoted by $\Delta(G)$ and $\delta(G)$. For $X \subset V$, we let $G[X]$ be the graph induced by X in G . Usually, G_n denotes a graph on n vertices. Finally, we call a subset $U \subset V$ *stable* (or *independent*) if there is no edge induced in U .

1.1 The constant density case

The subject of quasi-random graphs was introduced in the eighties by Thomason [20] and Chung, Graham and Wilson [8]. They realized the surprising fact that several important properties shared by almost all graphs are asymptotically equivalent in a deterministic sense. See also [1, 3, 10, 17] for related initial work in this area and [19] for a recent development.

These equivalent properties are satisfied almost surely by a random graph in which every edge is chosen independently with probability $p = 1/2$. In general one may consider a random graph $\mathcal{G}(n, p)$ on n vertices in which every edge is chosen independently with a constant probability $p \in (0, 1)$. Then one can show that the following properties hold for $G_n \in \mathcal{G}(n, p)$ asymptotically almost surely, that is, with probability tending to 1 as $n \rightarrow \infty$.

NSUB(k): For any graph H on k vertices, the number of labeled (not necessarily induced) copies of H in G_n is

$$N(H, G_n) = (1 + o(1))n^k p^e,$$

where e is the number of edges in H .

DISC: For all $X, Y \subset V(G_n)$ with $X \cap Y = \emptyset$, if $e(X, Y)$ denotes the number of edges between X and Y then

$$|e(X, Y) - p|X||Y|| = o(pn^2).$$

EIG: Let $A = (a_{x,y})_{x,y \in V(G_n)}$ denote the 0–1 adjacency matrix of G_n , with 1 denoting edges. Let λ_i ($1 \leq i \leq n$) be the eigenvalues of A and adjust the notation so that $\lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. Then

$$\lambda_1 = (1 + o(1))pn \quad \text{and} \quad |\lambda_2| = o(pn).$$

CYCLE(4): If C_4 denotes the 4-cycle, i.e., the cycle of length 4, then

$$N(C_4, G_n) = (1 + o(1))(pn)^4.$$

TUPLE(s): For all $r \in [s] = \{1, \dots, s\}$, we have

$$||N(x_1) \cap \dots \cap N(x_r)| - np^r| = o(p^r n),$$

for all but at most $o(n^r)$ r -element sets $\{x_1, \dots, x_r\} \subset V(G_n)$.

Above, s and k are arbitrary fixed constants. The following theorem holds (see [8] and [20]).

Theorem 1. *Let $k \geq 4$ be a fixed integer. Let G_n be a graph on n vertices and $(1 + o(1))p \binom{n}{2}$ edges for some fixed $p \in (0, 1)$. If G_n satisfies any of the properties NSUB(k), DISC, EIG, CYCLE(4), and TUPLE(2), then it satisfies all of them.*

Remark 2. Note that the property NSUB(k) depends on a parameter k . It is not hard to show that, for any k , property NSUB($k + 1$) implies property NSUB(k) (see Fact 45). Perhaps quite surprisingly, it follows from Theorem 1 that property NSUB(k) implies property NSUB($k + 1$) as well, as long as $k \geq 4$.

To make our assertions more precise, we may substitute the $o(1)$ terms that appear in the definitions of NSUB(k), DISC, EIG, CYCLE(4), and TUPLE(s) by a parameter $\varepsilon > 0$. We then obtain the properties for n -vertex graphs G_n given below. In what follows, unless explicitly stated otherwise, we let

$$p = p(n) = |E(G_n)| \binom{n}{2}^{-1}.$$

NSUB $_{\varepsilon}(k)$: For any graph H on k vertices, the number of labeled (not necessarily induced) copies of H in G_n satisfies

$$(1 - \varepsilon)n^k p^e < N(H, G_n) < (1 + \varepsilon)n^k p^e,$$

where e is the number of edges in H .

DISC $_{\varepsilon}$: For all $X, Y \subset V(G_n)$ with $X \cap Y = \emptyset$, if $e(X, Y)$ denotes the number of edges between X and Y then

$$|e(X, Y) - p|X||Y|| \leq \varepsilon p n^2.$$

EIG $_{\varepsilon}$: Let $A = (a_{x,y})_{x,y \in V(G_n)}$ denote the 0–1 adjacency matrix of G_n , with 1 denoting edges. Let λ_i ($1 \leq i \leq n$) be the eigenvalues of A and adjust the notation so that $\lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. Then

$$(1 - \varepsilon)pn < \lambda_1 < (1 + \varepsilon)pn \quad \text{and} \quad |\lambda_2| \leq \varepsilon pn.$$

CYCLE $_{\varepsilon}(4)$: We have

$$(1 - \varepsilon)(pn)^4 < N(C_4, G_n) < (1 + \varepsilon)(pn)^4.$$

TUPLE $_{\varepsilon}(s)$: For all $r \in [s] = \{1, \dots, s\}$, we have

$$||N(x_1) \cap \dots \cap N(x_r)| - np^r| < \varepsilon p^r n,$$

for all but at most $\varepsilon \binom{n}{r}$ r -element sets $\{x_1, \dots, x_r\} \subset V(G_n)$.

Remark 3.

(a) The equivalence between two properties in Theorem 1, say P and Q , should be understood in the following way. *Property P implies property Q for a sequence of graphs $\{G_n\}_{n=1}^{\infty}$* (we write $P \Rightarrow Q$ for $\{G_n\}_{n=1}^{\infty}$) if the following holds:

(*) *For all $\varepsilon > 0$, there exist $\delta > 0$ and n_0 such that any graph G_n with $n \geq n_0$ vertices satisfying P_{δ} satisfies Q_{ε} as well. Here, P_{δ} and Q_{ε} stand for P and Q with $o(1)$ replaced by δ and ε respectively.*

(b) We will write “ $P \Rightarrow Q$ ” to mean “ $P \Rightarrow Q$ for $\{G_n\}_{n=1}^{\infty}$ ” when the implicit reference to $\{G_n\}_{n=1}^{\infty}$ is clear from the context.

(c) Suppose we have a sequence of graphs $\{G_n\}_{n=1}^{\infty}$. We may then define the ‘density function’ $p = p(n)$ of this sequence by putting $p = p(n) = |E(G_n)| \binom{n}{2}^{-1}$ for all n . On the other hand, sometimes we prefer to think that we have a *given* function $p = p(n)$, and that our graph sequence $\{G_n\}_{n=1}^{\infty}$ is such that

$$|E(G_n)| = (1 + o(1))p \binom{n}{2}.$$

Although the relationship between $\{G_n\}_{n=1}^\infty$ and $p = p(n)$ in these two approaches are different, we may clearly ignore this small difference when considering implications of the form $P \Rightarrow Q$ with P and Q as above.

The investigation of quasi-randomness, for constant $p \in (0, 1)$, turned out to be a fruitful area with several applications in questions regarding random graphs and algorithms (see, e.g., [4], [9], [13], [15], [18], and [21]). Some of the open questions in this area deal with the problem of generalizing Theorem 1 to the case in which $p = o(1)$.

Before we proceed, we mention that in 1985 Thomason [20, 21] already considered the case in which $p = o(1)$. Our approach in this paper is different from the one taken by Thomason, who investigated pseudorandom properties with error terms that vanish together with p . Our approach is closer in spirit to the one in the recent paper by Chung and Graham [6].

1.2 The vanishing density case

In this section, we turn our attention to the study of quasi-randomness when $p = o(1)$. The first efforts towards this direction suggest that a generalization of Theorem 1 (which is valid when $p \in (0, 1)$ is constant) will not be straightforward. Indeed the quasi-random properties listed above are no longer equivalent when $p = o(1)$. For instance, property TUPLE(2) does not imply property NSUB(3), as we shall see in Proposition 6 below. However, some of these quasi-random properties are equivalent under more restrictive conditions.

Let TFSUB(k) be the property NSUB(k) restricted to triangle-free graphs H , that is, under TFSUB(k) we require the number of occurrences of triangle-free graphs H to be ‘correct’ in G_n (see Definition 12). For suitable values of p (see Theorems 10 and 18), the following diagram holds for “special families of graphs” such as the family BDD(C, t) and the family CG(C, t) (see Definitions 4 and 8).

$$\begin{array}{ccccc}
 \text{TFSUB}(k) & \implies & \text{CYCLE}(4) & & \\
 \updownarrow \text{Theorem 18} & & \downarrow \text{Chung-Graham} & & \\
 \text{TUPLE}(2) & \implies & \text{DISC} & \overset{\text{Chung-Graham}}{\iff} & \text{EIG}
 \end{array}$$

Note that the “missing” link in the above diagram is the implication $\text{DISC} \Rightarrow \text{TUPLE}(2)$. Although this implication does not hold in general (see [6] and [13]), it is possible that it holds under some natural conditions (such as BDD(C, t) and CG(C, t)). If this implication does hold under some special conditions, then the properties CYCLE(4), DISC, EIG, TFSUB(k) and TUPLE(2) would all be equivalent for sequences of graphs satisfying these conditions and BDD (we make this precise in Remark 21 below). For $p = o(1)$, the only known counterexamples to the implication $\text{DISC} \Rightarrow \text{TUPLE}(2)$ are graphs in which

the joint neighborhood of a few vertices is very large, that is, graphs for which the property $\text{BDD}(C, t)$ defined below fails.

Definition 4. Let constants $C > 1$ and $t \geq 1$ be given. We define $\text{BDD}(C, t)$ to be the family of all graphs G such that, if we let $n = |V(G)|$ and $p = |E(G)|/\binom{n}{2}$, then

- (i) $\deg_G(x) \leq Cpn$ for all $x \in V(G)$,
- (ii) for all $2 \leq r \leq t$ and for all distinct vertices $x_1, \dots, x_r \in V(G)$,

$$|N_G(x_1) \cap \dots \cap N_G(x_r)| \leq Cnp^r.$$

Remark 5. Note that $\text{BDD}(C, t+1) \subset \text{BDD}(C, t)$.

Since the property $\text{TFSUB}(k)$ is restricted to the counting of triangle-free graphs only, it is natural to ask whether this counting extends to graphs with triangles. The following proposition, Proposition 6, shows that there is no hope for such an extension for graphs out of the family BDD when $p = o(1)$ and even for graphs in BDD when $p = p(n)$ is of order $n^{-1/3}$.

Proposition 6.

- (A) For any $p = p(n) = o(1)$, there exists a graph sequence $\{G_i\}_{i=1}^\infty$, with $|V(G_i)| = n_i \rightarrow \infty$ as $i \rightarrow \infty$ and $|E(G_i)| \geq p(n_i)\binom{n_i}{2}$ for all i , for which the following holds:
 - (i) G_i is triangle-free for all $i \geq 1$,
 - (ii) $\{G_i\}_{i=1}^\infty$ satisfies properties DISC , EIG , and $\text{TUPLE}(2)$.
- (B) There exists a graph sequence $\{G_i\}_{i=1}^\infty$, with $|V(G_i)| = n_i \rightarrow \infty$ as $i \rightarrow \infty$ and $|E(G_i)| = (1/8 + o(1))n_i^{5/3}$, for which (i) and (ii) above hold and, furthermore,
 - (iii) $G_i \in \text{BDD}(128, 2)$ for all $i \geq 1$.

The proof of Proposition 6 will be discussed in Section 4.2.

Remark 7. It would be interesting to know if one can extend Proposition 6 to the existence of graphs G_i with $|V(G_i)| = n_i \rightarrow \infty$ and $p(n_i) = |E(G_i)|/\binom{n_i}{2} \gg n_i^{-1/3}$ such that G_i satisfies (i) and (ii) in Proposition 6 and $G_i \in \text{BDD}(C, 2)$ for all i for some fixed constant C .

Among other problems, the question of the equivalence of the properties EIG , DISC , and $\text{CYCLE}(4)$, in the sparse setting, was considered in [6] by Chung and Graham. Before we discuss their results, we introduce some terminology.

For any integer t and any two vertices u and v in a graph G , let $e_t(u, v)$ denote the number of paths of length t between u and v . Thus, we always have $e_1(u, v) \leq 1$ and $e_2(u, v) = |N(u) \cap N(v)|$.

Definition 8. Let $t \geq 2$ be an integer and let $C > 1$ be a fixed constant. Let $\text{CG}(C, t)$ denote the family of graphs G such that, putting $n = |V(G)|$ and $p = |E(G)|/\binom{n}{2}$, we have

- (i) $\deg_G(u) \leq Cpn$ for all $u \in V(G)$,
- (ii) $e_t(u, v) \leq Cp^t n^{t-1}$ for all $u, v \in V(G)$.

Remark 9. One can observe that

$$\text{CG}(C, t) \subset \text{CG}(C^2, t+1) \text{ and } \text{CG}(C, 2) = \text{BDD}(C, 2).$$

The next theorem follows from the results of Chung and Graham [6].

Theorem 10. *The implications*

$$\text{CYCLE}(4) \Rightarrow \text{EIG} \Rightarrow \text{DISC} \tag{1}$$

hold for any sequence $\{G_n\}_{n=1}^\infty$ with $|V(G_n)| = n$ and $|E(G_n)| = (1+o(1))p\binom{n}{2}$, as long as $p = p(n) \gg n^{-1/2}$.

Remark 11. Chung and Graham have in fact proved that the implication

$$\text{DISC} \Rightarrow \text{EIG}$$

holds even for fast decreasing functions $p = p(n)$, but assuming an extra hypothesis that can be expressed in terms of the classes $\text{CG}(C, t)$. We refer the interested reader to [6]. In the case in which p is constant, all the three properties in (1) are equivalent (see also Conjecture 20).

The study of quasi-random properties in a *random* setting was considered in [13]. We will discuss this approach in Section 4.3.

2 Statements of the main results

Now we turn our attention to the main goal of this paper. Complementing the work of Chung and Graham [6, 7], we will address the question as to how the property $\text{NSUB}(k)$ relates to the properties $\text{CYCLE}(4)$, DISC , EIG , and $\text{TUPLE}(2)$ in the sparse setting. We shall consider both undirected and directed graphs.

For the digraph case, we focus on the embedding of triangle-free digraphs into sparse pseudorandom digraphs satisfying certain extra conditions. We will only present the proofs of our main results in the undirected case, as they can be naturally extended to the directed case.

2.1 The undirected case

By Proposition 6 the implication “ $\text{TUPLE}(2) \Rightarrow \text{NSUB}(k)$ ” fails to be true for sequences of graphs with vanishing density. Thus, additional conditions are

needed in order to obtain any new relation between $\text{NSUB}(k)$ and the other properties. One such condition is to restrict the family of graphs G for which such a relation could exist. Another possibility is to weaken property $\text{NSUB}(k)$. This leads us to the following two adjustments:

- (i) As in the work of Chung and Graham [6] (see Theorem 10 and Remark 11 above), we restrict the domain to a special family of graphs G , namely, the family $\text{BDD}(C, t)$ introduced in Definition 4.
- (ii) We will also weaken the property $\text{NSUB}(k)$ and focus on counting the triangle-free subgraphs only.

Definition 12. Fix an integer $k \geq 4$. We say that a sequence of graphs $\{G_n\}_{n=1}^\infty$ with $|V(G_n)| = n$ has the property $\text{TFSUB}(k)$ if it satisfies the following condition:

(\dagger) For any triangle-free graph H on k vertices, the number of labeled (not necessarily induced) copies of H in G_n is

$$N(H, G) = (1 + o(1))n^k p^e,$$

where e is the number of edges in H , and $p = e(G_n) \binom{n}{2}^{-1}$.

Note that the only difference between the properties $\text{TFSUB}(k)$ and $\text{NSUB}(k)$ is the *triangle-freeness* condition. We need the following definitions before we may state our first main theorem.

Definition 13. For any graph H , we let

$$d_H = \max\{\delta(J) : J \subset H\}.$$

Remark 14. If a k -vertex graph H is triangle-free, then $d_H \leq k/2$.

Definition 15. For any graph H , we let

$$D_H = \min\{2d_H, \Delta(H)\}.$$

We may now state our key result.

Theorem 16 (Embedding Lemma). Suppose H is a triangle-free graph on k vertices and e edges. Let $\{G_n\}_{n=1}^\infty$ be a sequence of graphs with $|V(G_n)| = n$ for all n and with $p = p(n) = |E(G_n)| \binom{n}{2}^{-1}$ satisfying $p \gg n^{-1/D_H}$. Let $C > 1$ be a fixed constant and suppose that $G_n \in \text{BDD}(C, D_H)$ and G_n satisfies $\text{TUPLE}(2)$ for all n . More explicitly, for all n , we have

- (i) $\deg_{G_n}(x) \leq Cpn$ for all $x \in V(G_n)$,
- (ii) for all $2 \leq r \leq D_H$ and for all distinct vertices $x_1, \dots, x_r \in V(G_n)$,

$$|N_{G_n}(x_1) \cap \dots \cap N_{G_n}(x_r)| \leq Cnp^r,$$

(iii) for all but at most $o(n^2)$ pairs $\{x_1, x_2\} \subset V(G_n)$,

$$| |N_{G_n}(x_1) \cap N_{G_n}(x_2)| - np^2 | = o(np^2).$$

Then G_n contains $(1 + o(1))n^k p^e$ labeled copies of H .

Theorem 16 follows from Lemmas 29 and 30 (see Section 3). Roughly speaking, Theorem 16 states that property TUPLE(2) implies property TFSUB(k) as long as we restrict ourselves to graphs G in an appropriate class BDD(C, s), and the density of G is large enough. We propose the following conjecture.

Conjecture 17. *The parameter D_H occurring in Theorem 16 may be replaced by d_H .*

With Theorem 16 in hand, we may deduce the equivalence of some of the properties introduced in Section 1.1 for sparse graphs. For convenience, from now on $\{G_n\}_{n=1}^\infty$ denotes a sequence of graphs with $|V(G_n)| = n$.

Theorem 18. *Let a real number $C > 1$ and an integer $k \geq 4$ be fixed. Let $p = p(n)$ be a function of n with*

$$np^{\lfloor 2k/3 \rfloor} \gg 1. \tag{2}$$

Then properties TFSUB(k), CYCLE(4), and TUPLE(2) are equivalent for any sequence of graphs $\{G_n\}_{n=1}^\infty$ with

$$G_n \in \text{BDD}(C, \lfloor 2k/3 \rfloor) \tag{3}$$

for all n and $|E(G_n)| = (1 + o(1))p^{\binom{n}{2}}$.

The parameter $\lfloor 2k/3 \rfloor$ in conditions (2) and (3) in Theorem 18 comes from the fact that if we let

$$D(k) = \max_H D_H, \tag{4}$$

where the maximum is taken over all triangle-free graphs on k vertices, then (as proved in Fact 31) we have $D(k) = \lfloor 2k/3 \rfloor$ for all $k \geq 4$.

Remark 19. Note that if Conjecture 17 holds then the condition $np^{\lfloor 2k/3 \rfloor} \gg 1$ in Theorem 18 may be replaced by the weaker condition $np^{\lfloor k/2 \rfloor} \gg 1$.

We also propose the following conjecture.

Conjecture 20. *Let $C > 1$ be an arbitrary constant, and let $p = p(n) \gg n^{-1/2}$ be a function of n . Then the implication*

$$\text{DISC} \Rightarrow \text{TUPLE}(2)$$

holds for any sequence of graphs $\{G_n\}_{n=1}^\infty$ with $|E(G_n)| = (1 + o(1))p^{\binom{n}{2}}$ as long as $G_n \in \text{BDD}(C, 2)$ for all large enough n .

Remark 21. If Conjecture 20 holds, then we may add properties DISC and EIG to the collection of equivalent properties in Theorem 18. Indeed, this follows from the result of Chung and Graham, Theorem 10, stated above.

To embed general graphs (i.e., graphs that are not necessarily triangle-free) we need a stronger property, IND_{TUP}(s) ($s \geq 1$), defined as follows.

IND_{TUP}(s): For all $1 \leq r \leq s$ and all $0 \leq t \leq \binom{r}{2}$,

$$|\{X \in [V(G_n)]^r : e(X) = t, |N(X)| \not\sim np^r\}| = o(n^r p^t).$$

Remark 22. In the definition of IND_{TUP} above, the expression $o(n^r p^t)$ that appears on the right-hand side of the equation would perhaps more appropriately be $o(n^r p^t (1-p)^{\binom{r}{2}-t})$. However, since we are interested in the case in which $p = o(1)$ and $s = O(1)$, we may drop the $(1-p)^{\binom{r}{2}-t}$ factor, which is roughly equal to 1 for such values of p and s .

We may now state our result concerning the embedding of general, not necessarily triangle-free graphs.

Theorem 23. *Let $k \geq 3$ be an integer and let $C > 1$ be a fixed constant. Let $p = p(n) \gg n^{-1/(k-1)}$ be a function of n . Then, for any sequence of graphs $\{G_n\}_{n=1}^\infty$ with $G_n \in \text{BDD}(C, k-1)$ for all n and $|E(G_n)| = (1+o(1))p\binom{n}{2}$, we have*

- (i) NSUB($k+1$) \Rightarrow IND_{TUP}($k-1$),
- (ii) IND_{TUP}($k-1$) \Rightarrow NSUB(k).

Perhaps Theorem 23 may be strengthened to the following.

Conjecture 24. *Let $k \geq 3$ be an integer and let $C > 1$ be a fixed constant. Let $p = p(n) \gg n^{-1/(k-1)}$ be a function of n . Then the properties NSUB(k) and IND_{TUP}($k-1$) are equivalent for any sequence of graphs $\{G_n\}_{n=1}^\infty$ with $G_n \in \text{BDD}(C, k-1)$ for all n and $|E(G_n)| = (1+o(1))p\binom{n}{2}$.*

2.2 The directed case

In this section, we state our main result for directed graphs, Theorem 25. The proof of this result is discussed in Section 3.4. Let \vec{G} be a directed graph with set of vertices V and set of arcs \vec{E} . Thus $\vec{G} = (V, \vec{E})$, where $\vec{E} \subset V \times V \setminus \{(v, v) : v \in V\}$, and if $(u, v) \in \vec{E}$, then $(v, u) \notin \vec{E}$. We denote the out-degree (resp. in-degree) of a vertex $u \in V$ by $d^+(u)$ (resp. $d^-(u)$). We define

$$d^{++}(u, v) = \{w \in V : (u, w) \in \vec{E} \text{ and } (v, w) \in \vec{E}\},$$

and

$$N_{\vec{G}}(u) = \{v \in V : (u, v) \in \vec{E} \text{ or } (v, u) \in \vec{E}\}.$$

For any directed graph \vec{G} we let G be the undirected graph obtained from \vec{G} by transforming its arcs to “edges” (ignoring their orientation). With this convention, clearly $N_{\vec{G}}(u) = N_G(u)$.

We let $\overrightarrow{\text{BDD}}(C, t)$ be the family of all directed graphs \vec{G} such that $G \in \text{BDD}(C, t)$. Moreover, below, given a digraph \vec{H} , we shall consider the parameters d_H and D_H of the associated undirected graph H .

Let $\vec{G} = (V, \vec{E})$ be as above. We introduce the property $\overrightarrow{\text{TUPLE}}(2)$ which is analogous to the property $\text{TUPLE}(2)$ for undirected graphs.

$\overrightarrow{\text{TUPLE}}(2)$: \vec{G} satisfies the property $\overrightarrow{\text{TUPLE}}(2)$ if the following holds

(a) for all but $o(n)$ vertices $u \in V$,

$$d^+(u) = \frac{1}{2}pn(1 + o(1)),$$

(b)

$$\sum_{(u,v) \in V \times V} (d^{++}(u,v))^2 = n^2 \left(\frac{p^2 n}{4} \right)^2 (1 + o(1)).$$

Our embedding result for directed graphs is as follows.

Theorem 25. *Suppose that \vec{H} is a directed graph on k vertices and e arcs with H triangle-free. Let $\{\vec{G}_n\}_{n=1}^\infty$ be a sequence of directed graphs with $|V(\vec{G}_n)| = n$ for all n and with $p = p(n) = |E(\vec{G}_n)| \binom{n}{2}^{-1}$ satisfying $p \gg n^{-1/D_H}$. Suppose $\vec{G}_n \in \overrightarrow{\text{BDD}}(C, D_H)$ and \vec{G}_n satisfies $\overrightarrow{\text{TUPLE}}(2)$ for all n . Then \vec{G}_n contains*

$$N(\vec{H}, \vec{G}) = \frac{1}{2^e} n^k p^e (1 + o(1))$$

labeled copies of \vec{H} .

3 Proofs of the main results

3.1 Preliminaries

The following well-known result will be used often.

Lemma 26. *For all $\eta > 0$, there exists $\varepsilon_0 = \varepsilon_0(\eta) > 0$ such that, for any family of real numbers $\{a_i \geq 0: 1 \leq i \leq n\}$ satisfying the conditions*

$$(i) \sum_{i=1}^n a_i \geq (1 - \varepsilon_0)na,$$

$$(ii) \sum_{i=1}^n a_i^2 \leq (1 + \varepsilon_0)na^2,$$

we have

$$|\{i: a_i \sim_\eta a\}| > (1 - \eta)n.$$

Proof. Let $\eta > 0$ be given. We claim that $\varepsilon_0 = \eta^3/3$ will do. Let a_i ($1 \leq i \leq n$) be as in the statement of our lemma. Set $B = \{i: |a_i - a| \geq \eta a\}$. To prove the lemma, we have to show that $|B| < \eta n$.

From the definition of B , it follows that

$$\sum_{i=1}^n (a_i - a)^2 > |B|(\eta a)^2. \quad (5)$$

By hypothesis,

$$\begin{aligned} \sum_{i=1}^n (a_i - a)^2 &= \sum_{i=1}^n a_i^2 - 2a \sum_{i=1}^n a_i + \sum_{i=1}^n a^2 \\ &\leq (1 + \varepsilon_0)na^2 - 2a(1 - \varepsilon_0)na + na^2 = 3\varepsilon_0na^2. \end{aligned} \quad (6)$$

Combining (5) and (6), we obtain $|B|(\eta a)^2 < 3\varepsilon_0na^2$, which implies that $|B| < (3\varepsilon_0/\eta^2)n = \eta n$, and our lemma is proved. \square

3.2 Proof of Theorems 16 and 18

In this section, we shall prove Theorems 16 and 18. The proof of Theorem 18 is broken down into a few steps, and two of these steps will basically form the proof of Theorem 16 (see Section 3.2.4).

The proof of Theorem 18 involves the following components. Let $k \geq 4$ and $C > 1$ be given. Suppose

$$np^{d(k)} \gg 1, \quad (7)$$

where

$$d(k) = \max_H d_H,$$

and the maximum is taken over all triangle-free graphs H on k vertices. It is easy to see that, in fact, $d(k) = \lfloor k/2 \rfloor$. However, in what follows, we often prefer to write $d(k)$ instead of its explicit value.

Suppose $G_n \in \text{BDD}(C, D(k))$ for all n , where $D(k)$ is as defined in (4). Recall that $D(k) = \lfloor 2k/3 \rfloor$.

Remark 27. The reader may have noticed that our hypothesis on $p = p(n)$ above, namely (7), is weaker than the hypothesis in Theorem 18. It turns out that (7) is the natural hypothesis for the proof we shall present. However, as a simple argument shows, the condition that $\text{BDD}(C, D(k))$ should hold for G_n implies that, in fact, we have $pn^{D(k)} = pn^{\lfloor 2k/3 \rfloor} \gg 1$ (see (2)).

The proof of Theorem 18 is broken down as follows.

(a) Let $\text{NSUB}(C_4)$ be the property $\text{NSUB}(k)$ applied to $H = C_4$. Note that

$$\text{CYCLE}(4) = \text{NSUB}(C_4).$$

Since C_4 is a triangle-free graph and $\text{TFSUB}(k) \Rightarrow \text{TFSUB}(k-1)$ (see Fact 46), the implication

$$\text{TFSUB}(k) \Rightarrow \text{NSUB}(C_4) = \text{CYCLE}(4)$$

is immediate for any any sequence of graphs $\{G_n\}_{n=1}^\infty$ (recall $k \geq 4$).

(b) Fact 28 below, which may be proved by standard arguments, asserts that

$$\text{CYCLE}(4) = \text{NSUB}(C_4) \Rightarrow \text{TUPLE}(2),$$

for any sequence of dense enough graphs $\{G_n\}_{n=1}^\infty$.

(c) Lemma 29 (see below) tells us that

$$\text{TUPLE}(2) \Rightarrow \text{TUPLE}(d(k)).$$

(d) The major piece in the proof of Theorem 18 is the implication

$$\text{TUPLE}(d(k)) \Rightarrow \text{TFSUB}(k).$$

This implication is stated in an equivalent form in Lemma 30 and its proof constitutes the main task of this paper (Section 3.2.3).

(e) Finally the fact that $D(k) = \lfloor 2k/3 \rfloor$ is proved in Fact 31.

Steps (c) and (d) above basically constitute the proof of Theorem 16 (see Section 3.2.4).

Fact 28. *Let $C > 1$ be a constant and suppose $p = p(n)$ is such that $np^2 \gg 1$. Then the implication*

$$\text{CYCLE}(4) \Rightarrow \text{TUPLE}(2)$$

holds for any sequence of graphs $\{G_n\}_{n=1}^\infty$ with $|E(G_n)| = (1 + o(1))p \binom{n}{2}$.

Proof. Let $\{G_n\}_{n=1}^\infty$ be as in the statement of our result and suppose $\text{CYCLE}(4)$ holds. We have

$$\begin{aligned} \sum_{\{x,y\} \subset V, x \neq y} |N_{G_n}(x) \cap N_{G_n}(y)| &= \sum_{v \in V} \binom{\deg(v)}{2} \\ &\geq n \binom{n^{-1} \sum_{v \in V} \deg(v)}{2} = n \binom{(1 + o(1))pn}{2} \\ &= (1 + o(1)) \binom{n}{2} p^2 n. \end{aligned} \quad (8)$$

Observe that the number of labeled (not necessarily induced) copies of C_4 in G_n is

$$N(C_4, G_n) = 4 \sum_{\{x,y\} \subset V, x \neq y} \binom{|N_{G_n}(x) \cap N_{G_n}(y)|}{2}. \quad (9)$$

Since $\{G_n\}_{n=1}^\infty$ satisfies property CYCLE(4), we have

$$\sum_{\{x,y\} \subset V, x \neq y} \binom{|N_{G_n}(x) \cap N_{G_n}(y)|}{2} = \frac{1}{4}N(C_4, G_n) = \frac{1}{4}(1 + o(1))(pn)^4. \quad (10)$$

We now observe that the Cauchy–Schwarz inequality tells that

$$\begin{aligned} \sum_{\{x,y\} \subset V, x \neq y} |N_{G_n}(x) \cap N_{G_n}(y)|^2 &\geq \binom{n}{2}^{-1} \left\{ \sum_{\{x,y\} \subset V, x \neq y} |N_{G_n}(x) \cap N_{G_n}(y)| \right\}^2 \\ &\gg \sum_{\{x,y\} \subset V, x \neq y} |N_{G_n}(x) \cap N_{G_n}(y)|, \end{aligned} \quad (11)$$

where in the last inequality we used (8) and the fact that $p^2n \rightarrow \infty$ as $n \rightarrow \infty$. Now from (10) and (11), we obtain that

$$\sum_{\{x,y\} \subset V, x \neq y} |N_{G_n}(x) \cap N_{G_n}(y)|^2 = \frac{1}{2}(1 + o(1))(pn)^4. \quad (12)$$

Now Lemma 26 together with (8) and (12) imply that

$$|N_{G_n}(x) \cap N_{G_n}(y)| = (1 + o(1))p^2n,$$

for all but at most $o(n^2)$ pairs $\{x, y\} \subset V$. \square

We will sketch the proof of Lemma 29 (stated below) in Section 3.4, by deriving it as a corollary of Lemma 43. However, we mention that Lemma 29 was first proved in Luczak et al. [16]; the proof of Lemma 43 is a simple extension of the proof in [16] to the directed case.

Lemma 29. *Let $t \geq 2$ and $C > 1$ be fixed and suppose $p = p(n)$ satisfies $np^r \gg 1$. Let $\{G_n\}_{n=1}^\infty$ be a sequence of graphs with $G_n \in \text{BDD}(C, 2)$ for all n and $|E(G_n)| = (1 + o(1))p \binom{n}{2}$. Then the implication*

$$\text{TUPLE}(2) \Rightarrow \text{TUPLE}(r)$$

holds for $\{G_n\}_{n=1}^\infty$.

To complete the proof of Theorem 18, we need to prove Lemma 30 and Fact 31 below.

Let H be an arbitrary triangle-free graph on k vertices and e edges. Recall

$$d_H = \max_{J \subset H} \delta(J)$$

and

$$D_H = \min\{2d_H, \Delta(H)\}$$

(see Definitions 13 and 15).

Lemma 30. *Let $\delta > 0$, $C > 1$, and $k \geq 4$ be fixed. Let H be as above and let $p = p(n) = o(1)$ be a function of n satisfying $np^{D_H} \gg 1$. Then there exist $\varepsilon > 0$ and an integer n_2 for which the following holds. If a sequence of graphs $\{G_n\}_{n=1}^\infty$ with $|V(G_n)| = n$ is such that, for all n ,*

$$(i) \ G_n \in \text{BDD}(C, D_H),$$

$$(ii) \ p = p(n) = e(G_n) \binom{n}{2}^{-1},$$

(iii) $\text{TUPLE}_\varepsilon(d_H)$ holds for G_n ,

then

$$N(H, G_n) \sim_\delta n^k p^e$$

holds for all $n \geq n_2$.

The proof of Lemma 30 is somewhat involved and is delayed until Section 3.2.3. We finish this section with the statement of Fact 31.

Fact 31. *Let an integer $k \geq 1$ be given and let $D(k) = \max_H D_H$, where the maximum is taken over all triangle-free graphs H on k vertices. Then*

$$D(k) = \left\lfloor \frac{2k}{3} \right\rfloor.$$

Fact 31 is proved in Section 3.2.1 below.

3.2.1 Degenerate orderings

The following simple definition will be important.

Definition 32. *Let H be a graph. We say that H is d -degenerate if there is an ordering v_1, \dots, v_k of the vertices of H such that $\deg_{H_i}(v_i) \leq d$ for all $1 \leq i \leq k$, where $H_i = H[\{v_1, \dots, v_i\}]$ is the graph induced by $\{v_1, \dots, v_i\}$ in H . Moreover, if H is d -degenerate and this is certified by a certain ordering of the vertices of H , then we call this ordering a d -degenerate ordering of H .*

Remark 33. Let $d = d_H = \max\{\delta(J) : J \subset H\}$. Then H has a d -degenerate ordering. Indeed we can find such an ordering in the following way. First select a vertex $v \in V(H)$ such that $\deg_H(v) = \delta(H) \leq d_H$ (by the definition of d_H) and set $v_k = v$. Let $H_{k-1} = H - v_k$, then we repeat the same procedure on H_{k-1} and obtain a vertex v_{k-1} with $\deg_{H_{k-1}}(v_{k-1}) \leq \delta(H_{k-1}) \leq d_H$. Continuing in this way, we obtain the desired ordering of $V(H)$ after $k = |V(H)|$ steps. In fact, d_H is the smallest integer d for which H admits a d -degenerate ordering.

We may now prove Fact 31.

Proof. Suppose $k \geq 1$ is given, and let $D(k)$ be as in the statement of the fact. We may clearly suppose that $k \geq 2$.

First we show that $D(k) \geq \lfloor 2k/3 \rfloor$. It suffices to exhibit a triangle-free graph H on k vertices for which $D_H \geq \lfloor 2k/3 \rfloor$. We show that the complete

bipartite graph $H = K(\lceil k/3 \rceil, \lfloor 2k/3 \rfloor)$ with vertex classes of cardinality $\lceil k/3 \rceil$ and $\lfloor 2k/3 \rfloor$ will do. We have $d_H = \max_{J \subset H} \delta(J) \geq \delta(H) = \lceil k/3 \rceil$. Therefore $2d_H \geq 2\lceil k/3 \rceil \geq 2k/3 \geq \lfloor 2k/3 \rfloor$. Since $\Delta(H) = \lfloor 2k/3 \rfloor$, we have $D_H = \min\{2d_H, \Delta(H)\} \geq \lfloor 2k/3 \rfloor$.

Let us now show that $D(k) \leq \lfloor 2k/3 \rfloor$. To that end, let H be a triangle-free graph on k vertices. We show that $D_H \leq 2k/3$. Suppose $\Delta(H) > 2k/3$. Let u be a vertex of H with maximum degree, and suppose v_1, \dots, v_k is an ordering of the vertices of H with the last $\Delta(H)$ vertices $v_{k-\Delta(H)+1}, \dots, v_k$ forming the neighborhood of u in H . We claim that this is a $(\lceil k/3 \rceil - 1)$ -degenerate ordering of the vertices of H .

To see this, let $H_h = H[\{v_1, \dots, v_h\}]$ for all $1 \leq h \leq k$. Since H is triangle-free, every vertex v_h with $k - \Delta(H) + 1 \leq h$ has its neighborhood contained in the set $\{v_1, \dots, v_{k-\Delta(H)}\}$. Thus $\deg_h(v_h) \leq k - \Delta(H) < k/3$ for all $k - \Delta(H) + 1 \leq h \leq k$, and hence for all $1 \leq h \leq k$. Therefore $\deg_h(v_h) \leq \lceil k/3 \rceil - 1$ for all h and we do indeed have a $(\lceil k/3 \rceil - 1)$ -degenerate ordering as claimed. Hence $2d_H \leq 2(\lceil k/3 \rceil - 1) \leq 2k/3$, and hence $D_H = \min\{2d_H, \Delta(H)\} \leq 2k/3$, as required. \square

3.2.2 The extension lemma and a corollary

In this section, we shall establish a simple lemma, the *Extension Lemma*, and a corollary, Corollary 36. They will be used in the proofs of Theorems 16, 18, and 23.

Let H and G be graphs. In what follows, H will always have k vertices and e edges and G will always have n vertices. In this section, H is an arbitrary graph; in Sections 3.2.3 and 3.2.4, we shall consider triangle-free graphs H .

Let $\mathcal{E}(H, G)$ denote the set of all embeddings of H in G . Moreover, if $l \in [k]$ and $F = (v_1, \dots, v_l) \in V(H)^l$ and $X = (x_1, \dots, x_l) \in V(G)^l$, let $\mathcal{E}(H, G, F, X)$ denote the set of all embeddings $f \in \mathcal{E}(H, G)$ such that $f(v_i) = x_i$ for all $i \in [l]$. Clearly, we may always assume that the v_i ($1 \leq i \leq l$) and the x_i ($1 \leq i \leq l$) are all distinct. Recall that $F^{\text{set}} = \{v_1, \dots, v_l\}$ and $X^{\text{set}} = \{x_1, \dots, x_l\}$.

Below, for any graph H' and any l -tuple F of vertices of H' , we write $w(H', F)$ for the number of edges in H' that do not have both endpoints in F^{set} . That is,

$$w(H', F) = |E(H')| - |E(H'[F^{\text{set}}])|.$$

We now prove the following simple lemma.

Lemma 34 (Extension Lemma). *Let graphs G and H be given. Suppose $0 \leq l \leq \max\{2, d_H\}$, and let $F \in V(H)^l$ and $X \in V(G)^l$ be fixed. Let $C > 0$ be a constant and suppose $G \in \text{BDD}(C, D_H)$. Then*

$$|\mathcal{E}(H, G, F, X)| \leq C^{k-l} n^{k-l} p^{w(H, F)},$$

where $k = |V(H)|$, $n = |V(G)|$, and $p = e(G) \binom{n}{2}^{-1}$. In particular, if $F^{\text{set}} \subset V(H)$ is a stable set, then

$$|\mathcal{E}(H, G, F, X)| \leq C^{k-l} n^{k-l} p^e,$$

where $e = |E(H)|$.

In Claim 35 below, we prove the Extension Lemma under a stronger hypothesis. We then show that the hypothesis of this claim is satisfied even with the weaker assumption of the Extension Lemma.

Claim 35. *Let G, H, F and X be as in Lemma 34. Assume (in addition to the hypotheses of Lemma 34) that there exists a D_H -degenerate ordering v_1, \dots, v_k of H such that $F = \{v_1, \dots, v_\ell\}$. Then*

$$|\mathcal{E}(H, G, F, X)| \leq C^{k-l} n^{k-l} p^{w(H, F)},$$

where $k = |V(H)|$, $n = |V(G)|$, and $p = e(G) \binom{n}{2}^{-1}$.

Proof. Consider a D_H -degenerate ordering v_1, \dots, v_k of H with $F^{\text{set}} = \{v_1, \dots, v_\ell\}$. We shall prove that

(*) for all $l \leq h \leq k$, we have

$$|\mathcal{E}(H_h, G, F, X)| \leq C^{h-l} n^{h-l} p^{w(H_h, F)}, \quad (13)$$

where $H_h = H[\{v_1, \dots, v_h\}]$.

We prove (*) by induction on h . The case in which $h = l$ is clear. Now suppose that $l < h \leq k$ and that (13) holds for smaller values of h . We wish to prove (13). To that end, first observe that, by our choice of the ordering v_1, \dots, v_k of the vertices of H , we have $\deg_{H_h}(v_h) \leq D_H$. Therefore, as $G \in \text{BDD}(C, D_H)$, if we let $r = \deg_{H_h}(v_h)$, then any embedding of H_{h-1} can be extended in at most Cnp^r ways to an embedding of H_h . Using the induction hypothesis and the fact that $w(H_h, F) = w(H_{h-1}, F) + r$, we have

$$\begin{aligned} |\mathcal{E}(H_h, G, F, X)| &\leq Cnp^r |\mathcal{E}(H_{h-1}, G, F, X)| \\ &\leq Cnp^r \times C^{h-l-1} n^{h-l-1} p^{w(H_{h-1}, F)} = C^{h-l} n^{h-l} p^{w(H_h, F)}, \end{aligned}$$

verifying (13). This completes the induction step and assertion (*) follows by induction. Our lemma follows on setting $h = k$ in (13). \square

Proof of Lemma 34. To prove Lemma 34, it suffices to show that there exists a D_H -degenerate ordering v_1, \dots, v_k of H such that $F^{\text{set}} = \{v_1, \dots, v_\ell\}$. Indeed, we may then simply apply Claim 35. In the remainder of this proof, we prove the existence of such a D_H -degenerate ordering.

We distinguish the following two cases.

Case 1: $d_H = 1$

Fix a 1-degenerate ordering $L = v_1, \dots, v_k$ of H . By hypothesis, $|F| \leq \max\{2, d_H\} = 2$. If $F^{\text{set}} = \emptyset$, then our fixed 1-degenerate ordering L will do. If $F^{\text{set}} = \{v_i\}$, we consider the new ordering

$$L' = v_i, v_1, \dots, \widehat{v_i}, \dots, v_k,$$

where \widehat{x} means that the element x is omitted in the listing of L' . Since L is a 1-degenerate ordering, it follows that L' is a 2-degenerate ordering.

If $F^{\text{set}} = \{v_i, v_j\}$, let L_{v_i} and L_{v_j} be the set of vertices in the ‘left neighborhood’ of v_i and v_j in the ordering L , respectively. Clearly, $|L_{v_i} \cap L_{v_j}| \leq 1$ because L is a 1-degenerate ordering. This leads us to the following possibilities.

(i) $L_{v_i} \cap L_{v_j} = \emptyset$

Consider the ordering

$$L' = v_i, v_j, v_1, \dots, \widehat{v_i}, \dots, \widehat{v_j}, \dots, v_k.$$

Since $L = v_1, \dots, v_k$ is a 1-degenerate ordering and $L_{v_i} \cap L_{v_j} = \emptyset$, it is clear that L' is a 2-degenerate ordering as required.

(ii) $L_{v_i} \cap L_{v_j} \neq \emptyset$

Suppose $L_{v_i} \cap L_{v_j} = \{v_s\}$. Clearly, $s < i, j$. Consider the ordering

$$L' = v_i, v_j, v_s, v_1, \dots, \widehat{v_s}, \dots, \widehat{v_i}, \dots, \widehat{v_j}, \dots, v_k.$$

The vertex v_s has 2 left neighbors (v_i and v_j) in the ordering L' . Furthermore, since H is a forest (because $d_H = 1$), any vertex $u \notin \{v_i, v_j, v_s\}$ is joined to at most one vertex in $\{v_i, v_j, v_s\}$. Therefore the left degree of any vertex u in the ordering L' is at most 2. Consequently L' is a 2-degenerate ordering as required.

Case 2: $d_H \geq 2$

By Remark 33, our graph H has a d_H -degenerate ordering L . We now observe that if relocate the vertices in F at the beginning of that ordering, then we obtain a D_H -degenerate ordering of the vertices of H . To see this, let $L' = v_1, \dots, v_k$ be this ordering. If $D_H = \Delta(H)$, then clearly *any* ordering of $V(H)$ is a D_H -ordering. Thus suppose that $D_H = 2d_H$. As in Definition 32, let $H_h = H[\{v_1, \dots, v_h\}]$ ($0 \leq h \leq k$). Owing to the assumption that $l \leq \max\{2, d_H\} = d_H$, we have

$$\deg_{H_h}(v_h) \leq l + d_H \leq 2d_H = D_H.$$

This proves that there exist a D_H -degenerate ordering $L' = v_1, \dots, v_k$ of H such that $F = \{v_1, \dots, v_\ell\}$, as required. \square

The following notation will be used in the next corollary. Set

$$\mathcal{E}^{\text{ni}}(H, G) = \{f \in \mathcal{E}(H, G) : f \text{ is a non-induced embedding}\}.$$

Corollary 36. *Let $C > 1$, $k \geq 1$, and $\eta > 0$ be fixed and let $p = p(n) = o(1)$ be a function of n . Then there exists an integer n_1 such that, for any graph H with k vertices and any graph $G \in \text{BDD}(C, D_H)$ with $|E(G)| \leq pn^2$ and $n = |V(G)| \geq n_1$, we have*

$$|\mathcal{E}^{\text{ni}}(H, G)| < \eta n^k p^e, \tag{14}$$

where and $e = |E(H)|$.

Proof. Let η , p , H and G be as in the statement of the corollary. The case in which $k = 1$ is clear, and hence we suppose $k \geq 2$. To count non-induced embeddings of H in G , we select an edge $\{x, y\} \in E(G)$ and a pair u, v of distinct, non-adjacent vertices of H . By Lemma 34 applied to $F = (u, v)$ and $X = (x, y)$, the number of embeddings $f: V(H) \rightarrow V(G)$ such that $f(u) = x$ and $f(v) = y$ is at most $C^{k-2}n^{k-2}p^e$.

Since $\{x, y\} \in E(G)$ can be selected in at most pn^2 ways, the ordered pair X can be selected in $\leq 2pn^2$ ways. Similarly, F can be selected in $\leq 2\binom{k}{2}$ ways. Therefore

$$|\mathcal{E}^{\text{ni}}(H, G)| \leq 4pn^2 \binom{k}{2} C^{k-2}n^{k-2}p^e < 2k^2 C^{k-2}n^k p^{e+1}.$$

Since $p = o(1)$ and C , k , and $\eta > 0$ are constants, there exists an integer n_1 such that (14) holds for all $n \geq n_1$, as required. \square

3.2.3 Proof of Lemma 30

This section is devoted to the proof of Lemma 30. We start by introducing some notation and terminology. Let G_n be an n -vertex graph with $p = e(G_n)\binom{n}{2}^{-1}$. For every integer $r \geq 1$ and real $\varepsilon > 0$, we let

$$\mathcal{B}(\varepsilon, r) = \{X \in [V(G_n)]^r : |N_{G_n}(X) - np^r| \geq \varepsilon np^r\},$$

and

$$\mathcal{B}_{\text{stb}}(\varepsilon, r) = \{X \in \mathcal{B}(\varepsilon, r) : X \text{ is a stable set in } G_n\}.$$

A set $B \subset V(G_n)$ will be said to be ε -bad if $B \in \mathcal{B}_{\text{stb}}(\varepsilon, r)$ for some r with $1 \leq r \leq d_H$.

If G_n satisfies $\text{TUPLE}_\varepsilon(d_H)$, we have

$$|\mathcal{B}_{\text{stb}}(\varepsilon, r)| \leq |\mathcal{B}(\varepsilon, r)| < \varepsilon \binom{n}{r}$$

for all $r \leq d_H$.

Let us fix a triangle-free graph H as in the statement of Lemma 30. We shall also fix a d_H -degenerate ordering v_1, \dots, v_k of the vertices of H . As before, we let $H_h = H[\{v_1, \dots, v_h\}]$ ($1 \leq h \leq k$). The next definition introduces several important terms for our proof.

Definition 37. For (i)–(iii) below, we suppose that $1 < h \leq k$.

- (i) An embedding $f: V(H_{h-1}) \rightarrow V(G_n)$ is clean if the set $f(N_{H_h}(v_h))$ is not ε -bad; i.e., $f(N_{H_h}(v_h)) \notin \mathcal{B}_{\text{stb}}(\varepsilon, r)$ for any r with $1 \leq r \leq d_H$. Otherwise f is polluted. When we use the terms ‘clean’ and ‘polluted’, the value of ε will be clear from the context.

- (ii) Set

$$\mathcal{E}_{\text{poll}}(H_{h-1}, G_n) = \{f \in \mathcal{E}(H_{h-1}, G_n) : f \text{ is polluted}\}.$$

(iii) Finally, set

$$\mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n) = \{f \in \mathcal{E}(H_{h-1}, G_n) : f \text{ is clean and induced}\}.$$

Now we are ready to state another corollary of the Extension Lemma, Corollary 38 below. This corollary, along with Corollary 36, will be the key ingredients in the proof of Lemma 30.

Corollary 38. *Let $\varepsilon > 0$, $C > 1$, and $k \geq 4$ be fixed. Suppose $1 < h \leq k$ and set $r = \deg_{H_h}(v_h)$. If $G_n \in \text{BDD}(C, D_H)$ satisfies $\text{TUPLE}_\varepsilon(d_H)$, then*

$$|\mathcal{E}_{\text{poll}}(H_{h-1}, G_n)| \leq \varepsilon C^{h-r-1} n^{h-1} p^{e(H_{h-1})},$$

where $p = e(G_n) \binom{n}{2}^{-1}$. In particular, for any $\eta > 0$, $C > 1$, and k , there is an $\varepsilon > 0$ that guarantees that $|\mathcal{E}_{\text{poll}}(H_{h-1}, G_n)| \leq \eta n^{h-1} p^{e(H_{h-1})}$.

Proof. By definition, an embedding f of H_{h-1} in G_n is polluted if $f(N_{H_h}(v_h)) \in \mathcal{B}_{\text{stb}}(\varepsilon, r)$. Fix an r -tuple F such that $F^{\text{set}} = N_{H_h}(v_h)$. We have

$$\mathcal{E}_{\text{poll}}(H_{h-1}, G_n) = \bigcup_X \mathcal{E}(H_{h-1}, G_n, F, X),$$

where the union is taken over all r -tuples X such that $X^{\text{set}} \in \mathcal{B}_{\text{stb}}(\varepsilon, r)$. Therefore

$$|\mathcal{E}_{\text{poll}}(H_{h-1}, G_n)| \leq \sum_X |\mathcal{E}(H_{h-1}, G_n, F, X)|, \quad (15)$$

where the sum is over the same r -tuples X . Since $\text{TUPLE}_\varepsilon(d_H)$ holds for G_n and $r = \deg_{H_h}(v_h) \leq d_H$, the number of r -tuples X that we are summing over in (15) is at most εn^r . Observe also that $N_{H_h}(v_h)$ is a stable set in H_h , because $H_h \subset H$ is triangle-free. We now apply Lemma 34 to deduce from (15) that $|\mathcal{E}_{\text{poll}}(H_{h-1}, G_n)|$ is at most

$$\varepsilon n^r \times C^{h-r-1} n^{h-r-1} p^{e(H_{h-1})} = \varepsilon C^{h-r-1} n^{h-1} p^{e(H_{h-1})},$$

and our corollary follows. \square

We are now ready to prove Lemma 30. We start by outlining the idea of the proof.

Proof strategy for Lemma 30. The proof uses an inductive argument. To keep the induction step working, we need the Extension Lemma, Lemma 34. This lemma yields an upper bound on the number of those ‘‘copies’’ of H in G_n that contain a fixed copy of $H[F] \subset H$ for some $F \subset V(H)$.

Next, we use Corollary 36 to infer that most of the copies of H in G_n are induced copies. Then we further restrict the domain to a certain class of embeddings, called *clean* embeddings, and show that the number of *polluted* (i.e., not clean) embeddings is negligible. This enables us to reduce the proof of Lemma 30 to the special case when the embeddings of H in G_n are clean and induced.

Proof of Lemma 30. Throughout this proof, we suppose that $C > 1$ is a fixed constant and that $G_n \in \text{BDD}(C, D_H)$. We let $p = e(G_n) \binom{n}{2}^{-1}$, and suppose that $np^{d_H} \geq np^{D_H} \gg 1$. Recall that we have a fixed d_H -degenerate ordering v_1, \dots, v_k of the vertices of H , and that $H_h = H[\{v_1, \dots, v_h\}]$ ($1 \leq h \leq k$).

We shall prove by induction on h that

(**) for all $1 \leq h \leq k$ and all $\delta > 0$, there is $\varepsilon > 0$ such that if G_n satisfies $\text{TUPLE}_\varepsilon(d_H)$, then

$$|\mathcal{E}(H_h, G_n)| \sim_\delta n^h p^{e(H_h)}, \quad (16)$$

as long as n is sufficiently large.

Note that (16) clearly holds for any $\delta > 0$ for $h = 1$. Now suppose that $1 < h \leq k$ and that (16) holds for smaller values of h for all $\delta > 0$. Let $\delta > 0$ be given. We wish to show that (16) holds if G_n satisfies $\text{TUPLE}_\varepsilon(d_H)$ for small enough ε and n is large enough.

We start by showing the lower bound, that is, $|\mathcal{E}(H_h, G_n)| > (1 - \delta)n^h p^{e(H_h)}$. Let $\delta' = \min\{\delta/4, \delta/2C\}$, and let $\varepsilon' = \varepsilon'(\delta')$ be the value of ε given by the induction hypothesis to guarantee that

$$|\mathcal{E}(H_{h-1}, G_n)| \sim_{\delta'} n^{h-1} p^{e(H_{h-1})}, \quad (17)$$

as long as n is sufficiently large. Now put $\eta = \delta'/2$. Corollary 36 tells us that if n is large enough, then

$$|\mathcal{E}^{\text{ni}}(H_{h-1}, G_n)| \leq \eta n^{h-1} p^{e(H_{h-1})}. \quad (18)$$

Also, let $\varepsilon'' = \varepsilon''(\eta)$ be the value of ε whose existence is guaranteed in Corollary 38 to ensure that

$$|\mathcal{E}_{\text{poll}}(H_{h-1}, G_n)| \leq \eta n^{h-1} p^{e(H_{h-1})}. \quad (19)$$

We now let $\varepsilon = \min\{\varepsilon', \varepsilon'', \delta/8\}$, and claim that this choice of ε will do. Our induction step is reduced to proving this claim.

For future reference, observe that we have

$$(1 - 2\delta')(1 - 2\varepsilon) \geq 1 - \delta, \quad (20)$$

$$(1 + \delta')(1 + \varepsilon) \leq 1 + \frac{\delta}{2}, \quad (21)$$

and

$$\delta' C \leq \frac{\delta}{2}. \quad (22)$$

Let $r = \deg_{H_h}(v_h) \leq d_H$. Note that then $e(H_{h-1}) = e(H_h) - r$. By our choice of ε , if n is sufficiently large, then the number of embeddings in $\mathcal{E}(H_{h-1}, G_n)$ that are polluted or non-induced is

$$\leq 2\eta n^{h-1} p^{e(H_{h-1})} = \delta' n^{h-1} p^{e(H_{h-1})} = \delta' n^{h-1} p^{e(H_h) - r}$$

(see (18) and (19)). Hence, by (17), the number $|\mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)|$ of clean induced embeddings of H_{h-1} in G_n is such that

$$(1 - 2\delta')n^{h-1}p^{e(H_h)-r} < |\mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)| < (1 + \delta')n^{h-1}p^{e(H_h)-r}. \quad (23)$$

Given $f' \in \mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)$, we may estimate from below the number of embeddings $f \in \mathcal{E}(H_h, G_n)$ that extend f' as follows. Since f' is clean, by definition $f'(N_{H_h}(v_h)) \notin \mathcal{B}_{\text{stb}}(\varepsilon, r)$. Equivalently, either

(a) $f'(N_{H_h}(v_h))$ is not a stable set in G , or

(b) $|N_{G_n}(N_{H_h}(v_h)) - np^r| < \varepsilon np^r$ holds.

Since H is triangle-free, the set $N_{H_h}(v_h)$ is a stable set in H_h . Since f' is induced, the set $f'(N_{H_h}(v_h))$ is also a stable set and consequently (a) fails to hold. Thus (b) must hold, that is,

$$|N_{G_n}(N_{H_h}(v_h)) - np^r| < \varepsilon np^r. \quad (24)$$

Note that, to obtain an extension $f \in \mathcal{E}(H_h, G_n)$ of f' , we must simply select $f(v_h)$ in $N_{G_n}(f'(N_{H_h}(v_h))) \setminus f'(V(H_{h-1}))$. Consequently, the number of extensions of f' to embeddings of H_h in G_n is at least

$$\begin{aligned} & |N_{G_n}(f'(N_{H_h}(v_h))) \setminus f'(V(H_{h-1}))| \\ & \geq (1 - \varepsilon)np^r - (h - 1) \geq (1 - 2\varepsilon)np^r, \end{aligned} \quad (25)$$

where we used (24), the fact that $np^r \geq np^{d_H} \gg 1$, and that n is large. Combining (20), the lower bound in (23), and (25), we obtain that

$$|\mathcal{E}(H_h, G_n)| > (1 - 2\delta')n^{h-1}p^{e(H_h)-r}(1 - 2\varepsilon)np^r \geq (1 - \delta)n^h p^{e(H_h)}. \quad (26)$$

Now we need to show that $|\mathcal{E}(H_h, G_n)| < (1 + \delta)n^h p^{e(H_h)}$. Fix $f' \in \mathcal{E}(H_{h-1}, G_n)$. The number of extensions of f' to embeddings of H_h in G_n is bounded from above by

$$|N_{G_n}(f'(N_{H_h}(v_h)))|. \quad (27)$$

If, furthermore, $f' \in \mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)$, then we know that (24) holds and hence the quantity in (27) is $\leq (1 + \varepsilon)np^r$. Combining this fact with the upper bound in (23) and recalling (21), we obtain that the number of embeddings $f \in \mathcal{E}(H_h, G_n)$ whose restrictions to $V(H_{h-1})$ are in $\mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)$ is

$$\begin{aligned} & < (1 + \delta')n^{h-1}p^{e(H_h)-r}(1 + \varepsilon)np^r \\ & = (1 + \delta')(1 + \varepsilon)n^h p^{e(H_h)} \leq \left(1 + \frac{\delta}{2}\right)n^h p^{e(H_h)}. \end{aligned} \quad (28)$$

We already know that $|\mathcal{E}(H_{h-1}, G_n) \setminus \mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)| \leq \delta'n^{h-1}p^{e(H_h)-r}$. Since $r = \deg_{G_n}(v_h) \leq d_H \leq D_H$ and $G_n \in \text{BDD}(C, D_H)$, each such embedding f' gives rise to $\leq Cpn^r$ embeddings $f \in \mathcal{E}(H_h, G_n)$. Therefore, the

number of embeddings $f \in \mathcal{E}(H_h, G_n)$ whose restrictions to $V(H_{h-1})$ are not in $\mathcal{E}_{\text{clean}}^{\text{ind}}(H_{h-1}, G_n)$ is, by (22),

$$\leq \delta' n^{h-1} p^{e(H_h)-r} \times C n p^r \leq \frac{\delta}{2} n^h p^{e(H_h)}. \quad (29)$$

From (28) and (29), we deduce that

$$|\mathcal{E}(H_h, G_n)| < (1 + \delta) n^h p^{e(H_h)}. \quad (30)$$

Inequalities (26) and (30) complete our induction step, and hence (**) follows by induction. Lemma 30 follows on taking $h = k$ in (**). \square

3.2.4 Proof of Theorem 16

In this short section, we observe that we have already done all the work to prove Theorem 16. Indeed, let H and $\{G_n\}_{n=1}^\infty$ be as in the statement of Theorem 16. We first observe that we may boost hypothesis (iii) in the statement of that theorem to $\text{TUPLE}(d_H)$, by applying Lemma 29. But then we are in condition to apply Lemma 30. We leave the details to the reader.

3.3 Proof of Theorem 23

Throughout this section, H will be a (not necessarily triangle-free) graph on k vertices and e edges. Recall that we denote the set of embedding of H in a graph G by $\mathcal{E}(H, G)$. The set of *induced* embeddings of H in G will be denoted by $\mathcal{E}^{\text{ind}}(H, G)$, and the set of *non-induced* embeddings of H in G will be denoted by $\mathcal{E}^{\text{ni}}(H, G)$.

To prove Theorem 23, we need to prove the implications

$$\text{NSUB}(k+1) \Rightarrow \text{INDTUP}(k-1)$$

and

$$\text{INDTUP}(k-1) \Rightarrow \text{NSUB}(k),$$

for all appropriate sequences of graphs $\{G_n\}_{n=1}^\infty$. The implications above will be proved in Lemmas 39 and 41 below.

Lemma 39. *Let $k \geq 3$ and $C > 1$ be fixed. Let $p = p(n) = o(1)$ be a function of n satisfying $np^{k-1} \gg 1$. Then*

$$\text{NSUB}(k+1) \Rightarrow \text{INDTUP}(k-1)$$

for any sequence of graphs $\{G_n\}_{n=1}^\infty$ with $p = p(n) = e(G_n) \binom{n}{2}^{-1}$ and $G_n \in \text{BDD}(C, k-1)$ for all n .

Proof. We shall be somewhat sketchy in this proof. Let the sequence of graphs $\{G_n\}_{n=1}^\infty$ be as in the statement of our lemma. Suppose that $\text{INDTUP}(k-1)$ fails to hold. We will show that $\text{NSUB}(k+1)$ fails to hold as well. By definition

of $\text{INDTUP}(k-1)$, we know that there are integers $1 \leq r < k$ and $0 \leq t \leq \binom{r}{2}$ for which we have

$$|\text{Bad}^{\text{ind}}(r, t)| \neq o(n^r p^t), \quad (31)$$

where

$$\text{Bad}^{\text{ind}}(r, t) = \{X \subset V(G_n) : |X| = r, e(X) = t, \text{ and } |N(X) - np^r| \neq o(np^r)\}.$$

Given a graph F with r vertices and t edges, let $\mathcal{E}(F, G_n; \text{Bad}^{\text{ind}}(r, t))$ be the set of induced embeddings f of F in G_n with the image $f(V(F))$ of f in the family $\text{Bad}^{\text{ind}}(r, t)$. Formally,

$$\mathcal{E}(F, G_n; \text{Bad}^{\text{ind}}(r, t)) = \{f \in \mathcal{E}^{\text{ind}}(F, G_n) : f(V(F)) \in \text{Bad}^{\text{ind}}(r, t)\}.$$

Observe that there are at most $\binom{r}{t}$ graphs on r vertices and t edges that can be induced on $X \in \text{Bad}^{\text{ind}}(r, t)$. Hence we deduce from (31) that there is a graph F with r vertices and t edges such that

$$|\mathcal{E}(F, G_n; \text{Bad}^{\text{ind}}(r, t))| \neq o(n^r p^t). \quad (32)$$

Unwinding the definitions, we see that (32) means that the number of induced embeddings f of F in G_n failing to satisfy

$$|N(f(V(F)))| \sim np^r$$

fails to be $o(n^r p^t)$.

Suppose now that the vertices of F are u_1, \dots, u_r . Let u_{r+1} and u_{r+2} be two new vertices. We let F_1 be the graph obtained from F by adding u_{r+1} to F and joining it to all vertices in F . Moreover, we let F_2 be the graph obtained from F_1 by adding u_{r+2} to F_1 and joining it to all vertices in F . Note that u_{r+1} and u_{r+2} are not adjacent in F_2 . Finally, we let F_3 be obtained from F_2 by adding the edge $\{u_{r+1}, u_{r+2}\}$. For convenience, put $F_0 = F$.

In Claim 40 below, we prove that

$$|\mathcal{E}^{\text{ind}}(F_i, G_n)| \not\sim n^{|V(F_i)|} p^{|E(F_i)|},$$

for some i , $0 \leq i \leq 3$. Consequently $\text{NSUB}(k+1)$ fails, which is a contradiction. This contradiction proves Lemma 39. \square

Claim 40. *For some i , $0 \leq i \leq 3$, we have*

$$|\mathcal{E}^{\text{ind}}(F_i, G_n)| \not\sim n^{|V(F_i)|} p^{|E(F_i)|}.$$

Proof. Assume for a contradiction that the number of embeddings of F_i in G_n ($0 \leq i \leq 3$) is $\sim n^{|V(F_i)|} p^{|E(F_i)|}$. We will show that

$$|\mathcal{E}(F, G_n; \text{Bad}^{\text{ind}}(r, t))| = o(n^r p^t),$$

which would contradict (32).

Since $p = o(1)$, we may deduce from Corollary 36 that the number of *induced* embeddings $|\mathcal{E}^{\text{ind}}(F_i, G_n)|$ of F_i in G_n satisfies

$$|\mathcal{E}^{\text{ind}}(F_i, G_n)| \sim n^{|V(F_i)|} p^{|E(F_i)|}, \quad (33)$$

for all $0 \leq i \leq 3$.

For each induced embedding f of F to G_n , put

$$d(f) = |N(f(V(F)))|;$$

that is, $d(f)$ is the number of joint neighbors of the vertices in the image of f . Clearly, we have

$$|\mathcal{E}^{\text{ind}}(F_1, G_n)| = \sum \{d(f) : f \in \mathcal{E}^{\text{ind}}(F, G_n)\}. \quad (34)$$

Moreover, by (33) and the fact that $p = o(1)$, we have that

$$\begin{aligned} |\mathcal{E}^{\text{ind}}(F_2, G_n)| &\sim |\mathcal{E}^{\text{ind}}(F_2, G_n)| + |\mathcal{E}^{\text{ind}}(F_3, G_n)| \\ &= \sum \{d(f)(d(f) - 1) : f \in \mathcal{E}^{\text{ind}}(F, G_n)\}. \end{aligned} \quad (35)$$

Note that (33) and (34) imply that

$$\sum \{d(f) : f \in \mathcal{E}^{\text{ind}}(F, G_n)\} \sim n^{r+1} p^{t+r} \sim |\mathcal{E}^{\text{ind}}(F, G_n)| n p^r. \quad (36)$$

Since $n p^r \geq n p^{k-1} \gg 1$, we may deduce from (36) that

$$\sum_f d(f)^2 \geq \frac{1}{|\mathcal{E}^{\text{ind}}(F, G_n)|} \left(\sum_f d(f) \right)^2 \gg \sum_f d(f), \quad (37)$$

where all the sums above are over $f \in \mathcal{E}^{\text{ind}}(F, G_n)$. Combining (33), (35), and (37), we deduce that

$$\sum \{d(f)^2 : f \in \mathcal{E}^{\text{ind}}(F, G_n)\} \sim n^{r+2} p^{t+2r} \sim |\mathcal{E}^{\text{ind}}(F, G_n)| (n p^r)^2. \quad (38)$$

In view of (36) and (38), we may now simply apply Lemma 26 to deduce that

$$d(f) = |N(f(V(F)))| \sim n p^r$$

for $(1 - o(1)) |\mathcal{E}^{\text{ind}}(F, G_n)| \sim n^r p^t$ embeddings $f \in \mathcal{E}^{\text{ind}}(F, G_n)$. This means that

$$|\mathcal{E}(F, G_n; \text{Bad}^{\text{ind}}(r, t))| = o(n^r p^t).$$

However, as observed above, this contradicts (32). Thus the claim holds. \square

We now prove the implication

$$\text{INDTUP}(k-1) \Rightarrow \text{NSUB}(k),$$

for all appropriate sequences of graphs $\{G_n\}_{n=1}^\infty$. We in fact give a more precise assertion in Lemma 41 below.

Lemma 41. *Let $\delta > 0$, $C > 1$ and $k \geq 3$ be fixed. Let H be a (not necessarily triangle-free) graph on k vertices, and let $p = p(n) = o(1)$ be a function of n satisfying $np^{D_H} \gg 1$. Then there exist $\varepsilon > 0$ and an integer n_3 for which the following holds. If a sequence of graphs $\{G_n\}_{n=1}^\infty$ is such that, for all n ,*

- (i) $G_n \in \text{BDD}(C, D_H)$,
- (ii) $p = p(n) = e(G_n) \binom{n}{2}^{-1}$,
- (iii) $\text{INDTUP}_\varepsilon(d_H)$ holds for G_n ,

then the number of embedding of H in G_n is

$$N(H, G_n) \sim_\delta n^k p^e$$

for all $n \geq n_3$, where $e = |E(H)|$.

Proof. The proof of this lemma is very similar to the proof of Lemma 30, and hence we shall only sketch an informal proof.

We shall assume throughout that $C > 1$ and $k \geq 3$ are fixed constants and that $G_n \in \text{BDD}(C, D_H)$ for all n , where $\{G_n\}_{n=1}^\infty$ is as in the statement of our lemma. Let us also fix a graph H as in the statement of our lemma. As in the proof of Lemma 30, we shall also fix a d_H -degenerate ordering v_1, \dots, v_k of the vertices of H . As before, if $1 \leq h \leq k$, we shall write H_h for the graph $H[\{v_1, \dots, v_h\}]$ induced by $\{v_1, \dots, v_h\}$ in H .

We shall prove by induction on h that

- (†) for all $1 \leq h \leq k$, if property $\text{INDTUP}(d_H)$ holds for $\{G_n\}_{n=1}^\infty$, then

$$|\mathcal{E}(H_h, G_n)| \sim n^h p^{e(H_h)}. \quad (39)$$

Note that (†) is trivially true for $h = 1$. Now suppose that $1 < h \leq k$ and that (†) holds for smaller values of h . We need to show that (39) holds assuming that $\text{INDTUP}(d_H)$ holds.

By Corollary 36, we know that

$$|\mathcal{E}^{\text{ni}}(H_{h-1}, G_n)| = o(n^{h-1} p^{e(H_{h-1})}). \quad (40)$$

From the induction hypothesis and (40), we may deduce that

$$|\mathcal{E}^{\text{ind}}(H_{h-1}, G_n)| \sim n^{h-1} p^{e(H_{h-1})}. \quad (41)$$

We now need to introduce some notation. Let $r = \deg_{H_h}(v_h)$, and suppose that the neighborhood $N_{H_h}(v_h)$ of v_h in H_h induces t edges in H_h . Clearly, $N_{H_h}(v_h)$ induces t edges in H_{h-1} as well. As in the proof of Lemma 39, we put

$$\text{Bad}^{\text{ind}}(r, t) = \{X \subset V(G_n) : |X| = r, e(X) = t, \text{ and } |N(X) - np^r| \neq o(np^r)\}.$$

In words, $\text{Bad}^{\text{ind}}(r, t)$ is the family of the r -element sets of vertices of G_n that induce t edges in G_n and *fail* to have a joint neighborhood of cardinality $\sim np^r$.

Since $r = \deg_{H_h}(v_h) \leq d_H$ and we are assuming that $\text{INDTUP}(d_H)$ holds, we have

$$|\text{Bad}^{\text{ind}}(r, t)| = o(n^r p^t). \quad (42)$$

We now let

$$\mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t)) = \{f \in \mathcal{E}^{\text{ind}}(H_{h-1}, G_n) : f(N_{H_h}(v_h)) \in \text{Bad}^{\text{ind}}(r, t)\}.$$

We will need the following claim, Claim 42. We delay its proof until the end of this section.

Claim 42. *We have*

$$|\mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t))| = o(n^{h-1} p^{e(H_{h-1})}). \quad (43)$$

Assuming Claim 42, we proceed with the proof of Lemma 41.

If f is an embedding of H_{h-1} in G_n , let us write $d(f)$ for the number of extensions of f to embeddings of H_h in G_n . Note that

$$(\dagger) \text{ if } f \in \mathcal{E}^{\text{ind}}(H_{h-1}, G_n) \setminus \mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t)), \quad (44)$$

then

$$d(f) \sim np^r. \quad (45)$$

Observation (\dagger) , relation (41), and Claim 42 imply that

$$|\mathcal{E}(H_h, G_n)| \gtrsim n^{h-1} p^{e(H_{h-1})} \times np^r = n^h p^{e(H_h)}. \quad (46)$$

We now need to estimate $|\mathcal{E}(H_h, G_n)|$ from above. Note that any embedding f of H_{h-1} in G_n extends to $\leq Cnp^r$ embeddings of H_h in G_n , because we are assuming that $G_n \in \text{BDD}(C, D_H)$ and $r = \deg_{H_h}(v_h) \leq d_H \leq D_H$. In particular, if

$$f \in \mathcal{E}^{\text{ni}}(H_{h-1}, G_n) \cup \mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t)), \quad (47)$$

then $d(f) \leq Cnp^r$. Inequality (40) and Claim 42 imply that the number of embeddings f as in (47) is $o(n^{h-1} p^{e(H_{h-1})})$. It follows that the number of embeddings of H_h in G_n that extend embeddings f as in (47) is $o(n^h p^{e(H_h)})$.

Finally, we observe that if an embedding $f \in \mathcal{E}(H_{h-1}, G_n)$ is not as in (47), then it must be as in (44). Recalling (\dagger) , we see that the total number of embeddings of H_h in G_n is $\sim n^h p^{e(H_h)}$. The proof of the induction step is therefore complete, and hence (\dagger) follows by induction. Naturally, Lemma 41 follows by setting $h = k$ in (\dagger) . \square

Now we present the proof of Claim 42, which is a slight extension of the proof of Corollary 38.

Proof of Claim 42. By definition

$$\begin{aligned} \mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t)) \\ = \{f \in \mathcal{E}^{\text{ind}}(H_{h-1}, G_n) : f(N_{H_h}(v_h)) \in \text{Bad}^{\text{ind}}(r, t)\}. \end{aligned}$$

Fix an r -tuple F such that $F^{\text{set}} = N_{H_h}(v_h)$. By the above definition and the fact that G_n satisfies $\text{INDTUP}(k-1)$, we have

$$|\mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t))| = \sum_X |\mathcal{E}(H_{h-1}, G_n, F, X)|, \quad (48)$$

where the sum is over all r -tuples X such that $X^{\text{set}} \in \text{Bad}^{\text{ind}}(r, t)$. By (42), the number of r -tuples X that we are summing over in (48) is at most

$$r! \times o(n^r p^t) = o(n^r p^t).$$

For each r -tuple X , we apply the Extension Lemma (Lemma 34) to

$$\mathcal{E}(H_{h-1}, G_n, F, X)$$

and deduce from (48) that

$$\begin{aligned} |\mathcal{E}(H_{h-1}, G_n; \text{Bad}^{\text{ind}}(r, t))| &\leq o(n^r p^t) \times C^{(h-1)-r} n^{(h-1)-r} p^{e(H_{h-1})-t} \\ &= o(n^{h-1} p^{e(H_{h-1})}). \end{aligned} \quad (49)$$

This concludes the proof of Claim 42. \square

3.4 Remarks about Theorem 25 (the directed case)

We omit the proof of Theorem 25 (stated in Section 2.2) and make a few remarks about its connection to the undirected case.

Theorem 25 is the directed version of the Embedding Lemma (Theorem 16). Its proof goes along the lines of the the proof of the Embedding Lemma. That is, it uses the directed versions of the Extension Lemma (Lemma 34), Lemma 29, Lemma 30 and Corollary 36. However the proof of the directed versions of the Extension Lemma, Lemma 30 and Corollary 36 are very similar to the proofs for the undirected case. Thus we omit those proofs.

We shall rather state and prove Lemma 43, which is the directed analogue of Lemma 29. Then at the end of this section we briefly say how to deduce Lemma 29 from Lemma 43.

We follow the same notation as in the beginning of Section 2.2. Let $\vec{G} = (V, \vec{E})$ be a digraph, $\pi = (\pi_1, \dots, \pi_r) \in \{+, -\}^r$, and $(u_1, \dots, u_r) \in V^r$. We let $d^\pi(u_1, \dots, u_r) = |N^\pi(u_1, \dots, u_r)|$, where

$$\begin{aligned} N^\pi(u_1, \dots, u_r) = \{w \in V : \forall i \in [r], (u_i, w) \in \vec{E} \text{ if } \pi_i = + \\ \text{and } (w, u_i) \in \vec{E} \text{ if } \pi_i = -\}. \end{aligned} \quad (50)$$

Lemma 43. *Let $t \geq 2$ be an integer and let $\vec{G} = (V, \vec{E})$ be a digraph on n vertices satisfying the following conditions:*

(a) *for all but $o(n)$ vertices $u \in V$,*

$$d^+(u) = \frac{1}{2}pn(1 + o(1)),$$

(b)

$$\sum_{(u,v) \in V^2} (d^{++}(u,v))^2 = n^2 \left(\frac{p^2 n}{4} \right)^2 (1 + o(1)).$$

If $p = p(n) \gg n^{-1/t}$ and $\vec{G} \in \overline{\text{BDD}}(C, 2)$, then, for all $2 \leq r \leq t$, for all $\pi \in \{+, -\}^r$, and for all but $o(n^r)$ r -tuples $(u_1, \dots, u_r) \in V^r$, we have

$$d^\pi(u_1, \dots, u_r) = \frac{1}{2^r} p^r n (1 + o(1)).$$

Proof. First we show that

(c) *for all but $o(n)$ vertices $u \in V$, we have*

$$d^-(u) = \frac{1}{2}pn(1 + o(1)).$$

To that end, we first observe that

$$\sum_{a \in V} d^-(a) = \sum_{u \in V} d^+(u). \quad (51)$$

Condition (a) above and the fact that all vertices have degree $\leq Cpn$ imply that

$$\sum_{u \in V} d^+(u) = n \left(\frac{pn}{2} \right) (1 + o(1)) + o(n)Cpn = n \left(\frac{pn}{2} \right) (1 + o(1)). \quad (52)$$

Moreover, by the Cauchy-Schwarz inequality and (b) above, we have

$$\begin{aligned} \sum_{a \in V} d^-(a)^2 &= \sum_{(u,v) \in V^2} d^{++}(u,v) \leq n \left\{ \sum_{(u,v) \in V^2} d^{++}(u,v)^2 \right\}^{1/2} \\ &\leq n \left(\frac{pn}{2} \right)^2 (1 + o(1)). \end{aligned} \quad (53)$$

Lemma 26 and (51), (52), and (53) now imply that (c) above does indeed hold.

We may deduce from (c) that

$$\sum_{(u,v) \in V^2} d^{++}(u,v) = \sum_{a \in V} d^-(a)^2 \geq n^2 \left(\frac{p^2 n}{4} \right) (1 + o(1)). \quad (54)$$

Lemma 26, condition (b) and (54) now imply that

(d) for all but $o(n^2)$ pairs $(u, v) \in V^2$, we have

$$d^{++}(u, v) = \frac{1}{4}p^2n(1 + o(1)).$$

Similarly, we may deduce that

(e) for all but $o(n^2)$ pairs $(u, v) \in V^2$, we have

$$d^{--}(u, v) = \frac{1}{4}p^2n(1 + o(1)).$$

Indeed, this is a consequence of Lemma 26 and the identities

$$\sum_{(u,v) \in V^2} d^{--}(u, v) = \sum_{a \in V} d^+(a)^2$$

and

$$\sum_{(u,v) \in V^2} d^{--}(u, v)^2 = \sum_{(a,b) \in V^2} d^{++}(a, b)^2.$$

Having established the auxiliary facts (c)–(e), we are now in position to verify Lemma 43. For $\pi = (\pi_1, \dots, \pi_r) \in \{+, -\}^r$, let $P(\pi) = |\{i: \pi_i = +\}|$ and $Q(\pi) = r - P(\pi)$. We write $\mathbf{u} = (u_1, \dots, u_r)$ for a general element in V^r . We have

$$\sum_{\mathbf{u} \in V^r} d^\pi(\mathbf{u}) = \sum_{a \in V} d^-(a)^{P(\pi)} d^+(a)^{Q(\pi)}. \quad (55)$$

Condition (a) and property (c) deduced above and the fact that all vertices have degree $\leq Cpn$ allows us to conclude that the right-hand side of (55) is $\sim n(pn/2)^r = n^r(p^r n/2^r)$, so that

$$\sum_{\mathbf{u} \in V^r} d^\pi(\mathbf{u}) = n^r \left(\frac{1}{2^r} p^r n \right) (1 + o(1)). \quad (56)$$

We now observe that

$$\sum_{\mathbf{u} \in V^r} d^\pi(\mathbf{u})^2 = \sum_{(a,b) \in V^2} (d^{--}(a, b))^{P(\pi)} (d^{++}(a, b))^{Q(\pi)}. \quad (57)$$

Properties (d) and (e) deduced above and the fact that all pairs of vertices have joint degree $\leq Cp^2n$ allows us to conclude that the right-hand side of (57) is $\sim n^2(p^2n/4)^r = n^r(p^r n/2^r)^2$, so that

$$\sum_{\mathbf{u} \in V^r} d^\pi(\mathbf{u})^2 = n^r \left(\frac{1}{2^r} p^r n \right)^2 (1 + o(1)). \quad (58)$$

Finally, Lemma 26 and (56) and (58) imply that for all but $o(n^r)$ r -tuples $\mathbf{u} \in V^r$, we have

$$d^\pi(\mathbf{u}) = \frac{1}{2^r} p^r n (1 + o(1)),$$

which concludes the proof of Lemma 43. \square

Now we present a sketch of the proof of Lemma 29 (introduced in Section 3), based on Lemma 43. We start by restating Lemma 29 in the following equivalent form.

Lemma 44. *Suppose $t \geq 2$ and $C > 1$ are constants and $p = p(n)$ satisfies $np^t \gg 1$. Let $\{G_n\}_{n=1}^\infty$ be a sequence of graphs with $G_n \in \text{BDD}(C, 2)$ for all n and $|E(G_n)| = (1 + o(1))p\binom{n}{2}$. If*

(a) *for all but $o(n)$ vertices $u \in V(G_n)$,*

$$\deg_{G_n}(u) = pn(1 + o(1)),$$

(b) *for all but at most $o(n^2)$ pairs $\{x_1, x_2\} \subset V(G_n)$,*

$$|N_{G_n}(x_1) \cap N_{G_n}(x_2)| = p^2n(1 + o(1)),$$

then, for all $r \in [t]$, all but at most $o(n^r)$ r -element sets $\{x_1, \dots, x_r\} \subset V(G_n)$ are such that

$$|N_{G_n}(x_1) \cap \dots \cap N_{G_n}(x_r)| = p^r n(1 + o(1)).$$

Sketch of the proof of Lemma 44. Lemma 44 follows from Lemma 43. Suppose we are given a graph G_n as above; we then randomly orient its edges to get \vec{G}_n . One can easily show that the hypothesis of Lemma 43 holds almost surely for \vec{G}_n . Finally, note that if \vec{G}_n satisfies the conclusion of Lemma 43, then G_n satisfies the conclusion of Lemma 44. \square

4 Auxiliary facts and related work

4.1 General facts

We have used Facts 45 and 46 given below. Recall that, for two graphs X and Y , the set of all embeddings of X in Y is denoted by $\mathcal{E}(X, Y)$.

Fact 45. *Let $k \geq 1$ be a fixed integer. For any sequence of graphs $\{G_n\}_{n=1}^\infty$, we have*

$$\text{NSUB}(k+1) \Rightarrow \text{NSUB}(k).$$

Proof. Suppose $\{G_n\}_{n=1}^\infty$ satisfies $\text{NSUB}(k+1)$ and let $p = p(n) = |E(G_n)|\binom{n}{2}^{-1}$. To prove this fact, we have to show that, for any graph H on k vertices, we have

$$|\mathcal{E}(H, G_n)| = (1 + o(1))n^k p^e, \tag{59}$$

where $e = |E(H)|$. Given a graph H as above we construct H^+ where $V(H^+) = V(H) \cup \{u\}$ and $E(H^+) = E(H)$. By definition of H^+ , it follows that

$$|\mathcal{E}(H^+, G_n)| = N(H, G_n)(n - k). \tag{60}$$

By hypothesis, we know that $\{G_n\}_{n=1}^\infty$ satisfies $\text{NSUB}(k+1)$. Thus

$$|\mathcal{E}(H^+, G_n)| = (1 + o(1))n^{k+1} p^e. \tag{61}$$

Combining (60) and (61), we obtain (59). \square

Similarly, we may prove the following simple fact.

Fact 46. *Let $k \geq 1$ be a fixed integer. For any sequence of graphs $\{G_n\}_{n=1}^\infty$, we have*

$$\text{TFSUB}(k+1) \Rightarrow \text{TFSUB}(k).$$

4.2 Proof of Proposition 6

In this section, we shall sketch the proof of Proposition 6(A) and we shall prove Proposition 6(B) using a construction due to Alon [2].

Proof of Proposition 6(A). We only outline the proof of Proposition 6(A), because a similar result is proved in [13] (see Theorem B' in [13]). The graphs G_i satisfying properties (i) and (ii) above can be constructed from sparse random graphs whose triangles have been destroyed by the removal of a small fraction of the edges. Replacing each vertex of such a triangle-free “random like graph” by a stable set of appropriate cardinality and each edge by a complete bipartite graph yields suitable graphs G_i . \square

The proof of Proposition 6(B) is based on a family of graphs constructed by Alon [2].

4.2.1 Construction of Alon’s graph

Let $k > 1$ be an integer not divisible by 3 and let $F_k = \text{GF}(2^k)$ be the Galois field with 2^k elements. Depending on the context, we will think of the elements of F_k as polynomials over $\text{GF}(2)$ or as binary vectors of length k (whose entries are the coefficients of the corresponding polynomial representations). If u and v are two vectors, we will denote their concatenation by $u \circ v$.

For any $\alpha \in F_k - \{0\}$, we put the vector α in W_0 if the constant term of the polynomial α^7 is 0. Otherwise we put α in W_1 . Let $\Gamma = (\mathbb{Z}_2)^{3k}$ be the Abelian group with elements the binary vectors of length $3k$. Let $U_0 = \{w_0 \circ w_0^3 \circ w_0^5 : w_0 \in W_0\}$ and $U_1 = \{w_1 \circ w_1^3 \circ w_1^5 : w_1 \in W_1\}$. Note that

$$|U_0| = |W_0| = 2^{k-1} - 1 \quad \text{and} \quad |U_1| = |W_1| = 2^{k-1}.$$

In the following, we will use bold letters to denote vectors of length $3k$ to distinguish them from vectors of length k .

The Alon graph $G = G(\Gamma)$ is defined to be the Cayley graph on Γ with generating set

$$S = U_0 + U_1 = \{\mathbf{u}_0 + \mathbf{u}_1 : \mathbf{u}_0 \in U_0, \mathbf{u}_1 \in U_1\} \subset \Gamma.$$

In other words $V(G) = \Gamma = (\mathbb{Z}_2)^{3k}$ and $\mathbf{x}, \mathbf{y} \in V(G)$ form an edge in G if and only if $\mathbf{x} + \mathbf{y} \in S$. Let M_0 (resp. M_1) be the $3k \times (2^{k-1} - 1)$ (resp. $3k \times 2^{k-1}$) matrix whose columns are the vectors in U_0 (resp. U_1). Consider the matrix $M = [M_0, M_1]$. It turns out that M is the parity check matrix of a BCH code of designed distance 7. Alon showed that the graph G has the following properties:

- (a) G is triangle-free,
- (b) G is $d = |S| = 2^{k-1}(2^{k-1} - 1)$ -regular,
- (c) The second largest eigenvalue of the adjacency matrix of G has size $\Theta(2^k)$.

Roughly speaking, properties (a) and (b) follow from the fact any 6 columns in M are linearly independent over $\text{GF}(2)$. Property (c) is much more delicate, and depends on the Carlitz–Uchiyama bound for the Hamming weight of dual code words of BCH codes. We refer the reader to Alon [2] for details.

Proof of Proposition 6(B). From the above discussion, it follows that for each $k > 1$ not divisible by 3, we have an Alon graph G with the properties (a), (b) and (c) listed above. Let $\{G_i\}_{i=1}^\infty$ be the family of all such graphs G (ordered according to $|V(G_i)|$).

We prove Proposition 6(B) by showing that the family $\{G_i\}_{i=1}^\infty$ satisfies (i), (ii), and (iii) of Proposition 6(B).

Observe that (i) of Proposition 6(B) is simply (a) above. Next we prove (ii) of Proposition 6(B). For each G_i , we have (by definition) $n = |V(G_i)| = 2^{3k}$ and $d = pn = 2^{k-1}(2^{k-1} - 1)$. Thus, letting $k \rightarrow \infty$ yields

$$p = \left(\frac{1}{4} + o(1)\right) n^{-1/3}. \quad (62)$$

Let $A = (a_{x,y})_{x,y \in V(G_i)}$ denote the 0–1 adjacency matrix of the graph G_i , with 1 denoting edges. Let λ_j ($1 \leq j \leq n = 2^{3k}$) be the eigenvalues of A and adjust the notation so that $\lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. It follows from properties (b) and (c) above that

$$\lambda_1 = d = 2^{k-1}(2^{k-1} - 1) \quad \text{and} \quad |\lambda_2| = \Theta(2^k). \quad (63)$$

Hence $\{G_i\}_{i=1}^\infty$ satisfies EIG. By Fact 3 in [6], we have $\text{EIG} \Rightarrow \text{DISC}$. Consequently $\{G_i\}_{i=1}^\infty$ satisfies DISC as well. Now it remains to show that $\{G_i\}_{i=1}^\infty$ satisfies property TUPLE(2). Assume for a moment that $\{G_i\}_{i=1}^\infty$ satisfies property EIG(4), defined as follows:

$$\text{EIG(4): } \sum_{i=1}^n |\lambda_i|^4 = (1 + o(1))p^4 n^4.$$

By Fact 7 and Theorem 3 in [6], we have $\text{EIG(4)} \Rightarrow \text{CYCLE(4)}$. By Fact 28 in Section 3.2, we have $\text{CYCLE(4)} \Rightarrow \text{TUPLE(2)}$. Therefore, in order to show that $\{G_i\}_{i=1}^\infty$ satisfies TUPLE(2), it is enough to show that $\{G_i\}_{i=1}^\infty$ satisfies EIG(4).

By (63), there exists a constant C such that

$$\lambda_1^4 = d^4 = p^4 n^4 \quad (64)$$

and

$$\sum_{i=2}^n |\lambda_i|^4 < n(C2^k)^4 = C^4 2^{3k} \cdot 2^{4k} = o(p^4 n^4), \quad (65)$$

where in (65) we used that $p^4 n^4 = 2^{8k}$. By (64) and (65), our graph sequence $\{G_i\}_{i=1}^\infty$ satisfies property EIG(4). This concludes the proof of (ii) of Proposition 6(B).

To complete our proof of Proposition 6(B), it remains to show that the graphs G_i satisfy (iii). Since this will take some work, we state this fact as a separate lemma (see Lemma 47 below). \square

Lemma 47. $G_i \in \text{BDD}(128, 2)$ for all $i \geq 1$.

We will use the following simple fact in the proof of Lemma 47.

Fact 48. Suppose a_1 and $a_2 \in F_k$ with $a_1 \neq 0$ are given, and consider the system of equations

$$\begin{cases} x + y &= a_1 \\ x^3 + y^3 &= a_2. \end{cases} \quad (66)$$

System (66) has at most two pairs of solutions in F_k , namely $(x, y) = (\alpha, \beta)$ and $(x, y) = (\beta, \alpha)$ for some α and $\beta \in F_k$ with $\beta = \alpha + a_1 \neq \alpha$.

Proof. By substituting $x + a_1$ for y in the second equation of (66), we obtain the quadratic equation $a_1 x^2 + a_1^2 x + a_1^3 + a_2 = 0$, which has at most two solutions. If α is a solution to the latter equation, then so is $\beta = \alpha + a_1$, as a simple calculation shows. This implies that the solutions to (66) are as claimed. \square

Proof of Lemma 47. Since (by definition) G_i is d -regular, where $d = pn$, we have $\deg_{G_i}(\mathbf{x}) = pn$ for all $\mathbf{x} \in V(G_i)$. Thus, it remains to show that for any two vertices $\mathbf{x} \neq \mathbf{y}$ in $V(G_i)$, we have

$$|N_{G_i}(\mathbf{x}) \cap N_{G_i}(\mathbf{y})| \leq 128p^2n.$$

For $\mathbf{x} \neq \mathbf{y} \in V(G_i)$, the vertex $\mathbf{t} \in V(G_i)$ belongs to $N(\mathbf{x}) \cap N(\mathbf{y})$ if and only if there exist $\mathbf{s}, \mathbf{s}' \in S$ such that $\mathbf{x} + \mathbf{t} = \mathbf{s}$ and $\mathbf{y} + \mathbf{t} = \mathbf{s}'$, or, equivalently, $\mathbf{x} + \mathbf{y} = \mathbf{s} + \mathbf{s}'$. Consequently,

$$|N(\mathbf{x}) \cap N(\mathbf{y})| = |\{(\mathbf{s}, \mathbf{s}') \in S \times S : \mathbf{s} + \mathbf{s}' = \mathbf{x} + \mathbf{y}\}|.$$

Set $\mathbf{a} = \mathbf{x} + \mathbf{y} = a_1 \circ a_2 \circ a_3$ where a_1, a_2 , and a_3 are in F_k . Let $\mathbf{s} = (w_0 + w_1) \circ (w_0^3 + w_1^3) \circ (w_0^5 + w_1^5)$ and $\mathbf{s}' = (v_0 + v_1) \circ (v_0^3 + v_1^3) \circ (v_0^5 + v_1^5)$ where $v_0, w_0 \in W_0$ and $v_1, w_1 \in W_1$. Thus the equation $\mathbf{x} + \mathbf{y} = \mathbf{s} + \mathbf{s}'$ can be written as

$$\begin{cases} w_0 + w_1 + v_0 + v_1 &= a_1 \\ w_0^3 + w_1^3 + v_0^3 + v_1^3 &= a_2 \\ w_0^5 + w_1^5 + v_0^5 + v_1^5 &= a_3. \end{cases} \quad (67)$$

For any $f \in F_k$, let

$$\tilde{f} = f \circ f^3 \circ f^5 = \begin{pmatrix} f \\ f^3 \\ f^5 \end{pmatrix}.$$

We define

$$P = \{(\widetilde{w}_0 + \widetilde{w}_1, \widetilde{v}_0 + \widetilde{v}_1) : w_0, v_0 \in W_0 \text{ and } w_1, v_1 \in W_1 \text{ satisfy (67)}\}. \quad (68)$$

Observe that

$$|N(\mathbf{x}) \cap N(\mathbf{y})| = |\{(s, s') \in S \times S : s + s' = \mathbf{x} + \mathbf{y}\}| = |P|.$$

For each $z_0 \in W_0$, set

$$P(z_0) = \{(\widetilde{w}_0 + \widetilde{w}_1, \widetilde{v}_0 + \widetilde{v}_1) \in P : w_0 = z_0\}.$$

Since any 6 columns of M are linearly independent, a moment's thought shows that the sets $P(z_0)$ ($z_0 \in W_0$) are pairwise disjoint. Therefore

$$|N(\mathbf{x}) \cap N(\mathbf{y})| = |P| = \sum_{z_0 \in W_0} |P(z_0)|. \quad (69)$$

Let

$$T = \{z_0 \in W_0 : |P(z_0)| > 2\}. \quad (70)$$

We now state a claim that will be used to finish the proof of Lemma 47.

Claim 49. *With the same notation as above, the following holds.*

- (1) $|P(z_0)| \leq |W_1| = 2^{k-1}$ for all $z_0 \in W_0$.
- (2) $|T| \leq 2$.

The proof of Claim 49 is postponed to the end of Section 4.2. Now we are ready to finish the proof of Lemma 47. By (69) and Claim 49, we have

$$\begin{aligned} |P| &= \sum_{z_0 \in W_0} |P(z_0)| = \sum_{z_0 \in T} |P(z_0)| + \sum_{z_0 \in W_0 \setminus T} |P(z_0)| \\ &\leq |T| \cdot 2^{k-1} + |W_0| \cdot 2 \leq 2 \cdot 2^{k-1} + (2^{k-1} - 1) \cdot 2 < 4 \cdot 2^{k-1}. \end{aligned}$$

Now, it follows from (69) that

$$|N(\mathbf{x}) \cap N(\mathbf{y})| = |P| < 4 \cdot 2^{k-1} \leq 128p^2n,$$

because, as a quick calculation shows, $p^2n = 2^{k-4}(1 - 1/2^{k-1})^2 \geq 2^{k-6}$. This concludes the proof of Lemma 47, assuming Claim 49. \square

It remains to prove Claim 49.

Proof of Claim 49. We shall first prove (1) of Claim 49. Let $z_0 \in W_0$. If $|P(z_0)| \leq 1$ then we are done. Otherwise there exist at least two pairs $(\widetilde{z}_0 + \widetilde{w}'_1, \widetilde{v}'_0 + \widetilde{v}'_1)$, $(\widetilde{z}_0 + \widetilde{w}''_1, \widetilde{v}''_0 + \widetilde{v}''_1) \in P(z_0)$. From the definition of $P(z_0) \subset P$ and (67), it follows that

$$\widetilde{w}'_1 + \widetilde{v}'_0 + \widetilde{v}'_1 = \widetilde{w}''_1 + \widetilde{v}''_0 + \widetilde{v}''_1. \quad (71)$$

Since any 6 columns of the matrix M are linearly independent (see Alon [2]), each element in $\{\widetilde{w}'_1, \widetilde{w}''_1, \widetilde{v}'_0, \widetilde{v}''_0, \widetilde{v}'_1, \widetilde{v}''_1\}$ must occur an even number of times. Since $W_0 \cap W_1 = \emptyset$, $\{v'_0, v''_0\} \subset W_0$ and $\{v'_1, v''_1, w'_1, w''_1\} \subset W_1$, we have $\widetilde{v}'_0 = \widetilde{v}''_0$. In other words if z_0 is fixed then $\widetilde{v}_0 = v_0(z_0)$ is uniquely determined for all pairs $(\widetilde{z}_0 + \widetilde{w}_1, \widetilde{v}_0 + \widetilde{v}_1) \in P(z_0)$. By the definition of $P(z_0) \subset P$, we have

$$\widetilde{z}_0 + \widetilde{w}_1 + \widetilde{v_0(z_0)} + \widetilde{v}_1 = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}. \quad (72)$$

We distinguish the following two cases.

Case 1: $\widetilde{z}_0 + \widetilde{v_0(z_0)} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$

In this case, equation (72) implies that $w_1 = v_1$. Hence the elements of $P(z_0)$ are of the form $(\widetilde{z}_0 + \widetilde{w}_1, \widetilde{v_0(z_0)} + \widetilde{w}_1)$, where $w_1 \in W_1$ is arbitrary. Thus $|P(z_0)| \leq |W_1| = 2^{k-1}$. Hence (1) of Claim 49 holds in this case.

Case 2: $\widetilde{z}_0 + \widetilde{v_0(z_0)} \neq \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$

Set

$$\begin{pmatrix} a'_1 \\ a'_2 \\ a'_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \widetilde{z}_0 + \widetilde{v_0(z_0)} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (73)$$

Then (72) implies that $w_1 + v_1 = a'_1$, $w_1^3 + v_1^3 = a'_2$, and $w_1^5 + v_1^5 = a'_3$. Observe that these equations and (73) imply that $a'_1 \neq 0$ (indeed, otherwise $w_1 = v_1$, and we would have $a'_1 = a'_2 = a'_3 = 0$, which contradicts (73)). Now Fact 48 implies that the equations $w_1 + v_1 = a'_1$ and $w_1^3 + v_1^3 = a'_2$ are satisfied by at most two pairs (w_1, v_1) . Hence $|P(z_0)| \leq 2 \leq 2^{k-1}$, and (1) of Claim 49 is proven.

Now we shall prove (2) of Claim 49. If $z_0 \in T$ then $|P(z_0)| > 2$ and it follows from the above discussion that for any $(\widetilde{z}_0 + \widetilde{w}_1, \widetilde{v}_0 + \widetilde{v}_1) \in P(z_0)$, we have $\widetilde{v}_0 = \widetilde{v_0(z_0)}$ and $\widetilde{w}_1 = \widetilde{v}_1$. This observation combined with (72) implies that $z_0 + v_0(z_0) = a_1 \neq 0$ and $z_0^3 + (v_0(z_0))^3 = a_2$. Then Fact 48 implies that these two equations have at most two solution pairs $(z_0, v_0(z_0))$. Thus $|T| \leq 2$, proving (2) of Claim 49. \square

4.3 Connection to random graphs

4.3.1 Theorem 1 for subgraphs of random graphs

Proposition 6 tells us that the implications “DISC \Rightarrow NSUB(3)”, “TUPLE(2) \Rightarrow NSUB(3)” and “EIG \Rightarrow NSUB(3)” fail to be true when $p = o(1)$. However, counterexamples demonstrating this proposition are rare and “do not occur in random graphs”. Before stating a theorem turning this statement precise, we first introduce some definitions.

Let $\mathcal{G}(m, q)$ be the random graph on m vertices and edges chosen independently, each with probability q . Thus for a graph G on m vertices and e edges,

$$\text{Prob}(\mathcal{G}(m, q) = G) = q^e(1 - q)^{\binom{m}{2} - e}. \quad (74)$$

Sometimes, we will view $\mathcal{G}(m, q)$ as the set of all $2^{\binom{m}{2}}$ labeled graphs with m vertices and with the probability distribution given in (74).

For $G \in \mathcal{G}(m, q)$ and $c \in (0, 1)$, let $R_G(c)$ be the family of all subgraphs F of G (not necessarily induced and not necessarily spanning) such that

$$|E(F)| \geq cq \binom{m}{2}.$$

Fix integers $m \geq k \geq 4$ and let $\delta, c, \varepsilon > 0$ be given. Let $\mathcal{F} = \mathcal{F}(k, \delta, c, \varepsilon, m)$ be the family of m -vertex graphs G defined as follows:

$$G \in \mathcal{F} \quad \text{iff} \quad \begin{array}{l} \text{for any } F \in R_G(c) \text{ and any pair of properties} \\ P, Q \in \{\text{NSUB}(k), \text{DISC}, \text{EIG}, \text{TUPLE}(2)\}, \\ \text{if } F \text{ satisfies } P_\varepsilon, \text{ then } F \text{ satisfies } Q_\delta. \end{array}$$

Theorem 1 extends for “subgraphs of random graphs” if $p \rightarrow 0$ sufficiently slowly. More precisely, we have the following theorem (see [13] and [14]).

Theorem 50. *Let $k \geq 4$ and $\delta, c \in (0, 1)$ be given. Suppose that*

$$q \geq m^{-1/(k-1)}(\log m)^{100}.$$

Then there exists $\varepsilon > 0$ such that, for $G \in \mathcal{G}(m, q)$, we have

$$\text{Prob}(G \in \mathcal{F}(k, \delta, c, \varepsilon, m)) \rightarrow 1 \text{ as } m \rightarrow \infty.$$

4.3.2 Turán’s problem for subgraphs of random graphs

Theorem 16 implies the following extremal result for subgraphs of random graphs. As usual, we write $\chi(H)$ for the chromatic number of a graph H .

Theorem 51. *Let H be a triangle-free graph, and consider the random graph $\mathcal{G}(m, q)$, where $q \gg m^{-1/D_H}$. Suppose $\varepsilon > 0$ is fixed. Then, asymptotically almost surely, $G \in \mathcal{G}(m, q)$ satisfies the following property. If $F \subset G$ is any subgraph of G with*

$$|E(F)| \geq \left(1 - \frac{1}{\chi(H) - 1} + \varepsilon\right) q \binom{m}{2},$$

then F contains a copy of H .

For discussions on Turán type extremal problems for subgraphs of random graphs, see [11, Chapter 8], [12], [13, Section 1.4.2], and [14].

5 Concluding remarks

There are several questions that remain unsettled. We recall a few of them.

- (1) Although somewhat technical, the main open question is probably Conjecture 20.
- (2) As mentioned earlier, it would be interesting to know if one can extend Proposition 6 to the existence of graphs G_i with $|V(G_i)| = n_i$ and $p(n_i) = |E(G_i)|/\binom{n_i}{2} \gg n_i^{-1/3}$ such that G_i satisfies (i) and (ii) in Proposition 6 and $G_i \in \text{BDD}(C, 2)$ for some constant C for all i .
- (3) It would be interesting to extend Theorem 23 to Conjecture 24.

References

- [1] N. Alon, *Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory*, *Combinatorica* **6** (1986), no. 3, 207–219.
- [2] ———, *Explicit Ramsey graphs and orthonormal labelings*, *Electron. J. Combin.* **1** (1994), Research Paper 12, approx. 8 pp. (electronic).
- [3] N. Alon and F. R. K. Chung, *Explicit construction of linear sized tolerant networks*, *Discrete Mathematics* **72** (1988), no. 1-3, 15–19.
- [4] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma*, *J. Algorithms* **16** (1994), no. 1, 80–109.
- [5] Noga Alon and Joel H. Spencer, *The probabilistic method*, second ed., Wiley-Interscience [John Wiley & Sons], New York, 2000, With an appendix on the life and work of Paul Erdős.
- [6] F. R. K. Chung and R. L. Graham, *Sparse quasi-random graphs*, *Combinatorica*, to appear.
- [7] F. R. K. Chung and R. L. Graham, *Quasi-random tournaments*, *J. Graph Theory* **15** (1991), no. 2, 173–198.
- [8] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, *Combinatorica* **9** (1989), no. 4, 345–362.
- [9] Richard A. Duke, Hanno Lefmann, and Vojtěch Rödl, *A fast approximation algorithm for computing the frequencies of subgraphs in a given graph*, *SIAM J. Comput.* **24** (1995), no. 3, 598–620.
- [10] P. Frankl, V. Rödl, and R. M. Wilson, *The number of submatrices of a given type in a Hadamard matrix and related results*, *J. Combin. Theory Ser. B* **44** (1988), no. 3, 317–328.

- [11] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000.
- [12] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Foundations of Computational Mathematics (Berlin, Heidelberg) (F. Cucker and M. Shub, eds.), Springer-Verlag, January 1997, pp. 216–230.
- [13] Y. Kohayakawa and V. Rödl, *Regular pairs in sparse random graphs I*, submitted, 2001.
- [14] Y. Kohayakawa, V. Rödl, and M. Schacht, *The Turán theorem for random graphs*, submitted, 2002.
- [15] Y. Kohayakawa, V. Rödl, and L. Thoma, *An optimal algorithm for checking regularity (extended abstract)*, Proceedings of the 13th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2002), ACM/SIAM, 2002, pp. 277–286.
- [16] T. Łuczak, V. Rödl, and P. Sissokho, unpublished manuscript, 2000.
- [17] Vojtěch Rödl, *On universality of graphs with uniformly distributed edges*, Discrete Math. **59** (1986), no. 1-2, 125–134.
- [18] Vojtěch Rödl and Andrzej Ruciński, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc. **8** (1995), no. 4, 917–942.
- [19] Miklós Simonovits and Vera T. Sós, *Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs*, Combinatorica **17** (1997), no. 4, 577–596.
- [20] Andrew G. Thomason, *Pseudorandom graphs*, Random graphs '85 (Poznań, 1985), North-Holland, Amsterdam, 1987, pp. 307–331.
- [21] ———, *Random graphs, strongly regular graphs and pseudorandom graphs*, Surveys in Combinatorics 1987 (C. Whitehead, ed.), London Mathematical Society Lecture Note Series, vol. 123, Cambridge University Press, Cambridge–New York, 1987, pp. 173–195.