

k-wise independent random graphs

Noga Alon*

Asaf Nussboim[†]

Abstract

We study the k -wise independent relaxation of the usual model $\mathcal{G}(N, p)$ of random graphs where, as in this model, N labeled vertices are fixed and each edge is drawn with probability p , however, it is only required that the distribution of any subset of k edges is independent. This relaxation can be relevant in modeling phenomena where only k -wise independence is assumed to hold, and is also useful when the relevant graphs are so huge that handling $\mathcal{G}(N, p)$ graphs becomes infeasible, and cheaper random-looking distributions (such as k -wise independent ones) must be used instead. Unfortunately, many well-known properties of random graphs in $\mathcal{G}(N, p)$ are global, and it is thus not clear if they are guaranteed to hold in the k -wise independent case. We explore the properties of k -wise independent graphs by providing upper-bounds and lower-bounds on the amount of independence, k , required for maintaining the main properties of $\mathcal{G}(N, p)$ graphs: connectivity, Hamiltonicity, the connectivity-number, clique-number and chromatic-number and the appearance of fixed subgraphs. Most of these properties are shown to be captured by either constant k or by some $k = \text{poly}(\log(N))$ for a wide range of values of p , implying that random looking graphs on N vertices can be generated by a seed of size $\text{poly}(\log(N))$. The proofs combine combinatorial, probabilistic and spectral techniques.

*Schools of Mathematics and Computer Science, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel, and IAS, Princeton, NJ 08540, USA. Email: nogaa@post.tau.ac.il. Research supported in part by the Israel Science Foundation and by a USA-Israeli BSF grant.

[†]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. Email: asaf.nussbaum@weizmann.ac.il. Partly supported by a grant from the Israel Science Foundation.

1 Introduction

We study the k -wise independent relaxation of the usual model $\mathcal{G}(N, p)$ of random graphs where, as in this model, N labeled vertices are fixed and each edge is drawn with probability (w.p., for short) $p = p(N)$, however, it is only required that the distribution of any subset of k edges is independent (in $\mathcal{G}(N, p)$ all edges are mutually independent). These k -wise independent graphs are natural combinatorial objects that may prove to be useful in modeling scientific phenomena where only k -wise independence is assumed to hold. Moreover, they can be used when the relevant graphs are so huge that handling $\mathcal{G}(N, p)$ graphs is infeasible, and cheaper random-looking distributions must be used instead. However, what happens when the application that uses these graphs (or the analysis conducted on them) critically relies on the fact that random graphs are, say, almost surely connected? After all, k -wise independence is defined via ‘local’ conditions, so isn’t it possible that k -wise independent graphs will fail to meet ‘global’ qualities like connectivity? This motivates studying which global attributes of random graphs are captured by their k -wise independent counterparts.

Before elaborating on properties of k -wise independent graphs we provide some background on k -wise independence, on properties of random graphs, and on the emulation of huge random graphs.

1.1 Emulation of Huge Random Graphs

Suppose that one wishes to test the execution of some graph algorithm on random input graphs. Utilizing $\mathcal{G}(N, p)$ graphs requires resources polynomial in N , which is infeasible when N is huge (for example, exponential in the input length, n , of the relevant algorithms). A plausible solution is to replace $\mathcal{G}(N, p)$ by a cheaper ‘random looking’ distribution \mathcal{G}_N . To this end, each graph G in the support of \mathcal{G}_N is represented by a very short binary string (called seed) $s(G)$, s.t. evaluating edge queries on G can be done efficiently when $s(G)$ is known; then, sampling a graph from \mathcal{G}_N is done by picking the seed uniformly at random.

Goldreich, Goldwasser and the second author were the first to address this scenario in [23, 36]. They studied emulation by computationally pseudorandom graphs that are indistinguishable from $\mathcal{G}(N, p)$ from the view of any $\text{poly}(\log(N))$ -time algorithm that inspects graphs via edge-queries of its choice. They considered several prominent properties of $\mathcal{G}(N, p)$ graphs, and constructed computationally pseudorandom graphs that preserve many, though not all, of those properties (see the final paragraph of Section 2).

We consider replacing random graphs by k -wise independent ones. The latter can be sampled and accessed using only $\text{poly}(k \log(N))$ -bounded resources. This is achieved thanks to efficient constructions of discrete k -wise independent variables by Joffe [26], see also Alon, Babai and Itai [1]: the appearance of any potential edge in the graph is simply decided by a single random bit (that has probability p to attain the value 1). Such k -wise independent graphs were used by Naor, Tromer and the second author [35] to efficiently capture arbitrary first-order properties of huge $\mathcal{G}(N, p)$ graphs (see Section 3.6), and by [23, 36] as a building block for their main construction.

1.2 k -Wise Independent Random Variables

Distributions of discrete k -wise independent variables play an important role in computer science. Such distributions are mainly used for de-randomizing algorithms (and for some cryptographic applications). In addition, the randomness complexity of constructing k -wise independent variables was studied in depth, and in particular, the aforementioned constructions [26, 1] (based on degree k polynomials over finite fields) are known to provide essentially the smallest possible sample spaces. Our work is, however, the first systematic study of *combinatorial properties* of k -wise independent objects. Properties of various other k -wise independent objects (mainly percolation on \mathbb{Z}^d and on Galton-Watson trees) were subsequently explored by Benjamini, Gurel-Gurevich and Peled [6].

1.3 The Combinatorial Structure of Random Graphs

What are the principal attributes of random graphs that k -wise independent ones should maintain? Most theorems that manifest the remarkable structure of random graphs state that certain properties occur either almost surely (a.s. for short), or alternatively hardly ever, (namely, with probability tending either to 1 or to 0 as N grows to ∞). These results typically fall into one of the following categories.

Tight concentration of measure. A variety of prominent random variables (regarding random graphs) a.s. attain only values that are *extremely close* to their expectation. For instance, random graphs (with, say, constant p) a.s. have connectivity number $\kappa = (1 \pm o(1))pN$, clique number $c = (1 \pm o(1)) \frac{2 \log(pN)}{\log(1/p)}$ (Bollobás and Erdős [10], Matula [34], Frieze [22]) and chromatic number $\chi = (1 \pm o(1)) \frac{N \log(1/(1-p))}{2 \log(pN)}$ (Bollobás [9], Łuczak [33]).

Thresholds for monotone properties. For a given monotone increasing¹ graph property T , how large should $p(N)$ be for the property to hold a.s.? This question has been settled for many prominent properties such as connectivity (Erdős and Rényi [14]), containing a perfect matching (Erdős and Rényi [16, 17, 18]), Hamiltonicity (Pósa [37], Koršunov [29], Komlós and Szemerédi [30]), and the property of containing copies of some fixed graph H (Erdős and Rényi [15], Bollobás [8]). For these (and other) graph properties the sufficient density (for obtaining the property) is surprisingly small, and moreover, a threshold phenomenon occurs when by ‘slightly’ increasing the density from $\underline{p}(N)$ to $\bar{p}(N)$, the probability that T holds dramatically changes from $o(1)$ to $1 - o(1)$.² Thus, good emulation requires the property T to be guaranteed at densities as close as possible to the true $\mathcal{G}(N, p)$ threshold.

Zero-one laws. These well known theorems reveal that *any* first-order property holds either a.s. or hardly ever for $\mathcal{G}(N, p)$. A first-order property is any graph property that can be expressed by a single formula in the canonical language where variables stand for vertices and

¹Namely, any property closed under graph isomorphism and under addition of edges.

²Thresholds for prominent properties are often so sharp that $\bar{p} = (1 + o(1))\underline{p}$. Somewhat coarser thresholds were (later) established for *arbitrary* monotone properties by Bollobás and Thomason [11], and by Friedgut and Kalai [21].

the only relations are equality and adjacency (e.g. “having an isolated vertex” is specified by $\exists x \forall y \neg \text{EDGE}(x, y)$). These Zero-one laws hold for any fixed p (Fagin [19], Glebskii, Kogan, Liagonkii and Talanov [24]), and whenever $p(N) = N^{-\alpha}$ for a fixed irrational α (Shelah and Spencer [39]).

2 Our Contribution

We investigate the properties of k -wise independent graphs by providing upper bounds and lower bounds on the ‘minimal’ amount of independence, k_T , required for maintaining the main properties T of random graphs. The properties considered are: connectivity, perfect matchings, Hamiltonicity, the connectivity-number, clique-number and chromatic-number and the appearance of copies of a fixed subgraph H . We mainly establish upper bounds on k_T (where arbitrary k -wise independent graphs are shown to exhibit the property T) but also lower bounds (that provide specific constructions of k -wise independent graphs that fail to preserve T). Our precise results per each of these properties are discussed in Section 3, and proved in Section 5 (and the appendices). Interestingly, our results reveal a deep difference between k -wise independence and almost k -wise independence (a.k.a. (k, ϵ) -wise independence³). All aforementioned graph properties are guaranteed by k -wise independence (even for small $k = \text{poly}(\log(N))$), but are strongly violated by some almost k -wise independent graphs - even when $k = N^{\Omega(1)}$ is huge and $\epsilon = N^{-\Omega(1)}$ is tiny. For some properties of random graphs, T , our results demonstrate for the first time how to efficiently construct random-looking distributions on huge graphs that satisfy T .

Our Techniques & Relations to Combinatorial Pseudorandomness. For positive results (upper bounding k_T), we note that the original proofs that establish properties of $\mathcal{G}(N, p)$ graphs often fail for k -wise independent graphs. These proofs use a union bound over $M = 2^{\Theta(N)}$ undesired events, by giving a $2^{-\Omega(N)}$ upper-bound on the probability of each of these events.⁴ Unfortunately, there exist $\text{poly}(\log(N))$ -wise independent graphs where any event that occurs with positive probability, has probability $\geq 2^{-o(N)}$. Therefore, directly ‘de-randomizing’ the original proof fails, and alternative arguments (suitable for the k -wise independent case) are provided.

In particular, many properties are inferred via a variant of Thomason’s notion of ‘jumbledness’ [41] (mostly known in its weaker form as quasirandomness or pseudorandomness, as defined by Chung, Graham and Wilson [13], and related to the so called Expander Mixing Lemma and the pseudo-random properties of graphs that follow from their spectral properties, see [2]). For our purposes, α -jumbledness means that (as expected in $\mathcal{G}(N, p)$ graphs) for all vertex-sets U, V , the number of edges that pass from U to V should be $p|U||V| \pm \alpha\sqrt{|U||V|}$. Jumbledness and quasirandomness have been studied extensively (see [31] and its many references), and serve in Graph Theory as *the* common notion of resemblance to random graphs. In particular, $\mathcal{G}(N, p)$ graphs are known to exhibit (the best possible) jumbledness parameter, $\alpha = \Theta(\sqrt{pN})$. One of our main results (Theorem 1) demonstrates that k -wise independence for $k = \Theta(\log(N))$ is stronger than

³ (k, ϵ) -wise independence means that the joint distribution of any k potential edges is only required to be within small statistical distance ϵ from the corresponding distribution in the $\mathcal{G}(N, p)$ case.

⁴For instance w.r.t. connectivity, M is the number of choices for partitioning the vertices into 2 disconnected components.

jumbledness, in the sense that it guarantees the optimal $\alpha = \Theta(\sqrt{pN})$ even for tiny densities $p = \Theta(\frac{\ln(N)}{N})$. Therefore, prominent properties of k -wise independent graphs can be directly deduced from properties of jumbled graphs.

Proving Theorem 1 exploits a known connection between jumbledness and the eigenvalues of (a shifted variant of) the adjacency matrix of graphs, following the approach of Alon and Chung [2]. In particular, the analysis of Vu ([42], extending [20]) regarding the eigenvalues of random graphs is strengthened, in order to achieve optimal eigenvalues even for smaller densities p than those captured by [42]. This improvement implies, among other results, the remarkable fact that k -wise independent graphs for $k = \Theta(\log(N))$ preserve (up to constant factors) the $\mathcal{G}(N, p)$ sufficient density for connectivity.

More on Techniques & Relations to Almost k -Wise Independence. For negative results (producing random-looking graphs that defy a given property T of random graphs), the [23, 36] approach is to first construct some random-looking graph G , and later to ‘mildly’ modify G s.t. T is defied. This is done w.r.t. all graph properties considered here. For instance, the modification of choosing a random vertex and then deleting all its edges violates connectivity while preserving computational pseudorandomness. Unfortunately, such modifications fail to preserve k -wise independence (the resulting graphs are only almost k -wise independent). In contrast, most of our negative results exploit the fact that some constructions of k -wise independent bits produce strings with significantly larger probability than in the completely independent case. This is translated (by the construction in Lemma 5) to the unexpected appearance of some subgraphs (in k -wise independent graphs): either huge independent sets inside dense graphs or fixed subgraphs inside sparse graphs.

Comparison with Computational Pseudorandomness. Finally, k -wise independence guarantees all random graphs’ properties that were met by the (specific) computationally pseudorandom graphs of [23, 36]. In addition, only k -wise independence is known to capture (i) arbitrary first-order properties of $\mathcal{G}(N, p)$ graphs, (ii) high connectivity, (iii) strongest possible parameters of jumbledness, and (iv) almost regular $(1 \pm o(1))pN$ degree for all vertices, and $(1 \pm o(1))p^2N$ co-degrees for all vertex pairs. A single exception is that in [23, 36] the chromatic number of random graphs is achieved precisely, while here it is met only up to a constant factor. Importantly, our results hold for any k -wise independent graphs, (and in particular for the very simple and efficiently constructable ones derived from [26, 1]), whereas the approach of [23, 36] requires non-trivial modifications of the construction per each new property.

3 Combinatorial Properties of k -Wise Independent Graphs

We now survey our main results per each of the aforementioned graph properties T . Typically our arguments establish the following tradeoff: the smaller p is, the larger k should be to maintain T . Given this tradeoff we highlight minimizing k or, alternatively, minimizing p . The latter is motivated by the fact that the $\mathcal{G}(N, p)$ threshold for many central properties occurs at some $p^* \ll 1$. Minimizing p is subject to some reasonable choice of k , which is $k \leq \text{poly}(\log(N))$. Indeed, as the complexity of implementing k -wise independent graphs is $\text{poly}(k \log(N))$, we

get efficient implementations whenever $k \leq \text{poly}(\log(N))$ even when the graphs are huge and $N = 2^{\text{poly}(n)}$.⁵ In Section 3 whenever $k = \Theta(f(N))$ is said to suffice for capturing a property T whenever $p \geq \Theta(q(N))$, the meaning is that $k \geq cf(N)$ suffices for *some* constant c depending on T and on the constant hidden in the $\Theta(q(N))$ notation.

3.1 Connectivity, Hamiltonicity and Perfect Matchings (see Section 5.2)

The well known sufficient $\mathcal{G}(N, p)$ density for all these properties is $\sim \frac{\ln(N)}{N}$. For connectivity, this sufficient density is captured (up to constant factors) by $\Theta(\log(N))$ -wise independent graphs. Even $k = 4$ suffices for some $p = N^{-\frac{1}{2}+o(1)}$. Based on Hefetz, Krivelevich and Szabo's [25], Hamiltonicity (and hence perfect matchings) are guaranteed at $p \geq \Theta(\frac{\log^2(N)}{N})$ with $k \geq \Theta(\log(N))$, and at some $p \geq N^{-\frac{1}{2}+o(1)}$ with $k \geq 4$. On the other hand, some pair-wise independent graphs are provided that despite having constant density, are still a.s. disconnected and fail to contain any perfect matching.

3.2 High Connectivity (see Section 5.3)

The connectivity number, $\kappa(G)$, is the largest integer, ℓ , s.t. any pair of vertices is connected in G by at least ℓ internally vertex-disjoint paths. Since a typical degree in a random graph is $(1 \pm o(1))pN$, it is remarkable that $\mathcal{G}(N, p)$ graphs a.s. achieve $\kappa = (1 \pm o(1))pN$. Surprisingly, such optimal connectivity is guaranteed by $\Theta(\log(N))$ -wise independence whenever $p \geq \Theta(\frac{\log(N)}{N})$, and by $k \geq 4$ whenever $p \gg N^{-\frac{1}{3}}$.

3.3 Cliques and independent sets (see Appendix 7)

For $N^{-o(1)} \leq p \leq 1 - N^{o(1)}$ the independence number, I , of random graphs has a.s. only two possible values: either S^* or $S^* + 1$ for some $S^* = (1 - o(1))\frac{2 \log(pN)}{\log(1/(1-p))}$. This remarkable phenomenon is observed to hold by virtue of $\Theta(\log^2(N))$ -wise independence whenever p is bounded away from 0. On the other hand, k -wise independent graphs are provided with $k = \Theta(\frac{\log(N)}{\log \log(N)})$ where $I \geq (S^*)^{1+\Omega(1)}$ a.s. (for $k = \Theta(1)$, even huge $N^{\Omega(1)}$ independent sets may appear). For smaller densities, random graphs a.s. have $I \leq O(p^{-1} \log(N))$, while $\Theta(\log(N))$ -wise independence gives a weaker, yet useful, $I \leq O(\sqrt{N/p})$ bound whenever $p \geq \Omega(\frac{\log(N)}{N})$. By symmetry (replacing p with $1 - p$), analogous results to all the above hold for the clique number as well. Discussing the clique- and independence-number is deferred to the appendices, since the main relevant techniques are demonstrated elsewhere in the paper.

3.4 Coloring (see Section 5.5)

For $1/N \ll p \leq 1 - \Omega(1)$, the chromatic number χ of random graphs is a.s. $(1 + o(1))\frac{N \log(1/(1-p))}{2 \log(pN)}$. Given any $p \geq (\log(N))^{-O(1)}$, this $\mathcal{G}(N, p)$ lower-bound on χ is observed to hold for any

⁵Accessing the graphs via edge-queries is adequate only when $p \geq n^{-\Theta(1)}$ - otherwise a.s. no edges are detected by the $\text{poly}(n)$ inspecting algorithm. For smaller densities our study has thus mostly a combinatorial flavor.

$(\log(N))^c$ -wise independent graphs for some sufficiently large c . More surprisingly, $k = \Theta(\log(N))$ suffices to capture a similar upper-bound even for densities as small as $p = \Theta(\log(N)/N)$. Such upper-bounds are also implied by $k = 12$ for some larger densities $p = N^{-\Omega(1)}$. On the other hand, for $k = 2$ a huge $\chi = \Theta(N)$ might a.s. occur for constant ps (Theorem 4). The k -wise independent upper-bounds on χ are based on results of Alon, Krivelevich and Sudakov [3], [4] and of Johansson [27].

3.5 Thresholds for the Appearance of Subgraphs (see Section 5.4)

For a fixed (non-empty) graph H , consider the appearance of H -copies (*not necessarily* as an induced subgraph) in either a random or a k -wise independent graph. The $\mathcal{G}(N, p)$ threshold for the occurrence of H sub-graphs lies at $p_H^* \stackrel{\text{def}}{=} N^{-\rho}$, where the constant $\rho = \rho(H)$ is the minimum, taken over all subgraphs H' of H (including H itself), of the ratio $\frac{v(H')}{e(H')}$ (here, $v(H')$ and $e(H')$ respectively denote the number of vertices and edges in H'). Thus, no H -copies are found when $p \ll p^*$, while for any $p \gg p^*$, copies of H abound (Erdős and Rényi [15], Bollobás [8]). For any graph H , this $\mathcal{G}(N, p)$ threshold holds whenever $k \geq cv^4(H)$ (for some constant c), but as k is decreased to $\lfloor \frac{2}{\rho} \rfloor$, the $\mathcal{G}(N, p)$ threshold is defied: much sparser graphs exist where $p \ll p_H^*$ and yet copies of H are a.s. found. In particular, when $e(H) \geq \Omega(v^2(H))$, the threshold violation occurs at $k = \Omega(v(H))$.

3.6 First-Order Zero-One Laws (Previous Results)

A recent study (of Naor, Tromer and the second author [35]) considered capturing arbitrary depth- $D(N)$ properties of random graphs. These are graph properties expressible by a sequence of first-order formulas $\Phi = \{\phi_N\}_{N \in \mathbb{N}}$, with quantifier depth $\text{depth}(\phi_N) \leq D(N)$ (e.g. “having a clique of size $t(N)$ ” can be specified by $\phi_N = \exists x_1 \dots \exists x_{t(N)} \bigwedge_{i \neq j} (\text{EDGE}(x_i, x_j))$). A ‘threshold’ depth function $D^* = \frac{\log(N)}{\log(1/p)}$ was identified s.t. a graph sampled from any k -wise independent distribution simultaneously agrees with a random $\mathcal{G}(N, p)$ graph on all depth $(1 - o(1))D^*$ properties whenever $k \gg (D^*)^2$. In contrast, any efficiently computable graphs are strongly separated from $\mathcal{G}(N, p)$ by properties of only slightly higher depth $(1 + o(1))D^*$. These results are incomparable to the ones in the current paper, since most of the graph properties studied here require larger depth than D^* .

4 Preliminaries

Asymptotics. Invariably, $k : \mathbb{N} \rightarrow \mathbb{N}$, while $p, \epsilon, \delta, \gamma, \Delta : \mathbb{N} \rightarrow (0, 1)$. We often use $k, p, \epsilon, \delta, \gamma, \Delta$ instead of $k(N), p(N), \epsilon(N), \delta(N), \gamma(N), \Delta(N)$. Asymptotics are taken as $N \rightarrow \infty$, and some inequalities hold only for sufficiently large N . The $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ operators are ignored whenever insignificant for the asymptotic results. Constants c, \bar{c} are not optimized in expressions of the form $k = c \log(N)$ or $p = (\log(N))^{\bar{c}}/N^\Delta$, whereas the constant Δ is typically optimized.

Subgraphs. For a graph H , let $v(H)$ and $e(H)$ denote the number of vertices and edges in H . For vertex sets U, V let $e(U, V)$ denote the number of edges that pass from U to V (if $S =$

$U \cap V \neq \emptyset$, then any internal edge of S is counted twice). Similarly, we let $e(U) = e(U, U)$.

Random and k -Wise Independent Graphs. Throughout, graphs are simple, labeled and undirected. Given N, k, p as above then $\mathcal{G}^{k(N)}(N, p(N))$ (or $\mathcal{G}^k(N, p)$ for short) denotes some distribution over the set of graphs with vertex set $\{1, \dots, N\}$, where each edge appears w.p. $p(N)$, and the random variables that indicate the appearance of any $k(N)$ potential edges are mutually independent. We use the term ‘ k -wise independent graphs’ for a sequence of distributions $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ indexed by N .

Almost Sure Graph Properties. A graph property T , is any property closed under graph isomorphism. We say that ‘ T holds a.s. (almost surely) for $\mathcal{G}^k(N, p)$ ’ or that (abused notation) ‘ T holds for $\mathcal{G}^k(N, p)$ ’ whenever $\Pr_{\mathcal{G}^k(N, p)}[T] \xrightarrow{N \rightarrow \infty} 1$. Similar terminology is used for $\mathcal{G}(N, p)$ graphs.

Monotonicity in (k, p) . Since \bar{k} -wise independence implies k -wise independence for all $\bar{k} > k$ we may state claims for arbitrary $k \geq k'$ but prove them only for $k = k'$. When establishing monotone increasing properties we often state claims for arbitrary $p \geq p'$ but prove them only for $p = p'$. The latter is valid since for any $N, k, p > p'$, the process of sampling from any (independent) $\mathcal{G}^k(N, p), \mathcal{G}^k(N, p'/p)$ distributions and defining the final graph with edge-set being the intersection of the edge-sets of the two sampled graphs, clearly results in a $\mathcal{G}^k(N, p')$ distribution.

k -Wise Independent Random Variables. The term ‘ (M, k, p) -variables’ stands for any M binary variables that are k -wise independent with each variable having probability p of attaining value 1. Lemma 1 (proved in Section 6.1) adjusts the known construction of discrete k -wise independent variables of [26],[12], [1] to provide (M, k, p) -variables that induce some predetermined values with relatively high probability. Throughout, e_1 and e_0 resp. denote the number of edges and non-edges in a graph H .

Lemma 1 *Given $0 < p < 1$ with binary representation $p = 0.b_1\dots b_\ell$, and natural numbers e_0, e_1, M satisfying $e_0 + e_1 \leq M$, let $F = \max\{2^{\lceil \log_2 M \rceil}, 2^\ell\}$. Then there exists (M, k, p) -variables s.t. $\Pr[A] = F^{-k}$, where A denotes the event that the first e_0 variables receive value 0 while the next e_1 variables receive value 1.*

Tail Bounds for k -Wise Independent Random Variables. The following strengthened version of standard tail bounds (proved in Section 6.2) translates into smaller densities p for which monotone graph properties are established for k -wise independent graphs. Similar bounds were already obtained by Schmidt, Siegel and Srinivasan [40].

Lemma 2 *Let $X = \sum_{j=1}^M X_j$ be the sum of k -wise independent binary variables where $\Pr[X_j = 1] = \mu$ holds for all j . Let $\delta > 0$, and let k be even s.t. $\frac{M-k}{k} \mu(1-\mu) \geq 1$. Then*

$$\Pr[|X - \mathbb{E}(X)| \geq \delta \mathbb{E}(X)] \leq \left[\frac{2k(1-\mu)}{\delta^2 \mu M} \right]^{\frac{k}{2}}.$$

5 The properties of k -wise independent graphs

5.1 Degrees, Co-Degrees and Jumbledness

Lemma 3 (Achieving almost regular degrees) *In all k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ it a.s. holds that all vertices have degree $p(N-1)(1 \pm \epsilon)$ whenever $N \left[\frac{3k}{\epsilon^2 p N} \right]^{\lfloor k/2 \rfloor} \rightarrow 0$, and in particular when either*

1. $k \geq 4$, $N^{-1/2} \ll p \leq 1 - \frac{5}{N}$, and $1 \geq \epsilon \gg p^{-1/2} N^{-1/4}$; or
2. $k \geq 4 \log(N)$, $\frac{25 \log(N)}{N} \leq p \leq 1 - \frac{5 \log(N)}{N}$, and $1 \geq \epsilon \geq \sqrt{\frac{25 \log(N)}{p N}}$.

Proof. Fix a vertex v , and let X_w be the random variable that indicates the appearance of the edge $\{v, w\}$ in the graph. Thus, the degree of v is $X = \sum_{w \neq v} X_w$. Since X is the sum of $(N-1, k, p)$ -variables, Lemma 2 implies that the probability that v has an unexpected degree $X \neq p(N-1)(1 \pm \epsilon)$ is bounded by $\left[\frac{3k}{\epsilon^2 p N} \right]^{\lfloor k/2 \rfloor}$. Applying a union-bound over the N possible vertices v , gives that the probability of having *some* vertex with unexpected degree is bounded by $N \left[\frac{3k}{\epsilon^2 p N} \right]^{\lfloor k/2 \rfloor}$, which vanishes for the parameters in items 1 and 2. ■

Lemma 4 (Achieving almost regular co-degrees) *In all k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ it a.s. holds that all vertex pairs have co-degree $p^2(N-2)(1 \pm \gamma)$ whenever either*

1. $k \geq 12$, $N^{-\frac{1}{6}} \ll p \leq 1 - \frac{13}{N}$, and $1 \geq \gamma \gg p^{-1} N^{-\frac{1}{6}}$; or
2. $k \geq 12 \log(N)$, $\sqrt{\frac{73 \log(N)}{N}} \leq p \leq 1 - \frac{13 \log(N)}{N}$ and $1 \geq \gamma \geq \sqrt{\frac{73 \log(N)}{p^2 N}}$.

Proof. The proof is completely analogous to that of Lemma 3. Here the union-bound is over all $\binom{N}{2}$ vertex pairs $\{u, v\}$, and the co-degree of each $\{u, v\}$ is the sum of $(N-2, \lfloor \frac{k}{2} \rfloor, p^2)$ -variables. ■

The following definition is a modified version of the one in [41, 13], see also [2] and [5], Chapter 9.

Definition 1 (Jumbledness) *For vertex sets U, V , let $e(U, V)$ denote the number of edges that pass from U to V (internal edges of $U \cap V$ are counted twice). A graph is (p, α) -jumbled if $e(U, V) = p|U||V| \pm \alpha \sqrt{|U||V|}$ holds for all U, V .*

Theorem 1 (Achieving optimal jumbledness) *There exist absolute constants c_1, c_2, c_3 s.t. all k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ are a.s. (p, α) -jumbled whenever either:*

1. $k \geq 4$, $p \geq \Omega(\frac{1}{N})$ and $\alpha \gg \sqrt{p} N^{3/4}$; or
2. $k \geq \log(N)$, $\frac{c_1 \log(N)}{N} \leq p \leq 1 - \frac{c_2 \log^4(N)}{N}$ and $\alpha \geq c_3 \sqrt{p N}$.

Proof. The proof is based on spectral techniques and combines some refined versions of ideas from [2], [20] and [42], using the fact that traces of the k -th power of the adjacency matrix of a graph are identical in the k -wise independent case and in the totally random one. The details are lengthy and are thus deferred to Appendix 8.

5.2 Connectivity, Hamiltonicity and Perfect Matchings

Theorem 2 (Achieving connectivity) *There exists a constant c s.t. the following holds. All k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ are a.s. connected whenever either:*

- $k \geq 4$ and $p \gg \frac{1}{\sqrt{N}}$; or
- $k \geq 4 \log(N)$ and $p \geq \frac{c \ln(N)}{N}$.

Proof. Let U be a vertex-set that induces a connected component. Connectivity follows from having $|U| > 0.5N$ for all such U . The following holds a.s. for $\mathcal{G}^k(N, p)$. By Lemma 3, all vertices have degree $\geq 0.9pN$, so $e(U) \geq 0.9pN|U|$. By Theorem 1, all sets U satisfy $e(U) \leq p|U|^2 + \alpha|U|$ with $\alpha = O(\sqrt{pN}) = o(pN)$. Re-arranging gives $(0.9 - o(1))N \leq |U|$. ■

Theorem 3 (Achieving Hamiltonicity) *All k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ are a.s. Hamiltonian (and for even N contain a perfect matching) whenever either:*

- $k \geq 4$ and $p \geq \frac{\log^2(N)}{\sqrt{N}}$; or
- $k \geq 4 \log(N)$ and $p \geq \frac{\log^2(N)}{N}$.

Proof. Let $\bar{\Gamma}(V)$ denote the set of vertices $v \notin V$ that are adjacent to some vertex in the vertex-set V . By Theorem 1.1 in Hefetz, Krivelevich and Szabo's [25], Hamiltonicity follows from the existence of constants b, c such that a.s. (i) $|\bar{\Gamma}(V)| \geq 12|V|$ holds for all sets V of size $\leq bN$, and (ii) $e(U, V) \geq 1$ holds for all disjoint sets U, V of size $\frac{cN}{\log(N)}$. We remark that (unlike other asymptotic arguments in this paper), the sufficiency of (i) and (ii) might hold only for very large N . For (i), let $b = \frac{1}{170}$ and consider an arbitrary set V . By Theorem 1, a.s. all vertex-sets T have $e(T) \leq p|T|^2 + o(pN)|T|$. By Lemma 3 a.s. all the degrees are $(1 \pm o(1))pN$, so exactly $(1 \pm o(1))pN|V|$ edges touch V (where internal edges are counted twice). Let $T = V \cup \bar{\Gamma}(V)$, and assume that $|\bar{\Gamma}(V)| < 12|V|$. We get $(1 - o(1))pN|V| \leq e(T) \leq p(13|V|)^2 + o(pN)|V|$. Re-arranging gives $|V| > \frac{N}{170}$. Condition (i) follows. For (ii), by Theorem 1, a.s. all (equal-sized and disjoint) vertex-sets U, V have $e(U, V) \geq p|U||V| - O(\sqrt{pN})|U|$. If there is no edge between U and V , then $e(U, V) = 0$. Re-arranging gives $|U| \leq O(\sqrt{N/p}) \leq O(\frac{N}{\log(N)})$. Condition (ii) follows. ■

Theorem 4 (Failing to preserve connectivity) *There exist some pair-wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ where $p = 1/2$ that (i) are a.s. disconnected (and contain no Hamiltonian cycles), (ii) contain no perfect matchings with probability 1, and (iii) a.s. have clique number and chromatic number $(1/2 \pm o(1))N$ and independence number 2.*

Proof. Consider the graphs defined by partitioning all vertices into 2 disjoint sets V_0, V_1 where each V_j induces a clique, no edges connect V_0 to V_1 , and V_1 is chosen randomly and uniformly among all subsets of odd cardinality of the vertex set. Note that for every set of 4 vertices, there are 16 ways to split its vertices among V_0 and V_1 , and it is not difficult to check that if $N \geq 5$, then each of these 16 possibilities is equally likely. Therefore, any edge appears w.p. $\frac{1}{2}$, and any pair of edges (whether they share a common vertex or not) appears w.p. $\frac{1}{4}$. Still the graph is

connected iff all the vertices belong to the same V_j which happens only w.p. 2^{-N+1} (and only if N is odd). Since $|V_1|$ is odd, the graph contains no perfect matching. It is easy to verify that a.s. $|V_0|, |V_1| = (1/2 \pm o(1))N$ implying the last item. ■

5.3 High-connectivity

Theorem 5 (Achieving optimal connectivity) *There exists an absolute constant c , s.t. for all k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ the connectivity number is a.s. $(1 \pm o(1))pN$ when either*

- $k \geq 4$ and $p \gg N^{-\frac{1}{3}}$; or
- $k \geq \log(N)$ and $p \geq c \frac{\log(N)}{N}$.

Proof. The connectivity is certainly not larger than $(1 + o(1))pN$, as it is upper-bounded by the minimum degree. By Theorem 2.5 in Thomason's [41] $\kappa \geq d - \alpha/p$ holds for any (p, α) -jumbled graph with minimal degree $\geq d$. Thus, achieving $\kappa \gtrsim pN$, reduces to obtaining (i) $d = (1 \pm o(1))pN$, and (ii) $\alpha \ll pd$. Condition (i) a.s. holds by Lemma 3. By Theorem 1, we a.s. achieve $(c_3\sqrt{pN})$ -jumbledness for some constant c_3 , so condition (ii) becomes $p^2N \gg \sqrt{pN}$. This proves the first part of the theorem. To prove the second we note, first, that we may assume that $p \ll 1$ (since otherwise 4-wise independence suffices). Let S be a smallest separating set of vertices, assume that $|S|$ is smaller than $(1 - o(1))pN$, let U be the smallest connected component of $G - S$ and let W be the set of all vertices but those in $U \cup S$. Clearly $|W| \geq (\frac{1}{2} - o(1))N$. Note that $e(U, W) = 0$, but by jumbledness $e(U, W) \geq p|U||W| - c_3\sqrt{pN}|U||W|$. This implies, using the fact that $|W| > N/3$, that $|U| \leq \frac{3c_3^2}{p}$. Using jumbledness again, $e(U, S) \leq p|U||S| + c_3\sqrt{pN}|U||S|$ but as all degrees are at least $(1 - o(1))pN$, $e(U, S) \geq (1 - o(1))pN|S| - e(U) \geq (1 - o(1))pN|U| - p|U|^2 - c_3\sqrt{pN}|U| \geq |U|(1 - o(1))pN$, where here we used the fact that $|U| \leq O(1/p)$ and that $\sqrt{pN} = o(pN)$. This implies that either $p|U||S| \geq \frac{1}{2}|U|pN$, implying that $|S| \geq N/2 \gg pN$, as needed, or $c_3\sqrt{pN}|U||S| \geq \frac{1}{3}|U|pN$, implying that $|S| \geq \frac{1}{9c_3^2}|U|pN$ which is bigger than pN provided $|U| \geq 9c_3^2$. However, if $|U|$ is smaller, then surely $|S| \geq (1 - o(1))pN$, since all degrees are at least $(1 - o(1))pN$ and every vertex in U has all its neighbors in $U \cup S$. ■

5.4 Thresholds for the Appearance of Subgraphs

For a fixed non-empty graph H , let $\rho(H)$ and p_H^* be as in Section 3.5.

Observation 1 (Preserving the threshold for appearance of sub-graphs) *There exists a function $D(v) = (1 \pm o(1))\frac{v^4}{16}$ s.t. for any graph H with at most v vertices, and for all k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ with $k \geq D(v)$ the following holds. Let A denote the event that H appears in $\mathcal{G}^k(N, p)$ (not necessarily as an induced sub-graph). Then*

- If $p(N) \ll p_H^*(N)$ then $(\neg A)$ a.s. holds.
- If $p(N) \gg p_H^*(N)$ then A a.s. holds.

Proof. The proof (given in Appendix 6.3) applies Rucinski and Vince's [38] to specify a sufficiently large k for the original $\mathcal{G}(N, p)$ argument to hold. ■

Theorem 6 (Defying the threshold for appearance of sub-graphs) *For any (fixed) graph H that satisfies⁶ $\rho(H) < 2$, there exists k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ where $k = \lceil \frac{2}{\rho(H)} - 1 \rceil$ and $p(N) \ll p_H^*(N)$ s.t. H a.s. appears in $\mathcal{G}^k(N, p)$ as an induced sub-graph.*

Proof. Theorem 6 relies on Lemma 5. This lemma considers the appearance of the sub-graph H_N in $\mathcal{G}^k(N, p)$ where $\{H_N\}_{N \in \mathbb{N}}$ is any sequence of graphs (possibly) with unbounded order.

Lemma 5 (k -wise independent graphs with unexpected copies of sub-graphs) *Let $\{H_N\}_{N \in \mathbb{N}}$ be a sequence of graphs where H_N has exactly $S(N) < \sqrt{N}$ vertices, $e_1(N)$ edges and $e_0(N)$ non-edges. Assume that for each N there exists $((\binom{S(N)}{2}, k(N), p(N))$ -variables s.t. with probability $\Delta(N) \gg (S(N)/N)^2$ it holds that the first $e_0(N)$ variables attain value 0 and the next $e_1(N)$ variables attain value 1. Then there exist k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ that a.s. contain H_N -copies as induced sub-graphs.*

Proof (Lemma 5). Fix N , so $H = H_N, S = S(N), e_i = e_i(N), k = k(N), p = p(N), \Delta = \Delta(N)$. We construct graphs $\mathcal{G}^k(N, p)$ that a.s. contain H copies. Given the N vertices, let $\{V_j\}_{j=1}^M$ be any maximal collection of *edge-disjoint* vertex-sets, each of size $|V_j| = S$. For each j , decide the internal edges of V_j by some $((\binom{S}{2}, k, p)$ -variables s.t. H is induced by V_j with probability Δ . This can be done by appropriately defining which specific edge in V_j is decided by which specific variable. Critically, the constructions for distinct sets V_j are totally independent. The $R = \binom{N}{2} - M \binom{S}{2}$ remaining edges can be decided by any (R, k, p) -variables. The resulting graph is clearly k -wise independent.

The main point is that (i) the events of avoiding H -copies on the various sets V_j are totally independent (by the edge-disjointness of the V_j -s), and that (ii) in our k -wise independent case Δ is rather large (compared with the totally independent case). Thus, avoiding H -copies on *any* of the V_j -s is unlikely. Indeed, let B denote the event that no H -copies appear in the resulting graph, while B' only denotes the event that none of the V_j -s induces H . By Wilson's [44] and Kuzjurin's [32] we have $M = \Theta(N^2/S^2)$, so

$$\Pr[B] \leq \Pr[B'] = (1 - \Delta)^M \leq e^{-\Theta\left(\frac{\Delta N^2}{S^2}\right)},$$

which vanishes by our requirement that $\Delta \gg (S/N)^2$. ■ (Lemma 5)

Completing the proof of Theorem 6. For $v = v(H), \rho = \rho(H), p^* = p_H^*$, and some $1 \ll f(N) \leq N^{o(1)}$, define p s.t. p^{-1} is the minimal power of 2 that is larger than $\frac{f(N)}{p^*}$. As desired $p \ll p^*$. Let e_1 and e_0 respectively denote the number of edges and non-edges in H . With $M = \binom{v}{2}$ and $F = 1/p$, we apply Lemma 1 to produce (M, k, p) -variables s.t. with probability $\geq F^{-k}$ the first e_0 variables have value 0, and the remaining e_1 variables have value 1. By Lemma 5, the latter immediately implies the existence of k -wise independent graphs that a.s. contain H -copies as long as $F^k \ll (N/v)^2$. As $F = 1/p = N^{\rho+o(1)}$, this \ll requirement translates to $k\rho \lesssim 2$. ■ (Theorem 6)

⁶This condition rules out only graphs H that are a collection of disjoint edges. For such graphs $\rho(H) = 2$, so clearly no H -copies can be produced (even if $k = 1$) when $p(N) \ll p_H^*(N) = N^{-2}$.

5.5 The Chromatic Number

Observation 2 (Preserving the chromatic number lower bound) For any $c > 0$ there exists some $d > 0$, s.t. all k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ with $(\log(N))^{-c} \leq p \leq 1 - N^{-o(1)}$ and $k \geq d(\log(N))^{c+1}$ a.s. have chromatic number $\chi \geq \frac{N \log(1/(1-p))}{2 \log(pN)}$.

Proof. Let $I(G)$ denote the independence number of (a single) N -vertex graph G . Clearly, $\chi(G) \geq \frac{N}{I(G)}$, so observation 2 follows from the fact that a.s. $I \leq \frac{2 \log(pN)}{\log(1/(1-p))}$ which is precisely observation 3 in the Appendices. ■

Theorem 7 (Preserving the chromatic number upper bound) There exists an absolute constant c s.t. the following holds. All k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ with $p \leq 1/2$ a.s. have chromatic number $\chi \leq \frac{cN \log(1/(1-p))}{\log(pN)}$, whenever either:

1. $k \geq 12$ and $p \geq N^{-\frac{1}{75}}$; or
2. $k \geq \log(N)$ and $p \geq c \frac{\log(N)}{N}$.

Remark. No special effort was made to optimize the constants $\frac{1}{2}$ and $\frac{1}{75}$.

Proof (sketch). Since p is bounded from above and $\log(1/(1-p)) \xrightarrow{p \rightarrow 0} p/\ln(2)$, it suffices to show that a.s. $\chi \leq O(\frac{pN}{\log(pN)})$. Item 1 is based on Alon, Krivelevich and Sudakov's [3]. Specifically, choose $\delta = 1/25$, s.t. by item 1 in Lemma 3 (with $\epsilon = (\log(N))p^{-1/2}N^{-3/8}$) and by item 1 in Lemma 4 (with $\gamma = (\log(N))p^{-1}N^{-1/6}$), a.s. all the degrees are lower bounded by $pN(1 - p^{-1/2}N^{-3/8+o(1)}) \geq pN - N^{1-4\delta}$, and all co-degrees are upper bounded by $p^2N(1 + p^{-1}N^{-1/6+o(1)}) \leq p^2N - N^{1-4\delta}$. By Theorem 1.2 in [3], these conditions (with $\delta < 1/4$ and $p \geq N^{-\frac{\delta}{3}}$) imply that $\chi \leq \frac{4pN}{\delta \ln N} \leq O(\frac{pN}{\log(pN)})$.

Item 2 follows from jumbledness and the main result of Alon, Krivelevich and Sudakov in [4] (which is based on Johansson's [27]), by which any graph with maximum degree d in which every neighborhood of a vertex contains at most $d^{2-\beta}$ edges (for some constant β) has chromatic number $\chi \leq O(\frac{d}{\log d})$. ■

Acknowledgements. The second author wishes to thank Oded Goldreich for his encouragement, and Ori Gurel-Gurevich, Chandan Kumar Dubey, Ronen Gradwohl, Moni Naor, Eran Ofek, Ron Peled, and Ariel Yadin for useful discussions. We thank the anonymous referees of FOCS08 for their useful comments and for referring us to [40].

References

- [1] N. Alon, L. Babai, A. Itai. *A Fast and Simple Randomized Parallel Algorithm for the Maximal Independent Set Problem*. Journal of Algorithms 7, 567-583, 1986.
- [2] N. Alon, F. R. K. Chung. *Explicit construction of linear sized tolerant networks*. Discrete Math. 72, 15-19, 1988; (Proc. of the First Japan Conference on Graph Theory and Applications, Hakone, Japan, 1986.)

- [3] N. Alon, M. Krivelevich, B. Sudakov. *List Coloring of Random and Pseudo-Random Graphs*. *Combinatorica* 19 (1999), 453-472.
- [4] N. Alon, M. Krivelevich, B. Sudakov. *Coloring graphs with sparse neighborhoods*. *J. Combinatorial Theory, Ser. B* 77 (1999), 73-82.
- [5] N. Alon, J. Spencer. *The Probabilistic Method*. John Wiley, New York, 1992.
- [6] I. Benjamini, O. Gurel-Gurevich, R. Peled. *On k -wise independent distributions and boolean functions*. To appear.
- [7] B. Bollobás. *Random Graphs*. Academic Press, 1985.
- [8] B. Bollobás. *Random Graphs*. In *Combinatorics* (Swansea, 1981), Volume 52 of London. Math. Soc. Lecture Note Ser., 80102. Cambridge Univ. Press, 1981.
- [9] B. Bollobás. *The Chromatic Number of Random Graphs*. In *Combinatorica* 8 49-55, 1988.
- [10] B. Bollobás, P. Erdős. *Cliques in Random Graphs*. *Math Proc Camb Phil Soc* 80 (1976), 419-427.
- [11] B. Bollobás, A. Thomason. *Threshold Functions*. *Combinatorica* 7 (1986), 35-38.
- [12] B. Chor, O. Goldreich. *On the Power of Two-Point Based Sampling*. *J. Complexity* 5(1): 96-106 (1989).
- [13] F. R. K. Chung, R. L. Graham, R. M. Wilson. *Quasi-Random Graphs*. *Combinatorica* 9, 345-362, 1989.
- [14] P. Erdős, A. Rényi. *On Random Graphs I*. *Publicationes Mathematicae* 6 (1959), 290-297.
- [15] P. Erdős, A. Rényi. *On the Evolution of Random Graphs*. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17-61, 1960.
- [16] P. Erdős, A. Rényi. *On Random Matrices*. *Publicationes Mathematicae* 8 (1964), 455-461.
- [17] P. Erdős, A. Rényi. *On the Existence of a Factor of Degree One of a Connected Random Graph*. *Acta Mathematica* 17 (1966), 359-368.
- [18] P. Erdős, A. Rényi. *On Random Matrices ii*. *Studia Sci. Math. Hungar.* 3, 459-464, 1968.
- [19] R. Fagin. *Probabilities in Finite Models*, *Journal of Symbolic Logic*, Vol. 41, 50-58, 1969.
- [20] Z. Füredi, J. Komlos. *The eigenvalues of random symmetric matrices*. *Combinatorica* 1 (1981), 233-241.
- [21] E. Friedgut, G. Kalai. *Every Monotone Graph Property Has a Sharp Threshold*. *Proc. Amer. Math. Soc.* 124 (1996), 2993-3002.
- [22] A. Frieze. *On the Independence Number of Random Graphs*. *Discrete Math.* 81 171-175, 1990
- [23] O. Goldreich, S. Goldwasser, A. Nussboim. *On the Implementation of Huge Random Objects*. In *Proc. 44th IEEE Symposium on Foundations of Computer Science*, 68-79, 2003.
- [24] Y. V. Glebskii, D. I. Kogan, M. I. Liagonkii, V. A. Talanov. *Range and Degree of Realizability of Formulas in the Restricted Predicate Calculus*. *Cybernetics*, Vol. 5, 142-154, 1976.

- [25] D. Hefetz, M. Krivelevich, T. Szabo. *Hamilton Cycles in Highly Connected and Expanding Graphs*. Preprint.
- [26] A. Joffe. *On a Set of Almost Deterministic k -Wise Independent Random Variables*. Annals of Probability 2, 1961-1962, 1974.
- [27] A.R. Johansson. *Asymptotic Choice Number for Triangle Free Graphs*. DIMACS Technical Report 91-5.
- [28] S. Janson, T. Łuczak, A. Rucinski. *Random Graphs*. New York: Wiley, 2000.
- [29] A.D. Koršunov. *Solution of a Problem of Erdős and Rényi on Hamiltonian Cycles in Nonoriented Graphs*. Dokl. Akad. Nauk SSSR Tom 228(1976) 760-764.
- [30] J. Komlós, E. Szemerédi. *Limit Distributions for the Existence of Hamilton Circuits in a Random Graph*. Discrete Math. 43 (1983) 55-63.
- [31] M. Krivelevich, B. Sudakov. *Pseudo-random Graphs*. In More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies 15, Springer, 2006, 199-262.
- [32] Nikolai N. Kuzjurin. *On the difference between asymptotically good packings and coverings*. European J. Combin. 16 (1995), no. 1, 35-40.
- [33] T. Łuczak. *The Chromatic Number of Random Graphs*. Combinatorica(11),45-54,1991.
- [34] D.W. Matula. *The Largest Clique Size in a Random Graph*. Tech. Rep. Dept. Comp. Sci. Southern Methodist Univ., Dallas, 1976.
- [35] M. Naor, A. Nussboim, E. Tromer. *Efficiently Constructible Huge Graphs that Preserve First Order Properties of Random Graphs*. Proceedings of the 2'nd Theory of Cryptography Conference, 66-85, 2005.
- [36] A. Nussboim. *Huge Pseudo-Random Graphs that Preserve Global Properties of Random Graphs*. M.Sc. Thesis, Advisor: S. Goldwasser, Weizmann Institute of Science, 2003, <http://www.wisdom.weizmann.ac.il/~asafn/psdgraphs.ps>.
- [37] L. Pósa. *Hamiltonian Circuits in Random Graphs*. Discrete Math 14 (1976), 359-364.
- [38] A. Rucinski, A. Vince. *Strongly Balanced Graphs and Random Graphs*. J. Graph Theory 10 (1986) 251-264.
- [39] S. Shelah, J. H. Spencer. *Zero-One Laws for Sparse Random Graphs*, Journal of the American Mathematical Society, Vol. 1, 97-115, 1988.
- [40] J. P. Schmidt, A. Siegel, A. Srinivasan. *Chernoff-Hoeffding Bounds for Applications with Limited Independence*. SIAM J. Discrete Math. 8(2): 223-250 (1995).
- [41] A. Thomason. *Pseudo-Random Graphs*. Proceedings of Random Graphs, Annals of Discrete Mathematics 33, 307-331, 1987.
- [42] V. H. Vu. *Spectral norm of random matrices*. STOC 2005, 423-430.
- [43] E. Wigner. *On the Distribution of the Roots of Certain Symmetric Matrices*. Ann. of Math. 67, 325-328, 1958.
- [44] R. M. Wilson. *Decomposition of complete graphs into subgraphs isomorphic to a given graph*. Congressus Numerantium XV (1975), 647-659.

6 Appendix - Detailed Proofs

6.1 Modified Construction of k -Wise Independent Variables - Proving Lemma 1

Recall that given any prime power F , the original [26, 12, 1] construction considers the field \mathbb{F} with elements $\{0, \dots, F-1\}$, and for each element $j \in \mathbb{F}$, a random variable Z_j is defined, s.t. the Z_j s are k -wise independent, and each Z_j is uniformly distributed in $\{0, \dots, F-1\}$. We derive from those Z_j -s some (M, k, p) binary variables X_j , by setting (i) $X_j = 1$ iff $\frac{Z_j+1}{F} \geq 1-p$ for $j = 1, \dots, e_0$, and (ii) $X_j = 1$ iff $\frac{Z_j+1}{F} \leq p$ for $j = e_0 + 1, \dots, e_0 + e_1$. Evidently, the X_j -s are k -wise independent with $\Pr(X_j = 1) = p$. Recall that $Z_j \stackrel{\text{def}}{=} Q(j)$ with Q being a uniformly random degree k polynomial over F , and let B denote the event that the 0-polynomial was chosen. Since B implies A , we get $\Pr[A] \geq \Pr[B] = F^{-k}$. ■

6.2 k -Wise Independence Tail Bound - Proving Lemma 2

Let $\bar{X}_i \stackrel{\text{def}}{=} X_i - \mu$ and $\bar{X} \stackrel{\text{def}}{=} \sum_{i=1}^M \bar{X}_i$, so $X - \mathbb{E}(X) = \bar{X}$. Thus,

$$\begin{aligned} \Pr[|X - \mathbb{E}(X)| \geq \delta \mathbb{E}(X)] &= \Pr[|\bar{X}| \geq \delta \mathbb{E}(X)] \\ &= \Pr[\bar{X}^k \geq (\delta \mathbb{E}(X))^k] \leq \frac{\mathbb{E}(\bar{X}^k)}{(\delta \mathbb{E}(X))^k}, \end{aligned}$$

the last equality holds for any even positive k , while the \leq employs Markov's inequality.

We bound $E \stackrel{\text{def}}{=} \mathbb{E}(\bar{X}^k)$ using the expansion

$$\bar{X}^k = \sum_{\vec{d} \in D} \prod_{i=1}^M \bar{X}_i^{d_i},$$

where $D \stackrel{\text{def}}{=} \{\vec{d} = (d_1, \dots, d_M) \mid \sum_{i=1}^M d_i = k, d_i \geq 0\}$. The k -wise independence of the variables X_i , guarantees the k -wise independence of the \bar{X}_i s, so

$$E = \sum_{\vec{d} \in D} \mathbb{E}(\prod_{i=1}^M \bar{X}_i^{d_i}) = \sum_{\vec{d} \in D} \prod_{i=1}^M \mathbb{E}(\bar{X}_i^{d_i}).$$

Next, since $\mathbb{E}(\bar{X}_i) = 0$ we can ignore all terms where $d_i = 1$ for some i . Namely, we consider only terms $\Pi = \prod_{\ell=1}^j \bar{X}_{i_\ell}^{d_{i_\ell}}$ where for some $j \leq \frac{k}{2}$, it holds that precisely j variables appear and for each variable \bar{X}_{i_ℓ} in Π we have $d_{i_\ell} \geq 2$. Hence,

$$E \leq \sum_{j=1}^{\frac{k}{2}} \Psi_j, \tag{1}$$

whenever Ψ_j bounds the contribution of all terms Π with precisely j variables.

Strengthening standard versions of the inequality begins by taking

$$\Psi_j \stackrel{\text{def}}{=} \binom{M}{j} j^k [\mu(1-\mu)]^j. \tag{2}$$

Indeed, $\binom{M}{j}j^k$ clearly bounds the number of terms Π with precisely j variables, while $[\mu(1-\mu)]^j$ bounds the expectation of each term Π that has precisely j variables because

$$\begin{aligned}\mathbb{E}[(\bar{X}_i)^d] &= \mu(1-\mu)^d + (1-\mu)(-\mu)^d \\ &\leq \mu(1-\mu)^d + (1-\mu)(+\mu)^d \\ &\leq \mu(1-\mu)[(1-\mu)^1 + \mu^1] = \mu(1-\mu)\end{aligned}\quad (3)$$

(the final \leq applies the facts $0 \leq \mu, 1-\mu \leq 1$ and $d \geq 2$). Thus, multiplying over the j terms gives

$$\Pi_{\ell=1}^j \mathbb{E}(\bar{X}_{i_\ell}^{d_{i_\ell}}) \leq [\mu(1-\mu)]^j.$$

Observe that Ψ_j is maximized when $j = \frac{k}{2}$. Indeed,

$$\begin{aligned}\frac{\Psi_{j+1}}{\Psi_j} &= \frac{M-j}{j+1} \left(\frac{j+1}{j}\right)^k \mu(1-\mu) \\ &> \frac{M-k}{k} \mu(1-\mu) \geq 1\end{aligned}$$

(the concluding ≥ 1 holds by the lemma's assumption).

Thus, the maximal Ψ_j is

$$\begin{aligned}\Psi_{\frac{k}{2}} &= \binom{M}{\frac{k}{2}} \left(\frac{k}{2}\right)^k [\mu(1-\mu)]^{\frac{k}{2}} \\ &\leq \frac{M^{\frac{k}{2}}}{(\frac{k}{2})!} \left(\frac{k}{2}\right)^k [\mu(1-\mu)]^{\frac{k}{2}} \\ &\leq \frac{(eM)^{\frac{k}{2}}}{\sqrt{2\pi\frac{k}{2}}(\frac{k}{2})^{\frac{k}{2}}} \left(\frac{k}{2}\right)^k [\mu(1-\mu)]^{\frac{k}{2}} \\ &= \frac{[\frac{e}{2}Mk\mu(1-\mu)]^{\frac{k}{2}}}{\sqrt{\pi k}}\end{aligned}$$

(Stirling's approximation for $(\frac{k}{2})!$ implies the last \leq).

To summarize, all the above gives

$$\begin{aligned}\Pr[|X - \mathbb{E}(X)| \geq \delta \mathbb{E}(X)] &\leq \frac{k\Psi_{(k/2)}}{(\delta \mathbb{E}(X))^k} \\ &\leq \frac{\frac{k}{\sqrt{\pi k}} (\frac{e}{2}Mk\mu(1-\mu))^{\frac{k}{2}}}{(\delta \mu M)^k} \\ &\leq \sqrt{\frac{k}{\pi}} \left[\frac{\frac{e}{2}k(1-\mu)}{\delta^2 \mu M} \right]^{\frac{k}{2}}.\end{aligned}$$

The Lemma follows as it can be directly shown that for all k

$$\sqrt{\frac{k}{\pi}} \left(\frac{e}{2}\right)^{\frac{k}{2}} \leq 2^{\frac{k}{2}}. \blacksquare$$

6.3 Appearance of Subgraphs - Proving observation 1

We first consider only balanced graphs H , namely graphs where $\rho(H) \leq \rho(H')$ for any subgraph $H' \subseteq H$. The original $\mathcal{G}(N, p)$ -threshold proof [15] takes a fixed graph F as a parameter, and considers for each set T of $v(F)$ distinct vertices the random variable Y_T^F which indicates whether T spans F in the resulting graph. Thus $Y^F \stackrel{\text{def}}{=} \sum_T Y_T^F$ counts the number of sets that span F .

First, the authors of [15] consider a specific subgraph $H' \subseteq H$ s.t. $\rho(H) = \frac{v(H')}{e(H')}$ and show that $p \ll p_H^*$ implies that $\mathbb{E}(Y^{H'}) \ll 1$. In this case, H' rarely appears in $\mathcal{G}(N, p)$ graphs and so does H . On the other hand, whenever $p \gg p_H^*$, they show that $\mathbb{E}(Y^H) \gg 1$ and by Chebyshev's inequality it is deduced (only here the fact that H is balanced is used) that some H -copies appear. Thus, the entire argument applies only probabilities regarding either a single variable Y_T^F , or a pair $Y_T^F, Y_{T'}^F$ of variables, and relies only upon the independence of sets of $m \leq 2^{\binom{v(H)}{2}}$ edges.

For non-balanced graphs the $p \ll p_H^*$ part holds as for balanced ones. For $p \gg p_H^*$, we rely on the fact that for any graph H , there exists an extension graph $H \subseteq H''$ s.t. H'' is balanced and $\rho(H'') = \rho(H)$ (Rucinski and Vince [38]). Since $p \gg N^{-\rho(H)}$ means that $p \gg N^{-\rho(H'')}$, and since H'' is balanced, then $\mathcal{G}(N, p)$ graphs a.s. contain copies of H'' , and copies of H appear as well. This time the Chebyshev argument assumes only the independence of sets of $m \leq 2^{\binom{v(H'')}{2}}$ edges. Since by [38] there exists H'' as above with $v(H'') \leq (1 + o(1)) \frac{[v(H)]^2}{4}$, then $m = (1 \pm o(1)) \frac{[v(H)]^4}{16}$ suffices. ■

7 Appendix - The Independence Number of k -Wise Independent graphs

The following positive result follows the argument used to establish observation 1.

Observation 3 (Preserving random graphs' precise independence number) *Consider arbitrary k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ where $N^{-o(1)} \leq p(N) \leq 1 - N^{-o(1)}$, and let $I(N) = I(\mathcal{G}^k(N, p))$ denote the independence number of $\mathcal{G}^k(N, p)$. Then there exists a function $S^*(N, p) = (1 - o(1)) \frac{2 \log(pN)}{\log(1/(1-p))}$, s.t. if $k(N) \geq S^*(N, p) + 2$, then a.s. $I(N) \leq S^*(N, p) + 1$, and if $k(N) \geq \binom{S^*(N, p)}{2}$, then a.s. $I(N) \geq S^*(N, p)$.*

Proof. The classical proof ([10], [34]) of this claim for $\mathcal{G}(N, p)$ graphs considers for each set T of S distinct vertices (S being a parameter) the random variable Y_T^S which indicates whether T spans an independent set in the resulting graph. Thus $Y^S = \sum_T Y_T^S$ counts the total number of independent sets of size S . It is shown that for $S = S^* + 2$ then $\mathbb{E}(Y^S) \ll 1$ so a.s. the independence number $\leq S^* + 1$. On the other hand, for $S = S^*$ then $\mathbb{E}(Y^S) \gg (1)$, and by Chebyshev's inequality it is deduced that a.s. some independent sets of size S^* appear. This entire argument considers only probabilities regarding either a single variable Y_T^S (for the lower- and upper-bound on I), or a pair $Y_T^S, Y_{T'}^S$ of variables (for the lower-bound). Therefore, the upper-bound holds for all k -wise independent graphs with $k \geq S^* + 2$, and the lower-bound holds whenever $k \geq 2^{\binom{S^*}{2}}$. ■

We next provide our negative results. Since the complexity of known constructions of k -wise independent variables critically depend on the length, $\ell(p)$, of the binary representation of

$p = 0.b_1\dots b_\ell$, it is reasonable to focus on densities with bounded length. The argument used here was already applied in the context of Theorem 6.

Theorem 8 (K-wise independent graphs with huge independent sets) *Let $S, k : \mathbb{N} \rightarrow \mathbb{N}^+$ and $p : \mathbb{N} \rightarrow (0, 1)$ satisfy $S(N) \ll N^{\left(\frac{1}{k(N)+1}\right)}$ and $\ell(p(N)) \leq 2 \log(S(N))$. Then there exist k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ that a.s. contain independent sets of size $S(N)$.*

Proof. By Lemma 1, since $\ell(p) \leq 2 \log S$, we get $\left(\binom{S}{2}, k, p\right)$ -variables s.t. the probability that all variables receive value 0 is $\Delta \geq S^{-2k}$. From this, Lemma 5 gives k -wise independent graphs that a.s. contain independent sets of size S , whenever $S^2 \ll \Delta N^2$. ■

Corollary 1 *Let (S, k, p) be as in Theorem 8, with $\Omega(1) \leq p(N) \leq 1 - N^{-o(1)}$, and with S^* as in observation 3. Fix $c > 1$. Then there exist k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ where $k(N) \geq (1 - o(1)) \frac{\log(N)}{c \log \log(N)}$ that a.s. contain independent sets of size $S \gg (S^*(N))^c$.*

Proof. It suffices to provide an integer S s.t. : (i) $S \gg (S^*)^c$ (the desired outcome) and (ii) $S \ll N^{\left(\frac{1}{k+1}\right)}$ (the sufficient condition for applying Theorem 8). Such S clearly exists as long as

$$(S^*)^c \ll N^{\left(\frac{1}{k+1}\right)}. \quad (4)$$

Define r by $S^* = N^r$. Since $N^{\frac{1}{\log(N)}} = 2$, then any choice of $f(N) \gg 1$ gives $N^{\frac{f(N)}{\log(N)}} \gg 1$. Thus, equation (4) translates to having (for some $f(N) \gg 1$)

$$cr \leq \frac{1}{k+1} - \frac{f(N)}{\log(N)}. \quad (5)$$

Since p is bounded from 0, then $S^* \leq O(\log(N))$ so (again using $N^{\frac{1}{\log(N)}} = 2$)

$$cr = \frac{c \log(S^*)}{\log(N)} \leq \frac{c \log \log(N) + O(1)}{\log(N)}.$$

All this is valid in particular when $1 \ll f(N) \ll \log \log(N)$, so equation (5) becomes

$$\frac{1}{k+1} \geq cr + \frac{f(N)}{\log(N)} = (1 + o(1)) \frac{c \log \log(N)}{\log(N)}. \quad \blacksquare$$

The following upper bound for the independence number is larger than the bound of observation 3, yet holds for significantly smaller densities p .

Theorem 9 (Independence number upper bound) *There exist constants c_1, c_2 s.t. for any k -wise independent graphs $\{\mathcal{G}^k(N, p)\}_{N \in \mathbb{N}}$ the following a.s. holds. There are no independent sets of size S whenever either:*

1. $S \gg p^{-1/2} N^{3/4}$, $k \geq 4$ and $p \gg N^{-1/2}$; or
2. $S \geq c_1 \sqrt{\frac{N}{p}}$, $k \geq \log(N)$ and $p \geq \frac{c_2 \log(N)}{N}$.

Proof. By Theorem 1, α -jumbledness is a.s. achieved. For item 1 we have $\alpha \gg \sqrt{p} N^{3/4}$. For item 2 we have $\alpha = O(\sqrt{pN})$. Then, any vertex set U satisfies $e(U) \geq p|U|^2 - \alpha|U|$, so if U is independent, then $|U| \leq \alpha/p$. ■

8 Appendix - k-wise independence guarantees optimal jumbledness

This appendix is dedicated to proving Theorem 1. Given an N -vertex graph G , consider the complete graph \bar{G} , with weight $1 - p$ on any edge that appears in G , and weight $-p$ on any other edge and on any self loop. Let $A = A(\bar{G})$ denote the corresponding $N \times N$ matrix where $A_{u,w} = 1 - p$ whenever u, w are adjacent in G and $A_{u,w} = -p$ otherwise (including the case $u = w$).

Let $\lambda = \lambda(\bar{G})$ denote the largest eigenvalue in absolute value of A . By the argument in [2] G is λ -jumbled. Indeed, for any two sets of vertices U and W , if we let x_U and x_W denote the characteristic vectors of U and W , respectively, then $x_U^t A x_W = e(U, W) - p|U||W|$ and as the ℓ_2 -norm of $A x_W$ is at most $\lambda\sqrt{|W|}$, and that of x_U is $\sqrt{|U|}$, it follows by Cauchy-Schwarz that $|e(U, W) - p|U||W|| = |x_U^t A x_W| \leq \lambda\sqrt{|U||W|}$, as needed.

Let $\Gamma = (v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_R = v_0)$ be an arbitrary closed walk with R steps in \bar{G} . Throughout, Γ may repeat vertices and edges and may traverse self-loops. Let $W(\Gamma) = \prod_{j=0}^{R-1} A_{v_j, v_{j+1}}$, and let $X = \sum_{\Gamma} W(\Gamma)$.

By Wigner's trace argument [43], for any graph distribution, and any even $R \geq 4$ and $\omega \gg 1$, a.s.

$$\lambda \leq (\omega \mathbb{E}(X))^{1/R} \quad (6)$$

($\mathbb{E}(\cdot)$ stands for expectation). Thus, establishing the desired jumbledness reduces to bounding $\mathbb{E} = \mathbb{E}(X)$. We first fix t, j , and bound the contribution to \mathbb{E} of a single walk, Γ , that traverses exactly t vertices and j edges; later we bound the number of walks with such t, j . Let $\{e_1, \dots, e_j\}$ denote the set of all edges (excluding self-loops) used by Γ , where e_i is traversed precisely $q_i \geq 1$ times (we don't care how many times $e_i = \{u, w\}$ is traversed specifically from u to w or from w to u). As long as $k \geq R$ then the contribution of Γ to \mathbb{E} is bounded by $E(\Gamma) = \prod_{i=1}^j [p(1 - p)^{q_i} + (1 - p)(-p)^{q_i}]$. The latter equals 0 if some $q_i = 1$, so we focus on walks where each e_i is traversed at least twice. Then,

$$E(\Gamma) \leq \prod_{i=1}^j [p(1 - p)^2 + (1 - p)(-p)^2] < p^j. \quad (7)$$

Proving Theorem 1, item 1. Let $k = R = 4$. There are only 2 types of walks: Walks with 3 vertices contribute $O(p^2 N^3)$ to \mathbb{E} , and walks with 2 (or 1) vertices contribute only $O(pN^2)$, which is dominated by the 3-vertex walks' contribution. By (6), for any $\omega \gg 1$ a.s. $\lambda \leq (\omega(p^2 N^3))^{1/4}$. ■ (Item 1)

Proving Theorem 1, item 2. We adopt the approach of Füredi-Komlos-Vu [20, 42], who bound the number of walks with given (t, j) , by encoding the walks in a 1:1 manner, and then bound the number of code-words. We first describe their encoding scheme (Section 8.1) and later refine it (Section 8.2).

8.1 The Füredi-Komlos-Vu encoding

Fix Γ and consider the spanning-tree T of Γ , which consists of all the vertices visited by Γ and exactly those edges through which Γ visits a vertex for the first time. Edges (and consequently,

steps) in Γ are either **internal** ($e \in T$), or **external** ($e \notin T$). A step leading to a new vertex is called **positive**. A step traversing an internal edge for the 2nd time is called **negative**. Any other step is called **neutral** (thus, all (+) steps are internal, and neutral steps are either external, or pass through some internal edge for the i 'th time $i \geq 3$. Steps on self-loops are external). The encoding of Γ is composed of:

- A list of all t vertices visited by Γ , ordered by their first appearance.
- A string of length R , where the i 'th position encodes the i 'th step as follows.
 - Each positive step is encoded by (+).
 - Each negative step is encoded by (-).
 - Each neutral step ($u \rightarrow v$) is encoded by (v).

How is Γ retrieved from its encoding? The starting vertex is known, since the order in which the vertices appear in Γ is known. Assuming that the current position in the walk is known, then the next position is also known if the next step is either neutral or positive. Ambiguity is possible only when we are about to traverse a (-) step, and in addition the walk is currently at a **critical** vertex x . This means that the number of internal edges e_1, \dots, e_d that touch x , and have been traversed exactly once (up to this point) is ≥ 2 . For example, consider a walk starting with $1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 4 \rightarrow 5 \rightarrow 1$. At this point, $x = 1$ is critical since both edges $e_1 = \{1, 2\}$ and $e_2 = \{1, 4\}$ were traversed exactly once. If a (-) step immediately follows, it is not clear to which e_j this current (-) refers. The encoding in [20, 42] is modified in some way s.t. critical steps can be decoded un-ambiguously and the entire encoding-scheme becomes 1:1, as desired. By Theorem 1.5 in [42] this suffices for proving Item 2 in our theorem whenever $\Omega(\frac{\log^4(N)}{N}) \leq p \leq 1 - \Omega(\frac{\log^4(N)}{N})$. ■

For smaller p , we must refine the original encoding significantly.

8.2 Our refined encoding

We start with a simple observation. Throughout the analysis we set $k = R = \log(N)$, and let ℓ count the number of external edges in Γ . Let $\Phi_t, \Phi^\ell, \Phi_t^\ell$, resp. denote the contribution to \mathbb{E} of all walks with exactly t vertices, or exactly ℓ external edges, or exactly t vertices and ℓ external edges. Clearly, $\mathbb{E} = \sum_t \Phi_t$, $\mathbb{E} = \sum_\ell \Phi^\ell$, and $\mathbb{E} = \sum_{t,\ell} \Phi_t^\ell$. Since any of these sums has $\text{poly}(\log(N))$ summands, and since $R = \log(N)$ then $(\mathbb{E})^{1/R} = (1 + o(1))\Phi^{1/R}$ whenever Φ bounds the maximal term among all $\Phi_t, \Phi^\ell, \Phi_t^\ell$. It therefore suffices to show that $\Phi \leq (O(\sqrt{pN}))^R$.

We now give a high level description of our improved analysis. We keep 3 ingredients from [20, 42]: (i) The entire (ordered) list of vertices is provided. This contributes a $\Theta(N^t)$ multiplicative term to the bound on \mathbb{E} . (ii) Specific steps are encoded by a symbol from a fixed alphabet (the original alphabet is $\{+, -, \text{neutral}\}$; our final alphabet will be slightly larger). This contributes a $(\Theta(1))^R$ multiplicative term to \mathbb{E} . (iii) Since Γ traverses at least $t - 1$ edges, equation (7) bounds the contribution of each walk to \mathbb{E} by p^{t-1} . Thus, the combined contribution of (i)(ii) and (iii) to the bound on λ becomes (after taking the R 'th-root)

$$\Theta((pN)^{t/R}). \tag{8}$$

The latter partial encoding of (i)+(ii) is not 1:1, because of the neutral and the critical steps. Recall that there are $(t - 1)$ edges in T , and since (as mentioned above) all edges are traversed at least twice, then Γ has exactly $(t - 1)$ (+) steps, exactly $(t - 1)$ (-) steps, and $m = R - 2(t - 1)$ neutral steps. In [20, 42], the trivial t^m bound is used for the contribution of the neutral steps to the total number of code-words. We strengthen [20, 42], mainly by the following observations. First, we note that whenever ℓ is ‘very large’, then the entire contribution of all such paths to λ is negligible. Next (and perhaps most significantly), we show that whenever ℓ is not ‘very large’, the following holds: (i) Half of the neutral steps can be encoded very economically, reducing the t^m term from [20, 42] into roughly $t^{0.5m}$. (ii) All critical steps can be encoded so economically that their entire contribution is (almost) dominated by that of the neutral steps. Consequently, as $t \leq O(pN)$, we conclude that λ is bounded by (roughly) $\Theta(pN)^{\frac{t}{R}} \times O(pN)^{\frac{m}{2R}}$ (the first term stems from equation (8), the second from the neutral steps). Since $\frac{m}{2R} \sim 0.5 - \frac{t}{R}$, the latter bound becomes $O(\sqrt{pN})$, as desired. Details follow.

Handling the ‘non-typical’ walks (large ℓ). Clearly, the contribution of all (t, ℓ) -walks to \mathbb{E} is bounded by $B = p^\ell (pN)^{t-1} t^R N$. Indeed, $N^t t^R$ clearly bounds the number of walks, and by equation (7) the contribution of each walk is bounded by $p^{(t-1)+\ell}$. Let $\ell \geq 4 \log \log(N)$. Since $t - 1 \leq 0.5R$, $R = \log(N)$, and $pN \leq O(\log^4(N))$, we have $(pN)^{t-1} \leq \log(N)^{(4+o(1))(0.5 \log(N))} = \log(N)^{(2+o(1)) \log(N)}$, $N = \log(N)^{o(\log(N))}$, $t^R < \log(N)^{\log(N)}$, and $p^\ell \leq (N^{1-o(1)})^{4 \log \log(N)} = \log(N)^{(-4+o(1)) \log(N)}$. Consequently $B \leq \log(N)^{(-4+2+1+o(1)) \log(N)} \ll 1$. This concludes the treatment of the non-typical walks.

Handling the ‘typical’ walks (small ℓ). A new encoding is required to handle the typical walks. As before, the encoding includes the names of all t vertices, ordered by their first appearance, and all (+) and all non-critical (-) steps are simply encoded by (+) and (-) and decoded trivially. Our new perspective is thinking of the entire walk as composed of sequences of internal steps separated by external steps. We first encode the external steps economically, and prove that this enables to economically encode the critical steps as well. Next, we handle the internal sequences. We first provide some general observations regarding arbitrary internal sequences. Then, we describe how to handle the specific case of encoding a closed internal sequence. Finally, we generalize the latter to encoding open internal sequences as well.

8.2.1 Encoding external steps.

To exploit the small number of external edges, we add the following to the code.

- A list of all ℓ external edges e_1, \dots, e_ℓ .
- Each external step on e_i is encoded by (i, d) , where the bit d specifies the direction in which e_i is traversed.

Thus, encoding a single external step has only $2\ell = \Theta(\log \log(N))$ possible values. This improves upon [20, 42] where such steps are encoded by their end-vertex which might have $\Theta(\log(N))$ possible values.

8.2.2 Encoding critical steps.

Recall that a step \bar{s} is critical in Γ , if \bar{s} is taken from a vertex x , s.t. that x has $d \geq 2$ **critical edges** for \bar{s} . Critical edges are internal edges e_1, \dots, e_d that touch x and have been traversed exactly once up to \bar{s} . We will show that each e_i can be associated with a unique external edge e . This will enable us to encode \bar{s} using e which has only $\Theta(\log \log(N))$ possible values. Specifically, consider any critical edge $e_i = \{x, w\}$ for \bar{s} which is not the first edge leading to x in Γ . Consider the step s_i where e_i is traversed for the first time in Γ . Since x had already appeared in Γ , then $s_i = (x \rightarrow w)$. If we omit e_i from T , we partition T into 2 disjoint sub-trees: T_1 which contains x and T_2 which contains w . Since e_i is critical then the first time we return to x (after s_i), is not via $(w \rightarrow x)$. Thus, there must exist some external edge e (that connects T_2 to T_1) that is traversed between s_i and the first time we return to x . We call the first of these edges e the **external criticality edge (ECE)** of e_i , and denote it by $c(e_i, \bar{s})$. Clearly, different e_i -s have distinct ECEs, and in addition, at step \bar{s} all ECEs are well defined by previous steps in Γ . Consequently, the following encoding is un-ambiguous.

- Let \bar{s} be a critical step from x , with critical edges e_1, \dots, e_d , and external criticality edges $c(e_i, \bar{s})$. If \bar{s} traverses the first edge that leads to x in Γ we encode \bar{s} by $(-)$. Otherwise, we encode \bar{s} by the position of $c(e_i, \bar{s})$ in the list of external edges.

Note that not-critical negative steps are encoded by $(-)$ as before. This concludes the treatment of external and critical steps. We finally encode neutral internal steps.

8.2.3 Encoding an arbitrary sequence of internal steps.

Let $S = (s_1, \dots, s_q)$ be a ‘maximal’ sequence of internal steps. Here maximality means that (i) either the step previous to s_1 was external or that s_1 is the first step in the entire walk, and that (ii) either the next step after s_q is external, or that s_q is the last step in the entire walk (maximality does not mean that there are no longer internal sequences S' in Γ). We remark that, in general, some of the edges used by S may have been traversed before S started and some may be introduced by S for the first time.

Let x be the starting vertex of S . Fix some vertex $w \neq x$ visited by S and let $u = u(w)$ denote the **predecessor** of w in S (so the first time we reach w in S is via $(u \rightarrow w)$). Clearly, after each time we step $u \rightarrow w$, then the only way to return to u is by stepping $w \rightarrow u$ (otherwise we get a cycle from u to itself in T). Thus, when we pass e for the j 'th time during S we go forward $(u \rightarrow w)$ when j is odd and go backward $(w \rightarrow u)$ when j is even. We call this the **forward-backward observation**. Since the predecessor is uniquely determined by previous steps in Γ , we can encode backward steps very economically.

- A neutral-backward step is (economically) encoded by (nb) .
- A neutral-forward step $(u \rightarrow w)$ is (explicitly) encoded by (nf, w) .

Given this, we desire to demonstrate that many of the neutral steps in S go backward. We first handle the following simple case.

8.2.3.1. Encoding a closed sequence. Assume S is closed, namely, the end vertex of s_q is the starting vertex, x , of s_1 . We claim that at least half the neutral steps in S go backward. We actually prove the latter for every edge e in S . Let $\#f(e)$ and $\#b(e)$ resp. denote the number of forward and backward steps on an arbitrary edge e during S . Note that currently, not only neutral but also (+) and (-) steps are counted. We show that $\#f(e) = \#b(e)$. Indeed, otherwise, by the forward-backward observation the last step on e was a forward step $s_i = (u \rightarrow w)$, and clearly there exists a path from x to u in T . However, since S is closed there must exist another path in T from w to x that avoids stepping $(w \rightarrow u)$ - a contradiction. Now, let $\#nf(e)$ and $\#nb(e)$ count the number of neutral-forward and neutral-backward steps on e during S . By the above, there are at least 2 steps on e . There are 3 cases: (i) If e was never used prior to S the first step is $(+f)$, the second is $(-b)$. The next steps (if any exist) come in pairs of $(nf)(nb)$. (ii) If e was used at least twice prior to S , then all steps come in pairs of $(nf)(nb)$. (iii) If e was traversed exactly once prior to S , the first 2 steps are $(-f)(nb)$, and all consequent steps (if any exist) come in pairs of $(nf)(nb)$. In cases (i),(ii) we get $\#nf(e) = \#nb(e)$ and in case (iii) $\#nf(e) = \#nb(e) - 1$. Anyway, at least half of the neutral steps in a closed-internal sequence can be encoded economically. This concludes our analysis for closed sequences.

The problem is that for open sequences, S , it might hold that all $m(S)$ neutral steps in S are forward, and by the [20, 42] encoding-scheme these steps contribute a huge $t^{m(S)}$ factor to the bound on the number of code words. To overcome this, we use the following.

8.2.3.2. Encoding an open sequence. Let $x \neq y$ denote the start-vertex and end-vertex of some open maximal internal sequence S . Clearly, in T there exists a unique path $P = (x = x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_r = y)$. All the steps in S can be uniquely partitioned into 2 categories. (i) Steps that traverse P (either forward $(x_i \rightarrow x_{i+1})$ or backward $(x_{i+1} \rightarrow x_i)$). (ii) Entire sub-sequence S' , where each S' starts and ends at some path-vertex x_i , but never touch P at any other vertex other than x_i . Such S' is a closed internal sequence and is encoded as discussed in Section 8.2.3.1. We first modify the encoding simply s.t. path steps are explicitly encoded as such.

- Each positive-path step is encoded by $(+p)$.
- Each forward-neutral-path step is encoded by (npf) .
- Each backward-neutral-path step is encoded by (npb) .
- Each forward-negative-path step is encoded by $(-pf)$ (if the step \bar{s} is critical and has an external criticality edge, the index of this edge in the list of external edges is added to the encoding of \bar{s} as in Section 8.2.2).
- Each backward-negative-path step is encoded by $(-pb)$.

Clearly, **if** the entire path P is known, then the latter encoding suffices to decode any path-step, because on a path there is a unique forward-step and unique backward-step from each vertex. The question is how to recover the path itself. First, the end vertex y of P is well defined by the encoding. Indeed, if S is immediately followed by an external step \bar{s} , then the encoding of \bar{s} determines y . Otherwise, y is the last vertex in Γ , which is the (already known) first vertex of Γ . To recover the remaining vertices in P , we call a vertex on P either **old** or **new** according to

whether it appeared in Γ prior to S or not. If all vertices are old, since we know x and y , and since there is a unique path connecting x to y in T , then P is uniquely defined. Otherwise, some vertices in P are new. We claim that no new vertex is followed in P by an old vertex. Otherwise, the path P includes a step $x_i \rightarrow x_{i+1}$ where x_i is new but x_{i+1} is old. This means that before S started there was a path in T from x_{i+1} to x_1 (at each point in the walk, there exists a unique sub-tree of T that spans all the vertices traversed so far). Thus, P closes a cycle in T from x_{i+1} to itself - a contradiction. Therefore, if there are any new vertices there exists a unique **final old** vertex \bar{x} along P . If \bar{x} is known, then the path from x to \bar{x} is unique (since there is a unique path between any vertex-pair in T). In this case, the other part of P from \bar{x} to y is also well defined, because it consists only of new vertices (recall that the order in which new vertices appear in Γ is explicitly encoded). This covers all possible cases. Note that actually, if all steps on P are (+) and (-), then they are already uniquely decodable as before. Thus, the only addition required for decoding path steps is:

- Let S be an open internal sequence, with path P that contains at least a single new vertex and at least a single neutral step. Let \bar{x} be the final old vertex in P . Then the symbol \bar{x} is added to the encoding of the first neutral path step in S .

The main benefit here is that instead of encoding the end-point of each forward neutral step, it suffices to encode once the entire ‘direction’ of the path (this approach is similar to the [20, 42] encoding of critical steps).

Wrapping up. By all the above, the final encoding (including all aforementioned modifications) is 1:1 as desired. We currently fix any $\ell < 4 \log \log(N)$ and bound the contribution $E_{t,\ell}$ of all (t, ℓ) -walks to \mathbb{E} . Specifically, we bound the contribution of various parts in our encoding to $E_{t,\ell}$. Each contribution introduces a new multiplicative term to $E_{t,\ell}$. As before,

- Choosing the (list of ordered) t vertices to appear in Γ contributes $(N)_t = \Theta(N^t)$.
- The basic encoding of each step as some combination of positive/negative/neutral path/non-path forward/backward contributes $(\Theta(1))^R$.
- The contribution of each single walk Γ is $\Theta(p^{t-1+\ell})$.

We now consider the critical and neutral steps. Recall m is the total number of neutral steps. Let m_1 count the number of external steps. Let m_2 count the neutral steps in closed internal sequences. Let m_3 count the number of open internal sequences S s.t. their path $P = P(S)$ contains at least a single neutral step and at least a single new vertex.

- Choosing the ℓ external edges to appear in Γ contributes $\binom{t^2}{\ell} = O(t^{2\ell})$.
- By Section 8.2.1 encoding the external steps contributes $(2\ell)^{m_1}$.
- By Section 8.2.2 encoding the critical steps contributes at most $(\ell + 1)^{t-1}$.
- By Section 8.2.3.1 encoding the neutral steps in closed internal sequences contributes at most $t^{0.5m_2}$.

- By Section 8.2.3.2 encoding the neutral steps on the paths of open internal sequences contributes at most t^{m_3} .

Recall that to bound λ we are about to take the R 'th root of \mathbb{E} , and that we are willing to tolerate small Θ factors in the bound on λ . Since there are only $(\log(N))^{\Theta(1)}$ possible t, ℓ, m_1, m_2, m_3 , and since $(\log(N))^{\Theta(\frac{1}{R})} = 1 + o(1)$, then we may consider only the choice of t, ℓ, m_1, m_2, m_3 that maximizes the bound (on the contribution to \mathbb{E}). In addition, we may (i) Ignore the $(O(1))^R$ factor from encoding specific steps as a combination of $\{+, -, n, p, f, b\}$. (ii) Consider N^{t-1} instead of N^t (because $N^{\frac{1}{R}} = 2$). (iii) Ignore the $t^{2\ell}$ factor from the choice of external edges (since $t^{2\ell} < \log(N)^{8\log\log(N)} = 2^{o(R)}$). (iv) Replace the $(\ell + 1)^{t-1}$ term with $(\ell)^t$.

Combining all the remaining (un-ignored) terms yields the following expression

$$\Psi = (pN)^{t-1} t^{0.5m_2+m_3} p^\ell \ell^{m_1+t}.$$

As $m_1 + t < \log(N)$, and $p \geq N^{-1+o(1)}$, then $p^\ell \ell^{t+m_1} = 2^{-(1-o(1))\ell \log(N)} 2^{+\log(\ell) \log(N)} \leq 1$. Next, consider any open internal sequence S counted in m_3 . For any such S (except possibly the last one), there exists a unique neutral-external step that terminates S , so $m_3 \leq 0.5(m - m_1) + 1$. Thus, $0.5m_2 + m_3 \leq 0.5m + 1 = 0.5(R - 2(t - 1)) + 1$. Thus, $t < pN$ implies $\Psi < (pN)^{t-1} t^{0.5(R-2(t-1))+1} < (pN)^{0.5R+1}$. All the above gives $\lambda \leq \Theta(1)\Psi^{1/R} = \Theta(\sqrt{pN})$. ■