[Peter Allen](), [Julia Böttcher](), Hiệp Hàn, Yoshiharu Kohayakawa, and Yury Person

Powers of Hamilton cycles in pseudorandom graphs

Article (Accepted version)
(Refereed)

POWERS OF HAMILTON CYCLES IN PSEUDORANDOM GRAPHS

PETER ALLEN, JULIA BÖTTCHER, HIỆP HÀN, YOSHIHARU KOHAYAKAWA,
AND YURY PERSON

ABSTRACT. We study the appearance of powers of Hamilton cycles in pseudo-random graphs, using the following comparatively weak pseudorandomness notion. A graph G is $(\varepsilon, p, k, \ell)$-pseudorandom if for all disjoint X and $Y \subseteq V(G)$ with $|X| \geq \varepsilon p^k n$ and $|Y| \geq \varepsilon p^\ell n$ we have $e(X, Y) = (1 \pm \varepsilon) p |X||Y|$. We prove that for all $\beta > 0$ there is an $\varepsilon > 0$ such that an $(\varepsilon, p, 1, 2)$-pseudorandom graph on n vertices with minimum degree at least βpn contains the square of a Hamilton cycle. In particular, this implies that (n, d, λ)-graphs with $\lambda \ll d^{5/2} n^{-3/2}$ contain the square of a Hamilton cycle, and thus a triangle factor if n is a multiple of 3. This improves on a result of Krivelevich, Sudakov and Szabó [*Triangle factors in sparse pseudo-random graphs*, Combinatorica **24** (2004), no. 3, 403–426].

We also extend our result to higher powers of Hamilton cycles and establish corresponding counting versions.

1. INTRODUCTION AND RESULTS

The appearance of certain graphs H as subgraphs is a dominant topic in the study of random graphs. In the random graph model $G(n, p)$ this question turned out to be comparatively easy for graphs H of constant size, but much harder for graphs H on n vertices, i.e., *spanning* subgraphs. Early results were however obtained in the case when H is a Hamilton cycle, for which this question is by now very well understood [8, 21, 22, 23, 29].

When we turn to other spanning subgraphs H rather little was known for a long time, until a remarkably general result by Riordan [30] established good estimates for a big variety of spanning graphs H. In particular his result determines the threshold for the appearance of a spanning hypercube, and the threshold for the appearance of a spanning square lattice, as well as of the kth-power of a Hamilton cycle for $k > 2$. Here the *kth power* of H is obtained from H by adding all edges between distinct vertices of distance at most k in H. For the square of a Hamilton cycle the corresponding approximate threshold was only obtained recently by Kühn and Osthus [28].

Observe that the kth power of a Hamilton cycle contains $\lfloor n/(k + 1) \rfloor$ vertex disjoint copies of K_{k+1}, a so-called K_{k+1}-*factor*. It came as another breakthrough in the area and solved a long-standing problem when Johansson, Kahn and Vu [19] established the threshold for K_{k+1}-factors in $G(n, p)$ (or more generally of certain F-factors).

1.1. **Pseudorandom graphs.** Thomason [32] asked whether it is possible to single out some properties enjoyed by $G(n, p)$ with high probability that deterministically imply a rich collection of structural results that hold for $G(n, p)$. He thus initiated the study of pseudorandom graphs and suggested a deterministic property similar to the following notion of jumbledness. An n-vertex graph G is (p, β)-*jumbled* if

$$\left| e(A, B) - p|A||B| \right| \leq \beta \sqrt{|A||B|} \tag{1}$$

for all disjoint $A, B \subseteq V(G)$, where $e(A, B)$ is the number of edges in G with one endvertex in A and the other endvertex in B. The random graph $G(n, p)$ is with high probability (p, β)-jumbled with $\beta = O(\sqrt{pn})$, so this definition is justified. Moreover, this pseudorandomness notion indeed implies a rich structure (see, e.g., [11, 10, 14, 32]). However, for spanning subgraphs of general jumbled graphs (with a suitable minimum degree condition) not much is known.

One special class of jumbled graphs, which has been studied extensively, is the class of (n, d, λ)-graphs. Its definition relies on spectral properties. For a graph G with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ of the adjacency matrix of G, we call $\lambda(G) := \max\{|\lambda_2|, |\lambda_n|\}$ the *second eigenvalue* of G. An (n, d, λ)-*graph* is a d-regular graph on n vertices with $\lambda(G) \leq \lambda$. The connection between (n, d, λ)-graphs and jumbled graphs is established by the well-known *expander mixing lemma* (see, e.g., [7]), which states that if G is an (n, d, λ)-graph, then

$$\left| e(A, B) - \tfrac{d}{n}|A||B| \right| \leq \lambda(G) \sqrt{|A||B|} \tag{2}$$

for all disjoint subsets $A, B \subseteq V(G)$. Hence G is $\left(\tfrac{d}{n}, \lambda(G) \right)$-jumbled.

One main advantage of (n, d, λ)-graphs are the powerful tools from spectral graph theory which can be used for their study. Thanks to these tools various results concerning spanning subgraphs of (n, d, λ)-graphs G have been obtained. It turns out that already an almost trivial eigenvalue gap guarantees a spanning matching: if $\lambda \leq d - 2$ and n is even, then G has a perfect matching [26] (Stronger results in terms of other eigenvalues exist, see for example [12]). Moreover, if $\lambda \leq d(\log \log n)^2 / (1000 \log n \log \log \log n)$ then G has a Hamilton cycle [25]. The only other embedding result for spanning subgraphs of (n, d, λ)-graphs that we are aware of concerns triangle factors. Krivelevich, Sudakov and Szabó [27] proved that an (n, d, λ)-graph G with $3 \mid n$ and $\lambda = o\big(d^3/n^2 \log n\big)$ contains a triangle factor.

It is instructive to compare this last result with corresponding lower bound constructions. Krivelevich, Sudakov and Szabó also remarked that by using a blow-up of a construction of Alon [3] one can obtain for each $d' = d'(n')$ with $\Omega\big((n')^{2/3}\big) = d' \leq n'$ an (n, d, λ)-graph with $n = \Theta(n')$, $d = \Theta(d')$ and $\lambda = \Theta(d^2/n)$ which is triangle-free and thus contains no triangle factor. They conjectured that in fact (n, d, λ)-graphs are so symmetric that the upper bound on λ they proved for triangle factors can be improved, possibly all the way down to this lower bound. In this paper we bring the upper bound closer to the conjectured lower bound and establish more generally an embedding result for kth powers of Hamilton cycles (see Corollary 4).

1.2. **Our results.** The pseudorandomness notion we shall work with in this paper is weaker than that of (n, d, λ)-graphs, and in fact even weaker than jumbledness.

Definition 1. Suppose $\varepsilon > 0$ and $0 < p < 1$. Let k and ℓ with $k \leq \ell$ be positive integers. We call an n-vertex graph G $(\varepsilon, p, k, \ell)$-*pseudorandom* if

$$\left| e(X, Y) - p|X||Y| \right| < \varepsilon p|X||Y| \tag{3}$$

for any disjoint subsets $X, Y \subseteq V(G)$ with $|X| \geq \varepsilon p^k n$ and $|Y| \geq \varepsilon p^\ell n$.

It is easy to check that a graph which is $(p, \varepsilon^2 p^s n)$-jumbled is $(\varepsilon, p, k, \ell)$-pseudorandom for all k and ℓ with $k + \ell = 2s - 2$, but the jumbledness condition imposes tighter control on the edge density between (for example) linear sized subsets. An easy application of Chernoff's inequality and the union bound show that $G(n, p)$ is $(\varepsilon, p, k, \ell)$-pseudorandom with high probability if $p \gg (n^{-1} \log n)^{1/(\max\{k, \ell\} + 1)}$, while $G(n, p)$ only gets $(p, \varepsilon^2 p^{(k+\ell+2)/2} n)$-jumbled if $p \gg n^{-1/(k+\ell+1)}$. Our major motivation for using this weaker pseudorandomness condition is that it is all we require.

Our main result states that sufficiently pseudorandom graphs which also satisfy a mild minimum degree condition contain spanning powers of Hamilton cycles.

Theorem 2. *For every $k \geq 2$ and $\beta > 0$ there is an $\varepsilon > 0$ such that for any $p = p(n)$ with $0 < p < 1$ the following holds. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \beta p n$.*
  (a) *If G is $(\varepsilon, p, 1, 2)$-pseudorandom then G contains a square of a Hamilton cycle.*
  (b) *If G is $(\varepsilon, p, k - 1, 2k - 1)$-pseudorandom and $(\varepsilon, p, k, k + 1)$-pseudorandom then G contains a kth power of a Hamilton cycle.*

We remark that our proof of Theorem 2 also yields a deterministic polynomial time algorithm for finding a copy of the kth power of the Hamilton cycle. The proof technique (see Section 2.2 for an overview) is partly inspired by the methods used in [2] (which have similarities to those of Kühn and Osthus [28]).

It is immediate from the discussion above that our theorem implies the following result for jumbled graphs.

Corollary 3 (Powers of Hamilton cycles in jumbled graphs)**.** *For every $k \geq 2$ and $\beta > 0$ there is an $\varepsilon > 0$ such that for any $p = p(n)$ with $0 < p < 1$ the following holds. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \beta p n$.*
  (a) *If G is $(p, \varepsilon p^{5/2} n)$-jumbled then G contains a square of a Hamilton cycle.*
  (b) *If G is $(p, \varepsilon p^{3k/2} n)$-jumbled then G contains a kth power of a Hamilton cycle.*

As a consequence we also obtain a corresponding corollary for (n, d, λ)-graphs.

Corollary 4 (Powers of Hamilton cycles in (n, d, λ)-graphs)**.** *For all $k \geq 2$ there is $\varepsilon > 0$ such that for every (n, d, λ)-graph G,*
  (a) *if $\lambda \leq \varepsilon d^{5/2} n^{-3/2}$ then G contains a square of a Hamilton cycle,*
  (b) *if $\lambda \leq \varepsilon d^{3k/2} n^{1-3k/2}$ then G contains a kth power of a Hamilton cycle.*

In particular, under the conditions above, the graph G contains a spanning triangle factor and a spanning K_{k+1}-factor, respectively, if $3 \mid n$ and $(k + 1) \mid n$. Thus we improve on the result of Krivelevich, Sudakov and Szabó [27] for triangle factors and extend it to K_{k+1}-factors.

As remarked above even for $k = 2$ our upper bound for λ does not match the known lower bound. For $k > 2$ the situation gets even more complicated since 'good' lower bounds for the appearance of K_{k+1} (let alone kth powers of Hamilton cycles) in (n, d, λ)-graphs are not available. The best we can do is to observe that $G(n, p)$ with $(\log n/n)^{1/(k-\varepsilon)} \ll p \ll n^{-1/k}$ almost surely has no kth power of a Hamilton cycle, and that such a graph for any fixed $\varepsilon > 0$ is almost surely $(\varepsilon, p, k - 1 - \varepsilon, k - 1 - \varepsilon)$-pseudorandom.

1.3. **Counting.** Closely related to the question of the appearance of a certain subgraph in random or pseudorandom graphs is the question of how many copies of this subgraph are actually present. Janson [18], Cooper and Frieze [15], and Glebov and Krivelevich [17] studied this problem for Hamilton cycles in $G(n, p)$. Motivated by these results Krivelevich [24] recently turned to counting Hamilton cycles in

sparse (n, d, λ)-graphs G. He showed that for every $\varepsilon > 0$ and sufficiently large n, if $\lambda \leq d/(\log n)^{1+\varepsilon}$ and $\log \lambda \ll \log d - \log n / \log d$ then G contains $n!(d/n)^n (1+o(1))^n$ Hamilton cycles. This count is close to the expected number of labeled Hamilton cycles in $G(n, p)$ with $p = d/n$, which is $n!(d/n)^n$.

Krivelevich remarked that jumbled graphs may have isolated vertices and thus no Hamilton cycles at all. The same applies to our notion of pseudorandomness. If however, as in our main result, we combine this pseudorandomness with a minimum degree condition to avoid this obstacle, we do obtain a corresponding result concerning the number of Hamilton cycle powers in such graphs. Again, we obtain a count close to $p^{kn}n!$, which is the expected number of labeled copies of the kth power of a Hamilton cycle in $G(n, p)$. Note that (unlike Krivelevich) we do not provide a corresponding upper bound.

Theorem 5. *For every $k \geq 2$, β and $\nu > 0$ there is a constant $c > 0$, such that for every $\varepsilon = \varepsilon(n) \leq c/\log^2 n$ and $p = p(n)$ with $0 < p < 1$ the following holds. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \beta p n$. Suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom if $k = 2$, and $(\varepsilon, p, k-1, 2k-1)$-pseudorandom and $(\varepsilon, p, k, k+1)$-pseudorandom if $k > 2$. Then G contains at least $(1-\nu)^n p^{kn} n!$ copies of the kth power of a Hamilton cycle.*

With some minor modifications, this result follows from our proof of Theorem 2. For the sake of clarity, we sketch these modifications after detailing the proof of Theorem 2.

1.4. **Organisation.** The remainder of this paper is organised as follows. In Section 2 we give some basic definitions, outline our proof strategy, provide the main lemmas and use them to obtain Theorem 2. In Sections 3 and 4 we prove our three main lemmas. We sketch how to modify the proof of Theorem 2 to get Theorem 5 in Section 5, and close with some remarks and open problems in Section 6.

2. Main lemmas and proof of the main theorem

2.1. **Notation.** An s-tuple $(u_1, \ldots, u_s)$ of vertices is an ordered set of vertices. We often denote tuples by bold symbols, and occasionally also omit the brackets and write $\boldsymbol{u} = u_1, \ldots, u_s$. We write $V(\boldsymbol{u})$ for the set $\{u_1, \ldots, u_s\}$ of vertices in the tuple $\boldsymbol{u}$.

Given a graph H, the graph H^k, called the kth *power* of H, is the graph on $V(H)$ where two distinct vertices u and v are adjacent if and only if their distance in H is at most k.

For simplicity we also call the kth power of a path a k-*path*, and the kth power of a cycle a k-*cycle*. We will usually specify k-paths and k-cycles by giving the (cyclic) ordering of the vertices in the form of a vertex tuple. We say that the *start* s-tuple of a k-path $P = (u_1, \ldots, u_\ell)$ is $(u_s, \ldots, u_1)$, and the *end* s-tuple is $(u_{\ell-s+1}, \ldots, u_\ell)$ (the vertices $u_{s+1}, \ldots, u_{\ell-s}$ are said to be internal). In these definitions, we shall often have $s = k$.

For a given graph G let $N_X(x)$ be the set of neighbours of x in $X \subseteq V(G)$. For an ℓ-tuple $\boldsymbol{x}_\ell = (x_1, \ldots, x_\ell)$ of vertices let $N_X(x_1, \ldots, x_\ell)$ denote the common neighbourhood of $x_1, \ldots, x_\ell$ in X, and let $\deg_X(x_1, \ldots, x_\ell) = |N_X(x_1, \ldots, x_\ell)|$.

We say that $\boldsymbol{x}_\ell$ is (ϱ, p)-*connected* to a vertex set X if $x_1, \ldots, x_\ell$ forms a clique in G and

$$\deg_X(x_i, \ldots, x_\ell) \geq \varrho \left(\frac{p}{2}\right)^{\ell-i+1} |X| \tag{4}$$

for every $i \in [\ell] = \{1, \ldots, \ell\}$. To motivate this definition, note that the bound in (4) corresponds to the expected number of common neighbours of $(x_i, \ldots, x_\ell)$ in X in the random graph $G(n, p)$, up to a constant factor.

A vertex set $Y \subseteq X$ *witnesses* that $\boldsymbol{x}_\ell$ is (ϱ, p)-connected to X if for every $i \in [\ell]$ we have $\left| Y \cap N_X(x_i, \ldots, x_\ell) \right| \geq \varrho \left(\frac{p}{2} \right)^{\ell - i + 1} |X|$.

Remark 6. Since the sets $N_X(x_1, \ldots, x_\ell)$, $N_X(x_2, \ldots, x_\ell)$, ..., $N_X(x_\ell)$ are nested we have that if $\boldsymbol{x}_\ell$ is (ϱ, p)-connected to X, then there is a set $Y \subseteq X$ with $|Y| = \varrho p |X| / 2$ vertices which witnesses this connectedness.

In our proofs we shall additionally frequently make use of the following observation concerning our pseudorandomness notion.

Remark 7. If $0 < p \leq 1/2$ and $\varepsilon < 1/8$, and the n-vertex graph G is $(\varepsilon, p, k, \ell)$-pseudorandom, then G has a vertex y of degree at most $3n/4$. Furthermore, letting $X = V(G) \setminus (\{y\} \cup N(y))$ and $Y = \{y\}$ we see that the pseudorandomness condition (3) does not hold. It follows that $1 < \varepsilon p^\ell n$, or equivalently $p^\ell n > \varepsilon^{-1}$. A similar statement holds if $1/2 \leq p < 1$, taking $X = N(y)$. Thus assuming the n-vertex graph G to be $(\varepsilon, p, k, \ell)$-pseudorandom for any $0 < p < 1$ implicitly means we assume $p^\ell n > \varepsilon^{-1}$.

2.2. Outline of the proof.

Suppose that G is an $(\varepsilon, p, k-1, k)$-pseudorandom graph on n vertices. One crucial observation, which forms the starting point of our proof, is that it is relatively easy to find an almost spanning k-path in G. Indeed, it is not hard to check (see the Extension lemma, Lemma 8) that G contains copies of K_k and that typically such a K_k-copy is well-connected to the rest of the graph in the following sense. There are many vertices which extend this K_k-copy to a k-path on $k+1$ vertices. Iterating this argument we can greedily build a k-path P' covering most of G. Let L be the set of leftover vertices.

Thus, the true challenge is to incorporate the few remaining vertices into P' and to close P' into a k-cycle. To tackle the second of these tasks we will establish a Connection lemma (Lemma 12), which asserts that any two pairs of k-cliques in G which are sufficiently well-connected to a set U of vertices can be connected by a short k-path with interior vertices in U. At this point, if $k > 2$, we shall need to require that G be $(\varepsilon, p, k-1, 2k-1)$-pseudorandom.

For the first task, we make use of the *reservoir method* developed in [2] (see also [28] for a similar method). In essence, the fundamental idea of this method is to ensure that P' contains a sufficiently big proportion of vertices which are free to be taken out of P' and used otherwise. More precisely, we shall construct (see the Reservoir lemma, Lemma 10) a path P with the *reservoir property*: There is a subset R of $V(P)$, called the *reservoir*, such that for any $W \subseteq R$ there is a k-path in G whose vertex set is $V(P) \setminus W$ and whose ends are the same as those of P. We also call P a *reservoir path*. We then use the greedy method outlined above to extend P to an almost spanning k-path P'. For this step, if $k > 2$, we shall need to require that G be $(\varepsilon, p, k, k+1)$-pseudorandom.

With the reservoir property we are now in good shape to incorporate the leftover vertices L into P' (and then close the path into a cycle): We show, using a Covering lemma (Lemma 11), that we can find a k-path P'' in $L \cup R$ covering all vertices of L and using only a small fraction of R (this is possible because R is much bigger than L). Finally we connect both ends of P' and P'' using some of the remaining vertices of R with the help of the Connection lemma (again, this is possible because many vertices of R remain).

Now the only problem is that some vertices of R may be used twice, in P' and in P'' or the connections. But this is where the reservoir property comes into play. This property asserts that there is a k-path $\widetilde{P}$ which uses all vertices of P' except these vertices. Finally $\widetilde{P}$ and P'' together with the connections form the desired spanning k-cycle.

2.3. **Main lemmas.** The proof of Theorem 2 relies on four main lemmas, the Extension lemma, the Reservoir lemma, the Covering lemma and the Connection lemma, which we will state and explain in the following.

Our first lemma, the Extension lemma, states that in a sufficiently pseudorandom graph all well-connected k-tuples have a common neighbour which together with the last $k-1$ vertices of this k-tuple form again a well-connected k-tuple.

Lemma 8 (Extension lemma). *Given $k \geq 2$ and $\delta > 0$ there is an $\varepsilon > 0$ such that for all $0 < p < 1$, all $(\varepsilon, p, k-1, k)$-pseudorandom graphs G on n vertices, and all disjoint vertex sets L and R with $|L|, |R| \geq \delta n$ the following holds.*

Let $\boldsymbol{x} = (x_1, \ldots, x_k)$ be a k-tuple which is $(\frac{1}{8}, p)$-connected to both L and R. Then there is a vertex x_{k+1} of $L \cap N(x_1, \ldots, x_k)$ such that $(x_2, \ldots, x_{k+1})$ is $(\frac{1}{6}, p)$-connected to both L and R.

We stress that in this lemma we require and obtain well-connectedness to two sets L and R. This will enable us in the proof of Theorem 2 to extend a k-path alternatively using vertices of the leftover set L or the reservoir set R.

We remark moreover that the assumed $(\frac{1}{8}, p)$-connectedness is weaker than the $(\frac{1}{6}, p)$-connectedness in the conclusion. This is useful when we repeatedly apply the Extension lemma. It is possible to prove such a statement because the factor $\frac{1}{2}$ in the definition of connectedness allows for some leeway.

Since the proof of this lemma is short we give it straight away. We use the following lemma, which is a direct consequence of (3) and will frequently be used later as well.

Lemma 9. *If G is an $(\varepsilon, p, k, \ell)$-pseudorandom graph on n vertices and $X \subseteq V(G)$ satisfies $|X| \geq \varepsilon p^k n$, then less than $\varepsilon p^\ell n$ vertices $v \in V(G) \setminus X$ have $\deg_X(v) < (1-\varepsilon)p|X|$, and less than $\varepsilon p^\ell n$ vertices $v \in V(G) \setminus X$ have $\deg_X(v) > (1+\varepsilon)p|X|$.* $\square$

Proof of Lemma 8. Given k and δ we set

$$\varepsilon = \frac{\delta}{80 \cdot k \cdot 2^{k+3}}. \tag{5}$$

Because $\boldsymbol{x}$ is $(\frac{1}{8}, p)$-connected to L, for each $2 \leq i \leq k$ we have

$$\deg_L(x_i, \ldots, x_k) \geq \frac{1}{8} \left(\frac{p}{2}\right)^{k-i+1} |L|.$$

We claim that for each $2 \leq i \leq k$ there are less than $\varepsilon p^k n$ vertices which have less than $\frac{1}{6}(p/2)^{k-i+2}|L|$ neighbours in $N(x_i, \ldots, x_k) \cap L =: Y_i$. Indeed, we have $|Y_i| \geq \frac{1}{8}(\frac{p}{2})^{k-i+1}|L| \geq \frac{1}{8}(\frac{p}{2})^{k-i+1}\delta n > 10\varepsilon p^{k-1} n$. Now assume for contradiction that there is a set B_i of $\varepsilon p^k n$ vertices in $V(G)$ all of which have less than $\frac{1}{6}(p/2)^{k-i+2}|L| \leq \frac{2}{3}p|Y_i|$ neighbours in Y_i. Since $|B_i| \leq \frac{1}{10}|Y_i|$ and thus $|Y_i \setminus B_i| \geq \frac{9}{10}|Y_i| > \varepsilon p^{k-1} n$, this implies that each vertex in B_i has less than $\frac{2}{3}p|Y_i| \leq \frac{2}{3}p\frac{10}{9}|Y_i \setminus B_i| = \frac{20}{27}p|Y_i \setminus B_i|$ neighbours in $Y_i \setminus B_i$. This however contradicts Lemma 9 because G is $(\varepsilon, p, k-1, k)$-pseudorandom.

Similarly, less than $\varepsilon p^k n$ vertices have fewer than $\frac{1}{6}(p/2)|L|$ neighbours in L. The same calculations, replacing L with R, also hold. It follows that all but at most $2k\varepsilon p^k n$ vertices x_{k+1} of $N(x_1, \ldots, x_k) \cap L$ have the property that $(x_2, \ldots, x_{k+1})$ is $(\frac{1}{6}, p)$-connected to both L and R. Finally, since

$$\deg_L(x_1, \ldots, x_k) \geq \frac{1}{8}\left(\frac{p}{2}\right)^k |L| \geq \frac{\delta p^k n}{2^{k+3}} \overset{(5)}{>} 2k\varepsilon p^k n$$

there is indeed a vertex x_{k+1} with this property as desired. $\square$

Our second lemma allows us to construct the reservoir path P described in the outline, given a suitable reservoir R (see properties (a) and (d) of the lemma). In

addition, this lemma guarantees well-connectedness of the ends of this path to the reservoir and to the remaining vertices in the graph (see properties (b) and (c) of the lemma). This is necessary so that we can extend the reservoir path and later connect it to the path covering the leftover vertices L using R.

Lemma 10 (Reservoir lemma). *Given $k \geq 2$, $0 < \delta < 1/4$ and $0 < \beta < 1/2$ there exists an $\varepsilon > 0$ such that the following holds.*

Let $0 < p < 1$ and let $G = (V, E)$ be an n-vertex graph. Suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom if $k = 2$, and $(\varepsilon, p, k - 1, 2k - 1)$-pseudorandom and $(\varepsilon, p, k, k + 1)$-pseudorandom if $k > 2$. Let $R \subseteq V$ satisfy $\delta^2 n/(200k) \leq |R| \leq \delta n/(200k)$ and $\deg_{V \setminus R}(v) \geq \beta p n/2$ for all $v \in R$. Then there is a k-path P in G with the following properties.

(a) *$R \subseteq V(P)$, $|V(P)| \leq 50k|R|$, and all vertices from R are internal in P.*

(b) *The start and end k-tuples of P are $(\frac{1}{8}, p)$-connected to $V \setminus V(P)$.*

(c) *The start and end k-tuples of P are $(\frac{1}{2}, p)$-connected to R (and thus disjoint from R).*

(d) *For any $W \subseteq R$, there is a k-path with the vertex set $V(P) \setminus W$ whose start and end k-tuples are identical to those of P.*

Our third lemma enables us to cover the leftover vertices L with a k-path (see property (a)). This lemma allows us in addition to specify a set S to which the start and end tuples of this path have to maintain well-connectedness (see property (b)). When we cover the leftover vertices in the proof of the main theorem, S will be a big proportion of R and we will use the well-connectedness to connect the path covering L and the extended reservoir path.

Observe that the requirements and conclusions of Lemma 10 and Lemma 11 overlap substantially. In fact, we shall prove both lemmas together in Section 4.

Lemma 11 (Covering lemma). *Given $k \geq 2$, $0 < \delta < 1/4$ and $0 < \beta < 1/2$, there exists an $\varepsilon > 0$ such that the following holds.*

Let $0 < p < 1$ and let $G = (V, E)$ be an n-vertex graph. Suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom if $k = 2$, and $(\varepsilon, p, k - 1, 2k - 1)$-pseudorandom and $(\varepsilon, p, k, k + 1)$-pseudorandom if $k > 2$. Let L and S be disjoint subsets of $V(G)$ with $|L| \leq \delta n/(200k)$ and $|S| \geq \delta n$ such that $\deg_S(v) \geq \beta \delta p n/2$ for all $v \in L$. Then there is a k-path P contained in $L \cup S$ with the following properties.

(a) *$L \subseteq V(P)$ and $|V(P)| \leq 50k|L|$.*

(b) *The start and end k-tuples of P are in S and are $(\frac{1}{8}, p)$-connected to $S \setminus V(P)$.*

Our fourth and final main lemma allows us to connect two k-tuples with a short k-path.

Lemma 12 (Connection lemma). *For all $k \geq 2$ and $\delta > 0$ there is an $\varepsilon > 0$ such that the following holds.*

Let $0 < p < 1$ and let G be an n-vertex graph. Suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom if $k = 2$, and $(\varepsilon, p, k - 1, 2k - 1)$-pseudorandom if $k > 2$. Let $U \subseteq V(G)$ be a vertex set of size $|U| \geq \delta n$. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are two disjoint k-tuples which are (δ, p)-connected to U, then there exists a k-path P with ends $\boldsymbol{x}$ and $\boldsymbol{y}$ of length at most $7k$ such that $V(P) \subseteq U \cup V(\boldsymbol{x}) \cup V(\boldsymbol{y})$.

The proof of Lemma 12 can be found in Section 3. We remark that in the proof of Theorem 2 it is not especially important that the connecting k-path guaranteed by this lemma is of constant length. However, Lemma 12 is also used in the proof of Lemma 10, and in this proof we need that the connecting k-paths are of length independent of n.

2.4. **Proof of Theorem 2.** Using Lemmas 8, 10, 11 and 12 we can now prove our main theorem.

Proof. Given $k \geq 2$ and $0 < \beta < 1/2$, we set $\delta_{\mathrm{L}10} := \frac{1}{10}$, $\delta_{\mathrm{L}11} := \delta_{\mathrm{L}10}^2/(10^4 k)$, $\delta_{\mathrm{L}8} := \delta_{\mathrm{L}11}/(200k) \leq \delta_{\mathrm{L}10}^2/(200k)$ and $\delta_{\mathrm{L}12} = \frac{\beta}{16} \cdot \delta_{\mathrm{L}10}^2/(400k)$. We choose

$$\varepsilon \leq \frac{1}{7} \cdot \frac{\beta \delta_{\mathrm{L}10}^2}{6400 k^2 \cdot 2^k} \tag{6}$$

to be small enough to apply Lemma 8 with input k and $\delta_{\mathrm{L}8}$, to apply Lemma 10 with input k, $\delta_{\mathrm{L}10}$ and to apply Lemma 11 β and with input k, $\delta_{\mathrm{L}11}$ and β, and to apply Lemma 12 with input k and $\delta_{\mathrm{L}12}$.

Let $0 < p < 1$ and G be a graph on n vertices with minimum degree at least βpn. If $k = 2$, suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom. If $k \geq 3$, suppose that G is $(\varepsilon, p, k-1, 2k-1)$ and $(\varepsilon, p, k, k+1)$-pseudorandom. This ensures that we can apply Lemmas 10, 11 and 12.

Our first step now is to select an appropriate reservoir set.

Claim 13. *There is a set R, which we call reservoir set, such that*

  (i) $\delta_{\mathrm{L}10}^2 n/(200k) \leq |R| \leq \delta_{\mathrm{L}10} n/(200k)$,
  (ii) $\deg_R(v) \geq \frac{1}{2}\beta p |R|$ *for all $v \in V(G) \setminus R$ and*
  (iii) $\deg_{V(G) \setminus R}(v) \geq \frac{1}{2}\beta pn$ *for all $v \in R$.*

Proof. We start with an arbitrary set R' of $2 \cdot \delta_{\mathrm{L}10}^2 n/(200k)$ vertices. We remove from R' all vertices $v \in R'$ such that $\deg_{V(G) \setminus R'}(v) < 3\beta pn/4$ to obtain R''. Let R be obtained from R'' by adding all vertices v of $V(G) \setminus R''$ such that $\deg_{R''}(v) < \beta p |R''|$.

We first show that R satisfies property (i), by using that G is in particular $(\varepsilon, p, 0, 1)$-pseudorandom (For this proof we will require no stronger pseudorandomness). Since $|V(G) \setminus R'| > 3n/4 > \varepsilon n$ and $3\beta pn/4 \leq (1-\varepsilon)p(3n/4) < (1-\varepsilon)p|V(G) \setminus R'|$ we infer from Lemma 9 that

$$|R' \setminus R''| < \varepsilon pn. \tag{7}$$

Thus, clearly $|R' \setminus R''| < \delta_{\mathrm{L}10}^2 n/(200k)$, and hence $|R| \geq |R''| > \delta_{\mathrm{L}10}^2 n/(200k)$. Similarly $|R''| > \varepsilon n$ and $\beta p |R''| < (1-\varepsilon)p|R''|$ in conjunction with Lemma 9 implies that

$$|R \setminus R''| < \varepsilon pn, \tag{8}$$

and so $|R \setminus R''| \leq \delta_{\mathrm{L}10}^2 n/(200k)$ and hence $|R| = |R''| + |R \setminus R''| \leq |R'| + |R \setminus R''| \leq \varepsilon pn + 2 \cdot \delta_{\mathrm{L}10}^2 n/(200k) \leq \delta_{\mathrm{L}10} n/(200k)$. This yields property (i).

For (ii) observe that $|R \setminus R''| < \varepsilon pn$ and $|R| \geq \delta_{\mathrm{L}10}^2 n/(200k) \geq 2\varepsilon n$ implies $|R''| \geq \frac{1}{2}|R|$. Since $R'' \subseteq R$ we thus have by construction for each $v \in V(G) \setminus R$ that $\deg_R(v) \geq \beta p |R''| \geq \frac{1}{2}\beta p |R|$.

It remains to argue that R also satisfies (iii). By construction all vertices of R'' have at least $3\beta pn/4$ neighbours in $V(G) \setminus R'$, and thus by (7) at least $3\beta pn/4 - \varepsilon pn > \beta pn/2$ neighbours in $V(G) \setminus R$. All vertices of $R \setminus R''$, on the other hand, have at most $\beta p |R''| \leq \frac{1}{4}\beta pn$ neighbours in R'', and by (8) at most $\varepsilon pn \leq \frac{1}{4}\beta pn$ neighbours in $R \setminus R''$. Since $\delta(G) \geq \beta pn$, we conclude that every vertex of R has at least $\frac{1}{2}\beta pn$ neighbours in $V(G) \setminus R$. □

We now construct a reservoir path for this reservoir R by applying Lemma 10 with input k, $\delta_{\mathrm{L}10}$, β, p, G and R. Observe that this is possible by properties (i) and (iii) of Claim 13. Hence we obtain a k-path P in G which satisfies all four conclusions of Lemma 10. Let $\boldsymbol{u}$ be the start k-tuple of P, and $\boldsymbol{v}$ the end k-tuple. We conclude from (c) and (b) of Lemma 10 that $\boldsymbol{u}$ and $\boldsymbol{v}$ are $(\frac{1}{2}, p)$-connected to R and $(\frac{1}{8}, p)$-connected to $L_1 := V(G) \setminus V(P)$.

Our next step is to extend this reservoir path to an almost spanning k-path P' by repeatedly applying Lemma 8. For this purpose we let $t := |L_1| - \delta_{\mathrm{L11}}n/(200k)$ and apply Lemma 8 exactly t times with k, δ_{L8} and p to G. First we apply this lemma with sets L_1 and R, and the k-tuple $\boldsymbol{v} =: (v_1, \ldots, v_k)$. We obtain a vertex $v_{k+1} \in L_1 \cap N(v_1, \ldots, v_k)$ such that $(v_2, \ldots, v_{k+1})$ is $(\frac{1}{6}, p)$-connected to both L_1 and R. Let $L_2 := L_1 \backslash \{v_{k+1}\}$, and extend P by v_{k+1} to obtain $P_1 := (P, v_{k+1})$. Similarly for each $2 \leq i \leq t$ in succession we apply Lemma 8 with L_i, R and $(v_i, \ldots, v_{k+i-1})$ and obtain from this lemma an extending vertex v_{k+i} such that $(v_{i+1}, \ldots, v_{k+i})$ is $(\frac{1}{6}, p)$-connected to both L_i and R. We then let $L_{i+1} := L_i \backslash \{v_{k+i}\}$ and $P_i := (P_{i-1}, v_{k+i})$. We need to argue that these applications of Lemma 8 are possible. Indeed, by Claim 13 (i) and the choice of our constants we have $|R| \geq \delta_{\mathrm{L10}}^2 n/(200k) \geq \delta_{\mathrm{L8}}n$ and $|L_i| \geq |L_1| - t = \delta_{\mathrm{L11}}n/(200k) = \delta_{\mathrm{L8}}n$. Moreover, for $i > 1$ the k-tuple $(v_i, \ldots, v_{k+i-1})$ is $(\frac{1}{6}, p)$-connected to both L_{i-1} and R by construction. Since $|L_i| = |L_{i-1}| - 1$, the k-tuple $(v_i, \ldots, v_{k+i-1})$ is thus $(\frac{1}{8}, p)$-connected to L_i (for $i = 1$ the statement is guaranteed by Lemma 10 which constructed P).

What did we achieve so far? Let $P' := P_t$ and $L := V(G) \setminus V(P')$ be the set of leftover vertices at this point. Then

$$|L| \leq \delta_{\mathrm{L11}}n/(200k) \tag{9}$$

and by Claim 13 (ii) every vertex of L has at least $\frac{1}{2}\beta p|R|$ neighbours in R. By construction P' is a k-path extending the reservoir path P and covering all vertices of G but L. In addition, the start k-tuple $\boldsymbol{u}$ and end k-tuple $\boldsymbol{v}'$ of P' are both $(\frac{1}{8}, p)$-connected to R. Clearly this implies that these k-tuples are also $\left(\frac{\beta}{16}, p\right)$-connected to R, and in the following we will only work with this weaker conclusion.

Our next step will be to cover the leftover vertices L with a k-path P'' using the Covering lemma, Lemma 11. However, this needs some preparation. Recall that in Lemma 11 we can choose a vertex set S so that the k-path that this lemma constructs only uses vertices from L and S. As explained earlier we want to choose a big subset of the reservoir R as S. However, we need to bear in mind that we later want to connect the start $\boldsymbol{u}$ of P' and the end $\boldsymbol{v}''$ of P'' using only vertices from $U := R \setminus V(P'')$ with the help of the Connection lemma, Lemma 12 (similarly for the end $\boldsymbol{v}'$ of P' and the start $\boldsymbol{u}''$ of P''). But this lemma requires that $\boldsymbol{u}$ is well-connected to U. In order to guarantee this property we will now set aside a set $R_{\boldsymbol{u}} \subseteq R$ (and similarly a set $R_{\boldsymbol{v}'}$) of vertices which witness the well-connectedness of $\boldsymbol{u}$ to R and prevent these vertices from being used in P'' by setting $S = R \setminus (R_{\boldsymbol{u}} \cup R_{\boldsymbol{v}'})$.

More precisely, recall that the $\left(\frac{\beta}{16}, p\right)$-connectedness of $\boldsymbol{u}$ means that there is a set of $\frac{\beta}{16}(p/2)^k|R|$ common neighbours of $\boldsymbol{u}$ in R, a set of $\frac{\beta}{16}(p/2)^{k-1}|R|$ common neighbours of $(u_2, \ldots, u_k)$ in R, and so on. By Remark 6 there is a set $R_{\boldsymbol{u}}$ of $\frac{\beta}{16}(p/2)|R|$ vertices of R which witness that $\boldsymbol{u}$ is $\left(\frac{\beta}{16}, p\right)$-connected to R. Similarly, there is a set $R_{\boldsymbol{v}'}$ of $\frac{\beta}{16}(p/2)|R|$ vertices of R which witness that $\boldsymbol{v}'$ is $\left(\frac{\beta}{16}, p\right)$-connected to R. Moreover, the deletion of any set of at most $\frac{\beta}{32}(p/2)^k|R|$ vertices from $R_{\boldsymbol{u}}$ (or $R_{\boldsymbol{v}'}$) results in a set that still witnesses that $\boldsymbol{u}$ is $\left(\frac{\beta}{32}, p\right)$-connected to R.

Now let $S := R \setminus (R_{\boldsymbol{u}} \cup R_{\boldsymbol{v}'})$ and note that by part (i) of Claim 13 we have that

$$|S| \geq |R| - \frac{\beta}{16}p|R| \geq \frac{1}{2}|R| \geq \frac{\delta_{\mathrm{L10}}^2}{400k}n \geq \delta_{\mathrm{L11}}n. \tag{10}$$

Moreover, since every vertex of L has by Claim 13 (ii) at least $\frac{1}{2}\beta p|R|$ neighbours in R, we conclude from Claim 13 (i) that every vertex of L also has at least

$$\frac{\beta}{2}p|R| - \frac{\beta}{16}p|R| \geq \frac{7\beta}{16}p\frac{\delta_{\mathrm{L10}}^2}{200k}n \geq \frac{1}{2}\beta\delta_{\mathrm{L11}}pn \tag{11}$$

neighbours in S.

It follows from (9), (10) and (11) that we can apply Lemma 11 with input k, $\delta_{\mathrm{L}11}$, β, p, G, L and S. We obtain a k-path P'' with

$$\left|V(P'')\right| \le 50k|L| \overset{(9)}{\le} \frac{\delta_{\mathrm{L}11}n}{4} < \frac{1}{8}\cdot\delta_{\mathrm{L}10}^2\frac{n}{200k} \le \frac{1}{8}|R|\,,$$

which covers L and whose remaining vertices are in S. The start and end tuples $\boldsymbol{u}''$ and $\boldsymbol{v}''$ of P'' are $(\frac{1}{8},p)$-connected to $S\setminus V(P'')$, so $(\frac{1}{32},p)$-connected to $R\setminus V(P'')$ and in particular $(2\delta_{\mathrm{L}12},p)$-connected to $R\setminus V(P'')$. Let $R_{\boldsymbol{u}''}$ be a set of $\frac{1}{32}p|R\setminus V(P'')|$ vertices which witnesses the $(\frac{1}{32},p)$-connectedness of $\boldsymbol{u}''$ to $R\setminus V(P'')$.

It follows from the choice of $R_{\boldsymbol{u}}$ and $R_{\boldsymbol{v}'}$ that $\boldsymbol{u}$ and $\boldsymbol{v}'$ are $(\frac{\beta}{16},p)$-connected and hence $(2\delta_{\mathrm{L}12},p)$-connected to $R\setminus V(P'')$. Now we would like to apply Lemma 12 twice to connect the ends of P' and P'' such that the connections use vertices from $R\setminus V(P'')$. For this observe that $|R\setminus V(P'')| \ge \frac{7}{8}|R| \ge \frac{7}{8}\delta_{\mathrm{L}10}^2n/(200k) \ge 2\delta_{\mathrm{L}12}n$. Moreover, $\boldsymbol{u}$ and $\boldsymbol{v}''$ are both $(2\delta_{\mathrm{L}12},p)$-connected to $R\setminus V(P'')$. Hence we can apply Lemma 12 with k and $\delta_{\mathrm{L}12}$ to find a k-path C of length at most $7k$ connecting $\boldsymbol{u}$ and $\boldsymbol{v}''$ in $R\setminus V(P'')$. By Remark 7 we have $\varepsilon^{-1} < p^k n$ and hence we can use Claim 13 (i) to conclude that

$$|C| \le 7k \overset{(6)}{\le} \frac{1}{\varepsilon}\cdot\frac{\beta}{32}\cdot\frac{1}{2^k}\delta_{\mathrm{L}10}^2\frac{1}{200k} \le \frac{\beta}{32}\left(\frac{p}{2}\right)^k\delta_{\mathrm{L}10}^2\frac{n}{200k} \le \frac{\beta}{32}\left(\frac{p}{2}\right)^k|R|\,.$$

It follows that $R_{\boldsymbol{u}''}\setminus C$ and $R_{\boldsymbol{v}'}\setminus C$ still witness that $\boldsymbol{u}''$ and $\boldsymbol{v}'$, respectively, are $(\delta_{\mathrm{L}12},p)$-connected to $R\setminus\left(V(P'')\cup C\right)$. Hence we can apply Lemma 12 again to find C' connecting $\boldsymbol{u}''$ and $\boldsymbol{v}'$ in $R\setminus\left(V(P'')\cup C\right)$.

Finally, the graph obtained by concatenating P',C',P'',C certainly covers $V(G)$, and is almost a Hamilton k-cycle except that some vertices in R are used both in P' and elsewhere. But now we can appeal to the reservoir property (d) of the reservoir path P contained in P' to obtain a k-path P^* whose start and end tuples are those of P', and which uses exactly the vertices of P' not in $R\cap\left(C\cup C'\cup P''\right)$. The object obtained by concatenating P^*,C',P'',C then is the desired Hamilton k-cycle, and the proof is complete. $\qquad\square$

3. Proof of Lemma 12

In this section we prove the Connection lemma, Lemma 12. We treat the cases $k=2$ and $k\ge 3$ separately, and will first prove the case $k=2$.

3.1. **The Connection lemma for $k=2$.** The idea of the proof is as follows. We want to connect two pairs (x_1,x_2) and $(y_1,y_2)=:(x_2',x_1')$ which are (δ,p)-connected to a large set U of vertices, i.e. $|N(x_2)\cap U|,|N(x_2')\cap U| \ge \delta\frac{p}{2}|U|$, and $|N(x_1,x_2)\cap U|,|N(x_1',x_2')\cap U| \ge \delta\left(\frac{p}{2}\right)^2|U|$. For this we identify disjoint sets $X_3,\ldots,X_7$ in U and create many 2-paths $(x_1,x_2,\ldots,x_7)$ with $x_i\in X_i$ for $3\le i\le 7$ as follows. We let X_3 consist of $\Omega(p^2n)$ vertices in $N(x_1,x_2)$, X_4 of $\Omega(pn)$ vertices in $N(x_2)$, and X_5, X_6 and X_7 of $\Omega(n)$ vertices. Now any vertex $x_3\in X_3$ has the property that (x_1,x_2,x_3) is a 2-path, and most of these vertices have about the expected number of neighbours in X_4. Our pseudorandomness condition then implies that we can find $\Omega(pn)$ vertices $x_4\in X_4$ such that (x_1,x_2,x_3,x_4) is a 2-path. Similarly, $\Omega(n)$ vertices of X_5 are the end vertex of a 2-path from (x_1,x_2) through X_3 and X_4, and extending these paths further to X_6 we obtain that most vertices of X_6 are ends of 2-paths from (x_1,x_2).

Analogously we construct sets $X_3',\ldots,X_7'$ and 2-paths through these sets extending (x_1',x_2'). It remains to connect one of the 2-paths extending (x_1,x_2) and one extending (x_1',x_2'). It seems plausible that this should be possible because we have so many candidates for these 2-paths. However, so far we only know that most *vertices* in X_7 are ends of 2-paths from (x_1,x_2). But in order to connect

two 2-paths information merely about the final vertex of each of the paths is not enough, but we need information about the last *edge* of the paths. To this end we actually prove the following stronger property for X_6. We can find a subset Y_6 of $\Omega(pn)$ vertices x_6 in X_6 with the following property. There are $\Omega(p^2 n)$ vertices x_5 of X_5 such that (x_5, x_6) is the end of a 2-path from (x_1, x_2) – we call such edges $x_5 x_6$ *good*. Similarly we find $Y_6' \subseteq X_6'$ with analogous properties.

This stronger property then enables us to show that *almost all edges* from Y_6 to X_7 are ends of 2-paths from (x_1, x_2) and almost all edges from Y_6' to X_7' are ends of 2-paths from (x_1', x_2'). Since Y_6 and Y_6' are still only of size $\Omega(pn)$, we repeat this argument and obtain similar sets $Y_7 \subseteq X_7$ and $Y_7' \subseteq X_7'$ of size $\Omega(n)$, such that most edges from Y_7 to X_7' are ends of 2-paths from (x_1, x_2) and most edges from Y_7' to X_7 are ends of 2-paths from (x_1', x_2'). Since Y_7 and Y_7' are both large, we can then use the pigeonhole principle to find an edge between Y_7 and Y_7' which is the end of a 2-path both from (x_1, x_2) and (in the reverse direction) from (x_1', x_2'), and hence we find the desired 2-path connecting (x_1, x_2) and (x_2', x_1').

Proof of Lemma 12 for $k = 2$. Given $\delta > 0$, we set $\varepsilon = \delta^2 / 10^6$. Assume that G is $(\varepsilon, p, 1, 2)$-pseudorandom and $|U| \geq \delta n$. By Remark 7 this implies $p^2 |U| \geq 10^6 \delta^{-1}$. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be (δ, p)-connected to U. Our goal is to find a connection between $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{y} =: (x_2', x_1')$.

We first identify ten disjoint sets in U in which we will find our ten connecting vertices. We first choose $X_3 \subseteq (N(x_1, x_2) \cap U) \setminus \{x_1, x_2, x_1', x_2'\}$ and $X_3' \subseteq (N(x_1', x_2') \cap U) \setminus (\{x_1, x_2, x_1', x_2'\} \cup X_3)$, then $X_4 \subseteq (N(x_2) \cap U) \setminus (\{x_1, x_2, x_1', x_2'\} \cup X_3 \cup X_3')$ and $X_4' \subseteq (N(x_2') \cap U) \setminus (\{x_1, x_2, x_1', x_2'\} \cup X_3 \cup X_3' \cup X_4)$, and then pairwise disjoint subsets $X_5, X_6, X_7, X_5', X_6', X_7'$ of $U \setminus (\{x_1, x_2, x_1', x_2'\} \cup X_3 \cup X_3' \cup X_4 \cup X_4')$, such that

$$|X_3|, |X_3'| = \tfrac{1}{16} \delta p^2 |U|,$$
$$|X_4|, |X_4'| = \tfrac{1}{16} \delta p |U|,$$
$$|X_5|, |X_5'|, |X_6|, |X_6'|, |X_7|, |X_7'| = \tfrac{1}{10} |U|.$$

Here, the choice of $|X_3|$ (and similarly $|X_3'|$) is possible because (x_1, x_2) is (δ, p)-connected to U and so $|N(x_1, x_2) \cap U| \geq \delta p^2 |U|/4$. The choice of $|X_4|$ (and similarly $|X_4'|$) is possible because $|N(x_1) \cap U| \geq \delta p |U|/2$ and X_3, X_3' are small. The choice of the remaining sets is possible because all previously chosen sets are small. Note that since $p^2 |U|$ is large, all of these sets are large and rounding errors do not affect the validity of this argument.

By construction all vertices of X_3 form a 2-path with (x_1, x_2). We shall now extend these 2-paths to X_4, X_5, and so on. For this let

$$Y_3 := \left\{ y \in X_3 \colon \deg_{X_4}(y) \geq \delta p^2 |U|/20, \ \deg_{X_5}(y) \geq p|U|/20 \right\}$$

That is, the vertices in $y_3 \in Y_3$ have many 2-path extensions (x_1, x_2, y_3, x_4) into X_4 and they are good candidates for having many 2-paths which extend even further to X_5. Since $|X_4| \geq \delta p |U|/16 > \varepsilon p n$ and $\delta p^2 |U|/20 < (1 - \varepsilon) p |X_4|$ we can use Lemma 9 to infer that at most $\varepsilon p^2 n$ vertices of X_3 fail the first of these two conditions because G is $(\varepsilon, p, 1, 2)$-pseudorandom. Similarly, at most $\varepsilon p^2 n$ vertices fail the second condition, and hence

$$|Y_3| \geq |X_3| - 2\varepsilon p^2 n \geq \tfrac{1}{16} \delta p^2 |U| - 2\varepsilon p^2 \delta^{-1} |U| \geq \delta p^2 |U|/20.$$

Next, for each $y_3 \in Y_3$, we let

$$Y_4(y_3) := \left\{ y \in N_{X_4}(y_3) \colon \deg_{X_5}(y, y_3) \geq p^2 |U|/40, \ \deg_{X_6}(y) \geq p|U|/20 \right\}.$$

Observe that for each vertex $y_4 \in Y_4(y_3)$ we have that (x_1, x_2, y_3, y_4) is a 2-path and this 2-path is a good candidate for having many extensions to X_5 and X_6.

Again, since $|X_5 \cap N(y_3)| \geq \delta p|U|/20 > \varepsilon pn$ by the definition of Y_3 and $p^2|U|/40 < (1 - \varepsilon)p|X_5 \cap N(y_3)|$ we can use Lemma 9 to infer that at most $\varepsilon p^2 n$ vertices fail the first condition, and similarly for the second condition. So

$$|Y_4(y_3)| \geq |X_4 \cap N(y_3)| - 2\varepsilon p^2 n \geq \delta p^2 |U|/20 - 2\varepsilon p^2 n \geq \delta p^2 |U|/40 \,.$$

Analogously, for each $y_3 \in Y_3$ and $y_4 \in Y_4(y_3)$, we let

$$Y_5(y_3, y_4) := \left\{ y \in N_{X_5}(y_3, y_4) \colon \deg_{X_6}(y, y_4) \geq p^2|U|/40 \,, \deg_{X_7}(y) \geq p|U|/20 \right\}.$$

Similarly as before we have for each $y_5 \in Y_5(y_3, y_4)$ that $(x_1, x_2, y_3, y_4, y_5)$ is a 2-path and Lemma 9 implies $|Y_5(y_3, y_4)| \geq p^2|U|/40 - 2\varepsilon p^2 n \geq p^2|U|/80$.

For $y_3 \in Y_3$ we let $Y_5(y_3) := \bigcup_{y_4 \in Y_4(y_3)} Y_5(y_3, y_4)$, and set $Y_5 := \bigcup_{y_3 \in Y_3} Y_5(y_3)$ and claim that

$$|Y_5(y_3)| \geq p|U|/160 \qquad \text{and} \qquad |Y_5| \geq |U|/200 \,. \tag{12}$$

Indeed, for the first part let $y_3 \in Y_3$ be fixed, assume otherwise and consider the set $(N(y_3) \cap X_5) \setminus Y_5(y_3)$, which has cardinality at least $p|U|/20 - p|U|/160 = 7p|U|/160$. Since $|N(y_3) \cap X_4| \geq \delta p^2|U|/20$ by definition, we can thus use Lemma 9 to pick a vertex $y_4 \in X_4 \cap N(y_3)$ which is "typical" with respect to $N(y_3) \cap X_5$ and with respect to X_6, that is, which satisfies $|N_{X_5}(y_3, y_4) \setminus Y_5(y_3)| \geq p^2|U|/40$ and $\deg_{X_6}(y_4) \geq p|U|/20$. Hence, in particular, $y_4 \in Y_4(y_3)$. We now show that $N_{X_5}(y_3, y_4) \setminus Y_5(y_3)$, since it is big, contains a vertex from $Y_5(y_3, y_4) \subseteq Y_5(y_3)$, which yields a contradiction. For this we need to show that there is $y_5 \in N_{X_5}(y_3, y_4) \setminus Y_5(y_3)$ with $\deg_{X_6 \cap N(y_4)}(y_5) \geq p^2|U|/40$ and $\deg_{X_7}(y_5) \geq p|U|/20$. But by the definition of $y_4 \in Y_4(y_3)$ we have $|X_6 \cap N(y_4)| \geq p|U|/20$, hence the existence of such a vertex follows from Lemma 9. For the second part note that each $y_3 \in Y_3$ has at least $|Y_5(y_3)| \geq p|U|/160$ neighbours in Y_5, and thus we have $e(Y_3, Y_5) \geq |Y_3|p|U|/160$. By (3), we have $e(Y_3, Y_5) \leq (1 + \varepsilon)p|Y_3| \cdot |Y_5|$, and thus $|Y_5| \geq |U|/200$. Hence we have (12).

We next define good edges between X_5 and X_6. Let $y_5 \in Y_5$. For a neighbour $x_6 \in X_6$ of y_5, we call the edge $y_5 x_6$ *good* if $x_1 x_2 x_3 x_4 y_5 x_6$ is a 2-path for some $x_3 \in X_3$ and $x_4 \in X_4$. For each y_5 in Y_5 there are $y_3 \in Y_3$ and $y_4 \in Y_4$ such that $y_5 \in Y_5(y_3, y_4)$, which means $|N(y_4, y_5) \cap X_6| \geq p^2|U|/40$ by definition. So each vertex in Y_5 sends at least $p^2|U|/40$ good edges to X_6. Hence the average number of good edges incident to a vertex of X_6 is at least $|Y_5|p^2|U|/(40|X_6|) \geq p^2|U|/800$, where we used $|X_6| = |U|/10$ and (12). Let Z_6 be the set of those vertices in X_6 which are incident to at least $p^2|U|/1000$ good edges from Y_5. We will show that

$$|Z_6| \geq p|U|/300 \tag{13}$$

by using a double counting argument. Indeed, the total number $g(Y_5, X_6)$ of good edges from Y_5 to X_6 is at least $|Y_5|p^2|U|/40$. By definition of Z_6 each vertex in $X_6 \setminus Z_6$ is incident to less than $p^2|U|/1000$ good edges. Thus

$$|Y_5|p^2|U|/40 \leq g(Y_5, X_6) \leq e(Y_5, Z_6) + |X_6|p^2|U|/1000 \,.$$

Now (12) implies that the second summand can be bounded by $|X_6|p^2|U|/1000 = \frac{1}{200}|U| \cdot \frac{1}{50}p^2|U| \leq |Y_5|p^2|U|/50$. Hence $|Y_5|p^2|U|/200 \leq e(Y_5, Z_6) \leq |Y_5||Z_6|$ implying $|Z_6| \geq p^2|U|/200 > \varepsilon p^2 n$. This allows us to immediately obtain the desired bound (13) since we can now estimate $e(Y_5, Z_6) \leq (1 + \varepsilon)p|Y_5||Z_6|$ using (3), improving thus the lower bound on $|Z_6|$ by a factor of $p/(1 + \varepsilon)$.

We now let $Y_6 \subseteq Z_6$ be the set of those vertices with at least $p|U|/20$ neighbours in X_7. That is, Y_6 is the set of those vertices in X_6 which receive many good edges from Y_5 and have many neighbours in X_7. These are the vertices that we will continue to work with in the following. Lemma 9 gives a lower bound

$$|Y_6| \geq |Z_6| - \varepsilon p^2 n \overset{(13)}{\geq} p|U|/300 - \varepsilon p^2 n \geq p|U|/400 \,, \tag{14}$$

for the number of vertices in this set. However, this lower bound is only of order $\Omega(pn)$. Hence we iterate and define good edges between Y_6 and X_7 to obtain a linear sized set Y_7 with similar properties.

Given an edge $y_6 x_7$ from Y_6 to X_7, we call $y_6 x_7$ *good* if there is $y_5 \in Y_5$ such that $y_5 y_6$ is a good edge and y_5 is adjacent to x_7. By definition of Y_6, for $y_6 \in Y_6$ there are at least $p^2 |U|/1000 > \varepsilon p^2 n$ vertices of Y_5 which send good edges to y_6. It follows by (3) that at most εpn edges from y_6 to X_7 are not good, for each $y_6 \in Y_6$. Since vertices in Y_6 have at least $p|U|/20$ neighbours in X_7 we thus conclude that there are at least $|Y_6|(p|U|/20 - \varepsilon pn) \geq |Y_6| p|U|/40$ good edges from Y_6 to X_7. Let $Y_7 \subseteq X_7$ be the set of those vertices which are incident to at least $p^2 |U|/5000$ good edges (again, a bit less than the average, which is at least $|Y_6| p|U|/(40|X_7|) \geq p|U|/4000$). Applying a similar double counting argument as before, using (3) and (14), we obtain

$$|Y_7| \geq |U|/100 \, . \tag{15}$$

Let us examine the good edges leaving Y_7: We call an edge from $y_7 \in Y_7$ to $x_7' \in X_7'$ *good* if there is $y_6 \in Y_6$ such that $y_6 y_7$ is a good edge and y_6 is adjacent to x_7'. By definition of Y_7, for each $y_7 \in Y_7$ there are at least $p^2 |U|/5000$ good edges from Y_6 to y_7, and thus by (3) there are at most εpn edges from y_7 to X_7' which are not good. Observe that by definition any good edge $y_7 x_7'$ from Y_7 to X_7' is the last edge in a 2-path from $x_1 x_2$ to $y_7 x_7'$ using one vertex of each set $X_3, \ldots, X_6$.

Now we repeat the identical construction within the sets $X_3', \ldots, X_7'$, obtaining a set $Y_7' \subseteq X_7'$ of size at least $|U|/100$, where each vertex $y_7' \in Y_7'$ sends at most εpn edges to X_7 which are not good, and each good edge from y_7' to X_7 is the last edge in a 2-path from $x_1' x_2'$ using one vertex of each set $X_3', \ldots, X_6'$.

Finally we can apply the pigeon hole principle: By (3) there are at least

$$(1-\varepsilon)p|Y_7||Y_7'| \overset{(15)}{\geq} p|U|^2/20000 > \varepsilon pn^2 > \big(|Y_7| + |Y_7'|\big)\varepsilon pn$$

edges between Y_7 and Y_7', and in particular there is one edge $y_7 y_7'$ which is both good from Y_7 to Y_7' and good from Y_7' to Y_7. This yields a 2-path from $x_1 x_2$ to $x_2' x_1'$ using one vertex of each set $X_3, \ldots, X_6$, y_7, y_7', and one vertex of each set $X_6', \ldots, X_3'$, as desired. $\square$

3.2. **The Connection lemma for $k > 2$.** We use the same general strategy as in the $k = 2$ case. To connect the k-tuples $\boldsymbol{x}$ and $\boldsymbol{y}$ we start by constructing short k-paths from $\boldsymbol{x}$ step by step. In each step we look for many possible extensions of each of the k-paths constructed so far (so in step i all our k-paths will be of length i). Our goal is to continue until we reach a collection of k disjoint $\Omega(n)$-sized vertex subsets of U such that
($\star$) most copies of K_k with one vertex in each of the k sets are ends of k-paths leaving $\boldsymbol{x}$.
Repeating from $\boldsymbol{y}$, the pigeonhole argument then guarantees that one of these copies of K_k is also the end of a k-path leaving $\boldsymbol{y}$ in the reverse order, and thus we get the desired $\boldsymbol{x}$-$\boldsymbol{y}$ connection.

However, obtaining property ($\star$) is not straightforward. In fact $(\varepsilon, p, k - 1, k)$-pseudorandomness, a weaker pseudorandomness condition than we require, would be enough to guarantee that after $k + 1$ steps we get k-paths from $\boldsymbol{x}$ to a set of $\Omega(n)$ vertices. Thus after $k - 1$ further steps we get k disjoint $\Omega(n)$-sized subsets of U of vertices which are the ends of k-paths from $\boldsymbol{x}$, and we might hope that these sets also satisfy property ($\star$). However we are not able to show this with this weaker pseudorandomness condition.

Hence we resort to demanding $(\varepsilon, p, k-1, 2k-1)$-pseudorandomness. This allows us to show an inductive version of $(\star)$: at each step we maintain the property that most copies of K_k in the final k sets are ends of k-paths from $\boldsymbol{x}$.

The inductive argument as well as the pigeonhole argument in this proof rely on the following proposition, which states that in a sufficiently pseudorandom graph every collection of k sufficiently large disjoint vertex sets span roughly the expected number of k-cliques. We use the following definitions. For a graph G and disjoint subsets $V_1, \ldots, V_k$ of the vertex set $V(G)$ we denote by $K_k(V_1, \ldots, V_k)$ the set of all copies of K_k crossing $V_1, \ldots, V_k$, i.e., with one vertex in each of the sets $V_1, \ldots, V_k$. Given $p \in [0, 1]$, we define

$$\tilde{K}_k(V_1, \ldots, V_k) := p^{\binom{k}{2}} \prod_{i=1}^{k} |V_i|\,,$$

which we call the *expected number of k-cliques crossing $V_1, \ldots, V_k$*.

Proposition 14. *For each $0 < \mu \leq 1$ and integer $k \geq 1$ there exists $\varepsilon > 0$ such that for all $p \in (0, 1)$ the following holds. Suppose that $k \geq r \geq 2$ is an integer, and that $V_1, \ldots, V_r$ are pairwise disjoint vertex sets in an $(\varepsilon, p, k-1, 2k-2)$-pseudorandom graph G on n vertices such that $|V_i| \geq \mu p^{k-i} n$ for each $r \geq i \geq 1$. Then we have*

$$\left| K_r(V_1, \ldots, V_r) \right| = (1 \pm \mu) \tilde{K}_r(V_1, \ldots, V_r)\,.$$

We remark that the lower bound in this proposition requires only $(\varepsilon, p, k-2, k-1)$-pseudorandomness and that also the pseudorandomness requirement for the upper bound can undoubtedly be improved.

Proof of Proposition 14. Given $0 < \mu \leq 1$, we take $0 < \varepsilon_0 < 2^{-k}\mu$ small enough so that $(1 \pm 2k\varepsilon_0/\mu)^{\binom{k+1}{2}}$ is a sub-range of $1 \pm \mu$. Given $0 < \varepsilon < \varepsilon_0$, we will prove by induction on r the stronger statement

$$\left| K_r(V_1, \ldots, V_r) \right| = (1 \pm \tfrac{2k\varepsilon}{\mu})^{\binom{r+1}{2}} \tilde{K}_r(V_1, \ldots, V_r)$$

for disjoint sets $V_1, \ldots, V_r$ in an $(\varepsilon, p, k-1, 2k-2)$-pseudorandom graph G with $|V_i| \geq 2^{r-k}\mu p^{k-i} n$ for each i. The base case $r = 2$ is immediate from $(\varepsilon, p, k-2, k-1)$-pseudorandomness.

For the induction step, we split the vertices of V_1 into two classes: the typical vertices, whose degree into V_i is $(1 \pm \varepsilon)p|V_i|$ for each $2 \leq i \leq r$, and the remaining atypical vertices. Since $(1 - \varepsilon)p|V_i| \geq p|V_i|/2 \geq 2^{r-1-k}\mu p^{k-i+1} n$, for each typical vertex v we have by induction the estimate

$$\left| K_{r-1}\big(N_{V_2}(v), \ldots, N_{V_r}(v)\big) \right| = (1 \pm 2k\varepsilon/\mu)^{\binom{r}{2}} \left| \tilde{K}_{r-1}\big(N_{V_2}(v), \ldots, N_{V_r}(v)\big) \right|$$

$$= (1 \pm 2k\varepsilon/\mu)^{\binom{r}{2}} p^{\binom{r-1}{2}} |N_{V_2}(v)| \ldots |N_{V_r}(v)|$$

$$= (1 \pm 2k\varepsilon/\mu)^{\binom{r}{2}} p^{\binom{r}{2}} (1 \pm \varepsilon)^{r-1} |V_2| \cdots |V_r|\,,$$

which is the contribution of v to $\left| K_r(V_1, \ldots, V_r) \right|$. By Lemma 9 all but at most $2(r-1)\varepsilon p^{2k-2} n \leq 2k\varepsilon p^{k-1}\mu^{-1}|V_1|$ vertices of V_1 are typical. This clearly already yields the lower bound of our proposition.

To obtain the upper bound, it is then enough to show that the atypical vertices do not contribute too much. An atypical vertex certainly does not contribute more than $\left| K_{r-1}(V_2, \ldots, V_r) \right|$, which by induction is not more than

$$(1 + \tfrac{2k\varepsilon}{\mu})^{\binom{r}{2}} p^{\binom{r-1}{2}} |V_2| \cdots |V_r|\,.$$

Hence we get

$$
\begin{aligned}
\big|K_r(V_1,\ldots,V_r)\big| &\leq \big(1+\tfrac{2k\varepsilon}{\mu}\big)^{\binom{r}{2}}p^{\binom{r}{2}}(1+\varepsilon)^{r-1}|V_1|\cdots|V_r| \\
&\quad + \tfrac{2k\varepsilon}{\mu}p^{k-1}|V_1|\big(1+\tfrac{2k\varepsilon}{\mu}\big)^{\binom{r}{2}}p^{\binom{r-1}{2}}|V_2|\cdots|V_r| \\
&\leq \Big(\big(1+\tfrac{2k\varepsilon}{\mu}\big)^{\binom{r}{2}+r-1}+\tfrac{2k\varepsilon}{\mu}\big(1+\tfrac{2k\varepsilon}{\mu}\big)^{\binom{r}{2}}\Big)p^{\binom{r}{2}}|V_1|\cdots|V_r| \\
&\leq \big(1+\tfrac{2k\varepsilon}{\mu}\big)^{\binom{r+1}{2}}p^{\binom{r}{2}}|V_1|\cdots|V_r|
\end{aligned}
$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We now give the proof of the Connection lemma in the case $k > 2$, modulo a claim which encapsulates the inductive argument, whose proof we will provide subsequently.

Proof of Lemma 12 for $k > 2$. Let $k > 2$ and $0 < \delta \leq 1/(6k)$ be given. We set $\xi_{k+1} := \tfrac{1}{3}$, and for each $k+1 \geq i \geq 2$, we set

$$
\xi_{i-1} := \tfrac{1}{4}\xi_i^{k-1}3^{-\binom{k}{2}}. \tag{16}
$$

We choose

$$
\mu := \big(\tfrac{1}{10k}10^{-10k^2}\delta^2\xi_1\big)^2 \quad \text{and} \quad \varepsilon \leq \mu \tag{17}
$$

to be small enough for Proposition 14 with input μ and k. Let $0 < p < 1$ and G be an $(\varepsilon, p, k-1, 2k-1)$-pseudorandom graph on n vertices. Let U be a subset of $V(G)$ of size $|U| \geq \delta n$. Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ are disjoint k-tuples which are (δ, p)-connected to U.

We choose pairwise disjoint subsets $U_1,\ldots,U_{2k},U_1',\ldots,U_k'$ of U with

$$
|U_i|,|U_i'| = \tfrac{1}{3k}\delta^2(p/2)^{k-i+1}n \qquad \text{for } i \leq k, \text{ and} \tag{18}
$$

$$
|U_i| = \tfrac{1}{3k}\delta n \qquad\qquad\qquad \text{for } i > k \tag{19}
$$

as follows. We first choose the disjoint sets U_1 in $U \cap N(x_1,\ldots,x_k)$ and U_1' in $U \cap N(y_1,\ldots,y_k)$. From the remaining vertices in U we then choose the disjoint sets U_2 in $U \cap N(x_2,\ldots,x_k)$ and U_2' in $U \cap N(y_2,\ldots,y_k)$. We continue in this fashion, choosing for each $i \leq k$ the set U_i in $U \cap N(x_i,\ldots,x_k)$ and the set U_i' in $U \cap N(y_i,\ldots,y_k)$. Choosing these sets such that each set is disjoint from the previously chosen sets is possible by the (δ, p)-connectedness of $\boldsymbol{x}$ and $\boldsymbol{y}$ to U. Finally we choose in the remaining vertices of U disjoint sets $U_{k+1},\ldots,U_{2k}$ arbitrarily of the prescribed size. Further, for each $i \in [k]$ we let $U_{k+i}' := U_{2k-i+1}$. To summarise, we constructed $3k$ disjoint sets which we will use to construct k-paths: we will find many k-paths starting in $\boldsymbol{x}$ with one vertex in each of $U_1,\ldots,U_{2k}$ (that is why we chose $U_1,\ldots,U_k$ in the neighbourhood of vertices from $\boldsymbol{x}$), and many k-paths starting in $\boldsymbol{y}$ using $U_1',\ldots,U_{2k}'$. We will argue that, since $U_{k+1},\ldots,U_{2k}$ and $U_{2k}',\ldots,U_{k+1}'$ coincide, two of these k-paths join.

More precisely, for each $1 \leq i \leq k+1$, we call a k-clique $\boldsymbol{c}$ in $K_k(U_i,\ldots,U_{i+k-1})$ *good* (with respect to $\boldsymbol{x}$) if there is a k-path from $\boldsymbol{x}$ with one vertex in each of $U_1,\ldots,U_{i-1}$ followed by $\boldsymbol{c}$, in that order, and *bad* otherwise. We will use the following claim, whose proof we postpone.

Claim 15. *For each $1 \leq i \leq k+1$, all but at most $\xi_i \tilde{K}_k(U_i,\ldots,U_{i+k-1})$ of the k-cliques in $K_k(U_i,\ldots,U_{i+k-1})$ are good.*

This claim implies the desired statement. Indeed, by Claim 15 all but at most $\tfrac{1}{3}\tilde{K}(U_{k+1},\ldots,U_{2k})$ of the k-cliques in $K_k(U_{k+1},\ldots,U_{2k})$ are good with respect to $\boldsymbol{x}$. Similarly, for each $1 \leq i \leq k$ we call a clique in $K_k(U_i',\ldots,U_{i+k-1}')$ *good with respect to $\boldsymbol{y}$* if it is the end of a k-path from $\boldsymbol{y}$ using one vertex in

each of $U'_1, \ldots, U'_{i+k-1}$ in that order. By symmetry Claim 15 guarantees that also all but at most $\frac{1}{3}\tilde{K}(U_{k+1}, \ldots, U_{2k})$ of the k-cliques in $K_k(U'_{k+1}, \ldots, U'_{2k}) = K_k(U_{2k}, \ldots, U_{k+1})$ are good with respect to $\boldsymbol{y}$. By Proposition 14, there are at least

$$(1 - \mu)\tilde{K}(U_{k+1}, \ldots, U_{2k}) > \tfrac{2}{3}\tilde{K}(U_{k+1}, \ldots, U_{2k})$$

cliques in $K_k(U_{k+1}, \ldots, U_{2k})$, and therefore there must exist a clique which is both good with respect to $\boldsymbol{x}$ and to $\boldsymbol{y}$. Hence we obtain the desired $(\boldsymbol{x} - \boldsymbol{y})$-connecting k-path. $\qquad \square$

It remains to establish Claim 15, which we prove by induction on i.

Proof of Claim 15. For the base case $i = 1$, observe that by definition of the sets $U_1, \ldots, U_k$ there are no bad cliques in $K_k(U_1, \ldots, U_k)$.

For the induction step, assume $2 \leq i \leq k+1$. Let $W_0 := U_{i-1}, \ldots, W_k := U_{i+k-1}$. Suppose for contradiction that $K_k(W_1, \ldots, W_k)$ contains at least $\xi_i \tilde{K}_k(W_1, \ldots, W_k)$ bad cliques. We shall show that this implies at least $\xi_{i-1}\tilde{K}_k(W_0, \ldots, W_{k-1})$ bad cliques in $K_k(W_0, \ldots, W_{k-1})$, contradicting the induction hypothesis.

To this end we shall find many cliques $\boldsymbol{c}$ of size $k-1$ in $K_{k-1}(W_1, \ldots, W_{k-1})$ with the following two properties. Firstly, $\boldsymbol{c}$ has a set $C'_0(\boldsymbol{c})$ of common neighbours in W_0 of size at least $(1-\varepsilon)^{k-1}p^{k-1}|W_0|$ (i.e. almost the expected number). Secondly, there is a set $C_k(\boldsymbol{c})$ of vertices in W_k with $|C_k(\boldsymbol{c})| \geq \xi_i 3^{1-k}p^{k-1}|W_k|$ (i.e. a small but constant fraction of the average) such that $(\boldsymbol{c}, c_k)$ forms a bad clique for each $c_k \in C_k(\boldsymbol{c})$. If a $(k-1)$-clique $\boldsymbol{c}$ has these two properties we also say that $\boldsymbol{c}$ is a *normal clique*.

Claim 16. $K_{k-1}(W_1, \ldots, W_{k-1})$ *contains at least*

$$\prod_{j=1}^{k-1} \xi_i 3^{-j}p^{j-1}|W_j| = \xi_i^{k-1} 3^{-\binom{k}{2}} p^{\binom{k-1}{2}} |W_1| \cdots |W_{k-1}|$$

normal $(k-1)$-cliques.

Before proving this claim we argue that this implies the desired contradiction. Indeed, let $\boldsymbol{c}$ be a normal $(k-1)$-clique in $K_{k-1}(W_1, \ldots, W_{k-1})$. Then by definition we have

$$
\begin{aligned}
|C'_0(\boldsymbol{c})| &\geq (1-\varepsilon)^{k-1}p^{k-1}|W_0| \geq (1-\varepsilon)^{k-1}p^{k-1}\left(\tfrac{1}{2}|W_0| + \tfrac{1}{2}|U_1|\right) \\
&\overset{(18)}{\geq} (1-\varepsilon)^{k-1}p^{k-1}\left(\frac{1}{2}|W_0| + \frac{1}{2}\cdot\frac{\delta^2}{3k\cdot 2^k}p^k n\right) \\
&\overset{(17)}{\geq} \tfrac{1}{2}(1-\varepsilon)^{k-1}p^{k-1}|W_0| + \varepsilon p^{2k-1}n
\end{aligned}
$$

and

$$|C_k(\boldsymbol{c})| \geq \xi_i 3^{1-k}p^{k-1}|W_k| \overset{(19)}{\geq} \xi_i 3^{-k}\cdot\frac{1}{k}\delta\cdot p^{k-1}n \overset{(17)}{\geq} \varepsilon p^{k-1}n\,.$$

Thus, since G is $(\varepsilon, p, k-1, 2k-1)$-pseudorandom, Lemma 9 implies that at most $\varepsilon p^{2k-1}n$ vertices of $C'_0(\boldsymbol{c})$ do not have any neighbours in $C_k(\boldsymbol{c})$. It follows that the set $C_0(\boldsymbol{c})$ of vertices in $C'_0(\boldsymbol{c})$ which do have neighbours in $C_k(\boldsymbol{c})$ has size at least $\tfrac{1}{2}(1-\varepsilon)^{k-1}p^{k-1}|W_0|$.

Why are we interested in these edges $c_0 c_k$ between $C_0(\boldsymbol{c})$ and $C_k(\boldsymbol{c})$? By definition of $C_k(\boldsymbol{c})$ the k-clique $(\boldsymbol{c}, c_k)$ is a bad k-clique in $K_k(W_1, \ldots, W_k)$. Hence, since by definition of $C'_0(\boldsymbol{c}) \supseteq C_0(\boldsymbol{c})$ we have $c_0 \in N_{W_0}(\boldsymbol{c})$, the edge $c_0 c_k$ witnesses that also the k-clique $(c_0, \boldsymbol{c})$ must be bad (in $K_{k-1}(W_1, \ldots, W_{k-1})$). Because $\tilde{K}_k(W_0, \ldots, W_{k-1}) = p^{\binom{k}{2}}\prod_{i=0}^{k-1}|W_i|$ by definition, it therefore follows from

Claim 16 that we find at least

$$|C_0(\boldsymbol{c})|\cdot\xi_i^{k-1}3^{-\binom{k}{2}}p^{\binom{k-1}{2}}|W_1|\cdots|W_{k-1}|$$

$$\geq \tfrac{1}{2}(1-\varepsilon)^{k-1}p^{k-1}|W_0|\cdot\xi_i^{k-1}3^{-\binom{k}{2}}p^{\binom{k-1}{2}}|W_1|\cdots|W_{k-1}|$$

$$= \tfrac{1}{2}(1-\varepsilon)^{k-1}\xi_i^{k-1}3^{-\binom{k}{2}}\tilde{K}_k(W_0,\ldots,W_{k-1}) \overset{(16)}{>} \xi_{i-1}\tilde{K}_k(W_0,\ldots,W_{k-1})$$

bad cliques in $K_k(W_0,\ldots,W_{k-1})$, which is the desired contradiction.

Proof of Claim 16. We construct the normal $(k-1)$-cliques vertex by vertex in the following way. We first construct a set $Z_1 \subseteq W_1$ with $|Z_1| \geq \xi_i|W_1|/3$ and then for each $c_1 \in Z_1$ a set $Z_2(c_1) \subseteq N_{W_2}(c_1)$ with $|Z_2| \geq \xi_i p|W_2|/9$, and so on, in general constructing for $c_1 \in Z_1, c_2 \in Z_2(c_1),\ldots,c_{j-1} \in Z_{j-1}(c_1,\ldots,c_{j-2})$ a set $Z_j(c_1,\ldots,c_{j-1}) \subseteq N_{W_j}(c_1,\ldots,c_{j-1})$ with

$$|Z_j(c_1,\ldots,c_{j-1})| \geq \xi_i p^{j-1}|W_j|/3^j , \tag{20}$$

where j ranges from 1 to $k-1$, such that the following properties hold for each $c_j \in Z_j(c_1,\ldots,c_{j-1})$. Firstly, $(c_1,\ldots,c_j)$ is in at least

$$\xi_i 3^{-j}p^{\binom{k}{2}-\binom{j}{2}}|W_{j+1}|\ldots|W_k| \tag{21}$$

bad k-cliques in $K_k(W_1,\ldots,W_k)$. Secondly, for each $\ell \in \{0\}\cup\{j+1,\ldots,k\}$ the vertex c_j is *typical* with respect to $N_{W_\ell}(c_1,\ldots,c_{j-1})$, that is,

$$|N(c_j)\cap N_{W_\ell}(c_1,\ldots,c_{j-1})| = (1\pm\varepsilon)p|N_{W_\ell}(c_1,\ldots,c_{j-1})| . \tag{22}$$

Observe that by definition $(c_1,\ldots,c_j)$ form a clique for each $c_1 \in Z_1$, $c_2 \in Z_2(c_1)$, $\ldots$, $c_j \in Z_j(c_1,\ldots,c_{j-1})$. Moreover, successfully constructing all these sets proves Claim 16. Indeed, for each $c_1 \in Z_1$, $c_2 \in Z_2(c_1)$, $\ldots$, $c_{k-1} \in Z_{k-1}(c_1,\ldots,c_{k-2})$ the clique $(c_1,\ldots,c_{k-1})$ satisfies $|N_{W_0}(c_1,\ldots,c_{k-1})| \geq (1-\varepsilon)^{k-1}p^{k-1}|W_0|$ by (22), and $(c_1,\ldots,c_{k-1})$ is in at least $\xi_i 3^{-k+1}p^{k-1}|W_k|$ bad k-cliques in $K_k(W_1,\ldots,W_k)$ by (21), so $(c_1,\ldots,c_k)$ is normal. By (20) there are

$$\prod_{j=1}^{k-1}\xi_i 3^{-j}p^{j-1}|W_j| = \xi_i^{k-1}3^{-\binom{k}{2}}p^{\binom{k-1}{2}}|W_1|\cdots|W_{k-1}|$$

such $(k-1)$-cliques $(c_1,\ldots,c_{k-1})$.

It remains to show that we can construct sets $Z_j(c_1,\ldots,c_{j-1})$ satisfying (20), (21) and (22). We proceed by induction on j, where the base case and the inductive step use the same reasoning. So let $\boldsymbol{c} = (c_1,\ldots,c_{j-1})$ with $c_1 \in Z_1$, $\ldots$, $c_{j-1} \in Z_{j-1}(c_1,\ldots,c_{j-2})$ be fixed and assume that we constructed Z_1, $\ldots$, $Z_{j-1}(c_1,\ldots,c_{j-2})$ successfully. Now we consider $N_{W_j}(\boldsymbol{c})$ and first bound the size of the set $A_j \subseteq N_{W_j}(\boldsymbol{c})$ of vertices that violate (22) (where, as is usual, we adopt the convention that $N_{W_j}(\boldsymbol{c}) = W_j$ if $\boldsymbol{c}$ is empty, which happens in the base case $j=1$). By (22) in the induction hypothesis we have for each $\ell \in \{0\}\cup\{j,\ldots,k\}$ that

$$|N_{W_\ell}(\boldsymbol{c})| \overset{(22)}{=} (1\pm\varepsilon)p|N_{W_\ell}(c_1,\ldots,c_{j-2})| \overset{(22)}{=} (1\pm\varepsilon)^2 p^2|N_{W_\ell}(c_1,\ldots,c_{j-3})|$$

$$\overset{(22)}{=} \ldots \overset{(22)}{=} (1\pm\varepsilon)^{j-1}p^{j-1}|W_\ell| = (1\pm\varepsilon)^{j-1}p^{j-1}|U_{i-1+\ell}|$$

$$\overset{(18)}{\geq} (1\pm\varepsilon)^{j-1}p^{j-1}\frac{1}{3k}\delta^2\Big(\frac{p}{2}\Big)^{k-i-\ell+2}n = \frac{(1\pm\varepsilon)^{j-1}\delta^2}{3k\cdot 2^{k-i-\ell+2}}\cdot p^{k-\ell+j-i+1}n \tag{23}$$

and thus

$$|N_{W_\ell}(\boldsymbol{c})| \overset{(17)}{\geq} \sqrt{\mu}\,p^{k-\ell+j-1}n \overset{(17)}{\geq} \varepsilon p^{2k-2}n , \tag{24}$$

since $i \geq 2$, $j \leq k-1$, $\ell \geq 0$. Since G is $(\varepsilon, p, k-1, 2k-2)$-pseudorandom it follows from Lemma 9 that only $|A_j| < 2k\varepsilon p^{k-1}n$ vertices in $N_{W_j}(\boldsymbol{c})$ violate (22).

We now construct the desired set $Z_j(\boldsymbol{c})$ as follows. We choose among the vertices in $N_{W_j}(\boldsymbol{c}) \setminus A_j$ those $\xi_i p^{j-1}|W_j|/3^j$ vertices c_j which are together with $\boldsymbol{c}$ in the biggest number of bad k-cliques in $K_k(W_1, \ldots, W_k)$. This is possible since

$$|W_j| \overset{(18)}{\geq} \frac{1}{3k}\delta^2(p/2)^{k-i-j+2}n \geq 100k\sqrt{\mu}p^{k-j}n \geq 100k\varepsilon p^{k-j}n \qquad (25)$$

because $i \geq 2$, and by (23) we have $|N_{W_j}(\boldsymbol{c})| \geq (1-\varepsilon)^{j-1}p^{j-1}|W_j| \geq \frac{3}{4}p^{j-1}|W_j|$ and hence $|N_{W_j}(\boldsymbol{c}) \setminus A_j| \geq \frac{3}{4}p^{j-1}|W_j| - 2k\varepsilon p^{k-1}n \geq \frac{1}{2}p^{j-1}|W_j|$. By construction $Z_j(\boldsymbol{c})$ satisfies (20) and (22). In the remainder of this proof we will show that $Z_j(\boldsymbol{c})$ also satisfies (21).

For this purpose we next estimate how many k-cliques in $K_k(W_1, \ldots, W_k)$ use $\boldsymbol{c}$ and a vertex in A_j (which is an upper bound on the number of bad k-cliques that we "lose" to A_j). Observe that $\left|K_{k-j+1}\big(A_j, N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})\big)\right|$ is exactly the number of such k-cliques. In order to upper bound this quantity we want to apply Proposition 14 and so we must justify that the sets $A_j, N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})$ are large enough for this application. In fact, $|A_j|$ is not large enough, but we can rectify this by adding arbitrary vertices of $N_{W_j}(\boldsymbol{c})$ to obtain a set A_j' of size $\sqrt{\mu}|N_{W_j}(\boldsymbol{c})| \geq \mu p^{k-1}n$, where we used (24). By (24) we also have $|N_{W_\ell}(\boldsymbol{c})| \geq \mu p^{k-\ell+j-1}n$. Thus Proposition 14 implies that at most

$$\left|K_{k-j+1}\big(A_j, N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})\big)\right| \leq \left|K_{k-j+1}\big(A_j', N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})\big)\right|$$

$$\leq (1+\mu)\tilde{K}_{k-j+1}\big(A_j', N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})\big)$$

$$= (1+\mu)p^{\binom{k-j+1}{2}}\sqrt{\mu}|N_{W_j}(\boldsymbol{c})||N_{W_{j+1}}(\boldsymbol{c})|\cdots|N_{W_k}(\boldsymbol{c})|$$

$$\overset{(23)}{\leq} (1+\mu)p^{\binom{k-j+1}{2}}\sqrt{\mu}(1+\varepsilon)^{(k-j+1)(j-1)}p^{(k-j+1)(j-1)}|W_j|\cdots|W_k|$$

$$\overset{(17)}{\leq} \tfrac{1}{6}\xi_i 3^{1-j}p^{\binom{k}{2}-\binom{j-1}{2}}|W_j|\cdots|W_k|.$$

k-cliques in $K_k(W_1, \ldots, W_k)$ use $\boldsymbol{c}$ and a vertex in A_j.

From this together with (21) in the induction hypothesis we immediately get that the number of bad cliques in $K_k(W_1, \ldots, W_k)$ which use $\boldsymbol{c}$ and a vertex in $N_{W_j}(\boldsymbol{c}) \setminus A_j$ is at least $\frac{5}{6}\xi_i 3^{1-j}p^{\binom{k}{2}-\binom{j-1}{2}}|W_j|\cdots|W_k|$. We claim (and show below) that moreover at most half of these, i.e., at most

$$\tfrac{5}{12}\xi_i 3^{1-j}p^{\binom{k}{2}-\binom{j-1}{2}}|W_j|\cdots|W_k| \qquad (26)$$

bad cliques in $K_k(W_1, \ldots, W_k)$ use $\boldsymbol{c}$ and a vertex in $Z_j(\boldsymbol{c})$. Hence the vertex c_j in $N_{W_j}(\boldsymbol{c}) \setminus (A_j \cup Z_j(\boldsymbol{c}))$ which together with $\boldsymbol{c}$ is in the biggest number of bad k-cliques in $K_k(W_1, \ldots, W_k)$ is in at least

$$\tfrac{5}{12}\xi_i 3^{1-j}p^{\binom{k}{2}-\binom{j-1}{2}}|W_j|\cdots|W_k|/|N_{W_j}(\boldsymbol{c})| \overset{(23)}{\geq} \xi_i 3^{-j}p^{\binom{k}{2}-\binom{j}{2}}|W_{j+1}|\cdots|W_k|$$

such bad k-cliques. By construction of $Z_j(\boldsymbol{c})$ this is thus also true for the vertices c_j in $Z_j(\boldsymbol{c})$ and hence we get (21).

It remains to establish (26), which we obtain (similar as before) by bounding the size of $K_{k-j+1}\big(Z_j(\boldsymbol{c}), N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})\big)$ with the help of Proposition 14. Indeed, by (25) and (17) we have $|Z_j(\boldsymbol{c})| = \xi_i p^{j-1}|W_j|/3^j \geq (\xi_i/3^j)100k\sqrt{\mu}p^{k-1}n \geq \mu p^{k-1}n$ and hence this proposition implies

$$\left|K_{k-j+1}\big(Z_j(\boldsymbol{c}), N_{W_{j+1}}(\boldsymbol{c}), \ldots, N_{W_k}(\boldsymbol{c})\big)\right|$$

$$\leq (1+\mu)p^{\binom{k-j+1}{2}}|Z_j(\boldsymbol{c})||N_{W_{j+1}}(\boldsymbol{c})|\cdots|N_{W_k}(\boldsymbol{c})|$$

$$\overset{(23)}{\leq} (1+\mu)p^{\binom{k-j+1}{2}} \cdot \xi_i 3^{-j}p^{j-1}|W_j| \cdot (1+\varepsilon)^{(j-1)(k-j)}p^{(j-1)(k-j)}|W_{j+1}|\cdots|W_k|$$

$$\overset{(17)}{\leq} \tfrac{5}{12}\xi_i 3^{1-j}p^{\binom{k}{2}-\binom{j-1}{2}}|W_j|\cdots|W_k|$$

as desired. □

This concludes the proof of Claim 15. □

4. Proof of Lemma 10 and Lemma 11

In this section we will prove the following technical lemma, which implies both Lemma 10 and Lemma 11.

Lemma 17. *Given $k \geq 2$, $0 < \delta < 1/4$ and $0 < \beta < 1/2$ there exists an $\varepsilon > 0$ such that the following holds. Let $0 < p < 1$ and let G be an n-vertex graph. Suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom if $k = 2$; and $(\varepsilon, p, k-1, 2k-1)$-pseudorandom and $(\varepsilon, p, k, k+1)$-pseudorandom if $k > 2$. Let R and S be disjoint subsets of $V(G)$ with $|R| \leq \delta n/(200k)$ and $|S| \geq \delta n$ such that $\deg_S(v) \geq \beta \delta p n/2$ for all $v \in R$. Then there is a k-path P contained in $R \cup S$ with the following properties.*

(a) $R \subseteq V(P)$ and $|V(P)| \leq 50k|R|$.

(b) For any $W \subseteq R$, there is a k-path on $V(P) \setminus W$ whose start and end k-tuples are identical to those of P.

(c) The start and end k-tuples of the k-path P are in S and are $(\frac{1}{8}, p)$-connected to $S \setminus V(P)$.

(d) If $|R| \geq \delta^2 n/(200k)$, then the start and end k-tuples of P are $(\frac{1}{2}, p)$-connected to R.

In order to infer the Reservoir lemma, Lemma 10, from this lemma use R as given and set $S := V(G) \setminus R$. For the Covering lemma, Lemma 11 use S as given and set $R := L$.

For the proof of Lemma 17 we use the following definition. Given k, a *k-reservoir graph* is a graph which contains a spanning k-path P with the following extra property. There is a special vertex r, which we call the *reservoir vertex* of the reservoir graph, such that $V(P) \setminus \{r\}$ forms a k-path whose start and end k-tuples are identical to those of P. We also call these tuples the *start and end tuple of the reservoir graph*. To give a simple example, the triangle abc is a 1-reservoir graph, with $P = (a, c, b)$ and c being the reservoir vertex, and K_5 is a 2-reservoir graph. However, in our proof of Lemma 17 we shall need much sparser reservoir graphs. The following lemma states that such graphs exist.

Lemma 18 (Reservoir graph lemma). *For all $k \geq 2$, $\beta > 0$ and $0 < \delta < 1/4$ there exists an $\varepsilon > 0$ such that the following holds. Let G be an n-vertex graph, S a subset of $V(G)$ of size at least $\delta n/2$, and R^* a subset of $V(G)$ of size at least $\delta^2 n/(200k)$. Let r be a vertex of $V(G) \setminus S$ with at least $\beta \delta p n/8$ neighbours in S. For $k \geq 3$ suppose that G is $(\varepsilon, p, k, k+1)$-pseudorandom and for $k = 2$ suppose that G is $(\varepsilon, p, 1, 2)$-pseudorandom.*

Then there is a reservoir graph H in G whose reservoir vertex is r, whose remaining vertices are in S, and whose start and end k-tuple are $(\frac{1}{2}, p)$-connected to S and to R^*. Furthermore the reservoir graph has at most $\max(47, 2k+1)$ vertices.

Observe that this lemma allows us to specify the reservoir vertex r and a small vertex set S which contains the remaining constant number of vertices of the reservoir graph – the only requirement being that r has many neighbours in S. Moreover, the lemma guarantees well-connectedness of the start and end tuple to S and an additional vertex set R^*, which we shall use to obtain property (d) in the proof of Lemma 17.

We prove Lemma 18 at the end of this section and next show how it entails Lemma 17. We first briefly explain the idea. Roughly speaking, our goal is to construct for each $r \in R$ a reservoir graph which uses r as reservoir vertex and to

connect up all these reservoir graphs into a long path. Observe that the definition of a reservoir graph implies that the resulting long path is a reservoir path, that is, a k-path satisfying (b) of Lemma 17.

Proof of Lemma 17. Let $k \geq 2$, δ and β satisfy the conditions given in the statement of Lemma 17. We require $\varepsilon > 0$ to be sufficiently small to apply Lemma 18 with input k, β and δ, to apply Lemma 12 with input k and $\delta/4$, and smaller than $\beta\delta/(200k)$.

Let p, G, R and S be as in the statement of Lemma 17. Our approach now is as follows. We want to choose one vertex r_1 of R and use the Reservoir graph lemma, Lemma 18, to construct a k-reservoir graph H_1 with reservoir vertex r_1 and all other vertices in S. We then want to continue by choosing a second vertex r_2 of R, and repeat this procedure to obtain H_2, avoiding the vertices of H_1. Next we want to use the Connection lemma, Lemma 12, to connect the end tuple of H_1 to the start tuple of H_2, again within S. We will repeat this until finally we construct $H_{|R|}$ and connect it to $H_{|R|-1}$, and the result is the desired reservoir path P.

Before we can start we have to set up R^* for Lemma 17. Recall that the purpose of R^* will be to guarantee property (d). So, if we have $|R| \geq \delta^2 n/(200k)$, we set $R^* := R$; otherwise we set $R^* := V(G)$.

We now perform the first step of our procedure. Let r_1 be the vertex of R of lowest degree to S. By our assumptions we have $\deg_S(r_1) \geq \beta\delta pn/2$, hence we can apply Lemma 18. Let H_1 be the k-reservoir graph with reservoir vertex r_1 and remaining vertices in S guaranteed by this lemma. The start k-tuple of H_1 is $(\frac{1}{2}, p)$-connected to S, and so by the choice of ε, and Remark 7, it is also $(\frac{1}{8}, p)$-connected to $S \setminus V(H_1)$. In the following steps of our procedure, whose goal is to construct the reservoir path P, we have to be careful to avoid destroying this connectedness of the start k-tuple to $S \setminus P$ (in order to obtain Property (c) of Lemma 17). Hence we shall now fix witnesses of this connectedness and avoid using these vertices in the following. So let $Z \subseteq S \setminus V(H_1)$ be a vertex set that witnesses that the start k-tuple $(x_1, \ldots, x_k)$ of H_1 is $(\frac{1}{8}, p)$-connected to $S \setminus V(H_1)$ with $|Z| \leq \frac{1}{8}(p/2)|S|$, which is possible by Remark 6. We call the vertices in Z and $V(H_1)$ *used*, and all other vertices of $R \cup S$ *unused*.

Now for each $2 \leq i \leq |R|$ in succession, we perform the following procedure. Let S_i' be the unused vertices of S. Let r_i be an unused vertex of R with fewest neighbours in S_i'. We claim (and justify below) that r_i has at least $\beta\delta pn/8$ neighbours in S_i', and that $|S_i'| \geq \delta n/2 + 100k$. On this assumption, we can apply Lemma 18 to obtain a k-reservoir graph H_i with reservoir vertex r_i and remaining vertices in S_i'. By construction the end tuple of H_{i-1} is $(\frac{1}{2}, p)$-connected to S_i' and R^*, and hence by the choice of ε also $(\frac{1}{4}, p)$-connected to $S_i' \setminus V(H_i)$, as is the start tuple of H_i. Since $|S_i' \setminus V(H_i)| \geq \delta n/2$, we can apply Lemma 12 to find a connecting k-path of length at most $7k$ from the end tuple of H_{i-1} to the start tuple of H_i whose remaining vertices are in $S_i' \setminus V(H_i)$. We mark this k-path, r_i and $V(H_i)$ as used.

Assuming we successfully complete the above procedure to obtain a k-path P, Lemma 17 follows. Indeed, the path P then covers all vertices of R and certainly uses at most $50k|R|$ vertices, since the k-reservoir graph and connecting k-path that we construct at each step (except the first, where we only construct the k-reservoir graph) contain at most $50k$ vertices, hence we obtain property (a) of Lemma 17. Property (b) follows from the definition of a k-reservoir graph and the fact that we created for each vertex $r \in R$ one of these reservoir graphs with reservoir vertex r and connected them to form P. Property (c) follows by observing that $Z \cap V(P) = \emptyset$, witnessing that the start k-tuple of P is $(\frac{1}{8}, p)$-connected to $S \setminus V(P)$. Moreover, the end tuple of P is the end tuple of $H_{|R|}$, which by

construction and choice of ε is $(\frac{1}{8}, p)$-connected to $S \setminus V(P)$. Finally we have property (d) because each H_i is $(\frac{1}{2}, p)$-connected to R^*.

It remains only to justify our assumptions that r_i has at least $\beta\delta pn/8$ neighbours in S_i' and that $|S_i'| \geq \delta n/2 + 23k$ at each step i. The latter clearly follows from the facts that $|S| \geq \delta n$, that $|P| \leq 50k|R| \leq \delta n/4$, that $|Z| \leq \frac{1}{8}(p/2)|S|$ and that n is sufficiently large due to the choice of ε and Remark 7. For the former, suppose for contradiction that at step i in the above procedure, we find that the vertex r_i has less than $\beta\delta pn/8$ unused neighbours in S_i'. Since r_i has at least $\beta\delta pn/2$ neighbours in S, it follows that $i \geq \frac{3}{8}\beta\delta pn/(7k + \max(47, 2k+1)) > \beta\delta pn/(100k)$. Furthermore, at the step $i' := i - \beta\delta pn/(100k) + 1$, certainly r_i had at most $\beta\delta pn/8 + 50k(i - i') < \beta\delta pn/2$ unused neighbours, since only $50k(i' - i)$ vertices were used in the intervening steps. As the vertex r_i was not chosen at any of the steps $i' \leq j < i$, it follows that at each step r_j had at most as many unused neighbours as r_i, and in particular less than $\beta\delta pn/2$ unused neighbours. We conclude that each of the vertices $Q := \{r_{i'}, \ldots, r_i\}$ has less than $\beta\delta pn/2 < \delta pn/4$ neighbours in S_i', and thus $e(Q, S_i') \leq \delta pn|Q|/4 < (1-\varepsilon)p|Q||S|$. Since $|S| \geq \delta n/2 > \varepsilon n$, and $|Q| = i - i' = \beta\delta pn/(100k) > \varepsilon pn$, and since G is $(\varepsilon, p, 0, 1)$-pseudorandom, this is a contradiction to (3). It follows that our assumptions are justified, which completes the proof. $\qquad\square$

It remains to prove Lemma 18. Again we split the proof into the two cases $k = 2$, and $k \geq 3$. In the case $k \geq 3$ a rather straightforward construction works, which generalises the example of the triangle we gave above: Our k-reservoir graph has $2k + 1$ vertices, and consists simply of a $2k$-vertex k-path in S all of whose vertices are adjacent to r.

Proof of Lemma 18 for $k \geq 3$. We set $\varepsilon = \beta\delta^2/(1600k^2)$. Our goal is to construct a $2k$-vertex k-path in $S \cap N(r)$. Obviously, such a k-path together with r forms a k-reservoir graph with reservoir vertex r. So let $X_1, \ldots, X_{2k}$ be any collection of pairwise disjoint subsets of $S \cap N(r)$, each of size $\beta\delta pn/(16k)$, which we can choose because $|N(r) \cap S| \geq \beta\delta pn/8$. We now construct the desired k-path by choosing one vertex from each of the X_i. However, we have to bear in mind that we also want the start and the end k-tuple of this path to be $(\frac{1}{2}, p)$-connected to both R^* and S (which is why we impose conditions (27)–(30) below).

We shall call a vertex x *typical* with respect to a set Y if $|N(x) \cap Y| \geq (1-\varepsilon)p|Y|$, and atypical otherwise. Since G is $(\varepsilon, p, k, k+1)$-pseudorandom Lemma 9 implies that for each $|Y| \geq \varepsilon p^k n$ less than $\varepsilon p^{k+1} n$ vertices of G are atypical with respect to Y.

Firstly, for each $1 \leq i \leq k$, in order, we choose any $x_i \in X_i \cap N(x_1, \ldots, x_{i-1})$ which is typical with respect to each of the following at most $3k$ sets:

$$R^*, R^* \cap N(x_1), \ldots, R^* \cap N(x_1, \ldots, x_{i-1}), \tag{27}$$

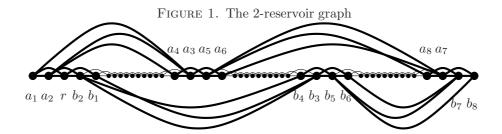$$S, S \cap N(x_1), \ldots, S \cap N(x_1, \ldots, x_{i-1}), \tag{28}$$

$$X_{i+1} \cap N(x_1, \ldots, x_{i-1}), \ldots, X_{k+1} \cap N(x_1, \ldots, x_{i-1}) \quad \text{and}$$

$$X_{k+2} \cap N(x_2, \ldots, x_{i-1}), \ldots, X_{k+i-1} \cap N(x_{i-1}), X_{k+i}.$$

Secondly, for each $k+1 \leq i \leq 2k$, in order, we choose an $x_i \in X_i \cap N(x_{i-k}, \ldots, x_{i-1})$ which is typical with respect to each of the following at most $3k$ sets:

$$R^*, R^* \cap N(x_{k+1}), \ldots, R^* \cap N(x_{k+1}, \ldots, x_{i-1}), \tag{29}$$

$$S, S \cap N(x_{k+1}), \ldots, S \cap N(x_{k+1}, \ldots, x_{i-1}) \quad \text{and} \tag{30}$$

$$X_{i+1} \cap N(x_{i-k}, \ldots, x_{i-1}), \ldots, X_{2k} \cap N(x_{i-k}, \ldots, x_{i-1}).$$

FIGURE 1. The 2-reservoir graph

Clearly vertices $x_1, \ldots, x_{2k}$ chosen in this way form a k-path. Moreover, choosing these vertices is possible for the following reasons. Since $|X_i| \geq \beta\delta pn/(16k)$ and $|S| \geq |R^*| \geq \delta^2 n/(200k)$ and by typicality in earlier steps, the smallest sets to which we require typicality are $X_\ell \cap N(x_1, \ldots, x_{k-1})$ and $X_\ell \cap N(x_{i-k}, \ldots, x_{i-1})$ for certain values of ℓ. Each of these sets involve the joint neighbourhood of $k-1$ vertices in one of the X_ℓ, hence (by typicality in earlier steps) these sets are of size at least $(1-\varepsilon)^{k-1}p^{k-1}|X_\ell| \geq (1-\varepsilon)^{k-1}p^k\beta\delta n/(16k) \geq \varepsilon p^k n$. Thus none of the sets to which we require typicality are smaller than $\varepsilon p^k n$. In addition, there are at most $3k$ sets to which we require typicality, forbidding at most $3k\varepsilon p^{k+1}n$ vertices for the choice of x_i, out of $\left((1-\varepsilon)p\right)^k|X_i| \geq (1-\varepsilon)^k\beta\delta^2 p^{k+1}n/(200k) > 4k\varepsilon p^{k+1}n$ vertices.

In order to show that $(x_k, \ldots, x_1)$ is moreover $(\frac{1}{2}, p)$-connected to R^* we need to check that $\deg_{R^*}(x_j, \ldots, x_1) \geq \frac{1}{2}(\frac{p}{2})^j|R^*|$ for each $j \in [k]$. Indeed, it follows from (27) that $\deg_{R^*}(x_j, \ldots, x_1) \geq (1-\varepsilon)^j p^j|R^*| > \frac{1}{2}p^j|R^*|$. Similarly, $(x_k, \ldots, x_1)$ is $(\frac{1}{2}, p)$-connected to S, and $(x_{k+1}, \ldots, x_{2k})$ is $(\frac{1}{2}, p)$-connected to both R^* and S. $\qquad\square$

In the case $k = 2$ we need to work with weaker pseudorandomness conditions and thus use a more involved construction, illustrated in Figure 1. In this construction there is a 2-path from a_1a_2 to b_7b_8 using all vertices in the left-to-right order, and a second, using all vertices but the reservoir vertex r, which starts $a_1a_2a_3a_4$ then goes "left" to $b_1b_2b_3b_4$, and so on to finish $b_6b_5b_7b_8$.

We will call the subgraph of the 2-reservoir graph induced by $\{a_1, \ldots, a_8\} \cup \{b_1, \ldots, b_8\} \cup \{r\}$ the *spine* of the reservoir graph (that is, the graph on the large vertices and with the thick edges in Figure 1). The construction of this graph uses similar ideas and a similar pigeonhole argument as the construction of a connecting 2-path for the Connection lemma.

Proof of Lemma 18 for $k = 2$. We let $\varepsilon \leq \beta^2\delta^2/10^6$ be small enough so that we can apply Lemma 12 with input $k = 2$ and $\delta/4$. Let G, S, R^* and $r \in V(G) \setminus S$ with

$$|S| \geq \frac{\delta n}{2}, \quad |R^*| \geq \frac{\delta^2 n}{400}, \quad |N_S(r)| \geq \frac{\beta\delta pn}{8}$$

be given. In this proof we call a vertex x *typical* with respect to a set Y if $\left|N_Y(x)\right| \geq (1-\varepsilon)p|Y|$.

Our goal is to find in S a copy of the 2-reservoir graph depicted in Figure 1. In a first step we now construct the five-vertex 2-path (a_1, a_2, r, b_2, b_1), sufficiently well-connected to S and R^*. For this we choose distinct vertices in the following

order:

$$a_1 \in N_S(r) \qquad \text{typical with respect to } S, R^*, N_S(r),$$
$$a_2 \in N_S(r, a_1) \qquad \text{typical with respect to } S, R^*, N_S(r), N_S(a_1), N_{R^*}(a_1),$$
$$b_2 \in N_S(r, a_2) \qquad \text{typical with respect to } S, N_S(r),$$
$$b_1 \in N_S(r, b_2) \qquad \text{typical with respect to } S, N_S(b_2).$$

This is possible by $(\varepsilon, p, 1, 2)$-pseudorandomness of G, since in each case we choose from a set of vertices of size at least

$$\min\left\{|N_S(r, a_1)|, |N_S(r, a_2)|, |N_S(r, b_2)|\right\} \geq (1 - \varepsilon)\tfrac{1}{8}\beta\delta p^2 n - 3 > \varepsilon p^2 n,$$

by typicality in earlier choices. Moreover, we require typical behaviour only to at most five sets each of size at least $(1-\varepsilon)\beta\delta^2 pn/400 > \varepsilon pn$, where the left hand side of this calculation is given by the lower bounds on $|N_S(r)|$ and $|N_{R^*}(a_1)|$. Now it is easy to check that (a_1, a_2, r, b_2, b_1) forms a 2-path in G, that (a_1, a_2), (a_2, a_1), (b_1, b_2) and (b_2, b_1) are $(\tfrac{1}{2}, p)$-connected to S, and (a_2, a_1) is $(\tfrac{1}{2}, p)$-connected to R^*.

Our second step is to construct the remainder of the spine of the reservoir graph. Let us first investigate the structure of this graph more closely and describe our strategy. First we will construct candidates for the induced subgraph of the spine on $a_1, \ldots, a_8$. Note that each edge of this subgraph is contained in one of the three 2-paths

$$(a_1, a_2, a_3, a_4), \quad (a_4, a_3, a_5, a_6), \quad (a_6, a_5, a_7, a_8).$$

Then we will construct candidates for the induced subgraph on $b_1, \ldots, b_8$. For this observe that the adjacencies among the vertices $a_1, \ldots, a_8$ and those among $b_1, \ldots, b_8$ are identical. Finally, in order to complete the spine, we need only in addition to guarantee that (a_8, a_7, b_7, b_8) is a 2-path. We will show that we can choose vertices among our candidates so that this is satisfied. When choosing the various candidates we have to keep in mind that we will want (b_7, b_8) to be $(\tfrac{1}{2}, p)$-connected to S and R^* as required in the conclusion of our lemma. In addition we want to complete the constructed spine to obtain the whole reservoir graph in S with the help of the Connection lemma. Hence we will also need that

$$\begin{aligned} (a_3, a_4), \quad (a_5, a_6), \quad (a_7, a_8) \\ (b_3, b_4), \quad (b_5, b_6), \quad (b_7, b_8) \end{aligned} \tag{31}$$

are $(\tfrac{1}{2}, p)$-connected to S.

We now first define in S twelve disjoint parts $X_3, \ldots, X_8, X_3', \ldots, X_8'$. We will then find candidates (with the required properties) for a_i in X_i and for b_i in X_i', $i \in \{3, \ldots, 8\}$. So choose

$$X_3 \subseteq N_S(a_1, a_2) \setminus \{a_1, a_2, b_1, b_2\} \qquad\qquad \text{with} \quad |X_3| = p^2|S|/20,$$
$$X_3' \subseteq N_S(b_1, b_2) \setminus \left(\{a_1, a_2, b_1, b_2\} \cup X_3\right) \qquad \text{with} \quad |X_3'| = p^2|S|/20,$$
$$X_4 \subseteq N_S(a_2) \setminus \left(\{a_1, a_2, b_1, b_2\} \cup X_3 \cup X_3'\right) \qquad \text{with} \quad |X_4| = p|S|/20,$$
$$X_4' \subseteq N_S(b_2) \setminus \left(\{a_1, a_2, b_1, b_2\} \cup X_3 \cup X_3' \cup X_4\right) \qquad \text{with} \quad |X_4'| = p|S|/20.$$

Note that these sets exist because (a_1, a_2) and (b_1, b_2) are $(\tfrac{1}{2}, p)$-connected to S. Further, choose

$$X_5, \ldots, X_8, X_5', \ldots, X_8' \subseteq S \setminus \left(\{a_1, a_2, b_1, b_2\} \cup X_3 \cup X_4 \cup X_3' \cup X_4'\right)$$

each of size $|S|/20$ and pairwise disjoint.

Let us now turn to the candidate sets for the vertices a_i. The construction of these sets is somewhat technically intricate, as we will proceed differently for different vertices a_i. It might help the reader to keep in mind though that in each

instance the main purpose of the definition of a candidate set is to guarantee the necessary adjacencies and well-connectedness.

We start with the candidate set for a_3. For its choice observe that a_3 is adjacent to a_4, a_5 and a_6 and recall that we want (a_3, a_4) to be $(\frac{1}{2}, p)$-connected to S. Thus let $A_3 \subseteq X_3$ be those vertices with at least $p|X_4|/2 = p^2|S|/40$ neighbours in X_4, at least $p|X_5|/2 = p|X_6|/2 = p|S|/40$ neighbours in each of X_5 and X_6, and at least $p|S|/2$ neighbours in S. By Lemma 9 we have $|A_3| \geq |X_3| - 4\varepsilon p^2 n > 0$. We fix a vertex $a_3 \in A_3$. By the definition of X_3 this a_3 is adjacent to a_1 and a_2 as required.

We next define the candidate set for a_4. We will choose this candidate set in $N_{X_4}(a_3)$, so again by definition, all candidates for a_4 will be adjacent to a_2 and a_3 as required. Moreover, a_4 should be adjacent jointly with a_3 to a_5, and (a_3, a_4) should be well-connected to S. So let $A_4 \subseteq N_{X_4}(a_3)$ be those vertices with at least $p^2|S|/80$ neighbours in $N_{X_5}(a_3)$, with at least $p|S|/2$ neighbours in S, and at least $p^2|S|/4$ neighbours in $N_S(a_3)$. By the definition of A_3 we have $|N_{X_5}(a_3)| \geq p|S|/40$ and $|N_S(a_3)| \geq p|S|/2$, and therefore Lemma 9 implies

$$|A_4| \geq |N_{X_4}(a_3)| - 3\varepsilon p^2 n \geq p^2|S|/40 - 3\varepsilon p^2 n \geq p^2|S|/80, \qquad (32)$$

where the second inequality follows from the definition of A_3. We do not choose a_4 from this candidate set immediately, but first define further candidate sets.

We will choose the candidate set for a_5 in $N_{X_5}(a_3)$ and we will select only vertices which have some neighbour in A_4. This will guarantee that a_5 is connected to a_3 as required and that, once we choose a vertex a_5 we can also choose a valid vertex a_4. In addition a_5 should be adjacent jointly with a_3 to a_6, adjacent to a_7 and a_8, and (a_5, a_6) should be $(\frac{1}{2}, p)$-connected to S. So let $A_5 \subseteq N_{X_5}(a_3)$ be those vertices with a neighbour in A_4, with at least $p^2|S|/80$ neighbours in $N_{X_6}(a_3)$, with at least $p|X_7|/2 = p|X_8|/2 = p|S|/40$ neighbours in both X_7 and X_8, and with at least $p|S|/2$ neighbours in S. Observe that by (32) the set A_4 is of size $\Omega(p^2 n)$ only, and so Lemma 9 only guarantees that at most $\varepsilon p n$ vertices in $N_{X_5}(a_3)$ do not have a neighbour in A_4. The remaining requirements for A_5 however involve only sets of size $\Omega(p n)$ because $|N_{X_6}(a_3)| \geq p|S|/40$ by the definition of A_3 and so Lemma 9 implies

$$|A_5| \geq |N_{X_5}(a_3)| - \varepsilon p n - 4\varepsilon p^2 n \geq p|S|/40 - \varepsilon p n - 4\varepsilon p^2 n \geq p|S|/80,$$

where the second inequality follows from the definition of A_3. Again, we will choose a_5 later, but next define for each possible choice of $a_5 \in A_5$ candidate sets $A_6(a_5)$ and $A_7(a_5)$ for a_6 and a_7.

For each $a_5 \in A_5$ we will now define a candidate set $A_6(a_5) \subseteq N_{X_6}(a_3, a_5)$, which again guarantees the correct adjacencies of a_6 to vertices with smaller indices. In addition, a_6 should be adjacent jointly with a_5 to a_7 and (a_5, a_6) should be $(\frac{1}{2}, p)$-connected to S. So for each $a_5 \in A_5$ let $A_6(a_5) \subseteq N_{X_6}(a_3, a_5)$ be those vertices with at least $p^2|S|/80$ neighbours in $N_{X_7}(a_5)$, at least $p|S|/2$ neighbours in S, and at least $p^2|S|/4$ neighbours in $N_S(a_5)$. Again, by the definition of A_3 and A_5 and Lemma 9 we have

$$|A_6(a_5)| \geq |N_{X_6}(a_3, a_5)| - 3\varepsilon p^2 n \geq p^2|S|/80 - 3\varepsilon p^2 n \geq p^2|S|/160.$$

Similarly, we have to guarantee that candidates for a_7 that are adjacent to a_5 have a valid choice for a neighbour a_6. Moreover, a_7 should be adjacent jointly with a_5 to a_8, and (a_7, a_8) should be $(\frac{1}{2}, p)$-connected to S. In addition we will guarantee that (a_7, a_8) is also $(\frac{1}{2}, p)$-connected to R^*. We do not actually need this for (a_7, a_8), but we will need it for (b_7, b_8); and since we want to construct the candidate sets for the subgraph on $b_3, \ldots, b_8$ analogously we will require it here. So for each $a_5 \in A_5$, let $A_7(a_5) \subseteq N_{X_7}(a_5)$ be those vertices with a neighbour in $A_6(a_5)$, with at least

$p^2|S|/80$ neighbours in $N_{X_8}(a_5)$, with at least $p|S|/2$ neighbours in S, and at least $p|R^*|/2$ neighbours in R^*. Because $|A_6(a_5)| \geq p^2|S|/160$ and $|N_{X_8}(a_5)| \geq p|S|/40$ by the definition of A_5, Lemma 9 implies again

$$|A_7(a_5)| \geq |N_{X_7}(a_5)| - \varepsilon pn - 3\varepsilon p^2 n \geq p|S|/40 - \varepsilon pn - 3\varepsilon p^2 n \geq p|S|/80\,.$$

Set $A_7 := \bigcup_{a_5 \in A_5} A_7(a_5)$. Then each vertex $a_5 \in A_5$ has at least $|A_7(a_5)| \geq p|S|/80$ neighbours in A_7, and so by (3) we have

$$|A_5|p|S|/80 \leq e(A_5, A_7) \leq (1+\varepsilon)p|A_5||A_7|\,,$$

and thus

$$|A_7| \geq |S|/160\,. \tag{33}$$

So we have a candidate set for a_7 which is of linear size. This will be crucial for the pigeonhole argument below.

Finally we can turn to the definition of candidate sets for a_8, which needs to be adjacent to a_5 and a_7 and such that (a_7, a_8) is well-connected to S and R^*. So for each $a_5 \in A_5$ and $a_7 \in A_7(a_5)$, let $A_8(a_5, a_7) \subseteq N_{X_8}(a_5, a_7)$ be those vertices with at least $p|S|/2$ neighbours in S, at least $p|R^*|/2$ neighbours in R^*, at least $p^2|S|/4$ common neighbours with a_7 in S and at least $p^2|R^*|/4$ common neighbours with a_7 in R^*. By Lemma 9 and the definition of A_5 and $A_7(a_5)$ we have

$$|A_8(a_5, a_7)| \geq |N_{X_8}(a_5, a_7)| - 4\varepsilon p^2 n \geq p^2|S|/80 - 4\varepsilon p^2 n \geq p^2|S|/160\,. \tag{34}$$

This defines all the candidate sets for the vertices a_i.

By the symmetry of the a_i and b_i we can carry out the same construction in the sets $X_3', \ldots, X_8'$ to obtain b_3 and candidate sets B_4, B_5, $B_6(b_5)$, $B_7(b_5)$, B_7, $B_8(b_5, b_7)$ for $b_4, \ldots, b_8$.

For completing the spine it now remains to use these candidate sets to find an edge $a_7 b_7$ such that we can choose valid vertices $a_3, \ldots, a_6, a_8$ and $b_3, \ldots, b_6, b_8$ in the respective candidate sets and moreover (a_8, a_7, b_7, b_8) forms a 2-path. Thus we would in particular like to require that the vertex b_7 in this edge has some neighbour in a candidate set for a_8 defined by a_7. However, the candidate sets for a_8 depend in addition on vertices $a_5 \in A_5$. This motivates the following definition. We call an edge $a_7 b_7$ with $a_7 \in A_7$ and $b_7 \in B_7$ *good with respect to* $a_1 a_2$ if b_7 has a neighbour in $A_8(a_5, a_7)$ for some $a_5 \in A_5$ with $a_7 \in A_7(a_5)$. Observe that by the definition of A_7 for each $a_7 \in A_7$, there indeed exists such an a_5. Similarly, $b_7 a_7$ is *good with respect to* $b_1 b_2$ if a_7 has a neighbour in $B_8(b_5, b_7)$ for some $b_5 \in B_5$ with $b_7 \in B_7(b_5)$.

Observe now that for each $a_5 \in A_5$ and $a_7 \in A_7(a_5)$ by (34) and Lemma 9 at most εpn vertices in B_7 have no neighbour in $A_8(a_5, a_7)$. Hence for all but at most εpn vertices b_7 in B_7 the edge $a_7 b_7$ is good with respect to $a_1 a_2$. Similarly for all but at most εpn vertices a_7 in A_7 the edge $b_7 a_7$ is good with respect to $b_1 b_2$. By (3) there are at least $(1-\varepsilon)p|A_7||B_7|$ edges between A_7 and B_7, of which all but at most $\varepsilon pn\big(|A_7| + |B_7|\big)$ are good with respect to both $a_1 a_2$ and $b_1 b_2$. Since

$$(1-\varepsilon)p|A_7||B_7| \overset{(33)}{\geq} p|S|^2/51200 > \varepsilon pn^2 \geq \varepsilon pn\big(|A_7| + |B_7|\big)$$

we can choose such a good edge $a_7 b_7$.

We will now complete this good edge and the already chosen a_3, b_3 to a copy of the spine. Since $a_7 b_7$ is good with respect to $a_1 a_2$ there exists a vertex $a_5 \in A_5$ such that $a_7 \in A_7(a_5)$ and b_7 has a neighbour a_8 in $A_8(a_5, a_7)$. Fix such vertices a_5 and a_8. Similarly, using goodness with respect to $b_1 b_2$, we fix b_5 and b_8. This readily implies by the definition of the candidate sets that $(a_8, a_7, b_8, b_7))$ forms a 2-path and that a_5 is adjacent to a_7 and a_8 (and to a_3), and similarly for b_5. Next, by the definition of $A_7(a_5)$ we have that a_7 has a neighbour a_6 in $A_6(a_5)$, which we fix. Again, this implies that a_6 is adjacent to a_5 and a_7. Moreover, by the definition

of A_5 we have that a_5 has a neighbour a_4 in A_4, which we fix and which is thus adjacent to a_5 and by the definition of A_4 to a_3 and a_2. Similarly, we fix b_6 and b_4.

This completes the construction of the spine on $a_1, \ldots, a_8, b_1, \ldots, b_8$. In this construction we have guaranteed that (a_1, a_2) and (b_7, b_8) are $(\frac{1}{2}, p)$-connected to R and S as required and that the pairs from (31) are $(\frac{1}{2}, p)$-connected to S.

In our final step, which will complete the reservoir graph, we will apply Lemma 12 three times, to connect (b_2, b_1) to (a_4, a_3), (a_5, a_6) to (b_4, b_3), and (b_5, b_6) to (a_8, a_7). More precisely let $S' = S \setminus \{a_1, \ldots, a_8, b_1, \ldots, b_8\}$. By Remark 7 we have $|S'| \geq \delta n/2 - 16 \geq \delta n/4$. We apply Lemma 12 with $k = 2$ and $\delta/4$ to find a ten-vertex squared path P_1 joining $b_2 b_1$ to $a_4 a_3$ in S', then again to find P_2 joining $a_5 a_6$ to $b_4 b_3$ in $S' \setminus P_1$, and once more to find P_3 joining $b_5 b_6$ to $a_8 a_7$ in $S' \setminus (P_1 \cup P_2)$. This is possible since all these pairs are $(\frac{1}{2}, p)$-connected to S, and, since $p^2|S| \geq 100$, even after removing the at most 47 vertices of the partially constructed reservoir graph we have $(\frac{1}{4}, p)$-connectedness.

To conclude we have that

$$a_1 a_2 r b_2 b_1 P_1 a_4 a_3 a_5 a_6 P_2 b_4 b_3 b_5 b_6 P_3 a_8 a_7 b_7 b_8$$

is a squared path on 47 vertices from $a_1 a_2$ to $b_7 b_8$ using r. On the other hand, by taking each P_i in the opposite direction to the previous squared path we obtain that

$$a_1 a_2 a_3 a_4 P_1 b_1 b_2 b_3 b_4 P_2 a_6 a_5 a_7 a_8 P_3 b_6 b_5 b_7 b_8$$

is a squared path from $a_1 a_2$ to $b_7 b_8$ which uses every vertex of the previous squared path except r. Hence we have constructed the desired 2-reservoir graph with reservoir vertex r. □

5. Enumerating powers of Hamilton cycles

To prove Theorem 5 we would ideally like to show that we can construct the kth power of a Hamilton cycle vertex by vertex, and that when we have t vertices remaining uncovered, we have at least $(1 - \nu)p^k t$ choices for the next vertex; then the theorem would follow immediately. However, we obviously do not construct kth powers of Hamilton cycles in this way: we have very little control over choice in constructing the reservoir paths and connecting paths. Moreover for the promised number $(1 - \nu)^n p^{kn} n!$ of Hamilton cycles powers even the Extension lemma, Lemma 8, does not provide the desired number of choices in the greedy portion of the construction where we do choose one vertex at a time. (We remark though that the proof of this lemma, together with the rest of our proof does immediately provide us with $c^n p^{(1-\nu)kn}((1 - \nu)n)!$ Hamilton cycle powers for some absolute constant $c > 0$.)

Thus we have to upgrade the Extension lemma in two ways. Firstly, we have to modify it to give us more choices in each step (after a few initial steps). Secondly, it turns out that to obtain the desired number of Hamilton cycles powers we have to apply the Extension lemma for longer, that is, the leftover set will in the end only contain $\mathcal{O}(n/(\log n)^2)$ vertices. Thus we have to change the Extension lemma to deal with this different situation. This comes at the cost of slightly tightening the pseudorandomness requirement.

It is not hard to check that such an upgrade is possible. In the lemma below we will guarantee that for an end k-tuple $\boldsymbol{x}$ of a k-path there are $\left(1 - \frac{\nu}{2k}\right) \deg_L(\boldsymbol{x})$ valid extensions, where L is the current set of leftover vertices. As we will argue below, this will provide us with the right number of Hamilton cycle powers if we can guarantee in addition that $\deg_L(\boldsymbol{x}) \geq \left((1 - \frac{\nu}{2k})p\right)^k |L|$. Recall however that we will want to use this lemma after constructing the reservoir path with the help of Lemma 10, which guarantees $(\frac{1}{8}, p)$-connectedness to L, a property which only

gives a weaker lower bound on $\deg_L(\boldsymbol{x})$ than desired. In order to overcome this shortcoming we will in the first few applications of the Counting version of the Extension lemma transform this $(\frac{1}{8}, p)$-connectedness to a stronger property which gives the desired bound. Conditions (ii) and (iii), and conclusions (b) and (c) take care of this.

Lemma 19 (Counting version of the Extension lemma)**.** *Given $k \geq 2$ and $\nu > 0$, if $C = 2^{k+23}k^4/\nu$ then the following holds. Let $0 < p < 1$ and G be an $\left(1/(C \log n)^2, p, k - 1, k\right)$-pseudorandom graph on n vertices. Let L and R be disjoint vertex sets with $|L|, |R| \geq n/(200k \log n)^2$. Suppose that there is $0 \leq j \leq k$ such that $\boldsymbol{x} = (x_1, \ldots, x_k)$ satisfies*

  (i) $\boldsymbol{x}$ *is $\left(\frac{1}{8}, p\right)$-connected to R,*

  (ii) $\deg_L(x_i, \ldots, x_k) \geq \frac{1}{8}\left(\frac{p}{2}\right)^{k-i+1}|L|$ *for each $1 \leq i \leq j$,*

  (iii) $\deg_L(x_i, \ldots, x_k) \geq \left((1 - \frac{\nu}{2k})p\right)^{k-i+1}|L|$ *for each $j < i \leq k$.*

Then at least $\left(1 - \frac{\nu}{2k}\right)\deg_L(x_1, \ldots, x_k)$ vertices $x_{k+1} \in N_L(x_1, \ldots, x_k)$ satisfy that

  (a) $(x_2, \ldots, x_{k+1})$ *is $\left(\frac{1}{6}, p\right)$-connected to R,*

  (b) $\deg_L(x_{i+1}, \ldots, x_{k+1}) \geq \frac{1}{8}\left(\frac{p}{2}\right)^{k-i+1}|L|$ *for each $1 \leq i \leq j - 1$,*

  (c) $\deg_L(x_{i+1}, \ldots, x_{k+1}) \geq \left((1 - \frac{\nu}{2k})p\right)^{k-i+1}|L|$ *for each $j - 1 < i \leq k$.*

Sketch of proof. In this proof we say that a vertex x is *typical* with respect to S if $\deg_S(x) \geq \left(1 - \frac{\nu}{2k}\right)p|S|$. Let $X_{k+1} \subseteq N_L(x_1, \ldots, x_k)$ be the set of all vertices typical with respect to $N_L(x_i, \ldots, x_k)$ for each $2 \leq i \leq k$, to $R \cap N(x_i, \ldots, x_k)$ for each $2 \leq i \leq k$, and to L and R. By Lemma 9 the number of vertices which fail any one of these conditions is at most $4knp^k/(C \log n)^2 < \nu p^k|L|/(2k \cdot 2^{-k-4})$. $\square$

Sketch of proof of Theorem 5. We follow essentially the proof of Theorem 2. We make the following alterations. We require the same pseudorandomness of G as in that theorem, *except* that we set $C = 2^{k+23}k^4/\nu$ and choose $\varepsilon \leq 1/(C \log n)^2$ such that any $n/(\log n)^2$-vertex induced subgraph of G still meets the pseudorandomness requirements of Theorem 2. By equation (3), this means we can choose $\varepsilon = \Theta\left(1/(\log n)^2\right)$.

We construct a reservoir set R of size between $n/(\log n)^2$ and $10n/(\log n)^2$, with the properties that all vertices of R have at least $\beta pn/2$ neighbours in $V(G) \setminus R$ and all vertices of $V(G) \setminus R$ have at least $\frac{1}{2}\beta p|R|$ neighbours in R. It is not hard to check that the same construction procedure works.

The application of Lemma 10 to obtain a reservoir path P of size $50k|R|$ covering R needs to change only in that we have to guarantee $(\frac{1}{2}, p)$-connection of the ends of P to R. The choice of ε obviously allows this.

We now use Lemma 19 instead of Lemma 8 to extend P greedily. In the first k steps we let the input j decrease from k to 1, while at the $(k+1)$st application and thereafter, we take $j = 0$. We continue until we can no longer use Lemma 19, i.e. at the step when we have constructed P' and the number of vertices not in $R \cup P'$ is less than $n/(200k \log n)^2$. We set L to be these leftover vertices.

The remainder of the proof is identical to that of Theorem 2: that is, we apply Lemma 11 to construct a path P'' covering L and using only vertices from $L \cup R$, and connect P'' and P' using Lemma 12. The choice of ε ensures that we can do this since $|L \cup R| > n/(\log n)^2$. We thus successfully construct the kth power of a Hamilton cycle in G.

Finally, by considering the choices only during the use of Lemma 19 with $j = 0$, we can estimate the number of kth powers of cycles which we construct are at least

$$\prod_{t=n/(200k\log n)^2}^{n-1000kn/(\log n)^2} \left(1 - \tfrac{\nu}{2k}\right)^k p^k t \geq \prod_{t=n/(200k\log n)^2}^{n-1000kn/(\log n)^2} \left(1 - \tfrac{\nu}{2}\right) p^k t$$

$$\geq \left(1 - \tfrac{\nu}{2}\right)^n p^{kn} n! n^{-1001kn/(\log n)^2}.$$

and since $n^{-1001kn/(\log n)^2} = \left(2^{-1001k/\log n}\right)^n > \left(1 - \tfrac{\nu}{2}\right)^n$ for sufficiently large n, the result follows. □

6. Concluding remarks

Hamilton cycles. For Hamilton cycles, a simple modification of our arguments for squared Hamilton cycles yields that $(\varepsilon, p, 0, 1)$-pseudorandom graphs with minimum degree βpn are Hamiltonian for sufficiently small $\varepsilon = \varepsilon(\beta)$. This bound is essentially best possible (for our notion of pseudorandomness) since the disjoint union of $G(n - pn, p)$ and K_{pn} is easily seen to be asymptotically almost surely $(\varepsilon, p, 0, 1 - \varepsilon)$-pseudorandom and have minimum degree at least $pn/2$. Although we do not know of any criterion for Hamiltonicity which implies this result, for more standard notions of pseudorandomness, which forbid the above somewhat unnatural construction, much stronger criteria are known, such as that in [25].

Improving the pseudorandomness requirements. It would be interesting to obtain stronger results on the pseudorandomness required to find kth powers of Hamilton cycles. We believe that a generalisation of our result for the $k = 2$ case is true.

Conjecture 20. *For all $k \geq 2$ the pseudorandomness requirement in Theorem 2 can be replaced by $(\varepsilon, p, k - 1, k)$-pseudorandomness.*

As remarked in the introduction even in the $k = 2$ case we do not know whether Theorem 2 is sharp. It would also be very interesting (albeit very hard) to find better lower bound examples than those mentioned in the introduction.

In the evolution of random graphs triangles, spanning triangle factors and squares of Hamilton cycles appear at different times: In $G(n, p)$ the threshold for triangles is $p = n^{-1}$, but only at $p = \Theta(n^{-2/3}(\log n)^{1/3})$ each vertex of $G(n, p)$ is contained in a triangle with high probability, which is also the threshold for the appearance of a spanning triangle factor [19]. Squares of Hamilton cycles on the other hand are with high probability not present in $G(n, p)$ for $p \leq n^{-1/2}$, and Kühn and Osthus [28] recently showed that for $p \geq n^{-1/2+\varepsilon}$ they are present. Our Theorem 2 is also applicable to random graphs, but the range is worse: $p \gg (\log n/n)^{1/3}$ for squares of Hamilton cycles and $p \gg (\log n/n)^{1/(2k)}$ for general kth powers of Hamilton cycles (recall that Riordan's result [30] implies the optimal bound $p \gg n^{-1/k}$ for $k \geq 3$).

Pseudorandom graphs behave differently. For (n, d, λ)-graphs it is known that there are triangle-free (n, d, λ)-graphs with $\lambda = cd^2/n$ for some c, but for 'small' c every vertex in an (n, d, λ)-graph is contained in a triangle (and, more generally, there exists a fractional triangle factor). This motivated Krivelevich, Sudakov and Szabó [27] to conjecture that indeed these graphs already contain a spanning triangle factor. We do not know whether triangle factors and squares of Hamilton cycles require differently strong pseudorandomness conditions.

Question 21. *Do spanning triangle factors and spanning 2-cycles appear for the same pseudorandomness requirements (up to constant factors)?*

Universality. For random graphs the study of when $G(n, p)$ asymptotically almost surely contains all spanning or almost spanning graphs with maximum degree bounded by a constant Δ was initiated in [6]. In this case $G(n, p)$ is also called universal for these graphs. The authors of [6] showed that $G(n, p)$ asymptotically almost surely contains all graphs on $(1 - \varepsilon)n$ vertices with maximum degree at most Δ if $p \geq Cn^{-1/\Delta} \log^{1/\Delta} n$. In [16] this result was extended to such subgraphs on n vertices. Recently, Conlon [13] announced that for the first of these two results he can lower the probability to $p = n^{-\varepsilon - 1/\Delta}$ for some (small) $\varepsilon = \varepsilon(\Delta) > 0$. The best known lower bound results from the fact that $p = \Omega(n^{-2/(\Delta+1)})$ is necessary for $G(n, p)$ to contain a $K_{\Delta+1}$-factor.

For pseudorandom graphs we were only recently able to establish universality results of this type, which follow from our work on a Blow-up lemma for pseudorandom graphs (see below). We can prove that $(p, cp^{\frac{3}{2}\Delta + \frac{1}{2}}n)$-jumbled graphs on n vertices with minimum degree βpn are universal for spanning graphs with maximum degree Δ [1]. We believe that these conditions are not optimal.

Question 22. *Which pseudorandomness conditions (plus minimum degree conditions) imply universality for spanning graphs of maximum degree Δ?*

It is worth noting that Alon and Capalbo [5] explicitly constructed almost optimally sparse universal graphs for spanning graphs with maximum degree Δ. These graphs have some pseudorandomness properties, but they also contain cliques of order $\log^2 n$, which random graphs of the same density certainly do not.

Additive structures in multiplicative subgroups. Alon and Bourgain [4] recently made use of properties of pseudorandom graphs in order to prove the following conjecture of Sun [31]. Given any prime $p \geq 13$, there is a cyclic ordering of the quadratic residues modulo p such that the sum of any two consecutive quadratic residues in the cyclic order is also a quadratic residue. In fact, Alon and Bourgain proved much more. It is not necessary to take the subgroup of $\mathbb{F}_p$ formed by the quadratic residues. Any sufficiently large multiplicative subgroup of $\mathbb{F}_p$ has the same property.

Theorem 23 (Alon and Bourgain [4], Theorem 1.2)**.** *There exists an absolute positive constant c such that for any prime power q and for any multiplicative subgroup A of the finite field $\mathbb{F}_q$ of size*

$$|A| = d \geq c \frac{q^{3/4}(\log q)^{1/2}(\log\log\log q)^{1/2}}{\log\log q}$$

there is a cyclic ordering $a_0, a_1, \ldots, a_{d-1}$ of the elements of A such that $a_i + a_{i+1} \in A$ for all i.

The proof of this theorem amounts to showing that a certain graph on vertex set A is pseudorandom and applying the Hamiltonicity result of Krivelevich and Sudakov [25] to find a Hamilton cycle in this graph, which defines the cyclic ordering. We can replace that result with Corollary 4 to obtain the following result, which in particular strengthens Proposition 1.6 of [4].

Corollary 24. *For each $k \geq 2$ there exists a positive constant c such that for any prime power q and multiplicative subgroup A of $\mathbb{F}_q$ of size*

$$|A| = d \geq \begin{cases} cq^{6/7} & k = 2 \\ cq^{(3k+1)/(3k+2)} & k \geq 3 \end{cases}$$

there is a cyclic ordering $a_0, a_1, \ldots, a_{d-1}$ of the elements of A such that $a_i + a_{i+j}$ is in A for all i and $1 \leq j \leq k$.

Blow-up lemmas. For dense graphs the Blow-up lemma [20] is a powerful tool for embedding spanning graphs with bounded maximum degree (versions of this lemma for certain graphs with a maximum degree not bounded by a constant have recently been developed in [9]). Already Krivelevich, Sudakov and Szabó [27] remark that their result on triangle factors in sparse pseudorandom graphs can be viewed as a first step towards the development of a Blow-up lemma for (subgraphs of) sparse pseudorandom graphs.

We see the results presented here as a further step in this direction. And in fact in recent work [1] we establish a blow-up lemma for spanning graphs with bounded maximum degree in sparse pseudorandom graphs. However, the pseudorandomness requirements for this more general result are more restrictive than those used here.

References

[1] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person, *Blow-up lemmas for sparse graphs*, in preparation.

[2] P. Allen, J. Böttcher, Y. Kohayakawa, and Y. Person, *Tight Hamilton cycles in random hypergraphs*, Random Structures Algorithms, to appear (`DOI: 10.1002/rsa.20519`).

[3] N. Alon, *Explicit Ramsey graphs and orthonormal labelings*, Electron. J. Combin. **1** (1994), Research Paper 12, 8 pp (electronic).

[4] N. Alon and J. Bourgain, *Additive patterns in multiplicative subgroups*, Geom. Funct. Anal. **24** (2014), no. 3, 721–739.

[5] N. Alon and M. Capalbo, *Sparse universal graphs for bounded-degree graphs*, Random Structures Algorithms **31** (2007), no. 2, 123–133.

[6] N. Alon, M. Capalbo, Y. Kohayakawa, V. Rödl, A. Ruciński, and E. Szemerédi, *Universality and tolerance (extended abstract)*, 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), IEEE Comput. Soc. Press, Los Alamitos, CA, 2000, pp. 14–21.

[7] N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., Hoboken, NJ, 2008, With an appendix on the life and work of Paul Erdős.

[8] B. Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 35–57.

[9] J. Böttcher, Y. Kohayakawa, A. Taraz, and A. Würfl, *An extension of the blow-up lemma to arrangeable graphs*, arXiv:1305.2059, 2013.

[10] F. Chung and R. Graham, *Sparse quasi-random graphs*, Combinatorica **22** (2002), no. 2, 217–244, Special issue: Paul Erdős and his mathematics.

[11] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), no. 4, 345–362.

[12] S. M. Cioabă, *Perfect matchings, eigenvalues and expansion*, C. R. Math. Acad. Sci. Soc. R. Can. **27** (2005), no. 4, 101–104.

[13] D. Conlon, talk at RSA 2013.

[14] D. Conlon, J. Fox, and Y. Zhao, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2014), 206–290.

[15] C. Cooper and A. M. Frieze, *On the number of Hamilton cycles in a random graph*, J. Graph Theory **13** (1989), no. 6, 719–735.

[16] D. Dellamonica Jr., Y. Kohayakawa, V. Rödl, and A. Ruciński, *An improved upper bound on the density of universal random graphs*, LATIN 2012: Theoretical informatics (Arequipa, 2012) (D. Fernández-Baca, ed.), Springer, 2012, pp. 231–242.

[17] R. Glebov and M. Krivelevich, *On the number of Hamilton cycles in sparse random graphs*, SIAM J. Discrete Math. **27** (2013), no. 1, 27–42.

[18] S. Janson, *The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph*, Combin. Probab. Comput. **3** (1994), no. 1, 97–126.

[19] A. Johansson, J. Kahn, and V. Vu, *Factors in random graphs*, Random Structures Algorithms **33** (2008), no. 1, 1–28.

[20] J. Komlós, G. N. Sárközy, and E. Szemerédi, *Blow-up lemma*, Combinatorica **17** (1997), no. 1, 109–123.

[21] J. Komlós and E. Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), no. 1, 55–63.

[22] A. D. Koršunov, *Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs*, Dokl. Akad. Nauk SSSR **228** (1976), no. 3, 529–532.

[23] _____, *Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in nonoriented graphs*, Diskret. Analiz (1977), no. 31 Metody Diskret. Anal. v Teorii Upravljajuscih Sistem, 17–56, 90.

[24] M. Krivelevich, *On the number of Hamilton cycles in pseudo-random graphs*, Electron. J. Combin. **19** (2012), no. 1, Paper 25, 14pp.

[25] M. Krivelevich and B. Sudakov, *Sparse pseudo-random graphs are Hamiltonian*, J. Graph Theory **42** (2003), no. 1, 17–33.

[26] _____, *Pseudo-random graphs*, More sets, graphs and numbers, Bolyai Soc. Math. Stud., vol. 15, Springer, Berlin, 2006, pp. 199–262.

[27] M. Krivelevich, B. Sudakov, and T. Szabó, *Triangle factors in sparse pseudo-random graphs*, Combinatorica **24** (2004), no. 3, 403–426.

[28] D. Kühn and D. Osthus, *On Pósa's conjecture for random graphs*, SIAM J. Discrete Math. **26** (2012), no. 3, 1440–1457.

[29] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), no. 4, 359–364.

[30] O. Riordan, *Spanning subgraphs of random graphs*, Combin. Probab. Comput. **9** (2000), no. 2, 125–148.

[31] Z. W. Sun, *Some new problems in additive combinatorics*, arXiv:1309.1679, 2013.

[32] A. Thomason, *Pseudo-random graphs*, Random graphs '85, Lect. 2nd Int. Semin., Poznań, Poland 1985, Ann. Discrete Math., vol. 33, 1987, pp. 307–331.

Department of Mathematics, London School of Economics, Houghton Street, London WC2A 2AE, U.K.

E-mail address: `p.d.allen@lse.ac.uk, j.boettcher@lse.ac.uk`

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil

E-mail address: `yoshi@ime.usp.br, hh@ime.usp.br`

Goethe-Universität, Institute of Mathematics, Robert-Mayer-Str. 10, 60325 Frankfurt, Germany

E-mail address: `person@math.uni-frankfurt.de`