An Algorithmic Version of the Blow-Up Lemma

János Komlós, Gabor N. Sarkozy,* Endre Szemerédi

[1]*Mathematics Department, Rutgers University, New Brunswick, NJ 08903; e-mail: komlos@math.rutgers.edu*

[2]*Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609; e-mail: gsarkozy@cs.wpi.edu*

[3]*Computer Science Department, Rutgers University, New Brunswick, NJ 08903; e-mail: szemered@cs.rutgers.edu*

Received 13 March 1997; revised 15 September 1997; accepted 17 November 1997

ABSTRACT: Recently we developed a new method in graph theory based on the regularity lemma. The method is applied to find certain spanning subgraphs in dense graphs. The other main general tool of the method, besides the regularity lemma, is the so-called blow-up lemma (Komlós, Sárközy, and Szemerédi [*Combinatorica*, **17**, 109–123 (1997)]. This lemma helps to find bounded degree spanning subgraphs in ε-regular graphs. Our original proof of the lemma is not algorithmic, it applies probabilistic methods. In this paper we provide an algorithmic version of the blow-up lemma. The desired subgraph, for an n-vertex graph, can be found in time $O(nM(n))$, where $M(n) = O(n^{2.376})$ is the time needed to multiply two n by n matrices with 0, 1 entires over the integers. We show that the algorithm can be parallelized and implemented in NC^5. Random Struct. Alg., 12, 297–312, 1998

Correspondence to: Gabor N. Sarkozy

*Part of this paper was written while Sarkozy was visiting MSRI Berkeley, as part of the Combinatorics Program. Research at MSRI is supported in part by NSF grant DMS-9022140.

1. INTRODUCTION

1.1. Notations and Definitions

All graphs are simple, that is, they have no loops or multiple edges. $v(G)$ is the number of vertices in G (order), $e(G)$ is the number of edges in G (size). $\deg(v)$ [or $\deg_G(v)$] is the degree of vertex v (within the graph G), and $\deg(v, Y)$ [or $\deg_G(v, Y)$] is the number of neighbors of v in Y. $\delta(G)$ and $\Delta(G)$ are the minimum degree and the maximum degree of G. $N(x)$ [or $N_G(x)$] is the set of neighbors of the vertex x, and $e(X, Y)$ is the number of edges between X and Y. A bipartite graph G with color-classes A and B and edges E will sometimes be written as $G = (A, B, E)$. For disjoint X, Y, we define the density,

$$d(X, Y) = \frac{e(X, Y)}{|X| \cdot |Y|}.$$

The density of a bipartite graph $G = (A, B, E)$ is the number,

$$d(G) = d(A, B) = \frac{|E|}{|A| \cdot |B|}.$$

For two disjoint subsets A, B of $V(G)$, the bipartite graph with vertex set $A \cup B$ which has all the edges of G with one endpoint in A and the other in B is called the pair (A, B).

A pair (A, B) is ε-regular if for every $X \subset A$ and $Y \subset B$ satisfying

$$|X| > \varepsilon |A| \quad \text{and} \quad |Y| > \varepsilon |B|,$$

we have

$$|d(X, Y) - d(A, B)| < \varepsilon.$$

A pair (A, B) is (ε, δ)-superregular if it is ε-regular and furthermore,

$$\deg(a) \geq \delta |B| \quad \text{for all } a \in A,$$

and

$$\deg(b) \geq \delta |A| \quad \text{for all } b \in B.$$

H is embeddable into G if G has a subgraph isomorphic to H, that is, if there is a one-to-one map (injection) φ: $V(H) \to V(G)$ such that $\{x, y\} \in E(H)$ implies $\{\varphi(x), \varphi(y)\} \in E(G)$.

As the model of computation we choose the weakest possible version of a *PRAM*, in which concurrent reads or writes of the same location are not allowed (*EREW*, see [21] for a discussion of the various *PRAM* models.) When researchers investigate the parallel complexity of a problem, the main question is whether a polyalgorithmic running time is achievable on a *PRAM* containing a polynomial number of processors. If the answer is positive then the problem and the corresponding algorithm are said to belong to class *NC* introduced in [34] (see also [10, 40]). When the running time is $O((\log n)^i)$, the algorithm is in NC^i.

1.2. An Algorithmic Version of the Blow-Up Lemma

In recent years the interaction between combinatorics and the theory of algorithms is getting stronger and stronger. It is therefore not surprising that there has been a significant interest in converting existing proofs into efficient algorithms. Many examples of this type can be found in [1, 32, 33]. Some of these are general methods, so algorithmic versions of these methods immediately imply efficient algorithms for several problems. One example is the Lovász local lemma whose algorithmic aspects were studied in [2, 7]. Another example is the regularity lemma [38]. The basic content of this lemma could be described by saying that every graph can, in some sense, be well approximated by random graphs. Some random graphs of a given edge density are much easier to treat than all graphs of the same edge-density, the regularity lemma helps uf to carry over results that are trivial for random graphs to the class of all graphs with a given number of edges. The lemma has numerous applications in various areas including combinatorial number theory [13, 39], computational complexity [19] and extremal graph theory [5, 6, 8, 9, 11, 12, 14, 15, 16, 35, 37]. Recently an NC^1-algorithmic version was given in [3].

During the past couple of years we have developed a new method in graph theory based on the regularity lemma (see [29]). We usually apply this method to find certain spanning subgraphs in dense graphs. Typical examples are spanning trees (Bolobás-conjecture, see [23]), Hamiltonian cycles or powers of Hamiltonian cycles (Pósa-Seymour conjecture, see [25, 26]) or H-factors for a fixed graph H (Alon–Yuster conjecture, see [28]). The other main general tool in the method, beside the regularity lemma, is the so-called blow-up lemma ([24]). This lemma helps to find bounded degree spanning subgraphs in ε-regular graphs. The rough idea of the original proof of this lemma was the following: we use a randomized greedy algorithm to embed most of the vertices of the bounded degree graph, and then finish the embedding by a König–Hall argument. Given the recent algorithmic version of the regularity lemma, the obvious question arises whether there is an algorithmic version also for the blow-up lemma. In this paper we give an affirmative answer to this question.

Theorem 1 (An algorithmic version of the blow-up lemma). *Given a graph R of order r and positive parameters δ, Δ, there exists an $\varepsilon > 0$ such that the following holds. Let N be an arbitrary positive integer, $n = rN$ and let us replace the vertices of R with pairwise disjoint N-sets $V_1, V_2, \ldots, V_r$ (blowing up). We construct two graphs on the same vertex-set $V = \bigcup V_i$. The graph $R(N)$ is obtained by replacing all edges of R with copies of the complete bipartite graph $K_{N,N}$, and a sparser graph G is constructed by replacing the edges of R with some (ε, δ)-superregular pairs. If a graph H with $\Delta(H) \le \Delta$ is embeddable into $R(N)$ then it is already embeddable into G. We can construct a copy of H in G in $O(n \times M(n))$ sequential time, where $M(n) = O(n^{2.376})$ is the time needed to multiply two n by n matrices with $0, 1$ entries over the integers. Furthermore, the algorithm can be parallelized and implemented in NC^5.*

Remark. For some very special cases of this theorem (e.g., a Hamiltonian path in a superregular pair) NC algorithms can be found in [36]. When using the blow-up lemma, we typically also need the following strengthened version: Given $c > 0$, there are positive functions $\varepsilon = \varepsilon(\delta, \Delta, r, c)$ and $\alpha = \alpha(\delta, \Delta, r, c)$ such that the blow-up lemma remains true if for every i there are certain vertices x to be embedded into V_i whose images are *a priori* restricted to certain sets $C_x \subset V_i$ provided that

i. Each C_x within a V_i is of size at least $c|V_i|$,
ii. The number of such restrictions within a V_i is not more than $\alpha|V_i|$.

Our proof is going to be similar to our original probabilistic proof, but we have to replace the probabilistic arguments with deterministic ones. In Section 2 we give a deterministic sequential algorithm for the embedding problem without considering implementation and time complexity issues. In Section 3 we show that the algorithm is correct. Implementation is discussed in Section 4. Finally, Section 5 contains various algorithmic applications.

2. THE ALGORITHM

The main idea of the algorithm is the following. We embed the vertices of H one-by-one by following a greedy algorithm, which works smoothly until there is only a small proportion of H left, and then it may get stuck hopelessly. To avoid that, we will set aside a positive proportion of the vertices of H as buffer vertices. Most of these buffer vertices will be embedded only at the very end by using a König–Hall argument.

2.1. Preprocessing

We will assume that $|V(H)| = |V(G)| = |\cup_i V_i| = n = rN$. We can also assume, without restricting generality, that $N \geq N_0(\delta, \Delta, r)$, for Theorem 1 is trivial for small N since ε-regularity with a small enough ε implies $G = R(N)$. Finally, we will assume for simplicity, that the density of every superregular pair in G is exactly δ. This is not a significant restriction, otherwise we just have to put everywhere the actual density instead of δ.

We will use the following parameters,

$$\varepsilon \ll \varepsilon' \ll \varepsilon'' \ll \varepsilon''' \ll \delta''' \ll \delta'' \ll \delta' \ll \delta,$$

where $a \ll b$ means that a is small enough compared to b. For simplicity we assume that any of these parameters multiplied with N gives an integer. For easier reading, we will mostly use the letter x for vertices of H, and the letter v for vertices of the host graph G. Given an embedding of H into $R(N)$, it defines an *assignment*,

$$\psi : V(H) \to \{V_1, V_2, \ldots, V_r\},$$

and we want to find an *embedding*,

$$\varphi : V(H) \to V(G), \qquad \varphi \text{ is one-to-one},$$

such that $\varphi(x) \in \psi(x)$ for all $x \in V(H)$. We will write $X_i = \psi^{-1}(V_i)$ for $i = 1, 2, \ldots, r$. Before we start the algorithm, we order the vertices of H into a sequence $S = (x_1, x_2, \ldots, x_n)$ which is more or less, but not exactly, the order in which the vertices will be embedded. Let $m = r\delta' N$. For each i, choose a set B_i of $\delta' N$ vertices in X_i such that any two of these vertices are at a distance at least 3 in H. (This is possible, for H is a bounded degree graph.) These vertices $b_1, \dots, b_m$ will be called the buffer vertices and they will be the last vertices in S.

S starts with the neighborhoods $N_H(b_1), N_H(b_2), \dots, N_H(b_m)$. The length of this initial segment of S will be denoted by T_0. Thus $T_0 = \sum_{i=1}^{m} |N_H(b_i)| \le \Delta m$. The rest of S is an arbitrary ordering of the leftover vertices of H. (When certain images are *a priori* restricted, as in the remark after the theorem, we also list the restricted vertices at the beginning of S right after the neighbors of the buffer vertices.)

2.2. Sketch of the Algorithm

In *Phase* 1 of the algorithm we will embed the vertices in S one-by-one into G until all nonbuffer vertices are embedded. For each x_j not embedded yet (including the buffer vertices) we keep track of an ever shrinking host set H_{t, x_j} that x_j is confined to at time t, and we only make a final choice for the location of x_j from H_{t, x_j} at time j. At time 0, H_{0, x_j} is the cluster that x_j is assigned to. For technical reasons we will also maintain another similar set, C_{t, x_j}, where we will ignore the possibility that some vertices are occupied already. Z_t will denote the set of occupied vertices. Finally we will maintain a set Bad_t of exceptional pairs of vertices.

In *Phase* 2, we embed the leftover vertices by using a König–Hall type argument.

2.3. Embedding Algorithm

At time 0, set $C_{0,x} = H_{0,x} = \psi(x)$ for all $x \in V(H)$. Put $T_1 = \delta'' n$.

Phase 1. For $t \ge 1$, repeat the following steps.

Step 1 (Extending the embedding). We embed x_t. Consider the vertices in H_{t-1, x_t}. We will pick one of these vertices as the image $\varphi(x_t)$ by using the selection algorithm (described below in Section 2.4).

Step 2 (Updating). We set

$$Z_t = Z_{t-1} \cup \{\varphi(x_t)\},$$

and for each unembedded vertex y (i.e., the set of vertices $x_j, t < j \le n$), set

$$C_{t,y} = \begin{cases} C_{t-1,y} \cap N_G(\varphi(x_t)) & \text{if } \{x_t, y\} \in E(H), \\ C_{t-1,y} & \text{otherwise}, \end{cases}$$

and

$$H_{t,y} = C_{t,y} \setminus Z_t.$$

We do not change the ordering at this step.

Step 3 (Exceptional vertices in G).

1. If $t \neq T_0$, then go to Step 4.
2. If $t = T_0$, then we do the following. We find the set (denoted by E_i) of those exceptional vertices $v \in V_i$, $1 \le i \le r$ for which v is not covered yet in the embedding and

$$\left|\{b\colon b \in B_i, v \in C_{t,b}\}\right| < \delta'' |B_i|.$$

We are going to change slightly the order of the remaining unembedded vertices in S. We choose a set E_H of unembedded nonbuffer vertices $x \in H$ of size $\sum_{i=1}^r |E_i|$ (more precisely $|E_i|$ vertices from X_i for all $1 \le i \le r$) with

$$H_{t,x} = H_{0,x} \setminus \left\{\varphi(x_j)\colon j \le t\right\} = \psi(x) \setminus \left\{\varphi(x_j)\colon j \le t\right\}.$$

Thus in particular, if $x \in X_i$, then $E_i \subset H_{t,x}$. For example, we may choose the vertices in E_H as vertices in H that are at a distance at least 3 from each other and any of the vertices embedded so far. We are going to show later in the proof of correctness that this is possible since H is a bounded degree graph and $\sum_{i=1}^r |E_i|$ is very small. We bring the vertices in E_H forward, followed by the remaining unembedded vertices in the same relative order as before. For simplicity we keep the notation $(x_1, x_2, \ldots, x_n)$ for the resulting order.

Step 4 (Exceptional vertices in H).

1. If T_1 does not divide t, then go to Step 5.
2. If T_1 divides t, then we do the following. We find all exceptional unembedded vertices $y \in H$ such that $|H_{t,y}| \le (\delta')^2 n$. Once again we slightly change the order of the remaining unembedded vertices in S. We bring these exceptional vertices forward (even if they are buffer vertices), followed by the nonexceptional vertices in the same relative order as before. Again, for simplicity we still use the notation $(x_1, x_2, \ldots, x_n)$ for the new order. Note that it will follow from the proof, that if $t \le 2T_0$, then we do not find any exceptional vertices in H, so we do not change the ordering at this step.

Step 5. If there are no more unembedded nonbuffer vertices left, then set $T = t$ and go to Phase 2, otherwise set $t \leftarrow t + 1$ and go back to Step 1.

Phase 2. Find a system of distinct representatives of the sets $H_{t,y}$ for all unembedded y (i.e., the set of vertices $x_j, T < j \le n$).

2.4. Selection Algorithm

We distinguish two cases.

Case 1. $x_t \notin E_H$.

We choose a vertex $v \in H_{t-1,x_t}$ as the image $\varphi(x_t)$ for which the following hold for all unembedded y with $\{x_t, y\} \in E(H)$,

$$(\delta - \varepsilon)|H_{t-1,y}| \le \deg_G(v, H_{t-1,y}) \le (\delta + \varepsilon)|H_{t-1,y}|, \tag{1}$$

$$(\delta - \varepsilon)|C_{t-1,y}| \le \deg_G(v, C_{t-1,y}) \le (\delta + \varepsilon)|C_{t-1,y}|, \tag{2}$$

and

$$(\delta - \varepsilon)|C_{t-1,y} \cap C_{t-1,y'}| \le \deg_G(v, C_{t-1,y} \cap C_{t-1,y'}) \le (\delta + \varepsilon)|C_{t-1,y} \cap C_{t-1,y'}|, \tag{3}$$

for at least a $(1-\varepsilon')$ proportion of the unembedded vertices y' with $\psi(y') = \psi(y)$ and $\{y, y'\} \notin \mathrm{Bad}_{t-1}$. Then we get Bad_t by taking the union of Bad_{t-1} and the set of all of those pairs $\{y, y'\}$ for which (3) does not hold for $v = \varphi(x_t)$, $C_{t-1,y}$, and $C_{t-1,y'}$. Thus note that we add at most $\Delta \varepsilon' N$ new pairs to Bad_t.

Case 2. $x_t \in E_H$.

If $x_t \in X_i$, then we choose an arbitrary vertex of E_i as $\varphi(x_t)$. Note that for all $y \in N_H(x_t)$, we have $C_{t-1,y} = \psi(y)$,

$$\deg_G(\varphi(x_t), C_{t-1,y}) = \deg_G(\varphi(x_t)) \ge \delta N > (\delta - \varepsilon)|C_{t-1,y}|, \tag{4}$$

and

$$\deg_G(\varphi(x_t), H_{t-1,y}) \ge \deg_G(\varphi(x_t)) - T_0 - |E_H| \ge \delta N - 2\Delta r \delta' N > \frac{\delta}{2} N. \tag{5}$$

Here we used superregularity and the fact that $|E_H| \lll \Delta m$ which will be shown later.

PROOF OF CORRECTNESS

The following claims state that our algorithm finds a good embedding of H into G.

Claim 1. *Phase* 1 *always succeeds*.

Claim 2. *Phase* 2 *always succeeds*.

If at time t, S is a set of unembedded vertices $x \in H$ with $\psi(x) = V_i$ (here and throughout the proof when we talk about time t, we mean *after* Phase 1 is executed for time t, so for example x_t is considered embedded at time t), then we define the bipartite graph U_t as follows. One color class is S, the other is V_i, and we have an edge between a $x \in S$ and a $v \in V_i$ whenever $v \in C_{t,x}$.

In the proofs of the above claims the following lemma will play a major role. First we prove the lemma for $t \le T_0$, from this we deduce that $|E_H|$ is small, then we prove the lemma for $T_0 < t \le T$.

Lemma 2. *We are given integers $1 \le i \le r$, $1 \le t \le T_0$, and a set $S \subset X_i$ of unembedded vertices at time t with $|S| \ge (\delta''')^2 |X_i| = (\delta''')^2 N$. If we assume that Phase 1 succeeded for all time t' with $t' \le t$, then apart from an exceptional set F of size at most $\varepsilon'' N$, for every vertex $v \in V_i$ we have*

$$\deg_{U_t}(v) = |\{x \colon x \in S, v \in C_{t,x}\}| \ge (1 - \varepsilon'') d(U_t)|S|, \qquad \left(\ge \frac{\delta^\Delta}{2}|S| \right).$$

Proof. In the proof of this lemma we will use the "defect form" of the Cauchy–Schwarz inequality (just as in the original proof of the regularity lemma): if

$$\sum_{k=1}^{m} X_k = \frac{m}{n} \sum_{k=1}^{n} X_k + D, \qquad (m \le n),$$

then

$$\sum_{k=1}^{n} X_k^2 \ge \frac{1}{n} \left(\sum_{k=1}^{n} X_k \right)^2 + \frac{D^2 n}{m(n-m)}.$$

Assume indirectly that the statement in Lemma 2 is not true, that is, $|F| > \varepsilon'' N$. We take an $F_0 \subset F$ with $|F_0| = \varepsilon'' N$. Let us write $\nu(t,x)$ for the number of neighbors (in H) of x embedded by time t. Then in U_t using the left side of (2) we get

$$\begin{aligned} e(U_t) = d(U_t)|S|\,|V_i| &= \sum_{v \in V_i} \deg_{U_t}(v) = \sum_{x \in S} \deg_{U_t}(x) \\ &= \sum_{x \in S} |C_{t,x}| \ge \sum_{x \in S} (\delta - \varepsilon)^{\nu(t,x)} N. \end{aligned} \tag{6}$$

We also have

$$\begin{aligned} &\sum_{x \in S} \sum_{x' \in S} \left| N_{U_t}(x) \cap N_{U_t}(x') \right| \\ &\quad = \sum_{x \in S} \sum_{x' \in S} |C_{t,x} \cap C_{t,x'}| \\ &\quad \le \sum_{x \in S} \sum_{x' \in S} (\delta + \varepsilon)^{\nu(t,x) + \nu(t,x')} N + |S| N + \Delta^2 |S| N + 2\,\Delta \varepsilon' N^3 \\ &\quad \le \sum_{x \in S} \sum_{x' \in S} (\delta + \varepsilon)^{\nu(t,x) + \nu(t,x')} N + 4\,\Delta \varepsilon' N^3. \end{aligned} \tag{7}$$

The error terms come from the following (x, x') pairs. For each such pair we estimate $|C_{t,x} \cap C_{t,x'}| \le N$. The first error term comes from the pairs where $x = x'$. The second error term comes from those pairs (x, x') for which $N_H(x) \cap N_H(x') \ne \emptyset$. The number of these pairs is at most $|S|\Delta(\Delta - 1) \le \Delta^2 |S|$. Finally we have the pairs for which $\{x, x'\} \in \mathrm{Bad}_t$. The number of these pairs is at most $2t\,\Delta \varepsilon' N \le 2\,\Delta \varepsilon' N^2$. For the good pairs we used the right side of (3).

Next we will use the Cauchy–Schwarz inequality with $m = \varepsilon'' N$ and the variables X_k, $k = 1, \dots, N$ we are going to correspond to $\deg_{U_t}(v)$, $v \in V_i$ (and the first m variables to degrees in F_0). Then we have

$$\begin{aligned} |D| &= \varepsilon'' \sum_{v \in V_i} \deg_{U_t}(v) - \sum_{v \in F_0} \deg_{U_t}(v) \\ &\geq \varepsilon'' \sum_{v \in V_i} \deg_{U_t}(v) - \varepsilon''(1-\varepsilon'') d(U_t)|S|N = (\varepsilon'')^2 \sum_{v \in V_i} \deg_{U_t}(v). \end{aligned} \tag{8}$$

Then using (6), (8) and the Cauchy–Schwarz inequality we get

$$\begin{aligned} &\sum_{x \in S} \sum_{x' \in S} \left| N_{U_t}(x) \cap N_{U_t}(x') \right| \\ &\quad = \sum_{v \in V_i} \left(\deg_{U_t}(v) \right)^2 \\ &\quad \geq \frac{1}{N} \left(\sum_{v \in V_i} \deg_{U_t}(v) \right)^2 + (\varepsilon'')^3 d(U_t)^2 N |S|^2 \\ &\quad \geq \frac{1}{N} \left(\sum_{x \in S} (\delta - \varepsilon)^{\nu(t,x)} N \right)^2 + (\varepsilon'')(\delta - \varepsilon)^{2\Delta} N |S|^2 \\ &\quad \geq \sum_{x \in S} \sum_{x' \in S} (\delta - \varepsilon)^{\nu(t,x) + \nu(t,x')} N + (\varepsilon'')^3 (\delta - \varepsilon)^{2\Delta} N |S|^2, \end{aligned}$$

which is a contradiction with (7), since $|S| \geq (\delta''')^2 N$,

$$(\varepsilon'')^3 (\delta - \varepsilon)^{2\Delta} (\delta''')^2 \gg 4 \Delta \varepsilon' \gg 4 \Delta \varepsilon,$$

and

$$\left((\delta + \varepsilon)^{\nu(t,x) + \nu(t,x')} - (\delta - \varepsilon)^{\nu(t,x) + \nu(t,x')} \right) \ll 4 \Delta \varepsilon. \qquad \blacksquare$$

An easy consequence of Lemma 2 is the following lemma.

Lemma 3. *In Step* 3 *we have* $|E_i| \leq \varepsilon'' N$ *for every* $1 \leq i \leq r$.

Proof. Indeed applying Lemma 2 with $t = T_0$ and $S = B_i$ [so we have $|S| = |B_i| = \delta' N > (\delta''')^2 N$] we get

$$(1 - \varepsilon'') d(U_t)|S| \geq \frac{\delta^{\Delta}}{2} |S| > \delta'' |S|,$$

and $E_i \subset F$. $\blacksquare$

From this we can prove Lemma 2 for $t > T_0$ with ε''' instead of ε''.

Lemma 4. *We are given integers* $1 \leq i \leq r$, $T_0 < t \leq T$, *and a set* $S \subset X_i$ *of unembedded vertices at time* t *with* $|S| \geq (\delta''')^2 |X_i| = (\delta''')^2 N$. *If we assume that Phase* 1 *succeeded for all time t' with $t' \le t$, then apart from an exceptional set F of size at most $\varepsilon''' N$, for every vertex $v \in V_i$ we have*

$$\deg_{U_t}(v) = |\{x, x \in S, v \in C_{t,x}\}| \ge (1-\varepsilon''')d(U_t)|S| \qquad \left(\ge \frac{\delta^{\Delta}}{2}|S| \right).$$

Proof. We only have to pay attention to the neighbors of the elements of E_H, otherwise the proof is the same as the proof of Lemma 2 with ε''' instead of ε''. In (6) we have no change, since as we noted in the selection algorithm,

$$|C_{t,x}| \ge (\delta - \varepsilon)^{\nu(t,x)} N$$

is also true for the neighbors of the elements of E_H.

In (7) we have more bad pairs, namely all pairs (x, x') where x or x' is a neighbor of an element of E_H. These give an additional error term of $2\Delta r \varepsilon'' |S| N^2$. However, the contradiction still holds, since

$$(\varepsilon''')^3 (\delta - \varepsilon)^{2\Delta} (\delta''')^2 \gg \varepsilon''. \qquad \blacksquare$$

An easy consequence of Lemmas 2 and 4 is the following lemma.

Lemma 5. *We are given integers $1 \le i \le r$, $1 \le t \le T$, a set $S \subset X_i$ of unembedded vertices at time t with $|S| \ge \delta''' |X_i| = \delta''' N$ and a set $A \subset V_i$ with $|A| \ge \delta''' |V_i| = \delta''' N$. If we assume that Phase 1 succeeded for all time t' with $t' \le t$, then apart from an exceptional set S' of size at most $(\delta''')^2 N$, for every vertex $x \in S$ we have*

$$|A \cap C_{t,x}| \ge \frac{|A|}{2N} |C_{t,x}|. \tag{9}$$

Proof. Assume indirectly that the statement is not true, i.e., there exists a set $S' \subset S$ with $|S'| > (\delta''')^2 N$ such that for every $x \in S'$ (9) does not hold. Once again we consider the bipartite graph $U_t = U_t(S', V_i)$. We have

$$\sum_{v \in A} \deg_{U_t}(v) = \sum_{x \in S'} |A \cap C_{t,x}| < \frac{|A|}{2N} \sum_{x \in S'} |C_{t,x}| = \frac{|A|}{2N} d(U_t)|S'|N.$$

On the other hand, applying Lemmas 2 or 4 for S' we get

$$\sum_{v \in A} \deg_{U_t}(v) \ge (1 - \varepsilon''') d(U_t) |S'| (|A| - \varepsilon''' N)$$

contradicting the previous inequality. $\blacksquare$

Finally we have

Lemma 6. *For every $1 \le t \le T$ and for every vertex y that is unembedded at time t, if we assume that Phase 1 succeeded for all time t' with $t' \le t$, then we have the following at time t,*

$$|H_{t,y}| > \delta'' N. \tag{10}$$

Proof. We apply Lemma 5 with S_t the set of all unembedded vertices in V_i at time t, and $A_t = V_i \setminus Z_t$ (all uncovered vertices). Then for all but at most $(\delta''')^2 N$ vertices $x \in S_t$ using (2) and (4) we get

$$|H_{t,x}| = |A_t \cap C_{t,x}| \geq \frac{|A_t|}{2N} |C_{t,x}| \geq \frac{\delta'}{4} (\delta - \varepsilon)^{\Delta} N \gg (\delta')^2 N, \tag{11}$$

if $|A_t| \geq (\delta'/2)N$. We will show next that in fact for $1 \leq t \leq T$, we have

$$|A_t| \geq |A_T| \geq (\delta' - \delta'')N \qquad \left(\geq \frac{\delta'}{2} N \right),$$

so (11) always holds. Assume indirectly that this is not the case, i.e., there exists a $1 \leq T' < T$ for which,

$$|A_{T'}| \geq (\delta' - \delta'')N \quad \text{but} \quad |A_{T'+1}| < (\delta' - \delta'')N.$$

From the above at any given time t for which $T_1 | t$ and $1 \leq t \leq T'$, in Step 4 we find at most $(\delta''')^2 N$ exceptional vertices in V_i. Hence, altogether we find at most

$$\frac{1}{\delta''} (\delta''')^2 N \ll \delta'' N,$$

exceptional vertices in V_i up to time T'. However, this implies that at time T' we still have many more than $(\delta' - \delta'')N$ unembedded buffer vertices in V_i, which in turn implies that $|A_{T'+1}| \gg (\delta' - \delta'')N$, a contradiction. Thus we have

$$|A_T| \geq (\delta' - \delta'')N, \qquad T \leq rN - r\delta' N + r\delta'' N,$$

at time T (or in the second phase) we have at least $(\delta' - \delta'')N$ unembedded buffer vertices in each V_i, and furthermore, for every $1 \leq t \leq T$ for all but at most $(\delta''')^2 N$ vertices $x \in S_t$ we have

$$|H_{t,x}| > (\delta')^2 N.$$

Let us pick an arbitrary $1 \leq t \leq T$ and an unembedded y at time [with $\psi(y) = V_i$]. We have to show that (10) holds. Let $k\,\delta'' n = kT_1 \leq t < (k+1)T_1$ for some $0 \leq k \leq T/T_1$. We distinguish two cases:

Case 1. y was not among the at most $(\delta''')^2 N$ exceptional vertices of V_i found in Step 4 at time kT_1. Then

$$|H_{t,y}| \geq \left(\left(\frac{\delta}{2} \right)^{\Delta} (\delta')^2 - r\delta'' \right) N.$$

Indeed, at time kT_1 we had $|H_{kT_1,y}| \geq (\delta')^2 N$. Until time t, $H_{t,y}$ could have been cut by $\geq (\delta/2)$-fraction [using (1) and (5)] up to at most Δ times, and precisely $t - kT_1 \leq T_1 = r\,\delta'' N$ new vertices were covered.

Case 2. y was among the most $(\delta''')^2N$ exceptional vertices of V_i found in Step 4 at time kT_1. Then

$$|H_{t,y}| \geq \left(\left(\frac{\delta}{2} \right)^{\Delta} (\delta')^2 - r\delta'' - r(\delta''')^2 \right) N,$$

since at time $(k-1)T$ (note that in this case we must have $k \geq 2$), y was not exceptional, and because the exceptional vertices were brought forward we have $t \leq kT_1 + r(\delta''')^2N$. Thus in both cases we have $|H_{t,y}| > \delta''N$. ■

Finally we show that the selection algorithm always succeeds in selecting an image $\varphi(x_t)$.

Lemma 7. *For every $1 \leq t \leq T$, if we assume that Phase 1 succeeded for all time t' with $t' < t$, then Phase 1 succeeds for time t.*

Proof. We only have to consider Case 1 in the selection algorithm. We choose a vertex $v \in H_{t-1,x_t}$ as the image $\varphi(x_t)$ which satisfies (1), (2), and (3). We have by Lemma 6,

$$|H_{t-1,x_t}| \geq \delta''N.$$

By ε-regularity we have at most $2\varepsilon N$ vertices in H_{t-1,x_t} which do not satisfy (1) and similarly for (2). For (3) we define an auxiliary bipartite graph B as follows. One color class W_1 is the vertices in H_{t-1,x_t} and the other class W_2 is the sets $C_{t-1,y} \cap C_{t-1,y'}$ for all pairs $\{y, y'\}$ where $\{x_t, y\} \in E(H)$, $\psi(y') = \psi(y)$, and $\{y, y'\} \notin \mathrm{Bad}_{t-1}$. We put an edge between a $v \in W_1$ and a $S \in W_2$ if inequality (3) is not satisfied for v and S. Let us assume indirectly that we have more than $\varepsilon'N$ vertices $v \in W_1$ with $\deg_B(v) > \varepsilon'|W_2|$. Then there must exist a $S \in W_2$ with

$$\deg_B(S) > (\varepsilon')N \gg \varepsilon N.$$

However, this is a contradiction with ε-regularity since

$$|S| \geq (\delta - \varepsilon)^{2\Delta} N \gg \varepsilon N.$$

Here we used the fact that the pair corresponding to S is not in Bad_{t-1}. Thus altogether we have at most $4\varepsilon N + \varepsilon'N \ll \delta''N$ vertices in H_{t-1,x_t} that we cannot choose and thus the selection algorithm always succeeds in selecting an image $\varphi(x_t)$, proving Claim 1. ■

Proof of Claim 2. We want to show that we can find a system of distinct representatives of the sets H_{T,x_j}, $T < j \leq n$, where the sets H_{T,x_j} belong to a given cluster V_i. To simplify notation, let us denote by Y the set of remaining vertices in V_i, and by X the set of remaining unembedded (buffer) vertices assigned to V_i. If $x = x_j \in X$ then write H_x for its possible location set H_{T,x_j} at time T. Also write

$M=|X|=|Y|$. The König–Hall condition for the existence of a system of distinct representatives obviously follows from the following three conditions,

$$|H_x| > \delta''' M \quad \text{for all } x \in X, \tag{12}$$

$$\left|\bigcup_{x\in S} H_x\right| \geq (1-\delta''')M \quad \text{for all subsets } S \subset X, |S| \geq \delta''' M, \tag{13}$$

$$\left|\bigcup_{x\in S} H_x\right| = M \quad \text{for all subsets } S \subset X, |S| \geq (1-\delta''')M. \tag{14}$$

Equation (12) is an immediate consequence of Lemma 6, (13) is a consequence of Lemma 2. Finally to prove (14), we have to show that every vertex in $Y \subset V_i$ belongs to at least $\delta'''|X|$ location sets H_x. However, this is trivial from the construction of the embedding algorithm, in Step 3 of Phase 1 we took care of the small number of exceptional vertices for which this is not true. This finishes the proof of Claim 2 and the proof of correctness. ■

4. IMPLEMENTATION

The sequential implementation is immediate. In Phase 1 we have $\leq n$ iterations, and it is not hard to see that one iteration can be implemented in $O(M(n))$ time. Phase 2 can be implemented in $O(rN^{5/2}) = O(nM(n))$ time by applying an algorithm for finding a maximum matching in a bipartite graph (see e.g., [20, 30]).

For the parallel implementation, our main tool is the NC^4 algorithm for the maximal independent set problem. A subset I of the vertices of a graph G is independent if there are no edges between any two vertices in I. An independent set I is *maximal* if it is not a proper subset of any other independent set. Karp and Wigderson ([22]) were the first to give an NC^4-algorithm for this problem. Better algorithms were later described in [4, 17, 18, 31]. We call this the maximal independent set (MIS) algorithm.

For the parallelization of Phase 1, we show that if α is a small enough constant and n' is the number of remaining unembedded vertices, then we can embed $\lfloor \alpha n' \rfloor$ vertices in parallel. First we pick these vertices by running the MIS algorithm on the following auxiliary graph. The vertices are the vertices of H, and we put an edge between two vertices, if either they are at a distance less than 3, or both vertices are embedded already. If in the maximal independent set that we find, we have a vertex that is embedded already (we can have only one such vertex), then we remove this vertex from the independent set. We keep $\lfloor \alpha n' \rfloor$ vertices from the remaining vertices. (Since H is a bounded degree graph, if α is sufficiently small then the size of *every* maximal independent set is much larger than $\lfloor \alpha n' \rfloor$.) These vertices are brought forward in the order in the embedding algorithm and we embed these vertices in parallel. For each such vertex we determine the set of vertices where it could be embedded by the embedding algorithm. Once again running the MIS algorithm on the appropriately defined auxiliary graph (we put an edge between two vertices in two different sets if they correspond to the same vertex), we can choose a distinct representative from these sets. Finally we embed each vertex to its representative. We iterate this procedure until the number of remaining unembedded vertices is $\leq (\log n)^5$, and then we embed these vertices sequentially. Thus Phase 1 can be implemented in $O(\log n(\log n)^4) = O((\log n)^5)$ parallel time.

For Phase 2 it remains to show, that if the bipartite graph U_t is defined as above between X and Y (i.e., there is an edge between $x \in X$ and $v \in Y$ if and only if $v \in H_x$), then we can construct a perfect matching in U_t in NC^4. For this purpose we obtain a maximal matching by running MIS on the linegraph of U_t. Then obviously for the remaining unmatched vertices, say $Z(X)$ and $Z(Y)$, $|Z(X)| = |Z(Y)|$, and $Z(X) \cup Z(Y)$ is an independent set. From (13) $|Z(X)| \leq \delta''' N$ follows. Furthermore, Lemmas 5 and 6 imply that, if we take $x \in Z(X)$, $v \in Z(Y)$, then there are many ($\gg \delta''' N$) internally vertex-disjoint alternating paths of length 5 between x and v. We define again an auxiliary graph in the following way. We pair up the vertices in $Z(X)$ and $Z(Y)$, and for each pair we have a set of ($\gg \delta''' N$) vertices corresponding to the internally vertex-disjoint alternating paths of length 5 between the two vertices in the pair. We put an edge between two alternating paths if they share a common vertex (so e.g., the alternating paths corresponding to one pair form a clique). Since the number of alternating paths for each pair is much more than the number of pairs, it is easy to see that every maximal independent set contains an alternating path for each pair. Thus running again MIS on the auxiliary graph, we can find vertex-disjoint alternating paths of length 5 between the pairs in $Z(X) \cup Z(Y)$. Changing the matching edges to nonmatching edges on these paths we get a perfect matching.

5. APPLICATIONS

In most applications of our method, the only nonconstructive parts are the regularity lemma and the blow-up lemma. Therefore, the existence proofs together with the NC^1 version of the regularity lemma and Theorem 1, provide several immediate algorithmic applications. Let us mention here two applications. Additional applications and the details of the proofs will appear in a forthcoming paper.

Theorem 8 (Existential version in [23], *NC*-version in [36]). *Let Δ and $0 < \delta < 1/2$ be given. Then there exists an n_0 with the following properties. If $n \geq n_0$, T is a tree of order n and maximum degree Δ, and G is a graph of order n and minimum degree at least $((1/2) + \delta)n$, then T is a subgraph of G. Furthermore, a copy of T in G can be found in $O(nM(n))$ sequential time as well as in NC^5.*

Theorem 9 (Existential version in [26] for $k = 3$ and in [27] in general). *For any positive integer k there exists an $n_0 = n_0(k)$ such that if G has order $n \geq n_0$ and minimal degree $\delta(G) \geq k/(k+1)n$, then G contains the kth power of a Hamiltonian cycle. Furthermore, a copy of the kth power of a Hamiltonian cycle can be found in $O(nM(n))$ sequential time as in NC^5.*

ACKNOWLEDGMENT

We are grateful to one of the referees and to the managing editor for many helpful suggestions and for pointing out an error in an earlier version of this paper.

REFERENCES

[1] N. Alon, Non-constructive proofs in combinatorics, *Proceedings of the International Congress of Mathematics, Kyoto 1990, Japan*, Springer-Verlag, Tokyo, 1991, pp. 1421–1429.

[2] N, Alon, A parallel algorithmic version of the local lemma, *Random Struct. Algorithms*, **2**, 367–378 (1991), *Proceedings of the Thirty-Second IEEE FOCS (1991)*, pp. 586–593.

[3] N. Alon, R. Duke, H. Lefman, V. Rödl, and R. Yuster, The algorithmic aspects of the regularity lemma, *Proceedings of the Thirty-Third IEEE FOCS (1993)*, pp. 479–481, *J. Algorithms*, **16**, 80–109 (1994).

[4] N. Alon, L. Babai, and A. Itai, A fast randomized parallel algorithm for the maximal independent set problem, *J. Algorithms*, **7**, 567–583 (1986).

[5] N. Alon and R. Yuster, Almost H-factors in dense graphs, *Graphs Combin.*, **8**, 95–102 (1992).

[6] N. Alon and R. Yuster, H-factors in dense graphs, *J. Combin. Theory Ser. B*, **66**, 269–282 (1996).

[7] J. Beck, An algorithmic approach to the Lovász local lemma, *Random Struct. Algorithms*, **2**, 343–365 (1991).

[8] B. Bollobás, P. Erdös, M. Simonovits, and E. Szemerédi, Extremal graphs without large forbidden subgraphs, in *Advances in Graph Theory*, Cambridge Combinatorial Conference, Trinity College, 1977, *Ann. Discrete Math.*, **3**, 29–41 (1978).

[9] V. Chvátal, V. Rödl, E. Szemerédi, and W. T. Trotter Jr., The Ramsey number of a graph with bounded maximum degree, *J. Combin. Theory Ser. B*, **34**, 239–243 (1983).

[10] S. A. Cook, Taxonomy of problems with fast parallel algorithms, *Inform. and Control*. **64**, 2–22 (1985).

[11] P. Erdös, P. Frankl, and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.*, **2**, 113–121 (1986).

[12] P. Franek and V. Rödl, On Erdös conjecture for multiplicities of complete subgraphs, *Graphs Combin.*, to appear.

[13] P. Frankl, R. L. Graham, and V. Rödl, On subsets of Abelian groups with no 3-term arithmetic progression, *J. Combin. Theory Ser. A*, **15**, 157–161 (1987).

[14] P. Frankl and Z. Füredi, Exact solution of some Turán-type problems, *J. Combin. Theory Ser. A*, **45**, 226–262 (1987).

[15] Z. Füredi, The maximum number of edges in a minimal graph of diameter 2, *J. Graph Theory*, **16**, 81–98 (1992).

[16] Z. Füredi, M. Goemans, and D. J. Kleitman, On the maximum number of triangles in wheel-free graphs, to appear.

[17] M. Goldberg and T. Spencer, A new parallel algorithm for the maximal independent set problem, *SIAM J. Comput.*, **18**, 419–427 (1989).

[18] M. Goldberg and T. Spencer, Constructing a maximal independent set in parallel, *SIAM J. Disc. Math.*, **2**, 322–328 (1989).

[19] A. Hajnal, W. Maass, and G. Turán, On the communication complexity of graph properties, *Proceedings of the Twentieth Symposium on Theory of Computing*, 1988, pp. 186–191.

[20] J. E. Hopcroft and R. M. Karp, An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs, *SIAM J. Comput.*, **2**, 225–231 (1973).

[21] R. Karp and V. Ramachandran, Parallel algorithms for shared memory machines, in *Handbook of Theoretical Computer Science*, J. Van Leeuven, Ed., North-Holland, Amsterdam, 1990, pp. 869–941.

[22] R. Karp and A. Wigderson, A fast parallel algorithm for the maximal independent set problem, *J. Assoc. Comput. Mach.*, **32**, 762–773 (1985).

[23] J. Komlós, G. N. Sárközy, and E. Szemerédi, Proof of a packing conjecture of Bollobás *Combinatorics, Probab. Comput.*, **4**, 241–255 (1995).

[24] J. Komlós, G. N. Sárközy, and E. Szemerédi, Blow-up lemma, *Combinatorica*, **17**, 109–123 (1997).

[25] J. Komlós, G. N. Sárközy, and E. Szemerédi, On the Pósa–Seymour conjecture, *J. Graph Theory*, to appear.

[26] J. Komlós, G. N. Sárközy, and E. Szemerédi, On the square of a Hamiltonian cycle in dense graphs, *Random Structures and Algorithms*, **9**, 193–211 (1996).

[27] J. Komlós, G. N. Sárközy, and E. Szemerédi, Proof of the Seymour conjecture for large graphs, submitted for publication.

[28] J. Komlós, G. N. Sárközy, and E. Szemerédi, Proof of the Alon–Yuster conjecture, in preparation.

[29] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, in *Paul Erdös is 80, Vol. 2., Proc. Colloq. Math. Soc. János Bolyai* 295–352 (1996).

[30] L. Lovász and M. D. Plummer, Matching theory, *Ann. Discrete Math.*, **29** (1986).

[31] M. Luby, A simple parallel algorithm for the maximal independent set problem, *SIAM J. Comput.*, **15**, 1036–1053 (1986).

[32] C. H. Papadimitriou, On graph theoretic lemmata and complexity classes, *Proceedings of the Thirty-First IEEE FOCS*, 1990, pp. 794–801.

[33] C. H. Papadimitriou, On the complexity of the parity argument and other inefficient proofs of existence, *J. Comput. System Sci.*, **47** (1993).

[34] N. J. Pippenger, On simultaneous resource bounds, *Proceedings of the Twentieth IEEE FOCS*, 1979, pp. 307–311.

[35] V. Rödl, On universality of graphs with uniformly distributed edges, *Discrete Math.*, **59**, 125–134 (1986).

[36] G. N. Sárközy, Fast parallel algorithms for finding Hamiltonian cycles and trees in graphs, Technical Report 93-81, DIMACS, Rutgers University, New Brunswick, NJ, 1993.

[37] M. Simonovits and V. T. Sös, Szemerédi's partition and quasirandomness, *Random Struct. Algorithms*, **2**, 1–10 (1991).

[38] E. Szemerédi, Regular partitions of graphs, in *Colloques Internationaux C.N.R.S.*, J.-C. Bermond, J.-C. Fournier, M. Las Vergnas, and D. Sotteau, Eds., 1978, pp. 399–401.

[39] E. Szemerédi, On a set containing no k elements in arithmetic progression, *Acta Arith.*, **XXVII**, 199–245 (1975).

[40] L. G. Valiant, Parallel computation, *Proceedings of the Seventh IBM Symposium on Mathematical Foundations of Computer Science*, 1982.