

# A DIRAC TYPE RESULT ON HAMILTON CYCLES IN ORIENTED GRAPHS

LUKE KELLY, DANIELA KÜHN AND DERYK OSTHUS

**ABSTRACT.** We show that for each  $\alpha > 0$  every sufficiently large oriented graph  $G$  with  $\delta^+(G), \delta^-(G) \geq 3|G|/8 + \alpha|G|$  contains a Hamilton cycle. This gives an approximate solution to a problem of Thomassen [21]. In fact, we prove the stronger result that  $G$  is still Hamiltonian if  $\delta(G) + \delta^+(G) + \delta^-(G) \geq 3|G|/2 + \alpha|G|$ . Up to the term  $\alpha|G|$  this confirms a conjecture of Häggkvist [10]. We also prove an Ore-type theorem for oriented graphs.

## 1. INTRODUCTION

An *oriented graph*  $G$  is obtained from a (simple) graph by orienting its edges. Thus between every pair of vertices of  $G$  there exists at most one edge. The *minimum semi-degree*  $\delta^0(G)$  of  $G$  is the minimum of its minimum outdegree  $\delta^+(G)$  and its minimum in-degree  $\delta^-(G)$ . When referring to paths and cycles in oriented graphs we always mean that these are directed without mentioning this explicitly.

A fundamental result of Dirac states that a minimum degree of  $|G|/2$  guarantees a Hamilton cycle in an undirected graph  $G$ . There is an analogue of this for digraphs due to Ghouila-Houri [9] which states that every digraph  $D$  with minimum semi-degree at least  $|D|/2$  contains a Hamilton cycle. The bounds on the minimum degree in both results are best possible. A natural question is to ask for the (smallest) minimum semi-degree which guarantees a Hamilton cycle in an oriented graph  $G$ . This question was first raised by Thomassen [20], who [22] showed that a minimum semi-degree of  $|G|/2 - \sqrt{|G|/1000}$  suffices (see also [21]). Note that this degree requirement means that  $G$  is not far from being a tournament. Häggkvist [10] improved the bound further to  $|G|/2 - 2^{-15}|G|$  and conjectured that the actual value lies close to  $3|G|/8$ . The best previously known bound is due to Häggkvist and Thomason [11], who showed that for each  $\alpha > 0$  every sufficiently large oriented graph  $G$  with minimum semi-degree at least  $(5/12 + \alpha)|G|$  has a Hamilton cycle. Our first result implies that the actual value is indeed close to  $3|G|/8$ .

**Theorem 1.** *For every  $\alpha > 0$  there exists an integer  $N = N(\alpha)$  such that every oriented graph  $G$  of order  $|G| \geq N$  with  $\delta^0(G) \geq (3/8 + \alpha)|G|$  contains a Hamilton cycle.*

A construction of Häggkvist [10] shows that the bound in Theorem 1 is essentially best possible (see Proposition 6).

In fact, Häggkvist [10] formulated the following stronger conjecture. Given an oriented graph  $G$ , let  $\delta(G)$  denote the minimum degree of  $G$  (i.e. the minimum number of edges incident to a vertex) and set  $\delta^*(G) := \delta(G) + \delta^+(G) + \delta^-(G)$ .

**Conjecture 2** (Häggkvist [10]). *Every oriented graph  $G$  with  $\delta^*(G) > (3n - 3)/2$  has a Hamilton cycle.*

---

D. Kühn was partially supported by the EPSRC, grant no. EP/F008406/1. D. Osthus was partially supported by the EPSRC, grant no. EP/E02162X/1 and EP/F008406/1.

Our next result provides an approximate confirmation of this conjecture for large oriented graphs.

**Theorem 3.** *For every  $\alpha > 0$  there exists an integer  $N = N(\alpha)$  such that every oriented graph  $G$  of order  $|G| \geq N$  with  $\delta^*(G) \geq (3/2 + \alpha)|G|$  contains a Hamilton cycle.*

Note that Theorem 1 is an immediate consequence of this. The proof of Theorem 3 can be modified to yield the following Ore-type analogue of Theorem 1. (Ore's theorem [19] states that every graph  $G$  on  $n \geq 3$  vertices which satisfies  $d(x) + d(y) \geq n$  whenever  $xy \notin E(G)$  has a Hamilton cycle.)

**Theorem 4.** *For every  $\alpha > 0$  there exists an integer  $N = N(\alpha)$  such that every oriented graph  $G$  of order  $|G| \geq N$  with  $d^+(x) + d^-(y) \geq (3/4 + \alpha)|G|$  whenever  $xy \notin E(G)$  contains a Hamilton cycle.*

A version for general digraphs was proved by Woodall [23]: every strongly connected digraph  $D$  on  $n \geq 2$  vertices which satisfies  $d^+(x) + d^-(y) \geq n$  whenever  $xy \notin E(D)$  has a Hamilton cycle.

Theorem 1 immediately implies a partial result towards a conjecture of Kelly (see e.g. [3]), which states that every regular tournament on  $n$  vertices can be partitioned into  $(n - 1)/2$  edge-disjoint Hamilton cycles. (A regular tournament is an orientation of a complete graph in which the indegree of every vertex equals its outdegree.)

**Corollary 5.** *For every  $\alpha > 0$  there exists an integer  $N = N(\alpha)$  such that every regular tournament of order  $n \geq N$  contains at least  $(1/8 - \alpha)n$  edge-disjoint Hamilton cycles.*

Indeed, Corollary 5 follows from Theorem 1 by successively removing Hamilton cycles until the oriented graph  $G$  obtained from the tournament in this way has minimum semi-degree less than  $(3/8 + \alpha)|G|$ . The best previously known bound on the number of edge-disjoint Hamilton cycles in a regular tournament is the one which follows from the result of Häggkvist and Thomason [11] mentioned above. A related result of Frieze and Krivelevich [8] states that every dense  $\varepsilon$ -regular digraph contains a collection of edge-disjoint Hamilton cycles which covers almost all of its edges. This immediately implies that the same holds for almost every tournament. Together with a lower bound by McKay [18] on the number of regular tournaments, the above result in [8] also implies that almost every regular tournament contains a collection of edge-disjoint Hamilton cycles which covers almost all of its edges.

Note that Theorem 3 implies that for sufficiently large tournaments  $T$  a minimum semi-degree of at least  $(1/4 + \alpha)|T|$  already suffices to guarantee a Hamilton cycle. (However, it is not hard to prove this directly.) It was shown by Bollobás and Häggkvist [5] that this degree condition even ensures the  $k$ th power of a Hamilton cycle (if  $T$  is sufficiently large compared to  $1/\alpha$  and  $k$ ). The degree condition is essentially best possible as a minimum semi-degree of  $|T|/4 - 1$  does not even guarantee a single Hamilton cycle.

Since this paper was written, we have used some of the tools and methods to obtain an exact version of Theorem 1 (but not of Theorems 3 and 4) for large oriented graphs [12] as well as an approximate analogue of Chvátal's theorem on Hamiltonian degree sequences for digraphs [17]. See [13] for related results about short cycles and pancyclicity for oriented graphs.

Our paper is organized as follows. In the next section we introduce some basic definitions and describe the extremal example which shows that Theorem 1 (and thus also Theorems 3 and 4) is essentially best possible. Our proof of Theorem 3 relies on the Regularity lemma for digraphs and on a variant (due to Csaba [6]) of the Blow-up lemma. These and other tools are introduced in Section 3, where we also give an overview of the proof. In Section 4

we collect some preliminary results. Theorem 3 is then proved in Section 5. In the last section we discuss the modifications needed to prove Theorem 4.

## 2. NOTATION AND THE EXTREMAL EXAMPLE

Before we show that Theorems 1, 3 and 4 are essentially best possible, we will introduce the basic notation used throughout the paper. Given two vertices  $x$  and  $y$  of an oriented graph  $G$ , we write  $xy$  for the edge directed from  $x$  to  $y$ . The order  $|G|$  of  $G$  is the number of its vertices. We write  $N_G^+(x)$  for the outneighbourhood of a vertex  $x$  and  $d_G^+(x) := |N_G^+(x)|$  for its outdegree. Similarly, we write  $N_G^-(x)$  for the inneighbourhood of  $x$  and  $d_G^-(x) := |N_G^-(x)|$  for its indegree. We write  $N_G(x) := N_G^+(x) \cup N_G^-(x)$  for the neighbourhood of  $x$  and use  $N^+(x)$  etc. whenever this is unambiguous. We write  $\Delta(G)$  for the maximum of  $|N(x)|$  over all vertices  $x \in G$ .

Given a set  $A$  of vertices of  $G$ , we write  $N_G^+(A)$  for the set of all outneighbours of vertices in  $A$ . So  $N_G^+(A)$  is the union of  $N_G^+(a)$  over all  $a \in A$ .  $N_G^-(A)$  is defined similarly. The oriented subgraph of  $G$  induced by  $A$  is denoted by  $G[A]$ . Given two vertices  $x, y$  of  $G$ , an  $x$ - $y$  path is a directed path which joins  $x$  to  $y$ . Given two disjoint subsets  $A$  and  $B$  of vertices of  $G$ , an  $A$ - $B$  edge is an edge  $ab$  where  $a \in A$  and  $b \in B$ , the set of these edges is denoted by  $E_G(A, B)$  and we put  $e_G(A, B) := |E_G(A, B)|$ .

Recall that when referring to paths and cycles in oriented graphs we always mean that they are directed without mentioning this explicitly. Given two vertices  $x$  and  $y$  on a directed cycle  $C$ , we write  $xCy$  for the subpath of  $C$  from  $x$  to  $y$ . Similarly, given two vertices  $x$  and  $y$  on a directed path  $P$  such that  $x$  precedes  $y$ , we write  $xPy$  for the subpath of  $P$  from  $x$  to  $y$ . A walk in an oriented graph  $G$  is a sequence of (not necessary distinct) vertices  $v_1, v_2, \dots, v_\ell$  where  $v_i v_{i+1}$  is an edge for all  $1 \leq i < \ell$ . The walk is *closed* if  $v_1 = v_\ell$ . A 1-factor of  $G$  is a collection of disjoint cycles which cover all the vertices of  $G$ . We define things similarly for graphs and for directed graphs. The *underlying graph* of an oriented graph  $G$  is the graph obtained from  $G$  by ignoring the directions of its edges.

Given disjoint vertex sets  $A$  and  $B$  in a graph  $G$ , we write  $(A, B)_G$  for the induced bipartite subgraph of  $G$  whose vertex classes are  $A$  and  $B$ . We write  $(A, B)$  where this is unambiguous. We call an orientation of a complete graph a *tournament* and an orientation of a complete bipartite graph a *bipartite tournament*. An oriented graph  $G$  is  $d$ -regular if all vertices have in- and outdegree  $d$ .  $G$  is *regular* if it is  $d$ -regular for some  $d$ . It is easy to see (e.g. by induction) that for every odd  $n$  there exists a regular tournament on  $n$  vertices. Throughout the paper we omit floors and ceilings whenever this does not affect the argument.

The following construction by Häggkvist [10] shows that Conjecture 2 is best possible for infinitely many values of  $|G|$ . We include it here for completeness.

**Proposition 6.** *There are infinitely many oriented graphs  $G$  with minimum semi-degree  $(3|G| - 5)/8$  which do not contain a 1-factor and thus do not contain a Hamilton cycle.*

**Proof.** Let  $n := 4m + 3$  for some odd  $m \in \mathbb{N}$ . Let  $G$  be the oriented graph obtained from the disjoint union of two regular tournaments  $A$  and  $C$  on  $m$  vertices, a set  $B$  of  $m + 2$  vertices and a set  $D$  of  $m + 1$  vertices by adding all edges from  $A$  to  $B$ , all edges from  $B$  to  $C$ , all edges from  $C$  to  $D$  as well as all edges from  $D$  to  $A$ . Finally, between  $B$  and  $D$  we add edges to obtain a bipartite tournament which is as regular as possible, i.e. the in- and outdegree of every vertex differ by at most 1. So in particular every vertex in  $B$  sends exactly  $(m + 1)/2$  edges to  $D$  (Figure 1).

It is easy to check that the minimum semi-degree of  $G$  is  $(m - 1)/2 + (m + 1) = (3n - 5)/8$ , as required. Since every path which joins two vertices in  $B$  has to pass through  $D$ , it follows

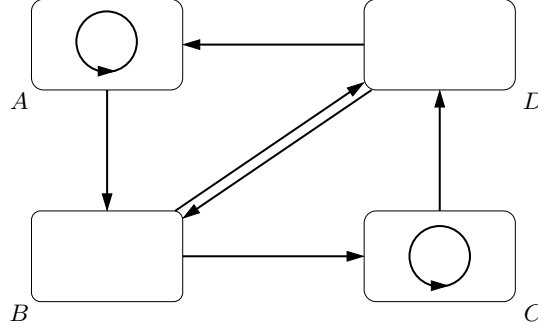


FIGURE 1. The oriented graph in the proof of Proposition 6.

that every cycle contains at least as many vertices from  $D$  as it contains from  $B$ . As  $|B| > |D|$  this means that one cannot cover all the vertices of  $G$  by disjoint cycles, i.e.  $G$  does not contain a 1-factor.  $\square$

### 3. THE DIREGULARITY LEMMA, THE BLOW-UP LEMMA AND OTHER TOOLS

**3.1. The Diregularity lemma and the Blow-up lemma.** In this section we collect all the information we need about the Diregularity lemma and the Blow-up lemma. See [16] for a survey on the Regularity lemma and [14] for a survey on the Blow-up lemma. We start with some more notation. The density of a bipartite graph  $G = (A, B)$  with vertex classes  $A$  and  $B$  is defined to be

$$d_G(A, B) := \frac{e_G(A, B)}{|A||B|}.$$

We often write  $d(A, B)$  if this is unambiguous. Given  $\varepsilon > 0$ , we say that  $G$  is  $\varepsilon$ -regular if for all subsets  $X \subseteq A$  and  $Y \subseteq B$  with  $|X| > \varepsilon|A|$  and  $|Y| > \varepsilon|B|$  we have that  $|d(X, Y) - d(A, B)| < \varepsilon$ . Given  $d \in [0, 1]$  we say that  $G$  is  $(\varepsilon, d)$ -super-regular if it is  $\varepsilon$ -regular and furthermore  $d_G(a) \geq (d - \varepsilon)|B|$  for all  $a \in A$  and  $d_G(b) \geq (d - \varepsilon)|A|$  for all  $b \in B$ . (This is a slight variation of the standard definition of  $(\varepsilon, d)$ -super-regularity where one requires  $d_G(a) \geq d|B|$  and  $d_G(b) \geq d|A|$ .)

The Diregularity lemma is a version of the Regularity lemma for digraphs due to Alon and Shapira [1]. Its proof is quite similar to the undirected version. We will use the degree form of the Diregularity lemma which can be easily derived (see e.g. [24]) from the standard version, in exactly the same manner as the undirected degree form.

**Lemma 7** (Degree form of the Diregularity lemma). *For every  $\varepsilon \in (0, 1)$  and every integer  $M'$  there are integers  $M$  and  $n_0$  such that if  $G$  is a digraph on  $n \geq n_0$  vertices and  $d \in [0, 1]$  is any real number, then there is a partition of the vertices of  $G$  into  $V_0, V_1, \dots, V_k$ , a spanning subdigraph  $G'$  of  $G$  and a set  $U$  of ordered pairs  $V_i V_j$  (where  $1 \leq i, j \leq k$  and  $i \neq j$ ) such that the following holds:*

- $M' \leq k \leq M$ ,
- $|V_0| \leq \varepsilon n$ ,
- $|V_1| = \dots = |V_k| =: m$ ,
- $d_{G'}^+(x) > d_G^+(x) - (d + \varepsilon)n$  for all vertices  $x \in G$ ,

- $d_{G'}^-(x) > d_G^-(x) - (d + \varepsilon)n$  for all vertices  $x \in G$ ,
- $|U| \leq \varepsilon k^2$ ,
- for every ordered pair  $V_i V_j \notin U$  with  $1 \leq i, j \leq k$  and  $i \neq j$  the bipartite graph  $(V_i, V_j)_G$  whose vertex classes are  $V_i$  and  $V_j$  and whose edge set is the set  $E_G(V_i, V_j)$  of all the  $V_i$ - $V_j$  edges in  $G$  is  $\varepsilon$ -regular,
- $G'$  is obtained from  $G$  by deleting the following edges of  $G$ : all edges with both endvertices in  $V_i$  for all  $i \geq 1$  as well as all edges in  $E_G(V_i, V_j)$  for all  $V_i V_j \in U$  and for all those  $V_i V_j \notin U$  with  $1 \leq i, j \leq k$  and  $i \neq j$  for which the density of  $(V_i, V_j)_G$  is less than  $d$ .

$V_1, \dots, V_k$  are called *clusters*,  $V_0$  is called the *exceptional set* and the vertices in  $V_0$  are called *exceptional vertices*.  $U$  is called the set of *exceptional pairs of clusters*. Note that the last two conditions of the lemma imply that for all  $1 \leq i, j \leq k$  with  $i \neq j$  the bipartite graph  $(V_i, V_j)_{G'}$  is  $\varepsilon$ -regular and has density either 0 or density at least  $d$ . In particular, in  $G'$  all pairs of clusters are  $\varepsilon$ -regular in both directions (but possibly with different densities). We call the spanning digraph  $G' \subseteq G$  given by the Diregularity lemma the *pure digraph*. Given clusters  $V_1, \dots, V_k$  and the pure digraph  $G'$ , the *reduced digraph*  $R'$  is the digraph whose vertices are  $V_1, \dots, V_k$  and in which  $V_i V_j$  is an edge if and only if  $G'$  contains a  $V_i$ - $V_j$  edge. Note that the latter holds if and only if  $(V_i, V_j)_{G'}$  is  $\varepsilon$ -regular and has density at least  $d$ . It turns out that  $R'$  inherits many properties of  $G$ , a fact that is crucial in our proof. However,  $R'$  is not necessarily oriented even if the original digraph  $G$  is, but the next lemma shows that by discarding edges with appropriate probabilities one can go over to a reduced oriented graph  $R \subseteq R'$  which still inherits many of the properties of  $G$ . (d) will only be used in the proof of Theorem 4.

**Lemma 8.** *For every  $\varepsilon \in (0, 1)$  there exist integers  $M' = M'(\varepsilon)$  and  $n_0 = n_0(\varepsilon)$  such that the following holds. Let  $d \in [0, 1]$  and let  $G$  be an oriented graph of order at least  $n_0$  and let  $R'$  be the reduced digraph and  $U$  the set of exceptional pairs of clusters obtained by applying the Diregularity lemma to  $G$  with parameters  $\varepsilon$ ,  $d$  and  $M'$ . Then  $R'$  has a spanning oriented subgraph  $R$  with*

- (a)  $\delta^+(R) \geq (\delta^+(G)/|G| - (3\varepsilon + d))|R|$ ,
- (b)  $\delta^-(R) \geq (\delta^-(G)/|G| - (3\varepsilon + d))|R|$ ,
- (c)  $\delta(R) \geq (\delta(G)/|G| - (3\varepsilon + 2d))|R|$ ,
- (d) if  $2\varepsilon \leq d \leq 1 - 2\varepsilon$  and  $c \geq 0$  is such that  $d^+(x) + d^-(y) \geq c|G|$  whenever  $xy \notin E(G)$  then  $d_R^+(V_i) + d_R^-(V_j) \geq (c - 6\varepsilon - 2d)|R|$  whenever  $V_i V_j \notin E(R) \cup U$ .

**Proof.** Let us first show that every cluster  $V_i$  satisfies

$$(1) \quad |N_{R'}(V_i)|/|R'| \geq \delta(G)/|G| - (3\varepsilon + 2d).$$

To see this, consider any vertex  $x \in V_i$ . As  $G$  is an oriented graph, the Diregularity lemma implies that  $|N_{G'}(x)| \geq \delta(G) - 2(d + \varepsilon)|G|$ . On the other hand,  $|N_{G'}(x)| \leq |N_{R'}(V_i)|m + |V_0| \leq |N_{R'}(V_i)||G|/|R'| + \varepsilon|G|$ . Altogether this proves (1).

We first consider the case when

$$(2) \quad \delta^+(G)/|G| \geq 3\varepsilon + d \quad \text{and} \quad \delta^-(G)/|G| \geq 3\varepsilon + d \quad \text{and} \quad c \geq 6\varepsilon + 2d.$$

Let  $R$  be the spanning oriented subgraph obtained from  $R'$  by deleting edges randomly as follows. For every unordered pair  $V_i, V_j$  of clusters we delete the edge  $V_i V_j$  (if it exists) with probability

$$(3) \quad \frac{e_{G'}(V_j, V_i)}{e_{G'}(V_i, V_j) + e_{G'}(V_j, V_i)}.$$

Otherwise we delete  $V_j V_i$  (if it exists). We interpret (3) as 0 if  $V_i V_j, V_j V_i \notin E(R')$ . So if  $R'$  contains at most one of the edges  $V_i V_j, V_j V_i$  then we do nothing. We do this for all unordered pairs of clusters independently and let  $X_i$  be the random variable which counts the number of outedges of the vertex  $V_i \in R$ . Then

$$\begin{aligned}
 \mathbb{E}(X_i) &= \sum_{j \neq i} \frac{e_{G'}(V_i, V_j)}{e_{G'}(V_i, V_j) + e_{G'}(V_j, V_i)} \geq \sum_{j \neq i} \frac{e_{G'}(V_i, V_j)}{|V_i| |V_j|} \\
 (4) \quad &\geq \frac{|R'|}{|G| |V_i|} \sum_{x \in V_i} (d_{G'}^+(x) - |V_0|) \\
 &\geq (\delta^+(G')/|G| - \varepsilon) |R| \geq (\delta^+(G)/|G| - (2\varepsilon + d)) |R| \stackrel{(2)}{\geq} \varepsilon |R|.
 \end{aligned}$$

A Chernoff-type bound (see e.g. [2, Cor. A.14]) now implies that there exists a constant  $\beta = \beta(\varepsilon)$  such that

$$\begin{aligned}
 \mathbb{P}(X_i < (\delta^+(G)/|G| - (3\varepsilon + d)) |R|) &\leq \mathbb{P}(|X_i - \mathbb{E}(X_i)| > \varepsilon \mathbb{E}(X_i)) \\
 &\leq e^{-\beta \mathbb{E}(X_i)} \leq e^{-\beta \varepsilon |R|}.
 \end{aligned}$$

Writing  $Y_i$  for the random variable which counts the number of inedges of the vertex  $V_i$  in  $R$ , it follows similarly that

$$\mathbb{P}(Y_i < (\delta^-(G)/|G| - (3\varepsilon + d)) |R|) \leq e^{-\beta \varepsilon |R|}.$$

Suppose that  $c$  is as in (d). Consider any pair  $V_i V_j \notin U$  of clusters such that either  $V_i V_j \notin E(R')$  or  $V_i V_j, V_j V_i \in E(R')$ . (Note that each  $V_i V_j \notin E(R) \cup U$  satisfies one of these properties.) As before, let  $X_i$  be the random variable which counts the number of outedges of  $V_i$  in  $R$  and let  $Y_j$  be the number of inedges of  $V_j$  in  $R$ . Similary as in (4) one can show that

$$(5) \quad \mathbb{E}(X_i + Y_j) \geq \frac{|R'|}{|G| |V_i|} \left( \sum_{x \in V_i} (d_{G'}^+(x) - |V_0|) + \sum_{y \in V_j} (d_{G'}^-(y) - |V_0|) \right).$$

To estimate this, we will first show that there is a set  $M$  of at least  $(1 - \varepsilon)|V_i|$  disjoint pairs  $(x, y)$  with  $x \in V_i, y \in V_j$  and such that  $xy \notin E(G)$ . Suppose first that  $V_i V_j, V_j V_i \in E(R')$ . But then  $(V_j, V_i)_G$  is  $\varepsilon$ -regular of density at least  $d$  and thus it contains a matching of size at least  $(1 - \varepsilon)|V_i|$ . As  $G$  is oriented this matching corresponds to a set  $M$  as required. If  $V_i V_j \notin E(R')$  then  $(V_i, V_j)_G$  is  $\varepsilon$ -regular of density less than  $d$  (since  $V_i V_j \notin U$ ). Thus the complement of  $(V_i, V_j)_G$  is  $\varepsilon$ -regular of density at least  $1 - d$  and so contains a matching of size at least  $(1 - \varepsilon)|V_i|$  which again corresponds to a set  $M$  as required. Together with (5) this implies that

$$\begin{aligned}
 \mathbb{E}(X_i + Y_j) &\geq \frac{|R'|}{|G| |V_i|} \sum_{(x, y) \in M} (d_{G'}^+(x) + d_{G'}^-(y) - 2|V_0|) \\
 &\geq \frac{|R'|}{|G| |V_i|} (c - 2(\varepsilon + d) - 2\varepsilon) |G| (1 - \varepsilon) |V_i| \geq (c - (5\varepsilon + 2d)) |R| \stackrel{(2)}{\geq} \varepsilon |R|.
 \end{aligned}$$

Similarly as before a Chernoff-type bound implies that

$$\mathbb{P}(X_i + Y_j < (c - (6\varepsilon + 2d)) |R|) \leq e^{-\beta \varepsilon |R|}.$$

As  $2|R|^2 e^{-\beta \varepsilon |R|} < 1$  if  $M'$  is chosen to be sufficiently large compared to  $\varepsilon$ , this implies that there is some outcome  $R$  which satisfies (a), (b) and (d). But  $N_{R'}(V_i) = N_R(V_i)$  for every

cluster  $V_i$  and so (1) implies that  $\delta(R) \geq (\delta(G)/|G| - (3\varepsilon + 2d))|R|$ . Altogether this shows that  $R$  is as required in the lemma.

If neither of the conditions in (2) hold, then (a), (b) and (d) are trivial and one can obtain an oriented graph  $R$  which satisfies (c) from  $R'$  by arbitrarily deleting one edge from each double edge. If for example only the first of the conditions in (2) holds, then (b) and (d) are trivial. To obtain an oriented graph  $R$  which satisfies (a) we consider the  $X_i$  as before, but ignore the  $Y_i$  and the sums  $X_i + Y_j$ . Again,  $N_{R'}(V_i) = N_R(V_i)$  for every cluster  $V_i$  and so (c) is also satisfied. The other cases are similar.  $\square$

The oriented graph  $R$  given by Lemma 8 is called the *reduced oriented graph*. The spanning oriented subgraph  $G^*$  of the pure digraph  $G'$  obtained by deleting all the  $V_iV_j$  edges whenever  $V_iV_j \in E(R') \setminus E(R)$  is called the *pure oriented graph*. Given an oriented subgraph  $S \subseteq R$ , the *oriented subgraph of  $G^*$  corresponding to  $S$*  is the oriented subgraph obtained from  $G^*$  by deleting all those vertices that lie in clusters not belonging to  $S$  as well as deleting all the  $V_iV_j$  edges for all pairs  $V_i, V_j$  with  $V_iV_j \notin E(S)$ .

In our proof of Theorem 3 we will also need the Blow-up lemma. Roughly speaking, it states the following. Let  $F$  be a graph on  $r$  vertices, let  $K$  be a graph obtained from  $F$  by replacing each vertex of  $F$  with a cluster and replacing each edge with a complete bipartite graph between the corresponding clusters. Define  $G$  similarly except that the edges of  $F$  now correspond to dense  $\varepsilon$ -super-regular pairs. Then every subgraph  $H$  of  $K$  which has bounded maximum degree is also a subgraph in  $G$ . In the original version of Komlós, Sárközy and Szemerédi [15]  $\varepsilon$  has to be sufficiently small compared to  $1/r$  (and so in particular we cannot take  $r = |R|$ ). We will use a stronger (and more technical) version due to Csaba [6], which allows us to take  $r = |R|$  and does not demand super-regularity. The case when  $\Delta = 3$  of this is implicit in [7].

In the statement of Lemma 9 and later on we write  $0 < a_1 \ll a_2 \ll a_3$  to mean that we can choose the constants  $a_1, a_2, a_3$  from right to left. More precisely, there are increasing functions  $f$  and  $g$  such that, given  $a_3$ , whenever we choose some  $a_2 \leq f(a_3)$  and  $a_1 \leq g(a_2)$ , all calculations needed in the proof of Lemma 9 are valid. Hierarchies with more constants are defined in the obvious way.

**Lemma 9** (Blow-up Lemma, Csaba [6]). *For all integers  $\Delta, K_1, K_2, K_3$  and every positive constant  $c$  there exists an integer  $N$  such that whenever  $\varepsilon, \varepsilon', \delta', d$  are positive constants with*

$$0 < \varepsilon \ll \varepsilon' \ll \delta' \ll d \ll 1/\Delta, 1/K_1, 1/K_2, 1/K_3, c$$

*the following holds. Suppose that  $G^*$  is a graph of order  $n \geq N$  and  $V_0, \dots, V_k$  is a partition of  $V(G^*)$  such that the bipartite graph  $(V_i, V_j)_{G^*}$  is  $\varepsilon$ -regular with density either 0 or  $d$  for all  $1 \leq i < j \leq k$ . Let  $H$  be a graph on  $n$  vertices with  $\Delta(H) \leq \Delta$  and let  $L_0 \cup L_1 \cup \dots \cup L_k$  be a partition of  $V(H)$  with  $|L_i| = |V_i| =: m$  for every  $i = 1, \dots, k$ . Furthermore, suppose that there exists a bijection  $\phi : L_0 \rightarrow V_0$  and a set  $I \subseteq V(H)$  of vertices at distance at least 4 from each other such that the following conditions hold:*

- (C1)  $|L_0| = |V_0| \leq K_1 dn$ .
- (C2)  $L_0 \subseteq I$ .
- (C3)  $L_i$  is independent for every  $i = 1, \dots, k$ .
- (C4)  $|N_H(L_0) \cap L_i| \leq K_2 dm$  for every  $i = 1, \dots, k$ .
- (C5) For each  $i = 1, \dots, k$  there exists  $D_i \subseteq I \cap L_i$  with  $|D_i| = \delta' m$  and such that for  $D := \bigcup_{i=1}^k D_i$  and all  $1 \leq i < j \leq k$

$$||N_H(D) \cap L_i| - |N_H(D) \cap L_j|| < \varepsilon m.$$

- (C6) If  $xy \in E(H)$  and  $x \in L_i, y \in L_j$  where  $i, j \neq 0$  then  $(V_i, V_j)_{G^*}$  is  $\varepsilon$ -regular with density  $d$ .
- (C7) If  $xy \in E(H)$  and  $x \in L_0, y \in L_j$  then  $|N_{G^*}(\phi(x)) \cap V_j| \geq cm$ .
- (C8) For each  $i = 1, \dots, k$ , given any  $E_i \subseteq V_i$  with  $|E_i| \leq \varepsilon' m$  there exists a set  $F_i \subseteq (L_i \cap (I \setminus D))$  and a bijection  $\phi_i : E_i \rightarrow F_i$  such that  $|N_{G^*}(v) \cap V_j| \geq (d - \varepsilon)m$  whenever  $N_H(\phi_i(v)) \cap L_j \neq \emptyset$  (for all  $v \in E_i$  and all  $j = 1, \dots, k$ ).
- (C9) Writing  $F := \bigcup_{i=1}^k F_i$  we have that  $|N_H(F) \cap L_i| \leq K_3 \varepsilon' m$ .

Then  $G^*$  contains a copy of  $H$  such that the image of  $L_i$  is  $V_i$  for all  $i = 1, \dots, k$  and the image of each  $x \in L_0$  is  $\phi(x) \in V_0$ .

The additional properties of the copy of  $H$  in  $G^*$  are not included in the statement of the lemma in [6] but are stated explicitly in the proof.

Let us briefly motivate the conditions of the Blow-up lemma. The embedding of  $H$  into  $G$  guaranteed by the Blow-up lemma is found by a randomized algorithm which first embeds each vertex  $x \in L_0$  to  $\phi(x)$  and then successively embeds the remaining vertices of  $H$ . So the image of  $L_0$  will be the exceptional set  $V_0$ . Condition (C1) requires that there are not too many exceptional vertices and (C2) ensures that we can embed the vertices in  $L_0$  without affecting the neighbourhood of other such vertices. As  $L_i$  will be embedded into  $V_i$  we need to have (C3). Condition (C5) gives us a reasonably large set  $D$  of ‘buffer vertices’ which will be embedded last by the randomized algorithm. (C6) requires that edges between vertices of  $H - L_0$  are embedded into  $\varepsilon$ -regular pairs of density  $d$ . (C7) ensures that the exceptional vertices have large degree in all ‘neighbouring clusters’. (C8) and (C9) allow us to embed those vertices whose set of candidate images in  $G^*$  has grown very small at some point of the algorithm. Conditions (C6), (C8) and (C9) correspond to a substantial weakening of the super-regularity that the usual form of the Blow-up lemma requires, namely that whenever  $H$  contains an edge  $xy$  with  $x \in L_i, y \in L_j$  then  $(V_i, V_j)_{G^*}$  is  $(\varepsilon, d)$ -super-regular.

We would like to apply the Blow-up lemma with  $G^*$  being obtained from the underlying graph of the pure oriented graph by adding the exceptional vertices. It will turn out that in order to satisfy (C8), it suffices to ensure that all the edges of a suitable 1-factor in the reduced oriented graph  $R$  correspond to  $(\varepsilon, d)$ -superregular pairs of clusters. A well-known simple fact (see the first part of the proof of Proposition 10) states that this can be ensured by removing a small proportion of vertices from each cluster  $V_i$ , and so (C8) will be satisfied. However, (C6) requires all the edges of  $R$  to correspond to  $\varepsilon$ -regular pairs of density precisely  $d$  and not just at least  $d$ . (As remarked by Csaba [6], it actually suffices that the densities are close to  $d$  in terms of  $\varepsilon$ .) The second part of the following proposition shows that this too does not pose a problem.

**Proposition 10.** *Let  $M', n_0, D$  be integers and let  $\varepsilon, d$  be positive constants such that  $1/n_0 \ll 1/M' \ll \varepsilon \ll d \ll 1/D$ . Let  $G$  be an oriented graph of order at least  $n_0$ . Let  $R$  be the reduced oriented graph and let  $G^*$  be the pure oriented graph obtained by successively applying first the Diregularity lemma with parameters  $\varepsilon, d$  and  $M'$  to  $G$  and then Lemma 8. Let  $S$  be an oriented subgraph of  $R$  with  $\Delta(S) \leq D$ . Let  $G'$  be the underlying graph of  $G^*$ . Then one can delete  $2D\varepsilon|V_i|$  vertices from each cluster  $V_i$  to obtain subclusters  $V'_i \subseteq V_i$  in such a way that  $G'$  contains a subgraph  $G'_S$  whose vertex set is the union of all the  $V'_i$  and such that*

- $(V'_i, V'_j)_{G'_S}$  is  $(\sqrt{\varepsilon}, d - 4D\varepsilon)$ -superregular whenever  $V_i V_j \in E(S)$ ,
- $(V'_i, V'_j)_{G'_S}$  is  $\sqrt{\varepsilon}$ -regular and has density  $d - 4D\varepsilon$  whenever  $V_i V_j \in E(R)$ .

**Proof.** Consider any cluster  $V_i \in V(S)$  and any neighbour  $V_j$  of  $V_i$  in  $S$ . Recall that  $m = |V_i|$ . Let  $d_{ij}$  denote the density of the bipartite subgraph  $(V_i, V_j)_{G'}$  of  $G'$  induced by  $V_i$  and  $V_j$ . So  $d_{ij} \geq d$  and this bipartite graph is  $\varepsilon$ -regular. Thus there are at most  $2\varepsilon m$  vertices



$v \in V_i$  such that  $||N_{G'}(v) \cap V_j| - d_{ij}m| > \varepsilon m$ . So in total there are at most  $2D\varepsilon m$  vertices  $v \in V_i$  such that  $||N_{G'}(v) \cap V_j| - d_{ij}m| > \varepsilon m$  for some neighbour  $V_j$  of  $V_i$  in  $S$ . Delete all these vertices as well as some more vertices if necessary to obtain a subcluster  $V'_i \subseteq V_i$  of size  $(1 - 2D\varepsilon)m =: m'$ . Delete any  $2D\varepsilon m$  vertices from each cluster  $V_i \in V(R) \setminus V(S)$  to obtain a subcluster  $V'_i$ . It is easy to check that for each edge  $V_i V_j \in E(R)$  the graph  $(V'_i, V'_j)_{G'}$  is still  $2\varepsilon$ -regular and that its density  $d'_{ij}$  satisfies

$$d' := d - 4D\varepsilon < d_{ij} - \varepsilon \leq d'_{ij} \leq d_{ij} + \varepsilon.$$

Moreover, whenever  $V_i V_j \in E(S)$  and  $v \in V'_i$  we have that

$$(d_{ij} - 4D\varepsilon)m' \leq |N_{G'}(v) \cap V'_j| \leq (d_{ij} + 4D\varepsilon)m'.$$

For every pair  $V_i, V_j$  of clusters with  $V_i V_j \in E(S)$  we now consider a spanning random subgraph  $G'_{ij}$  of  $(V'_i, V'_j)_{G'}$  which is obtained by choosing each edge of  $(V'_i, V'_j)_{G'}$  with probability  $d'/d'_{ij}$ , independently of the other edges. Consider any vertex  $v \in V'_i$ . Then the expected number of neighbours of  $v$  in  $V'_j$  (in the graph  $G'_{ij}$ ) is at least  $(d_{ij} - 4D\varepsilon)d'm'/d'_{ij} \geq (1 - \sqrt{\varepsilon})d'm'$ . So we can apply a Chernoff-type bound to see that there exists a constant  $c = c(\varepsilon)$  such that

$$\mathbb{P}(|N_{G'_{ij}}(v) \cap V'_j| \leq (d' - \sqrt{\varepsilon})m') \leq e^{-cd'm'}.$$

Similarly, whenever  $X \subseteq V'_i$  and  $Y \subseteq V'_j$  are sets of size at least  $2\varepsilon m'$  the expected number of  $X$ - $Y$  edges in  $G'_{ij}$  is  $d_{G'}(X, Y)d'|X||Y|/d'_{ij}$ . Since  $(V'_i, V'_j)_{G'}$  is  $2\varepsilon$ -regular this expected number lies between  $(1 - \sqrt{\varepsilon})d'|X||Y|$  and  $(1 + \sqrt{\varepsilon})d'|X||Y|$ . So again we can use a Chernoff-type bound to see that

$$\mathbb{P}(|e_{G'_{ij}}(X, Y) - d'|X||Y|| > \sqrt{\varepsilon}|X||Y|) \leq e^{-cd'|X||Y|} \leq e^{-4cd'(\varepsilon m')^2}.$$

Moreover, with probability at least  $1/(3m')$  the graph  $G'_{ij}$  has its expected density  $d'$  (see e.g. [4, p. 6]). Altogether this shows that with probability at least

$$1/(3m') - 2m'e^{-cd'm'} - 2^{2m'}e^{-4cd'(\varepsilon m')^2} > 0$$

we have that  $G'_{ij}$  is  $(\sqrt{\varepsilon}, d')$ -superregular and has density  $d'$ . Proceed similarly for every pair of clusters forming an edge of  $S$ . An analogous argument applied to a pair  $V_i, V_j$  of clusters with  $V_i V_j \in E(R) \setminus E(S)$  shows that with non-zero probability the random subgraph  $G'_{ij}$  is  $\sqrt{\varepsilon}$ -regular and has density  $d'$ . Altogether this gives us the desired subgraph  $G'_S$  of  $G'$ .  $\square$

**3.2. Overview of the proof of Theorem 3.** Let  $G$  be our given oriented graph. The rough idea of the proof is to apply the Diregularity lemma and Lemma 8 to obtain a reduced oriented graph  $R$  and a pure oriented graph  $G^*$ . The following result of Häggkvist implies that  $R$  contains a 1-factor.

**Theorem 11** (Häggkvist [10]). *Let  $R$  be an oriented graph with  $\delta^*(R) > (3|R| - 3)/2$ . Then  $R$  is strongly connected and contains a 1-factor.*

So one can apply the Blow-up lemma (together with Proposition 10) to find a 1-factor in  $G^* - V_0 \subseteq G - V_0$ . One now would like to glue the cycles of this 1-factor together and to incorporate the exceptional vertices to obtain a Hamilton cycle of  $G^*$  and thus of  $G$ . However, we were only able to find a method which incorporates a set of vertices whose size is small compared to the cluster size  $m$ . This is not necessarily the case for  $V_0$ . So we proceed as follows. We first choose a random partition of the vertex set of  $G$  into two sets  $A$  and  $V(G) \setminus A$  having roughly equal size. We then apply the Diregularity lemma to  $G - A$  in order to obtain clusters  $V_1, \dots, V_k$  and an exceptional set  $V_0$ . We let  $m$  denote the size of

these clusters and set  $B := V_1 \cup \dots \cup V_k$ . By arguing as indicated above, we can find a Hamilton cycle  $C_B$  in  $G[B]$ . We then apply the Diregularity lemma to  $G - B$ , but with an  $\varepsilon$  which is small compared to  $1/k$ , to obtain clusters  $V'_1, \dots, V'_\ell$  and an exceptional set  $V'_0$ . Since the choice of our partition  $A, V(G) \setminus A$  will imply that  $\delta^*(G - B) \geq (3/2 + \alpha/2)|G - B|$  we can again argue as before to obtain a cycle  $C_A$  which covers precisely the vertices in  $A' := V'_1 \cup \dots \cup V'_\ell$ . Since we have chosen  $\varepsilon$  to be small compared to  $1/k$ , the set  $V'_0$  of exceptional vertices is now small enough to be incorporated into our first cycle  $C_B$ . (Actually,  $C_B$  is only determined at this point and not yet earlier on.) Moreover, by choosing  $C_B$  and  $C_A$  suitably we can ensure that they can be joined together into the desired Hamilton cycle of  $G$ .

#### 4. SHIFTED WALKS

In this section we will introduce the tools we need in order to glue certain cycles together and to incorporate the exceptional vertices. Let  $R^*$  be a digraph and let  $\mathcal{C}$  be a collection of disjoint cycles in  $R^*$ . We call a closed walk  $W$  in  $R^*$  *balanced w.r.t.  $\mathcal{C}$*  if

- for each cycle  $C \in \mathcal{C}$  the walk  $W$  visits all the vertices on  $C$  an equal number of times,
- $W$  visits every vertex of  $R^*$ ,
- every vertex not in any cycle from  $\mathcal{C}$  is visited exactly once.

Let us now explain why balanced walks are helpful in order to incorporate the exceptional vertices. Suppose that  $\mathcal{C}$  is a 1-factor of the reduced oriented graph  $R$  and that  $R^*$  is obtained from  $R$  by adding all the exceptional vertices  $v \in V_0$  and adding an edge  $vV_i$  (where  $V_i$  is a cluster) whenever  $v$  sends edges to a significant proportion of the vertices in  $V_i$ , say we add  $vV_i$  whenever  $v$  sends at least  $cm$  edges to  $V_i$ . (Recall that  $m$  denotes the size of the clusters.) The edges in  $R^*$  of the form  $V_iv$  are defined in a similar way. Let  $G^c$  be the oriented graph obtained from the pure oriented graph  $G^*$  by making all the non-empty bipartite subgraphs between the clusters complete (and orienting all the edges between these clusters in the direction induced by  $R$ ) and adding the vertices in  $V_0$  as well as all the edges of  $G$  between  $V_0$  and  $V(G - V_0)$ . Suppose that  $W$  is a balanced closed walk in  $R^*$  which visits all the vertices lying on a cycle  $C \in \mathcal{C}$  precisely  $m_C \leq m$  times. Furthermore, suppose that  $|V_0| \leq cm/2$  and that the vertices in  $V_0$  have distance at least 3 from each other on  $W$ . Then by ‘winding around’ each cycle  $C \in \mathcal{C}$  precisely  $m - m_C$  times (at the point when  $W$  first visits  $C$ ) we can obtain a Hamilton cycle in  $G^c$ . Indeed, the two conditions on  $V_0$  ensure that the neighbours of each  $v \in V_0$  on the Hamilton cycle can be chosen amongst the at least  $cm$  neighbours of  $v$  in the neighbouring clusters of  $v$  on  $W$  in such a way that they are distinct for different exceptional vertices. The idea then is to apply the Blow-up lemma to show that this Hamilton cycle corresponds to one in  $G$ . So our aim is to find such a balanced closed walk in  $R^*$ . However, as indicated in Section 3.2, the difficulties arising when trying to ensure that the exceptional vertices lie on this walk will force us to apply the above argument to the subgraphs induced by a random partition of our given oriented graph  $G$ .

Let us now go back to the case when  $R^*$  is an arbitrary digraph and  $\mathcal{C}$  is a collection of disjoint cycles in  $R^*$ . Given vertices  $a, b \in R^*$ , a *shifted  $a$ - $b$  walk* is a walk of the form

$$W = aa_1C_1b_1a_2C_2b_2 \dots a_tC_tb_t$$

where  $C_1, \dots, C_t$  are (not necessarily distinct) cycles from  $\mathcal{C}$  and  $a_i$  is the successor of  $b_i$  on  $C_i$  for all  $i \leq t$ . (We might have  $t = 0$ . So an edge  $ab$  is a shifted  $a$ - $b$  walk.) We call  $C_1, \dots, C_t$  the cycles which are *traversed* by  $W$ . So even if the cycles  $C_1, \dots, C_t$  are not distinct, we say that  $W$  traverses  $t$  cycles. Note that for every cycle  $C \in \mathcal{C}$  the walk

$W - \{a, b\}$  visits the vertices on  $C$  an equal number of times. Thus it will turn out that by joining the cycles from  $\mathcal{C}$  suitably via shifted walks and incorporating those vertices of  $R^*$  not covered by the cycles from  $\mathcal{C}$  we can obtain a balanced closed walk on  $R^*$ .

Our next lemma will be used to show that if  $R^*$  is oriented and  $\delta^*(R^*) \geq (3/2 + \alpha)|R^*|$  then any two vertices of  $R^*$  can be joined by a shifted walk traversing only a small number of cycles from  $\mathcal{C}$  (see Corollary 14). The lemma itself shows that the  $\delta^*$  condition implies expansion, and this will give us the ‘expansion with respect to shifted neighbourhoods’ we need for the existence of shifted walks. The proof of Lemma 12 is similar to that of Theorem 11.

**Lemma 12.** *Let  $R^*$  be an oriented graph on  $N$  vertices with  $\delta^*(R^*) \geq (3/2 + \alpha)N$  for some  $\alpha > 0$ . If  $X \subseteq V(R^*)$  is nonempty and  $|X| \leq (1 - \alpha)N$  then  $|N^+(X)| \geq |X| + \alpha N/2$ .*

**Proof.** For simplicity, we write  $\delta := \delta(R^*)$ ,  $\delta^+ := \delta^+(R^*)$  and  $\delta^- := \delta^-(R^*)$ . Suppose the assertion is false, i.e. there exists  $X \subseteq V(R^*)$  with  $|X| \leq (1 - \alpha)N$  and

$$(6) \quad |N^+(X)| < |X| + \alpha N/2.$$

We consider the following partition of  $V(R^*)$ :

$$A := X \cap N^+(X), \quad B := N^+(X) \setminus X, \quad C := V(R^*) \setminus (X \cup N^+(X)), \quad D := X \setminus N^+(X).$$

(6) gives us

$$(7) \quad |D| + \alpha N/2 > |B|.$$

Suppose  $A \neq \emptyset$ . Then by an averaging argument there exists  $x \in A$  with  $|N^+(x) \cap A| < |A|/2$ . Hence  $\delta^+ \leq |N^+(x)| < |B| + |A|/2$ . Combining this with (7) we get

$$(8) \quad |A| + |B| + |D| \geq 2\delta^+ - \alpha N/2.$$

If  $A = \emptyset$  then  $N^+(X) = B$  and so (7) implies  $|D| + \alpha N/2 \geq |B| \geq \delta^+$ . Thus (8) again holds. Similarly, if  $C \neq \emptyset$  then considering the inneighbourhood of a suitable vertex  $x \in C$  gives

$$(9) \quad |B| + |C| + |D| \geq 2\delta^- - \alpha N/2.$$

If  $C = \emptyset$  then the fact that  $|X| \leq (1 - \alpha)N$  and (6) together imply that  $D \neq \emptyset$ . But then  $N^-(D) \subseteq B$  and thus  $|B| \geq \delta^-$ . Together with (7) this shows that (9) holds in this case too.

If  $D = \emptyset$  then trivially  $|A| + |B| + |C| = N \geq \delta$ . If not, then for any  $x \in D$  we have  $N(x) \cap D = \emptyset$  and hence

$$(10) \quad |A| + |B| + |C| \geq |N(x)| \geq \delta.$$

Combining (8), (9) and (10) gives

$$3|A| + 4|B| + 3|C| + 2|D| \geq 2\delta^- + 2\delta^+ + 2\delta - \alpha N = 2\delta^*(R^*) - \alpha N.$$

Finally, substituting (7) gives

$$3N + \alpha N/2 \geq 2\delta^*(R^*) - \alpha N \geq 3N + \alpha N,$$

which is a contradiction. □

As indicated before, we will now use Lemma 12 to prove the existence of shifted walks in  $R^*$  traversing only a small number of cycles from a given 1-factor of  $R^*$ . For this (and later on) the following fact will be useful.

**Fact 13.** *Let  $G$  be an oriented graph with  $\delta^*(G) \geq (3/2 + \alpha)|G|$  for some constant  $\alpha > 0$ . Then  $\delta^0(G) > \alpha|G|$ .*

**Proof.** Suppose that  $\delta^-(G) \leq \alpha|G|$ . As  $G$  is oriented we have that  $\delta^+(G) < |G|/2$  and so  $\delta^*(G) < 3n/2 + \alpha|G|$ , a contradiction. The proof for  $\delta^+(G)$  is similar.  $\square$

**Corollary 14.** *Let  $R^*$  be an oriented graph on  $N$  vertices with  $\delta^*(R^*) \geq (3/2 + \alpha)N$  for some  $\alpha > 0$  and let  $\mathcal{C}$  be a 1-factor in  $R^*$ . Then for any distinct  $x, y \in V(R^*)$  there exists a shifted  $x$ - $y$  walk traversing at most  $2/\alpha$  cycles from  $\mathcal{C}$ .*

**Proof.** Let  $X_i$  be the set of vertices  $v$  for which there is a shifted  $x$ - $v$  walk which traverses at most  $i$  cycles. So  $X_0 = N^+(x) \neq \emptyset$  and  $X_{i+1} = N^+(X_i^-) \cup X_i$ , where  $X_i^-$  is the set of all predecessors of the vertices in  $X_i$  on the cycles from  $\mathcal{C}$ . Suppose that  $|X_i| \leq (1 - \alpha)N$ . Then Lemma 12 implies that

$$|X_{i+1}| \geq |N^+(X_i^-)| \geq |X_i^-| + \alpha N/2 = |X_i| + \alpha N/2.$$

So for  $i^* := \lfloor 2/\alpha \rfloor - 1$ , we must have  $|X_{i^*}| = |X_{i^*}| \geq (1 - \alpha)N$ . But  $|N^-(y)| \geq \delta^-(R^*) > \alpha N$  and so  $N^-(y) \cap X_{i^*}^- \neq \emptyset$ . In other words,  $y \in N^+(X_{i^*}^-)$  and so there is a shifted  $x$ - $y$  walk traversing at most  $i^* + 1$  cycles.  $\square$

**Corollary 15.** *Let  $R^*$  be an oriented graph with  $\delta^*(R^*) \geq (3/2 + \alpha)|R^*|$  for some  $0 < \alpha \leq 1/6$  and let  $\mathcal{C}$  be a 1-factor in  $R^*$ . Then  $R^*$  contains a closed walk which is balanced w.r.t.  $\mathcal{C}$  and meets every vertex at most  $|R^*|/\alpha$  times and traverses each edge lying on a cycle from  $\mathcal{C}$  at least once.*

**Proof.** Let  $C_1, \dots, C_s$  be an arbitrary ordering of the cycles in  $\mathcal{C}$ . For each cycle  $C_i$  pick a vertex  $c_i \in C_i$ . Denote by  $c_i^+$  the successor of  $c_i$  on the cycle  $C_i$ . Corollary 14 implies that for all  $i$  there exists a shifted  $c_i$ - $c_{i+1}^+$  walk  $W_i$  traversing at most  $2/\alpha$  cycles from  $\mathcal{C}$ , where  $c_{s+1} := c_1$ . Then the closed walk

$$W' := c_1^+ C_1 c_1 W_1 c_2^+ C_2 c_2 \dots W_{s-1} c_s^+ C_s c_s W_s c_1^+$$

is balanced w.r.t.  $\mathcal{C}$  by the definition of shifted walks. Since each shifted walk  $W_i$  traverses at most  $2/\alpha$  cycles of  $\mathcal{C}$ , the closed walk  $W$  meets each vertex at most  $(|R^*|/3)(2/\alpha) + 1$  times. Let  $W$  denote the walk obtained from  $W'$  by ‘winding around’ each cycle  $C \in \mathcal{C}$  once more. (That is, for each  $C \in \mathcal{C}$  pick a vertex  $v$  on  $C$  and replace one of the occurrences of  $v$  on  $W'$  by  $vCv$ .) Then  $W$  is still balanced w.r.t.  $\mathcal{C}$ , traverses each edge lying on a cycle from  $\mathcal{C}$  at least once and visits each vertex of  $R^*$  at most  $(|R^*|/3)(2/\alpha) + 2 \leq |R^*|/\alpha$  times as required.  $\square$

## 5. PROOF OF THEOREM 3

**5.1. Partitioning  $G$  and applying the Diregularity lemma.** Let  $G$  be an oriented graph on  $n$  vertices with  $\delta^*(G) \geq (3/2 + \alpha)n$  for some constant  $\alpha > 0$ . Clearly we may assume that  $\alpha \ll 1$ . Define positive constants  $\varepsilon, d$  and integers  $M'_A, M'_B$  such that

$$1/M'_A \ll 1/M'_B \ll \varepsilon \ll d \ll \alpha \ll 1.$$

Throughout this section, we will assume that  $n$  is sufficiently large compared to  $M'_A$  for our estimates to hold. Choose a subset  $A \subseteq V(G)$  with  $(1/2 - \varepsilon)n \leq |A| \leq (1/2 + \varepsilon)n$  and such that every vertex  $x \in G$  satisfies

$$\frac{d^+(x)}{n} - \frac{\alpha}{10} \leq \frac{|N^+(x) \cap A|}{|A|} \leq \frac{d^+(x)}{n} + \frac{\alpha}{10}$$

and such that  $N^-(x) \cap A$  satisfies a similar condition. (The existence of such a set  $A$  can be shown by considering a random partition of  $V(G)$ .) Apply the Diregularity lemma (Lemma 7) with parameters  $\varepsilon^2$ ,  $d + 8\varepsilon^2$  and  $M'_B$  to  $G - A$  to obtain a partition of the vertex set of  $G - A$  into  $k \geq M'_B$  clusters  $V_1, \dots, V_k$  and an exceptional set  $V_0$ . Set  $B := V_1 \cup \dots \cup V_k$  and  $m_B := |V_1| = \dots = |V_k|$ . Let  $R_B$  denote the reduced oriented graph obtained by an application of Lemma 8 and let  $G_B^*$  be the pure oriented graph. Since  $\delta^+(G - A)/|G - A| \geq \delta^+(G)/n - \alpha/9$  by our choice of  $A$ , Lemma 8 implies that

$$(11) \quad \delta^+(R_B) \geq (\delta^+(G)/n - \alpha/8)|R_B|.$$

Similarly

$$(12) \quad \delta^-(R_B) \geq (\delta^-(G)/n - \alpha/8)|R_B|$$

and  $\delta(R_B) \geq (\delta(G)/n - \alpha/4)|R_B|$ . Altogether this implies that

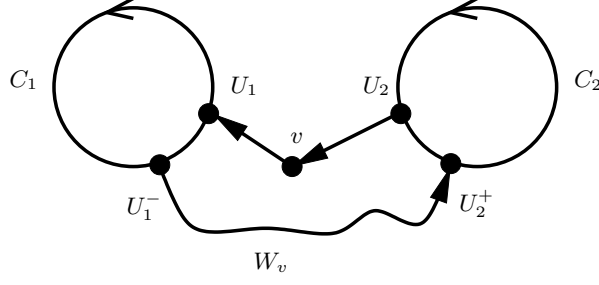
$$(13) \quad \delta^*(R_B) \geq (3/2 + \alpha/2)|R_B|.$$

So Theorem 11 gives us a 1-factor  $\mathcal{C}_B$  of  $R_B$ . We now apply Proposition 10 with  $\mathcal{C}_B$  playing the role of  $S$ ,  $\varepsilon^2$  playing the role of  $\varepsilon$  and  $d + 8\varepsilon^2$  playing the role of  $d$ . This shows that by adding at most  $4\varepsilon^2 n$  further vertices to the exceptional set  $V_0$  we may assume that each edge of  $R_B$  corresponds to an  $\varepsilon$ -regular pair of density  $d$  (in the underlying graph of  $G_B^*$ ) and that each edge in the union  $\bigcup_{C \in \mathcal{C}_B} C \subseteq R_B$  of all the cycles from  $\mathcal{C}_B$  corresponds to an  $(\varepsilon, d)$ -superregular pair. (More formally, this means that we replace the clusters with the subclusters given by Proposition 10 and replace  $G_B^*$  with its oriented subgraph obtained by deleting all edges not corresponding to edges of the graph  $G'_{\mathcal{C}_B}$  given by Proposition 10, i.e. the underlying graph of  $G_B^*$  will now be  $G'_{\mathcal{C}_B}$ .) Note that the new exceptional set now satisfies  $|V_0| \leq \varepsilon n$ .

Apply Corollary 15 with  $R^* := R_B$  to find a closed walk  $W_B$  in  $R_B$  which is balanced w.r.t.  $\mathcal{C}_B$ , meets every cluster at most  $2|R_B|/\alpha$  times and traverses all the edges lying on a cycle from  $\mathcal{C}_B$  at least once.

Let  $G_B^c$  be the oriented graph obtained from  $G_B^*$  by adding all the  $V_i V_j$  edges for all those pairs  $V_i, V_j$  of clusters with  $V_i V_j \in E(R_B)$ . Since  $2|R_B|/\alpha \ll m_B$ , we could make  $W_B$  into a Hamilton cycle of  $G_B^c$  by ‘winding around’ each cycle from  $\mathcal{C}_B$  a suitable number of times. We could then apply the Blow-up lemma to show that this Hamilton cycle corresponds to one in  $G_B^*$ . However, as indicated in Section 3.2, we will argue slightly differently as it is not clear how to incorporate all the exceptional vertices by the above approach.

Set  $\varepsilon_A := \varepsilon/|R_B|$ . Apply the Diregularity lemma with parameters  $\varepsilon_A^2$ ,  $d + 8\varepsilon_A^2$  and  $M'_A$  to  $G[A \cup V_0]$  to obtain a partition of the vertex set of  $G[A \cup V_0]$  into  $\ell \geq M'_A$  clusters  $V'_1, \dots, V'_\ell$  and an exceptional set  $V'_0$ . Let  $A' := V'_1 \cup \dots \cup V'_\ell$ , let  $R_A$  denote the reduced oriented graph obtained from Lemma 8 and let  $G_A^*$  be the pure oriented graph. Similarly as in (13), Lemma 8 implies that  $\delta^*(R_A) \geq (3/2 + \alpha/2)|R_A|$  and so, as before, we can apply Theorem 11 to find a 1-factor  $\mathcal{C}_A$  of  $R_A$ . Then as before, Proposition 10 implies that by adding at most  $4\varepsilon_A^2 n$  further vertices to the exceptional set  $V'_0$  we may assume that each edge of  $R_A$  corresponds to an  $\varepsilon_A$ -regular pair of density  $d$  and that each edge in the

FIGURE 2. Incorporating the exceptional vertex  $v$ .

union  $\bigcup_{C \in \mathcal{C}_A} C \subseteq R_A$  of all the cycles from  $\mathcal{C}_A$  corresponds to an  $(\varepsilon_A, d)$ -superregular pair. So we now have that

$$(14) \quad |V'_0| \leq \varepsilon_A n = \varepsilon n / |R_B|.$$

Similarly as before, Corollary 15 gives us a closed walk  $W_A$  in  $R_A$  which is balanced w.r.t.  $\mathcal{C}_A$ , meets every cluster at most  $2|R_A|/\alpha$  times and traverses all the edges lying on a cycle from  $\mathcal{C}_A$  at least once.

**5.2. Incorporating  $V'_0$  into the walk  $W_B$ .** Recall that the balanced closed walk  $W_B$  in  $R_B$  corresponds to a Hamilton cycle in  $G_B^c$ . Our next aim is to extend this walk to one which corresponds to a Hamilton cycle which also contains the vertices in  $V'_0$ . (The Blow-up lemma will imply that the latter Hamilton cycle corresponds to one in  $G[B \cup V'_0]$ .) We do this by extending  $W_B$  into a walk on a suitably defined digraph  $R_B^* \supseteq R_B$  with vertex set  $V(R_B) \cup V'_0$  in such a way that the new walk is balanced w.r.t.  $\mathcal{C}_B$ .  $R_B^*$  is obtained from the union of  $R_B$  and the set  $V'_0$  by adding an edge  $vV_i$  between a vertex  $v \in V'_0$  and a cluster  $V_i \in V(R_B)$  whenever  $|N_G^+(v) \cap V_i| > \alpha m_B/10$  and adding the edge  $V_i v$  whenever  $|N_G^-(v) \cap V_i| > \alpha m_B/10$ . Thus

$$|N_{R_B^*}^+(v) \cap B| \leq |N_{R_B^*}^+(v)|m_B + |R_B|\alpha m_B/10.$$

Hence

$$(15) \quad \begin{aligned} |N_{R_B^*}^+(v)| &\geq |N_G^+(v) \cap B|/m_B - \alpha|R_B|/10 \geq |N_G^+(v) \cap B||R_B|/|B| - \alpha|R_B|/10 \\ &\geq (|N_{G-A}^+(v)| - |V_0|)|R_B|/|G-A| - \alpha|R_B|/10 \\ &\geq (\delta^+(G)/n - \alpha/2)|R_B| \geq \alpha|R_B|/2. \end{aligned}$$

(The penultimate inequality follows from the choice of  $A$  and the final one from Fact 13.) Similarly

$$|N_{R_B^*}^-(v)| \geq \alpha|R_B|/2.$$

Given a vertex  $v \in V'_0$  pick  $U_1 \in N_{R_B^*}^+(v)$ ,  $U_2 \in N_{R_B^*}^-(v) \setminus \{U_1\}$ . Let  $C_1$  and  $C_2$  denote the cycles from  $\mathcal{C}_B$  containing  $U_1$  and  $U_2$  respectively. Let  $U_1^-$  be the predecessor of  $U_1$  on  $C_1$ , and  $U_2^+$  be the successor of  $U_2$  on  $C_2$ . (15) implies that we can ensure  $U_1^- \neq U_2^+$ . (However, we may have  $C_1 = C_2$ .) Corollary 14 gives us a shifted walk  $W_v$  from  $U_1^-$  to  $U_2^+$  traversing at most  $4/\alpha$  cycles of  $\mathcal{C}_B$ . To incorporate  $v$  into the walk  $W_B$ , recall that  $W_B$  traverses all those edges of  $R_B$  which lie on cycles from  $\mathcal{C}_B$  at least once. Replace one of the occurrences of  $U_1^- U_1$  on  $W_B$  with the walk

$$W'_v := U_1^- W_v U_2^+ C_2 U_2 v U_1 C_1 U_1,$$

i.e. the walk that goes from  $U_1^-$  to  $U_2^+$  along the shifted walk  $W_v$ , it then winds once around  $C_2$  but stops in  $U_2$ , then it goes to  $v$  and further to  $U_1$ , and finally it winds around  $C_1$ . The walk obtained from  $W_B$  by including  $v$  in this way is still balanced w.r.t.  $\mathcal{C}_B$ , i.e. each vertex in  $R_B$  is visited the same number of times as every other vertex lying on the same cycle from  $\mathcal{C}_B$ . We add the extra loop around  $C_1$  because when applying the Blow-up lemma we will need the vertices in  $V'_0$  to be at a distance of at least 4 from each other. Using this loop, this can be ensured as follows. After we have incorporated  $v$  into  $W_B$  we ‘ban’ all the 6 edges of (the new walk)  $W_B$  whose endvertices both have distance at most 3 from  $v$ . The extra loop ensures that every edge in each cycle from  $\mathcal{C}$  has at least one occurrence in  $W_B$  which is not banned. (Note that we do not have to add an extra loop around  $C_2$  since if  $C_2 \neq C_1$  then all the banned edges of  $C_2$  lie on  $W_v$  but each edge of  $C_2$  also occurs on the original walk  $W_B$ .) Thus when incorporating the next exceptional vertex we can always pick an occurrence of an edge which is not banned to be replaced by a longer walk. (When incorporating  $v$  we picked  $U_1^- U_1$ .) Repeating this argument, we can incorporate all the exceptional vertices in  $V'_0$  into  $W_B$  in such a way that all the vertices of  $V'_0$  have distance at least 4 on the new walk  $W_B$ .

Recall that  $G_B^c$  denotes the oriented graph obtained from the pure oriented graph  $G_B^*$  by adding all the  $V_i-V_j$  edges for all those pairs  $V_i, V_j$  of clusters with  $V_i V_j \in E(R_B)$ . Let  $G_{B \cup V'_0}^c$  denote the graph obtained from  $G_B^c$  by adding all the  $V'_0-B$  edges of  $G$  as well as all the  $B-V'_0$  edges of  $G$ . Moreover, recall that the vertices in  $V'_0$  have distance at least 4 from each other on  $W_B$  and  $|V'_0| \leq \varepsilon n / |R_B| \ll \alpha m_B / 20$  by (14). As already observed at the beginning of Section 4, altogether this shows that by winding around each cycle from  $\mathcal{C}_B$ , one can obtain a Hamilton cycle  $C_{B \cup V'_0}^c$  of  $G_{B \cup V'_0}^c$  from the walk  $W_B$ , provided that  $W_B$  visits any cluster  $V_i \in R_B$  at most  $m_B$  times. To see that the latter condition holds, recall that before we incorporated the exceptional vertices in  $V'_0$  into  $W_B$ , each cluster was visited at most  $2|R_B|/\alpha$  times. When incorporating an exceptional vertex we replaced an edge of  $W_B$  by a walk whose interior visits every cluster at most  $4/\alpha + 2 \leq 5/\alpha$  times. Thus the final walk  $W_B$  visits each cluster  $V_i \in R_B$  at most

$$(16) \quad 2|R_B|/\alpha + 5|V'_0|/\alpha \stackrel{(14)}{\leq} 6\varepsilon n / (\alpha|R_B|) \leq \sqrt{\varepsilon} m_B$$

times. Hence we have the desired Hamilton cycle  $C_{B \cup V'_0}^c$  of  $G_{B \cup V'_0}^c$ . Note that (16) implies that we can choose  $C_{B \cup V'_0}^c$  in such a way that for each cycle  $C \in \mathcal{C}_B$  there is subpath  $P_C$  of  $C_{B \cup V'_0}^c$  which winds around  $C$  at least

$$(17) \quad (1 - \sqrt{\varepsilon}) m_B$$

times in succession.

**5.3. Applying the Blow-up lemma to find a Hamilton cycle in  $G[B \cup V'_0]$ .** Our next aim is to use the Blow-up lemma to show that  $C_{B \cup V'_0}^c$  corresponds to a Hamilton cycle in  $G[B \cup V'_0]$ . Recall that  $k = |R_B|$  and that for each exceptional vertex  $v \in V'_0$  the outneighbour  $U_1$  of  $v$  on  $W_B$  is distinct from its inneighbour  $U_2$  on  $W_B$ . We will apply the Blow-up lemma with  $H$  being the underlying graph of  $C_{B \cup V'_0}^c$  and  $G^*$  being the graph obtained from the underlying graph of  $G_B^*$  by adding all the vertices  $v \in V'_0$  and joining each such  $v$  to all the vertices in  $N_G^+(v) \cap U_1$  as well as to all the vertices in  $N_G^-(v) \cap U_2$ . Recall that after applying the Diregularity lemma to obtain the clusters  $V_1, \dots, V_k$  we used Proposition 10 to ensure that each edge of  $R_B$  corresponds to an  $\varepsilon$ -regular pair of density  $d$  (in the underlying graph of  $G_B^*$  and thus also in  $G^*$ ) and that each edge of the union  $\bigcup_{C \in \mathcal{C}_B} C \subseteq R_B$  of all the cycles from  $\mathcal{C}_B$  corresponds to an  $(\varepsilon, d)$ -superregular pair.

$V'_0$  will play the role of  $V_0$  in the Blow-up lemma and we take  $L_0, L_1, \dots, L_k$  to be the partition of  $H$  induced by  $V'_0, V_1, \dots, V_k$ .  $\phi : L_0 \rightarrow V'_0$  will be the obvious bijection (i.e. the identity). To define the set  $I \subseteq V(H)$  of vertices of distance at least 4 from each other which is used in the Blow-up lemma, let  $P'_C$  be the subpath of  $H$  corresponding to  $P_C$  (for all  $C \in \mathcal{C}_B$ ). For each  $i = 1, \dots, k$ , let  $C_i \in \mathcal{C}_B$  denote the cycle containing  $V_i$  and let  $J_i \subseteq L_i$  consist of all those vertices in  $L_i \cap V(P'_{C_i})$  which have distance at least 4 from the endvertices of  $P'_{C_i}$ . Thus in the graph  $H$  each vertex  $u \in J_i$  has one of its neighbours in the set  $L_i^-$  corresponding to the predecessor of  $V_i$  on  $C_i$  and its other neighbour in the set  $L_i^+$  corresponding to the successor of  $V_i$  on  $C_i$ . Moreover, all the vertices in  $J_i$  have distance at least 4 from all the vertices in  $L_0$  and (17) implies that  $|J_i| \geq 9m_B/10$ . It is easy to see that one can greedily choose a set  $I_i \subseteq J_i$  of size  $m_B/10$  such that the vertices in  $\bigcup_{i=1}^k I_i$  have distance at least 4 from each other. We take  $I := L_0 \cup \bigcup_{i=1}^k I_i$ .

Let us now check conditions (C1)–(C9). (C1) holds with  $K_1 := 1$  since  $|L_0| = |V'_0| \leq \varepsilon_A n = \varepsilon n/k \leq d|H|$ . (C2) holds by definition of  $I$ . (C3) holds since  $H$  is a Hamilton cycle in  $G_{B \cup V'_0}^c$  (c.f. the definition of the graph  $G_{B \cup V'_0}^c$ ). This also implies that for every edge  $xy \in H$  with  $x \in L_i, y \in L_j$  ( $i, j \geq 1$ ) we must have that  $V_i V_j \in E(R_B)$ . Thus (C6) holds as every edge of  $R_B$  corresponds to an  $\varepsilon$ -regular pair of clusters having density  $d$ . (C4) holds with  $K_2 := 1$  because

$$|N_H(L_0) \cap L_i| \leq 2|L_0| = 2|V'_0| \stackrel{(14)}{\leq} 2\varepsilon n/|R_B| \leq 5\varepsilon m_B \leq dm_B.$$

For (C5) we need to find a set  $D \subseteq I$  of buffer vertices. Pick any set  $D_i \subseteq I_i$  with  $|D_i| = \delta' m_B$  and let  $D := \bigcup_{i=1}^k D_i$ . Since  $I_i \subseteq J_i$  we have that  $|N_H(D) \cap L_j| = 2\delta' m_B$  for all  $j = 1, \dots, k$ . Hence

$$||N_H(D) \cap L_i| - |N_H(D) \cap L_j|| = 0$$

for all  $1 \leq i < j \leq k$  and so (C5) holds. (C7) holds with  $c := \alpha/10$  by our choice  $U_1 \in N_{R_B^*}^+(v)$  and  $U_2 \in N_{R_B^*}^-(v)$  of the neighbours of each vertex  $v \in V'_0$  in the walk  $W_B$  (c.f. the definition of the graph  $R_B^*$ ).

(C8) and (C9) are now the only conditions we need to check. Given a set  $E_i \subseteq V_i$  of size at most  $\varepsilon' m_B$ , we wish to find  $F_i \subseteq (L_i \cap (I \setminus D)) = I_i \setminus D$  and a bijection  $\phi_i : E_i \rightarrow F_i$  such that every  $v \in E_i$  has a large number of neighbours in every cluster  $V_j$  for which  $L_j$  contains a neighbour of  $\phi_i(v)$ . Pick any set  $F_i \subseteq I_i \setminus D$  of size  $|E_i|$ . (This can be done since  $|D \cap I_i| = \delta' m_B$  and so  $|I_i \setminus D| \geq m_B/10 - \delta' m_B \gg \varepsilon' m_B$ .) Let  $\phi_i : E_i \rightarrow F_i$  be an arbitrary bijection. To see that (C8) holds with these choices, consider any vertex  $v \in E_i \subseteq V_i$  and let  $j$  be such that  $L_j$  contains a neighbour of  $\phi_i(v)$  in  $H$ . Since  $\phi_i(v) \in F_i \subseteq I_i \subseteq J_i$ , this means that  $V_j$  must be a neighbour of  $V_i$  on the cycle  $C_i \in \mathcal{C}_B$  containing  $V_i$ . But this implies that  $|N_{G^*}(v) \cap V_j| \geq (d - \varepsilon)m_B$  since each edge of the union  $\bigcup_{C \in \mathcal{C}_B} C \subseteq R_B$  of all the cycles from  $\mathcal{C}_B$  corresponds to an  $(\varepsilon, d)$ -superregular pair in  $G^*$ .

Finally, writing  $F := \bigcup_{i=1}^k F_i$  we have

$$|N_H(F) \cap L_i| \leq 2\varepsilon' m_B$$

(since  $F_j \subseteq J_j$  for each  $j = 1, \dots, k$ ) and so (C9) is satisfied with  $K_3 := 2$ . Hence (C1)–(C9) hold and so we can apply the Blow-up lemma to obtain a Hamilton cycle in  $G^*$  such that the image of  $L_i$  is  $V_i$  for all  $i = 1, \dots, k$  and the image of each  $x \in L_0$  is  $\phi(x) \in V'_0$ . (Recall that  $G^*$  was obtained from the underlying graph of  $G_B^*$  by adding all the vertices  $v \in V'_0$  and joining each such  $v$  to all the vertices in  $N_G^+(v) \cap U_1$  as well as to all the vertices in  $N_G^-(v) \cap U_2$ , where  $U_1$  and  $U_2$  are the neighbours of  $v$  on the walk  $W_B$ .) Using the fact that  $H$  was obtained from the (directed) Hamilton cycle  $C_{B \cup V'_0}^c$  and since  $U_1 \neq U_2$  for each



$v \in V'_0$ , it is easy to see that our Hamilton cycle in  $G^*$  corresponds to a (directed) Hamilton cycle  $C_B$  in  $G[B \cup V'_0]$ .

**5.4. Finding a Hamilton cycle in  $G$ .** The last step of the proof is to find a Hamilton cycle in  $G[A']$  which can be connected with  $C_B$  into a Hamilton cycle of  $G$ . Pick an arbitrary edge  $v_1v_2$  on  $C_B$  and add an extra vertex  $v^*$  to  $G[A']$  with outneighbourhood  $N_G^+(v_1) \cap A'$  and inneighbourhood  $N_G^-(v_2) \cap A'$ . A Hamilton cycle  $C_A$  in the digraph thus obtained from  $G[A']$  can be extended to a Hamilton cycle of  $G$  by replacing  $v^*$  with  $v_2C_Bv_1$ . To find such a Hamilton cycle  $C_A$ , we can argue as before. This time, there is only one exceptional vertex, namely  $v^*$ , which we incorporate into the walk  $W_A$ . Note that by our choice of  $A$  and  $B$  the analogue of (15) is satisfied and so this can be done as before. We then use the Blow-up lemma to obtain the desired Hamilton cycle  $C_A$  corresponding to this walk.

## 6. PROOF OF THEOREM 4

The following observation guarantees that every oriented graph as in Theorem 4 has large minimum semidegree.

**Fact 16.** *Suppose that  $0 < \alpha < 1$  and that  $G$  is an oriented graph such that  $d^+(x) + d^-(y) \geq (3/4 + \alpha)|G|$  whenever  $xy \notin E(G)$ . Then  $\delta^0(G) \geq |G|/8 + \alpha|G|/2$ .*

**Proof.** Suppose not. We may assume that  $\delta^+(G) \leq \delta^-(G)$ . Pick a vertex  $x$  with  $d^+(x) = \delta^+(G)$ . Let  $Y$  be the set of all those vertices  $y$  with  $xy \notin E(G)$ . Thus  $|Y| \geq 7|G|/8 - \alpha|G|/2$ . Moreover,  $d^-(y) \geq (3/4 + \alpha)|G| - d^+(x) \geq 5|G|/8 + \alpha|G|/2$ . Hence  $e(G) \geq |Y|(5|G|/8 + \alpha|G|/2) > 35|G|^2/64$ , a contradiction.  $\square$

The proof of Theorem 4 is similar to that of Theorem 3. Fact 16 and Lemma 8 together imply that the reduced oriented graph  $R_A$  (and similarly  $R_B$ ) has minimum semidegree at least  $|R|/8$  and it inherits the Ore-type condition from  $G$  (i.e. it satisfies condition (d) of Lemma 8 with  $c = 3/4 + \alpha$ ). Together with Lemma 17 below (which is an analogue of Lemma 12) this implies that  $R_A$  (and  $R_B$  as well) is an expander in the sense that  $|N^+(X)| \geq |X| + \alpha|R_A|/2$  for all  $X \subseteq V(R_A)$  with  $|X| \leq (1 - \alpha)|R_A|$ . In particular,  $R_A$  (and similarly  $R_B$ ) has a 1-factor: To see this, note that the above expansion property together with Fact 16 imply that for any  $X \subseteq V(R_A)$ , we have  $|N_{R_A}^+(X)| \geq |X|$ . Together with Hall's theorem, this means that the following bipartite graph  $H$  has a perfect matching: the vertex classes  $W_1, W_2$  are 2 copies of  $V(R_A)$  and we have an edge in  $H$  between  $w_1 \in W_1$  and  $w_2 \in W_2$  if there is an edge from  $w_1$  to  $w_2$  in  $R_A$ . But clearly a perfect matching in  $H$  corresponds to a 1-factor in  $R_A$ . Using these facts, one can now argue precisely as in the proof of Theorem 3.

**Lemma 17.** *Suppose that  $0 < \varepsilon \ll \alpha \ll 1$ . Let  $R^*$  be an oriented graph on  $N$  vertices and let  $U$  be a set of at most  $\varepsilon N^2$  ordered pairs of vertices of  $R^*$ . Suppose that  $d^+(x) + d^-(y) \geq (3/4 + \alpha)N$  for all  $xy \notin E(R^*) \cup U$ . Then any  $X \subseteq V(R^*)$  with  $\alpha N \leq |X| \leq (1 - \alpha)N$  satisfies  $|N^+(X)| \geq |X| + \alpha N/2$ .*

**Proof.** The proof is similar to that of Lemma 12. Suppose that Lemma 17 does not hold and let  $X \subseteq V(R^*)$  with  $\alpha N \leq |X| \leq (1 - \alpha)N$  be such that

$$(18) \quad |N^+(X)| < |X| + \alpha N/2.$$

Call a vertex of  $R^*$  *good* if it lies in at most  $\sqrt{\varepsilon}N$  pairs from  $U$ . Thus all but at most  $2\sqrt{\varepsilon}N$  vertices of  $R^*$  are good. As in the proof of Lemma 12 we consider the following partition of  $V(R^*)$ :

$$A := X \cap N^+(X), \quad B := N^+(X) \setminus X, \quad C := V(R^*) \setminus (X \cup N^+(X)), \quad D := X \setminus N^+(X).$$

(18) implies

$$(19) \quad |D| + \alpha N/2 > |B|.$$

Suppose first that  $|D| > 2\sqrt{\varepsilon}N$ . It is easy to see that there are vertices  $x \neq y$  in  $D$  such that  $xy, yx \notin U$ . Since no edge of  $R^*$  lies within  $D$  we have  $xy, yx \notin E(R^*)$  and so  $d(x) + d(y) \geq 3N/2 + 2\alpha N$ . In particular, at least one of  $x, y$  has degree at least  $3N/4 + \alpha N$ . But then

$$(20) \quad |A| + |B| + |C| \geq 3N/4 + \alpha N.$$

If  $|D| \leq 2\sqrt{\varepsilon}N$  then  $|A| + |B| + |C| \geq N - |D|$  and so (20) still holds with room to spare. Note that (19) and (20) together imply that  $2|A| + 2|C| \geq 3N/2 + 2\alpha N - 2|B| \geq 3N/2 - |B| - |D| \geq N/2$ . Thus at least one of  $A, C$  must have size at least  $N/8$ . In particular, this implies that one of the following 3 cases holds.

**Case 1.**  $|A|, |C| > 2\sqrt{\varepsilon}N$ .

Let  $A'$  be the set of all good vertices in  $A$ . By an averaging argument there exists  $x \in A'$  with  $|N^+(x) \cap A'| < |A'|/2$ . Since  $N^+(A) \subseteq A \cup B$  this implies that  $|N^+(x)| < |B| + |A \setminus A'| + |A'|/2$ . Let  $C' \subseteq C$  be the set of all those vertices  $y \in C$  with  $xy \notin U$ . Thus  $|C \setminus C'| \leq \sqrt{\varepsilon}N$  since  $x$  is good. By an averaging argument there exists  $y \in C'$  with  $|N^-(y) \cap C'| < |C'|/2$ . But  $N^-(C) \subseteq B \cup C$  and so  $|N^-(y)| < |B| + |C \setminus C'| + |C'|/2$ . Moreover,  $d^+(x) + d^-(y) \geq 3N/4 + \alpha N$  since  $xy \notin E(R^*) \cup U$ . Altogether this shows that

$$|A'|/2 + |C'|/2 + 2|B| \geq d^+(x) + d^-(y) - |A \setminus A'| - |C \setminus C'| \geq 3N/4 + \alpha N/2.$$

Together with (20) this implies that  $3|A| + 6|B| + 3|C| \geq 3N + 3\alpha N$ , which in turn together with (19) yields  $3|A| + 3|B| + 3|C| + 3|D| \geq 3N + 3\alpha N/2$ , a contradiction.

**Case 2.**  $|A| > 2\sqrt{\varepsilon}N$  and  $|C| \leq 2\sqrt{\varepsilon}N$ .

As in Case 1 we let  $A'$  be the set of all good vertices in  $A$  and pick  $x \in A'$  with  $|N^+(x)| < |B| + |A \setminus A'| + |A'|/2$ . Note that (19) implies that  $|D| > N - |X| - |C| - \alpha N/2 \geq \sqrt{\varepsilon}N$ . Pick any  $y \in D$  such that  $xy \notin U$ . Then  $xy \notin E(R^*)$  since  $R^*$  contains no edges from  $A$  to  $D$ . Thus  $d^+(x) + d^-(y) \geq 3N/4 + \alpha N$ . Moreover,  $N^-(y) \subseteq B \cup C$ . Altogether this gives

$$|A'|/2 + 2|B| \geq d^+(x) + d^-(y) - |A \setminus A'| - |C| \geq 3N/4 + \alpha N/2.$$

As in Case 1 one can combine this with (20) and (19) to get a contradiction.

**Case 3.**  $|A| \leq 2\sqrt{\varepsilon}N$  and  $|C| > 2\sqrt{\varepsilon}N$ .

This time we let  $C'$  be the set of all good vertices in  $C$  and pick  $y \in C'$  with  $|N^-(y) \cap C'| < |C'|/2$ . Hence  $|N^-(y)| < |B| + |C \setminus C'| + |C'|/2$ . Moreover, we must have  $|D| = |X| - |A| > \sqrt{\varepsilon}N$ . Pick any  $x \in D$  such that  $xy \notin U$ . Then  $xy \notin E(R^*)$  since  $R^*$  contains no edges from  $D$  to  $C$ . Thus  $d^+(x) + d^-(y) \geq 3N/4 + \alpha N$ . Moreover,  $N^+(x) \subseteq A \cup B$ . Altogether this gives

$$|C'|/2 + 2|B| \geq d^+(x) + d^-(y) - |A| - |C \setminus C'| \geq 3N/4 + \alpha N/2,$$

which in turn yields a contradiction as before.  $\square$

## 7. ACKNOWLEDGEMENT

We are grateful to Peter Keevash for pointing out an argument for the ‘shifted expansion property’ which is simpler than the one presented in an earlier version of this manuscript.

## REFERENCES

- [1] N. Alon and A. Shapira, Testing subgraphs in directed graphs, *Journal of Computer and System Sciences* **69** (2004), 354–382.
- [2] N. Alon and J. Spencer, *The Probabilistic Method* (2nd edition), Wiley-Interscience 2000.
- [3] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications*, Springer 2000.
- [4] B. Bollobás, *Random Graphs* (2nd edition), Cambridge University Press 2001.
- [5] B. Bollobás and R. Häggkvist, Powers of Hamilton cycles in tournaments, *J. Combin. Theory B* **50** (1990), 309–318.
- [6] B. Csaba, On the Bollobás–Eldridge conjecture for bipartite graphs, *Combin. Probab. Comput.* **16** (2007), 661–691.
- [7] B. Csaba, A. Shokoufandeh, and E. Szemerédi, Proof of a conjecture of Bollobás and Eldridge for graphs of maximum degree three, *Combinatorica* **23** (2003), 35–72.
- [8] A. Frieze and M. Krivelevich, On packing Hamilton cycles in  $\varepsilon$ -regular graphs, *J. Combin. Theory B* **94** (2005), 159–172.
- [9] A. Ghouila-Houri, Une condition suffisante d’existence d’un circuit hamiltonien, *C.R. Acad. Sci. Paris* **25** (1960), 495–497.
- [10] R. Häggkvist, Hamilton cycles in oriented graphs, *Combin. Probab. Comput.* **2** (1993), 25–32.
- [11] R. Häggkvist and A. Thomason, Oriented Hamilton cycles in oriented graphs, in *Combinatorics, Geometry and Probability*, Cambridge University Press 1997, 339–353.
- [12] P. Keevash, D. Kühn and D. Osthus, An exact minimum degree condition for Hamilton cycles in oriented graphs, submitted.
- [13] L. Kelly, D. Kühn and D. Osthus, Cycles of given length in oriented graphs, submitted.
- [14] J. Komlós, The Blow-up lemma, *Combin. Probab. Comput.* **8** (1999), 161–176.
- [15] J. Komlós, G. N. Sárközy, and E. Szemerédi, Blow-up lemma, *Combinatorica* **17** (1997), 109–123.
- [16] J. Komlós and M. Simonovits, Szemerédi’s Regularity Lemma and its applications in graph theory, *Bolyai Society Mathematical Studies 2, Combinatorics, Paul Erdős is Eighty (Vol. 2)* (D. Miklós, V. T. Sós and T. Szőnyi eds.), Budapest (1996), 295–352.
- [17] D. Kühn, D. Osthus and A. Treglown, Hamiltonian degree sequences in digraphs, submitted.
- [18] B. McKay, The asymptotic numbers of regular tournaments, Eulerian digraphs and Eulerian oriented graphs, *Combinatorica* **10** (1990), 367–377.
- [19] O. Ore, Note on Hamilton circuits, *American Math. Monthly* **67** (1960), 55.
- [20] C. Thomassen, Long cycles in digraphs with constraints on the degrees, in *Surveys in Combinatorics* (B. Bollobás ed.), *London Math. Soc. Lecture Notes* **38** (1979), 211–228, Cambridge University Press.
- [21] C. Thomassen, Long cycles in digraphs. *Proc. London Math. Soc.* **42** (1981), 231–251.
- [22] C. Thomassen, Edge-disjoint Hamiltonian paths and cycles in tournaments, *Proc. London Math. Soc.* **45** (1982), 151–168.
- [23] D. Woodall, Sufficient conditions for cycles in digraphs, *Proc. London Math. Soc.* **24** (1972), 739–755.
- [24] A. Young, Extremal problems for dense graphs and digraphs, Master’s thesis, School of Mathematics, University of Birmingham 2005.

Luke Kelly, Daniela Kühn & Deryk Osthus

School of Mathematics

University of Birmingham

Edgbaston

Birmingham

B15 2TT

UK

*E-mail addresses:* {kellyl,kuehn,osthus}@maths.bham.ac.uk