Arbitrary Orientations of Hamilton Cycles in Oriented Graphs

Luke Kelly

ltkelly@gmail.com

Abstract

We use a randomised embedding method to prove that for all $\alpha > 0$ any sufficiently large oriented graph G with minimum in-degree and out-degree $\delta^+(G), \delta^-(G) \geq (3/8 + \alpha)|G|$ contains every possible orientation of a Hamilton cycle. This confirms a conjecture of Häggkvist and Thomason.

1 Introduction

An *oriented graph* is a loop-free simple graph where each edge is given an orientation. A *directed graph (digraph)* is an oriented graph where we allow one edge in each direction between each pair of vertices, that is, we allow cycles of length 2. The *minimum semi-degree* $\delta^0(G)$ of an oriented graph G (or a digraph) is the minimum of its minimum outdegree $\delta^+(G)$ and its minimum indegree $\delta^-(G)$.

A fundamental result of Dirac states that a minimum degree of $|G|/2$ guarantees a Hamilton cycle in any undirected graph G on at least 3 vertices. Following this result several weaker conditions guaranteeing a Hamilton cycle have been found. One of the famous of these is Ore's theorem, which states that if $d(x) + d(y) \geq |G| \geq 3$ for all $x \neq y \in V(G)$ with $xy \notin E(G)$ then G contains a Hamilton cycle. In some sense the weakest possible condition of this type is Chvátal's theorem.[1] This gives a condition on the (ordered) degree sequence of a graph which forces a Hamilton cycle, such that for any (graphic) degree sequence not satisfying Chvátal's conditions there exists a graph with a degree sequence dominated by that sequence not containing a Hamilton cycle.

There is an analogue of Dirac's theorem for digraphs due to Ghouila-Houri [5] which states that every digraph D with minimum semi-degree at least $|D|/2$ contains a directed Hamilton cycle. As with Dirac's theorem, taking two disjoint cliques of as equal size as possible demonstrates that this minimum degree condition can not be improved.

---

[1]Whilst it is widely regarded as such, it should be noted that Chvátal's theorem does not quite imply Ore's theorem.

Thomassen [23] asked the natural question of whether there exists an analogous result for oriented graphs, where one expects to be able to obtain a weaker degree condition than the bounds needed for digraphs. Häggkvist [7] constructed an example in 1993 showing that a minimum semi-degree of $(3n - 4)/8$ was necessary and conjectured that this was also sufficient. With Thomason [9] he showed in 1997 that for any $\alpha > 0$ every sufficiently large oriented graph G with minimum semi-degree at least $(5/12 + \alpha)|G|$ has a directed Hamilton cycle. The author, together with Kühn and Osthus [14], finally confirmed in 2008 that, up to a linear error term, $3|G|/8$ is indeed the correct bound. Following this Keevash, Kühn and Osthus improved this to an exact result.

Theorem 1 (Keevash, Kühn and Osthus [11]). *There exists n_0 such that every oriented graph G on $n \geq n_0$ vertices with $\delta^0(G) \geq (3n - 4)/8$ contains a directed Hamilton cycle.*

Christofides, Keevash, Kühn and Osthus [3] have also since found an efficient algorithmic proof of (a generalisation of) this result.

Nash-Williams [20] conjectured a digraph analogue of Chvátal's theorem. This has recently been approximately confirmed by Kühn, Osthus and Treglown [18]. There also now exists a semi-exact degree condition result due to Christofides, Keevash, Kühn and Osthus [4].

It is natural to ask whether these bounds only give us directed Hamilton cycles or whether they give every possible orientation of a Hamilton cycle. Indeed this question was answered for digraphs, asymptotically at least, by Häggkvist and Thomason in 1995.

Theorem 2 (Häggkvist and Thomason [8]). *There exists n_0 such that every digraph D on $n \geq n_0$ vertices with minimum semi-degree $\delta^0(D) \geq n/2 + n^{5/6}$ contains every orientation of a Hamilton cycle.*

The question was asked originally for oriented graphs by Häggkvist and Thomason [9] who proved that for all $\alpha > 0$ and all sufficiently large oriented graphs G a minimum semi-degree of $(5/12 + \alpha)|G|$ suffices to give *any* orientation of a Hamilton cycle. They conjectured that $(3/8 + \alpha)|G|$ suffices, the same bound as for the directed Hamilton cycle up to the error term $\alpha|G|$. Whilst not asked explicitly before Häggkvist and Thomason's paper, there is some previous work of Thomason and Grant relevant to this area. Grant [6] proved in 1980 that any digraph D with minimum semi-degree $\delta^0(D) \geq 2|D|/3 + \sqrt{|D| \log |D|}$ contains an anti-directed Hamilton cycle, provided that n is even. (An anti-directed cycle is one in which the edge orientations alternate.) Thomason [22] showed in 1986 that every sufficiently large tournament contains every possible orientation of a Hamilton cycle (except possibly the directed Hamilton cycle if the tournament is not strong). The following theorem confirms the conjecture of Häggkvist and Thomason.

Theorem 3. *For every $\alpha > 0$ there exists an integer $n_0 = n_0(\alpha)$ such that every oriented graph G on $n \geq n_0$ vertices with minimum semi-degree $\delta^0(G) \geq (3/8 + \alpha)n$ contains every orientation of a Hamilton cycle.*
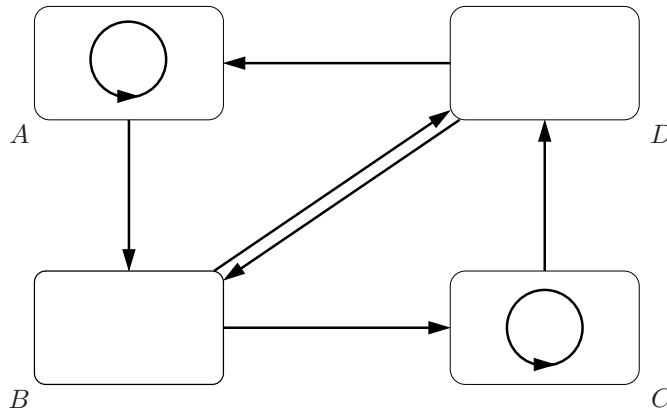
Figure 1: The oriented graph constructed in Proposition 4

1.1 Robust Expansion

The property underlying the proofs of all the recent Hamilton cycle results so far stated is *robust expansion.* This is a notion which was introduced by Kühn, Osthus and Treglown in [18] and has proved to be the correct notion of expansion in a digraph when dealing with this kind of question or when using the Diregularity lemma. Informally speaking, a digraph G is a robust outexpander if all subsets of $V(G)$ have outneighbourhoods larger than themselves unless they are very large or very small and, moreover, this still holds after the removal of a small number of edges.

Having a minimum semi-degree $\delta^0(G) \geq (3/8 + \alpha)|G|$ for some $\alpha > 0$, satisfying an approximate Ore-type condition or satisfying an approximate Chvátal condition imply robust outexpansion (see Lemma 11 in [18]). Hence an extension of Theorem 3 to robust outexpanders would imply an approximate Ore-type result and a Chvátal-type approximate result for arbitrary orientations of Hamilton cycles. The author believes it is likely that the argument given in this paper could be straight-forwardly extended to prove this.

1.2 Extremal Example

Häggkvist [7] constructed an example in 1993 giving a graph on $n = 8k - 1$ vertices with minimum semi-degree $(3n - 5)/8$ containing no Hamilton cycle and Keevash, Kühn and Osthus extended this to all n. This means that Theorem 1 is best possible and that Theorem 3 is best possible up to the linear error term. Interestingly, this example can be improved upon when considering arbitrary orientations. Hence the additive constant in Theorem 1 is not the correct bound when seeking any orientation of a Hamilton cycle, and it is an open question as to what the correct additional term should be.

Proposition 4. *There are infinitely many oriented graphs G with minimum semi-degree exactly $(3|G| - 4)/8$ which do not contain an anti-directed Hamilton cycle.*

Proof. Let $n := 8m + 4$ for some integer $m \in \mathbb{N}$. Let G be the oriented graph obtained from the disjoint union of two regular tournaments A and C on $2m+1$ vertices and sets B and D of $2m + 1$ vertices by adding all edges from A to B, all edges from B to C, all edges from C to D and 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 indegree and the outdegree of every vertex differs by at most 1. So in particular every vertex in B sends at least m edges to D. It is easy to check that the minimum semi-degree of G is $3m + 1 = (3n - 4)/8$, as required.

Let us try to construct an anti-directed Hamilton cycle in G and let us start in B with an edge going forwards. This edge can go either to C or to D. (Starting with an edge oriented backwards produces an identical argument and result.) The next edge must go backwards. It can go from C to either B or C. It can go from D to either B or C. So after two steps we can be in either B or C. Our next edge must go forwards. If we are in B our possible locations after the next two steps are B and C as before. From C we can go forwards either to C or to D. Both options repeat situations we have already met. In no case do we have a means to reach A whilst respecting the orientation of our anti-directed Hamilton cycle. Hence the longest anti-directed cycle in G has length at most $3n/4$ and we have no anti-directed Hamilton cycle as claimed. □

1.3 Pancyclicity

Recently the author, together with Kühn and Osthus, [13] showed that the minimum semi-degree condition in Theorem 1 gives not only a Hamilton cycle but a cycle of every possible length. It is natural to ask whether this can be extended to give all orientations of all cycles of all possible lengths. A simple probabilistic argument implies that Theorem 3 gives arbitrary orientations of any cycle of linear length (i.e. for all $\alpha > 0$, $\eta > 0$ every sufficiently large oriented graph G with $\delta^0(G) \geq (3/8 + \alpha)|G|$ contains every orientation of any cycle of length at least $\eta|G|$). It remains an open question as to whether the error term can be removed. The results on short cycles necessary to prove the exact pancyclicity result in [13] can (with the addition of an error term in the minimum semi-degree condition) also be extended to arbitrary orientations of cycles. In particular, the following theorem can be obtained.

Theorem 5. *Let $\alpha > 0$. Then there exists $n_0 = n_0(\alpha)$ such that if G is an oriented graph on $n \geq n_0$ vertices with minimum semi-degree $\delta^0(G) \geq (3/8 + \alpha)n$ then G contains a cycle of every possible orientation and of every possible length.*

A proof of Kelly, Kühn and Osthus of this result and extensions of the stronger results on some short cycles can be found in [13], along with a number of related open problems.

1.4 Overview of the Proof

The proof of Theorem 3 splits into two parts, both relying on the expansion properties that our minimum semi-degree condition implies. The cases are distinguished by the similarity of the Hamilton cycle C we are trying to embed to the standard orientation of a Hamilton cycle. It turns out that the correct measure, at least for this problem, of whether a cycle is close to a directed cycle is the number of pairs of consecutive edges with differeLnt orientations. Given an oriented graph C we call the subgraph induced by three vertices $x, y, z \in V(C)$ a *neutral pair* if $xy, zy \in E(C)$. Given an arbitrarily oriented cycle C on n vertices let $n(C)$ be the number of neutral pairs in C. Write C_n^* for the standard orientation of a cycle on n vertices. When there is no ambiguity we will merely write C^*.

The essential idea is to split the cycle up into alternating short and long paths and use the probabilistic method to find an approximate embedding of the long paths into a Hamilton cycle of the reduced graph created by applying a Regularity lemma for digraphs. We connect these paths up greedily using the short paths and then adjust the embedding to obtain something which, after the Blow-up lemma has been applied, gives us the desired orientation of a Hamilton cycle in our graph.

The case distinction comes in the manner in which we alter our embedding. In Section 7 we give the argument for cycles far from C^*, where we use the neutral pairs for our adjustments. In Section 8 we assume that we have few neutral pairs, and thus many long sections of C containing no changes in direction, and use these to adjust our embedding.

The number of exceptional vertices that the Diregularity lemma produces when applied directly is too great for the method used here and hence some technical difficulties are introduced. So we control the number of exceptional vertices by randomly splitting our oriented graph G. In still vague, but slightly more precise terms, the Diregularity lemma will for any $\varepsilon > 0$ give us a partition with the property of ε-regularity. It will also give us a set of 'exceptional vertices' which are in some sense badly behaved, but tells us that these make up at most an ε proportion of our vertices. Our method can only cope with $\eta n \ll \varepsilon n$ such vertices. Hence we split the vertices of our given graph G into two sets A and B of roughly equal size (satisfying some 'nice' properties). We apply the Regularity lemma to $G[B]$, giving us at most $\varepsilon|G|$ exceptional vertices V_0. We then apply the Diregularity lemma to $G[A \cup V_0]$ only this time not with parameter ε but with η. This gives us at most $\eta|G|$ exceptional vertices V_0'. We then consider $G_B := G[(B \setminus V_0) \cup V_0']$, which is ε-regular and has no exceptional vertices and $G_A := G - G_B$, which is η-regular and has $0 \ll \eta|G_A|$ exceptional vertices. Hence, at the cost of some technical work and having to stitch everything back together we will be able to control the number of exceptional vertices.

The next section contains much of the notation we use in this paper. In Section 3 we introduce the forms of the Diregularity lemma and Blow-up lemma that we need later. In Section 6 we prepare the oriented graph G and the cycle C for our approximate embedding and in Section 5 prove the main tool needed to do this. Following that in Section 6 we split into our two cases and in Section 7 (C is far from C^*) and Section 8 (C is close

to C^*) we prove Theorem 3.

2 Notation

Given two vertices x and y of a digraph 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 $d^+(x) := |N_G^+(x)|$ for its outdegree. Similarly, we write $N_G^-(x)$ for the inneighbourhood of x and $d^-(x) := |N_G^-(x)|$ for its indegree. Given $X \subseteq V(G)$ we denote $|N_G^+(x) \cap X|$ by $d_X^+(x)$, and define $d_X^-(x)$ similarly. We write $N_G(x) := N_G^+(x) \cup N_G^-(x)$ for the neighbourhood of x. We use $N^+(x)$ etc. whenever this is unambiguous. We write $\Delta(G)$ for the maximum of $|N(x)|$ over all vertices $x \in V(G)$. 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 directed subgraph of G induced by A is denoted by $G[A]$ and we write $e(A)$ for the number of its edges. $G - A$ denotes the digraph obtained from G by deleting A and all edges incident to A.

Given two vertices x, y of a digraph G, an *x-y path* is a path with any orientation which joins x to y. We call a path with the standard orientation a *directed path*. Given two subsets A and B of vertices of G, an *A-B edge* is an edge ab where $a \in A$ and $b \in B$. We write $e(A, B)$ for the number of all these edges. A *walk* in G is a sequence $v_1 v_2 \ldots v_\ell$ of (not necessarily distinct) vertices, where $v_i v_{i+1}$ or $v_{i+1} v_i$ is an edge for all $1 \le i < \ell$. The length of a walk W is $\ell(W) := \ell - 1$. The walk is *closed* if $v_1 = v_\ell$. Given two vertices x, y of G, the *distance $dist(x, y)$ from x to y* is the length of the shortest directed x-y path. The *diameter* of G is the maximum distance between any ordered pair of vertices.

We write $[k]$ for the set $\{1, 2, \ldots, k\}$. We write $0 < a_1 \ll a_2 \ll \ldots \ll a_k$ to mean that we can choose the constants $a_1, a_2, \ldots, a_k$ from right to left. More precisely, there are increasing functions $f_1, f_2, \ldots, f_{k-1}$ such that, given a_k, whenever we choose some $a_i \le f_i(a_{i+1})$, all calculations needed using these constants are valid.

3 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 [15] for a survey on the Blow-up lemma. We start with some more notation. The density of an undirected 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 *ε-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 *(ε, d)-super-regular* if it is ε-regular and furthermore $d_B(a) \ge (d - \varepsilon)|B|$ for all $a \in A$ and $d_A(b) \ge (d - \varepsilon)|A|$ for

all $b \in B$. (This is a slight variation of the standard definition of (ε, d)-super-regularity where one requires $d_B(a) \geq d|B|$ and $d_A(b) \geq d|A|$.)

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 6 (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 $V_0, V_1, \ldots, V_k$ and a spanning subdigraph G' of G such that the following holds:*

- $M' \leq k \leq M$,

- $|V_0| \leq \varepsilon n$,

- $|V_1| = \cdots = |V_k| =: m$,

- $d_{G'}^+(x) > d_G^+(x) - (d + \varepsilon)n$ *for all vertices $x \in G$,*

- $d_{G'}^-(x) > d_G^-(x) - (d + \varepsilon)n$ *for all vertices $x \in G$,*

- *for every ordered pair $V_i V_j$ 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 consists of all the V_i-V_j edges in G' is ε-regular and has density either 0 or at least d,*

- *for all $1 \leq i \leq k$ the digraph $G'[V_i]$ is empty.*

$V_1, \ldots, V_k$ are called *clusters*, V_0 is called the *exceptional set* and the vertices in V_0 are called *exceptional vertices*. Note that in G' all pairs of clusters are ε-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 $V_1, \ldots, V_k$ and the pure digraph G', the *reduced digraph* R' is the digraph whose vertices are $V_1, \ldots, V_k$ 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 ε-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. The following straightforward lemma, taken from a paper of Kelly, Kühn and Osthus [14], 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.

Lemma 7. *For every $\varepsilon \in (0,1)$ there exist integers $M' = M'(\varepsilon)$ and $n_0 = n_0(\varepsilon)$ such that the following holds. Let $d \in [0,1]$, let G be an oriented graph of order at least n_0 and let R' be the reduced digraph and G' the pure digraph obtained by applying the Diregularity lemma to G with parameters ε, 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) $\delta^0(R) \geq (\delta^0(G)/|G| - (6\varepsilon + 4d))|R|$.

The oriented graph R given by Lemma 7 is called the *reduced oriented graph*. The spanning oriented subgraph G^* of the pure digraph G' obtained by deleting all the V_i-V_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_i-V_j edges for all pairs V_i, V_j with $V_iV_j \notin E(S)$.

At various stages in our proof we will need some pairs of clusters to be not just regular but super-regular. The following well-known result tells us that we can indeed do this whilst maintaining the regularity of all other pairs.

Lemma 8. *Let $\varepsilon \ll d, 1/\Delta$ and let R be a reduced oriented graph of G as given by Lemmas 6 and 7. Let S be an oriented subgraph of R of maximum degree Δ. Then we can move exactly $2\Delta\varepsilon|V_i|$ vertices from each cluster into V_0 such that each pair (V_i, V_j) corresponding to an edge of S becomes $(2\varepsilon, d/2)$-super-regular and every pair corresponding to an edge of $R \setminus S$ becomes 2ε-regular with density at least $d - \varepsilon$.*

In our proof of Theorem 3 we will also need a consequence of the Blow-up lemma of Komlós, Sárközy and Szemerédi [17]. Roughly speaking, it says that an r-partite graph formed by r clusters such that all the pairs of these clusters are (ε, d)-super-regular behaves like a complete r-partite graph with respect to containing graphs of bounded maximum degree as subgraphs.

Lemma 9 (Blow-up Lemma, Komlós, Sárközy and Szemerédi [17])**.** *Given a graph F on $[k]$ and positive integers d and Δ there exists a positive real $\varepsilon_0 = \varepsilon_0(d, \Delta, k)$ such that the following holds for all positive numbers $\ell_1, \ldots, \ell_k$ and all $0 < \varepsilon \leq \varepsilon_0$. Let F' be the graph obtained from F by replacing each vertex $i \in F$ with a set V_i of ℓ_i new vertices and joining all vertices in V_i to all vertices in V_j whenever ij is an edge of F. Let G' be a spanning subgraph of F' such that for every edge $ij \in F$ the graph $(V_i, V_j)_{G'}$ is (ε, d)-super-regular. Then G' contains a copy of every subgraph H of F' with maximum degree $\Delta(H) \leq \Delta$. Moreover, this copy of H in G' maps the vertices of H to the same sets V_i as the copy of H in F', i.e. if $h \in V(H)$ is mapped to V_i by the copy of H in F', then it is also mapped to V_i by the copy of H in G'.*

The tool we shall actually use is the following consequence of the Blow-up lemma. The proof of it uses similar ideas to those in recent work of Christofides, Keevash, Kühn and Osthus [3].

Lemma 10. *Suppose that all the following hold:*

- $0 < 1/m \ll \varepsilon \ll d \ll 1$.

- $U_1, \ldots, U_k$ *are pairwise disjoint sets of size* m, *for some* $k \geq 6$, *and* G *is a digraph on* $U_1 \cup \ldots \cup U_k$ *such that each* $(U_i, U_{i+1})_G$ *is* (ε, d)*-super-regular (where by convention we consider* U_{k+1} *to be* U_1*);*

- $A_1, \ldots, A_k$ *are pairwise disjoint sets of vertices with* $(1 - \varepsilon)m \leq |A_i| =: m_i \leq m$ *and* H *is a digraph on* $A_1 \cup \ldots \cup A_k$ *which is a vertex-disjoint union of paths of length at least* 3, *where every edge going out of* A_i *ends in* A_{i+1} *for all* i;

- $S_1 \subseteq U_1, \ldots, S_k \subseteq U_k$ *are sets of size* $|S_i| = m_i$;

- *For each path* P *of* H *we are given vertices* $x_P, y_P \in V(G)$ *such that if the initial vertex* a_P *of* P *belongs to* A_i *then* $x_P \in S_i$ *and if the final vertex* b_P *of* P *belongs to* A_j *then* $y_P \in S_j$, *and the vertices* x_P, y_P *are distinct as* P *ranges over the paths of* H.

Then there is an embedding of H *into* $G_S := G[\bigcup S_i]$ *in which every path* P *of* H *is mapped to a path that starts at* x_P *and ends at* y_P.

The following immediate consequence of the Blow-up lemma is needed in the proof of Lemma 10.

Lemma 11. *For every* $0 < d < 1$ *and* $p \geq 4$ *there exists* $\varepsilon_0 > 0$ *such that the following holds for* $0 < \varepsilon < \varepsilon_0$. *Let* $U_1, \ldots, U_p$ *be pairwise disjoint sets of size* m, *for some* m, *and suppose* G *is a graph on* $U_1 \cup \ldots \cup U_p$ *such that each pair* (U_i, U_{i+1}), $1 \leq i \leq p-1$ *is* (ε, d)*-super-regular. Let* $f : U_1 \to U_p$ *be any bijective map. Then there are* m *vertex-disjoint paths from* U_1 *to* U_p *so that for every* $x \in U_1$ *the path starting at* x *ends at* $f(x) \in U_p$.

We also need the following random partitioning property of super-regular pairs which says that with high probability (i.e. with probability tending to 1 as $m \to \infty$) all new pairs created by a random partition of a super-regular pair are themselves super-regular.

Lemma 12. *Suppose that the following hold.*

- $0 < \varepsilon < \theta < d < 1/2$, $k \geq 2$ *and for* $1 \leq i \leq k$ *we have* $a_i, b_i > \theta$ *with* $\sum_{i=1}^{k} a_i = \sum_{i=1}^{k} b_i = 1$.

- $G = (A, B)$ *is an* (ε, d)*-super-regular pair with* $|A| = |B| = m$ *sufficiently large.*

- $A = A_1 \cup \ldots \cup A_k$ *and* $B = B_1 \cup \ldots \cup B_k$ *are partitions chosen uniformly at radnom with* $|A_i| = a_i m$ *and* $|B_i| = b_i m$ *for* $1 \leq i \leq r$.

Then with high probability (A_i, B_j) *is* $(\theta^{-1}\varepsilon, d/2)$*-super-regular for every* $1 \leq i, j \leq k$. $\square$

With these tools we can now prove Lemma 10.

of Lemma 10. Enumerate the paths of H as $P_1, \ldots, P_p$ and split them up arbitrarily into paths of length 3, 4 or 5 such that P_i becomes $P_{i,1}, \ldots, P_{i,q_i}$. Let $a_{i,j}$ and $b_{i,j}$ be the initial vertex and the final vertex of $P_{i,j}$ respectively. Then $a_{i,j} = b_{i,j-1}$ for $2 \leq j \leq q_i$. Let E_s consist of all $a_{i,j}$ belonging to the cluster A_s and similarly let F_s consist of all $b_{i,j}$ belonging to the cluster A_s. For each $a_{i,j} \in E_s$ pick a distinct vertex $x_{i,j} \in S_s$ and for each $b_{i,j} \in F_s$ pick a distinct vertex $y_{i,j} \in S_s$ such that if $a_{i,j} = b_{i,j-1}$ then $x_{i,j} = y_{i,j-1}$, $x_{i,1} = x_{P_{i,j}}$ and $y_{i,m_i} = y_{P_{i,j}}$. It is sufficient to show that there is an embedding of H in which each path $P_{i,j}$ is mapped to a path in G_S starting at $x_{i,j}$ and ending at $y_{i,j}$.

For a path $P_{i,j}$ encode whether each edge in $P_{i,j}$ goes forwards or backwards. If $P_{i,j}$ has length 3 then, writing `f` for an edge going from some A_ℓ to $A_{\ell+1}$ and `b` for an edge going from A_ℓ to $A_{\ell-1}$, t encodes one of the following $2^3 = 8$ possibilities:

<div align="center">

`fff ffb fbf fbb bff bfb bbf bbb.`

</div>

Similarly there are 2^4 possibilities for paths of length 4 and 2^5 for those of length 5. We divide the paths $P_{i,j}$ into $56k$ subcollections based on the orientations of their edges. It transpires that there are notational advantages in doing this by encoding the destination of each vertex relative to the first. More precisely, we divide the paths into subcollections $\mathcal{P}_{i,t}$ with $1 \leq i \leq k$, $3 \leq \ell \leq 5$ and

$$t : \{0, 1, \ldots, \ell\} \to \{-\ell, -\ell+1, \ldots, \ell\}$$

encoding one of the $2^3 + 2^4 + 2^5 = 56$ possibilities discussed above and the length $\ell = \ell(t)$ of the paths. Note that we always have $t(0) = 0$. For example, a path oriented `ffb` would have $t : (0, 1, 2, 3) \mapsto (0, 1, 2, 1)$. $\mathcal{P}_{i,t}$ contains all paths $P_{i,j}$ of length ℓ starting in A_i with each vertex in $P_{i,j}$ going to the cluster relative to A_i given by t.

Observe that as $|U_i \setminus S_i| \leq \varepsilon m$, every pair (S_i, S_{i+1}) is $(2\varepsilon, d/2)$-super-regular. We first use a greedy algorithm to sequentially embed those collections $\mathcal{P}_{i,t}$ containing at most $d^2 m$ paths. That is, we pick any $|\mathcal{P}_{i,t}|$ vertices in S_i to be the start of these paths, and then construct these paths by selecting any (distinct) neighbours of these vertices in the S_j appropriate for each vertex in each path. Each set S_i is met by at most 11×56 of the collections so at any stage in this process we have used at most $6 \times 11 \times 56 d^2 m$ vertices from any cluster U_i. As we have $d \ll 1$ the restriction of any pair (S_i, S_{i+1}) to the remaining vertices is still $(4\varepsilon, d/4)$-super-regular and so we can indeed do this.

Having embedded all the $\mathcal{P}_{i,t}$ containing few paths, we randomly split the remaining vertices so that for each large $\mathcal{P}_{i,t}$ we have sets $S_{i,t}^0 \subseteq S_{i+t(0)=i}$, $S_{i,t}^1 \subseteq S_{i+t(1)}$, $\ldots$, $S_{i,t}^\ell \subseteq S_{i+t(\ell)}$ each of size $|\mathcal{P}_{i,t}| > d^2 m$. By Lemma 12 for each large collection $\mathcal{P}_{i,t}$ and for all $0 \leq r \leq \ell-1$ the pair $(S_{i,t}^r, S_{i,t}^{r+1})$ if $t(r+1) > t(r)$ or the pair $(S_{i,t}^{r+1}, S_{i,t}^r)$ if $t(r+1) < t(r)$ is $(4d^{-2}\varepsilon, d/8)$-super-regular with high probability. Thus for sufficiently large m we can choose a partition with this property and apply Lemma 11 to embed each large $\mathcal{P}_{i,t}$ within its allocated sets. $\qquad\square$

4 Skewed Traverses and Shifted Walks

In this section we introduce some tools needed to tweak a random embedding of an arbitrarily oriented Hamilton cycle into a directed Hamilton cycle of the reduced oriented graph to make it correspond (in some sense) to the desired orientation of a Hamilton cycle in our original graph.

The following crucial result says that our minimum semi-degree condition implies outexpansion.

Lemma 13 (Kelly, Kühn, Osthus [14]). *Let R be an oriented graph with $\delta^0(R) \geq (3/8 + \alpha)|R|$ for some $\alpha > 0$. If $X \subset V(R)$ with $0 < |X| \leq (1 - \alpha)|R|$ then $|N^+(X)| \geq |X| + \alpha|R|/2$.*

Suppose that F is a Hamilton cycle (with the standard orientation) of the reduced oriented graph R and relabel the vertices of R such that $F = V_1 V_2 \ldots V_M$, where we let $M := |R|$. Create a new digraph R^* from R by adding all the exceptional vertices $v \in V_0$ to $V(R)$ and adding an edge vV_i (where V_i is a cluster containing m vertices) whenever $|N_{V_i}^+(v)| \geq cm$ for some given constant $c > 0$. (Recall that m denotes the size of the clusters.) The edges in R^* of the form $V_i v$ are defined in a similar way. Let G^c be the digraph 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 $V(G - V_0)$.

Let W be an assignment of the vertices of an arbitrarily oriented cycle C on n vertices to the vertices of R^* which respects edges (i.e. is a digraph homomorphism from C to R^*). We denote by $a(i)$ the number of vertices of C assigned to the cluster V_i. Observe that we can think of W either as a (possibly degenerate) embedding into G^c or as a closed walk in R^*. It will be useful to the reader to keep this duality in mind when reading the rest of the proof We say that an assignment W of C to R^* is γ-balanced if $\max_i |a(i) - m| \leq \gamma n$ and balanced if $a(i) = m$ for all $1 \leq i \leq M$. Furthermore, we say that an assignment (γ, μ)-corresponds to C if the following conditions hold.

- W is γ-balanced.

- Each exceptional vertex $v \in V_0$ has exactly one vertex of C assigned to it.

- In every $V_i \in V(R)$ at least $m - \mu n$ of the vertices of C assigned to V_i have both of their neighbours assigned to $V_{i-1} \cup V_{i+1}$.

We say that the assignment μ-corresponds to C if it $(0, \mu)$-corresponds to C.

Once we have found such an assignment we can, with some work, use Lemma 10 to show that it corresponds to a copy of C in G. Our immediate aim then is to find such a closed walk corresponding to C.

Given clusters V and V', a *skewed V-V' traverse* $T(V, V')$ is a collection of edges of the form

$$T(V, V') := V V_{i_1}, V_{i_1-1} V_{i_2}, V_{i_2-1} V_{i_3}, \ldots, V_{i_t-1} V'.$$
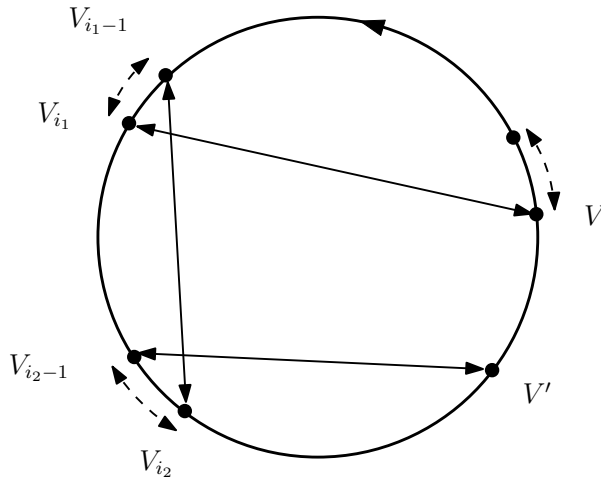
Figure 2: A skewed V-V' traverse

The *length* of a skewed traverse in the number of its edges minus one; so the length of the above skewed traverse is t. Suppose that we have a γ-balanced assignment W of C to R^* and that each vertex of R has many neutral pairs of C assigned to it. We would like to make this a balanced embedding by modifying W. Let V_i, V_j be clusters with $a(i) > m$ and $a(j) < m$. If $V_{i-1}V_j \in E(R)$ then we could replace one neutral pair assigned to $V_{i-1}V_iV_{i-1}$ in the embedding with $V_{i-1}V_jV_{i-1}$. This would reduce $a(i)$ by one and increase $a(j)$ by one. Repeating this process would give the desired balanced embedding. We can not guarantee though that $V_{i-1}V_j \in E(R)$ so we are forced to use skewed traverses to achieve the same effect, which we are able to show always exist under certain conditions. Let

$$V_{i-1}V_{i_1}, V_{i_1-1}V_{i_2}, V_{i_2-1}V_{i_3}, \ldots, V_{i_t-1}V_j.$$

be a skewed V_{i-1}-V_j traverse. Then replacing neutral pairs starting at $V_{i-1}, V_{i_1-1}, \ldots, V_{i_t-1}$ with the edges in the skewed V_{i-1}-V_j traverse we reduce $a(i)$ by one, increase $a(j)$ by one and crucially do not alter $a(k)$ for any $V_k \in V(R) \backslash \{V_i, V_j\}$. See Figure 2 for an illustration of this, where the dashed edges represent the neutral pairs which will be replaced by the solid edges representing the edges of the skewed traverse. We always assume that a skewed traverse has minimal length and thus that each vertex $V_i \in V(R)$ appears at most once as the first vertex of an edge in a skewed traverse.

Given vertices $V, V' \in V(R)$ and a Hamilton cycle F of R, a *shifted V-V' walk $S(V, V')$* is a walk of the form

$$S(V, V') := V \, V_{i_1} F V_{i_1-1} \, V_{i_2} F V_{i_2-1} \ldots V_{i_t} F V_{i_t-1} \, V',$$

where we write $V_i F V_j$ for the path

$$V_i F V_j := V_i V_{i+1} V_{i+2} \ldots V_j,$$

counting indices modulo $|F| = k$. (The case $t = 0$, and thus a walk VV', is allowed.) We say that W traverses F t times and always assume that a shifted walk $S(V, V')$ traverses F

as few times as possible. Its length is the length of the corresponding walk in R. Note that if we can find a skewed V-V' traverse then we can find a shifted V-V' walk.

The most important property of shifted walks is that the walk $W - \{V, V'\}$ visits every vertex in R an equal number of times. Observe also that by our minimality assumption each vertex V_i is visited at most one time from a vertex other than V_{i-1}. I.e. of the t times that V_i is visited at most one does not come from winding around F. This fact will be useful later when we try and bound the number of edges of an embedding not lying on the edges of F.

As with skewed traverses, we can use shifted walks to go from an approximate assignment W of a cycle C to a balanced assignment. Let V_i, V_j be clusters with $a(i) > m$ and $a(j) < m$. If $V_{i-1}V_j, V_jV_{i+1} \in E(R)$ then we could replace one section of W isomorphic to F by $V_{i-1}V_jV_{i+1}FV_{i-1}$, that is, replace $V_{i-1}V_iV_{i+1}$ by $V_{i-1}V_jV_{i+1}$. This new section has the same length as before and so would not alter the rest of W. Clearly we can not ensure that such edges always exist. Instead we use shifted walks and replace a section of the embedding that looks like $FF \ldots F$ with

$$S(V_{i-1}, V_j)S(V_j, V_{i+1})FV_{i-1}F \ldots FV_{i-1};$$

where the $F \ldots F$ in the new embedding contains the appropriate number of F to ensure that it is of exactly the same length as the section of the assignment it replaced. This is a shifted walk from V_{i-1} to V_j, then a shifted walk from V_j to V_{i+1} and then wind around F. By our definition of shifted walks each cluster will have the same number of vertices assigned to it (except V_{i-1}, V_i and V_j) and the total number of vertices assigned will not be altered. Clearly this method needs the cycle we're trying to embed to contain many long sections with no changes of orientation (and oriented in the same direction as F). In the case where the cycle we are trying to embed is close to C^*, the standard orientation of a cycle, we are indeed able to ensure this.

Corollary 14. *Let R be an oriented graph on k vertices with $\delta^0(R) \geq (3/8 + \alpha)k$ for some $\alpha > 0$ and let $F = V_1V_2 \ldots V_k$ be a directed Hamilton cycle of R. Define $r := \lceil 2/\alpha \rceil$. Then for any distinct $V, V' \in V(R)$ there exists the following.*

(i) A skewed V-V' traverse of length at most r.

(ii) A shifted V-V' walk traversing at most r cycles.

Proof. Let A_i be the set of vertices which can be reached from V by a skewed traverse of length at most i and let $A_i^- := \{V_i \in V(R) : V_{i+1} \in A_i\}$. If $|A_i| \geq (1-\alpha)k$ then $N^-(V') \cap A_i^- \neq \emptyset$ and we have a skewed V-V' traverse of length $i+1$. If $|A_i| \leq (1-\alpha)k$ then we can apply Lemma 13 (here we also need that $N^+(V) \neq \emptyset$) to get that $|A_{i+1}| \geq |A_i| + \alpha k/2$. Since $|A_{r-2}| > (1-\alpha)k > k - |N^-(V')|$ we again have $N^-(V') \cap A_{r-2}^- \neq \emptyset$ and hence the desired skewed traverse.

This skewed traverse also gives the desired shifted walk, merely 'wind around' F after each edge. $\qquad\square$

When linking together sections of our cycle we will sometimes need to find a path between two vertices which is not just short but is isomorphic to a path with given length and orientation. To do this we use the following lemma of Häggkvist and Thomason.

Lemma 15 (Häggkvist and Thomason [9]). *Let R be an oriented graph on k vertices with $\delta^0(R) \geq (3/8 + \alpha)k$ for some $\alpha > 0$. Let $4\lceil \log_2(1/\alpha) \rceil \leq k \leq \alpha k/4$ and let P be an arbitrarily oriented path of length k. Then, if k is large enough and $V, V' \in V(R)$ are distinct vertices, there exists a path from V to V' isomorphic to P.*

5   An approximate embedding lemma

Our main tool in our proof of Theorem 3 is the following probabilistic result which says that we can assign a series of paths P_i to the vertices of a small graph R such that each vertex of R is assigned approximately the same number of vertices. Furthermore, we show that if we have a collection of subpaths of the P_i we can assure that every vertex of R is assigned a reasonable number of the starting points of these. When we talk about 'greedily embedding an oriented path P_i around a cycle F given a starting point $V \in V(F)$' we mean the following. Assign the first vertex of P_i to V. Given an embedding of some initial segment of P_i which ends at $V' \in V(F)$ assign the next vertex of P_i to either the successor or the predecessor of V' in F according to the orientation of the edge in P_i.

Lemma 16. *Let R be an oriented graph on k vertices and let F be a Hamilton cycle in R. Let $\mathcal{P} = \{P_1, \ldots, P_s\}$ be a collection of arbitrarily oriented paths on t vertices and $\mathcal{Q}$ be a collection of pairwise disjoint oriented subpaths of the P_i. Then for any $\gamma > 0$ and sufficiently large s there exists a map $\phi : [s] \to V(R)$ such that if the paths are greedily embedded around F with the embedding of each $P(i)$ starting at $\phi(i)$ then the following holds. Define $a(i)$ to be the number of vertices in $\bigcup_{j=1}^s P_j$ assigned to V_i by this embedding and define $n(i, \mathcal{Q})$ to be the number of oriented subpaths in $\mathcal{Q}$ starting at V_i. Then for all $V_i \in V(R)$*

$$\left| a(i) - \frac{st}{k} \right| \leq \gamma st, \tag{1}$$

$$\left| n(i, \mathcal{Q}) - \frac{|\mathcal{Q}|}{k} \right| \leq \gamma st. \tag{2}$$

To prove it we need the following well-known probabilistic bound (see [19] for example).

Theorem 17. *Let X be a random variable determined by s independent trials $X_1, \ldots, X_s$ such that changing the outcome of any one trial can affect X by at most c. Then for any $\lambda > 0$,*

$$\Pr(|X - \mathbb{E}(X)| > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2c^2 s} \right).$$

of Lemma 16. We construct ϕ by picking each $\phi(i)$ independently and uniformly at random. Observe that the assignment of any one path P_i can change the number of vertices assigned to any vertex of R by at most t. Clearly $\mathbb{E}(a(i)) = st/k$. By Theorem 17 we have

$$\Pr(|a(i) - st/k| > \gamma st) \le 2\exp(-\frac{\gamma^2 s^2 t^2}{2t^2 s}) = 2\exp(-\frac{\gamma^2 s}{2}) < 1/(2k)$$

for $s \gg k$.

A similar argument gives that the probability that $n(i, \mathcal{Q})$ differs too much from the expected value is at most $1/(2k)$. Thus the probability that there exists V_i which does not have almost the expected number of vertices or almost the expected number of starting points of paths in $\mathcal{Q}$ assigned to it by ϕ is less than 1. So with positive probability a map constructed in this manner satisfies the conclusion of the lemma, and hence such a map exists. $\qquad\square$

6 Preparations for the Proof of Theorem 3

6.1 The Two Cases

We split into two cases depending on the number of neutral pairs. Let G be an oriented graph on n vertices with $\delta^0(G) \ge (3/8 + \alpha)n$ for some constant $0 < \alpha \ll 1$. Let C be an orientation of a cycle on n vertices with $n(C) =: \lambda n$ neutral pairs. Define the following hierarchy of constants.

$$0 < \varepsilon_1 \ll \varepsilon_2 \ll \varepsilon_3 \ll \varepsilon_4 \ll \varepsilon_5 \ll \varepsilon_6 \ll \alpha < 1.$$

Let $\mathcal{Q}$ be a maximal collection of neutral pairs all at a distance of at least 3 from each other.

If $\lambda \ll \varepsilon_4$ then let $\varepsilon := \varepsilon_6$, $\varepsilon_A := \varepsilon_5$ and $\varepsilon^* := \varepsilon_4$. The proof of this case is given in Section 8.

Otherwise we have $\lambda \gg \varepsilon_3$ and we set $\varepsilon := \varepsilon_3$, $\varepsilon_A := \varepsilon_2$ and $\varepsilon^* := \varepsilon_1$. The proof of this case is in Section 7.

The following two sections, where we partition G and C in preparation for our embedding, are common to both cases.

6.2 Preparing G for the Proof of Theorem 3

Define a positive constant d and integers M'_A, M'_B (all functions of α) such that

$$0 < \varepsilon^* \ll 1/M'_A \ll \varepsilon_A \ll 1/M'_B \ll \varepsilon \ll d \ll \alpha \ll 1.$$

Chernoff type bounds applied to a random partition of $V(G)$ show the existence of a subset $A \subset V(G)$ with $(1/2 - \varepsilon)n \le |A| \le (1/2 - \varepsilon)n$ such that every vertex $x \in V(G)$ satisfies

$$\frac{d^+(x)}{n} - \frac{\alpha}{10} \le \frac{|N_A^+(x)|}{|A|} \le \frac{d^+(x)}{n} + \frac{\alpha}{10} \tag{3}$$

and similarly for $d^-(x)$. Apply the Diregularity lemma (Lemma 6) with parameters ε^2, $d+8\varepsilon^2$ and M'_B to $G-A$ to obtain a partition of the vertex set of $G-A$ into $M_B := k \geq M'_B$ clusters $V_1, \ldots, V_k$ and an exceptional set V_0. Set $B := V_1 \cup \ldots \cup V_k$ and $m_B := |V_1| = \cdots = |V_k|$. Let $G'_B := G[B]$, let R_B denote the reduced oriented graph obtained by an application of Lemma 7 and let G^*_B be the pure oriented graph. By our choice of A we have $\delta^+(G-A)/|G-A| \geq \delta^+(G)/n - \alpha/9$ and a similar bound for δ^-. Hence we can apply Lemma 7 to obtain

$$\delta^0(R_B) \geq \left(\frac{\delta^0(G)}{n} - \frac{\alpha}{4} \right) |R_B| \geq \left(\frac{3}{8} + \frac{3\alpha}{4} \right) |R_B|. \tag{4}$$

So Theorem 1 gives us a Hamilton cycle F_B of R_B. Relabel the clusters of R_B so that $V_i V_{i+1} \in E(F_B)$ for all i where we let $V_{k+1} := V_1$. We now apply Lemma 8 with F_B playing the role of S, ε^2 playing the role of ε 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 ε-regular pair of density at least d (in the underlying graph of G^*_B) and that each edge in F_B corresponds to an (ε, d)-super-regular pair. Note that the new exceptional set now satisfies $|V_0| \leq \varepsilon n$.

Now apply the Diregularity Lemma with parameters $\varepsilon_A^2/4$, $d+2\varepsilon_A^2$ and M'_A to $G[A \cup V_0]$ to obtain a partition of the vertex set of $G[A \cup V_0]$ into $M_A := \ell \geq M'_A$ clusters $V'_1, \ldots, V'_\ell$ and an exceptional set V'_0. Let $A' := V'_1 \cup \cdots \cup V'_\ell$, let R_A denote the reduced oriented graph obtained from Lemma 7 and let G^*_A be the pure oriented graph. As before Lemma 7 implies that $\delta^0(R_A) \geq (3/8 + 3\alpha/4)|R_A|$ and so, as before, we can apply Theorem 1 to find a Hamilton cycle F_A of R_A. Then as before, Lemma 8 implies that by adding at most $\varepsilon_A^2 |A \cup V_0|$ further vertices to the exceptional set V'_0 we may assume that each edge of R_A corresponds to an ε_A-regular pair of density at least d and that each edge in F_A corresponds to an (ε_A, d)-super-regular pair. Finally define $G_B := G[B \cup V'_0]$ and $n_B := |G_B|$ and observe that we now have

$$|V'_0| \leq \varepsilon_A |A \cup V_0|/2 < \varepsilon_A n_B. \tag{5}$$

In both cases of our proof we now have

$$0 < \varepsilon^* \ll 1/M_A \ll \varepsilon_A \ll 1/M_B \ll \varepsilon \ll d \ll \alpha \ll 1.$$

6.3 Preparing C

We would like to divide C into a number of paths and use Lemma 16 to obtain an ε-balanced assignment of C to R. Since we have split our graph G into two parts, we have to split C into two paths P_A and P_B and embed these into (an oriented graph similar to) $G[A']$ and G_B respectively.

Define $r := 4\lceil \log_2(4/\alpha) \rceil$. Lemma 15 tells us that if P is an orientation of a path of length r then between any two distinct vertices in $V(R_B)$ or in $V(R_A)$ there exists a path isomorphic to P.

Define
$$s := \lfloor (\log n)^2 \rfloor, \quad t := \left\lfloor \frac{n - (s+1)(r-1)}{s+2} \right\rfloor - 1 \approx \frac{n}{(\log n)^2}.$$

Recall that $\mathcal{Q}$ is a maximal collection of neutral pairs in C all at a distance of at least 3 from each other. If $\mathcal{Q}$ is large, i.e. we are in the case where C is far from C^*, let v^* be a vertex in C such that the subpath of C of length $n/2$ following v^* and the subpath of C preceding v^* both contain at least $2|\mathcal{Q}|/5$ elements of $\mathcal{Q}$. Divide C into overlapping paths (by which we mean paths sharing endvertices)

$$C := Q_1 P_1 Q_2 P_2 \ldots Q_{s-1} P_{s-1} Q_s P_s Q^* P^*$$

where their lengths satisfy $\ell(P_i) = t$, $\ell(Q_i) = \ell(Q^*) = r$ and $2t \le \ell(P^*) < 3t$ and Q_1 starts at v^*. Let $s_B \in \mathbb{N}$ be such that

$$1 < n_B - s_B(t+r) < \ell(P^*)$$

and let

$$P_B := P_B^* Q_1 P_1 \ldots Q_{s_B} P_{s_B}$$

where P_B^* is an end-segment of P^* of such length as to ensure $\ell(P_B) + 1 = n_B$. Let

$$P_A := Q_1' P_1' \ldots Q_{s_A}' P_{s_A}' Q^* P_A^*$$

where $Q_i' := Q_{s_B + i}$, $P_i' := P_{s_B + i}$, $s_A := s - s_B$ and P_A^* is an initial-segment of P^* which overlaps P_B^* in exactly one place. Observe that we now have

$$n_B = s_B t + s_B r + \ell(P_B^*) + 1 = |V(P_B)| \tag{6}$$

and define

$$n_A := n - n_B = s_A t + (s_A + 1)r + \ell(P_A^*) + 1 = |V(P_A)| - 2.$$

7 Cycle is Far From C^*

7.1 Approximate Embedding

First we use the probabilistic tools in Section 5 to assign the paths P_i to the clusters of R_B in such a way as to ensure that all the clusters are assigned approximately the same number of vertices and the neutral pairs are relatively evenly distributed. Let $\mathcal{Q}_B \subset \mathcal{Q}$ consist of all neutral pairs from $\mathcal{Q}$ which are contained in the P_i and moreover are at a distance of at least three from the ends of the P_i. Apply Lemma 16 to R_B, $\mathcal{P}_B := \{P_1, P_2 \ldots, P_{s_B}\}$ and $\mathcal{Q}_B$ with ε^* as γ to obtain an embedding of the P_i into $V(R_B)$ with

$$\left| a(i) - \frac{s_B t}{M_B} \right| \le \varepsilon^* s_B t, \quad \left| n(i, \mathcal{Q}_B) - \frac{|\mathcal{Q}_B|)}{M_B} \right| \le \varepsilon^* s_B t.$$

for all $V_i \in V(R_B)$. (Recall that $a(i)$ is defined to be the number of vertices of P_B assigned to the cluster V_i by the embedding.) In a slight abuse of notation let $n(i)$ be the number of neutral pairs in $\mathcal{Q}_B$ starting at V_i. Note that

$$|a(i) - m_B| \overset{(6)}{\leq} \left| a(i) - \frac{s_B t}{M_B} \right| + \left| \frac{s_B r + 3t}{M_B} \right| \leq \left| a(i) - \frac{s_B t}{M_B} \right| + \varepsilon^* m_B. \tag{7}$$

The requirement that the neutral pairs in $\mathcal{Q}$ are at a distance of at least three from each other means that $|\mathcal{Q}| \geq n(C)/4$. By the observation in Section 6.3 we know that P_B contains at least $2|\mathcal{Q}|/5 \geq \lambda n/10$ neutral pairs. The paths Q_i and P_B^* together contain fewer than $s_B r + 3t$ neutral pairs and at most $4s_B$ neutral pairs can be in the P_i but within a distance of at most three from a Q_i. Thus for all i

$$n(i) \geq \frac{\lambda n}{10 M_B} - \varepsilon^* s_B t - (s_B r + 3t + 4s_B) \geq \frac{\lambda n_B}{6 M_B} - 2\varepsilon^* n_B \geq \frac{\lambda m_B}{7}.$$

For all $2 \leq i \leq s_B$ we can join P_{i-1} and P_i by a path in R_B isomorphic to Q_i using Lemma 15. Furthermore we can greedily extend P_1 backwards by a path isomorphic to $P_B^* Q_1$. This will increase $a(i)$ by at most $s_B r + 3t < \varepsilon^* m_B$ for n sufficiently large. We now have an assignment of P_B to the clusters of R_B which we can think of as a walk W_B in R_B.

7.2 Incorporating the Exceptional Vertices

Let G_B^c be the digraph obtained from the pure oriented graph G_B^* 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_B) and adding the vertices in V_0' as well as all the edges of G between V_0' and $V(G_B - V_0')$. Our next aim is to incorporate the exceptional vertices V_0' into the walk W_B. We do this by considering the following extension of R_B. Define $R_B^* \supseteq R_B$ to be the digraph formed by adding to R_B the vertices in V_0' and, for $v \in V_0'$ and $V_i \in V(R_B)$, the edge vV_i if $|N_G^+(v) \cap V_i| > \alpha m_B/10$ and $V_i v$ if $|N_G^-(v) \cap V_i| > \alpha m_B/10$.

Then for each $v \in V_0'$ pick an inneighbour $V_i \in V(R_B)$ and change the assignment of one neutral pair currently mapped to $V_i V_{i+1} V_i$ to $V_i v V_i$. We can always find such an inneighbour as (3) implies that each exceptional vertex sees at least a three-eighths proportion of the vertices in $V(G_B)$. This reduces $a(i+1)$ and $n(i)$ by one. Figure 3 contains an illustration of this, where we consider W_B as being in G_B^c and the dotted lines as the section of the embedding to be replaced by the solid lines. After doing this for every exceptional vertex we will have that for all $V_i \in V(R_B)$

$$|a(i) - m_B| \overset{(7)}{\leq} \left| a(i) - \frac{s_B t}{M_B} \right| + \varepsilon^* m_B$$

$$\leq (\varepsilon^* s_B t + \varepsilon_A m_B + |V_0'|) + \varepsilon^* m_B \overset{(5)}{<} 4\varepsilon_A n_B, \tag{8}$$
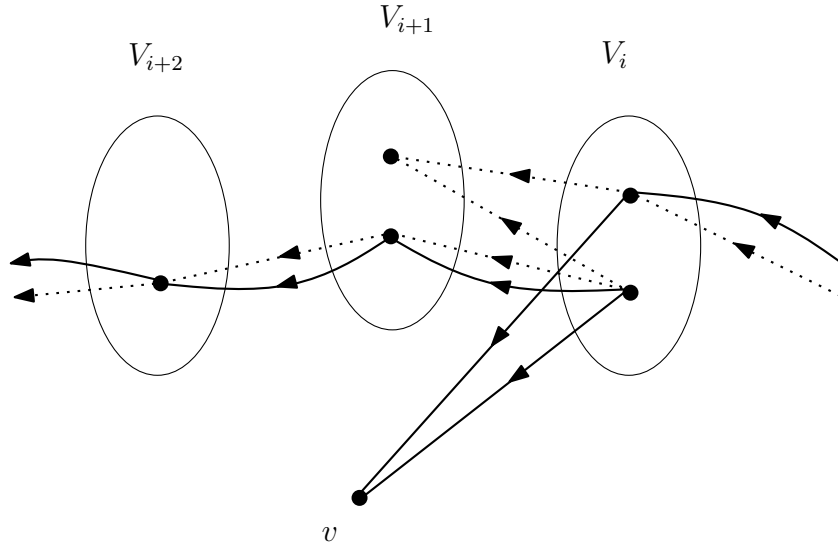
Figure 3: Incorporating an exceptional vertex when C is far from C^*.

where the second term in the second line comes from greedily embedding the Q_i. We also still have a reasonable number of neutral pairs starting at each cluster of R_B for all $V_i \in V(R_B)$:

$$ n(i) \geq \frac{\lambda m_B}{7} - |V_0'| > \frac{\lambda m_B}{7} - \varepsilon_A n_B > \frac{\lambda m_B}{8}. $$

Note that of the $a(i)$ vertices of P_B assigned to any $V_i \in V(R)$, at most $\varepsilon_A n_B + 2|V_0'| \leq 3\varepsilon_A n_B$ do not have their neighbours assigned to $V_{i-1} \cup V_{i+1}$, where the first term came from the Q_i and the second came from incorporating the exceptional vertices. Thus we currently have a $(4\varepsilon_A, 3\varepsilon_A)$-corresponding embedding of P_B into R_B^*.

7.3 Adjusting the Embedding

We now adjust W_B to obtain a $5\varepsilon_A M_B$-corresponding assignment of P_B to R_B^*; i.e. we adjust W_B to ensure that $a(i) = m_B$ for all $V_i \in V(R_B)$. Recall from Corollary 14 that between any two vertices in R_B there exists a skewed traverse of length at most $r' := \lceil 4/\alpha \rceil$. Then for each cluster $V_i \in V(R_B)$ with $a(i+1) > m_B$ pick $V_j \in V(R_B)$ with $a(j) < m_B$ and find a skewed V_i-V_j traverse of length $q \leq r'$:

$$ V_i V_{k_1}, V_{k_1-1} V_{k_2}, V_{k_2-1} V_{k_3}, \ldots, V_{k_q} V_{k_q-1}, V_{k_q-1} V_j. $$

As discussed in Section 4 we can use this skewed traverse to modify W_B to reduce $a(i+1)$ by one, increase $a(j)$ by one and leave the number of vertices assigned to every other cluster of R_B the same. We do this by, for every $0 \leq p \leq q$, replacing a neutral pair $V_{k_p-1} V_{k_p} V_{k_p-1}$ in W_B by $V_{k_p-1} V_{k_{p+1}} V_{k_p-1}$ where we define $V_{k_0-1} := V_i$ and $V_{k_{q+1}} := V_j$.

Since $\sum_{i=1}^{M_B} |a(i) - m_B| \leq 4\varepsilon_A M_B n_B$, doing this will consume at most $4\varepsilon_A M_B n_B$ neutral pairs starting at any vertex of R_B. This is fine though as for all $V_i \in V(R_B)$ we have $n(i) \geq \lambda m_B/8 \gg 4\varepsilon_A M_B n_B$. Each cluster V_i now has at most $3\varepsilon_A n_B + 4\varepsilon_A M_B n_B <$

$5\varepsilon_A M_B n_B$ vertices of P_B assigned to it that do not have both their neighbours assigned to $V_{i-1} \cup V_{i+1}$. Hence we have constructed a $5\varepsilon_A M_B$-corresponding embedding W_B of P_B into R_B^*.

7.4   Finding a copy of P_B in G_B

We will now use Lemma 10 to find a copy of P_B in G_B. To do this we use W_B to find an embedding W_B' of P_B into G_B such that

- Every vertex of W_B in V_0' is unchanged in W_B'.

- Each appearance of a cluster of R_B in W_B is replaced by a unique vertex in the corresponding cluster in G_B.

- Every edge of W_B which does not lie upon an edge of F_B is mapped to an edge of G_B.

First we split W_B into two digraphs W_B^1 and W_B^2. Let W_B^1 consist of all maximal walks

$$u_{i,1} u_{i,2} \ldots u_{i,\ell_i}$$

in W_B of length at least three whose edges all lie on F_B. Let W_B^2 consist of everything not in W_B^1. Then W_B^2 is a union of walks $v_{i,1} v_{i,2} \ldots v_{i,k_i}$, where we relabel if necessary to ensure that $u_{i,1} = v_{i-1,k_{i-1}}$ and $u_{i,\ell_i} = v_{i,1}$. In the next paragraph we will greedily find an embedding of W_B^2 into G_B which will satisfy the third requirement above.

The walks in W_B^2 are of one of three types. The first type comes from the incorporation of an exceptional vertex, in which we have an exceptional vertex $x \in V_0'$ and a cluster $V_i \in V(R_B)$ with $|N_G^-(x) \cap V_i| > \alpha m_B/10$. In this case we choose any two distinct vertices $u, v \in N_G^-(x) \cap V_i$, which we can do as there are at most $|V_0'| \ll \varepsilon m_B \ll \alpha m_B/10$ exceptional vertices. The second type comes from the paths Q_i and the path P_B^*. These we find in G_B^* (and hence in $G_B \supseteq G_B^*$) greedily. We can do so as the total length of the Q_i is at most $s_B r + 2t \ll \varepsilon m_B$ and all their edges are assigned to edges in R_B corresponding to ε-regular pairs of density at least d in G_B^*. The final type are pairs of edges ij, ji with $i, j \in V(R_B)$ which come from the skewed traverses used to ensure that the correct number of vertices of P_B were assigned to each vertex of R_B. There are at most $5\varepsilon_A M_B n \ll \varepsilon m_B$ of these and so we can again find these greedily. Note that our requirement that all the neutral pairs in $\mathcal{Q}$ are at a distance of at least three from each other and the ends of the P_i implies that we have now considered all possible walks in W_B^2. To satisfy the second condition above we simply assign each vertex of W_B not already assigned to a (distinct) vertex in the corresponding cluster in G_B. As W_B is balanced (i.e. W_B assigns exactly m_B vertices to each cluster) we can do this.

For all i let S_i consist of the vertices of $G_B - V_0'$ to which the vertices of W_B^1 that are not at the end of a path have been assigned. We can now apply Lemma 10 to $G_B - V_0'$ with W_B^1 as H, the $u_{i,1}$ and u_{i,ℓ_i} as the x_P and y_P respectively and the S_i as just defined. Combining this with the embedding of W_B^2 into G gives us a copy of P_B in G_B.

7.5 Finding a copy of C in G

Recalling how we 'chopped up' C at the start of this section, let $u, v \in V(G_B)$ be the vertices to which the endpoints of P_B were assigned. To complete the proof of this case we find a copy of P_A in $G_A := G[A' \cup \{u, v\}]$ starting at v and ending at u. We find a copy of P_A exactly as we found the copy of P_B with three differences. Firstly there are no exceptional vertices. Secondly, recalling that

$$P_A := Q_1' P_1' \ldots Q_{s_A}' P_{s_A}' Q^* P_A^*,$$

we require that the embeddings of Q_1' and P_A^* start and end at v and u respectively. Since Q_1' is long enough for Lemma 15 we can specify the cluster to which its initial vertex is assigned and use Lemma 15 to join it to P_1'. We embed P_A^* greedily and use Q^* and Lemma 15 to connect it with the rest of the embedding. Hence we can indeed start and end at the required vertices. This doesn't affect the constants in the rest of the proof. Since the number of exceptional vertices and the imbalances created by the approximate embedding are both small (and small as functions of M_A) we can proceed exactly as before and find the desired cycle C in G. The calculations work as before as a result of us only having two exceptional vertices. The equation (8) becomes

$$|a(i) - m_A| \leq \left| a(i) - \frac{s_A t}{M_A} \right| + \varepsilon^* m_A$$
$$\leq (\varepsilon^* s_A t + \varepsilon_A m_A + |\{u, v\}|) + \varepsilon^* m_A \leq 4\varepsilon_A m_A.$$

Hence from Section 7.4 we now have

$$\sum_{i=1}^{M_A} |a(i) - m_A| \leq 4\varepsilon_A M_A m_A,$$

which is fine as we will have that $n(i) \geq \lambda m_A / 8 \ll 4\varepsilon_A M_A m_A$ for all clusters $V_i' \in V(R_A)$.

8 Cycle is Close to C^*

Our argument closely follows that in the previous section, the difference being in the means of correcting imbalances. To correct imbalances we will need long sections of P_B with no changes in orientation. Define $\ell_B := \lceil \frac{4}{\alpha} \rceil M_B$, which is at least the maximum length of a shifted walk between two vertices in R_B. As before we split up C into P_A and P_B, the only difference being that we do not need a special vertex v^* this time. Let $\mathcal{Q}_B'$ consist be the largest possible collection of paths in P_B of length $3\ell_B$ all at a distance of at least 3 from each other, oriented in the same direction and containing no changes in orientation. We will call these *long runs*. There are at least

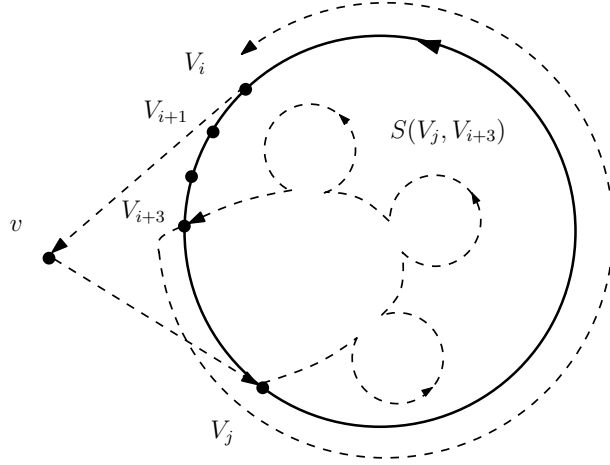$$m(P_B, \mathcal{Q}_B') \geq \frac{n_B}{3\ell_B + 6} - 2\lambda n \geq \frac{\alpha n_B}{14 M_B}$$

Figure 4: Incorporating an exceptional vertex when C is close to C^*.

of these in P_B. (We subtract $2\lambda n$ not λn as both neutral pairs $V_i V_{i+1} V_i$ and their inverse $V_i V_{i-1} V_i$ kill possible long runs.)

Let $\mathcal{Q}_B$ be the subset of $\mathcal{Q}'_B$ containing those long runs contained in the P_i, at a distance of at least 4 from the ends of all the P_i and all oriented in the same direction. We assume that these are all oriented in the same direction as F_B. Keeping only long runs oriented in one direction loses us at most half of them. The paths Q_i, the path $Q^* P_B^*$ and the 3 vertices neighbouring them in the P_i in each direction can intersect at most $2s + 2$ of the long runs and so, abusing notation slightly,

$$ m(\mathcal{P}_B) \geq \frac{\alpha n_B}{28 M_B} - 2s - 2 \geq \frac{\alpha n_B}{30 M_B} $$

for sufficiently large n, where we recall that $\mathcal{P}_B := \{P_1, P_2 \ldots, P_{s_B}\}$. Similarly defining $\ell_A := \lceil \frac{4}{\alpha} \rceil M_A$ and $\mathcal{Q}'_A$ and $\mathcal{Q}_A$ in the obvious way we have $m(\mathcal{P}_A) \geq \alpha(n_A)/30 M_A$.

Apply Lemma 16 to R_B, $\mathcal{Q}_B$ and $\mathcal{P}_B$ with ε^* as γ to obtain an embedding of the P_i into $V(R_B)$ with

$$ \left| a(i) - \frac{s_B t}{M_B} \right| \leq \varepsilon^* s_B t, \quad m(i) \geq \frac{\alpha n_B}{30 M_B^2} - \varepsilon^* s_B t \geq \frac{\alpha n_B}{32 M_B^2} \tag{9} $$

for all $V_i \in V(R_B)$, where we write $m(i)$ for the number of elements of $\mathcal{Q}_B$ whose initial vertex is assigned to $V_i \in V(R)$.

For all $2 \leq i \leq s_B$ we can join P_{i-1} and P_i by a path in R_B isomorphic to Q_i using Lemma 15. Furthermore we can greedily extend P_1 backwards by a path isomorphic to $P_B^* Q_1$. This will increase $a(i)$ by at most $s_B r + 2t \leq \varepsilon_A m_B$ for n sufficiently large. We now have an embedding of P_B into R_B which we can think of as a walk W_B in R_B.

Let G_B^*, G_B^c and R_B^* be defined exactly as in Section 7.2. Let $v \in V_0'$ be an exceptional vertex and let $V_i v, v V_j \in E(R_B^*)$. ($V_i$ and V_j exist by (3).) Take a long run in $\mathcal{Q}_B$ whose initial vertex is currently assigned to V_i. Since M_B divides ℓ_B it also ends at V_i. We cannot replace the long run simply by $V_i v V_j F_B \ldots F_B$ because this would not end at V_i. Thus it

would require us to alter the rest of our approximate embedding, possibly causing (9) to no longer hold. Instead we use shifted walks and a 'jump' to ensure that our modification incorporates v into our walk and does not alter $a(i)$ or $m(i)$ significantly for any cluster of R_B. We replace the long run starting at V_i with the following walk

$$V_i v V_j S(V_j, V_{i+3}) F_B F_B \ldots F_B V_i,$$

where $S(V_j, V_{i+3})$ is a shifted walk from V_j to V_{i+3}. The number of F_B is chosen so that the new section has exactly the same length as the long run it replaces. This is illustrated in Figure 4. This is a walk that goes out to v, back to V_j, follows a shifted walk to V_{i+3} and then winds around F until we have a walk of length $3\ell_B$ ending at V_i. This new walk visits V_{i+1} and V_{i+2} one time fewer than previously and V_j one time more. Observe that the shifted walk by definition visits every cluster in R_B the same number of times, which allows us to observe that we still end at V_i. Repeating this for each exceptional vertex creates a new assignment now satisfying

$$|a(i) - m_B| \overset{(6)}{\leq} \left| a(i) - \frac{s_B t}{M_B} \right| + \left| \frac{s_B r + 2t}{M_B} \right|$$
$$\leq (\varepsilon_A s_B t + \varepsilon_A m_B + |V_0'|) + \varepsilon_A m_B \leq 3\varepsilon_A n_B.$$

for all i. We also still have a reasonable number of long runs starting at each cluster.

$$m(i) \geq \frac{\alpha n_B}{32 M_B^2} - |V_0'| \geq \frac{\alpha n_B}{40 M_B^2}.$$

Note that of the $a(i)$ vertices of P_B assigned to $V_i \in V(R)$, at most

$$\varepsilon_A m_B + 4|V_0'| \leq 5\varepsilon_A n_B$$

do not have their neighbours assigned to $V_{i-1} \cup V_{i+1}$. The first term here comes from connecting the P_i and the second term from incorporating the exceptional vertices: each exceptional vertex has one direct edge to or from a given cluster in R_B and the shifted walk can add at most two edges outside F_B to each cluster. Thus we currently have a $(5\varepsilon_A, 3\varepsilon_A)$-corresponding assignment of P_B into R_B^*.

8.1 Correcting the imbalances

We now adjust our current assignment of P_B to R_B^* to obtain a $15\varepsilon_A$-corresponding assignment, i.e. we adjust W_B to ensure that $a(i) = m_B$ for all $V_i \in V(R_B)$. To do this we find a pair $V_i, V_j \in V(R_B)$ such that $a(i) > m_B$ and $a(j) < m_B$ and replace a long run starting at V_{i-1} with the following walk:

$$S(V_{i-1}, V_j) S(V_j, V_{i+1}) F_B \ldots F_B V_{i-1},$$

where the number of F_B is chosen to ensure that the new section has length $3\ell_B$. This walk removes the assignment of one vertex to V_i, assigns one extra vertex to V_j and does

not change the number of vertices assigned to all other clusters in R_B. Since $\sum_{i=1}^{M_B} a(i) = m_B M_B$ we can always find such a pair unless we have corrected all the imbalances. Each pair requires a long run and we still have at least $\alpha n_B / 40 M_B^2 \gg 3\varepsilon_A n_B$ of these starting at each cluster and so can indeed correct all the imbalances. This leaves us with a balanced assignment with at most

$$3\varepsilon_A n_B + 4 \cdot 3\varepsilon_A n_B = 15\varepsilon_A n_B$$

edges outside F_B from each vertex. Hence there are at most $15\varepsilon_A M_B n_B \ll \varepsilon m_B$ edges in total not in a path of length at least 3 all of whose edges lie on F_B or not lying entirely on F_B. This is exactly the same position as in Section 7.4. We can now proceed as before to first find a copy of P_B in G_B and then repeat the procedure with P_A (using $\mathcal{Q}_A$ not $\mathcal{Q}_B$) to find the desired cycle C in G. This completes this section and the proof of Theorem 3.

Acknowledgements

References

[1] N. Alon and A. Shapira, Testing subgraphs in directed graphs, *Journal of Computer and System Sciences* **69** (2004), 354–382.

[2] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications,* Springer 2000.

[3] D. Christofides, P. Keevash, D. Kühn and D. Osthus, Finding Hamilton cycles in robustly expanding digraphs, submitted.

[4] D. Christofides, P. Keevash, D. Kühn and D. Osthus, A semi-exact degree condition for Hamilton cycles in digraphs, submitted.

[5] A. Ghouila-Houri, Une condition suffisante d'existence d'un circuit hamiltonien, *C.R. Acad. Sci. Paris* **25** (1960), 495–497.

[6] D. Grant, Antidirected Hamiltonian cycles in digraphs, *Ars Combinatoria* **10** (1980), 205–209.

[7] R. Häggkvist, Hamilton cycles in oriented graphs, *Combin. Probab. Comput.* **2** (1993), 25–32.

[8] R. Häggkvist and A. Thomason, Oriented Hamilton cycles in digraphs, *Journal of Graph Theory* **20** (1995), 471–479.

[9] R. Häggkvist and A. Thomason, Oriented Hamilton cycles in oriented graphs, in *Combinatorics, Geometry and Probability*, Cambridge University Press (1997), 339–353.

[10] R. Häggkvist and C. Thomassen, On pancyclic digraphs, *J. Combinatorial Theory B* **20** (1976), 20–40.

[11] P. Keevash, D. Kühn and D. Osthus, An exact minimum degree condition for Hamilton cycles in oriented graphs, *Journal of the London Math. Soc.* **79** (2009), 144–166.

[12] L. Kelly, On cycles in dense directed graphs, PhD thesis, School of Mathematics, University of Birmingham, in preparation.

[13] L. Kelly, D. Kühn and D. Osthus, Cycles of given length in oriented graphs, submitted

[14] L. Kelly, D. Kühn and D. Osthus, A Dirac type result on Hamilton cycles in oriented graphs, *Combin. Probab. Comput.* **17** (2008), 689–709.

[15] J. Komlós, The Blow-up lemma, *Combin. Probab. Comput.* **8** (1999), 161–176.

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

[18] D. Kühn, D. Osthus and A. Treglown, Hamiltonian degree sequences in digraphs, submitted.

[19] M. Molloy and B. Reed, *Graph Colouring and the Probabilistic Method*, Springer 2002.

[20] C. St. J. A. Nash-Williams, Hamiltonian circuits, *Studies in Math.* **12** (1975), 301–360.

[21] M. Nathanson, The Caccetta-Häggkvist conjecture and additive number theory, *ArXiv Mathematics e-prints* (2006).

[22] A. Thomason, Paths and cycles in tournaments, *Trans. Amer. Math. Soc.* **296** (1986), 167–180.

[23] 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.

[24] A. Young, Extremal problems for dense graphs and digraphs, PhD thesis, School of Mathematics, University of Birmingham (2007).