
SUBGRAPHS OF WEAKLY QUASI-RANDOM ORIENTED GRAPHS

Omid Amini, Simon Griffiths & Florian Huc

Abstract. — It is an intriguing question to see what kind of information on the structure of an oriented graph D one can obtain if D does not contain a fixed oriented graph H as a subgraph. The related question in the unoriented case has been an active area of research, and is relatively well-understood in the theory of quasi-random graphs and extremal combinatorics.

In this paper, we consider the simplest cases of such a general question for oriented graphs, and provide some results on the global behavior of the orientation of D . For the case that H is an oriented four-cycle we prove: in every H -free oriented graph D , there is a pair $A, B \subseteq V(D)$ such that $e(A, B) \geq e(D)^2/32|D|^2$ and $e(B, A) \leq e(A, B)/2$. We give a random construction which shows that this bound on $e(A, B)$ is best possible (up to the constant). In addition, we prove a similar result for the case H is an oriented six-cycle, and a more precise result in the case D is dense and H is arbitrary. We also consider the related extremal question in which no condition is put on the oriented graph D , and provide an answer that is best possible up to a multiplicative constant. Finally, we raise a number of related questions and conjectures.

1. Introduction

Many results in graph theory show that local properties of a graph have global consequences. An example, which is particularly relevant to us, comes from the theory of quasi-randomness [7], and shows that if a large dense graph G does not contain a fixed graph H as a subgraph, then it must contain a large (linear-size) subgraph whose density is significantly different from that of the original graph. Thus, the restriction that G does not contain H as a subgraph forces G to exhibit irregular behaviour on a global scale. In another direction, it has been conjectured by Erdős and Hajnal [10] that for each graph H , there is a constant $\alpha > 0$ such that every graph G that does not contain an induced copy of H , must have a clique or independent set of size at least n^α . A partial result, proving the existence of a clique or independent set of size $\exp(c\sqrt{\log n})$, has been given [10] (see also [9]).

We prove results for oriented graphs which also show that large scale irregularities can be deduced from local properties. In this sense our results are quasi-randomness type results for oriented graphs. That said, they differ significantly in style from the work on quasi-randomness by Thomason [23, 24]; Chung, Graham and Wilson [7]; and Chung and Graham [6], which is of a more precise form and applies to dense graphs only. In fact, [6] is concerned with tournaments (orientations of complete graphs), and does indeed include a proof that excluding a fixed tournament as a subgraph (or even bounding the number of copies away from the expected value) does lead to consequences on a global scale (for example, vertex subsets X, Y with $|e(X, Y) - e(Y, X)| = \Omega(n^2)$). They actually prove much more – that eleven separate quasi-randomness conditions are equivalent. A more direct analogue of these results for general (dense) oriented graphs has been considered by the second author [15].

Throughout the article we write D for an oriented graph, n for the number of vertices of D , and $e(D)$ for the number of arcs (oriented edges) in D . For a pair of (non-necessarily disjoint) subsets $A, B \subseteq V$, we write $e(A, B)$ for the number of arcs from A to B , i.e., the size of the set $E(A, B) = \{\vec{xy} \in E(D) : x \in A, y \in B\}$. We say that a subgraph $E(A, B)$ is *biased* if $e(B, A) \leq e(A, B)/2$. (The choice of the fraction $\frac{1}{2}$ is rather arbitrary, it could be replaced by any $\eta \in (0, 1)$ without any significant effect on our results.) We write $bias(D)$ for the size of the largest biased subgraph in D :

$$bias(D) := \max \left\{ e(A, B) : A, B \subseteq V \text{ such that } e(B, A) \leq \frac{e(A, B)}{2} \right\}.$$

The parameter $bias(D)$ measures irregularities in the orientation of D . If $bias(D)$ is small, then one might say that D has a random-like orientation. For example, if D is obtained by orienting edges at random, then $bias(D) = O(n)$ with high probability (see Lemma 4.1). In general, however, $bias(D)$ may be as large as $e(D)$, the number of arcs of D . We now state our main result.

Theorem 1.1. — *There exists a constant $\varepsilon > 0$ such that every oriented graph D with $bias(D) < \varepsilon e(D)^2/n^2$ contains an oriented four-cycle.*

Remark. — This result is best possible, up to the choice of ε . In Section 4 we construct a family of oriented graphs D which have $bias(D) < K e(D)^2/n^2$ but do not contain oriented four-cycles, where K is a fixed constant.

Theorem 1.2. — *There exists a constant $\varepsilon > 0$ such that every oriented graph D with $bias(D) < \varepsilon e(D)^2/n^2$ contains an oriented six-cycle.*

We have no reason to believe that this result is best possible. In fact we conjecture a stronger result.

Conjecture 1.3. — *There exists a constant $\varepsilon > 0$ such that every oriented graph D with $bias(D) < \varepsilon e(D)^{3/2}/n$ contains an oriented six-cycle.*

We also have a more general conjecture concerning oriented cycles of all even lengths.

Conjecture 1.4. — *For each $k \geq 2$, there exists a constant $\varepsilon > 0$ such that every oriented graph D with $\text{bias}(D) < \varepsilon e(D)^{k/(k-1)}/n^{2/(k-1)}$ contains an oriented cycle of length $2k$.*

Remark. — The analogous question for oriented cycles of odd length is not so interesting. Indeed, a random orientation D of the complete bipartite graph with parts of order $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ has $\text{bias}(D)$ very small (of order n) while D does not contain any oriented cycle of odd length. Looking for odd length cycles may become interesting in the case the underlying graph is not arbitrary, cf. Concluding Remarks (Section 7).

In addition to guaranteeing the presence of an oriented cycle of prescribed length, the above conditions can also be used to guarantee many such cycles.

Theorem 1.5. — *There exist constants $c, \varepsilon > 0$ such that every oriented graph D with $\text{bias}(D) < \varepsilon e(D)^2/n^2$ contains at least $ce(D)^4/n^4$ oriented four-cycles.*

Theorem 1.6. — *There exist constants $c, \varepsilon > 0$ such that every oriented graph D with $\text{bias}(D) < \varepsilon e(D)^2/n^2$ contains at least $ce(D)^6/n^6$ oriented six-cycles.*

Remark. — The number of cycles obtained is (up to the constant) best possible. A random orientation of the random graph $G(n, e)$ (recall that $G(n, e)$ is a graph selected uniformly at random from the class of all graphs on n vertices with e edges) will generally obey the *bias* condition (see Lemma 4.1), and have $\Theta(e^4/n^4)$ oriented four-cycles and $\Theta(e^6/n^6)$ oriented six-cycles.

The above results are relatively general in the sense that they apply in the case D is sparse as well as the case D is dense, while they are restrictive in the sense that they only look for oriented four-cycles and oriented six-cycles. The next theorem does the opposite – restricting to the case D is dense, it counts copies of any oriented graph H .

In what follows we count homomorphic copies of subgraphs. Let H be an oriented graph on k vertices. We write $\text{hom}(H, D)$ for the number of homomorphic copies of H in D , i.e., the number of functions $\phi : V(H) \rightarrow V(D)$ such that $\phi(x)\vec{\phi}(y) \in E(D)$ for every arc $x\vec{y} \in E(H)$. Likewise, we write $\text{hom}(\bar{H}, D)$ for the number of homomorphic copies of the unoriented graph \bar{H} in D , i.e., the number of functions $\phi : V(H) \rightarrow V(D)$ such that for every edge $\{x, y\} \in E(\bar{H})$ either $\phi(x)\vec{\phi}(y)$ or $\phi(y)\vec{\phi}(x)$ is an arc of D .

Theorem 1.7. — *Let H be an oriented graph on k vertices and let D be an oriented graph with $\text{bias}(D) < \varepsilon n^2$. Then*

$$\text{hom}(H, D) \geq \frac{\text{hom}(\bar{H}, D)}{3^{e(H)}} - \frac{\varepsilon}{2} n^k.$$

Since one may easily bound the number of degenerate homomorphic copies of H , one easily deduces the following corollary.

Corollary 1.8. — *Let H be an oriented graph on k vertices and let D be an oriented graph with $\text{bias}(D) < \varepsilon n^2$. If $\text{hom}(\bar{H}, D) \geq 3^{e(H)} \varepsilon n^k$ and $n \geq 4\varepsilon^{-1}$, then D contains H as a subgraph.*

Since the main results of the article may be viewed as lower bounds on $\text{bias}(D)$ for H -free oriented graphs D (for given choices of H), it is natural to also consider the related extremal question:

What is the minimum value of $\text{bias}(D)$ over oriented graphs on n (non-isolated) vertices?

It will be seen from the results we obtain that this is indeed the most natural form of the extremal question. This question will be considered in detail in Section 6. The main results establish that the minimum value of $\text{bias}(D)$ over oriented graphs on n non-isolated vertices is $\Theta(n/\log n)$. We also prove that the answer is larger, $\Theta(n)$, in the case that the oriented graph is regular, or close to regular. In addition, we provide an algorithmic proof of the lower bound in the case D is regular.

The layout of the article is as follows. In Section 2 we prove Theorem 1.1 and Theorem 1.5. In Section 3 we prove Theorem 1.2 and Theorem 1.6. In Section 4 we describe a family of oriented graphs which show that Theorem 1.1 is best possible (up to the choice of ε). In Section 5 we consider the case of dense oriented graphs, and provide a proof of Theorem 1.7. In Section 6 we discuss extremal questions concerning the parameter $\text{bias}(D)$. Finally, we give in Section 7 a number of concluding remarks. These remarks include new questions and conjectures, together with further discussion of the conjectures stated above. We should note here that it is possible to obtain another proof of Corollary 1.8 (with some different (worse) constants) using Szemerédi's regularity lemma. This is discussed in Appendix A.

2. Oriented Four-Cycles

We begin by defining some notation. Let x be a vertex of an oriented graph D . We shall use the following notation:

$$\begin{aligned}\Gamma^+(x) &= \{y \in V : \vec{xy} \in E(D)\}, \\ \Gamma^-(x) &= \{y \in V : \vec{yx} \in E(D)\}, \\ \Gamma^{++}(x) &= \{y \in V : \exists z \in V \vec{xz}, \vec{zy} \in E(D)\}, \\ \Gamma^{--}(x) &= \{y \in V : \exists z \in V \vec{yz}, \vec{zx} \in E(D)\}.\end{aligned}$$

We let $d^+(x) = |\Gamma^+(x)|$ and $d^-(x) = |\Gamma^-(x)|$. We also define the following notation for joint-degrees, we let $d^{++}(x, u) = |\Gamma^+(x) \cap \Gamma^+(u)|$ and $d^{+-}(x, u) = |\Gamma^+(x) \cap \Gamma^-(u)|$.

We now prove a useful lemma. Recall that throughout D denotes an oriented graph on n vertices with e arcs.

Lemma 2.1. — *Let D be an oriented graph with $\text{bias}(D) \leq e(D)/2$. Then D contains at least $e(D)^2/8n$ paths of length two.*

Proof. — The number of paths of length two in D is

$$\sum_{y \in V} d^+(y)d^-(y).$$

Denote by Z the set of vertices y for which $d^-(y) > 2d^+(y)$. By summing over vertices $y \in Z$ one finds that $e(V, Z) > 2e(Z, V)$, and so, by the definition of $\text{bias}(D)$, we have that $e(V, Z) \leq \text{bias}(D) \leq e/2$. Therefore $e(V, Y) \geq e(D)/2$, where $Y = V \setminus Z$. We also have that $d^+(y) \geq d^-(y)/2$ for all $y \in Y$, so that

$$\sum_{y \in V} d^+(y)d^-(y) \geq \frac{1}{2} \sum_{y \in Y} d^-(y)^2 \geq \frac{1}{2n} \left(\sum_{y \in Y} d^-(y) \right)^2,$$

where the final inequality follows from an application of the Cauchy-Schwarz inequality. The proof is now complete, as $\sum_{y \in Y} d^-(y) = e(V, Y) \geq e(D)/2$. \square

We now turn to the proof of Theorem 1.1. We shall prove this theorem with $\varepsilon = 1/32$. Thus, throughout the proof we may assume that D has $\text{bias}(D) < e(D)^2/32n^2$. Note also the following useful formulation of the $\text{bias}(D) < \varepsilon e(D)^2/n^2$ property:

$$e(B, A) \geq e(A, B)/2 \quad \text{whenever} \quad e(A, B) \geq \varepsilon e(D)^2/n^2. \quad (\star)$$

Also we introduce a final piece of notation. For each vertex x , we write e_x for the number of paths of length two in D which start at x . Equivalently, $e_x = e(\Gamma^+(x), \Gamma^{++}(x))$.

Proof of Theorem 1.1. — Let D be an oriented graph with $\text{bias}(D) < e(D)^2/32n^2$. By Lemma 2.1, we know that there are at least $e(D)^2/8n$ paths of length two in D . Denote by W the set of vertices x with $e_x \geq e(D)^2/16n^2$. Since at most $e(D)^2/16n$ paths of length two start at vertices outside of W , we have that $\sum_{x \in W} e_x \geq e(D)^2/16n$.

For each vertex $x \in W$, we have $e(\Gamma^+(x), \Gamma^{++}(x)) = e_x \geq e(D)^2/16n^2$ and so, by (\star) , we have that $e(\Gamma^{++}(x), \Gamma^+(x)) \geq e_x/2$. Equivalently

$$\sum_{u \in \Gamma^{++}(x)} d^{++}(x, u) \geq \frac{e_x}{2}.$$

Summing over $x \in W$, we obtain

$$\sum_{x \in W} \sum_{u \in \Gamma^{++}(x)} d^{++}(x, u) \geq \sum_{x \in W} \frac{e_x}{2} \geq \frac{e(D)^2}{32n}.$$

We now consider a change in the order of summation.

$$\sum_{u \in V} \sum_{x \in \Gamma^{--}(u)} d^{++}(x, u) \geq \sum_{x \in W} \sum_{u \in \Gamma^{++}(x)} d^{++}(x, u) \geq \frac{e(D)^2}{32n}.$$

In particular this implies that for some $u \in V$ one has $\sum_{x \in \Gamma^{--}(u)} d^{++}(x, u) \geq e(D)^2/32n^2$. Equivalently, $e(\Gamma^{--}(u), \Gamma^+(u)) \geq e(D)^2/32n^2$. A final application of (\star) gives that

$$e(\Gamma^+(u), \Gamma^{--}(u)) \geq e(D)^2/64n^2 > 0.$$

This precisely gives us an oriented four-cycle. \square

In the above proof we use the property (\star) , together with the trick of changing the order of summation, to deduce the existence of a pair of vertices x, u between which there is a path of length two in each direction. This is not a rare occurrence, rather it is typical. Further, it is typical that the number of paths of length two in the two directions is of the same order. The following lemma proves this fact and allows us to deduce Theorem 1.5. Recall that $d^{+-}(x, u)$ denotes the number of paths of length two from x to u . Say that (x, u) is *unbalanced* if $d^{+-}(x, u) > 16d^{+-}(u, x)$, otherwise it is *balanced*. A path of length two $\vec{x}\vec{y}, \vec{y}\vec{u}$ is called *unbalanced* if (x, u) is unbalanced, otherwise it is *balanced*.

Lemma 2.2. — *Let D be an oriented graph with $\text{bias}(D) < \varepsilon e(D)^2/n^2$. Then the number of unbalanced paths of length two in D is at most $8\varepsilon e(D)^2/n$.*

Proof of Theorem 1.5. — Let $c = 1/4 \cdot 16^3$ and $\varepsilon = 1/128$. By Lemma 2.1, there are at least $e(D)^2/8n$ paths of length two in D . By Lemma 2.2, at most $8\varepsilon e(D)^2/n \leq e(D)^2/16n$ of these paths are unbalanced. Therefore there are at least $e(D)^2/16n$ balanced paths in D . Let C_x denote the number of oriented four-cycles containing the vertex x .

$$C_x = \sum_{u \in V} d^{+-}(x, u)d^{-+}(x, u) \geq \frac{1}{16} \sum_{u: (x, u) \text{ balanced}} d^{+-}(x, u)^2.$$

Summing this quantity over x , and applying the Cauchy-Schwarz inequality, one obtains

$$\sum_x C_x \geq \frac{1}{16} \sum_{x, u: (x, u) \text{ balanced}} d^{+-}(x, u)^2 \geq \frac{1}{16n^2} \left(\sum_{(x, u) \text{ balanced}} d^{+-}(x, u) \right)^2.$$

This sum counts exactly the number of balanced paths of length two and so is at least $e(D)^2/16n$. Thus, $\sum_x C_x \geq e(D)^4/16^3n^4$. The proof is now complete as this sum counts each oriented four-cycle four times. \square

Proof of Lemma 2.2. — Denote by f the number of unbalanced paths of length two, and suppose that $f \geq 8\varepsilon e(D)^2/n$. Let f_x denote the number of unbalanced paths of length two starting at x , and let W denote the set of vertices x for which $f_x \geq f/2n$. Since at most $f/2$ unbalanced paths of length two start at vertices outside of W , we have that at least $f/2$ unbalanced paths start inside W , i.e., $\sum_{x \in W} f_x \geq f/2$. For each vertex $x \in W$, we denote by U_x the set of vertices u for which (x, u) is unbalanced, we have that $e(\Gamma^+(x), U_x) = f_x \geq \varepsilon e(D)^2/n^2$. By (\star) we have that $e(U_x, \Gamma^+(x)) \geq f_x/2$, i.e., $\sum_{u: (x,u) \text{ unbalanced}} d^{++}(x, u) \geq f_x/2$. Thus,

$$\sum_{(x,u) \text{ unbalanced}} d^{++}(x, u) \geq \frac{1}{2} \sum_{x \in W} f_x \geq f/4.$$

Alternatively, denoting by X_u the set of vertices x for which (x, u) is unbalanced,

$$\sum_u e(X_u, \Gamma^+(u)) \geq f/4.$$

Let U denote those vertices u for which $e(X_u, \Gamma^+(u)) \geq f/8n$, and note that

$$\sum_{u \in U} e(X_u, \Gamma^+(u)) \geq f/8.$$

However, for each $u \in U$, we have by (\star) that $e(\Gamma^+(u), X_u) \geq e(X_u, \Gamma^+(u))/2$. And so

$$\sum_{u \in U} e(\Gamma^+(u), X_u) \geq f/16.$$

Reinterpreting this sum in terms of $d^{+-}(x, u)$ and recalling that

$$f = \sum_{(x,u) \text{ unbalanced}} d^{+-}(x, u),$$

we have that

$$\sum_{(x,u) \text{ unbalanced}} d^{+-}(u, x) \geq \frac{1}{16} \sum_{(x,u) \text{ unbalanced}} d^{+-}(x, u).$$

Thus, there exists an unbalanced pair (x, u) with $d^{+-}(u, x) \geq d^{+-}(x, u)/16$, a contradiction. \square

3. Oriented Six-Cycles

In this section we prove Theorem 1.6. Note that this immediately implies Theorem 1.2. First, a lemma showing that there are many paths of length two ending at vertices with out-degree at least $e(D)/8n$.

Lemma 3.1. — *Let D be an oriented graph with $\text{bias}(D) \leq e(D)/8$. Then D contains at least $e(D)^2/8n$ oriented paths of length two whose end point has out-degree at least $e(D)/8n$.*

Proof. — We say that an arc \vec{xy} is *good* if $d^-(x) \geq d^+(x)/2$ and $d^+(y) \geq d^-(y)/2$. It is *very good* if in addition $d^-(y) \geq e(D)/4n$. We first show that at least $e(D)/2$ arcs are very good. Let $Z_1 = \{x : d^+(x) > 2d^-(x)\}$, $Z_2 = \{y : d^-(y) > 2d^+(y)\}$ and $Z_3 = \{y : d^-(y) < e(D)/4n\}$. An arc is very good unless its start point is in Z_1 or its end point is in $Z_2 \cup Z_3$. Since $e(Z_1, V) > 2e(V, Z_1)$ we have from (\star) that $e(Z_1, V) \leq e(D)/8$. Similarly $e(V, Z_2) \leq e(D)/8$. Trivially $e(V, Z_3) < e(D)/4$. Hence, at least $e(D)/2$ arcs are very good. For each vertex $x \in V \setminus Z_1$, let $d_{vg}^+(x)$ denote the number of very good arcs with start point x . Thus, the number of paths of length two the second arc of which is very good, is at least

$$\sum_{x \in V \setminus Z_1} d^-(x) d_{vg}^+(x) \geq \frac{1}{2} \sum_{x \in V \setminus Z_1} d_{vg}^+(x)^2 \geq \frac{1}{2n} \left(\sum_{x \in V \setminus Z_1} d_{vg}^+(x) \right)^2 \geq \frac{e(D)^2}{8n},$$

where the second inequality follows from the Cauchy-Schwarz inequality. The proof is now complete, as a path of length two whose second arc \vec{xy} is very good has that $d^+(y) \geq d^-(y)/2 \geq e(D)/8n$. \square

For our proof of Theorem 1.6 we shall need some new notation. This notation will allow us to prove the existence of the constants $c, \varepsilon > 0$ without explicitly calculating them. Specifically, $\gamma(\varepsilon)$ denotes any decreasing function of ε which is positive for all sufficiently small $\varepsilon > 0$. While $\delta(\varepsilon)$ denotes any increasing function of ε which has limit 0 as $\varepsilon \rightarrow 0$. It should be understood that $\gamma(\varepsilon)$ and $\delta(\varepsilon)$ are not functions, but rather classes of functions, in the sense that the commonly used $O(\cdot)$ notation represents a class of functions rather than an individual function. To give some examples to clarify this notation, we note that for two arbitrary constants $c_1, c_2 > 0$, the function $-c_1\varepsilon + c_2$ is (of class) $\gamma(\varepsilon)$, while $-c_1\varepsilon$ is not $\gamma(\varepsilon)$ since for all non-negative (even sufficiently small) values of ε , $-c_1\varepsilon$ is negative. In the same way, for two arbitrary constants $c_1, c_2 > 0$, the function $c_1\varepsilon$ is (of class) $\delta(\varepsilon)$ while the function $c_1\varepsilon + c_2$ is not $\delta(\varepsilon)$. It should be now clear that $\gamma(\varepsilon) - \delta(\varepsilon) = \gamma(\varepsilon)$ and $\gamma(\varepsilon)/C = \gamma(\varepsilon)$ for any constant C , and also that $\gamma(\varepsilon)\gamma(\varepsilon) = \gamma(\varepsilon)$.

In this terminology, to prove Theorem 1.6 it suffices to show that every oriented graph D with $bias(D) < \varepsilon e(D)^2/n^2$ contains at least $\gamma(\varepsilon)e(D)^6/n^6$ oriented six-cycles.

Let D be such that $bias(D) < \varepsilon e(D)^2/n^2$. We state some consequences of previous results in our new notation. Lemma 3.1 tells us that D contains at least $\gamma(\varepsilon)e(D)^2/n$ paths of length two whose end point has out-degree at least $\gamma(\varepsilon)e(D)/n$. While Lemma 2.2 tells us that D contains at most $\delta(\varepsilon)e(D)^2/n$ unbalanced paths of length two. Together, these results imply that D contains $\gamma(\varepsilon)e(D)^2/n$ paths of length two which are balanced and which have end points of out-degree at least $\gamma(\varepsilon)e(D)/n$.

We are now in position to start our proof.

Proof of Theorem 1.6. — Let D be an oriented graph with $\text{bias}(D) < \varepsilon e(D)^2/n^2$. Say that a path of length two which is balanced and whose end point has out-degree at least $\gamma(\varepsilon)e(D)/n$ is a *great* path. We found above that D contains at least $\gamma(\varepsilon)e(D)^2/n$ great paths. We now form a weighted directed graph H on the same vertex set V as follows. Start with the complete directed graph (i.e., the directed graph with arcs \vec{xy} (and \vec{yx}) for all $x \neq y$), weight each arc \vec{xu} by $w_{\vec{xu}}$ the number of great paths from x to u , and delete arcs of weight zero. By the above result, the total weight of arcs in H is at least $\gamma(\varepsilon)e(D)^2/n$. Now divide the set of arcs of H into two categories of heavy and light arcs: a heavy arc is an arc of weight at least $e(D)/n$ while a light arc has weight less than $e(D)/n$. Since the total weight in H is at least $\gamma(\varepsilon)e(D)^2/n$, either the total weight of heavy arcs or the total weight of the light arcs is at least $\gamma(\varepsilon)e(D)^2/n$. We shall refer to these as Case I and Case II respectively, and denote by H' the subgraph consisting of arcs of appropriate weight (heavy arcs of weight $\geq e(D)/n$ in Case I and light arcs of weight $< e(D)/n$ in Case II), although for the time being the proof continues for the two cases in parallel.

For a weighted oriented graph and a vertex v , the out-weight (resp. in-weight) of v is the total weight of the out-going (resp. in-coming) arcs of v . Let H'' be the subgraph of H' consisting of all the arcs whose start vertex has out-weight at least $\gamma(\varepsilon)e(D)^2/n^2$. Since the total weight in H' is at least $\gamma(\varepsilon)e(D)^2/n$, a simple averaging argument shows that there is a function $\gamma(\varepsilon)$ such that the total weight of arcs of H'' is at least $\gamma(\varepsilon)e(D)^2/n$. For each vertex x with positive out-weight in H'' , we denote by e_x its out-weight; in particular, note that by the definition of H'' , $e_x \geq \gamma(\varepsilon)e(D)^2/n^2$.

Our proof will be based on a local argument concerning the distribution of weights of H'' around each vertex of positive out-weight. To simplify the presentation, fix such a vertex x and define a new (induced) weighting ω^x , or simply ω if there is no risk of confusion, on the vertices by setting ω_u equal to the weight of the arc \vec{xu} in H'' (if \vec{xu} is not an arc of H'' , simply define $\omega_u = 0$). Define the weight of a path of length two \vec{uy}, \vec{yv} in D to be $\omega_u\omega_v$. We note again that the weighting ω depends on the choice of x and is simply induced from the weighting of out-going arcs from x .

Claim. Under this weighting of the vertices, the total weight of paths of length two in D is at least $\gamma(\varepsilon)e_x^2e(D)^2/n^3$.

Before presenting the proof of this Claim, and to keep the continuity of the proof, we show how the Claim allows to finish the proof of Theorem 1.6. We will deduce from the Claim that there are at least $\gamma(\varepsilon)e(D)^6/n^6$ oriented six-cycles in D . We recall that, for a fixed vertex x and in terms of the original weighting of the arcs of H'' , the Claim is discussing the total weight of paths of length two in D where the weight of the path \vec{uy}, \vec{yv} is defined to be $w_{\vec{xu}}w_{\vec{xv}}$ (this now being the weight in H'' ; recall that $\omega_u = w_{\vec{xu}}$). Recall also that,

by the definition of the weighting of the arcs, if $w_{x\bar{v}} > 0$, then there are $w_{x\bar{v}}$ great paths of length two from x to v in D , and in particular this tells us that (x, v) is balanced, and so that in addition there are at least $\gamma(\varepsilon)w_{x\bar{v}}$ paths of length two from v to x . This in turn implies that each path of length two $\vec{u}\vec{y}, \vec{y}\vec{v}$ allows us to find $\gamma(\varepsilon)w_{x\bar{u}}w_{x\bar{v}}$ oriented six-cycles containing x (simply combine each of the $w_{x\bar{u}}$ paths of length two from x to u with the path $\vec{u}\vec{y}, \vec{y}\vec{v}$ followed by each of the $\gamma(\varepsilon)w_{x\bar{v}}$ paths of length two from v to x). Thus, writing C_x for the number of oriented six-cycles in D which contain x and using the Claim, we have that $C_x \geq \gamma(\varepsilon)e_x^2e(D)^2/n^3$. Summing C_x over all vertices x with positive out-weight in H'' , we obtain that $\sum_x C_x \geq \gamma(\varepsilon)\sum_x e_x^2e(D)^2/n^3$. By Cauchy-Schwartz, this is at least $\gamma(\varepsilon)(\sum_x e_x)^2e(D)^2/n^4$, and so, since the sum expresses the total weight of the arcs of H'' , $\sum_x C_x \geq \gamma(\varepsilon)e(D)^6/n^6$. This sum counts each oriented six-cycle at most six times, so the number of oriented six-cycles in D is $\gamma(\varepsilon)e(D)^6/n^6$, and Theorem 1.6 follows.

Thus, we are only left to prove the above Claim.

Proof of Claim. We recall that the weighting of vertices (induced by the fixed vertex x of positive out-weight) is simply defined in such a way that the weight ω_u of a vertex u is the weight $w_{x\bar{u}}$ of the arc $\vec{x}u$. We will divide the proof into two parts, depending on whether we are in Case I or in Case II.

(Recall that in Case I, the subgraph H' consists of all the heavy arcs, i.e., all the arcs of weight $\geq \gamma(\varepsilon)e(D)/n$, while in Case II, H' is the subgraph containing all the light arcs.)

The proof is simpler if we are in Case I. In this case, every vertex with positive weight has weight at least $e(D)/n$, so that every path of length two with positive weight has weight at least $e(D)^2/n^2$. It thus suffices to find at least $\gamma(\varepsilon)e_x^2/n$ paths of length two of positive weight. We denote by U the set of vertices of positive weight, i.e.,

$$U := \{u \in V(D) \mid \omega_u > 0\}.$$

Since x has positive out-weight in H'' , by the definition of H'' , it has out-weight $e_x \geq \gamma(\varepsilon)e(D)^2/n^2$. This implies that $e(\Gamma^+(x), U) = e_x \geq \gamma(\varepsilon)e(D)^2/n^2$ and so, trivially, $e(V, U) \geq e_x \geq \gamma(\varepsilon)e(D)^2/n^2$. Let now Z be the set of all vertices which are ‘‘unbalanced’’ with respect to U , namely, let

$$Z := \{y \mid e(\{y\}, U) > 2e(U, \{y\})\}.$$

Our condition on $bias(D)$ implies that $e(Z, U) < \varepsilon e(D)^2/n^2$. We infer that $e(Y, U) \geq \gamma(\varepsilon)e_x$, where $Y = V \setminus Z$, and so the number of paths of length two from U to U is at least

$$\sum_{y \in Y} e(\{y\}, U)e(U, \{y\}) \geq \frac{1}{2} \sum_{y \in Y} e(\{y\}, U)^2 \geq \frac{1}{2n} \left(\sum_{y \in Y} e(\{y\}, U) \right)^2 \geq \gamma(\varepsilon)e_x^2/n.$$

Which completes the proof in Case I.

For Case II, we divide the vertex set according to weights. For each $i \geq 1$, define

$$V_i := \left\{ u \mid \omega_u \in \left[\frac{e(D)}{2^i n}, \frac{e(D)}{2^{i-1} n} \right) \right\}.$$

Since we are in Case II, all the weights ω_u are less than $e(D)/n$ and thus, $V_+ = \bigsqcup_{i \geq 1} V_i$, where V_+ is the set of all the vertices of positive weight.

The idea in this case is roughly speaking as follows (to make this idea work, we have to restrict the subset of indices to some subset I , see below): To prove the claim, that the total weight of paths of length two in D with respect to the weighting ω is at least $\gamma(\varepsilon)e_x^2 e(D)^2/n^3$, we will only consider a subset of all the paths of length two from V_+ to V_+ , and decompose the set of all such paths into different sorts depending on the starting point of the path being in V_i and the end point of the path being in V_j , for $i, j \in I$ (for a subset I of the index set that we will define below). Let $e_2(V_i, V_j)$ be the number of paths of length two from V_i to V_j . By the definition of V_i , all the weights ω_u are ‘‘almost’’ uniform for $u \in V_i$, namely, $\frac{e(D)}{2^i n} \leq \omega_u < \frac{e(D)}{2^{i-1} n}$. It follows that the total contribution of the paths of length two from V_i to V_j , for $i, j \in I$, is $\Theta\left(\sum_{i, j \in I} \frac{e(D)^2}{2^{i+j-2} n^2} e_2(V_i, V_j)\right)$, and thus, we only need to show that $\sum_{i, j \in I} \frac{e(D)^2}{2^{i+j-2} n^2} e_2(V_i, V_j) \geq \gamma(\varepsilon)e_x^2 e(D)^2/n^3$. This is exactly what we will do in the following.

For each i , let $s_i = |V_i|$. Since by our assumption, x has positive out-weight in H'' , it must have out-weight $e_x \geq \gamma(\varepsilon)e(D)^2/n^2$. This implies that

$$e(\Gamma^+(x), V_+) = e_x \geq \gamma(\varepsilon)e(D)^2/n^2.$$

Given that for each i , ω_u is at most $\frac{e(D)}{2^{i-1} n}$, a simple summation shows that at most $ce(D)^2/n^2$ of these arcs go to sets V_i for which $s_i < ce(D)/n$, this for any constant $c > 0$. We infer, by the definition of the class $\gamma(\varepsilon)$, that for some sufficiently small constant c , we have

$$e(\Gamma^+(x), V_I) \geq \gamma(\varepsilon)e_x,$$

where $I = \left\{ i \mid s_i \geq ce(D)/n \right\}$ and $V_I = \bigcup_{i \in I} V_i$.

Since $2 \sum_{i \in I} s_i \frac{e(D)}{2^i n} \geq e(\Gamma^+(x), V_I)$, it follows that

$$(1) \quad \sum_{i \in I} \frac{s_i e(D)}{2^{i-1} n} \geq \gamma(\varepsilon)e_x.$$

For every vertex $u \in V_+$, the arc $\vec{x}u$ has positive weight in $H'' \subseteq H$, and by the definition of H , we know that u must have out-degree at least $\gamma(\varepsilon)e(D)/n$ in D . Therefore, for each $i \in I$, one has

$$e(V_i, V) = \sum_{u \in V_i} d^+(u) \geq |V_i| \gamma(\varepsilon) e(D)/n \geq c\gamma(\varepsilon)e(D)^2/n^2 = \gamma(\varepsilon)e(D)^2/n^2.$$

Define Y_i to be the set of “balanced” vertices with respect to V_i , namely,

$$Y_i = \left\{ y \mid e(V_i, \{y\}) \leq 2e(\{y\}, V_i) \right\}.$$

Using the definition of *bias*, we may bound by $\varepsilon e(D)^2/n^2$ the number of arcs going to vertices $y \notin Y_i$, deducing that

$$(2) \quad e(V_i, Y_i) \geq \gamma(\varepsilon) s_i e(D)/n \quad (\geq \gamma(\varepsilon) e(D)^2/n^2).$$

Consider the weighted bipartite simple graph F which has parts I and V , has an edge iy whenever $y \in Y_i$, and has a weight η_{iy} on the edge iy given by $\eta_{iy} = \min\{e(\{y\}, V_i), e(V_i, \{y\})\}$. Now for each walk of length two iy, yj in F , we know there are at least $\eta_{iy}\eta_{jy}$ paths of length two from V_i to V_j in D passing through y (there paths are constructed by $e(V_i, \{y\}) \geq \eta_{iy}$ from V_i to y and $e(\{y\}, V_j) \geq \eta_{jy}$ arcs from y to V_j). It follows that $e_2(V_i, V_j) \geq \sum_{y \in V} \eta_{iy}\eta_{jy}$.

Thus, to finish the proof of the claim, given that the weight ω_u of a vertex u in V_i is in the interval $[\frac{e(D)}{2^i n}, \frac{e(D)}{2^{i-1} n})$, and that the total weight of paths of length two from $\bigsqcup_{i \in I} V_i$ to $\bigsqcup_{i \in I} V_i$ is $\Theta\left(\sum_{i,j \in I} \frac{e(D)^2}{2^{i+j-2} n^2} e_2(V_i, V_j)\right)$, it suffices to prove that

$$(3) \quad \sum_{i,j \in I} \sum_{y \in V} \frac{e(D)^2}{2^{i+j-2} n^2} \eta_{iy}\eta_{jy} \geq \gamma(\varepsilon) e_x^2 e(D)^2/n^3.$$

Define the weight of a walk iy, yj in F to be $e(D)^2 \eta_{iy}\eta_{jy}/2^{i+j-2} n^2$. so that we will have to show that the total weight of walks of length two in F is at least $\gamma(\varepsilon) e_x^2 e(D)^2/n^3$. This follows immediately from the more general result (itself a trivial consequence of the Cauchy-Schwarz inequality) that for any graph with weighted vertices and edges, the total weight of walks of length two (where the walk uy, yv is assigned weight $\eta_u \eta_{uy} \eta_{yv} \eta_v$) is at least W^2/n , where W denotes $\sum_u \eta_u \sum_y \eta_{uy}$. In our case, in the bipartite graph F , the edge weights are η_{iy} for $i \in I$ and $y \in Y_i$, and the vertex weights are defined by $\eta_i = \frac{e(D)}{2^{i-1} n}$ for $i \in I$ and $\eta(y) = 0$ for $y \in V$. We have

$$\begin{aligned} W &= \sum_{i \in I} \eta_i \sum_y \eta_{iy} \\ &\geq \sum_{i \in I} \frac{e(D)}{2^{i-1} n} e(V_i, Y_i) && \text{(by the definition of } \eta_{iy}\text{)} \\ &\geq \frac{\gamma(\varepsilon) e(D)}{n} \sum_{i \in I} \frac{s_i e(D)}{2^{i-1} n}, && \text{(by Inequality (2))} \\ &\geq \gamma(\varepsilon) e_x e(D)/n && \text{(by Inequality (1))} \end{aligned}$$

The left term of (3) is at least $W^2/n \geq \gamma(\varepsilon)^2 e_x^2 e(D)^2/n^3 = \gamma(\varepsilon) e_x^2 e(D)^2/n^3$ and Inequality (3) follows. This completes the proof of the Claim (and so the proof of the theorem). \square

4. Examples

To prove that a result such as Theorem 1.1 is best possible up to the choice of constant, it would not suffice to produce just one oriented graph which does not contain an oriented four-cycle and has $bias(D) = Ke(D)^2/n^2$, for some constant K . Nor would it suffice to produce a class of examples that were all of the same density. The theorem applies across a large range of densities, so one must produce examples across a large range of densities. We provide in this section a wide class of examples of oriented graphs which do not contain oriented four-cycles and which have $bias(D) \leq Ke(D)^2/n^2$, where K is some fixed constant.

Our initial examples are obtained as random orientations of four-cycle free simple graphs. These will have approximately $n^{3/2}$ arcs. We will then obtain more dense examples as blow ups of these initial examples. Before we do this, we first prove a lemma concerning the value of $bias(D)$ (and certain variants) for randomly oriented graphs, prove a lemma concerning the value of $bias(D)$ when D is blow up of some other oriented graph, and recall a result concerning four-cycle free simple graphs.

We define a more general concept of *bias*. For $\gamma \in (0, 1)$, we say that a subgraph $E(A, B)$ is γ -biased if $e(B, A) \leq \gamma e(A, B)$, and we write $bias_\gamma(D)$ for the size of the largest γ -biased subgraph of D , so that,

$$bias_\gamma(D) = \max\{e(A, B) : A, B \subseteq V \text{ with } e(B, A) \leq \gamma e(A, B)\}.$$

Note that $bias(D)$ is of course $bias_{1/2}(D)$.

Lemma 4.1. — *Given $\gamma \in (0, 1)$, there exists $K_\gamma \in \mathbb{R}$ such that for every simple graph G on n vertices, there exists an oriented graph D obtained by orienting the edges of G with $bias_\gamma(D) < K_\gamma n$. Furthermore, a random orientation D of G has $bias_\gamma(D) < K_\gamma n$ with high probability.*

Proof. — For a pair $A, B \subseteq V(G)$ we write e_{AB} for the number of edges between A and B in the graph G . Let D be obtained from G by orienting its edges at random. From Chernoff's inequality [5] (see Lemma 6.6) we have

$$\mathbb{P}(e(B, A) \leq \gamma e(A, B)) \leq \exp(-c(\gamma)e_{AB}) \leq \exp(-c(\gamma)K_\gamma n),$$

where $c(\gamma)$ is a positive constant dependent on γ . We set $K_\gamma = 2/c(\gamma)$. We shall prove that $\mathbb{P}(bias_\gamma(D) \geq K_\gamma n) \leq 1.8^{-n}$. This proves the lemma. The event $bias_\gamma(D) \geq K_\gamma n$ can occur only if there is a pair $A, B \subseteq V(G)$ with $e_G(A, B) \geq K_\gamma n$ for which $e(B, A) \leq \gamma e(A, B)$. There are at most 4^n such pairs (A, B) and for each such pair the probability that $e(B, A) \leq \gamma e(A, B)$ is at

most $\exp(-c(\gamma)e_{AB}) \leq \exp(-c(\gamma)K_\gamma n) = \exp(-2n)$. Thus, by the union bound, $\mathbb{P}(\text{bias}_\gamma(D) \geq K_\gamma n) \leq 4^n \exp(-2n) \leq 1.8^{-n}$. \square

If an oriented graph D' contains no large biased subgraphs, then this property carries over, in a weakened form, to a blow-up D of D' . A blow-up of an oriented graph D' is defined as follows.

Definition. — Let D' be an oriented graph on $\{1, \dots, m\}$ and let $l \in \mathbb{N}$. The l -blow-up of D' is the oriented graph D with vertex set $V = V_1 \cup \dots \cup V_m$, where the sets V_i are disjoint and each of cardinality l , and with arc set $E(D) = \cup_{ij \in E(D')} B(V_i, V_j)$, where $B(V_i, V_j)$ represents the complete bipartite oriented graph on $V_i \cup V_j$ with all arcs going from V_i to V_j . Each V_i is called a *cell* of the blow-up.

The key result we need about blow-ups is,

Lemma 4.2. — *If $\text{bias}_{0.9}(D') < f$ and D is an l -blow-up of D' , then $\text{bias}(D) < 16fl^2$.*

Proof. — Let D be an l -blow-up of D' . We suppose that $\text{bias}(D) \geq 16fl^2$ and use this to show that $\text{bias}_{0.9}(D') \geq f$. Our assumption gives us that there exist sets $A, B \subseteq V(D)$ with $e(A, B) \geq 16fl^2$ and $e(B, A) \leq e(A, B)/2$. We use this irregularity between A and B , this bias in the direction from A to B , to find subsets $I, J \subseteq \{1, \dots, m\}$ such that in D' we have $e_{D'}(I, J) \geq f$ and $e_{D'}(J, I) \leq 0.9e(I, J)$, this will prove $\text{bias}_{0.9}(D') \geq f$ and so will complete the proof.

As a warm-up, we first consider the easier case when A and B are both unions of cells of the blow-up. In this case $A = \cup_I V_i$ and $B = \cup_J V_j$. Which gives

$$e(I, J) = \frac{e(A, B)}{l^2} \geq \frac{16fl^2}{l^2} = 16f \geq f,$$

while

$$e(J, I) = \frac{e(B, A)}{l^2} \leq \frac{e(B, A)}{2l^2} = \frac{e(I, J)}{2}.$$

Now for the general case, if there is a relative deficiency in the number of arcs from B to A , we identify the vertices of A responsible for this. Let

$$A' = \left\{ x \in A : e(\{x\}, B) \geq \frac{3}{2}e(B, \{x\}) \right\}.$$

Note that $e(B, A \setminus A') \geq 2e(A \setminus A', B)/3$ so that

$$e(A, B) \geq 2e(B, A) \geq 2e(B, A \setminus A') \geq \frac{4}{3}e(A \setminus A', B).$$

Thus, $e(A \setminus A', B) \leq 3e(A, B)/4$, and so $e(A', B) \geq e(A, B)/4 \geq 4fl^2$. Now, let $I = \{i : V_i \cap A' \neq \emptyset\}$ and let $A'' = \cup_I V_i$. By the homogeneity of parts of the

blow-up, we have for all $x \in A''$ that

$$e(\{x\}, B) \geq \frac{3}{2}e(B, \{x\}).$$

Also $e(A'', B) \geq e(A', B) \geq 4fl^2$. We now begin a similar procedure to find B'' . We let

$$B' = \left\{ y \in B : e(A'', \{y\}) > \frac{10}{9}e(\{y\}, A'') \right\}.$$

This implies that $e(B \setminus B', A'') \geq 9e(A'', B \setminus B')/10$, so that

$$e(A'', B) \geq \frac{3}{2}e(B, A'') \geq \frac{3}{2}e(B \setminus B', A'') \geq \frac{27}{20}e(A'', B \setminus B').$$

Thus, $e(A'', B \setminus B') \leq 20e(A'', B)/27$, and so

$$e(A'', B') \geq \frac{7}{27}e(A'', B) \geq \frac{1}{4}e(A'', B) \geq fl^2.$$

Let $J = \{j : V_j \cap B' \neq \emptyset\}$, and set $B'' = \cup_J V_j$. With a similar argument to that given previously we obtain that $e(A'', B'') \geq e(A'', B') \geq fl^2$ and

$$e(B'', A'') < \frac{9}{10}e(A'', B'').$$

This tells us that in D' we have $e(I, J) \geq fl^2/l^2 = f$, while

$$e(J, I) = \frac{e(B'', A'')}{l^2} < \frac{9e(A'', B'')}{10l^2} = \frac{9e(I, J)}{10}.$$

□

The final piece of information we need before stating our examples concerns the existence of large four-cycle free simple graphs. Let q be a prime power. The Erdős-Rényi graph G [12] has $V(G)$ being the set of points of the finite projective plane $PG(2, q)$ over the field of order q (so that $n = q^2 + q + 1$), and an edge between (x, y, z) and (x', y', z') if and only if $xx' + yy' + zz' = 0$. In fact this implies $e(G) = \frac{1}{2}q(q+1)^2$ and this graph does not contain a four-cycle. So for all n of the form $q^2 + q + 1$ (where q is a prime power), there is a graph G on n vertices with at least $\frac{1}{2}n^{3/2}$ edges which does not contain a four-cycle. We would like a result which holds for all n . We recall that Bertrand's Postulate states that for all $k \geq 2$, there is a prime between k and $2k$, combining this with the above example one may deduce the following.

Lemma 4.3. — *Given $n \geq 2$, there exists a graph G on n vertices with at least $\frac{1}{20}n^{3/2}$ edges which does not contain a four cycle.*

□

We may now state examples of oriented graphs with $bias(D) < Ke^2/n^2$ which do not contain an oriented four-cycle. Our first examples are obtained by considering random orientations of four-cycle free simple graphs. By the above lemma, there exists, for each n , a simple graph G on n vertices which is four-cycle free

and has $e(G) \geq n^{3/2}/20$. Let D be obtained by orienting the edges of G at random. Then D certainly cannot contain an oriented four-cycle and, by Lemma 4.1, with positive probability $\text{bias}(D) \leq K_{1/2}n \leq 400K_{1/2}e(D)^2/n^2$. In particular this gives us for all n an oriented graph on n vertices which does not contain an oriented four-cycle and for which $\text{bias}(D) \leq 400K_{1/2}e(D)^2/n^2$.

Our more general class of examples is obtained by considering blow-ups of the above examples. For a fixed constant K (in fact we take $K = 6400K_{0.9}$), we define for each pair of natural numbers m and l an oriented graph $D_{m,l}$ which contains no oriented four-cycle and has $\text{bias}(D) \leq Ke(D)^2/n^2$. The number of vertices of D will be $n = ml$, while $e(D)$, the number of arcs of D , will be of the order $m^{3/2}l^2$.

Fix a pair of natural numbers m and l . Let G be a four-cycle free simple graph on m vertices with at least $m^{3/2}/20$ edges. Let D' be an oriented graph obtained by orienting G and such that $\text{bias}_{0.9}(D') < K_{0.9}m$, the existence of such an oriented graph being assured by Lemma 4.1. Let $D = D_{m,l}$ be obtained as an l -blow-up of D' . We now have, by Lemma 4.2, that $\text{bias}(D) < 16K_{0.9}ml^2$. It is easily observed that $n = ml$ and $e \geq m^{3/2}l^2/20$, so that $ml^2 < 400e(D)^2/n^2$. Thus, $\text{bias}(D) < 6400K_{0.9}e(D)^2/n^2$, and, by inspection, D does not contain any oriented four-cycle.

As a demonstration of the generality of our class $(D_{m,l})_{m,l \in \mathbb{N}}$ of examples, note that for any pair n_0, e_0 with $e_0 \geq n_0^{3/2}$, there is a choice of m and l such that $D_{m,l}$ has approximately n_0 vertices and approximately e_0 arcs. Simply choose m to be an integer close to $n_0^4/400e_0^2$, choose a four-cycle free graph G with close to $m^{3/2}/20$ edges, and choose l to be close to $400e_0^2/n_0^3$.

5. The case D is dense

In this section we prove Theorem 1.7. In fact we shall prove a more general result, Proposition 5.1, which also counts homomorphic copies of partially oriented graphs. A partially oriented graph H is a graph which may have some of its edges oriented. We write $\vec{e}(H)$ for the number of edges of H that are oriented, e.g. if H is a simple graph then $\vec{e}(H) = 0$ and if H is an oriented graph then $\vec{e}(H) = e(H)$. We also introduce the notation $\bar{e}(A, B)$ for the total number of edges (whatever their orientation) between A and B . Note that if D is an oriented graph with $\text{bias}(D) < \varepsilon n^2$, then in particular the following holds in D :

$$(4) \quad e(B, A) \geq \frac{\bar{e}(A, B)}{3} - \frac{\varepsilon}{3}n^2 \quad \text{for all } A, B \subseteq V.$$

We now turn to Proposition 5.1. This proposition clearly implies Theorem 1.7.

Proposition 5.1. — Let D be an oriented graph on n vertices satisfying (4). Let H be a partially oriented graph on k vertices. Then

$$\text{hom}(H, D) \geq \frac{\text{hom}(\bar{H}, D)}{3^{\bar{e}(H)}} - (1 - 3^{-\bar{e}(H)}) \frac{\varepsilon}{2} n^k.$$

Remark. — To prove that a dense oriented graph with bias at most εn^2 contains H as a subgraph, one may use Szemerédi Regularity Lemma (cf proof in Appendix of the arxiv version of this article (arXiv:0911.3969)). However we believe that the use of Proposition 5.1 with the following direct proof is simpler and illustrates better the use of the bias parameter. Furthermore it gives better constants.

Proof. — We prove the proposition by induction on $\bar{e}(H)$. If $\bar{e}(H) = 0$, then $H = \bar{H}$, and so $\text{hom}(H, D) = \text{hom}(\bar{H}, D)$. For the general case, let H be an oriented graph on $\{1, \dots, k\}$ with $\bar{e}(H) \geq 1$. By relabelling if necessary (which does not affect the homomorphism count) we may assume that $\vec{12}$ is an arc (oriented edge) of H . Let H' be the partially oriented graph obtained by unorienting this edge. We now relate the quantities $\text{hom}(H', D)$ and $\text{hom}(H, D)$. For each $(x_3, \dots, x_k) \in V^{k-2}$, let $\text{Hom}(H', D; \dots, x_3, \dots, x_k)$ denote the set of homomorphisms ϕ of H' into D for which $\phi(i) = x_i$ for all $i = 3, \dots, k$. Similarly define $\text{Hom}(H, D; \dots, x_3, \dots, x_k)$. In fact, it is easy to characterise the homomorphisms $\phi \in \text{Hom}(H', D; \dots, x_3, \dots, x_k)$. A homomorphism $\phi \in \text{Hom}(H', D; \dots, x_3, \dots, x_k)$ must have $\phi(i) = x_i$ for $i = 3, \dots, k$, and must pick values for $\phi(1)$ and $\phi(2)$. Writing x_1 for $\phi(1)$, we know x_1 must join up appropriately to the vertices x_3, \dots, x_k . Specifically

- (i) $x_1 \vec{x}_i$ is an arc of D , for every arc $\vec{1i} : i \geq 3$ in H .
- (ii) $x_i \vec{x}_1$ is an arc of D , for every arc $\vec{i1} : i \geq 3$ in H .
- (iii) $x_1 x_i$ is an edge of \bar{D} , for every edge $1i : i \geq 3$ in \bar{H} .

Equivalently,

$$x_1 \in \bigcap_{i \geq 3: \vec{1i} \in E(H)} \Gamma^-(x_i) \cap \bigcap_{i \geq 3: \vec{i1} \in E(H)} \Gamma^+(x_i) \cap \bigcap_{i \geq 3: 1i \in E(\bar{H})} \Gamma(x_i).$$

We denote this set A . Similarly, writing x_2 for $\phi(2)$, there are similar restrictions on x_2 , which again are equivalent to demanding that x_2 belongs to a certain set, we denote this set B . Since H' has an unoriented edge between 1 and 2, we have a final condition - the condition that $x_1 x_2$ is an edge of \bar{D} . Hence for certain sets A and B , we have a one-to-one correspondence between homomorphisms $\phi \in \text{Hom}(H', D; \dots, x_3, \dots, x_k)$ and edges of \bar{D} between A and B .

Similarly, we may characterise the homomorphisms

$$\phi \in \text{Hom}(H, D; \dots, x_3, \dots, x_k).$$

Again we write x_1 and x_2 for $\phi(1)$ and $\phi(2)$. The restrictions $x_1 \in A$ and $x_2 \in B$ remain. However, on this occasion we require not only that there is some edge between x_1 and x_2 , but that there is an oriented edge from x_1 to x_2 . Thus, there is a one-to-one correspondence between homomorphisms $\phi \in \text{Hom}(H, D; \cdot, \cdot, x_3, \dots, x_k)$ and edges from A to B .

Thus, $|\text{Hom}(H', D; \cdot, \cdot, x_3, \dots, x_k)|$ and $|\text{Hom}(H, D; \cdot, \cdot, x_3, \dots, x_k)|$ are $\bar{e}(A, B)$ and $e(A, B)$ respectively, for some pair of subsets $A, B \subseteq V$. From our condition (4), we obtain that

$$|\text{Hom}(H, D; \cdot, \cdot, x_3, \dots, x_k)| \geq \frac{|\text{Hom}(H', D; \cdot, \cdot, x_3, \dots, x_k)|}{3} - \frac{\varepsilon}{3}n^2.$$

Since $\text{hom}(H, D)$ is the sum over $(x_3, \dots, x_k) \in V^{k-2}$ of

$$|\text{Hom}(H, D; \cdot, \cdot, x_3, \dots, x_k)|,$$

and similarly for $\text{hom}(H', D)$, we have that

$$\text{hom}(H, D) \geq \frac{\text{hom}(H', D)}{3} - \frac{\varepsilon}{3}n^k.$$

Having obtained this relation between $\text{hom}(H, D)$ and $\text{hom}(H', D)$, we require only an application of the induction hypothesis. As $\bar{e}(H') = \bar{e}(H) - 1$, an application of the induction hypothesis to H' gives $\text{hom}(H', D) \geq \text{hom}(\bar{H}, D)/3^{\bar{e}(H)-1} - (1 - 3^{1-\bar{e}(H)})\frac{\varepsilon}{2}n^k$. Combining this with the inequality proved above

$$\begin{aligned} \text{hom}(H, D) &\geq \frac{\text{hom}(\bar{H}, D)}{3^{\bar{e}(H)}} - \frac{1}{3} \left(1 - 3^{1-\bar{e}(H)}\right) \frac{\varepsilon}{2}n^k - \frac{\varepsilon}{3}n^k \\ &= \frac{\text{hom}(\bar{H}, D)}{3^{\bar{e}(H)}} - (1 - 3^{-\bar{e}(H)})\frac{\varepsilon}{2}n^k. \end{aligned}$$

□

6. The extremal problem concerning $\text{bias}(D)$

The main focus of the present article is the question of what structural information on an oriented graph D can be obtained from the knowledge that a certain oriented graph H is not a subgraph of D . More precisely, our results relate to the question:

What is the minimum value of $\text{bias}(D)$ over H -free oriented graphs D with n vertices and e arcs?

In this section we consider the related extremal question:

What is the minimum value of $\text{bias}(D)$ over oriented graphs on n (non-isolated) vertices?

It will be seen from the results we obtain that this is indeed the most natural form of the extremal question.

The definition of $bias(D)$ contains the constant $\frac{1}{2}$. As commented in the introduction, this choice is rather arbitrary, and all our results hold (up to a change of constant) if $\frac{1}{2}$ is replaced by some other constant $\eta \in (0, 1)$. The same is true for the results of this section, but since it is not trivial to deduce the bounds on $bias_\eta$ from those for $bias = bias_{1/2}$, we shall prove our bounds for $bias_\eta$, $\eta \in (0, 1)$. We shall state our lower bounds as lower bounds on the quantity $ow(D) := \max\{e(A, B) : A, B \subset V(D), e(B, A) = 0\}$, the size of the largest one-way subgraph in D . Since, trivially, $bias_\eta(D) \geq ow(D)$, this provides a lower bound on $bias_\eta(D)$, $\eta \in (0, 1)$. Our results are as follows.

Theorem 6.1. — *Every oriented graph D on n (non-isolated) vertices has $ow(D) \geq n/9 \lceil \log_2 n \rceil$.*

Remark. — This result is not difficult to prove, and we remark that it may be easily deduced from Lemma 2 of [13] (in fact the tight version of that lemma, whose proof is sketched after the proof of Lemma 2).

Theorem 6.2. — *For each $\eta \in (0, 1)$ there is a constant K_η such that, for every n , there is an oriented graph D on n (non-isolated) vertices with $bias_\eta(D) \leq K_\eta n / \log_2 n$.*

Remark. — This result is proved by considering a random orientation of an inhomogeneous random graph. The resulting graphs typically contain $\Theta(n^2/(\log n)^2)$ edges. Variants of this example could be produced with $\Theta(n^\alpha)$ edges, for any $\alpha \in (1, 2)$ (although the constant K_η would be dependent on α).

One surprising facet of our results is that a much larger one-way subgraph can be found in the case that D is out-regular (all out-degrees equal), irrespective of the degree: the lower bound $n/4$ follows immediately from the following proposition.

Proposition 6.3. — *Let D be an oriented graph with maximum out-degree Δ^+ , then $ow(D) \geq e(D)/4\Delta^+$.*

For oriented graphs that are close to out-regular, in the sense that Δ^+ is at most a constant multiple of the average out-degree, the above bound is of the same order as the upper bound (Lemma 4.1) on $bias_\eta(D)$ for D whose orientation is random. Simple applications of Chernoff bounds [5] (see Lemma 6.6) imply that a random orientation of the Erdős-Rényi random graph $G(n, p)$ (with $p = \omega(\log n/n)$) will be close to out-regular with high probability. From which we may deduce.

Corollary 6.4. — *Let D be a random orientation of the Erdős-Rényi random graph $G(n, p)$, with $p = \omega(\log n/n)$. Then $ow(D) = \Theta(n)$ with high probability.*

We also note the following cute lower bound, which follows from Lemma 6.3 and the trivial lower bound $ow(D) \geq \Delta^+$.

Proposition 6.5. — *Let D be an oriented graph, then $ow(D) \geq \sqrt{e(D)}/2$.*

Since there are many results in this section, we shall split into subsections. In subsection 6.1 we prove the lower bounds on $ow(D)$, Theorem 6.1 and Proposition 6.3. In subsection 6.2 we prove Theorem 6.2. We finish the section with an algorithmic proof of the lower bound $ow(D) \geq n/4$ for regular oriented graphs D .

We also draw here the reader's attention to the article of Brown, Erdős and Simonovits [4], which considers extremal problems for directed graphs, where parallel arcs in opposite directions are allowed.

Let us state here, the form of Chernoff's Inequality [5] that we shall use throughout the section.

Lemma 6.6 (Chernoff bound for the sum of Poisson trials)

Let $\beta \in [0, 1]$ and let $S_N = X_1 + \dots + X_N$ be a sum of independent random variables X_i taking values in $\{0, 1\}$, with $\mathbb{P}(X_i = 1) = p_i$ for each $i = 1, \dots, N$, so that the expectation of S_N is $\mu = \sum_i p_i$. Then, for all $\lambda \geq \mu$,

$$\mathbb{P}(S_N \geq (1 + \beta)\lambda) \leq \exp\left(\frac{-\beta^2\lambda}{3}\right).$$

In the other direction, for all $\lambda \leq \mu$,

$$\mathbb{P}(S_N \leq (1 - \beta)\lambda) \leq \exp\left(\frac{-\beta^2\lambda}{2}\right).$$

The usual statement of Chernoff's inequality is the $\lambda = \mu$ case of the above statement, a proof of which is given in [19]. The more general statement given here may be deduced from this by an easy monotonicity argument.

6.1. Lower bounds on $ow(D)$. — Throughout the section we shall write $B(A)$ for the set $\{v \in V : e(\{v\}, A) = 0\}$. Since $B(A)$ maximises $e(A, B)$ over the set of all sets B with $e(B, A) = 0$ we have that

$$ow(D) = \max_{A \subseteq V(D)} e(A, B(A)).$$

All we shall need for the following proofs is that, for any probability distribution on subsets A of $V(D)$, we have $ow(D) \geq \mathbb{E}(e(A, B(A)))$.

Proof of Proposition 6.3. — Let D be an oriented graph with maximum out-degree Δ^+ . Let $A \subseteq V$ be a subset selected at random, each vertex included in

A independently with probability p . To prove the proposition it suffices to show that, when p is chosen appropriately,

$$\mathbb{E}(e(A, B(A))) \geq \frac{e(D)}{4\Delta^+}.$$

Expressing $e(A, B(A))$ as $\sum_{\vec{xy} \in E} 1_{x \in A, y \in B(A)}$, exchanging the order of summation and expectation, and using that $\mathbb{E}(1_F) = \mathbb{P}(F)$ for any event F , we obtain

$$\mathbb{E}(e(A, B(A))) = \sum_{\vec{xy} \in E} \mathbb{P}(x \in A, y \in B(A)).$$

Fix an arc $\vec{xy} \in E(D)$. The event $y \in B(A)$ is exactly the event $\bigcap_{v \in \Gamma^+(y)} (v \notin A)$. Thus, this event is independent of the event $x \in A$, and the probability that it occurs is at least $(1-p)^{d^+(y)} \geq (1-p)^{\Delta^+(D)} \geq 1-p\Delta^+(D)$. From which we deduce,

$$\mathbb{E}(e(A, B(A))) \geq e(D)p(1-p\Delta^+).$$

The proof is now completed by taking $p = 1/2\Delta^+$. \square

In the case that D is a regular oriented graph, the above proof shows that $\mathbb{E}(e(A, B(A))) \geq n/4$ where A is selected by including each vertex in A independently with probability $p = 1/2d$ (where d is the regular in- and out-degree). In the general case, there may be a large range of degrees in D , and so there is no obvious choice of p . To prove Theorem 6.1, we first find a patch of vertices with approximately the same degree, and then proceed as above.

Proof of Theorem 6.1. — Define subsets V^+, V^- of the vertex set V by setting

$$V^+ = \{v : d^+(v) \geq d^-(v)\} \quad \text{and} \quad V^- = \{v : d^-(v) \geq d^+(v)\}.$$

Partition V^+ into $V_1^+, \dots, V_{\lceil \log_2 n \rceil}^+$ by defining

$$V_i^+ = \left\{ v \in V^+ : d^+(v) \in [n/2^i, n/2^{i-1}) \right\}.$$

Similarly partition V^- into $V_1^-, \dots, V_{\lfloor \log_2 n \rfloor}^-$. Since these $2\lceil \log_2 n \rceil$ sets cover V , one of them must have size at least $n/2\lceil \log_2 n \rceil$. By reversing the orientation of D if necessary (which has no effect on $ow(D)$), we may assume that $|V_i^-| \geq n/2\lceil \log_2 n \rceil$ for some i . Having chosen this patch of vertices V_i^- , we define $p = 2^i/3n$. We define the random subset A of V by including each vertex in A independently with probability p . Let $B = B(A) \cap V_i^-$. Arguing as in the proof of Proposition 6.3 above, we have, for each $y \in V_i^-$ and each arc \vec{xy} of D , that the events $x \in A, y \in B$ are independent and that $\mathbb{P}(y \in B) = (1-p)^{d^+(y)} \geq (1-p)^{d^-(y)} \geq 1-pd^-(y)$. Thus,

$$\mathbb{E}(e(A, B)) = \sum_{y \in V_i^-} \sum_{\vec{xy} \in E(D)} \mathbb{P}(x \in A, y \in B) \geq \sum_{y \in V_i^-} pd^-(y)(1-pd^-(y)).$$

By the choice of p we have that $pd^-(y) \in [\frac{1}{3}, \frac{2}{3}]$. Since $x(1-x) \geq \frac{2}{9}$ for all $x \in [\frac{1}{3}, \frac{2}{3}]$, we deduce that

$$\mathbb{E}(e(A, B)) \geq \frac{2|V_i^-|}{9} \geq \frac{n}{9\lceil \log_2 n \rceil}.$$

□

6.2. Proof of Theorem 6.2: An oriented graph with no large η -biased subgraph. — Fix $\eta \in (0, 1)$. To prove Theorem 6.2 we must prove, for some constant $K = K_\eta$, that for every $n \in \mathbb{N}$, there is an oriented graph D on n (non-isolated) vertices with $\text{bias}_\eta(D) \leq Kn/\log_2 n$. For notational convenience, we in fact prove the bound $\text{bias}_\eta(D) \leq K\lceil n/l \rceil$ where $l = \lfloor \log_2(n) \rfloor$.

We now define the random oriented graph D .

Let $V := \{1, \dots, n\}$. We partition V into subsets V_1, \dots, V_l of approximately equal size. Let $n = l\lfloor \frac{n}{l} \rfloor + q$, where q is an integers $0 \leq q < l$. We denote $r := \lceil \frac{n}{l} \rceil$. Let V_1, \dots, V_l be a partition of V such that:

$$|V_i| = r \quad \text{for } i = 1, \dots, q \quad \text{and} \quad |V_i| = \lfloor \frac{n}{l} \rfloor \quad \text{for } i = q + 1, \dots, l.$$

Let G be the random simple graph on vertex set V defined by including each edge xy , $x \in V_i, y \in V_j$, in G with probability $1/2^{i+j-1}$, independently of all other edges. Let D be obtained from G by orienting its edges at random.

Basic Properties of D .

- (i) Each arc \vec{xy} , $x \in V_i, y \in V_j, i \neq j$, has probability $1/2^{i+j}$ of being an arc of D .
- (ii) The event $\vec{xy} \in E(D)$ is independent of the event $\vec{uv} \in E(D)$ so long as $\{u, v\} \neq \{x, y\}$. Furthermore, the event $\vec{xy} \in E(D)$ is independent of all events F generated by the events $\vec{uv} \in E(D) : \{u, v\} \neq \{x, y\}$.
- (iii) While there is dependence between the events $\vec{xy} \in E(D)$ and $\vec{yx} \in E(D)$ (e.g., they are mutually exclusive), the bounds $0 \leq \mathbb{P}(\vec{xy} \in E(D) | F) \leq 1/2^{i+j-1}$ hold for any event F generated by the events $\vec{uv} \in E(D) : (u, v) \neq (x, y)$.

Showing that, with high probability, D has no isolated vertex is relatively straightforward, and may be easily checked by the reader.

We introduce now some new notation. We set $\delta = 1 - \eta$, $\gamma = (1 + \eta)/2$, $k = \lceil \log_2 \frac{64}{\delta} \rceil$, and $K = K_\eta = \lceil (32 + 2k)2^8\delta^{-2} \rceil$. An orientation F of an n -cycle has $\text{bias}_\eta(F) \leq n$. This example suffices for all n such that $n \leq Kn/\log_2 n$, i.e., $n \leq 2^K$. Fix $n \geq 2^K$, and let D be the random oriented graph defined above. The proof of Theorem 6.2 is completed by proving that with high probability $\text{bias}_\eta(D) \leq Kr$ (recall that $r = \lceil \frac{n}{l} \rceil$).

Thus, our task is to bound the probability that there is a pair $A, B \subseteq V$ with $e(A, B) > Kr$ and $e(B, A) \leq \eta e(A, B)$. If we consider this event separately for each pair A, B , we end up considering far too many events. So we restrict

our attention to a certain subset of this family of events. We define a family $\mathcal{R} = \mathcal{R}(\eta)$ of pairs (R, R^*) of subsets of V such that each biased set $E(A, B)$, for A and B subsets of V , provide two elements (R, R^*) and (T, T^*) of \mathcal{R} which do not belong to four type of events, concerning pairs of elements of \mathcal{R} . This will then reduce the problem to bounding the probabilities of these four events; we prove then these bounds and conclude the proof of Theorem 6.2.

To define the family \mathcal{R} below, we first need to associate an (increasing with respect to inclusion) function defined on the whole family of subsets of V . For a subset R of V , and $i = 1, \dots, l$, let $R_i := R \cap V_i$, $r_i := |R_i|$, and define $s(R)$ by

$$s(R) := \sum_{i=1}^l \frac{r_i}{2^i}.$$

Note that since each r_i is bounded by $r = \lceil \frac{n}{l} \rceil$, $s(R)$ itself is strictly smaller than r . Notice also that if $R \subseteq A$, then $s(R) \leq s(A)$. For $j = 1, \dots, 2l$, consider the collection of sub-intervals I_j of $[0, r)$ defined by

$$I_j := \left[2^{-j-1}r, 2^{-j+1}r \right).$$

By the choice of l (as the largest integer not exceeding $\log_2 n$), we infer that for every non-empty subset R of V , there is a j such that $s(R) \in I_j$.

For each $j = 1, \dots, 2l$, we denote by \mathcal{R}_j the family of pairs of subsets (R, R^*) such that $R \subseteq \bigcup_{i \leq j+k} V_i$, $R^* = R \cup \bigcup_{i > j+k} V_i$, and $s(R) \in I_j$. Formally,

$$\mathcal{R}_j := \left\{ \left(R, R \cup \bigcup_{i > j+k} V_i \right) : R \subseteq \bigcup_{i=1}^{j+k} V_i \quad \text{and} \quad s(R) \in \left[2^{-j-1}r, 2^{-j+1}r \right) \right\}.$$

The family \mathcal{R} is defined as the union of all the sets \mathcal{R}_j , i.e.,

$$\mathcal{R} := \bigcup_{j=1}^{2l} \mathcal{R}_j.$$

The importance of the family \mathcal{R} defined above is apparent in the following two lemmas, namely that, the pairs $(R, R^*) \in \mathcal{R}$ "cover" every non-empty subset of V , and the size of \mathcal{R} is sufficiently small.

Lemma 6.7. — *For every non-empty subset $A \subseteq V$, there is an element $(R, R^*) \in \mathcal{R}$ such that $R \subseteq A \subseteq R^*$*

Proof. — Let $1 \leq j \leq l$ be such that $s(A) \in [2^{-j}r, 2^{-j+1}r)$. Define $R = \bigcup_{i \leq j+k} (V_i \cap A)$ and $R^* = R \cup \bigcup_{i > j+k} V_i$. Since $\sum_{i \geq j+k+1} \frac{|V_i \cap A|}{2^i} < 2^{-j-k}r$, it follows that $2^{-j-1}r \leq s(R) < 2^{-j+1}r$, i.e., $(R, R^*) \in \mathcal{R}_j \subset \mathcal{R}$, and the claim follows. \square

Lemma 6.8. — $|\mathcal{R}| \leq \exp((k+8)r)$

Proof. — It suffices to prove the bound $|\mathcal{R}_j| \leq \exp((k+7)r)$ for all $1 \leq j \leq 2l$. Fix such a j . We have to prove that there are at most $\exp((k+7)r)$ sets $R \subset \bigcup_{i=1}^{j+k} V_i$ for which $s(R) \in [2^{-j-1}r, 2^{-j+1}r)$. In fact we will only use the upper bound $s(R) < 2^{-j+1}r$ to bound the cardinality of \mathcal{R}_j .

Define $R_i := R \cap V_i$ so that $r_i = |R_i|$. First remark that the bound $s(R) < 2^{-j+1}r$ immediately implies that

$$(5) \quad r_i \leq 2^{i+1-j}r \quad \text{for all } i \leq j+k.$$

What we shall actually prove is that the number of $(j+k)$ -tuples (R_1, \dots, R_{j+k}) with the property that $R_i \subseteq V_i$ and $r_i \leq 2^{i+1-j}r$, is bounded by $\exp((k+7)r)$. For $i = j-1, \dots, j+k$, the inequalities in (5) are automatically verified, i.e., $R_{j-1}, R_j, \dots, R_{j+k}$ can be any subset of $V_{j-1}, V_j, \dots, V_{j+k}$, respectively. There are $2^{(k+2)r} < \exp((k+2)r)$ ways to choose $R_{j-1}, R_j, \dots, R_{j+k}$.

The number of ways to choose the subsets R_1, \dots, R_{j-2} can be bounded as follows. There are at most $r^{j-2} \leq r^l \leq \exp(r)$ ways to choose the sequence of set sizes (r_1, \dots, r_{j-2}) , and for each choice of (r_1, \dots, r_{j-2}) , the number of sets (R_1, \dots, R_{j-2}) with corresponding cardinalities is bounded by the product of binomial coefficients $\binom{|V_i|}{r_i}$ for $i = 1, \dots, j-2$. By well known properties of binomial coefficients, and since $r_i \leq r/2$ by inequalities (5) for $i \leq j-2$, this latter product takes its maximum exactly when each r_i takes its maximum allowed value, i.e., for $r_i = 2^{i+1-j}r$ for each $1 \leq i \leq j-2$. We infer that the number of $(j-2)$ -tuples (R_1, \dots, R_{j-2}) with cardinality constraints above is bounded by

$$\begin{aligned} \exp(r) \prod_{i=1}^{j-2} \binom{r}{2^{i+1-j}r} &\leq \exp(r) \prod_{i=1}^{j-2} \left(e \frac{r}{2^{i+1-j}r} \right)^{2^{i+1-j}r} \\ &\leq \exp(r) \exp\left(\sum_{i=1}^{j-2} 2^{j-1-i}r\right) \exp\left(\sum_{i=1}^{j-2} (i+1-j)2^{j-1-i}r\right) \leq \exp(5r), \end{aligned}$$

and the lemma follows. Note that in the first inequality above, we used the well known bound $\binom{a}{b} \leq (ea/b)^b$ on binomial coefficients (using the fact that $x \log(y/x) \leq \sqrt{xy}$ for all positive reals x, y). \square

Lemma 6.8 shows that the number of pairs $(R, R^*), (T, T^*) \in \mathcal{R}$ is bounded by $\exp((2k+16)r)$. This is the main property we will use in the following. Namely, the four type of events that we are going to define below concern pairs of elements of \mathcal{R} , and have the property that the probability that one of these event happens for a given pair of elements in \mathcal{R} , is bounded by $\exp(-(32+2k)r)$. The bound on the total number of pairs then show that with (high) positive probability (as function of r) non of the four events happen. This will then complete the proof of Theorem 6.2 since as we said, we will show that the event there exists a biased $E(A, B)$ is contained in one of these four events.

In the following, we use the notation $\bar{e}(R, T)$ for the number of edges between R and T in the underlying graph G , so that $\bar{e}(R, T) = e(R, T) + e(T, R)$. We also write $\tilde{e}(R, R^*; T, T^*)$ for $e(R^*, T^*) - e(R, T)$. Consider the following four events concerning a given pair of elements (R, R^*) and (T, T^*) in \mathcal{R} .

F_I: For a given pair of elements $(R, R^*), (T, T^*) \in \mathcal{R}$ with $s(R)s(T) \geq Kr/4$,

$$\bar{e}(R, T) \leq s(R)s(T)/2.$$

F_{II}: For a given pair of elements $(R, R^*), (T, T^*) \in \mathcal{R}$ with $\bar{e}(R, T) \geq Kr/8$,

$$e(T, R) < \gamma e(R, T).$$

F_{III}: For a given pair of elements $(R, R^*), (T, T^*) \in \mathcal{R}$ with $s(R)s(T) \geq Kr/4$,

$$\tilde{e}(R, R^*; T, T^*) \geq \delta s(R)s(T)/4.$$

F_{IV}: For a given pair of elements $(R, R^*), (T, T^*) \in \mathcal{R}$ with $s(R)s(T) \leq Kr/4$,

$$e(R^*, T^*) > Kr.$$

The following lemma shows us that the event of interest to us is contained in the union of the above four type of events for pairs of elements of \mathcal{R} , and so combining the bounds on the probabilities of \mathbf{F}_* with Lemma 6.8 will conclude the proof of Theorem 6.2.

Lemma 6.9. — *The event that there is a pair $A, B \subseteq V$ with $e(A, B) > Kr$ and $e(B, A) \leq \eta e(A, B)$, is contained in the union $\mathbf{F}_I \cup \mathbf{F}_{II} \cup \mathbf{F}_{III} \cup \mathbf{F}_{IV}$ over all pair of elements $(R, R^*), (T, T^*)$ of \mathcal{R} .*

Proof. — Suppose none of the events $\mathbf{F}_I, \mathbf{F}_{II}, \mathbf{F}_{III}, \mathbf{F}_{IV}$ occur for any pair $(R, R^*), (T, T^*)$ in \mathcal{R} . We have to show that there is no pair $A, B \subseteq V$ with $e(A, B) > Kr$ and $e(B, A) \leq \eta e(A, B)$. For the sake of a contradiction, suppose $A, B \subseteq V$ be such a (biased) pair. Applying Lemma 6.7, let (R, R^*) and (T, T^*) be two elements of \mathcal{R} such that $R \subseteq A \subseteq R^*$ and $T \subseteq B \subseteq T^*$. We will consider the four events \mathbf{F}_* above for the pair $(R, R^*), (T, T^*)$, and the pair $(T, T^*), (R, R^*)$, and derive a contradiction.

- If $s(R)s(T) \leq Kr/4$, then, since \mathbf{F}_{IV} does not occur, we should have $e(A, B) \leq e(R^*, T^*) \leq Kr$, contradiction.
- If, on the other hand, $s(R)s(T) \geq Kr/4$, then, since none of $\mathbf{F}_I, \mathbf{F}_{II}, \mathbf{F}_{III}$ occur for the pair $(R, R^*), (T, T^*)$, and the pair $(T, T^*), (R, R^*)$, we have that $\bar{e}(R, T) > s(R)s(T)/2$, that $e(T, R) \geq \gamma e(R, T)$, that $e(R, T) \geq \gamma e(T, R)$, and that $\tilde{e}(R, R^*; T, T^*) < \delta s(R)s(T)/4$.

It follows that

$$e(B, A) \geq e(T, R) \geq \frac{1+\eta}{2} e(R, T) = \frac{1+\eta}{2} \left(e(R^*, T^*) - \tilde{e}(R, R^*; T, T^*) \right).$$

Note that $\bar{e}(R, T) = e(R, T) + e(T, R) \leq (1 + \frac{1}{\gamma})e(R, T)$. We infer that $e(R, T) \geq \frac{\gamma}{\gamma+1} \bar{e}(R, T) \geq e(R, T)/2 \geq e(R)e(S)/4$. From the bounds $e(R^*, T^*) \geq$

$e(R, T) \geq s(R)s(T)/4$ and $\tilde{e}(R, R^*; T, T^*) \leq \delta s(R)s(T)/4$, we deduce that

$$\tilde{e}(R, R^*; T, T^*) \leq \delta e(R^*, T^*)/2,$$

and so

$$e(B, A) \geq \frac{1+\eta}{2} \left(1 - \frac{\delta}{2}\right) e(R^*, T^*) > \eta e(R^*, T^*) \geq \eta e(A, B),$$

yielding again to a contradiction. \square

So to complete the proof of Theorem 6.2, by the choice of K , we are only left to prove the following lemmas.

Lemma 6.10 (Bounding $\mathbb{P}(\mathbf{F}_I)$). — For any pair $(R, R^*), (T, T^*) \in \mathcal{R}$ with $s(R)s(T) \geq Kr/4$,

$$\mathbb{P}\left(\bar{e}(R, T) \leq s(R)s(T)/2\right) \leq \exp(-Kr/32).$$

Lemma 6.11 (Bounding $\mathbb{P}(\mathbf{F}_{II})$). — For any pair $(R, R^*), (T, T^*) \in \mathcal{R}$ with $\bar{e}(R, T) \geq Kr/8$,

$$\mathbb{P}\left(e(T, R) < \gamma e(R, T)\right) \leq \exp(-Kr\delta^2/128).$$

Lemma 6.12 (Bounding $\mathbb{P}(\mathbf{F}_{III})$). — For any pair $(R, R^*), (T, T^*) \in \mathcal{R}$ with $s(R)s(T) \geq Kr/4$,

$$\mathbb{P}\left(\tilde{e}(R, R^*; T, T^*) \geq \delta s(R)s(T)/4\right) \leq \exp\left(\frac{-\delta^2 Kr}{256}\right).$$

Lemma 6.13 (Bounding $\mathbb{P}(\mathbf{F}_{IV})$). — For any pair $(R, R^*), (T, T^*) \in \mathcal{R}$ with $s(R)s(T) \leq Kr/4$,

$$\mathbb{P}\left(e(R, T) > Kr\right) \leq \exp(-Kr/3).$$

Proof of Lemma 6.10. — Recall that G is the random simple graph underlying D . The random variable $\bar{e}(R, T)$ may be expressed as $\sum_{\{x,y\}:x \in R, y \in T} 1_{\{x,y\} \in E(G)}$, a sum of independent $\{0, 1\}$ -valued random variables. Its mean may be bounded below by

$$\mathbb{E}(\bar{e}(R, T)) \geq \sum_{i=1}^l \sum_{j=1}^l \frac{r_i t_j}{2^{i+j}} = s(R)s(T) \geq \frac{Kr}{4}.$$

And so, by Chernoff's inequality, Lemma 6.6, we have

$$\mathbb{P}\left(\bar{e}(R, T) \leq s(R)t(R)/2\right) \leq \exp(-s(R)s(T)/8) \leq \exp(-Kr/32).$$

\square

Proof of Lemma 6.11. — Let $(R, R^*), (T, T^*) \in \mathcal{R}$ be such that $\bar{e}(R, T) \geq Kr/8$. Let $R' = R \setminus T$, let $T' = T \setminus R$, and let $S = R \cap T$. We may express $e(R, T)$ as $e(R', T') + e(R', S) + e(S, S) + e(S, T')$, and $e(T, R)$ as $e(T', R') + e(T', S) + e(S, S) + e(S, R')$. Let us introduce the notation $e'(R, T) = e(R', T') + e(R', S) + e(S, T') =$

$e(R, T) - e(S, S)$, and similarly $e'(T, R)$. Let $\bar{e}'(R, T) = e'(R, T) + e'(T, R)$. A few lines of calculation show that the event that $e(T, R) \leq \gamma e(R, T)$ is exactly the event that $(1 + \gamma)e'(T, R) \leq \gamma \bar{e}'(R, T) - (1 - \gamma)e(S, S)$. The mean of $e'(T, R)$ is $\mu = \bar{e}'(R, T)/2$, and so the above inequality becomes

$$e'(T, R) \leq \mu \left(1 - \frac{1 - \gamma}{1 + \gamma} - \frac{(1 - \gamma)e(S, S)}{\mu} \right).$$

Setting $\beta = (1 - \gamma)/(1 + \gamma) + (1 - \gamma)e(S, S)/\mu$, an easy calculation shows that $\beta\mu \geq (1 - \gamma)\bar{e}'(R, T)/(1 + \gamma)$, and so, from Chernoff's inequality, we deduce that

$$\begin{aligned} \mathbb{P} \left(e(T, R) \leq \gamma e(R, T) \right) &\leq \exp \left(- \left(\frac{1 - \gamma}{1 + \gamma} \right)^2 \frac{\bar{e}'(R, T)}{2} \right) \leq \exp \left(\frac{-Kr\delta^2}{64(1 + \gamma)^2} \right) \\ &\leq \exp \left(\frac{-Kr\delta^2}{128} \right). \end{aligned}$$

□

Proof of Lemma 6.12. — Let $(R, R^*), (T, T^*) \in \mathcal{R}$ be such that $s(R)s(T) \geq Kr/4$. The definition of \mathcal{R} implies that $s(R^* \setminus R) \leq \delta s(R)/32$, likewise $s(T^* \setminus T) \leq \delta s(T)/32$.

(Note that if $(R, R^*) \in \mathcal{R}_j$, then $s(R^* \setminus R) \leq 2^{-j-k}r = 2^{-j-1}r\delta/32 \leq s(R)\delta/32$.)

We let \tilde{P} be the set of pairs with one end point R^* and the other in T^* which do not have one end in R and the other in T . We may bound the expected value of $\tilde{e}(R, R^*; T, T^*)$ by

$$\sum_{\{u, v\} \in \tilde{P}} \mathbb{P}(1_{\{u, v\} \in E(G)}) \leq \sum_{i=1}^l \sum_{i'=1}^l \frac{r_i t_{i'}^* + r_i^* t_{i'} + r_i^* t_{i'}}{2^{i+i'-1}}.$$

The latter sum has value $2(s(R)s(T^*) + s(R^*)s(T) + s(R^*)s(T^*)) \leq (\delta/8 + \delta^2/512)s(R)s(T)$, and so $\mathbb{E}(\tilde{e}(R, R^*; T, T^*)) \leq 3\delta s(R)s(T)/8$. Applying Chernoff's inequality, Lemma 6.6, with $\lambda = 3\delta s(R)s(T)/16$ and $\beta = 1/2$ we obtain the bound

$$\mathbb{P} \left(\tilde{e}(R, R^*; T, T^*) \geq \frac{\delta s(R)s(T)}{4} \right) \leq \exp \left(\frac{-\delta^2 s(R)s(T)}{64} \right) \leq \exp \left(\frac{-\delta^2 Kr}{256} \right).$$

□

Proof of Lemma 6.13. — Let $(R, R^*), (T, T^*) \in \mathcal{R}$ be such that $s(R)s(T) \leq Kr/4$. An upper bound for $e(R, T)$ is given by $\sum_{\{u, v\}: u \in R, v \in T} 1_{\{u, v\} \in E(G)}$. This is a sum of independent $\{0, 1\}$ -valued random variables, and so, by Chernoff's inequality (Lemma 6.6), it suffices to prove that its expected value is at most $Kr/2$. This is easily proved,

$$\sum_{u \in R, v \in T} \mathbb{P}(1_{\{u, v\} \in E(G)}) \leq \sum_{i=1}^l \sum_{i'=1}^l \frac{r_i t_{i'}}{2^{i+i'-1}} = 2s(R)s(T) \leq \frac{Kr}{2}.$$

□

6.3. Algorithmic approach. — Proposition 6.3 implies that in every regular oriented graph D there are subsets $A, B \subseteq V$ with $e(A, B) \geq n/4$ and $e(B, A) = 0$. We now give a polynomial time algorithm which can find a pair $A, B \subseteq V$ with $e(A, B) \geq n/4$ and $e(B, A) = 0$.

Our algorithm will begin by building a sequence of sets A_1, \dots, A_n for which $e^{(2)}(A)$ is not too large. It then selects a set to be A and from that defines B .

```

1: Let  $t = 1$ , let  $v$  be an arbitrary vertex and let  $A_1 = \{v\}$ .
2: while  $t < n$  do
3:   Select a vertex  $u \in V \setminus A_t$  minimising  $e^{(2)}(A_t, \{u\})$ .
4:   Set  $A_{t+1} = A_t \cup \{u\}$ .
5:   Increase  $t$  by one.
6: end while
7: Let  $t = \lfloor n/2d \rfloor$ 
8: Let  $A = A_t$ 
9: Let  $B = \{v : e(\{v\}, A) = 0\}$ 
10: return  $A, B$ 

```

We now analyse the algorithm.

Lemma 6.14. — For each $t = 1, \dots, n-1$ the set A_t satisfies $e^{(2)}(A_t) \leq d^2(t^2 - 1)/n$.

Proof. — This is clearly true when $t = 1$. Suppose the bound fails: let t be minimal such that $e^{(2)}(A_{t+1}) > d^2((t+1)^2 - 1)/n$. Note that, by the definition of A_{t+1} , one has that

$$e^{(2)}(A_{t+1}) = e^{(2)}(A_t) + \min_{u \in V \setminus A_t} e^{(2)}(A_t, \{u\}).$$

Combining this with the observation that the sum of $e^{(2)}(A_t, \{u\})$ over vertices $u \in V \setminus A_t$ is precisely $e^{(2)}(A_t, V \setminus A_t) = 2d^2t - 2e^{(2)}(A_t)$, we obtain that

$$2d^2t - 2e^{(2)}(A_t) > (n-t) \left(e^{(2)}(A_{t+1}) - e^{(2)}(A_t) \right).$$

Straightforward calculation, and the use of the bounds we are assuming on $e^{(2)}(A_t)$ and $e^{(2)}(A_{t+1})$, yield that $t > n - 2$. Hence there cannot exist $t \in \{1, \dots, n-1\}$ for which the bound fails. \square

Let us define a quadratic equation $f(t) = dt - d^2(t^2 - 1)/n$. This quadratic obtains its maximum at $n/2d$.

Proposition 6.15. — Given a d -regular oriented graph D . The above polynomial time algorithm finds subsets $A, B \subseteq V$ satisfying $e(A, B) \geq n/4$ and $e(B, A) = 0$.

Proof. — Let $t = \lfloor n/2d \rfloor$. The algorithm outputs $A = A_t$ and $B = \{v : e(\{v\}, A) = 0\}$. It is immediate from the definition of B that $e(B, A) = 0$. We now prove the bound on $e(A, B)$. Let us note that each arc from A to $V \setminus B$ is the first arc of a path of length two from A to A . Thus $e(A, V \setminus B) \leq e^{(2)}(A)$. And so $e(A, B) = e(A, V) - e(A, V \setminus B) \geq d|A| - e^{(2)}(A)$. We now use the fact that A is obtained as A_t . This implies that $|A| = t$, and, by Lemma 6.14, that $e^{(2)}(A) \leq d^2(t^2 - 1)/n$. Substituting these values we obtain that $e(A, B) \geq f(t)$. Since f is a quadratic with its maximum at $n/2d$ its value in the range $(n/2d - 1, n/2d]$ is always at least $f(n/2d - 1)$. A simple calculation shows that $f(n/2d - 1) = n/4$, completing the proof. \square

7. Concluding Remarks

We would very much like to see a proof of Conjecture 1.4. By Theorem 1.1, the $k = 2$ case of this Conjecture is settled. Theorem 1.2 proves that every oriented graph with $\text{bias}(D) < \varepsilon e(D)^2/n^2$ contains an oriented six-cycle. This is weaker than the $k = 3$ case of the conjecture, which requires that every oriented graph with $\text{bias}(D) < \varepsilon e(D)^{3/2}/n$ contains an oriented six-cycle. If this result does hold then it is best possible. One can prove this by considering a random orientation of the largest (with respect to number of edges) simple graph (on n vertices) with girth greater than six, further examples being obtained as blow-ups of these first examples. The fact that these examples have $\text{bias}(D) = O(e(D)^{3/2}/n)$ relies on the fact that there is a constant $c > 0$ such that for all n , there is a simple graph G on n vertices which has girth greater than six and such that $e(G) \geq cn^{4/3}$. Such graphs may be constructed using the rank two geometries introduced by Tits [22].

Bondy and Simonovits [3] proved that a graph with girth greater than $2k$ has $O(n^{1+1/k})$ edges. It is widely believed that this result is best possible, i.e., it is believed that for each $k \geq 2$, there is a constant $c_k > 0$ such that for all n , there is a graph on n vertices with girth greater than $2k$ and with $e(G) \geq c_k n^{1+1/k}$. This is known to be true for $k = 2, 3, 5$. If it is true for all k , then by considering blow-ups of random orientations of graphs with girth greater than $2k$, we can find oriented graphs with $\text{bias}(D) = O(e(D)^{k/(k-1)}/n^{2/(k-1)})$ which contain no oriented $2k$ -cycles. In this case Conjecture 1.4, if true, is best possible.

Our final remark on Conjecture 1.4 is that it seems unlikely that a straightforward adaptation of our approach will give a proof. Our proofs focus on what happens locally to a given vertex. We examine the arcs from the out-neighbourhood of a vertex to the second out-neighbourhood. However, there are oriented graphs in which each component contains only $e(D)^2/n^2$ arcs, and so a condition of the form $\text{bias}(D) < \varepsilon e(D)^{k/(k-1)}/n^{2/(k-1)}$ (for $k \geq 3$) gives no information about a

particular component. Thus, any successful approach to Conjecture 1.4 must go beyond the aggregation of certain locally observed inequalities.

Up to this point we have never explicitly put any condition on the underlying graph. What happens if we do? This may change the problem considerably. Certainly the examples we have used up to this point have a certain structure to their underlying graph as well as to the orientation. One particular case that may be of interest is the case in which the underlying graph is random. The most general question that arises in this context is the following.

Question 7.1. — *Let H be an oriented graph and let $p = p(n)$ be some function of n (e.g. $p(n) = n^{1/2}$). Then for which function $b(n)$ do we have the following: with high probability, every orientation D of the Erdős-Rényi random graph $G(n, p)$ satisfying $\text{bias}(D) < b(n)$ contains a copy of H ?*

The two functions $b(n)$ and $p(n)$ are certainly correlated. In the extreme case, when $p(n)$ goes to 0 sufficiently slowly that there are (whp) many copies of H in $G(n, p)$, one may ask, similarly to the dense case, whether a condition as weak as $\text{bias}(D) < \varepsilon p n^2 \sim \varepsilon e(G(n, p))$ suffices to ensure that (whp) D contains a copy of H ? Let \bar{H} be the underlying unoriented graph of H with n_H vertices and e_H edges, and with upper density δ_H : this is by definition the maximum density of a subgraph of H , i.e., $\delta_H := \max_{U \subseteq V(H)} \frac{e(H[U])}{|U|}$. By the result of Erdős and Rényi [11] (for the balanced case) and Bollobás [2] (for the general case), there exists a threshold at $p_H = n^{-\frac{1}{\delta_H}}$ for the event that $G(n, p)$ contains a copy of H . Also, large deviation and concentration results for the number of homomorphic copies of H in $G(n, p)$ around $n^{n_H} p^{e_H}$ are known (e.g., see [20, 16, 26, 17]). This is roughly the expected number of copies of H in $G(n, p)$ modulo a constant factor imposed by the automorphisms of H . It is easy to see that, in order to have a result ensuring the existence of H as an oriented subgraph for an orientation of a random graph $G(n, p)$ with a condition as weak as $\text{bias}(D) < \varepsilon p n^2$, we should have $n^{n_H} p^{e_H} = \Omega(p n^2)$. So, based on some related questions on quasi-random unoriented graphs, it is reasonable to conjecture that the following should be true.

Conjecture 7.2. — *Let $\bar{\delta}_H = \max_{U \subseteq V(H), |U| \geq 3} \frac{e(H[U]) - 1}{|U| - 2}$ and $p(n) \gg n^{-\frac{1}{\bar{\delta}_H}}$. Then with high probability the following holds: an orientation D of $G(n, p)$ has an oriented copy of H as subgraph provided that $\text{bias}(D) \leq \varepsilon p n^2$.*

Indeed, the above conjecture follows from a conjectural version of the embedding lemma in the sparse case (c.f., conjecture 8 in [18] and conjecture 3.11 of [14]). This can be seen by mimicking the proof given in [14] of a related conjecture on the values of $\text{ex}(G(n, p), H)$ in the case $p n^{\frac{1}{\delta_H}} \rightarrow \infty$ assuming the sparse version of the embedding lemma (c.f., Conjecture 3.15 and Theorem 3.16 in [14]).

We now discuss other parameters that may be of interest. Can one ensure the presence of an oriented four-cycle by putting an upper bound on $\max\{e(A, B) - e(B, A) : A, B \subseteq V\}$? The answer is of course “Yes”, as it follows from Theorem 1.1 that an upper bound of the form $\varepsilon e(D)^2/n^2$ suffices. Surely this is not best possible. What is the largest upper bound that suffices? Another parameter that one might consider is the size of the largest eigenvalue. What is the largest upper bound on the largest eigenvalue that can ensure the presence of an oriented four-cycle? Finally, another parameter that may be of interest is the size of the largest one-way subgraph (c.f. Section 6): let $ow(D) = \max\{e(A, B) : A, B \subseteq V, e(B, A) = 0\}$. Can an upper bound on $ow(D)$ ensure the presence of an oriented four-cycle in D ? It is quite possible that a bound of the order of $e(D)^2/n^2$ suffices.

Conjecture 7.3. — *There exists a constant $\varepsilon > 0$ such that every oriented graph D with $ow(D) < \varepsilon e(D)^2/n^2$ contains an oriented four-cycle.*

Acknowledgment. The starting point for the research presented in this paper was a problem related to the Cacceta-Häggkvist conjecture which was asked by Pierre Charbit, and was communicated to the authors by Stéphan Thomassé. The authors thank both for fruitful discussions. Part of this research was done while O. A. was PhD student at École Polytechnique and Mascotte Project, S. G. was PhD student at Cambridge University, and F. H. was PhD student in Mascotte joint project at CNRS-UNSA-INRIA. The authors thank all these institutions for providing the possibility for several visits and for their hospitality.

References

- [1] N. Alon and A. Shapira, *Testing Subgraphs in Directed Graphs*, J. Comput. System Sci. **69**(3), 354–382, 2004.
- [2] B. Bollobás, *Threshold functions for small subgraphs*. Math. Proc. Cambridge Philos. Soc. **90**(2), 197–206, 1981.
- [3] J. A. Bondy and M. Simonovits, *Cycles of Even Length in Graphs*, J. Combin. Theory Ser. B **16**, 97–105, 1974.
- [4] W. G. Brown, P. Erdős and M. Simonovits, *Extremal Problems for Directed Graphs*, J. Combin. Theory Ser. B **15**, 77–93, 1973.
- [5] H. Chernoff, *A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations*, Ann. Math. Statistics **23**, 493–507, 1952.
- [6] F. R. K. Chung and R. L. Graham, *Quasi-Random Tournaments*, J. Graph Theory **15** (2), 173–198, 1991.
- [7] F. R. K. Chung, R. L. Graham and R. M. Wilson, *Quasi-Random Graphs*, Combinatorica **9**(4), 345–362, 1989.
- [8] R. Diestel, *Graph Theory*, 3rd edition. Springer-Verlag, Berlin, 2005.

- [9] P. Erdős and A. Hajnal, *On spanned subgraphs of graphs*, Beitrage zur Graphentheorie und deren Anwendungen, Kolloq. Oberhof (DDR), 80–96, 1977.
- [10] P. Erdős and A. Hajnal, *Ramsey-type theorems*, Combinatorics and complexity (Chicago, IL, 1987), Discrete Appl. Math. **25** (1–2), 37–52, 1989.
- [11] P. Erdős and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17–61, 1960.
- [12] P. Erdős and A. Rényi and V.T. Sós, *On a Problem of Graph Theory*, Studia Sci. Math. Hungar. **1**, 215–235, 1966.
- [13] J. Fox, P. Loh and B. Sudakov, *Large induced trees in K_r -free graphs*, J. Comb. Th. Ser. B **99**, 494–501, 2009.
- [14] S. Gerke and A. Steger, *The sparse regularity lemma and its applications*, Surveys in Combinatorics, London Math. Soc. Lecture Note Ser. **327**, 227–258, 2005.
- [15] S. Griffiths, *Quasirandom Oriented Graphs*, preprint.
- [16] S. Janson, T. Luczak and A. Ruciński, *An exponential bound on the nonexistence of a specified subgraph in a random graph*, Random graphs '87 (Poznań, 1987), 73–87, Wiley, Chichester, 1990.
- [17] S. Janson, T. Luczak and A. Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **142**(1), 61–92, 2004.
- [18] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Found. Comput. Math., 216–230, 1997.
- [19] M. Mitzenmacher and E. Upfal, *Probability and Computing*, CUP, 2005.
- [20] A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Probab. Theory Related Fields **78**, 1–10, 1988.
- [21] P. Turán, *On an extremal problem in graph theory* (in Hungarian), Mat Fiz Lapok **48**, 436–445, 1941.
- [22] J. Tits, *Théorème de Bruhat et sous-groupes paraboliques*, C. R. Math. Acad. Sci. Paris **254**, 2910–2912, 1962.
- [23] A. Thomason, *Pseudo-random graphs*, Ann. Discrete Math. **33**, 307–331, 1987.
- [24] A. Thomason, *Random graphs, strongly regular graphs and pseudo-random graphs*, Surveys in Combinatorics, London Math. Soc. Lecture Note Ser. **123**, 173–195, 1987.
- [25] E. Szemerédi, *On Sets of Integers Containing no k Elements in Arithmetic Progression*, Acta Arith. **27**, 199–245, 1975.
- [26] V. Vu, *A Large Deviation Result on the Number of Small Subgraphs of a Random Graph*, Combin. Probab. Comput. **10**(1), 74–94, 2001.

Appendix A

Dense Case Using Szemerédi's Regularity Lemma

In this appendix, we prove the following theorem.

Theorem A.1. — *Let $h \in \mathbb{N}$ be a fixed integer and H be an oriented graph on h vertices. Then for every $\nu \in (0, 1)$, there exists real numbers $\epsilon > 0$, $0 < \beta = \beta(H) < \frac{1}{2}$, and an integer N such that every oriented graph D on $n \geq N$ vertices, with βn^2 arcs with bias $\nu(D) \leq \epsilon\epsilon$, contains H as a subgraph.*

When H is an orientation of a bipartite graph, β can be any arbitrary small positive value.

The proof of Theorem A.1 is based on the use of Szemerédi's regularity lemma, so we need first to state it. Let us introduce the following notation: Let U and W be two disjoint subsets of the vertex set V of a given (oriented) graph D . The number of edges between U and W without taking into account the orientation, if any, is denoted by $\bar{e}(U, W)$. Recall that the *density* of the pair U, W is

$$d(U, W) = \frac{\bar{e}(U, W)}{|U||W|}.$$

Definition (Uniform Pairs). — *Let $\sigma \in (0, 1)$ be a real number. The pair (U, W) is called σ -uniform if, for all subsets $U' \subset U$ and $W' \subset W$ with $|U'| > \sigma|U|$ and $|W'| > \sigma|W|$, we have $|d(U', W') - d(U, W)| < \sigma$.*

A simple consequence of the σ -uniformity condition above is that bipartite subgraphs induced by uniform pairs must be roughly regular. More precisely we have the following lemma (the proof of which can be found for example in [8]).

Lemma A.2. — *Let (U, W) be a σ -uniform pair and let $d(U, W) = d$. Then we have*

$$|\{u \in U : |\Gamma(u) \cap W| > (d - \sigma)|W|\}| \geq (1 - \sigma)|U|$$

and similarly

$$|\{u \in U : |\Gamma(u) \cap W| < (d + \sigma)|W|\}| \geq (1 - \sigma)|U|.$$

Definition (Equipartitions). — *An equipartition of $V(G)$ into m parts is a partition V_1, \dots, V_m such that $\lfloor \frac{n}{m} \rfloor \leq |V_i| \leq \lceil \frac{n}{m} \rceil$, for $1 \leq i \leq m$, where $n = |V(G)|$. The partition is σ -uniform if (V_i, V_j) is σ -uniform for all but $\sigma \binom{m}{2}$ pairs $1 \leq i < j \leq m$.*

Lemma A.3 (Szemerédi Regularity Lemma [25]). — *Let σ be a real number in $(0, 1)$ and let l and N be two natural numbers. Then there exists $L = L(l, \sigma)$ such that every graph with at least N vertices has a σ -uniform equipartition into m parts for some $l \leq m \leq L$.*

We are now ready to prove the following directed version of the embedding lemma in the presence of an appropriate condition on the bias.

Lemma A.4. — *Let h and $m \in \mathbb{N}$ be two fixed integers with $m \geq h$, H be an oriented graph on h vertices, and ν a real number in $(0, 1)$. Let D be a digraph with e arcs, containing disjoint subsets V_1, \dots, V_h of size u or $u + 1$ of the vertex set V such that*

- (i) $u \geq \frac{n}{m}$;
- (ii) For each arc $(v_i, v_j) \in E(H)$, (V_i, V_j) is σ -uniform;
- (iii) For each $i \neq j$ with $(i, j) \in E(H)$, $d(V_i, V_j) \geq \eta$ for some positive real number $\eta > 0$;
- (iv) For each $i \neq j$ with $(i, j) \in E(H)$, $\text{bias}_\nu(D(V_i \cup V_j)) \leq \epsilon e$, where $\epsilon = \frac{\sigma^2(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)}{m^2}$.

Furthermore, suppose that $(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)^h \geq (h + 1)\sigma$, and finally $\sigma u \geq 1$. Then D contains an oriented subgraph isomorphic to H .

Proof. — Let $\{v_1, \dots, v_h\}$ be the set of vertices of H . We prove by induction on h that the following property holds for all $l = 0, \dots, h$:

Property $\mathcal{P}(l)$: There exists a sequence x_1, \dots, x_l so that

- I. for all $i = 1, \dots, l$, $x_i \in V_i$; and
- II. for every j , $l < j \leq h$, there is a set $X_j^l \subset V_j$ of potential candidates for x_j at stage l of size larger than $(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)^{N(j,l)}u$. More precisely, if we set $N^-(j, l) := \{x_i \mid 1 \leq i \leq l \text{ and } v_i v_j \in E(H)\}$, $N^+(j, l) := \{x_i \mid 1 \leq i \leq l \text{ and } v_j v_i \in E(H)\}$, and $N(j, l) = N^-(j, l) \cup N^+(j, l)$, then for any $y_j \in X_j^l$ and for every $x_i \in N(j, l)$, $x_i y_j \in E(D)$ or $y_j x_i \in E(D)$ provided that $x_i \in N^-(j, l)$ or $x_i \in N^+(j, l)$ respectively, and moreover, $|X_j^l| \geq (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)^{N(j,l)}u$.

It is clear that the property $\mathcal{P}(0)$ holds, for this we just take $X_j^0 = V_j$.

Suppose now that $\mathcal{P}(l)$ is true, we will prove that $\mathcal{P}(l + 1)$ also holds.

Let T^+ be the set of indices j for which we have an outgoing arc $v_{l+1}v_j$ from v_{l+1} to v_j , i.e.,

$$T^+ = \{j > l + 1 \mid v_{l+1}v_j \in E(H)\}.$$

Similarly, we define T^- for the set of incoming arcs:

$$T^- = \{j > l + 1 \mid v_j v_{l+1} \in E(H)\}.$$

We should select $x_{l+1} \in X_{l+1}^l$ in such a way to have arcs from x_{l+1} to the vertices of the set X_j^{l+1} in the appropriate direction, and such that updating X_j^l to X_j^{l+1} we respect the condition II of the property $P(l+1)$. In the following, let Y_j be the set of vertices of X_{l+1}^l which does not have enough arcs to X_j^l and hence are not good candidates for x_{l+1} . More precisely, for each $t \in T^+$ (resp. for each $t \in T^-$), we define the set Y_t as follows

$$Y_t := \{ y \in X_{l+1}^l \mid |\Gamma^+(y) \cap X_t^l| < (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l| \}$$

$$(\text{resp. } Y_t := \{ y \in X_{l+1}^l \mid |\Gamma^-(y) \cap X_t^l| < (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l| \}).$$

By property $\mathcal{P}(l)$ and the hypothesis, we have:

$$|X_t^l| \geq (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)^{N(j,l)}|V_t| \geq (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)^h|V_t|$$

$$\geq (h+1)\sigma|V_t| > \sigma|V_t| > \sigma u.$$

It follows from Definition A, the fact that $\text{bias}_\nu(X_t^l \cup Y_t) \leq \epsilon e$, and the choice of parameters that for all $t \in T^+ \cup T^-$, $|Y^t| \leq \sigma|V_{l+1}|$. To see this, suppose for the sake of a contradiction that this is not the case, i.e., $|Y_t| > \sigma|V_{l+1}|$. By the choice of ϵ and $|V_{l+1}| \geq \frac{n}{m}$, this means that $|Y_t||X_t^l| > \sigma^2 u^2 \geq \frac{\sigma^2}{m^2} n^2 \geq \frac{1}{(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)} \epsilon e$.

This means that $\epsilon e < (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l||Y_t|$.

From now on without loss of generality we will suppose that $t \in T^+$ (the other case follows similarly). Two cases can happen:

- if $e(X_t^l, Y_t) < \epsilon e$, then we have $e(X_t^l, Y_t) < (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l||Y_t| < \frac{1}{\nu}(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l||Y_t|$. By the definition of Y_t , we also have $e(Y_t, X_t^l) < (\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l||Y_t|$. It follows that $d(X_t^l, Y_t) = \frac{\bar{e}(X_t^l, Y_t)}{|X_t^l||Y_t|} < \eta - \sigma$. We saw that $X_t^l > \sigma|V_t|$ and by our assumption we have $Y_t > \sigma|V_{l+1}|$. These all will provide a contradiction to the σ -regularity, since $d(V_t, V_{l+1}) \geq \eta$.
- if $e(X_t^l, Y_t) \geq \epsilon e$ then by the condition on the bias, $e(Y_t, X_t^l) \geq \nu e(X_t^l, Y_t)$, and this means that $e(X_t^l, Y_t) \leq \frac{1}{\nu} e(Y_t, X_t^l) < \frac{1}{\nu}(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)|X_t^l||Y_t|$. So

$$\bar{e}(X_t^l, Y_t) = e(X_t^l, Y_t) + e(Y_t, X_t^l) \leq (1 + 1/\nu)(\eta/(1 + 1/\nu) - \sigma)|X_t^l||Y_t|.$$

This implies that $d(X_t^l, Y_t) \leq (1 + 1/\nu)(\eta/(1 + 1/\nu) - \sigma) < (\eta - \sigma)$ and this be again a contradiction as in the first case.

We infer that $Y_t \leq \sigma|V_{l+1}|$. Therefore $|X_{l+1}^l \setminus (\cup_{t \in T^- \cup T^+} Y^t)| \geq (\eta/(1 + 1/\nu) - \sigma)^{N(l+1,l)}|V_{l+1}| - |T^- \cup T^+| \sigma|V_{l+1}| \geq ((\eta/(1 + 1/\nu) - \sigma)^{N(l+1,l)} - h\sigma)|V_{l+1}| \geq \sigma u \geq 1$, and so the set $X_{l+1}^l \setminus (\cup_{t \in T^- \cup T^+} Y^t)$ is not empty. This means that we can select $x_{l+1} \in X_{l+1}^l \setminus (\cup_{t \in T^- \cup T^+} Y^t)$ in such a way to respect the condition I of $\mathcal{P}(l+1)$. By the choice of x_{l+1} , it is easy to verify that by taking $X_t^{l+1} = X_t^l \cap \Gamma(x_{l+1})$ for

$t \in T^- \cup T^+$ and $X_t^{l+1} = X_t^l$ for $t \notin T^- \cup T^+$, we respect also the condition II of $\mathcal{P}(l+1)$. And so $\mathcal{P}(l+1)$ is true. The above induction shows that $\mathcal{P}(l)$ holds for all $l, 0 \leq l \leq h$, so x_1, \dots, x_h can be found as desired. These vertices form a copy of H . \square

Proof of Theorem A.1. — We consider an oriented graph H on h vertices, a real number $\nu \in (0, 1)$, a sufficiently large integer number n ($n \geq N$, for N being fixed later), a digraph D on n vertices with $e = \beta n^2$ (β to be fixed later) arcs whose bias is at most $\nu \epsilon$ (ϵ to be fixed later). In what follows, for an oriented graph D , we denote by \overline{D} the corresponding non-oriented graph obtained by forgetting the orientation of D .

We now apply Szemerédi's regularity lemma to \overline{D} for l and some σ to be fixed later. We set l sufficiently large but fixed such that a graph G on at least l vertices with $(1 - \sigma) \binom{l}{2}$ edges contains a copy of \overline{H} as subgraph. We obtain a σ -uniform equipartition of $V(G)$ into m parts V_1, \dots, V_m with $l \leq m \leq L$. We require β to satisfy:

$$\beta n^2 \geq \eta \frac{n^2}{2} + \frac{n^2}{m} + ex(m, \overline{H}) \left(\frac{n}{m}\right)^2,$$

for some small enough η . When H is non bipartite, $ex(m, h) \geq (1 - \sigma) \binom{m}{2}$, and it gives a lower bound on β in terms of η and σ (as we will see η and σ are correlated in such a way that η small enough will imply σ small enough and $\sigma < \eta$). When H is bipartite, we can choose β as small as we want as long as η is also chosen small (it just implies n to be big enough).

The idea is that from this partition we construct a graph G , in which each vertex represents a part V_i of the corresponding partition. Two vertices of G are linked together if the corresponding parts have a density of at least η in between them (i.e., $(v_i, v_j) \in E(G)$ iff $d(V_i, V_j) \geq \eta$) for some $0 < \eta < 1$ which will be chosen according to the hypothesis of Lemma A.4. As long as η is not too big, there is some β that ensures us that $e(G) \geq ex(m, \overline{H})$. Consequently we get a copy of \overline{H} in G , wlog it uses the vertices $v_1, \dots, v_h, v_i \in V_i$, and we add $m - h$ vertices $v_{h+i} \in V_{h+i}$.

We will apply Lemma A.4 with appropriate values of ϵ, η and σ . We fix l large enough and we choose η small enough such that a graph G with $(1 - \eta^h) \binom{m}{2}$ edges contains a copy of \overline{H} for every $m \geq l$. Once l and η are fixed, we choose σ such that $\left(\frac{\eta}{1+\frac{1}{\nu}} - \sigma\right)^h \geq (h+1)\sigma$. Remark that such a positive σ always exists and $\sigma < \eta$. To see this, remark that the continuous function $f(x) = \left(\frac{\eta}{1+\frac{1}{\nu}} - x\right)^h - (h+1)x$ has a positive value at $x = 0$ and so for some small x close to zero it will still remain positive. We apply Szemerédi's regularity lemma to σ and l and \overline{D} on n vertices with βn^2 edges (β chosen as above with respect to this choice of σ and η) for n large enough. There exists L such that a partition to m parts exists

with $l \leq m \leq L$. We choose N large enough such that $\sigma \frac{N}{L} \geq 1$. Remark that $u \geq \frac{N}{L}$ in Lemma A.4 so that $\sigma u \geq 1$. Let us choose $\epsilon = \frac{\sigma^2(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)}{m^2}$. We claim that for these choices of parameters every oriented graph D on $n \geq N$ vertices with $\text{bias}(D) \leq \nu \epsilon e$ contains a copy of H . Let V_1, \dots, V_h be the parts of the equipartition which provide a copy of \overline{H} . We apply lemma A.4 to this data. We should verify that the conditions of lemma are verified:

- (i) We have easily $u \geq \frac{n}{m}$;
- (ii)&(iii) For each arc $(i, j) \in H$, the pair (V_i, V_j) is σ -uniform and $d(V_i, V_j) \geq \eta$. This is true because we find the copy of H in the graph G formed by the pairs which were σ -uniform and have the required density condition;
- (iv) For each arc (i, j) , $\text{bias}_\nu(D(V_i, V_j)) \leq \epsilon e$, where $\epsilon = \frac{\sigma^2(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)}{m^2}$. This is true because D has this property, and so $D(V_i, V_j)$ inherits this property from D .

Furthermore, the conditions $(\frac{\eta}{1+\frac{1}{\nu}} - \sigma)^h \geq (h+1)\sigma$ and $\sigma u \geq 1$ are also verified. So we can apply Lemma A.4 to find a copy of H in D . \square

A result from N. Alon and A. Shapira [1] says that if we have an oriented graph such that we have to remove ϵn^2 arcs before it becomes H -free, then it has many copies of H . The exact statement of their theorem is the following:

Theorem A.5 (Alon and Shapira [1]). — *For every fixed ϵ and h , there is a constant $c(\epsilon, h)$ with the following property: for every fixed digraph H of size h , and for every digraph D of a large enough size n , from which we need to delete ϵn^2 arcs to make it H -free, D contains at least $c(\epsilon, h)n^h$ copies of H .*

From this theorem and Theorem A.1, we may deduce the following Corollary:

Corollary A.6. — *For every h , every oriented graph H on h vertices, there exist $\epsilon > 0$, $0 < \beta < \frac{1}{2}$, $c(\epsilon, h, \beta)$ and an integer N such that every digraph D on $n \geq N$ vertices, with more than βn^2 arcs and with $\text{bias}(D) \leq \epsilon e$ contains $c(\epsilon, h)n^h$ copies of H .*

Proof. — (Sketch of proof) Let h be an integer and H an oriented graph on h vertices, Theorem A.1 gives ϵ , β and N such that every digraph D on $n \geq N$ vertices, with βn^2 arcs and with $\text{bias}_{0.4}(D) \leq \epsilon e$ contains H as a subgraph.

Setting $\epsilon' = \epsilon\beta/10$, every digraph D on $n \geq N$ vertices, with $(\beta + \epsilon')n^2$ arcs and with $\text{bias}(D) = \text{bias}_{0.5}(D) \leq \epsilon e$ contains H as a subgraph. On top of this, if we delete $\epsilon' n^2$ arcs, the remaining oriented digraph D' still contains H as a subgraph. Indeed D' has more than βn^2 arcs, and one can easily check that $\text{bias}_{0.4}(D') \leq \epsilon e$, so Theorem A.1 ensures the existence of a copy of H .

We may now apply Theorem A.5 which will imply the existence of $c(\varepsilon', h)n^h$ copies of H in D . Notice that if $\beta + \varepsilon' \geq 1/2$, we can change the value of ε' to some smaller value in order to have a number below $1/2$, and the reasoning remains unchanged. \square

In other words, the above corollary says that given h , there exist $\epsilon > 0$, $0 < \beta < \frac{1}{2}$, $c(\epsilon, h, \beta)$ and an integer N such that every digraph D on $n \geq N$ vertices, with βn^2 arcs and with $\text{bias}(D) \leq \epsilon$ contains the correct order of any orientation of any graph on h vertices.

OMID AMINI, CNRS-DMA, École Normale Supérieure, 45 Rue d'Ulm, 75230 Paris, France
E-mail : `oamini-at-dma.ens.fr`

SIMON GRIFFITHS, IMPA, Est. Dona Castorina 110, Jardim Botânico, Rio de Janeiro, Brazil
E-mail : `sgriff-at-impa.br`

FLORIAN HUC, CUI, Université de Genève, Geneva, Switzerland
E-mail : `florian.huc-at-unige.ch`