

On the Query Complexity of Testing Orientations for being Eulerian ^{*}

Eldar Fischer ^{**1}, Oded Lachish², Ilan Newman^{**3}, Arie Matsliah¹, and Orly Yahalom¹

¹ Computer Science Department, Technion - Israel Institute of Technology, Haifa 32000, Israel.

Emails: {eldar, ariem, oyahalom}.cs.technion.ac.il.

² Centre for Discrete Mathematics and its Applications (DIMAP), Warwick, UK.

Email: o.lachish@warwick.ac.uk.

³ Computer Science Department, Haifa University, Haifa 31905, Israel.

Email: ilan@cs.haifa.ac.il.

Abstract. We consider testing directed graphs for being Eulerian in the orientation model introduced in [15]. Despite the local nature of the property of being Eulerian, it turns out to be significantly harder for testing than other properties studied in the orientation model. We show a non-constant lower bound on the query complexity of 2-sided tests and a linear lower bound on the query complexity of 1-sided tests for this property. On the positive side, we give several 1-sided and 2-sided tests, including a sub-linear query complexity 2-sided test for general graphs. For special classes of graphs, including bounded-degree graphs and expander graphs, we provide improved results. In particular, we give a 2-sided test with constant query complexity for dense graphs, as well as for expander graphs with a constant expansion parameter.

1 Introduction

Property testing deals with the following relaxation of decision problems: Given a property \mathcal{P} , an input structure S and $\epsilon > 0$, distinguish between the case where S satisfies \mathcal{P} and the case where S is ϵ -far from satisfying \mathcal{P} . Roughly speaking, an input S is said to be ϵ -far from satisfying a property \mathcal{P} if more than an ϵ -fraction of its values must be modified in order to make it satisfy the property. Algorithms which distinguish with high probability between the two cases are called *property testers* or simply *testers* for \mathcal{P} . Furthermore, a tester for \mathcal{P} is said to be *1-sided* if it never rejects an input that satisfies \mathcal{P} . Otherwise, the tester is called *2-sided*. We say that a tester is *adaptive* if some of the choices of the locations for which the input is queried may depend on the returned values (answers) of previous queries. Otherwise, the tester is called *non-adaptive*. Property testing normally deals with problems involving a very large input or a costly retrieval procedure. Thus, the number of queries of input values, rather than the computation time, is considered to be the most expensive resource.

^{*} A full version is available at <http://www.cs.technion.ac.il/~oyahalom/EulerianOrientations.pdf>.

^{**} Research supported in part by an ISF grant number 1101/06.

Property testing has been a very active field of research since it was initiated by Blum, Luby and Rubinfeld [5]. The general definition of property testing was formulated by Rubinfeld and Sudan [25], who were interested mainly in testing algebraic properties. The study of property testing for combinatorial objects, and mainly for labelled graphs, began in the seminal paper of Goldreich, Goldwasser and Ron [12]. They introduced the *dense graph model*, where the graph is represented by an adjacency matrix, and the distance function is computed accordingly. For comprehensive surveys on property testing see [24, 8].

The dense graph model is in a sense too lenient, since for n -vertex graphs, the distance function allows adding and removing $o(n^2)$ edges, regardless of the number of actual edges in the graph. Thus, many interesting properties, such as connectivity in undirected or directed graphs, are trivially testable in this model, as all the graphs are close to satisfying the property. In recent years, researchers have studied several alternative models for graph testing, including the *bounded-degree* graph model of [13], in which a sparse representation of sparse graphs is considered, and the *general density* model (also called the *mixed* model) of [21] and [17]. In these models, the distance function allows edge insertions and deletions whose number is at most a fraction of the number of the edges in the original graph.

Property testing of directed graphs has also been studied in the context of the above models [1, 3]. Here we continue the study of testing properties of directed graphs in the orientation model, which started in [15] and followed in [14] and [7]. In this model, an underlying undirected graph $G = (V, E)$ is given in advance, and the actual input is an orientation \vec{G} of G , in which every edge in E has a direction. Our testers may access the input using edge queries. That is, every query concerns an edge $e \in E$, and the answer to the query is the direction of e in \vec{G} . An orientation \vec{G} of G is called ϵ -close to a property \mathcal{P} if it can be made to satisfy \mathcal{P} by inverting at most an ϵ -fraction of the edges of G , and otherwise \vec{G} is said to be ϵ -far from \mathcal{P} .

Note that the distance function in the orientation model naturally depends on the size of the underlying graph and is independent of representation details. Moreover, the testing algorithm may strongly depend on the structure of the underlying graph. The model is strict in that the distance function allows only edge inversions, but no edge insertions or deletions. On the other hand, we assume that our algorithms have a full knowledge of the underlying graph, whose size is roughly the same as the input size. Viewing the underlying graph as a parameter that the testing algorithm receives in advance, we say that the orientation model is an example of a *massively parameterized* model. Other examples of massively parameterized models appear in [20], where the property is represented by a known bounded-width branching program, in [9], where the input is a vertex-coloring of a known graph, and in other works.

In this paper we consider the property of being Eulerian, which was presented in [14] as one of the natural orientation properties whose query complexity was still unknown. A directed graph \vec{G} is called *Eulerian* if for every vertex v in the graph, the in-degree of v is equal to its out-degree. An undirected graph G has an Eulerian orientation \vec{G} if and only if all the degrees of G are even. Such an undirected graph is called Eulerian also. Throughout the paper we assume that our underlying undirected graph G is Eulerian. We note that it is common to require an Eulerian graph to be con-

nected. However, we may ignore this requirement, as all our algorithms and proofs work equally well whether G is connected or not. Moreover, as G is given as a parameter, its connectivity can be tested in a preprocessing stage.

Eulerian graphs and Eulerian orientations have attracted researchers since the dawn of graph theory in 1736, when Leonard Euler published his solution for the famous “Königsberg bridge problem”. Throughout the years, Eulerian graphs have been the subject of extensive research (e.g. [23, 18, 26, 19, 6, 2]; see [10, 11] for an extensive survey). Aside from their appealing theoretic characteristics, Eulerian graphs have been studied in the context of networking [16] and genetics [22].

Testing for being Eulerian in the orientation model is equivalent to the following problem. We have a known network (e.g. a communication network or a transportation system) where every edge can transport a unit of “flow” in both directions. Our goal is to know whether the network is “balanced”, or far from being balanced, where being balanced means that the number of flows entering every node in the network is equal to the number of flows exiting it. To examine the network, we detect the flow direction in selected individual edges, and this is deemed to be the expensive operation.

The main difficulty in testing orientations for being Eulerian arises from the fact that an orientation might have a small number of unbalanced vertices, and each of them with a small imbalance, and yet be far from being Eulerian. This is since trying to balance an unbalanced vertex by inverting some of its incident edges may violate the balance of its balanced neighbors. Thus, we must continue to invert edges along a directed path between a vertex with a positive imbalance and a vertex with a negative imbalance. We call such a path a *correction path*. A main component of our work is giving upper bounds for the length of the correction paths. We note that Babai [2] showed that the ratio between the diameter of digraphs and the diameter of their underlying undirected graphs is $\Omega(n^{1/3})$ for an infinite family of Eulerian graphs.

Our upper bounds are based on three “generic” tests, one 1-sided test and two 2-sided tests. Instead of receiving ϵ as a parameter, the generic tests receive a parameter p , which stands for the number of required correction paths in an orientation that is far from being Eulerian. We hence call these tests *p-tests*. We later derive ϵ -tests from the p -tests by proving two lower bounds for p . The first one gives an efficient test for dense graphs and the second one gives an efficient test for expander graphs. Finally, we show how to use variations of the expander tests for obtaining a 1-sided test and a 2-sided test for general graphs, using a decomposition (“chopping”) procedure into subgraphs that are roughly expanders. The 2-sided test that we obtain this way has a sub-linear query complexity for every graph. Unfortunately, our chopping procedure is adaptive and has an exponential computational time in $|E|$. All of our other algorithms are non-adaptive and their computational complexity is of the same order as their query complexity.

On the negative side, we provide several lower bounds. We show that any 1-sided test for being Eulerian must use $\Omega(m)$ queries for some graphs. For bounded-degree graphs, we use the toroidal grid to prove non-constant 1-sided and 2-sided lower bounds. These bounds are noteworthy, as bounded-degree graphs have a constant size witness for not being Eulerian, namely the edges incident with one unbalanced vertex. In contrast, the *st*-connectivity property, whose witness must include a cut in the graph, is testable with a constant number of queries in the orientation model [7]. In other testing

Table 1. Upper bounds

Result	1-sided tests	2-sided tests
Graphs with large d	$O\left(\frac{\Delta m}{\epsilon^2 d^2}\right)$	$\min\left\{\tilde{O}\left(\frac{m^3}{\epsilon^6 d^6}\right), \tilde{O}\left(\frac{\sqrt{\Delta} m}{\epsilon^2 d^2}\right)\right\}$
α -expanders (Section 4)	$O\left(\frac{\Delta \log(1/\epsilon)}{\alpha \epsilon}\right)$	$\min\left\{\tilde{O}\left(\left(\frac{\log(1/\epsilon)}{\alpha \epsilon}\right)^3\right), \tilde{O}\left(\frac{\sqrt{\Delta} \log(1/\epsilon)}{\alpha \epsilon}\right)\right\}$
General graphs (Section 6)	$O\left(\frac{(\Delta m \log m)^{2/3}}{\epsilon^{4/3}}\right)$	$\min\left\{\tilde{O}\left(\frac{\Delta^{1/3} m^{2/3}}{\epsilon^{4/3}}\right), \tilde{O}\left(\frac{\Delta^{3/16} m^{3/4}}{\epsilon^{5/4}}\right)\right\}$

models there are known super-constant lower bounds also for properties which have constant-size witness, e.g., [4] prove a linear lower bound for testing whether a truth assignment satisfies a known 3CNF formula. However, most of these bounds are for properties that have stronger expressive power than that of being Eulerian.

Table 2. Lower bounds

Result	1-sided tests	2-sided tests
General graphs (Section 2)	$\Omega(m)$	—
Bounded-degree graphs, non-adaptive tests	$\Omega(m^{1/4})$	$\Omega\left(\sqrt{\frac{\log m}{\log \log m}}\right)$
Bounded-degree graphs, adaptive tests	$\Omega(\log m)$	$\Omega(\log \log m)$

Tables 1 and 2 summarize our upper and lower bounds, respectively. Here and throughout the paper, we set $n = |V|$ and $m = |E|$, let Δ be the maximum vertex-degree in G , and set $d \stackrel{\text{def}}{=} m/n$. The tilde notation hides polylogarithmic factors. Due to space limitations, our upper bounds for dense graphs and lower bounds for bounded-degree graphs are omitted from this version, and most of the proofs are given as sketches.

2 Preliminaries and the 1-sided lower bound

In this section we introduce basic definitions, notations and lemmas to be used in the sequel. Throughout the paper, we assume a fixed and known underlying graph $G = (V, E)$ which is Eulerian, that is, for every $v \in V$, the degree $\deg(v)$ of v is even. Given an orientation $\vec{G} = (V, \vec{E})$ and a vertex $v \in V$, let $\text{indeg}_{\vec{G}}(v)$ denote the in-degree of v with respect to \vec{G} and let $\text{outdeg}_{\vec{G}}(v)$ denote the out-degree of v with respect to \vec{G} . We define the *imbalance* of v in \vec{G} as $\text{ib}_{\vec{G}}(v) \stackrel{\text{def}}{=} \text{outdeg}_{\vec{G}}(v) - \text{indeg}_{\vec{G}}(v)$. In the following, we sometimes omit the subscript \vec{G} whenever it is obvious from the context. We say that a vertex $v \in V$ is a *spring* in \vec{G} if $\text{ib}_{\vec{G}}(v) > 0$. We say that v is a *drain* in \vec{G} if $\text{ib}_{\vec{G}}(v) < 0$. If $\text{ib}_{\vec{G}}(v) = 0$ then we say that v is *balanced* in \vec{G} . We say that \vec{G} is *Eulerian* if all its vertices are balanced. Since all the vertices of G are of even degree, there always exists some Eulerian orientation \vec{G} of G .

Given a set $U \subseteq V$, let:
 $E(U) \stackrel{\text{def}}{=} \{\{u, v\} \in E \mid u, v \in U\}$ and $\vec{E}(U) \stackrel{\text{def}}{=} \{(u, v) \in \vec{E} \mid u, v \in U\}$,
 $\partial U \stackrel{\text{def}}{=} \{\{u, v\} \in E \mid u \in U, v \notin U\}$ and $\vec{\partial} U \stackrel{\text{def}}{=} \{(u, v) \in \vec{E} \mid u \in U, v \notin U\}$.
Given two disjoint sets $U, W \subseteq V$, let $E(U, W) \stackrel{\text{def}}{=} \{\{u, w\} \in E \mid u \in U, w \in W\}$
and $\vec{E}(U, W) \stackrel{\text{def}}{=} \{(u, w) \in \vec{E} \mid u \in U, w \in W\}$.

Lemma 1 *Suppose that \vec{H} is a knowledge graph that does not contain invalid cuts. Then \vec{H} is extensible to an Eulerian orientation $\vec{G} = (V, \vec{E}_G)$ of G . That is, $\vec{E}_H \subseteq \vec{E}_G$. Consequently, a witness that an orientation \vec{G} is not Eulerian must contain at least half of the edges of some invalid cut with respect to \vec{G} .*

Proof sketch. We extend the knowledge graph to an orientation of the entire graph by orienting the edges one by one. In each step we prove using counting arguments that if a certain orientation of an edge would invalidate one of the cuts, then orienting it in the other direction would not invalidate any of the other cuts. \square

Theorem 2. *There exists an infinite family of graphs for which every 1-sided test for being Eulerian must use $\Omega(m)$ queries.*

Proof. For every even n , let $G_n \stackrel{\text{def}}{=} K_{2, n-2}$, namely, the graph with a set of vertices $V = \{v_1, \dots, v_n\}$ and a set of edges $E = \{\{v_i, v_j\} \mid i \in \{1, 2\}, j \in \{3, \dots, n\}\}$. Clearly, G_n is Eulerian and $n = \Omega(m)$. Consider the orientation \vec{G}_n of G_n in which all the edges incident with v_1 are outgoing and all the edges incident with v_2 are incoming. Clearly, \vec{G}_n is $\frac{1}{2}$ -far from being Eulerian. According to Lemma 1, every 1-sided test must query at least half of the edges in some unbalanced cut (because otherwise it would clearly not obtain an invalid cut in the knowledge graph). However, one can easily see that every cut which does not separate v_1 and v_2 is balanced, while every cut which separates v_1 and v_2 is of size $n - 2 = \Omega(m)$. \square

Let \vec{G} be an orientation of G . Given a subgraph $\vec{H} = (V_H, \vec{E}_H)$ of \vec{G} (that is, a directed graph where $V_H \subseteq V$ and $\vec{E}_H \subseteq \vec{E}$) we define $\vec{G}_{\vec{H}} \stackrel{\text{def}}{=} (V, \vec{E}_{\vec{H}})$ to be the orientation of G derived from \vec{G} by inverting all the edges of \vec{H} . Namely, $\vec{E}_{\vec{H}} = \vec{E} \setminus \vec{E}_H \cup \{(v, u) \in (V_H)^2 \mid (u, v) \in \vec{E}_H\}$. We say that \vec{H} is a *correction subgraph* of \vec{G} if $\vec{G}_{\vec{H}}$ is Eulerian. Note that in such a case, \vec{G} is $|\vec{E}_H|/m$ -close to being Eulerian. Since we assume that G is Eulerian, there exists some correction subgraph \vec{H} for any \vec{G} . Furthermore, it is not difficult to show that any correction subgraph \vec{H} of \vec{G} has an acyclic subgraph which is also a correction subgraph of \vec{G} . Let S be the set of springs in \vec{G} and let T be the set of drains in \vec{G} . We say that a directed path $\vec{P} = \langle u_0, \dots, u_k \rangle$ in \vec{G} is a *spring-drain path* if $u_0 \in S$ and $u_k \in T$. It is easy to show that for any correction subgraph \vec{H} of \vec{G} , u_0 is a spring in \vec{H} and u_k is a drain in \vec{H} .

Lemma 3 *If \vec{G} is not Eulerian then any acyclic correction subgraph \vec{H} of \vec{G} is a union of $p = \frac{1}{4} \sum_{u \in V} |\text{ib}(u)|$ edge-disjoint spring-drain paths.*

Proof sketch. Suppose that \vec{G} is not Eulerian and let \vec{H} be an acyclic correction subgraph of \vec{G} . By definition, if we invert all the edges of \vec{H} in \vec{G} then we obtain an Eulerian orientation of G . It can easily be seen that, since \vec{G} is not Eulerian, \vec{H} contains a spring-drain path. We thus invert the edges of \vec{H} along one spring-drain path at a time, until we obtain an Eulerian orientation. One can see that \vec{H} is thus decomposed to the inverted paths. The value of p is computed by noting that by inverting a spring-drain path, we reduce the sum $\sum_{u \in V} |\text{ib}(u)|$ by exactly four. \square

Let p be some positive number. If every correction subgraph of an orientation \vec{G} is a union of at least p disjoint spring-drain paths, we say that \vec{G} is p -far from being Eulerian. An algorithm is called a p -test for being Eulerian if it accepts an Eulerian orientation with probability at least $2/3$ and rejects a p -far orientation with probability at least $2/3$. Similarly to ϵ -tests, if a p -test accepts every Eulerian orientation with probability 1 then it is called 1 -sided, and otherwise it is called 2 -sided.

Given $\beta > 0$, we say that a vertex v is β -small if $\deg(v) \leq \beta$ and β -big if $\deg(v) > \beta$. An orientation \vec{G} is called β -Eulerian if all the β -small vertices in V are balanced in \vec{G} . Note that for $\beta \geq \Delta$, \vec{G} is β -Eulerian if and only if \vec{G} is Eulerian. All our lemmas and observations for Eulerian orientations may be adapted to β -Eulerian orientations. In particular, we can show that modifying an orientation \vec{G} to become β -Eulerian requires inverting edges along at least $\frac{1}{4} \sum_{u \in V, \deg(u) \leq \beta} |\text{ib}(u)|$ spring-drain paths in which at least one of the spring and the drain is β -small. We call such paths β -spring-drain paths.

An algorithm is called a (p, β) -test for being Eulerian for some positive number p if it accepts a β -Eulerian orientation with probability at least $2/3$ and rejects an orientation that is p -far from being β -Eulerian with probability at least $2/3$. As usual, a (p, β) -test is said to be 1 -sided if it accepts every β -Eulerian orientation with probability 1. Otherwise, the test is said to be 2 -sided.

3 Generic tests

We present a p -test and two (p, β) -tests for being Eulerian. In later sections we devise several lower bounds on p for every orientation \vec{G} that is ϵ -far from being Eulerian, thus obtaining corresponding upper bounds on the tests below.

We begin with a simple 2 -sided p -test whose query complexity is independent of the maximum degree Δ . The algorithm uses probabilistic methods, as well as the characterization of p given in Lemma 3, in order to detect an unbalanced vertex with high probability. To simplify notation, we denote $\delta \stackrel{\text{def}}{=} \frac{p}{4m}$.

Algorithm 4 *SIMPLE-2*(\vec{G}, p):

- Repeat $\frac{4}{\delta}$ times independently:
 - Select an edge $e \in E(G)$ uniformly and query it. Denote the start vertex of e in \vec{E} by u and the end vertex of e in \vec{E} by v .
 - Query $\frac{16 \ln(12/\delta)}{\delta^2}$ edges incident with u uniformly and independently and reject if the sample contains at least $(1 + \delta) \frac{8 \ln(12/\delta)}{\delta^2}$ outgoing edges.

– Accept if the input was not rejected earlier.

Lemma 5 *SIMPLE-2 is a 2-sided p -test for being Eulerian with query complexity $\tilde{O}\left(\frac{1}{\delta^3}\right) = \tilde{O}\left(\frac{m^3}{p^3}\right)$.*

We next give a simple 1-sided (p, β) -test, which has a better query complexity than SIMPLE-2 for $\Delta \ll \frac{m^2}{p^2} \ln\left(\frac{m}{p}\right)$. Note that the test checks only β -small vertices for being unbalanced.

Algorithm 6 *GENERIC-1*(\vec{G}, p, β):

1. Repeat $\frac{\ln 3 m}{p}$ times independently:
 - Select an edge $e \in E(G)$ uniformly and query it. Denote the start vertex of e in \vec{E} by u and the end vertex of e in \vec{E} by v .
 - If $\deg(u) \leq \beta$ then query all the edges $\{u, w\} \in E$ and reject if u is unbalanced.
2. Repeat $\frac{\ln 3 m}{p}$ times independently:
 - Select an edge $e \in E(G)$ uniformly and query it. Denote the start vertex of e in \vec{E} by u and the end vertex of e in \vec{E} by v .
 - If $\deg(v) \leq \beta$ then query all the edges $\{w, v\} \in E$ and reject if v is unbalanced.
3. Accept if the input was not rejected by the above.

Lemma 7 *GENERIC-1 is a 1-sided (p, β) -test for being Eulerian with query complexity $O\left(\frac{\beta m}{p}\right)$. In particular, for $\beta = \Delta$, GENERIC-1 is a 1-sided p -test with query complexity $O\left(\frac{\Delta m}{p}\right)$.*

We conclude this section with a 2-sided (p, β) -test, which gives better query complexity than GENERIC-1 for $\beta \gg \log^2 m$ and better query complexity than SIMPLE-2 for $p \ll \frac{m}{\sqrt{\beta}}$. The main idea of the algorithm is to perform roughly $O((\log \beta)^2)$ testing stages, each designed to detect unbalanced β -small vertices whose degree and imbalance lie in a certain interval. In the following, \log denotes the logarithm with base 2.

Algorithm 8 *MULTISTAGE-2*(\vec{G}, p, β):

For $i = 1, \dots, \lceil \log \beta \rceil - 1$, do:

1. Let $V_i \stackrel{\text{def}}{=} \{u \in V \mid \deg(u) \in [2^i, 2^{i+1})\}$ and $n_i \stackrel{\text{def}}{=} |V_i|$.
2. Let $j = \lceil i/2 \rceil$. If $2^j \cdot n_i > \frac{2p}{(\log \beta)^2}$ then:
 - Sample $x_{ij} = \frac{\ln 12 (\log \beta)^2 2^{j+1} n_i}{2p}$ vertices in V_i uniformly and independently.
 - For every sampled vertex u , query all the edges incident with u , and reject if u is unbalanced.
3. For every $j \in \{\lceil i/2 \rceil + 1, \dots, i - 1\}$ such that $2^j \cdot n_i > \frac{2p}{(\log \beta)^2}$ do:
 - Sample $x_{ij} = \frac{\ln 12 (\log \beta)^2 2^{j+1} n_i}{2p}$ vertices in V_i uniformly and independently.
 - For every sampled vertex u , query $q_{ij} = 256 \cdot \ln(6(\log \beta)^2 x_{ij}) \cdot 2^{2(i-j)}$ edges adjacent to u , uniformly and independently, and reject if the absolute difference between the number of incoming and outgoing edges in the sample is at least $\frac{q_{ij}}{4 \cdot 2^{i-j}}$.

Accept if the input was not rejected earlier.

Lemma 9 *MULTISTAGE-2 is a 2-sided (p, β) -test for being Eulerian with query complexity $\tilde{O}\left(\frac{\sqrt{\beta} m}{p}\right)$. In particular, for $\beta = \Delta$, it is a 2-sided p -test for being Eulerian with query complexity $\tilde{O}\left(\frac{\sqrt{\Delta} m}{p}\right)$.*

4 Testing orientations of expander graphs

In this section we show how to apply our generic tests for expander graphs. A graph $G = (V, E)$ is called an α -expander for some $\alpha > 0$, if it is connected and for every $U \subseteq V$ such that $0 < |E(U)| \leq m/2$ we have $|\partial U| \geq \alpha|E(U)|$. Note that while the diameter of G is $O(\log_{(1+\alpha)} m)$, the “oriented-diameter” of \vec{G} is not necessarily low, even if we assume that the orientation is Eulerian, as was shown by [2].

In the following, $\log_b^{(k)}(x)$ denotes the k -nested logarithm with base b of x , i.e., $\log_b^{(1)}(x) \stackrel{\text{def}}{=} \log_b(x)$ and $\log_b^{(k+1)}(x) \stackrel{\text{def}}{=} \log_b(\log_b^{(k)}(x))$ for any natural $k \geq 1$.

Lemma 10 *Let G be an Eulerian α -expander and let $k \geq 1$ be a natural number such that $\log_{(1+\alpha/2)}^{(k-1)} m \geq \log_{(1+\alpha/2)}\left(\frac{4}{\epsilon}\right)$. Then: (1) Every non-Eulerian orientation \vec{G} of G contains a spring-drain path of length at most $\ell_k \stackrel{\text{def}}{=} 2 \cdot \log_{(1+\alpha/2)}^{(k)} m + 2 \cdot \log_{(1+\alpha/2)}\left(\frac{4}{\epsilon}\right)$; (2) Every orientation \vec{G} of G that is ϵ -far from being Eulerian is p_k -far from being Eulerian for $p_k \stackrel{\text{def}}{=} \frac{\epsilon m}{\ell_k}$.*

Proof sketch. We prove the lemma by induction on k . In each inductive step, we use the known bounds of ℓ_k and p_k to bound ℓ_{k+1} and p_{k+1} in an iterative manner. To prove Item 1 of the lemma for $k = 1$, let \vec{G} be a non-Eulerian orientation of \vec{G} . Consider a BFS traversal of \vec{G} starting from the set S of springs. For every $i \geq 0$, let L_i be the i th level of the traversal, where $L_0 = S$, and let $U_{<i} \stackrel{\text{def}}{=} \bigcup_{0 \leq j < i} L_j$ and $U_{\geq i} \stackrel{\text{def}}{=} \bigcup_{j \geq i} L_j$. For every $i > 0$, let f_i be the number of directed edges going from L_{i-1} to L_i . Let L_ℓ be the first level that contains a drain. By the expander property of G , for every $i > 0$ while $|E(U_{<i})| \leq m/2$ we have $|\partial(U_{<i})| \geq \alpha|E(U_{<i})|$. Note that for every $i \leq \ell$, the set $U_{<i}$ contains no drains, and all the directed edges that exit it are from L_{i-1} to L_i . Hence, for every $0 < i \leq \ell$ while $|E(U_{<i})| \leq m/2$, we have $f_i > \frac{1}{2}|\partial(U_{<i})| > \frac{\alpha}{2}|E(U_{<i})|$ and therefore $|E(U_{<i+1})| > \left(1 + \frac{\alpha}{2}\right)|E(U_{<i})|$. By induction, we have $|E(U_{<i})| > \left(1 + \frac{\alpha}{2}\right)^{i-1} f_1 \geq \left(1 + \frac{\alpha}{2}\right)^{i-1}$ for every $0 < i \leq \ell$ for which $|E(U_{<i})| \leq m/2$.

Now, if for every $0 < i \leq \ell$ we have $|E(U_{<i})| \leq m/2$, then clearly, $|E(U_{<\ell})| > \left(1 + \frac{\alpha}{2}\right)^{\ell-1}$, and hence $\ell = \ell_1 < \log_{(1+\alpha/2)} m$. Otherwise, let $r > 0$ be the minimal index for which $|E(U_{<r})| > m/2$. Using similar arguments to the above, we show that $|E(U_{\geq i-1})| > \left(1 + \frac{\alpha}{2}\right)^{\ell-i+1} |E(U_{\geq \ell})| \geq \left(1 + \frac{\alpha}{2}\right)^{\ell-i+1}$ for every $r \leq i \leq \ell$, which yields $\ell_1 < 2 \cdot \log_{(1+\alpha/2)} m$.

To prove Item 2 of the lemma for $k = 1$, let \vec{G} be an orientation of G that is ϵ -far from being Eulerian. While \vec{G} is not Eulerian, choose a shortest spring-drain path in

\vec{G} and invert all its edges. By Item 1, every chosen spring-drain path is of length at most ℓ_1 . Let \vec{H} be the union of the paths inverted. Clearly, \vec{H} is a correction subgraph of \vec{G} . As \vec{G} is ϵ -far from being Eulerian, \vec{H} contains at least ϵm edges, and thus it is necessarily a union of at least $p_1 = \frac{\epsilon m}{\ell_1}$ disjoint spring-drain paths. By Lemma 3, every correction subgraph of \vec{G} contains the same number of disjoint spring-drain paths, which completes the base case.

Assuming that the lemma holds for some natural $k \geq 1$, the proof of the lemma for $k + 1$ is very similar to that of the base case. However, we now know that $f_1 \geq p_k$ and $|E(U_{\geq \ell})| \geq p_k$, and so we use our known lower bound for p_k (instead of 1 in the base case). Item 1 is now proved using standard arithmetics, as well as the condition $\log_{(1+\alpha/2)}^{(k)} m \geq \log_{(1+\alpha/2)} \left(\frac{4}{\epsilon}\right)$. The proof of Item 2 is the same as for the base case. \square

Lemma 11 *Let G be an Eulerian α -expander. Let \vec{G} be an orientation of G that is ϵ -far from being Eulerian. Then \vec{G} is p -far from being Eulerian for $p = \Omega\left(\frac{\alpha \epsilon m}{\log(\frac{1}{\epsilon})}\right)$.*

Proof sketch. The proof considers the first natural number k such that the condition of Lemma 10 does not apply, namely $\log_{(1+\alpha/2)}^{(k)} m < \log_{(1+\alpha/2)} \left(\frac{4}{\epsilon}\right)$. The proof is similar to that of Lemma 10 for smaller k 's. However, since $\log_{(1+\alpha/2)}^{(k)} m$ is sufficiently small, we are able to give the stated upper bound, which is independent of k . \square

Substituting the lower bound for p of Lemma 11 in Lemmas 5, 7, and 9, we obtain the following theorem. Note that for a constant α , the query complexity of SIMPLE-2 depends only on ϵ .

Theorem 12. *Let G be an α -expander (for some $\alpha > 0$) with m edges and maximum degree Δ . Then:*

1. *SIMPLE-2* $\left(\vec{G}, \Omega\left(\frac{\alpha \epsilon m}{\log(1/\epsilon)}\right)\right)$ is a 2-sided ϵ -test for being Eulerian with query complexity $\tilde{O}\left(\left(\frac{\log(1/\epsilon)}{\alpha \epsilon}\right)^3\right)$.
2. *GENERIC-1* $\left(\vec{G}, \Omega\left(\frac{\alpha \epsilon m}{\log(1/\epsilon)}\right), \Delta\right)$ is a 1-sided ϵ -test for being Eulerian with query complexity $O\left(\frac{\Delta \log(1/\epsilon)}{\alpha \epsilon}\right)$.
3. *MULTISTAGE-2* $\left(\vec{G}, \Omega\left(\frac{\alpha \epsilon m}{\log(1/\epsilon)}\right), \Delta\right)$ is a 2-sided ϵ -test for being Eulerian with query complexity $\tilde{O}\left(\frac{\sqrt{\Delta} \log(1/\epsilon)}{\alpha \epsilon}\right)$.

5 Testing orientations of “lame” directed expanders

In this section we discuss a variation of the expander test, which will serve us in Section 6 for devising tests for general graphs. Given an orientation \vec{G} of G , we now test a subgraph $\vec{G}[U]$ of \vec{G} , induced by a subset $U \subseteq V$. We refer to the edges in $E(U)$ as the *internal edges* of $\vec{G}[U]$, and denote $m_U \stackrel{\text{def}}{=} |E(U)|$. We say that $\vec{G}[U]$ is *Eulerian* if and

only if all the vertices in U are balanced in \vec{G} . We say that $\vec{G}[U]$ is β -Eulerian if and only if all the β -small vertices in U are balanced in \vec{G} . Note that these definitions rely also on the edges in ∂U , which we will henceforth call *external edges*. We assume that the orientations of all the external edges are known, and furthermore, we use a distance function that does not allow inverting external edges. Namely, we will say that $\vec{G}[U]$ is ϵ -close to being Eulerian if and only if it has a correction subgraph of size at most ϵm_U which includes only internal edges. Otherwise, we say that $\vec{G}[U]$ is ϵ -far from being Eulerian. Note that we can view the external edges as comprising a knowledge graph (see Section 2). We always assume that all the cuts in \vec{G} are valid with respect to the orientation $\vec{\partial} U$ of the external edges. This condition ensures that $\vec{G}[U]$ can be made Eulerian (or β -Eulerian) by inverting internal edges only.

We will be interested in induced subgraphs $\vec{G}[U]$ that are “lame directed expanders”. Formally, given a subset $U \subseteq V$ and a parameter $\beta > 0$, we say that a cut (A, B) of U is a β -cut of U if $|E(B)| \geq |E(A)| \geq \beta$. Given $\alpha, \beta > 0$, we say that the subgraph $\vec{G}[U]$ of G is an (α, β) -expander if for every β -cut (A, B) of U :

$$|E(A, B)| - \left| |\vec{E}(V \setminus U, A)| - |\vec{E}(A, V \setminus U)| \right| \geq 2\alpha |E(A)|. \quad (1)$$

Lemma 13 *Let $\alpha, \beta, \epsilon > 0$ be parameters and let $U \subseteq V$ be such that $\vec{G}[U]$ is an (α, β) -expander. Denote $m_U \stackrel{\text{def}}{=} |E(U)|$ and $\Delta_U \stackrel{\text{def}}{=} \max\{\deg(u) \mid u \in U\}$. Assume that the external edges of U are known and do not induce an invalid cut. Then:*

1. *There exists a 1-sided ϵ -test for whether $\vec{G}[U]$ is Eulerian, GEN-1($\vec{G}[U], \alpha, \beta, \epsilon$), whose query complexity is $O\left(\frac{\Delta_U \log m_U}{\epsilon \alpha} + \frac{\beta \cdot \min\{\beta, \Delta_U\}}{\epsilon}\right)$.*
2. *There exists a 2-sided ϵ -test for whether $\vec{G}[U]$ is Eulerian, MULTI-2($\vec{G}[U], \alpha, \beta, \epsilon$), whose query complexity is $\tilde{O}\left(\frac{\sqrt{\Delta_U} \log m_U}{\epsilon \alpha} + \frac{\beta \cdot \sqrt{\min\{\beta, \Delta_U\}}}{\epsilon}\right)$.*

Proof sketch. GEN-1 is based on at most two calls to GENERIC-1 (Algorithm 6) and MULTI-2 is based on at most two calls to MULTISTAGE-2 (Algorithm 8). The parameters in these calls are computed by analyzing two possible cases in which $\vec{G}[U]$ is ϵ -far from being Eulerian.

In the first case, $\vec{G}[U]$ is $\frac{\epsilon}{2}$ -far from being 2β -Eulerian, which means that we need to invert many 2β -spring-drain paths in $\vec{G}[U]$ in order to make it 2β -Eulerian. Using an analysis similar to that used in the proof of Lemma 10 (with our condition for lame expansion instead of the condition for undirected expansion), we obtain a lower bound $p' = \Omega\left(\frac{\epsilon m_U}{\log m_U / \alpha + \beta}\right)$ for the number of these 2β -spring-drain paths. Thus, to take care of this case we call GENERIC-1 or MULTISTAGE-2 to test whether $\vec{G}[U]$ is $(p', 2\beta)$ -Eulerian. Note that p' differs from our bound for expander graphs in the addition of β to the denominator, which indicates an addition of β to the upper bound on the length of a correction path. This arises from the fact that the lame expansion condition applies only for β -cuts, and thus, it might not apply in the first and last β BFS layers.

As for the second case, if $\vec{G}[U]$ is ϵ -far from being Eulerian, but $\frac{\epsilon}{2}$ -close to being 2β -Eulerian, we consider a 2β -Eulerian orientation $\vec{G}'[U]$ that is $\frac{\epsilon}{2}$ -close to $\vec{G}[U]$. Clearly, $\vec{G}'[U]$ is $\frac{\epsilon}{2}$ -far from being Eulerian. However, since it is 2β -Eulerian, we can show that it can be made Eulerian by inverting edges along paths between β -big springs and β -big drains. We next use a similar analysis as for Lemma 10. However, since the spring and drain in each of our paths are 2β -big, it can be seen that all the cuts between our BFS layers are β -cuts, and thus, we obtain a lower bound $p'' = \Omega\left(\frac{\epsilon m_U}{\log m_U / \alpha}\right)$ for the number of spring-drain paths. Hence, to take care of the second case, we call GENERIC-1 or MULTISTAGE-2 to test whether $\vec{G}[U]$ is p'' -Eulerian (namely, we use $\beta = \Delta_U$).

The correctness of our algorithms now follows from Lemmas 7 and 9. The query complexity bounds are obtained from these lemmas, noting also that the second case discussed above is only possible for $\beta < \frac{\Delta_U}{2}$. \square

6 General tests based on chopping

We provide a 1-sided test and a 2-sided test as follows. Given an orientation \vec{G} of G , we show how to decompose \vec{G} into a collection of (α, β) -expanders with a relatively small number of edges that are outside the (α, β) -expanders, called henceforth *external edges*. We will find this “chopping” adaptively while querying external edges only. If we do not find a witness showing that \vec{G} is not Eulerian during the chopping procedure, then we sample a few (α, β) -expanders and test them using GEN-1 or MULTI-2 (see Lemma 13), obtaining a 1-sided test or a 2-sided test respectively.

Lemma 14 (The chopping lemma) *Given an orientation \vec{G} as input and parameters $\alpha, \beta > 0$, we can either find a witness showing that \vec{G} is not Eulerian, or find non-empty induced subgraphs $\vec{G}_i = (V_i, \vec{E}_i = \vec{E}(V_i))$ of \vec{G} (where $i = 1, \dots, k$ for some k), which we call (α, β) -components (or simply components), that satisfy the following:*

1. *The vertex sets V_1, \dots, V_k of the components are mutually disjoint.*
2. *$|\vec{E}_i| \geq \beta$ for $i = 1, \dots, k$.*
3. *All the components \vec{G}_i are (α, β) -expanders.*
4. *The total number of external edges satisfies $|\vec{E} \setminus \bigcup_{i=1, \dots, k} \vec{E}(V_i)| = O(\alpha m^2 \log m / \beta)$.*

During the chopping procedure, we query only external edges, i.e., edges that are not in any component G_i . The query complexity is in the same order also if we find a witness that \vec{G} is not Eulerian.

Proof sketch. The chopping procedure proceeds as follows. At first, we define $\vec{G} = \vec{G}[V]$ as our single component. Then, at each step, we decompose a component $\vec{G}[U]$ into two separate components $\vec{G}[A]$ and $\vec{G}[B]$, if (A, B) is a β -cut of U and Inequality (1) above does *not* apply. When decomposing, we query the edges of the cut (A, B) and mark them as external edges. Note that we need not query any additional edges to decide

on cutting a component, as all the required information is given by the underlying graph G and by the orientation of the external edges that were queried in previous steps. After each stage, we check whether the orientations of the edges queried so far invalidate any of the cuts in the graph (see Section 2), in which case we conclude that \vec{G} is not Eulerian and return the invalid cut.

The procedure terminates once there is no cut of any component that satisfies the chopping conditions. The components are clearly disjoint throughout the procedure. Since we only chopped components across β -cuts, every final component contains at least β edges. Moreover, note that a component is always chopped by the procedure unless all its β -cuts satisfy Inequality (1). Hence, if the algorithm terminates without finding a witness that \vec{G} is not Eulerian, then every G_i is an (α, β) -expander. It remains to prove the upper bound for the number of external edges and the query complexity of the chopping procedure.

Suppose that the chopping procedure has not found a witness that \vec{G} is not Eulerian. Consider a component U and a β -cut (A, B) of U whose edges were queried in some step of the lemma. Using the chopping criterion and the fact that all the cuts in the knowledge graph are valid, we obtain

$$\min \left\{ |\vec{E}(A, B)|, |\vec{E}(B, A)| \right\} < \alpha |E(A)|. \quad (2)$$

We refer to the edges in the minimal cut among $\vec{E}(A, B)$ and $\vec{E}(B, A)$ as *rare edges*, and to the edges in the other direction as *common edges*. We then prove that the total number of rare external edges is $O(\alpha m \log m)$, by “charging” a cost of α on every edge $e \in E(A)$. The proof uses Inequality (2) and the fact that, by definition, $|E(A)| \leq |E(B)|$. To complete the proof of the upper bound, we show that the ratio between the number of common edges and the number of rare edges is $O(m/\beta)$. This is done by observing that the multigraph defined by the components \vec{G}_i is Eulerian, and so decomposable into edge-disjoint directed cycles. Every cycle contains at least one rare edge because the subgraph of common edges is acyclic. The proof follows since the number of components is $O(m/\beta)$. Finally, it is easy to see that the query complexity is not larger in the case where the procedure terminates after finding an invalid cut. \square

Algorithm 15 $CHOP-1(\vec{G}, \epsilon, \alpha, \beta)$:

1. Use Lemma 14 (the chopping lemma) for finding (α, β) -components $\vec{G}_1, \dots, \vec{G}_k$ and querying their external edges, or reject and terminate if an invalid cut is found in the process.
2. Sample $3 \ln 3/\epsilon$ (α, β) -components \vec{G}_i randomly and independently, where the probability of selecting a component \vec{G}_i in a sample is proportional to $m_i \stackrel{\text{def}}{=} |E(V_i)|$.
3. Test every selected component \vec{G}_i using $GEN-1(\vec{G}_i, \alpha, \beta, \epsilon/2)$ (see Lemma 13). Reject if the test rejects for at least one of the components selected.
4. Accept if the input was not rejected by any of the above steps.

Theorem 16. *CHOP-1 is a 1-sided test for being Eulerian with query complexity $O\left(\frac{\alpha m^2 \log m}{\beta} + \frac{\Delta \log m}{\epsilon^2 \alpha} + \frac{\beta \cdot \min\{\beta, \Delta\}}{\epsilon^2}\right)$. In particular, for $\alpha = \frac{(\Delta \log m)^{1/3}}{(\epsilon m)^{2/3}}$ and $\beta = \frac{(\epsilon m \log m)^{2/3}}{\Delta^{1/3}}$, the query complexity is $O\left(\frac{(\Delta m \log m)^{2/3}}{\epsilon^{4/3}}\right)$.*

Finally, we obtain a similar 2-sided test, CHOP-2, by replacing the calls to GEN-1 in Step 3 of CHOP-1 with calls to MULTI-2 and using slightly different constants.

Theorem 17. *CHOP-2 is a 2-sided test for being Eulerian with query complexity $O\left(\frac{\alpha m^2 \log m}{\beta}\right) + \tilde{O}\left(\frac{\sqrt{\Delta} \log m}{\epsilon^2 \alpha} + \frac{\beta \cdot \sqrt{\min\{\beta, \Delta\}}}{\epsilon^2}\right)$. In particular, if $\Delta \leq (\epsilon m)^{4/7}$, then, for $\alpha = \frac{\Delta^{1/6}}{(\epsilon m)^{2/3}}$ and $\beta = \frac{(\epsilon m)^{2/3}}{\Delta^{1/6}}$, the query complexity is $\tilde{O}\left(\frac{\Delta^{1/3} m^{2/3}}{\epsilon^{4/3}}\right) = \tilde{O}\left(\frac{m^{6/7}}{\epsilon^{8/7}}\right)$. If $(\epsilon m)^{4/7} < \Delta \leq m$, then, for $\alpha = \frac{\Delta^{5/16}}{(\epsilon m)^{3/4}}$ and $\beta = \Delta^{1/8} \sqrt{\epsilon m}$, the query complexity is $\tilde{O}\left(\frac{\Delta^{3/16} m^{3/4}}{\epsilon^{5/4}}\right) = \tilde{O}\left(\frac{m^{15/16}}{\epsilon^{5/4}}\right)$.*

7 Concluding comments and open problems

We have shown a test with a sub-linear number of queries for all graphs. However, excepting the special cases of dense graphs and expander graphs, this should be only considered as a first step for this problem.

The procedure of our general test is surprisingly involved considering the problem statement. The question arises as to whether we can reduce the computational complexity from exponential to polynomial in m . Also, to make the test truly attractive, most of the calculations should be performed in a preprocessing stage, where the amount of calculations done while making the queries should ideally be also sub-linear in m .

Related to the preprocessing question is the unresolved question of adaptivity. We would like to think that a sub-linear query complexity *non-adaptive* test also exists for all graphs. Other adaptive versus non-adaptive gaps, such as the one concerning the 2-sided lower bounds, need also be addressed.

References

1. N. Alon and A. Shapira, Testing subgraphs in directed graphs, *J. Comput. Syst. Sci.* 69(3): 354–382, 2004 (a preliminary version appeared in Proc. of STOC 2003, 700–709).
2. L. Babai, On the diameter of Eulerian orientations of graphs, *Proceedings of the 17th SODA*, (2006), 822–831.
3. M. Bender and D. Ron, Testing properties of directed graphs: Acyclicity and connectivity, *Random Structures and Algorithms* (2002), 184–205.
4. E. Ben-Sasson, P. Harsha and S. Raskhodnikova, Some 3CNF properties are hard to test, *SIAM J. Computing*:35(1), 1–21, 2005 (a preliminary version appeared in Proc. 35th STOC, 2003).
5. M. Blum, M. Luby and R. Rubinfeld, Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences* 47 (1993), 549–595 (a preliminary version appeared in Proc. 22nd STOC, 1990).

6. G. R. Brightwell and P. Winkler, Counting Eulerian circuits is #P-complete, *Proc. 7th ALENEX and 2nd ANALCO 2005 (Vancouver BC)*, C. Demetrescu, R. Sedgewick and R. Tamassia, eds, SIAM Press, 259–262.
7. S. Chakraborty, E. Fischer, O. Lachish, A. Matsliah and I. Newman: Testing st -Connectivity, *Proceedings of the 11th RANDOM and the 10th APPROX (2007)*: 380–394.
8. E. Fischer, The art of uninformed decisions: A primer to property testing, *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, G. Paun, G. Rozenberg and A. Salomaa (editors), World Scientific Publishing (2004), Vol. I 229–264.
9. E. Fischer and O. Yahalom, Testing convexity properties of tree colorings, *Proc. of the 24th International Symposium on Theoretical Aspects of Computer Science (STACS 2007)*, LNCS 4393, Springer-Verlag 2007, 109–120.
10. H. Fleishner, *Eulerian graphs and related topics*, Part 1. Vol. 1. Annals of Discrete Mathematics 45, 1990.
11. H. Fleishner, *Eulerian graphs and related topics*, Part 1. Vol. 2. Annals of Discrete Mathematics 50, 1991.
12. O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *JACM* 45(4): 653–750 (1998).
13. O. Goldreich and D. Ron, Property testing in bounded degree graphs, *Algorithmica*, (2002), 32(2):302–343
14. S. Halevy, O. Lachish, I. Newman and D. Tsur, Testing properties of constraint-graphs, *Proceedings of the 22nd IEEE Annual Conference on Computational Complexity (CCC 2007)*, 264–277.
15. S. Halevy, O. Lachish, I. Newman and D. Tsur, Testing orientation properties, technical report, *Electronic Colloquium on Computational Complexity (ECCC)*, (2005), 153.
16. T. Ibaraki, A. V. Karzanov and H. Nagamochi, A fast algorithm for finding a maximum free multiflow in an inner Eulerian network and some generalizations, *Combinatorica* 18(1) (1988), 61–83.
17. T. Kaufman, M. Krivelevich and D. Ron, Tight bounds for testing bipartiteness in general graphs, *SICOMP*, (2004), 33(6):1441–1483.
18. L. Lovász, On some connectivity properties of Eulerian graphs, *Acta Math. Hung.* 28 (1976) 129–138.
19. M. Mihail, P. Winkler, On the number of Eulerian orientations of a graph, *Algorithmica* 16(4/5) (1996), 402–414.
20. I. Newman, Testing of Functions that have small width Branching Programs, *SIAM J. Computing*, 31(5):1557–1570, 2002 (a preliminary version appeared in Proc. 41st FOCS, 2000).
21. M. Parnas and D. Ron, Testing the diameter of graphs, *Random Struct. and Algorithms*, (2002), 20(2):165–183.
22. P. A. Pevzner, H. Tang and M. S. Waterman, An Eulerian path approach to DNA fragment assembly, *Proc. Natl. Acad. Sci. USA* 98, (2001), 9748–9753.
23. R. W. Robinson, Enumeration of Euler graphs, In *Proof Techniques in Graph Theory* (Ed. F. Harary), New York: Academic Press, 1969, 147–153.
24. D. Ron, Property testing (a tutorial), In: *Handbook of Randomized Computing* (S. Rajasekaran, P. M. Pardalos, J. H. Reif and J. D. P. Rolim eds), Kluwer Press (2001), Vol. II Chapter 15.
25. R. Rubinfeld and M. Sudan, Robust characterization of polynomials with applications to program testing, *SIAM J. Computing* 25 (1996), 252–271 (first appeared as a technical report, Cornell University, 1993).
26. W. T. Tutte, *Graph theory*, Addison-Wesley, New York, 1984.