On locally-characterized expander graphs (a survey)

Oded Goldreich*

December 19, 2024

Abstract

We consider the notion of a local-characterization of an infinite family of unlabeled bounded-degree graphs that contains at most one graph of each size. Such a local-characterization is defined in terms of a finite set of (marked) graphs yielding a generalized notion of subgraph-freeness, which extends the standard notions of induced and non-induced subgraph freeness.

We survey the work of Adler, Kohler and Peng (*32nd SODA* and *36th CCC*, 2021), which is pivoted at constructing locally-characterized expander graphs. The construction makes inherent use of the iterative and local nature of the Zig-Zag construction of Reingold, Vadhan, and Wigderson (*41st FOCS*, 2000). This yields a locally-characterizable graph property that cannot be tested (in the bounded-degree graph model) within a number of queries that does not depend on the size of the graph.

A preliminary version of this paper has been posted as TR24-013 of *ECCC*. The most significant revisions are in Section 3.

Contents

*Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. Email: oded.goldreich@weizmann.ac.il. This research was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702).

1 Overview

We consider the notion of a *local-characterization* of an infinite family of unlabeled *bounded-degree* graphs that contains at most one graph of each size. Such a local-characterization is defined in terms of a finite set of (marked) graphs yielding a generalized notion of subgraph-freeness, which extends the standard notions of induced and non-induced subgraph freeness.[1] Intuitively, this notion corresponds to forbidden neighbourhoods of constant distance; that is, we consider the set of graphs such that each vertex in them has a neighborhood that is not forbidden. In other words, a family of graphs $\mathcal{G}$ is locally-characterized by a finite set of marked graphs $\mathcal{F}$ if $\mathcal{G}$ equals the set of $\mathcal{F}$-free graphs (i.e., a graph G is in $\mathcal{G}$ if and only if for every $F \in \mathcal{F}$ the graph G is F-free).

We stress that, for every $\mathcal{F}$ as above, the set of $\mathcal{F}$-free graphs is a graph property; that is, it is closed under isomorphism. We also mention that this definition is related both to expressibility by first-order formula (cf. [1]) and to having a proximity-oblivious tester (cf. [5]). In light of this, one would expect that such graph properties would be testable (in the bounded-degree graph model) within query complexity that depends only on the proximity parameter; however, as shown by Adler, Kohler, and Peng [1], this is not the case. The latter result is proved by constructing a locally-characterizable graph property that consists solely of expander graphs (and using the fact that such a property cannot be tested within query complexity that depends only on the proximity parameter [2]). Actually, this graph property corresponds to an infinite sequence of unlabeled graphs that contains at most one graph of each size (equiv., all n-vertex graphs that have the property are isomorphic to a single n-vertex graph).

Theorem 1.1 (a locally-characterizable sequence of expander graphs [1]): *There exists a finite collection of marked graphs, denoted $\mathcal{F}$, such that the set of $\mathcal{F}$-free graphs is an infinite set of* (bounded-degree) *expander graphs. Furthermore, all n-vertex graphs that are $\mathcal{F}$-free are isomorphic to one another, and for some $d \in \mathbb{N}$ and each $m \in \mathbb{N}$ there exists an $\Theta(d^{4m})$-vertex $\mathcal{F}$-free graph.*

The fact that the set of $\mathcal{F}$-free n-vertex graphs are isomorphic to one another is quite striking, let alone the fact that they are expanders (which in important for the application to property testing). In fact, it is striking that an infinte set of connected graphs can be locally-characterized.

The proof of Theorem 1.1 is pivoted at the Zig-Zag construction of Reingold, Vadhan, and Wigderson [10]. Recall that they presented a sequence of graphs, $(G_i)_{i\in\mathbb{N}}$, such that $G_1 = H^2$ for some constant-size expander H (of degree d and second-eigenvalue at most $d/4$) and $G_i = G_{i-1}^2 ⓩ H$, where ⓩ denotes the Zig-Zag product. Furthermore, each vertex in G_{i-1} is replaced in G_i by a cloud of vertices of the same size as H, and edges of G_{i-1}^2 yield "connections" between the corresponding clouds in G_i (where the edges of the Zig-Zag product correspond to three-step walks on a graph that combine these connections with copies of H that are "placed" on each cloud). Loosely speaking, for each $m \in \mathbb{N}$, the proof of Theorem 1.1 (provided in [1]) identifies a graph that consists of the graphs $G_1, ..., G_m$ along with edges that connect each vertex of G_{i-1} to each vertex in the corresponding cloud of G_i. In fact, the construction is first presented in terms of directed edge-colored graphs, and gadgets are later used to yield undirected graphs (with no colors).

The point is that, while it is inconceivable that we can provide a local-characterization of G_i itself, it is quite conceivable that we can provide a local-characterization of G_i in terms of G_{i-1}. Indeed, the simple relation between G_{i-1} and G_i, already capitalized by Reingold [9], is pivotal here too. Furthermore, the relation between G_{i-1} and G_i can be enforced by constraints that are

[1] This notion, defined in [5], is reviewed in Section 2.2.

independent of i. These facts are the basis for a local charaterization of a graph that has $G_1, ..., G_m$ at its core. Needless to say, materializing this outline requires a careful implementation.

Application to property testing. We say that a property tester has size-oblivious (query) complexity if the number of queries that it makes depends only on the proximity parameter. The main result of [2] states that, in the bounded-degree graph model, every infinite property of graphs that has a tester of size-oblivious query complexity must contain an infinite hyperfinite subproperty. Recalling that hyperfinite graphs are the "extreme opposite" of expander graphs, it follows that the property asserted in Theorem 1.1 does not have a tester of size-oblivious complexity. Hence, we get

Corollary 1.2 (a locally-characterizable property that is not testable in the bounded degree graph model within size-oblivious complexity [1]): *There exists a finite collection of marked graphs, denoted $\mathcal{F}$, such that the set of $\mathcal{F}$-free graphs does not have a tester of size-oblivious query complexity* (in the bounded degree graph model). *In particular, this set has no proximity-oblivious tester* (in the bounded degree graph model).

Recall that a proximity-oblivious tester (for a graph property) is an oracle machine of constant query-complexity that accepts each graph in the property with probability 1, and rejects each graph that is not in the property with probability that is related to the (relative) distance of the graph from the property (cf. [3, Def. 1.7]). Hence, a proximity-oblivious tester for a property yields a tester of size-oblivious query complexity for that property (cf. [3, Thm. 1.9]).

Comment. In light of Theorem 1.1, one may wonder which sequences of graphs that contain at most one unlabelled graph of each size can be locally-characterized. As pointed out by Madhu Sudan, a simple example is any sequence of graphs such that each graph consists of isolated copies of some fixed graph H. Other examples, which utilize the strategy underlying the proof of Theorem 1.1, are presented in Section 4. Some of these graphs are expanders and some are not.

Organization. After recalling the background (Section 2), we turn to the core of this survey (Section 3), where we provide more details on the proof of Theorem 1.1. We end with some afterthoughts (Section 4).

2 Background

Throughout this text, we focus on bounded-degree graphs. While the contents of Section 2.1 is well-known and may be skipped, the notion of generalized subgraph freeness (reviewed in Section 2.2) is likely to be unfamiliar to most readers. For sake of self-containment, we also review the Zig-Zag product (in Section 2.3).

2.1 Testing in the Bounded-Degree Graph Model

(This model was introduced in [4] and is reviewed in [3, Chap. 9].)

The bounded-degree graph model refers to a fixed (constant) degree bound, denoted $d \geq 2$. In this model, a graph $G = (V, E)$ of maximum degree d is represented by the incidence function

$g : V \times [d] \to V \cup \{\perp\}$ such that $g(v, j) = u \in V$ if u is the j^{th} neighbor of v and $g(v, j) = \perp \notin V$ if v has less than j neighbors.[2] Distance between graphs is measured in terms of their foregoing representation; that is, as the fraction of (the number of) different array entries (over $d \cdot |V|$).

The tester is given oracle access to the representation of the input graph (i.e., to the incidence function g), where for simplicity we assume that $V = [n]$ for $n \in \mathbb{N}$. In addition, the tester is also given a proximity parameter ϵ and a size parameter (i.e., n). Recall that graph properties are sets of graphs that are closed under isomorphism.

Definition 2.1 (property testing in the bounded-degree graph model): *For a fixed $d \in \mathbb{N}$, a* tester for the graph property Π *is a probabilistic oracle machine T that, on input a* proximity parameter $\epsilon > 0$ *and* size parameter $n \in \mathbb{N}$, *and when given oracle access to an incidence function $g : [n] \times [d] \to [n] \cup \{\perp\}$, outputs a binary verdict that satisfies the following two conditions:*

1. *The tester accepts each graph $G = ([n], E)$ in Π with probability at least $2/3$; that is, for every $g : [n] \times [d] \to [n] \cup \{\perp\}$ representing a graph in Π* (and every $\epsilon > 0$)*, it holds that $\Pr[T^g(n, \epsilon) = 1] \geq 2/3$.*

2. *Given $\epsilon > 0$ and oracle access to any graph G that is ϵ-far from Π, the tester rejects with probability at least $2/3$; that is, for every $g : [n] \times [d] \to [n] \cup \{\perp\}$ that represents a graph that is ϵ-far from Π, it holds that $\Pr[T^g(n, \epsilon) = 0] \geq 2/3$, where the graph represented by g is* ϵ-far from Π *if for every $g' : [n] \times [d] \to [n] \cup \{\perp\}$ that represents a graph in Π it holds that $|\{(v, j) \in V \times [d] : g(v, j) \neq g'(v, j)\}| > \epsilon \cdot dn$.*

The tester is said to have one-sided error probability *if it always accepts graphs in Π; that is, for every $g : [n] \times [d] \to [n] \cup \{\perp\}$ representing a graph in Π* (and every $\epsilon > 0$)*, it holds that $\Pr[T^g(n, \epsilon) = 1] = 1$.*

The query complexity of a tester for Π is a function (of the parameters d, n and ϵ) that represents the number of queries made by the tester on the worst-case n-vertex graph of maximum degree d, when given the proximity parameter ϵ. Fixing d, we typically ignore its effect on the complexity (equiv., treat d as a hidden constant). Our focus here is on cases in which the query complexity depends only on the proximity parameter (i.e., size-oblivious query complexity).

2.2 Generalized Subgraph Freeness Properties

The notion of a *generalized subgraph-freeness*, which extends the standard notions of induced and non-induced subgraph freeness, was introduced in [5]. It is aimed to capture what one can see by exploring a constant-radius neighborhood of a vertex in a graph that has some predetermined graph property. The issue is that some vertices are fully explored (i.e., the explorer sees all their neighbors), whereas for other vertices (at the boundary of the exploration) the explorer may only encounter them but not all their neighbors (since it has not traversed their incident edges).

We shall actually consider the set of subgraphs that the explorer cannot encounter when exploring a graph that has the property, where these forbidden subgraphs are represented by *marked graphs*, which are graphs in which each vertex is marked either `full` or `semi-full` or `partial`. Intuitively, the marking `full` represent a vertex that is not at the boundary of the exploration, which means that all its incident edges were traversed. In contrast, vertices at the boundary are

[2] For simplicity, we adopt the standard convention by which the neighbors of v appear in arbitrary order in the sequence $(g(v, 1), ..., g(v, \deg(v)))$, where $\deg(v) \stackrel{\text{def}}{=} |\{j \in [d] : g(v, j) \neq \perp\}|$.

marked as `partial`, whereas the marking `semi-full` is inessential (and is included for sake of greater flexibility (see Footnote 3)).

Definition 2.2 (marked graphs, embedding, and generalized subgraph freeness): *A* marked graph *is a pair consisting of a graph and a marking of its vertices such that each vertex is marked either* `full` *or* `semi-full` *or* `partial`. *We say that a marked graph $F = ([h], A)$ can be* embedded *in a graph $G = ([N], E)$ if there exists a 1-1 mapping $\phi : [h] \to [N]$ such that for every $i \in [h]$ the following two conditions hold:*

1. *If i is marked* `full`*, then ϕ yields a bijection between the set of neighbors of i in F and the set of neighbors of $\phi(i)$ in G. That is, in this case $\Gamma_G(\phi(i)) = \phi(\Gamma_F(i))$, where $\Gamma_X(v)$ denotes the set of neighbors of v in the graph X, and $\phi(S) = \{\phi(v) : v \in S\}$.*

2. *If i is marked* `semi-full`*, then ϕ yields a bijection between the set of neighbors of i in F and the set of neighbors of $\phi(i)$ in the subgraph of G induced by $\phi([h])$. That is, in this case $\Gamma_G(\phi(i)) \cap \phi([h]) = \phi(\Gamma_F(i))$.*

3. *If i is marked* `partial`*, then ϕ yields an injection of the set of neighbors of i in F to the set of neighbors of $\phi(i)$ in G. That is, in this case $\Gamma_G(\phi(i)) \supseteq \phi(\Gamma_F(i))$.*

*The graph G is called F-*free *if F cannot be embedded in G* (i.e., there is no embedding of F in G that satisfies the foregoing conditions). *For a set of marked graphs $\mathcal{F}$, a graph G is called $\mathcal{F}$-*free *if for every $F \in \mathcal{F}$ the graph G is F-free.*

Indeed, the standard notion of (non-induced) subgraph freeness is a special case of generalized subgraph freeness, obtained by considering the corresponding marked graph in which all vertices are marked `partial`. Similarly, the notion of induced subgraph freeness is a special case of generalized subgraph freeness, obtained by considering the corresponding marked graph in which all vertices are marked `semi-full`.[3]

Marking vertices as `full` introduces a new type of constraint; specifically, this constraint mandates the non-existence of neighbors that are outside the embedding of the marked subgraph. For example, using vertices that are marked `full`, it is possible to disallow certain degrees in the graph (see Example 2.3). Thus, the generalized notion of subgraph freeness includes properties that are not hereditary (e.g., regular graphs), whereas induced and non-induced subgraph freeness are hereditary.

Example 2.3 (disallowing certain degrees via generalized subgraph freeness): *For every $d' \in \{0, 1, ..., d\}$, we can disallow vertices of degree d' by using a $(d'+1)$-vertex star in which the center is marked* `full` *and the d' leaves are marked* `partial`.

The foregoing example as well as the next one are actually used in the proof of Theorem 1.1. The following example refer to the case that we want to mandate that if the graph contains some fixed subgraph H' then it actually contains certain additional edges on the same vertices.

[3] Indeed, the `semi-full` marking (resp., the `partial` marking) can be avoided by emulating marked graphs by sets of mark graphs that use only `full` and `partial` (resp., `semi-full`) marking. Emulating the `partial` marking by `semi-full` marking is analogous to the emulation of non-induced subgraph freeness by induced subgraph freeness. As for emulating the `semi-full` marking, here we replace each marked graph F by a set of marked graphs $\mathcal{F}'$ such that each $F' \in \mathcal{F}'$ consists of a copy of F in which all `semi-full`-marked vertices are replaced by `full`-marked vertices and are connected to some auxiliary vertices, which are all marked `partial`. We stress that $\mathcal{F}'$ reflects all possible ways of connecting the newly `full`-marked vertices with the auxiliary vertices.

Example 2.4 (mandating some subgraph via generalized subgraph freeness): *Let* $H' = ([h], A')$ *be a subgraph of* $H = ([h], A)$*, and suppose that we want to enforce that every induced subgraph of* G *that contains* H' *also contains* H*. This can be obtained by requiring* G *to be* $\mathcal{F}$*-free, where* $\mathcal{F}$ *is the set of all marked* h*-vertex graphs that are consistent with* H' *but not with* H*. Specifically,* F *is in* $\mathcal{F}$ *if* F *is embedded in every* h*-vertex graph that contains* H' *but not* H*.*

For sake of completeness, we present the following definition, which we actually use only in headings.

Definition 2.5 (locally characterizable properties): *A graph property* Π *is called* locally characterizable *if there exists a finite set of marked graphs* $\mathcal{F}$ *such that* Π *equals the set of* $\mathcal{F}$*-free graphs.*

(We mention that Definition [5, Def. 5.2] is more general; it allows for a different set of marked graphs to be used for each graph-size as long as there is a uniform upper bound on the size of all marked graphs that are used.)

2.3 The Zig-Zag Product

(The Zig-Zag product was introduced and first studied in [10].)

Given a (big) D-regular graph $G = (V, E)$, and a (small) d-regular graph $H = ([D], F)$, their Zig-Zag product, denoted $G \circledz H$, consists of the vertex set $V \times [D]$, which is partitioned to D-vertex clouds such that the cloud that corresponds to vertex $v \in V$ is the set of vertices $C_v = \{(v, i) : i \in [D]\}$, and edges that correspond to certain 3-step walks (as detailed next).

Actually, it is instructive to first consider the graph, denoted $G \circledr H$, in which copies of H are placed on the clouds (i.e., for every $v \in V$ and $\{i, j\} \in F$ we place the intra-cloud edge $\{(v, i), (v, j)\}$), and edges of G connect the corresponding clouds by using corresponding edges; that is, if $\{u, v\} \in E$ is the i^{th} (resp., j^{th}) edge incident at u (resp., at v), then we place the inter-cloud edge $\{(u, i), (v, j)\}$. Note that each vertex in $G \circledr H$ has d intra-cloud edges and a single inter-cloud edge. Now, the edges of $G \circledz H$ correspond to 3-step walks in $G \circledr H$ that start with an intra-cloud edge, then take the (only available) inter-cloud edge, and lastly take some intra-cloud edge; that is, such a generic walk has the form $(v, i) \to (v, j) \to (w, k) \to (w, \ell)$, where $\{i, j\}, \{k, \ell\} \in F$ and $\{(v, j), (w, k)\}$ is an inter-cloud edge in $G \circledr H$ (i.e., $\{v, w\} \in E$ is the j^{th} edge incident at v and the k^{th} edge incident at w).

We shall assume that both G and H are connected and are not bipartite. In that case, it is clear that the graph $G \circledr H$ is also connected and non-bipartite, and it can be shown that also $G \circledz H$ has these properties. The main technical result of [10] asserts that the convergence rate of a random walk on $G \circledz H$ (a.k.a the relative second eigenvalue of the graph) can be upper-bounded in terms of the convergence rates of random walks on G and on H. A simple form of their bound asserts that $\lambda(G \circledz H) \leq \lambda(G) + \lambda(H)$, where $\lambda(X)$ denotes the convergence rate of a random walk on the graph X. Using $\lambda(H) \leq 1/4$, it follows that if $\lambda(G) \leq 1/2$, then $\lambda(G^2 \circledz H) \leq 1/2$.

The Zig-Zag construction. The last assertion is relevant to the Zig-Zag construction, which is captured by the equation $G_i = G_{i-1}^2 \circledz H$. Recall that we shall use $D = d^4$ and $G_1 = H^2$; hence, G_i is a d^2-regular d^{4i}-vertex graph.

3 On the Proof of Theorem 1.1

We shall first present a construction of a directed graph with edge-colors such that the corresponding underlying graph is an expander. This construction will be presented in terms of local conditions that the edges of the graph are required to satisfy, where the local conditions are enforced by forbidden neighborhoods of constant distance (akin those in Examples 2.3 and 2.4). Indeed, the forbidden neighborhoods correspond to directed and edge-colored versions of marked graphs, which are defined analogously to the Definition 2.2.

Recall that, for each $m \in \mathbb{N}$, we construct an $O(d^{4m})$-vertex graph that consists of the graphs $G_1, ..., G_m$ (of the Zig-Zag construction) along with edges that connect each vertex of G_{i-1} to each vertex in the corresponding cloud of G_i. Hence, the construction consists of two parts: (1) edges that represent the edges of the graphs $G_1, ..., G_m$, and (2) edges that form a d^4-ary tree in which each vertex in G_{i-1} is connected to all vertices of the corresponding cloud of G_i. Below, we show how this structure is enforced by postulates that can be expressed by local conditions.

The main part of the construction consists of directed edges that represents the edge-rotation functions of the graphs $G_1, ..., G_m$ in the Zig-Zag construction. Recall that the edge-rotation function of an undirected graph extend its adjacency function such that the pair (u, α) is mapped to the pair (v, β) if the α^{th} outgoing edge of u equals the β^{th} incoming edge of v (equiv., the α^{th} port of vertex u is connected to the β^{th} port of vertex v). In such a case, we shall color the directed edge (u, v) with the color (α, β).

Recall that H is a d-regular d^4-vertex graph and that $G_1 = H^2$ and $G_i = G_{i-1}^2 ⓩ H$ are d^2-regular d^{4i}-vertex graphs. For every $\alpha, \beta \in [d^2]$, we consider the edge-set $E_{\alpha,\beta}$ such that $(u, v) \in E_{\alpha,\beta}$ if for some i there exists an edge in G_i that connects the α^{th} port of vertex u to the β^{th} port of vertex v. Indeed, $E_{\alpha,\beta}$ is viewed as a set of directed edges that are colored (α, β), and we *postulate that* $(u, v) \in E_{\alpha,\beta}$ *if and only if* $(v, u) \in E_{\beta,\alpha}$. Letting $E \stackrel{\text{def}}{=} \bigcup_{\alpha,\beta\in[d^2]} E_{\alpha,\beta}$, we refer to $(u, v) \in E$ as an E-edge (and to $(u, v) \in E_{\alpha,\beta}$ as an $E_{\alpha,\beta}$-edge).

We stress that the foregoing anti-parallel postulate is a very minimal one and far more substantial conditions will be postulated about the E-edges by referring also to other edges that will induce a layered directed d^4-ary tree (with G_i identified with the i^{th} layer). Indeed, the actual structure of the graphs $G_1, ..., G_m$ will be enforced by relating each G_i to G_{i-1}.

As a warm-up, suppose that we want to augment the graph with auxiliary (colored) edges that will capture 2-step walks on the E-edges. In such a case, we introduce, for every $\alpha, \beta, \gamma, \delta \in [d^2]$, an edge-set $E'_{(\alpha,\gamma),(\delta,\beta)}$ such that $(u, w) \in E'_{(\alpha,\gamma),(\delta,\beta)}$ if and only if there exists v such that $(u, v) \in E_{\alpha,\beta}$ and $(v, w) \in E_{\gamma,\delta}$. (We stress that the latter condition is a local condition about the edge-sets $E'_{(\alpha,\gamma),(\delta,\beta)}$, $E_{\alpha,\beta}$ and $E_{\gamma,\delta}$; actually, we will use E' only as a shorthand.)[4]

As stated above, the structure of the graphs $G_1, ..., G_m$ is enforced by relating each G_i to G_{i-1}, for each $i \in [m]$, where we define G_0 to be the graph consisting of a single vertex. The first step towards enforcing this relation is the association of vertices in G_{i-1} with clouds of vertices in G_i such that each cloud contains d^4 vertices that are identified (equiv., ordered) within the cloud; that is, the d^4 ports of each vertex in G_{i-1}^2 are associated with distinct vertices of the corresponding cloud. This association is enforced by using edges that are directed from each vertex of G_{i-1} to the

[4] An alternative presentation may use E' explicitly. In such a presentation $G_i = G_{i-1}^2 ⓩ H$ is decomposed into $G'_{i-1} = G_{i-1}^2$ and $G_i = G'_{i-1} ⓩ H$. In this case (assuming we keep $G_1 = H^2$ at level 1), odd (resp., even) levels of the tree will consist of copies of the G_i's (resp., G'_i's), and tree edges of a different color will be used to connect vertices of G_i to their copy in G'_i. Such an alternative presentation makes the postulates that related G_i to G_{i-1} simpler, but this comes at a cost of a slightly more complicated postulates regarding the tree edges (which are treated next).

corresponding cloud of G_i such that these d^4 edges are assigned different colors. Specifically, for each $\sigma \in [d^4]$, we introduce a set of directed edges, denoted P_σ, and *postulate that each vertex has at most one outgoing P_σ-edge and at most one incoming P-edge*, where $P \stackrel{\text{def}}{=} \bigcup_{\sigma\in[d^4]} P_\sigma$. Indeed, $(u,v) \in P_\sigma$ implies that v is the σ^{th} vertex in the cloud associated with u, where u is the "parent" of v in the directed tree induced by P. Additional postulates are added to identify the vertices of G_0 and G_m; specifically:

1. Intuitively, we postulate that there exists a single vertex with no incoming P-edges; the graph G_0 will consist of this vertex.

   Actually, *we postulate that there exists at most one vertex with no incoming P-edges*; the existence of such a vertex will follows from the tree structure of the P-edges (see below).

2. We postulate that the P-outdegree of each vertex is either 0 or d^4 (equiv., *each vertex either has no outgoing P-edges or has at least d^4 outgoing P-edges*).

3. Intuitively, we postulate that all vertices that have no outgoing P-edges belong to the same G_i (and it will follow that $i = m$).

   Actually, we *postulate that vertices that are connected by E-edges have the same number of outgoing P-edges* (equiv., the same number of outgoing P_σ-edges, for every $\sigma \in [d^4]$). The fact that vertices with no outgoing P-edges are in G_m follows from the first item (i.e., for $i \geq 1$, the graph G_i cannot contain vertices with no incoming P-edges).

Combining the foregoing postulates with additional postulates that refer to E-edges, this implies that the P-edges form a directed d^4-ary tree such that all leaves are at the same distance from the root. We warn that establishing this tree structure is the most complex part of the proof of [1, Thm. 3.1].

Next, we *postulate that the E-edges between the d^4 vertices that neighbor the single vertex of P-indegree 0* (i.e., the vertex of G_0) *form a copy of H^2*. Specifically, recalling that these vertices are identified by their incoming P-edges, we postulate that $(u,v) \in E$ if and only if there exist $\sigma, \tau \in [d^4]$ such that u (resp., v) has an incoming P_σ-edge (resp., P_τ-edge) from G_0 and $\{\sigma, \tau\}$ is an edge in H^2. Furthermore, in this case $(u,v) \in E_{\alpha,\beta}$ if and only if the foregoing edge in H^2 uses the α^{th} port of u and the β^{th} port of v (for some $\alpha, \beta \in [d^2]$).

The main issue is relating the E-edges of $G_i = G_{i-1}^2 \textcircled{z} H$ to those of G_{i-1}, for $i > 1$. We stress that i itself cannot and is not referred to in this enforcement. Instead, we refer to any $(x,y) \in E$ such that x and y have outgoing P-edges and introduce conditions on the opposite endpoints of these P-edges; that is, we mandate E-edges among d of the P-neighbors of x (which reside in the cloud that replaces x) and d vertices of the P-neighbors of y (which reside in the cloud that replaces y). Specifically, for $(j,k) \in [d]^2 \equiv [d^2]$, we postulate that $(u,v) \in E_{(j,k),(k,j)}$ if and only if there exist $\sigma, \tau, \sigma', \tau' \in [d^4]$ and $(u',v') \in E'_{\sigma',\tau'}$ (i.e., u' and v' are connected by a 2-path colored (σ', τ') (see warm-up)) such that

1. $\{\sigma, \sigma'\}$ is an edge colored j in H.

2. $\{\tau, \tau'\}$ is an edge colored k in H.

3. $(u', u) \in P_\sigma$ and $(v', v) \in P_\tau$.

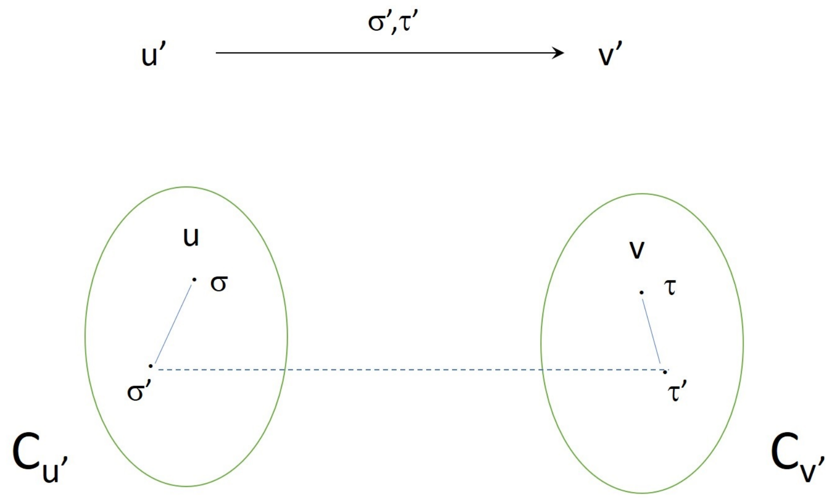

Figure 1: Vertex u (resp., v) is the σ^{th} (resp., τ^{th}) vertex in the cloud $C_{u'}$ (resp., $C_{v'}$) that replaces u' (resp., v'); these clouds are connected by an edge colored σ', τ'.

Intuitively, these conditions (on $(u,v) \in E$) imply that, for some i, the P-parent of u (denoted u') is connected in G_{i-1}^2 to the P-parent of v (denoted v'); see Figure 1. Furthermore, u (resp., v) is associated with the σ^{th} (resp., τ^{th}) vertex of the cloud of u (resp., v), and σ (resp., τ) neighbors σ' (resp., τ') in H. Moreover, for $\sigma \equiv (\alpha, \gamma) \in [d^2]^2$, vertices u' and v' are connected in G_{i-1} by a 2-path that uses the port $\alpha \in [d^2]$ of u' and the port $\gamma \in [d^2]$ of the intermediate vertex (whereas $\tau' = (\delta, \beta)$ such that δ and β are the ports used in walking this 2-path in the opposite direction). Indeed, the 2-paths referred to here are the edges of $E' \stackrel{\text{def}}{=} \bigcup_{\sigma',\tau' \in [d^4]} E'_{\sigma',\tau'}$, which were defined in the warm-up.

The foregoing description suggests that the P-edges of a graph that satisfies the listed postulates form a d^4-ary directed tree such that all leaves are at the same distance from the root, and that the subgraph (of E-edges) induced by the set of vertices that are at distance i from the root equals G_i. This is indeed the case, but proving the former fact (which refers to the P-edges) requires using also the postulates that refer to the E-edges.[5] This is core of the analysis provided in [1, Sec. 3.1]. Hence, we have

Lemma 3.1 (the postulated conditions determine a unique unlabeled directed n-vertex graph): *For $n = \sum_{i=0}^{m} d^{4i}$, an n-vertex* (unlabeled edge-colored) *directed graph satisfies the foregoing conditions*[6] *if and only if it consists of the graphs $G_0, G_1, ..., G_m$* (such that $G_1 = H^2$ and $G_i = G_{i-1}^2 \textcircled{z} H$)

[5]Specifically, the postulates that refer to P-edges only enforce that the corresponding graph consists of at most one directed tree and a collection of directed cycles such that each vertex on each cycle is a root of a directed tree. However, the postulates that refer to E-edges imply that no such cycles exist, and so the graph consists of a single directed tree.

[6]Following is a compilation of all the conditions.

- *The edge-rotation* (i.e., anti-parallel) *condition*: $(u,v) \in E_{\alpha,\beta}$ if and only if $(v,u) \in E_{\beta,\alpha}$.
- *The parental* (i.e., P-edges) *conditions*:
  - For every $\sigma \in [d^4]$ and every u, there exists at most one v such that $(u,v) \in P_\sigma$.
  - For every v, there exists at most one u such that $(u,v) \in P$.

that are connected by P-edges as outlined above (i.e., each vertex in G_{i-1} is connected by an P_σ-edge to the σ^{th} vertex in the corresponding cloud of G_i). *If $n > 1$ is not of the foregoing form, then no n-vertex graph satisfies these conditions.*

We note that the foregoing conditions can be enforced by forbidden neighborhoods of constant distance (akin those in Examples 2.3 and 2.4). Indeed, the forbidden neighborhoods correspond to directed and edge-colored versions of marked graphs, which are defined analogously to the Definition 2.2.

The foregoing n-vertex graph (consisting of $G_0, G_1, ..., G_m$ and the P-edges) has constant degree. We also observe that the corresponding undirected graph is an expander, by using the combinatorial notion of expansion. This is the case because each of the G_i's is an expander, whereas each vertex in G_{i-1} is connected to d^4 different vertices in G_i. Hence, for every set of vertices S and each i, letting S_i denote the vertices of S that reside in the d^{4i}-vertex graph G_i, we observe that if $|S_{i-1}| \leq d^{4(i-1)}/2$ then S_{i-1} contributes to the expansion inside G_{i-1} and otherwise S_{i-1} neighbors $d^4 \cdot |S_{i-1}| > d^{4i}/2$ vertices in G_i.

Lastly, we observe that the foregoing construction can be converted to the context of (simple) undirected graphs (with no edge-colors). This is done by replacing each color class and edge-direction by a different asymmetric gadget such that the gadgets are non-isomorphic and their vertices can be distinguished from the original ones. In particular, we may use gadgets that contain vertices of higher degree than the degree of the original vertices. The same transformation is applied to the (directed and edge-colored) forbidden neighborhoods that enforce the conditions imposed on the (directed and edge-colored) construction. This yields a corresponding finite set of marked graphs, denoted $\mathcal{F}$, that satisfies the following –

Proposition 3.2 (a locally-characterizable property of expander graphs): *For the foregoing set of mark graphs $\mathcal{F}$, the set of $\mathcal{F}$-free graphs is an infinite set of expander graphs. Furthermore, this set contains a single unlabeled $\Theta(d^{4m})$-vertex graph for every $m \in \mathbb{N}$.*

We mention that, by using additional constraints, one can force these expanders to be regular graphs. In fact, this is done in [1].

4 Digest and Afterthoughts

The proof of Lemma 3.1 does not refer to the fact that H and the G_i's are expanders. It actually uses only the fact that each of the G_i's is a connected graph and that G_i is derived from G_{i-1}

- There exists at most one v such that for every u it holds that $(u, v) \notin P$.
- For every u, the set $\{v : (u, v) \in P\}$ is either empty or has size d^4, where in the latter case for every $\sigma \in [d^4]$ there exists a (unique) v such that $(u, v) \in P_\sigma$.
- For every $(u, u') \in E$ it holds that $|\{v : (u, v) \in P\}| = |\{v' : (u', v') \in P\}|$.

- *The base graph* ("level one") *condition*: Let r denote the vertex that has no incoming P-edges, and v_σ be such that $(r, v_\sigma) \in P_\sigma$ (for $\sigma \in [d^4]$). Then, $(v_\sigma, v_\tau) \in E_{(j,k),(k,j)}$ if and only if there exists $\rho \in [d^4]$ such that $\{\sigma, \rho\}$ is an edge colored j in H and $\{\rho, \tau\}$ is an edge colored k in H.
- *The Zig-Zag condition*: For $(j, k) \in [d]^2 \equiv [d^2]$, we postulate that $(u, v) \in E_{(j,k),(k,j)}$ if and only if there exist $\sigma, \tau, \sigma', \tau' \in [d^4]$ and $(u', v') \in E'_{\sigma',\tau'}$ (i.e., u' and v' are connected by a 2-path colored (σ', τ') (see warm-up)) such that $\{\sigma, \sigma'\}$ is an edge colored j in H, $\{\tau, \tau'\}$ is an edge colored k in H, $(u', u) \in P_\sigma$, and $(v', v) \in P_\tau$.

by replacing vertices with clouds and replacing edges with specific bipartite graphs between the corresponding clouds. Specifically, the σ^{th} vertex of the cloud of u' is connected in G_i to the τ^{th} vertex of the cloud of v' if and only if u' and v' are connected in G_{i-1}^2 via ports σ' and τ' respectively, and some condition regarding $\sigma, \tau, \sigma', \tau' \in [d^4]$ holds. Hence, Lemma 3.1 remains valid also when replacing the recursion $G_i = G_{i-1}^2 \textcircled{z} H$ by the recursion $G_i = f(G_{i-1})$, provided that f satisfies the following conditions:

1. For some constant $c > 1$, each vertex of G' is mapped to a cloud of c vertices in $G = f(G')$, and the vertices of these clouds are ordered (equiv., identified by indices in $[c]$).

   We stress that each vertex of G belongs to a single cloud.

2. If G' is connected, then so is $G = f(G')$.

3. Let u and v be vertices in G such that u (resp., v) is the σ^{th} vertex of the cloud u' (resp., the τ^{th} vertex of the cloud v'). Then, there exists an edge between the α^{th} port of u and the β^{th} port of v if and only if the (unlabeled) constant-distance neighborhoods of u' and v' (in G') satisfy a condition that depends on σ, τ, α and β only. (The condition may refer to the port indices of additional vertices in G', but not to the labels of any of the vertices.)

In addition, G_1 may be an arbitrary constant-sized connected graph; if the number of vertices in G_1 is different from c, then the postulate regarding the outgoing P-edges of the root of the P-tree should be revised accordingly (i.e., the root of the P-tree would have a different degree (which matches G_1) than all internal vertices in the tree (which still have degree c)).

Note that the construction schema outlined above (which yields a graph consisting of $G_0, G_1, ..., G_m$ and P-edges from vertices of G_{i-1} to the corresponding clouds of G_i) always yields graphs of logarithmic diameter (by virtue of the P-edges). In Section 3, we used $G_1 = H^2$ and $f(G) = G^2 \textcircled{z} H$, for some c-vertex graph H (i.e., H is a sufficiently strong d-regular expander and $c = d^4$). These specific choices were important only for asserting that the G_i's are (bounded-degree) expanders, which was important for Proposition 3.2 and Corollary 1.2. Before turning to Corollary 1.2, we present a couple of other incarnations of the foregoing construction schema.

Example 4.1 (using paths and obtaining cycles): *Let $c, \ell \geq 3$ be integers and G_1 be the ℓ-vertex cycle. Then, for any 2-regular graph G, let $f(G)$ the 2-regular graph obtained from G by replacing each edge by a c-long path. Alternatively, consider replacing each vertex by a c-long path and connecting the first vertex of the path replacing u to the last vertex of the path replacing v if and only if the first port of u is connected in G to the second port of v. Then, this f satisfies the foregoing conditions, and applying the construction schema results in a sequence of graphs such that the i^{th} graph is a $\ell \cdot c^{i-1}$-vertex cycle.*

Needless to say, this incarnation of the construction schema is very different from the one presented in Section 3. In contrast, an incarnation that is very similar to the one presented in Section 3 follows.

Example 4.2 (using arbitrary d-regular connected graphs): *For any $d \geq 2$ and $t \geq 1$, let G_1 be an arbitrary $(d+1)$-regular connected graph, and H be an arbitrary connected $(d+1)^t$-vertex d-regular graph. In particular, for $d = 2$ the graph H is a 3^t-cycle, whereas for $t = 1$ the graph H is a $(d+1)$-clique. Then, letting $f(G) = G^t \textcircled{r} H$, where G is any $(d+1)$-regular graph and $\textcircled{r}$ denotes* *the replacement product* (outlined in Section 2.3)*, satisfies the foregoing conditions. Applying the construction schema results in a sequence of graphs such that the* i^{th} *graph is a* $(d+1)$*-regular* $(d+1)^{(i-1)\cdot t}\cdot|G_1|$*-vertex graph, where* $|G_1|$ *is the size of* G_1*.*

We stress that Example 4.2 makes sense even for $d = 2$ and $t = 1$; in that case, the graph resulting from the construction schema (applied to Example 4.2) is hyperfinite,[7] which implies that the corresponding graph property is testable within size-oblivious complexity (see [8]). In contrast, for adequate choice of d, H and G_1, even when using $t = 3$, the corresponding G_i's are expanders, and so the graph resulting from the construction schema is an expander,[8] which implies that the corresponding graph property is not testable within size-oblivious complexity (see [2]).

The last assertion leads ius to Corollary 1.2. Recall that Corollary 1.2 was proved by combining Proposition 3.2 with the fact that the corresponding graph property (which contains only expander graphs) does not contain an infinite hyperfinite subproperty, and so cannot have a tester of size-oblivious query complexity [2]. In the preleminary version of this survey, we expressed the belief that Corollary 1.2 can be proved also without relying on [2]. This belief was proved correct in [6], which establishes an explicit lower bound on the query complexity of testing the graph property that underlies Proposition 3.2. We comment that the proof of [6] extends to any family of expanders that results from the foregoing construction schema.

Acknowledgements

We are grateful to Irit Dinur, Madhu Sudan and Avi Wigderson for useful discussions.

References

[1] Isolde Adler, Noleen Kohler and Pan Peng. On Testability of First-Order Properties in Bounded-Degree Graphs and Connections to Proximity-Oblivious Testing. In `arXiv:2304.03810` [cs.LO]. Combines preliminary versions that appeared in *SODA21* and *CCC21*.

[2] Hendrik Fichtenberger, Pan Peng, and Christian Sohler. Every testable (infinite) property of bounded-degree graphs contains an infinite hyperfinite subproperty. In *30th SODA*, pages 714–726, 2019.

[7] That is, for every $\epsilon > 0$, omitting an ϵ fraction of the edges of the graph (which consists of $G_0, G_1, ..., G_m$ and the P-edges) yields a graph with connected components of size $O(1/\epsilon)$. The key observation here is that in case $t = 1$ there is an edge between two clouds of G_{i+1} (if and) only if there is an edge between the corresponding vertices of G_i. This observation allows to disconnect the subtree that are rooted at vertices of G_i by omitting all edges in G_i as well as all edges that "replace" these edges in $G_{i+1}, ..., G_m$. Specifically, letting $\ell = \log_{d+1}(1/\epsilon) + O(1)$, we omit all edges that are incidence at vertices in $G_0, G_1, ..., G_{m-\ell}$ (i.e., replacing these graphs by isolated vertices) and replacing $G_{m-\ell+1}, ..., G_m$ with $G'_{m-\ell+1}, ..., G'_m$ such that $G'_{m-\ell+i} = G'_{m-\ell+i-1} \circledR H$ for every $i \in [\ell]$ (and $G'_{m-\ell}$ consists of isolated vertices). In other words, the only remaining edges are those that have both endpoints in $G'_{m-\ell+1}, ..., G'_m$, whereas $G'_{m-\ell+1}$ consists of isolated copies of H. Thus, the remaining edges are internal to the $(\sum_{i\in[\ell]}(d+1)^i)$-sized connected components that evolve from the isolated vertices of $G'_{m-\ell}$. (Note that the size of $G_{m-\ell}$ is $(d+1)^{m-\ell-1}$ times the size of G_1.)

[8] The proof of expansion relies on $\lambda(G \circledR H) \leq 1 - \frac{(1-\lambda(H)^2)\cdot(1-\lambda(G))}{2}$ (see [7, Thm. 9.1 (iii)] and the comment at the bottom of [7, p. 510]) and on $\lambda(H) < 0.1$.

[3] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

[4] Oded Goldreich and Dana Ron. Property Testing in Bounded Degree Graphs. *Algorithmica*, Vol. 32 (2), pages 302–343, 2002.

[5] Oded Goldreich and Dana Ron. On Proximity Oblivious Testing. *SIAM Journal on Computing*, Vol. 40, No. 2, pages 534–566, 2011.

[6] Oded Goldreich. On the query complexity of testing local graph properties in the bounded-degree graph model. *ECCC*, TR24-047, 2024.

[7] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications *Bull. Amer. Math. Soc.*, Vol. 43, pages 439–561, 2006.

[8] Ilan Newman and Christian Sohler. Every Property of Hyperfinite Graphs Is Testable. *SIAM Journal on Computing*, Vol. 42 (3), pages 1095–1112, 2013.

[9] Omer Reingold. Undirected ST-Connectivity in Log-Space. In *Journal of the ACM*, Vol. 55 (4), pages17:1–17:24, 2008.

[10] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy Waves, the Zig-Zag Graph Product, and New Constant-Degree Expanders and Extractors. *Annals of Mathematics*, Vol. 155 (1), pages 157–187, 2001. Preliminary version in *41st FOCS*, pages 3–13, 2000.