TESTABILITY OF RELATIONS BETWEEN PERMUTATIONS

OREN BECKER, ALEXANDER LUBOTZKY, AND JONATHAN MOSHEIFF

Abstract. We initiate the study of property testing problems concerning relations between permutations. In such problems, the input is a tuple $(\sigma_1, \ldots, \sigma_d)$ of permutations on $\{1, \ldots, n\}$, and one wishes to determine whether this tuple satisfies a certain system of relations E, or is far from every tuple that satisfies E. If this computational problem can be solved by querying only a small number of entries of the given permutations, we say that E is *testable*. For example, when $d = 2$ and E consists of the single relation $\mathsf{XY} = \mathsf{YX}$, this corresponds to testing whether $\sigma_1\sigma_2 = \sigma_2\sigma_1$, where $\sigma_1\sigma_2$ and $\sigma_2\sigma_1$ denote composition of permutations.

We define a collection of graphs, naturally associated with the system E, that encodes all the information relevant to the testability of E. We then prove two theorems that provide criteria for testability and non-testability in terms of expansion properties of these graphs. By virtue of a deep connection with group theory, both theorems are applicable to wide classes of systems of relations.

In addition, we formulate the well-studied group-theoretic notion of stability in permutations as a special case of the testability notion above, interpret all previous works on stability as testability results, survey previous results on stability from a computational perspective, and describe many directions for future research on stability and testability.

1. Introduction

In this paper we study the testability of relations between permutations. We consider problems where several permutations are given in a black box form (e.g., as circuits or oracles), and one wishes to determine whether they satisfy a fixed system of relations E or are far from doing so. For example, suppose that E consists of the single formal relation $\mathsf{XY} = \mathsf{YX}$. The computational problem corresponding to E is to test whether two given permutations A and B, over the same finite set, commute. More precisely, the problem is to distinguish between the following two cases (i) $AB = BA$, and (ii) A and B are ε-far in the normalized Hamming metric (see (1.1) and Definition 1.5) from every pair of permutations A', B' that satisfies $A'B' = B'A'$.

Testing whether $AB = BA$ was shown in [10] to be achievable by an algorithm whose query and time complexities are both polynomial in $\frac{1}{\varepsilon}$. In particular, we say that the system $E = \{\mathsf{XY} = \mathsf{YX}\}$ is *testable* since it has a testing algorithm whose query complexity depends only on ε, and not[1] on the size n of the domain $\{1, \ldots, n\}$ of A and B. Here, *query complexity* counts queries of the form "what is $A(x)$" or "what is $B(x)$" for x in the domain of the permutations.

On the other hand, consider the system of relations $E = \{\mathsf{XZ} = \mathsf{ZX}, \mathsf{YZ} = \mathsf{ZY}\}$. Testing E amounts to testing, for three given permutations A, B and C, whether C commutes with both A and B. Theorem 3 in the present paper, together with [25], imply that this system is not testable (see Example 2.4). In other words, the query complexity of every testing algorithm for E must depend on n and not just on ε. Similarly, the system $E = \{\mathsf{XY}^2 = \mathsf{Y}^2\mathsf{X}\}$ is also not testable (this is also discussed in Example 2.4), even though it is superficially similar to the system $\{\mathsf{XY} = \mathsf{YX}\}$.

Oren Becker, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
Alexander Lubotzky, Einstein Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel
Jonathan Mosheiff, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA
E-mail addresses: oren.becker@gmail.com, alex.lubotzky@mail.huji.ac.il, jmosheif@cs.cmu.edu.

[1] More generally, testing algorithms with query complexity sublinear in n are a fascinating topic for future research, but the focus of the present paper is on testing with a constant number of queries. See also Section 2.2.3.

In this work we go far beyond the systems of relations in the examples above. We establish a framework and initiate a systematic study of the testability of systems of relations between permutations (also known as *equations in permutations*). Given such a system E, we naturally associate with it a certain infinite family of graphs GSol_E (see Section 1.2), which contains both finite and infinite graphs. Our main results give criteria for the testability and non-testability of E in terms of expansion properties of GSol_E.

Definition 1.1 (Measures of expansion)**.** Let G be a graph of bounded degree, with vertex set V. The *isoperimetric constant* of G is

$$\rho(G) = \inf\left\{ \frac{|E(X, V \setminus X)|}{|X|} \mid X \subseteq V \text{ is finite and nonempty} \right\}.$$

If G is finite, its *Cheeger constant* is

$$\alpha(G) = \inf\left\{ \frac{|E(X, V \setminus X)|}{|X|} \mid X \subseteq V \text{ and } 1 \leq |X| \leq \frac{|V|}{2} \right\}.$$

Here, $E(X, Y)$ denotes the set of edges between the sets X and Y.

Note that if G is finite then $\rho(G)$ is trivially 0 since the set $X = V$ has no expansion. However, some infinite graphs, such as the d-regular tree ($d \geq 3$), have a positive isoperimetric constant.

Our main positive theorem states that E is testable whenever GSol_E is nonexpanding in the isoperimetric sense.

Theorem 1 (Main positive theorem)**.** *If $\rho(G) = 0$ for every $G \in \mathrm{GSol}_E$ then E is testable.*

The notion of a *testable system of relations* is defined formally in Definition 1.6 in Section 1.1. The family of graphs GSol_E is defined in Section 1.2.

The aforementioned testability of $\{\mathsf{XY} = \mathsf{YX}\}$ is a narrow special case[2] of Theorem 1.

Our main negative theorem states that nonexpansion in the Cheeger sense is a necessary condition for testability.

Theorem 2 (Main negative theorem)**.** *Let FGSol_E denote the set of graphs in GSol_E that are finite and connected. If FGSol_E is infinite and $\inf\{\alpha(G) \mid G \in \mathrm{FGSol}_E\} > 0$ then E is non-testable.*

In Section 3 we will show how to associate a group[3] $\Gamma(E)$ with the system E. The expansion and isoperimetric constants appearing in Theorems 1 and 2 have been studied extensively in the framework of group theory, leading to numerous examples where the theorems are applicable, some of which are discussed in Section 3 and Appendix A.

We note that, while Theorems 1 and 2 apply to many systems of relations, they do not provide a complete classification. For example, the system consisting of the single relation $\mathsf{XY}^2 = \mathsf{Y}^3\mathsf{X}$ is known to satisfy neither the hypothesis of Theorem 1 nor that of Theorem 2, and the question of its testability remains open (see Problem 2.11). In the spirit of well-known classification theorems, such as those concerning constraint-satisfaction problems [13, 43] and efficient testability of the H-freeness property of a graph [3], a prominent objective of the present line of research is obtaining a complete characterization of testable systems of relations. We elaborate on this goal in Section 2.2.1. We now turn to providing the necessary framework for a precise statement of our results.

1.1. **A framework for systems of relations between permutations.** Fix a finite alphabet $S = \{s_1, \ldots, s_d\}$ throughout the introduction. Let $S^- = \left\{s_1^{-1}, \ldots, s_d^{-1}\right\}$ be the set of formal inverses of the letters in S, and write $S^{\pm} = S \cup S^-$.

[2] For $E = \{\mathsf{XY} = \mathsf{YX}\}$, it is not hard to verify directly that $\rho(G) = 0$ for all $G \in \mathrm{GSol}_E$, but in fact it suffices to verify that $\rho(C) = 0$, where C is the infinite grid in the plane. This suffices because C is the Cayley graph of the group $\mathbb{Z}^2$ (see Section 3).

[3] The family GSol_E is in fact the family of graphs whose connected components are Schreier graphs of the group $\Gamma(E)$.

Definition 1.2. A *relation* is a formal equation of the form $w_{i,1} = w_{i,2}$, where $w_{i,j}$ is a word over $S^{\pm}$. A *system of relations* is a finite set of relations.

Let $\mathrm{Sym}(n)$ denote the group of all permutations on $[n] := \{1 \dots, n\}$. Fix a system of relations $E = \{w_{i,1} = w_{i,2}\}_{i=1}^{r}$. We think of the letters in S as variables, and study the space of assignments $s_1 \leftarrow \sigma_1, \dots, s_d \leftarrow \sigma_d$ that satisfy all of the relations in E, where each σ_i is a permutation in $\mathrm{Sym}(n)$. In other words, we study the space of simultaneous solutions for E inside $(\mathrm{Sym}(n))^d$. More precisely, we are interested in testability problems related to this space of solutions. As an example, to fit the relation $\mathsf{XY} = \mathsf{YX}$ into this framework, we set $d = 2$, $S = \{s_1, s_2\}$, $r = 1$, $w_{1,1} = s_1 s_2$ and $w_{1,2} = s_2 s_1$. Thus $E = \{s_1 s_2 = s_2 s_1\}$ in this case. For notational convenience we sometimes denote $\mathsf{X} = s_1$, $\mathsf{Y} = s_2$ and $\mathsf{Z} = s_3$.

Definition 1.3. For a word w over $S^{\pm}$ and a tuple $\overline{\sigma} = (\sigma_1, \dots, \sigma_d) \in (\mathrm{Sym}(n))^d$, $n \in \mathbb{N}$, write $w(\overline{\sigma})$ for the permutation that results from applying the assignment $s_j \leftarrow \sigma_j$ to the word w.

Example 1.4. *If* $S = \{\mathsf{X}, \mathsf{Y}\}$, $w = \mathsf{XYX}^{-1}\mathsf{Y}^{-1}$ *and* $\overline{\sigma} = ((1\,2\,3), (1\,2)) \in (\mathrm{Sym}(3))^2$ *(where the permutations are given in cycle notation), then* $w(\overline{\sigma}) = (1\,2\,3)(1\,2)(1\,2\,3)^{-1}(1\,2)^{-1} = (1\,3\,2)$.

Let d_n^{H} denote[4] the normalized Hamming metric on $\mathrm{Sym}(n)$. That is,

$$d_n^{\mathrm{H}}(\sigma, \tau) = d^{\mathrm{H}}(\sigma, \tau) = \frac{1}{n}|\{x \in [n] \mid \sigma(x) \neq \tau(x)\}| \quad \forall \sigma, \tau \in \mathrm{Sym}(n) \ . \tag{1.1}$$

Definition 1.5. Let $n \in \mathbb{N}$. We say that $\overline{\sigma} \in (\mathrm{Sym}(n))^d$ is a *solution* for E in $(\mathrm{Sym}(n))^d$ (or that $\overline{\sigma}$ *satisfies* E) if $w_{i,1}(\overline{\sigma}) = w_{i,2}(\overline{\sigma})$ for each $1 \leq i \leq r$, and write $\mathrm{Sol}_E(n)$ for the set of solutions for E in $(\mathrm{Sym}(n))^d$. For $\varepsilon \geq 0$, let

$$\mathrm{Sol}_E^{\geq \varepsilon}(n) = \left\{ (\sigma_1, \dots, \sigma_d) \in (\mathrm{Sym}(n))^d \mid \sum_{j=1}^{d} d^{\mathrm{H}}(\sigma_j, \tau_j) \geq \varepsilon \quad \forall (\tau_1, \dots, \tau_d) \in \mathrm{Sol}_E(n) \right\} .$$

In the example where $S = \{\mathsf{X}, \mathsf{Y}\}$ and $E = \{\mathsf{XY} = \mathsf{YX}\}$, the space $\mathrm{Sol}_E(n)$ is the set of pairs $(A, B) \in (\mathrm{Sym}(n))^2$ such that $AB = BA$, and the space $\mathrm{Sol}_E^{\geq \varepsilon}(n)$ is the set of all pairs $(A, B) \in (\mathrm{Sym}(n))^2$ that satisfy $d^H(A, A') + d^H(B, B') \geq \varepsilon$ whenever $(A', B') \in (\mathrm{Sym}(n))^2$ is a commuting pair.

The following is the main novel definition of this paper:

Definition 1.6 (Testable system of relations)**.** An algorithm $\mathcal{M}$ that takes $n \in \mathbb{N}$ and a tuple $\overline{\sigma} = (\sigma_1, \dots, \sigma_d) \in (\mathrm{Sym}(n))^d$ as input is an (ε, q)*-tester* for E if it satisfies the following conditions:

- **Completeness:** if $\overline{\sigma} \in \mathrm{Sol}_E(n)$, the algorithm accepts with probability at least 0.99.
- ε**-soundness:** if $\overline{\sigma} \in \mathrm{Sol}_E^{\geq \varepsilon}(n)$, the algorithm rejects with probability at least 0.99.
- **Query efficiency:** the algorithm is only allowed to query q entries of $\sigma_1, \dots, \sigma_d$ and their inverses.

If for every $\varepsilon > 0$ there are $q = q(\varepsilon) \in \mathbb{N}$ and an (ε, q)-tester $\mathcal{M}_\varepsilon$ for $E = \{w_{i,1} = w_{i,2}\}_{i=1}^{r}$ then we say that E is $q(\varepsilon)$*-testable* (or just *testable*) and that $\varepsilon \mapsto \mathcal{M}_\varepsilon$ is a *family of testers* for E.

Note that in Definition 1.6 we allow the algorithm to have oracle access both to the entries of the given permutations and to the entries of their inverses. Oracle access to inverses of permutations is in fact not always necessary. Appendix B explains what can be done to circumvent the need to sample inverses.

In the sequel, it will be convenient to work with sets of *relators* rather than systems of relations, as explained below. A word w over the alphabet $S^{\pm}$ is *reduced* if it does not contain any subword of the form $s_i s_i^{-1}$ or $s_i^{-1} s_i$, $1 \leq i \leq d$. Write F_S for the set of reduced words over $S^{\pm}$ (in Appendix C we recall that F_S has a natural group structure, making it the *free group* on S). Every word w over $S^{\pm}$ is equivalent to a unique reduced word, obtained from w by repeatedly removing subwords of the form $s_i s_i^{-1}$ or $s_i^{-1} s_i$, $1 \leq i \leq d$.

[4] We shall omit the subscript n when it is clear from the context.

Let w be a word over $S^{\pm}$. Write $w = s_{i_1}^{e_1} \cdots s_{i_\ell}^{e_\ell}$, where $\ell \geq 0$, $e_j \in \{+1, -1\}$ and $1 \leq i_j \leq d$ for all $1 \leq j \leq \ell$. Then w has a formal inverse $w^{-1} := s_{i_\ell}^{-e_\ell} \cdots s_{i_1}^{-e_1}$.

A system of relations $E = \{w_{i,1} = w_{i,2}\}_{i=1}^{r}$ gives rise to a subset R_E of F_S, defined to be the set of reduced words equivalent to $w_{1,2}^{-1}w_{1,1}, \dots, w_{r,2}^{-1}w_{r,1}$. We say that R_E is the *set of relators* corresponding to E (in general, a set of relators is just a subset of F_S). For example, if $E = \{\mathsf{XZ} = \mathsf{ZX}, \mathsf{YZ} = \mathsf{ZY}\}$ then $R_E = \{\mathsf{X}^{-1}\mathsf{Z}^{-1}\mathsf{XZ}, \mathsf{Y}^{-1}\mathsf{Z}^{-1}\mathsf{YZ}\}$. Clearly, the system of relations E is equivalent to the system of relations $E' = \{w = 1 \mid w \in R_E\}$. Indeed, $\mathrm{Sol}_E(n) = \mathrm{Sol}_{E'}(n)$ and $\mathrm{Sol}_E^{\geq \varepsilon}(n) = \mathrm{Sol}_{E'}^{\geq \varepsilon}(n)$ for all $n \geq 1$ and $\varepsilon > 0$. It is generally more convenient to work with R_E rather than directly with E.

1.2. **A graph-theoretic view.** It will be beneficial to encode a tuple of permutations as an S-graph, defined below.

Definition 1.7. An S-graph is an edge-labelled directed (not necessarily finite) graph[5], where each directed edge is labelled by an element of S, and each vertex has exactly one outgoing and one incoming edge labelled s for each $s \in S$.

Given $n \in \mathbb{N}$, write $\mathcal{G}_S(n)$ for the set of S-graphs on the vertex set $[n]$. In particular, $\mathcal{G}_S(n)$ consists solely of finite S-graphs.

When referring to the connected components of an S-graph G, or to whether or not G is connected, we disregard edge orientation and labels, and treat G as an undirected graph.

Let G be an S-graph. For $s \in S$ and a vertex x, we write $s_G x = y$ for the unique vertex y such that $x \xrightarrow{s} y$ is an edge in G. We also define $s_G^{-1} y = x$. When there is no ambiguity about the graph in context, we write sx for $s_G x$ (for $s \in S^{\pm}$). For a word $w = w_t \cdots w_1$ ($w_i \in S^{\pm}$) we recursively define $w_G x = w_t(w_{t-1} \dots w_1 x)$ whenever $t > 1$. That is, $w_G x$ (or just wx) is the final vertex in the path that starts at x and follows t edges labelled according to w.

We next show how to encode a tuple $\overline{\sigma} = (\sigma_1, \dots, \sigma_d) \in (\mathrm{Sym}(n))^d$ as an S-graph $G_{\overline{\sigma}}$. Let $G_{\overline{\sigma}} \in \mathcal{G}_S(n)$ denote the S-graph with vertex set $[n]$ and edge set $\left\{x \xrightarrow{s_i} \sigma_i x \mid x \in [n], 1 \leq i \leq d\right\}$. Clearly, the map $\overline{\sigma} \mapsto G_{\overline{\sigma}} \colon (\mathrm{Sym}(n))^d \to \mathcal{G}_S(n)$ is a bijection.

Let $E = \{w_{i,1} = w_{i,2}\}_{i=1}^{r}$ be a system of relations over $S^{\pm}$. An S-graph G is said to belong to the class GSol_E if $(w_{i,1})_G x = (w_{i,2})_G x$ for every $1 \leq i \leq r$ and vertex x. Equivalently, $G \in \mathrm{GSol}_E$ if $w_G x = x$ holds for all $w \in R_E$ and every vertex x. In this case, we say that G *satisfies* E.

Example 1.8. *Let $S = \{\mathsf{X}, \mathsf{Y}\}$ and $E = \{\mathsf{XY} = \mathsf{YX}\}$. Let $m, n \geq 1$, and let G be the graph on the vertex set $V = \{0, \dots, m-1\} \times \{0, \dots, n-1\}$, with the following edges: for each $v = (a, b) \in V$, the X-labelled edge originating from v terminates at $(a+1, b)$, and the Y-labelled edge originating from v terminates at $(a, b+1)$ (where the addition is taken modulo m and n, respectively). Then G belongs to GSol_E (note that G can be embedded on a torus).*

Denote $\mathrm{GSol}_E(n) = \mathrm{GSol}_E \cap \mathcal{G}_S(n)$. Given $\overline{\sigma} \in (\mathrm{Sym}(n))^d$, note that $\sigma \in \mathrm{Sol}_E(n)$ if and only $G_{\overline{\sigma}} \in \mathrm{GSol}_E(n)$. In other words, $G_{\overline{\sigma}}$ satisfies E if and only if $\overline{\sigma}$ does the same. Importantly, the inclusion $\bigcup_{n \in \mathbb{N}} \mathrm{GSol}_E(n) \subset \mathrm{GSol}_E$ is strict since the latter set contains infinite S-graphs.

At this point, all notions in the statements of Theorems 1 and 2 have been fully defined. As these theorems demonstrate, the power of the correspondence $\mathrm{Sol}_E(n) \to \mathrm{GSol}_E(n)$, $n \in \mathbb{N}$, manifests in a connection between the testability of E and the geometry of the (finite and infinite) graphs in GSol_E.

[5] S-graphs are allowed to have self-loops and multiple edges, but as follows from the definition, two different edges directed from vertex x to vertex y must have distinct labels.

As we will be working extensively with the graph encoding of a tuple of permutations, it will be convenient to think of a tester as an algorithm whose input lies in $\mathcal{G}_S(n)$, rather than $(\mathrm{Sym}(n))^d$. We naturally define[6]

$$d^{\mathrm{H}}(G, G') = \sum_{s \in S} \frac{1}{n} |\{x \in [n] \mid s_G x \neq s_{G'} x\}| \quad \forall G, G' \in \mathcal{G}_S(n) \tag{1.2}$$

and

$$\mathrm{GSol}_E^{\geq \varepsilon}(n) = \left\{ G \in \mathcal{G}_S(n) \mid d^{\mathrm{H}}(G, G') \geq \varepsilon \ \forall G' \in \mathrm{GSol}_E(n) \right\} .$$

Note that

$$d^{\mathrm{H}}(G_{\overline{\sigma}}, G_{\overline{\tau}}) = \sum_{i=1}^{d} d^{\mathrm{H}}(\sigma_i, \tau_i)$$

for $\overline{\sigma} = (\sigma_1, \ldots, \sigma_d)$ and $\overline{\tau} = (\tau_1, \ldots, \tau_d)$ in $(\mathrm{Sym}(n))^d$, and that

$$\mathrm{GSol}_E^{\geq \varepsilon}(n) = \left\{ G_{\overline{\sigma}} \mid \overline{\sigma} \in \mathrm{Sol}_E^{\geq \varepsilon}(n) \right\} .$$

We can now restate Definition 1.6 in terms of S-graphs.

Definition 1.9 (Testable system of relations in terms of S-graphs)**.** An algorithm $\mathcal{M}$ that takes $n \in \mathbb{N}$ and an S-graph $G \in \mathcal{G}_S(n)$ as input is an (ε, q)-*tester* for $E = \{w_{i,1} = w_{i,2}\}_{i=1}^{r}$ if it satisfies the following conditions:

- **Completeness:** if $G \in \mathrm{GSol}_E(n)$, the algorithm accepts with probability at least 0.99.
- **ε-soundness:** if $G \in \mathrm{GSol}_E^{\geq \varepsilon}(n)$, the algorithm rejects with probability at least 0.99.
- **Query efficiency:** the algorithm is only allowed to make q queries, where each query is of the form "what is $s_G x$?", $x \in [n]$, $s \in S^{\pm}$.

If for every $\varepsilon > 0$ there are $q = q(\varepsilon) \in \mathbb{N}$ and an (ε, q)-tester $\mathcal{M}_\varepsilon$ for E then we say that E is $q(\varepsilon)$-*testable* (or just *testable*) and that $\varepsilon \mapsto \mathcal{M}_\varepsilon$ is a *family of testers* for E.

1.3. **The `Sample and Substitute` algorithm, a first attempt at defining a tester.** We turn to discussing algorithms for testing a given system of relations. Fix a system of relations E. The `Sample and Substitute` algorithm with repetition factor k (denoted SAS_k^E), defined below, is arguably the most natural candidate for an algorithm that tests whether an S-graph G lies in GSol_E.

Algorithm 1 Sample and Substitute for E with repetition factor k
Input: $n \in \mathbb{N}$ and $G \in \mathcal{G}_S(n)$

Sample $(w_1, x_1), \ldots, (w_k, x_k)$ uniformly and independently from $R_E \times [n]$.
if $w_j x_j = x_j$ in the graph G for all $1 \leq j \leq k$ **then**
    Accept.
else
    Reject.

For example, when $S = \{\mathsf{X}, \mathsf{Y}\}$ and $E = \{\mathsf{XY} = \mathsf{YX}\}$ (and so $R_E = \left\{\mathsf{X}^{-1}\mathsf{Y}^{-1}\mathsf{XY}\right\}$), the algorithm SAS_k^E randomly samples k vertices $x_1, \ldots, x_k$ from $[n]$. It then substitutes each x_j into the permutation of $[n]$ defined by $\mathsf{X}^{-1}\mathsf{Y}^{-1}\mathsf{XY}$ and checks whether the result is x_j. In other words, for each $1 \leq j \leq k$, the algorithm checks whether walking from x_j along the edge labelled Y and then the edge labelled X, and then in the reverse direction along the edge labelled Y and then the edge labelled X, leads back to x_j. The algorithm accepts if and only if this check succeeds for all $1 \leq j \leq k$. For this specific system of relations, the query complexity is $4k$, since, to compute $\mathsf{X}^{-1}\mathsf{Y}^{-1}\mathsf{XY}x_j$, the algorithm first needs to query the vertex $\mathsf{Y}x_i$, then the vertex $\mathsf{X}(\mathsf{Y}x_j)$, and so on, for a total of 4 queries. In terms of tuples of permutations (rather than graphs),

[6]In words, $d^{\mathrm{H}}(G, G')$ counts (with a $\frac{1}{n}$ normalization factor) pairs consisting of a vertex x and letter s, such that the respective s-labelled edges originating from x, in the graphs G and G', disagree on their target vertex.

the input is $(\sigma_1, \sigma_2) \in (\mathrm{Sym}(n))^2$, and the run of the algorithm is equivalent to sampling $x_1, \ldots, x_k$ uniformly and independently from $[n]$ and accepting if and only if $\sigma_1^{-1}\sigma_2^{-1}\sigma_1\sigma_2 x_j = x_j$ (equivalently, $\sigma_1\sigma_2 x_j = \sigma_2\sigma_1 x_j$) for all $1 \le j \le k$.

For a general system of relations E, the query complexity of SAS_k^E is $O_E(k)$ since, generalizing from the case $E = \{\mathsf{XY} = \mathsf{YX}\}$, the algorithm requires $|w_j|$ queries in order to compute $(w_j)_G x_j$, where we write $|w|$ for the length of a reduced word w. In particular, if k is independent of n then so is the query complexity. If E admits a family of testers consisting of SAS_k^E algorithms, we say that E is *stable*. This is formalized in the following definition, which is stated in combinatorial terms in [22, Definition 1] and in group-theoretic terms in [6, Theorem 4.2] and [25, Definition 1.1]. Here the definition is given in computational terms:

Definition 1.10. The system of relations E is *stable* if there is a function $k\colon \mathbb{R}_{>0} \to \mathbb{N}$ such that $\varepsilon \mapsto \mathrm{SAS}_{k(\varepsilon)}^E$ is a family of testers for E. In this case we also say that E is $k(\varepsilon)$-*stable*.

In particular, every stable system of relations is testable.

Note that for every $k \in \mathbb{N}$, the algorithm SAS_k^E has perfect completeness, namely, when $G \in \mathrm{GSol}_E(n)$, the algorithm always accepts. Thus, the question of whether a system E is stable is equivalent to the question of whether there is a function $\epsilon \mapsto k(\varepsilon)$ such that $\mathrm{SAS}_{k(\varepsilon)}^E$ is ε-sound for all $\varepsilon > 0$.

The question of which systems of relations are stable has recently received a lot of attention (see Section 2.1), yielding both positive and negative results. For example, the aforementioned testability of $E = \{\mathsf{XY} = \mathsf{YX}\}$ follows from the following stability result:

Example 1.11 ($\{\mathsf{XY} = \mathsf{YX}\}$ is stable)**.** *Consider the system $E = \{\mathsf{XY} = \mathsf{YX}\}$. The main theorem of [6] implies that there is a function $k\colon \mathbb{R}_{>0} \to \mathbb{N}$ such that $\varepsilon \mapsto \mathrm{SAS}_{k(\varepsilon)}^{\mathsf{XY}=\mathsf{YX}}$ is a family of testers for $\mathsf{XY} = \mathsf{YX}$. A later work [10] improved upon the result of [6] by showing that we may take $k(\varepsilon) = \left(\frac{1}{\varepsilon}\right)^{O(1)}$, where the implied constant is absolute, and by providing a constructive[7] proof. Thus, not only is E testable by* `Sample and Substitute`*, but moreover, this testing can be done efficiently.*

One natural question from the testing perspective is whether `Sample and Substitute` is *universal*, namely, does every testable system of relations have a family of `Sample and Substitute` testers? In other words, are testability and stability equivalent? As we shall see, Theorem 1 can be used to provide a negative answer to this question. In Section 3.1 we give a concrete example of a system of relations which is testable but not stable (see also Appendix A.1).

Fortunately, however, it turns out that a universal tester does exist. We turn now to describing this tester, which we name `Local Statistics Matcher`[8].

1.4. **Local Statistics Matcher, a universal tester for relations between permutations.** The idea behind Local Statistics Matcher is as follows: Let G be a finite S-graph. For a vertex x of G and $r \ge 1$, we consider the isomorphism class of the closed ball[9] $B_G(x, r)$ of radius r centered at x as a rooted directed edge-labelled graph (the root is x and the orientations and edge labels are inherited from G, but the relevant notion of a graph isomorphism ignores vertex labels). The graph G gives rise to a probability distribution on the set of such isomorphism classes: Write $N_{G,r}$ for the distribution of the isomorphism class of $B_G(x, r)$, where x is sampled uniformly from the set of vertices of G.

Given a graph $G \in \mathcal{G}_S(n)$, Local Statistics Matcher computes an approximation of the distribution $N_{G,r}$ (for a large enough r, independent of n), which we denote by $N_{G,r}^{\mathrm{Empirical}}$. The algorithm accepts if and only if there exists $G' \in \mathrm{GSol}_E(n)$ such that the distributions $N_{G,r}^{\mathrm{Empirical}}$ and $N_{G',r}$ are at most δ-far from each

[7] That is, the proof in [10] provides an explicit algorithm that, given a graph $G \in \mathcal{G}_S(n)$ that is accepted by $\mathrm{SAS}_{k(\varepsilon)}^{\mathsf{XY}=\mathsf{YX}}$ with high probability, produces a graph $G' \in \mathrm{GSol}_E(n)$ close to G.

[8] Readers familiar with Benjamini–Schramm convergence will recognize a close connection between `Local Statistics Matcher` and the Benjamini–Schramm metric.

[9] This ball is the graph whose vertex set V consists of all vertices wx of G, where w is a word of length at most r over $S^{\pm}$, and whose edge set consists of all the edges of G that start and end in V.

other (for a small enough $\delta > 0$, independent of n). In other words, Local Statistics Matcher accepts with high probability if and only if G and G' are close together under the Benjamini–Schramm metric [2, 12].

More precisely, Local Statistics Matcher is a parameterized family of algorithms, i.e., for k, P and δ (see below) we have a Local Statistics Matcher algorithm $\mathrm{LSM}^E_{k,P,\delta}$. As we shall see in Theorem 3, Local Statistics Matcher is universal in the following sense: the system E is testable if and only if for every $\varepsilon > 0$ there are k, P and δ such that $\mathrm{LSM}^E_{k,P,\delta}$ is an (ε, q)-tester for E, for some q which depends only on ε.

Rather than working with isomorphism classes of balls, it is more convenient (and essentially equivalent) to consider *stabilizers*:

Definition 1.12. For a vertex x of an S-graph G, let

$$\mathrm{Stab}_G(x) = \{w \in F_S \mid w_G x = x\} \ .$$

For $P \subset F_S$, write $N_{G,P}$ for the distribution (over the set of subsets of P) of $\mathrm{Stab}_G(x) \cap P$ when x is sampled uniformly from the set of vertices of G.

If $P = \{w \in F_S \mid |w| \le r\}$, $r \ge 1$, then $\mathrm{Stab}_G(x) \cap P$ determines the isomorphism class, as a rooted directed edge-labelled graph, of the closed ball $B_G(x, r/2)$ of radius $r/2$ centered at x in the graph G. Indeed, $\mathrm{Stab}_G(x) \cap P$ determines the set

$$\{(w_1, w_2) \in F_S \times F_S \mid |w_1|, |w_2| \le r/2, w_1 x = w_2 x\} \ . \tag{1.3}$$

because $w_1 x = w_2 x$ if and only if the reduced word equivalent to $w_2^{-1} w_1$ is in $\mathrm{Stab}_G(x) \cap P$ (when $|w_1|, |w_2| \le r/2$). The set (1.3) encodes the isomorphism class of $B_G(x, r/2)$.

Before formulating the algorithm we need some notation. Given a set A, we denote its power set by $\mathrm{Subsets}(A)$, and write $\mathrm{FinSubsets}(A)$ for the set of finite subsets of A. When A is finite, we write $\mathrm{U}(A)$ for the uniform distribution over A.

Denote the total-variation distance between two distributions θ_1, θ_2 over a finite set Ω by

$$d_{\mathrm{TV}}(\theta_1, \theta_2) \coloneqq \frac{1}{2} \sum_{x \in \Omega} |\theta_1(x) - \theta_2(x)| \ .$$

The `Local Statistics Matcher` algorithm for a system of relations E takes three parameters in addition to E: a *repetition factor* $k \in \mathbb{N}$; a finite *word set* $P \in \mathrm{FinSubsets}(F_S)$; and a *proximity parameter* $\delta > 0$. The algorithm, denoted $\mathrm{LSM}_{k,P,\delta}$, is defined as follows.

Algorithm 2 Local Statistics Matcher for E with with repetition factor k, word set $P \subset F_S$ and proximity parameter δ

Input: $n \in \mathbb{N}$ and $G \in \mathcal{G}_S(n)$

1: Sample $x_1, \dots, x_k$ uniformly and independently from $[n]$.
2: For each $1 \le j \le k$, compute the set $\mathrm{Stab}_G(x_j) \cap P$ by querying G.
3: Let $N^{\mathrm{Empirical}}_{G,P}$ be the distribution of $\mathrm{Stab}_G(x_j) \cap P$ where j is sampled uniformly from $[k]$.
4: **if**

$$\min\left\{d_{\mathrm{TV}}\left(N^{\mathrm{Empirical}}_{G,P}, N_{H,P}\right) \mid H \in \mathrm{GSol}_E(n)\right\} \le \delta \tag{1.4}$$

then
5: Accept.
6: **else**
7: Reject.

In our proof of the universaility of `Local Statistics Matcher`, we only make use of sets P of the form $P = \{w \in F_S \mid |w| \le r\}$, $r \ge 1$. This restriction does not hurt the universality of the algorithm. For sets P of this form, the algorithm can essentially be seen as sampling balls of radius $\frac{r}{2}$.

The query complexity of $\mathrm{LSM}^E_{k,P,\delta}$ is $O\big(k\sum_{w\in P}|w|\big)$, and in particular it is independent of n whenever the same is true for k, P and δ. Indeed, to determine $\mathrm{Stab}_G(x_j)\cap P$, $1\le j\le k$, it suffices to compute $w_G x_j$ for each $w\in P$, and this can be done in $|w|$ queries for any given w.

In contrast to `Sample and Substitute`, the `Local Statistics Matcher` algorithm is not necessarily perfectly complete. Namely, for some systems of relations E, the algorithm may reject (with some small probability) even if $G\in \mathrm{GSol}_E(n)$. This can happen in a run of `Local Statistics Matcher` only if $N^{\mathrm{Empirical}}_{G,P}$ is not a good approximation of $N_{G,P}$.

Remark 1.13. In a naive implementation of $\mathrm{LSM}^E_{k,P,\delta}$, the algorithm explicitly enumerates the elements of $\mathrm{GSol}_E(n)$ in order to compute the set of distributions $\{N_{H,P}\mid H\in \mathrm{GSol}_E(n)\}$. This results in running time exponential in n. While the present work focuses on query complexity, we discuss possible significant time complexity improvements in Section 2.2.4.

1.4.1. *Statistical distinguishability and the universality of* `Local Statistics Matcher`. To prove the universality of `Local Statistics Matcher`, we introduce the notion of a *statistically distinguishable* system of relations (Definition 1.14). We then show (Theorem 3) that for a system of relations E, the following implications hold:

E is testable $\implies$ E is statistically distinguishable $\implies$ E is testable by `Local Statistics Matcher`,

and thus the three statements are equivalent.

Definition 1.14. The system of relations E is *statistically distinguishable* if for every $\varepsilon>0$ there exist $P(\varepsilon)\in \mathrm{FinSubsets}(F_S)$ and $\delta(\varepsilon)>0$ such that $d_{\mathrm{TV}}\big(N_{G,P(\varepsilon)},N_{H,P(\varepsilon)}\big)\ge\delta(\varepsilon)$ for every $n\in\mathbb{N}$, $G\in\mathrm{GSol}_E(n)$ and $H\in\mathrm{GSol}^{\ge\varepsilon}_E(n)$. In this case we also say that E is $(P(\varepsilon),\delta(\varepsilon))$*-statistically-distinguishable*.

Remark 1.15. Given graphs $G,H\in\mathcal{G}_S(n)$, we sometimes refer to $d^{\mathrm{H}}(G,H)$ as their *global distance*, and to $d_{\mathrm{TV}}(N_{G,P},N_{H,P})$ (where $P\in\mathrm{FinSubsets}(F_S)$) as their *$P$-local distance*. Thus, E is statistically distinguishable if whenever H is ε-far from $\mathrm{GSol}_E(n)$ in the global metric, it is also δ-far from $\mathrm{GSol}_E(n)$ in the P-local metric, where δ and P may depend on ε.

Theorem 3. *Fix a system of relations E over an alphabet $S^{\pm}$. The following conditions are equivalent:*

(1) E is testable.
(2) There exist functions $k\colon\mathbb{R}_{>0}\to\mathbb{N}$, $P\colon\mathbb{R}_{>0}\to\mathrm{FinSubsets}(F_S)$ and $\delta\colon\mathbb{R}_{>}\to\mathbb{R}_{>0}$ such that $\varepsilon\mapsto\mathrm{LSM}^E_{k(\varepsilon),P(\varepsilon),\delta(\varepsilon)}$ is a family of testers for E.
(3) E is statistically distinguishable.

We prove Theorem 3 in Section 7. Note that, in addition to implying the universality of `Local Statistics Matcher`, Theorem 3 reduces the question of testability to the notion of statistical distinguishability. The latter is graph theoretic and geometric, rather than algorithmic. Our proofs of Theorems 1 and 2 rely on this reduction.

1.5. **Overview.** Section 2 contains a survey of previous work, and discusses directions for future research. Section 3 explains the powerful relationship between testability and group theory: each system of relations E gives rise to a group $\Gamma(E)$, and the foundational Proposition 3.2 says that the testability of E depends only on $\Gamma(E)$. Section 3 also reviews the relevant group-theoretic notions in a combinatorial language.

In Section 4 we prove Theorem 2 by a direct use of expansion in graphs. In Section 5 we prove Theorem 1 using deep results such as the Newman–Sohler Theorem [36] (see also [20]) and the Ornstein–Weiss Theorem [14, 38]. In Section 6 we prove Proposition 3.2.

The proofs of Theorems 1 and 2 and Proposition 3.2 use the reduction from testability to the geometry of graphs, as afforded by Theorem 3, which is proved in Section 7.

Appendix A.1 uses Theorem 1 to provide an example of a testable instable system of relations. Appendix A.2 gives an example of a system satisfying the hypothesis of Theorem 2.

2. A survey of previous work and directions for further research

2.1. **A survey of previous work on stability.** The notion of testability of relations and the computational perspective on stability are new concepts, introduced in this paper. Nevertheless, there is a lot of recent research of stability in permutations from group-theoretic and combinatorial points of view. Most of the existing literature on stability relies on the relation to group theory as explained in Section 3. We formulate these results in the graph-theoretic language developed in Section 1.2.

2.1.1. *The combinatorial definition of stability.* Stability, as in Definition 1.10, has several equivalent definitions (see [22, Definition 1], [6, Theorem 4.2], [25, Definition 1.1]). Below we recall one of them and prove the equivalence.

Definition 2.1 (Combinatorial definition of stability)**.** [22, Definition 1] Let E be a system of relations over $S^{\pm}$, and let $G \in \mathcal{G}_S(n)$. The *local defect* of G with respect to E is

$$L_E(G) \coloneqq \sum_{w \in R_E} \Pr_{x \sim \mathrm{U}([n])}[w_G x \neq x].$$

We say that E is *stable* if

$$\inf\left\{ L_E(G) \mid G \in \mathrm{GSol}_E^{\geq \varepsilon}(n), n \in \mathbb{N} \right\} > 0 \quad \forall \varepsilon > 0 \; .$$

Remark 2.2. We use the notion of local defect in the proof of Proposition 3.2 (see Section 6).

For $G \in \mathcal{G}_S(n)$, we have $G \in \mathrm{GSol}_E(n)$ if and only if $L_E(G) = 0$. Another way to state Definition 2.1 is: The system E is stable if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that $G \notin \mathrm{GSol}_E^{\geq \varepsilon}(n)$ whenever $n \in \mathbb{N}$, $G \in \mathcal{G}_S(n)$ and $L_E(G) < \delta$. That is (informally), E is stable if every graph which approximately satisfies E is close to a graph which fully satisfies E.

For $G \in \mathcal{G}_S(n)$, it is easy to see that the probability that SAS_1^E rejects G is $\frac{1}{|E|} L_E(G)$, and that SAS_k^E can be implemented by running SAS_1^E for k independent iterations and accepting if all iterations accept. These observations are used in the proof of the following claim.

Claim 2.3. A system of relations E is stable in the sense of Definition 2.1 if and only if it is stable in the sense of Definition 1.10.

Proof. Suppose that E is stable in the sense of Definition 2.1. For $\varepsilon > 0$, write

$$\delta(\varepsilon) = \frac{1}{|E|} \inf\left\{ L_E(G) \mid G \in \mathrm{GSol}_E^{\geq \varepsilon}(n), n \in \mathbb{N} \right\} > 0$$

and

$$k(\varepsilon) = \left\lceil \log_{1-\delta(\varepsilon)} 0.01 \right\rceil \; .$$

We claim that $\varepsilon \mapsto \mathrm{SAS}_{k(\varepsilon)}^E$ is a family of testers for E. Let $n \in \mathbb{N}$ and $G \in \mathrm{GSol}_E^{\geq \varepsilon}(n)$. Then SAS_1^E rejects G with probability $\frac{1}{|E|} L_E(G) \geq \delta(\varepsilon)$. Thus, the probability that SAS_k^E accepts G is at most

$$(1 - \delta(\varepsilon))^k \leq (1 - \delta(\varepsilon))^{\log_{1-\delta(\varepsilon)} 0.01} = 0.01 \; .$$

Conversely, suppose that E is stable in the sense of Definition 1.10, and let $\varepsilon \mapsto \mathrm{SAS}_{k(\varepsilon)}^E$ be a family of testers for E, $k \colon \mathbb{R}_{>0} \to \mathbb{N}$. Take $G \in \mathrm{GSol}_E^{\geq \varepsilon}(n)$. Write p_1 (resp. $p_{k(\varepsilon)}$) for the probability that SAS_1^E (resp. $\mathrm{SAS}_{k(\varepsilon)}^E$) rejects G. On one hand, $p_1 = \frac{1}{|E|} L_E(G)$ and $p_{k(\varepsilon)} \geq 0.99$. On the other hand, $p_{k(\varepsilon)} \leq k(\varepsilon) \cdot p_1$ by the union bound, and thus $L_E(G) \geq \frac{0.99|E|}{k(\varepsilon)}$. Hence

$$\inf\left\{ L_E(G) \mid G \in \mathrm{GSol}_E^{\geq \varepsilon}(n), n \in \mathbb{N} \right\} \geq \frac{0.99|E|}{k(\varepsilon)} > 0 \; .$$

$\square$

2.1.2. *Stability under the hypothesis of Theorem 1.* Let E be a system of relations satisfying the hypothesis of Theorem 1. That is, $\rho(G) = 0$ for every $G \in \mathrm{GSol}_E$. Theorem 1 says that E is testable. However, E is not necessarily stable. In fact, there is a group-theoretic criterion that determines whether a system E, satisfying the hypothesis of Theorem 1, is stable (see Theorem [9, Theorem 1.3(ii)][10]). See Section 3.1 and Appendix A.1 for an explicit example of a stable non-testable system.

Example 2.4 (Baumslag–Solitar relations)**.** *Fix $m, n \in \mathbb{Z}$, and consider the system $E_{m,n} = \{\mathsf{XY^m} = \mathsf{Y^nX}\}$, consisting of a single relation. Then $E_{m,n}$ is stable if and only if $|m| \le 1$ or $|n| \le 1$. Indeed:*

- *The case $m = n = 0$ is clear.*
- *If $m = 0$ and $n \neq 0$ then $E_{m,n}$ is stable by* [22, Theorem 2].
- *If $|m| = 1$ or $|n| = 1$ then $E_{m,n}$ is stable by* [9, Theorem 1.2(ii)].
- *If $|m|, |n| \ge 2$ and $|m| \neq |n|$ then $E_{m,n}$ is not stable by* [6, Example 7.3] *(see also* [9, Theorem 1.3(i)]*).*
- *If $|m|, |n| \ge 2$ and $|m| = |n|$ then $E_{m,n}$ is not stable by* [25, Corollary B(3)].

As for the testability of $E_{m,n}$, the stable cases are clearly testable. Testability in the case $|m|, |n| \ge 2$ and $|m| \neq |n|$ is an interesting open question (see Section 2.2.1).

In the remaining case, $|m|, |n| \ge 2$ and $|m| = |n|$, it turns out that $E_{m,n}$ is not testable. Indeed, by [25, Theorem D] *there are $\varepsilon_0 > 0$ and finite S-graphs in $\mathrm{GSol}^{\ge\varepsilon_0}_{E_{m,n}}$ with local statistics approximating a certain*[11] *infinite vertex-transitive S-graph H arbitrarily well. On the other hand, it is well-known that there are also finite S-graphs in $\mathrm{GSol}_{E_{m,n}}$ whose local statistics approximate the same infinite graph H arbitrarily well*[12]*. Thus $E_{m,n}$ is not statistically distinguishable*[13]*, and hence it is not testable by Theorem 3. In fact, $E_{m,n}$ is not even flexibly testable (see Section 2.1.3) because the argument above remains true even with a flexible definition of $\mathrm{GSol}^{\ge\varepsilon_0}_{E_{m,n}}$ by* [25, Theorem D].

By the above, the system of relations $E_{2,2} = \{\mathsf{XY^2} = \mathsf{Y^2X}\}$, mentioned in the beginning of Section 1, is not flexibly testable. By a similar argument, also based on [25, Theorem D] *and Theorem 3, the system $\{\mathsf{XY} = \mathsf{ZX}, \mathsf{YZ} = \mathsf{ZY}\}$ is also not flexibly testable.*

2.1.3. *Flexible stability (and testability).* A slightly weaker form of stability, called *flexible stability* [8, Section 4.4], has led to fascinating research. Here we define this notion in the language of the present paper. We also introduce a new, more general, notion of *flexible testability*.

For $G \in \mathcal{G}_S(n)$ and $G' \in \mathcal{G}_S(N)$, $N \ge n$, let

$$d^{\mathrm{H}}(G, G') = d^{\mathrm{H}}(G', G) = \sum_{s \in S} \frac{|\{x \in [n] \mid s_G x \neq s_{G'} x\}| + (N - n)}{N} .$$

Then d^{H} is a metric on the disjoint union $\coprod_{n \in \mathbb{N}} \mathcal{G}_S(n)$ [7, Lemma A.1], extending (1.2). Note that $d^{\mathrm{H}}(G, G') \in [0, |S|]$, and that if $d^{\mathrm{H}}(G, G')$ is close to 0 then $\frac{n}{N}$ is close to 1 (indeed, $1 - \frac{d^{\mathrm{H}}(G,G')}{|S|} \le \frac{n}{N} \le 1$). For a system of relations E over $S^{\pm}$, let

$$\mathrm{GSol}^{\ge\varepsilon,\mathrm{flex}}_E(n) = \left\{ G \in \mathcal{G}_S(n) \mid d^{\mathrm{H}}(G, G') \ge \varepsilon \quad \forall G' \in \coprod_{m \in \mathbb{N}} \mathrm{GSol}_E(m) \right\} .$$

Definition 2.5 (Flexibly-testable system of relations)**.** An algorithm $\mathcal{M}$ that takes $n \in \mathbb{N}$ and an S-graph $G \in \mathcal{G}_S(n)$ as input is an (ε, q)-*flexible-tester* for E if it satisfies the following conditions:

- **Completeness:** if $G \in \mathrm{GSol}_E(n)$, the algorithm accepts with probability at least 0.99.
- **ε-soundness:** if $G \in \mathrm{GSol}^{\ge\varepsilon,\mathrm{flex}}_E(n)$, the algorithm rejects with probability at least 0.99.

[10] Theorem 1.3(ii) of [9] characterizes stability among systems E for which the group $\Gamma(E)$ is amenable (see Section 3)

[11] Namely, H is the Cayley graph of the Baumslag–Solitar group $\mathrm{BS}(m, n)$

[12] This follows from the group $\mathrm{BS}(m, n)$ being residually finite when $|m| = |n|$ [31, Proposition 2.6(3)]

[13] This argument is similar to the proof of Theorem 5 using Lemma 4.3 in Section 4

- **Query efficiency:** the algorithm is only allowed to make q queries, where each query is of the form "what is $s_G x$?", $x \in [n]$, $s \in S^{\pm}$.

If for every $\varepsilon > 0$ there are $q = q(\varepsilon) \in \mathbb{N}$ and an (ε, q)-flexible-tester $\mathcal{M}_\varepsilon$ for E then we say that E is $q(\varepsilon)$-*flexibly-testable* (or just *flexibly testable*, or *testable under the flexible model*) and that $\varepsilon \mapsto \mathcal{M}_\varepsilon$ is a family of flexible testers for E.

Clearly, every testable system of relations is also flexibly testable. As shown in Example 2.4, not every system of relations is flexibly testable.

Definition 2.6 (Flexibly-stable system of relations)**.** We say that E is *flexibly stable* if there is a function $k\colon \mathbb{R}_{>0} \to \mathbb{N}$ such that $\varepsilon \mapsto \mathrm{SAS}^{E}_{k(\varepsilon)}$ is a family of flexible testers for E.

Clearly, if E is stable then it is flexibly stable. The converse is still open:

Problem 2.7. Is there a flexibly-stable system of relations that is not stable?

An open problem similar to Problem 2.7, in the context of testability, is stated and discussed in Section 2.2.7.

Remark 2.8. We note that the sufficient condition for instability in [9, Theorem 1.3(i)] is in fact sufficient for flexible instability (this follows directly from the proof given in [9] which was written before the notion of flexible stability was defined formally).

The following deep result provides examples of flexibly-stable systems, none of which are known to be stable.

Example 2.9. [27, Theorem 1.1] *For $g \geq 2$, the system $E_g = \{[s_1, s_2][s_3, s_4] \cdots [s_{2g-1}, s_{2g}] = 1\}$, consisting of a single relation, is flexibly stable. Here $[a, b] \coloneqq aba^{-1}b^{-1}$.*

More precisely, $\varepsilon \mapsto \mathrm{SAS}^{E_g}_{k(\varepsilon)}$ is a family of flexible testers for E_g for $k(\varepsilon) = O\big(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\big)$.

We refer the reader to [8, Section 4.4] and [25, Section 3] for further information regarding flexible stability.

2.1.4. *Query efficiency.* It is interesting to study the query complexity of a stable system of relations E in terms of ε. That is, how small can $k(\varepsilon)$ be such that $\varepsilon \mapsto \mathrm{SAS}^{E}_{k(\varepsilon)}$ is a family of testers for E.

When $|S| = d$, $d \geq 1$, and

$$E = \{ss' = s's \mid s, s' \in S\},$$

the main result of [10] (see Example 1.11) gives an upper bound on $k(\varepsilon)$ which is polynomial in $\frac{1}{\varepsilon}$. On the other hand, [10, Theorem 1.17] shows that $k(\varepsilon) \geq \Omega\Big(\big(\frac{1}{\varepsilon}\big)^d\Big)$ if $d \geq 2$. As for a lower bound on $k(\varepsilon)$ under the flexible model, we know that $k(\varepsilon) \geq \Omega\Big(\big(\frac{1}{\varepsilon}\big)^2\Big)$ if $d \geq 2$ (stated without proof in [10, Section 6]). In particular, the bound $k(\varepsilon) \leq O\big(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\big)$ in Example 2.9 does not hold when $g = 1$.

2.1.5. *Stability of an infinite system of relations.* Our definition of a system of relations E requires E to be finite. It is possible to define stability even when E is infinite: E is stable if $\varepsilon \mapsto \mathrm{SAS}^{E'(\varepsilon)}_{k(\varepsilon)}$ is a family of testers for E for some $k\colon \mathbb{R}_{>0} \to \mathbb{N}$ and $E'\colon \mathbb{R}_{>0} \to \mathrm{FinSubsets}(E)$. Examples of stable infinite systems, which are not equivalent to any finite system, are given in [42, Theorem 1.7], [29] and [30], using [9, Theorem 1.3(ii)].

2.1.6. *Matrices instead of permutations.* The general question of whether approximate solutions are close to solutions, in various contexts, was suggested by Ulam [39, Chapter VI]. Definition 2.1 puts the notion of stability in permutations into this framework, and follows the earlier notion of stability in matrices. The classical question in the latter context is whether for all $n \times n$ matrices A and B such that $\|AB - BA\| < \delta$ there are $n \times n$ matrices A' and B' such that $A'B' = B'A'$ and $\|A - A'\| + \|B - B'\| < \varepsilon$, where $\varepsilon = \varepsilon(\delta) \xrightarrow{\delta \to 0} 0$. The answer depends on the type of matrices considered (self-adjoint, unitary, etc.) and the matrix norm used. See the introduction of [6] for a short survey of stability of the relation $\mathsf{XY} = \mathsf{YX}$ in matrices, and [8, 15, 16, 19, 21, 24, 32, 35, 37] for newer works that consider more general relations.

2.2. **Directions for further research.** This work originated from the observation that stability, which has been studied extensively in the context of group theory, admits a natural equivalent definition in the language of property testing (Definition 1.10). This gave rise to the more general notion of testability of relations between permutations (Definition 1.6).

We hope that the present paper will stimulate more work on testability of relations. While Theorems 1 and 2 generate large families of testable and non-testable relations, many fundamental questions remain open. Below we present suggestions for further research (see Section 3 for several additional open problems).

2.2.1. *A complete characterization of the testable systems of relations.* The most direct goal of our line of research is to classify every system of relations as either testable or non-testable. We do not know whether this problem is computable, namely, it may be undecidable to determine, given a system of relations E as input, whether E is testable. Regardless of this potential obstacle, we seek a natural characterization for testability, and note that such a characterization need not necessarily be computable.

Problem 2.10. Obtain a natural geometric characterization for testable systems of relations (i.e., a characterization in terms of properties of the graphs GSol_E)

The present work advances us towards this goal via Theorems 1 and 2, which yield large natural sets of positive and negative examples, and hint at a complete characterization in terms of the geometry of the graph family GSol_E.

A natural approach to Problem 2.10 is concentrating on systems of relations that have evaded our methods so far. Perhaps the most prominent candidates are the Baumslag-Solitar systems of relations (see Example 2.4):

Problem 2.11. Let $m, n \in \mathbb{Z}$. Is $E_{m,n} = \{XY^m = Y^n X\}$ testable? If not, is it flexibly testable?

The only open cases in Problem 2.11 are when $|m|$ and $|n|$ are distinct and larger than 1 (see Example 2.4). In all of these open cases, $E_{m,n}$ is not flexibly stable (see Example 2.4 and Remark 2.8).

2.2.2. *Query efficiency.* A testable system of relations E, as in Definition 1.6, is a system that admits an (ε, q)-tester $\mathcal{M}_\varepsilon$ for all $\varepsilon > 0$, where q depends only on ε and E (but not on the input size n). From a computational perspective, it is desirable to derive explicit good upper bounds on q in terms of ε. Examples 1.11 and 2.9 establish results of this kind in the context of stability, as discussed in Section 2.1.4. It is interesting to study the quantitative aspect in the broader context of testability. One natural question is whether Theorem 1 can be extended to a quantitative statement.

Problem 2.12. Let E satisfy the hypothesis of Theorem 1. What is the minimum query complexity of a tester for E?

See Section 3.3 for a conjecture about Problem 2.12.

In line with Examples 1.11 and 2.9, it is interesting to find additional families of $\mathrm{poly}(1/\varepsilon)$-testable and $\mathrm{poly}(1/\varepsilon)$-stable systems of relations.

Problem 2.13. Which systems of relations are $O\big(\mathrm{poly}\big(\frac{1}{\varepsilon}\big)\big)$-testable?

The more refined question of the degree of polynomial stability or testability is also interesting. The following is still open.

Problem 2.14. What is the minimal[14] D such that $E = \{\mathsf{XY} = \mathsf{YX}\}$ is $O\Big(\big(\frac{1}{\varepsilon}\big)^D\Big)$-stable?

[14]More precisely, we ask about the infimum of the set of all D such that E is $O\Big(\big(\frac{1}{\varepsilon}\big)^D\Big)$-stable (we do not know if the minimum is attained).

As discussed in Section 2.1.4, the minimal D is at least 2. One can derive an explicit upper bound on D by following the proof of [10, Theorem 1.16].

Another question is whether there are systems of relations that are stable, but where an algorithm different than `Sample and Substitute` would yield better query complexity. For example:

Problem 2.15. Let D be minimal such that $E = \{\mathsf{XY} = \mathsf{YX}\}$ is $O\left(\left(\frac{1}{\varepsilon}\right)^D\right)$-stable. Is there some $D' < D$? such that E is $O\left(\left(\frac{1}{\varepsilon}\right)^{D'}\right)$-testable?

In the context of lower bounds, the following is also still open.

Problem 2.16. Is there a system of relations E which is testable, but not $O\left(\left(\frac{1}{\varepsilon}\right)^D\right)$-testable for any D? Similarly, is there E which is stable but not $O\left(\left(\frac{1}{\varepsilon}\right)^D\right)$-stable for any D?

2.2.3. *Allowing the query complexity to depend on n.* In this work, as in, e.g., [4,5,36], the query complexity of a tester must not depend on the input size. In general, however, testers whose query complexity depends on the input size in a sublinear fashion are widely studied (see, e.g., [23]). Allowing such testers raises many interesting questions, such as the following.

Problem 2.17. Consider the system of relations $E = E_3$, discussed in Section 3.2 and defined in Appendix A.2. This system satisfies the hypothesis of Theorem 2 and thus it is not testable. Is there a probabilistic algorithm that distinguishes between elements of $\mathrm{GSol}_{E_3}(n)$ and $\mathrm{GSol}^{\geq\varepsilon}_{E_3}(n)$, in the sense of Definition 1.6, by making only $q = q(\varepsilon, n)$ queries? Here q must be sublinear in n for a result to be interesting, and the question is how small can q be in terms of n (the same question is open for every other system E that satisfies the hypothesis of Theorem 2).

2.2.4. *Running time efficiency.* Another aspect that we do not pursue in this work is the time complexity of our testers. In the case of the Sample and Substitute algorithm, the time complexity is easily seen to be identical to the query complexity, so not much remains to be optimized.

On the other hand, as explained in Remark 1.13, a straightforward implementation of Local Statistics Matcher requires time exponential in n to determine whether Condition (1.4) holds. However, when the graphs in GSol_E are "well behaved", the set of distributions $\{N_{G,P} \mid G \in \mathrm{GSol}_E(n)\}$ may be structured enough to allow this condition to be checked much more efficiently, perhaps even in time independent of n.

Problem 2.18. For which testable systems of relations E, having $\varepsilon \mapsto \mathrm{LSM}^E_{k(\varepsilon),P(\varepsilon),\delta(\varepsilon)}$ as a family of testers, does there exist an implementation of $\mathrm{LSM}^E_{k(\varepsilon),P(\varepsilon),\delta(\varepsilon)}$ that is time-efficient in terms on n (or better, has running time independent of n)?

Additionally, despite the universality of Local Statistics Matcher (see Theorem 3), it may also be worthwhile to seek other, more time-efficient, testing algorithms.

In the context of Problem 2.18, it is worth noting that for a $k(\varepsilon)$-stable system E, $\varepsilon \mapsto \mathrm{LSM}^E_{k(\varepsilon),R_E,0}$ is a family of testers (here we plug R_E itself into LSM in the role of the word-set parameter P). Calculating the left-hand side of (1.4) in $\mathrm{LSM}^E_{k(\varepsilon),R_E,0}$ is trivial because N_{H,R_E} is the Dirac distribution concentrated on R_E for each $H \in \mathrm{GSol}_E(n)$, $n \in \mathbb{N}$. Thus, in this case the running time discussed in Problem 2.18 does not depend on n. Problem 2.18 asks if there are also testable instable systems with this property, or at least with running time depending weakly on n.

2.2.5. *Uniform testability.* Following [36, Definition 2.4], our notion of testability from Definition 1.6 is *nonuniform in ε,* in the sense that it requires a family of testers, one for each ε. In the more standard notion of testability, which we refer to as *uniform testability* (see [23, Definition 1.6]), there is just a single tester, and it takes ϵ as input. To upgrade our testability to uniform testability, we need the function $\epsilon \mapsto \mathcal{M}_\varepsilon$, that chooses the tester according to ε, to be computable.

We would like to understand which systems of relations are uniformly testable, and whether there is a gap between the uniform and nonuniform notions. In particular,

Problem 2.19. Can Theorem 1 be strengthened to prove uniform testability for some systems of relations?

This problem is further discussed in Section 3.4.

2.2.6. *POT-testability and fixed-radius sampling.* Proximity Oblivious Testability (POT) is a stronger form of testability. A system of relations E is *POT-testable* if it can be tested by repeating the same random boolean subroutine A for $k(\varepsilon)$ independent iterations, and accepting if the number of iterations in which A returns `true` is above a certain threshold. Notably, A itself must not depend on ε. See [23, Definition 1.7] for a precise definition of POT testability in a general context. A stable system of relations is POT-testable because SAS_k^E can be implemented by running SAS_1^E for k independent iterations, and accepting if all iterations accept.

A related notion is that of *fixed-radius sampling* (FRS). We say that a system of relations E is *FRS-testable* if it admits a family of testers $\varepsilon \mapsto \mathcal{M}_\varepsilon$ such that $\mathcal{M}_\varepsilon$, in its run on $G \in \mathcal{G}_S(n)$, repeats the same subroutine A for $k(\varepsilon)$ independent iterations and accepts or rejects according to the results of these iterations. Again, A must be independent of ε, but unlike the case of POT-testability, A does not have to be a boolean subroutine, and $\mathcal{M}_\varepsilon$ does not have to decide on its return value according to a threshold. The term *fixed radius* comes from the fact that each iteration of A on its input G may only examine a constant (i.e., independent of ε) number of vertices of the graph G. Without loss of generality, we may assume that it examines a constant number of balls in this graph of some fixed radius r. Thus, the tester $\mathcal{M}_\varepsilon$ has information about $k(\varepsilon)$ balls of radius r in the graph G.

Clearly, every POT-testable system of relations is also FRS-testable. In addition, if E is testable by the family $\varepsilon \mapsto \mathrm{LSM}^E_{k(\varepsilon),P,\delta(\varepsilon)}$ where P is constant, $k\colon \mathbb{R}_{>0} \to \mathbb{N}$ and $\delta\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, then E is FRS-testable.

We have thus established the hierarchy:

$$\text{stability} \implies \text{POT-testability} \implies \text{FRS-testability} \implies \text{testability}\,. \tag{2.1}$$

As discussed earlier, there are systems of relations that are testable but are not stable (see Section 3.1). Thus, the class of stable systems of relations is strictly contained in the class of testable systems of relations.

Problem 2.20. Determine which of the inclusions arising from (2.1), if any, are equalities.

2.2.7. *The flexible model.* Consider the flexible model of stability and testability as in Section 2.1.3. In line with Problem 2.7, we ask:

Problem 2.21. Are there flexibly-testable systems of relations which are not testable?

We note that Theorem 2 provides non-testable (and thus also instable) systems in the strict model, as in Definition 1.6, but the proof of Theorem 2 does not rule out flexible testability. Finding a flexibly-testable system satisfying the hypothesis of Theorem 2 will solve Problem 2.21.

Another possible source for a positive answer to Problem 2.21 is Example 2.9, which provides systems which are flexibly stable (and hence are flexibly testable). It is not known whether these systems are testable.

3. The connection to group theory

Most of the previous study of stability of relations between permutations has been done through a connection between stability and group theory. This section reviews this connection, introduces the connection between testability and group theory, and gives a brief introduction to the relevant group-theoretic notions.

Let E be a system of relations over $S^{\pm}$. Then E gives rise to a finitely-presented group $\Gamma(E)$ by means of a group presentation. For example, the system $E = \{\mathsf{XY} = \mathsf{YX}\}$ over $S = \{\mathsf{X}, \mathsf{Y}\}$ gives rise to the group $\Gamma(E) = \langle \mathsf{X}, \mathsf{Y} \mid \mathsf{XY} = \mathsf{YX} \rangle \cong \mathbb{Z}^2$. More generally, we define $\Gamma(E) \coloneqq \langle S \mid E \rangle$ (this is the group generated by S subject to the relations E; see Appendix C for a reminder about group presentations). The association $E \mapsto \Gamma(E)$ is many-to-one. That is, different systems of relations may give rise to isomorphic groups. The starting point of the group-theoretic approach to stability is the following observation from [6] (see also [10, Proposition 1.11]).

Proposition 3.1. [6, Section 3] *Let E_1 and E_2 be systems of relations (over possibly different sets of variables) such that the groups $\Gamma(E_1)$ and $\Gamma(E_2)$ are isomorphic. Then E_1 is stable if and only if E_2 is stable.*

In Section 6 we lay the foundations for the study of testability via group theory by proving the following analogue of Proposition 3.1.

Proposition 3.2. *Let E_1 and E_2 be systems of relations (over possibly different sets of variables) such that the groups $\Gamma(E_1)$ and $\Gamma(E_2)$ are isomorphic. Then E_1 is testable if and only if E_2 is testable.*

Propositions 3.1 and 3.2 suggest that one can study the stability and testability of the system E by studying the group $\Gamma(E)$. The following theorem is an example of this strategy.

Theorem 4. [6] *If the group $\Gamma(E)$ is abelian, then E is stable.*

Furthermore [10, Theorem 1.16]*, in this case $\varepsilon \mapsto \mathrm{SAS}^E_{k(\varepsilon)}$ is a family of testers for E, where $k(\varepsilon) \le C \cdot \left(\frac{1}{\varepsilon}\right)^D$. Here C depends on E, and D depends only on the isomorphism class of the group $\Gamma(E)$.*

Theorem 4 applies to the system of relations

$$E^d_{\mathrm{comm}} \coloneqq \{s_i s_j = s_j s_i \mid i, j \in [d]\} \ , \tag{3.1}$$

since $\Gamma\left(E^d_{\mathrm{comm}}\right) \cong \mathbb{Z}^d$ is abelian. In particular, the theorem applies to $E^2_{\mathrm{comm}} = \{\mathsf{XY} = \mathsf{YX}\}$.

Classical group properties can also be used to prove instability. For example, the instability of $E_{2,3} \coloneqq \left\{\mathsf{XY}^2 = \mathsf{Y}^3\mathsf{X}\right\}$ [22, Theorem 2] follows from properties of the group $\mathrm{BS}(2,3) \coloneqq \langle \mathsf{X}, \mathsf{Y} \mid \mathsf{XY}^2 = \mathsf{Y}^3\mathsf{X}\rangle$ (more preciesely, $E_{2,3}$ is not stable because $\mathrm{BS}(2,3)$ is sofic but not residually finite).

Our main theorems, namely, Theorems 1 and 2, can be formulated in group-theoretic terms, and these formulations allow us to find many systems of relations to which the theorems apply. For this we need the notions of amenability and property (τ) (the latter is a variant of the well-known Kazhdan property (T)). One of the reasons for the wealth of examples and applications of these notions is that each of them has many equivalent definitions. For groups of the form $\Gamma(E)$, i.e., for finitely-presented groups, we give simple definitions, using the isoperimetric quantities presented in the introduction, as follows[15].

The group $\Gamma(E)$ is amenable if and only if $\rho(G) = 0$ for every $G \in \mathrm{GSol}_E$. Thus Theorem 1 is equivalent to the following theorem:

Theorem 1' (Main positive theorem in group terms)**.** *If the group $\Gamma(E)$ is amenable then E is testable.*

It is worth noting that the Cayley graph C of the group $\Gamma(E)$ belongs to GSol_E, and that if $\rho(C) = 0$ then $\rho(G) = 0$ for each $G \in \mathrm{GSol}_E$. Thus the hypothesis of Theorem 1 can be reduced to an assumption about the Cayley graph of $\Gamma(E)$ only, rather than the family GSol_E of graphs.

The group $\Gamma(E)$ has property (τ) if and only if $\inf\{\alpha(G) \mid G \in \mathrm{FGSol}_E\} > 0$. Furthermore, FGSol_E is infinite if and only if the group $\Gamma(E)$ has infinitely many finite quotients. Thus, Theorem 2 is equivalent to the following theorem:

Theorem 2' (Main negative theorem in group terms)**.** *If the group $\Gamma(E)$ has property (τ) and infinitely many finite quotients then E is non-testable.*

The rest of this section discusses applications of Theorems 1 and 2 and other aspects of testability related to the connection to group theory.

[15] The definitions given here of amenability and property (τ) for a group of the form $\Gamma(E)$ refer to E itself. There are equivalent definitions of amenability and property (τ) for a group Δ that are intrinsic to the group and do not refer to a system of relations E such that $\Delta \cong \Gamma(E)$. See [17, Chapter 18], [11] and [33] for more information on amenability, property (T) and property (τ), respectively

3.1. **Theorem 1 yields testable instable systems of relations.** The class of amenable groups is vast and contains all solvable groups, and thus, by Theorem 1', E is testable whenever $\Gamma(E)$ is solvable. But even in this case, E is not necessarily stable. Indeed, [9, Theorem 1.3(ii)] provides a group-theoretic condition $(*)$ such that if $\Gamma(E)$ is amenable then E is stable if and only if $\Gamma(E)$ satisfies $(*)$. Using this characterization of stability, [9, Theorem 1.2(iii)] provides an instable system of relations E_p, for each prime number p, such that $\Gamma(E_p)$ is solvable. But E_p is testable by Theorem 1'. Thus we have infinitely many examples of instable testable systems of relations. An explicit description of E_p and $\Gamma(E_p)$ is given in Appendix A.1.

3.2. **Theorem 2 yields non-testable systems of relations.** The class of finitely-presented groups with property (τ) is also vast, and contains all finitely-presented groups that have property (T). For example, the finitely-presented group $\mathrm{SL}_m \mathbb{Z}$, $m \geq 3$, has infinitely many finite quotients and has property (T), and thus, by Theorem 2', E is non-testable if $\Gamma(E) \cong \mathrm{SL}_m \mathbb{Z}$. In Appendix A.2 we describe a system E_m such that $\Gamma(E_m) \cong \mathrm{SL}_m \mathbb{Z}$.

3.3. **Query efficiency and group theory.** Let E_1 and E_2 be systems of relations such that $\Gamma(E_1) \cong \Gamma(E_2)$. Proposition 3.1 states that if E_1 is stable then so is E_2. Furthermore, [10, Proposition 1.11] strengthens Proposition 3.1 by showing that if E_1 is $q(\varepsilon)$-stable then E_2 is $cq(\varepsilon)$-stable for a constant $c = c(E_1, E_2)$ (except for in the trivial case where $\mathrm{GSol}_{E_1}(n) = \mathcal{G}_S(n)$ for all $n \in \mathbb{N}$).

Assuming that E_1 is testable (and thus, so is E_2), Proposition 6.1 describes a quantitative relationship between the statistical-distinguishability parameters of E_1 and E_2. It would be interesting to find an explicit quantitative relationship between the query complexities required to test these systems of relations. Such a result may be relevant to the problems discussed in Section 2.2.2.

In the context of Problem 2.12, it is worthwhile to mention the *Følner function* [40] of the group $\Gamma(E)$ with respect to S. Informally, assuming that E satisfies the hypothesis of Theorem 1, the Følner function measures "how quickly" E does so. We conjecture that E as above is $q(\varepsilon)$-testable for some function $q(\varepsilon)$, given in terms of $|S|$, $\sum_{w\in R_E} |w|$ and the Følner function.

3.4. **Uniform testability and group theory.** For systems E_1 and E_2 such that $\Gamma(E_1) \cong \Gamma(E_2)$, if E_1 is uniformly testable (see Section 2.2.5) then the same is true for E_2. This follows from the method of Section 6 (see the explicit bounds in Proposition 6.1). It is interesting to study which systems of relations are uniformly testable. In particular, we would like to know if there is a uniform version of Theorem 1, as asked in Problem 2.19. We suspect that if the Følner function of $\Gamma(E)$ is computable then E is uniformly testable.

4. Proof of Theorem 2

Here we prove Theorem 2, which is equivalent, in light of Theorem 3, to the following result.

Theorem 5. *Fix a finite alphabet S. Let E be a system of relations over $S^{\pm}$ such that FGSol_E is infinite and*

$$\inf\{\alpha(G) \mid G \in \mathrm{FGSol}_E\} > 0. \tag{4.1}$$

Then E is not statistically distinguishable.

The rest of this section is devoted to proving Theorem 5. Our proof strengthens the argument of [8, Theorem 1.4], which proves, under similar assumptions, that E is not stable.

We start with a proof sketch. Let E be as in Theorem 5. For every $P \in \mathrm{FinSubsets}(F_S)$, we need to produce a sequence of graphs $(G_m)_{m=1}^{\infty}$, $G_m \in \mathcal{G}_S(n_m)$, $n_m \in \mathbb{N}$, such that

$$d_{\mathrm{TV}}\left(N_{G_m,P}, N_{G'_m,P}\right) \overset{m\to\infty}{\longrightarrow} 0 \tag{4.2}$$

for some $G'_m \in \mathrm{GSol}_E(n_m)$, but

$$d^{\mathrm{H}}(G_m, G''_m) \geq \varepsilon_0 \quad \forall m \in \mathbb{N}\ \forall G''_m \in \mathrm{GSol}_E(n_m)\ , \tag{4.3}$$

where $\varepsilon_0 > 0$ depends only on E. That is, G_m is close to some solution for E under the P-local metric, but far from every solution under the global metric (see Remark 1.15).

To produce G_m, we take a connected graph $\tilde{G}_m \in \mathrm{GSol}_E(n_m+1)$, $n_m \in \mathbb{N}$, $n_m \overset{m\to\infty}{\longrightarrow} \infty$, make a local change such that n_m+1 becomes an isolated vertex, and then remove the vertex n_m+1 and obtain $G_m \in \mathcal{G}_S(n_m)$. Then

$$d_{\mathrm{TV}}\left(N_{G_m,P}, N_{\tilde{G}_m,P}\right) \overset{m\to\infty}{\longrightarrow} 0 \ , \tag{4.4}$$

but, as shown by Lemma 4.1 below,

$$d^{\mathrm{H}}(G_m, G''_m) \geq \frac{1}{2|S|}\alpha(E) \quad \forall G''_m \in \mathrm{GSol}_E(n_m) \tag{4.5}$$

for all large enough m. Denote

$$\alpha(E) = \inf\{\alpha(G) \mid G \in \mathrm{FGSol}_E\}.$$

Set $\varepsilon_0 = \frac{1}{2}\alpha(E)$. By (4.1), $\varepsilon_0 > 0$, and thus (4.4) and (4.5) almost prove the desired (4.2) and (4.3), with the caveat that $\tilde{G}_m \in \mathrm{GSol}_E(n_m+1)$, while we need this graph to be in $\mathrm{GSol}_E(n_m)$. Lemma 4.3 helps us overcome this final difficulty. We now give a complete proof based on the proof sketch above.

For $n \geq 2$, define a function $\mathrm{res}_n \colon \mathcal{G}_S(n) \to \mathcal{G}_S(n-1)$ by letting $\hat{G} = \mathrm{res}_n(G)$ be the graph on the vertex set $[n-1]$ such that

$$s_{\hat{G}}x = \begin{cases} s_G x & s_G x \neq n \\ s_G s_G x & s_G x = n \end{cases}$$

for all $x \in [n-1]$ and $s \in S$. That is, the edge set of $\mathrm{res}_n(G)$ consists of all edges of $\hat{G}$ that are not incident to the vertex n, and all edges of the form $s_G^{-1}n \overset{s}{\longrightarrow} s_G n$ for all $s \in S$ such that $s_G n \neq n$.

The following lemma shows that $\mathrm{res}_n(G)$ is far from $\mathrm{GSol}_E(n-1)$ whenever E satisfies (4.1). The lemma and its proof can be seen as a combinatorial version of ideas from [8].

Lemma 4.1. *Let E be a system of relations over $S^{\pm}$. Take a connected graph $G \in \mathrm{GSol}_E(n)$, $n \geq 2$, and denote $\hat{G} = \mathrm{res}_n(G)$. Then*

$$d^{\mathrm{H}}\left(\hat{G}, G'\right) \geq \frac{\alpha(E)}{|S|} - \frac{1}{n-1} \quad \forall G' \in \mathrm{GSol}_E(n-1) \ .$$

Proof. Take $G' \in \mathrm{GSol}_E(n-1)$. Consider the product S-graph $G' \times G$. This is the S-graph on the vertex set $[n-1] \times [n]$ such that

$$s_{G'\times G}(x,y) = (s_{G'}x, s_G y) \ .$$

Write $X_1, \dots, X_c$ for the connected components of $G' \times G$, and let $D = \{(x,x) \mid x \in [n-1]\} \subset [n-1] \times [n]$. We claim that

$$|D \cap X_i| \leq \frac{1}{2}|X_i| \quad \forall 1 \leq i \leq c \ . \tag{4.6}$$

Let $1 \leq i \leq c$. If $D \cap X_i = \emptyset$ we are done. Otherwise, fix an arbitrary $(x,x) \in X_i$, $x \in [n-1]$. The connected component Y of x in G' clearly has at most $n-1$ vertices. On the other hand, the connected component of x in G has exactly n vertices because G is connected. Accordingly, $[F_S : \mathrm{Stab}_{G'}(x)] = |Y| \leq n-1$ and $[F_S : \mathrm{Stab}_G(x)] = n$ by the Orbit–Stabilizer Theorem[16], and so $\mathrm{Stab}_{G'}(x)$ is not contained in $\mathrm{Stab}_G(x)$. Thus, the inclusion in

$$\mathrm{Stab}_{G'\times G}((x,x)) = \mathrm{Stab}_{G'}(x) \cap \mathrm{Stab}_G(x) \subset \mathrm{Stab}_{G'}(x)$$

is strict, and therefore $m \coloneqq [\mathrm{Stab}_{G'}(x) : \mathrm{Stab}_{G'\times G}((x,x))]$ is at least 2. Using the Orbit–Stabilizer Theorem again,

$$|X_i| = [F_S : \mathrm{Stab}_{G'\times G}((x,x))] = m[F_S : \mathrm{Stab}_{G'}(x)] \geq 2|Y| \ .$$

[16]The theorem says, in our terminology, that the index $[F_S : \mathrm{Stab}_H(x)]$ of the subgroup $\mathrm{Stab}_H(x)$ of F_S is equal to the number of vertices of H for every connected S-graph H and vertex x of H.

Thus (4.6) follows since $|D \cap X_i| \leq |Y|$ (because Y is image of X_i under the projection onto the first coordinate $[n-1] \times [n] \to [n-1]$, and this projection is injective on D).

Since the graph X_i is a connected solution for E (namely, it is in FGSol_E), (4.6) implies that

$$\sum_{s \in S} |(s_{G' \times G}(D \cap X_i)) \setminus (D \cap X_i)| \geq \alpha(X_i)|D \cap X_i| \quad \forall 1 \leq i \leq c.$$

Thus

$$\begin{aligned} \sum_{s\in S} |(s_{G'\times G} D) \setminus D| &= \sum_{i=1}^{c} \sum_{s \in S} |(s_{G' \times G}(D \cap X_i)) \setminus (D \cap X_i)| \\ &\geq \sum_{i=1}^{c} \alpha(X_i)|D \cap X_i| \geq \alpha(E) \underbrace{|D|}_{=n-1} , \end{aligned} \tag{4.7}$$

where the equality on the first line follows since $X_1, \dots, X_c$ are disjoint and $s_{G'\times G} X_i = X_i$ for all $1 \leq i \leq c$.

On the other hand,

$$\begin{aligned} \sum_{s\in S} |(s_{G'\times G} D) \setminus D| &= \sum_{s \in S} |\{x \in [n-1] \mid s_{G'} x \neq s_G x\}| \\ &\leq \sum_{s\in S} \left|\left\{x \in [n-1] \mid s_{G'}x \neq s_{\hat{G}} x \text{ or } s_{\hat{G}} x \neq s_G x\right\}\right| \\ &\leq \sum_{s \in S} \left(\underbrace{\left|\left\{x \in [n-1] \mid s_{G'}x \neq s_{\hat{G}}x\right\}\right|}_{=(n-1)d^{\mathrm{H}}(\hat{G},G')} + \underbrace{\left|\left\{x \in [n-1] \mid s_{\hat{G}}x \neq s_G x\right\}\right|}_{\leq 1} \right) \\ &\leq |S|\left((n-1)d^{\mathrm{H}}\left(\hat{G}, G'\right) + 1\right) . \end{aligned} \tag{4.8}$$

The claim follows from (4.7) and (4.8). □

Let $G \in \mathcal{G}_S(n)$, write $\hat{G} = \mathrm{res}_n(G)$, and take $P \in \mathrm{FinSubsets}(F_S)$. Denote

$$C = \left\{v_G^{-1} n \mid v \text{ is a suffix of at least one } w \in P\right\} .$$

Then $w_G x = w_{\hat{G}} x$ for each $x \in [n-1] \setminus C$, and thus $\mathrm{Stab}_G(x) \cap P = \mathrm{Stab}_{\hat{G}}(x) \cap P$. Since $|C| \leq (\mathrm{TotalSize}(P))^2$, one can easily deduce the following lemma.

Lemma 4.2. *Let $G \in \mathcal{G}_S(n)$, write $\hat{G} = \mathrm{res}_n(G)$, and take $P \in \mathrm{FinSubsets}(F_S)$. Then*

$$d_{\mathrm{TV}}\left(N_{G,P}, N_{\hat{G},P}\right) \leq O_P\left(\frac{1}{n}\right) .$$

Let $P \in \mathrm{FinSubsets}(F_S)$. For an S-graph G, we view $N_{G,P}$ as a vector in $\mathbb{R}^{\mathrm{Subsets}(P)}$, belonging to the compact set $\mathrm{Prob}(\mathrm{Subsets}(P)) := \left\{f \colon \mathrm{Subsets}(P) \to [0,1] \mid \sum_{Q \in \mathrm{Subsets}(P)} f(Q) = 1\right\}$. For a sequence of S-graphs $(G_k)_{k=1}^{\infty}$, the sequence $(N_{G_k,P})_{k=1}^{\infty}$ is said to *converge* if it converges as a sequence of vectors in $\mathbb{R}^{\mathrm{Subsets}(P)}$. Intuitively, this means that the S-graphs $(G_k)_{k=1}^{\infty}$ tend toward having the same P-local statistics. This notion of convergence is used in the proof of Theorem 5 below. We use the fact that $\prod_{P \in \mathrm{FinSubsets}(F_S)} \mathrm{Prob}(\mathrm{Subsets}(P))$ is compact[17].

We also use the following lemma.

Lemma 4.3. [9, Lemma 7.6] *Let E be system of relations over $S^{\pm}$, and let $(G_k)_{k=1}^{\infty}$, $G_k \in \mathrm{GSol}_E(l_k)$, $l_k \in \mathbb{N}$, be a sequence of S-graphs such that the local statistics sequence $(N_{G_k,P})_{k=1}^{\infty}$ converges for each $P \in \mathrm{FinSubsets}(F_S)$. Take a sequence $(m_k)_{k=1}^{\infty}$ of integers such that $m_k \overset{k\to\infty}{\longrightarrow} \infty$. Then there is a sequence*

[17] This is equivalent to the fact that the space of S-graphs is compact under Benjamini–Schramm convergence, and to the fact that the space $\mathrm{IRS}(F_S)$ of invariant random subgroups is compact.

$(H_k)_{k=1}^{\infty}$, $H_k \in \mathrm{GSol}_E(m_k)$, *such that* $(N_{G_k,P})_{k=1}^{\infty}$ *and* $(N_{H_k,P})_{k=1}^{\infty}$ *converge to the same limit for each* $P \in \mathrm{FinSubsets}(F_S)$.

Proof of Theorem 5. By our assumption that FGSol_E is infinite (and thus contains graphs of unbounded cardinality), and the compactness of $\prod_{P\in\mathrm{FinSubsets}(F_S)} \mathrm{Prob}(\mathrm{Subsets}(P))$, there is a sequence of connected S-graphs $(G_k)_{k=1}^{\infty}$, $G_k \in \mathrm{GSol}_E(n_k)$, $n_k \overset{k\to\infty}{\longrightarrow} \infty$, such that $(N_{G_k,P})_{k=1}^{\infty}$ converges for each $P \in \mathrm{FinSubsets}(F_S)$. Let $\hat{G}_k = \mathrm{res}_{n_k}(G_k)$. By Lemma 4.2,

$$\lim_{k\to\infty} N_{\hat{G}_k,P} = \lim_{k\to\infty} N_{G_k,P} \ .$$

Write $\alpha = \alpha(E)$. By Lemma 4.1,

$$\hat{G}_k \in \mathrm{GSol}_E^{\geq \alpha/|S|-1/(n_k-1)}(n_k - 1) \qquad \forall k \in \mathbb{N} \ . \tag{4.9}$$

By applying Lemma 4.3 to the sequence $(G_k)_{k=1}^{\infty}$, with $m_k = n_k - 1$, we obtain a sequence of graphs $(H_k)_{k=1}^{\infty}$ such that

$$H_k \in \mathrm{GSol}_E(n_k - 1) \tag{4.10}$$

and

$$\lim_{k\to\infty} N_{H_k,P} = \lim_{k\to\infty} N_{G_k,P} \quad \forall P \in \mathrm{FinSubsets}(F_S) \ . \tag{4.11}$$

Thus

$$\lim_{k\to\infty} N_{H_k,P} = \lim_{k\to\infty} N_{\hat{G}_k,P} \quad \forall P \in \mathrm{FinSubsets}(F_S) \ . \tag{4.12}$$

The system E is not statistically distinguishable by (4.9), (4.10) and (4.12). □

5. Proof of Theorem 1

Let S be a finite alphabet and let E be a system of relations over $S^{\pm}$.

The proof of Theorem 1 relies on two deep results: The Ornstein–Weiss Theorem [38] (see also [14]) and the Newman–Sohler Theorem [36] (see also [20, Theorem 5]). These theorems are used together in [9, Proposition 6.8] in a simple manner to prove the following theorem (stated here in the language of the present paper).

Theorem 6. *Assume that* $\rho(G) = 0$ *for every* $G \in \mathrm{GSol}_E$. *Then for every* $\varepsilon > 0$ *there are* $r \in \mathbb{N}$ *and* $\delta > 0$ *such that* $d^{\mathrm{H}}(G, G') < \varepsilon$ *whenever* $n \in \mathbb{N}$, $G \in \mathrm{GSol}_E(n)$, $G' \in \mathcal{G}_S(n)$ *and* $d_{\mathrm{TV}}(N_{G,B_r}, N_{G',B_r}) < \delta$ *(where* $B_r = \{w \in F_S \mid |w| \leq r\}$*).*

Theorem 1 follows immediately from Theorems 3 and 6.

Proof of Theorem 1. Assume that $\rho(G) = 0$ for every $G \in \mathrm{GSol}_E$. By Theorem 6, there are functions $r\colon \mathbb{R}_{>0} \to \mathbb{N}$ and $\delta\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that E is $\big(B_{r(\varepsilon)}, \delta(\varepsilon)\big)$-statistically-distinguishable. Thus E is testable by Theorem 3. □

6. Proof of Proposition 3.2

Fix finite alphabets $S = \{s_1, \dots, s_{d_1}\}$ and $T = \{t_1, \dots, t_{d_2}\}$, and consider two systems of relations E_1 and E_2 over $S^{\pm}$ and $T^{\pm}$, respectively, such that the groups $\Gamma(E_1)$ and $\Gamma(E_2)$ are isomorphic. By Theorem 3, in order to prove Proposition 3.2, we need to prove that if E_2 is statistically distinguishable then so is E_1. This is achieved by Proposition 6.1 below.

Given a finite set of words P, denote

$$\mathrm{TotalSize}(P) = \sum_{x\in P} |x| \ .$$

Proposition 6.1. *If E_2 is $(P_2(\varepsilon),\delta_2(\varepsilon))$-statistically-distinguishable for $P_2\colon \mathbb{R}_{>0}\to \mathrm{FinSubsets}(F_T)$ and $\delta_2\colon \mathbb{R}_{>0}\to\mathbb{R}_{\geq 0}$, then E_1 is $(P_1(\varepsilon),\delta_1(\varepsilon))$-statistically-distinguishable for some $P_1\colon\mathbb{R}_{>0}\to\mathrm{FinSubsets}(F_S)$ and $\delta_1\colon\mathbb{R}_{>0}\to\mathbb{R}_{\geq 0}$ such that*

$$\mathrm{TotalSize}(P_1(\varepsilon)) = O_{E_1,E_2}(\mathrm{TotalSize}(P_2(\varepsilon))) \tag{6.1}$$

and

$$\delta_1(\varepsilon)=\Omega_{E_1,E_2}(\min(\delta_2(\varepsilon),\varepsilon)) \tag{6.2}$$

for every $\varepsilon>0$.

Explicitly,

$$P_1(\varepsilon)=\lambda_2(P_2(\varepsilon))\cup R_{E_1}$$

and

$$\delta_1(\varepsilon)=\min\left(\delta_2\left(\frac{\varepsilon}{2C_1}\right),\frac{\varepsilon}{2C_2}\right)\,,$$

for λ_2 as defined below and positive constants C_1 and C_2 that depend on E_1 and E_2.

The rest of this section is devoted to proving Proposition 6.1.

6.1. **A homomorphism view of $\mathcal{G}_S(n)$.** For $n\in\mathbb{N}$, let $\mathcal{H}_S(n)$ be the set of homomorphisms $F_S\to\mathrm{Sym}(n)$. For the sake of proving Proposition 6.1, it will be convenient to encode an S-graph in $\mathcal{G}_S(n)$ as a homomorphism in $\mathcal{H}_S(n)$, in a manner which we now describe.

For $G\in\mathcal{G}_S(n)$, let $f_G\in\mathcal{H}_S(n)$ be the $F_S\to\mathrm{Sym}(n)$ homomorphism that maps $w\in F_S$ to the permutation $i\mapsto w_G i$. Note that the map $G\mapsto f_G\colon\mathcal{G}_S(n)\to\mathcal{H}_S(n)$ is a bijection[18]

We translate some of the notions from the introduction to the language of homomorphisms. Given $f,g\in\mathcal{H}_S(n)$, define

$$d_n^{\mathrm{H}}(f,g)=d^{\mathrm{H}}(f,g)=\sum_{s\in S}\frac{1}{n}|\{x\in[n]\mid f(s)x\neq g(s)x\}|\,.$$

Let

$$\mathrm{HSol}_E(n)=\{f\in\mathcal{H}_S(n)\mid f(w)=\mathrm{id}\ \ \forall w\in R_E\}\,,$$

$$\mathrm{HSol}_E^{\geq\varepsilon}(n)=\left\{f\in\mathcal{H}_S(n)\mid d^{\mathrm{H}}(f,f')\geq\varepsilon\ \forall f'\in\mathrm{HSol}_E(n)\right\}\,.$$

For $x\in[n]$, let $\mathrm{Stab}_f(x)=\{w\in F_S\mid f(w)x=x\}$. For $P\in\mathrm{FinSubsets}(F_S)$, let $N_{f,P}$ be the distribution of $\mathrm{Stab}_f(x)\cap P$, where x is sampled uniformly from $[n]$. Then $\mathrm{HSol}_E(n)=\{f_G\mid G\in\mathrm{GSol}_E(n)\}$, $\mathrm{HSol}_E^{\geq\varepsilon}(n)=\left\{f_G\mid G\in\mathrm{GSol}_E^{\geq\varepsilon}(n)\right\}$, $\mathrm{Stab}_{f_G}(x)=\mathrm{Stab}_G(x)$ and $N_{f_G,P}=N_{G,P}$ for $G\in\mathcal{G}_S(n)$. We also denote $\mathrm{HSol}_E^{<\varepsilon}(n)\coloneqq\mathcal{H}_S(n)\setminus\mathrm{HSol}_E^{\geq\varepsilon}(n)$.

6.2. **The maps λ_1^* and λ_2^*.** To prove Proposition 6.1 we define two maps, $\lambda_1^*\colon\mathcal{H}_T(n)\to\mathcal{H}_S(n)$ and $\lambda_2^*\colon\mathcal{H}_S(n)\to\mathcal{H}_T(n)$, that behave nicely (see Section 6.3) with respect to the spaces $\mathrm{HSol}_{E_1}(n)$ and $\mathrm{HSol}_{E_2}(n)$, and also with respect to $\mathrm{HSol}_{E_1}^{<\varepsilon}(n)$ and $\mathrm{HSol}_{E_2}^{<\varepsilon}(n)$, $\varepsilon>0$.

We begin with the definitions of λ_1^* and λ_2^*. Write $\pi_1\colon F_S\twoheadrightarrow\Gamma(E_1)$ and $\pi_2\colon F_T\twoheadrightarrow\Gamma(E_2)$ for the quotient maps (see Appendix C), and fix an isomorphism $\theta\colon\Gamma(E_1)\to\Gamma(E_2)$. We "lift" the group isomorphisms $\theta\colon\Gamma(E_1)\to\Gamma(E_2)$ and $\theta^{-1}\colon\Gamma(E_2)\to\Gamma(E_1)$ to group homomorphisms

$$\lambda_1\colon F_S\to F_T \text{ and } \lambda_2\colon F_T\to F_S\,.$$

[18] This can be seen easily using the universal property of the free group F_S, which amounts to a bijection $(\mathrm{Sym}(n))^d\to\mathcal{H}_S(n)$, as recalled in Appendix C, and the bijection $\overline{\sigma}\mapsto G_{\overline{\sigma}}\colon(\mathrm{Sym}(n))^d\to\mathcal{G}_S(n)$ defined in Section 1.2.

More precisely, we fix homomorphisms λ_1 and λ_2 such that each of the two squares in the following diagram commutes (i.e., $\pi_2 \circ \lambda_1 = \theta \circ \pi_1$ and $\pi_1 \circ \lambda_2 = \theta^{-1} \circ \pi_2$):

(6.3)
$$\begin{array}{ccccc} F_S & \xrightarrow{\lambda_1} & F_T & \xrightarrow{\lambda_2} & F_S \\ \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi_2} & & \downarrow{\scriptstyle \pi_1} \\ \Gamma(E_1) & \xrightarrow{\theta} & \Gamma(E_2) & \xrightarrow{\theta^{-1}} & \Gamma(E_1) \end{array} .$$

Such maps λ_1 and λ_2 exist (but are generally not unique). Indeed, for $1 \le i \le d_1$ we set $\lambda_1(s_i) \in F_T$ to be an arbitrary word such that $\pi_2(\lambda_1(s_i)) = \theta(\pi_1(s_i))$, and note that λ_1 extends uniquely to a group homomorphism. We construct λ_2 similarly.

The maps λ_1 and λ_2 are not mutual inverses in general, but they enjoy inverse-like properties. By the commutativity of the diagram, $\pi_1 \circ \lambda_2 \circ \lambda_1 = \theta^{-1} \circ \pi_2 \circ \lambda_1 = \theta^{-1} \circ \theta \circ \pi_1 = \pi_1$. In particular, $(\lambda_2 \circ \lambda_1)(s_i) \in F_S$ and $s_i \in F_S$ belong to the same left coset of $\langle\langle R_{E_1} \rangle\rangle$ in F_S for each $1 \le i \le d_1$. That is,

(6.4)
$$(\lambda_2 \circ \lambda_1)(s_i) = s_i \prod_{j=1}^{m_i} v_{i,j} r_{i,j}^{\varepsilon_{i,j}} v_{i,j}^{-1} \ ,$$

where $m_i \ge 0$, $v_{i,j} \in F_S$, $r_{i,j} \in R_{E_1}$ and $\varepsilon_{i,j} \in \{\pm 1\}$. Set $Q_i \coloneqq \{v_{i,j} \mid 1 \le j \le m_i\}$.

For $n \in \mathbb{N}$, the homomorphisms $\lambda_1 \colon F_S \to F_T$ and $\lambda_2 \colon F_T \to F_S$ give rise to maps

$$\lambda_1^* \colon \mathcal{H}_T(n) \to \mathcal{H}_S(n) \text{ and}$$
$$\lambda_2^* \colon \mathcal{H}_S(n) \to \mathcal{H}_T(n)$$

given by

$$\lambda_1^* h = h \circ \lambda_1 \qquad \forall h \in \mathcal{H}_T(n) \text{ and}$$
$$\lambda_2^* f = f \circ \lambda_2 \qquad \forall f \in \mathcal{H}_S(n) \ .$$

Under the bijections $\mathcal{H}_S(n) \leftrightarrow \mathcal{G}_S(n)$ and $\leftrightarrow \mathcal{H}_T(n) \leftrightarrow \mathcal{G}_T(n)$ of Section 6.1, λ_1^* and λ_2^* give rise to maps between $\mathcal{G}_S(n)$ and $\mathcal{G}_T(n)$. For this section, the homomorphism view suffices, but the reader may find it instructive to spell out the definitions of λ_1^* and λ_2^* in terms of graphs.

6.3. **Properties of λ_1^* and λ_2^*.** We analyze the behavior of λ_1^* and λ_2^* with regard to both the global metric $d^{\mathrm{H}}(f,g)$, and the P-local metrics $d_{\mathrm{TV}}(N_{f,P}, N_{g,P})$, $P \in \mathrm{FinSubsets}(F_S)$ (see Remark 1.15). Due to symmetry, the claims in this section remain true if we swap the roles of S, E_1, λ_1 and T, E_2, λ_2.

6.3.1. *The global metric.*

Definition 6.2. Given $f \in \mathcal{H}_S(n)$, let $\mathrm{Bad}_{E_1}(f) = \{x \in [n] \mid \exists w \in R_{E_1}\ f(w)x \ne x\}$.

Remark 6.3. For $f \in \mathrm{HSol}_{E_1}(n)$, the probability distribution $N_{F,R_{E_1}}$ over $\mathrm{Subsets}(R_{E_1})$ assigns probability 1 to R_{E_1}. Thus,

$$d_{\mathrm{TV}}\big(N_{f,R_{E_1}}, N_{g,R_{E_1}}\big) = \Pr_{x \sim \mathrm{U}([n])}(\mathrm{Stab}_g(x) \cap R_{E_1} \ne R_{E_1}) = \frac{|\mathrm{Bad}_{E_1}(g)|}{n}$$

for all $g \in \mathcal{H}_S(n)$.

The following lemma shows that λ_1^* and λ_2^* enjoy inverse-like properties. More precisely, the lemma gives a tool for bounding the distance between f and $\lambda_1^*\lambda_2^* f$ for $f \in \mathcal{H}_S(n)$.

Lemma 6.4. *Let $f \in \mathcal{H}_S(n)$. Then $(\lambda_1^*\lambda_2^* f)(s_i)x = f(s_i)x$ for every $1 \le i \le d_1$ and $x \in [n]$ such that $x \notin \bigcup_{v \in Q_i}(f(v)\,\mathrm{Bad}_{E_1}(f))$.*

Proof. For $1 \le i \le d_1$ and $x \in [n]$,

$$\begin{aligned}(\lambda_1^*\lambda_2^* f)(s_i)x &= f((\lambda_2 \circ \lambda_1)(s_i))x \\ &= f(s_i)\left(\prod_{j=1}^{m_i} f(v_{i,j}) f\big(r_{i,j}^{\varepsilon_{i,j}}\big) f\big(v_{i,j}^{-1}\big)\right) x \;. && \text{by (6.4)}\end{aligned}$$

Thus, $(\lambda_1^*\lambda_2^* f)(s_i)x = f(s_i)x$ if

$$f\big(r_{i,j}^{\varepsilon_{i,j}}\big) f\big(v_{i,j}^{-1}\big)x = f\big(v_{i,j}^{-1}\big)x \qquad \forall 1 \le j \le m_i \;.$$

The latter condition is equivalent to

$$f(r_{i,j}) f\big(v_{i,j}^{-1}\big)x = f\big(v_{i,j}^{-1}\big)x \qquad \forall 1 \le j \le m_i \;.$$

This holds whenever $f(v)^{-1}x \notin \mathrm{Bad}_{E_1}(f)$ for all $v \in Q_i$, i.e., when $x \notin \bigcup_{v \in Q_i}(f(v)\,\mathrm{Bad}_{E_1}(f))$. $\square$

We conclude the following.

Corollary 6.5. *The image of* $\mathrm{HSol}_{E_1}(n)$ *under* λ_2^* *is contained in* $\mathrm{HSol}_{E_2}(n)$*. Furthermore, the restriction of* λ_2^* *to* $\mathrm{HSol}_{E_1}(n)$ *is a bijection* $\lambda_2^* |_{\mathrm{HSol}_{E_1}(n)} \colon \mathrm{HSol}_{E_1}(n) \to \mathrm{HSol}_{E_2}(n)$ *whose inverse is* $\lambda_1^* |_{\mathrm{HSol}_{E_2}(n)}$*.*

Proof. Let $f \in \mathrm{HSol}_{E_1}(n)$ and $v \in R_{E_2}$. Then $\pi_1(\lambda_2(v)) = \theta^{-1}(\pi_2(v)) = \theta^{-1}\big(1_{\Gamma(E_2)}\big) = 1_{\Gamma(E_1)}$ by (6.3). Thus $\lambda_2(v) \in \ker \pi_1 = \langle\!\langle E_1 \rangle\!\rangle$, and so $(\lambda_2^* f)(v) = f(\lambda_2(v)) = 1$. Hence $\lambda_2^* f \in \mathrm{HSol}_{E_2}(n)$. We proved that

$$\lambda_2^*(\mathrm{HSol}_{E_1}(n)) \subseteq \mathrm{HSol}_{E_2}(n) \;.$$

Similarly,

$$\lambda_1^*(\mathrm{HSol}_{E_2}(n)) \subset \mathrm{HSol}_{E_1}(n) \;.$$

Since $f \in \mathrm{HSol}_{E_1}(n)$ we have $\mathrm{Bad}_{E_1}(f) = \emptyset$. Consequently, Lemma 6.4 implies that $\lambda_1^*\lambda_2^* f = f$. Similarly, $\lambda_2^*\lambda_1^* h = h$ for $h \in \mathrm{HSol}_{E_2}(n)$, and the claim follows. $\square$

If $f \in \mathcal{H}_S(n)$ is not necessarily a solution for E_1, but the set $\mathrm{Bad}_{E_1}(f)$ is small, the following corollary shows that $\lambda_1^*\lambda_2^* f$ is close to f.

Corollary 6.6. *For* $f \in \mathcal{H}_S(n)$*,*

$$d^{\mathrm{H}}(f, \lambda_1^*\lambda_2^* f) \le \left(\sum_{i=1}^{d_1} |Q_i|\right) \frac{|\mathrm{Bad}_{E_1}(f)|}{n} \;.$$

Proof.

$$\begin{aligned} d^{\mathrm{H}}(f, \lambda_1^*\lambda_2^* f) &= \sum_{i=1}^{d_1} d^{\mathrm{H}}(f(s_i), (\lambda_1^*\lambda_2^* f)(s_i)) \\ &\le \sum_{i=1}^{d_1} \frac{1}{n} \left| \bigcup_{v \in Q_i} (f(v)\,\mathrm{Bad}_{E_1}(f)) \right| && \text{by Lemma 6.4} \\ &\le \sum_{i=1}^{d_1} \sum_{v \in Q_i} \frac{1}{n} |f(v)\,\mathrm{Bad}_{E_1}(f)| \\ &= \left(\sum_{i=1}^{d_1} |Q_i|\right) \frac{|\mathrm{Bad}_{E_1}(f)|}{n} \;. \end{aligned}$$

$\square$

Next, we study the interaction between λ_1^* and the Hamming metric.

Lemma 6.7. *Let* $h, h' \in \mathcal{H}_T(n)$*. Then* $d^{\mathrm{H}}(\lambda_1^* h, \lambda_1^* h') \le C d^{\mathrm{H}}(h, h')$*, where* $C = \Big(\sum_{i=1}^{d_1} |\lambda_1(s_i)|\Big)$*.*

Proof. For $w \in F_T$, let $A_w = \{x \in [n] \mid h(w)x = h'(w)x\}$ and $D_w = [n] \setminus A_w$. Write $\mathrm{suff}_j(w)$ for the suffix of length j of w (e.g., $\mathrm{suff}_2(ab^{-1}c^{-1}) = b^{-1}c^{-1}$). Write $\mathrm{last}_j(w)$ for the j-th letter of w, counting from the end (e.g., $\mathrm{last}_2(ab^{-1}c^{-1}) = b^{-1}$).

For $1 \le i \le d_1$,

$$
\begin{aligned}
& d^{\mathrm{H}}((\lambda_1^* h)(s_i), (\lambda_1^* h')(s_i)) \\
=& \frac{1}{n}\left|D_{\lambda_1(s_i)}\right| \\
\le& \frac{1}{n}\left|\bigcup_{j=1}^{|\lambda_1(s_i)|}\left(A_{\mathrm{suff}_{j-1}(\lambda_1(s_i))} \cap D_{\mathrm{suff}_j(\lambda_1(s_i))}\right)\right| \\
& \qquad \text{(since } D_w \subseteq \bigcup_{j=1}^{|w|} A_{\mathrm{suff}_{j-1}(w)} \cap D_{\mathrm{suff}_j(w)} \text{ for } w \in F_T\text{)} \\
\le& \frac{1}{n}\sum_{j=1}^{|\lambda_1(s_i)|}\left|A_{\mathrm{suff}_{j-1}(\lambda_1(s_i))} \cap D_{\mathrm{suff}_j(\lambda_1(s_i))}\right| \\
=& \frac{1}{n}\sum_{j=1}^{|\lambda_1(s_i)|}\left|A_{\mathrm{suff}_{j-1}(\lambda_1(s_i))} \cap h(\mathrm{suff}_{j-1}(\lambda_1(s_i)))^{-1} D_{\mathrm{last}_j(\lambda_1(s_i))}\right| \\
& \qquad \text{(since } x \in A_w \text{ implies } x \in D_{sw} \Leftrightarrow x \in h(w)^{-1}D_s \text{ for } w \in F_T, s \in S^{\pm}\text{)} \\
\le& \sum_{j=1}^{|\lambda_1(s_i)|} \underbrace{\frac{1}{n}\left|D_{\mathrm{last}_j(\lambda_1(s_i))}\right|}_{\le d^{\mathrm{H}}(h,h')} \\
\le& |\lambda_1(s_i)| d^{\mathrm{H}}(h, h') \ .
\end{aligned}
$$

Thus

$$
d^{\mathrm{H}}(\lambda_1^* h, \lambda_1^* h') = \sum_{i=1}^{d_1} d^{\mathrm{H}}((\lambda_1^* h)(s_i), (\lambda_1^* h')(s_i)) \le \left(\sum_{i=1}^{d_1} |\lambda_1(s_i)|\right) d^{\mathrm{H}}(h, h') \ .
$$

□

Lemma 6.8. *For $f \in \mathcal{H}_S(n)$ and $h \in \mathcal{H}_T(n)$,*

$$
d^{\mathrm{H}}(f, \lambda_1^* h) \le C_1 d^{\mathrm{H}}(\lambda_2^* f, h) + C_2 \frac{|\mathrm{Bad}_{E_1}(f)|}{n} \ ,
$$

where $C_1 = \sum_{i=1}^{d_1} |\lambda_1(s_i)|$ and $C_2 = \sum_{i=1}^{d_1} |Q_i|$.

Proof. By the triangle inequality, Corollary 6.6 and Lemma 6.7,

$$
\begin{aligned}
d^{\mathrm{H}}(f, \lambda_1^* h) &\le d^{\mathrm{H}}(f, \lambda_1^* \lambda_2^* f) + d^{\mathrm{H}}(\lambda_1^* \lambda_2^* f, \lambda_1^* h) \\
&\le \left(\sum_{i=1}^{d_1} |Q_i|\right) \frac{|\mathrm{Bad}_{E_1}(f)|}{n} + \left(\sum_{i=1}^{d_1} |\lambda_1(s_i)|\right) d^{\mathrm{H}}(\lambda_2^* f, h) \ .
\end{aligned}
$$

□

6.3.2. *The P-local metric.* For a probability distribution θ over a set Ω and a function $\varphi\colon \Omega \to \Omega'$, write $\varphi_* \theta$ for the distribution of $\varphi(x)$ when $x \sim \theta$.

Lemma 6.9. *Let θ and θ' be probability distributions over the finite sets Ω and Ω', respectively, and take a function $\varphi\colon \Omega \to \Omega'$. Then*

$$d_{\mathrm{TV}}(\varphi_*\theta, \varphi_*\theta') \le d_{\mathrm{TV}}(\theta, \theta')\ .$$

Proof. By the triangle inequality,

$$\begin{aligned} d_{\mathrm{TV}}(\varphi_*\theta, \varphi_*\theta') &= \frac{1}{2}\sum_{y\in\Omega'} |\varphi_*\theta(y) - \varphi_*\theta'(y)| = \frac{1}{2}\sum_{y\in\Omega'} \left| \sum_{\substack{x\in\Omega \\ f(x)=y}} \theta(x) - \sum_{\substack{x\in\Omega \\ f(x)=y}} \theta'(x) \right| \\ &\le \frac{1}{2}\sum_{y\in\Omega'} \sum_{\substack{x\in\Omega \\ f(x)=y}} |\theta(x) - \theta'(x)| = \frac{1}{2}\sum_{x\in\Omega} |\theta(x) - \theta'(x)| = d_{\mathrm{TV}}(\theta, \theta'). \end{aligned}$$

□

In the rest of the section we make frequent use of the notation $N_{G,P}$ (see Definition 1.12).

Corollary 6.10. *Let S be a finite alphabet and take $P, P^* \in \mathrm{FinSubsets}(F_S)$ such that $P^* \subseteq P$. Then*

$$d_{\mathrm{TV}}(N_{G,P^*}, N_{G',P^*}) \le d_{\mathrm{TV}}(N_{G,P}, N_{G',P})$$

for all $n \in \mathbb{N}$ and $G, G' \in \mathcal{G}_S(n)$.

Proof. The claim follows from Lemma 6.9 because $N_{G,P^*} = \varphi_* N_{G,P}$ and $N_{G',P^*} = \varphi_* N_{G',P}$ for $\varphi\colon \mathrm{Subsets}(P) \to \mathrm{Subsets}(P^*)$ given by $\varphi(M) = M \cap P^*$. □

Lemma 6.11. *Let $P_2 \in \mathrm{FinSubsets}(F_T)$. Then*

$$d_{\mathrm{TV}}\big(N_{\lambda_2^* f, P_2}, N_{\lambda_2^* f', P_2}\big) \le d_{\mathrm{TV}}\big(N_{f,\lambda_2(P_2)}, N_{f',\lambda_2(P_2)}\big)$$

for all $f, f' \in \mathcal{H}_S(n)$.

Proof. Define $\varphi\colon \mathrm{Subsets}(\lambda_2(P_2)) \to \mathrm{Subsets}(P_2)$ by

$$\varphi(V) = \{w \in P_2 \mid \lambda_2(w) \in V\} = \lambda_2^{-1}(V) \cap P_2 \quad \forall V \subset \lambda_2(P_2). \tag{6.5}$$

By Lemma 6.9, it suffices to show that $N_{\lambda_2^* f, P_2} = \varphi_* N_{f,\lambda_2(P_2)}$ and $N_{\lambda_2^* f', P_2} = \varphi_* N_{f',\lambda_2(P_2)}$. Thus, it is enough to prove that $\mathrm{Stab}_{\lambda_2^* h}(x) \cap P_2 = \varphi(\mathrm{Stab}_h(x) \cap \lambda_2(P_2))$ for all $h \in \mathcal{H}_S(n)$ and $x \in [n]$.

First, $\mathrm{Stab}_{\lambda_2^* h}(x) = \lambda_2^{-1}(\mathrm{Stab}_h(x))$ because $(\lambda_2^* h)(w)x = x \iff h(\lambda_2(w))x = x$ for $w \in F_T$. Thus,

$$\begin{aligned} \mathrm{Stab}_{\lambda_2^* h}(x) \cap P_2 &= \lambda_2^{-1}(\mathrm{Stab}_h(x)) \cap P_2 \\ &= \lambda_2^{-1}(\mathrm{Stab}_h(x) \cap \lambda_2(P_2)) \cap P_2 \\ &= \varphi(\mathrm{Stab}_h(x) \cap \lambda_2(P_2))\ . \end{aligned}$$

□

6.4. **Proof of Proposition 6.1.** Suppose that E_2 is $(P_2(\varepsilon), \delta_2(\varepsilon))$-statistically-distinguishable. We claim that E_1 is $(P_1(\varepsilon), \delta_1(\varepsilon))$-statistically-distinguishable, where

$$P_1(\varepsilon) = \lambda_2(P_2(\varepsilon)) \cup R_{E_1}$$

and

$$\delta_1(\varepsilon) = \min\left(\delta_2\left(\frac{\varepsilon}{2C_1}\right), \frac{\varepsilon}{2C_2}\right)\ ,$$

for C_1 and C_2 as in Lemma 6.8. First, note that

$$\begin{aligned}
\mathrm{TotalSize}(P_1(\varepsilon)) &= \sum_{w \in P_1(\varepsilon)} |w| \\
&\leq \mathrm{TotalSize}(R_{E_1}) + \sum_{w \in P_2(\varepsilon)} |\lambda_2(w)| \\
&\leq \mathrm{TotalSize}(R_{E_1}) + \sum_{w \in P_2(\varepsilon)} \underbrace{\max_{t \in T} |\lambda_2(t)|}_{=:C_3} |w| \\
&= \mathrm{TotalSize}(R_{E_1}) + C_3 \cdot \mathrm{TotalSize}(P_2(\varepsilon)) \ .
\end{aligned}$$

Hence (6.1) and (6.2) are satisfied. Take $\varepsilon > 0$, $f \in \mathcal{H}_S(n)$ and $f' \in \mathrm{HSol}_{E_1}(n)$ such that

$$d_{\mathrm{TV}}\left(N_{f,P_1(\varepsilon)}, N_{f',P_1(\varepsilon)}\right) < \delta_1(\varepsilon) \ .$$

It suffices to show that $f \in \mathrm{HSol}_{E_1}^{\leq\varepsilon}(n)$. Let h be an element of $\mathrm{HSol}_{E_2}(n)$ that minimizes $d^{\mathrm{H}}(\lambda_2^* f, h)$. Then $\lambda_1^* h \in \mathrm{HSol}_{E_1}(n)$ by Corollary 6.5, and thus it suffices to show that $d^{\mathrm{H}}(f, \lambda_1^* h) < \varepsilon$.

Now, $\lambda_2^* f' \in \mathrm{HSol}_{E_2}(n)$ by Corollary 6.5. Furthermore,

$$\begin{aligned}
d_{\mathrm{TV}}\left(N_{\lambda_2^* f, P_2(\varepsilon)}, N_{\lambda_2^* f', P_2(\varepsilon)}\right) &\leq d_{\mathrm{TV}}\left(N_{f,\lambda_2(P_2(\varepsilon))}, N_{f',\lambda_2(P_2(\varepsilon))}\right) && \text{by Lemma 6.11} \\
&\leq d_{\mathrm{TV}}\left(N_{f,P_1(\varepsilon)}, N_{f',P_1(\varepsilon)}\right) && \text{by Corollary 6.10} \\
&< \delta_1(\varepsilon) \leq \delta_2\left(\frac{\varepsilon}{2C_1}\right) ,
\end{aligned}$$

and thus $\lambda_2^* f \in \mathrm{HSol}_{E_2}^{<\frac{\varepsilon}{2C_1}}(n)$ since E_2 is $(P_2(\varepsilon), \delta_2(\varepsilon))$-statistically-distinguishable. In other words,

$$d^{\mathrm{H}}(\lambda_2^* f, h) < \frac{\varepsilon}{2C_1} \ . \tag{6.6}$$

Now,

$$\begin{aligned}
\frac{|\mathrm{Bad}_{E_1}(f)|}{n} &= d_{\mathrm{TV}}\left(N_{f,R_{E_1}}, N_{f',R_{E_1}}\right) && \text{by Remark 6.3} \\
&\leq d_{\mathrm{TV}}\left(N_{f,P_1(\varepsilon)}, N_{f',P_1(\varepsilon)}\right) && \text{by Corollary 6.10} \\
&< \delta_1(\varepsilon) \leq \frac{\varepsilon}{2C_2} \ . && (6.7)
\end{aligned}$$

Hence,

$$\begin{aligned}
d^{\mathrm{H}}(f, \lambda_1^* h) &\leq C_1 d^{\mathrm{H}}(\lambda_2^* f, h) + C_2 \frac{|\mathrm{Bad}_{E_1}(f)|}{n} && \text{by Lemma 6.8} \\
&< \varepsilon \ . && \text{by (6.6) and (6.7)}
\end{aligned}$$

7. Analysis and universality of the LSM Algorithm

Let $S = \{s_1, \ldots, s_d\}$ be an alphabet, and fix a system of relations E over $S^{\pm}$. The goal of this section is to prove Theorem 3, which follows immediately from the following two propositions.

Proposition 7.1 (Statistical distinguishability implies testability)**.** *If E is (P, δ)-statistically-distinguishable then $\varepsilon \mapsto \mathrm{LSM}^E_{k(\varepsilon), P(\varepsilon), \frac{\delta(\varepsilon)}{2}}$ is a family of testers for E, where $k(\varepsilon) = \left\lceil \frac{100 \cdot 2^{|P(\varepsilon)|}}{\delta(\varepsilon)^2} \right\rceil$.*

In particular, by the query-complexity analysis in Section 1.4, Proposition 7.1 implies that E is $O\left(\frac{\sum_{w \in P(\varepsilon)} |w| \cdot 2^{|P|}}{\delta(\varepsilon)^2}\right)$-testable.

Proposition 7.2 (Testability implies statistical distinguishability)**.** *Suppose that E is $q(\varepsilon)$-testable, $q \colon \mathbb{R}_{>0} \to \mathbb{N}$. Then, E is $\left(B_{2q(\varepsilon)} \cup E, q(\varepsilon)^{-c \cdot q(\varepsilon)}\right)$-statistically-distinguishable, where $c > 0$ is a universal constant.*

Here B_k denotes the ball of radius k in F_S, namely,

$$B_k = \{w \in F_S \mid |w| \le k\} \ .$$

7.1. **Proof of Proposition 7.1.** Assume that E is (P,δ)-statistically-distinguishable, $P\colon \mathbb{R}_{>0} \to \mathrm{FinSubsets}(F_S)$, $\delta\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. Fix $\varepsilon_0 > 0$ and denote $P_0 = P(\varepsilon_0)$ and $\delta_0 = \delta(\varepsilon_0)$. Let $k = \left\lceil \frac{100\cdot 2^{|P_0|}}{{\delta_0}^2} \right\rceil$. Fix $G \in \mathcal{G}_S(n)$ such that either $G \in \mathrm{GSol}_E(n)$ or $G \in \mathrm{GSol}_E^{\ge\varepsilon_0}(n)$. We wish to show that $\mathrm{LSM}^E_{k,P_0,\frac{\delta_0}{2}}$ returns a correct result in its run on G with probability at least 0.99.

If $G \in \mathrm{GSol}_E(n)$, set $A_0 = \mathrm{GSol}_E(n)$ and $A_1 = \mathrm{GSol}_E^{\ge\varepsilon_0}(n)$, and otherwise set $A_0 = \mathrm{GSol}_E^{\ge\varepsilon_0}(n)$ and $A_1 = \mathrm{GSol}_E(n)$. Then

$$d_{\mathrm{TV}}(N_{G,P_0}, N_{G',P_0}) \ge \delta_0 \quad \forall G' \in A_1 \tag{7.1}$$

because E is $(P(\varepsilon), \delta(\varepsilon))$-statistically-distinguishable.

Let $x_1, \dots x_k \in [n]$ denote the random variables sampled in Step 1 of a run of $\mathrm{LSM}^E_{k,P_0,\frac{\delta_0}{2}}$ (see Algorithm 2) on G. Denote $\overline{x} = (x_1,\dots,x_k) \in [n]^k$, and let $N^{\mathrm{Empirical}}_{\overline{x}}$ be the distribution, over $\mathrm{Subsets}(P_0)$, of $\mathrm{Stab}_G(x_i) \cap P_0$ where $i \sim \mathrm{U}([k])$. Recall that this run of $\mathrm{LSM}^E_{k,P_0,\frac{\delta_0}{2}}$ accepts G if and only if Condition (1.4) is satisfied, namely, if

$$\min\Big\{ d_{\mathrm{TV}}\Big(N^{\mathrm{Empirical}}_{\overline{x}}, N_{H,P_0}\Big) \mid H \in \mathrm{GSol}_E(n) \Big\} \le \frac{\delta_0}{2} \ . \tag{7.2}$$

In particular, the run returns a correct result on G whenever

$$d_{\mathrm{TV}}\Big(N^{\mathrm{Empirical}}_{\overline{x}}, N_{G,P_0}\Big) < \delta_0/2 \ . \tag{7.3}$$

Indeed, if $G \in \mathrm{GSol}_E(n)$, then (7.2) holds by (7.3), and so the run is accepting. If $G \in \mathrm{GSol}_E^{\ge\varepsilon_0}(n)$ then (7.2) does not hold by (7.1), (7.3) and the triangle inequality, and so the run is rejecting. Thus, it suffices to show that

$$\Pr_{\overline{x}\sim \mathrm{U}\left([n]^k\right)} \Big(d_{\mathrm{TV}}\Big(N^{\mathrm{Empirical}}_{\overline{x}}, N_{G,P_0}\Big) < \delta_0/2 \Big) \ge 0.99 \ .$$

For a fixed $Y \in \mathrm{Subsets}(P)$, the random variable $k \cdot N^{\mathrm{Empirical}}_{\overline{x}}(Y)\colon [n]^k \to \mathbb{Z}_{\ge 0}$ (that is, "k times $N^{\mathrm{Empirical}}_{\overline{x}}(Y)$") is distributed $\mathrm{Binomial}(k, N_{G,P_0}(Y))$ when $x \sim \mathrm{U}\Big([n]^k\Big)$. In particular, $\mathbb{E}_{\overline{x}\sim \mathrm{U}\left([n]^k\right)}\Big[k \cdot N^{\mathrm{Empirical}}_{\overline{x}}(Y)\Big] = kN_{G,P_0}(Y)$. Hence,

$$\begin{aligned}
\mathbb{E}_{\overline{x}\sim \mathrm{U}\left([n]^k\right)}\Big(\|N^{\mathrm{Empirical}}_{\overline{x}} - N_{G,P_\varepsilon}\|_2^2\Big) &= \sum_{Y \in \mathrm{Subsets}(P_0)} \underbrace{\mathbb{E}_{\overline{x}\sim \mathrm{U}\left([n]^k\right)}\Big(\Big(N^{\mathrm{Empirical}}_{\overline{x}}(Y) - N_{G,P_0}(Y)\Big)^2\Big)}_{=\frac{1}{k^2}\mathrm{Var}_{\overline{x}\sim \mathrm{U}\left([n]^k\right)}\left(kN^{\mathrm{Empirical}}_{\overline{x}}(Y)\right)} \\
&= \sum_{Y \in \mathrm{Subsets}(P_0)} \frac{N_{G,P_0}(Y)(1-N_{G,P_0}(Y))}{k} \\
&\le \frac{1}{k} \sum_{Y \in \mathrm{Subsets}(P_0)} N_{G,P_0}(Y) \\
&= \frac{1}{k} \ .
\end{aligned}$$

By the Cauchy–Schwartz inequality,

$$\begin{aligned}
d_{\mathrm{TV}}\Big(N^{\mathrm{Empirical}}_{\overline{x}}, N_{G,P_0}\Big) &= \frac{1}{2}\|N^{\mathrm{Empirical}}_{\overline{x}} - N_{G,P_0}\|_1 \le \frac{1}{2}|\mathrm{Subsets}(P_0)|^{\frac{1}{2}} \cdot \|N^{\mathrm{Empirical}}_{\overline{x}} - N_{G,P_0}\|_2 \\
&= \frac{1}{2} \cdot 2^{\frac{|P_0|}{2}} \cdot \|N^{\mathrm{Empirical}}_{\overline{x}} - N_{G,P_0}\|_2 \ .
\end{aligned}$$

Thus,

$$\Pr_{\overline{x}\sim \mathrm{U}\left([n]^k\right)}\left(d_{\mathrm{TV}}\left(N_{\overline{x}}^{\mathrm{Empirical}}, N_{G,P_0}\right) \geq \delta_0/2\right) \leq \Pr_{\overline{x}\sim \mathrm{U}\left([n]^k\right)}\left(\|N_{\overline{x}}^{\mathrm{Empirical}} - N_{G,P_0}\|_2^2 \geq 2^{-|P_0|}\delta_0^2\right)$$
$$\leq \frac{2^{|P_0|}}{k\delta_0^2} \qquad \text{(by Markov's inequality)}$$
$$\leq 0.01 \ .$$

7.2. **Testability implies statistical distinguishability.** We prove Proposition 7.2 in Section 7.2.5 after making the necessary preparations.

7.2.1. *A minimax principle.* The following lemma, a variant of *Yao's minimax inequality* [41], enables us to reduce Proposition 7.2 to a statement about deterministic, rather than randomized, algorithms.

Lemma 7.3. *Fix $\varepsilon > 0$ and $q \in \mathbb{N}$. The following conditions are equivalent:*

(1) There exists an (ε, q)-tester for E.
(2) For every distribution D on $\bigcup_{n\in\mathbb{N}}\left(\mathrm{GSol}_E(n) \cup \mathrm{GSol}_E^{\geq\varepsilon}(n)\right)$, there exists a deterministic algorithm $\mathcal{M}$, limited to making at most q queries of G, such that

$$\Pr_{G\sim D}[\mathcal{M} \textit{ solves the input } G \textit{ correctly}] \geq 0.99 \ .$$

Proof. Let $\mathcal{A}$ denote the set of all deterministic algorithms, with input in $\bigcup_{n\in\mathbb{N}}\mathcal{G}_S(n)$ and boolean output, limited to making at most q queries in a run. Let $\mathcal{B} = \bigcup_{n\in\mathbb{N}}\left(\mathrm{GSol}_E(n) \cup \mathrm{GSol}_E^{\geq\varepsilon}(n)\right)$. Let $\mathcal{D}_{\mathcal{A}}$ (resp. $\mathcal{D}_{\mathcal{B}}$) denote the set of all probability distributions over $\mathcal{A}$ (resp. $\mathcal{B}$). For $\mathcal{M} \in \mathcal{A}$, $G \in \mathcal{B}$, let

$$c_{\mathcal{M},G} = \begin{cases} 1 & \text{if } \mathcal{M} \text{ solves the input } G \text{ correctly} \\ 0 & \text{otherwise} \end{cases} .$$

A randomized algorithm can be viewed as a distribution over $\mathcal{A}$. Hence, the first condition in the statement of the proposition is equivalent to

$$\max_{D\in\mathcal{D}_{\mathcal{A}}}\min_{G\in\mathcal{B}}\Pr_{\mathcal{M}\sim D}[c_{\mathcal{M},G} = 1] \geq 0.99 \ , \tag{7.4}$$

while the second condition can be written as

$$\min_{D\in\mathcal{D}_B}\max_{\mathcal{M}\in\mathcal{A}}\Pr_{G\sim D}[c_{\mathcal{M},G} = 1] \geq 0.99 \ . \tag{7.5}$$

Observe that the left-hand side of (7.4) is equal to

$$\max_{D\in\mathcal{D}_{\mathcal{A}}}\min_{D'\in\mathcal{D}_{\mathcal{B}}}\Pr_{\substack{G\sim D\\ \mathcal{M}\sim D'}}[c_{\mathcal{M},G} = 1] \ . \tag{7.6}$$

Indeed, $\Pr_{\substack{G\sim D\\ \mathcal{M}\sim D'}}[c_{\mathcal{M},G} = 1]$ is linear in D', and hence the minimum in (7.4) is obtained at a vertex of the simplex $\mathcal{D}_B$. Similarly, the left-hand side of (7.5) is equal to

$$\min_{D'\in\mathcal{D}_B}\max_{D\in\mathcal{D}_{\mathcal{A}}}\Pr_{\substack{G\sim D\\ \mathcal{M}\sim D'}}[c_{\mathcal{M},G} = 1] \ . \tag{7.7}$$

Finally, (7.6) and (7.7) are equal by the von Neumann Minimax Theorem [18, Thm. 1]. ☐

7.2.2. *The* d_{TV} *metric.* We recall the following standard facts about the d_{TV} metric.

Fact 7.4 ([28, Prop. 4.2])**.** *Let θ and θ' be two distributions over the same finite base set Ω. Then*

$$d_{\mathrm{TV}}(\theta,\theta') = \max_{A\subset\Omega} |\theta(A) - \theta'(A)| \ .$$

Fact 7.5 (Coupling Lemma [28, Prop. 4.7])**.** *For any two distributions θ and θ' over the same finite base set Ω, there exists a distribution D over $\Omega\times\Omega$ such that:*

(1) The marginal distributions of D are equal, respectively, to θ and θ'.
(2)

$$d_{\mathrm{TV}}(\theta,\theta') = \Pr_{(x,y)\sim D}[x \neq y] \ .$$

Fact 7.6 ([28, Ex. 4.4])**.** *Let $\theta = \prod_{i=1}^t \theta_i$ and $\theta' = \prod_{i=1}^t \theta'_i$ be product measures. Then,*

$$d_{\mathrm{TV}}(\theta,\theta') \leq \sum_{i=1}^t d_{\mathrm{TV}}(\theta_i,\theta'_i)$$

7.2.3. *Partial S-graphs and runs of algorithms.*

Definition 7.7 (Partial S-graph)**.** Let S be a finite set. A *partial S-graph* is a directed graph H whose edges are labelled by S, such that for every $s\in S$, every vertex of H has at most one outgoing edge labelled s and at most one incoming edge labelled s.

We denote the vertex set of H by $V(H)$, and $E(H)$ will be the set of (S-labelled) edges in H. For vertices $x_1,x_2\in V(H)$ and $s\in S$, we say that $x_1\xrightarrow{s^{-1}}x_2$ is an edge of H if the same is true for $x_2\xrightarrow{s}x_1$. When computing the cardinality $|E(H)|$ of $E(H)$, the edges $x_2\xrightarrow{s}x_1$ and $x_1\xrightarrow{s^{-1}}x_2$ count as one edge.

We denote the set of all partial S-graphs H such that $V(H)\subset[n]$ by $\mathcal{PG}_S(n)$. Given $H\in\mathcal{PG}_S(n)$, we write $r(H) = |V(H)| - k$, where k is the number of connected components of H. For example, $r(H) = |V(H)| - 1$ if H is connected, and $r(H) = 0$ if H has no edges.

Definition 7.8 (Paths in a partial S-graph)**.** Let H be a partial S-graph, $x\in V(H)$ and $w\in F_S$. Suppose that the reduced form of w is given by $w_k\cdots w_1$ ($w_i\in S^{\pm}$ for each $1\leq i\leq k$). We write $w_Hx = y$, and say that w is an *H-path* for x, if there exist $x_0,\dots,x_k\in V(H)$ such that $x_0 = x$, $x_k = y$, and H contains a w_i-labelled edge $x_{i-1}\longrightarrow x_i$ for each $1\leq i\leq k$ (in this case $x_0,\dots,x_k$ are uniquely defined). If, in addition, the vertices $x_0,\dots,x_k$ are distinct, except for the possibility that $x_0 = x_k$, we say that w is a *simple H-path* for x.

Let

$$\mathrm{Paths}_H(x) = \{w\in F_S \mid w \text{ is an } H\text{-path for } x\} \ ,$$
$$\text{Simple-Paths}_H(x) = \{w\in F_S \mid w \text{ is a simple } H\text{-path for } x\} \ ,$$
$$\mathrm{Bipaths}_H(x) = \left\{w'^{-1}\cdot w \mid w,w'\in \text{Simple-Paths}_H(x)\right\}$$

and

$$\mathrm{PStab}_H(x) = \{w\in \mathrm{Paths}_H(x)\cap \mathrm{Bipaths}_H(x) \mid w_Hx = x\} \ .$$

The definition of $\mathrm{Bipaths}_H(x)$ ensures that the set $\mathrm{PStab}_H(x)$ contains all the information about simple paths emerging from x and ending at the same vertex. Indeed, for $w,w'\in\text{Simple-Paths}_H(x)$, we have $w_Hx = w'_Hx$ if and only if $w'^{-1}w\in\mathrm{PStab}_H(x)$.

Definition 7.9 (Inclusion of a partial S-graph)**.** Let $G\in\mathcal{G}_S(n)$ and $H\in\mathcal{PG}_S(n)$. Write $H\subset G$ if G contains all the labelled directed edges of H. Equivalently, $H\subset G$ if $w_Gx = w_Hx$ for all $x\in V(H)$ and $w\in\text{Simple-Paths}_H(x)$.

Remark 7.10. Crucially, in contrast to other common definitions of inclusion of graphs, here we care about vertex labels. For example, if $H\subset G$ and H has the edge $3\xrightarrow{s}8$, then G must also have the edge $3\xrightarrow{s}8$.

Let $\mathcal{M}$ be a deterministic algorithm that makes exactly q queries on its input $G \in \mathcal{G}_S(n)$. Let $H_{\mathcal{M},G} \in \mathcal{PG}_S(n)$ be the partial S-graph whose edge set is the set of edges of G queried by $\mathcal{M}$ in its run on G, with vertex set consisting of the vertices incident to these edges. More formally, in graph-theoretic terms each query of $\mathcal{M}$ is of the form: "what is the label of the vertex $s_G x$?" for given $s \in S^{\pm}$ and $x \in [n]$. We denote this query by (x, s). Denote the sequence of queries that $\mathcal{M}$ makes in its run on G by $\left(x^{\mathcal{M},G,1}, s^{\mathcal{M},G,1}\right), \ldots, \left(x^{\mathcal{M},G,q}, s^{\mathcal{M},G,q}\right)$. Then the vertex set of $H_{\mathcal{M},G}$ is $\left\{x^{\mathcal{M},G,1}, \ldots, x^{\mathcal{M},G,q}\right\} \cup \left\{s_G^{\mathcal{M},G,1} x^{\mathcal{M},G,1}, \ldots, s_G^{\mathcal{M},G,q} x^{\mathcal{M},G,q}\right\}$, and its edges are $x^{\mathcal{M},G,i} \overset{s^{\mathcal{M},G,i}}{\longrightarrow} s_G^{\mathcal{M},G,i} x^{\mathcal{M},G,i}$, $1 \le i \le q$.

Clearly $H_{\mathcal{M},G} \subset G$. The key observation is that if an S-graph $G' \in \mathcal{G}_S(n)$ contains $H_{\mathcal{M},G}$ then $\mathcal{M}$ has an identical run on G and on G'. Indeed, since $\mathcal{M}$ is deterministic, it always starts with the same query. Since both G and G' contain the edge $x^{\mathcal{M},G,1} \overset{s^{\mathcal{M},G,1}}{\longrightarrow} s_G^{\mathcal{M},G,1} x^{\mathcal{M},G,1}$, the answer to this query is the same in both runs. Consequently, the second query of $\mathcal{M}$ is also the same, and so on.

We denote

$$\mathcal{R}_{\mathcal{M}}(n) = \{H_{\mathcal{M},G} \mid G \in \mathcal{G}_S(n)\}$$

and

$$\mathcal{R}_{\mathcal{M},r}(n) = \{H \in \mathcal{R}_{\mathcal{M}}(n) \mid r(H) = r\} \tag{7.8}$$

for $r \ge 0$. For $G \in \mathcal{G}_S(n)$, the partial S-graph $H = H_{\mathcal{M},G}$ is the unique $H \in \mathcal{R}_{\mathcal{M}}(n)$ such that $H \subset G$. If $H_{\mathcal{M},G} = H_{\mathcal{M},G'}$ then $\mathcal{M}$ has identical runs on G and G'. In particular, these runs terminate with the same result. We proved:

Lemma 7.11. *Let $\mathcal{M}$ be a deterministic algorithm and let $G, G' \in \mathcal{G}_S(n)$. Suppose that $H_{\mathcal{M},G} = H_{\mathcal{M},G'}$. Then*

$$\mathcal{M} \text{ accepts } G \iff \mathcal{M} \text{ accepts } G' \ .$$

7.2.4. *Vertex relabelling of an S-graph.*

Definition 7.12 (S-graph relabelling)**.** Let $G \in \mathcal{G}_S(n)$ and let π be a permutation over $[n]$. The *vertex-relabelled graph* $G\pi \in \mathcal{G}_S(n)$ is the graph with vertex set $[n]$ and edge set $\left\{\pi^{-1}x \overset{s}{\longrightarrow} \pi^{-1}(s_G x) \mid x \in [n]\right\}$.

Remark 7.13. For each $w \in F_S$, the following diagram commutes:

$$\begin{array}{ccc} [n] & \xrightarrow{w_{G\pi}} & [n] \\ \big\downarrow \pi & & \big\downarrow \pi \\ [n] & \xrightarrow{w_G} & [n] \end{array} \ .$$

In other words, $w_{(G\pi)}x = \pi^{-1}(w_G \pi x)$ for $x \in V(G\pi)$.

We think of π as a relabelling function, in the sense that $G\pi$ is obtained from G by changing the name of each vertex $\pi(x) \in V(G)$ to x.

Remark 7.14. Let $G \in \mathcal{G}_S(n)$ and $\pi \in \mathrm{Sym}(n)$. It is straightforward to verify that $G \in \mathrm{GSol}_E(n)$ implies $G\pi \in \mathrm{GSol}_E(n)$, and that $G \in \mathrm{GSol}_E^{\ge\varepsilon}(n)$ implies $G\pi \in \mathrm{GSol}_E^{\ge\varepsilon}(n)$ ($\varepsilon > 0$).

Remark 7.15. For each $\pi \in \mathrm{Sym}(n)$, the map $G \mapsto G\pi \colon \mathcal{G}_S(n) \to \mathcal{G}_S(n)$ defines a right action of $\mathrm{Sym}(n)$ on $\mathcal{G}_S(n)$. For $G \in \mathcal{G}_S(n)$, the set $\{G\pi \mid \pi \in \mathrm{Sym}(n)\}$ is an orbit of this action. The distribution of $G\pi$ when $\pi \sim \mathrm{U}(\mathrm{Sym}(n))$ is the uniform distribution on this orbit.

7.2.5. *Proving Proposition 7.2* . We shall prove the following technical lemma.

Lemma 7.16. *Fix two S-graphs $G_0, G_1 \in \mathcal{G}_S(n)$. Fix $q \in \mathbb{N}$ with $1 \le q \le n^{\frac{1}{4}}$ and let $\mathcal{M}$ be a deterministic algorithm, limited to making at most q queries. Let θ_i, $i \in \{0, 1\}$, be the distribution (over $\mathcal{R}_{\mathcal{M}}(n)$) of the partial S-graph $H_{\mathcal{M},G_i\pi}$, where $\pi \sim \mathrm{U}(\mathrm{Sym}(n))$. Then*

$$d_{\mathrm{TV}}(\theta_1, \theta_2) \le (2q)^{q+2}\left(d_{\mathrm{TV}}\left(N_{G_0,B_{2q}}, N_{G_1,B_{2q}}\right) + cn^{-\frac{1}{4}}\right) \ ,$$

where $c > 0$ is a universal constant.

Before proving Lemma 7.16, we show that it implies Proposition 7.2.

Proof of Proposition 7.2. By the hypothesis, E is $q(\varepsilon)$-testable. Fix $\varepsilon > 0$ and write $q_\varepsilon = q(\varepsilon)$. Let $P = B_{2q_\varepsilon} \cup E$. Fix $G_0 \in \mathrm{GSol}_E(n)$ and $G_1 \in \mathrm{GSol}_E^{\geq\varepsilon}(n)$, and denote $\delta = d_{\mathrm{TV}}(N_{G_0,P}, N_{G_1,P})$. Our goal is to show that

$$\delta \geq q_\varepsilon^{-O(q_\varepsilon)} ,$$

where the implied constant is universal. We first deal with the case

$$n < \max\left\{\frac{(2q_\varepsilon)^{4q_\varepsilon+8}\cdot c^4}{0.057}, q_\varepsilon^4\right\} .$$

where c is the constant from Lemma 7.16. We note that, since $G_0 \in \mathrm{GSol}_E(n)$, we have $\mathrm{Stab}_{G_0,E}(x) = E$ for every $x \in [n]$. However, the same is not true for $\mathrm{Stab}_{G_1,E}(x)$ since $G_1 \notin \mathrm{GSol}_E(n)$. Since $E \subseteq P$, Corollary 6.10 yields (for the leftmost inequality)

$$\delta \geq d_{\mathrm{TV}}(N_{G_0,E}, N_{G_1,E}) \geq \Pr_{x\sim \mathrm{U}([n])}[\mathrm{Stab}_{G_0,E}(x) = E] - \Pr_{x\sim \mathrm{U}([n])}[\mathrm{Stab}_{G_1,E}(x) = E] \geq \frac{1}{n} \geq q_\varepsilon^{-O(q_\varepsilon)} .$$

We proceed, assuming that

$$n \geq \max\left\{\frac{(2q_\varepsilon)^{4q_\varepsilon+8}\cdot c^4}{0.057}, q_\varepsilon^4\right\} . \tag{7.9}$$

Let D be the distribution of the random S-graph $G_i\pi$, where i is the result of a fair coin flip, and π is independently sampled from $\mathrm{U}(\mathrm{Sym}(n))$ uniformly. Applying Lemma 7.3 to D yields a deterministic algorithm $\mathcal{M}$, limited to making q_ε queries, such that

$$\Pr_{G\sim D}[\mathcal{M} \text{ solves the input } G \text{ correctly}] \geq 0.99 . \tag{7.10}$$

Let θ_i, $i \in \{0,1\}$, be the distribution of the partial S-graph $H_{\mathcal{M},G_i\pi}$, where $\pi \sim \mathrm{U}(\mathrm{Sym}(n))$. We claim that

$$d_{\mathrm{TV}}(\theta_0, \theta_1) \geq 0.49 . \tag{7.11}$$

Before proving (7.11), we show that it implies the proposition. Note that every $H \in \mathcal{R}_{\mathcal{M}}(n)$ has $|E(H)| \leq q_\varepsilon \leq n^{\frac{1}{4}}$, where the rightmost inequality is due to (7.9). Hence,

$$\begin{aligned}
d_{\mathrm{TV}}(\theta_0,\theta_1) &\leq (2q_\varepsilon)^{q_\varepsilon+2}\left(d_{\mathrm{TV}}\left(N_{G_0,B_{2q}}, N_{G_1,B_{2q}}\right) + cn^{-\frac{1}{4}}\right) && \text{by Lemma 7.16}\\
&\leq (2q_\varepsilon)^{q_\varepsilon+2}\left(\delta + cn^{-\frac{1}{4}}\right) && \text{by Corollary 6.10}\\
&\leq (2q_\varepsilon)^{q_\varepsilon+2}\delta + 0.057^{1/4} . && \text{by (7.9)}
\end{aligned}$$

Together with (7.11), this yields

$$\delta \geq (2q_\varepsilon)^{-q_\varepsilon-2}\left(0.49 - \underbrace{0.057^{1/4}}_{<0.49}\right) \geq q_\varepsilon^{-O(q_\varepsilon)} ,$$

and the proposition follows.

We turn to proving (7.11). Fact 7.5 yields a distribution Q over $\mathcal{R}_{\mathcal{M}}(n)\times\mathcal{R}_{\mathcal{M}}(n)$ with respective marginal distributions θ_0 and θ_1, such that

$$\Pr_{(H_0,H_1)\sim Q}[H_0 \neq H_1] = d_{\mathrm{TV}}(\theta_0,\theta_1) .$$

Let $\overline{Q}$ be a distribution over $(\mathrm{Sym}(n))^2$ which generates a pair of permutations (π_0,π_1) by first sampling $(H_0,H_1) \sim Q$, and then independently sampling $\pi_0 \sim \mathrm{U}(\{\pi \in \mathrm{Sym}(n) \mid H_{\mathcal{M},G_0\pi} = H_0\})$ and $\pi_1 \sim$

$\mathrm{U}(\{\pi \sim \mathrm{Sym}(n) \mid H_{\mathcal{M},G_1\pi} = H_1\})$. Observe that both marginal distributions of $\overline{Q}$ are equal to the uniform distribution on $\mathrm{Sym}(n)$. Indeed, for $i \in \{0,1\}$ and $\alpha \in \mathrm{Sym}(n)$,

$$\begin{aligned}
\Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\pi_i = \alpha] &= \Pr_{(H_0,H_1)\sim Q}[H_i = H_{\mathcal{M},G_i\alpha}] \cdot \frac{1}{|\{\pi \in \mathrm{Sym}(n) \mid H_{\mathcal{M},G_i\pi} = H_{\mathcal{M},G_i\alpha}\}|} \\
&= \Pr_{H_i\sim\theta_i}[H_i = H_{\mathcal{M},G_i\alpha}] \cdot \frac{1}{|\{\pi \in \mathrm{Sym}(n) \mid H_{\mathcal{M},G_i\pi} = H_{\mathcal{M},G_i\alpha}\}|} \\
&= \frac{|\{\pi \in \mathrm{Sym}(n) \mid H_{\mathcal{M},G_i\pi} = H_{\mathcal{M},G_i\alpha}\}|}{|\mathrm{Sym}(n)|} \cdot \frac{1}{|\{\pi \in \mathrm{Sym}(n) \mid H_{\mathcal{M},G_i\pi} = H_{\mathcal{M},G_i\alpha}\}|} \\
&= \frac{1}{|\mathrm{Sym}(n)|} \ .
\end{aligned}$$

Furthermore,

$$\Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[H_{\mathcal{M},G_0\pi_0} \neq H_{\mathcal{M},G_1\pi_1}] = \Pr_{(H_0,H_1)\sim Q}[H_0 \neq H_1] = d_{\mathrm{TV}}(\theta_0,\theta_1) \ .$$

Then

$$\begin{aligned}
0.99 &\leq \Pr_{G\sim D}[\mathcal{M} \text{ correctly solves the input } G] && \text{by (7.10)} \\
&= \frac{1}{2}\left(\Pr_{\pi_0\sim\mathrm{U}(\mathrm{Sym}(n))}[\mathcal{M} \text{ correctly solves } G_0\pi_0] + \Pr_{\pi_1\sim\mathrm{U}(\mathrm{Sym}(n))}[\mathcal{M} \text{ correctly solves } G_1\pi_1]\right) \\
&= \frac{1}{2}\left(\Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ correctly solves } G_0\pi_0] + \Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ correctly solves } G_1\pi_1]\right) \\
&= \frac{1}{2}\Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ correctly solves exactly one of } G_0\pi_0 \text{ and } G_1\pi_1] \\
&\quad + \Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ correctly solves both } G_0\pi_0 \text{ and } G_1\pi_1] \\
&\leq \frac{1}{2} + \Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ correctly solves both } G_0\pi_0 \text{ and } G_1\pi_1] \\
&= \frac{1}{2} + \Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ accepts } G_0\pi_0 \text{ and rejects } G_1\pi_1] && \text{by Remark 7.14} \\
&\leq \frac{1}{2} + \Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[\mathcal{M} \text{ returns different results on } G_0\pi_0 \text{ and } G_1\pi_1] \\
&\leq \frac{1}{2} + \Pr_{(\pi_0,\pi_1)\sim\overline{Q}}[H_{\mathcal{M},G_0\pi_0} \neq H_{\mathcal{M},G_1\pi_1}] && \text{by Lemma 7.11} \\
&= \frac{1}{2} + d_{\mathrm{TV}}(\theta_0,\theta_1) \ ,
\end{aligned}$$

and (7.11) follows. □

The rest of this section is devoted to proving Lemma 7.16. The following lemma bounds the cardinality of the set $\mathcal{R}_{\mathcal{M},r}(n)$ (defined in (7.8)).

Lemma 7.17. *Let $\mathcal{M}$ be a deterministic algorithm, limited to making at most q queries per run. Let $n, r \in \mathbb{N}$. Then*

$$|\mathcal{R}_{\mathcal{M},r}(n)| \leq \binom{q}{r} n^r (2q)^{q-r} \ .$$

Proof. For $G \in \mathcal{G}_S(n)$ and $0 \leq i \leq q$, let H_i^G be the partial S-graph such that the edges of H_i^G are the edges of G queried by $\mathcal{M}$ in the first i queries, and the vertices of H_i^G are those that touch these edges. In particular, H_0^G is the empty partial S-graph, and $H_q^G = H_{\mathcal{M},G}$. Note that the query $\left(x^{\mathcal{M},G,i}, s^{\mathcal{M},G,i}\right)$ is determined by H_{i-1}^G. The answer to this query, for which there are at most n possibilities, depends on G.

Clearly, $r\big(H_0^G\big) = 0$ and $r\big(H_q^G\big) = r(H_{\mathcal{M},G})$. Furthermore, in each step, $r(H_i) = r(H_{i-1}) + 1$ or $r(H_i) = r(H_{i-1})$. Respectively, we say that the i-th step is *increasing* or *nonincreasing*. If the i-th step is nonincreasing then the vertex given as an answer to the i-th query is in $\big\{x^{\mathcal{M},G,j} \mid 1 \le j \le i\big\} \cup \big\{s^{\mathcal{M},G,j} x^{\mathcal{M},G,j} \mid 1 \le j \le i-1\big\}$. In particular, there are at most $2q$ possible answers to the query in a nonincreasing step. Write $\mathcal{R}_{\mathcal{M},r',i}(n) = \big\{H_i^G \mid G \in \mathcal{G}_S(n) \text{ and } r\big(H_i^G\big) = r'\big\}$. By the above,

$$|\mathcal{R}_{\mathcal{M},r',i}(n)| \le n|\mathcal{R}_{\mathcal{M},r'-1,i-1}(n)| + 2q|\mathcal{R}_{\mathcal{M},r',i-1}(n)| \quad \forall 1 \le i \le q \;\; \forall 0 \le r' \le r \ ,$$

and thus

$$|\mathcal{R}_{\mathcal{M},r}(n)| = |\mathcal{R}_{\mathcal{M},r,q}(n)| \le \binom{q}{r} n^r (2q)^{q-r} \ .$$

□

The next lemma provides a condition equivalent to $H \subset G$ when H is a connected partial S-graph.

Lemma 7.18 (An equivalent condition for inclusion of a partial S-graph)**.** *Fix $G \in \mathcal{G}_S(n)$ and $H \in \mathcal{PG}_S(n)$, where H is connected and nonempty. Treating H as undirected, fix an undirected spanning tree T for it. Fix some $y_0 \in V(H)$. For every $y \in V(H)$, fix a word $w^y \in \text{Simple-Paths}_H(y_0)$ such that $w^y_H y_0 = y$, and the simple H-path from y_0 to y, induced by w^y, proceeds along edges of T. Then, $H \subset G$ if and only if*

$$\text{Stab}_G(y_0) \cap \text{Bipaths}_H(y_0) = \text{PStab}_H(y_0) \tag{7.12}$$

and

$$w^y_G y_0 = y \quad \forall y \in V(H) \ . \tag{7.13}$$

Proof. Suppose that $H \subset G$. Then $w_G y_0 = w_H y_0$ for all $w \in \text{Simple-Paths}_H(y_0)$, and (7.13) follows. The $\supset$ inclusion in (7.12) is clear. To prove the $\subset$ inclusion, let $w \in \text{Stab}_G(y_0) \cap \text{Bipaths}_H(y_0)$ and write $w = v'^{-1}v$ where $v, v' \in \text{Simple-Paths}_H(y_0)$. Then $y_0 = w_G y_0 = v'^{-1}_G v_G y_0$, and thus $v'_G y_0 = v_G y_0$. Hence $v'_H y_0 = v_H y_0$, and therefore $w = v'^{-1}v \in \text{PStab}_H(y_0)$.

Conversely, suppose that (7.12) and (7.13) hold, and take an edge $y_1 \xrightarrow{s} y_2$ of H. Without loss of generality, assume that y_2 is not an internal vertex in the path, in T, from y_0 to y_1. Then, $sw^y_1 \in \text{Simple-Paths}_H(y_0)$, and so $(w^{y_2})^{-1} sw^{y_1} \in \text{Bipaths}_H(y_0)$. Note that $(sw^{y_1})_H y_0 = y_2 = w^{y_2}_H y_0$. Thus, $(w^{y_2})^{-1} sw^{y_1} \in \text{PStab}_H(y_0)$. By (7.12), $(sw^{y_1})_G y_0 = w^{y_2}_G y_0$. Consequently, (7.13) implies that $s_G y_1 = y_2$, so G contains the edge $y_1 \xrightarrow{s} y_2$. □

Next, for fixed $G \in \mathcal{G}_S(n)$ and connected $H \in \mathcal{PG}_S(n)$, we study

$$\Pr_{\pi \sim \text{U}(\text{Sym}(n))} (H \subset G\pi) \ . \tag{7.14}$$

We use Lemma 7.18. Fix some $y_0 \in V(H)$ and a set of words $\{w^y\}_{y \in V(H)}$ as in the lemma. The probability that

$$\text{Stab}_{G\pi}(y_0) \cap \text{Bipaths}_H(y_0) = \text{PStab}_H(y_0) \tag{7.15}$$

is $N_{G,\text{Bipaths}_H(y_0)}(\text{PStab}_H(y_0))$, since $\pi(y_0)$ is distributed uniformly on $[n]$. Conditioned on the event (7.15), the probability that

$$w^y_{G\pi} y_0 = y \quad \forall y \in V(H)$$

is approximately $n^{-(V(H)-1)}$. Lemma 7.19 below provides a more precise estimate of (7.14). Furthermore, the lemma shows that (7.14) does not change by much if we condition on the event $\pi \mid_L = f$ for a small subset L of $[n]$ and an injective function $f\colon L \to [n]$.

Lemma 7.19. *Fix $G \in \mathcal{G}_S(n)$, $H \in \mathcal{PG}_S(n)$, and suppose that H is connected and has at least one edge. Let y_0 be an arbitrary vertex of H. Then*

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G\pi) = \frac{(n-|V(H)|)!}{(n-1)!} N_{G,\mathrm{Bipaths}_H(y_0)}(\mathrm{PStab}_H(y_0)) \ . \tag{7.16}$$

Moreover, let $L \subset [n] \setminus V(H)$ and let $f\colon L \to [n]$ be an injective function. Then

$$\left|\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G\pi \mid \pi\mid_L = f) - p\right| \le \frac{(n-|L|-|V(H)|)!}{(n-|L|-1)!}\cdot\frac{|L|\cdot|V(H)|}{n} \ , \tag{7.17}$$

where $\pi_{|L}$ denotes the restriction of π to the set L and

$$p = \frac{(n-|V(H)|-|L|)!}{(n-1-|L|)!}\cdot N_{G,\mathrm{Bipaths}_H(y_0)}(\mathrm{PStab}_H(y_0)) \ .$$

Remark 7.20. In the setting of Lemma 7.19, when $L \le O(n^{1/4})$ and $|V(H)| \le O(n^{1/4})$, (7.17) yields

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G\pi \mid \pi\mid_L = f) = n^{-(V(H)-1)}\cdot\Big(N_{G,\mathrm{Bipaths}_H(y_0)}(\mathrm{PStab}_H(y_0)) \pm O\Big(n^{-\frac{1}{2}}\Big)\Big).$$

Proof of Lemma 7.19. Since (7.16) is a special case of (7.17), we only prove the latter. Let $D = \{\pi \in \mathrm{Sym}(n) \mid \pi\mid_L = f\}$, and note that the distribution of $\pi \sim \mathrm{U}(\mathrm{Sym}(n))$, conditioned on the event $\pi\mid_L = f$, is the uniform distribution on D.

Fix $y_0 \in V(H)$ and write $V(H) = \{y_0, y_1, \dots, y_{|V(H)|-1}\}$. For every $0 \le i \le |V(H)|-1$, fix a word $w^{y_i} \in \text{Simple-Paths}_H(y_0)$ such that $w^{y_i}_H y_0 = y_i$. Let

$$A = \{\pi \in \mathrm{Sym}(n) \mid \mathrm{Stab}_{G\pi}(y_0) \cap \mathrm{Bipaths}_H(y_0) = \mathrm{PStab}_H(y_0)\} \ .$$

For $0 \le i \le |V(H)|-1$, let

$$B_i = \{\pi \in \mathrm{Sym}(n) \mid w^{y_i}_{G\pi} y_0 = y_i\} \ .$$

By Lemma 7.18,

$$\Pr_{\pi\sim \mathrm{U}(D)}(H \subset G\pi) = \Pr_{\pi\sim \mathrm{U}(D)}\left(\pi \in A \cap \bigcap_{i=0}^{|V(H)|-1} B_i\right) \ . \tag{7.18}$$

Let

$$\overline{A} = \{\pi \in A \mid w^{y_i}_{G\pi} y_0 \notin L \text{ for all } 1 \le i \le |V(H)|-1\} \ .$$

Observe that

$$A \cap \bigcap_{i=1}^{|V(H)|-1} B_i \subset \overline{A} \ .$$

Indeed, $\pi \in B_i$ implies that $w^{y_i}_{G\pi} y_0 = y_i$, and in particular $w^{y_i}_{G\pi} y_0 \notin L$. Hence, since $\overline{A} \subseteq A$,

$$A \cap \bigcap_{i=1}^{|V(H)|-1} B_i = \overline{A} \cap \bigcap_{i=1}^{|V(H)|-1} B_i \ . \tag{7.19}$$

Next, we show that $\overline{A} \subset B_0$. It always holds that $w^{y_0}_H(y_0) = y_0$, so $w^{y_0} \in \mathrm{PStab}_H(y_0)$. Consequently, $\pi \in \overline{A}$ implies $w^{y_0} \in \mathrm{Stab}_{G\pi}(y_0)$, which implies $\pi \in B_0$. Therefore, by (7.18) and (7.19),

$$\Pr_{\pi\sim \mathrm{U}(D)}(H \subset G\pi) = \Pr_{\pi\sim \mathrm{U}(D)}\big(\pi \in \overline{A}\big)\cdot \prod_{i=1}^{|V(H)|-1} \Pr_{\pi\sim \mathrm{U}(D)}\left(\pi \in B_i \mid \pi \in \overline{A} \cap \bigcap_{j=0}^{i-1} B_j\right) \ . \tag{7.20}$$

We claim that

$$\left|\Pr_{\pi\sim \mathrm{U}(D)}\big(\pi \in \overline{A}\big) - N_{G,\mathrm{Bipaths}_H(y_0)}(\mathrm{PStab}_H(y_0))\right| \le \frac{|L|\cdot|V(H)|}{n} \tag{7.21}$$

and

$$\Pr_{\pi\sim \mathrm{U}(D)}\left(\pi\in B_i \mid \pi\in \overline{A}\cap\bigcap_{j=0}^{i-1} B_j\right) = \frac{1}{n-|L|-i} \quad \forall 1\le i\le |V(H)|-1 \ , \tag{7.22}$$

and thus (7.17) follows from (7.20).

To prove (7.21), let

$$W_1 = \{x\in[n] \mid \mathrm{Stab}_G(x)\cap \mathrm{Bipaths}_H(y_0) = \mathrm{PStab}_H(y_0)\}$$

and

$$W_2 = \{x\in[n] \mid w_G^{y_i}x\in f(L) \text{ for some } 1\le i\le |V(H)|-1\} \ ,$$

and note that $\pi\in\overline{A}$ if and only if $\pi y_0\in W_1\setminus W_2$. Also, observe that πy_0 is distributed uniformly on $[n]\setminus f(L)$ when $\pi\sim \mathrm{U}(D)$. Hence,

$$\begin{aligned}
\left|\Pr_{\pi\sim \mathrm{U}(D)}(\pi\in\overline{A}) - \frac{|W_1\setminus W_2|}{n}\right| &= \left|\Pr_{x\sim \mathrm{U}([n]\setminus f(L))}(x\in W_1\setminus W_2) - \frac{|W_1\setminus W_2|}{n}\right| \\
&= \left|\Pr_{x\sim \mathrm{U}([n]\setminus f(L))}(x\in W_1\setminus W_2) - \Pr_{x\sim \mathrm{U}([n])}(x\in W_1\setminus W_2)\right| \\
&\le d_{\mathrm{TV}}(\mathrm{U}([n]\setminus f(L)), \mathrm{U}([n])) \qquad \text{by Fact 7.4} \\
&= \frac{|L|}{n} \ . \qquad (7.23)
\end{aligned}$$

Now,

$$|W_1| = N_{G,\mathrm{Bipaths}_H(y_0)}(\mathrm{PStab}_H(y_0))\cdot n$$

and

$$0\le |W_2|\le (|V(H)|-1)\cdot |L| \ .$$

Thus

$$\left|\frac{|W_1\setminus W_2|}{n} - N_{G,\mathrm{Bipaths}_H(y_0)}(\mathrm{PStab}_H(y_0))\right| \le \frac{(|V(H)|-1)\cdot|L|}{n} \ ,$$

which implies (7.21) by virtue of (7.23) and the triangle inequality.

We turn to proving (7.22). First note that if $\pi\in\overline{A}$ then $w_{G\pi}^{y_0}y_0, w_{G\pi}^{y_1}y_0,\ldots,w_{G\pi}^{y_{|V(H)|-1}}y_0$ are distinct. Indeed, if $w_{G\pi}^{y_i}y_0 = w_{G\pi}^{y_j}y_0$ then $(w^{y_j})^{-1}w^{y_i}\in \mathrm{Stab}_G(y_0)\cap\mathrm{Bipaths}_H(y_0)$, and thus $(w^{y_j})^{-1}w^{y_i}\in\mathrm{PStab}_H(y_0)$ because $\pi\in\overline{A}$. Consequently,

$$y_j = w_H^{y_j}y_0 = w_H^{y_i}y_0 = y_i \ ,$$

and so $i=j$ since $y_0,\ldots,y_{|V(H)|-1}$ are distinct.

In particular, for $1\le i\le |V(H)|-1$, the event $\pi\in\overline{A}\cap\bigcap_{j=0}^{i-1}B_j$ implies that

$$\pi^{-1}(w_G^{y_i}(\pi y_0)) = w_{G\pi}^{y_i}y_0\notin\{y_0,\ldots,y_{i-1}\} \ ,$$

and so

$$w_G^{y_i}(\pi y_0)\notin\{\pi y_0,\ldots,\pi y_{i-1}\} \ . \tag{7.24}$$

Observe that, since

$$\pi\in B_j \iff w_G^{y_j}(\pi y_0) = \pi y_j,$$

the event $\pi\in\overline{A}\cap\bigcap_{j=0}^{i-1}B_j$ is determined solely by the restriction $\pi_{|\{y_0,\ldots,y_{i-1}\}}$. In other words, for a permutation $\pi\in D$ we have

$$\pi\in\overline{A}\cap\bigcap_{j=0}^{i-1}B_j \iff \pi_{|\{y_0,\ldots,y_{i-1}\}}\in K \tag{7.25}$$

where

$$K = \left\{ g\colon \{y_0, \dots, y_{i-1}\} \to [n] \setminus f(L) \mid g \text{ is injective and } \pi_{|y_0,\dots,y_{i-1}} = g \text{ implies } \pi \in \overline{A} \cap \bigcap_{j=0}^{i-1} B_j \right\} .$$

Fix some $g \in K$, and consider a random permutation $\pi \sim \mathrm{U}(D)$, conditioned on $\pi_{|\{y_0,\dots,y_{i-1}\}} = g$. Note that πy_i is distributed uniformly on the set $[n] \setminus (L \cup \{\pi y_0, \dots, \pi y_{i-1}\})$. Since $\pi \in \overline{A}$, we have $w_G^{y_i}(\pi y_0) \notin L$. Together with (7.24), it follows that

$$\begin{aligned} \Pr_{\pi\sim \mathrm{U}(D)}\left(\pi \in B_i \mid \pi_{|\{y_0,\dots,y_{i-1}\}} = g\right) &= \Pr_{\pi\sim \mathrm{U}(D)}\left(\pi y_i = w_G^{y_i}(\pi y_0) \mid \pi_{|\{y_0,\dots,y_{i-1}\}} = g\right) \\ &= \frac{1}{|[n] \setminus (L \cup \{\pi y_0, \dots, \pi y_{i-1}\})|} = \frac{1}{n-L-i} . \end{aligned}$$

Finally, (7.22) follows by virtue of (7.25). □

We next generalize Lemma 7.19 to the case in which H is not necessarily connected, provided that it does not have too many edges.

Lemma 7.21. *Fix $G \in \mathcal{G}_S(n)$ and $H \in \mathcal{PG}_S(n)$. Denote the connected components of H by $C_1, \dots, C_k$, and suppose that every connected component has at least one edge. Fix $x_i \in V(C_i)$ for each $1 \le i \le k$. Suppose that*

$$|E(H)| \le n^{\frac{1}{4}} . \tag{7.26}$$

Then,

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G\pi) = n^{-r(H)} \prod_{i=1}^{k} N_{G,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i)) + O\left(n^{-r(H)-\frac{1}{4}}\right) ,$$

where the implied constant is universal.

Proof. Clearly,

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G\pi) = \prod_{i=1}^{k} \Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(C_i \subset G\pi \mid \bigcup_{j=1}^{i-1} C_j \subset G\pi\right) ,$$

where by $C_i \subset G\pi$ we mean that $G\pi$ contains every edge in the connected component C_i of H. To bound the i-th term of this product, let $L_i = \bigcup_{j=1}^{i-1} C_j$, and note that the event $\bigcup_{j=1}^{i-1} C_j \subset G\pi$ is determined solely by the restriction $\pi_{|L_i}$. Indeed, $G\pi$ contains the edge $a \xrightarrow{s} b$ of C_i if and only if $s_G(\pi a) = \pi b$. Hence there exists a set D_i of injective functions $f\colon L_i \to [n]$, such that

$$\bigcup_{j=1}^{i-1} C_j \subset G\pi \iff \pi_{|L_i} \in D_i .$$

Thus, we can write

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(C_i \subset G\pi \mid \bigcup_{j=1}^{i-1} C_j \subset G\pi\right) = \sum_{f\in D_i} \Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(\pi_{|L_i} = f \mid \pi_{|L_i} \in D_i\right) \cdot \Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(C_i \subset G\pi \mid \pi_{|L_i} = f\right) . \tag{7.27}$$

By Remark 7.20,

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(C_i \subset G\pi \mid \pi_{|L_i} = f\right) = n^{-(|C_i|-1)} \cdot \left(N_{G,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i)) \pm O\left(n^{-\frac{1}{2}}\right)\right) , \tag{7.28}$$

where we used (7.26), and the fact that both $|C_i|$ and $|L_i|$ are bounded from above by $2|E(H)|$. Now, (7.27) and (7.28) yield

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(C_i \subset G\pi \mid \bigcup_{j=1}^{i-1} C_j \subset G\pi\right) = n^{-(|C_i|-1)} \cdot \left(N_{G,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i)) \pm O\left(n^{-\frac{1}{2}}\right)\right) .$$

Since $r(H) = \sum_{i=1}^{k}(|C_i|-1)$, and $N_{G,P}(Q) \le 1$ for every P and Q, it follows that

$$\begin{aligned}\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G\pi) &= \prod_{i=1}^{k} \Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}\left(C_i \subset G\pi \mid \bigcup_{j=1}^{i-1} C_j \subset G\pi\right)\\ &= n^{-r(H)}\left(\prod_{i=1}^{k} N_{G,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i)) + O\left(\sum_{i=1}^{k}\binom{k}{i}\cdot n^{-\frac{i}{2}}\right)\right) ,\end{aligned}$$

which yields the lemma since $k \le |E(H)| \le n^{\frac{1}{4}}$. ☐

Corollary 7.22. *Let $G_0, G_1 \in \mathcal{G}_S(n)$, and take $H \in \mathcal{PG}_S(n)$. Suppose that $|E(H)| \le n^{\frac{1}{4}}$, and that every connected component of H has at least one edge. Then,*

$$\left|\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G_0\pi) - \Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G_1\pi)\right| \le n^{-r(H)}\cdot|E(H)|\cdot d_{\mathrm{TV}}\left(N_{G_0,B_{2|E(H)|}}, N_{G_1,B_{2|E(H)|}}\right) + cn^{-r(H)-\frac{1}{4}} ,$$

for a universal constant $c > 0$.

Proof. Suppose that H has k connected components, and let $x_1, \dots, x_k$ be representative vertices of these components. For $j \in \{0,1\}$, Lemma 7.21 yields

$$\Pr_{\pi\sim \mathrm{U}(\mathrm{Sym}(n))}(H \subset G_j\pi) = n^{-r(H)}\prod_{i=1}^{k} N_{G_j,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i)) + O\left(n^{-r(H)-\frac{1}{4}}\right) . \tag{7.29}$$

Now,

$$\begin{aligned}&\left|\prod_{i=1}^{k} N_{G_0,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i)) - \prod_{i=1}^{k} N_{G_1,\mathrm{Bipaths}_H(x_i)}(\mathrm{PStab}_H(x_i))\right|\\ &\le d_{\mathrm{TV}}\left(\prod_{i=1}^{k} N_{G_0,\mathrm{Bipaths}_H(x_i)}, \prod_{i=1}^{k} N_{G_1,\mathrm{Bipaths}_H(x_i)}\right) && \text{by Fact 7.4}\\ &\le \sum_{i=1}^{k} d_{\mathrm{TV}}\left(N_{G_0,\mathrm{Bipaths}_H(x_i)}, N_{G_1,\mathrm{Bipaths}_H(x_i)}\right) && \text{by Fact 7.6}\\ &\le k \cdot d_{\mathrm{TV}}\left(N_{G_0,B_{2|E(H)|}}, N_{G_1,B_{2|E(H)|}}\right) && \text{by Corollary 6.10}\\ &\le |E(H)| \cdot d_{\mathrm{TV}}\left(N_{G_0,B_{2|E(H)|}}, N_{G_1,B_{2|E(H)|}}\right) && \text{since } k \le |E(H)|. \qquad (7.30)\end{aligned}$$

The claim follows from (7.29) and (7.30). ☐

Lemma 7.16 now follows from Lemma 7.17 and Corollary 7.22.

Proof of Lemma 7.16. Denote $\delta = d_{\mathrm{TV}}\left(N_{G_0,B_{2q}}, N_{G_1,B_{2q}}\right)$. Then,

$$
\begin{aligned}
d_{\mathrm{TV}}(\theta_0,\theta_1) &= \frac{1}{2}\sum_{H\in\mathcal{R}_{\mathcal{M}}(n)}\left|\Pr_{\pi\sim\mathrm{U}(\mathrm{Sym}(n))}(H_{\mathcal{M},G_0\pi}=H)-\Pr_{\pi\sim\mathrm{U}(\mathrm{Sym}(n))}(H_{\mathcal{M},G_1\pi}=H)\right| \\
&= \frac{1}{2}\sum_{H\in\mathcal{R}_{\mathcal{M}}(n)}\left|\Pr_{\pi\sim\mathrm{U}(\mathrm{Sym}(n))}(H\subset G_0\pi)-\Pr_{\pi\sim\mathrm{U}(\mathrm{Sym}(n))}(H\subset G_1\pi)\right| \\
&\leq \frac{1}{2}\sum_{H\in\mathcal{R}_{\mathcal{M}}(n)}\left(n^{-r(H)}\cdot|E(H)|\cdot d_{\mathrm{TV}}\big(N_{G_0,B_{2|E(H)|}},N_{G_1,B_{2|E(H)|}}\big)+cn^{-r(H)-\frac{1}{4}}\right) && \text{by Corollary 7.22} \\
&\leq \frac{1}{2}\sum_{H\in\mathcal{R}_{\mathcal{M}}(n)}\left(n^{-r(H)}\cdot|E(H)|\cdot\delta+cn^{-r(H)-\frac{1}{4}}\right) && \text{by Corollary 6.10} \\
&\leq \frac{1}{2}\sum_{H\in\mathcal{R}_{\mathcal{M}}(n)}\left(n^{-r(H)}\cdot q\cdot\delta+cn^{-r(H)-\frac{1}{4}}\right) \\
&\leq \frac{1}{2}\sum_{r=0}^{q}|\mathcal{R}_{\mathcal{M},r}(n)|\cdot\left(n^{-r}\cdot q\cdot\delta+cn^{-r-\frac{1}{4}}\right) \\
&\leq \frac{1}{2}\sum_{r=0}^{q}\binom{q}{r}(2q)^{q-r}\cdot\left(q\delta+cn^{-\frac{1}{4}}\right) && \text{by Lemma 7.17} \\
&\leq \frac{1}{2}\sum_{r=0}^{q}(2q)^{q}\left(q\delta+cn^{-\frac{1}{4}}\right) \\
&\leq (2q)^{q+2}\left(\delta+cn^{-\frac{1}{4}}\right)\,.
\end{aligned}
$$

□

Acknowledgements

O.B. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 803711). A.L. is supported by a grant of the Institute for Advanced Study and by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (Grant No. 692854). J.M. is partially supported by NSF grant CCF-1814603.

J.M. wishes to thank Venkatesan Guruswami for a helpful conversation, and Nicolas Resch and João Ribeiro for useful comments.

References

[1] Herbert Abels, *An example of a finitely presented solvable group*, Homological group theory (Proc. Sympos., Durham, 1977), 1979, pp. 205–211. ↑40

[2] David Aldous and Russell Lyons, *Processes on unimodular random networks*, Electron. J. Probab. **12** (2007), no. 54, 1454–1508. ↑7

[3] Noga Alon, *Testing subgraphs in large graphs*, Random Structures & Algorithms **21** (2002), no. 3-4, 359–370. ↑2

[4] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira, *A combinatorial characterization of the testable graph properties: It's all about regularity*, SIAM J. Comput. **39** (2009), no. 1, 143–167. ↑13

[5] Noga Alon and Asaf Shapira, *Every monotone graph property is testable*, SIAM J. Comput. **38** (2008), no. 2, 505–522. ↑13

[6] Goulnara Arzhantseva and Liviu Păunescu, *Almost commuting permutations are near commuting permutations*, J. Funct. Anal. **269** (2015), no. 3, 745–757. ↑6, 9, 10, 11, 14, 15

[7] Oren Becker and Michael Chapman, *Stability of approximate group actions: uniform and probabilistic*, available at https://arxiv.org/abs/2005.06652. ↑10
[8] Oren Becker and Alexander Lubotzky, *Group stability and Property (T)*, J. Funct. Anal. **278** (2020), no. 1, 108298, 20. ↑10, 11, 16, 17
[9] Oren Becker, Alexander Lubotzky, and Andreas Thom, *Stability and invariant random subgroups*, Duke Math. J. **168** (2019), no. 12, 2207–2234. ↑10, 11, 16, 18, 19, 39, 40
[10] Oren Becker and Jonathan Mosheiff, *Abelian Groups Are Polynomially Stable*, Int. Math. Res. Not. IMRN, to appear. ↑1, 6, 11, 13, 14, 15, 16
[11] Bachir Bekka, Pierre de la Harpe, and Alain Valette, *Kazhdan's property (T)*, New Mathematical Monographs, vol. 11, Cambridge University Press, Cambridge, 2008. ↑15
[12] Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. ↑7
[13] Hubie Chen, *A rendezvous of logic, complexity, and algebra*, ACM Comput. Surv. **42** (December 2009), no. 1. ↑2
[14] A. Connes, J. Feldman, and B. Weiss, *An amenable equivalence relation is generated by a single transformation*, Ergodic Theory Dynam. Systems **1** (1981), no. 4, 431–450 (1982). ↑8, 19
[15] Marius Dadarlat, *Obstructions to matricial stability of discrete groups and almost flat k-theory*, available at https://arxiv.org/abs/2007.12655. ↑11
[16] Marcus De Chiffre, Lev Glebsky, Alexander Lubotzky, and Andreas Thom, *Stability, cohomology vanishing, and nonapproximable groups*, Forum Math. Sigma **8** (2020), Paper No. e18, 37. ↑11
[17] Cornelia Druţu and Michael Kapovich, *Geometric group theory*, American Mathematical Society Colloquium Publications, vol. 63, American Mathematical Society, Providence, RI, 2018. With an appendix by Bogdan Nica. ↑15
[18] D.Z. Du and P.M. Pardalos, *Minimax and applications*, Nonconvex Optimization and Its Applications, Springer US, 1995. ↑27
[19] Sø ren Eilers, Tatiana Shulman, and Adam P. W. Sø rensen, *C^*-stability of discrete groups*, Adv. Math. **373** (2020), 107324, 41. ↑11
[20] Gábor Elek, *Finite graphs and amenability*, J. Funct. Anal. **263** (2012), no. 9, 2593–2614. ↑8, 19
[21] Dominic Enders and Tatiana Shulman, *Almost commuting matrices, cohomology, and dimension*, available at https://arxiv.org/abs/1902.10451. ↑11
[22] Lev Glebsky and Luis Manuel Rivera, *Almost solutions of equations in permutations*, Taiwanese J. Math. **13** (2009), no. 2A, 493–500. ↑6, 9, 10, 15
[23] Oded Goldreich, *Introduction to property testing*, Cambridge University Press, Cambridge, 2017. ↑13, 14
[24] Don Hadwin and Tatiana Shulman, *Stability of group relations under small Hilbert-Schmidt perturbations*, J. Funct. Anal. **275** (2018), no. 4, 761–792. ↑11
[25] Adrian Ioana, *Stability for product groups and property (τ)*, J. Funct. Anal. **279** (2020), no. 9, 108729, 32. ↑1, 6, 9, 10, 11
[26] D. L. Johnson, *Presentations of groups*, Second edition, London Mathematical Society Student Texts, vol. 15, Cambridge University Press, Cambridge, 1997. MR1472735 ↑41
[27] Nir Lazarovich, Arie Levit, and Yair Minsky, *Surface groups are flexibly stable*, available at https://arxiv.org/abs/1901.07182. ↑11
[28] D.A. Levin and Y. Peres, *Markov chains and mixing times*, MBK, American Mathematical Society, 2017. ↑28
[29] Arie Levit and Alexander Lubotzky, *Infinitely presented stable groups and invariant random subgroups of metabelian groups*, available at https://arxiv.org/abs/1909.11842. ↑11
[30] ———, *Uncountably many permutation stable groups*, available at https://arxiv.org/abs/1910.11722. ↑11

[31] Gilbert Levitt, *On the automorphism group of generalized Baumslag-Solitar groups*, Geom. Topol. **11** (2007), 473–515. ↑10
[32] Alexander Lubotzky and Izhar Oppenheim, *Non p-norm approximated groups*, available at https://arxiv.org/abs/1807.06790. ↑11
[33] Alexander Lubotzky and Andrzej Zuk, *On property (τ)*, available at http://www.ma.huji.ac.il/~alexlub/. ↑15
[34] John Milnor, *Introduction to algebraic K-theory*, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1971. Annals of Mathematics Studies, No. 72. ↑40
[35] Marco Antonio García Morales and Lev Glebsky, *Property of defect diminishing and stability*, available at https://arxiv.org/abs/1911.11752. ↑11
[36] Ilan Newman and Christian Sohler, *Every property of hyperfinite graphs is testable*, SIAM J. Comput. **42** (2013), no. 3, 1095–1112. ↑8, 13, 19
[37] Izhar Oppenheim, *Garland's method with banach coefficients*, available at https://arxiv.org/abs/2009.01234. ↑11
[38] Donald S. Ornstein and Benjamin Weiss, *Ergodic theory of amenable group actions. I. The Rohlin lemma*, Bull. Amer. Math. Soc. (N.S.) **2** (1980), no. 1, 161–164. ↑8, 19
[39] S. M. Ulam, *A collection of mathematical problems*, Interscience Tracts in Pure and Applied Mathematics, no. 8, Interscience Publishers, New York-London, 1960. ↑11
[40] A. Vershik, *Amenability and approximation of infinite groups*, 1982, pp. 311–330. Selected translations. ↑16
[41] Andrew Chi-Chih Yao, *Probabilistic computations: Toward a unified measure of complexity (extended abstract)*, 18th annual symposium on foundations of computer science, providence, rhode island, usa, 31 october - 1 november 1977, 1977, pp. 222–227. ↑27
[42] Tianyi Zheng, *On rigid stabilizers and invariant random subgroups of groups of homeomorphisms*, available at https://arxiv.org/abs/1901.04428. ↑11
[43] Dmitriy Zhuk, *A proof of the CSP dichotomy conjecture*, J. ACM **67** (2020), no. 5, 30:1–30:78. ↑2

Appendix A. An explicit description of certain systems of relations

A.1. **A system of relations which is testable but not stable.** Here we describe instable testable systems of relations using Theorem 1', as discussed in Section 3.1. Fix a prime number p. Let

$$S_p = \{d_2, d_3, s_{12}, s_{13}, s_{23}, s_{24}, s_{34}\} ,$$

and consider the following system of relations over the alphabet S_p:

$$\begin{aligned}
E_p =&\{d_2d_3 = d_3d_2, s_{12}s_{34} = s_{34}s_{12}\}\cup \\
&\{s_{23}s_{12} = s_{12}s_{23}s_{13}, s_{34}s_{23} = s_{23}s_{34}s_{24}\}\cup \\
&\{s_{13}s_{12} = s_{12}s_{13}, s_{13}s_{23} = s_{23}s_{13}, s_{24}s_{23} = s_{23}s_{24}\}\cup \\
&\{s_{24}s_{34} = s_{34}s_{24}, s_{13}s_{24} = s_{24}s_{13}\}\cup \\
&\{s_{12}d_2 = d_2s_{12}^p, s_{12}d_3 = d_3s_{12}\}\cup \\
&\{d_2s_{23} = s_{23}^pd_2, s_{23}d_3 = d_3s_{23}^p\}\cup \\
&\{s_{34}d_2 = d_2s_{34}, d_3s_{34} = s_{34}^pd_3\}.
\end{aligned}$$

As explained below, this system is testable by Theorem 1', yet instable due to [9, Theorem 1.3(ii)].

It was shown in [1] that $\langle S_p \mid E_p \rangle$ is a presentation of *Abels' group* A_p, which is defined as follows:

$$A_p = \left\{ \begin{pmatrix} 1 & * & * & * \\ 0 & p^m & * & * \\ 0 & 0 & p^n & * \\ 0 & 0 & 0 & 1 \end{pmatrix} \subset \mathrm{GL}_4 \mathbb{Z}[1/p] \mid m, n \in \mathbb{Z} \right\} .$$

Here $\mathbb{Z}[1/p]$ is the ring of rational number whose denominator is a power of p, and $\mathrm{GL}_4 \mathbb{Z}[1/p]$ is the group of 4×4 matrices with entries in $\mathbb{Z}[1/p]$ and determinant $\pm p^l$, $l \in \mathbb{Z}$.

An isomorphism $\langle S_p \mid E_p \rangle \xrightarrow{\sim} A_p$ is given by

$$d_2 \mapsto \begin{pmatrix} 1 & & & \\ & p & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \quad d_3 \mapsto \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & p & \\ & & & 1 \end{pmatrix}$$

$$s_{ij} \mapsto I + e_{ij} \ ,$$

where e_{ij} is the 4×4 matrix with 1 on the (i, j)-entry and 0 elsewhere.

The group A_p is solvable because it is contained in the group of upper triangular matrices. Hence A_p is amenable. In other words, E_p satisfies the hypothesis of Theorem 1', and so it is testable.

On the other hand, [9, Theorem 1.3(ii)] characterizes the systems of relations that are stable, among those that satisfy the hypothesis of Theorem 1'. By this characterization (which is given in terms of *invariant random subgroups* and *cosoficity*), E_p is instable [9, Corollary 8.7].

A.2. **A non-testable system of relations.** Let $m \geq 3$. Here, as discussed in Section 3.2, we use Theorem 2' to provide a non-testable system of relations E_m, over an alphabet $S_m^{\pm}$, such that $\Gamma(E_m) \coloneqq \langle S_m \mid E_m \rangle \cong \mathrm{SL}_m \mathbb{Z}$:

$$S_m = \{ s_{ij} \mid i, j \in [m], i \neq j \} \ ,$$

$$\begin{aligned} E_m = & \{ s_{ij} s_{kl} = s_{kl} s_{ij} \mid i, j, k, l \in [m], j \neq k, i \neq l \} \cup \\ & \{ s_{ij} s_{jk} = s_{ik} s_{jk} s_{ij} \mid i, j, k \in [m], i \neq j, j \neq k, k \neq i \} \cup \\ & \left\{ \left(s_{12} s_{21}^{-1} s_{12} \right)^4 = 1 \right\} \ . \end{aligned}$$

By [34, Corollary 10.3], $\langle S_m \mid E_m \rangle$ is a presentation of the group $\mathrm{SL}_m \mathbb{Z}$. The isomorphism given by

$$s_{ij} \mapsto I + e_{ij} \ ,$$

where e_{ij} is the $m \times m$ matrix with 1 on the (i, j)-entry and 0 elsewhere. The group $\mathrm{SL}_m \mathbb{Z}$ is well known to satisfy Property (T), and has infinitely many finite quotients. This means that E_m satisfies the hypothesis of Theorem 2', and therefore, it is not testable.

Appendix B. A note on sampling inverses of permutations

For a given permutation $\sigma \in \mathrm{Sym}(n)$ and $x \in [n]$, our model assumes that an algorithm $\mathcal{M}$ can read σx by making a single query. We also allow $\mathcal{M}$ to read $\sigma^{-1} x$ with a single query. It is also natural to study testability in a model that allows to read σx, but not $\sigma^{-1} x$, with a single query. Here we describe how every testable (resp. stable) system gives rise to a closely related system which is testable (resp. stable) even under the model where sampling inverses is not allowed (see Definitions 1.6 and 1.10).

Let E be a system of equations over $S^{\pm}$. We say the E is *inverseless* if all equations in E are over the alphabet S, that is, letters from S^{-1} are not used. For example, $E_2^{\mathrm{comm}} = \{\mathsf{XY} = \mathsf{YX}\}$ is inverseless. Thus, $\mathrm{SAS}_k^{E_2^{\mathrm{comm}}}$, $k \in \mathbb{N}$, can be implemented without sampling inverses (simply by checking whether $\mathsf{XY} x_j = \mathsf{YX} x_j$ rather than $\mathsf{X}^{-1}\mathsf{Y}^{-1}\mathsf{XY} x_j = x_j$ in Line 2 of Algorithm 1). Similarly, SAS_k^E can be implemented without sampling inverses whenever E is inverseless.

In general, E gives rise to an inverseless system of equations E' as follows. Extend the alphabet by defining $\overline{S} = \{\overline{s}_1, \ldots, \overline{s}_d\}$ and $S' = S \cup \overline{S}$. Define a system of equations $\tilde{E}$ over S' by starting from E and replacing each occurrence of each s_i^{-1}, $1 \le i \le d$, by $\overline{s}_i$. Finally, set $E' = \tilde{E} \cup \underbrace{\{s_i \overline{s}_i = 1 \mid 1 \le i \le d\}}_{=: E_{\mathrm{inv}}}$. Then E' is an inverseless system of equations, i.e., it is a system of equations over S' rather than $(S')^{\pm}$. Furthermore, E is stable if and only if E' is stable because the groups $\Gamma(E)$ and $\Gamma(E')$ are isomorphic and by Proposition 3.1.

Moreover, by Proposition 3.2, E is testable if and only if E' is. Additionally, the algorithm $\mathrm{LSM}^{E}_{k,P,\delta}$ (see Section 1.4) can be altered to avoid sampling inverses as follows: First run $\mathrm{SAS}^{E_{\mathrm{inv}}}_{k'}$ for a large enough k', and reject if $\mathrm{SAS}^{E_{\mathrm{inv}}}_{k'}$ rejects. Otherwise, run $\mathrm{LSM}^{E}_{k,P,\delta}$ without sampling inverses by replacing each query of $s_i^{-1}x$ by a query of $\overline{s}_i x$, and accept if $\mathrm{LSM}^{E}_{k,P,\delta}$ accepts.

Appendix C. A review of free groups and group presentations

Here we give a brief introduction to free groups and their universal property, and to group presentations (see [26]).

Let $S = \{s_1, \ldots, s_d\}$ be a set of letters. Write $S^{-1} = \{s_1^{-1}, \ldots, s_d^{-1}\}$ for the set of formal inverses of the letters in S, and let $S^{\pm} \coloneqq S \cup S^{-1}$. We define $\left(s_i^{-1}\right)^{-1} \coloneqq s_i$, and so $s \mapsto s^{-1}$ becomes an involution on $S^{\pm}$. Write $w_1 \sim w_2$ for words w_1 and w_2 over $S^{\pm}$ if there is a sequence of words $w_1 = u_0, \ldots, u_n = w_2$ such that u_{i+1} is obtained from u_i by adding or removing a *null subword*, i.e., a subword of the form ss^{-1}, $s \in S^{\pm}$. For example, $s_1 s_2 s_2^{-1} s_1 \sim s_1 s_1 s_3^{-1} s_3$. A word w over the alphabet $S^{\pm}$ is *reduced* if it does not contain a null subword as above. Every word w over $S^{\pm}$ is equivalent to a unique reduced word called the *reduced form* of w. For words w_1, w_2, w_1', w_2' over $S^{\pm}$ such that $w_1 \sim w_1'$ and $w_2 \sim w_2'$, we have $w_1 w_2 \sim w_1' w_2'$. In particular, the reduced forms of $w_1 w_2$ and $w_1' w_2'$ are equal.

The *free group* F_S over S is the set of reduced words over $S^{\pm}$, endowed with the following multiplication operation: for $w_1, w_2 \in F_S$, $w_1 \cdot w_2$ is the unique reduced word equivalent to the concatenation $w_1 w_2$. We sometimes abuse notation and view a word w over $S^{\pm}$, not necessarily reduced, as an element of F_S. We do so only when making statements where only the equivalence class of w matters.

It is worth noting the following alternative way to define F_S, although we are not using it in this paper. One can realize the free group F_S as the group of $\sim$-equivalence classes of (not necessarily reduced) of words over $S^{\pm}$, with the multiplication defined by $[w_1] \cdot [w_2] \coloneqq [w_1 w_2]$, where $[w]$ denotes the $\sim$-equivalence class of a word w. This point of view also makes it easy to define *group presentations*, i.e., to define the group $\langle S \mid R \rangle$ for a set of words R over $S^{\pm}$: We define an equivalence class $\sim_R$ on the set of words over $S^{\pm}$ just as we defined $\sim$, only that we consider all subwords of the form $vr^{\pm 1}v^{-1}$ to be null subwords, for every word v over $S^{\pm}$ and $r \in R$, in a addition to the null subwords ss^{-1}, $s \in S$. We then let $\langle S \mid R \rangle$ be the group of $\sim_R$-equivalence classes, with the multiplication of classes defined as in F_S above. We now go back to thinking of F_S as the set of reduced words and define $\langle S \mid R \rangle$ in a different, equivalent, way.

The free group F_S has the following important property, known as *the universal property of* F_S: for every group Γ and function $f \colon S \to \Gamma$, there is exactly one group homomorphism $\tilde{f} \colon F_S \to \Gamma$ such that $\tilde{f}(s_i) = f(s_i)$ for all $1 \le i \le d$. The map $f \mapsto \tilde{f}$ is a bijection from the set of functions $S \to \Gamma$ to the set of homomorphisms $F_S \to \Gamma$.

In particular, for a group Γ generated by $\gamma_1, \ldots, \gamma_d \in \Gamma$, there is a unique homomorphism $\pi \colon F_S \to \Gamma$ such that $\pi(s_i) = \gamma_i$ for each $1 \le i \le d$. The map π is surjective because its image $\pi(F_S)$ contains a generating set for Γ. That is, every group Γ generated by d elements is a quotient of F_S. For example, $\Gamma_0 \coloneqq \mathbb{Z}^2$ is a quotient of $F_{\{s_1, s_2\}}$ as exhibited by the unique homomorphism $\pi_0 \colon F_{\{s_1,s_2\}} \to \mathbb{Z}^2$ sending $s_1 \mapsto (1,0)$ and $s_2 \mapsto (0,1)$.

The kernel $\ker \pi$ of $\pi \colon F_S \to \Gamma$ is a normal subgroup of F_S. It is often useful to have a subset R of $\ker \pi$ such that $\ker \pi = \langle\langle R \rangle\rangle$. Here $\langle\langle R \rangle\rangle$ denotes the normal closure of R in F_S, that is, the smallest normal subgroup of F_S that contains R. Concretely, $\langle\langle R \rangle\rangle$ consists of all elements of the form $\prod_{i=1}^{m} v_i r_i^{\varepsilon_i} v_i^{-1}$, where

$m \geq 0$, $v_i \in F_S$, $r_i \in R$ and $\varepsilon_i \in \{\pm 1\}$. In the cases of interest of this paper, $\ker \pi = \langle\langle R \rangle\rangle$ where R is a finite set. For $\pi_0 \colon F_{\{s_1,s_2\}} \to \mathbb{Z}^2$ as in the example, it can be shown that $\ker \pi_0 = \langle\langle R_0 \rangle\rangle$ for $R_0 = \{s_1^{-1} s_2^{-1} s_1 s_2\}$.

The surjective homomorphism $\pi \colon F_S \to \Gamma$ gives rise to an isomorphism $F_S / \ker \pi \xrightarrow{\sim} \Gamma$. If $\ker \pi = \langle\langle R \rangle\rangle$ for $R \subset \ker \pi$, we write $\langle S \mid R \rangle$ for the quotient group $F_S / \ker \pi$ and say that the group $\langle S \mid R \rangle$, which is isomorphic to Γ, is a *presentation* of Γ. For a system of relations E, we write $\langle S \mid E \rangle$ for $\langle S \mid R_E \rangle$ (where R_E is as in the introduction).

For example,

$$\langle s_1, s_2 \mid s_1 s_2 = s_2 s_1 \rangle = \langle s_1, s_2 \mid s_1^{-1} s_2^{-1} s_1 s_2 \rangle \cong \mathbb{Z}^2 \ .$$

The group $\langle S \mid R \rangle$ is a *finite presentation* if S and E are finite sets. In this paper we also write $\Gamma(E)$ for $\langle S \mid E \rangle$ when the set S is understood from the context.