Constant time testability of FOMOD on finitary graphs

Jenny Stimpson

Otto-Friedrich-Universität Bamberg

Tuesday 3rd March 2026, AIMoTh Passau

Joint work with Isolde Adler

Outline

Property Testing: Closeness

We are interested in simple undirected graphs.

A **graph property** is a subclass of the class of all graphs which is closed under isomorphism.

Let $\varepsilon \in (0, 1)$. We call a graph G ε-**close** to property $\mathcal{P}$ iff there is a graph $G' \in \mathcal{P}$ such that G' can be reached from G by performing at most εdn edge modifications. Otherwise G is ε-**far** from $\mathcal{P}$.

In particular graphs can only be close if they have the same number of vertices.

Property Testing: The Bounded Degree Model

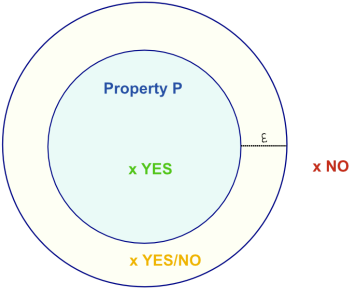
In property testing we sacrifice the algorithms ability to view the entire input in order to achieve sublinear running times.

In the bounded degree model we assume that for any input graph G:

- There is a degree bound d s.t. every vertex of our input graph has degree at most d.
- We access the graph by making queries to an **oracle function** $Q(v, i)$ which returns the i^{th} neighbour of vertex v in constant time.
- The number of vertices n is in the input.

Property Testing: ε-Testers and Testability

An ε-**tester** for a property $\mathcal{P}$ is an algorithm which, with probability $2/3$, returns TRUE if the input graph has the property, or FALSE if the input graph is ε-far from having the property.



A property $\mathcal{P}$ is **testable** if for every $\varepsilon \in (0, 1)$, there is an ε-tester for $\mathcal{P}$ which has constant query complexity in the size of the input graph.

Motivation

We are interested in **algorithmic meta-theorems** for property testing.

Newman & Sohler (2011)

Every property of bounded tree-width graphs is testable with constant query complexity.

Adler & Harwath (2018)

Every CMSO definable property on the class K_d^t of all graphs with tree-width $\leq t$ and degree bound $\leq d$ is testable with constant query complexity and polylogarithmic running time.

Motivation

Open Problem

Can constant running time be achieved for CMSO definable properties on K_d^t?

Towards this goal we investigate the case where we also bound the tree-depth of inputs: finitary graphs.

Theorem

Let $c \in \mathbb{N}$. Every CMSO definable property on the class of c-finitary, graphs is testable in constant time.

Motivation

Open Problem

Can constant running time be achieved for CMSO definable properties on K_d^t?

Towards this goal we investigate the case where we also bound the tree-depth of inputs: finitary graphs.

Theorem

Let $c \in \mathbb{N}$. Every CMSO definable property on the class of c-finitary, graphs is testable in constant time.

For brevity I will present only the result for FO.

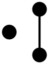
Finitary Graphs

A graph G is called c-**finitary** if each connected component of G contains at most c vertices.

The class of all c-finitary, d-bounded degree graphs is denoted $\mathcal{C}_d^c$.

The number of pairwise non-isomorphic connected components appearing in members of $\mathcal{C}_d^c$ is finite and denoted $M(c, d)$.

Motivating Example

Consider the property  -freeness.

Graphs which **do not** contain both an isolated vertex and an isolated edge.

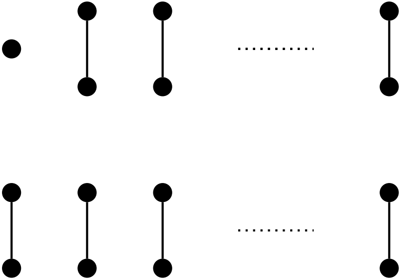
This is expressible in FO by the sentence

$$\psi := \neg\Big(\exists x\Big(\forall y(\neg Exy)\Big) \\ \wedge \exists x \exists y\Big(Exy \wedge \forall z\big((Exz \to z = y) \wedge (Eyz \to z = x)\big)\Big)\Big).$$

As it is similar to subgraph freeness it seems trivial to test in constant time on the class $\mathcal{C}_1^2$ - but is it?

Motivating Example

Consider the following two graphs in $\mathcal{C}_1^2$ (supposing each contains a "very large" number of vertices):



G_{odd}: (Is ε-far from P_ψ.)

G_{even}: (Has property P_ψ.)

To a tester seeing a constant size sample these graphs will look identical (whp).

Motivating Example

We must use the number of vertices n in the input graph.

ε-tester for $P_\psi \cap \mathcal{C}_1^2$

Input: Oracle access to a graph G, the size n of G.

1. Sample a sufficiently large, constant number of vertices and explore the components they lie in.
2. Accept if no edges are seen in the sample.
3. Accept if no singleton vertices are seen and n is even.
4. Reject otherwise.

Component Vectors

Component Histogram Vector: A vector $\text{chv}_c(G)$ of length $M(c, d)$ whose ith entry is the number of connected components of G with component type t_i.

A k-**capped component histogram vector** (k-CCHV) $\overline{a}$ is a vector of length $M(c, d)$ with entries from the set

$$\{0, 1, 2, \ldots, k-1, \text{“} \geq k\text{”}\}$$

A d-bounded degree, c-finitary graph G is said to *satisfy* a k-capped component histogram vector $\overline{a}$ if for all $i \in \{1, \ldots, M(c, d)\}$, either

$$\overline{a}[i] = \text{chv}_c(G)[i], \text{ or } \overline{a}[i] = \text{“} \geq k\text{”} \text{ and } \text{chv}_c(G)[i] \geq k.$$

Hanf Sentences

A **Hanf sentence** of FO with **radius** r is a sentence:

$$\exists^{\geq m} x(\tau_r(x)) := \exists x_1 \exists x_2, \ldots, \exists x_m \left(\bigwedge_{1 \leq i < j \leq m} x_i \neq x_j \wedge \bigwedge_{1 \leq i \leq m} \tau_r(x_i) \right)$$

"There exist at least m vertices x such that the r-ball around x has isomorphism type τ."

Hanf Normal Form

A sentence of FO is in **Hanf normal form** if it is a Boolean combination of Hanf sentences.

Hanf's Locality Theorem for FO

Let $d \in \mathbb{N}$. For every sentence φ of FO and every $d \in \mathbb{N}$ there is an FO sentence ψ in Hanf normal form such that $\varphi \equiv_d \psi$.

Component Normal Form on $\mathcal{C}_d^c$

From Hanf normal form we can make a series of transformations which preserve truth on graphs in $\mathcal{C}_d^c$ to reach **component normal form**.

This expresses a property P_φ as a finite collection of k-CCHVs of which every $G \in P_\varphi$ must satisfy at least one.

Component Normal Form

Let $c, d \in \mathbb{N}$, $c \geq 1$. For every satisfiable FOMOD sentence φ there exist a $k \in \mathbb{N}$ and a finite set X of k-capped component histogram vectors such that $P_\varphi \cap \mathcal{C}_d^c$ contains exactly the c-finitary, d-bounded degree graphs which satisfy at least one of the histogram vectors in X, i.e.

$$P_\varphi \cap \mathcal{C}_d^c = \{G \in \mathcal{C}_d^c \mid G \text{ satisfies } \overline{a} \text{ for some } \overline{a} \in X\}.$$

Obtaining Component Normal Form

A sketch of the translation from HNF to CNF on graphs in $\mathcal{C}_d^c$:

1. Begin with an FOMOD sentence in HNF which is satisfiable in $\mathcal{C}_d^c$.
2. Rearrange into disjunctive normal form of Hanf sentences.
3. By substitution, increase the radius of each Hanf sentence to $c-1$.
4. Let k be the largest m such that a Hanf sentence $\exists^{\geq m}x(c-1,\tau)$ appears. For each smaller m replace $\exists^{\geq m}x(c-1,\tau)$ with $\exists^{=m}x(c-1,\tau) \vee \cdots \vee \exists^{=k-1}x(c-1,\tau) \vee \exists^{\geq k}x(c-1,\tau)$.
5. Each conjunctive clause now exactly describes a collection of k-CCHVs and we take the union of these as our CNF.

Component Normal Form

For our result

Theorem (FO)

Let $c \in \mathbb{N}$. Every FO definable property on the class of c-finitary graphs is testable in constant time.

it is sufficient to prove it only for properties whose CNF is **exactly one** k-CCHV, $\bar{u}$.

Rare and Frequent Components

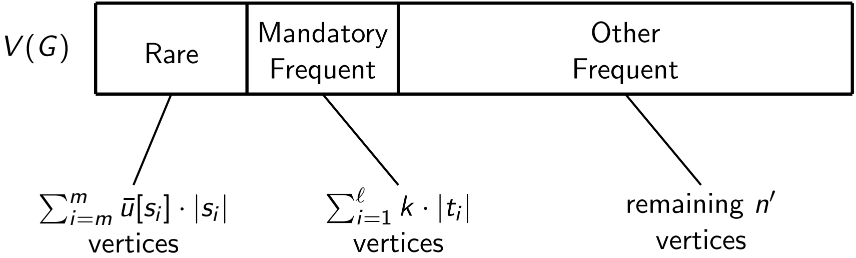
Given a particular k-CCHV $\bar{u}$ and $n \in \mathbb{N}$ we divide possible component types t_i into "rare" and "frequent" by their corresponding entries in $\bar{u}$.

$$\text{Component } t_i \text{ is called } \begin{cases} \text{rare} & \text{if } \bar{u}[t_i] \in \{1, \ldots, k-1\}, \\ \text{frequent} & \text{if } \bar{u}[t_i] = \text{``} \geq k \text{''}. \end{cases}$$

We rename the frequent components $t_1, \ldots, t_\ell$ and the rare components $s_1, \ldots, s_m$.

Graphs Satisfying a k-CCHV

Let $\bar{u}$ be a k-CCHV.



Testing Algorithm

Let P_φ be a property who's component normal form is a single k-CCHV, $\bar{u}$.

Define $g := \gcd\{|t_1|, \ldots, |t_\ell|\}$.

ε-tester for $P_\varphi \cap \mathcal{C}_d^c$

1. Sample a sufficiently large number of vertices, and explore the components they lie in.
2. If a rare component is seen during sampling reject.
3. Otherwise compute $n' = n - \sum_{i=m}^{m} \bar{u}[s_i] \cdot |s_i| - \sum_{i=1}^{\ell} k \cdot |t_i|$ and accept iff $g | n'$.

This algorithm has constant running time.

Correctness - Yes Instances

Suppose $G \in P_\varphi$. Then G satisfies $\bar{u}$.

The number of vertices of G in rare components is constant, and thus we will not see them during sampling (whp).

The remaining n' vertices of G are in the disjoint union of frequent components.

Since g divides $|t_i|$ for every frequent component t_i, we see $g|n'$ so the tester accepts.

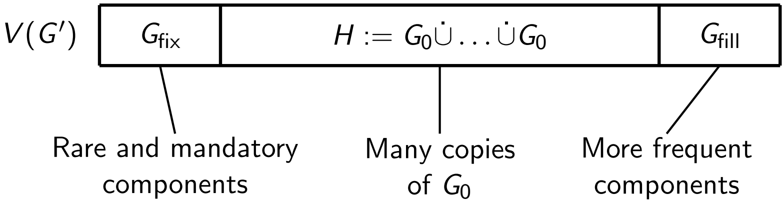
Correctness - ε-far Instances

Suppose that G is ε-far from $P_\varphi \cap \mathcal{C}_d^c$, but the algorithm accepts.

We reach contradiction by constructing a graph $G' \in P_\varphi \cap \mathcal{C}_d^c$ such that G' is ε-close to G.

Let G_0 be the graph consisting of all vertices we saw during sampling and their connected components.

G' has 3 sections:



Frobenius Coin Theorem

How do we know some G_{fill} exists on the right number of vertices?

Given a finite collection of coin denominations, which values can you give exact change for?

Frobenius Coin Theorem (I. Schur)

For every setwise coprime set of integers $\{a_1, \dots, a_n\}$, there exists a largest integer $\mathcal{F}$ which cannot be expressed as an integer linear combination of $a_1, \dots, a_n$.

Corollary

An immediate consequence of this is that for every set of integers $\{a_1, \dots, a_n\}$ there exists a largest multiple of $g := \gcd(a_1, \dots, a_n)$ which cannot be expressed as an integer linear combination of $a_1, \dots, a_n$.

Correctness - ε-far Instances

Let F be the largest multiple of g which is not a conical combination of the integers $|t_1|, \dots, |t_\ell|$.

We choose H to be the largest number of copies of G_0 such that

$$n' - |H| > F.$$

By assumption $g|n'$, the desired size $|G_{\text{fill}}|$ is a multiple of g greater than F. By the corollary such a G_{fill} exists.

Thus $G' \in P_\varphi \cap \mathcal{C}_d^c$.

Finally we need to check that G' is ε-close to G.

Correctness - ε-far Instances

As we sampled sufficient vertices, G_0 (and thus H) are $\frac{\varepsilon}{2}$-close* to G by a result of Newman & Sohler (2011).

As our graph is sufficiently large, most vertices in G' are in H. Some calculations show that G' and H are $\frac{\varepsilon}{2}$-close*.

Finally we conclude that G' and G are ε-close, a contradiction. □

Correctness - ε-far Instances

As we sampled sufficient vertices, G_0 (and thus H) are $\frac{\varepsilon}{2}$-close* to G by a result of Newman & Sohler (2011).

As our graph is sufficiently large, most vertices in G' are in H. Some calculations show that G' and H are $\frac{\varepsilon}{2}$-close*.

Finally we conclude that G' and G are ε-close, a contradiction. □

*For brevity here I am amalgamating two closeness measures!

Conclusion and Outlook

- We have seen that FO is constant time testable on $\mathcal{C}_d^c$.
- Since MSO and FO have equivalent expressive power on graphs of bounded tree depth, it follows immediately that MSO is also constant time testable on $\mathcal{C}_d^c$.
- The same result can be proven for FOMOD (and therefore CMSO) with some careful adaptations. In particular Hanf normal form for FOMOD was given by Nurmonen (2000).

Open Problem

Can constant running time be achieved for CMSO definable properties on K_d^t?