

# Testing Graph Properties with the Container Method

Eric Blais

David R. Cheriton School of Computer  
Science

University of Waterloo  
Waterloo, Canada

*eric.blais@uwaterloo.ca*

Cameron Seth

David R. Cheriton School of Computer  
Science

University of Waterloo  
Waterloo, Canada

*cjmpseth@uwaterloo.ca*

---

**Abstract-** We establish nearly optimal sample complexity bounds for testing the  $\rho$ -clique property in the dense graph model. Specifically, we show that it is possible to distinguish graphs on  $n$  vertices that have a  $\rho n$ -clique from graphs for which at least  $\epsilon n^2$  edges must be added to form a  $\rho n$ -clique by sampling and inspecting a random subgraph on only  $\tilde{O}(\rho^3/\epsilon^2)$  vertices.

We also establish new sample complexity bounds for  $\epsilon$ -testing  $k$ -colorability. In this case, we show that a sampled subgraph on  $\tilde{O}(k/\epsilon)$  vertices suffices to distinguish  $k$ -colorable graphs from those for which any  $k$ -coloring of the vertices causes at least  $\epsilon n^2$  edges to be monochromatic.

The new bounds for testing the  $\rho$ -clique and  $k$ -colorability properties are both obtained via new extensions of the graph container method. This method has been an effective tool for tackling various problems in graph theory and combinatorics. Our results demonstrate that it is also a powerful tool for the analysis of property testing algorithms.

## I. Introduction

The paper discusses the problem of testing if a graph contains a large clique or is  $k$ -colorable by examining only a small fraction of the graph. It introduces the concept of a canonical tester, which samples a subset of vertices and examines the induced subgraph to determine if the graph has

a certain property. The sample complexity for testing the  $\rho$ -CLIQUE property is shown to be  $(\tilde{O}(\rho^3/\epsilon^2))$ , which is nearly optimal. This result has been significant implications, especially in the small clique regime, where the sample complexity is sublinear in  $(n)$  for certain values of  $\rho$  and  $\epsilon$ .

## A. Testing Cliques

The problem of testing the  $\rho$ -CLIQUE property in graphs, which focuses on determining whether the graph contains a large clique on  $\rho n$  vertices. The problem can be analyzed in two regime - Large Clique Regime where  $\rho$  is a constant, meaning the size of the clique is a fixed proportion of the graph's vertices. Small Clique Regime where  $\rho$  is a function of the graph size  $(n)$ , typically when  $\rho$  is small relative to  $n$ .

Here the sample complexity refers to how many vertices need to be sampled from a graph to test if it has a  $\rho n$ -clique or not. Goldreich, Goldwasser, and Ron [1] were the first to consider the question and showed that  $S_{\rho\text{-CLIQUE}}(n, \epsilon) = \tilde{O}(\rho/\epsilon^4)$ . In the large clique regime, it shows that  $\epsilon$ -testing  $\rho$ -clique property is sublinear in  $n$  for all  $\rho = \omega(n^{-1/7})$  when  $\epsilon = \Omega(\rho^2)$  - i.e., when the tester must distinguish graphs with  $\rho n$ -cliques from those whose  $\rho n$ -subgraphs all have constant density bounded by 1. Then there is an improved upperbound of  $S_{\rho\text{-CLIQUE}}(n, \epsilon) = \tilde{O}(\rho^4/\epsilon^3)$  provided by Feige, Langberg, and Schechtman [2].

Additionally, they established a lower bound of  $S_{\rho\text{-CLIQUE}}(n, \epsilon) = \tilde{\Omega}(\rho^3/\epsilon^2)$ , meaning any testing algorithm will require at least this many samples

to work efficiently.

The authors present a new upper bound on the sample complexity of testing the  $\rho$ -CLIQUE property that nearly matches the previously established lower bound, up to polylogarithmic factors. This implies that their result is close to the optimal solution for this problem. The results are significant in property testing, where the goal is to determine graph properties by examining a small sample, rather than the entire graph.

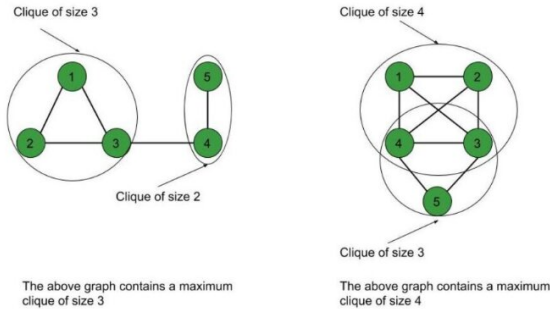


Figure 1: Graph Clique  
[3]

**Theorem 1.** *Sample Complexity of the  $\rho$ -CLIQUE Property is  $S_{\rho\text{-CLIQUE}}(n, \epsilon) = \tilde{O}(\frac{\rho^3}{\epsilon^2})^2$ .* This theorem establishes the sample complexity for testing the  $\rho$ -CLIQUE property in a graph. Specifically, it shows that Sample Complexity  $S_{\rho\text{-CLIQUE}}(n, \epsilon) = \tilde{\Omega}(\rho^3/\epsilon^2)$ . This result implies that you can test whether a graph contains a large clique or is far from having a large clique by inspecting a relatively small number of vertices. The bound is near-optimal, matching known lower bounds up to polylogarithmic factors.

**Theorem (1').** [Alternative Formulation] For any  $n, k = k(n) < n$ , and  $\delta := \delta(n) > 0$  a randomized algorithm can distinguish between Graphs containing  $k$ -cliques and Graphs where every  $k$ -subgraph contains at most  $(1 - \delta)k^2$  edges. Here we only need to inspect a random subset of  $|S| = O(n/(\delta^2 k) \ln^3(n/(\delta^2 k)))$  vertices. This formulation shows that if  $k$  grows faster than  $\ln^3 n$ , you can distinguish between these two types of graphs by examining a sublinear portion of the graph. The result generalizes the previous findings Previous bounds showed sublinear sample complexity

for specific values of  $k$  by Feige, Langberg, and Schechtman [2]. The results build on these earlier works related to Planted  $k$ -Clique Detection, showing similar sample complexity bounds for the Densest  $k$ -Subgraph problem given by Racz and Schiffer [4] and Huleihel, Mazumdar, and Pal [5]. For algorithms that can query any vertex pair (as opposed to sampling vertices), Theorem 1' also provides a nearly optimal upper bound on the number of queries needed  $O(n^2/(\delta^4 k^2) \ln^6(n/(\delta^2 k)))$ .

## B. Testing Colorability

The discussion focuses on the problem of testing the  $k$ -COLORABLE property for graphs, where a graph is  $k$ -colorable if its vertices can be colored with  $k$  colors such that no two adjacent vertices share the same color. For  $k = 2$ , which corresponds to testing bipartiteness, nearly optimal bounds  $S_{2\text{-COLORABLE}}(n, \epsilon) = \Theta(1/\epsilon)$  [1], [6] are known. For  $k \geq 3$ , there is low chromatic number regime. When  $k$  is constant, Rodl and Duke [7] demonstrated that  $S_{k\text{-COLORABLE}}(n, \epsilon)$  is constant when both  $k$  and  $\epsilon$  are constants. However, this result is limited to very large values of  $k$  when  $\epsilon$  is small. For  $k$  as a function of  $n$ , Goldreich, Goldwasser, and Ron (1998) [1] provided that  $S_{k\text{-COLORABLE}}(n, \epsilon) = \tilde{O}(k^2/\epsilon^3)$  for all  $k \geq 3$ . This shows  $k$ -colorability is testable with sublinear complexity for small  $\epsilon$  when  $k = o(n^{1/5})$ . Alon and Krivelevich (2002) [6] improved the bound to  $S_{k\text{-COLORABLE}}(n, \epsilon) = \tilde{O}(k/\epsilon^3)$ , showing the  $k$ -colorability is testable with sublinear complexity for all  $k = o(n^{1/3})$ . Sohler (2012) [8] provided that  $S_{k\text{-COLORABLE}}(n, \epsilon) = \tilde{O}(k^6/\epsilon)$  for low chromatic numbers, where the bound is optimal when  $k$  is constant. In the polychromatic regime, the bounds of Alon and Krivelevich are stronger for large  $k$  and small  $\epsilon$ .

**Theorem 2.** *The  $k$ -COLORABLE property has sample complexity  $S_{k\text{-COLORABLE}}(n, \epsilon) = \tilde{O}(k/\epsilon)$ .* This theorem states that the sample complexity for testing  $k$ -COLORABLE property is  $S_{k\text{-COLORABLE}}(n, \epsilon) = \tilde{O}(k/\epsilon)$ . This result unifies and improves upon previous results, demonstrating that sublinear sample complexity is sufficient to distinguish  $k$ -colorable graphs

from graphs where any  $k$ -coloring results in at least  $\delta/k$  monochromatic edges for  $k = o(\sqrt{n})$  and constant  $\delta > 0$ . The lower bound for testing  $k$ -colorability in the polychromatic regime is  $\omega(1/\epsilon)$  [AK02]. This indicates a quadratic gap between the upper bound provided by Theorem 2 and the best-known lower bound for large  $k$ .

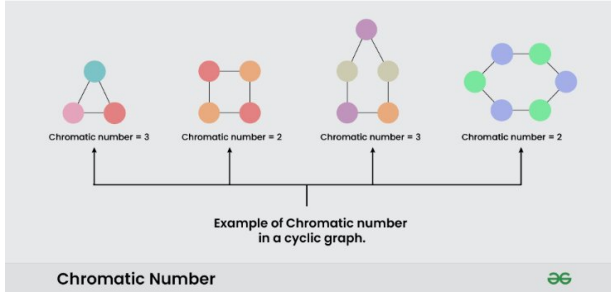


Figure 2: Graph Color [9]

### C. The Graph Container Method

The text describes the use of the graph container method to address challenges in property testing, specifically for testing  $\rho$ -cliques and  $k$ -colorability. The method involves finding a smaller collection of "containers" for independent sets within a graph. Each independent set is a subset of at least one container, and each container is small and sparse. The goal is to simplify the analysis of property testing by focusing on these containers.

#### Testing $\rho$ -independence

The method is applied to test  $\rho$ -independent sets, which is equivalent to testing  $\rho$ -cliques in the complement graph. Key technical lemmas provide tight bounds on the size of containers based on the size of their fingerprints. The probability that a sample contains a large independent set is bounded using these containers and their fingerprints.

#### Testing $k$ -colorability

Similar to  $\rho$ -independence, the method is applied to test  $k$ -colorability, but a different argument is used. The text hints at the application

of the hypergraph container method for partition properties of hypergraphs, which might offer new insights for property testing in other domains.

It is discussed how the query complexity of adaptive testers (which can query any edge) compares to the canonical tester (which samples vertices and queries edges within the induced subgraph). For large cliques, Theorem 1 indicates that the number of edge queries for the canonical tester is close to the query complexity of the best adaptive tester when  $\epsilon = \Omega(\rho^2)$ . For  $k$ -colorability, the open question is whether adaptivity can improve query complexity, particularly for  $\epsilon$  smaller than  $\epsilon = \Omega(\rho^2)$ .

Theorem 1' provides a bound on the sample complexity of the Densest  $k$ -Subgraph problem, which also implies results on time complexity. For distinguishing graphs with  $k$ -cliques from those with  $k$ -subgraphs of density at most  $(1 - \delta)$ , sampling  $s = O(n^{\delta^2 k} \ln^3(n^{\delta^2 k}))$  vertices suffices. Checking whether the induced subgraph contains a clique of size  $O(\delta^{-2} \ln^3(n^{\delta^2 k}))$  results in quasipolynomial time complexity  $n^{O(\ln^3 n)}$ .

### D. Discussion

The discussion highlights several open problems related to the graph container method and its application to property testing:

#### 1. Property Testing with the Container Method:

The method is effective for testing cliques and colorability and may extend to other 0-1 graph partition properties [10]. Future work could explore its application to more general graph properties and hypergraph partition properties.

#### 2. Query Complexity:

The relationship between adaptive testers querying any edge and canonical testers sampling vertices is still open. Theorem 1 suggests that the query complexity for testing large cliques is close to the best adaptive testing algorithm when  $\epsilon = \Omega(\rho^2)$ . Questions remain about whether adaptivity can reduce queries further, especially for testing bipartiteness with  $O(n^2)$  queries [11].

### 3. Time Complexity:

Theorem 1' provides bounds for the Densest  $k$ -Subgraph problem, showing that distinguishing graphs with  $k$ -cliques from those with lower density subgraphs can be done with sampling  $s = O(n^{\delta^2 k} \ln^3(n^{\delta^2 k}) / \delta^2 k)$  vertices and checking for cliques of size  $O(\delta^{-2} \ln^3(n^{\delta^2 k} / n))$ , resulting in quasipolynomial time complexity  $n^{O(\ln^3 n)}$ . Improvements in time complexity for property testing could be achieved using the container method.

## II. The Container Method

The container method involves a process to study independent sets in graphs through a specific greedy algorithm, which is a variant of the approach initially proposed by Kleitman and Winston [12]. This method uses fingerprints and containers to manage and analyze independent sets.

---

### Algorithm 1: FINGERPRINT & CONTAINER GENERATOR

---

**Input:** A graph  $G = (V, E)$  and an independent set  $I \subseteq V$

- 1 Initialize  $F_0 \leftarrow \emptyset$  and  $C_0 \leftarrow V$ ;
- 2 **for**  $t = 1, 2, \dots, |I|$  **do**
- 3      $v_t \leftarrow$  the vertex in  $I \setminus F_{t-1}$  with the largest degree in  $G[C_{t-1}]$ ;  
       // Add  $v_t$  to the fingerprint
- 4      $F_t \leftarrow F_{t-1} \cup \{v_t\}$ ;  
       // Remove all the neighbors of  $v_t$  from the container
- 5      $C_t \leftarrow C_{t-1} \setminus \{w \in C_{t-1} : (v, w) \in E\}$ ;  
       // And remove all vertices with higher degree than  $v_t$  in  $G[C_{t-1}]$
- 6      $C_t \leftarrow C_t \setminus \{w \in C_{t-1} : \deg_{G[C_{t-1}]}(w) > \deg_{G[C_{t-1}]}(v_t)\}$ ;
- 7 **return**  $F_1, \dots, F_{|I|}$  and  $C_1, \dots, C_{|I|}$ ;

---

### A. Basic Properties of Container

#### 1. Containment Relation:

$$F_1 \subseteq F_2 \subseteq F_3 \subseteq F_4 \subseteq \dots \subseteq F_{|I|} = I \subseteq C_{|I|} \subseteq C_{|I|-1} \subseteq \dots \subseteq C_2 \subseteq C_1 \quad (1)$$

Each fingerprint is a subset of the independent set  $I$ , and each container is a superset.

### 2. Container Consistency:

There is a proposition about this topic.

**Proposition 3.** For any  $t$ ,

$$C_T(F_t(I)) = C_t(I)$$

This means that the container for  $F_t(I)$  is the same as the container for  $I$  at the same stage.

### 3. Degree Bound:

There is also a proposition about the maximum degree bound about the graph container method.

**Proposition 4.** The maximum degree in  $G[C_t]$  is bounded by

$\Delta(G[C_t]) \leq (n/t)$ . This follows from the fact that vertices are excluded based on their degrees, ensuring that no vertex in  $C_t$  has a degree exceeding  $n/t$  in the included subgraph. This method and its properties are valuable in understanding the structure of independent sets and can be applied to various problems in graph theory and combinatorics.

### B. Graph Container Lemmas

In the case of testing the  $\rho$ -CLIQUE property, we first change our focus so that we instead look at graphs that are  $\epsilon$ -far from having the  $\rho$ -INDEPENDENT SET property. We show that for each such graph  $G$ , every independent set in  $G$  is a subset of a container of size strictly less than  $\rho n$ . In fact, our result shows more: there is a container for  $I$  of size  $(\rho - \alpha)n$  where  $\alpha$  is at least as large as  $\epsilon/\rho$  times the length of the corresponding fingerprint (up to log factors).

**Lemma 5** (*Graph Container Lemma I*). For a graph  $G = (V, E)$  that is  $\epsilon$ -far from the  $\rho$ -INDEPSET property, every independent set  $I$  is a subset of a container  $C_t$ , where size of the container bounded as :

$$|C_t| \leq (\rho - t \cdot \frac{\epsilon}{8\rho \ln(2\rho)})n \quad (2)$$

Additionally, the induced subgraph  $G[C_t]$  contains at most  $\epsilon n^2$  edges.

**Lemma 6** (*Graph Container Lemma II*). For graphs that are  $\epsilon$ -far from being  $k$ -colorable, if  $I_1, I_2, \dots, I_K$  are independent sets in  $G$ , then there exists an index  $t \leq 4$  such that: The lemmas focus on bounding the size of containers that cover independent sets and the trade-off between container size and the corresponding fingerprints.

$$\bigcup_{i_1}^k C_t(I_i) \leq (1 - t \cdot \frac{\epsilon}{4 \ln(1/\epsilon)})n \quad (3)$$

Additionally, a Container Shrinking Lemma is introduced, showing that many vertices are excluded in each round of a greedy algorithm when containers contain a large sparse subgraph.

### III. TESTING CLIQUES AND INDEPENDENT SETS

We complete the proof of Theorem 1 in this section. The proof we describe applies to both the  $\rho$ -CLIQUE and the  $\rho$ -INDEPSET properties (by considering the complement of the tested graph) but the latter fits in more naturally with the graph container method, so in the rest of the section we discuss the result entirely in terms of independent sets.

#### A. Container Shrinking Lemma

**Lemma 7** (*Container Shrinking Lemma*). Let  $G = (V, E)$  be a graph on  $n$  vertices that is  $\epsilon$ -far from the  $\rho$ -INDEPSET property. Let  $I$  be an independent set in  $G$ , and let  $t < |I|$  be an index for which the  $t$ -th container  $C_t$  of  $I$  has cardinality  $|C_t| \geq \rho n$ . For any  $\alpha > 0$ , if there exists a subset  $C \subseteq C_{t+1}$  of  $(\rho - \alpha)n$  vertices where  $G[C]$  contains at most  $\frac{\epsilon}{4}n^2$  edges, then:

$$|C_{t+1} \setminus C| \leq \left(1 - \frac{\epsilon}{4\rho\alpha}\right) |C_t \setminus C|$$

We can now complete the proof of the claim that  $G[C_t \setminus C]$  contains at least  $\frac{\epsilon}{4\rho\alpha} |C_t \setminus C|^2$  edges. Let  $R$  be chosen uniformly at random from all subsets of  $C_t \setminus C$  of size  $\alpha n$ . Since  $G$  is  $\epsilon$ -far from having independent sets of size  $\rho n$ , the induced subgraph  $G[R \cup C]$  contains at least  $\epsilon n^2$  edges. By the lemma hypothesis,  $G[C]$  contains at most

$\frac{\epsilon}{4}n^2$  edges. Finally, each edge connecting a vertex in  $C$  to one in  $C_t \setminus C$  is included in  $G[R \cup C]$  with probability  $\frac{|R|}{|C_t \setminus C|}$ . So by the maximum degree of the vertices in  $C$  in the graph  $G[C_t]$ , the expected number of edges between  $C$  and  $R$  is at most  $\frac{\epsilon}{4\alpha} |R|n = \frac{\epsilon}{4}n^2$ . Therefore, the expected number of edges in  $G[R]$  is at least:

$$\frac{\epsilon}{2}n^2 = \frac{\epsilon}{2\alpha^2} |R|^2 \geq \frac{\epsilon}{\alpha^2} \binom{|R|}{2}$$

#### B. . Proof of Graph Container Lemma I

**Lemma 5** (*Graph Container Lemma I*)

*Proof.* Let  $C$  denote the first container in the sequence  $C_1, C_2, \dots$  that contains at most  $\frac{\epsilon}{4}n^2$  edges. The existence of  $C$  is guaranteed because  $C_{|I|+1}$  is an independent set and has no edges. Define  $\alpha$  such that  $|C_{|I|+1}| = (p - \alpha)n$ . The value of  $\alpha$  is bounded by  $\frac{\epsilon}{4p} \leq \alpha \leq p$ , where the lower bound on  $\alpha$  is due to the fact that  $G$  is  $\epsilon$ -far from the  $\rho$ -INDEPSET property. Define  $t^*$  to be the largest index for which  $|C_t| \geq \rho n$ . First observe that  $t^* \leq |I|$  because otherwise  $C_{t^*}$  is an independent set, which contradicts the fact that  $G$  is  $\epsilon$ -far from the  $\rho$ -INDEPSET property. Hence, by Lemma 7 applied to all values of  $t = 0, 1, 2, \dots, t^* - 1$ ,  $|C_{t^*} \setminus C| \leq (1 - \frac{\epsilon}{4\rho\alpha})^{t^*} n$  and since  $|C_{t^*} \setminus C| \geq \alpha n$ , we conclude that  $t^* \leq \frac{4\rho\alpha}{\epsilon} \ln(1/\alpha) \leq \frac{4\rho\alpha}{\epsilon} \ln(2\rho/\epsilon)$ .  $\square$

#### C. Proof of Theorem 1

**Lemma 8** (Chernoff's Bound). Let  $X$  be drawn from the hypergeometric distribution  $H(N, K, n)$  where  $n$  elements are drawn without replacement from a population of size  $N$  where  $K$  elements are marked and  $X$  represents the number of marked elements that were drawn. Then for any  $\theta \geq E[X]$ ,  $\Pr[X \geq \theta] \leq \exp\left(-\frac{(\theta - E[X])^2}{\theta + E[X]}\right)$ .

**Theorem 1** (*Precise formulation*).

The sample complexity of the  $\rho$ -INDEPSET property is

$$\mathcal{S}_{\rho\text{-INDEPSET}}(n, \epsilon) = O\left(\frac{\rho^3}{\epsilon^2} \ln^3\left(\frac{1}{\epsilon}\right)\right)$$

*Proof.* Let  $S$  be a random set of  $s = c \cdot \frac{\rho^3}{\epsilon^2} \ln^3(\frac{1}{\epsilon})$  vertices drawn uniformly at random from  $V$  without replacement, where  $c$  is a large enough constant. Note that for clarity of presentation, in the

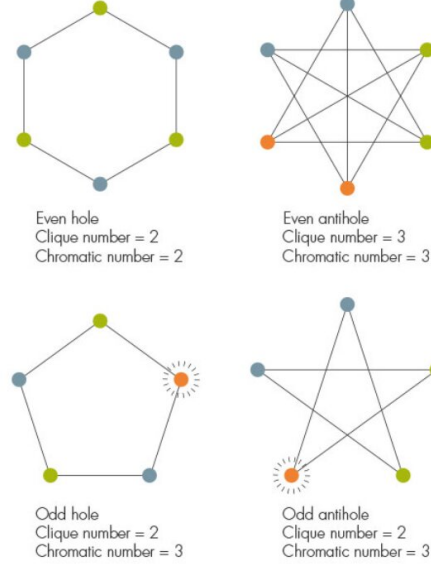


Figure 3: Graph color and clique  
[13]

rest of the proof we ignore all integer rounding issues as they do not affect the asymptotics of the final result. If  $G$  contains a  $\rho n$  independent set, then  $S$  contains at least  $\rho s$  vertices from this independent set with probability at least  $\frac{1}{2}$ , since the number of such vertices follows a hypergeometric distribution, and the median of this distribution is at least  $\rho s$  [14][15]. In the remainder of the proof we upper bound the probability that  $S$  contains a  $\rho s$ -independent set when  $G$  is  $\epsilon$ -far from containing a  $\rho n$ -independent set. For the rest of the argument, let us call a container  $C_t$  small when its cardinality is bounded by  $|C_t| \leq (\rho - t \cdot \frac{\epsilon}{8\rho \ln(2\rho/\epsilon)})n$  the expression (2) in the conclusion of Lemma 5. For the better understanding of the proof of the theorem. The related mathematical expressions with  $i_1, i_2, i_3, \dots \in [s]$  :

$$\begin{aligned}
E[X] &\leq \left( \rho - \frac{t\epsilon}{8\rho \ln\left(\frac{2\rho}{\epsilon}\right)} \right) (s - t) \\
&< \rho s - \frac{t\epsilon s}{8\rho \ln\left(\frac{2\rho}{\epsilon}\right)} \\
&\leq \rho s - t - \frac{t\epsilon s}{16\rho \ln\left(\frac{2\rho}{\epsilon}\right)},
\end{aligned}$$

$$\begin{aligned}
Pr[X \geq \rho s - t] &\leq \exp\left(-\frac{(\rho s - t - E[X])^2}{\rho s - t + E[X]}\right) \\
&\leq \exp\left(-\frac{t^2 \epsilon^2 s}{512\rho^3 \ln^2\left(\frac{2\rho}{\epsilon}\right)}\right).
\end{aligned}$$

Therefore, by applying a union bound over all values  $t \leq \frac{8\rho^2 \ln(\frac{2\rho}{\epsilon})}{\epsilon}$  and all possible choices of  $t$  indices from  $[s]$ ...

$$\begin{aligned}
&\sum_t \binom{s}{t} \exp\left(-\frac{t^2 \epsilon^2 s}{512\rho^3 \ln^2\left(\frac{2\rho}{\epsilon}\right)}\right) \\
&\leq \sum_t \exp(t \ln s - \frac{t^2 \epsilon^2 s}{512\rho^3 \ln^2\left(\frac{2\rho}{\epsilon}\right)}) < \frac{1}{3}
\end{aligned}$$

□

#### IV. TESTING K-COLORABILITY

##### A. Proof of Graph Container Lemma II

**Lemma 6**(Graph Container Lemma II). Let  $G$  be  $\epsilon$ -far from  $k$ -colorable, and let  $I_1, \dots, I_k$  be independent sets in  $G$ . Then, there exists  $t \leq \frac{4}{\epsilon}$  such that

$$\left| \bigcup_{i=1}^k C_t(I_i) \right| \leq \left( 1 - t \cdot \frac{\epsilon}{4 \ln\left(\frac{1}{\epsilon}\right)} \right) n. \quad (3)$$

Proof. Assume on the contrary that for all  $t \leq 4/\epsilon$ ,

$$\left| \bigcup_{i=1}^k C_t(I_i) \right| > \left( 1 - t \cdot \frac{\epsilon}{4 \ln(\frac{1}{\epsilon})} \right) n. \quad (4)$$

$\forall t = 1, \dots, \frac{4}{\epsilon}$ , Let  $A_t$  be the set of vertices that are contained in at least one container at the  $(t-1)$ -th step, but not contained in any container at step  $t$  of the container procedure. In other words, let

$$A_t = \{v \in V : v \in \bigcup_{i \in [k]} C_{t-1}(I_i) \text{ and } v \notin \bigcup_{i \in [k]} C_t(I_i)\}.$$

The sum of the all edges in each of the sets  $V_1, \dots, V_k$  can be upper bounded by

$$\frac{\epsilon n^2}{4} + |A_1| \cdot n + \sum_{t=2}^{4/\epsilon} |A_t| \cdot \frac{n}{t-1} \quad (5)$$

Again the sum of the all edges in each of the sets  $V_1, \dots, V_k$  can be upper bounded by

$$\frac{\epsilon n^2}{4} + \frac{\epsilon n^2}{4 \ln(1/\epsilon)} + \frac{\epsilon n^2}{4 \ln(1/\epsilon)} \sum_{t=2}^{4/\epsilon} \frac{1}{t-1} < \epsilon n^2,$$

### B. Proof of Theorem 2

**Theorem 2** (Precise formulation). The sample complexity of the  $k$ -COLORABLE property is

$$S_{k\text{-COLORABLE}}(n, \epsilon) = O\left(\frac{k}{\epsilon} \ln^2\left(\frac{1}{\epsilon}\right)\right).$$

*Proof.*  $\forall t$ , there are at most  $\binom{s}{k} \leq s^{tk}$  ways to choose  $k$  sets of at most  $t$  distinct indices as above. So by applying a union bound argument over all  $t \leq \frac{4}{\epsilon}$  and over all possible  $4k$  choices of indices, the probability that any set of at most  $\frac{4k}{\epsilon}$  vertices in  $S$  form the  $k$  fingerprints of a family of containers with a small union from which we sample all the remaining vertices is at most

$$\sum_{t=1}^{4/\epsilon} s^{tk} \left(1 - \frac{t\epsilon}{4 \ln(1/\epsilon)}\right)^{s-tk} \\ \leq \sum_{t=1}^{4/\epsilon} \exp\left(tk \ln(s) - \frac{t\epsilon s}{8 \ln(1/\epsilon)}\right) < \frac{1}{3}$$

□

In conclusion, the container method provides a powerful framework for efficiently testing graph properties, particularly for determining whether a graph is far from containing large independent sets. By leveraging the shrinking of containers and probabilistic tools like Chernoff bounds, this method reduces the computational complexity involved in property testing. The approach ensures that only a small sample of vertices needs to be considered while maintaining high accuracy. Furthermore, the method's versatility makes it applicable to a range of other combinatorial optimization problems, offering a promising direction for future research in graph theory and algorithms.

## References

- [1] O. Goldreich, S. Goldwasser, and D. Ron, "Property testing and its connection to learning and approximation," in *Proceedings of the 37th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 339–348, IEEE Computer Society, 1998.
- [2] U. Feige, O. Langberg, and S. Schechtman, "Testing the chromatic number," in *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 652–661, ACM, 2004.
- [3] GeeksforGeeks, "Graph color." <https://images.app.goo.gl/CEXh41qL11Yam2KJ9>, 2024. Accessed: 2024-09-16.
- [4] G. Racz and A. Schiffer, "The complexity of testing the planted clique problem," *SIAM Journal on Computing*, vol. 48, no. 3, pp. 1–24, 2019.
- [5] T. Huleihel, R. Mazumdar, and A. Pal, "On the sample complexity of testing the planted clique," *Journal of Machine Learning Research*, vol. 22, no. 55, pp. 1–28, 2021.
- [6] N. Alon and I. Krivelevich, "Testing  $k$ -colorability," *Random Structures Algorithms*, vol. 21, no. 3, pp. 330–346, 2002.

- [7] V. Rödl and D. Duke, “On the sample complexity of testing  $k$ -colorability,” *Combinatorica*, vol. 5, no. 2, pp. 143–155, 1985.
- [8] C. Sohler, “Testing  $k$ -colorability: The case of low chromatic numbers,” *Theoretical Computer Science*, vol. 450, pp. 113–122, 2012.
- [9] GeeksforGeeks, “Graph clique.” <https://images.app.goo.gl/EhLqjbww1f2cNwjt7>, 2024. Accessed: 2024-09-16.
- [10] O. Nakar and D. Ron, “Testing graph properties with a fixed number of queries,” *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2500–2518, 2018.
- [11] A. Bogdanov and W. Li, “Optimal bounds for property testing of graphs,” *Journal of Computer and System Sciences*, vol. 76, no. 5, pp. 387–401, 2010.
- [12] D. J. Kleitman and K. J. Winston, “On the number of graphs without 4-cycles,” *Discrete Mathematics*, vol. 41, no. 2, pp. 167–172, 1982.
- [13] Quanta Magazine, “A quantum computer that breaks with conventional wisdom.” <https://www.quantamagazine.org/wp-content/uploads/2015/10/011.jpg>, 2015. Accessed: 2024-09-17.
- [14] A. Neu, *Title of the Book*. Publisher Name, 1970.
- [15] A. KB, “Title of the article,” *Journal Name*, vol. 10, no. 2, pp. 100–120, 1980.