Testing graph isomorphism

Sotnikov Dmitry

Sub linear Algorithms Seminar 2008
The problem definition

- Given two graphs G, H on n vertices distinguish the case that they are isomorphic from the case that they are not isomorphic is very hard.
  - The best algorithm is known today to solve the problem has run time $2^{O(\sqrt{n \log n})}$ for graphs with n vertices.
  - And almost the subgraph isomorphism problem is NP complete.

- So how can we do something in sub linear time that we don’t know if any polynomial algorithm is exist?
The problem definition

- So if we can not to solve very hard problem we try to solve another problem that is easier but have correlation with original problem.
The problem definition

- **Won’t** talk about the run time
- **Will** talk about the number of queries that is necessary to distinguish between the two cases: graphs G and H are isomorphic and they are not.
The gap

- We allow gap: we want that the algorithm give us the correct answer if the graphs G and H are isomorphic or when they are $\varepsilon$-far from being isomorphic, and we don’t care what algorithm says in another cases.
The problem

- How many queries are required to distinguish between the case that two graphs $G$ and $H$ on $n$ vertices are isomorphic, and the case that they are $\varepsilon$-far.
The probability algorithm

And the second facilitation is that we allow some inaccuracy, we allow the algorithm to give the right answer with probability >2/3 instead of always saying the correct answer.
Remainder

- And now we can use something that we see in the beginning of the seminar: Property Testing.
Combinatorial property testing

For a fixed \( \varepsilon > 0 \) and a fixed property \( P \), distinguish using as few queries as possible (and with probability at least 2/3) between the case that an input of length \( m \) satisfies \( P \), and the case that the input is \( \varepsilon \)-far (with respect to an appropriate metric) from satisfying \( P \).
Notation and preliminaries

- All graphs considered here are undirected and with neither loops nor parallel edges.
- We denote by $[n]$ the set $\{0, 1, \ldots, n-1\}$.
- For a vertex $v$, $N(v)$ denotes the set of $v$'s neighbors.
Notation and preliminaries

- Given a permutation $\sigma : [n] \to [n]$, and a subset $U$ of $[n]$, we denote by $\sigma(U)$ the set $\{\sigma(i) : i \in U\}$.
- Given a subset $U$ of the graph $G$, we denote by $G(U)$ the induced subgraph of $G$ on $U$. 

Notation and preliminaries

- We denote by $G(n,p)$ the random graph where each pair of vertices forms an edge with probability $p$, independently of each other.
The property definition

Graph Isomorphism

Graph $G$ and $H$ are isomorphic if there exists a permutation $\pi : V_G \rightarrow V_H$, such that it preserves the edge relationship, i.e.

$$(x, y) \in E_G \iff (\pi(x), \pi(y)) \in E_H$$
Graph isomorphism

\[ \pi(1) = 1 \quad \pi(2) = 3 \]
\[ \pi(3) = 2 \quad \pi(4) = 4 \]
Graph isomorphism

$G$ and $H$ are not isomorphic because the degree of vertex 1 in $H$ is 3 and we don’t have any vertex of same degree in $G$. 
The permutation distance definition

For permutation \( \pi : V_G \to V_H \), the distance between \( G \) and \( H \) under \( \pi \), \( d_\pi \) is symmetric distance between the adjacent matrices of \( \pi(G) \) and \( H \) divided by \( \binom{n}{2} \).

The graphs distance definition

The distance between \( G \) and \( H \) is:

\[
d(G, H) = \min_{\pi} d_\pi (G, H)
\]
The distance

\[ n = 4 \text{ so } \binom{n}{2} = 6 \]

\[
\begin{align*}
\pi(1) &= 1 & \pi(2) &= 3 \\
\pi(3) &= 2 & \pi(4) &= 4 \\
d_\pi(G,H) &= 0
\end{align*}
\]

\[
\begin{align*}
\sigma(1) &= 1 & \sigma(2) &= 2 \\
\sigma(3) &= 3 & \sigma(4) &= 4 \\
d_\sigma(G,H) &= 4/6
\end{align*}
\]

\[ d(G,H) = 0 \]
Input definition

- Formally, our inputs are two functions 
  \[ g: \{1,2,\ldots,\binom{n}{2}\} \rightarrow \{1,0\} \quad \text{and} \quad h: \{1,2,\ldots,\binom{n}{2}\} \rightarrow \{1,0\} \]
  which represents the edge sets of two corresponding graphs \( G \) and \( H \) over vertex set \( V=\{1,\ldots,n\} \).

- We assumes that \( G \) and \( H \) are “dense” graphs.

- A query consisting of finding whether two vertices \( u,v \) of \( G \) (or \( H \)) form an edge of \( G \) (or \( H \)) or not.
The questions we want to answer

1. Given two input graphs $G$ and $H$, how many queries to $G$ and $H$ are required to test that the two graphs are isomorphic?

2. Given a graph $G$, which is known in advance (and for which any amount of preprocessing is allowed), and an input graph $H$, how many queries to $H$ are required to test that $H$ is isomorphic to $G$?
## Summary of the results

<table>
<thead>
<tr>
<th></th>
<th>Upper bound</th>
<th>Lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>One sided error, one graph known</td>
<td>$\Theta(n)$</td>
<td>$\Omega(n)$</td>
</tr>
<tr>
<td>One sided error, both graphs unknown</td>
<td>$\Theta(n^{3/2})$</td>
<td>$\Omega(n^{3/2})$</td>
</tr>
<tr>
<td>Two sided error, one graph known</td>
<td>$\Theta(n^{1/2})$</td>
<td>$\Omega(n^{1/2})$</td>
</tr>
<tr>
<td>Two sided error, both graphs unknown</td>
<td>$\Theta(n^{5/4})$</td>
<td>$\Omega(n)$</td>
</tr>
</tbody>
</table>
Knowledge chart

- Given a query set $Q$ of the adjacency matrix $A$ of the graph $G=(V,E)$ on $n$ vertices, we define the knowledge chart $I(G,Q)$ of $G$ as the subgraph of $G$ known after making the set $Q$ of queries to $A$. 
Knowledge chart

For a fixed $q$, $0 \leq q \leq \binom{n}{2}$, and G, we define $I(G,q)$ as the set of knowledge charts $\{I(G,Q) : |Q| = q\}$. For example, note that $|I(G,0)| = |I(G,\binom{n}{2})| = 1$. 
Knowledge chart

\[ G \]

\[ Q = \{ (11,12), (11,8), (11,5), (12,8), (12,2), (2,3), (2,8), (2,6), (3,6), (3,7), (3,5), (5,6), (6,7), (7,8) \} \]
Knowledge packability (Consistence)

- A knowledge-packing of two knowledge chart $I(G, Q(G)), I(H, Q(H))$ where $G$ and $H$ are graphs with $n$ vertices, is a bijection $\pi$ of the vertices of $G$ into vertices of $H$ such that for all $u, v \in V(G)$, if $(v, u) \in E(G) \cap Q$ then $(\pi(v), \pi(u)) \notin Q \setminus E(H)$, and if $(v, u) \in Q \setminus E(G)$ then $(\pi(v), \pi(u)) \notin E(H) \cap Q$
Knowledge packability

$Q=\{(1,2),(2,3),(3,6)\}$
$I(G,Q)$ and $I(H,Q)$ are knowledge packable

$K=\{(1,2),(2,3),(4,5)\}$
$I(G,K)$ and $I(H,K)$ are knowledge not packable
One-sided error

A property testing algorithm has one-sided error probability if it accepts inputs that satisfy the property with probability 1. We also call such testers one-sided error testers.
Lemma 1

Any one-side error isomorphism tester, after completing its queries \( Q(G), Q(H) \), must always accept \( G \) and \( H \) if the corresponding knowledge charts \( I(G, Q(G)) \), \( I(H, Q(H)) \) on which the decision is based are knowledge-packable. In particular, if some \( G, H \) and \( 0 \leq q \leq \binom{n}{2} \), any \( I(G, Q(G)) \in I(G, q) \) and \( I(H, Q(H)) \in I(H, q) \) are knowledge-packable, then every one-side error isomorphism tester which is allowed to ask at most \( q \) questions must always accept \( G \) and \( H \).
Lemma’s proof

This is true, since if the knowledge charts $I(G,Q(G))$ and $I(H,Q(H))$ are packable, it means that there is an extension $G'$ of $G$’s restriction to $Q(G)$ to a graph that is isomorphic to $H$. In other words, given $G'$ and $H$ as inputs, there is a positive probability that the isomorphism testers obtained $I(G',Q(G)) = I(G,Q(G))$ and $I(H,Q(H))$ after completing its queries, and hence, a one-sided error tester must always accept in this case.
One-sided Tester

- Usually, one-sided testers look at some query set \( q \) of the input, and accept iff the restriction of the input to \( Q \) is extensible to some input satisfying the property.

- The main idea is to prove that if the input is far from satisfying the property, then with high probability its restriction \( Q \) will provide the evidence for it.
One-side testing of two unknown graphs

In this section we prove the following:

If both graphs are unknown the query complexity of one-side isomorphism testers is $\Theta(n^{3/2})$ (up to coefficient depending only on the distance parameter $\varepsilon$)
The upper bound

Algorithm 1

1. For both graphs $G,H$ construct the query sets $Q(G), Q(H)$ respectively by choosing every possible query with probability $\frac{\ln n}{\epsilon n}$, independently of other queries.

2. If $|Q(G)|$ or $|Q(H)|$ is larger than $1000n^{3/2} \sqrt{\frac{\ln n}{\epsilon}}$ accept without making the queries. Otherwise make the chosen queries.

3. If there is a knowledge-packing of $I(G, Q(G))$ and $I(H, Q(H))$, accept. Otherwise reject.
Query complexity

Clearly, the query complexity of the Algorithm 1 is $O(n^{3/2} \sqrt{\log n})$. 
Correctness of the Algorithm 1

- Now we need to prove that the Algorithm 1 accepts with probability 1 if G and H are isomorphic, and if G and H are e-far from being isomorphic, Algorithm 1 rejects with probability $> \frac{2}{3}$ (we will show something stronger, we will show that probability to reject if G and H are e-far is $1-o(1)$).
Proof

Assume first that $G$ and $H$ are isomorphic, and let $\pi$ be one of the isomorphisms between them. Obviously $\pi$ is also a knowledge-packing for any pair of knowledge charts of $G$ and $H$. Hence, if the algorithm did not accept in the second stage, then it will accept in the third stage.
Proof continue

Now we turn to the case where $G$ and $H$ are $\varepsilon$-far from being isomorphic.

First I want to show that the probability that algorithm terminates in step 2 is $o(1)$.

- We chose the queries to $Q$ i.i.d with probability $\sqrt{\frac{\ln n}{\varepsilon n}}$, so the expectation of the $|Q|$ is:

$$E[|Q|] = n^2 \sqrt{\frac{\ln n}{\varepsilon n}} = n^{3/2} \sqrt{\frac{\ln n}{\varepsilon}}$$

- And now by Marcov bound we get:

$$\Pr[|Q| \geq t \cdot E[|Q|]] < \frac{1}{t} \Rightarrow \Pr[|Q| \geq 1000n^{3/2} \sqrt{\frac{\ln n}{\varepsilon}}] < \frac{1}{1000}$$
Proof continue

- So we can assume without harming the correctness that the algorithm reaches step 3.
- Since $G$ and $H$ are $\varepsilon$-far from being isomorphic, every possible bijection $\pi$ of their vertices has a set $E_\pi$ of at least $\varepsilon n^2$ of the $G$’s vertices such that:

\[(u, v) \in E_\pi \Rightarrow (u, v) \in E(H) \oplus E(G)\]
Proof continue

Let \((u,v)\) be one such pair. The probability that 
\((u,v)\) or \((\pi(u),\pi(v))\) were not queried in some 
graph is \(1 - \frac{\ln n}{\varepsilon n}\).

Using the union bound and the fact that the 
queries where chosen independently, we bound the 
probability of not revealing at least one such pair 
in both graphs for all possible bijections by 
\(n!(1 - \frac{\ln n}{\varepsilon n})^{\varepsilon n^2}\)

This bound satisfies 
\(n!(1 - \frac{\ln n}{\varepsilon n})^{\varepsilon n^2} \leq n! e^{-n\ln n} = n! n^{-n} = o(1)\)
The lower bound

Here we construct a pair $G,H$ of $1/100$-far graphs on $n$ vertices, such that every knowledge chart from $I(G,n^{3/2}/200)$ can be packed with every knowledge chart $I(H,n^{3/2}/200)$ and hence by Lemma 1, any one-sided algorithm which is allowed to use at most $n^{3/2}/200$ queries must always accept $G$ and $H$. 
The lower bound

- Note that this hold for non-adaptive as well as adaptive algorithms, since we actually prove that there is no certificate of size $n^{3/2}/200$ for the non-isomorphism of these graphs.
Lemma 2

For every large enough $n$ there are two graphs $G$ and $H$ on $n$ vertices, such that:
1. $G$ is $1/100$-far from being isomorphic to $H$.
2. Every knowledge chart from $I(G, n^{3/2}/200)$ can be knowledge-packed with any knowledge chart from $I(H, n^{3/2}/200)$.
Lemma’s proof:

- We set both $G$ and $H$ to be the union of a complete bipartite graph with set of isolated vertices.

- Formally, $G$ has three vertex sets $L$, $R_f$, $R_e$, where $|L| = n/2$, $|R_f| = 26n/100$ and $|R_e| = 24n/100$, and it has the following edges: $\{(u,v) : u \in L \land v \in R_f\}$

- $H$ has the same structure, but with $|R_f| = 24n/100$ and $|R_e| = 26n/100$. 
Example of the $G$ and $H$ graphs
Lemma’s proof:

- Let count the edges of the both graphs:
  - $|E(G)| = \frac{n}{2} \times 26\frac{n}{100} = 13n^2 / 100$
  - $|E(H)| = \frac{n}{2} \times 24\frac{n}{100} = 12n^2 / 100$

- G is $1/100$-far from being isomorphic to H, so G and H satisfy the first part of the Lemma 2.
The Lemma’s second condition prove

- To prove the second condition we will show that for all possible query sets \( Q(G), Q(H) \) of size \( n^{3/2}/200 \) there exist sets \( Y_H \in Re(H) \) and \( Y_G \in Rf(G) \) that satisfy following:
  - \( |Y_G| = |Y_H| = n/50 \)
  - The knowledge charts \( I(G,Q(G)) \) and \( I(H,Q(H)) \) restricted to \( L(G) \cup Y_G \) and \( L(H) \cup Y_H \) can be packed in a way that pairs vertices from \( L(G) \) with vertices from \( L(H) \).
Finding $Y_G$ and $Y_H$

**Legend:**
- The queried non-edge
- The queried edge
Remark

Note that there is a trivial algorithm that distinguishes between the two graphs in $O(n)$ queries by sampling vertices and checking their degrees. However, such an algorithm has two-sided error. Any one-sided error algorithm must find evidence to the non-isomorphism of the graphs, i.e. two knowledge charts that cannot be packed (in the sense that there is no isomorphism consistent with them).
Proving the existence of $Y_G$ and $Y_H$

- For all $v \in V(G)$ we define the query degree as $d_Q(v) = |\{(v,u) : u \in V(G) \land (v,u) \in Q(G)\}|$

- We also denote by $N_Q(v)$ and $\overline{N}_Q(v)$ the sets
  
  $N_Q(v) = \{u : (v,u) \in E(G) \cap Q(G)\}$
  
  $\overline{N}_Q(v) = \{u : (v,u) \in Q(G) \setminus E(G)\}$

- Note: $d_Q(v) = |\overline{N}_Q(v)| + |N_Q(v)|$
Proving the existence of $Y_G$ and $Y_H$

Since $|Q(G)|, |Q(H)| \leq n^{3/2}/200$ there must be two sets of vertices $D_G \in Rf(G)$ and $D_H \in Re(H)$ both of size $n/10$, such that:

$$\forall v \in D_G : d_Q(v) \leq n^{1/2}/2$$

$$\forall v \in D_H : d_Q(v) \leq n^{1/2}/2$$
Proving the existence of $Y_G$ and $Y_H$

- We define the arbitrary pairing set
  
  $B_D = \{(v_G^1, u_H^1), (v_G^2, u_H^2), ..., (v_G^{n/10}, u_H^{n/10})\}$

  of $D_G$'s and $D_H$'s elements.

- And we choose a bijection $B_L : L(G) \rightarrow L(H)$ uniformly at random.
Proving the existence of $Y_G$ and $Y_H$

- Now we want to show what with some positive probability, there are at least $n/50$ consistent (packable) pairs in $B_D$.
- We define

$$Y = \{(v_G, u_H) \in B_D : B_L(N_Q(v_G)) \cap \overline{N_Q}(u_H) = \emptyset\}$$

as the set of consistent pairs, and show that

$$\Pr[|Y| \geq n/50] > 0$$
Proving the existence of $Y_G$ and $Y_H$

For a specific pair $\{v \in D_G, u \in D_H\}$

\[
\Pr_{B_v}\left[B_L(N_Q(v)) \cap \overline{N_Q(u)} = \emptyset\right] \geq \prod_{i=0}^{n^{1/2}/2-1} \left(1 - \frac{n^{1/2}/2}{n/2 - i}\right) \geq \\
\geq (1 - \frac{2n^{1/2}}{n})^{n^{1/2}/2} \geq e^{-1} \geq 1/3
\]

\[E[|Y|] \geq |D_G|/3 \geq n/50\]
Proving the existence of $Y_G$ and $Y_H$

\[ Y_G = \{ u : \exists v \in V(H) \text{ such that } (u,v) \in Y \} \]

\[ Y_H = \{ v : \exists u \in V(G) \text{ such that } (u,v) \in Y \} \]
One-sided testing where one of the graphs is known in advance.

In this section we prove the following:

If only one graph is unknown and the second is known in advance the query complexity of one-side isomorphism testers is $\Theta(n)$ (up to coefficient depending only on the distance parameter $\varepsilon$).
The upper bound
Algorithm 2

1. Construct the query set Q respectively by choosing every possible query from H with probability $\frac{\ln n}{\varepsilon n}$, independently of other queries.

2. If $|Q|$ is larger than $\frac{10n\ln n}{\varepsilon}$ accept without making the queries. Otherwise make the chosen queries.

3. If there is a knowledge-packing of $I(G, [V(G)]^2)$ and $I(H, Q)$, accept. Otherwise reject.
Query complexity

- Clearly, the query complexity of the Algorithm 2 is $O(n \log n)$. 
Correctness of the Algorithm 2

- Home work exercise 1: To prove the correctness of the Algorithm 2.
The lower bound

- As before, to give a lower bound on one-side error algorithms it is sufficient to show that for some $G$ and $H$ (where $G$ is in advance known graph) that are far, no “proof” of their non-isomorphism can be provided with $\Omega(n)$ queries.
Lemma 3

If some $G, H$ where $G$ is known in advance, and some fixed $0 \leq q \leq \binom{n}{2}$, $I(G, [V(G)]^2)$ is knowledge-packable with every $I(H,Q) \in I(H,q)$ then every one sided error isomorphism testers which is allowed to ask at most $q$ queries must always accept $G$ and $H$. 
Lemma’s prove

- Home work exercise 2: To prove the Lemma 3.
The lower bound

- Here we construct a pair \(G,H\) of 1/5-far graphs on \(n\) vertices, such that every knowledge chart \(I(H,Q) \in I(H, \frac{n}{4})\) can be packed with knowledge chart \(I(G, [V(G)]^2)\), and hence by Lemma 3, any one-sided algorithm which is allowed to use at most \(n/4\) queries must always accept \(G\) and \(H\).
The graphs structure

- We set $G$ to be a disjoint union of $K_{n/2}$ and $n/2$ isolated vertices, and set $H$ to be a completely edgeless graph.
Example of the $G$ and $H$ graphs

$G$

$H$

$n/2$

$n$

$n/2$

$n/2$
The lower bound proof

- Clearly, just by difference in the edge count, $G$ is $1/5$-far from being isomorphic to $H$. But since $n/4$ queries cannot involve more than $n/2$ vertices from $H$ (all isolated), and $G$ has $n/2$ isolated vertices, the knowledge chart are packable.
Restriction

For a distribution $\text{Dist}$ over inputs, where each inputs is a function $f : D \rightarrow \{0, 1\}$, and for a subset $Q$ of the domain $D$, we define the restriction $\text{Dist}|_Q$ of $\text{Dist}$ to $Q$ to be the distribution over functions of the type $g : Q \rightarrow \{1, 0\}$, that results from choosing a random function $f : D \rightarrow \{0, 1\}$ according to the distribution $\text{Dist}$, and then setting $g$ to be $f|_Q$ the restriction of $f$ to $Q$. 
Variation distance

Given two distributions $D_1$ and $D_2$ of binary function from $Q$, we define the variation distance between $D_1$ and $D_2$ as follows:

$$d(D_1, D_2) = \frac{1}{2} \sum_{g: Q \rightarrow \{0,1\}} |\Pr_{D_1}[g] - \Pr_{D_2}[g]|$$

where $\Pr_D[g]$ denotes the probability that a random function chosen according to $D$ is identical to $g$. 
Lemma 4 (Without proof)

Suppose there exists a distribution $DP$ on inputs over $D$ that satisfy a given property $P$, and a distribution $DN$ on inputs that are $\varepsilon$-far from satisfying the property, and suppose further that for any $Q \subset D$ of size $q$, the variation distance between $DP|_Q$ and $DN|_Q$ is less than $1/3$. Then it is not possible for a non-adaptive algorithm making $q$ (or less) queries to $\varepsilon$-test for $P$. 

Lemma 5 (Without proof)

Suppose there exists a distribution $\mathcal{D}_P$ on inputs over $D$ that satisfy a given property $P$, and a distribution $\mathcal{D}_N$ on inputs that are $\epsilon$-far from satisfying the property, and suppose further that for any $Q \subset D$ of size $q$, and any $g : Q \rightarrow \{1,0\}$, we have
\[
\Pr_{\mathcal{D}_P|_Q}[q] < \frac{3}{2} \Pr_{\mathcal{D}_N|_Q}[q]
\]
Then it is not possible for a non-adaptive algorithm making $q$ (or less) queries to $\epsilon$-test for $P$.

The conclusion also holds if we have
\[
\Pr_{\mathcal{D}_N|_Q}[q] < \frac{3}{2} \Pr_{\mathcal{D}_P|_Q}[q]
\]
Two-sided testing where one of the graphs is known in advance

- The query complexity of two-sided error isomorphism testers is $\Theta(\sqrt{n})$ if one of the graphs is known in advance, and the other needs to be queried.

- We will proof only the lower bound.
The lower bound

- Any isomorphism tester that makes at most $\frac{\sqrt{n}}{2}$ queries to $H$ cannot distinguish between the case that $G$ and $H$ are isomorphic and the case that they are $1/32$-far from being isomorphic, where $G$ is known in advance.
Clone of graph

Given a graph $G$ and a set $W$ of $n/2$ vertices of $G$, we define the clone $G^{(W)}$ of $G$ in the following way:

- The vertex set of $G^{(W)}$ is defined as:
  $$V(G^{(W)}) = W \cup \{w' : w \in W\}$$

- The edge set of $G^{(W)}$ is defined as:
  $$\{(v,u) : (v,u) \in E(G)\} \cup$$
  $$\{(v',u) : (v,u) \in E(G)\} \cup$$
  $$\{(v',u') : (v,u) \in E(G)\}$$
Lemma 6

Let $G \sim G(n, 1/2)$ be a random graph. With probability $1 - o(1)$ the graph $G$ is such that for every sub set $W \subset V(G)$ of size $n/2$, the clone $G^{(W)}$ of $G$ is $1/32$-far from being isomorphic to $G$. 
Lemma’s proof

- Let $G$ be random graph according to $G(n,1/2)$, and let $W \subset V(G)$ be an arbitrary subset of $G$’s vertices of size $n/2$.

- First I show that for an arbitrary bijection $\sigma : (G^{(w)}) \to V(G)$ the graphs $G^{(W)}$ and $G$ are $1/32$-close under $\sigma$ with probability at most $2^{-\Omega(n^2)}$. 
Lemma’s proof

We split the bijection $\sigma: (G^{(w)}) \rightarrow V(G)$ into two injections $\sigma_1: W \rightarrow V(G)$ and $\sigma_2: V(G^{(w)}) \setminus W \rightarrow V(G) \setminus \sigma_1(W)$. Note that either $|W \setminus \sigma_1(W)| \geq n/4$ or $|W \setminus \sigma_2(W')| \geq n/4$. Assume without loss of generality that the first case holds, and let $U$ denote the set $W \setminus \sigma_1(W)$.
Lemma’s proof

Since every edge in G is chosen at random with probability 1/2, the probability that for some pair $u, v \in U$ either $(u,v)$ is an edge in G and $(\sigma(u), \sigma(v))$ is not an edge in G or $(u,v)$ is not an edge in G and $(\sigma(u), \sigma(v))$ is an edge in G is exactly $\frac{1}{2}$. 
Lemma’s proof

Therefore, using Chernoff bound we get that the probability that in the set $U$ there are less than $\binom{n}{2}/32$ such pairs is at most $2^{-\Omega(n^2)}$. 
Lemma’s proof

There are at most $n!$ possible bijections and $\binom{n}{n/2}$ possible choices for $W$, so using the union bound, the probability that for some $W$ the graph $G \sim G(n,1/2)$ is not $1/32$-far from being isometric to $G^{(w)}$ is $o(1)$. 
The distribution definition

- **DP**: A permutation of $G$, chosen uniformly at random.
- **DN**: A permutation of $G^{(W)}$, where both $W$ and the permutation are chosen uniformly at random.
**The lower bound proof**

- **According to Lemma 5 and Lemma 6, it is sufficient to show that the distribution DP and DN restricted to a set of $\sqrt{n}/4$ queries are close.**

- **In particular, we intend to show that for any $Q \subset D = V^2$ of size $\sqrt{n}/4$ and any $q:Q \rightarrow \{0,1\}$ we have $\Pr_{DP|Q}[q] < 3/2 \Pr_{DN|Q}[q]$**
Observation

- For a set $U$ of $G^{(w)}$'s vertices, define the event $E_u$ as the event that there is no pair of copies $w, w'$ of any one of $G$'s vertices in $U$.

- For a given set $Q$ of pairs of vertices, let $U_q$ be the set of all vertices that belong to some pair in $Q$. 
Observation

- Then the distribution \( DN\big|_Q \) conditioned on the event \( E_{VQ} \) and the unconditioned distribution \( DP\big|_Q \) are identical.
Observation’s proof

- In DN, if no two copies of any vertex were involved in the queries, then the source vertices of the $H$ are in fact a uniformly random sequence (with no repetition) of the vertices of $G$. This is the same as the unconditioned distribution induced by DP.
Lemma 7

For a fixed set $Q$ of at most $\sqrt{n} / 4$ queries and the corresponding $U = U_Q$ of vertices, the probability that event $E_U$ did not happen is at most $\frac{1}{4}$. 
Lemma’s Proof

- The bound on $|Q|$ implies that $|U| \leq \sqrt{n}/2$
- Now we examine the vertices in $U$ as if we add them one by one.
  - The probability that the vertex $v$ that added to $U$ is a copy of some vertex $u$ that was already inserted to $U$ is at most $\sqrt{n}/2n$
  - Hence, the probability that eventually we have two copies of the same vertex in $U$ is at most $\sqrt{n}/2n \cdot \sqrt{n}/2 = 1/4$
The lower bound proof

From the lemma 7 and the Observation we get:

for any \( g : Q \rightarrow \{1, 0\} \) we have

\[
\Pr_{DN|Q}[g] = \Pr[E_U] \cdot \Pr_{DP|Q}[g] + (1 - \Pr[E_U]) \cdot \Pr_{DP|Q}[g] \geq \\
\geq \Pr[E_U] \cdot \Pr_{DP|Q}[g] \geq \frac{2}{3} \Pr_{DP|Q}[g]
\]
Corollary

- It is not possible for any algorithm (adaptive or not) making $\sqrt{n}/4$ (or less) queries to test for isomorphism between a known graph and a graph that needs to be queried.