Graphs Connectivity
Directed and Undirected Graphs

Abner Chih-Yi Huang[1]

least upadte at November 24, 2006

[1]Graduate student of M.S. degree CS program of Algorithm and Biocomputing Laboratory, National Tsing Hua University. BarrosH@gmail.com

Outline

- Some terminologies
- Testing Connectivity of Graphs
  - Directed graph
  - Undirected graph
- Related work
  - Undirected bounded graph
  - Directed graph
- Other kind of tester.
- Summary

Disscus paper

- Topic: Testing Connectivity of Graphs
- Author: Hsuan Chang Chen
- Source: Master thesis, Graduate Institute of Computer Science and Information Engineering , NTU, 2006.

Property Tester I

Definiton

A **Tester** T for property P is an algorithm for given ϵ and input x,

- if P holds for x, then $\Pr[T(x,\epsilon) = \text{YES}] \geq \frac{2}{3}$
- if $d(x,P) \geq \epsilon$, then $\Pr[T(x,\epsilon) = \text{NO}] \geq \frac{2}{3}$

where the $d(\cdot,\cdot)$ is the distance function. $\Diamond$

Property Tester II

- Query complexity:

  We aim at spending time that is sublinear in or even independent of the size of graph.

- Distance function $d(\cdot, \cdot)$ `Hamming dist., Edit dist. etc.`
  - ϵ-`close`
  - ϵ-`far`

Testing Connectivity of Graphs I.

- How to solve this problem?
- What is the time complexity?
- Good enough?

Testing Connectivity of Graphs II.

- Graph representation matters!
  - Is the former algorithm good for adjacent matrix?
  - Is the former algorithm good for adjacent linked list?
- What are the suitable algorithms for such graph representations?

Testing Connectivity of Graphs III.

- Graph representation matters for property testing, too!
  - for adjacent matrix?
  - for adjacent linked list?
- It is not trivial! `Different graph representation might make the same problem un-testable.`

Testing Connectivity of Graphs VI.

- Related Work
  - Goldreich et. al., 1997[4], 2002[5] $O\left(\frac{\log^2(1/\epsilon d)}{\epsilon}\right)$, Undirected bounded degree model.
  - Blender et. al., 2000[1], 2002[2] $\Omega\left(\sqrt{n}\right)$, adjacency matrix.
- Chen's Contributions
  - On adjacency matrix
  - No restriction on degree.
  - Time complexity $O\left(\frac{1}{\epsilon \log(1-\epsilon)}\right)$

For Undirteced Graphs I.

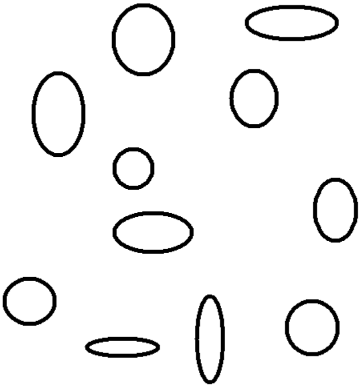


Figure: The disconnected graph and connected components

For Undirteced Graphs II.

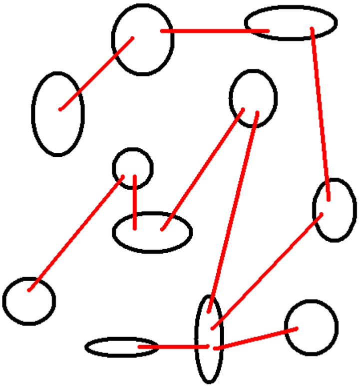


Figure: The connected graph with as spanning tree of connected components

For Undirteced Graphs III.

Lemma

If a graph G, is ϵ-far from the class of connected graphs, then it has more than $\epsilon\binom{n}{2}$ connected components.

Thus its distance measure is

$$\frac{\#\text{ of adding edges}}{\#\text{ of edges of } K_n}$$

For Undirteced Graphs IV.

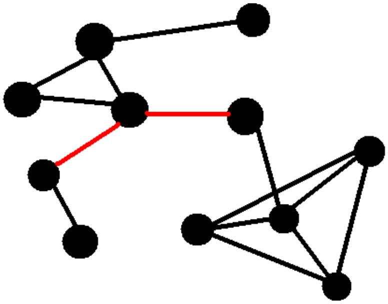


Figure: $\binom{11}{2} = 55, 55 \times \epsilon = 3, \therefore \epsilon = \frac{3}{55}$.

For Undirteced Graphs V.

Lemma

If a graph G, is ϵ-far from the class of connected graphs, then it has at least ϵn connected components each containning less than $\lceil \frac{2}{n-3}(\frac{1}{\epsilon} - 1) \rceil$ vertices.

For Undirteced Graphs VI.

Proof.

Assume it is wrong

$$
\begin{aligned}
n &\geq \{\epsilon\binom{n}{2} + 1 - \epsilon n\} \cdot \lceil \frac{2}{n-3}(\frac{1}{\epsilon} - 1)\rceil + \epsilon n \\
&> (\frac{\epsilon n(n-1)}{2} - \epsilon n) \cdot \frac{2}{n-3}(\frac{1}{\epsilon} - 1) + \epsilon n \\
&= \epsilon n(\frac{1}{\epsilon} - 1) + \epsilon n \\
&= n,
\end{aligned}
$$

For Undirteced Graphs VII.

Roughly speaking, we want to find some small connected components as the evidence of disconnectivity.

Algorithm

- $S \leftarrow \varnothing, M \leftarrow \lceil \log_{1-\epsilon} \frac{1}{3} \rceil, X \leftarrow \lceil \frac{2}{n-3}(\frac{1}{\epsilon} - 1) \rceil$
- While $|S| < M$ do
  - pick u from V then add it to S. Perform BFS from u and stop as reach X vertices or run out of vertices (REJECT)
- ACCEPT

For Undirteced Graphs VIII.

Theorem

If a graph G is connectivity, then it must accept. If G is ϵ-far from the class of connected graphs, then it has reject with probability at least $\frac{2}{3}$. The query complexity and time complexity are $O(\frac{1}{\epsilon \log(1-\epsilon)})$.

Obviously it is one-side error:

$$P \leq (1-\epsilon)^{\lceil \log_{1-\epsilon} \frac{1}{3} \rceil} \leq \frac{1}{3}$$

For Undirteced Graphs IX.

Since it costs n to find all neigbhors on adjacency matrix for each vertex, we have time complexity:

$$
\begin{aligned}
& \lceil \log_{1-\epsilon} \frac{1}{3} \rceil \cdot \lceil \frac{2}{n-3}(\frac{1}{\epsilon} - 1) \rceil \cdot n \\
< \;& [\frac{1}{\log_{\frac{1}{3}}(1-\epsilon)} + 1] \cdot [\frac{2}{n-3}(\frac{1}{\epsilon} - 1) + 1] \cdot n \\
= \;& O(\frac{1}{\epsilon \log(1-\epsilon)})
\end{aligned}
$$

For Digraphs I.

Recall some terminolgies of Digraph

- strongly connected: $\forall u, v \in V$, $\exists$ a directed path $u \longrightarrow v$ in G,
- source: node with only out-degree,
- sink: node with only in-degree,
- isolation: node without in-degree and out-degree,
- transferrer: node with in-degree and out-degree,

For Digraphs II.

Lemma

If a graph G, is ϵ-far from the class of strongly connected graphs, then the number of `source`, `sink`, *and* `isolation` *components in G will be larger than ϵn^2.*

Note: Chen assumed that all components are strongly connected.

For Digraphs III.

The idea of proof is similar to undirected one.
if no **transferrer**, the case is the follwoing

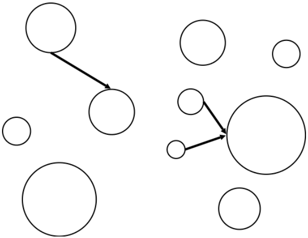


Figure: strongly connected components without transferrers

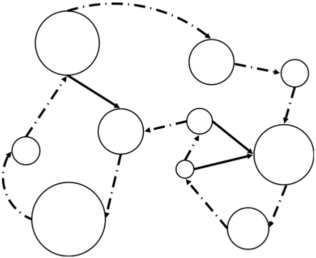


Figure: As directed cycle of strongly connected components

For Digraphs IV.

If **transferrers** exist, the case is the follwoing

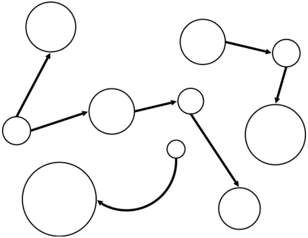


Figure: strongly connected components of several types

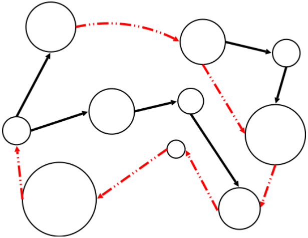


Figure: As directed cycle of strongly connected components which are not

For Digraphs V.

Lemma

If a graph G, is ϵ-far from the class of strongly connected graphs, then it has at least ϵn strongly connected components each containing less than $\lceil \frac{1}{n-2}(\frac{1}{\epsilon} - 1) \rceil$ vertices.

For Digraphs VI.

Proof.

Assume it is wrong

$$
\begin{aligned}
n &\geq \{\epsilon n^2 + 1 - \epsilon n\} \cdot \lceil \frac{1}{n-2}(\frac{1}{\epsilon} - 1) \rceil + \epsilon n \\
&> (\epsilon n^2 - \epsilon n) \cdot \frac{1}{n-2}(\frac{1}{\epsilon} - 1) + \epsilon n \\
&= \frac{n-1}{n-2}(n - \epsilon n) + \epsilon n \\
&> (n - \epsilon n) + \epsilon n \\
&= n
\end{aligned}
$$

For Digraphs VII.

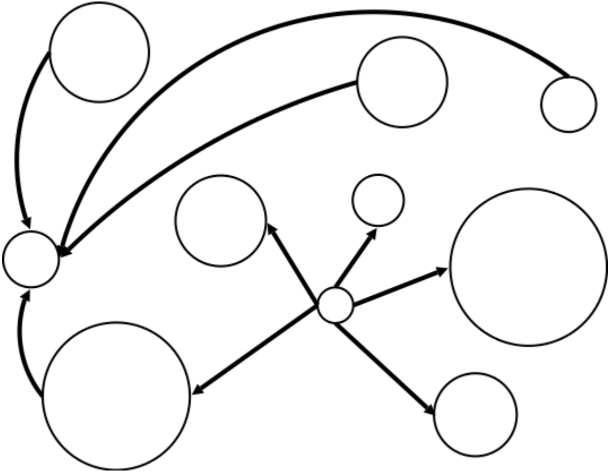
Roughly speaking, we want to find some small connected components as the evidence of disconnectivity.

But we need to check two directions, in-degree and out-degree.

Algorithm

- $S \leftarrow \varnothing, M \leftarrow \lceil \log_{1-\epsilon} \frac{1}{3} \rceil, X \leftarrow \lceil \frac{1}{n-2}(\frac{1}{\epsilon} - 1) \rceil$
- While $|S| < M$ do
  - pick u from V then add it to S. Perform BFS from u using in-degree and stop as reach X vertices or run out of vertices (REJECT)
  - If it reach X vertices
    - Perform BFS from u using out-degree and stop as reach X vertices or run out of vertices (REJECT)
- ACCEPT

Why do we need to check two directions?



For Digraphs VIII.

Theorem

If a digraph G is connectivity, then it must accept. If G is ϵ-far from the class of connected graphs, then it has reject with probability at least $\frac{2}{3}$. The query complexity and time complexity are $O(\frac{1}{\epsilon \log(1-\epsilon)})$.

Obviously it is one-side error:

$$P \leq (1-\epsilon)^{\lceil \log_{1-\epsilon} \frac{1}{3} \rceil} \leq \frac{1}{3}$$

For Digraphs IX.

Since it costs n to find all neigbhors on adjacency matrix for each vertex, we have time complexity:

$$
\begin{aligned}
& \lceil \log_{1-\epsilon} \frac{1}{3} \rceil \cdot 2 \cdot \lceil \frac{1}{n-2} (\frac{1}{\epsilon} - 1) \rceil \cdot n \\
< \; & [\frac{1}{\log_{\frac{1}{3}}(1-\epsilon)} + 1] \cdot [\frac{2}{n-2} (\frac{1}{\epsilon} - 1) + 2] \cdot n \\
= \; & O(\frac{1}{\epsilon \log(1-\epsilon)})
\end{aligned}
$$

Some questions!

Is Chen's algorithm a property tester?

- For given undirected disconnected graph $G(V, E)$, $|V| = n, |E| = m$, one add $k - 1$ edges to make G connected, therefore

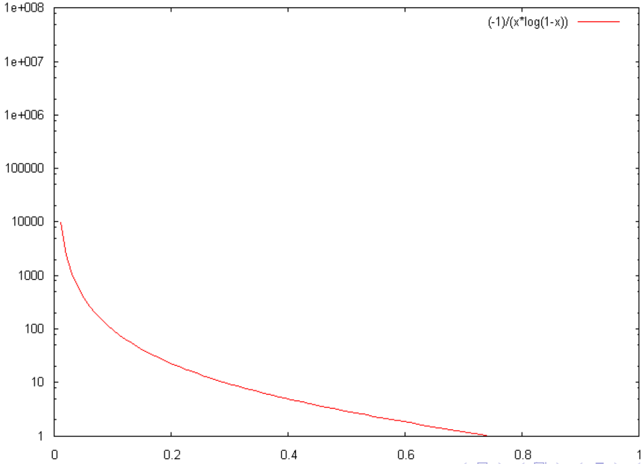
$$\epsilon \binom{n}{2} = \epsilon \frac{n(n-1)}{2} = k$$

$$\epsilon = \frac{2(k)}{n(n-1)} \leq \frac{2}{n}$$

It is not independent from input.

Some questions!

If $\epsilon \to 0$, then $\frac{1}{\epsilon} \to \infty$, but $\frac{1}{\log(1-\epsilon)} \to -\infty$.



Some questions!

What's the contribution of previous wroks?

- Goldreich et. al., 1997[4], 2002[5]. Their algorithm could test the k-connectivity, and in $O\left(\frac{\log^2(1/\epsilon d)}{\epsilon}\right)$ for $k = 1$ on bounded degree model.
- $\text{dist}(G, P) = 2\rho_d(G, P)/dn$, i.e., $\rho_d(G, P)$ the edit distance to closest connected graph P.

Goldreich et. al. 2002

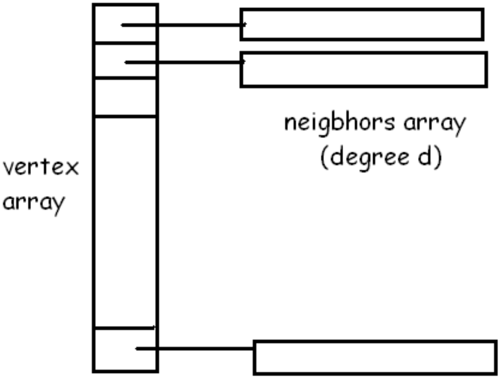


Figure: bounded degree model

Goldreich et. al. 2002

Connectivity Testing Algorithm

1. For i from 1 to $\log(8/(\epsilon d))$ do:
   - Uniformly and independently select $m_i = 32\log(8/(\epsilon d))/(2^i \cdot \epsilon \cdot d)$ vertices in the graph.
   - For each vertex s selected, perform a BFS starting from s until 2^i vertices have been reached or no more new vertices can be reached (REJECT).
2. ACCEPT.

Goldreich et. al. 2002

Lemma

Let $d \geq 2$. If a graph G is ϵ-far from the class of n-vertex connected graphs with maximum degree d, then it has more than $(\epsilon/4)dn$ connected components.

Goldreich et. al. 2002

Lemma

If a graph G is ϵ-far from the class of n-vertex connected graphs of degree bound $d \geq 2$, then G has at least $dn\epsilon/8$ connected components each containing less than $8/(d\epsilon)$ vertices.

Goldreich et. al. 2002

Lemma

If G is ϵ-far from the class of connected graphs with maximum degree d, then Algorithm rejects it with probability at least 2/3. The query complexity and running time of the algorithm are $O\left(\frac{\log^2(1/\epsilon d)}{\epsilon}\right)$

Goldreich et. al. 2002

PROOF:
Since it is ϵ-far, there are at least $dn\epsilon/8$ connected components. Let B_i be the set of connected components in G which contain at most $2^i - 1$ vertices and at least 2^{i-1} vertices.

Goldreich et. al. 2002

The probability that a uniformly selected vertex resides in one of these components is at least

$$\frac{2^{i-1}|B_i|}{n} \geq \frac{2^{i-1}}{n}\frac{dn\epsilon}{(8\log(8/(\epsilon d)))} = \frac{2}{m_i}$$

Thus $1-(1-\frac{2}{m_i})^{m_i} \geq 1-e^{-2} \geq \frac{2}{3}$.

Goldreich et. al. 2002

Roughly speaking, for each turn,

$$
\begin{aligned}
O(m_i \times \text{Time}_{BFS}) &= 32\frac{\log(\frac{8}{\epsilon d})}{(2^i \cdot \epsilon \cdot d)} \cdot 2^i d \\
&= \frac{32\log(\frac{8}{\epsilon d})}{\epsilon}
\end{aligned}
$$

, Thus it costs $O\left(\frac{\log^2(1/\epsilon d)}{\epsilon}\right)$.

End of Proof

Blender et. al., 2002

Blender et. al. proposed two algorithms for testing connectivity on digraph.

1. The informations of outgoing edges and incoming edges of a vertex are available.

   Note that Chen's algorithm for digraph is this type.

2. Only the information of outgoing edges of a vertex is available.

Blender et. al., 2002

- The first one is very similar to Chen's algorithm, but it is based on bounded degree graph
- The time complexity is in $O(\frac{1}{\epsilon^2 d})$,
- Chen found a better "small size"

Are those algorithms testers?

Is the ϵ independnet from G? No!

Goldreich:

$\epsilon\frac{dn}{2} \leq n - 1 \Rightarrow \epsilon < \frac{2}{d}, \epsilon\frac{dn}{2} \geq 1 \Rightarrow \epsilon \geq \frac{2}{dn}$

$$O\left(\frac{\log^2(1/\epsilon d)}{\epsilon}\right) \Rightarrow \left(\frac{d\log^2(d)}{2}\right), dn\log^2(\frac{n}{2})$$

Review the Property Tester

Property tester behaves as an algorithm with constant parameter ϵ.

Chen's ϵ is not independnet, too!

Does there exist indepednet tester?

Testing Monotonicity

Q: Dose a sequence be sorted?

- Any deterministic decision algorithm runs in $\Omega(n)$ to read the input and make a decision.
- Ergun et. al. 2000[3] proposed an algorithm to solve it in sublinear time.

Ergun et. al. 2000

Monotonicity Testing Algorithm

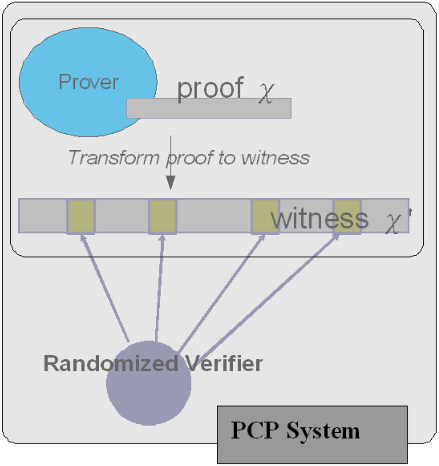
- For j from 1 to $O(1/\epsilon)$ do:
  - Query x_i from sequence $x_1 x_2 \cdots x_n$, uniformly at random.
  - Perform binary search for x_i. If the search does not found x_i, return "No".
- "YES"

Ergun et. al. 2000

- Their algorithm runs in time $O(\frac{\log n}{\epsilon})$ since each binary search costs $O(\log n)$.
- It is in **BPP**.
- ϵ-far for ϵn elements that BSearch fails,
- $(1-\epsilon)^{c/\epsilon} < e^{c/\epsilon} < 1/3$ for some c
- ϵ is independnet from n.

Summary

- Graph representations,
- Randomized sampling,
- Query complexity, `Oblivious Tester` [6]
- `Parameter` ϵ `matters! Decision version PTAS?`



Bibliography I

BENDER, M. A., AND RON, D.
Testing acyclicity of directed graphs in sublinear time.
In *ICALP '00: Proceedings of the 27th International Colloquium on Automata, Languages and Programming* (London, UK, 2000), Springer-Verlag, pp. 809–820.

Bibliography II

- BENDER, M. A., AND RON, D.
  Testing properties of directed graphs: acyclicity and connectivity.
  Random Structures and Algorithms 20, 2 (2002), 184–205.

- ERGUN, F., KANNAN, S., KUMAR, S. R., RUBINFELD, R., AND VISWANATHAN, M.

Bibliography III

Spot-checkers.
Journal of Computer and System Sciences 60, 3 (June 2000), 717–751.

GOLDREICH, O., AND RON, D.
Property testing in bounded degree graphs.
In *STOC '97: Proceedings of the twenty-ninth annual ACM symposium on*

Bibliography IV

Theory of computing (New York, NY, USA, 1997), ACM Press, pp. 406–415.

Goldreich, O., and Ron, D.
Property testing in bounded degree graphs.
Algorithmica 32, 2 (2002), 302–343.

Bibliography V

- SHAPIRA, A.
  Graph Property Testing and Related Problems.
  PhD thesis, School of Computer Science at Tel-Aviv University, 2006.
  Thesis Prepared Under the Supervision of Prof. Noga Alon.