236358
Distributed Graph Algorithms

Spring 2017
Class by Keren Censor-Hillel
Decision Tasks

• Is there a triangle in the graph?
• Is there a 4-cycle in the graph?
• Is there a $k$-cycle in the graph?
Distributed Decision

- **Distributed Decision:**
  - If \( G \) satisfies \( P \), all nodes output \textcolor{red}{true}.
  - If \( G \) does not satisfy \( P \), at least one node outputs \textcolor{red}{false}.

- Significance of allowing only one node to indicate false output (compare to global tasks).
Decision Tasks

- Is there a triangle in the graph?
- Is there a $4$-cycle in the graph?
- Is there a $k$-cycle in the graph?

- In LOCAL?
  - $O(k)$ rounds
- In CONGEST?
  - depends
3,4-cycles

4-cycles
• Can be solved in $O(\sqrt{n})$ rounds
• Admits an $\Omega(\sqrt{n}/\log n)$ lower bound

3-cycles
• Seems hard
• Distributed property testing
4-Cycle Detection

**Fact 1:** There is a constant $c$ such that every graph with $cn^{3/2}$ edges contains a 4-cycle

**Theorem:** There is an $O(\sqrt{n})$-round algorithm for detecting 4-cycle freeness
4-Cycle Detection Algorithm

- \( T = a \sqrt{n} \), for \( a \geq 2c+2 \) (c from Fact 1)
- \( N^{\text{heavy}}(v) = \{ u \in N(v) \mid d(u) > T \} \)
- \( h(v) = |N^{\text{heavy}}(v)| \)

1. If \( d(v) \leq T \), send \( N(v) \) to \( N(v) \)
2. If \( d(v) > T \)
3. If \( h(v) > T \), output \text{false}
4. Else, \( h(v) \leq T \), send \( N^{\text{heavy}}(v) \) to \( N(v) \)
5. If received \( z \) from \( u,w \) in \( N(v) \), output \text{false}
6. Output \text{true}
Complexity

The only communication is in Lines 1 and 4, each requires $T = O(\sqrt{n})$ rounds.
Correctness

**Claim 1:** If $h(v) > T$ then there is a 4-cycle.

**Proof:**

- Number of edges touching $N^{\text{heavy}}(v)$ is at least $h(v)T/2$
- By **Fact 1**, if there are $c \cdot h(v)^{3/2}$ edges within $N^{\text{heavy}}(v)$ then there is a 4-cycle
- Otherwise, at least $h(v)T/2 - c \cdot h(v)^{3/2}$ edges are from a node in $N^{\text{heavy}}(v)$ to a node outside of $N^{\text{heavy}}(v)$
Correctness

Claim 1: If $h(v) > T$ then there is a 4-cycle.

Proof: Otherwise, at least $h(v)T/2 - c \cdot h(v)^{3/2}$ edges are from a node in $N^{\text{heavy}}(v)$ to a node outside of $N^{\text{heavy}}(v)$

- $h(v)T/2 - c \cdot h(v)^{3/2} = h(v)(T/2 - c \cdot h(v)^{1/2})$
  \[ \geq h(v)(a\sqrt{n}/2 - c\sqrt{n}) \]
  \[ \geq a\sqrt{n} (a\sqrt{n}/2 - c\sqrt{n}) \]
  \[ = (a^2/2 - a \cdot c)n > 2n \]

- So there is at least one shared neighbor that is not $v$ itself
Correctness

• **Proof of Theorem**: By **Claim 1**, there is a node that outputs true only if there is a 4-cycle.

• Need to show that if there is a 4-cycle, there is a node that outputs true. By a case analysis.

![Diagram with nodes a, b, c, d connected to form a 4-cycle](image)
Correctness

A light node sends all of its neighbors

A heavy node sends all of its heavy neighbors
Lower Bound for Detecting 4-Cycles

**Fact 2:** There is a constant $c'$ and a graph $G_{-4}$ with $c'n^{3/2}$ edges that does not contain a 4-cycle

**Theorem:** Every algorithm for detecting 4-cycles requires $\Omega(\sqrt{n}/\log n)$ rounds

- Add edges of $G_{-4}$ depending on input
- 4-cycle iff inputs not disjoint
- input size $k = \Theta(n^{3/2})$
- cut size $C = \Theta(n)$
- Requires $R = \Omega(\sqrt{n}/\log n)$ rounds
3-Cycle Detection

• Can solve in $\Delta$ rounds
  – Exchange list of neighbors
  – Linear in $n$, in the worst case

• Faster solutions?
  – New randomized sub-linear algorithm

• Lower bounds?
  – Alice-Bob framework does not work (each triangle is known to a player having two of its vertices)
Property Testing

• LCA – local computation algorithms
• Provide answers without reading the entire input

Property testing:
  – Make queries
  – Output true if $G$ satisfies property $P$ holds
  – Output false if $G$ is far from satisfying property $P$
  – Otherwise, arbitrary output allowed
Triangle-Freeness

• $G$ is triangle-free if there are no triangles in $G$

• Different than the property of having triangles
  – Which is harder
When is $G$ far from satisfying $P$?

• Two graphs $G$ and $G'$ obtained from each other by inserting or deleting $\epsilon n^2$ edges are $\epsilon$-far

• $G$ is $\epsilon$-far from satisfying $P$ if $G'$ does not satisfy $P$, for every $G'$ that is $\epsilon$-far from $G$
Distributed Property Testing

- Distributed Property Testing:
  - If $G$ satisfies $P$, all nodes output $\text{true}$
  - If $G$ is $\varepsilon$-far from satisfying $P$, at least one node outputs $\text{false}$
  - Otherwise, arbitrary outputs allowed

- Significance of allowing only one node to indicate false output (compare to global tasks)
Triangle-Freeness

**Theorem:** There is a distributed triangle-freeness testing algorithm that completes in $O(1/\varepsilon^2)$ rounds

**Algorithm:**

- For $32/\varepsilon^2$ rounds
  - Pick two neighbors $w_1 \neq w_2$ independently at random
  - Ask $w_1$ if $w_2$ is its neighbor
  - Return **false** if answered yes
- Return **true**
Analysis

• Let $b = 2\sqrt{\frac{m}{\varepsilon}}$

• **Heavy nodes**: $B = \{v \in V \mid d(v) \geq b\}$

• **Light nodes**: $V \setminus B = \{v \in V \mid d(v) < b\}$

• An edge $e = \{u,v\}$ is
  
  – **Light**, if $v$ or $u$ not in $B$
  
  – **Heavy**, otherwise: $u, v \in B$

• $H = \{u,v\} \in E \mid u \in B, v \in B\}$ is the set of heavy edges

• $T = \text{the set of light edges in triangles (subset of } E \setminus H)$
Number of Heavy Edges

• **Claim 1**: $|H| \leq \varepsilon m/2$

• **Proof**: Each heavy edge connects two heavy nodes (in $B$), so: $|H| \leq |B|(|B|-1)/2 < |B|^2/2$.

• $|B|b \leq 2m$, because the sum of degrees is at most twice the number of edges.

• Math: $|B| \leq 2m/b = 2m/2\sqrt{m/\varepsilon} = \sqrt{\varepsilon m}$

• Combine: $|H| \leq |B|^2/2 \leq \varepsilon m/2$
Number of Light Edges

• **Claim 2**: If $G$ is $\varepsilon$-far from being triangle-free, then $|T| > \varepsilon m/2$.

• **Proof**:  
  • If $G$ is $\varepsilon$-far from being triangle-free, then there are at least $\varepsilon m$ triangle edges.  
  • By **Claim 1**, $|H| \leq \varepsilon m/2$  
  • Hence, $|T| > \varepsilon m/2$
Matched Edges

• \( A = \{(v, w_1) \in E \mid v \in V \setminus B\} \)
• \( AT = A \cap T \)

• Edge \((v, w_1) \in AT\) is matched if \((v, w_2)\) is in the same triangle as \((v, w_1)\)

• If \((v, w_1) \in AT\) is matched then \(\{v, w_1, w_2\}\) is a triangle that is detected by \(v\)

• \( Y = \) the number of matched edges
Number of Matched Edges

• **Claim 3**: If $G$ is $\varepsilon$-far from being triangle-free, then $E[Y] \geq \varepsilon^2/8$

• **Proof**:
  • For $e=\{v,w_1\}$ in $AT$, $Y_e$ indicates $e$ is matched
  • $E[Y|AT] = E[ \sum_{e \in AT} Y_e | AT]$
    $$= \sum_{e \in AT} \Pr[e \text{ is matched}] \geq |AT|/b$$
  • Because probability of choosing a good $w_2$ is at least $1/b$
Number of Matched Edges

• **Claim 3**: If $G$ is $\varepsilon$-far from being triangle-free, then $E[Y] \geq \varepsilon^2/8$

• **Proof (cont.)**: For any $e$, $X_e$ indicates $e$ is in $A$
• $\sum_{e \in T} X_e = |AT|$
• Then, $E[|AT|] = E[\sum_{e \in T} X_e] = \sum_{e \in T} E[X_e] = \sum_{e \in T} \Pr[e \in A] \geq |T|/b$
• Because $e$ in $T$ gets picked by a light endpoint with probability at least $1/b$
• By **Claim 2**, $E[|AT|] > \varepsilon m/2b$
Number of Matched Edges

• **Claim 3**: If $G$ is $\varepsilon$-far from being triangle-free, then $E[Y] \geq \varepsilon^2/8$

• **Proof (cont.)**:

- $E[Y] = E_{AT}[E[Y|AT]] \geq E[|AT|/b]$
  \[ \geq \varepsilon m/2b^2 \geq \varepsilon m/2 \cdot 4(m/\varepsilon) = \varepsilon^2/8. \]
Wrap-up

• **Theorem**: There is a distributed triangle-freeness testing algorithm that completes in $O(1/\varepsilon^2)$ rounds

• **Proof**:  
  • If $G$ is triangle-free, all nodes return true
Wrap-up

• **Theorem**: There is a distributed triangle-freeness testing algorithm that completes in $O(1/\varepsilon^2)$ rounds

• **Proof (cont.)**: If $G$ is $\varepsilon$-far from being triangle-free:
  • $Z_{i,v}$ indicates that $v$ detects a triangle in iteration $i$
  • By **Claim 3**, $E[\sum_{v \in V} Z_{i,v}] = E[Y] \geq \varepsilon^2/8$
  • $Z = \sum_{i=1,\ldots,32/\varepsilon^2} \sum_{v \in V} Z_{i,v}$
  • $E[Z] = E[\sum_{i=1,\ldots,32/\varepsilon^2} \sum_{v \in V} Z_{i,v}] = \sum_{i=1,\ldots,32/\varepsilon^2} E[\sum_{v \in V} Z_{i,v}] \geq \sum_{i=1,\ldots,32/\varepsilon^2} \varepsilon^2/8 = 4$
Wrap-up

• **Theorem**: There is a distributed triangle-freeness testing algorithm that completes in $O(1/\varepsilon^2)$ rounds

• **Proof (cont.)**:
  
  • Chernoff bound: $\Pr[Z<(1-\delta)E[Z]] < (e^{-\delta}/(1-\delta)^{(1-\delta)})^{E[Z]}
  
  • Here $E[Z] \geq 4$, and take $\delta=3/4$
  
  • $\Pr[Z<1] \leq \Pr[Z<(1-\delta)E[Z]] < (e^{-3/4}/(1/4)^{(1/4}))^4$
    
    $= 4/e^3 < 1/3$

  • If $G$ is $\varepsilon$-far from being triangle-free, with probability at least $2/3$ there is a node that outputs false
Distributed Property Testing

• Faster triangle-freeness testing
• Additional subgraphs
• Cycles, bipartiteness