Distributed Detection of Cycles

Pierre Fraigniaud, Dennis Olivetti

GSSI, L'Aquila and IRIF, Paris

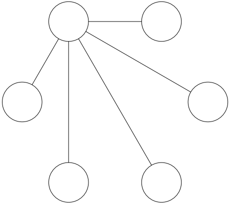
Outline

- Property Testing
- Distributed Property Testing
- Testing of C_k freeness

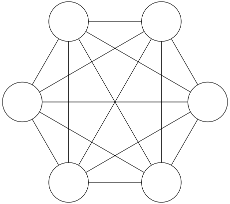
Property testing

- Given a property P
- Given a graph G
- Decide:
  - Does G satisfy the property P?
  - Is G far from satisfying the property P?
- The input is huge:
  - Only a small part of the input can be seen
  - We want sublinear algorithms

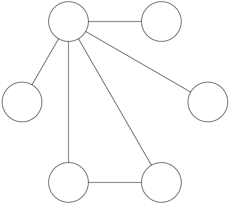
Example: 2 colorability



2 colorable



Far from being 2 colorable



Almost 2 colorable

How to measure how far is a graph from satisfying a property?

Let $G = (V, E)$, $n = |V|$, $m = |E|$. Let ϵ be a small constant in $(0, 1)$.
There exist two distinct models:

How to measure how far is a graph from satisfying a property?

Let $G = (V, E)$, $n = |V|$, $m = |E|$. Let ϵ be a small constant in $(0, 1)$. There exist two distinct models:

Dense model

A graph is ϵ-far from satisfying a property if at least ϵn^2 edges should be added or removed from G in order to make the property hold.

How to measure how far is a graph from satisfying a property?

Let $G = (V, E)$, $n = |V|$, $m = |E|$. Let ϵ be a small constant in $(0, 1)$. There exist two distinct models:

Dense model

A graph is ϵ-far from satisfying a property if at least ϵn^2 edges should be added or removed from G in order to make the property hold.

Sparse model

A graph is ϵ-far from satisfying a property if at least ϵm edges should be added or removed from G in order to make the property hold.

Complexity

- The complexity is measured in number of queries
- Different type of queries are allowed:
  - Give me the id of a random node and its degree
  - Give me the i-th neighbor of node x
  - Are nodes x and y neighbors?

Definition

Property Tester (1 sided error)

A tester for a graph property P is a randomized algorithm A that is required to accept or reject any given network instance, under the following two constraints:

- G satisfies $P \Rightarrow A$ accepts G
- G is ϵ-far from satisfying $P \Rightarrow Pr[A \text{ rejects } G] \geq \frac{2}{3}$

Subgraph freeness

We want to know if G does not contain any copy of a subgraph H, or if it contains many copies of H, being H some small graph (e.g. K_5).

- Easy in the dense model (using the Graph Removal Lemma)

Lemma

H freeness can be tested in constant time, for any H of constant size.

- Hard in the sparse model

Lemma [Alon, Kaufman, Krivelevich, Ron '08]

Testing triangle freeness requires $\Omega(n^{\frac{1}{3}})$ queries.

Distributed property testing

Definition

A distributed tester for a graph property P is a distributed randomized algorithm A that satisfies the following conditions:

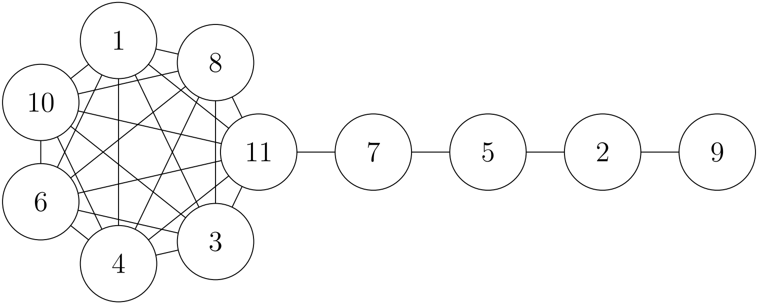
- G satisfies $P \Rightarrow$ every node outputs "accept"
- G is ϵ-far from satisfying $P \Rightarrow$
  $\Pr[\text{at least one node outputs "reject"}] \geq \frac{2}{3}$

The Congest Model

- All nodes start the computation at the same round
- The computation proceeds in phases
- At each phase each node:
  - sends (possibly different) messages to its neighbors
  - receives messages sent by its neighbors
  - performs some local computation

The main constraint of the Congest model is that the exchanged messages should be small, typically $O(\log n)$.

Knowing the 2-hop neighborhood is hard



State of the art

Lemma [Censor-Hillel, Fischer, Schwartzman, Vasudev '16]

Any ϵ-tester for the dense model (for a non-disjointed property) that makes q queries can be converted to a distributed ϵ-tester that requires $O(q^2)$ rounds in the distributed setting.

State of the art

Lemma [Censor-Hillel, Fischer, Schwartzman, Vasudev '16]

Any ϵ-tester for the dense model (for a non-disjointed property) that makes q queries can be converted to a distributed ϵ-tester that requires $O(q^2)$ rounds in the distributed setting.

[Censor-Hillel, Fischer, Schwartzman, Vasudev '16]

- Triangle freeness can be tested in $O(1/\epsilon^2)$
- Cycle freeness can be tested $O(\log n/\epsilon)$
- Cycle freeness requires at least $\Omega(\log n)$
- Bipartiteness can be tested in in $O(poly(\log \frac{n}{\epsilon}/\epsilon))$ in bounded degree graphs

State of the art

Lemma [Censor-Hillel, Fischer, Schwartzman, Vasudev '16]

Any ϵ-tester for the dense model (for a non-disjointed property) that makes q queries can be converted to a distributed ϵ-tester that requires $O(q^2)$ rounds in the distributed setting.

[Censor-Hillel, Fischer, Schwartzman, Vasudev '16]

- Triangle freeness can be tested in $O(1/\epsilon^2)$
- Cycle freeness can be tested $O(\log n/\epsilon)$
- Cycle freeness requires at least $\Omega(\log n)$
- Bipartiteness can be tested in in $O(poly(\log \frac{n}{\epsilon}/\epsilon))$ in bounded degree graphs

[Fraigniaud, Rapaport, Salo, Todinca '16]

- H-freeness can be tested in constant time for any H s.t. $|V(H)| \leq 4$

Results

Theorem

There exists an ϵ-tester for C_k freeness, for any constant $k \geq 3$, that requires $O(\frac{1}{\epsilon})$ rounds in the CONGEST model.

Results

Theorem

There exists an ϵ-tester for C_k freeness, for any constant $k \geq 3$, that requires $O(\frac{1}{\epsilon})$ rounds in the CONGEST model.

Procedure:

- Choose an edge u.a.r.
- Check if there is a cycle of length k passing through that edge
  - It can be done deterministically

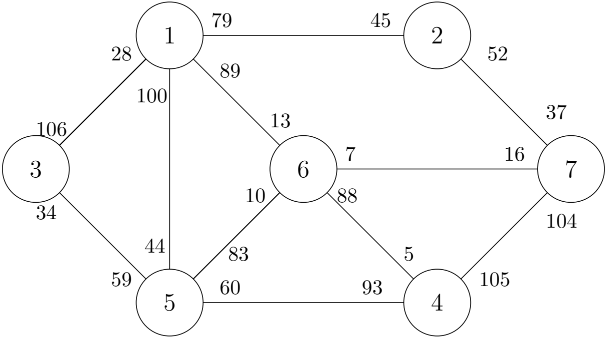
Choose an edge at random

Lemma [Fraigniaud, Rapaport, Salo, Todinca '16]

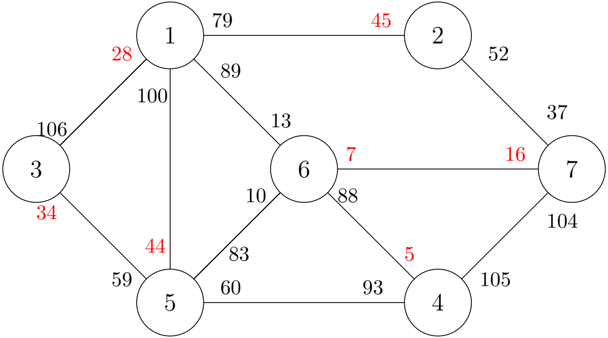
Let H be any graph. Let G be an m-edge graph that is ϵ-far from being H-free. Then G contains at least $\epsilon m/|E(H)|$ edge-disjoint copies of H.

This implies that by choosing a random edge we have probability $\Omega(\epsilon)$ to choose an edge that is part of some copy of H.

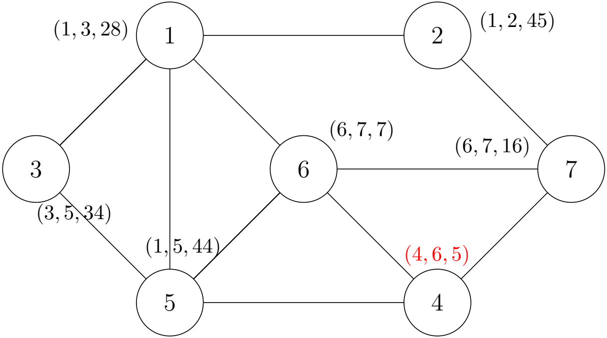
Choose an edge at random



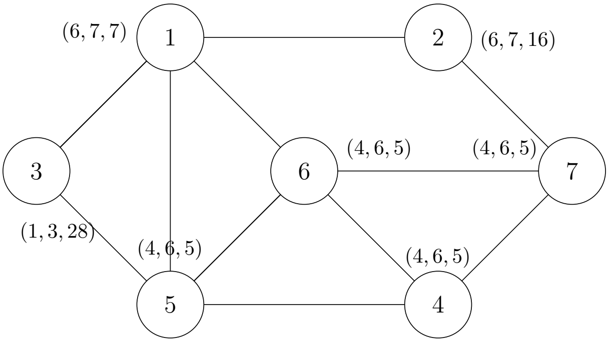
Choose an edge at random



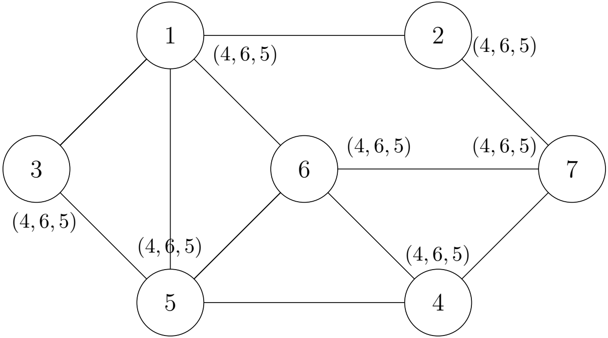
Choose an edge at random



Choose an edge at random



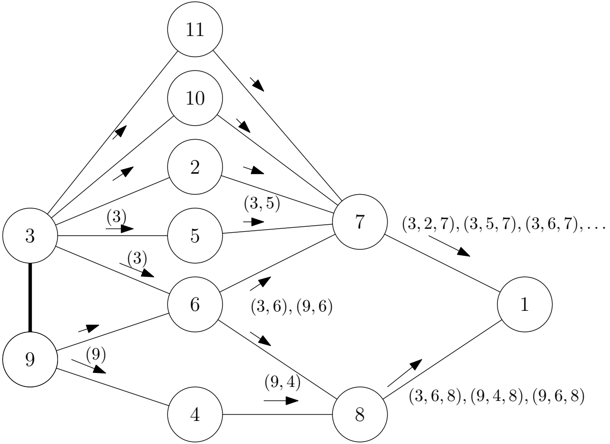
Choose an edge at random



Check the presence of a cycle

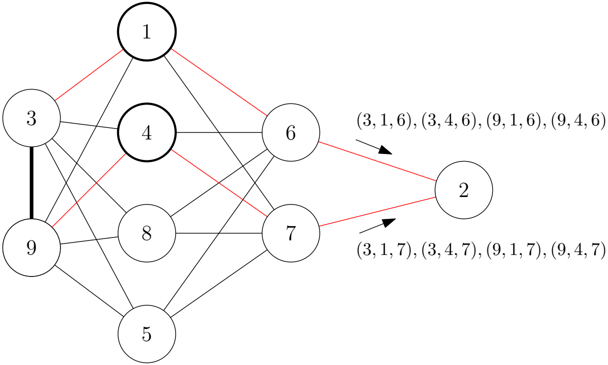
Append and forward:

- Nodes at distance 2 could potentially receive $\Theta(n)$ messages
- Not feasible in the CONGEST model



C_7 detection

- The partial solution can be sparsified
- For C_7, 3 subpaths (for each inital node) are enough



Sparsification of the intermediate solution

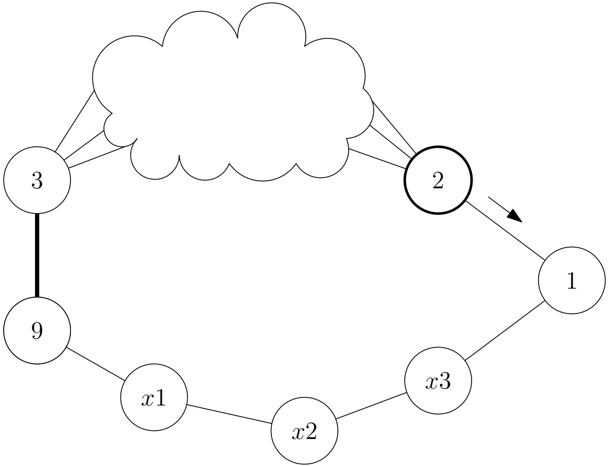
Lemma [Erdős, Hajnal, Moon '64]

Let V be a set of size n, and consider two integer parameters p and q. For any set $F \subseteq \mathcal{P}(V)$ of subsets of size at most p of V, there exists a *compact* (p, q)*-representation* of F, i.e., a subset $\hat{F}$ of F satisfying:

1. For each set $C \subseteq V$ of size at most q, if there is a set $L \in F$ such that $L \cap C = \emptyset$, then there also exists $\hat{L} \in \hat{F}$ such that $\hat{L} \cap C = \emptyset$;
2. The cardinality of $\hat{F}$ is at most $\binom{p+q}{p}$, for any $n \geq p + q$.

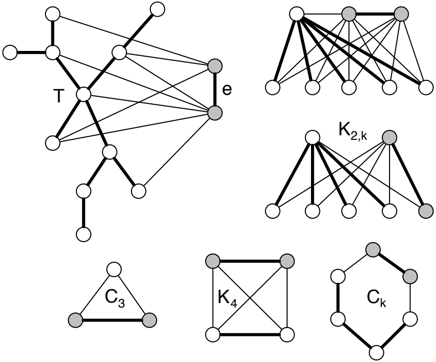
In other words, the number of subpaths that must be forwarded at each round do not depend on the size of the graph.

Sparsification of the intermediate solution



- Node 2 should send at least one sequence that does not contain $x1, x2$ and $x3$
- A constant number of sequences are enough

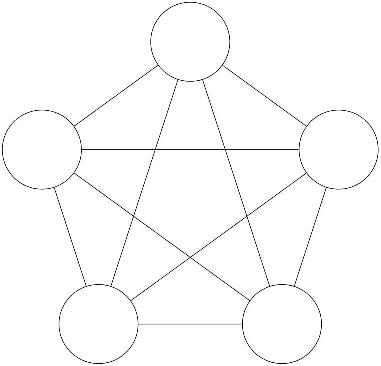
Tree + 1 edge



[Fraigniaud, Montealegre, Olivetti, Rapaport, Todinca '17]

There exists an ϵ-tester for H freeness, for any graph H of constant size composed by a tree, an edge, and arbitrary connections between the endpoints of the edge and the nodes of the tree, that requires $O(\frac{1}{\epsilon})$ rounds in the CONGEST model.

Open problems



Does there exist an ϵ-tester for K_5-freeness?

Thank you