Graph-Constrained Group Testing

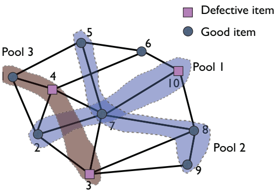
Mahdi Cheraghchi et al.

presented by Ying Xuan

May 20, 2010

Table of contents

Applications call for connected pools

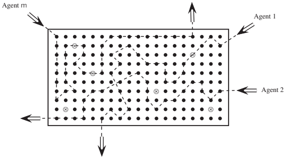


- detection of congested links in IP networks or all-optical networks using probes.
- detection of dead nodes or links in sensor networks using testing packets.
- detection of infected individuals using human agents.

All need the testing pools to be walks by the probe/packet/agent.

Four Problem Variations for detecting defected vertices

Given a Undirected non-weighted graph $G = (V, E)$ with $|V| = n$ and at most $d \ll n$ defected vertices; find the d-disjunct matrix standing for the testing pools.



- (**Fixed Testing Entrances**) all the probes starting from r designated vertices (entry), no constraint on the exit;
- (**Fixed Testing Exit**) all the probes stops at a designated sink node (exit), no constraint on the entrance;
- **Fixed Testing Entrances and Exit**;
- **No constraints on Entrances and Exit**

Similar for detecting detected edges.

Necessary Constraints on the underlying graph

(D, c)-uniform

$D \leq deg(v) \leq cD$ for special parameters $D, c > 1$ and $\forall v \in V$.

$(\frac{1}{2}cn)^2$-mixing time

The smallest integer $T(n) = t$ such that a random walk of length t starting at $\forall v \in V$ ends up having a distribution μ' with

$$\|\mu' - \mu\|_\infty = \max_{i \in \Omega} \|\mu(i) - \mu'(i)\| < (\frac{1}{2}cn)^2$$

Equivalent Constraints

Specially, the graph can either be the following two kinds

- a random graph $G(n, \frac{c^2 d \log^2 n}{n})$;
- any graph with conductance

$$\Phi(G) := \min_{S \subseteq V : \sum_{v \in S} deg(v) \leq |E|} \frac{E(S, \bar{S})}{\sum_{v \in S} deg(v)} = \Omega(1)$$

if we need $T(n) = O(\log n)$ (can be relaxed).

Algorithms

Construct each row of the testing matrix independently from a walk by letting each walked through vertices as 1, others as 0. The d-disjunct matrices with probability $1 - o(1)$ for different problem variations are:

- **(Fixed Testing Entrances)** $m_1 \times |V|$: each walk starts from a designated entry vertex, having t_1 hops.
- **(Fixed Testing Exit)** $m_4 \times |V|$: each walk starts from an arbitrary vertex, and ends at the designated exit vertex.
- **(Fixed Testing Entrances and Exit)** $m_3 \times |V|$: each walk starts from a designated entry vertex, and ends at the designated exit vertex.
- **(No constraints on Entrances and Exit)** $m_2 \times |V|$: each walk starts from an arbitrary vertex, having t_2 hops.

Parameter	Value
D_0	$O(c^2 dT^2(n))$
m_1, m_2	$O(c^4 d^2 T^2(n) \log(n/d))$
m_3	$O(c^8 d^3 T^4(n) \log(n/d))$
m_4	$O(c^9 d^3 DT^4(n) \log(n/d))$
t_1	$O(n/(c^3 dT(n)))$
t_2	$O(nD/(c^3 dT(n)))$

Three probabilities

Definition

Consider a random walk $W := (v_0, v_1, ..., v_t)$ of length t where all these vertices form a Markov chain. Define three probabilities related to W:

- π_v the probability that W passes any single node v;
- $\pi_{v,A}$ the probability that W of length t passes node v, but none of the vertices in A where $A \subseteq V$ and $v \notin A$.
- $\pi^u_{v,A}$ the probability that W with sink (exit) u passes node v, but none of the vertices in A where $A \subseteq V$ and $v \notin A$.

Why we need (D, c)-uniform?

Lemma

Denote by μ the stationary distribution of G, then for each $v \in V$, $\mu(v) \in [\frac{1}{cn}, \frac{c}{n}]$.

Proof.

$(D, c) - uniform \Rightarrow \; D \leq deg(v) \leq cD \Rightarrow \; nD \leq 2|E| = sum_v deg(v) \leq ncD$
property: a random walk on any graph that is not bipartite converges (finite number of steps) to a stationary distribution $\mu(v) = \frac{deg(v)}{2|E|}$ □

Apparently, this is loose, so D, c-uniform can be relaxed for specific topology.

Why we need δ-mixing time

Lemma

$$\pi = \Omega(\frac{t}{cnT(n)})$$

Proof.

Assume the random walk $W = \{w_0, w_1, \cdots, w_{t/T(n)}\}$ with $w_i = v_{iT(n)}$ (scale to $T(n)$), from the definition of δ-mixing time, where $\delta = (\frac{1}{2}cn)^2$, we can see

$$\begin{aligned} Pr[w_0 \neq v, w_1 \neq v, \cdots, w_t \neq v] &\leq (1 - 1/cn + \delta)^{t/T(n)} \\ &\leq (1 - 1/2cn)^{2t/T(n)} \\ &\leq exp(-t/(cnT(n))) \\ &\leq 1 - \Omega(t/cn(T(n))) \end{aligned}$$

If $\mu(v)$ can be tightened, δ can be enlarged, so that t could be smaller, so the matrix will have smaller row weight. □

What do we need to lower bound $\pi_{v,A}$

Idea: we don't want the walk to enter the set A within t steps, so we can upper bound the probability of each vertices being passed for more than $k > 1$ times and being passed within the first h steps. Can we get h larger enough than t so we can avoid passing the vertices in A? Not that straightforward.

Lemma

There is a $k = O(c^2 T(n))$ such that for every $v \in V$, the probability that W passes v more than k times is at most $\pi_v/4$

Lemma

For any walk W, if v is not a designated entrance vertex, then the probability that W visits v within the first h steps is at most h/D.

Lower Bounding $\pi_{v,A}$

Theorem

For the first algorithm (Fixed Testing entries) with D_0 and t_1 mentioned above. Let $v \in V$ and $A \subseteq V$ be a subset of at most d vertices in G such that $v \in A$ and $A \cap \{v\}$ does not include any of the designated vertices $s_1, s_2, \cdots s_r$, then

$$\pi_{v,A} = \Omega(\frac{1}{c^4 d T^2(n)})$$

Proof

Proof.

- $\mathcal{G}$:= event that W hits v no more than $k = O(c^2T(n))$ times and never within the first $2T(n)$ steps. $\Rightarrow$ $Pr[\mathcal{G}] \geq 1 - 2T(n)/D - O(t/cnT(n))$;
- $\mathcal{B}$:= event that W hits some vertex in $A \Rightarrow$ $\pi_{v,A} \geq Pr[\neg\mathcal{B}, v \in W, \mathcal{G}]$;
- upperbound $Pr[\mathcal{B}|v \in W, \mathcal{G}]$;
- lowerbound $\pi_{v,A}$.

□

upperbound $Pr[\mathcal{B}|v \in W, \mathcal{G}]$

Proof.

- fix $i > 2T(n)$ and $v_i = v$, i.e. assume W visits v after $2T(n)$ steps;
- divide the walk into four parts W_1, W_2, W_3, W_4 with intervals $(0, T(n)), (T(n)+1, i-T(n)-1), (i-T(n), i+T(n)), (i+T(n)+1, t)$;
- bound $\mathcal{B}$ for each node in each interval, and get loose union bound for each i value as $Pr[\mathcal{B}|v_i, \mathcal{G}] \leq 1.1(\frac{6dT(n)}{D} + \frac{4dct}{n})$
- since W hits v no more than k times, consider $t > 2T(n)$ events $v_i = v$ for $i = [2T(n)+1, t]$, their intersection is empty. Since $v \in W$ is the union of these events, we have a union bound

$$Pr[\mathcal{B}|v \in W, \mathcal{G}] = O(c^2 T(n)(\frac{6dT(n)}{D} + \frac{4dct}{n}))$$

□

Main Theorem

Correctness of the first algorithm

The first algorithm returns a $O(c^4 D^2 T^2(n) \log(n/d)) \times n$ d-disjunct matrix for $D > O(c^d T^2(n))$ and $t = O(n/(c^3 d T(n)))$.

Proof

- $X_i :=$ the i^{th} row has 1 at column v and all 0 at $|A| < d$ columns, so $E[X_i] = Pr[X_i = 1] = \pi_{v,A}$;
- failure probability for all $v \in V$ and d-subset A is $p_f \leq \sum_{v,A}(1 - \pi_{v,A})^m \leq exp(d \log \frac{n}{d}) \left(1 - \Omega(\frac{1}{c^4 d T^2(n)})\right)^m = o(1)$

Relaxations

- (D, c)-uniform;
- δ-mixing time;
- calculation of the failure probability;

Possible Usages

- study on specific topologies instead of arbitrary graph;
- divide the graph into multiple subgraphs that satisfy the graph constraint;

The End

Q & A