Group Testing and Graph Constrained Group Testing

Venkatesh Saligrama
Boston University

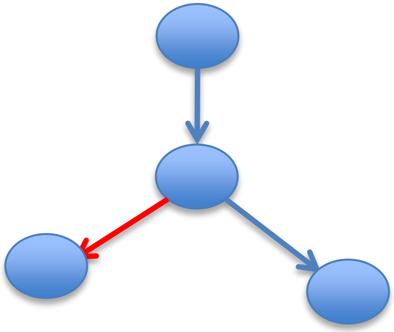
Workshop on Frontiers of Controls, Games & Network Science with Civilian and Military Applications, UT Austin

Based on collaborations with G. Atia, M. Cheraghchi, A. Karbasi, S. Mohajer

Applications of Group Testing

- DNA Screening (clones) [Du,Hwang'00]

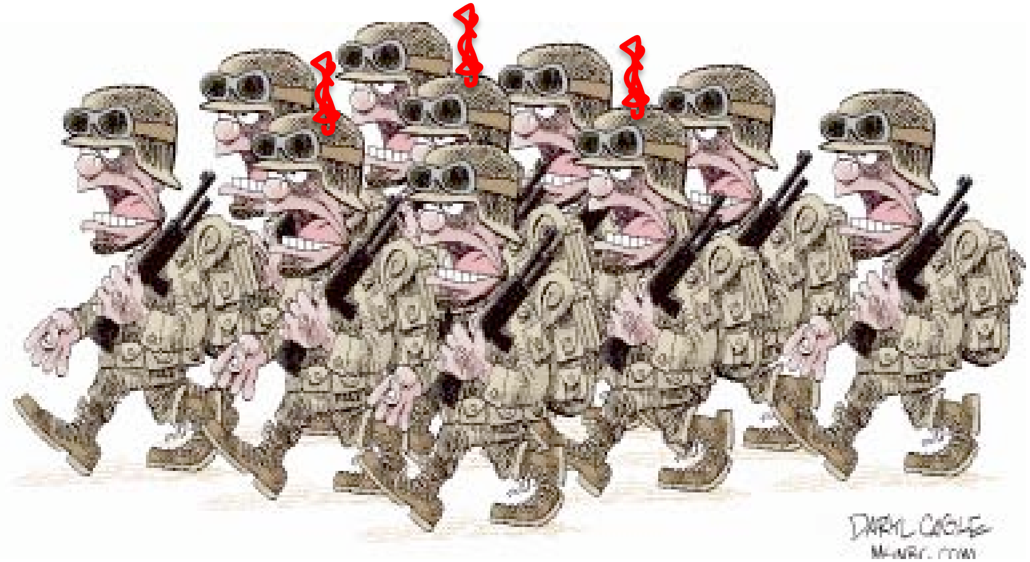
- Streaming [Gilbert-Strauss 08]
  - Telephone Calls
- Compressed Sensing [Muthukrishnan 05]



- Congested IP Link [Nguyen-Thiran07]
- Spectrum Violation in Cognitive Radio [Atia-S-S 08]

What is Group testing?

- K soldiers in a large population, N, have disease
  - Detect by testing pooled blood samples
    - Compressed sensing in 1940s



Illustrative Example

Patients (Blood Samples)

		1	2	3					N	TEST Results
	POOL 1	1	0	0	1	1	1	0	0	-ve
	POOL 2	0	0	0	0	0	0	1	0	-ve
	POOL 3	0	1	1	0	0	1	1	0	+ve
					⋮					
Patients {1, 3} →	POOL T	1	0	1	0	0	0	0	0	+ve

- **Goal:**
  - Find defectives
  - How many tests are required?

Non-Adaptive Group Testing

- Given
  - N items: $R_1, R_2, R_3, \ldots, R_N \in \{0,1\}$
  - $\leq$ K defectives: $R_m = 1$
- Test Matrix: $\mathbf{X} = [X_{mn}]$,
  - $X_{mn} = 1 \rightarrow$ Put nth item in pool m
- Binary Output (Y_m) - mth test
  - $Y_m = 1$ is mth pool contains a defective
  - Noisy Case

$$Y_m = \bigvee_{n=1}^{N} X_{mn} R_n$$

- Goal:
  - Find defectives
  - Min. number of tests required

Noisy Cases

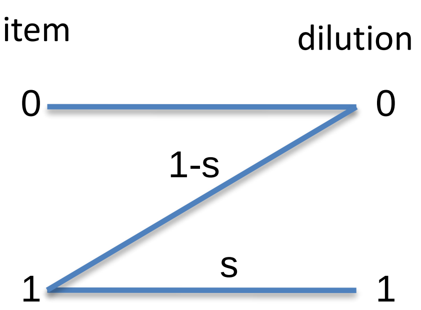
- Additive noise (False Positives)

$$Y_m = \bigvee_{n=1}^{N} X_{mn} R_n \bigvee W_m$$

W_m Bernoulli(q)

- Dilution Effect

$$Y_m = \bigvee_{n=1}^{N} \mathcal{Z}(X_{mn} R_n)$$



  - Blood Sample is Diluted
  - Link Losses are random [Nguyen-Thiran07]

Problem Setup

- N Items, Defectives $\leq$ K, **X** Test Matrix, T tests
- Defective set: $s \in \mathcal{D}$
- Decoder $g : Y^T \rightarrow \mathcal{D}$
- Errors:
  - Average Error, Worst-Case, …

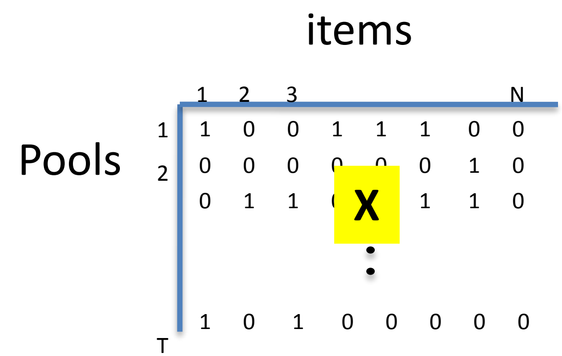
$$\lambda_s(\mathbf{X}) = \mathbf{1}(g(Y^T) \neq s|s)$$

$$\bar{\lambda}_s(\mathbf{X}) = \text{Prob}(g(Y^T) \neq s|s))$$

$$\frac{1}{|\mathcal{D}|} \sum_s \lambda_s(\mathbf{X}) \qquad \max_s \lambda_s(\mathbf{X})$$

Problems

- Problem 1
  - X arbitrary test matrix
  - Can design any test matrix
  - Lot of work …. (next slide)



- Problem 2:
  - X structured test Matrix
  - Test matrix comes from a constraint set (Graph)
  - Min. # of tests
    - New …

Problem 1: Arbitrary Test Matrix

- Problem 1: find best matrix → min. tests
  - Noiseless Setting
    - Combinatorial Group Testing: Du, Hwang'2000
    - Superimposed codes (Kautz and Singleton'64)
    - Deterministic designs (Dyachkov and Rykov'83)(Ruszinko'94)(Erdos'85)(Ngyuen'88)(Porat'08)
    - Random Designs (Dyachkov'76,'82)(Sebo'85)(Macula'96)
    - Compressed sensing and approximate identification(Gilbert'08)
    - Two-Stage Disjunctive Testing: (Berger-Levenshtein 2002)
  - Noisy Setting/Probabilistic Method
    - Information Theory [Atia-Saligrama 09]
      - K – Defectives → Put nth element in mth pool with prob 1/K
      - Multi-User Comms → Mutual Information formula

Main Result [Atia-S 09]

- Theorem (average error → 0 asymptotically) if:

$$T = \max_i \left\{ \frac{\log \binom{N-K}{i} \binom{K}{i}}{I(X_{(i)}; X_{(K-i)}, Y)} \right\}$$ sufficiency

$$T \geq \frac{\log \binom{N}{K}}{I(X_{(K)}, Y)}$$ Necessity

- X_{mn} generated i.i.d. Bernoulli p
- $X_{(K)}$ corresponds to collection associated with K defective items
- $X_{(i)}$ subset of i defective items in K (that are mis-classified)
- Necessity: FANO Bound

Results: Arbitrary Tests/Pools[Atia-S 09]

- N items; K defectives; T Pools/Tests

$$T = \Theta\left(K\log(N)\right)$$

Noiseless - average error

$$T = \Theta\left(\frac{K\log(N)}{(1-q)}\right)$$

q: false alarm probability

$$T = O\left(\frac{K\log(N)}{(1-s)^2}\right)$$

s: Prob. defective diluted in a pool

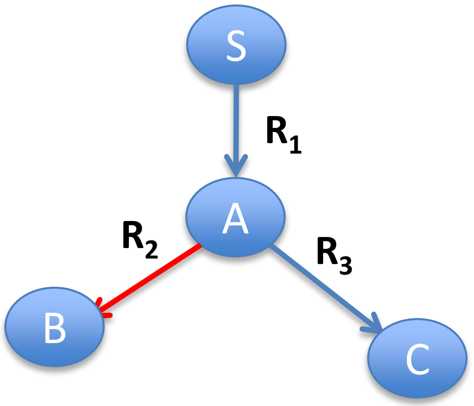
$$T = \Theta\left(K^2\log(N)\right)$$

Noiseless - Exact Reconstruction
Contrast with Compressed Sensing

Problem 2: Structured Test Matrix

- Boolean Congested IP Link Problem
  - link congested if trans. rate is small
  - path congested if at least one link congested

$$Y_m = \bigvee_{n=1}^{N} X_{mn} R_n$$



Links

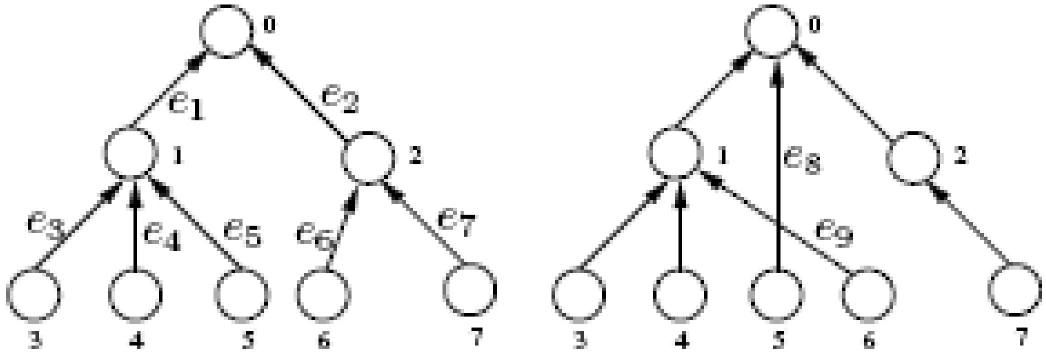
	R_1	R_2	R_3
Path 1	1	1	0
Path 2	1	0	1
	0	1	1

- Summary:
  - Xmn = 1 only when link n is on path m
  - Cannot design test matrix arbitrarily

Structured Tests: Lossy Links in SNETs

- Test Matrix is a Tree

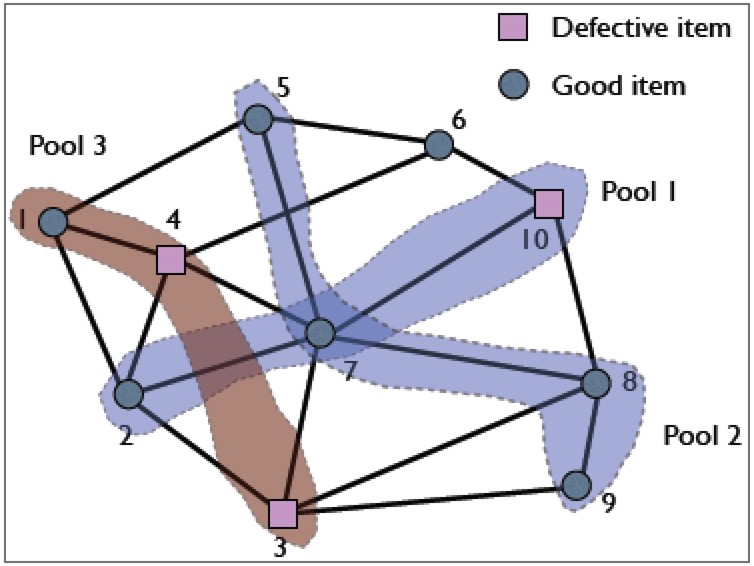
- Different routing tree used each time instant
  - Sensor node failures, ad-hoc trans, …



[Cheraghchi, Karbasi, Mohajer, Saligrama 10]

Graph Constrained Group Testing

- Graph G = (V, E)
- Items:
  - vertices $v \in V$, edges $e \in E$
- Defectives $\leq$ k items
- Pools:
  - items visited by uniform random walk graph of length t starting at source
  - Random Walk (RW) from source to dest.
- Goal:
  - Min. tests reqd to locate defectives

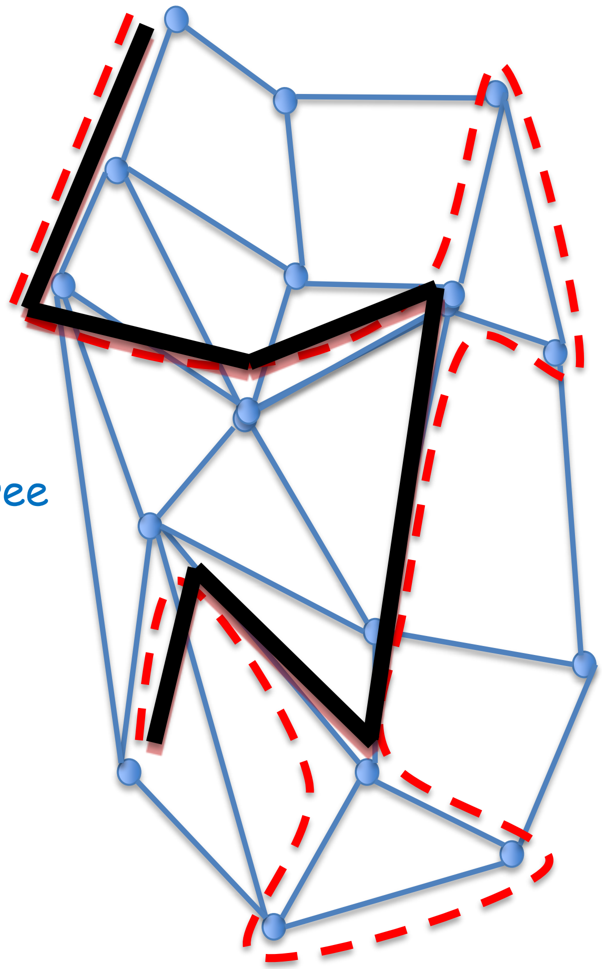


$$P_{ij} = \begin{cases} \frac{1}{d_i} & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

Wilson 96

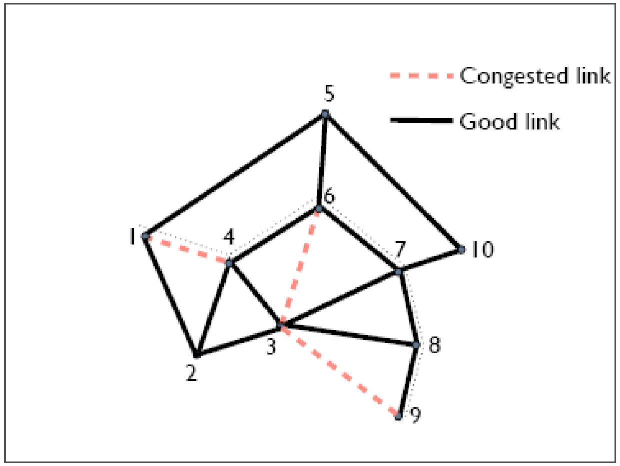
Connection: Loop Erased RW

- Random Rooted Spanning Trees
  - Path generation: erase cycles in the order generated
  - Uniform distribution over all spanning trees
- Stopping time
  - Do same as above for any stopping time
  - Prob that specific rooted tree is a subset of spanning tree
- Our setup
  - Vertices tested only once
    - Our pool is a tree
    - tests are whether a tree contains defectives
  - Similar but more elaborate argument for edges



Solution

- Links (36), (39),(41) are congested
  - When is a test informative ? ``good link" (56)
  - RW passes (56) but not through congested links.
  - Error if there is no path such that this happens for all good links



RW: Random Walk

$\mathcal{A} = \{\text{Congested Links}\}$

$\pi_{v,A} = \text{Prob}(v \in RW, RW \cap \mathcal{A} = \emptyset)$

$Error(v, \mathcal{A}) = (1 - \pi_{v,\mathcal{A}})^T$

Links

	R_1	R_2	R_3
Path 1	1	1	0
Path 2	1	0	1
	0	1	1

Solution: Simple Cases

- N items, K Defectives

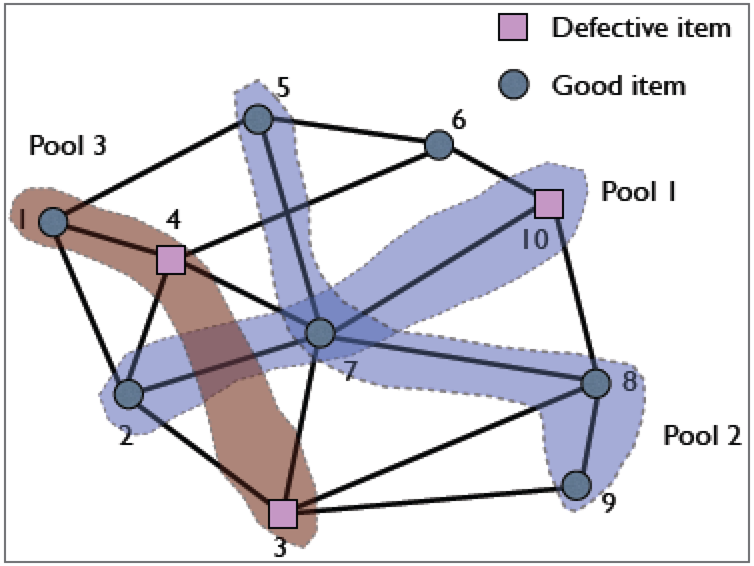
$$Error(v, \mathcal{A}) = (1 - \pi_{v,\mathcal{A}})^T$$

- Completely connected graph
  - Source S, RW Length t= O(N/K)
  - T ≈ O(K log(N)) (avg error)
    - Vertices or edges
  - Source (S) → Destination (D)

$$\pi_{v,\mathcal{A}} = O\left(\frac{1}{K^2}\right)$$

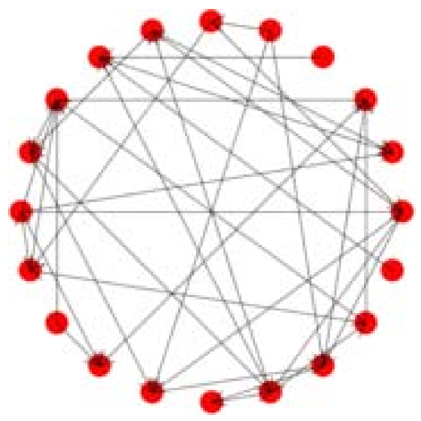
→ T = O(K² log(n))

- Why? Order S→ v → D → A



More Involved Cases

- Erdos-Renyi Graph
  - $G(N,p)$, $p = \log(N)/N$
    - Expander whp
  - Mixing time $\Lambda_N = O(\log(N))$
    - Time to reach ``equilibrium dist''



- Degree $D \in [D_1, D_2]$ graph
  - $D >> d$, Mixing time Λ_N
- Pools:
  - RW of length t from source
  - RW from Source(s) to Destination(s)

Results: Structured Tests [CKM-S 10]

- N items; K defectives; T Pools/Tests (Erdos-Renyi)

$$T = O\left(K \log^2(N)\right)$$

Noiseless - average error

$$T = O\left(\frac{K \log^2(N)}{(1-q)}\right)$$

q: false alarm probability

$$T = O\left(\frac{K \log^2(N)}{(1-s)^2}\right)$$

s: Prob. defective diluted in a pool

$$T = O\left(K^2 \log(N)\right)$$

Source → Destination Pools

Results: Structured Tests [CKM-S 10]

- N items; K defectives; T Pools/Tests ($D \in [D_1, D_2]$)

Λ_N : Mixing Time

$$T = O\left(K\Lambda_N^2 og(N)\right)$$

Noiseless - average error

$$T = O\left(\frac{K\Lambda_N^2 \log(N)}{(1-q)}\right)$$

q: false alarm probability

$$T = O\left(\frac{K\Lambda_N^2 \log(N)}{(1-s)^2}\right)$$

s: Prob. defective diluted in a pool

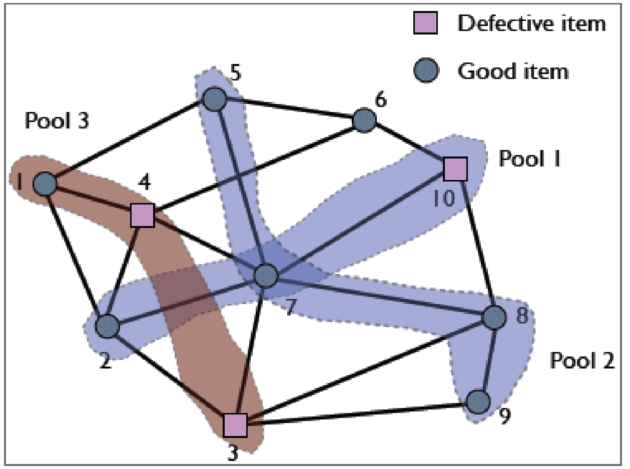
$$T = \Theta\left(K^2\Lambda_N^2 \log(N)\right)$$

Source → Destination Pools

Proof Idea (RW of length t)

- Observe: $Error(v, \mathcal{A}) = (1 - \pi_{v,\mathcal{A}})^T$

- If we want T = O(K Λ_N log(N))
  - we need $\pi_{v,\mathcal{A}} = \Omega\left(\frac{1}{K}\right)$
  - But $\pi_v = \Omega\left(\frac{t}{N\Lambda_N}\right)$
  - Can length of RW? $t = \Omega\left(\frac{N\Lambda_N}{K}\right)$



Proof Idea

- ``Lower bound"

$$\pi_{v,\mathcal{A}} \geq \pi_v(1 - \pi_{\mathcal{A}|v \in RW})$$

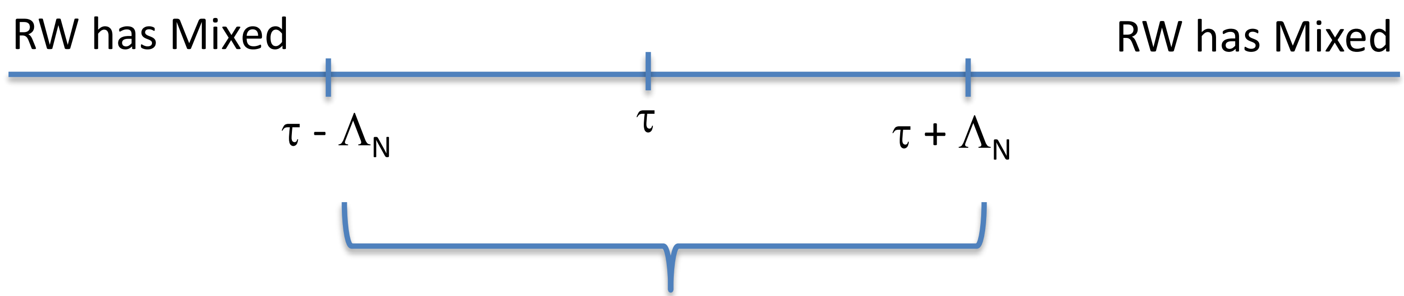
- # visits to v $\leq$ O(Λ_N)
  - Markov inequality

- Upper bound for (not) visiting *A* given each time v visited

Upper bound for visiting A in time t

- Suppose v visited at time τ

$$\pi_{\mathcal{A}|v_\tau=v} \le O\left(\frac{Kt}{N}\right) \qquad\qquad \pi_{\mathcal{A}|v_\tau=v} \le O\left(\frac{Kt}{N}\right)$$



$$\pi_{\mathcal{A}|v_\tau=v} \le O\left(\Lambda_N \frac{d}{D}\right)$$

- But # visits to v can be $O(\Lambda_N)$, so

$$\pi_{\mathcal{A}|v\in RW} = O\left(\Lambda_N \frac{Kt}{N}\right) \Rightarrow t = \Omega\left(\frac{N}{K\Lambda_N}\right)$$

Result

- RW of length t

$$\pi_{\mathcal{A}|v \in RW} = O\left(\Lambda_N \frac{Kt}{N}\right) \Rightarrow t = \Omega\left(\frac{N}{K\Lambda_N}\right)$$

$$\pi_{v,\mathcal{A}} \approx \pi_v = \Omega\left(\frac{t}{N\Lambda_N}\right) = \Omega\left(\frac{1}{K\Lambda_N^2}\right)$$

- RW from source (S) to destination (D)
  - Prob(start at S visit v then D then A)
  - Prob(start at S visit v but not D & A)
  - Prob(start at v visit D but not A)

$$\pi_{v,\mathcal{A}} = \Omega\left(\frac{1}{K^2\Lambda_N^2}\right)$$

Comments (Optimality?)

- Dependence on K defectives:
  - $O(K)$, $O(K^2)$ etc
  - Lower bounds using unstructured tests, complete graphs
- Mixing time ?
  - Appears fundamental.
    - Grids, Cycle
- Can we improve these results?

Extensions 1

- Average error
- Symmetric Graphs
- Rare Events
- Weighted Graphs

Conclusions

- Graph Constrained Group Testing
  - Tests/Pools constrained to be paths
- Pools generated by Random Walks
  - RW of fixed length
  - RW Source to Destination
- Expanders:
  - Fixed Length → # Tests = $O(K \log^2(N))$
  - RW S→ D → #Tests = $O(K^2 \log^2(N))$
- References
  - CKM-S: http://arxiv.org/abs/1001.1445
  - Atia-S: http://arxiv.org/abs/0907.1061