Unconstraining graph-constrained group testing

Bruce Spang*, Mary Wootters

September 20, 2019

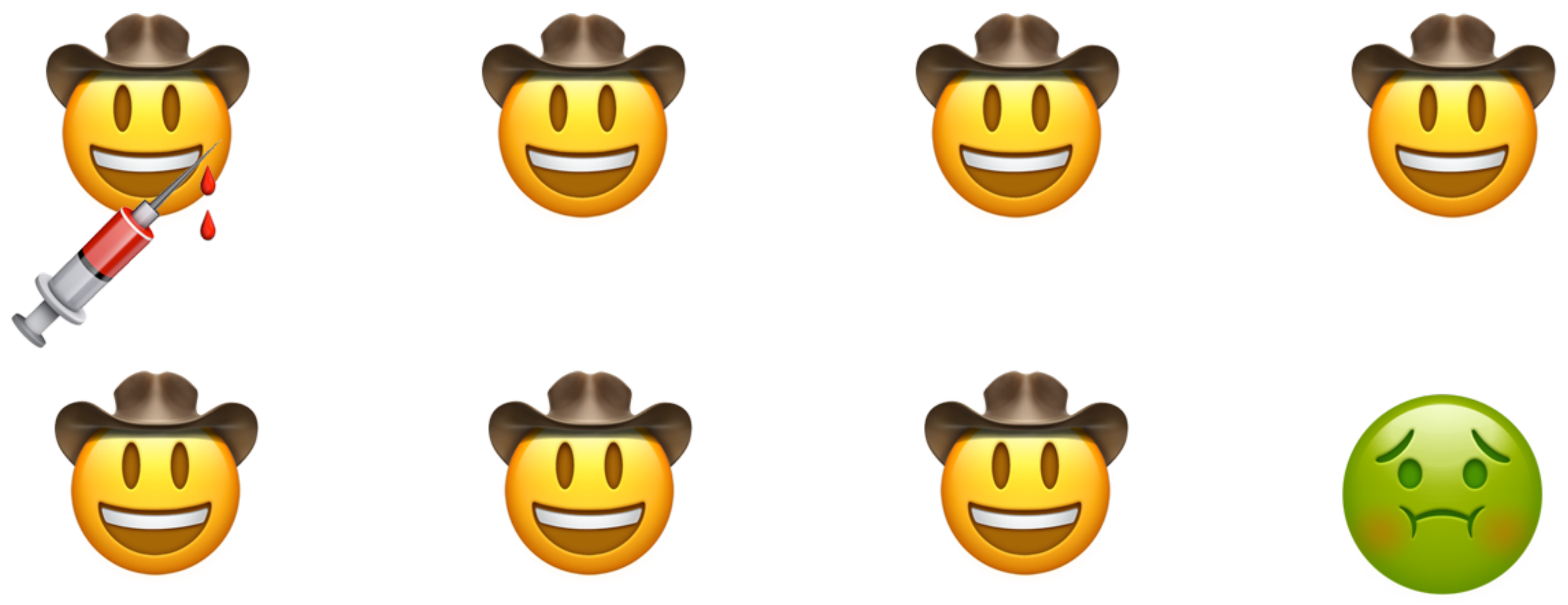
3. Unconstraining
2. graph-constrained
1. group testing

Group Testing

The setting is World War II…

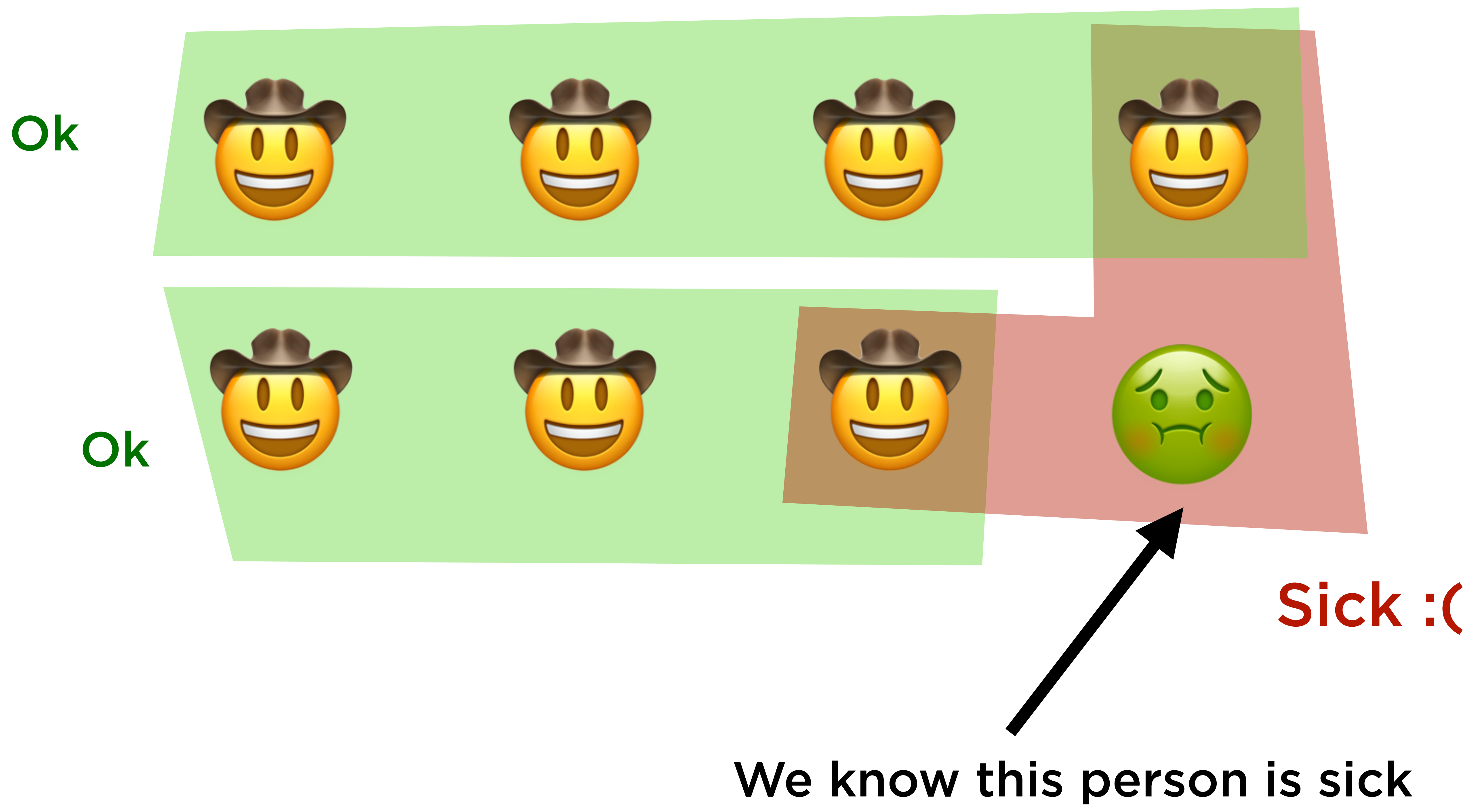
Sick :(

Group Testing

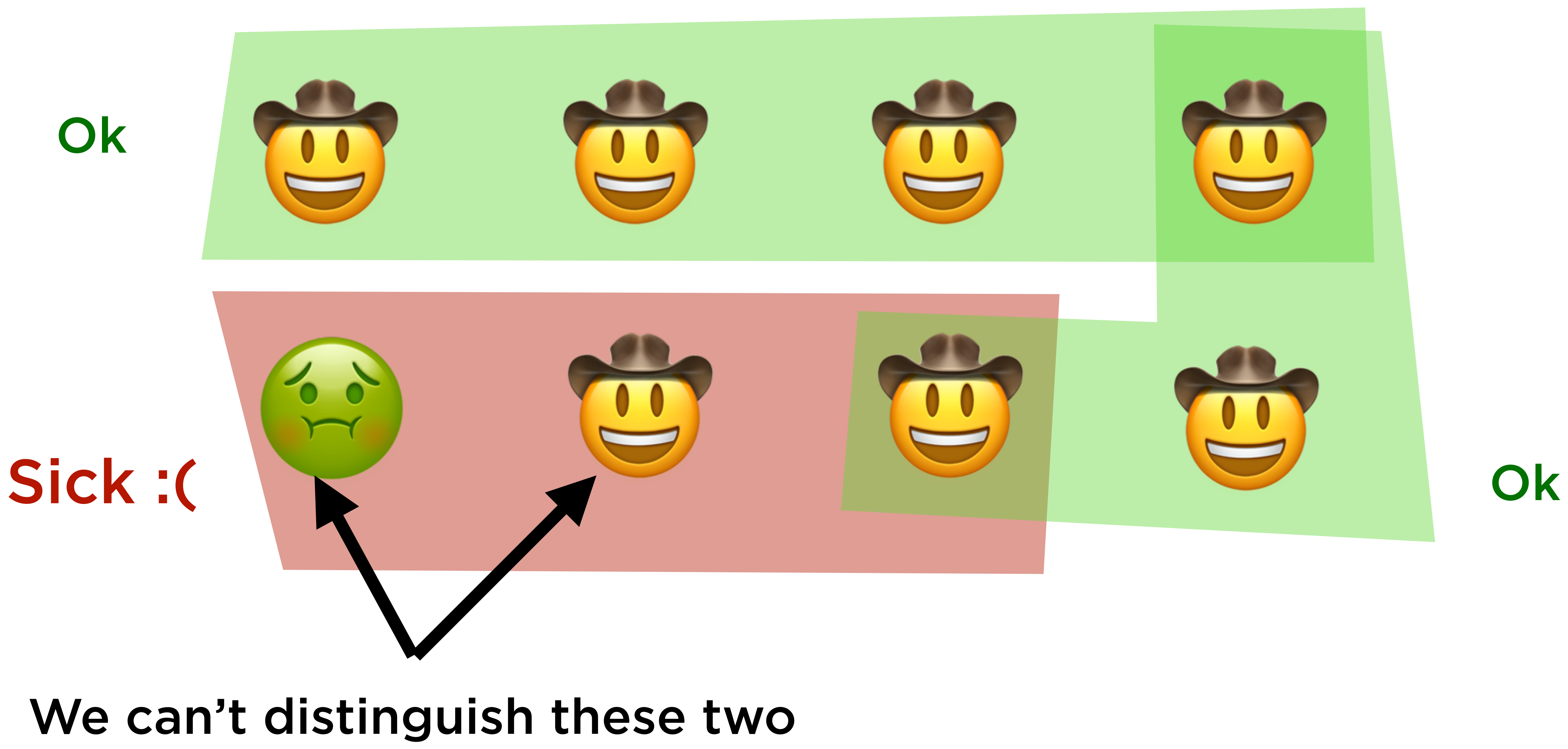


Sick :(

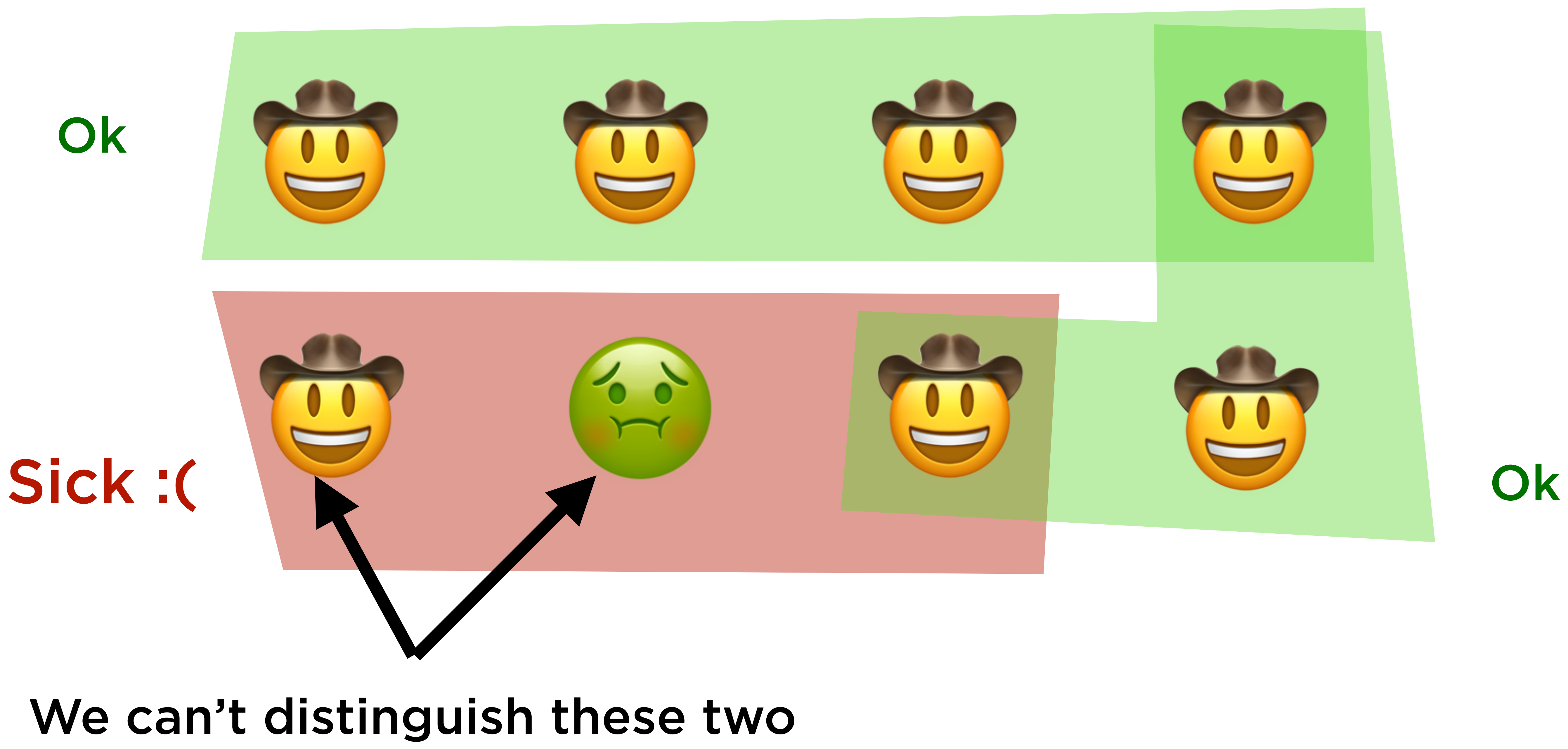
Don’t need individual tests



Need to carefully design tests



Need to carefully design tests



Group Testing Problem

We have m items, at most d of which are defective.

Definition: A test returns whether a subset of items includes any defectives or not.

Problem: Construct a set of tests which can identify any set of at most d defective items.

Some known results

Include each person in a test with probability $1/(d+1)$

$O(d^2 \log m)$ easy construction

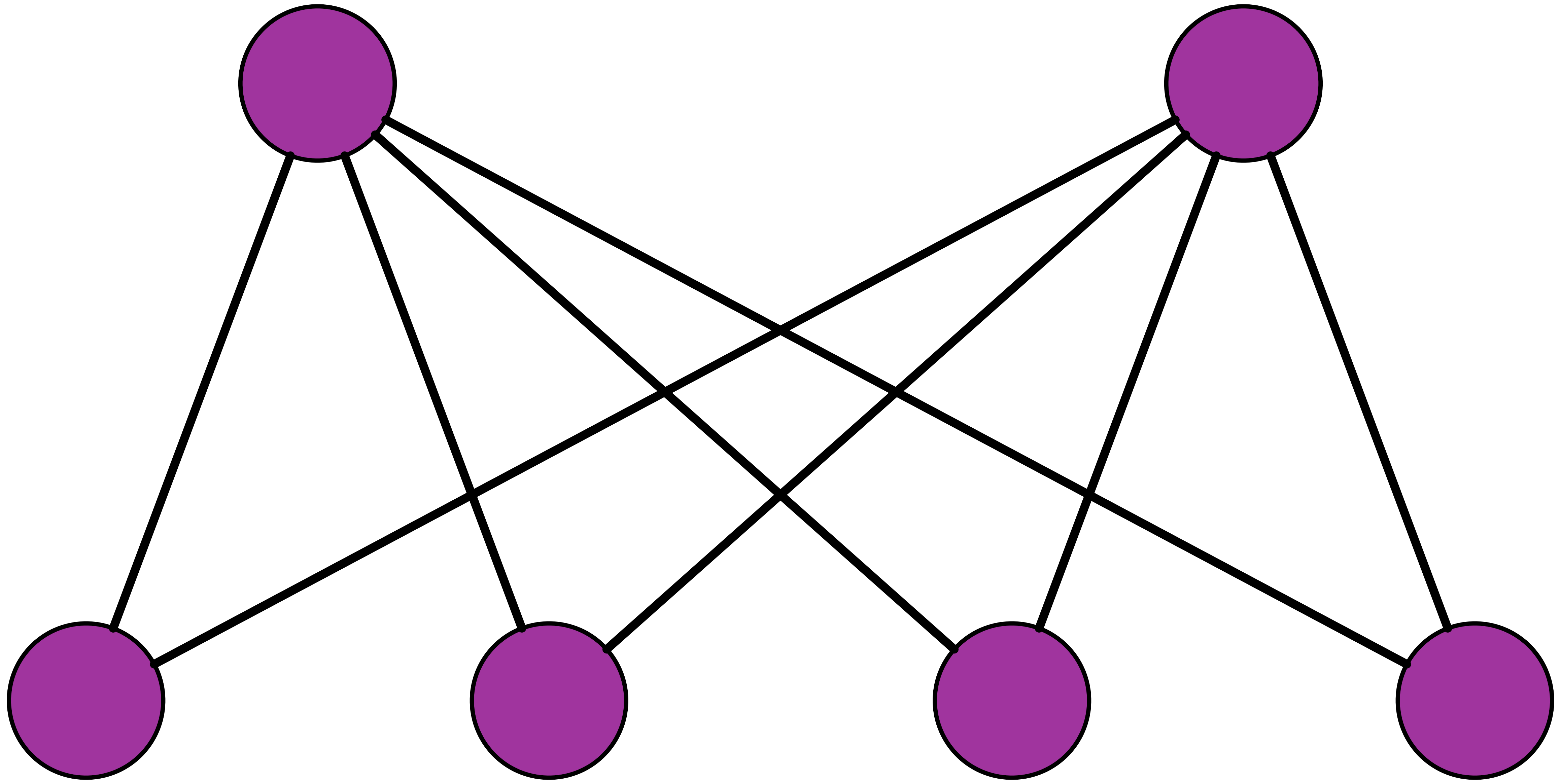
$\Omega(d^2 \log_d m)$ lower bound [Dyachkov-Rykov 82]

Random is pretty close to optimal

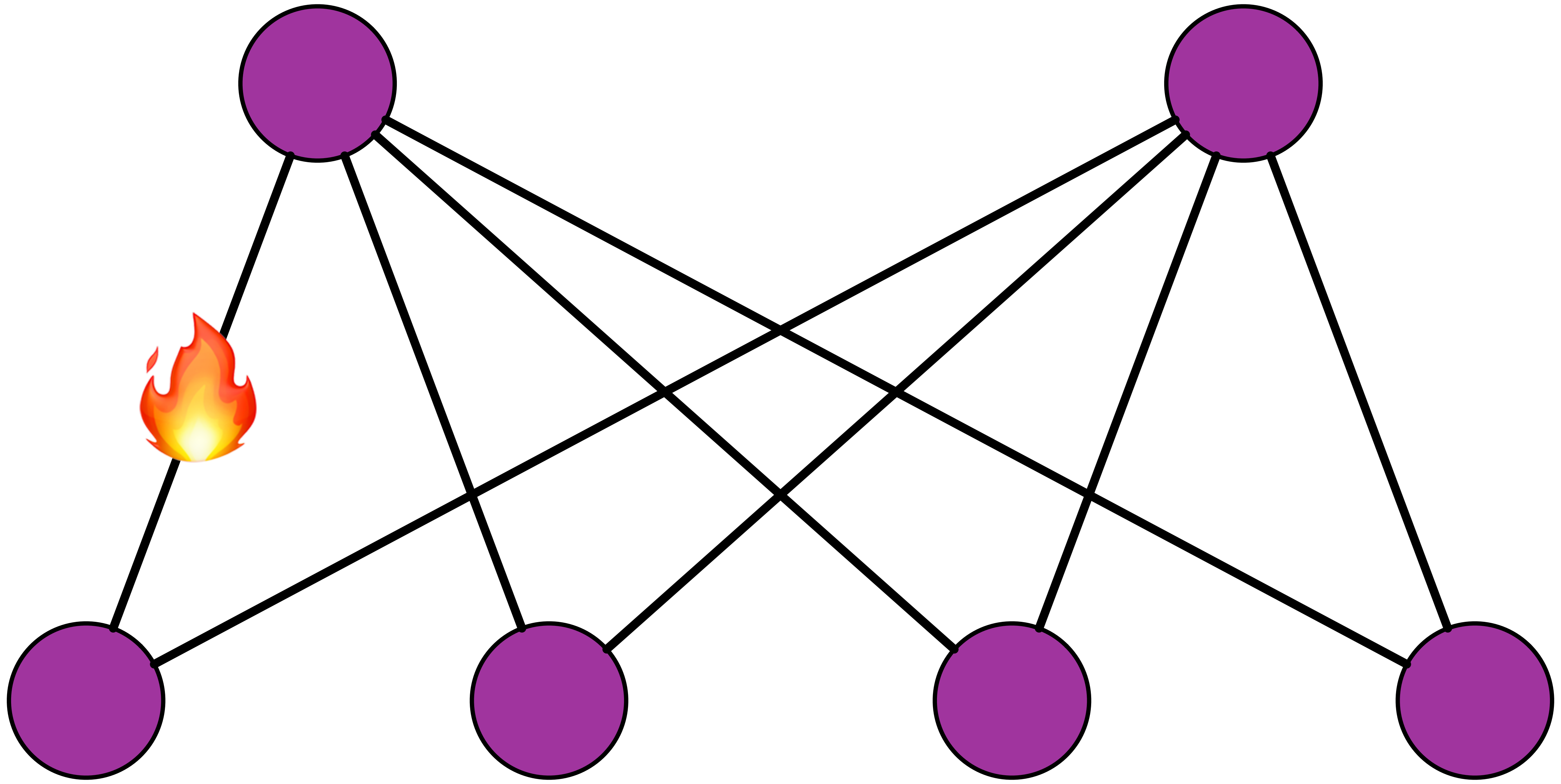
check citation style

3. Unconstraining
2. graph-constrained
1. group testing

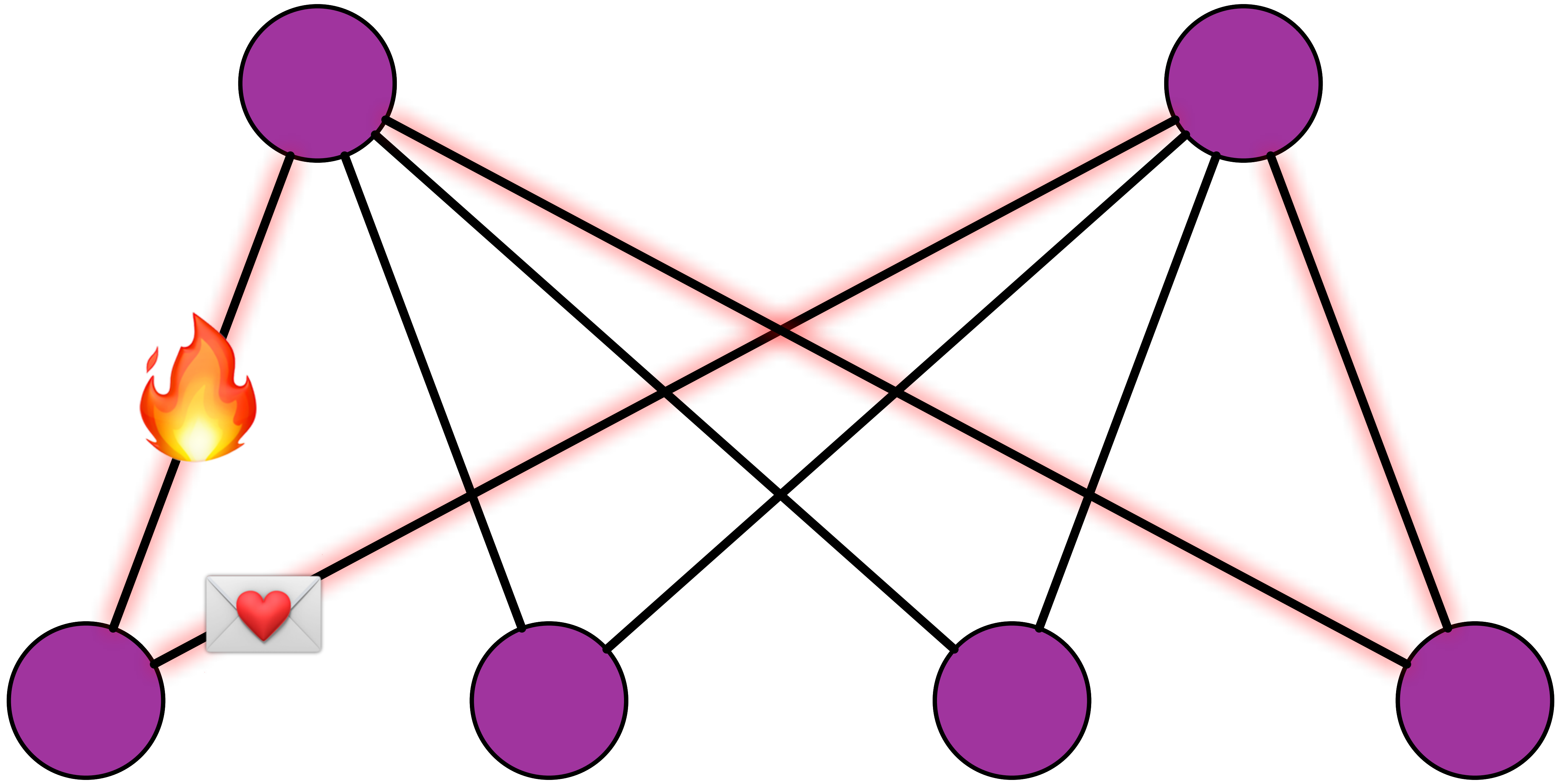
A network



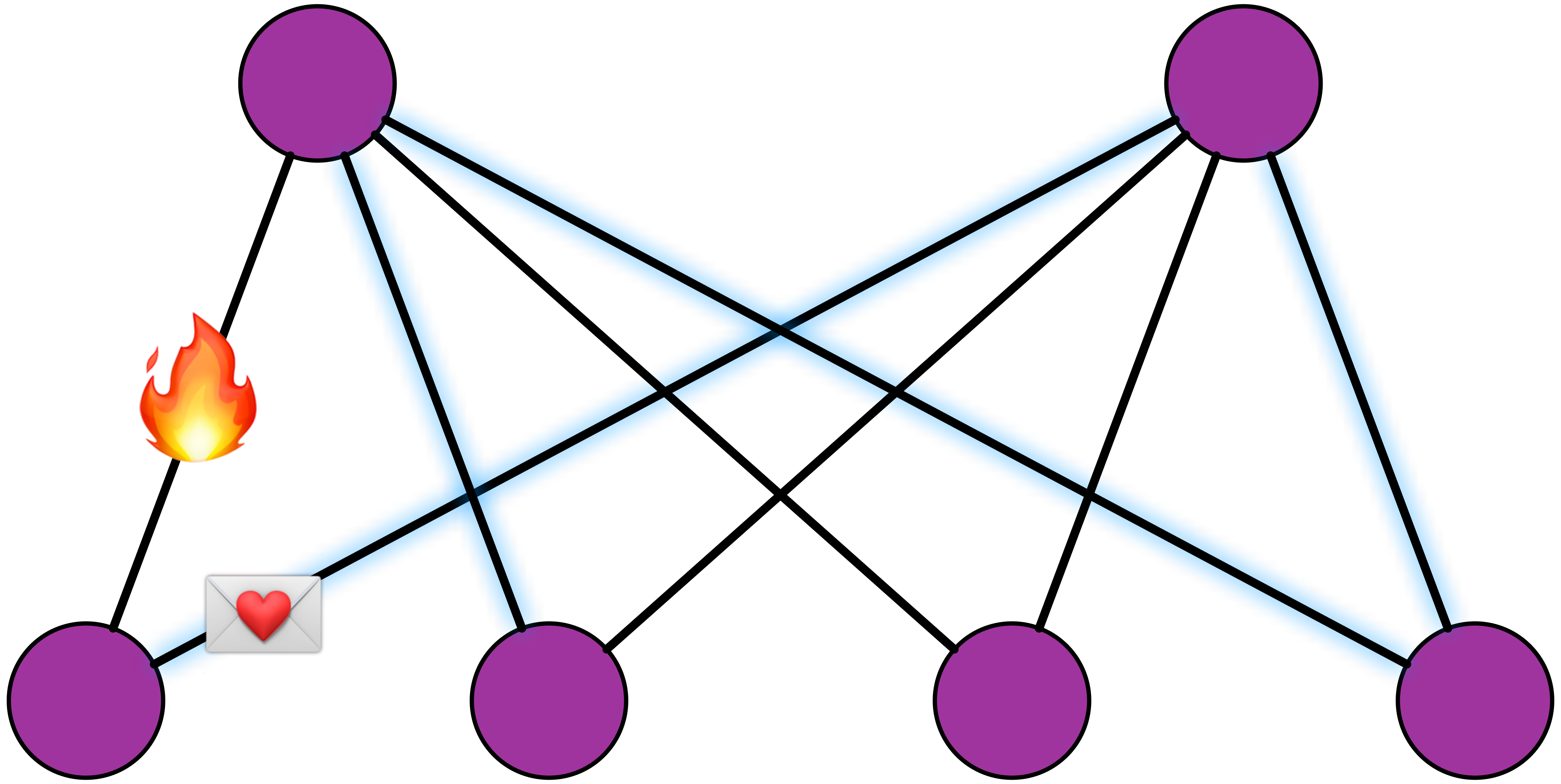
A network, failing



Finding failures



Finding failures



Graph-constrained Problem

We have a graph $G=(V,E)$ with n nodes and m edges, at most d edges are defective.

Definition: A *graph-constrained test* returns whether any edges in a connected subset of edges are defective or not.

Problem: Construct a set of graph-constrained tests which can identify any set of at most d defective edges.

Our informal result

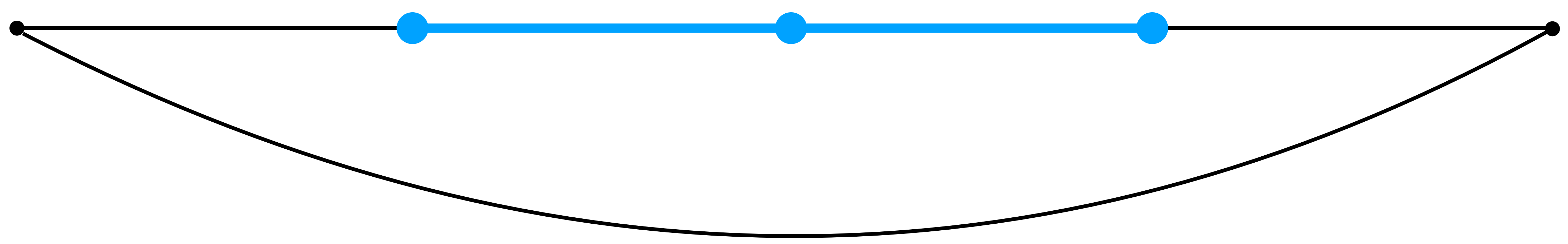
“You can do this nearly-optimally for lots of graphs (more than previously known)”

This seems surprising

For some graphs, these constraints matter a lot

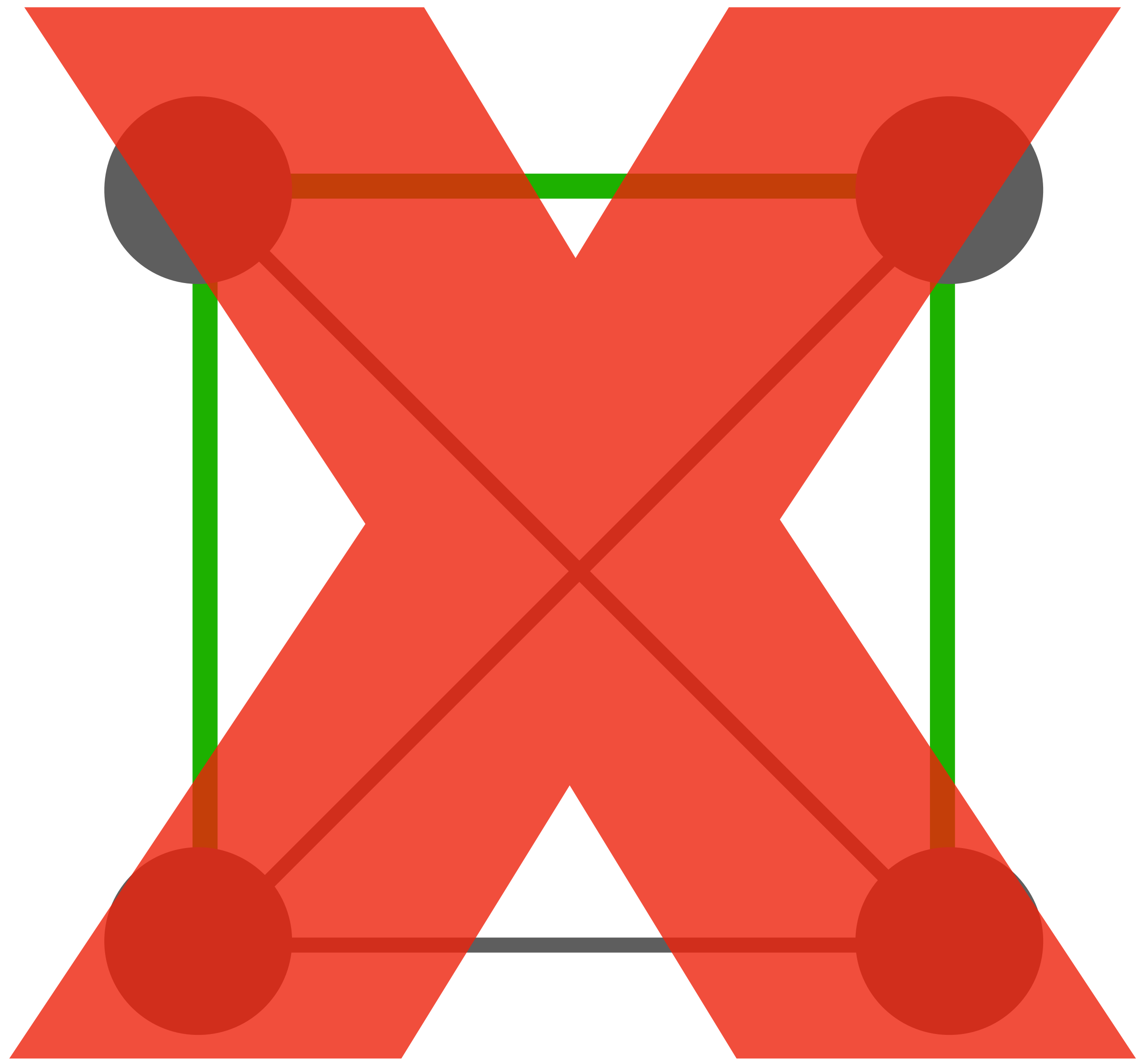
Theorem [Harvey et al 2007]: For the cycle graph on *n* nodes, at least *n*/2 tests required

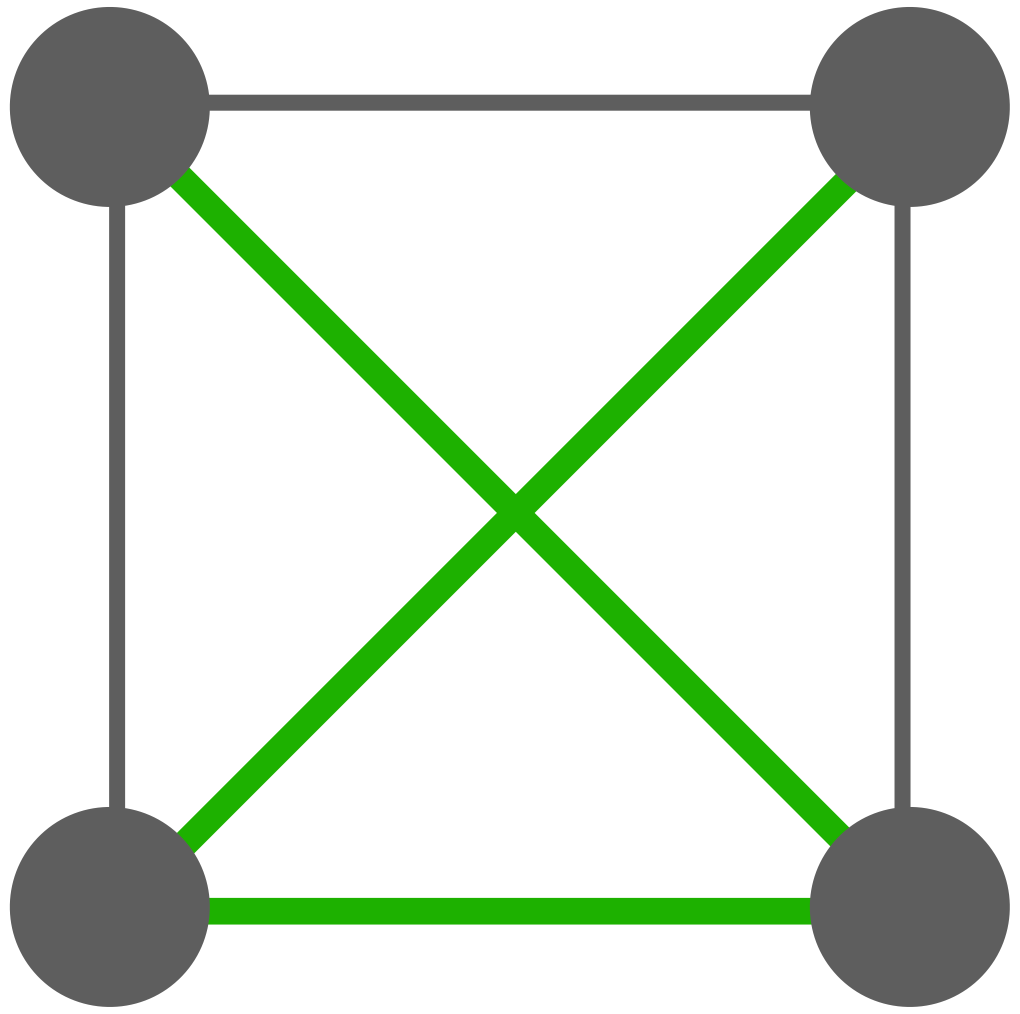
Proof: Each neighboring pair of edges must be separated by some test. Each test is a path and can only separate two pairs. There are about n pairs.

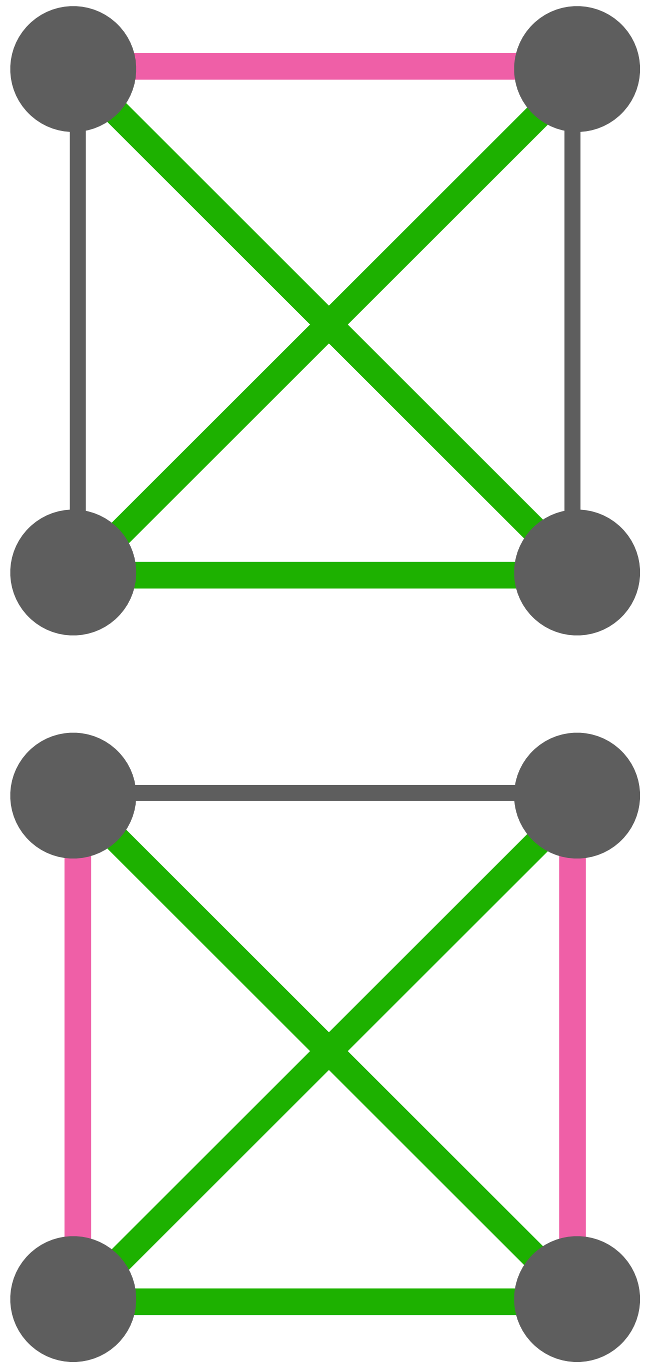


Harvey et al 2007

Most general result: for any graph with more than *d* edge-disjoint spanning trees, can use O(d^3 log m) tests to identify at most *d* defective edges.

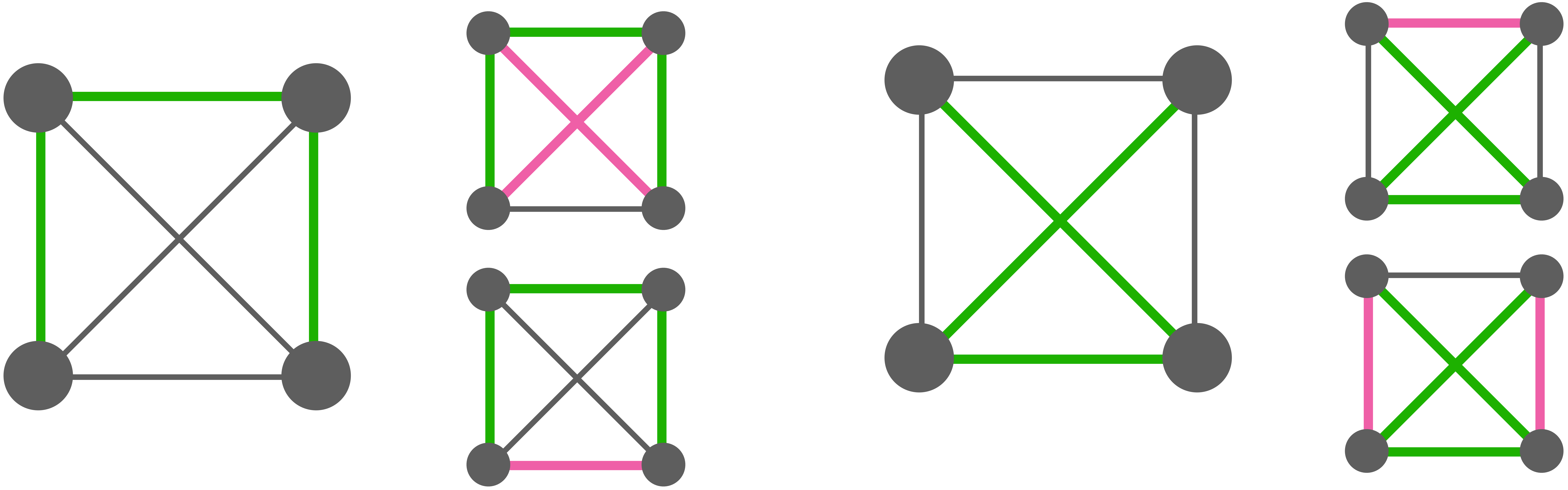






Harvey et al 2007

Most general result: for any graph with more than *d* edge-disjoint spanning trees, can use O(d^3 log m) tests to identify at most *d* defective edges.



Cheraghchi et al 2010

- Current state of the art
- Each test is a random walk on the graph
- For certain graphs, can do it in $O(\tau^2 d^2 \log m)$ tests!

D-regular graphs, $D \geq 6d \log^2 n$

Cheraghchi et al 2010

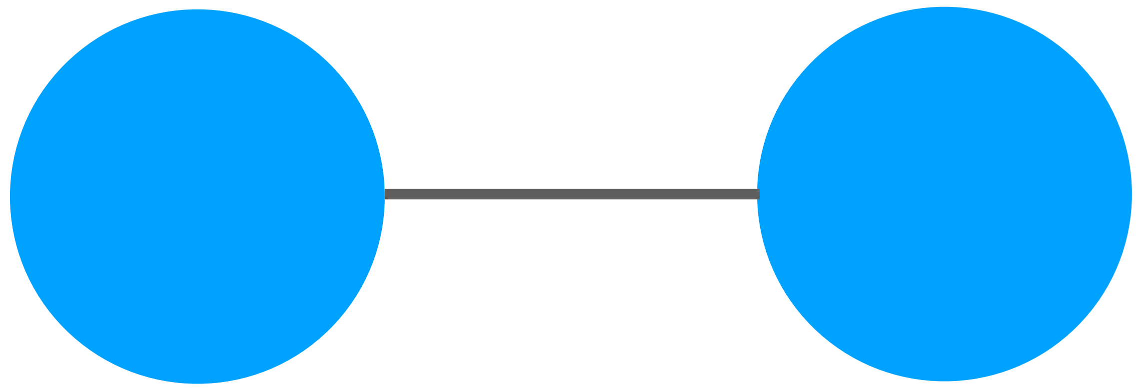
Good parts

- Optimal for a complete graph!
- Good expanders are nearly optimal: off by $O(\log^2 m)$

Limitations

- Degree requirement of log n means it can't deal with constant-degree expanders
- Barbell feels like it should work but doesn't:

n/2



n/2

Summary

Problem: group testing, each test is connected subgraph

Lower bound: $\Omega(d^2 \log_d m)$

Gaps:

- Constant mixing time: none
- Expanders: $O(\log^2 m)$
- Barbell: $O(m)$

3. Unconstraining
2. graph-constrained
1. group testing

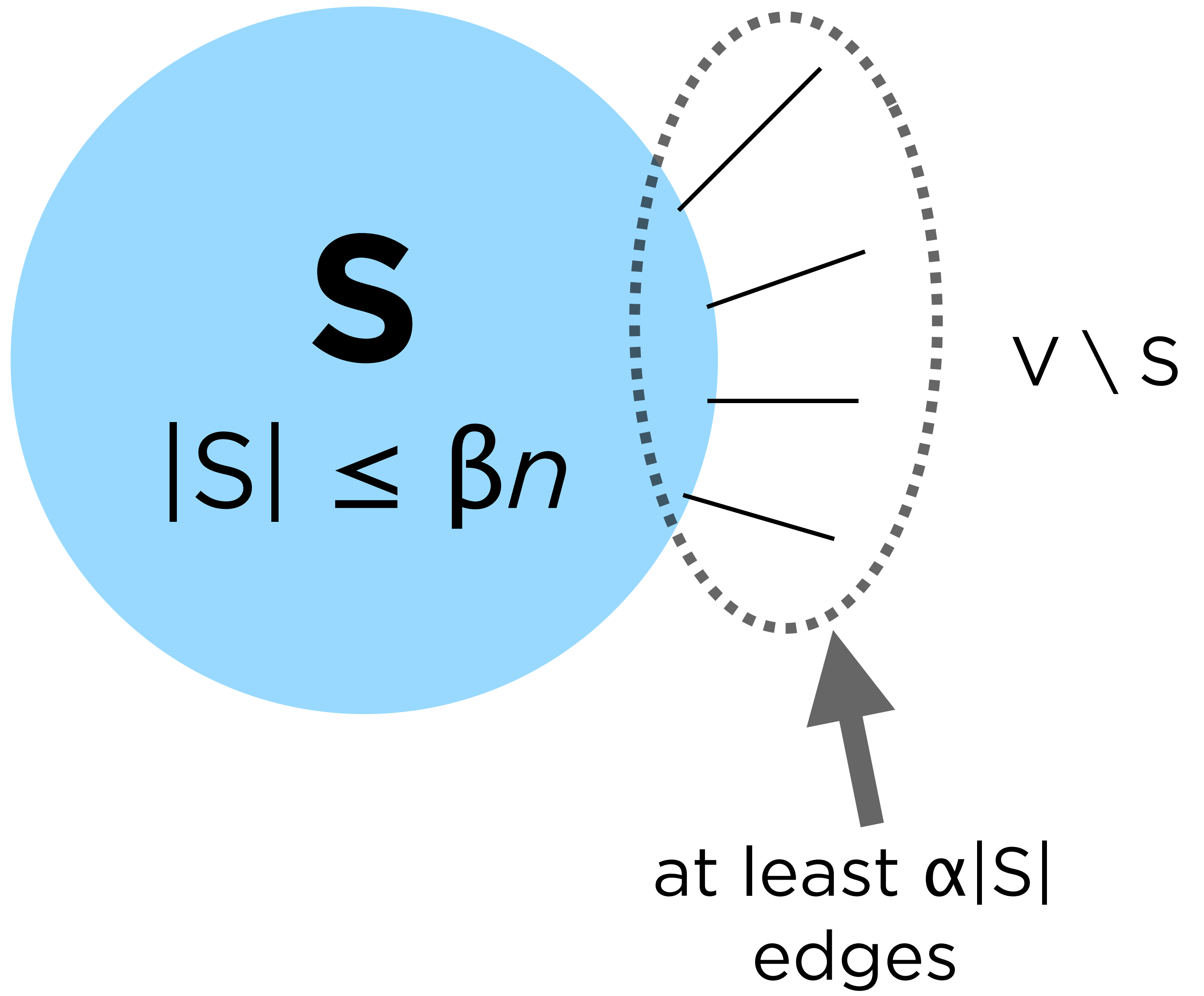
Informal result

If a graph *sufficiently well-enough connected*, we can find any set of *d* defective edges using $O(d^2 \log m)$ tests

Same as *unconstrained* group testing

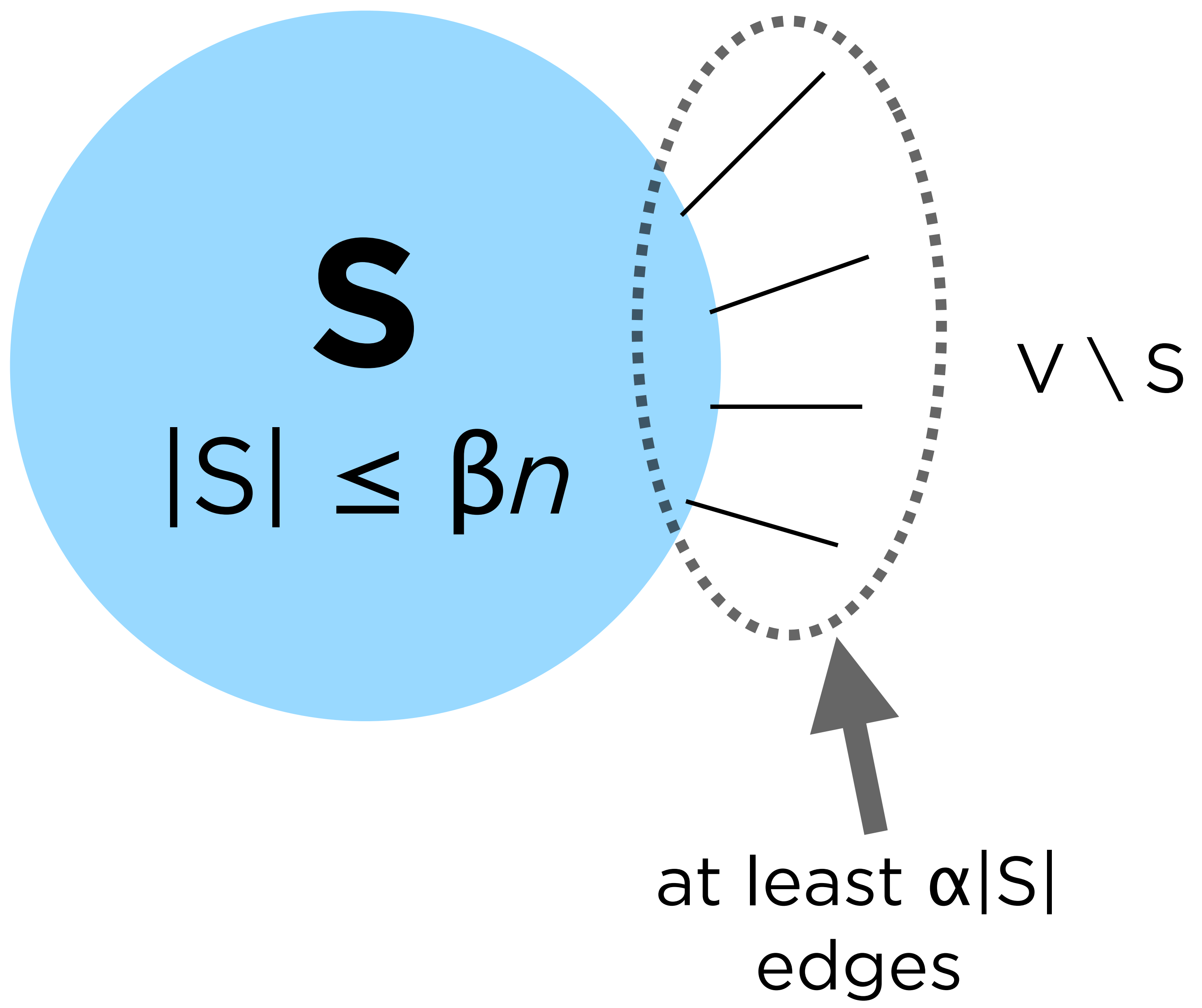
(β,α)-expanders

All sets S of size at most βn have a boundary of at least α|S| edges



Examples of (β,α)-expanders

Graph	β	α
Edge expander	1/2	
Complete	1/2	n/2
Barbell	1/4	n/4



Main Theorem

Let $G = (V,E)$, $|V| = n$, $|E| = m$ be a (β,α)-expander, and

$d \geq 0$ where

$$\alpha \geq d/2 + O(1).$$

Then there exists a set of $O(\beta^{-1}d^2 \log m)$ tests that

identify any set of at most d defective edges.

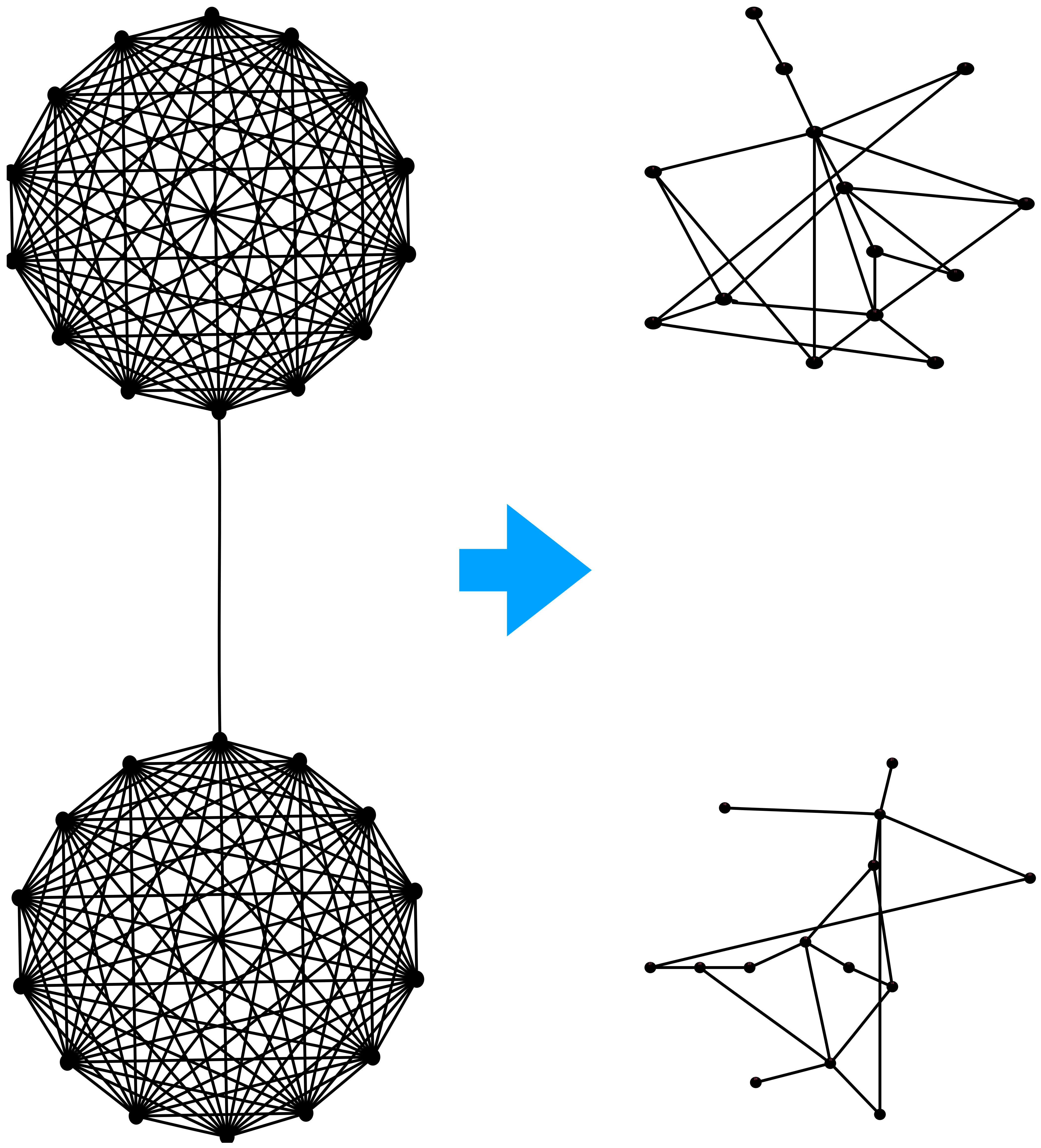
Special Cases

Graph	Source	Number of Tests	At most $d \le d_0$ failures
Complete	**[CKMS10]**	**$O(d^2 \log m)$**	**$d_0 = \Omega(m)$**
	Our work	**$O(d^2 \log m)$**	**$d_0 = \Omega(m)$**
D-regular expander	[HPW+07]	$O(d^3 \log m)$	**$d_0 = \Omega(D)$**
	[CKMS10]	$O(d^2 \log^3 m)$	$d_0 = \Omega(D/\log^2 m)$
	Our work	**$O(d^2 \log m)$**	**$d_0 = \Omega(D)$**
Erdös-Rényi Graph G(n,D/n)	[CKMS10]	$O(d^2 \log^3 m)$	$d_0 = \Omega(D/\log^2 m)$
	Our work	**$O(d^2 \log m)$**	**$d_0 = \Omega(D)$**
Barbells	[HPW+07]	$O(d^3 \log m)$	$d_0 = 1$
	[CKMS10]	$O(m)$	**$d_0 = m$**
	Our work	**$O(d^2 \log m)$**	**$d_0 = \Omega(m)$**

Algorithm

For 1...T:

- Include each edge with probability p ~ 1/d
- Use connected components larger than β*n*



Proof outline

- Recall from earlier, if we just pick edges with probability ~1/d, we win if the resulting graph is connected.

- If K ~ d log n, just pick each edge with probability ~1/d, the resulting graph is connected by [Karger 94]

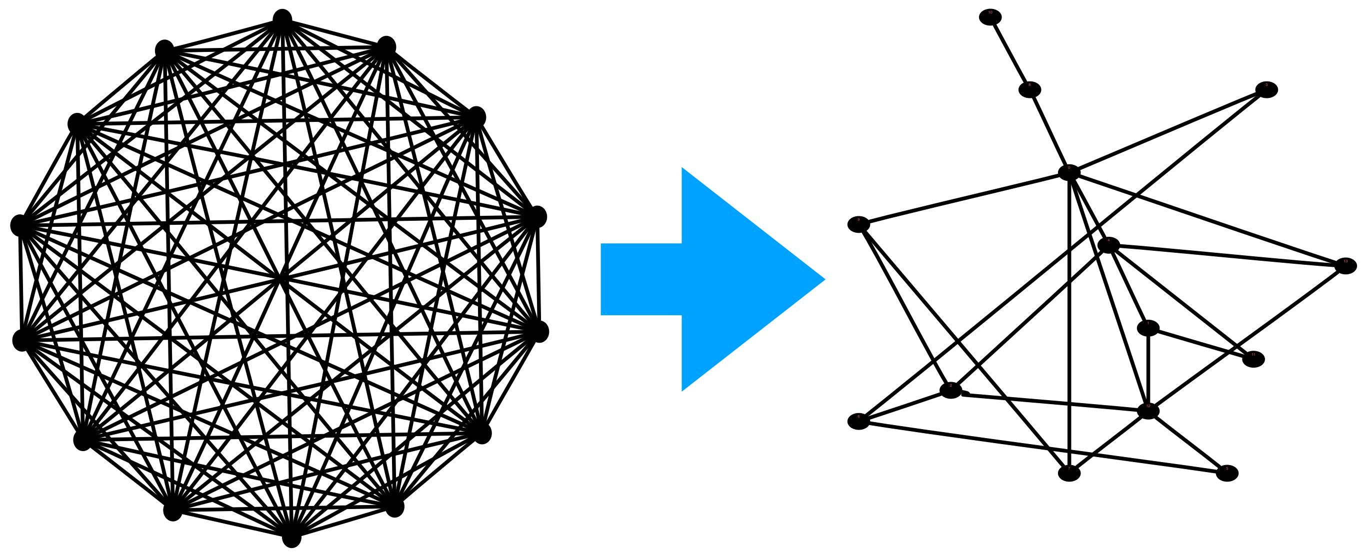
- We show *most* of the graph is connected when we pick each edge with probability ~1/d

Giant components

Model: Fix a graph, keep each edge independently with probability *p.*

Lots of previous work shows large connected components *exist* above some value of *p:*

- [Erdös-Rènyi 59] $p = \frac{1+\epsilon}{n}$
- Expanders $p = \frac{1+\epsilon}{\alpha}$



We show something stronger:

- *For each edge*, the probability the edge is picked and included in a giant component is at least pε/8

Open problems

- Result for an arbitrary graph (start with a hypercube)
- Find a deterministic algorithm

Thank you!

Technical Result

Let *G*=(*V*,*E*) be a (β,α)-expander, 0 < p < 1, *G*(*p*) be the subgraph of *G* constructed by including each edge independently with probability *p*, and *C*(*u*,*v*) be the connected component of *G*(*p*) including edge (*u*,*v*).

If $p \geq \frac{1+\epsilon}{\alpha}$,

then for all $(u, v) \in E$

$$\mathbb{P}(|C(u,v)| \geq \beta n) \geq \epsilon/2$$