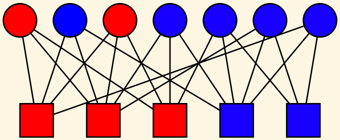
Group Testing

Amin Coja-Oghlan

TU Dortmund, Informatik 2

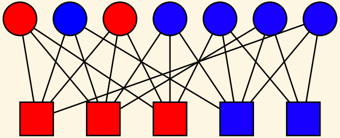
The problem



Group testing [D43,DH93]

- n =population size, $k = n^\theta$ = #infected, m = #tests
- all tests are conducted in parallel
- how many tests are necessary to identify the infected?

The problem

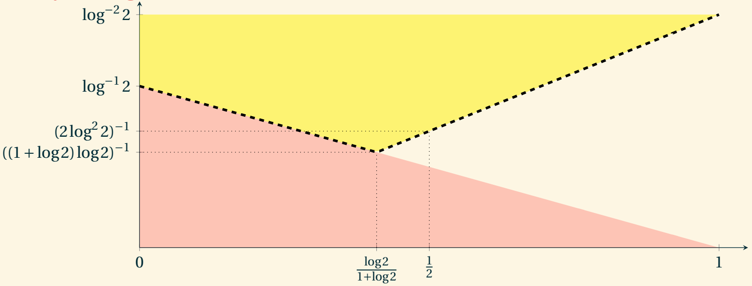


Impossible–hard–easy

depending on the number of tests, the task may be

- information-theoretically impossible
- possible but computationally “hard”
- computationally easy

Random hypergraphs



Theorem

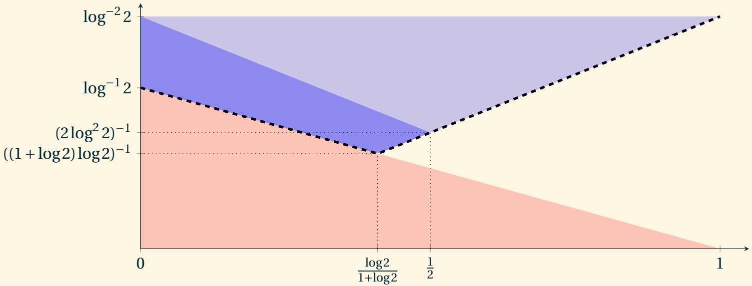
Let

$$m_{\mathrm{rnd}} = \max\left\{\frac{1-\theta}{\log 2}, \frac{\theta}{\log^2 2}\right\} k \log n \qquad \text{where} \quad k \sim n^{\theta}$$

The inference problem on the random hypergraph

- is insoluble if $m < (1-\varepsilon)m_{\mathrm{rnd}}$ [JAS16]
- reduces to hypergraph VC if $m > (1+\varepsilon)m_{\mathrm{rnd}}$ [COGHKL19]

The SPIV algorithm

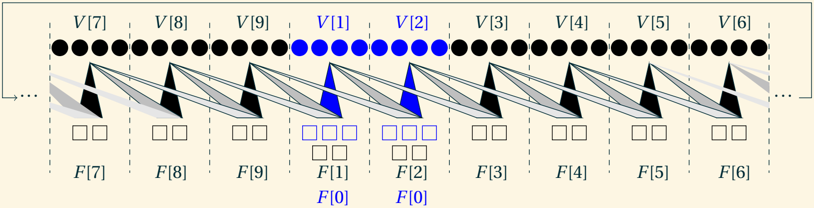


Theorem [COGHKL19]

There exist a test design and an efficient algorithm SPIV that succeed w.h.p. for

$$m \sim m_{\mathrm{rnd}} = \max\left\{\frac{1-\theta}{\log 2}, \frac{\theta}{\log^2 2}\right\} k \log n$$

The SPIV algorithm

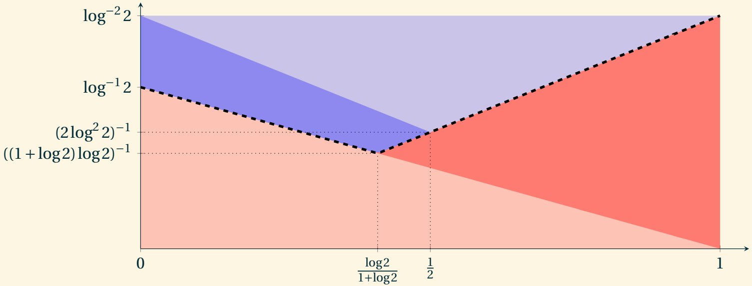


Spatial coupling

- a ring comprising $1 \ll \ell \ll \log n$ compartments
- individuals join tests within a sliding window of size $1 \ll s \ll \ell$
- extra tests at the start facilitate DD
- algorithm based on Belief Propagation

inspired by low-density parity check codes *[KMRU10]*

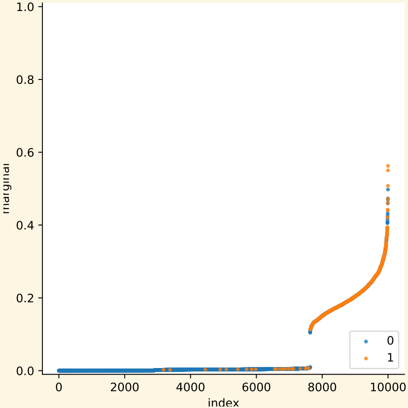
A matching lower bound



Theorem [COGHKL19]

Identifying the infected individuals is information-theoretically impossible with $(1-\varepsilon)\, m_{\mathrm{rnd}}$ tests.

Experiments



- Belief Propagation posteriors
- orange: infected; blue: healthy
- $n = 10^4$, $m = 1600$, $k = 500$, $\Delta = 2$
- false positive rate 0.01; false negative rate 0.02

Summary

- a randomised construction
- tight information-theoretic and algorithmic bounds
- inference via Belief Propagation