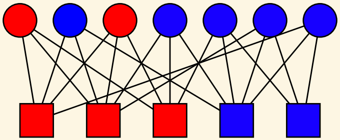
Disordered systems and random graphs 3

Amin Coja-Oghlan
Goethe University

based on joint work with Dimitris Achlioptas, Oliver Gebhard, Max Hahn-Klimroth, Joon Lee, Philipp Loick, Noela Müller, Manuel Penschuck, Guangyan Zhou

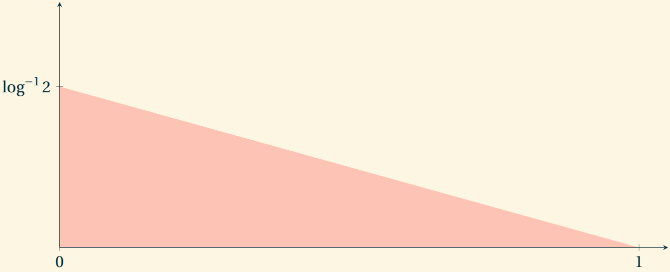
The problem



Group testing [D43,DH93]

- n =population size, $k = n^\theta$ = #infected, m = #tests
- all tests are conducted in parallel
- how many tests are necessary...
- ...information-theoretically?
- ...algorithmically?

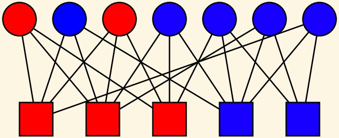
Information-theoretic lower bounds



- if $k \sim n^{\theta}$ we need

$$2^m \geq \binom{n}{k} \qquad \Rightarrow \qquad m \geq \frac{1-\theta}{\log 2} \cdot k \log n$$

Random hypergraphs



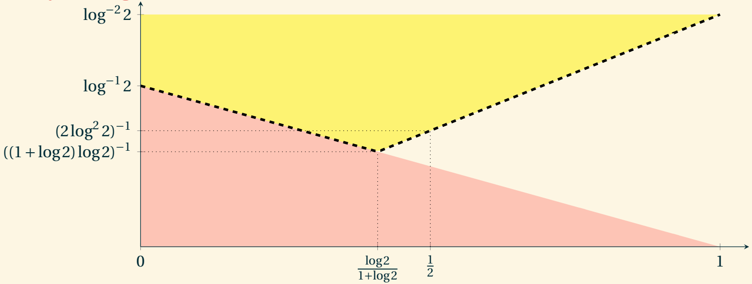
A randomised test design [JAS16,A17]

- a random Δ-regular Γ-uniform hypergraph with

$$\Delta \sim \frac{m \log 2}{k}, \qquad \Gamma \sim \frac{n \log 2}{k}$$

- the choice of Δ, Γ maximises the entropy of the test results

Random hypergraphs



Theorem

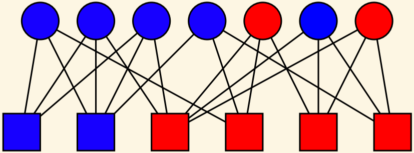
Let

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

The inference problem on the random hypergraph

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

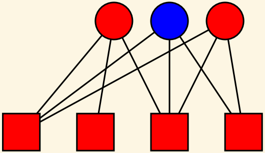
Greedy algorithms



DD: Definitive Defectives [ABJ14]

- declare all individuals in negative tests uninfected
- check for positive tests with just one undiagnosed individual
- declare those individuals infected
- declare all others uninfected
- $\rightsquigarrow$ *may produce false negatives*

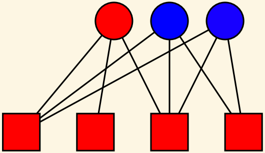
Greedy algorithms



DD: Definitive Defectives [ABJ14]

- declare all individuals in negative tests uninfected
- check for positive tests with just one undiagnosed individual
- declare those individuals infected
- declare all others uninfected
- $\rightsquigarrow$ *may produce false negatives*

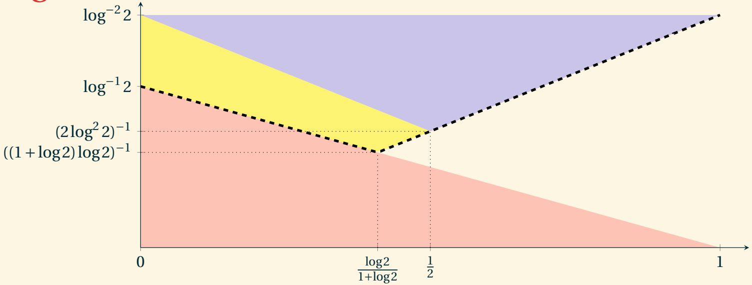
Greedy algorithms



DD: Definitive Defectives [ABJ14]

- declare all individuals in negative tests uninfected
- check for positive tests with just one undiagnosed individual
- declare those individuals infected
- declare all others uninfected
- $\rightsquigarrow$ *may produce false negatives*

Greedy algorithms



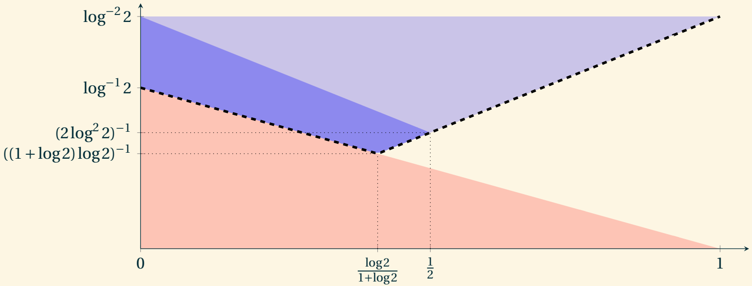
Theorem

Let

$$m_{\mathrm{DD}} = \frac{\max\{1-\theta, \theta\}}{\log^2 2} k \log n$$

- if $m > (1+\varepsilon) m_{\mathrm{DD}}$, then both DD succeeds [ABJ14]
- if $m < (1-\varepsilon) m_{\mathrm{DD}}$, then DD and other algorithms fail [COGHKL19]

The SPIV algorithm

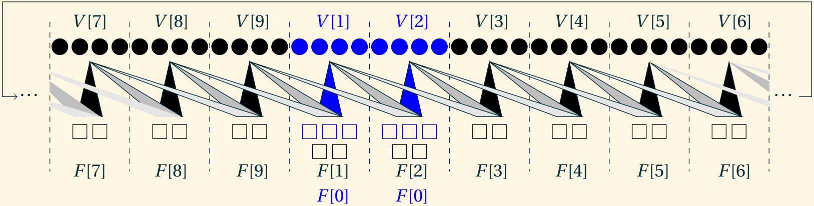


Theorem [COGHKL19]

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

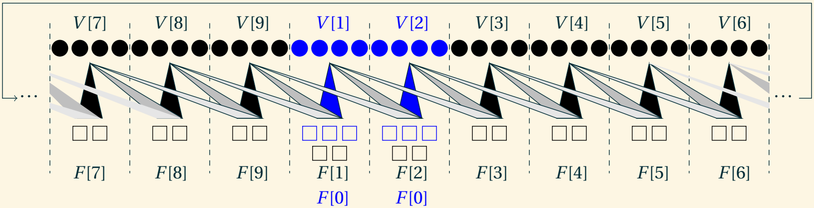


Spatial coupling

- 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

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

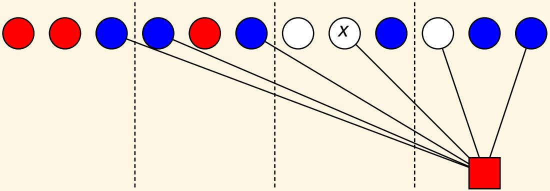
The SPIV algorithm



The algorithm

- run DD on the s seed compartments
- declare all individuals that appear in negative tests uninfected
- tentatively declare infected k/ℓ individuals with max score W_x
- combinatorial clean-up step

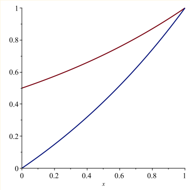
The SPIV algorithm

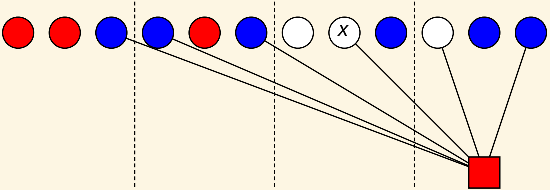


Unexplained tests

- let $W_{x,j}$ be the number of 'unexplained' positive tests $j-1$ compartments to the right of x

The SPIV algorithm

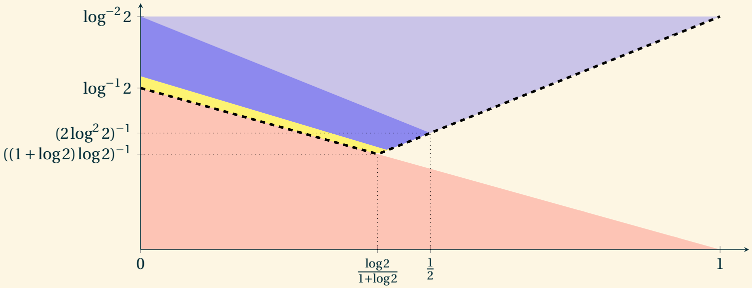




Unexplained tests

- if x is infected, then $W_{x,j} \sim \mathrm{Bin}(\Delta/s, 2^{j/s-1})$
- if x is uninfected, then $W_{x,j} \sim \mathrm{Bin}(\Delta/s, 2^{j/s} - 1)$

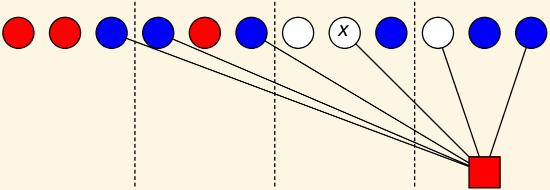
The SPIV algorithm



The score: first attempt

- just count unexplained tests
- we find the large deviations rate function of $\sum_{j=1}^{s-1} W_{x,j}$
- unfortunately, we will likely misclassify $\gg k$ individuals

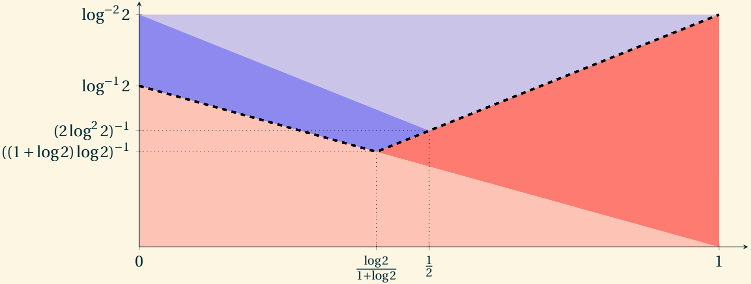
The SPIV algorithm



The score: second attempt

- consider a weighted sum $W_x = \sum_{j=1}^{s-1} w_j W_{x,j}$
- Lagrange optimisation $\rightsquigarrow$ optimal weights $w_j = -\log(1 - 2^{-j/s})$
- only $o(k)$ misclassifications

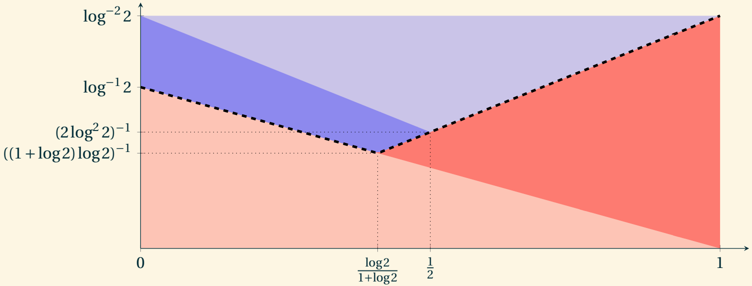
A matching lower bound



Theorem [COGHKL19]

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

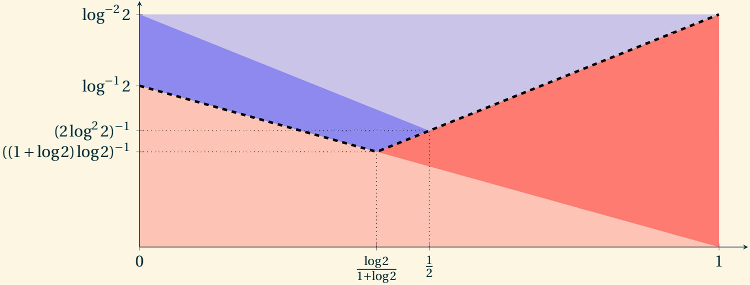
A matching lower bound



Proof strategy

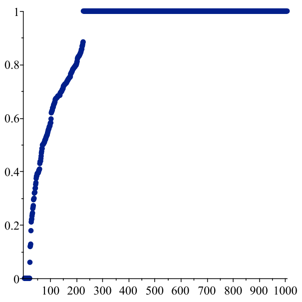
- *Dilution:* it suffices to consider $\theta = 1 - \delta$
- *Regularisation:* optimal designs are approximately regular
- *Positive correlation:* probability of being disguised [MT11,A18]
- *Probabilistic method:* disguised individuals likely exist

Group testing: summary



- optimal efficient algorithm `SPIV` based on spatial coupling
- matching information-theoretic lower bound
- existence of an adaptivity gap

Linear group testing via Belief Propagation



Linear group testing

- non-adaptive testing impossible when $k = \Theta(n)$ [A19]
- Belief Propagation leads to a promising multi-stage scheme
- currently only experimental results

References

- M. Aldridge: Individual testing is optimal for nonadaptive group testing in the linear regime. IEEE Trans Inf Th **65** (2019)
- M. Aldridge, O. Johnson, J. Scarlett: Group testing: an information theory perspective (2019)
- A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, P. Loick: Optimal group testing. COLT 2020
- D. Donoho, A. Javanmard, A. Montanari: Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. IEEE Trans Inf Th **59** (2013)
- R. Dorfman: The detection of defective members of large populations. Annals of Mathematical Statistics **14** (1943)
- S. Kudekar, T. Richardson, R. Urbanke: Spatially coupled ensembles universally achieve capacity under Belief Propagation. IEEE Trans Inf Th **59** (2013)
- P. Zhang, F. Krzakala, M. Mézard, L. Zdeborová: Non-adaptive pooling strategies for detection of rare faulty items. ICC 2013