Information theoretic and algorithmic aspects of binary and quantitative group testing in the sublinear regime

Oliver Gebhard

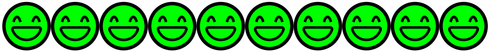
Goethe-University Frankfurt

gebhard@math.uni-frankfurt.de

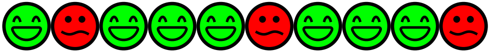
Joint work with
A. Coja-Oghlan, M.Hahn-Klimroth, P. Loick
and
M.Hahn-Klimroth, D. Kaaser, P. Loick

Overview

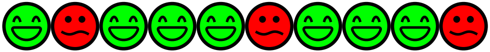
Basic setup



Basic setup

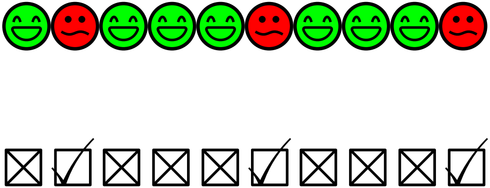


Basic setup

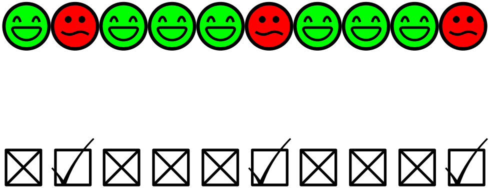


- **Goal**:Are we able to identify the sick individuals?

Basic setup



Basic setup



- **Goal**:Are we able to reduce the number of tests?

Basic setup

Basic setup

- Binary: Is a sick individual contained?
- Quantitative: Number of sick individuals?

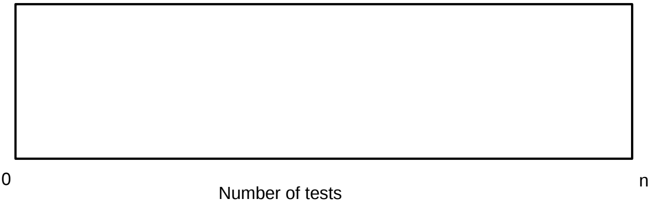
Basic setup

- Options to choose the underlying testing procedure
    1. Number of stages
    2. Pooling procedure

Basic setup

- Options to choose the underlying testing procedure
    1. Number of stages
    2. Pooling procedure
- **To Do:** Rigorous Analysis of the choice

Basic setup

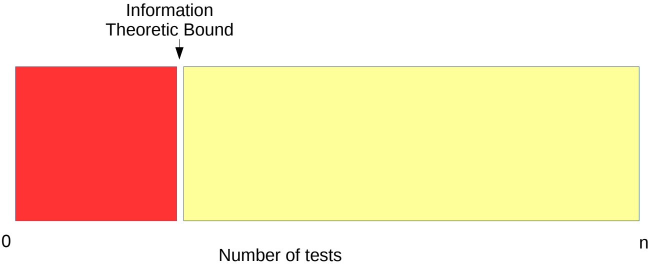


0

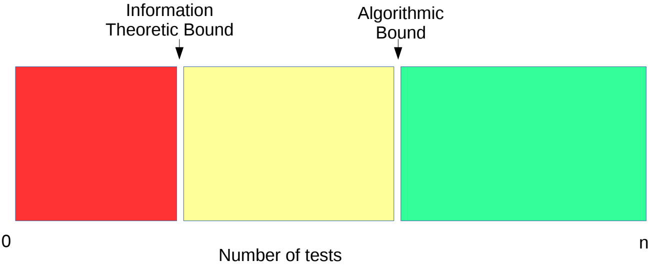
n

Number of tests

Basic setup

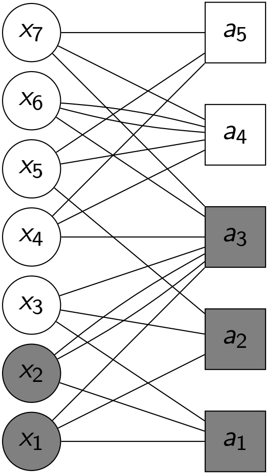


Basic setup


Information Theoretic Bound
Algorithmic Bound
0
n
Number of tests

Binary group testing

Binary group testing



Binary group testing: Information Theory

- Counting Bound implies $m > k \cdot log_2(n/k)$.

Binary group testing: Information Theory

- Counting Bound implies $m > k \cdot log_2(n/k)$.
- Baldassini et al: Adaptive testing strategies achieve this bound.

Binary group testing: Information Theory

- Counting Bound implies $m > k \cdot log_2(n/k)$.
- Baldassini et al: Adaptive testing strategies achieve this bound.
- **Question:** Is non-adaptive group testing able to achieve the bound as well?

Binary group testing: Information Theory

- Assume a constant weight testing scheme

Binary group testing: Information Theory

- Assume a constant weight testing scheme
- The underlying graph structure:

Binary group testing: Information Theory

- Assume a constant weight testing scheme
- The underlying graph structure:
  - Bipartite Factor Graph
  - Fixed variable nopde degree

Binary group testing: Information Theory

- Assume a constant weight testing scheme
- The underlying graph structure:
    - Bipartite Factor Graph
    - Fixed variable nopde degree

Theorem 1 [CGHL18]

Let $m_{inf} = \frac{n^{\theta} log(n)}{min\{1, \frac{1-\theta}{\theta} log(2)\} log(2)}$

1. $m < (1-\epsilon)m_{inf}$: No algorithm exists to output the right configuration for the constant weight pooling
2. $m > (1+\epsilon)m_{inf}$: There exist an algorithm, which outputs the right configuration w.h.p.

Binary group testing: Information Theory

- Lower Bound: Derive the value m^*, below which infected/uninfected individuals occur that may swap status without harming the test result

Binary group testing: Information Theory

- Lower Bound: Derive the value m^*, below which infected/uninfected individuals occur that may swap status without harming the test result
- Upper Bound: Derive the value m^{**} above which no satisfying assignment beside the original one exists

Binary group testing: Information Theory

- Lower Bound: Derive the value m^*, below which infected/uninfected individuals occur that may swap status without harming the test result
- Upper Bound: Derive the value m^{**} above which no satisfying assignment beside the original one exists
    1. Small overlap argument: Analyse probability that configuration with certain overlap fulfills the test result
    2. Large overlap argument: Analyse probability that positive (negative) tests stay positive (negative)

Binary group testing: Information Theory

- Lower Bound: Derive the value m^*, below which infected/uninfected individuals occur that may swap status without harming the test result
- Upper Bound: Derive the value m^{**} above which no satisfying assignment beside the original one exists
    1. Small overlap argument: Analyse probability that configuration with certain overlap fulfills the test result
    2. Large overlap argument: Analyse probability that positive (negative) tests stay positive (negative)
- m^*, m^{**} set conditions to derive the $m_{inf}(n, \theta)$ as stated

Binary group testing: Algorithmic Aspect

- Allemann: multi-stage algorithm at the predicted lower bound

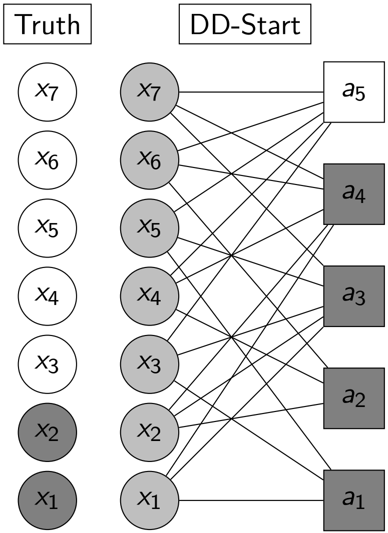
Binary group testing: Algorithmic Aspect

- Allemann: multi-stage algorithm at the predicted lower bound
- Aldridge, Scarlett et. al.: Sub-optimal non-adaptive strategies available

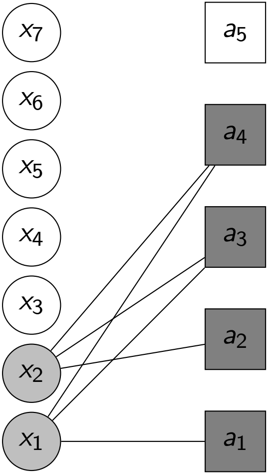
Binary group testing: Algorithmic Aspect

- Allemann: multi-stage algorithm at the predicted lower bound
- Aldridge, Scarlett et. al.: Sub-optimal non-adaptive strategies available
- Most promising algorithms: SCOMP, Definite Defective

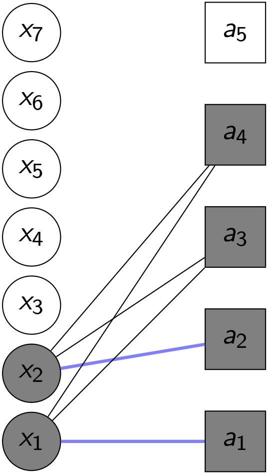
Binary group testing: Algorithmic Aspect



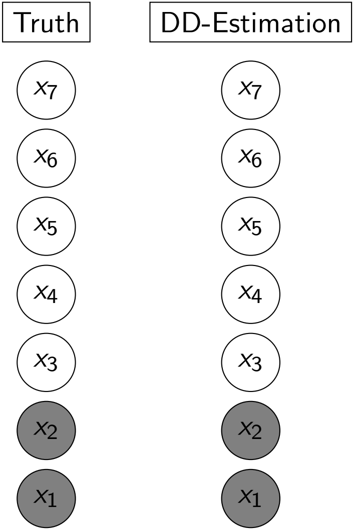
Binary group testing: Algorithmic Aspect



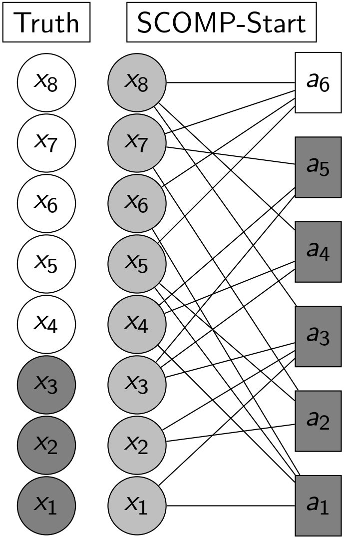
Binary group testing: Algorithmic Aspect



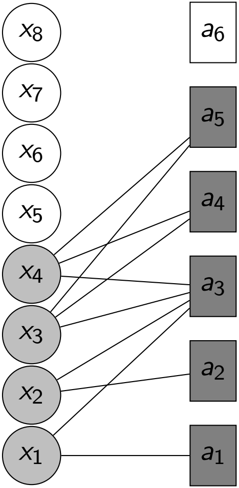
Binary group testing: Algorithmic Aspect



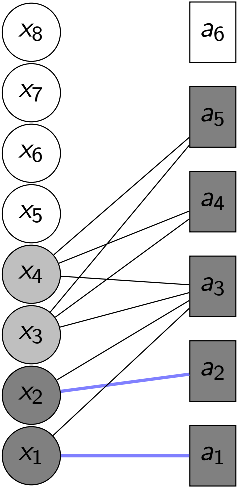
Binary group testing: Algorithmic Aspect



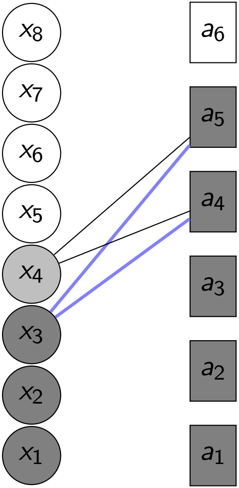
Binary group testing: Algorithmic Aspect



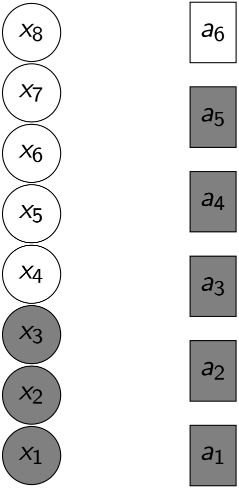
Binary group testing: Algorithmic Aspect



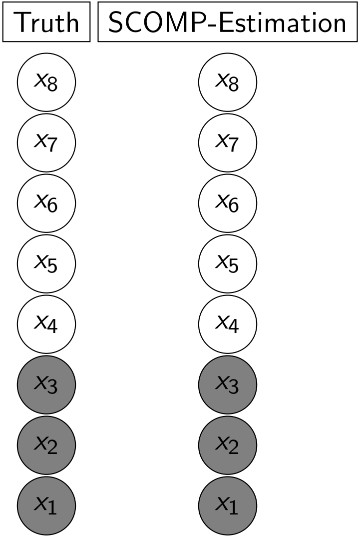
Binary group testing: Algorithmic Aspect



Binary group testing: Algorithmic Aspect



Binary group testing: Algorithmic Aspect



Truth

SCOMP-Estimation

Binary group testing: Algorithmic Aspect

- **Conjecture:** SCOMP outperforms the Definite Defective

Binary group testing: Algorithmic Aspect

- **Conjecture:** SCOMP outperforms the Definite Defective
- We refute the conjecture by showing that the algorithms fail at the same point

Binary group testing: Algorithmic Aspect

- **Conjecture:** SCOMP outperforms the Definite Defective
- We refute the conjecture by showing that the algorithms fail at the same point

Theorem 2 [CGHL18]

Let $m_{\mathrm{alg}}(n,\theta) = \frac{k\log(n/k)}{\min\{1, \frac{1-\theta}{\theta}\log^2 2\}}$, $0 < \theta < 1$ and $\epsilon > 0$:
For $m < (1-\epsilon)m_{alg}(n,\theta)$, both SCOMP and DD fail to output $\boldsymbol{\sigma}$ w.h.p.

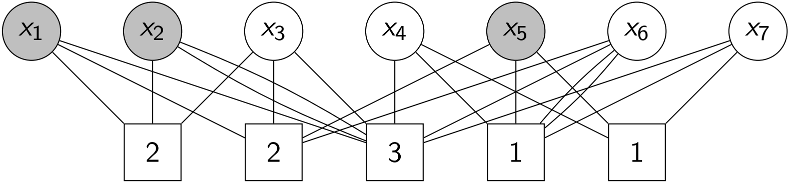
Binary group testing: Algorithmic Aspect

- DD work iff every infected individual is in at least one test with only uninfected individuals that are themselves in at least one test with only other uninfected individuals
- Additional SCOMP step is a Greedy Vertex Cover
- Show w.h.p.: Local structure of infected/ uninfected individuals in remaining graph look the same
- SCOMP fails in first step w.h.p.

$\Rightarrow$ SCOMP and DD have the same algorithmic threshold

Quantitative group testing

Quantitative group testing



Quantitative group testing

- Counting bound: $(k+1)^m \geq \binom{n}{k}$.

Quantitative group testing

- Counting bound: $(k+1)^m \geq \binom{n}{k}$.
- Alaoui et. al.: For the linear case Information Theoretic phase transition and efficient algorithm at lower bound established

Quantitative group testing

- Counting bound: $(k+1)^m \geq \binom{n}{k}$.
- Alaoui et. al.: For the linear case Information Theoretic phase transition and efficient algorithm at lower bound established
- For sub-linear regime: Information Theory not entirely understood (Djackov) and only sub-optimal algorithms available (Karimi et. al.) .

Quantitative group testing

- Counting bound: $(k+1)^m \geq \binom{n}{k}$.
- Alaoui et. al.: For the linear case Information Theoretic phase transition and efficient algorithm at lower bound established
- For sub-linear regime: Information Theory not entirely understood (Djackov) and only sub-optimal algorithms available (Karimi et. al.) .
- **Contribution:**
  1. Establish sharp phase transition in the sublinear regime (i.e. we show the achievability).
  2. Introduce a Greedy Algorithm that outperforms the best known one in certain sparsity levels

Quantitative group testing: Information Theory

Theorem 3 [GHKL19]

Suppose that $0 < \theta < 1, k = n^{\theta}$ and $\epsilon > 0$ and let

$$m_{inf} = 2\frac{1-\theta}{\theta}k$$

1. For $m < (1-\epsilon)m_{inf}(n,\theta)$: No algorithm that outputs $\boldsymbol{\sigma}$ exists
2. For $m > (1+\epsilon)m_{inf}(n,\theta)$: An algorithm exists that outputs $\boldsymbol{\sigma}$ with high probability

Quantitative group testing: Information Theory

- Derive a m^* s.t. for $m > m^*$ no second satisfying assignment exists

Quantitative group testing: Information Theory

- Derive a m^* s.t. for $m > m^*$ no second satisfying assignment exists
- Analyse it for small and high overlaps

Quantitative group testing: Information Theory

- Derive a m^* s.t. for $m > m^*$ no second satisfying assignment exists
- Analyse it for small and high overlaps
  1. Large overlap: Analyse changes via Balls and Bins
  2. Small overlap: Analyse changes as returning random walk

Quantitative group testing: Information Theory

- Derive a m^* s.t. for $m > m^*$ no second satisfying assignment exists
- Analyse it for small and high overlaps
  1. Large overlap: Analyse changes via Balls and Bins
  2. Small overlap: Analyse changes as returning random walk
- Analysis establishes conditions for m_{inf}

Quantitative group testing: Algorithmic aspect

Maximum-Neighborhood Algorithm [GHKL19]

Input: $\mathcal{G}, y, k$
Output: Estimation $\tilde{\vec{\sigma}}$ for $\vec{\sigma}$.

1. For every x_i for $i \in [n]$ calculate $\Psi_i = \sum_{j \in \partial x_i} y_j$
2. Set $\Psi_i' := \Psi_i \cdot \frac{m/2}{\Delta_i}$
3. Order the individuals i in decreasing order due to Ψ_i'
4. Declare the first k ordered individuals as infected, declare the other individuals as healthy

Quantitative group testing: Algorithmic aspect

Maximum-Neighborhood Algorithm [GHKL19]

Input: $\mathcal{G}, y, k$
Output: Estimation $\tilde{\vec{\sigma}}$ for $\vec{\sigma}$.

1. For every x_i for $i \in [n]$ calculate $\Psi_i = \sum_{j \in \partial x_i} y_j$
2. Set $\Psi_i' := \Psi_i \cdot \frac{m/2}{\Delta_i}$
3. Order the individuals i in decreasing order due to Ψ_i'
4. Declare the first k ordered individuals as infected, declare the other individuals as healthy

Theorem 4 [GHKL19]

Define $m_{MN}(n, \theta) = \frac{1+\sqrt{\theta}}{1-\sqrt{\theta}} k \log(n/k)$. The MN-Algorithm outputs the correct configuration w.h.p. if $m > m_{MN}$

Quantitative group testing: Algorithmic aspect

- Seperate the distribution of Ψ_i for uninfected and infected individuals.

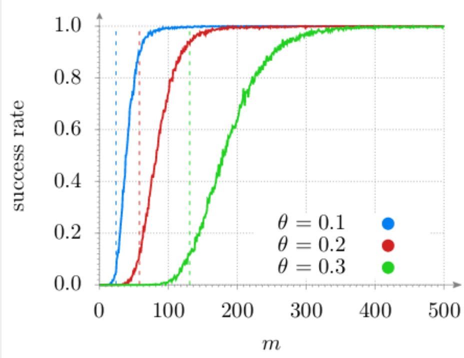
Quantitative group testing: Algorithmic aspect

- Seperate the distribution of Ψ_i for uninfected and infected individuals.
- Union bound over all k infected and $n - k$ uninfected individuals.

Quantitative group testing: Algorithmic aspect

- Seperate the distribution of Ψ_i for uninfected and infected individuals.
- Union bound over all k infected and $n-k$ uninfected individuals.
- Applying the Chernoff Bound and optimizing w.r.t. the seperating-parameter gives the (sufficient) condition on m_{MN}.

Quantitative group testing: Algorithmic aspect



Recovery success rate

- $n = 1000$ individuals.
- 1000 rounds of simulation.
- Dashed lines: Asymptotic prediction of the required number of tests.

Thank you for your attention