IEEE
Information
Theory Society

UCLA

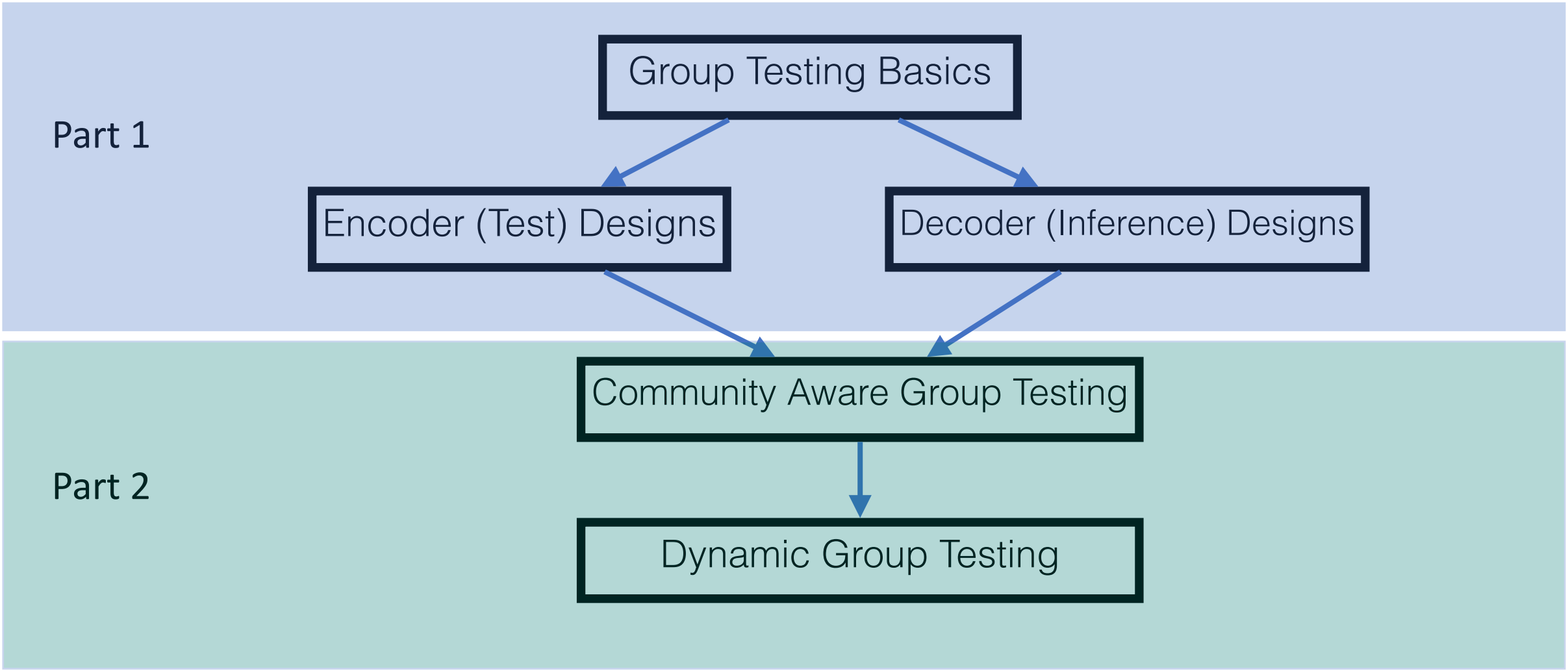
Community Aware Group Testing

Part II

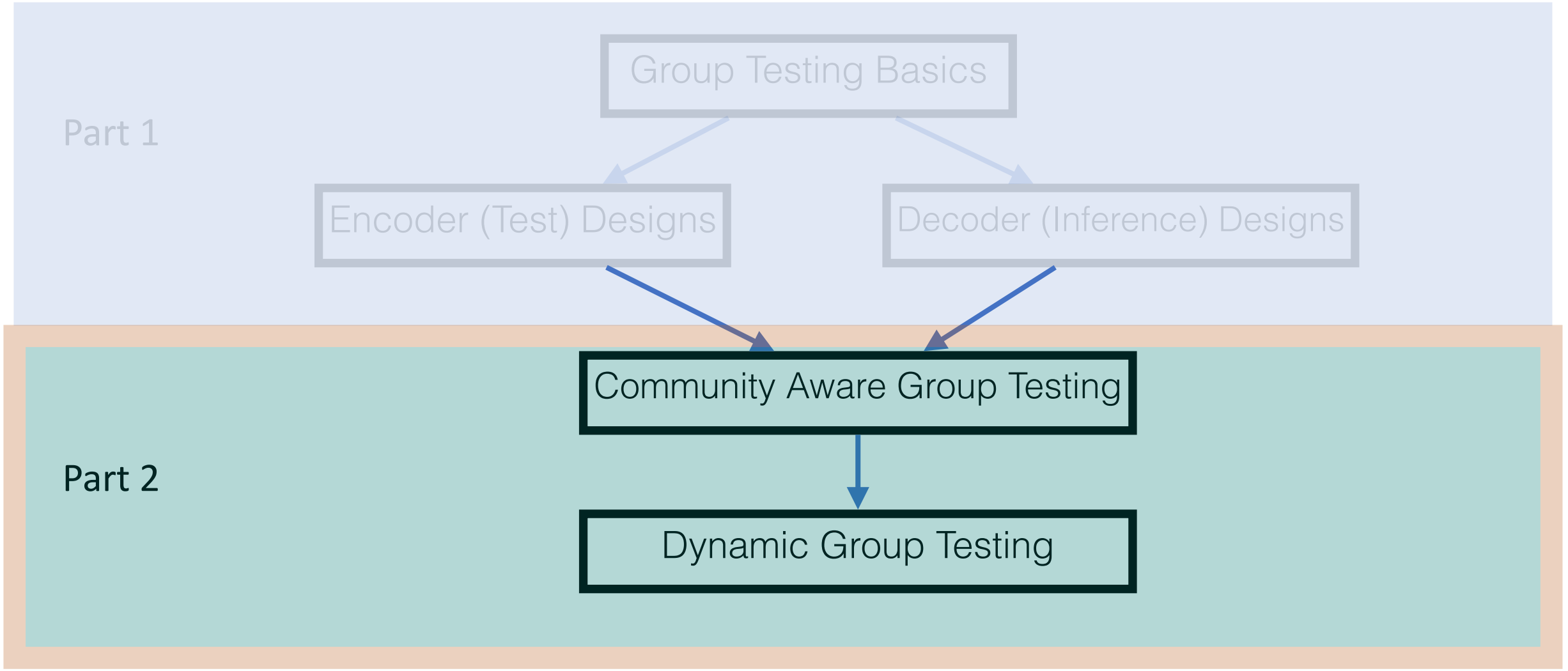
Christina Fragouli

North America School of Information Theory, August 2022

Outline


Part 1
Group Testing Basics
Encoder (Test) Designs
Decoder (Inference) Designs
Part 2
Community Aware Group Testing
Dynamic Group Testing

Outline


Part 1
Group Testing Basics
Encoder (Test) Designs
Decoder (Inference) Designs
Part 2
Community Aware Group Testing
Dynamic Group Testing

Thanks to:

Sundar Rajan
Meta (Facebook)

Pavlos Nikolopoulos
EPFL

NSF CIF Award 2146828
Group Testing for Epidemics Control

IEEE BITS THE INFORMATION THEORY MAGAZINE

Group Testing for Community Infections

Abstract

Group testing is the technique of pooling together diagnostic samples in order to increase the efficiency of medical testing. Traditionally, works in group testing assume that the infections are i.i.d. However, contagious diseases like COVID-19 are governed by community spread and hence the infections are correlated. This survey presents an overview of recent research progress that leverages the community structure to further improve the efficiency of group testing. We show that taking into account the side-information provided by the community structure may lead to significant savings—up to 60% fewer tests compared to traditional test designs. We review lower bounds and new approaches to encoding and decoding algorithms that take into account the community structure and integrate group testing into epidemiological modeling. Finally, we also discuss a few important open questions in this space

Testing against Coronavirus

enables:

- better epidemiological models
- managing lockdowns
- a “safer” opening after a lockdown

- Has to be *large-scale*

We need *good* testing

- Reliable
    - Very high sensitivity and specificity rates to enable targeted interventions
- Large-scale
    - Test large populations like neighborhoods or cities, etc.
- Continual
    - Several thousands or millions of tested people per country per day

Group (or pooled) testing

Key idea:

- Pool multiple samples together
- Test them all with one test
    - If test is positive, then at least one individual is infected
    - If test is negative, then *no* individuals are infected

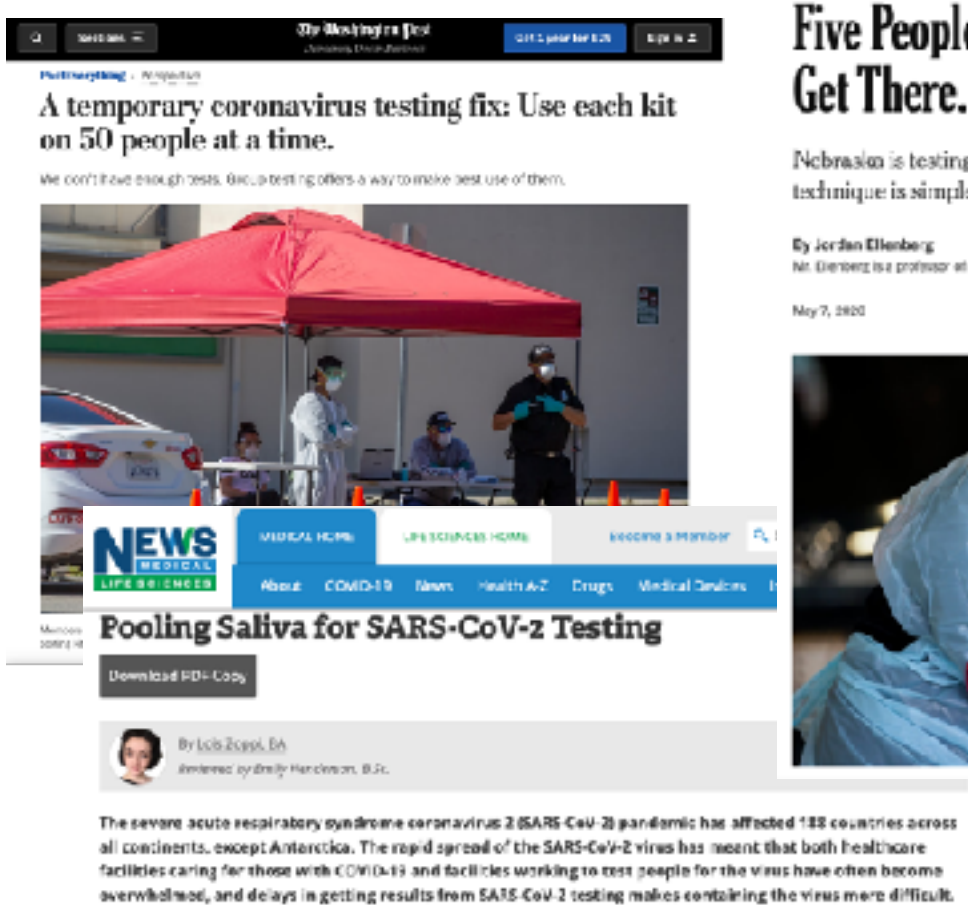
Group testing

- Applicable to almost any type of test
- Achieves high throughput (especially in low prevalence)
    - few tests to test a large population
    - *large-scale*, *continual* testing becomes possible
- Improves *reliability* of noisy testing
    - without sacrificing tests for repetition

Group testing is being rediscovered…

Academia has investigated this for decades.

But now, on the news!



Observation

All known group-testing schemes assume *independent* infections,

but…infections are *correlated*.

Observation

Can we further improve group-testing efficiency,
in terms of *#tests* and/or *reliability*,
if we consider *correlated* infections
based on a *known* community structure?

Community-Aware Group Testing Literature

2020

- P. Nikolopoulos, et al. Group testing for connected communities
- J. Zhu, K. Rivera, and D. Baron, “Noisy pooled PCR for virus testing

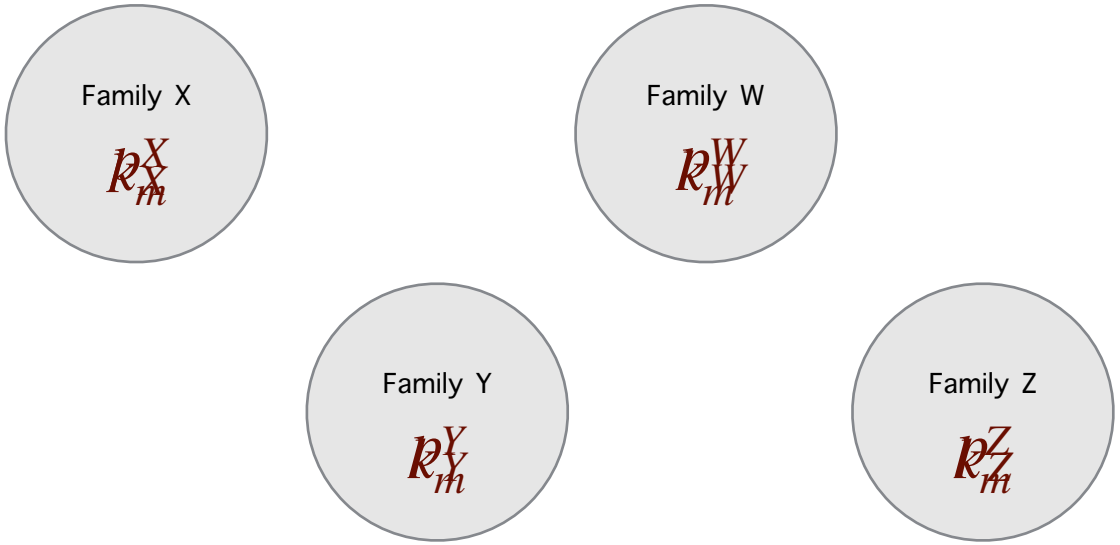
2021

- Ahn et al., "Adaptive Group Testing on Networks with Community Structure,"
- Arasli et al. “Group testing with a graph infection spread model,
- S. Sihag, A. Tajer and U. Mitra, "Adaptive Graph-Constrained Group Testing
- P. Nikolopoulos et al. "Group testing for overlapping communities,
- S. R. Srinivasavaradhan et al. ”An entropy reduction approach to continual testing
- Gabyrs et al. AC-DC: Amplification Curve Diagnostics for Covid-19 Group Testing,
- …

2022 (many more, eg 13 papers at ISIT 2022)

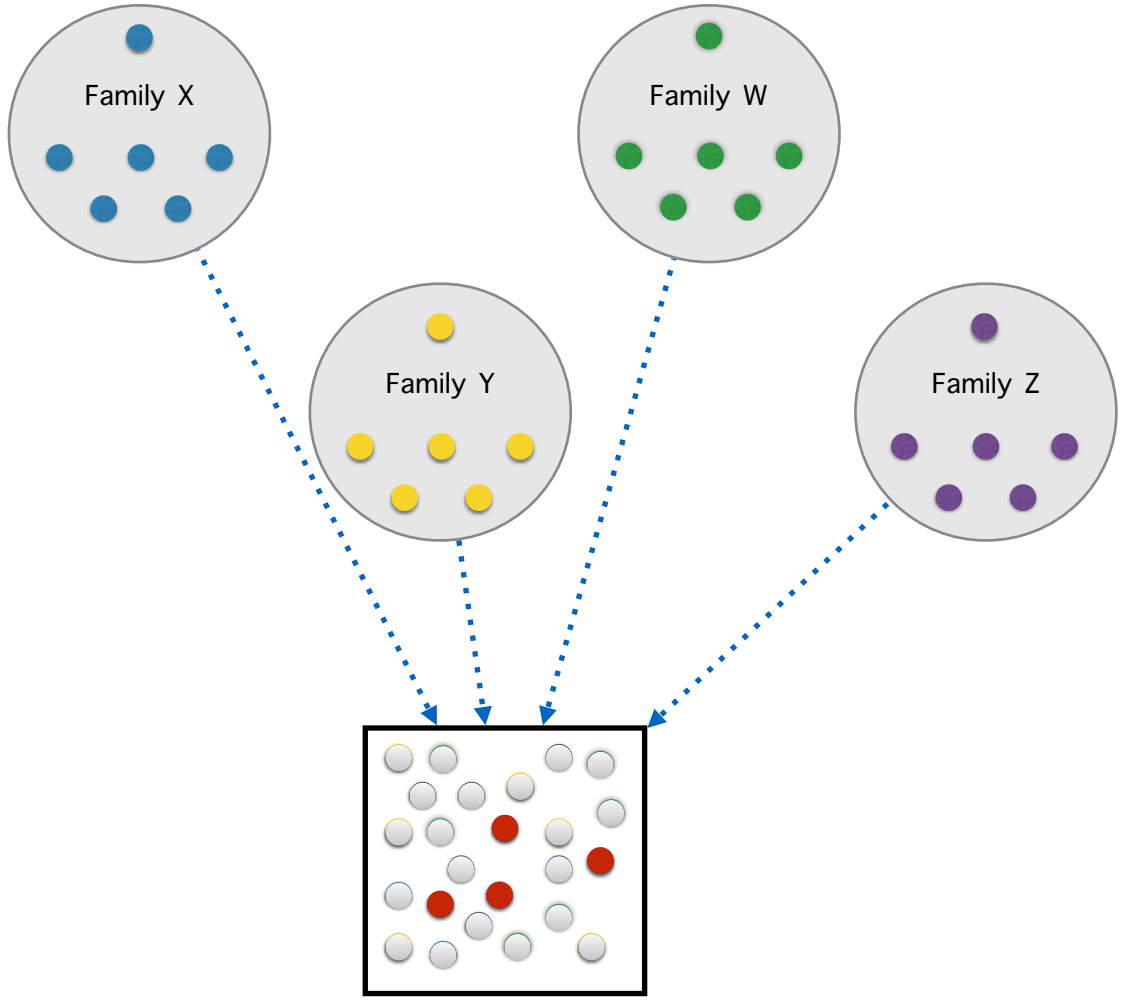
Community and infection model

- Combinatorial:
  - k_f infected families
  - k_m^j infected members/ infected family j
- Probabilistic:
  - each family j is infected w.p. q
  - each family member of an infected family is infected w.p. p_j



Traditional group testing (adaptive)

- “views” only a single set of members
- performs adaptive or nonadaptive testing
- identifies infected individuals (with or without errors)



Known performance bounds

Counting bound: $T \geq \log_2 \binom{n}{k}$ (which is $\ll n$)

** n = population size, k = #of infected people

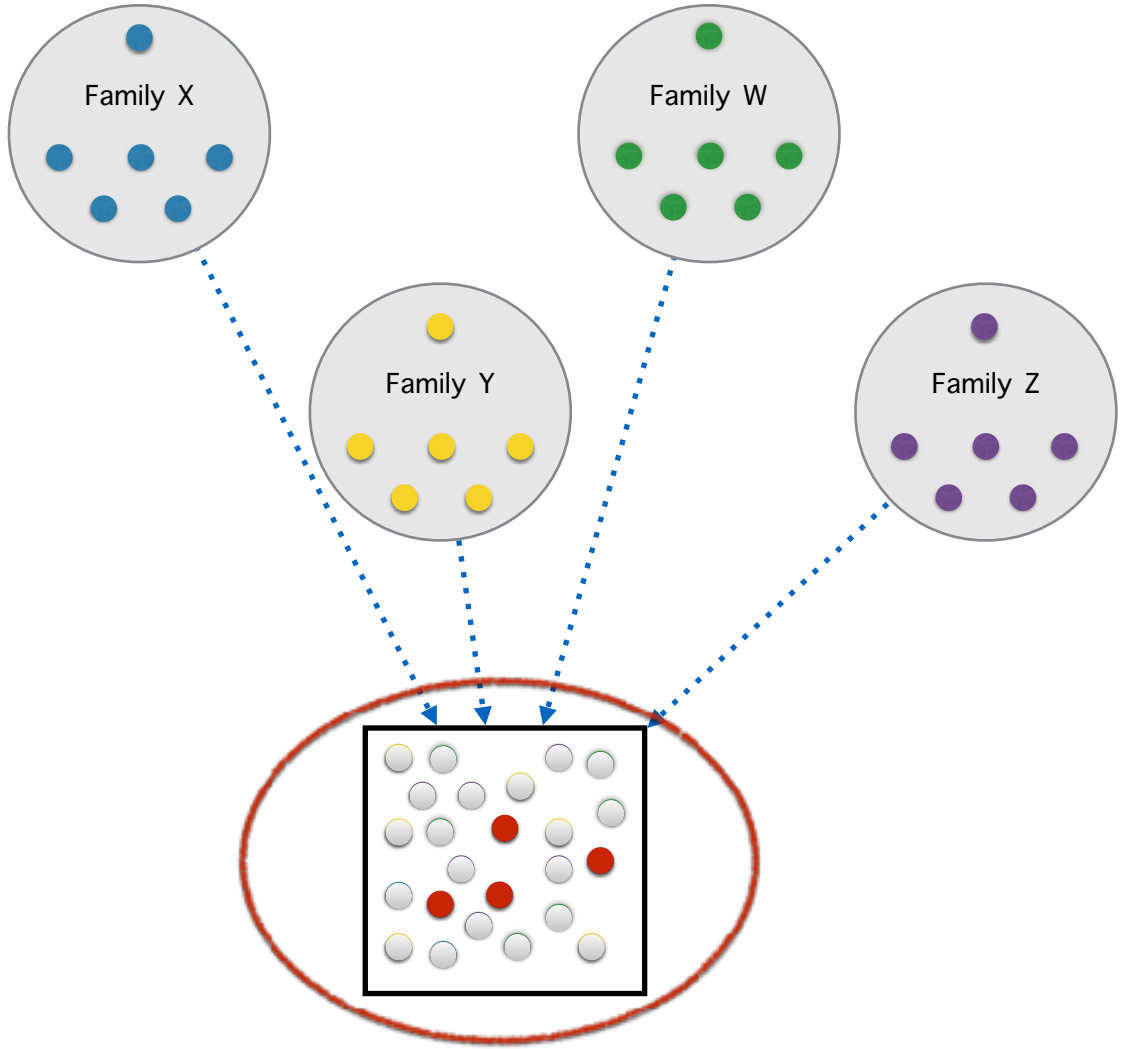
- Adaptive group testing
  - achieves the bound
  - no detection errors

- Nonadaptive group testing
  - may achieve the bound in cases
  - asymptotically vanishing errors

But…is this the best we can do?

- Group testing needs *≥13 tests* (or ≥20, if k is unknown)

- An algorithm that takes structure into account would need *only 7 tests (~50% improvement)*



New lower bounds

- Combinatorial

$$T \geq \log_2 \binom{\#families}{k_f} + k_f \log_2 \binom{familySize}{k_m}$$

k_m *times better than the counting bound*

- Probabilistic (asymptotically equivalent)

$$T \geq \#families \cdot h_2(q) + q \cdot n \cdot h_2(p)$$

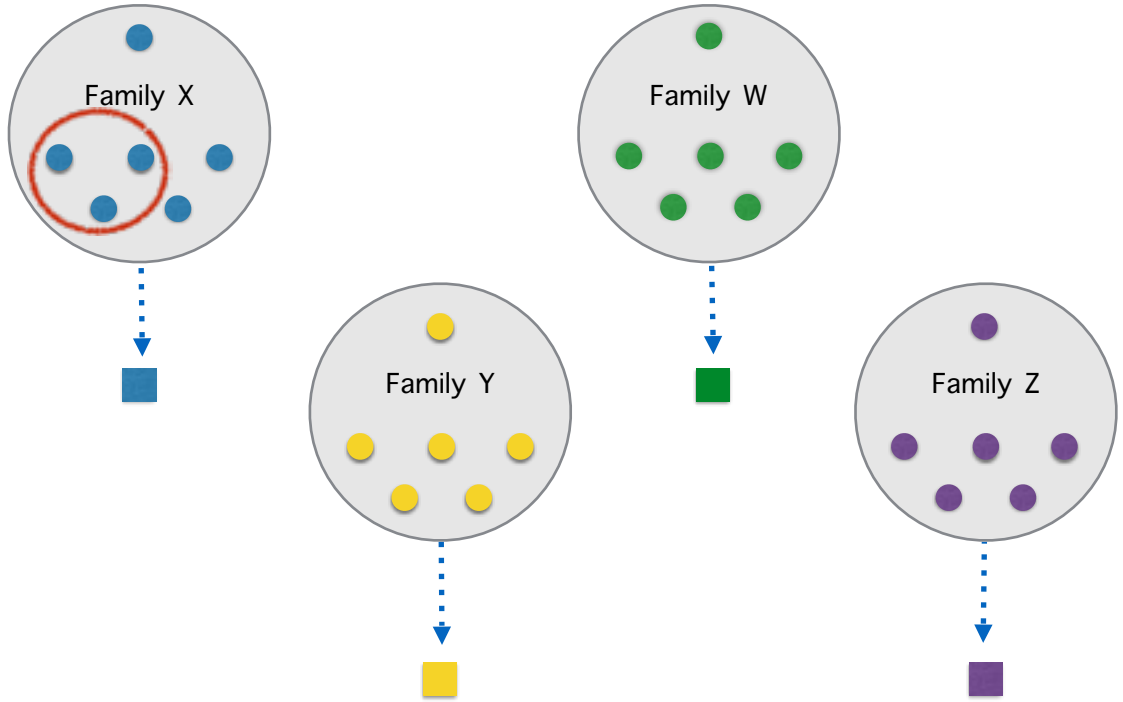
Adaptive Algorithm

- P. Nikolopoulos, S. R. Srinivasavaradhan, T. Guo, C. Fragouli, S. N. Diggavi, "Group testing for connected communities", In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, PMLR 130:2341-2349, 2021 (ArXiv version June 2020).

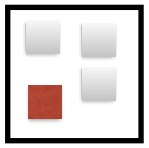
Idea 1: First identify infected families

For each family:

- Select R representatives
- Pool together their samples
- Group test the pooled samples



- Identify infected families
  e.g. blue family



Group testing

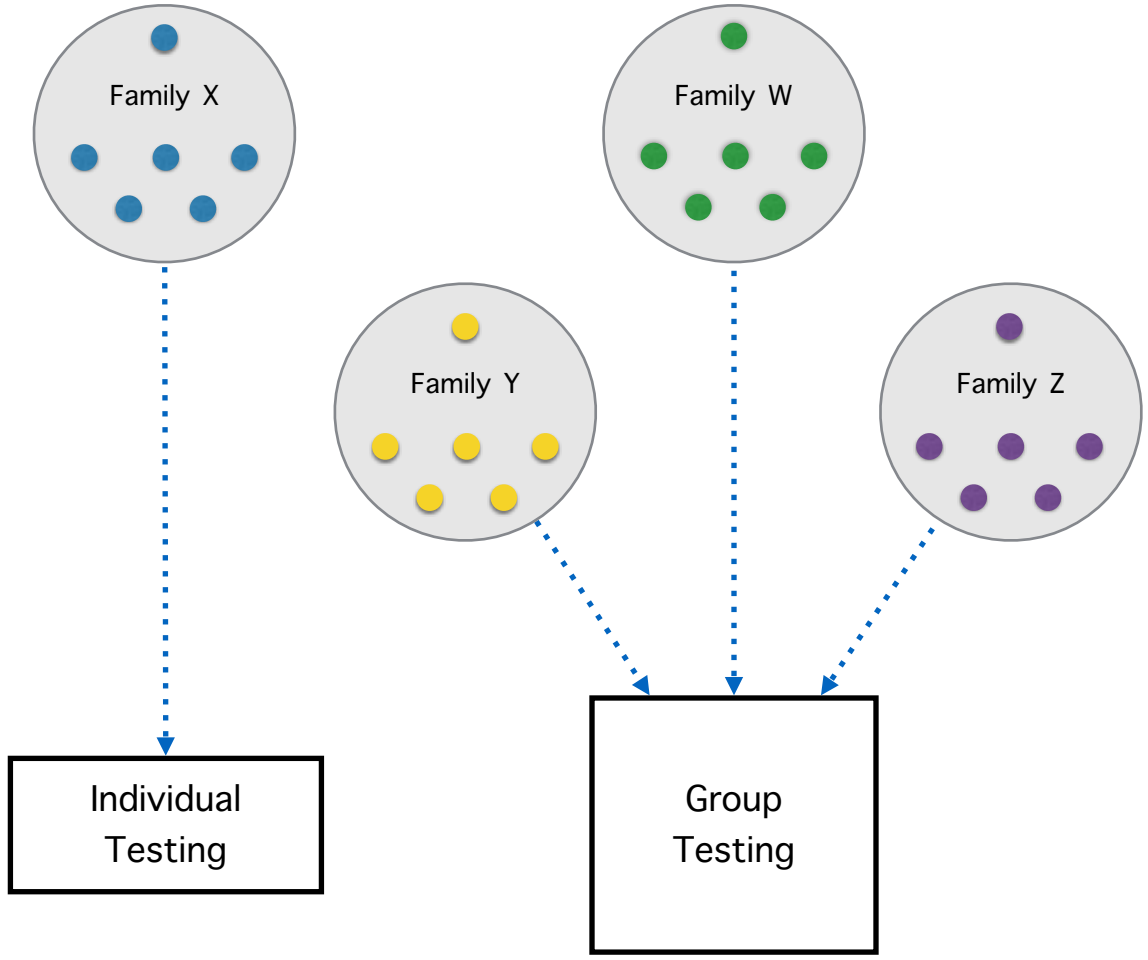
Idea 2: Test infected families based on their regime

If pooled sample is positive (*heavily* infected families):

- Test all members individually

If pooled samples is negative (*not/lightly* infected families):

- Group test all members together



Properties

- Completely adaptive in 2 stages, without errors
- Lower bound is asymptotically achieved when:
  $p_j \to 1$ (or $k_m^j \to familySize$)
- $\mathbb{E}[T] \ll k\log_2 n$ (= performance of classic binary splitting)

Nonadaptive algorithm

- P. Nikolopoulos, S. R. Srinivasavaradhan, T. Guo, C. Fragouli, S. N. Diggavi, "Group testing for connected communities", In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, PMLR 130:2341-2349, 2021 (ArXiv version June 2020).

Idea 1: Testing matrix in 2 parts

$$G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

- G_1 identifies not infected families,
  G_2 identifies infected members of infected families

- G_1 - encoding: use one pooled samples from each family (R=familySize)

- G_1 - decoding: COMP

Idea 2: Specific design for G_2

- Encoding: a family shares tests with only few other families

$$\mathbf{G}_2 = \begin{bmatrix} I_M & 0_{M\times M} & 0_{M\times M} & I_M & 0_{M\times M} & 0_{M\times M} \\ 0_{M\times M} & I_M & 0_{M\times M} & 0_{M\times M} & I_M & 0_{M\times M} \\ 0_{M\times M} & 0_{M\times M} & I_M & 0_{M\times M} & 0_{M\times M} & I_M \end{bmatrix}$$

- Decoding:
  - members from not-infected families are identified from G_1 and are eliminated from the decoding of G_2
  - all other members are identified using COMP decoding

Properties

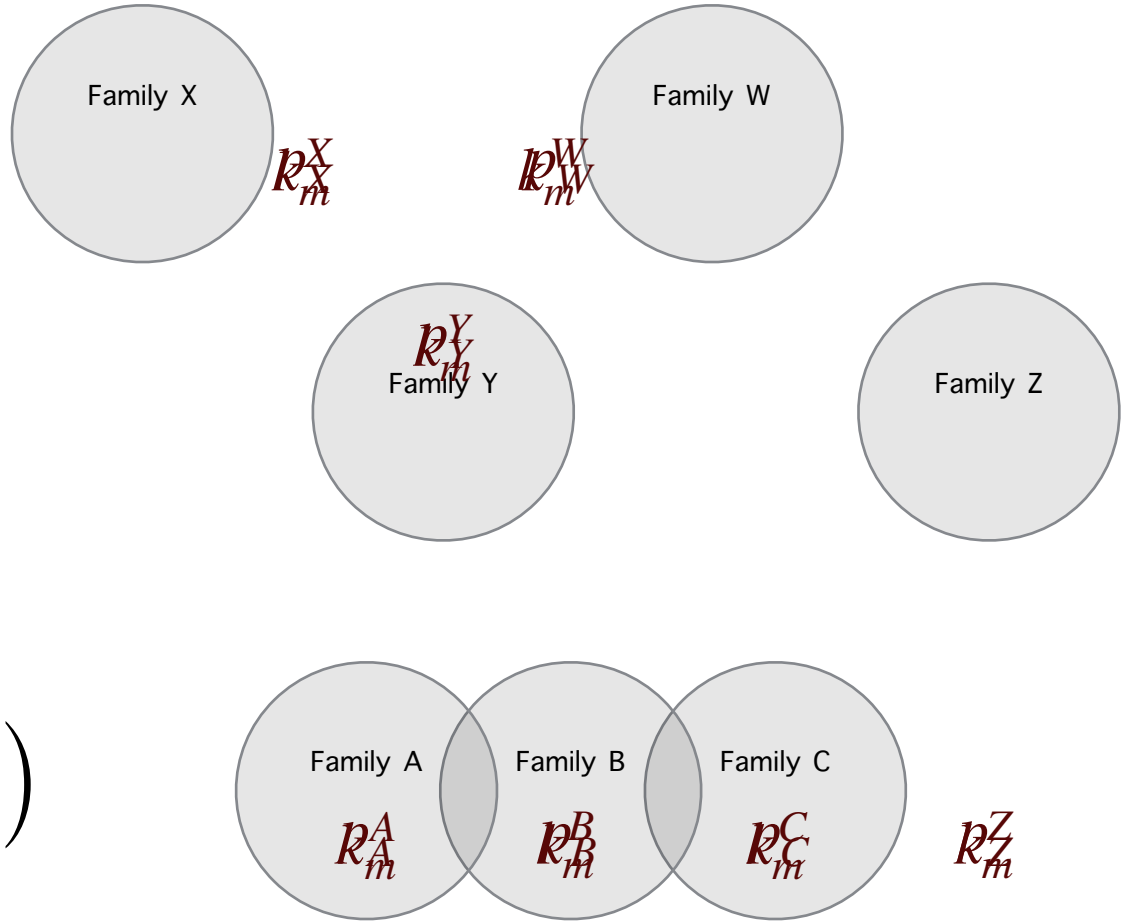
- No zero-error decoding
    - would require $T_2 > n$
- No false negatives
    - due to COMP decoding
- False positives are minimized, if block rows have the same #families

$$\mathbf{G}_2 = \begin{bmatrix} I_M & 0_{M\times M} & 0_{M\times M} & I_M & 0_{M\times M} & 0_{M\times M} \\ 0_{M\times M} & I_M & 0_{M\times M} & 0_{M\times M} & I_M & 0_{M\times M} \\ 0_{M\times M} & 0_{M\times M} & I_M & 0_{M\times M} & 0_{M\times M} & I_M \end{bmatrix}$$

Community and infection model — *overlapping*

- Combinatorial:
  - k_f infected families
  - k_m^j infected members/ infected family j, some of which are *shared*
- Probabilistic:
  - each family j is infected w.p. q
  - if infected, its infection prob. is $p_j > 0$
  - member i is infected w.p. that *increases* with the number of the infected families $\mathcal{I}$ it belongs to:

$$p_i = 1 - \prod_{j \in \mathcal{I}} \left(1 - p_j\right)$$



Contributions

- New information-theoretic lower bounds for #of tests

In realistic setups:

scale linearly with #of infected families

- Community-aware test design (encoding)
  - adaptive and non-adaptive
- Community-aware decoding
  - loopy belief propagation

In our simulations:

30-65% fewer tests than traditional GT

may achieve the lower bounds

Community-aware encoding (adaptive test design)

Key ideas

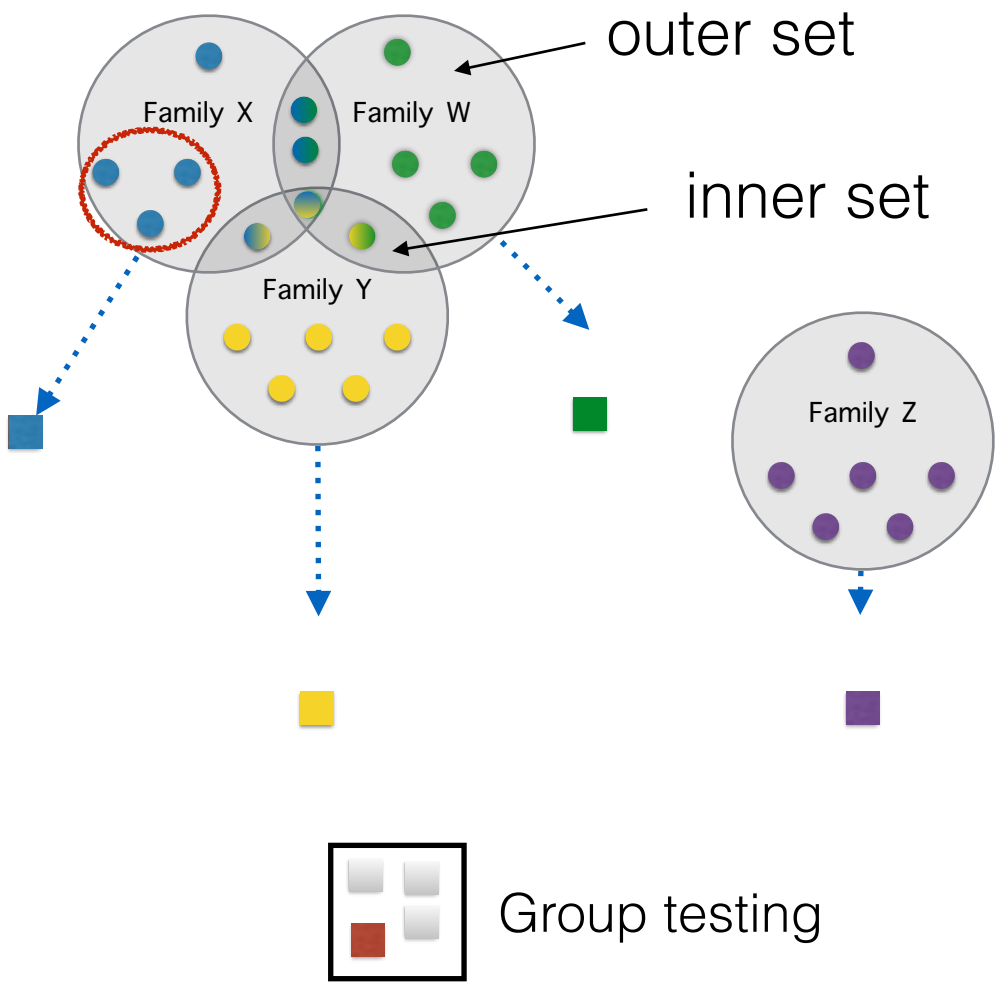
- Infer how heavily/lightly infected families are
    - heavily: infection rate >= θ
    - lightly: infection rate < θ
    - e.g. θ = 38%
- Test based on the infection regime
    - heavily infected —> individual testing
    - lightly infected —> traditional GT (e.g. binary splitting)

Idea 1: Infer infection regime with 1 test/family

For each family:

- Select R representatives from each "outer" set
- Pool together their samples
- Group test the pooled samples

- Identify infected ones



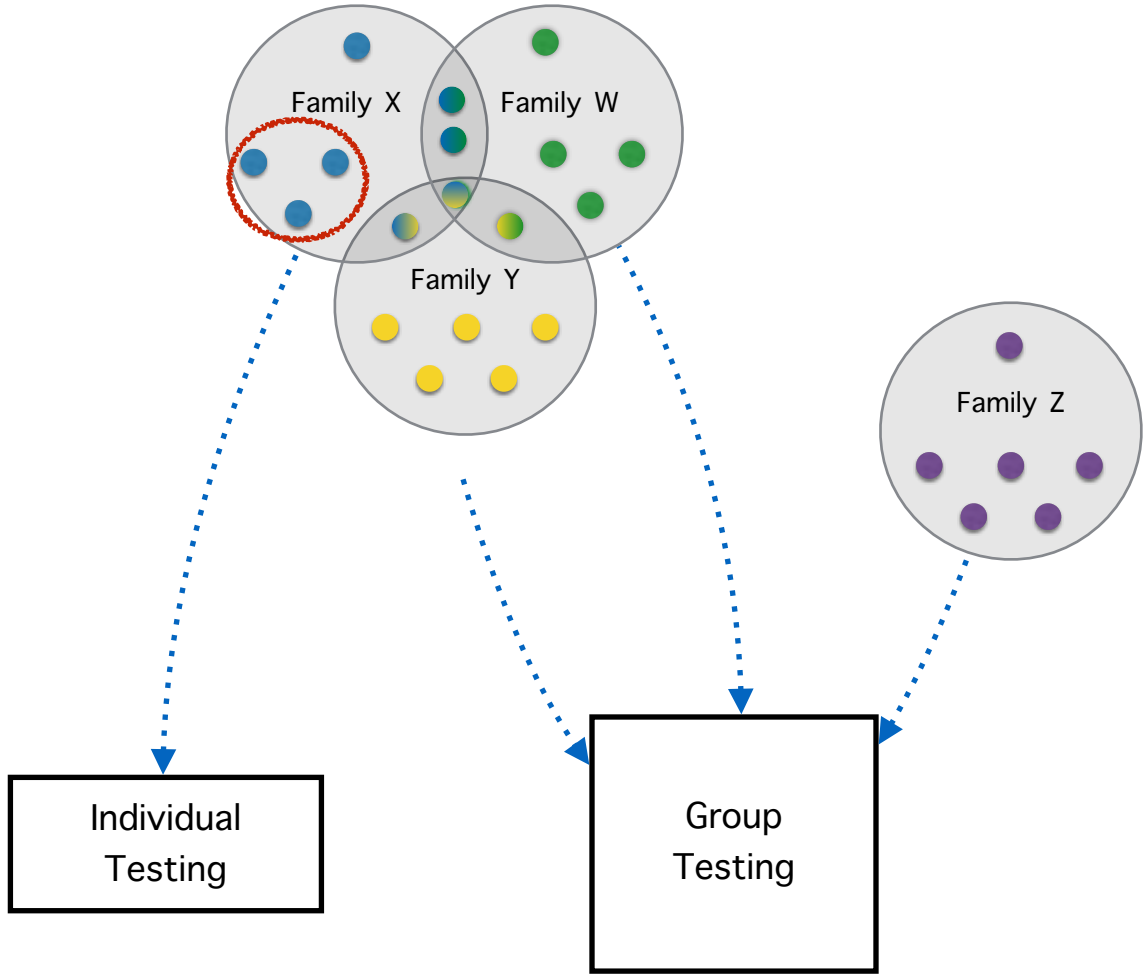
Idea 2: Test based on the infection regime

If pooled sample is positive (*heavy* infection):

- Test all members individually

If pooled sample is negative (*light* infection):

- Group test all members together



Properties

- Fully adaptive, without errors
- Asymptotically achieves the lower bound when:
  - $p_j \to \frac{1}{2}$
    - shared members are few (“low overlap”)

Community-aware encoding (nonadaptive test design)

Properties (nonadaptive)

- Tests are split in 2 parts (similarly to adaptive)
- All tests are pre-determined
- Errors, but members are grouped into tests s.t.: $FN = 0$

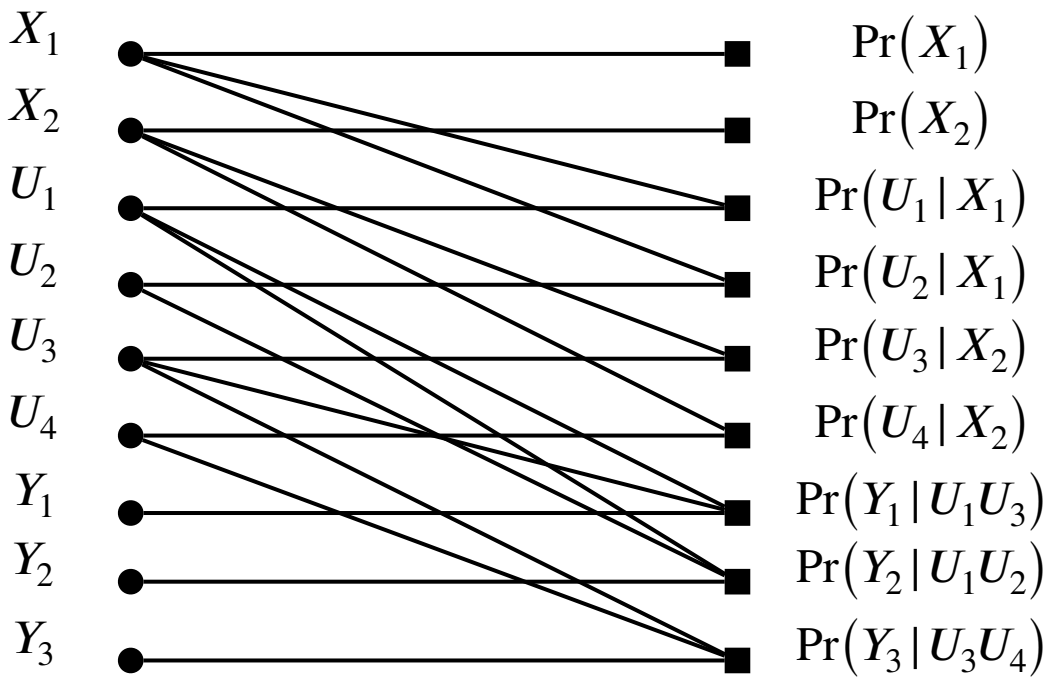
Other things considered

- 2-stage algorithm:
    - much lower error rates than nonadaptive
    - ≤3% in numerical experiments
- Noisy tests
    - noise model = z-channel
    - comparison of 2-stage algorithm with noisy tests against repetition testing
    - examples with simple decoders yield exponentially smaller error rates

Community-aware decoding

Decoding using community structure side-information

- The loopy belief propagation (LBP) can easily be configured to account for the side-information provided by the community structure and estimate the infection statuses.
- First, we construct the *factor graph,* which represents the joint distribution as a factorization (product) of marginals. Each left node corresponds to a variable and each right node to a factor in this factorization.



Example factor graph with 2 families, 4 individuals and 3 tests.

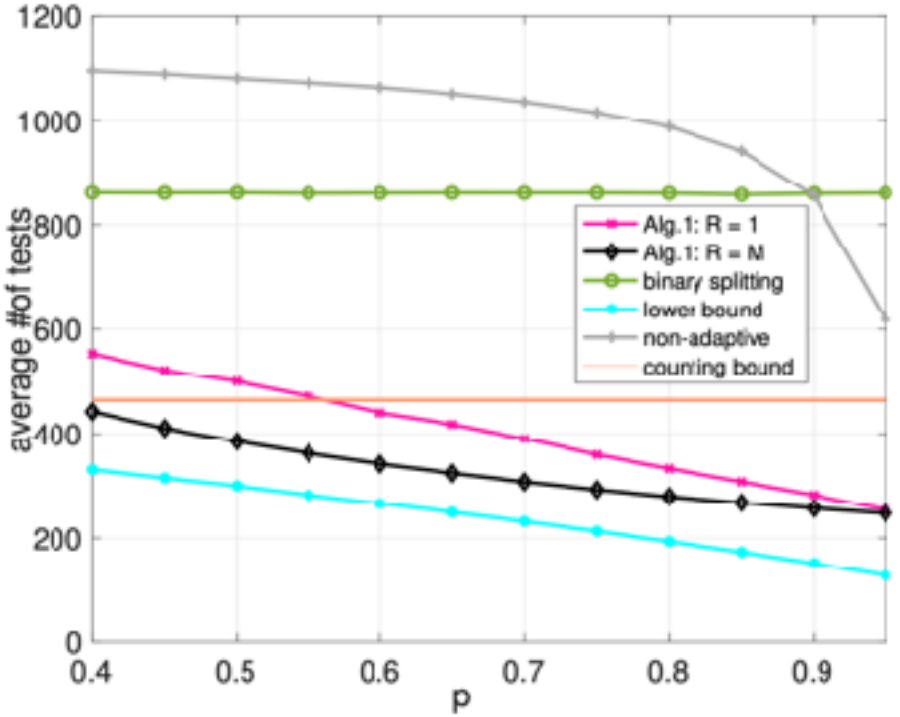
Evaluation

Experimental setup non-overlaping families

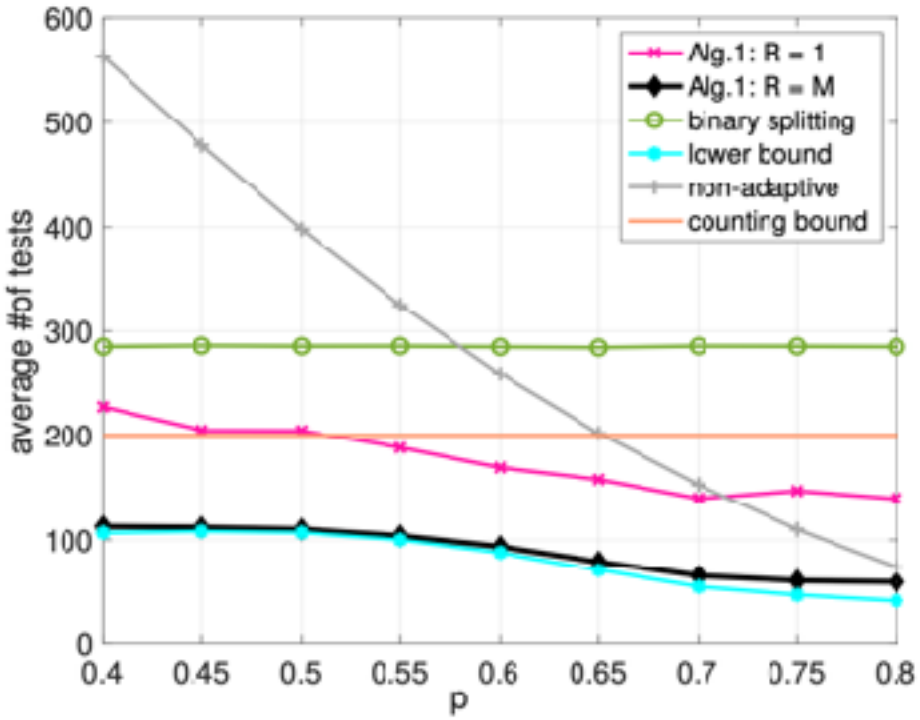
- 2 different use cases for the community structure:
    - neighborhood: 200 families, 5 members each
    - university: 20 classes, 50 students each
- 2 different infection regimes:
    - sparse: (expected) 32 infected people out of 1000
    - linear: (expected) 100 infected people out of 1000

Results - #tests

Linear regime

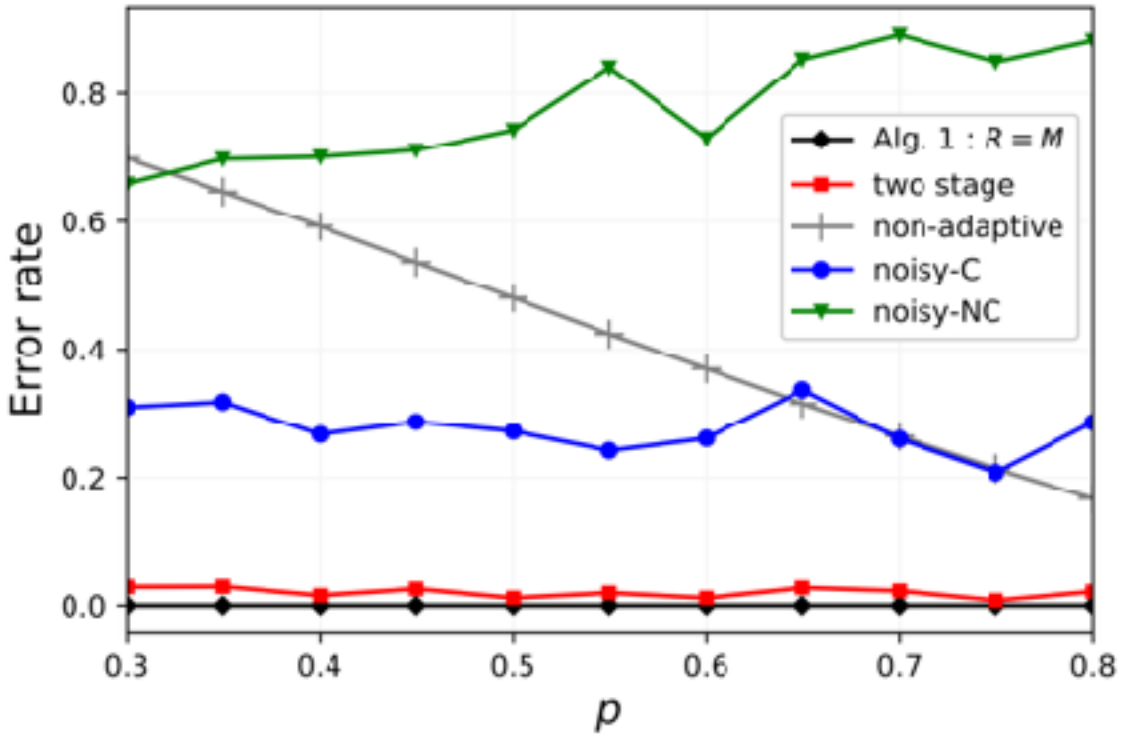


Sparse regime



- Improvement: 50% to 85%.

Results - error rate



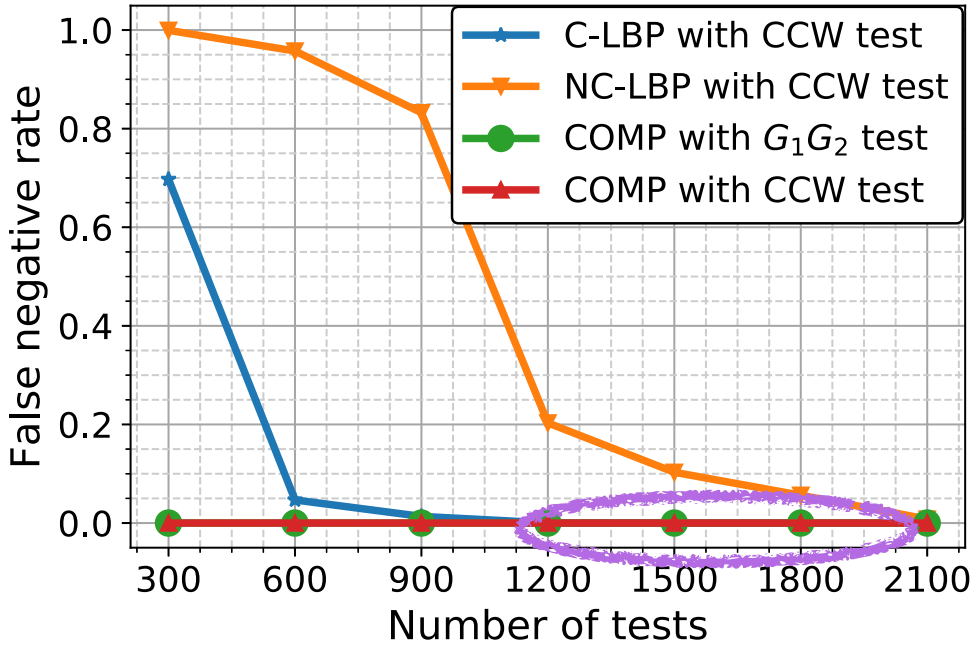
Experimental setup for overlapping families

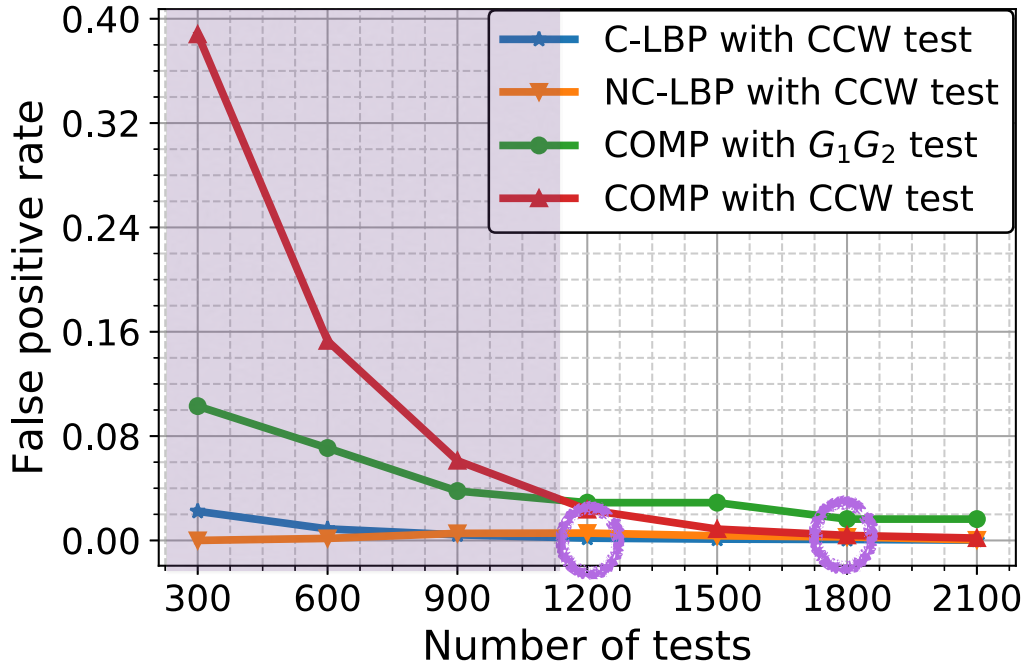
- 100 different communities, where:
  - 3000 members belong to ~200 overlapping families of various sizes
  - each family is infected w.p. $q = 0.05$
  - infection rates $p_j \in [0.3,0.9]$
  - $5\,\%$ of members are infected on average—*examined other rates, too*
- Compare w.r.t. #tests and error rates

Results - avg #tests for adaptive algorithms

- Binary splitting
    - ~15.2x the lower bound
- Our algorithm
    - ~ 5.5x the lower bound, improvement: 63%
    - always below the counting bound

Results - avg error rate for nonadaptive algorithms



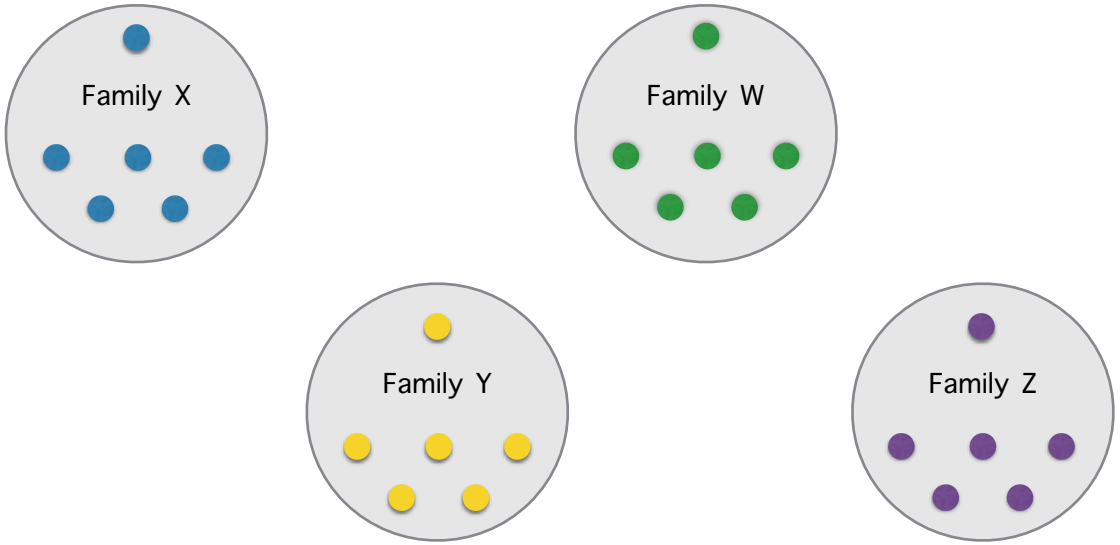


- If 0 FN-rate required and tests are few —> community-aware encoding
- If more tests available —> community-aware LBP

Other models

Stochastic Block Model

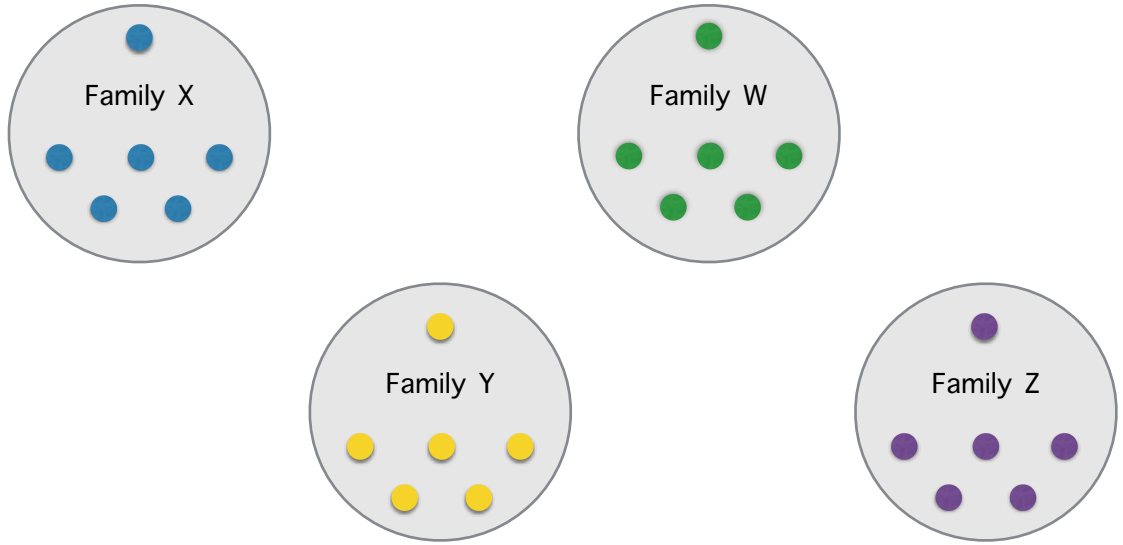
- n vertices divided into m families
  - k members in each family
- Seed selection: initially infect each vertex with prob. p
- Each seed:
  - infects members of the same family w.p. p_1
  - infects members of other families w.p. p_2



- S. Ahn, W. -N. Chen and A. Özgür, "Adaptive Group Testing on Networks with Community Structure," *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 1242-1247, doi: 10.1109/ISIT45174.2021.9517888.

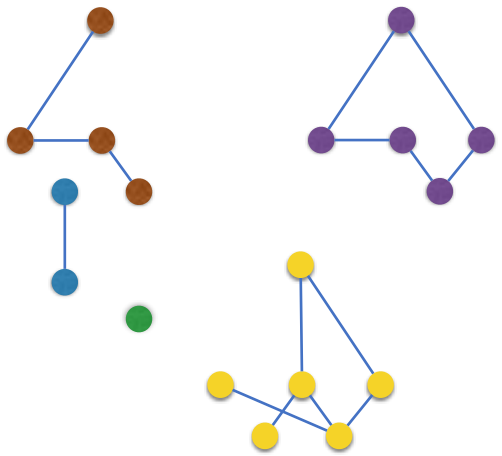
Adaptive Algorithm

- Pool together all the members of each family and use binary splitting to identify which families have at least one infected member.
- Perform binary splitting within each infected family to identify infected members.



Cluster formation tree model

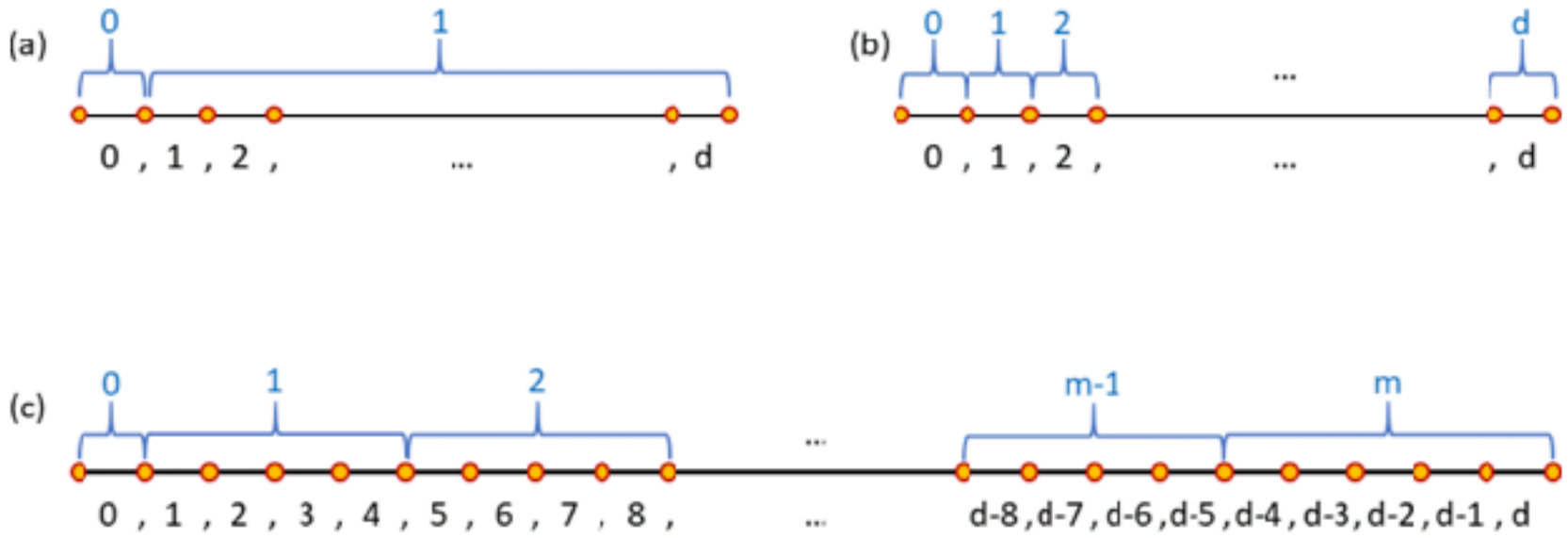
- Create a random connection graph - any two vertices connected with probability p .
- if there is a path between two vertices, their status is equal (both infected or not)
- patient ``zero" selected to be infected.
- work analyzes algorithms for cluster formation trees (a specific way to form clusters).



- B. Arasli and S. Ulukus, "Group testing with a graph infection spread model," 2021, arXiv:2101.05792.

Semiquantitative group testing (SGT)

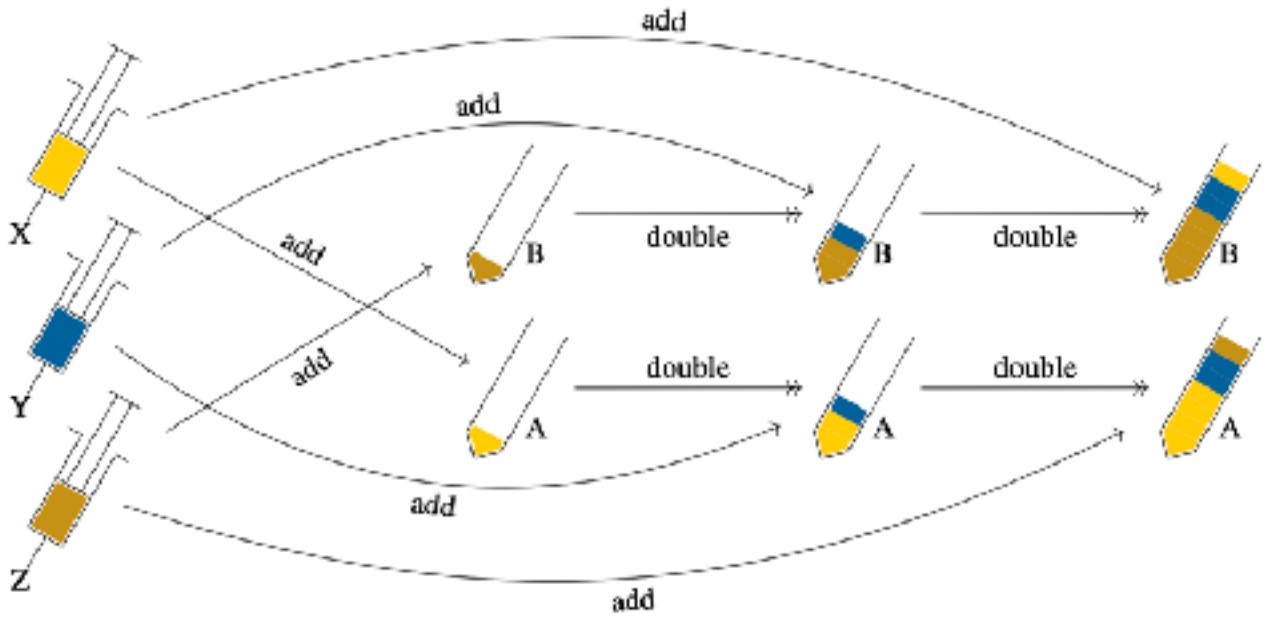
- The result of the test is a nonbinary value that depends on the number of defectives through a fixed set of thresholds.



AC-DC: Amplification Curve Diagnostics for Covid-19 Group Testing, Ryan Gabrys, Srilakshmi Pattabiraman, Vishal Rana, João Ribeiro, Mahdi Cheraghchi, Venkatesan Guruswami, Olgica Milenkovic, ArXiv 2021

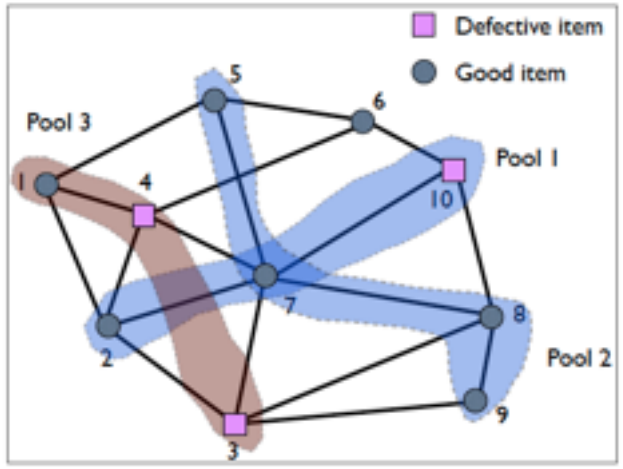
Tropical group testing

- PCR operates in cycles where at each cycle the concentration of DNA in each “tube” doubles.
- Idea: leverage SGT and delay by adding specimens into each tube at different cycles.



Graph Constrained Group Testing

Each pool test must conform to the constraints imposed by a graph
eg: can only pool together vertices on the same path



M. Cheraghchi, A. Karbasi, S. Mohajer and V. Saligrama, "Graph-constrained group testing," *2010 IEEE International Symposium on Information Theory*, 2010

S. Sihag, A. Tajer and U. Mitra, "Adaptive Graph-Constrained Group Testing," in *IEEE Transactions on Signal Processing*, vol. 70, pp. 381-396, 2022, doi: 10.1109/TSP.2021.3137026.

Take-away messages

We can *significantly* improve group-testing efficiency,
in terms of #tests and reliability
if we take into account the *community structure*
in test *designs* or *decoding*.

- Simplified first examples but approach extends broadly
- Many aspects are still open, but first results are promising

Dynamic Case

Group testing can help, but…

Traditional group-testing assumes
a *static* setting

while…infections are *dynamic*.

We need to focus on *dynamic* infections

Can we identify all new infections every day
using *fewer* tests than complete individual testing?

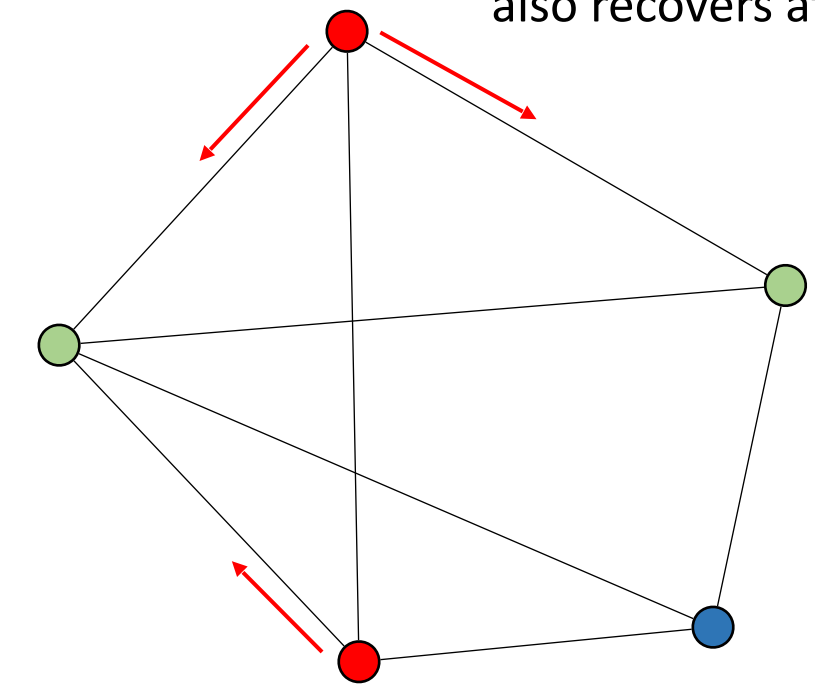
Can we reuse existing GT algorithms
or we need to develop new ones?

S. R. Srinivasavaradhan, P. Nikolopoulos, C. Fragouli and S. Diggavi, "An entropy reduction approach to continual testing," *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 611-616, doi: 10.1109/ISIT45174.2021.9518188.

SIR stochastic network model

- Infected nodes can transmit the infection to a susceptible neighbor.
- Infected nodes can get recovered, and stay recovered forever.
- The time for an infection to be transmitted over an edge and time for an infected person to recover are exponentially distributed. We call the mean of these distributions as the “rate” of infection and recovery.

Transmits the disease to each neighbor at rate β, also recovers at rate γ

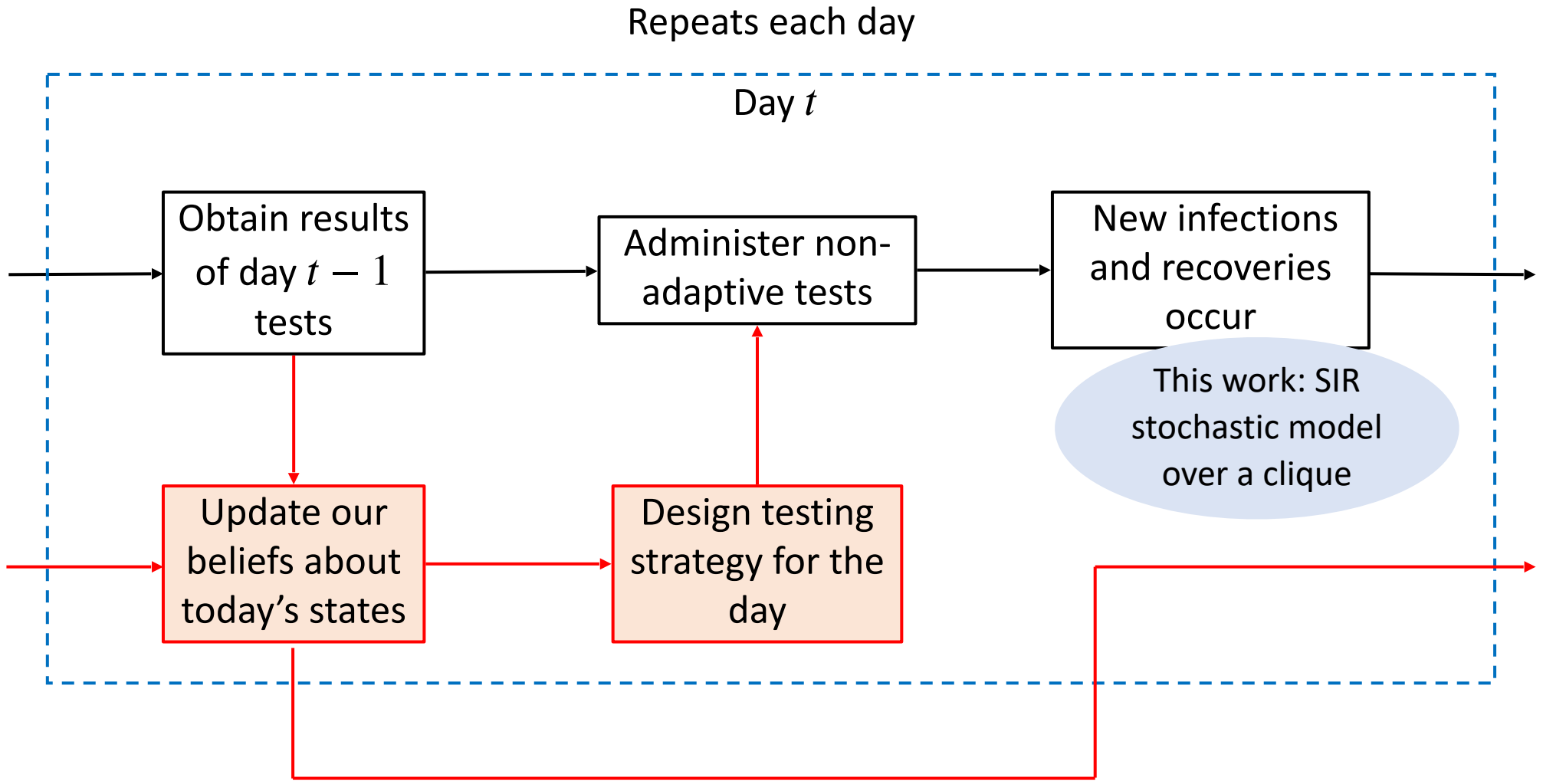


Susceptible

Infected

Recovered

State estimation on SIR network model (no intervention)



Dynamic testing on SIR network model

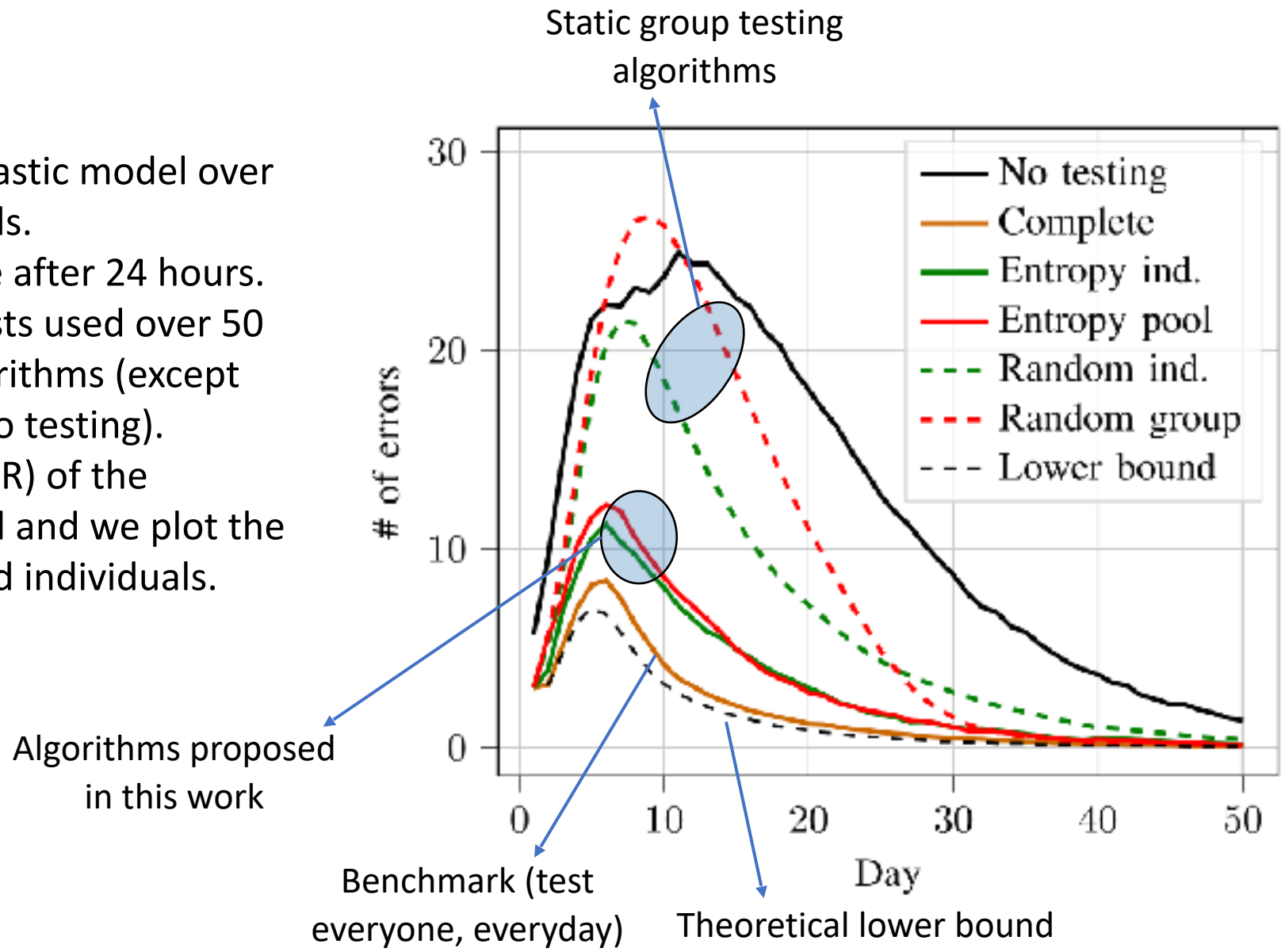
Here: SIR stochastic model over a clique

- Clique is representative of a well-mixed and closely-knit community.
- Simplest network one needs to understand before delving into sophisticated ones.
- Our results naturally extend to networks represented by multiple, non-overlapping cliques.
- E.g. Schools, universities, office spaces.
- In our simulations, we further enrich our model with an exogenous input that can infect individuals i.i.d with some small probability.

- S. R. Srinivasavaradhan, P. Nikolopoulos, C. Fragouli and S. Diggavi, "An entropy reduction approach to continual testing," *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 611-616, doi: 10.1109/ISIT45174.2021.9518188.

Illustrative experimental results (no intervention)

- Simulation of SIR stochastic model over a clique of 50 individuals.
- Testing results available after 24 hours.
- The total number of tests used over 50 days is fixed for all algorithms (except complete testing and no testing).
- Each day the states (S,I,R) of the individuals are decoded and we plot the number of misidentified individuals.



Entropy lower bound

- As a first question, we ask how many tests are required to achieve same performance as testing every individual everyday?
- Denote by $U_i^t := \mathbb{1}\{X_i^t = 1\}$, the random variable which tracks if an individual is in the infected state at a given time.
- Theorem: Suppose you test at times $\{t_1, t_2, t_3, \dots\}$, where $t_l > t_{l-1}\ \forall\ l$, then the expected number of tests needed to identify all infected individuals at time t_l:

$$\mathbb{E}\left[T^{t_l}\right] \geq H\left(U_1^{t_l}, U_2^{t_l}, \dots, U_n^{t_l} \;\middle|\; X_1^{t_{l-1}}, X_2^{t_{l-1}}, \dots, X_n^{t_{l-1}}\right), \tag{1}$$

where n is the size of the population.

- Contribution: exactly calculate the entropy term in (1) for the SIR clique model.
- Observation: the number of tests to use each day must depend on the uncertainty in the state of the system.

Main algorithmic idea – entropy reduction

- Known idea, time and again employed in adaptive group testing.
- Each test must give maximal information.
- E.g. Laminar algorithm from [Li2014] and binary search to find exactly one infected individual in a population.
- The number of tests to use each day must depend on the uncertainty in the system – more uncertain we are about the states, more the tests we require.

[Li2014] Li, Tongxin, et al. "Group testing with prior statistics." *2014 IEEE International Symposium on Information Theory*. IEEE, 2014.

Entropy reduction – individual testing

- Each day, update the marginal distribution of states $\Pr(X_i^{(t)} = x)$, where $X_i^{(t)}$ is the state of the individual i at time t and $x \in \{S, I, R\}$.
- Given T tests, find the T individuals whose marginal probability of being infected $\Pr(X_i^{(t)} = I)$ is closest to 0.5 and test them individually.
- Each day we choose a number of tests proportional to the uncertainty in the system $H\left(U_1^{t_l}, U_2^{t_l}, \ldots, U_n^{t_l} \;\middle|\; X_1^{t_{l-1}}, X_2^{t_{l-1}}, \ldots, X_n^{t_{l-1}}\right)$.

Entropy reduction – pooled testing

- Similar strategy when the pooled tests are constrained to have non-overlapping pools.
- Each day, update the marginal distribution of states $\Pr(X_i^{(t)} = x)$, where $X_i^{(t)}$ is the state of the individual i at time t and $x \in \{S, I, R\}$.
- Given T tests, find T non-overlapping subsets (pools) of individuals such that the marginal probability of each pool being infected is as close to 0.5 as possible. This problem is related to the subset-sum problem and is not easy.
- We propose a heuristic where we first pick out the T individuals whose probability of infection is closest to 0.5 and add individuals to move the probability of infection closer to 0.5.
- Each day we choose a number of tests proportional to the uncertainty in the system

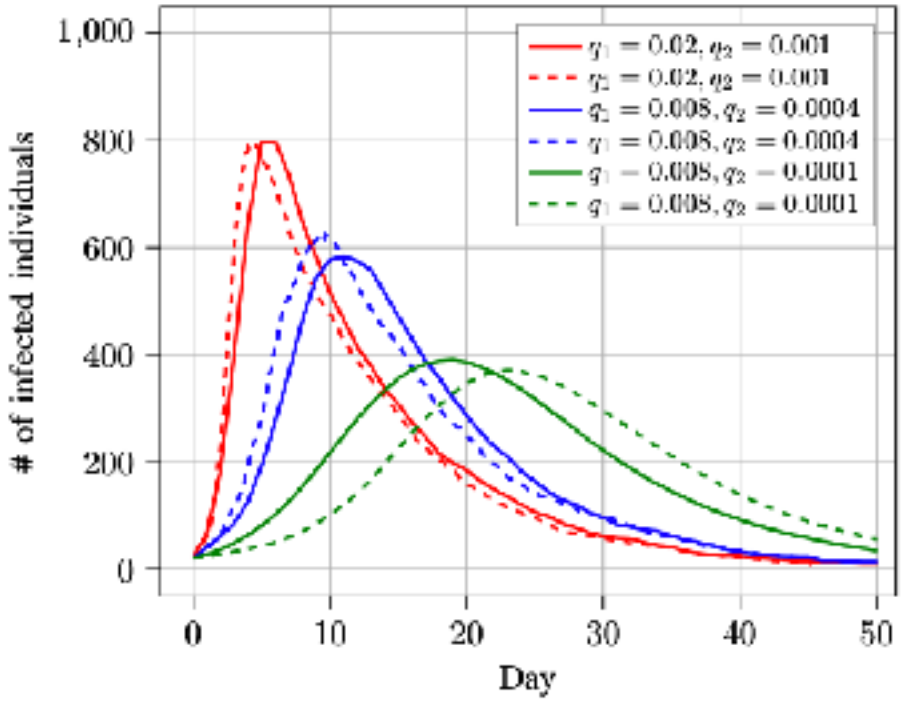
$$H\left(U_1^{t_l}, U_2^{t_l}, \ldots, U_n^{t_l} \;\middle|\; X_1^{t_{l-1}}, X_2^{t_{l-1}}, \ldots, X_n^{t_{l-1}}\right).$$

This example:

- We demonstrated that the knowledge of the dynamics of disease spread increases the efficiency and performance of testing.
- On the other hand, we also derived lower bounds on the performance of any testing strategy.
- It remains to see how the idea of entropy reduction can be applied to group testing where each individual participates in multiple tests.
- Analysis of testing and interventions challenging on continuous model.

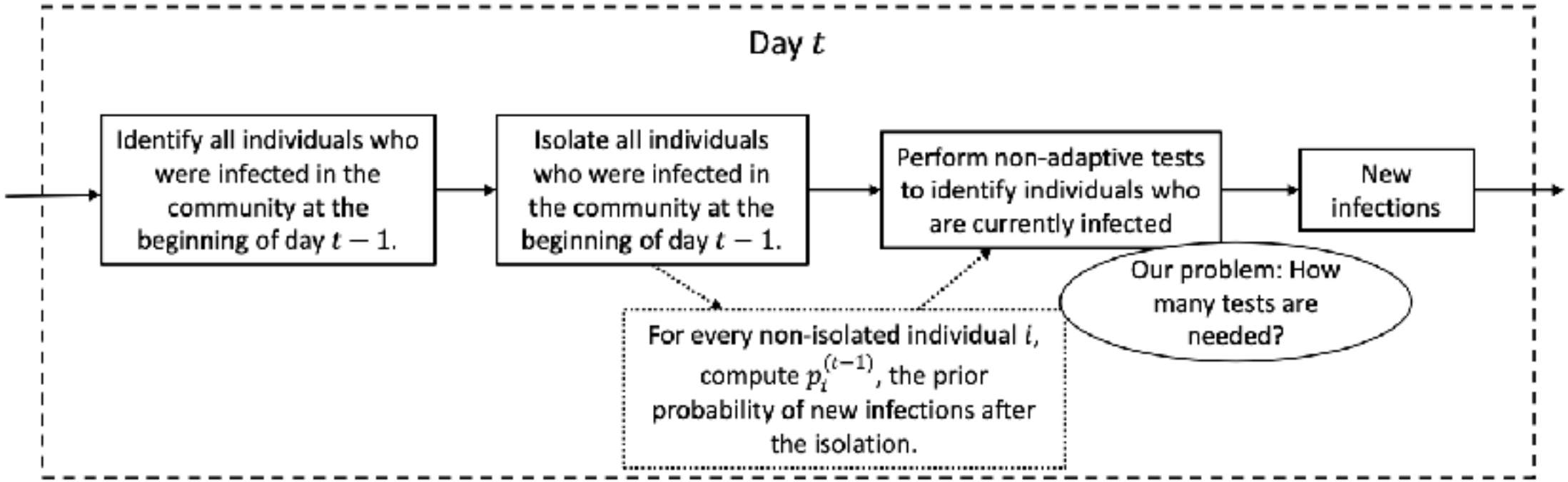
A discrete-time SIR network model

- inherits basic SIR properties
  - infections go from infected to susceptible
  - infections and recoveries are stochastic
  - common disease progression patterns
- but, it is *discrete-time*:
  - no individual can get infected and recovered in the same day
  - more amenable to analysis under testing delays and quarantines



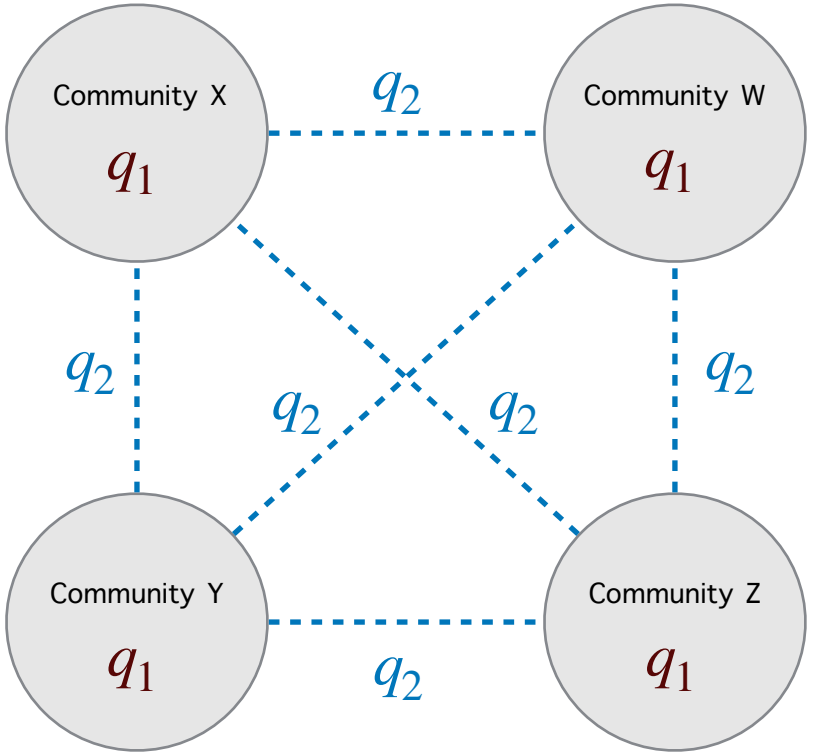
- S. R. Srinivasavaradhan, P. Nikolopoulos, C. Fragouli and S. Diggavi, "Dynamic group testing to control and monitor disease progression in a population," *arXiv e-prints*,arXiv:2106.10765. June 2021.

Dynamic testing and intervention strategy



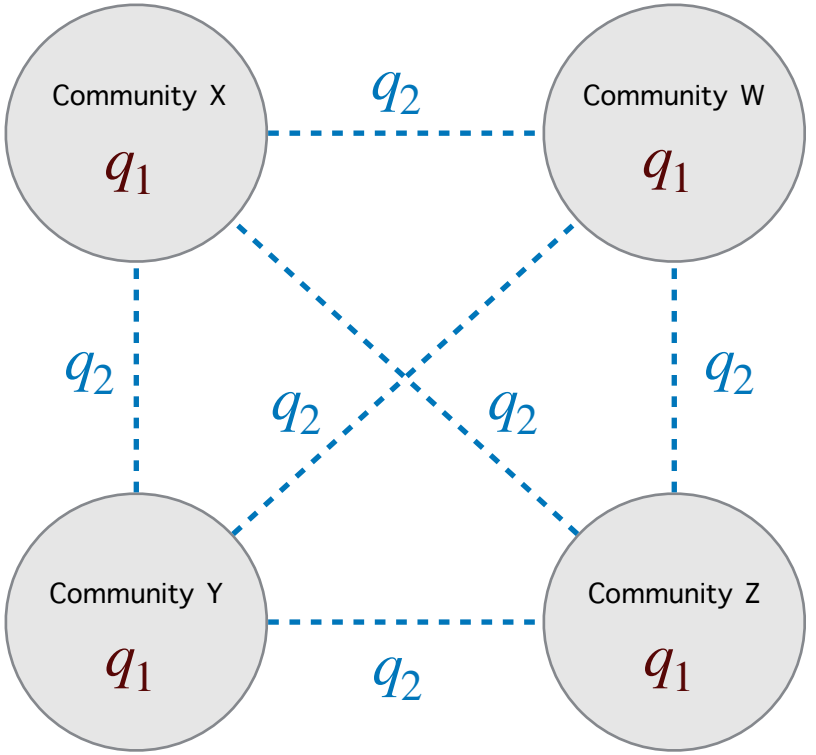
Community and infection model (discrete-time SIR)

- N individuals partitioned into communities of same size C
- Initial infections —> i.i.d. w.p. p_{init}
- Every-day infections follow a *stochastic block model*
  - Each infected individual infects a susceptible of the *same* community w.p. q_1
  - Each infected individual infects a susceptible of a *different* community w.p. q_2
  - Each infected recovers independently w.p. r



An every-day testing problem

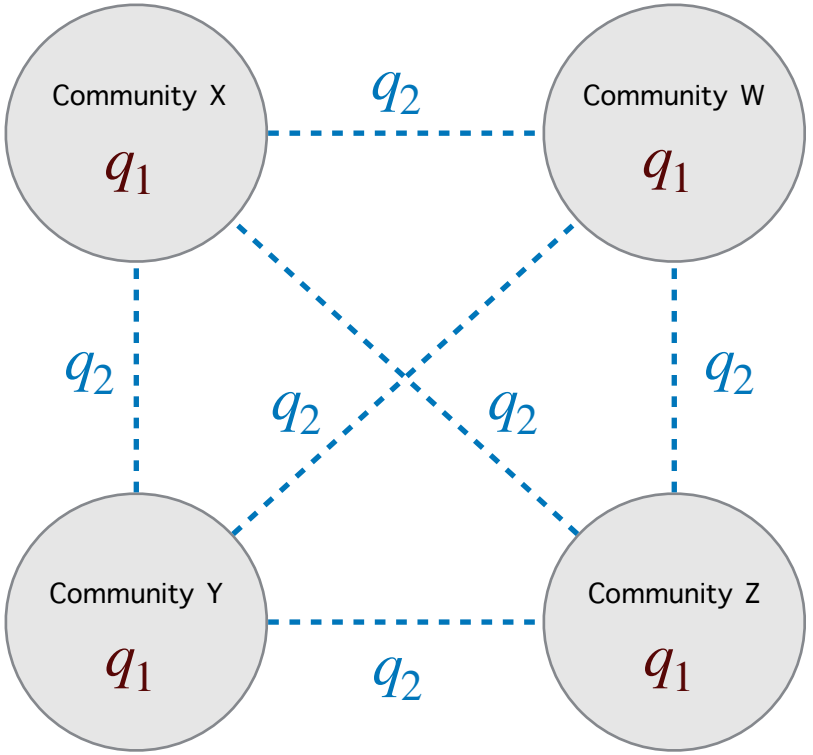
- Assumptions:
  - testing delay = 1 day (as with PCR)
  - only individuals tested positive isolate
  - isolated ones return after recovery
- Strategy:
  - test every day, isolate the next day, re-test
- Questions:
  - how many non-adaptive group tests are needed to identify all infections each day?
  - optimal designs?



Reduction to the static case with prior statistics

- Let $I_j^{(t)}$ be the number of infected individuals inside community j, on day t
- Infection probability for each susceptible individual in j, on day t:

$$p_j^{(t)} = 1 - (1 - q_1)^{I_j^{(t)}} (1 - q_2)^{\sum_{j' \neq j} I_{j'}^{(t)}}$$



- If all $I_j^{(t)}, \forall j$ are known,
  susceptible individuals are infected *independently* w.p. $p_j^{(t)}$

Overview of results

- Lower bound on the number of tests for non-identical priors
    - alternative of the information theoretic lower bound
- *Under constraints about priors*, existing GT algorithms are order-optimal
- These constraints about priors are satisfied for specific p_{init}, q_1, q_2, N, C

- *Existing GT algorithms are order-optimal in specific dynamic settings*

Result 1: Lower bound

Theorem (non-adaptive static case with non-identical priors):

If $p_i \leq 0.5$, to achieve a probability of error $\to 0$ as $N \to \infty$, we need:

$$T(\mathbf{p}) = \Omega\left(\min\{Np_{min}\log N, N\}\right), \qquad \text{where } p_{min} \triangleq \min_{i\in[N]} p_i \ .$$

Theorem 2.1 from [*W.H.Bay, E.Price, and J.Scarlett, "Optimal Non-Adaptive Probabilistic Group Testing in General Sparsity Regimes," 2020*]:

Consider the i.i.d. group testing setting. If $p \leq 0.5$, any group testing strategy having probability of error $\to 0$ as $N \to \infty$ must use:

$$T = \Omega\left(\min\{Np\log N, N\}\right) \text{ tests.}$$

Result 1: Lower bound

Theorem (non-adaptive static case with non-identical priors):

If $p_i \leq 0.5$, to achieve a probability of error $\rightarrow 0$ as $N \rightarrow \infty$, we need:

$$T = \Omega\left(\min\{Np_{min}\log N, N\}\right), \qquad \text{where } p_{min} \triangleq \min_{i\in[N]} p_i \ .$$

Proof outline:

- Consider $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ and $\mathbf{p_{min}} = (p_{min}, p_{min}, \ldots, p_{min})$
- Under the optimal decoder, any test design that achieves the minimum error probability for $\mathbf{p}$, achieves an even smaller error probability for $\mathbf{p_{min}}$ => $T(\mathbf{p_{min}}) \leq T(\mathbf{p})$
- Apply the theorem from the i.i.d. case and conclude

Result 1: Lower bound

Theorem (non-adaptive static case with non-identical priors):

If $p_i \leq 0.5$, to achieve a probability of error $\to 0$ as $N \to \infty$, we need:

$$T(\mathbf{p}) = \Omega\left(\min\{Np_{min}\log N, N\}\right), \qquad \text{where}$$
$$p_{min} \triangleq \min_{i\in[N]} p_i \,.$$

A useful corollary:

Let $\eta \in [1,\infty)$ be a fixed constant w.r.t. N, and $\dfrac{\max_i p_i}{\min_i p_i} \leq \eta$, then

$$T(\mathbf{p}) = \Omega\left(\min\{Np_{mean}\log N, N\}\right) = \Omega\left(\min\{Np_{max}\log N, N\}\right)$$

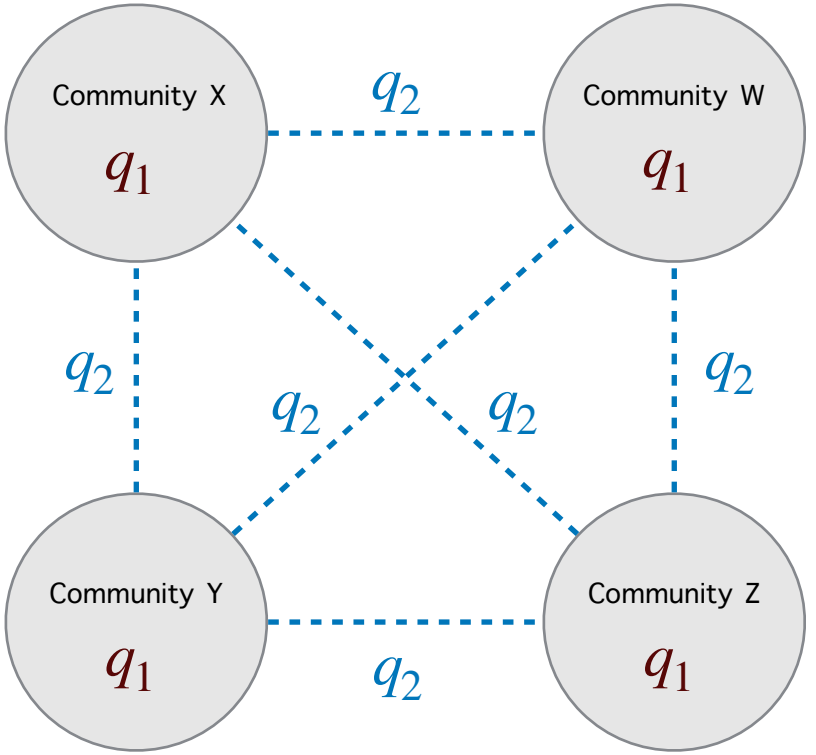
Result 2: Known GT algorithms can be order-optimal

- Coupon collector algorithm (CCA) achieves error probability $2N^{-\delta}$ with $T \leq 4e(1+\delta)Np_{mean}\log N$
  - If $Np_{mean}\log N < N$, *CCA* is order-optimal
  - else individual testing is order-optimal

- Randomized designs for iid priors w.p. p' achieve vanishing error probability with $T = O(Np'\log N)$
  - If $Np_{max}\log N < N$, *randomized* designs for $p' = p_{max}$ are order-optimal
  - else individual testing is order-optimal

Result 3: Bridging the gap with the dynamic case

- N individuals partitioned into communities of same size C
- Initial infections —> i.i.d. w.p. p_{init}
- Every-day infections follow a *stochastic block model*
    - Each infected individual infects a susceptible of the *same* community w.p. q_1
    - Each infected individual infects a susceptible of a *different* community w.p. q_2



Result 3: Bridging the gap with the dynamic case

Theorem (dynamic case):

- If $p_{init} \leq 0.5$, $q_1 \leq \dfrac{1-1/\sqrt{2}}{C}$, $q_2 = \dfrac{1-1/\sqrt{2}}{N}$, then $p_j^{(t)} \leq 0.5$, on any day t
- If $\dfrac{q_1}{q_2} \leq \eta$, then $\dfrac{\max_i p_i}{\min_i p_i} \leq \eta$, on any day t

- CCA and randomized designs are order-optimal in such scenarios
- Result holds not just for the specific model

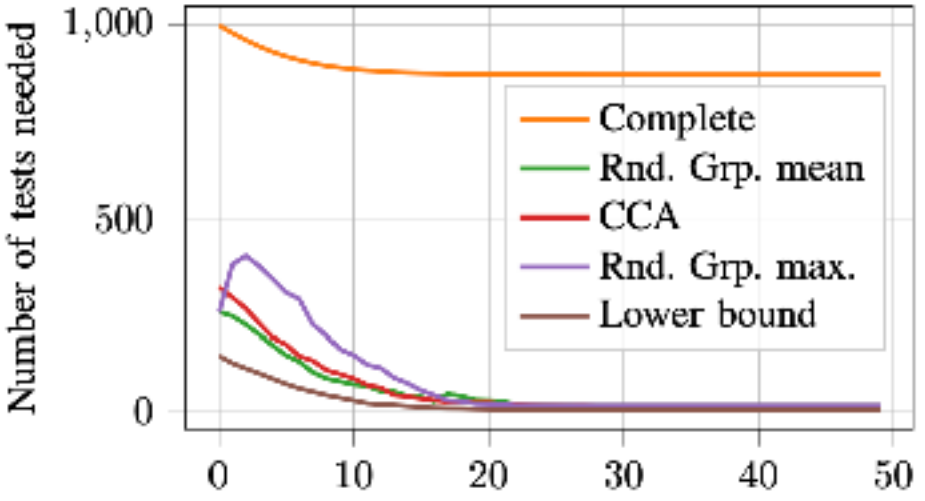
Experimental results

Setup

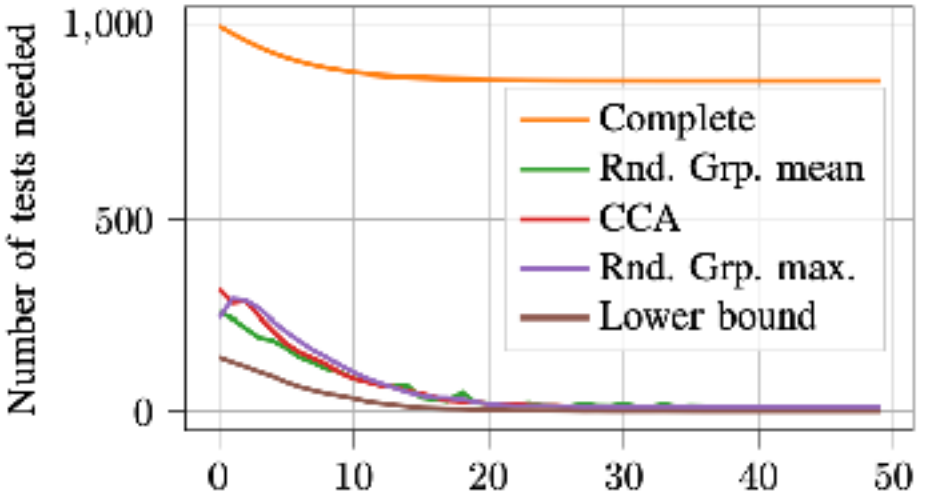
- CCA vs Randomized (with p_{mean} or p_{max})
- Metric: average min #tests over 200 trajectories

Results

- Randomized (with p_{mean}) is the best
- All GT strategies need ~80% fewer tests than complete testing



(a) $(N, C, p_{\text{init}}, q_1, q_2) = (1000, 20, 0.02, 0.03, 0.0004)$.



(b) $(N, C, p_{\text{init}}, q_1, q_2) = (1000, 50, 0.02, 0.012, 0.0004)$.

In this work:

- Considered a dynamic community and infection model (inspired by SIR)
- Showed that this reduces to a static case if all infections are known each day
- Provided a lower bound on the number of tests each day
- Provided conditions for which existing GT algorithms are order-optimal
- Numerically verified the performance of CCA and randomized designs over 200 random dynamic settings

  —> reduced #tests by 80% compared to individual testing

Some Open Questions

- Benefits from community structure? What if limited knowledge?
- What are trade-offs we can achieve given a small number of tests? (eg number of tests is limited below what is needed for independence?)
- What if we have additional information (non binary test outputs?) and constraints (graph constraints? different types of tests?)
- What are good epidemiology models to use?
- Can data-based techniques help to optimize the designs?
- …

Thank you!