Group Testing on a Network

AAAI'21

Arlei Silva, Ambuj Singh

Computer Science Department
University of California, Santa Barbara, CA.

[1]Research funded by NSF IIS-1817046, DTRA HDTRA1-19-1-0017 and a UCSB Office of Research VCR COVID-19 Seed Grant.

COVID-19: Timeline

Deutsche Welle

China investigates SARS-like virus as dozens struck by pneumonia

Chinese health authorities on Tuesday said they are investigating 27 cases of viral pneumonia in central Hubei province, amid online ...

Dec 31, 2019

NBC News

China reports first death from outbreak of mystery virus

China reports first death from outbreak of mystery virus. Cases of a new type of coronavirus mystery flu have stoked fears of an outbreak not ...

Jan 11, 2020

AP Associated Press

Washington man is 1st in US to catch new virus from China

SEATTLE (AP) — The U.S. on Tuesday reported its first case of a new and potentially deadly virus circulating in China, saying a Washington state resident who ...

Jan 21, 2020

CNBC

Italy's hospitals reach breaking point as Rome boosts funding to help virus-hit economy

Lombardy is the epicenter of Italy's outbreak with 1254 cases of the virus, 38 deaths and where 10 towns remain under lockdown. Hospitals in Lombardy are ...

Mar 3, 2020

Bloomberg

U.S. Deaths Could Reach 200,000; Spain Toll Rises: Virus Update

U.S. Deaths Could Reach 200,000; Spain Toll Rises: Virus Update. Bloomberg News. March 28, 2020, 2:44 PM PDT Updated on March 29, 2020, 2 ...

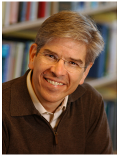
Mar 28, 2020

Wall Street Journal

Testing Is Our Way Out

Testing Is Our Way Out. Returning to normal is too dangerous. Lockdowns are unsustainable. Let's save lives without a depression. By Paul ...

Apr 2, 2020

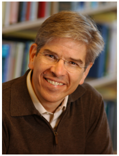


Paul Romer.

Population-scale testing:
''The intervention is based on: (1) test every individual (2) repeatedly, and (3) isolation of infected individuals". Around 23M tests/day, $75B. Taibale et al.[1]

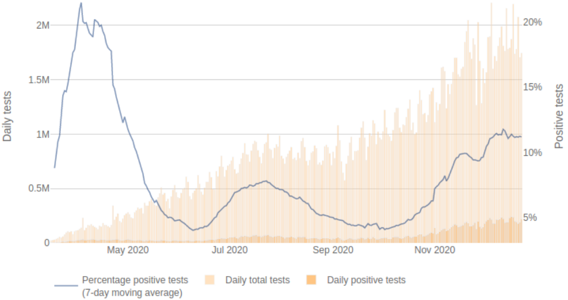
[1] *[TRL20] Population-scale testing can suppress the spread of COVID-19*. Taipale, Romer and Linnarsson. Preprint, 2020.

[2] The COVID tracking project.



Paul Romer.

Population-scale testing:
''The intervention is based on: (1) test every individual (2) repeatedly, and (3) isolation of infected individuals". Around 23M tests/day, \$75B. Taibale et al.[1]



[1] *[TRL20] Population-scale testing can suppress the spread of COVID-19*. Taipale, Romer and Linnarsson. Preprint, 2020.

[2] The COVID tracking project.

Vox

The CDC's struggle to get coronavirus test kits out, explained

We don't know because the US has been extremely slow to roll out diagnostic testing for the Covid-19 disease. It's unclear if there's a specific policy or decision to ...

Mar 6, 2020

Politico

Coronavirus testing hits dramatic slowdown in U.S.

The number of coronavirus tests analyzed each day by commercial labs in the U.S. plummeted by more than 30 percent over the past week, even though new ...

Apr 14, 2020

NPR

Coronavirus Testing Machines Are Latest Bottleneck In ...

Coronavirus testing in the U.S. has run into a number of snags, from a lack of nasal swabs to not enough chemicals needed to run the tests. Now there's a new ...

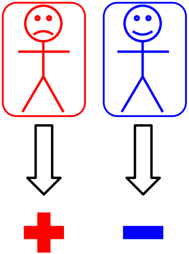
May 28, 2020

qRT-PCR test ($100)

1. Swab collection;
2. Transport media;
3. RNA purification;
4. Reverse transcription and quantitative PCR.

(a) Collection

(b) PCR



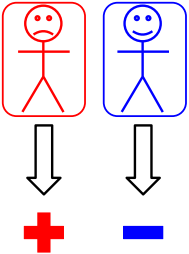
(c) Individual testing.

qRT-PCR test ($100)

1. Swab collection;
2. Transport media;
3. RNA purification;
4. Reverse transcription and quantitative PCR.

(a) Collection

(b) PCR



(c) Individual testing.

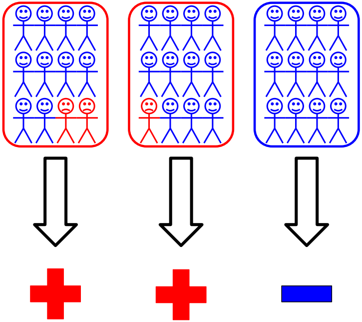
Robert Dorfman.

''Suppose that after the individual blood sera are drawn they are pooled in groups of, say, five and that groups rather than individual sera are subjected to chemical analysis". Dorfman, 1943.[1]

[1] [Dor43] Dorfman. *The Detection of Defective Members of Large Populations.*

''Suppose that after the individual blood sera are drawn they are pooled in groups of, say, five and that groups rather than individual sera are subjected to chemical analysis". Dorfman, 1943.[1]

Robert Dorfman.

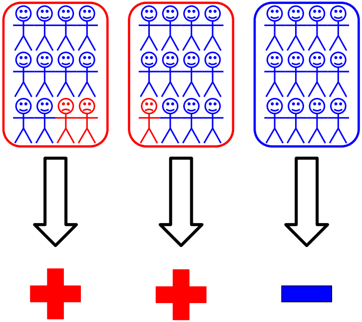


Group testing: Stage 1

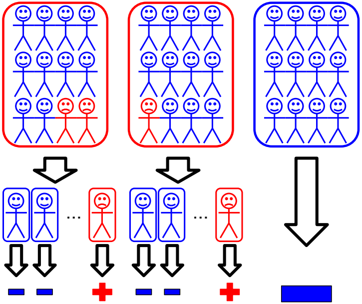
[1] [Dor43] Dorfman. *The Detection of Defective Members of Large Populations.*

"Suppose that after the individual blood sera are drawn they are pooled in groups of, say, five and that groups rather than individual sera are subjected to chemical analysis". Dorfman, 1943.[1]

Robert Dorfman.



Group testing: Stage 1

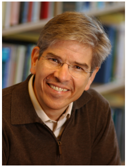


Group testing: Stage 2

[1] [Dor43] Dorfman. *The Detection of Defective Members of Large Populations.*

Group Testing

Prevalence	Group Size	Savings (%)
1	11	80
5	5	57
10	4	41
30	3	1



$ 75B

Groups assembled at random

In practice, group size bounded by dillution

CNN

Fauci says task force 'seriously considering' new testing strategy

Anthony Fauci, the nation's top infectious disease expert, said Friday that the White House coronavirus task force is "seriously considering" pool testing for Covid- ...

Jun 26, 2020

Used in the US, China, Germany and India

Literature

Original paper, two-round testing (1943) [Dor43]

Groups assembled at random.

Later applications in production lines, computer networks etc.

Books on group testing:

Combinatorial group testing and its applications [DHH00]
Pooling designs and non-adaptive group testing [DH06]

The benefit of correlated groups:

FAST: a Feasible, Accurate and Speedy Test Strategy for COVID-19 [FLJY20]

Group Testing in a Pandemic: The Role of Frequent Testing, Correlated Risk, and Machine Learning [AKOW20]

Community aware group testing [NGFD20]

Literature

Original paper, two-round testing (1943) [Dor43]

Groups assembled at random.

Later applications in production lines, computer networks etc.

Books on group testing:

Combinatorial group testing and its applications [DHH00]
Pooling designs and non-adaptive group testing [DH06]

The benefit of correlated groups:

FAST: a Feasible, Accurate and Speedy Test Strategy for COVID-19 [FLJY20]

Group Testing in a Pandemic: The Role of Frequent Testing, Correlated Risk, and Machine Learning [AKOW20]

Community aware group testing [NGFD20]

How to group?

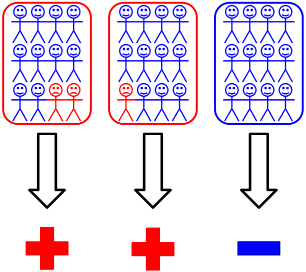
Can network science improve the cost-effectiveness of group testing in an epidemic?

Group Testing on a Network

1. Massive group testing;
2. Contact tracing;

Group Testing on a Network

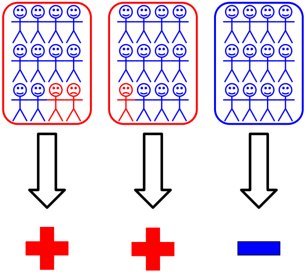
1. Massive group testing;
2. Contact tracing;



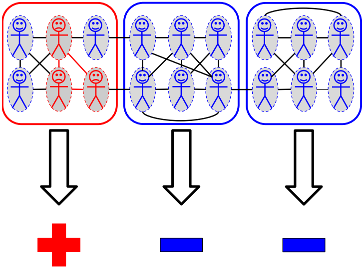
Group testing.

Group Testing on a Network

1. Massive group testing;
2. Contact tracing;



Group testing.



Group testing on a network.

Problem Definition

Contact network: Graph G(V,E,W)

Individuals as nodes V; Contacts as edges E;

Infection probabilities as weights $W : E \rightarrow [0, 1]$

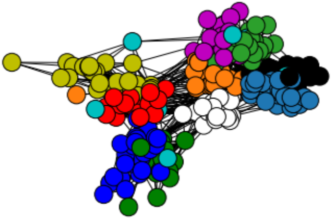


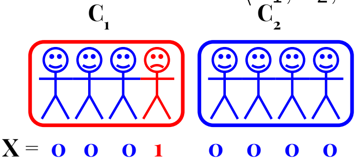
Figure: Contact network from a primary school.[2]

<u>GTN</u>: Given a transmission network G, epidemic parameters θ and a maximum group size k, partition the vertices V into groups $\{C_1, \ldots C_m\}$ such that $|C_i| \leq k$, $\forall i$, and the expected number of tests $\sigma(G, \theta, C)$ to screen V is minimized.

[2] [SVB$^+$11] *High-resolution measurements of face-to-face contact patterns in a primary school*. Stehlé et al. PloS one, 2011.

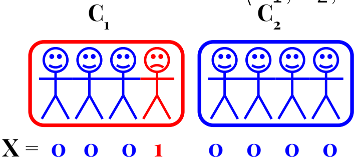
Why grouping matters?

Multivariate Bernoulli outcomes $X = \langle X_1, X_2, \ldots X_n \rangle$, $n = |V|$



Why grouping matters?

Multivariate Bernoulli outcomes $X = \langle X_1, X_2, \ldots X_n \rangle$, $n = |V|$

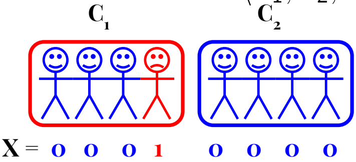


$$\text{Expected tests } \sigma(G, \theta, C) = m + \sum_{j=1}^{m} |C_j| \times \textbf{Prob}(\sum_{v \in C_j} X_v \geq 1)$$

$$= n + m - \sum_{j=1}^{m} |C_j| \times \mathbb{E}[\prod_{v \in C_j} (1 - X_v)]$$

Why grouping matters?

Multivariate Bernoulli outcomes $X = \langle X_1, X_2, \ldots X_n \rangle$, $n = |V|$



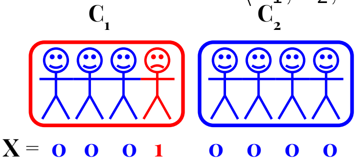
$$\text{Expected tests } \sigma(G, \theta, C) = m + \sum_{j=1}^{m} |C_j| \times \textbf{Prob}(\sum_{v \in C_j} X_v \geq 1)$$

$$= n + m - \sum_{j=1}^{m} |C_j| \times \mathbb{E}[\prod_{v \in C_j} (1 - X_v)]$$

Random groups: $\mathbb{E}[\prod_{v \in C_m} (1 - X_v)] = (1 - p)^k$

Why grouping matters?

Multivariate Bernoulli outcomes $X = \langle X_1, X_2, \ldots X_n \rangle$, $n = |V|$



$$\text{Expected tests } \sigma(G, \theta, C) = m + \sum_{j=1}^{m} |C_j| \times \textbf{Prob}(\sum_{v \in C_j} X_v \geq 1)$$

$$= n + m - \sum_{j=1}^{m} |C_j| \times \mathbb{E}[\prod_{v \in C_j} (1 - X_v)]$$

Random groups: $\mathbb{E}[\prod_{v \in C_m} (1 - X_v)] = (1 - p)^k$

Correlated groups: Need to account for covariances of X_v's

$k = 2$: $\mathbb{E}[(1 - X_u)(1 - X_v)] = 1 - p_u - p_v + p_u p_v + \textbf{cov}_{u,v}$

Hardness of GTN

<u>GTN</u>: Given a transmission network G, epidemic parameters θ and a maximum group size k, partition the vertices V into groups $\{C_1, \ldots C_m\}$ such that $|C_i| \leq k$, $\forall i$, and the expected number of tests $\sigma(G, \theta, C)$ to screen V is minimized.

Hardness of GTN

<u>GTN</u>: Given a transmission network G, epidemic parameters θ and a maximum group size k, partition the vertices V into groups $\{C_1, \ldots C_m\}$ such that $|C_i| \leq k$, $\forall i$, and the expected number of tests $\sigma(G, \theta, C)$ to screen V is minimized.

GTN is NP-hard

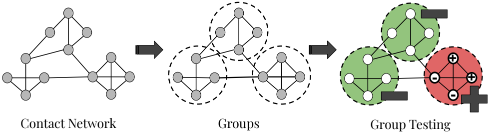
Reduction from 3-partition

Computing $\sigma(G, \theta, C)$ is #P-hard

Reduction from influence maximization

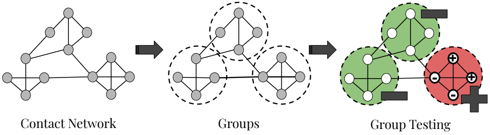
Heuristics: Topology vs Sampling-based

Topology-based: Graph clustering fixed size clusters



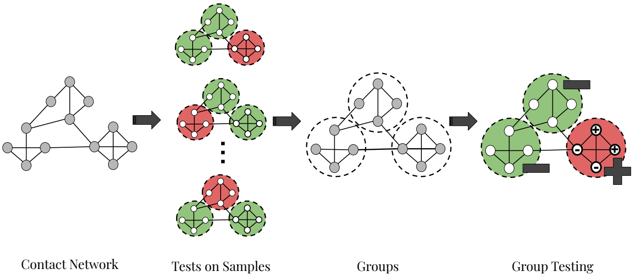
Heuristics: Topology vs Sampling-based

Topology-based: Graph clustering fixed size clusters



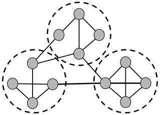
Sampling-based:

- Sample infection process z times based on parameters θ;
- Cluster nodes (bounded size k) minimizing tests over samples.



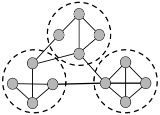
Heuristics: Grouping algorithm

Goal is to minimize score Δ



Heuristics: Grouping algorithm

Goal is to minimize score Δ



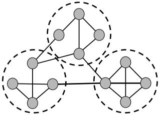
Greedy:

Initializes each vertex as a cluster;

Merges two clusters minimizing Δ

Heuristics: Grouping algorithm

Goal is to minimize score Δ



Greedy:

Initializes each vertex as a cluster;

Merges two clusters minimizing Δ

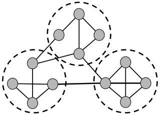
Kernighan-Lin (KL):

Initializes clusters;

Swaps vertices between clusters minimizing Δ

Heuristics: Grouping algorithm

Goal is to minimize score Δ



Greedy:

Initializes each vertex as a cluster;

Merges two clusters minimizing Δ

Kernighan-Lin (KL):

Initializes clusters;

Swaps vertices between clusters minimizing Δ

Four algorithms: $\{$Topology,Sampling$\}\times\{$Greedy,KL$\}$

Experiments

	$\lvert V \rvert$	$\lvert E \rvert$	**time**
High School (HS)	326	2,141	1 week
Erdos-Renyi (ER)	500	2,500	-
Gowalla (GW)	1,899	3,565	7 months

Table: Real contact data.

Network SIR simulations

Metric: average tests/person over 100 simulations

Baselines:

- Random [Dor43];
- Origami [KW08];
- Modularity [GN02].

Tests/person, 4% prevalence

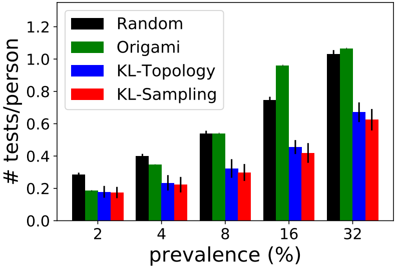
	Method	HS	GW	ER
No network	Random	.40 ± .01	.38 ± .01	.39 ± .01
	Origami	.35 ± .00	.34 ± .00	**.33 ± .00**
Topology	Modularity	.27 ± .04	.36 ± .01	.38 ± .02
	Greedy	.23 ± .05	.32 ± .01	.36 ± .02
	KL	.23 ± .05	.32 ± .01	.36 ± .02
Sampling	Greedy	.23 ± .05	.21 ± .03	.37 ± .02
	KL	**.22 ± .05**	**.20 ± .02**	.36 ± .02

Tests/person, 4% prevalence

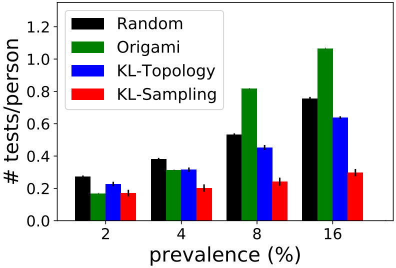
	Method	HS	GW	ER
No network	Random	.40 ± .01	.38 ± .01	.39 ± .01
	Origami	.35 ± .00	.34 ± .00	**.33 ± .00**
Topology	Modularity	.27 ± .04	.36 ± .01	.38 ± .02
	Greedy	.23 ± .05	.32 ± .01	.36 ± .02
	KL	.23 ± .05	.32 ± .01	.36 ± .02
Sampling	Greedy	.23 ± .05	.21 ± .03	.37 ± .02
	KL	**.22 ± .05**	**.20 ± .02**	.36 ± .02

KL-Sampling outperforms Random and Origami for most of the datasets and by up to 40% and 33%, respectively.

Tests/person, varying prevalence

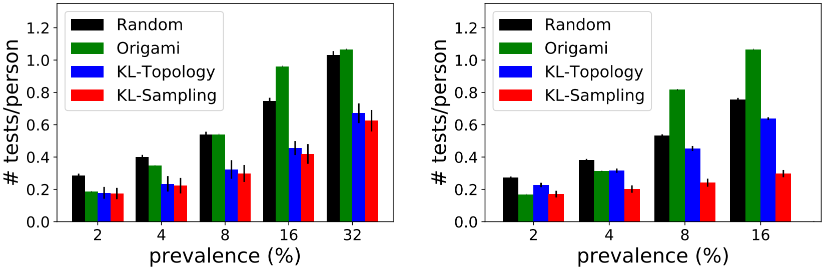


(a) High School (HS)



(b) Gowalla (GW)

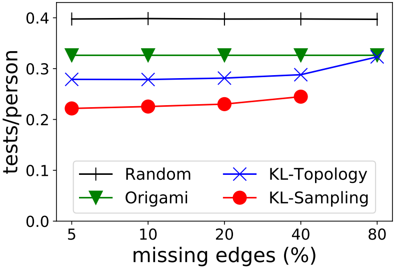
Tests/person, varying prevalence



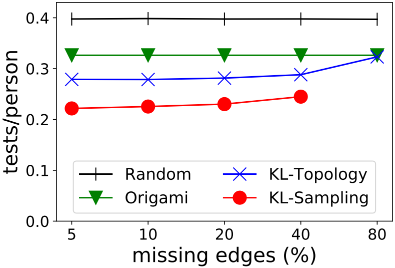
(a) High School (HS) (b) Gowalla (GW)

KL-Sampling outperforms all competing approaches for values of prevalence beyond 2% and is still effective when prevalence reaches 32%.

Tests/person, missing edges



Tests/person, missing edges



Savings are quite robust to missing edges.

Conclusions

We have proposed group testing on a network. We have formalized the problem, characterized its computational hardness, and proposed heuristics for it.

Experiments show that our approaches:

- Save up to 33% of resources for a prevalence of 4%;
- Are still effective for higher values of prevalence (32%);
- Are robust to missing edges in the transmission network.

Group Testing on a Network

AAAI'21

Arlei Silva, Ambuj Singh

Computer Science Department
University of California, Santa Barbara, CA.

References I

[AKOW20] Ned Augenblick, Jonathan T Kolstad, Ziad Obermeyer, and Ao Wang.
Group testing in a pandemic: The role of frequent testing, correlated risk, and machine learning.
Technical report, National Bureau of Economic Research, 2020.

[DH06] Dingzhu Du and Frank Hwang.
Pooling designs and nonadaptive group testing: important tools for DNA sequencing.
World Scientific, 2006.

[DHH00] Dingzhu Du, Frank K Hwang, and Frank Hwang.
Combinatorial group testing and its applications, volume 12.
World Scientific, 2000.

[Dor43] Robert Dorfman.
The detection of defective members of large populations.
The Annals of Mathematical Statistics, 14(4):436–440, 1943.

[FLJY20] Linjiajie Fang, Shen Ling, Bing-Yi Jing, and Qing Yang.
Fast: a feasible, accurate and speedy test strategy for covid-19.
medRxiv, 2020.

[GN02] Michelle Girvan and Mark EJ Newman.
Community structure in social and biological networks.
Proceedings of the national academy of sciences, 99(12):7821–7826, 2002.

[KW08] Raghunandan M Kainkaryam and Peter J Woolf.
poolhits: A shifted transversal design based pooling strategy for high-throughput drug screening.
BMC bioinformatics, 9(1):256, 2008.

[NGFD20] Pavlos Nikolopoulos, Tao Guo, Christina Fragouli, and Suhas Diggavi.
Community aware group testing.
arXiv preprint arXiv:2007.08111, 2020.

References II

[SVB+11] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, et al.
High-resolution measurements of face-to-face contact patterns in a primary school.
PloS one, 6(8):e23176, 2011.

[TRL20] Jussi Taipale, Paul Romer, and Sten Linnarsson.
Population-scale testing can suppress the spread of covid-19.
medRxiv, 2020.

Image Credits

Slides 3 and 7: Photo of Paul Romer. The original uploader was Doerrb at English Wikipedia.
CC-BY-SA-2.5,2.0,1. Source: https://commons.wikimedia.org/wiki/File:Paul_Romer_in_2005.jpg
Slide 5: Swab collection. Raimond Spekking. CC BY-SA 4.0. Source: https://citylimits.org/2020/04/13/to-test-or-not-to-test-the-latest-on-new-yorks-covid-19-screening-saga/
Slide 5: PCR machine. Isabella Apriyana. CC-BY-4.0. Source:
https://commons.wikimedia.org/wiki/File:Loading_PCR_mixture_to_thermocycler_machine.jpg
Slide 6: Photo of Robert Dorfman. Author unknown. License unknown. Source:
https://www.hetwebsite.net/het/profiles/dorfman.htm