QuTE: decentralized multiple testing on sensor networks with FDR control

A. Ramdas, J. Chen, M. Wainwright, M. Jordan

Xiang Zhang

Outline

- Motivation & Background
  - FDR control: Distributed VS Centralized
- QuTE: Query-Test-Exchange
  - Methodology
  - Performance guarantee: proof of two theorems
  - Multi-round QuTE
- Simulation
  - Erdos-Renyi random graph
  - Planar grid
- Discussion

Background

- FDR(False Discovery Rate)

$$FDR = \mathbb{E}\left[\frac{V}{\max\{R, 1\}}\right]$$

$$V: \text{\# of fasle discoveries}$$
$$R: \text{\# total discoveries}$$

- BH procedure
  - Order the *p*-values: $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(n)}$
  - Reject the smallest $\widehat{k}$ *p*-values: $\widehat{k} = \max\{k : P_{(k)} \leq \frac{k}{n}\alpha\}$
  - Control at level $FDR \leq \alpha$

Motivation

- Centralized control
  - A central entity with access to all ***N*** *p*-values
  - Perform BH procedure
- Decentralized control
  - Wireless senor networks
  - Low power --> limited communication range
  - Each sensor has access to only a subset of *p*-values
  - Overall FDR control ?

Query-Test-Exchange (QuTE)

- Problem setup
  - Network represented by a graph: $G = (\mathcal{V}, \mathcal{E})$
  - Agent *a* is responsible for testing: $\mathcal{H}_a = \{H_{a,i}\}_{i=1}^{n_a}$
  - True null: $\mathcal{H}_a^0$
  - Total # of hypotheses: $N = \sum_{a\in\mathcal{V}} n_a$
  - Overall true null: $\mathcal{H}^0 = \bigcup_{a\in\mathcal{V}} \mathcal{H}_a^0$
- Communication model
  - Nodes only communicate with their neighbors
  - Perfect *p*-value transmission: no noise & quantization
  - Static topology, *p*-value, etc.

QuTE---Methodology

- Initialization
  - Computes local *p*-vector: $\mathbf{P}_a = (P_{a,1}, P_{a,2}, \cdots, P_{a,n_a})$
- Step 1: **Query**
  - Each agent queries its neighbors for their *p*-vectors
  - Denote $\mathcal{S}_a$ as the set of *p*-values agent *a* gets after query
- Step 2: **Test**
  - Agent *a* runs BH procedure on the $|\mathcal{S}_a|$ *p*-values
  - Target level $\boxed{\alpha_a := \alpha \frac{|\mathcal{S}_a|}{N}}$
  - Reject all *p*-values smaller than $\alpha_a \frac{\widehat{k}^a}{|\mathcal{S}_a|} = (\alpha \frac{|\mathcal{S}_a|}{N}) \frac{\widehat{k}^a}{|\mathcal{S}_a|} = \alpha \frac{\widehat{k}^a}{N}$
- Step 3: **Exchange**
  - Exchange rejection decisions among neighbors

QuTE---Methodology

- **Query**
- **Test**
  - Each agent/node tests all the local p-values & those available at its neighbors (assume that no hypothesis is tested at two or more nodes, i.e., $\mathcal{H}_a$'s are non-overlapping)
  - There might be *computation redundancy (explain later)*
- **Exchange**
  - An agent rejects any local hypothesis rejected by its own test, or by the tests of any of its neighbors
  - Hypothesis rejected by any node will be rejected in the overall FDR control

QuTE---Performance guarantee

Theorem 1 *Suppose that the p-values are independent, or positively dependent. Then for any graph topology, the QuTE algorithm achieves FDR control at level* $\alpha\frac{|\mathcal{H}_0|}{N}$.

- Performing local BH results in overall FDR control
- Each agent only gather information from neighbors
- Two extremes

  >> empty graph: Bonferroni test ($\alpha\frac{|\mathcal{S}_a|}{N} = \frac{\alpha}{N}$)

  >> complete graph: centralized BH

QuTE---Performance guarantee

Theorem 1 *Suppose that the p-values are independent, or positively dependent. Then for any graph topology, the QuTE algorithm achieves FDR control at level* $\alpha\frac{|\mathcal{H}_0|}{N}$.

Proof: denote $\widehat{k}^{qute}$ as the overall # of rejections

define $\widehat{k}^{(a)} = \max_{s\in\text{Neighbors}(a)\cup\{a\}} \widehat{k}^s$

>> maximum # of rejections among *a* and its neighbors

[Claim] $P_{a,j}$ is rejected iff. $P_{a,j} \leq \alpha\frac{\widehat{k}^{(a)}}{N}$ (since is rejected if it is locally rejected, or rejected by any of *a*'s neighbors)

QuTE---Performance guarantee

Theorem 1 *Suppose that the p-values are independent, or positively dependent. Then for any graph topology, the QuTE algorithm achieves FDR control at level* $\alpha \frac{|\mathcal{H}_0|}{N}$.

Proof: [Claim] $P_{a,j}$ is rejected iff. $P_{a,j} \leq \alpha \frac{\widehat{k}^{(a)}}{N}$.

$$FDR = \mathbb{E}\left[\frac{V}{R}\right] = \mathbb{E}\left[\frac{\sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \mathbb{1}\left\{P_{a,j} \leq \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\widehat{k}^{qute}}\right]$$

$$\stackrel{(1)}{=} \sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \mathbb{E}\left[\frac{\mathbb{1}\left\{P_{a,j} \leq \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\widehat{k}^{qute}}\right]$$

$$\stackrel{(2)}{\leq} \sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \mathbb{E}\left[\frac{\mathbb{1}\left\{P_{a,j} \leq \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\widehat{k}^{(a)}}\right] \qquad (2) : \widehat{k}^{(a)} \leq \widehat{k}^{qute}, \forall a \in \mathcal{V}$$

QuTE---Performance guarantee

Proof:

$$
\begin{aligned}
FDR &= \mathbb{E}\left[\frac{V}{R}\right] = \mathbb{E}\left[\frac{\sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \mathbb{1}\left\{P_{a,j} \le \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\widehat{k}^{qute}}\right] \\
&\overset{(1)}{=} \sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \mathbb{E}\left[\frac{\mathbb{1}\left\{P_{a,j} \le \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\widehat{k}^{qute}}\right] \\
&\overset{(2)}{\le} \sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \mathbb{E}\left[\frac{\mathbb{1}\left\{P_{a,j} \le \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\widehat{k}^{(a)}}\right] \\
&\overset{(3)}{=} \sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \frac{\alpha}{N}\mathbb{E}\left[\frac{\mathbb{1}\left\{P_{a,j} \le \alpha\frac{\widehat{k}^{(a)}}{N}\right\}}{\alpha\frac{\widehat{k}^{(a)}}{N}}\right] \\
&\overset{(4)}{\le} \sum_{a\in\mathcal{V}}\sum_{j\in\mathcal{H}_a^0} \frac{\alpha}{N}.1 \\
&= \sum_{a\in\mathcal{V}} \alpha\frac{|\mathcal{H}_a^0|}{N} \\
&\overset{(5)}{=} \alpha\frac{|\mathcal{H}^0|}{N}
\end{aligned}
$$

[Corollary]

Under independence positive dependence, this item is no greater than 1. [ref 1,2]

$$(5) : \sum_{a\in\mathcal{V}} |\mathcal{H}_a^0| = |\mathcal{H}^0|$$

IMAGINE THE UNIVERSITY OF UTAH

QuTE---Multiple Rounds

- Run the QuTE algorithm ***c*** (>1) times
  - Each node/agent can query *p*-vectors from all the other nodes within distance ***c***.
  - Equivalent to enlarging the communication range
- Performance guarantee

Theorem 2 *Suppose that the p-values are independent, or positively dependent. Then the multi-step QuTE algorithm with c>1 rounds of communication guarantees that* $FDR \leq \alpha \frac{|\mathcal{H}^0|}{N}$.

>> Proof is essentially the same as Theorem 1

Simulation

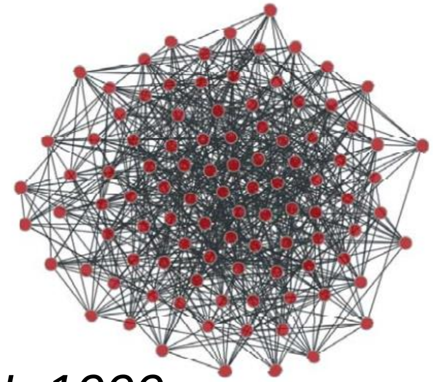
- Generation of *p*-values

$$\begin{aligned} X &\sim \mu + \mathcal{N}(0,1) \\ P &= 1 - \Phi(X) \end{aligned}$$

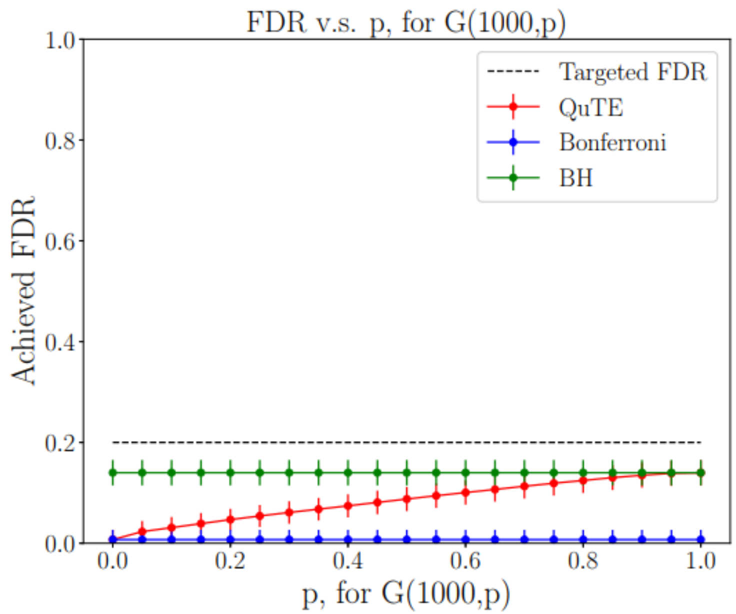
- $\mu = 0$ for nulls and $\mu > 0 (\mu = 2)$ for alternatives
- Target FDR: $\alpha = 0.2$
- Null-proportion: $\pi_0 = \frac{|\mathcal{H}^0|}{N} = 0.7$
- Erdos-Renyi random graph G(*N, p*)
  - Any pair of vertices are connected w.p. *p*, in total *N* vertices
- Planar grid
  - Rectangle and square grid

Simulation

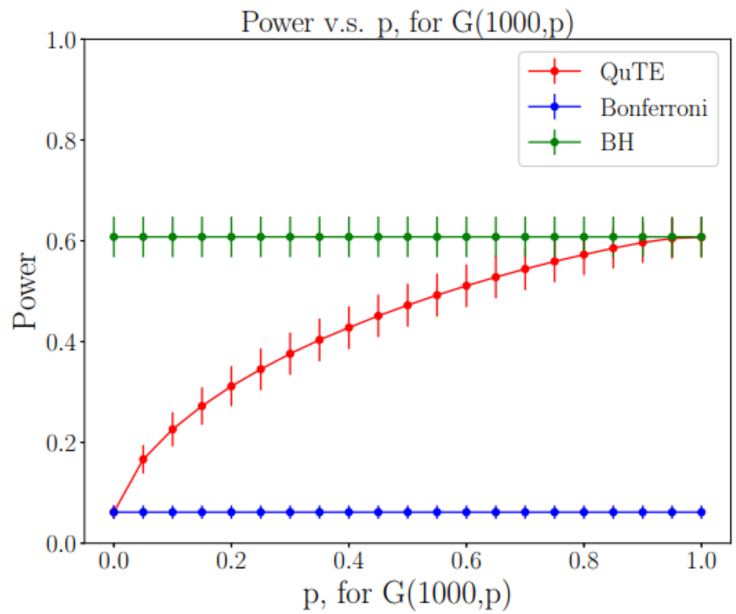
- Erdos-Renyi random graph G(*N, p*)
  - Any pair of vertices are connected w.p. *p*, in total *N=1000* nodes, each testing one hypothesis



(a)



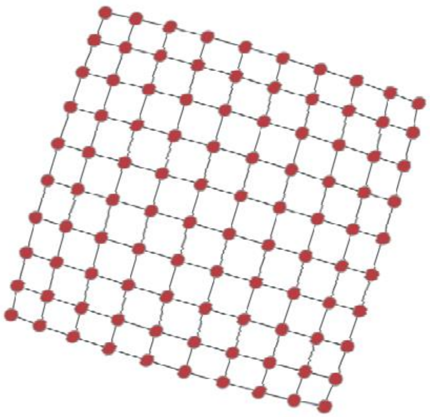
(b)



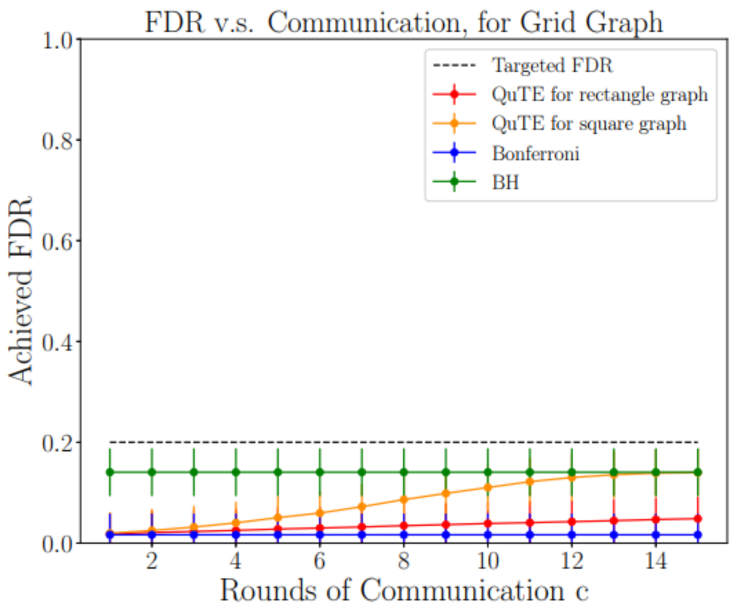
(c)

Simulation

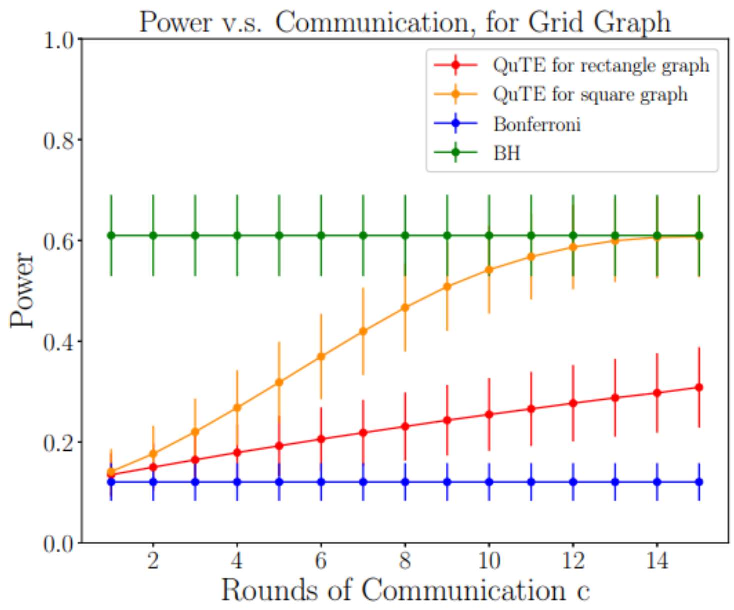
- Planar grid
  - *N*=256 nodes, each testing one hypothesis
  - Square grid 16*16
  - Rectangular grid 2*128



(a)



(b)



(c)

IMAGINE THE UNIVERSITY OF UTAH

Discussion

- Quantization error of *p*-values

  >> Sensors with limited computation ability--> accuracy

  >> Wireless channels are noisy

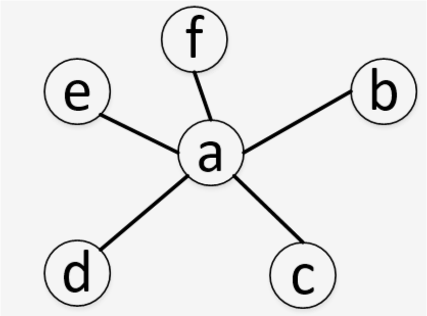
  >> What is the effect of quantization?

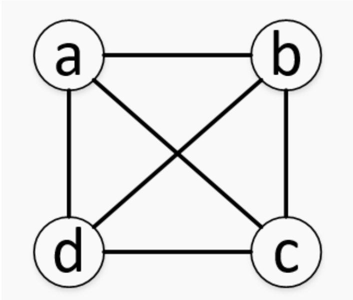
- Computation redundancy

  >> As long as there exists one single node that are connected to all the other nodes, QuTE becomes centralized BH procedure

  >> For a complete graph, all nodes perform the same BH multiple times--> unnecessary power consumption

  >> Sampling





References

[1] Barber, Rina Foygel, and Aaditya Ramdas. "The p-filter: multilayer false discovery rate control for grouped hypotheses." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.4 (2017): 1247-1268.

[2] Blanchard, Gilles, and Etienne Roquain. "Two simple sufficient conditions for FDR control." *Electronic journal of Statistics* 2 (2008): 963-992.

Thank you

- Questions?