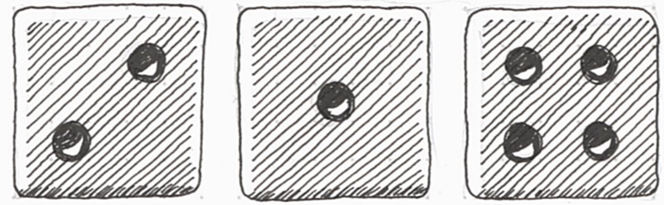
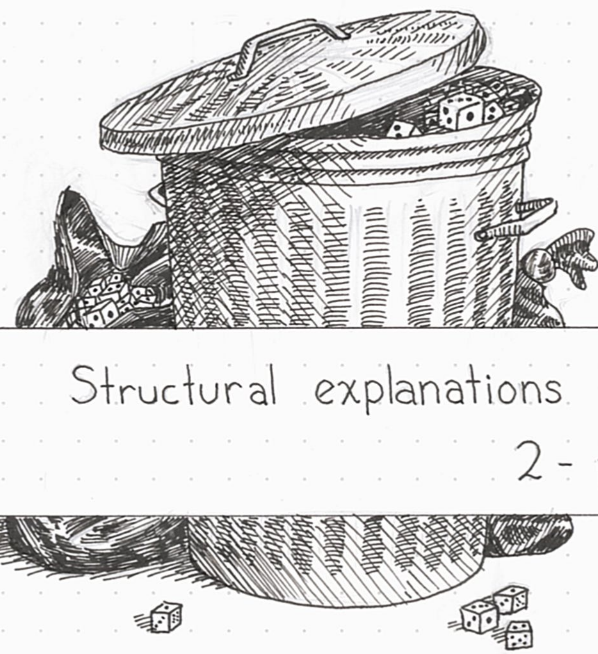
Understanding

RANDOMNESS



Without

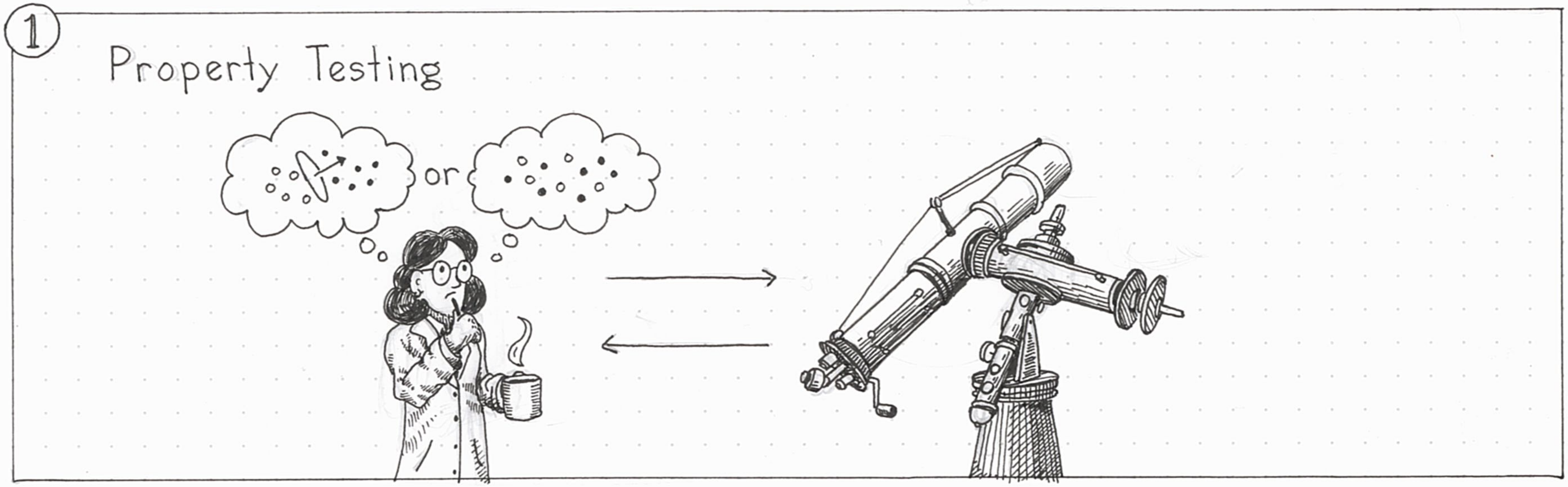
RANDOMNESS

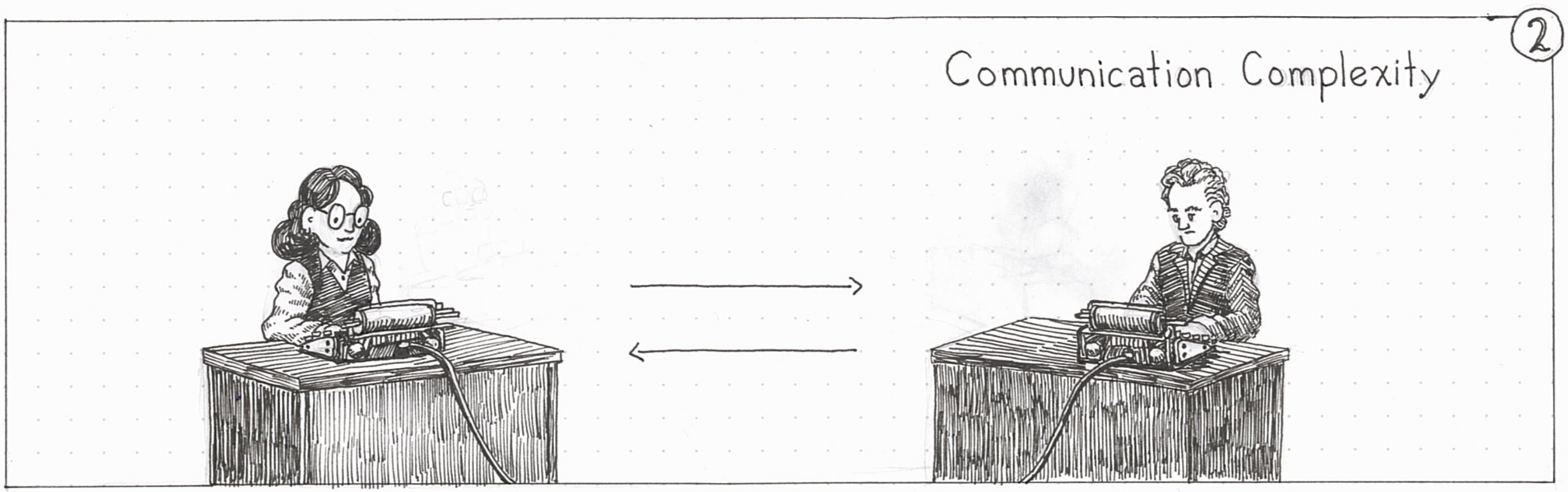


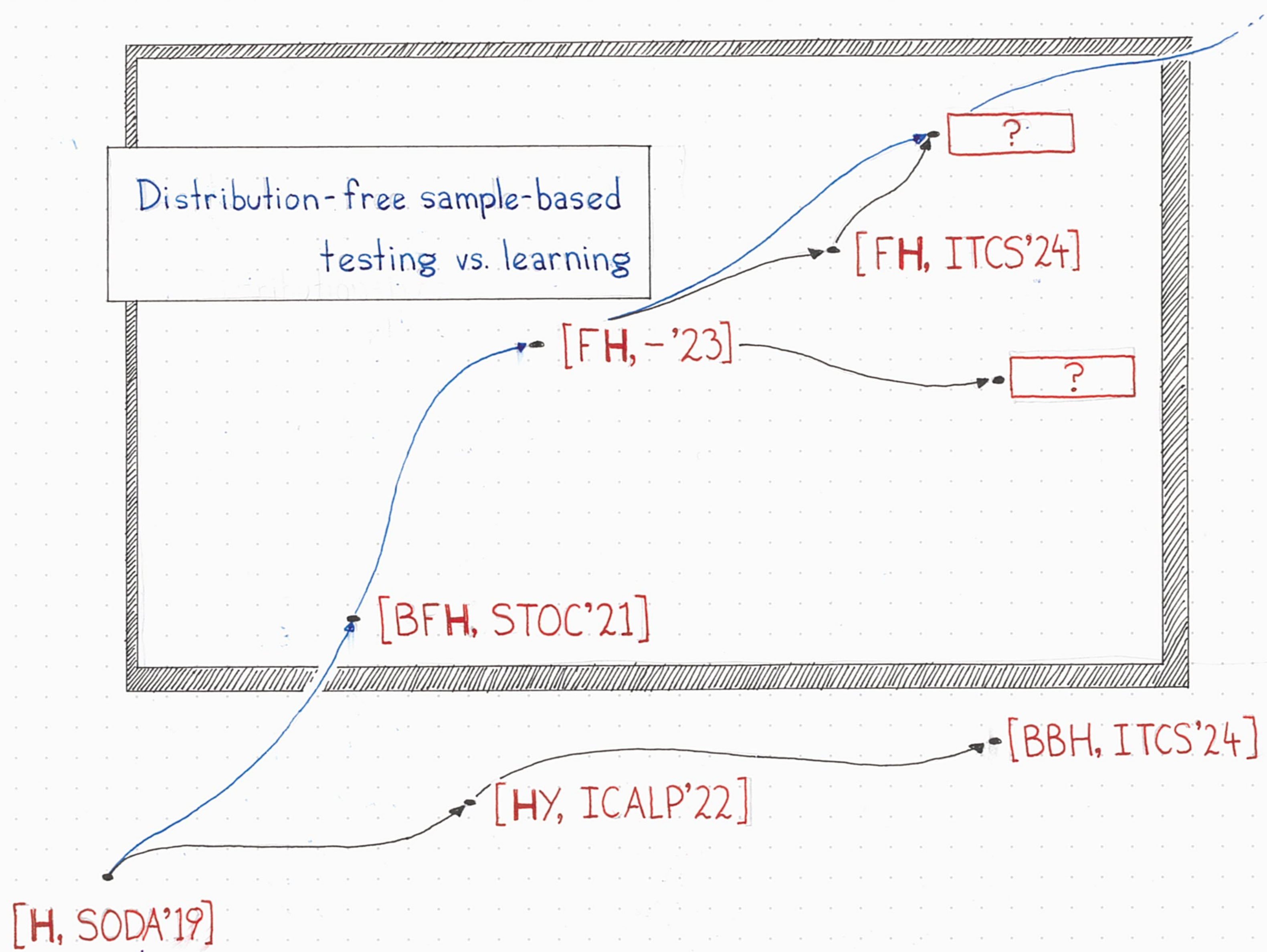
Structural explanations of the power of randomized 2-party protocols

Nathan Harms (EPFL)

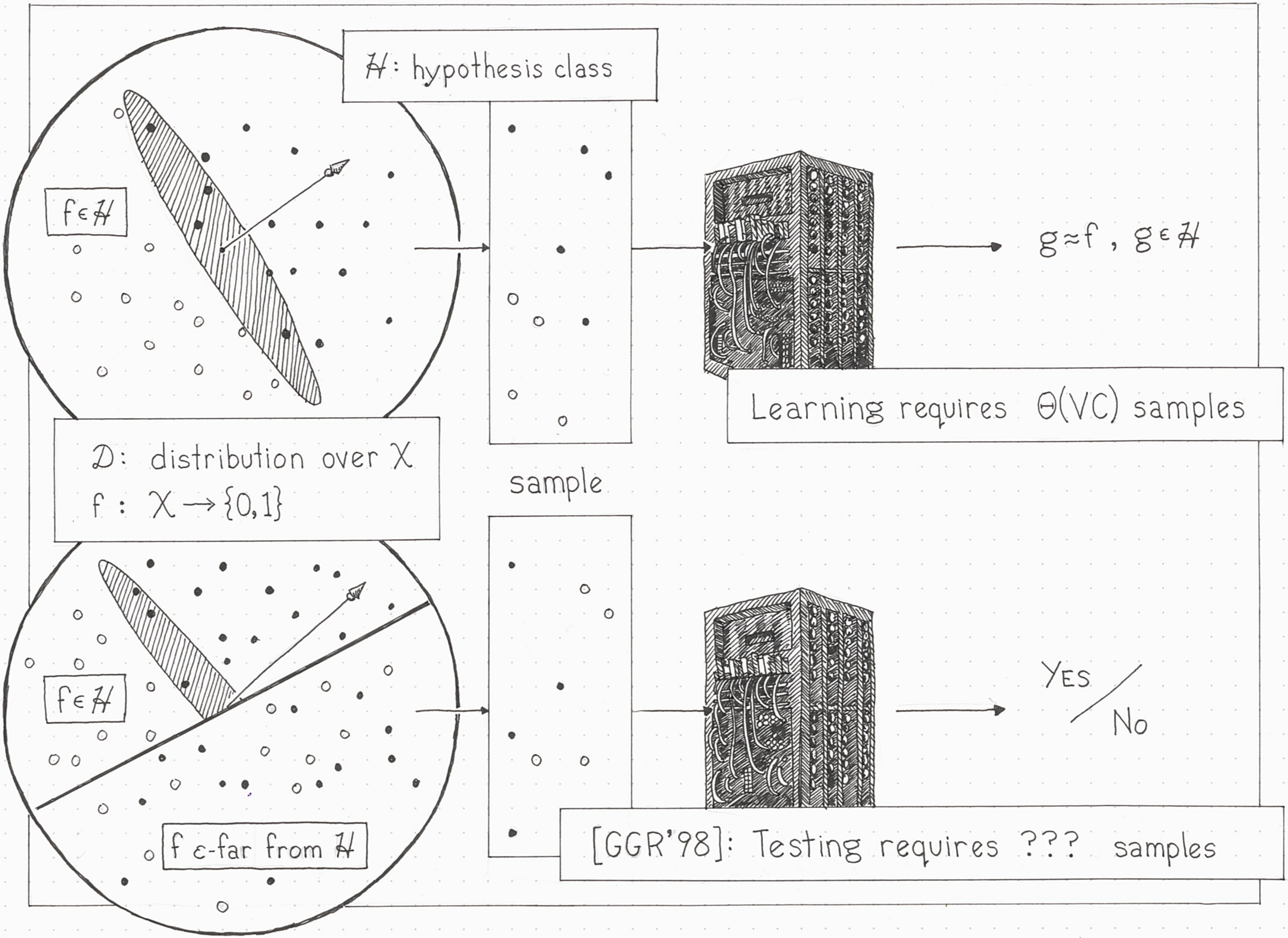
Two-Party Decision Problems





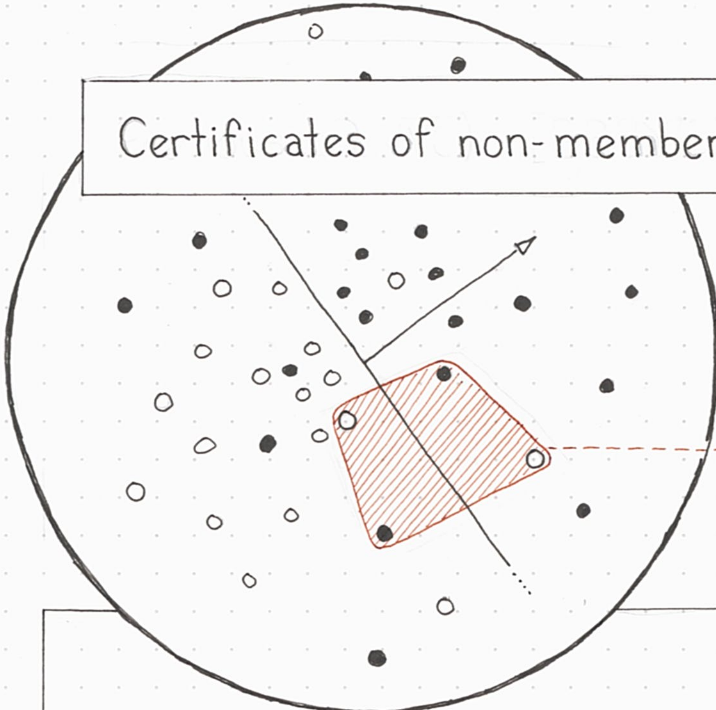


Distribution-free sample-based testing vs. learning
[FH, ITCS'24]
?
[FH, –'23]
?
[BFH, STOC'21]
[BBH, ITCS'24]
[HY, ICALP'22]
[H, SODA'19]



$\mathcal{H}$: hypothesis class
$f \in \mathcal{H}$
$\mathcal{D}$: distribution over $\mathcal{X}$
$f: \mathcal{X} \to \{0,1\}$
sample
$g \approx f$, $g \in \mathcal{H}$
Learning requires $\Theta(VC)$ samples
$f \in \mathcal{H}$
f ε-far from $\mathcal{H}$
YES / NO
[GGR'98]: Testing requires ??? samples

Certificates of non-membership



<u>LVC</u>: minimum certificate size

"Large certificates": LVC ≈ VC

↳ halfspaces, intersections etc.

Prior: $\Omega(\sqrt{VC})$ for classes with LVC ≈ VC [GR'16]

<u>Theorem</u>: $\Omega(VC/\log VC)$ for classes with LVC ≈ VC [BFH'21]

Tight for "<u>symmetric classes</u>"!

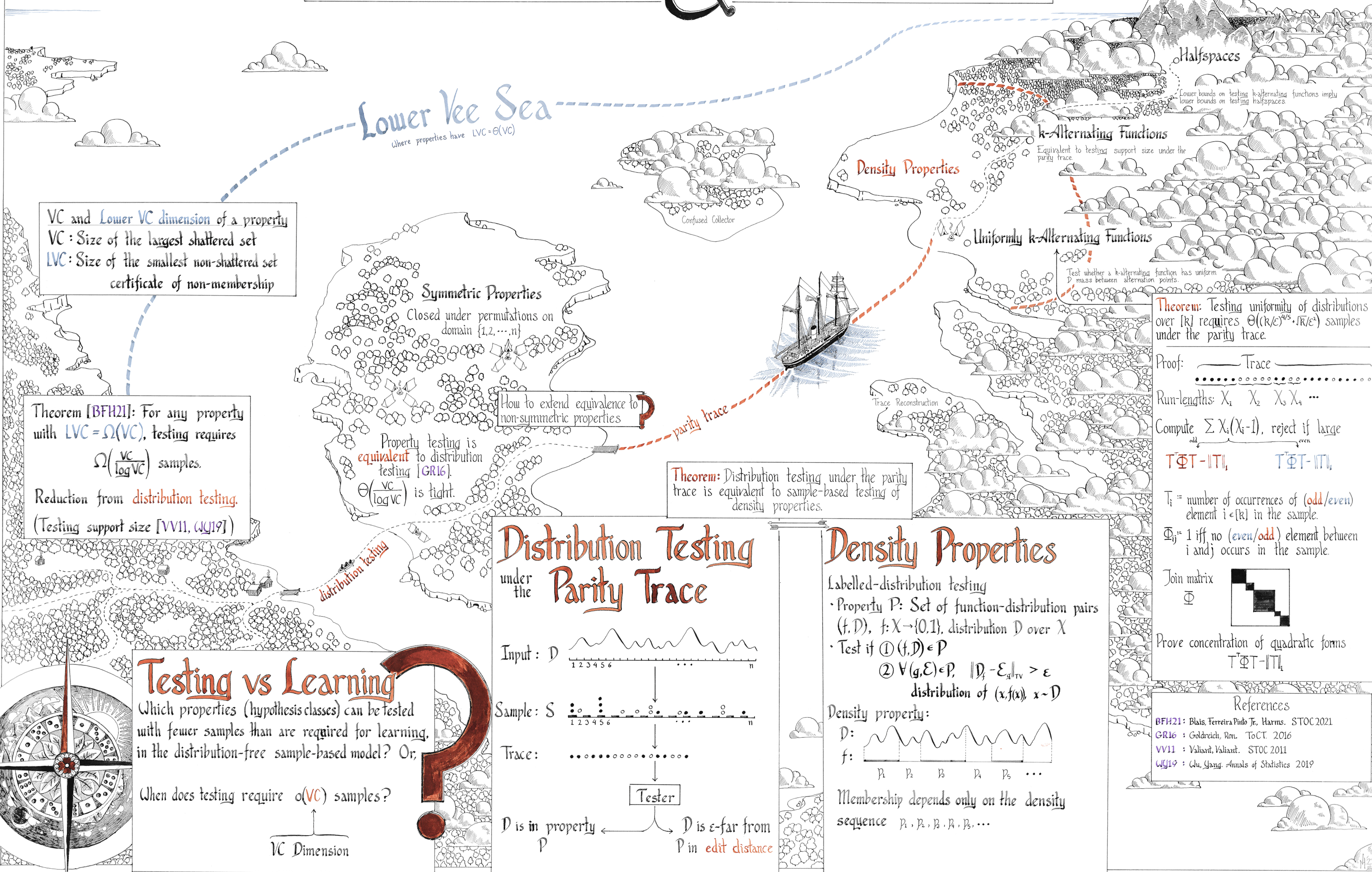
[GR'16] ↓ ↑ [BFH'21]

Distribution testing

Deeper connection? ⟶

Renato Ferreira Pinto Jr., University of Waterloo
Nathan Harms, EPFL

Distribution Testing & Testing vs Learning



VC and Lower VC dimension of a property

VC : Size of the largest shattered set

LVC : Size of the smallest non-shattered set certificate of non-membership

Theorem [BFH21]: For any property with $LVC = \Omega(VC)$, testing requires $\Omega\left(\frac{VC}{\log VC}\right)$ samples.

Reduction from distribution testing.

(Testing support size [VV11, WY19])

Theorem: Distribution testing under the parity trace is equivalent to sample-based testing of density properties.

Testing vs Learning

Which properties (hypothesis classes) can be tested with fewer samples than are required for learning, in the distribution-free sample-based model? Or,

When does testing require $o(VC)$ samples?

VC Dimension

Distribution Testing under the Parity Trace

Input : D

1 2 3 4 5 6 … n

Sample : S

1 2 3 4 5 6 … n

Trace :

Tester

D is in property P

D is ε-far from P in edit distance

Density Properties

Labelled-distribution testing

- Property P: Set of function-distribution pairs (f, D), $f: X \to \{0,1\}$, distribution D over X
- Test if ① $(f, D) \in P$
  ② $\forall (g, \mathcal{E}) \in P$, $\|D_f - \mathcal{E}_g\|_{TV} > \varepsilon$
  D_f: distribution of $(x, f(x))$, $x \sim D$

Density property:

D:

f:

p_1 p_2 p_3 p_4 p_5 …

Membership depends only on the density sequence $p_1, p_2, p_3, p_4, p_5, \dots$

Theorem: Testing uniformity of distributions over $[k]$ requires $\Theta((k/\varepsilon)^{4/5} + \sqrt{k}/\varepsilon^2)$ samples under the parity trace.

Proof: Trace

Run-lengths: X_1 X_2 X_3 X_4 …

Compute $\sum X_i(X_i - 1)$, reject if large

odd: $T^T \Phi T - \|T\|_1$ even: $T^T \Phi T - \|T\|_1$

T_i := number of occurrences of (odd/even) element $i \in [k]$ in the sample.

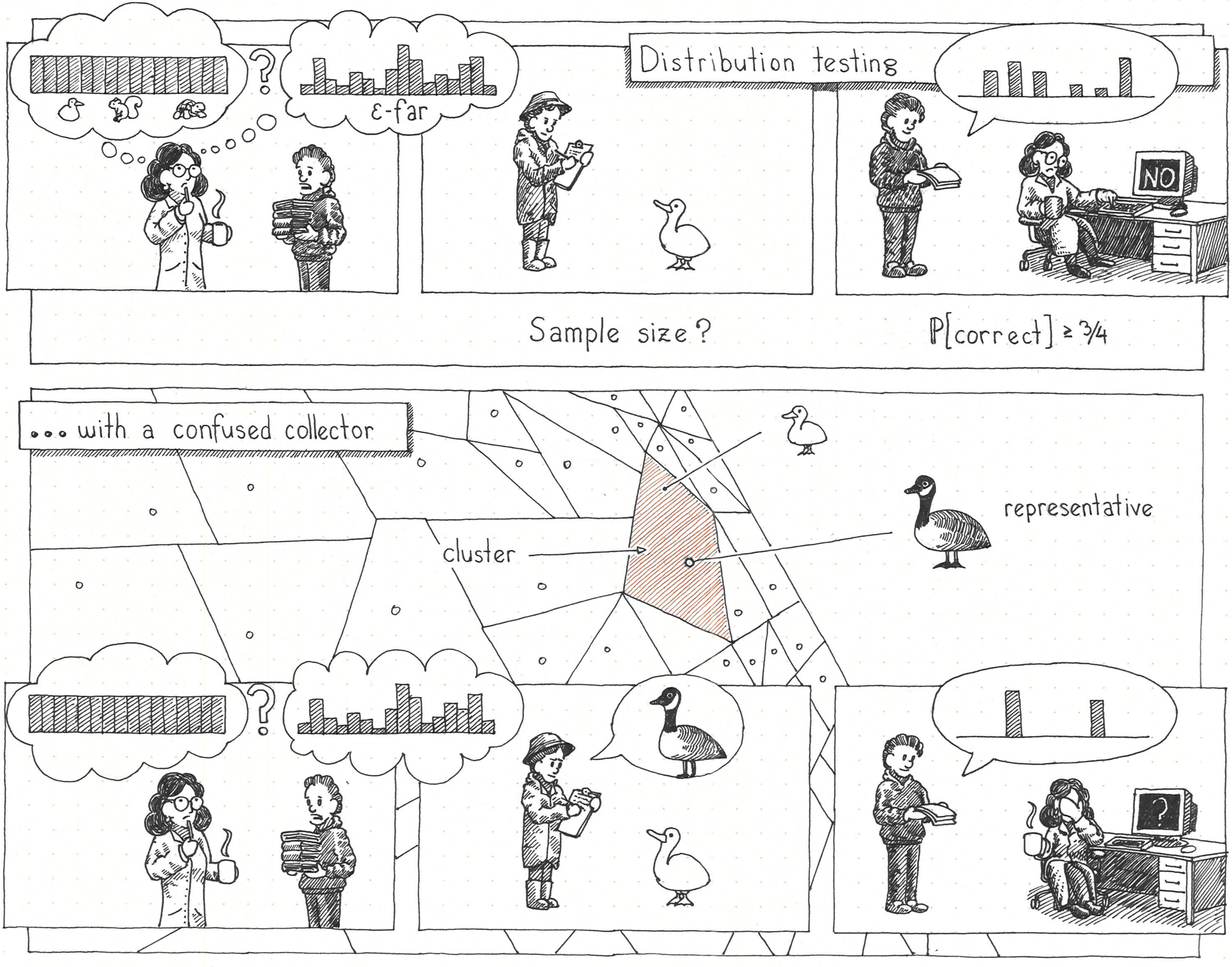
Φ_{ij} := 1 iff no (even/odd) element between i and j occurs in the sample.

Join matrix Φ

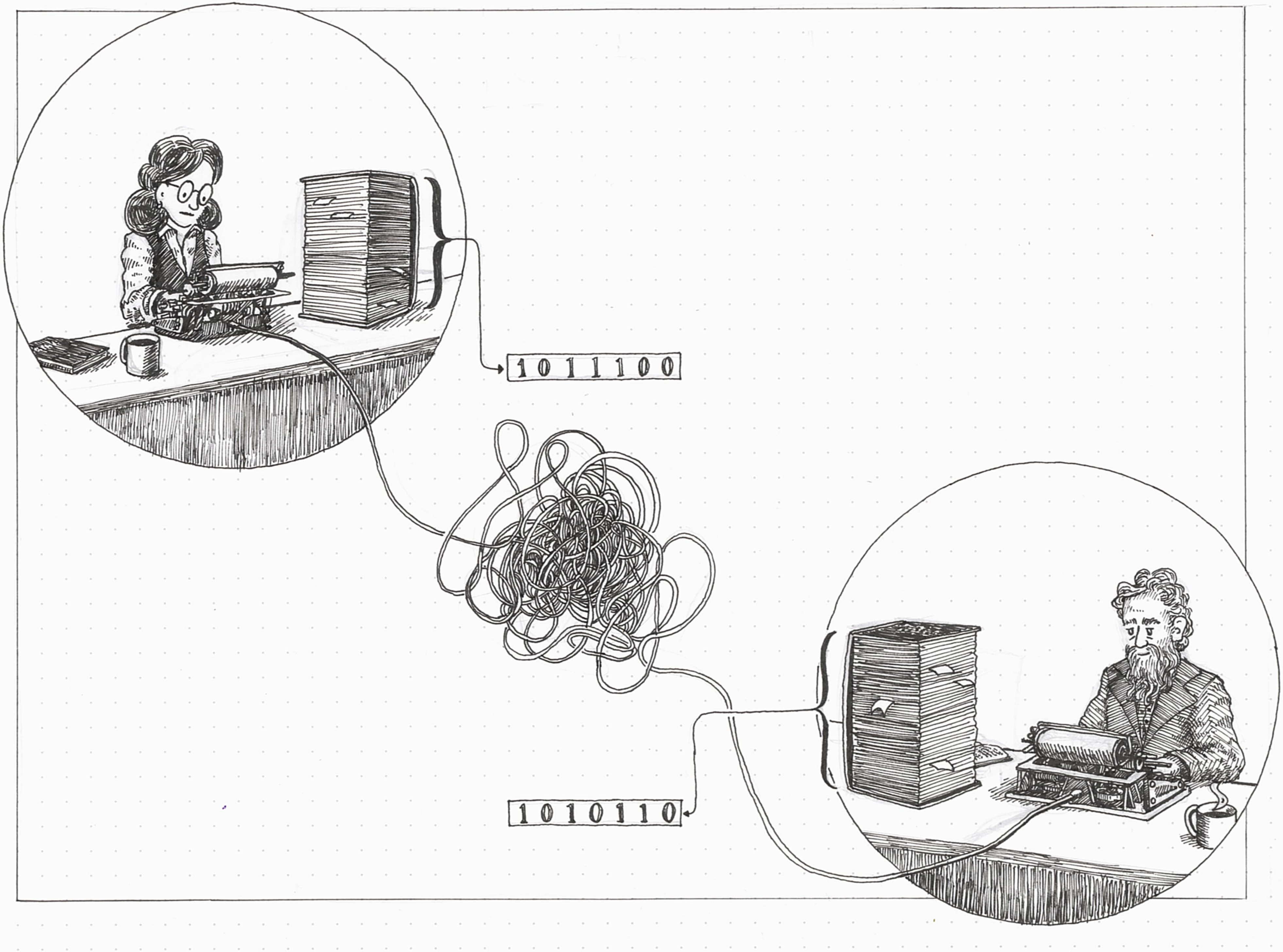
Prove concentration of quadratic forms $T^T \Phi T - \|T\|_1$

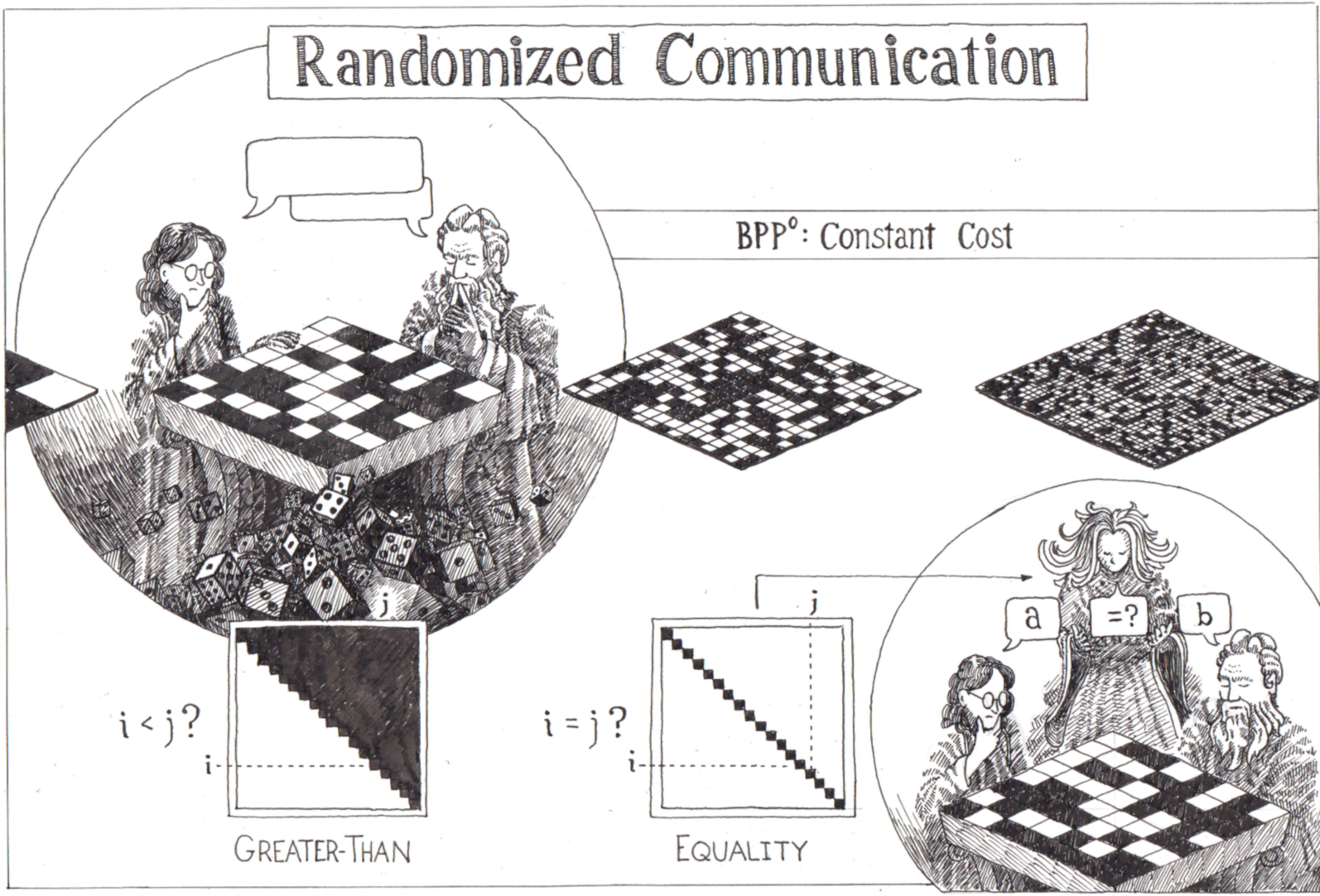
References

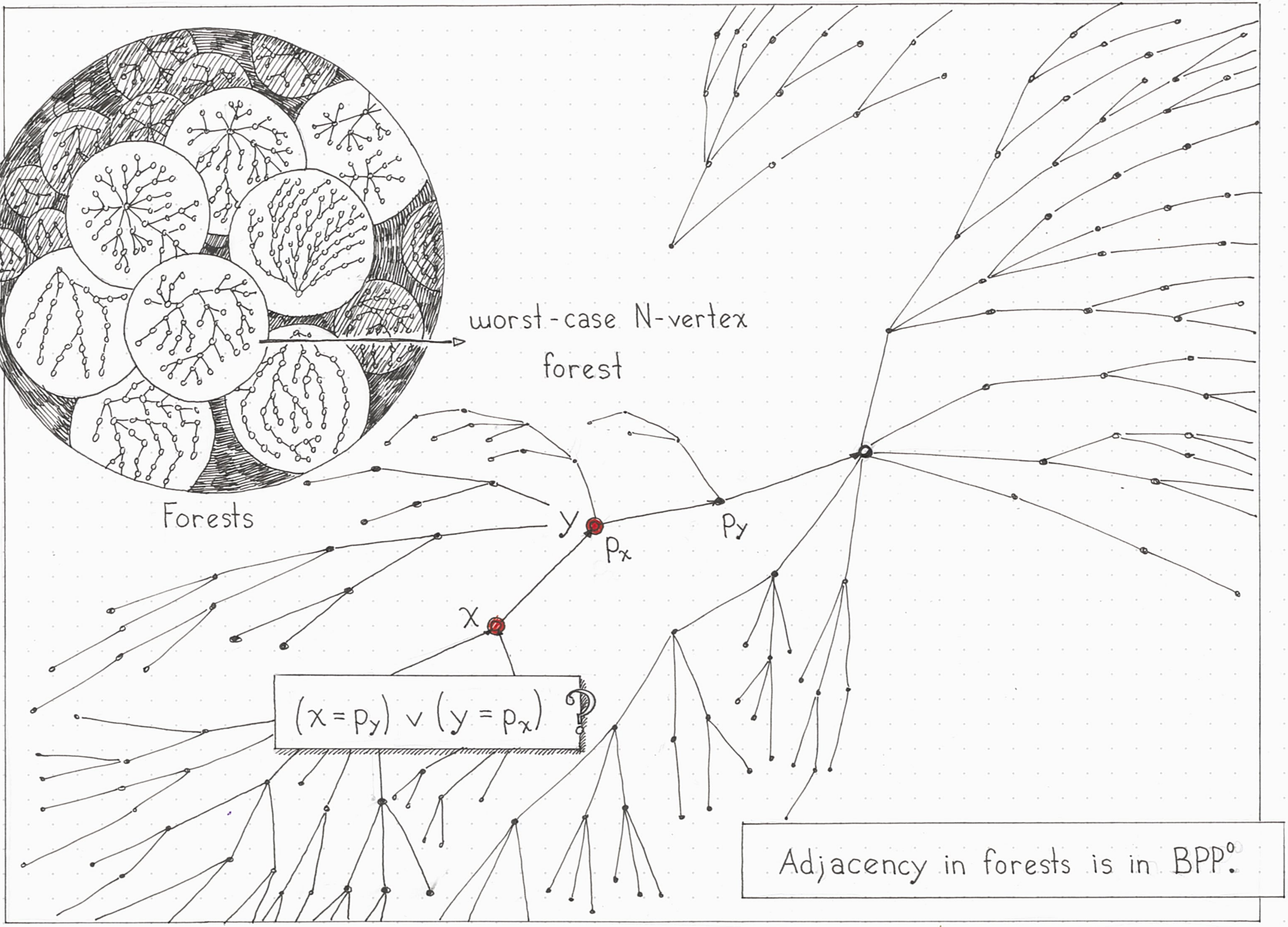
BFH21 : Blais, Ferreira Pinto Jr., Harms. STOC 2021
GR16 : Goldreich, Ron. ToCT. 2016
VV11 : Valiant, Valiant. STOC 2011
WY19 : Wu, Yang. Annals of Statistics 2019



?
ε-far
Distribution testing
NO
Sample size?
ℙ[correct] ≥ 3/4
... with a confused collector
cluster
representative
?
?


1011100
1010110


Randomized Communication
BPP°: Constant Cost
j
i < j ?
i
GREATER-THAN
j
i = j ?
i
EQUALITY
a
=?
b


worst-case N-vertex forest
Forests
y
p_x
p_y
x
$(x = p_y) \vee (y = p_x)$?
Adjacency in forests is in BPP°.

Why?

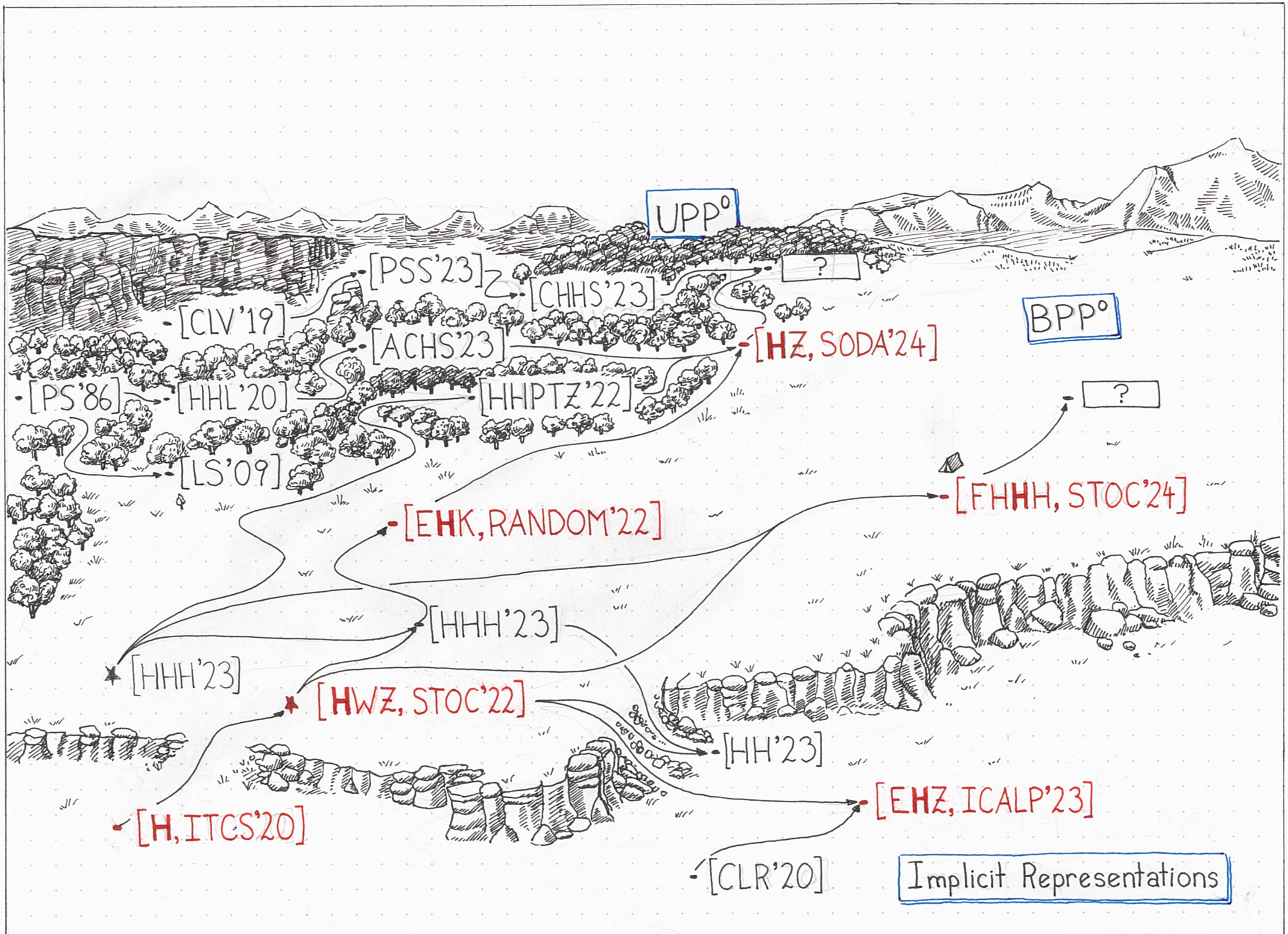
① Power of randomness

- Most extreme case
- Most basic lower bound
- "Fine-grained" understanding
- Standard techniques fail
- Complete characterization?

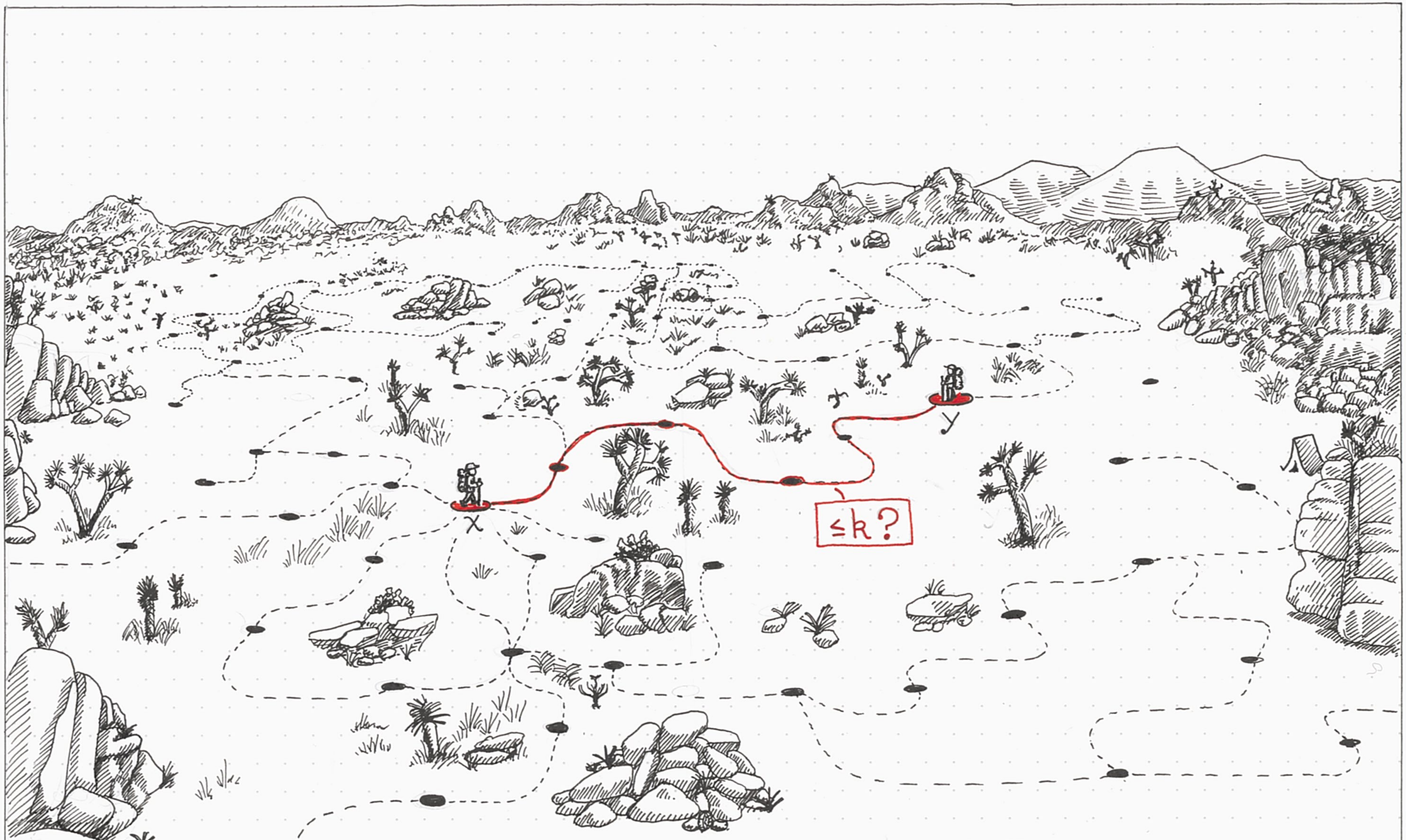
② Connections

- Matrix representations
- Structural graph theory
  - → New concepts
  - → New techniques
- Algebra
- Learning theory

③ It is cool



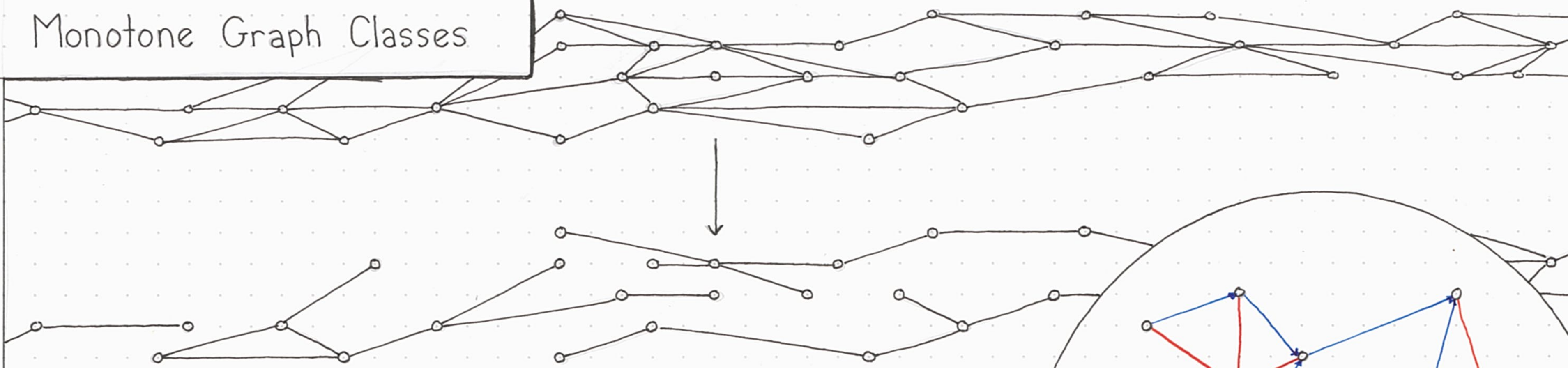
UPP°
BPP°
?
?
[PSS'23]
[CHHS'23]
[CLV'19]
[ACHS'23]
[HZ, SODA'24]
[PS'86]
[HHL'20]
[HHPTZ'22]
[LS'09]
[EHK, RANDOM'22]
[FHHH, STOC'24]
[HHH'23]
[HHH'23]
[HWZ, STOC'22]
[HH'23]
[EHZ, ICALP'23]
[H, ITCS'20]
[CLR'20]
Implicit Representations

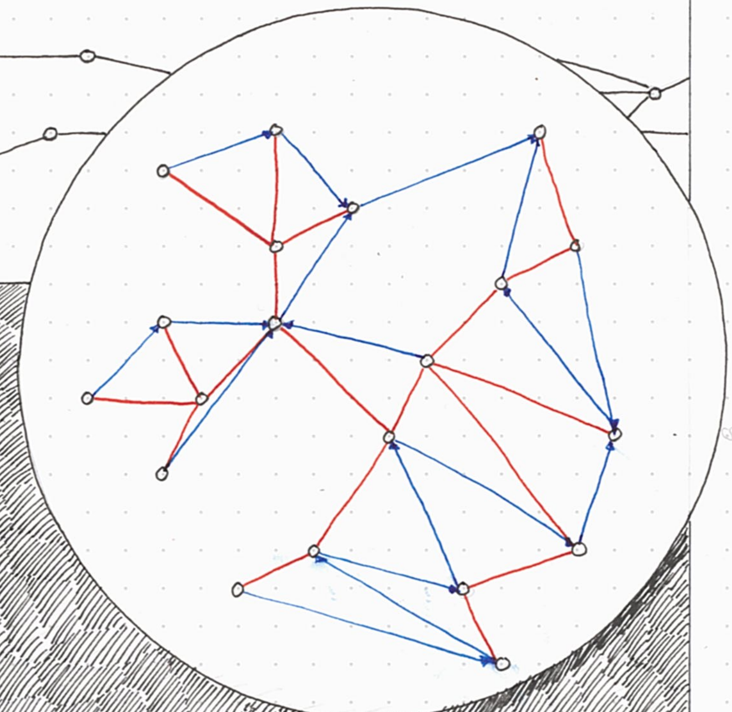


Theorem: $\forall$constants k, deciding $\mathrm{dist}(x,y) \leq k$ in planar graphs is in BPP^0. [H'20, HWZ'22]

$$\phi(x,y) = \exists\, p_0, \ldots, p_k \in V : x = p_0 \wedge y = p_k \wedge \big(p_0 = p_1 \vee E(p_0, p_1)\big) \wedge \cdots \wedge \big(p_{k-1} = p_k \vee E(p_{k-1}, p_k)\big)$$

Monotone Graph Classes

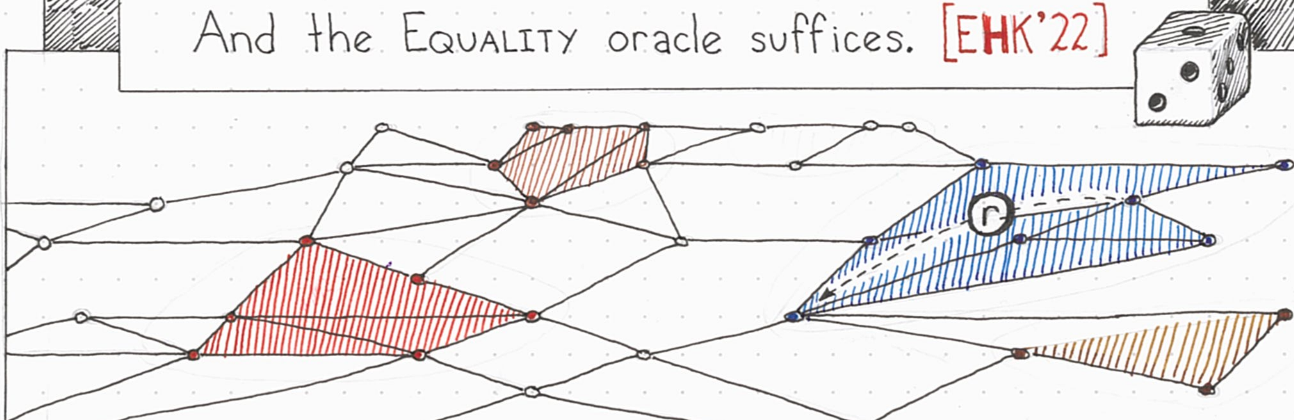




<u>Theorem</u>: If $\mathcal{G}$ is a monotone graph class:

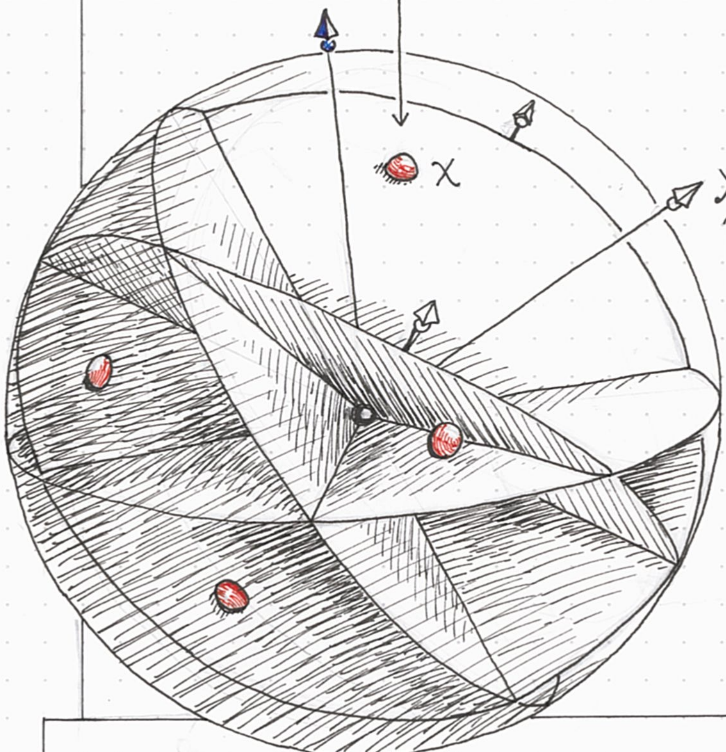
1. Adjacency is in BPP^0
   $\Leftrightarrow$ $\mathcal{G}$ has bounded arboricity.
2. Distance k is in BPP^0 $\forall$ constants k
   $\Leftrightarrow$ $\mathcal{G}$ has bounded expansion.

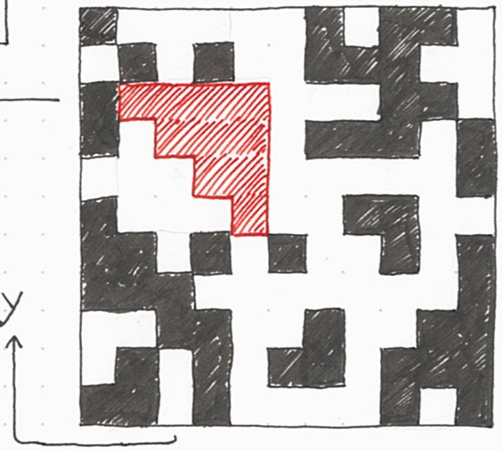
And the Equality oracle suffices. [EHK'22]

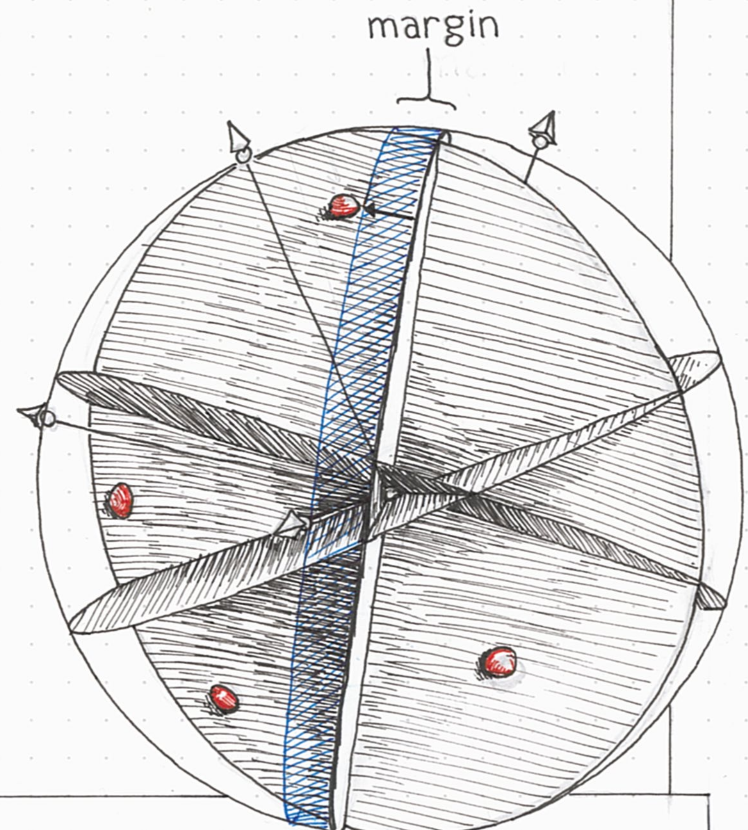


contraction has average degree $f(r)$. [e.g. NO'12]

Matrix Representations







$\mathrm{UPP}^0 \cap \mathrm{BPP}^0$?

Constant dimension:

- Equivalent to UPP^0 [PS'86]
- Can have randomized cost $\Theta(\log N)$ in dimension 3 [HHL'22, ACHS'23]

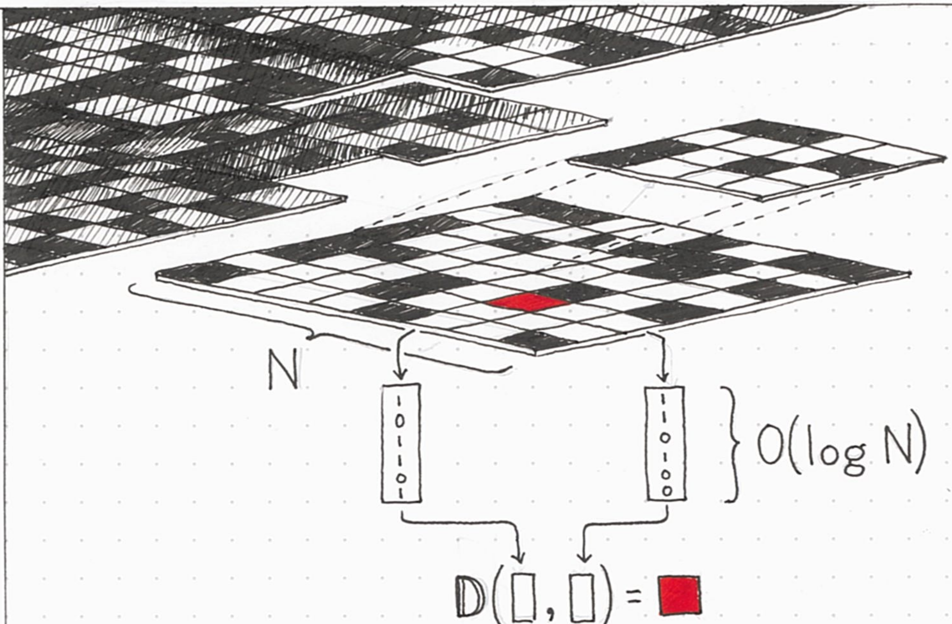
Stable: largest GREATER-THAN has constant size

Constant margin:

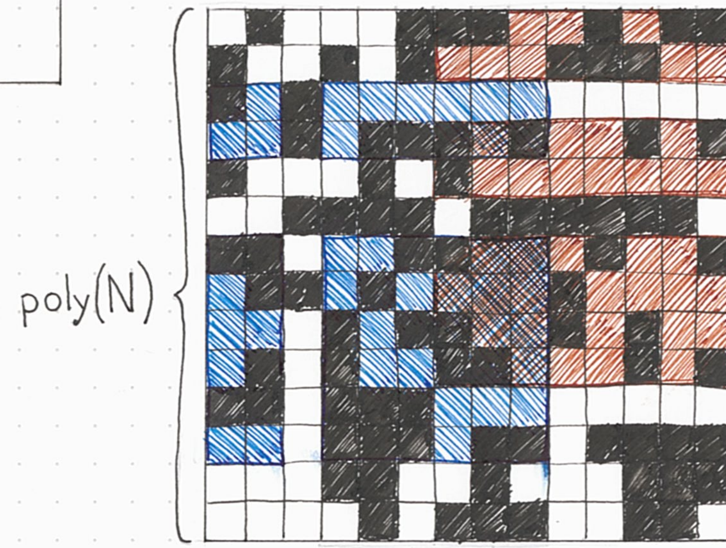
- Equivalent to BPP^0 [LS'09]
- Stability is necessary

Theorem: For dimension 3, stability is sufficient for BPP^0. [HZ'24]

Implicit Representations (Adjacency Labels)

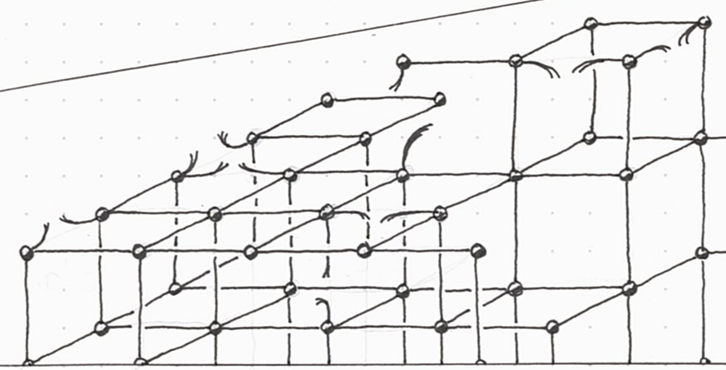


≡



Universal graph

<u>Open</u>: Representations for UPP^0
[S'03, F'19, A'22]

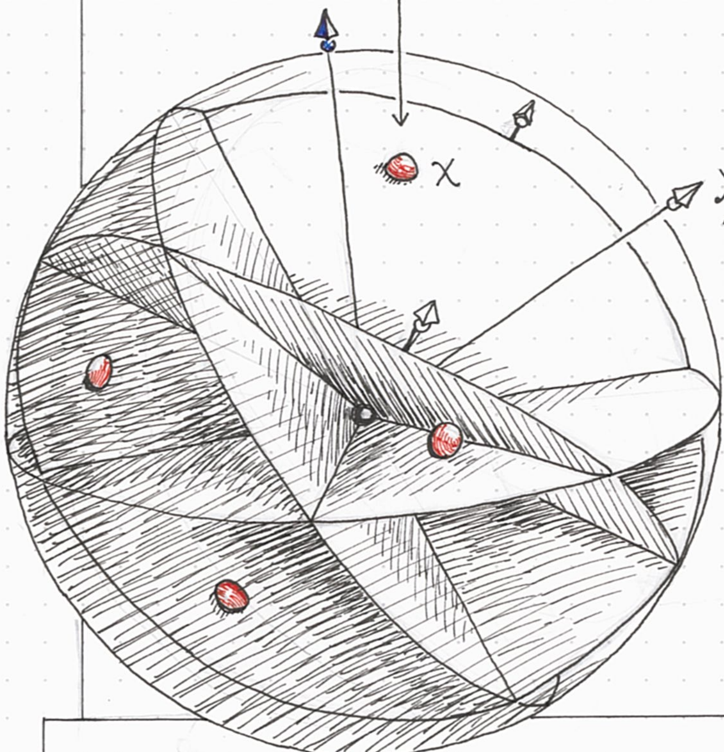


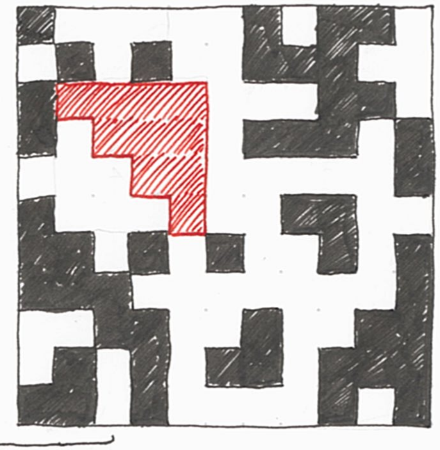
BPP^0 ≡ Adjacency "sketch" ⇒ Implicit representation
[H'20, HWZ'22]

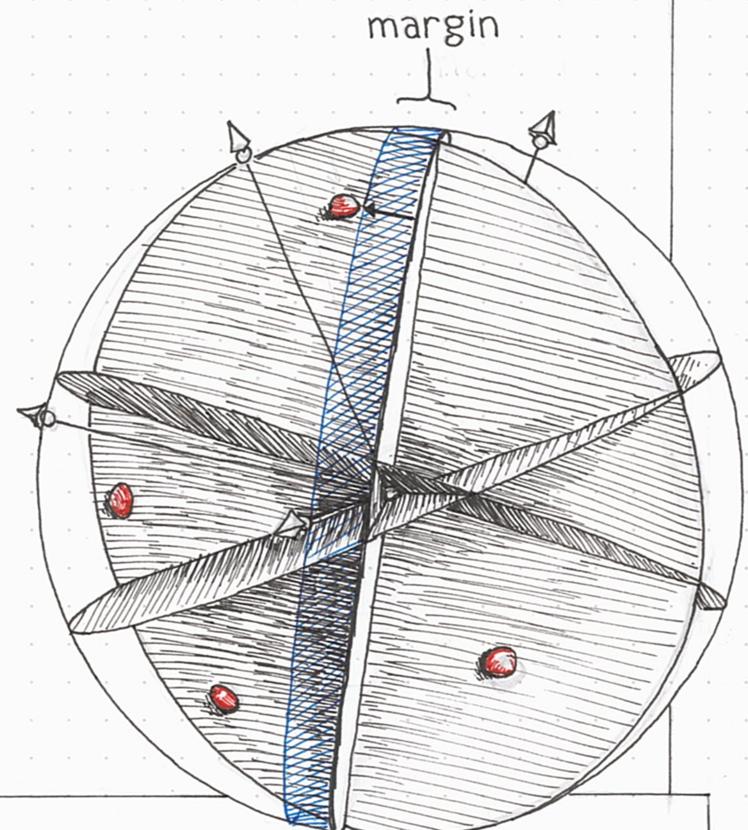
↳ + additive combinatorics, hashing ⟶

<u>Theorem</u>: Optimal adjacency labels for subgraphs of Cartesian products.
[EHZ'23] [CLR'20, AAL'21]

Matrix Representations







$UPP^0 \cap BPP^0$?

Constant dimension:

- Equivalent to UPP^0 [PS'86]
- Can have randomized cost $\Theta(\log N)$ in dimension 3 [HHL'22, ACHS'23]

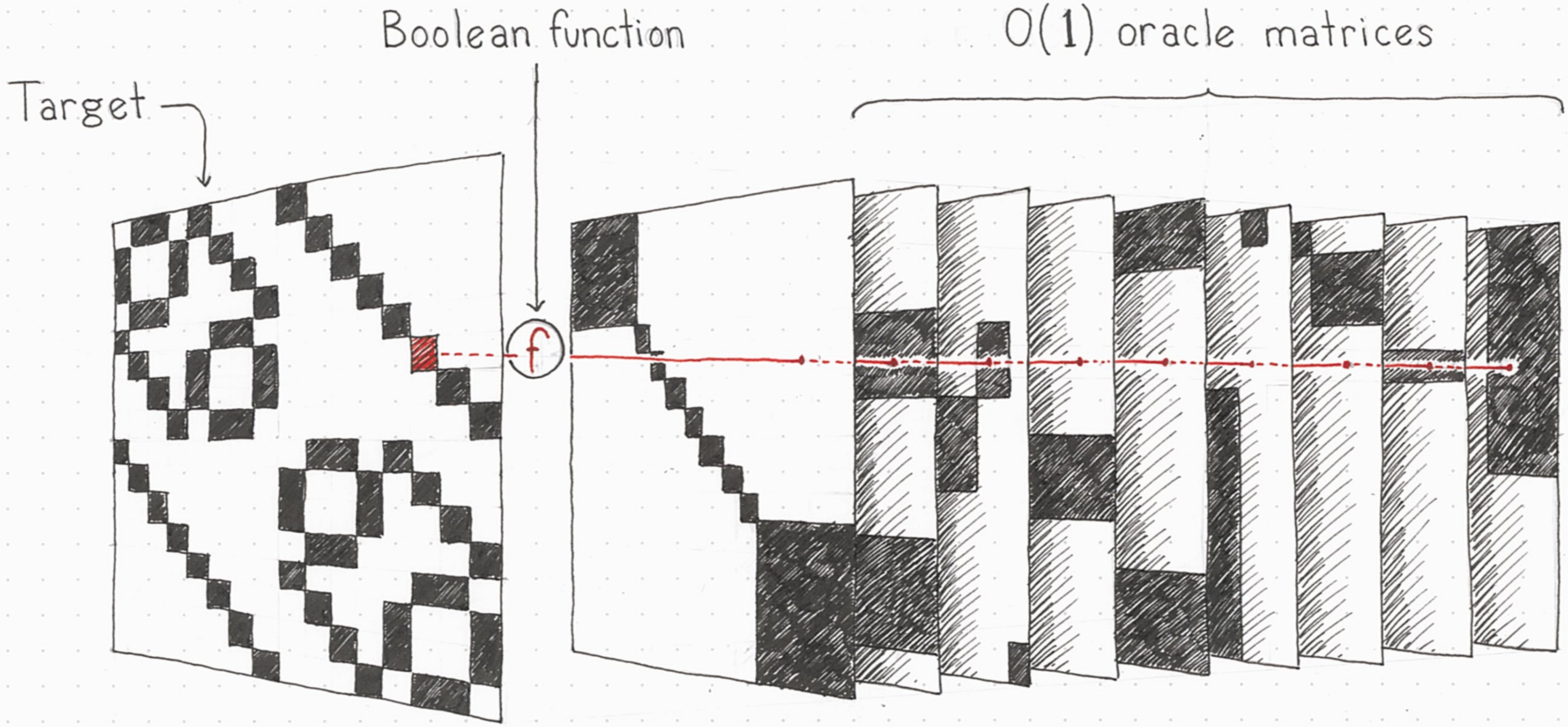
Stable: largest GREATER-THAN has constant size

Constant margin:

- Equivalent to BPP^0 [LS'09]
- Stability is necessary

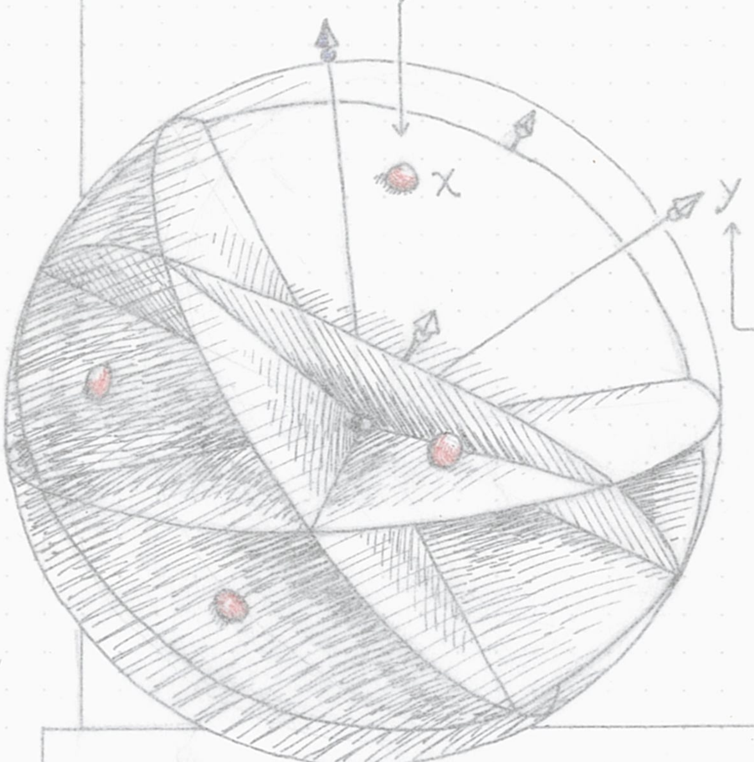
Theorem: For dimension 3, stability is sufficient for BPP^0. [HZ'24]

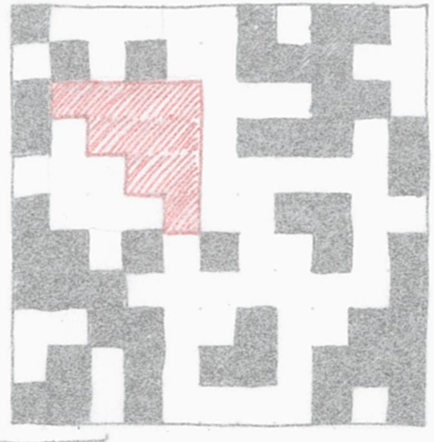
Is Equality "complete" for BPP^0?

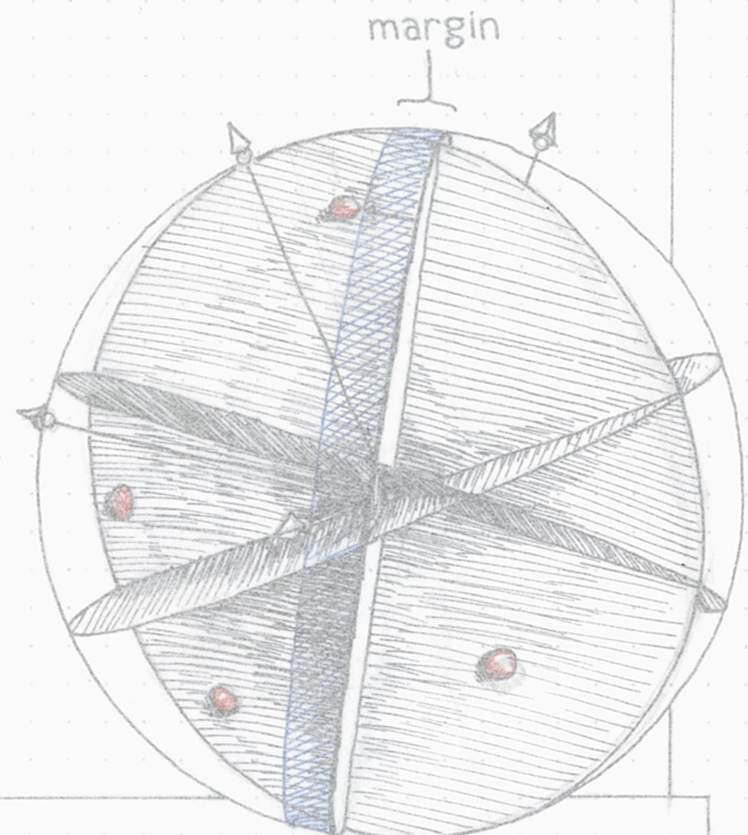


<u>Theorem</u>: 1-Hamming Distance does not reduce to Equality. [HWZ'22, HHH'23]

Matrix Representations







$\mathsf{UPP}^0 \cap \mathsf{BPP}^0 = (\mathsf{P}^{\mathsf{EQ}})^0$?

Constant dimension:

- Equivalent to UPP^0 [PS'86]
- Can have randomized cost $\Theta(\log N)$ in dimension 3 [HHL'22, ACHS'23]

Stable: largest GREATER-THAN has constant size

Constant margin:

- Equivalent to BPP^0 [LS'09]
- Stability is necessary

Theorem: For dimension 3, stability is sufficient for BPP^0. [HZ'24]

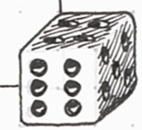
... Then, is 1-Hamming Distance complete for BPP^0?

All previous techniques fail ...
[HWZ'22, HZ'24, HHH'23, CLV'19, CHHS'23, PSS'23]

But... <u>Theorem</u>: 2-Hamming Distance does not reduce to 1-Hamming Distance [FHHH'24]

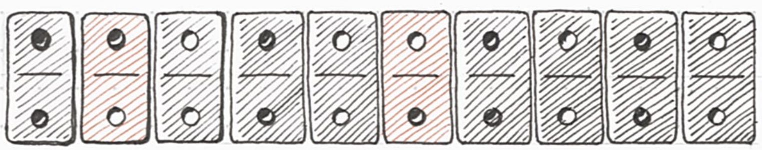
<u>Theorem</u>: There is no complete problem for BPP$^{\mathrm{o}}$.

<u>Theorem</u>: The k-Hamming Distance problems form an infinite hierarchy within BPP$^{\mathrm{o}}$. [FHHH'24]



① Every oracle in BPP$^{\mathrm{o}}$ is <u>stable</u>. →

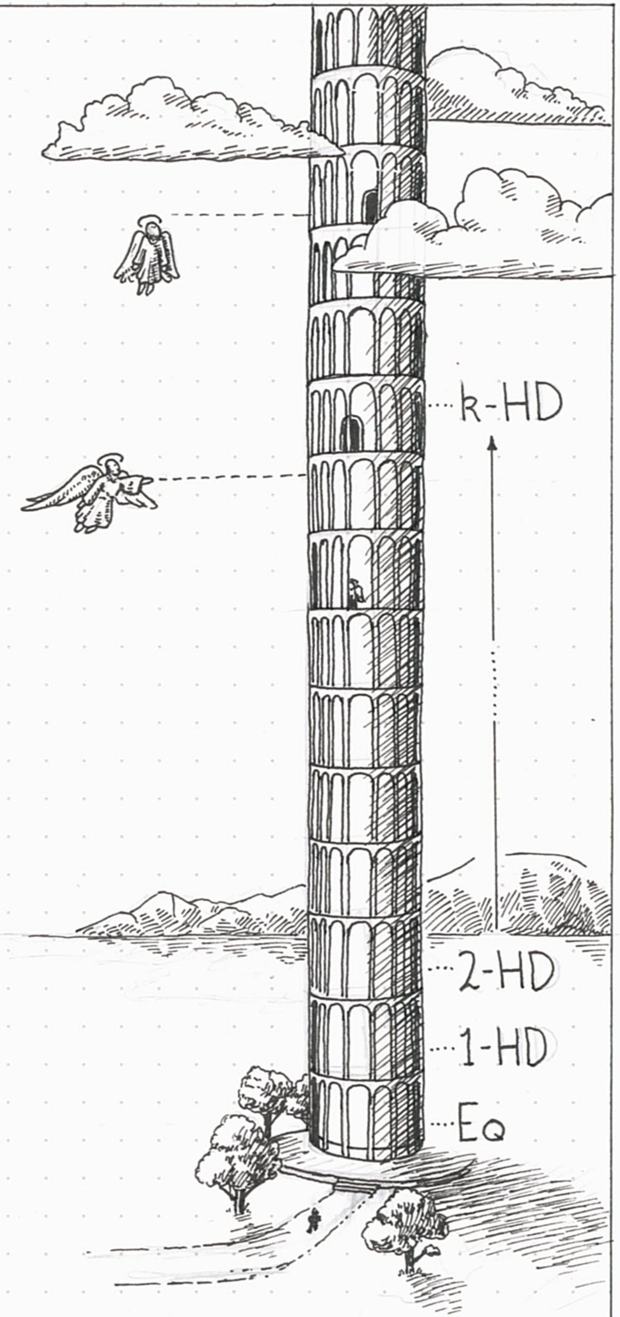
② k-Hamming Distance is permutation-invariant:



If the oracle queries are <u>not</u> permutation-invariant:

→ Hypergraph Ramsey theorem ⇒ stability is violated.

③ <u>One query</u> computes distance k vs. $k+2$ for all weight $n/2$ strings. Stability is violated.



End