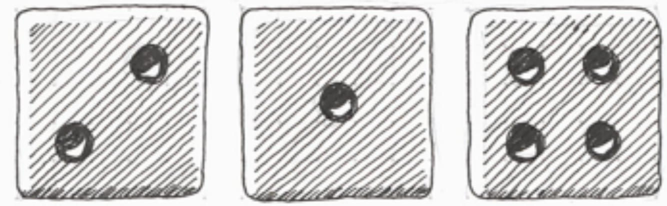
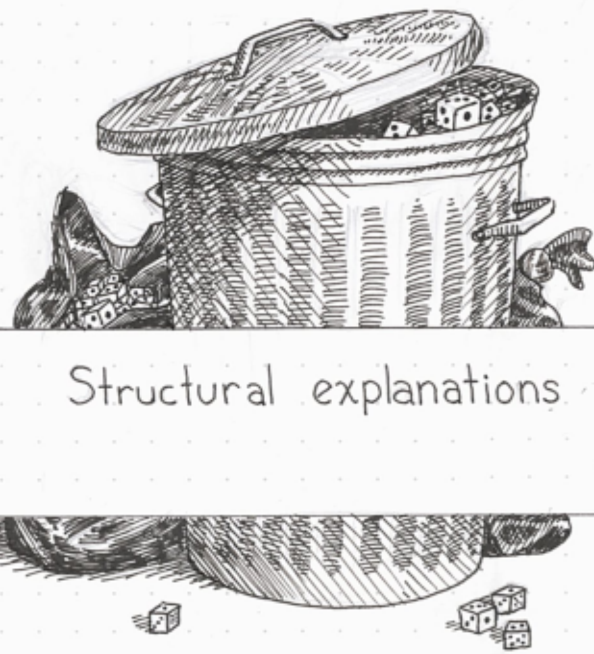
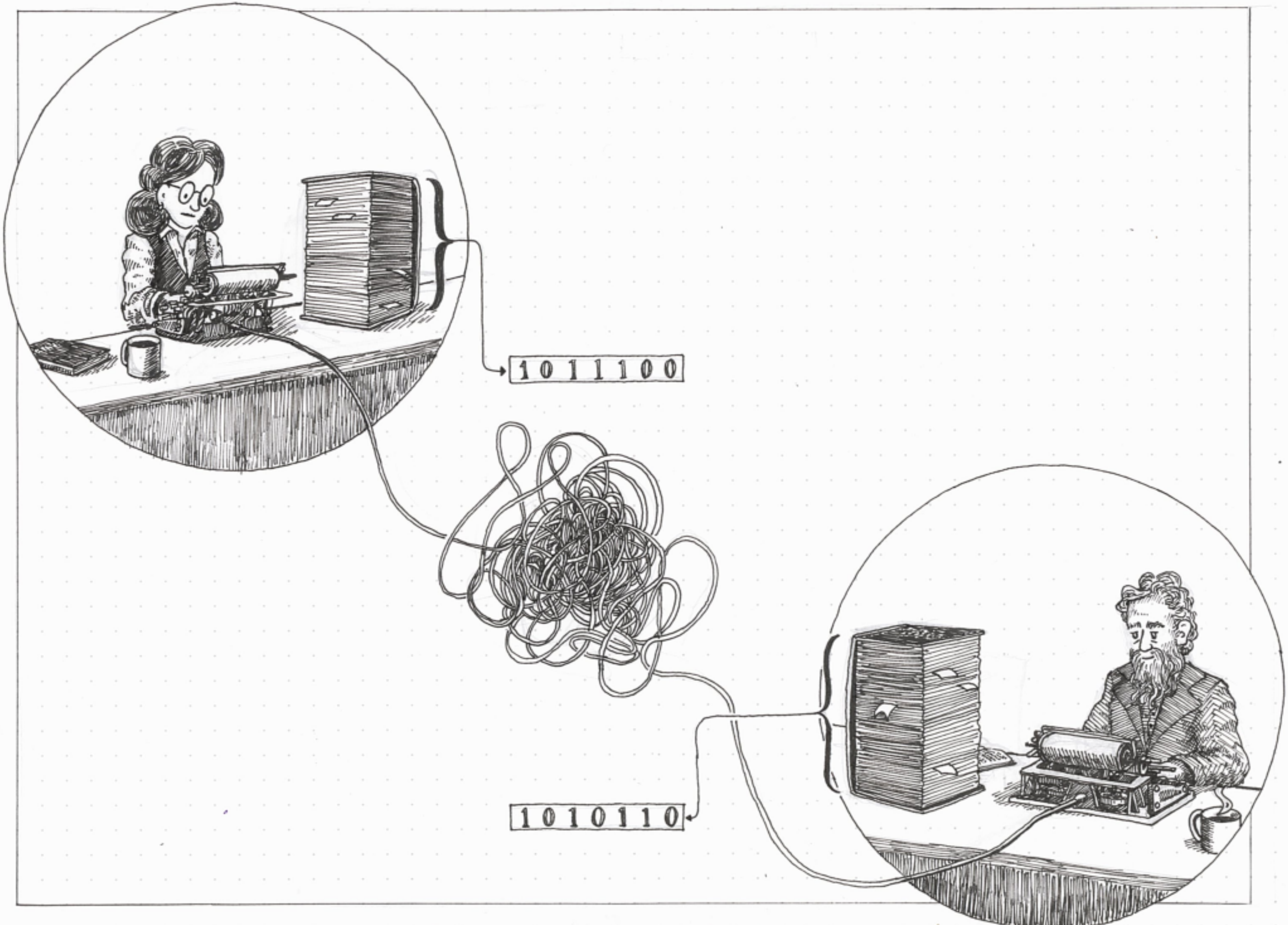
Understanding RANDOMNESS  Without RANDOMNESS

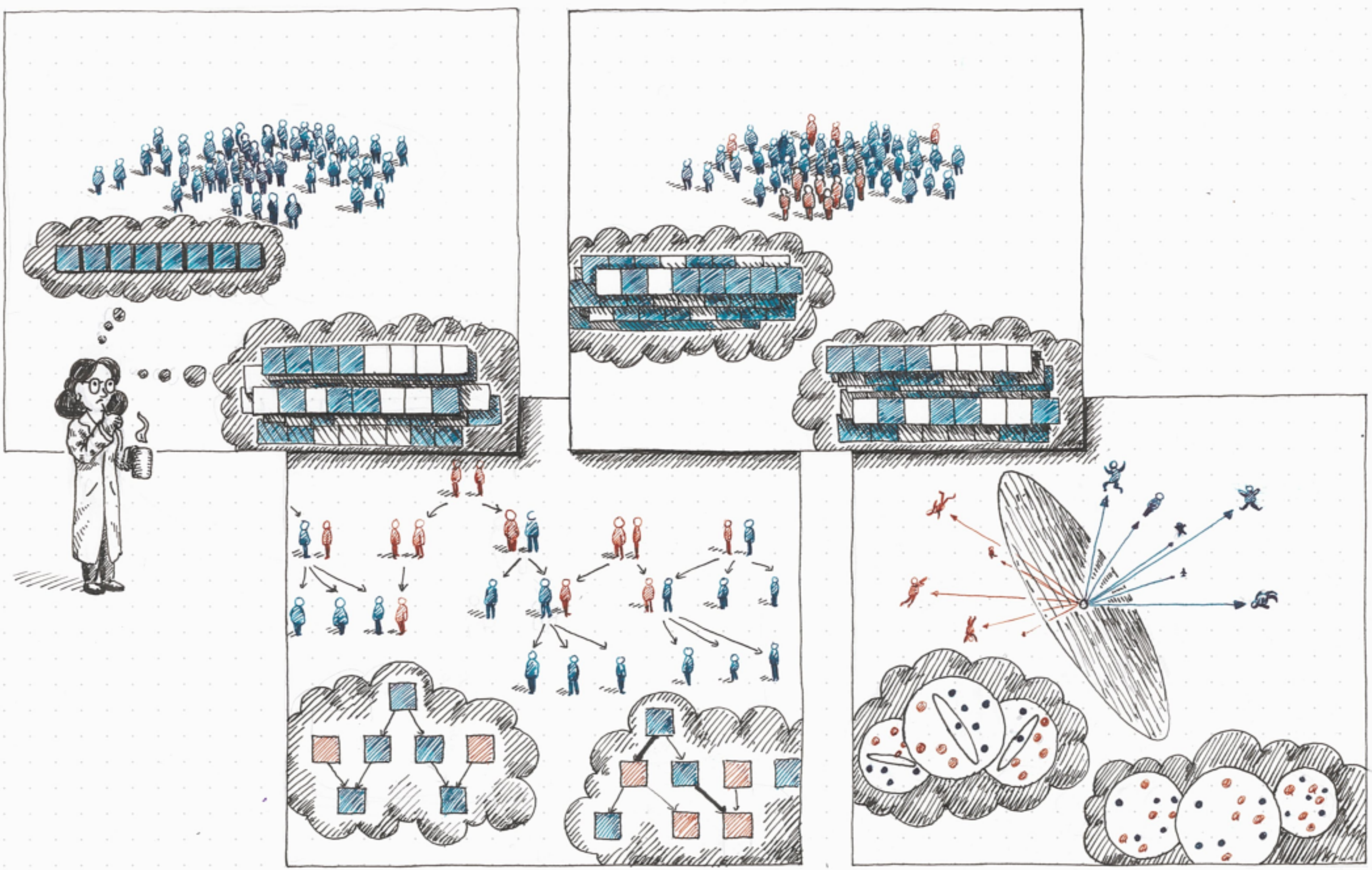


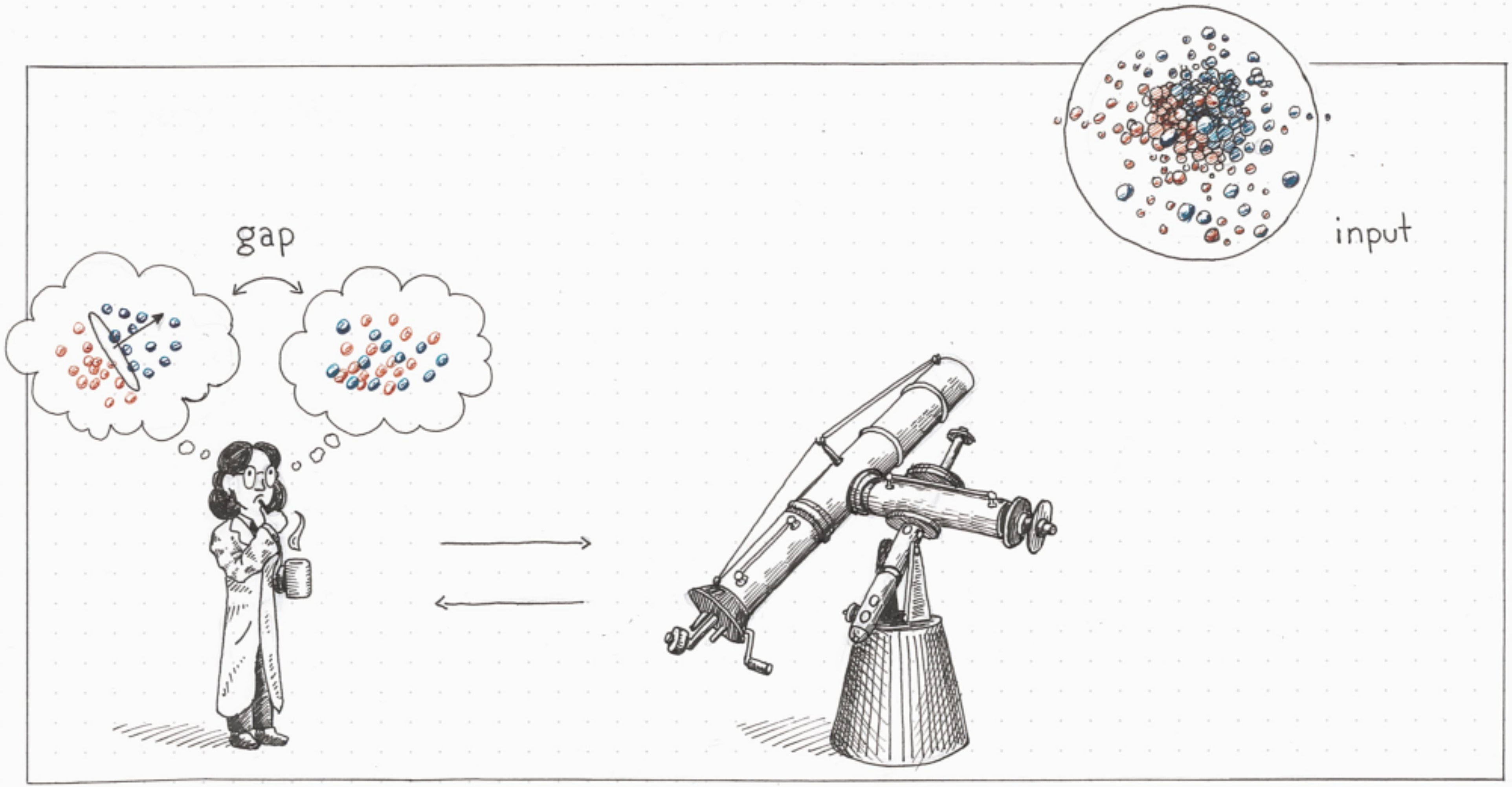
Structural explanations of the power of randomized algorithms

Nathan Harms (EPFL)

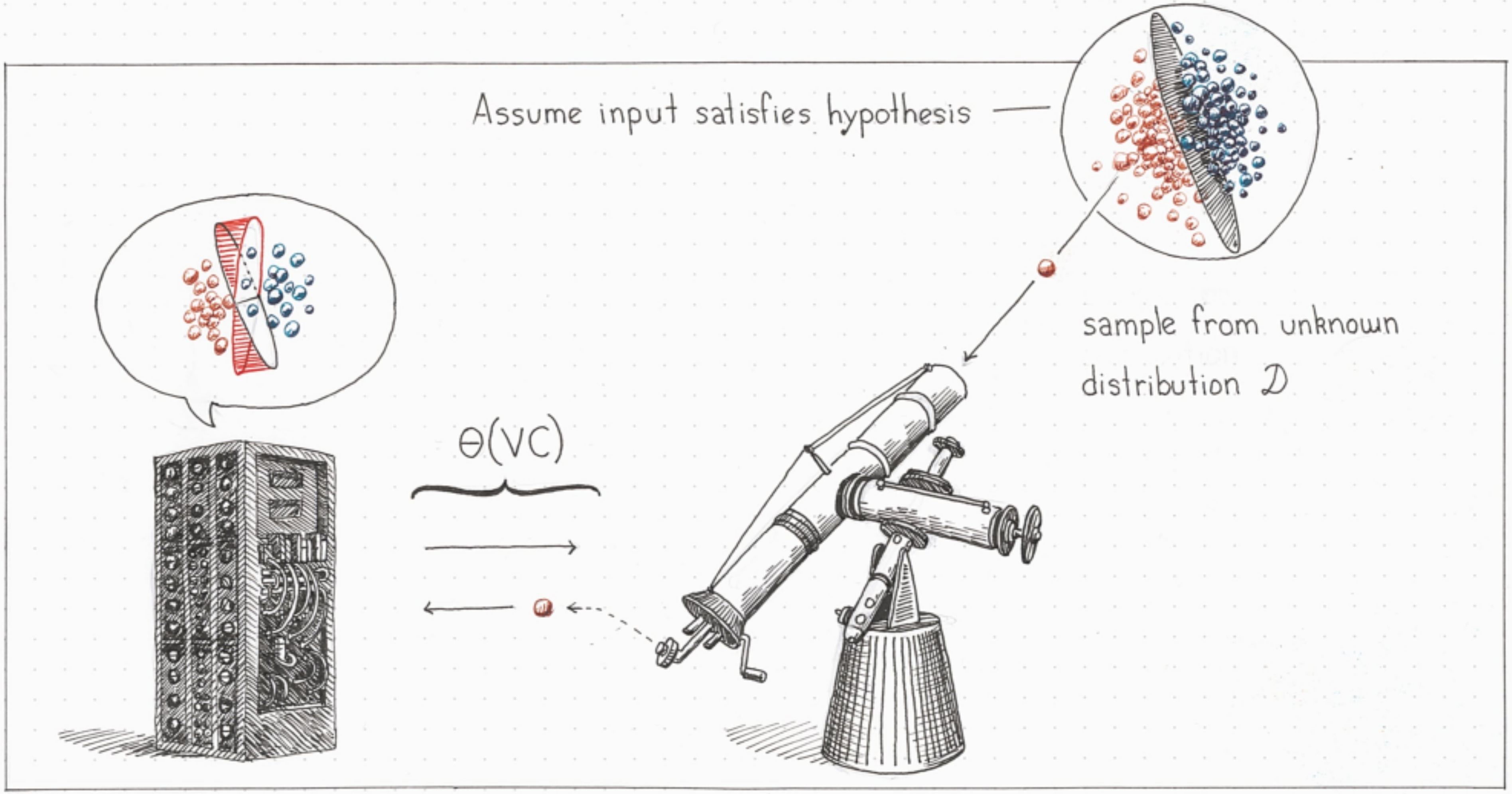


Communication Complexity



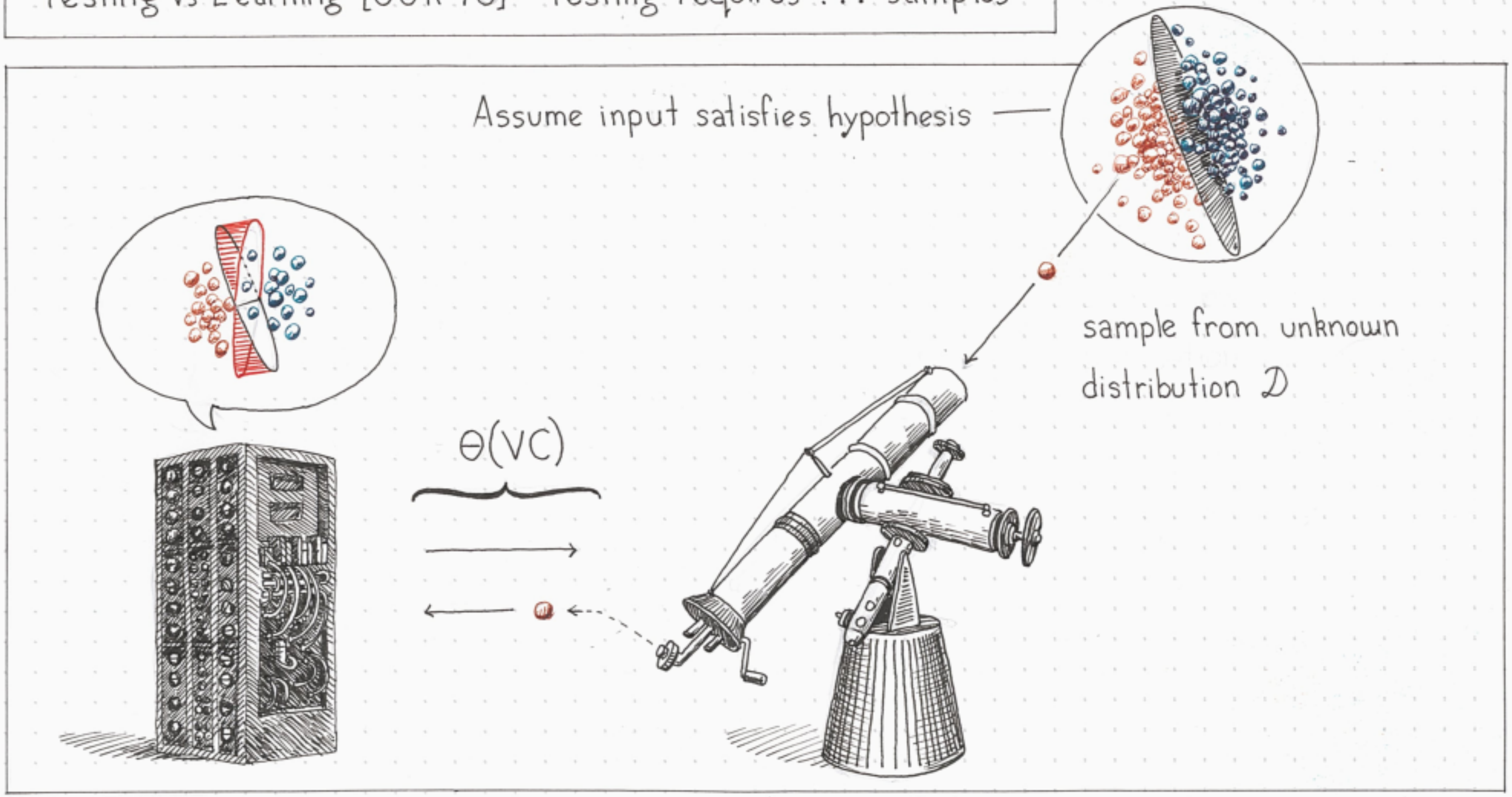


Property Testing

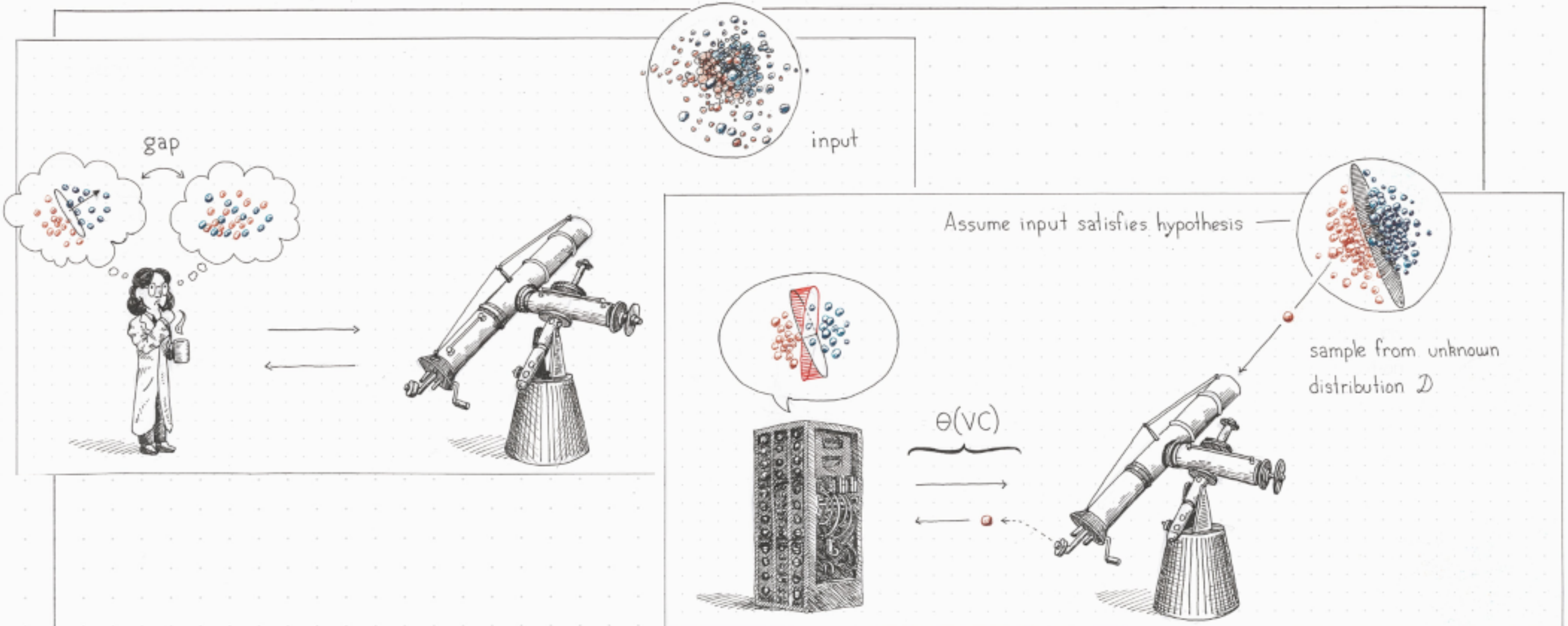

Assume input satisfies hypothesis
sample from unknown distribution $\mathcal{D}$
$\Theta(\mathrm{VC})$

Learning

Testing vs Learning [GGR'98]: Testing requires ??? samples

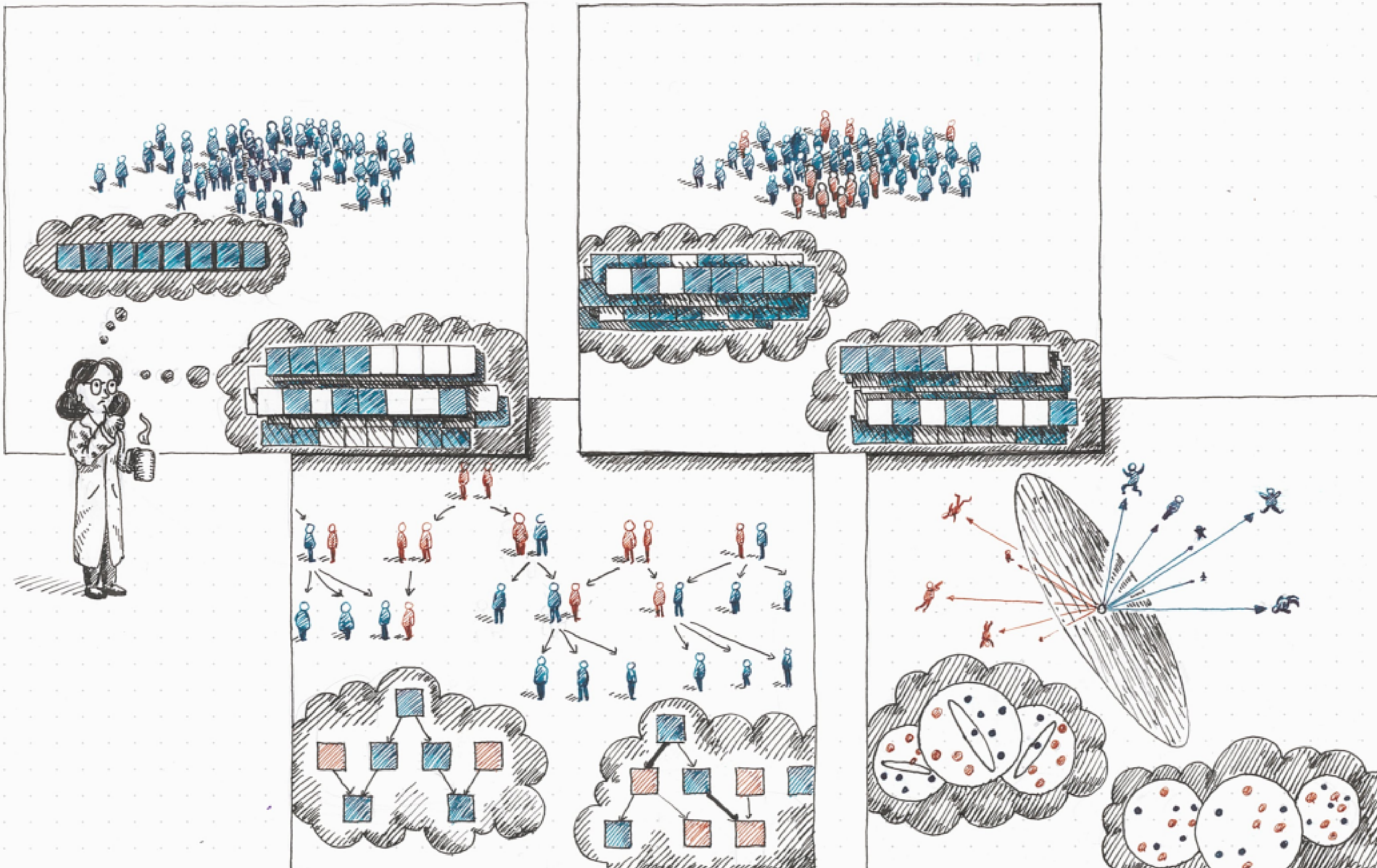


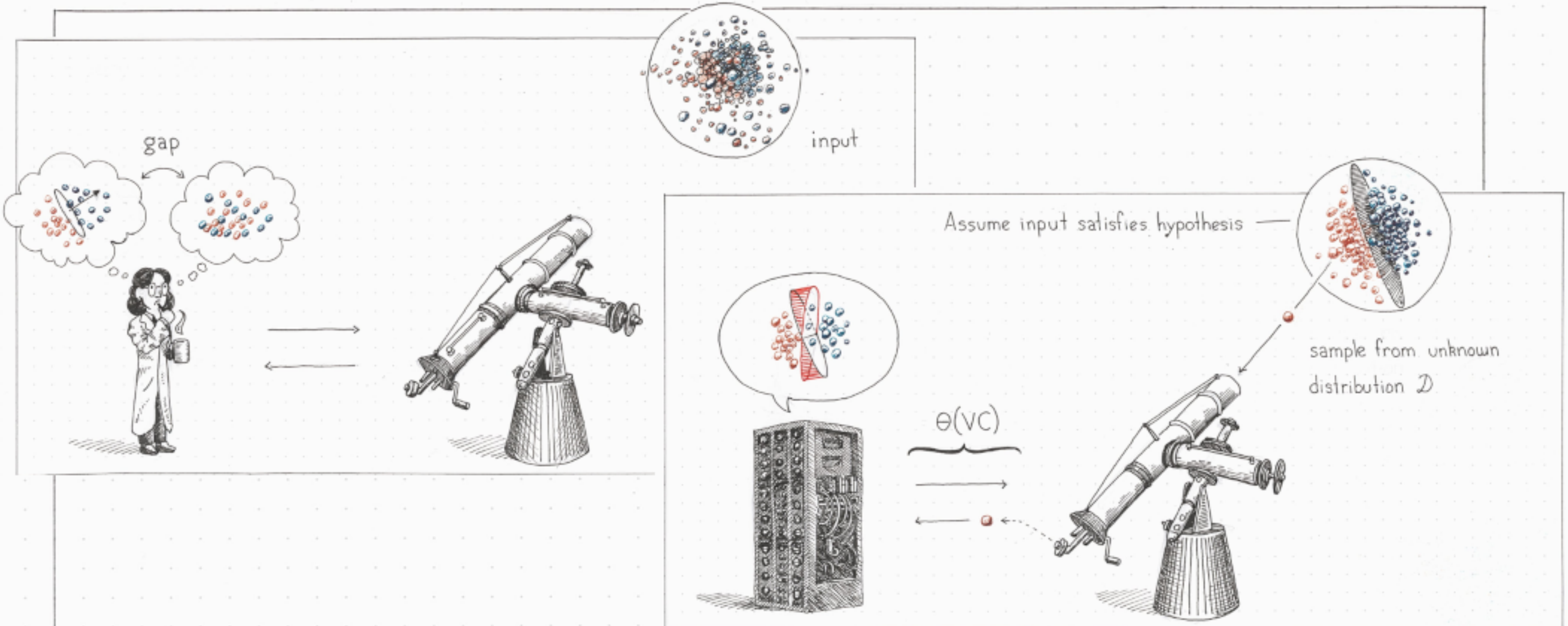
Learning



Theorem: Hypotheses with "large certificates" → $\approx$VC require $\Omega(VC/\log VC)$ samples to test. [BFH'21]

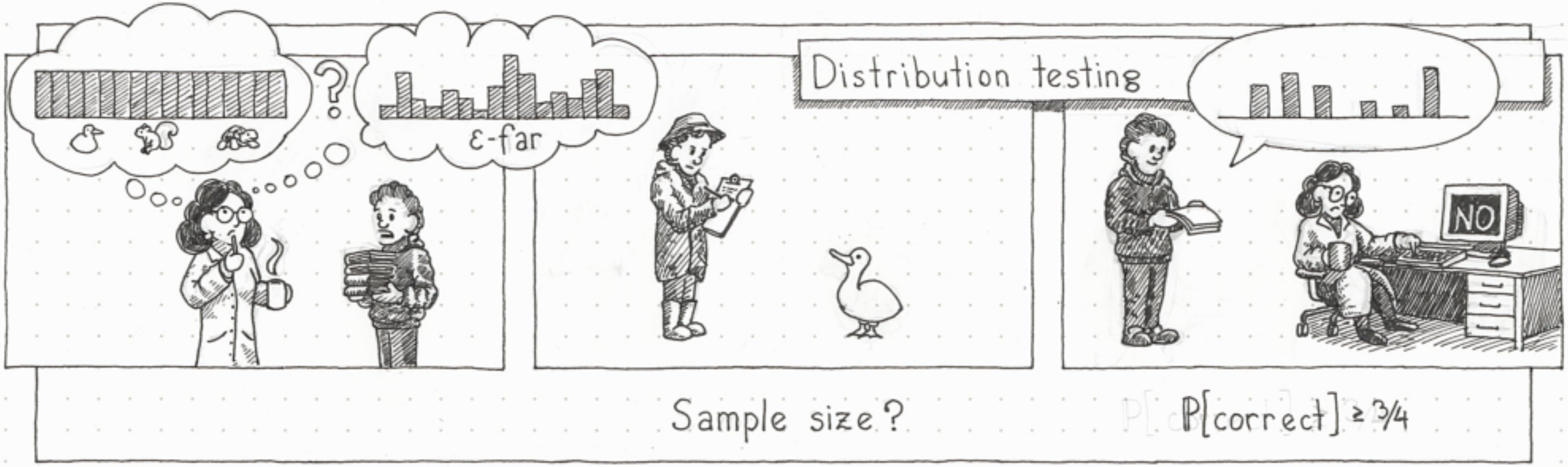
Sometimes tight! — Deeper connection to distribution testing? →



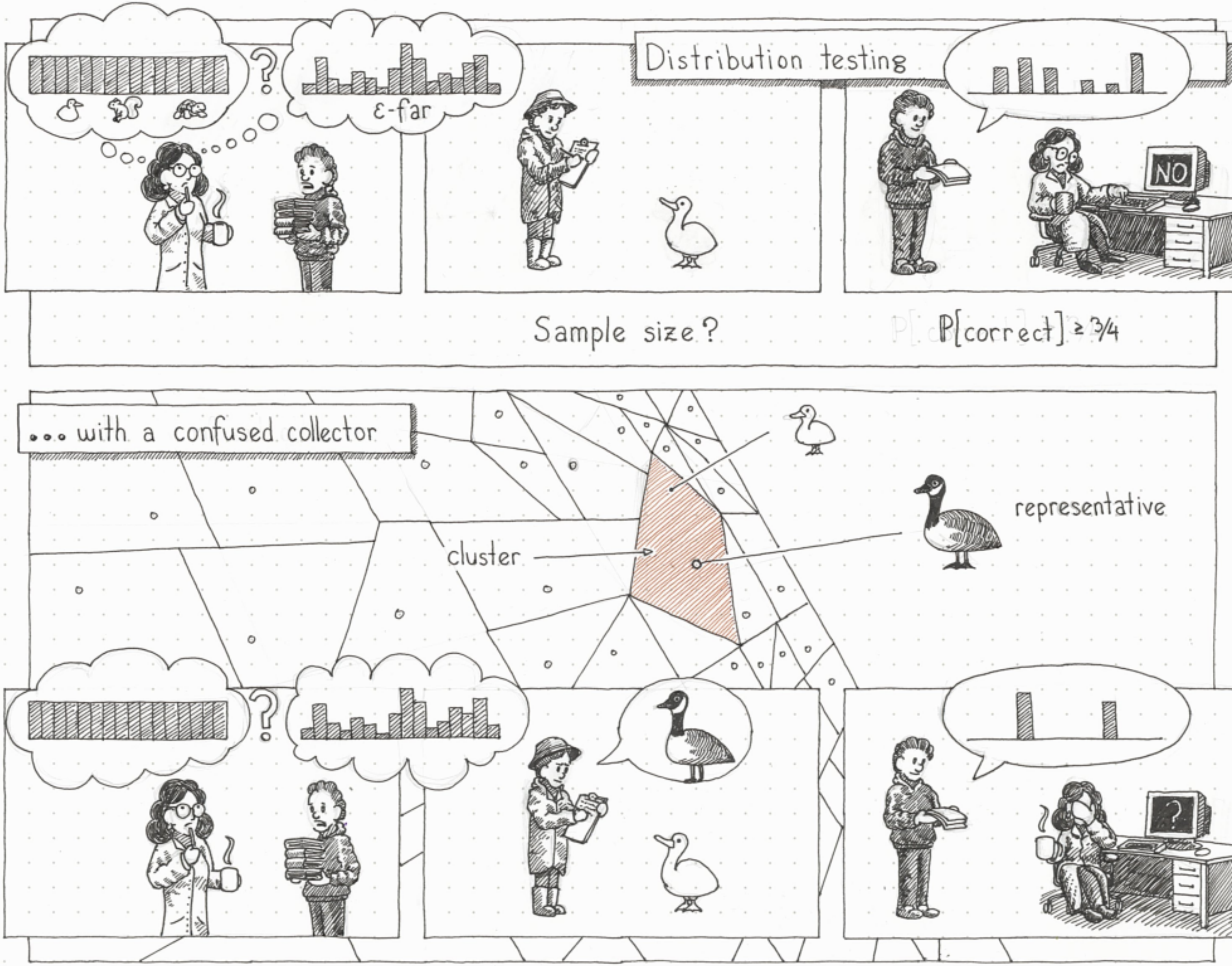


Theorem: Hypotheses with "large certificates" → $\approx VC$
require $\Omega(VC/\log VC)$ samples to test. [BFH'21]

Sometimes tight! — Deeper connection to distribution testing? →


Distribution testing
ε-far
NO
Sample size?
P[correct] ≥ 3/4

... with a confused collector
cluster
representative

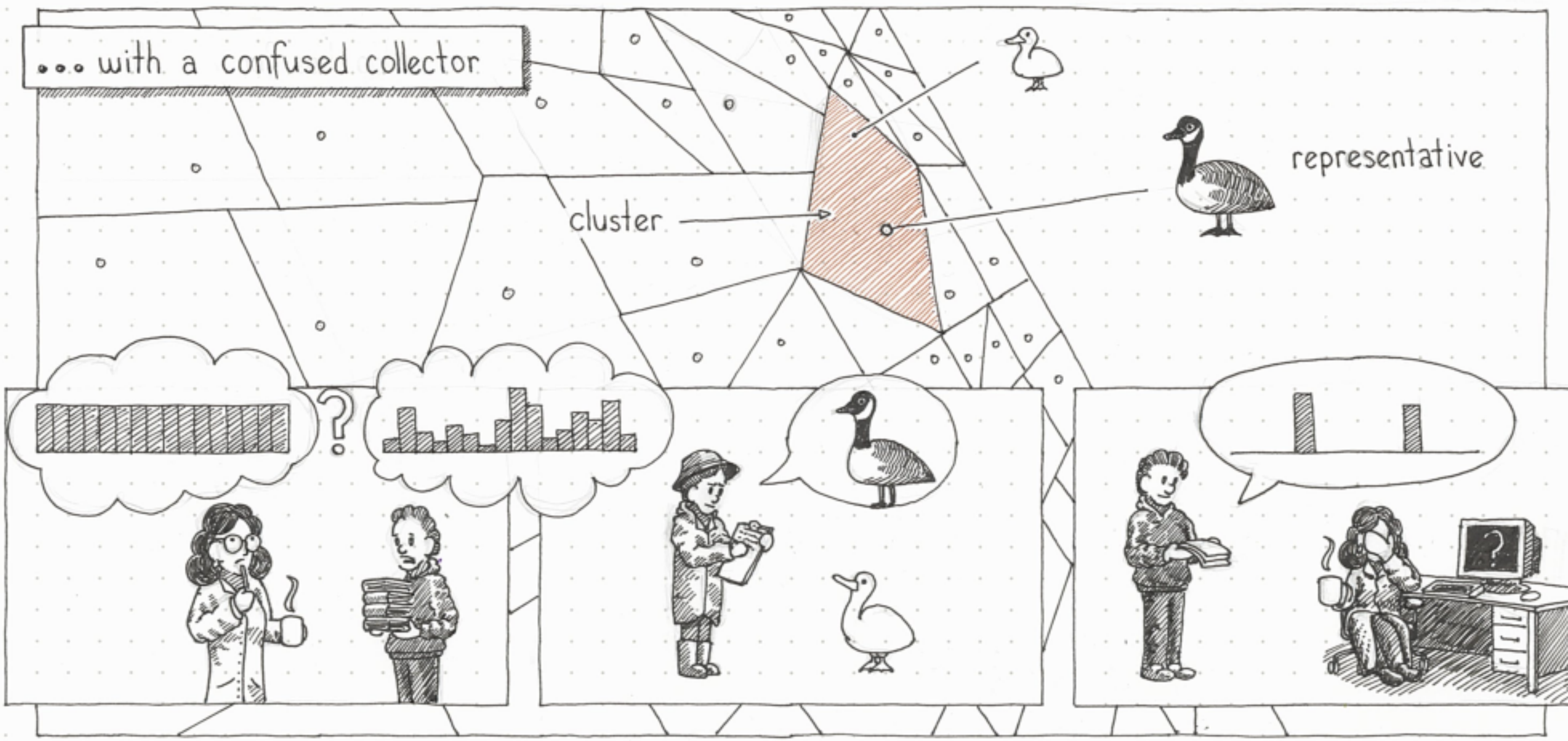

Distribution testing
?
ε-far
NO
Sample size?
ℙ[correct] ≥ 3/4
... with a confused collector
cluster
representative
?
?

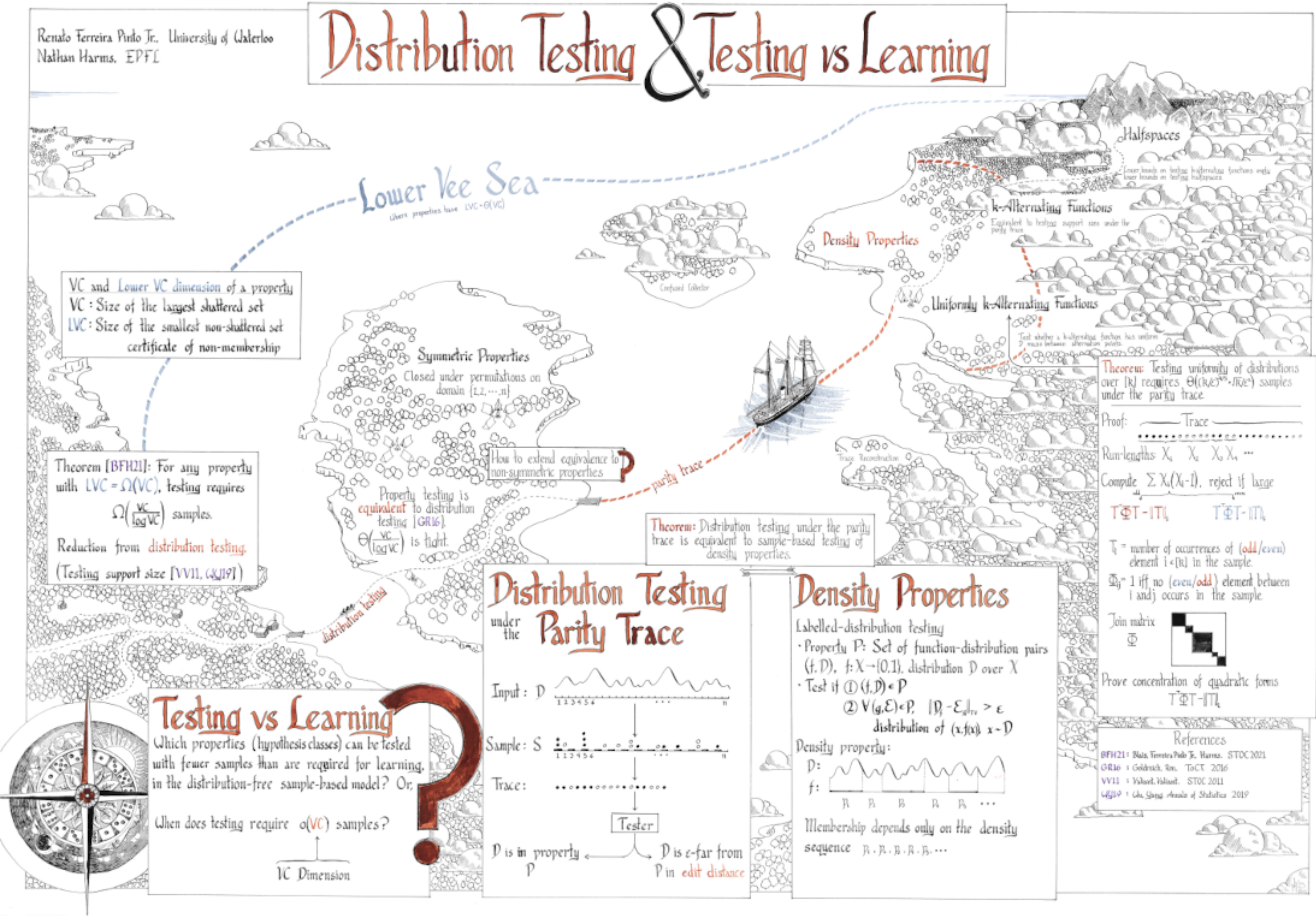
Results: [FH'24]

Algorithms for

random clustering

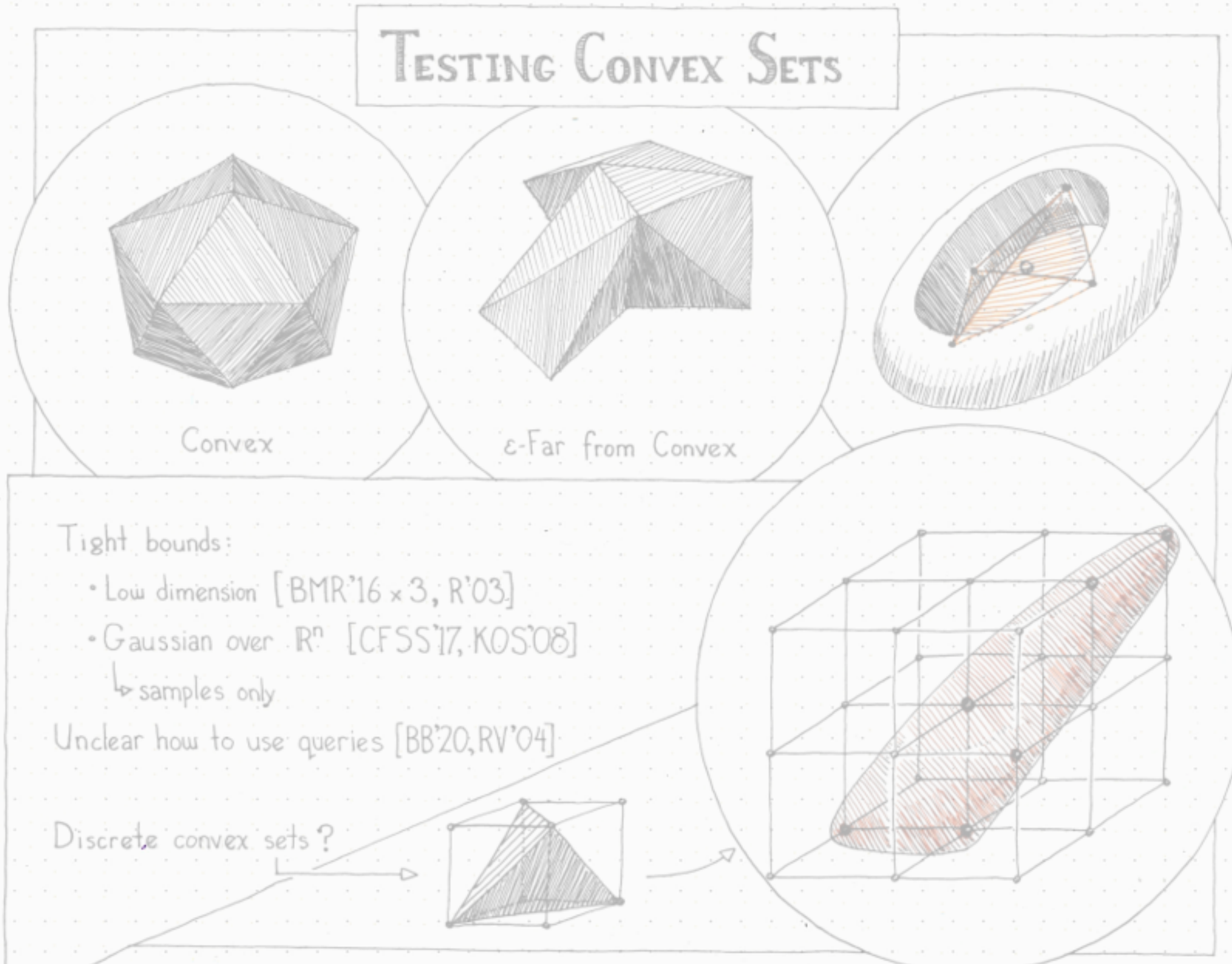
adversarial clustering!

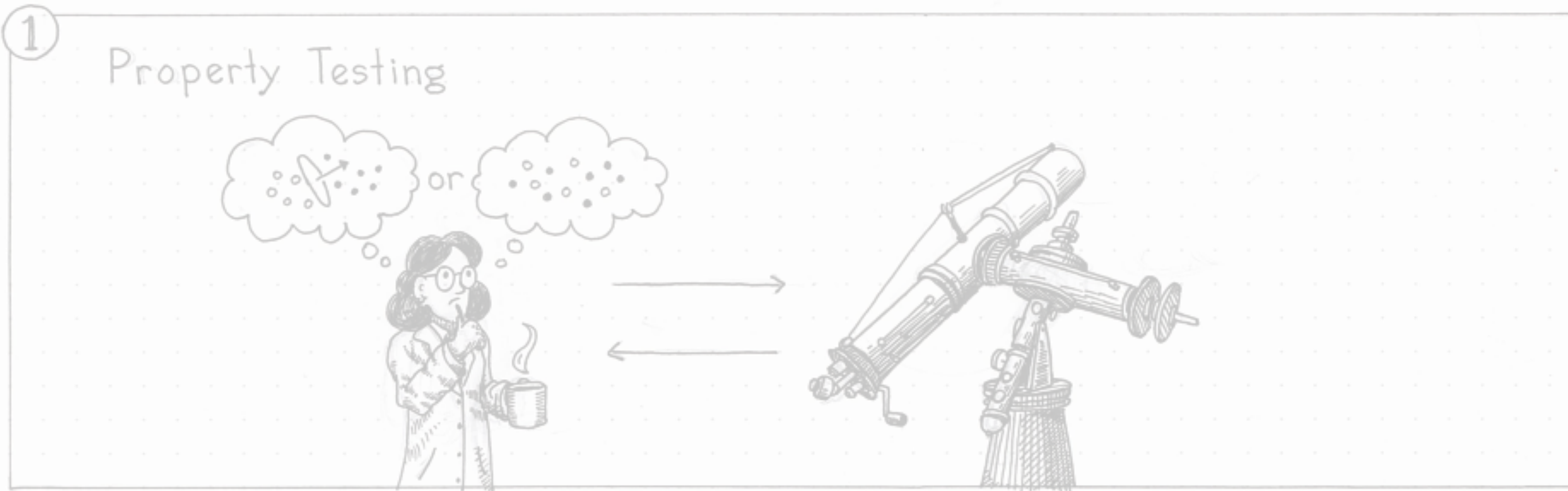


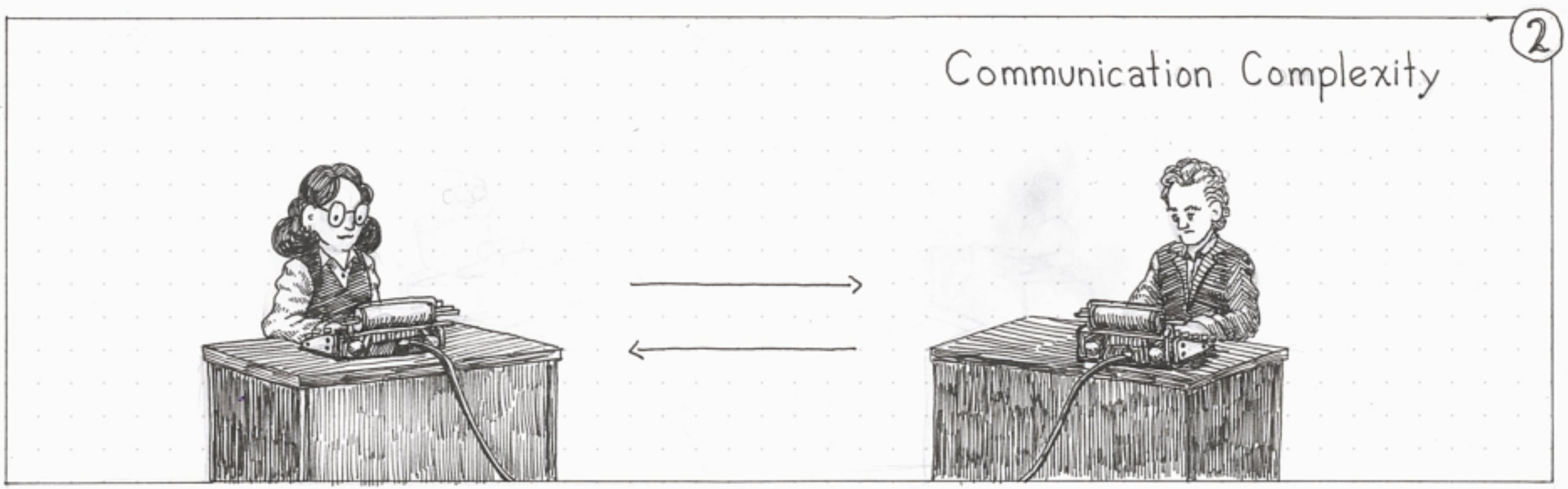


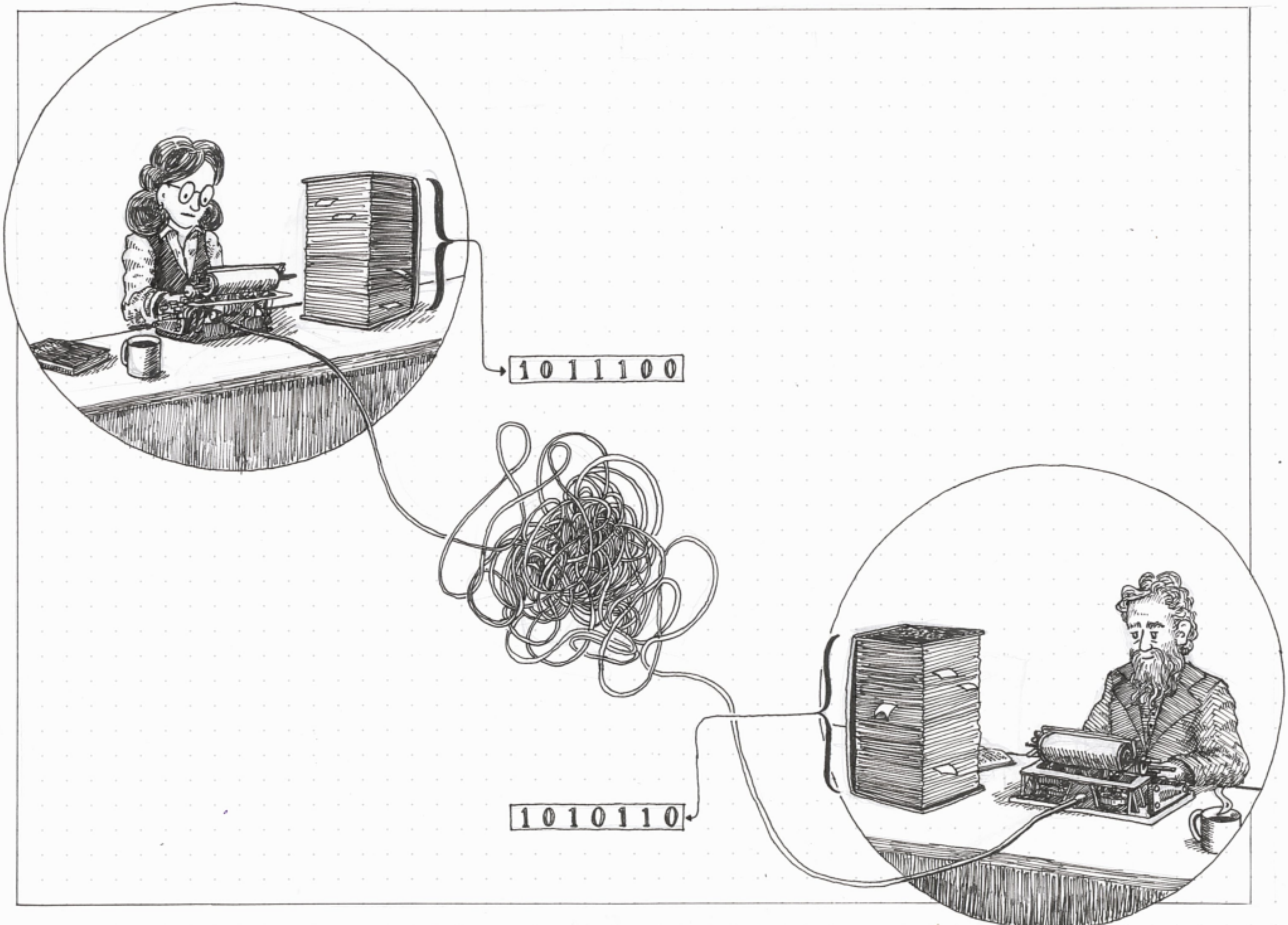
Renato Ferreira Pinto Jr., University of Waterloo
Nathan Harms, EPFL
Distribution Testing & Testing vs Learning
Lower Vee Sea
Where properties have LVC = Θ(VC)
VC and Lower VC dimension of a property
VC: Size of the largest shattered set
LVC: Size of the smallest non-shattered set
certificate of non-membership
Theorem [BFH21]: For any property with LVC = Ω(VC), testing requires Ω(VC / log VC) samples.
Reduction from distribution testing.
(Testing support size [VV11, WY19])
Symmetric Properties
Closed under permutations on domain {1,2,…,n}
How to extend equivalence to non-symmetric properties
Property testing is equivalent to distribution testing [GR16].
Θ(VC / log VC) is tight.
distribution testing
parity trace
Confused Collector
Density Properties
Halfspaces
Lower bounds on testing k-alternating functions imply lower bounds on testing halfspaces
k-Alternating Functions
Equivalent to testing support size under the parity trace
Uniformly k-Alternating Functions
Test whether a k-alternating function has uniform D mass between alternation points.
Trace Reconstruction
Theorem: Distribution testing under the parity trace is equivalent to sample-based testing of density properties.
Theorem: Testing uniformity of distributions over [k] requires Θ((k/ε)^{4/5} + √k/ε²) samples under the parity trace
Proof: Trace
Run-lengths: X₁ X₂ X₃ X₄ …
Compute Σ X_i(X_i − 1), reject if large
odd even
TᵀΦT − ‖T‖₁   TᵀΦT − ‖T‖₁
T_i = number of occurrences of (odd/even) element i ∈ [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 TᵀΦT − ‖T‖₁
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 → {0,1}, distribution D over X
· Test if ① (f, D) ∈ P
② ∀(g, E) ∈ P, ‖D_f − E_g‖_TV > ε
distribution of (x, f(x)), x ~ D
Density property:
D:
f:
P₁ P₂ P₃ P₄ P₅ …
Membership depends only on the density sequence P₁, P₂, P₃, P₄, P₅, …
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
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

Left out... [H, SODA'19], [HY, ICALP'22], [BBH, ITCS'24]




1
Property Testing
or


2
Communication Complexity



Communication Complexity

Why?

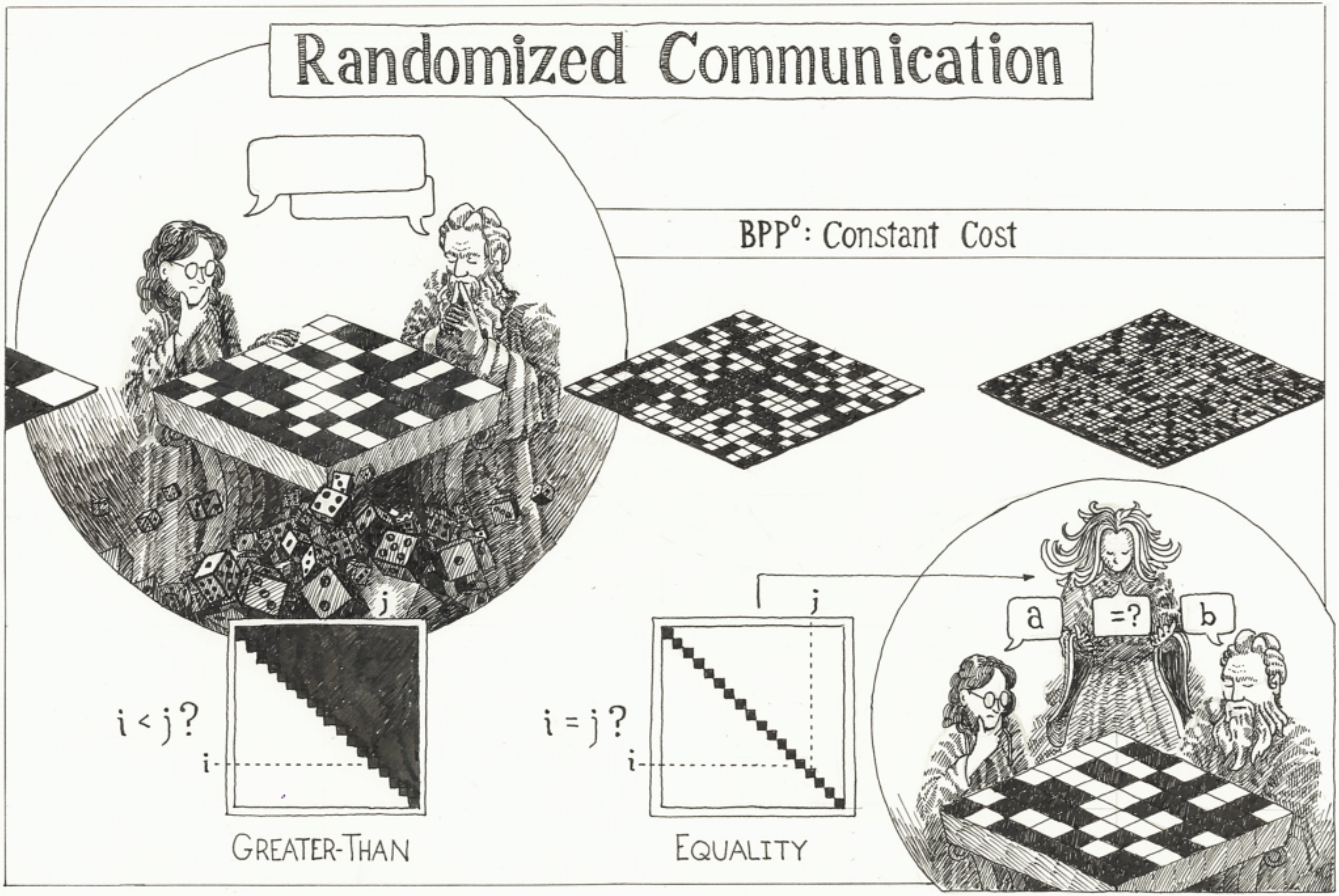
① Power of randomness

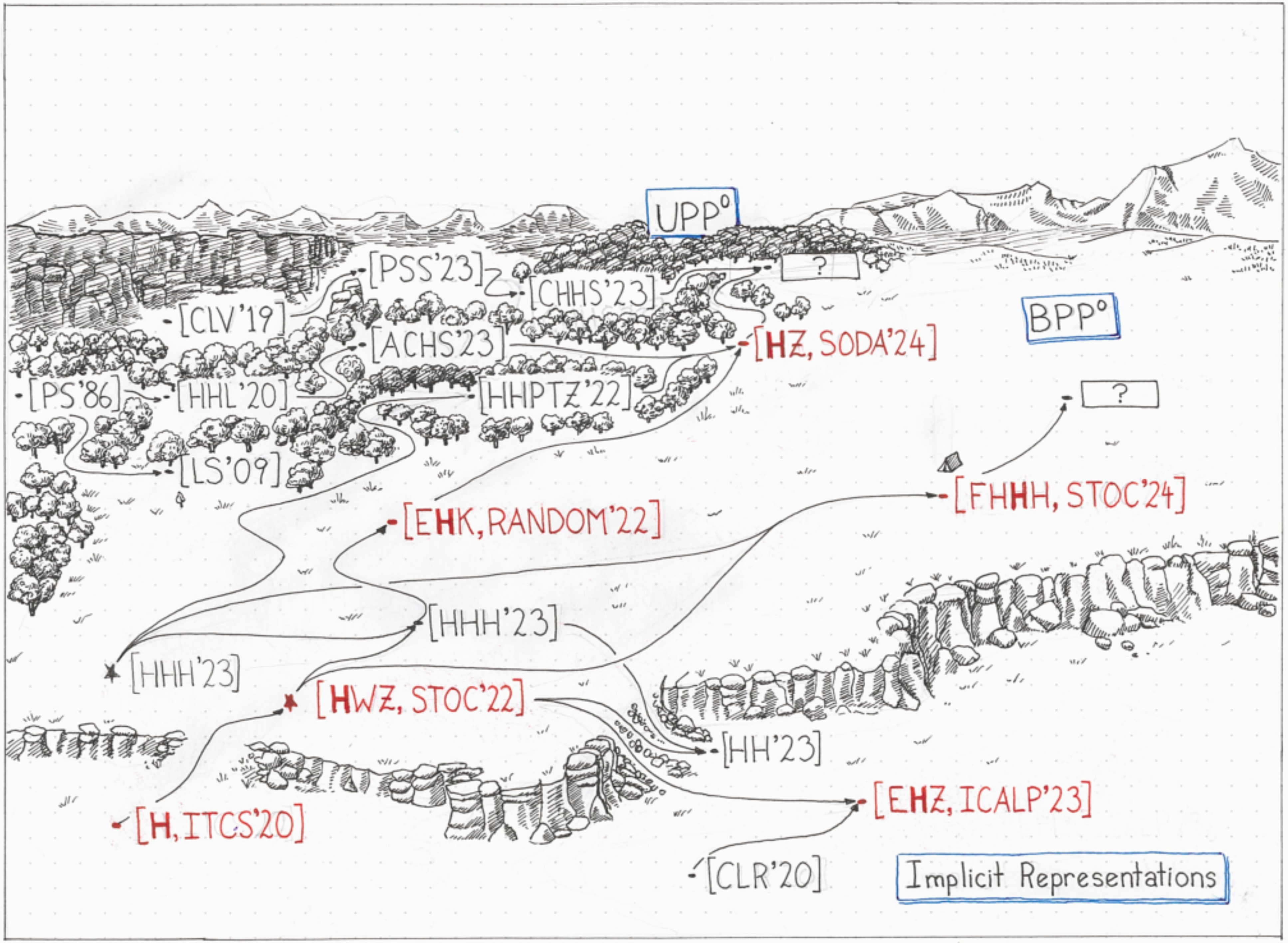
- Most extreme case
- Most basic lower bound
- "Fine-grained" understanding
- Standard techniques fail
- Complete characterization?
- Most evident structure

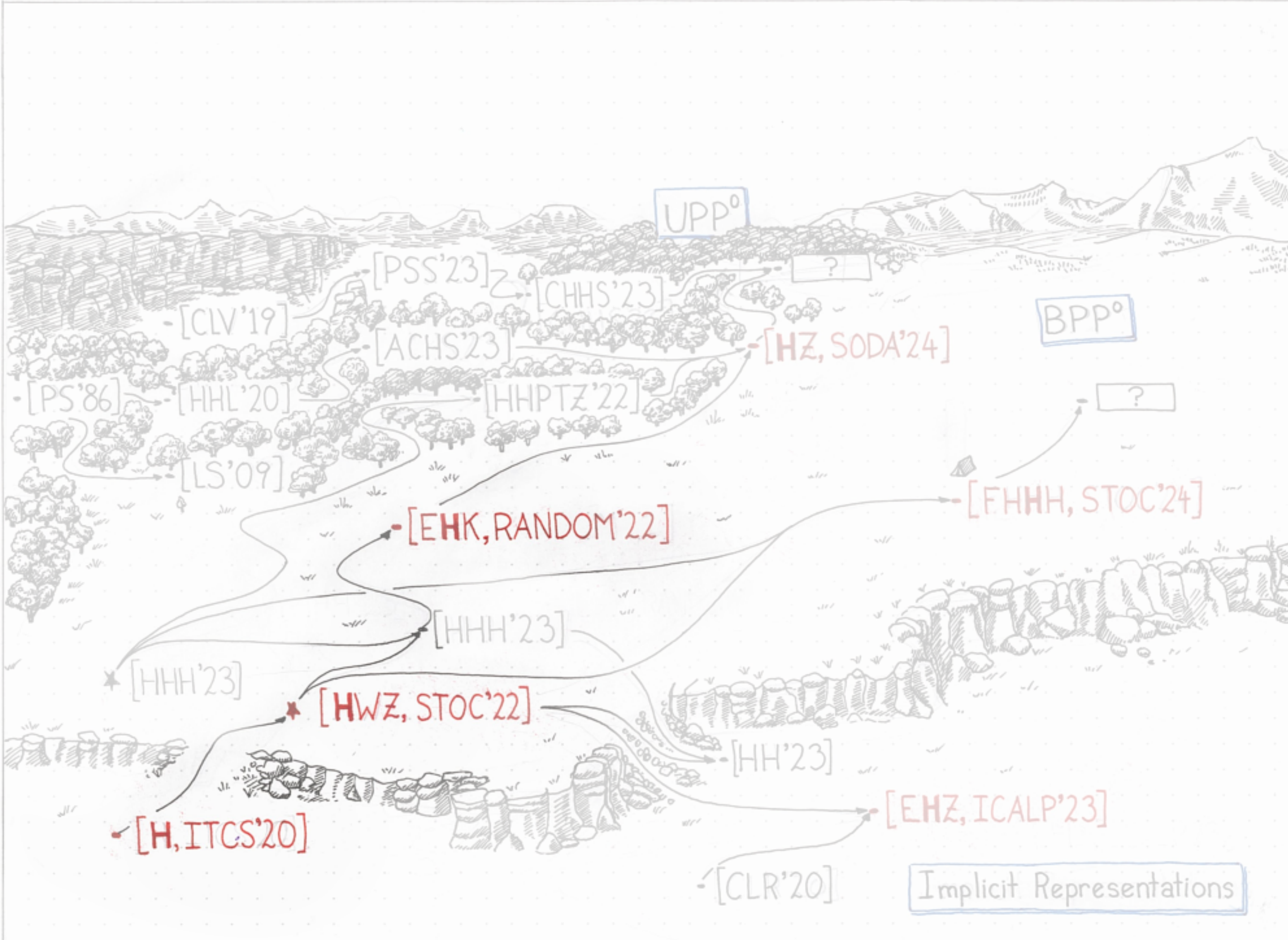
② Connections

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

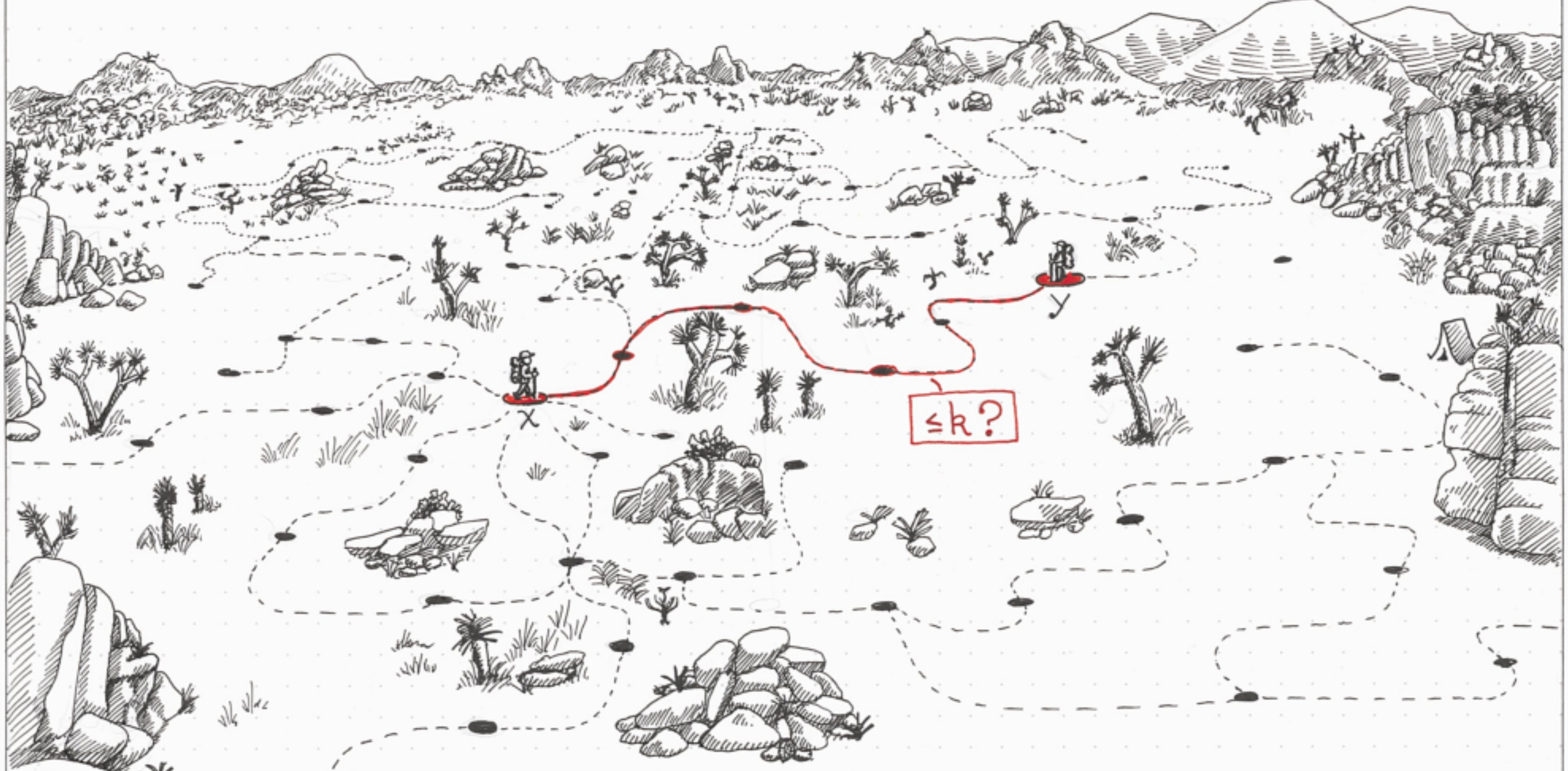
③ It is cool


Randomized Communication
BPP⁰: Constant Cost
a
=?
b
j
i < j?
i
GREATER-THAN
j
i = j?
i
EQUALITY


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



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

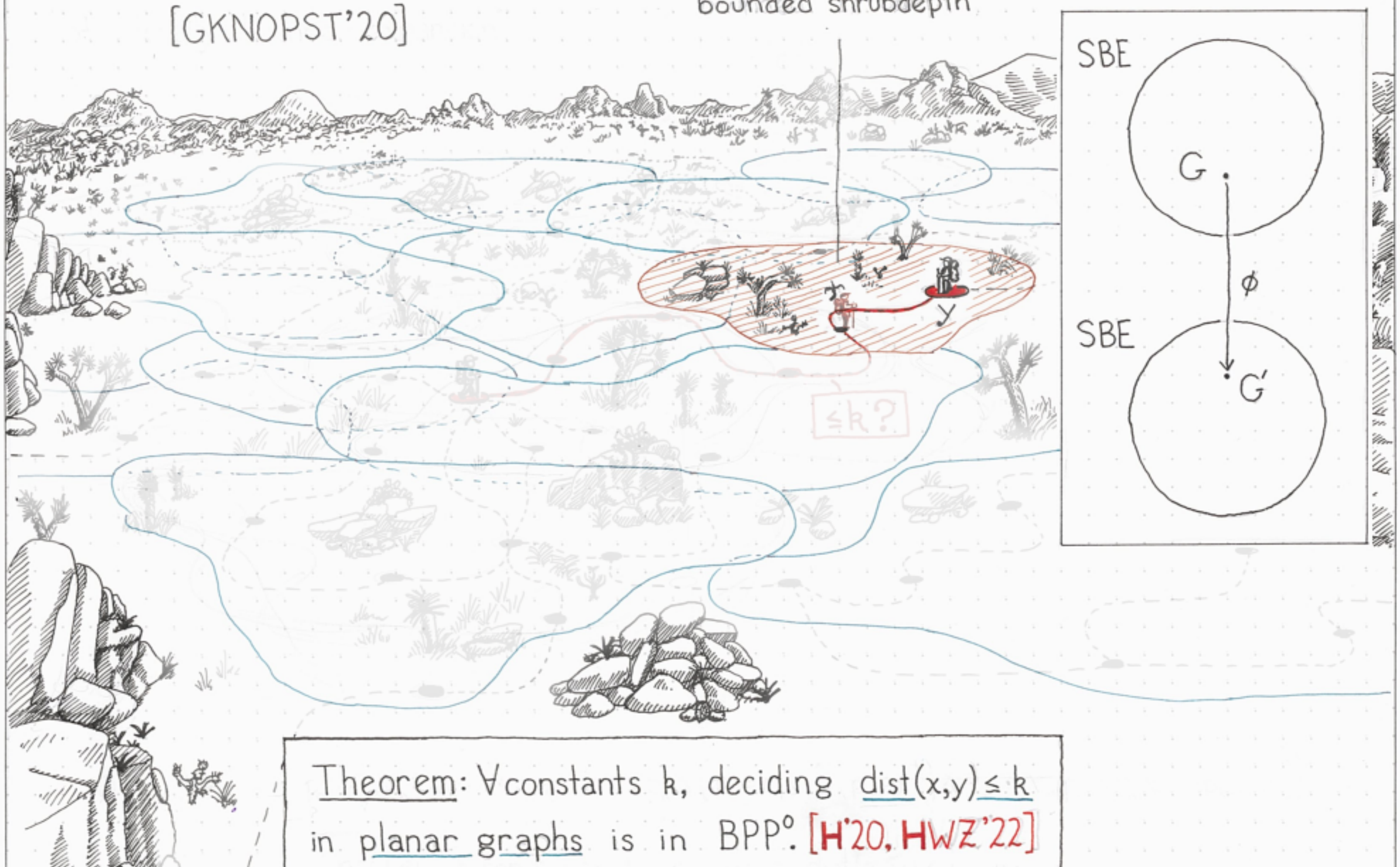


Theorem: ∀ constants k, deciding $\mathrm{dist}(x,y) \le k$ in planar graphs is in BPP^0. [H'20, HWZ'22]

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

"Structurally bounded expansion"

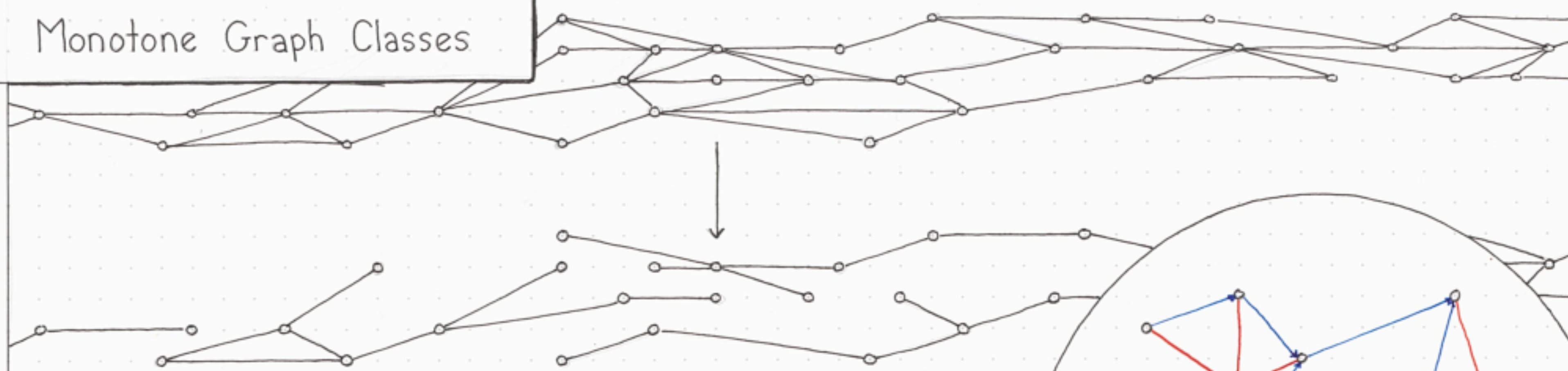
[GKNOPST'20]

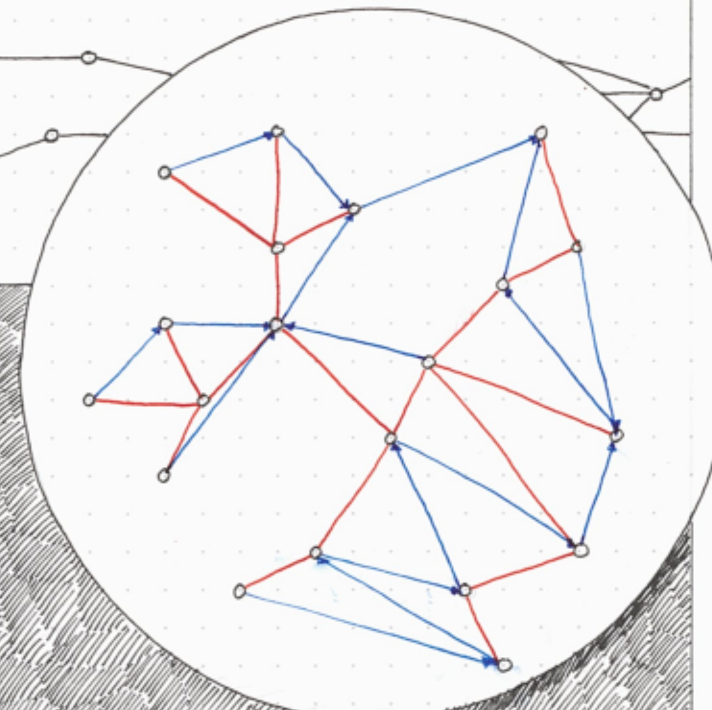


Theorem: $\forall$ constants k, deciding $\mathrm{dist}(x,y) \leq k$ in planar graphs is in BPP^0. [H'20, HWZ'22]

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

Monotone Graph Classes

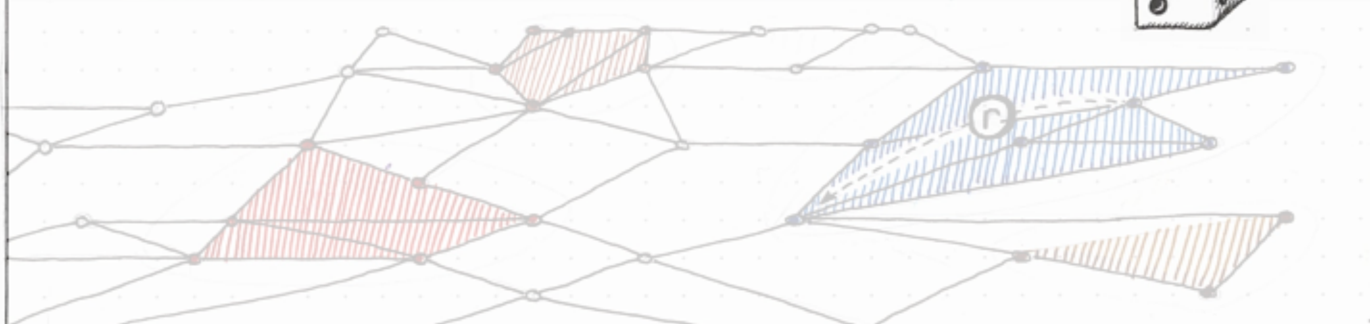




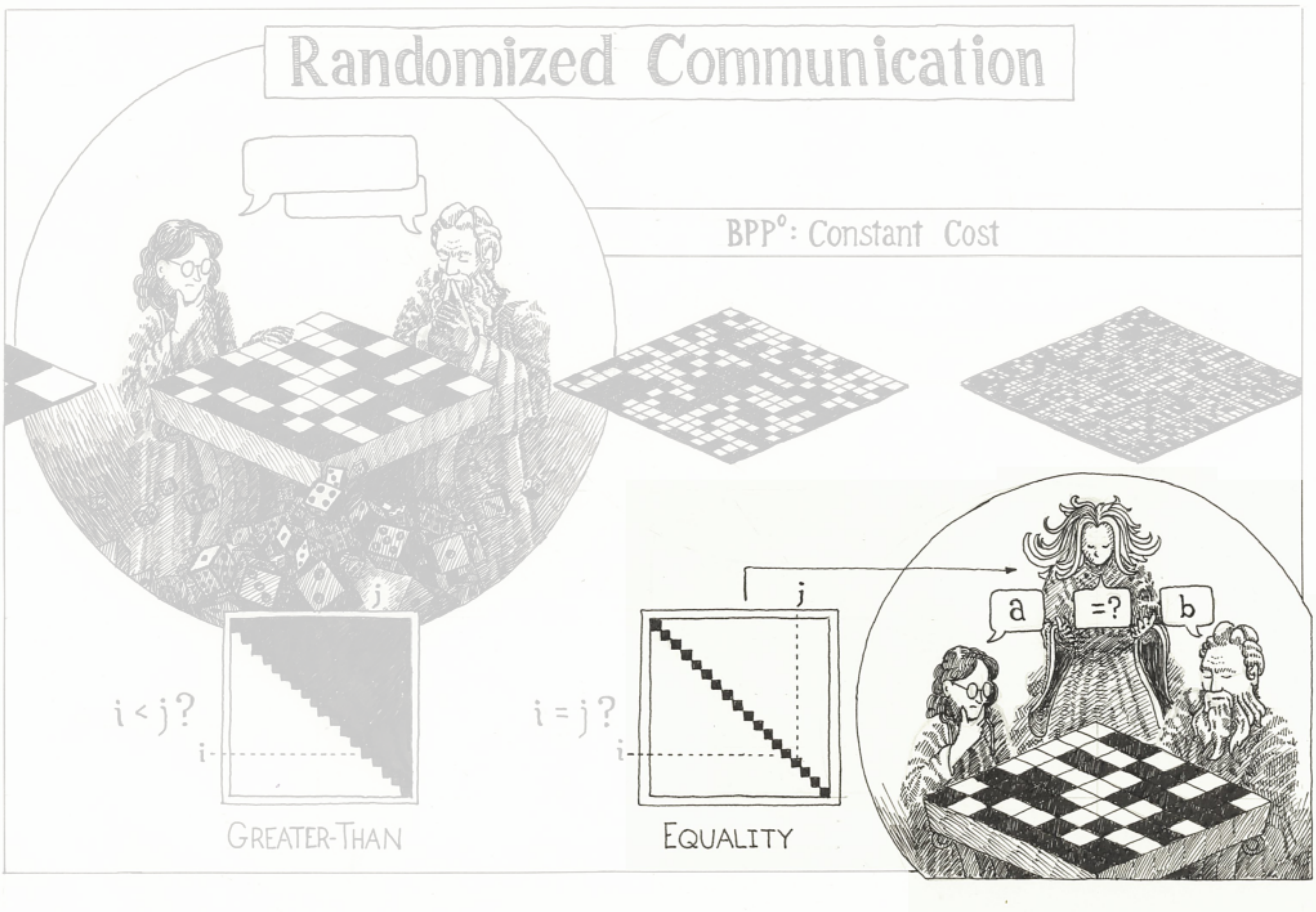
<u>Theorem</u>: If $\mathcal{G}$ is a monotone graph class:

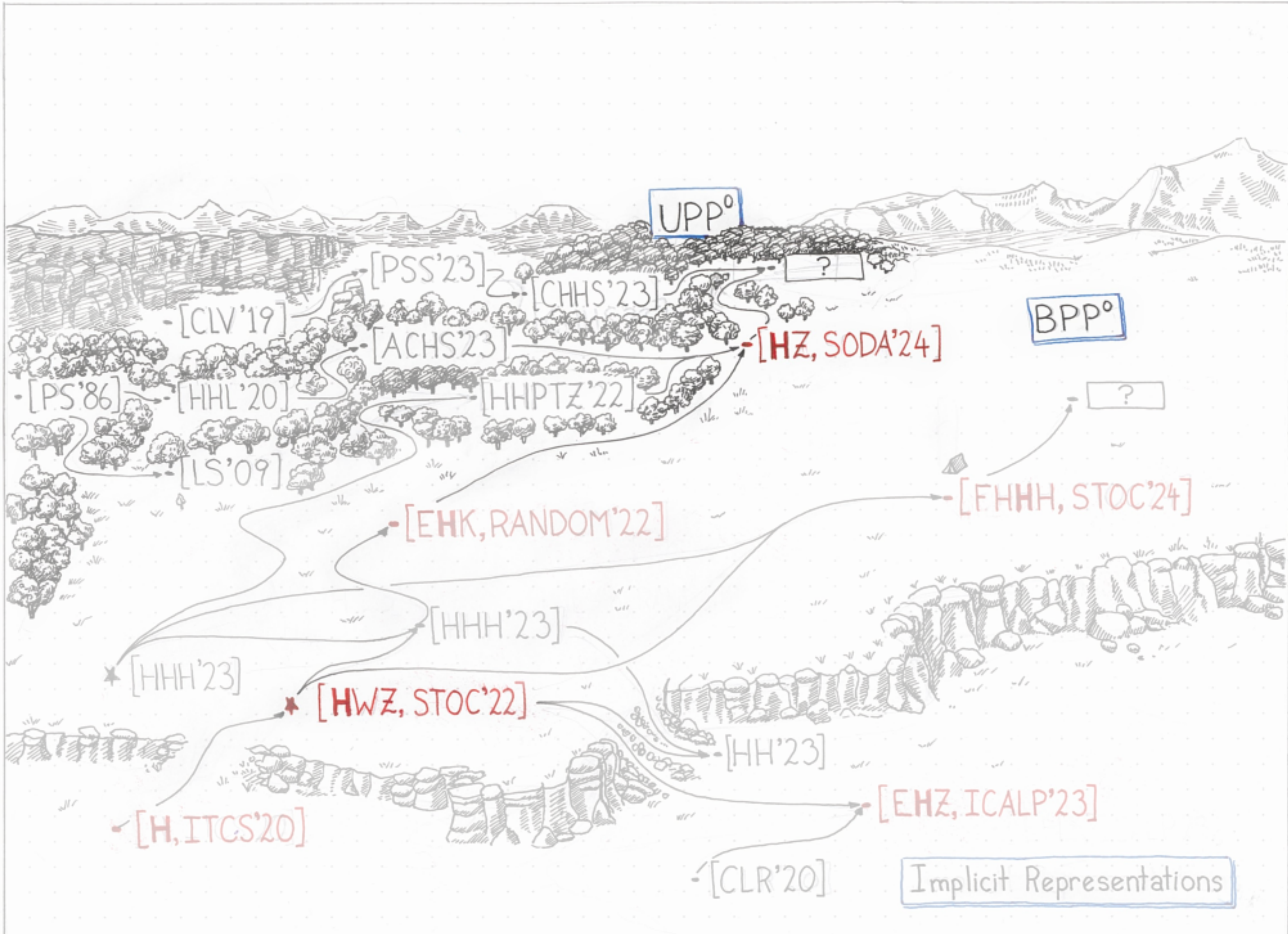
1. Adjacency is in BPP^0
   $\iff$ $\mathcal{G}$ has bounded arboricity.
2. Distance k is in BPP^0 $\forall$constants k
   $\iff$ $\mathcal{G}$ has bounded expansion.

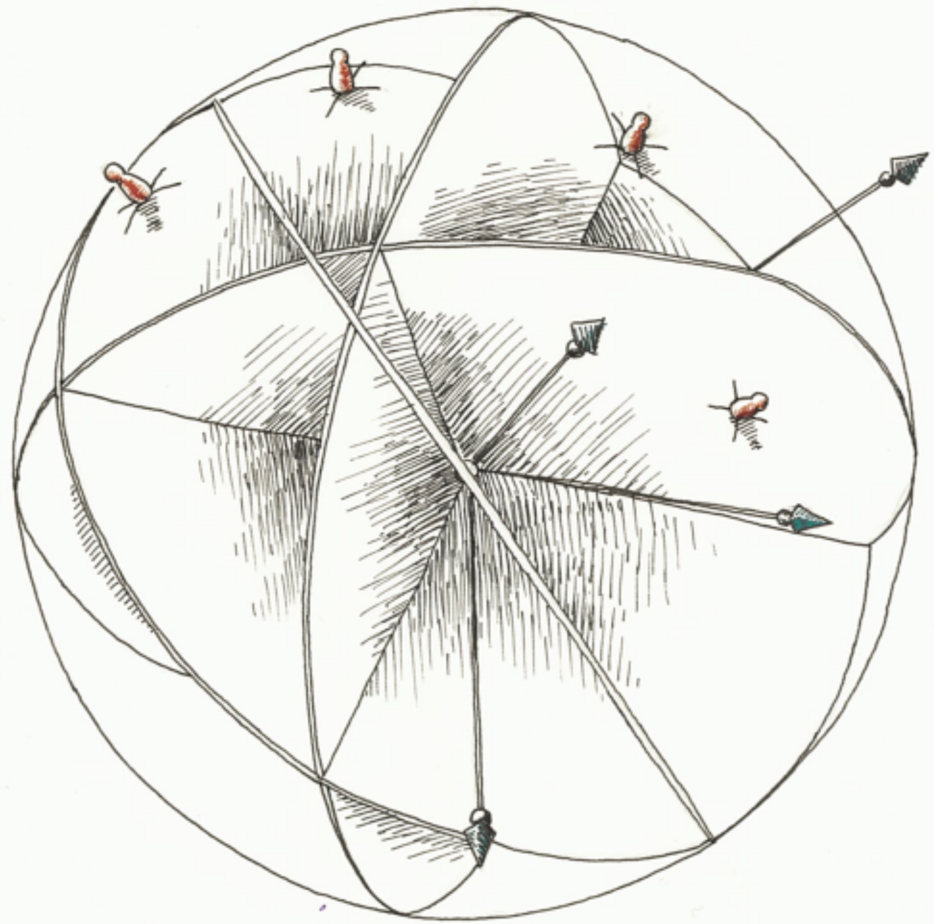
And the EQUALITY oracle suffices. [EHK'22]



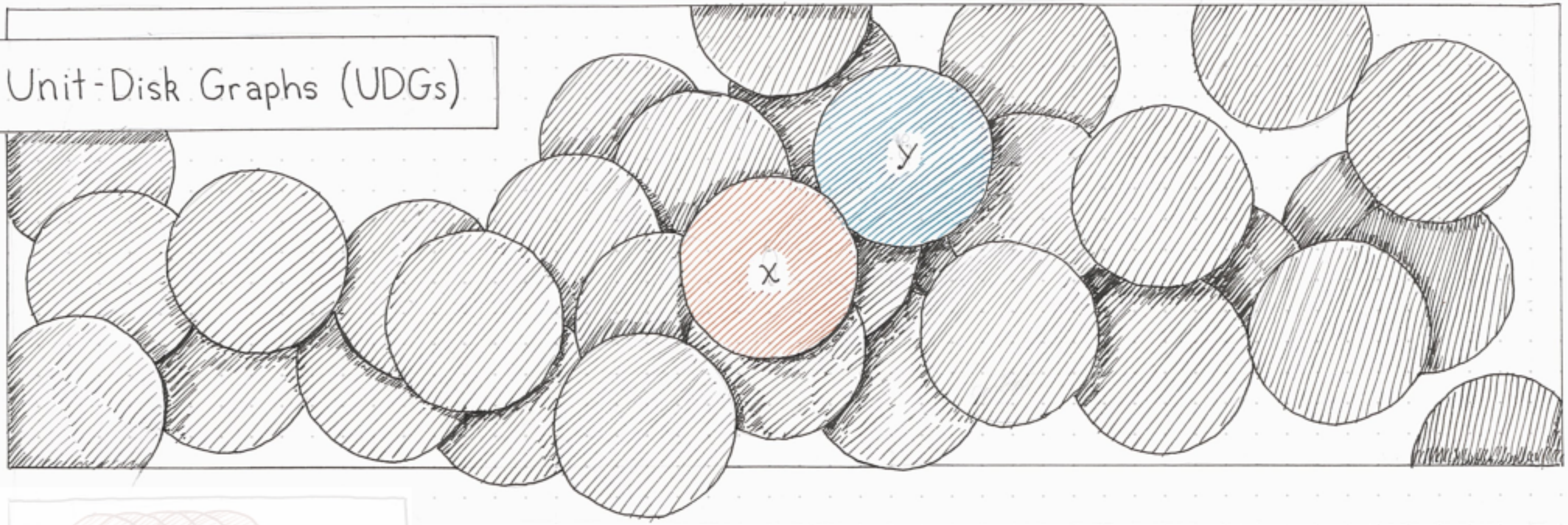
contraction has average degree f(r). [e.g. NO'12]


Randomized Communication
BPP⁰: Constant Cost
a
=?
b
j
i < j?
i
GREATER-THAN
i = j?
j
i
EQUALITY


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



Unit-Disk Graphs (UDGs)



remove

Theorem: UDGs have constant-cost protocols for adjacency iff they are stable. [HZ'24]

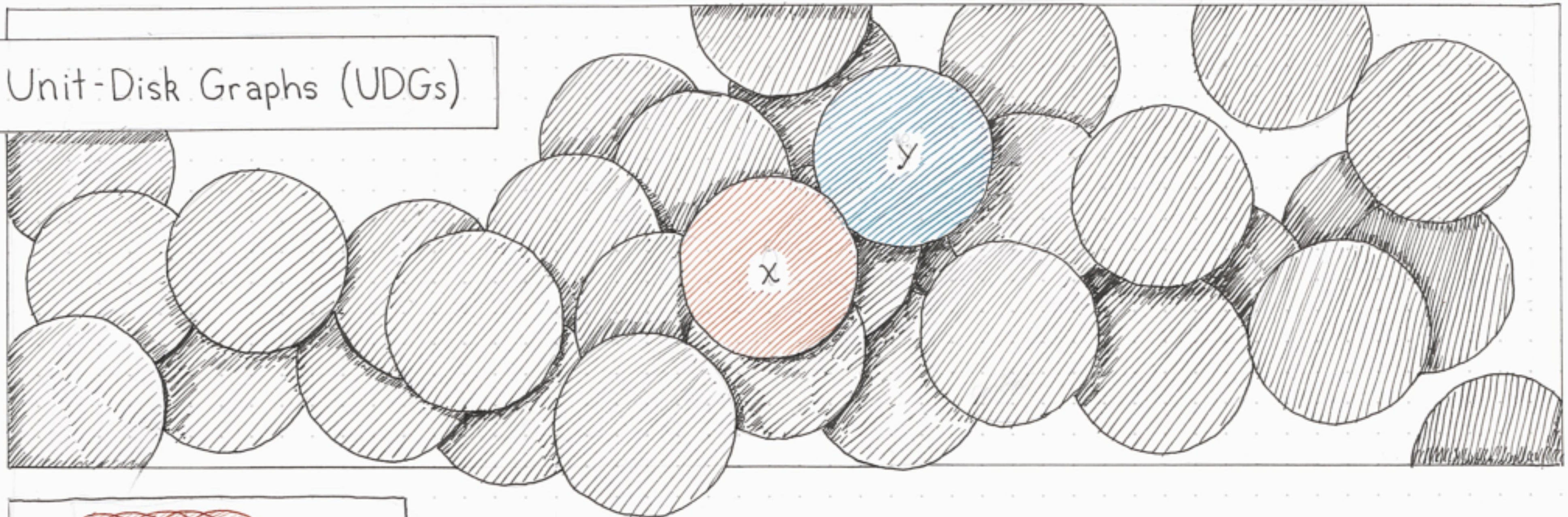
Proof:

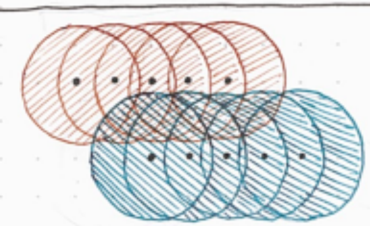
-free

Ramsey's theorem
↓
Recurse on subgraph,
shrink Greater-Than

[OPS'22]: "Gyárfás decomposition"

Unit-Disk Graphs (UDGs)





GREATER-THAN

remove

Theorem: UDGs have constant-cost protocols for adjacency iff they are stable. [HZ'24]

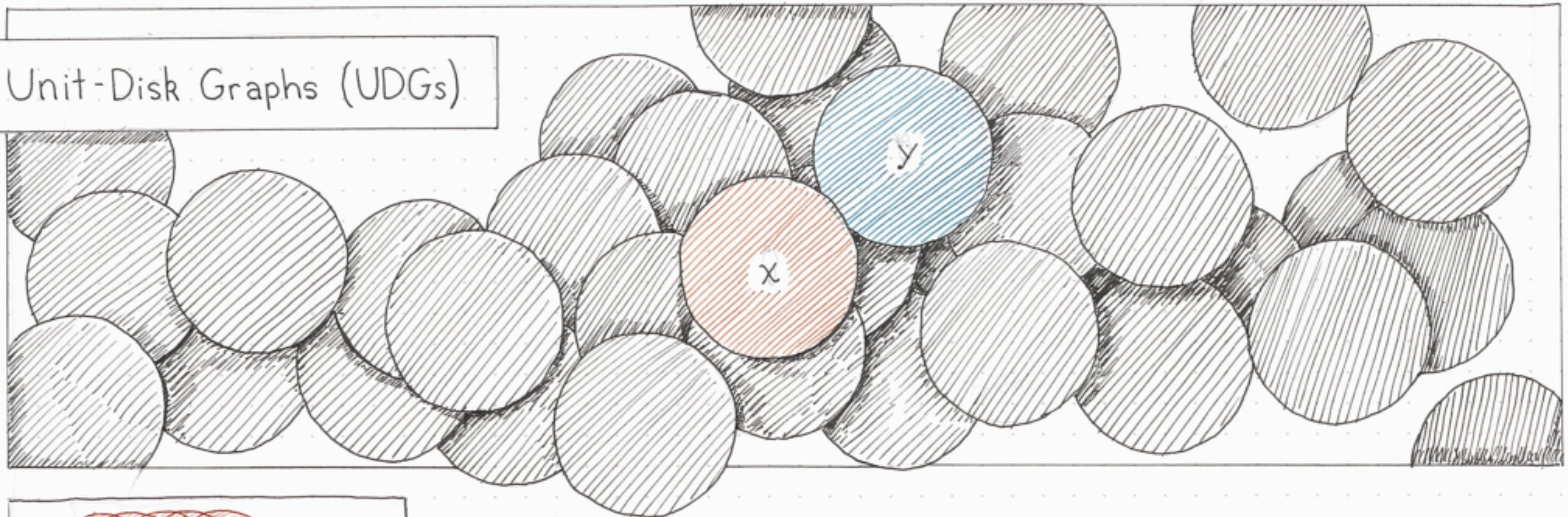
Proof:

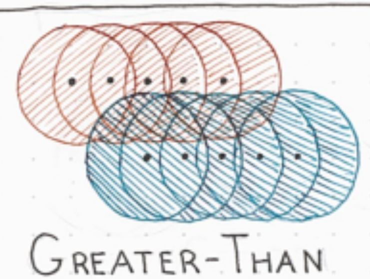
, - free

Ramsey's theorem
↓
Recurse on subgraph, shrink GREATER-THAN

[OPS'22]: "Gyárfás decomposition"

Unit-Disk Graphs (UDGs)





remove

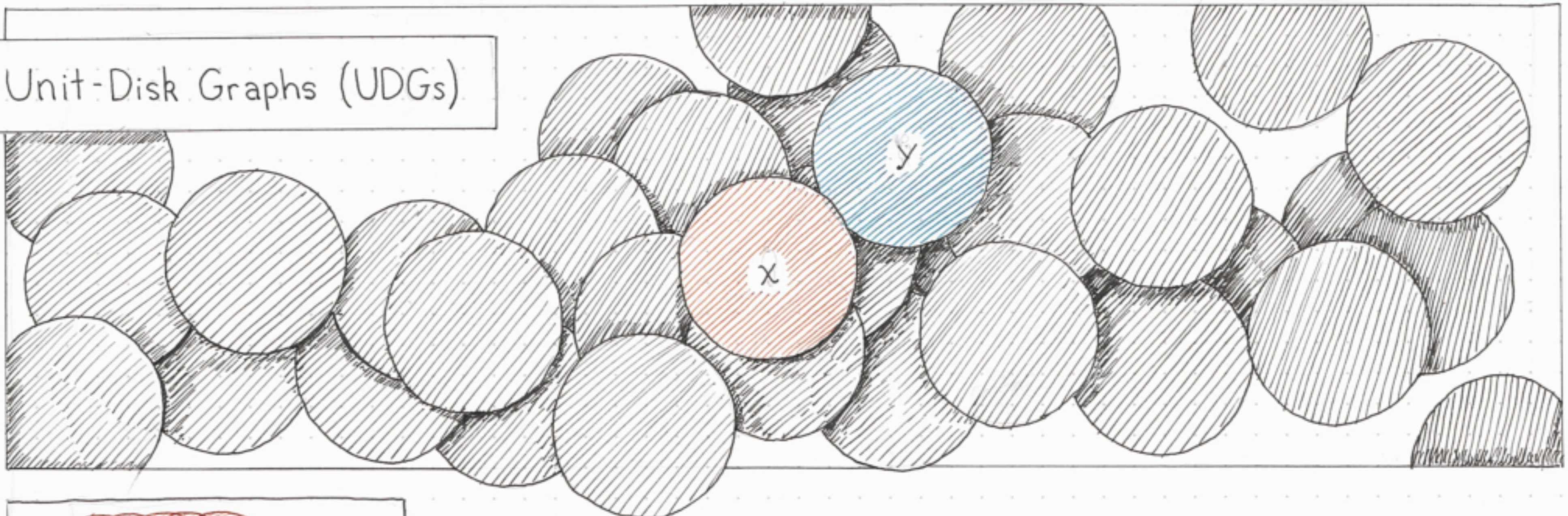
Theorem: UDGs have constant-cost protocols for adjacency iff they are stable. [HZ'24]

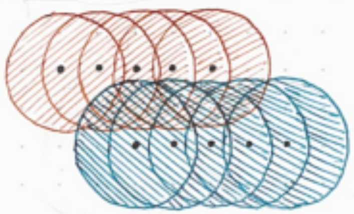
Proof:

Ramsey's theorem
↓
Recurse on subgraph, shrink GREATER-THAN

[OPS'22]: "Gyárfás decomposition"

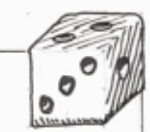
Unit-Disk Graphs (UDGs)



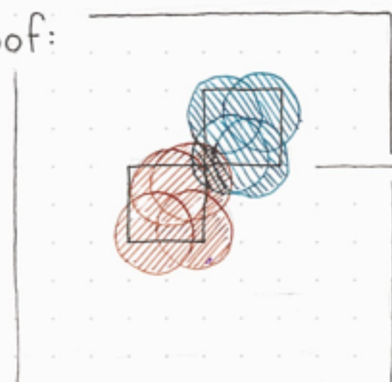


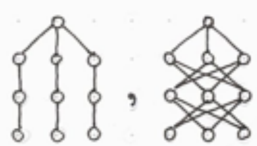
Greater-Than

remove → **Theorem**: UDGs have constant-cost protocols for adjacency iff they are stable. [HZ'24]



Proof:



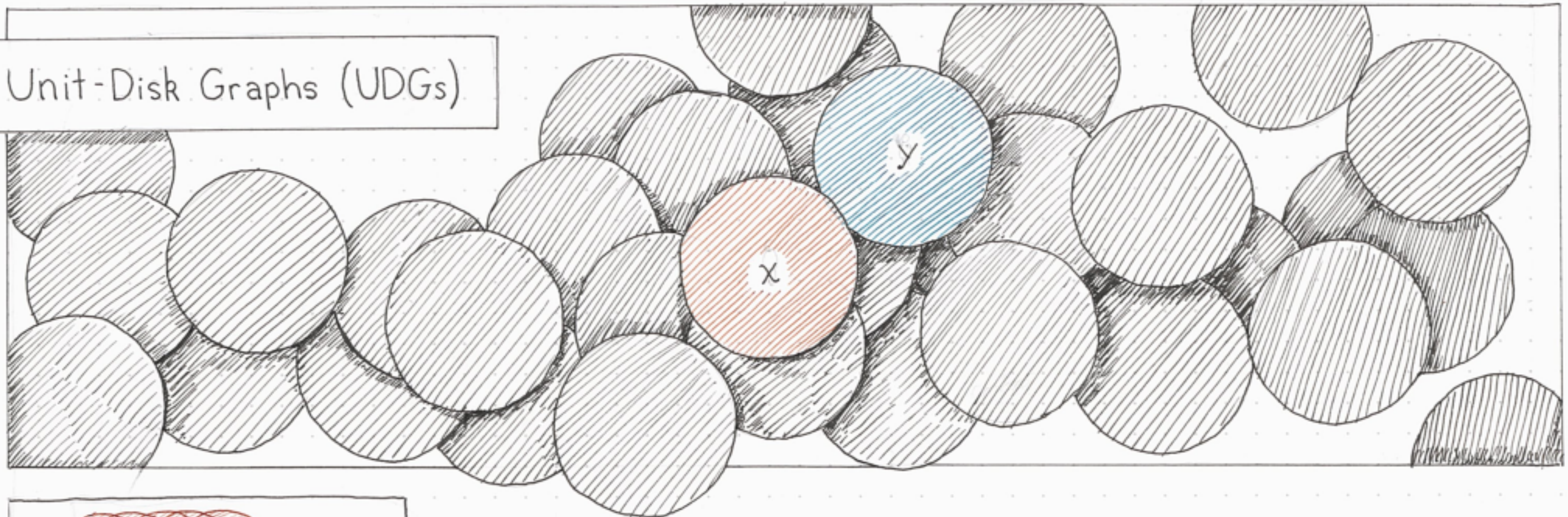


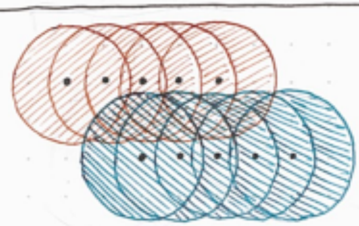
- free →

Ramsey's theorem
↓
Recurse on subgraph,
shrink Greater-Than

[OPS'22]: "Gyárfás decomposition"

Unit-Disk Graphs (UDGs)





Greater-Than

— remove →

Theorem: UDGs have constant-cost protocols for adjacency iff they are stable. [HZ'24]

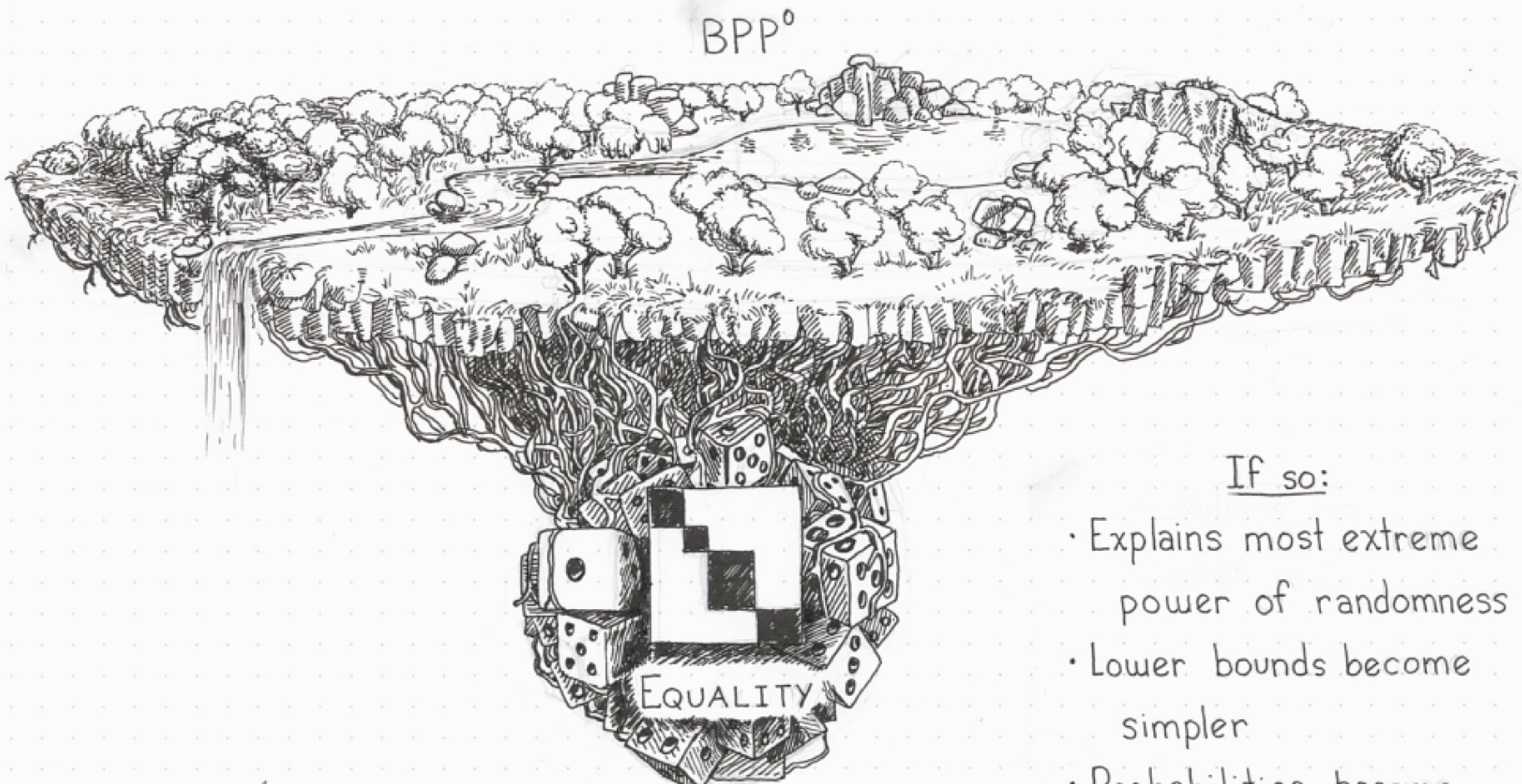
Works also for:

- Dimension 3 halfspaces
- Interval graphs
- Permutation graphs
- etc...

$\Big\} \in \mathrm{UPP}^0$

Only uses Equality! →

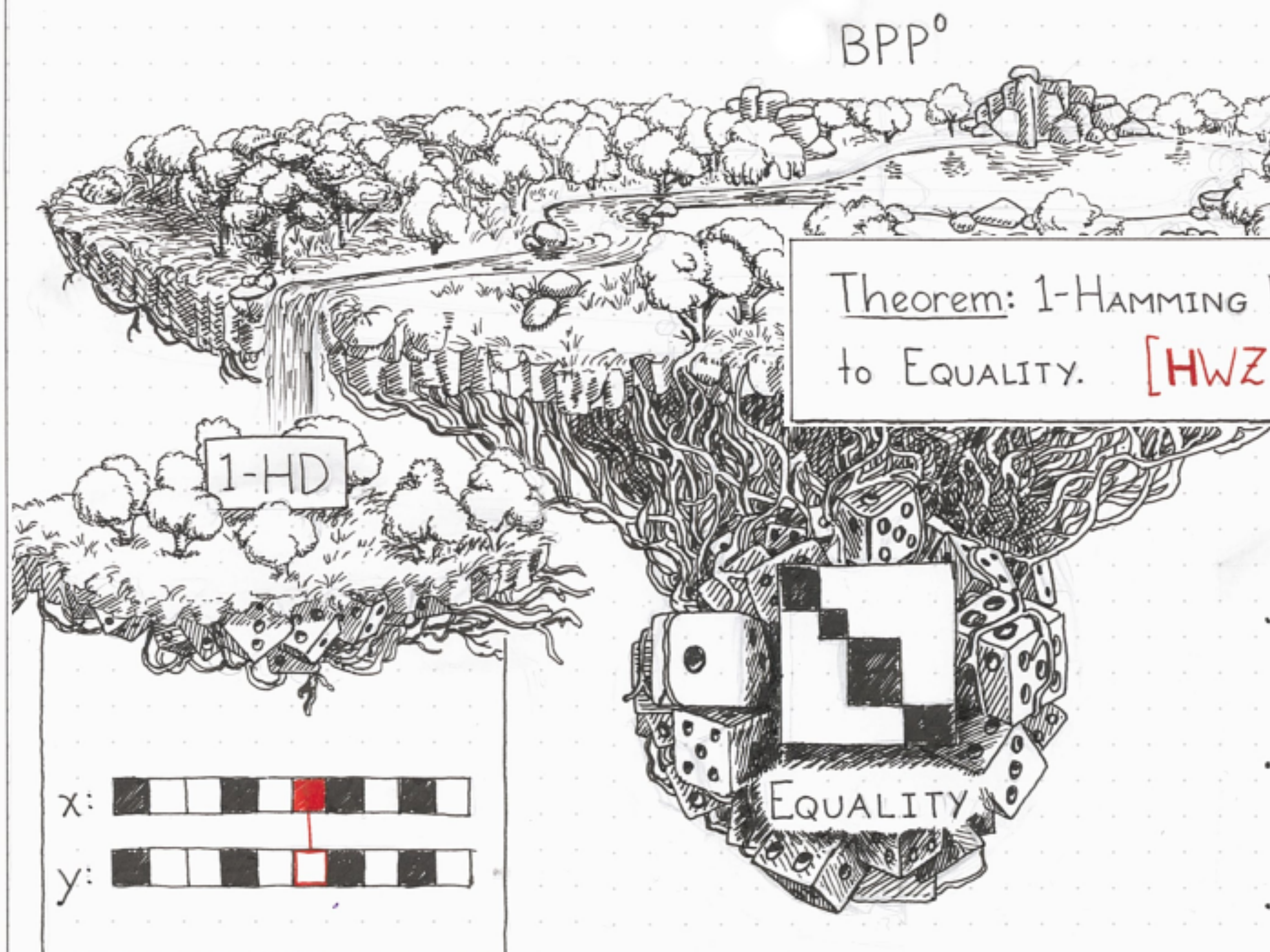
Is EQUALITY "complete" for constant-cost communication?



<u>If so:</u>

- Explains most extreme power of randomness
- Lower bounds become simpler
- Probabilities become unnecessary

Is EQUALITY "complete" for constant-cost communication?



Theorem: 1-HAMMING DISTANCE does not reduce to EQUALITY. [HWZ'22, HHH'23]

If so:

- Explains most extreme power of randomness
- Lower bounds become simpler
- Probabilities become unnecessary

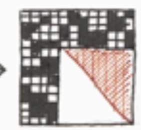
... Then, is 1-Hamming Distance complete for BPP^0?

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

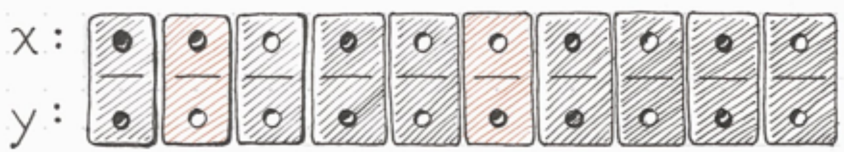
But... Theorem: 2-Hamming Distance does not reduce to 1-Hamming Distance [FHHH'24]

Theorem: There is no complete problem for BPP^0.

Theorem: The k-HAMMING DISTANCE problems form an infinite hierarchy within BPP^0. [FHHH'24]

① Every oracle in BPP^0 is stable. →  } $O(1)$

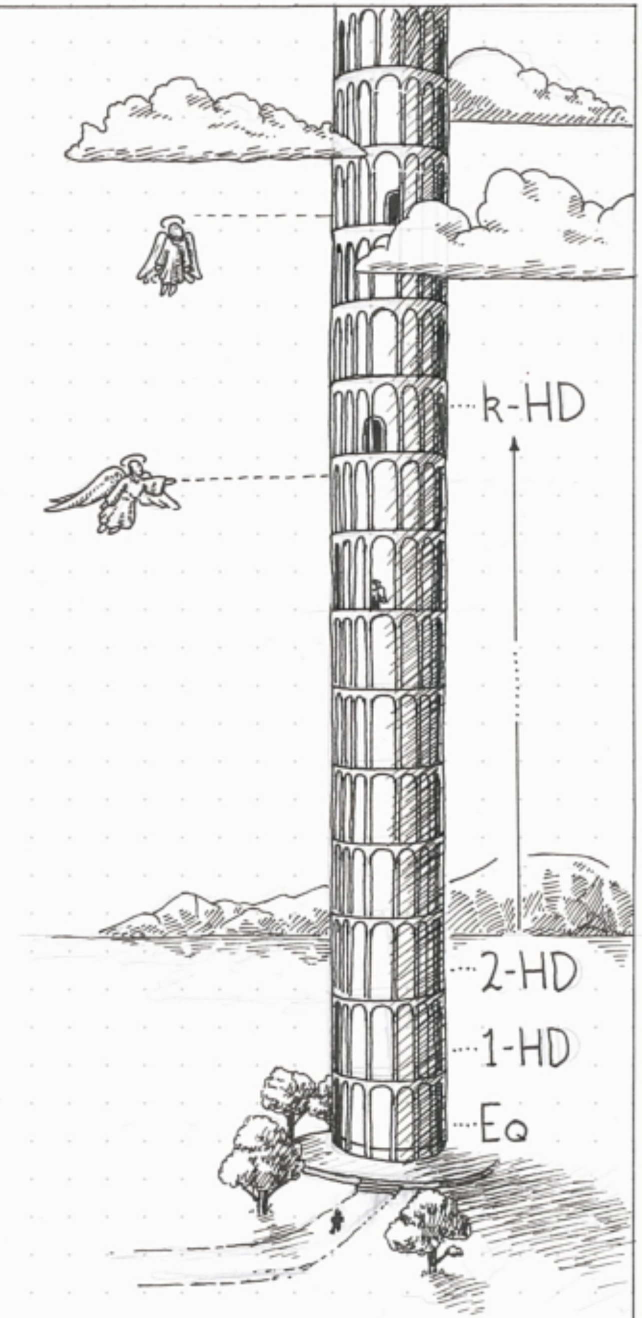
② k-HAMMING DISTANCE is permutation-invariant:



If the oracle queries are not permutation-invariant:

→ Hypergraph Ramsey theorem ⇒ stability is violated.

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



Future work

Testing vs. learning

distribution testing

confused collector

multiple collectors

random clusters (trees?)

?

small certificates?

parity trace

trace reconstruction*

k-alternating — ? → halfspaces

? → $\boxed{\text{VC-like theory}}$

Constant-cost communication

$\mathsf{BPP}^0 \not\subseteq \mathsf{UPP}^0$

model theory?

other types

dimension 4 → $\mathsf{UPP}^0 \cap \mathsf{BPP}^0 = \mathsf{EQ}$

stability

hereditary problems

$\boxed{\text{complete characterization?}}$

finitely-defined problems

rectangle size

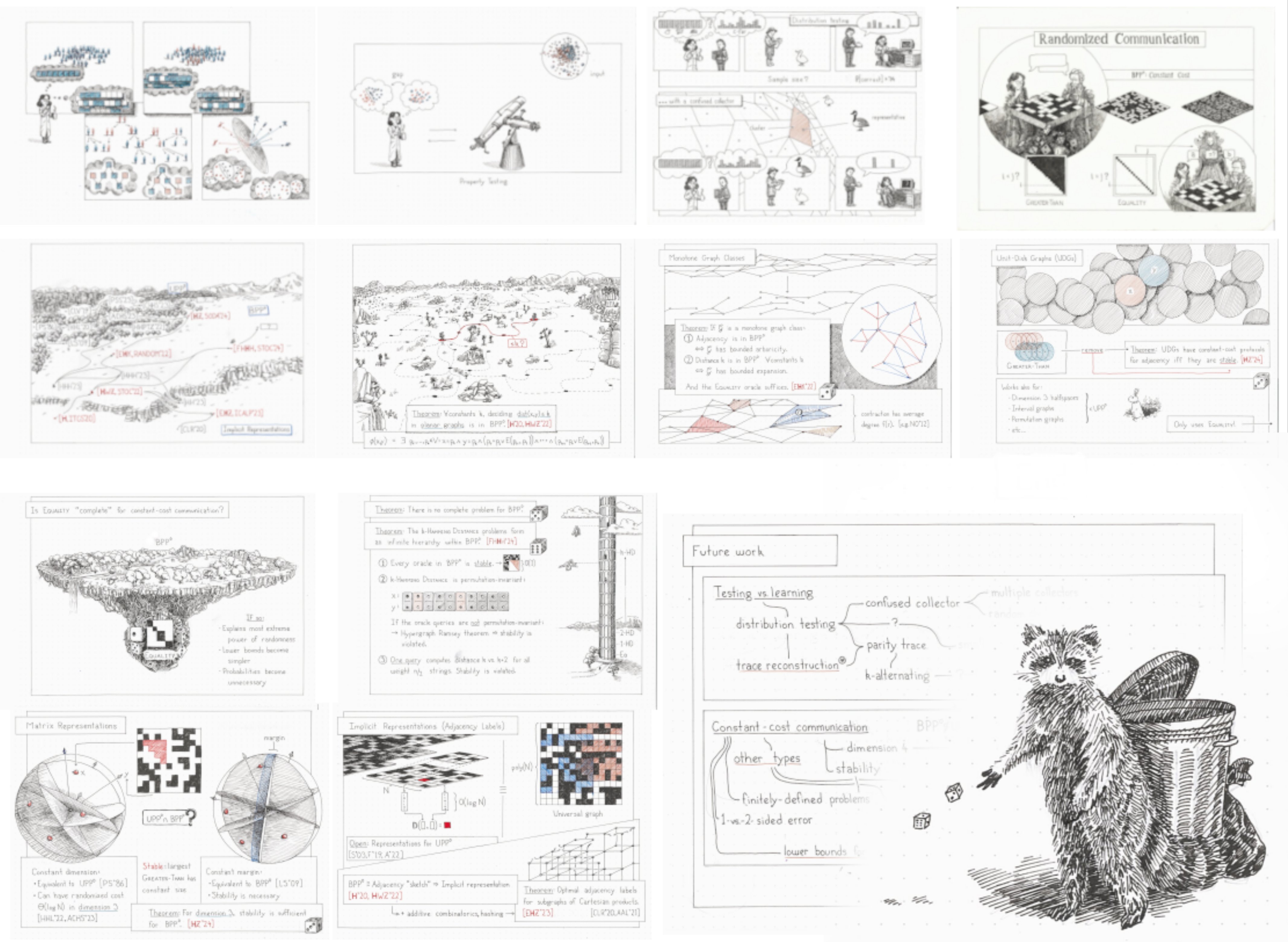
gap-hamming

insight into higher complexity?

randomized "log-rank"?

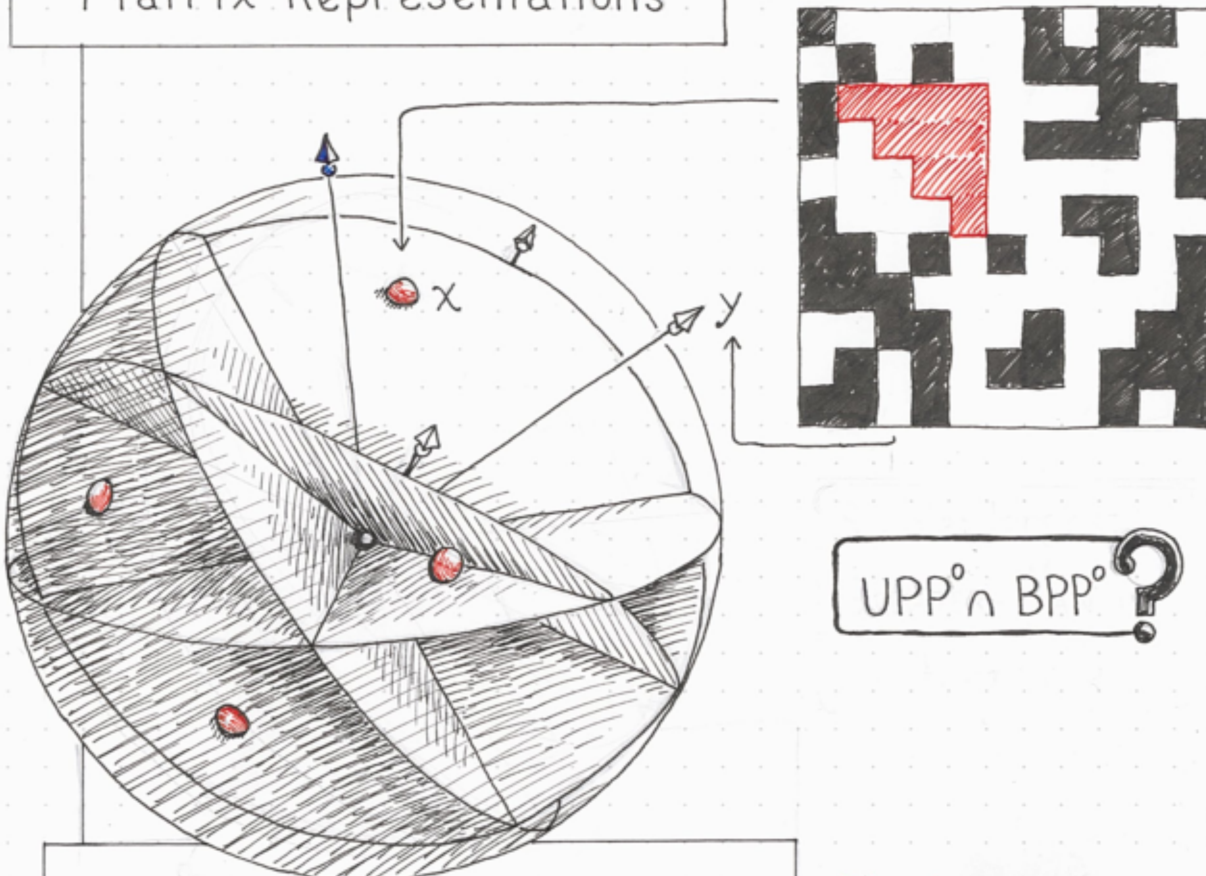
1-vs.-2-sided error

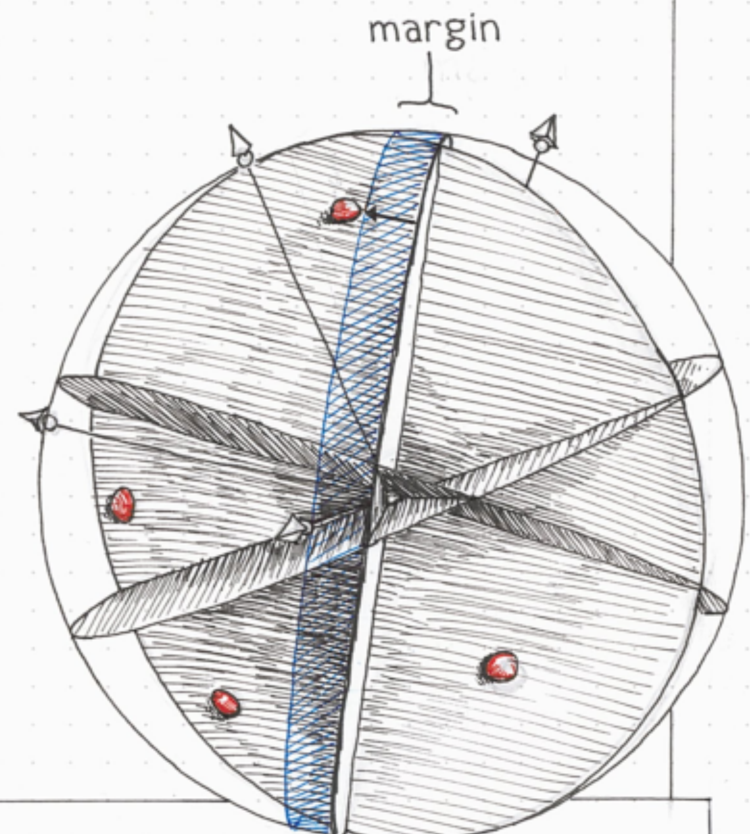
lower bounds for implicit representations



Property Testing
Randomized Communication
Equality
Monotone Graph Classes
Unit-Disk Graphs (UDGs)
Is Equality "complete" for constant-cost communication?
Matrix Representations
Implicit Representations (Adjacency Labels)
Universal graph
Future work
Testing vs learning
distribution testing
confused collector
multiple collectors
?
parity trace
trace reconstruction
k-alternating
Constant - cost communication
other types
dimension
stability
finitely-defined problems
1-vs-2- sided error
lower bounds

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]

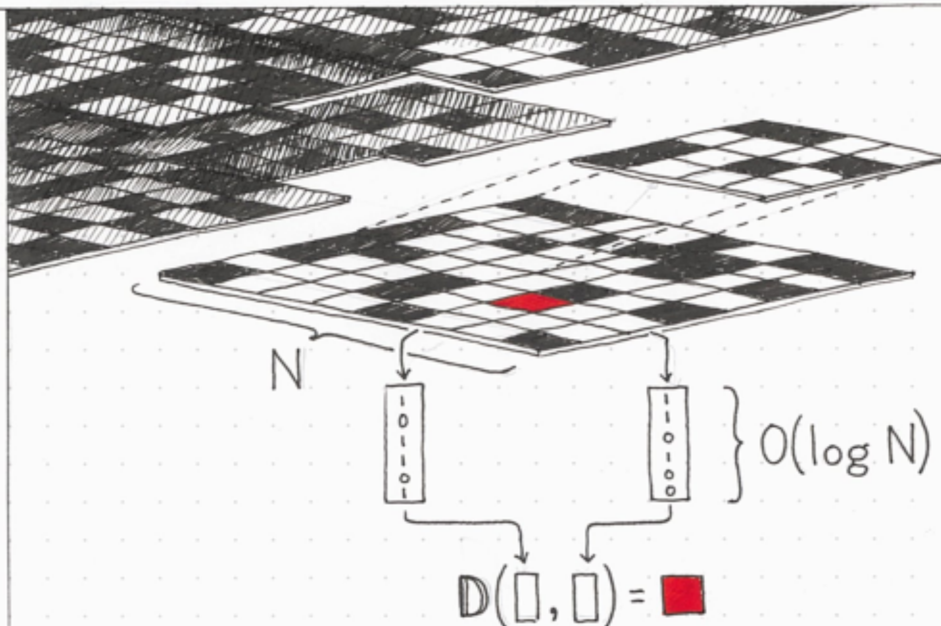
Stable: largest GREATER-THAN has constant size

Constant margin:

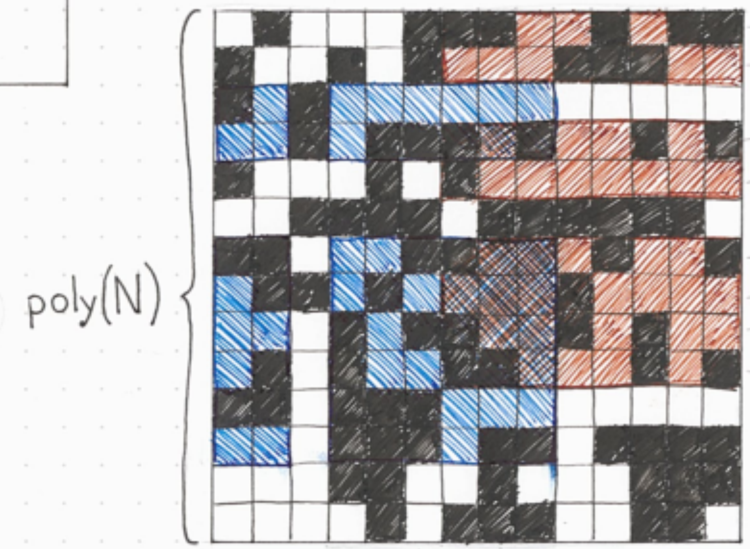
- Equivalent to BPP^0 [LS'09]
- Stability is necessary

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

Implicit Representations (Adjacency Labels)

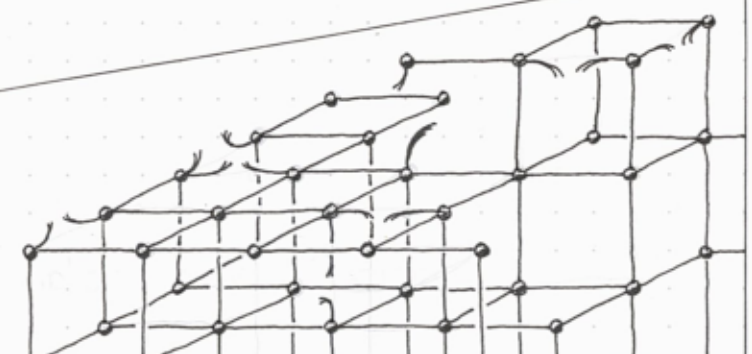


$\equiv$



Universal graph

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



$BPP^0 \equiv$ Adjacency "sketch" $\Rightarrow$ Implicit representation
[H'20, HWZ'22]

$\hookrightarrow$ + additive combinatorics, hashing $\rightarrow$

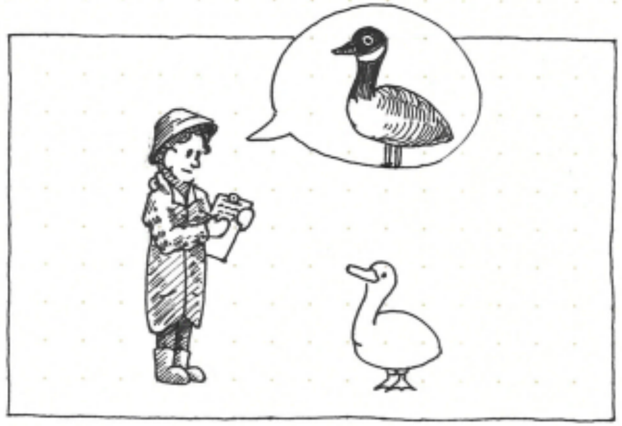
Theorem: Optimal adjacency labels for subgraphs of Cartesian products.
[EHZ'23] [CLR'20, AAL'21]

Part I: Adversarial Clusters

Impossible!

1. Restrict clustering to $\mathcal{U}$

2. Replace TV with earth-mover (EMD)
Domain → metric space

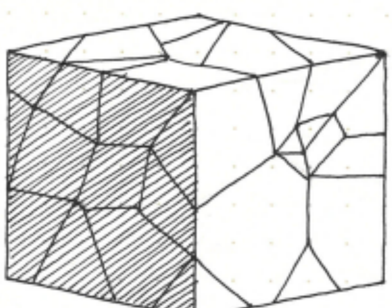


3. Queries

4. Cluster-Reject

$\mathcal{U}$: Universe

E.g.



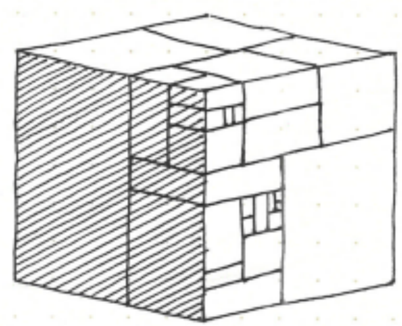
Convex

$G \subseteq \mathcal{U}$: "Good" clusterings

not allowed!

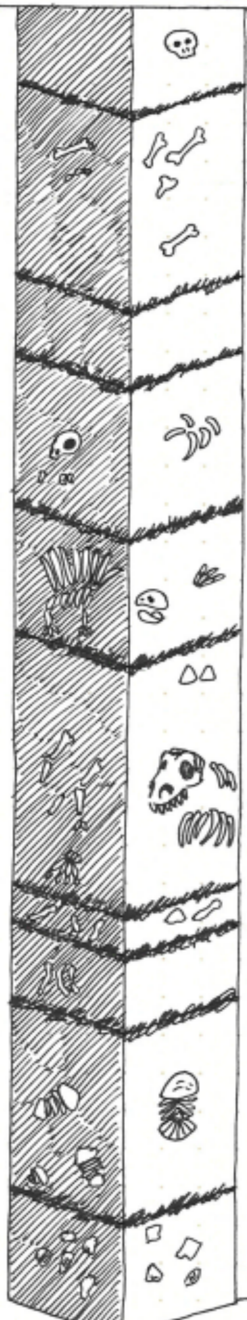
depends on distribution

High-probability of low diameter boxes, decision trees



Result: learning cells is not necessary

Part II: Random Clusters



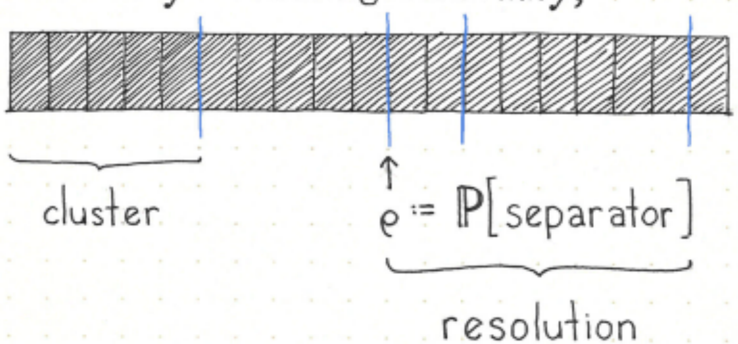
Motivations:

- Environmental randomness
- Randomized classifier training

Naïve (known clusters):

$$O\left(\frac{\sqrt{n}}{\rho^{3/2}\,\varepsilon^{2}}\right)$$

We study: Testing uniformity,



Without queries, $\left(\rho \geq \tilde{\Omega}\left(n^{-1/5}\varepsilon^{-4/5}\right)\right)$

$$\tilde{O}\left(\frac{\sqrt{n}}{\rho^{3/2}\,\varepsilon^{2}}\right)$$

With queries

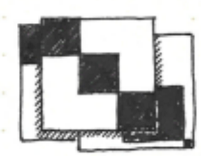
$$O\left(\frac{\sqrt{n}}{\rho\,\varepsilon^{2}}\right)$$ Use [VV'17]

Standard:

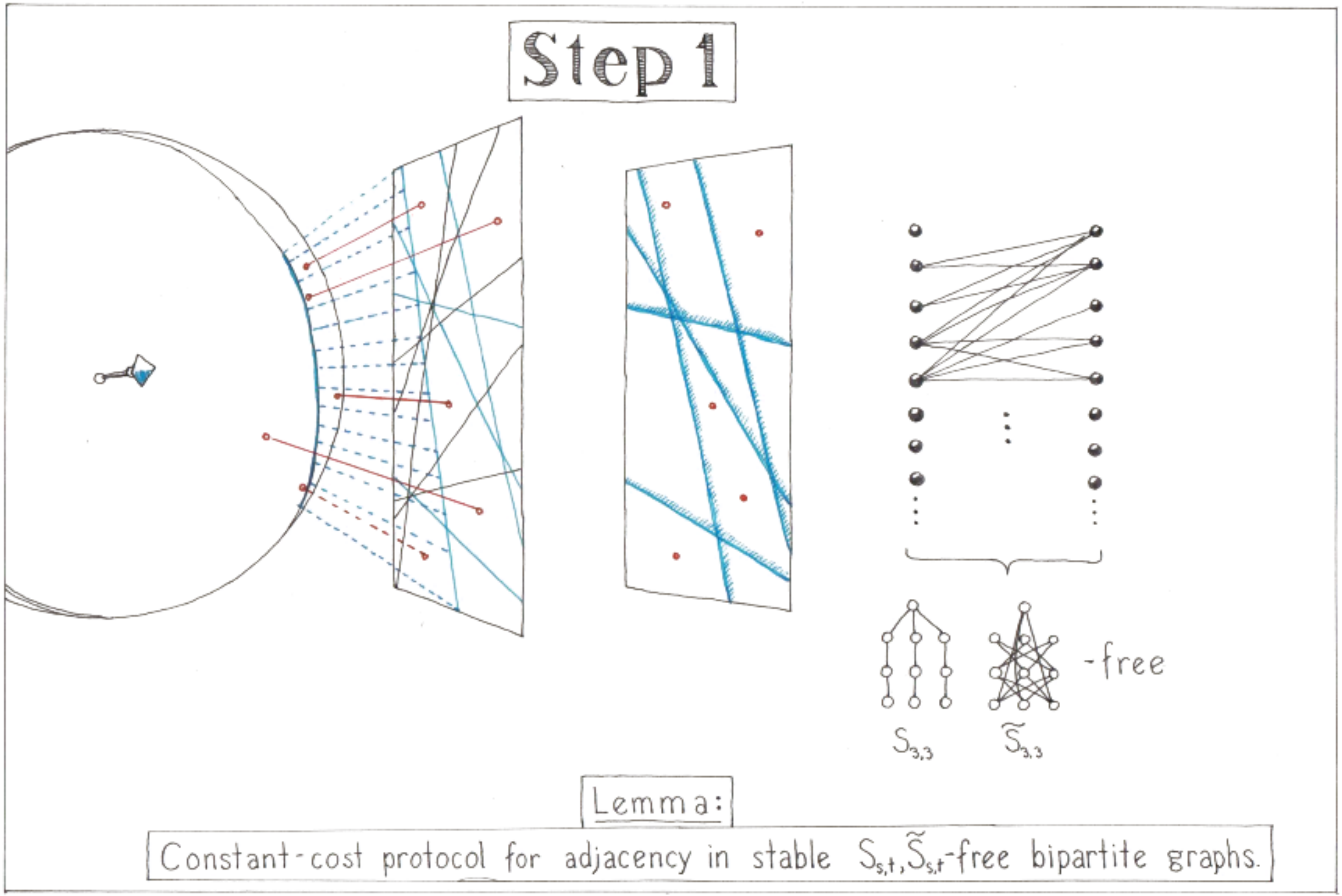
X: histogram 

Analyze $X^{T} I X - \|X\|_1$

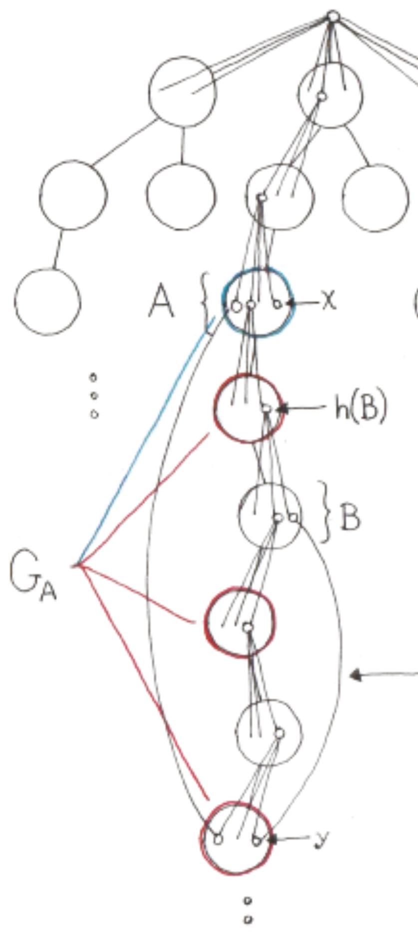
Now:



Analyze $X^{T} \Phi X - \|X\|_1$


Step 1
-free
$S_{3,3}$
$\widetilde{S}_{3,3}$
Lemma:
Constant-cost protocol for adjacency in stable $S_{s,t}, \widetilde{S}_{s,t}$-free bipartite graphs.

Step 2



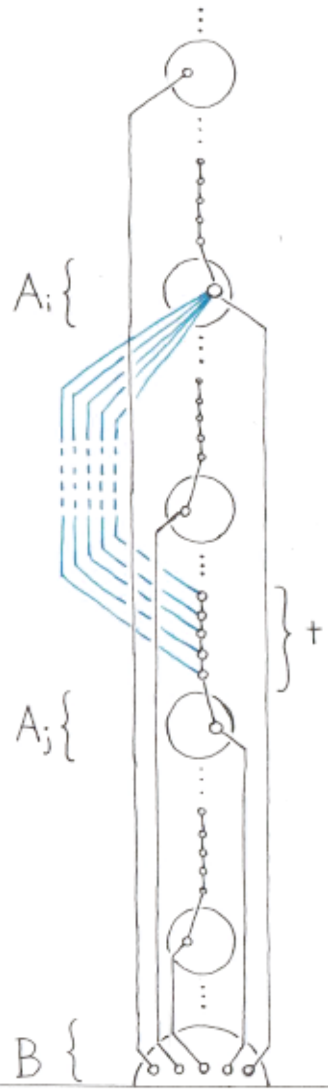
hook: • adjacent to all $v \in B$
• adjacent to nothing below B

All edges are to ancestors or descendants

Gyárfás decomposition [POS'22]

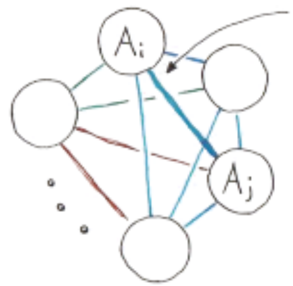
Players agree on ancestor bag A
⇒ "Recurse on G_A or $\widetilde{G}_A$ with smaller GREATER-THAN."

- Stable ⇒ $O(1)$ recursions
- $S_{s,t}$, $\widetilde{S}_{s,t}$-free ⇒ always $S_{s,t}$-free



Step 3

<u>Lemma</u>: Stable, $S_{s,t}$-free $\Rightarrow$ each bag has edges to $O(1)$ ancestor bags.



Color: induced subgraph encoding

Ramsey's theorem + case analysis on each color: find large GREATER-THAN, or $S_{s,t}$.

Players need $O(1)$ EQUALITY queries to agree on the ancestor, or output "non-adjacent."