The long path to $\sqrt{d}$ monotonicity testers

C. Seshadhri (Seshadhri Comandur)

Dept. of Computer Science & Engg
UC Santa Cruz

Amazon Scholar
AWS

Thanks to my teachers

Manindra Agrawal

Bernard Chazelle

Mike Saks

Tamara Kolda

Ali Pinar

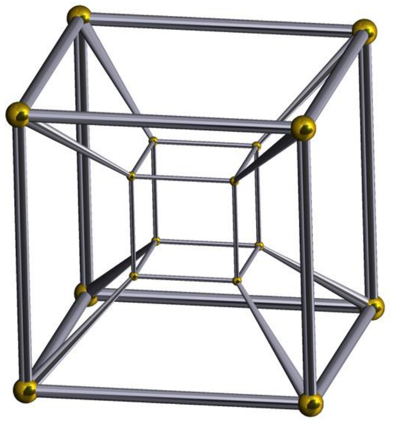
Thanks to my collaborators

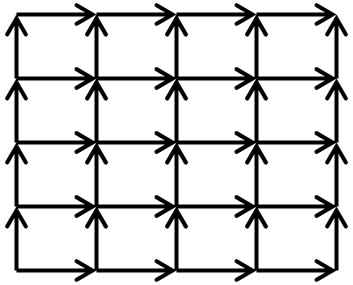
Hadley Black

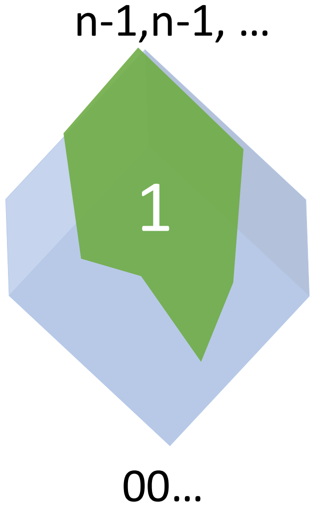
Deeparnab Chakrabarty

Monotone hypergrid functions

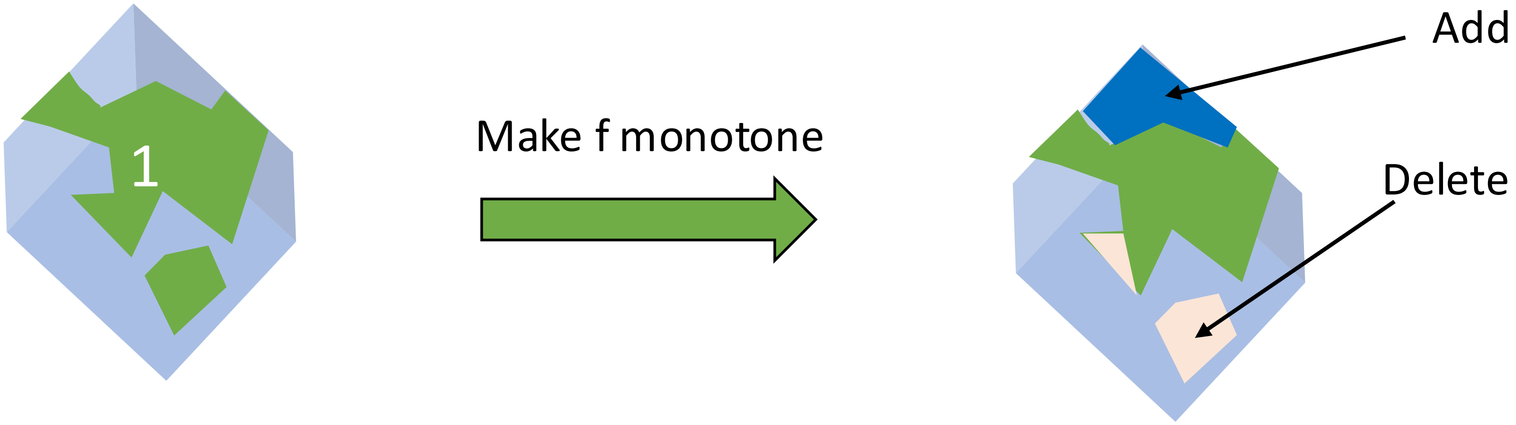
- $f: D \to \{0,1\}$
- **Hypergrid:** $[n]^d$
  (includes line, tesseract, etc.)
- **n = 2, Boolean Hypercube**: $\{0,1\}^d$
- $x \prec y$ iff $x_i \leq y_i$ for all i
- f is monotone: if x ≺ y, f(x) ≤ f(y)
- Equivalently, think of f as indicator for set







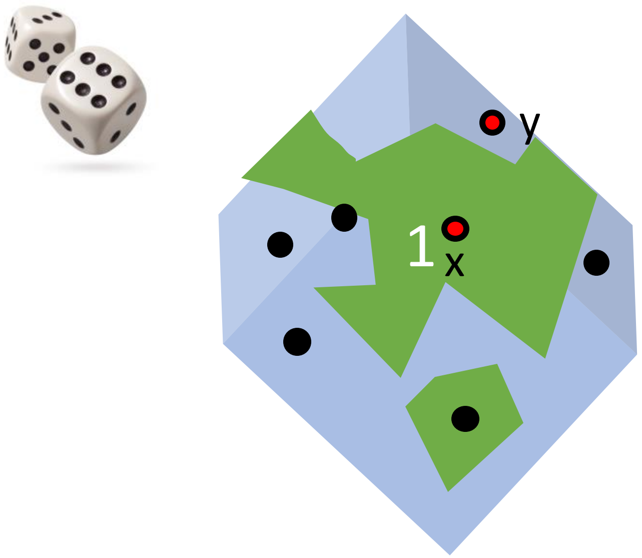
The distance to monotonicity



- Distance to monotonicity = (min changes to make set monotone)/n^d
- ε_f in [0,1)
  - Amen

- $\varepsilon_f = \min_{g \text{ monotone}} |f - g|_0 / n^d$

Monotonicity testing



$x \prec y$

$f(x) = 1$

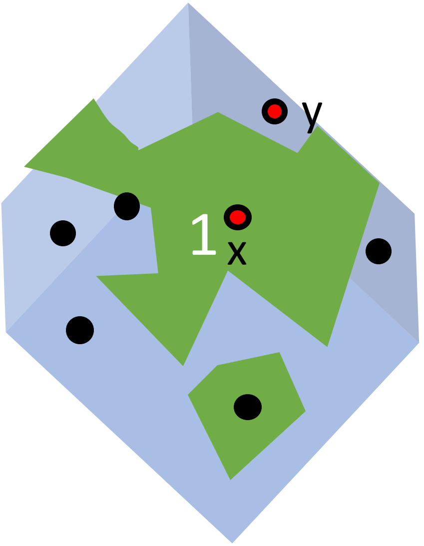
$f(y) = 0$

- [Bshouty-Tamon96, Lange-Vasilyan23] Learning monotone functions needs $> \exp(\sqrt{d})$
- Given ε and query access to f:

  Distinguish monotone ($\varepsilon_f = 0$) vs far from monotone ($\varepsilon_f > \varepsilon$) whp

- One-sided tester: given f such that $\varepsilon_f > \varepsilon$, discover a “violation” whp
- Non-adaptive: all queries are made in advance

Big question



$x \prec y$

$f(x) = 1$

$f(y) = 0$

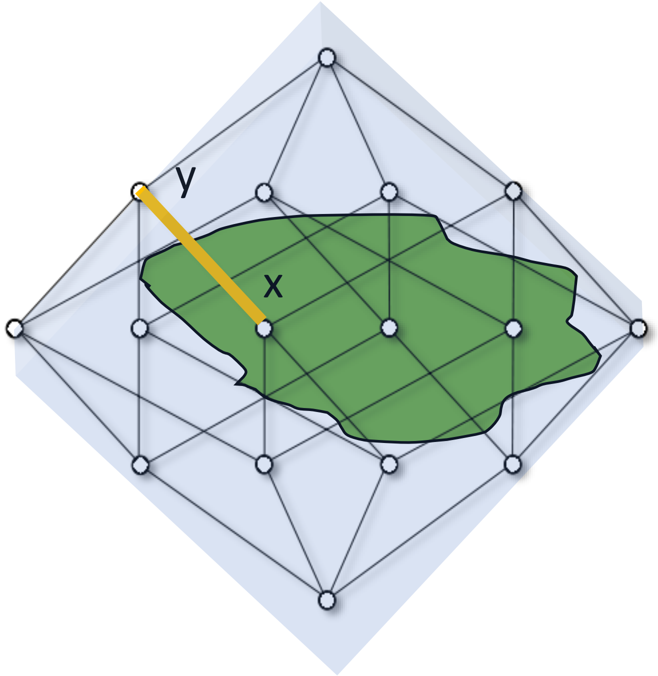
- [Goldreich-Goldwasser-Lehman-Ron-Samorodnitsky 00, Raskhonikova 99, Dodis-Goldreich-Lehman-Raskhodnikova-Ron-Samorodnitsky 99]

What is the (non-adaptive) complexity of monotonicity testing?

- Hypergrid domain, Boolean range

- > 20 papers and two decades of history

The Edge Tester



- [GGLRS00, DGGLRS00]
- Sample an edge of hypercube (x, y) u.a.r
- Query $f(x), f(y)$
- Reject if violation

Theorem: the probability of finding violation is > ε_f /d

So if $\varepsilon_f > \varepsilon$, $O(d/\varepsilon)$ samples suffice to detect violation whp

- [DGGLRS00] $O(d \log^2 d /\varepsilon)$ monotonicity tester for hypergrids

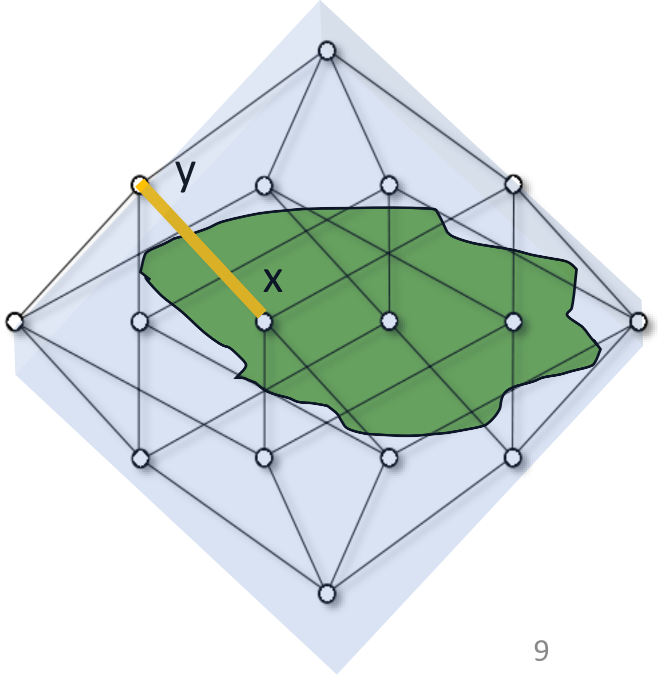
Can we beat d?

- [Blais-Brody-Matulef 12] $\Omega(d)$ lower bound when range is real
- [Chakrabarty-S 13a, Chakrabarty-S 13b]

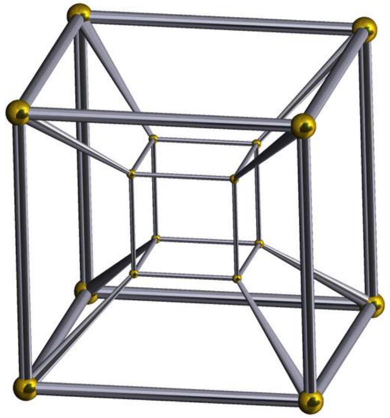
  Hypergrid domain, arbitrary range, the complexity is $\Theta(d \log n / \varepsilon)$
- Non-adaptive, one-sided upper bound. Adaptive, two-sided lower bound
  - Finis. End of story
  - Optimal tester for hypercube is edge tester

- Does Boolean range make complexity lower?
  - [Fischer-Lehman-Newman-Raskhodnikova-Rubinfeld 02]

  $\Omega(\sqrt{d})$ one-sided lower bound
  - Can one get o(d) complexities?



The mystery of root d, for hypercubes



- [Chakrabarty-S 13] $d^{7/8}$ query tester
- [Chen-Servedio-Tan 14] $d^{5/6}$ query tester

- [Khot-Minzer-Safra 15] $\sqrt{d}$ query tester
  - Essentially, same tester as before. All in the analysis

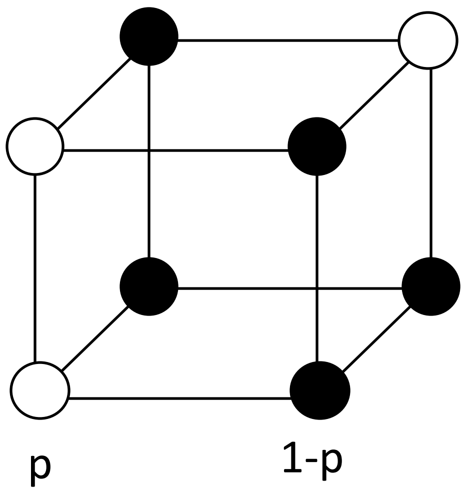
- [Chen-De-Servedio-Tan 15, Chen-Waingarten-Xie 17] $\Omega(\sqrt{d})$ two-sided non-adaptive lower bound

- The theory of Directed Isoperimetric Theorems
  - Seemingly hypercube specific; can they generalize?

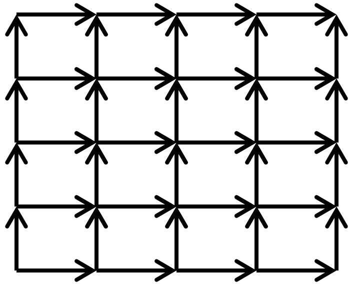
Getting to root d, for hypergrids

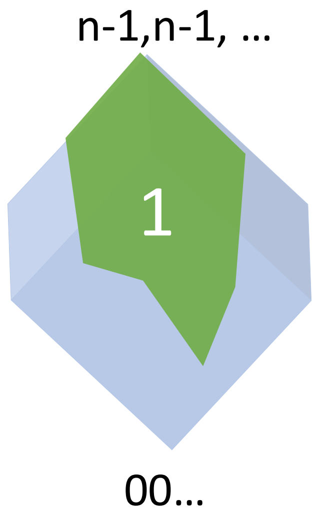
- [DGGLRS00] O(d $\log^2$d /ε) monotonicity tester for hypergrid
- [Berman-Raskhodnikova-Yaroslavtsev 17] O(d log d)
- [Black-Chakrabarty-S 18, BCS 20] $d^{5/6}$ query tester
- [Braverman-Khot-Kindler-Minzer 23, BCS 23] $poly(n)\sqrt{d}$ query tester
- [BCS 24] $d^{1/2+o(1)}$ query tester
  - All results have different testers!

The generality of hypergrids



- **Hypergrid:** $[n]^d$
  Results for uniform distribution automatically translate to any product distribution
  - Otherwise uniform distribution on hypercube looks like special case





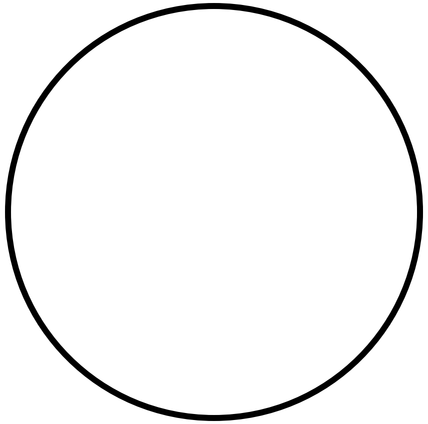
- Can even set n = ∞, so domain is R^d
  - Monotonicity testing for measurable functions over product distributions

Why did the baby grid not get any sugar?

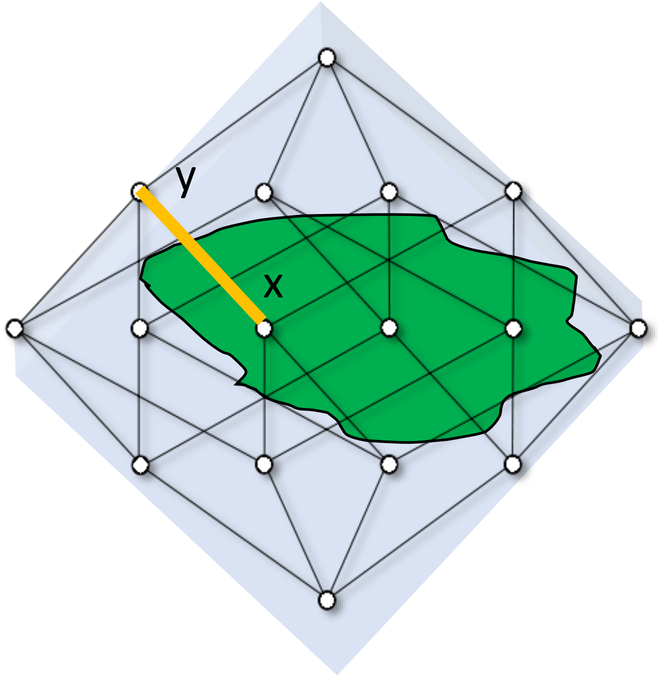
Because it would become a hypergrid.

Directed Isoperimetry

Surfaces, volumes, and why o(d) is possible



The Edge Tester



- [GGLRS00, DGLRRS99]
- Sample an edge of hypercube (x, y) u.a.r
- Query $f(x), f(y)$
- Reject if violation

Theorem: the probability of finding violation is $> \varepsilon_f / d$

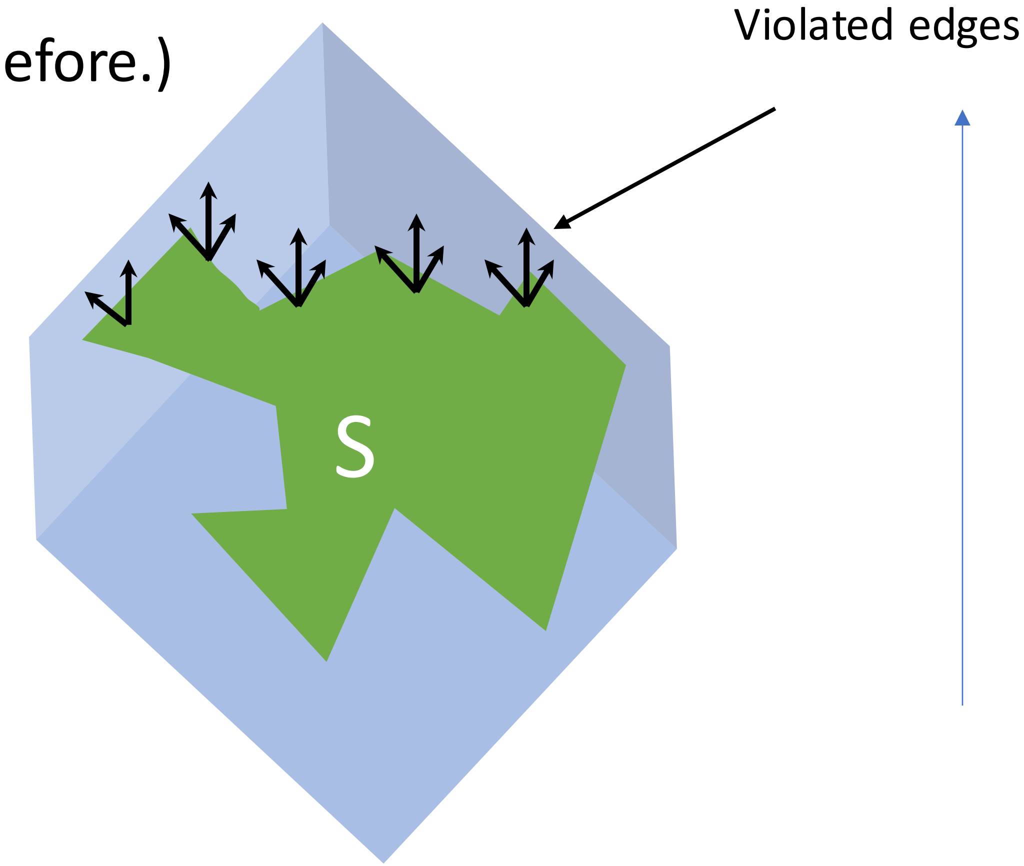
A Geometric Interpretation

Abusing notation, S is the set of 1s. (We still use ε_f as before.)

Theorem: The number of **violated** hypercube edges is = $\Omega(\varepsilon_f \cdot 2^d)$

$$\mathrm{Inf}_f^+ = \frac{\left|\vec{E}(S, S^c)\right|}{2^d}$$

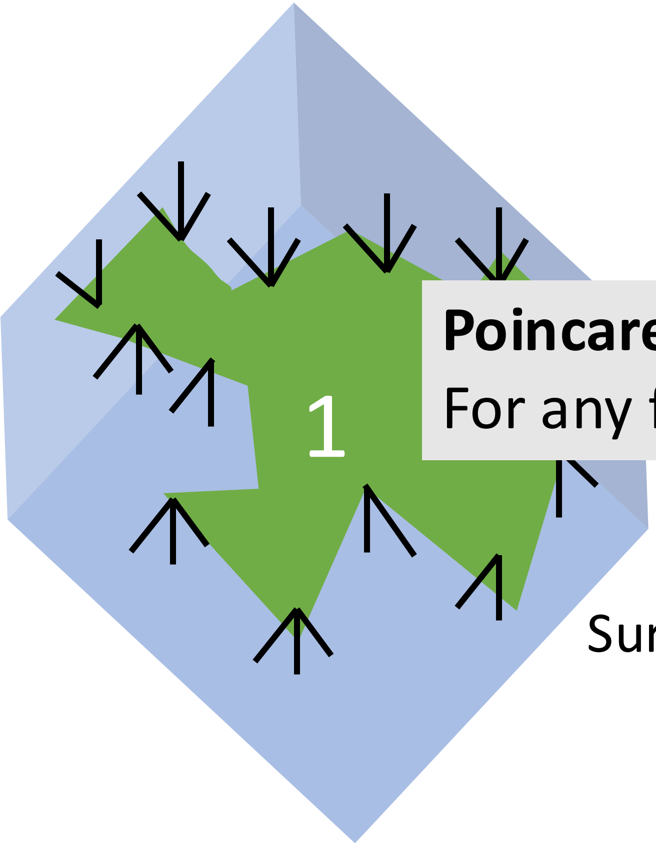
Theorem: $\mathrm{Inf}_f^+ = \Omega(\varepsilon_f)$

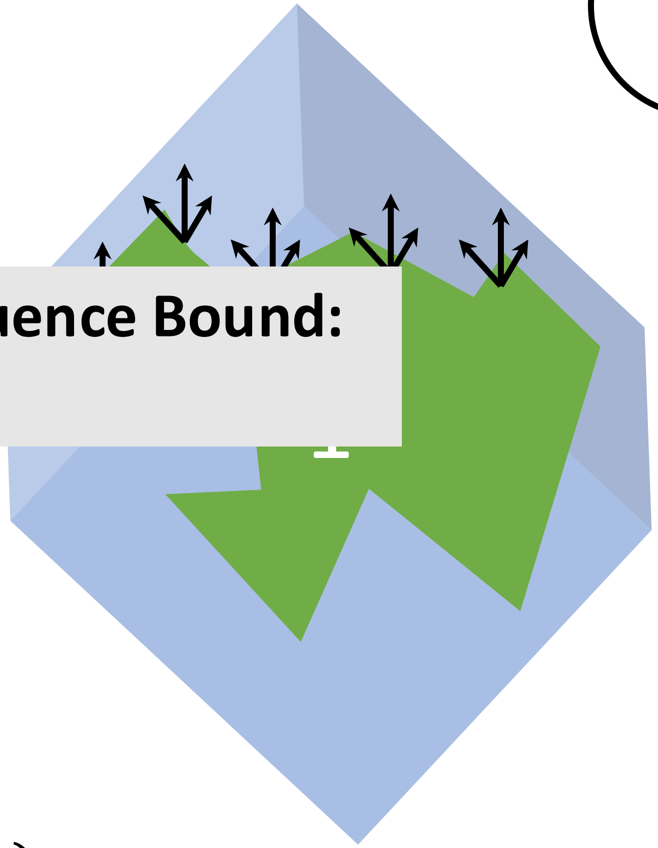


A far from monotone function

Undirected Isoperimetry: Poincare

$$\mathrm{Inf}_f = \frac{|E(S,S^c)|}{2^d}$$





Poincare Inequality/Edge Expansion/Influence Bound:

For any f, $\mathrm{Inf}_f = \Omega(var_f)$

Surface area > (Function of) Volume

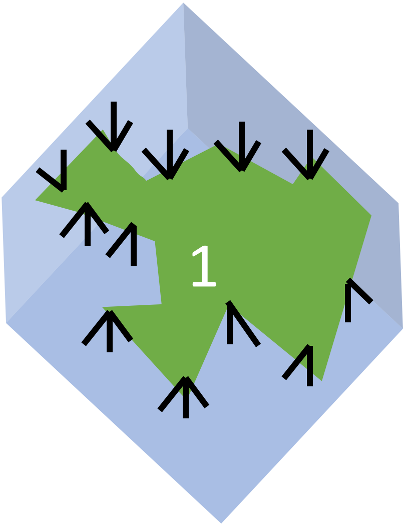
$$var_f = \frac{\#1's}{2^d} \cdot \frac{\#0's}{2^d} = \frac{|S|}{2^d} \cdot (1 - \frac{|S|}{2^d})$$

Influential Edges

GGLRS = Directed Poincare

Undirected Hypercube

- **Poincare:** $\mathrm{Inf}_f = \Omega(var_f)$

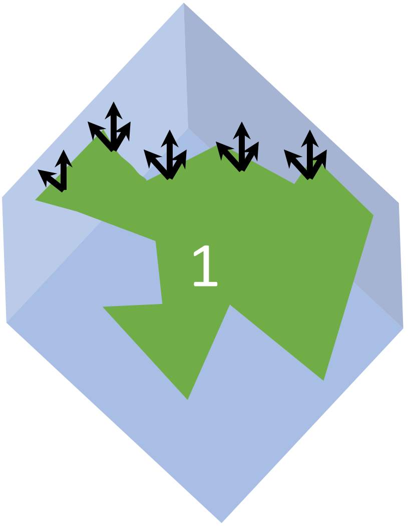


Directed Hypercube

- **GGLRS:** $\mathrm{Inf}_f^+ = \Omega(\varepsilon_f)$

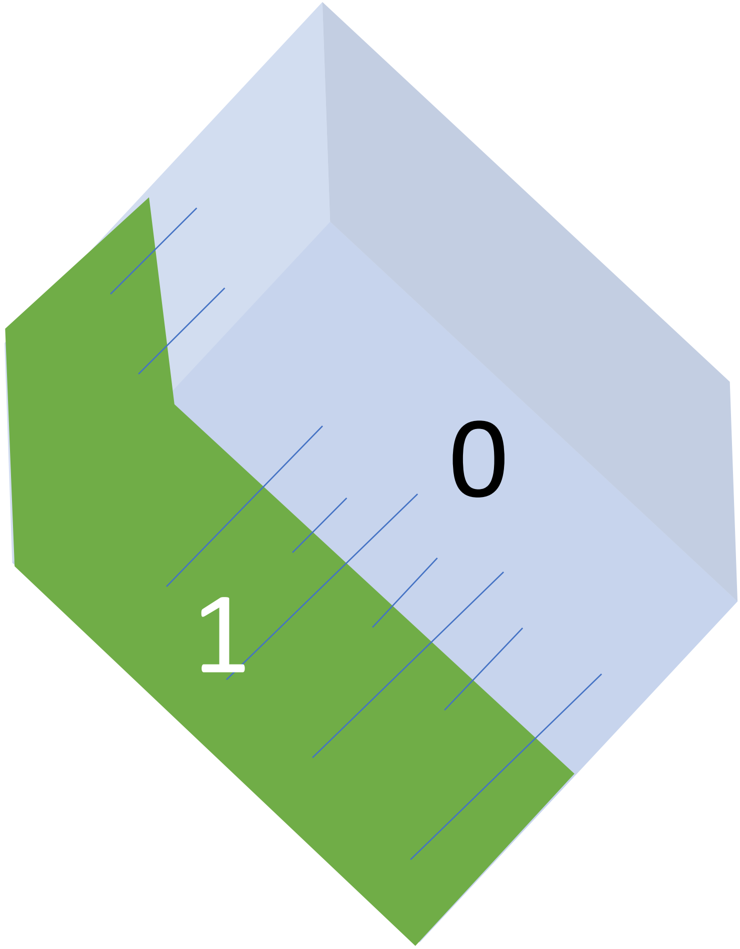
Directed expansion/
surface area

Directed volume…?



(Directed) Poincare is tight

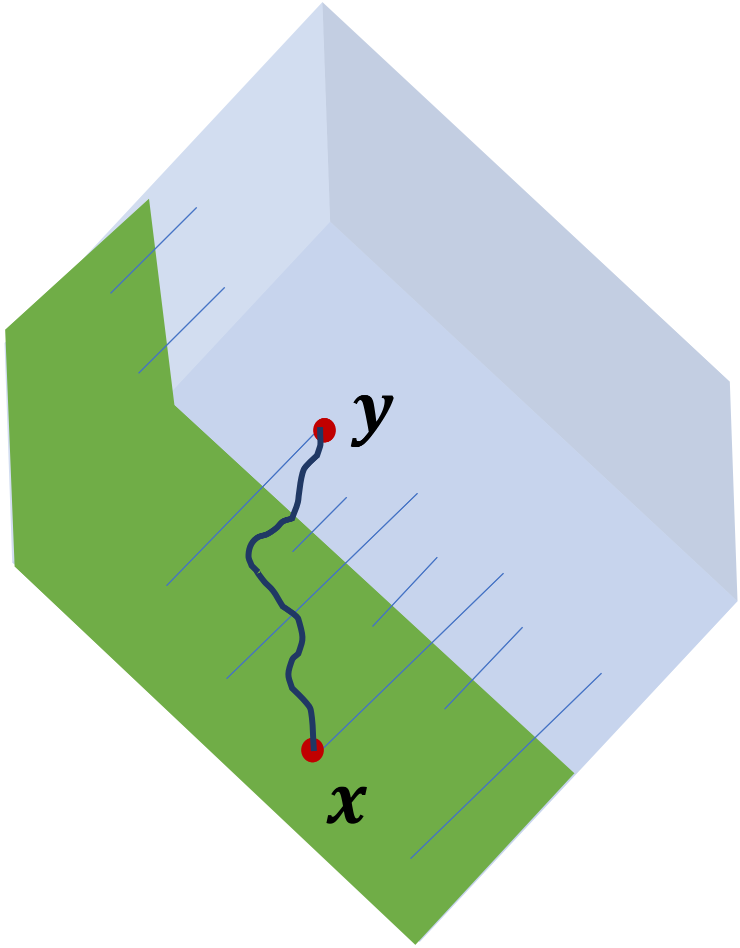
- Distance = $\frac{1}{2}$
- Number of viol edges = 2^{d-1}
    - Total edges = $d \cdot 2^{d-1}$
- Edge tester needs $\Omega(d)$ queries to catch violation



Bypassing the (anti)-Dictator

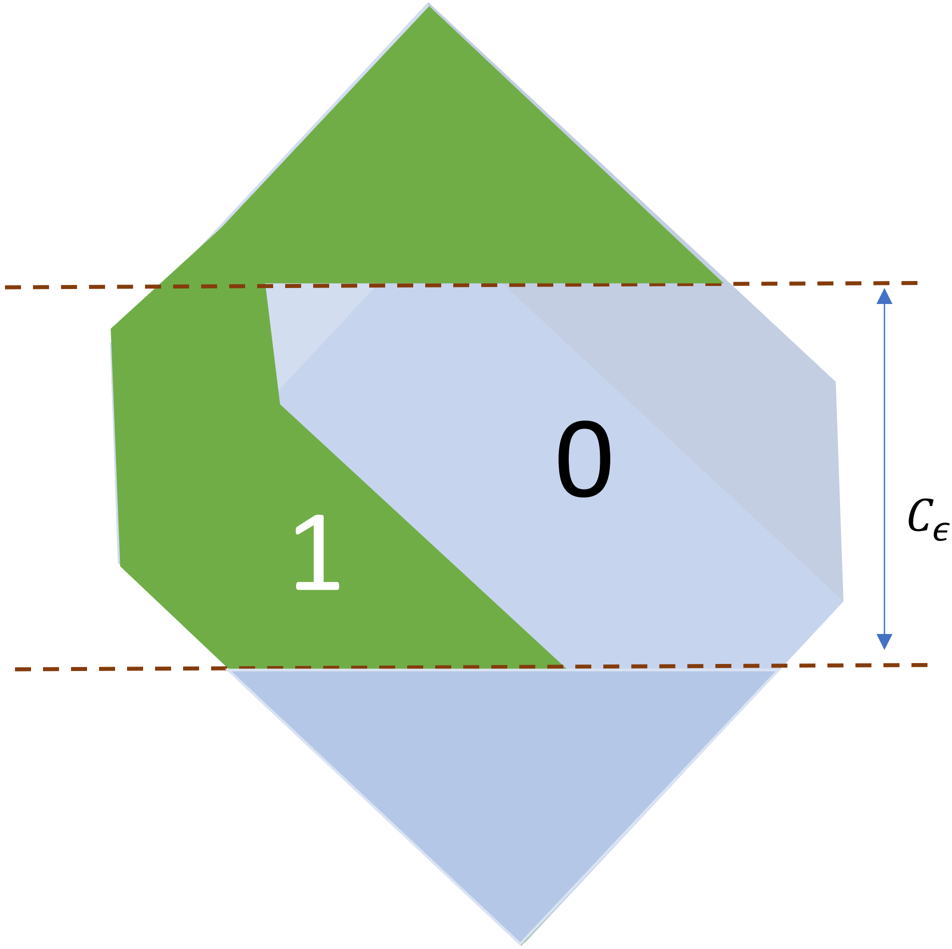
- Sample x
- Walk “up” to get y
- Query $f(x), f(y)$ and Test.

How long should we walk?



Square Root of Dimension

- Sample x
- Walk "up" to get y
- Query $f(x), f(y)$ and Test.



Can't walk more than $\approx \sqrt{d}$

Analysis for the (capped) Anti-Di

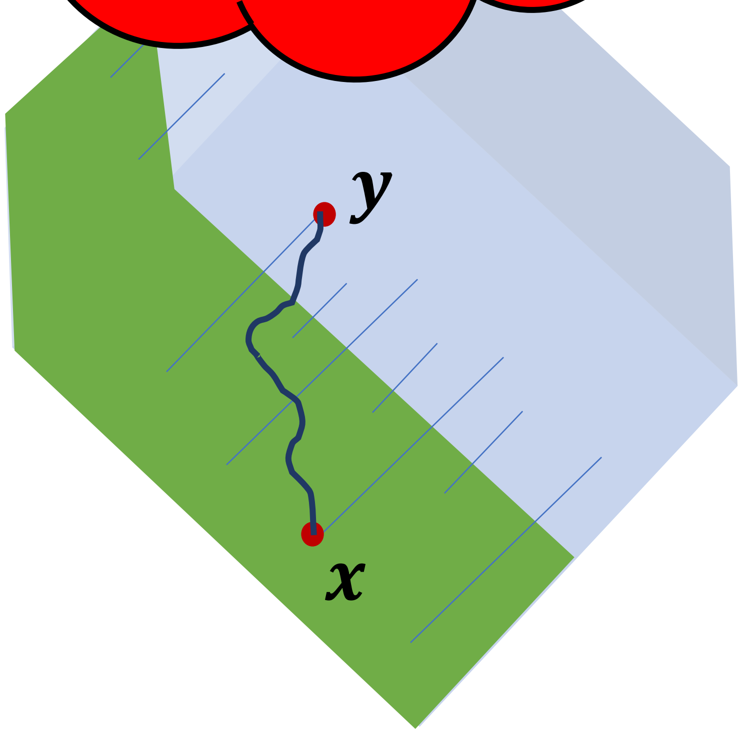
The root to root d is taking a longer root.

- Sample x
- Walk “up” to get y
- Query $f(x), f(y)$ and Test.

- Prob[x is green] $\approx 1/2$
- Prob[crossing] $\approx \sqrt{d} \cdot \frac{1}{d} \approx \frac{1}{\sqrt{d}}$

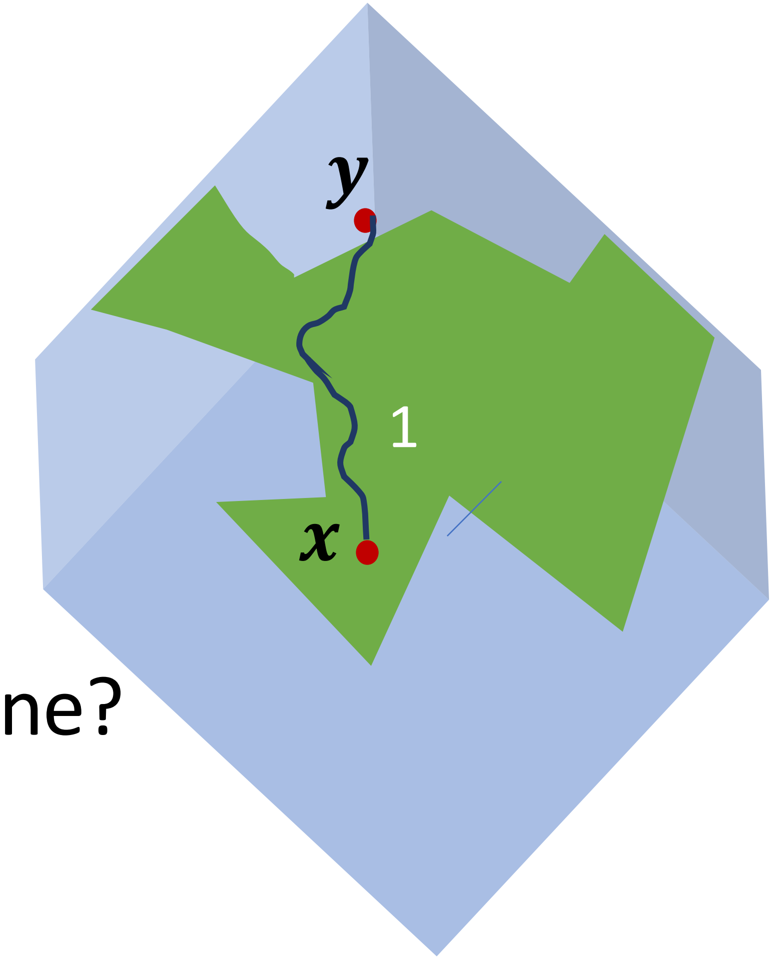
Number of steps

Chance of crossing in each step



How to analyze the "path tester"

Need structural insight
"far from monotone" functions/sets



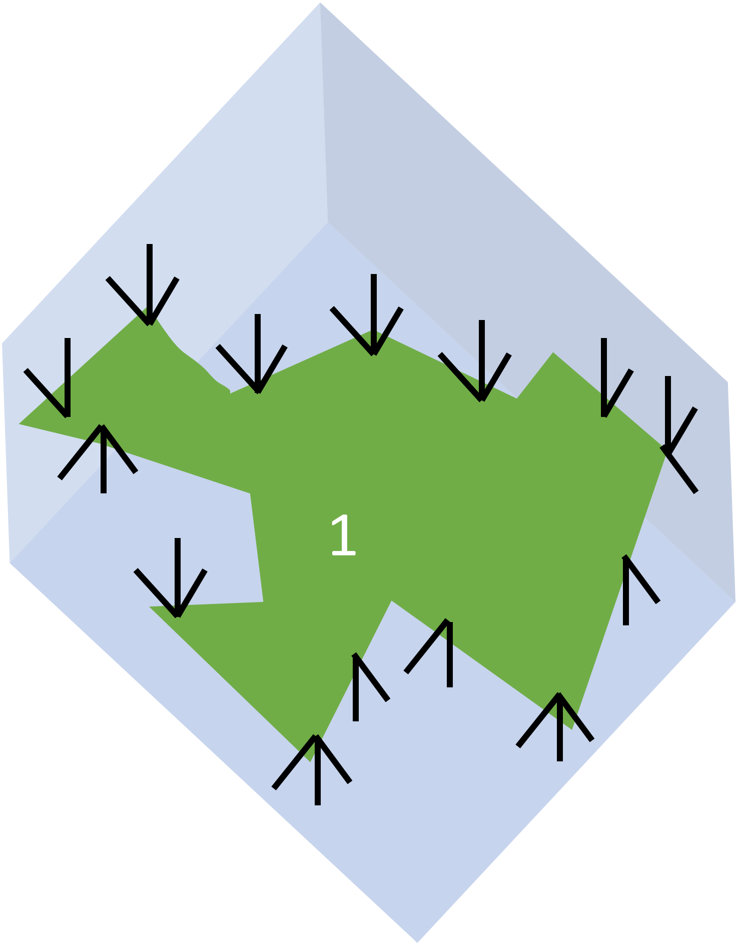
How to "escape" S that is far from monotone?

Directed Isoperimetric Theorems

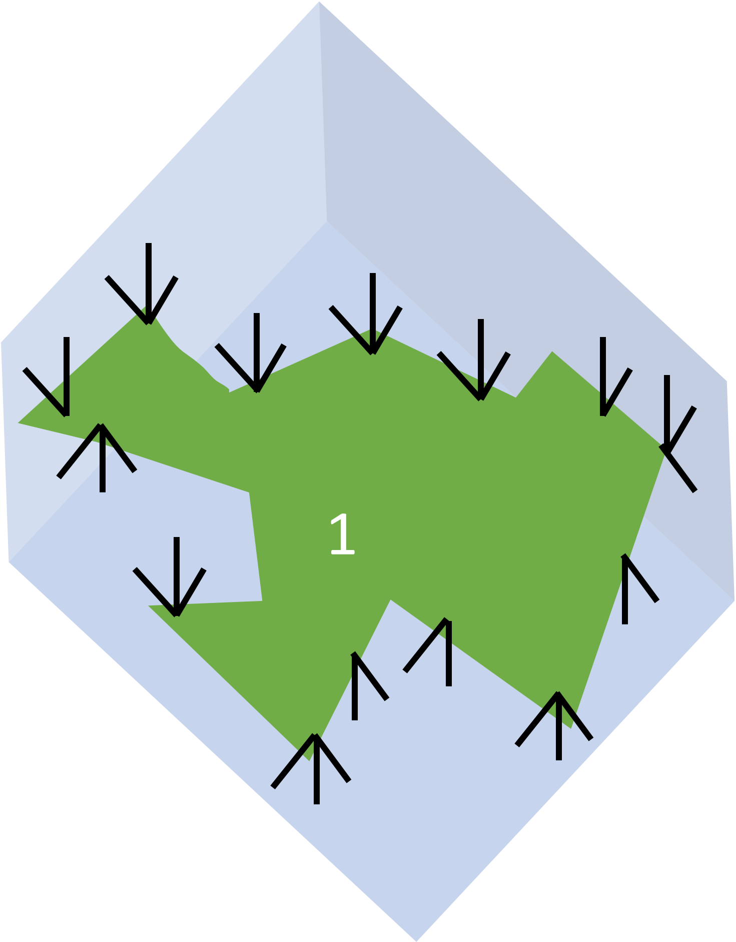
Undirected Hypercube	Directed Hypercube	Query
• Poincare: $\mathrm{Inf}_f = \Omega(var_f)$	• [DGLRRS00, GGLRS00]: $\mathrm{Inf}_f^+ = \Omega(\varepsilon_f)$	d
• [Margulis74]: $\mathrm{Inf}_f \cdot \Gamma_f = \Omega(var_f^2)$	• [Chakrabarty-S 13]: $\mathrm{Inf}_f^+ \cdot \Gamma_f^+ = \Omega(\varepsilon_f^2)$	$d^{5/6}$
• [Talagrand 92]: $\mathrm{E}[\sqrt{\mathrm{Inf}(x)}] = \Omega(var_f)$	• [Khot-Minzer-Safra 15]: $\mathrm{E}[\sqrt{\mathrm{Inf}^+(x)}] = \Omega(\varepsilon_f)$	$\sqrt{d}$

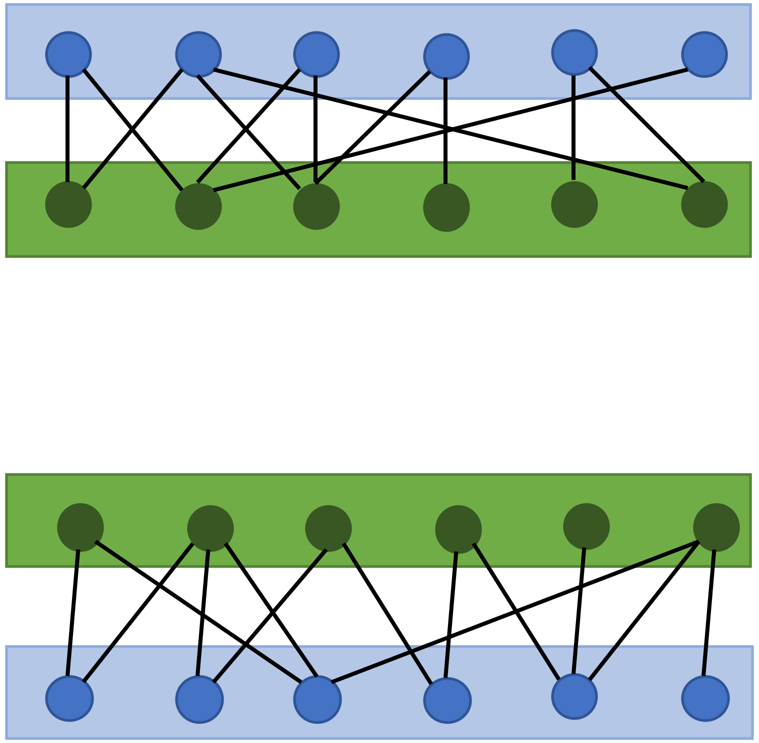
How to measure the boundary?

Boundary as bipartite graph

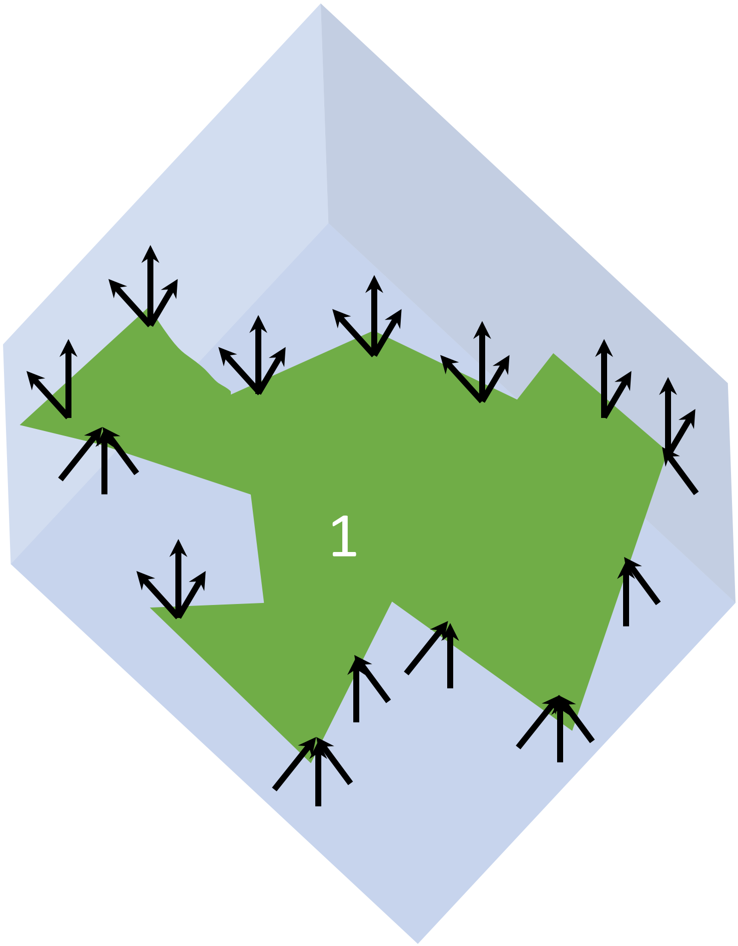


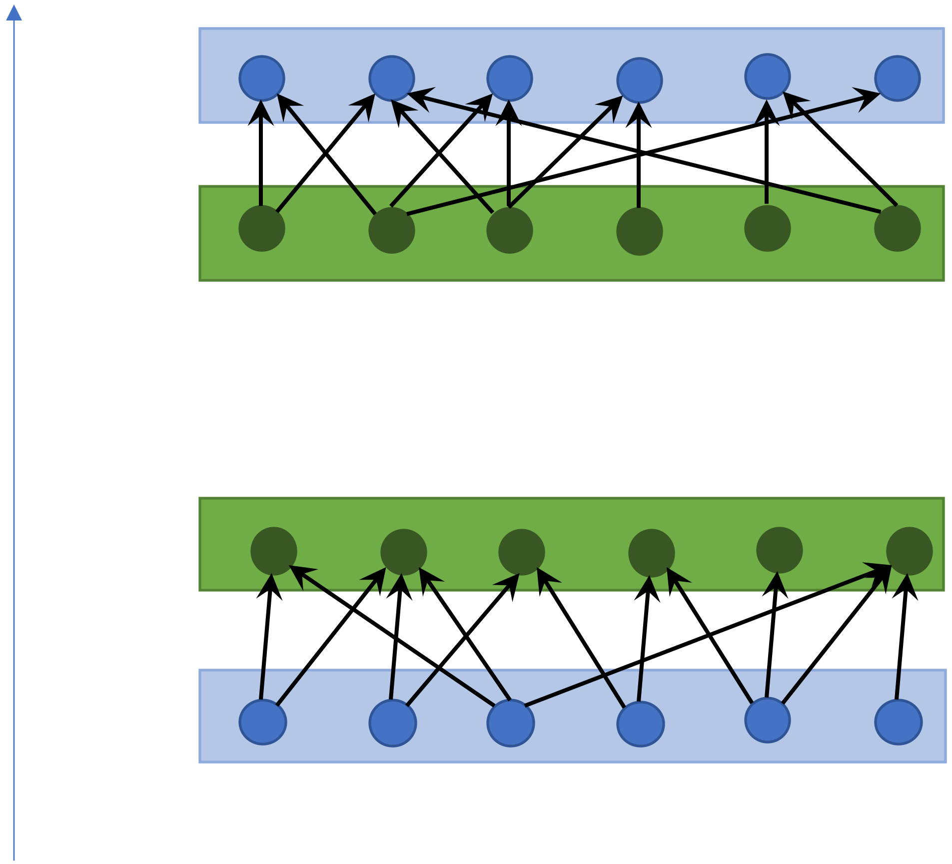
Boundary as bipartite graph



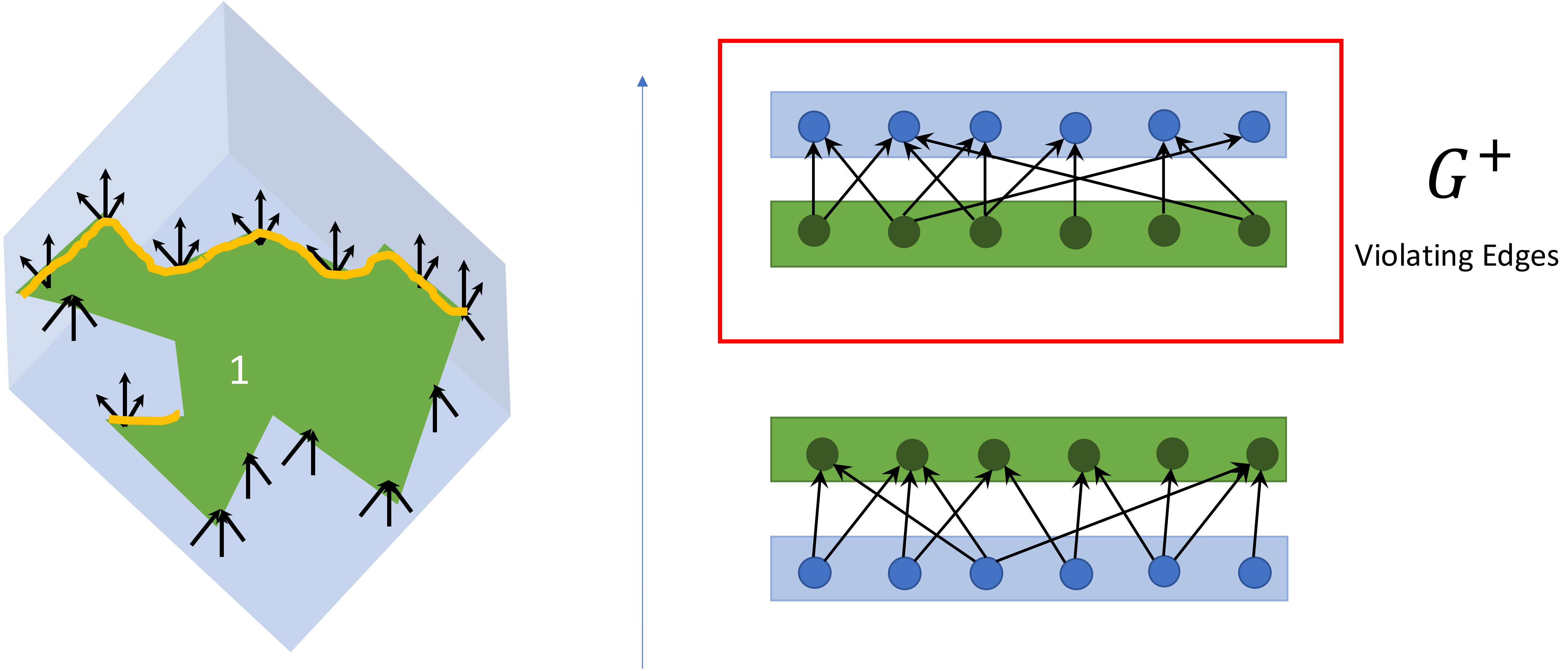


Directed boundary

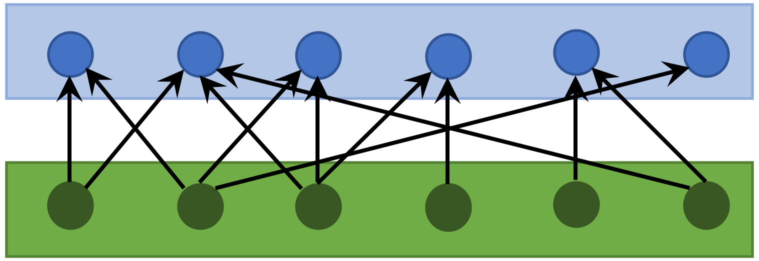




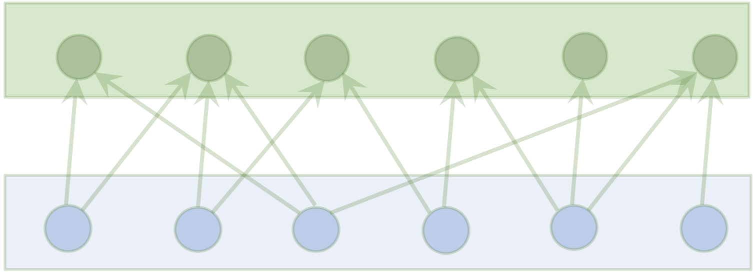
Directed boundary



Directed Influence



$\mathrm{Inf}^+(x) =$ out-degree of x in <u>directed</u> G^+

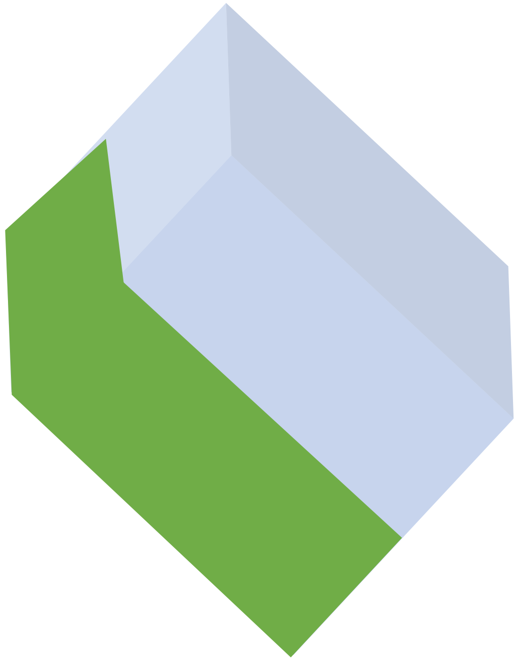


$$\mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}^+(x)] = \frac{E^+(S, S^c)}{2^{\mathrm{d}}} = I_f{}^+$$

$$\mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)] = I_f$$

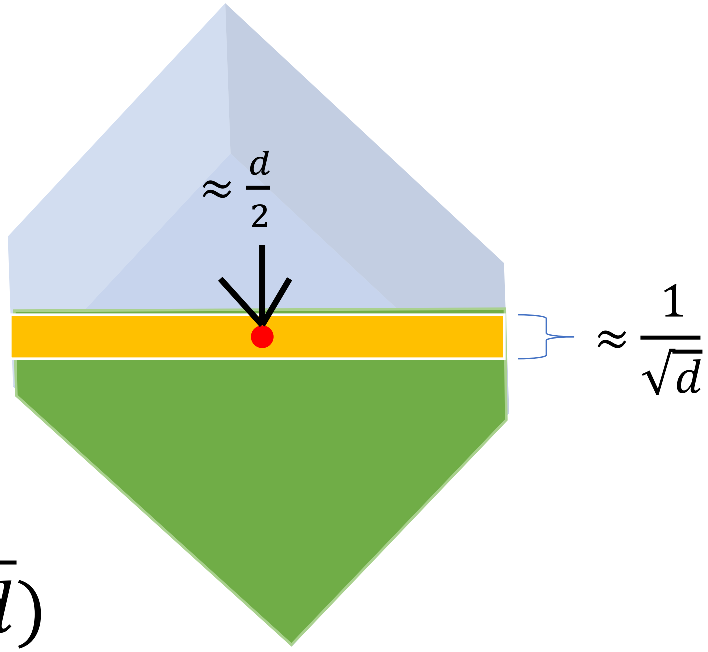
A careful look into influence

Poincare: $I_f := \mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)] = \Omega(var_f)$



$I_f = \Theta(1)$

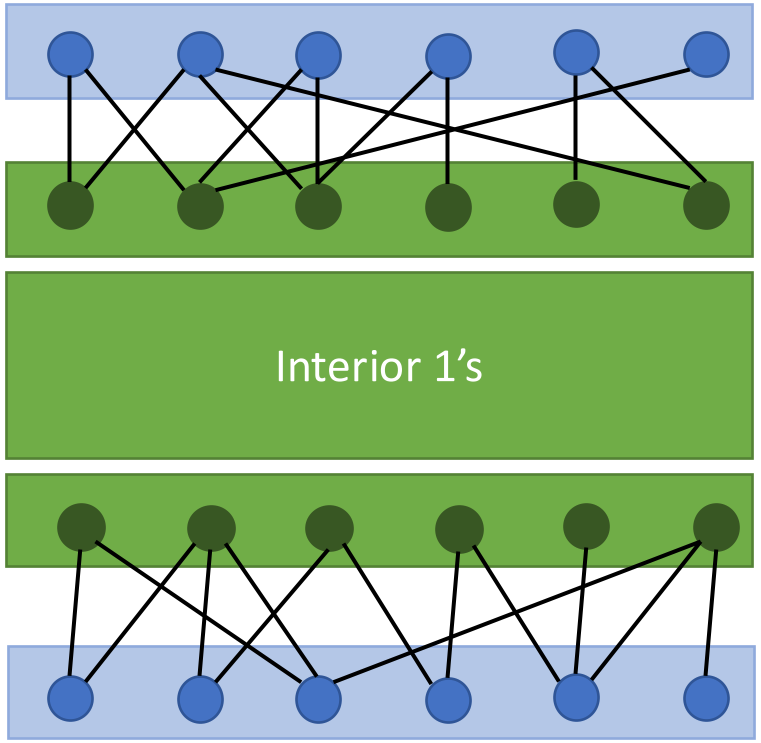
anti-dictator



$I_f = \Theta(\sqrt{d})$

anti-majority

Consider the Vertex Boundary

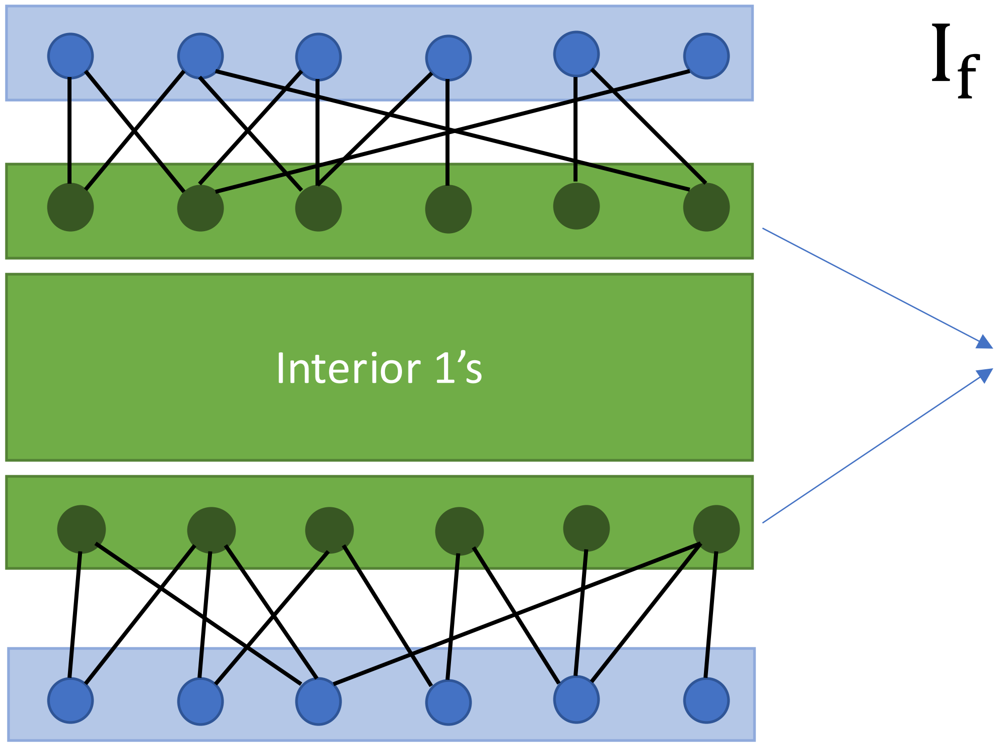


$$\mathrm{I_f} \;\coloneqq\; \mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)]$$

“Edge boundary”

“Vertex Boundary”

Consider the Vertex Boundary



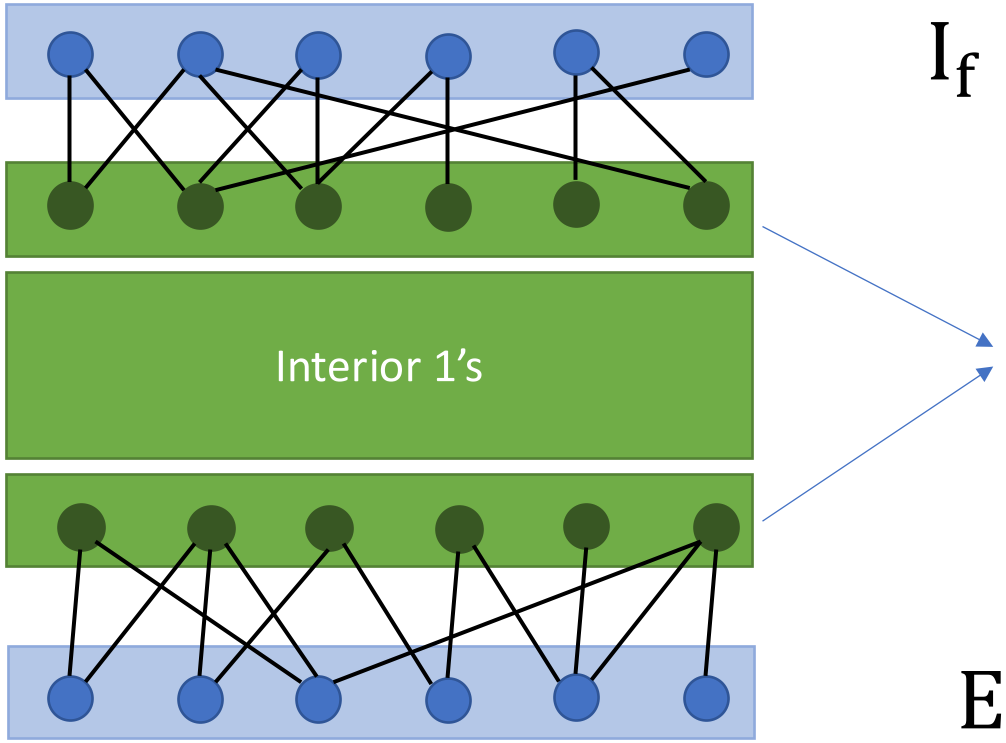
$$\mathrm{I_f} \; := \; \mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)]$$

"Edge boundary"

"Vertex Boundary"

$\frac{1}{2^d} \cdot$ $\{x$: $\mathbf{1}$ such that it has at least one influential edge incident on it$\}$

Consider the Vertex Boundary



$$\mathrm{I_f} \;\coloneqq\; \mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)]$$

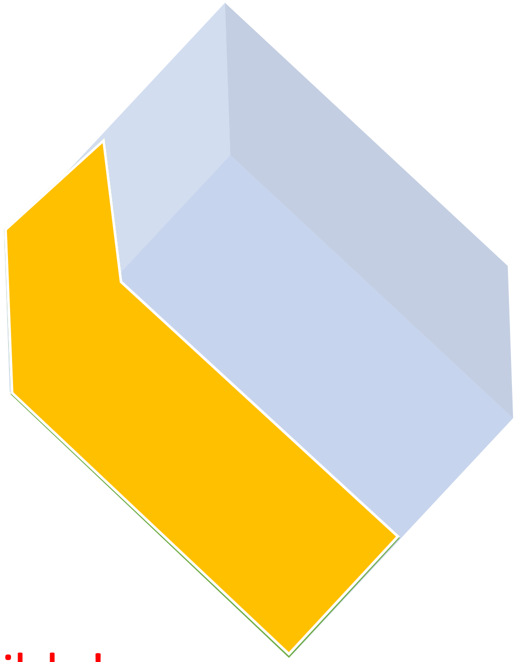
"Edge boundary"

"Vertex Boundary"

$\frac{1}{2^d} \cdot$ $\{x$: **1** such that it has at least one influential edge incident on it$\}$

$$\mathrm{Exp}_{x:\mathbf{1}}[\mathbb{I}\{\mathrm{Inf}(x) > 0\}]$$

Vertex Boundaries



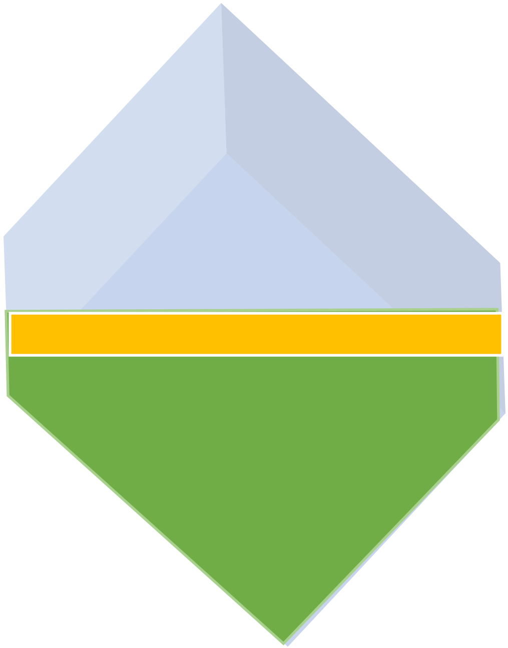
Lowest possible!
(By Poincare)

anti-dictator

Small Edge boundary $= \Theta(1)$

Every vertex in the boundary

Large Vertex boundary $= \ \Theta(1)$



anti-majority

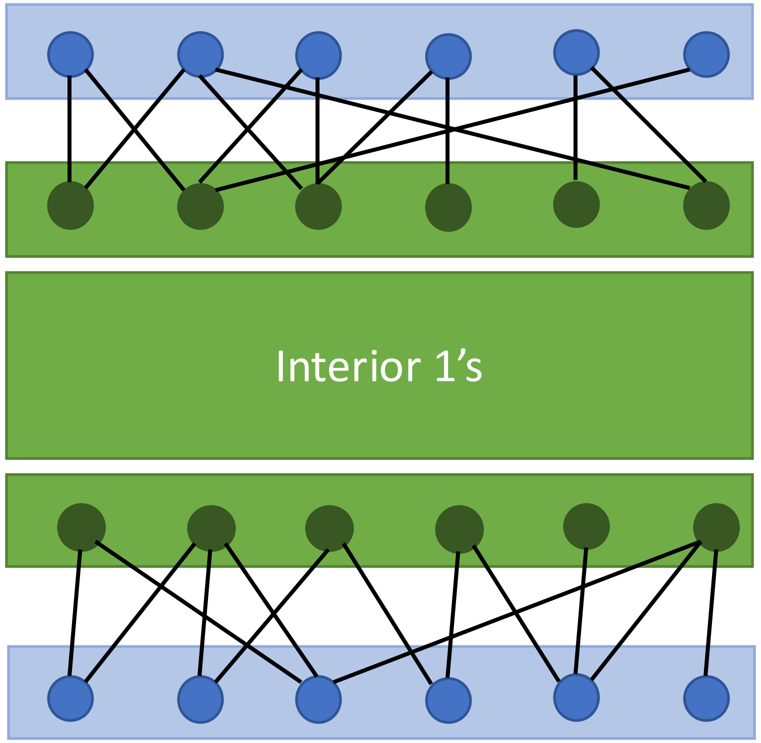
Large Edge boundary $= \ \Theta(\sqrt{d})$

Only $1/\sqrt{d}$ fraction in boundary

Lowest possible!
(By Harper)

Small Vertex boundary $= \Theta(1/\sqrt{d})$

Margulis



$$\mathrm{I}_f \;\coloneqq\; \mathrm{Exp}_{x:1}[\mathrm{Inf}(x)]$$ *“Edge boundary”*

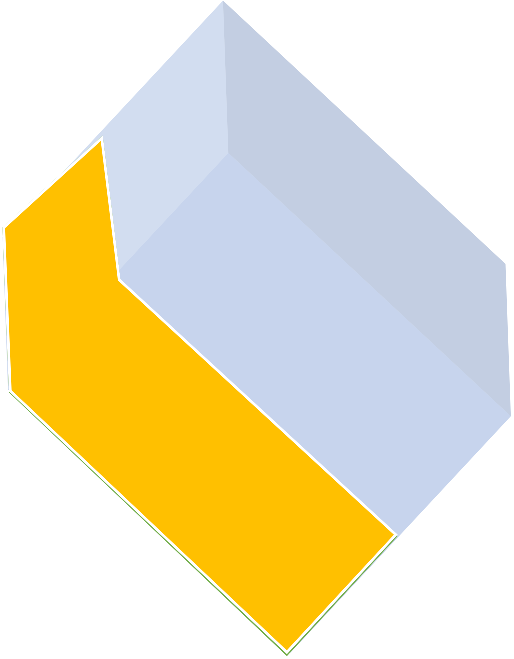
$$\Gamma_f \;\coloneqq\; \mathrm{Exp}_{x:1}[\mathbb{I}\{\mathrm{Inf}(x) > 0\}]$$ *“Vertex Boundary”*

Margulis 1974:

$$I_f \cdot \; \Gamma_f \; = \Omega(var_f^2)$$

Both cannot be simultaneously small!

Vertex Boundaries

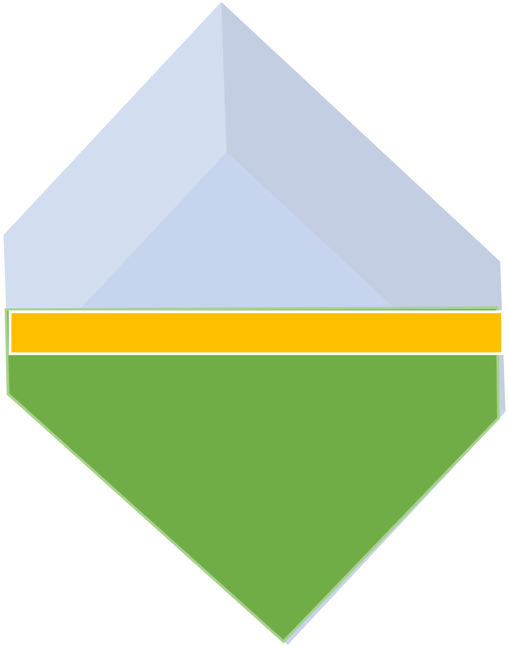


anti-dictator

Small $\mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)] = 1$

Every vertex in the boundary

Large $\mathrm{Exp}_{x:\mathbf{1}}[\mathbb{I}\{\mathrm{Inf}(x) > 0\}] = 1$



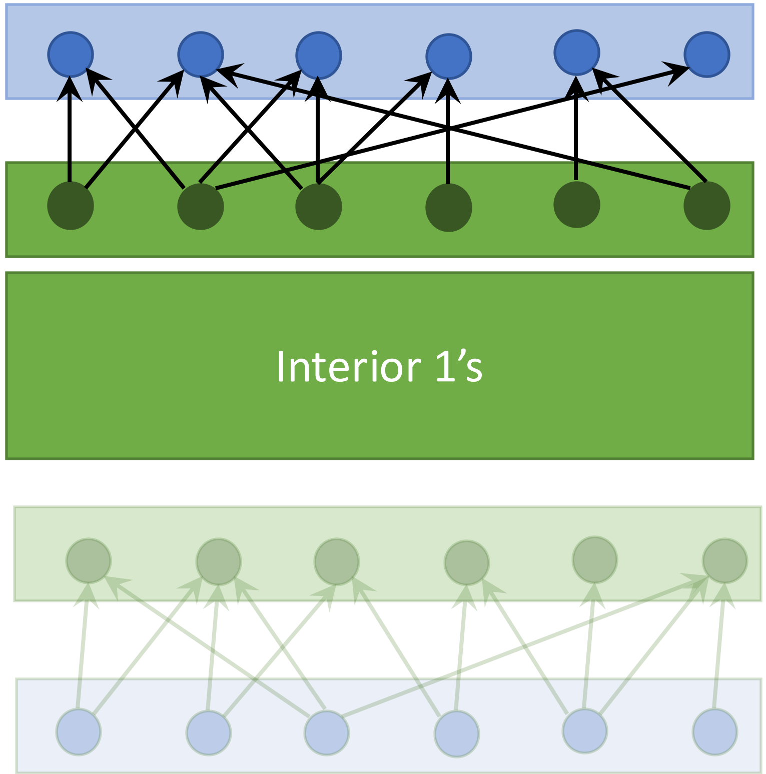
anti-majority

Large $\mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}(x)] = \Theta(\sqrt{d})$

Only $1/\sqrt{d}$ fraction in boundary

Small $\mathrm{Exp}_{x:\mathbf{1}}[\mathbb{I}\{\mathrm{Inf}(x) > 0\}] = \Theta(1/\sqrt{d})$

Directed Margulis



$$\mathrm{I}_{\mathrm{f}}^{+} \;:=\; \mathrm{Exp}_{x:\mathbf{1}}[\mathrm{Inf}^{+}(x)]$$

"Directed Edge boundary"

$$\Gamma_{f}^{+} \;:=\; \mathrm{Exp}_{x:\mathbf{1}}[\mathbb{I}\{\mathrm{Inf}^{+}(x) > 0\}]$$

"Dir. Vrtx. boundary"

[Chakrabarty-S 13]:

$$I_{f}^{+} \cdot \; \Gamma_{f}^{+} \; = \Omega(\varepsilon_{f}^{2})$$

Analysis f the Anti-Dictator

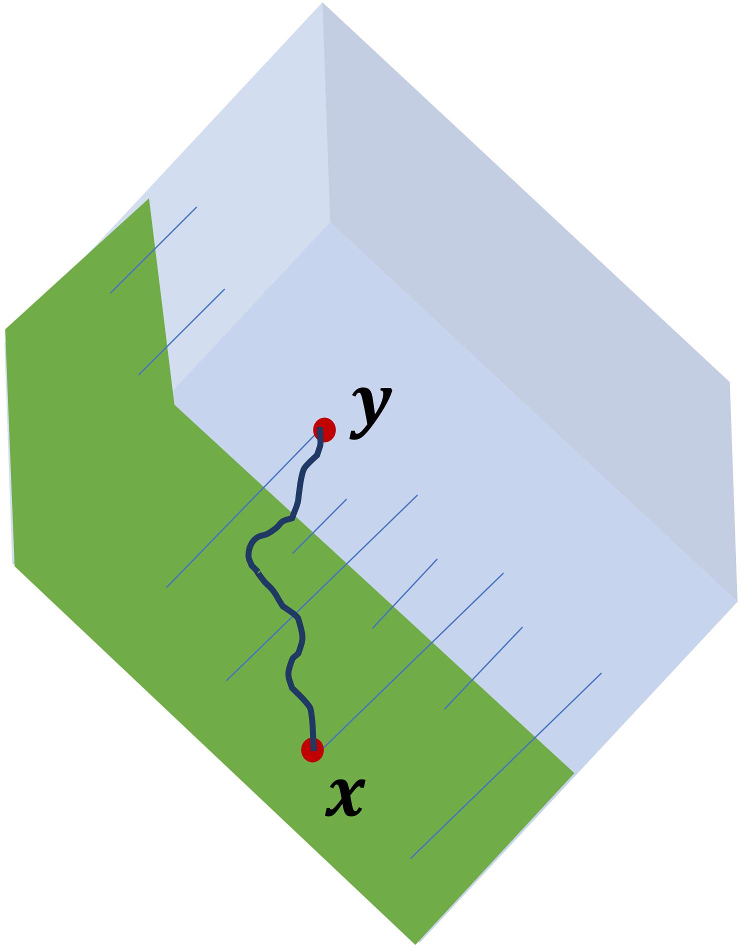
By directed Margulis, if $I_f^+ = O(1)$, constant fraction of vertices on directed boundary.

Analysis like this should work?

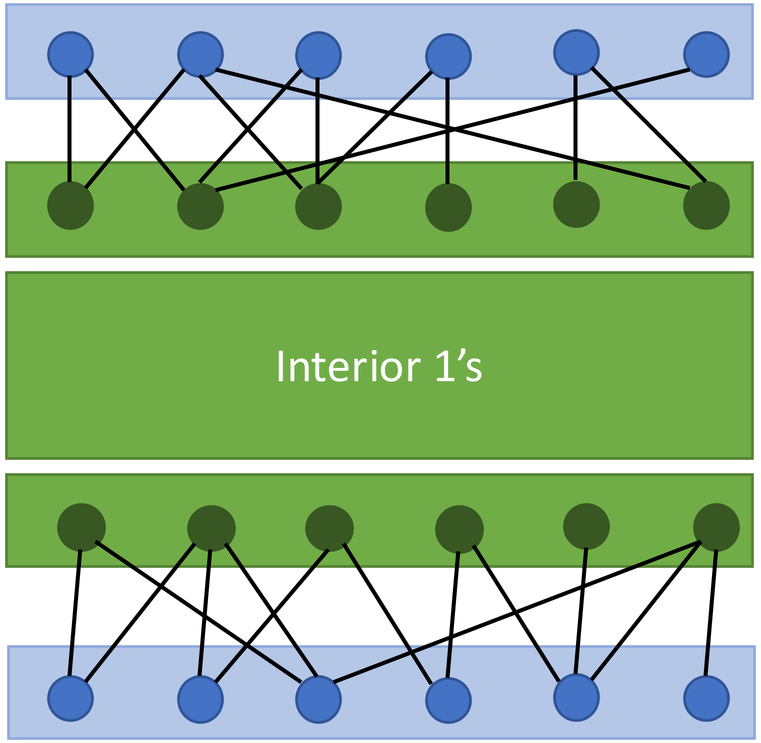
- Prob[x is green] ≈ 1/2
- Prob[crossing] $\approx \sqrt{d} \cdot \frac{1}{d} \approx \frac{1}{\sqrt{d}}$

Number of steps

Chance of crossing in each step



The Talagrand theorem



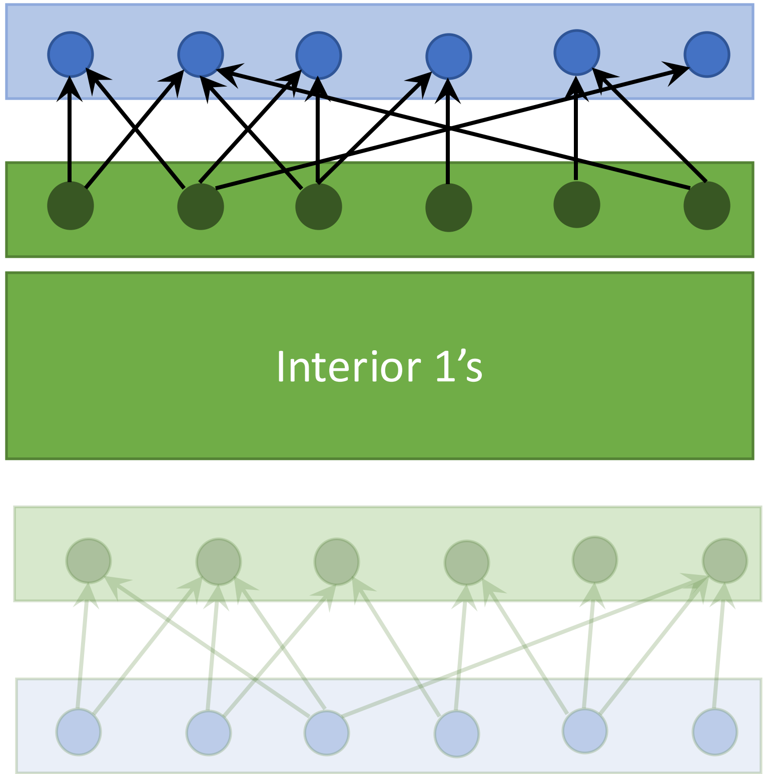
Talagrand 1993: *"Notion of Surface Area"*

$$\mathrm{Tal}_f \;\coloneqq\; \mathrm{Exp}_{x:\mathbf{1}}[\sqrt{\mathrm{Inf}(x)}]$$

$$\mathrm{Tal}_f = \Omega(var_f)$$

Implies Margulis (by Cauchy-Schwartz), which also implies Poincare

Khot-Minzer-Safra



$$\mathrm{Tal}_f^+ \;:=\; \mathrm{Exp}_{x:1}[\sqrt{\mathrm{Inf}^+(x)}]$$

$$= \Omega(\varepsilon_f)$$

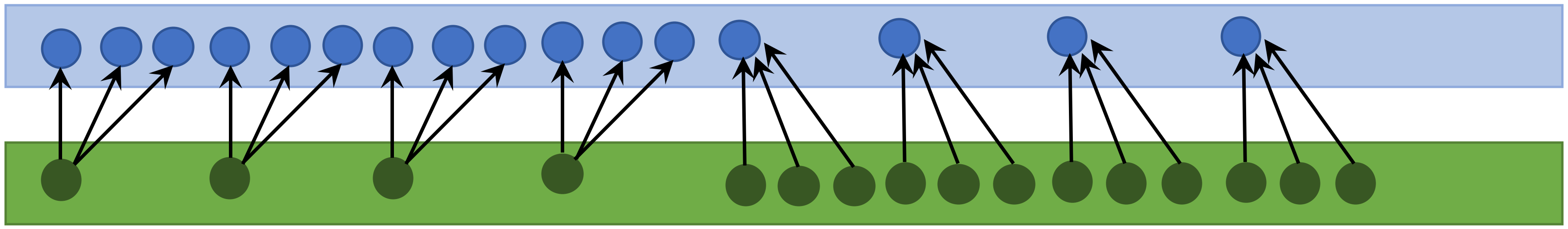
Actually, for any edge bicoloring ψ,

$$\mathrm{Tal}_f^+ \;:= \min_\psi \mathrm{E}_x[\sqrt{\mathrm{Inf}_\psi^+(x)}]$$

KMS lost a log factor, which [Pallavoor-Raskhodnikova-Waingarten 22] removed

Robustness

$$\mathrm{Tal}_f^+ := \mathrm{Exp}_{x:1}\left[\sqrt{\mathrm{Inf}^+(x)}\right] = \Omega(1) \qquad \mathrm{Tal}_f^+ := \mathrm{Exp}_{x:0}\left[\sqrt{\mathrm{Inf}^+(x)}\right] = \Omega(1)$$



2^d vertices

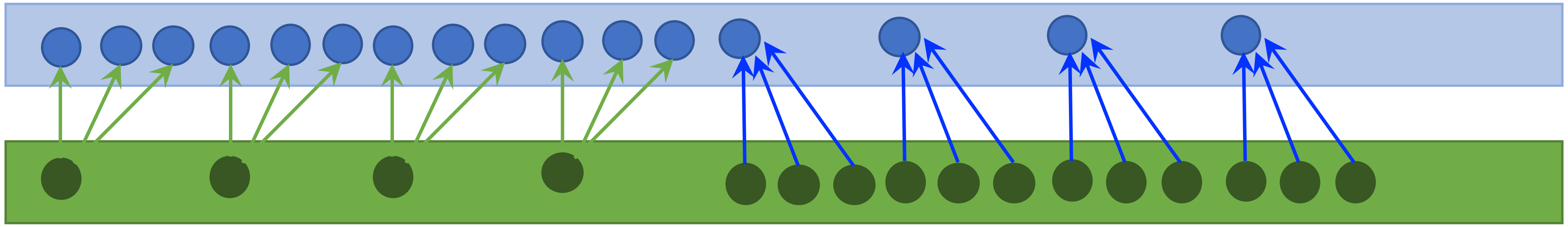
2^d edges

Adversary ψ assigns each edge to 0 or 1, to minimize Talagrand “surface area”

$$\mathrm{Tal}_f^+ := \min_\psi \mathrm{E}_x[\sqrt{\mathrm{Inf}_\psi^+(x)}]$$

Robustness

$\mathrm{Tal}_f^+ \ := \ \mathrm{Exp}_{x:1}\left[\sqrt{\mathrm{Inf}^+(x)}\right] = \ \Omega(1)$ $\quad$ $\mathrm{Tal}_f^+ \ := \ \mathrm{Exp}_{x:0}\left[\sqrt{\mathrm{Inf}^+(x)}\right] = \ \Omega(1)$



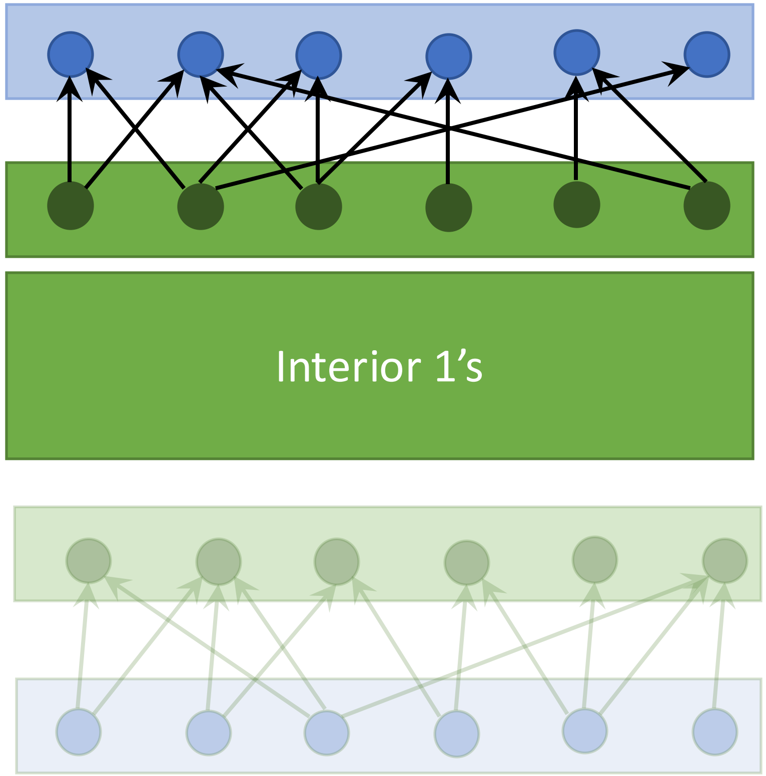
2^d vertices

2^d edges

Adversary ψ assigns each edge to 0 or 1, to minimize Talagrand "surface area"

$$\mathrm{Tal}_f^+ \ := \min_\psi \mathrm{E}_x\left[\sqrt{\mathrm{Inf}_\psi^+(x)}\right] = \frac{2^d}{d}\cdot\sqrt{d} = O(1/\sqrt{d})$$

Khot-Minzer-Safra



Actually, for any edge bicoloring ψ,

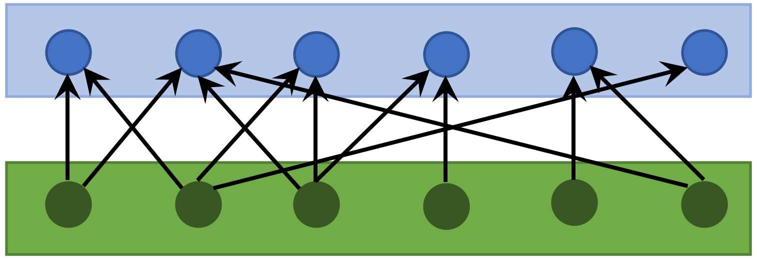
$$\mathrm{Tal}_f^+ := \min_\psi \mathrm{E}_x\left[\sqrt{\mathrm{Inf}_\psi^+(x)}\right] = \Omega(\varepsilon_f)$$

KMS lost a log factor, which [Pallavoor-Raskhodnikova-Waingarten 22] removed

But what does it mean?

$$\mathrm{Tal}_f^+ := \min_\psi \mathrm{E}_x[\sqrt{\mathrm{Inf}_\psi^+(x)]} = \Omega(\varepsilon_f)$$

Assume $\varepsilon_f = \Omega(1)$



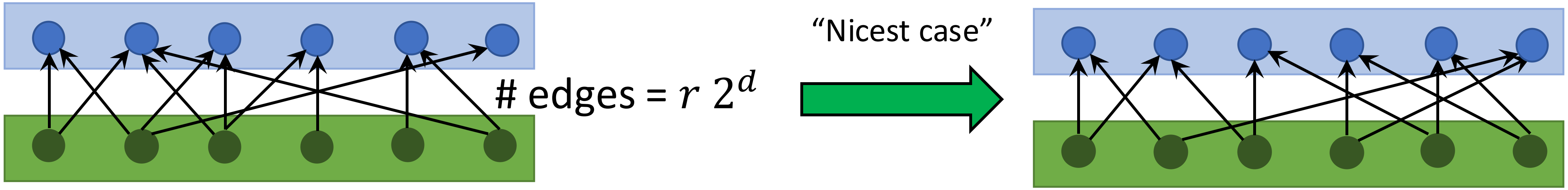
vertices = $r\ 2^d$ $\quad I_f^+ \cdot \Gamma_f^+ = \Omega(1)$ Dir. Margulis

$I_f^+ = \Omega(1)$ Dir. Poincare

vertices = $2^d/r$

For some r > 1, $I_f^+ = r$ and $\Gamma_f^+ = 1/r$

The most regular boundary



edges = $r\ 2^d$

“Nicest case”

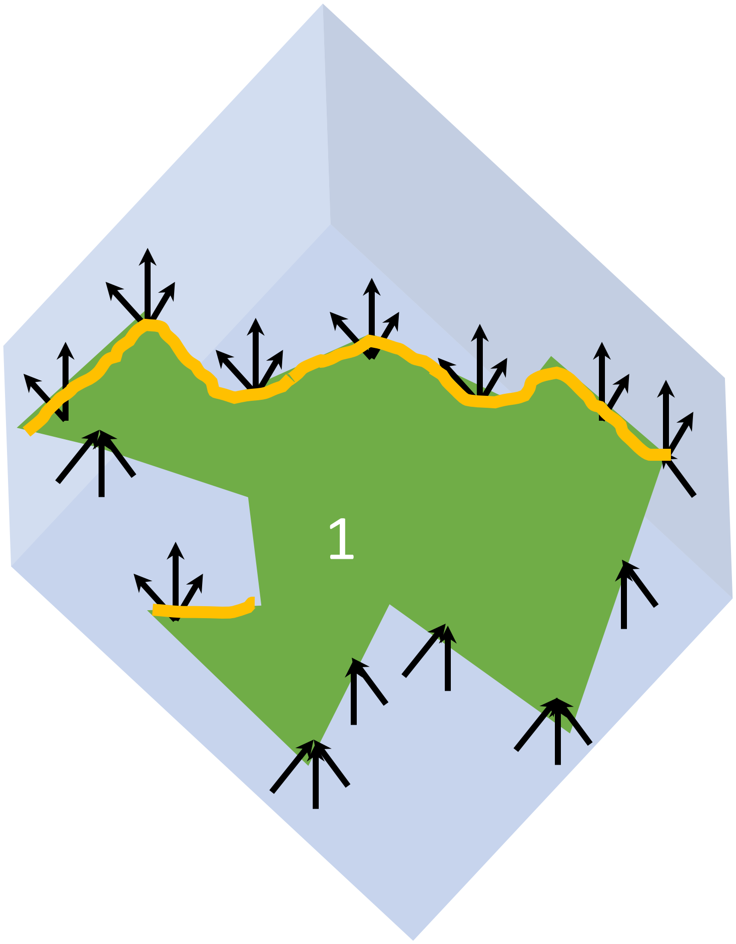
vertices = $2^d/r$

vertices = $2^d/r$

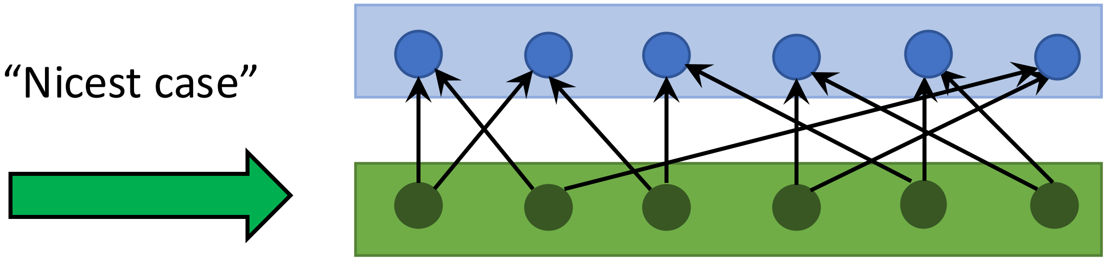
All degrees = r^2

- Robust Talagrand theorem of KMS implies that’s exactly what happens!

Boundaries are always regular!



Assume $\varepsilon_f = \Omega(1)$



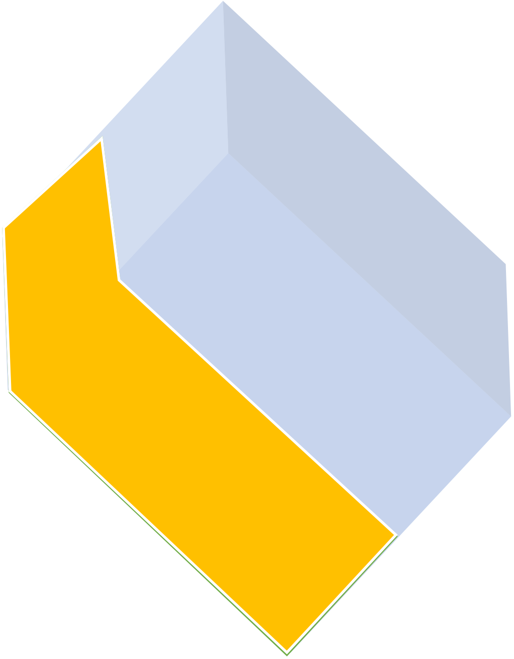
vertices = $2^d/r$

All degrees = r^2

There exists r > 1 such that boundary contains regular bipartite graph with these parameters.

- (Up to $d^{o(1)}$ factors)

Our usual examples

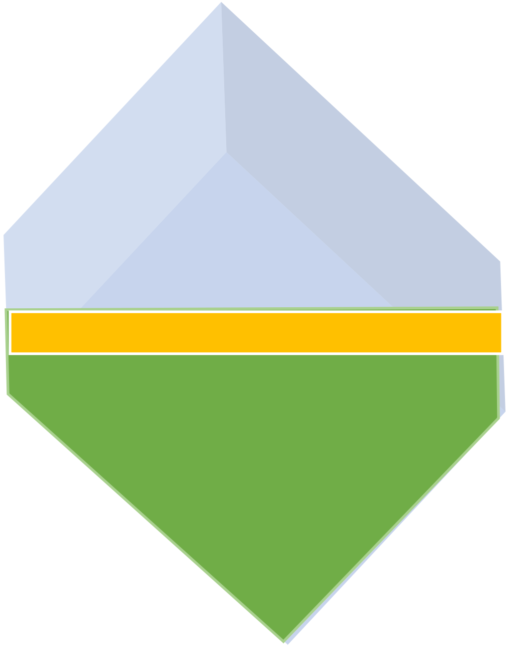


anti-dictator

2^d vertices on boundary,
each with degree 1

Boundary size = $2^d/r$

All degrees = r^2



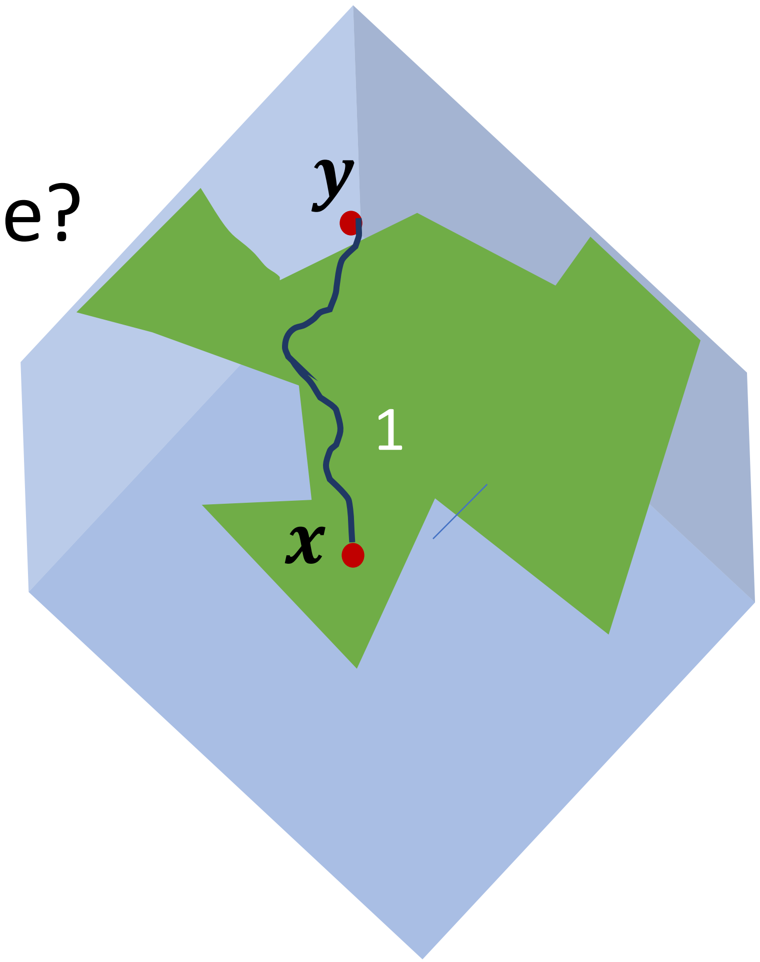
anti-majority

$2^d/\sqrt{d}$ vertices on
boundary,
each with degree d

But what about monotonicity testing?

How to analyze the “path tester”

How to “escape” S that is far from monotone?



Be persistent!

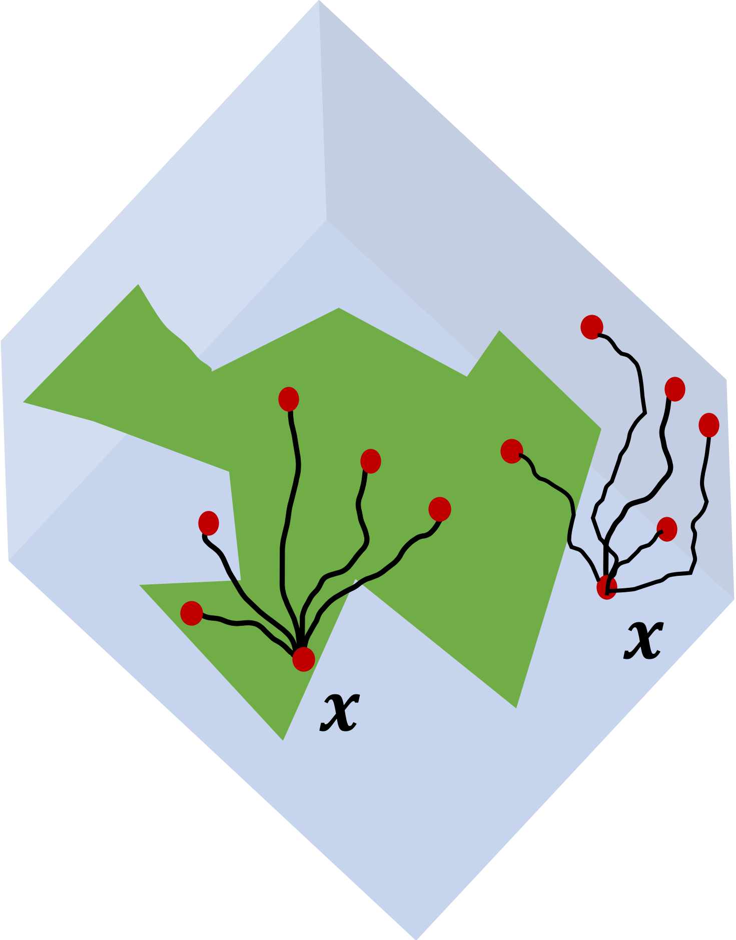
- x is ℓ-persistent, if ℓ-length (directed walk) stays within f(x)-region whp

Pr[single step changes value] = $\frac{I_f}{d}$

Pr[one of ℓ-steps changes value] $\leq \ell \cdot \frac{I_f}{d}$

Fraction of NON ℓ-persistent vertices $= O\left(\ell \cdot \frac{I_f}{d}\right)$

$$\xrightarrow{I_f = O(\sqrt{d})} O\left(\frac{\ell}{\sqrt{d}}\right)$$



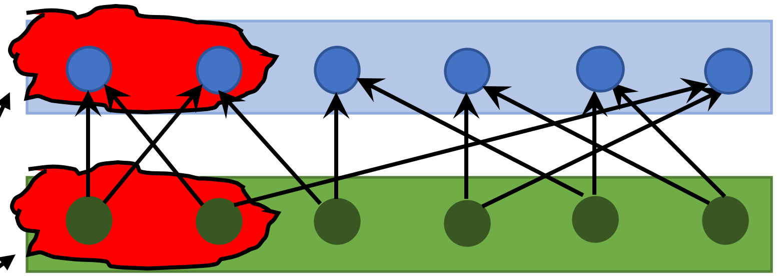
If $I_f \gg \sqrt{d}$ then $I_f^+ \gg \sqrt{d}$

Edge tester itself good

The analysis, in one slide

By (robust) Directed Talagrand, boundary is

Non- ℓ-persistent

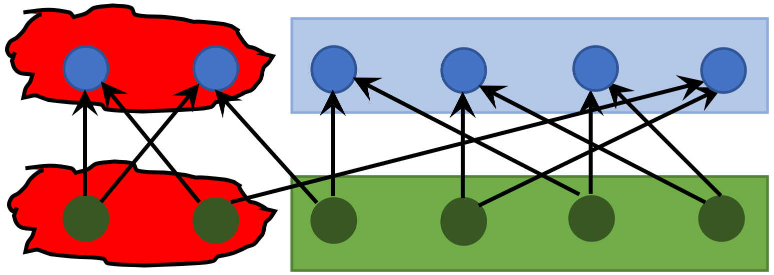


Frac. non-persistent = $O\left(\frac{\ell}{\sqrt{d}}\right)$

Set $\ell \ll \sqrt{d}/r$
so frac. $\ll 1/r$

vertices = $2^d/r$

All degrees = r^2



By regularity of boundary

vertices ≈ $2^d/r$

All degrees ≈ r^2

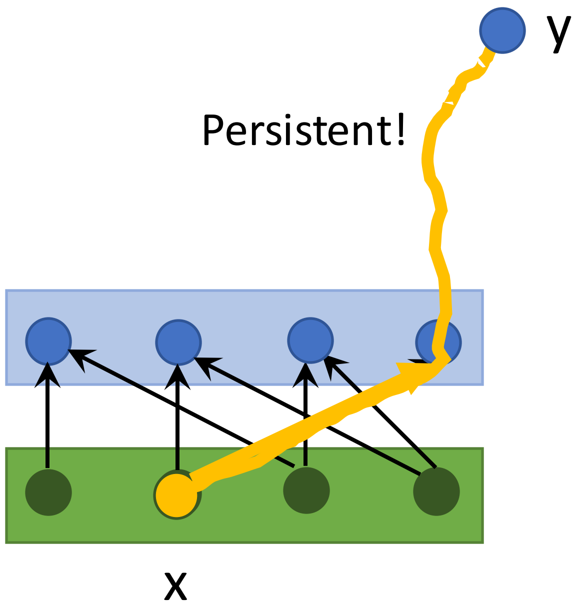
All ℓ-persistent

The analysis, in ~~one~~ two slides

With prob ≈ $1/r$, start with f(x) = 1 in this part

With prob ≈ $r^2/\ell = r/\sqrt{d}$, relevant bit is flipped

When both happen, f(y) = 0



By regularity of boundary

\# vertices ≈ $2^d/r$

All degrees ≈ r^2

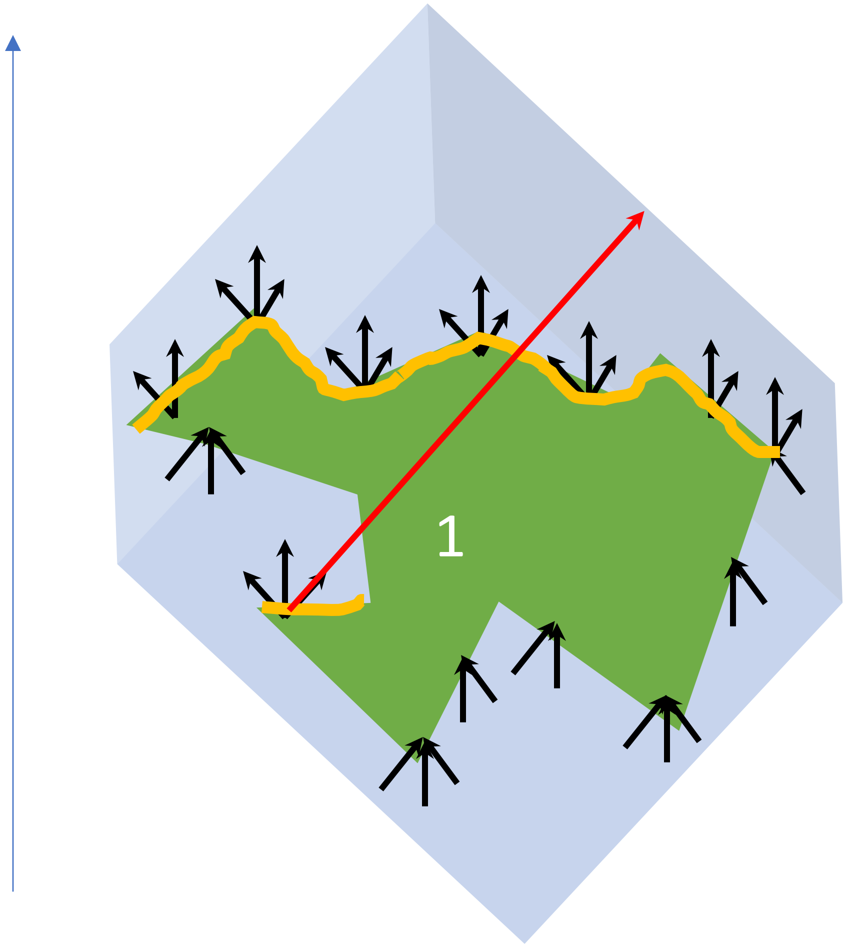
All ℓ-persistent

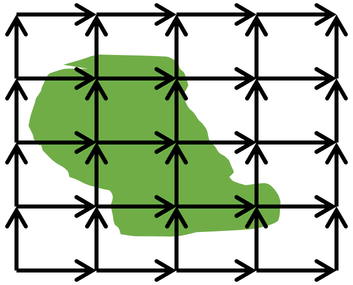
$$\ell = o(\sqrt{d}/r)$$

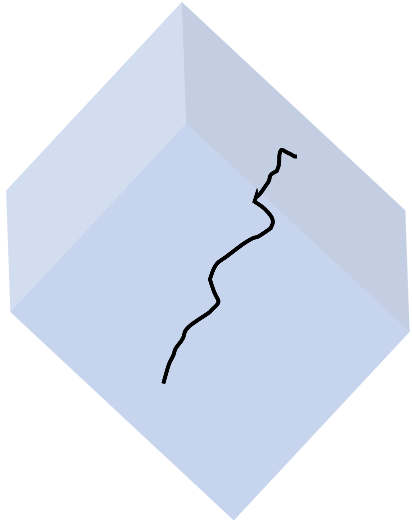
Total prob ≈ $\frac{1}{r} \cdot \frac{r}{\sqrt{d}} = \frac{1}{\sqrt{d}}$

The challenge of hypergrids

What is a boundary?





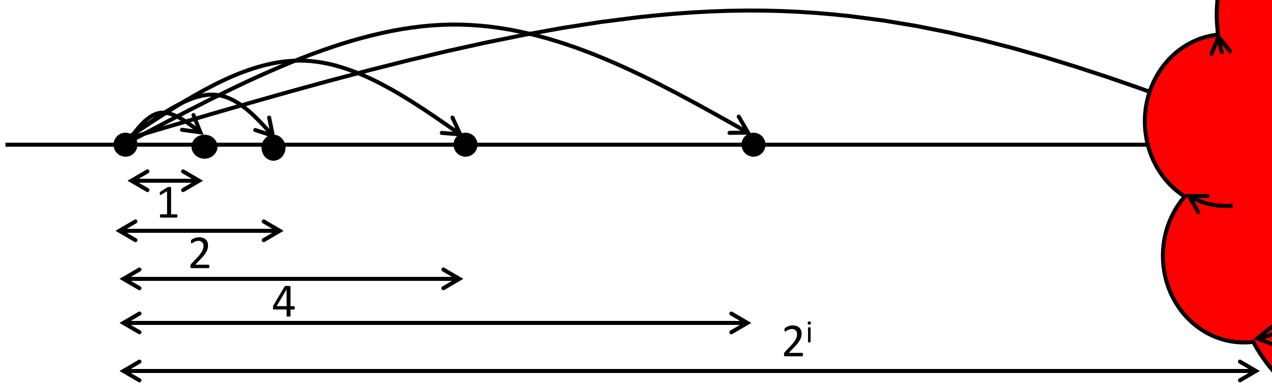


Move in a direction. But for “how much”?

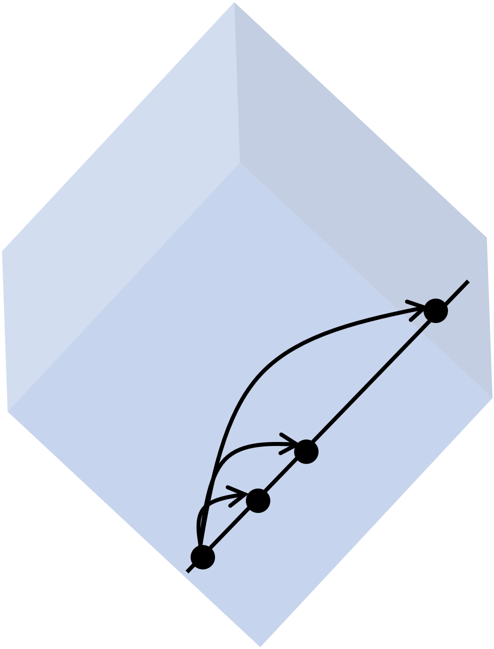
What is influence?

How to define a path tester? What are upward random walks?

The augmentation view



Does directed Talagrand hold?

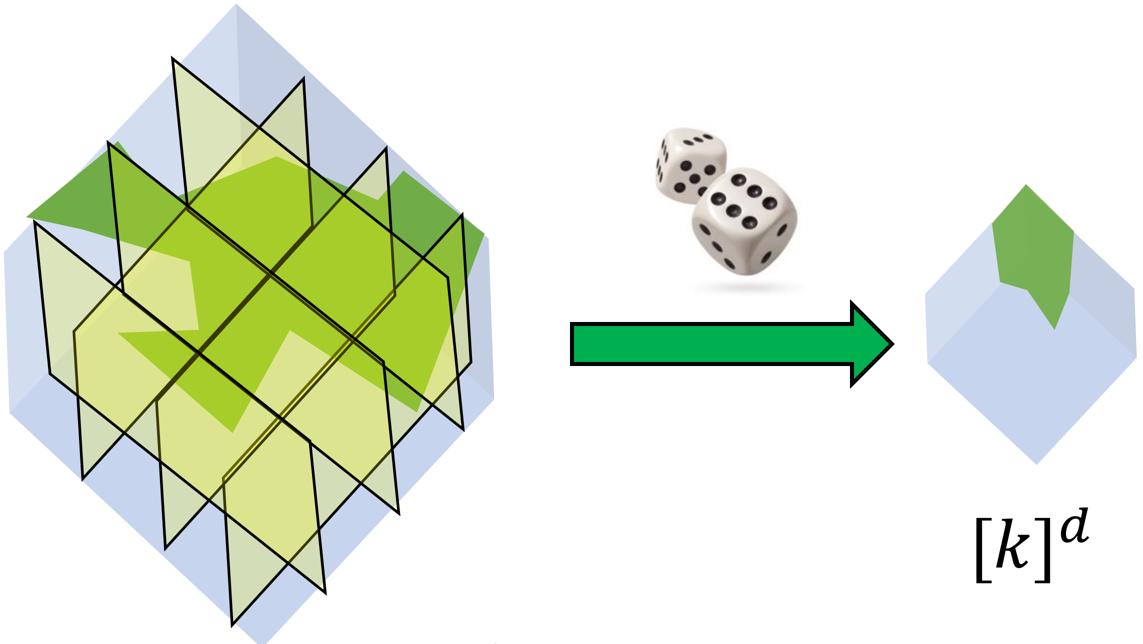


- Degree is now $d \log n$
  - Can we treat it a hypercube/product structure?
- [BCS18] Directed Margulis holds
  - Define boundaries in the graph theoretic way
  - $d^{5/6} \log n$ query tester, using path tester

Directed Margulis proofs more amenable to alternate domains.

[KMS15] very intricate proof; doesn't seem to generalize

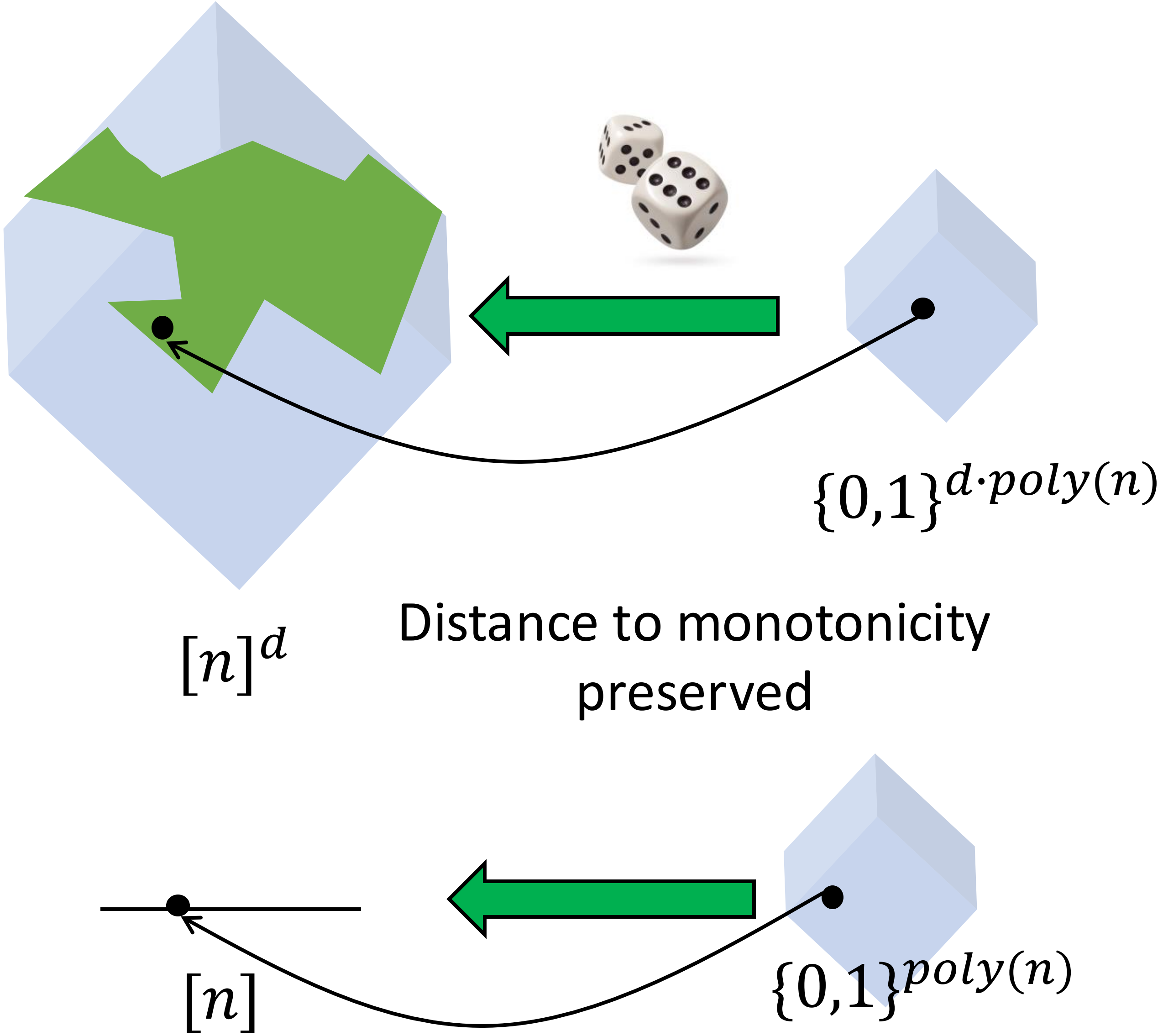
Domain reduction



Is distance to monotonicity preserved?

- Set k=2, and reduce to hypercubes…?
- [BCS20] If k << d, sampled function can be close to monotone
- [BCS20] If k = poly(d), distance is preserved
  - So we can assume n = poly(d)
  - Even reduces from continuous
- [Harms-Yoshida 22] Downsampling

The embedding method

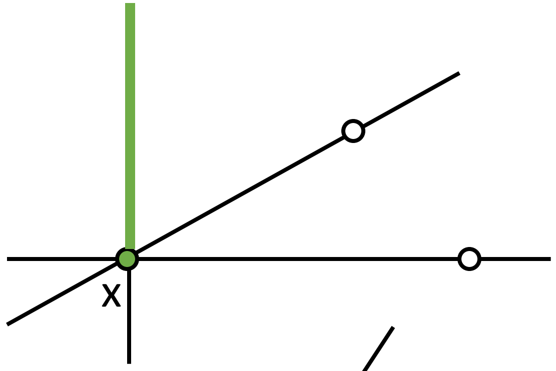


Distance to monotonicity preserved

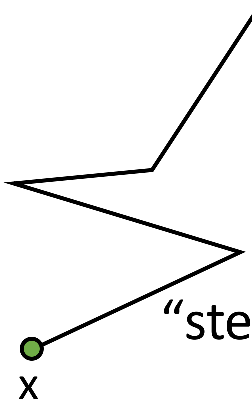
- Embedding [n] into hypercube
- [Braverman-Kindler-Khot-Minzer 23]

- First $\sqrt{d}$ tester for any n > 2
  - poly(n) dependence necessary

Doing isoperimetry



Dimension contains violation



“step” uar in [x$_i$, n]

$\mathrm{Inf}^+(x) =$ # dimensions where x is in violation

[BCS23–1] Tal_f^+

$:= \; \mathrm{Exp}_{x:1}\left[\sqrt{\mathrm{Inf}_f^+(x)}\right]$

$= \Omega(\varepsilon_f / \log n)$

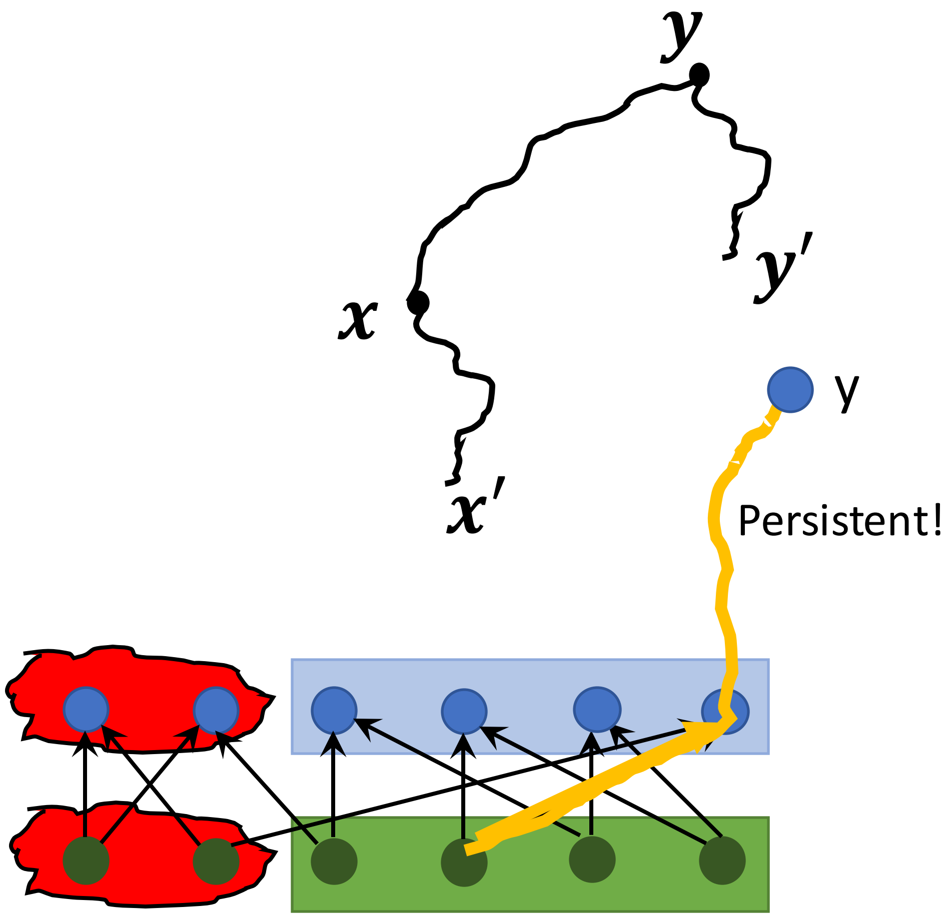
- Also prove the robust version
- Direct tester analysis leads to $n\sqrt{d}$ query tester
  - Path picks uniform random step in each direction

The final step

[BCS 23-2] Getting $d^{\frac{1}{2}+o(1)}$ query tester

- Tester does correlated walks
  - Heavily used in analysis
  - Standard tester probably works

- How to perform this analysis and not lose n factor?
  - New combinatorial tools to analyze the walk using the [BCS23-1] Talagrand theorem

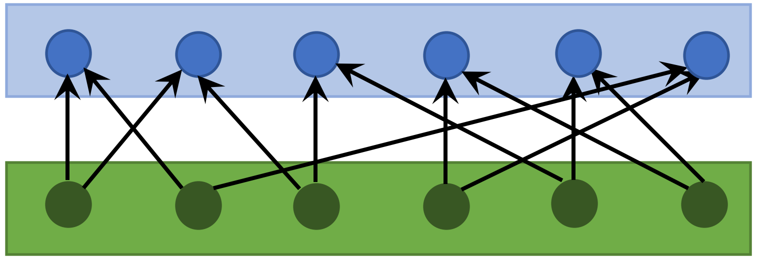


(Mostly) end of story

vertices = $2^d/r$

All degrees = r^2

$$\min_{\psi} \mathrm{E}_x\left[\sqrt{\mathrm{Inf}_{\psi}^{+}(x)}\right] = \Omega(\varepsilon_f / \log n)$$



- $d^{o(1)}$ loss annoying, comes from one specific calculation
- $d^{\frac{1}{2}+o(1)}$ non-adaptive one-sided monotonicity tester
  - Ends a 24 year odyssey
- Nearly matching (two-sided) lower bound

Some odds and ends

- Get rid of the $d^{o(1)}$ factor

- Are hypergrids strictly harder than hypercubes?
  - Only (log d) factors apart

- [Black-Kalemaj-Raskhodnikova 23] Tester for larger ranges, for hypercube
  - We can probably apply the methods for hypergrids

The big question

What is the adaptive complexity of monotonicity testing for hypercubes/grids?

- [Blais-Belovs 16] $\Omega\left(d^{1/4}\right)$ lower bound
- [Chen-Waingarten-Xie 18] $\Omega\left(d^{1/3}\right)$ lower bound
  - Every example has matching adaptive upper bound

- [Chakrabarty-S 16] $O\left(I_f\right)$ query adaptive tester
  - Proves adaptivity gap, for natural families

More on directed isoperimetry

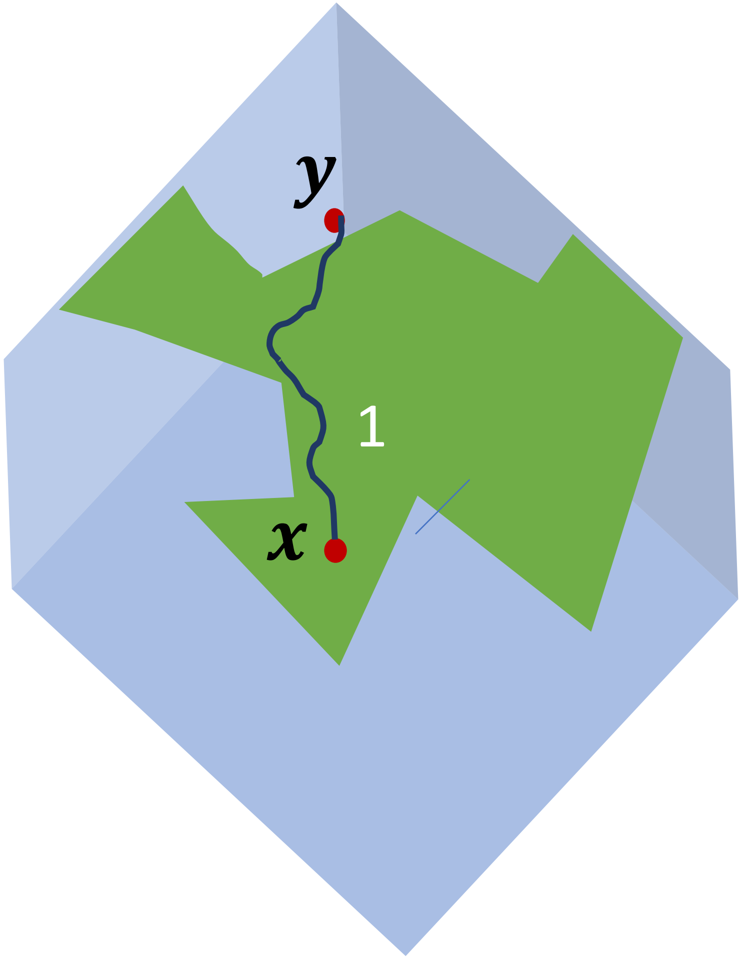
- [Black-Kalemaj-Raskhodnikova 23] Directed Talagrand for real-range

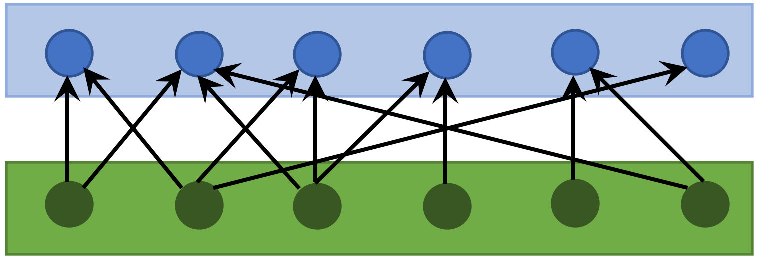
- [Pinto 23, Pinto 24] Differential functions $f\colon [0,1]^d \to [0,1]$, with access to derivatives
  - Draws connections to optimal transport theory, based on undirected Poincare/Talagrand inequalities
- [Canonne-Chen-Levi-Kamath-Waingarten 21] Applications of robustness concept to distribution testing
- [Yoshida 24] Directed isoperimetry for general posets

Generalizing hypercube results

- [Pallavoor-Raskhodnikova-Waingarten 20] Estimating distance to monotonicity, $\sqrt{d}$-approximations in poly(d) time for hypercube
  - Can we get results for hypergrids? Also implies results for product distributions…?

- [Berman-Raskhodnikova-Yaroslavtsev 14] L_p testing





Thank you!

Here, no bad pun