

# Testing thresholds in random geometric graphs

Siqi Liu

**UC Berkeley**

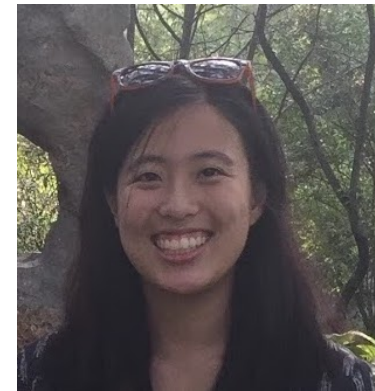
Joint work with



Sidhanth Mohanty  
**UC Berkeley**



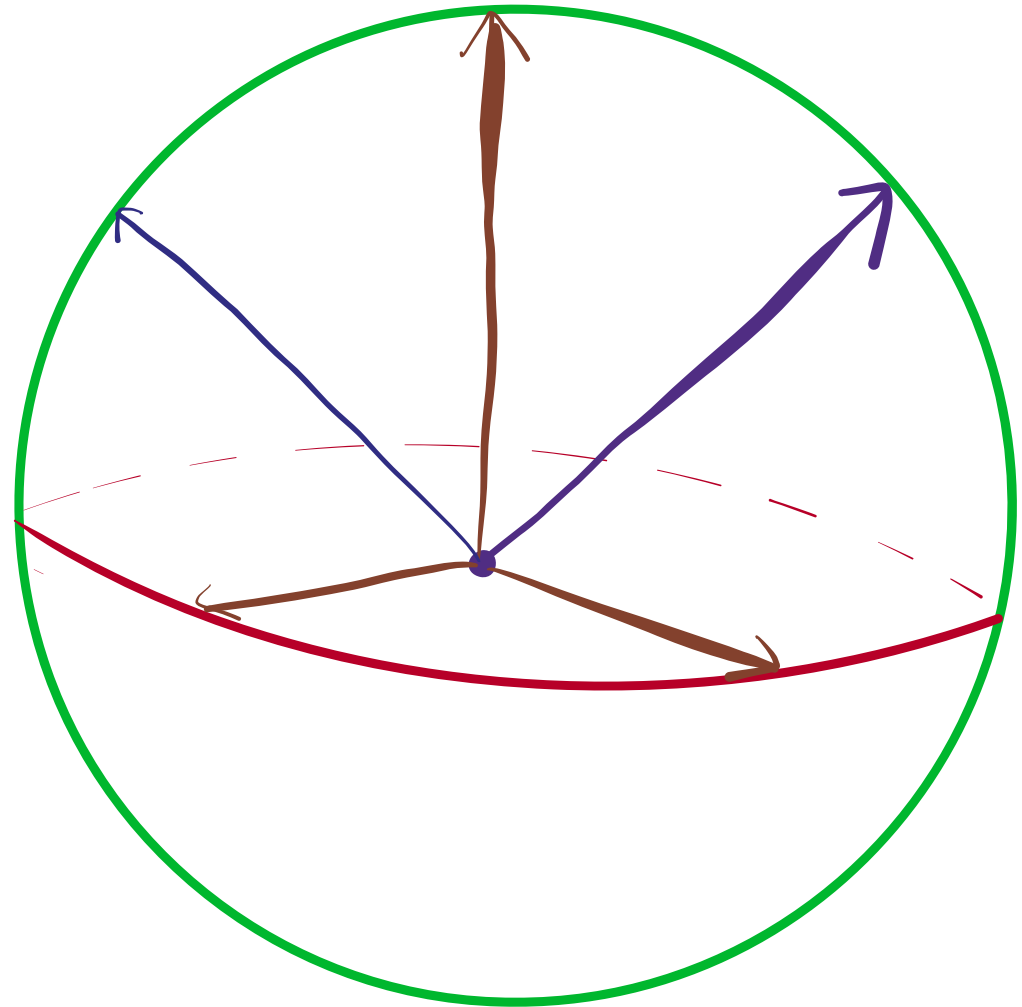
Tselil Schramm  
**Stanford**



Elizabeth Yang  
**UC Berkeley**

# Random geometric graphs

$$v_1, \dots, v_n \sim \mathbb{S}^{d-1}$$

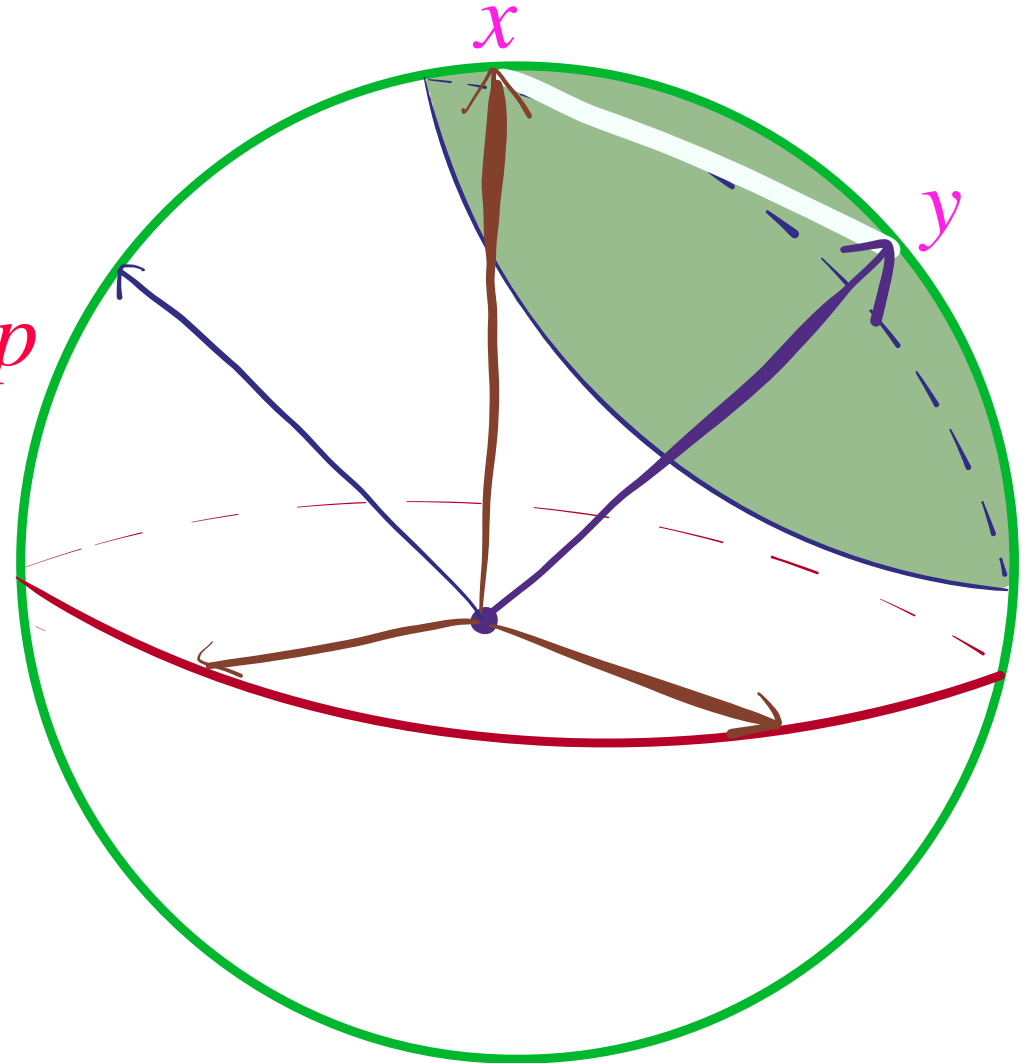


# Random geometric graphs

$$v_1, \dots, v_n \sim \mathbb{S}^{d-1}$$

$xy$  edge iff  $\langle v_x, v_y \rangle \geq \tau(p)$

$xy$  edge with probability  $p$

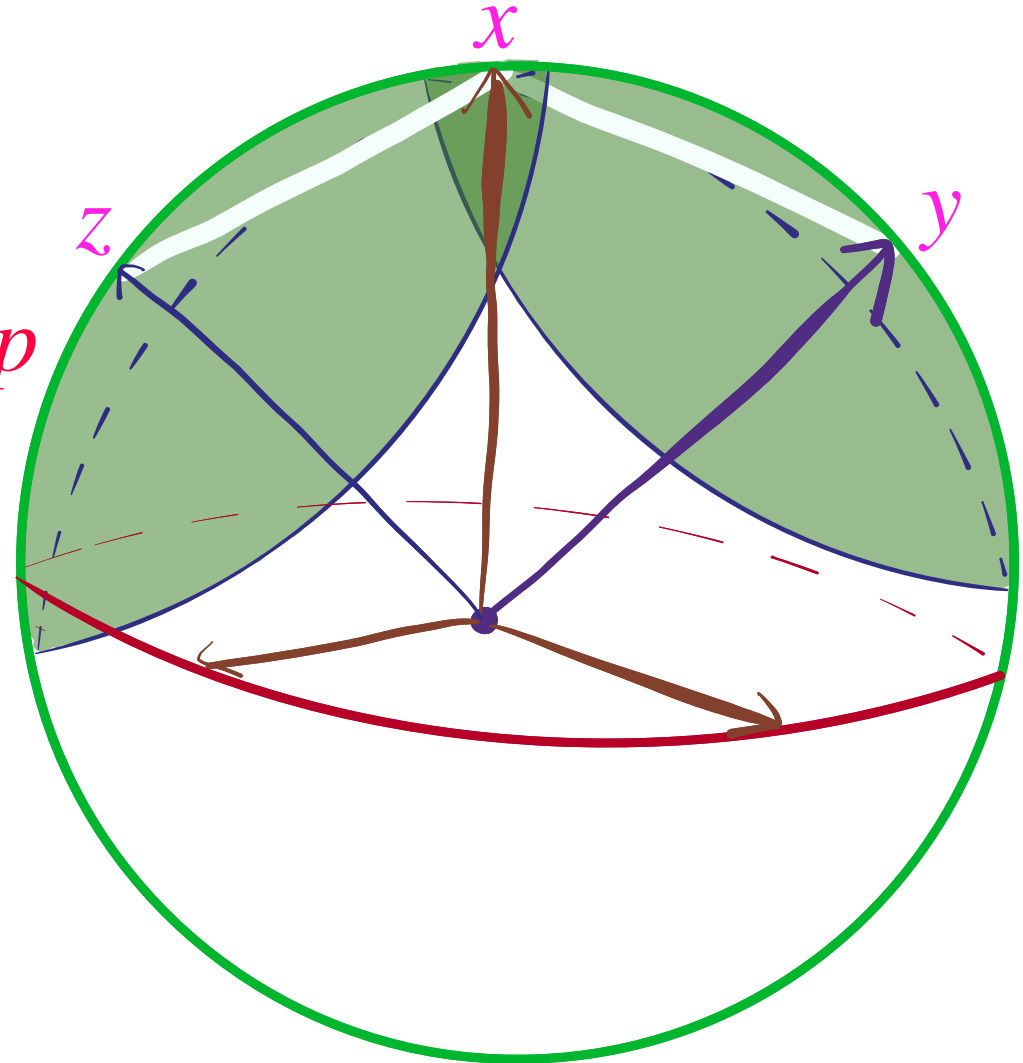


# Random geometric graphs

$$v_1, \dots, v_n \sim \mathbb{S}^{d-1}$$

$xy$  edge iff  $\langle v_x, v_y \rangle \geq \tau(p)$

$xy$  edge with probability  $p$



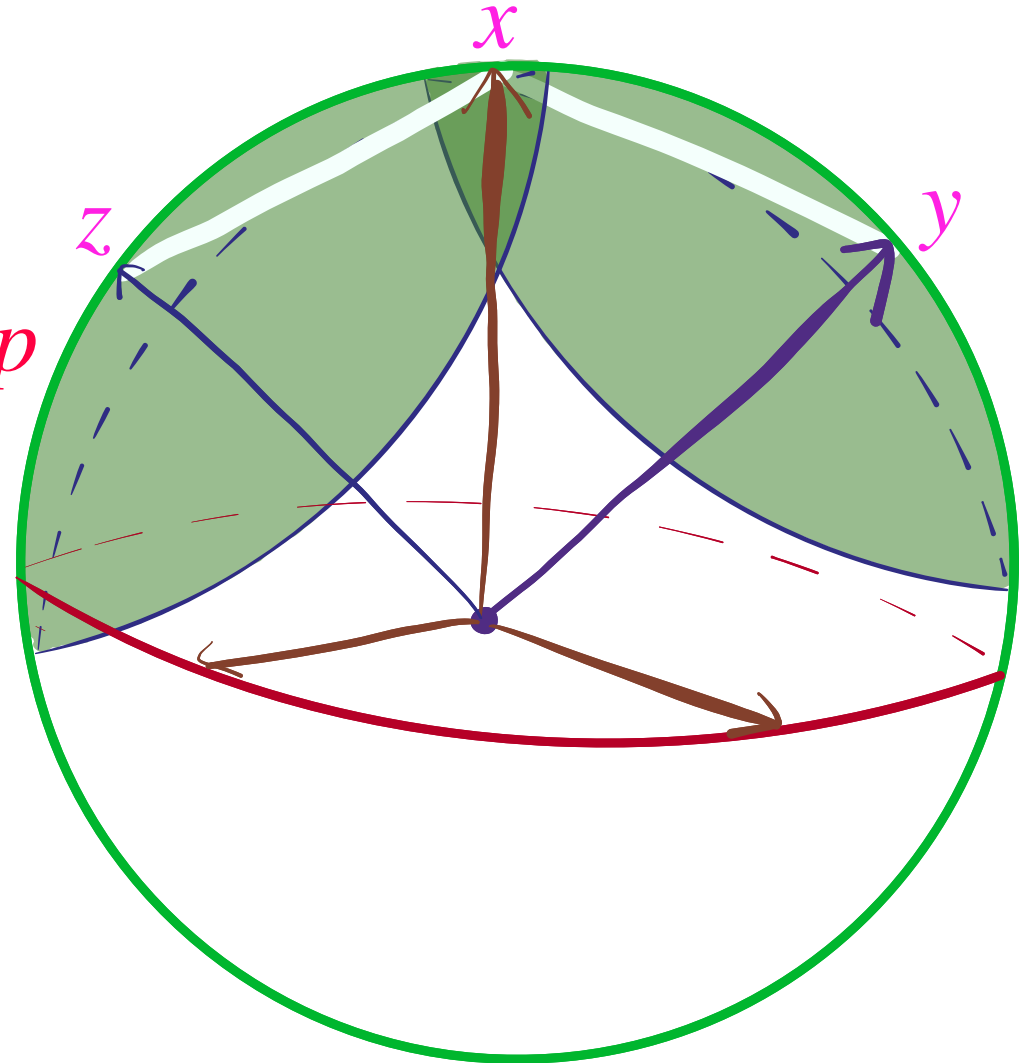
# Random geometric graphs

$$v_1, \dots, v_n \sim \mathbb{S}^{d-1}$$

$xy$  edge iff  $\langle v_x, v_y \rangle \geq \tau(p)$

$xy$  edge with probability  $p$

$$\tau(p) \approx \sqrt{\frac{\log \frac{1}{p}}{d}}$$

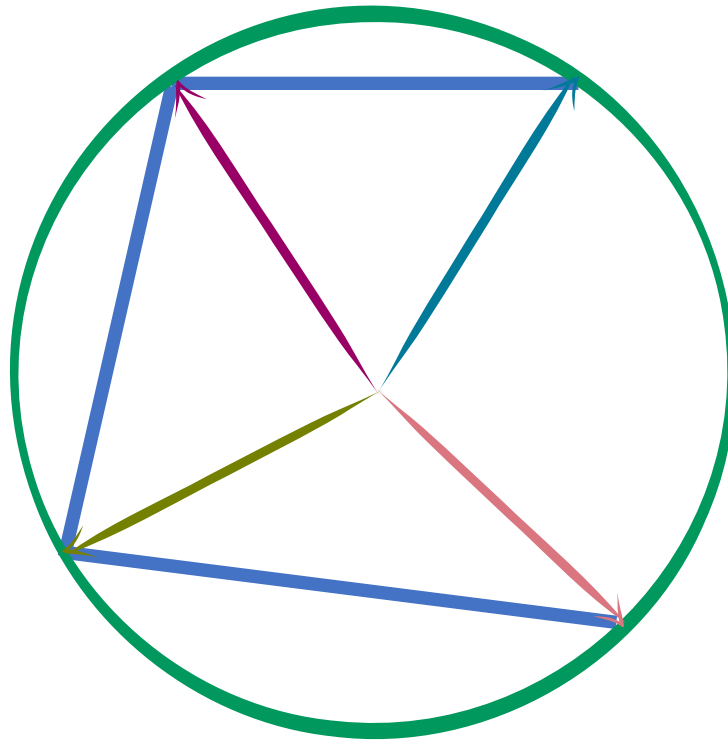


# Random geometric graphs

$$v_1, \dots, v_n \sim \mathbb{S}^{d-1}$$

$$G = \{xy : \langle v_x, v_y \rangle \geq \tau(p)\}$$

Edge  $xy$  exists with probability  $p$



$$G \sim \text{Geo}_d(n, p)$$

# Key motivating property

$$G \sim \text{Geo}_d(n, p)$$

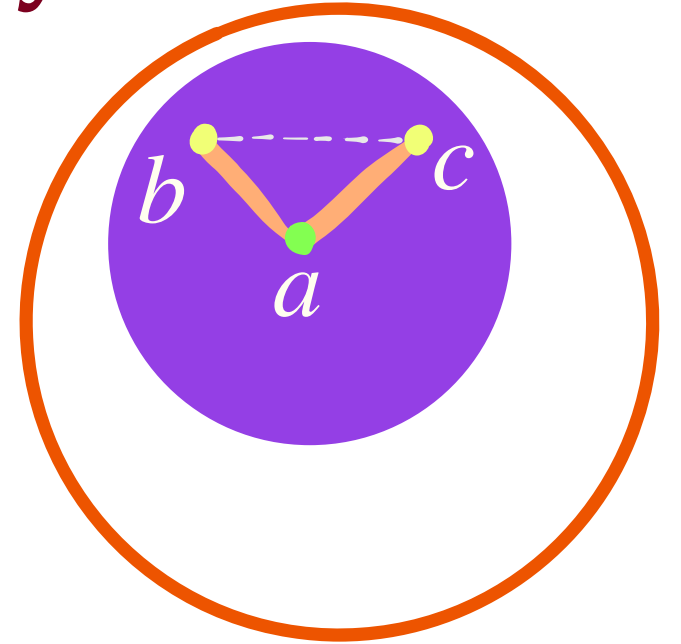
$$\Pr[bc \text{ edge} \mid ab, ac] > p$$

Better approximates real world networks?

More faithful test bed for algorithms  
on graphs (clustering, partitioning etc.)

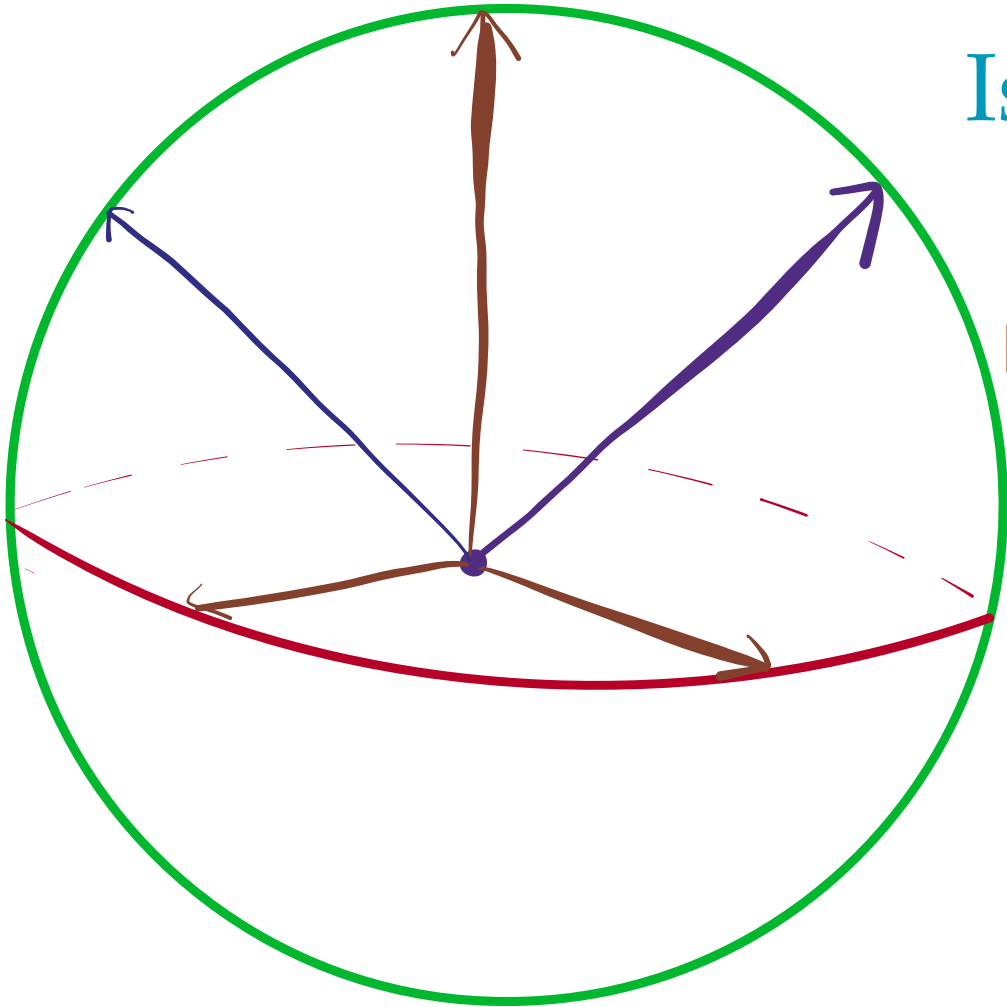
Probabilistic method

Examples of high-dimensional expanders  
(expanders with expanding neighborhoods)

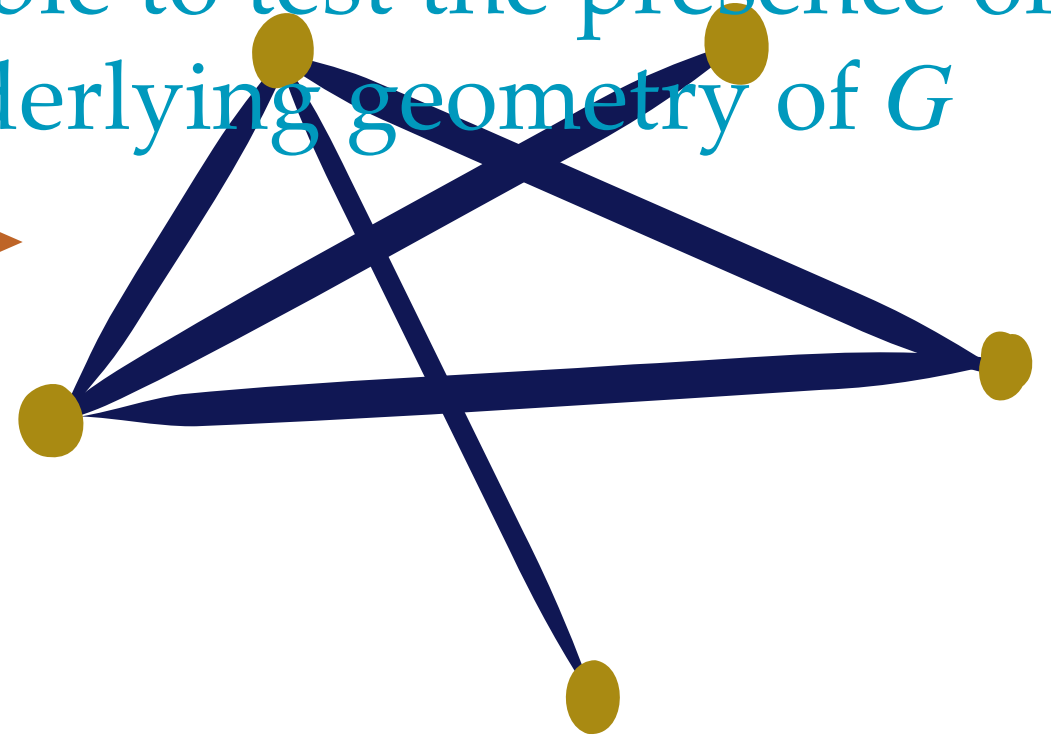
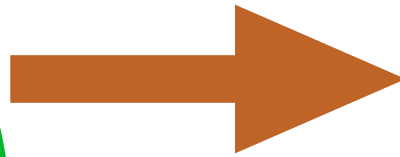


# Question 0 about random geometric graphs

$$G \sim \text{Geo}_d(n, p)$$



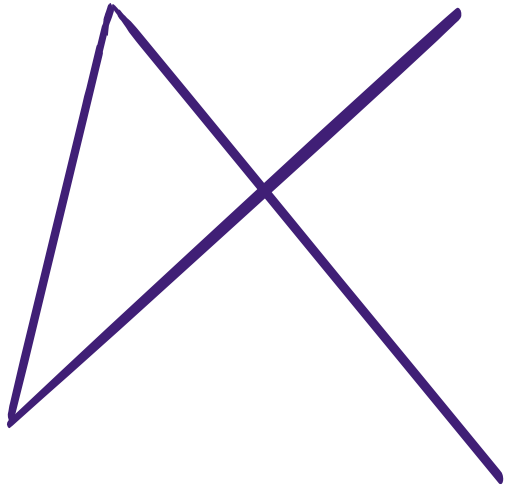
Is it possible to test the presence of the underlying geometry of  $G$





$$G \sim \mathbf{G}(n, p)$$

$$\forall xy \in \binom{[n]}{2} : xy \in G \text{ w.p. } p$$

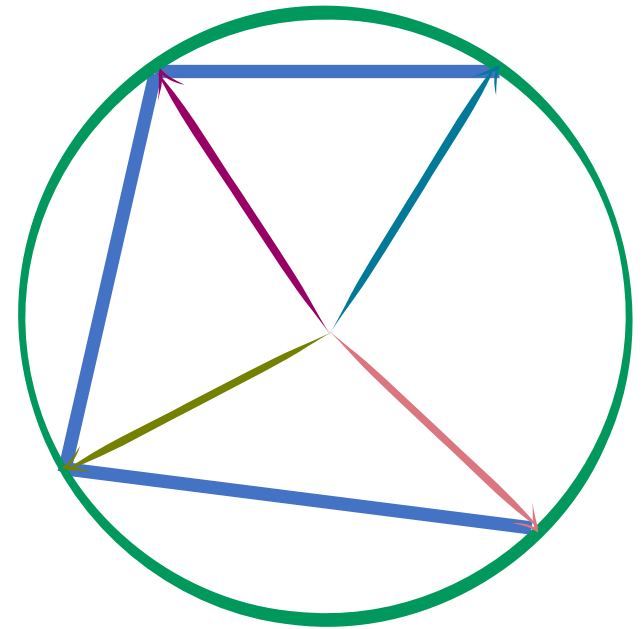


vs.

$$G \sim \mathbf{Geo}_d(n, p)$$

$$v_1, \dots, v_n \sim \mathbb{S}^{d-1}$$

$$G = \{xy : \langle v_x, v_y \rangle \geq \tau(p)\}$$



Given  $G$  distinguish between models

# Known

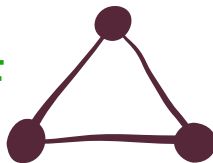
Distinguish  $G(n, p)$  vs.  $\text{Geo}_d(n, p)$

Bubeck-Ding-Eldan-Rácz'16

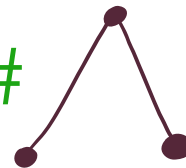
when  $d \ll H(p)^3 n^3$

distinguishable via

$c_1 \cdot \#$



$+c_2 \cdot \#$



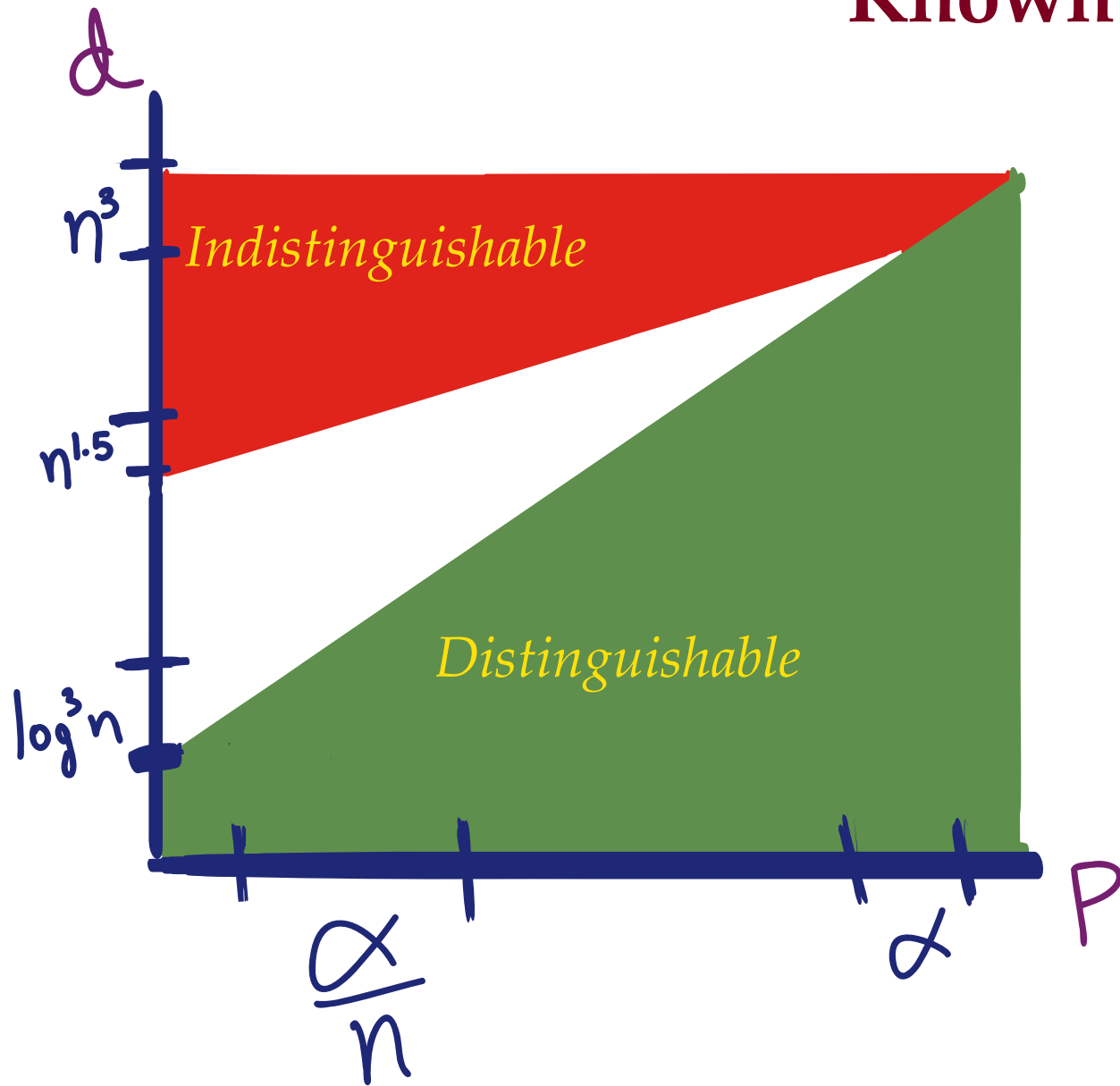
$+c_3 \cdot \#$



Brennan-Bresler-Nagaraj'20

indistinguishable when  $d \gg H(p)^3$ , distinguishable when  $d \ll H(p)^3 n^{7/2}$

# Known



Distinguishable when

$$d \ll H(p)^3 n^3$$

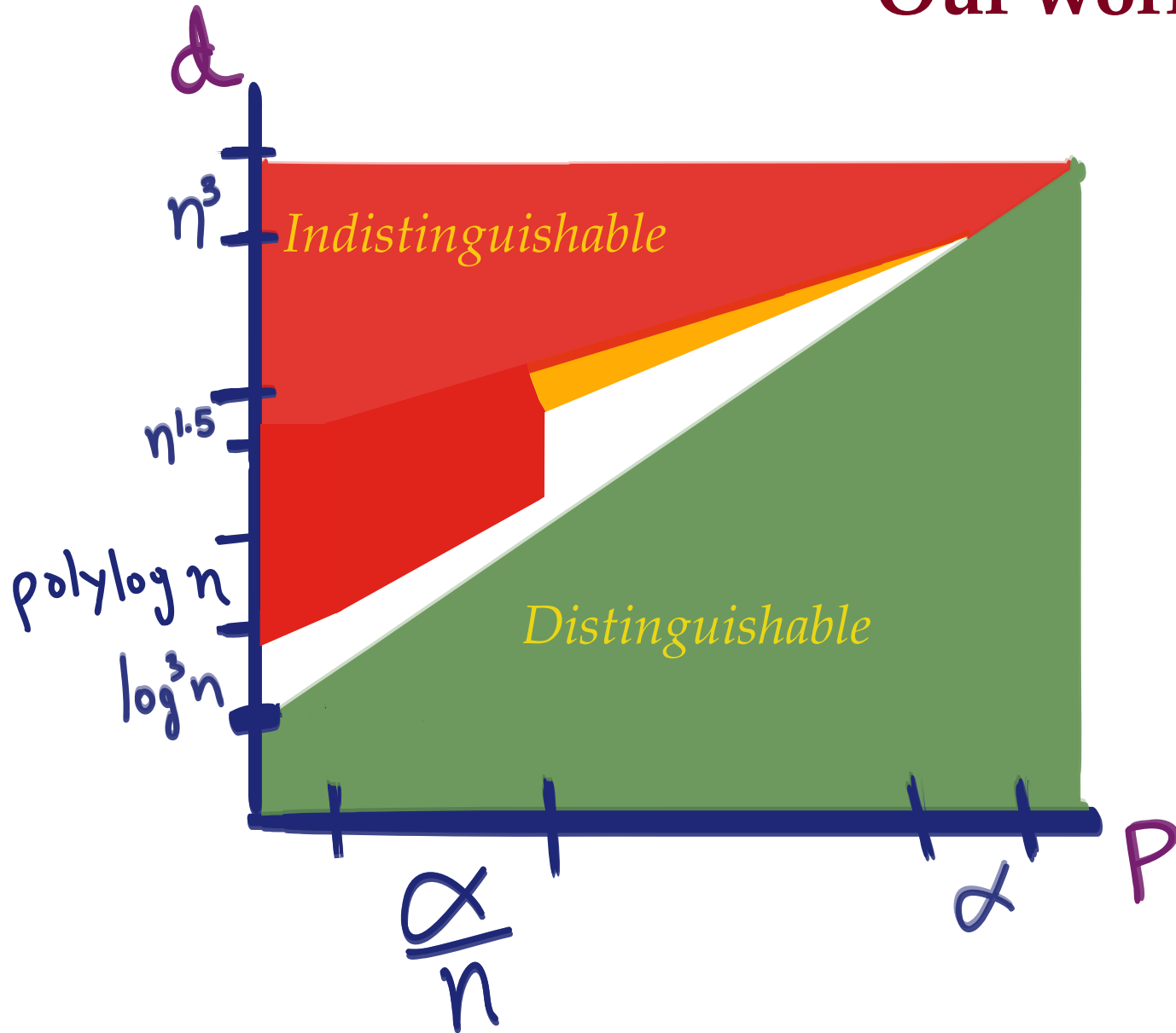
when  $p = \frac{\alpha}{n}$ ,  $d \ll \log^3 n$

Indistinguishable when

$$d \gg \min\{H(p)n^3, H(p)^2 n^{7/2}\}$$

when  $p = \frac{\alpha}{n}$ ,  $d \gg n^{1.5}$

# Our work



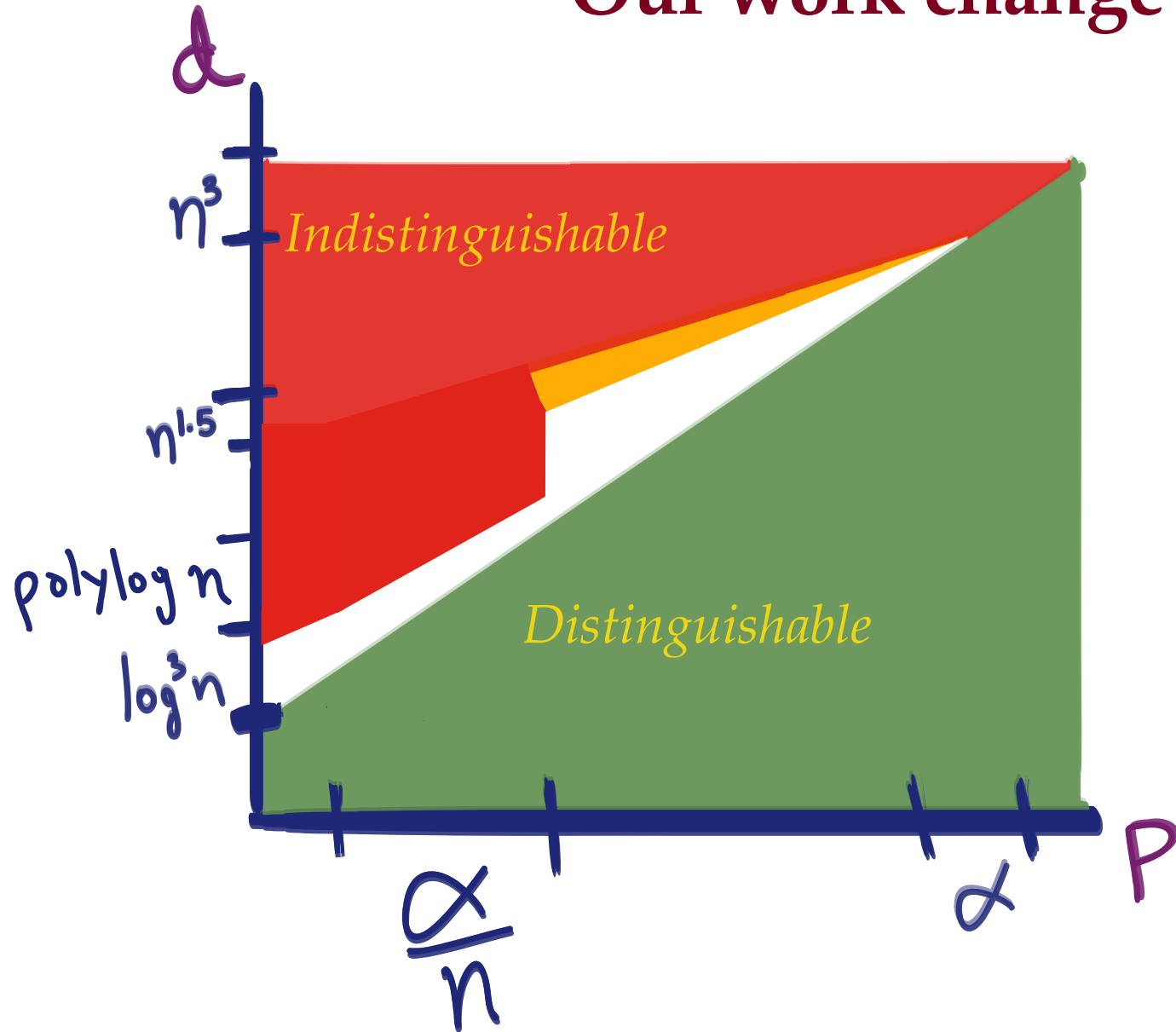
L-Mohanty-Schramm-Yang'22:

Indistinguishable when

$$d \gg H(p)^2 n^3 \text{ for all } p$$

$$\text{when } p = \frac{\alpha}{n}, d \gg \log^{36} n$$

# Our work change the figure



L-Mohanty-Schramm-Yang'22:

Indistinguishable when

$$d \gg H(p)^2 n^3 \text{ for all } p$$

Loose by  $\frac{1}{H(p)}$  factor

$$\text{when } p = \frac{\alpha}{n}, d \gg \log^{36} n$$

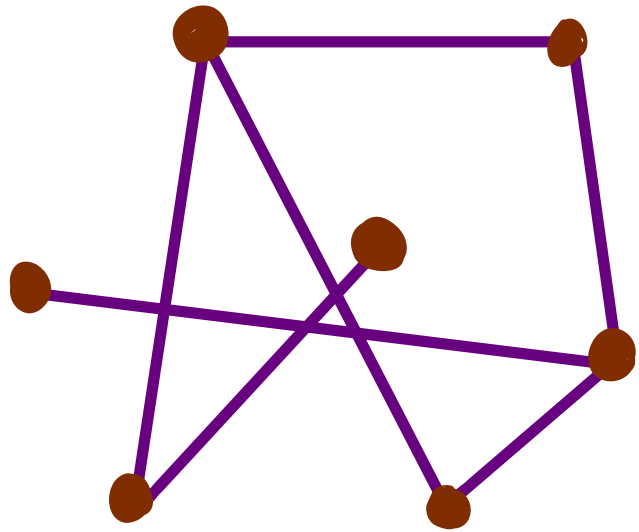
Tight up to  $\text{polylog}(n)$  factors

Theorem [LMSY'22]: Let  $p = \frac{\alpha}{n}$ ,  $\alpha = \Theta(1)$ ,  $d \gg \log^{36} n$

$$d_{\text{TV}}(\mathbf{G}(n, p), \mathbf{Geo}_d(n, p)) \leq o_n(1)$$

To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

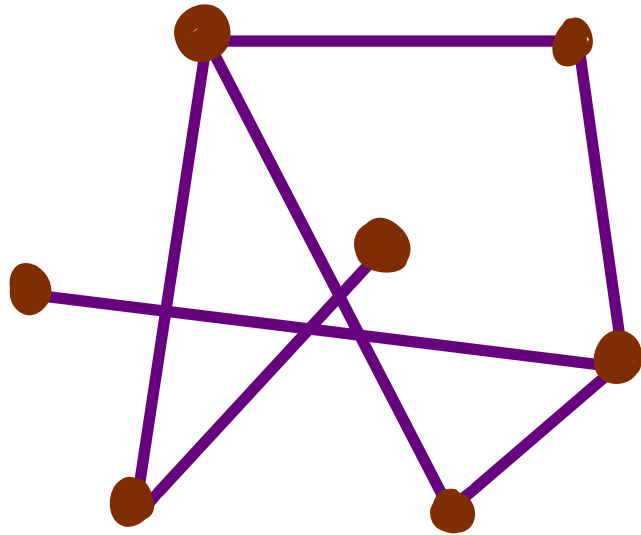
Views of sampling graphs



$G_{\leq t}$

To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

Views of sampling graphs



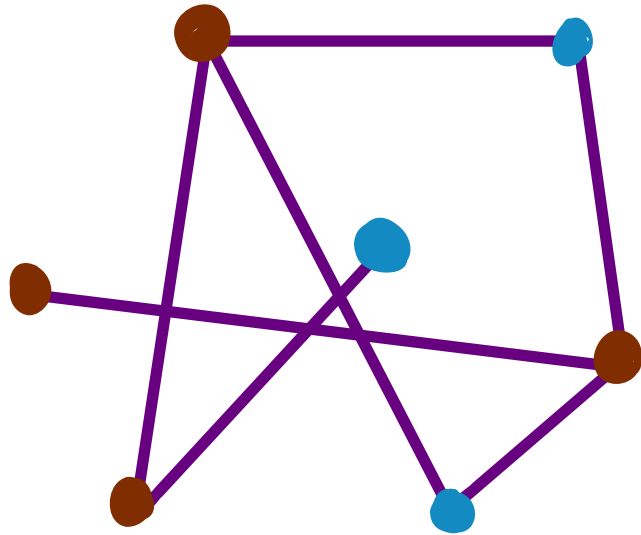
●  $t + 1$

$G_{\leq t}$



To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

Views of sampling graphs



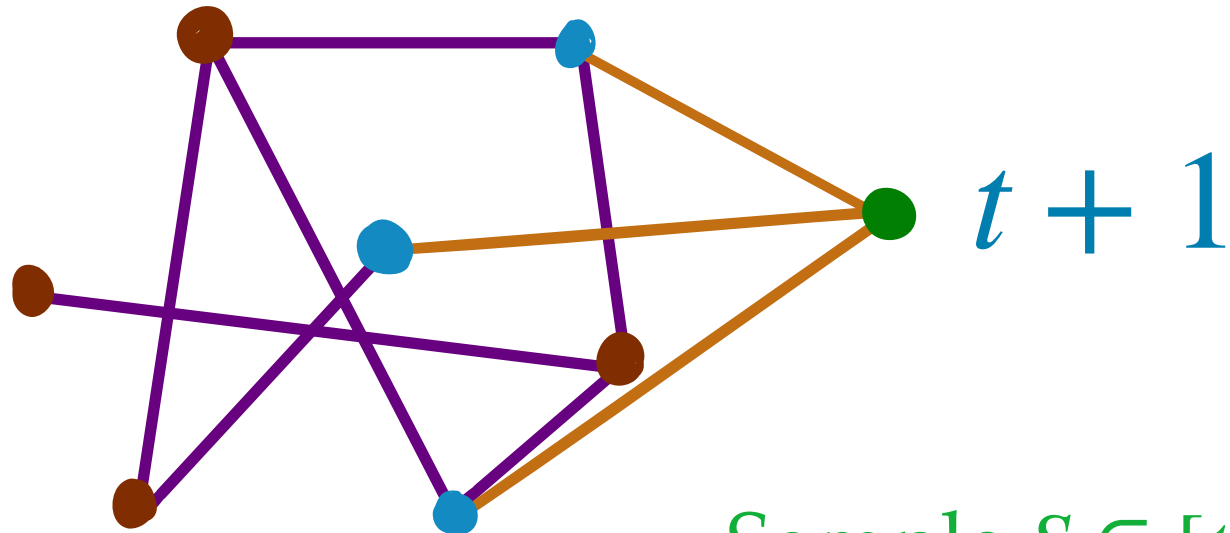
$G_{\leq t}$

●  $t + 1$

Sample  $S \subseteq [t] \sim \begin{cases} \text{Nbr}(\mathbf{G}(n, p)) \\ \text{Nbr}(\text{Geo}_d(n, p)) \mid G_{\leq t} \end{cases}$

To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

Views of sampling graphs



$G_{\leq t+1}$

Sample  $S \subseteq [t] \sim \begin{cases} \text{Nbr}(\mathbf{G}(n, p)) \\ \text{Nbr}(\text{Geo}_d(n, p)) \mid G_{\leq t} \end{cases}$

Connect  $t + 1$  to  $S$

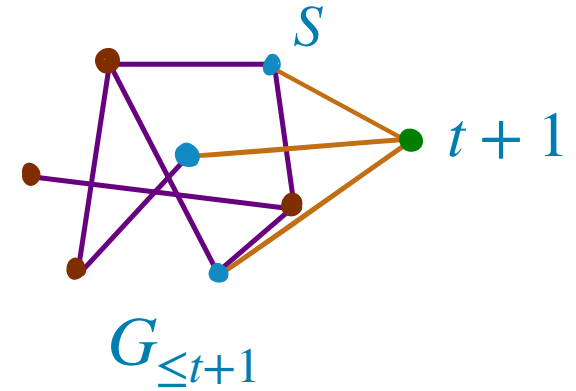
To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

Views of sampling graphs

Nbr( $\mathbf{G}(n, p)$ )

Choose each  $i \in [t]$  independently with probability  $p$

$S$  chosen with probability  $p^{|S|}(1-p)^{t-|S|}$



To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

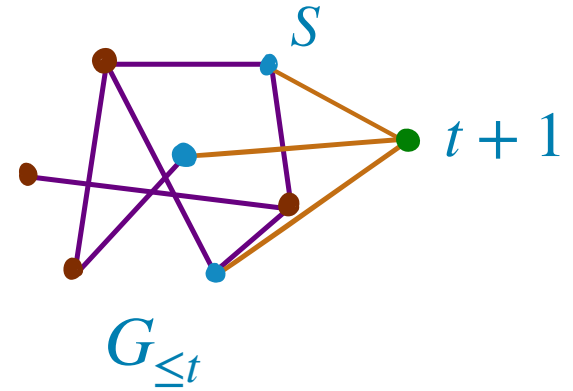
Views of sampling graphs

$\text{Nbr}(\text{Geo}_d(n, p)) \mid G_{\leq t}$

$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid \text{gg}(v_1, \dots, v_t) = G_{\leq t}$$

$$v_{t+1} \sim \mathbb{S}^{d-1}$$

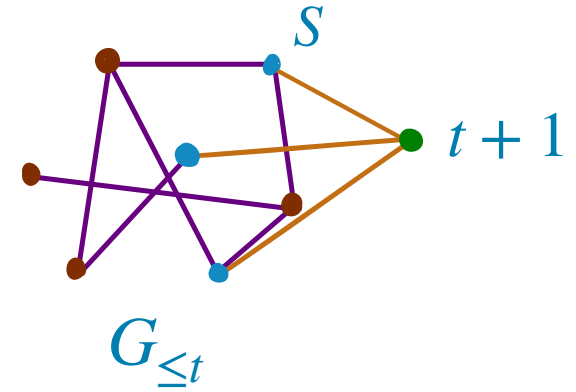
$$S := \{i \in [t] : \langle v_i, v_{t+1} \rangle \geq \tau(p)\}$$



To prove:  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p)) \leq o(1)$

By standard argument suffices to show

W.h.p.  $G_{\leq t} \sim \text{Geo}_d(n, p), S \sim \text{Nbr}(\mathbf{G}(n, p))$



$$\frac{\Pr_{\text{Nbr}(\text{Geo}_d(n, p)) | G_{\leq t}}[S]}{\Pr_{\text{Nbr}(\mathbf{G}(n, p))}[S]} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$$

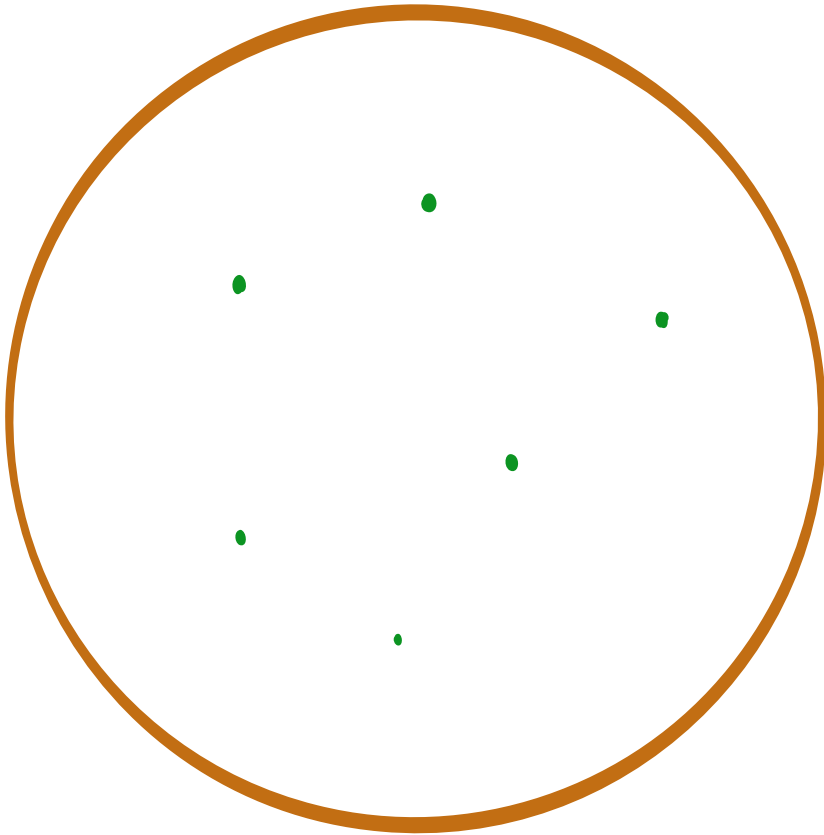
Standard argument = Pinsker's inequality + tensorization of relative entropy

What is  $\Pr$   $[S]$ ?  
 $\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}$

To prove:  $d_{\text{TV}}(\text{G}(n,p), \text{Geo}_d(n,p)) \leq o(1)$

Suffices to show

$$\frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}}[S]}{\Pr_{\text{Nbr}(\text{G}(n,p))}[S]} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$$



$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(v_1, \dots, v_t) = G_{\leq t}$$

What is  $\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}}[S]$ ?

To prove:  $d_{\text{TV}}(\text{G}(n,p), \text{Geo}_d(n,p)) \leq o(1)$

Suffices to show

$$\frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}}[S]}{\Pr_{\text{Nbr}(\text{G}(n,p))}[S]} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$$



$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid \text{gg}(v_1, \dots, v_t) = G_{\leq t}$$

Choose  $S$  with probability  $\text{Area}(\text{Splinter}(S))$

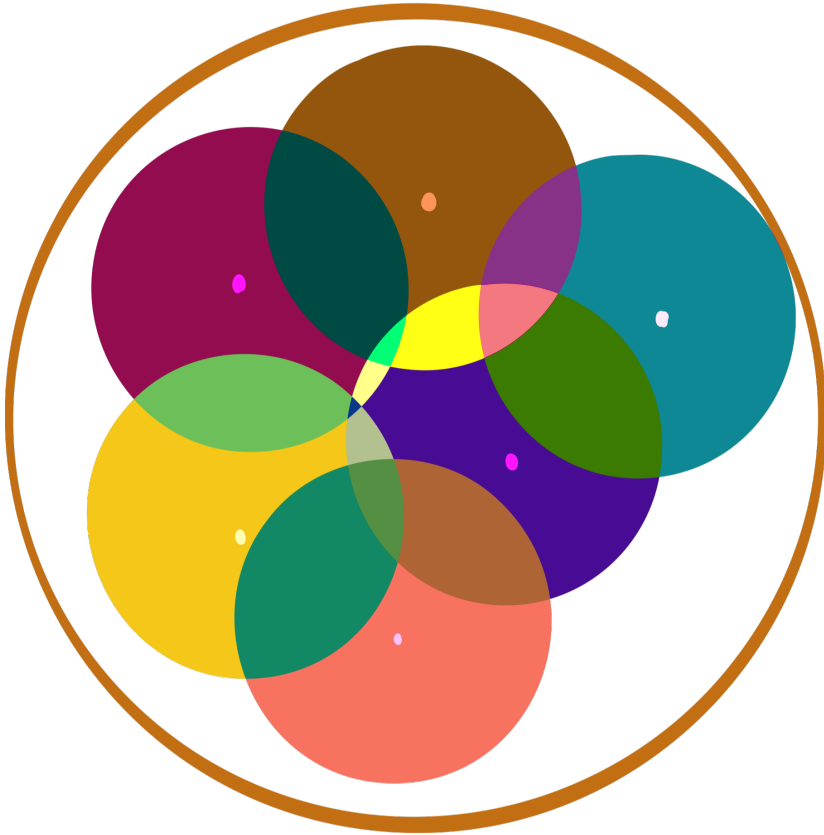
$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$

What is  $\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S]$ ?

To prove:  $d_{\text{TV}}(\text{G}(n,p), \text{Geo}_d(n,p)) \leq o(1)$

Suffices to show

$$\frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S]}{\Pr_{\text{Nbr}(\text{G}(n,p))} [S]} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$$



$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid \text{gg}(v_1, \dots, v_t) = G_{\leq t}$$

Choose  $S$  with probability  $\text{Area}(\text{Splinter}(S))$

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$

$$\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S] = \mathbb{E}_{v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid \text{gg}(v_1, \dots, v_t) = G_{\leq t}} \left[ \text{Area}(\text{Splinter}(S)) \right]$$



What is  $\Pr_{\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}} [S]$ ?

$$\Pr_{\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}} [S] = \mathbb{E}_{v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(v_1, \dots, v_t) = G_{\leq t}} [\text{Area}(\text{Splinter}(S))]$$

$$\Pr_{\text{Nbr}(G(n,p))} [S] = \mathbb{E}_{v_1, \dots, v_t \sim \mathbb{S}^{d-1}} [\text{Area}(\text{Splinter}(S))]$$

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$



First attempt:

Try to show  $\text{Area}(\text{Splinter}(S))$  concentrates under the randomness of  $v_1, \dots, v_t$

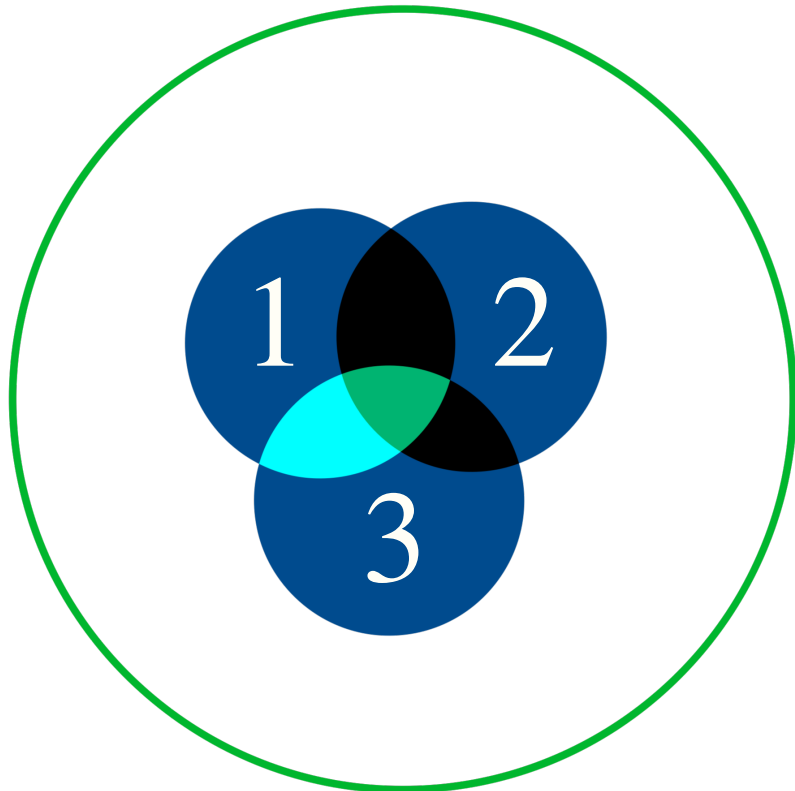
What is  $\Pr [S]?$   
 $\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}$

Example:  $S = \{1,2\}$       $\bar{S} = \{3\}$

$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(G_{\leq t}) = G$$

Choose  $S$  with probability  $\text{Area}(\text{Splinter}(S))$

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$



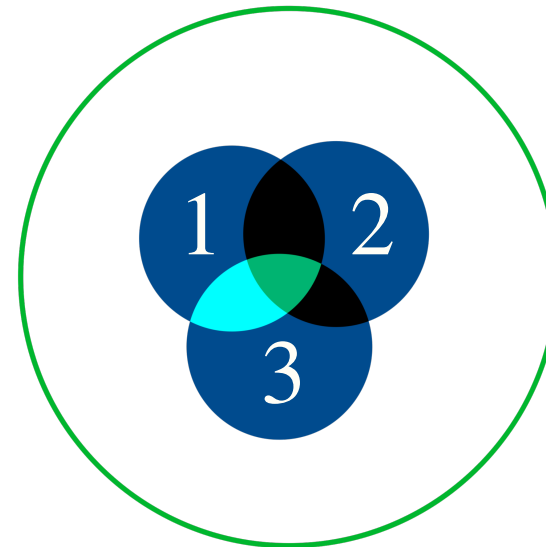
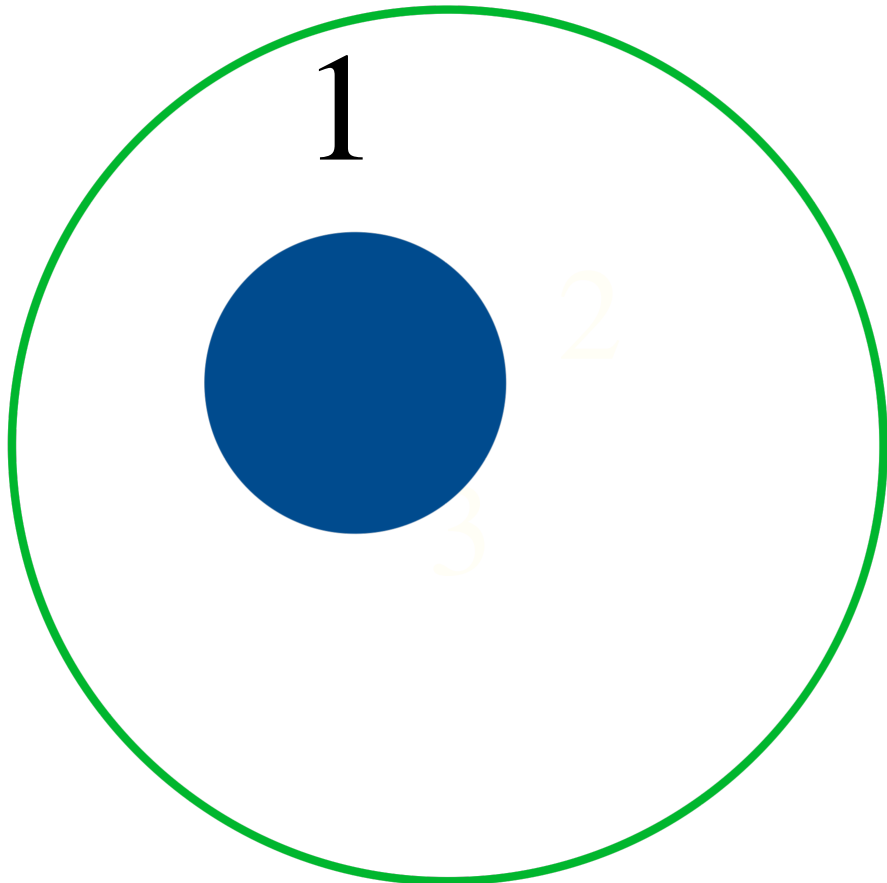
What is  $\Pr [S]?$   
 $\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}$

Example:  $S = \{1,2\}$       $\bar{S} = \{3\}$

$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(G_{\leq t}) = G$$

Choose  $S$  with probability  $\text{Area}(\text{Splinter}(S))$

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$



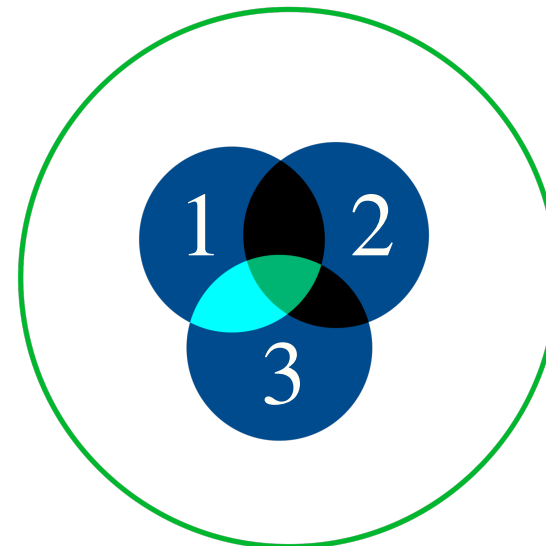
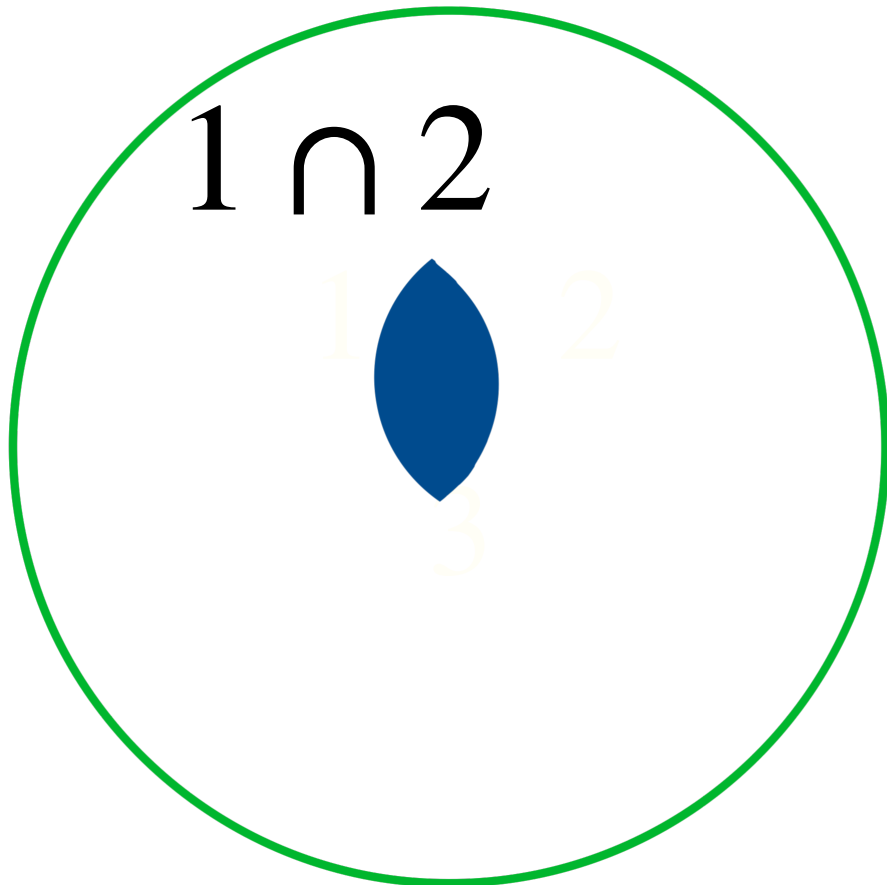
What is  $\Pr [S]?$   
 $\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}$

Example:  $S = \{1,2\}$       $\bar{S} = \{3\}$

$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(G_{\leq t}) = G$$

Choose  $S$  with probability  $\text{Area}(\text{Splinter}(S))$

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$



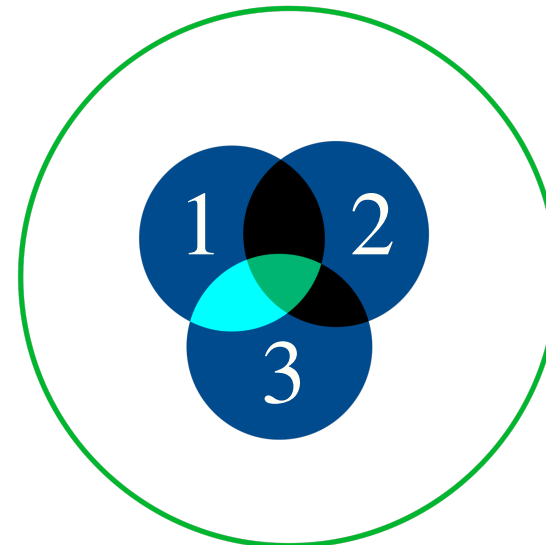
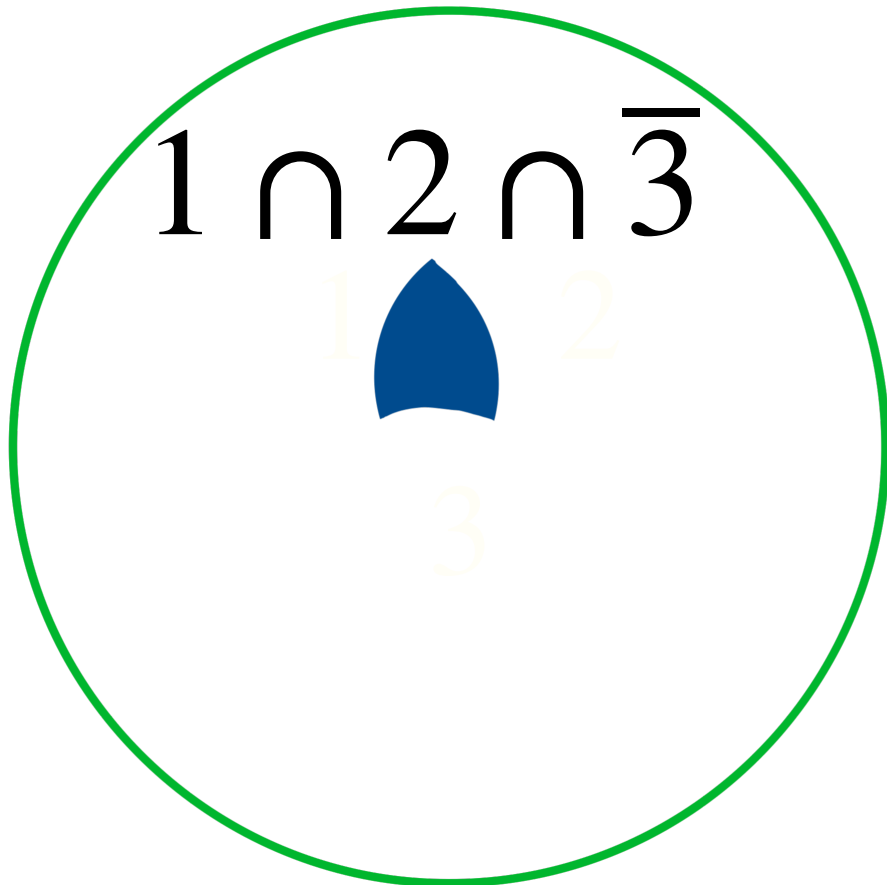
What is  $\Pr [S]?$   
 $\text{Nbr}(\text{Geo}_d(n,p)) | G_{\leq t}$

Example:  $S = \{1,2\}$       $\bar{S} = \{3\}$

$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(G_{\leq t}) = G$$

Choose  $S$  with probability  $\text{Area}(\text{Splinter}(S))$

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$



# Concentration of Area (Splinter( $S$ ))

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$



# Concentration of Area (Splinter( $S$ ))

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$

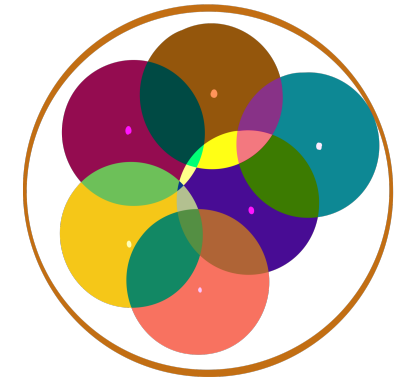
More general question  $L \subseteq \mathbb{S}^{d-1}$ ,  $w \sim \mathbb{S}^{d-1}$



# Concentration of Area (Splinter( $S$ ))

$$\text{Splinter}(S) := \bigcap_{i \in S} \text{cap}(v_i) \cap \bigcap_{i \in \bar{S}} \overline{\text{cap}(v_i)}$$

More general question  $L \subseteq \mathbb{S}^{d-1}$ ,  $w \sim \mathbb{S}^{d-1}$



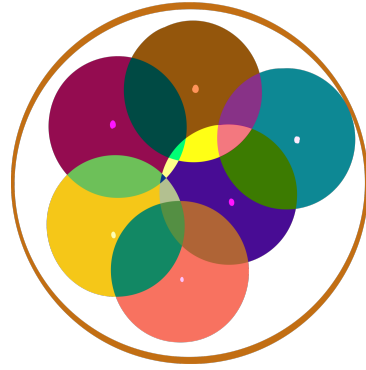
Concentration of  $\text{Area}(L \cap \text{cap}(w))$ ?

$$\mathbb{E}[\text{Area}(L \cap \text{cap}(w))] = p \cdot \text{Area}(L)$$



$$L \subseteq \mathbb{S}^{d-1}, w \sim \mathbb{S}^{d-1}$$

Concentration of Area (Splinter( $S$ ))



$$\text{Theorem: } \text{Area}(L \cap \text{cap}(w)) \in \left( 1 \pm \tilde{O} \left( \sqrt{\frac{\log \frac{1}{\text{Area}(L)}}{d}} \right) \right) \cdot p \cdot \text{Area}(L)$$

$$\text{Area}(L \cap \overline{\text{cap}}(w)) \in \left( 1 \pm \tilde{O} \left( p \sqrt{\frac{\log \frac{1}{\text{Area}(L)}}{d}} \right) \right) \cdot (1 - p) \cdot \text{Area}(L)$$



For most  $S$  :  $\text{Area}(\text{Splinter}(S)) \in \left( 1 \pm \tilde{o}\left(\frac{1}{\sqrt{d}}\right) \right) \Pr_{\text{Nbr}(\text{G}(n,p))} [S]$

$$\text{Implies } \frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S]}{\Pr_{\text{Nbr}(\text{G}(n,p))} [S]} = 1 \pm \tilde{o}\left(\frac{1}{\sqrt{d}}\right)$$

$$\text{We need } \frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S]}{\Pr_{\text{Nbr}(\text{G}(n,p))} [S]} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$$

Requirement met when  $d \gg n \text{ polylog } n$

But we care about  $d \geq \text{polylog } n$

For most  $S$  :  $\text{Area}(\text{Splinter}(S)) \in \left( 1 \pm \tilde{o}\left(\frac{1}{\sqrt{d}}\right) \right) \Pr_{\text{Nbr}(G(n,p))} [S]$

We need  $\frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S]}{\Pr_{\text{Nbr}(G(n,p))} [S]} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$

$$\Pr_{\text{Nbr}(\text{Geo}_d(n,p))|G_{\leq t}} [S] = \mathbb{E}_{v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(v_1, \dots, v_t) = G_{\leq t}} \left[ \text{Area}(\text{Splinter}(S)) \right]$$



Studied concentration of  $\text{Area}(\text{Splinter}(S))$  under randomness of  $v_1, \dots, v_t$

$\mathbb{E}_{v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(v_1, \dots, v_t) = G_{\leq t}}$  is key

Need to study concentration of  $\mathbb{E}_{v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(v_1, \dots, v_t) = G_{\leq t}} [\text{Area}(\text{Splinter}(S))]$   
under randomness of  $G_{\leq t}$

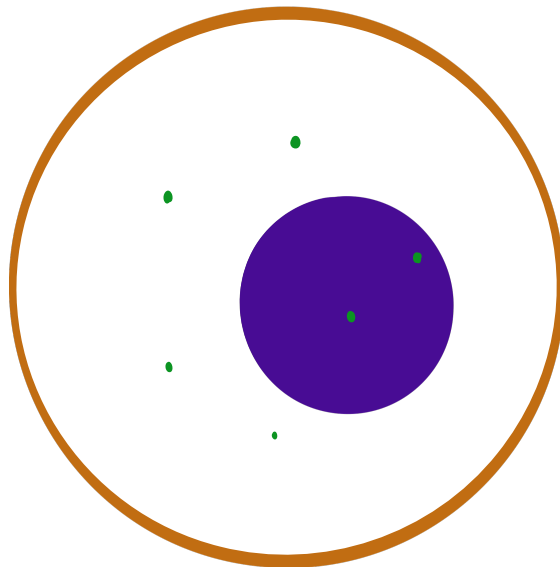
New goal: show  $\frac{\Pr_{\text{Nbr}(t+1)|G_{\leq t}}[\supseteq S]}{p^{|S|}} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$  w.h.p. over  $G_{\leq t}$

$$\text{Nbr}(t+1) | G_{\leq t}$$

$$v_{t+1} \sim \mathbb{S}^{d-1}$$

$$v_1, \dots, v_t \sim \mathbb{S}^{d-1} | \text{gg}(v_1, \dots, v_t) = G_{\leq t}$$

$$\text{Nbr}(t+1) := \{i : v_i \in \text{cap}(v_{t+1})\}$$



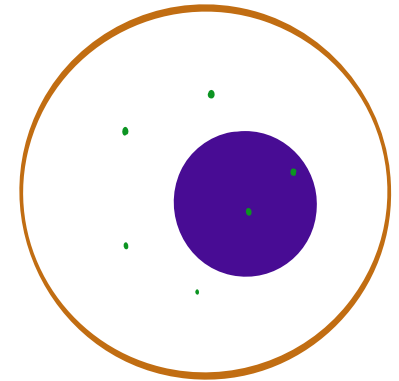
New goal: show  $\frac{\Pr_{\text{Nbr}(t+1)|G_{\leq t}}[\supseteq S]}{p^{|S|}} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$  w.h.p. over  $G_{\leq t}$

Suppose  $(v_i)_{i \in S} | G_{\leq t}$  i.i.d. & uniform

$$\Pr_{\text{Nbr}(t+1)|G_{\leq t}}[\supseteq S] = p^{|S|}$$

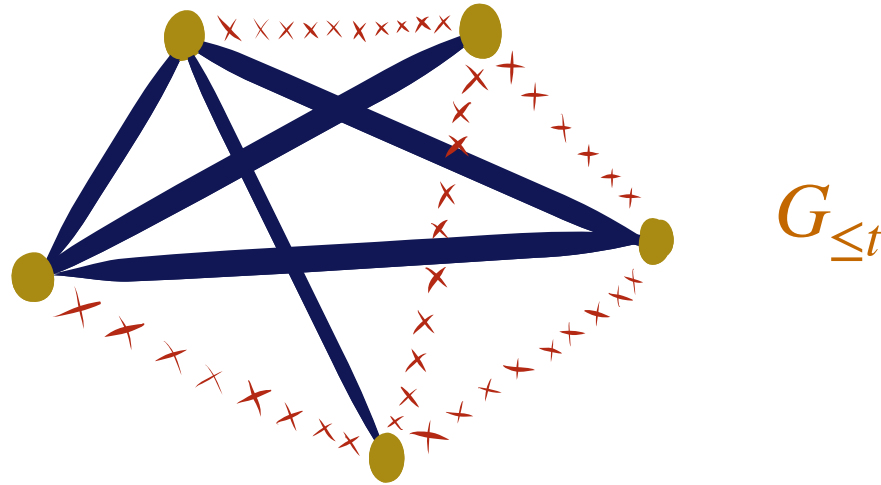
Will show  $(v_i)_{i \in S} | G_{\leq t}$  w.h.p i.i.d. and approximately uniform

$\text{Nbr}(t+1) := \{i : v_i \in \text{cap}(v_{t+1})\}$



Goal: understand marginals  $(v_i)_{i \in S} | G_{\leq t}$

Goal: understand marginals  $(v_i)_{i \in S} \mid G_{\leq t}$



Constraint Satisfaction Problem on  $G_{\leq t}$

$$\forall ij \text{ edge: } \langle v_i, v_j \rangle \geq \tau(p)$$

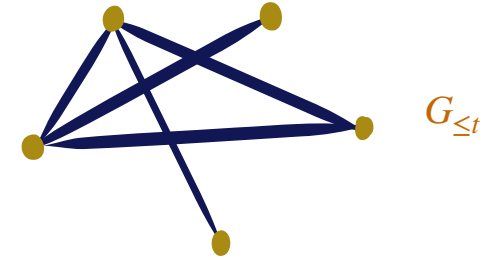
$$\forall ij \text{ non edge: } \langle v_i, v_j \rangle < \tau(p)$$

$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid G_{\leq t}$  uniform on solutions to above CSP

Constraint Satisfaction Problem on  $G_{\leq t}$

$$\forall ij \text{ edge: } \langle v_i, v_j \rangle \geq \tau(p)$$

~~$$\forall ij \text{ non edge: } \langle v_i, v_j \rangle < \tau(p)$$~~



$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid G_{\leq t}$  uniform on solutions to above CSP

Goal: understand marginals  $(v_i)_{i \in S} \mid G_{\leq t}$

Technical simplification for talk: Drop non-edge constraints!

## Constraint Satisfaction Problem on $G_{\leq t}$

$$\forall ij \text{ edge: } \langle v_i, v_j \rangle \geq \tau(p)$$

$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid G_{\leq t}$  uniform on solutions to above CSP

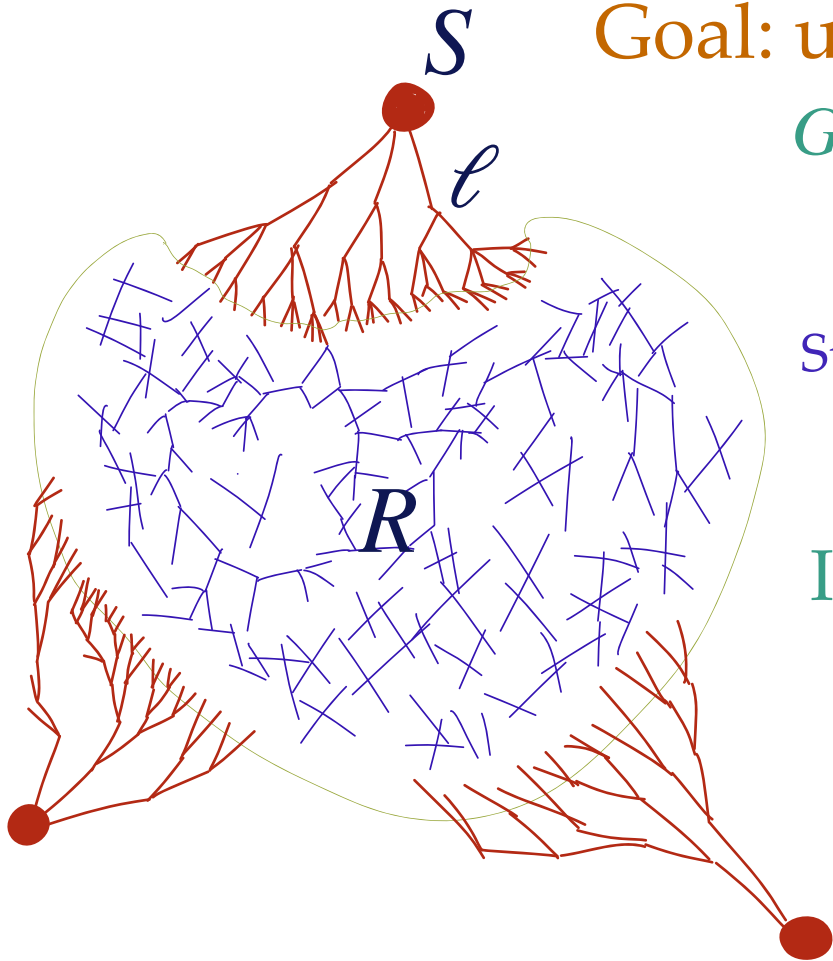
**Goal:** understand marginals  $(v_i)_{i \in S} \mid G_{\leq t}$

$G_{\leq t}$  sparse, locally-treelike      vertices in  $S$  pairwise far

Strategy: show for a typical fixing of vectors in purple part  
 $(v_i)_{i \in S}$  independent and uniform

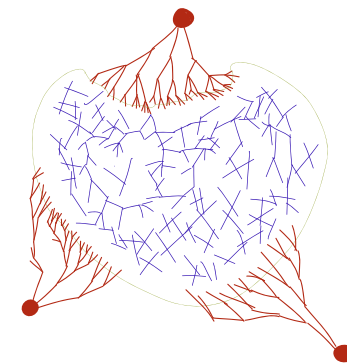
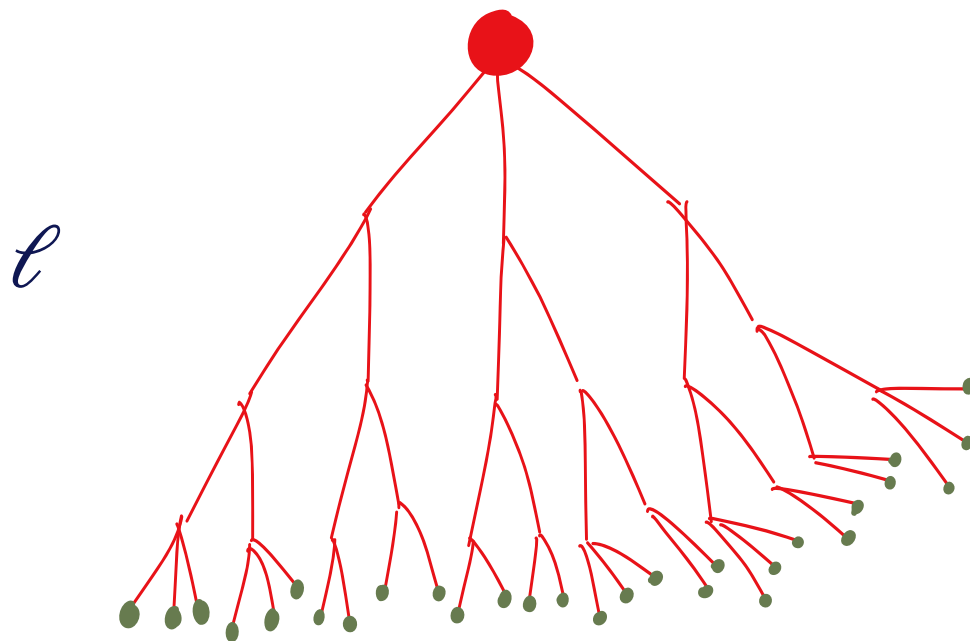
Independence for free because pieces are disjoint!

Need to show (approximate) uniformity of  $v_i$





Need to show (approximate) uniformity of  $v_i$



Green leaves receive "typical" vector assignment

Will show: Red root conditioned on leaves is approximately uniform

Strategy: compute distribution of  $v_{\text{root}} \mid (v_i)_{i \in \text{Leaves}}$  via belief propagation

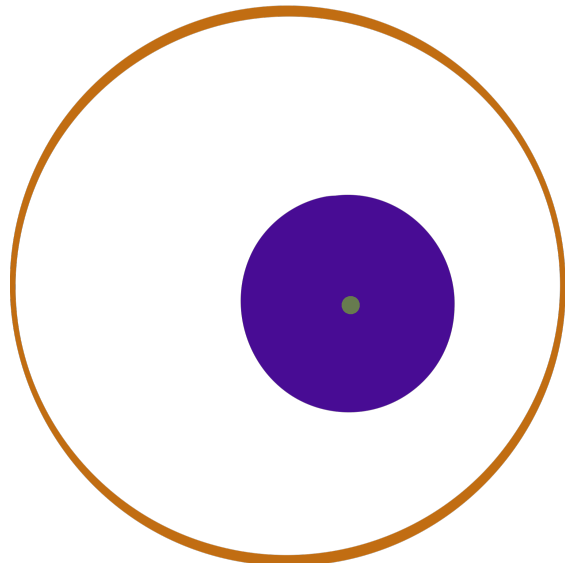
# Illustration on special case: tree is length- $\ell$ path

Green leaf has vector  $v_{\text{leaf}}$

Will show:  $v_{\text{root}} | v_{\text{leaf}}$  is approximately uniform

Distribution  $f_{\text{Parent}(x)} | f_x$   $\left\{ \begin{array}{l} \text{Sample } w \sim f_x \\ \text{Walk to random vector in cap}(w) \end{array} \right.$

$\ell$



$P$  Markov operator

$$f_y := \text{PDF of } v_y \quad f_{\text{Parent}(x)} = P \cdot f_x$$

$$f_{\text{root}} = P^\ell \cdot f_{\text{Leaf}}$$

Uniform stationary for  $P$

Exhibit contraction properties of  $P$

# Exhibit contraction properties of $P$

Key lemma: For "smooth"  $f$

"nice" means density  $\leq 1/p$

$$d_{\text{TV}}(Pf, \text{Unif}) \leq \tilde{O}\left(\frac{1}{\sqrt{d}}\right) d_{\text{TV}}(f, \text{Unif})$$

$$d_{\text{TV}}(f_{\text{root}}, \text{Unif}) \leq \tilde{O}\left(\frac{1}{\sqrt{d}}\right)^{\ell-1}$$

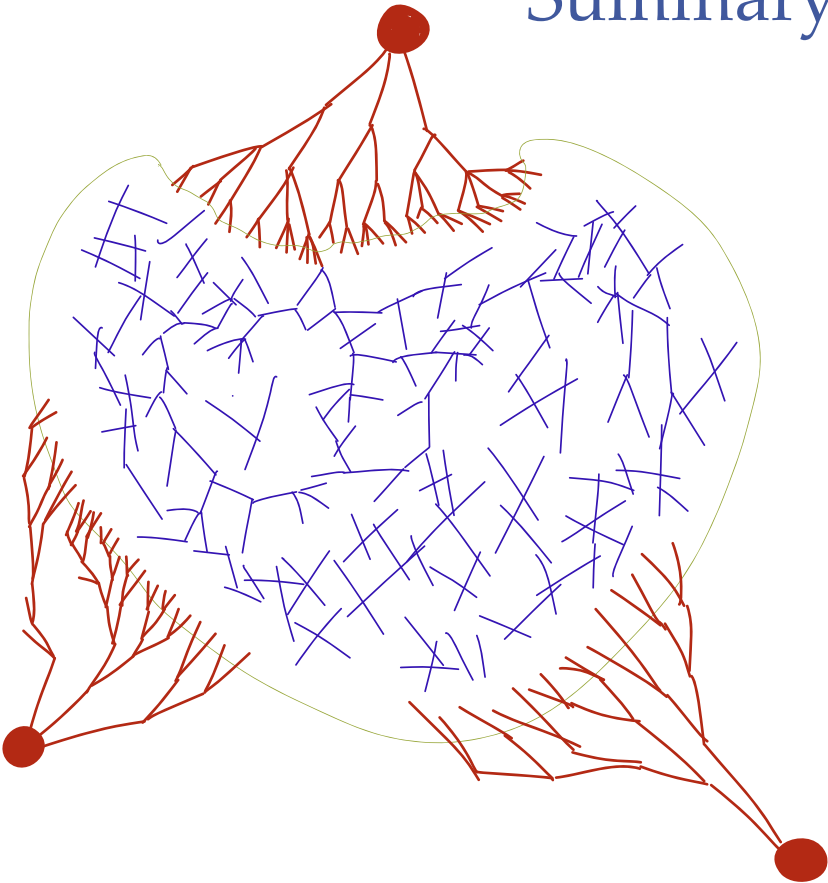
when  $d \geq \text{polylog } n$ , set  $\ell = \frac{\log n}{\log \log n}$ , and then TV distance is  $o\left(\frac{1}{\sqrt{n}}\right)$

## Constraint Satisfaction Problem on $G_{\leq t}$

$$\forall ij \text{ edge: } \langle v_i, v_j \rangle \geq \tau(p)$$

$v_1, \dots, v_t \sim \mathbb{S}^{d-1} \mid G_{\leq t}$  uniform on solutions to above CSP

Summary:  $(v_i)_{i \in S}$  independent and  $o\left(\frac{1}{\sqrt{n}}\right)$ -close to uniform



$$\frac{\Pr_{\text{Nbr}(\text{Geo}_d(n,p)) \mid G_{\leq t}}[\supseteq S]}{p^{|S|}} = 1 \pm o\left(\frac{1}{\sqrt{n}}\right)$$

$$d_{\text{TV}}(\mathbf{G}(n,p), \text{Geo}_d(n,p)) \leq o(1)$$

*Thank you! Questions?*

