A matrix Chernoff bound for strongly Rayleigh distributions

and

spectral sparsifiers from a few random spanning trees

Speaker	**Rasmus Kyng**	Harvard University
Joint work	with Zhao Song	UT Austin

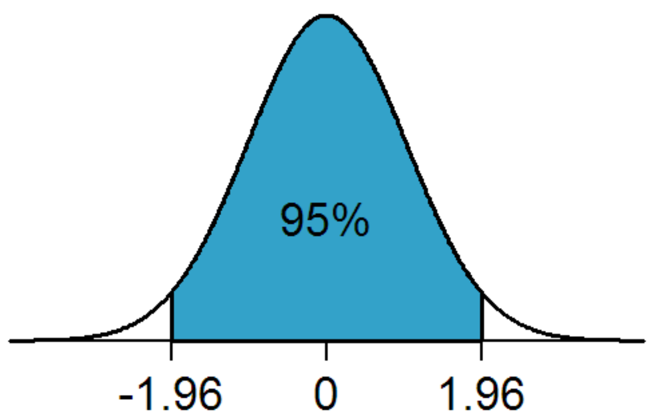
October 2018

Concentration of scalar random variables

Independent random $X_i \in \mathbb{R}$

$X = \sum_i X_i$

Is $X \approx \mathbb{E}X$ with high probability?



Concentration of scalar random variables

Chernoff inequality

Independent random $X_i \in \mathbb{R}_{\geq 0}$

$X = \sum_i X_i$

1. $\mathbb{E}X = \mu$
2. $|X_i| \leq r$

E.g. if $\varepsilon = 0.5, r = 1$ and $\mu = 10\log(1/\tau)$

gives

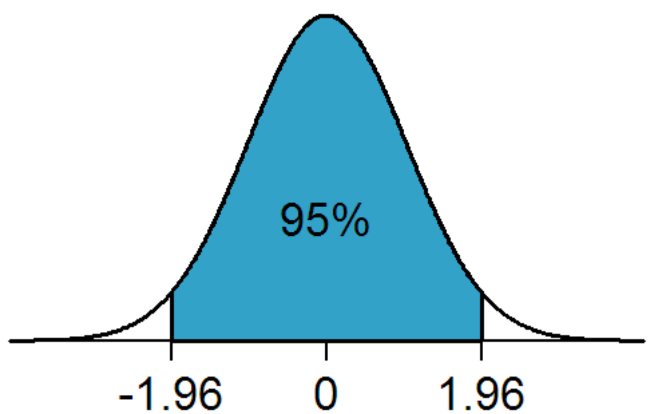
$$\mathbb{P}[|X - \mu| > \varepsilon\mu] \leq 2\exp\left(-\frac{\mu\varepsilon^2}{r(2+\varepsilon)}\right)$$

Concentration of random matrices

Independent random $\boldsymbol{X}_i \in \mathbb{R}^{d \times d}$

$\boldsymbol{X} = \sum_i \boldsymbol{X}_i$

Is $\boldsymbol{X} \approx \mathbb{E}\boldsymbol{X}$ with high probability?



Concentration of random matrices

Matrix Chernoff [Tropp '11]

Independent random $\boldsymbol{X}_i \in \mathbb{R}^{d \times d}$, positive semi-definite

$\boldsymbol{X} = \sum_i \boldsymbol{X}_i$

1. $\|\mathbb{E}\boldsymbol{X}\| = \mu$
2. $\|\boldsymbol{X}_i\| \leq r$

E.g. if $\varepsilon = 0.5, r = 1$ and $\mu = 10 \log(d/\tau)$

gives

$$\mathbb{P}[\|\boldsymbol{X} - \mathbb{E}\boldsymbol{X}\| > \varepsilon] \leq d \, 2\exp\left(-\frac{\mu\varepsilon^2}{r(2+\varepsilon)}\right)$$

[Rudelson '99, Ahlswede-Winter '02]

What if variables are not independent?

$X \in \{0,1\}$ random variable

$Y = X$

$X + Y$ not concentrated, 0 or 2

$Z = 1 - X$

$X + Z$ very concentrated, always 1

Negative dependence:

X makes Z less likely and vice versa

What if variables are not independent?

$\xi \in \{0,1\}^m$ random variable

Negative pairwise correlation

For all pairs $i \neq j$

$$\xi(i) = 1 \quad \Rightarrow \quad \text{lower prob. of } \xi(j) = 1$$

Formally $\mathbb{P}[\xi(j) = 1 | \xi(i) = 1] \leq \mathbb{P}[\xi(j) = 1]$

What if variables are not independent?

$\xi \in \{0,1\}^m$ random variable

Negative correlation

For all $S \subseteq [m]$

$\mathbb{P}[\forall i \in S. \xi(i) = 1] \leq \Pi_{i \in S} \mathbb{P}[\xi(i) = 1]$

Can we get a Chernoff bound? Yes.

If ξ AND $\bar{\xi}$ (negated bits) are negatively correlated,

Chernoff-like concentration applies to $\sum_i \xi(i)$

[Goyal-Rademacher-Vempala ‘09, Dubhashi-Ranjan ’98]

Strongly Rayleigh distributions

A class of negatively dependent distributions

[Borcea-Branden-Liggett '09]

$\xi \in \{0,1\}^m$ random variable

Many nice properties

Negative pairwise correlation

Negative association

Closed under conditioning, marginalization

Strongly Rayleigh distributions

A class of negatively dependent distributions

[Borcea-Branden-Liggett '09]

$\xi \in \{0,1\}^m$ random variable

Examples:

Uniformly sampling k items without replacement

Random spanning trees

Determinantal point processes, volume sampling

Symmetric exclusion processes

Strongly Rayleigh distributions

A class of negatively dependent distributions

[Borcea-Branden-Liggett '09]

$\xi \in \{0,1\}^m$ random variable

$\boldsymbol{k}$**-homogeneous Strongly Rayleigh:**

$|\{i : \xi(i) = 1\}| = k$ always

Concentration of random matrices

Strongly Rayleigh matrix Chernoff [K. & Song '18]

Fixed $\boldsymbol{A}_i \in \mathbb{R}^{d\times d}$, positive semi-definite

$\xi \in \{0,1\}^m$ is k-homogeneous strongly Rayleigh

Random $\boldsymbol{X} = \sum_i \xi(i)\boldsymbol{A}_i$

1. $\|\mathbb{E}\boldsymbol{X}\| = \mu$
2. $\|\boldsymbol{A}_i\| \leq r$

E.g. if $\varepsilon = 0.5, r = 1$ and $\mu = 10\log(d/\tau)\log(k)$

gives

$$\mathbb{P}[\|\boldsymbol{X} - \mathbb{E}\boldsymbol{X}\| > \mu\varepsilon] \leq d\, 2\exp\left(-\frac{\mu\varepsilon^2}{r(\log k+\varepsilon)}\right)$$

Scalar version: Peres-Pemantle '14

Concentration of random matrices

Strongly Rayleigh matrix Chernoff [K. & Song '18]

Fixed $\boldsymbol{A}_i \in \mathbb{R}^{d\times d}$, positive semi-definite

$\xi \in \{0,1\}^m$ is k-homogeneous strongly Rayleigh

Random $\boldsymbol{X} = \sum_i \xi(i)\boldsymbol{A}_i$

1. $\|\mathbb{E}\boldsymbol{X}\| = \mu$
2. $\|\boldsymbol{A}_i\| \leq r$

E.g. if $\varepsilon = \log(k), r = 1$ and $\mu = 10\log(d/\tau)$

gives

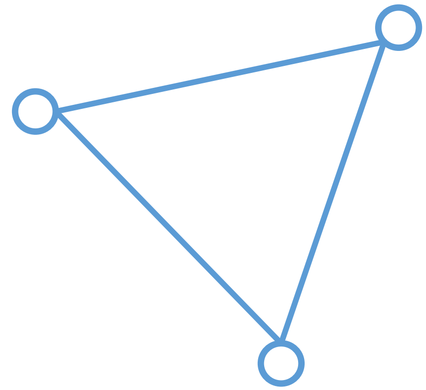
$$\mathbb{P}[\|\boldsymbol{X} - \mathbb{E}\boldsymbol{X}\| > \mu\varepsilon] \leq d\, 2\exp\left(-\frac{\mu\varepsilon^2}{r(\log k+\varepsilon)}\right)$$

Scalar version: Peres-Pemantle '14

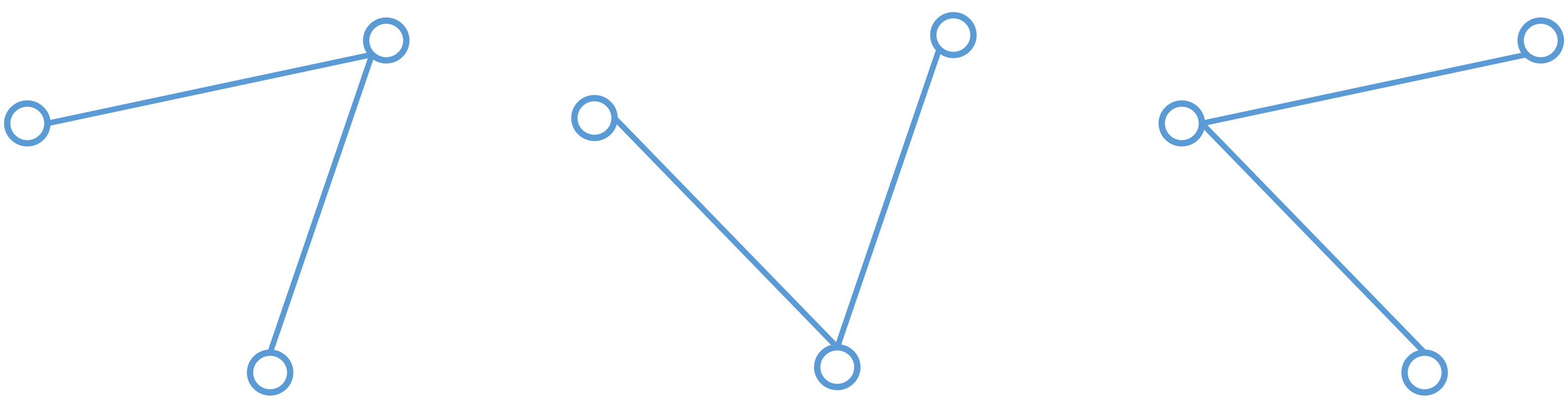
An application:
Graph approximation using random spanning trees

Spanning trees of a graph

Graph $G = (V, E, w)$
Edge weights $\quad w: E \to \mathbb{R}_+$
$n = |V|$

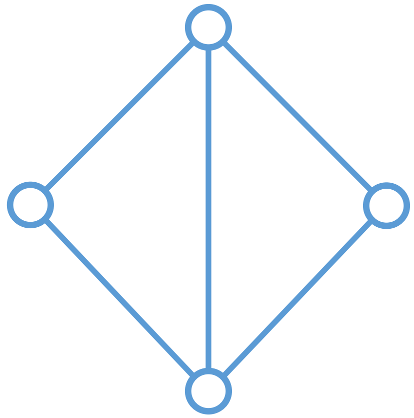


Spanning trees of G

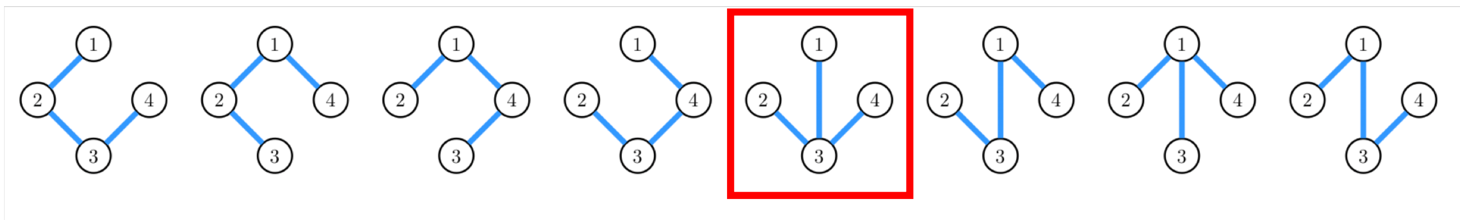


Random spanning trees

Graph G



Tree distribution



Pick a random tree?

Random spanning trees

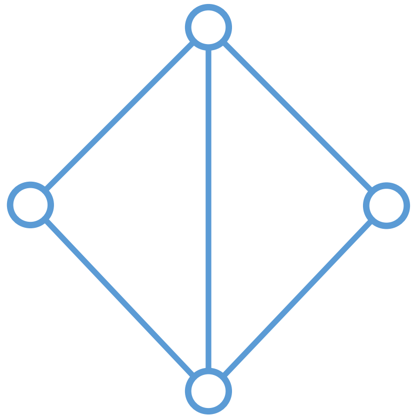
Does the sum of a few random spanning trees resemble the graph?

E.g. is the weight across each cut similar?

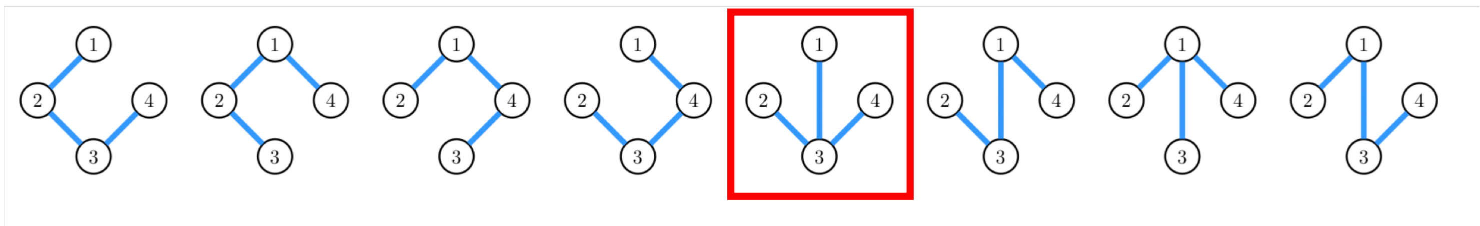
Starter question:
Are the edge weights similar in expectation?

Random spanning trees

Graph G



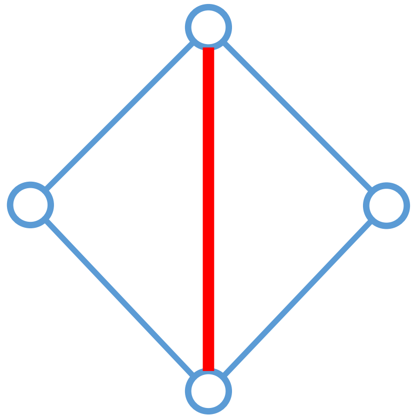
Tree distribution



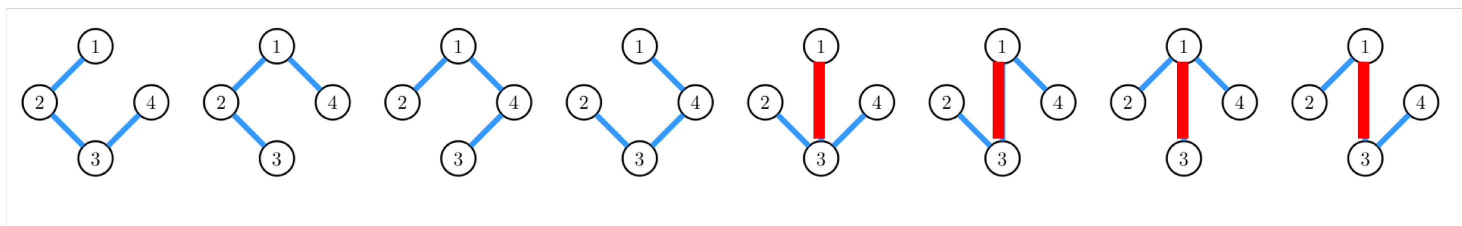
Pick a random tree

Random spanning trees

Graph G



Tree distribution



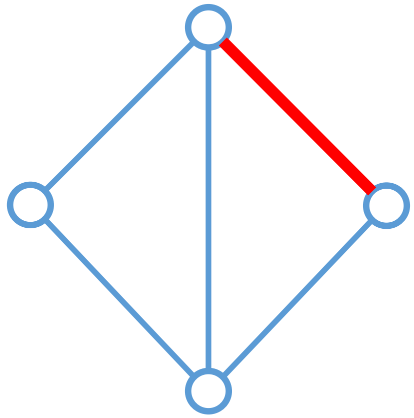
Pick a random tree

$$p_e = 1/2$$

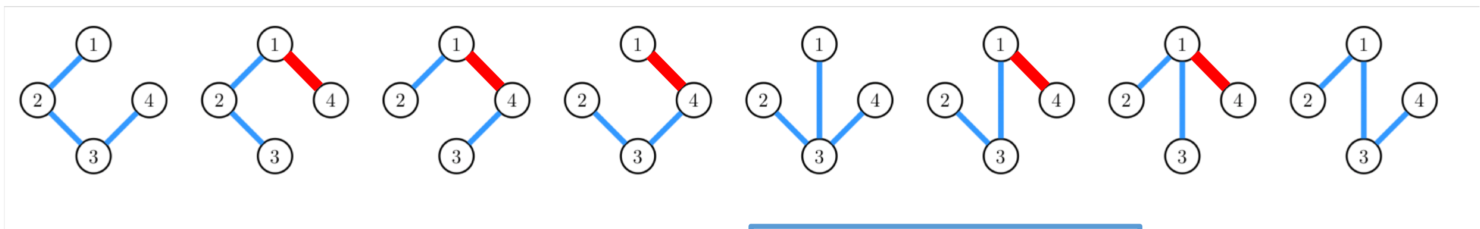
p_e: probability of edge present

Random spanning trees

Graph G



Tree distribution

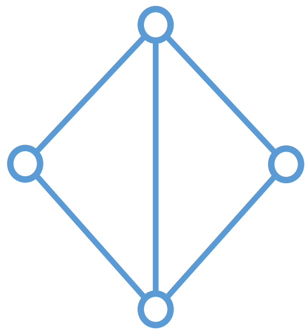


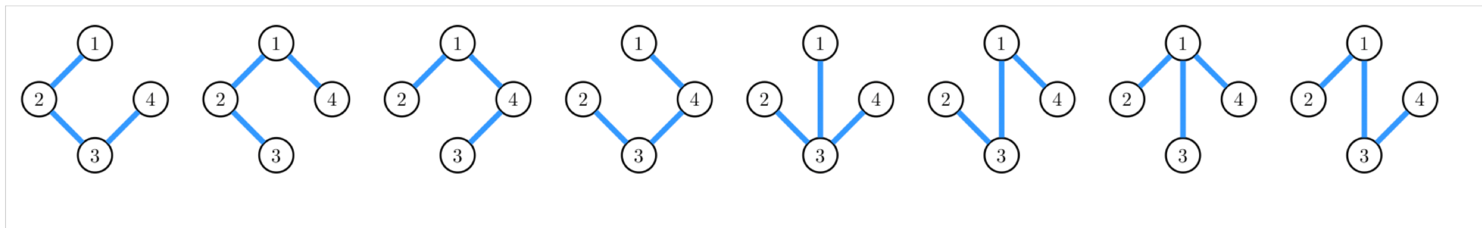
Pick a random tree

$p_e = 5/8$

p_e: probability of edge present

Random spanning trees





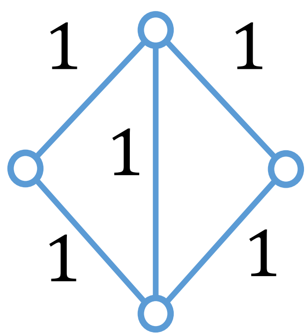
Getting the expectation right:

$$w_T(e) = \begin{cases} \frac{1}{p_e} w_G(e) & \text{w. probability } p_e \\ \mathbf{0} & \text{o.w.} \end{cases}$$

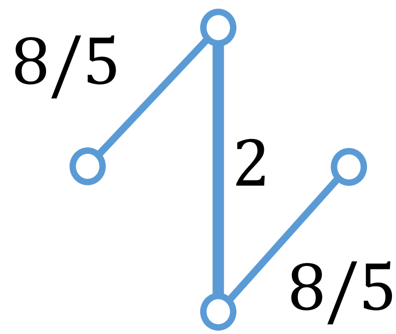
$$\mathbb{E} w_T(e) = p_e \cdot \frac{1}{p_e} w_G(e) = w_G(e)$$

Reweighted random spanning trees

Original weights

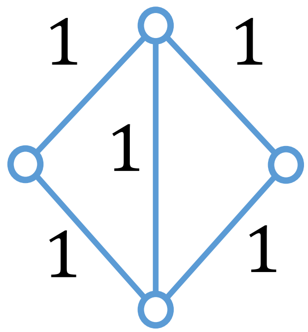


Tree weights

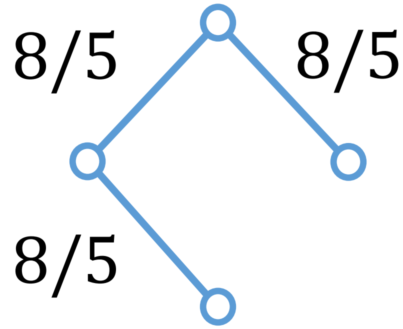


Reweighted random spanning trees

Original weights

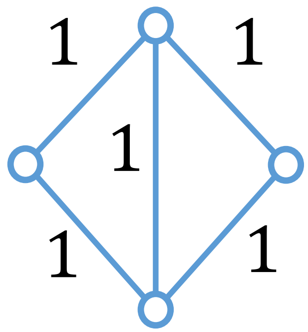


Tree weights

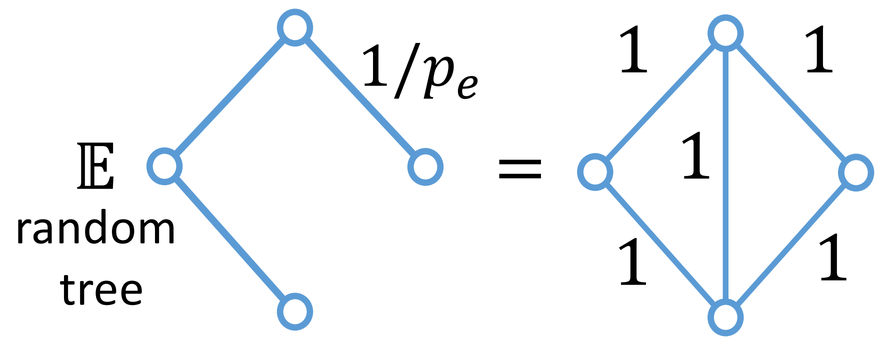


Reweighted random spanning trees

Original weights

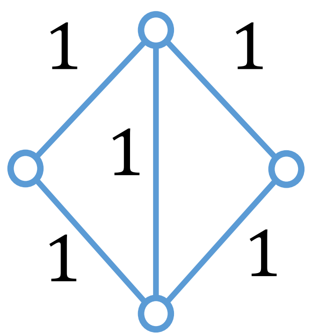


Tree weights

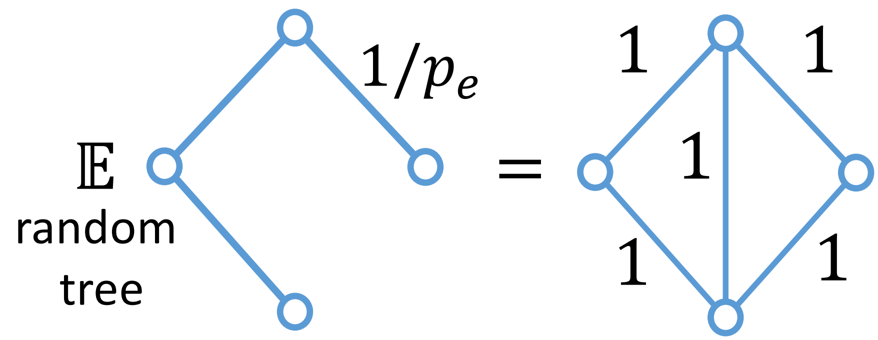


Reweighted random spanning trees

Original weights

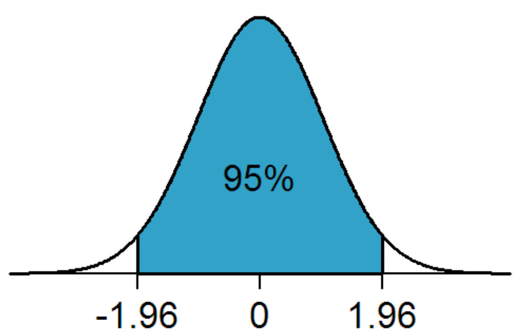


Tree weights



The **average** weight over trees equals the original weight

Does the tree "behave like" the original graph?



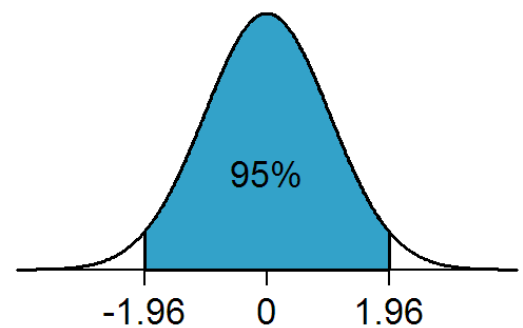
Preserving cuts?

Given cut $S \subseteq V$,

$$w_G(S, \bar{S}) = \sum_{(a,b) \in E \cap S \times \bar{S}} w_{ab}$$

Want for all $S \subseteq V$

$$w_T(S, \bar{S}) \approx w_G(S, \bar{S})$$



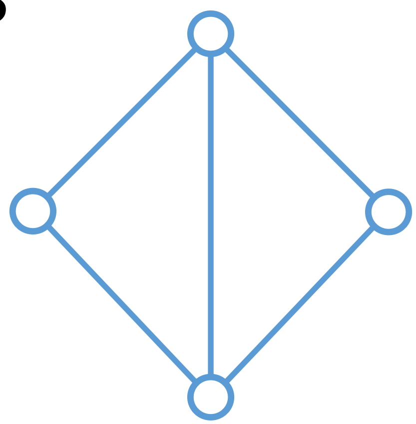
with high probability?

Too much to ask of one tree!

How many edges are necessary?

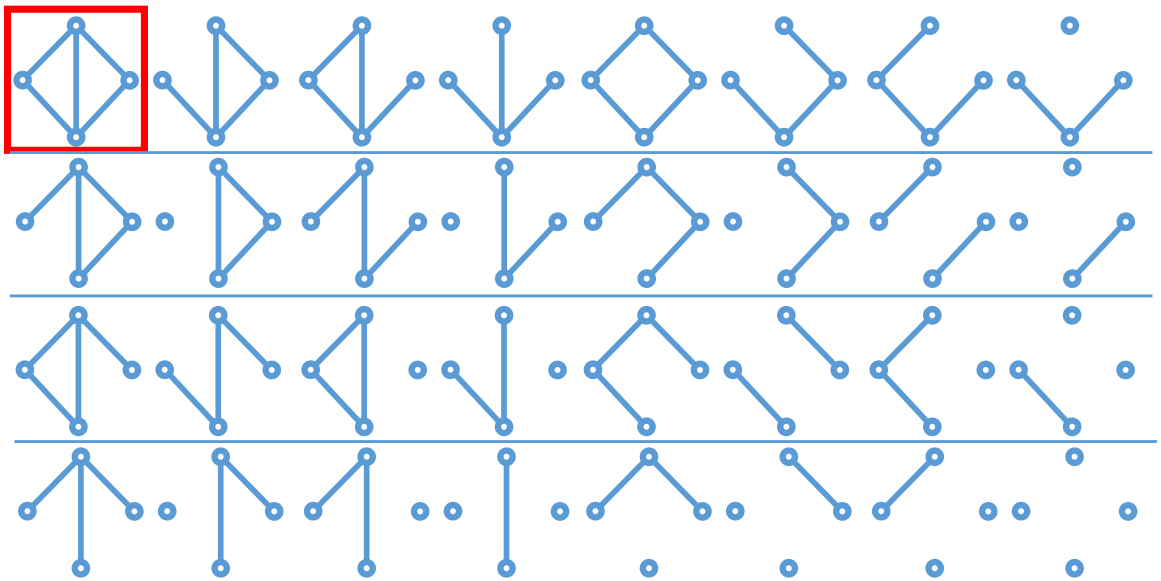
Independent edge samples

G **graph**



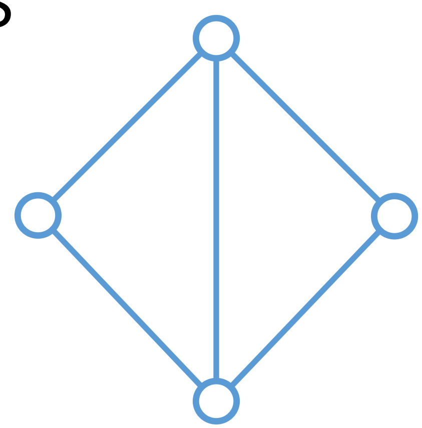
Flip a coin for each edge to decide if present

H **random graph, independent edges**



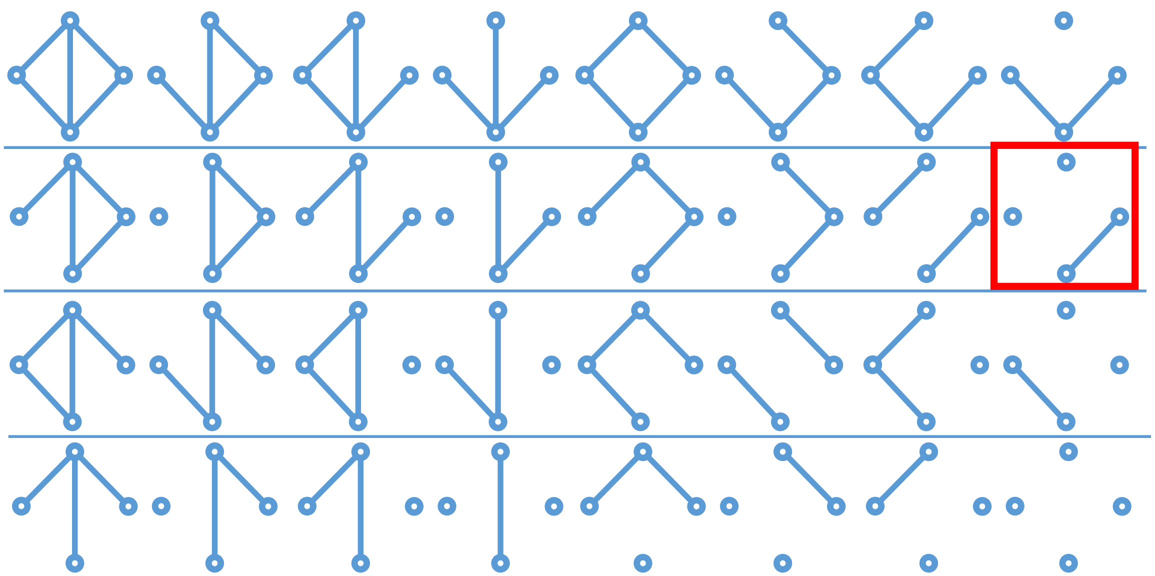
Independent edge samples

G **graph**

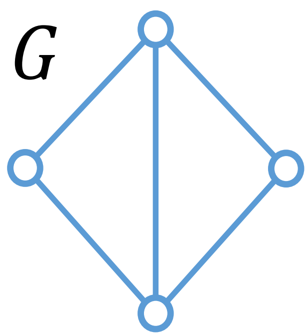


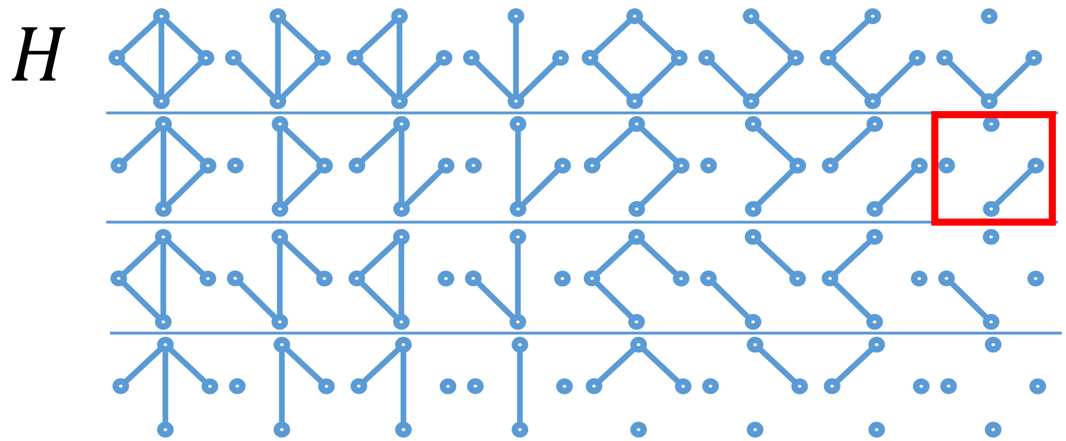
Flip a coin for each edge to decide if present

H **random graph, independent edges**



Independent edge samples





Getting the expectation right:

$$w_H(e) = \begin{cases} \frac{1}{p_e} w_G(e) & \text{w. probability } p_e \\ 0 & \text{o.w.} \end{cases}$$

$$\mathbb{E} w_H(e) = p_e \cdot \frac{1}{p_e} w_G(e) = w_G(e)$$

Preserving cuts?

Benczur-Karger '96

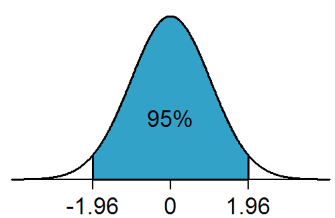
Sample edges independently with

"well-chosen" coin probabilities p_e,

s.t. H has on average ~~$O(\varepsilon^{-2} n \log^2 n)$~~ $O(\varepsilon^{-2} n \log n)$

Edges then w.h.p. for all cuts $S \subseteq V$

$$(1-\varepsilon) w_G(S, \bar{S}) \leq w_H(S, \bar{S}) \leq (1+\varepsilon) w_G(S, \bar{S})$$



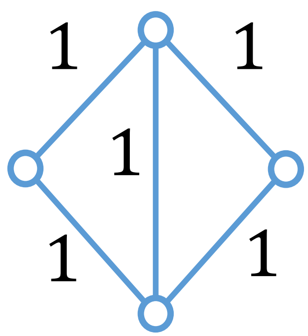
Proof sketch

Count #cuts of each size

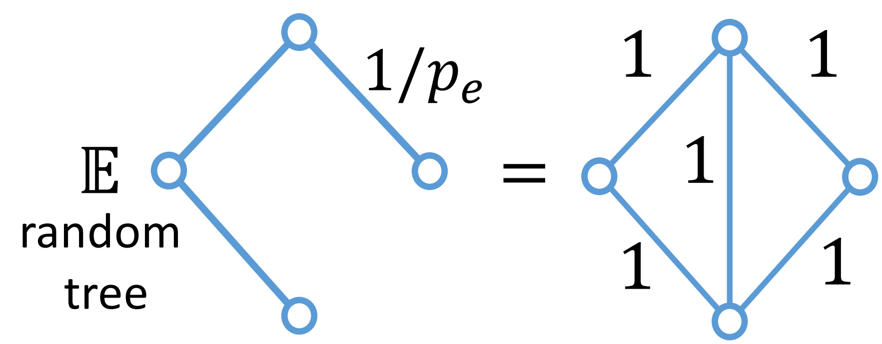
Chernoff concentration bound per cut

Reweighted random tree

Original weights

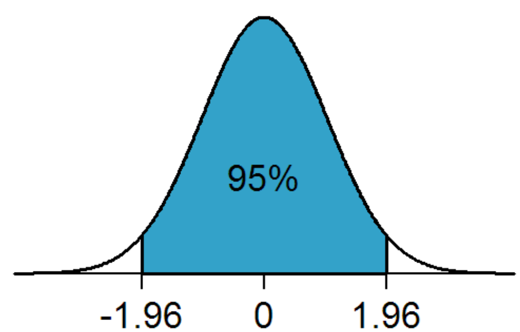


Tree weights



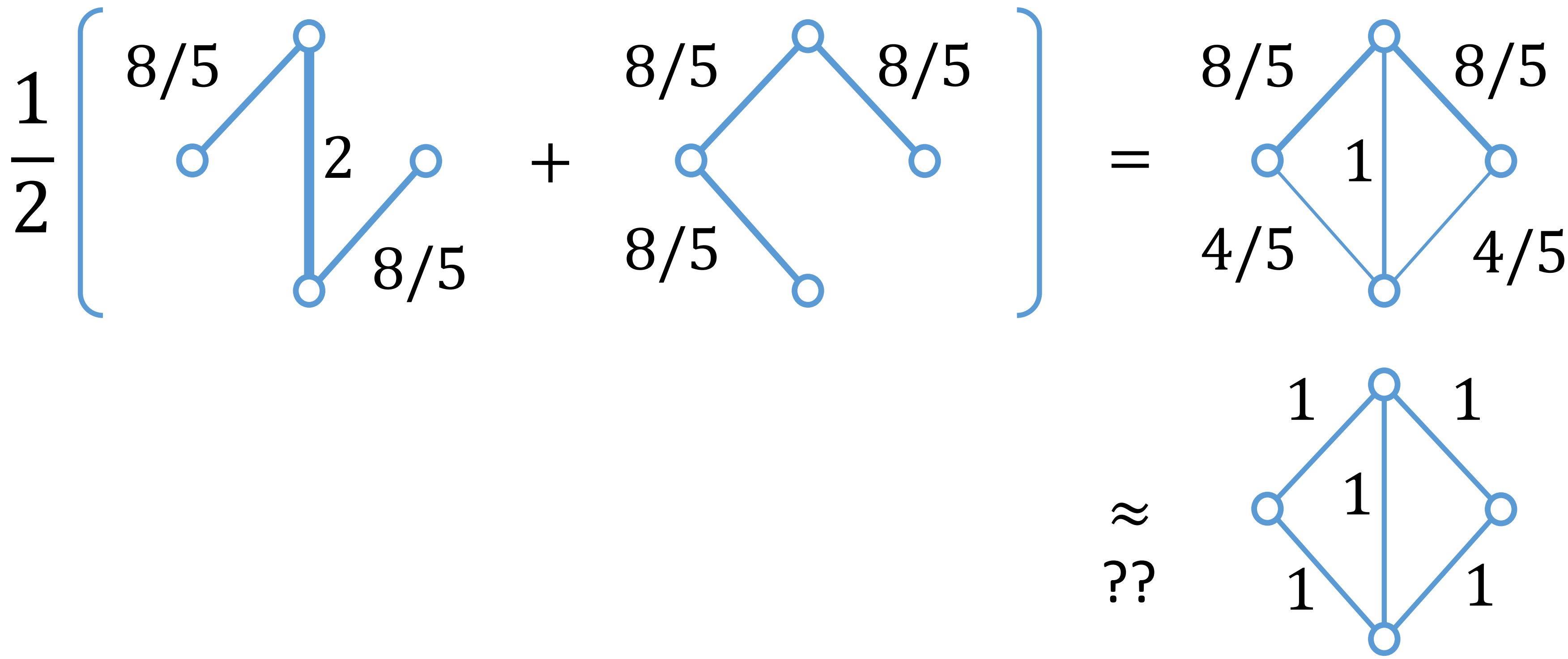
The **average** weight over trees equals the original weight

Does the tree "behave like" the original graph?



Combining trees

Maybe it’s better if we average a few trees?



Preserving cuts?

Fung-Harvey & Hariharan-Panigrahi '10

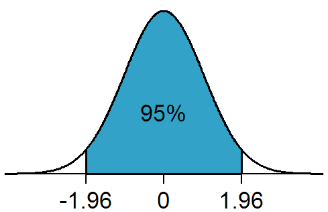
Let $H = \frac{1}{t}\sum_{i=1}^{t} T_i$ be the average of $t = O(\varepsilon^{-2}\log^2 n)$

reweighted random spanning trees of G

then w.h.p. for all cuts $S \subseteq V$

$$(1-\varepsilon)w_G(S,\bar{S}) \le w_H(S,\bar{S}) \le (1+\varepsilon)w_G(S,\bar{S})$$

Proof sketch



Benczur-Karger cut counting

Scalar Chernoff works for negatively correlated variables

Preserving cuts?

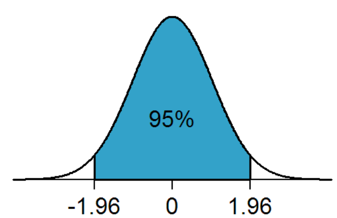
Goyal-Rademacher-Vempala '09

Given an unweighted bounded degree graph G,

let $H = \frac{1}{t}\sum_{i=1}^{t} T_i$ be the average of $O(1)$ unweighted random spanning trees of G

then w.h.p. for all cuts $S \subseteq V$

$$\Omega(1/\log n) w_G(S, \bar{S}) \leq w_H(S, \bar{S}) \leq w_G(S, \bar{S})$$



Proof sketch

Benczur-Karger cut counting + first tree gets small cuts
Scalar Chernoff works for negatively correlated variables

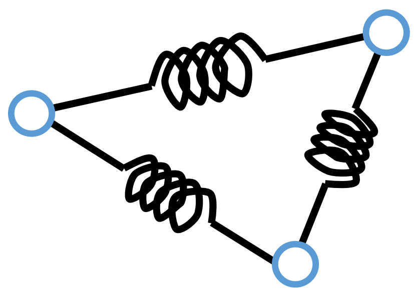
Preserving more than cuts: Matrices and quadratic forms

Laplacians: It's springs!

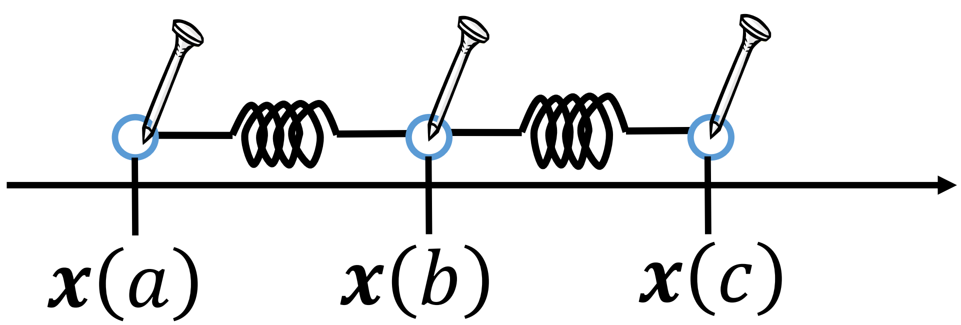
Weighted, undirected graph $G = (V, E, w)$, $w: E \to \mathbb{R}_+$

The Laplacian $\boldsymbol{L}$ is a $|V| \times |V|$ matrix describing G

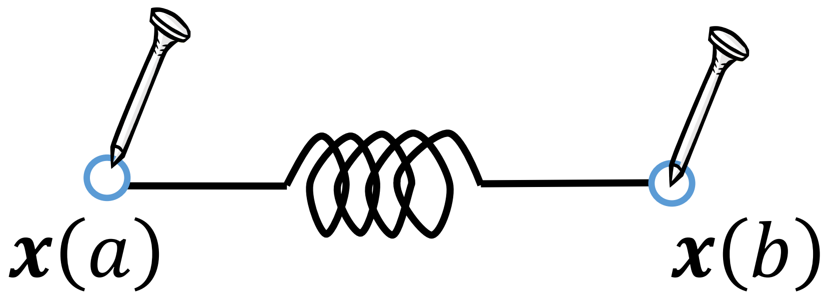
On each edge (a, b), put a spring between the vertices.



Nail down each vertex a at position $\boldsymbol{x}(a)$ along the real line.



Laplacians: It's springs!



Length $= |\boldsymbol{x}(a) - \boldsymbol{x}(b)|$

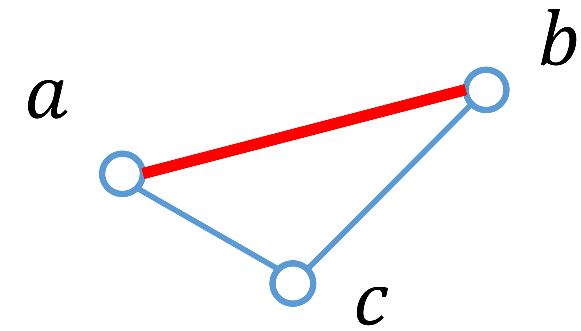
Energy $=$ spring const. $\cdot$ (length)$^2 = w_{ab}\big(\boldsymbol{x}(a) - \boldsymbol{x}(b)\big)^2$

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{(a,b)\in E} w_{ab}\big(\boldsymbol{x}(a) - \boldsymbol{x}(b)\big)^2$$

Laplacians

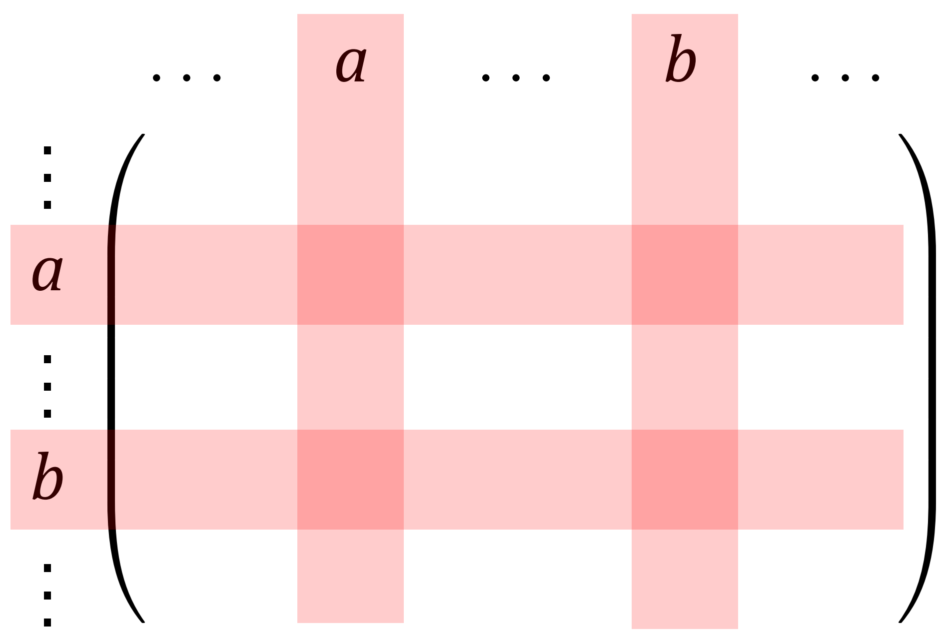
$$\boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x} = \sum_{(a,b)\in E} w_{ab}\big(x(a) - x(b)\big)^2$$

$$= \sum_{(a,b)\in E} \boldsymbol{x}^\top \boldsymbol{L}_{(a,b)} \boldsymbol{x}$$



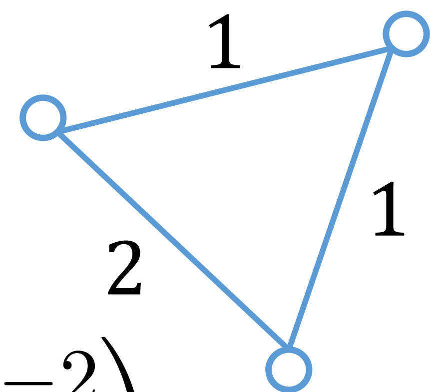
$$\boldsymbol{L} = \sum_{(a,b)\in E} \boldsymbol{L}_{(a,b)}$$ "baby Laplacian" per edge

$$\boldsymbol{L}_{(a,b)} = w_{(a,b)}$$



Laplacian of a graph

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$



$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & -1 & -2 \\ -1 & 2 & -1 \\ -2 & -1 & 3 \end{pmatrix}$$

Preserving matrices?

Suppose H is a random weighted graph s.t.
for every edge e, $\mathbb{E} w_H(e) = w_G(e)$.

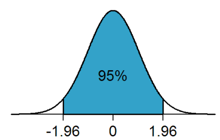
Then $\mathbb{E}\boldsymbol{L}_H = \boldsymbol{L}_G$

Does $\boldsymbol{L}_H$ "behave like" $\boldsymbol{L}_G$?

Preserving quadratic forms?

For all $\boldsymbol{x} \in \mathbb{R}^V$

$$(1-\epsilon)\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x} \le \boldsymbol{x}^\top \boldsymbol{L}_H \boldsymbol{x} \le (1+\epsilon)\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$$



with high probability?

Useful?

Since $\mathbf{1}_S^\top \boldsymbol{L}_G \mathbf{1}_S = w_G(S, \bar{S})$

implies cuts are preserved by letting $\boldsymbol{x} = \mathbf{1}_S$.

Quadratic form crucial for solving linear equations

Preserving quadratic forms?

Spielman-Srivastava ’08 (a la Tropp)

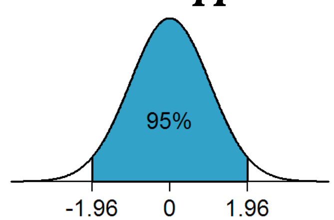
Sample edges independently with

“well-chosen” coin probabilities p_e,

s.t. H has on average $O(\varepsilon^{-2} n \log n)$ edges

then w.h.p. for all $\boldsymbol{x} \in \mathbb{R}^V$

$$(1-\varepsilon)\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x} \leq \boldsymbol{x}^\top \boldsymbol{L}_H \boldsymbol{x} \leq (1+\varepsilon)\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$$



Proof sketch

Bound spectral norm of sampled edge “baby Laplacians”

Matrix Chernoff concentration

What sampling probabilities?

Spielman-Srivastava ’08

“well-chosen” coin probabilities

$$p_e \propto \max_{\boldsymbol{x}} \frac{\boldsymbol{x}^\top \boldsymbol{L}_e \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x}}$$

What is the marginal probability of edges being present in a random spanning tree?

Also proportional to $\max_{\boldsymbol{x}} \frac{\boldsymbol{x}^\top \boldsymbol{L}_e \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x}}$ (!)

Random spanning trees similar to sparsification?

Preserving quadratic forms?

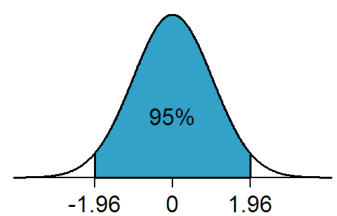
K.-Song '18

Let $H = \frac{1}{t}\sum_{i=1}^{t} T_i$ be the average of $t = O(\varepsilon^{-2}\log^2 n)$

reweighted random spanning trees of G

then w.h.p. for all $\boldsymbol{x} \in \mathbb{R}^V$

$$(1-\varepsilon)\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x} \le \boldsymbol{x}^\top \boldsymbol{L}_H \boldsymbol{x} \le (1+\varepsilon)\boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$$



Proof sketch

Bound norms of sampled matrices (immediate via SS'08)

Strongly Rayleigh matrix Chernoff concentration

Random spanning trees

$$\boldsymbol{x}^\top \frac{1}{t}\sum_{i=1}^{t} \boldsymbol{L}_{T_i}\boldsymbol{x} \approx_\varepsilon \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}, \quad t = \varepsilon^{-2}\log^2 n$$

Lower bound (K.-Song ’18)

$t = \Omega(\varepsilon^{-2}\log n)$ needed for ε-spectral sparsifier

Open question

Right number of logs?

Guess: $O(\varepsilon^{-2}\log n)$ trees

Random spanning trees

More results (K.-Song '18)

$\boldsymbol{x}^\top \boldsymbol{L}_T \boldsymbol{x} \leq O(\log n)\, \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$ for all $\boldsymbol{x}$ w.h.p.

$\Rightarrow$ in ε-spectrally connected graphs

random tree is $O(\varepsilon \log n)$-spectrally thin

Lower bounds

In some graphs, w. prob. $\geq 1 - e^{-0.4n}$ there exists $\boldsymbol{x}$ s.t.

$$\boldsymbol{x}^\top \boldsymbol{L}_T \boldsymbol{x} \nleq \frac{1}{8} \frac{\log n}{\log \log n}\, \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$$

and for some $\boldsymbol{y}$, $\boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y} \nleq \boldsymbol{y}^\top \boldsymbol{L}_T \boldsymbol{y}$

In a ring graph, there exists $\boldsymbol{x}, \boldsymbol{y}$ s.t.

$$\boldsymbol{x}^\top \boldsymbol{L}_T \boldsymbol{x} \nleq \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x} \quad \text{and} \quad \frac{1}{n-2}\, \boldsymbol{y}^\top \boldsymbol{L}_G \boldsymbol{y} \nleq \boldsymbol{y}^\top \boldsymbol{L}_T \boldsymbol{y}$$

Proving the strongly Rayleigh matrix Chernoff bound

An illustrative case

$\boldsymbol{x}^\top \boldsymbol{L}_T \boldsymbol{x} \leq O(\log n)\, \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$ for all $\boldsymbol{x}$ w.h.p.

Loewner order

$\boldsymbol{A} \preccurlyeq \boldsymbol{B}$ iff for all $\boldsymbol{x}$ $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}$

$\boldsymbol{x}^\top \boldsymbol{L}_T \boldsymbol{x} \leq O(\log n)\, \boldsymbol{x}^\top \boldsymbol{L}_G \boldsymbol{x}$ for all $\boldsymbol{x}$

$\boldsymbol{L}_T \preccurlyeq O(\log n) \boldsymbol{L}_G$

Proof strategy?

Convert problem to Doob martingales

Matrix martingale concentration

Control effect of conditioning via coupling

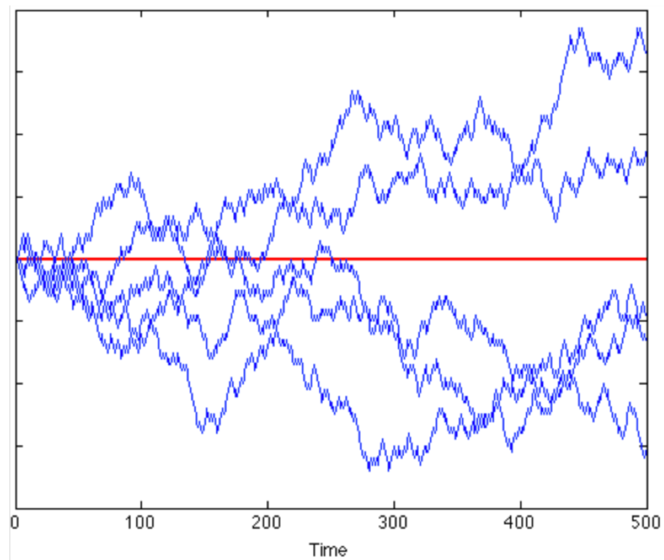
Norm bound from coupling

Variance bound: coupling symmetry + shrinking marginals

What is a martingale?

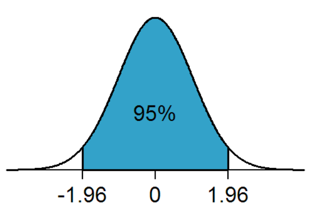
A sequence of random variables $Y_0, \dots, Y_k$ s.t.

$\mathbb{E}[Y_i | Y_0, \dots, Y_{i-1}] = Y_{i-1}$



Many concentration bounds for independent random variables can be generalized to the martingale case,

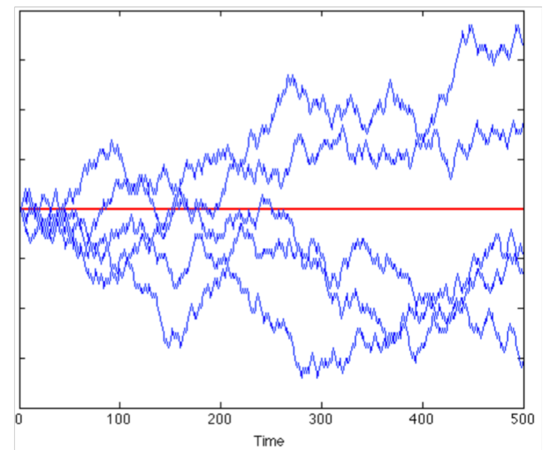
to show $Y_k \approx Y_0$ w.h.p.



Concentration of martingales

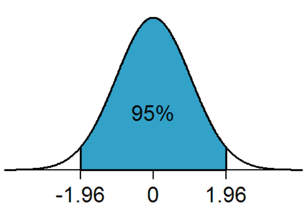
Why do martingales exhibit concentration?

Each difference is zero mean, conditional on previous outcomes $\mathbb{E}[Y_i - Y_{i-1} | Y_0, \dots, Y_{i-1}] = 0$



If each difference $Y_i - Y_{i-1}$ is small, then

$$Y_k - Y_0 = \sum_i Y_i - Y_{i-1} \approx 0$$



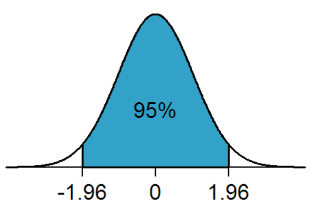
Doob martingales

Random variables $\gamma_1, \dots, \gamma_k$ NOT indep.

Goal: Prove concentration for $f(\gamma_1, \dots, \gamma_k)$

where f is ``stable'' under small changes to $\gamma_1, \dots, \gamma_k$

$f(\gamma_1, \dots, \gamma_k) \approx \mathbb{E} f(\gamma_1, \dots, \gamma_k)$?



Also need $\gamma_1, \dots, \gamma_k$ stable under conditioning

Doob martingales

Pick random outcome $\gamma_1, \dots, \gamma_k$ from distribution

$$Y_0 = \mathbb{E}[f(\gamma_1, \dots, \gamma_k)]$$

$$Y_1 = \mathbb{E}[f(\gamma_1, \dots, \gamma_k) | \gamma_1]$$

$$Y_2 = \mathbb{E}[f(\gamma_1, \dots, \gamma_k) | \gamma_1, \gamma_2]$$

$$\vdots$$

$$Y_k = \mathbb{E}[f(\gamma_1, \dots, \gamma_k) | \gamma_1, \gamma_2, \dots, \gamma_k] = f(\gamma_1, \dots, \gamma_k)$$

$$\mathbb{E}Y_1 = \mathbb{E}_{\gamma_1}\left[\mathbb{E}[f(\gamma_1, \dots, \gamma_k) | \gamma_1]\right] = \mathbb{E}[f(\gamma_1, \dots, \gamma_k)] = Y_0$$

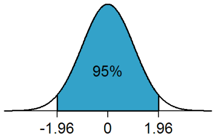
$$\mathbb{E}[Y_i - Y_{i-1} | \text{prev. steps}] = 0$$

Martingale!

Despite $\gamma_1, \dots, \gamma_k$

NOT independent

Show $Y_k \approx Y_0$, i.e. $f(\gamma_1, \dots, \gamma_k) \approx \mathbb{E}f(\gamma_1, \dots, \gamma_k)$



Our Doob martingale

Reveal one edge of tree at a time

Let γ_i denote the index of the ith edge of the tree

Pick random tree as $T = \gamma_1, \gamma_2, \dots, \gamma_{n-1}$

$\boldsymbol{L}_T = f(\gamma_1, \gamma_2, \dots, \gamma_{n-1})$

$\boldsymbol{Y}_0 = \mathbb{E}[\boldsymbol{L}_T]$

$\boldsymbol{Y}_1 = \mathbb{E}[\boldsymbol{L}_T | \gamma_1]$

$\vdots$

$\boldsymbol{Y}_{n-1} = \mathbb{E}[\boldsymbol{L}_T | \gamma_1, \gamma_2, \dots, \gamma_{n-1}] = \boldsymbol{L}_T$

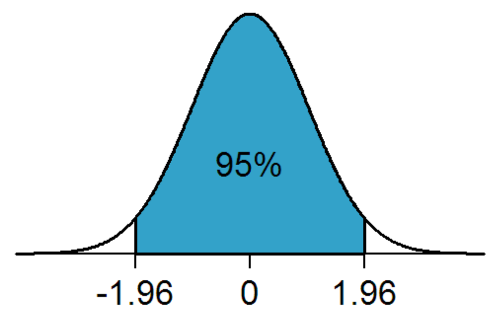
$\mathbb{E}[\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1} | \text{prev. steps}] = \boldsymbol{0}$

Our Doob martingale

Want to show

$\boldsymbol{Y}_{n-1} = \boldsymbol{L}_T$ is close to $\boldsymbol{Y}_0 = \mathbb{E}[\boldsymbol{L}_T]$

$\boldsymbol{Y}_{n-1} - \boldsymbol{Y}_0 = \sum_i \boldsymbol{Y}_i - \boldsymbol{Y}_{i-1}$



Matrix martingale concentration?

Matrix Freedman (Tropp '11)

Norm $\|\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1}\| \leq 1$

Variance $\|\sum_i \mathbb{E}[(\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1})^2 | \text{ prev. steps}]\| \leq O(\log n)$

implies w.h.p

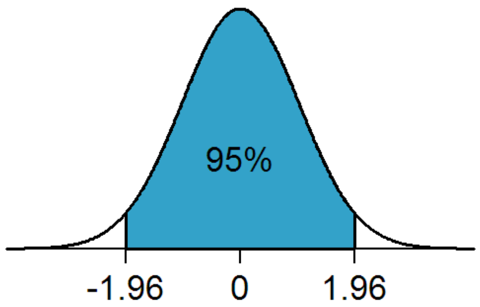
$\boldsymbol{L}_T \preccurlyeq O(\log n) \boldsymbol{L}_G$

Our Doob martingale

Want to show

$\boldsymbol{Y}_{n-1} = \boldsymbol{L}_T$ is close to $\boldsymbol{Y}_0 = \mathbb{E}[\boldsymbol{L}_T]$

$\boldsymbol{Y}_{n-1} - \boldsymbol{Y}_0 = \sum_i \boldsymbol{Y}_i - \boldsymbol{Y}_{i-1}$



Matrix martingale concentration?

Matrix Freedman (Tropp '11)

Norm $\|\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1}\| \leq 1$

Variance $\|\sum_i \mathbb{E}[(\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1})^2 | \text{prev. steps}]\| \leq O(\log n)$

difficult

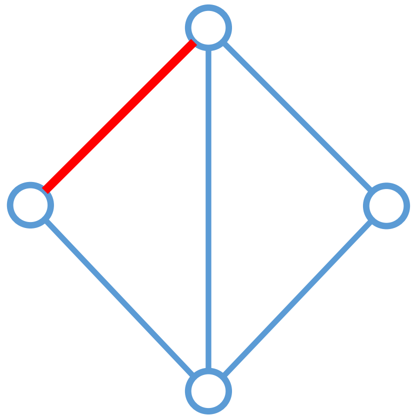
implies w.h.p

$\boldsymbol{L}_T \preccurlyeq O(\log n)\boldsymbol{L}_G$

How can we understand $\boldsymbol{Y}_{42} = \mathbb{E}[\boldsymbol{L}_T | \gamma_1, \ldots, \gamma_{42}]$?

How does conditioning change the distribution?

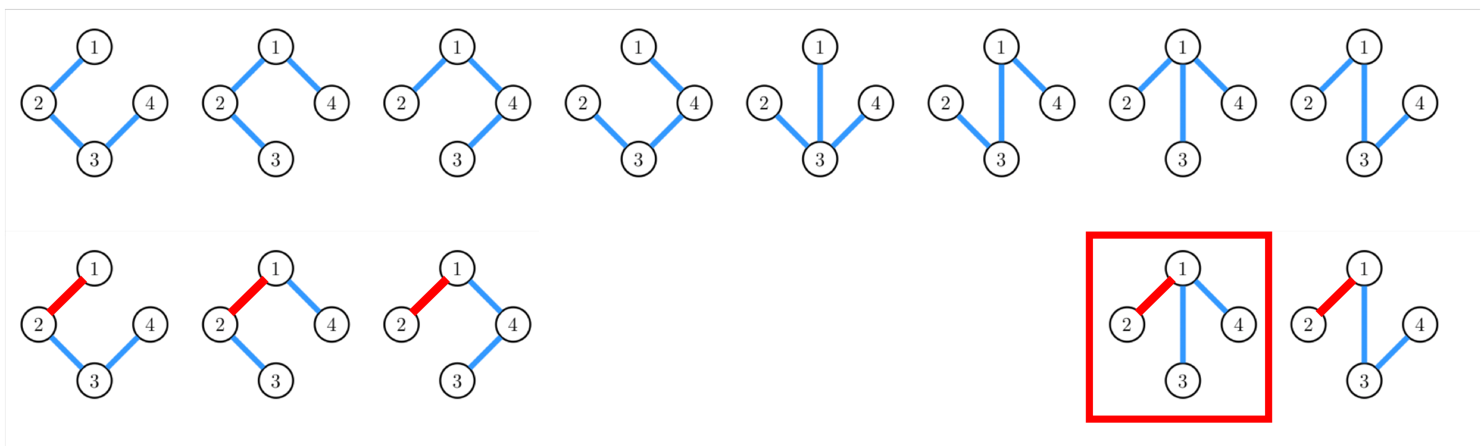
Graph



Tree distribution

All

Conditional



Pick a random tree, conditional on red edge present?

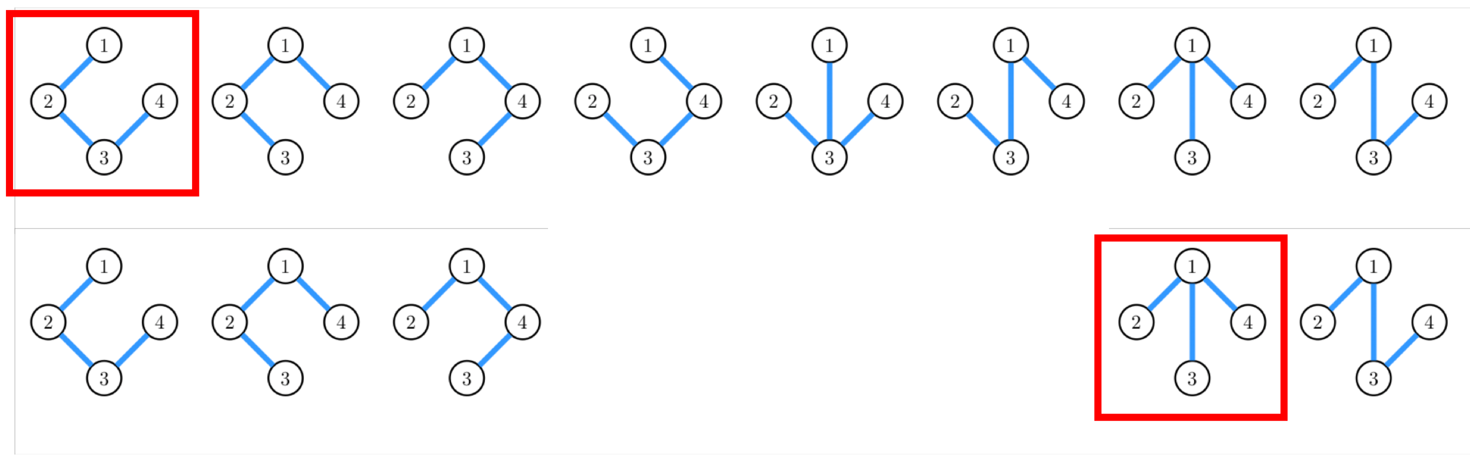
How does conditioning change the distribution?

How similar are the distributions?

Tree distribution

All

Conditional



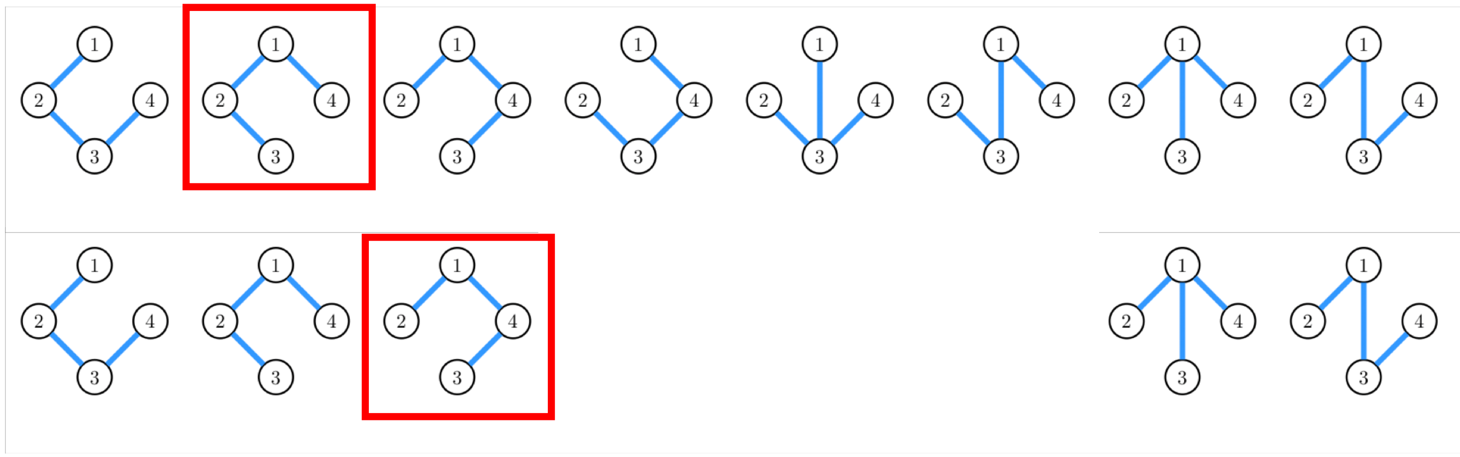
How does conditioning change the distribution?

How similar are the distributions?

Tree distribution

All

Conditional



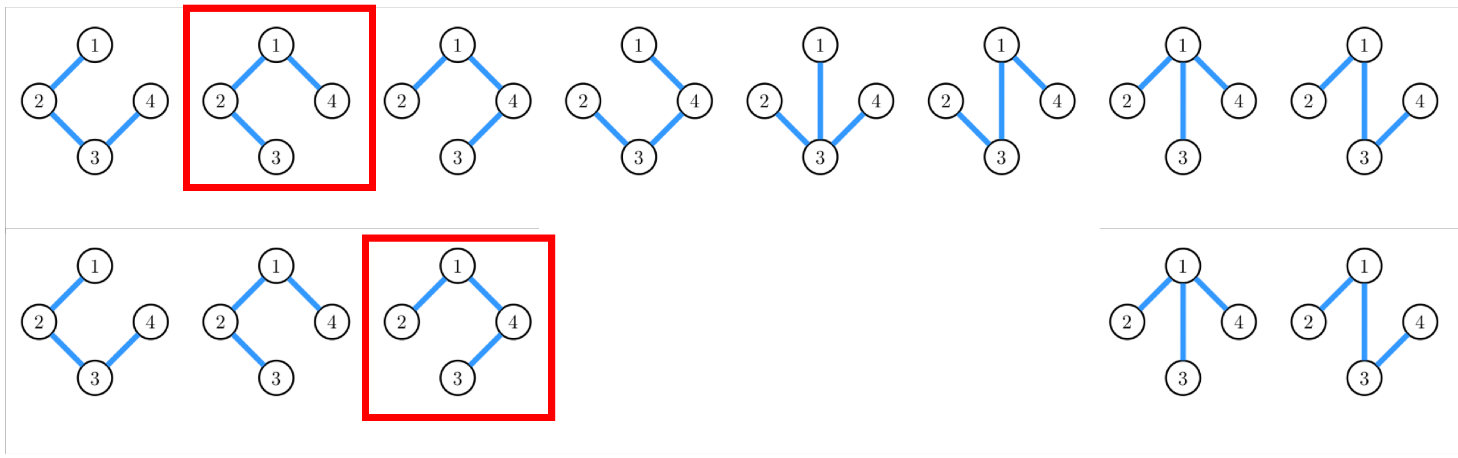
How does conditioning change the distribution?

How similar are the distributions?

Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above

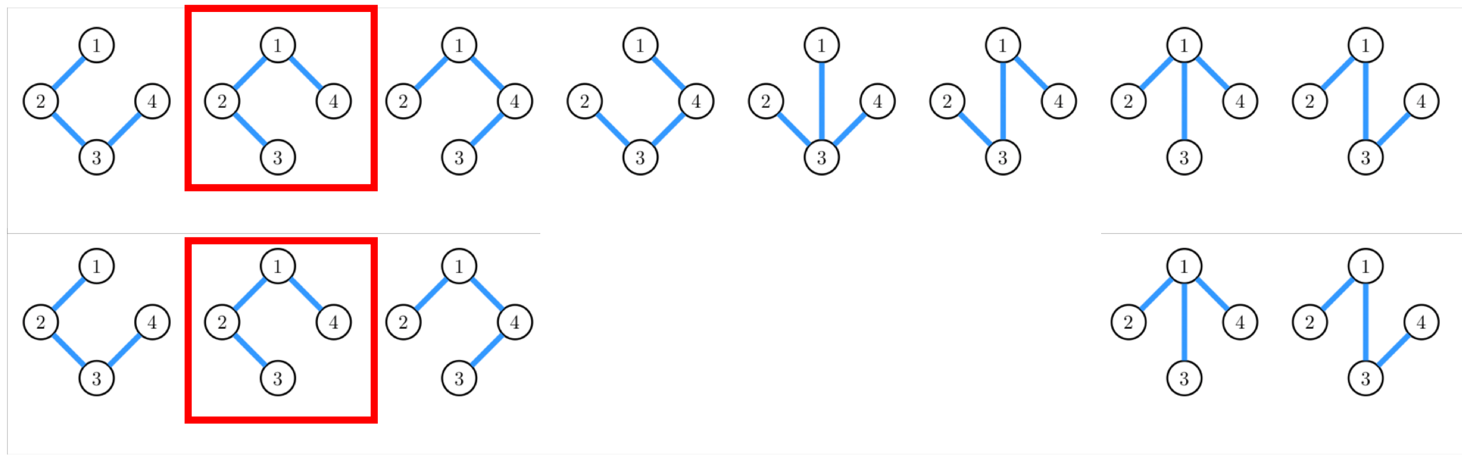
How does conditioning change the distribution?

How similar are the distributions?

Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above

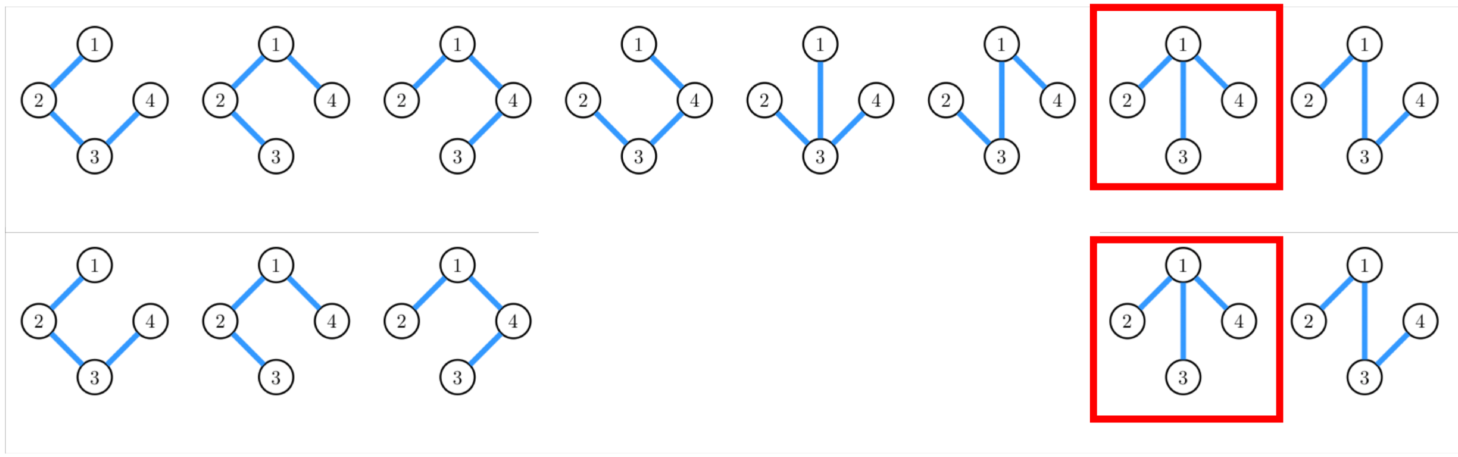
How does conditioning change the distribution?

How similar are the distributions?

Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above

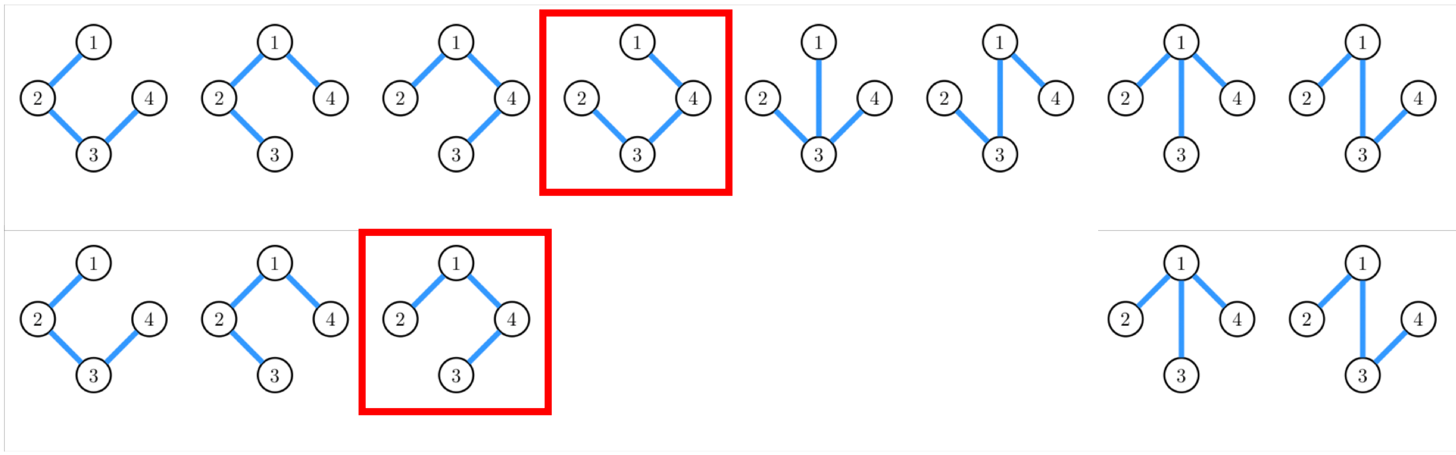
How does conditioning change the distribution?

How similar are the distributions?

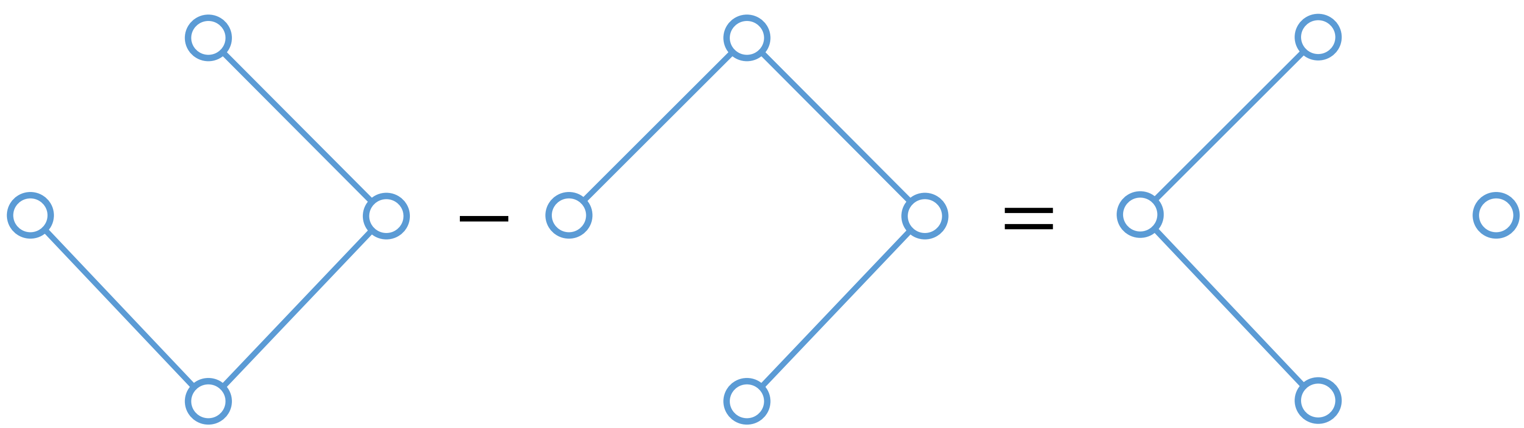
Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above



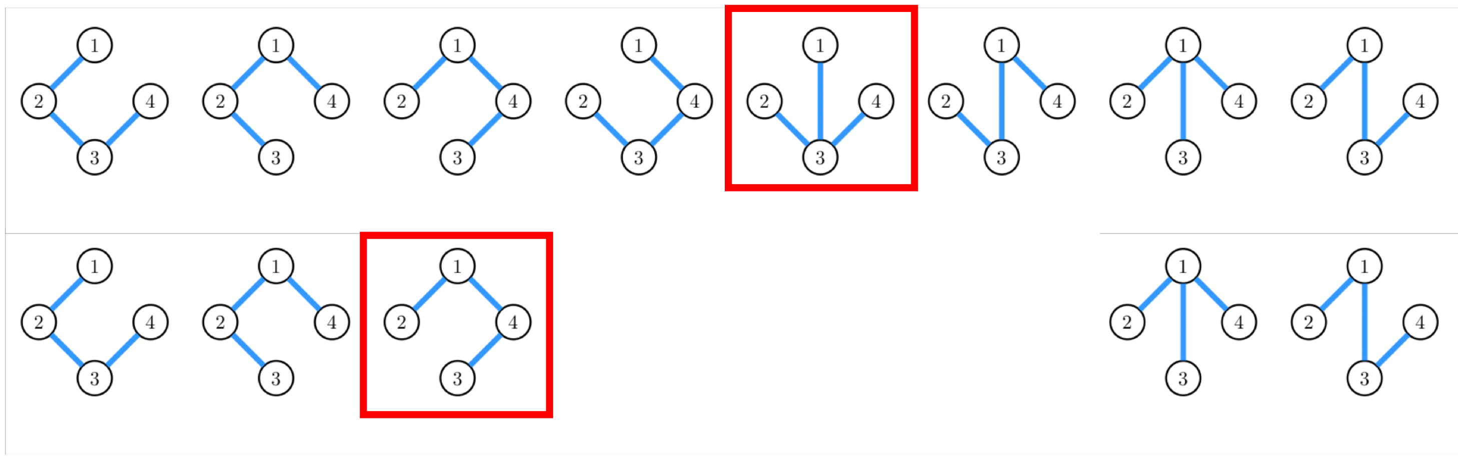
How does conditioning change the distribution?

How similar are the distributions?

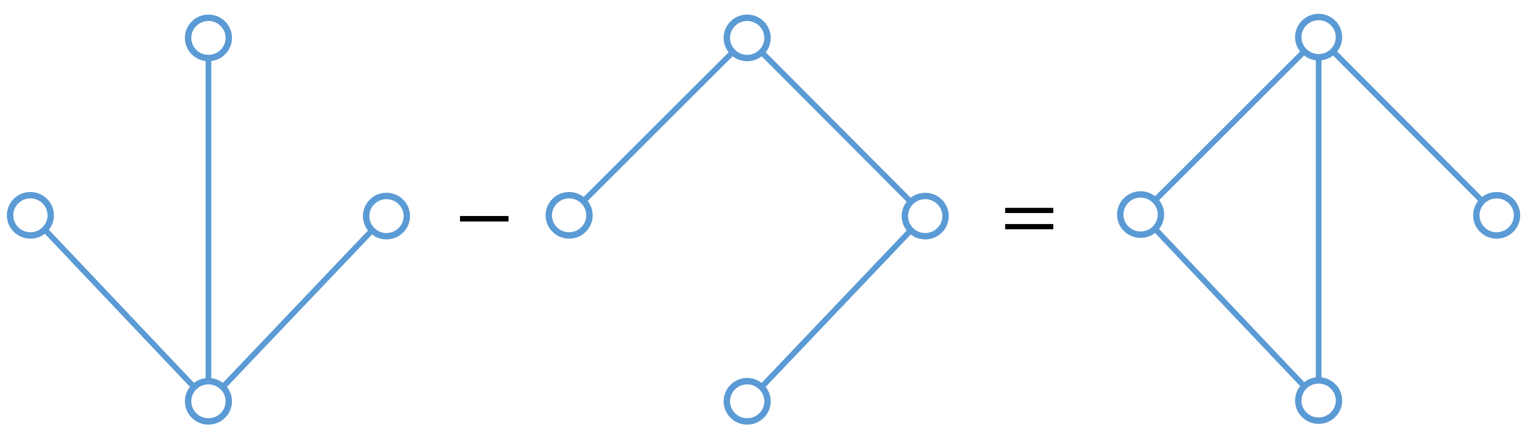
Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above



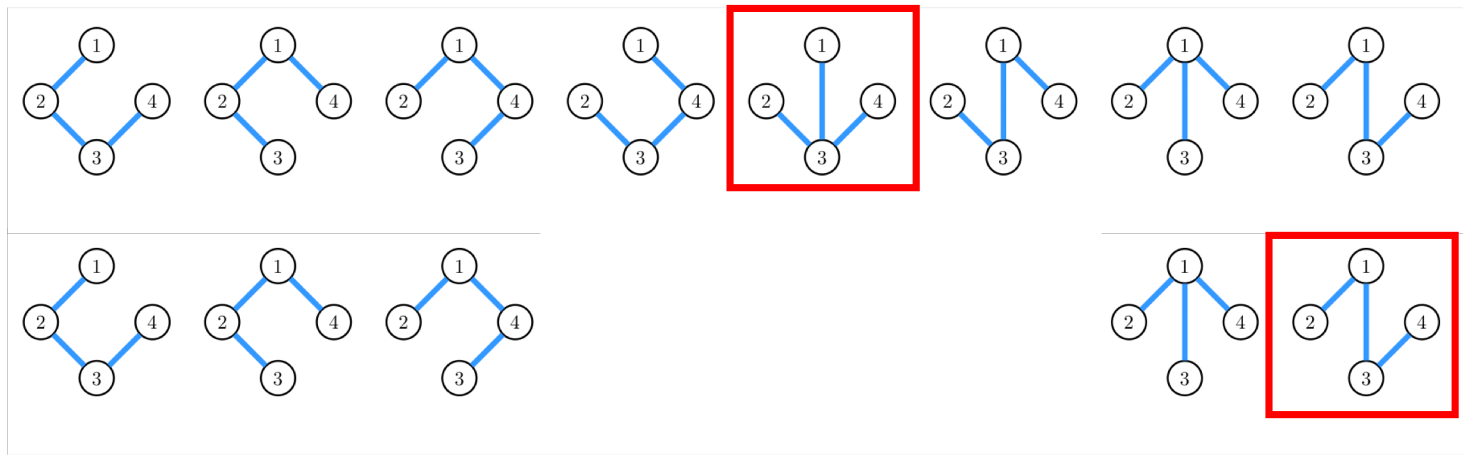
How does conditioning change the distribution?

How similar are the distributions?

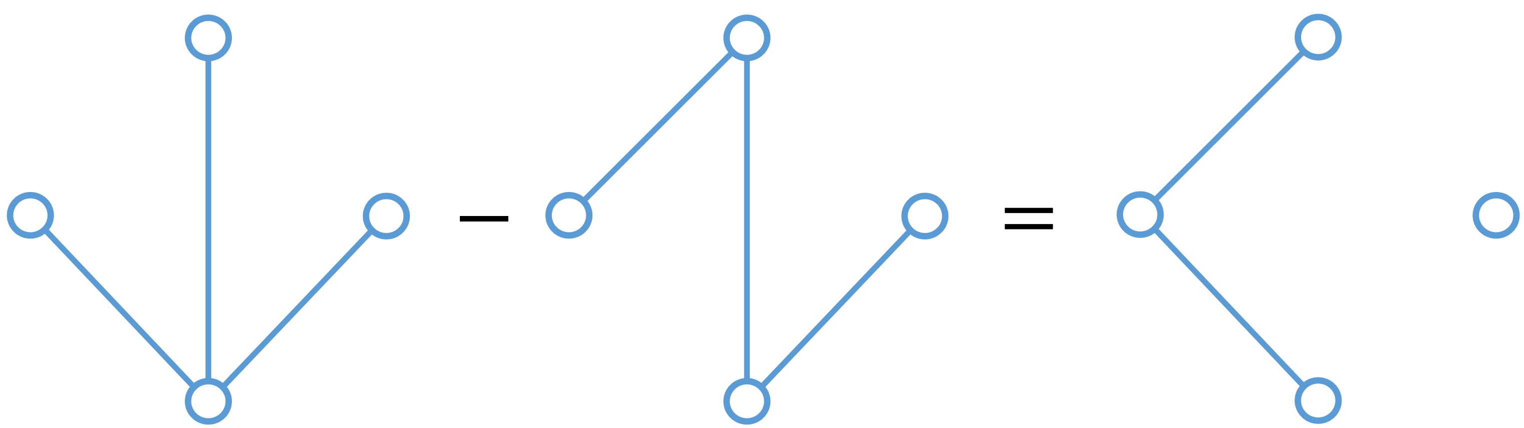
Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above



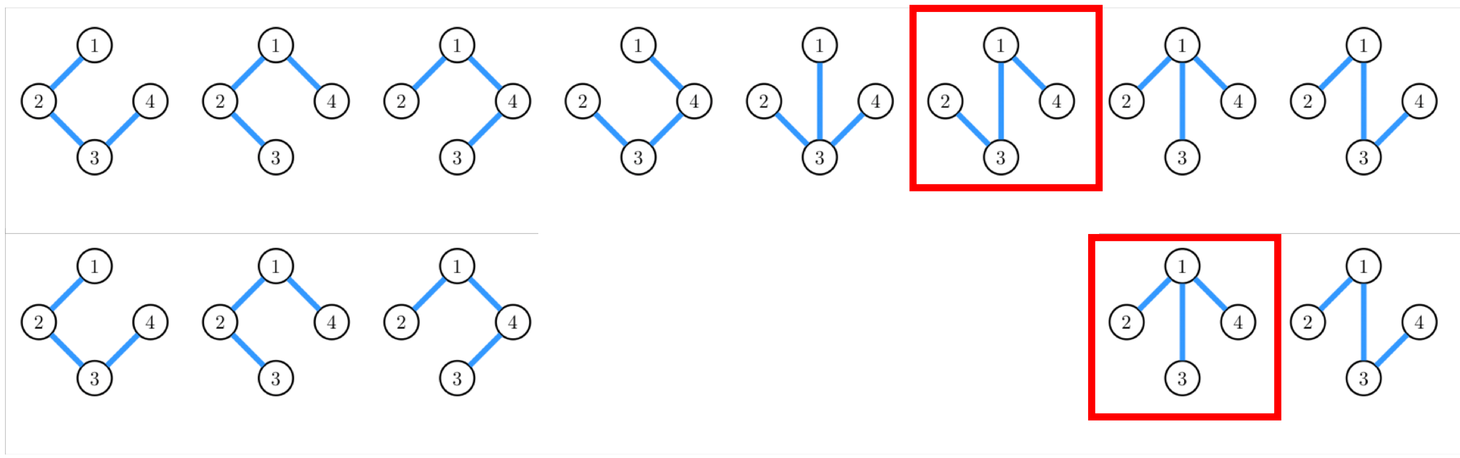
How does conditioning change the distribution?

How similar are the distributions?

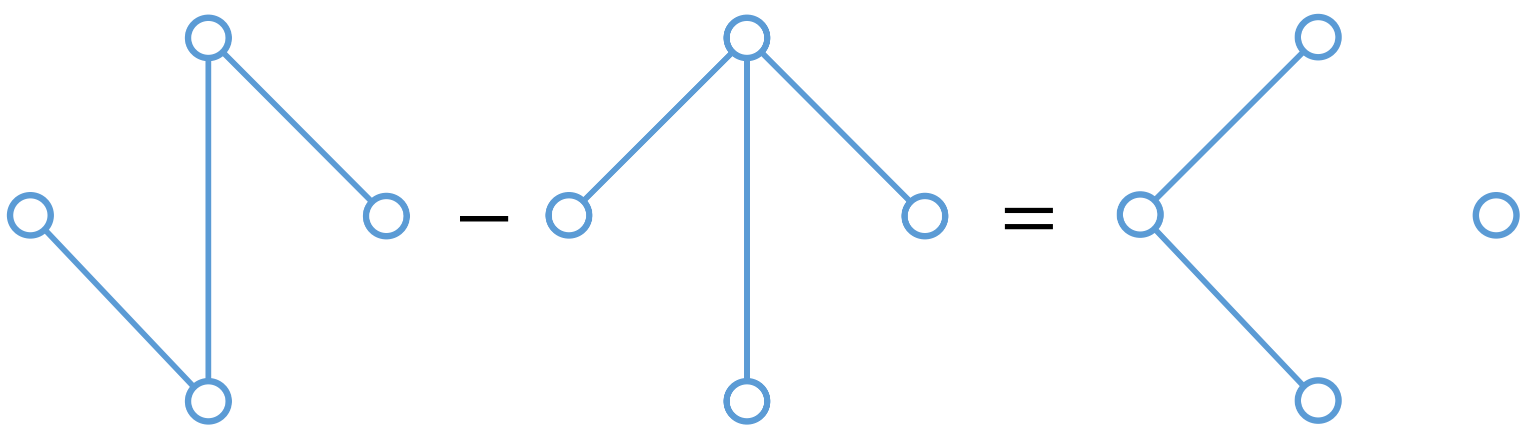
Tree distribution

All T

Cond. T'



Coupling Pick pair (T, T') with marginals as above

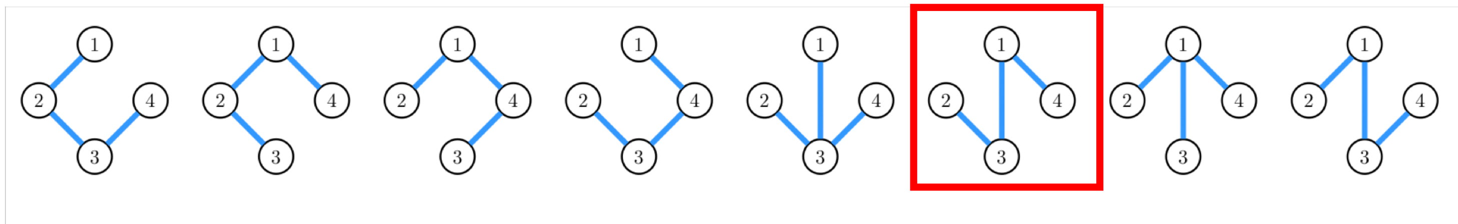


How does conditioning change the distribution?

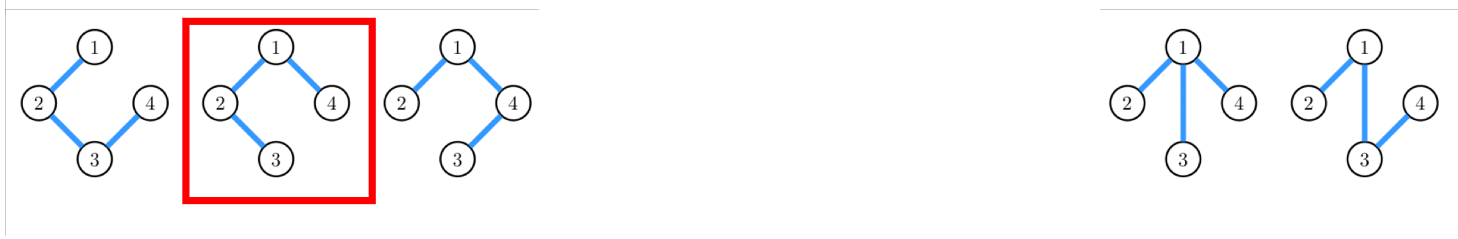
How similar are the distributions?

Tree distribution

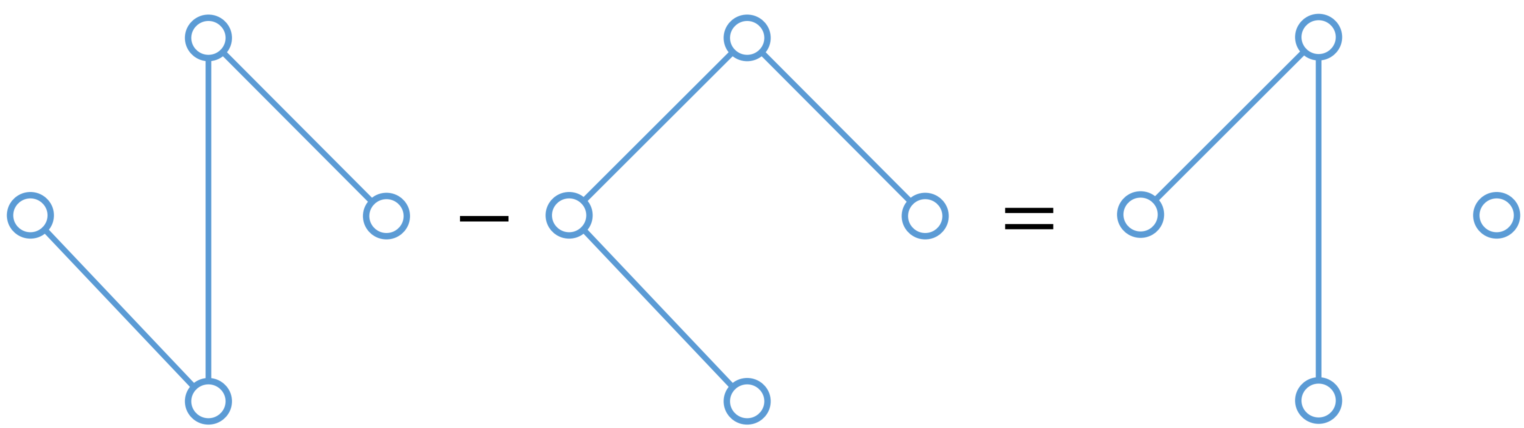
All T



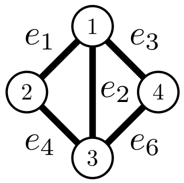
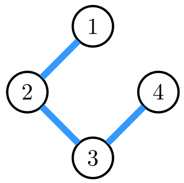
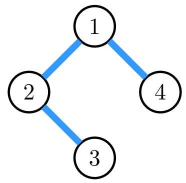
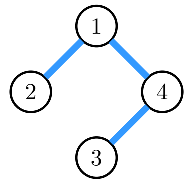
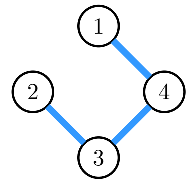
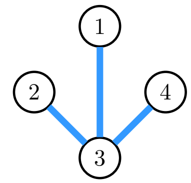
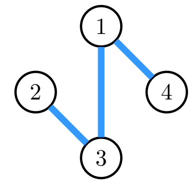
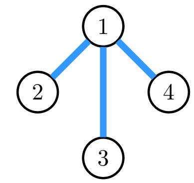
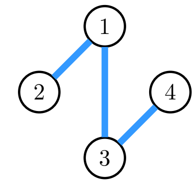
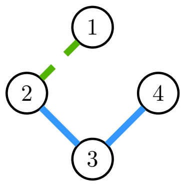
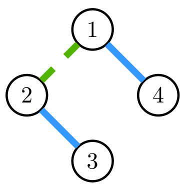
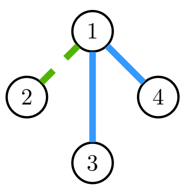
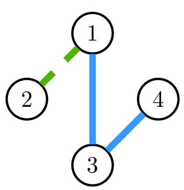
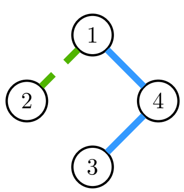
Cond. T'



Coupling Pick pair (T, T') with marginals as above



Coupling table – difference ≤ 2

								
	$e_1, \frac{1}{8}$			$e_3, \frac{1}{40}$	$e_2, \frac{2}{40}$			
		$e_1, \frac{1}{8}$		$e_6, \frac{1}{40}$		$e_2, \frac{2}{40}$		
						$e_4, \frac{3}{40}$	$e_1, \frac{1}{8}$	
					$e_4, \frac{3}{40}$			$e_1, \frac{1}{8}$
			$e_1, \frac{1}{8}$	$e_4, \frac{3}{40}$				

Good couplings

Stochastic covering property

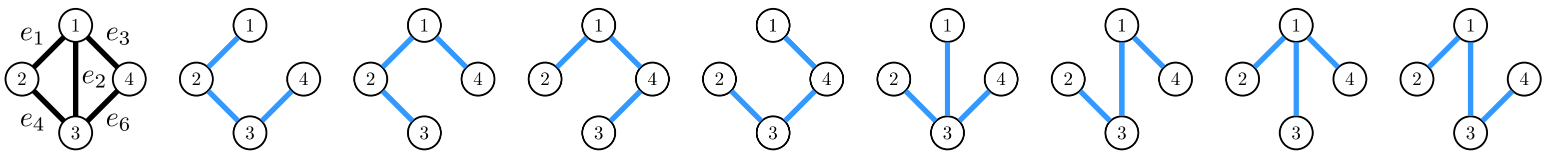
For k-homogenous strongly Rayleigh distributions
a coupling of (T, T') with difference ≤ 2 always exists.

[Borcea-Branden-Liggett '09]
[Peres-Pemantle '14]

Coupling table has more structure

Alice, Bob, Charlie want to form a tree,

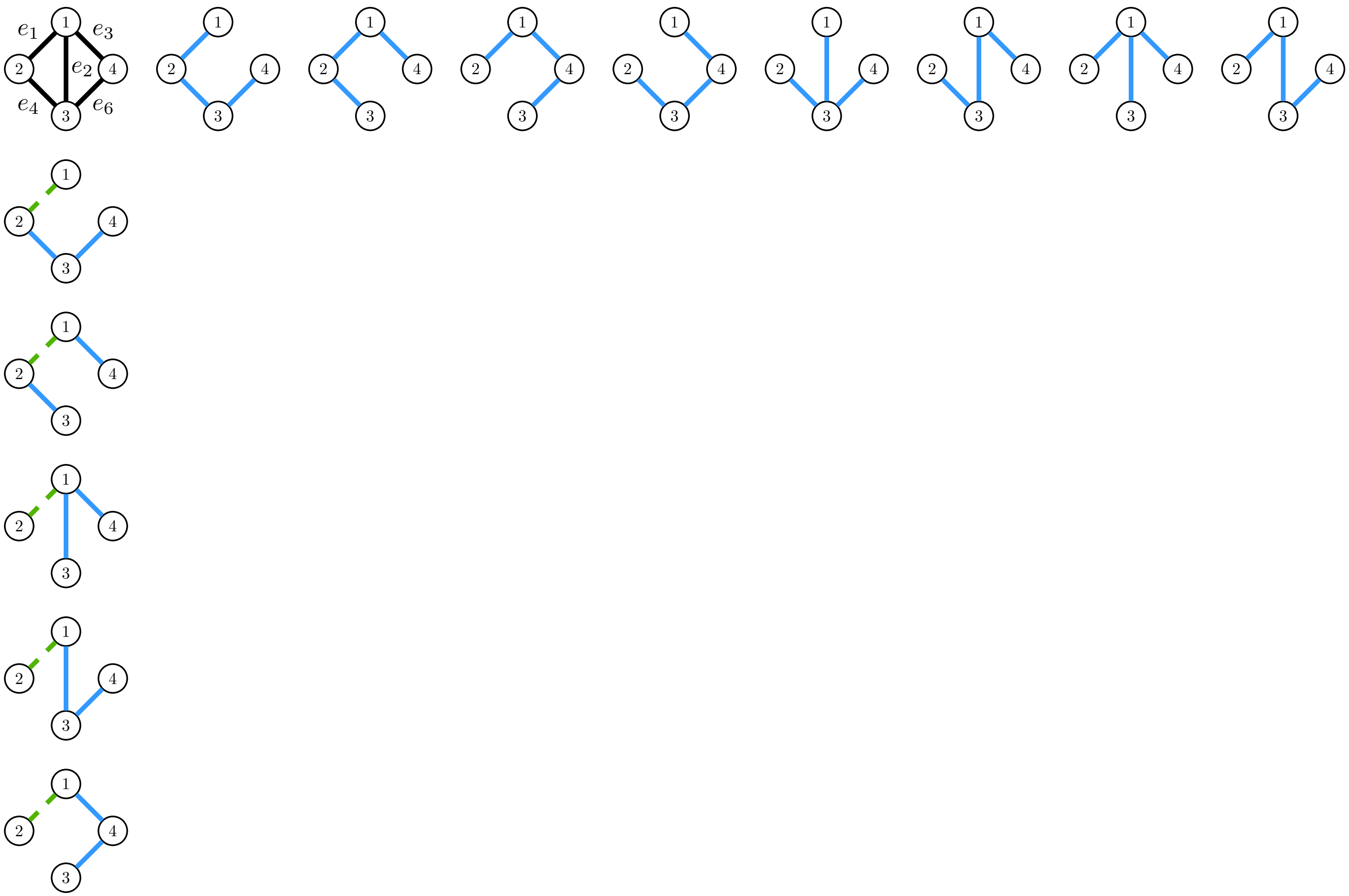
by each selecting one edge: $\gamma_1, \gamma_2, \gamma_3$



Coupling table has more structure

Alice picks γ_1

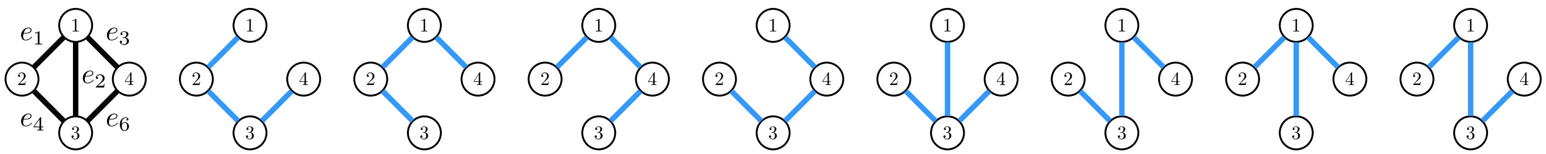
How much does this restrict Bob and Charlie?

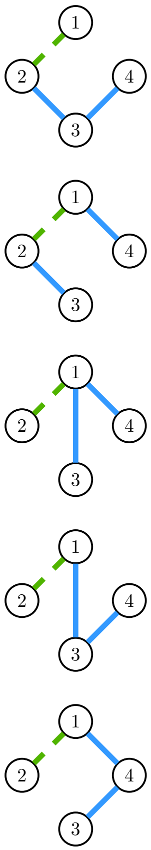


Coupling table has more structure

Alice picks γ_1 

How much does this restrict Bob and Charlie?



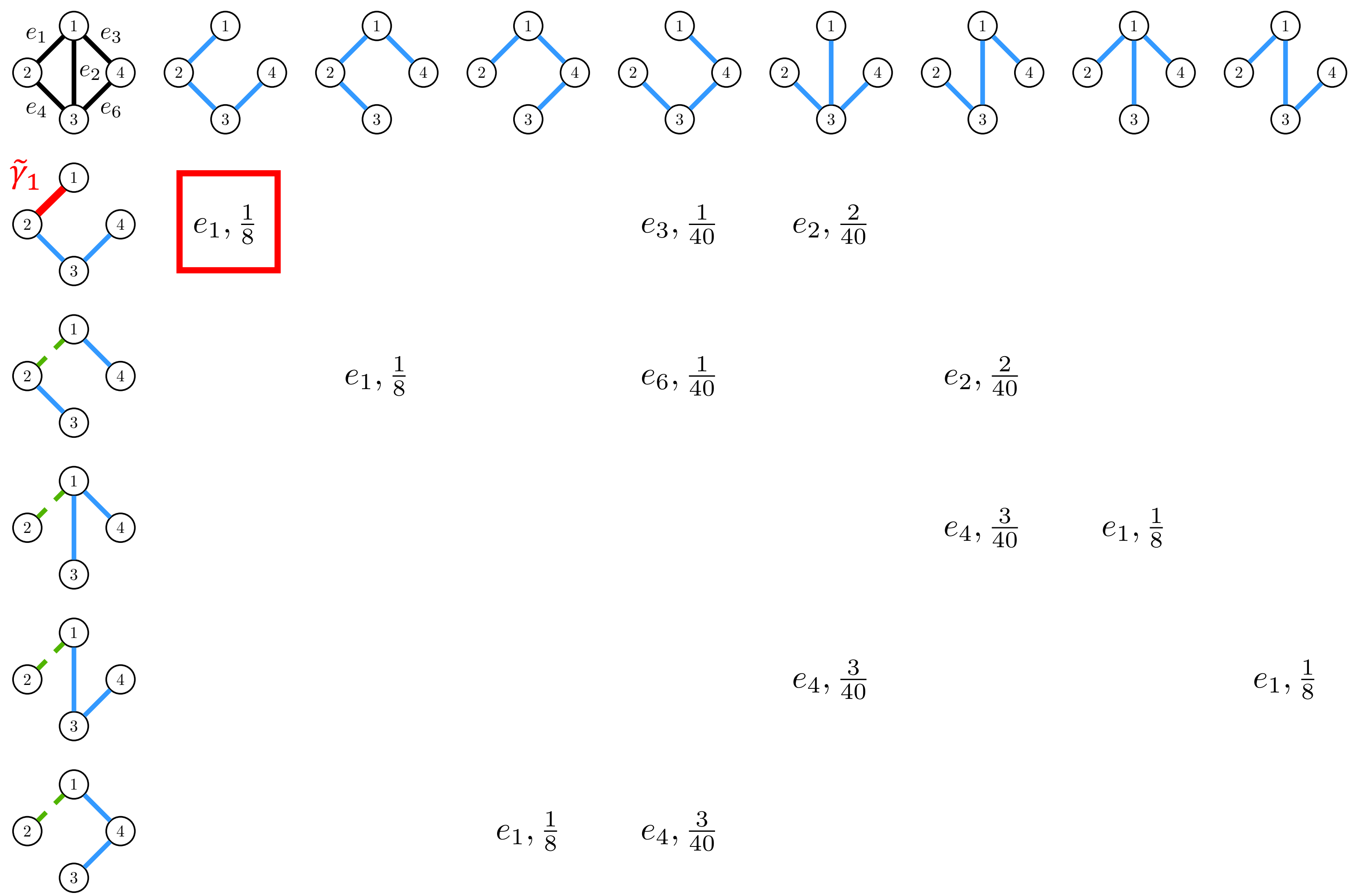


After Bob and Charlie choose their edges,
they enlist Anna to pick an extra edge, $\tilde{\gamma}_1$

Anna can choose that edge s.t.
Anna, Bob, & Charlie, obtain original distribution

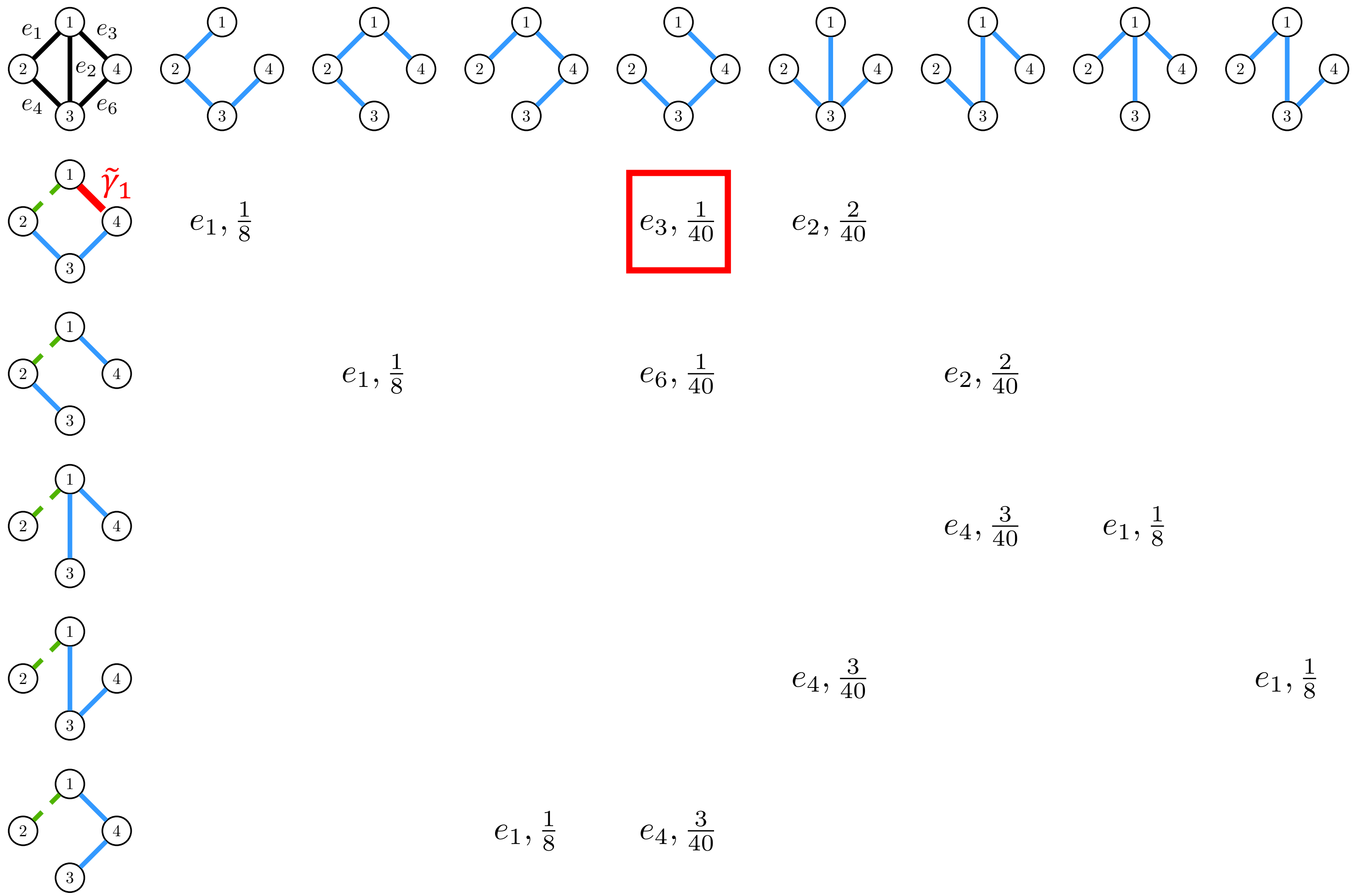
Coupling table has more structure

Recover the original distribution by adding a “make-up edge” to conditional distribution



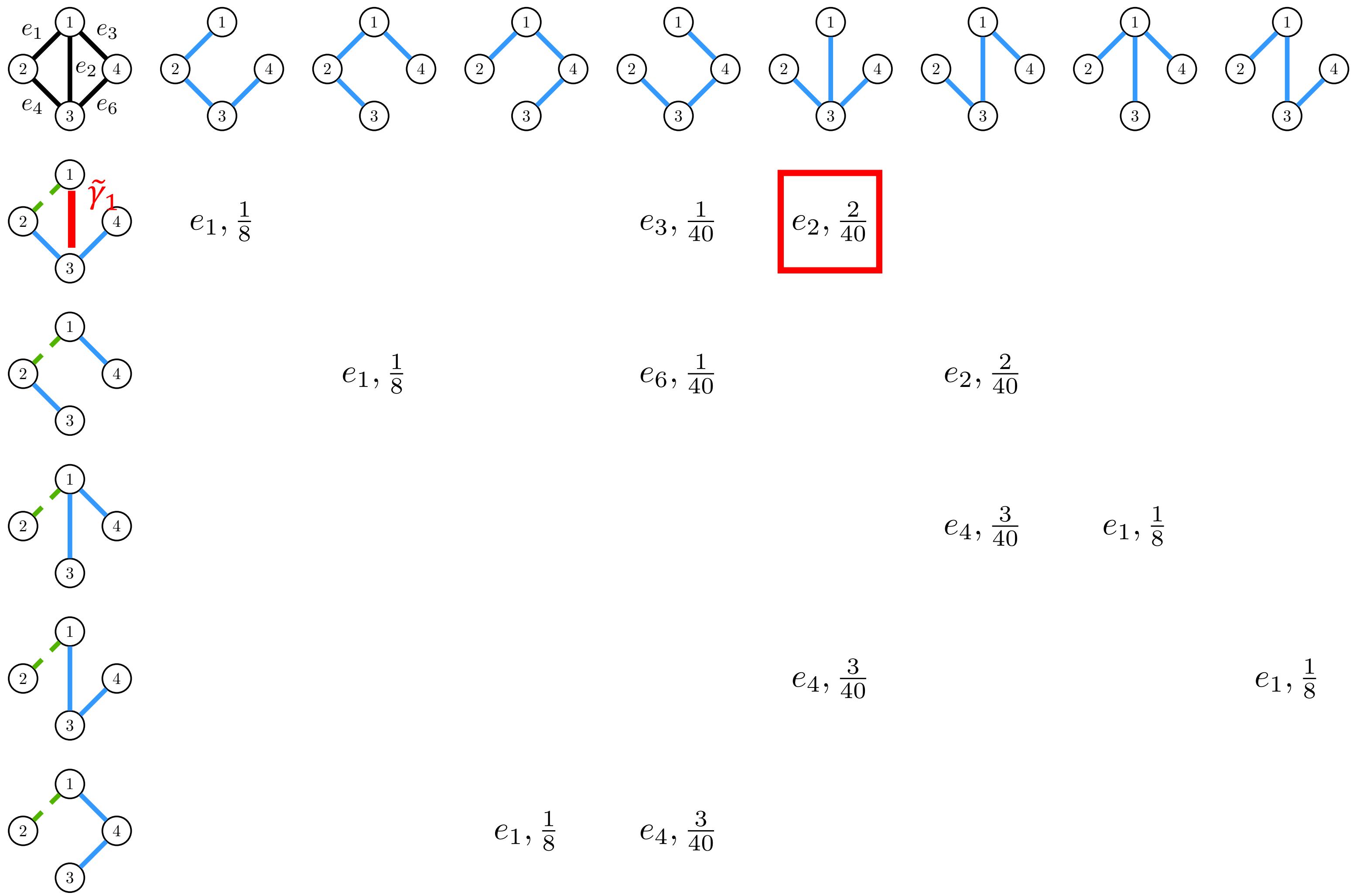
Coupling table has more structure

Recover the original distribution by adding ≤ 1 "make-up edge" to conditional distribution



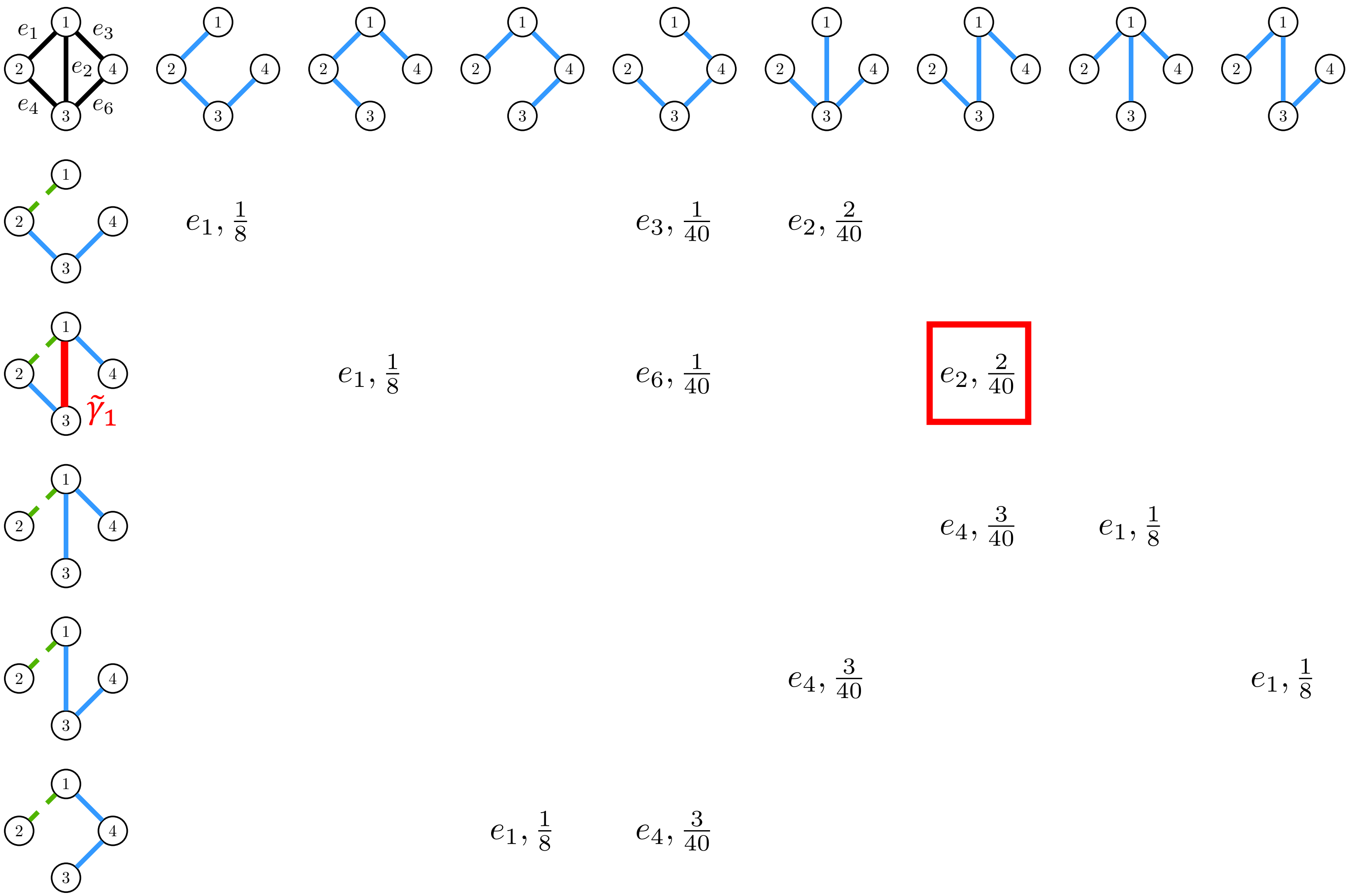
Coupling table has more structure

Recover the original distribution by adding ≤ 1 “make-up edge” to conditional distribution



Coupling table has more structure

Recover the original distribution by adding ≤ 1 "make-up edge" to conditional distribution



Coupling table has more structure

Alice, Bob, Charlie

γ_1

Conditional distribution

Anna, Bob, Charlie

$\tilde{\gamma}_1$ "makeup edge"

Original distribution (!)

$$\boldsymbol{Y}_1 - \boldsymbol{Y}_0$$

$$= \mathbb{E}[\boldsymbol{L}_T | \gamma_1] - \mathbb{E}[\boldsymbol{L}_T]$$

$$= \mathbb{E}[\text{Alice} + \text{Bob} + \text{Charlie} | \text{Alice}] - \mathbb{E}[\text{Anna} + \text{Bob} + \text{Charlie} | \text{Alice}]$$

$$= \text{Alice} - \mathbb{E}[\text{Anna} | \text{Alice}]$$

$$= \boldsymbol{L}_{\gamma_1} - \mathbb{E}\left[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1\right]$$

Coupling table has more structure

Alice, Bob, Charlie
γ_1

Conditional distribution

Anna, Bob, Charlie
$\tilde{\gamma}_1$ "makeup edge"

Original distribution (!)

$$\boldsymbol{Y}_1 - \boldsymbol{Y}_0$$
$$= \mathbb{E}[\boldsymbol{L}_T | \gamma_1] - \mathbb{E}[\boldsymbol{L}_T]$$

$$= \|\boldsymbol{Y}_1 - \boldsymbol{Y}_0\|$$
$$= \boldsymbol{L}_{\gamma_1} - \mathbb{E}[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1]$$
$$\le \max(\|\boldsymbol{L}_{\gamma_1}\|, \|\mathbb{E}[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1]\|)$$
$$\le \max(\|\boldsymbol{L}_{\gamma_1}\|, \mathbb{E}[\|\boldsymbol{L}_{\tilde{\gamma}_1}\| | \gamma_1])$$
$$\le \max_e \|\boldsymbol{L}_e\| \le 1$$

Coupling table has more structure

Alice, Bob, Charlie
γ_1

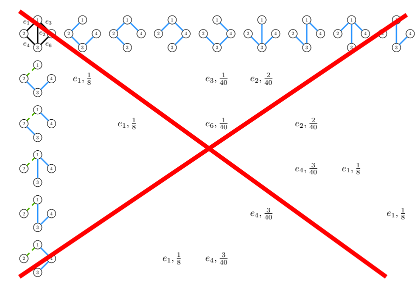
Conditional distribution

Anna, Bob, Charlie
$\tilde{\gamma}_1$ "makeup edge"

Original distribution (!)

$$\mathbb{E}[\boldsymbol{Y}_1 - \boldsymbol{Y}_0] = \boldsymbol{0}$$

$$\mathbb{E}[\boldsymbol{Y}_1 - \boldsymbol{Y}_0] = \mathbb{E}\left[\boldsymbol{L}_{\gamma_1} - \mathbb{E}\left[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1\right]\right] = \boldsymbol{0}$$



So

$$\mathbb{E}\left[\boldsymbol{L}_{\gamma_1}\right] = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1\right]\right]$$

Alice

Anna, "makeup edge"

Important symmery

Coupling table has more structure

Alice, Bob, Charlie
γ_1

Conditional distribution

Anna, Bob, Charlie
$\tilde{\gamma}_1$ "makeup edge"

Original distribution (!)

$$
\begin{aligned}
\mathbb{E}[(\boldsymbol{Y}_1 - \boldsymbol{Y}_0)^2] &= \mathbb{E}\left[\left(\boldsymbol{L}_{\gamma_1} - \mathbb{E}[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1]\right)^2\right] \\
&= \mathbb{E}\left[\boldsymbol{L}_{\gamma_1}^2 + \mathbb{E}[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1]^2\right] \\
&\preccurlyeq \mathbb{E}\left[\boldsymbol{L}_{\gamma_1} + \mathbb{E}[\boldsymbol{L}_{\tilde{\gamma}_1} | \gamma_1]\right] \\
&\preccurlyeq 2\mathbb{E}[\boldsymbol{L}_{\gamma_1}]
\end{aligned}
$$

Coupling table has more structure

$$\mathbb{E}[(\boldsymbol{Y}_1 - \boldsymbol{Y}_0)^2] \preccurlyeq 2\mathbb{E}\left[\boldsymbol{L}_{\gamma_1}\right]$$

$$\mathbb{E}\left[\boldsymbol{L}_{\gamma_1}\right] = \frac{1}{n-1}\mathbb{E}[\boldsymbol{L}_T] = \frac{1}{n-1}\boldsymbol{L}_G$$

So $\mathbb{E}[(\boldsymbol{Y}_1 - \boldsymbol{Y}_0)^2] \preccurlyeq \frac{2}{n-1}\boldsymbol{L}_G$

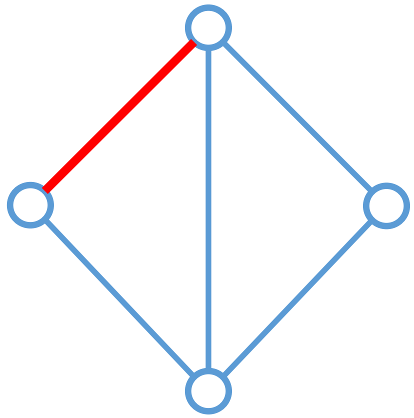
Later steps

What about $\boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$?

Boils down to bounding $\mathbb{E}\left[\boldsymbol{L}_{\gamma_t} \mid \gamma_1, \gamma_2, \ldots, \gamma_{t-1}\right]$

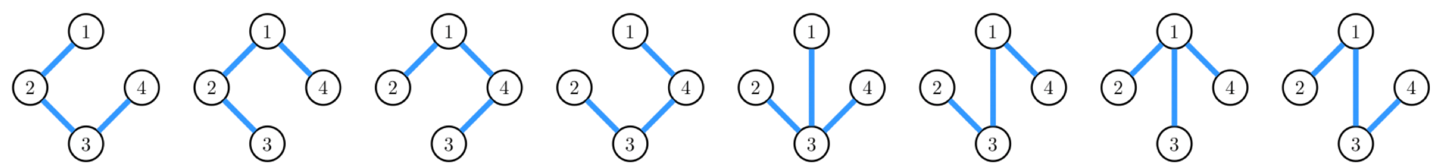
How does conditioning change the distribution?

Graph



Tree distribution

All

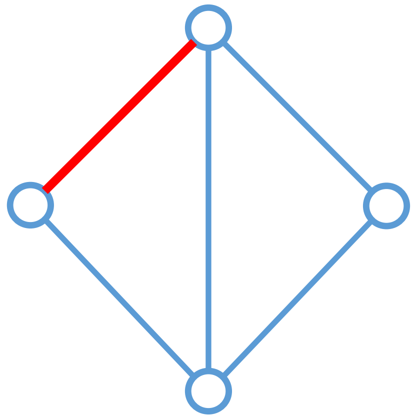


Conditional

Pick a random tree, conditional on red edge present?

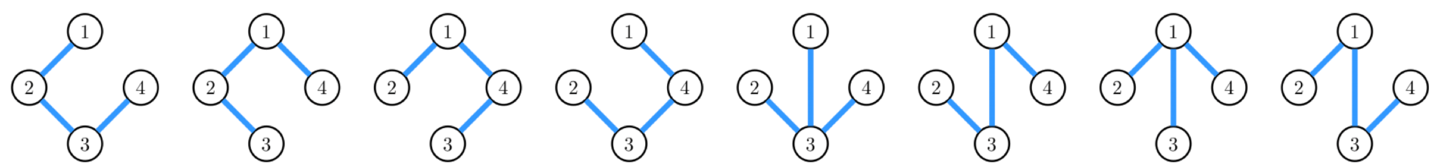
How does conditioning change the distribution?

Graph



Tree distribution

All



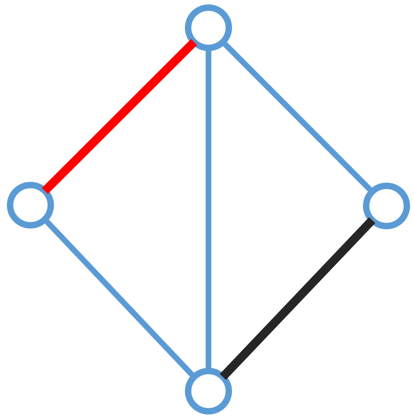
Conditional

“Shrinking Marginals Lemma”

All other edges become less likely

How does conditioning change the distribution?

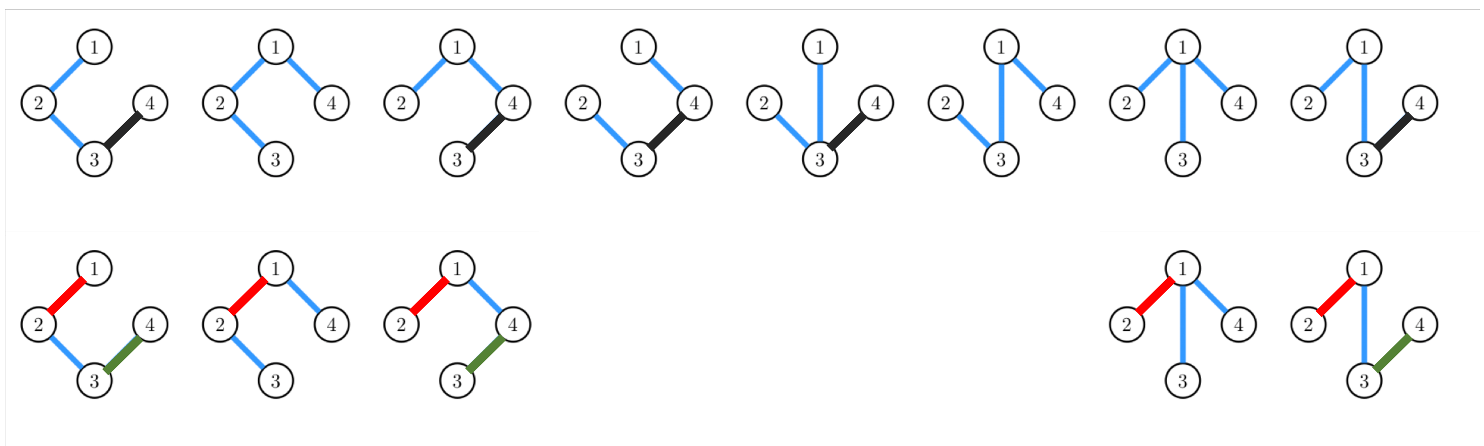
Graph



Tree distribution

All

Conditional

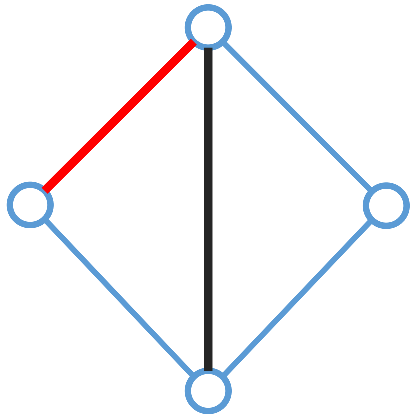


"Shrinking Marginals Lemma"

All other edges become less likely $\frac{5}{8} = 0.625$ vs $\frac{3}{5} = 0.6$

How does conditioning change the distribution?

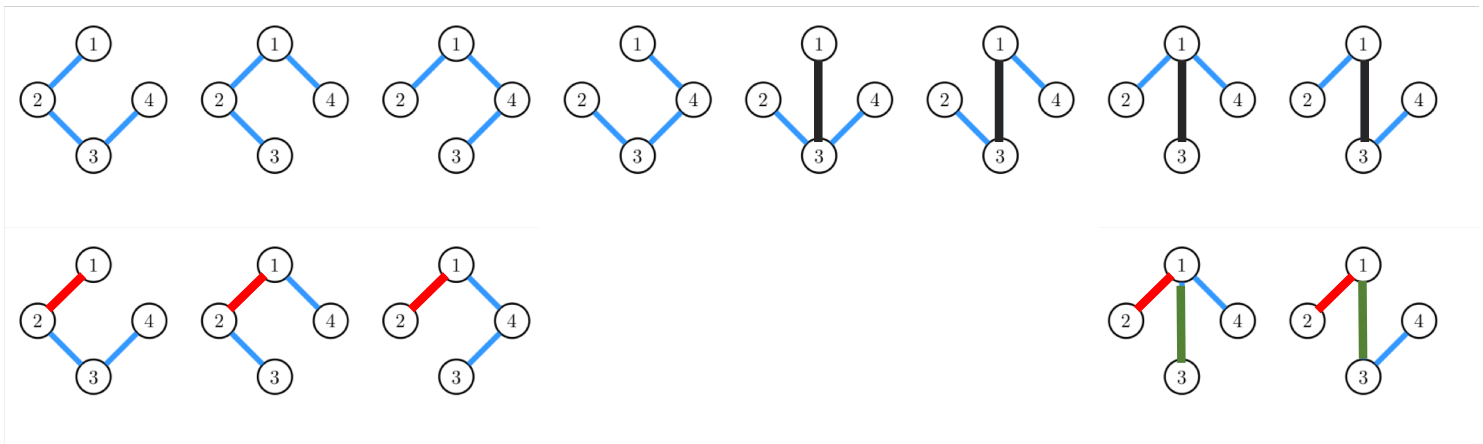
Graph



Tree distribution

All

Conditional



"Shrinking Marginals Lemma"

All other edges become less likely $\frac{4}{8} = 0.5$ vs $\frac{2}{5} = 0.4$

Later steps

$\mathbb{E}[\text{remaining edges} | \gamma_1, \gamma_2, \ldots, \gamma_{i-1}] \preccurlyeq \mathbb{E}[\boldsymbol{L}_T] = \boldsymbol{L}_G$

Shrinking marginals

$\mathbb{E}\left[\boldsymbol{L}_{\gamma_i} | \gamma_1, \gamma_2, \ldots, \gamma_{i-1}\right] \preccurlyeq \frac{1}{n-i} \boldsymbol{L}_G$

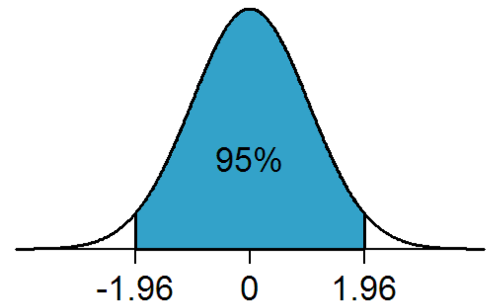
$\sum_i \mathbb{E}[(\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1})^2 | \text{prev. steps}] \preccurlyeq \sum_i \frac{2}{n-i} \boldsymbol{L}_G = O(\log n)\, \boldsymbol{L}_G$

Our Doob martingale

Want to show

$\boldsymbol{Y}_{n-1} = \boldsymbol{L}_T$ is close to $\boldsymbol{Y}_0 = \mathbb{E}[\boldsymbol{L}_T\ ]$

$\boldsymbol{Y}_{n-1} - \boldsymbol{Y}_0 = \sum_i \boldsymbol{Y}_i - \boldsymbol{Y}_{i-1}$



Matrix Freedman (Tropp '11)

Norm $\|\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1}\| \leq 1$

Variance $\|\sum_i \mathbb{E}[(\boldsymbol{Y}_i - \boldsymbol{Y}_{i-1})^2 | \text{ prev. steps}]\| \leq O(\log n)$

implies

$\boldsymbol{L}_T \preccurlyeq O(\log n) \boldsymbol{L}_G$ w.h.p

Concentration of random matrices

Strongly Rayleigh matrix Chernoff [K. & Song '18]

Fixed $\boldsymbol{A}_i \in \mathbb{R}^{d\times d}$, positive semi-definite

$\xi \in \{0,1\}^m$ is k-homogeneous strongly Rayleigh

Random $\boldsymbol{X} = \sum_i \xi(i)\boldsymbol{A}_i$

1. $\|\mathbb{E}\boldsymbol{X}\| = \mu$
2. $\|\boldsymbol{A}_i\| \leq r$

gives

$$\mathbb{P}[\|\boldsymbol{X} - \mathbb{E}\boldsymbol{X}\| > \mu\varepsilon] \leq d\; 2\exp\left(-\frac{\mu\varepsilon^2}{r(\log k+\varepsilon)}\right)$$

Open Questions

Our bound for k-homogeneous strongly Rayleigh

$$\mathbb{P}[\|\boldsymbol{X} - \mathbb{E}\boldsymbol{X}\| > \mu\epsilon] \leq d\, 2\exp\left(-\frac{\mu\varepsilon^2}{r(\log k + \varepsilon)}\right)$$

Remove the $\log k$?

Remove homogeneity condition

Find more applications

Show $\log n$ sparsifier from $O(1)$ spanning trees?

Thanks!