CUTOFF PHENOMENA IN RANDOM WALKS ON RANDOM REGULAR GRAPHS

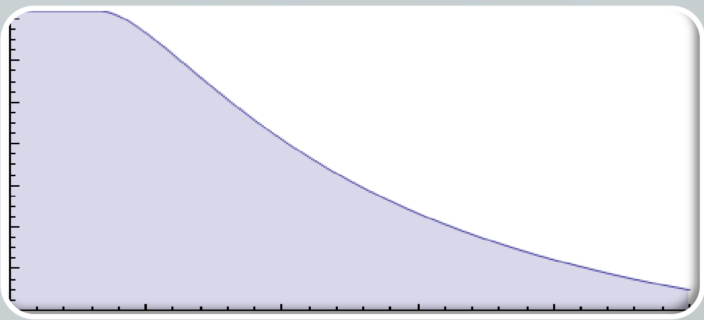
Eyal Lubetzky

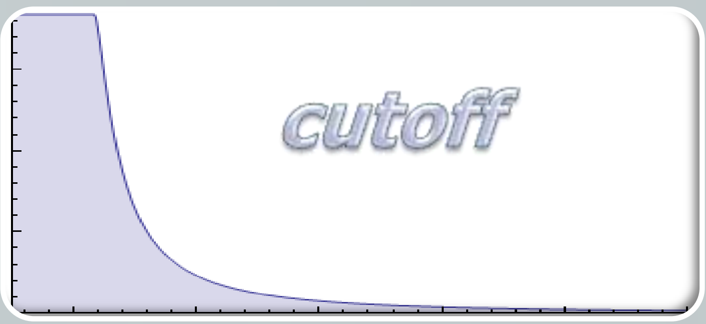
Microsoft Research

Joint work with Allan Sly (UC Berkeley)

The Cutoff Phenomenon

- Describes a sharp transition in the convergence of finite ergodic Markov chains to stationarity.



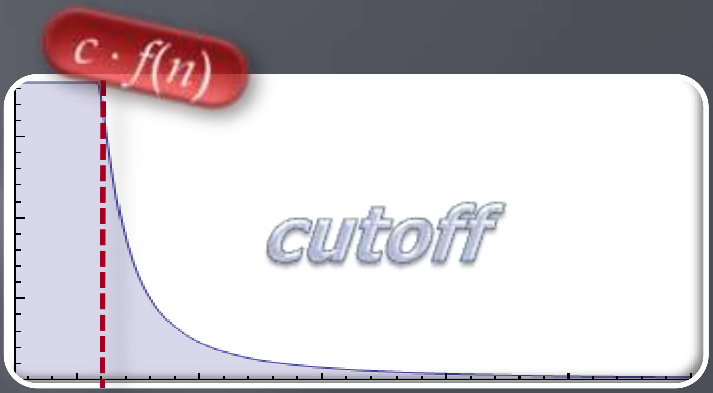


Steady convergence
it takes a while to reach distance ½ from π, then a while longer to reach distance ¼, etc.

Abrupt convergence
the distance from π quickly drops from 1 to 0

Why is cutoff important?

- Consider an MCMC sampler (e.g., heat-bath Glauber dynamics for the Ising Model) with a mixing-time of order $f(n)$.
- Cutoff ⇔ ∃ some $c > 0$ so that:
  - We must run the chain for at least $\sim c \cdot f(n)$ steps to get anywhere near stationarity.
  - Running it any longer than that is essentially redundant.



- Proofs usually require (and thus provide) a deep understanding of the chain (its reasons for mixing).
- Many natural chains are *believed* to have cutoff, yet proving cutoff can be extremely challenging.

Stochastic Ising model

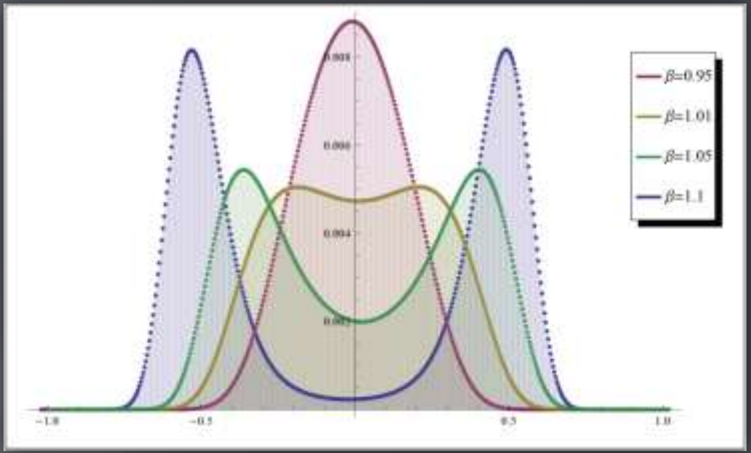
- Underlying geometry: finite graph $G = (V,E)$.
- Set of possible configurations: $\Omega = \{\pm 1\}^V$ (*spins*).
- Probability of a configuration $\sigma \in \Omega$ is given by the *Gibbs distribution* (no external field):

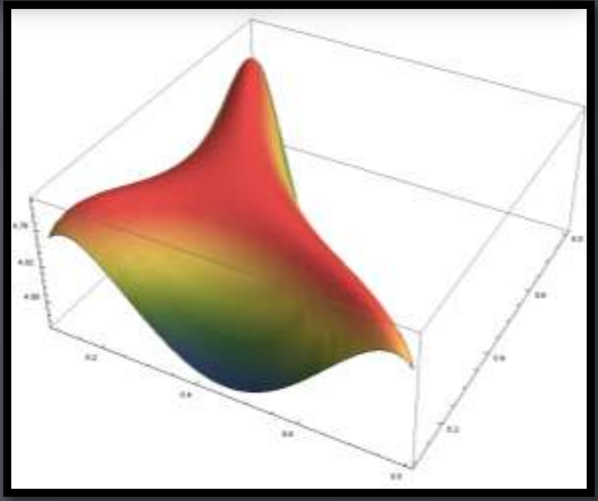
$$\mu_G(\sigma) = \frac{1}{Z(\beta)} \exp\left(\beta \sum_{xy \in E} \sigma(x)\sigma(y)\right)$$

- β = inverse-temperature: as $\beta \uparrow$ ⟹ μ_G favors configurations with aligned neighboring spins.
- Heat-bath Glauber dynamics for μ_G (MC on Ω): Choose $x \in V$ u.a.r. and update its spin according to μ_G conditioned on remaining spins.

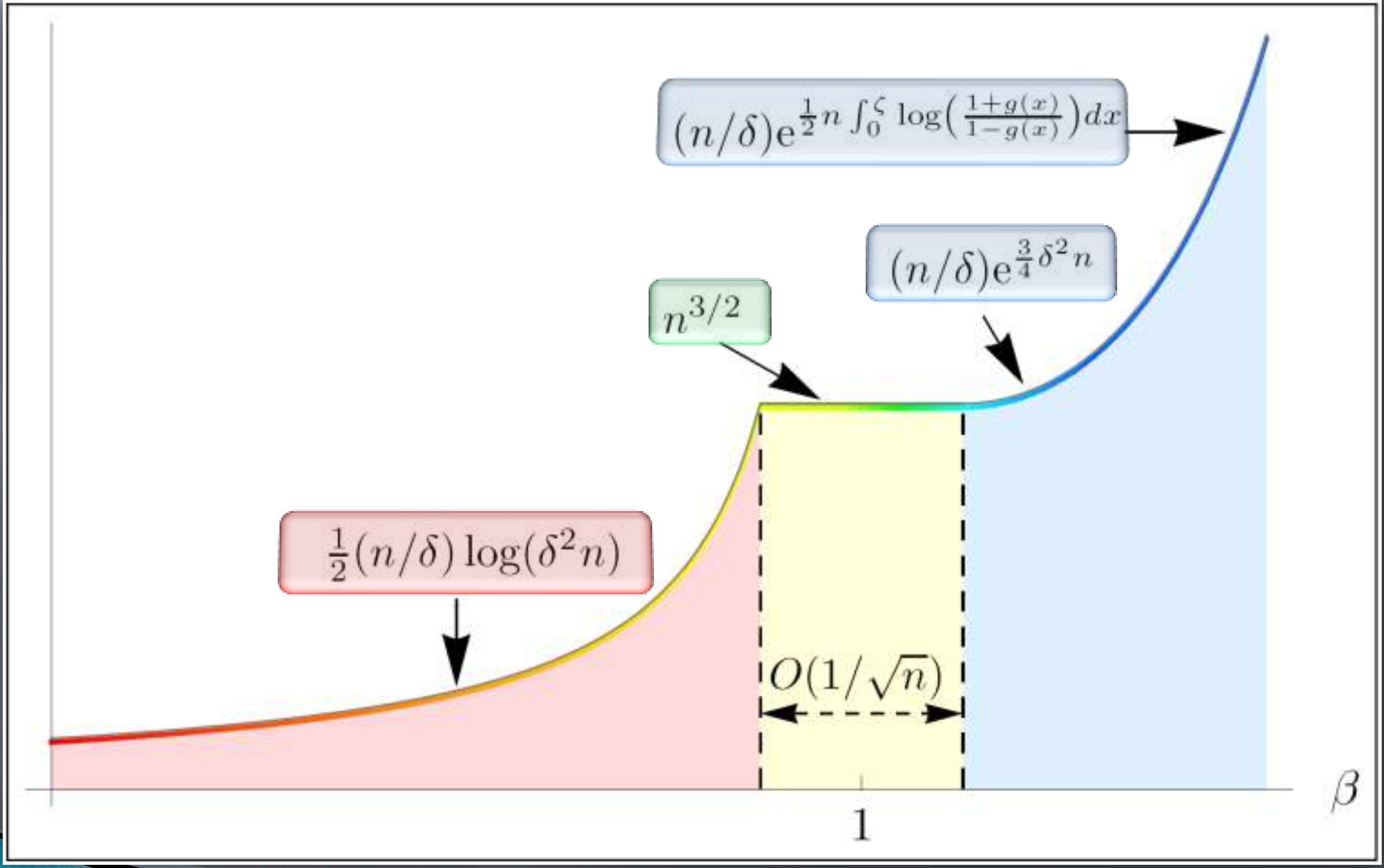
Cutoff for Stochastic Ising

- Ising model on the complete graph:
  - High temperature: [Levin, Luczak, Peres '08]
  - Complete picture: [Ding, L. , Peres '10]
- Key element in analysis:
  - Birth-and-death chains: [Ding, L. , Peres '09]
- Extensions to *q*-state Potts model :
  - [Cuff, Ding, L., Louidor, Peres, Sly]

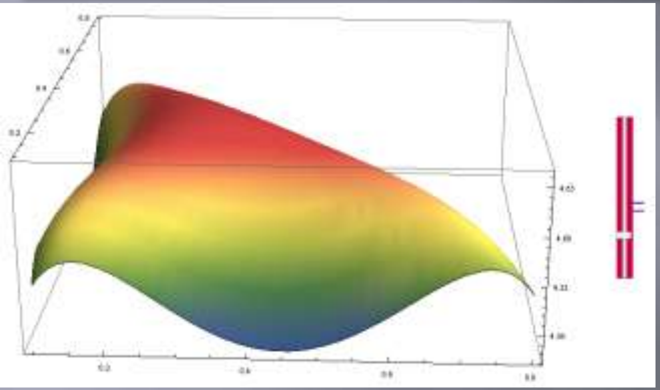




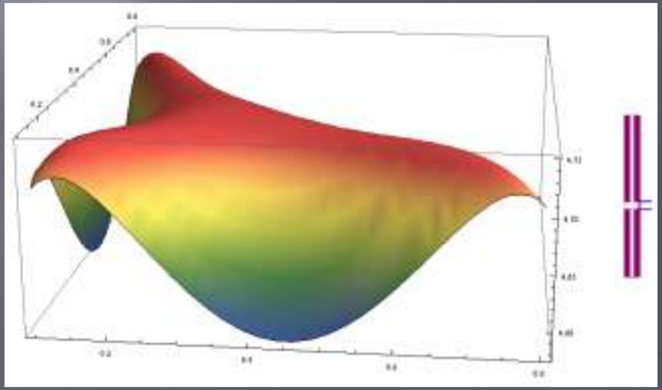
Curie-Weiss model:
Scaling window in the mixing-time evolution



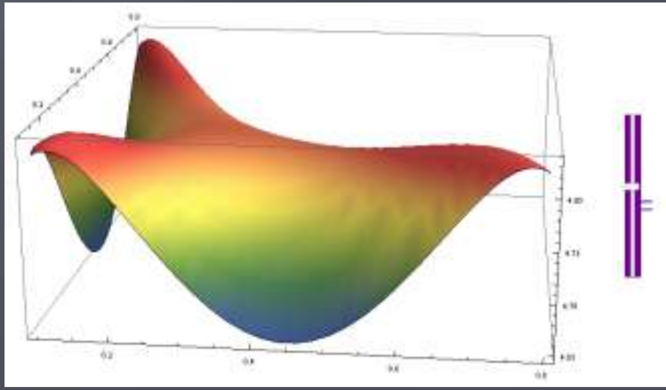
Mean-field Potts model, q=3



$\beta < \beta_m \approx 2.746$: Fast mixing with cutoff
$\beta = \beta_m$: Power law mixing ($n^{4/3}$)



Between (β_m, β_c):
Exponential mixing, but Fast “essential mixing” with cutoff



$\beta \geq \beta_c \approx 2.773$: Exponential mixing (well known)

Cutoff Definition

- The total-variation mixing time of (X_t) w.r.t. some $0<\varepsilon<1$ is $t_{\text{mix}}(\varepsilon)=\min\left\{t : d_{\text{TV}}(t)<\varepsilon\right\}$.
- A family $\left(X_t^n\right)$ has *cutoff* if the following holds:

$$\lim_{n\to\infty}\frac{t_{\text{mix}}(\varepsilon)}{t_{\text{mix}}(1-\varepsilon)}=1 \quad \text{for any } 0<\varepsilon<1.$$

- A sequence (w_n) is called a *cutoff window* if

$$w_n = o(t_{\text{mix}}(\tfrac{1}{4})) ,$$

$$t_{\text{mix}}(\varepsilon)-t_{\text{mix}}(1-\varepsilon)=O_\varepsilon\left(w_n\right) \quad \text{for any } 0<\varepsilon<1 .$$

$$d_{\text{TV}}(t)=\max_{x\in\Omega}\sup_{A\subset\Omega}\left|\mathbf{P}_x(X_t\in A)-\pi(A)\right|$$

Cutoff for Random Walks on $\mathbb{G}(n,d)$

- Consider the *Simple Random Walk* (SRW) on a uniformly chosen 3-regular graph on n vertices.
- Well known: the mixing-time is **whp** $O(\log n)$.
- New precise results include:

$3\log_2 n + O((\log n)^{1/2})$

SRW on random 3-regular n-vertex graph **whp** has

$$t_{\text{mix}}(s) = 3\log_2 n - \left(2\sqrt{6} + o(1)\right)\Phi^{-1}(s)\sqrt{\log_2 n}\ .$$

$\log_2 n + O(1)$

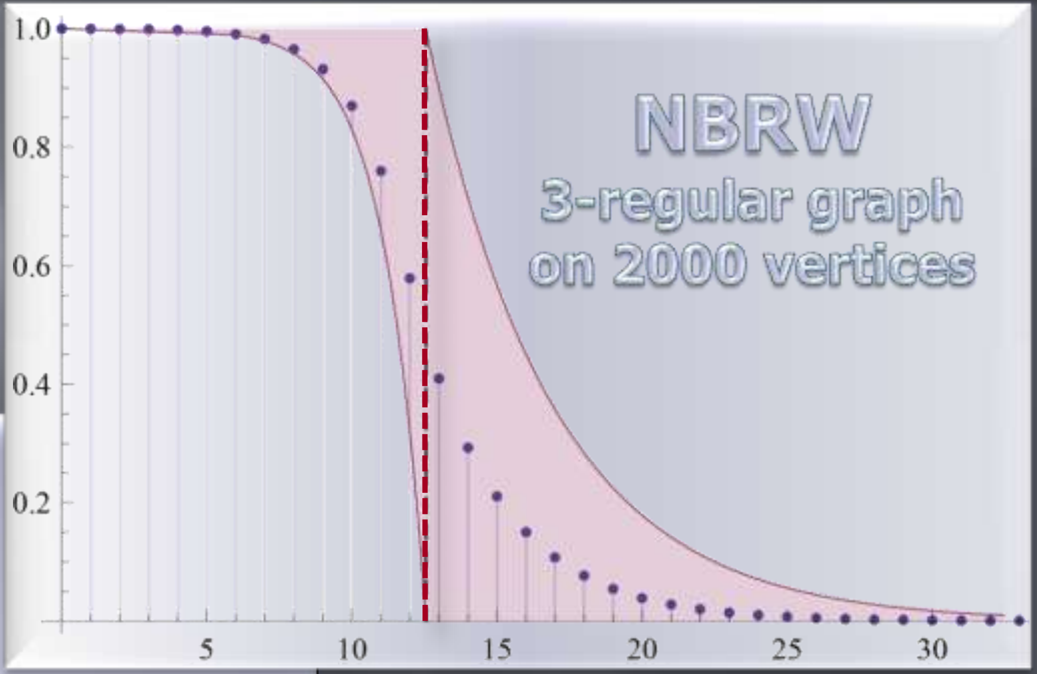
If we forbid *backtracking* (NBRW) then **whp**

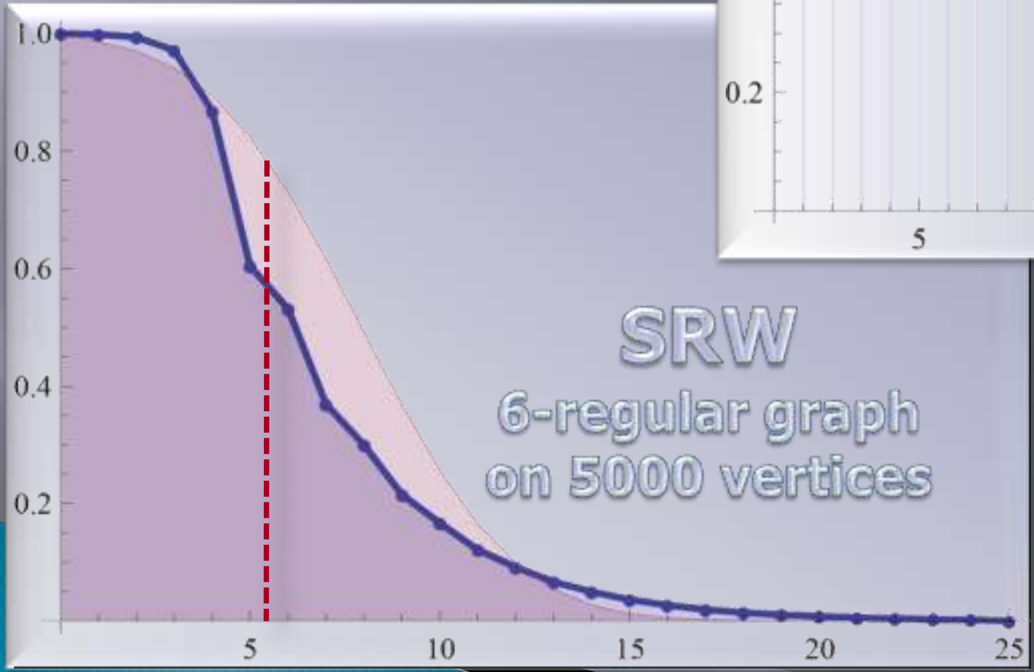
$$t_{\text{mix}}(1-\varepsilon) \geq \lceil \log_2(3n) \rceil - \lceil \log_2(1/\varepsilon) \rceil,$$

$$t_{\text{mix}}(\varepsilon) \leq \lceil \log_2(3n) \rceil + 3\lceil \log_2(1/\varepsilon) \rceil + 4\ .$$

Simulations of RWs / NBRWs

- Graphs sampled via the pairing model:





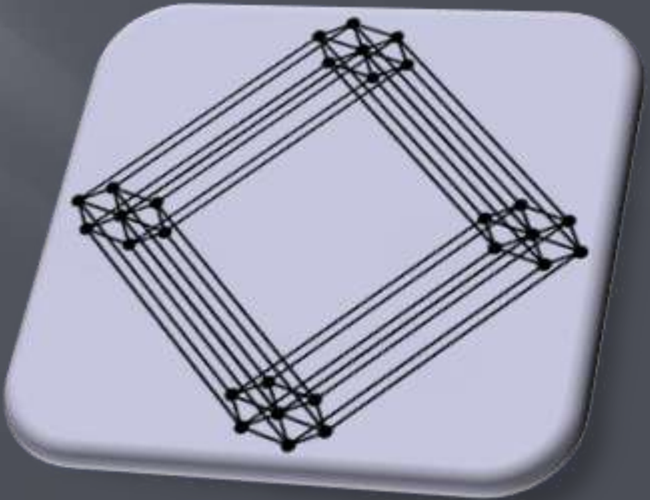
Cutoff History

- Discovered:
  - Random transpositions on S_n [Diaconis, Shahshahani '81]
  - RW on the hypercube, Riffle-shuffle [Aldous '83]
- Named "Cutoff Phenomenon" in the top-in-at-random shuffle analysis [Diaconis, Aldous '86] .
- Cutoff for RW on finite groups ([Saloff-Coste '04]).
- Unfortunately: relatively few rigorous examples, compared to many important chains that are *believed* to exhibit cutoff.

Example: RW on a hypercube

bipartite

- Vertices are binary vectors $\{0,1\}^n$ and there is an edge between any pair of vectors with Hamming distance 1.
- Lazy chain: holds its position with probability ½ . Here: uniformly choose a coordinate and a {0,1} update.
- Projecting onto the Hamming weights gives the classical "Ehrenfest's Urn" (a birth & death chain).
- The Coupon Collector approach:
  $t_{\text{mix}}(\varepsilon) \leq n\log n + c_\varepsilon n,$ whereas
  $t_{\text{mix}}(1-\varepsilon) \geq \frac{1}{2} n \log n - c'_\varepsilon n.$
- [Aldous '83]: lower bound is tight: ½ n log n + $O(n)$ steps suffice!



Determining cutoff

- Merely deciding whether or not there is cutoff can already be highly involved (Diaconis [‘96]).
- In 2004, Peres suggested the “product-condition”
  - $\text{gap} \cdot t_{\text{mix}}(\tfrac{1}{4}) \to \infty$ ⇐ ?⇒ *Cutoff*

  as a cutoff criterion.
  - Necessary for cutoff in a reversible chain.
  - Not always sufficient… ([Aldous ’04, Pak ’06])
- Peres nevertheless conjectured that for many natural chains, cutoff occurs iff $\text{gap} \cdot t_{\text{mix}}(\tfrac{1}{4}) \to \infty$;
  cf. [Chen, Saloff-Coste ’07],
  [Diaconis, Saloff-Coste ’06], [Ding, L., Peres ’09].

SRW on expander graphs

- *Expanders* of fixed degree d : graphs where λ is uniformly bounded away from d.

2nd largest (in abs. value) eigenvalue of the adj. matrix

- SRW has rapid (logarithmic) convergence to stationarity.
- Numerous applications, e.g., de-randomization, space efficient algorithms, etc.
- [Chen, Saloff-Coste '08]: cutoff when measuring convergence under other notions (e.g., L^2–norm).
- Total-variation cutoff (L^1-norm) for any family of transitive expanders remains open.

SRW on random regular graphs

- SRW on $G \sim \mathbb{G}(n,d)$ for fixed $d \geq 3$ **whp** has
  - $t_{\mathrm{mix}}(1/4) = \Theta(\log n)$; gap $\geq c > 0$.
- According to the product-criterion of Peres, this chain should exhibit cutoff **whp**.
- [Berestycki, Durret '08]: studied SRW on $\mathbb{G}(n,3)$, showing that at time $c \log_2 n$, the walk is at distance $\sim \left(\frac{c}{3} \wedge 1\right) \log_2 n$ from its starting point.
- <u>Conjecture</u> [Durrett '07]:

> The mixing time of the lazy RW on the random 3-regular n-vertex graph is asymptotically $6 \log_2 n$.

New results: SRW

- We confirm the above conjecture of Durrett, as well as Peres' product-criterion for $\mathbb{G}(n,d)$:
- Theorem [L., Sly]:

Let $G \sim \mathbb{G}(n,d)$ for $d \geq 3$ fixed. Then **whp**, the SRW on G has cutoff at $\frac{d}{d-2}\log_{d-1} n$ with window of order $\sqrt{\log n}$.

Furthermore, for any fixed $0 < s < 1$, **whp** :

$$t_{\mathrm{mix}}(s) = \frac{d}{d-2}\log_{d-1} n - \left(\frac{2\sqrt{d(d-1)}}{(d-2)^{3/2}} + o(1)\right)\Phi^{-1}(s)\sqrt{\log_{d-1} n}$$

where Φ is the c.d.f. of the standard normal.

The non-backtracking walk

- *Non-Backtracking Random Walk* (NBRW) : does not traverse the same edge twice in a row.
- Reveals the actual behavior of SRWs on $\mathbb{G}(n,d)$: cutoff occurs earlier, with **constant** window!
- Theorem [L., Sly]:

Let $G \sim \mathbb{G}(n,d)$ for $d \geq 3$ fixed. Then **whp**, the NBRW on G has cutoff at $\log_{d-1}(dn)$ with $O(1)$ window.

More precisely, for any fixed $0 < \varepsilon < 1$, **whp** :

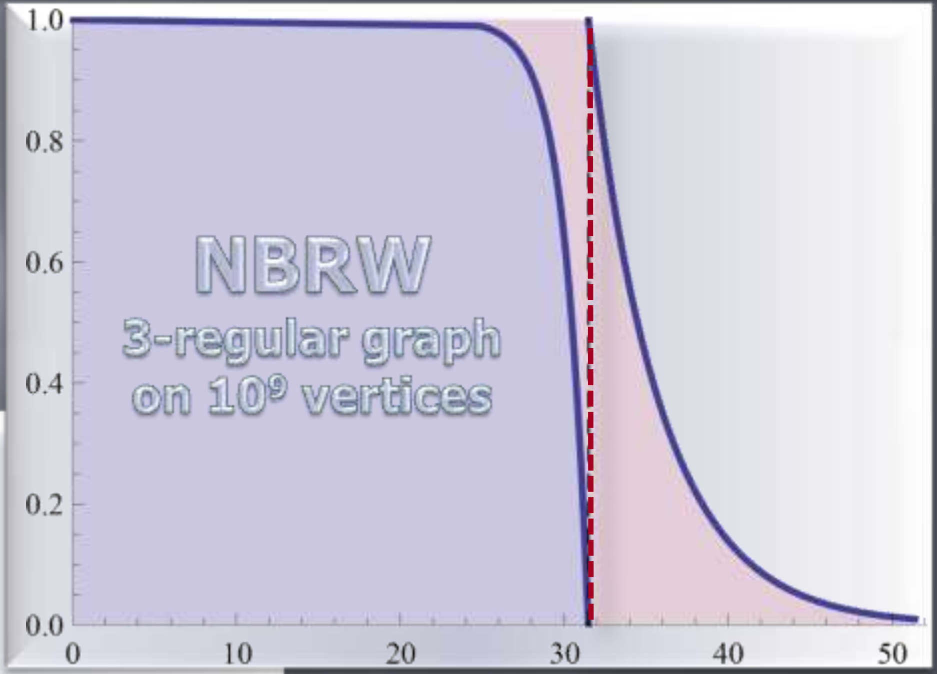
$$t_{\mathrm{mix}}(1-\varepsilon) \geq \lceil \log_{d-1}(dn) \rceil - \lceil \log_{d-1}(1/\varepsilon) \rceil ,$$

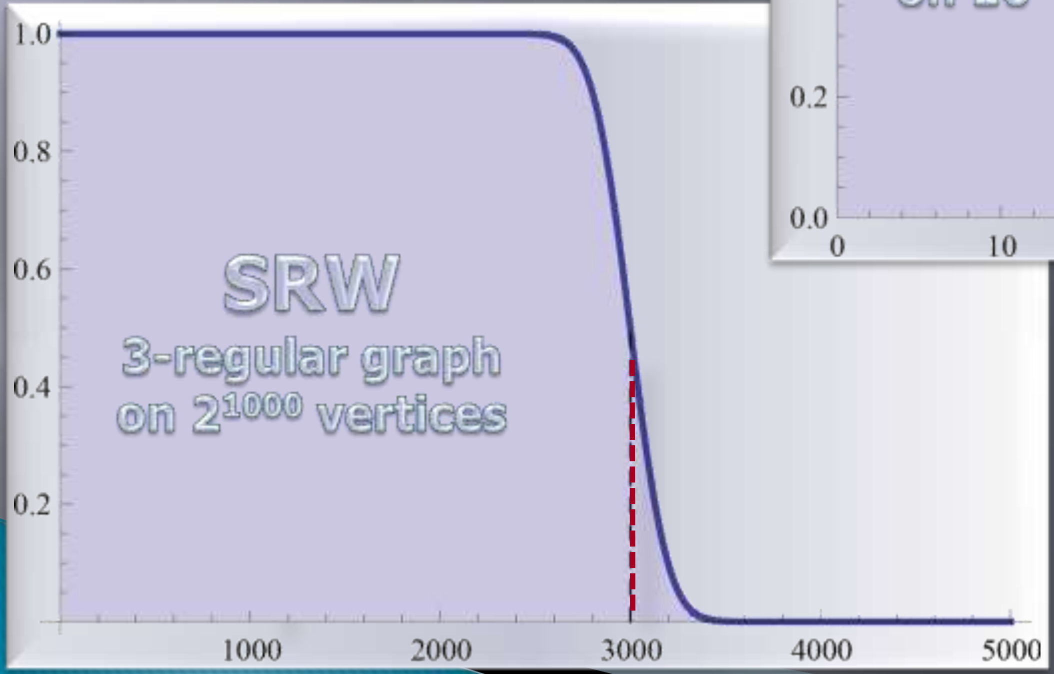
$$t_{\mathrm{mix}}(\varepsilon) \leq \lceil \log_{d-1}(dn) \rceil + 3\lceil \log_{d-1}(1/\varepsilon) \rceil + 4 .$$

Window logarithmic in $1/\varepsilon$

Asymptotic behavior of the RWs

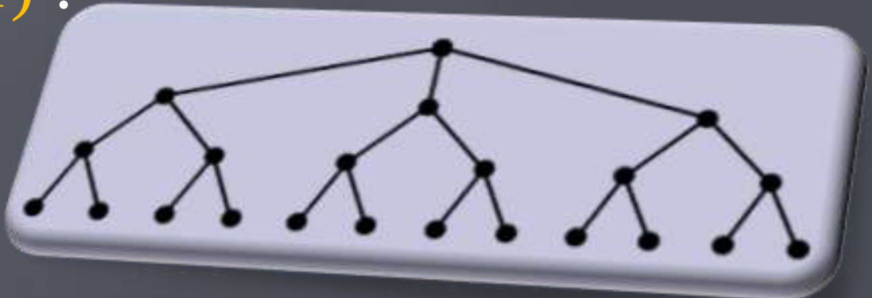
- Bounds on $d_{\mathrm{TV}}(t)$ following the above theorems:





Insight: cutoff for SRW & NBRW

- Consider a d-regular tree, rooted at the starting point of the RW, where the walk mixes precisely upon hitting one of the leaves.
- NBRW cannot backtrack up the tree
$\Rightarrow$ reaches a leaf after precisely $\log_{d-1} n$ steps.
- Height of SRW ~ biased 1D RW with speed $\frac{d-2}{d}$
$\Rightarrow$ expected hitting time to a leaf = $\frac{d}{d-2}\log_{d-1} n$ with std. dev. of $O(\sqrt{\log n})$.



NBRWs *do* mix faster

- Formally: MC on the set of $d\,n$ directed edges. In most applications: project onto the n vertices.
- [Alon, Benjamini, L., Sodin ’08]: compared the *mixing rate* (spectral parameter) of this projection vs. the SRW on expanders (“NBRWs mix faster”).
- Unclear how that spectral data translates into a direct comparison of the two mixing times.
- New results: NBRW is indeed $d/(d\text{-}2)$ times faster, even for the original chain (on directed edges).
- Moreover, we pinpoint the cutoff location up to a window of order $\log_{d\text{-}1}(1/\varepsilon)$.

Graphs with unbounded degree

- ***Q***: Recalling the cutoff window of $\log_{d-1}(1/\varepsilon)$ for NBRWs, what would happen if $d \to \infty$?

  A: "non-mixed" $\to$ "mixed" transition in 2 steps!

- Theorem [L., Sly]:

Let $G \sim \mathbb{G}(n,d)$ for $d = n^{o(1)}$, where $d \to \infty$ with n . Then for any fixed $0 < s < 1$, the NBRW on G **whp** satisfies

$$t_{\text{mix}}(s) \in \left\{ \lceil \log_{d-1}(dn) \rceil, \lceil \log_{d-1}(dn) \rceil + 1 \right\} .$$

Cutoff within two steps!

- Here d is largest possible: if $d = n^{\delta}$ for some $\delta > 0$ then $t_{\text{mix}} = O(1)$ and we cannot discuss cutoff.

Analogous result for SRW

- Corollary [L., Sly]:

  The SRW on $G \sim \mathcal{G}(n,d)$ for $d = n^{o(1)}$, $d \to \infty$ with n , **whhp** has cutoff at $\frac{d}{d-2}\log_{d-1} n$ with window $\sqrt{\log n / d \log d}$. If also $\frac{\log\log n}{\log n} d \to \infty$, the results for NBRWs apply.

- Window becomes narrower with d, and as it turns $o(1)$, the SRW coincides with the NBRW.
- $\Rightarrow$ For $\mathcal{G}(2^m, m)$ we expect to have cutoff **whp** at $\sim (\log 2) m / \log m$ with a 2 step window!
- Compare to the hypercube: there the lazy RW has cutoff at ½ $m \log m$ with window of $O(m)$ …

Proof ideas

Thm 1 (SRW)

- Exploration Process via the configuration model.
- Burn in period for a new locally-tree-like starting pt.
- Analyze local geometry: neighborhoods & cuts.

Thm 2 (NBRW)

- <u>O(1)-window challenge</u>: only $O(1)$ burn-in allowed, and typical cuts have $O(1)$ size (no large deviation argument).
- <u>Solution</u>: amplify the cuts analysis with a one vs. many *Poissonization* argument; delicate analysis of local geometry.

Thm 3 ($d \gg 1$)

- Obtain error bounds to beat ~ $\exp(-d^2)$ probability of the configuration graph producing a non-simple graph.

Additional results: testing cutoff

- How can we tell whether our $G \sim \mathbb{G}(n, d)$ is "typical" and the RW indeed exhibits cutoff?
- We provide a randomized algorithm that given any $0 < \varepsilon < ½$ has
  - Runtime: $\tilde{O}_\varepsilon(n\, t_{\mathrm{mix}}(\varepsilon))$ [optimal up to poly-log factors].
  - Returns estimates so that **whp**:

$$\begin{cases} t_{\mathrm{mix}}(\varepsilon) \le \tilde{t}(\varepsilon) \le t_{\mathrm{mix}}(\tfrac{\varepsilon}{2}) \\ t_{\mathrm{mix}}(1-\tfrac{\varepsilon}{2}) \le \tilde{t}(1-\varepsilon) \le t_{\mathrm{mix}}(\varepsilon) \end{cases}$$

- Cutoff $\Leftrightarrow \dfrac{t_{\mathrm{mix}}(\varepsilon)}{t_{\mathrm{mix}}(1-\varepsilon)} \to 1 \Leftrightarrow \dfrac{\tilde{t}(\varepsilon)}{\tilde{t}(1-\varepsilon)} \to 1$

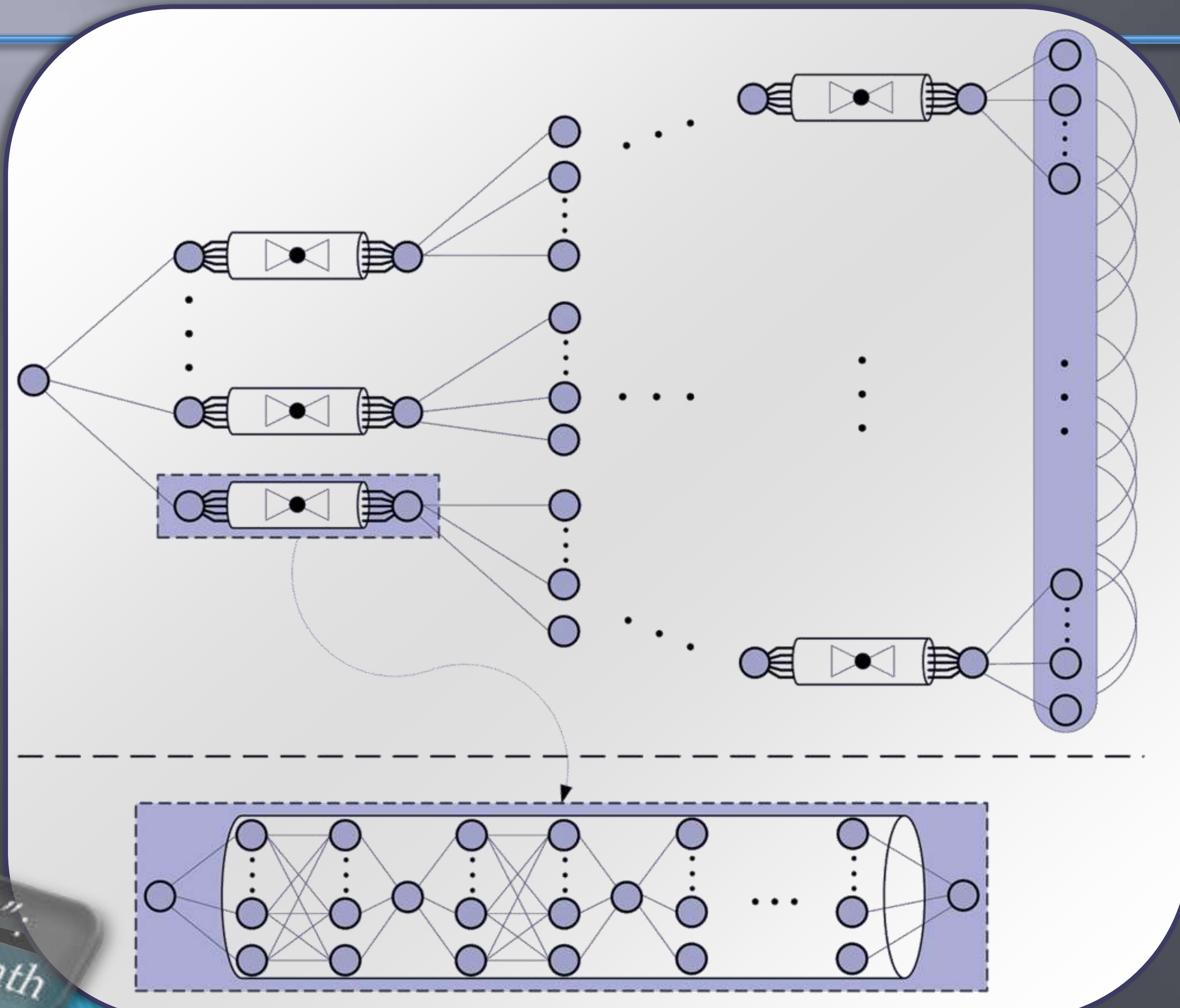
Explicit constructions: SRW cutoff

- Mimicking the structure of a typical $G \sim \mathcal{G}(n, d)$ $\Rightarrow$ explicit construction of d-regular graphs where the SRW exhibits cutoff (at essentially any prescribed location):

For any fixed $d \geq 3$ and any sequence t_n of order between $(\log n, n^2)$, there $\exists$ an explicit family of d-regular graphs where the SRW has cutoff at t_n .

Order of $t_{\text{mix}}(¼)$ is always at least $\log n$ and at most n^2…

Explicit cutoff construction (ctd.)



"cylinder": a d-regular path

Recent progress: Ising on lattices

- Theorem [L., Sly]:

Let $\beta_c = \frac{1}{2}\log(1+\sqrt{2})$ be the critical inverse-temperature for the Ising model on $\mathbb{Z}^2$. Then the continuous-time Glauber dynamics for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^2$ with periodic boundary conditions at $0 \le \beta < \beta_c$ has cutoff at $(1/\lambda)\log n$, where λ is the spectral gap of the dynamics on the infinite volume lattice.

- Analogous result holds for *any* dimension $d \ge 1$… [Previously: even pre-cutoff was only known for the "simpler" 1D case (there with a factor of 2)]

Recent progress on lattices (ctd.)

- Main result hinges on an L^1-L^2 reduction, enabling the application of log-Sobolev inequalities.
- Extending this method gives further results on:
  - Arbitrary external field and non-uniform interactions.
  - Boundary conditions (including free, all-plus, mixed).
  - Other lattices (e.g., triangular, graph products).
  - Other models:
    Anti-ferromagnetic Ising; Gas Hard-core
    Potts (ferro./anti-ferro.); Coloring; Spin-glass.

Open problems

- *Does the* SRW *on any family of transitive 3-regular expanders exhibit cutoff?*
  - *Specifically, does this hold for LPS-expanders?*
- *How does the* NBRW *behave on such a family of graphs (e.g., cutoff pt., window etc.)?*

THANK YOU.