ARCC WORKSHOP: SHARP THRESHOLDS FOR MIXING TIMES

SCRIBES: SHARAD GOEL, ASAF NACHMIAS, PETER RALPH AND DEVAVRAT SHAH

Day I

1. Speaker I: Persi Diaconis

Definition 1 (Sharp Threshold)**.** A Markov chain with transition matrix K and stationary distribution π is said to have (a_n, b_n) cut-off iff

$$\|K_x^\ell - \pi\|_{TV} \to f(c),$$

for $\ell = a_n + cb_n$, such that $b_n/a_n \to 0$ and

$$f(c) \to \begin{cases} 0 & \text{if } c \to \infty \\ 1 & \text{if } c \to -\infty \end{cases}.$$

Take any connected graph of n vertices, say $G = (n, E)$. Graph G can induce transpositions on deck of n cards as follows: each edge $(i, j) \in E$, induces transposition (exchange position) between cards i and j. The Markov chain over n cards is induced by pick a node at random every time and then selecting an edge incident on the node at random and inducing the corresponding transposition on the deck of cards. (See work of Diaconis-Saloff-Coste for detailed definition and upper bound on mixing times, See work of Wilson for lower bound). The following is conjectured:

Open Question 1. All connected graph induce random walk on the deck of cards with sharp threshold.

Riffle shuffles (and other shuffles) can be abstractly seen as walk on hyperplane chambers (See work by Brown-Diaconis). All eigenvalues of such random walks are known ! The following is conjectured:

Open Question 2. All such random walks on hyperplane chambers have cut-off.

Let $B = (B_1, \ldots, B_n)$ be such that $B_1 \le \cdots \le B_n$. Let $S_B = \{\pi \in S_n : \pi(i) \ge B_i\}$. Consider a random walk on S_B. Choose a random transposition every time. If application of random transposition retains the resulting permutation in S_B then perform it else don't do anything.

Open Question 3 (Hanlon's Walk)**.** All the eigenvalues of the above walks are known for any B. Show that there exists a cutoff.

Date: December 20-23, 2004.

2. Speaker II: Yuval Peres

Consider a sequence of Markov chain (parametrized by state-space size n) with stationary distribution π such that $\max_{x,y} \pi(x)/\pi(y) \leq \beta < \infty$. Let T_{rel} be relaxation time which is the inverse of spectral gap of the Markov chain and let $T_{mix}(\epsilon)$ be ϵ mixing time (with respect to total variation distance). The following is conjectured:

Conjecture 1. $T_{rel}/T_{mix}(0.5) \to 0$ as $n \to \infty$ if and only if cut-off is sharp, that is, for arbitrarily small $\epsilon > 0$

$$\frac{T_{mix}(\epsilon)}{T_{mix}(1-\epsilon)} \to 1.$$

Some speculations by speaker for the conjecture being true are:

1. Mathew's bound on cover time in terms of maximal hitting time.
2. Aldous' theorem that says that covering time is concentrated if the average hitting is smaller order than the average covering time.
3. (Vague) Connection with weak-mixing of ergodic transformation.

P. Diaconis' speculation: Coupon collector's problem, sharp cut-off in mixing times and existence of concentration in the sense of Talagrand's results are correlated phenomenon.

3. Speaker III: Laurent Saloff-Coste

Consider random transpositions for a deck of n cards. It is known that with respect to ℓ_2 norm, the cut-off is $0.5n \log n$. It is also known that $n \log n$ is cut-off with respect to ℓ_∞ norm. The following is conjectured:

Open Question 4. With respect to all ℓ_p norm, $2 < p < \infty$, the cut-off is $0.5n \log n$. Also, for random-to-random permutatations $0.75n \log n$ is conjectured to be cut-off.

Consider random walk on group $SL_n(\mathbb{F}_q)$ with some small set of generators. The following is conjectured:

Conjecture 2. It is known that this random walk has spectral gap bounded away from 0 for all n, that is, $SL_n(\mathbb{F}_q)$ is expander. It is conjectured that there is cutoff.

4. Speaker IV: Ben Morris

It is known that Thorp shuffling has mixing time $O(poly(\log n))$ for deck of n cards when $n = 2^d$ for some $d \in \mathbb{N}$ (See work by Morris). The following is conjectured:

Open Question 5. For any n, the mixing time is $O(poly(\log n))$. Further, it is $\Theta(\log^2 n)$.

5. Speaker V: David Aldous

(Notes contributed by Aldous)

Consider a finite state irreducible continuous-time Markov chain. Write π for its stationary distribution. For each initial state x there is a first time τ_x such that the variation distance between $P_x(X_t \in \cdot)$ and $\pi(\cdot)$ is at most $1/4$, say. Define τ as the median of τ_x when x is distributed as π. Define τ_2 as the relaxation time ($= 1/(\text{spectral gap})$) of the chain. Here is a slight modification of Conjecture 1 of Yuval Peres.

Conjecture 3. Consider a sequence of chains such that $\tau/\tau_2 \to \infty$ as $n \to \infty$. Then for any fixed $\varepsilon > 0$, as $n \to \infty$

$$\pi\{x : ||P_x(X_{(1-\varepsilon)\tau} \in \cdot) - \pi(\cdot)|| > 1 - \varepsilon\} \to 1$$
$$\pi\{x : ||P_x(X_{(1+\varepsilon)\tau} \in \cdot) - \pi(\cdot)|| < \varepsilon\} \to 1$$

Possible proof strategy. Here is an idea for a rather abstract proof strategy, though no doubt a more concrete proof would be preferable. Consider a continuous-time general-space Markov process with some stationary distribution π, and set

$$d_x(t) = ||P_x(X_t \in \cdot) - \pi(\cdot)||.$$

Then consider the random function

$$F_\infty(t) = d_X(t) \text{ when } X \text{ has dist. } \pi. \tag{1}$$

define $F_n(t)$ similarly for a n-state chain.

Conjecture 4. Consider a sequence of finite-state chains, and assume (by scaling the time unit) that each has $\tau = 1$. Suppose $F_\infty(t)$ is a subsequential weak limit of $F_n(t)$. Then either F_∞ is the step function that jumps from 1 to 0 at time 1, or it is the function (1) for some Markov process.

This should imply the Peres conjecture, as follows. We want to prove the limit is the step function; if not, it is F_∞ for some process. But the Peres-conjecture hypothesis says $\tau_2/\tau \to 0$, implying the limit process would have $\tau_2 = 0$, which is impossible, or at least contradicts having a non-degenerate F_∞.

Why might Conjecture 4 be true? In a n-state chain write

$$g_n(x, y; t) = P_x(X_t = y)/\pi_y.$$

Take an infinite sequence $(\xi_i, i \geq 1)$ which are i.i.d. with distribution π. Write

$$G_n(i, j; t) = g_n(\xi_i, \xi_j; t)$$

and then regard $G_n(i, j)$ as a random function. Note that the distribution of F_n above is determined by the distribution of $G_n(1, 2)$. The hypothesis of Conjecture 2 is that, in a subsequence, F_n converges in distribution to F_∞, and by a compactness argument we can suppose that (in some appropriate sense) the random infinite matrices $(G_n(i, j))$ converge in distribution as $n \to \infty$ to a random infinite matrix $(G_\infty(i, j))$. By construction, the random marices $(G_n(i, j))$ have a certain exchangeability property called *joint exchangeability*, which persists in the limit $(G_\infty(i, j))$. By the Aldous-Hoover-Kallenberg theory (see Aldous, "Exchangeablility and Related Topics", Springer LNM 1117, 1985) there is a representation which is essentially of the form

$$G_\infty(i, j; t) = g_\infty(\xi_i, \xi_j; t)$$

where the (ξ_i) are i.i.d. uniform$(0, 1)$ and g_∞ is a measurable function. Thus is would be enough to show that there exists some reversible Markov process on state-space $(0, 1)$ with uniform stationary distribution and whose transition density is $g_\infty(x, y; t)$.

Showing this is no doubt technically challenging; but the rough idea is that the Chapman-Kolmogorov equations for the n-state chain are preserved in the and that such equations serve to define a Markov process. A technical issue is that it might be the case that $F_n(t) \to 1$ for $t < 1/2$, say, in which case we (informally) lack absolute continuity and have $g_\infty(x, y; t)$ identically zero for $t < 1/2$. Thus the issue is to reconstruct a Markov process from its large-time transition densities.

6. Panel Discussion

The panel discussion generated various suggestions from panel and audience. The following are some of the central ones:

1. Show that expanders have sharp cut-off.
2. Characterize operations on Markov chains that retain the sharp thresholds.
3. Use optimal strong stationary time and its concentration property to characterize the sharp cut-off. Similarly one can use "threshold rule" of Lovasz-Winkler.

Day II

7. Speaker I: David Revelle

Theorem 1 (Peres and Revelle). *Let $\{G_n\}$ be a sequence of regular graphs for which $|G_n| \to \infty$ and the maximal hitting time satisfies $t^* \leq K|G_n|$ for some constant K. Then there are constants c_1, c_2 depending on ϵ and K such that*

$$c_1|G_n|(T_{rel}(G_n) + \log|G_n|) \leq \tau(\epsilon, \mathbb{Z}_2 \wr G_n) \leq c_2|G_n|(T_v(G_n) + \log|G_n|).$$

Open Question 6. For the walks in Theorem 1 it is known that there is an L^∞ cut-off. What is the exact value of this threshold?

Open Question 7. For $\mathbb{Z}_2 \wr \mathbb{Z}_n$ there is no total variation cut-off, and for $\mathbb{Z}_2 \wr \mathbb{Z}_n^2$ there is total variation cut-off at $\Theta(n^2 \log^2 n)$. What is the behavior of $\mathbb{Z}_2 \wr \mathbb{Z}_n^d$ for $d \geq 3$?

Open Question 8. For $d \geq 3$, the total variation mixing time for $\mathbb{Z}_2 \wr \mathbb{Z}_n$ is bounded between the expected cover time $\mathbb{E}C$ and $\mathbb{E}C/2$. What is the correct answer?

Open Question 9. What can be said if $\mathbb{Z}_2$ is replaced by $\mathbb{Z}_k$ in the above walks?

8. Speaker II: Laurent Saloff-Coste

Theorem 2 (Saloff-Coste). *Fix $1 < p \leq \infty$, and consider a family $\{K_n\}$ of reversible chains with spectral gaps λ_n and L^p mixing times $t_n = T_n^p(\epsilon)$. Then the following are equivalent:*

1. $\lambda_n t_n \to \infty$
2. *cut-off at time t_n in L^p*
3. *(t_n, λ_n^{-1}) cut-off in L^p*

Open Question 10. The example of Aldous shows that the above theorem cannot hold in L^1. But does it hold for separation cut-off and/or entropy cut-off?

9. Open Question Session

Open Question 11 (Winkler). For a directed graph it is computational easy to exactly count the number of Eulerian circuits. Although this problem is $\#P$-complete in the case of undirected graphs, Brightwell and Winkler propose a Markov chain that if rapidly mixing would provide an efficient scheme to approximate this quantity. Question: What is the mixing time of their chain?

Open Question 12 (Babai)**.** Does there exist a family of bounded degree, vertex transitive graphs $G_n = (V_n, E_n)$ with $|V_n| \to \infty$, and a sequence d_n satisfying

1. $|S_{d_n}(v_0)| = |B_{d_n}(v_0) \setminus B_{d_n-1}(v_0)| \geq |V_n|/2$
2. 99% of the vertices in $S_{d_n}(v_0)$ have no outward neighbor
3. $\text{Diam}(G_n) \geq 1.9d_n$

Open Question 13 (Tetali)**.** For several combinatorial objects that are enumerated by the Catalan numbers, there are corresponding Markov chains with stationary distributions uniform over those sets of objects. What is the mixing time for those chains? Here are three concrete examples:

- Consider the following walk on the triangulations of a regular n-gon: Choose a quadrilateral uniformly at random and flip the diagonal. The total variation mixing time for this walk satisfies
$$n^{3/2} \leq \tau_{TV} \leq n^5 \log n.$$
The upper bound was shown by Tetali and McShine, and Aldous conjectured that $n^{3/2}$ is the correct order mixing time.
- Consider the following walk on non-negative lattice paths from $(0,0)$ to $(2n,0)$ with steps $(1,1)$ and $(1,-1)$: Uniformly choose a local peak or valley and flip its orientation. Wilson shows that the total variation mixing time is $\Theta(n^3 \log n)$. Is there a sharp threshold?
- Consider the following walk on partitions of a regular $2n$-gon by non-intersecting chords: Pick a pair of chords uniformly at random and swap the vertices they connect if the result is a non-intersecting partition. It is known the total variation mixing time is $O(n^4 \log n)$, but is conjectured that the mixing time is $\Theta(n^2 \log n)$.

Open Question 14 (Peres)**.** The semi-random random transposition walk on S_n proceeds as follows: At each time t, choose $i \in \{0, \dots n\}$ uniformly at random and swap the cards in positions i and $j = t \bmod n$. The mixing time is $\Theta(n \log n)$. Is there cut-off? More generally, consider the walk in which position i is chosen at random, and position j is chosen either deterministically or with randomness independent of the choice of i. For this family of walks, Mossel, Peres and Sinclair show an $n \log n$ upper bound. Is there a matching lower bound?

Open Question 15 (Wilson)**.** Fix a finite group G, and integer d. Choose a random set of d elements of G. What is the typical (e.g. median) mixing time? Is it true that the hypercube has largest typical mixing time among groups of a given size?

Open Question 16 (Randall)**.** Start with an $n \times n$ grid, and pick a square uniformly at random. If no edges are present in the square, add all of the edges of the square with probability λ. If all edges are present, remove the square. The stationary distribution satisfies $\pi(\gamma) \sim \lambda^{|\gamma|/4}$. Is this chain rapidly mixing for small λ.

Open Question 17 (Saloff-Coste)**.** Is the Peres Conjecture true for convolutions on the cycle?

Open Question 18 (Saloff-Coste)**.** Consider the Burnside group $B(3,d)$ with d generators, and all elements having order 3. This groups has order approximately $3^{\binom{d}{3}}$. Since the spectral gap is approximately $1/d$ (see Stong), this immediately gives an upper bound on mixing time of order d^4. Comparing the size of the generating set to the size of the group, mixing time is lower bounded by $d^3/\log d$. What is the mixing time for this walk?

Open Question 19 (Saloff-Coste)**.** Find an explicit set of $2d$ elements in $\mathbb{Z}_2^d$ whose associated walk reaches stationarity after order d steps. It is known that the walk generated by a typical set will mix in this time.

Open Question 20 (Morris)**.** For simple random walk on a path with m edges, relaxation time satisfies $1/\lambda \sim 2m^2/\pi^2$. Is this (asymptotically) maximal among graphs with m edges? What if we restrict to trees with m edges?

Day III

10. Speaker I: Peter Winkler

Given distributions ρ and ν on the state space of a Markov chain, a **stopping rule from** ρ **to** ν is a random stopping time Γ for which the Markov chain begun at ρ and stopped at Γ is distributed as ν, and is independent of Γ. It is **optimal** if $E\Gamma$ is minimal over all such stopping rules. Lovász and Winkler described four general characterizations of optimal stopping rules with different properties. One of these characterizations, a **threshold rule**, has the additional property that it is the most concentrated about the (common) mean of all optimal stopping rules, in the sense that it minimizes the maximum-time-to-stop (as well as in many other senses). In the context of mixing of finite markov chains, it makes sense to consider optimal stopping rules from the "worst" initial state to the stationary distribution. In this context:

Open Question 21. Is there a connection between the properties of the "threshold" optimal stopping rule and the behavior of the total variation mixing time? In a family of markov chains, does some property of the former imply the existence of a threshold in the latter?

Open Question 22. Each of the four characterizations of optimal stopping rules require the computation of some value for each state in the chain, depending on the initial and final distributions, a process which takes polynomial time in the size of the state space. Is the behavior of the optimal stopping rule robust under small errors in these calculations?

11. Speaker II: David Wilson

Wilson conjectures that for several interesting families of markov chains, the decay behavior of total variation distance, considered as (for instance) the size of the state space grows, is governed by the asymptotic ($t \to \infty$) rate of decay of total variation distance. Also, the same is true for separation distance. For an exact formulation, numerical evidence, and a relationship to the second eigenvector, see **Mixing Times of Lozenge Tilings and Card Shuffling Markov Chains**, `math.PR/0102193`, a much better exposition than we can hope to give in these notes.

12. Speaker III: Jim Fill

A method for building strong stationary times (stopping rules, as described above, that don't depend on the time elapsed), as described by Diaconis and Fill, involves coupling the chain to a "dual" chain, in which the stopping rule for the first chain is exactly when the dual chain hits some absorbing state. Denote the original chain X_t, the dual chain X_t^*, and the distribution of the original chain conditioned on the history of the dual chain

$$\mathcal{L}(X_t|\mathcal{F}_t^*) = \Lambda(X_t^*, \cdot).$$

Theorem 3 (Diaconis & Fill)**.** *Given a MC with transition kernel K, stationary distribution π, and distribution ρ_t at time t, there exists a dual chain with transition kernel K^* and initial distribution ρ_0^* iff*

$$\rho_0 = \rho_0^* \Lambda \qquad \text{and} \tag{2}$$
$$\Lambda K = K^* \Lambda. \tag{3}$$

Note that this implies $\Lambda K^t = (K^*)^t \Lambda$ for all t.

Connection of set-valued duals to the Evolving Set process: If we want the dual chain to have as it's state space the set of nonempty subsets of the original state space, then equation (3) is:

$$\sum_{x \in x^*} \frac{\pi(x)}{\pi(x^*)} K(x,y) = \sum_{y^* \ni y} K^*(x^*, y^*) \frac{\pi(y)}{\pi(y^*)}$$

Claim: Let G be the evolving set kernel corresponding to the time-reversal of K, as described by Morris and Peres. Then

$$K^*(x^*, y^*) = \frac{\pi(y^*)}{\pi(x^*)} G(x^*, y^*),$$

the kernel corresponding the evolving set process conditioned to absorb on the whole state space, satisfies (3). Also, if X_t is a markov chain and $\check{X}_t$ is the corresponding evolving set process, the two are related via Sigmund duality: $\mathbb{P}_{x^*}(\check{X}_1 \ni y) = \mathbb{P}_y(X_1 \in x^*)$.

Excercise: See work by Morris and Peres for evolving sets; and by Diaconis and Fill, for strong stationary duality; see if you can fill in the details!

13. Speaker IV: Persi Diaconis

Session on random vs. systematic updates

The "first modern Markov chain" was described by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller in 1953, and until recent work by Kannan, Mahoney, and Montenegro, very little was known about it's mixing behavior. The behavior at "high" density is still an open question. The chain is the following: Begin by putting n labeled, nonoverlapping disks of radius ϵ in the unit square. Sequentially, remove each disk, and replace it in a position uniformly amoung those where it doesn't overlap any other disk.

Metropolis et.al., quite naturally, defined one step of the chain to be one systematic "scan" through all the n disks — but a probabilist, at least for ease of analysis, might define one step by moving a random disk. In this sort of chain, at least, it seems intuitive that the systematic updates would at least save a factor of $\log(n)$ in the mixing time, by analogy to the coupon-collector problem.

Open Question 23. Does this chain mix significantly faster than the chain defined by moving random disks?

However, in many cases where one might think systematic updates would save significant amounts of time, this has proved not to be true. For example:

Define a distribution on permutations of n elements by: fix a permutation x_0, let $d(x,y)$ be the minimal number of adjacent permutations needed to get from x to y, and let $\pi(x) \propto \theta^{-d(x,y)}$, for some $\theta \in [0,1]$. Consider the chain used by the Metropolis-Hastings algorithm, with moves as adjacent transpositions, to generate samples from this distribution. Bejamini, Berger, Hoffman, and Mossel showed that if the position considered to update at each step was random, then the chain mixes in $\Theta(n^2)$ steps. Diaconis and Ram showed that if the positions considered proceeded sequentially $1 \ldots n \ldots 1 \ldots$ then the chain (where one scan is one step) mixes in $\Theta(n)$ steps — so the number of adjacent transitions used are of the same order.

Open Question 24. Does the same behavior hold using a different metric $d(\cdot,\cdot)$?

However, generating uniform 3-colorings of a tree using the graph-coloring chain give different results. If we begin at an empty tree, and update random nodes, then the mixing time is exponential in the maximum degree, while a systematic (well–chosen) scan requires only one pass through the tree.

Open Question 25. Is there a chain for which some update orderings are significantly better than random updates, but others are worse?

In some situations, there are natural schemes that lie qualitatively between random and systematic updates: for example, the cyclic-to-random shuffle introduced by Thorpe and studied by Mossel, Peres, and Sinclair who showed a $\Theta(n \log n)$ mixing time.

Open Question 26. Is there a good way to order update schemes by "randomness"? In a way that orders mixing time as well?

Open Question 27. Is there an example where all systematic updates take significantly longer than random updates to mix?

Open Question 28. Concretely, is there a set of generators $S = \{s_k\}$ for $\{0,1\}^d$ such that (lazy?) symmetric random walk on $\{0,1\}^d$ using these generators takes more steps to mix than $|S| \log |S|$, but the chain induced on $\{0,1\}^d$ by moving through the generators in order, and deciding whether to do the move by a coin flip, takes significantly longer?

Day IV

14. Speaker I: Dana Randall

Randall and Winkler have recently proved $\Theta(n^3 \log n)$ mixing for the one-dimensional version of the chain one n disks in the unit square described above.

Open Question 29. Does spatial discretization add a factor of $\log n$? Or not?

15. Speaker II: Tom Hayes

Open Question 30. There is a folk conjecture that mixing for the graph-coloring chain with at least as many colors as (max. degree plus 2) occurs in $O(n \log n)$. Is this true?

There has been much work done proving weaker results of this form — see work by Jerrum, Vigoda and Hayes.

16. Speaker III: Gèrard Ben Arous

Motivated by work on aging and random energy landscapes, Ben Arous raised the following question:

Take simple random walk on the hypercube in d dimensions, and let τ_A be the hitting time to some randomly chosen subset A of the hypercube. What is the joint distribution of (τ_A, X_{τ_A}), in the $d \to \infty$ limit? More specifically, if A is a uniformly chosen random subset of size M, then when is X_{τ_A} uniform, uniformly up to order $(NM)^{-1}$?

This last, specific question, was solved by Ben Morris immediately following the workshop. It is left as an exercise.

17. Speaker IV: Yuval Peres

Consider Glauber dynamics for the ferromagnetic Ising model on any finite graph, with free boundary conditions, a model with stationary measure

$$\mathbb{P}(\sigma) \propto e^{\beta \sum_{i\sim j} \sigma_i \sigma_j},$$

where the dynamics consist of forgetting about a spin at random, and resampling from the marginal distribution, conditioned on the others.

Conjecture 5. The relaxation time (1/spectral gap) is increasing in β.

Note: The condition of free boundary conditions is important to the conjecture. All bounds that have been shown on relaxation time are monotone in β... but there is no proof that the value is.

A stronger conjecture concerns the extended model:

$$\mathbb{P}(\sigma) \propto e^{J_{ij} \sum_{i\sim j} \sigma_i \sigma_j}.$$

Conjecture 6. Relaxation time is increasing in each J_{ij}.

This implies, amoung other things, that adding an edge to the graph will cause the chain to relax more slowly. The same should be true for any notion of mixing.

To review the state of the conjecture concerning cutoff in total variation:

Conjecture 7. For random walk on a family of transitive graphs, $T_{rel}/T_{mix} \to \infty$ implies existence of a cutoff in total variation.

Open Question 31. Does the above conjecture hold on general graphs, if we measure total variation distance with respect to a stationary start (as opposed to worst-case start)?

The (conjectured, but not carefully checked) counterexample to the conjecture in "worst-case" form is: Consider random walk on $[-n, +n] \subset \mathbb{N}$, where on $[-n, 0]$ the walk is biased towards zero at speed γ_-, and on $[0, +n]$ the walk is biased towards zero at speed γ_+. As $n \to \infty$, each half viewed on it's own has a sharp cutoff in total variation at $n\gamma_-$ and $n\gamma_+$, respectively. Now connect the node at $+n/2$, say, to the node at $-n$, so the walk begun at $+n$ and proceeding towards zero has roughly equal chance of proceeding down the positive integers, or jumping to $-n$. If $\gamma_- > \gamma_+$, then as $n \to \infty$, then $\|p^t(n, \cdot) - \pi\|$ will have two, distinct, sharp drops.

The following are suggested special cases to prove (or disprove!) the conjecture for:

Open Question 32. Does conjecture 7 hold for convolution on the cycle?

Open Question 33. Does conjecture 7 hold for abelian groups?

Open Question 34. Does conjecture 7 hold for transitive expanders? Here the hypotheses are automatically satisfied — need only check for cutoff.

Saloff-Coste showed that the width of the cutoff window for L^p ($p > 1$) mixing is $O(1/gap)$. Peres, feeling courageous, conjectures:

Conjecture 8. The cutoff window for total variation mixing has width $O(1/gap)$.

Open Question 35. Under the same conditions as conjecture 7, is there a cutoff for mixing measured by relative entropy? What about separation distance?