Convergence to Equilibrium in Local Interaction Games and Ising Models

Andrea Montanari[*] Amin Saberi[†]

November 17, 2008

Abstract

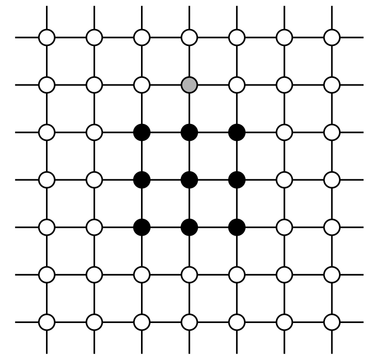
Coordination games describe social or economic interactions in which the adoption of a common strategy has a higher payoff. They are classically used to model the spread of conventions, behaviors, and technologies in societies. Here we consider a two-strategies coordination game played asynchronously between the nodes of a network. Agents behave according to a noisy best-response dynamics.

It is known that noise removes the degeneracy among equilibria: In the long run, the "risk-dominant" behavior spreads throughout the network. Here we consider the problem of computing the typical time scale for the spread of this behavior. In particular, we study its dependence on the network structure and derive a dichotomy between highly-connected, non-local graphs that show slow convergence, and poorly connected, low dimensional graphs that show fast convergence.

[*] Departments of Electrical Engineering and Statistics, Stanford University
[†]Department of Management Science and Engineering, Stanford University

1 Introduction

The unprecedented growth of online social network and their increasing role in the spread of knowledge, behaviors and new technologies have given rise to a wealth of interesting questions. Is it possible to explain the emergence of a new phenomenon based on the dynamics of the interaction among individuals [Klein07, Young93]?



As an example consider a two-dimensional grid and assume that each node adopts the new behavior (call it $+1$, the alternative being -1) if at least two of its neighbors have already adopted it. It is then easy to see that no finite set of $+1$'s can influence the whole grid, and in fact the influence of any finite set of $+1$'s is limited to the smallest rectangle that circumscribes them. For instance, the group of black nodes in the figure on the right does not expand further.

Now, consider the same dynamics with a small noise, i.e. assume that, with some small probability ϵ, agents do not follow the pre-established rule. This can have dramatic effects. If the gray node in the figure switches to $+1$ by mistake, then a new layer may be added to the group of black nodes at no extra (probability) cost. Of course, the reverse can happen: the block of $+1$'s can be eroded because of noise. However if the initial block is *large enough* (and under some technical assumptions) the former mechanism will prevail [NeS91, NeS92]. The important point is that 'large enough' means here larger than some *constant* quantity, and that influence spreads at some positive velocity. This phenomenon was first discovered in statistical physics, under the name of 'nucleation' and received an intense attention in the mathematical physics literature over the last 30 years [OV04, Bov03].

Similar models were developed independently within the context of evolutionary game theory. For example consider a simple game in which every individual placed in a network has to make a decision between two alternatives. The payoff of an action for each person is proportional to the number of its neighbors who are taking the same action. These games, known as coordination games, have been studied extensively for modeling the emergence of technologies and social norms [Young93, Morr00, Klein07, Blu93]. The main conclusion of this line of work is that adding a small random perturbation to best response dynamics creates an evolutionary force that drives the system towards a particular equilibrium in which all players take the same action.

In real-world networks stochasticity is unavoidable. As a consequence, we can expect the players to eventually achieve coordination on a particular equilibrium, irrespective of the initial state. The present paper characterizes the *rate of convergence* for such dynamics in terms of explicit graph quantities. It thus provide the first step in a longer term program aimed at developing approximation algorithms to estimate convergence to Nash equilibria.

Our characterization is expressed in terms of tilted cutwidth and tilted cut of the graph that are dual quantities. The former provides a path to the $+1$ equilibrium that gives an upper bound on the converge time. The latter corresponds to a bottleneck along the highest separating set in the space of configurations. We show that tilted cut and tilted cutwidth coincide for the 'slowest' subgraph and the convergence time is exponential in this graph parameter.

The proof uses an argument similar to [DV76, DSC93, JS89] to relate hitting time to the spectrum of an appropriate transition kernel. The convergence time is then estimated in terms of the most likely path from the worst-case initial configuration. It turns out that the most likely path is the one that implies the lowest decrease of probability in stationary measure. A delicate argument using the submodularity of the potential function shows that there exists a monotone increasing path with this property. In order to prove the characterization in terms of tilted cut we study the 'slowest' eigenvector and show that it is monotone using a fixed point argument. We then approximate the eigenvector with a characteristic function.

The above result allows us to estimate the convergence time for specific graphs through their isoperimetric function. For example in interaction graphs that can be embedded in low dimensional spaces, the dynamics converges in a very short time. On the other hand, for a wide class of bounded degree graphs such as random regular graphs or certain small-world networks the convergence may take as long as exponential in the number of nodes.

Related work

There is a very interesting line of work in mathematical physics leading to very sharp estimates of the convergence times of specific models: mainly two and three dimensional grids [BC96, BM02]. Berger et al. [BK+05] compute the mixing time of a similar dynamics in terms of cutwidth of the graph using different techniques from the current paper.

In the game theory literature, one of the criticisms of Nash equilibria is that its multiplicity makes it hard to predict the outcome of a play. How do players *learn* to play a specific equilibrium, and which one do they *select*? For example, the grid graph described above shows that the coordination game can have several equilibria. There is a vast literature in evolutionary game theory for resolving this problem especially in the context of coordination games [KMR93, Young93, Ell93, Blu93, FL98].

The importance of estimating convergence times was first stressed in the pioneering work of Ellison [Ell93]. He argued that the long-run equilibrium is relevant only if the convergence time is reasonably small. Ellison studied the rate of convergence for two extreme interaction graphs: a complete graph and a graph obtained by placing individuals on a cycle and connecting all pairs of distance smaller than some given range. He showed that the dynamics converges very slowly for the former model and very quickly for the latter. Based on this observation, he concluded that when the interaction is global the outcome is determined by historic factors. In contrast, when players "interact with small sets of neighbors," we can assume that evolutionary forces may determine the outcome.

Our result implies that the key property of the network that captures the rate of convergence is not the number of nodes each agent interacts with, or the number of edges of the graph. This can be proved for a large class of (non-reversible) noisy best-response dynamics including the one of [Ell93].

2 Definitions

A game is played in periods $t = 1, 2, 3, \dots$ among a set V of players. Each player $i \in V$ has two alternative strategies as $x_i \in \{+1, -1\}$. Let $\underline{x} = \{x_i : i \in V\}$. The payoff matrix A is a 2×2- matrix illustrated in the figure. The players interact on an undirected graph $G = (V, E)$. The payoff of player i is $\sum_{j \in \partial i} A(x_i, x_j)$, where ∂i is the set of neighbors of vertex i.

a, a	c, d
d, c	b, b

The payoff matrix A defines a coordination game which means $a > d$ and $b > c$. It is easy to verify that for every i, the best response strategy is $\text{sign}(h_i + \sum_{j \in \partial i} x_j)$, where $h_i = \frac{a-d-b+c}{a-d+b-c} |\partial i| \equiv \rho\, |\partial i|$, with $|\partial i|$ the degree of node i. We assume that $a - b > d - c$, so that $h_i > 0$ for all $i \in V$ of non-vanishing degree. Harsanyi and Selten [HS88] named + the "risk-dominant" equilibrium, as it minimizes the utility loss due to a change in the opponent strategy. Notice that this does not coincide, in general, with the payoff dominant equilibrium.

Noisy best-response dynamics is specified by a one-parameter family of Markov chains $\mathbb{P}_\beta\{\cdots\}$ indexed by β. The parameter $\beta \in \mathbb{R}_+$ determines how noisy is the dynamics, with $\beta = +\infty$ corresponding to the noise-free case. Two type of updates are naturally defined:

(1) Synchronous updates. At each step of the chain, each player draws a new strategy y_i conditionally on its neighbor's strategies $x_{\partial i}$ at the previous time step. The conditional distribution is denoted by $p_{i,\beta}(y_i|\underline{x}_{\partial i})$.

(2) Asynchronous updates. Each node i updates its value at the arrival time of an independent Poisson clock of rate 1. The conditional distribution of the new strategy is again denoted as $p_{i,\beta}(y_i|\underline{x}_{\partial i})$.

The dynamics of [Ell93] is recovered by the following transition probabilities. Let $y_i^* = \text{sign}(h_i + \sum_{j\in\partial i} x_j)$. Then for every player i, $p_{i,\beta}(y_i^*|\underline{x}_{\partial i}) = 1 - e^{-\beta}$ and $p_{i,\beta}(-y_i^*|\underline{x}_{\partial i}) = e^{-\beta}$.

A considerable simplification is achieved for the so-called *heath bath* or *Glauber* kernel

$$p_{i,\beta}(y_i|\underline{x}_{\partial i}) = \left(1 + e^{-2\beta K_i(\underline{x}) y_i}\right)^{-1} \tag{1}$$

where $K_i(\underline{x}) = h_i + \sum_{j\in\partial i} x_j$. . This is also known as logit update rule which is standard in the discrete choice literature [M74]. It has also been used to model subjects' empirical choice behavior in laboratory situations [MS94, MP95]. In this context it has been studied by Blume [Blu93]. The corresponding Markov chain is reversible with respect to the stationary distribution $\mu_\beta(\underline{x}) \propto \exp(-\beta H(\underline{x}))$, with

$$H(\underline{x}) \;\; = \;\; -\sum_{(i,j)\in E} x_i x_j - \sum_{i\in V} h_i x_i\,, \tag{2}$$

in the case of asynchronous dynamics. This is the energy function of the Ising model; an analogous expression can be written for synchronous updates. In both the above models the stationary distribution for large β concentrates around the all-(+1) configuration. In other words, these dynamics predict that in the long run, the play will converge to the risk-dominant equilibrium.

In the following we will often adopt the equivalent representation of configurations as subsets of vertices $S \subseteq V$, whereby $i \in S$ if and only if $x_i = +1$, and, with a slight abuse of notation, we shall denote by $H(S)$ the corresponding energy. If $|S|_h \equiv \sum_{i\in S} h_i$, then $H(S) - H(\emptyset) = 2\,\text{cut}(S, V\setminus S) - 2|S|_h$. It is important to notice that $H(\,\cdot\,)$ is submodular.

Our aim is to determine whether this prediction is realized in a reasonable time. To this end, we let T_+ denote the hitting time to the all-(+1) configuration, and define the *typical hitting time* for $\underline{+1}$ as

$$\tau_+(G;\underline{h}) = \sup_{\underline{x}} \inf\left\{t \geq 0:\, \mathbb{P}_\beta^{\underline{x}}\{T_+ \geq t\} \leq e^{-1}\right\}\,. \tag{3}$$

For the sake of brevity, we will often refer to this as the hitting time, and drop its arguments.

3 Main results

Our first step is to express the large-β (low-noise) behavior of $\tau_+(G;\underline{h})$ in terms of graph-theoretical quantities. Let $n = |V|$ be the number of players. Given $\underline{h} = \{h_i : i \in V\}$, and $U \subseteq V$, we let $|U|_h \equiv \sum_{i\in U} h_i$. We define the *tilted cutwidth of* G as

$$\Gamma(G;\underline{h}) \equiv \min_{S:\emptyset\to V} \max_{t\leq n} \left[\text{cut}(S_t, V\setminus S_t) - |S_t|_h\right]\,. \tag{4}$$

Here the min is taken over all *linear orderings* of the vertices $i(1),\dots,i(n)$, with $S_t \equiv \{i(1),\dots,i(t)\}$. Note that if for all i, $h_i = 0$, the above is equal to the cutwidth of the graph.

Given a collection of subsets of V, $\Omega \subseteq 2^V$ such that $\emptyset \in \Omega$, $V \notin \Omega$, we let $\partial\Omega$ be the collection of couples $(S, S\cup\{i\})$ such that $S \in \Omega$ and $S\cup\{i\} \notin \Omega$. We then define the *tilted cut* of G as

$$\Delta(G;\underline{h}) \equiv \max_{\Omega} \min_{(S_1,S_2)\in\partial\Omega} \max_{i=1,2} \left[\text{cut}(S_i, V\setminus S_i) - |S_i|_h\right]\,, \tag{5}$$

the maximum being taken over *monotone* sets Ω (i.e. such that $S \in \Omega$ implies $S' \in \Omega$ for all $S' \subseteq S$). It thus coincide

It is known that, in the case $h_i = 0$, the mixing time of Glauber dynamics is at most exponential in the cutwidth of G [BK+05]. The following result provides a generalization to the case $h_i > 0$ of interest here, in the limit of large β. Since $\Gamma(G; \underline{h})$ (as well as $\Delta(G; \underline{h})$) is decreasing in $\underline{h}$, the upper bound is smaller than the one for the $h_i = 0$ case.

Theorem 3.1. *Given an induced subgraph $F \subseteq G$, let $\underline{h}^F$ be defined by $h_i^F = h_i + |\partial i|_{G \setminus F}$, where $|\partial i|_{G \setminus F}$ is the degree of i in $G \setminus F$. For reversible asynchronous dynamics we have $\tau_+(G; \underline{h}) = \exp\{2\beta \Gamma_*(G; \underline{h}) + o(\beta)\}$, where*

$$\Gamma_*(G; \underline{h}) = \max_{F \subseteq G} \Gamma(F; \underline{h}^F) = \max_{F \subseteq G} \Delta(F; \underline{h}^F) \, . \tag{6}$$

Note that tilted cutwidth and tilted cut are dual quantities. The former corresponds the maximal energy height along the lowest path to the + equilibrium. The latter is the lowest energy along the highest separating set in the space of configurations. A natural strategy for estimating $\Gamma_*(G; \underline{h})$ consists in lower bounding $\Delta(F; \underline{h}^F)$ by exhibiting a monotone set $\Omega \subseteq 2^{V(F)}$, and upper bounding $\Gamma(F; \underline{h}^F)$ by exhibiting a linear ordering of $V(F)$. The above theorem shows that tilted cut and cutwidth coincide for the 'slowest' subgraph of G and if the h_i's are non-negative. The hitting time is exponential in this graph parameter.

The two characterizations above are exact but it is highly non-trivial to compute them. In the rest of this section, we will show how the above theorem implies the known results for special classes of graphs. Then, we relate tilted cutwidth to graph expansion and derive a dichotomy between the hitting time on expanders versus locally connected graphs. In the end, we show how to use algorithms for sparsest cuts to find the approximately optimal linear ordering as defined in tilted cutwidth.

The cases treated by Ellison are easily understood within the present framework. In order to derive a lower bound for the complete graph, with $h_i = h$ for all $i \in V$, one can restrict attention to $F = G$ and for that graph define Ω to be the family of all sets with cardinality at most $n/2$.

$$\Gamma^*(K_n; \underline{h}) \geq \min_{|S| = n/2} [\mathrm{cut}(S, V \setminus S) - |S|_h] = (n-h)^2/4 + O(n) \, . \tag{7}$$

The second example studied by Ellison is a $2k$-regular graph resulting from connecting all vertices of distance at most k in a cycle. In that graph, the maximum is again achieved for $F = G$, and the natural linear ordering of the cycle yields $\Gamma(G; h) \leq 4k^2$.

It is also straightforward to recover the result of Young [Young95] from the above theorem. Indeed, the hypotheses of [Young95] are equivalent to the existence of a sequence $S_1, \dots, S_T \subseteq V$ such that $H(S_t) = \min_{S' \subseteq S_t} H(S') \leq 0$ and $|S_i| \leq k$. By flipping vertices along this sequence and using the submodularity of $H(\,\cdot\,)$, it follows that $\Gamma(F; \underline{h}^F) \leq k^2$.

3.1 Relation to graph expansion

The following Lemma links the isoperimetric function of G (and its subgraphs) to the hitting time. It is particularly useful when analyzing specific graph families.

Lemma 3.2. *For $\theta \in \mathbb{R}$ define $J(\theta) = [\theta - h_{\max}, \theta + h_{\max}]$. Assume that there exist constants α and $\gamma < 1$ such that for any subset of vertices $U \subseteq V$, and any θ such that there exists $S \subseteq U$ with $|S|_h \in J(\theta)$, we have*

$$\mathrm{cut}(S, U \setminus S) \leq \alpha \, |S|^\gamma \, , \tag{8}$$

for at least one such S. Then $\Gamma_*(G;\underline{h}) \le A(\alpha,\gamma,h_{\max})\, h_{\min}^{-1/(1-\gamma)} \log\max(2, h_{\min}^{-1})$.

Conversely, assume there exists $U \subseteq V(G)$, such that for $i \in U$, $|\partial i \cap (V \setminus U)| \le b$, and the subgraph induced by U is a (δ,λ) expander. Then $\Gamma_*(G;\underline{h}) \ge (\lambda - h_{\max} - b)\lfloor \delta |U| \rfloor$.

In words, the hitting time is dominated by highly connected subgraphs of G, that are loosely tied to the rest of the graph. On the other hand, an upper bound on the isoperimetric function leads to upper bounds on the hitting time.

In order to gain some intuition we consider a few interesting graph models:

(a) *Finite-range d-dimensional networks.* The graph G is a d-dimensional range-K network if we can associate to each of its vertices $i \in V$ a position $x_i \in \mathbb{R}^d$ such that, (1) whenever $(i,j) \in E$, $d_{\rm Eucl}(x_i, x_j) \le K$ (here $d_{\rm Eucl}(\cdots)$ denotes Euclidean distance); (2) Any cube of volume v contains at most $2\,v$ vertices. We will also say that G is *embeddable* in this case.

(b) *Small world networks.* Again, the vertices are those of a d-dimensional grid of side $n^{1/d}$. Two vertices i, j are connected by an edge if they are nearest neighbors. Further, each vertex i is connected to k other vertices $j(1), \dots, j(k)$ drawn independently with distribution $P_i(j) = C(n)|i-j|^{-r}$.

(c) *Random regular graphs of degree k.*

Theorem 3.3. *The following statements hold with high probability:*

If G is a d-dimensional finite-range graph, and $h_{\min} > 0$, then $\Gamma_*(G;\underline{h}) = O(1)$.

If G is a small world network with $r \ge d$, and $h_{\max} \le k-d-5/2$, then $\Gamma_*(G;\underline{h}) = \Omega(\log n/\log\log n)$.

If G is a small world network with $r < d$, and $h_{\max}$ is small enough, then $\Gamma_*(G;\underline{h}) = \Omega(n)$.

If G is a random k-regular graph, and $h_{\max} < k-2$, then $\Gamma_*(G;\underline{h}) = \Omega(n)$.

These qualitatively distinct behaviors correspond to different mechanisms by which consensus spreads in these networks. In finite-range networks, the process is initiated in a relatively compact region taking value $+1$. If this is large enough (which happens with positive probability), it spreads through the whole graph. This is possible because of the bias provided by $h_{\min} > 0$. Indeed the proof of this statement implies an upper bound of the form $\Gamma(G;\underline{h}) = O(h_{\min}^{-(d-1)} \log(1/h_{\min}))$.

In small-world networks with $r \ge d$ the process is similar, but the spread of $+1$'s is blocked in its very last stages by small, highly connected regions of size roughly $(\log n)$. Finally, small-world networks with $r < d$ and random regular graphs are expanders and convergence is extremely slow.

All the above statements take the form of a tradeoff between how 'well-connected' is G and how biased is the dynamics (the latter being measured by $h_{\min}$). In the case of well-connected graphs it is not hard to prove upper bounds on $\Gamma_*(G;\underline{h})$ for large enough $\underline{h}$. For instance, in the case of k-regular graphs $\Gamma_*(G;\underline{h}) = O(1)$ if $h_{\min} \ge k$.

3.2 Approximating tilted cut and tilted cutwidth

The maximization over Ω in Eq. (5) for computing tilted cut is highly non-trivial. Here we obtain a class of lower bounds by restricting Ω to essentially subsets with a given cardinality. The following result shows the 'loss' resulting from this restriction is bounded, under appropriate conditions. On the other hand, it implies that algorithms for computing sparse cuts find approximately optimal orderings corresponding to a tilted cutwidth.

Theorem 3.4. *Assume that, for some L_1, L_2, with $L_2 \ge h_{\max}$ and for every induced subgraph $F \subseteq G$, we have*

$$\min_{|S|_h \in [L_1, L_2]} \left[\mathrm{cut}(S, V(F)\setminus S) - |S|_{h^F}\right] \le L_1\,, \tag{9}$$

where it is understood that $\emptyset \neq S \subseteq V(F)$. If, for every subset of vertices U, with $|U|_h \leq L_2$, the induced subgraph has cutwidth upper bounded by C, then $\Gamma(G; 4\underline{h}) \leq C + L_1 + L_2$.

It is interesting to compare this result with the analysis of contagion models [Morr00]. In that case contagion takes place if there exists an ordering of the vertices $i(1)$, $i(2)$, ... such that, assuming $x_{i(1)} = +1$, $x_{i(2)} = +1$,... $x_{i(t)} = +1$, the best response for $i(t+1)$ is strategy $+1$. Theorem 3.4 allows to replace single vertices, by 'blocks' as long as they have bounded size and bounded cutwidth.

Assuming that a 'good' path to consensus exists, can it be found efficiently? By using a simple generalization of Feige and Krauthgamer's [FK02] $O(\log^2 n)$ approximation algorithm for finding the sparsest cut of a given cardinality, we have the following

Remark 3.5. *If $G = (V, E)$ satisfies equation (9), it is possible to find an ordering $i_1, i_2, \dots, i_n$ of V in polynomial time so that for every $S_t = \{i_1, i_2, \dots i_t\}$, and $L = L_1 + L_2 + C$*

$$\text{cut}(S_t, V \setminus S_t) = O(|S_t|_h \log^2 n + L \log n).$$

3.3 Nonreversible and synchronous dynamics

In this section we consider a general class of Markov dynamics over $\underline{x} \in \{+1, -1\}^V$. An element in this class is specified by $p_{i,\beta}(y_i|\underline{x}_{\partial i})$, with $p_{i,\beta}(+1|\underline{x}_{\partial i})$ a non-decreasing function of the number $\sum_{j \in \partial i} x_j$. Further we assume that $p_i(+1|\underline{x}_{\partial i}) \leq e^{-2\beta}$ when $h_i + \sum_{j \in \partial i} x_j < 0$. Note that the synchronous Markov chain studied in KMR [KMR93] and Ellison [Ell93] is a special case in this class.

Denote the hitting time of all (+1)-configuration in graph G with $\tau_+(G)$ as before.

Proposition 3.6. *Let $G(V, E)$ be a k-regular graph of size n such that for $\lambda, \delta > 0$, every $S \subset V, |S| \leq \delta n$ has vertex expansion at least λ. Then for any noisy-best response dynamics defined above, there exists a constant $c = c(\lambda, \delta, k)$ such that $\tau_+(G; \underline{h}) \geq \exp\{\beta c n\}$ as long as*

$$\lambda > \frac{3k}{4} + \frac{\max_i h_i}{2}.$$

Note that random regular graphs satisfy the condition of the above proposition as long as h_i's are small enough. The proof of the proposition is by simply considering the evolution of one dimensional chain indicating the number of +1 vertices.

Proposition 3.7. *Let G be a d-dimensional grid of size n and constant $d \geq 1$. For any synchronous or asynchronous noisy-best response dynamics defined above, there exists constant c such that $\tau_+(G; \underline{h}) \leq \exp\{\beta c\}$.*

The above proposition can be proved by a simple coupling argument very similar to that of Young [Young93]. We will leave its details to a more complete version of the paper. The above two propositions show that for a large class of noisy best-response dynamics including the one considered in [Ell93], the degrees of vertices are not the key property dictating the rate of convergence.

4 Proofs

4.1 Theorem 3.1

It is a basic result in the theory of reversible Markov chains with exponentially small transition rates, that hitting time are related to 'energy barriers.'

Lemma 4.1. *Consider a Markov chain with state space $\mathcal{S}$ reversible with respect to the stationary measure $\mu_\beta(x) = \exp(-\beta H(x) + o(\beta))$, and assume that, if $p_\beta(x, y) = \exp(-\beta V(x, y) + o(\beta))$.*

Let $A = \{x : H(x) \leq H_0\}$ be non-empty, and define the typical hitting time for A as in Eq. (3), with $+$ replaced by A. Then $\tau_A = \exp\{\beta\widetilde{\Gamma}_A + o(\beta)\}$ where

$$\widetilde{\Gamma}_A = \max_{z \notin A} \min_{\omega: z \to A} \max_{t \leq |\omega|-1} [H(\omega_t) + V(\omega_t, \omega_{t+1}) - H(z)] \ , \tag{10}$$

and the min *runs over paths $\omega = (\omega_1, \omega_2, \dots, \omega_T)$ in configuration space such that $p_\beta(\omega_t, \omega_{t+1}) > 0$ for each t.*

The proof can be obtained by building on known results, for instance Theorem 6.38 in [OV04]. These however typically apply to exit times from local minima of $H(x)$. We provide a simple proof based on spectral arguments in Appendix B.

For the sake of clarity, we split the proof of Theorem 3.1 in two parts: first the characterization in terms of tilted cutwidth (i.e. the first identity in Eq. (6)); then the one in terms of tilted cut (second identity in Eq. (6)).

Proof. (Theorem 3.1, Tilted cutwidth). Notice that Glauber dynamics satisfies the hypotheses of Lemma 4.1, with $H(\underline{x}) = H(\underline{x})$ given by Eq. (2). In this case, for any allowed transition $\underline{x} \to \underline{y}'$, $H(\underline{x}) + V(\underline{x}, \underline{y}) = \max(H(\underline{x}), H(\underline{y}))$. As a consequence, we can drop the factor $V(\cdots)$ in Eq. (10). We thus obtain $\tau_+ = \exp(\beta \max_{\underline{z}} \widetilde{\Gamma}_+(\underline{z}) + o(\beta))$ where

$$\widetilde{\Gamma}_+(\underline{z}) = \min_{\omega: \underline{z} \to \underline{+1}} \max_{t \leq |\omega|-1} [H(\omega_t) - H(\underline{z})] \ . \tag{11}$$

An upper bound is obtained by restricting the minimum to monotone paths. It is not hard to realize that the result coincides with $2\Gamma(F; \underline{h}^F)$ where F is the subgraph induced by vertices i such that $z_i = -1$. It is far less obvious that the optimal path can indeed be taken to be monotone.

It is convenient to use the representation of the path $\omega = (\underline{x}_0 = \underline{z}, \underline{x}_1, \dots, \underline{x}_{|\omega|-1} = \underline{+1})$ as a sequence of subsets of vertices: $\omega = (S_0 = S, S_1, \dots, S_{|\omega|-1} = V)$. We will consider a more general class of paths whereby $S_t \setminus S_{t-1} = \{v\}$ or $S_t \subset S_{t-1}$, and let $G(\omega) = \max_t [H(S_t) - H(S_0)]$.

Let us start by considering the optimal initial configuration We claim that if $B \in \arg\max_S \min_{\omega: S \to V} G(\omega)$ is such an optimal configuration, then for every $A \subset B$, $H(A) \geq H(B)$. Indeed, suppose $H(A) < H(B)$. By prepending B to any path $\omega : A \to V$, we obtain a path $\omega' : B \to V$ with $G(\omega') < G(\omega)$. Therefore $\min_{\omega': B \to V} G(\omega') < \min_{\omega: A \to V} G(\omega)$ which is a contradiction.

Among all paths that achieve the optimum, choose the path ω that minimizes the potential function $f(\omega) = |\omega|^2 |V| - \sum_{S_i \in \omega} |S_i|$. Intuitively, f puts a very high weight on shorter paths and then paths with larger sets. We will prove that, with this choice, ω is monotone.

For the sake of contradiction, suppose ω is not monotone. Let S_k be the set with the smallest index such that $S_{k+1} \subset S_k$. Partition $S_k \setminus S_{k+1}$ into two subsets $R = (S_k \setminus S_{k+1}) \cap S_0$ and $T = (S_k \setminus S_{k+1}) \setminus S_0$. Without loss of generality assume that for $1 \leq i \leq k$, $S_i = \{1, 2, \cdots i\} \cup S_0$. Let $v_1 \leq v_2 \cdots \leq v_t$ be the elements of T in the order of their appearance in ω.

For a subset $A \subset T$, and $i \leq k$ define the marginal value of subset A at position i to be $M(A, i) = H(S_i \setminus A) - H(S_i)$. Since H is submodular, $M(A, i)$ is non-decreasing with i as long as $A \subset S_i$. Because of our claim about the initial condition, we have, in particular,

$$M(R, 0) = H(S_0) - H(S_0 \setminus R) \geq 0 \,. \tag{12}$$

The crucial lemma below is proved in Appendix C.

Lemma 4.2. *One of the following two statements is correct: Case (I) There exists a subset $T' \subset T$ such that for all i, $M(T', i) \le 0$; Case (II) $M(T \cup R, k) \ge 0$.*

We are now ready to finish the proof. Suppose the first statement of the lemma is correct. We construct a new path ω' by removing the vertices of T' from the sequence $1, 2, \cdots, t$ in the beginning of ω and also removing T' from T. Since ω' is shorter than ω, we only need to argue that $G(\omega') \le G(\omega)$. This is obvious because for every $i \le k$, $H(S_i \setminus T') - H(S_i) = M(T', i) \le 0$.

In the second case, we construct another path by changing S_{k+1}. First note that since ω is minimizing the potential function, $S_{k+2} = S_{k+1} \cup \{v\}$ for some v that is not in S_k. Now note that by replacing S_{k+1} with $S_k \cup \{v\}$ we obtain a path with a higher value of the potential function and at most the same barrier. This is because

$$H(S_{k+1} \cup \{v\}) - H(S_k \cup \{v\}) \ge H(S_{k+1}) - H(S_k) = M(T \cup R, k) \ge 0\,. \tag{13}$$

□

The second part of the proof exploits the well known fact that Glauber dynamics is monotone for the Ising model. Given initial conditions $\underline{x}(0)$ and $\underline{x}'(0) \succeq \underline{x}(0)$, the corresponding evolutions can be coupled in such a way that $\underline{x}'(t) \succeq \underline{x}(t)$ after any number of steps.

Proof. (Theorem 3.1, Tilted cut). By monotonicity of Glauber dynamics $\Gamma_*(G; \underline{h}) \ge \Gamma_*(F; \underline{h}^F)$ for any induced subgraph $F \subseteq G$. Theorem 4.1 implies $\Gamma_*(F; \underline{h}^F) \ge \Delta(F; \underline{h}^F)$: indeed given a path $\omega = (S_0, S_1, \dots, S_{|\omega|-1} = V)$ this must have at least one step in $\partial\Omega$. Hence $\Gamma_*(G; \underline{h}) \ge \max_F \Delta(F; \underline{h}^F)$.

We need to prove $\Gamma_*(G; \underline{h}) \le \Delta(F; \underline{h}^F)$ for at least one induced subgraph F. Fix F to be a subgraph which achieves the maximum in Eq. (6) (i.e. $\arg\max \Gamma(F; \underline{h}^F)$). Notice that, to leading exponential order, the hitting time in F is the same as in G, i.e. $\Gamma_*(F; \underline{h}^F) = \Gamma_*(G; \underline{h})$.

Let $p_\beta(\underline{x}, \underline{y})$ be the transition probabilities of Glauber dynamics on F, and $p^+_\beta(\underline{x}, \underline{y})$ the kernel restricted to $\{+1, -1\}^{V(F)} \setminus \{\underline{+1}\}$. By this we mean that we set $p^+_\beta(\underline{x}, \underline{+1}) = p^+_\beta(\underline{+1}, \underline{y}) = 0$. Denote by P^+_β the matrix with entries $p^+_\beta(x, y)$ and by ψ_0 its eigenvector with largest eigenvalue. By Perron-Frobenius Theorem, we can assume $\psi_0(\underline{x}) \ge 0$. We claim that $\psi_0(\underline{x})$ is monotonically decreasing in $\underline{x}$. Indeed consider the transformation $\psi \mapsto T(\psi) \equiv P^+_\beta \psi / ||P^+_\beta \psi||_{2,\mu}$. This is a continuous mapping from the set of unit vectors in $L^2(\mu)$ onto itself. Further, if ψ is monotone and non-negative, $T(\psi)$ is monotone an non-negative as well (the first property follows from monotonicity of the dynamics). The set of non-negative and monotone unit vectors in $L^2(\mu)$ is homeomorphic to a simplex. By Brouwer fixed point theorem, T has at least one fixed point that is non-negative and monotone, which therefore coincides with ψ_0 by Perron-Frobenius.

Lemmas B.1 and E.1 imply that there exists $\Omega = \{x \in \mathcal{S} :\ \psi_0(\underline{x}) > b\}$, such that

$$\tau_+(F; \underline{h}^F) \le C_n(1+\beta)\, \frac{\sum_{\underline{x} \in \Omega} \mu(\underline{x})}{\sum_{(\underline{x},\underline{y}) \in \partial\Omega} \mu(\underline{x}) p^+_\beta(\underline{x}, \underline{y})}\,. \tag{14}$$

for some β-independent constant C_n. Using $\tau_+(F; \underline{h}^F) = \exp\{2\beta\Gamma_*(F; \underline{h}^F) + o(\beta)\}$ and the large β asymptotics of $\mu(\underline{x})$, $p^+_\beta(\underline{x}, \underline{y})$ we get

$$\Gamma_*(F; \underline{h}^F) \le \min_{(S_1,S_2) \in \partial\Omega} \max_{i=1,2} [\mathrm{cut}(S_i, V \setminus S_i) - |S_i|_h] + o_\beta(1)\,. \tag{15}$$

Since $\psi_0(\underline{x})$ is monotone, Ω is monotone as well and therefore the last inequality implies the thesis.
□

4.2 Theorem 3.3

Proof. (Lemma 3.2). By Theorem 3.1, it is sufficient to find an upper bound for $\Gamma(\tilde{F};\underline{h}^{\tilde{F}})$ for every induced subgraph $\tilde{F}$. By monotonicity of $\Gamma(\tilde{F};\underline{h})$ with respect to $\underline{h}$, $\Gamma(\tilde{F};\underline{h}^{\tilde{F}}) \leq \Gamma(\tilde{F};\underline{h})$. We will upper bound $\Gamma(\tilde{F};\underline{h})$ by showing Eq. (9) holds for any induced subgraph $F \subseteq \tilde{F}$.

First notice that, for any U and for any θ, there exists $S \subseteq U$ such that $|S|_h \in J(\theta)$ and

$$\mathrm{cut}(S, U \setminus S) - \frac{1}{4}|S|_h \leq \alpha h_{\min}^{-\gamma}|S|_h^{\gamma} - \frac{1}{4}|S|_h \leq A'(\alpha,\gamma)\, h_{\min}^{-\gamma/(1-\gamma)}\,, \tag{16}$$

where $A'(\alpha,\gamma) = \max(\alpha x^{\gamma} - x/4 :\, x \geq 0)$. Take $L_1 = A'(\alpha,\gamma)\, h_{\min}^{-\gamma/(1-\gamma)}$ and $L_2 = L_1 + 2h_{\max}$. By Eq. (16)

$$\min_{|S|_h \in [L_1, L_2]} \left[\mathrm{cut}(S, V(F)\setminus S) - \frac{1}{4}|S|_h\right] \leq L_1\,.$$

Finally the cutwidth of any set S with $|S|_h \leq L_2$ is upper bounded by $\alpha|S|^{\gamma}\log|S|$ (using [LR99] and Eq. (8)) which is at most $C = A''(\alpha,\gamma,h_{\max})\, h_{\min}^{-1/(1-\gamma)} \log\max(2, h_{\min}^{-1})$. The thesis thus follows by applying Theorem 3.4.

To prove the lower bound we use Theorem 3.1 again. Let F be the subgraph induced by U. By monotonicity of $\Delta(G;\underline{h})$ with respect to $\underline{h}$, for $t = \lfloor \delta|U|\rfloor$, we have

$$\Delta(F;\underline{h}^F) \geq \Delta(F; h_{\max} + k) \geq \min_{|S|=t}\left[\lambda|S| - (h_{\max}+k)|S|\right]\,.$$

which implies the thesis. □

We notice in passing that the estimates in the second part of this proof could be improved by using more specific arguments instead of directly applying Theorem 3.1.

For the proof of theorem 3.3, we need to estimate the isoperimetric function of finite range d-dimensional graphs. This can be done by an appropriate relaxation.

Given a function $f : V \to \mathbb{R}$, $i \mapsto f_i$, and a set of non-negative weights w_i, $i \in V$, we define

$$||f||_w^2 \equiv \sum_{i\in V} w_i\, f_i^2\,, \qquad ||\nabla_G f||^2 \equiv \sum_{(i,j)\in E} |f_i - f_j|^2\,. \tag{17}$$

We then have the following generalization of Cheeger inequality.

Lemma 4.3. *assume there exists two vertex sets $\Omega_1 \subseteq \Omega_0 \subseteq V$ and a function $f : V \to \mathbb{R}$ such that: (1) $f_i \geq |f_j|$ for any $i \in \Omega_1$ and any $j \in V$; (2) $f_i = 0$ for $i \in V \setminus \Omega_0$; (3) $L_1 \leq |\Omega_1|_w \leq |\Omega_0|_w \leq L_2$; (4) $||\nabla_G f||^2 \leq \lambda\, ||f||_h^2$. Then there exists $S \subseteq V$ with $L_1 \leq |S|_w \leq L_2$*

$$\mathrm{cut}(S, V\setminus S) \leq \sqrt{4\lambda \max_{i\in V}\{|\partial i|/h_i\}}\;\; |S|_h\,. \tag{18}$$

The proof of this Lemma is deferred to Appendix A.

Proof. (Theorem 3.3) *Finite-range d dimensional networks.* We need to prove that, for each induced subgraph G', $\Gamma(G';\underline{h}^{G'}) = O(1)$. By Theorem 3.4, it is sufficient to show that, for any induced and connected subgraph F, there exists a set S of bounded size such that $\mathrm{cut}(S, V(F)\setminus S) - \frac{1}{4}|S|_{(h)^F} \leq 0$, with $h_i' = h_i/4$. If the original graph is embeddable, any induced subgraph is embeddable as well. Since $h_i^F \geq h_i$, the thesis follows by proving that for any embeddable graph G, we can find a set of vertices S of bounded size with $\mathrm{cut}(S, V\setminus S) \leq |S|_{h/4}$.

We will construct a function f with bounded support such that $||\nabla_G f||^2 \le \lambda ||f||^2$ with $\lambda = \min_{i\in V}\{\frac{h_i}{16|\partial i|}\}$. In order to achieve this goal, consider the d-dimensional of G and partition $\mathbb{R}^d$ in cubes $\mathcal{C}$ of side ℓ to be fixed later. Denote by $\mathcal{C}_0$ the cube maximizing $\sum_{i:x_i\in\mathcal{C}} h_i$, and let C_j, $j = 1,\dots 3^d - 1$ be the adjacent cubes. Let $f_i = \varphi(x_i)$, where for $x \in \mathbb{R}^d$, we have

$$\varphi(x) = \left[1 - \frac{d_{\rm Eucl}(x,\mathcal{C})}{\ell}\right]_+ . \tag{19}$$

Notice that $|\nabla\varphi(x)| \le 1/\ell$ and $|\nabla\varphi(x)| > 0$ only if $x \in \mathcal{C}_j$, $j = 1,\dots 3^{d-1}$. Since $|f_i - f_j| \le |\nabla\varphi|\ \ ||x_i - x_j||$ we have

$$\begin{aligned} ||\nabla_G f||^2 &\le \left(\frac{K}{\ell}\right)^2 \sum_{i\in V} |\partial i|\ \mathbb{I}(x_i \in \cup_{j=1}^{3^d-1}\mathcal{C}_j) \le \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|\partial i|/h_i\} \sum_{i\in V} h_i\ \mathbb{I}\Big(x_i \in \cup_{j=1}^{3^d-1}\mathcal{C}_j\Big) \\ &\le 3^d\left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|\partial i|/h_i\} \sum_{i\in V} h_i\ \mathbb{I}\Big(x_i\in\mathcal{C}_0\Big) \le 3^d\left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|\partial i|/h_i\} ||f||_h^2 . \end{aligned} \tag{20}$$

The thesis follows by choosing $\ell = 2^{d+2}K\max_{i\in V}\{|\partial i|/h_i\}$.

Small world networks with $r \ge d$. Let U be a subset of vertices forming a cube of side ℓ, and G_U a $(\varepsilon, k-5/2)$, k-regular expander with vertex set U. Such a graph exists for all ℓ large enough and ε small enough by [Kah92]. Call A_U the event that the subgraph induced by long-range edges in U coincides with G_U, and no long-range edge from $i \in V\setminus U$ is incident on U.

Under A_U, the subgraph G_U satisfies the hypotheses of Lemma 3.2, second part, with $b = d$. Therefore $\Gamma_*(G;\underline{h}) \ge (k - 5/2 - h_{\max} - d)\lfloor\varepsilon\ell^d/4\rfloor$. The thesis thus follows if we can prove the existence of U with volume $\ell^d = \Omega(\log n/\log\log n)$ such that A_U is true.

Fix one such cube U. The probability that the long range edges inside U induce the expander G_U is larger than $(C(n)\ell^{-r})^{k\ell^d}$. On the other hand, for any vertex $i \in U$, the probability that no long range edge from $V\setminus U$ is incident on U is lower bounded as

$$\prod_{j\in V\setminus i}\big[1 - C(n)|i-j|^{-r}\big]^k \ge \exp\Big\{-3k\,C(n)\sum_{j\in V\setminus i}|i-j|^{-r}\Big\}$$

where we used the lower bound $1 - x \ge e^{-3x}$ valid for all $x \le 1/2$, together with the fact that $C(n) \le 1/2d$ (which follows by considering the $2d$ nearest neighbors). From the definition of $C(n)$, the last expression is lower bounded by e^{-3k}, whence

$$\mathbb{P}\{A_U\} \ge \big[C(n)e^{-3}\ell^{-r}\big]^{k\ell^d} .$$

Let S denote a family of (n/ℓ^d) disjoint subcubes, and denote by N_S the number of such subcubes for which property A_U holds. Then $\mathbb{E}[N_S] = (n/\ell^d)\mathbb{P}\{A_U\}$. Using the above lower bound together with the fact $C(n) \ge C_{r,d} > 0$ for $r > d$ and $C(n) \ge C_{*,d}/\log n$ for $r = d$, it follows that there exists $a, b > 0$ such that $\mathbb{E}[N_S] = \Omega(n^a)$ if $ell^d \le b\log n/\log\log n$.

The proof if finished by noticing that, for $U\cap U' =$, $\mathbb{P}\{A_U \cap A_{U'}\} \le \mathbb{P}\{A_U\cap A_{U'}\}$, whence $\mathrm{Var}(N_S) \le \mathbb{E}[N_S]$. The thesis follows applying Chebyshev inequality to N_S.

Small world networks with $r < d$. It is proved in [Fla06] that these graphs are with high probability expanders. The thesis follows from Lemma 3.2.

Random regular graphs. It is well known that a random k-regular graph is with high probability a $k - 2 - \delta$ expander for all $\delta > 0$ [Kah92]. The thesis follows again from Lemma 3.2. □

Acknowledgement. We would like to thank Daron Acemoglu, Glenn Ellison, Fabio Martinelli, Roberto Schonmann, Eva Tardos, Maria Eulalia Vares, and Peyton Young for helpful discussions and pointers to the literature.

References

[BK+05] N .Berger, C. Kenyon, E. Mossel and Y. Peres, *Glauber dynamics on trees and hyperbolic graphs,* Probab. Theory Relat. Fields, **131** (2005) 311-340.

[Blu93] L. E. Blume, *The Statistical Mechanics of Best-Response Strategy Revision*, Games and Economic Behavior **11** (1995) 111-145.

[Bov03] A. Bovier, *Markov Processes and Metastability* Lecture notes TUB, Summer 2006.

[BC96] G. Ben Arous and R. Cerf, *Metastability of the three dimensional Ising Model on a torus at very low temperature*, Electron. J. Probab. **1** (1996).

[BM02] A. Bovier and F. Manzo, J *Metastability in Glauber Dynamics in the Low-Tempersture Limit: Beyond Exponential Asymptotics*, J. Stat. Phys. **107** (2002) 757-779.

[DSC93] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains, Annals of Applied Probability* 3 (1993), pp. 696730.

[DV76] M. D. Donsker and S. R. S. Varadhan *On the principal eigenvalue of second-order elliptic differential operators* Commun. Pure Appl. Math. 29 (1976) 595-621

[Ell93] G. Ellison, *Learning, Local Interaction, and Coordination*, Econometrica **61** (1993) 1047-1071.

[FK02] U. Feige and R. Krauthgamer, *A Polylogarithmic Approximation of the Minimum Bisection.* SIAM J. Comput. (SIAMCOMP) 31(4):1090-1118 (2002).

[FL98] D. Fudenberg, D. K. Levine, *The Theory of Learning in Games* , MIT Press, 1998.

[Fla06] A. D. Flaxman, *Expansion and lack thereof in randomly perturbed graphs*, Proc. of Web Alg. Workshop, 2006.

[KMR93] M. Kandori, H. J. Mailath and F. Rob, *Learning, Mutation, and Long Run Equilibria in Games*, Econometrica **61** (1993) 29-56.

[Klein07] J. Kleinberg, *Cascading Behavior in Networks: Algorithmic and Economic Issues*, in 'Algorithmic Game Theory', N. Nisan et al. ed., Cambridge University Press, Cambridge 2007.

[GHS70] R. B. Griffiths, C. A. Hurst and S. Sherman, *Concavity of Magnetization of an Ising Ferromagnet in a Positive External Field*, J. Math Phys. 11 (1970) 790-795.

[HS88] J. C. Harsanyi and R. Selten, *A General Theory of Equilibrium Selection in Games,* MIT Press, Cambridge (1988).

[JS89] Alistair Sinclair, Mark Jerrum: Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains. Inf. Comput. (IANDC) 82(1):93-133 (1989).

[Kah92] N. Kahale, *On the second eigenvalue and linear expansion of regular graphs*, Foundations of Computer Science, Pittsburgh, October 1992, 296-303.

[LR99] T. Leighton , S. Rao, *Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms,* Journal of the ACM (JACM), v.46 n.6, p.787-832, Nov. 1999.

[M74] D. McFadden, *Conditional Logit Analysis of Qualitative Choice Behavior*, Frontiers in Econometrics, 1974.

[MP95] R. McKelvey, T. Palfrey, *Quantal Response Equilibria for Normal Form Games*, Games and Economic Behavior, 10: 6-38, 1995.

[MS94] D. Mookherjee, B. Sopher, *Learning Behavior in an Experimental Matching Pennies Game*, Games and Economic Behavior 7: 62-91, 1994.

[Morr00] S. Morris, *Contagion*, Rev. Econ. Stud. **67** (2000), 57-78.

[NeS91] E. J. Neves and R. H. Schonman, *Critical Droplets and Metastability for a Glauber Dynamics at Very Low Temperatures*, Commun. Math. Phys. **137** (1991) 209-230.

[NeS92] E. J. Neves and R. H. Schonman, *Behavior of droplets for a class of Glauber dynamics at very low temperatures*, Prob. Theor. Relat. Fields **91** (1992) 331-354.

[OV04] E. Olivieri and M. E. Vares, "Large deviations and metastability," Cambridge University Press, Cambridge, 2004.

[Young93] H. P. Young, *The Evolution of Conventions*, Econometrica **61** (1993) 57-84.

[Young95] H. P. Young, *The diffusion of innovation in social networks*, Santa Fe Inst. Working paper, 1999.

A Proof of Lemma 4.3

Assume without loss of generality that $\max\{|f_i| \,:\, i \in V\} = 1$, whence $f_i = 1$ for $i \in \Omega_1$. We use the same trick as in the proof of the standard Cheeger inequality

$$||\nabla_G f||^2 = \sum_{(i,j)\in E} (f_i - f_j)^2 \geq \frac{\left(\sum_{(i,j)\in E} |f_i^2 - f_j^2|\right)^2}{\sum_{(i,j)\in E} (f_i + f_j)^2}\,. \tag{21}$$

The denominator is upper bounded by

$$4 \sum_{i\in V} |\partial i|\, f_i^2 \leq 4 \max \left|\frac{|\partial i|}{h_i}\right| \;\; ||f||_h^2\,. \tag{22}$$

The argument in parenthesis at the numerator is instead equal to

$$\sum_{(i,j)\in E} \int_0^1 \left|\mathbb{I}(f_i^2 > z) - \mathbb{I}(f_j^2 > z)\right| \mathrm{d}z = \int_0^1 \mathrm{cut}(S_z, V\setminus S_z)\, \mathrm{d}z \tag{23}$$

where $S_z = \{i \in V :\; f_i^2 > z\}$. The quantity above is lower bounded by

$$\min_{z\in[0,1]} \frac{\mathrm{cut}(S_z, V\setminus S_z)}{|S_z|_h} \int_0^1 |S_z|_h\, \mathrm{d}z = \min_{z\in[0,1]} \frac{\mathrm{cut}(S_z, V\setminus S_z)}{|S_z|_h}\, ||f||_h\,. \tag{24}$$

Let $S = S_{z_*}$ where z_* realizes the above minimum (the function to be minimized is piecewise constants and right continuous hence the minimum is realized at some point). Notice that $\Omega_1 \subseteq S_z \subseteq \Omega_0$ for all $z \in [0,1]$, and thus we have in particular $L_1 \le |S|_w \le L_2$. Further, form the above

$$\lambda \geq \frac{||\nabla_G f||^2}{||f||_h^2} \geq \frac{1}{4} \min \left| \frac{h_i}{|\partial i|} \right| \left\{ \frac{\text{cut}(S, V \setminus S)}{|S|_h} \right\}^2 \tag{25}$$

which finishes the proof. □

B Hitting times at low temperature: proof of Lemma 4.1

We consider a general setting of Lemma 4.1: a discrete time Markov chain with state space $\mathcal{S}$, transition probabilities $p_\beta(x,y)$, reversible with respect to the stationary distribution $\mu(x)$. Given $A \subseteq \mathcal{S}$ define $p_\beta^A(x,y) = p_\beta(x,y)$ if $x,y \in \mathcal{S} \setminus A$ and $p_\beta^A(x,y) = 0$ otherwise. Notice by reversibility the eigenvalues of p_β^A are real, and smaller than 1. We assume that p_β^A is irreducible and aperiodic.

The lower bound in the next lemma is due to Donsker and Varadhan [DV76]: we nevertheless propose an elementary proof.

Lemma B.1. *If $1-\lambda_{0,A}$ is the largest eigenvalue of p_β^A, then*

$$\frac{1}{\log(1/(1-\lambda_{0,A}))} \le \tau_A \le \frac{1}{\log(1/(1-\lambda_{0,A}))} \left\{ 1 + \frac{1}{2} \max_{x \in \mathcal{S} \setminus A} \log \frac{1}{\mu(x)} \right\} .$$

Proof. Let P_A denote the matrix with entries $p_\beta^A(x,y)$, and $f(x)$ be the characteristic function of $\mathcal{S} \setminus A$. Then $\mathbb{P}_x \{T_A > t\} = P_A^t f(x)$, whence

$$\sqrt{\mu(x)}\, \mathbb{P}_x\{T_A > t\} \le \sqrt{\sum_x \mu(x) \mathbb{P}_x\{T_A > t\}^2} = ||P_A^t f||_{\mu,2} \le (1-\lambda_{0,A})^t ,$$

which proves the upper bound. To prove the lower bound, let $\psi_0(x)$ denote the eigenvector of P_A, with eigenvalue $\lambda_{0,A}$ and notice that by Perron-Frobenius theorem, it has non-negative entries. Therefore

$$\max_x \mathbb{P}_x\{T_A > t\}\, (\psi_0, f)_\mu \ge \sum_x \mu(x) \psi_0(x) \mathbb{P}_x\{T_A > t\} = (1-\lambda_{0,A})^t (\psi_0, f) .$$

□

Proof. (Lemma 4.1). Due to Lemma B.1, it is sufficient to prove that $\lambda_{0,A} = \exp\{-\beta\widetilde{\Gamma}_A + o(\beta)\}$. To this end we use the well known variational characterization of eigenvalues

$$\lambda_{0,A} = \inf_\varphi \frac{\text{Dir}(\varphi)}{\mathbb{E}(\varphi^2)} , \qquad \text{Dir}(\varphi) \equiv \frac{1}{2} \sum_{x,y} \mu(x) p_\beta(x,y) (\varphi(x) - \varphi(y))^2 . \tag{26}$$

Here the inf is taken over functions non-vanishing functions $\varphi : \mathcal{S} \setminus A \to \mathbb{R}$.

A lower bound can be obtained by comparison. More precisely, for each $z \in \mathcal{S} \setminus A$, let $\omega^{(z)}$ be a path or allowed transition from z to A. Proceeding along the lines of [JS89, DSC93], one obtains that $\lambda_{0,A} \geq 1/\max_{x,y} C(x,y;\omega)$, where, for each allowed transition $x \to y$, we defined the associated congestion as

$$C(x,y;\omega) = \frac{1}{\mu(x)p_\beta(x,y)} \sum_{z:\omega^{(z)} \ni (x,y)} \mu(z)|\omega^{(z)}|\,.$$

The thesis then follows by choosing the path $\omega^{(z)}$ in such a way to achieve the minimum in Eq. (10) and taking the limit $\beta \to \infty$.

To get an upper bound, define the boundary ∂B of a configuration B, as the subset of couples (x,y) such that $p_\beta(x,y) > 0$ and $x \in B$, while $y \notin B$. Notice that from Eq. (10) it follows that there exists a set $B \subseteq \mathcal{S} \setminus A$ such that

$$\widetilde{\Gamma}_A = \min_{(x,y)\in\partial B} [H(x) + V(x,y)] - \min_{z\in B} H(z)\,.$$

The proof is completed by taking φ in Eq. (26) to be the characteristic function of B. □

C Proof of Lemma 4.2

Construct the following partitioning of T into $T_1 = \{v_1, v_2, \cdots v_{i_1-1}\}$, $T_2 = \{v_{i_1}, v_{i_1+1}, \cdots v_{i_2-1}\}$ $\cdots T_r = \{v_{i_{r-1}} \cdots v_k\}$ in such a way that for every $T_j = \{v_{i_{j-1}}, \cdots v_{i_j-1}\}$ and $i_{j-1} < l < i_j$, $M(T_j, v_l - 1) = M(\{v_{i_{j-1}} \cdots v_{l-1}\}, v_l - 1) < 0$ and for $l = i_j$, $M(T_j, v_l - 1) \geq 0$.

Such a partition can be obtained the following way. Start with $j = 1$ and iteratively add v_i's to the current set T_j. If $M(T_j, v_i - 1) \geq 0$, increment j and add v_i and the next vertices to the new subset.

Let $T_r = \{v_s, \cdots, v_t\}$ be the last subset in the above sequence. We claim that if $M(T_r, k) < 0$ then $M(T_r, i) < 0$ for all $i \geq v_s$. For every $s \leq j \leq t$ and every i between v_j and v_{j+1} by supermodularity $M(T_r, i) = M(\{v_l, \cdots v_j\}, i) \leq M(\{v_l, \cdots v_j\}, v_{j+1} - 1) < 0$. The same argument goes for $v_t \leq i \leq k$. In that case the lemma is correct for $T' = T_r$.

If $M(T_r, k) \geq 0$, we will show that the second statement of the lemma is true. For that, we need to write the H function for all sets $T_1, \cdots T_r$ explicitly. For a set T_j and $l = i_j$

$$M(T_j, v_l - 1) = 2\left[\mathrm{cut}(T_j, \{1, 2, \cdots v_l - 1\}) - \mathrm{cut}(T_j, \{v_l, v_l + 1, \cdots n\}) + \sum_{i\in T_j} h_i\right] \geq 0\,. \tag{27}$$

One can write a similar equation $j = l$ by replacing $v_l - 1$ with k. Equation (12) gives a similar inequality for R. Adding up these inequalities for all j and R and noting that the contribution of every edge with both ends in $\cup_j T_j \cup R$ cancels out, we get

$$M(T \cup R, k) \geq \sum_{j=1}^{l-1} M(T_j, v_{i_j} - 1) + M(T_l, k) + M(R, 0) \geq 0. \tag{28}$$

□

D Proof of Theorem 3.4

Proof. (Theorem 3.4). Partition V into subsets $R_1, R_2, \cdots, R_l$ by letting $V_0 \equiv V$ and defining recursively

$$R_t = \arg\min_{S \in \Omega_t} \{\mathrm{cut}(S, V_t \setminus S) - |S|_{h^{V_t}}\}$$

where $V_t = V \setminus \cup_{s=1}^{t-1} R_s$ and Ω_t is the set of all subsets $S \subseteq V_t$ such that $L_1 \le |S|_h \le L_2$. With an abuse of notation, we wrote $\underline{h}^{V_t}$ for $\underline{h}^{G(V_t)}$ ($G(V_t)$ being the subgraph induced by V_t). Explicitly, for any $j \in V_t$, $(h^{V_t})_j = h_j + |\partial j|_{V \setminus V_t}$.

Continue this process until no such set S can be found, and let $R_l = V_l$ be the residual set. Notice that, since $L_2 \ge h_{\max}$, we necessarily have $|R_l|_h < L_1$. By applying Eq. (9) to $F = G(V_t)$, we have

$$\mathrm{cut}(R_t, V_t \setminus R_t) \le |R_t|_{h^{V_t}} + L_1 \le |R_t|_{h^{V_t}} + |R_t|_h = |R_t|_{2h} + \mathrm{cut}(R_t, V \setminus V_t)\,. \tag{29}$$

Notice that $\mathrm{cut}(R_t, V_t \setminus R_t) - \mathrm{cut}(R_t, V \setminus V_t) = \mathrm{cut}(\cup_{s=1}^{t} R_s, V_{t+1}) - \mathrm{cut}(\cup_{s=1}^{t-1} R_s, V_t)$. By summing up this relation, we have, for all $1 \le t < l$,

$$\mathrm{cut}(\cup_{s=1}^{t} R_s, V \setminus \cup_{s=1}^{t} R_s) \le \sum_{s=1}^{t} |R_s|_{2h} = |\cup_{s=1}^{t} R_s|_{2h}.$$

For each R_t, consider a linear arrangement of the induced subgraph that achieves its cutwidth. Construct a linear arrangement of V by concatenating the above linear arrangement of each R_t in the order $t = 1, 2, \ldots, l$. We will show that this ordering gives us the desired upper bound on the tilted cutwidth of G. Let $S = \cup_{s=1}^{t-1} R_s \cup R$ where $R \subset R_t$ for some t between 1 and l. Then

$$\begin{aligned}
\mathrm{cut}(S, V \setminus S) &\le \mathrm{cut}(\cup_{s=1}^{t-1} R_s, V \setminus \cup_{s=1}^{t-1} R_s) + \mathrm{cut}(R_t, V \setminus V_t) + \mathrm{cutwidth}(R_t) \\
&\le \mathrm{cut}(\cup_{s=1}^{t-1} R_s, V \setminus \cup_{s=1}^{t-1} R_s) + \mathrm{cut}(R_t, V \setminus V_t) + |R_t|_h + L_1 + C \\
&\le 2\,\mathrm{cut}(\cup_{s=1}^{t-1} R_s, V \setminus \cup_{s=1}^{t-1} R_s) + L_1 + L_2 + C \\
&\le 2|\cup_{s=1}^{t-1} R_s|_{2h} + L_1 + L_2 + C\,.
\end{aligned}$$

□

E Eigenvectors and barriers

As in the last appendix, we consider here a general Markov chain with state space $\mathcal{S}$, and let $A \subseteq \mathcal{S}$ a subset of configurations.

Lemma E.1. *Let $\psi_0 : \mathcal{S} \to \mathbb{R}$ be the unique eigenvector of P_A with eigenvalue $1 - \lambda_{0,A}$ and assume (without loss of generality by Perron-Frobenius theorem) $\psi_0(x) \ge 0$. Then there exists $b \ge 0$ such that, letting $B = \{x \in \mathcal{S} :\ \psi_0(x) > b\}$, we have*

$$\frac{1}{|\mathcal{S}|} \frac{\sum_{(x,y)\in\partial B} \mu(x) p_\beta(x,y)}{\sum_{x\in B} \mu(x)} \le \lambda_{0,A} \le \frac{\sum_{(x,y)\in\partial B} \mu(x) p_\beta(x,y)}{\sum_{x\in B} \mu(x)} \tag{30}$$

Proof. The upper bound follows immediately by substituting $\varphi(x) = \mathbb{I}(x \in B)$ in the variational principle (26).

In order to prove the lower bound, let $0 = \psi^{(0)} < \psi^{(1)} \le \cdots \le \psi^{(N)}$ be the points in the image of $\psi_0(\,\cdot\,)$ (obviously $N \le \mathcal{S}$). For any (x, y) such that $\psi_0(x) = \psi^{(i)}$, $\psi_0(y) = \psi^{(j)}$, with $i < j$, we have $(\psi_0(x) - \psi_0(y))^2 \ge \sum_{l=i}^{j-1} (\psi^{(l+1)} - \psi^{(l)})^2$. Therefore, by letting $B_l = \{x \in \mathcal{S} : \psi_0(x) \ge \psi^{(l)}\}$, we have

$$\mathrm{Dir}(\psi_0) \ge \sum_{l=1}^{N} W(l)\,(\psi^{(l)} - \psi^{(l-1)})^2\,, \qquad W(l) \equiv \sum_{(x,y)\in\partial B_l} \mu(x) p_\beta(x,y)\,. \tag{31}$$

On the other hand, $(\psi^{(i)})^2 \le i \sum_{l=1}^{i} (\psi^{(l)} - \psi^{(l-1)})^2$. If $M(l) \equiv \sum_x \mu(x) \mathbb{I}(\psi_0(x) = \psi^{(l)}) = \mu(B_l) - \mu(B_{l-1})$

$$\mathbb{E}(\psi_0^2) = \sum_{i=0}^{N} M(i)\,(\psi^{(i)})^2 \le \sum_{l=1}^{N} \Big(\sum_{i=l}^{N} i\, M(i) \Big)\,(\psi^{(l)} - \psi^{(l-1)})^2\,. \tag{32}$$

Therefore

$$\lambda_{0,A} = \frac{\mathrm{Dir}(\psi_0)}{\mathbb{E}(\psi_0^2)} \ge \inf_{1\le l\le N} \frac{W(l)}{\sum_{i=l}^{N} i\, M(i)}\,, \tag{33}$$

which implies the thesis. □