Convergence to Equilibrium in Local Interaction Games

Andrea Montanari* and Amin Saberi†

April 2, 2009

Abstract

We study a simple game-theoretic model for the spread of an innovation in a network. The diffusion of the innovation is modeled as the dynamics of a coordination game in which the adoption of a common strategy between players has a higher payoff.

Classical results in game theory provide a simple condition for the innovation to spread through the network. The present paper characterizes the rate of convergence as a function of graph structure. In particular, we derive a dichotomy between well-connected (e.g. random) 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 networks and their increasing role in the spread of knowledge, behaviors and new technologies has 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, Young01]?

One way to address this question is to use epidemic or independent cascade models. In epidemic models, the underlying assumption is that people adopt an innovation when they come in contact with others who have already adopted it, that is, innovations spread much like epidemics.

The focus of this paper is on game-theoretic models that are based on the notion of utility maximization rather than exposure. The basic hypothesis is that, when adopting a new behavior, each individual makes a rational choice to maximize his or her payoff in a coordination game. In these games players adopt a strategy when enough of their neighbors in the social network have adopted it, that is, innovations spread because there is an incentive to conform.

For example, consider a simple game in which each individual placed in a network has to make a decision between two alternatives. The payoff of an action for each player increases linearly with the number of its neighbors who are taking the same action. Agents revise their strategies asynchronously each time choosing the strategy with the best payoff with probability close to 1.

Such noisy best-response dynamics have been studied extensively as a simple model for the emergence of technologies and social norms [HS88, KMR93, Young93, Young01]. The main result in this line of work is that the combination of random experimentation (noise) and the myopic attempts of players to increase their utility by coordination (best response) creates a powerful evolutionary force. The dynamics drives the system towards a particular equilibrium in which all players take the same action. The analysis also offers a simple condition (known as risk dominance) that determines whether an innovation introduced in the network will eventually become widespread.

The present paper characterizes the *rate of convergence* for such dynamics in terms of explicit graph quantities. Suppose a superior (risk dominant) technology is introduced as a new alternative. How long does it take for it to become widespread in the network?

Our characterization is expressed in terms of quantities that we name tilted cutwidth and tilted cut of the graph. These can be seen as duals of each other: The former provides a path to the risk-dominant equilibrium that implies an upper bound on the convergence time. The latter corresponds to a bottleneck along the separating set in the space of configurations with lowest probability. We prove a duality theorem that shows that tilted cut and tilted cutwidth coincide for the 'slowest' subgraph. 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 a suitable transition matrix. The convergence time is then estimated in terms of the most likely path from the worst-case initial configuration. A key contribution of this paper consists in proving that there exists a monotone increasing path with this property. This indicates that the risk dominant strategy indeed *spreads through the network*, i.e. an increasing subset of players adopt it over time. 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 sparse graphs such as random regular graphs or power-law networks or certain small-world networks the convergence may take as long as exponential in the number of nodes.

This observation also highlights an important difference between game theoretic and epidemic models. In epidemic models, the innovation spreads very quickly in well-connected networks. Moreover, existence of high degree nodes, expedites the rate of diffusion significantly [Lig05]. The striking difference between the dynamics of this model and the result of our analysis gives the first rigorous evidence that the aggregate behavior of the diffusion is indeed very sensitive to the dynamics of the interaction of individuals. This may suggest that assuming that diffusion of a virus, a new technology, or a new political or social belief have the same "viral" behavior may be misleading. And this intuition can be quite important when it comes to making predictions or developing algorithms for spreading or containing them (see also [Watts] for a related discussion).

Related work

Kandori, Mailath and Rob [KMR93] studied the noisy best response dynamics and showed that it converges to the equilibrium in which every agent takes the same strategy. Harsanyi and Selten [HS88] named this the *risk dominant* strategy (see next section for a definition).

The role of graph structure and its interplay with convergence times was first emphasized by Ellison [Ell93]. In his pioneering work, Ellison considered two types of structures for the interaction network: a complete graph, and a one-dimensional network obtained by placing individuals on a cycle and connecting all pairs of distance smaller than some given constant. Ellison proved that, on the first type of graph structure, convergence to the risk dominant equilibrium is extremely slow (exponential in the number of players) and for practical purposes, not observable. On the contrary, convergence is relatively fast on linear network and the risk dominant equilibrium is an important predictive concept in this case. Based on this observation, Ellison concludes that when the interaction is global the outcome is determined by historic factors. In contrast, when players "interact with small sets of neighbors," evolutionary forces may determine the outcome.

Even though this result has received a lot of attention in the economic theory (for example see detailed expositions in books by Fudenberg [FL98] and by Young [Young01]) the conclusion of [Ell93] has remained rather imprecise. The contribution of the current paper is to precisely derive the graph quantity that captures the rate of convergence. Our results make a different prediction on models of social networks that are well-connected but sparse. We also show how to interpret Ellison's result by defining a geometric embedding of graphs.

Most of our results are based on a reversible Markov chain model for the dynamics. Blume [Blu93] already studied the same model within a social science context rederiving the results by Kandori et al. [KMR93]. In Section 4.1 we consider generalizations to a broad family of non-reversible dynamics.

Finally, we refer to the next two sections for a comparison with related work within mathematical physics and Markov Chain Monte Carlo theory.

2 Definitions

A game is played in periods $t = 1, 2, 3, \dots$ among a set V of players, with $|V| = n$. The players interact on an undirected graph $G = (V, E)$. Each player $i \in V$ has two alternative strategies denoted by $x_i \in \{+1, -1\}$. The payoff matrix A is a 2×2-matrix illustrated in the figure. Note that the game is symmetric. The payoff of player i is $\sum_{j \in N(i)} A(x_i, x_j)$, where $N(i)$ is the set of neighbors of vertex i.

We assume that the game defined by matrix A is a coordination game, i.e. the players obtain a higher payoff from adopting the same strategy as their opponents. More precisely, we have $a > d$ and $b > c$.

a	c
d	b

Let $N_+(i)$ and $N_-(i)$ be the set of neighbors of i adopting strategy $+1$ and -1 respectively. The best strategy for a node i is $+1$ if $(a-d)N_+(i) \geq (b-c)N_-(i)$ and it is -1 otherwise. For the convenience of notation, let us define $h = \frac{a-d-b+c}{a-d+b-c}$ and $h_i = h|N(i)|$ where $N(i)$ is the set of neighbors of i. In that case, every node i has a threshold value h_i such that the best response strategy can be written as $\text{sign}(h_i + \sum_{j \in N(i)} x_j)$.

We assume that $a - b > d - c$, so that $h_i > 0$ for all $i \in V$ with non-zero degree. In other words, when the number of neighbors of node i taking action $+1$ is equal to the number of its neighbors taking action -1, the best response for i is $+1$. Harsanyi and Selten [HS88] named $+1$ the "risk-dominant" action because it seems to be the best strategy for a node that does not have any information about its neighbors. Notice that it is possible for h to be larger than 0 even though $b > a$. In other words, the risk-dominant equilibrium is in general distinct from the "payoff-dominant equilibrium", the equilibrium in which all the players have the maximum possible payoff.

It is easy to verify that coordination games belong to the class of potential games. As a consequence, best response dynamics always converges to one of the pure Nash equilibria. In this paper, we study *noisy* best response dynamics. In this dynamics, when the players revise their strategy they choose the best response action with probability close to 1. Still, there is a small chance that they choose the alternative strategy with inferior payoff.

More formally, a 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 or best-response dynamics.

We assume that each node i updates its value at the arrival time of an independent Poisson clock of rate 1. The probability that node i take action y_i is proportional to $e^{\beta y_i (h_i + \sum_{j \in N(i)} x_j)}$. More precisely, the conditional distribution of the new strategy is

$$p_{i,\beta}(y_i | \underline{x}_{N(i)}) = \frac{e^{\beta y_i (h_i + \sum_{j \in N(i)} x_j)}}{e^{\beta (h_i + \sum_{j \in N(i)} x_j)} + e^{-\beta (h_i + \sum_{j \in N(i)} x_j)}}. \quad (1)$$

Note that this is equivalent to the best response dynamics $y_i = \text{sign}(h_i + \sum_{j \in N(i)} x_j)$ for $\beta = \infty$. The above chain is called *heat bath* or *Glauber* dynamics for the Ising model. It is also known as logit update rule which is the standard model in the discrete choice literature [M74, MS94, MP95].

Let $\underline{x} = \{x_i : i \in V\}$. The corresponding Markov chain is reversible with the stationary distribution $\mu_\beta(\underline{x}) \propto \exp(-\beta H(\underline{x}))$ where

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

For large β, the stationary distribution concentrates around the all-$(+1)$ configuration. In other words, this dynamics predicts that the $+1$ equilibrium or the Harsanyi-Selten's risk-dominant equilibrium is the likely outcome of the play in the long run.

The above was observed in by Kandori et al. [KMR93] and Young [Young93] for a slightly different definition of noisy-best response dynamics (see Section 4.1 and Appendix A). Their result has been studied and extended as a method for refining Nash equilibria in games. Also it has been used as a simple model for studying formation of social norms and institutions and diffusion of technologies. See [FL98] for the former and [Young01] for an exposition of the latter.

Our aim is to determine whether the convergence to this equilibrium is realized in a reasonable time. For example, suppose the behavior or technology corresponding to action -1 is the widespread action in the network. Now the technology or behavior $+1$ is offered as an alternative. Suppose $a > b$

and $c = d = 0$ so the innovation corresponding to $+1$ is clearly superior. The above dynamics predict that the innovation corresponding to action $+1$ will eventually become widespread in the network. We are interested to characterize the networks on which this innovation spreads in a reasonable time.

To this end, we let T_+ denote the hitting time or convergence time to the all-$(+1)$ configuration, and define the *typical hitting time* for $\underline{+1}$ as

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

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

Relations with MCMC theory and statistical physics

The reversible Markov chain studied in this paper coincides with the Glauber dynamics for the Ising model, and is arguably one of the most studied Markov chains of the same type. Among the few general results, Berger et al. [BK+05] proved an upper bound on the mixing time for $h = 0$ in terms of the cutwidth of the graph. Their proof is based a simple but elegant canonical path argument. Because of our very motivation, we must consider $h > 0$. It is important to stress that this seemingly innocuous modification leads to a dramatically different behavior. As an example, for a graph with a d-dimensional embedding (see below for definitions and analysis), the mixing time is $\exp\{\Theta(n^{(d-1)/d})\}$ for $h = 0$, while for any $h > 0$ is expected to be polynomial. This difference is not captured by the approach of [BK+05]: adapting the canonical path argument to the case $h > 0$ leads to an upper bound of order $\exp\{\Theta(n^{(d-1)/d})\}$. We will see below that the correct behavior is instead captured by our approach.

There is another important difference with respect to the MCMC literature. In modeling economic phenomena, it is well accepted that players respond in the same way to the same environment, thus ruling out any non-deterministic dynamics. Starting with [KMR93], this assumption was slightly relaxed: instead of strictly deterministic dynamics, the focus has passed to the noiseless ($\beta \to \infty$) limit of a family of stochastic dynamics. We will take the same point of view. As in the social science literature, our estimates of the convergence time will take the form $\tau_+(G;\underline{h}) = \exp\{\beta\Gamma_*(G) + o(\beta)\}$. The constant $\Gamma_*(G)$ will then be estimated for large graphs. In particular, even if $\Gamma_*(G)$ is upper bounded by a constant, the $o(\beta)$ term can hide n-dependent factors.

Focusing on the $\beta \to \infty$ limit is also important if we want to model the spread of a new technology, behavior or product. In these contexts we rarely witness an individual changing opinion several times, or a fraction of the population oscillating back and forth. This is instead what happens with the Markov chain (1) for bounded β. On the other side, for $\beta \to \infty$ the players switches from the -1 to the $+1$ equilibrium along a well defined evolution sequence.

The same point of view (namely studying Glauber dynamics in the $\beta \to \infty$ limit) has been explored within mathematical physics to understand 'metastability.' This line of research has lead to sharp estimates of the convergence time (more precisely, of the constant $\Gamma_*(G)$) when the graph is a two- or three-dimensional grid [NeS91, NeS92, BC96, BM02]. It is natural to ask how robust these results are when the graph is perturbed: we will answer to this question in several cases.

3 Main results: Specific graph families

The main result of this paper is to derive the graph theoretical quantity that captures the low noise behavior of the hitting time. In order to build intuition, we will start with some familiar and natural models of social networks:

(a) *Random graphs.* Including random regular graphs of degree $k \geq 3$, random graphs with a fixed degree sequence with minimum degree 3, and random graphs in preferential-attachment model with minimum degree 2 [MPS06, GMS03].

(b) *d-dimensional networks.* We say that the graph G is *embeddable* in d dimensions or is a d-dimensional range-K graph if one can associate to each of its vertices $i \in V$ a position $\xi_i \in \mathbb{R}^d$ such that, (1) $(i,j) \in E$ implies $d_{\text{Eucl}}(\xi_i, \xi_j) \leq K$ (here $d_{\text{Eucl}}(\cdots)$ denotes Euclidean distance); (2) Any cube of volume v contains at most $2\,v$ vertices.

(c) *Small-world networks.* The vertices of this graph 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 $P_i(j) = C(n)|i-j|^{-r}$.

Theorem 3.1. *As $\beta \to \infty$, the convergence time is $\tau_+(G) = \exp\{2\beta\Gamma_*(G) + o(\beta)\}$ where*

(i) If G is a random k-regular graph with $k \geq 3$, a random graphs with a fixed degree sequence with minimum degree 3 or a preferential-attachment graph with minimum degree 2, then for h small enough, $\Gamma_*(G) = \Omega(n)$.

(ii) If G is a d-dimensional graph with bounded range, then for all $h > 0$, $\Gamma_*(G) = O(1)$.

(iii) If G is a small world network with $r \geq d$, and h is such that $\max_i h_i \leq k - d - 5/2$, then with high probability $\Gamma_*(G) = \Omega(\log n / \log\log n)$.

(iv) If G is a small world network with $r < d$, and h is small enough, then with high probability $\Gamma_*(G) = \Omega(n)$.

The basic implication of the above theorem is that if the underlying interaction or social network is well-connected, i.e. if it resembles a random-regular graph or a power-law graph, then the $+1$ action spreads very slowly in the network. On the other hand, if the interaction is restricted only to individuals that are geographically close, then convergence to $+1$ equilibrium is very fast.

The proof of is based on relating the convergence time to the isoperimetric function of G.

Lemma 3.2. *Let G be a graph with maximum degree Δ. Assume that there exist constants α and $\gamma < 1$ such that for any subset of vertices $U \subseteq V$, and for any $k \in \{1, \dots, |U|\}$*

$$\min_{S \subseteq U, |S|=k} \text{cut}(S, U \setminus S) \leq \alpha\, |S|^\gamma \,. \tag{4}$$

Then there exists constant $A = A(\alpha, \gamma, h, \Delta)$ such that $\Gamma_*(G) \leq A$.

Conversely, for a graph G with degree bounded by Δ, assume there exists a subset $U \subseteq V(G)$, such that for $i \in U$, $|N(i) \cap (V \setminus U)| \leq M$, and the subgraph induced by U is a (δ, λ) expander, i.e. for every $k \leq \delta|U|$,

$$\min_{S \subseteq U, |S|=k} \text{cut}(S, U \setminus S) \geq \lambda\, |S| \,. \tag{5}$$

Then $\Gamma_*(G) \geq (\lambda - h\Delta - M)\lfloor \delta |U| \rfloor$.

In words, an upper bound on the isoperimetric function of the graph leads to an upper bound on the hitting time. On the other hand, highly connected subgraphs that are loosely connected to the rest of the graph can slow down the convergence significantly.

It is not hard to derive the proof of Theorem 3.1 (*i*), (*iii*), and (*iv*) from the above lemma. Random regular graphs, preferential-attachment graphs, and and small-world networks with $r < d$ have constant expansion. Small-world networks with $r \geq d$ contain a small, highly connected regions of size roughly $O(\log n)$. In fact, the proof of this part of theorem is based on identifying an expander of this size in the graph.

For part (*ii*) of Theorem 3.1 note that, roughly speaking, in networks with dimension d, the number of edges in the boundary of a ball that contains v vertices is of order $O(v^{1-1/d})$. Therefore the first part of Lemma 3.2 should give an intuition on why the convergence time is fast. The actual proof is significantly less straightforward because we must control the isoperimetric function of *every* subgraph of G (and there is no monotonicity with respect to the graph). The proof is presented in Section 5.

So far we assumed that $h_i = h|N(i)|$. This choice simplifies the statements but is not technically needed. In the next section, we will consider a *generic graph* G and generic values of $h_i \geq 0$.

4 Results for general graphs

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

$$\Gamma(G;\underline{h}) \equiv \min_{S:\emptyset\to V}\max_{t\leq n}\left[\mathrm{cut}(S_t, V\setminus S_t) - |S_t|_h\right] . \tag{6}$$

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

$$\Delta(G;\underline{h}) \equiv \max_{\Omega}\min_{(S_1,S_2)\in\partial\Omega}\max_{i=1,2}\left[\mathrm{cut}(S_i, V\setminus S_i) - |S_i|_h\right] , \tag{7}$$

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

Theorem 4.1. *Given an induced subgraph $F \subseteq G$, let $\underline{h}^F$ be defined by $h_i^F = h_i + |N(i)|_{G\setminus F}$, where $|N(i)|_{G\setminus F}$ 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*

$$\Gamma_*(G;\underline{h}) = \max_{F\subseteq G}\Gamma(F;\underline{h}^F) = \max_{F\subseteq G}\Delta(F;\underline{h}^F) . \tag{8}$$

Note that tilted cutwidth and tilted cut are dual quantities. The former corresponds the maximum increase in the potential function $H(\underline{x})$ along the lowest path to the $+1$ equilibrium. The latter is the lowest value of potential function along the highest separating set in the space of configurations. The above theorem shows that tilted cut and cutwidth coincide for the 'slowest' subgraph of G provided that the h_i's are non-negative. This identity is highly non-trivial: for instance the two expressions in Eq. (8) do not coincide for all subgraphs F. The hitting time is exponential in this graph parameter.

Monotonicity of the optimal path. The linear ordering in Eq. (6) corresponds to an evolution path leading to the risk-dominant (all $+1$) equilibrium from a different equilibrium. Characterizing the optimal path provides insight on the typical process by which the network converges to the $+1$ equilibrium [OV04]. For instance, if all optimal paths include a certain configuration $\underline{x}$, then the network will pass through the state $\underline{x}$ on its way to the new equilibrium, with probability converging to 1 as $\beta \to \infty$. Remarkably in Eq. (6) it is sufficient to optimize over linear orderings, instead of generic paths in $\{+1,-1\}^V$. This is suggestive of the fact that the convergence to the risk-dominant equilibrium is realized by a monotone process: the new $+1$ strategy effectively spreads as a new behavior is expected to spread. A similar phenomenon was indeed proved in the case of two- and three-dimensional grids [NeS91, NeS92, BC96]. Here we provide rigorous evidence that it is indeed generic.

In Appendix A we compare Theorem 4.1. with results in the economics literature.

4.1 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 $p_{i,\beta}(y_i|\underline{x}_{N(i)})$, with $p_{i,\beta}(+1|\underline{x}_{N(i)})$ a non-decreasing function of the number $\sum_{j\in N(i)} x_j$. Further we assume that $p_i(+1|\underline{x}_{N(i)}) \le e^{-2\beta}$ when $h_i + \sum_{j\in N(i)} x_j < 0$. 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 $\tau_+(G;\underline{h})$ as before.

Proposition 4.2. *Let $G = (V,E)$ be a k-regular graph of size n such that for $\lambda, \delta > 0$, every $S \subset V, |S| \le \delta n$ 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 $\tau_+(G;\underline{h}) \ge \exp\{\beta c n\}$ as long as $\lambda > (3k/4) + (\max_i h_i/2)$.*

Note that random regular graphs satisfy the condition of the above proposition as long as h_i's are small enough. This proposition can be proved by considering the evolution of one dimensional chain tracking the number of $+1$ vertices.

Proposition 4.3. *Let G be a d-dimensional grid of size n and constant $d \ge 1$. For any synchronous or asynchronous noisy-best response dynamics defined above, there exists constant c such that $\tau_+(G;\underline{h}) \le \exp\{\beta c\}$.*

The above proposition can be proved by a simple coupling argument very similar to that of Young [Young06]. We will leave its details to a more complete version of the paper. Together, these 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.

5 Proofs

5.1 Relating the rate of convergence to tilted cutwidth

We start by proving Theorem 4.1. The first part of the proof relates the hitting time of $+1$ to the evolution of the potential function H. The main intuition of the lemma is as follows: the dynamics has a tendency to decrease the value of potential function H. However to reach the set A from z, it may be necessary to go through configurations that have high values of H. These configurations create a barrier and the hitting time is related exponentially to the height the path with the smallest barrier.

Lemma 5.1. *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 $A = \{x : H(x) \le H_0\}$ 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

$$\widetilde{\Gamma}_A = \max_{z\notin A} \min_{\omega: z\to A} \max_{t\le|\omega|-1} [H(\omega_t) + V(\omega_t,\omega_{t+1}) - H(z)] \ , \tag{9}$$

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 of this lemma 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 E.

For the sake of clarity, we split the proof of Theorem 4.1 in two parts: first the characterization in terms of tilted cutwidth (i.e. the first identity in Eq. (8)); the one in terms of tilted cut (second identity in Eq. (8)) is deferred to Appendix B.

Proof. (Theorem 4.1, Tilted cutwidth). Notice that Glauber dynamics satisfies the hypotheses of Lemma 5.1, with $H(\underline{x})$ 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. (9). We thus obtain $\tau_+ = \exp(\beta \max_{\underline{z}} \widetilde{\Gamma}_+(\underline{z}) + o(\beta))$ where

$$\widetilde{\Gamma}_+(\underline{z}) = \min_{\omega:\underline{z}\to\underline{+1}} \max_{t\le|\omega|-1} \left[H(\omega_t) - H(\underline{z})\right] . \qquad (10)$$

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. The first part of the proof is dedicated to proving that.

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 $A \subset B$, $H(A) \ge H(B)$. 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 $1 \le i \le k$, $S_i = \{1, 2, \cdots i\} \cup S_0$. Let $v_1 \le v_2 \cdots \le v_t$ be the elements of T in the order of their appearance in ω.

For a subset $A \subset T$, and $i \le k$ 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,

$$M(R, 0) = H(S_0) - H(S_0 \setminus R) \ge 0 . \qquad (11)$$

The crucial lemma below is proved in Appendix F.

Lemma 5.2. *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

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

□

5.2 Deriving the rate of convergence for specific graph families

In this section, we present the proof of Theorem 3.1. The first step is to relate the tilted cut and tilted cutwidth to isoperimetric functions of the graph. Such a relation is provided by Lemma 3.2, which is proved in Appendix C.

Further, it will be useful to replace paths in the configuration space by bolder steps.

Theorem 5.3. *Assume that, for some L_1, L_2, with $L_2 \geq h_{\max}$ and for every induced subgraph $F \subseteq G$, we have*

$$\min_{|S|_h \in [L_1, L_2]} [\mathrm{cut}(S, V(F) \setminus S) - |S|_{h^F}] \leq L_1\,, \tag{12}$$

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 $\Gamma(G; 4\underline{h}) \leq C + L_1 + L_2$.

The proof of this result is in Appendix D, while further discussion and comparison with the literature can be found in Section D.1.

Finally, we need to estimate the isoperimetric function of 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

$$||f||_w^2 \equiv \sum_{i \in V} w_i\, f_i^2\,, \qquad ||\nabla_G f||^2 \equiv \sum_{(i,j) \in E} |f_i - f_j|^2\,. \tag{13}$$

We then have the following generalization of Cheeger inequality.

Lemma 5.4. *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 $L_1 \leq |S|_w \leq L_2$*

$$\mathrm{cut}(S, V \setminus S) \leq \sqrt{4\lambda \max_{i \in V}\{|N(i)|/h_i\}}\ \ |S|_h\,. \tag{14}$$

The proof of this Lemma is deferred to Appendix G.

Proof. (Theorem 3.1)

Random 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]. Also, it is known that for small constant λ, random graphs with a fixed degree sequence with minimum degree 3, and random graphs in preferential attachment model with minimum degree 2 have expansion λ with high probability [GMS03, MPS06]. The thesis follows from Lemma 3.2.

d-dimensional networks. We need to prove that, for each induced subgraph G', $\Gamma(G'; \underline{h}^{G'}) = O(1)$. By Theorem 5.3, 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 $\text{cut}(S, V \setminus S) \leq |S|_{h/4}$.

We will construct a function f with bounded support such that $||\nabla_G f||^2 \leq \lambda ||f||^2$ with $\lambda = \min_{i\in V}\{\frac{h_i}{16|N(i)|}\}$. 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 $\sum_{i:\xi_i\in\mathcal{C}} h_i$, and let C_j, $j = 1, \dots 3^d - 1$ be the adjacent cubes. Let $f_i = \varphi(\xi_i)$, where for $x \in \mathbb{R}^d$, we have

$$\varphi(x) = \left[1 - \frac{d_{\text{Eucl}}(x, \mathcal{C})}{\ell}\right]_+ . \tag{15}$$

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

$$\begin{aligned} ||\nabla_G f||^2 &\leq \left(\frac{K}{\ell}\right)^2 \sum_{i\in V} |N(i)|\, \mathbb{I}(\xi_i \in \cup_{j=1}^{3^d-1}\mathcal{C}_j) \leq \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|N(i)|/h_i\} \sum_{i\in V} h_i\, \mathbb{I}\Big(\xi_i \in \cup_{j=1}^{3^d-1}\mathcal{C}_j\Big) \\ &\leq 3^d \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|N(i)|/h_i\} \sum_{i\in V} h_i\, \mathbb{I}\Big(\xi_i \in \mathcal{C}_0\Big) \leq 3^d \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|N(i)|/h_i\} ||f||_h^2 . \end{aligned} \tag{16}$$

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

Small world networks with $r \geq d$. 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 $\Gamma_*(G; \underline{h}) \geq (k - 5/2 - h_{\max} - d)\lfloor \varepsilon\ell^d/4 \rfloor$. 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

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

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

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

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 $C(n) \geq C_{r,d} > 0$ for $r > d$ and $C(n) \geq C_{*,d}/\log n$ for $r = d$, it follows that there exists $a, b > 0$ such that $\mathbb{E}[N_S] = \Omega(n^a)$ if $\ell^d \leq b\log n/\log\log n$.

The proof if finished by noticing that, for $U \cap U' =$, $\mathbb{P}\{A_U \cap A_{U'}\} \leq \mathbb{P}\{A_U \cap A_{U'}\}$, whence $\text{Var}(N_S) \leq \mathbb{E}[N_S]$. 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.

□

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

[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).

[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.

[BM02] A. Bovier and F. Manzo, *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.

[GMS03] C. Gkantsidis, M. Mihail, and A. Saberi, *Throughput and Congestion in Power-Law Graphs,* SIGMETRICS 2003.

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

[JS89] A. Sinclair, M. 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.

[Lig05] T. M. Liggett, *Interacting particle systems*, Springer, 2005.

[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.

[MPS06] M. Mihail, C. Papadimitriou, and A. Saberi, *On Certain Connectivity Properties of the Internet Topology,* Journal of Computer and System Sciences, 2006.

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

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

[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.

[Watts] D. Watts, *Challenging the Influentials Hypothesis*, WOMMA Measuring Word of Mouth, 2007.

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

[Young01] H. P. Young, *Individual strategy and social structure: an evolutionary theory of institutions*, Princeton University Press, 2001.

[Young06] H. P. Young, *The diffusion of innovation in social networks*, in “The economy as an evolving complex system, III: current perspectives and future directions”, L. Blume, S. N. Durlauf Eds. Oxford University Press, 2006

A Comparison with results in the economics literature

Ellison [Ell93] originally considered a Markov chain with transition rates slightly different from the ones of Glauber dynamics. At each time step, each node i updates its strategy to the best response one $\text{sign}(h_i + \sum_{j \in N(i)} x_j)$ with probability $1 - e^{-\beta}$ and to the opposite one with probability $e^{-\beta}$. In other words, the probability of making a mistake is independent of the loss in utility. In Section 4.1 we discuss a class of general models including Ellison's Markov chain.

On the other hand, it is interesting to consider the implications of Theorem 4.1 for the [Ell93] are easily analyzed 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 $n/2$. By evaluating Eq. (7) we get $\Gamma_*(K_n; \underline{h}) \geq (n-h)^2/4 + O(n)$. 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 $\Gamma(G; h) \leq 4k^2$.

Young [Young06] studied instead Glauber dynamics, and proved a sufficient condition for fast convergence at large β. This work introduces a slightly different notion of convergence time, and proves that convergence to the risk dominant equilibrium is fast for 'close-knit' families graphs. Namely, he defines (for δ a small positive constant)

$$\tau_+(G, \delta; \underline{h}) = \sup_{\underline{x}} \inf \left\{ t \geq 0 : \mathbb{P}_\beta^{\underline{x}}\{\sum_{i \in V} x_i(t) \geq (1-\delta)n\} \geq 1 - \delta \right\} . \tag{17}$$

Further, graph G is said to be '(r, v)-close-knit' if each vertex belongs to at least one set of vertices S such that $|S| \leq v$ and, for every $S' \subseteq S$:

$$d(S', S) \geq r \sum_{i \in S'} |N(i)| , \tag{18}$$

where $d(S', S)$ is the number of edges between a vertex in S' and a vertex in S. A family $\mathcal{F}$ of graphs is said to be close-knit if, for every $r \in (0, 1/2)$ there exists a $v = v(r)$ such that every graph in the family is $(r, v(r))$ close-knit.

Theorem A.1 (Young, 2006)**.** *Consider a symmetric* 2×2 *game with a risk-dominant equilibrium, and let* $\mathcal{F}$ *be a close-knit family of graphs. Then there exists* β_* *and* $\tau_*(\beta, \delta, v(\,\cdot\,))$ *such that, for any* $\beta > \beta_*$ *and any graph in the family*

$$\tau_+(G, \delta; \underline{h}) \leq \tau_*(\beta, \delta, v(\,\cdot\,)) . \tag{19}$$

Notice that the conclusions of this theorem are not directly comparable with our results, in that it provides a finite-β upper bound, but does not estimate the $\beta \to \infty$ behavior. Further, the definition of hitting time is slightly different from ours and from the one of [Ell93]. On the other hand, it is easy to use Theorem 5.3 to show that, for G belonging to a close-knit family $\tau_+(G; \underline{h}) = \exp\{\beta\Gamma_*(G) + o(\beta)\}$ with $\Gamma_*(G)$ upper bounded by a constant independent of the graph size. Indeed, if G is (r, v) close-knit with r close enough to $1/2$, then there exists a sequence $S_1, \dots, S_T \subseteq V$ such that $H(S_t) = \min_{S' \subseteq S_t} H(S') \leq 0$ and $|S_i| \leq v$. By flipping vertices along this sequence and using the submodularity of $H(\,\cdot\,)$, it follows that $\Gamma(F; \underline{h}^F) \leq v^2$.

B The convergence rate in terms of tilted cut

The second part of the proof of Theorem 4.1 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 4.1, Tilted cut). By monotonicity of Glauber dynamics $\Gamma_*(G;\underline{h}) \geq \Gamma_*(F;\underline{h}^F)$ for any induced subgraph $F \subseteq G$. Theorem 5.1 implies $\Gamma_*(F;\underline{h}^F) \geq \Delta(F;\underline{h}^F)$: 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 $\Gamma_*(G;\underline{h}) \geq \max_F \Delta(F;\underline{h}^F)$.

We need to prove $\Gamma_*(G;\underline{h}) \leq \Delta(F;\underline{h}^F)$ for at least one induced subgraph F. Fix F to be a subgraph which achieves the maximum in Eq. (8) (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 $\psi_0(\underline{x}) \geq 0$. 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 E.1 and H.1 imply that there exists $\Omega = \{x \in \mathcal{S} : \psi_0(\underline{x}) > b\}$, such that

$$\tau_+(F;\underline{h}^F) \leq C_n(1+\beta)\,\frac{\sum_{\underline{x}\in\Omega}\mu(\underline{x})}{\sum_{(\underline{x},\underline{y})\in\partial\Omega}\mu(\underline{x})p^+_\beta(\underline{x},\underline{y})}\,. \tag{20}$$

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

$$\Gamma_*(F;\underline{h}^F) \leq \min_{(S_1,S_2)\in\partial\Omega}\max_{i=1,2}[\mathrm{cut}(S_i, V\setminus S_i) - |S_i|_h] + o_\beta(1)\,. \tag{21}$$

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

C Isoperimetric functions, proof of Lemma 3.2

Proof. (Lemma 3.2). By Theorem 4.1, 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. (12) holds for any induced subgraph $F \subseteq \tilde{F}$.

Let $h_{\min} = \min_i h_i$ and $h_{\max} = \max_i h_i \leq h\Delta$. First notice that, for any U and for any k, there exists $S \subseteq U$ such that $|S| = k$ and

$$\mathrm{cut}(S, U\setminus S) - \frac{1}{4}|S|_h \leq \alpha h_{\min}^{-\gamma}|S|_h^\gamma - \frac{1}{4}|S|_h \leq A'(\alpha,\gamma)\, h_{\min}^{-\gamma/(1-\gamma)}\,, \tag{22}$$

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. (22)

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

Finally the cutwidth of any set S with $|S|_h \le L_2$ is upper bounded by $\alpha|S|^\gamma \log|S|$ (using [LR99] and Eq. (4)) which is at most $C = A''(\alpha,\gamma,h_{\rm max})\, h_{\rm min}^{-1/(1-\gamma)} \log\max(2, h_{\rm min}^{-1})$. The thesis thus follows by applying Theorem 5.3.

To prove the lower bound we use Theorem 4.1 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

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

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 4.1.

D Proof of Theorem 5.3

Proof. (Theorem 5.3). 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 $j \in V_t$, $(h^{V_t})_j = h_j + |N(j|_{V\setminus V_t}$.

Continue this process until no such set S can be found, and let $R_l = V_l$ be the residual set. Notice that, since $L_2 \ge h_{\rm max}$, we necessarily have $|R_l|_h < L_1$. By applying Eq. (12) to $F = G(V_t)$, we have

$$\mathrm{cut}(R_t, V_t\setminus R_t) \le |R_t|_{h^{V_t}} + L_1 \le |R_t|_{h^{V_t}} + |R_t|_h = |R_t|_{2h} + \mathrm{cut}(R_t, V\setminus V_t)\,. \tag{23}$$

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

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

□

D.1 Approximating tilted cut and tilted cutwidth

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 5.3 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 D.1. *If $G = (V, E)$ satisfies equation (12), 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 $L = L_1 + L_2 + C$*

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

E Hitting times for small β: proof of Lemma 5.1

We consider a general setting of Lemma 5.1: 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 $p^A_\beta(x, y) = p_\beta(x, y)$ if $x, y \in \mathcal{S} \setminus A$ and $p^A_\beta(x, y) = 0$ otherwise. Notice by reversibility the eigenvalues of p^A_β are real, and smaller than 1. We assume that p^A_β 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 E.1. *If $1 - \lambda_{0,A}$ is the largest eigenvalue of p^A_β, 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 $p^A_\beta(x, y)$, and $f(x)$ be the characteristic function of $\mathcal{S} \setminus A$. Then $\mathbb{P}_x \{T_A > t\} = P^t_A f(x)$, whence

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

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 5.1). Due to Lemma E.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

$$\lambda_{0,A} = \inf_\varphi \frac{\mathrm{Dir}(\varphi)}{\mathbb{E}(\varphi^2)}, \qquad \mathrm{Dir}(\varphi) \equiv \frac{1}{2} \sum_{x,y} \mu(x) p_\beta(x, y) (\varphi(x) - \varphi(y))^2 . \tag{24}$$

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. (9) 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. (9) 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. (24) to be the characteristic function of B. □

F Proof of Lemma 5.2

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 $l = i_j$

$$M(T_j, v_l - 1) = 2 \left[\text{cut}(T_j, \{1, 2, \cdots v_l - 1\}) - \text{cut}(T_j, \{v_l, v_l + 1, \cdots n\}) + \sum_{i \in T_j} h_i \right] \geq 0 \,. \tag{25}$$

One can write a similar equation $j = l$ by replacing $v_l - 1$ with k. Equation (11) 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

$$M(T \cup R, k) \geq \sum_{j=1}^{l-1} M(T_j, v_{i_j} - 1) + M(T_l, k) + M(R, 0) \geq 0. \tag{26}$$

□

G Proof of Lemma 5.4

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

$$||\nabla_G f||^2 = \sum_{(i,j)\in E} (f_i - f_j)^2 \geq \frac{\left(\sum_{(i,j)\in E} |f_i^2 - f_j^2|\right)^2}{\sum_{(i,j)\in E} (f_i + f_j)^2} \, . \tag{27}$$

The denominator is upper bounded by

$$4 \sum_{i\in V} |N(i)| \, f_i^2 \leq 4 \max \left| \frac{|N(i)|}{h_i} \right| \ \ ||f||_h^2 \, . \tag{28}$$

The argument in parenthesis at the numerator is instead equal to

$$\sum_{(i,j)\in E} \int_0^1 \left| \mathbb{I}(f_i^2 > z) - \mathbb{I}(f_j^2 > z) \right| \mathrm{d}z = \int_0^1 \mathrm{cut}(S_z, V \setminus S_z) \, \mathrm{d}z \tag{29}$$

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

$$\min_{z\in[0,1]} \frac{\mathrm{cut}(S_z, V \setminus S_z)}{|S_z|_h} \int_0^1 |S_z|_h \, \mathrm{d}z = \min_{z\in[0,1]} \frac{\mathrm{cut}(S_z, V \setminus S_z)}{|S_z|_h} \, ||f||_h \, . \tag{30}$$

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 $L_1 \leq |S|_w \leq L_2$. Further, form the above

$$\lambda \geq \frac{||\nabla_G f||^2}{||f||_h^2} \geq \frac{1}{4} \min \left| \frac{h_i}{|N(i)|} \right| \left\{ \frac{\mathrm{cut}(S, V \setminus S)}{|S|_h} \right\}^2 \tag{31}$$

which finishes the proof. □

H 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 H.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) $\psi_0(x) \geq 0$. Then there exists $b \geq 0$ such that, letting $B = \{x \in \mathcal{S} : \ \psi_0(x) > b\}$, we have*

$$\frac{1}{|\mathcal{S}|} \frac{\sum_{(x,y)\in\partial B} \mu(x) p_\beta(x,y)}{\sum_{x\in B} \mu(x)} \leq \lambda_{0,A} \leq \frac{\sum_{(x,y)\in\partial B} \mu(x) p_\beta(x,y)}{\sum_{x\in B} \mu(x)} \tag{32}$$

Proof. The upper bound follows immediately by substituting $\varphi(x) = \mathbb{I}(x \in B)$ in the variational principle (24).

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

$$\mathrm{Dir}(\psi_0) \ge \sum_{l=1}^{N} W(l)\, (\psi^{(l)} - \psi^{(l-1)})^2\,, \qquad W(l) \equiv \sum_{(x,y)\in\partial B_l} \mu(x) p_\beta(x, y)\,. \tag{33}$$

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

$$\mathbb{E}(\psi_0^2) = \sum_{i=0}^{N} M(i)\ (\psi^{(i)})^2 \le \sum_{l=1}^{N} \Big(\sum_{i=l}^{N} i\, M(i) \Big)\ (\psi^{(l)} - \psi^{(l-1)})^2\,. \tag{34}$$

Therefore

$$\lambda_{0,A} = \frac{\mathrm{Dir}(\psi_0)}{\mathbb{E}(\psi_0^2)} \ge \inf_{1\le l\le N} \frac{W(l)}{\sum_{i=l}^{N} i\, M(i)}\,, \tag{35}$$

which implies the thesis. $\square$