Supplementary Information for: The Spread of Innovations in Social Networks

Andrea Montanari* and Amin Saberi†

* Departments of Electrical Engineering and Statistics, Stanford University

†Department of Management Science and Engineering, Stanford University

We provide the supplementary information for the article "The Spread of Innovations in Social Networks", in the Proceedings of the National Academy of Sciences.

This document mainly contains the proofs of lemmas and theorems stated in the paper. In the end, we also give a short comparison between the results of this paper and a few well-known results in economics.

1 Proof of Theorem 3

For the sake of clarity, we split the proof of Theorem 3 into two parts: first, we present the characterization in terms of tilted cutwidth (i.e. the first identity in Eq. (9)). This is possible by relating the convergence time to the evolution of the potential function H. The characterization in terms of tilted cut (second identity in Eq. (9)) is then presented in Section 1.2.

1.1 Relating the rate of convergence to tilted cutwidth

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

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

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

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 of this lemma can be obtained by building on known results, for instance Theorem 6.38 in [22]. These however typically apply to exit times from local minima of $H(x)$. We provide a simple proof based on spectral arguments in Section 4.1.

Proof. (Theorem 3, Tilted cutwidth). Notice that Glauber dynamics satisfies the hypotheses of Lemma 6, 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. (10). We thus obtain $\tau_+ = \exp(\beta\max_{\underline{z}}\widetilde{\Gamma}_+(\underline{z}) + o(\beta))$ where

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

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. The next lemma proves that and finishes the proof of the first part of the Theorem. $\square$

Lemma 7. *Suppose z is a worst case starting point, i.e. it is a state that obtains the maximum in equation (10). Then there exists a monotone path, a path that only flips -1 to $+1$ from z to $\underline{+1}$, that obtains the minimum in equation (11).*

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

Let us start by examining the worst-case initial configuration. We claim that if B is such a configuration, i.e., $B \in \arg\max_S \min_{\omega: S \to V} G(\omega)$, 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,

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

In order to finish the proof, we consider two cases. The next lemma shows that these two cases are exhaustive.

Lemma 8. *One of the following two statements is correct: Case (I) There exists a subset $T' \subset T$ such that for all i, $M(T', i) \leq 0$; Case (II) $M(T \cup R, k) \geq 0$.*

Proof. 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 $l = i_j$

$$M(T_j, v_l - 1) = 2\left[\mathrm{cut}(T_j, \{1, 2, \cdots v_l - 1\}) - \mathrm{cut}(T_j, \{v_l, v_l + 1, \cdots n\}) + \sum_{i \in T_j} h_i\right] \geq 0\,. \tag{13}$$

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

$$M(T \cup R, k) \geq \sum_{j=1}^{l-1} M(T_j, v_{i_j} - 1) + M(T_l, k) + M(R, 0) \geq 0. \tag{14}$$

□

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 $G(\omega') \leq G(\omega)$. This is obvious because for every $i \leq k$, $H(S_i \setminus T') - H(S_i) = M(T', i) \leq 0$.

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

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

□

1.2 The convergence rate in terms of tilted cut

The second part of the proof of Theorem 3 exploits the fact that Glauber dynamics is monotone for the Ising model. In other words, 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, Tilted cut). By the monotonicity of Glauber dynamics $\Gamma_*(G; \underline{h}) \geq \Gamma_*(F; \underline{h}^F)$ for any induced subgraph $F \subseteq G$. Lemma 6 implies $\Gamma_*(F; \underline{h}^F) \geq \Delta(F; \underline{h}^F)$: indeed given a path $\omega = (S_0, S_1, \ldots, S_{|\omega|-1} = V)$ 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. (9) (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 theorem.

Lemmas 11 and 12 imply that there exists $\Omega = \{x \in \mathcal{S} : \psi_0(\underline{x}) > b\}$, such that

$$\tau_+(F; \underline{h}^F) \leq C_n(1 + \beta)\, \frac{\sum_{\underline{x} \in \Omega} \mu(\underline{x})}{\sum_{(\underline{x}, \underline{y}) \in \partial\Omega} \mu(\underline{x}) p^+_\beta(\underline{x}, \underline{y})}\,. \tag{15}$$

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

$$\Gamma_*(F; \underline{h}^F) \leq \min_{(S_1, S_2) \in \partial\Omega} \max_{i=1,2} [\mathrm{cut}(S_i, V \setminus S_i) - |S_i|_h] + o_\beta(1)\,. \tag{16}$$

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

2 The relationship with the isoperimetric functions

In order to prove Theorem 1, we have to relate $\Gamma^*(G)$ and in particular tilted cutwidth to the connectivity of the graph. The following Lemma as well as Lemma 2 mainly establish this goal.

Lemma 9. *Assume that, for some L_1, L_2, with $L_2 \geq h_{\max}$ and for every induced subgraph $F \subseteq G$, we have*

$$\min_{|S|_h \in [L_1, L_2]} [\mathrm{cut}(S, V(F) \setminus S) - |S|_{h^F}] \leq L_1 \,, \tag{17}$$

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 $+1$. Lemma 9 allows to replace single vertices, by 'blocks' as long as they have bounded size and bounded cutwidth.

Proof. 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 $L_1 \leq |S|_h \leq L_2$. 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 $L_2 \geq h_{\max}$, we necessarily have $|R_l|_h < L_1$. By applying Eq. (17) to $F = G(V_t)$, we have

$$\mathrm{cut}(R_t, V_t \setminus R_t) \leq |R_t|_{h^{V_t}} + L_1 \leq |R_t|_{h^{V_t}} + |R_t|_h = |R_t|_{2h} + \mathrm{cut}(R_t, V \setminus V_t) \,. \tag{18}$$

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 $1 \leq t < l$,

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

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

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

$\square$

2.1 Proof of Lemma 2

Proof. (Lemma 2). By Theorem 3, 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. (17) 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

$$\mathrm{cut}(S, U \setminus S) - \frac{1}{4}|S|_h \leq \alpha h_{\min}^{-\gamma} |S|_h^{\gamma} - \frac{1}{4}|S|_h \leq A'(\alpha, \gamma)\, h_{\min}^{-\gamma/(1-\gamma)}\,, \tag{19}$$

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. (19)

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

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. (5)) 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 Lemma 9.

But before presenting the proof, let us start with a simple and intuitive argument to show that the dynamics will be very slow to converge when the underlying graph is an expander. Suppose the starting point is the configuration in which everyone is taking the action -1. Obviously, before reaching the state all $+1$, the dynamics has to go through at least one state in which half of the vertices are taking the action $+1$ and half of the vertices are taking the action -1. Because of the high expansion of the graph, such a configuration will have a very high value of H. Applying Lemma 6, we are done. The calculations below use Theorem 3 directly to make the above intuition more precise.

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

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

which implies the thesis. □

3 Proof of Theorem 1

In this section we finally derive the rate of convergence for specific graph families proving Theorem 1. In most cases, our proof is a simple application of Lemma 9. However, for d-dimensional graphs, we have to estimate the isoperimetric function. 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

$$||f||_w^2 \equiv \sum_{i \in V} w_i\, f_i^2\,, \qquad ||\nabla_G f||^2 \equiv \sum_{(i,j) \in E} |f_i - f_j|^2\,. \tag{20}$$

We then have the following generalization of Cheeger's inequality.

Lemma 10. *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 $L_1 \leq |S|_w \leq L_2$*

$$\mathrm{cut}(S, V \setminus S) \leq \sqrt{4\lambda \max_{i\in V}\{|N(i)|/h_i\}}\ \ |S|_h\,. \tag{21}$$

The proof of this Lemma is deferred to the end of this section.

Proof. (Theorem 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 [24, 23]. The thesis follows from Lemma 2.

d-dimensional networks. We need to prove that, for each induced subgraph G', $\Gamma(G'; \underline{h}^{G'}) = O(1)$. By Lemma 9, 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 $h^F_i \geq h_i$, 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 $||\nabla_G f||^2 \leq \lambda ||f||^2$ 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

$$\varphi(x) = \left[1 - \frac{d_{\mathrm{Eucl}}(x, \mathcal{C})}{\ell}\right]_+ . \tag{22}$$

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

$$\begin{aligned} ||\nabla_G f||^2 &\leq \left(\frac{K}{\ell}\right)^2 \sum_{i\in V} |N(i)|\, \mathbb{I}(\xi_i \in \cup_{j=1}^{3^d-1}\mathcal{C}_j) \leq \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|N(i)|/h_i\} \sum_{i\in V} h_i\, \mathbb{I}\Big(\xi_i \in \cup_{j=1}^{3^d-1}\mathcal{C}_j\Big) \\ &\leq 3^d \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|N(i)|/h_i\} \sum_{i\in V} h_i\, \mathbb{I}\Big(\xi_i \in \mathcal{C}_0\Big) \leq 3^d \left(\frac{K}{\ell}\right)^2 \max_{i\in V}\{|N(i)|/h_i\} ||f||_h^2\,. \end{aligned} \tag{23}$$

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

Small world networks with $r < d$. It is proved in [25] that these graphs are with high probability expanders. The thesis follows from Lemma 2.

□

3.1 Proof of Lemma 10

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

$$||\nabla_G f||^2 = \sum_{(i,j) \in E} (f_i - f_j)^2 \geq \frac{\left(\sum_{(i,j)\in E} |f_i^2 - f_j^2|\right)^2}{\sum_{(i,j)\in E} (f_i + f_j)^2} . \tag{24}$$

The denominator is upper bounded by

$$4 \sum_{i \in V} |N(i)|\, f_i^2 \leq 4 \max \left| \frac{|N(i)|}{h_i} \right| \;\; ||f||_h^2 . \tag{25}$$

The argument in parenthesis at the numerator is instead equal to

$$\sum_{(i,j)\in E} \int_0^1 \left|\mathbb{I}(f_i^2 > z) - \mathbb{I}(f_j^2 > z)\right| \mathrm{d}z = \int_0^1 \mathrm{cut}(S_z, V \setminus S_z)\, \mathrm{d}z \tag{26}$$

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

$$\min_{z \in [0,1]} \frac{\mathrm{cut}(S_z, V \setminus S_z)}{|S_z|_h} \int_0^1 |S_z|_h \, \mathrm{d}z = \min_{z \in [0,1]} \frac{\mathrm{cut}(S_z, V \setminus S_z)}{|S_z|_h} \; ||f||_h . \tag{27}$$

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

$$\lambda \geq \frac{||\nabla_G f||^2}{||f||_h^2} \geq \frac{1}{4} \min \left| \frac{h_i}{|N(i)|} \right| \left\{ \frac{\mathrm{cut}(S, V \setminus S)}{|S|_h} \right\}^2 \tag{28}$$

which finishes the proof. □

4 Eigenvectors and barriers

For the next two Lemmas, consider a general setting of: a discrete time Markov chain with state space $\mathcal{S}$, transition probabilities $p_\beta(x,y)$, reversible with respect to the stationary distribution $\mu(x)$.

4.1 Hitting times for large β: proof of Lemma 6

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 [8]: we nevertheless propose an elementary proof.

Lemma 11. *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 6). Due to Lemma 11, 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

$$\lambda_{0,A} = \inf_\varphi \frac{\mathrm{Dir}(\varphi)}{\mathbb{E}(\varphi^2)}\,, \qquad \mathrm{Dir}(\varphi) \equiv \frac{1}{2}\sum_{x,y}\mu(x)p_\beta(x,y)(\varphi(x)-\varphi(y))^2\,. \tag{29}$$

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 [10, 9], one obtains that $\lambda_{0,A} \ge 1/\max_{x,y} C(x,y;\omega)$, 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. (29) to be the characteristic function of B. $\square$

Like the last Lemma, consider a general Markov chain with state space $\mathcal{S}$, and let $A\subseteq\mathcal{S}$ a subset of configurations.

Lemma 12. *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*

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

Proof. The upper bound follows immediately by substituting $\varphi(x)=\mathbb{I}(x\in B)$ in the variational principle (29).

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

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

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

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

Therefore

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

which implies the thesis. $\square$

5 Non-reversible and Synchronous Dynamics

5.1 Proof of Proposition 4

For the sake of simplicity, we shall focus here on asynchronous dynamics. It is straightforward to analyze synchronous dynamics using the same technique.

Let S_t denote the subset of vertices adopting strategy $+1$ at time t, and denote its size by $X_t=|S_t|$. Clearly $X_0=0$, $X_t\geq 0$ and $|X_{t+1}-X_t|\leq 1$. The proof is mainly based on showing that for $X_t\leq n\delta$, X_t is stochastically dominated by a birth-and-death chain with negative drift.

For $b \in \{0, \dots, k\}$, call $n_-(b)$ the number of vertices $i \in V \backslash S_t$, with b neighbors in S_t. Analogously, let $n_+(b)$ be the number of vertices $i \in S_t$, with b neighbors in S_t. If we let $q \equiv \lfloor k(1-h)/2 \rfloor$ (this is the maximum number of $+1$ neighbors such that the best response is still -1), it is elementary to show

$$\begin{aligned}
\mathbb{P}\{X_{t+1)} = \ell + 1 | S_t; X_t = \ell\} &\le \frac{1}{n} \sum_{b=q+1}^{k} n_-(b) + e^{-2\beta}\,,\\
\mathbb{P}\{X_{t+1)} = \ell - 1 | X_t = \ell\} &\ge \frac{1}{n} \sum_{b=0}^{q} n_+(b) - e^{-2\beta}\,.
\end{aligned}$$

In both cases, the first term is the probability of picking a vertex for which the best response corresponds to changing strategy from -1 to $+1$ (first bound) or from $+1$ to -1 (second bound). The term $e^{-2\beta}$ bounds the probability of non-best-response choices.

Since the degree of all vertices is k, we have $\sum_{b=1}^{k} n_-(b)b \le k\ell$. Further assuming $X_t = \ell \le \delta n$, we also have $\sum_{b=1}^{k} n_-(b) \ge \lambda\ell$. It follows that

$$q \sum_{b=q+1}^{k} n_-(b) \le \sum_{b=1}^{k} n_-(b)(b-1) \le (k-\lambda)\ell\,,$$

whence

$$\mathbb{P}\{X_{t+1} = \ell + 1 | S_t; X_t = \ell\} \le \left(\frac{k-\lambda}{q}\right)\frac{\ell}{n} + e^{-2\beta}\,.$$

A similar argument yields, for all $1 \le \ell \le n\delta$,

$$\mathbb{P}\{X_{t+1} = \ell - 1 | S_t; X_t = \ell\} \ge \left(1 - \frac{k-\lambda}{q+1}\right)\frac{\ell}{n} - e^{-2\beta}\,.$$

We therefore have

$$\mathbb{E}\{X_{t+1} - X_t | S_t; X_t = \ell\} - \left(1 - 2\frac{k-\lambda}{q}\right)\frac{\ell}{n} + 2e^{-2\beta}\,.$$

where, recalling that $\lambda \ge 4k/5$, we have for all h small enough

$$\alpha = 1 - 2\frac{k-\lambda}{q} \ge 1 - \frac{2k}{5\lfloor k(1-h)/2 \rfloor} > 0\,.$$

The proof is completed with the following general Lemma:

Lemma 13. *Let $\{X_t\}_{t\in\mathbb{N}}$ be a sequence of non-negative random variables adapted to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ such that*

$$\mathbb{E}[X_{t+1} | \mathcal{F}_t] \le (1 - \alpha/n)\, X_t + \epsilon\,, \tag{34}$$

$|X_{t+1} - X_t| \le 1$, and $X_0 = 0$. If $T_* = \inf\{\, t \ge 0 : \ \ X_t \ge n\delta\}$ and $\epsilon < \alpha\delta$, then there exists $K > 0$ such that, with high probability, $T_* \ge e^{K\,n}$. Further any K can be taken to be any number smaller than $(\delta\alpha - \epsilon)/\alpha(1 + \alpha + \epsilon)^2$.

Proof. Define

$$Z_t = \left(1 - \frac{\alpha}{n}\right)^{-t} X_t + \frac{n\epsilon}{\alpha}\left[1 - \left(1 - \frac{\alpha}{n}\right)^{-t}\right]. \tag{35}$$

It is then easy to check that $\{Z_t\}_{t\in\mathbb{N}}$ is submartingale with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$, and $Z_0 = 0$. Further

$$|Z_{t+1} - Z_t| \le (1 + \alpha + \epsilon)\left(1 - \frac{\alpha}{n}\right)^{-t}. \tag{36}$$

The thesis thus follows from

$$\mathbb{P}\{T_* \le e^{nk}\} \le \sum_{t=0}^{\exp(nK)} \mathbb{P}\{X_t \ge n\delta\} \le \sum_{t=0}^{\exp(nK)} \mathbb{P}\{Z_t \ge n(\delta - \epsilon/\alpha)(1-\alpha/n)^{-t}\}, \tag{37}$$

and applying Azuma-Hoeffding inequality to the latter. □

5.2 Proof of Proposition 5

Given a vertex i in the grid, let $\mathsf{B}_i(\ell)$ be the cube of side $(2\ell+1)$ centered at i. Following [27], consider the *censored* dynamics in which vertices $j \in V \setminus \mathsf{B}_i(\ell)$ never change strategy and are frozen to the value $x_j = -1$. Denoting by $\mathbb{P}_{i,\ell}\{\,\cdot\,\}$ probabilities under this modified dynamics, it follows from the monotonicity of the transition rates that

$$\mathbb{P}\{x_i(t) = +1\} \ge \mathbb{P}_{i,\ell}\{x_i(t) = +1\},$$

Using union bound over i, it is therefore sufficient to show that for some $t \le \exp\{c\beta + o(\beta)\}$

$$\mathbb{P}_{i,\ell}\{x_i(t) = +1\} \ge 1 - \frac{1}{10n}. \tag{38}$$

By the same argument as in [11, Theorem 1], the stationary measure $\mu_\beta(\underline{x})$ for the censored dynamics satisfies

$$\mu_\beta(\underline{x} = \underline{+1}) = 1 - O(e^{-2\beta}), \tag{39}$$

for all ℓ that is large enough. Fix such an ℓ. It follows condition (38) is satisfied in the stationary state for all $\beta \ge C\log n$ (with C a big enough constant). It is therefore sufficient to show that the censored dynamics has mixing time bounded by $\exp\{\beta c + o(\beta)\}$. This is immediate, since the censored dynamics is an irreducible Markov chain over a space with $2^{(2\ell+1)^d}$ states so its spectral gap is at least $\exp\{-A\beta(2\ell+1)^d\}$ for some numerical constant A.

6 Comparison with results in the economics literature

Ellison [11] 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 $\mathrm{sign}(h_i + \sum_{j\in N(i)} x_j)$ 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 6 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 3 for the [11] 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. (8) 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 [27] 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)

$$\tau_+(G, \delta; \underline{h}) = \sup_{\underline{x}} \inf \left\{ t \geq 0 : \mathbb{P}^{\underline{x}}_{\beta}\{\sum_{i \in V} x_i(t) \geq (1-\delta)n\} \geq 1 - \delta \right\} . \tag{40}$$

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 $S' \subseteq S$:

$$d(S', S) \geq r \sum_{i \in S'} |N(i)| , \tag{41}$$

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 14 (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*

$$\tau_+(G, \delta; \underline{h}) \leq \tau_*(\beta, \delta, v(\,\cdot\,)) . \tag{42}$$

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 [11]. On the other hand, it is easy to use Lemma 9 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 $\Gamma(F; \underline{h}^F) \leq v^2$.

7 Alternative convergence criteria

The convergence time $\tau_+(G; \underline{h})$ was defined as the typical time for $\underline{x}$ to converge to the all-$(+1)$ configuration. It is important to investigate whether our conclusions are robust with respect to modifications of this notion of convergence. To this purpose, we defined in the main text $\tau_\epsilon(G; \underline{h})$ to be the typical time for a fraction $(1 - \epsilon)$ of the nodes to adopt the $+1$ strategy.

Formally, define T_ϵ to be the hitting time for the event $\{\sum_{i \in V} x_i(t) \geq n(1 - 2\epsilon)\}$. In analogy with $\tau_+(G; \underline{h})$, we let

$$\tau_\epsilon(G; \underline{h}) = \sup_{\underline{x}} \inf \left\{ t \geq 0 : \mathbb{P}^{\underline{x}}_{\beta}\{T_\epsilon \geq t\} \leq e^{-1} \right\} . \tag{43}$$

In particular $\tau_+(G;\underline{h}) = \tau_{\epsilon=0}(G;\underline{h})$. Further notice that this definition is only marginally different from the one by Young [27], provided in Eq. (40). The main difference is that we define typicality always with respect to the same probability threshold e^{-1} while Young uses instead the same small number ϵ.

It turns out that most of our results are robust, in the sense that the $\tau_\epsilon(G;\underline{h}) = \exp\{2\Gamma_*(G;\underline{h},\epsilon)\beta + o(\beta)\}$ where, for $\Gamma_*(G;\underline{h},\epsilon)$ differs from $\Gamma_*(G;\underline{h}) = \Gamma_*(G;\underline{h},0)$ at most by a multiplicative constant, for all ϵ small enough. More explicitly, one can find a $\epsilon_* > 0$ (and independent of n) such that this hold for $\epsilon < \epsilon_*$. Notice that this claim corresponds to the correct notion of robustness: neglecting a small positive fraction of the agents does not change significantly the convergence time. Also notice that Figures 3 and 4 in the main text clearly illustrate this point. Complete convergence to the risk dominant equilibrium takes place in a time that is not larger than twice $\tau_{1/2}(G;\underline{h})$ (the time at which strategy $+1$ becomes majority).

In order to prove the robustness of our rigorous results, it is convenient to to state a theorem that will play the role of 3.

Theorem 15. *Let $\tau_A(G,\underline{h})$ be the typical hitting time to $A = \{\underline{x} : \underline{x} \leq H_0\}$, defined as in (43) with T_ϵ replaced by T_A. For reversible asynchronous dynamics we have $\tau_A(G;\underline{h}) = \exp\{2\beta\Gamma_A(G;\underline{h}) + o(\beta)\}$, where*

$$\Gamma_A(G;\underline{h}) = \max_{\underline{z}\in\{+1,-1\}^V} \Gamma_{A,G}(\underline{z};\underline{h}). \tag{44}$$

and, for any $\underline{z} \in \{+1,-1\}^V$,

$$\Gamma_{A,G}(\underline{z};\underline{h}) = \min_{\omega:\underline{z}\to A} \max_{t\leq|\omega|-1} \left[H(\omega_t) - H(\underline{z})\right] . \tag{45}$$

Proof. The proof follows immediately from Lemma 6, using the fact that, for any allowed transition $\underline{x} \to \underline{y}$, $H(\underline{x}) + V(\underline{x},\underline{y}) = \max(H(\underline{x}), H(\underline{y}))$. □

Notice that Theorem 15 is neither as elegant nor as easy to use as Theorem 3: it does not address directly the convergence time $\tau_\epsilon(G,\underline{h})$, and does not provide a purely graph-theoretical characterization. Nevertheless, it is sufficient for establishing the robustness of Theorem 1 in main article, as stated below.

Theorem 16 (Robust version of Theorem 1 in the article). *Consider the reversible asynchronous noisy best response dynamics on graph G and let $\tau_\epsilon(G) = \tau_\epsilon(G;\underline{h})$ with $h_i = h|N(i)|$ be the typical convergence time to $(1-\epsilon)$ majority. As $\beta \to \infty$, we have $\tau_\epsilon(G) = \exp\{2\beta\Gamma(G;\epsilon) + o(\beta)\}$. Further, for each the graph sequences below, there exist $\epsilon_* > 0$ such that the following happens for $\epsilon < \epsilon_*$:*

(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, $\Gamma(G;\epsilon) = \Omega(n)$.

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

(iii) If G is a small world network with $r \geq d$, then $\Gamma(G;\epsilon) = \Omega(\log(1/\epsilon)/\log\log(1/\epsilon))$.

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

The proof is completely analogous to the one of Theorem 1 and we will only describe the main differences. Notice that in the above statement only the estimate at point (iii) changed with respect to the complete agreement case. Further, this changed only through the replacement of $\log n$ with $\log(1/\epsilon)$, which is rather natural.

Before proving Theorem 16, we state and prove a generalization of Lemma 2.

Lemma 17 (Robust version of Lemma 2). *For a graph G with degree bounded by Δ, assume there exist disjoint subsets $U \subseteq V(G)$, such that for $i \in U$, $|N(i) \cap (V \setminus U)| \le M$, and the subgraph induced by U is a (δ, λ) expander, i.e., for every $k \le \delta|U|$,*

$$\phi(G_U; k) \ge \lambda. \tag{46}$$

Let $T_{U,\epsilon}$ be the hitting time for the event $\{\sum_{i \in U} x_i(t) \ge (1-\delta)|U|\}$ and

$$\tau_{U,\epsilon}(G) = \sup_{\underline{x}} \inf \left\{ t \ge 0 : \mathbb{P}^{\underline{x}}_{\beta}\{T_{U,\epsilon} \ge t\} \le e^{-1} \right\}. \tag{47}$$

Then there exist $\epsilon_* = \epsilon_*(\lambda, h, \delta, \Delta, M)$ such that for all $\epsilon < \epsilon_*$, we have $\Gamma_U(G; \epsilon) \ge (\lambda - h\Delta - M)\lfloor \delta|U| \rfloor$.

Proof. By the same monotonicity arguments as in Lemma 2, we can reduce the problem to the case in which all the vertices in $V \setminus U$ are frozen to $+1$. This is equivalent (up to renaming of the vertex set) to assuming $V = U$, and increasing the bias from $\underline{h}$ to uh^F.

We claim that, for all ξ that is small enough, we can choose $\epsilon_1(\xi)$, $\epsilon_2(\xi)$ (depending on λ, Δ,... but not on the graph) such that

$$\{\underline{x} : \sum_{i \in U} V_i \ge (1-2\epsilon_1)|V|\} \subseteq A(\xi) \equiv \{\underline{x} : H(\underline{x}) \le H(\underline{+1}) + \xi|V|\} \subseteq \{\underline{x} : \sum_{i \in V} x_i \ge (1-2\epsilon_2)|V|\}. \tag{48}$$

Further $\epsilon_{1,2}(\xi) \to 0$ as $\xi \to 0$. The first inclusion follows from the bounded degree and bounded h_i property. The second from the expansion and non-negativity of h_i (the latter argument was for instance already used in the proof of Proposition 4).

Therefore we can lower bound $\tau_\epsilon(G; \underline{h})$ by some $\tau_{A(\xi)}(G; \underline{h})$ for a suitable choice of $\xi(\epsilon) > 0$. For the latter we can use Theorem 15, which implies $\tau_{A(\xi)}(G; \underline{h}) \ge \exp\{\beta \Gamma_{A(\xi),G}(\underline{-1}; \underline{h}) + o(\beta)\}$. Let $\Omega \subseteq 2^V$ be a collection of subsets of V, such that $\emptyset \in \Omega$, $V \setminus A(\xi) \notin \Omega$. Using notations already introduced in the main text, we have the following lower bound in terms of cuts

$$\Gamma_{A(\xi),G}(\underline{-1}; \underline{h}) \ge \min_{(S_1,S_2) \in \partial\Omega} \max_{i=1,2} \left[\mathrm{cut}(S_i, V \setminus S_i) - |S_i|_h \right], \tag{49}$$

At this point the thesis follows by using the same set Ω that is already used in the proof of Lemma 2. (One has to simply check that $V \setminus A(\xi) \notin \Omega$ via (48).) □

Proof of Theorem 16. Clearly $\tau_\epsilon(G; \underline{h}) \le \tau_\delta(G; \underline{h})$ because the event $\{\sum_{i \in V} x_i(t) \ge (1-2\epsilon)n\}$ necessarily precedes the one $\{x_i(t) = +1 \text{ for all } i \in V\}$. Hence claim (ii) is immediate from the analogous claim on $\tau_+(G)$.

Claims (i) and (iv) follow from appropriate construction of expander subgraphs that have vertex set $U = V$. The constructions remain the same as in the proof of Theorem 1 in the main text. The resulting lower bounds on mixing time are robust by Lemma 17.

Finally, claim (iii) can be proved by repeating the same construction as in Theorem 1 in the main text, but keeping ℓ bounded. More precisely, one can construct $\Omega(n\epsilon)$ disjoint subgraphs with volume $\ell^d = \Omega(\log(1/\epsilon)/\log\log(1/\epsilon))$. By Lemma 17, it takes $\exp(C\ell^d)$ time before in any of them the majority switches to $+1$. This implies the desired lower bound. □

Additional References

[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.

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