# How Homophily Affects the Speed of Learning and Best-Response Dynamics

# Online Appendices

Benjamin Golub    Matthew O. Jackson

## Appendix 1 Proofs of Results

### A Islands Homophily

We begin by proving the fact that spectral homophily is equal to the more simply defined islands homophily in the special case of the islands model.

Proposition 1. If $(\mathbf{P}, \mathbf{n})$ is an islands network with parameters $(m, p_s, p_d)$, then

$$h^{\text{islands}}(m, p_s, p_d) = h^{\text{spec}}(\mathbf{P}, \mathbf{n}).$$

**Proof of Proposition 1**: Note

$$\mathbf{F}(\mathbf{P}, \mathbf{n}) = \frac{p_d}{p_s + (m-1)p_d}\mathbf{E}_m + \frac{p_s - p_d}{p_s + (m-1)p_d}\mathbf{I}_m,$$

where $\mathbf{E}_m$ denotes the $m$-by-$m$ matrix of ones and $\mathbf{I}_m$ denotes the $m$-by-$m$ identity matrix. The eigenvalues of this matrix can be computed directly. The only nonzero eigenvalue of the first matrix is

$$\frac{mp_d}{p_s + (m-1)p_d}$$

with multiplicity 1; and adding

$$\frac{p_s - p_d}{p_s + (m-1)p_d}\mathbf{I}_m$$

just shifts all the eigenvalues by adding to them the constant multiplying the identity. Thus, the second-largest eigenvalue of $\mathbf{F}(\mathbf{P}, \mathbf{n})$ (after the eigenvalue 1) is

$$\frac{p_s - p_d}{p_s + (m-1)p_d}.$$

Theorem 2 gives the convergence in probability of $\mathbf{T}(\mathbf{A}(\mathbf{P}, \mathbf{n}))$ to this expression. Simple algebra shows that this is the same as $h^{\text{islands}}(n)$. ∎

## B The Main Technical Results

For the main results, we begin by fixing some notation and reviewing some important background results. Given a vector $\mathbf{s}$, define

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{s}} = \sum_i (v_i w_i) s_i.$$

This is just the Euclidean dot product weighted by the entries of the vector $\mathbf{s}$. Note also that the weighted norm introduced in Section II.D.4 can be written as $\|\mathbf{v}\|_{\mathbf{s}}^2 = \langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{s}} = \sum_i v_i^2 s_i$.

For any stochastic matrix $\mathbf{T}$ (a nonnegative matrix in which every row sums to 1), let

$$1 = \lambda_1(\mathbf{T}), \ldots, \lambda_n(\mathbf{T})$$

denote the eigenvalues of $\mathbf{T}$ ordered from greatest to least by magnitude.[1] Recall that $\mathbf{s}(\mathbf{A})$ is defined by $s_i(\mathbf{A}) = d_i(\mathbf{A})/D(\mathbf{A})$, the degree of agent $i$ divided by the sum of degrees in the network. In the proof of Lemma 2, we use the fact that $\mathbf{T}(\mathbf{A})$ is self-adjoint relative to $\langle \cdot, \cdot \rangle_{\mathbf{s}(\mathbf{A})}$. That is, for every $\mathbf{v}$ and $\mathbf{w}$, we have

$$\langle \mathbf{T}(\mathbf{A})\mathbf{v}, \mathbf{w} \rangle_{\mathbf{s}(\mathbf{A})} = \langle \mathbf{v}, \mathbf{T}(\mathbf{A})\mathbf{w} \rangle_{\mathbf{s}(\mathbf{A})},$$

which is immediate to check. This implies, by a standard theorem from linear algebra, that $\mathbf{T}(\mathbf{A})$ is diagonalizable, and, moreover, that there is a basis of eigenvectors orthogonal to each other under $\langle \cdot, \cdot \rangle_{\mathbf{s}(\mathbf{A})}$.

With these basic observations in hand, we prove Lemma 2.

Lemma 2. Let $\mathbf{A}$ be connected, $\lambda_2(\mathbf{T}(\mathbf{A}))$ be the second-largest eigenvalue in magnitude of $\mathbf{T}(\mathbf{A})$, and $\underline{s} := \min_i d_i(\mathbf{A})/D(\mathbf{A})$ be the minimum relative degree. If $\lambda_2(\mathbf{T}(\mathbf{A})) \neq 0$, then for any $0 < \varepsilon \leq 1$:

$$\left\lfloor \frac{\log(1/(2\varepsilon)) - \log(1/\underline{s}^{1/2})}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}))|)} \right\rfloor \leq \mathrm{CT}(\varepsilon; \mathbf{A}) \leq \left\lceil \frac{\log(1/\varepsilon)}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}))|)} \right\rceil.$$

If $\lambda_2(\mathbf{T}) = 0$, then for every $0 < \varepsilon < 1$ we have $\mathrm{CT}(\varepsilon; \mathbf{A}) = 1$.

**Proof of Lemma 2**: In the proof, we fix $\mathbf{A}$ and drop it as an argument; we also drop the argument $\mathbf{T}$ of the eigenvalues, as it is fixed throughout.

We first show that, if $\lambda_2 \neq 0$, then

$$\mathrm{CT}(\varepsilon) \leq \left\lceil \frac{\log(1/\varepsilon)}{\log(1/|\lambda_2|)} \right\rceil.$$

Take any $\mathbf{b} \in [0,1]^n$, which should be thought of as the initial beliefs $\mathbf{b}(0)$. Let $\mathbf{U}_i$ be the projection onto the eigenspace of $\mathbf{T}$ corresponding to $\lambda_i$. Note that under $\langle \cdot, \cdot \rangle_{\mathbf{s}}$, these eigenspaces are orthogonal, as mentioned above. Define $\mathbf{U} = \sum_{i=2}^n \mathbf{U}_i$. This is the projection *off* the eigenspace corresponding to $\lambda_1 = 1$. Recall also that $\mathbf{U}_1 = \mathbf{T}^\infty$ (see Meyer [2000] for

[1] Recall that 1 is a largest eigenvalue of any stochastic matrix. See Meyer (2000) for details.

details). Then:

$$
\begin{aligned}
\|(\mathbf{T}^t - \mathbf{T}^\infty)\mathbf{b}\|_{\mathbf{s}}^2 &= \left\| \sum_{i=2}^{n} \lambda_i^t \mathbf{U}_i \mathbf{b} \right\|_{\mathbf{s}}^2 && \text{spectral theorem applied to } \mathbf{T} \\
&= \sum_{i=2}^{n} |\lambda_i|^{2t} \|\mathbf{U}_i \mathbf{b}\|_{\mathbf{s}}^2 && \text{orthogonality of the spectral projections} \\
&\le |\lambda_2|^{2t} \sum_{i=2}^{n} \|\mathbf{U}_i \mathbf{b}\|_{\mathbf{s}}^2 \\
&= |\lambda_2|^{2t} \left\| \sum_{i=2}^{n} \mathbf{U}_i \mathbf{b} \right\|_{\mathbf{s}}^2 && \text{orthogonality of the spectral projections} \\
&= |\lambda_2|^{2t} \|\mathbf{U}\mathbf{b}\|_{\mathbf{s}}^2 && \text{definition of } \mathbf{U} \\
&\le |\lambda_2|^{2t} \|\mathbf{b}\|_{\mathbf{s}}^2 && \text{projections are contractions} \\
&\le |\lambda_2|^{2t} \sum_{i=1}^{n} s_i && \mathbf{b} \in [0,1]^n \text{ and definition of } \langle \cdot, \cdot \rangle_{\mathbf{s}} \\
&= |\lambda_2|^{2t} && \text{entries of } \mathbf{s} \text{ sum to 1.}
\end{aligned}
$$

Assume $\lambda_2 \neq 0$. Under this assumption, if

$$t \ge \frac{\log(1/\varepsilon)}{\log(1/|\lambda_2|)},$$

then

$$\|(\mathbf{T}^t - \mathbf{T}^\infty)\mathbf{b}\|_{\mathbf{s}} \le \varepsilon,$$

from which the bound follows upon observing that $\mathrm{CT}(\varepsilon)$ must be an integer. The above calculations also show that when the second eigenvalue is identically 0, then consensus time must be 1.

Now we show that, if $\lambda_2 \neq 0$, then

$$\left\lfloor \frac{\log(\underline{s}^{1/2}/(2\varepsilon))}{\log(1/|\lambda_2|)} \right\rfloor \le \mathrm{CT}(\varepsilon).$$

Let $\mathbf{w}$ be an eigenvector of $\mathbf{T}$ corresponding to $\lambda_2$, scaled so that $\|\mathbf{w}\|_{\mathbf{s}}^2 = \underline{s}/4$. Then the maximum entry of $\mathbf{w}$ is at most $1/2$ and the minimum entry is at least $-1/2$. Consequently, if we let $\mathbf{e}$ denote the column vector of ones and define $\mathbf{b} = \mathbf{w} + \mathbf{e}/2$, then $\mathbf{b} \in [0,1]^n$. Now, using the fact that $\mathbf{e}$ is a right eigenvector corresponding to $\lambda_1 = 1$ and spectral projections are orthogonal, it follows that:

$$
\begin{aligned}
\|(\mathbf{T}^t - \mathbf{T}^\infty)\mathbf{b}\|_{\mathbf{s}}^2 &= |\lambda_2|^{2t} \|\mathbf{U}_2 \mathbf{w}\|_{\mathbf{s}}^2 \\
&= |\lambda_2|^{2t} \|\mathbf{w}\|_{\mathbf{s}}^2 \\
&= \frac{\underline{s}}{4} |\lambda_2|^{2t}.
\end{aligned}
$$

Therefore, if

$$t \leq \frac{\log(\underline{s}^{1/2}/(2\varepsilon))}{\log(1/|\lambda_2|)},$$

then

$$\|(\mathbf{T}^t - \mathbf{T}^\infty)\mathbf{b}\|_{\mathbf{s}} \geq \varepsilon,$$

from which the remaining bound follows upon observing that $\mathrm{CT}(\varepsilon)$ must be an integer. ∎

We can tighten this bound under the assumptions of Definition 3.

**Proposition A.5.** Consider a sufficiently dense sequence of multi-type random networks satisfying the three regularity conditions in Definition 3. The regularity conditions imply that there exist constants $\alpha$ and $\beta$ such that, for high enough $n$, (i) $\min_k n_k/n \geq \alpha > 0$; and (ii) $[\min_k d_k(\mathbf{P},\mathbf{n})]/[\max_k d_k(\mathbf{P},\mathbf{n})] \geq \beta > 0$. Then, for any $\delta > 0$, for high enough $n$, with probability at least $1-\delta$

$$\left\lfloor \frac{\log(1/\sqrt{8}\varepsilon) - \log(1/\sqrt{\alpha\beta})}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P},\mathbf{n})))|)} \right\rfloor - 1 \leq \mathrm{CT}(\varepsilon; \mathbf{A}(\mathbf{P},\mathbf{n})) \leq \left\lceil \frac{\log(1/\varepsilon)}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P},\mathbf{n})))|)} \right\rceil.$$

Thus, for small $\varepsilon$, the consensus time $\mathrm{CT}(\varepsilon;\mathbf{A}(\mathbf{P},\mathbf{n}))$ is approximately proportional to $\frac{\log(1/\varepsilon)}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P},\mathbf{n})))|)}$. Essentially the same proof establishes that, under the same assumptions,

$$\left\lfloor \frac{\log(1/\sqrt{8}\varepsilon) - \log(1/\sqrt{\alpha\beta})}{\log(1/|h^{\mathrm{spec}}(\mathbf{P},\mathbf{n})|)} \right\rfloor - 1 \leq \mathrm{CT}(\varepsilon; \mathbf{A}(\mathbf{P},\mathbf{n})) \leq \left\lceil \frac{\log(1/\varepsilon)}{\log(1/|h^{\mathrm{spec}}(\mathbf{P},\mathbf{n})|)} \right\rceil.$$

The proof of the proposition appears below, after the proof of Proposition A.6, which introduces key machinery.

## The Representative Agent Theorem

Theorem 2 and Proposition A.5 require related machinery, which we will develop and apply in this section. First, we introduce some notation.

Throughout the appendix, we often drop the arguments $(\mathbf{P},\mathbf{n})$ on the matrix $\mathbf{A}$ and other random matrices obtained from it, keeping in mind that these are random objects that depend on the realization of the multi-type random network $\mathbf{A}(\mathbf{P},\mathbf{n})$. Let $\mathbf{D}(\mathbf{A})$ denote the diagonal matrix whose $(i,i)$ entry is $d_i(\mathbf{A})$, the degree of agent $i$. Let $\mathbf{R}$ be the $n$-by-$n$ matrix given by $R_{ij} = P_{k\ell}$ if $i \in N_k$, $j \in N_\ell$. The expected degree of node $i$ is $w_i := \sum_j R_{ij}$. We let $w_{\min} = \min_i w_i$ be the minimum expected degree, $w_{\max} = \max_i w_i$ be the maximum expected degree, and $\bar{w} = \frac{1}{n}\sum_i w_i$ be the average expected degree.

Let $V = \sum_i w_i$ be the sum of expected degrees and $v = \sum_i d_i(\mathbf{A})$ the sum of realized degrees.

For any matrix $\mathbf{T}$, let $\|\mathbf{T}\| = \sup_{\|\mathbf{v}\|=1}\langle \mathbf{v}, \mathbf{T}\mathbf{v}\rangle$, where the inner product is the standard (i.e., unweighted) Euclidean dot product. Similarly, an unadorned norm $\|\mathbf{v}\|$ will refer to $\langle \mathbf{v},\mathbf{v}\rangle = \sum_i v_i^2$. Define $\mathbf{E}$ to be the $n \times n$ matrix of ones, and let

(7) $$\mathbf{J} = \mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2} - v^{-1}\mathbf{D}(\mathbf{A})^{1/2}\mathbf{E}\mathbf{D}(\mathbf{A})^{1/2}$$

and

$$\mathbf{K} = \mathbf{D}(\mathbf{R})^{-1/2}\mathbf{R}\mathbf{D}(\mathbf{R})^{-1/2} - V^{-1}\mathbf{D}(\mathbf{R})^{1/2}\mathbf{E}\mathbf{D}(\mathbf{R})^{1/2}. \tag{8}$$

Note that $\mathbf{D}(\mathbf{R})$ is a diagonal matrix whose $(i, i)$ entry is $w_i$.

Now we observe a basic fact from linear algebra:

**Fact 1.** $\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}$ and $\mathbf{T}(\mathbf{A}) = \mathbf{D}(\mathbf{A})^{-1}\mathbf{A}$ are similar matrices, so that they have the same eigenvalues, and that $v^{-1}\mathbf{D}(\mathbf{A})^{1/2}\mathbf{E}\mathbf{D}(\mathbf{A})^{1/2}$ is the summand of the spectral decomposition of $\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}$ corresponding to the eigenvalue 1. The same reasoning applies when we replace $\mathbf{A}$ by $\mathbf{R}$ and $v$ by $V$.

Why is this true? One gets from $\mathbf{D}(\mathbf{A})^{-1}\mathbf{A}$ to $\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}$ by the following transformation: multiplying on the left by $\mathbf{D}(\mathbf{A})^{1/2}$ and on the right by $\mathbf{D}(\mathbf{A})^{-1/2}$. The eigenvalue 1 summand of the spectral decomposition of $\mathbf{T}(\mathbf{A})$ is $v^{-1}\mathbf{E}\mathbf{D}(\mathbf{A})$ and applying this same transformation to that matrix yields the claim. The argument in the second case is analogous.

Next, Theorem 2 can be reduced to the following proposition, which will also be useful for proving Proposition A.5.

**Proposition A.6.** If $w_{\min}/\log^2 n$ is high enough, then with probability at least $1 - \delta$ we have $\|\mathbf{J} - \mathbf{K}\| < \delta$.

Before proving the proposition, we show why Theorem 2 is a consequence. Recall the statement of this theorem:

THEOREM 2. Consider a sequence of multi-type random networks described by $(\mathbf{P}, \mathbf{n})$ that is sufficiently dense and satisfies the conditions of no vanishing groups and comparable densities from Definition 3. Then for large enough $n$,

$$|\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P}, \mathbf{n}))) - \lambda_2(\mathbf{F}(\mathbf{P}, \mathbf{n}))| \leq \delta,$$

with probability at least $1 - \delta$.

**Proof of Theorem 2 using Proposition A.6**: It is clear that $\mathbf{D}(\mathbf{R})^{-1}\mathbf{R}$ has the same eigenvalues as $\mathbf{F}$, so to prove the claim, it suffices to prove that the former matrix has second eigenvalue close enough to that of $\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}$.

By Fact 1, we know that $\|\mathbf{J}\|$ is the second-largest eigenvalue in magnitude of the matrix $\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}$, and $\|\mathbf{K}\|$ is the second-largest eigenvalue in magnitude of the matrix $\mathbf{D}(\mathbf{R})^{-1/2}\mathbf{R}\mathbf{D}(\mathbf{R})^{-1/2}$. Thus, by the triangle inequality, if we can show that with probability at least $1-\delta$ we have $\|\mathbf{J}-\mathbf{K}\| < \delta$, then the proof is done. This is the content of Proposition A.6. ∎

We will turn to the proof of Proposition A.6 right after stating the following lemma that is essential to it and several other results. It is from the proof of Theorem 3.6 of Chung, Lu, and Vu (2004), and follows from the Chernoff inequality.

**Lemma A.4.** Fix any $\delta > 0$. If $w_{\min}/\log n$ is large enough, the following statement holds with probability at least $1 - \delta$ for all $i$ simultaneously: $|d_i - w_i| < \delta w_i$.

**Proof of Proposition A.6**: Write

$$\mathbf{J}-\mathbf{K}=\mathbf{B}+\mathbf{C}+\mathbf{U}+\mathbf{M} \qquad \text{where} \qquad \begin{aligned} B_{ij} &= \frac{A_{ij}}{\sqrt{d_i d_j}}\left(1-\frac{\sqrt{d_i d_j}}{\sqrt{w_i w_j}}\right) \\ C_{ij} &= \frac{A_{ij}-R_{ij}}{\sqrt{w_i w_j}} \\ U_{ij} &= \frac{\sqrt{w_i w_j}}{V}\left(1-\frac{\sqrt{d_i d_j}}{\sqrt{w_i w_j}}\right) \\ M_{ij} &= (V^{-1}-v^{-1})\sqrt{d_i d_j}. \end{aligned}$$

By the triangle inequality,

$$\|\mathbf{J}-\mathbf{K}\| \le \|\mathbf{B}\|+\|\mathbf{C}\|+\|\mathbf{U}\|+\|\mathbf{M}\|,$$

so it suffices to bound the pieces individually.

Now we list two lemmas from Chung, Lu, and Vu (2004), which will be used in the following argument. Their paper deals with a special case of the multi-type random graph model in which $R_{ij}=w_i w_j \rho$ for a constant $\rho$, but their arguments that we use rely only on very simple features of the expected entries of the adjacency matrix which also hold without change in our setting. Only one step of a proof is at all different, and we discuss that below in the proof of Lemma A.5.

**Lemma A.5.** Fix any $\delta>0$. Then there exists some $c>0$, independent of $n$, so that if $w_{\min}/\log^2 n$ is high enough, with probability at least $1-\delta$:

$$\|\mathbf{C}\| \le \frac{2c}{\sqrt{\bar{w}}}+\frac{c\log n}{\sqrt{w_{\min}}}.$$

**Proof of Lemma A.5**: The only step of the proof of this last lemma that does not work exactly as in the proofs of Theorems 3.2 and 3.6 of Chung, Lu, and Vu (2004) is their equation (3.2). Let $C^r_{ij}$ denote the $(i,j)$ entry of $\mathbf{C}^r$. This step asserts (in our notation) that for $t\ge 2$, we have

$$\mathbb{E}[C^t_{ij}] \le \frac{(1-R_{ij})R_{ij}+(-R_{ij})^t(1-R_{ij})}{(w_i w_j)^{t/2}} \le \frac{R_{ij}}{(w_i w_j)^{t/2}} \le \frac{w_i w_j/V}{(w_i w_j)^{t/2}} \le \frac{1/V}{(w_{\min})^{t-2}}.$$

We will prove a slightly weaker statement that is still sufficient to make the rest of the proof go through unchanged. The step which is slightly different is the penultimate inequality. We will show that, for large enough $n$,

$$R_{ij} \le c\cdot w_i w_j/V \tag{9}$$

for some real number $c\ge 0$, and conclude that

$$\mathbb{E}[C^t_{ij}] \le \frac{c/V}{(w_{\min})^{t-2}},$$

which suffices for the Chung-Lu-Vu proof.

We can rewrite (9) as

$$R_{ij}\left(\sum_{i',j'} R_{i'j'}\right) \le c\left(\sum_{j'} R_{ij'}\right)\left(\sum_{i'} R_{i'j}\right). \tag{10}$$

To show that this is true, we proceed as follows. Due to the *no vanishing groups* condition, we can find some constant $c_1 > 0$ so that for large enough $n$ we have $\sum_{j'} R_{ij'} \ge c_1 n R_{ij}$ and $\sum_{i'} R_{i'j} \ge c_1 n R_{ij}$, since the groups of agents $i$ and $j$ must both grow at a rate of at least $n$. Thus,

$$n^2 R_{ij}^2 \le c_1^{-2}\left(\sum_{j'} R_{ij'}\right)\left(\sum_{i'} R_{i'j}\right). \tag{11}$$

As a result of the *comparable densities* condition, we have a constant $c_2$ so that

$$\sum_{i',j'} R_{i'j'} \le c_2 n^2 R_{ij},$$

from which we deduce

$$R_{ij}\left(\sum_{i',j'} R_{i'j'}\right) \le c_2 n^2 R_{ij}^2. \tag{12}$$

Combining (12) with (11) above, we conclude that (10) holds for large enough $n$, as desired. ∎

It follows that, if $w_{\min}/\log^2 n$ is high enough, we have $\|\mathbf{C}\| < \delta/4$ with probability at least $1-\delta/4$.

**Lemma A.6.** Fix any $\delta > 0$. If $w_{\min}/\log^2 n$ is high enough, the following statement holds with probability at least $1-\delta$:

$$\|\mathbf{M}\| \le \frac{1}{\sqrt{\bar{w}}}.$$

It follows that, when $w_{\min}/\log^2 n$ is high enough, we have $\|\mathbf{M}\| < \delta/4$ with probability at least $1-\delta/4$.

To bound $\|\mathbf{B}\|$ and $\|\mathbf{U}\|$, we will use Lemma A.4 and two simple facts about the matrix norm, whose proofs are immediate. Let $\mathrm{abs}(\mathbf{X})$ denote the matrix whose $(i,j)$ entry is $|X_{ij}|$.

**Lemma A.7.** 1. For any matrix $\mathbf{X}$, $\|\mathbf{X}\| \le \|\mathrm{abs}(\mathbf{X})\|$.

2. Suppose there are two nonnegative matrices, $\mathbf{X}$ and $\mathbf{Y}$ and a constant $c > 0$ such that for each $i, j$, we have $Y_{ij} < cX_{ij}$. Then $\|\mathbf{Y}\| \le c\|\mathbf{X}\|$.

To show that, with probability at least $1-\delta$, we have $\|\mathbf{B}\| < \delta/4$, define $\hat{\mathbf{B}} = \mathrm{abs}(\mathbf{B})$;

by Lemma A.7(1), it suffices to show $\|\hat{\mathbf{B}}\| < \delta/4$. Note

$$\hat{B}_{ij} = \frac{A_{ij}}{\sqrt{d_i d_j}} \left| 1 - \frac{\sqrt{d_i d_j}}{\sqrt{w_i w_j}} \right|.$$

By Lemma A.4, we have with probability at least $1 - \delta/4$ that

$$\left| 1 - \frac{\sqrt{d_i d_j}}{\sqrt{w_i w_j}} \right| < \delta/4$$

and so, noting that

$$\|\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}\| = 1$$

and using Lemma A.7(2), the claim is proved.

Precisely the same argument works to show that, with probability at least $1 - \delta/4$, we have $\|\mathbf{U}\| < \delta/4$, with $V^{-1}\mathbf{D}(\mathbf{R})^{1/2}\mathbf{E}\mathbf{D}(\mathbf{R})^{1/2}$, which also has norm 1, playing the role of $\mathbf{D}(\mathbf{A})^{-1/2}\mathbf{A}\mathbf{D}(\mathbf{A})^{-1/2}$.

Combining all the bounds shows that, with probability at least $1-\delta$, we have $\|\mathbf{J}-\mathbf{K}\| < \delta$, as desired.

This completes the proof of the proposition. ∎

The results established so far in this section will now be used to prove Proposition A.5.

**Proof of Proposition A.5**: We will reuse for new purposes some of the variable names used solely inside the proof of Proposition A.6, but the variables defined for the whole subsection will be unchanged.

The upper bound is a direct consequence of Lemma 2. For the lower bound, define $\zeta = \varepsilon^2$. We will show that there is an initial vector of beliefs $\mathbf{b}$ so that $\|\mathbf{T}(\mathbf{A})^t\mathbf{b} - \mathbf{T}(\mathbf{A})^\infty\mathbf{b}\|^2_{\mathbf{s}(\mathbf{A})} \geq \zeta$ for at least $z - 1$ steps, where

$$z := \left\lfloor \frac{\log(1/\sqrt{8}\varepsilon) - \log(1/\sqrt{\alpha\beta})}{\log(1/|\lambda_2(\mathbf{T}(\mathbf{A}))|)} \right\rfloor.$$

This will suffice for the proof. The reason for the transformation is to make notation less cumbersome by working with the square of the norm rather than norm itself.

Let $\mathbf{R}$ be as above but now let $\mathbf{C} = \mathbf{D}(\mathbf{R})^{-1}\mathbf{R}$ and $\mathbf{T} = \mathbf{T}(\mathbf{A})$. That is, $\mathbf{C}$ is the version of $\mathbf{T}$ in the world where realizations of link random variables are replaced by their expectations. Also, note that we can write $z$ above equivalently as

$$z = \left\lfloor \frac{\log(1/8\zeta) - \log(1/\alpha\beta)}{2\log(1/|\lambda_2(\mathbf{T})|)} \right\rfloor.$$

There are three steps to the proof. In Step 1, we show that for $\mathbf{C}^t\mathbf{b}$ to converge within $2\zeta$ of its limit when distance is measured by $\|\cdot\|^2_{\mathbf{s}(\mathbf{A})}$ takes at least $z - 1$ steps for some $\mathbf{b}$. In Step 2, we use Proposition A.6 to show that for any $\eta > 0$, we can find a high enough $n$ so that $\|\mathbf{T} - \mathbf{C}\| < \eta$ with probability at least $1 - \eta$. In Step 3, we rely on Step 2 to show that, if $\eta$ is chosen small enough, then $\mathbf{C}^t\mathbf{b}$ and $\mathbf{T}^t\mathbf{b}$ are at most $\zeta$ apart for at least $z-1$ steps under the inner product $\langle\cdot,\cdot\rangle_{\mathbf{s}(\mathbf{A})}$. This proves the proposition.

**Step 1.** Let $\mathbf{v}$ be a right eigenvector of $\mathbf{C}$ corresponding to eigenvalue $\hat{\lambda}_2 := \lambda_2(\mathbf{F})$ (this is also the second eigenvalue in magnitude of $\mathbf{C}$ by Fact 1). By multiplying $\mathbf{v}$ by a scalar if necessary, we may assume that the entry with largest magnitude is 1/2. By the assumption of interior homophily, $\hat{\lambda}_2$ is bounded away from 0 for all $n$. Given this and the fact that $\mathbf{C}$ is constant on a given type, it follows that $\mathbf{v}$ is constant on a given type. Thus, by definition of $\alpha$ above, that there are at least $\alpha n$ entries in $\mathbf{v}$ equal to 1/2. And from this it follows, by the definition of $\mathbf{s}(\mathbf{R})$ and the definition of $\beta$ above, that

$$\langle\mathbf{v},\mathbf{v}\rangle_{\mathbf{s}(\mathbf{R})} \geq n\alpha\cdot\left(\frac{1}{2}\right)^2\cdot\frac{w_{\min}}{nw_{\max}} \geq \frac{\alpha\beta}{4}.$$

Setting $b_i = v_i + 1/2$, we see as at the end of the proof of Lemma 2 that

$$\|\mathbf{C}^t\mathbf{b} - \mathbf{C}^\infty\mathbf{b}\|^2_{\mathbf{s}(\mathbf{R})} \geq \frac{\alpha\beta}{4}|\lambda_2(\mathbf{C})|^{2t},$$

which yields the lower bound on convergence time we want with $\lambda_2(\mathbf{C})$ instead of $\lambda_2(\mathbf{T})$. But in view of the assumption of interior homophily and Theorem 2, for high enough $n$ we can replace $\mathbf{C}$ by $\mathbf{T}$ and lose at most an additive factor of 1 in the bound.

**Step 2.** Recall that $\mathbf{C} = \mathbf{D}(\mathbf{R})^{-1}\mathbf{R}$ and $\mathbf{T} = \mathbf{T}(\mathbf{A})$. Also write:

$$\mathbf{U} = \mathbf{D}(\mathbf{A})^{-1/2}\mathbf{J}\mathbf{D}(\mathbf{A})^{1/2}$$

and

$$\mathbf{M} = \mathbf{D}(\mathbf{R})^{-1/2}\mathbf{K}\mathbf{D}(\mathbf{R})^{1/2}.$$

Recalling the definitions of $\mathbf{J}$ and $\mathbf{K}$ from (7) and (8) above, we see that

$$\mathbf{T} - \mathbf{C} = v^{-1}\mathbf{E}\mathbf{D}(\mathbf{A}) - V^{-1}\mathbf{E}\mathbf{D}(\mathbf{R}) + \mathbf{U} - \mathbf{M}.$$

So by the triangle inequality, it suffices to bound $\|v^{-1}\mathbf{E}\mathbf{D}(\mathbf{A}) - V^{-1}\mathbf{E}\mathbf{D}(\mathbf{R})\|$ and $\|\mathbf{U}-\mathbf{M}\|$. Fix a $\gamma > 0$. By Lemma A.4, if $w_{\min}/\log^2 n$ is high enough, the following event occurs with probability at least $1-\gamma$ for all $i$ simultaneously: $|d_i - w_i| < \gamma w_i$. (Given the assumptions of this proposition, high enough $n$ ensures the condition of the lemma is met.) Call this event $E_1$. Thus, on $E_1$,

$$\|v^{-1}\mathbf{E}\mathbf{D}(\mathbf{A}) - V^{-1}\mathbf{E}\mathbf{D}(\mathbf{R})\| < \gamma,$$

and so it suffices to take care of the other term.

By Proposition A.6, we know that if $w_{\min}/\log^2 n$ is high enough, then on an event $E_2$ of probability at least $1-\gamma$ we have $\|\mathbf{J}-\mathbf{K}\| < \gamma$. As above, for high enough $n$ the condition is met. Now let

$$\mathbf{F} = \mathbf{D}(\mathbf{A}) - \mathbf{D}(\mathbf{R}),$$

$$\mathbf{G} = (\mathbf{D}(\mathbf{R}) + \mathbf{F})^{1/2} - \mathbf{D}(\mathbf{R})^{1/2},$$

and

$$\mathbf{H} = (\mathbf{D}(\mathbf{R}) + \mathbf{F})^{-1/2} - \mathbf{D}(\mathbf{R})^{-1/2}.$$

Observe that

$$\begin{aligned}
\|\mathbf{U} - \mathbf{M}\| &= \|(\mathbf{D}(\mathbf{R}) + \mathbf{F})^{-1/2}\mathbf{J}(\mathbf{D}(\mathbf{R}) + \mathbf{F})^{1/2} - \mathbf{D}(\mathbf{R})^{-1/2}\mathbf{K}\mathbf{D}(\mathbf{R})^{1/2}\| \\
&= \|(\mathbf{D}(\mathbf{R})^{-1/2} + \mathbf{H})\mathbf{J}(\mathbf{D}(\mathbf{R})^{1/2} + \mathbf{G}) - \mathbf{D}(\mathbf{R})^{-1/2}\mathbf{K}\mathbf{D}(\mathbf{R})^{1/2}\| \\
&= \|\mathbf{D}(\mathbf{R})^{-1/2}(\mathbf{J} - \mathbf{K})\mathbf{D}(\mathbf{R})^{1/2} + \mathbf{D}(\mathbf{R})^{-1/2}\mathbf{J}\mathbf{G} + \mathbf{H}\mathbf{J}\mathbf{D}(\mathbf{R})^{1/2} + \mathbf{H}\mathbf{J}\mathbf{G}\| \\
&\leq \|\mathbf{D}(\mathbf{R})^{-1/2}(\mathbf{J} - \mathbf{K})\mathbf{D}(\mathbf{R})^{1/2}\| + \|\mathbf{D}(\mathbf{R})^{-1/2}\mathbf{J}\mathbf{G}\| \\
&\quad + \|\mathbf{H}\mathbf{J}\mathbf{D}(\mathbf{R})^{1/2}\| + \|\mathbf{H}\mathbf{J}\mathbf{G}\|.
\end{aligned}$$

Using Lemma A.4 and standard series approximation arguments, for high enough $n$ we can ensure $\|\mathbf{G}\| \leq \gamma\|\mathbf{D}(\mathbf{R})^{1/2}\|$ and $\|\mathbf{H}\| \leq \gamma\|\mathbf{D}(\mathbf{R})^{-1/2}\|$ on an event $E_3$ of probability at least $1 - \gamma$. Using the fact that $\|\mathbf{J}\| \leq 1$, the Cauchy-Schwartz inequality yields that each of the middle two terms above is bounded by $\gamma$. For the last term, note that

$$\|\mathbf{H}\mathbf{J}\mathbf{G}\| \leq \gamma^2\|\mathbf{D}(\mathbf{R})^{1/2}\| \cdot \|\mathbf{D}(\mathbf{R})^{-1/2}\| \leq \gamma^2 \cdot \left(\frac{w_{\max}}{w_{\min}}\right)^{1/2} \leq \frac{\gamma^2}{\beta^{1/2}}.$$

So it suffices to take care of the first term in our expansion of $\|\mathbf{U} - \mathbf{M}\|$ above. This is accomplished by noticing that, on $E_2$,

$$\begin{aligned}
\|\mathbf{D}(\mathbf{R})^{-1/2}(\mathbf{J} - \mathbf{K})\mathbf{D}(\mathbf{R})^{1/2}\| &\leq \frac{w_{\max}^{1/2}}{w_{\min}^{1/2}} \cdot \|\mathbf{J} - \mathbf{K}\| \\
&\leq \frac{1}{\beta^{1/2}} \cdot \|\mathbf{J} - \mathbf{K}\| && \text{definition of } \beta \\
&\leq \frac{\gamma}{\beta^{1/2}} && \text{definition of } E_2.
\end{aligned}$$

Together, these facts show that for high enough $n$, on $E_1 \cap E_2 \cap E_3$, which occurs with probability at least $1 - 3\gamma$, we have

$$\|\mathbf{T} - \mathbf{C}\| \leq \gamma + \frac{(1 + \gamma)\gamma}{(1 - \gamma)\beta^{1/2}} + 2\gamma + \frac{\gamma^2}{\beta^{1/2}}.$$

By choosing $\gamma$ so that the right-hand side is less than $\eta$ and $3\gamma < \eta$ (to take care of the probability), the step is complete.

**Step 3.** Fix $t \leq z - 1$. Write $\mathbf{T} = \mathbf{C} + \mathbf{Y}$, where $\|\mathbf{Y}\| \leq \eta$. Note that

$$(\mathbf{T} + \mathbf{Y})^t = \mathbf{T}^t + \sum_{q=1}^{t} \mathbf{X}_q,$$

where $\mathbf{X}_q$ is a sum of $\binom{t}{q}$ terms, each of which is a product of $q$ copies of $\mathbf{Y}$ and $t - q$ copies of $\mathbf{T}$ in some order. By the fact that $\|\mathbf{T}\| = 1$ and $\|\mathbf{Y}\| \leq \eta$, we have $\|\mathbf{X}_q\| \leq \binom{t}{q}\eta^q$ for each

$q \geq 1$. Then, by the triangle inequality,

$$\left\| \sum_{q=1}^{t} \mathbf{X}_q \right\| \leq \sum_{q=1}^{t} \binom{t}{q} \eta^q \leq \frac{2^{z-1}\eta}{1-\eta}.$$

Thus,

$$\mathbf{Y}_t := \|\mathbf{C}^t - \mathbf{T}^t\| \leq \frac{2^{z-1}\eta}{1-\eta}.$$

Take $\mathbf{b}$ and $\mathbf{v}$ to be the vectors constructed in Step 1. Then, for high enough $n$,

$$\begin{aligned}
\|\mathbf{T}^t\mathbf{b} - \mathbf{T}^\infty\mathbf{b}\|^2_{\mathbf{s}(\mathbf{A})} &= \langle \mathbf{T}^t\mathbf{v}, \mathbf{T}^t\mathbf{v}\rangle_{\mathbf{s}(\mathbf{A})} \\
&= \langle (\mathbf{C}^t + \mathbf{Y}_t)\mathbf{v}, (\mathbf{C}^t + \mathbf{Y}_t)\mathbf{v}\rangle_{\mathbf{s}(\mathbf{A})} \\
&\geq \langle \mathbf{C}^t\mathbf{v}, \mathbf{C}^t\mathbf{v}\rangle_{\mathbf{s}(\mathbf{A})} + 2\langle \mathbf{Y}_t\mathbf{v}, \mathbf{C}^t\mathbf{v}\rangle_{\mathbf{s}(\mathbf{A})} \\
&\geq (1-\eta)\langle \mathbf{C}^t\mathbf{v}, \mathbf{C}^t\mathbf{v}\rangle_{\mathbf{s}(\mathbf{R})} \\
&\quad + 2\langle \mathbf{Y}_t\mathbf{v}, \mathbf{C}^t\mathbf{v}\rangle_{\mathbf{s}(\mathbf{A})} && \text{Lemma A.4} \\
&\geq 2(1-\eta)\zeta + 2\langle \mathbf{Y}_t\mathbf{v}, \mathbf{C}^t\mathbf{v}\rangle_{\mathbf{s}(\mathbf{A})} && \text{definition of } z \\
&\geq 2(1-\eta)\zeta - 2\|\mathbf{Y}_t\mathbf{v}\|_{\mathbf{s}(\mathbf{A})} \cdot \|\mathbf{C}^t\mathbf{v}\|_{\mathbf{s}(\mathbf{A})} && \text{Cauchy-Schwartz} \\
&\geq 2(1-\eta)\zeta - 2\|\mathbf{Y}_t\mathbf{v}\|_{\mathbf{s}(\mathbf{A})} && \text{see below} \\
&\geq 2(1-\eta)\zeta - 2\|\mathbf{Y}_t\mathbf{v}\| && \text{def'n of } \|\cdot\|_{\mathbf{s}(\mathbf{A})} \\
&\geq 2(1-\eta)\zeta - \frac{2^z\eta}{1-\eta}.
\end{aligned}$$

The step whose explanation is missing is straightforward; no entries in $\mathbf{v}$ have magnitude exceeding $1/2$ and multiplication by the stochastic matrix $\mathbf{C}$ preserves this property. Since $\mathbf{s}(\mathbf{R})$ is a probability distribution (i.e. has nonnegative entries summing to 1), the inequality $\|\mathbf{C}^t\mathbf{v}\|_{\mathbf{s}(\mathbf{R})} \leq 1$ holds by definition of the inner product. Choosing $\eta$ so that $2(1-\eta)\zeta - \frac{2^z\eta}{1-\eta} > \zeta$, the proof is complete. ∎

# C Consequences

The main consequence of the machinery above is, of course, Theorem 1.

Theorem 1. Consider a sufficiently dense sequence of multi-type random networks satisfying the three regularity conditions in Definition 3. Then, for any $\gamma > 0$:

$$\mathrm{CT}\left(\frac{\gamma}{n}; \mathbf{A}(\mathbf{P}, \mathbf{n})\right) \approx \frac{\log(n)}{\log(1/|h^{\mathrm{spec}}(\mathbf{P}, \mathbf{n})|)}.$$

**Proof of Theorem 1**: We will index objects by $n$ to make explicit the sequence of networks and use the notation $\mathbf{A}(n) = \mathbf{A}(\mathbf{P}(n), \mathbf{n}(n))$ and $\mathbf{T}(n) = \mathbf{T}(\mathbf{A}(n))$.

First, note that with a probability going to 1, $\mathbf{A}$ is connected: apply Proposition 4 noting that $h(n)$ is bounded away from 1 so that its condition (iv) applies, and (i)-(iii) apply given the islands model and $w_{\min} \geq \log^2(n)$. Thus, we can apply Lemma 2.

By the assumed regularity conditions, the minimum expected degree divided by the maximum expected degree remains bounded. Lemma A.4 above guarantees that this is true of the realized degrees, too, with a probability tending to 1 as $n$ grows. Thus, the expression $-\log(1/\underline{s}^{1/2})$ of Lemma 2 can be bounded below by $-\frac{1}{2}\log(cn)$ for some positive constant $c$. In view of this, we can conclude from Lemma 2 that with a probability going to 1

$$\left\lfloor \frac{\log(n/(2\gamma)) - \frac{1}{2}\log(cn)}{-\log(|\lambda_2(\mathbf{T}(n))|)} \right\rfloor \leq \mathrm{CT}(\gamma/n; \mathbf{A}(n)) \leq \left\lceil \frac{\log(n/\gamma)}{-\log(|\lambda_2(\mathbf{T}(n))|)} \right\rceil.$$

This implies that with a probability going to 1:

$$\left\lfloor \frac{\frac{1}{2}\log(n) - \log(2\gamma) - \frac{1}{2}\log(c)}{-\log(|\lambda_2(\mathbf{T}(n))|)} \right\rfloor \leq \mathrm{CT}(\gamma/n; \mathbf{A}(n)) \leq \left\lceil \frac{\log(n) - \log(\gamma)}{-\log(|\lambda_2(\mathbf{T}(n))|)} \right\rceil. \tag{13}$$

Next, applying Theorem 2,

$$|\lambda_2(\mathbf{T}(n)) - h^{\mathrm{spec}}(\mathbf{P}(n), \mathbf{n}(n))| \xrightarrow{p} 0. \tag{14}$$

Since $h^{\mathrm{spec}}(\mathbf{P}(n), \mathbf{n}(n))$ is bounded away from 1, by (14), we deduce that for any $1 > \delta > 0$, with a probability going to 1

$$\frac{1-\delta}{-\log(|h^{\mathrm{spec}}(\mathbf{P}(n), \mathbf{n}(n))|)} \leq \frac{1}{-\log(|\lambda_2(\mathbf{T}(n)|)} \leq \frac{1+\delta}{-\log(|h^{\mathrm{spec}}(\mathbf{P}(n), \mathbf{n}(n))|)}.$$

The result then follows from (13) the fact that $(\log n)/(\log c) \to \infty$. ∎

Proposition 2. Consider a sequence of multi-type random networks $(\mathbf{P}, \mathbf{n})$ and another $(\mathbf{P}', \mathbf{n})$ where $\mathbf{P}' = c\mathbf{P}$ for some $c > 0$. Under the conditions of Theorem 1, the ratio of consensus times

$$\frac{\mathrm{CT}\left(\frac{\gamma}{n}; \mathbf{A}(\mathbf{P}, \mathbf{n})\right)}{\mathrm{CT}\left(\frac{\gamma}{n}; \mathbf{A}(\mathbf{P}', \mathbf{n})\right)}$$

converges in probability to 1.

**Proof of Proposition 2**: Observe that $\mathbf{F}(\mathbf{P}, \mathbf{n})$ is invariant to the density adjustment in the statement of the proposition; thus, applying Proposition A.5, we conclude the claim in the corollary. ∎

Corollary 2. Consider a sufficiently dense sequence of two-group random networks (as described above) satisfying the three regularity conditions in Definition 3. Then for any $\gamma > 0$:

$$\mathrm{CT}\left(\frac{\gamma}{n}; \mathbf{A}(\mathbf{P}, \mathbf{n})\right) \approx \frac{\log(n)}{\log(1/|h^{\mathrm{two}}(p_s, p_d, \mathbf{n})|)}.$$

**Proof of Corollary 2**: To apply Theorem 2, we need to compute the second eigenvalue of

$$\mathbf{F}(\mathbf{P}) = \begin{bmatrix} \frac{n_1 p_s}{n_1 p_s + n_2 p_d} & \frac{n_2 p_d}{n_1 p_s + n_2 p_d} \\ \frac{n_1 p_d}{n_1 p_d + n_2 p_s} & \frac{n_2 p_s}{n_1 p_d + n_2 p_s} \end{bmatrix},$$

or

$$\mathbf{F}(\mathbf{P}) = \begin{bmatrix} \frac{n_1 p_s}{n p_1} & \frac{n_2 p_d}{n p_1} \\ \frac{n_1 p_d}{n p_2} & \frac{n_2 p_s}{n p_2} \end{bmatrix}.$$

This has a second eigenvalue of

$$\frac{n_1 p_s}{n p_1} - \frac{n_1 p_d}{n p_2},$$

which is also equal to

$$-\frac{n_2 p_d}{n p_1} + \frac{n_2 p_s}{n p_2}.$$

Thus, the second eigenvalue is also equal to

$$\frac{1}{2}\left(\frac{n_1 p_s - n_2 p_d}{n p_1} + \frac{n_2 p_s - n_1 p_d}{n p_2}\right).$$

This can be rewritten as

$$\frac{1}{2}\left(\frac{2n_1 p_s - n p_1}{n p_1} + \frac{2n_2 p_s - n p_2}{n p_2}\right),$$

or

$$\frac{1}{2}\left(\frac{2n_1}{n}\frac{p_s - p_1}{p_1} + \frac{2n_2}{n}\frac{p_s - p_2}{p_2} + \frac{2n_1 + 2n_2 - 2n}{n}\right).$$

This simplifies to

$$\frac{n_1}{n}\frac{p_s - p_1}{p_1} + \frac{n_2}{n}\frac{p_s - p_2}{p_2} = \frac{n_2}{n} h_1 + \frac{n_1}{n} h_2.$$

The result then follows from Theorem 2. ∎

## D Reducing to the Dynamics of Representative Agents

**Proof of Proposition 3**: The proof proceeds in three steps. First, we show that the consensus limit of $\mathbf{b}(t)$ is, for large enough $n$, within $\delta/2$ of the consensus limit of $\overline{\mathbf{b}}(t)$ with probability at least $1-\delta/2$. Second, we note that, with probability at least $1-\delta/2$, there is some fixed time $T$ such that for all large enough $n$, $\mathbf{b}(t)$ for $t \geq T$ is permanently within $\delta/2$ of its consensus limit. Thus, the claim of the proposition holds for times $t \geq T$. Third, we show that the claim of the proposition holds for times $t < T$ by recalling $\mathbf{C}$ from the proof of Proposition A.5 and writing

$$\overline{\mathbf{b}}(t) = \mathbf{C}^t \mathbf{b}(0)$$
$$\mathbf{b}(t) = \mathbf{T}^t \mathbf{b}(0),$$

then using the fact that we can ensure that the spectral norm of $\mathbf{C}^t\mathbf{b} - \mathbf{T}^t\mathbf{b}$ is small for any given finite number of steps, which follows from the argument of Step 2 of Proposition A.5.

**Step 1.** By several applications of the Chernoff inequality to the Bernoulli random variables $A_{ij}$, it follows that, with probability at least $1-\delta/4$, the following inequalities hold simultaneously for all pairs $(i, k)$, where $k$ is the type of $i$:

$$\left| \frac{d_i(\mathbf{A}(\mathbf{P},\mathbf{n}))}{D(\mathbf{A}(\mathbf{P},\mathbf{n}))} - \frac{d_k(\mathbf{Q}(\mathbf{P},\mathbf{n}))}{D(\mathbf{Q}(\mathbf{P},\mathbf{n}))} \right| < \delta/4.$$

By the weak law of large numbers, with probability at least $1 - \delta/(4m)$, the average belief of agents in group $k$ is within $\delta/(4m)$ of the expected belief $\mu_k$ of an agent in group $k$ (recall that $m$ is thet total number of groups). Combining these observations, it follows that, with probability at least $1 - \delta/2$, the consensus belief of the society,

$$\sum_i \frac{d_i(\mathbf{A}(\mathbf{P},\mathbf{n}))}{D(\mathbf{A}(\mathbf{P},\mathbf{n}))} b_i(0)$$

is within $1 - \delta/2$ of the consensus belief of the representative agents, namely

$$\sum_k \frac{d_k(\mathbf{Q}(\mathbf{P},\mathbf{n}))}{D(\mathbf{Q}(\mathbf{P},\mathbf{n}))} \mu_k.$$

**Step 2.** This step follows directly from Proposition A.5 after noting that the $\ell_2$ norm $\|\cdot\|_{\mathbf{e}/n}$ used in the statement of the present proposition is within a constant factor of the weighted $\ell_2$ norm $\|\cdot\|_{\mathbf{s}(\mathbf{A})}$ used in the definition of CT, since the ratio of any two degrees is bounded by a constant (indepenent of $n$) with high probability.

**Step 3.** The only subtlety to note that is not given in the sketch above is that, once again, it does not matter whether we refer to the weighted $\ell_2$ norm or an unweighted one in defining the spectral norm, for the same reason given above in Step 2. ∎

## E What Spectral Homophily Measures

The proof of Lemma 3 uses the Courant-Fischer variational characterization of the second-largest eigenvalue of $\mathbf{F}(\mathbf{P},\mathbf{n})$. We saw above that $\mathbf{F}(\mathbf{A})$ is similar to

$$\mathbf{D}(\mathbf{P},\mathbf{n})^{-1/2}\mathbf{F}(\mathbf{P},\mathbf{n})\mathbf{D}(\mathbf{P},\mathbf{n})^{-1/2},$$

where $\mathbf{D}(\mathbf{P},\mathbf{n})^{-1/2}$ is a diagonal matrix with the degree of group $k$ in position $(k, k)$. Since this matrix is symmetric, we know that the eigenvalues of $\mathbf{F}(\mathbf{P},\mathbf{n})$ are all real. Letting $\beta_1(\mathbf{F}(\mathbf{P},\mathbf{n})) \geq \beta_2(\mathbf{F}(\mathbf{P},\mathbf{n})) \geq \cdots \geq \beta_n(\mathbf{F}(\mathbf{P},\mathbf{n}))$ be the eigenvalues of $\mathbf{F}(\mathbf{P},\mathbf{n})$ ordered as real numbers, the Courant-Fischer result is as follows.

**Proposition A.7.**

$$\beta_n(\mathbf{F}(\mathbf{P},\mathbf{n})) = \inf_{\mathbf{0}\neq\mathbf{v}\in\mathbb{R}^m} \left\{ \frac{\langle \mathbf{v}, \mathbf{F}(\mathbf{P},\mathbf{n})\mathbf{v}\rangle_{\mathbf{s}}}{\langle \mathbf{v}, \mathbf{v}\rangle_{\mathbf{s}}} \right\}. \tag{15}$$

$$\beta_2(\mathbf{F}(\mathbf{P},\mathbf{n})) = \sup_{\substack{\mathbf{0}\neq\mathbf{v}\in\mathbb{R}^m \text{ s.t.} \\ \langle \mathbf{v},\mathbf{e}\rangle=0}} \left\{ \frac{\langle \mathbf{v}, \mathbf{F}(\mathbf{P},\mathbf{n})\mathbf{v}\rangle_{\mathbf{s}}}{\langle \mathbf{v}, \mathbf{v}\rangle_{\mathbf{s}}} \right\}, \tag{16}$$

where the inner product everywhere is defined by $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{s}} = \sum_{k \in M} v_k w_k s_k$, and where $s_k = \frac{d_k(\mathbf{Q}(\mathbf{P},\mathbf{n}))}{\sum_\ell d_\ell(\mathbf{Q}(\mathbf{P},\mathbf{n}))}$.

LEMMA 3. If $\mathbf{Q}(\mathbf{P}, \mathbf{n})$ is connected (viewed as weighted network), then

$$|h^{\text{spec}}(\mathbf{P}, \mathbf{n})| \geq |\operatorname{DWH}(\mathbf{P}, \mathbf{n})|.$$

**Proof of Lemma 3**:

The arguments $\mathbf{P}, \mathbf{n}$ will be dropped throughout, and we will use $\mathbf{d}$ to refer to the vector with $d_k = d_k(\mathbf{Q}(\mathbf{P}, \mathbf{n})) = \sum_\ell Q_{k\ell}$. We will construct a $\mathbf{v}$ satisfying $\langle \mathbf{v}, \mathbf{e} \rangle_{\mathbf{d}} = 0$ so that the absolute value of the quantity $\langle \mathbf{v}, \mathbf{T}\mathbf{v} \rangle_{\mathbf{d}} / \langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{d}}$ is equal to $|\operatorname{DWH}(B)|$. Since $|\lambda_2| = \max\{|\beta_2|, |\beta_n|\}$, this shows that $|\lambda_2| \geq |\operatorname{DWH}(B)|$ using Proposition A.7.[2] Since $B$ is arbitrary, that proves the lemma (recalling the definition of $h^{\text{spec}}(\mathbf{P}, \mathbf{n})$).

Set $r = |B|$ and define

$$v_k = \begin{cases} \frac{1}{r d_k} & \text{if } k \in B \\ -\frac{1}{(m-r) d_k} & \text{if } k \notin B. \end{cases}$$

Let $D = \sum_i d_i$ and note

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{d}} &= \sum_{k \in M} v_k^2 \cdot d_k \\ &= \sum_{k \in B} \frac{d_k}{(r d_k)^2} + \sum_{i \in Bc} \frac{d_k}{((m-r) d_k)^2} \\ &= \frac{1}{r^2} \sum_{k \in B} \frac{1}{d_k} + \frac{1}{(m-r)^2} \sum_{k \in B^c} \frac{1}{d_k}. \end{aligned} \tag{17}$$

Also,

$$\begin{aligned} \langle \mathbf{v}, \mathbf{T}\mathbf{v} \rangle_{\mathbf{d}} &= \sum_{k \in M} v_k \left( \sum_\ell F_{k\ell} v_\ell \right) d_k \\ &= \sum_{k \in M} \sum_{\ell \in M} v_k F_{k\ell} v_\ell d_k \\ &= \frac{1}{r^2} \sum_{k,\ell \in B} F_{k\ell} F_{\ell k} + \frac{1}{(n-r)^2} \sum_{k,\ell \in B^c} F_{k\ell} T_{\ell k} - \frac{2}{r(n-r)} \sum_{k \in B, \ell \in B^c} F_{k\ell} F_{\ell k}. \end{aligned}$$

Dividing $\langle \mathbf{v}, \mathbf{T}\mathbf{v} \rangle_{\mathbf{d}}$ by $\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{d}}$ and using the definition of $W$ yields the result. ∎

[2] Note that $\mathbf{s}(\mathbf{A})$ and $\mathbf{d}(\mathbf{A})$ differ only by a normalization and so this does not affect the results.

## Appendix 2 Linear Updating with Persistent Opinions

In our baseline linear updating model of communication set out in Section II.D, agents update beliefs based exclusively on the current beliefs of their neighbors, and possibly their own. Let us consider the alternative updating rule under which each agent always places positive weight on his or her own initial position – the one that he or she held at time $t = 0$. Let us fix a network $\mathbf{A}$ and write $\mathbf{T}$ for $\mathbf{T}(\mathbf{A})$. This rule is given by:

$$\mathbf{b}(t+1) = (1-\alpha)\mathbf{b}(0) + \alpha\mathbf{T}\mathbf{b}(t). \tag{18}$$

From this one can deduce directly that

$$\mathbf{b}(t) = \left[(1-\alpha)\sum_{r=0}^{t-1}(\alpha\mathbf{T})^r + (\alpha\mathbf{T})^t\right]\mathbf{b}(0). \tag{19}$$

This model differs from the baseline model (in which $\alpha = 0$) in the sense that beliefs need not converge to consensus, regardless of the structure of the updating matrix $\mathbf{T}$. Nevertheless, it will still be useful to consider as a benchmark the consensus beliefs that the agents *would* converge to if they updated without any persistent weight on their own initial positions. These hypothetical consensus beliefs are given by

$$\mathbf{c} = \mathbf{T}^\infty\mathbf{b}(0).$$

In the baseline model, our proxy for differences of opinion was the consensus time. In this case, a more direct measure is available. Define the *disagreement distance* as

$$\mathrm{DD}(\mathbf{A}) = \limsup_t \max_{\mathbf{b}(0)\in[0,1]} \|\mathbf{b}(t) - \mathbf{c}\|_{\mathbf{s}(\mathbf{A})}.$$

The quantity $\|\mathbf{b}(t) - \mathbf{c}\|_{\mathbf{s}(\mathbf{A})}$ gives the average magnitude by which a message sent along a random link in the network at time $t$ differs from the hypothetical consensus. The quantity $\mathrm{DD}(\mathbf{A})$ measures how large this difference can be (persistently) in the worst case (compare with the discussion of consensus time in Section II.D.4).

We have the following result.

**Proposition A.8.**

$$\frac{1-\alpha}{1-\alpha\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P},\mathbf{n})))}\cdot\frac{\underline{s}}{4} \leq \mathrm{DD}(\mathbf{A}(\mathbf{P},\mathbf{n})) \leq \frac{1-\alpha}{1-\alpha\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P},\mathbf{n})))}, \tag{20}$$

where $\lambda_2(\mathbf{T})$ is the second-largest eigenvalue in magnitude of $\mathbf{T}$, and $\underline{s}$ is the smallest entry of the vector $\mathbf{s}(\mathbf{A})$ defined by $\mathbf{s}(\mathbf{A}) = \left(\frac{d_1(\mathbf{A})}{D(\mathbf{A})}, \dots, \frac{d_n(\mathbf{A})}{D(\mathbf{A})}\right)$.

The proof appears below. This result (and the proof) may be compared with Lemma 2, which makes a similar statement about the consensus time. The lower bound in this proposition has the deficiency that it is proportional to $\underline{s}$, which decays as $n$ grows. To remedy this, suppose now that $\mathbf{A}(\mathbf{P},\mathbf{n})$ is a multi-type random graph and $n$ is large. Then under the regularity conditions of Definition 3, it is possible to obtain an analogue of Proposition A.5 in Appendix 1, which makes the bounds well-behaved even as $n$ grows. That is, using an argument analogous to the proof of Proposition A.5, one can drop the $\underline{s}$, replace $\lambda_2(\mathbf{T}(\mathbf{A}(\mathbf{P},\mathbf{n})))$ by $h^{\text{spec}}(\mathbf{P},\mathbf{n})$, and get upper and lower bounds that match except for a constant; we will not repeat the derivations. The conclusion is that $\text{DD}(\mathbf{A})$ behaves like

$$\frac{1-\alpha}{1-\alpha h^{\text{spec}}(\mathbf{P},\mathbf{n})}.$$

The key comparative statics are as follows: the maximum sustainable disagreement distance is increasing and convex in homophily, attaining a maximum of 1 as homophily reaches the maximum value of 1. As $\alpha$ increases (which corresponds to agents putting less weight on initial beliefs), the disagreement distance decreases.

**Proof of Proposition A.8**: Let us drop the arguments $\mathbf{A}(\mathbf{P},\mathbf{n})$ and keep in mind that all matrices are random variable.s Defining $\mathbf{S}=\mathbf{T}^\infty$, we can write $\mathbf{c}$ as

$$\mathbf{c}=\left[(1-\alpha)\sum_{r=0}^{t-1}(\alpha\mathbf{S})^r+(\alpha\mathbf{S})^t\right]\mathbf{b}(0).$$

Define $\mathbf{V}=\mathbf{T}-\mathbf{S}$, and note that, since $\mathbf{S}$ is the summand of the spectral decomposition of $\mathbf{T}$ corresponding to eigenvalue 1, we have $\mathbf{T}^r=\mathbf{S}+\mathbf{V}^r$ (recall the proof of Lemma 2). Thus,

$$\|\mathbf{b}(t)-\mathbf{c}\|_{\mathbf{s}(\mathbf{A})}=\left\|\left[(1-\alpha)\sum_{r=0}^{t-1}\alpha^r\mathbf{V}^r+\alpha^t\mathbf{V}^t\right]\mathbf{b}(\mathbf{0})\right\|_{\mathbf{s}(\mathbf{A})}.$$

Now, the spectral radius of $\mathbf{V}$ is $|\lambda_2|$. From this it follows that if $\mathbf{b}(0)\in[0,1]^n$, then

$$\|\mathbf{b}(t)-\mathbf{c}\|_{\mathbf{s}(\mathbf{A})}\le(1-\alpha)\sum_{r=0}^{t-1}\alpha^r|\lambda_2|^r+\alpha^t|\lambda_2|^t.$$

Taking $t\to\infty$ furnishes the upper bound.

For the other inequality, let $\mathbf{w}$ be an eigenvector of $\mathbf{T}$ corresponding to $\lambda_2$, scaled so that $\|\mathbf{w}\|_{\mathbf{s}}^2=\underline{s}/4$. Then the maximum entry of $\mathbf{w}$ is at most $1/2$ and the minimum entry is at least $-1/2$. Consequently, if we let $\mathbf{e}$ denote the column vector of ones and define $\mathbf{b}(0)=\mathbf{w}+\mathbf{e}/2$, then $\mathbf{b}(0)\in[0,1]^n$. Then we have

$$\begin{aligned}\|\mathbf{b}(t)-\mathbf{c}\|_{\mathbf{s}(\mathbf{A})}&=\left\|\left[(1-\alpha)\sum_{r=0}^{t-1}\alpha^r\mathbf{V}^r+\alpha^t\mathbf{V}^t\right]\mathbf{b}(\mathbf{0})\right\|_{\mathbf{s}(\mathbf{A})}\\&=\left[(1-\alpha)\sum_{r=0}^{t-1}\alpha^r\lambda_2^r+\alpha^t\lambda_2^t\right]\frac{\underline{s}}{4}.\end{aligned}$$

Once again, taking $t\to\infty$ furnishes the lower bound. ∎

While Proposition A.8 shows that the persistence of disagreement in this model is closely related to the extent of homophily, it may be desirable to have a closer analog of Theorem 1, in which the convergence *speed* is related to the spectral homophily. The following result shows that this is possible. Let $\mathbf{b}(\infty)$ denote $\lim_{t\to\infty}\mathbf{b}(t)$.

**Proposition A.9.** For any $t \geq 0$,

$$\frac{1-\alpha}{1-\alpha|\lambda_2|} \cdot (\alpha|\lambda_2|)^t \cdot \frac{\underline{s}}{4} \leq \sup_{\mathbf{b}(0)\in[0,1]^n} \|\mathbf{b}(t) - \mathbf{b}(\infty)\|_{\mathbf{s}(\mathbf{A})} \leq \frac{1-\alpha}{1-\alpha|\lambda_2|} \cdot (\alpha|\lambda_2|)^t,$$

where $\lambda_2(\mathbf{T})$ is the second-largest eigenvalue in magnitude of $\mathbf{T}$, and $\underline{s}$ is the smallest entry of the vector $\mathbf{s}(\mathbf{A})$ defined by $\mathbf{s}(\mathbf{A}) = \left(\frac{d_1(\mathbf{A})}{D(\mathbf{A})}, \dots, \frac{d_n(\mathbf{A})}{D(\mathbf{A})}\right)$.[3]

Although we can no longer discuss the speed of convergence to a consensus, as one is no longer reached, we can still measure the speed at which beliefs converge to their limit, or more precisely, how big the differences between the current beliefs and the limit beliefs are over time. The worst-case difference between the step $t$ beliefs and the limiting ones (which is what the quantity

$$\sup_{\mathbf{b}(0)\in[0,1]^n} \|\mathbf{b}(t) - \mathbf{b}(\infty)\|_{\mathbf{s}(\mathbf{A})}$$

measures) is decaying in $\alpha|\lambda_2|^t$. Once again, under the regularity conditions of Definition 3, we can drop the $\underline{s}$, replace $\lambda_2(\mathbf{T})$ by $h^{\text{spec}}(\mathbf{P}, \mathbf{n})$, and get upper and lower bounds that match except for a constant. The techniques to do this are the same as those of Proposition A.5. Thus, the rate of convergence is again determined by spectral homophily, with faster convergence when it is lower. However, the quantity that determines the rate is also proportional to $\alpha$, meaning that the less weight agents place on their initial beliefs, the faster convergence is.

**Proof of Proposition A.9**: Earlier we noted that

$$\mathbf{b}(t) = \left[(1-\alpha)\sum_{r=0}^{t-1}(\alpha\mathbf{T})^r + (\alpha\mathbf{T})^t\right]\mathbf{b}(0).$$

Define $\mathbf{V} = \mathbf{T} - \mathbf{S}$, where $\mathbf{S}$ is, as in Proposition A.8, the first (eigenvalue 1) term in the spectral decomposition of $\mathbf{T}$, which is also equal to $\mathbf{T}^\infty$. Then we have

$$\mathbf{b}(\infty) - \mathbf{b}(t) = \left[(1-\alpha)\sum_{r=t}^{\infty}(\alpha\mathbf{V})^r\right]\mathbf{b}(0).$$

Using the same spectral arguments as in the proof of Lemma 2, and using the fact that the spectral norm of $\mathbf{V} - \mathbf{I}$ is at most 2, we find that for any $\mathbf{b}(0) \in [0,1]^n$

$$\|\mathbf{b}(t) - \mathbf{b}(\infty)\|_{\mathbf{s}(\mathbf{A})} \leq (1-\alpha)\sum_{r=t}^{\infty}(\alpha|\lambda_2|)^r = \frac{1-\alpha}{1-\alpha|\lambda_2|} \cdot (\alpha|\lambda_2|)^t,$$

where we are dropping the argument $\mathbf{T}(\mathbf{A})$ on the eigenvalue $\lambda_2$.

For the other inequality, let $\mathbf{w}$ be an eigenvector of $\mathbf{T}$ corresponding to $\lambda_2$, scaled so that $\|\mathbf{w}\|_{\mathbf{s}}^2 = \underline{s}/4$. Then the maximum entry of $\mathbf{w}$ is at most $1/2$ and the minimum entry is at least $-1/2$. Consequently, if we let $\mathbf{e}$ denote the column vector of ones and define $\mathbf{b}(0) = \mathbf{w} + \mathbf{e}/2$,

[3] Recall the $\|\cdot\|$ notation from Section II.D.4.

then $\mathbf{b}(0) \in [0,1]^n$. Then, using the same techniques as in Lemma 2, we have

$$\begin{aligned}
\|\mathbf{b}(t) - \mathbf{b}(\infty)\|_{\mathbf{s}(\mathbf{A})} &= \left\| \left[ (1-\alpha) \sum_{r=t}^{\infty} (\alpha \mathbf{V})^r \right] \mathbf{b}(0) \right\|_{\mathbf{s}(\mathbf{A})} \\
&= (1-\alpha) \sum_{r=t}^{\infty} (\alpha|\lambda_2|)^r \cdot \frac{s}{4} \\
&= \frac{1-\alpha}{1-\alpha|\lambda_2|} \cdot (\alpha|\lambda_2|)^t \cdot \frac{s}{4}.
\end{aligned}$$

This completes the proof. ∎