Online Appendix: Not for Publication

APPENDIX E: RGG-ER MIXTURES

For the given metric space (Ω, d), we denote $B(i, r)$ to be the open ball centered at $i \in \Omega$ with radius $r > 0$. The model is *Euclidean* if $\Omega \subseteq \mathbb{R}^h$ is an open set and $d(i,j) := \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$. The results in this section uses an Euclidean model with $h = 2$ and uniform Poisson intensity; $f(i) = 1$ for all $i \in \Omega$. However, all results are easily generalizable for any intensity function f, and non-Euclidean models (we clarify this below) with higher dimensions. For any measurable $A \subseteq \Omega$ define the random variable $n_A = \{\text{number of nodes } i \in A\}$. The Poisson point process assumption implies that $n_A \sim \text{Poisson}(\lambda\mu(A))$, where $\mu(\cdot)$ is the Borel measure over $\mathbb{R}^h$. For any potential node $j \in \Omega$, define $d_j(A) := \{\text{number of links } j \text{ has with nodes } i \in A\}$. $d_j = d_j(\Omega)$ denotes the total number of links j has (i.e., its degree).

Define $\boldsymbol{\nu} := \mathbb{E}\{n_A\}$ with $A = B(i,r)$ as the expected number of nodes in a "local neighborhood," which is $\boldsymbol{\nu} := \lambda\pi r^2$ in the Euclidean model with $h = 2$.[36] Define also the volume of Ω simply as its measure; i.e., $\text{vol}(\Omega) := \mu(\Omega)$. It is also useful to define $\omega := \text{vol}(\Omega)/\boldsymbol{\nu}$, so that the expected number of nodes on the graph can be expressed as $\mathbb{E}[n_\Omega] = \lambda\text{vol}(\Omega) = \boldsymbol{\nu} \times \omega$.

A *local clan* is a non-trivial clan $C \subset V$ (i.e., with $\#C \geq 2$) where the probability of a link forming between any pair $\{i,j\} \in C$ is α. A necessary condition for C to be a local clan is that $C \subset L := B\left(i, \frac{r}{2}\right)$ for some $i \in \Omega$. With the above definitions, $C \subseteq L$ is a local clan if $\#C \geq 2$ and, for all $j \in C$, $d_j(L) \geq d_j(\Omega \backslash L)$.

For a given $j \in L$, and a given number of nodes N_L in L, the number of links j has to other nodes in L is distributed $\text{Binomial}(\alpha, N_L)$. Since $N_L \sim \text{Poisson}(\lambda\mu(L))$ then $d_j(L) \sim \text{Poisson}(\alpha \times \nu(L))$, where $\nu(L) = \lambda\mu\left[B\left(i, \frac{r}{2}\right)\right] = \lambda\pi r^2/4$.[37]

Define $H_j := B(j,r)\backslash L$ and $K_j := \Omega\backslash H_j$. Because of the assumptions on the mixed RGG model,

$$\mathbb{P}(j \text{ has link with given } i \in L) = \mathbb{P}(j \text{ has link with } i \in H_j) = \alpha,$$

and

$$\mathbb{P}(j \text{ has link with given } i \in K_j) = \beta.$$

Also, because these sets are disjoint and $\Omega = L \cup H_j \cup K_j$, $d_j = d_j(L) + d_j(H_j) + d_j(K_j)$, and $d_j(L), d_j(H_j)$ and $d_j(K_j)$ are independent Poisson random variables, with distributions

$$d_j(L) =_d d(L) \sim \text{Poisson}\left(\alpha \times \frac{\boldsymbol{\nu}}{4}\right), \tag{E.1}$$

$$d_j(H_j) =_d d(H_j) \sim \text{Poisson}\left(\alpha \times \frac{3}{4}\boldsymbol{\nu}\right), \text{ and} \tag{E.2}$$

$$d_j(K_j) =_d d(K_j) \sim \text{Poisson}(\beta \times (\omega - 1)\boldsymbol{\nu}), \tag{E.3}$$

where "$=_d$" stands for equality in distribution, using that **(a)** uniform intensity implies $d_j(A) = d(A)$ for any

[36] If $h > 2$, $\boldsymbol{\nu} := \lambda \times \left(R\sqrt{\pi}\right)^h / \Gamma(1 + h/2)$

[37] If $h \geq 2$, we have that in the euclidean model, $\lambda\mu(L) = 2^{-h}\boldsymbol{\nu}$.

A, **(b)** $N_{H_j} \sim$ Poisson $(\lambda\mu(H_j))$ and $N_{K_j} \sim$ Poisson $(\lambda\mu(K_j))$, where $\mu(H_j) = \mu(B(j,r)) - \mu(B(i,r/2)) = (3/4)\pi r^2$ and $\mu(K_j) = \text{vol}(\Omega) - [\mu(H_j) + \mu(L)] = (\omega - 1)\pi r^2$. It is useful to work with the random variable

$$d_j(\Omega \backslash L) := d_j(H_j) + d_j(K_j) =_d d(\Omega \backslash L) \sim \text{Poisson}\left[\boldsymbol{\nu}\left(\frac{3}{4}\alpha + (\omega - 1)\beta\right)\right]. \tag{E.4}$$

The goal of this section is to provide lower and upper bounds for the event

(E.5)

$$\mathcal{B}_L := \{g = (V,E) : C = V \cap L \text{ is a local clan}\} = \left\{g = (V,E) : \#C \geq 2 \text{ and } \bigwedge_{j \in C} \{d_j(L) \geq d_j(\Omega \backslash L)\}\right\}$$

the problem of course being that, even though $d_j(L)$ and $d_j(\Omega \backslash L)$ are independent, the same variables across agents $j \in C$ may very well not be (and usually will not be).

Given $i \in \Omega$ and $\hat{r} > r$, an *annulus* $\mathbf{An}(i,r,\hat{r}) \subset \Omega$ is the ring between the outer ball for radius r' and the inner ball with radius r, i.e.,

$$\mathbf{An}(i,r,\hat{r}) := \left\{j \in \Omega : \ r \leq d(i,j) < r'\right\} = B(i,\hat{r}) \backslash B(i,r).$$

The most important fact to keep in mind for Proposition A.1 is that, in the Euclidean model, the distributions of $d_j(L), d_j(G_j)$ and $d_j(H_j)$ given by equations E.1, E.2 and E.3 **do not depend on the chosen node** j. This is the key property that allows us to obtain bounds on the probability of the existence of clans that do not depend on the particular nodes drawn in L.

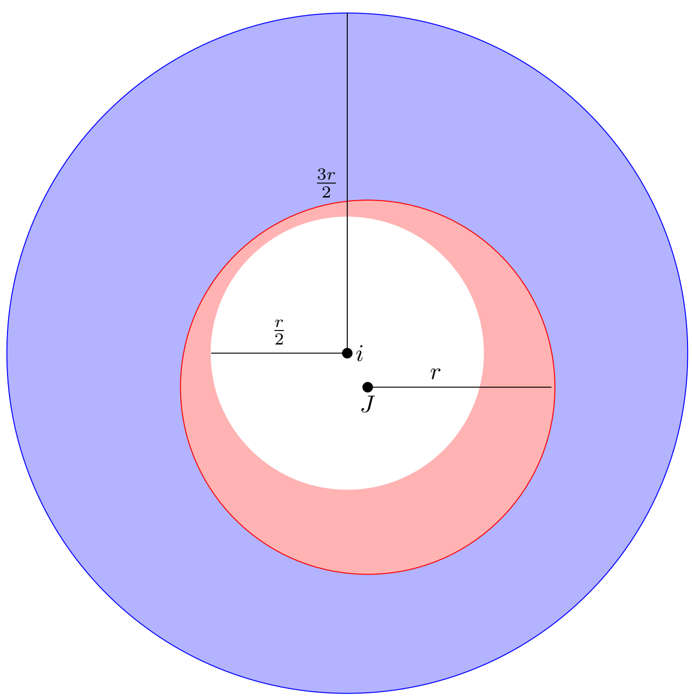

Figure 7: $H_J = B(J,r) \setminus B(i,\frac{r}{2})$; $H^* = B(i,\frac{3r}{2}) \setminus B(i,\frac{r}{2})$; $L = B(i,\frac{r}{2})$

Proof of Proposition A.1 : We develop some notation. We denote $d_j(A) \mid V$ as the number of nodes j has with nodes $i \in V \cap A$, conditional on a realized, finite set of nodes $V \subseteq \Omega$. Also, if X, Y are random variables, we use "$X \trianglerighteq Y$" to denote first order stochastic dominance of X. Let $H^* = \mathbf{An}\left(i, \frac{r}{2}, \frac{3}{2}r\right)$ and $K^* := \Omega \backslash \{L \cup H^*\}$. Conditional on a realization of V, define $d^* \mid V$ as the the number of links that a potential node would have if it has a probability α of forming links with nodes in H^*(i.i.d across nodes in H^*), and a probability β of forming links with nodes in K^* (again, i.i.d across nodes in K^*). This can be sumarized as

$$d^* \mid V = \text{Binomial}(\alpha, n_{H^*}) + \text{Binomial}(\beta, n_{K^*}), \tag{E.6}$$

where $n_A := \#\{V \cap A\}$ is the number of realized nodes in set $A \subseteq \Omega$.[38] Equation E.6 also implies that, integrating over V, we get that $d^* \sim \text{Poisson}(\alpha\nu(H^*) + \beta\nu(K^*))$. This implies

$$\mathbb{E}(d^*) = \alpha\nu(H^*) + \beta\nu(K^*) = \lambda\left\{\alpha\mu\left(\mathbf{A}\left(i, \frac{r}{2}, \frac{3}{2}r\right)\right) + \beta\left(\mu[\Omega] - \mu(L) + \mu\left(\mathbf{A}\left(i, \frac{r}{2}, \frac{3}{2}r\right)\right)\right)\right\} =$$

$$\lambda\left\{\alpha\left(\frac{9}{4} - \frac{1}{4}\right)\pi r^2 + \beta\left(\omega\pi r^2 - \frac{1}{4}\pi r^2 - \left(\frac{9}{4} - \frac{1}{4}\right)\pi r^2\right)\right\} =$$

$$\lambda\pi r^2\left[2\alpha + \beta\left(\omega - \frac{9}{4}\right)\right] = \left[2\alpha + \beta\left(\omega - \frac{9}{4}\right)\right] \times \boldsymbol{\nu}$$

using that $\mu[\mathbf{A}(i, r, \hat{r})] = \left(\hat{r}^2 - r^2\right)\pi$ and $\mu(\Omega) = \omega\lambda\pi r^2$ by the definition of ω, as we have seen above.

We first show the lower bound A.1 in 5 steps.

Step 1: *For any $j \in L$, $H_j \subseteq H^*$.*

To show this, first we show that $B(j, r) \subseteq B\left(i, \frac{3}{2}r\right)$. Take $x \in B(j, r)$, so that $d(j, x) < r$. Then $d(x, i) < d(x, j) + d(j, i) < r + \frac{1}{2}r = \frac{3}{2}r$ using that $j \in B\left(i, \frac{r}{2}\right)$. Then $H_j = B(j, r) \backslash B(i, r/2) \subseteq B\left(i, \frac{3}{2}r\right) \backslash B(i, r/2) = A^*$, as we wanted to show.

Step 2: *For any realization of V, have $d^* \mid V \trianglerighteq d_j(\Omega \backslash L) \mid V$ for all $j \in C = V \cap L$.*

We provide a more heuristic proof for this statement. Define $K^* = \Omega \backslash \{L \cup H^*\}$. Because for all $j \in C$, $H_j \subseteq H^*$ and also $K_j \supseteq K^*$. Defining $Z_j := H^* \backslash H_j$ we can decompose $\Omega \backslash L$ as

$$\Omega \backslash L = H_j \cup Z_j \cup K^*,$$

which are disjoint sets. Now, according to the RGG model, a node j has a probability α to make a link with any node in H_j since $H_j = B(j, r) \backslash L$, but has probability $\beta \leq \alpha$ to make a link with nodes in $Z_j \cup K^*$. Therefore, conditional on V,

$$d_j(\Omega \backslash L) \mid V \sim \text{Binomial}\left(\alpha, n_{H_j}\right) + \text{Binomial}\left(\beta, n_{Z_j}\right) + \text{Binomial}(\beta, n_{K^*}),$$

where we use the fact that $Z_j \cup K^* = K_j$ and hence j has probability β of making successful links there. Meanwhile, for d^*,

$$d^* \mid V \sim \text{Binomial}\left(\alpha, n_{H_j}\right) + \text{Binomial}\left(\alpha, n_{H_j}\right) + \text{Binomial}(\beta, n_{K^*}),$$

since $Z_j \subseteq H^*$. Therefore, since $\alpha \geq \beta$, $d^* \mid V \trianglerighteq d_j(\Omega \backslash L) \mid V$

[38]We use the convention that, if $n_A = 0$, then $\text{Binomial}(\gamma, n_A) = 0$ with probability 1, for any $\gamma \in [0, 1]$).

Step 3: *Suppose we condition on the realized subgraph* $g_C = (C, E_C)$*. Then,*

$$\mathbb{P}\left(\mathcal{B}_L \mid g_C\right) \geq \prod_{j \in C} F^*\left[d_j\left(L\right)\right],$$

where $F^*\left(\cdot\right)$ *is the cdf of* d^**.*

Given g_C, the in-degrees $d_j\left(L\right)$ are known values. $y_j := d_j\left(\Omega \backslash L\right)$ are independent random variables, conditional on the realization of C, since **(a)** they are independent conditional on V and **(b)** the realization of nodes in $\Omega \backslash L$ is independent of g_C. Therefore, conditioning on both $g_C = (C, E_C)$ and $V \backslash C$ (i.e., taking expectations over the links with nodes in $V \backslash C$),

$$\text{(E.7)} \qquad \mathbb{P}\left\{g : \bigwedge_{j \in C} \left\{d_j\left(L\right) \geq d_j\left(\Omega \backslash L\right)\right\} \mid g_C, V \backslash C\right\} = \prod_{j \in C} \mathbb{P}\left[g : d_j\left(\Omega \backslash L\right) \leq d_j\left(L\right) \mid V \backslash C, g_C\right] \geq$$

$$\geq \prod_{j \in C} \mathbb{P}\left[d^* \leq d_j\left(L\right) \mid V\right],$$

where we use the fact that $d^* \mid V \trianglerighteq d_j\left(\Omega \backslash L\right) \mid V$ for all $j \in C$. Moreover, (n_{H^*}, n_{K^*}) are sufficient statistics for the conditional distribution of d^*, and hence $d^* \mid V$ is also independent of C and of $d_j\left(L\right)$ for all $j \in C$ (and hence $d^* \mid V = d^* \mid (V \backslash C)$). Therefore, using E.7 and taking expectations over $V \backslash C$,

$$\mathbb{P}\left(\mathcal{B}_L \mid g_C\right) = \mathbb{E}_{V \backslash C}\left\{\mathbb{P}\left\{g : \bigwedge_{j \in C} \left\{d_j\left(L\right) \geq d_j\left(\Omega \backslash L\right)\right\} \mid g_C, V \backslash C\right\}\right\} \geq$$

$$\mathbb{E}_{V \backslash C}\left\{\prod_{j \in C} \mathbb{P}\left[g : d_j\left(\Omega \backslash L\right) \leq d_j\left(L\right) \mid V \backslash C, g_C\right]\right\} = \prod_{j \in C} \mathbb{E}_{V \backslash C}\left\{\mathbb{P}\left[g : d_j\left(\Omega \backslash L\right) \leq d_j\left(L\right) \mid V \backslash C, g_C\right]\right\} \geq$$

$$\mathbb{E}_{V \backslash C}\left\{\prod_{j \in C} \mathbb{P}\left[d^* \leq d_j\left(L\right) \mid V \backslash C\right]\right\} = \prod_{j \in C} \mathbb{E}_{(n_{H^*}, n_{K^*})}\left[\mathbb{P}\left(d^* \leq d_j\left(L\right) \mid n_{H^*}, n_{K^*}\right)\right] =$$

$$\prod_{j \in C} F^*\left[d_j\left(L\right)\right],$$

where we use the independence of y_j conditional on g_C, the fact that (n_{H^*}, n_{K^*}) are sufficient statistics for $d^* \mid (V \backslash C)$, and that F^* is the cdf of the unconditional Poisson distribution of d^* that we derived above.

Step 4: *Given* $n_L = \#\left\{C\right\} \geq 2$*, we have that*

$$\mathbb{P}\left(\mathcal{B}_L \mid n_L\right) \geq F^*\left(n_L - 1\right)^{n_L} \times \alpha^{n_L(n_L-1)/2}$$

.

Given n_L, we want to get a lower bound on the probability that, for n_L independent random draws $d_j^* \sim \text{Poisson}\left[\left(2\alpha + \left(\omega - \frac{9}{4}\right)\beta\right) \times \boldsymbol{\nu}\right]$, we have that $d_j\left(L\right) \geq d_j^*$ for all $j \in C$. One of these potential subgraphs is a clique, where $d_j(C) = n_L - 1$ for all $j \in C$. Since $g_C \mid n_L$ is an Erdos-Renyi graph with parameter $p = \alpha$, we know that the probability that g_C is a clique is $\alpha^{\binom{n_L}{2}} = \alpha^{n_L(n_L-1)/2}$. Therefore,

$$\mathbb{P}\left\{g : \bigwedge_{j\in C} \{d_j(L) \geq d_j(\Omega\backslash L)\} \mid n_L\right\} = \sum_{g_C:\#C=n_L} \mathbb{P}\left\{g : \bigwedge_{j\in C} \{d_j(L) \geq d_j(\Omega\backslash L)\} \mid g_C\right\} \times \mathbb{P}(g_C \mid n_L) \geq$$

$$\sum_{g_C:\#C=n_L} \prod_{j\in C} F^*[d_j(L)] \times \mathbb{P}(g_C \mid n_L) \geq \left[\prod_{j\in C} F^*(n_L - 1)\right] \times \alpha^{n_L(n_L-1)/2} =$$
$$F^*(n_L-1)^{n_L} \times \alpha^{n_L(n_L-1)/2}$$

Step 5: $\mathbb{P}(\mathcal{B}_L) \geq \sum_{n=2}^{\infty} \left(\frac{\boldsymbol{\nu}}{4}\right)^n \frac{e^{-\boldsymbol{\nu}/4}}{n!} F^*(n-1)^n \times \alpha^{n(n-1)/2}$.

The previous result implies $\mathbb{P}(\mathcal{B}_L) \geq \mathbb{P}(n_L \geq 2) \times \mathbb{E}_{n_L}\left\{F^*(n_L-1)^{n_L} \times \alpha^{n_L(n_L-1)/2} \mid n_L \geq 1\right\}$. The fact that $n_L \sim \text{Poisson}(\boldsymbol{\nu}/4)$ gives us the desired expression.

Step 6: $\mathbb{P}(\mathcal{B}_L) \leq \sum_{d=1}^{\infty} \left(\alpha\frac{\boldsymbol{\nu}}{4}\right)^d \frac{e^{-\alpha\boldsymbol{\nu}/4}}{n!} \hat{F}(d)$.

The upper bound comes simply from the fact that the event $\mathcal{B}_L \subseteq \{d_j(L) \geq d_j(\Omega\backslash L)\}$ for a particular $j \in C$. Since $d_j(A) =_d d(A)$ for any $j \in \Omega$ and any measurable $A \subseteq \Omega$ we obtain the upper bound using E.4. For it to be a local clan, we need $n_L \geq 2$ which necessarily implies that $d_j(L) \geq 1$ for any $j \in C$. Moreover, since $d_j(\Omega\backslash L)$ is independent of both n_L and $d_j(L)$, we get that

$$\mathbb{P}(\mathcal{B}_L) \leq \mathbb{P}\{d_j(\Omega\backslash L) \leq d_j(L) \mid d_j(L) \geq 1\} = \mathbb{P}\left[\hat{F}(d_j(L)) \mid d_j(L) \geq 1\right] =$$
$$\sum_{d=1}^{\infty} \left(\alpha\frac{\boldsymbol{\nu}}{4}\right)^n \frac{e^{-\alpha\boldsymbol{\nu}/4}}{d!} \hat{F}(d)$$

where $\hat{F}$ is the cdf of $d_j(\Omega\backslash L) \sim \text{Poisson}\left[\left(\frac{3}{4}\alpha + (\omega-1)\beta\right) \times \boldsymbol{\nu}\right]$, and using the fact that $d_j(L) \sim \text{Poisson}(\alpha\boldsymbol{\nu}/4)$. *Q.E.D.*

An obvious corollary of this Proposition is that, for $\alpha \in (0,1)$, the probability of finding a local clan is strictly positive in any local neighborhood over Ω.

APPENDIX F: GENERALIZING THE COARSE DEGROOT MODEL

In this section we generalize our environment where individuals can send finer beliefs, though not continuous messages. The naive agents are still DeGroot in the sense that they average their neighbors' and own prior messages and transmit a coarse estimate of their updated belief (finer than the binary case but more granular than the continuous case). We generalize the notion of clan in the appropriate way and show that our main result, Theorem 1, directly extends. An incomplete information model of Bayesian and coarse DeGroot agents still is asymptotically inefficient in that there are misinformation traps if the share of such clans is non-vanishing.

F.1. *Generalizing Clans*

We generalize our setup to allow for a richer set of messages. We maintain the assumptions on the i.i.d. nature of signals. Let $\mathbb{P}(\theta \mid s_i)$ be the (random) posterior belief over $\theta \in \{0,1\}$ given the observed signal. In the first period, agents only observe their own signal and choose $a_{i,0} = 1$ whenever $\mathbb{P}(\theta = 1 \mid s_i) > 1/2$, $a_{i,0} = 0$ whener $\mathbb{P}(\theta = 1 \mid s_i) < 1/2$, and $a_{i,0} \in \{0,1\}$ if $\mathbb{P}(\theta = 1 \mid s_i) = 1/2$. We make the assumption that $\mathbb{P}(a_{i,0} = 1 - \theta \mid \theta) > 0$ for all $\theta \in \{0,1\}$ (i.e., every agent can have a wrong guess in the first round).

The main difference relative to the model presented in the body of the paper is a richer, yet discrete, set of messages that agents can communicate with their neighbors. Formally, agents can communicate their beliefs encoded on a sparse set of messages $\mathcal{M} = \left\{0, \frac{1}{m-1}, \frac{2}{m-1}, \ldots, \frac{m-2}{m-1}, 1\right\}$, where $m \in \mathbb{N}$ is the number of messages. Our original model in the experiment has $m = 2$. If agent i has beliefs $b_{i,t} = \mathbb{P}(\theta = 1 \mid \mathcal{I}_{i,t})$ she communicates the belief

$$b^c_{i,t} := \operatorname*{argmin}_{b \in \mathcal{M}} |b_{i,t} - b| \tag{F.1}$$

Agents then form next period beliefs $b_{i,t+1}$ according to the DeGroot averaging formula:

$$b_{i,t+1} = \frac{1}{1+d_i} \sum_{j \in N_i \cup i} b^c_{j,t} \tag{F.2}$$

Finally, actions are based according to the average of communicated beliefs agent i observes:

$$a_{i,t} = \begin{cases} 1 & \text{if } b_{i,t} > \frac{1}{2} \\ 0 & \text{if } b_{i,t} < \frac{1}{2} \\ \in \{0,1\} & \text{if } b_{i,t} = \frac{1}{2} \end{cases}$$

We now generalize the notion of "clans" for arbitrarily number of messages m. We say that a set of nodes $C \subseteq V$ is γ*–cohesive* if, for all agents $i \in C$,

$$\frac{d_C(i) + 1}{d(i)} \geq \gamma.$$

That is, every agent in the set has at least a fraction γ of links to the inside of the set (including herself). We use the name γ–cohesive because its definition is nearly identical to the definition of p–cohesive sets in Morris (2000).

For any set of nodes C, we can define its cohesiveness as

$$\gamma(C) := \min_{i \in C} \frac{d_C(i) + 1}{d(i)}$$

Without loss of generality, assume that the true state of the world is $\theta = 0$. Let $b_0^{\min} = \min_{i \in C} b^c_{i,0}$ be the lowest communicated beliefs about θ held in the group at $t = 0$.

Proposition F.1 For $m \in \mathbb{N}$, let b^+_m and b^-_m be the closest beliefs in $\mathcal{M}$ to $\frac{1}{2}$ (other than itself).[39] Suppose beliefs are such that $q_{0,m} = \mathbb{P}_s\left(\mathbb{P}(\theta = 1 \mid s) > b^+_m \mid \theta = 0\right)$ and $q_{1,m1} = \mathbb{P}_s\left(\mathbb{P}(\theta = 0 \mid s) < b^-_m \mid \theta = 1\right)$ are

[39] If m is odd, then $b^+_m = \frac{1}{2} + \frac{1}{m-1}$and $b^-_m = \frac{1}{2} - \frac{1}{m-1}$. If m is even, then $b^+_m = \frac{m}{2(m-1)}$and $b^-_m = \frac{m-2}{2(m-1)}$

positive for all m. If a set $C \subseteq V$ is such that

$$\gamma(C) > \underline{\gamma}_m := \begin{cases} 1 - \frac{1}{m} & \text{if } m \text{ is even} \\ 1 - \frac{1}{m+1} & \text{if } m \text{ is odd} \end{cases} \tag{F.3}$$

then C gets stuck at time $T = 0$ with probability at least $q_m = \left(q_{0,m}^{|C|} + q_{1,m}^{|C|}\right)/2$.

It is easy to see that we can rewrite condition (F.3) as

$$d_i(C) \geq \begin{cases} (m-1) \times d_i(V \backslash C) \text{ for all } i \in C & \text{if } m \text{ is even} \\ m \times d_i(V \backslash C) \text{ for all } i \in C & \text{if } m \text{ is odd.} \end{cases} \tag{F.4}$$

We define a m*–cohesive clan* to be a subset of nodes C that satisfies the condition of (F.4).

For $m = 2$ this is just the definition of a clan as in the body of the paper. For $m > 2$, this gives us the proper generalization of clans. That is, agents in C must have a large enough fraction of their links to other nodes in C. For example, if $\mathcal{M} = \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}$ then $\gamma(C) > \underline{\gamma}_m$ if and only if $d_i(C) \geq 3d_i(V \backslash C)$ for all $i \in C$ (i.e., all nodes in C have at least 3 times as many links to nodes in C as to nodes outside of C).

Proof: For simplicity of exposition, let's assume $\theta = 0$, so agents are stuck if $a_{i,t} = 1$ for all $t \geq 1$. For a signal $s \in \mathcal{S}$, let $b(s) = \mathbb{P}(\theta = 1 \mid s)$. Without loss of generality, let us redefine signals $s = p(s)$, so that if $s_i > \frac{1}{2}$ then $a_{i,0} = 1$. Define $c_m := 1/2(m-1)$. It has the property that if $\left|b_i - \frac{k}{m-1}\right| < c_m$ then $b_i^c = \frac{k}{m-1}$.

Fix $k \leq m$ such that $\frac{k}{m-1} > \frac{1}{2}$. We want to first find conditions under which we have $b_{i,t} \geq \frac{k}{m-1}$ for all $t \geq 0$, for all $i \in C$. If this is the case, then $a_{i,t} = 1$ for all $t \geq 0$, and the set C gets stuck from the begining. First, we need that $b(s_i) = \mathbb{P}(\theta = 1 \mid s_i) > \frac{k}{m-1} - c_m$ for all $i \in C$, so that $b_{i,0}^c \geq \frac{k}{m-1}$. Then, the belief for period 1 for agent i is

$$\begin{aligned} b_{i,1} &= \frac{1}{1+d_i} \sum_{j \in N_i \cup i} b_{i,0}^c = \frac{1}{1+d_i} \sum_{j \in C} b_{i,0}^c + \frac{1}{1+d_i} \sum_{j \in N_i \cap (V \backslash C)} b_{i,0}^c \\ &\geq \frac{1}{1+d_i} \sum_{j \in C} \frac{k}{m-1} = \frac{d_i(C)+1}{1+d_i} \times \frac{k}{m-1} \geq \gamma(C) \frac{k}{m-1}. \end{aligned}$$

Therefore, if $(\gamma(C) - 1) \times \frac{k}{m-1} > -c_m$ then

$$b_{i.1} - \frac{k}{m-1} \geq (\gamma(C) - 1) \times \frac{k}{m-1} > -c_m$$

and hence $b_{i,1} > \frac{k}{m-1} - c_m$. So $b_{i,1}^c \geq \frac{k}{m-1}$ again and therefore all agents in C get stuck from this point on.

We can simplify this inequality using the definition of c_m:

$$(\gamma(C) - 1) \times \frac{k}{m-1} > -c_m \iff (1 - \gamma(C)) \times \frac{k}{m-1} < \frac{1}{2(m-1)} \iff$$

$$\gamma(C) + \frac{1}{2k} > 1 \tag{F.5}$$

which gives a lower bound of $1 - 1/2k$ on the cohesion of set C.

Observe that, if this property holds for $b(s) = \frac{k}{m-1}$, it also holds for any $b(s') = \frac{k'}{m-1}$ with $k' \geq k$. Therefore, the least restrictive condition is to have signals that are just above $\frac{1}{2}$. If m is odd, then we know that at $k = \frac{m-1}{2}$ the signal $\frac{k}{m-1} = \frac{m-1}{2(m-1)} = \frac{1}{2}$, so the least message that is larger than $\frac{1}{2}$ is $\frac{1}{2}+\frac{1}{m-1} = \frac{m+1}{2(m-1)}$, and hence the least signal corresponds to $k_m^* := (m+1)/2$. So, if m is odd, condition (F.5) written at the least message strictly greater than half is

$$\gamma(C) + \frac{1}{2\left(\frac{m+1}{2}\right)} > 1 \iff \gamma(C) > 1 - \frac{1}{m+1}.$$

If m is instead even, we know that the least message greater than half is $\frac{m}{2(m-1)}$, so $k_m^* = m/2$ and condition (F.5) evaluated at this message is

$$\gamma(C) + \frac{1}{2\frac{m}{2}} > 1 \iff \gamma(C) > 1 - \frac{1}{m}.$$

To summarize, if we define $\underline{\gamma}_m = 1 - m^{-1}$ when m is odd and $= 1 - (1+m)^{-1}$ when even, then if

1. $p(s_i) = \mathbb{P}(\theta = 1 \mid s_i) > \frac{k_m^*}{m-1} - c_m$ for all $i \in C$
2. $\gamma(C) > \underline{\gamma}_m$

we have $a_{i,t} = 1$ for all $t \geq 0$.

Finally, we check the probability of this happening. Notice that in both cases we have $p(s_i) \geq b_m^+$ so agents get stuck on 1 when the true state was $\theta = 0$ with probability at least $q_{0,m}^{|C|}$. Redoing the argument for when the true state is $\theta = 1$ finishes the proof. *Q.E.D.*

F.2. *Addendum to Theorem 1*

We can now extend Theorem 1 to the case of coarse DeGroot with message space $\mathcal{M}$. A similar result follows, but now for the case of m–cohesive clans rather than 1–cohesive clans as in the body of the paper.

THEOREM F.1 Suppose $G_n = (V_n, E_n)$ with $|V_n| = n$ is such that signals are i.i.d. across agents, and either (1) signals are binary, with $\mathbb{P}(s = \theta \mid \theta) = p > \frac{1}{2}$ or (2) posterior distributions $\mathbb{P}(\theta \mid s_i)$ are non-atomic in s for $\theta \in \{0,1\}$. Take the incomplete information model, where non-Bayesian agents are coarse DeGroot types with m messages. Then the incomplete information model may not be asymptotically efficient.

In particular, suppose there exist $k < \infty$ such that such that

$$X_n^{(m)} := \#\left\{i \in V_n : i \text{ is in a set } C \text{ of size } k : \gamma(C) > \underline{\gamma}_m\right\}/n$$

is positive in the limit. Then the model is not asymptotically efficient.

PROOF: This is identical to proof of Theorem 1. The only difference is in the definition of "acceptable" sets. Instead of (1–cohesive) clans used in Theorem 1, we use $\underline{\gamma}_m$ − cohesive sets, which are themselves m–cohesive clans, so $X_n^{(m)} \leq X_n$ from Theorem 1. *Q.E.D.*

APPENDIX G: CLANS ARE ESSENTIAL FOR STUCKNESS

Theorem 1 shows that if a sequence of networks has a non-negative fraction of agents in clans, then a non-vanishing fraction of nodes that get stuck. We show that the converse is also true: if a node is stuck, then this node is necessarily a part of a clan whose members got stuck as well. Thus, clans are essential for stuckness.

Suppose there is a network G with n agents who receive signals $s \in \{0,1\}^n$. Agents form beliefs and take actions according to the binary coarse DeGroot model.

The learning process converges (under signals $s \in \{0,1\}^n$) if there exists $T \in \mathbb{N}$ and $a^\infty(s) \in \{0,1\}$ such that

$$a_{i,t} = a_i^\infty(s) \text{ for all } i \in V, t \geq T.$$

PROPOSITION G.1 Take a network G and initial signals $s \in \{0,1\}^n$ such that the learning process converges to $a^\infty(s)$. If $i \in V$ gets stuck, that is $a_i^\infty = 1-\theta$, then there is a set $C_i \subseteq V$ such that

1. C_i is a clan,
2. $i \in C_i$,
3. the clan C is stuck: $\forall j \in C_i$ we have $a_j^\infty(s) = 1-\theta$.

So if there is a stuck node in the limit, then i is a member of a clan that becomes stuck.

PROOF: The proof is rather trivial. Suppose learning converges. Now, for $a_i^\infty(s) = 1-\theta$ we need that most of i's neighbors also get stuck. Define

$$D(i) := \left\{ j_1 \in V : (ij_1) \in E \text{ and } a_{j_1}^\infty(s) = 1-\theta \right\}.$$

Therefore, every node $j_2 \in D(i)$ is also stuck. Using the same argument, there must exist (for each $j_1 \in D(i)$) another set of stuck nodes connected to them:

$$D(j_1) = \left\{ j_2 \in V : (j_1 j_2) \in E \text{ and } a_{j_2}^\infty(s) = 1-\theta \right\}.$$

Now, the sets $D(i)$ and $D(j_1)$ may not be disjoint: if C_i is a stuck clique they would be identical. Following in this fashion, recursively, we can define the clan C_i as just the unions of all these sets

$$C_i = D(i) \cup \left(\bigcup_{j_1 \in D(i)} D(j) \cup \left(\bigcup_{j_2 \in D(j_1)} D(j_2) \ \ldots \right) \right)$$

Notice that by construction C_i is a clan comprised of stuck nodes and $i \in C$, finishing the proof. *Q.E.D.*

APPENDIX H: PROTOCOL OVERVIEW

Researchers welcome the participants. Provide each person an identification number 1-7 and explain to them that for the purposes of this game, they will be associated with this number.

Today we will play three short games in which you will have the opportunity to earn some money.

H.1. *Introduction*

Bags

In this game, there will be two identical bags. *Show bags.*

The first bag has 5 blue balls and 2 yellow balls. The second bag has 5 yellow balls and 2 blue balls. Therefore, the first bag will be called the “blue bag” and the second bag will be called the “yellow bag.”

Check comprehension of the participants. To check for comprehension of the participants, mix the two bags and randomly pick one. Then, take all the balls out and ask one of the participants at random to identify the bag that was randomly chosen.

Objective of the game

As shown, before starting each game we will mix up the bags and randomly select one. We will use this bag for the entire game. The objective of this experiment is that in each round of each game you guess correctly which bag was randomly chosen for that game. That is, in each round of the game, each of you will make a new guess. For this we will provide you some information in a sequential way.

Game

In the first round, we will privately show each of you a ball which was drawn from the chosen bag. This one will be your first guess of the color of the bag. We will go to each of you, randomly draw one ball, and show it to you only. Then, we will put that ball back in bag, go to the next person, randomly draw a new ball, and show it only to him or her.

This way nobody but you will be able to see the color of that ball. Therefore, each of you will see only the color of one ball drawn from the same chosen bag.

Do demonstrations.

Draw a ball of the color of the bag from a bag few times for different people and make sure they realize that, given that is all they know, the optimal thing to do is to guess that the bag is of the color of the ball that was drawn.

Further, from the second round on, we will show you what your neighbors' guesses in the previous round were. This extra information should help you to make a better guess of the color of the chosen bag because your neighbors probably got a different ball drawn from the same bag.

How we will decide who is whose neighbor and how we will show you their previous round guesses?

To show you who are your neighbors in each game, we will give each of you these maps. *Give to each of them an empty demo network topology as in Figure 8.*

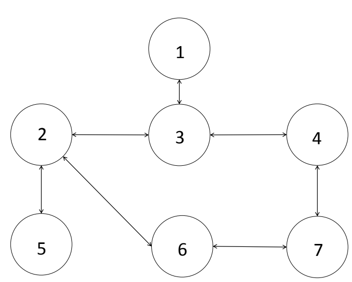

Figure 8: Demo network topology

This map shows people 1-7. Can you find yourself on the map? *Ask each participate to locate their batch number in the map.*

The arrows between numbers on the map show who your neighbors are. *Ask each participate to locate their neighbors in the map.*

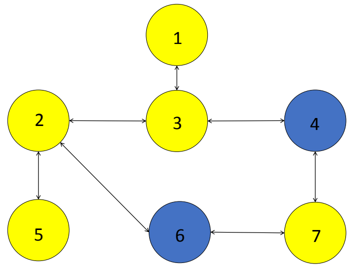

Figure 9: Demo example

To see how we will show you neighbors' previous round guess consider the following example. *Show the demo example in Figure 9.*

Imagine that the true bag is the yellow one. In the example, 5 people got yellow balls in the beginning of the game (individuals 1,2,3,5 and 7) and 2 got blue balls (4 and 6). Naturally, the first guesses of the individuals that got the yellow ball is that the bag is the yellow one, and for the ones that got the blue ball, they guess that the bag is the blue one.

Then, before you have to make a guess in Round 2 each of you will receive the following sheet. *Hand out to each participant the example of their initial signals. For individual 4, this is shown in Figure 10.*

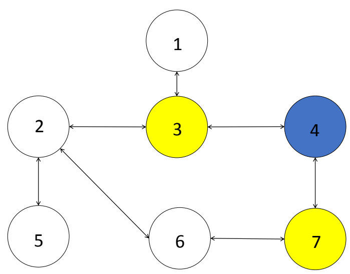

Figure 10: Example of initial signals for individual 4

As you can see, each sheet has colored the guess that your neighbors made. However, you cannot see the guess of those individuals that are not your neighbors.

So, in Round 2, for example, whose Round 1 guess can person 1 see and how are these? What about person 2?

Demonstrate and quiz for comprehension.

Teaching them how they can extract information from their neighbors to make an educated guess of the bag color

Now, let us see how you can use the information from these maps to make a guess in this round. First, look at individual 4 (*see Figure 10*): he got a blue ball, but in the second round, he realizes that individuals 3 and 7 guessed yellow balls, which indicates that they likely got yellow balls. Therefore, given the information he has, individual 4 guesses that the bag is the yellow one. Likewise, individual 6 realizes that individuals 2 and 7 got yellow balls, and then he guesses that the bag is the yellow bag.

Also, see that the neighbors of individuals 1,2,3 and 5 have seen the same thing they have seen, and therefore their decision remains the same, and they guess again that the bag is yellow.

For individual 7 however, things are different (*See example in Figure 11*): he observed a yellow ball, but now he realizes that individuals 4 and 6 got blue balls in the beginning of the game. Then, given his information, individual 7 changes his guess to the blue bag.

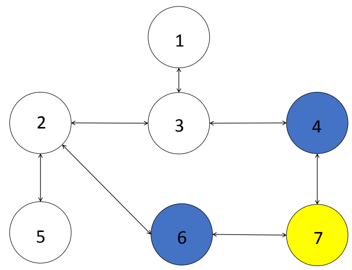

Figure 11: Example of initial signals for individual 7

All these guesses are reflected in the following sheet. *(Show example in Figure 12).*

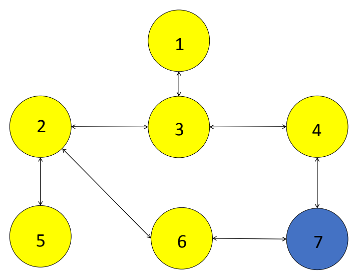

Figure 12: Example of all guesses after observing the initial signal of all neighbors

So, before individuals have to make a guess in Round 3 each of them receives the following sheet. *Hand out each people's worksheet of the Demo Network Topology for the above example of Round 2 choices. In Figure 13 is the example for individual 7.*

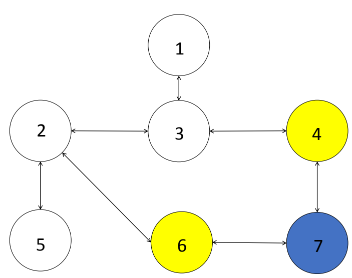

Figure 13: Example of all guesses of individual 7' neighbors after observing the signal of all their neighbors

So, we see that these sheets are like the previous round ones. Then, why would individuals make a different choice from the one that they make in the second round? They now have more information about what their neighbors saw in the first round.

For example, let us consider the case of individual 7, which is the interesting one.

In round 3, individual 7 sees that, even though individuals 4 and 6 saw a blue ball in the first round, they changed their guess to yellow in the second round. The only way this can be is if they saw at least two other individuals getting a yellow ball drawn from the bag in the first period. Thus, it has to be that individuals 2 and 3 also got a yellow ball drawn in the first period.

Then, individual 7 now knows that 2,3 and himself got yellow balls, while 4 and 6 got blue balls. Therefore, he or she should switch back to guessing that the true bag is the yellow bag. Then, in the end, everybody guesses yellow, which is the actual bag.

Demonstrate and quiz for comprehension

Payment

Remember, we will play 3 different games, each of them with some rounds. You will have a "payment bag." In each round, we will record your guess about the color of the bag and the game number on a chip. We will put that chip in the payment bag. After each game we will show the right bag. Then, at the end of the game we will draw a chip from your payment bag. If the color of the chip matches the color of the right bag of the round that is written in your chip you will earn Rs 100 in addition to your Rs 20 participation fee.

Therefore, you can see that it is very important that in each round you make your best guess about what bag you think we have used to draw to balls of the first round. Check comprehension.

Ending

The game will have an indeterminate number of rounds. On average, each game will last 5 rounds, but we could finish it in 1, 2, 3, 4, 5, 6, 7, etc. It will be very clear when a game is over and a new one begins or we are done with the experiment.

H.2. *Game structure*

Now we will play three games that are exactly as the example that we have shown you before with the only distinction that in each of these games the neighbors that you will have may be different. At the beginning of each game we will give you a map so that you can see who your neighbors are.

Remember that during these games you will not be able to talk or even look at any of the other people you are playing with. We will put you in separate corners of the room / house / building. Please cooperate in this aspect. For this experiment, it is very important that you only use the information that you are given to make your guess of the bag.

Experiment 1

So, for our first game we will shuffle the two bags and we will select one bag to play the game with. We will put aside the other bag. *Show that we randomly select one bag, put away the other bag.*

Also, we will tell you who will be your neighbors for this game. We will provide you with a map that contains this information, much like we did in the demo. *Give experimentees the Game 1 Network Topology.*

In the first round, we will come to each of you, privately, and draw a ball from the bag. You will have to make a guess of the color of the bag. Did we draw the bag from the "blue bag" or "yellow bag"? We will take a chip which has the color of the bag you guessed, write down the number of the game (1 in this case), and put this one in your payment bag.

You cannot talk and tell people the color of the ball shown to you. You alone can know the color of the ball that was shown to you in Round 1. The way you will guess is to simply touch the color blue or color yellow on the cardboard that we will show in every guess. *Show cardboard.*

After the first round, we will no longer show anyone any balls from the bag. Further, in rounds 2, 3, 4, etc., you will have to make a guess of the color of the bag. Did we draw the bag from the "blue bag" or "yellow bag"? We will pick and chip of the color of your guess, write down the number of the game (again 1) and put it in a bag.

In order to make a guess about the bag from which we drew the ball we showed you, you can use the following information. You will not only know which ball that was shown to you in Round 1 but you will also know the guess of bag that your neighbors did in previous rounds. In order to show you this information will hand over sheets like the ones we showed you in the demonstration. Until the end of each game you will keep the sheet shown to you in all previous rounds.

Remember, you cannot talk. The way you will guess is to simply touch the color blue or color yellow on the cardboard that we will show you with the guesses of your neighbors in the previous round. *Show cardboard again.*

Now we will send you to separate locations in the room. *Accompany each experimentee to his/her corresponding location in the room.*

Experimenters run the experiment

The experiment is over.

The bag where the Round 1's balls were drawn from was the [bag color] bag. *Show the bag where the Round 1's balls were drawn from.*

Experiment 2

Again, for our second game we will reshuffle the two bags and we will select one bag to play the game with. We will put aside the other bag. *Show that we select one bag, put away the other bag.*

Again we will tell you who your neighbors will be in this game. We will hand out a new map containing this information. *Give experimentees the Game 2 Network Topology.*

Game 2 will be as in game 1. In the first round we will come to each of you, privately, and draw a ball from the bag. After you guess the color of the bag, we will pick a chip with such color and write down the number of the game (2 in this case), and place the chip in your bag. .

Remember, you cannot talk and tell people the color of the ball shown to you. You alone know the color of the ball that was shown to you in Round 1.

From the second round on we will show you the guess of bag that your neighbors did in previous rounds. Using this information, you will have to make a guess of the color of the bag in each round.

Remember, you cannot talk. The way you will guess is to simply touch the color blue or color yellow on the cardboard that we will show you with the guesses of your neighbors in the previous round.

Now we will send you again to your corresponding location in the room. *Accompany subjects their corresponding location.*

Experimenters run the experiment.

The bag where the Round 1's balls were drawn from was the [bag color] bag. *Show the bag where the Round 1's balls were drawn from.*

Experiment 3

Again, for our third game we will reshuffle the two bags and we will select one bag to play the game with. We will put aside the other bag. *Show that we select one bag, put away the other bag.*

Again we will tell you who will be your neighbors for this game. We will hand out a new map containing this information. *Give experimentees the Game 3 Network Topology.*

Game 3 will be as in games 1 and 2. In the first round we will come to each of you, privately, and draw a ball from the bag. After you guess the color of the bag, we will pick a chip with such color and write down the number of the game (2 in this case), and place the chip in your bag.

Remember, you cannot talk and tell people the color of the ball shown to you. You are the only one that can know the color of the ball that was shown to you in Round 1.

From the second round on we will show you the guess of bag that your neighbors did in previous rounds. Using this information you will have to make a guess of the color of the bag in each round.

Remember, you cannot talk. The way you will guess is to simply touch the color blue or color yellow on the cardboard that we will show you with the guesses of your neighbors in the previous round.

Now we will send you again to your corresponding location in the room. *Accompany subjects their corresponding location.*

Experimenters run the experiment.

The bag where the Round 1's balls were drawn from was the [bag color] bag. *Show the bag where the Round 1's balls were drawn from.*

Payment

Line up all participants and draw a chip from each of their bags. If the color of the chip and game number written on it coincide with of color the bag that was randomly chosen for the game pay the participant INR 100 / MXN 150. Otherwise, simply play the participant INR 20 / MXN 50, which is the participation fee. Make sure all participants sign receipts acknowledging that they received the payment.