Rational DeGroot Learning

By Lyle Goodyear and Markus Mobius*

We study a game of social learning on networks in which agents choose how much attention to pay to their neighbors before engaging in DeGroot-style averaging. Each agent minimizes the variance of her long-run belief by choosing how to weight her neighbors, subject to a minimum attention constraint $\epsilon \geq 0$ on every edge. This framework occupies a middle ground between fully rational Bayesian updating and naive DeGroot updating: agents strategically adjust for the "double-counting" problem identified by DeMarzo, Vayanos and Zwiebel (2003) while retaining the simple averaging rule. We characterize the best response using a universal score vector, which ranks all agents by their marginal informational value, and all agents agree on this ranking. Intuitively, agents should pay less attention to neighbors who are connected to the most influential agents through many short paths, formalizing the double-counting intuition. Turning to equilibrium, we show that without attention constraints, every strongly connected equilibrium achieves the first-best (uniform influence), though echo chambers may persist as coordination failures. With the constraint $\epsilon > 0$, a strictly convex reformulation yields a unique equilibrium influence vector. The first-best outcome, where all agents have uniform influence, is achievable whenever $\epsilon \leq \epsilon^*$, a network-dependent threshold we characterize using a transportation problem with a natural economic interpretation in terms of supplies and demands of attention. For undirected networks, ϵ^* equals the largest feasible minimum attention level, so the first-best is always achievable. For directed networks, ϵ^* can be strictly smaller; above ϵ^*, we show that the welfare cost of forced attention on an edge and the cost of missing edges are determined by the universal score vector.

Social networks are a major mechanism through which people aggregate information and develop an understanding of the world. The DeGroot model is a canonical means of studying how people share opinions and come to consensus in a social network. In the model, agents are treated as "naive" updaters, repeatedly taking a simple weighted average of their neighbors' beliefs to form their own opinion.

DeMarzo, Vayanos and Zwiebel (2003) demonstrate that naive agents are subject to a "double-counting" bias, wherein a single source of information may reach

* Goodyear: Stanford University (lyleg@stanford.edu); Mobius: Microsoft Research (mobius@microsoft.com). We thank Matthew Jackson, Arun Chandrasekhar, Aviad Rubinstein, and Stephen Kirkland for comments and suggestions.

an agent through multiple paths, and the agent treats the information as novel each time. While real people are certainly more naive than perfect Bayesians, we argue that they are rarely as naive as the uniform-weight DeGroot updater. Human social learning heuristics appear in our day-to-day language. We may ask for a "second opinion" to diversify our streams of information, place extra emphasis on an "independent source" who shares no relationships with our friends, or discount the opinion of a neighbor influenced by an echo chamber. Humans clearly have an intuitive sense of how to weigh the opinions of their social connections.

We develop a model that occupies a middle ground between fully rational Bayesian updating and simple naive DeGroot updating. Agents observe the network structure, then simultaneously choose how much weight to place on each neighbor. Updating proceeds by DeGroot averaging with these chosen weights, fixed across periods. Each agent minimizes the variance of her long-run belief. We impose a minimum attention constraint: each agent must place at least $\epsilon \geq 0$ weight on every neighbor, reflecting the difficulty of completely ignoring a social connection. When $\epsilon = 0$, agents may sever links entirely; when $\epsilon > 0$, every edge carries positive weight. The parameter ϵ captures the strength of the social environment: colleagues in a shared office, students in a dorm, or users on a social media platform who cannot fully control their feed.

Illuminating Example

Consider agent 0 in the following undirected network, displayed in Fig. 1. Agent 0 must divide her attention between agents $\{1, 2, 4, 6\}$. Because agents 2 and 1 share a mutual connection (agent 3), agent 0 will eventually hear 3's opinion through two separate paths. On the other hand, each of agents 4 and 6 provides access to a unique source of information. Intuitively, agent 0 would therefore like to place more weight on 4 and 6 together than 1 and 2. Our framework formalizes this intuition and characterizes the optimal attention allocation for every agent in the network.

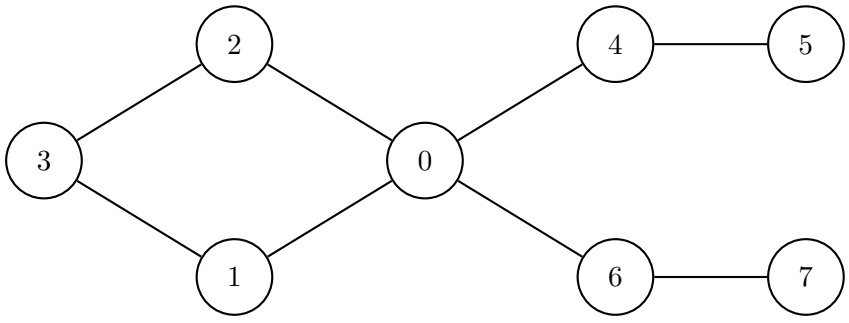


Figure 1. Example undirected network.

Overview of Results

The central object in our analysis is the *score vector*, which quantifies how informationally redundant each agent's perspective is in the current consensus: agents who are "close" (in a random-walk sense) to the most influential parts of the network have high scores, while agents who are far from the influential core have low scores. In the standard DeGroot setting on undirected graphs (where agents distribute weight equally among their neighbors), the score reduces to the negative covariance between an agent's effective resistance to a randomly sampled node and that node's influence. All agents agree on the score ranking. If two agents share two common neighbors, they agree on which of the two deserves more weight relative to the other.

Using the score vector, we characterize a single agent's best response to a fixed weighted network. An agent's optimal strategy equalizes scores among all neighbors she voluntarily listens to, and gives zero (or minimum) weight to neighbors whose scores exceed this common level. Agents shift weight away from high-score neighbors who echo the mainstream consensus and towards low-score neighbors whose perspectives are novel.

We then turn to equilibrium. In the unconstrained game ($\epsilon = 0$), agents *can* fully correct the double-counting bias. We prove that every Nash equilibrium in which the network stays strongly connected achieves the first-best outcome (uniform social influence across all agents) and that these are the only strict equilibria. Furthermore, these equilibria exist; network topology never harms accuracy in the best case. Reducible (fragmented) equilibria can also arise: when the network admits a partition into equal-size, internally connected communities with limited cross-group links, agents may sustain echo chambers in which each group achieves consensus independently with strictly higher variance.

Imposing minimum attention $\epsilon > 0$ on edges eliminates echo chambers but introduces a new source of inefficiency: agents are forced to listen to neighbors whose information is redundant. We show that for $\epsilon > 0$, the game admits a strictly convex reformulation (following Breen and Kirkland (2015)) with a unique equilibrium influence vector π^*. A network-dependent threshold ϵ^* separates two regimes. When $\epsilon \leq \epsilon^*$, the first-best is still achievable and is the unique equilibrium outcome; agents have enough free attention to correct for double-counting. We characterize ϵ^* using a transportation problem with a natural economic interpretation: each agent has a supply of discretionary attention and an absorption capacity for incoming attention, and ϵ^* is the point at which the most constrained subset's supply exhausts the capacity of its collective neighborhood. For undirected networks, ϵ^* equals the largest feasible minimum attention level ($1/d_{\max}$), so the first-best is always achievable whenever the constraint is feasible. For directed networks, ϵ^* can be strictly smaller: asymmetries in the flow of attention create bottlenecks that prevent uniform influence even when the constraint is feasible.

When $\epsilon > \epsilon^*$, the minimum attention constraints bind and the equilibrium

influence is no longer uniform. We derive the shadow price on each forced edge and show it is proportional to the gap between the forced neighbor's score and the scores of voluntary neighbors, multiplied by the influence of the listening agent. The welfare cost of forced attention on an edge thus depends on two factors: how informationally redundant the forced neighbor is relative to the rest of the neighborhood, and how much influence the listening agent has, which determines how much her reallocation would move the consensus. We also derive shadow prices on missing edges, which tell a network designer the marginal value of allowing attention on a new link. The score vector serves as connective tissue between an individual agent's best response and the welfare properties of the full equilibrium.

Related Literature

Our paper contributes to a research program studying the interaction between network structure and information aggregation under DeGroot updating. Golub and Jackson (2010) established the foundational result that with exogenous weights, the DeGroot process achieves asymptotic wisdom of crowds (consensus converges to the truth as the network grows) if and only if the most influential agent's influence vanishes in the limit. Their analysis takes the weight matrix as given. Acemoglu, Bimpikis and Ozdaglar (2014) endogenize the network itself: agents form costly communication links before exchanging information, and asymptotic learning depends on the resulting network's diameter. Their model endogenizes link formation but not the weights on existing links. Como, Fagnani and Proskurnikov (2025) endogenize one dimension of the weight matrix: each agent chooses her self-confidence (weight on her self-loop), taking off-diagonal relative weights as fixed. They show that this limited strategic choice is sufficient for agents to achieve the Pareto optimum. Our paper endogenizes the entire weight vector: each agent chooses how to distribute attention across all of her neighbors and herself. The key novel element is the minimum attention constraint $\epsilon > 0$. In practice, agents have substantial but not complete control over how they weigh their social connections (they cannot fully tune out a coworker or a family member). The minimum attention constraint parameterizes this partial agency, from full control at $\epsilon = 0$ to increasingly restricted reallocation as ϵ grows. This allows us to trace how welfare deteriorates as agents lose control over their attention, rather than treating relative weights as exogenously fixed as in Como, Fagnani and Proskurnikov (2025).

Application

The motivating application is social media. Online platforms create directed networks (users follow accounts but are not necessarily followed back) and algorithmic curation amplifies already-popular voices, effectively forcing attention toward high-score nodes. Our results predict that such environments are where

the welfare costs of forced attention are largest: directed structure pushes ϵ^* below $1/d_{\max}$, and algorithmic amplification pushes the effective ϵ above ϵ^*. The model provides a language for distinguishing the welfare cost of specific forced connections from the cost of missing ones, a distinction that enriches the standard recommendation to "diversify your feed".

Roadmap

Section I introduces the model. Section II derives the best-response characterization and the score vector. Section III analyzes equilibrium: first without forced attention (Section III.A), then with the minimum attention constraint (Section III.B), including the dual problem and shadow prices. Proofs not in the main text appear in the Mathematical Appendix.

I. Model

We consider a finite directed weighted network of n agents. The relationships between agents are captured by the $n \times n$ adjacency matrix D, where $D_{ij} = 1$ if i listens to j and 0 otherwise. Let $E = \{(i,j) : D_{ij} = 1\}$ be the set of edges. Let $N(i) = \{j : (i,j) \in E\}$ be the "neighborhood" of i, the set of agents who i is connected to. Agent i may divide her attention heterogeneously among $N(i)$. Let M denote the weighted adjacency matrix where $0 \leq M_{ij} \leq 1$ denotes the amount of attention i gives to j. In particular $(i,j) \notin E$ implies $M_{ij} = 0$, but it may be that $M_{ij} = 0$ while $(i,j) \in E$. In other words, i may pay 0 attention to someone she is nonetheless connected to. Finally, we require that each row of M sums to 1, so that agents must allocate exactly 1 unit of weight among their neighbors.

In each period $t \in \{0, 1, 2, \dots\}$, agents update their beliefs by taking a weighted average of their neighbors' beliefs. The vector of agents' beliefs at time t is denoted $b^{(t)}$, and the updating rule is

$$b^{(t)} = Mb^{(t-1)}, \quad b^{(t)} = M^t b^{(0)}, \tag{1}$$

where $b^{(0)}$ is the vector of initial beliefs. We assume that M is fixed across periods, so the updating rule does not depend on the period. At time $t = 0$, agents receive their initial belief $b_i^{(0)}$, where $b_i^{(0)} = \mu + \eta_i$ for some true state of nature μ. The η_i's are drawn iid from a normal distribution with mean 0 and variance σ^2. Such a model of belief updating is referred to as the "DeGroot" model.

We assume throughout that D is strongly connected and that every agent has a self-loop: $(i,i) \in E$ for all i. Strong connectivity ensures that information can flow between any pair of agents. Self-loops are a natural assumption in our setting: agents can always hear themselves. Together with strong connectivity, they ensure aperiodicity whenever all edges carry positive weight (as in the $\epsilon > 0$ game of Section III.B).

For any stochastic matrix M, define

$$M^\infty = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} M^t.$$

This limit always exists. When M is aperiodic (as guaranteed in the $\epsilon > 0$ game), it coincides with the pointwise limit $\lim_{t\to\infty} M^t$. Agent i's long-run belief is

$$\bar{b}_i = (e_i^\top M^\infty) b^{(0)},$$

a weighted average of the initial beliefs with weights given by the i-th row of M^∞. When M is irreducible, every row of M^∞ equals the stationary distribution $\pi^\top$, the unique left eigenvector satisfying $\pi^\top M = \pi^\top$ and $\mathbf{1}^\top \pi = 1$, and all agents share the same long-run belief $\pi^\top b^{(0)}$. When M is reducible (which can arise even on a strongly connected D if agents zero out edges), agents in different recurrent classes reach different long-run beliefs.

In the standard DeGroot model, agents place equal weight on each neighbor, so

$$M_{ij} = \frac{D_{ij}}{|N(i)|}. \tag{2}$$

Instead, we study the setting where agents may freely distribute weight among their neighbors.

The Game

In our setting, agent i has control over her row of M. She must choose a vector $M_{i*} \geq 0$ where $\sum_j M_{ij} = 1$ such that $M_{ij} = 0$ if $(i,j) \notin E$. Agent i aims to minimize the mean squared error of her long-run belief:

$$\mathbb{E}[(\bar{b}_i - \mu)^2] = \mathrm{Var}(\bar{b}_i) = \sigma^2 \|e_i^\top M^\infty\|^2,$$

since $\bar{b}_i$ is an unbiased estimator of μ (it is a convex combination of the $b_j^{(0)}$, each with mean μ). Her payoff is $u_i(M) = -\sigma^2 \|e_i^\top M^\infty\|^2$. When M is irreducible, $e_i^\top M^\infty = \pi^\top$ for all i, so agents share a common payoff $u_i = -\sigma^2 \|\pi\|^2$ and agent i's problem reduces to

$$\min_{M_{i*}} \|\pi(M)\|_2^2 \quad \text{s.t.} \quad \sum_j M_{ij} = 1,\ M_{ij} = 0\ \forall (i,j) \notin E. \tag{3}$$

II. Best Response

In this section, we analyze a single individual's best response to a fixed network *when the current weights define an irreducible matrix*, so that the graph remains

strongly connected. Irreducibility unlocks a suite of tools from the theory of ergodic Markov chains. Furthermore, we will show in Section III that in the unconstrained game, efficient equilibria coincide with irreducible equilibria, so our tools will be directly applicable. Let M_{-i} denote the matrix M with row i set to $\mathbf{0}^\top$. For the remainder of this section, we take M_{-i} as fixed and consider agent i's best response to M_{-i}. Furthermore, let $F(w) = \|\pi(w)\|_2^2$ be the objective.

We work with feasible first-order changes of the row-stochastic matrix M, with a fixed starting row i. For a fixed row i, a valid perturbation has the form $dM = e_i\alpha^\top$ with $\alpha \in \mathcal{A} :- \{\alpha : \mathbf{1}^\top\alpha = 0, supp(\alpha) \subseteq N(i)\}$ where *supp* denotes the support of a vector, the set of its non-zero entries. In other words, α preserves stochasticity and may only affect edges to nodes in i's neighborhood.

Differentiating the stationary equation yields

$$d\pi^\top M + \pi^\top dM = d\pi^\top \implies \pi^\top dM = d\pi^\top(I - M). \tag{4}$$

Let $Z = (I - M + \mathbf{1}\pi^\top)^{-1}$ be the fundamental matrix of M Schweitzer (1968). Z satisfies the identity $(I - M)Z = I - \mathbf{1}\pi^\top$. Combining this identity with (4) yields

$$\pi^\top dMZ = d\pi^\top(I - \mathbf{1}\pi^\top). \tag{5}$$

Finally, since we require $\pi^\top\mathbf{1} = 1$, it follows that $d(\pi^\top\mathbf{1}) = 0$ and so

$$d\pi^\top = \pi^\top dMZ \implies d\pi = \pi_i Z^\top\alpha \tag{6}$$

using $dM = e_i\alpha^\top$. The same may be derived from (17) in Schweitzer (1968).

With $d\pi$ in hand, we may now differentiate the objective $\|\pi\|^2 = \pi^\top\pi$. Define $s = Z\pi$. Then

$$(d\pi)^\top\pi + \pi^\top d\pi = 2\pi^\top d\pi = 2\pi_i\pi^\top Z^\top\alpha = 2\pi_i s^\top\alpha. \tag{7}$$

The directional derivative of F along any feasible perturbation α is $2\pi_i s^\top\alpha$. Since the constraint $\mathbf{1}^\top\alpha = 0$ defines the feasible subspace, the projected gradient onto $\mathcal{A}$ is obtained by subtracting the component along $\mathbf{1}$, yielding:

$$\nabla_i F = \begin{cases} 2\pi_i(s_j - \overline{s}_{N(i)}), & j \in N(i) \\ 0, & j \notin N(i), \end{cases} \tag{8}$$

where $\overline{s}_{N(i)} = \frac{1}{|N(i)|}\sum_{k\in N(i)} s_k$ is the mean score across i's neighborhood. The direction of steepest descent is therefore

$$-\nabla_i F = -2\pi_i(s|_{N(i)} - \overline{s}_{N(i)}\mathbf{1}), \tag{9}$$

where $s|_{N(i)}$ denotes the restriction of s to the coordinates in $N(i)$.

The best infinitesimal move is to take weight *from* neighbors with $s_j > \overline{s}$ and

give it *to* those with $s_j < \overline{s}$, proportional to how far on either side of $\overline{s}$ each neighbor is. A key observation is that the direction of steepest-descent is *independent* of the best-responding agent; that is, $s = Z\pi$ may be thought of as a list of scores determining the marginal benefit of giving weight to any individual agent, and each agent sees the same list of scores. An agent must then project those scores onto their neighborhood to determine the direction of movement (by comparing to the neighborhood average); while the exact directions will therefore naturally differ between agents, the *ranking* of agents is fixed. If $k, l \in N(i)$ and $k, l \in N(j)$, then i and j will always agree on which of k or l gets more weight.

A. Graph-Theoretic Interpretation of the Score Vector

The score vector $s = Z\pi$ admits a graph-theoretic interpretation that clarifies how an agent's score is determined by her position in the network structure. Using the relationship between the fundamental matrix and mean first passage times (Appendix A.A1), we can write

$$s_k = c - \sum_j \pi_j^2 m_{kj}, \tag{10}$$

where m_{kj} is the mean first passage time from k to j in the random walk defined by M, and c is a graph-dependent constant. Intuitively, the score of agent k decreases with her expected distance from high-influence nodes. Agents who are close (in a random-walk sense) to the most influential parts of the network have high scores; agents who are far from them have low scores. Since agent i shifts weight toward low-score neighbors, she is directing attention toward those neighbors who are farthest from the parts of the network that already dominate the consensus.

The score takes a particularly clean form in the standard DeGroot setting: agents in an undirected graph place uniform weight on each of their neighbors, yielding $M_{ij} = \frac{D_{ij}}{|N(i)|}$. Such a setting may be exactly modeled by the natural random walk on an undirected graph, which unlocks a suite of useful analytical tools (in particular because M is a reversible Markov chain). In Appendix A.A2, we show that s_i is a decreasing affine function of $Cov_\pi(\pi_J, R_{iJ})$:

$$s_i = a - b \cdot Cov_\pi(\pi_J, R_{iJ}), \quad b > 0, \tag{11}$$

where a and b are graph-dependent constants and R_{iJ} is the effective resistance from i to a randomly sampled node J (drawn with probability according to π). Effective resistance measures path redundancy between two nodes: it is small when many short paths connect them and large when connections are long and sparse. An agent with low effective resistance to the influential core hears the same opinions repeatedly through many channels. This is exactly the double-counting problem. Her covariance is negative (high influence correlates with low

resistance), giving her a high score: neighbors will shift weight away from her. Conversely, an agent with high resistance to the core receives information through few, independent paths. Her covariance is positive, giving her a low score: neighbors will shift weight toward her, precisely because her perspective is not already echoing through the network. Once agents begin reweighting, reversibility of M is generally lost, but the qualitative insight persists: the score vector identifies which agents' information is a redundant echo of popular opinions and which is genuinely novel.

B. Conditions for Best Response

We have described the gradient of $F(w) = \|\pi(w)\|_2^2$ with respect to agent i's outgoing edges, and in particular the global "scores" of each player that determine the degree to which her neighbors wish to re-weight her in the direction of steepest descent. We will now characterize the set of weights agent i places on her neighbors in a best response, fixing the weights in the rest of the graph.

The Karush-Kuhn-Tucker (KKT) conditions provide first-order necessary conditions for the optimality of i's weight vector w. Our constraints are

$$\sum_j w_j = 1, \quad w_j \geq 0, \quad \forall j. \tag{12}$$

Form the Lagrangian function:

$$\mathcal{L}(w, \eta, \rho) = F(w) + \eta(\sum_j w_j - 1) - \sum_{j \in N(i)} \rho_j w_j \tag{13}$$

for KKT multipliers η and ρ_j. The KKT conditions are as follows. **Stationarity**:

$$\nabla \mathcal{L}(w, \eta, \rho) = 0 \iff \frac{\partial \mathcal{L}}{\partial w_j} = 2\pi_i s_j + \eta - \rho_j = 0 \quad \forall j. \tag{14}$$

Primal Feasibility:

$$\sum_j w_j = 1, \quad w_j \geq 0 \ \forall j. \tag{15}$$

Dual Feasibility:

$$\rho_j \geq 0 \quad \forall j. \tag{16}$$

Complementary Slackness:

$$\rho_j w_j = 0 \quad \forall j. \tag{17}$$

From the KKT conditions, we derive the following rule for when a set of weights

is optimal. Consider any minimizer w^* of $F(w)$. For agent $j \in N(i)$, if $w_j^* > 0$ then j is given positive weight and $\rho_j = 0$. Therefore the stationarity condition becomes

$$2\pi_i s_j + \eta = 0 \implies s_j = -\frac{\eta}{2\pi_i}. \tag{18}$$

Define $\tau_i = -\frac{\eta}{2\pi_i}$ and observe that τ_i does not depend on j. It follows s_j is the same for any neighbor j for which $w_j^* > 0$, or equivalently that each positively-weighted neighbor has the same "score" τ_i in any optimal vector of weights.

Now consider j for which $w_j^* = 0$. Then $\rho_j = 2\pi_i s_j + \eta \geq 0$, so $s_j - \tau_i \geq 0$. It follows that any neighbor j who receives zero weight will have a score s_j, which is at least the common score τ_i shared by positively-weighted neighbors.

We summarize these first-order conditions below:

Proposition 1. *Consider agent i. For any KKT point w^* of $F(w)$ (with M irreducible), we have that*

$$\exists \tau_i \text{ s.t. } \begin{cases} w_j^* > 0 \implies s_j = \tau_i, \\ w_j^* = 0 \implies s_j \geq \tau_i, \end{cases} \qquad \sum_{j \in N(i)} w_j^* = 1, \quad w_j^* \geq 0 \ \forall j \in N(i). \tag{19}$$

In other words, in any best response, agent i allocates weight such that positively-weighted agents have the minimal score, and zero-weighted agents have at least the minimal score. Proposition 1 gives us a clean method to verify whether a given candidate vector w^* satisfies the first-order conditions: compute $s(w^*)$ and check whether the aforementioned statement holds.

While the conditions in Proposition 1 characterize the KKT points of $F(w)$, the following proposition states that any such KKT point is also a local minimum.

Proposition 2. *Let w^* be a KKT point of $\min F(w)$ subject to $\sum_j w_j = 1$ and $w_j \geq \epsilon$ for all $j \in N(i)$, for any $\epsilon \geq 0$. For any nonzero feasible perturbation α with $supp(\alpha) \subseteq N(i)$ and $\mathbf{1}^\top \alpha = 0$, either the α-directional first derivative is strictly positive, or it vanishes and the second derivative is strictly positive. In particular, w^* is a strict local minimum of $F(w)$.*

The proof appears in Appendix A.A3. Since F is not convex in w (Appendix A.A4), KKT points could in principle be saddle points. Proposition 2 rules this out: every first-order stationary point of agent i's problem is a strict local minimum. This ensures that best responses are locally well-determined and will be used in Section III.B.

III. Equilibrium

A. *Equilibrium Without Forced Attention*

We begin by characterizing equilibrium in the unconstrained game, where agents may place zero weight on any neighbor ($\epsilon = 0$). Since agents may willingly sever links, the resulting matrix M need not be irreducible, and we must consider deviations beyond the irreducible region studied in Section II.

Recall from Section I that agent i's long-run belief is $(e_i^\top M^\infty)b^{(0)}$. Her payoff is

$$u_i(M) = -\sigma^2 \|e_i^\top M^\infty\|^2,$$

the negative variance of her long-run belief. When M is irreducible, every row of M^∞ equals $\pi^\top$, so $u_i = -\sigma^2\|\pi\|^2$ for all i and the analysis of Section II applies directly. When M is reducible, agents in different recurrent classes reach different long-run beliefs and may have different payoffs.

We will show that the set of irreducible, first-best efficient, and strict Nash equilibria coincide. The key tool is an identity linking each agent's score to the distance of her influence from the norm of the influence vector.

Lemma 1 (Score-Influence Identity)**.** *For any irreducible M with stationary distribution π, fundamental matrix Z, and score vector $s = Z\pi$:*

$$s_i - (Ms)_i = \pi_i - \|\pi\|^2 \quad \forall i. \tag{20}$$

PROOF:

$$(I - M)s = (I - M)Z\pi = (I - \mathbf{1}\pi^\top)\pi = \pi - \|\pi\|^2 \cdot \mathbf{1},$$

where we used the defining property $(I - M)Z = I - \mathbf{1}\pi^\top$. The left side is $s - Ms$. □

Theorem 1 (Characterization of Efficient Equilibria)**.** *For a Nash equilibrium M of the unconstrained game on a strongly connected graph D with self-loops, the following are equivalent:*

(i) M is irreducible.

(ii) M achieves the first-best: $u_i = -\sigma^2/n$ for all i.

(iii) M is a strict NE.

Moreover, such equilibria exist: any irreducible doubly stochastic M supported on D is a strict NE.

PROOF:

Since σ^2 is a positive constant, we work with the variance $\|\pi\|^2$ throughout the proof; minimizing variance is equivalent to maximizing payoff.

(i) $\Longrightarrow$ (ii). Since any Nash equilibrium is in particular a local optimum over perturbations that preserve irreducibility, the first-order conditions of Proposition 1 hold at any irreducible NE.

At $\epsilon = 0$, Proposition 1 partitions each agent's neighbors into chosen ($M_{ij} > 0$, score $s_j = \tau_i$) and unchosen ($M_{ij} = 0$, score $s_j \geq \tau_i$). Since every positively-weighted neighbor has score exactly τ_i:

$$(Ms)_i = \sum_j M_{ij} s_j = \sum_{j: M_{ij} > 0} M_{ij} \tau_i = \tau_i.$$

This uses the fact that at $\epsilon = 0$, there are no forced neighbors. Every positive weight is chosen voluntarily.

This equality $\tau_i = (Ms)_i$ is specific to the unconstrained game. At $\epsilon > 0$, forced neighbors contribute score above τ_i, making $(Ms)_i > \tau_i$. This is precisely the mechanism by which forced attention creates welfare costs, which we will revisit in Section III.B.

Self-loops guarantee $i \in N(i)$, so Proposition 1 gives $s_i \geq \tau_i$. By Lemma 1:

$$\pi_i - \|\pi\|^2 = s_i - (Ms)_i = s_i - \tau_i \geq 0.$$

Therefore $\pi_i \geq \|\pi\|^2$ for all i.

Summing over all agents: $1 = \sum_i \pi_i \geq n\|\pi\|^2$, giving $\|\pi\|^2 \leq 1/n$. The reverse inequality $\|\pi\|^2 \geq 1/n$ holds for any probability vector by Cauchy-Schwarz, with equality if and only if $\pi = \frac{1}{n}\mathbf{1}$. Therefore $\pi = \frac{1}{n}\mathbf{1}$.

(ii) $\Longrightarrow$ (iii). If every agent's variance is $1/n$, then $\pi = \frac{1}{n}\mathbf{1}$ and M is doubly stochastic. Any non-trivial unilateral deviation by agent i produces M' that is not doubly stochastic: column j of M' sums to $1 - M_{ij} + M'_{ij} \neq 1$ for at least one j. Agent i's long-run belief vector $e_i^\top (M')^\infty$ is a stationary distribution of M'. If it equaled $\frac{1}{n}\mathbf{1}$, then M' would be doubly stochastic, a contradiction. Therefore $e_i^\top (M')^\infty \neq \frac{1}{n}\mathbf{1}$, and Cauchy-Schwarz gives variance $> 1/n$. All deviations are strictly worse.

(iii) $\Longrightarrow$ (i). Suppose M is a reducible NE with recurrent classes $C_1, \ldots, C_k$ ($k \geq 2$). Each C_a defines an irreducible sub-chain on a strongly connected subgraph of D with self-loops. If any agent in C_a had a profitable deviation supported on $N(i) \cap C_a$, it would be profitable in the full game (it keeps weight within C_a, leaving other classes unaffected). So agents in C_a are at a NE of the sub-game, and the (i) $\Longrightarrow$ (ii) argument applied within C_a gives $\pi_{C_a} = \frac{1}{|C_a|}\mathbf{1}$ with variance $1/|C_a|$.

Let C_a be a smallest class. By strong connectivity of D, some $i \in C_a$ has a D-edge to $j \notin C_a$. When i deviates to place positive weight on j, the class C_a becomes transient: since C_a was communicating, every node in C_a can reach i

through unchanged internal edges, and i now has an edge out. No subset of C_a remains closed. The surviving recurrent classes C_b ($b \neq a$) are unaffected. Agent i is absorbed into surviving classes with probabilities $p_b \geq 0$, $\sum_{b\neq a} p_b = 1$. Since all surviving classes have size $\geq |C_a|$:

$$\sum_{b\neq a} \frac{p_b^2}{|C_b|} \leq \frac{\sum_{b\neq a} p_b^2}{|C_a|} \leq \frac{1}{|C_a|}$$

The deviation variance is at most $1/|C_a|$. Agent i has a deviation that is not strictly worse, so M is not strict.

Existence. At doubly stochastic M: $\pi = \frac{1}{n}\mathbf{1}$, so $u_i = -\sigma^2/n$ for all i. For any deviation M'_{i*} by agent i (producing M', irreducible or reducible), agent i's payoff is $u_i(M') = -\sigma^2\|e_i^\top (M')^\infty\|^2$. Since $e_i^\top (M')^\infty$ is a probability vector of length n, Cauchy-Schwarz gives $\|e_i^\top (M')^\infty\|^2 \geq 1/n$. Therefore $u_i(M') \leq -\sigma^2/n = u_i(M)$, and no deviation is profitable.

To see that irreducible doubly stochastic matrices supported on D exist: for each off-diagonal edge $(i,j) \in E$ with $i \neq j$, the strong connectivity of D provides a directed path from j to i, which together with (i,j) forms a directed cycle. The corresponding cyclic permutation matrix is doubly stochastic and supported on D. Including the identity matrix (which covers all self-loops) and taking a convex combination with strictly positive coefficients yields an irreducible doubly stochastic matrix with $M_{ij} > 0$ for all $(i,j) \in E$. □

The theorem provides a complete characterization of efficient equilibria: they exist, they are the only strict equilibria, and they are precisely the irreducible doubly stochastic matrices on D.

The proof of (i) $\Longrightarrow$ (ii) clarifies the important role of self-loops in efficiency. At an irreducible NE, Proposition 1 gives $s_j \geq \tau_i$ for all $j \in N(i)$. Self-loops ensure $i \in N(i)$, so in particular $s_i \geq \tau_i$. Without self-loops, Proposition 1 constrains only the score of i's neighbors, not i's own score, so the identity cannot be used to bound π_i. With them, the bound $s_i \geq \tau_i$ translates to $\pi_i \geq \|\pi\|^2$, which when imposed simultaneously across all agents forces uniform influence. That self-weight is the binding force echoes Como, Fagnani and Proskurnikov (2025), where control over self-confidence alone is sufficient for efficiency.

The set of efficient equilibria is typically large: the irreducible doubly stochastic matrices on D form a convex set. Imposing minimum attention $\epsilon > 0$ in Section III.B perturbs the equilibrium set and pins down a unique equilibrium influence vector.

Theorem 1 leaves open whether reducible Nash equilibria exist. We now show that reducible Nash equilibria can arise, and that they have a sharp economic interpretation as echo chambers sustained by coordination failure.

Example. Consider the network of four agents in Fig. 2.

The setting where A and B listen equally to each other (and similarly C and D), ignoring the cross-group edges, is a reducible Nash equilibrium with recurrent

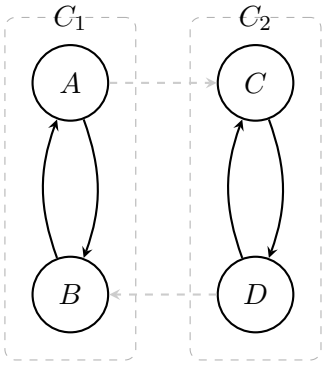


FIGURE 2. A REDUCIBLE NASH EQUILIBRIUM ON FOUR AGENTS. SOLID ARROWS INDICATE EDGES RECEIVING POSITIVE WEIGHT; DASHED GRAY ARROWS INDICATE EDGES PRESENT IN D BUT RECEIVING ZERO WEIGHT.

classes $C_1 = \{A, B\}$ and $C_2 = \{C, D\}$. Since agents in each class interact only with each other, their strategies must form a Nash equilibrium of the two-player sub-game on their group. Theorem 1 applied to this sub-game implies that the within-class stationary distribution is uniform, giving every agent a payoff of $-\sigma^2/2$. To verify this is a NE: agents B and C have no edges outside their class, so any deviation by B and C will break within-class double stochasticity and strictly worsen their payoff. It remains to check the bridge agents A and D.

If A shifts any positive weight to C, the class $\{A, B\}$ becomes transient. Both A and B are absorbed into $\{C, D\}$, the sole surviving recurrent class. Because agent A traded the two signals $\{A, B\}$ for two different signals $\{C, D\}$, her post-deviation payoff is still $-\sigma^2/2$. She is exactly indifferent. The same logic applies to D: shifting weight to B dissolves $\{C, D\}$ into $\{A, B\}$ with the same payoff for all agents. Neither bridge agent has a strict incentive to deviate.

But if both A and D shift weight $t > 0$ to their respective cross-group neighbors, the resulting matrix is doubly stochastic and irreducible, giving every agent a payoff of $-\sigma^2/4$. The welfare gain (access to four signals instead of two) is not attainable by individual deviations: each agent's individual deviation completely washes out her group's initial information. Only with the symmetric joint deviation can the society achieve first-best consensus.

The example illustrates three structural features: two groups of equal size, each internally well-connected, with cross-group edges that route each bridge agent to exactly one other group. The following proposition shows that these three conditions on the graph structure are sufficient to sustain reducible Nash equilibria.

Proposition 3 (Echo Chamber Characterization)**.** *Let D be a strongly connected graph with self-loops. A reducible NE with $k \geq 2$ recurrent classes exists if D admits a partition $V = C_1 \cup \cdots \cup C_k$ satisfying:*

(a) $|C_1| = \cdots = |C_k|$,

(b) each C_a's internal edges in D form a strongly connected subgraph, and

(c) no agent has edges in D to members of two or more classes distinct from her own

PROOF:

Given a partition satisfying (a)-(c), construct within-class doubly stochastic blocks on each C_a (these exist by the same cycle-permutation argument used in Theorem 1). Set all cross-class edge weights to zero. Within-class deviations are suboptimal by Theorem 1 applied to the sub-game. Agents with no cross-class edges in D have no other deviations available. It remains to check bridge agents, those with an edge in D leaving their class.

By condition (c), any bridge agent $i \in C_a$ has cross-class edges leading to members of exactly one other class, say C_b. If i places positive weight on any such edge, C_a becomes transient (since C_a was communicating internally, every member of C_a can reach i, who now has an edge out). Agent i is absorbed into C_b, and her post-deviation payoff becomes $-\sigma^2/|C_b| = -\sigma^2/|C_a|$ by condition (a). She is exactly indifferent, so no deviation is strictly profitable. □

The proposition identifies three sufficient conditions under which a network can sustain information fragmentation. Each has a distinct economic interpretation. Condition (a) requires that the groups be of equal size. A smaller group's bridge agent would strictly prefer to join a larger group, which offers a more diversified consensus; equal-size is the knife-edge condition at which each bridge agent trades her group's signals for the same number of different signals with zero net gain. Condition (b) requires that each community is internally cohesive enough to sustain independent consensus. Condition (c) requires that bridge agents face limited outside options. An agent who can access two distinct communities other than her own benefits from splitting attention across them: her payoff $-\sigma^2 \sum p_b^2/|C_b|$ is strictly convex in the absorption probabilities, so concentrating attention in one class is strictly worse than diversifying across two when classes are the same size.

Furthermore, a single graph can support multiple fragmentation patterns. If D admits several distinct partitions satisfying (a)-(c), each supports a different reducible NE, all coexisting alongside the efficient equilibria from Theorem 1. The coordination failure at each reducible NE has the same structure: bridge agents are indifferent between sustaining and dissolving their echo chamber, and the welfare improvement from connecting the communities requires coordinated deviation.

In the unconstrained game, we showed that the first-best consensus is achievable on any strongly connected directed graph with self-loops. The welfare cost of fragmentation into echo chambers is real, but this is purely a coordination failure. The network structure itself does not force poor consensus in equilibrium.

The echo chambers of Proposition 3 are sustained by a particular mechanism: bridge agents zero out their cross-group edges, causing fragmentation. In practice, social connections rarely afford this degree of control. In most settings, the graph D represents communication channels that cannot be completely severed: in-person coworkers are unable to avoid each other completely, students in the same dorm will run into each other, and family members get together regularly for the holidays. In the following section, we formalize this minimum attention constraint, investigating equilibria efficiency in the setting where each edge must carry some positive weight.

B. Equilibrium With Minimum Attention Constraints

The previous section showed that without any constraints on attention, strategic agents in a network achieve the first-best at any irreducible equilibrium, but they may also fragment into smaller echo chambers at a cost to consensus accuracy. We now impose a minimum attention constraint, reflecting the difficulty of completely ignoring a social connection: each agent must place at least ϵ weight on every neighbor in D. With the constraint, every edge in D carries positive weight, and every Nash equilibrium is irreducible.

Irreducibility simplifies the payoff structure considerably. When M is irreducible, every row of M^∞ equals $\pi^\top$, so agents share the same payoff $u_i = -\sigma^2\|\pi\|^2$. Finding Nash equilibria therefore reduces to a single optimization problem: minimize $\|\pi(M)\|^2$ over the set of feasible weight matrices.

We now formally describe this optimization problem. Given D, the adjacency matrix of a strongly connected directed graph, let S_D denote the set of stochastic matrices whose underlying directed graphs are exactly D, and whose positive entries are all at least $\epsilon > 0$. That is,

(21)
$$S_D = \{M \in \mathbb{R}^{n\times n} \mid M \geq 0, M\mathbf{1} = \mathbf{1}, (i,j) \in E \iff M_{ij} \geq \epsilon,\ (i,j) \notin E \iff M_{ij} = 0\}.$$

S_D is equivalent to the set of strategy profiles of the game—in any particular $M \in S_D$, each agent has chosen some way to weigh her neighbors' opinions such that they sum to 1 and no individual neighbor receives less than ϵ weight. With $\epsilon > 0$, every $M \in S_D$ is irreducible, so all agents share the payoff $u_i = -\sigma^2\|\pi(M)\|^2$. Since each agent controls only her own row of M and the constraint set is row-separable, M is a Nash equilibrium if and only if no unilateral change to a single row decreases $\|\pi(M)\|^2$, that is, if and only if each row globally minimizes $\|\pi(M)\|^2$ given the others.

The objective $\|\pi(M)\|^2$ is not convex in the weight matrix M, even when restricted to a single agent's row (see Appendix $A.A4$ for a counterexample). We therefore cannot guarantee that the local minima characterized in Section II are global. In this section, we provide a strictly convex formulation of the problem, use it to demonstrate that there is a unique equilibrium consensus, then characterize the threshold of minimum attention above which the first-best payoff is not

achievable in this unique consensus.

Let $M \in S_D$ and suppose $\pi^\top M = \pi^\top$ for some $\pi \in \mathbb{R}^n_+$ such that $\pi^\top \mathbf{1} = 1$. We make use of the following observation from Breen and Kirkland (2015): Let $X = \text{diag}(\pi)$ and consider the matrix $B = XM$. Then $B\mathbf{1} = \pi$ and $\mathbf{1}^\top B = \pi^\top$. It follows that a positive probability vector π is a stationary vector of some $M \in S_D$ if and only if there exists a nonnegative B such that these constraints hold. As noted in Breen and Kirkland (2015), the matrix B is a member of the symmetric transportation polytope $\mathcal{T}(\pi)$, the set of nonnegative matrices whose rows and columns both sum to π.

Note that $B_{ij} = \pi_i M_{ij}$. Treating π and B as our decision variables, the constraints become

$$B\mathbf{1} = \pi, \quad \mathbf{1}^\top B = \pi^\top, \quad B_{ij} \geq \epsilon \pi_i \; \forall (i,j) \in E, \quad B_{ij} = 0 \; \forall (i,j) \notin E. \tag{22}$$

The objective $\|\pi\|^2$ is strictly convex in π, though it does not depend on B. Since the constraints (22) are linear in (π, B), minimizing a strictly convex function of π over a convex feasible set yields a unique minimizer π^*. The matrix B^* need not be unique: any B consistent with π^* and the constraints achieves the same value. We can construct an $M \in S_D$ of which the solution π is a stationary vector by taking the resulting B and setting $M_{ij} = B_{ij}/\pi_i$. Solving the convex problem will give us the unique set of optimal social influences π^* and some set of weights that admits π^*.

Before connecting the convex program to the game's equilibrium conditions, we confirm that equilibria exist.

Proposition 4 (Existence of Nash Equilibrium)**.** *For any ϵ such that S_D is nonempty (equivalently, $\epsilon \leq 1/d_{\max}$), the game has a Nash equilibrium. In particular, the matrix $\hat{M} = \text{diag}(\pi^*)^{-1} B^*$ constructed from any optimal B^* of the convex program* (22) *is a Nash equilibrium.*

PROOF:

With $\epsilon > 0$, every $M \in S_D$ is irreducible, so all agents share the payoff $-\sigma^2 \|\pi(M)\|^2$; the game has identical interests and is in particular an exact potential game with potential $-\|\pi(M)\|^2$. The Breen-Kirkland equivalence already established above implies that the feasible influence vectors of the convex program are exactly $\{\pi(M) : M \in S_D\}$, so the convex minimizer π^* achieves the global minimum of $\|\pi(M)\|^2$ over S_D. The matrix $\hat{M}$ lies in S_D: its entries satisfy $\hat{M}_{ij} = B^*_{ij}/\pi^*_i \geq \epsilon$ on edges (since $B^*_{ij} \geq \epsilon \pi^*_i$), equal zero on non-edges, and its rows sum to one (since $B^*\mathbf{1} = \pi^*$ and $\pi^* > 0$). Any unilateral deviation replaces one row and keeps the matrix in S_D, so it cannot reduce the common objective below its global minimum. Therefore each agent's row is a best response and $\hat{M}$ is a Nash equilibrium. □

To connect the equilibria of the game to this convex program, recall that each agent best-responds, and characterize that best response by its first-order conditions, the ϵ-analogue of Proposition 1. Fixing M^*_{-i}, agent i minimizes

$F(w_i) = \|\pi(w_i, M^*_{-i})\|_2^2$ subject to $w_{ij} \geq \epsilon$ for all $j \in N(i)$ and $\sum_j w_{ij} = 1$. The Lagrangian takes the same form as in Proposition 1 with ϵ replacing 0 as the lower bound:

$$L(w, \eta, \rho) = F(w) + \eta\Big(\sum_j w_{ij} - 1\Big) - \sum_{j \in N(i)} \rho_j(w_{ij} - \epsilon). \tag{23}$$

Since $\partial F/\partial w_{ij} = 2\pi_i^* s_j^*$, stationarity gives $2\pi_i^* s_j^* + \eta - \rho_j = 0$ for all $j \in N(i)$, with $\rho_j \geq 0$ the multiplier on $w_{ij} \geq \epsilon$. Setting $\tau_i = -\eta/(2\pi_i^*)$ gives $\rho_j = 2\pi_i^*(s_j^* - \tau_i)$, and dual feasibility yields $s_j^* \geq \tau_i$ for all $j \in N(i)$. Complementary slackness $\rho_j(w_{ij}^* - \epsilon) = 0$ separates voluntary neighbors ($w_{ij}^* > \epsilon$, hence $\rho_j = 0$ and $s_j^* = \tau_i$) from floored neighbors ($w_{ij}^* = \epsilon$, hence $s_j^* \geq \tau_i$):

$$M_{ij}^* > \epsilon \implies s_j^* = \tau_i, \qquad M_{ij}^* = \epsilon \implies s_j^* \geq \tau_i. \tag{24}$$

The minimizer π^* above is unique as the optimum of the convex program. The following lemma shows it is also the influence of *every* Nash equilibrium, so the equilibrium influence is unique without qualification and every equilibrium is efficient.

Lemma 2 (Efficiency and uniqueness of constrained equilibria)**.** *Fix $\epsilon > 0$. Every Nash equilibrium M^* on S_D attains the global minimum of $\|\pi\|^2$ over S_D. Consequently the equilibrium influence vector π^* is unique and equals the minimizer of the convex program* (22)*, and every Nash equilibrium achieves this π^*.*

The proof constructs dual variables for the convex program from the equilibrium first-order conditions (24) and verifies that they satisfy KKT; convexity then forces global optimality. Because the same construction also determines the shadow prices on forced edges, we defer the proof to Theorem 2 below.

The unconstrained minimum of $\|\pi\|^2$ over the simplex is $1/n$, achieved at $\pi_n = (1/n)\mathbf{1}$. By Lemma 2 the equilibrium influence equals this unique optimum, so the equilibrium achieves the first-best $\pi^* = \pi_n$ whenever π_n is feasible. Substituting $\pi = (1/n)\mathbf{1}$ into the constraints and writing $M = nB$, feasibility reduces to the existence of a doubly stochastic matrix M supported on D with $M_{ij} \geq \epsilon$ on all edges. Crucially, whether such a matrix exists depends on the interaction between ϵ and the network structure.

Define $\epsilon^*(D)$ as the largest minimum attention level that can sustain uniform influence:

$$\epsilon^*(D) = \max\{\epsilon \geq 0 : \exists \text{ DS } M \text{ with support} \subseteq D \text{ with } M_{ij} \geq \epsilon \ \forall (i,j) \in E\}. \tag{25}$$

The equilibrium achieves the first-best if and only if $\epsilon \leq \epsilon^*$. That $\epsilon^* > 0$ follows from the same cycle-permutation construction in the existence proof of Theorem 1: a strictly positive convex combination of cyclic permutation matrices covering

every edge of D is doubly stochastic with all edge entries bounded away from zero. The threshold ϵ^* thus partitions the parameter space into two regimes. When $\epsilon \leq \epsilon^*$, agents can fully correct for double-counting regardless of the network topology. When $\epsilon > \epsilon^*$, the network structure forces some agents to receive disproportionate influence, and the equilibrium variance $\|\pi^*\|^2$ is strictly larger than the first-best $1/n$.

How does the network structure determine ϵ^*? We characterize the threshold separately for undirected and directed networks, as the two cases differ sharply.

When D is undirected (with self-loops), the threshold takes a simple form.

Proposition 5 (First-Best Threshold for Undirected Graphs)**.** *For any connected undirected graph D with self-loops, $\epsilon^* = 1/d_{\max}$, where $d_{\max} = \max_i |N(i)|$, the maximum feasible ϵ.*

PROOF:

The upper bound is immediate: the node with degree $d_{\max}$ must place at least ϵ on each neighbor, requiring $\epsilon \leq 1/d_{\max}$. For the converse, consider the weight matrix in which each agent assigns weight ϵ to their off-diagonal neighbors and the remainder to themselves. Since D is undirected, every node's in-degree equals its out-degree, so this matrix is symmetric and hence doubly stochastic. All entries are at least ϵ whenever $\epsilon \leq 1/d_{\max}$. □

On undirected networks, the first-best is therefore achievable whenever the minimum attention constraint is feasible at all. Network topology affects convergence speed (see the Remark below) but never equilibrium accuracy.

The undirected proof exploits a symmetry: every node's in-degree equals her out-degree, so assigning uniform weight to each neighbor produces a doubly stochastic matrix automatically. On directed graphs, this symmetry breaks. A node may receive forced attention from many sources while having few outgoing edges, pushing ϵ^* strictly below $1/d_{\max}$.

Example. Consider the following example directed network in Fig. 3.

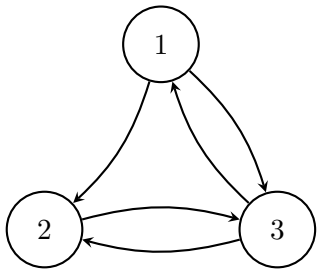


FIGURE 3. A DIRECTED GRAPH WITH $\varepsilon^* = 1/4 < 1/3 = 1/d_{\max}$.

Here, $d_{\max} = 3$ and $1/d_{\max} = 1/3$, but $\epsilon^* = 1/4$. The bottleneck is agent 2. She is popular: agents 1 and 3 each have to send at least ϵ weight to her, capping

her self-weight at $M_{22} \leq 1 - 2\epsilon$ for double stochasticity to hold. The remaining weight (which she can't place on herself) is $M_{23} = 1 - M_{22} \geq 2\epsilon$. This must all go to agent 3, who is equally popular. Agent 3 thus receives at least ϵ from agent 1, at least 2ϵ from agent 2, and at least ϵ from herself, totaling at least 4ϵ. When $\epsilon > 1/4$, column 3 cannot sum to 1; agent 3's incoming attention exceeds her capacity, and no doubly stochastic matrix respecting the constraints exists. The equilibrium influence is uneven: mandatory attention to high-degree agents 2 and 3 inflates their influence beyond $1/n$, and the welfare loss $\|\pi^*\|^2 - 1/n$ becomes strictly positive. If agent 2 had an edge to agent 1, who has in-degree only 2 and thus more room to absorb attention, agent 2's extra attention could be routed there instead. The missing edge creates the bottleneck that forces unequal influence in equilibrium.

More generally, ϵ^* is determined by the most constrained subset of agents in the network. For each agent, formally define each member's discretionary supply $q_i = 1 - \epsilon \cdot d_i^{out}$ (attention remaining after meeting minimum obligations) and absorption capacity $r_i = 1 - \epsilon \cdot d_i^{in}$ (room left in column i after accounting for forced incoming attention). As ϵ increases, both quantities shrink, but at different rates depending on the degree structure. The first-best consensus breaks at the smallest ϵ for which some subset of agents' supplies exceeds the capacity of their collective out-neighborhood.

Proposition 6 (Constrained Subgroups)**.** *Let D be a strongly connected directed graph with self-loops and fix $\epsilon > 0$. The first-best $\pi^* = \pi_n$ is achievable if and only if $\epsilon \leq 1/\bar{d}$ and for all $T \subseteq N$,*

$$\sum_{i \in T} q_i \leq \sum_{j \in N^{out}(T)} r_j, \tag{26}$$

where $\bar{d} = \max_i \max(d_i^{out}, d_i^{in})$ is the largest degree (in- or out-) and $N^{out}(T)$ is the one-step out-neighborhood of T (which contains T itself, since every agent has a self-loop).

The proof, which reduces the problem to a transportation problem on supplies and demands of attention, appears in Appendix *A.A*5.

The threshold ϵ^* is the largest ϵ at which the condition 26 holds for every subset. Equivalently, it is the point at which the most constrained subset's attention supply exactly exhaust the capacity of its collective neighborhood. In the example, the binding subset is $T = \{2\}$: agent 2's supply $q_2 = 1 - 2\epsilon$ equals the combined capacity $2(1 - 3\epsilon)$ of agents 2 and 3 at $\epsilon = 1/4$. On undirected graphs, in-degree equals out-degree at every node, so supply and capacity are balanced across every subset and the condition reduces to $\epsilon \leq 1/d_{\max}$.

Remark. Even within the first-best regime, the worst-case convergence time can be made arbitrarily slow. Consider the maximum self-listening matrix M from the proof of Proposition 5: weight ϵ on each neighbor, the remainder on oneself. This matrix is symmetric and hence reversible. The relaxation time $t_{rel}(M) =$

$1/(1-\lambda_2(M))$, where $\lambda_2(M)$ is the second-largest eigenvalue of M, measures the characteristic timescale of convergence to consensus. Applying the Cheeger inequality for reversible chains (Levin, Peres and Wilmer (2008)) yields the lower bound $t_{rel} \geq 1/(2\epsilon \cdot h(D))$, where $h(D)$ is the Cheeger constant measuring the sparsity of the worst bottleneck in D. The intuition is direct: the only information flow between the two sides of a bottleneck is ϵ units of attention per crossing edge, so consensus across the bottleneck slows in proportion to $1/\epsilon$. The tension between consensus accuracy and convergence speed is a natural direction for future work.

C. Dual Problem

Having demonstrated that first-best consensus is preserved in the $\epsilon \leq \epsilon^*$ regime, we now turn to $\epsilon > \epsilon^*$. In this regime, the minimum attention constraints prevent uniform influence, and the equilibrium π^* deviates from π_n. We derive the Lagrange dual of the convex formulation to determine which constraints are responsible for this deviation and what each costs in terms of consensus accuracy. The dual variables will provide a characterization of the equilibrium structure, including a partition of each agent's neighborhood into voluntarily weighted neighbors and those receiving forced weight, a threshold rule governing attention allocation, and shadow prices for missing edges that can inform network design.

The Lagrange dual of the convex formulation (22) assigns a multiplier α_i to each row-sum constraint and β_i to each column-sum constraint (the full derivation appears in Appendix A.A6). Stationarity of the Lagrangian with respect to B_{ij} yields the edge price

$$\lambda_{ij} = \alpha_i + \beta_j \quad \forall\, (i,j) \in E, \tag{27}$$

the shadow price on the minimum attention constraint $M_{ij} \geq \epsilon$. On non-edges, the corresponding multiplier is $\mu_{ij} = -(\alpha_i + \beta_j)$. Since $\lambda_{ij} \geq 0$, we require $\alpha_i + \beta_j \geq 0$ on all edges. The primal problem satisfies Slater's condition, so strong duality holds and the dual optimum is attained.

Complementary slackness on the minimum attention constraints yields, for each edge $(i,j) \in E$:

$$\lambda_{ij}(\epsilon\pi_i^* - B_{ij}^*) = 0. \tag{28}$$

Condition (28) partitions each agent i's neighborhood into three sets:

(29)

$$\begin{aligned}
C(i) &= \{j \in N(i) : M_{ij}^* > \epsilon\} && \text{(voluntary: } \lambda_{ij} = 0\text{, constraint slack)},\\
F(i) &= \{j \in N(i) : M_{ij}^* = \epsilon,\ \lambda_{ij} > 0\} && \text{(forced: constraint binds at positive cost)},\\
G(i) &= \{j \in N(i) : M_{ij}^* = \epsilon,\ \lambda_{ij} = 0\} && \text{(marginal: constraint binds at zero cost)},
\end{aligned}$$

where $M_{ij}^* = B_{ij}^*/\pi_i^*$. A voluntary neighbor receives attention beyond the minimum. A forced neighbor receives exactly the minimum, and the positive multiplier

$\lambda_{ij} > 0$ reflects the welfare cost of this obligation. Agent i would prefer to give j even less weight. A marginal neighbor also receives exactly the minimum, but at zero cost: agent i would voluntarily choose exactly ϵ weight even without the constraint.

D. Shadow Prices on Forced Attention

We now show that the dual variables of the convex formulation are determined by the score vector $s^* = Z^*\pi^*$ from Section II. This connects the dual problem to the best-response analysis: the welfare cost of each minimum attention constraint can be read off from the network structure through the score vector.

Recall from the constrained first-order conditions (24) that each agent's voluntarily weighted neighbors share a common score τ_i, while forced neighbors have score $s_j^* \geq \tau_i$. Agent i allocates her discretionary weight, the $1 - \epsilon|N(i)|$ units left after meeting the minimum attention constraint, to neighbors whose scores equal τ_i. Neighbors who are given exactly ϵ weight are either marginal (the constraint binds at zero cost) or forced (the constraint binds at positive cost). Forced neighbors are over-represented in the consensus, and i would prefer to listen to them less but cannot. The multiplier $\rho_{ij} = 2\pi_i^*(s_j^* - \tau_i)$ measures the individual welfare cost of this obligation in agent i's problem. We are now ready to state the main result.

Theorem 2 (Score representation of the dual variables)**.** *At any Nash equilibrium M^* with $\epsilon > 0$, the edge prices from the convex formulation satisfy*

$$\lambda_{ij} = \alpha_i + \beta_j = 2(s_j^* - \tau_i) \quad \forall (i,j) \in E, \tag{30}$$

where $s^* = Z^*\pi^*$ is the equilibrium score vector and τ_i is the threshold defined above. The dual variables decompose as

$$\beta_j = 2s_j^* + c, \qquad \alpha_i = -2\tau_i - c, \tag{31}$$

for a global gauge constant $c \in \mathbb{R}$.

PROOF:

Let M^* be any Nash equilibrium with social influence π^* and scores $s^* = Z^*\pi^*$. The first-order conditions (24) hold at every agent. For an arbitrary constant c, define

$$\lambda_{ij} = 2(s_j^* - \tau_i), \quad \alpha_i = -2\tau_i - c, \quad \beta_j = 2s_j^* + c, \quad \nu_i = 0, \quad \gamma = -2\|\pi^*\|^2,$$

and $\mu_{ij} = -(\alpha_i + \beta_j)$ on non-edges. We verify these satisfy the KKT conditions of the convex program (22) at the primal (π^*, B^*) with $B^* = \text{diag}(\pi^*)M^*$.

Dual feasibility. $\lambda_{ij} = 2(s_j^* - \tau_i) \geq 0$ by (24); $\nu_i = 0 \geq 0$.

Stationarity in B. On edges, $\alpha_i + \beta_j - \lambda_{ij} = (-2\tau_i - c) + (2s_j^* + c) - 2(s_j^* - \tau_i) = 0$; on non-edges, $\alpha_i + \beta_j + \mu_{ij} = 0$ by definition.

Complementary slackness. On voluntary edges, (24) gives $s_j^* = \tau_i$, so $\lambda_{ij} = 0$. On floored edges, $B_{ij}^* = \epsilon\pi_i^*$, so the constraint is binding. In both cases $\lambda_{ij}(\epsilon\pi_i^* - B_{ij}^*) = 0$. Since $\pi^* > 0$ and $\nu_i = 0$, we also have $\nu_i\pi_i^* = 0$.

Stationarity in π. Write $M_{ij}^* = \epsilon + (M_{ij}^* - \epsilon)$, where the second term is positive only on voluntary edges, which carry $s_j^* = \tau_i$. Then

$$(M^*s^*)_i = \epsilon \sum_{j \in N(i)} s_j^* + \tau_i\big(1 - \epsilon|N(i)|\big),$$

so $(s_i^* - \tau_i) - \epsilon\sum_{j\in N(i)}(s_j^* - \tau_i) = s_i^* - (M^*s^*)_i = \pi_i^* - \|\pi^*\|^2$ by Lemma 1. Substituting into the π-stationarity condition (A38):

$$2\pi_i^* - 2\Big[(s_i^* - \tau_i) - \epsilon \sum_{j \in N(i)} (s_j^* - \tau_i)\Big] + \gamma = 2\|\pi^*\|^2 + \gamma,$$

which vanishes for every i with $\gamma = -2\|\pi^*\|^2$.

All KKT conditions hold. Since the program is convex, (π^*, B^*) is a global minimizer; strict convexity in π gives uniqueness. This establishes Lemma 2: every Nash equilibrium achieves π^*. The construction directly gives (30) and (31), with c the gauge freedom from the linear dependence between the row-sum and column-sum constraints.

Interpretation

Theorem 2 bridges the best-response analysis developed in Section II and the full game's convex formulation. The dual prices α_i and β_j are determined by the score vector $s^* = Z^*\pi^*$ and the combinatorial structure it encodes, characterized in Section II.

Recall from Section II that at an unconstrained best response, each agent equalizes scores among her chosen neighbors. To accomplish this, she shifts weight away from high-score neighbors and towards low-score neighbors until all voluntarily weighted neighbors share the common score τ_i. The binding minimum attention constraint $\epsilon > \epsilon^*$ prevents full equalization. Forced neighbors have scores above τ_i; agent i would like to reduce their weight further but cannot. The edge price $\lambda_{ij} = 2(s_j^* - \tau_i)$ is precisely the gap between the forced neighbor j's score and the level at which i has equalized her chosen neighbors. The welfare cost of a forced edge is large when that neighbor's score is far above the desired equalization level τ_i.

The welfare cost of the floor on edge (i, j) is $\rho_{ij} = \lambda_{ij}\pi_i^*$: relaxing the constraint $M_{ij} \geq \epsilon$ by δ decreases the equilibrium variance by approximately $\rho_{ij}\delta$. A planner who could selectively reduce the minimum attention requirement on specific edges should target those with the largest $\rho_{ij} = 2\pi_i^*(s_j^* - \tau_i)$: the welfare cost of a forced edge depends on both the score gap (how redundant the forced neighbor is) and the agent's influence (how much her reallocation moves the con-

sensus). The minimum attention parameter ϵ reflects the difficulty of ignoring a social connection: agents who share a classroom, workplace, or living space will inevitably absorb some of each other's views.

Consider a company that requires all employees to attend a weekly all-hands meeting led by the CEO. The CEO's perspective is already widespread in the organization, thus they represent a high-score or *redundant* node. Employees who interact with the CEO will have heard many of the same talking points from their other coworkers. The all-hands meeting forces every employee to pay some minimum attention to the most redundant source of information in the network; a planner might instead make the all-hands optional for some employees, freeing attention that could be voluntarily redirected towards colleagues in distant parts of the organization.

The planner does not need to tell agents who to listen to. She only needs to remove constraints that force attention toward over-represented sources. Agents' own selfish optimization will redirect the freed attention toward under-represented sources.

For $(i, j) \notin E$, the multiplier on the zero-flow constraint is $\mu_{ij} = 2(\tau_i - s_j^*)$, following from the established fact that $\mu_{ij} = -(\alpha_i + \beta_j)$. This gives the rate of decrease in $\|\pi^*\|^2$ per unit of flow B_{ij} allowed on the absent edge. Adding optional flow on a missing link from i to j is valuable when $s_j^* < \tau_i$: agent j's score is below i's equalization target, meaning j is more under-represented than any of i's current chosen neighbors. If i could listen to j, she would voluntarily shift weight toward j and away from her existing neighbors, bringing all of their scores closer. When $s_j^* \geq \tau_i$, the shadow price is non-positive; even if the edge were available, the planner would choose not to use it, since j is no less over-represented than i's current chosen neighbors.

It is worth noting that the shadow price on missing edges answers a narrower question than the natural one a network designer would face. It measures the value of making flow *possible* on a currently absent edge, without requiring that ϵ attention flow across that edge. If the planner were to add a new edge in the usual sense (that is, with the minimum attention constraint), the obligation could outweigh the informational benefit. Even when $s_j^* < \tau_i$, the ϵ floor may force more attention than is optimal.

Returning to the workplace example: suppose the company is considering connecting an employee at its headquarters to one in a satellite office through a cross-office working group. The shadow price tells the planner whether the satellite employee's perspective is under-represented enough (low s_j^*) relative to the headquarters employee's current options (τ_i) for the connection to be beneficial. But the working group also requires ongoing participation (a minimum attention cost). If the satellite office's perspective is only marginally under-represented, the information gain may not be worth diverting the headquarters employee's attention from her other colleagues.

Example. We return to the directed bottleneck at $\epsilon = 3/10 > \epsilon^* = 1/4$

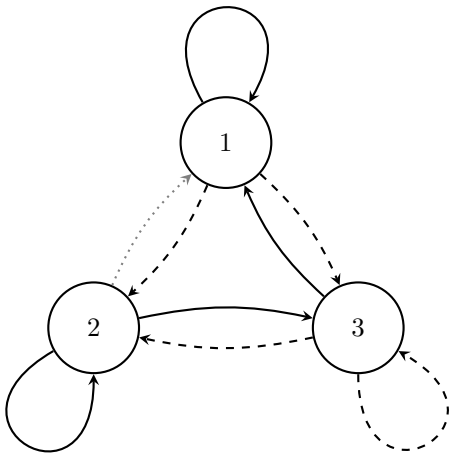


FIGURE 4. EQUILIBRIUM AT $\epsilon = \frac{3}{10} > \epsilon^* = \frac{1}{4}$. SOLID: VOLUNTARY ($\lambda_{ij} = 0$). DASHED: FORCED AT ϵ ($\lambda_{ij} = \frac{5}{19}$). DOTTED: MISSING EDGE ($\mu_{21} = \frac{5}{19}$).

to illustrate how the score vector identifies the source of welfare costs through shadow prices.

The network has an asymmetry: agents 2 and 3 receive attention from every direction, while agent 1 is heard only by herself and agent 3. The score vector captures this: $s^* \approx (0.245, 0.377, 0.377)$. Agents 2 and 3 are well-represented in the consensus and "close" to each other (in a random-walk sense), so their information is redundant. Agent 1 is only reachable through agent 3 and so has a low score; increasing weight on her perspective would be the most valuable for consensus.

At equilibrium, agents 1 and 3 each place their discretionary weight (0.1 units remaining after meeting the minimum constraint) on agent 1, and the minimum constraint is binding on their other edges. Agent 1 is the lowest-score neighbor available to both, so agent 3 wants to shift weight towards agent 1, and agent 1 wants to shift weight toward herself. However, the minimum weight constraint mandates $\epsilon = 0.3$ to agents 2 and 3, diverting attention toward those neighbors who are already over-represented in the consensus. The resulting influence is $\pi^* = (5/19, 13/38, 15/38)$, with agents 2 and 3 receiving disproportionate influence.

Fig. 4 displays the example graph with forced edges as dashed, voluntary edges as solid, and the missing edge as dotted. The edge price $\lambda = 5/19$ on each forced edge is the score gap between the forced neighbor and the outgoing agent's equalization level. The welfare benefit of relaxing a specific forced edge is $\rho_{ij}\delta = \lambda_{ij}\pi_i^*\delta$ per unit of relaxation δ. This benefit is larger for agent 3's forced

edges ($\rho = 75/722$) than agent 1's ($\rho = 25/361$): agent 3 has more influence, so her reallocation moves more weight in the consensus.

The missing edge $(2, 1)$ reveals the network structure's effect on welfare. Agent 2 cannot hear agent 1 at all, and the shadow price $\mu = 5/19$ tells the planner this correction would be valuable: agent 1's score is well below the score level of agent 2's neighbors, so agent 2 would shift weight toward 1 if the link existed. Naturally, the missing edge that constrained ϵ^* in Section III.B is the same one the shadow prices identify as costly above ϵ^*.

IV. Concluding Remarks

This paper studied a game of strategic attention allocation on networks, where agents choose how to weight their neighbors before engaging in DeGroot averaging. The unifying mathematical object is a score vector that ranks every agent by her marginal informational value. All agents agree on this ranking. Without minimum attention constraints, agents can fully correct the double-counting bias and achieve uniform influence at every strongly connected equilibrium, though echo chambers may also arise due to coordination failure. With minimum attention constraints, a network-dependent threshold on the minimum attention level separates two regimes: below the threshold, agents retain enough free attention to achieve the first-best; above it, forced attention on costly edges drives the network away from uniform influence. The score vector quantifies edge-by-edge costs through shadow prices in the game's dual.

Limitations

Several modeling choices limit the scope of our results. We assume homogeneous signal variance: all agents receive equally informative signals, enabling the common payoff that simplifies our analysis in the constrained game. We suspect our analytical framework can be easily adapted to heterogeneous signals given an appropriate choice of objective, as in Como, Fagnani and Proskurnikov (2025). Furthermore, we assume agents choose weights once before updating begins, rather than adapting weights over time in response to the beliefs they observe. We assume agents observe the full network structure when choosing weights, an unrealistic assumption for large networks. Finally, we assume independent signals, so the score vector only captures redundancy in the network. With correlated signals, a neighbor's information can be redundant even without double-counting through paths in the network, and optimal attention allocation would depend on the signal covariance structure and network topology together.

Future work

Convergence speed. Our analysis characterizes equilibrium accuracy (the variance $\|\pi^*\|^2$) but does not examine speed of convergence to consensus. A potentially large set of weight matrices can implement equilibria with the same

social influence vector, but they can differ substantially in convergence speed. The Cheeger bound in Section III.B implies that worst-case relaxation grows proportional to $1/\epsilon$ across sparse bottlenecks, suggesting that equilibria which are payoff-equivalent may be far from equivalent in practice. More broadly, future work could address the setting where agents can trade some consensus accuracy for faster convergence, characterizing the Pareto frontier between accuracy and mixing time.

Comparative statics on ϵ. Our results characterize the equilibrium at each fixed ϵ but say little about how equilibrium changes as ϵ varies. How does the influence vector π^* move as ϵ crosses the threshold ϵ^*? Is the transition smooth or discontinuous? How does the set of forced edges evolve? The convex reformulation makes $\pi^*(\epsilon)$ a well-defined function, enabling such comparative statics. A full analysis would clarify how welfare degrades as agents lose control over their attention.

Large-network asymptotics. Golub and Jackson (2010) studied wisdom of crowds in large networks, but our objects of analysis (ϵ^*, the score vector) are defined for finite networks. Whether the two-regime structure survives $n \to \infty$, and how ϵ^* scales with network size and structure are open questions.

Endogenous network formation. We take the network D as given and study weight allocation on a fixed set of edges. A natural extension allows agents to form or sever links alongside choosing weights. The shadow prices on missing edges provide a starting point: they identify which absent connections would be most valuable for consensus accuracy on the margin.

Value of strategic attention. A natural benchmark is naive DeGroot updating with uniform weights. The gap $\|\pi_{\text{uni}}\|^2 - \|\pi^*\|^2$ measures how much welfare strategic weight choice recovers on a given network. On regular graphs, uniform weights already produce doubly stochastic matrices, so the gap is zero. On irregular networks with high-degree hubs, uniform weights concentrate influence, so the gap can be substantial. Characterizing which network structures benefit most from strategic attention allocation is an open question.

REFERENCES

Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar. 2014. "Dynamics of information exchange in endogenous social networks." *Theoretical Economics*, 9(1): 41–97.

Breen, Jane, and Stephen Kirkland. 2015. "Stationary vectors of stochastic matrices subject to combinatorial constraints." *The Electronic Journal of Linear Algebra*, 28: 25–53.

Como, Giacomo, Fabio Fagnani, and Anton Proskurnikov. 2025. "Wisdom of Crowds Through Myopic Self-Confidence Adaptation." arXiv:2506.18195 [math].

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel. 2003. "Persuasion Bias, Social Influence, and Unidimensional Opinions*." *The Quarterly Journal of Economics*, 118(3): 909–968.

Gale, David. 1957. "A theorem on flows in networks." *Pacific Journal of Mathematics*, 7(2): 1073–1082.

Ghosh, Arpita, Stephen Boyd, and Amin Saberi. 2008. "Minimizing Effective Resistance of a Graph." *SIAM Review*, 50(1): 37–66.

Golub, Benjamin, and Matthew O. Jackson. 2010. "Naïve Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics*, 2(1): 112–149.

Grinstead, Charles Miller, and James Laurie Snell. 1997. *Introduction to Probability.* American Mathematical Soc. Google-Books-ID: 14oq4uWGCkwC.

Levin, D.A., Y. Peres, and E.L. Wilmer. 2008. *Markov Chains and Mixing Times.* American Mathematical Soc.

Schweitzer, Paul J. 1968. "Perturbation theory and finite Markov chains." *Journal of Applied Probability*, 5(2): 401–413.

Tetali, Prasad. 1991. "Random walks and the effective resistance of networks." *Journal of Theoretical Probability*, 4(1): 101–109.

Mathematical Appendix

A1. MFPT Interpretation of Score Vector

The fundamental matrix Z of an ergodic Markov chain M may be used to compute the *mean first passage time* (MFPT) m_{ij} in a weighted random walk from any node to any other node in the graph (Grinstead and Snell (1997)):

$$m_{kj} = \frac{Z_{jj} - Z_{kj}}{\pi_j} \implies Z_{kj} = Z_{jj} - \pi_j m_{kj}, \tag{A1}$$

where π is the stationary vector of M. We then have

$$s_k = \sum_j \pi_j Z_{kj} = \sum_j \pi_j Z_{jj} - \sum_j \pi_j^2 m_{kj}, \tag{A2}$$

where $\sum_j \pi_j Z_{jj}$ is constant across entries of s.

A2. Score Vector in Reversible Networks

As DeMarzo, Vayanos and Zwiebel (2003) point out, the social influence vector in the reversible setting has a simple form:

$$\pi_i = \frac{d_i}{\sum_j d_j}, \tag{A3}$$

where d_i is the degree of agent i. That is, social influence is directly proportional to degree. Since this is a reversible Markov chain, we may also make use of a formula due to Tetali (1991) who write the mean first passage times in terms of *effective resistance*:

$$m_{ij} = \frac{1}{2}\sum_k d_k(R_{ij} + R_{jk} - R_{ik}). \tag{A4}$$

The effective resistance from i to j, denoted R_{ij}, is the voltage across i and j when a unit current is applied at i, treating the network as an electrical circuit with a 1-Ω resistor on each edge (Ghosh, Boyd and Saberi (2008)). Intuitively, R_{ij} is small when there are many short distinct paths between i and j, and R_{ij} is large when there are fewer, longer paths between i and j. The effective resistance measure captures our "double-counting" intuition well—the presence of many short paths between two nodes will cause them to hear each other's opinions many times.

Applying Tetali's formula to $\sum_j \pi_j^2 m_{ij}$:

$$s_i = const - \sum_j \pi_j^2 \frac{1}{2}\sum_k d_k(R_{ij} + R_{jk} - R_{ik}). \tag{A5}$$

Note that we can move terms involving R_{jk} to the *const* since they don't depend on i:

$$s_i = const - \frac{1}{2}\sum_j \pi_j^2 \sum_k d_k(R_{ij} - R_{ik}). \tag{A6}$$

We also have that d_k is exactly $\pi_k \sum_j d_j$ in this setting, so we will also suppress this constant:

$$s_i = const - const \sum_{j,k} \pi_j^2 \pi_k (R_{ij} - R_{ik}). \tag{A7}$$

Notice that

$$\sum_{j,k} \pi_j^2 \pi_k (R_{ij} - R_{ik}) = \sum_{j,k} \pi_j^2 \pi_k R_{ij} - \sum_{j,k} \pi_j^2 \pi_k R_{ik} \tag{A8}$$

$$= \sum_{j,k} \pi_j^2 \pi_k R_{ij} - \sum_{j,k} \pi_k^2 \pi_j R_{ij} = \sum_{j,k} (\pi_j^2 \pi_k - \pi_k^2 \pi_j) R_{ij}, \tag{A9}$$

where we re-indexed the second term. Then we have

$$\sum_j (\pi_j^2 \sum_k \pi_k - \pi_j (\sum_k \pi_k^2)) R_{ij} = \sum_j (\pi_j^2 - \pi_j (\sum_k \pi_k^2)) R_{ij}, \tag{A10}$$

where we used the fact that $\sum_k \pi_k = 1$. Finally, we have

$$\sum_j \pi_j^2 R_{ij} - (\sum_k \pi_k^2) \sum_j \pi_j R_{ij}. \tag{A11}$$

Now suppose we sample a node randomly according to π. Let $\mathbb{E}_\pi$ denote the expectation wrt the π-distribution, and let J be our sampled index. Then

$$\mathbb{E}_\pi[\pi_J R_{iJ}] = \sum_j \pi_j (\pi_j R_{ij}) = \sum_j \pi_j^2 R_{ij} \tag{A12}$$

$$\mathbb{E}_\pi[\pi_J] = \sum_j \pi_j \pi_j = \sum_j \pi_j^2 \tag{A13}$$

$$\mathbb{E}_\pi[R_{iJ}] = \sum_j \pi_j R_{ij}. \tag{A14}$$

Therefore (A11) becomes

$$\mathbb{E}_\pi[\pi_J R_{iJ}] - \mathbb{E}_\pi[\pi_J] \mathbb{E}_\pi[R_{iJ}] = Cov_\pi(\pi_J, R_{iJ}). \tag{A15}$$

Substituting back, we have

$$s_i = a - b \cdot Cov_\pi(\pi_J, R_{iJ}), \quad b > 0, \tag{A16}$$

for graph-dependent constants a and b, where $Cov_\pi(\pi_J, R_{iJ})$ is the covariance between the social influence of J and the effective resistance from i to J for a randomly selected J (sampled from π, that is, with probability exactly equal to social influence). Agent i will want to place more marginal weight on agents j whose $Cov_\pi(\pi_J, R_{jJ})$ is large; these agents tend to have high resistance to high influence parts of the graph, and likewise low resistance to low influence parts. Meanwhile, i will remove weight from agents who have low resistance to high influence parts and high resistance to low influence parts.

A3. Proof of Proposition 2

PROOF:

The proof is written for the case $\epsilon = 0$; the argument for $\epsilon > 0$ is identical with the inactive set redefined as $I = \{j : w_j^* = \epsilon\}$. For readability, we will suppress the starting point w^* in our notation throughout the proof, abbreviating $F(w^*)$, $Z(w^*)$, and $\pi(w^*)$.

From earlier, we have that $d\pi = \pi_i Z^\top \alpha$ and that the α-directional derivative of F is $\nabla_\alpha F = 2\pi_i s^\top \alpha$, where $s = Z\pi$. We will first need to derive the second α-directional derivative. Since $Z = (I - M + \mathbf{1}\pi^\top)^{-1}$ depends smoothly on M, the standard identity $d(A^{-1}) = -A^{-1}(dA)A^{-1}$ gives

$$dZ = ZdMZ - Z\mathbf{1}(d\pi)^\top Z. \tag{A17}$$

Substitute $dM = e_i\alpha^\top$:

$$dZ = Z(e_i\alpha^\top)Z - Z\mathbf{1}(d\pi^\top Z). \tag{A18}$$

Transpose both sides:

$$dZ^\top = Z^\top(\alpha e_i^\top)Z^\top - Z^\top d\pi \mathbf{1}^\top Z^\top. \tag{A19}$$

Right-multiply both sides by α:

$$dZ^\top\alpha = Z^\top(\alpha e_i^\top)Z^\top\alpha - Z^\top d\pi \mathbf{1}^\top Z^\top\alpha. \tag{A20}$$

Note that $e_i^\top Z^\top\alpha$ is a scalar, namely $(Z^\top\alpha)_i$. Then the first term is $Z^\top(\alpha e_i^\top)Z^\top\alpha = (Z^\top\alpha)_i(Z^\top\alpha)$. We can also show that the second term vanishes. Note that $\mathbf{1}^\top Z^\top\alpha = (Z\mathbf{1})^\top\alpha$. Z satisfies $(I - M)Z = I - \mathbf{1}\pi^\top$, and we can right-multiply by $\mathbf{1}$ to see that $(I - M)(Z\mathbf{1}) = 0$, or that $Z\mathbf{1}$ is in the null space of $I - M$. Since M is row-stochastic, we have $M\mathbf{1} = \mathbf{1}$, so $\mathbf{1}$ is a right-eigenvector of M with eigenvalue 1. The Perron-Frobenius theorem tells us that the eigenvalue 1 is simple, so the right-eigenspace corresponding to 1 is span$\{\mathbf{1}\}$. We conclude that the null space of $I - M$ is span$\{\mathbf{1}\}$, so $Z\mathbf{1} = c\mathbf{1}$ for some scalar constant c. Then $(Z\mathbf{1})^\top\alpha = (c\mathbf{1})^\top\alpha = c\mathbf{1}^\top\alpha = 0$ by the feasibility condition $\mathbf{1}^\top\alpha = 0$. We conclude that $(dZ^\top)\alpha = (Z^\top\alpha)_i(Z^\top\alpha)$.

As we showed earlier, the first α-directional derivative of $F = \pi^\top\pi$ is $2\pi_i s^\top\alpha$. To obtain the second directional derivative, note that

$$\pi'' = \pi_i' Z^\top\alpha + \pi_i(dZ^\top)\alpha = 2\pi_i(Z^\top\alpha)_i Z^\top\alpha. \tag{A21}$$

Therefore

$$F'' = 2\|\pi'\|_2^2 + 2\pi^\top\pi'' \tag{A22}$$

$$2\|\pi'\|_2^2 = 2\pi_i^2\|Z^\top\alpha\|_2^2 \tag{A23}$$

$$2\pi^\top\pi'' = 2\pi^\top(2\pi_i(Z^\top\alpha)_i Z^\top\alpha) = 4\pi_i(Z^\top\alpha)_i\pi^\top Z^\top\alpha = 4\pi_i(Z^\top\alpha)_i s^\top\alpha. \tag{A24}$$

Then

$$F'' = 2\pi_i^2\|Z^\top\alpha\|_2^2 + 4\pi_i(Z^\top\alpha)_i s^\top\alpha. \tag{A25}$$

We now need to be slightly more careful with our choice of direction α. Define the active neighbor set $A = \{j : w_j^* > 0\}$ and inactive neighbor set $I = \{j : w_j^* = 0\}$ at w^*; A is the set of neighbors who i places positive weight on at w^* and I is the set of neighbors who i does not listen to at w^*. From the KKT conditions, we know that $s_j = \tau_i$ for $j \in A$ and $s_j \geq \tau_i$ for $j \in I$ for some $\tau_i \in \mathbb{R}$. A direction is feasible if $\mathbf{1}^\top\alpha = 0$ and $\alpha_j \geq 0$ for all $j \in I$, that is, if α does not shift weight away from neighbors who already have 0 weight. For any such feasible α:

$$s^\top\alpha = \sum_{j\in A}\alpha_j\tau_i + \sum_{j\in I}\alpha_j s_j = \tau_i\sum_{j\in N(i)}\alpha_j + \sum_{j\in I}\alpha_j(s_j - \tau_i) = \sum_{j\in I}\alpha_j(s_j - \tau_i) \geq 0. \tag{A26}$$

Thus for feasible α, $F' = 2\pi_i s^\top\alpha \geq 0$. We can further split the inactive set I into $I_{\text{str}} = \{j \in I : s_j > \tau_i\}$ and $I_{\text{eq}} = \{j \in I : s_j = \tau_i\}$. Define the critical cone:

$$\mathcal{C} = \left\{\alpha : \mathbf{1}^\top\alpha = 0,\ \alpha_j = 0\ \forall j \in I_{\text{str}},\ \alpha_j \geq 0\ \forall j \in I_{\text{eq}}\right\}. \tag{A27}$$

That is, $\mathcal{C}$ is the set of feasible directions that *do not* activate any inactive neighbors. Suppose $\alpha \notin \mathcal{C}$. Then there exists some $j \in I_{\text{str}}$ such that $\alpha_j > 0$. Therefore

$$s^\top\alpha = \sum_{j\in I}\alpha_j(s_j - \tau_i) \geq \alpha_j(s_j - \tau_i) > 0, \tag{A28}$$

so $F' = 2\pi_i s^\top\alpha > 0$, or in other words the first-order directional derivative is positive.

Next, suppose $\alpha \in \mathcal{C}$. Then any inactive neighbor $j \in I$ with $\alpha_j > 0$ satisfies $s_j = \tau_i$. For $j \in A$, we also have $s_j = \tau_i$. Then $s^\top\alpha = \tau_i\sum_j\alpha_j = 0$ and the first-order change vanishes. The cross-term $4\pi_i(Z^\top\alpha)_i s^\top\alpha$ in F'' also vanishes, leaving $F'' = 2\pi_i^2\|Z^\top\alpha\|_2^2$.

Since Z is invertible (following from ergodicity), its null-space is exactly $\{0\}$ Therefore $Z^\top\alpha = 0$ if and only if $\alpha = 0$, so $F'' = 2\pi_i^2\|Z^\top\alpha\|_2^2 > 0$ for our choice of α. We have shown for every feasible $\alpha \neq 0$, either $F' > 0$ or $F' = 0$ and $F'' > 0$, so w^* must be a strict local minimum. □

A4. Non-convexity of F

Consider the four-agent network where agent 1's full weight vector is (w_1, w_2, w_3, w_4) with $\sum_k w_k = 1$ and $w_k \geq 0$. We restrict to $w_1 = 0$, parametrizing the remaining weights as $w = (a, b, c) = (w_2, w_3, w_4)$:

$$M(w) = \begin{bmatrix} 0 & a & b & c \\ 0.05 & 0.05 & 0.90 & 0 \\ 0.05 & 0.90 & 0.05 & 0 \\ 0.90 & 0.05 & 0.05 & 0 \end{bmatrix}, \qquad a, b, c \geq 0, \quad a + b + c = 1. \tag{A29}$$

Let $F(a, b, c) = \|\pi(M(a, b, c))\|^2$. Take $w^{(1)} = (0, 0, 1)$ and $w^{(2)} = (\frac{1}{2}, 0, \frac{1}{2})$, with midpoint $\overline{w} = (\frac{1}{4}, 0, \frac{3}{4})$. Evaluating:

$$F(w^{(1)}) = 0.25, \quad F(w^{(2)}) \approx 0.395434, \quad F(\overline{w}) \approx 0.333467$$

Convexity would require $F(\overline{w}) \leq \frac{1}{2}F(w^{(1)}) + \frac{1}{2}F(w^{(2)}) = 0.322717$. Since $0.333467 > 0.322717$, F is not convex on the region where $w_1 = 0$. Since the restriction of a convex function to any convex subset of its domain remains convex, non-convexity on $\{w_1 = 0\}$ implies non-convexity of F over agent 1's full simplex, holding the rest of the network fixed.

A5. Proof of Proposition 6

The first-best is achievable if and only if a doubly stochastic matrix M supported on D with $M_{ij} \geq \epsilon$ exists on all edges, as established in the main text. We show this is equivalent to the condition stated in Proposition 6.

Reduction to a transportation problem.

Suppose we have a doubly stochastic M with $M_{ij} \geq \epsilon$ for all edges. Let $\tilde{M}_{ij} = M_{ij} - \epsilon$. Then $\tilde{M}_{ij} \geq 0$ and

$$\sum_{j \in N(i)} \tilde{M}_{ij} = \sum_{j \in N(i)} (M_{ij} - \epsilon) = 1 - \epsilon \cdot d_i^{out} = q_i \tag{A30}$$

$$\sum_{i:(i,j) \in E} \tilde{M}_{ij} = \sum_{i:(i,j) \in E} (M_{ij} - \epsilon) = 1 - \epsilon \cdot d_j^{in} = r_j. \tag{A31}$$

Conversely, given nonnegative flows $\tilde{M}$ with these row and column sums, setting $M_{ij} = \tilde{M}_{ij} + \epsilon$ on edges recovers a doubly stochastic matrix with entries at least ϵ. Therefore the problem is equivalent to finding a nonnegative matrix supported on D with row sums q_i and column sums r_j. This is a balanced transportation problem, since $\sum q_i = n - \epsilon|E| = \sum r_j$.

Application of the Gale-Hoffman Feasibility Theorem. The condition

$\epsilon \leq 1/\bar{d}$ ensures $q_i \geq 0$ (since $\epsilon d_i^{out} \leq 1$) and $r_j \geq 0$ (since $\epsilon d_j^{in} \leq 1$) for all agents. We embed the transportation problem in a bipartite graph with supplier nodes $\{1, \ldots, n\}$ and demander nodes $\{1', \ldots, n'\}$, where $N' = \{1, \ldots, n\} \cup \{1', \ldots, n'\}$. Wire edges such that supply i connects to demander j' if and only if $(i, j) \in E$. We want to find edge flows from the suppliers to the demanders such that all supply is exhausted and all demand is met.

By the feasibility theorem of Gale (1957), the demand is satisfiable if and only if for every subset S of nodes, the excess demand of the complement $S' = N' \setminus S$ does not exceed the capacity of the cut from S to S'. Suppose S contains some suppliers $T \subseteq \{1, \ldots, n\}$ and some demanders $U \subseteq \{1', \ldots, n'\}$. Since capacities are infinite on existing edges and zero on non-edges, the cut capacity $c(S, S')$ is infinite whenever any supplier in T has an edge to a demander outside U; equivalently, whenever $N^{out}(T) \not\subset U$. The condition is therefore vacuous unless $N^{out} \subseteq U$, in which case $c(S, S') = 0$. The condition thus requires that the net excess demand of the complement does not exceed zero:

$$-\sum_{i \notin T} q_i + \sum_{j \notin U} r_j \leq 0. \tag{A32}$$

Using balance ($\sum q_i = \sum r_j$), this is equivalent to $\sum_{i \in T} q_i \leq \sum_{j \in U} r_j$. Since this must hold for all $U \supseteq N^{out}(T)$, and the right side is minimized when $U = N^{out}(T)$, the binding condition is

$$\sum_{i \in T} q_i \leq \sum_{j \in N^{out}(T)} r_j \quad \forall\, T \subseteq N. \tag{A33}$$

□

A6. Dual Variable Characterization

We introduce dual variables for each constraint in (22):

- $\alpha_i \in \mathbb{R}$ for the row-sum constraint $\sum_j B_{ij} = \pi_i$,
- $\beta_j \in \mathbb{R}$ for the column-sum constraint $\sum_i B_{ij} = \pi_j$,
- $\gamma \in \mathbb{R}$ for the simplex constraint $\pi^\top \mathbf{1} = 1$,
- $\lambda_{ij} \geq 0$ for each minimum attention constraint $\epsilon\pi_i - B_{ij} \leq 0$ on edges $(i, j) \in E$,
- $\nu_i \geq 0$ for each nonnegativity constraint $-\pi_i \leq 0$,
- $\mu_{ij} \in \mathbb{R}$ for each zero-flow constraint $B_{ij} = 0$ on non-edges $(i, j) \notin E$.

Lagrangian

The Lagrangian is

(A34)
$$\mathcal{L}(\pi, B, \alpha, \beta, \gamma, \lambda, \nu, \mu) = \|\pi\|^2 + \sum_i \alpha_i \left(\sum_j B_{ij} - \pi_i \right) + \sum_j \beta_j \left(\sum_i B_{ij} - \pi_j \right) + \gamma(\pi^\top \mathbf{1} - 1) + \sum_{(i,j)\in E} \lambda_{ij}(\epsilon\pi_i - B_{ij}) + \sum_i \nu_i(-\pi_i) + \sum_{(i,j)\notin E} \mu_{ij} B_{ij}.$$

Stationarity in B_{ij}.

Differentiating $\mathcal{L}$ with respect to B_{ij} and setting the result to zero yields:

(A35)
$$\frac{\partial \mathcal{L}}{\partial B_{ij}} = \begin{cases} \alpha_i + \beta_j - \lambda_{ij} = 0 & \text{if } (i,j) \in E, \\ \alpha_i + \beta_j + \mu_{ij} = 0 & \text{if } (i,j) \notin E. \end{cases}$$

These will become the constraints in the dual problem. On edges, we may solve for the inequality multiplier:

(A36)
$$\lambda_{ij} = \alpha_i + \beta_j, \quad (i,j) \in E.$$

On non-edges, we solve for the equality multiplier:

(A37)
$$\mu_{ij} = -(\alpha_i + \beta_j), \quad (i,j) \notin E.$$

Since $\lambda_{ij} \geq 0$, the stationarity conditions on edges impose the constraint $\alpha_i + \beta_j \geq 0$ for all $(i,j) \in E$. The absent edge multipliers μ_{ij} are unrestricted in sign and fully determined by α and β.

Stationarity in π_i.

Differentiating $\mathcal{L}$ with respect to π_i and setting it to zero gives

(A38)
$$\frac{\partial \mathcal{L}}{\partial \pi_i} = 2\pi_i - \alpha_i - \beta_i + \gamma + \epsilon \sum_{j:(i,j)\in E} \lambda_{ij} - \nu_i = 0,$$

where the $\epsilon \sum_j \lambda_{ij}$ term comes from the minimum-attention constraints $\epsilon\pi_i - B_{ij} \leq 0$ and $-\nu_i$ from the nonnegativity constraint $-\pi_i \leq 0$.