

Network Games Seminar

Lecture 2: The spread of information across a network

1 Introduction

We first looked at why and how individuals might form networks. We saw that network formation is influenced by:

- the form of each individual's utility function (the value that they do or do not get from being connected to other people), and
- the way a 'stable' link is defined (do both parties need to want the link, or just one?)

We also looked at networks from an outside perspective, and evaluated the 'efficiency' of a network, which is defined by the aggregate 'welfare' (sum of each agent's utility). We saw that there are efficient networks that are stable – some agent in the network would like to change one or more of their links – and generally the reverse is also true, that there can be stable networks which are inefficient (do not maximize welfare).

Lastly, we looked at network formation through a theoretical computer science perspective and saw that for at least one way of defining the network formation problem, that forming an optimal network is NP-hard; we do not know of a quick algorithm to solve such a problem, and if we did, then we would be solving at least one other very hard algorithm problem (the 'minimum dominating set' problem).

This week, we will step away from the formation of networks and instead look at one common use for networks: spreading information. There are many situations in life where many people have local (or 'private') information about a particular thing, but they may want to connect with other people in order to pool their information, or simply verify that their information is correct.

What we will be examining is shared by both disciplines: repeated local updating on a graph. We will start with the model of averaging information from DeGroot (1974), then connect that to Golub and Jackson (2010) and their "wisdom of the crowds" result. We will then look across to TCS and see similar math in the distributed averaging and gossip algorithms of Boyd et al. (2006).

2 DeGroot Learning

We have the following set-up: there are n agents in a network. There is some number $\theta \in \mathbb{R}$ that every agent is trying to guess. Agent i has an estimate $x_i(t)$ at time t . All together, their estimates form the vector $\vec{x}(t) = [x_1(t) \dots x_n(t)]$.

For now, we will forgo a graph and instead think about an "influence matrix" T . T_{ij} will be the influence (or weight) that agent i places on the estimate from agent j . T will thus be a row-stochastic matrix.

Definition 1 (Row-stochastic matrix). *A matrix $T = (T_{ij})$ is row-stochastic if, for every row i , $T_{ij} \geq 0$ and $\sum_{j=1}^n T_{ij} = 1$*

Recall that stochastic is a fancy way of saying random, and the constraints we have mean that each row is a valid probability distribution. If we were writing T based on a graph, we would likely make $T_{ij} = 0$ if i and j were not neighbors.

The model of DeGroot (1974) is that each agent updates their own estimate by averaging the estimates she is hearing.

$$x(t+1) = Tx(t). \quad (1)$$

Equivalently,

$$x_i(t+1) = \sum_{j=1}^n T_{ij}x_j(t).$$

Using the properties of stochastic matrices, we know that:

$$x(t) = T^t x(0). \quad (2)$$

Thus the long-run outcome is determined by powers of the influence matrix.

2.1 When does consensus occur?

By consensus, I mean under what conditions for T and $x(0)$ is it the case that after an infinite amount of time, every agent reaches the same estimate? In order to answer this question, we will need to slightly reformulate our problem.

Suppose that you are dropped into this network at random, right next to some agent. At each time step t , the agent you are standing next to tells you the agents that he is listening to, and how much he believes them. With his beliefs as your probability distribution, you randomly choose an agent, and go to them at $t+1$. The path that you take is now a stochastic process, and T is now your transition matrix – the matrix that defines how you move between agents (states). Using this formulation, we require two qualities: T must be irreducible, and T must be aperiodic.

Definition 2 (Irreducibility). *A transition matrix T is irreducible if, for all agents i , if you start at i you have a positive probability of reaching every other agent within a finite number of moves. Mathematically, $\forall i, j, \exists t \geq 0 : T_{ij}^t > 0$.*

If we drew a directed graph between the agents (vertices) where there is an edge from i to j if $T_{ij} > 0$, then the above would give us strongly connected graph: a directed graph where there is at least one directed path between every pair of vertices.

Irreducibility guarantees that there is a ‘stationary distribution’ for this Markov chain. A stationary distribution means that, eventually, our journey around the agents will be entirely predictable, and there will be a single probability distribution that, regardless of what agent we are currently at, will accurately describe our chances of moving between agents.

However, we need something even stronger. Even with a stationary distribution, it may be the case that I simply go back and forth between two agents, who at time t believe the other completely, and thus take on their opinion. Thus at $t+1$, they have flipped opinions, and at $t+2$, they have flipped opinions again. (Ex: person A estimates $\theta = 2$, person B estimates $\theta = 5$. At $t = 0$, we have $(A, B) = (2, 5)$. At $t = 1$, they flip, and we have $(5, 2)$. At $t = 2$, they flip again...)

The other property we need is that T is aperiodic. What does this mean? Starting at agent i , there can be multiple possible return trips. We may be able to go from i to j to k back to i (3 steps), or from i to j to i (2 steps), or simply from i to i (1 step).

The period of an agent (state) in a Markov chain is the greatest common divisor of all of its return times. In the above example, the greatest common divisor is 1. The significance of aperiodicity is that each row of the transition matrix T will eventually be equal to each other. This will give us consensus, as we described it above.

Definition 3 (Aperiodic). *A transition matrix T is aperiodic if the period of every agent (state) i is 1. Mathematically, the period of state i is $d(i) := \gcd\{n \geq 1 : T_{ii}^n > 0\}$.*

An irreducible and aperiodic Markov chain is called ‘ergodic’, and is extremely nice to work with.

Theorem 1 (Stationary distribution of an ergodic chain). *Suppose T is the transition matrix of an irreducible and aperiodic Markov chain. Then there is a unique vector $s \in \mathbb{R}^n$ which satisfies:*

$$\forall i, s_i > 0 ; \sum_i s_i = 1 ; s \cdot T = s$$

and $T^t \rightarrow \mathbf{1} \cdot s$, i.e. T^t converges to a matrix where each row is s

This gives us consensus: we know that $x(t) = T^t \cdot x(0)$. The above tells us $x(t) \rightarrow \mathbf{1} \cdot s \cdot x(0)$. Thus each agent will reach the same opinion, the weighted average $\sum_{i=1}^n s_i x_i(0)$.

3 Example: a star

Suppose each agent chooses who to connect with (perhaps the result of some network formation rule), and then of the people that agent is connected with, she gives each of them equal weight. We can then find the stationary ‘influence’ of agent i as follows:

$$s_i = \frac{d_i}{\sum_j d_j}. \quad (3)$$

For an n -node star, the center has degree $n - 1$ and the $n - 1$ leaves each have degree 1. Since total degree is $2(n - 1)$,

$$s_{\text{center}} = \frac{1}{2}, \quad s_{\text{leaf}} = \frac{1}{2(n - 1)}.$$

One person therefore receives half of the total long-run weight, no matter how large the network becomes.

4 From consensus to learning

We continue with our unknown $\theta \in \mathbb{R}$. Each agent begins with a noisy signal

$$x_i(0) = \theta + \varepsilon_i, \quad (4)$$

where $\mathbb{E}[\varepsilon_i] = 0$, $\text{Var}(\varepsilon_i) = \sigma_i^2$, and the errors are independent across agents.

If everyone was connected with everyone else, and trusted everyone equally, then the law of large numbers would make the average estimate increasingly accurate as n grows. But in many cases we neither are connected to everyone nor believe everyone equally.

Using what we found above, we know that the limiting consensus is

$$\hat{\theta} = \sum_i s_i x_i(0) = \theta + \sum_i s_i \varepsilon_i.$$

This estimate is unbiased, but its variance is

$$\text{Var}(\hat{\theta}) = \sum_i s_i^2 \sigma_i^2. \quad (5)$$

If we assume none of the variances are unbounded, then there is some maximal $\bar{\sigma}^2$ such that

$$\text{Var}(\hat{\theta}) \leq \bar{\sigma}^2 \sum_i s_i^2.$$

This brings us to a main result of Golub and Jackson (2010): society will learn the truth when no small set of agents retains too much influence as the society grows.

Proposition 1 (Vanishing individual influence implies accurate aggregation). *Consider a sequence of societies ordered by their population, n . Suppose the initial errors are independent, mean zero, and have uniformly bounded variances. If*

$$\max_i s_i^{(n)} \longrightarrow 0,$$

then the DeGroot consensus converges in probability to the true state θ .

Proof. Because $\sum_i s_i = 1$,

$$\sum_i s_i^2 \leq \left(\max_i s_i \right) \sum_i s_i = \max_i s_i.$$

Therefore (5) implies

$$\text{Var}(\hat{\theta}) \leq \bar{\sigma}^2 \max_i s_i \rightarrow 0.$$

Since $\hat{\theta}$ is unbiased, Chebyshev's inequality gives, for every $\varepsilon > 0$,

$$\Pr(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{\text{Var}(\hat{\theta})}{\varepsilon^2} \rightarrow 0.$$

□

This short proof captures a deep economic idea. A large group is not automatically wise. It becomes wise only if its aggregation process prevents any one person's noise from receiving persistent weight.

4.1 Network structure and wisdom

Under our equal-neighbor updating rule, for an undirected network, influence is proportional to degree as in (3). Thus a simple sufficient condition for a growing society to learn the true parameter value is

$$\frac{\max_i d_i}{\sum_j d_j} \longrightarrow 0. \quad (6)$$

In plain language, this means that the most connected individual must account for a vanishing share of all connections. As we saw above, the star fails this test because the center's degree share is always $1/2$.

5 Convergence time of distributed averaging

So far, we have only been considered with whether a network ‘eventually’ learns the truth, given the conditions they start under. Now we will ask how quickly a network can learn the truth; how much communication is needed to reach a given amount of agreement, irregardless of accuracy? We will turn to the pairwise averaging model of Boyd et al. (2006) and derive an upper bound on convergence time.

5.1 The objective and the gossip rule

Suppose each of n nodes initially holds a number $x_i(0)$. The objective is for every node to individually approach the global average

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i(0).$$

This is a number determined by the initial data. Whether each node individually reaches this average is different from whether the network as a whole ‘learns’ the unknown state.

For comparison, a fixed DeGroot matrix T converges to this average if it is primitive and *doubly stochastic*: both its rows and its columns sum to one. The stationary distribution is then uniform.

In Boyd, we no longer have a static transition matrix. Instead, the ‘gossip rule’ will give us a dynamic transition matrix, i.e. it is (potentially) different at each time t .

Let $G = (N, E)$ be a connected undirected graph. At each time step, independently of earlier steps, we choose an unordered edge $\{i, j\}$ with probability q_{ij} . The selected pair replaces its values by their average:

$$x_i(t+1) = x_j(t+1) = \frac{x_i(t) + x_j(t)}{2}.$$

All other nodes keep their current values. We will use the simple case of uniform probability of selecting any edge, $q_{ij} = 1/|E|$.

Here, one step means *one pairwise communication event*, not a round in which all nodes update simultaneously. In the paper’s asynchronous clock model, each of n nodes has a rate-one clock, so there are on average n such events per unit of elapsed time. We will count events throughout.

Note that updating does not change the goal $\bar{x}$. Each pairwise update preserves the total sum between those nodes:

$$\frac{x_i + x_j}{2} + \frac{x_i + x_j}{2} = x_i + x_j.$$

Thus the total sum, and hence the average, remains unchanged. Thus if the network reaches consensus, we know it will reach $\bar{x}$. We just want to know how quickly that will happen.

5.2 How much does one interaction help?

Define the disagreement vector and its squared length:

$$y(t) = x(t) - \bar{x}\mathbf{1}, \quad V(t) = \|y(t)\|_2^2 = \sum_{k=1}^n (x_k(t) - \bar{x})^2.$$

Because $\bar{x}$ is the average, $\sum_k y_k(t) = 0$.

If $\{i, j\}$ is selected, only y_i and y_j change. They both become $(y_i + y_j)/2$. Therefore

$$\begin{aligned} V(t) - V(t+1) &= y_i^2 + y_j^2 - 2 \left(\frac{y_i + y_j}{2} \right)^2 \\ &= \frac{1}{2} (y_i - y_j)^2 \\ &= \frac{1}{2} (x_i(t) - x_j(t))^2. \end{aligned}$$

Every interaction weakly reduces disagreement. An interaction between nodes with very different estimates makes substantial progress; one between nodes that already agree makes no progress.

Connectivity matters because V can stop decreasing across *all* available edges only when every neighboring pair agrees. On a connected graph, that means everyone agrees. Independent repeated selection with positive edge probabilities allows those useful interactions to occur.

5.3 The matrix that determines convergence speed

We can represent the operation of averaging two nodes with a matrix. Let e_i be the column vector with a one in position i and zeros elsewhere. The matrix for averaging nodes i and j is

$$W^{ij} = I - \frac{1}{2}(e_i - e_j)(e_i - e_j)^T.$$

This matrix replaces entries i, j by their pairwise average, and leaves the others unchanged. Each such matrix W^{ij} is symmetric, doubly stochastic, and idempotent (applying it twice has the same effect as applying it once):

$$(W^{ij})^T = W^{ij}, \quad (W^{ij})^2 = W^{ij}.$$

Write $W(t)$ for the randomly selected update matrix. Then

$$x(t+1) = W(t)x(t), \quad y(t+1) = W(t)y(t).$$

We can now define the *expected update matrix*

$$\bar{W} = \mathbb{E}[W(t)] = \sum_{\{i,j\} \in E} q_{ij} W^{ij}.$$

The probabilities q_{ij} are fixed, so $\bar{W}$ does not depend on t .

Combining

$$V(t+1) = \|y(t+1)\|_2^2 = y(t+1)^T y(t+1), \text{ and } y(t+1) = W(t)y(t)$$

Then conditional on the current disagreement,

$$\mathbb{E}[V(t+1) \mid y(t)] = y(t)^T \bar{W} y(t).$$

Thus the expected update matrix controls the level of (squared) disagreement.

Now to proceed, we will need to talk about the eigenvalues of $\bar{W}$. First, let us understand those of each W^{ij} . For any $x(t)$, we can decompose it in the following way:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \dots \end{pmatrix} = \begin{pmatrix} \frac{x_1+x_2}{2} \\ \frac{x_1+x_2}{2} \\ 0 \\ \dots \end{pmatrix} + \begin{pmatrix} \frac{x_1-x_2}{2} \\ \frac{x_2-x_1}{2} \\ 0 \\ \dots \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ x_3 \\ \dots \end{pmatrix}.$$

What happens when we do $W^{ij}x(t)$ is we get rid of the second column, and leave all others untouched. This implies that the eigenvalues of W^{ij} are $1, 0, 1, 1, \dots$. The matrix $\bar{W}$ is the expectation of W^{ij} s, so it's eigenvalues must be between 0 and 1. We can thus order them as:

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n \geq 0.$$

Since every pairwise W^{ij} does nothing to a vector which is already in full agreement ($c\mathbf{1}$), we know that the expected $\bar{W}$ will also be such that: $\bar{W}\mathbf{1} = \mathbf{1}$, implying that one eigenvalue (the largest) is 1, with a unit eigenvector of $\mathbf{1}/\sqrt{n}$. The disagreement vector has no component of agreement ($y(t) = x(t) - \bar{x}\mathbf{1}$), so $y(t)^T\mathbf{1} = 0$. Thus, we can write

$$y(t) = \sum_{k=2}^n a_k v_k$$

a sum of the remaining orthonormal eigenvectors. Then

$$y(t)^T \bar{W} y(t) = y(t)^T \sum_{k=2}^n \lambda_k a_k v_k = \sum_{k=2}^n \sum_{l=2}^n a_k \lambda_l a_l (v_k^T v_l) = \sum_{k=2}^n \lambda_k a_k^2 \leq \lambda_{(2)} \sum_{k=2}^n a_k^2 = \lambda_{(2)} V(t).$$

Proposition 2 (Expected disagreement contracts over time). *Under the independent pairwise gossip rule (the above situation),*

$$\mathbb{E}[V(t)] \leq \lambda_{(2)}^t V(0).$$

Proof. We already showed that $\mathbb{E}[V(t+1) \mid y(t)] \leq \lambda_{(2)} V(t)$. You can continually take the expectation to get the result. $\square$

Thus $\lambda_{(2)}$ measures the slowest direction of disagreement. If it is close to one, some patterns of disagreement decay slowly. We can then define the *spectral gap*, $\gamma = 1 - \lambda_{(2)}$; a larger gap gives a stronger contraction bound per communication event.

5.4 Turning contraction into a convergence-time bound

Fix an accuracy level $0 < \varepsilon < 1$ and an allowed failure probability $0 < \delta < 1$. We want

$$\Pr(\|y(t)\|_2 \geq \varepsilon \|y(0)\|_2) \leq \delta.$$

This means the Euclidean disagreement has shrunk to an ε fraction of its initial size, with probability at least $1 - \delta$. If $y(0) = 0$, the system already agrees.

Markov's inequality says that for a nonnegative random variable Z and $a > 0$,

$$\Pr(Z \geq a) \leq \frac{\mathbb{E}[Z]}{a}.$$

Apply it to $Z = V(t)$:

$$\begin{aligned} \Pr(V(t) \geq \varepsilon^2 V(0)) &\leq \frac{\mathbb{E}[V(t)]}{\varepsilon^2 V(0)} \leq \frac{\lambda_{(2)}^t}{\varepsilon^2} \leq \delta \\ \lambda_{(2)}^t &\leq \delta \varepsilon^2 \rightarrow t \ln \lambda_{(2)} \leq \ln \delta + 2 \ln \varepsilon \\ t &\geq \frac{\ln \delta + 2 \ln \varepsilon}{\ln \lambda_{(2)}} \rightarrow t \geq \frac{\ln(1/\delta) + 2 \ln(1/\varepsilon)}{\ln(1/\lambda_{(2)})} \end{aligned}$$

Using $-\log \lambda_{(2)} \geq 1 - \lambda_{(2)} = \gamma$, we can rewrite the above using our spectral gap:

$$t \geq \frac{\log(1/\delta) + 2 \log(1/\varepsilon)}{\gamma}$$

5.5 How the graph affects speed

To compare two graphs, we need to compare their spectral gaps. A larger gap gives a smaller sufficient number of events.

We will compare the complete graph and the cycle, each with $n \geq 4$ nodes and uniform selection of edges.

Complete graph. In the complete graph, every pair of nodes can communicate. There are $\binom{n}{2}$ edges, so each pair is selected with probability $1/\binom{n}{2}$.

For a particular node i , the probability of being selected is $2/n$. When selected, it keeps half of its own value. Otherwise, it keeps its entire value. Therefore,

$$\overline{W}_{ii} = \left(1 - \frac{2}{n}\right) + \frac{2}{n} \cdot \frac{1}{2} = 1 - \frac{1}{n}.$$

For $j \neq i$, node i receives half of node j 's value only when their edge is selected, so

$$\overline{W}_{ij} = \frac{1}{\binom{n}{2}} \cdot \frac{1}{2} = \frac{1}{n(n-1)}.$$

To see how this matrix acts on disagreement, recall that $\sum_j y_j = 0$. Thus $\sum_{j \neq i} y_j = -y_i$, giving

$$(\overline{W}y)_i = \left(1 - \frac{1}{n}\right) y_i + \frac{1}{n(n-1)} \sum_{j \neq i} y_j = \left(1 - \frac{1}{n-1}\right) y_i.$$

Hence every disagreement direction has the same eigenvalue:

$$\lambda_2 = \dots = \lambda_n = 1 - \frac{1}{n-1}, \quad \gamma = \frac{1}{n-1}.$$

Substituting into our convergence bound, it is sufficient to take

$$t \geq (n-1) [\log(1/\delta) + 2 \log(1/\varepsilon)].$$

For fixed accuracy and failure probability, this sufficient number of events grows linearly with n .

Cycle. Now arrange the nodes in a cycle. Each node can communicate only with its two neighbors, and each of the n edges is selected with probability $1/n$.

With probability $1 - \frac{2}{n}$, neither of a node's neighbors is chosen by the update matrix and so the node remains unchanged. The node has a $\frac{1}{n}$ chance of being averaged by either neighbor. Hence,

$$(\overline{W}y)_i = \left(1 - \frac{2}{n}\right) y_i + \frac{1}{n} \cdot \frac{1}{2} y_{i-1} + \frac{1}{n} \cdot \frac{1}{2} y_{i+1} = \left(1 - \frac{1}{n}\right) y_i + \frac{1}{2n} y_{i-1} + \frac{1}{2n} y_{i+1},$$

The second-largest eigenvalue of this matrix is

$$\lambda_2 = 1 - \frac{1 - \cos(2\pi/n)}{n} \Rightarrow \gamma = \frac{1 - \cos(2\pi/n)}{n} \approx \frac{2\pi^2}{n^3}.$$

To see where the cosine comes from, consider the vector

$$v_i = \cos\left(\frac{2\pi(i-1)}{n}\right).$$

This describes one full wave around the cycle. The identity $\cos(a - h) + \cos(a + h) = 2 \cos(h) \cos(a)$ gives

$$v_{i-1} + v_{i+1} = 2 \cos(2\pi/n) v_i.$$

Substituting into the expected update rule,

$$(\overline{W}v)_i = \left(1 - \frac{1}{n}\right) v_i + \frac{1}{2n}(v_{i-1} + v_{i+1}) = \left(1 - \frac{1 - \cos(2\pi/n)}{n}\right) v_i.$$

Thus v is an eigenvector. More generally, the cycle's eigenvalues are $1 - [1 - \cos(2\pi k/n)]/n$ for $k = 0, \dots, n-1$. The largest is 1 at $k = 0$; the second-largest occurs at $k = 1$ and $k = n-1$, giving

$$\lambda_2 = 1 - \frac{1 - \cos(2\pi/n)}{n}.$$

Our bound therefore gives the sufficient condition

$$t \geq \frac{n}{1 - \cos(2\pi/n)} [\log(1/\delta) + 2 \log(1/\varepsilon)].$$

For large n , the right-hand side is approximately

$$\frac{n^3}{2\pi^2} [\log(1/\delta) + 2 \log(1/\varepsilon)].$$

For fixed accuracy and failure probability, this sufficient number of events grows cubically with n .

Why does the cycle have a weaker bound? Recall that an interaction across edge $\{i, j\}$ reduces squared disagreement by

$$\frac{1}{2}(y_i - y_j)^2.$$

On a cycle, estimates can vary gradually around the network: high in one region and low in another, with only small differences between neighbors. Every available interaction then makes little progress, even though overall disagreement is substantial.

On the complete graph, nodes in the high and low regions can communicate directly. There is no disagreement direction that is protected by having only similar-valued neighbors.

Both graphs converge to the same global average. However, our sufficient communication bound grows like n for the complete graph and like n^3 for the cycle, holding ε and δ fixed.

References

- Boyd, Stephen et al. (2006). “Randomized Gossip Algorithms”. In: *IEEE Transactions on Information Theory* 52.6, pp. 2508–2530. DOI: [10.1109/TIT.2006.874516](https://doi.org/10.1109/TIT.2006.874516).
- DeGroot, Morris H. (1974). “Reaching a Consensus”. In: *Journal of the American Statistical Association* 69.345, pp. 118–121. DOI: [10.1080/01621459.1974.10480137](https://doi.org/10.1080/01621459.1974.10480137).
- Golub, Benjamin and Matthew O. Jackson (2010). “Naïve Learning in Social Networks and the Wisdom of Crowds”. In: *American Economic Journal: Microeconomics* 2.1, pp. 112–149. DOI: [10.1257/mic.2.1.112](https://doi.org/10.1257/mic.2.1.112).