**MS&E 336/CS 366: Computational Social Choice. Aut 2021-22**
Course URL: `http://www.stanford.edu/~ashishg/msande336/index.html`.
Instructor: Ashish Goel, Stanford University.

**Lecture 14, 11/03/2021. Scribed by Zhihao Jiang.**

# 1 Polarization in Opinion Dynamics

This lecture is mainly based on [1].

An opinion dynamics process is said to be *polarizing* if the variance of the opinions increases with time.

*Homophily* refers to there is greater interaction between like-minded individuals. Empirical studies have shown that homophily results in polarization. A popular belief is the folowing:

$$\text{Homophily} \rightarrow \text{Social corroboration} \rightarrow \text{polarization}$$

However, while Homophily is clearly captured by DeGroot ("my neighbours influence me") and Hegselmann-Krause ("people with opinions close within a radius $r$ influence me"), which we learned in previous lectures, they are in fact depolarizing. What can possibly explain the polarization that we currently see?

To explain polarization, at least the way we have defined it, Homophily should come with additional assumptions, such as Biased assimilation. Biased assimilation is a behaviour empirically demonstrated by psychologists [2], where people easily accept ideas that align with their beliefs, while they easily disbelieve ideas that are in contradiction with their opinions.

## 1.1 Urn dynamics

We use urn dynamics as a simplified, stylized model of how other people's opinion impacts a voter. Consider two voters, $i$ and $j$. Each voter has their own urn, containing some mixture of red and blue balls. Denote the fraction of red balls for voter $i$ at time $t$ to be $x_i(t)$. This fraction represents voter $i$'s preference for red.

Each time, voter $i$ draws 2 balls, one ball from voter $j$'s urn and one ball from their own urn. If the color of the two drawn balls matches, voter $i$ puts another ball of that color into their urn, and discards a random preexisting ball to keep the total number constant. If the color of the two drawn balls does not match, the drawn balls are returned to their respective urns and no further changes are made. This is an example of biased assimilation, only letting another person's opinion influence your own opinion if it matches your preexisting beliefs.

We assume there are $n$ balls in each urn, and $i$ only has one neighbor $j$. At time $t$, the probability that $i$ introduce a new red ball is $x_i(t)x_j(t)$, and the probability that $i$ introduce a new blue ball is $(1 - x_i(t)) \cdot (1 - x_j(t))$. When $i$ introduce a new ball, one preexisting ball is removed, where a red ball is removed with probability $x_i(t)$. As a result, we have

$$\begin{aligned}
x_i(t+1) &= x_i(t) + \frac{1}{n}\cdot Pr(\text{a red ball is introduced}) - \frac{1}{n}\cdot Pr(\text{a red ball is removed}) \\
&= x_i(t) + \frac{1}{n}\cdot Pr(\text{a red ball is introduced}) - \frac{1}{n}\cdot x_i(t)\cdot Pr(\text{a random ball is removed}) \\
&= x_i(t) + \frac{1}{n}\cdot x_i(t)x_j(t) - \frac{1}{n}\cdot x_i(t)\cdot (x_i(t)x_j(t) + (1-x_i(t))(1-x_j(t))) \qquad (1)
\end{aligned}$$

## 1.2 Formal definition of Attitude Polarization

We begin by defining a notion of other people's opinions. $S_i(t)$ is the weighted average of the opinions of $i$'s neighbors, or more formally,

$$S_i(t) = \frac{\sum_{j\neq i} w_{ij}x_j(t)}{\sum_{j\neq i} w_{ij}}$$

where we only have the assumption

$$w_{ii} > 0$$

With these definitions, we can define the notion of biased assimilation.

$$x_i(t+1) = \frac{w_{ii}x_i(t) + (x_i(t))^b S_i(t)}{w_{ii} + (x_i(t))^b S_i(t) + (1-x_i(t))^b(1-S_i(t))} \qquad (2)$$

where $b$ is the bias factor, which allows additional control of how strongly the voter weights other's opinions relative to their own.

If $b = 0$, we get DeGroot.

If $b = 1$, we get urn dynamics, because we can replace $x_j(t)$ by $S_j(t)$ in (1), and let

$$w_{ii} = n - (x_i(t)S_j(t) + (1-x_i(t))(1-S_j(t)))$$

in (2).

## 1.3 Two island network

Consider a graph with two very well connected components $V_1, V_2$. Edges within these components have the weight $p_{same}$. These two components are sparsely connected by a smaller number of edge, each with weight $p_{diff}$. Assume $x_i(0) = x_0, \forall i \in V_1$, and $x_i(0) = 1 - x_0, \forall i \in V_2$, where $1/2 < x_0 < 1$.

Let $h = p_{same}/p_{diff}$, and we have

**Theorem 1.1 ([1])** *If $|V_1| = |V_2|$, then*

1. *(Polarization) If $b \geq 1$, $\forall i \in V_1$, $\lim_{t\to\infty} x_i(t) = 1$, and $\forall i \in V_2$, $\lim_{t\to\infty} x_i(t) = 0$.*
2. *(Persistent disagreement) If $1 > b \geq \frac{2}{h+1}$, there exists a unique $\hat{x} \in (\frac{1}{2}, 1)$ such that $\forall i \in V_1$, $\lim_{t\to\infty} x_i(t) = \hat{x}$, and $\forall i \in V_2$, $\lim_{t\to\infty} x_i(t) = 1 - \hat{x}$.*
3. *(Consensus) If $b < \frac{2}{h+1}$, then for all $i \in V_1 \cup V_2$, $\lim_{t\to\infty} x_i(t) = \frac{1}{2}$.*

One corollary is that the urn dynamic leads to polarization in two island network if the two components are equally large.

### 1.4 Connection to recommender systems

There are two kinds of recommender systems

1. Best result: Look at all the books I have bought, or all the people I follow, and recommend to me the one that is the most similar to this set (e.g. the Twitter example).
2. Most similar to one item: Look at a random book I bought, or a random person I followed, and recommend one that is most similar to this one person/book.

The first one polarizes our reading choices, like biased assimilation. The second one is similar to persistent disagreement — does not polarize.

### 1.5 Schelling segregation

Schelling segregation is a surprising result where even mild homophily leads to segregation. We begin with a grid, where agents of 2 different types reside on the grid points. The initial placement of the agents is random, and some grid points may be unoccupied. Of 4 possible neighbors (N,S,E,W), each agent prefers to have at least two neighbors be the same type as itself. If this condition is not satisfied, the agent will move, switching to an unoccupied spot selected uniformly at random from spots that do satisfy this property, if such a point exists. This process is iterated over many time steps, and over time, the two types become quite separated. A simulation can be found at:

`http://nifty.stanford.edu/2014/mccown-schelling-model-segregation/`

The surprising aspect of this model is that the realized segregation is much stronger than the preferences that created it. Even though agents only want at least half their neighbors to be like them, this drives behavior that creates a much higher percentage of neighbors to be the same type. As one might presume, if the preferences are increased such that agents want at least 3 neighbors to be the same, the resulting segregation is even more pronounced.

This very simple model shows how residential segregation might arise in spite of only mild preferences of the residents. Despite the simplicity of the model, which neglects heterogeneity amongst houses, preferences, etc., it describes a broader emergent pattern that is seen in practice. Schelling shared the Nobel Prize in economics in 2005, for this and other similar work.

## References


[1] Pranav Dandekar, Ashish Goel, and David T. Lee. *Biased assimilation, homophily, and the dynamics of polarization.* PNAS, 2013.

[2] Charles G. Lord, Lee Ross, and Mark R. Lepper. *Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.* Journal of Personality and Social Psychology, 1979.