MS&E 336/CS 366: Computational Social Choice. Winter 2019-20
Course URL: `http://www.stanford.edu/~ashishg/msande336.html`.
Instructor: Ashish Goel, Stanford University.

Lecture 5, 1/27/2020. Scribed by Arnaud Autef.

6 Approval voting

Each voter provides a set of approved candidates, with no ranking. Candidates with the most number of approvals wins (ties broken arbitrarily). Is this a good rule? It depends on assumptions about the preferences of voters (their utility function):

- **Assumption 1:** Voters have *dichotomous* preferences, *i.e* for each candidate, they are "ok" with this person being selected or "not ok". For instance, some voters of the Democratic primary might have the following dichotomous preferences: equally favorable to any candidate having fair chance to win against Trump, equally unfavorable to any candidate unlikely to win against Trump.
- **Assumption 2:** Voters truly have rankings as preferences, but are forced by the "Approval voting" procedure to collapse them into "ok" and "not ok" candidates. In the previous example: a Democratic voter may have strongly ordered preferences, with say Elizabeth Warren > Bernie Sanders > Joe Biden > Andrew Yang, etc...

6.1 Approval voting with dichotomous preferences

If we assume that voters have *dichotomous* preferences, the following properties hold for Approval voting.

- Approval voting is a scoring rule that assigns 1 to preferred candidates and 0 to others.
- Approval voting is equivalent to finding the Condorcet winner. Thus, the Borda and the Copeland rule become *united* under this specific dichotomous assumption for voters' preferences. This observation is significant: without the dichotomous preferences assumption, Borda voting is not Condorcet consistent and hence fundamentally different from the Copeland method.
- *Strategy proofness*: with the additional assumption that voters neither approve of all candidates or disapprove of all candidates, Approval voting is strategy-proof. On Figure 1 we picture the preferences of a voter and notice that it is not in their interest to change their reporting.

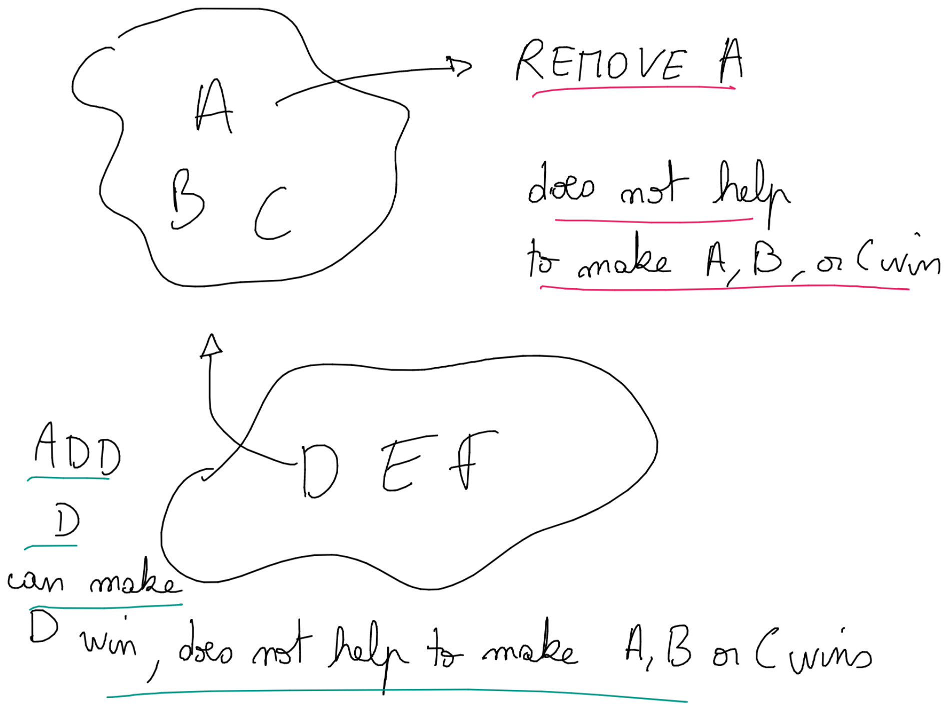


Figure 1: Illustration of strategy-proofness for a voter with Dichotomous preferences. The voter pictured has true preferences: approving of A, B, C and disapproving of D, E, F. If voter removes A when reporting their preferences, it cannot help to make A, B or C win; furthermore, it can hurt if it makes D, E or F win. If voter adds D when reporting their preferences, it cannot help to make A, B or C win, and can hurt if then D wins.

7 Opinion Dynamics

Opinion Dynamics (OD) models are concerned with modelling changes in the opinion of an agent over time t. An example use case in computational social choice would be too:

1. Compare different OD models an compare them with respect to how well they fit actual data.
2. In the context of the OD model that best fits the data, analyze the properties of a voting protocol we are interested in.

Formally, in the context of opinion dynamics, we consider:

- N *agents*
- Each agent has an *opinion*, assumed to lie on some continuous space (say in the interval $[0, 1]$).
- Opinions evolve over time t, which is assumed to be discrete.
- $x_i(t)$ is the opinion of agent i at time t.
- $\mathbf{x}(t)$ denotes the vector of all agent opinions at time t:

$$\mathbf{x}(t) = (x_1(t),\ ...,\ x_N(t))$$

- Evolution of opinion dynamics are governed by a *model* f that maps $\mathbf{x}(t)$ (opinions at time-step t) to $\mathbf{x}(t+1)$ (opinions at the next time-step $t+1$):

$$\mathbf{x}(t+1) = f(\mathbf{x}(t))$$

7.1 DeGroot Dynamics model

The DeGroot model is a very influential opinion dynamics model. It assumes that the opinion of agent i at time step $t+1$ is a weighted average of the opinions of its *neighbors* at time step t, in addition to its own opinion at time step t.

Formally, a $N \times N$ adjacency matrix W represents the DeGroot model. For any $1 \leq i,\ j \leq N$, $W_{i,j}$ can be interpreted as how much weight agent i gives to j's opinion. W has the following properties:

1. $\forall 1 \leq i \leq N,\ W_{i,i} > 0$. Agent i is always influenced by itself.
2. $\forall i, j,\ W_{i,j} \geq 0$. Weights are non negative.
3. The W matrix is *row stochastic*: $\forall i,\ \sum_j W_{i,j} = 1$. The weights given by agent i to its neighbors' opinions sum to 1.

Opinions at time t are then derived from opinions at time $t-1$ using:

$$\mathbf{x}(t) = W\mathbf{x}(t-1)$$

W **represents a graph between agents**.

- Vertices are the N agents.
- If $W_{i,j} > 0$, opinion of agent j influences opinion of agent i. Thus, we draw an edge from j to i in the graph, $j \to i$.

Theorem: Convergence. If the graph represented by W is a single strongly connected component, opinions converge to a single common opinion z as $t \to +\infty$.

$$\exists z \in [0,1],\ \mathbf{x}(t) \to_{t \to +\infty} (z,\ ...,\ z)$$

Figure 2 illustrates this with a simple example for W where opinions converge.

7.2 Hegselmann-Krause Dynamics model

The Hegselmann-Krause model is also fundamental in opinion dynamics. Given some radius $r > 0$, opinions of agents are influenced by agents with opinions within a radius r of theirs. Formally, opinion $x_i(t+1)$ of user i at time step $t+1$ is the average of opinions $x_j(t)$ of users j at time step t that verify $|x_j(t) - x_i(t)| \leq r$ (*i.e* users j whose opinion at time step t are close to agent i's opinion at time step t by at most r). In particular, agent i is always influenced by its past opinion and the Hegselmann-Krause model can be summarized by the following equation:

$$\forall i,\ x_i(t+1) = \frac{1}{|B(i,t,r)|} \sum_{j \in B(i,t,r)} x_j(t)$$

Where $B(i,t,r) := \{j \text{ such that} |x_j(t) - x_i(t)| \leq r\}$.

Theorem: Convergence. The Hegselmann-Kraus OD model converges in a *finite* number of step, and at most $O(N^3)$.

Opinions need not to converge to a single opinion z. There exists a set Z of $m \leq N$ distinct final opinion. Such that every agent i's opinion converges to a final opinion $z_i \in Z$ as t goes to infinity. Formally:

$$\forall i,\ x_i(t) \to_{t \to +\infty} z_i \in Z$$

With $\forall z \neq z' \in Z,\ |z - z'| > r$, *i.e* final opinions are at least r apart. Indeed, by definition, opinions do not change anymore at convergence.

Lower bound: It is possible to find an example setting where at least $\Omega(N^2)$ steps are necessary to converge.

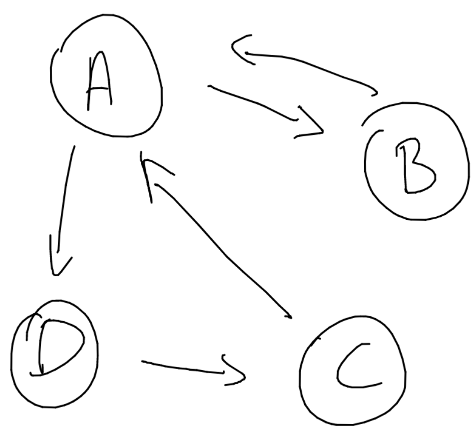
Closing the gap between the lower bound found and the convergence theorem is a challenging *open problem.*

Remark: The DeGroot and the Hegselmann-Krause models are "TV" models that are easy to describe and motivate but more complex models may be required to accurately model actual data.

W matrix

$$
\begin{array}{c|c}
 & \begin{array}{cccc} A & B & C & D \end{array} \\
\begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix}
\end{array}
$$

Underlying graph :



$$
\underset{X(0)}{\begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}} \xrightarrow[X(t)]{t \to +\infty} \underset{X(\infty)}{\begin{bmatrix} 0.8 \\ 0.8 \\ 0.8 \\ 0.8 \end{bmatrix}}
$$

Figure 2: Example for W

7.3 Polarization

Along the lines of Dandekar et al [1], an opinion dynamics process is said to be **polarizing** if the *variance* of the opinions increases with time. The variance is defined formally as follows:

$$\mathbb{V}\text{ar}(t) := \sum_{i=1}^{N} x_i^2(t) - \left(\sum_{i=1}^{N} x_i(t)\right)^2$$

Homophily refers to the tendency for people to have (non-negative) ties with people who are similar to themselves in socially significant ways. A common meme, specially in the popular media, is then 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”), they are in fact *depolarizing*. For instance, in the DeGroot OD model, if we take a simple example where W is *diagonal by block*, it becomes a doubly stochastic matrix. Then one can show easily that the variance decreases at every time step, because it is a convex function of the x_i's.

How to explain this contradiction with our common belief?

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 (see Lord et al [2] for a reference), where people easily accept ideas that align with their beliefs, while they easily disbelieve ideas that are in contradiction with their opinions.

References

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

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