Technical University of Crete
School of Electrical and Computer Engineering
MSc Program: **MLDS**
Course: **Reinforcement Learning**
Instructor: T. Spyropoulos
Student **Chivintar Amenty**

Reducing Opinion Polarization over a Social Network
Phase I
Report Delivery Date: May 26, 2025

This report is the first part of an examination of simple influence dynamics over a social network. First, a random graph and an echo-chamber graph are implemented. Each node has an opinion attribute assigned. Opinions are bound in $[-1, 1]$ with a step of 0.5. Then, a recommendation-driven influence model is applied to each graph. The analysis is followed by Value Iteration algorithm with the goal to minimize the total polarization as fast as possible. The Python package `networkx` was used.

TASK 1. The Environment

The environment is designed in a very simple MDP setting. The states are represented by tuples of nodes' opinions $(S_0, S_1, ..., S_N)$. Actions are represented in a list structure where each element is a tuple of (directed) edges (j, i) "node i is influenced by node j". Hence, the valid actions are all bidirectional edges of the graph (i.e. $[(0, 1), (1, 0), ...]$. Total polarity S_{tot} of the graph is measured as $\sum_{n=0}^{N} |S_n|$. The rewards are chosen to be $-S_{tot}$, plus some additional cases to encourage improvement (check `.ipynb` file).
As terminal states, those for which all nodes have the same opinion value are chosen. Naturally in those states there is no way for change to happen. For the case that the opinion is 0, a positive terminating reward is given, whereas all other terminal states yield negative terminal reward.
For this task, the main function are `create_random_graph()` and `create_echo_chambers()`. Some example graphs are shown in Fig. 1.

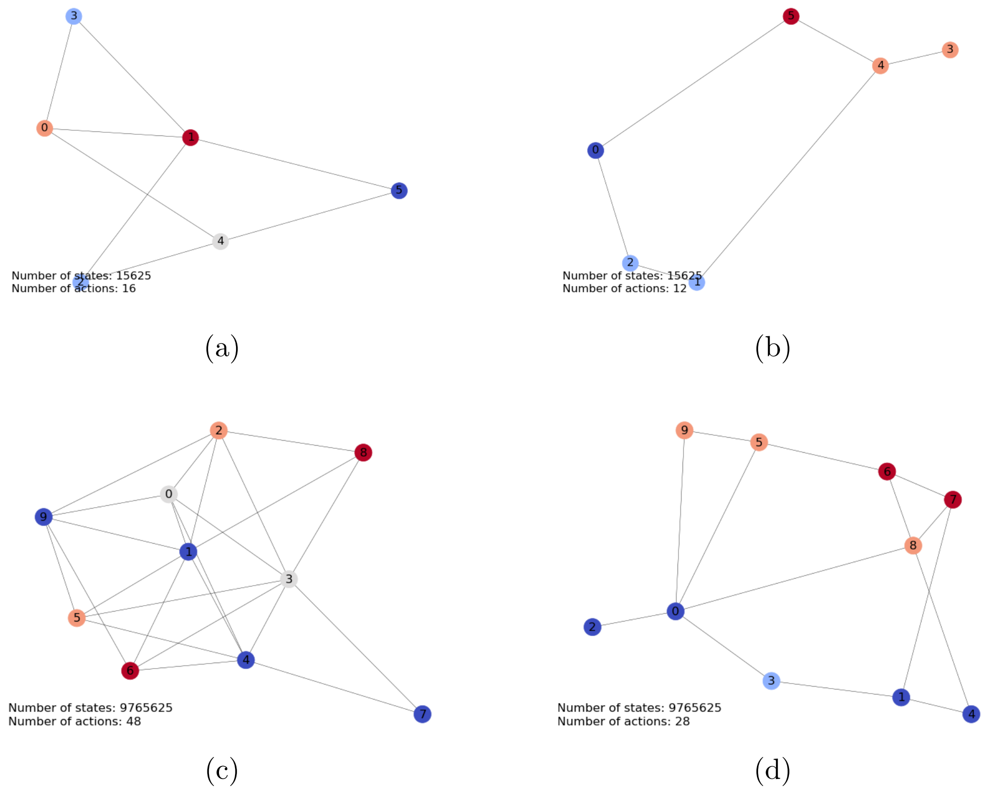


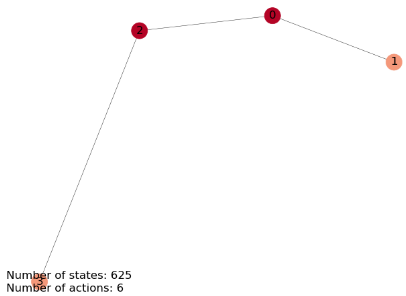
Figure 1: (a), (c) random nets, (b), (d) echo-chamber random nets with six and ten nodes respectively.

TASK 2. Baseline Implementation

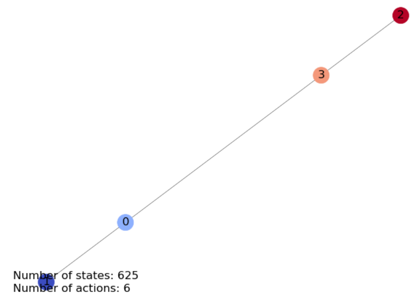
In Fig. 2 two toy networks are used: (a) corresponds to a random one and (b) corresponds to an echo chamber. A random policy is shown below each, with respect to the total polarization over time. It is clear that the behaviour is random, frequently arriving to a terminal state, and for this small number of nodes, in less than 10 steps most of the time. Subfigures (e) and (f) show a biased policy where node i selects node j with the closest agreement. It follows that in echo chamber graphs with high initial agreement within groups and fewer edges it is possible that one group converts the other or remains polarized. Of course, the environment is stochastic and this shall be considered into account.

TASK 3. Value Iteration

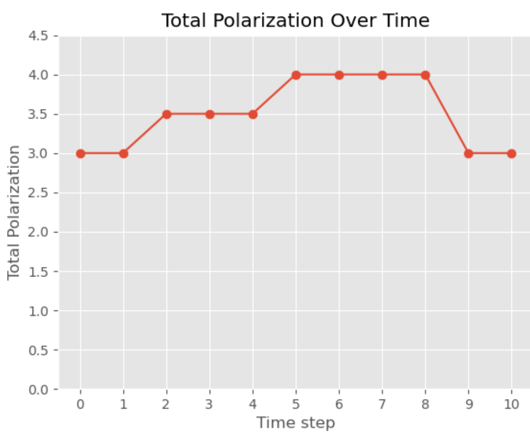
In Fig. 3 we see the performance of the optimal policy as calculated with the Value Iteration algorithm for our toy networks. Optimal value vector $V^*(s)$ for all states s (and



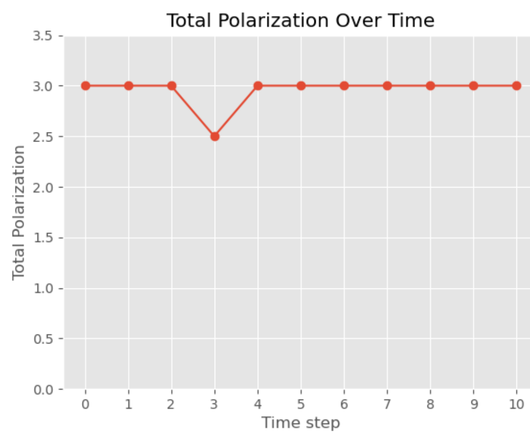
(a) graph GA



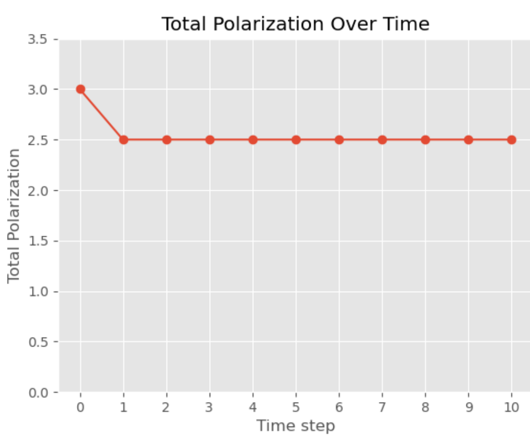
(b) graph GB



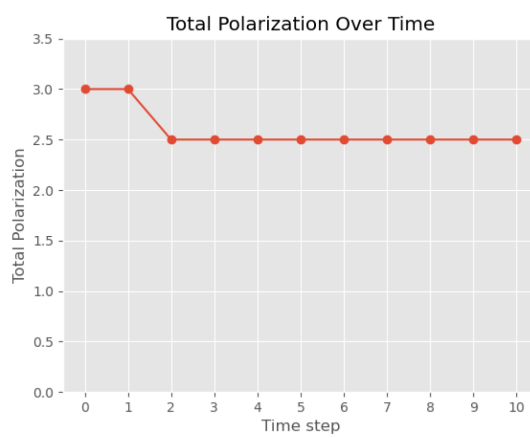
(c) random policy for GA



(d) random policy for GB

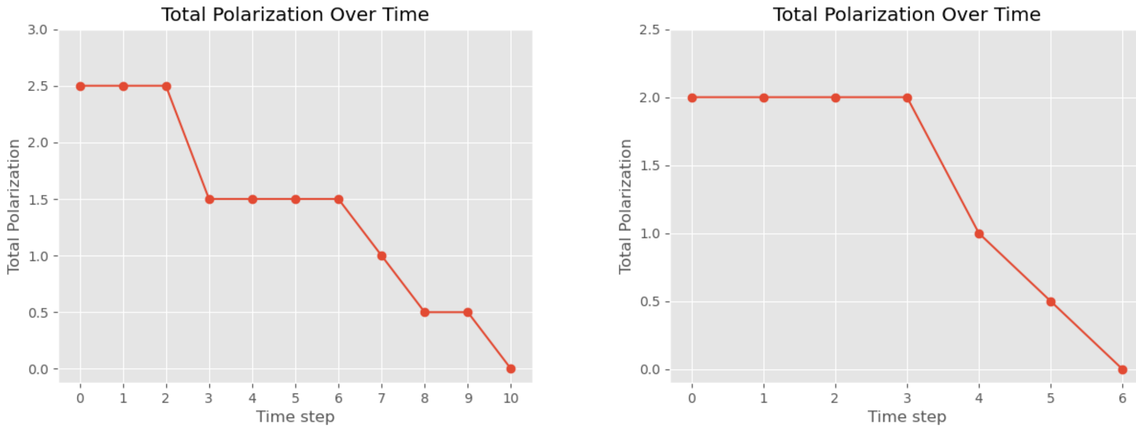


(e) biased policy for GA



(f) biased policy for GB

Figure 2: For graph (a), performance of random policy (c) and of biased policy (V.I.) (e). Similar plots for graph (b) below it.



(a) V^* for GA

(b) V^* for GB

Figure 3: Performance with optimal policy found with the Value Iteration algorithm. (a): corresponds to GA, (b): corresponds to GB (from Fig. 2) both converged in 80 iterations.

for graph (a), Fig.2) has the form:
`array([-24. , -29.66666667, 8. , 3.91666667, -1. , -29.66666667, -39.83333333, ... ])`
Computational time for 4-node nets takes ≈1 second, for 5-node nets takes ≈10 seconds, while it increases exponentially for bigger networks.

For validation of the algorithm, the optimal policy for the simplest 3-node graph is derived. With initial state $s^{(0)} = (-0.5, 0, 0.5)$, and nodes connected linearly: 0–1–2. The initial total polarization is $S_{tot}^{(0)} = 1$. In Fig. 4 the by-hand calculations are shown.They match exactly the actions chosen by the agent (check second to last cell in `.ipynb`), and convergence occurs in 2–3 steps. Final total polarization is 0, and terminal reward is the highest.

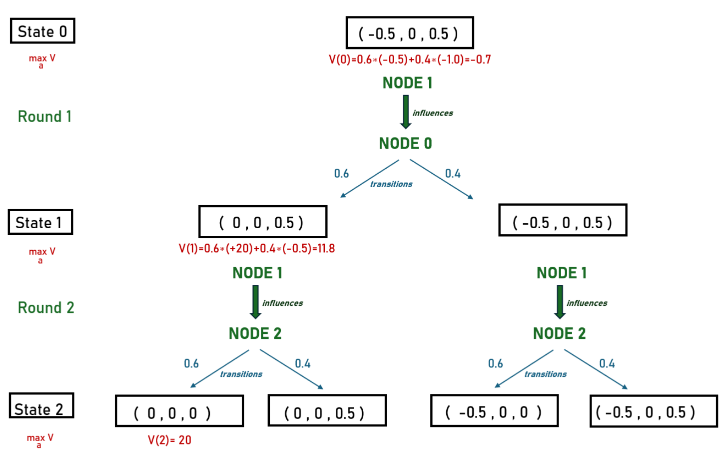


Figure 4: Validation with `gamma = 0`