Spread of Extremist Ideas in Social Networks with Censorship

Dogukan Yucel, Taggart Bonham

Department of Mathematics, Dartmouth College
MATH 76: The Mathematics of Misinformation
Advisor: Dr. Feng Fu

The spread of extremist ideas is a significant problem facing contemporary societies. This paper investigates how these ideas spread and turn into extremist movements with respect to various population parameters. It utilizes graph theory and stochastic games to evaluate the efficacy of censorship, punishment and social welfare in reducing the number of extremists. It was found that for each society there is a relationship between welfare and censorship, such that extreme ideas are reduced.

Introduction: Extremist movements such as terrorism, neo-Nazism or any kind of political extremism have been influential in social dynamics and have led to large-scale social consequences in recent history. A mathematical modeling of the spread of the ideas would provide us with a better framework to understand the spread of the ideas, and thus will enable us to generate effective ways to deal with extremists and reduce our communities' vulnerability to extremist movements.

In this paper, we consider the limitations in modeling the spread of extreme ideas in social networks. When a new idea is introduced to a society dominated by certain ideologies, individuals tendency to accept or reject the idea is dependent on the success of the new ideology against the predominant ones. We can think of the success of the idea as the incentives it provides the individuals with. However, for extremist ideas, these incentives are difficult to quantify. For the purpose of this paper, we don't focus on why people choose a certain idea, but how these ideas spread given certain incentives.

	E	P	N
E	α	β	-c
P	β	δ	δ
N	δ	δ	δ

Figure 1: payoff matrix, A

	E	P	N
E	1	w	w
P	w	1	1
N	w	1	1

Figure 2: probability matrix, W

Methods: We use game theory to model the transition between possible strategies for individuals in a population. Since we are only interested in the spread of extremist ideas and movements, we denote all of the other ideologies as non-extremists. In the context of our model, it's hard to imagine a clear transition from being a non-extremist to extremist, therefore we introduce an intermediate strategy: passive extremism. These individuals are contaminated with the idea and would turn into extremists under the right conditions, however, they don't share their ideas or partake in any extremist activities. We then have a 3-strategy game where we denote the strategies:

Extremist (E): Accepted the idea and acts accordingly. Might share the idea with a certain probability (w) and forms a coalition with another extremist or potential extremist if the idea is shared.

Passive extremist (P): Contaminated with the idea and might accept it if the conditions are suitable. Can be recruited by the extremists and gets a benefit in that case. Passive otherwise.

Non-extremist (N): All the other individuals in the population.

We assume in our model that extremists are homophilic. They share their extremist ideas conservatively with everyone who is not an extremist and interact with each other in clusters. We assume that since passive extremist don't partake in any activities, extremists can't tell them apart from non-extremists. Therefore, we have the probability matrix W (figure 2) that models the interactions between individuals in the population. We use the payoff matrix A (figure 1) to model the results of the interactions. In this case:

$$0 \leq \delta \leq \alpha \leq \beta \leq c \ \ and \ \ 0 \leq w \leq 1$$

δ is the benefit individuals get from their regular daily interactions. We interpret this parameter as an indicator for population welfare. α is the intrinsic value of the idea to the extremists. β is the benefit

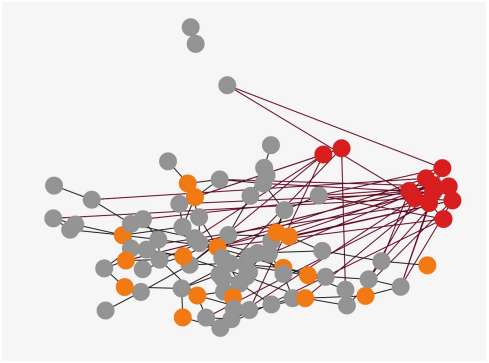

Figure 3: Social Network

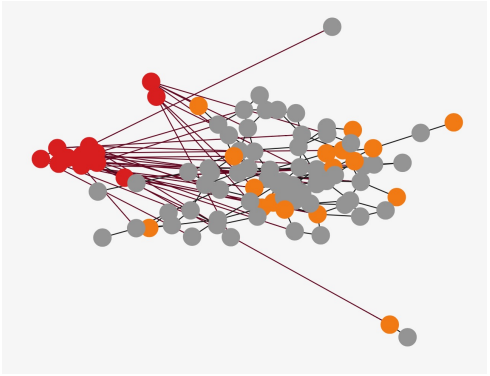

Figure 4: Social Network

of recruiting an extremist or converting to an extremist. We assume that when an extremist shares his controversial ideas with a non-extremist, there is a chance he gets detained. $-c$ represents the intrinsic value the extremist assigns to the chance of getting caught. The probability parameter w is an indicator for the censorship on extreme ideas, since under high levels of censorship, controversial ideas are less likely to be communicated publicly. For simplicity, we assume:

$$censorship \propto (1 - w)$$

Experimental Setup: We assume that our population is not well-mixed as most social networks are found to be homophilic. In the case of extremism, we expect extremists to form connections in clusters and be highly homophilic especially when sharing their controversial ideas. Therefore, we need to introduce a graph that represents the social networks accurately. We used the Python package *networkx* to create graphs for our simulation. The graphical structure we simulated was two inter-connected relaxed caveman graphs, one for the general population and one for the extremists. In this design, there are smaller, inter-connected clusters that then generate links with outside groups based on a probability parameter p (figure 3).

The general population of non-extremists and passive extremists was constructed of $N = 100$ nodes, where based on parameter n they were randomly initiated as either non-extremist or passive. Their connections were re-weighted with $p = .6$ to generate a social structure that reasonably mimics human connections. Given that their communication is uncensored, the edge weight of this sub-graph is 1,

meaning that all messages transfer perfectly through the graph.

The extremist population was constructed of $N * .1 = 10$ nodes, where based given their internal organization, they are likely highly interconnected. Thus, their connections were re-weighted with $p = .9$ to generate a social structure that reasonably mimics extremist groups. We assume that given that they know who else is an extremist and can communicate between each other, their edge weights in this subgraph is also 1.

The two populations were then merged, such that each extremist connected with between two to four members of the general population, where the edge weight is parameter $0 < w < 1$, meaning the level of censorship the extremist population encounters when attempting to recruit from the general population. As seen in figures 3 and 4, each time we run the simulation, the output has the same parameters, but the connections are generated randomly based on the parameters we've set.

We updated strategies based on a stochastic imitation model. At each time click, a random node is selected to update it's strategy. It either stays with the same strategy or switches with probability proportional to the fitness of the neighbors. Fitness of each node is calculated based on:

$$f_i = \sum_{j \in \text{neighbors}} A_{\text{type of i, type of j}} * (\text{edge})_{ij} \qquad \text{where A is the payoff matrix}$$

This randomly selected node then chooses strategy i with probability:

$$p_i = \frac{f_i}{f_i + \sum_{j \in \text{neighbors}} f_j}$$

This process is run for $t = 1000$ time clicks to ensure that it will be in its final form. At this point, the frequency of extremists ($\frac{\text{number of extremists}}{N}$) is calculated and stored.

Results: For the heat map (figure 5), we considered 20x20 levels of censorship and cost (c) and averaged the frequency of extremists over 100 trials for each entry in the 20x20 matrix. We applied the same process for welfare and censorship (figure 6). Since we've used stochastic updating strategies, we needed a large number of trials averaged (n=100) to see trends with respect to variables. For both graphs, the heat of the square corresponds to the average frequency of extremists at the end of the 100 trials. Red represents higher frequencies and white represents lower frequencies of extremists. The pink bar

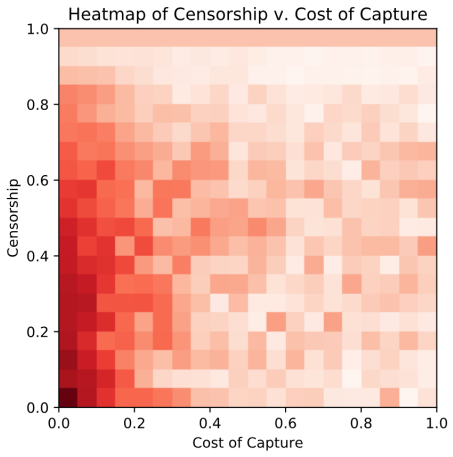

Figure 5: Freq. of Extremists (c vs. w)

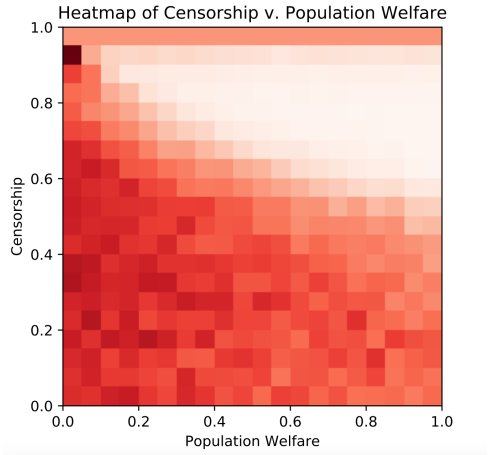

Figure 6: Freq. of Extremists (δ vs. w)

at the top represents the initial frequency of extremists since when censorship = 1 (w = 0), there are no interactions between the extremists and the general population. Thus, their numbers remain constant.

We first attempted to understand the relationship between censorship, punishment and the spread of extremism (Figure 5). Cost of capture or punishment is modeled by parameter $-c$. Here we assume that when an extremist shares his controversial ideas with a non-extremist, he has a chance of getting caught. c represents the intrinsic value the extremist assigns to this risk. In these trials, we concluded rather trivially that as the cost of getting caught increased, the rate of extremism decreased. Since there is a limit to what the cost of capture can be, it's not useful to just suggest that increasing the punishment would solve the problem. The y-axis on this heat map is $1 - w$ which represents the level of censorship in the population (1 = high censorship). For this heat map, we don't observe a clear relationship between censorship and frequency of extremists at the end. At low levels of cost, we observe that censorship decreases the vulnerability to extremists, but it requires extremely high levels of censorship (~ 1) which is not feasible in real life situations.

We then attempted to understand the relationship between censorship, population welfare and the outcome of extremists (Figure 6). For these trials, the population welfare is modeled by parameter $0 \leq \delta \leq \alpha$. This heat map is more interesting as it shows a clear direction to reduce the population's vulnerability to extremists. There is a nearly linear line between the population welfare and the censor-

ship needed to reduce the outcome of extremists ($(1 - w) + \delta =$ constant). This heat map suggests that theoretically, increasing both the censorship and welfare would reduce the vulnerability to extremists. However, it's hard to imagine a situation in real life where welfare can be increased in the existence of high levels of censorship. Therefore, there exists a trade-off between welfare and censorship.

Conclusion: Based on the experimental outcome, we draw several conclusions about the efficacy of tweaking various parameters in an attempt to curb the spread of extreme ideas. We first conclude that the results demonstrate that increasing punishment for extreme ideas limits their spread. However, there is a limit to the extent of reasonable punishment in societies. Since c is the intrinsic value the extremist assesses to the chance of getting caught, we can also try to improve our national security apparatus. However, this also comes at a cost and it's difficult to ensure extremely high levels of security without violating people's rights. The other solutions involve increasing the censorship on extremist ideas and increasing the population welfare. While we observe that increasing both would theoretically yield the best results based on our model, this is not feasible in real life. Lack of freedom of expression and high levels of censorship is typically correlated with lower levels of population welfare, which would in turn create the necessary conditions for extreme ideas to take root. Thus, based on our trials, we observe a relatively stable boundary between extremists taking hold of the population and being eliminated. Therefore, the minimum censorship level for a given society can be determined such that extremist ideas cannot spread, while minimizing the negative effect on the welfare of the population. Then, efforts can be made to increase the population welfare which is a healthier solution for our problem. Based on our assumptions, there exists an ideal level of censorship on extreme ideas for each society.

References

1. Acemoglu, D., Ozdaglar, A. (2011). Opinion Dynamics and Learning in Social Networks.Dynamic Games and Applications,1(1), 3-49.
2. Aleroud, A., Gangopadhyay, A. (2016). Multimode Co-Clustering for Analyzing Terrorist Networks.Information Systems Frontiers, 1-22.
3. Bergstrom, T. C. (2003). The Algebra of Assortative Encounters and the Evolution of Cooperation.International Game Theory Review,5(03), 211-228.
4. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G. (2000). Mixing Beliefs among Interacting Agents.Advances in Complex Systems,3(01n04), 87-98.
5. Eshel, I., Cavalli-Sforza, L. L. (1982). Assortment of Encounters and Evolution of Cooperativeness.Proceedings of the National Academy of Sciences,79(4), 1331-1335.
6. Evans, T., Fu, F. (2018). Opinion Formation on Dynamic Networks: Identifying Conditions for the Emergence of Partisan Echo Chambers.arXiv preprint arXiv:1807.01252.
7. Fu, Feng., Nowak, M. A., Christakis, N. A., Fowler, J. H. (2012). The Evolution of Homophily.Scientific Reports,2, 845.
8. Gundabathula, V. T., Vaidhehi, V. (2018). An Efficient Modelling of Terrorist Groups in India using Machine Learning Algorithms.Indian Journal of Science and Technology,11(15).
9. Helfstein, S. (2012).Edges of radicalization: Ideas, Individuals and Networks in Violent Extremism. Military Academy West Point, NY, Combating Terrorism Center.
10. Jackson, M. O., Watts, A. (2002). On the Formation of Interaction Networks in Social Coordination Games.Games and Economic Behavior,41(2), 265-291.
11. Mahmood, T., Rohail, K. (2012, October). Analyzing Terrorist Incidents to Support Counter-Terrorism-Events and Methods. InRobotics and Artificial Intelligence (ICRAI), 2012 International Conference on(pp. 149-156). IEEE.

12. Malm, A., Nash, R., Moghadam, R. (2016). Social Network Analysis and Terrorism.The Handbook of the Criminology of Terrorism, 221-231.

13. Nizamani, S., Memon, N. (2011, September). Evolution of Terrorist Network using Clustered Approach: A Case Study. InIntelligence and Security Informatics Conference (EISIC), 2011 European(pp. 116-122). IEEE.

14. Pilny, A. N. (2015).Social Movements as Networks of Communication Episodes(Doctoral dissertation, University of Illinois at Urbana-Champaign).