Stubborn Agents in Networks with Homophily
An Experiment

Tilman Jacobs

June 28, 2024

Abstract

I present an experimental design to study the susceptibility of different types of networks to the influence of stubborn agents. Subjects are shown a private signal and asked to make a guess about the true state. After, they are shown the guesses of their neighbours and asked to guess again. A certain number of subjects are designated stubborn agents. Their objective is not truth seeking but to steer consensus as close as possible towards their initial belief. I formulate research hypotheses about their effect on different networks and lay out how to answer them using the data obtained from implementing my experimental design.

1 Introduction

When making important decisions agents regularly rely on their social networks. Job seekers (Montgomery, 1991), voters (Beck et al., 2002), and consumers (Trusov et al., 2009) all use their networks to help inform their decisions. Unfortunately, not all interactions in an agents social network occur in good faith. Advertisers must not necessarily be incentivized to provide accurate information to a consumer and political operatives will support their candidates regardless of the voters best interests. Agents that do not participate in the general social learning process but stubbornly stick to their initial beliefs whilst influencing others are fittingly dubbed "stubborn agents" or "zealots" in the social learning literature. These kinds of agents where first formalized by Mobilia (2003); Mobilia et al. (2007) and have since been studied in multiple models. Wu and Huberman (2004) consider the effect of stubborn agents in a network where agents update their beliefs by averaging neighbors beliefs. Chinellato et al. (2007) provide a more general analysis and Acemoğlu et al. (2013) study a general network with continuous opinions. Yildiz et al. (2013) study the optimal placement of stubborn agents in a network with binary opinions and find that the presence of opposing stubborn agents prevents consensus.

Apart from agents actions the social learning literature has isolated other, structural network characteristics that influence opinions and behaviour. The two main features of importance are that agents or nodes in a network often exhibit large differences in the number of their connections and that in many real-world networks agents associate disproportionately with those who are similar to them which is referred to as Homophily (Golub and Jackson, 2012; Bessi et al., 2016). Empirically, homophily relates to the concepts of echo chambers and polarization (Tsang and Larson, 2016; Barberá, 2020; Baumann et al., 2020; Del Vicario et al., 2016) where only receiving information from those with the same beliefs can create feedback loops and deeply entrench those beliefs.

In this paper I am concerned with studying how stubborn agents and the presence of homophily interact. On their own both are known to hamper consensus and sometimes create insular communities of homogenous beliefs detached from the outside world (Acemoglu et al., 2010; Golub and Jackson, 2012). Intuition tells us that a stubborn agent placed in a community that rarely receives or considers outside information will have a strong impact on it's belief. To describe the propagation of beliefs and information in networks I use a model referred to as deGroot updating where agents update their beliefs by averaging the beliefs of all other agents available to them. In Section 2, I use the literature on optimal stubborn agent placement as well as a number of simulations in networks with low, medium and high homophily to derive several hypotheses about the effect of stubborn agents in networks with homophily. In Section 3, I present an experimental Design to test these hypotheses and in Section 4 I lay out how to use the data obtained from implementing this design.

Related Literature

For a comprehensive survey of the literature on learning in networks see Golub and Sadler (2017) and for experimental works refer to Choi et al. (2016). Early investigations of belief formation in

networks include Choi et al. (2005); Kearns et al. (2012) The experiment proposed in this paper is most closely related to the one by Choi et al. (2023). The authors investigate the effect of homophily and connectivity on consensus formation using three canonical networks. They find that both factors have negative effects on consensus formation. However, they do not consider any heterogenous agents.Grimm and Mengel (2020) study belief formation in differently structured networks in a laboratory environment. They find that the network structure matters significantly to the propagation of beliefs. They also consider which models of learning are most appropriate in different contexts. The results support the predictions of deGroot updating. However, they focus on small networks with few nodes. Chandrasekhar et al. (2020) study different modes of belief formation in networks. They find evidence for both Bayesian agents and deGroot updating in real world environments. Coutts (2019) tests how financial incentives affect beliefs in the laboratory. Finally, Nyarko and Schotter (2002) propose and study an elicitation procedure to directly observe beliefs in the laboratory.

2 Theory and Hypotheses

After Gale and Kariv (2003) and Choi et al. (2023) the basic model has binary states, signals and guesses. The set of agents is $N = \{1, 2, \ldots, n\}$. Further, the world is in one of two possible states $\omega \in \{0, 1\}$. Agents believe both states to be equally likely a priori and time is discrete. In period zero a private signal $s_i \in 0, 1$ is shown to each agent. The signal is noisy but informative meaning the probability that s_1 corresponds to the true state ω is $p \in (0.5, 1)$. After receiving their signal each agent updates their a priori beliefs about ω accordingly. In each proceeding period t agents make a binary guess $a_{i,t} \in \{0, 1\}$. Agents receive a payoff of one if they guess correctly and zero otherwise. Agents are connected by a network or graph $G(N, E)$. N is the set vertices and corresponds to the agents. E is the set of edges and decides whether a pair of agents is connected. If a two agents i and j are connected or $(i, j) \in E$ they observe each others guesses in each round. An agent i's neighborhood on the graph is $N(i) = \{j \in N : (i, j) \in E\}$. Thus, in a period t an agent i is aware of the guesses of all agents $j \in N(i)$ from periods 1 through $t - 1$.

Since it is always optimal to guess ones signal in round one agents can deduce their neighbors signal from their first period guesses. Depending on the information about the network structure agents can also make inference about neighbors' neighbors' signals in the following rounds and so on. An agent that updates their beliefs according to Bayes' rule and takes into account all available information is called a Bayesian Agent. However, such calculations are highly complex and computationally infeasible in many scenarios. Due to these complications I will follow Choi et al. (2023) and consider agents that act similar to the updating rule established by Degroot (1974) instead. Agents simply average the previous period's guesses of their neighbors, including their own. If the average is greater than $\frac{1}{2}$ they adopt the belief $\omega = 1$, if the average is less the new belief is

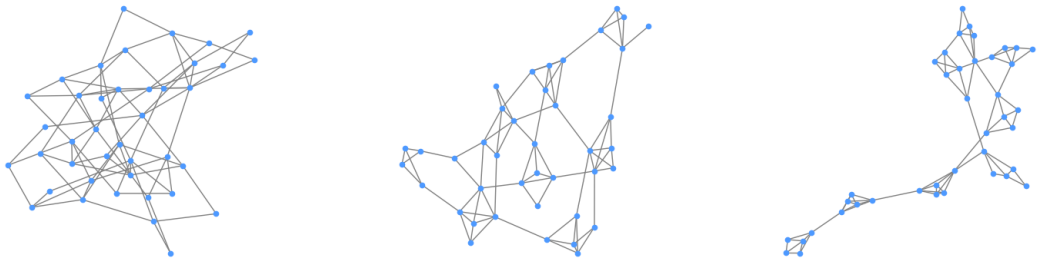
$\omega = 0$. If the average is exactly $\frac{1}{2}$ the agent does not change their belief. Formally I write

$$a_{i,t} = \begin{cases} 1, & \text{if } \mu_{i,t} > \frac{1}{2}, \\ 0, & \text{if } \mu_{i,t} < \frac{1}{2}, \\ a_{i,t}, & \text{otherwise,} \end{cases}$$

$$\text{where } \mu_{i,t} = \frac{1}{|N(i)|+1} \left(\sum_{j \in N(i)} a_{j,t-1} + a_{i,t-1} \right)$$

Chandrasekhar et al. (2020) test the prevalence of Bayesian agents in different real world scenarios and find that the share of Bayesian individuals is generally less than half and as low as 10 percent in some scenarios. Other researchers finds that the predictions of deGroot updating are consistent with experimental results.Thus, deGroot updating is a reasonable assumption. In a departure from the model I now assume that there is a subset of individuals $S \subset N$. An agent $i \in S$ is called a stubborn agent. Stubborn agents differ from the other agents in the following ways. First, a stubborn agent receives a fully informative signal, that is they know the true state. Second, a stubborn agents payoff is one if they guess the wrong state and zero otherwise. Lastly, stubborn agents never change their beliefs or guesses. The last characteristic is a direct consequence of the first two, assuming the agents behave rationally.

To account for the effect of different degrees of homophily I will consider three different types of networks. To model a real-world network with low homophily I will consider an Erdős-Renyi network. For higher degrees I will consider two types of Stochastic block networks with increasing homophily. Examples of these three networks on 40 vertices are given in Figure 1. The Erdős-Renyi network includes 80 edges on 40 vertices. Consequently it has average degree four. Over a thousand such randomly generated networks the average diameter was 5.537 and average clustering 0.01. A stochastic Block network is characterized by a number of communities or blocks with a high density of links and low connectivity between groups. The two Stochastic Block networks in this paper contain 8 blocks of five vertices each. In their generation the linking probability within a block is 0.85. For the lower homophily Stochastic Block graph the outside-of-group linking probability is 0.28 while in the higher homophily one it is 0.14. These specifications resulted in average degree 4.3973 (3.9077), diameter 6.647 (8.639) and clustering coefficient 0.4734 (0.6022) for the low (high) homophily Stochastic Block networks.



(a) Erdős-Renyi network

(b) Stochastic Block network (low homophily)

(c) Stochastic Block network (high homophily)

Figure 1: Networks

The theory of learning under homophily clearly shows that higher homophily adversely affects convergence and the speed of learning (Golub and Jackson, 2012). Choi et al. (2023) find experimental evidence congruent with these predictions. In isolation stubborn agents too have negative effects on these parameters. The literature on echo chambers in political processes suggest that there might be compounding effects when both are present (Tsang and Larson, 2016). To substantiate this and to formulate concrete, experimentally testable hypotheses I run simulations of deGroot updating on the networks described above. Each network is generated 100000 times with signals drawn with quality $p = 0.7$ and without stubborn agents. Then the process is repeated with five nodes being randomly chosen to be stubborn. To measure the learning in the network I adapt a metric from Choi et al. (2023) to fit the stubborn agent environment.

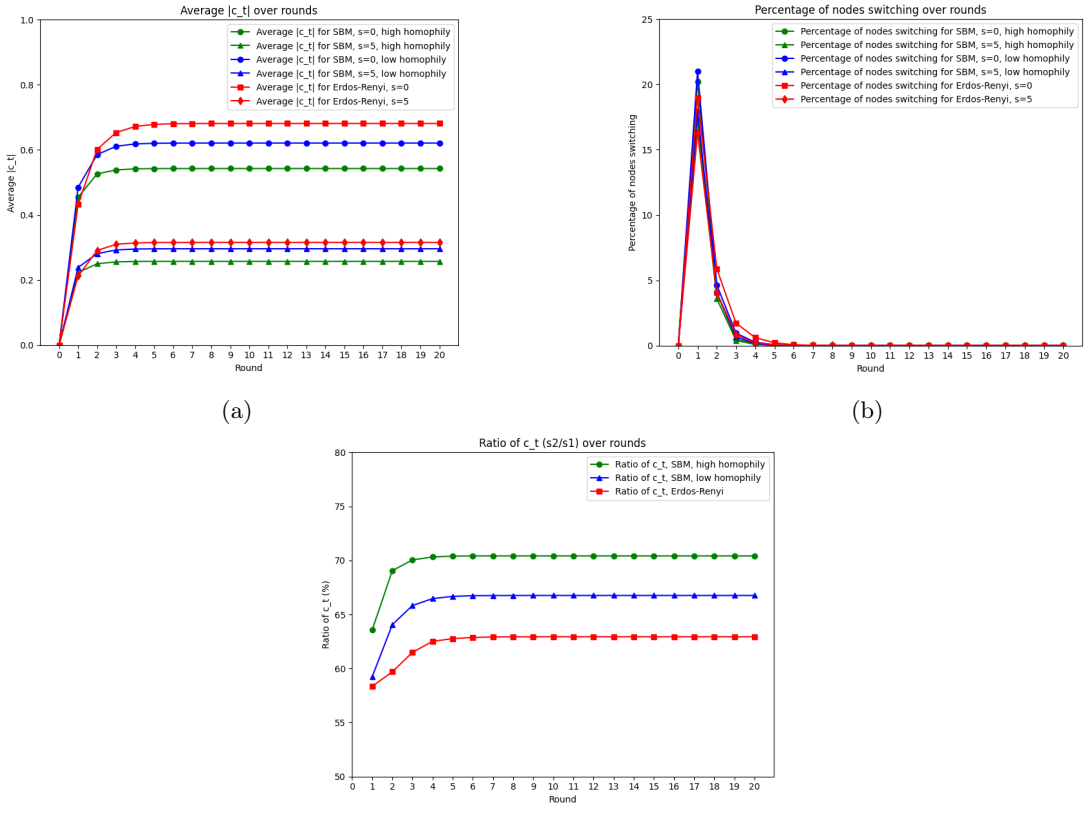
$$c_t = \begin{cases} (n_t - n_0)/(n - s - n_0), & \text{if } n_t \geq n_0 \\ (n_t - n_0)/n_0, & \text{if } n_t > n_0, \end{cases}$$

where n_t represents the amount of correct guesses in period t, n_0 the number of correct signal received in the beginning and s is the number of stubborn agents. The metric measures the average guess's movement towards or away from the correct consensus. The denominator serves to normalize it by the maximal potential movement in either direction. If the network reaches a false consensus c_t is -1 and if the correct consensus is reached it is 1. To measure the impact of the stubborn agents on the learning process I define another variable r_t which simply gives the percentage decrease in the c_t variable associated with a change in the number of stubborn agents present.

$$r_t = (1 - c_{t,s_1}/c_{t,s_2}) \cdot 100$$

where s_1, s_2 denote the number of stubborn agents in the two scenarios to be compared. The results of the simulations are illustrated in Figure 2. Figure 2a gives distributions of the averages of the absolute changes in learning for all six distinct setups. In line with the theoretical predictions

and results from the literature increasing homophily is negatively correlated with the amount of learning the networks support. As is consistent with deGroot learning the large majority of the learning occurs in the few first rounds after which opinions converge towards an equilibrium in all three network types regardless of the presence of stubborn agents. Figure 2b further supports this as it shows that on average after the fifth round opinions are static in all networks. Figure c shows the distribution of the average r_t values accross the three networks. Congruent with a compounding of effects the influence of stubborn agents is highest in the high homophily Stochastic block networks where the estimated average decrease in the learning coefficient after 20 rounds is 70 percent. In the lower homophily Stochastic Block networks the decrease is 67 percent and in the Erdős-Renyi networks, which have the lowest homophily, it is only 62 percent. An interesting curiosity is that if homophily is taken to it's extremes in the Stochastic Block setup the r_t coefficient appeared to be declining. This might be due to the fact that those kind of networks are more likely to be disconnected or extremely insular. Therefore, the effect of one or more stubborn agents could easily be isolated from the rest of the network. Such considerations are however beyond the scope of this paper.



(a) (b)

(c)

Figure 2: Simulation Results

On the basis of the simulations and theory laid out above I formulate the following hypotheses.

H1 Even when stubborn agents are present subjects guesses converge across all networks.

H2 The presence of stubborn agents leads to the breakdown of consensus in all networks.

H3 The effect of stubborn agents is stronger in the high homophily Stochastic Block network than it's low homophily counterpart and is weakest in the Erdős-Renyi network.

Intuitively, since the Stochastic Block networks are separated into a number of blocks that have relatively less contact to the remaining network they also have access to less information. Therefore we would assume that they are less likely to reach consensus on their own. This is only exacerbated when we reduce the number of connections between these blocks as we do in the high homophily Stochastic Block. In the Erdős-Renyi network there are no such blocks and information access is much broader. Thus, consensus is more likely to be achieved. As for the effect of stubborn agents the insular nature of the Stochastic Blocks means that a stubborn agents reported guesses represent a greater share of the overall information available to a neighboring agent. This agent then receives lower quality information and is in turn more likely to develop or sustain false beliefs. If a stubborn agent happens to occupy the only connection of a block to the rest of the network it effectively disconnects this block. In such a case the amount of information available to it's constituting agents.

3 Experimental Design

Due to the large scale of the proposed networks and the need to repeatedly run the experiment, implementing it will be resource intensive. First and foremost a large number of subjects will be needed. Choi et al. (2023) ran an experiment on three equally sized networks which I will follow closely with the appropriate adaptations to measure the impact of stubborn agents. They recruited 480 individuals. All choices about the further specifications of the experiment must be made with the available number of subjects in mind. Regardless of its size the pool of subjects should be split into 12 groups of equal size. Ideally, each group would contain 40 subjects as in the simulations but smaller groups of for example 20 members would be workable. If a smaller group size is chosen rerunning all simulations at that size and reformulating the hypotheses where necessary would be advisable for internal consistency.

At the start of the experiment a large number network will be generated according to the specifications laid out in Section 2 and the one amongst them most clsoely matching the average characteristics: average degree, diameter and clustering will be chosen. Each subject will then be identified with one node in the networks described. Five subjects will be designated *stubborn*. To send a noisy signal and induce appropriate beliefs non stubborn subjects will be told about the existence of a bag containing ten balls of either red or green color. The distribution of which is either ω_1: seven red and three green or ω_2: three red and ten green. The two possible states will be denoted *red bag* and *green bag* respectively and are equally likely a priori. After non stubborn subjects privately observe a ball being drawn from the true bag and stubborn subjects are informed of the true state all subjects are prompted to report a guess about whether the bag is green or red.

They are then shown the guesses of all subjects that are their neighbors in the network. Crucially, they cannot distinguish stubborn neighbors from those who are not stubborn. Each round of the experiments should contain at least ten periods of subjects guessing and updating their beliefs. Since subjects must be incentivized to guess correctly one of their guesses will be chosen at random at the end of each round. If this guess matches the true state the subject will receive a monetary reward in line with local standards and the expected duration of the experiment. If the guess does not match, the subject will receive no reward for this round. Subjects designated stubborn will simply receive their reward based on whether they consistently reported false guesses or not. As mentioned above, for usable data to be collected each round should have subjects guess and update their beliefs at least ten times. Thus, it might be necessary to shrink the subject pool to be able to afford more rounds. A session concludes after four rounds of the experiment on the same network. For each network four sessions should be run, totalling 12 sessions all together. Subjects should only participate in one session each. I assume one session would take between one and two hours to complete given that subjects receive thirty seconds to submit their guesses. For reference Choi et al. (2023) report their sessions taking 1.5 hours in their very similarly structured experiment.

4 Pre-Analysis Plan

In line with my hypotheses there are a number of parameters of interest to be recorded during the experiment. To decide whether our evidence supports hypothesis **H1** it suffices to record the number of switches that occur in every round. In the simulations I observed that agents do not change their reported beliefs after round five on average. If this is also reflected in the experimental data I can conclude that the distribution of beliefs does indeed converge. Accepting hypothesis **H1** would add to the experimental evidence for deGroot learning in real world scenarios as fast convergence is one of its core predictions. To decide about the validity of hypothesis **H2** I compare the fraction of consensus reached in the respective networks when stubborn agents are present to the base case of when they are not. Another important statistic is the number of subjects that do not guess their initial signal in the first round. As this is clearly irrational for non stubborn subjects a large share of subjects behaving like this would force us to question all other results since I strongly rely on rationality in the theory.

To measure the effects on consensus I simply measure the previously discussed variable c_t in the experimental data. This will allow us to both test previous experimental findings about the effect of homophily on convergence as well as assess stubborn wether stubborn agents do break down consensus in real world networks. To test the joint effect I need to record the r_t variable. I will use OLS to regress the consensus outcomes on dummy variables for the different network types and for the presence of stubborn agents. To directly test whether homophily and stubborn agents interact congruently with **H3** I need to add an interaction term.

To further test whether our deGroot learning assumption was valid I will use the experimental data to compute the predicted guesses in each period. This allows me to compare the predictions

and actual guesses. For comparison random guessing will match deGroot learning in about 60 percent of cases.

References

D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis) information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.

D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar. Opinion Fluctuations and Disagreement in Social Networks. *Mathematics of Operations Research*, 38(1):1–27, Feb. 2013. ISSN 0364-765X, 1526-5471. 10.1287/moor.1120.0570.

P. Barberá. Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, pages 34–55, 2020.

F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini. Modeling Echo Chambers and Polarization Dynamics in Social Networks. *Physical Review Letters*, 124(4):048301, Jan. 2020. ISSN 0031-9007, 1079-7114. 10.1103/PhysRevLett.124.048301.

P. A. Beck, R. J. Dalton, S. Greene, and R. Huckfeldt. The social calculus of voting: Interpersonal, media, and organizational influences on presidential choices. *American political science review*, 96(1):57–73, 2002.

A. Bessi, F. Petroni, M. D. Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi. Homophily and polarization in the age of misinformation. *The European Physical Journal Special Topics*, 225(10):2047–2059, Oct. 2016. ISSN 1951-6355, 1951-6401. 10 .1140/epjst/e2015-50319-0.

A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri. Testing Models of Social Learning on Networks: Evidence From Two Experiments. *Econometrica*, 88(1):1–32, 2020. ISSN 0012-9682. 10.3982/ ECTA14407.

D. D. Chinellato, M. A. M. de Aguiar, I. R. Epstein, D. Braha, and Y. Bar-Yam. Dynamical Response of Networks under External Perturbations: Exact Results, Nov. 2007.

S. Choi, D. Gale, and S. Kariv. Behavioral aspects of learning in social networks: An experimental study. In *Experimental and Behavorial Economics*, pages 25–61. Emerald Group Publishing Limited, 2005.

S. Choi, E. Gallo, and S. Kariv. Networks in the laboratory. 2016.

S. Choi, S. Goyal, F. Moisan, and Y. Y. T. To. Learning in Networks: An Experiment on Large Networks with Real-World Features. *Management Science*, 69(5):2778–2787, May 2023. ISSN 0025-1909, 1526-5501. 10.1287/mnsc.2023.4680.

A. Coutts. Testing models of belief bias: An experiment. *Games and Economic Behavior*, 113: 549–565, 2019.

M. H. Degroot. Reaching a Consensus. *Journal of the American Statistical Association*, 69(345): 118–121, Mar. 1974. ISSN 0162-1459, 1537-274X. 10.1080/01621459.1974.10480137.

M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*, 6 (1):37825, 2016.

D. Gale and S. Kariv. Bayesian learning in social networks. *Games and economic behavior*, 45(2): 329–346, 2003.

B. Golub and M. O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012.

B. Golub and E. Sadler. Learning in social networks. *Available at SSRN 2919146*, 2017.

V. Grimm and F. Mengel. Experiments on belief formation in networks. *Journal of the European Economic Association*, 18(1):49–82, 2020.

M. Kearns, S. Judd, and Y. Vorobeychik. Behavioral experiments on a network formation game. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 690–704, Valencia Spain, June 2012. ACM. ISBN 978-1-4503-1415-2. 10.1145/2229012.2229066.

M. Mobilia. Does a Single Zealot Affect an Infinite Group of Voters? *Physical Review Letters*, 91 (2):028701, July 2003. ISSN 0031-9007, 1079-7114. 10.1103/PhysRevLett.91.028701.

M. Mobilia, A. Petersen, and S. Redner. On the role of zealotry in the voter model. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08029, 2007.

J. D. Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American economic review*, 81(5):1408–1418, 1991.

Y. Nyarko and A. Schotter. An experimental study of belief learning using elicited beliefs. *Econometrica*, 70(3):971–1005, 2002.

M. Trusov, R. E. Bucklin, and K. Pauwels. Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing*, 73(5):90–102, Sept. 2009. ISSN 0022-2429, 1547-7185. 10.1509/jmkg.73.5.90.

A. Tsang and K. Larson. The echo chamber: Strategic voting and homophily in social networks. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 368–375, 2016.

F. Wu and B. A. Huberman. Social Structure and Opinion Formation, July 2004.

E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary Opinion Dynamics with Stubborn Agents. *ACM Transactions on Economics and Computation*, 1(4):1–30, Dec. 2013. ISSN 2167-8375, 2167-8383. 10.1145/2538508.