Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-5 (2020) 307–310

Rethinking the Micro-Foundation of Opinion Dynamics: Rich Consequences of an Inconspicuous Change ⋆

Wenjun Mei * **Francesco Bullo** ** **Ge Chen** *** **Julien M. Hendrickx** **** **Florian Dörfler** †

* *Automatic Control Laboratory, ETH Zurich, 8006 Zurich, Switzerland (e-mail: wmei@ethz.ch)*

** *Center for Control, Dynamical Systems, and Computation, University of California, Santa Barbara, CA 93106, USA, (e-mail: bullo@engineering.ucsb.edu)*

*** *Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100190, China (e-mail: chenge@amss.ac.cn)*

**** *Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université catholique de Louvain, Louvain-la-Neuve B-1348, Belgium (e-mail: julien.hendrickx@uclouvain.be)*

† *Automatic Control Laboratory, ETH Zurich, 8006 Zurich, Switzerland (e-mail: dorfler@ethz.ch)*

Abstract: Mathematical modeling plays a fundamental role in understanding how social influence shapes individuals' opinions. Although most opinion dynamics models assume that individuals update their opinions by averaging others' opinions, we point out that the weighted-averaging mechanism features a non-negligible unrealistic implication. We propose a new micro-foundation of opinion dynamics, i.e., the weighted-median mechanism, in the framework of cognitive dissonance theory and resolves the shortcomings of weighted averaging. Validation via empirical data indicates that the weighted-median mechanism significantly outperforms the weighted-averaging mechanism in predicting individual opinion shifts. Compared with the averaging-based opinion dynamics, the weighted-median model, despite its simplicity in form, replicates more realistic features of opinion dynamics, and exhibits richer phase-transition behavior depending on more delicate and robust network structures. The novel weighted-median model opens up a new line of research and renovates our understanding of the opinion formation process.

Keywords: Opinion dynamics, Social networks, Multi-agent systems

1. INTRODUCTION

Nowadays public opinion formation is deeply influenced by social networks and faces unprecedented challenges such as radicalization, echo chambers, and misinformation. Mathematical modeling of opinion dynamics plays a fundamental role in gaining reliable understanding of how empirically observed macroscopic sociological phenomena emerge from certain microscopic social-influence mechanisms and social network structures. Due to the complicated nature of interpersonal influences, the key challenge in building predictive and tractable mathematical models is to identify the proper micro-foundation of opinion dynamics.

Most existing deterministic opinion dynamics models originate from the classic *DeGroot model* (French Jr., 1956; DeGroot, 1974), given by the equation below:

$$x_i(t+1) = \text{Mean}_i(x(t); W) = \sum_{j=1}^{n} w_{ij} x_j(t), \qquad (1)$$

for any $i \in \{1, \dots, n\}$, where $x_i(t)$ is individual i's opinion at time t. The matrix $W = (w_{ij})_{n\times n}$ is row-stochastic and defines a directed weighted graph $\mathcal{G}(W)$, referred to as the *influence network*. Despite its mathematical elegance, the DeGroot model (1) leads to overly-simplified predictions, e.g., the system reach consensus as long as the influence network satisfies some very mild connectivity conditions. To capture the phenomenon of persistent disagreement, various extensions have been proposed, including the DeGroot model with absolutely stubborn individuals (Acemoglu et al., 2013), the bounded-confidence model with interpersonal influences truncated according to opinion distances (Hegselmann and Krause, 2002), and the Friedkin-Johnsen model with individual prejudice (Friedkin and Johnsen, 1990). These extensions are also based on weighted averaging opinion updates. Despite being successful in generating persistent disagreement and being mathematically sophisticated, none of these aforementioned models fully captures other prominent features of opinion dynamics supported by empirical studies and everyday experi-

⋆ We acknowledge the financial support of the U. S. Army Research Laboratory and the U. S. Army Research Office under grant numbers W911NF-15-1-0577 and W911NF-16-1-0005, the National Key Basic Research Program of China under grant number 2016YFB0800404, as well as the ETH Zurich funds.

10.1016/j.ifacol.2021.04.187

ence, e.g., the connection between social marginalization and opinion radicalization (McCauley and Moskalenko, 2008) and lower consensus likelihoods in larger groups (Hare, 1952).

2. THE WEIGHTED-MEDIAN OPINION DYNAMICS: DERIVATION AND EMPIRICAL VALIDATION

The bottleneck in predictive power met by the aforementioned models inspires us to retrospect the very foundation of opinion dynamics. Here we point out that the weighted-averaging mechanism, adopted as the micro-foundation by DeGroot model and all its extensions, features a non-negligibly unrealistic implication, which is manifested by the following simple example: Suppose an individual i's opinion is influenced by individuals j and k via the weighted-averaging mechanism, i.e.,

$$x_i(t+1) = x_i(t) + w_{ik}\big(x_k(t) - x_i(t)\big) + w_{ij}\big(x_j(t) - x_i(t)\big).$$

The equation above implies that whether individual i's opinion moves towards $x_k(t)$ or $x_j(t)$ is determined by whether $w_{ik}|x_k(t) - x_i(t)|$ is larger than $w_{ij}|x_j(t) - x_i(t)|$. That is, the "attractiveness" of any opinion $x_j(t)$ to individual i is proportional to the opinion distance $|x_j(t) - x_i(t)|$.

We resolve such an unrealistic feature and propose a new micro-foundation of opinion dynamics in the framework of the cognitive dissonance theory in psychology: Individuals in a group experience cognitive dissonance from disagreement and attempt to reduce such dissonance by changing their opinions (Festinger, 1957; Matz and Wood, 2005). Therefore, opinion updates can be viewed as individuals' attempts to minimize such cognitive dissonance, the most parsimonious form of which is

$$x_i(t+1) \in \operatorname{argmin}_z \sum_j w_{ij}|z - x_j(t)|^\alpha, \text{ for } i \in \{1, \dots, n\},$$

with $\alpha > 0$. For example, $\alpha = 2$ for the DeGroot model. An exponent $\alpha > 1$ ($\alpha < 1$ resp.) implies that individuals are more sensitive to distant (nearby resp.) opinions. In the absence of any widely-accepted psychological theory in favor of $\alpha > 1$ or $\alpha < 1$, in this paper we adopts the neutral hypothesis $\alpha = 1$. One could easily check that, for generic weights, the best-response dynamics

$$x_i(t+1) = \operatorname{argmin}_z \sum_{j=1}^{n} w_{ij}|z - x_j(t)|$$

lead to a weighted-median opinion update mechanism. We formally define the weighted-median opinion dynamics as follows.

Definition 1. (Weighted-median opinion dynamics). Given any row-stochastic influence matrix W and any initial condition $x(0) \in \mathbb{R}^n$, at each time t, randomly pick one node i and update their opinion via the following equation

$$x_i(t+1) = \operatorname{Med}_i\big(x(t); W\big),$$

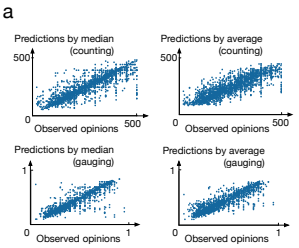
where $\operatorname{Med}_i(x(t); W)$ is the weighted-median of $x(t)$ associated with the weights given by the i-th row W. To be more specific, $\operatorname{Med}_i(x(t); W) = y \in \{x_1(t), \dots, x_n(t)\}$ such that

$$\sum_{j:\, x_j(t)<y} w_{ij} \le \frac{1}{2}, \quad \text{and} \quad \sum_{j:\, x_j(t)>y} w_{ij} \le \frac{1}{2}.$$

If such y is not unique, then let $\operatorname{Med}_i(x(t); W)$ be the weighted-median that is the closest to $x_i(t)$.

After careful examinations, once could conclude that the well-posedness and uniqueness of $\operatorname{Med}_i(x(t); W)$ is guaranteed.

Empirical validation on a longitudinal dataset (Vande Kerckhove et al., 2016) shows that the weighted-median mechanism



b

Counting Games, 3rd-round opinions

Predictions by	Median error rate	95% confidence interval	MER
Median	0.0714	[0.0667, 0.0769]	0.1776
Average	0.1331	[0.1230, 0.1408]	0.2332

Gauging Games, 3rd-round opinions

Predictions by	Median error	95% confidence interval	MAE
Median	0.0200	[0.0200, 0.0200]	0.0454
Average	0.0400	[0.0375, 0.0425]	0.0621

Fig. 1. Comparison between the weighted-median and the weighted-averaging mechanisms via empirical data analysis for a set of online experiments (Vande Kerckhove et al., 2016). Panel a are the scatter plots between the participants' observed answers at the 3rd rounds and the predictions by median and average respectively. Panel b presents the corresponding prediction errors/error rates, their 95% confidence intervals computed by the *binomial distribution method* (Bland, 2015), and mean error rate (MAE) or mean absolute-value error (MAE).

enjoys significantly lower errors than the weighted-averaging mechanism in predicting individual opinion shifts. This dataset is collected in a set of online human-subject experiments. Every single experiment involves 6 anonymous individuals, who sequentially answer 30 questions within tightly limited time. The questions are either guessing the proportion of a certain color in a given image (*gauging game*), or guessing the number of dots in certain color in a given image (*counting game*). For each question, the 6 participants give their answers for 3 rounds. After each round, they will see the answers of all the 6 participants as feedback and possibly alter their opinions based on this feedback. The dataset records, for each experiment, the individuals' opinions in each round of the 30 questions.

Since the participants are anonymous, it is reasonable to assume that the participants uniformly assign weights to each other. For each question, at each round, we use the average and the median of the participants' current answers respectively to predict their answers in the next round. For counting games, we randomly sample 18 experiments from the dataset, in which 71 participants give answers to all the 30 questions at each round. We apply the average and the median rules respectively to predict their answers in the 3rd round of each question, based on the participants' answers in the 2nd round, and then compare the *error rates* of the predictions, defined as follows:

$$\text{error rate} = \frac{\big|\text{prediction} - \text{true value}\big|}{\text{true value}}.$$

For the gauging games, we randomly sampled 21 experiments, in which 55 participants answers all the 30 questions at each round. Since these answers are already in percentages, we directly compare the magnitudes of errors. As Figure 1 shows, in counting games, the median error rate of the predictions by median is a stunning 46.36% lower than that of the predictions by average. In gauging games, the median error of the predictions by median is even 50% lower than that of the predictions by average. Results regarding the opinion shifts from the 1st rounds to the 2nd rounds yield to similar conclusions.

3. COMPARATIVE NUMERICAL STUDIES

Comparative numerical studies indicate that the weighted-median opinion dynamics replicate various non-trivial realistic features of opinion dynamics whereas the DeGroot model and

its extensions fail to. The models in comparison include the DeGroot model with absolutely stubborn individuals, the Friedkin-Johnsen model, and the networked bounded-confidence model, all with randomized model parameters.

Among all the opinion dynamics models compared in this section, our weighted-median model is the only one showing that peripheral nodes on influence networks are more vulnerable to extreme opinions. We independently simulate different models for 1000 times on a randomly generated scale-free network with 2000 nodes. The initial opinions are uniformly randomly generated from $[-1, 1]$ and opinions are classified into 4 categories: extreme ($[-1, -0.75) \cup (0.75, 1]$), radical ($[-0.75, -0.5) \cup (0.5, 0.75]$), biased ($[-0.5, -0.25) \cup (0.25, 0.5]$), and moderate ($[-0.25, 0.25]$). For each opinion dynamics model, we estimate the in-degree centrality distributions for individuals holding different categories of opinions at the steady states via the 1000 independent simulations. As indicated by Figure 2a, only in our weighted-median model, the in-degree centrality distributions for different categories of opinions are clearly separated, and the empirical probability density of the most extreme opinions decays the fastest as the in-degree increases. We further compute the individuals' extremist focuses, defined as the ratio of extreme opinion holders among their neighbors, at the final steady states, and plot the individuals' two-dimension distributions over the in-degree centrality and the extremist focus. As indicated by Figure 2b and c, compared with the entire population, extreme opinion holders in the weighted-median model tend to have higher extremist focus, which implies that extremists form small-sized clusters in peripheral areas of the influence network. This feature is consistent with the real-world data on Twitter user networks (Benigni et al., 2017). In this dataset, some Twitter accounts are labelled as ISIS supporters for posting pro-ISIS materials. For these ISIS supporters, the two-dimension distribution over the ISIS focus, i.e., the ratio of ISIS supporters among social neighbors, and the in-degree centrality, see Figure 2f, looks very similar to Figure 2c predicted by the weighted-median opinion dynamics, but very different to the predictions by other models, see Figure 2d and e.

Moreover, simulations on small-world networks show that, among all the models in comparison, only the weighted-median model and the networked bounded-confidence model reflect the realistic feature that larger networks have lower likelihoods of reaching consensus (Hare, 1952), see Figure 3a and 3b. Moreover, as shown by Figure 3c and 3d, for the weighted-median model and the bounded-confidence model, with fixed network sizes and link densities, the likelihoods of reaching consensus increase as the networks become less clustered (corresponding with larger parameter β). For the other models in comparison, network features such as size and clustering coefficient play no role in determining the probability of reaching consensus. Instead, these models predict either almost-sure consensus or almost-sure disagreement, as shown in Figure 3b.

As we see in the section, our weighted-median model exhibits various desired realistic features of opinion dynamics, which the other models in comparison fail to fully capture. The physical intuition behind it is that, the other models in comparison are all based on the weighted-averaging opinion updates, which implies overly large "attractive forces" between distant opinions driving the system to consensus. In order to resist such overly strong tendency to consensus, these models have to either introduce additional individual-level dynamics independent of network structure or artificially truncate the attraction of

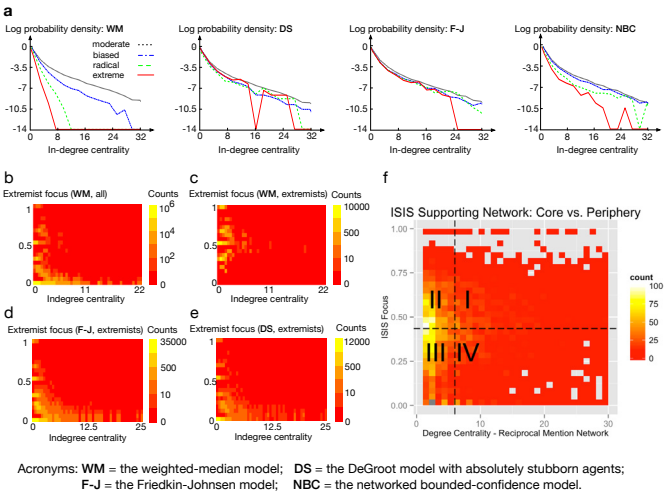


Fig. 2. Distributions of opinions with different degrees of extremeness in influence networks at final steady states. Panel a shows different models' predictions of the in-degree centrality distributions for individuals with various levels of extremeness at the steady states. Panel b and c are the two-dimension distributions over the extremist focus and in-degree centrality for the entire population and the extremists respectively, predicted by the weighted-median model. Panel d and e are the corresponding two-dimension distributions for extremists predicted by other models. Panel f is Figure 5 in a previous paper (Benigni et al., 2017), licensed under Creative Commons CC0 public domain dedication (CC0 1.0). This figure plots the empirical distribution of randomly sampled Twitter users over in-degree and the ISIS focus (the ratio of social neighbors who support the ISIS terrorists).

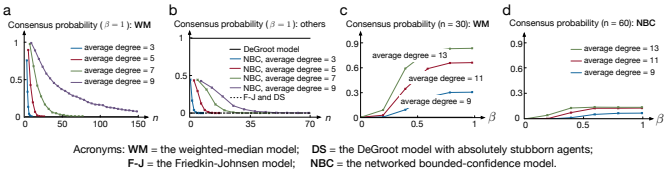


Fig. 3. Different models' predictions on the effects of network size and clustering on the probability of reaching consensus. For each model, the consensus probabilities are estimated via 5000 independent simulations on randomly generated Watts-Strogatz small-world networks. This generative random network model has three main parameters: the network size n, the average degree, and the rewiring probability β. The smaller β, the more clustered the network being generated.

distant opinions. By resolving the unrealistic feature of overly large attractions of distant opinions, in the weighted-median model, the effects of some delicate network structures naturally emerge. In the next section, we present some analytical results on how certain delicate network structures shape the behavior of the weighted-median opinion dynamics.

4. ANALYTICAL RESULTS

In this section, we characterize the equilibrium set of the weighted-median opinion dynamics and establish its almost-sure convergence, as well as the conditions for asymptotic con-

sensus and persistent disagreement. We refer to the technical report (Mei et al., 2020) for all the proofs.

The dynamical behavior of the weighted-median model depends on the following two important graph-theoretic concepts: *cohesive sets* and *decisive links*. The former was first proposed in (Morris, 2000) while the latter is novel. They are formally defined as follows.

Definition 2. (Cohesive set). Given a $n \times n$ row-stochastic influence matrix W, a set $M \subset \{1, \dots, n\}$ of nodes is a cohesive set on the influence network $\mathcal{G}(W)$ if, for any $i \in M$, $\sum_{j \in M} w_{ij} \geq 1/2$. Moreover, if M is cohesive and there does not exist any $i \in \{1, dots, n\} \backslash M$ such that $M \cup \{i\}$ is cohesive, then M is s maximal cohesive set in $\mathcal{G}(W)$.

Definition 3. (Decisive links). Given a row-stochastic matrix W and the associated influence network $\mathcal{G}(W)$, let $N_i = \{j \in \{1, \dots, n\} \,|\, w_{ij} \neq 0\}$. A link (i, j) is decisive if there exists a subset $\theta \subseteq N_i$ satisfying: 1) $j \in \theta$; 2) $\sum_{k \in \theta} w_{ik} > 1/2$; 3) $\sum_{k \in \theta \backslash \{j\}} w_{ik} < 1/2$. Otherwise, (i, j) is indecisive.

Given these definitions, the main analytical results regarding the weighted-median opinion dynamics are stated as follows.

Theorem 4. (Equilibrium set). Given an influence network $\mathcal{G}(W)$, where W is a row-stochastic matrix, $x^* \in \mathbb{R}^n$ is an equilibrium of the weighted-median opinion dynamics defined in Definition 1 if and only of, for any $y \in \mathbb{R}$, the nodes set $\{i \in \{1, \dots, n\} \,|\, x_i^* \geq y\}$ is either empty or a maximal cohesive set in $\mathcal{G}(W)$.

Theorem 5. (Convergence and phase transition). Consider the weighted-median opinion dynamics given by Definition 1 on an influence network $\mathcal{G}(W)$, where W is row-stochastic. Denote by $\mathcal{G}_{\text{decisive}}(W)$ the subgraph of $\mathcal{G}(W)$ with all the indecisive out-links removed. The following statements hold,

(1) for any initial condition $x_0 \in \mathbb{R}^n$, the solution $x(t)$ almost surely converges to an equilibrium x^* in finite time;
(2) if the only maximal cohesive set of $\mathcal{G}(W)$ is V, then, for any initial condition $x_0 \in \mathbb{R}^n$, the solution $x(t)$ almost surely converges to a consensus state;
(3) if the graph $\mathcal{G}(W)$ has a maximal cohesive set $M \neq V$, then there exists a subset of initial conditions $X_0 \subseteq \mathbb{R}^n$ with non-zero Lebesgue measure in $\mathbb{R}^n$ such that, for any $x_0 \in X_0$, there is no update sequence along which the solution converges to consensus; and
(4) If $\mathcal{G}_{\text{decisive}}(W)$ has no globally reachable node, then, for any initial condition $x_0 \in \mathbb{R}^n$, the solution $x(t)$ almost surely reaches a non-consensus fixed point in finite time.

From the analytical results above, one could observe that cohesive sets and decisive links, as some delicate structure of the influence networks, are crucial in determining the dynamical behavior of the weighted-median opinion dynamics. Compared with the network connectivity, which determines the consensus-disagreement phase transition in DeGroot model, cohesive sets and decisive links are more robust to uncertainty or perturbation of the influence networks. The addition/removal of a link with arbitrarily small weight could completely change the connectivity property of a graph and thus change the long-term behavior of the DeGroot model. However, generically, such perturbations has no qualitative effect on the weighted-median model since a link with very small weight is often indecisive.

5. CONCLUSION

In this paper, we point out that most of the existing opinion dynamics models are based on an unrealistic micro-foundation: the weighted-averaging opinion updates. By resolving this unrealistic feature, we propose a new microscopic mechanism of opinion dynamics, i.e., the weighted-median mechanism. The new model, despite its simplicity, reflects various real-world features of opinion dynamics, while some other widely-studied models fail to. The weighted-median model also exhibits richer dynamical behavior that depends on more delicate and robust network structures. Our work opens up a new research direction and inspires researchers to think about some fundamental problems about opinion dynamics.

REFERENCES

Acemoglu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. *Mathematics of Operation Research*, 38(1), 1–27. doi: 10.1287/moor.1120.0570.

Benigni, M.C., Joseph, K., and Carley, K.M. (2017). Online extremism and the communities that sustain it: Detecting the isis supporting community on twitter. *PloS one*, 12(12), e0181405.

Bland, M. (2015). *An Introduction to Medical Statistics*. Oxford University Press.

DeGroot, M.H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121. doi: 10.1080/01621459.1974.10480137.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

French Jr., J.R.P. (1956). A formal theory of social power. *Psychological Review*, 63(3), 181–194. doi:10.1037/h0046123.

Friedkin, N.E. and Johnsen, E.C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4), 193–206. doi:10.1080/0022250X.1990.9990069.

Hare, A.P. (1952). A study of interaction and consensus in different sized groups. *American Sociological Review*, 17(3), 261–267. doi:10.2307/2088071.

Hegselmann, R. and Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulations. *Journal of Artificial Societies and Social Simulation*, 5(3). URL `http://jasss.soc.surrey.ac.uk/5/3/2.html`.

Matz, D.C. and Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology*, 88(1), 22–37. doi: 10.1037/0022-3514.88.1.22.

McCauley, C. and Moskalenko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence*, 20(3), 415–433. doi: 10.1080/09546550802073367.

Mei, W., Bullo, F., Chen, G., Hendrickx, J., and Dörfler, F. (2020). Rethinking the micro-foundation of opinion dynamics: Rich consequences of an inconspicuous change. *arXiv preprint arXiv:1909.06474*.

Morris, S. (2000). Contagion. *The Review of Economic Studies*, 67(1), 57–78. doi:10.1111/1467-937X.00121.

Vande Kerckhove, C., Martin, S., Gend, P., Rentfrow, P.J., Hendrickx, J.M., and Blondel, V.D. (2016). Modelling influence and opinion evolution in online collective behaviour. *PLoS One*, 11(6), 1–25. doi:10.1371/journal.pone.0157685.