The Normalized Friedkin-Johnsen Model (A Work-in-progress Report)

Xi Chen, Jefrey Lijffijt, and Tijl De Bie

IDLab, Ghent University, Ghent, Belgium
{ xi.chen, jefrey.lijffijt, tijl.debie}@ugent.be

Abstract. The formation of opinions in a social context has long been studied by sociologists. A well-known model is due to Friedkin and Johnsen (further referenced as the *FJ model*), which assumes that individuals hold an immutable *internal opinion* while they *express* an opinion that may differ from it but is more in agreement with the expressed opinions of their friends. Formally, the *expressed opinion* is modeled as the weighted average of the individual's internal opinion and the expressed opinions of their neighbors. This model has been used in recent research originating from the computer science community, studying the origination and reduction of conflict on social networks, how echo chambers arise and can be burst, and more.
Yet, we argue that the FJ model in its elementary form is not suitable for some of these purposes. Indeed, the FJ model entails that the more friends one has, the less one's internal opinion matters in the formation of one's expressed opinion. Arguing that this may not be realistic, we propose a modification of the FJ model that normalizes the influence of one's friends and keeps the influence of one's internal opinion constant. This normalization was in fact suggested by Friedkin and Johnsen, but it has been ignored in much of the recent computer science literature.
In this work-in-progress report, we present the details of the normalized model, and investigate the consequences of this normalization, both theoretically and empirically.

Keywords: Social networks · opinion formation · conflict · controversy.

1 Introduction and Motivation

How people form their opinions has long been the subject of research in the field of social sciences [6, 7]. More recently, such models for opinion formation and dynamics (e.g., [2]) have been used by computer scientists and computational social scientists to study how to quantify and control notions of controversy, disagreement, polarization and conflict on social networks[5, 12], e.g. by manipulation the opinions of a small set of particular individuals, or by locally changing the network structure [15, 9, 14]. Opinion formation models serve as the fundamental part of these studies.

Background. Many opinion formation models have been proposed and studied based on the influence through social interactions [1, 11, 4, 16, 8, 10]. The

Friedkin-Johnsen (FJ) Model [7] is a very popular extension of the DeGroot's Model [6] that is used often [9, 14, 3]. In the model, individuals are assumed to have two types of opinions: the *internal opinion* and the *expressed opinion*. The internal opinions are assumed to be immutable, and represent individuals' innate opinion about matters. In the absence of any influence by others, this is the opinion an individual would express. However, the actual expressed opinion *will* be affected by one's friends/neighbors (e.g. due to a desire for social acceptance), and is modeled as the weighted average of the individual's own internal opinion and their neighbors' expressed opinions. The opinions are formed through continuous averaging in the model. Later on, the expressed opinion vector in FJ Model was interpreted as the Nash equilibrium in the social game of opinion formation, in which people get social costs as payoffs [2].

Motivation. A feature of the FJ model is that an individual's internal opinion matters less the more friends that individual has (or the stronger those friendships are). This may not be realistic, and for this reason Friedkin and Johnsen themselves suggested that the influence of a friend's expressed opinion on one's own expressed opinion should be normalized. This would ensure that the relevance of one's internal opinion is independent on the number of friends and strength of these friendships.

Yet, this normalization, which is important in particular in studies that investigate how to engineer the connectivity of the network so as to achieve a certain goal (e.g. reducing some measure of conflict, maximizing some measure of influence, etc.), is often ignored in recent work.

In this short work-in-progress paper, we study the relevance of the normalization. First, we make the normalization explicit by proposing a minor variant of the FJ model: the Normalized Friedkin-Johnson (NFJ) model. Then, we investigate theoretically how NFJ model differs qualitatively from the FJ model. In particular, we focus on a recently discovered conservation law of conflict [3], which stated that for opinions that follow the FJ model, the sum of a measure for internal conflict, for external conflict, and controversy sums to a constant. We show that this conservation law no longer holds under the NFJ model. Finally, we investigate empirically how the NFJ and FJ models yield different quantifications for important measures of conflict.

2 The Normalized Friedkin-Johnsen model, and a theoretical analysis

This section contains the details of the proposed model, but first we need to introduce some notation.

Notation. Let $G = (V, E, w)$ be a network, where $V = \{1, \ldots, n\}$ is the set of nodes, $E \in V \times V$ is the set of $m = |E|$ edges, and w is a weight function mapping an edge $e \in E$ onto its weight $w(e) \geq 0$. We denote with $\mathbf{W}$ the weighted adjacency matrix (with zero diagonal), defined by $w_{ij} = w(i, j)$ iff $\{i, j\} \in E$ and $w_{ij} = 0$ otherwise. With $N(i)$ we denote the set of neighboring nodes of node i: $N(i) \triangleq \{j \in V | (j,i) \in E\}$ (i.e., node j is a friend who has influence on node i in social networks). Let $\mathbf{e}$ denote the vector of ones of appropriate size. Furthermore, let $\mathbf{d} \triangleq \mathbf{W}^T\mathbf{e}$ denote the vector containing the weighted (in-)degrees of all nodes, and $\mathbf{D} \triangleq \text{diag}(\mathbf{d})$ the diagonal degree matrix. Then the Laplacian matrix is defined as $\mathbf{L} \triangleq \mathbf{D} - \mathbf{W}$. Note here the notations are related to in-degrees of nodes in directed networks, and they correspond to degrees (either in-degree or out-degree) for undirected networks.

2.1 The Normalized Friedkin-Johnsen model

Before discussing the NFJ model, we first discuss two logical predecessors: a model due to DeGroot, and the vanilla FJ model.

DeGroot's model [6] formalizes opinion formation as a repeated averaging process of one's opinion with one's neighbors. In the model, every person $i \in V$ updates his/her opinion $s_i(t+1)$ at time $t+1$ as the weighted sum of their own opinion (with weight w_{ii}) and those of the neighbours (with weight w_{ij} for neighbor j) at time t. Note that w_{ii} is independent from any w_{ij}, and represents the node's believe in its own opinion. Given an undirected weighted graph $G = (V, E, w)$, the updating rule is defined as:

$$s_i(t+1) = \frac{w_{ii}s_i(t) + \sum_{j \in N(i)} w_{ij}s_j(t)}{w_{ii} + \sum_{j \in N(i)} w_{ij}}. \tag{1}$$

In 1990, Friedkin and Johnsen extended DeGroot's model to have two different kinds of opinions [7]: a fixed internal opinion s_i, which is private to each individual, and a public expressed opinion z_i. The expressed opinions are the weighted sum of the node's own internal opinion and the expressed opinions of the neighbors:

$$z_i = \frac{w_{ii}s_i + \sum_{j \in N(i)} w_{ij}z_j}{w_{ii} + \sum_{j \in N(i)} w_{ij}}. \tag{2}$$

Expressed in matrix-vector notation, and with $w_{ii} = 1$ (a common assumption in the literature), this equation is solved by (3) below at equilibrium [2]:

$$\mathbf{z} = (\mathbf{L} + \mathbf{I})^{-1}\mathbf{s}. \tag{3}$$

In the proposed NFJ model, we consider that the influence from neighbors should be normalized by the number of neighboring nodes or the total strength of the incident edges, because people's internal opinions will not be less important if they have more friends. We discuss it for directed graphs – undirected graphs can be regarded as a special case (note that most of the existing literature focuses on undirected networks only). In directed networks, only the incoming edges contribute to the opinion formation process. We consider edge $(i,j) \in E$ as the edge from node i to node j, so the element $d_i \triangleq \sum_{j \neq i} w_{ji}$ of $\mathbf{d}$ is the (weighted) in-degree of node i.

In the proposed NFJ model, the expressed opinion is updated as follows:

$$z_i = \begin{cases} s_i, & \text{if } d_i = 0 \\ \frac{a s_i + \frac{\sum_{j \in N(i)} w_{ji} z_j}{d_i}}{a+1}, & \text{otherwise.} \end{cases} \tag{4}$$

Thus, in the NFJ model, it is assumed that each node puts the same weight $w_{ii} = a$ (instead of $w_{ii} = 1$) on its internal opinion, independently of the network weights – i.e. independently of the number and weights of incoming edges. Note that when $d_i = 1$, the node follows exactly the updating rule in the vanilla FJ model. Assuming that $d_i \neq 0$ for all i, the set of linear Equations (4) is solved by Equation (5) below, where $\mathbf{K} = \frac{1}{a}\mathbf{D}^{-1}\mathbf{L}^T$ is a normalized Laplacian:

$$\mathbf{z} = (\mathbf{K} + \mathbf{I})^{-1}\,\mathbf{s}. \tag{5}$$

2.2 Implications of the normalization on the quantification of conflict in networks

Based on FJ Model, several conflict measures have been proposed in the recent computer science literature. Four measures in particular were highlighted in [3]:

- Internal Conflict ic $(= \sum_i (s_i - z_i)^2)$ quantifies the extend to which individuals' internal and expressed opinions differ.
- External Conflict ec $(= \sum_{(i,j)\in E} w_{ij} (z_i - z_j)^2)$ quantifies the extend to which the expressed opinions of neighbors are in disagreement with each other.
- Controversy c $(= \sum_i z_i^2)$ does not depend on the network structure, and simply quantifies how much the opinion varies across the individuals in the network.
- Resistance r $(= \sum_i s_i z_i)$ is the inner product between expressed and internal opinion vectors, and also the sum of external conflict and controversy.

Matrix expressions for these quantities in terms of $\mathbf{s}$ and $\mathbf{z}$ are shown in Table 1. These measures were proposed for undirected networks. It was shown in [3] that

Table 1: Conflict Measures based on FJ Model

Name	$\mathbf{z}$	$\mathbf{s}$
ic	$\mathbf{z}^T\mathbf{L}^2\mathbf{z}$	$\mathbf{s}^T(\mathbf{L}+\mathbf{I})^{-1}\mathbf{L}^2(\mathbf{L}+\mathbf{I})^{-1}\mathbf{s}$
ec	$\mathbf{z}^T\mathbf{L}\mathbf{z}$	$\mathbf{s}^T(\mathbf{L}+\mathbf{I})^{-1}\mathbf{L}(\mathbf{L}+\mathbf{I})^{-1}\mathbf{s}$
c	$\mathbf{z}^T\mathbf{z}$	$\mathbf{s}^T(\mathbf{L}+\mathbf{I})^{-2}\mathbf{s}$
r	$\mathbf{z}^T\mathbf{s}$	$\mathbf{s}^T(\mathbf{L}+\mathbf{I})^{-1}\mathbf{s}$

the first three together give rise to a conservation law of conflict, indicating that reducing one kind of conflict implies that another must be increased. Formally:

$$ic + 2ec + c = \mathbf{s}^T\mathbf{s}. \tag{6}$$

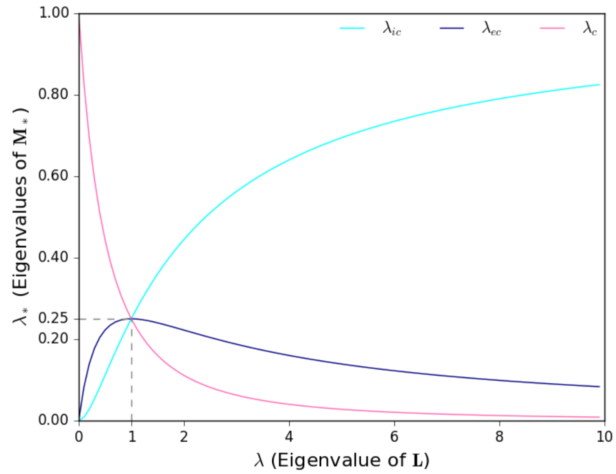


Fig. 1: Conservation Law

Note that the expressions in the right column of Table 1 are all quadratic forms $\mathbf{s}^T\mathbf{M}_*\mathbf{s}$ for some middle matrix $\mathbf{M}_*$ that depends on the conflict measure of interest $* \in \{ic, ec, c, r\}$ (e.g., $(\mathbf{L}+\mathbf{I})^{-1}\mathbf{L}(\mathbf{L}+\mathbf{I})^{-1}$ for ec). The middle matrices are all scalar functions of $\mathbf{L}$, such that they share the same eigenvectors, and their eigenvalues can be expressed as a scalar function of the eigenvalues of $\mathbf{L}$. Figure 1 illustrates this relation, with λ representing an eigenvalue of $\mathbf{L}$ while λ_* is the eigenvalue of one of the middle matrices. The conservation law is reflected in a similar relation amongst the eigenvalues of the three middle matrices, and this for any eigenvalue λ of the Laplacian.

This figure also illustrates that for a given $\mathbf{s}$, a conflict measure will be larger if $\mathbf{s}$ aligns better with an eigenvectors of $\mathbf{L}$ for which λ_* is larger. Arguably the most interesting measure is ec, which increases at first and then decreases when $\mathbf{s}$ becomes less smooth (i.e., with larger eigenvalues)[1]. We will discuss this in greater detail with experimental results in the next section.

As a first theoretical analysis of the NFJ model as compared with the FJ model, it is interesting to investigate whether this conservation law still holds in the NFJ model. We start from the conflict measures, and then investigate this only for directed networks. Referring to their definitions based on the FJ model, we define the three conflict measures in the conservation law as follows:

$$ic = \sum_i (s_i - z_i)^2, \quad ec = \frac{1}{a}\sum_{i,j}\frac{w_{ij}}{d_j}(z_i - z_j)^2, \quad c = \sum_i z_i^2$$

In the NFJ model, the conflict measures are very similar to the ones in [3] (i.e., ic and c stay the same). However, ec is different because the importance of the

[1] Here smooth/low-frequency represents that the close-by nodes hold similar opinions - corresponding to smaller eigenvalus, and high-frequency means nodes differ more with nodes around in their opinions, which corresponds with larger eigenvalues.

opinion differences over existing edges is also normalized by the in-degrees of the incident nodes. Based on Equation (5), the three measures in the new model are expressed in matrix-vector form as in Table 2, where $\mathbf{N}_{ec}$ is

Table 2: Conflict Measures based on NFJ Model

Name	$\mathbf{z}$	$\mathbf{s}$
ic	$\mathbf{z}^T\mathbf{K}^T\mathbf{K}\mathbf{z}$	$\mathbf{s}^T\left(\mathbf{K}^T+\mathbf{I}\right)^{-1}\mathbf{K}^T\mathbf{K}\left(\mathbf{K}+\mathbf{I}\right)^{-1}\mathbf{s}$
ec	$\mathbf{z}^T\mathbf{N}_{ec}\mathbf{z}$	$\mathbf{s}^T\left(\mathbf{K}^T+\mathbf{I}\right)^{-1}\mathbf{N}_{ec}\left(\mathbf{K}+\mathbf{I}\right)^{-1}\mathbf{s}$
c	$\mathbf{z}^T\mathbf{z}$	$\mathbf{s}^T\left(\mathbf{K}^T+\mathbf{I}\right)^{-1}\left(\mathbf{K}+\mathbf{I}\right)^{-1}\mathbf{s}$

$$\mathbf{N}_{ec} = \frac{1}{a}\operatorname{diag}\left(\mathbf{D}^{-1}\mathbf{W}\mathbf{e} + \mathbf{W}\mathbf{D}^{-1}\mathbf{e}\right) - \frac{1}{a}\left(\mathbf{D}^{-1}\mathbf{W} + \mathbf{W}\mathbf{D}^{-1}\right). \tag{7}$$

The definition of ec in the new model is inspired by the conservation law of conflict. After finding that the conservation law no longer holds in the NFJ model, we introduce an additional term, denoted as x shown in Equation (8) below, such that the law can be restored. It is equivalent to finding two matrices $\mathbf{N}_{ec}$ and $\mathbf{N}_x$, which sum to $\mathbf{K}^T + \mathbf{K}$.

$$ic + ec + c + x = \mathbf{s}^T\mathbf{s} \tag{8}$$

So we have $\mathbf{N}_{ec}$ as in Equation (7), and $\mathbf{N}_x$ below

$$\mathbf{N}_x = \frac{1}{a}\operatorname{diag}\left(\mathbf{D}^{-1}\mathbf{W}\mathbf{e} - \mathbf{W}\mathbf{D}^{-1}\mathbf{e}\right), \tag{9}$$

$$x = \mathbf{z}^T\mathbf{N}_x\mathbf{z} = \frac{1}{a}\sum_i z_i^2 \sum_{j\neq i}\left(\frac{w_{ji}}{d_i} - \frac{w_{ij}}{d_j}\right). \tag{10}$$

If x cannot be interpreted as a relevant measure of conflict, it can be seen as an opportunity for eliminating conflict: it is then conceivable that the network can be edited (e.g. by adding or removing edges, or by changing weights) so as to reduce all of ic, ec, and c while increasing x. I.e., the sum of the three conflict measures can be minimized by maximizing x. According to Equation (8), x can be expressed as in Equation (10). It shows that the network edits for conflict optimization (i.e., maximizing x) should consider both how opinionated nodes are (i.e., the values of z_i^2) and the importance of the node's influence on all its neighbors (i.e., the value of $\sum_{j\neq i}\frac{w_{ij}}{d_j}$ since $\sum_{j\neq i}\frac{w_{ji}}{d_i} = 1$). Meanwhile, when it comes to comparing the amount of conflict between networks of similar sizes, x indicates that the more opinionated nodes are of minor importance in influencing their neighbors (i.e., small $\sum_{j\neq i}\frac{w_{ij}}{d_j}$), the less total conflict (i.e., $ic+ec+c$) there will be. The interpretation of x, and on how it can be maximized, are subject of our current research.

3 Discussion and Experiments

This section discusses the difference of the NFJ model to the original model, using synthetic as well as real-world networks, which are of varying sizes.

3.1 Opinion Formation

We start from a very simple network as shown in Figure 2, and assign each node with internal opinions where green means $s_i = 1$ and red represents $s_i = -1$. According to Table 3, node 1 and node 7, which are the centers of the two star-subgraphs, have expressed opinions opposite to the internal ones in the old model, while they remain on their "original side" in the new model. It is clear that the normalization can have a big impact in this opinion formation model, and it corresponds to the suggested assumption in the original FJ Model [7] as $\sum_j w_{ij} = 1$ and $w_{ij} \in [0, 1]$. Surprisingly, it is usually neglected in works based on this model.

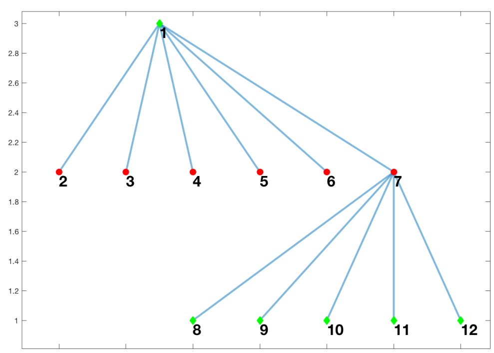


Fig. 2: Network Example 1

Not only in the static case, the NFJ model ensures people value their own internal opinions with an importance that is independent of the environment, such that the number of friends or the strength of these friendships will not affect people's adherence to their own opinion.

3.2 Quantifying Conflict

In addition to the conflict eliminator in the conservation law, we will give evidence that this normalized model is different from the original one in terms of conflict measures (i.e., here we focus on external conflict ec as it was arguably the most interesting measure [3]). This model seems to be better as it preserves

Table 3: Expressed Opinions at Equilibrium ($a = 1$)

Node	1	2	3	4	5	6
s	1	-1	-1	-1	-1	-1
z_{FJ}	-0.27	-0.64	-0.64	-0.64	-0.64	-0.64
z_{NFJ}	0.33	-0.33	-0.33	-0.33	-0.33	-0.33
Node	7	8	9	10	11	12
s	-1	1	1	1	1	1
z_{FJ}	0.27	0.64	0.64	0.64	0.64	0.64
z_{NFJ}	-0.33	0.33	0.33	0.33	0.33	0.33

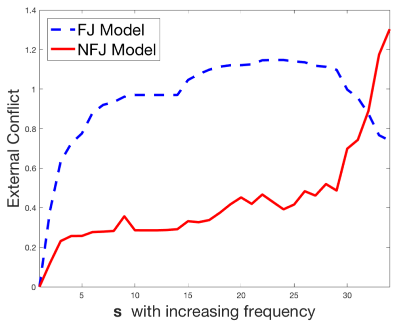
the controversial discussion within social networks, instead of "diminishing" it with too much opinion averaging. We consider different sizes of synthetic random networks and real-world social networks: 1) the Karate network of friendships between 34 members [17]; 2) a Watts-Strogatz random network with the small world property of 500 nodes; 3) and a real-world Facebook social network containing friend circles [13] of 4039 nodes.

In the original FJ Model, external conflict increases first and then it decreases slowly as the eigenvalues of the Laplacian matrix $\mathbf{L}$ increases, shown in Figure 1. In other words, when the vector of internal opinions $\mathbf{s}$ aligns with the eigenvectors of increasing frequency, ec reaches the maxima at a certain point (i.e., $\lambda = 1$). However, the higher the frequency on the graph for $\mathbf{s}$, the more the conflicts there should be in the network because this is how controversy arised. It shows the real conflict between people holding "opposite" (potentially differing) opinions, because high-frequency $\mathbf{s}$ means more differences over existing edges.

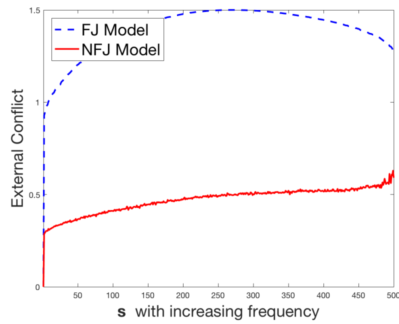
In the experiments, we use the eigenvectors of the network Laplacian matrix $\mathbf{L}$ as the internal opinion vector $\mathbf{s}$, which correspond to different frequencies (i.e., eigenvalues). In order to make a clearer comparison between both models, we scale the magnitude of the edge weights. We can see that the old model has decreased amount of conflict for high-frequency $\mathbf{s}$ since every node is influenced by more neighbors holding opposite opinions. This is due to too much opinion averaging.

On the contrary, from Figure 3, we can see that the high-frequency internal opinions correspond to larger external conflict if we use the new model. This is because the NFJ model limits the overall amount of external influence by the normalization, thus the opinions are not over-averaged and the conflict measure reflects the "real" (i.e., internal) opinion divergence to some extend. Note that conflict is what exists between people holding opposite opinions internally. It means even if they express themselves differently, one of them should realize the other is on the same side with him/her internally. Therefore, the more people differ from their neighbors on the graph in terms of internal opinions (i.e., $\mathbf{s}$ shows higher frequency), the more conflict there should be. It is consistent with the results of our proposed NFJ model.

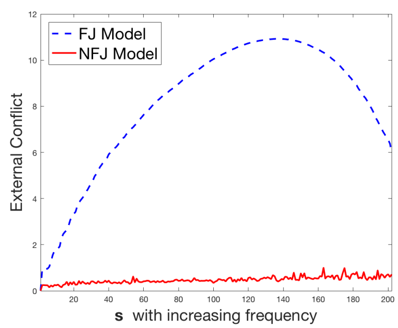
This report only presents a first look at the normalized Friedkin-Johnsen (NFJ) Model, and there are a lot of interesting tasks to be done in the near



(a) Karate ($n = 34$, $m = 78$)



(b) small world random network ($n = 500$, $m = 1500$)



(c) Facebook Network ($n = 4039$, $m = 88234$)

Fig. 3: Conflict Comparison in Networks

future. For example, the evolution of opinion dynamics, network conflict risk problems under the new model, the discussion on the parameter a (i.e., the self-appraisal [12]), networks with different type of nodes (e.g., introducing stubborn nodes who only express their own internal opinions), and so on.

Also, instead of doing nomalization, which discounts the neighboring influences, we can switch the sign of the moderation. In other words, one instance could be that two very opinionated people who hate each other will never moderate the opinions of the other person, on the contrary, their opinions will be reinforced through the connection. Therefore, it leads to a non-linear model of opinion formation. Another study direction is considering higher dimensions of opinions because different issues do not necessarily correspond to different social networks. People within a social network communicate about various issues, and their attitudes on one issue may have influence on other issues, which means minimizing conflict on one issue might actually increase conflict on another. A higher dimensional opinion vector seems to be closer to people's daily life and is more interesting for future study.

Acknowledgements. This work was supported by the ERC under the EU's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agr. no. 615517, FWO (project no. G091017N, G0F9816N), the EU's Horizon 2020 research and innovation programme and the FWO under the Marie Sklodowska-Curie Grant Agr. no. 665501.

References

1. Berger, R.L.: A necessary and sufficient condition for reaching a consensus using degroot's method. Journal of the American Statistical Association **76**(374), 415–418 (1981)
2. Bindel, D., Kleinberg, J., Oren, S.: How bad is forming your own opinion? Games and Economic Behavior **92**, 248–265 (2015)
3. Chen, X., Lijffijt, J., De Bie, T.: Quantifying and minimizing risk of conflict in social networks. In: Proceedings of 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2018)
4. Cox, J.T.: Coalescing random walks and voter model consensus times on the torus in zd. The Annals of Probability pp. 1333–1366 (1989)
5. Dandekar, P., Goel, A., Lee, D.T.: Biased assimilation, homophily, and the dynamics of polarization. Proceedings of the National Academy of Sciences **110**(15), 5791–5796 (2013)
6. DeGroot, M.H.: Reaching a consensus. Journal of the American Statistical Association **69**(345), 118–121 (1974)
7. Friedkin, N.E., Johnsen, E.C.: Social influence and opinions. Journal of Mathematical Sociology **15**(3-4), 193–206 (1990)
8. Galam, S.: Minority opinion spreading in random geometry. The European Physical Journal B-Condensed Matter and Complex Systems **25**(4), 403–406 (2002)
9. Gionis, A., Terzi, E., Tsaparas, P.: Opinion maximization in social networks. In: Proceedings of the 2013 SIAM International Conference on Data Mining. pp. 387–395. SIAM (2013)
10. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence: Models, analysis and simulation. Journal of Artificial Societies and Social Simulation **5**, 1–24 (2002)
11. Holley, R.A., Liggett, T.M.: Ergodic theorems for weakly interacting infinite systems and the voter model. The annals of probability pp. 643–663 (1975)
12. Jia, P., MirTabatabaei, A., Friedkin, N.E., Bullo, F.: Opinion dynamics and the evolution of social power in influence networks. SIAM review **57**(3), 367–397 (2015)
13. Leskovec, J., Mcauley, J.J.: Learning to discover social circles in ego networks. In: Advances in neural information processing systems. pp. 539–547 (2012)
14. Matakos, A., Terzi, E., Tsaparas, P.: Measuring and moderating opinion polarization in social networks. Data Mining and Knowledge Discovery **31**(5), 1480–1505 (2017)
15. Musco, C., Musco, C., Tsourakakis, C.E.: Minimizing polarization and disagreement in social networks. arXiv preprint arXiv:1712.09948 (2017)
16. Sznajd-Weron, K., Sznajd, J.: Opinion evolution in closed community. International Journal of Modern Physics C **11**(06), 1157–1165 (2000)
17. Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of anthropological research **33**(4), 452–473 (1977)