



# Bounded Confidence Gossip Algorithms for Opinion Formation and Data Clustering

Nguyen, Linh Thi Hoai

Wada, Takayuki

Masubuchi, Izumi

Asai, Toru

Fujisaki, Yasumasa

---

(Citation)

IEEE Transactions on Automatic Control, 64(3):1150-1155

(Issue Date)

2018-06-01

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90007329>



# Bounded Confidence Gossip Algorithms for Opinion Formation and Data Clustering

Linh Thi Hoai Nguyen, Takayuki Wada, *Member, IEEE*,  
 Izumi Masubuchi, *Member, IEEE*, Toru Asai, *Member, IEEE*,  
 and Yasumasa Fujisaki, *Member, IEEE*

**Abstract**—This paper presents a bounded confidence gossip algorithm for describing the process of opinion formation over a communication network. Each agent in the network keeps a time-varying opinion vector (or state) which represents its opinion about a set of matters. A common confidence threshold is set for all of the agents. The states of agents in the network will be updated time by time according to an iterative procedure: At each time, (i) one agent is chosen randomly, then it chooses one of its neighbors on the communication graph to contact with; (ii) they exchange their states; and (iii) if they have different states and the distance between their states is strictly smaller than the confidence threshold, they update their states as the average of the two. This algorithm converges almost surely to some equilibrium point such that any two adjacent agents either have the same state or have distinct states whose distance is no less than the confidence threshold. This is called the constant confidence threshold algorithm. An increasing confidence threshold algorithm, which repeats the constant confidence threshold algorithm several times with increasing confidence threshold, is also proposed. The algorithm is also convergent almost surely to some equilibrium point. Applicability of the method to clustering problems is shown through numerical examples.

**Index Terms**—opinion dynamics; bounded confidence; gossip algorithm; distributed computation; convergence analysis; data clustering;

## I. INTRODUCTION

This study is motivated by the current active research on public opinion dynamics and its vital importance in a variety of fields such as economics, management, social science, engineering and so on. Our objective is twofold. First is to get insights on the opinion forming processes in social science. The second is to apply the insights to management and engineering problems, especially problems concerning energy management system.

The understanding of the opinion dynamics and mechanism of forming different patterns of opinion formation—for instance, consensus, clustering, or fragmentation—can be achieved using mathematical models.

In the literature, many models addressing opinion dynamics have been introduced. The very early works on this problem are those of French [1] and Harary [2]. Then a lot of researches focused on consensus and how to reach it, e.g., [3]–[5]. Several other models based on binary values leading to uniformity of opinions are also studied in [6], [7]. However, the uniformity does not reflect fully the complex phenomena of public opinion formation, where, besides consensus, clustering and fragmentation are also observed frequently. Furthermore, binary-valued based models are also not valid for scenarios in social networks for which continuous spectrum of opinions is required.

In order to study the variety of opinion formation on real social networks, a few approaches are presented. One approach is to take stubbornness of agents into account [8], [9]. Another approach is to

introduce a notion of bounded confidence to characterize the opinion ranges in opinion forming processes of agents in the networks. The latter one is known to be suitable for studying the clustering phenomena in public opinion formation. We consider models following this approach in the present paper.

Krause [10] proposes a bounded confidence model. It is then further studied in [11]–[14]. In this model, agents synchronously update their opinions by averaging all opinions in their confidence bounds. However, the clock synchronization itself is a very difficult problem in dealing with multi-agent systems. Motivated by various random variations of the Krause model for opinion dynamics and gossip algorithm in an endogenously changing environment, in [15], Touri and Langbort propose a framework for the study of endogenously varying random averaging dynamics whose evolution suffers from history-dependent sources of randomness. It is shown that under general assumptions, such dynamics is almost surely convergent.

In [16], Deffuant et al. introduce another bounded confidence model which helps avoiding the difficulty of synchronous update. In the literature, it is often called DW model. It describes opinion dynamics for the evolution of continuous-valued opinions among a finite group of individuals. Each individual has  $[0, 1]$ -valued opinions. At each time step, two individuals are sampled. If their opinions differ at most by a confidence threshold, the two opinions become closer to each other through a process governed by a confidence factor. The model is then studied by means of numerical simulations and heuristic arguments without rigorous proofs.

The paper [13] gives a rigorous proof for the almost sure convergence of the general Deffuant model for scalar opinion with complete communication graph. The result is based on the important fact that the number of clusters is non-decreasing. However, this is no longer true for the vector opinion case.

This paper establishes a theoretical analysis for the DW models in the case where the confidence factor is set to  $1/2$  with vector opinions. More precisely, we consider a network of agents among which there is an undirected communication graph. Each agent keeps a real-valued vector representing its opinions about a set of issues. We first introduce the so-called constant confidence threshold algorithm for updating opinions of agents through communicating with each other: At each time, (i) one agent is chosen randomly, then it chooses one of its neighbor on the communication graph to contact with; (ii) they exchange their states; and (iii) if they have different states and the distance between their states is strictly smaller than the confidence threshold, they update their states as the average of the two. It is shown that, for any initial opinion profile, the model converges almost surely with respect to randomly chosen interacting pairs. Furthermore, any convergent opinion profile has the property that any two distinct opinions of adjacent agents differ at least by a confidence threshold. The algorithm results in the clustering among opinion profiles. We then propose another algorithm with increasing bounded confidence which represents a hierarchical clustering algorithm (see e.g., [17]).

The fact that the algorithms often reveal clustering phenomenon of opinion formation suggests that we may apply the algorithms to multi-dimensional data clustering problem. Clustering problems aim at partitioning a set of data objects into groups based on their similarity. Clustering can be considered to be one of the most important unsupervised learning problems. It is the main task of exploratory data mining, statistical data analysis, machine learning, pattern recognition, and information retrieval. Clustering techniques find their applications in a wide rank of fields such as management, economic, marketing, engineering, medicine and so on.

The preliminary versions of this paper appeared in conferences [19]–[21]. In our first paper [19] on bounded confidence

Linh Thi Hoai Nguyen, Takayuki Wada, and Yasumasa Fujisaki are with Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan  
 e-mail: {nth.linh, t-wada, fujisaki}@ist.osaka-u.ac.jp

Izumi Masubuchi is with Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada, Kobe 657-8501, Japan  
 email: msb@harbor.kobe-u.ac.jp

Toru Asai is with Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8603, Japan  
 email: asai@nuem.nagoya-u.ac.jp

via gossip algorithms, we consider a scalar opinion dynamics over complete communication graph and we propose a distributed data clustering algorithm based on bounded confidence via gossip algorithms. We consider a scalar opinion dynamics over general communication graph in [21]. We extend this result to a vector opinion case over complete communication graph [20]. As we described in the explanation on [13], proving almost sure convergence in the vector opinion case is completely different from that in the scalar opinion case. In fact, although we employ covariance of state variable for proving almost sure convergence in a vector case opinion [20], this approach is not enough to show it. In this paper, we therefore deal with vector opinion dynamics over general communication graph and we provide a complete proof on almost sure convergence.

*Paper Outline:* The organization of this paper is as follows. In Section II, we first introduce the opinion forming process on social networks. Then we propose two bounded confidence gossip algorithms for exchanging and updating states of agents in the network through pairwise interaction. The first algorithm with constant confidence threshold and the second with increasing confidence threshold. Section III is devoted to proving the almost sure convergence of the algorithms. Section IV gives some numerical examples to verify the theoretical results on clustering of opinion. We make some concluding remarks in Section V.

*Notation:* The set of natural numbers is denoted by  $\mathbb{N}$ . The  $d$ -dimensional real vector space is denoted by  $\mathbb{R}^d$ . The identity matrix of an appropriate size is denoted by  $\mathbf{I}$ , and the matrix whose elements are all ones is denoted by  $\mathbf{1}$ . Let  $\cdot^T$  be the transpose operator of a matrix or a vector. Let  $\|\cdot\|_2$  and  $\|\cdot\|_F$  be Euclidean norm and Frobenius norm, respectively of a vector or a matrix. The probability of an event is denoted by  $\mathbb{P}(\cdot)$ .

## II. OPINION FORMING ALGORITHMS

We consider a communication network  $G = (\mathcal{V}, \mathcal{E})$  where  $\mathcal{V} = \{1, 2, \dots, n\}$  denotes the set of agents and  $\mathcal{E}$  denotes the set of edges. We suppose that the communication graph is undirected. For simplicity, we assume that the graph is connected, otherwise we consider each connected component separately. Let the set of the agent  $i$ 's neighbor be  $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$ , and let its cardinality be  $|\mathcal{N}_i|$ . Because of the connectivity of communication graph, we have  $|\mathcal{N}_i| > 0$  for every  $i \in \mathcal{V}$ .

The state of each agent  $i \in \mathcal{V}$ , at time  $k$ , is  $\mathbf{x}_i(k) \in \mathbb{R}^d$ . Let its initial state be denoted by  $\mathbf{x}_i(1) \in \mathbb{R}^d$ . In the following, we discuss two bounded confidence gossip algorithms describing the opinion forming process of agents in the network through communication: one with constant confidence threshold and the other with increasing confidence threshold. Our goal is to reach consensus at all channels. In fact, this is useful for distributed data clustering.

Given a confidence threshold  $\delta$  which is a positive constant, the algorithm with constant confidence threshold is as follows.

Let  $X(k) = [\mathbf{x}_1(k) \mathbf{x}_2(k) \cdots \mathbf{x}_n(k)]^T$  be the system state at time  $k$ . That is, each row of  $X(k)$  contains the state of an agent at time  $k$ . Suppose that the probability for choosing any  $i \in \mathcal{V}$  at each step is uniform, that is,  $1/n$ . The probability for  $i$  to choose any of its neighbor to contact with is  $1/|\mathcal{N}_i|$ . Assume that, at time  $k$ , the edge  $(i, j) \in \mathcal{E}$  are chosen with probability  $(1/|\mathcal{N}_i| + 1/|\mathcal{N}_j|)/n$ , then we set

$$W(i, j, X(k)) = \begin{cases} W_{ij} & \text{if } 0 < \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 < \delta, \\ \mathbf{I} & \text{otherwise,} \end{cases} \quad (1)$$

where

$$W_{ij} = \mathbf{I} - \frac{1}{2}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \quad (2)$$

---

### Algorithm 1 Constant Confidence Threshold

---

```

1: procedure CONSTANT_CONFIDENCE( $X(1), \delta, \kappa$ )
    $\kappa$ : number of iterations
2:   for  $k = 1 : \kappa$  do
3:     An agent  $i$ 's clock ticks.
4:     The agent  $i$  contacts one of its neighbors, say  $j$ .
5:      $i$  and  $j$  exchange their states  $\mathbf{x}_i(k), \mathbf{x}_j(k)$ .
6:     if  $0 < \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 < \delta$  then
7:        $\mathbf{x}_i(k+1) \leftarrow \frac{\mathbf{x}_i(k) + \mathbf{x}_j(k)}{2}$ 
8:        $\mathbf{x}_j(k+1) \leftarrow \frac{\mathbf{x}_i(k) + \mathbf{x}_j(k)}{2}$ 
9:     else
10:       $\mathbf{x}_i(k+1) \leftarrow \mathbf{x}_i(k)$ 
11:       $\mathbf{x}_j(k+1) \leftarrow \mathbf{x}_j(k)$ 
12:     end if
13:      $\mathbf{x}_\ell(k+1) \leftarrow \mathbf{x}_\ell(k), \forall \ell \in \mathcal{V} \setminus \{i, j\}$ 
14:   end for
15: end procedure

```

---

with  $\mathbf{e}_i = [0 \cdots 0 \ 1 \ 0 \cdots 0]^T$  is a unit vector with the  $i$ th component equals to one.

Using the above notations, we can rewrite Algorithm 1 in compact form as

$$X(k+1) = W(i(k), j(k), X(k))X(k), \quad (3)$$

where  $i(k)$  and  $j(k)$  are random sequences which describe the selected agents.

Let us consider the following simple example to see how the algorithm works.

*Example 1:* Consider a system of three agents  $\mathcal{V} = \{1, 2, 3\}$  on a complete graph, that is,  $\mathcal{E} = \{(1, 2), (1, 3), (2, 3)\}$ , with an initial opinion profile  $X(0) = [0 \ 0.4 \ 1]^T$  and confidence threshold  $\delta = 0.7$ . If agent 1's clock ticks and agent 1 contacts agent 2, that is,  $W(0) = W_{12}$ , the system's opinion profile at the end of the first time slot is  $X(1) = [0.2 \ 0.2 \ 1]^T$  which is an equilibrium point of the system. On the other hand, suppose that agent 2's clock ticks and agent 2 contacts agent 3, that is,  $W(0) = W_{23}$ . This drives the state at the end of the first time slot to  $X(1) = [0 \ 0.7 \ 0.7]^T$ . Otherwise, if agent 3's clock ticks and agent 3 contacts agent 1, that is,  $W(0) = W_{13}$ , the system's opinion profile at the end of the first time slot remains as  $X(1) = [0 \ 0.4 \ 1]^T$ . Since the probability which these exists a time slot  $k$  such that  $W(k) = W_{12}$  or  $W(k) = W_{23}$  is 1, all the system state converges to  $[0.2 \ 0.2 \ 1]^T$  or  $[0 \ 0.7 \ 0.7]^T$  with probability 1. This phenomenon holds in a general case. We will show the almost sure convergence of all the system states in Section III.

The above example also illustrates the randomness of stationary opinion profiles achieved due to the randomness of choosing interaction pairs.

The increasing confidence threshold algorithm is described in Algorithm 2.

*Remark 1:* Algorithm 1 is quite simple. Nevertheless, when we choose a quite large  $\delta$ , we usually get total consensus. Conversely, if we choose a quite small  $\delta$ , we have so many clusters (see Example 3). Using Algorithm 2, we can gradually decrease the number of clusters to the amount we want. In other words, Algorithm 2 helps us to control the number of clusters. Of course, the computational cost for Algorithm 2 is more than that of Algorithm 1.

The following section gives a rigorous proof for the almost sure convergence of the constant confidence threshold algorithm.

**Algorithm 2** Increasing Confidence Threshold

---

```

1: procedure INCREASING CONFIDENCE( $X(1), d, \Delta\delta, \kappa$ )
   $d > 0$ : upper bound for confidence threshold
   $\Delta\delta$ : increment of confidence threshold  $\delta$ 
   $\kappa$ : numbers of iterative steps of the constant confidence threshold
  algorithm for each value of confidence threshold
2:    $\delta \leftarrow 0$ .
3:   while  $\delta \leq d - \Delta\delta$  do
4:      $\delta \leftarrow \delta + \Delta\delta$ .
5:     CONSTANT CONFIDENCE( $X(1), \delta, \kappa$ ).
6:      $X(1) \leftarrow X(\kappa)$ .
7:   end while
8: end procedure

```

---

## III. CONVERGENCE ANALYSIS

Let us first study the structure of equilibrium set. First we give the definition of equilibrium point.

*Definition 1:* A point  $X^* = [\mathbf{x}_1^* \mathbf{x}_2^* \dots \mathbf{x}_n^*]^T \in \mathbb{R}^{n \times d}$  is an equilibrium point of (3) if it satisfies

$$X^* = W(i, j, X^*)X^* \quad \forall (i, j) \in \mathcal{E}. \quad (4)$$

Then we have the following lemma.

*Lemma 1:* Any equilibrium point  $X^*$  of (3) has the form

$$\mathbf{x}_i^* = \mathbf{x}_j^* \quad \text{or} \quad \|\mathbf{x}_i^* - \mathbf{x}_j^*\|_2 \geq \delta \quad \forall (i, j) \in \mathcal{E}. \quad (5)$$

*Proof:* Let  $X^*$  be an equilibrium point of (3). Suppose that there exist  $i$  and  $j$  such that  $0 < \|\mathbf{x}_i^* - \mathbf{x}_j^*\|_2 < \delta$ . According to (4),

$$\begin{aligned}
(\mathbf{x}_i^*)^T &= \mathbf{e}_i^T X^* = \mathbf{e}_i^T W_{ij} X^* \\
&= \mathbf{e}_i^T \left( \mathbf{I} - \frac{1}{2}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \right) X^* \\
&= (\mathbf{x}_i^*)^T - \frac{1}{2} \mathbf{e}_i^T (\mathbf{e}_i - \mathbf{e}_j) (\mathbf{x}_i^* - \mathbf{x}_j^*)^T \\
&= \frac{1}{2} (\mathbf{x}_i^* + \mathbf{x}_j^*)^T
\end{aligned}$$

or equivalently  $\mathbf{x}_i^* = \mathbf{x}_j^*$ . This contradicts the above assumption. It follows that  $X^*$  must have the form (5).

Conversely, for any point  $X^*$  of the form (5),  $W_{ij} = \mathbf{I}$  for any way of choosing  $(i, j) \in \mathcal{E}$ , then condition (4) is always satisfied. This completes the proof.  $\blacksquare$

The structure of the equilibrium set suggests the following notion of convergence which will be studied in the remaining of this section.

*Definition 2:* Given a confidence threshold  $\delta > 0$ , an algorithm given by the iterative process over a communication network  $G = (\mathcal{V}, \mathcal{E})$

$$X(k+1) = F(X(k)) \quad \text{for } k = 1, 2, \dots$$

is said to achieve  $\delta$ -clustering convergence if the system starting from any initial state  $X(1)$  eventually converges to some state  $X^*$  of the form (5).

Let us define

$$f(X(k)) = \|X(k)\|_F^2,$$

where  $\|\cdot\|_F$  is the Frobenius norm of a matrix. We have the following property of the sequence  $\{f(X(k))\}$ .

*Lemma 2:* There exists  $\lim_{k \rightarrow \infty} f(X(k))$ .

*Proof:* If the update at time  $k+1$  happens between two agents  $i, j$  such that  $0 < \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 < \delta$ , then

$$\begin{aligned}
f(X(k)) - f(X(k+1)) &= \|\mathbf{x}_i(k)\|_2^2 + \|\mathbf{x}_j(k)\|_2^2 - 2 \left\| \frac{\mathbf{x}_i(k) + \mathbf{x}_j(k)}{2} \right\|_2^2 \\
&= \frac{\|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2^2}{2},
\end{aligned} \quad (6)$$

which is strictly positive. Otherwise  $f(X(k+1)) = f(X(k))$ . It follows that  $\{f(X(k))\}$  is a non-increasing sequence. This, together with the fact that  $\{f(X(k))\}$  is bounded from below by 0, establishes the lemma.  $\blacksquare$

For any  $\varepsilon < \delta$ , given  $k \in \mathbb{N}$ , we define

$$\begin{aligned}
\Omega_k &= \{ \forall (i, j) \in \mathcal{E}, \\
&\quad \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 \leq \varepsilon \text{ or } \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 \geq \delta \}, \\
\Omega &= \bigcup_{k \in \mathbb{N}} \Omega_k.
\end{aligned}$$

The following theorem is the key result for proving the almost sure convergence of the constant confidence threshold algorithm.

*Theorem 1:*

$$\mathbb{P}(\Omega) = 1.$$

*Proof:* Let  $\Omega_k^c$  be the complement of  $\Omega_k$ , that is

$$\Omega_k^c = \{ \exists (i, j) \in \mathcal{E}, \varepsilon < \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 < \delta \}.$$

We have that

$$\begin{aligned}
\mathbb{P}(\Omega) &= 1 - \mathbb{P}(\Omega^c) = 1 - \mathbb{P} \left( \left( \bigcup_{k \in \mathbb{N}} \Omega_k \right)^c \right) \\
&= 1 - \mathbb{P} \left( \bigcap_{k \in \mathbb{N}} \Omega_k^c \right).
\end{aligned} \quad (7)$$

We claim that the second term of the right hand side of (7) equals zero. We are done if we can prove this claim. Suppose otherwise that

$$\mathbb{P} \left( \bigcap_{k \in \mathbb{N}} \Omega_k^c \right) > 0. \quad (8)$$

We define  $A_k$  is the event that the pair  $(i, j) \in \mathcal{E}$  with  $\varepsilon < \|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|_2 < \delta$  is chosen at  $k$ .

Lemma 2 implies that for any sufficiently large  $k$ ,

$$f(X(k)) - f(X(k+1)) < \frac{\varepsilon^2}{2}. \quad (9)$$

Suppose that there exists some random finite instant  $k_0$  which is sufficiently large, the update happens between the two adjacent agents  $i$  and  $j$  satisfying  $\varepsilon < \|\mathbf{x}_i(k_0) - \mathbf{x}_j(k_0)\|_2 < \delta$ , according to (6),

$$f(X(k_0)) - f(X(k_0+1)) = \frac{\|\mathbf{x}_i(k_0) - \mathbf{x}_j(k_0)\|_2^2}{2} > \frac{\varepsilon^2}{2}.$$

The above inequality contradicts (9). This implies that, from some

large enough  $k_0$ ,  $\mathbb{P}(A_{k_0}) = 0$ . We therefore see that

$$\begin{aligned}
 & \mathbb{P}\left(\bigcap_{k \in \mathbb{N}} \Omega_k^c\right) \\
 & \leq \mathbb{P}\left(\bigcap_{k \geq k_0} \left[\Omega_k^c \cap (A_k \cup A_k^c)\right]\right) \\
 & \leq \mathbb{P}\left(\bigcap_{k \geq k_0} (\Omega_k^c \cap A_k^c)\right) \\
 & = \mathbb{P}\left(\left(\bigcap_{k \geq k_0} \Omega_k^c\right) \cap \left(\bigcap_{k \geq k_0} A_k^c\right)\right) \\
 & = \mathbb{P}\left(\bigcap_{k \geq k_0} \Omega_k^c\right) \mathbb{P}\left(\bigcap_{k \geq k_0} A_k^c \mid \bigcap_{k \geq k_0} \Omega_k^c\right) \\
 & = 0.
 \end{aligned} \tag{10}$$

Here the notion  $\mathbb{P}(X \mid Y)$  denotes the conditional probability of random variable  $X$  given  $Y$ . Due to the assumption (8), the conditional probability above is well defined. The last equality is deduced from the fact that the probability for  $A_k^c$  to happen at infinite steps, given by  $\bigcap_{k_0 \in \mathbb{N}} \Omega_k^c$ , is zero. We see that (10) contradicts (8). This follows that

$$\mathbb{P}\left(\bigcap_{k \in \mathbb{N}} \Omega_k^c\right) = 0.$$

This completes the proof.  $\blacksquare$

In the following we state and prove the main result of this paper.

**Theorem 2:** For any given  $X(1)$  and a given confidence threshold  $\delta$ , the proposed algorithm achieves  $\delta$ -clustering convergence almost surely with respect to choosing interaction pair at each iterative step.

*Proof:* Theorem 1 holds true for any small  $\varepsilon > 0$ . Let  $\varepsilon$  tend to zero. Then we get the theorem.  $\blacksquare$

Theorem 2 ensures that the opinion profile, obtained from the averaging update between random chosen pairs of agents, eventually converges to one such that any two adjacent agents either have the same opinion or different opinions whose distance is no less than the confidence threshold  $\delta$ . As a consequence, when the number  $\kappa$  of iterations is chosen large enough,  $X(\kappa)$ , obtained from Algorithms 1 and 2 would be close enough to some stable state of the form (5). Notice that the above results hold regardless connectivity of the graph, while the algorithms have the following property if the graph is complete.

**Corollary 1:** Suppose that the communication graph is complete, then for any given initial opinion profile and a given confidence threshold  $\delta$ , the proposed algorithm drives the opinions of agents to a stable state where any two opinions are either the same or differ no less than the confidence threshold, with probability 1.

#### IV. APPLICATIONS TO OPINION FORMATION AND DISTRIBUTED DATA CLUSTERING

**Example 2 (Opinion Formation):** We consider a network of 100 agents. The communication graph over the network is a random geometric graph, namely, the 100 agents are located uniformly and independently in the unit square, and each pair of agents is connected if their Euclidean distance is no more than some critical distance. The communication graph is depicted in Fig. 1, where the critical distance is chosen to be 0.15.

Initial opinion profiles used in the simulations are taken randomly from  $(0, 1)^{100}$ . We take  $\delta = 0.35$ . Fig. 2 shows a trajectory of opinion profile according to time from 1 to  $2 \times 10^5$ . Fig. 3 demonstrates the

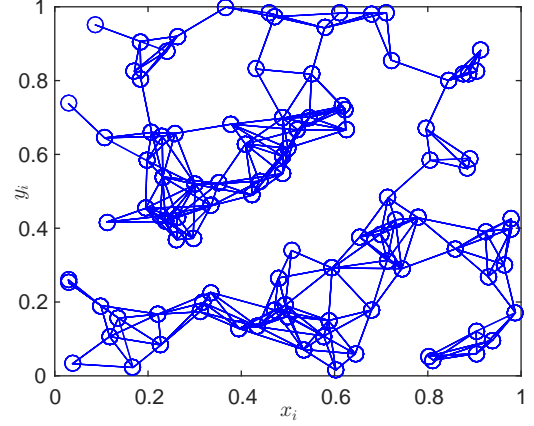


Fig. 1. Communication graph in Example 2

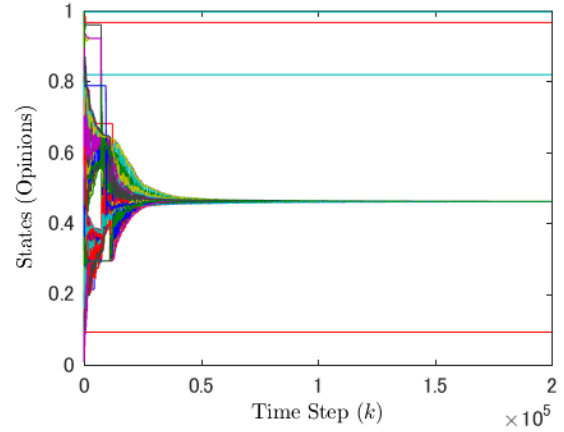


Fig. 2. One trajectory of opinion profile with  $\delta = 0.35$

final opinion profile of the trajectory at  $k = 2 \times 10^5$  accompanied by the communication graph. In this figure, the markers are at  $(x_i, y_i, z_i)$ ,  $i = 1, 2, \dots, 100$ , where the positions of agents are represented by the  $x$  and  $y$  coordinates, and their opinions are shown by the  $z$  coordinate. The solid lines connecting markers represent edges of the communication graph. The result shows the clustering phenomenon in the convergent opinion profile. We see that any two adjacent agents either have nearly same opinions or very far opinions whose distance is at least  $\delta$ .

**Example 3 (Data Clustering):** We consider a network consisting of 100 agents with complete communication graph. We use the same initial opinion profile as that in Example 2. Figs. 4 and 5 demonstrate the convergence results with different values of confidence threshold  $\delta = 0.5, 0.05$ .

We see that, with  $\delta = 0.5$ , all the opinions form only one  $\delta$ -cluster at all time in Fig. 4. Then, the system state gets to averaging consensus. On the other hand, in the other simulation with  $\delta = 0.05$ , we get several clusters in final opinion profile (Fig. 5).

Fig. 6 illustrates a process resulting from the Algorithm 2. Here the confidence threshold increases from 0.05, to 0.1, to 0.15, and then to 0.2. For each value of confidence threshold, there are  $10^4$  iteration steps.

Both examples ensure the clustering phenomena of public opinion resulting from the two opinion forming algorithms. Example 3 implies that we are able to apply our algorithms to the clustering algorithms. Since our algorithm is according to distributed manner,

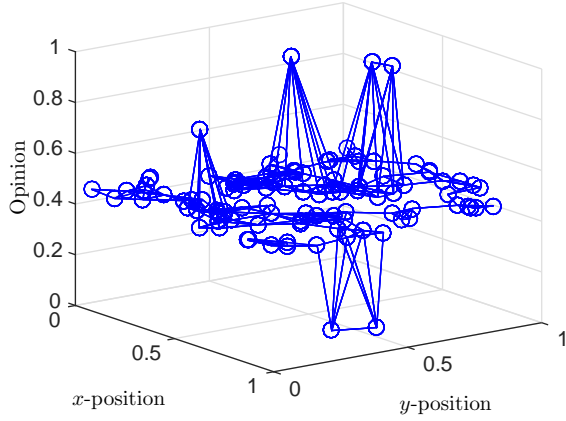


Fig. 3. Convergent opinion profile with  $\delta = 0.35$

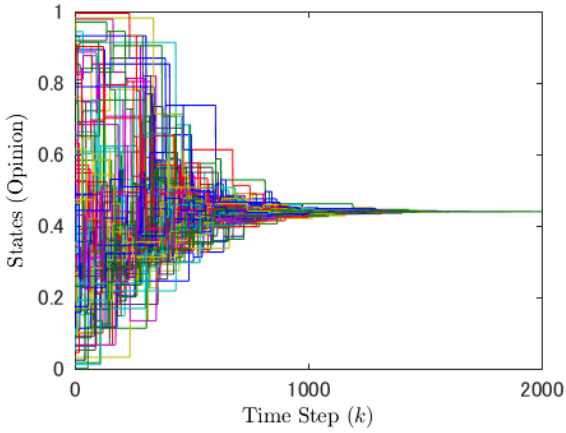


Fig. 4. Averaging convergence with  $\delta = 0.5$

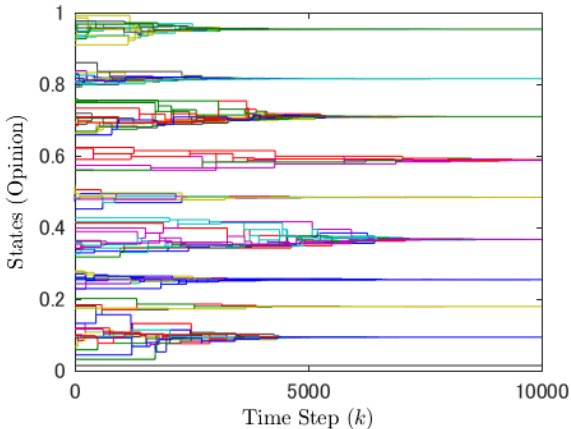


Fig. 5. Clustering convergence with  $\delta = 0.05$

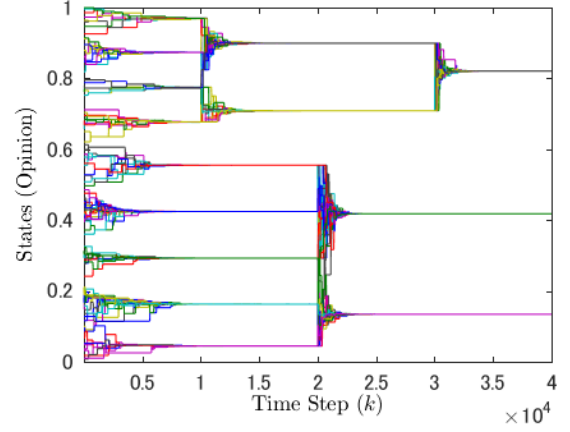


Fig. 6. Increasing confidence threshold with increment 0.05

computational cost for each update is  $O(n)$ . That is, if the number of data are large, we can apply our distributed algorithms for clustering. This is a key feature of them. Furthermore, if we select appropriate  $\delta$ , we can adjust the number of clustering. In fact, when we selected  $\delta$  as 0.5, we obtained only 1 cluster (Fig. 4). However,  $\delta = 0.05$  leads to 10 clusters (Fig. 4). In addition to it, we can adjust the number of clusters on line by increasing  $\delta$  gradually (Fig.6).

## V. CONCLUDING REMARKS

In this paper, we have proposed two gossip algorithms describing the opinion formation process over a network of agents through pairwise interaction. The algorithms are shown to converge almost surely with respect to choosing interaction pairs for any initial opinion profile. More precisely, the algorithms reveal clustering formation of opinion profile.

## ACKNOWLEDGMENT

This work was supported by JST CREST Grant Number JP-MJCR15K2, Japan.

## REFERENCES

- [1] J. R. P. French, "A formal theory of social power," *Psychol. Rev.*, vol. 63, pp. 181-194, 1956.
- [2] F. Harary, "A criterion for unanimity in French's theory of social power," *Studies in Social Power*, Oxford, England: Univers. Michigan, vol. ix, pp. 168-182, 1959.
- [3] S. Chatterjee and E. Seneta, "Toward consensus: some convergence theorems on repeated averaging," *J. Appl. Prob.*, vol. 14, pp. 89-97, 1977.
- [4] M. H. De Groot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, pp. 118-121, 1974.
- [5] H. Föllmer, "Random economies with many interacting agents," *J. Math. Econ.*, vol. 1, pp. 51-62, 1974.
- [6] B. Latané and A. Nowak, "Self-organizing social systems: Necessary and sufficient conditions for the emergence of clustering, consolidation, and continuing diversity," *Progress in Communication Sciences*, pp. 43-74, 1997.
- [7] K. Sznajd-Weron and J. Sznajd, "Opinion evolution in closed community," *Int. J. Mod. Phys. C*, vol. 11, pp. 1157-1166, 2000.
- [8] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *J. Math. Soc.*, vol. 15, pp. 193-206, 1990.
- [9] N. E. Friedkin and E. C. Johnsen, "Social influence networks and opinions change," *Advances in Group Processes*, vol. 16, pp. 1-29, 1999.
- [10] U. Krause, "A discrete nonlinear and non-autonomous model of consensus formation," In Selaydi, J. Popena and J. Rakowski (Eds.), *Communications in Differential Equations*, Amsterdam: Gordon and Breach Publ., pp. 227-236, 2000.

- [11] J. Dittmer, "Consensus formation under bounded confidence," *Nonlinear Anal.-Theor.*, vol. 47, pp. 4615-4622, 2001.
- [12] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation (JASSS)*, vol. 5 (3), pp. 1-33, 2002.
- [13] J. Gómez-Serrano, C. Graham, and J. L. Boudec, "The bounded confidence model of opinion dynamics," *Mathematical Models and Methods in Applied Sciences*, vol. 22, no. 2, pp. 115007-1 - 115007-46, 2012.
- [14] G. Kou, Y. Zhao, Y. Peng, and Y. Shi, "Multi-level opinion dynamics under bounded confidence," *Plos ONE*, vol. 7 (9), e43507, pp. 1-10, 2012.
- [15] B. Touri and C. Langbort, "On endogenous random consensus and averaging dynamics," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 241-248, 2014.
- [16] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Adv. Complex Syst.*, Vol. 3, pp. 87-98, 2000.
- [17] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, Vol. 2, pp. 241-254, 1967.
- [18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, Vol. 39, No. 1, pp. 1-38, 1977.
- [19] L. T. H. Nguyen, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki, "Opinion formation under bounded confidence via gossip algorithm," *Proceeding of the 54th IEEE Conference on Decision and Control*, Osaka, Japan, December 15-18, 2015, pp. 2223-2228.
- [20] L. T. H. Nguyen, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki, "Gossip algorithms for clustering problems," *Proceedings of the 41st Annual Conference of the IEEE Industrial Electronics Society*, Yokohama, Japan, November 9-12, 2015, pp. 589-594.
- [21] L. T. H. Nguyen, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki, "A bounded confidence gossip algorithm over communication graph for clustering opinions," *Proceedings of the 2nd SICE International Symposium on Control Systems*, Nagoya, Japan, March 7-10, 2016, paper ID 2A3-3.
- [22] S. Baker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart\*: an open dataset and tools for enabling research in sustainable homes," *Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability*, Beijing, China, August 2012.
- [23] D. L. Davies, D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1 (2), pp. 224-227.
- [24] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Computational and Applied Mathematics*, Vol. 20, pp. 53-65, 1987.
- [25] R. N. Shepard, "Representation of structure in similarity data: problems and prospects," *Psychometrika*, Vol. 39, pp. 373-421, 1974.