EUSIPCO 2013 1569744751

CLUSTER FORMATION OVER ADAPTIVE NETWORKS WITH SELFISH AGENTS

Chung-Kai Yu, Mihaela van der Schaar, and Ali H. Sayed

Electrical Engineering Department
University of California, Los Angeles

ABSTRACT

We examine the problem of adaptation and learning over networks with selfish agents. In order to motivate agents to cooperate, we allow the agents to select their partners according to whether they can help them reduce their utility costs. We divide the operation of the network into two stages: a cluster formation stage and an information sharing stage. During cluster formation, agents evaluate a long-term combined cost function and decide on whether to cooperate or not with other agents. During the subsequent information sharing phase, agents share and process information over their sub-networks. Simulations illustrate how the clustering technique enhances the mean-square-error performance of the agents over non-cooperative processing.

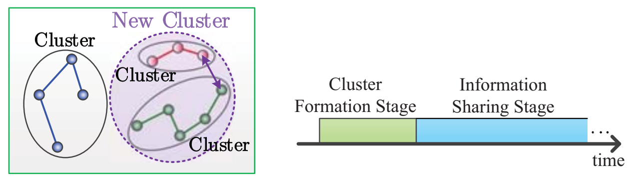
***Index Terms*—** Adaptive networks, cluster formation, selfish agents, diffusion strategy, mean-square-error.

1. INTRODUCTION

In prior works on distributed estimation over networks, agents were modeled as cooperative players that exchange information willingly. Several distributed strategies have been developed to enable the decentralized processing of information among cooperating agents, such as the consensus strategy (e.g., [1, 2]) and the diffusion strategy (e.g., [3–5]). In this work, we study networks where agents can behave in a selfish manner. In this case, agents share information with their neighbors only if they believe that cooperation is beneficial for their long-term interests.

One way to motivate cooperation among selfish agents is to allow them to decide with whom to cluster and share information. The clustering concept is widely studied in the social sciences and game theory (e.g., [6–10]). It enables agents to drive their cooperative behavior by selecting their partners according to whether they can help them reduce their utility costs. For adaptive networks, the challenge is to select utility functions that can drive the clustering operation. Recent results on the performance of adaptive networks [11] can be exploited to great effect for this purpose.

This work was supported in part by NSF Grants CCF-1011918 and NSF CSR-1016081. Emails: {ckyuna, mihaela, sayed}@ee.ucla.edu



(a) Cluster formation. (b) Timeline.

Fig. 1: (a) Selfish agents establish new links to form a larger cluster. (b) Timeline illustrates two stages of cluster formation and information sharing.

In the formulation studied in this paper, the objective of the agents is to estimate a common parameter of interest by relying on local measurements and on local interactions. We divide the operation of the network into two stages. The first stage is the cluster formation phase and the second stage is the information sharing and processing phase. During cluster formation, agents meet randomly in pairs following a random pairing protocol [12]. This situation could occur, for example, due to an exogenous matcher or the mobility of the agents. Based on some prior reference knowledge about mutual clusters, each agent then evaluates the expected cost of its possible actions and decides on whether to propose cooperation to the other agent. If both agents agree on cooperation, then they establish a link and become part of the same larger cluster. We illustrate cluster formation and the timeline involved in Figure 1. Once clusters are formed, the agents can then proceed to solve the estimation task in a distributed manner by cooperating within their sub-networks. We assume there exist harsh punishments to prevent agents from deviating from the agreement of information sharing, such as to permanently isolate the deviant agents.

2. INFORMATION SHARING STRUCTURE

2.1. Reference Knowledge and Transmission Cost

Consider a network with N selfish agents. During the cluster formation stage, pairs of agents, say, agents k and ℓ, randomly meet and exchange some preliminary knowledge, denoted by $\mathbb{K}_k$ and $\mathbb{K}_\ell$, respectively. Based on $\mathbb{K}_k$ and $\mathbb{K}_\ell$, the agents decide on whether they want to become part of the same cluster. Membership in the same cluster implies that the agents would agree to cooperate with each other during the information sharing stage. During this second phase, agents share information denoted by $\mathbb{I}_{k,i}$ and $\mathbb{I}_{\ell,i}$ at time i. Obviously, the sharing of the information $\mathbb{I}_{k,i}$ with agent ℓ bears some transmission cost for agent k, which is denoted by $c_{k\ell} > 0$ and assumed to be known by agent k. Likewise, $c_{\ell k} > 0$ represents the cost for agent ℓ when it shares information with agent k. In the subscripts ℓk, the first letter represents the source agent and the second letter represents the destination agent. We set $c_{kk} = 0$.

2.2. Agreement to Cluster

When agent k first meets agent ℓ during the cluster formation stage, agent k chooses an action $\boldsymbol{\alpha}_{k\ell} \in \{0, 1\}$ based on their shared preliminary knowledge $\mathbb{K}_k$ and $\mathbb{K}_\ell$ (as described further ahead in Sec. 3).The action $\boldsymbol{\alpha}_{k\ell} = 1$ means that agent k proposes to agent ℓ that they become part of the same cluster, and the action $\boldsymbol{\alpha}_{k\ell} = 0$ means that agent k does not want to cluster with agent ℓ. Agent ℓ's action, $\boldsymbol{\alpha}_{\ell k} \in \{0, 1\}$, is defined in a similar manner. The agreement to cluster must be consensual, i.e., both agents need to propose $\boldsymbol{\alpha}_{k\ell} = 1$ and $\boldsymbol{\alpha}_{\ell k} = 1$. This situation can be represented by the indicator value defined by:

$$\boldsymbol{I}_{k\ell} = \boldsymbol{I}_{\ell k} \triangleq \boldsymbol{\alpha}_{k\ell} \cdot \boldsymbol{\alpha}_{\ell k} \tag{1}$$

Thus, $\boldsymbol{I}_{k\ell} = 1$ means that both agents have agreed to become part of the same cluster so that agent k will share information $\mathbb{I}_{k,i}$ with agent ℓ during the information sharing stage, and vice-versa. On the other hand, $\boldsymbol{I}_{k\ell} = 0$ means that agents k and ℓ do not wish to cluster. We set $\boldsymbol{I}_{kk} = 1$.

2.3. Diffusion Strategy

During the information sharing stage, agents will share information to solve a distributed estimation task, such as estimating and tracking some parameter vector of interest, which we denote by $w^o \in \mathbb{C}^{M\times 1}$. In this context, the information $\mathbb{I}_{k,i}$ to be shared by agent k refers to its estimate of w^o at time i, which we denote by $\boldsymbol{w}_{k,i}$. At each time instant i during the information sharing stage, each agent k in the network is assumed to have access to a scalar measurement $\boldsymbol{d}_k(i) \in \mathbb{C}$ and a $1 \times M$ regression vector $\boldsymbol{u}_{k,i} \in \mathbb{C}^{1\times M}$ with a covariance matrix $R_{u,k} \triangleq \mathbb{E}\boldsymbol{u}_{k,i}^*\boldsymbol{u}_{k,i} > 0$. The data are assumed to be related via the linear regression model:

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}_k(i) \tag{2}$$

where $\boldsymbol{v}_k(i) \in \mathbb{C}$ is measurement noise with variance $\sigma_{v,k}^2$ and is independent of all other variables. Models of the form (2) are common in applications and can be used to model several scenarios of interest: parameter and channel estimation, target tracking, system modeling, data regression, etc. Agents in the network update their estimates of w^o based on their own data $\boldsymbol{d}_k(i)$ and $\boldsymbol{u}_{k,i}$, and on estimates from their neighbors. Two prominent classes of distributed strategies that can be used to compute the estimates $\boldsymbol{w}_{k,i}$ in a distributed and online manner are consensus strategies [1, 2] and diffusion strategies [3–5]. In this work, we focus on diffusion strategies since they have been shown to have superior mean-square-error performance and stability properties [13]. There are several variants of diffusion adaptation. We employ the adapt-then-combine (ATC) formulation, where agents update their estimates according to the following recursive construction:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^*[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}] \tag{3}$$

$$\boldsymbol{w}_{k,i} = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\psi}_{\ell,i} \tag{4}$$

where the symbol $\mathcal{N}_k$ denotes the set of neighbors of agent k, including k itself; these are agents that can share information directly with k. As illustrated in Figure 2, it is obvious that agents in $\mathcal{N}_k$ should belong to the cluster of agent k, denoted by $\mathcal{C}_k$, i.e., $\mathcal{N}_k \subset \mathcal{C}_k$. The cluster of agent k includes two types of agents: (a) those agents which agent k has decided to cluster with and, therefore, has direct links to them, and (b) agents which agent k has a path through other intermediate agents to connect with. In other words, the set $\mathcal{C}_k$ represents a connected sub-network that includes k and its immediate neighborhood in addition to other agents. Formally, the cluster set $\mathcal{C}_k$ is constructed as follows. Representing the connection topology graphically, we connect two agents k and ℓ by an edge if $\boldsymbol{I}_{k\ell} = 1$. Then, the cluster $\mathcal{C}_k$ is the maximally connected subnetwork containing agent k. In this way, for any other agent in $\mathcal{C}_k$, there will exist at least one path connecting agent k to it either directly by an edge, or by means of a path passing through other intermediate agents.

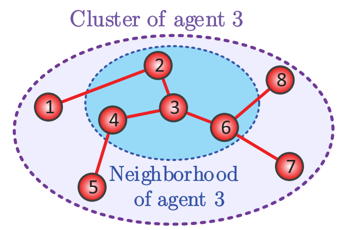


Fig. 2: The neighborhood of agent 3 is $\mathcal{N}_3 = \{2, 3, 4, 6\}$ and the cluster of agent 3 is $\mathcal{C}_3 = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

The parameter μ_k in (3) is a positive step-size factor, which is assumed to be sufficiently small and identical for all agents, i.e., $\mu_k \equiv \mu \ll 1$. Sufficiently small step-sizes ensure mean-square stability of the diffusion strategy [3, 4, 14]. In the first step (3), an intermediate estimate $\boldsymbol{\psi}_{k,i}$ is determined by adjusting the existing estimate $\boldsymbol{w}_{k,i-1}$ using local data. The second step (4) uses non-negative coefficients $\{a_{\ell k}\}$ to combine the estimates from the neighbors. The coefficients $\{a_{\ell k}\}$ are required to satisfy:

$$a_{\ell k} \geq 0,\ a_{\ell k} = 0 \text{ if } \ell \notin \mathcal{N}_k \tag{5}$$

We collect the coefficients $\{a_{\ell k}\}$ into an $N \times N$ matrix A. In this work, although unnecessary, we assume that A is doubly-stochastic, i.e., the entries on each of its rows and columns add up to one, such as selecting A to be the Laplacian com-

Table 1: Cost values for all four combinations of actions by the selfish agents.

	$\boldsymbol{\alpha}_{k\ell} = 0$	$\boldsymbol{\alpha}_{k\ell} = 1$
$\boldsymbol{\alpha}_{\ell k} = 0$	$\text{MSD}_k(\mathcal{C}_k) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$ $\text{MSD}_\ell(\mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q}$	$\text{MSD}_k(\mathcal{C}_k) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$ $\text{MSD}_\ell(\mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q}$
$\boldsymbol{\alpha}_{\ell k} = 1$	$\text{MSD}_k(\mathcal{C}_k) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$ $\text{MSD}_\ell(\mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q}$	$\text{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} + \beta_k c_{k\ell}$ $\text{MSD}_\ell(\mathcal{C}_k \cup \mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q} + \beta_\ell c_{\ell k}$

bination rule [14, 15] or the Metropolis combination rule [15, 16]. Then, we have

$$A^T \mathbb{1} = \mathbb{1}, \quad A\mathbb{1} = \mathbb{1} \tag{6}$$

where the notation $\mathbb{1}$ denotes a vector with all its entries equal to one. In the context of algorithm (3)-(4), the information to be shared between neighbors are the intermediate estimates $\boldsymbol{\psi}_{\ell,i}$.

During the cluster formation stage, the cluster dynamics is evolving and, therefore, $\mathcal{C}_k$ is dependent on time during this phase. When two agents k and ℓ first meet randomly at some time i, prior to the adaptation stage involving (3)-(4), the reference knowledge $\mathbb{K}_k$ and $\mathbb{K}_\ell$ that they share is assumed to consist of the agents that belong to their clusters and their respective noise variances:

$$\mathbb{K}_k \triangleq \left\{(q, \sigma_{v,q}^2) | q \in \mathcal{C}_k\right\} \tag{7}$$

When two agents decide to cluster, then their cluster sets are merged and all agents in these sets become part of the same larger cluster. As such, whenever two agents meet and they are not members of the same cluster, then their cluster sets are necessarily disjoint.

3. COMBINED COST FOR CLUSTERING AGREEMENT

In the cluster formation stage, when two agents k and ℓ meet randomly, they select their actions $\{\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k}\}$ based on their assessment of a long-term expected return as follows. Each agent k employs a combined cost function that takes into account the cost of communicating with agent ℓ *and* the contribution of agent ℓ towards the estimation task (i.e., whether it will help reduce the steady-state mean-square error). The combined cost function for agent k depends on the actions by both agents and on their existing clusters:

$$J_k(\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k} | \mathcal{C}_k, \mathcal{C}_\ell) \triangleq \begin{cases} \text{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) + \beta_k \left(\sum\limits_{q \in \mathcal{N}_k \cup \{\ell\}} \boldsymbol{I}_{kq} c_{kq} \right), & \text{if } (\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k}) = (1,1) \\ \text{MSD}_k(\mathcal{C}_k) + \beta_k \left(\sum\limits_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} \right), & \text{otherwise} \end{cases} \tag{8}$$

where β_k is a normalization parameter, and MSD_k denotes the steady-state mean-square-deviation (MSD) measure for agent k:

$$\text{MSD}_k \triangleq \lim_{i \to \infty} \mathbb{E} \|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \tag{9}$$

in terms of the error vector $\widetilde{\boldsymbol{w}}_{k,i} \triangleq w^o - \boldsymbol{w}_{k,i}$. Moreover, the notation $\text{MSD}_k(\mathcal{C}_k)$ for cluster $\mathcal{C}_k$ is used to denote the MSD value that would be attained by agent k if its cluster is $\mathcal{C}_k$. In Table 1, we summarize the resulting cost values for the agents under their respective actions.

Let us now explain how the MSD values in (8) can be evaluated. Consider an arbitrary agent k and a cluster set $\mathcal{C}_k$ of size K. Under the assumption that the regressors $\boldsymbol{u}_{k,i}$ are spatially and temporally independent and that the step-size μ is sufficiently small, it holds that for the doubly-stochastic A, we have the following expression (refer to Equations (89) and (97) in [11] or Equation (32) in [17]):

$$\text{MSD}_k(\mathcal{C}_k) \approx \frac{\mu M}{2} \cdot \frac{1}{K^2} \sum_{q \in \mathcal{C}_k} \sigma_{v,q}^2 \tag{10}$$

Suppose agent k meets agent ℓ with cluster $\mathcal{C}_\ell$ of size L. We note that one of two situations will occur: $\mathcal{C}_k = \mathcal{C}_\ell$ or $\mathcal{C}_k \bigcap \mathcal{C}_\ell = \emptyset$. In the trivial case that $\mathcal{C}_k = \mathcal{C}_\ell$, we have

$$\text{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) = \text{MSD}_k(\mathcal{C}_k) \tag{11}$$

since agents k and ℓ have the same cluster. For $\mathcal{C}_k \bigcap \mathcal{C}_\ell = \emptyset$, if agents k and ℓ fail to reach agreement, which means $\boldsymbol{I}_{k\ell} = 0$, then we again obtain $\text{MSD}_k(\mathcal{C}_k)$ for (10). On the other hand, if they successfully reach agreement ($\boldsymbol{I}_{k\ell} = 1$), then

$$\text{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) \approx \frac{\mu M}{2} \cdot \frac{1}{(K+L)^2} \sum_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2 \tag{12}$$

In this way, the combined cost values in (8) are given by:

$$J_k(\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k} | \mathcal{C}_k, \mathcal{C}_\ell) = \begin{cases} \frac{\mu M}{2} \cdot \frac{1}{(K+L)^2} \sum\limits_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2 + \beta_k \sum\limits_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} + \beta_k c_{k\ell}, & \text{if } (\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k}) = (1,1) \\ \frac{\mu M}{2} \cdot \frac{1}{K^2} \sum\limits_{q \in \mathcal{C}_k} \sigma_{v,q}^2 + \beta_k \sum\limits_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}, & \text{otherwise} \end{cases} \tag{13}$$

Then, agents choose the actions that minimize their combined cost function (13). Once $\boldsymbol{I}_{kl} = 1$, agents k and ℓ start sharing estimates in the information sharing stage. To prevent agents from deviating from the agreement, we punish the deviant agents in the following manner: if any agent k violates the agreement to cooperate with agent ℓ, agent ℓ broadcasts this misbehavior to its neighbors and from there to their neighbors and agents will stop sharing estimates with agent k permanently.

We remark that the individual actions of agents could impact the combined cost values of other agents in the same cluster. However, individual actions do not worsen the marginal combined costs of other agents in a cluster. To see this, if no larger clustering (no new agreement) occurs, the combined cost of every agent in a cluster remains the same. If a new clustering agreement of agents, say, k and ℓ, is made, the MSD costs of other agents reduce but there is no addition communication cost required by them, and thus their combined costs reduce.

4. CLUSTER FORMATION PROCESS

The following lemma characterizes the conditions for cluster formation.

Lemma 1. *Agents k and ℓ reach agreement to cluster ($\boldsymbol{I}_{k\ell} = 1$), when the following two conditions are met:*

$$\frac{\sum_{q\in\mathcal{C}_k}\sigma_{v,q}^2}{K^2} - \frac{\sum_{q\in\mathcal{C}_k\cup\mathcal{C}_\ell}\sigma_{v,q}^2}{(K+L)^2} > \frac{2}{\mu M}\beta_k c_{k\ell} \tag{14}$$

and

$$\frac{\sum_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2}{L^2} - \frac{\sum_{q\in\mathcal{C}_k\cup\mathcal{C}_\ell}\sigma_{v,q}^2}{(K+L)^2} > \frac{2}{\mu M}\beta_\ell c_{\ell k} \tag{15}$$

Proof. From Table 1, we first note that if agent ℓ selects $\boldsymbol{\alpha}_{\ell k} = 0$, then it is indifferent to agent k selecting $\boldsymbol{\alpha}_{k\ell} = 0$ or 1. On the other hand, in the case of $\boldsymbol{\alpha}_{\ell k} = 1$, if we have

$$\begin{aligned} J_k(\boldsymbol{\alpha}_{k\ell} = 0, &\boldsymbol{\alpha}_{\ell k} = 1|\mathcal{C}_k,\mathcal{C}_\ell) \\ &> J_k(\boldsymbol{\alpha}_{k\ell} = 1, \boldsymbol{\alpha}_{\ell k} = 1|\mathcal{C}_k,\mathcal{C}_\ell) \end{aligned} \tag{16}$$

then agent k should choose $\boldsymbol{\alpha}_{k\ell} = 1$ to obtain a lower combined cost. Therefore, condition (16) ensures the best strategy for agent k to be $\boldsymbol{\alpha}_{k\ell} = 1$. Using (13) we can rewrite (16) as

$$\begin{aligned} &\frac{\mu M}{2}\left(\frac{\sum_{q\in\mathcal{C}_k}\sigma_{v,q}^2}{K^2}\right) + \beta_k \sum_{q\in\mathcal{N}_k} \boldsymbol{I}_{kq}c_{kq} \\ &> \frac{\mu M}{2}\left(\frac{\sum_{q\in\mathcal{C}_k\cup\mathcal{C}_\ell}\sigma_{v,q}^2}{(K+L)^2}\right) + \beta_k \sum_{q\in\mathcal{N}_k} \boldsymbol{I}_{kq}c_{kq} + \beta_k c_{k\ell} \end{aligned} \tag{17}$$

which is equivalent to (14). Similarly, we can obtain condition (15) to ensure $\boldsymbol{\alpha}_{\ell k} = 1$ from agent ℓ's perspective. □

Note that when conditions (14) and (15) hold, the dominant strategies for agents k and ℓ become $\boldsymbol{\alpha}_{k\ell} = 1$ and $\boldsymbol{\alpha}_{\ell k} = 1$. On the other hand, when either one of conditions (14) or (15) fails to hold, agents have no incentive to cluster. In this case, $(\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k}) = (1,1)$ will not be chosen, which results in $\boldsymbol{I}_{k\ell} = 0$. We assume agents k and ℓ select $(\boldsymbol{\alpha}_{k\ell}, \boldsymbol{\alpha}_{\ell k}) = (0,0)$ if equalities occur in (14) and (15). From Lemma 1, we know that clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ unite if both conditions (14) and (15) hold. Furthermore, we observe that low weighted transmission costs, $\beta_k c_{k\ell}$ and $\beta_\ell c_{\ell k}$, facilitate the formation of the united cluster. Now, let us consider networks with uniform $\beta_k = \beta_\ell \equiv \beta$ and $c_{k\ell} = c_{\ell k} \equiv c$. If every agent further has the same noise variance, we obtain the following result.

Lemma 2. *If the noise variances across the network are uniform, i.e., $\sigma_{v,q}^2 \equiv \sigma_v^2$, then the following condition guarantees the cluster formation $\mathcal{C}_k \cup \mathcal{C}_\ell$:*

$$\frac{K+L}{\sigma_v^2}\frac{2}{\mu M}\beta c < \min\left\{\frac{L}{K}, \frac{K}{L}\right\} \tag{18}$$

Proof. For agent k, it follows from (14) that we must have

$$\frac{L}{K} > \frac{K+L}{\sigma_v^2}\frac{2}{\mu M}\beta c \tag{19}$$

Similarly, for agent ℓ it follows from (15) that we must have

$$\frac{K}{L} > \frac{K+L}{\sigma_v^2}\frac{2}{\mu M}\beta c \tag{20}$$

Combining both results, we obtain (18). □

Therefore, if we want to facilitate the formation of larger clusters, Lemma 2 suggests to maximize the right-hand side of (18), which occurs when $K = L$ and the maximum value becomes equal to one. In other words, larger clustering is more likely to occur for clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ of equal sizes. Now, let us examine the case in which the clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ have the same sizes but their agents have heterogeneous noise variances.

Lemma 3. *If clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ have the same sizes, i.e., $K = L$, then the following condition guarantees the cluster formation $\mathcal{C}_k \cup \mathcal{C}_\ell$:*

$$\frac{8}{\mu M}\beta c < \min\left\{\frac{1}{K}(3\bar{\sigma}_k^2 - \bar{\sigma}_\ell^2), \frac{1}{L}(3\bar{\sigma}_\ell^2 - \bar{\sigma}_k^2)\right\} \tag{21}$$

where

$$\bar{\sigma}_k^2 \triangleq \frac{1}{K}\sum_{q\in\mathcal{C}_k}\sigma_{v,q}^2 \;\; \text{and} \;\; \bar{\sigma}_\ell^2 \triangleq \frac{1}{L}\sum_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2 \tag{22}$$

are the average noise variances of $\mathcal{C}_k$ and $\mathcal{C}_\ell$, respectively.

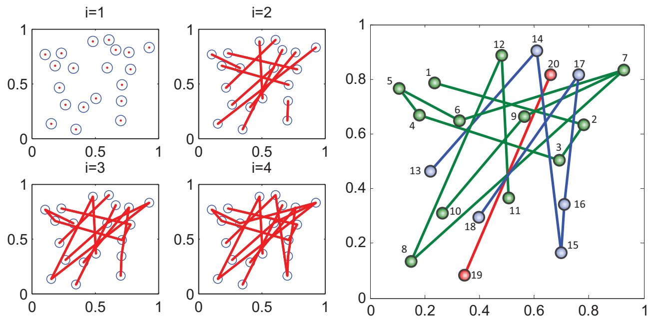
Proof. For agent k, we conclude from (14) that we must have:

$$3\frac{\sum_{q\in\mathcal{C}_k}\sigma_{v,q}^2}{K} - \frac{\sum_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2}{L} > \frac{8K}{\mu M}\beta c \tag{23}$$

Similarly, for agent ℓ it must hold that

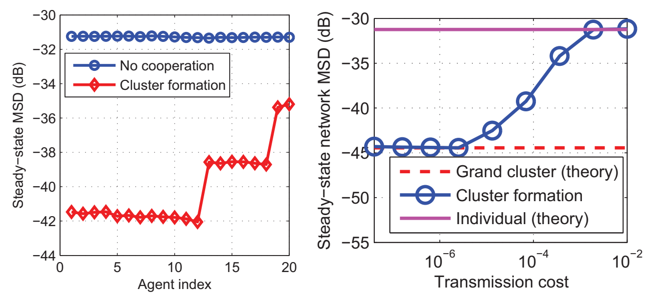
$$3\frac{\sum_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2}{L} - \frac{\sum_{q\in\mathcal{C}_k}\sigma_{v,q}^2}{K} > \frac{8L}{\mu M}\beta c \tag{24}$$

Combining both conditions, we obtain (21). □



(a) Evolution of clusters. (b) Final clusters.

Fig. 3: Cluster formation with $c = 5 \times 10^{-5}$ and $\sigma_v^2 = -6$ (dB).



(a) Agents in three clusters. (b) Effect of transmission cost.

Fig. 4: Simulations of steady-state network MSD.

Again, the maximum of the term on the right-hand side of (21) occurs when

$$(3L + K)\bar{\sigma}_k^2 = (3K + L)\bar{\sigma}_L^2 \tag{25}$$

Therefore, for clusters of equal sizes, two clusters with the same (weighted) average noise variance will be more likely to unite.

5. SIMULATION RESULTS

In our simulations, we consider a network with 20 agents. During the first 10 time instants, agents are uniformly and randomly paired. Then, agents proceed to cooperate within their clusters to solve the estimation problem. The length of w^o is $M = 3$ and we randomly choose its entries and normalize them to satisfy $\|w^o\| = 1$. The regressor $\{\boldsymbol{u}_{k,i}\}$ is zero-mean and $R_{u,k}$ is diagonal with entries uniformly generated between [0,1]. The background noise $\boldsymbol{v}_k(i)$ is temporally white and spatially independent Gaussian distributed with zero-mean and assumed to be uniform with variance $\sigma_{v,k}^2 = \sigma_v^2 = -6$ (dB). We set $\mu = 0.005$, $\beta_k = \beta = 1$, and $c_{k\ell} = c = 5 \times 10^{-5}$ for all agents.

Figure 3(a) shows the topology evolution from $i = 1$ to 4. We observe that agents gradually form clusters to maximize their own utilities. The final topology with three disjoint clusters is shown in Figure 3(b). Cooperating over the resulting sub-networks, agents start to share estimates and run algorithm (3)-(4). We simulate the corresponding steady-state MSD in Figure 4(a) where agents are indexed and grouped according to their clusters. We observe that through clustering, every agent is able to achieve better estimation performance than if the agents were to act independently of the other agents by running their own individual LMS recursions. Figure 4(b) shows the effect of transmission cost to the cluster formation and thus to the steady-state network MSD.

6. REFERENCES

[1] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and trade-offs," *IEEE J. Sel. Topics in Signal Process.*, vol. 5, no. 5, pp. 674–690, Aug. 2011.

[2] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[3] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, July 2008.

[4] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[5] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4289–4305, Aug. 2012.

[6] J. P. Kahan and A. Rapoport, *Theories of Coalition Formation*, Hillsdale, NJ: L. Erlbaum Associates, 1984.

[7] W. A. Gamson, "A theory of coalition formation," *Amer. Sociolog. Rev., 26 (3)*, pp. 373–382, 1961.

[8] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations*, Cambridge University Press, 2008.

[9] H. Park and M. van der Schaar, "Coalition-based resource reciprocation strategies for P2P multimedia broadcasting," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 557–567, Sept. 2008.

[10] H. Park and M. van der Schaar, "Coalition based resource negotiation for multimedia applications in informationally decentralized networks," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 765–779, June 2009.

[11] X. Zhao and A. H. Sayed, "Performance limits for distributed estimation over LMS adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5107–5124, Oct. 2012.

[12] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[13] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[14] A. H. Sayed, "Diffusion adaptation over networks," in *E-Reference Signal Processing,* R. Chellapa and S. Theodoridis, editors, Elsevier, 2013. Also available online as arXiv:1205.4220 [cs.MA], May 2012.

[15] D. S. Scherber and H. C. Papadopoulos, "Locally constructed algorithms for distributed computations in ad-hoc networks," in *Proc. Information Processing Sensor Networks (IPSN)*, Berkeley, CA, Apr. 2004, pp. 11–19.

[16] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, pp. 65–78, 2003.

[17] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.