LEARNING BY WEAKLY-CONNECTED ADAPTIVE AGENTS

Bicheng Ying and Ali H. Sayed

Department of Electrical Engineering
University of California, Los Angeles

ABSTRACT

In this paper, we examine the learning mechanism of adaptive agents over weakly-connected graphs and reveal an interesting behavior on how information flows through such topologies. The results clarify how asymmetries in the exchange of data can mask local information at certain agents and make them totally dependent on other agents. A leader-follower relationship develops with the performance of some agents being fully determined by other agents that can even be outside their immediate domain of influence. This scenario can arise, for example, from intruder attacks by malicious agents or from failures by some critical links. The findings in this work help explain why strong-connectivity of the network topology, adaptation of the combination weights, and clustering of agents are important ingredients to equalize the learning abilities of all agents against such disturbances. The results also clarify how weak-connectivity can be helpful in reducing the effect of outlier data on learning performance.

Index Terms— Weakly-connected graphs, distributed strategies, Pareto optimality, leader-follower relationship.

1. INTRODUCTION AND RELATED WORK

Consider a network consisting of N agents connected by a topology. We assign a pair of nonnegative weights, $\{a_{k\ell}, a_{\ell k}\}$, to the edge connecting any two agents k and ℓ. The scalar $a_{\ell k}$ is used by agent k to scale the data it receives from agent ℓ and similarly for $a_{k\ell}$. The network is said to be connected if paths with nonzero scaling weights can be found linking any two distinct agents in both directions. The network is said to be *strongly–connected* if it is connected with at least one self-loop, meaning that $a_{kk} > 0$ for some agent k [1–3].

For each agent k, the combination coefficients are assumed to satisfy:

$$a_{\ell k} \geq 0, \quad \sum_{\ell \in \mathcal{N}_k} a_{\ell k} = 1, \quad a_{\ell k} = 0 \text{ if } \ell \notin \mathcal{N}_k \tag{1}$$

where $\mathcal{N}_k$ denotes the set of neighbors of agent k. We collect the coefficients $\{a_{\ell k}\}$ into an $N \times N$ matrix $A = [a_{\ell k}]$. Then, condition (1) implies that A satisfies $A^\mathsf{T}\mathbb{1} = \mathbb{1}$, so that A is a left-stochastic matrix. Additionally, the strong connectivity of the network implies that A is a primitive matrix (see Lemma 6.1 from [1] or Lemma 8.5.4 from [4]). It then follows from the Perron-Frobenius Theorem [4, 5] that A will have a single eigenvalue at one, with all other eigenvalues lying strictly inside the unit circle. Moreover, if we let p denote the right-eigenvector of A corresponding to its single eigenvalue at one, and normalize its entries to add up to one, then all entries of p will be strictly positive, meaning that p satisfies:

$$Ap = p, \quad \mathbb{1}^\mathsf{T} p = 1, \quad p \succ 0 \tag{2}$$

We refer to p as the Perron eigenvector of A.

We associate with each agent k a twice-differentiable and convex cost function, denoted by $J_k(w) \in \mathbb{R}$, with independent variable $w \in \mathbb{R}^M$. We assume at least one of these costs is strongly-convex. The agents are assumed to run a collaborative distributed strategy of the consensus [6–8] or diffusion type [1–3]. It is sufficient for our purposes in this work to illustrate the results by using the following adapt-then-combine (ATC) diffusion form [1]:

$$\begin{aligned} \boldsymbol{\psi}_{k,i} &= \boldsymbol{w}_{k,i-1} - \mu_k \, \widehat{\nabla_{w^\mathsf{T}} J}_k(\boldsymbol{w}_{k,i-1}) &(3)\\ \boldsymbol{w}_{k,i} &= \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \, \boldsymbol{\psi}_{\ell,i} &(4) \end{aligned}$$

where $\mu_k > 0$ is the step-size used by agent k, the vectors $\{\boldsymbol{\psi}_{k,i}, \boldsymbol{w}_{k,i}\}$ denote iterates at agent k at time i, and the quantity $\widehat{\nabla_w J}_k(\boldsymbol{w}_{k,i-1})$ denotes an approximation for the true gradient vector of $J_k(w)$. The difference between the approximate and true gradient vectors is called gradient noise. We represent the step-sizes as scaled multiples of the same factor $\mu_{\max}$, namely, $\mu_k = \tau_k \mu_{\max}$ with $0 < \tau_k \leq 1$. We also define the vector $q = \text{diag}\{\mu_k\} \cdot p$, with entries q_k, and introduce the strongly-convex weighted aggregate cost:

$$J^{\text{glob},\star}(w) \;\triangleq\; \sum_{k=1}^{N} q_k J_k(w) \tag{5}$$

We denote its unique minimizer by $w^\star$ and the error at each agent relative to $w^\star$ by $\widetilde{\boldsymbol{w}}_{k,i} = w^\star - \boldsymbol{w}_{k,i}$. It was shown in [1, 9] that $w^\star$ serves as a Pareto optimal solution for the network. Specifically, under some reasonable conditions on the cost functions and the gradient noise process, it holds that (see Theoream 9.1 of [1]):

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \;=\; O(\mu_{\max}) \tag{6}$$

where the notation $\alpha = O(\mu)$ means $|\alpha| < c|\mu|$ for some constant $c > 0$. Furthermore, if we let MSD_k denote the size of the mean-square deviation, $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$, in steady-state to first-order in $\mu_{\max}$, and let MSD_{av} denote the average MSD value across all N agents, then it was also shown in [1, 9] that these measures are given by (see Lemma 11.3 of [1]):

$$\text{MSD}_k = \text{MSD}_{\text{av}} = \frac{1}{2}\text{Tr}\left[\left(\sum_{k=1}^{N} q_k H_k\right)^{-1}\left(\sum_{k=1}^{N} q_k^2 G_k\right)\right] \tag{7}$$

This work was supported in part by NSF grants CCF-1011918 and ECCS-1407712. Email: {ybc,sayed}@ucla.edu

where the matrix quantities $\{H_k, G_k\}$ correspond to the Hessian matrix of the cost function and to the covariance matrix of the gradient noise process, respectively, at agent k:

$$H_k \;\; \triangleq \;\; \nabla_w^2 \; J_k(w^\star) \tag{8}$$

$$G_k \;\; \triangleq \;\; \lim_{i\to\infty} \mathbb{E}\left[\boldsymbol{s}_{k,i}(w^\star)\boldsymbol{s}_{k,i}^{\mathsf{T}}(w^\star) \,|\, \boldsymbol{\mathcal{F}}_{i-1} \right] \tag{9}$$

and $\boldsymbol{s}_{k,i}$ denotes the gradient noise process defined as

$$\boldsymbol{s}_{k,i}(\boldsymbol{w}_{i-1}) \;\triangleq\; \widehat{\nabla_{w^{\mathsf{T}}} J}_k(\boldsymbol{w}_{i-1}) \;-\; \nabla_{w^{\mathsf{T}}} J_k(\boldsymbol{w}_{i-1}) \tag{10}$$

In (9), the symbol $\boldsymbol{\mathcal{F}}_{i-1}$ denotes the filtration corresponding to all past iterates $\{\boldsymbol{w}_{\ell,j},\ j \le i-1\}$ across all agents $\ell = 1, 2, \ldots, N$.

2. WEAKLY-CONNECTED NETWORKS

We now examine how these results are modified for *weakly*-connected networks. In doing so, some interesting patterns of behavior are revealed. In loose terms, weakly-connected topologies consist of collections of sub-networks with information flowing only in one direction among some of them. This scenario does not only arise as the result of intruder attacks or asymmetric information exchanges, but can also be the result of failures by some critical links that render the network topology weakly-connected. Among other applications, such networks can be used to model stubborn agent behavior or authoritarian behavior over social networks [10, 11]. Figure 1 illustrates one particular example consisting of three sub-networks, with the number of their agents being denoted by N_1, N_2, and N_3.

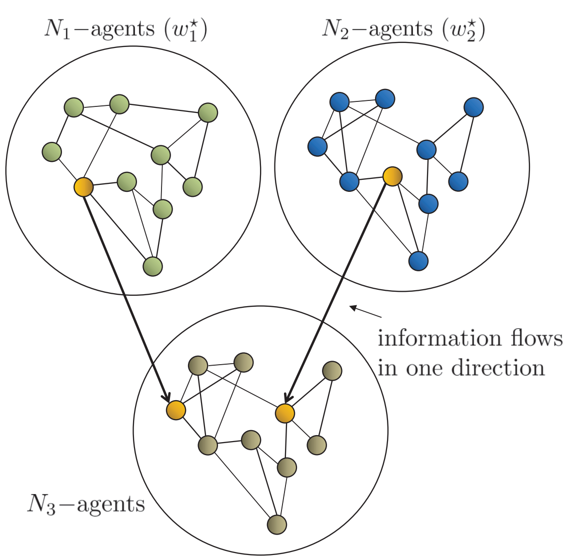

Fig. 1. Illustration of a weakly connected network consisting of three sub-networks.

In the figure, each of the two top sub-networks is strongly-connected and does not receive information from any other sub-network (self-loops are not indicated in the figure). Each of these sub-networks has its own combination policy, denoted by $\{A_1 \in \mathbb{R}^{N_1\times N_1},\ A_2 \in \mathbb{R}^{N_2\times N_2}\}$, and its own Perron vector, now denoted by $\{p_1, p_2\}$. Therefore, if each of these sub-networks were to run the diffusion strategy (3)–(4), then each one of them will independently converge in the mean-square-error sense towards its own Pareto solution, denoted by $\{w_1^\star, w_2^\star\}$. The same figure shows a third sub-network in the bottom, and which appears at the *receiving* end relative to the other sub-networks. The figure shows two arrows emanating in one direction from the top sub-networks towards the bottom sub-network. Therefore, this third sub-network is influenced by the behavior of the top sub-networks, while it does not feed any information back to them. We would like to examine whether the limiting behavior of this third sub-network is ultimately dictated by the two top sub-networks or whether it can still exhibit independent behavior based on its own local data.

More generally, consider a network consisting of a collection of S stand-alone strongly-connected sub-networks. Each of these sub-networks does not receive information from any other sub-network and they can therefore run their diffusion strategies independently of the other sub-networks. We further assume that the network contains a second collection of R sub-networks where some agents in these sub-networks receive information from agents in the first collection. If we refer to Figure 1, then $S = 2$ and $R = 1$. The total number of agents in the network is still denoted by N and it is equal to the sum of the number of agents across all sub-networks.[1]

We collect all weighting coefficients $\{a_{\ell k}\}$ from across all edges into a large $N \times N$ combination matrix $A = [a_{\ell k}]$. Without loss of generality, we assume the agents are numbered with the agents from the union of all S strongly-connected sub-networks coming first, followed by the agents from the remaining R sub-networks. In this way, the matrix A will exhibit the following upper block-triangular structure (if A is not in this irreducible form, then a permutation transformation of the form $P^{\mathsf{T}}AP$ can transform A into the desired form [13, Ch. 8]):

$$\begin{array}{c} \overbrace{\hphantom{A_1 \;\; 0 \;\; \cdots \;\; 0}}^{\text{Subnetworks:}1,2,\ldots,S} \quad \overbrace{\hphantom{A_{1,S+1} \;\; A_{1,S+2} \;\; \cdots \;\; A_{1,S+R}}}^{\text{Subnetworks:}S+1,S+2,\ldots,S+R} \end{array}$$
$$\left[\begin{array}{cccc|cccc} A_1 & 0 & \ldots & 0 & A_{1,S+1} & A_{1,S+2} & \ldots & A_{1,S+R} \\ 0 & A_2 & \ldots & 0 & A_{2,S+1} & A_{2,S+2} & \ldots & A_{2,S+R} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_S & A_{S,S+1} & A_{S,S+2} & \ldots & A_{S,S+R} \\ \hline 0 & 0 & \ldots & 0 & A_{S+1} & A_{S+1,S+2} & \ldots & A_{S+1,S+R} \\ 0 & 0 & \ldots & 0 & 0 & A_{S+2} & \ldots & A_{S+2,S+R} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & A_{S+R} \end{array}\right] \tag{11}$$

In the above expression, the matrices $\{A_1, \ldots, A_S\}$ are the left-stochastic primitive matrices corresponding to the S strongly-connected sub-networks. We denote the size of each matrix A_s by N_s. Likewise, the matrices $\{A_{S+1}, \ldots, A_{S+R}\}$ in the lower right-most block contain the internal combination coefficients for

[1] We remark that our definition of weakly-connected networks is more strict than the terminology used in graph theory. There, a directed graph is called weakly-connected if replacing all of its directed edges with undirected edges produces a connected (undirected) graph [12]. This definition would also include strongly-connected networks as special cases. Our definition is meant to focus exclusively on networks that are truly weakly-connected in that they induce an asymmetric flow of information among some of its components.

the R collection of sub-networks. For example, A_{S+1} contains the coefficients that appear on the edges within sub-network $S+1$; this matrix is *not* left-stochastic because it does not contain all the combination coefficients that are used by the agents within sub-network $S+1$. The entries in the right-most upper block of A contain the combination weights for the edges that emanate from the S sub-networks towards the R sub-networks. For example, for the network shown in Figure 2, one possibility for the combination matrix A is:

$$A = \left[\begin{array}{ccccc|ccc} 0.2 & 0.2 & 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.4 & 0.1 & 0 & 0 & 0.2 & 0 & 0.4 \\ 0.3 & 0.4 & 0.1 & 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.3 & 0.3 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.7 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0.2 & 0.3 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.5 & 0.3 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.2 & 0.1 \end{array}\right] \quad (12)$$

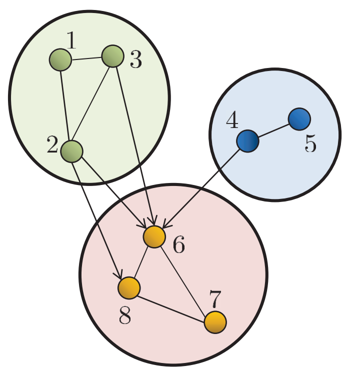

Fig. 2. A weakly connected network consisting of three sub-networks and the corresponding combination policy (12).

3. STEADY-STATE DYNAMICS

This section summarizes the main results that characterize the learning process over weakly-connected networks. Proofs are omitted due to space limitations. We first need to examine the limit of A^n as $n \to \infty$, for matrices A that have the irreducible structure (11). We denote the block structure of A from (11) by

$$A \triangleq \left[\begin{array}{c|c} T_{\mathsf{SS}} & T_{\mathsf{SR}} \\ \hline 0 & T_{\mathsf{RR}} \end{array}\right] \quad (13)$$

where, for example, T_{SS} is block diagonal and consists of the left-stochastic and primitive entries $\{A_1, A_2, \ldots, A_s\}$, while T_{RR} is block upper triangular when $R > 1$ or full matrix when $R = 1$. The block T_{SR} represents the influence of the S sub-networks on the R sub-networks.

Lemma 1 (LIMITING POWER OF A) *Let the Perron eigenvectors of the S strongly-connected sub-networks be denoted by $\{p_s,\ s = 1, 2, \ldots, S\}$. It then holds that:*

$$A_\infty \triangleq \lim_{n\to\infty} A^n = \left[\begin{array}{c|c} \Theta & \Theta W \\ \hline 0 & 0 \end{array}\right] \quad (14)$$

where the matrices Θ and W are defined by

$$W \triangleq T_{\mathsf{SR}}(I - T_{\mathsf{RR}})^{-1} \quad (15)$$

$$\Theta \triangleq \text{blockdiag}\left\{p_1 \mathbb{1}_{N_1}^\mathsf{T}, \ldots, p_S \mathbb{1}_{N_S}^\mathsf{T}\right\} \quad (16)$$

■

The factor W that appears on the right-hand side of (14) is a left-stochastic matrix and admits the following useful interpretation. It can be shown that T_{RR} is a stable matrix and, hence, we can write

$$W = T_{\mathsf{SR}} + T_{\mathsf{SR}}T_{\mathsf{RR}} + T_{\mathsf{SR}}T_{\mathsf{RR}}^2 + \ldots \quad (17)$$

Now note that the first term in (17) represents the information that is transferred from group S into group R, while the second term in (17) represents how this information is transformed internally within group R after one step, and similarly for the subsequent terms in (17) involving higher-order powers of T_{RR}.

3.1. Limit Points for Group S of Sub-Networks

Let $\{w_s^\star,\ \ s = 1, 2, \ldots, S\}$ denote the Pareto optimal solutions for the strongly-connected sub-networks in group S. Consider sub-network s from this group. It has N_s agents and its agents' step-sizes will be denoted by $\{\mu_{s,k}\}$, with the first subscript referring to the sub-network and the second subscript referring to the agent. Likewise, the Perron vector of sub-network s is denoted by p_s with individual entries $\{p_{s,k}\}$. The associated scaled weights are denoted by:

$$q_{s,k} \triangleq \mu_{s,k} p_{s,k}, \quad k = 1, 2, \ldots, N_s \quad (18)$$

Now, the Pareto solution, $w_s^\star$, that corresponds to sub-network s is the unique solution to the following algebraic equation:

$$\sum_{k=1}^{N_s} q_{s,k} \nabla_{w^\mathsf{T}} J_{s,k}(w_s^\star) = 0 \quad (19)$$

where the $\{J_{s,k}(w)\}$ denote the cost functions that are associated with the agents k within sub-network s. Each agent in the sub-network s will converge towards $w_s^\star$ within $O(\mu_{\max})$. In other words, the limit point will be uniform within each sub-network; though the limit points can be distinct across the sub-networks. Collecting the Pareto solutions $\{w_s^\star\}$ from across the S sub-networks, we find that the limiting points for all agents within group S are described by the following extended vector:

$$\mathcal{W}^\star \triangleq \text{col}\{\mathbb{1}_{N_1} \otimes w_1^\star,\ \ldots,\ \mathbb{1}_{N_s} \otimes w_S^\star\} \quad (20)$$

where notation $\mathbb{1}_{N_1}$ means all-one-vector with length N_1, and notation $\otimes$ represents the kronecker product.

3.2. Limit Points for Group R of Sub-Networks

Now consider an arbitrary sub-network r from group R. It turns out that, contrary to the uniform behavior observed for group S, each agent within the sub-network r will converge to an *individual* limit point and the values of these points are determined by the sub-networks in group S (as shown below by (23)).

We denote the limiting value for each agent k in sub-network r by $w_{r,k}^\bullet$, for $k = 1, 2, \ldots, N_r$. In this way, the collection of limit points for each sub-network r will be

$$\mathcal{W}_r^\bullet = \text{col}\{w_{r,1}^\bullet,\ \ldots,\ w_{r,N_r}^\bullet\} \quad \text{(sub-network } r) \quad (21)$$

and the collection of limit points for all R sub-networks is

$$w^{\bullet} \triangleq \text{col}\{w_1^{\bullet}, w_2^{\bullet}, \ldots, w_R^{\bullet}\} \tag{22}$$

We can identify the limit vector $w^{\bullet}$ for the sub-networks from group R in terms of the Pareto solutions of the sub-networks from group S. The argument is omitted but the key result is that it will hold that:

$$w^{\bullet} = \mathcal{W}^{\mathsf{T}} w^{\star} \tag{23}$$

where $\mathcal{W} = W \otimes I_M$. Recall that W is left-stochastic, which implies that if all S sub-networks happen to have the same limit point, say, $w^{\star}$, then all agents in group R will also converge to this same limit point. More generally, if sub-networks within group S have different limit points, $w_s^{\star}$, then each agent in group R will usually converge towards a different limit point, $w_{r,k}^{\bullet}$. We can assess the steady-state mean-square deviation for each agent, defined as the value of the error variance relative to its limit point, $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$, as $i \to \infty$ and for sufficiently small step-sizes. For agents in group R, we introduce the $S \times 1$ column vector:

$$c_k = \text{blockdiag}\{\mathbb{1}_{N_1}^{\mathsf{T}}, \mathbb{1}_{N_2}^{\mathsf{T}}, \ldots, \mathbb{1}_{N_S}^{\mathsf{T}}\} \cdot [W]_{:,k} \tag{24}$$

where the notation $[W]_{:,k}$ refers to the column of W corresponding to agent k. Recall that W is a left-stochastic matrix and, therefore, the entries of c_k will add up to one. In addition, as expression (26) below reveals, the square of the $s-$th entry of c_k will measure the influence of sub-network s from group S on the performance of agent k from group R.

Theorem 1 (MSD PERFORMANCE) *Assume agent k belongs to a sub-network s from group S. Then, all agents within sub-network s achieve the same MSD level given by*

$$\text{MSD}_s = \frac{1}{2}\text{Tr}\left[\left(\sum_{k=1}^{N_s} q_{s,k} H_{s,k}\right)^{-1}\left(\sum_{k=1}^{N_s} q_{s,k}^2 G_{s,k}\right)\right] \tag{25}$$

where we are using the notation $\{H_{s,k}, G_{s,k}\}$, with a subscript s, to denote the Hessian and covariance matrices (8)–(9) for agent k within sub-network s. Assume, on the other hand, that agent k belongs to a sub-network r from group R. Then, in this case, it holds that

$$\text{MSD}_k = \sum_{s=1}^{S} c_k^2(s) \cdot \text{MSD}_s \tag{26}$$

where $c_k(s)$ denotes the $s-$th entry of vector c_k. That is, the performance of any agent from group R is given by a weighted combination of the MSD performance levels of the sub-networks from group S. ■

In conclusion, the findings established in this article help explain why strong-connectivity of the network topology [1,2] adaptation of the combination weights [1,3] and clustering of agents [14–16] are important ingredients to safeguard against such pitfalls. The results also clarify how weak-connectivity is helpful in reducing the effect of outlier data when all agents in a network are interested in the same objective, as illustrated next.

4. APPLICATION EXAMPLE

We consider an example based on the topology from Fig. 2 to illustrate that a weakly-connected network can sometimes achieve better performance than strongly-connected networks in the presence of outliers in the data. The objective is to determine an elliptical curve that separates the data into two classes: class +1 consists of data that are concentrated inside the curve and class -1 consists of data that are concentrated outside the curve. We assume about 10% of the data available to sub-networks R are outliers. The outlier data belong to class +1 but are located away from origin. Obviously, since we are dealing with a weakly-connected network, agents in group S will not be affected by these outliers.

We assume each agent employs the logistic cost function:

$$J_k(w) = \rho\|w\|^2 \; + \; \mathbb{E}\left\{\ln[1 + e^{-\boldsymbol{\gamma}_k \boldsymbol{h}_k^{\mathsf{T}} w}]\right\} \tag{27}$$

where $\boldsymbol{\gamma}_k$ represents the class label $\{+1, -1\}$, ρ is a regularization parameter, and the feature vector $\boldsymbol{h}_k$ is chosen as follows in terms of the $(\boldsymbol{x}, \boldsymbol{y})$ coordinates of each data point [17, 18]:

$$\begin{aligned} &\boldsymbol{h}_k(1) = 1, && \boldsymbol{h}_k(2) = \boldsymbol{x}, && \boldsymbol{h}_k(3) = \boldsymbol{y} \\ &\boldsymbol{h}_k(4) = \boldsymbol{x}^2, && \boldsymbol{h}_k(5) = \boldsymbol{y}^2, && \boldsymbol{h}_k(6) = \boldsymbol{x}\boldsymbol{y} \end{aligned}$$

The gradient vector of $J_k(w)$is approximated by using

$$\widehat{\nabla_{w^{\mathsf{T}}} J}_k(\boldsymbol{w}_{k,i-1}) = \frac{-\,\boldsymbol{\gamma}_k(i)\boldsymbol{h}_{k,i}}{1 + e^{\boldsymbol{\gamma}_k(i)\boldsymbol{h}_{k,i}^{\mathsf{T}}\boldsymbol{w}_{k,i-1}}} + \rho\boldsymbol{w}_{k,i-1} \tag{28}$$

We compare the performance of the weakly-connected topology against a strongly-connected (actually, fully-connected) network with combination matrix: $A = \frac{1}{8}\mathbb{1}_8\mathbb{1}_8^{\mathsf{T}}$. In Fig. 3, we observe that the black elliptic curve, which is the result obtained by the strongly-connected network, is larger than the green elliptic curve obtained by the weakly-connected network. Comparing both boundary curves, we find that the black curve includes a larger proportion of class -1 data, which will be mistakenly inferred as belonging to class +1.

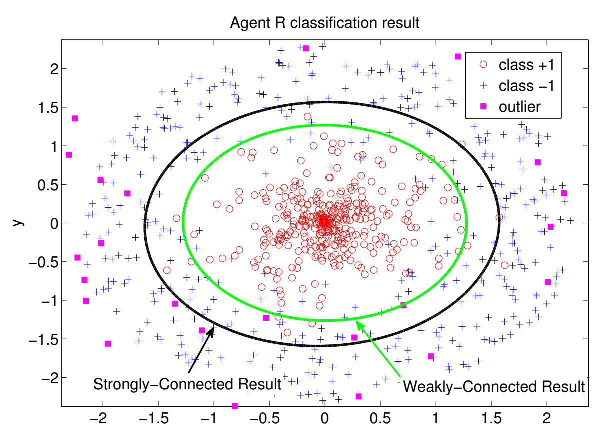

Fig. 3. Logistic classification result using an elliptic separation curve. In the simulation, sub-networks R suffer from outlier data.

5. REFERENCES

[1] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4–5, pp. 311–801, NOW Publishers, Boston-Delft, 2014.

[2] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

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

[4] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 2003.

[5] S. U. Pillai, T. Suel, and S. Cha, "The Perron–Frobenius theorem: Some of its applications," *IEEE Signal Processing Magazine*, vol. 22, no. 2, pp. 62–75, 2005.

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

[7] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[8] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed consensus algorithms in sensor networks: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, 2009.

[9] J. Chen and A. H. Sayed, "Distributed Pareto optimization via diffusion strategies," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 2, pp. 205–220, April 2013.

[10] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, vol. 1, no. 1, pp. 3–49, 2011.

[11] E. Yildiz, D. Acemoglu, A. Ozdaglar, A. Saberi, and A. Scaglione, "Binary opinion dynamics with stubborn agents," *ACM Trans. Econ. Comput.*, vol. 1, no. 4, pp. 19:1–19:30, Dec. 2013.

[12] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, Macmillan London, 1976.

[13] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, PA, 2000.

[14] X. Zhao and A. H. Sayed, "Distributed clustering and learning over networks," *Submitted for publication*. Also available as arXiv:1409.6111[cs.MA], Sep. 2014.

[15] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4144, Aug. 2014.

[16] X. Zhao and A. H. Sayed, "Clustering via diffusion adaptation over networks," in *Proc. International Workshop on Cognitive Information Processing (CIP)*, Parador de Baiona, Spain, May 2012, pp. 1–6.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 1, Springer, NY, 2006.

[18] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 4th edition, 2008.