**Preprints of the 19th World Congress**
**The International Federation of Automatic Control**
**Cape Town, South Africa. August 24-29, 2014**

# Focused First-Followers Accelerate Aligning Followers with the Leader in Reaching Network Consensus ⋆

**Ming Cao** * **Alex Olshevsky** ** **Weiguo Xia** ***

* *Faculty of Mathematics and Natural Sciences, ITM, University of Groningen, the Netherlands (e-mail: m.cao@rug.nl).*
** *Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana Champaign, USA (e-mail: aolshev2@illinois.edu)*
*** *ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Sweden, (e-mail: weiguo@kth.se)*

**Abstract:** This paper proposes and analyzes a new strategy to accelerate the process of reaching consensus in leader-follower networks. By removing or weakening specific *directed* couplings pointing to the first followers from the other followers, we prove that all the followers' states converge faster to that of the leader. This result is in sharp contrast to the well known fact that when the followers are coupled together through *undirected* links, removing or weakening links always decelerate the converging process. Simulation results are provided to illustrate this subtle, yet somewhat surprising, provably correct result.

*Keywords:* convergence speed, leader-follower network, accelerating consensus

## 1. INTRODUCTION

Although variations of consensus algorithms as a typical class of distributed coordination algorithms for multi-agent networks have been studied extensively in the past years {Jadbabaie et al. (2003); Olfati-Saber et al. (2007); Cao et al. (2008a); Ren and Cao (2011)}, one central issue that remains open and demands more in-depth investigation is how to accelerate the convergence speed for such algorithms when the scales of the networks are large {Blondel et al. (2009); Olshevsky and Tsitsiklis (2009); Cao et al. (2008b); Nedic and Ozdaglar (2010); Olshevsky and Tsitsiklis (2011)}. In {Xiao and Boyd (2004)}, centralized semidefinite programming is utilized to find the optimal combinations of the weights for distributed weighted averaging; modified gossiping algorithms have been discussed when shift-registers are installed to store each agent's past values {Cao et al. (2006); Oreshkin et al. (2010); Liu et al. (2013)}. Various heuristics, e.g. using prediction mechanisms, have been discussed as well {Zhang et al. (2009)}.

Researchers have been especially interested in the process of aligning followers with the leaders in multi-agent networks when some agents are taking the role of leaders that guide the followers to reach consensus {Jadbabaie et al. (2003); Cao et al. (2008b); Scardovi and Sepulchre (2009); Ni and Cheng (2010)}. Since in leader-follower networks, the consensus, if reached in the end, will be the state of the leader, it is relatively easier to characterize how fast a follower's state approaches its asymptotic value. As a result, deeper insight can be obtained into what affects the convergence speed of the consensus algorithms and how one might come up with new strategies to accelerate the converging processes.

In this paper, we study the problem of accelerating the process of reaching consensus in leader-follower networks. We propose a new strategy based on an intuitive idea rooted in understanding the hierarchical differences for the followers in terms of their topological distances to the leaders in the network. We pay special attention to those followers who can sense or acquire directly the leaders' information and call such followers *first followers.* Although in *undirected* multi-agent networks, stronger or more links between followers always accelerate convergence {Xiao and Boyd (2004)}, in *directed* networks, the convergence speed changes in more complicated fashions {Cao et al. (2008b)}. For the first followers, their information about the leader's state is the most critical for them to reach consensus while their information about the other followers may very likely differ or even contradict that of the leaders. Based on this observation, we claim that if the first followers are more focused on the leaders and cut or weaken their links to the other followers, the convergence process of *all* the followers may get accelerated. Although this claim is plausible at least intuitively, to prove it rigourously requires quite some involved arguments. It is the main contribution of this paper to prove mathematically how the new simple strategy accelerates reaching network consensus.

The rest of the paper is organized as follows. In Section 2, we introduce the leader-follower network model and formulate the convergence speed problem. In Section 3,

⋆ The authors are listed in alphabetic order. The work of Cao is supported in part by the European Research Council (ERC-StG-307207) and the EU INTERREG program under the auspices of the SMARTBOT project.

we give rigorous proof to show that the convergence is accelerated when directed links pointing to the first followers and starting from the other followers are removed or weakened. In Section 4, we give a simulation example to illustrate the theoretical results.

## 2. PROBLEM FORMULATION

Consider a network of $N+1$ agents that consists of a leader, labeled by 0, and $N$ followers, labeled by $1, \ldots, N$. The couplings between the followers are equally weighted (we will relax this later) but *not* necessarily symmetric, and thus need to be described by a directed graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$, whose vertex set is $\mathcal{V} = \{1, 2, \ldots, N\}$. There are no self-couplings for the agents and correspondingly no self-loops in $\mathbb{G}$. So an element $a_{ij}$ of the nonnegative adjacency matrix $A = (a_{ij})_{N \times N}$ of $\mathbb{G}$ is one if and only if $i \neq j$ and $(j, i) \in \mathcal{E}$. In the rest of the paper, we assume that $\mathbb{G}$ is strongly connected and that the leader can influence at least one follower directly. Let $d_i = \sum_{j=1, j \neq i}^{N} a_{ij}$ be the in-degree of vertex $i$ and the diagonal matrix $D = \text{diag}(d_1, \ldots, d_N)$ be the in-degree matrix. The Laplacian matrix $L$ of $\mathbb{G}$ is given by $L = D - A$.

The state of the leader $x_0 \in \mathbb{R}$ is fixed at a constant value $c$, which in real networks may correspond to a preferred migration direction of a flock of birds, a strong opinion of a leading figure of an online community, or a prescribed rendezvousing location of a team of robots. All the followers have to interact with the other peer agents so that the values of their states $x_i \in \mathbb{R}$, $1 \leq i \leq N$, converge to $x_0 = c$ as the networked system evolves. To be more specific, the updating rules of the followers are described by

$$\dot{x}_i = \sum_{j=0, j \neq i}^{N} a_{ij}(x_j - x_i), \quad i = 1, \ldots, N, \tag{1}$$

where $a_{i0} = \epsilon > 0$ if the leader influences agent $i$ directly and zero otherwise, and $a_{ij} = 1$ if $j \neq i, 0$ and $(j, i) \in \mathcal{E}$ and zero otherwise. While it has been well established that when $\mathbb{G}$ is connected, the values of all the agents will converge exponentially fast to $x_0$ {Jadbabaie et al. (2003); Cao et al. (2008b)}, the convergence can slow down significantly as the scale of the network grows. So it is of great interest to explore how to speed up the convergence using as little global information as possible, which rules out the approaches optimizing the topologies of $\mathbb{G}$ globally in order to maximize the smallest nonzero eigenvalues of the Laplacian matrices of the leader-follower networks.

In this paper, we focus on the potential role of the *first followers*, those who have direct access to the information about the leader, in accelerating the convergence speed. It is the goal to prove rigourously that when the first followers focus more on the leader and pay less attention to the other followers, the values of all the followers will converge faster to that of the leader. Note that this underscores the subtle fact that stronger or more couplings between the followers do *not* necessarily lead to faster convergence.

Towards this end, we rearrange the labels of the followers such that the first followers are labeled by $1, \ldots, l$, where $l \geq 1$ is the number of the first followers, and the other followers are labeled by $l+1, \ldots, N$. Note that we have used $\epsilon > 0$ to denote the gain that the first followers can choose reflecting their attention to the leader; and the $\epsilon$'s are the same for all the first followers. Hence, the dynamics of the followers can be written into

$$\begin{cases} \dot{x}_i = \displaystyle\sum_{j=1, j \neq i}^{N} a_{ij}(x_j - x_i) + \epsilon(x_0 - x_i), \quad i = 1, \ldots, l; \\ \dot{x}_i = \displaystyle\sum_{j=1, j \neq i}^{N} a_{ij}(x_j - x_i), \quad i = l+1, \ldots, N. \end{cases} \tag{2}$$

Let $y_i = x_i - x_0$ be the difference between the state of the follower $i$ and that of the leader, then the dynamics of the $y_i$-system are

$$\begin{cases} \dot{y}_i = \displaystyle\sum_{j=1, j \neq i}^{N} a_{ij}(y_j - y_i) - \epsilon y_i, \quad i = 1, \ldots, l, \\ \dot{y}_i = \displaystyle\sum_{j=1, j \neq i}^{N} a_{ij}(y_j - y_i), \quad i = l+1, \ldots, N, \end{cases} \tag{3}$$

or in a compact form

$$\dot{y} = -(L + E)y, \tag{4}$$

where $y = [y_1 \ \cdots \ y_N]^T$ and $E$ is the $N$-dimensional diagonal matrix whose first $l$ elements are $\epsilon$ and the rest are zero.

Since $\mathbb{G}$ is strongly connected, $-(L+E)$ is Hurwitz and thus $y$ converges to zero as time goes to infinity. So the states of all the followers converge to that of the leader and the convergence speed is determined by that eigenvalue of $-(L+E)$ that has the largest real part. In the next section, we explain and prove our idea that by making the first followers more focused on the influence from the leader and less distracted by the influence from the other followers, one may accelerate the convergence of the states of all the followers.

## 3. MAIN RESULT

We start with the case when there is only one first follower and then look into the more general case when there are multiple first followers.

### *3.1 The case with one first-follower*

Now the dynamics of the $y$-systems (3) and (4) become

$$\begin{cases} \dot{y}_1 = \displaystyle\sum_{j=2}^{N} a_{1j}(y_j - y_1) - \epsilon y_1, \\ \dot{y}_i = \displaystyle\sum_{j=1, j \neq i}^{N} a_{ij}(y_j - y_i), \quad i = 2, \ldots, N, \end{cases} \tag{5}$$

and

$$\dot{y} = -(L + E)y, \tag{6}$$

where $E = \text{diag}(\epsilon, 0, \ldots, 0)$. The idea of accelerating the convergence is illustrated by a four-follower example shown in Fig. 1. The first-follower, agent 1, was originally influenced by both the leader and the followers 3 and 4; by deleting the edges $(3, 1)$ and $(4, 1)$ in the graph, we make agent 1 get more focused on using the information from the leader.

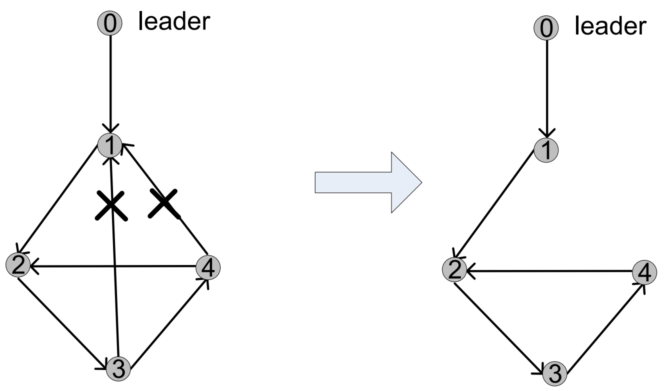


Fig. 1. An illustrative example showing that the first-follower neglects the information received from the other followers.

Before looking into the property of the solution to system (5), we first introduce some lemmas and notions that will be used often in the rest of the paper.

Let $M = (m_{ij})_{N\times N}$ be a real matrix. We write $M \geq 0$ if $m_{ij} \geq 0,\ i,j = 1,\ldots,N$, and such a matrix $M$ is called a *nonnegative* matrix.

*Lemma 1.* (Horn and Johnson (1985)) If $M \geq 0$ is a nonnegative matrix, then the spectral radius $\rho(M)$ of $M$ is an eigenvalue of $M$ and there is a nonnegative vector $x \geq 0,\ x \neq 0$, such that $Mx = \rho(M)x$. If $M$ is nonnegative and irreducible, then $\rho(M)$ is a simple eigenvalue of $M$ and there is a positive vector $x > 0$ such that $Mx = \rho(M)x$.

*Lemma 2.* (Horn and Johnson (1985)) Let $M_1$ be an irreducible nonnegative matrix. If $M_2 \geq 0$ and $M_2 \neq 0$ then $\rho(M_1 + M_2) > \rho(M_1)$.

*Lemma 3.* (Horn and Johnson (1985)) Suppose $M \in \mathbb{R}^{N\times N}$ and $M \geq 0$. Then

$$\min_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij} \leq \rho(M) \leq \max_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij}. \qquad (7)$$

*Lemma 4.* If $M \geq 0$ and $M$ is irreducible, and $\min_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij} < \max_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij}$, then

$$\min_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij} < \rho(M) < \max_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij}. \qquad (8)$$

*Proof.* Let $\max_{1\leq i\leq N} \sum_{j=1}^{N} m_{ij} = \alpha$ and construct a new matrix $B$ with $b_{ij} = \alpha \frac{m_{ij}}{\sum_{j=1}^{N} m_{ij}}$. Then $B \geq M$, and $\sum_{j=1}^{N} b_{ij} = \alpha$ for all $i = 1,\ldots,N$, implying $\rho(B) = \alpha$. Since $B - M \geq 0$, $B - M \neq 0$, and $M$ is irreducible, from Lemma 2, one knows $\rho(M) < \rho(B) = \alpha$. The lower bound can be established in a similar manner. □

*Lemma 5.* (Horn and Johnson (1985)) If $M \in \mathbb{R}^{N\times N}$ and $M \geq 0$, then

$$\rho(M) = \max_{\substack{x\geq 0\\ x\neq 0}} \min_{x_i\neq 0} \frac{(Mx)_i}{x_i} = \max_{\substack{x\geq 0\\ x^T x=1}} \min_{x_i\neq 0} \frac{(Mx)_i}{x_i}, \qquad (9)$$

where $(Mx)_i$ is the $i$th element of the vector $Mx$.

*Lemma 6.* Let $M \in \mathbb{R}^{N\times N}$ be an irreducible nonnegative matrix. Then there is a unique vector $x^* \in \{x|x \geq 0,\ x^T x = 1\}$ such that $x^* > 0$ and

$$\rho(M) = \min_i \frac{(Mx^*)_i}{x_i^*} = \frac{(Mx^*)_i}{x_i^*},\ i = 1,\ldots,N, \qquad (10)$$

and for any $y \in \{x|x \geq 0,\ x^T x = 1\},\ y \neq x^*$,

$$\rho(M) > \min_{y_i\neq 0} \frac{(My)_i}{y_i}. \qquad (11)$$

The proof is omitted here due to the length limit.

An $N \times N$ real matrix $M$ with nonnegative off-diagonal elements $m_{ij},\ i \neq j$, is called *essentially nonnegative* {Cohen (1981)}. Such an $M$ has an eigenvalue $r(M)$, called the *dominant eigenvalue*, which is real and greater than or equal to the real part of any other eigenvalue of $M$. This can be seen by adding a positive scalar multiple of the identity matrix $\alpha I$ to $M$ such that $M + \alpha I$ is nonnegative. From Lemma 1, one knows that $M + \alpha I$ has an eigenvalue $\rho(M + \alpha I)$, which is real and greater than or equal to the real part of any other eigenvalue of $M + \alpha I$. Letting $r(M) = \rho(M + \alpha I) - \alpha$, one obtains the dominant eigenvalue of $M$.

It is obvious that both $-L$ and $-(L + E)$ are essentially nonnegative. Let $d_{\max} = \max_{1\leq i\leq N} d_i$, $\alpha = d_{\max} + \epsilon + 1$, and $P = -(L + E) + \alpha I$. $P$ is nonnegative and has an eigenvalue $\rho(P)$. Since the graph $\mathbb{G}$ is strongly connected, the matrices $L$ and $P$ are irreducible. Since $E \geq 0,\ E \neq 0$, in view of Lemma 2, we know

$$\rho(P) < \rho(P + E) = \rho(-L + \alpha I) = \alpha.$$

Thus

$$r(-(L + E)) = \rho(P) - \alpha < 0,$$

which implies that the eigenvalues of $-(L + E)$ all have negative real parts. The solution $y$ to system (6) converges to 0 as $t \to \infty$. Thus $r(-(L + E))$ is one of the indices characterizing the convergence speed.

When agent 1 only uses the information about the leader and neglects the information about the other followers, system (5) becomes

$$\begin{cases} \dot{y}_1 = -\epsilon y_1, \\ \dot{y}_i = \displaystyle\sum_{j=1, j\neq i}^{N} a_{ij}(y_j - y_i), \quad i = 2,\ldots,N, \end{cases} \qquad (12)$$

or in a compact form

$$\dot{y} = -(\bar{L} + E)y, \qquad (13)$$

where $\bar{L}$ is the same as $L$ except that the first row of $\bar{L}$ is equal to $\mathbf{0}$. Let the Laplacian matrix $L$ be partitioned into

$$L = \begin{bmatrix} l_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix},$$

where $l_{11} = d_1$, $L_{21}, L_{12}^T \in \mathbb{R}^{N-1}$ and $L_{22} \in \mathbb{R}^{(N-1)\times(N-1)}$. We then have

$$\bar{L} = \begin{bmatrix} 0 & \mathbf{0} \\ L_{21} & L_{22} \end{bmatrix},\ -(L + E) = \begin{bmatrix} -l_{11} - \epsilon & -L_{12} \\ -L_{21} & -L_{22} \end{bmatrix},$$

$$-(\bar{L} + E) = \begin{bmatrix} -\epsilon & \mathbf{0} \\ -L_{21} & -L_{22} \end{bmatrix}.$$

The following theorem discloses the relationship between the dominant eigenvalue of $-(\bar{L} + E)$ and $-(L + E)$, which guarantees a faster convergence speed for system (13) compared with system (6).

*Theorem 1.* If the graph $\mathbb{G}$ is strongly connected, then

$$\max\{-\epsilon, -1\} \le r(-(\bar{L}+E)) < r(-(L+E)) < 0. \quad (14)$$

*Proof.* Let $d_{\max} = \max_{1\le i\le N} d_i$, $\alpha = d_{\max} + \epsilon + 1$, $P = -(L+E) + \alpha I$ and $\bar{P} = -(\bar{L}+E) + \alpha I$. Therefore, $P$ and $\bar{P}$ are nonnegative matrices. In view of Lemma 1, $\rho(P)$ and $\rho(\bar{P})$ are eigenvalues of $P$ and $\bar{P}$, respectively. To prove (14), it suffices to prove

$$\max\{\alpha - \epsilon, \alpha - 1\} \le \rho(\bar{P}) < \rho(P) < \alpha. \quad (15)$$

Since $P$ is irreducible and

$$\alpha - \epsilon = \min_{1\le i\le N} \sum_{j=1}^{N} p_{ij} < \max_{1\le i\le N} \sum_{j=1}^{N} p_{ij} = \alpha,$$

from Lemma 4, we know that $\alpha - \epsilon < \rho(P) < \alpha$. It is easy to see from the structure of $\bar{P}$ that the spectrum of $\bar{P}$ is $\sigma(\bar{P}) = \{\alpha - \epsilon, \sigma(\bar{P}_{22} = -L_{22} + \alpha I)\}$. Suppose the graph $\bar{\mathbb{G}}$, obtained by deleting vertex 1 and the edges starting or ending at this vertex in $\mathbb{G}$, is still connected (otherwise apply the arguments to each connected component of $\bar{\mathbb{G}}$ respectively). It follows that $\rho(\bar{P}) = \max\{\alpha - \epsilon, \rho(\bar{P}_{22})\}$ and there is a positive vector $y > 0$, $y^T y = 1$ such that $\bar{P}_{22} y = \rho(\bar{P}_{22}) y$ and $\rho(\bar{P}_{22}) = \min_{y_i \ne 0} \frac{(\bar{P}_{22} y)_i}{y_i}$.

Let $x^*$ be the unique positive vector satisfying $x^* \ge 0$, $x^{*T} x^* = 1$ such that (10) is satisfied for $P$. Let $z = [0, y^T]^T$. Since $P_{22} = -L_{22} + \alpha I = \bar{P}_{22}$ and $z \ne x^*$, from (11) in Lemma 6, one has

$$\rho(P) = \min_{x_i^* \ne 0} \frac{(Px^*)_i}{x_i^*} = \max_{\substack{x\ge 0\\ x^T x = 1}} \min_{x_i \ne 0} \frac{(Px)_i}{x_i}$$

$$> \min_{z_i \ne 0} \frac{(Pz)_i}{z_i} = \min_{y_i \ne 0} \frac{(\bar{P}_{22} y)_i}{y_i} = \rho(\bar{P}_{22}).$$

Combining with the fact that $\rho(P) > \alpha - \epsilon$, we arrive at the conclusion that $\rho(\bar{P}) < \rho(P)$.

From

$$\alpha - 1 = \min_{1\le i\le N-1} \sum_{j=1}^{N-1} (\bar{P}_{22})_{ij} \le \rho(\bar{P}_{22})$$

$$\le \max_{1\le i\le N-1} \sum_{j=1}^{N-1} (\bar{P}_{22})_{ij} \le \alpha,$$

we conclude that the inequality (15) holds. □

In Fig. 1, we have removed all the edges pointing to vertex 1 in $\mathbb{G}$, which gives rise to the special structure of the matrix $-(\bar{L}+E)$ which has only one nonzero element in the first row. Now we show that if we only remove some of the edges pointing to vertex 1, but not all of them, i.e., agent 1 only neglects some of the information received from the other followers, the convergence process will still be accelerated. For example, in Fig. 2, we illustrate the situation when only the edge $(3,1)$ is removed from the graph.

Assume that there are more than one edge pointing to vertex 1 in $\mathbb{G}$ and only one of these edges is deleted, say $(k,1)$. The resulted graph after deleting $(k,1)$ is still strongly connected and the corresponding Laplacian matrix is denoted by $\bar{L}$. Let $P = -(L+E) + \alpha I$ and $\bar{P} = -(\bar{L}+E)+\alpha I$. Since only the $(1,1)$ and $(1,k)$ elements of $P$ and $\bar{P}$ are different, one has

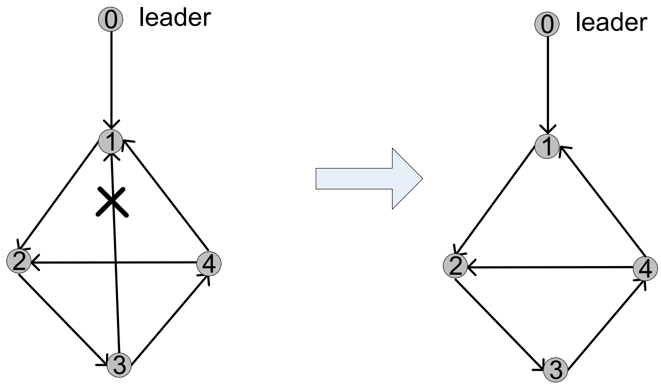


Fig. 2. The edge $(3,1)$ is deleted to accelerate the convergence.

$$\bar{p}_{11} = p_{11} + p_{1k},\ \bar{p}_{1k} = 0,\ \bar{p}_{1j} = p_{1j},\ \text{for } j \ne 1, k,$$

and

$$\bar{p}_{ij} = p_{ij},\ i = 2, \ldots, N,\ j = 1, \ldots, N.$$

We now show that $r(-(\bar{L}+E)) < r(-(L+E))$, which is equivalent to $\rho(\bar{P}) < \rho(P)$.

*Theorem 2.* Assume that the graph $\mathbb{G}$ is strongly connected. Then

$$\max\{-\epsilon, -1\} \le r(-(\bar{L}+E)) < r(-(L+E)) < 0. \quad (16)$$

The proof of this theorem makes use of the following lemma.

*Lemma 7.* Let $M = (m_{ij})_{N\times N}$ be a nonnegative irreducible matrix. Assume that there exists some positive number $r$ such that for some $k$, $1 \le k \le N$, $\sum_{j=1}^{N} m_{kj} < r$ and for all $i \ne k$, $1 \le i \le N$, $\sum_{j=1}^{N} m_{ij} = r$. Assume $x$ is a positive vector such that $Mx = \rho(M)x$. Then $x_k < \min_{\substack{i\ne k\\ i=1,\ldots,N}} \{x_i\}$.

For a nonnegative irreducible matrix $M$ that satisfies the conditions specified in Lemma 7, we are able to write equation (9) into

$$\rho(M) = \max_{\substack{x\ge 0\\ x^T x=1}} \min_{x_i\ne 0} \frac{(Mx)_i}{x_i} = \max_{\substack{x>0, x^T x=1\\ x_k < \min\{x_i\}\\ i\ne k}} \min_{i} \frac{(Mx)_i}{x_i}. \quad (17)$$

In virtue of Lemma 7, $\bar{x}_1^* < \min_{i=2,\ldots,N} \{\bar{x}_i^*\}$ if $\bar{x}^*$ is a positive eigenvector associated with the eigenvalue $\rho(\bar{P})$ of $\bar{P}$. The proof of Theorem 2 makes use of this fact and (17) and is omitted here due to length limit.

Theorem 2 also shows that when more and more edges from the other followers to the first follower are removed, $r(-(L+E))$ becomes smaller and smaller and thus the convergence is always accelerated. So we have proved the following theorem.

*Theorem 3.* Consider the two leader-follower networks (6) and (13) with one first follower, the latter is obtained from the former by letting the first follower neglect the information from one or more of its neighboring followers. The convergence speed of (13) is faster than that of (6).

In the next subsection, we study the more general case when there are multiple first followers.

*3.2 The case with multiple first followers*

Now the number of the first followers $l > 1$. If we remove in $\mathbb{G}$ some or all the edges starting from the other followers to the first followers, then $r(-(L+E))$ decreases and the convergence is accelerated. The proof is similar to that of Theorem 2 and relies on the following result, which is a general form of Lemma 7.

*Lemma 8.* Let $M = (m_{ij})_{N\times N}$ be a nonnegative irreducible matrix and let $\mathcal{N}$ be a nonempty subset of $\{1,\ldots,N\}$. Assume that there exists some positive number $r$ such that

$$\begin{cases} \sum_{j=1}^{N} m_{ij} < r, & i \in \mathcal{N}, \\ \sum_{j=1}^{N} m_{ij} = r, & i \in \{1,\ldots,N\}\backslash\mathcal{N}. \end{cases}$$

Let $x$ be the positive vector such that $Mx = \rho(M)x$. Then $\max_{i\in\mathcal{N}}\{x_i\} < \min_{i\in\{1,\ldots,N\}\backslash\mathcal{N}}\{x_i\}$.

*Remark 1.* The above discussions have assumed that the weights of the edges are all 1 and shown that by removing some of the edges starting from the other followers to the first followers, the convergence of the leader-follower network is accelerated. In fact, the weights of the edges can be any positive numbers. When the weights of the chosen edges starting from the other followers to the first followers decrease, the convergence process is also accelerated. We will develop a general result on accelerating the convergence in the full-length version of the paper.

*Remark 2.* In addition, one can gain insight into how the convergence process in the leader-follower network might get slowed down. The reasons might be that new couplings have been established from the other followers to the first followers, or the weights of such couplings have been increased.

*Remark 3.* If we remove one or some of the edges between the first followers, the convergence is not guaranteed to be accelerated, and actually can be decelerated. We give the following example to illustrate.

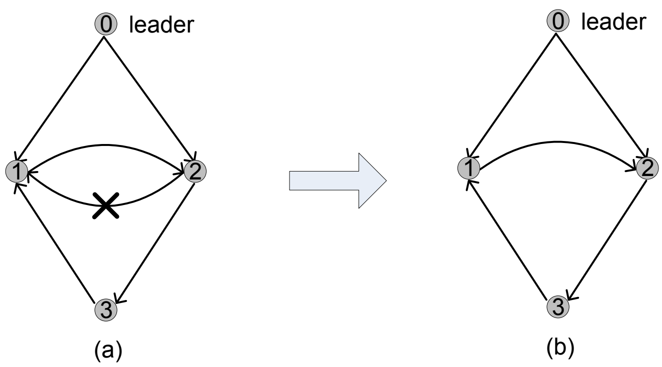


Fig. 3. The edge $(2,1)$ between two first-followers is deleted.

*Example.* Consider a leader-follower network shown in Fig. 3(a). The corresponding Laplacian matrix $L$ and $L+E$ are

$$L = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},\ L+E = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 2 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

We have taken $\epsilon$ to be 1. The dominant eigenvalue $r(-(L+E))$ of $-(L+E)$ is $-0.5858$. After removing the edge $(2,1)$, the corresponding Laplacian matrix $\bar{L}$ and $\bar{L}+E$ are

$$\bar{L} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},\ \bar{L}+E = \begin{bmatrix} 2 & 0 & -1 \\ -1 & 2 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

We see that the dominant eigenvalue $r(-(\bar{L}+E))$ of $-(\bar{L}+E)$ is $-0.5344$, which is greater than that of $-(L+E)$. The convergence process is decelerated when we delete one edge between two first followers.

*Remark 4.* A direct implication of Theorem 2 is that if the graph that describes the communication topology between agents is directed and weighted, then stronger connectivity of the graph might actually slow down the convergence. This is in sharp contrast with the case when the graph is undirected and unweighted, for which more edges between vertices or higher edge weights always accelerate the convergence.

Theorem 2 has investigated the variation of the dominant eigenvalue of $-(L+E)$ in the process of removing the edges from the other followers to the first followers in $\mathbb{G}$. For the other eigenvalues, they may not monotonically decrease or increase. Since the trace of $-(L+E)$ increases and the dominant eigenvalue $r(-(L+E))$ decreases in the process of removing these specific edges, the sum of the other eigenvalues of $-(L+E)$ excluding $r(-(L+E))$ will increase. In addition, one can show that the magnitude of the product of the other eigenvalues of $-(L+E)$ excluding $r(-(L+E))$ will decrease. In the next section, we look at how the other eigenvalues change through a simulation example.

## 4. SIMULATION EXAMPLE

In this section, we give a simulation example to validate the theoretical results obtained in the previous section. Consider a leader-follower network given by Fig. 4 and the associated Laplacian matrix of $\mathbb{G}$ is

$$L = L_0 = \begin{bmatrix} 3 & -1 & 0 & -1 & -1 \\ -1 & 3 & -1 & 0 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & -1 & 0 & -1 & 2 \end{bmatrix}.$$

Let $\epsilon = 1$ in Eq. (5) and thus $E = \text{diag}(1,0,0,0,0)$. Denote the Laplacian matrix of the follower network by $L_1$ after removing the edge $(2,1)$, denote the Laplacian matrix by $L_2$ after further removing the edge $(4,1)$ and the Laplacian matrix by $L_3$ if $(5,1)$ is furthermore removed. From Theorem 2, we know that the dominant eigenvalue $r(-(L_i+E))$ decreases in this process, namely,

$$r(-(L_{i+1}+E)) < r(-(L_i+E)),\ i = 0,1,2.$$

The real parts of the eigenvalues $\lambda(-(L_i+E))$ of the matrices $-(L_i+E)$, $i = 0,\ldots,3$, are shown in Fig. 5, where $\text{Re}(\lambda(-(L_i+E)))$ denotes the real part of the eigenvalue $\lambda(-(L_i+E))$. From the figure, we can clearly see that the dominant eigenvalue is decreasing and thus the convergence is accelerated. But for the other eigenvalues, there seems to be no clear rule explaining how they change. Though the sum of the other eigenvalues increases, the real part of each individual eigenvalue does not necessarily increase. For example, the real part of the eigenvalue with the largest modulus does not increase all the time as can be seen from Fig. 5.

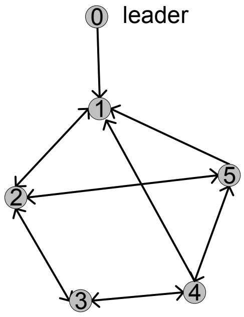


Fig. 4. A leader-follower network with three directed edges pointing to the first-follower 1 from the other followers.

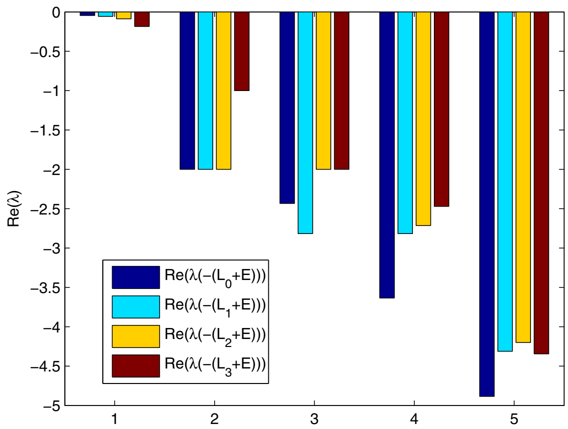


Fig. 5. The changes of the five eigenvalues when edges pointing to the first follower are removed in sequence.

## 5. CONCLUSION

We have proposed a new strategy to accelerate the convergence to consensus in leader-follower networks by making the first followers more focused on their information about the leader. Rigorous proof has been provided to show that the dominant eigenvalue of the system matrix decreases in the process of removing the links pointing to the first followers from the other followers. A simulation example has been used to validate the theoretical results.

We are looking into more complicated agent models to check whether our proposed strategy still works. We are also interested in investigating more comprehensively how the hierarchical ranking of the followers in terms of their topological relationships in the network affects the convergence speed when all the followers' states are converging to that of the leader.

## ACKNOWLEDGEMENTS

The third author would like to thank Fan Zhang for helpful discussions.

## REFERENCES

Blondel, V.D., Hendrickx, J.M., and Tsitsiklis, J.N. (2009). On Krause's multi-agent consensus model with state-dependent connectivity. *IEEE Transactions on Automatic Control*, 54, 2586–2597.

Cao, M., Morse, A.S., and Anderson, B.D.O. (2008a). Reaching a consensus in a dynamically changing environment: A graphical approach. *SIAM Journal on Control and Optimization*, 47, 575–600.

Cao, M., Morse, A.S., and Anderson, B.D.O. (2008b). Reaching a consensus in a dynamically changing environment: Convergence rates, measurement delays and asynchronous events. *SIAM Journal on Control and Optimization*, 47, 601–623.

Cao, M., Spielman, D.A., and Yeh, E.M. (2006). Accelerated gossip algorithms for distributed computation. In *Proc. of the 44th Annual Allerton Conference*, 952–959.

Cohen, J.E. (1981). Convexity of the dominant eigenvalue of an essentially nonnegative matrix. *Proceedings of the American Mathematical Socity*, 81, 657–658.

Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge, U.K.

Jadbabaie, A., Lin, J., and Morse, A.S. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48, 985–1001.

Liu, J., Anderson, B.D.O., Cao, M., and Morse, A.S. (2013). Analysis of accelerated gossip algorithms. *Automatica*, 49, 873–883.

Nedic, A. and Ozdaglar, A. (2010). Convergence rate for consensus with delays. *Journal of Global Optimization*, 47, 437–456.

Ni, W. and Cheng, D. (2010). Leader-following consensus of multi-agent systems under fixed and switching topologies. *Systems and Control Letters*, 59, 209–217.

Olfati-Saber, R., Fax, J.A., and Murray, R.M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95, 215–233.

Olshevsky, A. and Tsitsiklis, J.T. (2009). Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48, 33–55.

Olshevsky, A. and Tsitsiklis, J.T. (2011). A lower bound on distributed averaging on the line graph. *IEEE Transactions on Automatic Control*, 56, 2694–2698.

Oreshkin, B.N., Coates, M.J., and Rabbat, M.G. (2010). Optimization and analysis of distributed averaging with short node memory. *IEEE Transactions on Signal Processing*, 58, 2850–2865.

Ren, W. and Cao, Y. (2011). *Distributed coordination of multi-agent networks: Emergent problems, models, and issues*. Springer-Verlag, London.

Scardovi, L. and Sepulchre, R. (2009). Synchronization in networks of identical linear systems. *Automatica*, 45, 2557–2562.

Xiao, L. and Boyd, S. (2004). Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53, 65–78.

Zhang, H., Chen, M., and Zhou, T. (2009). Improve consensus via decentralized predictive mechanisms. *Europhyscis Letters*, 86, 40011.