Effect of Community Structure of Network on Distributed Estimation

Xiaodan Shao[1], Feng Chen[1, a]

[1]College of Electronics and Information Engineering, Southwest University, Chongqing, 400715, P.R. China

[a]Email address: Fengchen.uestc@gmail.com (F. Chen)

Keywords: Diffusion LMS, community structure, distributed estimation, estimation performance.

Abstract. The diffusion strategies have been widely studied for distributed estimation over adaptive networks. This paper investigates the impacts of community structure of network on the performance of the adaptive-then-combine diffusion LMS. The study covers different local community structures, while the performance is analyzed according to the transient and steady state mean-square errors. Simulations demonstrate that the performance of distributed estimation have regular changes along with the variety of number of community, which indicates that the network community indeed plays an important role in the distributed estimation.

Introduction

Distributed estimations have become popular for parameter estimation in wireless networks and applications. In the Diffusion strategies, the information is processed locally and simultaneously at all nodes while the nodes communicate with all their neighbors to share their data intermediate estimates with them [1, 2, 3]. This paper focus on the adaptive-then-combine (ATC) diffusion LMS algorithm, which has been proved to be superior to the other diffusion LMS algorithms [3].

When information spreads through a network, the local dynamics will be related to the local community structures. The network motif is the basic community structure of complex networks. Motifs are small subgraphs that occur significantly more frequently in real networks than expected by chance alone, and are detected purely by topological analysis [4]. They have been found in a wide variety of networks [5, 6]. Due to the widespread existence of network motifs, it is interesting to know how the distributed estimation performs over varying motifs. Other similar studies can be found in the effect of unbalanced triangles in social networks [7]. This paper analyzes the impacts of network communities on the performance of distributed estimation. Simulation results show that the estimation performance is largely dependent on the number of triangular motifs.

Diffusion Least-Mean Squares Algorithm

Consider a connected network with N nodes. $i \in \{1, 2, \mathbf{K}, N\}$ is the node index and t is the time index. To proceed with the analysis, we assume a liner measurement model as follows:

$$d_i(t) = \boldsymbol{u}_{i,t}{}^T \boldsymbol{w}^0 + v_i(t). \tag{1}$$

Where $\boldsymbol{w}^0$ is a deterministic but unknown $M \times 1$ vector, $d_i(t)$ is a scalar measurement of some random process , $\boldsymbol{u}_{i,t}$ is the $M \times 1$ regression vector at time t with zero mean and covariance, $v_i(t)$ is the random noise signal at time t with zero mean and variance. Taking Adapt-then-Combine Diffusion LMS (ATC) [6] as an example, the parameter updating rule for node i is

$$\boldsymbol{x}_{i,t+1} = \boldsymbol{w}_{i,t} + m_i \boldsymbol{u}_{i,t+1}(d_i(t+1) - \boldsymbol{u}_{i,t+1}{}^T \boldsymbol{w}_{i,t}). \tag{2}$$

$$\boldsymbol{w}_{i,t+1} = \sum_{j \in N_i} a_{i,j} \boldsymbol{x}_{j,t+1}. \tag{3}$$

Where N_i denotes the neighboring nodes set of node i (including node i itself), m_i is the step-

© 2016. The authors - Published by Atlantis Press

size at node i, $a_{i,j}$ are liner weights satisfying the following conditions [2].

$$\sum_{j \in N_i} a_{i,j} = 1, \text{ and } a_{i,j} = 0 \text{ for all } j \notin N_i. \tag{4}$$

The performance of the diffusion LMS algorithm is usually evaluated by the mean-square-deviation (MSD) and the excess-mean-square-error (EMSE) [1,3]. The MSD and EMSE for the whole network are computed by (5) and (6), where $h_i(t)$ and $z_i(t)$ are the MSD and EMSE for each node i at time instant t.

$$h(t) = \frac{1}{N}\sum_{i=1}^{N} h_i(t) = \frac{1}{N}\sum_{i=1}^{N} E\left\|\boldsymbol{w}^0 - \boldsymbol{w}_{i,t}\right\|_2^2. \tag{5}$$

$$z(t) = \frac{1}{N}\sum_{i=1}^{N} z_i(t) = \frac{1}{N}\sum_{i=1}^{N} E\left\|\boldsymbol{u}_{i,t}^{\ T}\left(\boldsymbol{w}^0 - \boldsymbol{w}_{i,t}\right)\right\|_2^2. \tag{6}$$

Network Motifs

Motifs represent the state of the local sub-communities. This paper only considers the common size-3 motifs and size-4 motifs in an undirected graph depicted in Fig. 1a and 1b, and will refer to them using the corresponding codes.

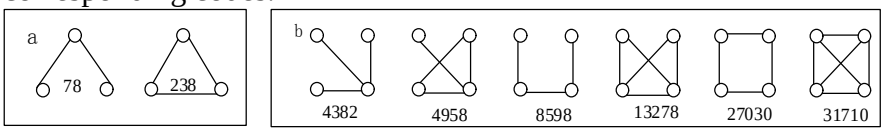


Fig.1. Motifs representation in an undirected graph. a. The two types of motifs of size 3. b. All existing motifs of size 4.

For any network, F is one of its motifs, we define the number of motif F including a node i is the motif degree of the node i expressed as k_i^F. We can further obtain the average motif degree of all nodes in the network as (7). Intuitively, the greater the average motif degree is, the more motifs the network will include.

$$k = \frac{1}{N}\sum_{i=1}^{N} k_i^F \tag{7}$$

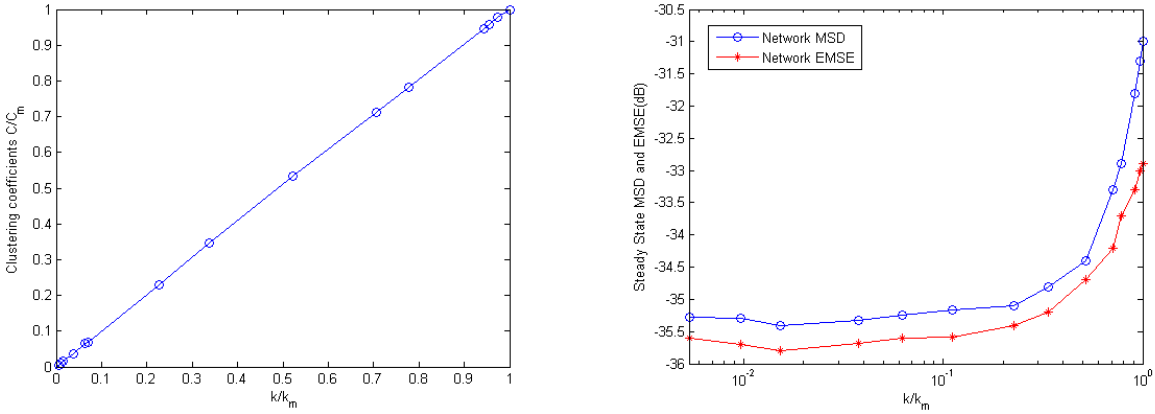
Simulation Results

To systematically investigate the effect of motifs on distributed estimation, we have carried out a number of experiments.

Effect of Size-3 Motifs. We randomly generate a network topology. The step-size of the LMS is set to $m_i = 0.03$. For a stringent comparison, our simulations are performed by starting with networks that have the same nodes $N = 1000$ and edges $L = 3000$ .To begin with, the triangular motif 238 is considered. Fig.2 (a) plots the normalized clustering coefficient C/C_m of network vs. normalized average motif degree k/k_m, where C_m and k_m denote those of the fixed scale network including the most triangular motifs. The figure shows that the network with the bigger normalized clustering coefficient has the higher normalized average motif degree about triangular motifs.

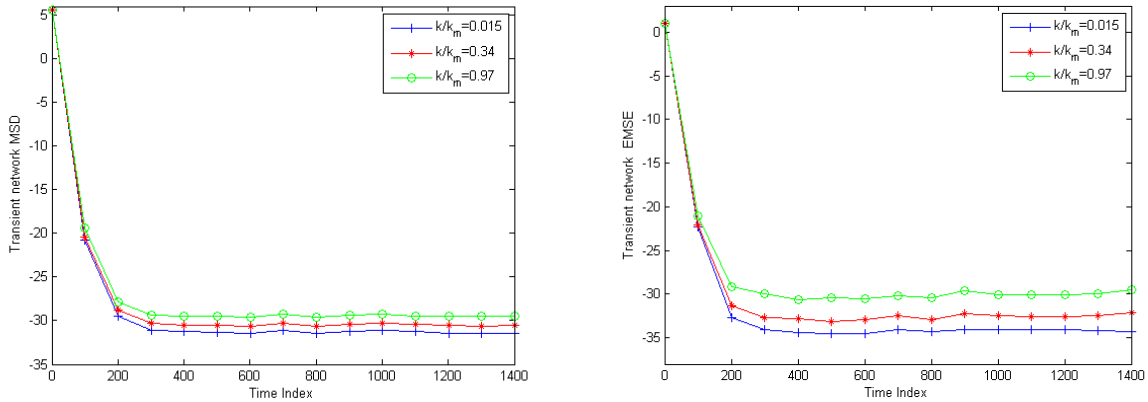
The ATC diffusion LMS is applied to estimate the vector of interest $\boldsymbol{w}^0$ over different networks with varying average motif degrees of triangular motifs. Their corresponding steady-state MSD and EMSE are computed by averaging the last 300 samples after 1400 iterations over 100 experiments. The results are depicted in Fig.2 (b). It is interesting to notice that a similar changing trend can be

observed as compared with the normalized clustering coefficient given in Fig.2 (a). Steady state MSD and EMSE increase with the increasing of normalized average motif degree. Initially, increase the value of k/k_m does not affect the estimation performance significantly (see $k/k_m < 0.015$); when k/k_m increases from 0.015 to 0.23, the MSD and EMSE increase slightly; afterwards, with the increase of the normalized average motif degree, the MSD and EMSE increase sharply. Fig.2 (b) shows that the MSD and EMSE reach their optima around $k/k_m = 0.015$.



(a) (b)

Fig.2. (a) Normalized clustering coefficient (C/C_m) for a range of normalized average motif degree of triangular motifs (k/k_m). (b) The steady state MSD and EMSE against k/k_m.



(a) Network MSD (b) Network EMSE

Fig.3. Transient performances comparison with different normalized average motif degree of triangular motifs (k/k_m).

Fig.3 (a) and (b) plot the transient MSD and EMSE of the entire network in dB for different normalized average motif degree of triangular motifs. From Fig.3, we notice that the network with the highest k/k_m has the worst MSD and EMSE performance. However, by reducing the number of triangular motifs (decreasing the value of k/k_m), the accuracy of the estimation is significantly improved. Then, we do the same experiments for the chain motif (78), but do not find regular changes.

Effect of Size-4 Motifs. By applying the above same analytical method on normalized average motif degree of size-4 motifs, we obtain the respective steady-state MSD of the diffusion LMS in Fig.4. Looking at Fig.1, the six motifs can be divided in two categories: motifs with triads (4958, 13278, 31710) and with no triads (4382, 8598, 27030) in their structure. From Fig.4, we note that the steady state MSD increases with the increasing of normalized average motif degree k/k_m over motifs of size-4 that have triadic closures, and furthermore the motif with more triadic closures has

stronger impact on the mean-square errors. However, increase the value of k/k_m about motifs with no closed triangles does not affect the estimation performance meaningfully.

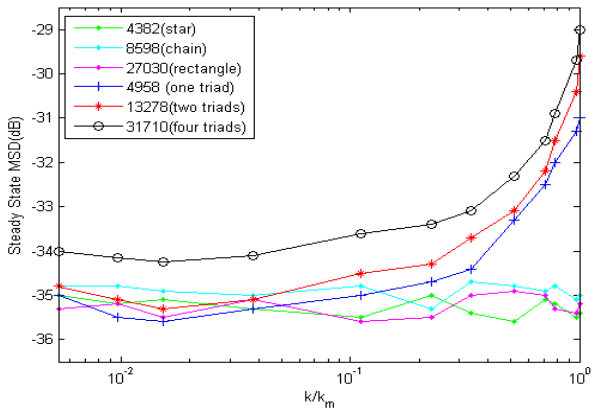


Fig.4. The comparison of steady-state MSD between motifs of size-4 that have triadic closures and motifs that do not have any closed triangles in their structure.

It is found that motifs with triadic closures does contribute to reducing the performance of distributed estimation. Fig.2 (a) demonstrates that the prevalent occurrence of triads also represents the high clustering coefficient, while big clustering coefficient results in excessive reuse data in the exchange of intermediate estimates. These results are consistent with the distributed estimation over complex network [8]. The results for the same properties can be observed in other size motifs.

Conclusions

This paper have investigated the performance of the ATC diffusion LMS algorithm over varying communities from the viewpoints of transient and steady state mean-square errors. The results have shown that the motifs with triadic closures does lead to reducing the performance of distributed estimation. In other words, the fewer motifs with triadic closures in the same scale networks, the better the performance of the distributed estimation. This study also provides some guidelines on how to select links in the traffic restricted network such that the qualities of estimates are optimized.

Acknowledgements

This work was supported in part by the Fundamental Research Funds for the Central Universities un der grant number XDJK2014C018 and Doctoral Fund of Southwest University (No.SWU113067).

References

[1] Cattivelli, C.G. Lopes, A.H. Sayed: IEEE Trans. Signal Process. 2008, 56(5):1865–1877.

[2] C.G. Lopes, A.H. Sayed: IEEE Trans. Signal Process. 2008, 56 (7):3122–3136.

[3] Cattivelli, A.H. Sayed: IEEE Transactions on Signal Processing, 2010, 58(3):1035-1048.

[4] R. Milo, S. Shen-Orr, S. Itzkovitz, et al. Science, 2002, 298(5594):824–827.

[5] A.M. Krieg,,et al. Nature International Weekly Journal of Science,1995, 374(6522):546-549.

[6] S.S. Shen-Orr, R. Milo, S. Mangan: Nature Genetics, 2002, 31(1):64-68.

[7] T.Antal. Physical Review E Statistical Nonlinear&Soft Matter Physics, 2005, 72(3):1191-1201.

[8] Ying. Liu, Chunguang Li, W.K.S Tang: Information Sciences, 2012, 197:91–104.