

Analysis and Application of the Relationship between Information Output and Network Topology

Liu Bing⁺

Beijing Electronic Science and Technology Institute, Beijing, China

Abstract. The information output in networks with high node density is influenced by the correlation among the data from neighboring nodes. Such networks include Internet of Things, Wireless Sensor Networks and so on. To improve the efficiency of the information collection, the relationship between information output and network topology is analyzed in order to give the disciplines in network deployment. The proofs and numerical computations verify the conclusions. They can also be used in the protocol design such as Media Access Control.

Keywords: information output; network topology; network protocol

1. Introduction

In the application of some networks, such as Internet of Things (IoT), Wireless Sensor Networks (WSNs) and Ad Hoc Networks, the terminals are always collaborated to accomplish a specific task [1]. In order to guarantee the quality of application, the nodes are usually distributed in a high density way. This makes the data acquired from adjacent nodes have certain correlation [2], which can be used to eliminate the redundant data, thus reducing the transmission and improving the efficiency of the networks. Such technologies include Data Aggregation [3], Distributed Source Coding (DSC) [4], etc.

On the other hand, network layout, including the node position and the network organizational structure, has important influence on the performance and efficiency of the networks. In such networks, data transmissions are often in the form of multiple hops. There exists the cluster, tree or other forms of data flows [5]. In some scenarios the node location and network structure can be accurately designed. And in other scenarios, such as aircraft spread and animal carrying, it also should be considered according to certain principles. So the following issues are often considered: In order to collect enough information, how many nodes should be deployed in the area? And how can it be done with better efficiency? In the cluster and tree structures, the upper nodes (cluster heads or superior nodes) should reserve the resources (mostly the time slots) for the lower nodes. How should the resources be distributed for each node (the simplest average assignment is not the most efficient way obviously)?

The above issues are about the relationship between the information output and the network topology. And it is discussed in this paper in order to improve the effectiveness of information collection. The certification and verification of some basic conclusions are presented by the mathematical proofs and numerical calculations. These basic principles can be also used in the design of Media Access Control (MAC) and other network protocols.

2. Problem Analysis

⁺ Corresponding author.
E-mail address: leoliu76@outlook.com.

The problems will be discussed under certain correlation model. For clarity, it will begin with the topology in one dimension, and the mathematical proofs are given. In the cases of two-dimensional plane and three-dimensional space, they are discussed by numerical calculations.

2.1. Network model

Discuss the continuous case and assume that nodes collect data of the Gaussian variables [6]. The data of nodes can be seen as a random vector $X^{(n)} = (X_1, X_2, \dots, X_n)$, of which the probability density function is

$$p(x^{(n)}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x^{(n)} - \mu^{(n)})\Sigma^{-1}(x^{(n)} - \mu^{(n)})'\right) \quad (1)$$

Σ is the covariance matrix (σ_{ij}),

$$\sigma_{ij} = \int_{R^n} (x_i - \mu_i)(x_j - \mu_j)p(x^{(n)})dx^{(n)} \quad (2)$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{21} & \sigma_2^2 & \cdots & \sigma_2\sigma_n\rho_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_n\sigma_1\rho_{n1} & \sigma_n\sigma_2\rho_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \quad (3)$$

The differential entropy of the entire network is

$$H(X^{(n)}) = \frac{n}{2} \log(2\pi e |\Sigma|^{1/n}) \quad (4)$$

To model the correlation which decreases with the increasing distance, assumption is giving as below:

$$\rho_{ij} = \rho_{ji} = e^{-\theta d_{ij}} \quad (5)$$

θ is the impact factor, which represents the influence degree of the distance to the correlation. The greater θ is, the more intensely correlation changes with distance [7].

2.2. Deployment principles

The following will answer to a few basic questions and give proofs to some conclusions. They can be referred in the deployment of networks.

Conclusion 1: in a given area, more nodes give greater amount of useful information.

It is correct obviously. If a node is joined in the network, the incremental information is greater than zero since the data is not determinate. More nodes mean greater information collected. Of course, the higher node density means the greater data correlation and the information increment will be smaller.

Conclusion 2: When the node number is fixed, the maximum amount of information in the network is achieved when nodes are uniformly distributed.

Proof: It is proved in one dimensional case. Assume that n nodes X_1, X_2, \dots, X_n are distributed in a line with the length L , in the order from left to right as shown in Fig.1.

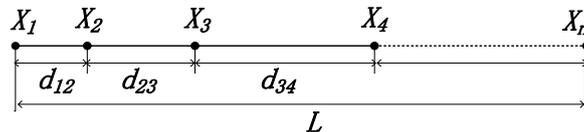


Fig. 1: Node distribution in one dimension

$d_{i,j}$ presents the distance between X_i and X_j .

$$\sum_{k=i}^{j-1} d_{k,k+1} = d_{i,j} \quad (6)$$

So $d_{1,n} = L$.

For the convenience of discussion, assume that the data variances of all nodes are equal, then

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2 e^{-\theta d_{12}} & \dots & \sigma^2 e^{-\theta d_{1n}} \\ \sigma^2 e^{-\theta d_{12}} & \sigma^2 & \dots & \sigma^2 e^{-\theta d_{2n}} \\ \dots & \dots & \dots & \dots \\ \sigma^2 e^{-\theta d_{1n}} & \sigma^2 e^{-\theta d_{2n}} & \dots & \sigma^2 \end{bmatrix} \quad (7)$$

This is a symmetric matrix.

Under the condition of $\sum_{i=1}^{n-1} d_{i,i+1} = L$, calculate the extreme value of $|\Sigma|$.

$$|\Sigma| = \begin{vmatrix} \sigma^2 & \sigma^2 e^{-\theta d_{12}} & L & \sigma^2 e^{-\theta d_{1n}} \\ \sigma^2 e^{-\theta d_{12}} & \sigma^2 & L & \sigma^2 e^{-\theta d_{2n}} \\ L & L & L & L \\ \sigma^2 e^{-\theta d_{1n}} & \sigma^2 e^{-\theta d_{2n}} & L & \sigma^2 \end{vmatrix} = \sigma^{2n} \begin{vmatrix} 1 & e^{-\theta d_{12}} & L & e^{-\theta d_{1n}} \\ e^{-\theta d_{12}} & 1 & L & e^{-\theta d_{2n}} \\ L & L & L & L \\ e^{-\theta d_{1n}} & e^{-\theta d_{2n}} & L & 1 \end{vmatrix} \quad (8)$$

Assume $a_1 = e^{-\theta d_{12}}$, $a_2 = e^{-\theta d_{23}}$, ..., $a_{n-1} = e^{-\theta d_{n,n-1}}$, Then

$$\begin{vmatrix} 1 & e^{-\theta d_{12}} & L & e^{-\theta d_{1n}} \\ e^{-\theta d_{12}} & 1 & L & e^{-\theta d_{2n}} \\ L & L & L & L \\ e^{-\theta d_{1n}} & e^{-\theta d_{2n}} & L & 1 \end{vmatrix} = \begin{vmatrix} 1 & a_1 & L & a_1 a_{n-1} \\ a_1 & 1 & L & a_2 L a_{n-1} \\ L & L & L & L \\ a_1 a_{n-1} & a_2 L a_{n-1} & L & 1 \end{vmatrix} \quad (9)$$

Do the elementary row transformation. Except the first line, line i is added by the product of line 1 and $-a_i$. Then the matrix becomes into a triangular matrix

$$\begin{vmatrix} 1 & a_1 & L & a_1 a_{n-1} \\ 0 & 1-a_1^2 & L & L \\ L & L & L & L \\ 0 & 0 & L & 1-a_{n-1}^2 \end{vmatrix} = \sum_{i=1}^{n-1} (1-a_i^2) \quad (10)$$

and

$$\prod_{i=1}^{n-1} a_i = e^{-\theta L} \quad (11)$$

It is a constant.

$|\Sigma|$ achieves the maximum when $a_1 = a_2 = \dots = a_{n-1}$ when $d_{12} = d_{23} = \dots = d_{n,n-1}$, and the nodes are uniformly distributed.

Conclusion 3: The output of network information is not only related to the number of nodes, but also the size of network.

The case of one dimension is also calculated:

$$|\Sigma|_{\max} = \sigma^{2n} (1 - e^{-\frac{2\theta L}{n-1}})^{n-1} \quad (12)$$

By the formula (4)

$$H_{\max}(X^{(n)}) = \frac{n}{2} \log(2\pi e |\Sigma|_{\max}^{1/n}) = \frac{n}{2} \log\left(2\pi e \sigma^2 (1 - e^{-\frac{2\theta L}{n-1}})^{\frac{n-1}{n}}\right) = \frac{n}{2} \log 2\pi e \sigma^2 + \frac{n-1}{2} \log(1 - e^{-\frac{2\theta L}{n-1}}) \quad (13)$$

Obviously, when the other parameters are invariant, the differential entropy increases with the increase of L .

2.3. Numerical calculation

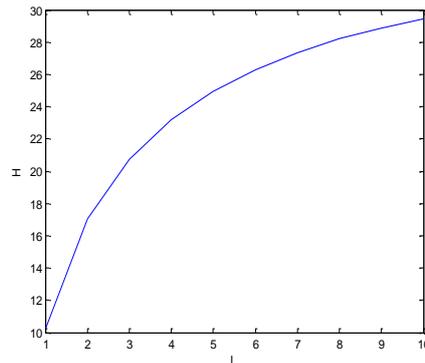


Fig. 2: Relation of information output and the network size in one dimension

The following are the numerical calculation using MATLAB about the conclusions. It makes the conclusions more clearly, and more complex situations are studied.

In Fig. 2 is the one dimensional case. The number of nodes is 16 and the data variance of each node is 1. With the increase of L , the node deployment is sparser, and the total amount of information increases. The reason is that the correlation between nodes decreases with larger distance.

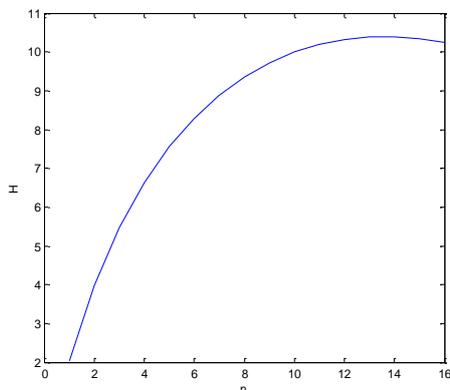


Fig. 3: Relation of information output and the number of nodes in one dimension

Fig. 3 shows the variation of information output with the number of nodes, in the case that the network size is constant (the value of L is 1) and the data variance of each node is 1. As expected, the amount of information should be monotone increasing. The turning point in the figure is not the fault of the conclusions. The reason is that the calculated item here is the differential entropy and it is not the absolute value of information. The differential entropy is the difference between the absolute entropy with the ‘infinity’. For continuous data, each additional sampling brings out infinite amount of information. But for the differential entropy, the simple summation corresponds to the difference with the bigger ‘infinity’. So there may be the case that the differential entropy decreases with the number of nodes.

Though the relation cannot be directly seen from the figure, some meaningful information can be found. In the initial stage, when the nodes are sparse, adding nodes can give obvious increase to the amount of information. In such case the differential entropy is greater, even compared with ‘bigger’ infinity. Only when the node density becomes higher, and the differential entropy turns negative (whose absolute value is still the infinity), differential entropy with the number of nodes appears to decline.

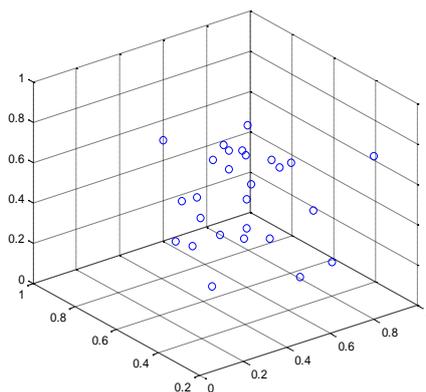


Fig. 4: Normal distribution of nodes in three dimensions

The relationship between the information output and the nodes distribution in three dimension case is discussed in the following. The nodes are deployed in space with the normal distribution. That is, the distances from the nodes to the centre of the space are normally distributed with the variance of σ_s . The greater variance means the distribution is more uniform. Fig. 4 is an example of such distribution.

The following is about the relation of information output with the impact factor θ or distribution variance σ_s in three dimension case. Except the two parameters talked respectively, others are listed in Table I.

Table 1: Parameters in Simulation

| Parameters | Value |
|---|-----------------------|
| Network Area | $1 \times 1 \times 1$ |
| Number of Nodes (N) | 27 |
| Variance of distribution (σ_s) | 0.1 |
| Variance of data (σ) | 1 |
| Impact factor (θ) | 1 |

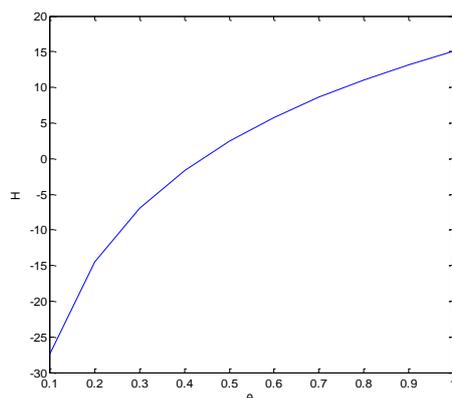


Fig. 5: Relation of information output and impact factor in three dimensions

In Fig. 5, with the same distribution, the greater the impact factor θ is, the more intensely data correlation among the nodes changes with distance. So the data correlation under the same distance is smaller, and the information throughput shows a trend of increase.

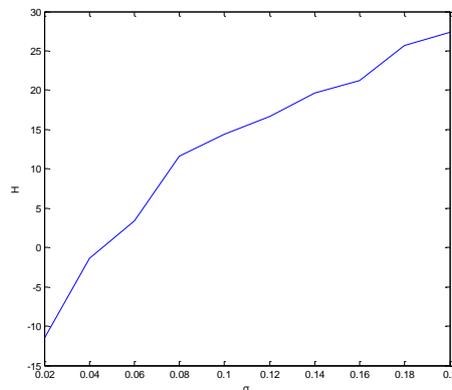


Fig. 6: Relation of information output and distribution uniformity in three dimensions

In Fig. 6, with other parameters constant, the larger σ_s is, the more uniformly nodes distributed, and the greater the information output is, which is the same as our predictions.

3. Application Examples

There are a lot of practical applications of the basic principles in network layout.

If the number of nodes is pre-decided, in order to collect more information, nodes should be uniformly deployed in the area. If the needed information output is fixed, the number of nodes deployed should be calculated firstly by a reverse method.

In the networks, senior nodes (cluster heads or upper nodes in the tree) should reserve resources (mostly the slots) to the lower nodes, basing on their information output. And the output can be estimated by the topological space and the number of nodes. So the former conclusions can be used to design protocols such

as dynamic MAC for networks with tree structures, in which the slots reserved by the parent nodes to the child nodes depend on the topological space and the node number of lower layers [8].

4. Conclusion

With the application of Internet of Things, wireless sensor network and so on, more and more networks that have high node density emerge, facing at more complex structure and protocol design problems. In order to improve the effectiveness of information collection, this paper discusses the relationship of information output and topology, proves and validates some basic principles, and gives the basis for some practical applications.

5. References

- [1] L. Mainetti, L. Patrono, and A. Vilei. Evolution of wireless sensor networks towards the internet of things: a survey. *Proceedings of the 19th International Conference on Software Telecommunications and Computer Networks*, Split, Croatia, 2011, pp. 1-6.
- [2] S. Meguerdichian, F. Koushanfar, M. Potkonjak, et al. Coverage problems in wireless ad hoc sensor networks. *Proc. IEEE INFOCOM*, 2001, pp. 1380-1387.
- [3] T. He, B. M. Blum, J. A. Stankovic, et al. AIDA: Adaptive application independent data aggregation in wireless sensor networks. *ACM Transactions on Embedded Computing system (Special Issue on Dynamically Adaptable embedded Systems)*, 2004, 3(3), pp. 426-457.
- [4] D. Slepian, J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 1973, IT-19(4), pp. 471-480.
- [5] M. Alowish, Y. Takano, Y. Shiraishi, et al. Performance Evaluation of a Cluster Based Routing Protocol for VANETs. *Journal of Communications*. Vol. 12, No. 2, 2017, pp. 137-144.
- [6] T. M. Cover, and J. A. Thomas. *Elements of Information Theory*. 2nd Edition. John Wiley & Sons, Inc, 2006.
- [7] M. C. Vuran, and I. F. Akyildiz. Spatial Correlation-Based Collaborative Medium Access Control in Wireless Sensor Networks. *IEEE/ACM Transactions on Networking*. Vol. 14, No. 2. April, 2006, pp. 316-329.
- [8] H. Pei, X. Li, W. M. Matt, X. Ning. The Evolution of MAC Protocols in Wireless Sensor Networks: A Survey. *IEEE Communicatians surveys & Tutorials*, vol.15, No.1, first quarter. 2013, pp. 101-120.