Fast Topology Inference of Wireless Networks Based on Hawkes Process

Yehui Song[1], Jiachen Sun[2], Guoru Ding[3]
{Songyh564@sina.com[1], sun_jiachen@outlook.com[2], dr.guoru.ding@ieee.org[3]}

College of Communications Engineering, Army Engineering University,
Nanjing 210007, China[1,2,3],

Abstract. Based on the reasonable equivalent assumption of communication events, using the Hawkes process to model the information interaction in wireless communication networks is an emerging direction in the field of non-cooperative topology inference. At present, topology inference algorithms based on the Hawkes process mainly use a fixed sample size for inference, considering only its reliability, but not regarding its effectiveness. In this paper, we consider introducing a sample size as a new performance indicator. For small sample size scenarios in wireless networks, a kind of fast topology inference algorithm is proposed, which uniformly represents parameters belonging to different dimensions, and thoroughly mines topological information from different batches to increase the speed and effectiveness of inference. Experimental simulations show that compared with the existing algorithm, our algorithm has better performance in small sample size scenarios.

Keywords: Fast topology inference, Wireless networks, Hawkes process, Effectiveness

1 Introduction

With the widespread application of wireless networks, the importance of intelligent analysis of network behaviors is becoming increasingly prominent. In the analysis of networks behaviors, learning and reasoning about the connectivity of unknown networks is a fundamental problem [1]. Through topology inference, we can mine the connection relationships between nodes to achieve the visual network's topology, and further improve our knowledge of the observed networks. For wireless networks management, topology inference can provide information support for managers to troubleshoot and enhance network security and robustness. For monitoring uncooperative wireless networks, especially in the field of electronic warfare, topology inference can help us identify and combat key nodes and critical links in enemy networks. Further, we can mine intelligence from topological information, infer the enemy's combat intentions, and realize the information superiority on the battlefield. But whether it is the topology inference for our network or the enemy networks, this requires timeliness, and we need to get the topology of the observed networks as quickly as possible. These observations motivate us to research fast topology inference in the framework of wireless networks.

The existing research on topology inference can be divided into two categories: cooperative topology inference and non-cooperative topology inference. Cooperative topology inference refers to the fact that sensor nodes are part of the perceived network and can reason out the topology of the network through message exchange within the network. Non-cooperative

topology inference refers to the fact that sensor nodes do not belong to the perceived network, and can reason out the topology of the network just only through passive detection of the network.

With the development of wireless networks, in recent years, researches on non-cooperative topology inference are receiving more and more attention [2]. Based on a reasonable equivalent assumption for communication events: the receiving nodes will act accordingly when they receive the signal from the sending nodes, using the Hawkes process to model the information interaction in wireless communication networks is an emerging direction in the field of non-cooperative topology inference. The Hawkes process is widely used in financial transactions, social network analysis, bioinformatics and many other fields [3]. In the field of communication, [4] models the information transmission process as the Hawkes process based on reasonable assumptions of communication equivalence. Based on [4], [5] further proposes a method named Low Cost Paths for Acyclic Graphs (LCPAG) to discover event chains.Then, [6] considers the wireless channel, develops a physical model for external topology awareness, and proposes a wireless channel-oriented topology sensing method based on the Hawkes process. However, the above works still have limitations: firstly, the fixed sample size is used for inference, only considering the reliability of the algorithm, without its effectiveness; secondly, there is a lack of research on small sample size scenarios in wireless networks, so it is necessary to improve the use efficiency of sample size. Thirdly, the choice of the global threshold will greatly affect the reasoning effect, and there are problems in selection in practical application.

Therefore, in view of the above problems, this paper reconsiders topology inference algorithm based on Hawkes process by considering both reliability and effectiveness. The main contributions of this paper are summarized as follows:

-   Considering the effectiveness of the topological inference algorithm, the sample size is introduced, and a new performance index for the topological inference algorithm is proposed.

-   Based on the idea of sequential detection, we propose a fast topology inference algorithm that uniformly represents parameters belonging to different dimensions and fully mines the topological information obtained from inference between different batches to improve the speed and effectiveness of inference.

-   Through experimental simulations, we prove that the proposed algorithm has better performance in small sample size scenarios compared with existing algorithms.

The rest of this paper is organized as follows. Section 2 introduces the system model. In Section 3, the fast topology inference algorithm based on Hawkes process for wireless channels is proposed. Then, in Section 4, we present simulation results that demonstrate the advantages of the algorithm in small sample size scenarios. In addition, Section 5 presents conclusions and future prospects.

2 System Model

2.1 Single-node Sensing Model

As shown in **Figure 1**, this paper considers a single-node sensing model. The target wireless network V is a wireless communication network of Unmanned Aerial Vehicles (UAVs), where the sensor S has the ability of signal sensing and topology inference.
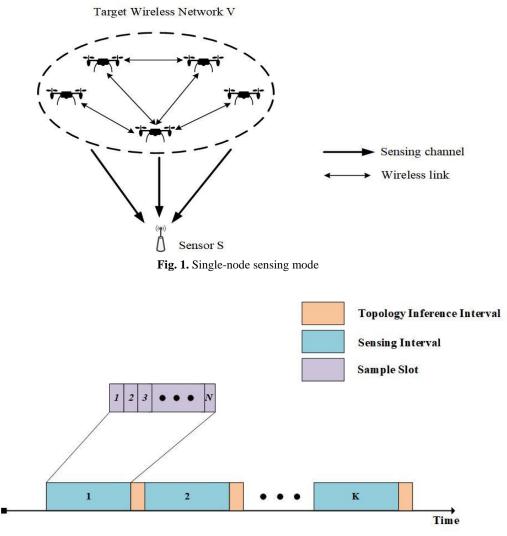
Fig. 1. Single-node sensing mode

Fig. 2. Working frame of sensor S

The working frame of the sensor S is shown in **Figure 2**. It means that there are N sample slots in a sensing interval and the data in a sensing interval is used for once topology inference. Once a sensing is over, the sensor will make inferences based on all existing data. If it meets the requirements, stop perception, otherwise continue until the inference effect reaches the requirements or the maximum number of perceptions is reached.

To simplify the problem, we make assumptions [6]: the transmission channel in the target wireless network V is the ideal channel, and the sensing channel outside the network is the AWGN channel, and the sensor S as an external observer has only two kinds of information available: the node transmitting signal (called the event sequence) and the time at which the signal was transmitted (called the time sequence).

For this model, our goal is to use the sensed event sequence and time sequence to reason about the topology of the target wireless network.

2.2 Mathematical Model of Information Interaction

The Hawkes process is a point process that has autoregressive dependence on past events, where the intensity of an event at any time is a function of the most recent event in the process. From this, we can discover the causality or correlation between events through the Hawkes process. In the field of wireless communications, it is natural that information exchange is performed by each node by transmitting signals. Now we will emit the signal as an event. Then, we can assume that when there is information interaction between two nodes, events on one node are likely to cause a response on the other node. Through such equivalent assumptions, we can model communication events between nodes as Hawkes processes, and then infer the topology of the communication network by solving key parameters.

For an event, given the time when the event occurred in the past t_k, then the intensity of the occurrence of the event at the moment t is

$$\lambda(t) = \mu(t) + A\sum_{k=1}^{K} \gamma(t - t_k), \tag{1}$$

where the parameter $\mu(t) \geq 0$ represents the basic intensity of the occurrence of the event; the parameter $A \geq 0$ represents the degree of self-motivation of the event, that is, how much the event occurs at the moment t_k has an impact on the event occurs at the moment t; the kernel function $\gamma(t)$ represents a time relationship between the corresponding events, which is a known, causal, non-negative, and integrable function. Under normal circumstances, we believe that the basic basis of an event does not change with time, that is, the probability of a new event occurring at any time is certain, so the formula (1) can become

$$\lambda(t) = \mu + A\sum_{k=1}^{K} \gamma(t - t_k). \tag{2}$$

In a wireless communication network, for a node, the event of transmitting a signal is not only related to its own behavior, but also related to the behavior of other corresponding nodes. The one-dimensional Hawkes model can be easily extended to a process that contains multiple subprocesses. At this time, the intensity of the occurrence of a subprocess is not only affected by its own behavior, but also by the behavior of other subprocesses. Then, we can use the multidimensional Hawkes process to model the communication process in a wireless network, that is, a node transmitting a signal in the network can be see as a subprocess. Further, for a process with N subprocesses, according to the formula (2), the occurrence intensity of the ith subprocess is

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{N} A_{ij} \sum_{k \in K_j} \gamma(t - t_k), \tag{3}$$

where K_j represents the set of events in the jth subprocess; μ_i represents the basic intensity of the ith subprocess, we also think that it does not change with time; A_{ij} represents the degree of response of the ith subprocess to the jth subprocess, and $A_{ij}=0$ represents the occurrence of the jth

subprocess has no effect on the ith subprocess; and the larger $A_{ij}>0$ becomes, the more likely the ith subprocess would occur because of the jth subprocess. But it should be noted that A_{ij} is not a probability value. In addition, we can call A adjacency matrix.

3 Fast topology inference in wireless networks

3.1 Determination of Hawkes Process Parameters

For the topology inference,the most critical thing is to get the parameters A_{ij} and μ_i, which we can use to infer whether there is a communication relationship between the two nodes. In many cases, the parameters are unknown, so we need to reason about the parameters based on the observations within a certain period, that is, time sequences and event sequences, to estimate the parameters. Naturally, we choose the maximum likelihood estimation method to determine the parameters. Then, in $t \in [0,T]$, the negative likelihood function of the ith sub-process is

$$L_i\left(\mu_i, A_{ij}\right) = \int_0^T \lambda_i\left(t\right) - \sum_{k \in K_i} \log \lambda_i\left(t_k\right) \qquad j = 1, 2, ... N .\tag{4}$$

The maximum likelihood estimation method requires us to minimize this convex function to estimate the parameters A_{ij} and μ_i.From [5], we choose to use quasi-Newton method for iterative solution. The data required for the solution of $L_i\left(\mu_i, A_{ij}\right)$ is only derived from the independent data set related to the ith subprocess, so the parameter estimate of each subprocess is independent. Therefore, in $t \in [0,T]$, the negative likelihood function of a process containing N subprocesses is

$$L(\mu, A) = \sum_{i \in N} L_i\left(\mu_i, A_{ij}\right) \qquad j = 1, 2, ... N .\tag{5}$$

3.2 Fast topology inference

The solution of the maximum likelihood estimate for each node depends only on an independent subset of the data associated with it. That is, the data set used for the solution of $A_{ij}(j = 1, 2, ..., N)$ is the same and is independent of the data set used for the solution of $A_{kj}(k \neq i, j = 1, 2, ..., N)$. Therefore, $A_{ij}(j = 1, 2, ..., N)$ of different nodes are not numerically comparable, so it is more reasonable to perform optimization individually rather than global optimization. For a node, the closer A_{ij} is to 0, the weaker the connection with another node is, and vice versa. Based on this premise, the traditional binary hypothesis is:

$$\begin{cases} H_0 : A_{ij} \leq \alpha & \text{(j=1,.....,N)} \\ H_1 : A_{ij} > \alpha & \text{(j=1,.....,N)} \end{cases},\tag{6}$$

where H_0 represents there is no connection between the node i and the node j, and H_1 represents connection exists between the two nodes. In this case, it is more challenging to select the global threshold. This paper is based on the idea of expanding the tendency by accumulation, for a fixed node, and put forward a unified statistic for the same batch:

$$Z_{ij}^k = \frac{A_{ij}^k - \pi_i^k}{\pi_i^k} , \tag{7}$$

where A_{ij}^k indicates the communication intensity value obtained from the sample size of the node i in the kth batch, and π_i^k indicates the average communication intensity value between the node i and other nodes which may have the link with the node i in the kth batch. When threshold of the binary hypothesis is set to 0, Z_{ij}^k indicates the tendency of that A_{ij}^k obey H_0 or H_1. Then, by accumulation and calculation:

$$Z_{ij}^k = \sum_{m=1}^{k} Z_{ij}^m , \tag{8}$$

this tendency will continue to become apparent while avoiding the limitations of setting global thresholds by artificial experience and reducing the need for sample size.

Since it is impossible to know the actual communication relationship inside the wireless network in advance, it isn't very elementary to precisely quantify the performance of the inferred topology. For this problem, in the actual operation, we propose to analyze the adjacency matrix A^{k-1} obtained from the previous batch as the actual adjacency matrix and then compare the adjacency matrix obtained from the current batch A^k with A^{k-1} to calculate the false alarm and missed alarm of the current batch.When the false alarm and missed alarm have little change compared with the previous batch, the sampling can be stopped, and the topology obtained by reasoning can be given.

The fast topology inference algorithm is summarized as follows: **Algorithm 1**. The main components of the algorithm are: Hawkes process parameter determination (lines 4-6), the unification of parameter values and the accumulation of the expansion tendency (lines 7-8), the decision of adjacency matrix (lines 10-16)and determining whether to stop sampling (line 3). As for the traditional topology inference algorithm, please refer to the paper [5].

Specifically introduce the key functions in **Algorithm 1**. The objective function in the line 5-6 is to use the pre-processed data set to estimate the required parameters by the maximum likelihood estimation method. It should be emphasized again that the parameter set of each node is estimated separately. π_i^k in the line 7 refers to the mean of all non-zero values in the parameter set related to node i in the kth batch. Due to the unified and cumulative processing of the parameters, we can fixedly choose 0 as the decision threshold, as shown in lines 10-16.

Algorithm 1 Fast topology inference algorithm based on Hawkes process

---

Input: The number of nodes in the network, N_{node} ; The set of events (the source node ID), E_{node}; The set of slots in which there is a signal sent, E_{slot}; Number of samples, N_{sample}; Threshold of false alarm probability ratio, α ; Threshold of miss alarm probability, ratio, β ;

Output: Recovered influence matrix, A;

1: Initialize μ ,A

2: Sort E_{node} and E_{slot} ;

3: for each $k \in [1, N_{sample}]$ **and** $\dfrac{P_f^k}{P_f^{k-1}} < \alpha$ **and** $\dfrac{P_m^k}{P_m^{k-1}} < \beta$

4:   **for** each $i \in [1, N_{node}]$ **do**

5: $(\mu_i^k, A_{ij}^k) = \underset{\mu,A}{\arg\min}\, L_i(\mu, A)$

6:   where $L_i(\mu, A) = \displaystyle\int_0^T \lambda_i(t)dt - \sum_{k \in K_i} \log \lambda_i(n_k)$

7:   **calculate** π_i^k , $Z_{ij}^k = Z_{ij}^{k-1} + \dfrac{A_{ij}^k - \pi_i^k}{\pi_i^k}$

8: $A_{ij}^k = Z_{ij}^k$

9:   **end for**

10:   **for** each $i, j \in [1, N_{node}]$ **do**

11:   **if** $A_{ij}^k > 0$ **then**

12: $A_{ij}^k = 1$

13:   **Else if** $A_{ij}^k \leq 0$ **then**

14: $A_{ij}^k = 0$

15:   **end if**

16:   **end for**

17:   **compare** A^k and A^{k-1} **then calculate** P_f^k and P_m^k

18: end for

---

4 Performance Evaluation

4.1 Simulation Preparation

This parer considers that the target wireless network has a complex distributed topology, as shown in **Figure 3**. The network has 25 nodes. A fact can be concluded from the topology

diagram that there are only 8% node pairs in communication relationship. Because there is no actual data, we chose to use MATLAB to simulate the process of signal transmission between nodes and sensing of sensor to generate the event sequence and time sequence required for inference. It should be noted that, because AWGN is considered in sensor sensing channel, the sample data obtained has noise samples, that is, there are random event values and time values. This paper mainly considers the small sample size scenario. Set the number of times that 24 pairs of nodes communicate with each other in 96 in 1 batch, and there are 10 batches in all. Other specific simulation parameters are shown in **Table 1**.

Table 1. Simulation Parameters

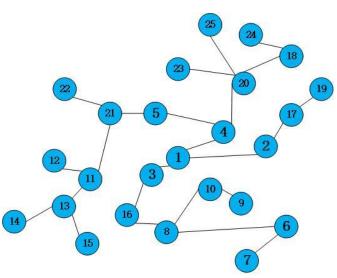| Parameters	Value
Distribution range of the network	1km*1km
Noise spectral density	-174dBm/Hz
Noise bandwidth	10M
Signal power	-5 dBm
Signal decision threshold	$4*10^{-14}$W

Fig. 3. Target wireless network topology in simulation

4.2 Simulation Results

In order to facilitate observation and analysis, a total of five batches of samples were taken for reasoning. The traditional topology inference algorithm and fast topology inference algorithm were used for simulation, and the results are shown in **Figure 4**. It can be found that, under the same sample size, the inference effect of the fast topology inference algorithm is significantly better than the traditional topology inference algorithm. In other words, the proposed algorithm significantly improves the effectiveness of samples and the speed of inference.
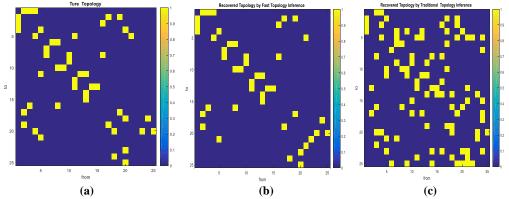
| (a) | (b) | (c) |

Fig. 4. Comparison of simulation results based on different algorithms, in which the yellow represents that the two nodes are connected and the blue represents that the two nodes are disconnected. The subfigure **(a)** is the true topology; the subfigure **(b)** is the recovered topology by fast topology inference algorithm; the subfigure **(c)** is the recovered topology by traditional topology inference algorithm.

This paper firstly uses the probability of missed alarm and false alarm to measure the effect of topology inference, namely:

$$FNR = \frac{FN}{FN+TP} = \frac{FN}{P} \tag{9}$$

and

$$FPR = \frac{FP}{FP+TN} = \frac{FP}{N(N-1)-P} . \tag{10}$$

Among them, *FNR* represents the probability of missed alarm and *FPR* represents the probability of false alarm. When there are *N* nodes in the target wireless network, assuming that the total number of links in which there is a communication relationship is *P*, the total number of links in which there is no communication relationship is naturally *N(N-1)-P*.

Given the actual topology, the MonteCarlo simulation method is used to calculate the false alarm and missed alarm, and the number of repetitions is 1000 times. The results are shown in **Figure 5**. It can be found by comparison that the proposed fast topology inference algorithm is significantly superior to traditional topology inference algorithm in both false alarm and missed alarm indicators.

Based on Bayesian criteria, a comprehensive cost function *C* can be introduced as follow:

$$C = C_{10} * FPR * P_0 + C_{01} * FNR * P_1, \tag{11}$$

where C_{10} represents the cost of the false alarm, and C_{01} represents the cost of missed alarm. These two parameters generally take 1. And P_0 and P_1 are the prior probabilities of link and non-link between nodes in the network. It can be understood that the smaller the cost become, the better the inference performance.

From **Figure 6**, it can be more intuitively found that in a small sample size scenario, the cost of the fast topology inference algorithm is far less than the traditional topology inference algorithm, that is, the comprehensive performance of the fast topology inference algorithm is greatly improved.
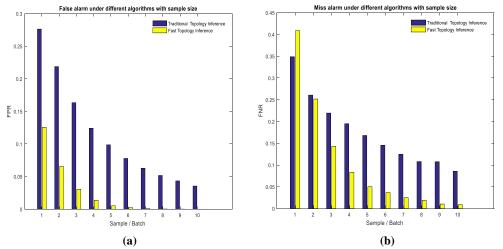
Fig. 5. Comparison of specific performance of different algorithms, in which the blue represents the traditional topology inference algorithm and the yellow represents that the fast topology inference algorithm. The subfigure (a) shows the false alarm varies with sample size; the subfigure (b) shows the missed alarm varies with sample size.
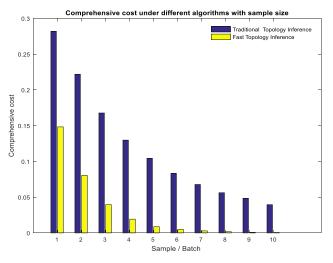

Fig. 6. Comparison of the comprehensive cost of different algorithms, in which the blue represents the traditional topology inference algorithm and the yellow represents that the fast topology inference algorithm.

4.3 Performance analysis

For fast topology inference algorithms, it is necessary to make a balance between inference performance and sample size. Therefore, the sample size is introduced to form a new comprehensive cost function. Assuming that the sampling is homogeneous, each batch of

samples will consume a same cost, then the total cost function of topology inference after the kth batch of samples becomes:

$$C = C_{10} * P_f^k * P_0 + C_{01} * P_m^k * P_1 + cE\{k\}, \tag{12}$$

where P_f^k is the false alarm after the kth inference, P_m^k is the missed alarm after the kth inference, and $E\{\bullet\}$ represents the average value under the average distribution. Therefore, the optimal choice for the sample size is:

$$\inf_{k \in \mathcal{K}}[C_{10} * P_f^k * P_0 + C_{01} * P_m^k * P_1 + cE\{k\}]. \tag{13}$$

Constraint (13) provides a way to find the fastest topology inference.

5 Summary and Outlook

In this paper, we study the fast topology inference for small sample size scenarios in wireless networks. Under the single-node sensing model, by considering both reliability and effectiveness, we propose a fast topology inference algorithm based on the Hawkes process, and give a sample selection formula for the fastest topology inference. Simulation results show that the proposed algorithm outperforms traditional recognition algorithms in small sample size scenarios. In the next step, we can conduct research in the framework of distributed sensing and further explore the mathematical logic of the fastest topology inference.

References

[1] T. Bouchoucha, C. Chuah and Z. Ding.: Topology Inference of Unknown Networks Based on Robust Virtual Coordinate Systems. IEEE/ACM Transactions on Networking. vol. 27, no. 1, pp. 405-418(2019)

[2] G. B. Giannakis, Y. Shen and G. V. Karanikolas. :Topology Identification and Learning over Graphs: Accounting for Nonlinearities and Dynamics. Proceedings of the IEEE. vol. 106, no. 5, pp. 787-807(2018)

[3] M. Piggott, V. Solo.: Non-Negative Online Estimation for Hawkes Process Networks. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB. pp. 6333-6337(2018)

[4] M. G. Moore, M. A. Davenport.: Analysis of wireless networks using Hawkes processes. 2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Edinburgh. pp. 1-5(2016)

[5] M. G. Moore, M. A. Davenport.: A Hawkes' eye view of network information flow. 2016 IEEE Statistical Signal Processing Workshop (SSP). Palma de Mallorca. pp. 1-5(2016)

[6] Z. Liu, J. Sun, F. Shen, G. Ding, Q. Wu.: Topology sensing in wireless networks by leveraging symmetrical connectivity. Machine Learning and Intelligent Communications. pp.49-57(2019)