IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 17, 2019, accepted October 11, 2019, date of publication November 6, 2019, date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2952059

Research on Communication Network Structure Mining Based on Spectrum Monitoring Data

CHANGKUN LIU, XINRONG WU, LEI ZHU, CHANGHUA YAO, LU YU, LEI WANG, HAOREN FAN, AND TING PAN

College of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China

Corresponding authors: Xinrong Wu (njwxr@163.com), Lei Zhu (zhulei_paper@126.com), Changhua Yao (ych2347@163.com), Lu Yu (yulu_mail@263.net), and Lei Wang (iponly@126.com)

This work was supported in part by the Natural Science Foundation of Jiangsu Province of China under Grant SBK2019020930 and Grant BK20191329, in part by the Postdoctoral Science Fund of China under Grant 2018M633684, and in part by the National Natural Science Foundation of China under Grant 61971439.

ABSTRACT The physical characteristics of the massive spectrum signals carrying the communication information and the statistical laws of these characteristics also potentially reflect the communication behavior of the communication individuals and the intelligence information related to the communication behavior. Intercepting and cracking signal content usually faces enormous difficulties and costs, and more often, we are not able to crack the encrypted signal content. However, by studying the physical features extracted from the spectrum monitoring signals and the statistical laws of these features, it is also possible to dig out the hidden relationships between communication individuals and even the communication network structure, so as to analyze the communication behaviors of the communication individuals. Based on the characteristics of carrier frequency, bandwidth, power, signal monitoring time and direction information of spectrum monitoring signals, this paper identifies each spectrum signal and studies the distribution characteristics and statistical laws of massive spectrum monitoring signals in the column coordinate system. Due to the clustering of the spectrum signals generated by the sources in the power, monitoring time and direction, and the correlation of the spectrum signals generated by the two parties in the communication process, based on the improved density clustering algorithm, this paper proposes a method for mining the communication relationship between communication individuals from the spectrum monitoring data, and guesses and constructs the communication network structure by matching the communication individual with the communication relationship. Finally, we analyze the communication network structure mined from the spectrum monitoring data.

INDEX TERMS Spectrum monitoring data, communication network structure, communication relationship discovery, data mining, density clustering.

I. INTRODUCTION

With the rapid development of wireless communication, the scarcity of spectrum is becoming more and more prominent. It is increasingly urgent to strengthen the monitoring and analysis of spectrum signals and the management of electromagnetic spectrum [1]–[3]. As the medium of information transmission, it is of great significance to study the spectrum signal [4]. At present, the mining and analysis of massive spectrum monitoring signals mainly focus on a spectrum situation display, signal feature extraction, signal classification and other aspects [5]. The physical characteristics of the massive spectrum signals carrying the communication information and the statistical laws of these characteristics also potentially reflect the communication behavior of the communication individuals and the intelligence information related to the communication behavior. However, it is almost empty to research the communication relationship, communication network structure and communication behavior of communication individuals from the massive spectrum signals only according to the physical characteristics of spectrum signals and the statistical rules of these characteristics, without relying on cracking the signal content.

The associate editor coordinating the review of this manuscript and approving it for publication was Imran Ahmed.

In the fields of anti-terrorism, military communications, communication investigation, communication security and so on, intercepting and cracking signal content usually faces enormous difficulties and costs, and more often, we are not able to crack the encrypted signal content. Therefore, it is difficult to obtain the communication behavior of the communication individual based on the signal content in some specific scenarios. However, the physical characteristics of the spectrum signals are difficult to encrypt, and these features are easy to obtain. By studying the physical characteristics extracted from the spectrum signals and the statistical laws of these characteristics, it is also possible to dig out the communication relationships between communication individuals, the communication network and other hidden information, so as to further analyze and obtain the communication behavior of the communicating individuals.

The communication relationship reflects the communication connections among communication individuals. We can further speculate and construct the communication network structure by analyzing the statistical rules of communication time, communication duration, communication times, the sequence of communication, the communication direction and other characteristics of the communication individual. The communication network reflects the communication relationship and communication behavior among communication nodes (communication individuals) in the monitoring area. Through the research on network connectivity, network communication path, and other information, it is possible to analyze the network structure and hierarchy, and estimate the node level in the network.

Due to environmental factors, the wireless channel has a high bit error rate. The data link layer usually adopts the stop-and-wait ARQ (Automatic Repeat Request) protocol to ensure reliable data transmission [6]. Therefore, each time when the sender sends a data frame, the receiver need to reply to the feedback information for confirmation, so that the spectrum signals generated by the two parties are related in time. On the other hand, the continuity of communication makes the spectrum monitoring signal continuous in time. Even if the source position changes, the direction, and power of the spectrum signal detected in a short time remain relatively stable. Spectral signals show clustering property in three dimensions of direction, power and monitoring time.

In order to mine the communication network from the spectrum monitoring data, this paper first discusses the characteristics of the spectrum monitoring data and the factors affecting the spectrum monitoring data. Secondly, features such as signal frequency, bandwidth, signal power and signal direction are extracted from spectrum monitoring data to identify each spectrum signal. In the column coordinate system constructed by signal direction, signal power and signal monitoring time, we study the distribution characteristics and statistical rules of mass spectrum moni-toring signals. The spectrum signals are clustered by the improved OPTICS (Ordering Points to Identify the Clustering Structure) algorithm, and clustering sets represent the spectrum signal sets generated by different communication individuals in different communication processes. Then, through the communication network structure mining method proposed in this paper, we discovered the communication relationship between the stations and represented the network nodes in the polar coordinate system, thus constructing the communication network structure. Finally, we analyze the network structure to obtain the communication behavior information of the station.

The experimental results show that the method has good adaptability to the massive spectrum monitoring signals, and can mine the communication relationship between the source nodes from the spectrum monitoring data, and infer the communication network structure. In addition, the method realizes the communication behavior research of the communication individual through statistical analysis of the network structure and the number of node communication.

The contributions of this paper are summarized as follows: Firstly, this paper mines the communication relationship between the communication individuals and the commun-ication network structure from the spectrum monitoring data, and obtain the communication behavior of the communication individuals. Secondly, the research in this paper is not to rely on communication content and frame structure, but by mining the statistical laws of spectrum signals to obtain the communication relationship between communication indivi-duals and the communication network structure. Besides, it lays foundation for further study on analysis of commun-ication behavior rules and provides a new perspective of analysis and mining of massive spectrum monitoring data. Thirdly, we process the spectrum monitoring data in the cylindrical coordinate system, change the ε-neighborhood in the OPTICS algorithm to the (ε, h)-neighbor domain suitable for the cylindrical coordinate system, and study the relation-ship between ε, *MinPts* and h.

II. RELATED WORK

Although a large number of literatures have conducted in-depth studies on spectral signals, these studies have focused more on the characteristics and information of the spectral signals themselves, such as the estimation of spectral signal related parameters [7]–[9], signal detection [10], [11], ano-maly detection based on signal characteristics [12]–[15], monitoring and management of spectral signals [1]–[3], and spectrum sensing [16]–[19], spectrum decision [20], [21] and other related research. For the massive spectrum signals generated by communication, it is not deep enough that the research on mining the connection between spectral signals and the communication relationship between the communi-cation individuals that generate spectral signals and analyzing the behavior characteristics of these communication individuals.

The existing research on the communication behavior of wireless communication individual mainly relies on monitoring or eavesdropping to crack the content of the intercepted spectrum signal [22]–[24], and analyzes com-munication behavior and intention according to the content of the

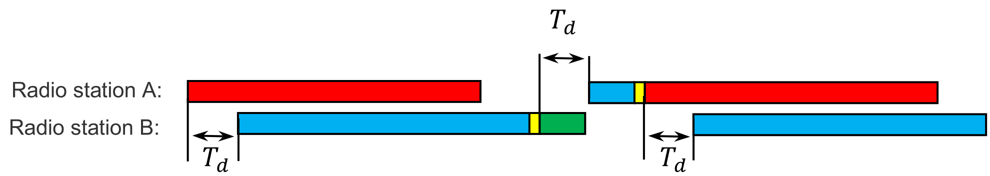


FIGURE 1. Schematic diagram of the transmission and reception process of a data frame based on stop-and-wait ARQ.

signals [4], [5]. But more often, these methods do not capture important communications that have been encrypted. In order to avoid the cost and difficulty of cracking the signal content, and to avoid the applicability of the analysis methods based on communication content and prior knowledge (communication protocol frame format, etc.) in specific scenarios, the research in this paper is not to rely on communication content and frame structure, but by mining the statistical laws of spectrum signals to obtain the communication relationship between communication individuals and the communication network structure. Besides, it lays foundation for further study on analysis of communication behavior rules.

On the other hand, based on the physical characteristics of the signal, the research on mining massive spectrum monitoring data is not deep enough. Akyildiz *et al.* [3] analyzed the differences between the two communication modes of frequency hopping communication and fixed frequency communication, and classified the spectrum signals by making classification rules to find the communication relationship between different classification sets. However, this method only works with ideal and complete spectrum monitoring data, and cannot effectively analyze the missing spectrum monitoring data. Pan *et al.* [25] used multiple monitoring devices to monitor spectral signals, and mined the communication relationship between information sources, combined with signal fading model to locate information sources. However, the fading of signals is different in different environments, and there is a large error in the location of the geographical location of the source. Liu *et al.* [5] proposed a method based on improved density clustering to mine the communication relationship in spectrum monitoring data, which provides an idea for the analysis of spectrum monitoring data. However, for further analysis, the article does not realize the identification of network nodes, so it is impossible to speculate and construct the communication network structure. Comparing the above three methods, we propose a method for mining the communication relationship between communication individuals and the structure of communication networks from spectrum monitoring data. It is a further study on mining spectrum monitoring data and analyzing the communication behavior of communication individuals.

Finally, due to the manifold distribution characteristics of the spectrum data, in the process of data mining, we use the method of density clustering such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [26]–[28], OPTICS [29] to process the data.

III. THE DATA PREPROCESSING

This section firstly discusses the influence of communication protocol and scanning period of monitoring equipment on spectrum monitoring data. Then, based on the characteristics of spectrum monitoring data, we propose the features of identifying communication relationship and speculating communication network, and express the distribution of data in the column coordinate system.

A. FACTORS AFFECTING SPECTRUM MONITORING DATA

1) THE IMPACT OF COMMUNICATION PROTOCOLS ON SPECTRUM MONITORING DATA

Since the wireless channel is easily interfered by the environment, with a high bit error rate, in order to ensure reliable transmission of data, the radio station that adopts the half-duplex communication mode usually adopts the stop-and -wait ARQ protocol at the data link layer. Fig. 1 shows the length of time occupied by the sending of information frame and the reply of confirmation frame in the communication process based on the stop-and-wait ARQ protocol. The red rectangle indicates the duration of the information frame sent by the station, the green indicates the duration of the confirmation frame (error pattern) sent by the station, the blue indicates the duration of the station receiving the information frame (or the confirmation frame), the yellow indicates the conversion time of transmission and reception, and the blank interval indicates the propagation delay T_d. Therefore, for a pair of communication stations, the monitored spectrum signal set is jointly generated by both the transmitting station and the receiving station, that is, corresponding to two sources. The carrier frequency of the frequency hopping communication continuously changes, and the information is not transmitted when the channel is switched. However, the carrier frequency of fixed frequency communication remains unchanged, and the amount of data detected in the same time is larger.

2) THE IMPACT OF SCAN CYCLE ON SPECTRUM MONITORING DATA

For monitoring equipment, the scanning period is affected by the monitoring range and the monitoring scanning rate. Different scanning periods correspond to monitoring data of different densities. Fig. 2 shows the amount of data collected by the monitoring device based on different scanning periods, where the green rectangle indicates the duration of the spectrum signal propagating to the monitoring device, and purple

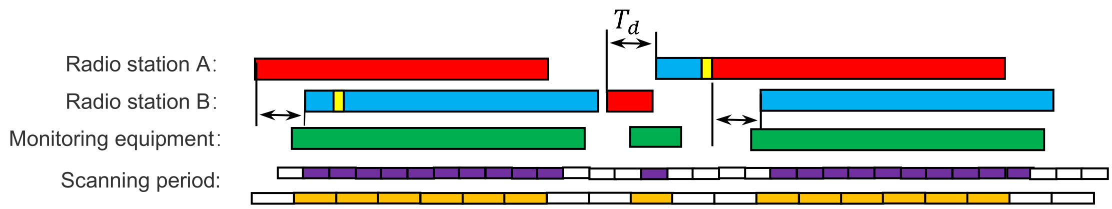


FIGURE 2. Schematic diagram of monitoring data density of fixed-frequency communication spectrum.

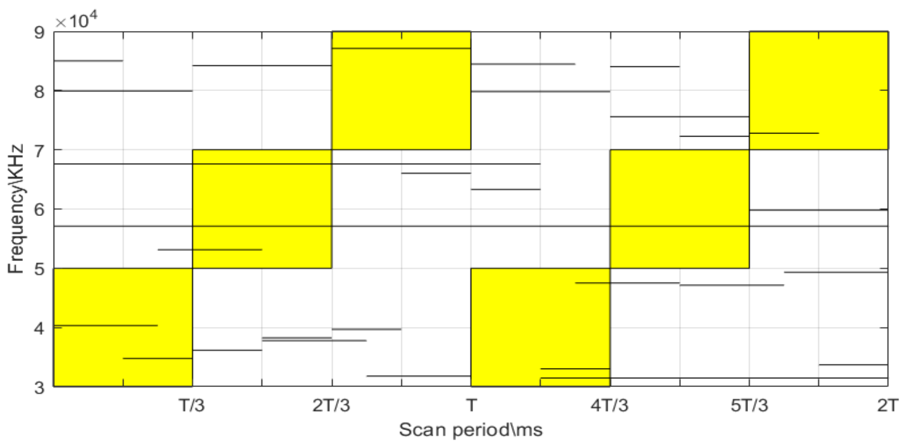


FIGURE 3. Schematic diagram of monitoring the frequency hopping signal.

and orange indicate the monitoring conditions corresponding to different scanning periods. Obviously, the smaller the scan period is, the more spectrum data will be detected. For fixed-frequency communication, the main reasons affecting data distribution are propagation delay and scanning period.

On the other hand, for frequency hopping communication, there will be more missing spectral signals. In order to resist interference, the carrier frequency of frequency hopping communication is constantly changing. The characteristics of this hopping also largely avoid the monitoring of monitoring equipment. Fig. 3 shows the monitoring of frequency hopping communication. The abscissa is time, in units of scan cycles, and the ordinate is frequency. The yellow rectangle in the Fig.3 indicates the frequency range monitored within one scan bandwidth, and the horizontal lines with different lengths represent different frequency hopping signals. The monitoring equipment monitors within the range of 30-90MHz, and the scanning bandwidth (corresponding to the height of the yellow rectangle in the figure) is 20MHz. One scanning period corresponds to three yellow rectangles. The signal in the white area is undetected, so the actual scanned signal contains a large number of missing signals. This lack results in a smaller data density of monitoring data and uneven distribution of data. In the data processing process, it is necessary to consider a data processing method that can accommodate data missing.

B. FEATURE SELECTION

Due to the interaction and transmission of information, there is a communication relationship between communication individuals, which constitutes the basis of the communication network. In order to mine the communication relationship and communication network among sources from the spectrum monitoring data, this paper firstly classifies the spectrum signals by clustering method based on the characteristics of spectrum signals. Each cluster set corresponds to the spectrum signal set generated by the source in their respective communication. Then we replace the source with the clustering set, and determine the communication relationship between the source nodes according to the distribution characteristics of the data of the clustering set in time. Finally, we construct a communication network based on the communication relationship between nodes, so as to mine the communication network from the spectrum monitoring data.

1) SIGNAL POWER

The signal power represents the distance information of the station relative to the monitoring device. Due to the error of the monitoring equipment and the fading during signal propagation, the monitored signal power exhibits a normal distribution. If the position of the station is relatively fixed, the power of the monitored spectral signal exhibits a stable distribution. Even if the radio is moving, the power of the spectrum signal monitored in a short period of time exhibits a stable distribution or change.

2) SIGNAL DIRECTION (ANGLE)

The signal direction (angle) represents the direction information of the station relative to the monitoring device. Due to the error of the monitoring equipment and the fading during signal propagation, the monitored signal power exhibits a normal distribution. If the position of the station is relatively fixed, the angle of the monitored spectral signal exhibits a stable distribution. Even if the radio is moving, the angle of the spectrum signal monitored in a short period of time exhibits a stable distribution or change.

3) SIGNAL MONITORING TIME

Since wireless communication has a large error rate, the station usually performs error control based on stopping waiting for the ARQ protocol. The sender sends the data frame, and the receiver sends the acknowledgement frame immediately, so the spectrum signals generated by both parties are roughly the same in the time range. On the other hand, the spectrum monitoring signal appears continuously in time, so that the spectrum signal exhibits a stream pattern in the time domain.

Features such as carrier frequency, signal bandwidth, signal power, time of signal occurrence and signal direction that extracted from spectrum monitoring data carry important information of spectrum signals and can uniquely identify spectrum monitoring signals. Therefore, we take them as features of mining communication relations and speculating communication network structure.

C. FEATURE REPRESENTATION

Let the spectrum monitoring data set be $X = \{x_1, x_2, \cdots, x_i, \cdots, x_n\}^T$, where $x_i = \{f_i, B_i, \theta_i, P_i, t_i\}$, f_i represents the signal frequency, B_i represents the signal bandwidth, θ_i represents the signal direction, P_i represents the signal power, and t_i represents the signal monitoring time. In order to study the clustering properties of spectral data more intuitively, this paper introduces a cylindrical coordinate system to describe the distribution of data. Let the spectrum monitoring data set be $Y = \{y_1, y_2, \cdots, y_j, \cdots, y_n\}^T$, where $y_j = \{\theta_j, P_j, t_j$. Fig. 4 shows the distribution of spectral data generated by a pair of communication stations in a cylindrical coordinate system. Obviously, the data shows clustering and communication directivity. The data set Y is projected into the polar coordinate system in the cylindrical coordinate system, and we obtain the data set $Z = \{z_1, z_2, \cdots, z_k, \cdots, z_n\}^T$, where $z_k = \{\theta_k, P_k\}$. Fig. 5 shows the distribution of the data set Z in a polar coordinate system, the origin representing the position of the monitoring station. The distribution of data in polar coordinates indicates radio stations relative position to the monitoring equipment and the relative position among stations. This provides a node location for the construction of the communication network, although it is not a real geographical location.

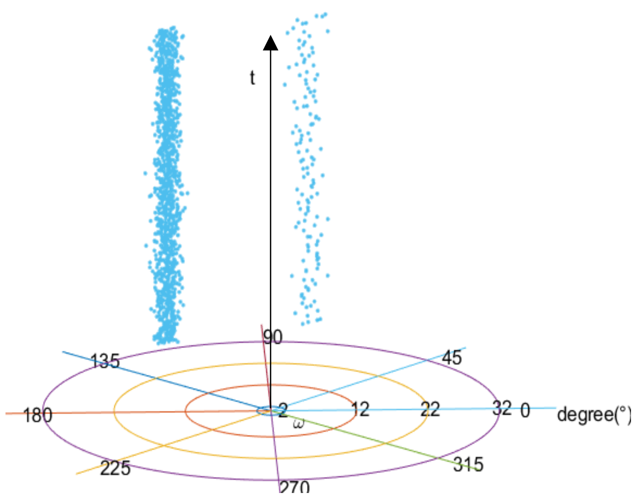


FIGURE 4. The distribution of spectrum signal in the cylindrical coordinates.

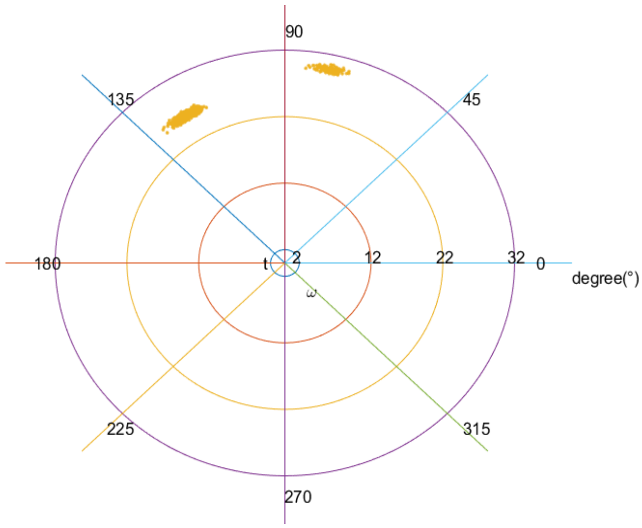


FIGURE 5. The distribution of spectrum signal in the polar coordinates.

IV. COMMUNICATION RELATIONSHIP DISCOVERY

A. COMMUNICATION RELATIONSHIP MINING METHOD

Mining the communication relationship between the sources from the spectrum monitoring data is to classify the monitored spectrum signals according to the characteristics of the signals. In this process, the spectrum signals generated by each communication station during each communication process are separated from the spectrum monitoring data. The spectrum data of the classification set represents the spectrum signal generated by the radio station in a communication. Based on the classification results, the classification sets of spectrum data with similar time range are matched to mine the communication relations.

B. DENSITY CLUSTERING

Signal power, signal direction, and signal monitoring time can uniquely identify the monitored spectrum signal. Because of the propagation delay and path loss and the error of the monitoring equipment, the monitored spectrum signal has errors in signal power and signal direction. These errors result in the approximate normal distribution of signal power and direction. Signal monitoring time represents the time when the signal appears. Spectrum monitoring data is acquired

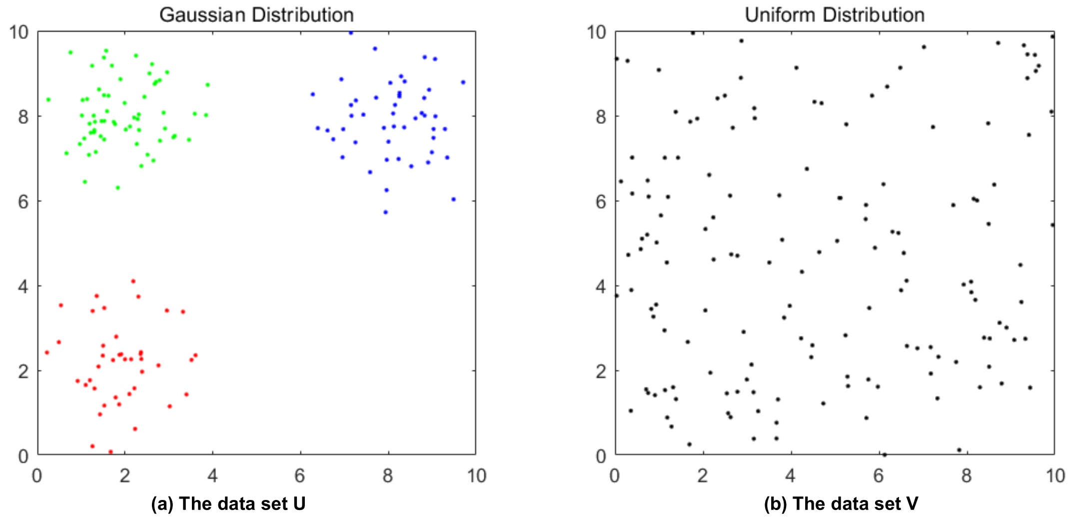


FIGURE 6. The distribution of data.

based on scanning of monitoring equipment, and continuous communication causes the monitored spectral signals to be continuous in time. The data represented by signal power, signal direction and signal monitoring time exhibit manifold clustering, as shown in Fig. 4. On the other hand, because the monitored data is missing and confusing, it also determines to classify the data by clustering.

The data of data set Y exhibits manifold characteristics. In the dimension of time, since the scanning period is constant, the spacing of the generated data is relatively stable, so scaling the spacing of the data in the time dimension does not change the clustering characteristics of the data. On the other hand, we study the distribution law of data in the cylindrical coordinate system. Based on the characteristics of the cylindrical coordinate system, this paper changes the spherical ε-neighborhood in the original OPTICS algorithm to a columnar neighborhood, and the neighborhood is defined as:

$$N_\varepsilon\left(y_j\right) = \left\{y_i \in D \,|\, dist\left(y_i - y_j\right) \leq \varepsilon\right\} \tag{1}$$

where

$$dist\left(y_i, y_j\right) = \sqrt{(\theta_i - \theta_j)^2 + \left(P_i - P_j\right)^2 + \delta(t_i - t_j)} \tag{2}$$

$$\delta\left(t\right) = \begin{cases} 0, & |t| \leq h \\ \infty, & |t| > h \end{cases} \tag{3}$$

h is the threshold of the time difference between the data, which determines the height of the columnar field. ε determines the bottom area of the columnar neighborhood. After defining the column neighborhood $N_\varepsilon\left(y_j\right)$, the value of $MinPts$ needs to be further determined, and ε and h are estimated to determine the range of the neighborhood.

Daszykowski *et al.* [30] proposes that the selection of $MinPts$ value in the neighborhood depends on the number of objects in the data. In addition to this, the distribution characteristics of the data and additional information about the data cluster can also be used to define $MinPts$.

Based on the value of the preset $MinPts$, we estimate ε and h. Daszykowski *et al.* [30] optimizes the neighbor-hood radius ε by estimating the data set with the same dimension as the research data but uniformly distributed within the experimental range, regardless of the distribution of objects in the data set. As shown in Fig. 6, the data set U contains m data points and follows the normal distribution. The data set V is uniformly distributed and is the same as the data dimension, the number of data, and the experimental range of the data set U. Selecting the optimal neighborhood radius ε for the data set U is to calculate the distance of each object in the data set V to its $MinPts - th$ neighbor, sort the m calculated distances in ascending order, and then select a distance equal to 95% as ε.

Inspired by the literature [30], in order to estimate the column neighborhood, this paper combines the distribution of data objects to estimate the columnar neighborhood of the data set with the same data dimension but uniformly distributed within the experimental range. In cylindrical coordinate system, the data set Y presents local manifold distribution, and different clustering sets have similar density and distribution characteristics. In the communication process, the duration of the acknowledgment message sent by the receiving station is less than the length of time that the transmitting station transmits the information. Within the same time, the number of signals sent by the receiving station is monitored to be small, and the density of the receiving station spectrum data in the cylindrical coordinates is small, as shown in the Fig. 4. The difference in spectral monitoring data density determines the columnar neighborhood formed by ε and h based on the cluster set of the smaller density of the receiving stations, and such columnar neighborhood is still valid for dense data.

Let the spectrum signal set generated by a certain receiving station be R $= \{\theta_i, P_i, t_i\}$, where $i = 1, 2, \cdots, m$. For a more intuitive representation, the data set R is transformed into a three-dimensional cartesian coordinate system to obtain $R' = \{x_i, y_i, t_i\}$ by the formula (4).

$$\begin{cases} x_i = P_i cos\theta_i \\ y_i = P_i sin\theta_i \end{cases} \tag{4}$$

The range occupied by the data set R' in space is denoted as V_R. Let R'' be a data set with the same dimensions and experimental range as the data set R', but subject to uniform distribution. The average range occupied by each object in R'' can be expressed as $\frac{V_R}{m}$, where

$$V_R = \pi \max_{1\le i,j\le m} \frac{1}{4}\left[\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2\right] \cdot \left(\max_{1\le i,j\le m} t_q - \min_{1\le i,j\le m} t_p\right) \tag{5}$$

$$MinPts \cdot \frac{V_R}{m} \le 2h\pi\varepsilon^2 \le (MinPts + 1) \cdot \frac{V_R}{m} \tag{6}$$

$2h\pi\varepsilon^2$ represents the range occupied by the cylindrical (ε, h)-neighborhood, and $MinPts \cdot \frac{V_R}{m}$ represents the average range corresponding to $MinPts$ points in the neighborhood of each object. Based on the given $MinPts$, eq(6) determines the relationship between h and ε and the range of the columnar (ε, h)-neighborhood.

C. MATCH CLUSTERS TO DETERMINE communication RELATIONSHIPS

The spectrum monitoring data is classified by the improved OPTICS algorithm, and each cluster set represents the spectrum signal set generated by the station in one communication, as shown in Fig 4. Based on the stop-and -wait ARQ, the communicating parties maintain the trans-mission and acknowledgement of the data frames during the communication. Therefore, for two stations with communication relationship, the distribution of the generated spectrum signals is similar in the time range, that is, the initial signal time and the end time corresponding to the two cluster sets are similar. Therefore, the communication relationship of signal sources can be confirmed according to the distribution of time.

Time complexity analysis: The time complexity of the OPTICS algorithm is $O(n^2)$, the time complexity of calculating the center position of the cluster set U_l is $O(n)$, and the time complexity of calculating the time range of the cluster set U_l is $O(n)$, and the time complexity of communication relationship matching is $O(n)$, so the time complexity of algorithm 1 is $O(n^2)$.

V. THE NETWORK STRUCTURE MINING AND ANALYSIS

A. CONJECTURE OF COMMUNICATION NETWORK STRUCTURE

Mining the communication network in the spectrum monitoring data is to classify the spectrum monitoring data by clustering method. Then the relative position of the cluster set in the cylindrical coordinate system is taken as the node of the network. Finally, based on the communication relationship between clustering sets, we connect nodes to build the network and record the communication direction.

Algorithm 1 The Communication Relationship Discovery Algorithm

Input: data set $Y = \{y_1, y_2, \cdots, y_j, \cdots, y_n\}^T$, where $y_j = \left\{\theta_j, d_j, t_j\right\}$.
ε, $MinPts$, h

Output: Signal spectrum set V corresponding to the com-munication relationship
The centroid position $(\bar{\theta}_i, \bar{P}_i)$ of the source
Communication direction
Communication sequence

1: According to the distance defined by formula (2) (3), use the OPTICS algorithm to cluster the data to obtain the clustering set $U = \{U_1, U_2, U_3, \cdots, U_l, \cdots\}$ of spectr-um signals
2: Calculate the centroid position $(\bar{\theta}_i, \bar{P}_i)$ of the cluster set U_l projected to the polar coordinate system
3: Sort the objects of the cluster set U_l according to time, and extract the initial time and end time of signals of the cluster set U_l
4: Calculate the time range of the data in the cluster set U_l
5: Matching the cluster set U_l to discover the communi-cation relationship of information interaction
6: *if* the initial time of U_l is close to that of U_j
7:     *if* the end time of U_l is close to that of U_j
8:         There is a communication relationship between U_l and U_j.
9:         Compare the number of data of U_l and U_j, the number of receivers is small, while the sender is large.
10:         $V_k = \{U_l, U_j\}$ is the spectrum set correspondding to the communication relationship
11:     *end if*
12: *end if*
13: Output spectrum signal sets corresponding to differ-ent communication relationships in the cylindrical coordinate system

In the process of building the communication network, the nodes of the network must first be determined. The data set $Z = \{z_1, z_2, \cdots, z_j, \cdots, z_n\}^T$ represents the direction and power information of each spectral signal in the spectrum monitoring data, where $z_j = \left\{\theta_j, P_j\right\}$. In the polar coordinate system, data set Z describes the relative position information of the spectral signals, and the data presents the clustering distribution. The DBSCAN algorithm implements a division of the data set $Z = \{C_1, C_2, \cdots, C_p, \cdots, C_m, D$, where $p = 1, 2, 3, \cdots m$. The data distribution of the cluster set C_p represents the relative position of the source in the polar coordinate system, and D is the set of abnormal points. The centroid neighborhood of each cluster set C_p represents the relative position of the source and acts as a node of the communication network. The centroid position $\bar{C_p}$ of the cluster set $C_p = \{c_{p1}, c_{p2}, \cdots, c_{pi}, \cdots, c_{pk}\}$ (where $c_{pi} = (\theta_{pi}, P_{pi})$)

is expressed as:

$$(\bar{\theta_p}, \bar{P_p}) = \frac{1}{k}\sum_{i=1}^{k} c_{pi} \tag{7}$$

In order to record and study the communication relationship and gradual change process of the communication network in different time periods, we divide the data set Y into $\mathrm{Y} = \{\mathrm{Y}_1, \mathrm{Y}_2, \cdots, \mathrm{Y}_i, \cdots\}$ according to the time interval $t_{interval}$. It should be emphasized that the seg-mentation of data set Y is necessary. Only in this way can we intuitively analyze the communication relationships, communication sequences, network connectivity, paths, and communication directions in different time periods. Based on Algorithm 1, we mine the communication relationship of Y_i, and record the communication direction, communication sequence, and calculate the relative position of the source in polar coordinates $(\bar{\theta_l}, \bar{P_l})$, where $l = 1, 2, \cdots\cdots$. In order to correctly match the source relative position $(\bar{\theta_l}, \bar{P_l})$ in Y_i with the network node $(\bar{\theta_p}, \bar{P_p})$, we set the neighborhood range of $\bar{C_p}$:

$$\begin{aligned} &\mathrm{N}_r\left((\bar{\theta_p}, \bar{P_p})\right) \\ &\quad = \{(\theta_i, P_i) | d\left[\left(\bar{\theta_p}, \bar{P_p}\right), (\theta_i, P_i)\right] \leq r\} \end{aligned} \tag{8}$$

$$\begin{aligned} &d\left[\left(\bar{\theta_p}, \bar{P_p}\right), (\theta_i, P_i)\right] \\ &\quad = \sqrt{(P_i cos\theta_i - \bar{P_p} cos\bar{\theta_p})^2 + (P_i sin\theta_i - \bar{P_p} sin\bar{\theta_p})^2} \end{aligned} \tag{9}$$

If the relative position $\left(\bar{\theta_l}, \bar{P_l}\right)$ of the source in Y_i is within the neighborhood of $\bar{C_p}$, $(\bar{\theta_l}, \bar{P_l})$ is considered to represent the network node, and the nodes are connected according to the communication relationship. Each Y_i corresponds to a communication network Ω_i with a time range of $t_{interval}$, and Ω_i records the communication within this time range. The change of the communication network Ω_i to Ω_{i+1} corresponds to the change of the node communication relationship and the evolution of the network with time. The communication network $\Omega_i (i = 1, 2, 3, \cdots\cdots)$ is superimpo-sed to form the communication network Ω corresponding to the data set Y.

It should be noted that when Y is divided according to the time interval $t_{interval}$, continuous communication spectrum monitoring data belonging to one class may be divided into adjacent subsets Y_j, $j = i, i+1, i+2, \cdots\cdots$. Therefore, when Ω_j is combined, it is necessary to treat the same communication relationship continuously distributed in the adjacent network Ω_j as one, to ensure an accurate number of communication relationships.

Time complexity analysis: the time complexity of DBSCAN algorithm for data set Z clustering is $\mathrm{O}(n^2)$, the time complexity of computing network nodes is $\mathrm{O}(n)$, and the time complexity of communication relation discovery algorithm is $\mathrm{O}(n^2)$. Therefore, the total time complexity of algorithm 2 is $\mathrm{O}(n^2)$. In the future work, targeted or dynamic segmentation of the data set is an effective way to reduce the cost of the algorithm.

Algorithm 2 Conjecture of Communication Network structure

Input: data set $\mathrm{Y} = \{y_1, y_2, \cdots, y_j, \cdots, y_n\}^T$, where
$y_j = \{\theta_j, P_j, t_j\}$
$\varepsilon_1, MinPts_1$
$t_{interval}$
$\varepsilon_2, Minpts_2, h, r$
Output: network node coordinates $(\bar{\theta_p}, \bar{P_p})$
Communication network Ω
Communication order

1: Project the data set Y on the polar coordinate system plane to obtain a data set $\mathrm{Z} = \{z_1, z_2, \cdots, z_j, \cdots, z_n\}^T$, where $\mathrm{z}_j = \{\theta_j, P_j\}$.
2: Cluster the data set Z with the DBSCAN algorithm in polar coordinates, $\varepsilon = \varepsilon_1$, $MinPts = MinPts_1$. The cluster set is $\mathrm{Z} = \{C_1, C_2, \cdots C_p, \cdots, C_m, \mathrm{D}\}$
3: Calculate the centroid coordinates $(\bar{\theta_p}, \bar{P_p})$ of C_p as a network node according to formula (7)
4: According to the time interval $t_{interval}$, the data set Y is divided to obtain $\mathrm{Y} = \{\mathrm{Y}_1, \mathrm{Y}_2, \cdots, \mathrm{Y}_i, \cdots, \mathrm{Y}_n\}$
5: *for* $i = 1 : n$
6: According to algorithm 1, $MinPts = MinPts_2$, $\varepsilon = \varepsilon_2$, obtain the communication relationship, communication direction, communication order from Y_i, centroid position $(\bar{\theta_q}, \bar{P_q})$ of U_q in the polar coordinate system, $\mathrm{q} = 1, 2, \cdots\cdots$.
7: Match $\left(\bar{\theta_q}, \bar{P_q}\right)$ with the network node $(\bar{\theta_p}, \bar{P_p})$ according to formula (8)(9)
8: According to the communication relationship, connect the network nodes and get the communication network Ω_i
9: *end for*
10: Combine Ω_i to get the communication network Ω
11: Visual representation in the polar coordinate system

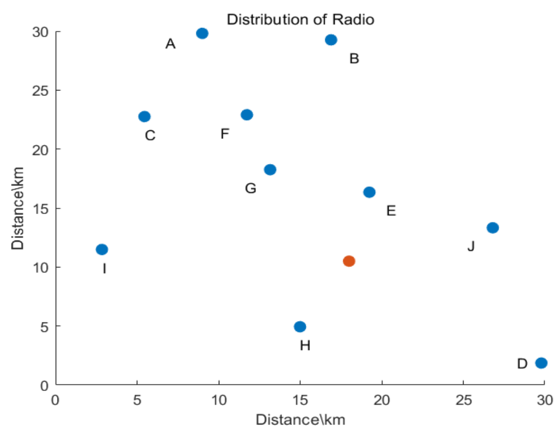


FIGURE 7. Location distribution of stations and monitoring equipment.

B. COMMUNICATION NETWORK STRUCTURE ANALYSIS

For the communication network mined from the spectrum monitoring data, this paper analyzes the nodes and structure of the network according to the characteristics of the network.

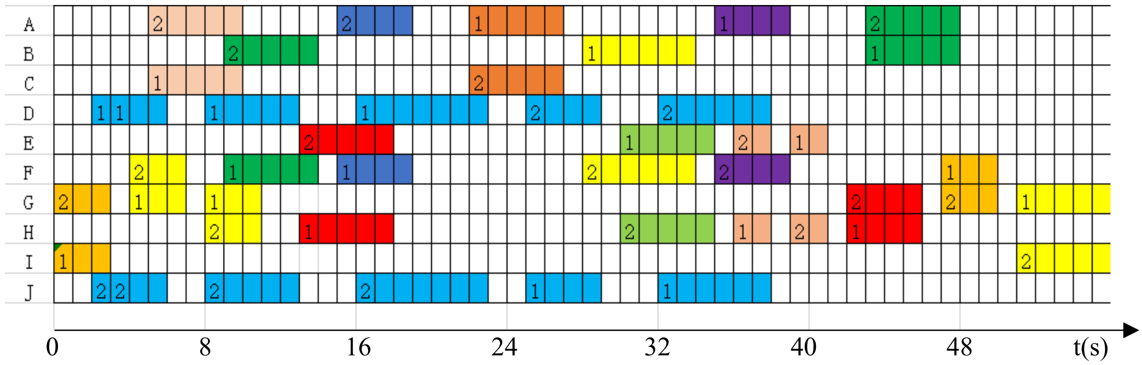


FIGURE 8. The time distribution of radio communication.

Referring to the basic idea of PageRank [32] algorithm, for network nodes, if a node is connected with other nodes, this node is very important in the network, that is, the PageRank value of the node will be relatively high. If a node with a high PageRank value is connected to one of the other nodes, the PageRank value of the node to which it is connected will increase accordingly [32]. Therefore, the analysis of the network structure needs to consider the statistical laws of nodes and edges.

Assume that the communication network mined from spectrum monitoring data is G = {V, E}, where V = {v_1, v_2, $\cdots\cdots$} represents the source node, E = {(e_1, p_1), (e_2, p_2), $\cdots$,(e_i, p_i), $\cdots$} represents the communication relationship among nodes, e_i represents the edge, and p_i represents the number of connections. In addition, we define $d^+(v_i)$ to indicate times the source is the sender, $d^-(v_i)$ to indicate times the source is the receiver, and $d(v_i)$ to indicate times the source participates in the communication. By analyzing the statistical characteristics of the number of nodes participating in the communication, the hierarchical positions of different nodes in the network are investigated.

In military communication networks, information is usually transmitted step by step, and communication nodes of different levels have different communication behavior characteristics. On the other hand, each node within the sub-network has a close relationship with other nodes, and communication between sub-networks may be carried out through higher-level communication. Although the nodes inside the network are all interoperable, the actual communication range and permissions of the nodes at different levels are different. For example, in a military communication network, each node is interconnected, but the communication range of the class radio stations in the communication network is limited by a company, and the information flowing to a higher level usually needs to be transmitted upwards step by step. Similarly, the command is issued step by step.

Network connectivity can be used to analyze the network structure. The path characterizes the direction of information transfer and the time sequence of communication between nodes, as well as the depth of the network. Analysis of the network communication path can identify critical paths, key nodes, and regional sub-network.

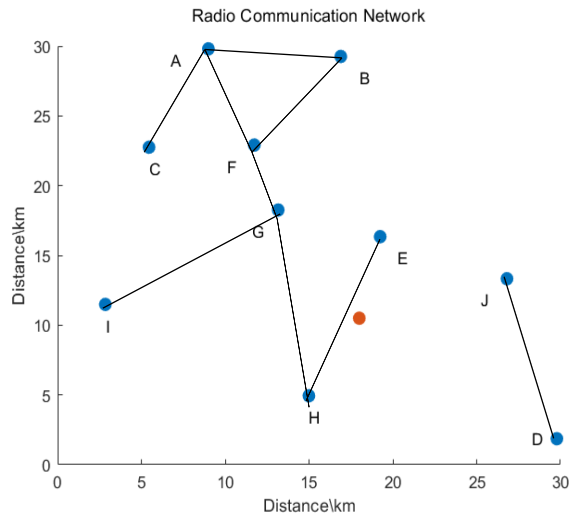


FIGURE 9. The communication network.

VI. SIMULATION EXPERIMENT

A. SCENE SETTING AND DATA COLLECTION

In the region of 30km in width and 30km in depth, 10 radio stations were randomly set as the experimental information sources, among which station D and J carried out fixed-frequency communication, and other stations carried out frequency-hopping communication. The spectrum range of radio communication is 30-90MHz, the scanning bandwidth of monitoring equipment is 20MHz, and the scanning rate is 80GHz/s. Fig. 7 shows the distribution of radio stations and monitoring equipment, where blue dots representing radio stations and red dots representing monitoring equipment.

Based on the radio and monitoring equipment set in the Fig. 7, we simulated the communication between the radio stations, monitored the spectrum signals through the monitoring equipment, and then mined the communication relationship and communication network from the spectrum monitoring data. Finally, we analyzed the structure of the communication network. In the communication model of the

TABLE 1. The communication sequence.

The serial number	Communication individuals and direction	Initial time of communication	End time of communic-ation	Communi-cation duration	Communicat-ion mode	Communicat-ion network	Station operating power
1	I → G	0''	3''	3s	Frequency hopping	Y	50W
2	D → J	2''	6''	4s	Fixed frequency	X	50W
3	G → F	4''	7''	3s	Frequency hopping	Y	50W
4	C → A	5''	10''	5s	Frequency hopping	Y	50W
5	D → J	8''	13''	5s	Fixed frequency	X	50W
6	G → H	8''	11''	3s	Frequency hopping	Y	50W
7	F → B	9''	14''	5s	Frequency hopping	Y	50W
8	H → E	13''	18''	5s	Frequency hopping	Y	50W
9	F → A	15''	19''	4s	Frequency hopping	Y	50W
10	D → J	16''	23''	7s	Fixed frequency	X	50W
11	A → C	22''	27''	5s	Frequency hopping	Y	50W
12	J → D	25''	29''	4s	Fixed frequency	X	50W
13	B → F	28''	34''	6s	Frequency hopping	Y	50W
14	E → H	30''	35''	5s	Frequency hopping	Y	50W
15	J → D	32''	38''	6s	Fixed frequency	X	50W
16	A → F	35''	39''	4s	Frequency hopping	Y	50W
17	H → E	36''	38''	2s	Frequency hopping	Y	50W
18	H → E	39''	41''	2s	Frequency hopping	Y	50W
19	H → G	42''	46''	4s	Frequency hopping	Y	50W
20	B → A	43''	48''	5s	Frequency hopping	Y	50W
21	F → G	47''	50''	3s	Frequency hopping	Y	50W
22	G → I	51''	56''	5s	Frequency hopping	Y	50W

simulation experiment, information is transmitted from the upper node to the terminal step by step, and the terminal generates feedback. In the communication of each pair of stations, we reduced the communication time to accom-modate communication among more stations in the monitoring data. The communication relationship and communication sequence of communication individuals are shown in Table 1. Fig. 8 shows the sequence of radio communication in Table 1 in time, where different colors represent the communication between different radio stations, "1" represents the sending state of radio station, and "2" represents the receiving state of radio station.

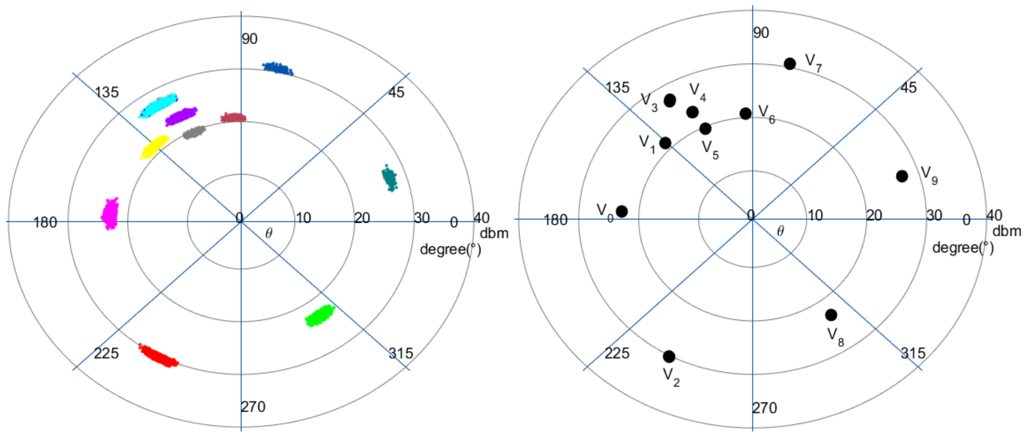
Based on the geographic location of communication nodes (stations) and the simulated communication between them, the actual communication network model is shown in Fig. 9.

TABLE 2. The monitoring data characteristics of electromagnetic spectrum.

X	Frequency	Power	Time	angle
X_1	49400KHz	29.56dbm	10''000	137.2°
X_2	53300KHz	28.74dbm	10''016	136.9°
…	…	…	…	…
X_i	73200KHz	20.47dbm	99''000	209.7°
…	…	…	…	…

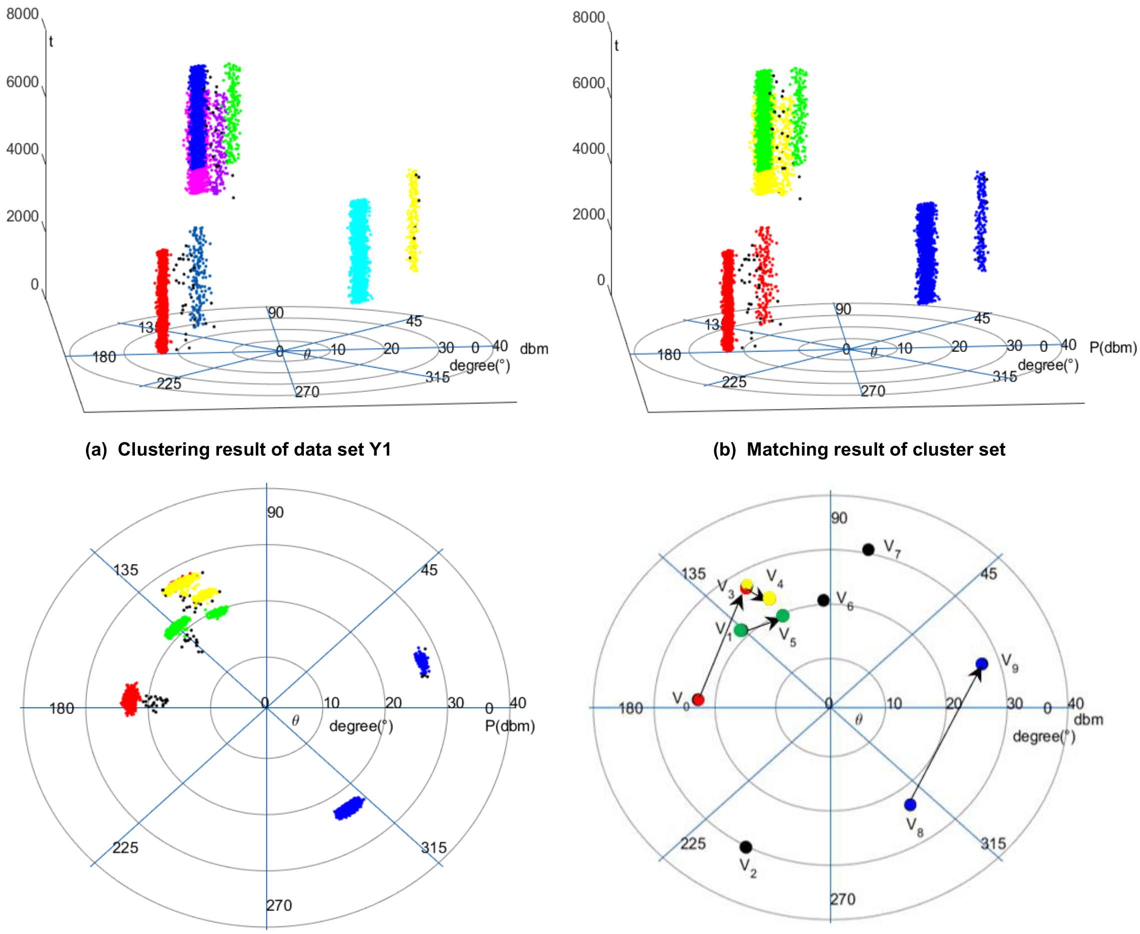
B. SPECTRUM MONITORING DATA DESCRIPTION

After preprocessing the spectrum monitoring data, we obtain the data set X for communication behavior research, which contains the following characteristics: signal center frequen-cy point, signal power, signal monitoring time and



(a) (b)

FIGURE 10. Distribution of relative positions of sources in the polar coordinate system.



(a) Clustering result of data set Y1

(b) Matching result of cluster set

(c) The distribution of cluster sets in polar coordinates

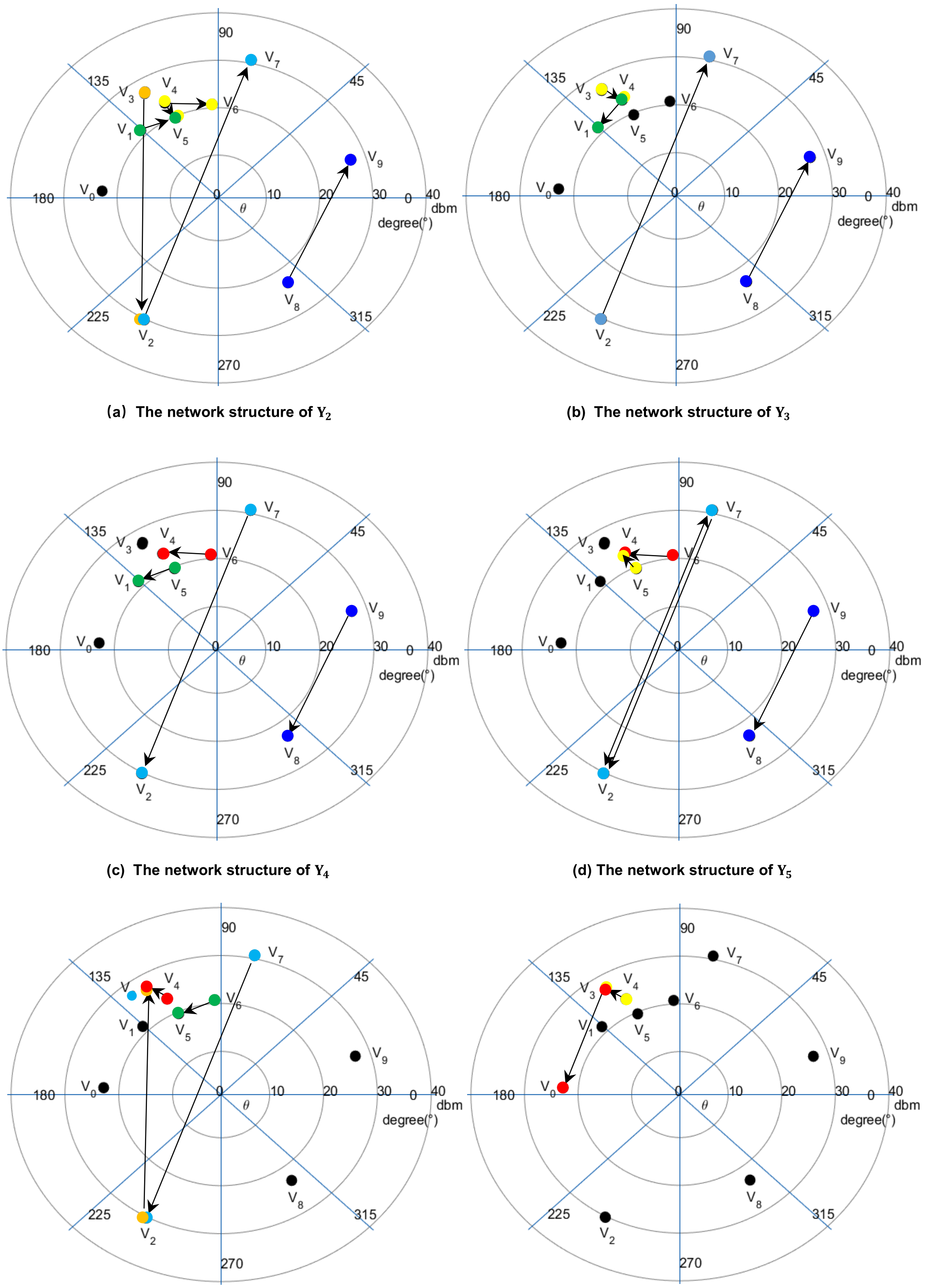
(d) Network structure

FIGURE 11. The process of mining the communication network structure of $\mathbf{Y_1}$.

signal direction. Table 2 shows the format of the spectrum monitoring data set X.

C. ANALYSIS OF EXPERIMENT RESULT

For the data set $Y = \{y_1, y_2, \cdots, y_j, \cdots, y_n\}^T$, where $y_j = \{\theta_j, d_j, t_j\}$, we set $t_{interval} = 8$ s and divide the spectral data Y of 56s into 7 segments, namely $Y = \{Y_1, Y_2, \cdots, Y_i, \cdots\}$, $i = 1, 2, 3, \cdots, 7$. According to Algorithm 2, this paper mine the structure of communication network from spectrum monitoring data. Fig. 10(a) shows the projection of data set Y on polar coordinates, that is, the relative position of source is marked by signal power and direction in polar coordinates.



(a) The network structure of Y_2

(b) The network structure of Y_3

(c) The network structure of Y_4

(d) The network structure of Y_5

(e) The network structure of Y_6

(f) The network structure of Y_7

FIGURE 12. The segmented netork structure corresponding to Y_i. The dots marked with color indicate communication stations, and the straight lines with arrow indicate the communication direction. Nodes and connections form a communication network snapshot of Y_i. For example, in (a), the communication nodes are v1, v2, v3, v4, v5, v6, v7, v8, v9, and the communication paths are v1-v5, v4-v5, v5-v6, v3-v2-v7, V8-v9.

Cluster sets of different colors represent the distribution of signals generated by different sources. Fig. 10(b) shows the relative positions of the centroid neighborhoods of the respective cluster sets of Fig. 10(a), which are used as nodes of the network.

We take Y_1 to demonstrate the communication network structure mining. Fig. 11 shows the mining process of communication relationship and communication network structure of Y_1. Fig. 11 (a) shows the clustering results of the data set Y_1 in the column coordinates composed of signal power, signal direction and signal monitoring time, where different colors represent different clustering sets, that is, clustering sets of different colors correspond to spectrum signal sets generated by different radio stations in the communication process. Since the spectrum signals generated by radio stations with a communication relation-ship are very similar in time, we match the clustering setaccording to the time range of the cluster sets to determine the communication relationship (corresponding to the communication relationship between radio stations). Fig. 11(b) shows the matching result of the cluster set in Fig. 11(a), and the cluster sets with a communication relation are labeled with the same color. Fig. 11 (a) (b) show the discovery process of the communication relation of Y_1. In the section IV, we described in detail the communication relationship method, in which the data clustering method is the improved OPTICS algorithm.

To build a communication network structure, we need identify nodes and edges. We projected the data in Fig. 11 (b) into the polar coordinate system to obtain the distribution of spectral data in the polar coordinate system, as shown in Fig.11 (c).Then we calculated each centroid of the cluster set of Fig. 11 (c) and matched them with the coordinates of the nodes in Fig. 10(b).The matched nodes are labeled with the same color in Fig. 11 (d). Finally, we connected nodes of the same color according to the communication relation, so as to form the construction network topology structure of Y_1, as shown in Fig. 11 (d). The arrow indicates the communication direction.

Fig. 12 shows the different communication network structure corresponding to subsets of data set Y. Finally, we merge all the network structure snapshots to form the network structure of Y, as shown in Fig. 13.

D. NETWORK STRUCTURE ANALYSIS

As shown in Fig. 13, in the monitoring range, v_8 and v_9 are stations that communicate independently, and do not communicate with other nodes, which can be regarded as network F; and other nodes constitute network G. From the comparison of the various figures in Fig. 12, it can be found that the network G is mainly divided into three paths:

Path 1: $v_0 \leftrightarrow v_3 \leftrightarrow v_4 \leftrightarrow v_5 \leftrightarrow v_1$

Path 2: $v_0 \leftrightarrow v_3 \leftrightarrow v_4 \leftrightarrow v_6 \rightarrow v_5$

Path 3: $v_0 \leftrightarrow v_3 \leftrightarrow v_2 \leftrightarrow v_7$

In network G, information transfer starts from node v_0, and then passes through v_3 to other nodes. As nodes v_1, v_4, v_5, v_6 communicate closely, they constitute the sub-network G_1,

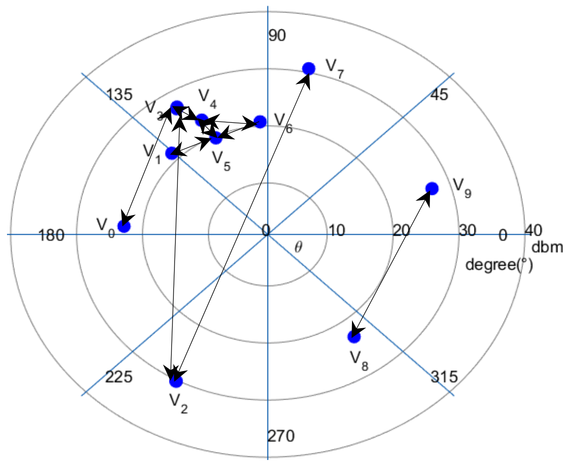


FIGURE 13. **The communication network structure.**

where v_4 is the core node of sub-network G_1. v_2 and v_7 constitute subnetwork G_2. v_3 is the key node connected with two subnets.v_0 is the initial node of communication, and v_1 and v_7 are the terminal nodes of communication. From the statistical analysis of the nodes, $d(v_3) = 6, d(v_4) = 6, d(v_5) = 5, d(v_2) = 4$, these nodes can be regarded as important nodes for network communication.

In summary, the level of the network G can be divided into 4 layers. v_0 is the beginning of communication, which can be regarded as the highest node in communication; v_3 is the connection point of sub-networks G_1 and G_2, which is the intermediate node of information exchange, as the second-level node; v_4 and v_2 are regarded as the core nodes of the sub-network, which are used to organize the communication inside the sub-network and serve as the third-level nodes. As terminal nodes, v_1, v_5, v_6 and v_7 are the fourth-level nodes of the network.

VII. CONCLUSION

As a medium carrying communication information, the physical characteristics of the spectrum signal itself and the statistical laws of certain features also potentially reflect the communication behavior of the communication individual and the intelligence information related to the communication content. Therefore, the in-depth research and analysis of massive spectrum is of great significance. In order to avoid the difficulty and cost of cracking the signal content, this paper mines and analyzes the communication behavior of the communication individual from the physical charac-teristics of the spectrum monitoring signal and the statistical laws of these characteristics.

This paper first discusses the characteristics of spectrum monitoring data, and then obtains the frequency, signal power, signal bandwidth, signal monitoring time, signal direction and other characteristics of the signal from the spectrum monitoring data to uniquely identify the spectrum signal. Then we study the distribution characteristics and statistical laws of the spectrum signals in the cylindrical coordinate system composed of signal power, signal monitoring time and signal power. Through the spectrum monitoring signal mining method proposed in this paper, the communication relationship and communication network between communication individuals are extracted from the spectrum monitoring data. The experimental results show that the method has good adaptability to the massive spectrum monitoring signals, and can mine the communi-cation relationship between the source nodes from the spectrum monitoring data, and infer the communication network structure. Finally, we realize the research of individual communication behavior through the statistical analysis of network structure and node communication quantity.

The research of this paper realizes the mining of the communication relationship between the sources and the communication network structure from the spectrum monitoring data. Through the analysis of network connectivity, communication direction, communication times, communication order and other characteristics, we obtain the hierarchical structure of the network, the hierarchical position of different communication individuals in the network, and realize the analysis of communication behavior of communication individuals in the monitoring area. The research method proposed in this paper provides a new method for the analysis of spectrum monitoring data, and it can realize the acquisition of hidden intelligence in military communication, investigation and other related fields, and has practical application value.

REFERENCES

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *Int. J. Wireless Mobile Comput.*, vol. 8, no. 2, pp. 153–165, Apr. 2015.

[2] T. Manku and O. Kravets, "Wireless spectrum monitoring and analysis," U.S. Patent 9 143 968, Nov. 22, 2015.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[4] C. Liu, X. Wu, C. Yao, L. Zhu, Y. Zhou, and H. Zhang, "Discovery and research of communication relation based on communication rules of ultrashort wave radio station," in *Proc. IEEE ICBDA*, Suzhou, China, Mar. 2019, pp. 112–117, doi: 10.1109/ICBDA.2019.8713220.

[5] C. Liu, X. Wu, L. Zhu, C. Yao, L. Yu, L. Wang, W. Tong, and T. Pan, "The communication relationship discovery based on the spectrum monitoring data by improved DBSCAN," *IEEE Access*, vol. 7, pp. 121793–121804, Aug. 2019, doi: 10.1109/ACCESS.2019.2938296.

[6] R. Fantacci, "Performance evaluation of some efficient stop-and-wait techniques," *IEEE Trans. Commun.*, vol. 40, no. 11, pp. 1665–1669, Nov. 1992, doi: 10.1109/26.179927.

[7] X. Zhang and S. Zhang, "Parameter estimation of multiple frequency-hopping radar signals," in *Proc. ISCID*, Hangzhou, China, Dec. 2017, pp. 99–102, doi: 10.1109/ISCID.2017.124.

[8] Y. Qi, L.-X. Lu, and K. Zhang, "Frequency-hopping period estimation based on binary-sum in frequency domain," in *Proc. WCSN*, Wuhan, China, Dec. 2014, pp. 91–94, doi: 10.1109/WCSN.2014.25.

[9] X. Zhang, X. Hu, and X. Dong, "A joint algorithm of parameters estimation for frequency-hopping signal based on sparse recovery," in *Proc. WCSP*, Nanjing, China, Oct. 2017, pp. 1–5.

[10] G. Cui, J. Liu, H. Li, and B. Himed, "Signal detection with noisy reference for passive sensing," *Signal Process.*, vol. 108, pp. 389–399, Mar. 2015.

[11] C.-I. Chang, "Multiparameter receiver operating characteristic analysis for signal detection and classification," *IEEE Sensors J.*, vol. 10, no. 3, pp. 423–442, Mar. 2010, doi: 10.1109/JSEN.2009.2038120.

[12] R. Sreeraj, W. Meert, V. Lenders, and S. Pollin, "SAIFE: Unsupervised wireless spectrum anomaly detection with interpretable features," in *Proc. IEEE DySPAN*, Seoul, South Korea, Oct. 2018, pp. 1–9.

[13] L. Zhang, G. Ding, and Q. Wu, "Detecting abnormal power emission for orderly spectrum usage," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1989–1992, Feb. 2019.

[14] H. Alipour, Y. B. Al-Nashif, P. Satam, and S. Hariri, "Wireless anomaly detection based on IEEE 802.11 behavior analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2158–2170, Oct. 2015.

[15] M. Ahmed, "Abnormal behavior detection of jamming signal in the spectrum using a combination of compressive sampling and intelligent bivariate k-means clustering technique in wideband cognitive radio systems," in *Proc. IEEE ICEE*, Boumerdes, Algeria, Dec. 2015, pp. 1–4.

[16] A. Ali and W. Hamouda, "Advances on spectrum sensing for cognitive radio networks: Theory and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1277–1304, 2nd Quart., 2016.

[17] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: A survey," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 74–81, Apr. 2013.

[18] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[19] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," *EURASIP J. Adv. Signal Process*, vol. 2010, Jan. 2010, Art. no. 381465.

[20] M. T. Masonta, M. Mzyece, and N. Ntlatlapa, "Spectrum decision in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1088–1107, 3rd Quart., 2013.

[21] W.-Y. Lee and I. F. Akyildiz, "A spectrum decision framework for cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 161–174, Feb. 2011, doi: 10.1109/TMC.2010.147.

[22] Y. Zeng and R. Zhang, "Wireless information surveillance via proactive eavesdropping with spoofing relay," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1449–1461, Dec. 2016.

[23] Y. Han, L. Duan, and R. Zhang, "Jamming-assisted eavesdropping over parallel fading channels," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2486–2499, Sep. 2019, doi: 10.1109/TIFS.2019.2901821.

[24] J. Xu, L. Duan, and R. Zhang, "Proactive eavesdropping via cognitive jamming in fading channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2790–2806, May 2017, doi: 10.1109/TWC.2017.2666138.

[25] T. Pan, X. Wu, C. Yao, Y. Zhou, and X. Lu, "Communication behavior structure mining based on electromagnetic spectrum analysis," in *Proc. IEEE TAIC*, Chongqing, China, May 2019, pp. 1611–1616, doi: 10.1109/ITAIC.2019.8785571.

[26] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for dis-covering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc KDD*, Aug. 1996, pp. 226–231.

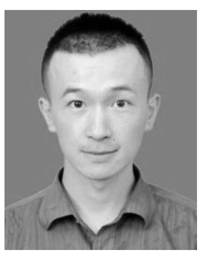
[27] G. H. Shah, "An improved DBSCAN, a density based clustering algorithm with parameter selection for high dimensional data sets," in *Proc. NUiCONE*, Ahmedabad, India, Dec. 2012, pp. 1–6.

[28] A. Smiti and Z. Eloudi, "Soft DBSCAN: Improving DBSCAN clustering method using fuzzy set theory," in *Proc. IEEE HSI*, Sopot, Poland, Jun. 2013, pp. 380–385.

[29] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," *ACM SIGMOD. Rec.*, vol. 28, no. 2, pp. 49–60, Jun. 1999.

[30] M. Daszykowski, B. Walczak, and D. L. Massart, "Looking for natural patterns in data: Part 1. Density-based approach," *Chemometrics Intell. Lab. Syst.*, vol. 56, no. 2, pp. 83–92, Feb. 2001.

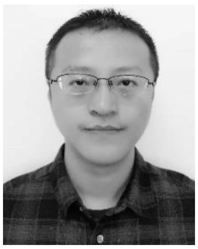
[31] M. Franceschet, "PageRank: Standing on the shoulders of giants," *Commun. ACM*, vol. 54, no. 6, pp. 1–9, Feb. 2010.



CHANGKUN LIU received the B.S. degree in information and computing science from Shandong University, in 2017. He is currently a Graduate Student with the Army Engineering University of PLA. His research interests include data mining and machine learning.

XINRONG WU received the M.S. degree from the Communication and Engineering Institute, in 1996. She is currently with the Institute of Communications Engineering, Army Engineering University of PLA. She has authored more than 25 scientific articles. Her current research interests include communication networks and network security.

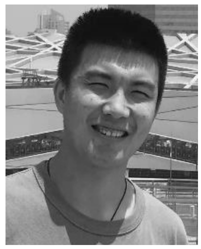
LEI ZHU received the Ph.D. degree from the College of Communications Engineering, PLA University of Science and Technology, China, in 2002. He is currently a Professor of system engineering. He has authored more than 35 scientific articles. His research interests include network planning and system simulation.



CHANGHUA YAO received the B.S. degree in automation from Zhejiang University, in 2005, and the Ph.D. degree in communications and information systems from the PLA University of Science and Technology, in 2016. He has been a Postdoctoral Researcher with the Army Engineering University of PLA, since 2017. His research interests include UAV communication, wireless networks, network security, opportunistic spectrum access, distributed optimization, data analysis, artificial intelligence, and machine learning.

LU YU received the Ph.D. degree from Southeast University, in 2007. She is currently with the Institute of Communications Engineering, Army Engineering University of PLA. She has authored more than 25 scientific articles. Her current research interests include machine learning, image understanding, and pattern recognition.

LEI WANG received the Ph.D. degree in military operational research from the PLA University of Science and Technology, China, in 2014. He holds a Postdoctoral position in communication and information system and an Engineer of optimization and system engineering with the College of Communications Engineering, Army Engineering University of PLA, China. He has authored more than 25 scientific articles. His research interests include knowledge engineering, data mining, artificial intelligence, and network planning.



HAOREN FAN received the B.S. degree in computer science and technology from Shihezi University, in 2017. He is currently a Graduate Student with the Army Engineering University of PLA. His research interest includes deep reinforcement learning.

TING PAN received the B.S. degree in information and computing science from the Wuhan University of Science and Technology, in 2017. She is currently a Graduate Student with the Army Engineering University of PLA. Her research interests include data mining and machine learning.