RELIABLE SPECTRUM HOLE DETECTION IN SPECTRUM-HETEROGENOUS MOBILE COGNITIVE RADIO NETWORKS VIA SEQUENTIAL BAYESIAN NON-PARAMETRIC CLUSTERING

by

ALIREZA ZAEEMZADEH
B.Sc. University of Tehran, 2014

A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science
in the Department of Electrical and Computer Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2017

Major Professor: Nazanin Rahnavard

ABSTRACT

In this work, the problem of detecting radio spectrum opportunities in spectrum-heterogeneous cognitive radio networks is addressed. Spectrum opportunities are the frequency channels that are underutilized by the primary licensed users. Thus, by enabling the unlicensed users to detect and utilize them, we can improve the efficiency, reliability, and the flexibility of the radio spectrum usage. The main objective of this work is to discover the spectrum opportunities in time, space, and frequency domains, by proposing a low-cost and practical framework.

Spectrum-heterogeneous networks are the networks in which different sensors experience different spectrum opportunities. Thus, the sensing data from sensors cannot be combined to reach consensus and to detect the spectrum opportunities. Moreover, unreliable data, caused by noise or malicious attacks, will deteriorate the performance of the decision-making process. The problem becomes even more challenging when the locations of the sensors are unknown.

In this work, a probabilistic model is proposed to cluster the sensors based on their readings, not requiring any knowledge of location of the sensors. The complexity of the model, which is the number of clusters, is automatically inferred from the sensing data. The processing node, also referred to as the base station or the fusion center, infers the probability distributions of cluster memberships, channel availabilities, and devices' reliability in an online manner. After receiving each chunk of sensing data, the probability distributions are updated, without requiring to repeat the computations on previous sensing data. All the update rules are derived mathematically, by employing Bayesian data analysis techniques and variational inference.

Furthermore, the inferred probability distributions are employed to assign unique spectrum opportunities to each of the sensors. To avoid interference among the sensors, physically adjacent devices should not utilize the same channels. However, since the location of the devices is not

known, cluster membership information is used as a measure of adjacency. This is based on the assumption that the measurements of the devices are spatially correlated. Thus, adjacent devices, which experience similar spectrum opportunities, belong to the same cluster. Then, the problem is mapped into a energy minimization problem and solved via graph cuts. The goal of the proposed graph-theory-based method is to assign each device an available channel, while avoiding interference among neighboring devices. The numerical simulations illustrates the effectiveness of the proposed methods, compared to the existing frameworks.

This humble work is dedicated to my

Father and Mother,

whose sacrificial love and encouragements made my accomplishments possible, and to my amazing sisters,

Rokhsareh and Afsaneh.

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Dr Nazanin Rahnavard of the Department of Electrical Engineering at University of Central Florida. She was always keen to find out how I am proceeding and always encouraged me to keep moving forward. She consistently gave me insightful comments, while allowing this thesis to be my own work.

I also want to acknowledge my friends and colleagues in Communications and Wireless Networking Laboratory, Dr. Behzad Shahrasbi and Mr. Mohsen Joneidi. Their valuable help and comments helped me a lot during this process. It was a fantastic opportunity to work with such amazing and knowledgeable people.

Finally, last but by no means least, I am grateful to my parents, who have supported me unconditionally throughout my life. I owe it all to them. I am also grateful to my loving and sweet siblings, Rokhsareh and Afsaneh, who have provided me through moral and emotional support in my life. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Radio spectrum is known to be one of the most valuable resources of the modern era [1]. However, as indicated by many reports, the licensed frequency channels often experience low utilization. Thus, motivated by such striking observations, the 90-year-old spectrum policies are being reexamined. The cognitive radio (CR) paradigm is seen as the remedy to this low usage of the radio spectrum in licensed band and the spectrum scarcity in unlicensed bands [1–3]. The main idea is to allow the unlicensed users to utilize the spectrum, provided that they will not interfere with the operation of licensed users [3]. Thus, it is the responsibility of the unlicensed users to sense their surrounding environment and intelligently adapt to it.

The main objective of CR devices is to find the time intervals and areas in which the licensed users are not active, which is referred to as *spectrum opportunity detection* in this work. *Spectrum heterogeneity* among devices is due to the fact that devices in different locations might experience different spectrum opportunities. An efficient CR framework should be able to find and utilize spectrum opportunities both in temporal and spatial domains.

However, the cost of integrating intelligence into CR devices is often neglected by researchers [4]. A CR framework is not practical unless it can be implemented with inexpensive devices and its overall cost of dynamic spectrum usage is realistic. In this work, to develop a low-cost CR framework, it is assumed that:

- devices do not have processing units and cannot perform computationally intensive tasks
- devices are not capable of extracting complex features of environment,
- devices are not equipped with location-finding technologies,

- devices might report *faulty measurements*, either unintentionally or maliciously,
- each device is able to sense at most one frequency channel at each sensing time,
- and there exist missing entries among the sensing data, either because of network and device failures or because of energy and bandwidth constraints.

Advancements in microelectronics have paved the way for deploying a large number of sensors in data collection tasks. However, since we are collecting measurements from a large number of low-cost sensors, there might be measurements with low accuracy [5, 6]. In sensor networks, some devices may report unreliable data to the processing node unintentionally or maliciously. This may occur because of low battery, network failures, physical obstruction of the scene, and malicious attacks. Thus, the processing node should not simply aggregate measurements from all sensors. It is more efficient to disregard the outlier measurements and draw conclusions based on reliable measurements.

In this work, a spectrum opportunity detection framework is proposed to detect opportunities both in *time* and *space*, using measurements collected from a network of low-cost devices. Here, since the sensing devices do not have processing power, they have to be capable of communicating with a central processing node, which is referred to as the *base station* (BS). The BS has to aggregate data from different sensors and draw conclusions on availability of different frequency bands for each device. The problem becomes more challenging when the location of the devices is unknown, spectrum opportunities vary over time and space, and SUs might report faulty measurements.

The main contribution of this work is addressing the spectrum opportunity detection problem in a spectrum-heterogeneous network of low-cost and possibly unreliable devices. The novelty of the proposed framework is two-fold. First, a novel Bayesian data analysis technique is proposed that allows us to cluster the devices solely based on their sensing data. Each cluster contains devices

that are experiencing similar spectrum opportunities. Simultaneously, the spectrum availability and reliability of each device are inferred. It is worthwhile to note that the BS is not provided with any prior information on the number of clusters, number of licensed users, and location of licensed and unlicensed devices. Second, we propose a graph-theory-based method to exploit the inferred information and detect the spectrum opportunities for each device.

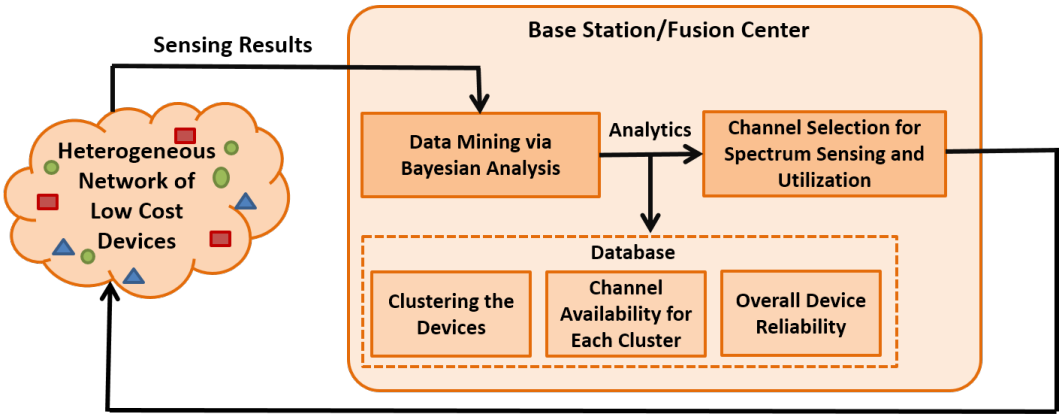

Figure 1.1: Overall Architecture of the Proposed Framework

After receiving each set of new measurements, the probability distributions of the desired variables are updated using the closed form update rules. The update rules are derived mathematically using variational inference techniques and are shown to be computationally light-weight. Figure 1.1 illustrates the overall architecture of the proposed framework.

1.1 Literature Review

Most of the prior work focuses on finding spectrum opportunities in either time [7–9] or space [10–12]. Recently, there has been an effort to tackle the problem of spatio-temporal spectrum opportunity detection. However, most studies are only capable to work either in the presence of only a single PU or need to be aware of the locations [13–20]. For instance, authors in [20] exploit

interpolation techniques to construct a spatio-temporal spectrum maps from the measurements of a network of devices with known locations. The problem of discovering the white spaces is discussed in [12]. In this approach, the fusion center aggregates the data from randomly distributed sensors with binary measurements and known locations to find the spectrum opportunities within a geographical area. Intelligent cooperative spectrum sensing is proposed in [21]. This method uses Bayesian inference to find the spectrum holes by considering the spatial correlation among the observations. However, the main focus of the work is on finding the spectrum opportunities, not allocating the resources. Moreover, temporal correlation of the observations is not taken into account.

Most of the methods to facilitate the database-enabled spectrum awareness are based on constructing a spatio-temporal spectrum map, also referred to as radio environment map (REM), [20, 22–31]. REM contains information about the channel status at different locations, times, and frequencies. REM construction methods can be categorized into *direct* and *indirect* methods. The direct methods interpolate the sensor readings to estimate the signal strength at different locations. On the other hand, the indirect methods localize the transmitters, then by exploiting the propagation models, the REM is reconstructed [32, 33]. Construction of REM requires location information about all or a large subset of sensing devices. Furthermore, most of the techniques are computationally intensive.

Recently, in [34], a cluster-based spectrum sensing is devised to discover the spectrum opportunities in time and space without knowing the location of the devices. However, the update rules to predict the channel availability are mostly heuristic and the reliability of the measurements is not taken into the account.

In the proposed problem, since we are collecting information from tiny and low-cost sensors, we *do not have access to the location information* of any of the devices. Hence, we cannot construct a

map containing the channel status information. Here, we propose to employ Bayesian analysis to extract information from the received measurements. The proposed framework neither requires any prior knowledge about the location of the devices nor assumes any propagation model. In addition, it gives us the opportunity to update the decisions in an online fashion. Since the information of the past measurements is integrated into the past decisions, we do not need to maintain a long history of measurements.

CHAPTER 2: BACKGROUND

In this chapter, we present the necessary background that facilitates understanding of this thesis. First, we briefly introduce the concept of cognitive radio networking. Then, fundamentals of Bayesian data analysis and multi-label graph cuts are discussed.

2.1 Cognitive Radio Networks

Recent advances in wireless communications and microelectronic devices are leading the trend of research toward cognitive radios (CRs). The main feature of CRs is the opportunistic usage of spectrum. CR systems try to improve the spectrum efficiency by using the spectrum holes in frequency, time, and space domains [2, 35]. This means that unlicensed users are allowed to utilize the spectrum, provided that their transmissions do not interfere with the communication of primary licensed users [3]. The fundamental components of CR systems that allow them to avoid interference are spectrum sensing and resource allocation.

The introduction of unlicensed users inevitably deteriorates the quality of service of the primary users. Thus, the objective is to minimize such adverse impacts. To this end, CR frameworks should have the following functionalities: spectrum awareness, analysis and decision, and spectrum exploitation [36]. An intelligent CR should be able to analyze the spectrum data to make decisions. Furthermore, the CR exploits the spectrum opportunities by tuning its operating frequency channel, transmission power, modulation, coding, etc.

Spectrum awareness techniques can either be blind or non-blind. The non-blind approaches need specific information about the licensed users such as the location of the transmitters, transmission power, or modulation. Some of the database enabled techniques as well as beacon transmission-

based methods are examples of non-blind spectrum awareness approaches. Moreover, based on the sensing capabilities of the devices, the spectrum awareness methods can be broadly categorized into:

- **Database Assisted**: In this scheme, the parameters of the primary network such as operating frequency, location, coverage areas, and time are stored in a database. The devices can query the database and obtain information about the radio environment. Thus, the devices are not required to have sensing capability.

- **Spectrum Sensing**: The devices are assumed to be able to detect the presence of the primary signal, by exploiting some signal processing techniques such as energy detection, matched filter detection, eigenvalue-based sensing, etc. Cooperative spectrum sensing, in which a network of devices collaborate to improve the performance, is seen as a promising approach [36, 37].

- **Environmental Parameter Estimation**: In this approach, the devices are able to estimate radio environment parameters such as number and location of licensed transmitters. To achieve this additional level of awareness, the devices need to be equipped with more advance sensing technologies.

- **Waveform Parameter Estimation**: The devices can estimate parameters of the primary signals such modulation and coding schemes.

There are many approaches that combine different methods to achieve higher levels of cognition. For instance, the approaches that use spectrum sensing capability of the devices to maintain and update the database [25, 27, 32, 38, 39] The proposed approach falls into the category of blind centralized cooperative spectrum sensing. This means that the spectrum measurements are collected

and analyzed by a central processing node, without requiring any information from the network of the licensed users.

On the other hand. *spectrum exploitation* methods can be categorized into three different families:

- **Interweave**: For interweave communication, the unlicensed network must find the spectrum holes in temporal, spatial, and spectral domains. Spectrum holes are the frequency channels, areas, and time intervals that the signal from the licensed users is absent. Hence, there is no constraint on the transmission power of the secondary network and no interference occurs in the ideal case.

- **Underlay**: In this paradigm, coexistence of the licensed and unlicensed users are allowed, only if the interference caused by the unlicensed network can be tolerated by the licensed users. The unlicensed users that are transmitting on a shared frequency channel must guarantee that their interference will not exceed a certain threshold.

- **Overlay**: In this scheme, to compensate the interference, the secondary network transmits the signal of the licensed user along with its own signal. In this approach, the unlicensed users should be aware of the waveform parameters of the licensed users, which makes it difficult to implement.

Again, there are approaches that combine interweave and underlay methods to increase the efficiency of the spectrum utilization [40, 41]. Interested readers may refer to [36] for a detailed survey on different CR techniques.

2.2 Bayesian Data Analysis

In this work, Bayesian data analysis is employed to make inferences from observed data. Flexibility, generality, and simplicity of the Bayesian framework make it a useful tool to tackle complex problems. The most desirable feature of Bayesian inference is the quantification of uncertainty, which is realized by using probability distributions to explain the variables [42]. In most practical problems in physical sciences, we wish to find out the cause of a given effect. In other words, it is desired to determine the input of a system after observing the output. Specifically, in this work, the effects, or the observable quantities, are the sensing data measured by the devices. On the other hand, the cause of such effects, or the hidden variables, are the true occupancy status of the channels and the reliability of the sensing devices.

Our goal is to infer the cause given the effect. In Bayesian data analysis, we need to set up a probability model $\mathbb{P}\{\boldsymbol{D}, \boldsymbol{H}\}$, which is the joint probability distribution of all observable data $\boldsymbol{D}$ and hidden variables $\boldsymbol{H}$. The model should mirror the characteristics of the system, as well as the data collection process. In our model, we have considered the spatial correlation among the measurements and the sensing performance of each device on different frequency channels. In Chapter 3, we will discuss our proposed model in details. The joint probability distribution can be written as the product of two distributions:

$$\mathbb{P}\{\boldsymbol{D}, \boldsymbol{H}\} = \mathbb{P}\{\boldsymbol{D}|\boldsymbol{H}\}\mathbb{P}\{\boldsymbol{H}\}.$$

$\mathbb{P}\{\boldsymbol{D}|\boldsymbol{H}\}$ and $\mathbb{P}\{\boldsymbol{H}\}$ are often referred to as the *sampling distribution* and the *prior distribution*, respectively.

The next step in Bayesian data analysis is the inference, which is calculating an appropriate *posterior distribution*. This means that we need to find the distribution of the hidden variables,

which we are ultimately interested in, given the observations. Mathematically, we aim to calculate $\mathbb{P}\{\boldsymbol{H}|\boldsymbol{D}\} \propto \mathbb{P}\{\boldsymbol{D}|\boldsymbol{H}\}\mathbb{P}\{\boldsymbol{H}\}$.

These simple expressions summarize the fundamentals of Bayesian inference. However, in most practical problems, the inference procedure is not very straight forward, due to complexity of the model. The number of the integrations can grow exponentially with the number of hidden and observable variables. This makes the impossible to solve practically. Thus, a method of approximation should be exploited to find the posterior distribution. In variational inference [43], the posterior distribution is approximated by a family of distributions for which the calculations are tractable. Thus, the objective is to find the best approximate of distribution in a computationally feasible manner. In Chapter 3, the model, i.e., $\mathbb{P}\{\boldsymbol{D}, \boldsymbol{H}\}$, is developed and in Chapter 4 and Chapter 5 necessary computations to find $\mathbb{P}\{\boldsymbol{H}|\boldsymbol{D}\}$ are performed.

2.3 Multi-Label Graph Cuts

After inferring the quantities of the interest using Bayesian data analysis, the extracted information is used to discover the spectrum opportunities. In Chapter 6, it is shown that the problem can be mapped to a multi-label graph cuts problem. In computer vision, Graph cuts are often used to assign labels to the pixels. Such labels oftentimes should satisfy a certain constraint, such as smoothness or piecewise smoothness. Such problems are naturally formulated as energy minimization [44]. The labels are assigned by minimizing the energy

$$E(\boldsymbol{h}) = E_{data}(\boldsymbol{h}) + E_{constraint}(\boldsymbol{h}),$$

where $\boldsymbol{h}$ is the set of assigned labels. In this expression, $E_{data}(\boldsymbol{h})$ quantifies how well the labeling matches the observed data. Moreover, $E_{constraint}(\boldsymbol{h})$ measures the deviation from the constraint,

e.g., smoothness. The energy function can typically be rewritten in the form of:

$$E(\boldsymbol{h}) = \sum_{n} D(h_n) + \sum_{n,n'} C(h_n, h_{n'}), \tag{2.1}$$

where n and n' index the pixel, or in general the *objects* that we are trying to label. Here, the first term represents $E_{data}(\boldsymbol{h})$ and $D(h_n)$ measures the disagreement between the observed data for n^{th} object and it label h_n. On the other hand, the second term is the penalty coming from the constraint by considering pairwise interactions between each object pair $\{n, n'\}$. For instance, in computer vision, the adjacent pixels should have similar labels, due to smoothness assumption.

It is shown that this problem is NP-hard, but can be approximated by using graph cuts [44–47]. A cut is a set of edges that removing them separates the graph into disjoint subgraphs. Furthermore, the cost of a cut is the sum of its edge weights. Thus, a minimum cut is the cut that has the lowest cost among all the possible cuts.

Now, to minimize the energy function introduced in 2.1, the problem is seen as a minimum cut problem. The structure of the graph is as follows: There exist a vertex for each object and for each label. The weight of the edge between an object and a label is determined by, not necessarily equal to, $D(h_n)$. Furthermore, the weight of the edges between object edges comes from $C(h_n, h_{n'})$. There are no edges between the labels. The objective is to partition the graph such that in each partition there exist a subset of object to objects and a label node.

In Chapter 6, it is shown that the problem of spectrum opportunity detection can be mapped into a graph cut problem. In our problem, the objects are the devices and the labels are the channels. The energy, i.e., the cost function that we are trying to minimize, comes from the information extracted by the Bayesian data analysis.

CHAPTER 3: MODELING THE SYSTEM

In this chapter, the system model and the overall architecture of the network is discussed. Then, we will talk about how the observations can be modeled using probability distributions. It is assumed that the network consists of N stationary SUs trying to find the spectrum opportunities among M frequency channels. The SUs are able to sense one channel at each sensing time and report it to the BS. Each SU measures the energy in the sensing channel and compares it with a threshold, coming from a standard [1]. It is known that when there is no prior information concerning the PUs' signal, energy detector achieves the best performance [48]. If the measured energy in m^{th} channel by n^{th} SU is greater than the threshold, the SU will report the channel as unavailable ($y_{nm}(t) = 0$), otherwise it will report the channel as available ($y_{nm}(t) = 1$). However, due to shadowing, device failure, or malicious attacks, a device might report faulty measurements on a subset of channels.

We also assume that there exists a common control channel (CCC) between the SUs and the BS. The BS assigns the sensing channels using the CCC and synchronizes the SUs to operate on a frame-by-frame structure [49]. At each frame, there exist a quite interval when the SUs stop their transmission, listen to the assigned channel, and send their measurements to the BS via CCC. Then, by aggregating the sensing data from all the devices, the BS infers the availability of channels and discovers the spectrum opportunities. The spectrum opportunity detection procedure is discussed in detail in Chapter 6.

It is also assumed that an unknown number of licensed PUs are utilizing the bands. Each PU is transmitting on a subset of channels and the activity of the PUs is assumed to be independent of each other. To model the dynamic behavior of the PUs, a two state Markov chain model is exploited [50]. The PUs alternate between active and inactive states. If the PU is active, it utilizes an unknown subset of channels.

3.1 Generative Model

Here, the generative model is presented, describing the observations using the hidden variables. A generative model specifies the probability distribution of all the variables, including the observed and hidden variables. In this work, the observed variables are the sensing data collected from the sensors. The hidden variables are parameters that are desired to be inferred from the observations, such as availability of the channels, reliability of devices, and the clusters. The goal of the inference algorithm is to infer the hidden variables given the observations. In Chapter 4, the details of inference and evolution of distributions over time will be discussed.

Specifically, in our model, the following hidden variables are defined:

1. Channel-specific reliability $u_{nm} \in \{0, 1\}$, which is either 0 or 1 and describes the reliability of measurements of n^{th} SU on m^{th} channel. A device might have different performances on different channels due to frequency selective fading or device failure. If $u_{nm} = 1$ ($u_{nm} = 0$), the device n is reliable (unreliable) on channel m.

2. Device reliability $r_n \in [0, 1]$, denoting the overall trustworthiness of the device. For small values of r_n, the device is more prone to reporting faulty data. A generally reliable device will report trustworthy measurements on most of the channels.

3. Cluster membership $g_n \in \{1, 2, 3, \ldots\}$, indicating the cluster that the n^{th} SU belongs to. Each cluster contains the devices experiencing similar spectrum opportunities. This prior captures the *spatial correlation* among the SUs. We will use l to index the clusters.

4. Channel availability for each cluster $c_{lm} \in [0, 1]$, describing the probability that channel m is available for SUs in cluster l.

Moreover, the observed variable is y_{nm}. The SU n will report $y_{nm} = 1$, if it senses the channel

m as available and $y_{nm} = 0$ otherwise. In this model, clustering the SUs helps us to model correlation among the sensing outcomes. It is assumed that there exist clusters of SUs that agree on the availability of different frequency bands. Number of the clusters is assumed to be unknown a priori and will be inferred from the data.

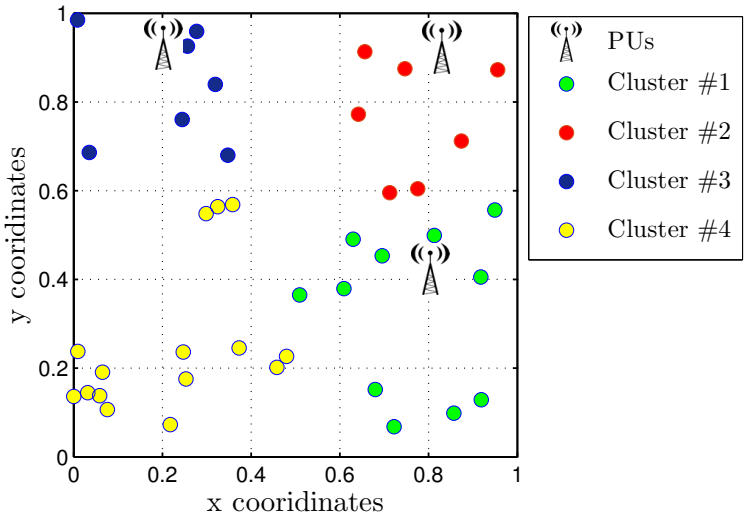

Figure 3.1: Outcome of the clustering algorithm in the presence of 3 PUs.

To illustrate the idea, Figure 3.1 shows the clustered devices in a simple scenario. It is easy to notice that the devices with same cluster membership are more likely to be close to each other. In Chapter 6, this knowledge will be used to avoid assigning same spectrum opportunities to neighboring SUs, to reduce the chance of interference. Also, channel-specific reliability and device reliability model the faulty measurements and devices. Without them, all the observations would be assumed to be reliable and the faulty measurements would easily reduce the accuracy of the inference algorithm.

Figure 3.2 illustrates the graphical representation of the generative model, describing the dependencies among the hidden and observed variables. Each white circle represents a hidden variable in our model and the gray circle is the observed variable, which is y_{nm}. For simplicity, each plate

groups the variables that repeat together in the model and the sequence inside the plate indicates the number of repetitions. This means that, for example, instead of having N nodes in the graph representing the devices reliability, one for each device, there exist only one node in the graph denoted by r_n, which is repeated N times.

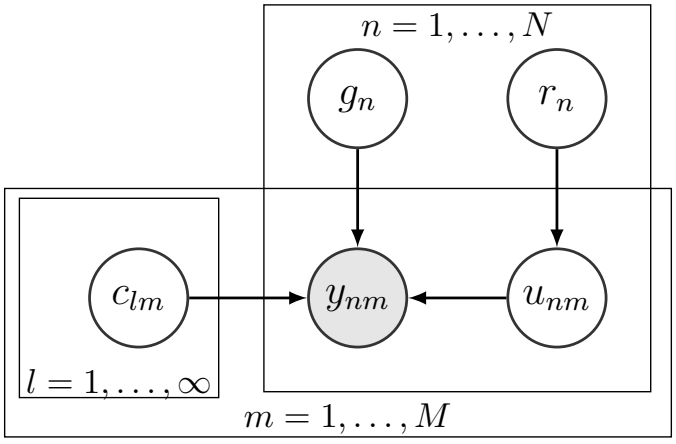

Figure 3.2: Graphical representation of the generative model.

For now, since the exact number of clusters is not known, it is assumed that we have an infinite number of candidate clusters. However, an upper bound will be set for number of clusters in Chapter 4.1.

The arrows in Figure 3.2 indicate the dependency among the variables. The observations made by each device on a specific channel depends on its cluster membership, channel availability for the cluster, and its reliability on the channel. Also, according to the model, the observations are independent of the device reliability, given the channel-specific reliability. This is intuitive since if we know the reliability of the device on a specific channel, we do not need the reliability of the device on the other channels or the overall reliability to describe the observations on that channel.

The proposed model can be formulated as follows:

$$
\begin{aligned}
g_n &\sim \text{Discrete}(\boldsymbol{\pi}_n) && && n = 1, \ldots, N \\
r_n &\sim \text{Beta}(b_n^1, b_n^0) && && n = 1, \ldots, N \\
u_{nm} &\sim \text{Bernoulli}(r_n) && && n = 1, \ldots, N, m = 1, \ldots, M \\
c_{lm} &\sim \text{Beta}(a_{lm}^1, a_{lm}^0) && && n = 1, \ldots, N, l = 1, \ldots, \infty \\
y_{nm} &\sim u_{nm}\,\text{Bernoulli}(c_{g_n,m}) && + (1 - u_{nm})\,\text{Bernoulli}(1 - c_{g_n,m}) && n = 1, \ldots, N, m = 1, \ldots, M
\end{aligned}
\tag{3.1}
$$

The hidden variable g_n is modeled with a discrete distribution and $\boldsymbol{\pi}_n = [\pi_{n,1}, \pi_{n,2}, \ldots]$ contains the cluster membership probabilities for SU n. This means that the distribution Discrete($\boldsymbol{\pi}_n$) generates the value $g_n = l$ with probability $\pi_{n,l}$, i.e., $\mathbb{P}\{g_n = l\} = \pi_{n,l}$ and $\sum_l \pi_{n,l} = 1$ for $n = 1, \ldots, N$. Number of the clusters will be determined by capturing the correlation among the SUs.

The variable for the device reliability, r_n, is modeled with a Beta distribution with parameters b_n^1 and b_n^0. This is the natural choice since r_n is used as the parameter of the Bernoulli distribution of the channel specific reliability and the conjugate prior for a Bernoulli distribution is Beta distribution.

The channel-specific reliability is modeled as a Bernoulli distribution with parameter r_n. This implies that if a device is generally reliable, i.e., r_n close to 1, it is more likely to be reliable on different channels. This prior links the performance of the device on different channels and reduces the chance of overfitting the channel-specific reliability.

The channel availability is also defined as a random variable sampled from a Beta distribution with parameters a_{lm}^1 and a_{lm}^0. This is also because c_{lm} is later used as the parameter of the Bernoulli distribution that describes the observations.

Finally, the observed variable, y_{nm}, which is the reported occupancy status of channel m by SU n, is defined as the summation of two Bernoulli distributions. In words, if the SU is reliable on a channel, $u_{nm} = 1$, the distribution would be $\mathrm{Bernoulli}(c_{g_n,m})$. This means that y_{nm} will be sampled from a Bernoulli distribution with true parameter of channel availability, i.e., $c_{g_n,m}$. Otherwise, for $u_{nm} = 0$, it will be sampled from $\mathrm{Bernoulli}(1 - c_{g_n,m})$.

The goal of the BS is to infer the distribution of the hidden variables r_n, u_{nm}, g_n, and c_{lm} given the observations y_{nm}. The current state of the network can summarized using the distributions of hidden random variables. In Chapter 4, we will discuss how the distributions are initialized and updated after receiving the observations.

CHAPTER 4: INFERENCE VIA SEQUENTIAL BAYESIAN UPDATING

In the Bayesian approach, the goal is to infer the distribution of the hidden variables given the observations, i.e., $\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}|\boldsymbol{y}\}$, which is known as the posterior distribution. For compactness of notation, we set $\boldsymbol{y} = \{y_{nm}\}$, $\boldsymbol{g} = \{g_n\}$, $\boldsymbol{r} = \{r_n\}$, $\boldsymbol{u} = \{u_{nm}\}$, and $\boldsymbol{c} = \{g_{lm}\}$. Using the Bayes rule, the posterior distribution can be written as:

$$\begin{aligned}\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}|\boldsymbol{y}\} &\propto \mathbb{P}\{\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\} \\ &\propto \mathbb{P}\{\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}.\end{aligned}$$

$\mathbb{P}\{\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}$ can be calculated using the model described in (3.1). Specifically, the last expression in (3.1), can be used to build $\mathbb{P}\{\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}$, known as the likelihood of the observations, and the other terms represent our prior belief, $\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}$. The posterior distribution is the updated distribution of the hidden variables after receiving the observations.

In sequential Bayesian updating, the prior knowledge of the model is represented as the prior distribution, which is the distribution of the hidden variables before collecting data. After observing the first set of measurements, the posterior distribution is determined using the Bayes rule. Then, the posterior distribution can be used as the prior when the next set of observations becomes available. Thus, the problem boils down to updating the distribution of the hidden variables using the observations. In this approach, all the information is stored in the current distribution of the hidden variables. Hence, old observations and distributions can be ignored. This helps us obtain update rules that are not computationally burdensome. Moreover, since the distribution can be updated using a single measurement of a single SU or a complete set of measurements from all the channels and all the SUs, the BS can easily handle missing entries and different rates of data stream.

In the following, we will present how the distribution of hidden variables are initialized and how they are updated after receiving each set of measurements.

4.1 Initialization

As mentioned earlier, the initialization of the distribution of the hidden variables reflects our prior knowledge. Thus, there is no universal recipe to initialize the parameters. Also, it is clear that as the BS collects more measurements, the effect of the initialization becomes less and less significant.

The parameter sets a_{lm}^1 and a_{lm}^0 determine the distribution of channel availability for each cluster. In this work, no prior information is assumed on the channels availability and the truth will be inferred completely *a posteriori*. Thus, the parameters are initialized as $a_{lm}^1 = 1$ and $a_{lm}^0 = 1,\ \forall l, m$. This choice of parameters results in a uniform distribution for the probability of the channel availability.

To set the parameters for the device reliability, it is plausible to assume that on average at least half of the measurements are reliable. Also, since there is no information regarding which devices might report faulty data, the parameters of all the devices should be initialized in a similar manner. In our numerical experiments, we initialize $b_n^1 = 3$ and $b_n^0 = 1,\ \forall l, m$, which means on average 75% of the measurements are reliable. This is because the expected value of a random variable with distribution $\mathrm{Beta}(b_n^1, b_n^0)$ is $\frac{b_n^1}{b_n^1 + b_n^0}$

To initialize the parameters for cluster membership, stick-breaking construction is used [51]. The class of stick-breaking priors are widely used in different classification problems, where the complexity of the model is unknown. In general, the complexity of model (number of the clusters) can be either bounded or unbounded. Although, the model description in (3.1) presented with an unbounded number of clusters, it is reasonable and also more practical to assume an upper bound,

L_{max}, for the number of candidate clusters. Here, the number of candidate clusters can be assumed to be no larger than the number of devices, i.e., $L_{max} = N$.

The weights of a finite dimensional stick-breaking prior with concentration parameter α can be formulated as $\eta_l = \rho_l \prod_{i=1}^{l-1}(1-\rho_i),\ \forall l$, where ρ_l are independent random variables drawn from $\mathrm{Beta}(1,\alpha)$. Setting $\rho_{L_{max}} = 1$ will result in a finite dimensional prior [51]. Thus, the cluster membership distribution is set as $\boldsymbol{\pi}_n = \boldsymbol{\eta}$ for all the devices, where $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_{L_{max}}]$.

Parameter α represents the prior information on degree of correlation among the SUs. Specifically, according to the defined stick-breaking prior, it is known that the probability that two SUs belong to the same group is equal to $\frac{1}{1+\alpha}$ [52]. Thus, as we increase the value of α, less SUs may end up in the same group. For initialization, we set $\alpha = L_{max}$, which indicates a low degree of correlation among SUs.

To make the model more accurate, Figure 4.1 shows the graphical model including the stick-breaking prior. As a result, the Bayes rules should be rewritten as

$$\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho} | \boldsymbol{y}\} \propto \mathbb{P}\{\boldsymbol{y} | \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\} \mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\},$$

where $\boldsymbol{\rho} = \{\rho_l\}$ denotes the set of stick-breaking variables, which is the same for all the SUs. In this model, $\boldsymbol{\rho}$ links the information of cluster membership from different SUs, captures the level of correlation among SUs, and helps us cluster SUs without knowing the number of clusters. Without ρ, the variables of cluster membership g_n for different SUs would be independent of each other, which is not the case in our application.

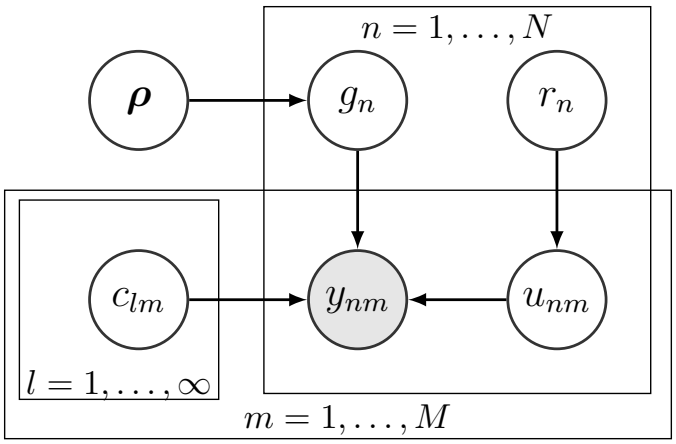

Figure 4.1: Graphical model including the stick-breaking prior.

4.2 Sequential Updating

As the new observations become available, the BS has to update the posterior distributions. Using the model defined in (3.1) and depicted in Figure 4.1, the joint distribution of the hidden and observed variables can be presented as:

$$\begin{aligned}\mathbb{P}\{\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\} = &\prod_{n,m} \mathbb{P}\{y_{nm}|g_n, u_{nm}, c_{g_n m}\} \prod_{n} \mathbb{P}\{g_n|\boldsymbol{\rho}\}\mathbb{P}\{\boldsymbol{\rho}\} \\ &\prod_{n,m} \mathbb{P}\{u_{nm}|r_n\}\mathbb{P}\{r_n|b_n^1, b_n^0\} \prod_{l,m} \mathbb{P}\{c_{lm}|a_{lm}^1, a_{lm}^0\},\end{aligned} \tag{4.1}$$

Now, given the observations, the goal is to infer the distribution of cluster membership, channel availability, channel-specific reliability, and device reliability. Mathematically speaking, we want to calculate $\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}|\boldsymbol{y}\}$ in a timely manner. However, it is not practically possible to directly find $\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}|\boldsymbol{y}\}$. This is due to the fact that the number of integrals, which are needed to be evaluated, grow exponentially.

To handle the intractable integrals arising in the inference procedure, *variational inference* is often employed [43, 52, 53] . In variational inference, the posterior distribution is approximated by a family of distributions, for which the calculations are tractable. The approximate distribution is often assumed to be fully factorized over all the hidden variables. Specifically, we can define the fully factorized variational distribution $\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\}$ as:

$$\begin{aligned}\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\} = &\prod_{n} \mathbb{Q}\{g_n|\hat{\boldsymbol{\pi}}_n\} \prod_{n,m} \mathbb{Q}\{u_{nm}|\tau_{nm}\} \\ &\prod_{n} \mathbb{Q}\{r_n|\hat{b}_n^1, \hat{b}_n^0\} \prod_{l,m} \mathbb{Q}\{c_{lm}|\hat{a}_{lm}^1, \hat{a}_{lm}^0\} \prod_{l} \mathbb{Q}\{\rho_l|\hat{\gamma}_l^1, \hat{\gamma}_l^0\},\end{aligned} \tag{4.2}$$

where $\hat{\boldsymbol{\pi}}_n$, $\hat{b}_n^1$, $\hat{b}_n^0$, $\hat{a}_{lm}^1$, $\hat{a}_{lm}^0$, τ_{nm}, $\hat{\gamma}_l^1$, and $\hat{\gamma}_l^0$ are the parameters of the factorized distributions. Our goal is to obtain $\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\}$ such that it approximates the posterior distribution $\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}|\boldsymbol{y}\}$.

Specifically, in variational inference, the goal is to obtain a distribution $\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\}$ that maximizes the likelihood of the observations. It is easy to show that the log-likelihood of the observations can be written as [54, Chapter 10]:

$$\begin{aligned}\ln(\mathbb{P}\{\boldsymbol{y}\}) \geq &\mathcal{L}(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\}) + \mathrm{KL}(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\} \| \mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}|\boldsymbol{y}\}), \\ \text{where} \quad \mathcal{L}(\mathbb{Q}) = &\int \mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\} \ln(\mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}|\boldsymbol{y}\}) - \int \mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\} \ln(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}) \\ &\triangleq \mathbb{E}\{\ln(\mathbb{P}\{\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\})\} - \mathbb{E}\{\ln(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\})\},\end{aligned}$$

and the expected value is with respect to variational distribution $\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}\}$. $\mathrm{KL}(.\|.)$ is the Kull-backLeibler (KL) divergence of two distributions. The equality occurs when $\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\} = \mathbb{P}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}|\boldsymbol{y}\}$, which means KL-divergence is zero. This means $\mathcal{L}(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\})$ is the largest possible lower bound that can be attained. However, since the variational distribution is limited to the family of fully factorized distributions, the maximum of the lower bound cannot be achieved.

Thus, the problem boils down to maximizing the $\mathcal{L}(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\})$ to find the best approximate posterior distribution. To do that, at each step, the lower bound is maximized with respect to only one of the factorized distributions, i.e., $\mathbb{Q}\{g_n|\hat{\boldsymbol{\pi}}_n\}$, $\mathbb{Q}\{u_{nm}|\tau_{nm}\}$, $\mathbb{Q}\{r_n|\hat{b}_n^1, \hat{b}_n^0\}$, $\mathbb{Q}\{c_{lm}|\hat{a}_{lm}^1, \hat{a}_{lm}^0\}$, or $\mathbb{Q}\{\rho_l|\hat{\gamma}_l^1, \hat{\gamma}_l^0\}$, keeping all the other distributions fixed. The procedure is repeated until convergence. Each step results in the update rule for one of the variables. Since $\mathcal{L}(\mathbb{Q}\{\boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\})$ is concave with respect to each of the factorized distributions, convergence is guaranteed [54]. The derivations of the updating rules are presented in Chapter 5.

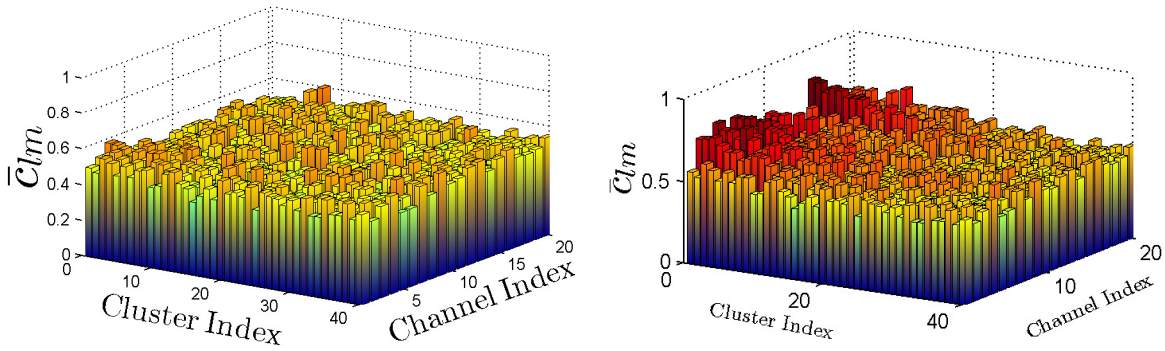

Figure 4.2: Expected value of channel availability (a) after 2 time frames and (b) after 20 time frames.

At each time frame, when the BS receives new observations, it employs the update rules and updates the distribution of the hidden variables. To illustrate the idea, Figure 4.2 shows the expected value of the channel availability for different clusters, i.e. $\bar{c}_{lm} = \mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{c_{lm}\}$, for a scenario similar to Figure 3.1. At the first time frame $\bar{c}_{lm}$, is equal to 0.5 for all the channels and the clusters, according to the initialization discussed in Chapter 4.1, which indicates unbiased estimate of channel availability in absence of further information. It is also worthwhile to mention that since the number of clusters the candidate clusters L_{max} is set to larger than the number of actual clusters, most of the clusters in Figure 4.2 are empty. Thus, the belief on channel availability of these clusters are not updated. On the hand, it is easy to see that, for the non-empty clusters, as the BS

collects more measurements, uncertainty decreases and the spectrum opportunities are revealed.

This sequential updating allows us to propagate information without requiring to store the old observations or repeating computations. In Chapter 6, the inferred distributions are exploited to assign channels for sensing and to detect the spectrum opportunities.

CHAPTER 5: DERIVATIONS OF UPDATE RULES

In this chapter, the derivations of the update rules for the inference algorithm are presented. As discussed in Chapter 4.2, the posterior distribution is approximated by a family of distributions for which the calculations are tractable. We restrict the distribution by fully factorizing it over all the hidden variables. This method is referred to as *naive mean field* approach [53].

For simplicity of notation, let us denote the whole set of hidden variables with

$$\boldsymbol{Z} = \{\{g_n\}, \{r_n\}, \{u_{nm}\}, \{c_{lm}\}, \{\rho_l\}\}$$

. In (4.2), $\boldsymbol{Z}$ is divided into disjoint groups $Z_i, i = 1, \ldots,$ where each Z_i is representing one of the hidden variables in $\boldsymbol{Z}$. By maximizing the lower bound $\mathcal{L}(\mathbb{Q}\{\boldsymbol{Z}\})$, the variational distribution of each partition $\mathbb{Q}\{\boldsymbol{Z}_i\}$ is given by [54, Chapter 10]:

$$\ln(\mathbb{Q}\{\boldsymbol{Z}_i\}) = \mathbb{E}_{j\neq i}\{\ln(\mathbb{P}\{\boldsymbol{y}, \boldsymbol{Z}\})\} + const, \tag{5.1}$$

where $\mathbb{E}_{j\neq i}\{.\}$ is the expectation with respect to distributions $\mathbb{Q}\{\boldsymbol{Z}_j\}, j \neq i$. Then by plugging in $\mathbb{P}\{\boldsymbol{y}, \boldsymbol{Z}\} = \mathbb{P}\{\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\}$ from (4.1) and employing the exponential form of the distributions, the variational distributions can be obtained. The constant value is determined by normalizing the distribution.

It is worthwhile to state that if $x \sim \text{Bernoulli}(p)$, then

$$\ln(\mathbb{P}\{x\}) = \ln(\frac{p}{1-p})x + \ln(1-p) \tag{5.2}$$

and if $x \sim \mathrm{Beta}(b^1, b^0)$, we have

$$\begin{aligned} \ln(\mathbb{P}\{x\}) &= (b^1 - 1)\ln(x) + (b^0 - 1)\ln(1-x) + const \\ \mathbb{E}\{\ln(x)\} &= \psi(b^1) - \psi(b^1 + b^0), \\ \mathbb{E}\{\ln(1-x)\} &= \psi(b^0) - \psi(b^1 + b^0), \end{aligned} \tag{5.3}$$

where $\psi(.)$ is digamma functions. Here, using the observations $\boldsymbol{y}$ and the prior distribution, the update rules to obtain the approximate posterior distributions are presented.

5.1 Channel Availability

Using (5.1), to update the channel availability, we have:

$$\ln(\mathbb{Q}\{c_{lm}\}) = \mathbb{E}\{\ln(\mathbb{P}\{\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{c}, \boldsymbol{\rho}\})\} + const.$$

Employing (5.2) and (5.3) and integrating out all the variables except c_{lm}, we will have:

$$\begin{aligned} &\ln(\mathbb{Q}\{c_{lm}\}) = const + \ln(\mathbb{P}\{c_{lm}|a^1_{lm}, a^0_{lm}\}), \\ &+ \sum_n \mathbb{E}_{\mathbb{Q}\{g_n\}}\mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{\ln(\mathbb{P}\{y_{nm}|u_{nm}, g_n, c_{lm}\})\} \\ &= const + (a^1_{lm} - 1)\ln(c_{lm}) + (a^0_{lm} - 1)\ln(1 - c_{lm}) \\ &+ \sum_n \hat{\pi}_{n,l}[\mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}(\ln(\frac{c_{lm}}{1-c_{lm}})y_{nm} + \ln(1-c_{lm})) \\ &+ (1 - \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\})(\ln(\frac{1-c_{lm}}{c_{lm}})y_{nm} + \ln(c_{lm}))]. \end{aligned}$$

In this expression, the summations are taken over the devices that have reported a measurement on m^{th} device. The first two terms come from the prior knowledge on the channel status and are

written using (5.3). Moreover, the summation is aggregating the information from observations, considering the reliability of each measurement and the cluster membership of each device. The terms inside the summation are written using the exponential form of Bernoulli distribution, as described in (5.2).

This expression can be further written in the form of $(\hat{a}^1_{lm}-1)\ln(c_{lm})+(\hat{a}^0_{lm}-1)\ln(1-c_{lm})+const$, which is a Beta distribution with parameters:

$$\hat{a}^1_{lm} = a^1_{lm} + \sum_n \hat{\pi}_{n,l}[\mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}y_{nm} + (1 - \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\})(1 - y_{nm})]$$
$$\hat{a}^0_{lm} = a^0_{lm} + \sum_n \hat{\pi}_{n,l}[\mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}(1 - y_{nm}) + (1 - \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\})y_{nm}].$$

Here, $\mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{.\}$ is expectation with respect to $\mathbb{Q}\{u_{nm}\}$ and $\mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}$ can be calculated using $\mathbb{Q}\{u_{nm}\}$, which will be discussed shortly.

5.2 Device Reliability

Similarly, to update the device reliability, for all n, we have:

$$\begin{aligned}\ln(\mathbb{Q}\{r_n\}) &= const + (b^1_n - 1)\ln(r_n) + (b^0_n - 1)\ln(1 - r_n) \\ &+ \sum_m \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{\ln(\frac{r_n}{1 - r_n})u_{nm} + \ln(1 - r_n)\} \\ &= const + \ln(r_n)(b^1_n + \sum_m \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\} - 1) \\ &+ \ln(1 - r_n)(b^0_n + \sum_m [1 - \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}] - 1),\end{aligned}$$

By comparing this expression to the exponential form of the Beta distribution, it is easy to see that $\mathbb{Q}\{r_n\}$ is a Beta distribution with parameters $\hat{b}^1_n = b^1_n + \sum_m \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}$ and $\hat{b}^0_n = b^0_n + M -$

$\sum_m \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\}$.

5.3 Channel-Specific Reliability

Again, by integrating out all the variables except u_{nm}:

$$\ln(\mathbb{Q}\{u_{nm}\}) = const + \mathbb{E}_{\mathbb{Q}\{r_n\}}\{\ln(\mathbb{P}\{u_{nm}|r_n\})\} + \mathbb{E}_{\mathbb{Q}\{g_n\}}\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(\mathbb{P}\{y_{nm}|u_{nm}, g_n, c_{lm}\})\}$$

By employing (5.2) and (5.3), we have:

$$\begin{aligned}\ln(\mathbb{Q}\{u_{nm}\}) = const &+ u_{nm}\mathbb{E}_{\mathbb{Q}\{r_n\}}\{\ln(r_n)\} + (1-u_{nm})\mathbb{E}_{\mathbb{Q}\{r_n\}}\{\ln(1-r_n)\} \\ &+ u_{nm}[y_{nm}\sum_l \hat{\pi}_{n,l}\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(c_{lm})\} + (1-y_{nm})\sum_l \hat{\pi}_{n,l}\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(1-c_{lm})\}] \\ &+ (1-u_{nm})[y_{nm}\sum_l \hat{\pi}_{n,l}\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(1-c_{lm})\} + (1-y_{nm})\sum_l \hat{\pi}_{n,l}\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(c_{lm})\}].\end{aligned}$$

This update rule, like the other ones, boils down to simple expressions, as the observations are either 0 or 1. It is intuitive that, at any time frame, the BS cannot update u_{nm}, if device n has not reported a measurement on channel m. Thus, the update rule is employed for each pair of SUs and channels that the BS has received a new observation.

To update the distribution, the expression is evaluated for $u_{nm} = 0$ and $u_{nm} = 1$. Since it is shown that $\mathbb{Q}\{c_{lm}\}$ and $\mathbb{Q}\{r_n\}$ are Beta distributions, $\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(c_{lm})\}$, $\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(1-c_{lm})\}$, $\mathbb{E}_{\mathbb{Q}\{r_n\}}\{\ln(r_n)\}$, and $\mathbb{E}_{\mathbb{Q}\{r_n\}}\{\ln(1-r_n)\}$ can be calculated using (5.3). The summations are comparing the observations with the expected value of y_{nm}, coming from the model, and soft counting the agreements and disagreements.

After normalizing the probabilities to have a valid Bernoulli distribution, the parameter of the distribution can be updated as $\tau_{nm} = \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\{u_{nm}\} = \mathbb{Q}\{u_{nm} = 1\}$.

5.4 Stick-Breaking Variables

Using (5.1), the update rule can be derived as:

$$\ln(\mathbb{Q}\{\rho_l\}) = const + \ln(\mathbb{P}\{\rho_l|\gamma_l^1, \gamma_l^0\}) + \sum_n \mathbb{E}_{\mathbb{Q}\{\boldsymbol{g}_n\}}\{\ln(\mathbb{P}\{g_n|\boldsymbol{\rho}\})\}.$$

Notice that, as mentioned in Chapter 4.1, $\mathbb{P}\{\rho_l|\gamma_l^1, \gamma_l^0\}$ is initialized with $\mathrm{Beta}(1, \alpha)$. $\mathbb{P}\{g_n|\boldsymbol{\rho}\}$ can be written in terms of $\boldsymbol{\rho}$, by utilizing the stick-breaking construction rules. Also, $\mathbb{P}\{\rho_l|\gamma_l^1, \gamma_l^0\}$ can be rewritten using (5.3). Thus, we have:

$$\begin{aligned}\ln(\mathbb{Q}\{\rho_l\}) = const &+ (\gamma_l^1 - 1)\ln(\rho_l) + (\gamma_l^0 - 1)\ln(1 - \rho_l) \\ &+ \sum_n \sum_{i=1}^{L_{max}} \hat{\pi}_{n,i}[\ln(\rho_i) + \sum_{j=1}^{i-1} \ln(1 - \rho_j)],\end{aligned}$$

which can be written in form of $(\hat{\gamma}_l^1 - 1)\ln(\rho_l) + (\hat{\gamma}_l^0 - 1)\ln(1 - \rho_l) + const$. This is also a Beta distribution with parameters:

$$\begin{aligned}\hat{\gamma}_l^1 &= \gamma_l^1 + \sum_n \hat{\pi}_{n,l}, \\ \hat{\gamma}_l^0 &= \gamma_l^0 + \sum_n \sum_{i=l+1}^{L_{max}} \hat{\pi}_{n,i}.\end{aligned}$$

5.5 Cluster Membership

The probability of device n belonging to cluster l, i.e., $\hat{\pi}_{n,l}$, is updated using this expression:

$$\begin{aligned}\ln(\mathbb{Q}\{g_n = l\}) = const &+ \mathbb{E}_{\mathbb{Q}\{\boldsymbol{\rho}\}}\{\ln(\mathbb{P}\{g_n = l|\boldsymbol{\rho}\)\}, \\ &+ \sum_m \mathbb{E}_{\mathbb{Q}\{u_{nm}\}}\mathbb{E}_{\mathbb{Q}\{c_{lm}\}}\{\ln(\mathbb{P}\{y_{nm}|u_{nm}, g_n, c_{lm}\})\}.\end{aligned}$$

Again, by using stick-breaking construction rules and (5.2):

$$\begin{aligned}\ln(\mathbb{Q}\{g_n = l\}) = const + \mathbb{E}_{\mathbb{Q}\{\rho_l\}}\{\ln(\rho_l)\} + \sum_{i=1}^{l-1} \mathbb{E}_{\mathbb{Q}\{\rho_l\}}\{\ln(1-\rho_l)\} \\ + \sum_{m}[\tau_{nm}\mathbb{E}_{\mathbb{Q}\{\boldsymbol{c}_{lm}\}}\{\ln(c_{lm})y_{nm} + \ln(1-c_{lm})(1-y_{nm})\} \\ + (1-\tau_{nm})\mathbb{E}_{\mathbb{Q}\{\boldsymbol{c}_{lm}\}}\{\ln(1-c_{lm})y_{nm} + \ln(c_{lm})(1-y_{nm})\}],\end{aligned}$$

where, the summation is over the channels that device n has reported a measurement. $\mathbb{E}_{\mathbb{Q}\{\boldsymbol{c}_{lm}\}}\{\ln(c_{lm})\}$, $\mathbb{E}_{\mathbb{Q}\{\boldsymbol{c}_{lm}\}}\{\ln(1-c_{lm})\}$, $\mathbb{E}_{\mathbb{Q}\{\rho_l\}}\{\ln(\rho_l)\}$, and $\mathbb{E}_{\mathbb{Q}\{\rho_l\}}\{\ln(1-\rho_l)\}$ can be evaluated using (5.3) and the distributions $\mathbb{Q}\{g_n\}$ are normalized to represent a valid probability distribution.

The derived update rules are employed alternatively to update the parameters of the variational distributions, until convergence is achieved. Generally, the per iteration complexity of the update rules is $\mathcal{O}(NML_{max})$. However, considering the fact that there are at most N measurements at each time frame, the complexity of the inference algorithm becomes $\mathcal{O}(N\max\{M, L_{max}\})$.

CHAPTER 6: SPECTRUM OPPORTUNITY DISCOVERY VIA MULTI-LABEL GRAPH CUTS

As mentioned earlier, it is assumed that the network is performing on a frame-by-frame structure. At the beginning of each frame, the BS assigns a channel to each SU. Then, if the channel is sensed as empty, the SU utilizes the channel. Moreover, the SU reports the outcome of the sensing to the BS. Here, we present the procedure for discovering the spectrum opportunities, using the distribution of the hidden variables.

In order to increase the number of *unique* spectrum opportunity in space and to reduce the risk of interference, same spectrum opportunity should not be appointed to neighboring SUs. Thus, the BS has to assign spectrum opportunities to SUs such that the channels are more likely to be available; and the same channel is not assigned to physically adjacent SUs.

Since the location information of the SUs is not available, cluster membership information is exploited as a measure of adjacency. This is based on the assumption that measurements from neighboring SUs are correlated. Thus, SUs with correlated measurements will end up in the same cluster. Figure 3.1 shows the outcome of the proposed clustering for a sample environment. In the figure, the n^{th} SU is assigned a cluster l^*, where $l^* = \max_l \mathbb{Q}\{g_n = l\}$. It is easy to notice that the neighboring nodes are more likely to be in the same cluster.

In this chapter, we present how the spectrum opportunities can be assigned to SUs by mapping the problem to multi-label energy minimization via graph cuts [45]. First, we need to quantify the probability that a channel is available for each of the SUs. Let t_{nm} be the true occupancy status of the channel m for device n, which clearly might be different from the observed value y_{nm}. If channel m is available for device n, t_{nm} will be equal to 1, otherwise it will be 0. Then, $\mathbb{Q}\{t_{nm}\}$,

which is the inferred variational distribution for t_{nm}, can be calculated as the marginal distribution of $\mathbb{Q}\{t_{nm}, g_n, c_{lm}\}$, given by:

$$\begin{aligned}
\mathbb{Q}\{t_{nm}=1\} &= \sum_{l=1}^{L_{max}} \int_{c_{lm}} \mathbb{Q}\{t_{nm}=1, g_n=l, c_{lm}\} dc_{lm} \\
&= \sum_{l=1}^{L_{max}} \int_{c_{lm}} \mathbb{Q}\{t_{nm}=1|g_n=l, c_{lm}\}\mathbb{Q}\{g_n=l\}\mathbb{Q}\{c_{lm}\} dc_{lm} \\
&\overset{(a)}{=} \sum_{l=1}^{L_{max}} \mathbb{Q}\{g_n=l\} \int_{c_{lm}} c_{lm}\mathbb{Q}\{c_{lm}\} dc_{lm} \\
&\overset{(b)}{=} \sum_{l=1}^{L_{max}} \hat{\pi}_{n,l} \frac{\hat{a}^1_{lm}}{\hat{a}^1_{lm}+\hat{a}^0_{lm}}.
\end{aligned} \tag{6.1}$$

Here, (a) uses the fact that $\mathbb{P}\{t_{nm}=1|g_n=l, c_{lm}\} = c_{lm}$. This comes from the definition of the c_{lm}, which is the probability of channel m being available for devices in cluster l. Moreover, (b) exploits the fact that $\mathbb{Q}\{c_{lm}\}$ is a Beta distribution with parameters $\hat{a}^1_{lm}$ and $\hat{a}^0_{lm}$, which is shown in Chapter 5.

Now, let $h_n \in \{1, \ldots, M\}$ be the channel detected as the spectrum opportunity for device n and $\boldsymbol{h} = \{h_1, \ldots, h_N\}$ be the set of spectrum opportunities assigned to all of the SUs. The goal of the BS is to find the best set of channels $\boldsymbol{h}^* = \{h^*_1, \ldots, h^*_N\}$, which are most likely to be *unique spectrum opportunities*. For that, the BS has to ensure that the h^*_n is available for device n and is not assigned to its neighbors.

Accordingly, the following objective function can be defined:

$$\boldsymbol{h}^* = \arg\max_{\boldsymbol{h}} \prod_{n=1}^{N} \mathbb{P}\{t_{nh_n}=1\} \prod_{\substack{\forall n_i, n_j \\ h_{n_j} \neq h_{n_i}}} \mathbb{P}\{g_{n_i} = g_{n_j}\}. \tag{6.2}$$

The first term is the probability that all the assigned channels are available. The second term

indicates the probability that the SUs in the same cluster are assigned to different channels. This way, the BS makes sure that the SUs that are spatially close to each other will be assigned different channels. Using the negative log-likelihood, the objective can be further rewritten as:

$$\boldsymbol{h}^* = \arg\min \sum_{n=1}^{N} -\log(\mathbb{P}\{t_{nh_n} = 1\}) + \sum_{n_i=1}^{N}\sum_{n_j=1}^{N} -\log(\mathbb{P}\{g_{n_i} = g_{n_j}\})\delta(h_{n_j} \neq h_{n_i}). \tag{6.3}$$

$\delta(h_{n_j} \neq h_{n_i})$ is equal to 1, when $h_{n_j} \neq h_{n_i}$ and 0 other wise. It is known that the optimization problems with similar cost function as the problem formulated in (6.3) can be solved via graph cuts [44–47].

Graph cuts are often used in computer vision to assign each pixel of an image a label, while ensuring that the similar pixels are assigned the same label [44, 46]. Likewise, in our problem, the BS wants to assign each SU a channel, while assigning the nonadjacent SUs the same channel.

A cut divides a graph into disjoint sub-graphs by removing the edges of the graph. A minimum graph cut $\mathcal{C}$ is a cut that minimizes the cost of the cut $|\mathcal{C}|$, which is defined as the sum of weights of edges removed by the cut.

As an example, consider the simple network depicted in Figure 6.1(a). This network consists of 5 SUs, shown by circles, and 3 channels, represented by squares. This network can be represented as a graph of 8 vertices, containing one vertex for each SU and one vertex for each channel. The edge between each pair of SUs, denotes the probability that this pair of SUs does not belong to that same cluster. Thus, if a pair of SUs are likely to be in the same cluster, the weight of the edge will be small and it will be more likely to be removed by the graph cut algorithm. Moreover, the weight of the edge that links SU n to channel m represents the probability that channel m is not available for device n. Hence, if a channel is available for an SU, it will not easily be removed from the graph.

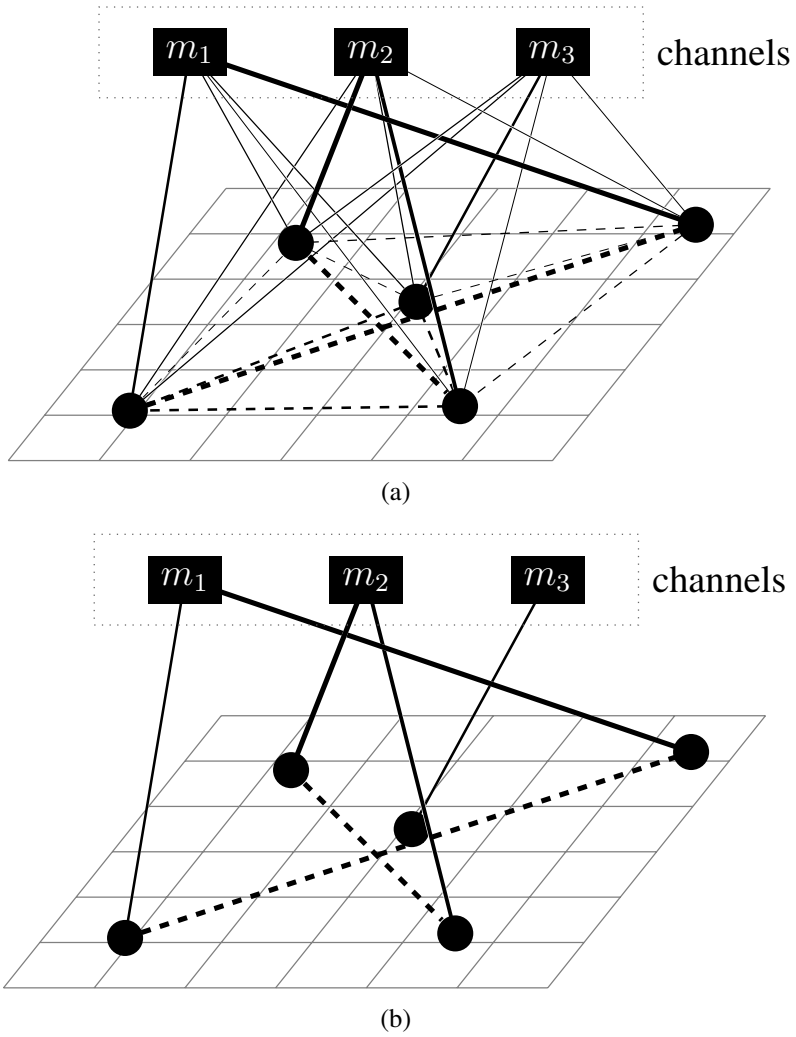

Figure 6.1: (a) An example of spectrum opportunity assignment using multi label energy minimization. The thicker lines represent the edges with larger wrights. (b) A multiway cut on the graph consisting of the channels and the SUs. In (a), each SU is connected to all the other SUs and all the channels. A multiway cut divides the SUs into disjoint groups and assigns each group a unique channel.

In Figure 6.1(b), an example of channel assignment is depicted. The graph is partitioned into three disjoint subsets, one for each channel. Such partitioning is achieved by removing edges of the original graph via multi-label graph cut.

Specifically, to minimize (6.3), each channel-SU edge is weighted by $K + \log(\mathbb{P}\{t_{nh_n} = 1\})$,

where K is a constant greater than $\max\{\log(\mathbb{P}\{t_{nh_n} = 1\})\}, \forall n$, to make the weights positive. This is because removing an edge with negative weight decreases the cost of the cut. Moreover, each SU-SU edge is weighted by $-\log(\mathbb{P}\{g_{n_i} = g_{n_j}\})$.

Let h_n^* be the channel assigned to SU n. Using the defined weights, the cost of removing all the channel-SU edges to SU n, except the edge to h_n^*, would be

$$\sum_{m \neq h_n^*} K + \log(\mathbb{P}\{t_{nm} = 1\}).$$

Similarly, the cost of removing edges between SU n and all the SUs that are not assigned the same channel is:

$$\sum_{\substack{\forall n' \\ h_n^* \neq h_{n'}^*}} -\log(\mathbb{P}\{g_n = g_{n'}\}).$$

Hence, the total cost of the cut can be written as:

$$|\mathcal{C}| = \sum_{n=1}^{N} \sum_{m \neq h_n^*} K + \log(\mathbb{P}\{t_{nm} = 1\}) - \sum_{n_i=1}^{N} \sum_{n_j=1}^{N} \log(\mathbb{P}\{g_{n_i} = g_{n_j}\})\delta(h_{n_j}^* \neq h_{n_i}^*). \tag{6.4}$$

This cost can be further rewritten as:

$$\begin{aligned} |\mathcal{C}| =& N(M-1)K + \sum_{n=1}^{N} \sum_{m=1}^{M} \log(\mathbb{P}\{t_{nm} = 1\}) - \sum_{n=1}^{N} \log(\mathbb{P}\{t_{nh_n^*} = 1\}) \\ & - \sum_{n_i=1}^{N} \sum_{n_j=1}^{N} \log(\mathbb{P}\{g_{n_i} = g_{n_j}\})\delta(h_{n_j}^* \neq h_{n_i}^*). \end{aligned} \tag{6.5}$$

Since, the first two terms do not depend on the cut, a set $\boldsymbol{h}^*$ that minimizes the cost of the cut $|\mathcal{C}|$ will also minimize (6.3), which is equivalent to maximizing the objective function in (6.2). Thus, after inferring the distributions of the hidden variables, the BS can perform a graph cut algorithm, using the defined graph and weights, to find the spectrum opportunities.

CHAPTER 7: RESULTS

In this chapter, the simulation results are presented to quantify the performance of the proposed framework. The simulation parameters are set as follows, unless otherwise is stated. By default, we consider a square area with size 50 (distance unit)^2, with $N = 15$ stationary SUs and 4 stationary PUs uniformly at random distributed over the space. For simplicity, it is assumed that a channel will be unavailable if the SU is in the circular transmission range of a PU that is utilizing the channel. Each SU device is in the interference range of a PU or SU if they are closer than 2 distance units. Using such parameters, each SU has to share the spectrum with about 2.7 other SUs and 0.8 PUs on average.

We set the number of spectrum bands to 20 ($M = 20$) and each PU is transmitting on 6 channels, randomly selected at the beginning of the simulation. To model the PU's behavior, a two-state Markov chain is adapted. The PUs utilize the channels $\lambda = 0.5$ of the time and switch from inactive state to active state with probability 0.1.

At each time frame, SUs can sense and report the occupancy status of only one channel, which is specified by the BS. Furthermore, the contamination ratio is set to be $\beta = 0.2$, which means 20% of the measurements from the SUs are randomly drawn from a Bernoulli distribution with parameter 0.5. The BS collects the measurements, updates the posterior distributions, discovers the spectrum opportunities, and assigns channels for sensing in the next frame.

The performance of the proposed algorithm is compared with existing non-cooperative and cooperative methods such as *greedy adaptive learning* [48], *exponential-weight algorithm for exploration and exploitation* (Exp3) [55], and *cluster-based coordinated multiband spectrum sensing* (Cluster-CMSS) [34]. The code for the Cluster-CMSS method was provided by the authors. The performances are compared according to a defined time-averaged normalized success rate (TNSR).

Success rate indicates what ratio of discovered channels is in fact available and unique in a neighborhood of SUs. Thus, if an available channel is detected as a spectrum opportunity for several SUs, which are in interference range of each other, this spectrum opportunity is considered as spectrum opportunity only for one of the SUs. The success rate is averaged over $T = 100$ time slots and normalized by the maximum achievable success rate. Color-sensitive graph coloring (CSGC), introduced in [13], is employed to find the maximum achievable success rate. Location of all the SUs and the true occupancy status of channels are provided as the inputs of CSGC. Moreover, the algorithm introduced in [45] is exploited to find the minimum graph cut and discover the spectrum opportunities.

Choosing success rate as the performance metric makes the results as general as possible and is a common practice [13, 34, 48]. A higher TNSR leads to improved throughput of individual SUs, overall throughput of the network, packet loss rate, and transmission delay. It is also worthwhile to mention that the energy cost of the proposed method is same as the centralized competitors. Moreover, for a fair comparison, number of observations collected by the BS at each time slot is the same for all the algorithms, i.e., one channel per SU. This means that the cost of sensing is the same for all the approaches.

In our first numerical experiment, we start by examining the performance of the mentioned methods for different network topologies. Using the default setting defined above, Figure 7.1 shows the performance of different algorithms for 50 different topologies generated randomly. It is easy to notice that the proposed Bayesian clustering method outperforms the competitors for all the cases, which illustrates the fact that the proposed method is not sensitive to the network topology and is able to perform well for different topologies. To quantify the performance, Table 7.1 shows the mean, standard deviation, and relative standard deviation of different methods. The table illustrates that the proposed method has a low variation, while having a high average TNSR.

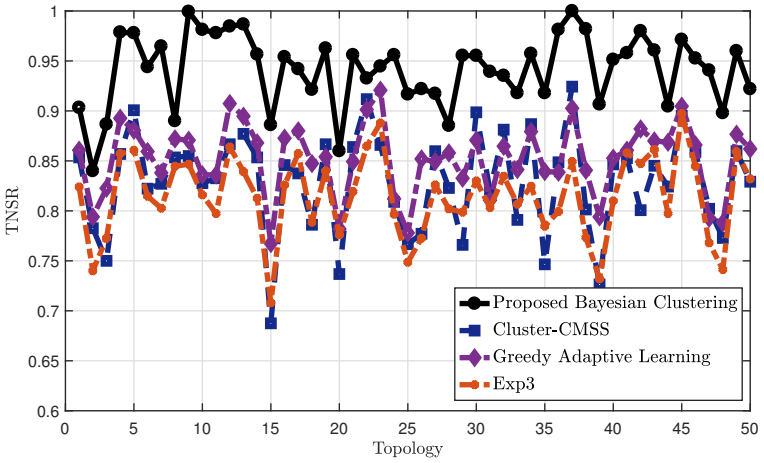

Figure 7.1: Performance of different frameworks for various topologies.

Table 7.1: Performance of different spectrum opportunity detection algorithms

Method	Mean	SD	RSD (%)
Proposed Bayesian Clustering	**0.94**	**0.036**	**3.83**
Cluster CMSS [34]	0.83	0.050	6.15
Greedy Adaptive Learning [48]	0.85	0.035	4.13
Exp3 [55]	0.81	0.042	5.18

To cluster the SUs and to draw conclusions about the availability of the channels, the proposed method exploits spatial correlation among the observations. This means that, to achieve a performance gain, there should exist some level of spatial correlation in the network.

Since the activity of the PUs are assumed to be uncorrelated, the SUs sensing the same PUs will report similar observations. Thus, as the number of SUs per each PU increases, the observations will be more correlated. Figure 7.2 illustrates the performance gain for different levels of spatial correlation, averaged over sufficiently large Monte Carlo trials. In this experiment, for a fixed number of SUs, i.e., $N = 15$, number of PUs is decreased from 15 to 1.

The effect of decreasing the number of PUs is two-fold. First, by decreasing the number of PUs,

more channels become available and it is easier to discover and assign the spectrum opportunities. This explains the improvement in the performance of all the methods. Also, for fewer number of PUs, the observations of the SUs are more correlated. Hence, the performance gap between the proposed method and other methods becomes increasingly larger as the number of PUs decreases.

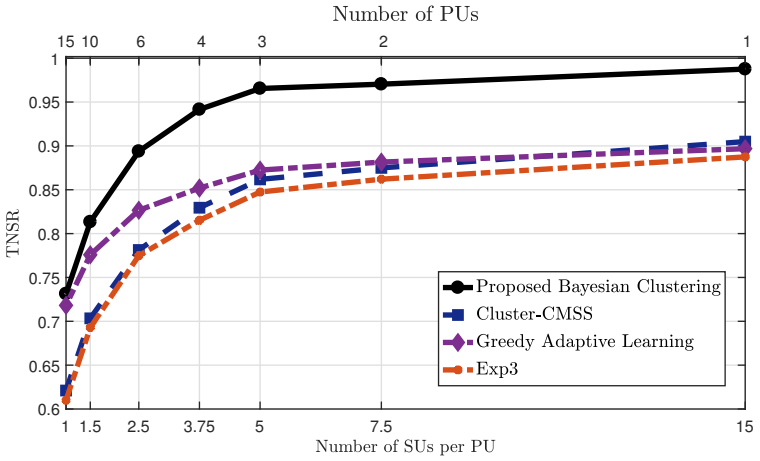

Figure 7.2: Spectrum opportunity detection performance for different levels of spatial correlation for $N = 15$. Performance gain is achieved when the number SUs is greater than the number of PUs, i.e., there exist some level of correlation among observations.

When the number of PUs is equal to the number of SUs, meaning that on average we have a PU for each SU, the observations are not correlated. Therefore, no performance gain is achieved by considering the correlation among the SU. However, as soon as the number of SUs for each PU becomes greater than 1, the proposed algorithm is able to employ the correlation to improve the spectrum opportunity detection performance, without knowing the location of SUs.

To study the influence of the utilization factor on the performance of the algorithm, Figure 7.3 compares the success rate of different algorithms for different utilization factors, i.e, λ. As it was expected, due to spectrum opportunity scarcity, the success rate of all the methods deteriorates for higher utilization factors.

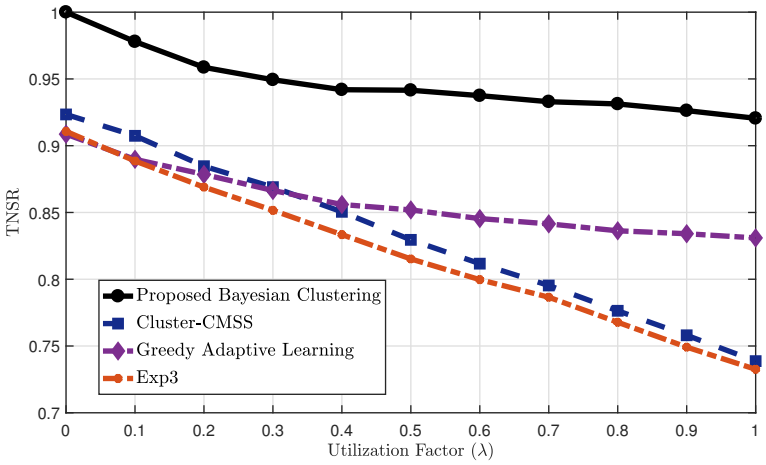

Figure 7.3: Spectrum opportunity detection performance for different utilization factors, i.e., λ, in presence of 4 PUs for environment size of 50 ($\sqrt{50} \times \sqrt{50}$) and interference range of 2.

It is also evident that the proposed method is only method that achieves an optimal performance for the best-case scenario, i.e., $TNSR = 1$, for $\lambda = 0$. This is because the proposed method is able to exploit the spatial correlation to avoid interference among the SUs. In such scenario, all the SUs are in the same cluster, due to similarity of the observations. Thus, a unique spectrum channel is assigned to each of them.

Density of the SU devices is also an important factor in the performance of any spectrum opportunity detection framework. Although, for larger number of SUs, the fusion center receives more data. However, as the number of SUs increases, for a fixed size of environment and a fixed transmission range, the scarcity of unique spectrum opportunities becomes more considerable, degrading the spectrum utilization. As an example, for $N = 25$, using the default environment size of$\sqrt{50}$ distance unit ($\approx 3.5\times$ interference range), each SU is sharing the spectrum channels with about 4.74 other SUs and 0.8 PUs, which are transmitting on 6 channels. Figure 7.4 shows that for large number of SUs, the success rate of different spectrum opportunity detection methods decreases.

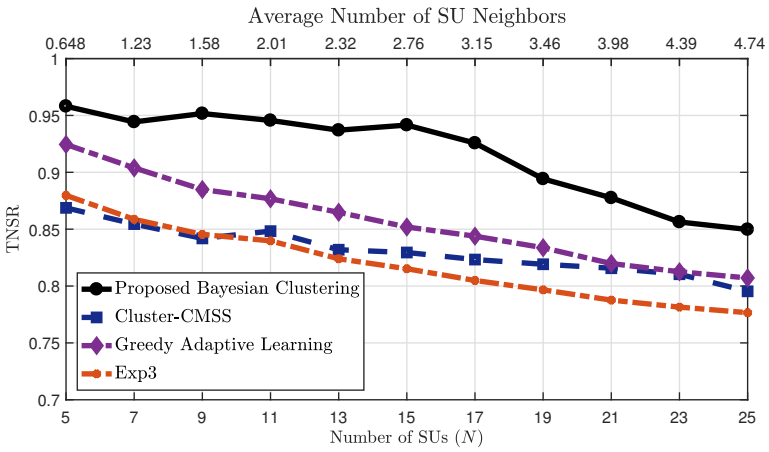

Figure 7.4: Spectrum opportunity detection performance with varying number of secondary users. There are 4 PUs, each transmitting on a subset of 6 channels out of $M = 20$ channels.

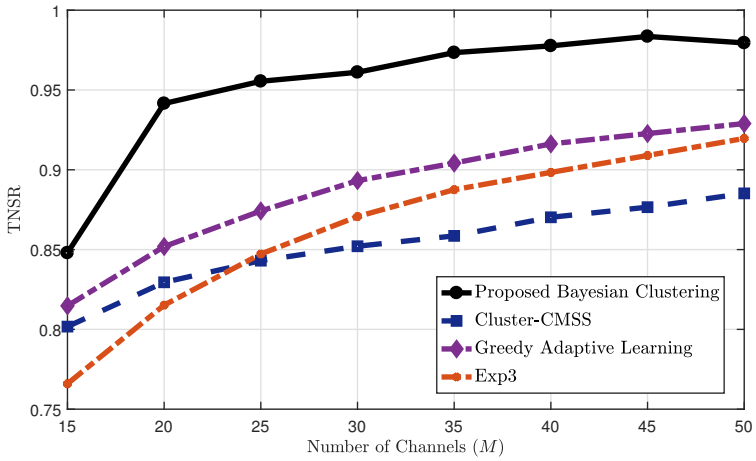

Figure 7.5: Spectrum opportunity detection performance for different number of channels and fixed number of SUs and PUs.

Unavailable location information puts certain limits on the performance of any cooperative spectrum opportunity detection algorithm. Two SUs might be far away from each other, while experiencing the exact same spectrum opportunities. Thus, in the absence of additional side information, there is no way to handle such scenario and have a perfect spectrum reuse. This issue is particularly

of more importance in large-scale networks with many SUs. For such networks, exploiting a single BS to aggregate the data from the entire network reduces the success rate. Hence, data form each region of the network should be processed by a local BS, also referred to as cluster head.

Finally, Figure 7.5 illustrates the fact that for fixed number of SUs and PUs, i.e., fixed level of spatial correlation, performance gain achieved by the proposed method becomes larger, as the spectrum opportunities increases. This is particularly due to the fact that the proposed method is able to aggregate data from different SUs to draw conclusions about the network and to discover more spectrum opportunities. This is one of the main advantages of the cooperative methods in general.

CHAPTER 8: CONCLUSIONS

In this thesis, we considered the problem of discovering the spectrum opportunities in spectrum-heterogeneous networks. It is further assumed that location information is absent and some of the measurements might be erroneous. These assumptions were due to the fact that we aim to achieve spectrum awareness by employing a network of low-cost devices. Thus, the devices have limited sensing capabilities.

We proposed a centralized, cooperative, Bayesian inference framework to extract information from the measurements, assuming the presence of faulty data and correlation among the measurements of different SUs. To construct the model, we considered the data collection process, by modeling device-specific and channel-specific reliabilities, as well as characteristics of the system, by modeling spatial correlation among the measurements.

To solve the problem and obtain the posterior distributions, which contain information on variables of interest, a sequential updating formulation is utilized. The update rules are derived mathematically by employing variational inference. After inferring the cluster membership and channel availability information, it is shown that multi-label graph cuts can be employed to discover the unique spectrum opportunities. The simulation results suggest that the proposed algorithm outperforms the existing cooperative and non-cooperative methods, whenever the measurements of different SUs are spatially correlated.

On the other hand, unavailable location information limits the efficiency of any spectrum exploitation method. Without knowing the location, perfect spatial frequency reuse is impossible. As a preliminary example, assume a wide area network without any primary licensed users. The measurements of the SUs are exactly the same, but they might be far away from each other. In the absence of further information, it is impossible to assign the frequency channels perfectly. Thus,

an interesting future research direction is to design a decentralized or distributed framework. In such frameworks, the observed data is processed locally by a cluster head, instead of processing the data from the entire network. This improves the efficiency of spatial spectrum reuse, by limiting the size of network. However, this comes at a cost. For less number of devices, the fusion center or the cluster head receives less data. Thus, as the size of the network decreases, the processing node will have less spectrum awareness, but needs to solve an easier problem for channel assignment. Hence, another interesting research question is to determine the best size of the network cells or the optimal number of processing nodes in a network.

LIST OF REFERENCES

[1] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, pp. 328–337, 11 2005.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[3] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE journal on selected areas in communications*, vol. 23, no. 2, pp. 201–220, 2005.

[4] M. B. H. Weiss, M. Altamimi, and L. Cui, "Spatio-temporal spectrum modeling: Taxonomy and economic evaluation of context acquisition," *Telecommunications Policy*, vol. 36, no. 4, pp. 335–348, 2012.

[5] V. Kekatos and G. B. Giannakis, "From Sparse Signals to Sparse Residuals for Robust Sensing," *Signal Processing, IEEE Transactions on*, vol. 59, pp. 3355–3368, 7 2011.

[6] P. J. Huber, *Robust statistics*. Springer, 2011.

[7] Z. Khan, J. J. Lehtomki, L. A. DaSilva, E. Hossain, M. Latva-Aho, J. J. Lehtomäki, L. A. DaSilva, E. Hossain, and M. Latva-Aho, "Opportunistic Channel Selection by Cognitive Wireless Nodes Under Imperfect Observations and Limited Memory: A Repeated Game Model," *IEEE Transactions on Mobile Computing*, vol. 15, pp. 173–187, 1 2016.

[8] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum

sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, 2009.

[9] C. S. Hyder, B. Grebur, L. Xiao, and M. Ellison, "ARC: adaptive reputation based clustering against spectrum sensing data falsification attacks," *IEEE Transactions on mobile computing*, vol. 13, no. 8, pp. 1707–1719, 2014.

[10] B. L. Mark and A. O. Nasif, "Estimation of maximum interference-free power level for opportunistic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2505–2513, 2009.

[11] H. Li, "Cooperative Spectrum Sensing via Belief Propagation in Spectrum-Heterogeneous Cognitive Radio Systems," in *2010 IEEE Wireless Communication and Networking Conference*, pp. 1–6, 4 2010.

[12] R. Vaze and C. R. Murthy, "Multiple Transmitter Localization and Whitespace Identification Using Randomly Deployed Binary Sensors," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, pp. 358–369, 12 2016.

[13] C. Peng, H. Zheng, and B. Y. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *Mobile Networks and Applications*, vol. 11, pp. 555–576, 8 2006.

[14] Z. Zhao, Z. Peng, S. Zheng, and J. Shang, "Cognitive radio spectrum allocation using evolutionary algorithms," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 4421–4425, 9 2009.

[15] T. Do and B. L. Mark, "Joint spatial–temporal spectrum sensing for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 7, pp. 3480–3490, 2010.

[16] Q. Wu, G. Ding, J. Wang, and Y. D. Yao, "Spatial-Temporal Opportunity Detection for Spectrum-Heterogeneous Cognitive Radio Networks: Two-Dimensional Sensing," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 516–526, 2 2013.

[17] Z. Wei, Z. Feng, Q. Zhang, and W. Li, "Three Regions for Space–Time Spectrum Sensing and Access in Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2448–2462, 2015.

[18] J. A. Bazerque and G. B. Giannakis, "Distributed Spectrum Sensing for Cognitive Radio Networks by Exploiting Sparsity," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1847–1862, 3 2010.

[19] S. Maleki, P. Ciblat, S. Chatzinotas, B. S. M. R., and B. Ottersten, "Cooperative Estimation of Power and Direction of Transmission for a Directive Source," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, pp. 343–357, 12 2016.

[20] S. Debroy, S. Bhattacharjee, and M. Chatterjee, "Spectrum Map and Its Application in Resource Management in Cognitive Radio Networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, pp. 406–419, 12 2015.

[21] X.-L. Huang, F. Hu, J. Wu, H.-H. Chen, G. Wang, and T. Jiang, "Intelligent Cooperative Spectrum Sensing via Hierarchical Dirichlet Process in Cognitive Radio Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 771–787, 5 2015.

[22] K. Connelly, Y. Liu, D. Bulwinkle, A. Miller, and I. Bobbitt, "A toolkit for automatically constructing outdoor radio maps," in *International Conference on Information Technology: Coding and Computing (ITCC), 2005.*, vol. 2, pp. 248–253, 4 2005.

[23] D. Denkovski, V. Atanasovski, L. Gavrilovska, J. Riihijarvi, and P. Mahonen, "Reliability of a radio environment Map: Case of spatial interpolation techniques," in *7th International ICST*

Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2012, pp. 248–253, 6 2012.

[24] E. Dall?Anese, J. A. Bazerque, and G. B. Giannakis, "Group sparse Lasso for cognitive network sensing robust to model uncertainties and outliers," *Physical Communication*, vol. 5, no. 2, pp. 161–172, 2012.

[25] S. Grimoud, B. Sayrac, S. Ben Jemaa, and E. Moulines, "An algorithm for fast REM construction," in *Sixth International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2011*, pp. 251–255, 6 2011.

[26] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN), 2012*, pp. 422–433, IEEE, 2012.

[27] G. Vanhoy, H. Volos, C. E. C. Bastidas, and T. Bose, "A Spatial Interpolation Method for Radio Frequency Maps Based on the Discrete Cosine Transform," in *IEEE Military Communications Conference, MILCOM 2013*, pp. 1045–1050, 11 2013.

[28] V. Atanasovski, J. van de Beek, A. Dejonghe, D. Denkovski, L. Gavrilovska, S. Grimoud, P. Mahonen, M. Pavloski, V. Rakovic, J. Riihijarvi, and B. Sayrac, "Constructing radio environment maps with heterogeneous spectrum sensors," in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*, pp. 660–661, 5 2011.

[29] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative Spectrum Sensing for Cognitive Radios Using Kriged Kalman Filtering," *IEEE Journal of Selected Topics in Signal Processing,*, vol. 5, pp. 24–36, 2 2011.

[30] Z. Wei, Q. Zhang, Z. Feng, W. Li, and T. A. Gulliver, "On the construction of radio environ-

ment maps for cognitive radio networks," in *IEEE Wireless Communications and Networking Conference (WCNC), 2013*, pp. 4504–4509, IEEE, 2013.

[31] J. Ojaniemi, J. Kalliovaara, A. Alam, J. Poikonen, and R. Wichman, "Optimal field measurement design for radio environment mapping," in *47th Annual Conference on Information Sciences and Systems (CISS), 2013*, pp. 1–6, 3 2013.

[32] T. Farnham, "{REM} based approach for hidden node detection and avoidance in cognitive radio networks," in *IEEE Global Communications Conference (GLOBECOM), 2012*, pp. 1391–1397, 12 2012.

[33] M. Zafer, B. J. Ko, and I.-H. Ho, "Transmit Power Estimation Using Spatially Diverse Measurements Under Wireless Fading," *IEEE/ACM Transactions on Networking,*, vol. 18, pp. 1171–1180, 8 2010.

[34] B. Shahrasbi, N. Rahnavard, and A. Vosoughi, "Cluster-CMSS: A Cluster-Based Coordinated Spectrum Sensing in Geographically Dispersed Mobile Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2016.

[35] Q. Zhao, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy," in *in IEEE Signal Processing Magazine*, pp. 79–89, 2007.

[36] S. K. Sharma, T. E. Bogale, S. Chatzinotas, B. Ottersten, L. B. Le, and X. Wang, "Cognitive Radio Techniques Under Practical Imperfections: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 1858–1884, 2015.

[37] 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 Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[38] H. B. Yilmaz, T. Tugcu, F. Alago?z, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *Communications Magazine, IEEE*, vol. 51, pp. 162–169, 12 2013.

[39] A. Zaeemzadeh, M. Joneidi, B. Shahrasbi, and N. Rahnavard, "Missing spectrum-data recovery in cognitive radio networks using piecewise constant Nonnegative Matrix Factorization," vol. 2015-Decem, pp. 238–243, IEEE, 10 2015.

[40] X. Jiang, K. K. Wong, Y. Zhang, and D. Edwards, "On hybrid overlay-underlay dynamic spectrum access: Double-threshold energy detection and markov model," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 4078–4083, 2013.

[41] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "A Hybrid Cognitive Transceiver Architecture: Sensing-Throughput Tradeoff," in *Proceedings of the 9th International Conference on Cognitive Radio Oriented Wireless Networks*, pp. 143–149, 2014.

[42] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science, 3rd ed., 2013.

[43] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.

[44] Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision.," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1124–1137, 9 2004.

[45] Y. Boykov, O. Veksler, and R. Zabih, "Efficient Approximate Energy Minimization via Graph Cuts," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1222–1239, 11 2001.

[46] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class Segmentation and Object Localization with Superpixel Neighborhoods," in *Proceedings of the International Conference on Computer Vision*, 10 2009.

[47] V. Kolmogorov and R. Zabih, "What Energy Functions can be Minimized via Graph Cuts?," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 147–159, 2 2004.

[48] C. Tekin, S. Hong, and W. Stark, "Enhancing cognitive radio dynamic spectrum sensing through adaptive learning," in *MILCOM 2009-2009 IEEE Military Communications Conference*, pp. 1–7, IEEE, 2009.

[49] A. De Domenico, E. C. Strinati, and M.-G. Di Benedetto, "A survey on MAC strategies for cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 1, pp. 21–44, 2012.

[50] B. Shahrasbi and N. Rahnavard, "A clustering-based coordinated spectrum sensing in wideband large-scale cognitive radio networks," in *2013 IEEE Global Communications Conference (GLOBECOM)*, pp. 1101–1106, IEEE, 2013.

[51] H. Ishwaran and L. F. James, "Some further developments for stick-breaking priors: finite and infinite clustering and classification," *Sankhy{\=a}: The Indian Journal of Statistics*, pp. 577–592, 2003.

[52] G.-J. Qi, C. C. Aggarwal, J. Han, and T. Huang, "Mining collective intelligence in diverse groups," in *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, (Republic and Canton of Geneva, Switzerland), pp. 1041–1052, ACM, International World Wide Web Conferences Steering Committee, 2013.

[53] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational

inference," *Foundations and Trends{®} in Machine Learning*, vol. 1, no. 1-2, pp. 1–305, 2008.

[54] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[55] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002.