Whitespace Measurement and Virtual Backbone Construction for Cognitive Radio Networks: From the Social Perspective

Shouling Ji[†], Zhipeng Cai[‡], Meng Han[‡], and Raheem Beyah[†]
[†]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA
[‡]Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA
Email: sji@gatech.edu, zcai@cs.gsu.edu, mhan7@student.gsu.edu, rbeyah@ece.gatech.edu

***Abstract*—The existing works on analyzing/utilizing spectrum whitespace in Cognitive Radio Networks (CRNs) are either empirical studies lacking of theoretical guarantee, or local primary network information based inducing inaccurate analysis and estimation, or overlooking the spectrum whitespace details. Therefore, we propose to systematically analyze the spectrum whitespace in CRNs from a social network perspective. Our main contributions include four parts. First, we propose a novel metric named *centrality score* to measure the active weights of Primary Users (PUs) by considering each PU's topological importance in the primary network and the global primary network running and traffic information. Subsequently, based on the centrality scores of PUs, we derive the whitespace in CRNs under different social patterns of primary activities. Since we consider both the primary network topological structure and the global network running and traffic information, our whitespace analysis is more accurate compared with the existing works. Third, according to our whitespace analysis, we design a Virtual Backbone (VB) construction algorithm, which aims to improve the spectrum utilization efficiency in CRNs. Finally, we conduct extensive simulations to validate our whitespace analysis and the VB construction algorithm. The simulation results demonstrate that our social attributes based whitespace analysis can accurately characterize the whitespace in CRNs and the VB construction algorithm significantly improves the performance of the existing VB-based CRN protocols.**

I. Introduction

Recently, with the fast progresses of wireless applications and devices, communications over unlicensed (free) spectrum become more and more crowded. On the other hand, according to the report from FCC [1], the utilization of licensed spectrum (assigned to some companies and agents) varies from 15% to 85% temporally and geographically, which is very inefficient. This necessitates a new communication paradigm, named Cognitive Radio Networks (CRNs), which enables unlicensed users, also named Secondary Users (SUs), to sense and learn the communication environment and then access the licensed spectrum without causing unacceptable interference to the licensed users, also named Primary Users (PUs).

Ever since the CRN paradigm is introduced, numerous efforts have been spent on the design and management issues for CRNs. One of the most fundamental issues is to understand how much spectrum bandwidth, named *whitespace*, is available for SUs. Accurate analysis on the whitespace is important for routing and scheduling protocol design, topology control, performance analysis, *etc*. Therefore, it is significant to conduct dedicated research on the whitespace analysis in CRNs.

978-1-4673-7331-9/15/$31.00

However, to the best of our knowledge, most existing works either provide some empirical evaluations on the available whitespace without systematical and rigorous mathematical analysis [2][3], or simply assume the available spectrum whitespace is a variable overlooking the details [8]-[11]. For some other works [12]-[16], the whitespace analysis based on simplified primary activity assumption and local primary network information is presented. Nevertheless, in practical, the available whitespace for SUs is not only related to the local primary network information but also affected by the global primary network running/traffic information, *e.g.* the traffic flow in the primary network.

To remedy the gap, in this paper, we study the spectrum whitespace for CRNs by proposing a mathematical analysis framework based on a very novel social networking approach. Specifically, we consider two social attributes of PUs: *centrality* of PUs within the primary network and *social pattern* followed by primary activities. In graph theory and social graph analysis, centrality measures the topological importance of a vertex in a graph [17]. A node with high centrality implies it has stronger connection with other nodes in a network and is more likely to serve as an intermediate relay node for other data flows. The social pattern of PUs describes the distributions of primary activities in the time and space domains [6][7]. For instance, according to empirical studies [4][5], the active distribution of cell phone users and cellular network users (PUs) can be approximated by Gaussian distribution. By exploiting centrality and social pattern of PUs, we (i) take both local primary network information and global primary network running/traffic information into consideration, and develop a mathematical model to accurately calculate the activity distribution of PUs; and (ii) propose a mathematical framework to estimate the spectrum whitespace under different primary traffic/activity patterns. To the best of our knowledge, this is the first work applying a social networking approach on whitespace analysis.

In summary, our main contributions are as follows.

First, we study the activity distribution of PUs from a social networking perspective. Based on the *betweenness centrality* importance of each PU in the primary topology, we propose a novel *centrality score* of each PU to measure its active intensity, which jointly considers the impacts from primary topological structure as well as primary network running and traffic information. Furthermore, we propose to incorporate the social pattern of primary activities into whitespace analysis, which can improve the accuracy of whitespace estimation.

Second, based on the introduced centrality score and the social pattern of primary activities, we propose a mathematical framework to analyze the whitespace for SUs under three typical primary activity social patterns, namely *independently and identically distribution* (i.i.d.), *Poisson distribution*, and *Gaussian distribution*. In the framework, a method to conduct whitespace estimation based on primary social attributes is provided. Furthermore, both the lower and upper bounds on whitespace estimation are derived under these three distributions.

Third, in terms of the primary social attributes concerned whitespace analysis, we propose a Spectrum Altitudes based Virtual Backbone (SAVB) construction algorithm, where the *spectrum altitude* can measure the average available whitespace at each SU. In SAVB, the SUs having more available whitespace are more likely to serve as Virtual Backbone (VB) nodes, and thus the spectrum utilization efficiency can be significantly improved. In addition, we theoretically analyze the correctness and performance of SAVB.

Finally, we conduct extensive simulations to examine our social network based whitespace estimation framework and the performance of SAVB. The results indicate that our framework can accurately approximate the whitespace for SUs, and SAVB can significantly improves network throughput compared with a recently published VB construction algorithm.

The rest of this paper is organized as follows. In Section II, we summarize the related work. In Section III, we introduce the primary and secondary network model. In IV, we introduce the applied social attributes and derive an accurate primary activity distribution based on PU centrality. The spectrum whitespace analysis under different primary social patterns is presented in Section V. We give a VB construction application of our whitespace analysis in Section VI, followed by the simulation results in Section VII. Finally, we conclude our paper in Section VIII.

II. Related Work

A. Empirical Studies on Whitespace

In [2], McHenry *et al.* conducted research on Chicago spectrum occupancy measurement in all bands between 30 MHz and 3000 MHz. The measurements were taken for a two-day period and it reveals that 82.6% (about 2850 MHz) of the spectrum [30 MHz, 3000 MHz] is whitespace. This implies the licensed spectrum utilization is very inefficient and the CRN paradigm is reasonable and meaningful. Similar measurement on bands [30 MHz, 2900 MHz] is also conducted in downtown Washington, DC, where both the government and commercial spectrum usage is heavy [3]. Again, the results demonstrate that about 62% of the licensed spectrum is whitespace. Based on a dataset collected over three weeks, Willkomm *et al.* [4] analyzed the usage in cellular bands. They also discussed the implications of their results on enabling dynamic spectrum access in these bands. Recently, Chen *et al.* in [5] presented a spectrum measurement study with data concurrently collected in spectrum bands between 20 MHz and 3 GHz at four locations in China. Based on their analysis, the channel vacancy durations follow an exponential-like distribution. However, in the aforementioned empirical studies, detailed whitespace analysis for a specific secondary link in a specific time slot is not provided. In addition, a rigorous mathematical framework for analyzing spectrum whitespace in CRNs is desired.

B. Local PU Information based Whitespace Analysis

In [12], Ji *et al.* studied the unicast and broadcast scheduling issues for CRNs. They also presented a whitespace analysis method based on local information of an independently and identically distributed primary network. Similarly, when studying the data collection and broadcasting problems for CRNs, the authors in [13][14][15] analyzed the spectrum whitespace for SUs considering the local information of Poisson distributed primary networks. In [16], Huang *et al.* derived the spectrum availability for SUs based on the historical primary activity information, followed by designing a spectrum mobility aware routing algorithm.

C. Other Works

There are also some works employ the spectrum whitespace as a parameter/variable yet overlooking the details [8]-[11]. In [8], Chowdhury and Akyildiz proposed a distributed routing algorithm for CRNs, where the spectrum available probability (which is equivalent to whitespace) is defined as a pre-known parameter. In [9], Zhang *et al.* studied the opportunistic spectrum scheduling for mobile CRNs in whitespace, which is also taken as a pre-known value. In [10][11], the authors studied the joint routing and frequency-domain scheduling for CRNs. Similarly, they use a variable to model the whitespace, which is reasonable. However, if we take a CRN as an integrated system, overlooking the whitespace details, *e.g.* the available spectrum opportunity/whitespace for specific SU communication pairs, may only derive suboptimal results.

III. System Model

We consider a secondary network coexisting with a primary network deployed in a square area of size A. They share the same time, space, and spectrum.

Primary Network: The primary network consists of N Poisson distributed PUs with density λ_p, denoted by S_i $(1 \leq i \leq N)$. The transmission and interference radii of PUs are assumed to be R and R_I, respectively. Therefore, the primary network can be modeled by a graph $G_p = (V_p, E_p)$, where $V_p = \{S_i \mid 1 \leq i \leq N\}$ is the node set and E_p involves all the possible links formed by the nodes in V_p.

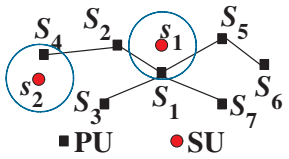

Fig. 1. A motivation example (the circles show the interference range of SUs).

Without loss of generality, we assume the primary network is connected (or intermittently connected for some time if instantaneous connection cannot be achieved). Time is assumed to be slotted with each time slot of length τ. Since we focus on whitespace analysis, we assume the licensed spectrum is an abstract spectrum with bandwidth W. Note that, this assumption is reasonable since our analysis can be directly extended to the multiple spectrum bands scenario by applying it on each spectrum band individually. Furthermore, we assume the primary network is dense scaling, *i.e.* $\sqrt{A}\cdot\sqrt{\frac{\log N}{N}} = O(1)$. Equivalently, we assume $A = \frac{cN}{\log N}$, where c is an adjustable user-defined constant value.

Secondary Network: The secondary network consists of n randomly deployed SUs, denoted by s_i $(1 \leq i \leq n)$. The transmission and interference radii of SUs are r and r_I, respectively. Consequently, the secondary network can also be modeled by a graph $G_s = (V_s, E_s)$, where $V_s = \{s_i \mid 1 \leq i \leq n\}$ is the node set and E_s is the set of links formed by the nodes in V_s. We also assume the secondary network G_s is connected and dense scaling, *i.e.* $\sqrt{A}\cdot\sqrt{\frac{\log n}{n}} = O(1)$.

Note that, for a secondary communication pair s_i and s_j such that $D(s_i, s_j) \leq r$, a spectrum opportunity exists from s_i to s_j if (i) no PU is receiving within the disk centered at s_i with radius r_I, denoted by $\odot(s_i, r_I)$ and (ii) no PU is transmitting within the disk $\odot(s_j, R_I)$. Evidently, the spectrum opportunity in the secondary network is asymmetric, *i.e.* a spectrum opportunity exists from s_i to s_j does not imply a spectrum opportunity from s_j to s_i.

IV. PUs' Social Attributes

In this section, we discuss PUs' social attributes which can be employed in accurate whitespace analysis for CRNs.

A. Betweenness Centrality

The existing works based on local primary topological information to estimate whitespace may induce inaccuracy even in data homogenous CRNs, where all the PUs have the same data generation rate. This is because the network running status and traffic information are overlooked. A motivation example is shown in Fig.1, which shows a CRN consisting of 7 PUs and 2 SUs. Now, suppose each PU randomly generates a data packet to be transmitted to another PU with some probability. Intuitively and statistically, from Fig.1, we can see that S_1 is busier (more active from the view of SUs) than other PUs since it carries more forwarding tasks. From the graph theory/social network perspective, this is because S_1 is more crucial in the primary topological structure. Furthermore, this implies the whitespace for s_1 is less than that of s_2 although the number of PUs within their interference range is the same $(|I_i| = |I_j|)$ and all the PUs have the same data generation rate. Consequently, local primary network information based whitespace estimation may induce inaccuracy. This motivates us to find a more reasonable metric to accurately measure the activities of PUs.

We propose to exploit *betweenness centrality* to measure the activities of PUs. In graph theory and social network analysis, *centrality* is defined as a quantity measuring the topological importance of a node in a graph [17]. There are many ways to define centrality in a graph, among which the *betweenness centrality* of a node is defined as the number of the shortest paths passing this node. Similarly, for a PU S_i, we define its betweenness centrality, denoted by β_i, in the primary network as the number of the shortest data transmission paths passing S_i. In the case that multiple shortest paths exist between a pair of PUs, we only count one of them which is randomly chosen among all the shortest paths. Note that, the "shortest path" here can be defined based on the number of hops from a source to a destination. To be more general, it can also be defined based on some other metrics by assigning different weights on primary links. Therefore, the betweenness centrality definition of PUs can be very general and is meaningful.

From the definition of betweenness centrality, we can see that it (i) incorporates the overall network topological and traffic information into consideration since the data transmission paths are single/multi-hop and the paths of all the communication pairs are taken into consideration; and (ii) indicates the importance of a SU within the primary topological structure, as well as the potential traffic at this node. A PU with high betweenness centrality implies this PU is more crucial in the primary network and more likely to relay data for other communication flows. Taking Fig.1 as an example, the betweenness centralities for S_1 and S_4 are $\beta_1 = 21$ and $\beta_4 = 6$, respectively. Evidently, the betweenness centrality of S_1 is much higher than that of S_4, which confirms the intuition that S_1 is more active than S_4.

Based on the betweenness centrality of PUs, we define a novel *centrality score* for S_i, denoted by α_i, to measure how active a PU in the primary network is:

$$\alpha_i = \frac{\mathcal{E}}{C_N^2 \cdot H} \cdot \beta_i, \tag{1}$$

where $\mathcal{E}$ is the expected number of active PUs, C_N^2 indicates the number of all the shortest paths in a primary network, and H is the average number of PUs on all the shortest paths. The reasonableness of defining α_i is as follows. Since there are $\mathcal{E}$ (expected) active PUs in the primary network, the average number of active PUs on each shortest path is $\frac{\mathcal{E}}{C_N^2}$. Furthermore, since the average number of PUs on each path is H, then the weight of a PU on each shortest path to be active is $\frac{\mathcal{E}}{C_N^2 \cdot H}$. Finally, for S_i, there are β_i shortest paths passing it, which implies the active weight of S_i is $\alpha_i = \frac{\mathcal{E}}{C_N^2 \cdot H} \cdot \beta_i$.

From the definition of centrality score, we can conclude that α_i can be viewed as the active probability of S_i. On the other hand, it is still possible for $s_i > 1$ in some extreme case e.g. a star topological primary network. If the primary network consists of N PUs where S_i is the network center (star center), and the expected number of active PUs is $\mathcal{E} = \log N$, then when $N \to +\infty$, $\alpha_i > 1$. Therefore, when $\alpha_i > 1$, it is almost sure that S_i is active (transmitting/receiving data), and thus we can assume that S_i is active with probability 1. Furthermore, from the definition of centrality score, it can be viewed as that we redistribute the active weights of PUs based on their betweenness centralities in the primary network, *i.e.* the PU with high betweenness centrality has a larger active weight α_i. This is reasonable since S_i with large β_i is more likely to relay data for other traffic flows besides its own traffic.

After employing the centrality score to redistribute the active weights of PUs, the total number of primary active nodes can also be preserved as shown in the following lemma, *i.e.* we do not reduce the activities of the primary network. We omit the proof of Lemma 1 due to the space limitation.

Lemma 1: $\sum_i E[\alpha_i] = \mathcal{E}$.

Based on the definition of centrality score and Lemma 1, we can see that both the overall primary network topological structure and the primary running/traffic information are considered. Therefore, employing the centrality score metric to measure the activity distribution of PUs is more reasonable and accurate. This is also validated by the simulation results in Section VII.

B. Social Activity Pattern

Recent empirical studies show that the PUs' activities usually follow some social patterns [2]-[7]. For instance, in the primary network with cell phone users, *e.g.* AT&T users, some evident social pattern can be found in weekdays: (i) PUs are more active during time slot [10:00, 22:00] (represented in a 24-hour manner); and (ii) PUs are more active in the working area in the daytime and in the residential area in the evening/night and early morning time. Similar social patterns can be found in the cellular networks [4]. This demonstrates that PUs' activities follow some social patterns in the time and space domains. Consequently, instead of assuming the activity distribution of PUs as homogenous over time, we propose to analyze spectrum whitespace considering the social pattern of PUs' activities, which can improve the accuracy of whitespace estimation.

To be specific, as revealed in the existing empirical studies, some primary network's activity generally follows a Poisson distribution in the time domain, *e.g.* the GSM900 uplink service [5], while some others follow a Gaussian distribution, *e.g.* the primary networks consisting of cell phone users and cellular network users [4], mobile device holders [6], WiFi users [7], *etc.* In practical, the social patterns of PUs' activities are application based. Therefore, in this paper, we consider typical activity distribution patterns. Note that, the analysis framework can be extended to other activity distribution patterns directly or with some modification. The detailed analysis is shown in the next section.

V. Whitespace Analysis

In this section, we derive the spectrum whitespace for SUs based on centrality scores of PUs under different primary activity social patterns. As analyzed before, the primary activity social pattern is application dependent in practice. Therefore, we study the whitespace under three typical primary social activity patterns, namely the *independently and identically distribution* (i.i.d.) denoted by $\mathcal{I}(p)$ where p is the probability of a PU to be active in a time slot, the *Poisson distribution* denoted by $\mathcal{P}(\lambda)$ where λ indicates the activity intensity of PUs over the network deployment area, and the *Gaussian distribution* denoted by $\mathcal{N}(\mu, \sigma^2)$ where μ and σ are the *mean* and *standard deviation* of the distribution, respectively.

A. Whitespace under $\mathcal{I}(p)$

Under the i.i.d. activity pattern, each PU is active with probability p. Then, we can conclude that the expected number of active PUs during each time slot is $\mathcal{E} = Np$. It follows that the centrality score of S_i is $\alpha_i^{\mathcal{I}} = \frac{\mathcal{E}}{C_N^2 \cdot H} \cdot \beta_i = \frac{Np}{C_N^2 \cdot H} \cdot \beta_i$. Based on the definition, centrality score indicates the active weight of each PU when the primary network is running. Therefore, for a stable primary network, it is reasonable to approximate the active probability of S_i, denoted by $p_i^{\mathcal{I}}$, under the i.i.d. activity pattern as $p_i^{\mathcal{I}} = \max\{p, \min\{1, \alpha_i^{\mathcal{I}}\}\}$. From the definition of $p_i^{\mathcal{I}}$, we can see that if $\min\{1, \alpha_i^{\mathcal{I}}\} > p$, S_i takes many forwarding tasks besides its own generated traffic. In this case, we use $\min\{1, \alpha_i^{\mathcal{I}}\}$ to characterize the active probability of S_i, which is more accurate. Otherwise, if $\min\{1, \alpha_i^{\mathcal{I}}\} < p$, then the active weight of S_i during the primary network running time is small which implies S_i will not take too many forwarding tasks. Consequently, employing p to represent the active probability of S_i is more accurate. In the following, we derive the whitespace for SUs based on PUs' centrality scores and activity pattern $\mathcal{I}(p)$. Here, we assume the primary network is *non-trivial*, which means the primary network has more than one PU ($N \geq 2$).

Let P_{ij} be the shortest path between S_i and S_j and D_{ij} be the number of PUs on P_{ij}. Then, define $D = \max\{D_{ij} \mid S_i, S_j \in V_p\}$, *i.e.* D is the maximum possible number of PUs on a shortest path. Then, we have the following lemma, which indicates the lower and upper bounds of $\alpha_i^{\mathcal{I}}$. The proof of Lemma 2 is omitted due to the space limitation.

Lemma 2: $\max\{\frac{2p}{D}, \Omega(\frac{p \log N}{N})\} \leq \alpha_i^{\mathcal{I}} \leq \frac{Np}{2}$.

Based on Lemma 2, let $\alpha_{\max}^{\mathcal{I}} = \frac{Np}{2}$ and $\alpha_{\min}^{\mathcal{I}} = \max\{\frac{2p}{D}, \Omega(\frac{p \log N}{N})\}$ under $\mathcal{I}(p)$. Then, it is straightforward to obtain the lower and upper bounds of $p_i^{\mathcal{I}}$ as follows.

Lemma 3: $\max\{p, \min\{1, \alpha_{\min}^{\mathcal{I}}\}\} \leq p_i^{\mathcal{I}} \leq \max\{p, \min\{1, \alpha_{\max}^{\mathcal{I}}\}\}$.

Let $W_{uv}^t(\mathcal{I}(p))$ denote the spectrum whitespace available for the communication from SU s_u to SU s_v ($D(s_u, s_v) \leq r$) by the primary network in time slot t. Based on Lemma 3, we derive $W_{uv}^t(\mathcal{I}(p))$ and its upper and lower bounds as follows.

Theorem 1: Let $p^{\mathcal{I}}_{\max} = \max\{p, \min\{1, \alpha^{\mathcal{I}}_{\max}\}\}$ and $p^{\mathcal{I}}_{\min} = \max\{p, \min\{1, \alpha^{\mathcal{I}}_{\min}\}\}$. Then, $W^t_{uv}(\mathcal{I}(p)) = W \cdot \prod_{S_i \in \odot(s_u, r_I) \cup \odot(s_v, R_I)} (1 - p^{\mathcal{I}}_i)$ and $W \cdot (1 - p^{\mathcal{I}}_{\max})^{c_2 \log n} \leq W^t_{uv}(\mathcal{I}(p)) \leq W \cdot (1 - p^{\mathcal{I}}_{\min})^{c_1 \log n}$, where $c_1 = \min_{\xi > 0} \frac{2 \log N + \lambda_p a_{uv}(e^{\xi} - 1)}{\xi \log N}$ and $c_2 = \max_{\xi < 0} \frac{2 \log N + \lambda_p a_{uv}(e^{\xi} - 1)}{\xi \log N}$ are some constant values, and a_{uv} is the size of network deployment area covered by disk $\odot(s_u, r_I)$ or $\odot(s_v, R_I)$ (*i.e.* $a_{uv} = |(\odot(s_u, r_I) \cup \odot(s_v, R_I)) \cap A|$) which is also a constant value.

Proof: To guarantee there is a spectrum opportunity from s_u to s_v, there should not be any primary receiver within disk $\odot(s_u, r_I)$ or any primary transmitter within disk $\odot(s_v, R_I)$, *i.e.* $\forall S_i \in \odot(s_u, r_I) \cup \odot(s_v, R_I)$, S_i is inactive (since the primary network may employ an ACK based protocol, and we assume a PU is active if it is a transmitter or receiver). Then, the spectrum opportunity from s_u to s_v is $\prod_{S_i \in \odot(s_u, r_I) \cup \odot(s_v, R_I)} (1 - p^{\mathcal{I}}_i)$, followed by $W^t_{uv} = W \cdot \prod_{S_i \in \odot(s_u, r_I) \cup \odot(s_v, R_I)} (1 - p^{\mathcal{I}}_i)$.

Now, we derive the lower and upper bounds of W^t_{uv}. Let a_{uv} be the size of the network deployment area covered by $\odot(s_u, r_I)$ or $\odot(s_v, R_I)$. Evidently, $\max\{\frac{\pi r_I^2}{4}, \frac{\pi R_I^2}{4}\} \leq a_{uv} \leq \pi(r_I^2 + R_I^2)$, which implies a_{uv} is a constant value. Furthermore, let X be a random variable denoting the number of PUs in a_{uv}. Then, for $\xi > 0$ applying the Chernoff bound, we have $\Pr(X \geq c_1 \log N) \leq \min_{\xi > 0} \frac{E[e^{\xi X}]}{e^{\xi c_1 \log N}} = \min_{\xi > 0} e^{\lambda_p a_{uv}(e^{\xi} - 1) - \xi c_1 \log N} = e^{-2 \log n} \leq \frac{1}{n^2}$. According to the Borel-Cantelli Lemma, $X \geq c_1 \log N$ with probability 1. Similarly, $X \leq c_2 \log N$ with probability 1. Then, the probability of having a spectrum opportunity from s_u to s_v is lower bounded by $(1 - p^{\mathcal{I}}_{\max})^{c_2 \log n}$ and upper bounded by $(1 - p^{\mathcal{I}}_{\min})^{c_1 \log n}$, followed by $W \cdot (1 - p^{\mathcal{I}}_{\max})^{c_2 \log n} \leq W^t_{uv}(\mathcal{I}(p)) \leq W \cdot (1 - p^{\mathcal{I}}_{\min})^{c_1 \log n}$. □

B. Whitespace under $\mathcal{P}(\lambda)$

When the primary activity social pattern follows a Poisson distribution, we assume the active PUs are distributed according to a two-dimensional Poisson process with density λ, denoted by $\mathcal{P}(\lambda)$. Then, the centrality score of S_i is $\alpha^{\mathcal{P}}_i = \frac{\mathcal{E}}{C_N^2 \cdot H} \cdot \beta_i = \frac{\lambda A}{C_N^2 \cdot H} \cdot \beta_i$. Similar as the scenario that PUs' activity pattern follows $\mathcal{I}(p)$ and based on the centrality score $\alpha^{\mathcal{P}}_i$, we can approximate the active probability of S_i, denoted by $p^{\mathcal{P}}_i$, under $\mathcal{P}(\lambda)$ as $p^{\mathcal{P}}_i = \max\{\frac{c\lambda}{\log N}, \min\{1, \alpha^{\mathcal{P}}_i\}\}$, where $\frac{c\lambda}{\log N}$ (comes from $\frac{\lambda A}{N}$) is the average/expected active probability of S_i while $\alpha^{\mathcal{P}}_i$ is the active weight of S_i considering the primary network topological structure, running status, as well as the traffic information. Then, before deriving the whitespace, we have the following lemma indicating the lower and upper bounds of $\alpha^{\mathcal{P}}_i$ under $\mathcal{P}(\lambda)$. The proof of Lemma 4 is omitted due to the space limitation.

Lemma 4: $\max\{\frac{2c\lambda}{D \log N}, \Omega(\frac{2c\lambda}{N})\} \leq \alpha^{\mathcal{P}}_i \leq \frac{\lambda A}{2}$.

Let $\alpha^{\mathcal{P}}_{\max} = \frac{\lambda A}{2}$ and $\alpha^{\mathcal{P}}_{\min} = \max\{\frac{2c\lambda}{D \log N}, \Omega(\frac{2c\lambda}{N})\}$ under $\mathcal{P}(\lambda)$. Then, based on Lemma 4, it is easy to obtain the following lemma indicating the lower and upper bounds of $p^{\mathcal{P}}_i$.

Lemma 5: $\max\{\frac{c\lambda}{\log N}, \min\{1, \alpha^{\mathcal{P}}_{\min}\}\} \leq p^{\mathcal{P}}_i \leq \max\{\frac{c\lambda}{\log N}, \min\{1, \alpha^{\mathcal{P}}_{\max}\}\}$.

Let $p^{\mathcal{P}}_{\max} = \max\{\frac{c\lambda}{\log N}, \min\{1, \alpha^{\mathcal{P}}_{\max}\}\}$ and $p^{\mathcal{P}}_{\min} = \max\{\frac{c\lambda}{\log N}, \min\{1, \alpha^{\mathcal{P}}_{\min}\}\}$ be the upper and lower bounds of the active probability of PUs under $\mathcal{P}(\lambda)$. We can derive the spectrum whitespace $W^t_{uv}(\mathcal{P}(\lambda))$ and its lower and upper bounds for a secondary link from s_u to s_v in time slot t as follows.

Theorem 2: $W^t_{uv}(\mathcal{P}(\lambda)) = W \cdot \prod_{S_i \in \odot(s_u, r_I) \cup \odot(s_v, R_I)} (1 - p^{\mathcal{P}}_i)$ and $W \cdot (1 - p^{\mathcal{P}}_{\max})^{c_2 \log n} \leq W^t_{uv}(\mathcal{P}(\lambda)) \leq W \cdot (1 - p^{\mathcal{P}}_{\min})^{c_1 \log n}$, where $p^{\mathcal{P}}_i = \max\{\frac{c\lambda}{\log N}, \min\{1, \alpha^{\mathcal{P}}_i\}\}$, and c_2 and c_1 are the same as specified in Theorem 1.

Proof: Considering the number of PUs within $\odot(s_u, r_I) \cup \odot(s_v, R_I)$ is upper bounded by $c_2 \log N$ and lower bounded by $c_1 \log N$ (see the proof in Theorem 1), this theorem holds based on Lemma 5. □

C. Whitespace under $\mathcal{N}(\mu, \sigma^2)$

Now, we derive the whitespace for a secondary link under the scenario that the active probability of a PU follows a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ over a time period T. For instance, for a primary network consisting of cell phone users, the active probability of PUs generally follows a Gaussian distribution on each day (in this case, we can assume $T = 24$ hours since cell phone users follow similar pattern everyday). Now, let $f(x)$ and $F(x)$ be the *probability density function* (pdf) and the *Cumulative Distribution Function* (CDF) of $\mathcal{N}(\mu, \sigma^2)$, respectively. Then, the expected active probability p_e of a PU s_i during a time slot t ($t \in T$) is $p^t_e = \int_{(t-1)\tau}^{t\tau} f(x)dx = F(t\tau) - F((t-1)\tau) = \Phi(\frac{t\tau - \mu}{\sigma}) - \Phi(\frac{(t-1)\tau - \mu}{\sigma})$, where $\Phi(\cdot)$ is the CDF of the *standard normal distribution* $\mathcal{N}(0, 1)$.

Consequently, the centrality score of S_i under $\mathcal{N}(\mu, \sigma^2)$ in time slot t is $\alpha^{\mathcal{N}}_i = \frac{\mathcal{E}}{C_N^2 \cdot H} \cdot \beta_i = \frac{N \cdot p^t_e}{C_N^2 \cdot H} \cdot \beta_i$. Similar as the aforementioned activity pattern scenario, we can approximate the active probability of S_i, denoted by $p^{\mathcal{N}}_i$, in time slot t under $\mathcal{N}(\mu, \sigma^2)$ as $p^{\mathcal{N}}_i = \max\{p^t_e, \min\{1, \alpha^{\mathcal{N}}_i\}\}$.

Now, we derive the whitespace for a secondary link from s_u to s_v in t under $\mathcal{N}(\mu, \sigma^2)$, denoted by $W^t_{uv}(\mathcal{N}(\mu, \sigma^2))$. First, we show the lower and upper bounds of $\alpha^{\mathcal{N}}_i$ in the following lemma, which can be proven by similar techniques in Lemma 2 and Lemma 4.

Lemma 6: $\max\{\frac{2p^t_e}{D}, \Omega(\frac{p^t_e \log N}{N})\} \leq \alpha^{\mathcal{N}}_i \leq \frac{Np^t_e}{2}$.

Denote $\alpha^{\mathcal{N}}_{\max} = \frac{Np^t_e}{2}$ and $\alpha^{\mathcal{N}}_{\min} = \max\{\frac{2p^t_e}{D}, \Omega(\frac{p^t_e \log N}{N})\}$ under $\mathcal{N}(\mu, \sigma^2)$. It is straightforward to obtain the bounds of $p^{\mathcal{N}}_i$ as follows.

Lemma 7: $\max\{p^t_e, \min\{1, \alpha^{\mathcal{N}}_{\min}\}\} \leq p^{\mathcal{N}}_i \leq \max\{p^t_e, \min\{1, \alpha^{\mathcal{N}}_{\max}\}\}$.

Let $p^{\mathcal{N}}_{\max} = \max\{p^t_e, \min\{1, \alpha^{\mathcal{N}}_{\max}\}\}$ and $p^{\mathcal{N}}_{\min} = \max\{p^t_e, \min\{1, \alpha^{\mathcal{N}}_{\min}\}\}$ be the upper and lower bounds of $p^{\mathcal{N}}_i$ under $\mathcal{N}(\mu, \sigma^2)$, respectively. Then, based on Lemma 7, we can derive $W^t_{uv}(\mathcal{N}(\mu, \sigma^2))$ and its lower and upper bounds

under $\mathcal{N}(\mu, \sigma^2)$ as follows. Theorem 3 can be proved by similar techniques as in Theorem 1 and Theorem 2.

Theorem 3: $W^t_{uv}(\mathcal{N}(\mu, \sigma^2)) = W \cdot \prod_{S_i \in \odot(s_u, r_I) \cup \odot(s_v, R_I)} (1 - p^{\mathcal{N}}_i)$ and $W \cdot (1 - p^{\mathcal{N}}_{\max})^{c_2 \log n} \leq W^t_{uv}(\mathcal{N}(\mu, \sigma^2)) \leq W \cdot (1 - p^{\mathcal{N}}_{\min})^{c_1 \log n}$, where $p^{\mathcal{N}}_i = \max\{p^t_e, \min\{1, \alpha^{\mathcal{N}}_i\}\}$ is the PU active probability in t under $\mathcal{N}(\mu, \sigma^2)$, and c_1 and c_2 are the same as specified in Theorem 1.

VI. Virtual Backbone Construction

Our whitespace analysis has many potential applications in aiding the design, management, and analysis of CRNs. For instance, (i) it can be used in dynamical spectrum access algorithm design. With accurate available whitespace as preliminary knowledge, smarter spectrum access algorithms can be developed; (ii) it can be applied to routing and scheduling algorithm design. Instead of simply assuming a fixed available whitespace which may only derive suboptimal solutions, proper routing and scheduling algorithms can be designed to improve the spectrum utilization efficiency; (iii) it can guide people to design a more reasonable heterogenous secondary network in terms of available spectrum whitespace instead of deploying a homogenous secondary network, which may induce high cost; *etc.* In this section, we demonstrate one potential application of our whitespace analysis by proposing a VB construction framework with the objective of maximizing the spectrum utilization efficiency, followed by analyzing the performance of the constructed VB.

A. VB Construction Algorithm

A VB B of G_s is a subset of V_s such that (i) $\forall s_u \in V_s$, either $s_u \in B$ or $N_u \cap B \neq \emptyset$; and (ii) $G_s[B]$ (the subgraph of G_s on B) is connected. Trivially, V_s itself is a VB. However, a large size VB usually induces more storage, routing, and management cost. Therefore, as in traditional wireless networks, we would like to seek a VB with small size. On the other hand, considering the spectrum dynamics in CRNs, we also want to improve the whitespace utilization efficiency. Consequently, the SUs with more available spectrum whitespace are preferred to serve as VB nodes. Aiming at the above two objectives, we propose to study the whitespace-aware VB construction problem.

In Section V, we analyze the whitespace available for secondary links based on social scores of PUs under different primary activity social patterns. Now, we define a metric ζ_u, named *spectrum altitude*, to measure the average available whitespace at each SU s_u in a time period T as follows.

$$\zeta_u = (\sum_{t=1}^{T} \sum_{v \in N_u} W^t_{uv}(X))/(T\Delta_s), \tag{2}$$

where X is $\mathcal{I}(p)$ (respectively, $\mathcal{P}(\lambda)$ or $\mathcal{N}(\mu, \sigma^2)$) when the primary activity pattern follows the i.i.d. (respectively, Poisson distribution or Gaussian distribution), and Δ_s is the maximum degree of G_s. From the definition of spectrum altitude, we can see that a high ζ_u implies s_u has more neighbors and large available spectrum whitespace from s_u to its neighbors in N_u on average. Consequently, the s_u with a high ζ_u is more suitable for disseminating data to and gathering/aggregating data from its neighbors, *i.e.* more suitable to serve as a VB node.

Algorithm 1: SAVB.

$M = \emptyset$, $V'_s = V_s$;
while $V'_s \neq \emptyset$ **do**
  $M = M \cup \{s_u\}$ s.t. $\zeta_u = \max\{\zeta_v | s_v \in V'_s\}$;
  $\forall s_v \in V'_s$, **if** $\exists s_u \in M$ *s.t.* $s_v = s_u \vee s_v \in N_u$ **then**
    $V'_s = V'_s \setminus \{s_v\}$;
$C' = \{s_u | s_u \notin M \wedge |N_u \cap M| \geq 2\}$;
$C = \{s_u\}$ where
$s_u = \arg\max_{s_v \in C'} (\sum_{t=1}^{T} \sum_{s_w \in N_v \cap M} W^t_{uv}(X))/T$;
$C' = C' \setminus \{s_u\}$;
while $C' \neq \emptyset$ **do**
  $C = C \cup \{s_u\}$ where $s_u \in C'$ and satisfies (i)
  $s_u \in C_1 = \{s_v | N_v \cap \bigcup_{s_w \in C} (N_w \cap M) \neq \emptyset\}$; (ii)
  $s_u \in C_2 = \{s_v | (N_v \cap M) \nsubseteq \bigcup_{s_w \in C} (N_w \cap M)\}$; and
  (iii) $s_u = \arg\max_{s_v \in C_1 \cap C_2} (\sum_{t=1}^{T} \sum_{s_w \in N_v \cap M} W^t_{uv}(X))/T$;
  $\forall s_v \in C'$, **if** $(N_v \cap M) \subseteq \bigcup_{s_u \in C} N_u$ **then**
    $C' = C' \setminus \{s_v\}$;
return $B = M \cup C$;

Based on the Spectrum Altitudes (SA) of SUs, we design a VB construction algorithm, denoted by SAVB, as shown in Algorithm 1. In SAVB, $|\cdot|$ denotes the cardinality of a set. The basic idea of SAVB is as follows. We first seek a Maximal Independent Set (MIS) $M \subseteq V_s$ for G_s, which is defined as $\forall s_v \in V_s$, either $s_v \in M$ or $N_v \cup M \neq \emptyset$. When determining M, a greedy strategy is employed with the criteria of maximum spectrum altitude. Once the s_u with the maximum ζ_u in V'_s is included in M, then s_u and all its neighbors are removed from V'_s. According to this process, the SUs with high spectrum altitudes are more likely to be included in M. After finding an MIS M, a connector set C is seeked to connect the independent nodes in M. When determining the connector set C, SAVB at each step selects the s_u satisfying (i) $s_u \in C_1 = \{s_v | N_v \cap \bigcup_{s_w \in C} (N_w \cap M) \neq \emptyset\}$; (ii) $s_u \in C_2 = \{s_v | (N_v \cap M) \nsubseteq \bigcup_{s_w \in C} (N_w \cap M)\}$; and (iii) $s_u = \arg\max_{s_v \in C_1 \cap C_2} (\sum_{t=1}^{T} \sum_{s_w \in N_v \cap M} W^t_{uv}(X))/T$. From this criteria, we can see that condition (i) guarantees s_u can connect to $G_s[C \cup (\bigcup_{s_w \in C} (N_w \cap M))]$, *i.e.* the SUs in M that have already been connected by the connectors in the current C; condition (ii) guarantees that including s_u in C can make more independent nodes be connected, *i.e.* make independent nodes in $\bigcup_{s_w \in C} (N_w \cap M)$ and independent nodes

in $(N_u \cap M) \setminus \bigcup_{s_w \in C} (N_w \cap M)$ to be connected; and condition (iii) indicates the SU that can connect more independent SUs in M and with more spectrum whitespace to the independent SUs on average has priority to be chosen. Note that, to remove redundant connectors, if a SU can only connect some independent nodes which have already been connected in M, it will be removed from the candidate set C' without further consideration. The correctness proof and performance analysis of SAVB is given in the following subsection.

B. Correctness Proof and Performance Analysis of SAVB

First, we show the correctness of SAVB in Theorem 4.

Theorem 4: SAVB derives a VB for G_s within finite time.

Proof: SAVB consists of two phases where an MIS M is determined in the first phase and a connector set C is identified to connect the SUs in M in the second phase. The first phase terminates in polynomial time since $|M| \leq |V_s| = n$ and one s_u is added to M in each iteration. Similarly, the second phase to identify C can also terminate in polynomial time since $|C| \leq |C'| \leq |V_s \setminus M| \leq n$ and one s_u is added to C in each iteration. Consequently, SAVB terminates in finite time. In addition, B returned by SAVB is a VB. The reasons are as follows. From the first phase of SAVB, it is straightforward that $\forall s_v \in V_s$, either $s_v \in M$ or $N_v \cap M \neq \emptyset$. From the second phase of SAVB, if $\exists s_w \in M \wedge s_w \notin G_s[C \cup (\bigcup_{s_v \in C} (N_v \cap M))]$, then as long as G_s is connected (which is an assumption), we can always find some $s_u \in C'$ satisfying the three conditions in line 10 of Algorithm 1 and add that s_u to C. Since both G_s and $G_s[C' \cup M]$ (C' is specified in line 6 of Algorithm 1) is connected, the obtained $G[C \cup M]$ is connected. □

From Algorithm 1, both the independent nodes and the connectors in B are likely to be SUs with high spectrum altitudes. Therefore, letting all the nodes in B to serve as a VB to take more data forwarding, disseminating, and gathering tasks is reasonable and can improve spectrum utilization efficiency. Now, we analyze the size of B. First, we introduce a typical geometric property of disk packing as follows.

Lemma 8: (Groemer Inequality [18]) Suppose that $\mathcal{C}$ is a compact convex set and $\mathcal{U}$ is a set of points with mutual distance of at least one. Then, $|\mathcal{C} \cap \mathcal{U}| \leq \frac{2area(\mathcal{C})}{\sqrt{3}} + \frac{peri(\mathcal{C})}{2} + 1$, where $area(\mathcal{C})$ and $peri(\mathcal{C})$ are the area and perimeter of $\mathcal{C}$ respectively.

Based on Lemma 8, we can derive the upper bound of $|M|$ as follows. Due to the space limitation, the proof of Lemma 9 is omitted.

Lemma 9: $|M| \leq \frac{2A}{\sqrt{3}r^2} + \frac{2\sqrt{A}}{r} + 1$.

Based on Lemma 9, we can prove the following theorem which indicates the lower and upper bounds of B.

Theorem 5: $|M| + \lceil \frac{|M|-1}{4} \rceil \leq |B| \leq \frac{4A}{\sqrt{3}r^2} + \frac{4\sqrt{A}}{r} + 1$.

Proof: From SAVB, whenever a new connector is included in C, at least one independent node in M is connected to $G_s[C \cup (\bigcup_{s_v \in C} (N_v \cap M))]$. Furthermore, the first connector added to C connects at least two nodes in M. Therefore, we have $|C| \leq |M| - 1$, followed by $|B| \leq 2|M| - 1 \leq \frac{4A}{\sqrt{3}r^2} + \frac{4\sqrt{A}}{r} + 1$. On the other hand, starting from any node in M, whenever a new connector is included in C, at most 4 independent nodes can be connected to $G_s[C \cup (\bigcup_{s_v \in C} (N_v \cap M))]$ based on the geometry property of independent nodes (which can be proven by applying Lemma 8 to show the upper bound of the number of points with mutual distance of at least r within a disk of radius r). Hence, we have $|C| \geq \lceil \frac{|M|-1}{4} \rceil$, followed by $|B| \geq |M| + \lceil \frac{|M|-1}{4} \rceil$. □

VII. Simulations

In this section, we validate our betweenness centrality based and primary activity social pattern concerned whitespace estimation, as well as the proposed spectrum altitudes based VB construction algorithm SAVB. We consider a randomly deployed secondary network coexisting with a primary network. Time is assumed to be slotted with each time slot normalized to 1. The primary bandwidth is normalized to $W = 100$. We define the *SU density*, denoted by ρ, as the number of SUs within a unit area, *i.e.* $\rho = n/A$. All the other simulation settings are the same as defined in Section III and the detailed system parameters are specified in each group of simulations. In addition, each group of simulations is repeated for 100 times and the results are the average values.

For SAVB, we examine its unicast throughput under different primary activity distributions. The compared algorithm is a recently published VB based unicast algorithm US [12]. In US, a VB is first constructed on top of a cell-based network partition. Then, the unicast is carried out on the VB. Here, we compare SAVB with the VB construction in [12] by employing the same scheduling method in [12] for fairness.

A. Social Network based Whitespace Estimation

To examine the Social Network based Whitespace Estimation (SNWA) method, we consider a primary network having the same topological structure as shown in Fig.1 deployed in a square area of size $A = 10 \times 10$. The locations of PUs are as follows: $S_1(5.5, 5.5)$, $S_2(4.5, 7.5)$, $S_3(2.5, 2.5)$, $S_4(2.5, 8.0)$, $S_5(6.5, 8.0)$, $S_6(8.5, 9.0)$, and $S_7(8.5, 3.0)$. In the primary network, we assume each PU randomly initiates a data communication session with another randomly selected PU destination in each time slot. The data generation rate (equivalent to PU activity) is specified under each primary activity distribution. The secondary network consists of 100 nodes each with coordinates (x, y) where x, y are positive integers and $1 \leq x, y \leq 10$. The SUs are assigned IDs with $1 \leq i \leq 100$. We examine our whitespace estimation under the circumstances where the primary activity social pattern follows i.i.d., Poisson distribution, and Gaussian distribution, respectively. Particularly, for the Gaussian distribution, we employ the user behaviors in the MIT Reality data trace [6] to model the activities of PUs. As shown in [6], the users' (which can be viewed as PUs) behaviors in MIT Reality generally follow a Gaussian distribution $\mathcal{N}(10.5, 3.6^2)$ (24-hour notation manner) during each per-day-based periodical cycle. The simulation results are shown in Fig.2(a)-(f), where SNWE

represents our social network based whitespace estimation scheme and Actual denotes the statistically actual available whitespace at each SU.

Fig.2(a) and Fig.2(b) show the average whitespace at each SU of Actual and SNWE under the i.i.d. $\mathcal{I}(p)$, where (a) shows the whitespace when $p = 0.05$ and (b) shows the whitespace at SU $s_0(0,0)$ and $s_{55}(5,5)$ when p increases. From Fig.2(a), we can see that our estimation is a little lower than the actual average whitespace at almost all the SUs. This is because we take PU active probability as the larger one between p and the centrality score $\alpha_i^{\mathcal{I}}$, thus our centrality score based whitespace estimation is conservative. The benefit of such a conservative estimation can help to reduce the chance of SUs causing unacceptable interference to PUs. Fortunately, our whitespace estimation is very close to the actual situation. On average, the difference between SNWE and Actual is less than 2.7%. As can be seen in Fig.2(b), the whitespace at $s_0(0,0)$ and $s_{55}(5,5)$ decreases when p increases. The reason comes from the fact that a large p implies more primary activities followed by fewer spectrum opportunities for SUs. From Fig.2(b), we can also see that the whitespace at $s_0(0,0)$ is larger than that of $s_{55}(5,5)$. This is because the primary activities around $s_{55}(5,5)$ are heavier than those around $s_0(0,0)$, which can be derived from the topological structure of the primary network.

Fig.2(c) and Fig.2(d) show the whitespace of Actual and SNWE under the Poisson distribution $\mathcal{P}(\lambda)$, where (c) shows the whitespace when $\lambda = 0.01$ and (d) shows the whitespace at $s_0(0,0)$ and $s_{55}(5,5)$ when λ increases. Similar as the i.i.d. scenario, (i) our estimation is a little lower than the actual situation as shown in Fig.2(c), and the difference is within 7.9% on average; and (ii) the whitespace at $s_0(0,0)$ and $s_{55}(5,5)$ decreases when the primary active intensity λ increases. This is because fewer spectrum opportunities are left for SUs.

When the primary active probability follows the Gaussian distribution $\mathcal{N}(10.5, 3.6^2)$, the whitespace at each SU when *Time Index* $= 4$ (the time that the primary activity is lighter) and *Time Index* $= 16$ (the time that the primary activity is heavier) is shown in Fig.2(e), and the whitespace at $s_0(0,0)$ and $s_{55}(5,5)$ at different time is shown in Fig.2(f). From Fig.2(e), we can see that when the primary activity is lighter, almost the entire spectrum is available for SUs. In this case, our whitespace estimation is also the same as the actual situation. When the primary activity is heavier, our whitespace estimation demonstrates similar trend as shown in the i.i.d. and Poisson distribution. On average, the difference between SNWE and Actual is within 6.66%. From Fig.2(f), the whitespace change over time at $s_0(0,0)$ and $s_{55}(5,5)$ generally shows an "inverse" Gaussian distribution curve. This is because the active probability of PUs follows a Gaussian distribution. Same as before, more whitespace is available at $s_0(0,0)$ due to the deployment of PUs.

B. SAVB's Throughput under $\mathcal{I}(p)$, $\mathcal{P}(\lambda)$, and $\mathcal{N}(\mu, \sigma^2)$

In this subsection, we examine the performance of our SAVB in the unicast communication environment under differ-

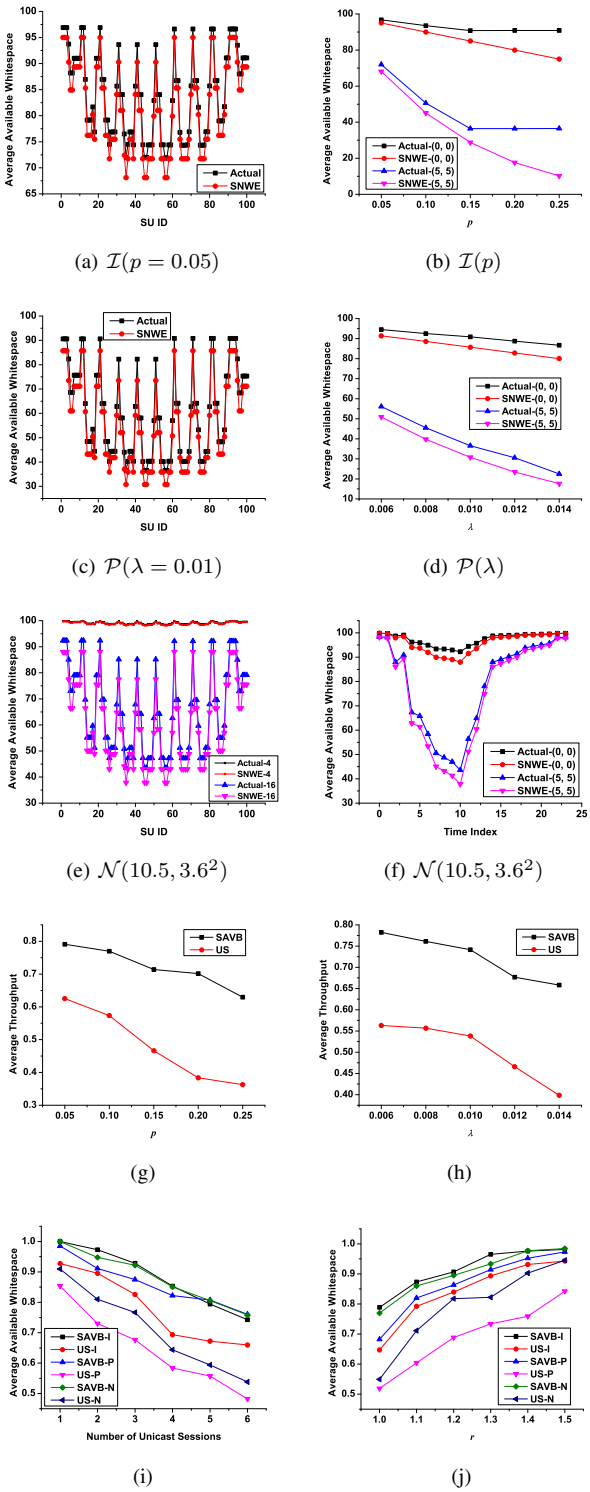

Fig. 2. Average whitespace at each SU and SAVB's throughput under $\mathcal{I}(p)$, $\mathcal{I}(p)$, and $\mathcal{N}(10.5, 3.6^2)$. The default settings are as follows: $l = 6$,$p = 0.05$, $\lambda = 0.01$, $X = Y = 10$, $\rho = 5.0$, $r = 1.0$, $r_I = 1.5$, $R_I = 3.0$, $W = 100$, $\Gamma = 10$, and $T = 1000$.

ent primary activity distributions. For simplicity, we assume the primary network has the same topological structure and settings as in the previous subsection. However, the secondary network now is randomly deployed in the primary network area with SU density ρ. Furthermore, we assume the secondary network randomly generates l unicast sessions in each simulation. For each unicast session, the source randomly generates $x \in [\frac{W}{2}, W]$ amount of traffic every Γ time slots, which is defined as the *data generation interval*. The network *throughput* is defined as the ratio between the amount of data been successfully received by unicast destinations and the amount of data been generated by unicast sources within a time window T.

We check the throughput of SAVB when the primary activity follows i.i.d. $\mathcal{I}(p)$, Poisson distribution $\mathcal{P}(\lambda)$, and Gaussian distribution $\mathcal{N}(10.5, 3.6^2)$ as shown in Fig.2(g)-(l), where "-I" denotes the i.i.d. scenario, "-P" denotes the Poisson distribution scenario, and "-N" denotes the Gaussian distribution scenario. As shown in Fig.2(g) and (h), when p and λ increases, the average throughput of SAVB and US decreases. This is because more primary activities induce fewer spectrum opportunities for SUs. Thus, the throughput of both algorithms decreases in T. Fig.2(i) shows that when the number of unicast sessions l increases, the throughput of both algorithms decreases under all the three distributions. The reasons are (i) a small l implies less communication traffic in the secondary network, and thus high success delivery ratio can be achieved; while (ii) when l becomes large, more data is generated in the secondary network followed by more interference and collisions, and thus lower data delivery ratio is induced. From Fig.2(j), we can tell when r increases, the average number of hops (transmission times) of each unicast session decreases, which implies more data can be delivered within time T. Therefore, the throughput of SAVB and US increases under all the distributions.

As indicated in Fig.2(g)-(j), SAVB outperforms US by achieving higher network throughput under all the three distributions. The reasons are as follows. US constructs a VB based on a network partition method and randomly selects the backbone node within each cell. Therefore, the available whitespace information is overlooked which may induce an inefficient VB followed by low throughput. However, SAVB constructs a spectrum altitude concerned VB, which considers the primary topological structure, the primary activity pattern, as well as the primary running and traffic status. Additionally, the available whitespace at each SU is considered during the VB construction process. Consequently, spectrum utilization efficiency is significantly improved in SAVB.

VIII. Conclusion

In this paper, we study whitespace analysis and VB construction for CRNs from a novel social network perspective. First, based on the betweenness centrality importance of each PU in the primary network, we define a novel centrality score for each PU to accurately measure its active intensity. Subsequently, based on centrality scores of PUs, we propose a mathematical framework to derive the whitespace for SUs under three typical primary activity social patterns. The lower and upper bounds of our whitespace estimation are also provided under different primary activity distributions. Third, in terms of our primary social attributes concerned whitespace estimation, we propose a Spectrum Altitudes based Virtual Backbone (SAVB) construction algorithm which can significantly improve spectrum utilization efficiency. Finally, extensive simulations are conducted to validate the whitespace analysis and the VB construction algorithm. The simulation results indicate that the social network based whitespace estimation can accurately approximate the whitespace for SUs, and SAVB can achieve higher network throughput compared with a recently published VB construction algorithm.

Acknowledgment

This work was partly supported by the National Science Foundation (NSF) under grant No. CNS-1252292.

References

[1] FCC, ET Docket No 03-237 Notice of Proposed Rule Making and Order, December 2003.

[2] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, Chicago Spectrum Occupany & Analysis and a Long-Term Studies Proposal, TAPAS 2006.

[3] M. McHenry, Spectrum White Space Measurements, New America Foundation Broadband Forum, 2003.

[4] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, Primary User Behavior in Cellular Networks and Implications for Dynamic Spectrum Access, IEEE Communications Magazine, 2009.

[5] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li, Mining Spectrum Usage Data: a Large-scale Spectrum Measurement Study, Mobicom 2009.

[6] N. Eagle and A. Pentland, Reality Mining: Sensing Complex Social Systems, Personal and Ubiquitous Computing, Vol. 10, No. 4, pp. 255-268, 2006.

[7] M. McNett and G. Voelker, Access and Mobility of Wireless PDA Users, ACM SIGMOBILE Mobile Computing and Communications Review, 2005.

[8] K. R. Chowdhury and I. F. Akyildiz, CRP: A Routing Protocol for Cognitive Radio Ad Hoc Networks, JSAC 2011.

[9] L. Zhang, K. Zeng, and P. Mohapatra, Opportunistic Spectrum Scheduling for Mobile Cognitive Radio Networks in White Space, WCNC 2011.

[10] Y. Shi, Y. T. Hou, S. Kompella, and H. D. Sherali, Maximizing Capacity in Multihop Cognitive Radio Networks under the SINR Model, IEEE Transactions on Mobile Computing, Vol. 10, No. 7, pp. 954-967, 2011.

[11] Y. T. Hou, Y. Shi, and H. D. Sherali, Spectrum Sharing for Multi-hop Networking with Cognitive Radios, IEEE Journal on Selected Areas in Communications, Vol. 26, No. 1, pp. 146-155, 2008.

[12] S. Ji, A. S. Uluagac, R. Beyah, and Z. Cai, Practical Unicast and Convergecast Scheduling Schemes for Cognitive Radio Networks, Journal of Combinatorial Optimization (JCO), Vol. 26, No. 1, pp(s): 161-177, 2013.

[13] Z. Cai, S. Ji, J. (S.) He, and A. G. Bourgeois, Optimal Distributed Data Collection for Asynchronous Cognitive Radio Networks, ICDCS 2012.

[14] Z. Cai, S. Ji, J. (S.) He, L. Wei, and A. G. Bourgeois, Distributed and Asynchronous Data Collection in Cognitive Radio Networks with Fairness Consideration, IEEE Transactions on Parallel and Distributed Systems (TPDS), 2013.

[15] S. Ji, R. Beyah, and Z. Cai, Minimum-Latency Broadcast Scheduling for Cognitive Radio Networks, IEEE SECON 2013.

[16] X. Huang, D. Lu, P. Li, and Y. Fang, Coolest path: spectrum mobility aware routing metrics in cognitive ad hoc networks, ICDCS 2011.

[17] Y. Zhu, B. Xu, X. Shi, and Y. Wang, A Survey of Social-based Routing in Delay Tolerant Networks: Positive and Negative Social Effects, IEEE Communications Surveys and Tutorials, Vol. 15, No. 1, pp. 387-401, 2013.

[18] H. Groemer, Über die Einlagerung von Kreisen in einen konvexen Bereich, Math. Z., Vol. 73, pp. 285-294, 1960.