

Graph-Based Joint User-Centric Overlapped Clustering and Resource Allocation in Ultra Dense Networks

Yan Lin, *Student Member, IEEE*, Rong Zhang, *Senior Member, IEEE*, Chunguo Li, *Senior Member, IEEE*, Luxi Yang, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—With the increase of access point (AP) density and the exponential growth of mobile devices supported by ultra dense networks (UDNs), user-centric overlapped clustering relying on AP cooperation is becoming a promising design principle for guaranteeing the quality of service (QoS) required by each UE. This ambitious goal can be achieved by the joint design of user-centric overlapped clustering and resource allocation. In this context, both the traffic-load balancing and the limited availability of orthogonal resource blocks (RBs) have to be carefully considered in UDNs. To tackle these challenges, we formulate a joint user-centric overlapped clustering and resource allocation problem with the goal of maximizing the system’s spectral efficiency (SE). To efficiently solve this problem, it is decoupled into two subproblems, which can be solved independently with the aid of our graph-theoretical framework. We first develop a novel distributed three-stage user-centric clustering solution, which is aware of the APs’ traffic-loads. Then, based on the overlapped clusters constructed, we propose a novel two-stage graph-based user-centric resource allocation scheme for mitigating the resultant inter-cluster interference, despite the limited availability of orthogonal RBs. Our numerical results confirm the benefits of the proposed framework and show that it outperforms the benchmark solutions in terms of both its per area aggregated user rate (PAAR) and user rate.

Index Terms—UDN, user-centric, overlapped clustering, resource allocation, graph theory.

I. INTRODUCTION

A. Motivation and Scopes

To meet the ever-increasing tele-traffic requirements imposed by the proliferation of mobile devices in the wake of the 5G era, the regular heterogeneous networks consisting of macro-cells and several small cells are incapable of meeting the imminent tele-traffic demands. Hence the APs are allocated more densely for increasing the capacity [1] [2]. Then, there is always an access point (AP) in the close proximity of the user equipment (UE) [3]–[5]. However, in such ultra dense networks (UDNs), the beneficial cooperation of APs relying on their joint transmission strategy is required for mitigating the deleterious effects of interference. Naturally, significant challenges arise due to the overlapped coverage of APs. Hence, a user-centric clustering based architecture [6] is introduced to deal with the above-mentioned problems in UDNs, where the APs are grouped together into clusters (or virtual cells) to

fully exploit the benefits of the joint transmission technique of [7]. Nevertheless, in order to mitigate the cluster-edge effects, where the cluster-edge UEs suffer from potential inter-cluster interference, the user-centric overlapped clustering philosophy of [8] can be employed for substantially enhancing the performance of all UEs in cooperative networks. Conceptually, the user-centric overlapped clustering design principle prompts the UEs to become actively involved in cooperation in order to guarantee the quality of service (QoS) required by each UE, regardless of their locations and allows the clusters to be overlapped. Technically, this ambitious objective may be achieved by the joint design of user-centric overlapped clustering and resource allocation.

To elaborate, in high-density scenarios, each UE is cooperatively served by a number of APs, whose coverage is potentially overlapped. This concept may be capable of reducing the cooperation overhead. However, there are at least two challenging practical issues in UDNs. Firstly, when forming AP clusters, the maximum traffic-load of each AP has to be considered and shared with the aid of load balancing; secondly, there is a limited availability of orthogonal resource blocks (RBs) that may not be sufficient for all UEs as a result of their dense congregation. Hence, it is particularly challenging to jointly design user-centric overlapped clustering and orthogonal RB allocation in UDNs under the above constraints. Although the greedy exhaustive search algorithm is capable of finding the optimal solution, the associated computational complexity tends to increase exponentially both with the number of the APs and of the UEs. Hence its complexity is excessive in UDNs. As an interesting alternative, graph theory [9] [10] has been extensively applied as a benefit of their potential simplicity. For example, bipartite graph matching methods have been applied in wireless networks [11] to solve both resource allocation problems [12]–[15] and user association problems [16]. Additionally, the classic graph coloring approach has been widely employed for resource allocation [17]–[22]. However, to the best of our knowledge, the existing graph theory based algorithms cannot be directly used for solving the realistic user-centric overlapped clustering and resource allocation problems, since they neglect the practical traffic-loads and RB constraints.

Against the above backdrop, to effectively mitigate both the intra-cluster and inter-cluster interference as well as to maximize the benefits of AP cooperation, we propose a novel framework by jointly considering the user-centric overlapped

The financial support of the 863 Program of China under Grant 2015AA01A703, that of the China Scholarship Council and that of the European Research Council’s Advanced Fellow Grant under the Beam-Me-Up project is gratefully acknowledged.

clustering and resource allocation in UDNs, subject to the traffic-load constraint and to the limited number of RBs. Theoretically, this is a non-convex and nonlinear integer programming problem, which is NP-hard. Hence, it is impractical to seek optimal solutions through exhaustive search, because it would impose an excessive computational complexity. As a result, we resort to decoupling the joint problem into two subproblems, each of which may be solved efficiently. To be specific, the user-centric overlapped clustering subproblem is used for determining the serving AP set for any UE in order to fully exploit the benefits of cooperation, while the user-centric resource allocation subproblem is designed for mitigating the resultant inter-cluster interference in order to maximize the overall spectral efficiency (SE). As far as the user-centric overlapped clustering subproblem is concerned, the existing clustering methods have not considered the maximum affordable traffic-load constraint and the associated resource allocation problem. Therefore, in order to decouple the user-centric resource allocation subproblem, we adopt the coverage distance as the criterion of determining the association between APs and UEs. In our work, we conceive a novel three-stage distributed user-centric overlapped clustering solution. We first propose a modified Gale-Shapley (GS) technique-based algorithm [23] relying on bipartite graph matching to arrive at the first anchoring stage, which is a many-to-one matching problem. Then we continue with the exploration and the determination stages so as to complete the clustering process, where the final association actually relies on a many-to-many matching.

Next, upon the construction of user-centric overlapped clusters, we invoke graph theory to translate the underlying resource allocation subproblem into a corresponding graph coloring problem. Graph coloring has been used for solving resource allocation problems for decades. However, there are main differences between our problem and the general graph coloring problems, at least in three aspects. Firstly, in contrast to the traditional graph coloring problems that minimize the number of colors in use [24], our problem focuses on maximizing the utility, given a fixed number of colors. In other words, our problem is that of finding a solution that makes full use of the affordable number of colors for maximizing the utility. Secondly, instead of allocating channels relying on orthogonal RBs using exclusive colors [19], the vertices in our graph represent the user-centric clusters and allow the reuse of colors, since the available number of colors is lower than the number of vertices. Finally, the topology of the constructed graph representing our problem is more complex owing to the overlapped clusters in the particularly dense deployment of APs and UEs considered. Hence, the existing graph-based approaches cannot be directly applied in our problem. In order to further mitigate the inter-cluster interference and to improve the SE of UDNs, we design a novel two-stage resource allocation scheme, where the vertices are first sorted and the vertex associated with a higher degree is colored preferentially.

The main contributions of this paper can be summarized as follows:

- A novel framework is proposed for jointly designing user-

centric overlapped clustering and resource allocation in practical UDNs, which is the first attempt to carefully consider both the traffic-load and the limited number of RBs as constraints.

- With the aid of a graph-theoretical framework, two independent subproblems are formulated, and then a distributed user-centric overlapped clustering scheme and a practical low-complexity resource allocation scheme are conceived, respectively.
- The benefits of the user-centric overlapped clustering architecture are quantified, demonstrating that the proposed solution offers superior performance in terms of both its per area aggregated user rate (PAAR) and its user rate.

The rest of this paper is organized as follows. Section II introduces the system model and formulates the joint user-centric overlapped clustering and resource allocation problem. Then, Section III develops the proposed user-centric overlapped clustering solution based on bipartite graph matching and analyzes both the stability as well as the complexity. Section IV presents the proposed user-centric resource allocation solution based on graph coloring and discusses the associated complexity. Our numerical results are provided in Section V and our conclusions are offered in Section VI.

Notation: Matrices and column vectors are presented in bold capital letters \mathbf{A} and bold lower-case letters \mathbf{a} respectively. $|\mu|$ denotes the cardinality of a set μ . $\mathbb{C}^{N \times M}$ is the space of all $N \times M$ matrices with complex entries. $\binom{b}{a}$ denotes a combination of a elements taken from b different elements.

B. Related Works

Clustering design relying on AP cooperation is an important and timely problem [7], [25]–[27]. In order to satisfy each UE's QoS, regardless of their locations, the user-centric clustering design has attracted substantial research interests. In [28], the authors considered user-centric clustering, where each UE may be served by several cooperating APs only relying on channel state information (CSI) exchanged among the cooperating APs. Then, an efficient algorithm was proposed for solving the joint AP clustering and the linear beamforming design problem relying on partially shared CSI. To further investigate the user-centric overlapped clustering challenges, Kim *et al.* [8] compared it to the user-centric non-overlapped clustering scheme and designed a so-called virtual cell based beamforming scheme. It was shown that as expected, user-centric overlapped clustering outperforms its non-overlapped clustering counterpart in terms of all the edge users' throughput. As a further advance, Kang and Kim [29] put forward a user-centric overlapped clustering method by adopting anchor-based transmit precoding, which provides a superior average user throughput as well as cell-edge user throughput compared to the conventional static non-overlapped clustering. Another interesting proposal of Nie *et al.* [30] was the energy-efficient user-centric overlapped clustering solution based on tier-specific received signal strength, where the impact of fading coefficients on the clustering thresholds was highlighted. As an explicit benefit of AP cooperation, user-centric overlapped clustering has been found eminently

suitable for UDNs. For example, Garcia *et al.* [31] designed a user-centric adaptive clustering method for maximizing each AP's normalized outage capacity. It has been shown that the performance of joint transmission in UDNs relying on user-centric adaptive clustering is capable of enhancing the system performance by beneficially adapting the coordination to match each UE's specific condition. However, none of the above sophisticated solutions have considered the traffic-load constraint of each AP and the associated resource allocation problem of UDNs.

Indeed, the previous research advances reveal that combining the clustering method with appropriate resource allocation has the potential of simultaneously mitigating both the inter-cell and intra-cell interference [32]–[34]. In the context of the user-centric overlapped clustering framework, there have also been some research efforts focused on the resource allocation problem. For instance, in [24], an orthogonal training resource allocation scheme was developed based on the user-centric overlapped clustering model with the objective of minimizing the overall training overhead. The joint user-centric overlapped clustering and resource allocation problem of UDNs was investigated in [35] and [36]. More specifically, Liu *et al.* [35] considered a joint user-centric overlapped clustering and resource allocation problem in UDNs from a rather different perspective. A distributed two-step scheme was conceived for approximating the joint optimization by a pair of subproblems, which confirmed the superiority of joint user-centric overlapped clustering and resource allocation design. However, the user-centric overlapped clustering problem is based on the traffic-load information, rather than on the maximum traffic-load constraint. Additionally, the resource allocation subproblem did not consider the limited number of RBs, as a constraint. Hua *et al.* [36] studied the resource allocation problem in the context of wireless back-haul links as well as joint user-centric overlapped clustering and beamforming in the access links in order to maximize the weighted sum rate under specific back-haul resource constraints in UDNs. Nevertheless, Hua *et al.* have not considered traffic-load-aware clustering neither have they dealt with the limited availability of orthogonal RBs in UDNs, which constitute the main focus of our paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We consider a downlink two-tier heterogeneous UDN that consists of a macro BS (MBS) and a dense set of multi-antenna APs having the index set of $\mathcal{L} = \{1, \dots, L\}$. In order to provide an increased data rate and improved service control, the C/U¹ split architecture of Ge *et al.* [4] is adopted in our UDN, where the APs are deployed in a hot-spot area and the macro-cell provides always-on coverage. The over-sailing MBS controls the UE handover, while the APs are responsible for transmitting the UE data. The set of single-antenna UEs is represented by the set $\mathcal{K} = \{1, \dots, K\}$. Thus, only co-tier interference is considered in our system model, while ignoring

the cross-tier interference. We assume that all APs and UEs are uniformly distributed in the macro-cell's coverage area A , and the density of APs is comparable to or possibly even higher than that of the UEs.

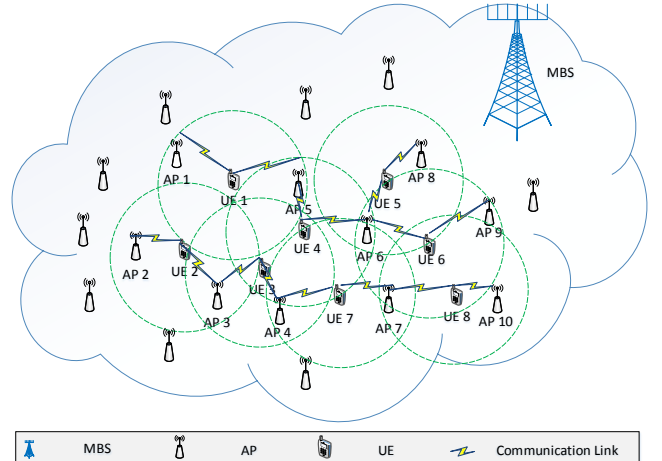


Fig. 1: An example of user-centric overlapped clustering in a heterogeneous UDN with one MBS and several APs. Each circle represents a user-centric virtual cell formed by AP clusters.

In the user-centric architecture, the MBS assists each UE in selecting multiple APs to form a cluster that performs joint transmission. In this paper, we consider the Maximum Ratio Transmission (MRT) based joint scheme. As a result of the densely distributed APs, each AP may simultaneously belong to a number of different clusters. An example of the user-centric overlapped clustering in heterogeneous UDNs relying on a single MBS and several densified APs is illustrated in Fig. 1, where each UE is served by a potentially overlapped cluster of cooperating APs and each AP cluster forms a user-centric virtual cell. For instance, observe in Fig. 1 that UE 2 is cooperatively served by AP 2 and AP 3, while AP 3 can simultaneously serve UE 2 and UE 3. In this case, the clusters or virtual cells belonging to UE 2 and UE 3 are overlapped. Typically, each UE may be associated with multiple APs relying on joint transmission, and each AP may also serve multiple UEs using orthogonal RBs, where the maximum number of UEs that AP j could serve, i.e. the AP's traffic-load, is denoted by S_j . In practice, if the AP is not associated with any UE, it is assumed to be inactive, where the AP-to-UE association can only be established, if they are within the coverage distance threshold d_t . All APs share the same spectrum, which is partitioned into multiple orthogonal RBs of the set of $\mathcal{N} = \{1, \dots, N\}$. Considering that the number of orthogonal RBs is lower than the number of UEs in the UDN, we assume that a RB can be reused for serving more than one cluster and that only one RB can be assigned to each cluster².

²Note that each cluster also could be allocated a certain number of RBs and all APs within that cluster will reuse that set of RBs via joint transmission. Furthermore, these RBs may be (partially) reused across different clusters. In this case, we will assume that a certain number of RBs constitute a group of RBs and then allocate these RBs as a group.

¹C-plane (control plane) and U-plane (user plane).

Let us define the network as the sets and variables of

- \mathcal{C}_k : the serving AP cluster of UE k ;
- \mathcal{I}_k : the set of interfering UEs, which use the same RB as \mathcal{C}_k ;
- M_j : the number of antennas at AP j ;
- p_j : the total transmit power of AP j ;
- $\mathbf{g}_{j,k,n}^H \in \mathbb{C}^{1 \times M_j}$: the normalized downlink channel vector of the link spanning from AP j to UE k on RB n ;
- $\mathbf{w}_{j,k,n} \in \mathbb{C}^{M_j \times 1}$: the normalized beamforming vector used by AP j to UE k on RB n ;
- $q_{j,k,n}$: the data symbol transmitted from AP j to UE k on RB n satisfying $\mathbb{E}\{|q_{j,k,n}|^2\} = 1$;
- $z_{k,n}$: the additive white Gaussian noise (AWGN) of RB n at UE k .

B. Channel Model

In general, the downlink channel gains $\{\mathbf{g}_{j,k,n}^H\}$ between AP j and UE k on RB n include both the small-scale fading $\{\mathbf{h}_{j,k,n}^H\}$ between AP j and UE k on RB n , as well as the path loss $\{\xi_{j,k}\}$ and shadow fading $\{\zeta_{j,k}\}$ between AP j and UE k , modelled as $\mathbf{g}_{j,k,n}^H = \mathbf{h}_{j,k,n}^H \sqrt{\xi_{j,k} \zeta_{j,k}}$. According to [2], the main differences between regular cellular networks and UDNs manifest themselves both in terms of the path loss model and spatially correlated channel model, plus the AP to UE density. First of all, as for the large-scale fading model, the standard inverse-power path loss model is not capable of accurately characterizing UDNs [37] [38]. Hence, in this paper we follow the practical two-segment 3GPP path loss model of [39], where the path loss includes both the line-of-sight (LoS) and the non-line-of-sight (NLoS) transmissions with a certain probability. Let $d_{j,k}$ represent the 'two-dimensional distance' (termed as coverage distance hereafter) between AP j and UE k , while ψ denotes the absolute antenna height difference between any AP-UE pair. Then we consider the average path loss between AP j and UE k , which is based on the following function

$$\xi_{j,k} = \begin{cases} D_L z_{j,k}^{-\theta_L}, & \text{LoS Prob: } Pr_L(z_{j,k}), \\ D_{NL} z_{j,k}^{-\theta_{NL}}, & \text{NLoS Prob: } 1 - Pr_L(z_{j,k}). \end{cases} \quad (1)$$

Herein, $z_{j,k} = \sqrt{d_{j,k}^2 + \psi^2}$ is the 'three-dimensional distance' between AP j and UE k , while D_L and D_{NL} denote the LoS and NLoS path losses at a unit reference distance, respectively. Moreover, θ_L and θ_{NL} represent the LoS and NLoS path loss exponents, respectively. Additionally, the LoS path is partitioned into two segments as follows

$$Pr_L(z_{j,k}) = \begin{cases} 1 - 5\exp(-\phi_1/z_{j,k}), & 0 < z_{j,k} \leq \bar{d}, \\ 5\exp(-z_{j,k}/\phi_2), & z_{j,k} > \bar{d}, \end{cases} \quad (2)$$

where ϕ_1 , ϕ_2 and \bar{d} are the shape parameters ensuring the continuity of $Pr_L(z_{j,k})$.

Additionally, in terms of the small-scale fading model between APs and UEs, we assume that there is a high probability of LoS between APs and UEs in UDNs. Hence, the Rician fading channel model may be more appropriate than Rayleigh

fading in this scenario. Therefore, we adopt the distance-dependent Rician small-scale fading model used in [2], where the distance-dependent K-factor is modelled as $K_R = \frac{Pr_L}{1-Pr_L}$ with Pr_L being the probability of LoS transmission. Bearing in mind that a high spatial correlation of the downlink channels between APs and UEs may exist in practical UDNs, we model the Rician fading channel vectors composed of two parts as [40],

$$\mathbf{h}_{j,k,n}^H = \sqrt{K_R/(K_R+1)} \bar{\mathbf{h}}_{j,k,n}^H + \sqrt{1/(K_R+1)} \tilde{\mathbf{h}}_{j,k,n}^H (\boldsymbol{\Omega}_j)^{1/2}, \quad (3)$$

where $\bar{\mathbf{h}}_{j,k,n} \in \mathcal{C}^{M_j \times 1}$ is the deterministic LOS component, satisfying $|\bar{\mathbf{h}}_{j,k,n}|^2 = 1$, while $\tilde{\mathbf{h}}_{j,k,n} \in \mathcal{C}^{M_j \times 1}$ is the complex-valued Gaussian random vector with identically and independently distributed (i.i.d.) entries of zero mean and unit variance. Furthermore, $\boldsymbol{\Omega}_j \in \mathcal{C}^{M_j \times M_j}$ represents the transmitter's spatial correlation matrix at AP j , and $\tilde{\mathbf{h}}_{j,k,n}$ denotes the Rayleigh fading channel with independent fading coefficients between the downlink transmitter of AP j to UE k on RB n . For simplicity, we use the popular exponential correlation model [41] formulated as $[\boldsymbol{\Omega}_j]_{u,v} = \rho_{ss}^{|u-v|}$, $u, v = 1, \dots, M_j$, where $0 \leq \rho_{ss} \leq 1$ is the Rayleigh fading correlation coefficient, which is assumed to be the same for different APs.

Furthermore, in these UDNs, the shadowing correlations cannot be neglected, since dense APs are prone to correlated shadow fading, which degrades the performance due to the associated diversity gain reduction [2] [42] [43]. To this end, we also consider the practical correlated shadow fading model adopted in [39], wherein the shadow fading $\{\zeta_{j,k}\}$ expressed in dB can be modelled as a zero-mean Gaussian random variable with a standard deviation of σ_{sd} and the correlation coefficient between any pair of APs is denoted by $0 \leq \rho_{sd} \leq 1$.

C. SINR

Let $\mathbf{X} = [x_{j,k}]$ having $(L \times K)$ elements be the association matrix, which is defined as

$$x_{j,k} = \begin{cases} 1, & \text{if AP } j \text{ is in the cluster } \mathcal{C}_k, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Let $\mathbf{Y} = [y_{k,n}]$ having $(K \times N)$ elements be the RB allocation matrix, which is defined as

$$y_{k,n} = \begin{cases} 1, & \text{if RB } n \text{ is allocated to the cluster } \mathcal{C}_k, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

In our hypothesis, the total power transmitted from each AP is shared equally amongst all of its associated UEs. Then, the power transmitted from AP j to each associated UE is represented as $p_j / \sum_{\kappa \in \mathcal{K}} x_{j,\kappa}$. Upon assuming flat Rayleigh fading for each RB, the signal received at UE k on RB n is given by

$$s_{k,n} = \sum_{j \in \mathcal{C}_k} \sqrt{\frac{p_j}{\sum_{\kappa \in \mathcal{K}} x_{j,\kappa}}} \mathbf{g}_{j,k,n}^H \mathbf{w}_{j,k,n} q_{j,k,n} + \sum_{m \in \mathcal{I}_k} \sum_{i \in \mathcal{C}_m} \sqrt{\frac{p_i}{\sum_{\kappa \in \mathcal{K}} x_{i,\kappa}}} \mathbf{g}_{i,k,n}^H \mathbf{w}_{i,m,n} q_{i,m,n} + z_{k,n}, \quad (6)$$

where the first item is the useful signal received from the APs in \mathcal{C}_k by joint transmission, while the second and third items denote the inter-cluster RB-reuse-induced interference and the AWGN. The power of the AWGN at each UE on any RB is assumed to be the same, which is denoted by σ^2 . Accordingly, the signal to interference-plus-noise ratio (SINR) at UE k , when RB n is allocated is given by

$$\gamma_{k,n}(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{j \in \mathcal{C}_k} \frac{p_j}{\sum_{\kappa \in \mathcal{K}} x_{j,\kappa}} |\mathbf{g}_{j,k,n}^H \mathbf{w}_{j,k,n}|^2}{\sigma^2 + \sum_{m \in \mathcal{I}_k} \sum_{i \in \mathcal{C}_m} \frac{p_i}{\sum_{\kappa \in \mathcal{K}} x_{i,\kappa}} |\mathbf{g}_{i,k,n}^H \mathbf{w}_{i,m,n}|^2}. \quad (7)$$

Herein, \mathcal{C}_k is the function of the k -th column of \mathbf{X} , while \mathcal{I}_k is the function of \mathbf{Y} . Mathematically, we have $\mathcal{C}_k = \{j \mid x_{j,k} = 1, \forall j \in \mathcal{L}\}$ and $\mathcal{I}_k = \{m \mid \exists n \in \mathcal{N}, \text{ s.t. } y_{k,n} = y_{m,n}, \forall m \in \mathcal{K}\}$.

Again, in this paper, we employ MRT beamforming at each AP, thus the normalized beamforming vector is given by $\mathbf{w}_{j,k,n} = \mathbf{g}_{j,k,n} / \|\mathbf{g}_{j,k,n}\|$. Due to having overlapped and densified clusters, different UEs in their own clusters sharing the same AP should be assigned mutually orthogonal RBs for mitigating the inter-cluster interference.

D. Problem Formulation

Accordingly, the user rate of UE k is represented by

$$R_k(\mathbf{X}, \mathbf{Y}) = \sum_{n=1}^N y_{k,n} \log_2 [1 + \gamma_{k,n}(\mathbf{X}, \mathbf{Y})]. \quad (8)$$

The aim of this paper is to maximize the system-wide SE by jointly designing the user-centric overlapped AP clusters with mutually orthogonal RBs under the constraint of having a limited number of RBs ($K > N$) for our UDN. Hence, the optimization problem can be formulated as

$$(P0): \max_{\mathbf{X}, \mathbf{Y}} \sum_{k=1}^K \sum_{n=1}^N y_{k,n} \log_2 [1 + \gamma_{k,n}(\mathbf{X}, \mathbf{Y})] \quad (9a)$$

$$\text{s.t. } x_{j,k} = \{0, 1\}, \forall j, \forall k, \quad (9b)$$

$$\sum_{k=1}^K x_{j,k} \leq S_j, \forall j, \quad (9c)$$

$$y_{k,n} = \{0, 1\}, \forall k, \forall n, \quad (9d)$$

$$\sum_{n=1}^N y_{k,n} = 1, \forall k, \quad (9e)$$

where eq.(9b) and eq.(9c) indicate that any given AP j can only serve at most S_j UEs. Eq.(9d) and eq.(9e) imply that each UE can only be assigned a single RB. It is noteworthy that $\gamma_{k,n}$ in eq.(9a) is related not only to the association matrix \mathbf{X} , but also to the allocation matrix \mathbf{Y} due to the serving AP cluster \mathcal{C}_k and the interfering UE cluster \mathcal{I}_k contaminating UE k .

It can be observed that (P0) is a non-linear binary integer programming problem, which has no exact solution in polynomial time. More explicitly, owing to the high densities of APs and UEs, the exhaustive search has an excessive computational complexity, which is hence impractical. Since

the subproblems of user-centric overlapped clustering and user-centric resource allocation are inter-dependent, their joint optimization will impose an extremely high computational complexity, especially in the scenario of UDNs having high AP and UE densities. Hence, we may decouple the problem into two subproblems: user-centric overlapped clustering and user-centric RB allocation. To be specific, the user-centric overlapped clustering process is used for determining the serving AP set for any UE in order to fully exploit the benefits of cooperation, while the user-centric RB allocation is designed for mitigating the resultant inter-cluster interference in order to maximize the overall SE. In the following, we will propose low-complexity user-centric solutions for both problems with the aid of a graph-theoretical framework, which will be elaborated on in the next sections.

III. USER-CENTRIC OVERLAPPED CLUSTERING SOLUTION

To reduce the coordination overhead imposed, each UE is cooperatively served by a limited number of potentially overlapped APs. In UDNs, the user-centric overlapped cluster formation is dynamic in nature, depending on the specific locations of APs and UEs. Motivated by this, we propose a coverage distance based user-centric overlapped clustering scheme using bipartite graph matching in this section. Then we analyze the stability and complexity of the proposed solution and compare it to the optimal user-centric overlapped clustering solution.

A. Coverage Distance Based User-Centric Overlapped Clustering

In our work, we consider realistic practical constraints when forming AP clusters, such as the maximum traffic-load of each AP and the maximum coverage distance of each AP. Since the coverage distance metric facilitates a clean and direct technique of determining the topology, we will adopt it as the criterion of determining the association relationship between UEs and APs. Although the coverage distance-based criterion is not related to the channel conditions, its benefits are compelling. Firstly, the computational complexity is significantly reduced, because the user-centric overlapped clustering subproblem can be decoupled from the resource allocation subproblem, which will be discussed in Section IV. As shown in Fig. 2(a), in a nutshell, the proposed coverage distance based user-centric overlapped clustering solution may be divided into the following three stages.

1) *Stage I: Anchoring Stage*: In this stage, each UE can be assigned at most one anchor AP. However, given that AP j can serve no more than $S_j (> 1)$ UEs according to the constraint (9c), this process corresponds to a *many-to-one matching* problem of classic bipartite graph theory. It is worth noting that choosing the anchor AP constitutes a *one-to-one matching* problem if $S_j = 1, \forall j$, which can be solved by the GS algorithm. Inspired by it, we propose a modified GS-based algorithm to assign anchor APs to more than one UE for solving the underlying many-to-one matching problem.

To complete this matching process, the preferences of the players (i.e. UEs and APs) have to be well defined. As

mentioned above, we choose the coverage distance $\mathbf{d} = \{d_{j,k}\}$ between the APs and UEs as the preference profile for both the APs and UEs. The modified GS-based algorithm invoked for choosing the anchor AP is illustrated in Algorithm 1-Stage I. As described in Algorithm 1, each UE first sends a request to the AP with its first preference (i.e. the nearest AP) and this AP collects all requests. If the number of requests for AP j is higher than the maximum traffic-load S_j , AP j will have to sort these UEs according to its own preference (i.e. the coverage distance) and then only satisfy the top S_j requests. The remaining UEs' requests are thus rejected and those UEs will continue to send their requests based on their second preference. At the AP side, its candidate set determined during the previous round would be updated accordingly, should more preferable UE requests have appeared. The process implies that the best UE candidate set is retained throughout all rounds of handshake. The iterations are repeated, until each UE is assigned to an anchor AP or rejected by all preferred APs. Considering the scenario of Fig. 1 as an example, we assign the anchor AP for each UE through the proposed modified GS-based algorithm, as shown in Fig. 2(a)-(i).

2) *Stage II: Exploration Stage:* After the above pre-association stage, if any under-loaded AP could be associated with more UEs, it will explore its potential association with other UEs within its coverage of d_t , regardless of its traffic-load limit and any preference. In the example of Fig. 1, all APs explore their potential association within the coverage of d_t , as shown by the dashed line in Fig. 2(a)-(ii).

3) *Stage III: Confirmation Stage:* With the above exploration association results, each overloaded AP will first sort these UEs according to the preferences and then determine the final association in the sorted set, until reaching its traffic-load limit. Hence, the user-centric AP clusters are formed for each UE and the resultant association constitutes a many-to-many matching. As an example of Fig. 1, we assume that the maximum traffic-load limit is set to 3 UEs for each AP. From the explored association results seen in Fig. 2(a)-(ii), we find that AP 6 is overloaded and hence it will sort the associated UEs (i.e. UE 4, UE 5 and UE 7) according to the coverage distance. Note that the pre-association result remains unchanged. Since UE 7 is farther than UE 4 and UE 5 from AP 6, we finally confirm UE 4 and UE 5 as the final association results for AP 6 and delete the pre-association relationship between UE 7 and AP 6, as shown in Fig. 2(a)-(iii).

To sum up, the three-stage distributed user-centric overlapped clustering algorithm proposed is presented in Algorithm 1.

B. Analysis of the Proposed User-Centric Overlapped Clustering Solution

In the following, we analyze both the stability and the computational complexity of the proposed solution.

1) *Stability:* Before proving the stability of the solution, we have the following definitions [11] [44]. First, formally, a many-to-one matching is stated as follows :

Algorithm 1 Proposed Coverage Distance-based User-Centric Overlapped Clustering Solution

- 1: Calculate $\{d_{j,k}\}$ for $\forall j \in \mathcal{L}$ and $\forall k \in \mathcal{K}$;
 - Stage I: Anchoring Stage (Modified GS-based algorithm)**
 - 2: Each UE $k \in \mathcal{K}$ proposes to its nearest AP within d_t which has not rejected it before.
 - 3: Each AP $j \in \mathcal{L}$ keeps all the requests until the number of requests exceeds S_j . If it exceeds, AP j keeps the S_j nearest UEs and rejects the others.
 - 4: The rejected UEs repeat step 2 and step 3 until each UE $k \in \mathcal{K}$ is assigned an anchor AP or rejected by all preferred APs.
 - 5: Set $X_{j^*,k} = 1$ for any UE k and its anchor AP j^* .
 - Stage II: Exploration Stage**
 - 6: **for** $\forall j \in \mathcal{L}$
 - 7: **if** $\sum_k X(j,k) < S_j$, **then**
 - 8: Explore to associate with other UEs within d_t ;
 - 9: Set the corresponding association elements in \mathbf{X} .
 - 10: **endif**
 - 11: **endfor**
 - Stage III: Confirmation Stage**
 - 12: **for** $\forall j \in \mathcal{L}$
 - 13: **if** $\sum_k X(j,k) > S_j$, **then**
 - 14: Sort the explored associated UEs and keep them until reaching S_j .
 - 15: Set the corresponding association elements in \mathbf{X} .
 - 16: **endif**
 - 17: **endfor**
-

Definition 1. A many-to-one matching μ is a mapping from the set $\mathcal{K} \cup \mathcal{L}$ into the set of all subsets of $\mathcal{K} \cup \mathcal{L}$, so that for every UE $k \in \mathcal{K}$ and AP $j \in \mathcal{L}$ we have:

- $|\mu(k)| \leq 1$, $\forall k$, and $\mu(k) = \emptyset$ if UE k is not matched to any AP;
- $|\mu(j)| \leq S_j$, $\forall j$, and $\mu(j) = \emptyset$ if AP j is not matched to any UE;
- $k \in \mu(j)$ if and only if $\mu(k) = j$.

In essence, a matching game is a double-sided assignment problem between two disjoint sets of players in which the players of each set would like to be matched to the players of the other set. A *preference* relationship \succ is defined as a *complete* and *transitive* binary relationship between the elements of a given set [45]. In general, we denote by \succ_m the preference relationship of player m , where $b \succ_m a$ indicates that player m prefers b to a .

Definition 2. A matching μ is blocked by a UE-AP pair (k, j) if $j \succ_k \mu(k)$ and $k \succ_j \mu(j)$.

Definition 3. A matching μ is stable if there is no UE-AP pair (k, j) , so that $\mu(k) \neq j$ where $j \succ_k \mu(k)$ and $k \succ_j \mu(j)$, i.e. not blocked by any UE-AP pair.

Proposition 1. The Stage I of Algorithm 1 is a stable assignment.

Proof: We prove the proposition by the method of contra-

diction, i.e. the pair of conditions $j \succ_k \mu(k)$ and $k \succ_j \mu(j)$ cannot be established at the same time for any $k \in \mathcal{K}$ and any $j \in \mathcal{L}$. Assume that there exists a pair (k, j) , which satisfies both $j \succ_k \mu(k)$ and $k \succ_j \mu(j)$. Due to the condition $k \succ_j \mu(j)$, a matching has already been formed between UE $\mu(j)$ and AP j based on the predefined matching rules. As a result of $\mu(j) \neq k$, we infer that UE k prefers AP $\mu(k)$ to AP j , i.e. we have $\mu(k) \succ_k j$, which contradicts to our assumption. Consequently, the matching μ leads to a stable matching, since no blocking pair exists. \square

2) *Computational Complexity*: Let the number of candidate APs within the coverage distance of d_t for a given UE k be denoted by L_k^c and the number of candidate UEs within the coverage distance of d_t for a given AP j be denoted by K_j^c . In Stage I of Algorithm 1, for each handshake iteration, the maximum complexity order is given by $\mathcal{O}(\sum_{k=1}^K L_k^c \log L_k^c + \sum_{j=1}^L K_j^c \log K_j^c)$ in terms of polynomial time. As a matter of fact, Stage I will eventually be terminated after a finite number of iterations, because the number of both UEs and APs is finite and no UE will request to be paired more than once with the same AP. In Stage II, if all APs are underloaded, they will explore the association with other UEs with the maximum complexity of $\sum_{j=1}^L K_j^c$. As for Stage III, the explored associated UEs have to be sorted for each overloaded AP, and hence it has the maximum computational complexity of $\sum_{j=1}^L K_j^c \log K_j^c$. Therefore, all in all, the maximum algorithmic complexity order of Algorithm 1 is in polynomial time $\mathcal{O}(\sum_{k=1}^K L_k^c \log L_k^c + \sum_{j=1}^L K_j^c \log K_j^c)$.

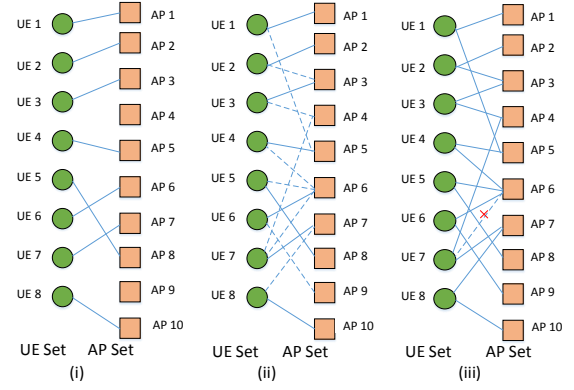
C. Optimal User-Centric Overlapped Clustering Solution

Let us now discuss the optimal user-centric solution as our benchmark, where the optimal user-centric problem (P0) is solved in two stages. First, each UE sends its requests to all APs within d_t . Then, the APs search through all possible user associations in order to compare the sum of all throughput requests received so far to the maximum traffic-load. Note that the above process of exhaustively searching for the optimal user association is intricately coupled with the associated resource allocation problem.

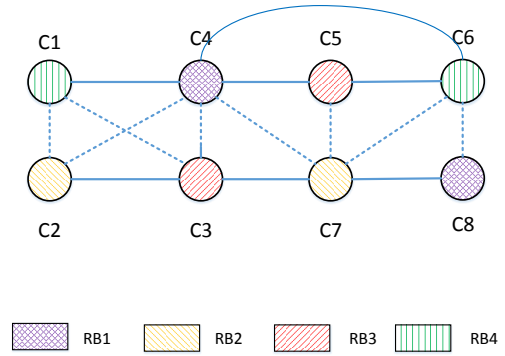
Let us now analyze the computational complexity of the optimal clustering solution. In fact, there are $\sum_{\tau=0}^{S_j} \binom{K_j^c}{\tau}$ possible user association combinations for AP l . Note that the number of UEs associated with AP j is at most $\min\{S_j, K_j^c\}$. Thus, even without considering the complexity of comparing the SE, the computational complexity order of the exhaustive search is given by

$$\mathcal{O} \left(\prod_{j=1}^L \sum_{\tau=0}^{S_j} \binom{K_j^c}{\tau} \right). \quad (10)$$

Naturally, the complexity of the optimal solution increases exponentially both with the number of APs and UEs, making it impractical in the scenario of UDNs. Additionally, all APs have to share their association information in a centralized manner. Therefore, the proposed low-complexity coverage distance-based user-centric clustering solution is much more practical in the scenario of UDNs.



(a) An example of the user-centric overlapped clustering formation procedure. (i) Anchoring Stage ; (ii) Exploration Stage ; (iii) Confirmation Stage. (Each dashed line represents the explored association and the cross between UE 7 and AP 6 represents the deleted explored association.)



(b) An example of the user-centric graph modelling and coloring. (Each solid line represents the edge formed in the overlapped clusters and each dashed line denotes the edge formed in the non-overlapped clusters.)

Fig. 2: The constructed graph model for **user-centric** clustering and resource allocation.

IV. USER-CENTRIC RESOURCE ALLOCATION SOLUTION

After forming the user-centric overlapped AP clusters, the next goal is to solve the user-centric resource allocation problem of UDNs. In the context of graph theory, each UE represents a vertex and the edge between two vertices indicates the interference between two UEs. It can be readily seen that the underlying resource allocation subproblem associated with a given user-centric cluster formation can be translated to a graph coloring problem, but this graph coloring problem cannot be solved directly by the existing graph-based techniques. However, we emphasize that the basic ideas of random coloring and sequential coloring [46] are widely adopted in various resource allocation problems for comparison, such as in [19] [47] [48] and [24], respectively. Hence, we can exploit the basic ideas of random coloring and sequential coloring in our user-centric resource allocation subproblem based on user-centric overlapped clustering UDNs. This allows us to obtain a pair of appealingly simple solutions as follows: 1) random allocation (RA): allocating random RBs to each user-centric

cluster; 2) sequential allocation (SA): sequentially assigning a limited number of RBs to each user-centric cluster according to a predefined sequence. Nevertheless, the performance of the RA and SA solutions may not be impressive since they do not take into account the channel conditions, when performing resource allocation. Therefore, in the following, we propose a novel graph-based user-centric resource allocation solution, which is constituted by two stages: 1) graph model construction and 2) graph coloring. Specifically, first the interference graph model is constructed based on the pre-defined vertices as well as edges, and then the original optimization problem is transformed into a graph coloring problem. More particularly, we note that the vertices associated with higher interference tend to degrade the system performance more gradually, hence they should be preferentially assigned orthogonal resources. Thus, a graph coloring algorithm based on two stages, namely sorting and coloring, is proposed. An example of a graph model based on Fig. 2(a) is given in Fig. 2(b), where the graph consists of 15 edges and 8 vertices associated with 4 colors.

A. Graph Model Construction

In order to construct the graph model, first we define an overlap matrix \mathbf{Z} of size $(K \times K)$ to indicate whether the clusters \mathcal{C}_k and \mathcal{C}_m overlap, where the binary flag $z_{k,m} = 1$ indicates that they do overlap and 0 that they do not. Since the value of $\sum_{j \in \mathcal{L}} x_{j,k} x_{j,m}$ ($k, m \in \mathcal{K}$) indicates the number of common APs in the clusters of \mathcal{C}_k and \mathcal{C}_m , we have

$$z_{k,m} = \begin{cases} 1, & \sum_{j \in \mathcal{L}} x_{j,k} x_{j,m} > 0, \\ 0, & \sum_{j \in \mathcal{L}} x_{j,k} x_{j,m} = 0. \end{cases} \quad (11)$$

Let the graph be denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of vertices $\{C_1, \dots, C_K\}$ corresponding to the clusters $\{\mathcal{C}_1, \dots, \mathcal{C}_K\}$ respectively and \mathcal{E} is the set of edges that connects any two vertices. The degree of the edge $\theta(e)$ is defined as the number of vertices in edge e , and for simplicity we only consider the case of $\theta(e) = 2$. The degree of the C_k -th vertex is denoted by $\vartheta(C_k)$, which equals to the number of edges involving vertex C_k . Naturally, the edge exists between two overlapped clusters, which can be indicated by the entries of \mathbf{Z} . Additionally, we also have to construct the edge between non-overlapped clusters according to the signal to interference ratio, where an edge is assigned when the SIR is below a certain threshold. To this end, let $U_{k,n}$ and $F_{m,k,n}$ denote the signal for cluster \mathcal{C}_k on RB n and the interference from cluster \mathcal{C}_m to cluster \mathcal{C}_k on RB n respectively, which may be formulated as:

$$U_{k,n} = \sum_{j \in \mathcal{C}_k} \frac{p_j}{\sum_{\kappa \in \mathcal{K}} x_{j,\kappa}} |\mathbf{g}_{j,k,n}^H \mathbf{w}_{j,k,n}|^2, \quad (12)$$

$$F_{m,k,n} = \sum_{i \in \mathcal{C}_m} \frac{p_i}{\sum_{\kappa' \in \mathcal{K}} x_{i,\kappa'}} |\mathbf{g}_{i,k,n}^H \mathbf{w}_{i,m,n}|^2. \quad (13)$$

To sum up, we can construct the graph \mathcal{G} as follows:

- First, the cluster set $\{C_1, \dots, C_K\}$ formed represents the vertices $\mathcal{V} = \{C_1, \dots, C_K\}$ in the graph \mathcal{G} . As shown in

Fig. 2(b), each vertex corresponds to the cluster formed in Fig. 2(a).

- Next, search for all the overlapped clusters and construct the edges according to the matrix \mathbf{Z} : if $z_{k,m} = 1$, an edge between vertex C_k and vertex C_m is formed. For instance, both UE 2 and UE 3 are served by AP 3, thus there is an edge between the vertices C_2 and C_3 , as shown by the solid line in Fig. 2(b).
- Finally, search for all the interferers from the non-overlapped clusters and construct the edge between vertex C_k and vertex C_m , provided that for any RB n we have:

$$\frac{U_{k,n}}{F_{m,k,n}} < \delta_g, \quad (14)$$

where $\delta_g \geq 0$ is the threshold selected to reflect the severity of the interference between two clusters. As an example of Fig. 1, we construct the edges among the non-overlapped clusters, as shown by the dashed line in Fig. 2(b).

Naturally, the value of the threshold δ_g will have an impact on the construction of the graph model. When $\delta_g = 0$, this implies that there are no edges between the non-overlapped clusters.

B. Graph Coloring Algorithm

Having constructed the graph model, a specific color in the graph \mathcal{G} corresponds to a RB, and the process of coloring vertices is equivalent to allocating orthogonal RBs for the communication links within clusters. As a matter of fact, maximizing the overall throughput can be approximated by minimizing the interference imposed on any UE by the careful exploitation of the limited number of orthogonal RBs. To this end, the vertices connected to the same edge cannot be colored by the same color and each color can be reused at least once. Therefore, based on this idea, the resource allocation subproblem becomes a graph coloring problem under a limited number of colors.

Since there is no computationally efficient technique of finding the optimal solution, we have to resort to an exhaustive search. Since we do not have sufficient RBs, the vertex having the maximum degree should be colored preferentially. This is because the vertex having a higher degree has a higher impact on the performance of the communication links in the system. Thus, we may opt for coloring the vertex having the highest degree first and then allocate other vertices to other colors in order to avoid the interference. As illustrated in Algorithm 2, the proposed solution can be divided into two stages: the sorting stage and coloring stage. To elaborate, in the sorting stage, the vertex having the maximum degree is repeatedly searched, and then the degree of the other vertices connected to it has to be reduced by one. In the following coloring stage, the first N vertices having the highest degree are colored first, and then the remaining vertices select a random color which is different from the adjacent vertices. Consider Fig. 2(b) as an example. Based upon the graph model constructed in Section IV-A, the degrees of the 8 vertices can be obtained as $\{\vartheta(C_k)\}_{k=1}^8 = \{3, 3, 4, 6, 3, 4, 5, 2\}$. Then, according to the

decreasing order of $\vartheta(C_k)$, the sorted vertex set is ordered as $C4 \rightarrow C7 \rightarrow C3 \rightarrow C6 \rightarrow C1 \rightarrow C2 \rightarrow C5 \rightarrow C8$. When $N = 4$, the first 4 vertices (i.e. C4, C7, C3, C6) are successively colored with RB1, RB2, RB3, and RB4. Then the remaining 4 vertices (i.e. C1, C2, C5, C8) choose a random color, which is different from that of the adjacent vertices, respectively. In this example, vertex C1 may be colored with RB2 and RB4. If RB4 is chosen randomly to be allocated to vertex C1, vertex C2 can only choose RB2, which is different from the colors of the adjacent vertices. Similar processes are performed also for the remaining vertices C5 and C8, which are finally allocated RB3 and RB1, respectively.

Algorithm 2 Proposed Graph-based User-Centric Resource Allocation Algorithm

- 1: **Stage I: Sorting Stage**
 - 2: Calculate the degree $\vartheta(C_k)$ for any vertex $C_k \in \mathcal{V}$ in graph \mathcal{G} .
 - 3: **repeat**
 - 4: Search for the vertex C_{k^*} with the maximum degree and put it in the sorted set \mathcal{S} .
 - 5: Set $\vartheta(C_k) = \vartheta(C_{k^*}) - 1$ for each unsorted vertex C_k connected with C_{k^*} .
 - 6: **until** All vertices in \mathcal{G} are sorted.
 - 7: **Stage II: Coloring Stage**
 - 8: Assign to the first N vertices in \mathcal{S} with N kinds of colors respectively.
 - 9: Set $C_k = N + 1$.
 - 10: **repeat**
 - 11: Choosing one color n from the set \mathcal{N} randomly;
 - 12: Decide whether all other vertices with color n connect with vertex C_k :
If no, then allocate color n to vertex C_k ; otherwise generating color n' randomly in remaining colors until find the suitable color.
 - 13: Set $C_k = C_k + 1$.
 - 14: **until** All vertices in \mathcal{G} are colored.
-

C. Analysis of the Proposed User-Centric Resource Allocation Solution

In the following, we analyze the properties of the proposed user-centric resource allocation solution including its computational complexity.

1) *Properties*: In graph theory, the chromatic polynomial counts the number of ways a graph can be colored using no more than a given number of colors. The chromatic number of a graph is the smallest number of colors required for coloring the vertices, so that no two adjacent vertices share the same color. Assuming that the *chromatic number* of graph \mathcal{G} is $\chi(\mathcal{G})$, we have the following properties [9]:

Lemma 1. *If the graph has K vertices, then $1 \leq \chi(\mathcal{G}) \leq K$.*

Lemma 2. *(Theorem 17.1 in [9]) If the largest vertex degree of the graph \mathcal{G} is $\Delta(\mathcal{G})$, then $\chi(\mathcal{G}) \leq \Delta(\mathcal{G}) + 1$.*

Lemma 1 illustrates that assigning distinct colors to distinct vertices always yields a proper coloring. Lemma 2 states that

for a greedy coloring algorithm, every graph can be colored with the aid of just one more color than the maximum vertex degree. Then, upon combining Lemma 1 and 2, we infer the following Remark 1.

Remark 1. *In the case of making full use of all RBs for maximizing the system's utility, the available number of RBs N should obey $\Delta(\mathcal{G}) + 1 \leq N < K$. In other words, if the number of RBs is less than $\Delta(\mathcal{G}) + 1$, it may be lower than the chromatic number and hence it may cause serious interference due to having an insufficient number of RBs.*

Remark 1 indicates that our proposed solution exhibits superiority under certain conditions imposed on the number of RBs. Additionally, compared to RA/SA, our proposed solution performs well in high-interference scenarios, since it is adaptive to different channel conditions.

2) *Computational Complexity*: As far as graph construction is concerned, the complexity order of edge formation is $\mathcal{O}(LNK^2)$. In Algorithm 2, the complexity of graph coloring is on the order of $\mathcal{O}(K^2)$. For comparison, the complexity of exhaustive search for the optimal solution is $\mathcal{O}(N^K)$ and the complexity of RA or SA is $\mathcal{O}(K)$. The complexity of the optimal exhaustive search will increase exponentially with the number of UEs. Hence, in UDNs, compared to the optimal exhaustive search solution, our proposed solution efficiently reduces the complexity.

V. NUMERICAL RESULTS

In this section, our aim is to characterize the performance of the proposed framework under different conditions via numerical simulations. We focus our attention on a squared area, wherein L APs and K UEs are uniformly distributed and share N orthogonal RBs. By default, L and K are both set to 100, and N is set to 50. For simplicity, the maximum traffic-load for each AP is set to be the same, denoted by S , and we set $S = 8$ by default. The deterministic LOS component characterized by $\mathbf{h}_{j,k,n} = \mathbf{1}_{M_j \times 1}$ is employed. The other simulation parameters are listed in Table I and all results are generated by 5000 Monte Carlo runs. In terms of the performance metric, we define the PAAR expressed in [b/s/Hz/km²], which quantifies the aggregated user rate (i.e. $\sum_{k \in \mathcal{K}} R_k$) normalized by the macro-cell's coverage area A .

For fair comparison, we consider the user-centric non-overlapped clustering case ($S = 1$) into consideration for comparison with our proposed user-centric overlapped clustering ($S > 1$) solution (Algorithm 1). Based on the user-centric clusters constructed, we compare the proposed user-centric resource allocation (Algorithm 2) to a pair of appealingly simple solutions (RA/SA), as well as to the following classic benchmarkers:

- UFR (Unity Frequency Reuse) [49]: allowing all user-centric clusters to reuse all the resources;
- WP (Welsh-Powell) [50]: minimizing the number of colors using the classic greedy algorithm.

Note that we are unable to evaluate the performance of both optimal solutions for the two subproblems due to their extremely high computational complexity in UDNs. To be

specific, the optimal user-centric overlapped clustering solution has an excessive complexity, and additionally it is also intricately coupled with the associated resource allocation problem. Explicitly, the optimal resource allocation solution relying on exhaustive search arrives at the complexity of $\mathcal{O}(50^{100})$ in the default simulation settings. Therefore, in the following we do not rely on the optimal solutions for benchmarking.

A. Impact of the APs' traffic-load constraint on the performance

Primarily, we evaluate and compare the performance of the proposed solution and the aforementioned schemes versus the maximum traffic-load of APs. The first point to observe from Fig. 3 is that all the overlapped solutions ($S > 1$) result in a reduced PAAR, as the maximum traffic-load of APs increases. This is a result of the two-segment path loss model and of the equal power sharing, which is related to the actual traffic-load. To be specific, the increased maximum AP traffic-load will require the serving APs of the UE to approximately share their fixed total power for serving all other UEs, while the new serving APs involved contribute less to the UE rate due to the two-segment path loss model. Another interesting observation is that only our proposed overlapped solution outperforms the non-overlapped clustering scenario ($S = 1$), as shown in Fig. 3. By contrast, the RA/SA/UFR/WP non-overlapped clustering solutions yield an even better performance than the corresponding overlapped clustering, **except for the RA one when $S = 2$** . This trend is due to the fact that our proposed user-centric **overlapped** solution **benefits from much more AP cooperation than the corresponding non-overlapped clustering relying on the designed resource allocation**, while the other **overlapped** solutions suffer from the severe overlapping-induced inter-cluster interference.

TABLE I: SIMULATION PARAMETERS

Macro-cell Coverage Area	$A = 200\text{m} \times 200\text{m}$
AP Coverage Area Radius	$d_t = 50\text{ m}$
Subcarrier Bandwidth	180 KHz
Path Loss Model [39]	$\theta_L = 2.09,$ $\theta_{NL} = 3.75,$ $D_L = 10^{-10.38},$ $D_{NL} = 10^{-14.54},$ $\phi_1 = 156\text{ m},$ $\phi_2 = 30\text{ m},$ $\bar{d} = \phi_1 / \ln(10),$ $\psi = 8.5\text{ m}$
Rician Fading K-factor	$K_R = 32$
Rayleigh Fading Correlation Coefficient	$\rho_{ss} = 0.7$
Shadow Fading Standard Deviation	$\sigma_{sd} = 10\text{dB}$
Shadow Fading Correlation Coefficient	$\rho_{sd} = 0.5$
Noise Power Density (5 dB figure)	-174 dBm/Hz
AP Transmit Power	$p_j = 30\text{ dBm}, \forall j$
Interference Threshold	$\delta_g = 10\text{ dB}$
Number of Antennas at AP	$M_j = 8, \forall j$

This implies that our proposed solution takes advantage of the overlapped clustering with the aid of our improved resource allocation. In the example of Fig. 3, when $S = 8$, the average UE rate reaches 12.3656 [b/s/Hz] and the average sum rate of 100 UEs reaches 1.2366×10^3 [b/s/Hz], thus the PAAR in the squared area of 0.04 [km²] becomes 3.0914×10^4 [b/s/Hz/km²]. We can see from the above observations that the performance advantage is a function of the value of the maximum AP traffic-load.

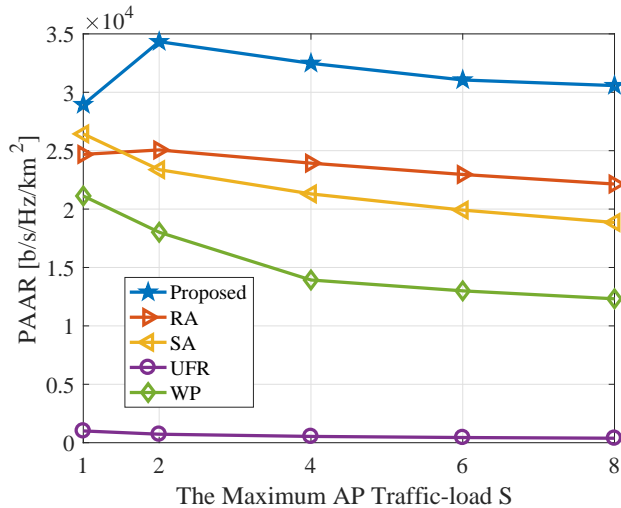


Fig. 3: PAAR versus the maximum traffic-load of APs S . The abbreviations are defined as follows: RA (Random Allocation), SA (Sequential Allocation), UFR (Unity Frequency Reuse) and WP (Welsh-Powell).

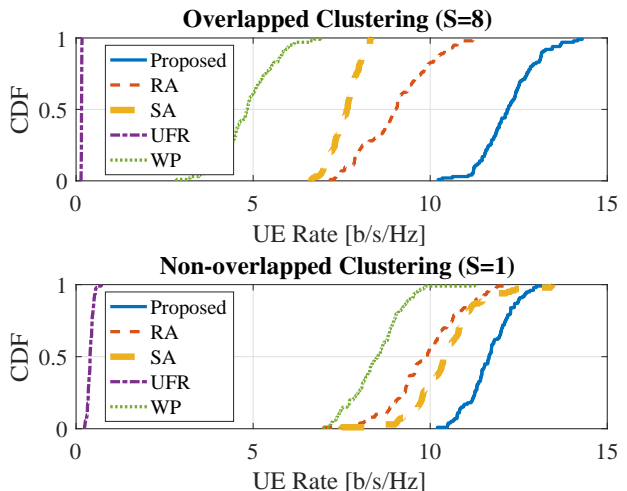


Fig. 4: CDF of the UE rate. The abbreviations are defined as follows: RA (Random Allocation), SA (Sequential Allocation), UFR (Unity Frequency Reuse) and WP (Welsh-Powell).

Next, in order to further analyze the user-centric non-overlapped and overlapped clustering, we compare the CDF of the UE rate for the aforementioned solutions in Fig. 4. As

expected, we can observe that the UE rate of the proposed solution is better than that of the other solutions, when the user-centric clusters are allowed to overlap. Furthermore, the proposed overlapped clustering solution achieves a better performance in terms of the UE rate than the corresponding non-overlapped clustering solution. Although all the other non-overlapped clustering solutions avoid the high inter-cluster interference and obtain a better throughput gain than the corresponding overlapped clustering solutions, our proposed joint user-centric overlapped clustering and resource allocation solution beneficially exploits the AP cooperation and mitigates the inter-cluster interference for enhancing the UE rate.

B. Impact of AP and UE density on the performance

The performance is further investigated here as a function of the AP and UE density in the network. Fig. 5 plots the PAAR versus the AP density under different number of RBs. On one hand, it is clearly observed from the upper subfigure that the PAAR of our proposed solution is increased with the AP density, when $N = 100$. This implies that denser APs lead to having more cooperating APs ready to serve more UEs. Hence, we achieve an increased PAAR, if the number of RBs is sufficiently high for all UEs. On the other hand, we observe from the lower subfigure of Fig. 5 that when $N = 50$, the PAAR is first increased upto 10 APs and then decreased slightly. Finally, upon further increasing the number of APs it becomes near-constant as a function of the AP density for our proposed solution. The reason behind this trend is that at a low density of APs, only some UEs can be served owing to the coverage distance threshold and due to the maximum AP traffic-load constraint. Then, upon increasing the number of APs, more UEs are served and thus the PAAR is increased, provided that the number of available RBs is sufficient for the served UEs. Nevertheless, once the number of served UEs exceeds the number of available RBs (around $L = 10$), the inter-cluster interference will degrade the UE rate attained. As a consequence, the PAAR may be reduced upon increasing the AP density, albeit more UEs are served. Furthermore, when the number of APs is further increased, the UEs will have more cooperating APs to be served by. However, this also results in an increased inter-cluster interference with the limited availability of RBs. In conclusion, the proposed solution still outperforms all the benchmarks, which is a result of its superior resource allocation capability. Additionally, it is noteworthy that even if the PAAR is a near-constant function of the AP density at a high density, the increased density of APs is beneficial in terms of serving more UEs.

Fig. 6 depicts the PAAR versus the number K of UEs for a fixed number of APs for the different solutions. Naturally, the increasing density of UEs will lead to higher interference owing to the smaller distance among them with the limited number of RBs. Additionally, more UEs will share the fixed total power in the context of equal power sharing under our maximum AP traffic-load constraint, which will degrade the UE rate. Despite this, with more UEs served by AP cooperation, the overall PAAR is expected to increase. Indeed, we observe in Fig. 6 that the PAAR is increased for all the

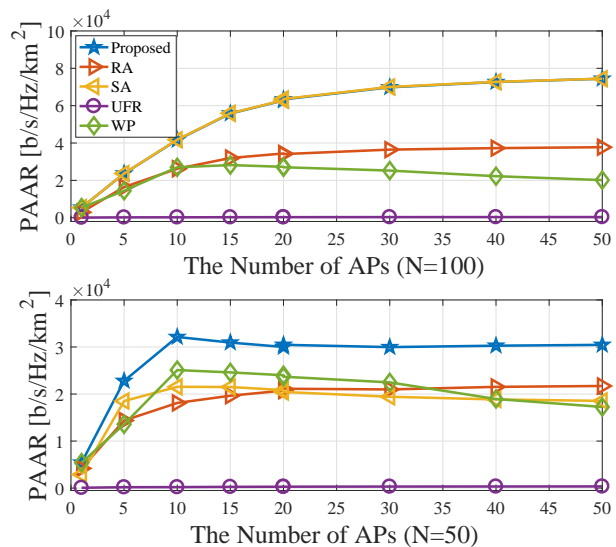


Fig. 5: PAAR versus the number of APs L . The abbreviations are defined as follows: RA (Random Allocation), SA (Sequential Allocation), UFR (Unity Frequency Reuse) and WP (Welsh-Powell).

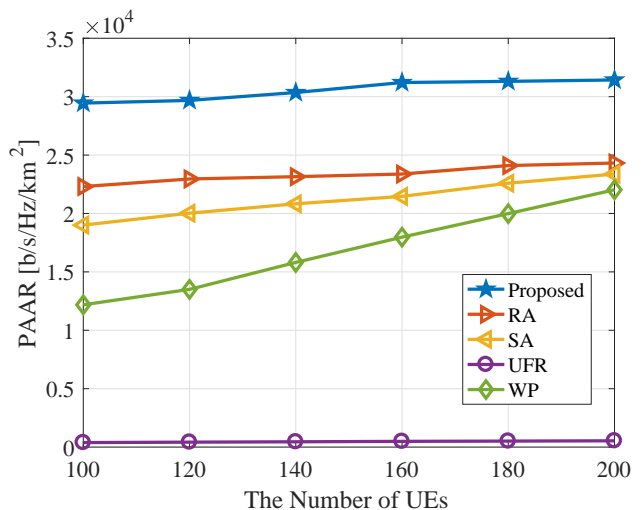


Fig. 6: PAAR versus the number of UEs K . The abbreviations are defined as follows: RA (Random Allocation), SA (Sequential Allocation), UFR (Unity Frequency Reuse) and WP (Welsh-Powell).

solutions except for the UFR solution based on overlapped clustering, when the UEs become denser. Moreover, our proposed solution outperforms the other algorithms as a benefit of mitigating the inter-cluster interference, which is an explicit benefit of enhancing the PAAR through strategically allocating orthogonal RBs. We conclude from Fig. 5 and Fig. 6 that the proposed solution achieves better PAAR gains than the other solutions in UDNs.

C. Impact of the number of RBs on the performance

In this subsection, the PAAR and the CDF of the UE rate versus the number of orthogonal RBs is investigated. It can be seen from Fig. 7 that increasing the number of orthogonal RBs has a beneficial effect on the PAAR of the proposed solution compared to the other solutions, indicating the superiority of our solution. This demonstrates that our proposed solution makes better use of the limited number of orthogonal RBs, hence achieving a higher PAAR. Furthermore, it can be clearly observed in terms of the WP solution that its PAAR first increases then remains near-constant, since it completes the graph coloring process with the aid of the minimum number of colors to ensure that no two adjacent vertices share the same color. In other words, as Remark 1 suggested, in order to obtain a high PAAR, the available number of RBs should be no less than the chromatic number. As also expected, the proposed solution achieves significant gains, when N is higher than 30 in Fig. 7.

The CDF of the UE rate versus the number of RBs is depicted in Fig. 8, which clearly shows that there is an increase in the UE rate as N increases for the joint user-centric overlapped clustering and resource allocation solution. By comparison, the proposed solution operating in the non-overlapped clustering scenario achieves a worse performance than the overlapped clustering, when the number of RBs is much higher than the chromatic number of the graph constructed. Therefore, the results of Fig. 7 and Fig. 8 demonstrate that the benefits of our proposed solution are a function of the number of RBs, which can provide us with some insights as to how we should select the number of RBs.

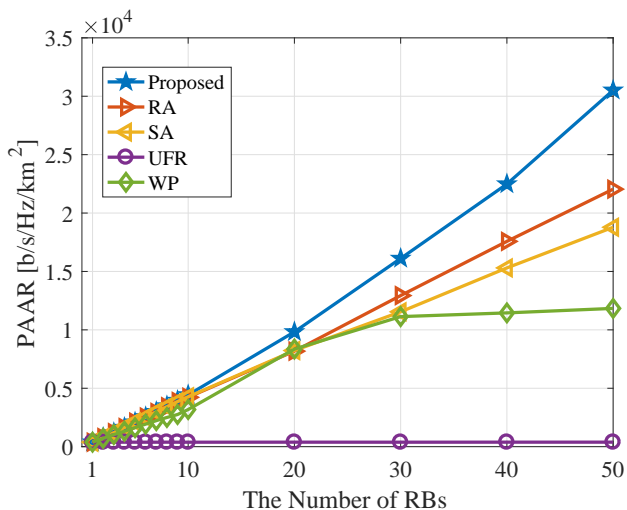


Fig. 7: PAAR versus the number of RBs N . The abbreviations are defined as follows: RA (Random Allocation), SA (Sequential Allocation), UFR (Unity Frequency Reuse) and WP (Welsh-Powell).

VI. CONCLUSIONS

A joint user-centric overlapped clustering and resource allocation framework was proposed for UDNs aiming for

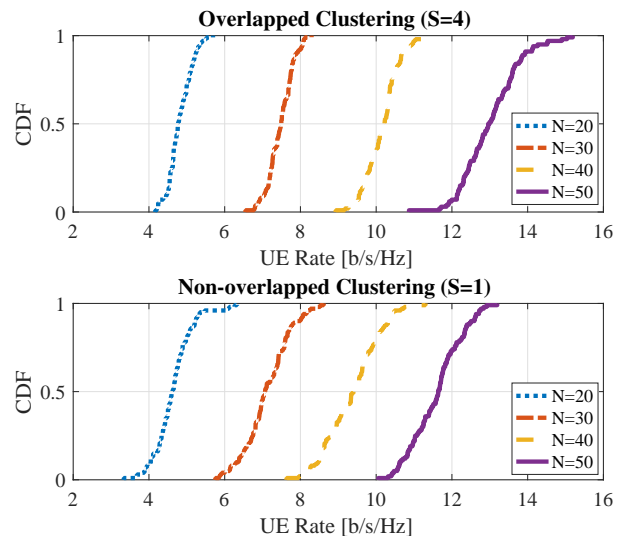


Fig. 8: CDF of the UE rate for different number of RBs N . ($S = 4$ versus $S = 1$)

maximizing the SE under practical AP traffic-load and RB constraints. Based on a graph-theoretic framework, we decoupled the problem into two independent subproblems, and then put forward a distributed user-centric overlapped clustering scheme and a practical low-complexity resource allocation scheme, respectively. Our numerical results quantified the benefits of two user-centric clustering architectures, which are predominantly owing to AP cooperation. Our proposed user-centric resource allocation solution offers superior PAAR and user rate under a range of practical system settings. For future work, we can consider the user-centric clustering solution including AP cooperation and user grouping. Instead of RB allocation aiming for maximizing the SE, multi-component objective functions can also be considered in UDNs. Finally, the users' specific motion-trajectories may also be considered.

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [2] D. López-Pérez, M. Ding, H. Claussen, and A. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2078–2101, Fourthquarter 2015.
- [3] A. Gotsis, S. Stefanatos, and A. Alexiou, "Ultra dense networks: The new wireless frontier for enabling 5G access," *IEEE Veh. Technol. Magazine*, vol. 11, no. 2, pp. 71–78, June 2016.
- [4] X. Ge, S. Tu, G. Mao, C. X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–79, February 2016.
- [5] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 2016.
- [6] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 78–85, April 2016.
- [7] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint transmission in heterogeneous networks," *IEEE Trans. on Commun.*, vol. 62, no. 11, pp. 4134–4146, Nov 2014.

- [8] J. Kim, H. W. Lee, and S. Chong, "Virtual cell beamforming in cooperative networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1126–1138, June 2014.
- [9] R. J. Wilson, *Introduction to graph theory*. Addison Wesley Longman Limited, 1996.
- [10] N. Deo, *Graph theory with applications to engineering and computer science*. Courier Dover Publications, 2017.
- [11] S. Bayat, Y. Li, L. Song, and Z. Han, "Matching theory: Applications in wireless commun." *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 103–122, Nov 2016.
- [12] Z. Zhou, M. Dong, K. Ota, and Z. Chang, "Energy-efficient context-aware matching for resource allocation in ultra-dense small cells," *IEEE Access*, vol. 3, pp. 1849–1860, 2015.
- [13] L. Lu, D. He, G. Y. Li, and X. Yu, "Graph-based robust resource allocation for cognitive radio networks," *IEEE Trans. on Signal Process.*, vol. 63, no. 14, pp. 3825–3836, July 2015.
- [14] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided device-to-device communication under channel uncertainties: A stable matching approach," *IEEE Trans. on Commun.*, vol. 63, no. 10, pp. 3882–3897, Oct 2015.
- [15] Z. Zhou, K. Ota, M. Dong, and C. Xu, "Energy-efficient matching for resource allocation in D2D enabled cellular networks," *IEEE Trans. on Veh. Technol.*, vol. PP, no. 99, pp. 1–1, 2016.
- [16] S. Sekander, H. Tabassum, and E. Hossain, "Decoupled uplink-downlink user association in multi-tier full-duplex cellular networks: A two-sided matching game," *IEEE Trans. on Mobile Comput.*, vol. PP, no. 99, pp. 1–1, 2016.
- [17] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Resource allocation for D2D communication underlaid cellular networks using graph-based approach," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 10, pp. 7099–7113, Oct 2016.
- [18] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference graph-based resource allocation (InGRA) for D2D communications underlying cellular networks," *IEEE Trans. on Veh. Technol.*, vol. 64, no. 8, pp. 3844–3850, Aug 2015.
- [19] R. Y. Chang, Z. Tao, J. Zhang, and C. C. J. Kuo, "Multicell OFDMA downlink resource allocation using a graphic framework," *IEEE Trans. on Veh. Technol.*, vol. 58, no. 7, pp. 3494–3507, Sept 2009.
- [20] T. Yang, R. Zhang, X. Cheng, and L. Yang, "Graph Coloring Based Resource Sharing (GCRS) Scheme for D2D Communications Underlying Full-Duplex Cellular Networks," *IEEE Trans. on Veh. Technol.*, vol. 9545, no. c, pp. 1–1, 2017.
- [21] Y. Meng, J. D. Li, H. Y. Li, and P. Liu, "Graph-based user satisfaction-aware fair resource allocation in OFDMA femtocell networks," *IEEE Trans. on Veh. Technol.*, vol. 64, no. 5, pp. 2165–2169, 2015.
- [22] Y. Meng, J. Li, H. Li, and M. Pan, "A transformed conflict graph-based resource-allocation scheme combining interference alignment in OFDMA femtocell networks," *IEEE Trans. on Veh. Technol.*, vol. 64, no. 10, pp. 4728–4737, 2015.
- [23] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.
- [24] Z. Chen, X. Hou, and C. Yang, "Training resource allocation for user-centric base station cooperation networks," *IEEE Trans. on Veh. Technol.*, vol. 65, no. 4, pp. 2729–2735, April 2016.
- [25] S. Akoum and R. W. Heath, "Interference coordination: Random clustering and adaptive limited feedback," *IEEE Trans. on Signal Process.*, vol. 61, no. 7, pp. 1822–1834, April 2013.
- [26] K. Huang and J. G. Andrews, "An analytical framework for multicell cooperation via stochastic geometry and large deviations," *IEEE Trans. on Inf. Theory*, vol. 59, no. 4, pp. 2501–2516, April 2013.
- [27] F. Baccelli and A. Giovanidis, "A stochastic geometry framework for analyzing pairwise-cooperative cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 2, pp. 794–808, Feb 2015.
- [28] M. Hong, R. Sun, H. Baligh, and Z. Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, February 2013.
- [29] H. S. Kang and D. K. Kim, "User-centric overlapped clustering based on anchor-based precoding in cellular networks," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 542–545, March 2016.
- [30] W. Nie, F. C. Zheng, X. Wang, W. Zhang, and S. Jin, "User-centric cross-tier base station clustering and cooperation in heterogeneous networks: Rate improvement and energy saving," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1192–1206, May 2016.
- [31] V. Garcia, Y. Zhou, and J. Shi, "Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 8, pp. 4297–4308, Aug 2014.
- [32] A. Abdelnasser, E. Hossain, and D. I. Kim, "Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 3, pp. 1628–1641, March 2014.
- [33] A. Hatoum, R. Langar, N. Aitsaadi, R. Boutaba, and G. Pujolle, "Cluster-based resource management in OFDMA femtocell networks with QoS guarantees," *IEEE Trans. on Veh. Technol.*, vol. 63, no. 5, pp. 2378–2391, Jun 2014.
- [34] L. Liang, W. Wang, Y. Jia, and S. Fu, "A cluster-based energy-efficient resource management scheme for ultra-dense networks," *IEEE Access*, vol. 4, pp. 6823–6832, 2016.
- [35] L. Liu, V. Garcia, L. Tian, Z. Pan, and J. Shi, "Joint clustering and inter-cell resource allocation for CoMP in ultra dense cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2015, pp. 2560–2564.
- [36] C. Hua, Y. Luo, and H. Liu, "Wireless backhaul resource allocation and user-centric clustering in ultra-dense wireless networks," *IET Commun.*, vol. 10, no. 15, pp. 1858–1864, 2016.
- [37] X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models," *IEEE Trans. on Commun.*, vol. 63, no. 5, pp. 1881–1894, May 2015.
- [38] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 3, pp. 2365–2380, March 2016.
- [39] M. Ding and D. López-Pérez, "On the performance of practical ultra-dense networks: The major and minor factors," in *IEEE Int. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2017, pp. 1–8.
- [40] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge university press, 2003.
- [41] S. L. Loyka, "Channel capacity of mimo architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sept 2001.
- [42] S. Lu, J. May, and R. J. Haines, "Efficient modeling of correlated shadow fading in dense wireless multi-hop networks," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, April 2014, pp. 311–316.
- [43] M. Ding, M. Zhang, D. López-Pérez, and H. Clausen, "Correlated shadow fading for cellular network system-level simulations with wrap-around," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2015, pp. 2245–2250.
- [44] A.E.Roth and M.Sotomayor, *Two Sided Matching: A study in Game-Theoretic Modeling and Analysis*. Cambridge, UK: Cambridge University Press, 1991.
- [45] A. B. MacKenzie and L. A. DaSilva, *Game Theory for Wireless Engineers*. Morgan & Claypool, 2006.
- [46] D. W. Matula and L. L. Beck, "Smallest-last ordering and clustering and graph coloring algorithms," *Journal of the ACM (JACM)*, vol. 30, no. 3, pp. 417–427, 1983.
- [47] Q. Zhang, X. Zhu, L. Wu, and K. Sandrasegaran, "A coloring-based resource allocation for OFDMA femtocell networks," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, April 2013, pp. 673–678.
- [48] J. Qiu, G. Ding, Q. Wu, Z. Qian, T. A. Tsiftsis, Z. Du, and Y. Sun, "Hierarchical resource allocation framework for hyper-dense small cell networks," *IEEE Access*, vol. 4, pp. 8657–8669, 2016.
- [49] F. Jin, R. Zhang, and L. Hanzo, "Fractional frequency reuse aided twin-layer femtocell networks: Analysis, design and optimization," *IEEE Trans. on Commun.*, vol. 61, no. 5, pp. 2074–2085, May 2013.
- [50] D. J. A. Welsh and M. B. Powell, "An upper bound for the chromatic number of a graph and its application to timetabling problems," *The Computer Journal*, vol. 10, no. 1, pp. 85–86, 1967.