

NOMA-Relevant Clustering and Resource Allocation for Proportional Fair Uplink Communications

Mylene Pischella, Didier Le Ruyet

► **To cite this version:**

Mylene Pischella, Didier Le Ruyet. NOMA-Relevant Clustering and Resource Allocation for Proportional Fair Uplink Communications. IEEE wireless communications letters, IEEE comsoc, 2019, 8 (3), pp.873-876. 10.1109/LWC.2019.2897796 . hal-02446283

HAL Id: hal-02446283

<https://hal-cnam.archives-ouvertes.fr/hal-02446283>

Submitted on 20 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NOMA-Relevant Clustering and Resource Allocation for Proportional Fair Uplink Communications

Mylene Pischella, *Senior Member, IEEE* and Didier Le Ruyet, *Senior Member, IEEE*

Abstract—This letter focuses on clustering and Resource Block (RB) allocation in multi-carrier uplink networks using Power-Domain Non Orthogonal Multiple Access (PD-NOMA). The optimization objective is time-based Proportional Fairness. PD-NOMA is applied per frame and RB only if the achieved individual data rates are larger with PD-NOMA than with Time Division Multiple Access. The clustering algorithm recursively uses maximum weight matching to build clusters per RB. Given the clusters, RB allocation is then optimal. The proposed algorithm provides a good trade-off between reduced overall network latency and rate improvement compared to two reference algorithms.

I. INTRODUCTION

Uplink Power-Domain Non Orthogonal Multiple Access (PD-NOMA) is a major technique for future dense cellular networks to multiplex the signal of several users on the same radio resources [1]. However, its performance strongly depends on the clusters of users that are allocated on Resource Blocks (RB) and decoded with Successive Interference Cancellation (SIC) at the Base Station (BS). Most previous work on uplink PD-NOMA clustering optimization aims at maximizing the sum data rate [2]–[4]. In [2], clusters of size 2 are formed with the Hungarian algorithm. In [3], the clustering problem is expressed as a maximum weight matching problem in a general graph. Edges weight are defined to match users with large channel gain disparity: the weight on the edge between users (k, j) with channel gains (g_k, g_j) , is proportional to $\max\{g_k, g_j\} / \min\{g_k, g_j\}$. In [4], a low-complexity clustering algorithm groups users depending on their channel gains. K users are ordered by descending value of their channel gains, noted $\{g_k\}_{0 \leq k \leq K-1}$ and such that $g_0 \geq g_1 \geq \dots \geq g_{K-1}$. Then the p^{th} cluster, with $p \in \{0, \dots, \lfloor K/C \rfloor - 1\}$, contains users whose indexes are equal to p modulo $(\lfloor K/C \rfloor)$. In [5], clusters are allocated on RB according to the Proportional Fair objective using exhaustive search. Proportional Fairness (PF) is also pursued in [6] with clusters of size 2 through power allocation optimization. In all these papers, PD-NOMA is used even if it does not lead to any data rate improvement for some clusters and channels realizations, compared to transmitting interference-free with Time Division Multiple Access (TDMA).

The authors are with CNAM CEDRIC/LAETITIA, 292 rue Saint-Martin, 75003 Paris, France (contact: mylene.pischella@cnam.fr).

In this letter, we address the problem of clustering users in clusters of any size, and allocating clusters to RB in multi-carrier uplink PD-NOMA. The optimization objective is time-based Proportional Fairness. Contrary to [3], [4], matching users only according to their channel gains is then not consistent with the optimization objective. Compared to [5], the proposed clustering and RB algorithms are less complex as we do not use exhaustive search on all possible clusters, but only select clusters where the individual data rates of all users are larger with PD-NOMA than with TDMA. This constraint, referred to as 'NOMA relevance' in this letter, is fundamental to ensure that PD-NOMA actually brings some additional gain. Contrary to [7], it is not enforced through power allocation for predefined clusters, but through the determination of clusters that verify this constraint with fixed power levels.

II. SYSTEM MODEL

We consider K users aiming at transmitting to one BS using uplink multi-carrier single-antenna PD-NOMA. At the BS, users signal are decoded with SIC. The decoding order is by descending channel gains [4]. We focus on the optimization of clustering and RB allocation in a given frame f , composed of T Time Transmit Intervals (TTI) and N RB. In the following, the frame index is removed in order to simplify notations.

g_k^l is the channel gain from user k to the BS in RB l and frame f , assumed constant during the frame. It includes frequency-flat fading, path loss and shadowing. N_0 is the noise power in the RB bandwidth B . P is the transmit power per user and RB. All channel gains are perfectly known by the BS, which is responsible of the allocation. The set of users is denoted \mathcal{S}_K and the set of RB \mathcal{S}_N . F_{\max} is the total number of frames required for all users to achieve a minimum amount of transmitted data, denoted as D_{\min} . We refer to F_{\max} as the overall network latency, as it is proportional to the required delay for all users to achieve their Quality of Service constraint D_{\min} . We assume that a user can be active in only one TTI per frame and RB.

Let Θ be the set of all possible (not disjoint) clusters containing 1 up to C users in \mathcal{S}_K . Its cardinality is equal to $|\Theta| = N_C = \sum_{m=1}^C \binom{K}{m}$. We consider a specific cluster $\theta \in \Theta$ chosen for transmission in RB l . The BS receives the sum of the signals transmitted by all users in θ . It then uses SIC to recursively decode the signal of the user with largest

channel gain, and then remove it from the received summed signal. Consequently, the interference on the signal of user $k \in \theta$ active in RB l comes from the users with lower channel gains in set $\bar{\theta}_k^l = \{m \in \theta \mid g_m^l \leq g_k^l\}$. It is equal to:

$$I_{k,\theta}^l = \sum_{m \in \bar{\theta}_k^l} g_m^l P \quad (1)$$

Then the data rate of user k in RB l using PD-NOMA in cluster θ is:

$$R_{k,\text{NOMA},\theta}^l = \log_2 \left(1 + \frac{g_k^l P}{I_{k,\theta}^l + N_0} \right) \quad (2)$$

On the contrary, if the signal of user k is transmitted in TDMA, it does not suffer from any interference. However, the required number of frames to make all users of cluster θ transmit orthogonally is equal to the number of users in θ , denoted as $|\theta|$. Assuming that the allocated RB for users in θ is still RB l with TDMA, and that channels remain constant during $|\theta|$ consecutive frames, the data rate of user $k \in \theta$ is then:

$$R_{k,\text{TDMA},\theta}^l = \frac{1}{|\theta|} \log_2 \left(1 + \frac{g_k^l P}{N_0} \right) \quad (3)$$

where the term $1/|\theta|$ comes from the fact that user k transmits in one frame out of $|\theta|$. Eq. (2) and (3) show that the data rate of user k in RB l and cluster θ is larger with PD-NOMA than with TDMA if the following constraint is verified:

$$\left(1 + \frac{g_k^l P}{I_{k,\theta}^l + N_0} \right)^{|\theta|} \geq 1 + \frac{g_k^l P}{N_0} \quad (4)$$

with $I_{k,\theta}^l$ the SIC interference given by equation (1). If constraint (4) holds for all users k in cluster θ , then the individual data rates of each user in θ is larger with PD-NOMA than with TDMA. In this case, all users benefit from PD-NOMA. We refer to the set of constraints (4) for all users $k \in \theta$ as the NOMA relevance constraints.

III. NOMA CLUSTERING AND RESOURCE ALLOCATION

A. Optimization problem

Let a_θ^l be a Boolean that indicates if cluster θ is active in RB l . If $a_\theta^l = 1$, then all users in cluster θ are allocated in one TTI of frame f and in RB l using PD-NOMA. \mathbf{a} is the corresponding allocation matrix with dimension $N_C \times N$. Let $b_{k,\theta}^l$ be a Boolean indicating if user k belongs to cluster θ and RB l . The optimization problem enforces that all users in the same cluster should verify the NOMA relevance constraint. Consequently, if eq. (4) is not true, then $b_{k,\theta}^l$ is necessarily equal to 0, as expressed next in constraint (6e). \mathbf{b} is the final clustering matrix with dimension $K \times N_C \times N$.

Clustering and resource allocation are performed according to the PF optimization objective. PF is well-known for achieving good trade-offs between large sum rates and low complexity, contrary to completely unfair optimization objectives (aiming at maximizing the sum rate) leading to high latency and to completely fair objective (aiming at maximizing

the minimum rate) leading to low sum rates. In this letter, it is applied in time-domain, by maximizing in each frame f the sum of the weighted data rates of all users, where the weight of user k , denoted as w_k , is inversely proportional to the cumulative amount of data transmitted by user k up to frame $(f-1)$: $w_k = 1/D_{k,\text{cum}}^{f-1}$. The cumulative amount of transmitted data after resource allocation in frame f is defined as follows:

$$D_{k,\text{cum}}^f = D_{k,\text{cum}}^{f-1} + \sum_{l \in \mathcal{S}_N} \sum_{\theta \in \Theta} B \delta_{\text{TTI}} b_{k,\theta}^l \times a_\theta^l \times R_{k,\theta}^l \quad (5)$$

where δ_{TTI} is the TTI duration.

Then the PD-NOMA PF optimization problem in frame f is:

$$\max_{\mathbf{a}, \mathbf{b}} \sum_{\theta \in \Theta} \sum_{k \in \mathcal{S}_K} \sum_{l \in \mathcal{S}_N} b_{k,\theta}^l \times a_\theta^l \times w_k R_{k,\theta}^l \quad (6a)$$

$$\text{s.t.} \quad \sum_{\theta \in \Theta} b_{k,\theta}^l \leq 1 \quad \forall (k, l) \in \mathcal{S}_K \times \mathcal{S}_N \quad (6b)$$

$$\sum_{k \in \mathcal{S}_K} b_{k,\theta}^l \leq C \quad \forall (\theta, l) \in \Theta \times \mathcal{S}_N \quad (6c)$$

$$\sum_{\theta \in \Theta} a_\theta^l \leq T \quad \forall l \in \mathcal{S}_N \quad (6d)$$

$$b_{k,\theta}^l = 0 \text{ if (4) is not true, } \forall (k, \theta, l) \in \mathcal{S}_K \times \Theta \times \mathcal{S}_N \quad (6e)$$

where 's.t.' means 'subject to' and $R_{k,\theta}^l$ is computed with eq. (2) with the interference given by (1) that depends on the other users in cluster θ , and may potentially be equal to 0. Constraint (6b) expresses that each user should belong to at most one cluster per RB. Constraint (6c) means that the maximum number of users per cluster should be equal to C . Constraint (6d) expresses that up to T different clusters can be allocated per RB with one cluster allocated per TTI. In order to circumvent the complexity of problem (6), we propose a heuristic algorithm that proceeds in two stages: a) first, optimize \mathbf{b} per RB ; b) then, allocate clusters on RB to optimize \mathbf{a} .

B. First stage: clustering algorithm

Problem (6) is firstly written independently per RB $l \in \mathcal{S}_N$ without taking into account constraint (6d). Let \mathbf{b}^l be the clustering binary matrix of size $K \times N_C$ for RB l and let us define the PF metric of user k in cluster θ and RB l as:

$$\alpha_{k,\theta}^l = w_k R_{k,\theta}^l \quad (7)$$

Then the optimization problem (6) per RB becomes the following clustering determination problem:

$$\max_{\mathbf{b}^l} \sum_{\theta \in \Theta} \sum_{k \in \mathcal{S}_K} b_{k,\theta}^l \alpha_{k,\theta}^l \quad (8a)$$

$$\text{s.t.} \quad \sum_{\theta \in \Theta} b_{k,\theta}^l \leq 1 \quad \forall k \in \mathcal{S}_K \quad (8b)$$

$$\sum_{k \in \mathcal{S}_K} b_{k,\theta}^l \leq C \quad \forall \theta \in \Theta \quad (8c)$$

$$b_{k,\theta}^l = 0 \text{ if (4) is not true, } \forall (k, \theta) \in \mathcal{S}_K \times \Theta \quad (8d)$$

Problem (8) is solved by recursively matching users with graph-theoretical techniques in $(C-1)$ steps. Users are added one by one in each cluster θ if they verify the NOMA relevance constraint (4), taking into account the interference from the users already present in cluster θ . We assume that channel gains are ordered by descending values: $g_0 \geq g_1 \geq \dots \geq g_{K-1}$ and we define for $p \in \{1, \dots, C\}$ the set of $K_c = \lfloor \frac{K}{C} \rfloor$ users $\Gamma_p = [(C-p)K_c, \dots, (C-p+1)K_c - 1]$. Ordering users in such a way allows to only test one constraint (4) when adding one user in a cluster.

1) *First step:* In the first step, the users with the lowest $2K_c$ channel gains, belonging to set $\Gamma_1 \cup \Gamma_2$, are matched to form clusters of size 2. A graph $\mathcal{G}_2^l = \{\Gamma_1 \cup \Gamma_2, \mathcal{E}_2^l\}$ with vertices representing users in $\Gamma_1 \cup \Gamma_2$ is built. Its edges \mathcal{E}_2^l are defined as follows: vertices $(k, j) \in (\Gamma_1 \cup \Gamma_2)^2$ with $k \leq j$ are connected by an edge if constraint (4) is verified for both users. As $g_k^l \geq g_j^l$, constraint (4) for user k is:

$$\left(1 + \frac{g_k^l P}{g_j^l P + N_0}\right)^2 \geq 1 + \frac{g_k^l P}{N_0} \quad (9)$$

whereas constraint (4) is always true for user j because it does not suffer from any interference and the SINR is positive. If (9) is verified, then (k, j) is the i^{th} potential cluster, denoted as $\theta_{2,i}^l$. The weight associated to vertice (k, j) is equal to the summed PF metric α of users in $\theta_{2,i}^l$, with α given by eq. (7):

$$\beta_{\theta_{2,i}^l}^l = \alpha_{k,\theta_{2,i}^l}^l + \alpha_{j,\theta_{2,i}^l}^l \quad (10)$$

Let $\beta_{\theta_2}^l$ be the vector of weights associated to all vertices in graph \mathcal{G}_2^l . Then the following theorem holds:

Theorem 1: The optimal solution of problem (8) for clusters of maximum size $C = 2$ is equal to the solution $\Theta_2^l = \{\theta_{2,s[0]}^l, \dots, \theta_{2,s[N_2^l-1]}^l\}$ of the maximum weight matching algorithm in the general undirected graph \mathcal{G}_2^l with weights $\beta_{\theta_2}^l$, where $s[i]$ is the i^{th} cluster selected by the maximum weight matching algorithm applied on graph \mathcal{G}_2^l .

Proof: The proof is straightforward due to the construction of graph \mathcal{G}_2^l [8]. We can notice that $N_2^l < N_2$ because all original

2) *Second step:* In the second step, a new graph \mathcal{G}_3^l is defined. Its vertices are equal to the clusters in set Θ_2^l selected in the first step and to the users with larger channel gains in set Γ_3 . In graph \mathcal{G}_3^l , a user m in Γ_3 is connected with a cluster (k, j) in Θ_2^l by an edge if all users (m, k, j) jointly verify NOMA relevance. Constraint (4) for the user m with the largest channel gain is:

$$\left(1 + \frac{g_m^l P}{g_k^l P + g_j^l P + N_0}\right)^3 \geq 1 + \frac{g_m^l P}{N_0} \quad (11)$$

For user k , eq. (4) is necessarily verified since eq. (9) holds and the SINR is positive. The same reasoning applies for user j who does not suffer from any interference. Then $(m, k, j) = \theta_{3,i}^l$ is the potential i^{th} cluster of size 3. In graph \mathcal{G}_3^l , the weight associated to the edge between vertices m and (k, j) is:

$$\beta_{\theta_{3,i}^l}^l = \alpha_{m,\theta_{3,i}^l}^l + \alpha_{k,\theta_{3,i}^l}^l + \alpha_{j,\theta_{3,i}^l}^l \quad (12)$$

\mathcal{G}_3^l is a bipartite graph because sets Θ_2^l and Γ_3 are disjoint.

Lemma 1: A feasible solution of problem (8) for clusters of size $C = 3$ is equal to the solution Θ_3^l of the maximum weight matching algorithm in the undirected bipartite graph \mathcal{G}_3^l with weights $\beta_{\theta_3}^l$ given by eq. (12).

Proof: The proof is similar to that of Theorem 1, except that the solution may then be sub-optimal because of the recursive nature of the algorithm.

3) *Next $(C-3)$ steps:* Finally in the next steps, $p \in \{4, \dots, C\}$, one user with larger channel gain indexed in Γ_p is matched with a cluster of size $(p-1)$ obtained at the previous recursion in a similar way as in the second step. It is then obvious that Lemma 1 is still valid for clusters of size p .

4) *Remark on single-carrier case with $C = 2$:* For the single-carrier case with $C = 2$, RB allocation is directly solved by clustering, as expressed by Theorem 2:

Theorem 2: Let Θ_2^l be the list of clusters of size 2 given by Theorem 1 and ordered such that $\beta_{\theta_{2,s[0]}^l}^l \geq \beta_{\theta_{2,s[1]}^l}^l \geq \dots \geq \beta_{\theta_{2,s[N_2^l-1]}^l}^l$. Then the optimal solution of problem (6) when $S_N = \{l\}$ for clusters of maximum size $C = 2$ is equal to the T first clusters $\{\theta_{2,s[0]}^l, \dots, \theta_{2,s[T-1]}^l\}$.

Proof: Constraint (6d) in problem (6) is verified by selecting T clusters. These clusters maximize the objective function (6a) and they fulfill constraints (6b)-(6c)-(6e) according to Theorem 1. This completes the proof.

C. Second stage: RB allocation

In the second stage, the clusters obtained at the first stage are allocated on RB according to the PF optimization objective. The weight of any cluster $\theta \in \Theta$ in RB l is equal to the summed PF weights of the users in the cluster, assuming that the cluster has been selected in RB l in the first stage of the algorithm. It is defined as follows:

$$\gamma_{\theta}^l = \sum_{k \in S_K} b_{k,\theta}^l \times w_k R_{k,\theta}^l = \begin{cases} \beta_{\theta}^l & \text{if } \left(\sum_{k \in S_K} b_{k,\theta}^l\right) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

With this definition, problem (6) becomes:

$$\max_{\mathbf{a} \in \{0,1\}^{N_C \times N}} \sum_{\theta \in \Theta} \sum_{l \in S_N} a_{\theta}^l \gamma_{\theta}^l \quad (14a)$$

$$\text{s.t. } \sum_{\theta \in \Theta} a_{\theta}^l \leq T \quad \forall l \in S_N \quad (14b)$$

Problem (14) is a binary linear programming problem. It is optimally solved by a sorting algorithm, that iteratively chooses the best cluster and RB according to:

$$(\theta^*, l^*) = \arg \max_{(\theta \in \Theta, l \in S_N)} \gamma_{\theta}^l \quad (15)$$

Then $a_{\theta^*}^{l^*} = 1$ and cluster θ^* is removed from the set of clusters that can be allocated in RB l^* by setting $\gamma_{\theta^*}^{l^*} = 0$. Moreover, when the number of clusters allocated in frame f and RB l^* reaches T , $\gamma_{\theta}^{l^*}$ is set to 0 for all the clusters that have not yet been allocated in RB l^* .

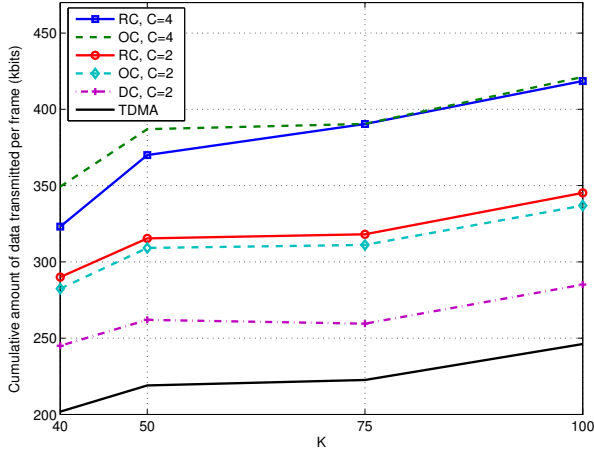


Fig. 1. Cumulative amount of data transmitted per frame after F_{\max}

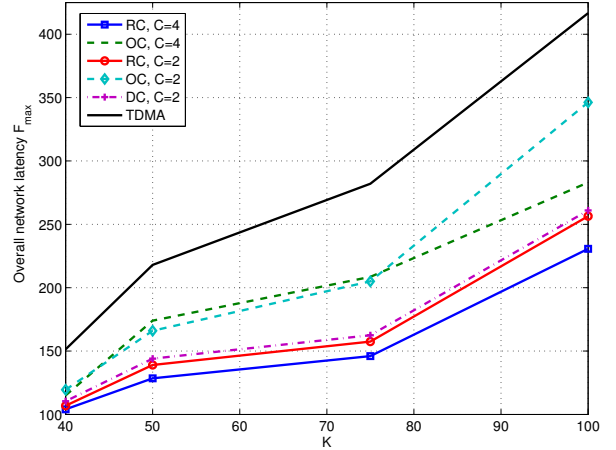


Fig. 2. Overall network latency F_{\max}

D. Complexity analysis

The clustering algorithm has a polynomial complexity in $\mathcal{O}(N(2K_c)^{2.5} + N(C-2)K_c^{1.5})$, where the first term corresponds to the complexity of maximum weight matching in a general undirected graph and the other terms to the complexity of maximum weight matching in a bipartite graph [8]. The RB allocation complexity is that of the sorting algorithm: $\mathcal{O}(N'_c N \log_2(N'_c N) + N'_c N)$, where N'_c is the largest number of feasible clusters per RB after the first stage. $N'_c < N_c$ due to constraints (6b) and (6e). The overall complexity is dominated by the first maximum weight matching step in $\mathcal{O}(N(2K_c)^{2.5})$ for large K and low C .

IV. SIMULATION RESULTS

The clustering algorithm proposed in section III-B, called 'RC' for 'Relevant Clustering', is compared with two other clustering algorithms presented in section I: Disparity Clustering (DC) from [3], and Ordered Clustering (OC) from [4]. The RB allocation algorithm from section III-C is used after clustering with RC, DC and OC. The simulation parameters are the following: the BS has an omnidirectional antenna with radius $R = 1$ km. Users are uniformly distributed within the cell. The noise spectral density is -174 dBm/Hz, B is equal to 180 kHz and P to 5 mW. Log-normal shadowing has a standard deviation of 9 dB and the path loss model is LTE at 2.6 GHz: $L_{dB}(r) = 128.1 + 37.6 \log_{10}(r(\text{km}))$. The i.i.d. flat fading is Rayleigh distributed. We consider $N = 25$ RB, $T = 10$ TTIs per frame with $\delta_{\text{TTI}} = 1$ ms. The overall network latency is equal to the minimum number of frames F_{\max} required so that all K users transmit $D_{\min} = 10$ kbits.

Fig. 1 and Fig. 2 show that DC is less efficient in terms of achieved data rates, but that it requires approximately the same overall network latency than the proposed RC algorithm. Opposite conclusions stand for OC. With RC, data rates and network latency are both improved by using clusters of larger size. The overall latency however almost does not decrease when $C = 4$ compared to $C = 2$ for low values of K with OC,

because the users in clusters of size 4 may have very similar channel gains, and consequently suffer from large interference. Finally, when $C = 2$, the overall latency with TDMA is at least twice that of PD-NOMA. Consequently, the global power consumption is then not larger with PD-NOMA than with TDMA. This conclusion no longer stands when $C = 4$, but the rate increase then largely compensates for the power increase.

V. CONCLUSIONS

This letter has detailed a graph-based clustering and resource allocation algorithm for uplink PD-NOMA guaranteeing NOMA relevance. Simulation results showed that the proposed algorithm provides good performances both in terms of overall network latency and sum data rates, contrary to several reference algorithms.

REFERENCES

- [1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, Oct 2017.
- [2] M. A. Sedaghat and R. R. Müller, "On user pairing in uplink noma," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3474–3486, May 2018.
- [3] A. Celik, R. M. Radaydeh, F. S. Al-Qahtani, A. H. A. El-Malek, and M. S. Alouini, "Resource allocation and cluster formation for imperfect noma in dl/ul decoupled hetnets," in *2017 IEEE Globecom Workshops (GC Wkshps)*, Dec 2017, pp. 1–6.
- [4] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (noma) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.
- [5] X. Chen, A. Benjebbour, A. Li, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (noma)," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, May 2014, pp. 1–5.
- [6] F. Liu, P. Mhnen, and M. Petrova, "Proportional fairness-based user pairing and power allocation for non-orthogonal multiple access," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Aug 2015, pp. 1127–1131.
- [7] J. A. Oviedo and H. R. Sadjadpour, "A fair power allocation approach to noma in multiuser siso systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7974–7985, Sept 2017.
- [8] Z. Galil, "Efficient algorithms for finding maximum matching in graphs," *Computing surveys*, vol. 18, no. 1, pp. 23–38, March 1986.