Clustering for Small Cells

Applying data mining, sensor, and graph theory schemes in Ultra Dense Networks

Stefanos Falangitis, Panagiotis Spapis, Panagis Magdalinos, George Beinas, Nancy Alonistioti
National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
Athens, Greece
{sfalangitis, pspapis, panagis, gbeinas, nancy}@di.uoa.gr

***Abstract*—The huge increase in the number of small cells as well as their unplanned deployment has increased the need for identification of schemes for their coordination. Up to now several grouping schemes have been proposed for the coordination of the femtocells mainly linked to specific network problems, such as load balancing, interference coordination, power control etc. In general, these solutions are problem and topology specific and have benefits for the network and the user in specific deployments. In terms of this paper we propose a revolutionary approach for the femtocell grouping problem; we use clustering algorithms coming from data mining, sensor networks, and graph theory for grouping femtocells. We evaluate these mechanisms against their ability to form clusters in a realistic environment and come up with a quantity and quality analysis of their applicability in 5G networks.**

Keywords—clustering, data mining, sensor networks, graph theory, femtocells, 5G

I. Introduction

Traffic analysis the past years indicates that both wireless traffic and number of connected devices will increase significantly over the past years. More specifically, the traffic will have increased 10 times from 2011 to 2015 [1], whereas the number of connected devices will increase to 100 billions including computers, sensors and actuators, residential applications etc. [2].

3GPP has timely identified the significant traffic volume increase and the augmented users' demands, and draw three directions for covering their needs. These directions are related to improvement of spectral efficiency, the exploitation of the unused spectrum, and the extra layering. The first two directions are related to Coordinated Multiple Point (CoMP) transmission using sophisticated MIMO techniques and cooperation among spectrum license holders for dynamic spectrum usage respectively. The third one is related to the addition of extra layers, via installation of small cells such as micro, pico, and femto cells. The uncontrolled installation and proliferation of small cells is responsible for several issues that often occur in such networks, like uncontrolled interference, waste of energy, and resources in general, etc.; this implies that the coordination of the small cells is required. Several coordination schemes have been proposed in the literature either centralized ones or decentralized ones [3]. All the schemes however assume the presence of grouping algorithm at least for the identification of the neighborhoods, where the configuration actions will take place. The research community has proposed various grouping schemes; the grouping is referred as clustering.

Several nodes' clustering techniques have been proposed in the literature focusing on diverse objectives; the node clustering schemes may be focusing on formation of clusters for load balancing and interference mitigation, emergency communication links to backhaul, knowledge distribution etc. (e.g., [8][9][15]. On the other hand, the idea of grouping of observations is related to the identification of the similarities among observations, and the visualization of these similarities; which has been largely studied in data mining application area and graph theory. In this paper we are following an evolutionary approach. We suggest using algorithms from several scientific application areas, namely data mining, sensor networks, and graph theory algorithms, for clustering of LTE small cells. The small cells' clustering is based on simple connectivity measurements and on link quality measurements related to how well a node senses its neighboring ones.

The rest of this paper is organized as follows: Section II presents state of the art algorithms related to small cells clustering; Section III presents the selected algorithms and their key characteristics whereas Section IV provides a qualitive and a quantitive analysis of the selected algorithms in future ultra-dense environment. Section V concludes the findings of this paper.

II. Related Work

As described above, the clustering as a functionality has attracted the interest of the research community; several schemes related to resource allocation and efficient network have been assuming clustering schemes that enable them to disseminate information, enforce actions or even make decisions related to interference management [8], mobility management and handovers [15] or even control actions (such as synchronization [9]).

Regarding the clustering functionality in wireless networks, the clustering proposals try to exploit the connectivity measurements among the nodes and end-up to problem specific solutions. In [10] the authors focus on the vast deployment of femtocells while taking into account quality of service (QoS) requirements; each Femtocell Access Point (FAP) calculates the number of interfering femtocells (called interference degree) and transmits this information to its direct neighbors. The FAP with the highest interference degree becomes Cluster-

Head (CH) while the rest of the FAPs become Cluster-Member (CM). In [11] the authors propose femtocells' clustering for the efficient resource allocation inside each cluster based on a Semi-Definite Programming (SDP) algorithm that examines the channel gain among FAPs when grouped in the same cluster. This process is executed by the femto gateway (FGW), which is responsible to calculate the optimal cluster assignment and this information is sent to the femtocells through wired backhaul; the CH election is also executed at the FGW. In [12] the authors perform femtocell clustering for energy savings while maintaining a certain Quality of Service. The main clustering formation criteria are the femtocell distribution density and the RSRP; each base station calculates its local weight based on the neighbors RSRPs. The solution is similar to ad-hoc networks' schemes and the procedure is repeated until all nodes are either FBS Leaders (FLs) or FBS Members (FMs).

In [14] and [16] the authors propose clustering for handling interference. In the first scheme the clusters are one hop formations and the CH is the most interfering node. In the [16] they provide an interference graph for grouping the nodes; the nodes that interfere the most form a cluster.

Finally, in [13] the concepts of Physical Cluster and Virtual Cluster are being introduced in order to collaboratively allocate resources among multiple femtocells. The Physical Cluster formation is based on the spatial correlation between any two femtocells and femtocells are grouped into the same cluster if their location is within a threshold with respect to a threshold distance. On the other hand Virtual Clusters contain nodes using the same channel but in different locations.

III. Proposed Scheme

The previous section described state of the art mechanisms for clustering of nodes. The initial rational for clustering, comes from the need for identification of the similarities among observations, and the visualization of these similarities; which ended up to data mining clustering techniques. Inspired by these approaches other scientific areas attempted to identify similarities among observations (or vectors) no matter what they represent, for the minimization of the communication cost, the extraction of hierarchies and specific roles, etc. The examination of a variety of clustering algorithms from different scientific fields and the selection of the most suitable per case will enable the capturing of the most important objective(s) of the application field (e.g., minimization of energy/power consumption, reliability, etc.). A generic categorization of clustering schemes could be considered the following:

- Data Mining clustering schemes: the algorithms comprising this family have as main target the grouping of data objects into groups with similar objects; the grouping results in the extraction of useful observations.
- Sensor – Ad hoc networks: the algorithms of this family concern the grouping of nodes for minimization of communication cost, information fusion, and identification of specific roles in the cluster. Also the exploitation of local information may be available; thus no global knowledge is required for the clustering of nodes.
- Graph theory based schemes: in such schemes the similarities among the observations are weighted links; attributes from graph theory are used in order to group entities based on their connections into similar groups. Such schemes are in general used in biomedicine.

Each category focuses on grouping entities based on the similarities that can be found. This distinction among clustering algorithms is based on the application field of each mechanism. The following sub-sections provide a brief description of some of the most representative algorithms of each category.

A. Data Mining Algorithms

The analysis of huge amount of data has led to the need for the development of clustering schemes. As mentioned afore, the algorithms of this family target to the grouping of data objects into similar groups so as to extract meaningful sets of objects. Typically, these objects are database entries describing key features of the set member; the clustering procedure is executed simultaneously on the entire dataset. Depending on the algorithm, the clustering process has variable time constraints and cluster assignments. The procedure may lead to clusters' formation, hierarchies of clusters, etc. Typical examples of algorithms of this field are the k-Means and Hierarchical clustering that are being analyzed in the following paragraphs.

1) k-Means

The k-Means algorithm is a well-known partitioning method 0. Such methods partition n objects into k partitions ($k \leq n$). Each partition represents a cluster and the clusters are formed so as to optimize a similarity function (e.g., distance). Respectively, k-Means takes as input the parameter k and partitions a set of n objects into k clusters having as objective high intra-cluster similarity and low inter-cluster similarity.

Initially, k objects are selected and they represent a cluster center. All other objects are associated with the cluster to which it is the most similar based on the distance between the current center and the object. After the partitioning, the new centers of the clusters are computed (each center is not necessarily an object). All objects re-associate themselves with the new centers. This procedure is repeated until the computed cluster centers remain the same and so the cluster assignments of objects are permanent.

k-Means is a centralized algorithm with its main application being in data analysis. It comprises a simple approach for cluster formation. The partitioning scheme is produced in a fast and at the same time fine-grained way, even for a large number of objects. The main drawbacks of the algorithm are found on the random selection of the initial cluster centers and its dependency on the number of clusters as an input parameter. Finally, the output partitions are in the form of hyper-spheres which may not always result to the best partitioning.

2) Hierarchical clustering

Hierarchical methods group objects into a tree of clusters 0. This structure consists of levels of hierarchy. The tree is created in two ways: either as a top-down (divisive) approach or as a bottom-up (agglomerative).

Hierarchical divisive clustering has as input n objects and groups them in the same cluster. In each iteration a cluster is split based on the criteria set (e.g., split the cluster from which will emerge two clusters with the least similarity between them). Gradually, smaller clusters are formed and the algorithm finishes either when all objects belong to single-object clusters or when certain termination condition is satisfied (e.g., number of clusters, minimum distance threshold between clusters). At high levels of hierarchy the split decision is difficult due to the comparison of all the objects. As a result this approach has rarely been applied.

Hierarchical agglomerative clustering is the complete opposite process. To begin with, each object forms a single-object cluster. The algorithm continues by merging two clusters with the highest similarity at each step (e.g., choose two clusters with the smallest distance between them). This process is repeated until all objects belong to the same group or until certain criteria are met (e.g., number of cluster, maximum distance threshold between clusters). Unlike divisive method, agglomerative clustering is greatly applied and lots of variations have been developed. Their diversification resides in their definition of between-cluster similarity.

B. Sensor - Ad hoc networks Algorithms

The deployment of sensors in many application areas is a common practice for information gathering. This implies that there is need for information gathering and fusion in an efficient manner (i.e., computational and communication cost). In such decentralized schemes, the focus is paid in the formation of clusters using partial knowledge of the network (e.g., one-hop away connectivity) and focusing on the efficiency of the approach. The main roles in the majority of these algorithms are:

- Cluster-Head (CH): is responsible for gathering and transmitting the information acquired from all nodes belonging to the same cluster, as well as the discovery of routes to the core network.
- Cluster-Member (CM): periodically sends information to the CH and acts as intermediate for routing purposes.

Apart from sensors, such algorithms are also applied to ad-hoc topologies with mobile nodes. Nodes' mobility imposes the need for cluster maintenance phase is introduced for adapting the clusters' structure. The studied mechanism that will be presented is 3-hop between adjacent CHs (3hBAC) [5] 3-hop between adjacent cluster-heads (3hBAC) algorithm focuses on grouping nodes into small groups of radius up to 2 hops, while limiting the number of clusters produced and keeping the network structure as stable as possible. Throughout the maintenance phase excessive formation of small clusters is avoided. The cluster formation phase of the algorithm is as follows:

- at the beginning all nodes exchange with their direct neighbors their degree (i.e., the number of nodes they sense),
- for the member role assignment, every node has a Head-Priority (HP) attribute.
- one random node (without defined state) checks if its degree is higher than the degree of all its neighbors. The elected node becomes CH, and all the neighboring nodes (2-hop away radius) are being restricted from becoming CHs.

The previous process takes place until all nodes have been either CHs or CMs. A special type of CM may occur that of nodes that are two hops away from CHs and are not able becoming CHs on their own. Similar procedure takes place when two clusters may be merged due to nodes' mobility.

C. Graph Theory approaches

The graph-based clustering methods take advantage of the graph's structure and properties so as to create partitions with certain conditions satisfied such as intra-cluster connectivity, minimum cluster diameter, etc. Each observation is linked with the rest of the observations with a weight. Such weight captures the similarity between two observations; the entire set of entities can then be modeled as a directed weighted graph. The aforementioned modeling is used in biomedicine where there is a need for using analytical methodology for the analysis and exploitation of the information contained in genes. The formed groups consist of observations with probable transactions among them and improbable ones among observations from different clusters. The base example for this family of clustering algorithms is Markov Clustering, which profits of the graph's structure and produces clusters with high internal similarity.

Markov Clustering partitions a graph by simulating multiple random walks inside the graph [6]. The main idea resides in the fact that by randomly visiting nodes, the number of times strongly connected nodes are visited is much higher than nodes with weak paths between them.

The first action is to create the graph's adjacency matrix. Regarding the importance and cost of paths (i.e., edge weight) both weighted and un-weighted matrices are acceptable and regarding the direction of connections among nodes, the input matrix can be either directed or undirected. Thereafter, a column-stochastic probability matrix M is calculated and a two-step repeated procedure is initiated. The conclusion of this process results to a probabilistic transition matrix whose every column represents a cluster.

This algorithm includes optional steps. Before initiating the two-steps repeated procedure, self-loops with magnitude m may be added to every node. This option strengthens the connection between neighbors with many paths connecting them. One more option, during the inflation step, is pruning of values smaller than a threshold t and it is executed before rescaling and after raising all elements to power r. Some transitions are improbable due to small values. This can be seen as two nodes that do not belong in the same cluster. As a result, the value can be set to zero and the magnitude of this link is distributed proportionally to all other probable links. This way, the algorithm converges faster to a steady state.

IV. CLUSTERING ALGORITHMS' EVALUATION

A. Quantitive Comparative Analysis

The presented algorithms have been evaluated in a typical multi-floor environment, with several rooms. This environment could comprise an office environment, a university building, or even a shopping mall. At each room we assume that a femto BS is located in a random location.

Simulation Parameters overview	
Number of floors	3
Number of rooms per floor	20
Number of HeNBs per room	1
Transmit power	23.0 dBm
Simulation time	60 s
Simulation time unit	0.1 s

TABLE 1 - SIMULATION PARAMETERS

The objective of the proposed analysis is to group the femto BS in a concise manner for enabling their cooperation. Such cooperation could be related to their cooperation for power control and energy efficiency, interference mitigation/compensation, or even load balancing among the BS. For the evaluation of the consistency of the clusters of each approach we have used the modularity metric. Modularity measuring the quality of the formed clusters, i.e. the division of a network into communities [7], which can be expressed by the following equation:

$$Q = \frac{1}{2L}\sum_{i,j}\left(A_{ij} - \frac{\deg_i \deg_j}{2L}\right)\delta(c_i, c_j) = \sum_{c=1}^{C}\left[\frac{l_c}{L} - \left(\frac{\deg_c}{2L}\right)^2\right] \quad (1)$$

where A_{ij} is the number of edges between nodes i,j; C represents the number of clusters; l_c is the total number of edges joining nodes of cluster c; $\deg_c$ is the sum of the degrees of the nodes of c; L denotes the number of edges of the network and $\delta(x,y) = 1$ when both x and y belong to the same cluster, otherwise $\delta(x,y) = 0$.

The modularity of a partition is a scalar value between − 1 and 1 that measures the density of links inside communities as compared to links between communities.

Figure1 presents the top view of all the three flours of the simulation environment. The different colors capture the positioning of each femto BS in each flour, black for the first, yellow for the second and red for the third. The femtocells have been placed in a random position, one in each room.

Assuming only the femtocells simple connectivity links among the nodes at sensitivity levels of the wireless cards of the nodes (i.e., -120 dB) all femtocells end up to a single cluster in all four methods. This occurs because all femtocells in the building sense each other. This implies that for each specific problem, we should propose different sensitivity levels, depending on the problem that is being addressed. Additionally, we should consider weights in each link; the RSRP measurements are used for this purpose, capturing how strongly a femtocell hears its neighbors.

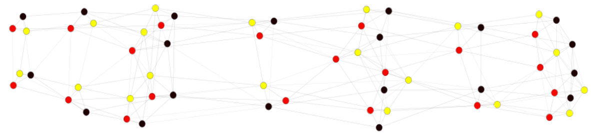

FIGURE1 - INITIAL TOPOLOGY REPRESENTATION

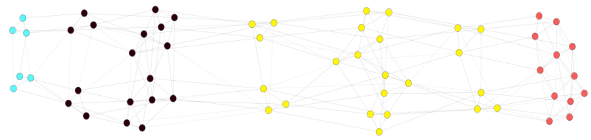

FIGURE2 – K-MEANS CLUSTERS FOR LINK SENSITIVITY THRESHOLD -80DB

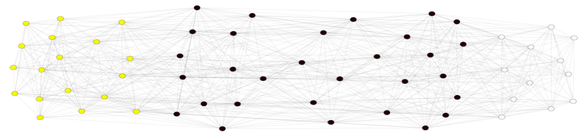

FIGURE3 – K-MEANS CLUSTERS FOR LINK SENSITIVITY THRESHOLD -90DB

Figure2 presents the formed clusters considering femtocells that are strongly sensed by each other. We assume as links among the femtocells, cells that are sensed with RSRP above of -80dB. This linking is suitable for interference mitigation and/or coordination because we link femtocells that are highly affecting each other. Figure3 presents a similar analysis with considered connections of -90 dB and above; such linking is more suitable for RRM problems (i.e., handovers, load balancing, power control etc.). These two clustering outputs were chosen based on the distinct and at the same time different outputs they produced. Network's representation on figure3 differs due to high nodes' connectivity.

Sensitivity Threshold	Algorithm family	Algorithm	No Clusters	Modularity
-80dB	Data mining	k-Means	4	0.497
		HAC	4	0.476
	Sensor	3hBAC	7	0.432
	Graph theory	MCL	6	0.172
-90dB	Data mining	k-Means	3	0.216
		HAC	3	0.185
	Sensor	3hBAC	2	0.273
	Graph theory	MCL	2	0.267

TABLE 2 - SIMULATION ANALYSIS

For the higher sensitivity thresholds we observe that smaller clusters are being formed. The reason for that is the fact that the connectivity grid related only nodes that are being sensed with high received signal power. We observe that with high power set threshold values, the clusters are more concise, based on the modularity metric, due to the well-formed previously mentioned connectivity graph. This is also the reason for the increased number of clusters in the first case. What should also be mentioned is that in general the formed clusters come from all three flours of the topology. This is a big benefit of the proposed approach, because all four schemes are not topologically biased, but only consider pure network links. Out of the four clustering schemes, in sparse connectivity graphs, the data mining schemes perform better, whereas in denser network environments the sensor and graph theory schemes outperform the data mining ones.

B. *Qualitative Comparative Analysis*

Based on the analysis of the algorithms of the previous section we have ended up with a set of conclusions regarding the problems that could be addressed with the considered solutions.

Table 3 summarizes our findings related to the applicability of the studied algorithms with regards to three main aspects, namely whether mobility during setup supported, the algorithms' (intuitive) complexity, and the required information. More specifically, we should highlight that only the sensor algorithms (i.e., 3hBAC) handle light mobility during the setup, whereas the core algorithms of the other categories do not support integration of new nodes during or after the setup of clusters. Derivatives of the data mining and graph theory algorithms have maintenance phase which enables the incorporation of new nodes after the cluster setup. In terms of the required calculations, the 3hBAC algorithm is the simplest among all the algorithms of the experimentation, whereas the data mining algorithms (k-Means, HAC) have moderate complexity and the graph one (MCL) is even more complex. The computational comparison among data mining algorithms and algorithms from graph theory derives from the fact that both require global knowledge (i.e., vector distance calculation and matrix multiplication).

Algorithm Family	***Algorithm***	***Mobility during setup***	***Complexity***	***Information required***
Data mining algorithms	k-Means	Not supported	Moderate	Global
	HAC	Not supported	Moderate	Global
Sensor algorithms	3hBAC	Light mobility supported	Simple	Local
Graph theory algorithms	MCL	Not Supported	Complex	Global

TABLE 3 - QUALITATIVE ANALYSIS OF THE CLUSTERING MECHANISMS

V. CONCLUSION

Clustering schemes in future network environments is a topic that has attracted the interest of the research community the recent years. This is mainly due to the massive deployment of micro, pico, and femto cells. In this paper, we followed a revolutionary approach and we proposed the use of algorithms coming from different research fields to networks. Therefore, we evaluated the application of a variety of clustering algorithms in small cells. In particular, we studied how deployed femtocells in ultra-dense networks should be grouped into concise clusters. The outcome of this experimentation is that all algorithms perform very well in terms of conciseness of the built clusters. Additionally depending on the problem (e.g., interference management, power control, RRM, etc.) we should follow different clustering methodology. The next steps of our work rely on the combination of these algorithms with the previously mentioned problems so as to evaluate their effectiveness on clustered topologies. Furthermore, femtos' mobility shall be considered given that the introduction of nomadic cells becomes really attractive.

ACKNOWLEDGMENT

This work has been performed in the framework of the FP7 project ICT-317669 METIS. The authors would like to acknowledge the contributions of their colleagues. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein.

REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015, Whitepaper

[2] David Soldani, "On the Advanced 5G Network Infrastructure for the Future Internet", Presentation in IEEE ICC 2013

[3] S. A. Saad, M. Ismail, and R. Nordin, "A Survey on Power Control Techniques in Femtocell Networks," Journal of Communications, vol. 8, no. 12, pp. 845-854, 2013. doi: 10.12720/jcm.8.12.845-854

[4] Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann (2011)

[5] Yu, J. Y., Chong, P. H. J., "3hBAC (3-hop between Adjacent Clusterheads): a Novel Non-overlapping Clustering Algorithm for Mobile Ad Hoc Networks", in Proc. IEEE Pacrim03, 31821, 1, (2003)

[6] Stijn van Dongen, Graph Clustering by Flow Simulation. PhD thesis, University of Utrecht, May 2000.

[7] U. Brandes,D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, D. Wagner, "On Modularity Clustering," IEEE Transactions on Knowledge and Data Engineering, vol.20, no.2, pp.172,188, Feb. 2008, doi: 10.1109/TKDE.2007.190689

[8] X. Zhu, M. Wang, Q. Zhang, X. Li: An Efficient Spectrum Allocation Scheme for OFDMA Femtocell Networks

[9] A. H. Abdalla, R. A. Saeed, M. K. Hasan, S. Islam, O. O. Khalifa: Cluster-Based multihop synchronization scheme for femtocell network, IIUM Engineering Journal (Volume 13, No. 2, 2012, Pages 163 to 175)

[10] Hatoum, A.; Langar, R.; Aitsaadi, N.; Boutaba, R.; Pujolle, G., "Cluster-based Resource Management in OFDMA Femtocell Networks with QoS Guarantees," Vehicular Technology, IEEE Transactions on , vol.PP, no.99, pp.1,1

[11] Abdelnasser, A.; Hossain, E.; Kim, D., "Clustering and Resource Allocation for Dense Femtocells in a Two-Tier Cellular OFDMA Network," Wireless Communications, IEEE Transactions on , vol.PP, no.99, pp.1,14

[12] Wei Li; Wei Zheng; Yuanbao Xie; Xiangming Wen, "Clustering based power saving algorithm for self-organized sleep mode in femtocell networks," Wireless Personal Multimedia Communications (WPMC), 2012 15th International Symposium on , vol., no., pp.379,383, 24-27 Sept. 2012

[13] Xin Tao, Zhifeng Zhao, Rongpeng Li, Jacques Palicot, Honggang Zhang, "Downlink interference minimization in cooperative cognitive LTE-femtocell networks", EURASIP Journal on Wireless Communications and Networking 2013,

[14] Hatoum, A.; Aitsaadi, N.; Langar, R.; Boutaba, R.; Pujolle, G., "FCRA: Femtocell Cluster-Based Resource Allocation Scheme for OFDMA Networks," Communications (ICC), 2011 IEEE International Conference on , vol., no., pp.1,6, 5-9 June 2011

[15] K.Sethom Ben Reguiga, F Mhiri and R.Bouallegue. Article: Handoff Management in Green FEMTOCELL Network. International Journal of Computer Applications 27(4):1-7, August 2011. Published by Foundation of Computer Science, New York, USA.

[16] Dan HU, Hong-jia LI, Xiao-dong XU, Xiao-feng TAO, "Inter-femtocell interference coordination in 3D in-building scenario", Elsevier, The Journal of China Universities of Posts and Telecommunications, April, 2012