# A Multi-stage Graph Approach for Efficient Clustering in Self-Organized Wireless Sensor Networks

Abhishek Karpate and Hesham H. Ali
*Department of Computer Science, College of Information Science and Technology*
*University of Nebraska at Omaha, Omaha, NE 68182, U.S.A.*

Keywords: Wireless Sensor Networks, Self-Organized Networks, Graph Modeling, Clustering, Multi-Stage Graph Algorithms, Energy-Aware Solutions.

Abstract: With the rapid increase in applications utilizing the current advancements of wireless sensor networks, a number of problems related to self-organization, energy-awareness and network organizations have attracted many researchers in the field. Various groups have proposed grouping the sensors into clusters and design communication routes in two levels as a way to improve communication cost and better organize networks of large sensors. In this paper, we propose a new approach to cluster wireless sensors and identify cluster heads using multi-stage graph algorithms. The approach takes advantage of the optimally associated with finding matching solutions in multi-stage graph networks. The proposed solution is designed to accommodate networks with different sizes and levels of density. We tested the algorithm using different types of networks and measure the quality of the key parameters as compared to those obtained by traditional greedy heuristics. Obtained results show that the multi-stage graph approach produces better network organization and better cluster head selection which leads to be more efficient self-organized networks.

## 1 INTRODUCTION

Wireless sensor networks (WSNs) typically consist of a large number of sensors; sensors are small wireless devices having limited resources like energy, processing speed and storage. With the recent technology advances it is possible to produce small and low cost sensors making it economically feasible to deploy sensors in large numbers. Sensors measure ambient conditions or measure certain environmental parameters and report it to processing nodes. Instead of each individual sensor being always active and directly reporting to the processing node, the sensors in the WSNs could be clustered in a way where different sensors play different roles. Clustering provides network scalability and network topology stability in addition to possible energy saving attributes. Due to the various schemes employed in clustering, there is reduction in communication overhead and interferences among the sensor nodes (Karaki et al. 2004, Mhatre et al. 2004, Jiang et al. 2009).

There are various ways in which clustering schemes can be classified. Clustering schemes are categorized depending on what objective the cluster intends to attain and what main algorithmic technique it employs. This includes dominating-Set based clustering, low-maintenance clustering, mobility-aware clustering, energy-efficient clustering, load balancing clustering, or combined based metrics clustering (Yu and Chong 2005). The clustering scheme has also been classified according to key cost associated parameters like explicit control message for clustering, ripple effect of re-clustering, stationary assumption for cluster formation, constant computation round and communication complexity (Yu and Chong 2005). Clustering in networks also depends on the type of network that is being considered. An alternate way to classify clustering in ad-hoc networks is based on the type of networks they are used to cluster; single-hop or multi-hop, location based or non-location based, synchronous or asynchronous (depending on the network topology) and stationary nodes or mobile nodes (Wei and Chan 2006). Clustering is performed on sensor networks which are either homogeneous, where all the sensor nodes are identical in built and functionality, or heterogeneous, when the network consists of sensors which differ from each other in built or functionality. Both categories of networks have to deal with the overhead of cluster construction process.


Karpate A. and Ali H..
A Multi-stage Graph Approach for Efficient Clustering in Self-Organized Wireless Sensor Networks.
DOI: 10.5220/0005244700560062
In *Proceedings of the 4th International Conference on Sensor Networks* (SENSORNETS-2015), pages 56-62
ISBN: 978-989-758-086-4

In this paper, we propose a new model for distributed clustering in heterogeneous networks that focuses on allocating sensors to cluster heads such that the total communication cost for the entire system is minimized. The suggested model also provides flexibility by which the density of networks and number of resource handling capacity of a device can be significantly varied.

# 2 PROPOSED MODEL

In a heterogeneous sensor network, self-organization continues to be a prominent feature due to increase in the complexity in managing the network as most of the routing paths are dynamically decided. Several models have been proposed to take energy awareness into consideration (Chamam and Pierre 2009, Zhange et al. 2007, Cardei et al. 2005). Using cluster heads in wireless sensor networks has been utilized to coordinate the process of collecting and reporting data. In such networks, there exist many sensors that measure the ambient conditions or collect various parameters and report the information they have sensed to cluster heads. The cluster heads, in turn, process the information it receives from all the sensors and further report it to the processing or sink nodes (Yu et al. 2011, Nayer and Ali 2004, Nayer and Ali 2008). The model we propose focuses on handling the problem of clustering sensor nodes to the cluster heads and further clustering heads to the sink nodes. The operation of node clustering can be modelled by graph clustering, which groups vertices of a graph into clusters based on certain criteria. Graph clustering can be broadly divided into two categories: global clustering and local clustering (Yu et al. 2011, Camilus et al 2008, Karpate and Ali 2011). The difference between the two types of clustering being that in global clustering every vertex on a graph is allocated to a cluster and in local clustering only a certain subset of vertices is allocated to a cluster. Applications like WSNs usually use global clustering.

## 2.1 The Multi-Stage Graph Model

Multi-stage graphs are usually used in cases where there is a connected graph optimization problem having several stages with each stage contains a set of nodes. The edges of the graph are used to connect nodes in different stages. There are no edges between nodes of the same stage or non-adjacent stage. The entire WSN can be modelled by a multi-stage graph having three stages as shown in Figure 1 where the first stage of nodes represents the set of sensors, the second stage represents the set of cluster heads and the third stage represents the set of sink nodes. An edge connects two nodes if that particular sensor (or cluster head) can communicate with the particular cluster head (or sink node). The weight on the edge represents the distance between the two nodes. The network can be represented by the multi-stage graph as follows:

$G = (N \cup C \cup S, E1 \cup E2)$ where,

$N \rightarrow$ Set of sensor nodes
$C \rightarrow$ Set of cluster heads
$S \rightarrow$ Set of sink nodes
$E1 \rightarrow$ Set of edges connecting N and C
$E2 \rightarrow$ Set of edges connecting C and S

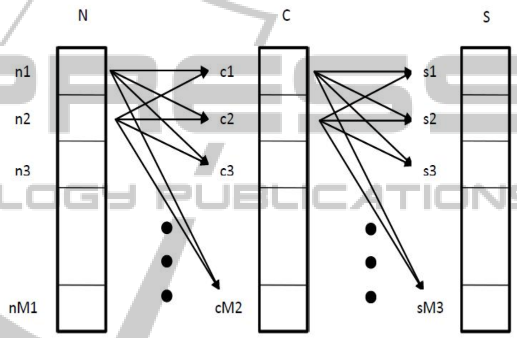


Figure 1: Multi-stage graph representation of Heterogeneous Sensor Network.

As shown in the figure, there are M1 sensor nodes, M2 cluster head nodes and M3 sink nodes. There exists an edge that connects n1 to c1, c2, c3 up to cM2 and all the edges have respective weights associated with them representing the distance between the two nodes. In the allocation process, the entire system is considered and considering the maximum number of sensors that can be allocated to a cluster head, appropriate edges are shortlisted and accordingly each sensor is allocated to some particular cluster head. A network could have more than one sink node depending on the size of the network; if the set of cluster heads and sink nodes are considered depending on the distance between a particular cluster head and sink nodes, it is clustered with one of the sink nodes.

# 3 THE MATCHING ALGORITHM

The task of allocating a sensor to a cluster head and allocating a cluster head to a sink node can be translated to a maximum matching problem. Maximum matching algorithm gives us an indepen-

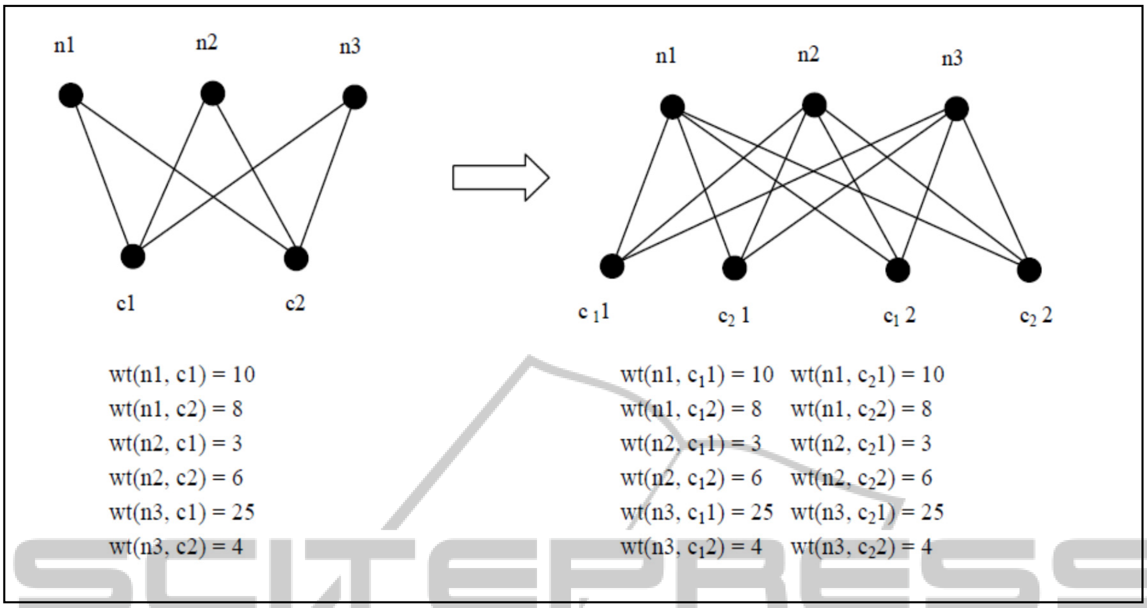


Figure 2: Expanded graph with replicated nodes.



dent edge set with no common vertices such that the combined weight of the edges selected is the maximum possible for that graph. Our problem is to find an optimal allocation of every resource in a stage to a resource in the next stage. In terms of graph theory, our problem can be termed as a minimum matching problem that is a set of independent edges where the combined weight is as minimum as possible. Consider any stage of the multi-stage graph like the first stage where the sensors report to cluster heads. Assuming a cluster head can handle data from *t* sensors. Replicate the set of cluster heads a number of *t* times such that for every sensor, every cluster head it was connected to is replicated t times. If we implement the minimum-matching algorithm in this case, we would have an optimal allocation of sensors to cluster heads. This concept will become clearer from the example in Figure 2, where t is given a value of 2 in the example.

On applying the algorithm on the expanded version, the set of independent edges selected based on weights are (n2, $c_1$1), (n3, $c_1$2) and (n1, $c_2$2). The total cost of allocation is 15. The execution is implemented as shown in algorithm 1 and algorithm2. The expanded graph is created by Agorithm1 and then is subjected to the proposed matching algorithm, described in Algorithm2 to attain the most effective allocation.

The proposed algorithms help in finding the most economical assignment of sensors to cluster head as a whole and also the most economical assignment of cluster heads to sink nodes. In terms of time complexity, the main algorithm is based on finding optimal matching in Bipartite graphs which is a fast algorithm that is quadratic in the number of nodes.

## 4 SIMULATION RESULTS

To illustrate the performance of the proposed algorithm, we have compared the outputs produced by the algorithm with the outputs produced using a robust/greedy graph approach. The greedy approach has a decent track record for dealing with the clustering problem. Using the greedy approach, at any stage when two sets of nodes are considered, a minimum weight edge is selected at each step. Both the models take the following inputs: number of sensors deployed, number of cluster heads deployed and the number of sink nodes available. The quality and hardware superiority of the networking device deployed determines how much data and from how many devices can it handle the data from.

For example the better the superiority of a cluster head the more number of sensors can report to it. Both the algorithms implemented permit the discussed flexibility by taking the maximum number of sensors a cluster head can handle and the maximum number of cluster heads a sink node can handle. The impact of the number of sensors deployed is also considered. Better redundancy and coverage can be expected with the increase in the number of sensors being deployed. In the figures below, we use the term robust algorithm to refer to the greedy approach and the combinatorial

1. Start.
2. Randomly distribute the sensors, cluster heads and sink nodes in the area being monitored.
3. Generate two arrays
   a. Array 1 representing the distance between sensor and cluster head with sensors representing rows and cluster heads representing columns.
   b. Array 2 representing the distance between cluster head and sink nodes with cluster heads representing rows and sink nodes representing columns
4. If the number of rows is greater than the number of columns in any of the array.
   a. Replicate the number of columns such that the required condition is satisfied.
   b. Replace the original array with the modified array.
5. Apply algorithm 2 to the arrays.

Algorithm 1: Proposed Combinatorial Algorithm.

1. Start
2. Set markedUncoveredZero flag to false.
3. Create new zeros.
   a. For each row, subtract its smallest element from all its elements.
   b. For each column, subtract its smallest element from all its elements.
4. Assign lone zeros.
5. Consider the following cases:
   a. If all the rows have been assigned.
      i. Then stop.
      ii. Display the allocation.
   b. If the matrix is not fully covered.
      i. Assign an uncovered zero.
      ii. Set markedUncoveredZero flag to true.
      iii. Go to step 4.
   c. If the matrix is fully covered.
      i. If markedUncoveredZero flag is false
         1. Create new zeros by subtracting the value of the smallest uncovered cost from all the uncovered costs.
         2. Add the smallest uncovered value to all the double-covered costs.
         3. Go back to step 4.
      ii. Else
         1. Mark all unassigned rows.
         2. Mark all unmarked columns that have zero in the marked rows.
         3. Mark all unmarked rows that have assignments in the marked columns.
         4. Repeat 2 & 3 until no changes are observed.
         5. Create new zeros by subtracting the value of the smallest uncovered cost from all the uncovered costs.
         6. Add the smallest uncovered value to all the double-covered costs.
         7. Go back to step 4.

Algorithm 2: Graph expansion Algorithm

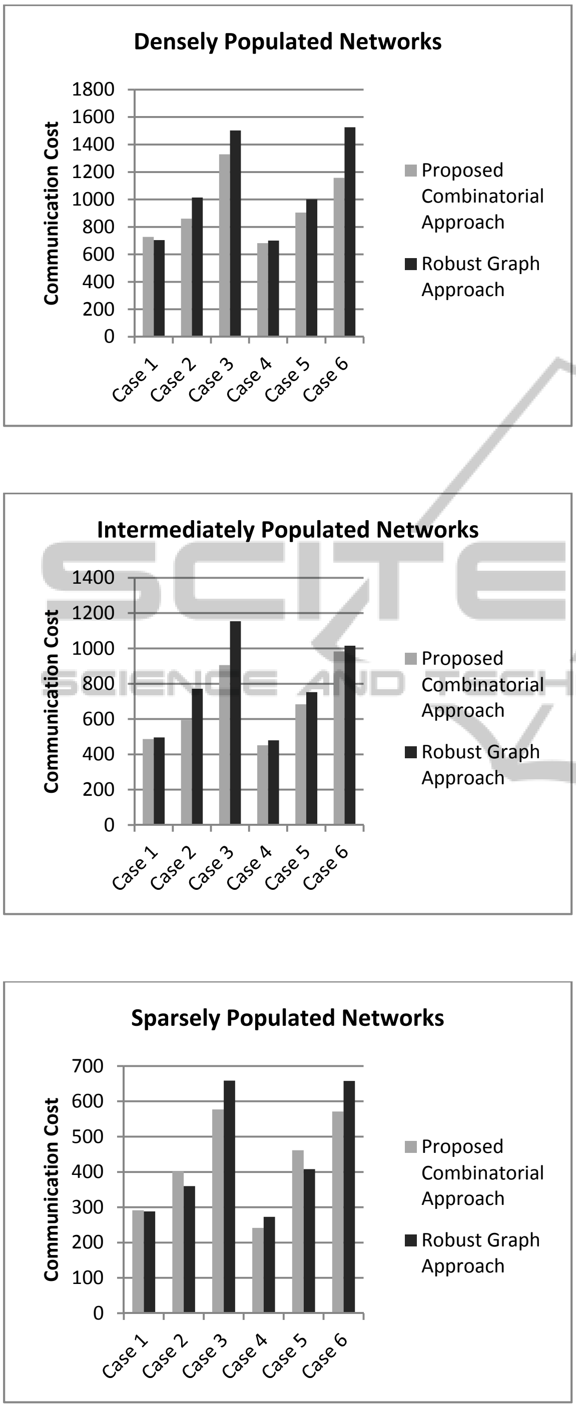


Figure 3: Graphs representing the communication cost for sensor - cluster head communication.

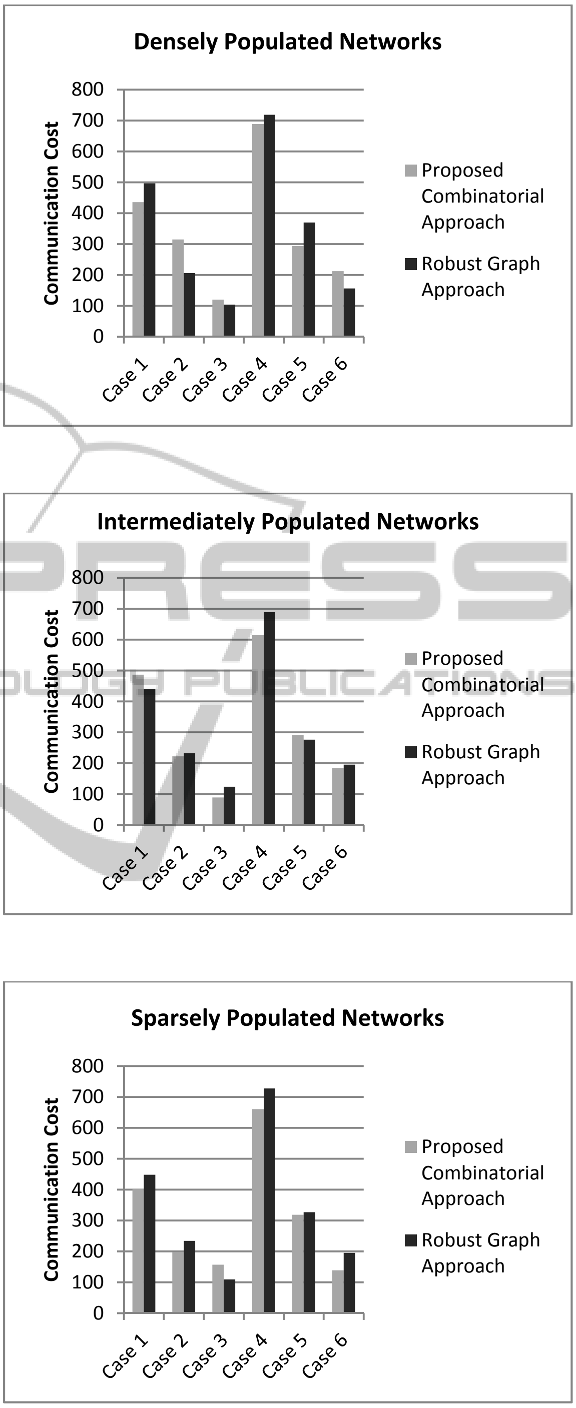


Figure 4: Graphs representing the communication cost for cluster head - sink communication.

algorithm to refer to the multi-stage graph-matching algorithm.

In our simulations there are three types of networks – densely populated networks, moderately populated networks and sparsely populated networks. A network is dense if it has more than 50 sensors, networks having sensors between 25 and 50 are moderately populated networks and networks having 25 or less sensors are sparsely populated networks.

The cost of communication between any two devices is directly proportional to the distance between them. As the distance between the two devices increases it is expected that the communication cost between the two will increase. The simulations are carried out assuming the cost of communication is one unit for every one meter. Sensors, Cluster heads and sink nodes are randomly distributed on the area to be monitored. The proposed algorithm is carried out on all types of network and the results obtained. Figures 3 and 4 show the two costs the network had to incur. Figure 3 compares the communication cost between sensor and cluster head for both the approaches for all the three networks under similar conditions and Figure 4 does the same for the communication cost between cluster heads and sink nodes. To test the networks under different conditions, the count of the number of cluster heads and number of sink nodes is changed in different cases. Case 1 has 20 cluster heads and 5 sink nodes are deployed. Case 2 -10 cluster heads and 5 sink nodes; case 3 – 5 cluster heads and 5 sink nodes; case 4 – 20 cluster heads and 2 sink nodes; case 5 – 10 cluster heads and 2 sink nodes and case 6 – 5 cluster heads and 2 sink nodes are deployed.

From the above graphs, we can observe that in most of the cases the proposed algorithm gives better solutions as compared to the robust/greedy graph approach. Another major advantage of the proposed algorithm over the robust approach is that it does not follow a greedy strategy by making choices based on a global overview. The robust approach makes choices that look the best at that moment. Although most of the times the attained optimal solution by the robust approach maybe at par with our proposed solution, it could fail at critical conditions and hence is not so reliable.

## 5 CONCLUSIONS

In this paper we propose a graph theoretic approach to efficiently form a weight-based cluster formation algorithm for wireless sensor networks. The network is self-organized such that any particular sensor while determining which cluster head to report not only considers its physical distance from the available cluster heads but also considers the receiving capacity of the available cluster heads and the physical distance of all other unallocated sensors from the available cluster heads. The same approach is used while determining which cluster head should report to which sink node. In this attempt we manage to find an allocation that consider the system as a whole and provides gives energy-aware solutions. The efficiency of the algorithm is measured by comparing it with a robust graph approach, which is a greedy approach for solving the problem. The efficiency of the proposed combinatorial optimization algorithm is better than that of the greedy algorithm in most of the cases; it also avoids the local short-sighted issues that are dealt with while using the greedy approach. One of the disadvantages it has over the robust graph approach is that it takes a longer time to process the best allocation of resources. The graph theoretic model can be further enhanced to increase the level of redundancy by deciding how many resources can a particular resource report to, this feature is very useful when the area being monitored is of high priority and if due to certain factors, some signals or data is lost it can be retrieved from an alternate source.

## REFERENCES


Al-Karaki, J. N., and Kamal, A. E., 2004. *Routing techniques in wireless sensor networks: a survey.* Wireless communications, IEEE, 11(6), 6-28.

Mhatre, V. and Rosenberg, C., 2004. *Homogeneous vs Heterogeneous Clustered Sensor Networks: A Comparative Study*, IEEE International Conference on Communications.

Jiang, C., Yuan, D., and Zhao, Y., 2009. *Towards clustering algorithms in wireless sensor networks-a survey.* In Wireless Communications and Networking Conference, IEEE.

Yu, J. Y., and Chong, P. H. J., 2005. *A survey of clustering schemes for mobile ad hoc networks.* IEEE Communications Surveys and Tutorials, 7(1-4), 32-48.

Wei, D., and Chan, H. A., 2006. *Clustering ad hoc networks: Schemes and classifications.* In Sensor and Ad Hoc Communications and Networks, 2006. SECON'06. 2006 3rd Annual IEEE Communications Society on (Vol. 3, pp. 920-926). IEEE.

Yu, C. H., Qin, S., Alava, M., and Tirkkonen, O., 2011. *Distributed graph clustering for application in wireless networks.* In Self-Organizing Systems (pp. 92-103). Springer Berlin Heidelberg.

Camilus, K. S., Govindan, V. K., and Sathidevi, P. S., 2008. *A Robust Graph Theoretic Approach for Image Segmentation.* In Signal Image Technology and Internet

Based Systems, 2008. SITIS'08. IEEE International Conference on (pp. 382-386). IEEE.
Karpate, A., and Ali, H. H., 2011. *An energy-aware genetic algorithm for managing self-organized wireless sensor networks.* In Wireless Days (WD), 2011 IFIP (pp. 1-6). IEEE.
Nayer, S. I. and Ali, H. H., 2004. *On Employing Distributed Algorithms and Evolutionary Algorithms in Managing Wireless Sensor Networks*. International Workshop in Wireless Ad Hoc, Sensor Networks.
Nayer, S. I., and Ali, H. H., 2008. *A dynamic energy-aware algorithm for self-optimizing wireless sensor networks.* Lecture Notes in CS 5343, IWSOS 08 (pp. 262-268), Springer Berlin Heidelberg.
Chamam, A., and Pierre, S., 2009. *On the planning of wireless sensor networks: Energy-efficient clustering under the joint routing and coverage constraint. Mobile Computing.* IEEE Transactions on, 8(8), 1077-1086.
Zhang, M., Chan, M. C. and Anand, A. L., 2007. *Coverage Protocol for Wireless Sensor Networks Using Distance Estimates. Sensor, Mesh and Ad Hoc Communications and Networks,* IEEE SECON 07, 183 – 192, 2007.
Cardei, M., Thai, M. T., Li, Y., and Wu, W., 2005. *Energy-efficient target coverage in wireless sensor networks.* In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE (Vol. 3, pp. 1976-1984). IEEE.