

Evolutionary design of wireless sensor networks based on complex networks

Andre S. Ruela ^{#1}, Raquel S. Cabral ^{*2}, Andre L. L. Aquino ^{#3}, Frederico G. Guimaraes ^{#4}

[#] *Department of Computer - Federal University of Ouro Preto
Ouro Preto, MG, Brazil*

¹ andrebardo@gmail.com

³ alla@iceb.ufop.br

⁴ frederico.g.guimaraes@gmail.com

^{*} *Department of Electrical Engineering - Federal University of Minas Gerais
Belo Horizonte, MG, Brazil*

² raquelcabral@gmail.com

Abstract—This work proposes a genetic algorithm for designing a wireless sensor network based on complex network theory. We develop an heuristic approach based on genetic algorithms for finding a network configuration such that its communication structure presents complex network characteristics, e.g. a small value for the average shortest path length and high cluster coefficient. The work begins with the mathematical model of the hub location problem, developed to determine the nodes which will be configured as hubs. This model was adopted within the genetic algorithm. The results reveal that our methodology allows the configuration of networks with more than a hundred nodes with complex network characteristics, thus reducing the energy consumption and the data transmission delay.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) represent an emerging technology that allows the monitoring and control of physical and environmental variables and conditions, such as temperature, sound, light, vibration, pressure, movement and pollution. A WSN consists of a great number of wireless autonomous devices, called sensor nodes, which work in a cooperative way to perform many different functions. These characteristics make the WSNs a promising technology in a wide range of application domains, for instance, biotechnology, industry, public health, and transportation ([1], [2]).

Despite its potential applicability, a WSN has several resource restrictions, such as low computational power, reduced bandwidth, and limited energy source. Therefore, algorithms for WSNs need to be carefully designed. Thus, sending a large amount of data can be problematic, causing excessive delay in response time, invalidating the data. Moreover, a large traffic on the network can diminish its lifetime. Due to these restrictions, in some cases, it is necessary to adopt specific infrastructure designs to balance network requirements keeping its functionality.

Generally, the phenomenon monitored is reported through to sink considering a specific routing strategy. Examples of routing strategies are depicted in Figs. 1–3. A simple naive, but inefficient, way of propagating information through the network is flooding (Fig. 1). In this case, the information is flooded to all sensors until it reaches the sink node [3]. This

strategy causes unnecessary communication, consequently, a large energy consumption and a high response time to deliver the data.

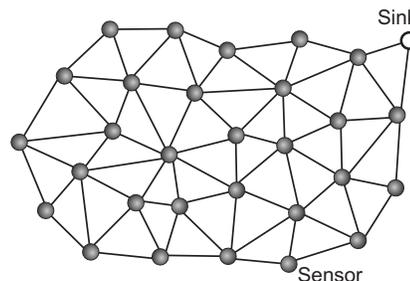


Fig. 1. Flooding propagation.

A common alternative to flooding is tree routing (Fig. 2), a simple and low-overhead routing protocol. Using a tree routing, each sensor is configured to send its data only to a specific sensor node, denoted father node. The choice of which node will be the father depends on the policy established by the application, in general, the shortest path policy is used [4]. The major drawback of tree routing is the increased hop counts as compared with more sophisticated path search protocols. However, there is a significant energy consumption because the link is kept, i.e., all non father nodes perceive the propagated data, this situation can be seen as light gray links in Fig. 2.

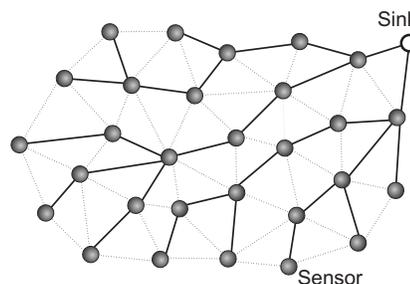


Fig. 2. Routing tree propagation.

Additionally, considering applications that use thousands of

nodes working independently and together [5], some strategies based on tree routing might not be scalable. An alternative routing strategy based on complex network characteristics (Fig. 3), consists in setting some sensor nodes as hubs, these ones using a communication radius greater than used by normal nodes. Normal nodes propagate their data to a given hub using a normal link frequency, and the hubs propagate the data to the sink node using a special link frequency. In both cases it is used a multi-hop communication. The use of these hub nodes lead to important characteristics of complex networks: a small average shortest path length between all sensors and the sink; and a high cluster coefficient, see [6]. This complex network characteristics help us saving network resources, avoiding excessive communication, and reducing the time to data delivery.

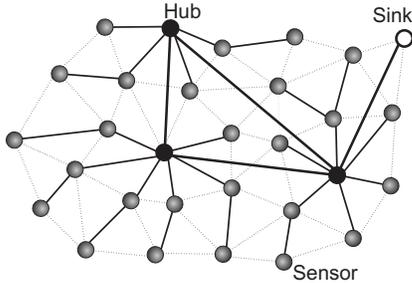


Fig. 3. Complex network propagation.

In fact, the proposed approach gives rise to a complex network, which are networks with irregular, complex and dynamic structure [7]. The theory of complex networks provides a mathematical framework for analyzing a number of real-world networks that otherwise could not be addressed with the available traditional models. The theory of complex networks can be useful in the study of WSNs, given some of its peculiar characteristics, such as the quantity, increase, and distribution of nodes in the network.

In this paper, we present a genetic algorithm for designing the logical topology for data propagation in a WSN. The goal here is to produce a logical topology that presents a high clustering coefficient and a small average shortest path length, which are two independent metrics from the complex network theory for quantifying structural features of the network. Exact methods are limited to small size networks, but for networks with hundreds or thousands of sensor nodes, the genetic algorithm becomes more interesting than the exact method from a practical point of view.

This paper is organized as follows. Section II shows the problem definition and its mathematical formulation. Section III, presents the solution based on genetic algorithms. Section IV, discusses about the results obtained using the genetic algorithm. Finally, Section V concludes the paper.

II. PROBLEM DEFINITION

Consider a WSN as a regular graph $G = (V, E)$, where V represents the set of sensor nodes and E the set of edges, representing the logical links between nodes. With this, the

problem addressed in this work can be stated as follows:

Problem statement: *Given a WSN, find the nodes $v \in V$ that should be reconfigured as hubs, generating a new set of edges E' , such that a new network $G' = (V, E')$ presents complex network characteristics.*

For instance, the complex network characteristics considered are only: the small shortest path length and a high cluster coefficient. However, the main hypothesis considered over the problem is:

Main hypothesis: *By using a WSN with complex network characteristics it is possible to minimize the energy consumption and time in data delivery.*

A. Problem Formulation

In this subsection, we present the mathematical formulation of the single allocation problem in sensor networks, which is an approach to the problem stated above.

Given a network with a set of nodes, the problem consists of finding the nodes that will be reconfigured as hub nodes and also the logical links that should be established in order to minimize the total cost. Let N be the set of normal nodes and H be the set of hub nodes, such that $N \cup H = V$ and $N \cap H = \emptyset$. The parameters of our mathematical model are:

- ϕ_i communication demand, i.e., the total amount of data that node i must send to the sink;
- r basic communication radius;
- d_{ij} distance between node i and node j ;
- c_{ij} fixed communication cost per data unit from node i to node j ;
- a_j fixed installation cost of node j as a hub node. It is inversely proportional to the distance between j and the sink, i.e., the higher the distance from j to the sink, the lower the installation cost;

The decision variables of the mathematical model are:

- $z_i \in \{0, 1\}$: $z_i = 1$ if node i is defined as hub, and $z_i = 0$, otherwise;
- $q_{ij} \in \{0, 1\}$: $q_{ij} = 1$ if there is a logical link between nodes i and j , and $q_{ij} = 0$, otherwise;

A nonlinear integer programming formulation of the problem defined above is given by:

$$z^* = \arg \min C_I + C_P \quad (1)$$

subject to:

$$\sum_{j \in V} q_{ij} = 1, \forall i \in N \quad (2)$$

$$\sum_{j \in H} q_{ij} \leq 1, \forall i \in H \quad (3)$$

$$q_{ij} \leq z_j, \forall i \in N, \forall j \in H \quad (4)$$

$$d_{ij} q_{ij} \leq r, \forall i \in N, \forall j \in H \quad (5)$$

$$d_{ij} \leq 2r, \forall i \in H, \forall j \in H \quad (6)$$

$$z, q \in \{0, 1\} \quad (7)$$

where the installation cost C_I is given by:

$$C_I = \sum_{i \in V} a_i z_i,$$

and the propagation cost C_P is given by:

$$C_P = \sum_{i \in H} \left(\sum_{j \in N} \phi_j c_j q_{ji} + \phi_i \right) \times \left(\sum_{k \in H} (c_{ik} q_{ik} + c_{k0}) \right)$$

The objective function (1) gives the total cost for establishing the hub network. This total cost includes the installation cost C_I and the propagation cost C_P . The first term in C_P represents the total amount of data that the hub $i \in H$ must send to the sink, which is multiplied by the cost of sending the data to the sink through hub k , where k may be equal to i or may be a different hub. This gives two or three hops from any node in the sensor network to the sink.

The objective function (1) is subject to following constraints:

- (2) Guarantee that a node $i \in N$ is connected to only one hub.
- (3) Guarantee that data from the hub i is either routed through one hub or directly to the sink.
- (4) Ensure that data from node $i \in N$ is only routed through a hub node.
- (5) Ensure that the distance between the node i and a hub node j is less than or equal to the communication radius.
- (6) Ensure that all hubs are within the range of all other hubs, using twice the value of the communication radius.
- (7) Restrict the values of the integer variables z_i and q_{ij} to assume either 0 or 1.

III. SOLUTION BASED ON GENETIC ALGORITHMS

In the context of combinatorial optimization problems, many metaheuristic techniques have emerged in the last thirty years [8], including Genetic Algorithms. In this paper, we present a GA to solve the problem presented in the previous section, the design of the logical routing structure in a WSN based on complex network theory.

A. Initial considerations

Our GA uses binary encoding to represent the network configuration, representing the nodes that are configured as hubs. Each individual $p_t^{(i)}$ in the population P_t is represented by a binary vector of V bits, where V is the number of sensor nodes, $p_t^{(i)}[k] = 0$ indicates that the node k is not configured as a hub and $p_t^{(i)}[k] = 1$ indicates that the node k is configured as a hub in the network structure encoded by this individual. Therefore, the encoding scheme represents a candidate routing structure.

We can define the following heuristics in order to eliminate the decision variables q_{ij} in the GA:

- Given an individual $p_t^{(i)} \in P_t$, we obtain the set of nodes in N and the set of hubs in H ;

- As a general rule, every node $i \in N$ sends its data to the hub $j \in H$ with the smallest cost c_{ij} , usually the closest one. In this way, we automatically have the corresponding $q_{ij} = 1$;
- From each one of the hubs and the sink, the routing is obtained using Dijkstra's algorithm, respecting the maximum number of hops allowed.

With these simplifications, the variables q_{ij} are implicitly calculated from the hub allocation provided by an individual $p_t^{(i)}$, thus simplifying the computation of the propagation cost C_P . The computation of C_I is simply given by the summation of 1's in $p_t^{(i)}$. Moreover, constraints (2)-(4) can be neglected in the model. We consider only the following constraints:

- The distance between a node $i \in N$ and its associated hub should be smaller than the communication radius r , in other words, every node $i \in N$ should have a hub within its communication range.
- The distance from every hub to each other should be smaller than twice the communication radius.

B. Simplified model

The considerations given above lead to the following model used for the GA:

$$\min C_I + C_P \quad (8)$$

subject to:

$$d_{ij} \leq r, \forall i \in N, j = \arg \min_h \{c_{ih}, h \in H\} \quad (9)$$

$$d_{ij} \leq 2r, \forall i, j \in H \quad (10)$$

If one of these constraints is violated, then the value of the objective function is penalized giving the fitness function:

$$f(p_t^{(k)}) = C_I + C_P + w \sum_{i \in N} \max [d_{ij} - r, 0]^2 + w \sum_{u, v \in H} \max [d_{uv} - 2r, 0]^2, \quad (11)$$

with $j = \arg \min_h \{c_{ih}, h \in H\}$.

C. Basic operators

The genetic algorithm implemented use tournament selection for the reproduction, two-point crossover with probability $\rho_c = 0.8$ and bit-flip mutation with probability $\rho_m = 0.1$, see ([9], [10]) for further details.

The genetic algorithm for network configuration (Algorithm 1) considers:

- **Line 2:** Initialization operation.
- **Line 3:** The random initialization of a population with μ individuals that encode candidate configurations for the problem. At each generation t , the population undergoes the usual genetic operators, producing new candidate solutions for the problem.
- **Lines 4-12:** The loop that control the maximum number of interactions.
- **Line 5:** The new fitness is calculated.

Algorithm 1: Genetic algorithm for network configuration

Data: V – Number of nodes, \mathbf{a} – Installation cost vector, \mathbf{D} – distance matrix, \mathbf{C} – propagation cost matrix, ϕ – communication demand vector, μ – population size, ρ_m – mutation rate and ρ_c – crossover rate.

Result: solution with the set of hub nodes, and its related total cost.

```
1 begin
2    $t \leftarrow 0$  {generation counter}
3    $P_t = \{p_t^{(1)}, \dots, p_t^{(\mu)}\} \leftarrow \text{Initialize}(V, \mu)$ 
4   while  $t < t_{\max}$  do
5      $F_t \leftarrow \text{Evaluate}(P_t, \mathbf{a}, \mathbf{D}, \mathbf{C}, \phi)$ 
6      $t^*, p_{best}^* \leftarrow \text{Update Best}(P_t, F_t)$ 
7      $P'_t \leftarrow \text{tournament selection}(P_t, F_t)$ 
8      $P''_t \leftarrow \text{Crossover}(P'_t, \rho_c)$ 
9      $P'''_t \leftarrow \text{Mutation}(P''_t, \rho_m)$ 
10     $P_{t+1} \leftarrow P'''_t$ 
11     $t \leftarrow t + 1$ 
12  end
13 end
```

- **Line 6:** The best population is updated.
- **Line 7:** Individuals are selected for the reproduction stage using binary tournament selection, in which two individuals are randomly selected from P_t , and compete against each other. The one with the best fitness value wins the tournament and goes to P'_t , thus having a chance to reproduce. The best individuals are more likely to be selected for reproduction, but poor individuals also have some chance to be selected.
- **Lines 8–9:** A new population is formed from the selected individuals, using the crossover and mutation operators.
- **Lines 10–11:** Incremental operations.

It should be emphasized that, although the genetic operators are random, the genetic algorithm is far from being a purely random search, because the selection operator has a deterministic component that guides the searching process.

IV. COMPUTATIONAL RESULTS

The genetic algorithm searches for a WSN configuration that meets the complex network characteristics, i.e., a network having a small average path length between every node and the sink and a large clustering coefficient. Thus the data traffic is reduced, and consequently the energy consumption and the data delivery time are reduced. Initially, we present some general assumptions for the genetic algorithm evaluation:

- **Simulations:** The simulations were performed with the algorithm implemented in Java. The genetic algorithm was executed considering a population of $\mu = 50$ individuals, for 500 generations. The number of necessary simulations is given by [11],

$$\text{rounds} = \left(\frac{100 z s}{p \bar{X}} \right)^2,$$

where z is a constant of value 1.96, s is the standard deviation found in the first simulations, \bar{X} is the average of the obtained values and p is the percentage of the average that we want to get as deviation, which in this case was 5%.

We consider 30 rounds with random topologies, and for each topology we executed the genetic algorithm 33 times and the results are presented with symmetrical asymptotic confidence interval of 95%. The tests are executed in a machine Intel Quadcore 2.4GHz with 2GB RAM.

- **Network topology:** The network density is kept constant, the area is

$$A = \pi r^2 |S|/d,$$

where r is the radio range, $|S|$ is the number of sensor nodes and d is network density (arbitrarily chosen with the value 31.4791).

The nodes started the execution with the same hardware configuration, at the end the hub nodes reconfigure the radio range based on the infrastructure solution. Thus, the final solution has a heterogeneous WSN.

- **Resultant configuration:** Considering $G = (V, E)$ the initial WSN graph and $G_b = (V, E_b)$ the graph return by the genetic algorithm. The resultant configuration, used in a real network will be the graph $G' = (V, E')$, where $E' = E \cup E_b$. Thus, the cluster coefficient and the average shortest path length are calculated considering those graphs. However, the number of installed hubs and the time to compute the results is showed.

Table I shows the results of the genetic algorithm with the variation of the number of nodes (128, 256, 512, 1024). The evaluated parameters are: (i) number of installed hubs; (ii) execution time in seconds; (iii) cluster coefficient comparing the initial regular graph G and the complex graph G' ; and (iii) average shortest path length comparing the initial regular graph G and the complex graph G' .

TABLE I
RESULTS CONSIDERING THE VARIATION OF THE NUMBER OF NODES (#).

#	Hubs	Time(s)	Cluster coefficient		Shortest path	
			regular	complex	regular	complex
128	11	44	0.689	0.691	2.550	2.659
256	24	353	0.657	0.650	3.381	2.729
512	68	4394	0.633	0.655	3.739	2.619
1024	252	49119	0.624	0.706	3.558	2.475

The results presented in Table I show that, with our genetic algorithm, it is possible to build a WSN with two specific complex network characteristics, the high cluster coefficient and the low average shortest path length. As we can see, the cluster coefficient of the graph G' is roughly the same or slightly higher than that of the original regular graph G , and the average shortest path length of the graph G' for the networks with sizes (256, 512, 1024) was reduced by 20%, 30%, and 31% respectively, when compared with the original regular graph G . These complex network characteristics

required the installation of 11, 24, 68, and 252 hubs¹ for the networks with sizes (128, 256, 512, 1024), respectively, which gives a percentage of about 9%, 9%, 13%, and 25% of nodes installed as hubs.

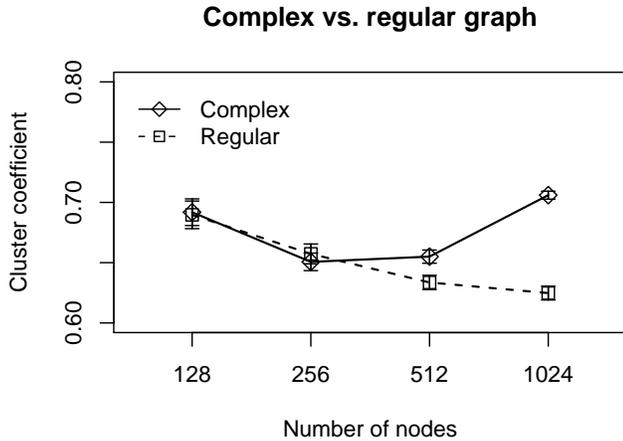


Fig. 4. Cluster coefficient analysis.

Specifically, in Fig. 4 the cluster coefficient results are analyzed. In this case, when the number of nodes in the network is increased, the cluster coefficient of the resultant graph G' slightly increases with the number of nodes, while that of the graph G always decreases. The installation of hubs increases this coefficient because the normal nodes are connected only to hub nodes. Considering the WSN aspects, a high cluster coefficient reduces data delivery delays and unnecessary energy consumption by concentrating the data sensing in a given hub node. Nevertheless, undesired interferences and link layer processing are avoided because the communication between hubs employs a different frequency range.

In Fig. 5 the average shortest path length results are presented. In this case, when the number of nodes in the network increases, the average shortest path length slightly decreases in the resultant graph G' , while it increases in the graph G . This occurs because the installation of hubs with a high radio range diminishes the shortest path in the network. Considering the WSN aspects, a low average shortest path length avoids, mainly, the data delivery delay. The drawback, in this case, is that when the radio range is modified more energy is consumed, but considering the global energy consumption this approach can actually save energy.

Finally, Fig. 6 shows the ratios of the cluster coefficient and the average shortest path length between the complex and regular graphs. In this Figure, CC and SP represent the cluster coefficient and the average shortest path length. In this case, when the number of nodes in the network is increased, the ratio of the average shortest path length is decreased while the ratio of the cluster coefficient remains close to one. This

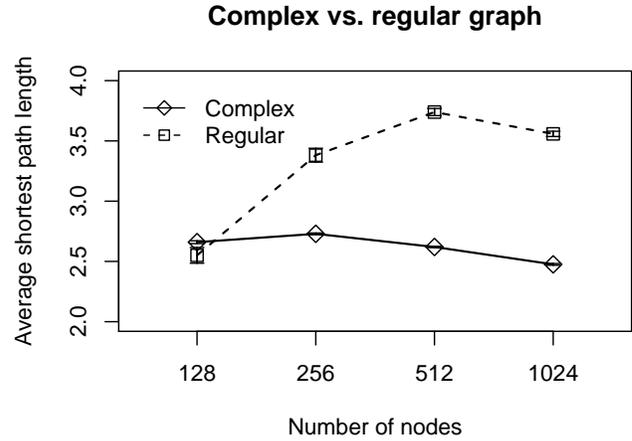


Fig. 5. Average shortest path length analysis.

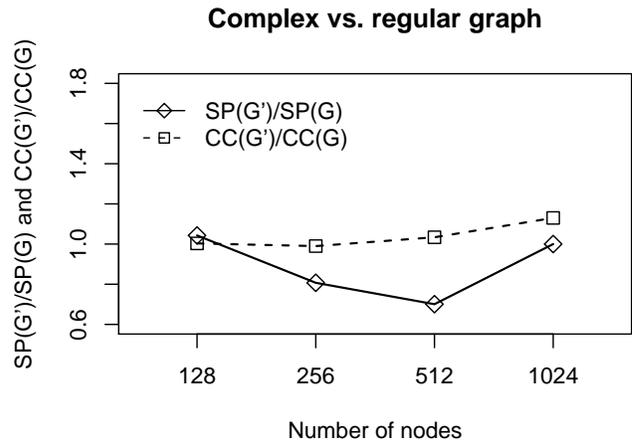


Fig. 6. Ratios of the cluster coefficient and the average shortest path length between the complex and regular graphs.

behavior is evidence that the WSN obtained in this work can be a small world network [12], [13] or a scale free network [14], [15], but this assumption should be investigated more deeply.

V. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This work presented a genetic algorithm for designing the logical topology for data propagation in a WSN. The goal was to produce a logical topology that presents a high clustering coefficient and a small average shortest path length, which are two independent metrics from the complex network theory for quantifying structural features of the network. For a network with 1024 nodes, the genetic algorithm was satisfactory, obtaining a network that satisfies the complex network characteristics.

The results showed that our genetic algorithm can find a WSN design with two specific complex network characteris-

¹Nodes with differentiated radio range.

tics, the high cluster coefficient and the low average shortest path length. This was highlighted in our results that showed that the cluster coefficient of the resultant graph is the same or slightly higher when compared to the original regular graph, and the average shortest path length of the resultant graph, in our specific scenario, was reduced by 20% or 30% when compared to the original regular graph.

This complex network strategy is important in WSNs because a high cluster coefficient avoids the data delivery delay and unnecessary energy consumption by concentrating the data sensing in a given hub node. Interferences and link layer processing are avoided when different communication frequencies are used between hubs. The low average shortest path length avoids, mainly, the data delivery delay but on the other hand more local energy is consumed. This discussion shows the truthfulness of the *Main hypothesis* presented in Section II. Finally, features of more specific complex networks, e.g. small world or scale free networks, can be easily incorporated in the design of WSNs.

As future work we intend to implement an exact solution to show the efficiency of the genetic algorithm. Additionally, we intend to simulate the solutions provided by the genetic algorithm using available network simulators, in order to emphasize the advantages when using our solution design. Finally, we intend to implement a specific local search operator to improve the efficiency and robustness of the genetic algorithm implemented.

ACKNOWLEDGES

This work is partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under the grant number 477292/2008-9.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.
- [2] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in *13th IEEE Mediterranean Conference on Control and Automation (MED'05)*. Hawaii, USA: IEEE Computer Society, June 2005, pp. 719–724.
- [3] M. Maroti, "Directed flood-routing framework for wireless sensor networks," in *5th ACM/IFIP/USENIX International Conference on Middleware (Middleware'04)*, vol. 78. Toronto, Ontario, Canada: ACM, October 2004, pp. 99–114.
- [4] W. Qiu, E. Skafidas, and P. Hao, "Enhanced tree routing for wireless sensor networks," *Ad Hoc Networks*, vol. 7, no. 3, pp. 638–650, May 2009.
- [5] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *26th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)*, vol. 4. Salt Lake City, Utah, USA: IEEE Computer Society, June 2001, pp. 2033–2036.
- [6] G. Sharma and R. Mazumdar, "Hybrid sensor networks: A small world," in *6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobHoc'05)*. Urbana-Champaign, Illinois, USA: ACM, May 2005, pp. 366–377.
- [7] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks : Structure and dynamics," *Physics Reports*, vol. 424, pp. 175–308, 2006.
- [8] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268–308, September 2003.
- [9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, 1989.
- [10] M. Mitchell, *An Introduction to Genetic Algorithms*, ser. Complex Adaptive Systems. The MIT Press, 1998.
- [11] R. K. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*, C. P. Association, Ed. John Wiley & Sons, April 1991.
- [12] M. E. J. Newman and D. J. Watts, "Scaling and percolation in the small-world network model," *Physical Review E*, vol. 60, pp. 7332–742, 1999.
- [13] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440–442, 1998.
- [14] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.
- [15] A.-L. Barabasi and E. Bonabeau, "Scale-free networks," *Scientific American*, pp. 50–59, 2003.