Computer Networks 110 (2016) 154–167

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

On the deployment of large-scale wireless sensor networks considering the energy hole problem

Heitor S. Ramos[a,b,1,*], Azzedine Boukerche[b,1], Alyson L.C. Oliveira[a], Alejandro C. Frery[a,1], Eduardo M.R. Oliveira[c,1], Antonio A.F. Loureiro[c,1]

[a] Institute of Computing, Federal University of Alagoas, Maceió, AL, Brazil
[b] DIVA Research center, University of Ottawa, Ottawa, ON, Canada
[c] Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil

ARTICLE INFO

Article history:
Received 12 February 2016
Revised 17 August 2016
Accepted 18 September 2016
Available online 19 September 2016

Keywords:
Wireless sensor networks
Network topology
Complex networks
Stochastic point process
Heterogeneous wireless sensor networks

ABSTRACT

Heterogeneous sensor networks have been proposed to address some fundamental limits and performance issues present in homogeneous Wireless Sensor Networks (WSNs). Questions such as the number of high-end sensors should be used, and how to deploy them, need proper assessment. In this work, we propose a novel model capable of representing a wide variety of scenarios, from totally random to planned stochastic node deployment in both homogeneous and heterogeneous sensor networks. In particular, this model encompasses networks with the characteristics of small-world networks. Using only about 3% of high-end sensors, and deploying nodes by using the slightly attractive model defined herein, we observe improved characteristics of the network topology, such as: (i) low average path length, (ii) high clustering coefficient, and (iii) improved relay task distribution among sensors. We also provide a guide for deploying nodes in order to improve the network lifetime, showing that the aforementioned model can be used to diminish the energy hole effect. Moreover, we evaluate a topological metric, namely Sink Betweenness, suitable for characterizing the relay task of a node.

1. Introduction

Node deployment and the consequent induced topology both play an important role in the design of Wireless Sensor Networks (WSNs). Many important properties such as coverage, connectivity, data fidelity, and lifetime are directly influenced by the way nodes are placed in the sensor field.

Most WSN models in the literature assume that the network is comprised of homogeneous nodes; that is, that all sensors have the same capabilities in terms of energy, processing, memory, and communication. However,Yarvis et al. [37] show that homogeneous ad hoc networks suffer from fundamental limitations, and hence, exhibit poor network performance such as end-to-end success rate, latency and energy consumption. Another class of WSN models assumes the existence of different sets of nodes with different capabilities. For instance, suppose there are two types: the first is comprised of a small number of powerful high-end sensors (H-sensors), and the second is comprised of a large number of low-end sensors (L-sensors). In this case, we have a Heterogeneous Sensor Network model [37]. The H-sensors have higher powerful transmitters, and consequently have a greater communication radius; in addition, they present a higher battery capacity.

Wu et al. [36] showed that the lifetime of a uniformly deployed WSN is strongly limited by the sensors at the first hop from the sink, an issue known as the “energy hole problem”. This problem ensues from the relay task, which concentrates more intensively on nodes placed close to the sink. The energy hole problem is also present in heterogeneous networks. In this case, it appears in the neighborhood of each H-sensor and the sink.Wu et al. [36] concluded that simply randomly increasing the number of nodes cannot elongate the network lifetime as desired when a totally random deployment is used. They show that the entire network lifetime can be improved by spreading more nodes near the sink.

An important task in the development of energy-aware solutions for WSNs is the design of efficient techniques for the creation of heterogeneous network topologies with specific properties. Complex networks [27] can be used to model networks with

* Corresponding author.
E-mail addresses: heitor@ic.ufal.br, ramosh@acm.org (H.S. Ramos), boukerch@site.uottawa.ca (A. Boukerche), alysonlcoliveira@gmail.com (A.L.C. Oliveira), acfrery@pq.cnpq.br (A.C. Frery), edumucelli@dcc.ufmg.br (E.M.R. Oliveira), loureiro@dcc.ufmg.br (A.A.F. Loureiro).

[1] This work is partially supported by NSERC DIVA Network, Canada Research Chair program, and the National Council for Scientific and Technological Development (CNPq), Brazil.

http://dx.doi.org/10.1016/j.comnet.2016.09.013
1389-1286/

certain non-trivial topological features, such as heavy-tailed degree distribution, high clustering coefficients, community structures at different scales, and evidence of a hierarchical structure, among others. The two most well-known examples of complex networks are those of scale-free and small-world networks. In a scale-free network, a vertex degree obeys a power law distribution, while a small-world network has a high clustering coefficient and a small path length [27]. Small-world networks present interesting characteristics with regards to data communication [17], such as a shorter average path length and a higher clustering coefficient (which can improve fault-tolerance properties). To create a network with small-world features, the designer should add a small number of long-range links, known as shortcuts.

In this work, we introduce a novel deployment model for WSNs. This model can represent topologies that exhibit desirable characteristics for WSNs. Furthermore, an adequate deployment can properly address the energy hole problem. We also evaluate a centrality metric, namely Sink Betweenness, introduced in our previous work [28,30,31], which is able to characterize the energy hole problem, and can be used in the design of algorithms for WSNs.

This work is organized as follows. Section 2 discusses the related work that motivates this research. Section 3 presents a centrality metric used to characterize topologies represented by the M^2P^2 model, and introduces the M^2P^2 deployment model proposed herein. Section 4 presents the evaluation of the M^2P^2 model. Finally, Section 5 presents the concluding remarks and future directions for this work.

2. Related work

Social network analysis (SNA) metrics have been used in the design of ad hoc networks. Katsaros et al. [19] surveyed a variety of protocols that take advantage of centrality metrics and community formation to improve network performance. They concluded that SNA metrics played a key role for the advancement of the network performance, especially when the communication was opportunistic in nature.

Helmy [17] and Guidoni et al. [15] exploited the possibility of implementing long-distance links between few nodes to construct WSNs that presented small world characteristics, such as small path lengths and high clustering coefficients, and hence, improved their performance. Those studies considered that long distance links could be established between any two nodes, and then, used only a few of them to create the small world characteristics. In contrast, we use less high powerful nodes, because we plan their deployment a priori. As we will see, our deployment is planned but not deterministic. We also provide a guide for deploying real-world WSNs following our model and preserving the improved features.

Matthias et al. [24] explored the small-world effect by building a topology control algorithm that worked with local information, instead of computing metrics such as clustering coefficients and average path-lengths that require global data. In contrast, we do not tackle the topology control problem, but instead plan deployments to build topologies that feature small-world properties.Vázquez-Rodas and de la Cruz Llopis [34] proposed a new technique for topology control based on centrality metrics, borrowed from social network analyses such as degree, closeness and Betweenness.

Younis and Akkaya [38] presented a survey on strategies and techniques for node placement in WSNs. They also propose a classification for different deployment methods. The first criterion is whether a node is static or mobile. Two deployment strategies are considered for static nodes: controlled and random. A controlled deployment is usually appropriated to indoor applications, or whenever the designer is able to specify the placement of all nodes, whereas random location is usually pursued in applications in which the designer is unable to precisely place the sensor nodes. The latter scenario assumes that sensors will be randomly placed; for instance, they can be dropped from a helicopter or an airplane. The authors [38] also suggested that the deployment can be optimized for the following metrics: (i) area coverage, (ii) network connectivity, (iii) network longevity, and (iv) data fidelity. Meanwhile nodes can assume the following roles: sensor, relay, cluster-head, and base station.

Topology influences the core characteristics of WSNs. Despite this, studies involving different topologies are seldom found in the literature. Actually, uniform random placement (URP) is the most widely used deployment strategy in WSN simulations [35,38]. To bridge this gap, we define a novel deployment model as a stochastic point process, i.e., a collection of random variables capable of describing the location of a number of points in a region of the space. For the sake of simplicity, let us assume that we are interested in stochastic point processes on the compact window $W = [0, \ell]^2 \subset \mathbb{R}^2$, where ℓ is the side length of the sensor field. A fixed number of n points obeys a URP distribution on W if they are placed independently of each other. A sample from such a process can be built observing outcomes from $2n$ independent identically distributed random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$, obeying the uniform law on $[0, \ell]$, say $x_1, \ldots, x_n, y_1, \ldots, y_n$, and then placing the n points on coordinates $(x_i, y_i)_{1 \le i \le n}$.Younis and Akkaya [38] stated that the URP assumption can be unrealistic or even undesirable for WSN scenarios.

Hoydis et al. [18] presented a study on the effects of the topology on the local throughput capacity of the slotted aloha MAC protocol in the context of ad hoc networks, and observed that URP deployment negatively affected network capacity and performance. They used cluster-based point processes [5] to conclude that simulations and analytic calculations done under the URP hypothesis may lead to incomplete findings and insights. They pointed out the necessity of further studies and consideration of other deployment models beyond the URP.

Strübe et al. [33] discussed the importance of better representation of real-world deployments such as fragile hardware and misbehaving software. They suggested analyzing the actual environment conditions of a deployed network, and mapping them to a simulator. Our article proposes a deployment model that can be useful for this purpose, i.e., a wide variety of real-life network topologies can be represented by this model, and can be embedded to a simulation tool. Ducrocq et al. [11] showed that the topology can greatly affect the routing performance, and studied this phenomenon using geographic routing algorithms [8,26]. They showed that the topology can influence up to 25% of the delivery ratio and average route length, and up to 100% of the overall cost of transmissions.

Haenggi et al. [16] presented a tutorial that showed a study involving the modeling of random node placement based on, among others, a Poisson point process. They argued that stochastic geometry and random graph theory are indispensable tools for the of wireless networks analysis, and that such tools lead to analytic results on a number of concrete and important problems. For instance, they applied those techniques to model and quantify interference, connectivity, outage probability, throughput, and capacity of wireless networks deployed as Poisson point process.

Li and Mohapatra [20] and Mohapatra [25] presented one of the first mathematical models towards the characterization of the energy hole problem. They considered sensor nodes distributed following the URP law (see Section 2) in a circular region, divided in concentric coronas. They studied the impact of four factors on the network performance: node density, hierarchical deployment, source bit rate, and traffic compression, and they showed that the use of hierarchical deployment and data compression could miti-

gate the energy hole, whereas increasing the bit rate led to worse results.

Wu et al. [36] showed that it is possible to achieve nearly balanced energy depletion by increasing the density in geometric progression from the outer to the inner coronas. Based on this fact, they proposed a nonuniform node distribution strategy: the Q-Model (see Section 3.2). The model proposed in this work encompasses the Q-model, among others, as a particular case.

Chatterjee and Das [10] and Mahmud and Wu [23] proposed multi-sink deployment to tackle the problem of the design of large scale wireless sensor networks. The first article presented a technique for optimizing the number of clusters and their diameter and, thus, increasing the network lifetime. They applied a graph theoretical approach based on a random graph decomposition, and they showed that their technique generated topologies with diameters that, in most cases, approached the specified bounds. In the second article, two different problems were studied: (i) deploying k static sinks such that the lifetime is maximized, and (ii) deploying k mobile sinks such that the lifetime is maximized. This study resulted in high energy savings for both cases when using multiple sinks.

Abo-Zahhad et al. [1] and Liu et al. [21] explored mobile sink approaches to tackle the energy hole problem. In the first, the sink follows a path that minimizes the total dissipated energy, while second uses mobile sink nodes in the edge of the network to increase coverage. Although these approaches are interesting, mobile and multi-sink schemes are not feasible in all scenarios.

In contrast from the aforementioned approaches, in this work we present a novel stochastic deployment model for WSNs that leads to planned but not totally controlled (i.e., nondeterministic) topologies that are able to represent a wide variety of WSNs. This new model encompasses many other models present in the literature, and is capable of properly addressing the energy hole problem. We assess a wide variety of scenarios that can be described by this new model in terms of (i) coverage, (ii) connectivity, (iii) small-world characteristics, and (iv) energy hole behavior. We show that with a planned stochastic deployment, the generated topology improves the network performance by means of a better average path length (shorter paths) and a higher clustering coefficient, and reduces the energy hole problem with the addition of only 3% of H-sensors.

3. Topology-based characterization and modeling of WSNs

3.1. Topological characterization: the Sink Betweenness metric

In the WSN context, we aim to find strong relationships between topological metrics and network metrics. For instance, in this work we investigate the relationship between energy depletion and topological metrics that defines the centrality of a node in a specific context.

Distributed inference of topological metrics is always desirable for the design of topology-aware algorithms. In this context, energy-efficient distributed inference is a challenge that requires attention. In this work, we are interested in studying a centrality metric that presents nifty characteristics for WSNs. We refer to [28,30] for an energy-efficient distributed algorithm to calculate our proposed metric.

3.1.1. Centrality metrics

Consider a network whose topology is represented by the graph $G(\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V} = \{v_1, \ldots, v_n\}$ is the set of $n = |\boldsymbol{V}|$ nodes, and $\boldsymbol{E}$ is the set of edges.

Several centrality metrics are based on graph features such as the distance between vertices, *Closeness*, degree, *Eccentricity*, neighborhood importance, *Eigenvector*, and *Hub Score* (see [22] for an example). Another widely used concept in centrality metrics is the graph shortest path; for example, the *Shortest-path Betweenness Centrality* [12] calculates the centrality of vertex i, based on the proportion of geodesics between any pair of vertices that fall on i (with respect to the total number of geodesics in the graph).

It is difficult to locate and count geodesics in large networks [12], and computational resources are limited in WSNs. The most efficient centralized algorithm for calculating the Betweenness has running time $O(n|\boldsymbol{E}| + n^2 \log n)$ for weighted graphs, and $O(n|\boldsymbol{E}|)$ for non-weighted graphs.

The Betweenness of node v is defined as:

$$B(v) = \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{1}$$

where σ_{st} is the number of shortest paths from s to t, $\{s, t\} \in \boldsymbol{V}$, and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through $v \in \boldsymbol{V}$, $s \neq v \neq t$ and $s \neq t$. The rationale behind the Betweenness metric is to indicate node centrality by means of the number of shortest paths from all nodes to all others that pass through that node.

In WSN scenarios, communication typically takes place between sensor nodes and the sink node, and vice versa. In order to consider this characteristic, we adopt a new centrality metric, namely Sink Betweenness (SBet), [28,32], which considers only the shortest paths that include the sink as one of the terminal nodes. It is defined, for every $v \in V$, as

$$\mathrm{SBet}(v) = \sum_{\substack{v,t \in \boldsymbol{V} \\ v \neq t}} \frac{\sigma_{ts}(v)}{\sigma_{ts}}, \tag{2}$$

where s is the sink, σ_{st} is the number of shortest paths from t to the sink node, and $\sigma_{ts}(v)$ is the number of shortest paths from t to the sink that pass through a vertex v. In contrast to Betweenness, the SBet metric indicates centrality by means of all shortest paths from all nodes towards the sink node that passes through that node.

For the sake of simplicity, in this work we consider that WSNs can be represented by non-weighted graphs. In some scenarios, it is more appropriate to use weighted graphs, and both Betweenness and Sink Betweenness can easily be modified to support such a feature.

3.1.2. Evaluation models

In the following section, we assess whether centrality metrics are capable of capturing the energy consumption behavior in a variety of WSN scenarios. Other metrics commonly used in complex networks theory fail to represent the energy consumption in WSNs scenarios [31]; thus, in this work we only study the Betweenness and SBet. For this, we estimate the correlation between the spent energy of the nodes and the centrality metrics that interest us. We used Spearman's rank correlation, a non-parametric statistical measure of dependence between two variables, which is shown to be robust, and is recommended if data are not necessarily Gaussian [7].

We evaluate the performance of these two centrality metrics, considering a realistic wireless channel model. We refer to [39] for more details about the wireless channel. For this assessment, we considered that nodes are deployed following the URP model.

We also consider a simple MAC protocol that intends to alleviate collisions, and therefore keeps high delivery rates to the sink node. Observe that under intense traffic, the energy hole problem might be not well-characterized; this is due to a high packet loss rate, even in nodes that are at a large distance from the sink. Thus, nodes close to the sink will not receive the lost packets, and will not deplete energy relaying them. Such saturated environments are usually avoided by the network designer, and we do not consider

Table 1
Simulation scenarios.

Parameter	Value
sink node	(center-most or a random node)
network size	$n \in$ {100, 200, 300, 400}nodes
deployment model	URP
simulation time	3000 s
data rate	1 packet/m
sensor field	$100 \times 100\text{m}^2$
collision model	additive
routing model	random tree
app. model	continuous data collection
comm. radius	15 m
tx power	−10 dB
sensor model	Mica 2 CC 1000

this situation. The MAC protocol we employ performs carrier sense, exponential back-off scheme and random off-set transmissions to alleviate collisions.

To deliver the data, we use a random tree: flooding begins in the sink node, and each node then stores its neighbors upon receiving the setup packet. Every time a node is ready to transmit a packet, it randomly chooses one of its neighbors along a shortest path to the sink.

The application represents the usual situation in which sensor nodes periodically report information to the base station, but not collectively at the same time. Each node has a parameter called *sampling rate*, all initialized with the same value t_{sr}: this is the expected time between sensed values, to be reported by a sensor node. In order to avoid collisions as much as possible, each sensor node chooses a random value to transmit its ith value uniformly in the time interval $((i-1)t_{sr}, it_{sr}]$. Other data collection models, such as event-driven and query-based applications [2], are not considered in this work. For those models, the results of the correlation between the centrality metrics and energy consumption are considered only for the subgraph formed by the nodes that are used to collect and relay data. Thus, one can expect similar results to those presented herein, but restricted to the subset of nodes that are used.

3.1.3. Evaluation scenarios

Whenever analytic results are unavailable; the Monte Carlo approach is useful for estimating quantities of interest by simulation. Such estimation approximates the result of an integral, e.g. the expected value of a quantifier, by the average of representative random samples. The relevance of this class of procedures can be attested to by computing confidence intervals and other measures of variability. In the following simulation, we estimate the correlation between the energy consumption and two centrality metrics, namely, Betweenness and SBet.

Table 1 presents our evaluated simulation scenarios. The Monte Carlo experiment we performed was replicated independently 30 times, each scenario indexed by the parameters shown in Table 1. We empirically found this number of replications sufficient for hypothesis testing sample mean differences at the 95% level.

The radio power is set to provide 15 m of communication radius under the UDG (Unit Disk Graph) model. This radius induces graphs whose nodes typically have 11 and 49 neighbors, on average, for 100 and 400 nodes, respectively. Thus, we assess typical to high density WSNs.

We used the R package version 3.2.2 [29] for statistical analysis, the Omnet++ simulator version 3.3*p*1 for discrete event simulation, and the Castalia version 2.3*b*[2] for WSN models. The wireless channel, the MAC models, and the Mica 2 CC 1000 radio module are already available in Castalia; the routing and application models were implemented as specified.

3.1.4. Relation between the Sbet metric and energy consumption

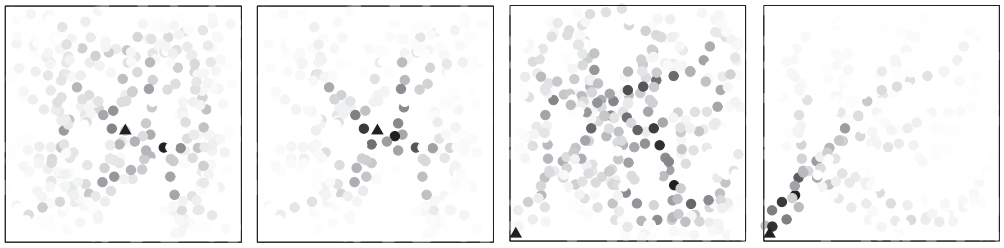
We studied the following centrality metrics, borrowed from the theory of complex networks: Betweenness [13], (SBet, eigenvector centrality, closeness, degree centrality, Kleinberg's hubscore centrality, and Kleinberg's authority centrality (see [22]). All those metrics indicate the centrality of a node by using different approaches. Betweenness and SBet are related to shortest paths; degree and eigenvector centrality account for the number of nodes' connections; but eigenvector centrality gives more importance to nodes that are connected to important nodes. Closeness indicates that as a node's centrality increases, its distance from all other nodes decreases. Kleinberg's hub score and authority centralities are equal for undirected networks, and consider a node as an authority if it is linked to hubs and vice-versa, that is, a node is a hub if it is linked to authorities. These metrics consider the importance of the connections, not only the number of connections.

We investigated how those metrics relate to the distribution of energy spent by nodes in all scenarios derived from Table 1. Even though all the aforementioned metrics have been investigated, only Betweenness and SBet present good results, as they capture the energy consumption behavior of the relay task when nodes use shortest paths to communicate to the sink. We evaluated these metrics by using the Spearman correlation between all aforementioned metrics and the energy consumed in the node. In the following, we present the results of our simulations, omitting the values of all other metrics, which did not appear useful in our studies.

Fig. 1 shows a typical panorama of the distribution of Betweenness and SBet. The gray level of a node is proportional to its Betweenness or SBet. Points become darker with greater Betweenness or SBet. The sink node is represented by a triangle. Figs. 1(a) and 1(b) show the Betweenness and SBet when the sink node is positioned at the center of the network. Observe that both Betweenness and SBet identify the nodes that concentrate more routes toward the sink node. Notice, also, that SBet is more selective, and presents high values only in nodes that participate in more paths to the sink. Betweenness presents more nodes with high values far from the sink, once it considers paths among all nodes. Those figures offer an indication of why SBet is more likely to be related to energy depletion than Betweenness. When the sink is located far from the center, Betweenness fails to represent the nodes that participate along more paths to the sink, and lacks the ability to characterize the energy hole problem.

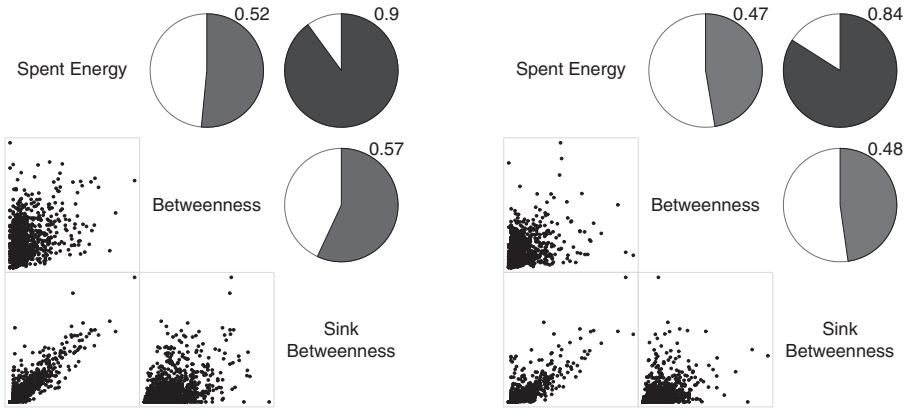
Fig. 2 shows the correlograms for 100 and 400 nodes, with the sink both centered and randomly placed. The pies depicted above the diagonal illustrate the correlation between the spent energy, Betweenness, and SBet, whereas the figures below the diagonal depict the scatterplots between those metrics. All correlograms correspond to a realistic channel with collision. Notice that SBet presents higher correlation values than Betweenness. In those cases, the Betweenness fails to represent the energy consumption due to the low correlation. As we can observe, the SBet's correlations with the spent energy are 0.9, 0.84, 0.9 and 0.83 for 100 nodes with a centrally located sink, 100 nodes with a randomly placed sink, 400 nodes withe sink located in the center and 400 nodes with the sink randomly placed, respectively. Observe also that the Betweenness' correlations with the spent energy are only 0.52, 0.47, 0.34 and 0.31 for the same scenarios. These results indicate that the SBet metric is much more appropriate than the Betweenness for capturing the energy consumption related to the relay task. This is also reflected in the scatterplots: SBet relates more linearly to the spent energy than Betweenness, in every considered scenario.

[2] http://castalia.npc.nicta.com.au.

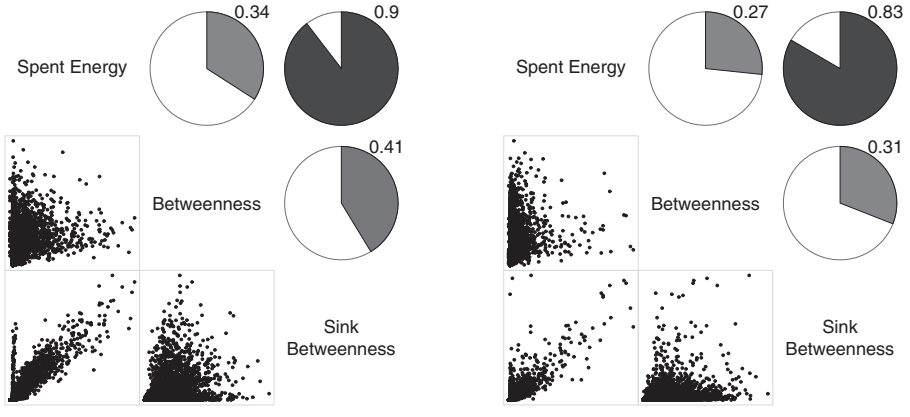


(a) Betweenness (b) SBet (c) Betweenness (d) SBet

Fig. 1. Examples of Betweenness and SBet maps when sink is in the center of the sensor field (first line) and at the left-bottom corner (second line). The sink is represented by a triangle.



(a) 100 nodes (C) (b) 100 nodes (R)



(c) 400 nodes (C) (d) 400 nodes (R)

Fig. 2. Correlograms for tree routing, 100 and 400 nodes, centered (C) and randomly placed (R) sink.

In the following sections we will formally define the M^2P^2 topology model proposed herein. For the sake of completeness, we will present an introduction to the stochastic point process theory, which is necessary to fully understand the proposed model. We also present a theoretical analysis of the energy consumption of this model.

3.2. Topology model: the M^2P^2 process

3.2.1. Stochastic point process

A stochastic spatial point process χ defined on a space W, equipped with a σ-algebra $\mathcal{B}$ and a probability measure β, is a finite subset of W that describes the location of a number of points in a region of the space, for all $\{\xi\} \in W$, $\xi \in \mathcal{B}$ and $\beta(\{\xi\}) \in \mathcal{B}$ [6]. The state space of χ is the set of all finite point configurations $\Omega = \cup_{i=0}^{\infty}\{x \subseteq W : n(x) = i\}$, where $n(x)$ denotes the number of points in x; for $i = 0$ we have the empty point configuration $x = \emptyset$. We equip Ω with the smallest σ-algebra making the mappings $n_B(x) = n(x \cap B)$ measurable for all $B \in \mathcal{B}$. Let ν denote a Poisson point process on W with intensity measure $\lambda\beta$, where $\lambda > 0$ is a parameter. In other words, if χ follows ν, then $n(\chi)$ is Poisson distributed with mean λ, and conditionally on $n(\chi) = i$, the i points in χ are independent and each point has distribution β. Specifically, consider a circular space $W_{r_s}^c$, where $r_s, c \in \mathbb{R}^+$ is the radius of the sensor field, and its center, respectively, $\mathcal{B}$ as the Borel

sets, and λ as the uniform distribution, in which case ν is a standard Poisson Process.

If the number of points n is the outcome of N, a random variable following the Poisson distribution with parameter (mean) $\lambda > 0$, i.e., $\Pr(N = k) = e^{-\lambda}\lambda^k/k!$ for every $k \in \mathbb{N}_0$ and $N(\omega) = n$, $\omega \in \Omega$ is an arbitrary event, i.e., points are placed according to a binomial point process on W, we then have a Poisson point process with intensity λ on W. This distribution process is regarded as one of the basic tools in the theory and practice of point process, since it describes complete randomness.

Several properties stem from the aforementioned constructive definition provided for Poisson point processes, some of them being equivalent definitions as, for instance, the following two:

PPP1 The number of points in every compact set $A \subset W$, denoted by $C(A)$, follows a Poisson distribution with mean $\lambda\Lambda(A)$, where $\lambda > 0$ is called "intensity" and $\Lambda(A)$ is the area of A.

PPP2 If $A_1, A_2, \ldots, A_m$ are disjoint subsets of W, then $C(A_1), C(A_2), \ldots, C(A_m)$ are collectively independent random variables.

An important generalization is obtained by varying the intensity λ suitably on W. In order to do so, we define the bounded positive function $\lambda : W \to \mathbb{R}_+$, called "intensity function", and replace property PPP1 above by the following:

PPP1' The number of points in every compact set $A \subset W$, denoted by $C(A)$, follows a Poisson distribution with mean $\beta = \int_A \lambda(u)du$.

Whenever λ is not a constant, we have an inhomogeneous Poisson point process.

In the WSN context, such an inhomogeneous process can be used to specify some areas with a greater concentration of points (sensors) by using the intensity parameter in an appropriate manner.

Thus, by construction, the intensity function will tend to concentrate more points in regions near both the sink and the H-sensors, with fewer points in other regions.

3.2.2. The M^2P^2 model

Let us introduce the Multilevel Marked Point Process (M^2P^2). First we place m H-sensors on W, and we then deploy the remaining $n - m$ sensors "close" to them. We denote the coordinates of the m H-sensors by $\boldsymbol{h} = \{(hx_1, hy_1), \ldots, (hx_m, hy_m)\}$ (how these sensors are placed will be described later).

In this work, we adopt two different strategies for deploying the L-sensor nodes considering the H-sensors. The first strategy is the Q-model proposed by Wu et al. [36] to deal with the energy hole problem in homogeneous WSNs. We adapted this model to accomplish the heterogeneous WSNs needs, and consider the intensity function λ as:

$$\lambda(i) = \frac{1}{q^i}, \tag{3}$$

where $q \geq 1$, and $i = 1, 2, 3, \ldots$ is the corona index, i.e., the distance, in hops, to the sink node. This model concentrates more and more nodes close to the H-sensors as q increases, following a geometric progression with common ratio q, i.e., the expected number of nodes N_i in the corona C_i is $q = N_i/N_{i+1}$.

The second model, namely the P-model, considers the following intensity function λ:

$$\lambda(x, y) = \begin{cases} a, \text{ if } d((x, y), (hx_i, hy_i)) \leq r_c,\ 1 \leq i \leq m, \\ 1, \text{ otherwise.} \end{cases} \tag{4}$$

where $a \geq 1$ is the attractiveness parameter, d is any distance measure, and r_c is the communication radius of the L-sensors. In the remainder of this work we employ the Euclidean distance, but any suitable distance measure may be used to enhance realism.

Notice that a stochastic point process defined by an intensity function λ, as the one in Equation (4), has an overall mean intensity given by $\int_W \lambda$. If $a > 1$, then it is more likely to have points around the m coordinates where there is an H-sensor; if A_1 belongs to the area of influence of an H-sensor and A_2 does not, but $\Lambda(A_1) = \Lambda(A_2)$, there will be on average a more sensors in the former than in the latter subset. As defined, two or more H-sensors that are arbitrarily close will behave as a single H-sensor for the deployment of L-sensors.

We have described two processes for deploying the L-sensors. We denote such processes by $\Phi(n - m, \nu, \boldsymbol{h})$, where ν indexes the model: q for the Q-model and a for the P-model.

In heterogeneous WSNs, H-sensors are useful for providing long-range shortcuts and diminishing the number of hops required to reach the sink node. They have a high-powerful radio that is able to communicate in long-range distances, and a high-capacity battery that increases their lifetime. These features make the H-sensors more expensive than the other nodes. Ideally, the H-sensors would be deployed in such a way that the number of H-sensors in close proximity to one another is diminished, thus decreasing the total amount of H-sensors required for creating the appropriate shortcuts.

The SSI (Simple Sequential Inhibition) stochastic point process [5] is a convenient model for the repulsive deployment of sensors. This process is defined on a circle W by the maximum number of m points and an inhibition distance d. The first of the m points is placed in W, obeying a binomial process. At each subsequent iteration, a new point is placed in W, and it is accepted only if all other previous points lie further than d; otherwise it is rejected. The procedure stops either when the m points have been placed, or when a maximum number of iterations has been reached. If $d > \ell/m^{1/2}$, where ℓ is the side of the square circumscribed inside the circular sensor field, it will be impossible to place all the points in W. Smaller inhibition distances do not guarantee that there will be all the n points, unless d is negligible. This is the process that places at most m non-overlapping disks of radii $d/2$ on W. In some richer repulsive point processes there is no strict inhibition, such as the Strauss process [5]; the SSI will suffice for our purposes, because we only want to avoid the overlap of the regions of influence of the H-sensors. It will be denoted by $H(m, 2r)$, for *hardcore*.

We now are ready to define the Multilevel Marked Point Process M^2P^2.

Definition 1. (**M** ^{2}P^2(m, n, a, r_c, r_i) on W). Consider a number $m \geq 1$ of H-sensors over a total of $n > m$ sensors, the intensity ν of L-sensors on a circle or radius $r_c > 0$ centered at each H-sensor (r_c is the communication radius among L-sensors) and inhibition radius $r_i > 0$ among H-sensors. Thus, M^2P^2 is a compounded process of m samples of $H(m, 2r_i)$ (the H-sensors) and $n - m$ samples of $\Phi(n - m, \nu, \boldsymbol{h})$ (the L-sensors).

Sampling from this process can be achieved in two steps. First, consider a sample from an $H(m, r_i)$ process with exactly m points: the coordinates of the m H-sensors. Second, return the outcome of an inhomogeneous binomial point process with intensity function λ, such as those defined in Equations (3), and (4), using as $\boldsymbol{h}$ the m coordinates obtained in the first step, and take a sample of $n - m$ points by using $\Phi(n - m, \nu, \boldsymbol{h})$.

A sample from the M^2P^2 process is a set of marked points. The connectivity radii among L- and H-sensors, r_c and r_{ch} respectively, induce a network topology. Our first approach for tackling the problem of modeling WSNs is the $\mathcal{C}$ model presented in [14]. The M^2P^2 is far more general than the $\mathcal{C}$ model, and encompasses heterogeneous networks.

The M^2P^2 process defined herein is a general model capable of generating a wide variety of WSN topologies, from random to planned stochastic node deployment, and from homogeneous to heterogeneous networks. This model can be implemented in network simulation tools to take advantage of the M^2P^2 expressivity for evaluating protocols under different topologies.

3.2.3. Energy consumption analysis of inner models

Considering that both Q- and P-models represent inhomogeneous Poisson stochastic processes, we are able to analytically evaluate their energy consumption behavior by estimating the number of nodes expected to be deployed in each corona. If we consider the application as the one described in Section 3.1.2, in which sensors report their sensed data periodically to the base station, we are able to estimate the total number of messages transmitted by each corona. Consequently, we are also able to estimate the average number of messages transmitted by each sensor. For the sake of simplicity, we show only the homogeneous scenario, but this analysis can be easily extended to the heterogeneous scenario.

Following the aforementioned property PPP1' of Poisson Processes, the expected number of nodes in each corona, $\mathbb{E}(n_i)$ is expressed by

$$\mathbb{E}(N_i) \propto \int_{\Lambda(C_i)} \lambda(r)dr,$$

where $\Lambda(Ci)$ is the area of the ith corona, and r is the distance to the sink node.

For instance, the URP model can be represented by a Poisson point process with $\lambda = 1$. Thus, the expected number of nodes in each corona is

$$\mathbb{E}_{URP}(N_i) = N\frac{\int_{\Lambda(C_i)} dr}{\int_W dr}, \tag{5}$$

where N is the total number of nodes, and W is the sensing area. Observe that for this simple case, the expected number of nodes is the ratio between the area of the corona to the total area.

For the P-model, we have a similar situation, but we consider two different regions. The first corona has $\lambda = a$ and all others have $\lambda = 1$. Thus, the expected number of nodes per corona is

$$\mathbb{E}_P(N_i) = \begin{cases} N\frac{\int_{\Lambda(C_i)} adr}{\int_W dr}, & \text{if } i = 1 \\ N\frac{\int_{\Lambda(C_i)} dr}{\int_W dr}, & \text{otherwise.} \end{cases} \tag{6}$$

Considering the Q-model, we have

$$\mathbb{E}_Q(N_{i+1}) \equiv \frac{N_i}{q}, \tag{7}$$

with

$$\mathbb{E}_Q(N_1) = \frac{N}{1 + \sum_{i=2}^{|C|} q^{-1}}, \tag{8}$$

where $|C|$ is the total number of coronas.

Consider, for instance, 1000 sensors with communication radius $r_c = 50$ m deployed on a circular field with radius $R = 500$ m. In this scenario, the area of the coronas are represented by the circular sectors formed by each corona, $\mathbb{E}(N_i)$ is proportional to the volume of the cylinder sector formed by the area of the corona, and height is defined by the intensity function λ. In our examples, the intensity function λ is either constant, as in the URP model, or constant in each corona, but this method can be applied to any function λ. To illustrate the method, consider that each node transmits γ own messages and the nodes of corona C_i must transmit the messages from the nodes of corona C_{i+1}, and so on. Thus, the expected number of messages transmitted by each node is

$$\mathbb{E}(M) = \gamma\left(1 + \frac{N_{i+1}}{N_i}\right). \tag{9}$$

We used Equations (5), (6), (7) and (9) with the parameters of the aforementioned network, and used $\gamma = 10$ to produce the results presented in Fig. 3. As expected, the Q-model outperforms the P-model and URP regarding energy consumption. Observe that this situation represents the best case of the energy consumed per node to perform data transmission, while in this analysis we do not consider retransmissions whatsoever, and we assume that the total number of messages originated in corona N_{i+1} will be equally distributed with the nodes of corona N_i. Notice that in the ideal scenario, only the Q-model equally divides the relay task among the nodes in the same corona. In the following sections we analyze, by means of simulation, the average case for many different scenarios, and show that the P-model may be preferable in some situations.

4. Evaluation of the M^2P^2 model

In this work, we evaluate the M^2P^2 model to assess some properties of the deployments presented in Section 3.2, namely: (i) coverage, (ii) energy hole behavior, and (iii) clustering coefficient and average path length.

We consider homogeneous and heterogeneous WSNs with sensors with the same sensing capability (r_s), and two levels of transmission (r_c and r_{ch}) ranges, which are, for the sake of simplicity, perfect circles. H-sensors have higher data throughput than the L-sensors, but a comparable sensing capability. H-sensors have a radio that operates in two different channels; they are able to achieve long-distance communication only with other H-sensors, and they use short-distance communication in order to send data to and receive data from L-sensors.

Coverage plays an important role in the design of WSNs. It indicates how the network covers the sensor field, and it has a strong influence on the quality of the information reported by the WSN [14].

A simple way of estimating the coverage of a topology is to calculate the area of the intersection of the circles centered at each sensor within the sensor field.

We only consider nodes inside the connected component to which the sink belongs, since only their data will be reported. No data fusion is considered in the following analysis.

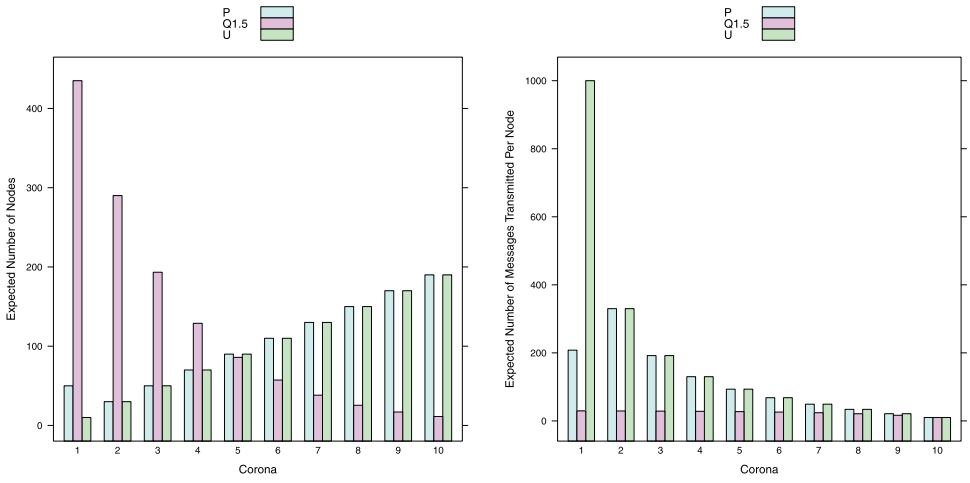
Small-world networks share characteristics of both regular and random graphs, presenting high values of clustering coefficients (similar to regular networks), and small values of the average shortest path length (similar to random networks) [17,22,27].

High clustering coefficients make the topology more fault-tolerant, due to the high density of cycles of order three in the network. Thus, if a node fails, is a greater number of neighbors that can recover from the failure [9]. Networks with high clustering coefficients typically increase the probability of routing loops. Random topologies, typically presented in WSNs, are prone to routing loops; hence, the overlying routing protocol should address this issue.

Small values of the average shortest paths indicate that the shortcuts provided by the H-sensors tend to reduce latency in data communication [15]. Thus, as stated by Helmy [17], small-world networks present desirable characteristics for WSNs.

We used the SBet metric to evaluate the energy hole behavior of the topologies generated by our deployment model, as we have showed that it is highly correlated to energy consumption (see Section 3.1.4).

Table 2 presents the simulation scenarios we evaluated. Each scenario was replicated independently 30 times, for the aforementioned reasons. Observe that the parameters are quite different from those in Table 1, because now we are interested in large scale networks. We simulate those large networks with and without H-sensors, to stress the benefits of using a heterogeneous approach



(a) $\mathbb{E}(N_i)$ (b) $\mathbb{E}(M)$

Fig. 3. Expected number of nodes per corona/Expected number of messages transmitted per node with respect to three different models.

Table 2
Simulation scenarios.

Parameter	Value
sink node	1 (center-most node)
network size	{1000, 2000, 3000} nodes
L-sensors' comm. radius	50 m
H-sensors' comm. radius	300 m
number of H-sensors	{0, 20, 30, 40} nodes
Q-model deployment parameter (q)	{1, 1.5, 2, 2.5, 3, 4, 5}
P-model deployment parameter (a)	{1, 3, 5, 7, 10, 15, 30}
sensing radius	30 m
sensor field	circular with radius of 500 m

for such large networks. It is worth mentioning that some deployment model parameters (q and a) listed in that table were evaluated only either in homogeneous or heterogeneous networks in order to find the most suitable parameters for each scenario. We also simulate the P-model with $a = 1$ to produce a URP deployment used as baseline in our evaluation. In the following, we used a Monte Carlo experiment to estimate the mean value and the confidence intervals of (i) coverage, (ii) SBet metric, (iii) clustering coefficient, and (iv) average path length, for a wide variety of scenarios of homogeneous and heterogeneous networks and different topology models, namely, URP, Q-model and P-model.

Except for the homogeneous networks, H-sensors are placed by a repulsive deployment with inhibition radius between r_c and $r_s/m^{1/2}$, where r_c is the L-sensor communication radius, and r_s is the sensor field radius. Additional details about the inhibition radius are provided in Section 4.4. The sink node is the most central H-sensor of the sensor field.

We use R language and programming environment version 3.3.0 [29] for node deployment and estimation of topological quantities. The following sections discuss the simulation results.

4.1. Coverage

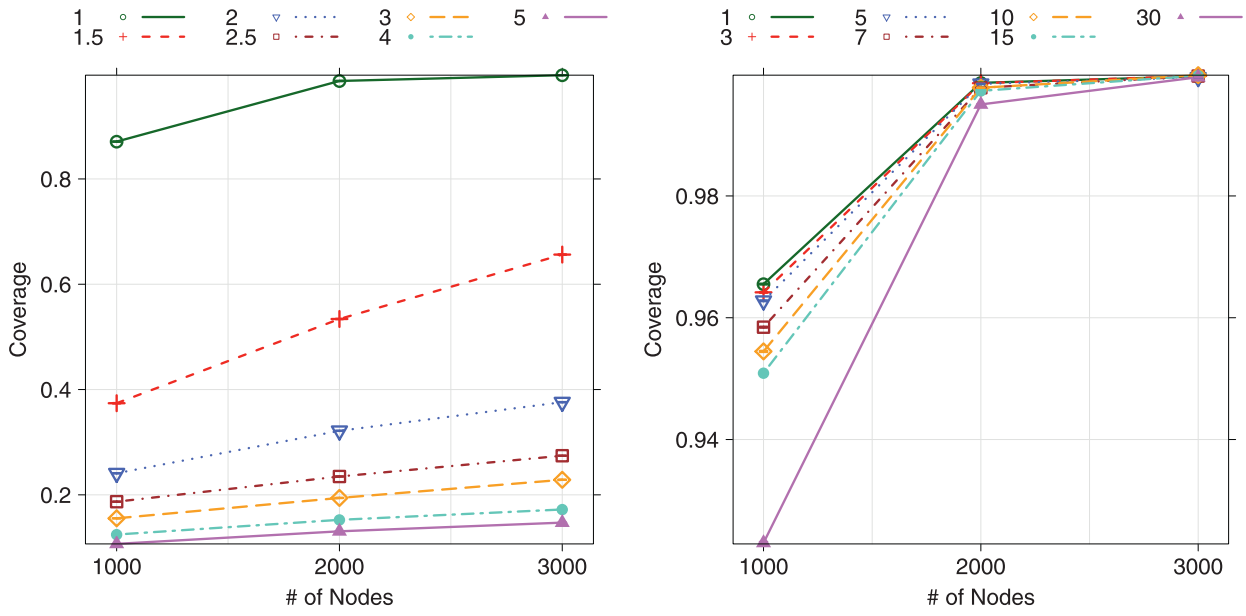
Fig. 4 shows the coverage provide respectively by the Q- and P-models for homogeneous networks, with respect to the number of nodes. The curves depict the behavior of each parameter of the deployment model. Usually, higher coverage means better deployment. However, there is a trade-off between the coverage and the energy balance by distributing more nodes in the coronas near the sink node. Thus, we conduct this first experiment to guide the choice of the parameter for each model. We observe that in Fig. 4(a), the best coverage for the Q-model is achieved when $q = 1$. However, as we will see in Section 4.2, this is the value that presents the most uneven energy balance (highest SBet). When considering the homogeneous case, $q = 1$ is the only value that guarantees at least 85% coverage.[3] This occurs because the network diameter (highest distance to the sink) is high. As the Q-model aggressively concentrates more nodes near the coronas close to the sink, we expect only a few sensor nodes in the farthest coronas. We repeat this analysis for the other model, and based on Fig. 4(b), we chose $a = 5$ for the P-model. It is worth mentioning that the P-model preserves high coverage, above 90% for all situations. This is because this model only changes the first corona (nearest to the sink node), and preserves the node distribution for all other coronas. Therefore, with 1000 nodes, the P-model shows similar coverage as the Q-model achieves with 2000 nodes.

Fig. 5 shows the comparison between the aforementioned two models, along with the URP model, for both homogeneous and heterogeneous scenarios. The URP model is used as a baseline. We see in Fig. 5(a) that the P- and URP models present similar results. This means that considering the parameters we chose for the two models, the P-model ensures almost the same coverage results as the URP model, while the Q-model requires more nodes to achieve similar results.

In Fig. 5(b) we show the results for heterogeneous networks, with 20, 30 and 40 H-sensors deployed as described in Section 3.2. We conduct a similar evaluation as the one described for homogeneous networks, to choose the most suitable parameter values. For the P-model we chose $a = 3$, for Q-model the value was $q = 1.5$. When H-sensors are used, we observe that the P-model presents slightly higher coverage only when the number of H-sensors is small, for instance, 20 and 30. This occurs because the P-model tends to concentrate more nodes in the first corona around the H-sensors. When there are more H-sensors, more nodes will be placed around them, and fewer nodes will be left to cover the rest of the sensor area. The Q-model scales better than the P-model with the number of H-sensors, since the transition is smoother and the network diameter decreases.

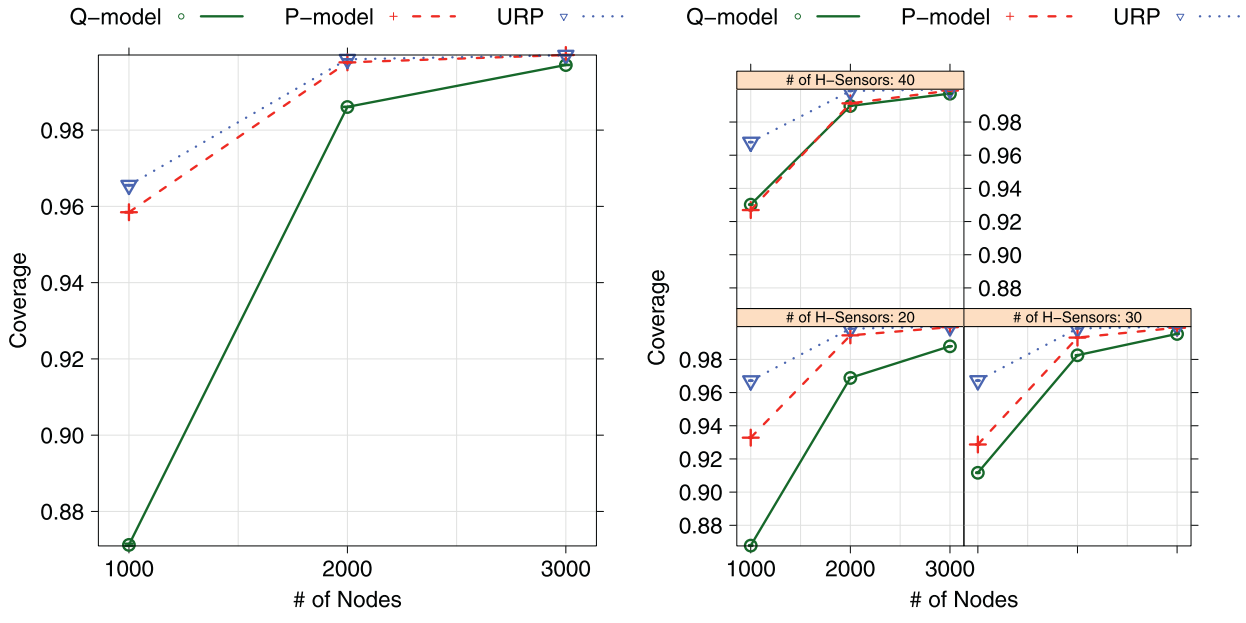
The main goal of these deployment models is to alleviate the energy hole problem without compromising the coverage. Thus, in the next section we present the energy balancing properties of each model to guide us in the choice of the most suitable model.

[3] This is an arbitrary choice, any other coverage guarantee can be chosen here.



(a) Q-model (b) P-model

Fig. 4. Coverage for the homogeneous scenarios as a function of the number of nodes for each deployment model.



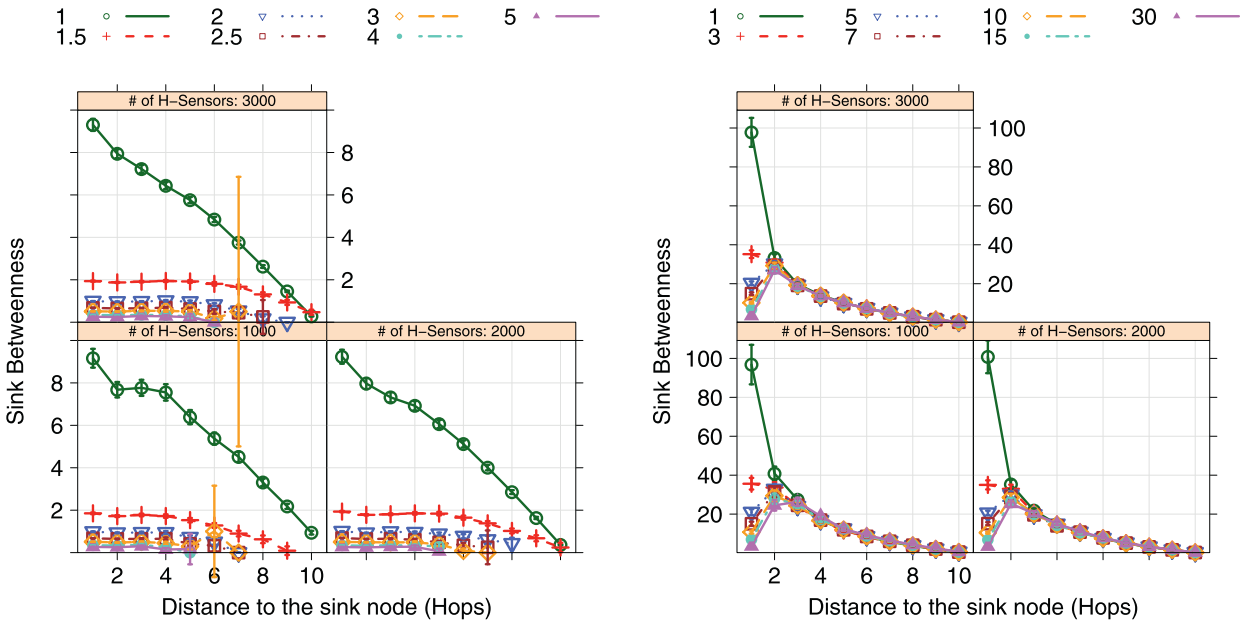
(a) Homogeneous (b) Heterogeneous

Fig. 5. Coverage model comparison for the homogeneous and heterogeneous scenarios with respect to the number nodes.

4.2. Energy balancing

The lifetime of an individual sensor is defined by the total time it can receive and transmit data until it runs out of energy. The entire network lifetime is a more sophisticated subject, and it is beyond the scope of this work; we refer to [4] for further details. The transmission and reception of data dominate the energy consumption of each sensor and, thus, we disregard any other energy spent in tasks as sensing, processing and sensor idleness [35]. In these simulations, we are trying to evaluate the extent to which the relay task on each assessed topology is fairly distributed. For this analysis, we used the results presented in Section 3.1, as we concluded that SBet is able to characterize the relay task, and can therefore characterize the energy depletion of this task.

Fig. 6 presents the SBet metric for the Q- and P-models for homogeneous networks as a function of the distance to the sink node, in hops. The curves show the behavior of each parameter of the deployment model. Along with the coverage results, we also use the results shown in Fig. 6 to guide the choice of parameter for each deployment model. Fig. 6(a) shows the results for the Q-model: when $q = 5$, SBet reaches the smallest values. We also notice that this value leads to the lowest coverage. Therefore, we use the same criteria that was used to evaluate the coverage to choose $q = 1$, because although it leads to the highest SBet, it meets our coverage requirements. In Fig. 6(b), we observe that the P-model is



(a) Q-model (b) P-model

Fig. 6. SBet for the homogeneous scenarios as a function of the number of the distance to the sink.

capable of alleviating the energy hole problem in the first corona by increasing parameter a. When $a = 15$, the average SBet value of the first corona is three times lower than the second corona. This occurs because, because all other coronas are unaffected, even though the expected number of nodes in the first corona increases. In our scenarios, the total number of nodes is enough to cover the sensor field, even with the most attractive parameter value. This may change in severely sparse networks, affecting the coverage. We adopt $a = 5$ as the most suitable value, because it meets the coverage requirements while decreasing the SBet of the first corona to low values. Observe that the SBet values of the Q-model are much lower than when the P-model is used. The implications of this fact are the Q-model is better indicated to alleviate the energy hole in homogeneous networks, while preserving coverage. It is worth noticing that with $q = 1$, we expect the same number of nodes per corona.

Fig. 7 shows the SBet values for nodes in the homogeneous and heterogeneous scenarios as a function of the distance to the sink node. We now want to compare the evaluated models. Similarly to the coverage evaluation, we conducted an assessment to choose the parameter values. Fig. 7(a) shows the SBet as a function of the distance to the sink node for 1000, 2000 and 3000 nodes and the homogeneous scenario. Observe that, as expected, the URP model presents the highest SBet, and the use of the Q- and P-models decreases it. We observe that the Q-model presents the best results among the models herein evaluated, as it presents the lowest Sbet values. The P-model is only capable of decreasing the SBet of the first corona; the remaining coronas are similar to the URP model. Observe that the pattern showed in Fig. 7(a) resembles Fig. 3(b). This similarity means that both SBet and the analytic approach presented in Section 3.2.3, are able to describe energy consumption in these networks, besides, we observe that the simulations confirm the analytic approach.

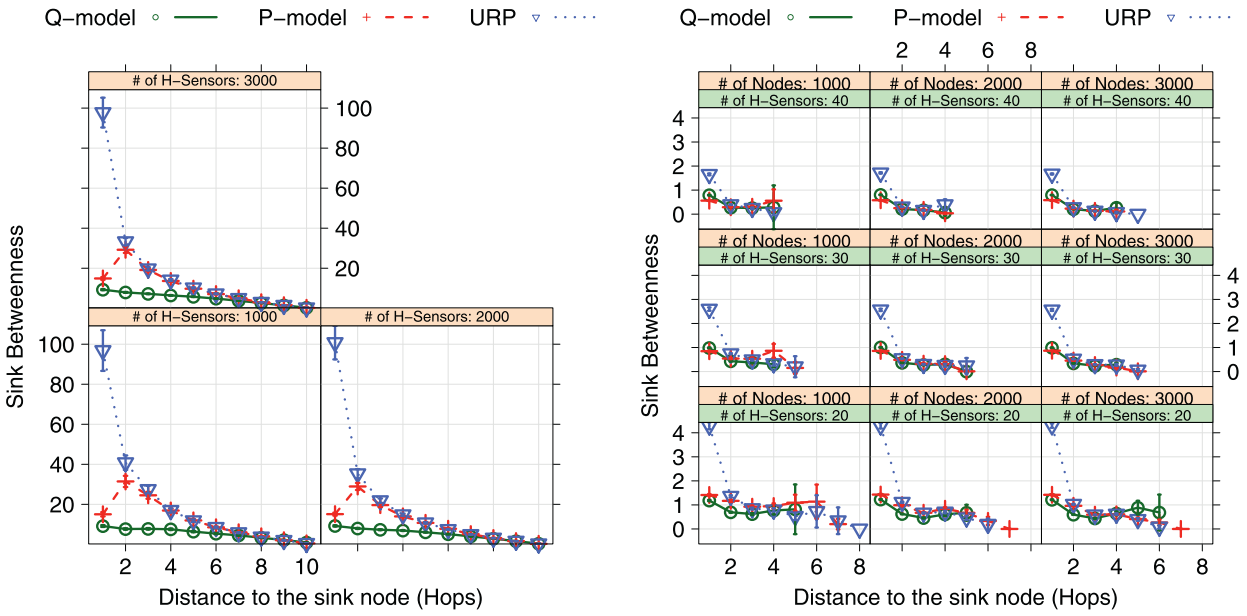
Fig. 7(b) shows the SBet in heterogeneous networks with 20, 30, and 40 H-sensors as functions of the distance to the sink. Observe that an increase in the number of H-sensors leads to just a slight decrease of the SBet, and that an increase in the total number of nodes does not considerably change the SBet values.

Notice that the adoption of the H-sensor approach leads to a substantial decrease of the SBet, when compared to the homogeneous scenario. For instance, twenty H-sensors lead to SBet values about ten times smaller than in the homogeneous case. Among all deployment models, the P-model is the one that leads to the highest decrease of the SBet for heterogeneous networks, closely followed by the Q-model. As the mere presence of H-sensors helps alleviate the energy hole problem, the deployment model is responsible for fine tuning. We observe that with forty H-sensors, even the URP model presents SBet values close to the others. Since the main goal is to use the lowest possible number of H-sensors, the adoption of a heterogeneous network with twenty H-sensors and the P-model decreases the SBet fifty times when compared to homogeneous networks with the URP model, and four times when compared to heterogeneous networks with the URP model.

4.3. Clustering coefficient and average path length

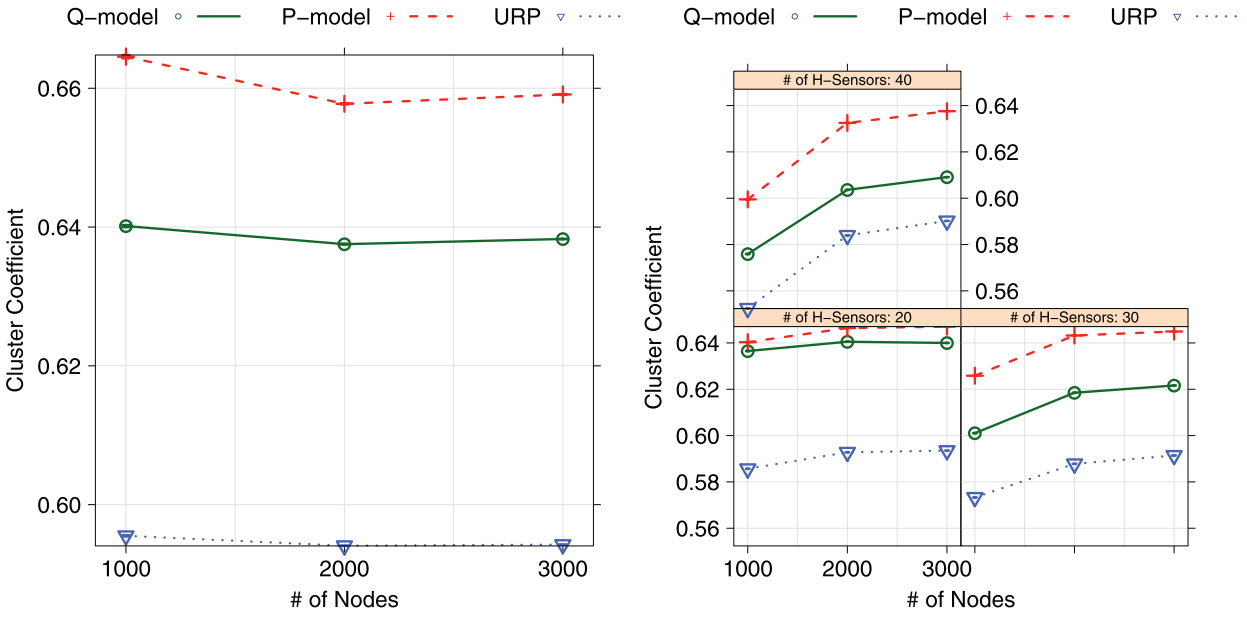
Fig. 8 shows the clustering coefficient as a function of the number of nodes for homogeneous and heterogeneous networks with the same aforementioned parameters. Fig. 8(a) shows the homogeneous case. Observe that the P-model leads to the highest clustering coefficient. We observe a similar behavior when a heterogeneous network is adopted, as shown in Fig. 8(b). Overall, the cluster coefficient of heterogeneous networks is slightly higher than in homogeneous networks. The clustering coefficient is only slightly affected by the presence of H-sensors. These results show that the P-model is most likely to favor fault-tolerance strategies.

Fig. 9 shows the average path length for both homogeneous and heterogeneous networks as a function of the number of nodes, when the two deployment models and the URP are used with the same aforementioned parameter values. Although the adoption of heterogeneous network design does not offer a high impact on the behavior of the cluster coefficient, the average path length of the heterogeneous networks is consistently lower than in the homogeneous case. We can explain this behavior by observing that heterogeneous networks add shortcuts that decrease the path lengths. We also observe that increasing the number of H-sensors decreases the average path length. The Q-model leads to the lowest average



(a) Homogeneous (b) Heterogeneous

Fig. 7. SBet model comparison for the homogeneous and heterogeneous scenarios as a function of the distance to the sink.



(a) Homogeneous (b) Heterogeneous

Fig. 8. Clustering coefficient.

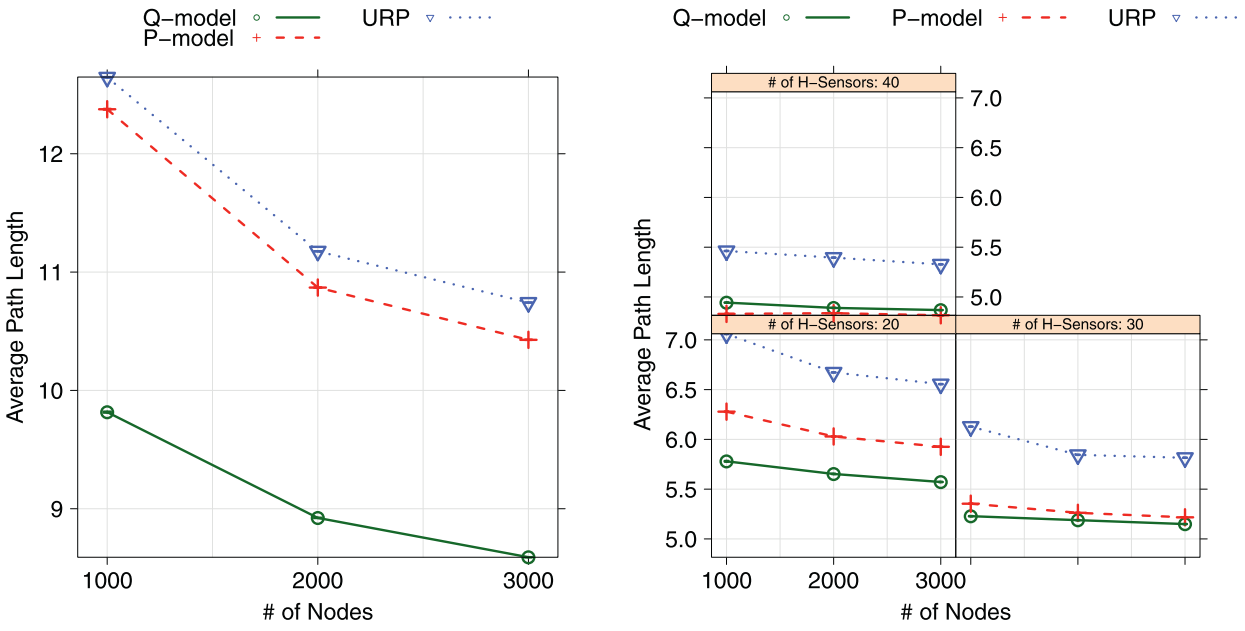
path length in the homogeneous case, as well as in the heterogeneous case.

4.4. A guide to a stochastic planned deployment

This section discusses the parameters for the M^2P^2 process that were used in our simulation, and how such parameters should be chosen for the description of real world situations.

The $M^2P^2(m, n, a, r_c, r_{ch}, r_i)$ process on W has the following parameters:

- The area W, where the process takes place; the user should describe the actual geometry of interest. Only connected windows make sense.
- The distance measure among sensors. For the sake of simplicity, we used the UDG model, but other models could be used. The communication radii should be carefully specified as a function of the communication channel; this distance specifies r_c and r_{ch}. More general distances can be used, such as those that take obstacles into account.
- Number (n) and types of sensors required for precise, lasting and economic data acquisition and delivery; the most expensive m H-sensors, aim to improve network performance, while the $n - m$ L-sensors are primarily devoted to data collection and first-level data relay.
- The inhibition parameter r_i specifies the minimum distance at which H-sensors are allowed to lie. Overlapping H-sensors is



(a) Homogeneous (b) Heterogeneous

Fig. 9. Average path length.

redundant and wastes resources, so the repulsive placement we proposed enhances network performance in an economic manner.

- The intensity parameter, for instance *s*, *a*, and *i*, which describes the expected proportion of L-sensors around an H-sensor; outside this area.

As previously mentioned, the inhibition radius r_i grants that the areas of influence of H-sensors do not overlap and, at the same time, allows the placement of all the m H-sensors on W. The first condition, which grants that the areas of influence do not overlap, requires $r_i \geq r_{ch}$, while the last condition, which assures the placement of all H-sensors, imposes $2r_i < \ell/m^{1/2}$, where ℓ is the side of the square circumscribed inside of the sensor field. The most repulsive process, namely the one with $2r_i \approx \ell/m^{1/2}$, yields H-sensors deployed in an almost regular grid. Observe that the knowledge of the exact location of neither H- nor L-sensors is required for the deployment. The stochastic point process we provide is a description of a physical inexpensive procedure for node deployment, not an *a priori* specification.

The intensity parameter *s*, *a* or *i* specifies the expected number of L-sensors around each H-sensor (assuming non-overlapping areas of influence); it can be chosen, among other criteria, using our simulation results.

5. Final remarks

We proposed a novel modeling solution capable of representing a wide variety of scenarios, from totally random to planned stochastic node deployment, in heterogeneous sensor networks. This model can represent WSNs with characteristics of small-world networks, and can address the energy hole problem.

We showed that by using only about 2% of H-sensors (20 out of 1000) and deploying nodes by using the P- and Q-models to distribute the L-sensors around the H-sensors deployed with a repulsive model, we observe important characteristics of the network topology, such as low average path length and high clustering coefficient. Observing the results for coverage, SBet and small world parameters for the two models evaluated herein, we notice that the Q-model is the most suitable for addressing the energy hole in homogeneous networks, while the P-model achieves the highest coverage and overcomes all other models for heterogeneous networks. The P-model acts only in the first corona, while the Q-model acts in all coronas. Regarding the small world properties, the P-model leads to higher cluster coefficient, while Q-model leads to lower average path length, characteristics that are desirable for WSNs.

Moreover, we showed that the SBet is a suitable metric for characterizing the relay task of a node. This metric's ability, in contrast to the relative insensitivity of the classical Betweenness, suggests other possibilities. SBet can be used in a wide variety of applications, both in the design and operation of WSNs. For instance, the designer can assess the best deployment strategy in order to create graphs with a more appropriate SBet distribution. Such an assessment should improve the understanding and management of the network lifetime, since the energy consumption becomes more evenly distributed among the nodes. Studies in that direction require only spatial point process generators (in order to model the deployment models), and tools for graph analysis; therefore there is no need for either complex discrete event simulators or network models. Both are provided by R, a free, multiplatform software environment for statistical computing and graphics, which exhibits excellent numerical properties [3].

We also envision the following research lines: the quantification of the relationship between the metrics used herein, such as the clustering coefficient, the average path length, and the SBet with the fault-tolerance properties, latency and network lifetime, respectively; the introduction of fault-tolerance schemes based on the proposed model and metric; the use of topology control schemes, based on the SBet, to diminish the possibility of interference on nodes that were attractively deployed around the H-sensors and the sink, and the use of SBet to improve the routing performance in WSNs.

References

[1] M. Abo-Zahhad, S. Ahmed, N. Sabor, S. Sasaki, Mobile sink based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks, IEEE Sens. J. 15 (8) (2015) 4576–4586. doi: 10.1109/JSEN.2015.2424296.

[2] J.N. Al-Karaki, A.E. Kamal, Routing techniques in wireless sensor networks: a survey, IEEE Wirel. Commun. 11 (6) (2004) 6–28, doi:10.1109/MWC.2004.1368893.

[3] M. Almiron, E.S. Almeida, M. Miranda, The reliability of statistical functions in four software packages freely used in numerical computation, Braz. J. Probab. Stat. 23 (2) (2009) 107–119, doi:10.1214/08-BJPS017. URL http://www.imstat.org/bjps/.

[4] G. Anastasi, M. Conti, M. Di Francesco, A. Passarella, Energy conservation in wireless sensor networks: a survey, Ad Hoc Networks 7 (3) (2009) 537–568, doi:10.1016/j.adhoc.2008.06.003. URL http://linkinghub.elsevier.com/retrieve/pii/S1570870508000954.

[5] A. Baddeley, Spatial Point Processes and Their Application, in: W. Weil (Ed.), Stochastic Geometry, Lecture Notes in Mathematics, 1892, Springer, Berlin, 2006, pp. 1–75, doi:10.1007/978-3-540-38175-4_1.

[6] K.K. Berthelsen, J. Møller, A primer on perfect simulation for spatial point processes, Bull. Braz. Math. Soc. 33 (3) (2002) 351–367, doi:10.1007/s005740200019.

[7] D.G. Bonett, T.A. Wright, Sample size requirements for estimating pearson, kendall and spearman correlations, Psychometric 65 (1) (2000) 23–28.

[8] A. Boukerche, B. Turgut, N. Aydin, M.Z. Ahmad, L. Blni, D. Turgut, Routing protocols in ad hoc networks: a survey, Comput. Netw. 55 (13) (2011) 3032–3080, doi:10.1016/j.comnet.2011.05.010. URL http://www.sciencedirect.com/science/article/pii/S1389128611001654.

[9] M.R. Brust, D. Turgut, C.H.C. Ribeiro, M. Kaiser, Is the clustering coefficient a measure for fault tolerance in wireless sensor networks? in: 2012 IEEE International Conference on Communications (ICC), IEEE, 2012, pp. 183–187, doi:10.1109/ICC.2012.6364474. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6364474.

[10] P. Chatterjee, N. Das, Multiple sink deployment in multi-hop wireless sensor networks to enhance lifetime, in: Applications and Innovations in Mobile Computing (AIMoC), 2015, 2015, pp. 48–54, doi:10.1109/AIMOC.2015.7083829.

[11] T. Ducrocq, M. Hauspie, N. Mitton, On the impact of network topology on wireless sensornetworks performances: Illustration with geographic routing, in: Proceedings of the 10th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks, in: PE-WASUN '13, ACM, New York, NY, USA, 2013, pp. 141–144, doi:10.1145/2507248.2507269.

[12] L.C. Freeman, A set of measures of centrality based on Betweenness, Sociometry 40 (1) (1977) 35–41.

[13] L.C. Freeman, Centrality in social networks conceptual clarification, Social Networks 1 (3) (1979) 215–239, doi:10.1016/0378-8733(78)90021-7.

[14] A.C. Frery, H.S. Ramos, J. Alencar-Neto, E.F. Nakamura, A.A.F. Loureiro, Data driven performance evaluation of wireless sensor networks, Sensors (Basel) 10 (3) (2010) 2150–2168, doi:10.3390/s100302150. URL http://www.mdpi.com/1424-8220/10/3/2150/.

[15] D.L. Guidoni, R.A.F. Mini, A.A.F. Loureiro, On the design of resilient heterogeneous wireless sensor networks based on small world concepts, Comput. Netw. 54 (8) (2010) 1266–1281, doi:10.1016/j.comnet.2009.10.021.

[16] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, M. Franceschetti, Stochastic geometry and random graphs for the analysis and design of wireless networks, IEEE J. Sel. Areas Commun. 27 (7) (2009) 1029–1046, doi:10.1109/JSAC.2009.090902. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5226957.

[17] A. Helmy, Small worlds in wireless networks, IEEE Commun. Lett. 7 (10) (2003) 490–492.

[18] J. Hoydis, M. Petrova, P. Maehoenen, The effects of topology on the local throughput of ad hoc networks, Ad Hoc Sens. Wireless Netw. 7 (3–4) (2009) 337–347.

[19] D. Katsaros, N. Dimokas, L. Tassiulas, Social network analysis concepts in the design of wireless ad hoc network protocols, IEEE Network 24 (6) (2010) 23–29, doi:10.1109/MNET.2010.5634439. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5634439.

[20] J. Li, P. Mohapatra, Analytical modeling and mitigation techniques for the energy hole problem in sensor networks, Pervasive Mob. Comput. 3 (3) (2007) 233–254.

[21] H. Liu, Y. Zhang, H. Liu, X. Su, Inhomogeneous distribution strategy based on mobile sink nodes in wireless sensor networks, Wireless Pers. Commun. (2015) 1–16, doi:10.1007/s11277-015-2400-8.

[22] Luciano, F.A. Rodrigues, G. Travieso, V.P.R. Boas, L.d.F. Costa, P.R. Villas Boas, Characterization of complex networks: A Survey of measurements, Adv. Phys. 56 (0001–8732) (2007) 167–242.

[23] S. Mahmud, H. Wu, Lifetime aware deployment of k base stations in wsns, in: Proceedings of the 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, in: MSWiM '12, ACM, New York, NY, USA, 2012, pp. 89–98, doi:10.1145/2387238.2387256.

[24] R.B. Matthias, H.C.R. Carlos, D. Turgut, S. Rothkugel, LSWTC: a local small-world topology control algorithm for backbone-assisted mobile ad hoc networks, in: Local Computer Networks (LCN), 2010 IEEE 35th Conference on, 2010, pp. 144–151, doi:10.1109/LCN.2010.5735688.

[25] P. Mohapatra, An Analytical Model for the Energy Hole Problem in Many-to-One Sensor Networks, in: VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference, 2005., 4, IEEE, 2005, pp. 2721–2725, doi:10.1109/VETECF.2005.1559043.

[26] A. Mostefaoui, M. Melkemi, A. Boukerche, Localized routing approach to bypass holes in wireless sensor networks, IEEE Trans. Comput. 63 (12) (2014) 3053–3065, doi:10.1109/TC.2013.180.

[27] M. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.

[28] E.M.R. Oliveira, H.S. Ramos, A.A.F. Loureiro, Centrality-based Routing for Wireless Sensor Networks, in: 2010 IFIP Wireless Days, IEEE, 2010, pp. 1–5, doi:10.1109/WD.2010.5657731. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5657731.

[29] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016. URL https://www.R-project.org/.

[30] H.S. Ramos, A.C. Frery, A. Boukerche, E.M.R. Oliveira, A.A.F. Loureiro, Topology-related metrics and applications for the design and operation of wireless sensor networks, ACM Trans. Sen. Netw. 10 (3) (2014) 53:1–53:35, doi:10.1145/2512328.

[31] H.S. Ramos, D. Guidoni, A. Boukerche, E.F. Nakamura, A.C. Frery, A.A. Loureiro, Topology-related modeling and characterization of wireless sensor networks, in: Proceedings of the 8th ACM Symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks - PE-WASUN'11, ACM Press, New York, New York, USA, 2011, p. 33, doi:10.1145/2069063.2069070. URL http://dl.acm.org/citation.cfm?doid=2069063.2069070.

[32] H.S. Ramos, E.M.R. Oliveira, A. Boukerche, A.A. Loureiro, Characterization and mitigation of the energy hole problem of many-to-one communication in wireless sensor networks, in: ICNC'12: Procedings of the IEEE International Conference on Computing, Networking and Communications, 2012, pp. 954–958.

[33] M. Strübe, F. Lukas, B. Li, R. Kapitza, Drysim: Simulation-aided deployment-specific tailoring of mote-class wsn software, in: Proceedings of the 17th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, in: MSWiM '14, ACM, New York, NY, USA, 2014, pp. 3–11, doi:10.1145/2641798.2641838.

[34] A. Vázquez-Rodas, L.J. de la Cruz Llopis, Topology control for wireless mesh networks based on centrality metrics, in: Proceedings of the 10th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks, in: PE-WASUN '13, ACM, New York, NY, USA, 2013, pp. 25–32, doi:10.1145/2507248.2507257.

[35] D.M. Wang, B. Xie, D.P. Agrawal, Coverage and lifetime optimization of wireless sensor networks with gaussian distribution, IEEE Trans. Mob. Comput. 7 (12) (2008) 1444–1458.

[36] X. Wu, G. Chen, S.K. Das, Avoiding energy holes in wireless sensor networks with nonuniform node distribution, IEEE Trans. Parallel Distrib. Syst. 19 (5) (2008) 710–720. URL http://doi.ieeecomputersociety.org/10.1109/TPDS.2007.70770.

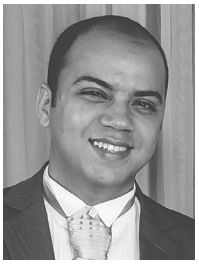
[37] M.D. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, S. Singh, Exploiting heterogeneity in sensor networks, in: INFOCOM '05: Proceedings of the 24th IEEE International Conference on Computer Communications, IEEE, 2005, pp. 878–890.

[38] M. Younis, K. Akkaya, Strategies and techniques for node placement in wireless sensor networks: A Survey, Ad Hoc Networks 6 (4) (2008) 621–655, doi:10.1016/j.adhoc.2007.05.003.

[39] M. Zuniga, B. Krishnamachari, Analyzing the transitional region in low power wireless links, in: In First IEEE International Conference on Sensor and Ad hoc Communications and Networks (SECON), 2004, pp. 517–526.

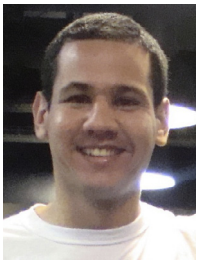
Heitor S. Ramos is graduated in Electrical Engineering from the Federal University of Campina Grande (UFCG), Brazil, his master in Computing Modeling from the Federal University of Alagoas (UFAL), Brazil, and his PhD in Computer Science from the Federal University of Minas Gerais. His research interests rely on wireless networks, sensors networks, and mobile and ad hoc networks. He is currently a Professor at the Institute of Computing of the Federal University of Alagoas (UFAL), Maceió, Brazil.

Azzedine Boukerche (FIEEE, FCAE, FEiC FAAAS) is a Full Professor and holds a Canada Research Chair Tier-1 position at the University of Ottawa. He is the founding director of PARADISE Research Laboratory at Ottawa and NSERC-DIVA Research of Excellence Centre. He is a Fellow of the Canadian Academy of Engineering. His current research interests include wireless ad hoc and sensor networks, wireless networks, mobile and pervasive computing, wireless multimedia, QoS service provisioning, performance evaluation and modeling of large-scale distributed systems, distributed computing, large-scale distributed interactive simulation, and parallel discrete event simulation. He is a holder of the IEEE Canada Gotelieb Meda Award, an Ontario Early Research Excellence Award (previously known as Premier of Ontario Research Excellence Award), Ontario Distinguished Researcher Award, University of Ottawa Research Excellence Award, and Glinski Research Excellence Award.



Alyson L. C. Oliveira received his bachelor's degree in Computer Science and master in Computing Modeling from the Federal University of Alagoas (UFAL), Brazil.

Alejandro C. Frery graduated in Electronic and Electrical Engineering from the Universidad de Mendoza, Argentina. His M.Sc. degree was in Applied Mathematics (Statistics) from the Instituto de Matematica Pura e Aplicada (Rio de Janeiro) and his Ph.D. degree was in Applied Computing from the Instituto Nacional de Pesquisas Espaciais (Sao José dos Campos, Brazil). He is currently with the Instituto de Computação, Universidade Federal de Alagoas, Maceió, Brazil. His research interests are statistical computing' and stochastic modeling.



Eduardo M. R. Oliveira holds a BSc, a MSc from the Federal University of Minas Gerais (UFMG), Brazil, and a PhD degree in Computer Science from Ecole Polytechnique, France.

Antonio A. F. Loureiro received his B.Sc. and M.Sc. degrees in computer science from the Federal University of Minas Gerais (UFMG), Brazil, and the Ph.D. degree in computer science from the University of British Columbia, Canada. Currently, he is a professor of computer science at UFMG, where he leads the research group in wireless sensor networks and ubiquitous computing. His is a regular Visiting Professor at PARADISE Research Laboratory, University of Ottawa. His main research areas are wireless sensor networks, urban sensing, ubiquitous computing, and distributed algorithms. In the last 15 years he has published over 100 papers in international conferences and journals related to those areas, and also presented tutorials at international conferences. He is the recipient of the prestigeous IEEE ComSoc AHSN Technical Achievement Award.