Structural Changes in Data Communication in Wireless Sensor Networks

Raquel S. Cabral 1 *, Andre L. L. Aquino 2 †, Alejandro C. Frery 2 ‡, Osvaldo A. Rosso 2,3 §, Jaime A. Ramírez 1 ¶

Abstract: Wireless sensor networks are an important technology for making distributed autonomous measures in hostile or inaccessible environments. Among the challenges they pose, the way data travel among them is a relevant issue since their structure is quite dynamic. The operational topology of such devices can often be described by complex networks. In this work, we assess the variation of measures commonly employed in the complex networks literature applied to wireless sensor networks. Four data communication strategies were considered: geometric, random, small-world, and scale-free models, along with the shortest path length measure. The sensitivity of this measure was analyzed with respect to the following perturbations: insertion and removal of nodes in the geometric strategy; and insertion, removal and rewiring of links in the other models. The assessment was performed using the normalized Kullback-Leibler divergence and Hellinger distance quantifiers, both deriving from the Information Theory framework. The results reveal that the shortest path length is sensitive to perturbations.

PACS (2008): 89.75.-k, 89.70.-a, 02.50.-r

Keywords: complex networks • structural measures • stochastic quantifiers • information theory quantifiers

© Versita sp. z o.o.

1. Introduction

Wireless Sensor Networks (WSNs) are an emerging technology that allows the monitoring of physical variables, such as temperature, sound, light, vibration, pressure or movement [1]. A WSN consists of a large number of wireless autonomous devices, called “sensor nodes”, “sensors” or “nodes.” These entities work in a cooperative way...
sensing the environment, communicating among them, and taking decisions. Such networks are a promising technology in a wide range of applications, for instance, biotechnology, industry, public health, and transportation [2]. They are of particular interest for monitoring hostile or inaccessible environments [3].

The physical variables are monitored and stored in the sensor, and propagated to a sink node. The sink is a management node responsible for processing the data and delivering it to an external user [4]. The communication uses the nodes between the sensor source and sink in an ad-hoc fashion. The two most common communication strategies are based on data flooding and on complex networks. In the flooding based, the data communication starts from a node to its direct neighbors, then each neighbor re-propagates to the next neighbors, and so on. Each node propagates each information once and this process repeats until the data arrives to the sink node. Fig. 1(a) illustrates this communication strategy.

In complex network based, the communication depends on the model used to characterize it (random, small-world or scale-free, for example). So, specific topological properties of complex networks are used to determine this kind of communication. Fig 1(b) shows a complex network based on scale-free model. An important issue in this model is the presence of nodes with high degree known as “hubs”. The communication is towards the nodes closer to the sink or to a hub. The hubs can propagate the information to distant nodes and, consequently, the information flow is concentrated on them.

Figure 1. Examples of WSNs communications

The WSN communication strategies can be characterized by a set of measures. These measures describe different features on the network, such as connectivity, centrality, cycles or distances [5]. Example of measures are the average shortest path length, clustering coefficient, network diameter and betweenness. In this work, we use the shortest path length, because it is directly related to the network energy consumption and sudden changes in this measure may result in an increase or decrease in energy consumption.

Energy availability is a critical feature in WSN [1]. Moreover, classical routing solutions in WSNs employ the shortest paths [6]. If the shortest paths change, the routing solutions may become inefficient.
The communication structure in WSNs is dynamic, i.e., it changes over time. It is, therefore, important to analyze how this reflects on the measures. The most common changes are: addition and inactivity of nodes, in the flooding based communication; and insertion, removal and rewiring of links, in the complex network based communication [7]. These changes have a great impact over real WSNs design. The infrastructure mechanisms usually are designed considering topological characteristics. For instance, in a routing algorithm based on shortest path length if a node is removed the routing behaviour is affected.

In this way, the main question of this work is:

“What is the impact of changes or perturbations on these strategies, as measured by the shortest path length?”

This work presents the analysis of communication strategies in WSNs by means of analyzing the variation of measures. We analyze flooding, random, small-world, and scale-free networks, and the shortest path length. The variation of measures was analyzed with respect to the insertion and removal of nodes in flooding; and with respect to insertion, removal and rewiring of links in the strategy based in complex networks. Stochastic quantifiers, namely the normalized Kullback-Leibler divergence and Hellinger distance [8], were used to quantify the variation of the measure. The results reveal to which extent the measure is influenced by the perturbations considered: the shortest path length exhibits a clear dependence on the type and intensity of the perturbation.

The paper is organized as follow. Section 2 presents the related work. Section 3 analyses communication strategies in WSNs. Section 4 presents the results of measures behavior and the quantifiers performance. Finally, conclusions and future directions are presented in Section 5.

2. Related work

This section presents a brief review about WSNs communication strategies that use complex networks concepts, and stochastic quantifiers applied to networks.

2.1. WSNs communication strategies

Helmy [9] described WSNs as spatial graphs that tend to be much more clustered and with higher path lengths than random graphs. He showed that it is possible to reduce the path length of wireless networks with the addition of a few short cut links.

Guidoni et al. [10] proposed on-line models to design heterogeneous sensor networks with small world features. The proposed model takes into account the data communication flow in this kind of networks to create shortcuts towards the sink in such a way that the communication between the sink and the sensor nodes is optimized. The network presents better small world features when the shortcuts are created, and an interesting trade off between energy and communication latency is observed.

Ruela et al. [11] improved the data communication by using hubs, an approach based on scale-free networks. The
hubs introduce useful characteristics of complex networks, e.g. small average shortest path length between all sensors and the sink, and high clustering coefficient. This strategy saves resources, avoiding excessive communication and, consequently, reducing the time to data delivery.

2.2. Information Theory quantifiers

Wang et al. [12] used entropy to provide an average measure of networks heterogeneity since it measures the diversity of the link distribution.

Boas et al. [7] analyzed the effect of perturbations in complex networks. They chose measures based on the fact that the network characterization is made from samples rather than from the entire network. They applied three perturbations: link addition, link removal, and link rewiring.

Carpi et al. [13] proposed a new quantifier based on Information Theory for the analysis of dynamic network evolution. It is used to compute changes in topological randomness for degree distribution of the network. The quantifier, a statistical complexity measure, is obtained by the product of the normalized Shannon entropy and the normalized Jensen-Shannon distance.

In this paper, we use two stochastic quantifiers, a divergence and a distance, to quantify the changes in WSNs. We show that it is possible to identify the strength of the perturbations that leads to a breakdown of the network properties, i.e., a kind of phase transition. Moreover, we evaluate the flooding communication strategy considering node perturbations. These aspects are not addressed in the related works mentioned and constitute the main contributions.

3. Analysis of communication strategies in WSNs

The WSN physical connectivity can be described as an undirected graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N\}$ is the set of sensor nodes, and $E = \{(v_i, v_j) : 1 \leq i, j \leq N\}$ is the set of links between nodes. Each link is determined by the communication geometry. In our case it is a circumference, i.e., the neighborhood of node $v_i$ is formed by all nodes $v_j$ at most $\mathcal{R}$ units away from $v_i$. The data employ this physical connectivity to propagate, but obeying the rules of a protocol. The protocol may impose restrictions on the use of the available links $E$, resulting in a communication subgraph $G = (G, E)$, such that $E \subset E$ and, therefore, $G \subset G$.

The number of connections of a node is its degree $\kappa$. An important graph characteristic is its degree distribution, i.e., the collection of all its degrees.

The flooding based communication employs all the available links, so $G = G$. The flooding starts propagating the data from $v_i$ to all its neighbors, then each neighbor re-propagates to its neighbors, and so on. Each node propagates each unit of information once. This process repeats until the data arrives to the sink node.

The communication based on complex networks employs a subset of links $E \subset E$ such that data flow towards the nodes which are closest to the sink. The source nodes always send the information to that neighbor which reduces
the distance to the sink. Each node propagates the information following the same criterion. This process repeats
until the data arrives to the sink node.

In some cases the path to the sink is not the optimal, because it is not always immediate to find the optimal
routing solution in distributed scenarios. These scenarios offer only local information, a limitation which often
prevents finding global optima. Specific measures of complex networks are used to characterize this kind of
communication, for instance, the shortest path length [14, 15]. A shortest path is any path that connects two
nodes and has minimal length. It is an important measure in communication strategies.

There are several complex networks models. In this work, we use the random, small-world, and scale-free models:

**Random model** The probability to connect each pair of nodes is the same. There are two ways to build a random
graph [16]: \( N \) nodes and exactly \( M \) links, and \( N \) nodes and the probability \( 0 < p_c < 1 \) to connect each pair
of nodes. In the first description, \( M \) links are uniformly distributed among the \( N(N-1)/2 \) possibilities. In
the second description, which is the one we adopted here, we start with a totally disconnected graph and
then connect each pair of nodes with probability \( p_c \). In this case, the probability of observing \( 0 \leq k \leq N-1 \)
connections in each node follows a Binomial distribution with \( N-1 \) trials and probability of success \( p_c \),
leading to \( (N-1)p_c \) as mean degree.

**Small-world model** A communication strategy is small-world when the communication network presents a high
clustering coefficient and a small shortest path length [17]. There are several ways to build such strategy;
in this work, we use the Watts-Strogatz model. This model starts with a circular regular topology with \( N \) nodes,
each one connected to the \( k \) nearest neighbors in each direction (right or left in circular topology).
Then, each link is randomly “rewired” with probability \( p_r \) [5], i.e., if the current link is \((v_i, v_j)\) and there is
no link \((v_i, v_k)\), rewiring consists in deleting \((v_i, v_j)\) and creating \((v_i, v_k)\).

**Scale-free model** A communication strategy is scale-free when the communication network displays a power law
degree distribution \( p(\kappa) \sim \kappa^{-\lambda}, \kappa > 0 \), with \( 2 < \lambda < 3 \). The main feature of this topology is the presence
of some nodes with high degree, often called “hubs”. To generate this topology we use the Barabási-Albert
scale-free model that starts with a small number of nodes with \( m_0 \) links, and in the next step a new node
with \( m \) links is added to the network (with \( m \leq m_0 \)). The probability connection is linearly proportional
to the node degree [18].

The main question stated in this work is: “What is the impact of changes or perturbations on these strategies, as
measured by the shortest path length?” To answer this question, we need to quantify the variation of the measure
when perturbations occur. The most common changes, or perturbations, are: the addition and inactivity of nodes,
in the flooding based communication; and insertion, removal and rewiring of links, in the complex network based
communication.

The stochastic nature of the aforementioned quantifiers suggests the use of techniques deriving from Information
Theory in order to assess their change. The normalized Kullback-Leibler divergence and Hellinger distance are
two quantities suitable for describing the difference between distributions [8].

Consider the discrete random variables $X$ and $Y$ defined on the same sample space $\Omega = \{\xi_1, \xi_2, \ldots, \xi_n\}$. The distributions are characterized by their probability functions $p, q : \Omega \to [0, 1]$, where $p(\xi_i) = \Pr(X = \xi_i)$ and $q(\xi_i) = \Pr(Y = \xi_i)$. A metric $D$ between these two distributions is a quantifier obeying:

1. $D(p, p) = 0$, reflexivity;
2. $D(p, q) \geq 0$, non-negativity;
3. $D(p, q) = D(q, p)$, commutativity;
4. $D(p, q) \leq D(p, r) + D(r, q)$, triangle inequality for any other probability function $r$ defined on the same probability space.

A distance is not required to satisfy property 4, and a divergence is only required to satisfy properties 1 and 2 [19]. Assuming $q(\xi) > 0$ for every event $\xi \in \Omega$, the Kullback-Leibler divergence is defined as

$$D_{KL}(p, q) = \sum_{\xi \in \Omega} p(\xi) \log \frac{p(\xi)}{q(\xi)}. \quad (1)$$

The Hellinger distance does not impose positivity on the probabilities; it is defined as

$$D_H(p, q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{\xi \in \Omega} \left(\sqrt{p(\xi)} - \sqrt{q(\xi)}\right)^2} = 1 - \sum_{\xi \in \Omega} \sqrt{p(\xi)q(\xi)}. \quad (2)$$

In order to make the Kullback-Leibler divergence (an unbounded positive quantity) and the Hellinger distance (which is confined to the $[0, 1]$ interval) comparable, in the remainder of this work we will use the normalized Kullback-Leibler distance defined as $D_{KL}(p, q) = 1 - \exp\{-D_{KL}(p, q)\}$.

### 4. Simulation results

This section presents the simulation study about the complex network measures behavior of WSNs using two quantities presented previously.

#### 4.1. Methodology

Simulation assumptions and parameters were:

**Network parameters** In order to simulate a sparse WSNs, we used $N = 1000$ nodes deployed in an $L^2 = 100 \times 100$ area. The communication radius of each node was $R = 5$ units in strategies based in flooding communication. With these values, we got a network density approximated 1.5 obtained through $d = \pi R N / L^2$, where $d$ represents the number of neighbours of each nodes [20]. The probability of connection
in the random model was $p_c = 0.06$, which implies an average degree equal to 6. The nearest neighbors and probability rewiring in the small-world model were $k = 3$ and $p_r = 0.3$, respectively. The number of links added in each step in the scale-free model was $m = 1$. The above parameters were chosen according to Boas et al. [7].

**Perturbations** In the strategies based in flooding communication we performed: (i) *nodes addition*, nodes were randomly added to the network; and (ii) *nodes removal*, nodes were randomly removed from the network, that represents the node inactivity. The perturbations were performed on $\{1\%, 2\%, \ldots, 10\%\}$ of the total number of nodes. In the complex networks based communication, we performed tree types of links perturbations [7]: (i) *link removal*, links were randomly removed from the network; (ii) *link addition*, two unconnected nodes were randomly selected, and a new link was established; and (iii) *link rewiring*. In this last case, the perturbations were performed in $\{1\%, 2\%, \ldots, 10\%\}$ of the total number of links.

**Normalization** for each network, the normalized histogram $H$, also known as histogram of proportions, was obtained with 200 bins of equal width. The $D_{KL}$ diverges for $q(\xi) = 0$ and $p(\xi) \neq 0$, as defined in equation (1). In order to avoid the division by zero, a small positive constant $\delta = 0.001$ was added to each bin, and then the histogram is normalized to add 1 [7, 21]. The original histogram is used to compute the Hellinger distance once it does not impose the positivity restriction on the probabilities.

**General parameters** for each communication strategy, we generated 10 different networks and for each network 100 different perturbations were made. In this way, we are able to present the mean results with symmetrical asymptotic confidence intervals at the 95% significance level.

**Computational resources** we performed the evaluation using the R platform [22], on an Intel(R) Core(TM) i5 CPU 760 2.80 GHz with 7 GB RAM, running Ubuntu 12.04 (64 bits). The igraph library was used to generate and modify the graphs [23].

Figures 2 and 3 present the variation of the shortest path length, for the different communication strategies and the perturbations considered. Each plot presents two quantifiers: the normalized Kullback-Leibler divergence $D_{KL}$ (denoted as “△”), and the Hellinger distance $D_H$ (denoted as “◦”), as functions of the level (intensity) of the perturbation.

### 4.2. Flooding communication

Figure 2 shows the shortest path variation in flooding communication; from left to right, node addition and node removal. Both addition and removal alter the shortest path length, the bigger the level of perturbation the stronger the change. The Hellinger distance is more sensitive to changes in the shortest path length than the normalized Kullback-Leibler divergence. The variability of both measures has some dependence on the level of perturbation; it increases with the level of perturbation but then stabilizes. The removal of nodes has stronger
impact on the shortest path length than node addition (the former variation is steeper than the latter); this effect is more noticeable with higher levels of perturbation. This may be explained by the fact that when new nodes are added within the communication range of a node, the shortest path is little affected. The shortest path length changes when nodes are removed because many links are lost and some of them belonged to the shortest paths.

The quantifiers are, thus, sensitive to both node addition and removal in this scenario, but with different responses. The Hellinger distance is the one which varies most, reaching, in mean, 0.032 for the addition and 0.054 for the most intense removal of nodes (10%).

The results presented are important to WSNs designers. Based on this study, a more efficient topology control can be applied. For instance, when 10% of nodes are added the shortest path length changes. This change indicates that adding more than 10% of nodes in the original network could be harmful. When 5% of nodes die (i.e. are removed) the shortest path length changes. This information can be used by the designer to calibrate the duty cycling operation, where each node periodically switches between sleeping mode and awake mode [24].

### 4.3. Complex network communication

Figure 3 presents the results in the random, small-world and scale-free communication models (top to bottom rows) and the three types of perturbations: link addition, link removal and link rewiring (left to right columns). The shortest path length is sensitive to the link addition and removal applied in the three communication models, and both quantifiers ($D_{KL}$ and $D_H$) behave alike: the stronger the perturbation, the more the quantifier changes in direct proportion.
Regarding these two perturbations, again, the Hellinger distance exhibits more intense variations with respect to the perturbation than the normalized Kullback-Leibler divergence; notice that the confidence intervals do not overlap. Table 1 shows the mean value of the most intense perturbation (10%); this corresponds to the rightmost point of the plots of Figure 3. The Hellinger distance is consistently and significantly higher than the Kullback-Leibler divergence by a factor of, approximately, two, in all the situations where link addition and removal are applied.

The behavior of link addition and removal is the same in the tree models. Both affect the shortest path length, since these perturbations increase or decrease the distance between nodes. In particular, networks where all nodes have the same degree (regular graphs) may become small-world with just a few reconnections [17], i.e., randomly adding or removing links may result in connecting nodes which are far away, reducing the shortest path length. Link rewiring has a different behavior. Although the quantifiers exhibit differences, see Table 1, they are negligible in the random model. Link rewiring alters small-world models, but its effect seems constant, i.e., independent of the intensity of the perturbation. Regarding the scale-free model, there is a strong variation of both measures when small perturbations are applied, but the change tends to stabilize, i.e., saturates, soon after. The former is the only case where the normalized Kullback-Leibler divergence is bigger than the Hellinger distance.

The main feature of these networks is the presence of hubs, which makes them sensitive to this kind of perturbation: the shortest path length alters radically whenever a link involving hubs is added or removed. Additionally, the removal of some links makes the network disconnected.

Table 1. Mean most intense perturbation (10%) in the Hellinger distance and normalized Kullback-Leibler divergence.

|                      | Hellinger distance | Normalized Kullback-Leibler divergence |
|----------------------|--------------------|----------------------------------------|
|                      | Link addition | Link removal | Link rewiring | Link addition | Link removal | Link rewiring |
| Small-world model    | 0.122    | 0.119     | 0.073     | 0.095    | 0.071     | 0.033     |
| Scale-free model     | 0.105    | 0.082     | 0.261     | 0.076    | 0.043     | 0.378     |
| Random Model         | 0.092    | 0.105     | 0.006     | 0.052    | 0.058     | 10^{-4}   |

All models analysed represent a routing topology generated by some management strategies. Generally, these strategies combine the shortest path length with other QoS parameter (energy, delay or priority). The results reveal that when 5% of the links are added, or removed, the shortest path length changes. This information could be used to calibrate the management strategies to avoid the interference of QoS parameter during the routing generation, once this task considers just the links addition or removal.

5. Conclusion

The analysis of the variability of measures in WSNs communication strategies provides important information. It gives an insight of the behavior of the network when it is perturbed and it helps us in the design of appropriate solutions for every application.
In this paper, we used the normalized Kullback-Leibler divergence and the Hellinger distance to compare three communication models: flooding, random, small-world and scale-free (the last three are members of complex network based strategies). We performed two types of perturbations in the flooding-based strategy: addition and removal of nodes. In the complex network strategy, we used three types of perturbation: addition, removal and rewiring of links. We analyzed how the shortest path length changes with respect to different levels of each perturbation.

---

**Figure 3.** Hellinger distance and normalized Kullback-Leibler divergence for the shortest path length to the random network model, small-world network model and scale-free network model.
The shortest path length is sensitive to these changes, and in most situations it alters accordingly to the intensity of the perturbation. The analysis allows identifying the relationship between the strength of the perturbations and the change of the shortest path length. The use of quantifiers that involve logarithms or ratios may not be a good choice for this kind of characterization, because the occurrence of zeros leads to numerical problems and, possibly, to incorrect interpretation of network changes.

In addition, more efficient topology control or routing strategies in WSNs can be proposed. For instance the duty cycling or routing operations can be based on shortest path length sensibility. In the specifics scenarios treated, these mechanism could be calibrated when 5% – 10% of the nodes or links are added or removed. The impact that the observed results have on the design and operation of WSNs was commented.

**Acknowledgement**

This work is partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) and the Research Foundation of the State of Alagoas (FAPEAL). O. A. Rosso gratefully acknowledges support from CONICET, Argentina and CNPq fellowship, Brazil.

**References**

[1] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., IEEE Communications Magazine 40 (2002) 102.

[2] Arampatzis, T., Lygeros, J., and Manesis, S., A survey of applications of wireless sensors and wireless sensor networks, in *13th IEEE Mediterranean Conference on Control and Automation (MED’05)*, pages 719–724, Hawaii, USA, 2005, IEEE Computer Society.

[3] Buratti, C., Conti, A., Dardari, D., and Verdone, R., Sensors 9 (2009) 6869.

[4] Frery, A. C., Ramos, H. S., Alencar-Neto, J., Nakamura, E., and Loureiro, A. A. F., Sensors 10 (2010) 2150, Special Issue: Wireless Sensor Network and Its Application in Advanced Computer Science.

[5] Costa, L. D., Rodrigues, F. A., Travieso, G., and Boas, P. R. V., Advances in Physics 56 (2008) 167.

[6] Akkaya, K. and Younis, M., Ad Hoc Networks 3 (2005) 325.

[7] Boas, P. R. V., Rodrigues, F. A., Travieso, G., and da F Costa, L., Journal of Statistical Mechanics: Theory and Experiment 2010 (2010) P03009.

[8] Nascimento, A. D. C., Cintra, R. J., and Frery, A. C., IEEE Transactions on Geoscience and Remote Sensing 48 (2010) 373.

[9] Helmy, A., IEEE Communications Letters 7 (2003) 490.

[10] Guidoni, D. L., Mini, R. A., and Loureiro, A. A., Computer Networks 54 (2010) 1266.

[11] Ruela, A. S., da Silva Cabral, R., Aquino, A. L. L., and Guimares, F. G., International Journal of Natural
Computing Research 1 (2010) 33.

[12] Wang, B., Tang, H., Guo, C., and Xiu, Z., Physica A 363 (2006) 591.

[13] Carpi, L. C., Rosso, O. A., Saco, P. M., and Ravetti, M. G., Physics Letters A 375 (2011) 801.

[14] Newman, M. E. J. and Watts, D. J., Physical Review E 60 (1999) 7332.

[15] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U., Physics Reports 424 (2006) 175.

[16] Erdős, P. and Rényi, A., Publication of the Mathematical Institute of the Hungarian Academy of Sciences 5 (1960) 17.

[17] Watts, D. J. and Strogatz, S. H., Nature 393 (1998) 440.

[18] Barabsi, A.-L. and Albert, R., Science 286 (1999) 509.

[19] Cha, S.-H. and Srihari, S. N., Pattern Recognition 35 (2002) 1355.

[20] Aquino, A. L., Junior, O. S., Frey, A. C., de Albuquerque, E. L., and Mini, R. A., IEEE Transactions on Computers (2013) in press (doi.ieeecomputersociety.org/10.1109/TC.2012.229).

[21] Cabella, B. C. T. et al., Brazilian Journal of Physics 38 (2008) 20.

[22] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0.

[23] Csárdi, G. and Nepusz, T., InterJournal Complex Systems (2006) 1695.

[24] Lai, S. and Ravindran, B., Efficient opportunistic broadcasting over duty-cycled wireless sensor networks, in 29th IEEE Conference on Computer Communications (INFOCOM’2010), San Diego, USA, 2010.