Is the Clustering Coefficient a Measure for Fault Tolerance in Wireless Sensor Networks?

Matthias R. Brust*, Damla Turgut*, Carlos H.C. Ribeiro† and Marcus Kaiser‡

*Department of Electrical Engineering and Computer Science
University of Central Florida, Orlando FL
Email: {mbrust,turgut}@eecs.ucf.edu

†Computer Science Division
Technological Institute of Aeronautics, Brazil
Email: carlos@ita.br

‡School of Computing Science, Newcastle University, UK and
Department of Brain and Cognitive Sciences, Seoul National University, Korea
Email: m.kaiser@ncl.ac.uk

***Abstract*—Distributed systems such as the Internet and wireless sensor networks must provide a high degree of resilience against errors and attacks. Besides steps that increase reliability of data and resources of the network, the topology structure itself plays a crucial role in the efficacy of the fault-tolerance behavior. The network topology is a supportive factor to reduce or avoid malfunction behavior of the system after a strike on a strategic node or a random failure of a node. For a self-organizing topology with numerous nodes, it is necessary to have a local fault tolerance measure available instead of collecting information of the entire network to adjust the topology locally when needed. The *local clustering coefficient* measure determines the degree of how strong connected the neighbors of a node are. The correlation between the clustering coefficient and fault tolerance is an open research problem.**

In this paper, we propose the clustering coefficient as a local metric for fault tolerance, in particular for wireless sensor networks. We describe how to increase the clustering coefficient by (a) exclusively adding and (b) exclusively removing links to a wireless sensor network topology. Simulation results indicate that the clustering coefficient is correlated to the fault tolerance of the system.

I. Introduction

Wireless sensor networks (WSNs) are composed of sensor nodes, which transmit their data through hop-by-hop wireless communication [1]. The nodes of a WSN are tiny devices equipped with several sensors, efficient computing and memory units, and wireless networking adapters.

The combination of sensing capabilities and wireless communication enables the nodes to gather and process environmental data locally, aggregate and forward them to a gateway node that evaluates the data [2]. A wireless sensor network can be formed spontaneously whenever devices are in transmission ranges of each other. Joining and leaving of nodes occur dynamically, particularly when dealing with mobility. The flexibility of WSNs allows diverse application scenarios: to monitor the flow of the cars in traffic, under water to monitor seismic activities, swarm behavior monitoring of animals, environmental observations or search and rescue scenarios [3].

Due to their distributed nature, wireless sensor networks introduce challenging concepts on how the information is computed and communicated. The WSNs must compensate the lack of authority by self-organization and the locality principle. The lack of central nodes in the network requires data has to be handled on the node itself. To learn more about the network, the node must gather information from its neighboring nodes, i.e. nodes that are directly (1-hop) or through intermediate nodes (k-hop) connected with this node. Algorithms that process data on the sensor nodes exchanging data locally without generating a global state are defined as *local algorithms* [4] and are an essential concept of self-organizing systems.

An outstanding characteristic of any self-organizing system is its capability to react and adapt to unforeseen influences such as sudden failures preventing the system to behave unexpectedly [5]. Therefore, the fault tolerance is an important facet for WSNs.

The nodes of a WSNs can locally control the network topology by increasing transmission ranges or dropping links from the neighboring lists. Changes in the topology can affect the network's receptivity to failures. In order to reach this goal, it is important to have a local measure available, which validates the impact on the topology in real-time.

In this paper, the clustering coefficient metric is analyzed for its efficiency to measure the fault tolerance of a WSN. The clustering coefficient (C) measures the degree of how strongly the neighbors of a node are clustered and is 2-locally defined [6]. The contribution of this paper is the demonstration of the correlation between the clustering coefficient and the fault tolerance for WSNs. Fault tolerance is measured by the impact of arbitrary failures on the characteristic shortest path length in a network. This is an appropriate measure for fault

tolerance since an high impact on the path length subsequently requires more resources for communication [7]. Simulation results demonstrate that a high C leads to a more resilient network in terms of sudden and arbitrary link failures and even targeted attack.

II. Related Work

Wireless sensor networks suffer in practice from unpredictable factors such as limited battery lifetime, interference, noise and temporary link failures or even targeted attacks.

Topology control algorithms reduce the number of links to optimize energy consumption while reducing interferences. However, reducing links decreases in turn the degree of connectivity, which makes the topology more susceptible to failures.

A few related protocols have been proposed to control the network's topology in order to guarantee fault tolerance, i.e. connectivity while minimizing energy consumption.

The *LMST* (Local Minimum Spanning Tree) protocol [8] aims to find the minimum transmission range where all devices in the network are still connected by constructing minimum spanning trees locally. The *LMST* protocol has a message complexity of $O(n)$.

The Cone-Based Topology Control ($CBTC$) protocol [9] is concerned with connectivity in static network topologies. The optimization criteria are preserving connectivity and minimizing energy consumption by removing energy inefficient links. Additionally, *CBTC* has been extended to provide fault tolerance in terms of k-connectivity [10].

The Fault-tolerant Local Spanning Subgraph (*FLSSk*) protocol is proposed by [11]. The protocol is fully localized and preserves k-node connectivity while maintaining bi-directionality.

Fault tolerance can also be analyzed on the overall impact of failures or attacks on the network performance in terms of characteristic path length between a pair of nodes. Basagni et al. [12] states that a high C supports local information spreading as well as a decentralized infrastructure. Networks with a high C show a faster global response on local impacts whereas local sensitivity implies low fault tolerance. On the other hand, Latora et al. [7] defines the efficiency measure and argues that the local efficiency measure reveals how much a system is fault tolerant, since it shows "how efficient the communication is between the first neighbors of i when i is removed". They also show that the local efficiency measure is directly related to C implying that a high C correlates to an increased fault tolerance.

Originally, C is introduced in combination with the characteristic path length measure in order to describe the behavior of small-world networks [6], which are known to be remarkably resilient against failures [7].

Despite the large use in the field of network analysis, the impact of a high C is ambiguously discussed in the literature above. Therefore, in this paper, we investigate the significance of C to the fault tolerance property of a network.

III. Fault Tolerance Optimization Approach

A. System model and fault tolerance

All nodes considered are stationary and have the same transmission range r.

The communication graph for WSNs is constructed as follows. Let $V \in \mathbb{R}^2$ be a set of nodes in the 2-dimensional bounded region with side length l. The nodes are deployed uniformly at random. The links E of the symmetric Euclidean graph $G{=}(V, E)$ fulfill the condition that for any pair $u, v \in V$ of nodes, $dist\,(u,v) \leq r \implies \{u, v\} \in E$ and $dist\,(u,v) > r \implies \{u, v\} \notin E$.

The neighborhood of a node v is formally defined as a subgraph S that consists of all nodes adjacent to v. The 2-neighborhood or 2-hop neighbors of a node v is then the subgraph that consists of all nodes adjacent to any of the nodes in S, but not including the nodes of S. This can be generalized for k-neighborhood or k-hop neighbors. Increasing k often implies, an exponential increase of the message complexity. For that reason k should remain low.

Calculating the fault tolerance using a local algorithm is in many ways is appealing. By knowing the local fault tolerance degree, actions can be taken on a node to change the topology to increase further the fault tolerance level. Therefore, a local measure that can be correlated to the global fault tolerance degree of a system is a characteristic with practical implications.

In this paper, the clustering coefficient measure is analyzed on its correlation to the fault tolerance in a WSN. The clustering coefficient measures the degree of how strongly nodes are clustered in a network [6]. The *local* clustering coefficient C_v of a node v with k_v neighbors is $C_v = \frac{|E(\Gamma_v)|}{k_v(k_v-1)}$ where $|E\,(\Gamma_v)|$ is the number of links in the neighborhood of v and $k_v(k_v-1)$ is the total number of possible links in the neighborhood of v. The *global* clustering coefficient C of a graph $G = (V, E)$ is then the average of all local clustering coefficients in the network denoted as $C = \frac{1}{n}\sum_v CC$, where n is the number of nodes in G denoted as $n = |V|$. It is called as clustering coefficient, C, in this paper. A high C means that the network consists of a high number of locally clustered nodes, i.e. C reflects the probability that a randomly chosen pair of nodes $v_1, v_2 \in V$ that are connected $(v_1, v_2) \in E$ have a mutual neighbor $v_3 \in V$ with $(v_1, v_3) \in E$ and $(v_2, v_3) \in E$.

In this paper, fault tolerance is measured by the impact of failures on the characteristic path length. The characteristic path length L is the mean of the means of all the shortest path lengths connecting each node to all other nodes. That is, given length of the shortest path between two nodes $d(v_1, v_2)$ for $\forall v_1, v_2 \in V$, the characteristic path length L is $\frac{1}{n(n-1)}\sum_{i,j} d(v_i, v_j)$. The clustering coefficient reflects local characteristics of a network, how well neighbors of a node can reach each other (local efficiency), while the characteristic path length is a global characteristic, how well any node in the network can be reached (related to global efficiency) [Latora2001b].

B. Topology management

The approach is to generate topologies of WSNs with different clustering coefficients and measure the impact of sudden link failures on the path length.

The most complicating factor for manipulating the topology of a WSN is the limited transmission range of the devices that prevent adding arbitrary links between two devices, unless they are in transmission range. Besides this physical restriction, the usage of maximum transmission range additionally affects the network capacity due to an increased number of interferences and a high energy consumption [13].

Two approaches are considered to analyze the characteristics of the clustering coefficient in a WSN: (1) removing dedicated links (for maximum transmission range) and (2) increasing transmission range (for configuration with energy-efficient setting of the transmission range).

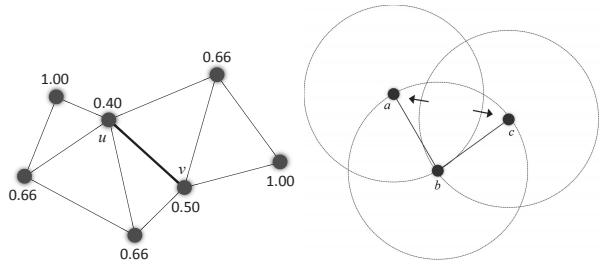

Figure 1. Increasing the C: Removing dedicated links (left) and increasing the transmission range r (right).

1) Removing selected links: The removal of selected links increases the clustering coefficient C globally (Fig. 1) [14].

The algorithm described in [14]verifies for each link (u, v) if its removal increases C. In this case, the link (u, v) is considered as a candidate for removal until the 2-hop node neighborhood of the set $\{u, v\}$ performed the same task.

In the next step, nodes exchange the candidate with their neighbors. In case that the neighbors have the same remove-candidate and the link (v, u) is removed; otherwise the link (v, u) is not removed.

Connectivity is guaranteed by the criterion that removing (u, v) requires at least one neighbor of u to be connected to one neighbor of v.

2) Gradually increasing transmission range: Although a removal of links as described in III-B1 is a practical approach, in many cases - in particular in sparse networks - a removal can cause a disconnected network. Therefore, approaches have to be considered that increase the the number of links.

The transmission range r of the devices can be set to a minimum range $r = min$ required to establish communication in a network. In practice, min is often set to a value close to 0. According to the fault tolerance criteria or threshold, devices increase r gradually until the criteria is reached or the transmission range is the maximum transmission range max.

IV. Simulation Study

A. Settings and metrics

For each simulation, 130 nodes with a transmission range of $50m$ are deployed uniformly at random in a squared area with a side length of $350m$. A connected geometric random graph is generated according to the graph construction described in Section III-A such that nodes which are in transmission range of each other are neighbors. Preservation of connectivity is guaranteed by a local condition of the optimization policy.

In order to gain connectivity, the average node degree k must support a connected network w.h.p., e.g. for geometric random graphs $k > 6$ [15] and $k > \frac{\ln n}{n}$ for Erdős–Rényi model [16].

Sudden failure and targeted attack on the network are simulated by following procedures.

- *Sudden failure*: removing uniformly at random a fraction of p links in the network.
- *Targeted attack*: removing a fraction of p links in the network uniformly at random around a randomly chosen node u with the condition that at least one node of the link must be in the Euclidean distance r to u.

In order to measure the impact of a relatively small influence on the network, the fraction of links to be removed is set to $p = 1\%$.

In the next step, d and d' are determined by measuring L before and after a sudden failure as well as a targeted attack for following cases:

G: Initial network (d, d')
G_o: C-optimized network (d_o, d'_o)
G_s: Initial network type but with same number of nodes and links as in (2) ($\bar{d}_s$, $\bar{d}'_s$)

After taking the average of each value, the impact i of the failures on the topology is measured by $i = \bar{d}' - \bar{d}$, $i_o = \bar{d}'_o - \bar{d}_o$ and $i_s = \bar{d}'_s - \bar{d}_s$. i, i_o and i_s.

For the second scenario discussed in Section III-B it is assumed that the transmission range is set to a low value (e.g. energy-efficient value). This situation makes it possible to increase the transmission range gradually to catch neighboring nodes increasing the number of links. In the simulation, the transmission range is initially set to 0 and increased by 1 until almost a complete graph is created. At each step, C is calculated. 250 simulation runs are executed for each setting.

B. Results and discussion

Table I
Clustering coefficient C, characteristic path length L, and average node degree k for the initial network G, optimized network G_o, and a network with same number of link as the optimized network G_s in case of a failure and attack.

Failure	G	G_o	G_s	Attack	G	G_o	G_s
C	0.63	0.70	0.62	C	0.63	0.70	0.62
L	5.32	6.10	6.57	L	5.06	5.80	6.02
k	8.74	7.08	6.99	k	8.55	6.96	6.99

A cumulative sum is created from each of the impact differences i, i_o and i_s, and the resulting vector for failures

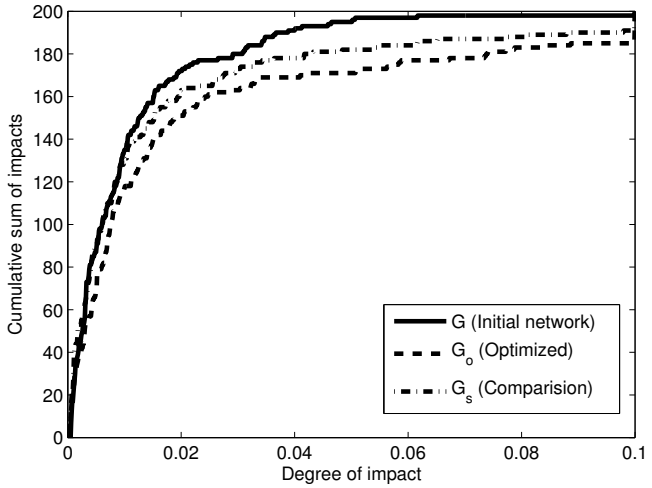

Figure 2. Test of robustness to failures with respect to the characteristic path length. The impact differences i, i_o and i_s are shown: the initial network G has been optimized by removing dedicated links and increasing C (G_o). A third network has been included into the graph, in which number of links corresponds to that of the optimized network (G_s).

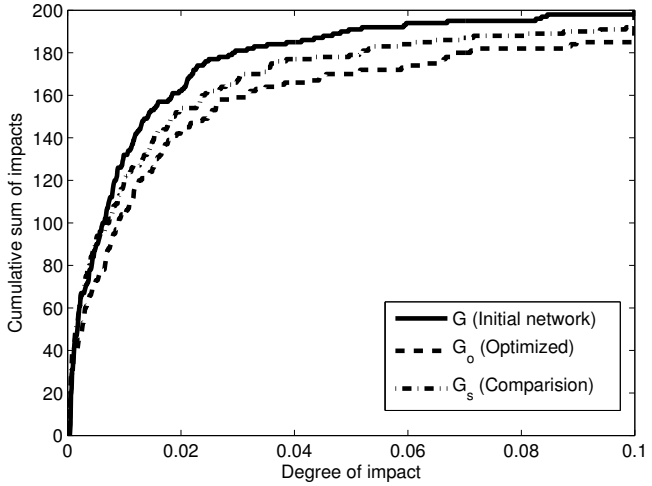

Figure 3. Similar to Fig. 2 is the test of robustness to attacks with respect to the characteristic path length. The impact differences i, i_o and i_s are shown: the initial network G has been optimized by removing dedicated links and increasing C (G_o). A third network has been included into the graph, in which number of links corresponds to that of the optimized network (G_s).

is shown in Fig. 2 and for attacks in Fig. 3. Corresponding values for C, L and k for each network are shown in Table I. Fig. 2 and Fig. 3 show the cumulative sum of the number of appearances of each impact degree. For all networks, there are many small impacts, which is illustrated by the rapid increase of the curves to about 0.4. From then on, the curves are getting flatter, illustrating that impacts with higher values are less frequent than smaller impacts.

Observing the difference between the impact for the original and the optimized networks, our approach reduces the total impact by 50% for small network changes (cumulative sum 0.02) and by 20% for larger network changes (cumulative sum 0.1). Therefore, our approach is especially useful for small network failures, which occur more frequently than large-scale failures involving the whole network.

Interestingly the behavior for a sudden failure and a targeted attack is comparable. This indicates that spatially localized node removal under the attack condition is still insufficient to destroy communication in local neighborhoods. Therefore, wider node removal at a regional spatial scale seems necessary to reach a larger effect than for a random failure.

All curves follow the same behavior; the initial network G suffers a more total impact than the network with optimized C (by removing links) (G_o). G_o does not grow as fast as G. How can this be explained?

For the network G which has a lower C than G_o or G_s, it can be observed that a high number of failures caused a small increase of L. However, big changes in the L value due to link failures are rare. In contrast, results with the clustering coefficient optimization enabled have considerably smaller values that change L. This means that certain numbers of impacts are small, but a higher number of impacts are higher compared to the network with lower clustering coefficient.

Although the number of higher impacts are higher for the high C network, the total sum of the impact is always lower than the lower C network. Interestingly this is valid, although the optimized network has about 15% to 20% fewer numbers of links, which could be expected to make the network more receptive to sudden failures. This effect is against our intuitive understanding that more resources imply higher fault tolerance. By removing resources from the network, we would expect a decrease of the tolerance against faults.

One important aspect of a high C is its stabilizing effect of sudden failures on L. In contrast to the initial network topology, a network with an optimized C consists of different node types: normal nodes and hubs. All nodes tend to cluster their direct neighbors and few of these clustered nodes have extra links connecting their clusters to other clusters. Assuming that the network will be affected by randomly occurring failures, the probability to hit a hub is smaller than to hit a normal node. In other words, a randomly occurring failure is unlikely to increase L, thus keeping L more stable (in terms of percent of change in regard to the number of failures). In contrast, due to the low clustering coefficient, a random network, for example suffers steady increases of L with random failures [17].

The network G_o is more tolerant to faults. However, due to the optimization procedure, links have been removed and therefore the average node degree k changed. In order to find out if the solely change in k is relevant to the increased fault tolerance and not the structural optimization, the same simulation has been done for a non-optimized network that has the same k as G_o and results have been included (G_s) into Fig. 3 and Table I. The figure shows that even if the resources are the same, the optimization towards a higher C has a significant impact on the fault tolerance of the network.

For a network configuration where the transmission range is set to a minimum value to take into account energy efficiency, new connections to the communication graph of the WSN

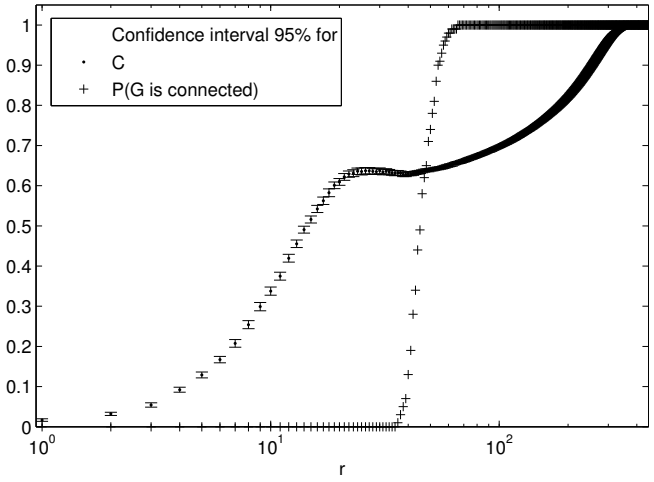

Figure 4. The effect on the clustering coefficient C and the probability P of a fully connected network when increasing transmission range r from 0 to the maximum.

can be added by increasing the transmission range up to its maximum value. In the simulation, the transmission range is increased gradually from 0 to l, forming almost a complete graph at the end. Fig. 4 shows how C behaves when increasing the transmission range. Interestingly, C increases very quickly in the beginning until it comes to a plateau and then later increases to the value 1 which is the case for constructing a complete graph. This indicates that a stronger local neighborhood is the best strategy for increasing the transmission range for sparse networks with otherwise low transmission range. For denser networks with higher transmission range, connecting nodes at the regional and global level, outside local neighborhoods, seems more efficient for further increasing the transmission range.

In Fig. 4, consider C for transmission ranges $r = 30$ and $r = 57$ are almost the same and form a plateau, which has practical impact on the choice of the transmission range. If the network designer has to take into account a fault tolerant network and energy efficiency, the compromise can be found in choosing the values around the plateau.

V. Conclusion

The results gained in this paper suggest a strong correlation between the clustering coefficient C and the fault tolerance of a network.

Networks with a higher C do not suffer drastic changes of characteristic path length L when the network gets affected. The higher C is, the fewer nodes with extra links exist, thus the probability to hit a hub by random failure is smaller. This makes a network with its higher C more tolerant of sudden failures.

The clustering coefficient is particularly attractive as this measure can be determined 2-locally instead of requiring information of a partition of the network.

The results motivate to analyze the clustering coefficient with additional fault tolerance metrics to infer their correlations. Additionally, a gradual increase of the area of impact and the fraction of links may further contribute to understanding.

Acknowledgment

Matthias R. Brust gratefully acknowledges support by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) under grant AP.PIPE 1 2010/51435-9. The authors thank Mustafa İlhan Akbaş for helpful discussions. Carlos H.C. Ribeiro is grateful to CNPq (Grant no. 301228/97-3-NV) and FAPESP (Grant no. 2010/11334-9). Marcus Kaiser acknowledges support by the WCU program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R32-10142) and by EPSRC (EP/G03950X/1) and the CARMEN e-science project (http://www.carmen.org.uk) funded by EPSRC (EP/E002331/1). (http://www.biological-networks.org/2012-03-10-2837/). Ok, next try (http://www.biological-networks.org/2012-03-10-2837).

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, pp. 2292–2330, Aug. 2008.

[2] M. R. Brust, M. I. Akbas, and D. Turgut, "Multi-hop localization system for environmental monitoring in wireless sensor and actor networks," *Wiley Journal of Concurrency and Computation: Practice and Experience*, Aug. 2011.

[3] M. I. Akbas, M. R. Brust, and D. Turgut, "Local Positioning for Environmental Monitoring in Wireless Sensor and Actor Networks," in *Proceedings of the 35th Annual IEEE Conference on Local Computer Networks (LCN)*, pp. 822–829, 2010.

[4] R. Wattenhofer, "Sensor Networks: Distributed Algorithms Reloaded - Or Revolutions?," in *13th Colloquium on Structural Information and Communication Complexity (SIROCCO)*, pp. 24—-28, 2006.

[5] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, July 2000.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–2, June 1998.

[7] V. Latora and M. Marchiori, "Efficient Behavior of Small-World Networks," *Physical Review Letters*, vol. 87, no. 19, p. 4, 2001.

[8] J. Cartigny, F. Ingelrest, D. Simplot-Ryl, and I. Stojmenović, "Localized LMST and RNG based minimum-energy broadcast protocols in ad hoc networks," *Ad Hoc Networks*, vol. 3, pp. 1–16, Jan. 2005.

[9] L. E. Li and P. Sinha, "Throughput and energy efficiency in topology-controlled multi-hop wireless sensor networks," in *Proceedings of the Second ACM International Conference on Wireless Sensor Networks and Applications*, pp. 132–140, 2003.

[10] M. Bahramgiri, M. Hajiaghayi, and V. S. Mirrokni, "Fault-Tolerant and 3-Dimensional Distributed Topology Control Algorithms in Wireless Multi-hop Networks," *Wireless Networks*, vol. 12, pp. 179–188, Dec. 2005.

[11] N. Li and J. C. Hou, "FLSS: a fault-tolerant topology control algorithm for wireless networks," *Proceedings of the 10th annual international conference on Mobile computing and networking - MobiCom '04*, p. 275, 2004.

[12] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovi, *Mobile Ad Hoc Networking*. Hoboken, NJ, USA: John Wiley & Sons, Inc., June 2004.

[13] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, Mar. 2000.

[14] M. R. Brust, C. H. Ribeiro, D. Turgut, and S. Rothkugel, "LSWTC: A Local Small-World Topology Control Algorithm for Backbone-Assisted Mobile Ad hoc Networks," in *Proceedings of the IEEE Conference on Local Computer Networks (LCN)*, pp. 144–151, 2010.

[15] F. Xue and P. Kumar, "The Number of Neighbors Needed for Connectivity of Wireless Networks," *Wireless Networks*, vol. 10, pp. 169–181, Mar. 2004.

[16] R. Diestel, *Graph Theory (Graduate Texts in Mathematics)*. Springer, 2nd ed., 2006.

[17] A.-L. Barabási, "Scale-free networks: a decade and beyond.," *Science (New York, N.Y.)*, vol. 325, pp. 412–3, July 2009.