

Performance of Wireless Sensor Topologies Inspired by *E. coli* Genetic Networks

Bhanu K. Kamapantula*, Ahmed Abdelzaher*, Preetam Ghosh*, Michael Mayo[†], Ed Perkins[‡] and Sajal K Das[‡]

**Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284.*

Email: kamapantulbk@mymail.vcu.edu, abdelzaher@myemail.vcu.edu, pghosh@vcu.edu

[†]Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS 39180.

Email: Michael.L.Mayo@usace.army.mil, Edward.J.Perkins@usace.army.mil

[‡]Department of Computer Science & Engineering

The University of Texas at Arlington, Arlington, TX 76019. Email: das@uta.edu

Abstract—Wireless Sensor Networks (WSNs) form a critical component in modern computing applications; given their size, ability to process and communicate information, and to sense stimuli, they are a promising part of The Internet of Things. However, they are also plagued by reliability and node failure problems. Here we address these problems by using *E. coli* Gene Regulatory Networks (GRNs)—believed to be robust against signaling disruptions, such as gene failures—to study the transmission properties of randomly-generated WSNs and transmission structures derived from these genetic networks. Selection of sink nodes is crucial to the performance of these networks; here we introduce two sink-node selection techniques: a Motif-based, and a Highest Degree-based approach. Using NS-2 simulations, the performance of both networks is evaluated under varying channel loss models. Successful packet receipts are compared among these networks, which are shown to be higher using GRNs for the communication structure, rather than randomly generated WSNs. This work paves the way for future development of fault-tolerant and robust WSN deployment and routing algorithms.

Keywords—Gene Regulatory Networks, Wireless Sensor Networks, NS-2, Robustness

I. INTRODUCTION

Wireless Sensor Networks (WSNs) gather information from the deployed environment, which is processed and communicated to nearby nodes using a minimum of hardware: transmitters, receivers, a controller, and low-storage memory units. Large-scale WSNs are useful in military applications to monitor enemy targets, in disaster management to deliver critical environmental information, and in agricultural climate-monitoring applications. Despite these capabilities, they do not operate completely free of problems. Significant issues include transmission inconsistencies, channel noise, frequent hardware maintenance, reprogramming difficulties, and node failures. These issues increase the financial and energetic costs associated with more widespread implementation of such networks; reducing these costs requires breakthroughs in automated maintenance and repair, more efficient energy storage and use, and a focus on reducing error and mitigating sensor and packet damage.

There are several parallels between genetic and sensor networks that motivate our discussions. Through a process termed transcription, genes process stimuli in the

form of varying transcription factor levels—proteins responsible for activating/deactivating genes—by producing mRNA molecules directly from the nucleotide sequence of the given gene locus. These transcription products may serve as the activating factors for other genes; so, genes “communicate” with one another by processing incoming signals (varying transcription factor levels) into output signals (the mRNA) used as input for the activation/deactivation of other genes. The network mapping communication between genes in living tissues is termed the gene regulatory network. Similarly, wireless sensor networks (herein WSNs) are nets of communicating sensor motes, whereby hardware is responsible for processing incoming signals (the packets) into out-going messages (packet forwarding). Since living cells are able to adapt to disruptions to genetic “signals” due, in part, to the evolved network topology, we hypothesize that a deployed sensor network architecture based on GRN topologies will adopt similarly “robust” signal-transmission properties.

Although a mapping between gene regulatory networks and wireless sensor networks has already been proposed [9], the signal routing scheme and a simulation platform should be identified as part of our work here. Adaptive routing schemes have been well studied in WSNs, with respect to resilience against channel errors, node failures or congestion [14]. Some current MAC protocols ensure reliable multi-hop data forwarding; however, NS2-based simulations revealed that IEEE 802.15.4 possess very low packet-transmission reliability with power management enabled for energy conservation [3]. Other routing schemes serve as transport protocols in WSNs [20], while experimental validation strategies are enforced to improve the reliability [12], [21]. Validation in the Motes [6] and Sunspot [1] is supported by link-quality estimation schemes that evaluate the stability and reliability of routes [4], [7], and locate alternate routes when some sensors fail [10].

Here we employ NS-2 simulations [11] to evaluate and assess the transmission properties of the wireless sensor networks adapted from the gene regulatory network of *E. coli*. To meet these goals, we first discuss the generation of random and biological network models (Section II). Simulating packet transmission between two distal regions

Table I
ABUNDANCE OF MOTIFS IN *E. coli* GENETIC NETWORKS.

Name	Symbol	Abundance
Feed-forward loop		1860
Uplinked mutual dyad		878
Downlinked mutual dyad		34
Three node feed-back loop		12
Feed-forward and -back loop		10
Feed-back with two mutual dyads		8
Fully connected triad		3

of the network requires the identification of sink nodes, i.e. nodes that receive packets, but do not forward them to others. To this end, three selection methods are proposed in sections III and IV), and discussed in section V. Finally, section Section VI discusses the implications of these results to the direction of future work in the sensor network field.

II. GENERATING MODEL NETWORKS

To account for variations in deployment environments, such as geography, terrain, and foliage, many wireless sensor network models are typically generated randomly and their collective properties studied. So, the sensor network's transmission skeleton is typically modeled as an Erdős-Renyi random graph, which has been well-studied in the wireless sensor and graph-theoretical fields. Here, we term these randomly deployed networks Random Wireless Sensor Networks (RWSNs), which serve as the benchmark network for the present study.

A WSN module in the Python programming language [8] is used to generate these models. Here, two different nodes within the network are chosen at random, and a link is place between them with a probability independent of node degree. Networks of varying size, ranging from 10 to 500 nodes, were generated using this protocol, although networks with 100, 150, 200, 250 and 300 nodes were chosen for demonstration purposes as representing "medium" sized sensor networks.

An entirely different type of network is represented by the gene regulatory network of *E. coli*, wherein the probability to find a node of degree k , $p(k)$, follows a power-law $p(k) \sim k^{-\gamma}$ with $\gamma = 2.1 \pm 0.3$ [19]—although, new research restricts all realizable scale-free networks to $\gamma \geq 2$ [5]. Model GRNs of varying size are generated using the software GeneNetWeaver, originally developed as a platform to test the capability of differing gene network inference

algorithms [17], providing directed sub-networks of user-defined size sampled from the full *E. coli* transcriptional network. For simplicity, we ignore the direction of links between genes resulting from GeneNetWeaver, and no directionality is assigned to the edges of the RWSNs used here. The implications of directed graphs on the WSN transmission characteristics will be considered elsewhere.

All scale-free networks are sparse [5], and many biological networks, including the genetic network of *E. coli*, are scale-free. This is quite different from randomly generated networks, wherein their degree distributions are Poisson-distributed. For a meaningful comparison between RWSN and GRN models, we must ensure that both networks possess an equivalent number of nodes and edges. Hence, edges were removed from WSNs ensuring network connectivity such that they match the number of edges in the corresponding GRNs. Similarly, if the number of edges in a generated RWSN is less than that of a GRN, edges were added between two randomly generated nodes ensuring no self edges exist (i.e. closed loops that begin and end with the same node).

To utilize these generated model networks for packet-transmission scenarios, a source and sink/receptor node must be specified; transmission originates with the source node(s), and terminates at the sink. Due to the biological origin of the scale-free networks provided by GeneNetWeaver, the choice of sink node is not straight-forward. In the next section we discuss two selection methods presented and used here to overcome this problem.

III. SELECTING SINK NODES

Sink node selection is critical to the implementation of NS-2 simulations that model packet-transport across the sensor network. One method to choose such nodes in the scale-free *E. coli* genetic network is to appeal to its topological structure. A power-law degree distribution in the *E. coli* network (independent of link directionality), $p(k) \sim k^{-\gamma}$, ensures that very few nodes have large numbers of nearest neighbors, $\lim_{k \rightarrow \infty} p(k) = 0$, but also guarantees most nodes are of low degree, $\lim_{k \rightarrow 1} p(k) = 1$. Nodes with drastically larger numbers of nearest neighbors than the average are termed "hubs" [2]; their strong coupling to the network implies they have a greater chance of participating

Table II
SINK NODE SELECTION IN THE GRN-DERIVED NETWORKS: MOTIF- AND HIGHEST DEGREE-BASED METHODS.

Network size	HC	HD	MB
20	3, 2, 1	3, {1,2}, 8	3
25	1, 13, 3	1, {13,14}, 3	1
30	14, 1, 8	14, {1,8}, 4	{1,14}
35	2, 29, 22	2, 29, 22	N/A
40	12, 16, 15	12, 16, 15	{12,16}
45	14, 18, 17	{3,14}, 18, 13	{14,18}
50	3, 16, 19	3, 16, 19	3

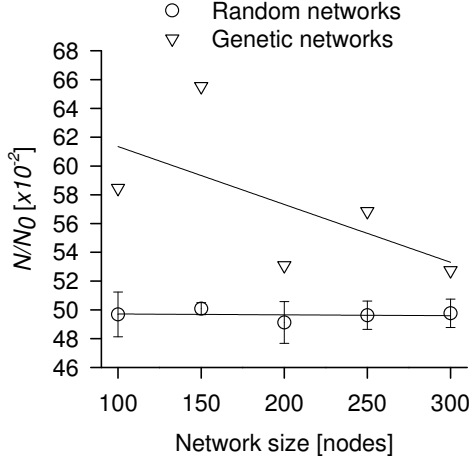


Figure 1. The fraction of sent packets arriving at the sink, N/N_0 , is measured using the Highest Degree (HD) method in GRN-derived sensor networks. Sinks are chosen at random in the random wireless networks (RWSNs).

in paths connecting its opposing sides, making them a natural sink node candidates. This scheme is termed Highest Degree, or HD.

Another scheme selecting sink nodes for genetic networks is based on the observation that they are composed of repeating elementary substructures, termed motifs [16], that serve specific functions. Various motifs have been identified previously in biological networks. Table I illustrates common motifs and their abundance in subnetworks of the largest connected component of *E. coli*'s genetic network. Three node motifs, specifically the most frequently observed feed-forward loop (FFL), may play a critical role establishing robustness of the network, i.e., its ability to preserve function despite external or internal perturbations [13]. The FFL is known to support critical functions, such as a delaying signal response times in target genes, or generating signal pulses [15]. Based on this evidence, we hypothesize that a sensor node most frequently involved in a feed-forward loop of the network's communication structure is a superior choice of sink node for the transmission scenarios studied here. This scheme is termed the motif-based approach, or MB.

Table III
DEGREES OF SINK NODES CHOSEN USING THE HIGHEST DEGREE (HD) METHOD FOR GRN-DERIVED MODEL NETWORKS OF VARYING SIZE.

Network Size	Degree of Selected Node
100	70
150	127
200	97
250	146
300	135

IV. SIMULATING TRANSMISSION IN MODEL NETWORKS

Based on the HD method, sink nodes were selected in the GRN-derived sensor networks if they were the highest degree nodes in the model networks. To implement the MB method for these networks, as described above, we observed all possible three-node motifs in the GRN-derived model network using the MAVisto software package [18]. The number of motifs to which a node belongs, termed motif coverage, is tabulated and the network node with the highest coverage is so chosen as the sink. In addition to the MB method, a more general method in which a node is involved in *any* three-node motif is termed the highest coverage (HC) method.

A. Identification of sink nodes in GRN-derived networks

We compared the HC, HD, and MB methods, and these results are tabulated in Table II. Nodes maintain their identifying labels between all model networks, and nodes identified using the highest motif coverage (HC) method (represented in 2nd column) are boldened. Similarly, nodes identified using the highest degree (HD) method (represented in 3rd column) are boldened. Finally, node labels surrounded by braces, {}, indicate they have either the same degree or the same feed-forward loop coverage using the MB method.

Table IV
SINK NODE SELECTION FOR RANDOM WIRELESS NETWORKS.

Network Size	Number Generated as Sink Node	% of Packets Received	
100	5, 0, 31, 18, 98, 90, 14, 35, 57, 99 , 68	49.8947,	50.8403,
		49.896,	50.5219,
		49.4781,	49.8956,
		50.3132,	50.5219,
		50.4167,	51.2712 ,
150	115, 136 , 80, 8, 32, 19, 116, 57, 145, 13, 149	49.152542	50.282485 ,
		50,	49.812734,
		49.812030,	49.717514,
		50,	49.718574,
		50.187265,	50,
200	58, 36, 50, 1, 83, 55, 88, 23, 47, 11 , 199	50.187265,	51.190476
		49.938195,	50,
		49.752475,	49.938195,
		49.783549,	50,
		49.783549,	50,
250	217, 42, 99, 184, 197, 104 , 166, 85, 73, 22, 249	49.907235,	50.248138 ,
		50.031948	
		49.958506,	49.875724,
		50.124069,	50.124069,
		50,	50.207125 ,
300	111, 67, 158, 283, 71, 34, 17 , 210, 299, 249, 221	50.082644,	49.834437,
		49.792874,	49.917081,
		48.970840	
		50.137849,	49.763593,
		50.019646,	49.921197,
		49.960691,	49.822485,
		50.157604 ,	50.039277,
		49.478748,	49.941060,
		49.960629	

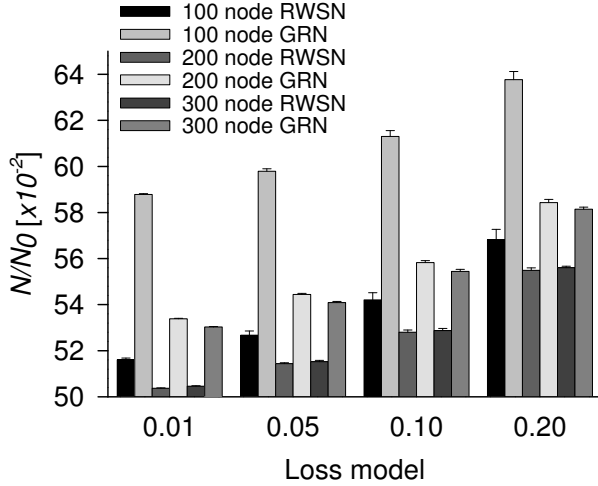


Figure 2. Fraction of successful packet transmissions, N/N_0 , as measured across GRN-derived and random sensor networks of varying size under differing loss models.

In each comparison, the node boldened is considered the sink node for that particular network.

Because we observed that nodes with highest degree also qualify as nodes with highest motif coverage (using both methods of motif coverage, HC and MB), from hereon we disregard the other methods described above, using only those nodes identified as sink nodes according to the HD method of selection. Although this equivalence between nodes identified across all methods is intuitive—i.e., hubs are strongly connected to the rest of the network, participating in the most sub-networks—we should emphasize that further experimentation, either by numerics or analytical methods, are required to establish this observation as fact. The chosen sink nodes based on this method are reported in Table III, and are used to generate the results in the next section.

B. NS-2 implementation

Transmission scenarios are evaluated using NS-2 simulations. A duplex-link is established between the nodes (based on input files) that have edges in RWSNs/GRNs. A packet interval of 5×10^{-4} ms is used. To reduce performance limitations of the network due to bandwidth restrictions, and emphasize topologically-relevant transmission characteristics, we assigned a bandwidth of 1 Mb is to each link. Note, however, that bandwidth may be significantly reduced from this value under experimental conditions. Ten simulations are executed for every category of observation using a flooding routing protocol. Information is transmitted at periodic intervals and successful packet receipts are recorded at the sink node. A packet-loss model is used to evaluate the performance of each network of fixed size, in which packet-loss of 1% to 20% is considered.

C. Identification of sink nodes in RWSNs

We first report the packet receipt characteristics of GRN derived networks and RWSNs using the same criterion (i.e., the HD method) for selecting the sink node. Figure 1 illustrates that GRN derived networks are more robust than RWSNs with this criteria under varying network size.

For performance assessment of RWSNs, sink nodes are chosen at random because their topological structure cannot be used with the sink-selection methods outlined above for GRNs. Here, a random number is generated from 0 to n , where $n = 100, 150, 200, 250$, or 300 is the number of nodes in the model network. Ten random numbers were generated for each model network; given that each node of the network is labeled $i = 1, \dots, n$, these random numbers select nodes that are each considered to be a sink. Next, separate simulations were conducted with each of these choices for sink nodes and the packet receipt rates were observed. The results have been reported in Table IV. Note that sink nodes labeled 99, 136, 11, 104, 17 perform best (highest packet receipt rate) for the 100, 150, 200, 250 and 300 node networks respectively. However, nodes 68, 149, 199, 249, 299 have the highest degree for these random networks. These nodes (with highest degree) and the corresponding fraction of successful packet receipts are italicized; whereas, the nodes with the best performance are boldened. So, nodes with highest degree are not necessarily the best choices for sink nodes in RWSNs. The best performing sink nodes among the 10 options considered (the results in bold) used to exemplify signal transmission across random networks contrasted against GRN-derived networks, as described below.

V. RESULTS

A. GRN-derived model networks improve transmission reliability

Using NS-2 simulations, we find the ratio, N/N_0 , of packets received at the aforementioned sink nodes, N , to the total number of packets sent from the source, N_0 , is consistently higher in the sensor networks derived from the *E. coli* transcriptional network than the random wireless networks modeled here, as illustrated by Fig. 1. This result holds for all simulated packet loss scenarios, including those with (i) no packet loss model implemented (i.e., Fig. 1); and (ii) loss models ranging from 0.01 to 0.2, wherein 1% to 20% of transmitted packets are dropped during transmission to adjacent nodes, and illustrated in Fig. 2.

Moreover, Fig. 3 illustrates the fraction of packets dropped, $1 - N/N_0$, varies only weakly across an order of magnitude of the packet-loss (1-20%). Interestingly, the different between random and genetic networks is almost constant with respect to dropped packets, as demonstrated by Fig. 3, across the entire span of loss models studied here, with random networks performing least favorably. Note the general trend in both network types to decrease the fraction

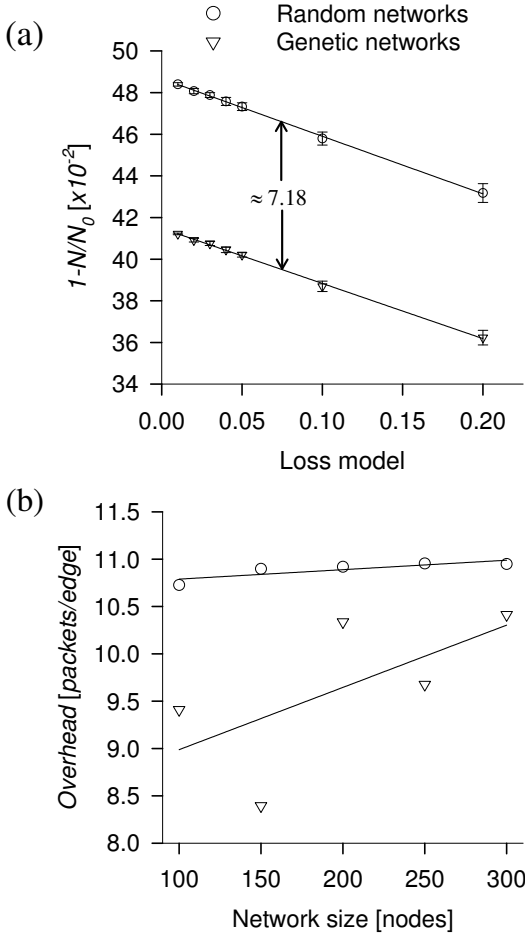


Figure 3. (a) Number of packets lost across both random and GRN-derived sensor networks, and (b) average overhead.

of lost packets (or increase the number of successful transits) cannot decrease monotonically with increasing packet loss (i.e., $\lim_{loss \rightarrow 1} 1 - N/N_0 \neq 0$).

B. Average overhead is higher in random sensor networks

The average overhead is the averaged ratio of packets sent across the network to its total edges. We measure this by computing the average number of packets sent across the 10 simulations, and then dividing this number by the total number of network edges. Since the number of edges in both GRN- and random-derived sensor networks were equal, the overhead of both GRN- and random-derived networks can be directly compared. As demonstrated by Fig. 3, RWSNs provide a higher average overhead than their GRN-derived model network counterparts, although an increase in the number of packets sent in the random network is almost exactly matched by an increase in its number of edges. This should be contrasted with the GRNs, in which the overhead increases with increasing network size. This is likely a result of the network topology—an added node is not very likely

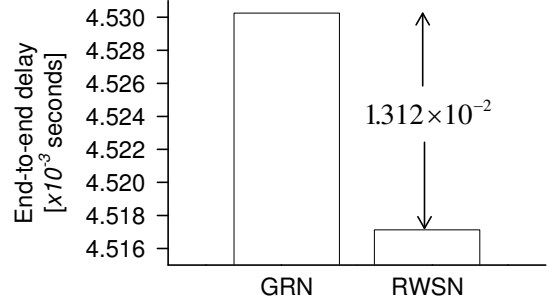


Figure 4. Average end-to-end delay in GRN-derived and random sensor networks.

to contribute many additional edges, so the number of total edges grows slower than the number of sent packets.

C. End-to-end delay increases for GRN-derived sensor networks

The end-to-end delay, defined as the elapsed time for a packet to successfully reach the sink, was measured in all simulations for both GRN- and random-derived sensor networks, and an arithmetic average of these values was used to assess performance in both types of network topology considered here. Packets dropped according to the loss models were not considered in the averaging procedure. Figure 4 shows that the end-end delay in the GRN-derived networks is slightly higher than the RWSNs, suggesting that GRN-derived sensor networks may host improved transmission efficiency (i.e. low packet-loss rates), but are prone to higher signal transmission delays. This feature may disappear upon further analysis using other scale-free networks.

VI. CONCLUSION

When the *E. coli* transcriptional network is taken as the result of a sensor “deployment” scheme, transmission scenarios on these networks under variable packet-loss conditions result in smaller packet-loss rates, but also exhibit greater transmission delays than a purely random deployment. On the one hand, it appears that packets more reliably find their way across the network; on the other hand, it takes them longer to do so. This effect arises directly from the topology itself, because the randomly-generated comparison networks were assigned an equal number of links, with all other conditions equal. Scale-free networks are necessarily sparse, possessing few hubs [5]; so, short direct paths linking

Table V
FEEDFORWARD LOOP MOTIF COUNT IN GRN-DERIVED SENSOR NETWORKS, RWSNs.

Size	Motif abundance (GRNs)	Motif abundance (RWSNs)
100	390	138
150	192	80
200	1424	880
250	330	202
300	2200	904

opposing sides of the network might also be rare, leading to longer transits.

Transmission time across the network depends on the proper selection of a sink node, of which we proposed three methods: a highest degree (HD), a feed-forward loop motif-based (MB), and a generalized motif-based method that emphasizes the role a node plays in each motif of the network, termed highest coverage (HC). Interestingly, all three methods identify the *same node* as the best candidate for sink. This result, however, is intuitive: the hubs of the GRN-derived sensor networks are strongly connected to the rest of the network, giving them the highest chance of participating in the most sub-networks, or motifs.

Although we have demonstrated that GRN-derived sensor networks out-perform those of randomly-generated ones with respect to packet-loss rates, but at the expense of longer transmission delays, it remains undetermined how the directionality of GRNs should affect the transmission scenarios in our simulations presented here. Other work [9] suggests that directionality of the GRN links have, for example, implications for packet collision events, and future work will incorporate these directional links into the NS-2 simulations.

Finally, biological networks, such as the *E. coli* transcriptional network considered here, evolve in response to environmental pressures, yet their topological structures resist homogenization from random mutations of gene bases that occur during every cell division. Transferring these biological principles of network growth into the engineering domain may provide new insights into the design of novel transmission networks that preserve function when exposed to realistic factors such as noise, damage, or targeted attacks.

ACKNOWLEDGEMENTS

This work was supported by grant number NSF-1143737, and the US Army's Environmental Quality and Installations 6.1 basic research program. The Chief of Engineers approved his material for publication.

REFERENCES

- [1] Link quality routing protocol (lqrp), 2009.
- [2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, January 2002.
- [3] G. Anastasi, M. Di Francesco, M. Conti, and V. Neri. Reliability and energy efficiency in multi-hop ieee 802.15.4/zigbee wireless sensor networks. In *Proc. of the 15th IEEE Symposium on Computers and Communications (ISCC'10)*, pages 336–341, Riccione, Italy, June 2010.
- [4] D. Couto, D. Aguayo, J. Bicket, and R. Morris. A high-throughput path metric for multi-hop wireless routing. *Wireless Networks*, 11:419–434, 2005.
- [5] C.I. Del Genio, T. Gross, and K.E. Bassler. All scale-free networks are sparse. *Phys. Rev. Lett.*, 107:178701, Oct 2011.
- [6] R. Fonseca. The multihopLQI protocol, 2009.
- [7] R. Fonseca, O. Gnawali, K. Jamieson, and P. Levis. Four bit wireless link estimation. In *Proceedings of the 6th Workshop on Hot Topics in Networks (HotNets)*, Atlanta, GA, November 2007.
- [8] Python Software Foundation. Core python programming. www.python.org, 1991.
- [9] P. Ghosh, M. Mayo, V. Chaitankar, T. Habib, E. Perkins, and S.K. Das. Principles of genomic robustness inspire fault-tolerant wsn topologies: A network science based case study. In *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2011 IEEE International Conference on, pages 160–65, March 2011.
- [10] O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis. Collection tree protocol. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 1–14, Berkeley, CA, November 2009.
- [11] Information Sciences Institute. Ns-2. <http://isi.edu/nsnam/ns>.
- [12] S. Kim, R. Fonseca, and D. Culler. Reliable transfer on wireless sensor networks. In *Proc. of IEEE SECON 2004*, pages 449–459, Santa Clara, CA, October 2004.
- [13] H. Kitano. Towards a theory of biological robustness. *Mol. Sys. Biol.*, 3:137–44, 2007.
- [14] J. Li and P. Mohapatra. Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. *Pervasive Mobile Computing*, 3:233–254, 2007.
- [15] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 100(21):11980–11985, 2003.
- [16] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [17] T. Schaffter, D. Marbach, and D. Floreano. Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(5):2263–70, 2011 Aug 15.
- [18] F. Schreiber and H. Schwöbbermeyer. MAVisto: a tool for exploration of network motifs. *Bioinformatics*, 21:3572–74, 2005.
- [19] A. Vázquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z. N. Oltvai, and A.-L. Barabási. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proc. Natl. Acad. Sci. USA*, 101(52):17940–17945, 2004.
- [20] C. Wang, K. Sohraby, B. Li, M. Daneshmand, and Y. Hu. A survey of transport protocols for wireless sensor networks. *IEEE Network*, 20:34–40, 2006.
- [21] A. Woo, T. Tong, and D. Culler. Taming the underlying challenges of reliable multihop routing in sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 14–27, Los Angeles, CA, November 2003.