Empirical Prediction of Packet Transmission Efficiency in Bio-Inspired Wireless Sensor Networks

Ahmed F. Abdelzaher*, Bhanu K. Kamapantula*, Preetam Ghosh* and Sajal K. Das†
*Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia 23284
Email: abdelzaheraf@vcu.edu, kamapantulbk@vcu.edu, pghosh@vcu.edu
†Department of Computer Science and Engineering, The University of Texas at Arlington,
Arlington, Texas 76019. Email: das@uta.edu

***Abstract*—Biological networks (specifically, genetic regulatory networks) exhibit an optimized sparse topology and are known to be robust to various external perturbations. We have earlier utilized such networks, particularly, the gene regulatory network of E. coli, for constructing smart communication structures in bio-inspired Wireless Sensor Networks (WSNs) having high packet transmission efficiency. In this paper, we present machine learning approaches to relate the graph topology based characteristics of such bio-inspired WSNs to their network-level robustness in terms of average packet transmission efficiency. In particular, we generate a support vector regression model using the graph metric features as input data. The model predicts the percentage of packets received by the highest degree sink node and a theoretical estimate for the overall network robustness.**

***Index Terms*—wireless sensor network; genetic regulatory network; feedforward loop; bi-fan; support vector regression**

I. Introduction

In military applications, where monitoring enemy targets is the primary objective, large-scale WSNs are deployed to identify foreign object movements. WSNs are also capable of capturing climate data: temperature, sunlight, wind speed and humidity, which is crucial for agricultural monitoring and are also used widely in many other application areas. However, with the increase of their emplacement structure, issues like node failures and channel noise are harder to detect and minimize in real-time operations. These issues increase the financial and energetic costs of maintaining such networks as well as increasing the delay for their reparation.

To avoid such problems, advances must be made in increasing WSN efficiency measures such as ensuring the relay of packets from remote sensors to sink nodes with minimal packet loss. While much work has investigated network transmission issues such as multi-path interference, line-of-sight obstruction, channel inhomogeneity[1], node failures and congestion [2], we have illustrated in our earlier works ([3], [4]) that particular natural network topologies (i.e., gene regulatory networks) operating as smart routing topologies in bio-inspired WSNs, demonstrate more efficient packet transmission rates than that of randomly deployed sensor nets generated using the popular Erdös-Rényi model [5]. Particularly, sub-graphs extracted from the genetic regulatory network (GRN) of the bacterium Escherichia coli (E. coli), that is broadly used as a model prokaryotic framework in the life sciences, were used to construct the routing topologies of bio-inspired WSNs; comprehensive NS-2 [13] simulations demonstrated that such bio-inspired WSNs exhibit higher packet receipt rates at the sinks under varying channel noise and node failure scenarios.

Similarities between GRNs and WSNs can be explained through the biological function of transcription, where genes process signals from neighboring nodes in the form of transcription factors and forward them to downstream nodes of a GRN. Therefore genes communicate by sending signals much the same way as in a WSN, and such genetic signals withstand various node/edge disruption scenarios arising inside cells due to both external and internal perturbations – a feature termed biological robustness ([6], [7]), which has been primarily attributed to the optimized GRN topologies responsible for routing genetic signals. The process is similar to WSNs where sensors receive packets from neighbors based on user-defined routing protocols with packet forwarding instructions to other destination points (sinks). This motivated our earlier works in activating GRN-derived routing topologies [3], [4] in randomly deployed bio-inspired WSNs. The goal of this paper is to assess the specific graph theoretic features of such GRN topologies and understand their contributions toward achieving network robustness, measured in terms of packet transmission efficiency using extensive NS-2 [13] simulations.

Internal disturbances in a communication network (WSN in our case) can occur due to link failure, link congestion or node failure [8]. Consequently researchers have yet to establish a unified measure for robustness, however, many works suggest that robustness should be quantified by the systems behavior post a particular disturbance. For example, [9] suggests that the network efficiency, an inverse function of the average shortest path length [10], should be analyzed for single node deletions. Other work includes the assessment of connection failures [11] and fractional inactivation [12]. In this paper, we focus on the evaluation of the topological features of the bio-inspired WSN which makes the system perform its primary function reliably, that of packet delivery to destination sink nodes.

In this regard, our contributions in this paper can be summarized as follows. We present one measure of robustness in bio-inspired WSNs (that adopt GRN-based routing topologies) as the ability for each node in the network to deliver packets to the sink node with minimal packet loss. We generated several bio-inspired WSN topologies and used comprehensive NS-2 [13] simulations to correlate their topological features to the overall network transmission efficiency. Finally, we propose a support vector regression model using well-grounded optimization techniques to generate the optimal hyper plane, to predict the packet transmission efficiency with minimal

978-1-4673-5119-5/12/$31.00

validation error, using these network features as inputs.

The paper is organized as follows: we first discuss the different robustness measures used in previous works (Section II) followed by the generation of the different sample data extracting GRNs out of E. coli and conducting NS-2 simulations to quantify their robustness (Section III); we then present the different parameters that can affect the outputs of our model (Section IV); comparison of the model prediction values with the true output values of the NS-2 simulations are shown in Section V; finally, we discuss the significance and implications of this study to the sensor network domain in Section VI.

II. Robustness Measures

Here, we discuss some graph theoretic metrics that have been traditionally used earlier to assess network robustness.

A. Network Diameter

The network diameter [14] (D) is the longest shortest path d_{ij} between any pair of nodes i and j

$$\mathrm{D} = Max(d_{ij}). \tag{1}$$

A robust network is more likely to maintain its diameter post attacks. Attacks on edges or nodes increase the diameter by deleting components that constitute part of the shortest paths between pairs of nodes. In some cases, deleting central nodes or edges can result in splitting the network into two separate components which makes the diameter infinite.

B. Average Shortest Path (ASP)

The Average Shortest Path represents the average of all the shortest paths d_{ij} between all pairs of nodes,

$$\mathrm{ASP} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} \tag{2}$$

were n is the number of nodes in the network. This property represents the mean value of the networks characteristic for two nodes to communicate with each other. Since attacks increase d_{ij}, network disruptions causing edge deletion increase the ASP. The larger the ASP, the greater the chance for transmitted packets to be discarded due to end to end delays.

C. Network Efficiency (NE)

The network efficiency can be determined by a function of the inverse sum of the shortest distances [15],

$$\mathrm{NE} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1;\neq i}^{n} \frac{1}{d_{ij}}. \tag{3}$$

When nodes and edges are deleted, the network efficiency is expected to decrease due to the increase in d_{ij}.

III. Generating Sample Data

We examine the dependency of the network robustness, determined by the efficiency of packet transmission (herein the Robustness Scores), on the topological metrics of the bio-inspired WSN. The main network skeleton which embodies all sub-networks extracted in this study is the GRN of E. coli, wherein the probability to locate a node of degree k, $p(k)$, obeys a power-law $p(k)\ k^{-\gamma}$, wherein $\gamma = 2.1 \pm 0.3$ [16], although new research restricts all realizable scale-free networks to $\gamma \geq 2$ [17]. A power-law degree distribution in the E coli GRN guarantees that the majority of nodes are of low degree (low number of nearest neighbors) $\lim_{k\to 1} p(k) = 1$, while preserving the biological properties that few nodes exhibit relatively very high degree, $\lim_{k\to\infty} p(k) = 0$, referred to as "hubs" [18].

GRNs of varying sizes, $50 \leq n \leq 1477$, were rendered for this study, out of which; 289 were used to solve our learning problem, 58 were used as test data, and the others for controlled experiments that will be discussed elsewhere. Random GRN extraction was accomplished using the GeneNetWeaver tool [19], originally acknowledged as an application to evaluate network inference algorithms, rendering directed bio-inspired sub-graphs of user defined network magnitudes. For employing NS-2 simulations, we ignore the directionality of the interconnections and adopt the Highest Degree method for selecting the appropriate sink nodes as discussed in [4]. In order to avoid computational overheads, we deleted self-loops that were associated with regulatory gene nodes capable of self up-regulation/down-regulation; note that such self-loops are crucial in the overall functionality of GRNs however, they do not contribute to WSN routing performance in any way. The generated graphs are composed of a set of source nodes S and destination nodes d, such that $(S \cap d) \subseteq N$, were N denotes the set of nodes of the sub-graphs. We denote a binary relationship "Connects" as;

$$R_c = \{< a, b > | \exists\ a\, direct\, link\, between\, a\, and\, b\}, \tag{4}$$

wherein $a \in S$ and $b \in d$. Hence, for $< a, b > \in R_c$, the value of the adjacency matrix $A_{ij} = 1$ and $A_{ji} = 1$ should hold true for a connection, and 0 otherwise. Based on source-destination input files of the different substrate bio-inspired networks, a bidirectional-link is established between connected nodes in the GRNs. Packet receipt rates are recorded at the sink nodes for packet-loss models of 2%, 4%, 20%, 40%, 50%, and 60%. Other specifications of the simulation set-up include: assigning link bandwidths of 1 Mb, Queue length of 1, a packet interval of $5 \times 10^{-4} ms$ and the packet receipt rates at the sink are averaged across ten different simulation runs in each case.

Based on the observation that the entire E. coli GRN is composed of 23 connected components, together forming 1565 nodes and 3758 interconnections, we confine our sub-networks extraction to the GRN's largest connected component (LC), composed of 1477 nodes and 3671 interconnections. From here on, we denote the robustness score associated with LC, $S_{LC} = percentage\, of\, packets\, received\, in\, the\, NS - 2\ simulations\, with\, the\, different\, loss\, models.$

IV. Features of the Model

Our premise is that the fundamental structural properties of bio-inspired WSNs influence the overall source-sink packet transmission efficiency. In the next sub-section, 5 GRNs of size $n = 100$ and different total number of links are used for our demonstration of the effects of different sink node selections, while in the next two sub-sections ten "randomized networks" are derived from each GRN, together adding up to 40 density-controlled networks. Such randomized networks

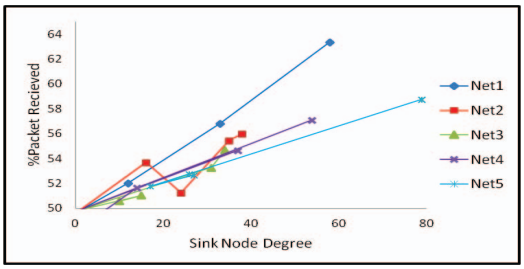

Fig. 1. Scatter plot of percentage packets received vs. the degrees of the sink nodes selected for networks $1-5$.

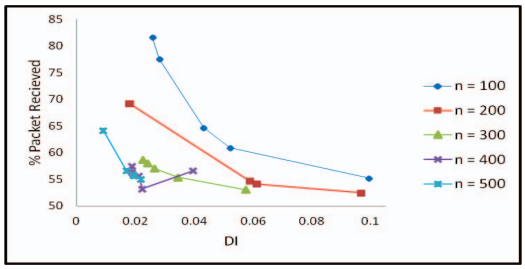

Fig. 2. Scatter plot of percentage packets received vs. the degree index for different networks of sizes $100-500$

are derived from a GRN, using the switch based method from [20], that preserves the individual properties (in/out degrees) of the nodes, but end up having not the same original neighbors as in the GRN, thereby affecting the final overall structure of the networks. Furthermore, all sample features have been determined computationally; their concepts are described in the following subsections.

A. The Degree Index

We define the Degree Index of a directed network, as the ratio of the Average Nodal Degree Eq. (5) to the highest degree Eq. (6) as follows;

$$K_{avg} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} \tag{5}$$

$$K_{max} = Max\left[\sum_{m=1}^{n} A_{m1}, \sum_{m=1}^{n} A_{m2}, ..., \sum_{m=1}^{n} A_{mn}\right]. \tag{6}$$

As shown in Figure 1, the packet transmission performance of the different networks depend heavily upon the degree of the sink node selected; specifically Hub nodes, having relatively larger degrees than the average, usually exhibit better packet receipt rates when selected as sink nodes. Furthermore, note that the average packet receipt rates does not monotonically increase with the selection of sinks having higher degrees (as shown in networks 5 and 2) although, the highest degree node almost always gives best performance.

Based on this observation, we hypothesize that the Degree Index is a suitable metric that can characterize S_{LC}. We denote the Degree Index as;

$$\text{DI} = \frac{K_{avg}}{K_{max}}, \tag{7}$$

by which we can evaluate the relative importance of choosing one node as sink over the other in the entire network. Therefore as $DI \rightarrow 1$, we can deduce that the network becomes more tightly connected, such that the nodes would gain less significance in comparison to one another to act as sinks (i.e. the network does not rely heavily on fewer potential sink nodes). Furthermore, as illustrated in Figure 2, as DI increases for any particular sub-graph, the percentage of packets received decrease making the benefits of sink node selection (having highest degree) largely redundant.

B. The Network Density

The Network Density is traditionally a measure of the territorial occupation of a communication network, calculated as the ratio of the sum of the edge lengths to the surface area occupied by the network grid [21]. Since our simulations do not account for edge weights, we simplistically consider the nodes to be equidistant, having unit lengths of one. Our measure accounts for how many links occupy the adjacency matrix (A_{ij}) grid and determine the Network Density as follows:

$$\text{ND} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} \tag{8}$$

where, ND ranges between $(n-1)/n^2$ for a Star configuration (with a single hub, provided every other node is solely connected to the hub with one edge), and $1-(1/n)$ for a fully connected sub-graph excluding the self-loops.

We observe that tightly connected graphs does not necessarily have higher packet receipt rates. In Table I, we present the average packet receipt rates of 40 different randomized graphs to prove our hypothesis; 10 of which were drawn out of 4 different 100 node GRNs with different values of network densities. Note that the performance of the networks having the same ND are comparable because the properties of the sink nodes selected after the randomizations are still preserved in the sub-graphs.

TABLE I
AVERAGE PACKET RECEIPT RATES FOR NETWORKS HAVING DIFFERENT DENSITIES

Networks	Density ($\times 10^{-3}$)	% Packets Recieved
1-10	5.4	33.33
11-20	17.9	50.00
21-30	19	52.17
31-40	25	50.00

C. The Motif Index

Before dwelling into the definition of the Motif Index, we must acknowledge that these repetitive "motif" substructures have significant contributions to WSN performance and functionality (as we have shown earlier in [4] and separately at [3]), as well as affecting robustness in biological networks [6]. Although various types of motifs have been identified previously in biological networks, we consider the "most significant" ones for our model: the Feed-Forward Loop (FFL) and the Bi-fan (BF) [20]. These two motifs significantly outnumber similar sub-structures when mined from the GRN of E. coli in comparison to other randomized networks and hence they are believed to have a significance in biological networks in general. Furthermore, FFLs are notable for their ability to deliver vital functions such as delay responses times in genes, irreversible speed up, or create pulses [23]. Similarly

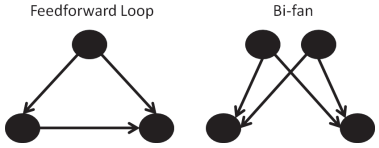

Fig. 3. The FFL and BF motif structures.

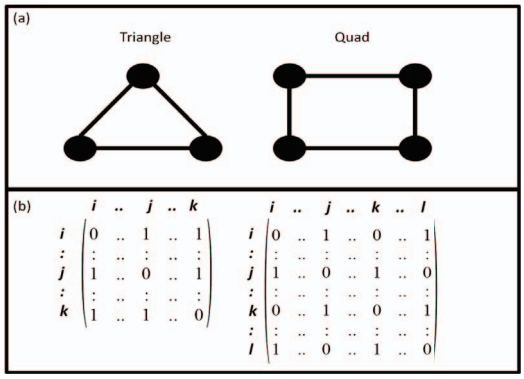

Fig. 4. (a) Actual (undirected) structures considered for the Motif Index. (b) The adjacency matrix of the structures considered for the Motif Index

BFs are the building blocks of dense overlapping regulons, which are considered to be the backbone for GRNs, sharing global functions such as: stress response, nutrient metabolism, or bio-synthesis of key classes of cellular components [24].

Figure 3 shows the FFL and BF structures as mentioned above, for which we hypothesize that their relative abundance in the network should make an important feature to consider in our regression model. In our data sets, we observed that the FFL counts accompany larger BF counts for any bio-inspired network considered. In order to account for both, we converted their counts into a normalized ratio of one motif abundance to the other, which also reduces our features' dimensionality by one parameter. Since we ignore directions in our simulated networks, we apply the same to the Motif Index and consider the occurrences of quadrilaterals and triangles (corresponding to BFs and FFLs respectively) in the networks as shown in Figure 4(a).

To calculate the motif counts, we consider every pattern: $N_T = (< i,j > \cap < j,k > \cap < i,k >) \in R_c$ and $N_Q = (< i,j > \cap < i,l > \cap < k,j > \cap < k,l >) \in R_c$ for every i,j,k and $l < n$, for the Triangle and Quadrilateral structures respectively. We also motivate our hypothesis based on the fact that such patterns are ideal for considering cluster formations based on the number of nodes that participate in forming them and the ones that do not [25], [26]. For an undirected non weighted network stored in an adjacency matrix A, these counts can be determined mathematically as;

$$N_T = \frac{1}{6}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}[A_{ij} \cap A_{ik} \cap A_{jk}] \tag{9}$$

$$N_Q = \frac{1}{8}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}[A_{ij} \cap A_{il} \cap A_{jk} \cap A_{kl}] \tag{10}$$

An illustration of the motif patterns in the adjacency matrix is given in Figure 4(b). Note that in Eq. (9), the occurrence of triangles is divided by 6 to avoid redundancy caused by the symmetry of the triangle pattern, similarly with Eq. (10) we divide by 8. We can hence calculate the Motif Index as:

$$\text{MI} = \frac{N_Q}{N_T + N_Q} \tag{11}$$

which will account for effects of the motif ratios to the packet transmission efficiency of the networks considered.

D. The Sink Coverage

The sink coverage measures the percentage of nodes that have a direct link to the sink node, using K_{max} of Eq. (6),

$$\text{SC} = \frac{K_{max}}{n}. \tag{12}$$

When node a tries to send packets to node b through node c lying along the path d_{ab}, packets are queued at c before they get forwarded to b, which in return can be dropped if packets exceed the queue length at c. However if c did not exist in the d_{ab} path, packets will not be discarded due to multi-hops; SC is a feature that captures such scenarios.

E. The Hub Nodes Density

The density of the hub nodes measures the territorial occupation of the adjacency matrix grid by the higher degree nodes as a fraction of the total number of edges. We hypothesize that the hub nodes are the hot spot traffic management zones as they have more packets hopping through them. We can determine this quantity as follows:

$$\text{HND} = \frac{1}{l_{total}}\sum_{i}^{n_h}\left[2\sum_{j=g_i}^{n_g}A_{ij} + \sum_{j=h_i}^{n_h}A_{ij}\right] \tag{13}$$

were n_h is the number of hub nodes, g_i is the index of nodes outside the set of hub nodes, h_i is the index of the hub nodes and l_{total} is the total number of edges in the network.

F. Outliers

After generating the datasets, we observed that in very few cases the packet transmission rates were exceptionally high, i.e., $S_{LC} > 90\%$ for smaller networks, $n \leq 80$. After visually analyzing these networks using GeneNetweaver, we realized that these minor cases exhibit star topologies. Since we choose the highest degree node to be the sink (the central node), packets are directly forwarded from its neighbors without suffering from the possibility of being dropped due to multi-hop queues and transmission errors. However these cases never occur in GRNs of large sizes and need not be considered as a part of the training dataset for our model. For our subsequent experiments for data generation, only one network of size $n = 50$ was discarded wherein the average packet receipt rates was more than seven times that of its standard deviation. Therefore, for simplicity we restricted our model to predict S_{LC} for bio-inspired WSNs that do not posses a perfect star topology. This was in fact achieved by adding the following restriction in generating GRNs:

$$MI \geq 0, \tag{14}$$

This guarantees that $N_Q > 0$ (or $N_T > 0$), resulting in a non-star topology.

V. The Model

A. Training

After generating the packet transmission rates using NS-2 simulations, the values of the loss models were stored in a y vector of $s \times 1$ dimension for s subgraphs. If X is a $s \times 5$ matrix, then we can denote $X^{(s)}$ as the sth row of X, which holds the values of the training data (or topological features) of the sth training sample (or network). X and y are used as input parameters for generating the model's weights vector (w) and the bias (b) using LibSVM[27]. Once, the model is generated, any approximation O_i can be predicted with any input X_i , using $O_i = X_i w + b$, and the influences of the different features on the models output are ranked by analyzing the magnitudes of their corresponding weights. Other parameters include setting the svm type to an epsilon-SVR with a polynomial kernel. The training was repeated several times with varying values of the hyper parameters; C (the quadratic optimization constraint), γ (the order of the polynomial kernel) and ϵ ($\pm$ error range), and the parameters which yields the minimum error were used for the prediction of the test data. Furthermore, as we observe that the values recorded from the different loss models are exactly the same for a queue length of 1, there was no need to account for any difference in the loss models. The model set-up and results with varying queue lengths in the NS-2 simulations will be discussed elsewhere.

B. Validation

We first checked the parameters that gave us the least 10 k-fold cross validation error, for which training is assumed 10 times at random on 90% of the sample space and prediction is performed on the remaining 10% during each iteration. The error percentage of the difference between the desired average packet receipt rates and the predicted values is calculated by,

$$error = 100 \times \frac{\|y - y_{apx}\|}{\|y\|}, \tag{15}$$

were y_{apx} represents the model's approximated (predicted) values. The error is accumulated after each iteration and then averaged across the 10 folds to give us the cross validation error. Using the same parameters, training is performed across the whole sample space and prediction was again performed on the whole data set to calculate the data set error and similarly with the test data set.

Figure 6 shows the errors for the model generated using all 5 features (DI, ND, MI, SC and HND) versus the errors for the models generated using a combination of 1, 2, 3 and 4 features. In Figure 6 we display the errors which correspond to the combination of features that yield the least cross validation error. Note that the use of all features would give the least summed error of all three (cross, test and total error). In general, we have observed (results not shown) that the relative importance of the features used in our model can be ranked in descending order as follows: SC, MI, ND, HND and DI, as shown in Figure 5 (in this we divide the weights by the maximum, which is SC).

The plots of the models prediction dependency on the tolerance parameter ϵ is displayed in Figure 7 and 8. A

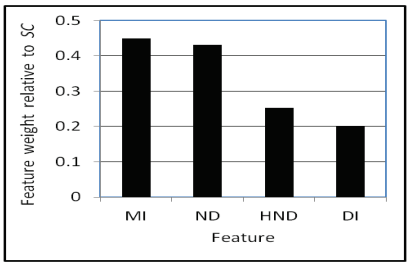

Fig. 5. Bar chart showing the relative importance of the features.

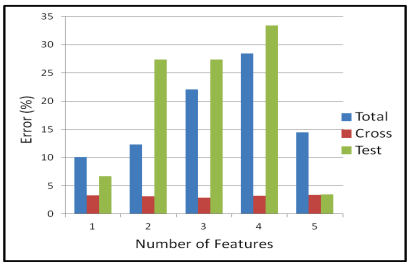

Fig. 6. Bar chart showing the Total data, Cross validation and Test data errors vs. the number of features

tolerance value of $\epsilon = 1.0$ gives us the best correspondence between the actual and predicted network robustness values for both the sample and test datasets.

The correlation between the different traditional robustness measures (diameter, ASP and NE) with the predicted packet transmission efficiency from our model is displayed in Figure 9 for different networks considered. It can be observed that both diameter and ASP have an inverse relationship to the average packet receipt rates while NE is directly proportional to the packet transmission efficiency. As both the diameter and ASP metrics deal with the shortest paths in the network, it is expected that larger shortest paths would incur higher packet drops (due to multi-hop communication) and vice versa resulting in the inverse relationship illustrated here. However, NE considers the inverse of the shortest paths and is expected to be directly proportional to our robustness measure of packet transmission efficiency.

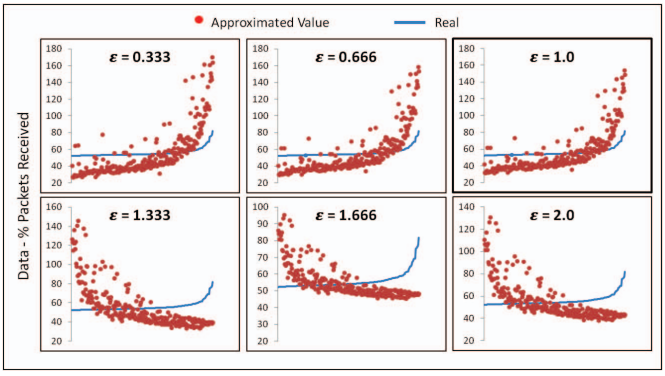

Fig. 7. Plots of the real and approximated values for the sample space data % packets received using different tolerance ranges ϵ

VI. Conclusion

In this paper, we present a standard robustness metric of average packet receival rates at the sink in a bio-inspired wireless

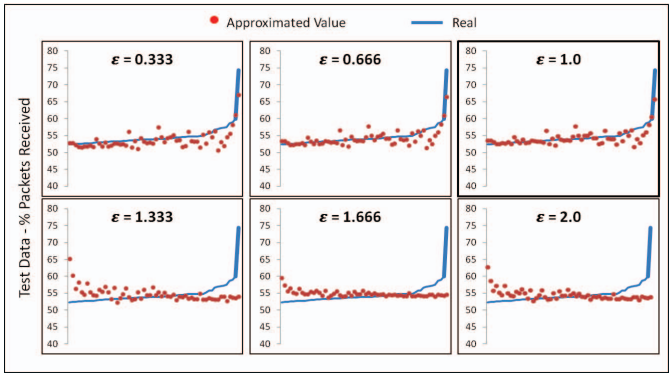

Fig. 8. Plots of the real and approximated values for the test data % packets received using different tolerance ranges ϵ

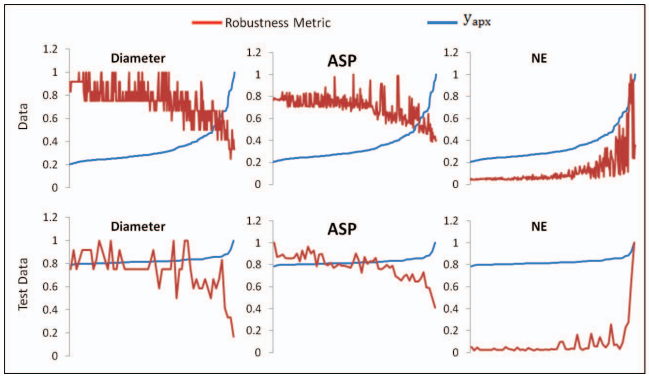

Fig. 9. Plots of the normalized robustness measures and the normalized approximated values for % packets received

sensor network. We compare our proposed robustness metric to traditional metrics for network robustness involving graph diameter, average shortest paths and network efficiency; the packet transmission efficiency metric is inversely proportional to the graph diameter and average shortest path in the network and directly proportional to the network efficiency.

We simulated the bio-inspired WSNs using NS-2 to generate the test and sample dataset for various model networks and then proposed a support vector machine model to correlate the average packet receival rates to the following five topological features of the such bioinspired WSNs: degree index, network density, motif index, sink coverage and hub node density. Extensive experiments of both sample and test dataset illustrate that using all five features yields the minimum error over using any other combination of a selected set of these features. It was also observed that the features can be ranked as follows in order of their importance in providing packet transmission robustness in WSNs: sink node coverage, motif index, network density, the hub nodes density and degree index.

Our work makes an important contribution to the evaluation of packet transmission efficiency in any network considering its graph topology and provides a theoretical model to predict network robustness based on the five identified features.

REFERENCES

[1] C. Savarese, J.M. Rabaey, K. Langendoen. Robust Positioning Algorithms for Distributed Ad-Hoc Wireless Sensor Networks. In *ATEC '02 Proceedings of the General Track of the annual conference on USENIX Annual Technical Conference*, pp 317–327, 2011.

[2] J. Li, and P. Mohapatra. Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. In *Pervasive and Mobile Computing*, pages 233–254, 2007.

[3] 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*, 2011, pp. 160-65.

[4] B.k. Kamapantula, A. Abdelzaher, P. Ghosh, M. Mayo, E. Perkins, and S.K. Das. Performance of wireless sensor topologies inspired by E. coli genetic networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pp 302–307, 2012.

[5] P. Erdös , A. Rényi. On the Evolution of Random Graphs. *The Mathematical Institue of the Hungarian Academy of Sciences*, 1960.

[6] H. Kitano. Towards a theory of biological robustness. *Mol Syst Biol.*, 2007.

[7] H. Kitano. Biological Robustness. *Nature Reiews Genetics 4*, 826-837, 2004.

[8] S. Eum, S. Arakawa, M. Murata. Toward bio-inspired network robustness - Step 1. Modularity. In *Bio-Inspired Models of Network, Information and Computing Systems, 2007. Bionetics 2007. 2nd*, pages 84–87, 2007.

[9] T. Feyessa, M. Bikdash. Measuring nodal contribution to global network robustness. In *Southeastcon, 2011 Proceedings of IEEE*, pp 131–135, 2011.

[10] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda Error and attack tolerance of complex networks. In *Physica A: Statistical Mechanics and its Applications*, Vol. 340, 388–394, 2004.

[11] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin. Resilience of the Internet to random breakdowns. In *Phys. Rev. Lett 85*, 2000.

[12] V. Ágoston, P. Csermely, S. Pongor. Multiple weak hits confuse complex systems: A transcriptional regulatory network as an example. In *Phys. Rev. E*, Vol. 71, No. 5, 2005.

[13] Information Sciences Institute. Ns-2. http://isi.edu/nsnam/ns.

[14] A.K. Ng*, J. Efstathiou. Structural Robustness of Complex Networks. Presented at *NetSci2006*, 2006.

[15] V. Latori, M. Marchiori. The Architecture of Complex Systems. In *Phys. Rev. Lett 87*, 2001.

[16] 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. In *Proc. Natl. Acad. Sci. USA*, Vol. 101 no. 52, 2004.

[17] C.I. Del Genio, T. Gross, K.E. Bassler. All scale-free networks are sparse. In *Phys Rev Lett.*, 107:178701, 2011.

[18] R. Albert, H. Jeong, A.L. Barabási Error and attack tolerance of complex networks. *Nature 406*, 378–382, 2000.

[19] 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.

[20] 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.

[21] L. Beauguitte, C. Ducruet. Scale-Free and Small-World Networks in Geographical Research: A Critical Examination. In *17th European Colloquium on Theoretical and Quantitative Geography*, Athénes, 2011.

[22] P. Hovareshti, H. Chen, J.S. Baras Motif-based Topology Design for Effective Performance by Networks of Mobile Autonomous Vehicles. In *Proceedings of the Eighth International Conference on Complex Systems (ICCS 2011)* , pp. 571–573, 2011.

[23] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci.*, 100(21):11980–11985, 2003.

[24] U. Alon. An Introduction to Systems Biology Design Principles of Biological Circuits. 2007.

[25] G. Fagiolo. Clustering in Complex Directed Networks. *Phy. Rev. E.*, 2007.

[26] D. Barmpoutis, R.M. Murray. Networks with the Smallest Average Distance and the Largest Average Clustering. arXiv:1007.4031v1, 2010.

[27] C. Chang, C. Lin LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 2011.