Network Complexity Models for Automated Cyber Range Security Capability Evaluations

Relating Network Complexity to Defensive Difficulty to Enable Comprehensive Evaluation

Thomas J. Klemas and Lee Rossey
SimSpace Corporation
tom@simspace.com

***Abstract*—For any organization to maintain a strong cyber security posture, it is important to test readiness and capabilities of cyber security teams and the tools that they use. In order to design and conduct experiments to assess performance of defensive cyber security teams and tools, it is crucial to either ensure the test range accurately represents the real environment in which the defensive teams or tools normally would operate or to ensure that testing is conducted across a suite of test ranges that provides comprehensive coverage of the potential real-life network environments. In this paper, we present a novel network complexity scoring framework that is designed to capture the set of the network attributes that have the principal impact on the performance of defenders and defensive tools and to differentiate networks according to defensive difficulty.**

Keywords – Network; network complexity; network theory, graph theory; degree; connectivity; connectance; centrality; degree centrality; hyper-edge; hypergraph; information theory, eigen decomposition; eigen vector; eigen value; eigen centrality; Bonacich centrality.

I. INTRODUCTION

No longer a newly emerging issue, cyber security is a continuing and rapidly growing challenge, facing all organizations, whether small, large, public, or private. Thus, added to conventional risk management, there is an imperative to manage cyber risk by a combination of building and maintaining a strong cyber security readiness posture, as well as other approaches, such as cyber insurance. Cyber security readiness depends on knowledge skills, regular training of cyber security defenders, in addition to the organization's information technology architecture, cyber security policies, enforcement of these policies, defensive technologies, and many other contributing elements of cyber readiness. To train cyber security staff and develop defensive skills, many organizations have initiated regular red-blue challenges, with real or automated "red" cyber adversaries attacking a virtual organization that "blue" cyber defenders are tasked to defend, on cyber ranges that are intended to emulate the organization's real networks. In addition, cyber defensive technology companies validate and demonstrate their tools, similarly, on cyber ranges. In either of these emerging applications, it is critical to develop a notion of the complexity of the network on which the red-blue gaming or performance testing is conducted in order to understand cyber defensive performance.

Many notions of complexity have been explored heretofore and Wikipedia, [11], has a nice overview of several of these, but an immediate observation from examining the Wikipedia page is that there is great variation in the definition across applications. Most specific previous descriptions of complexity and research in modeling network complexity, such as the research presented in [2][3][7], was primarily focused in other application areas either non-specific to or other than cyber security, so that work could not be directly leveraged for our purposes. In addition, a number of the earlier efforts are primarily qualitative in nature, such as [1], and therefore did not align with our objectives for this research. Some previous efforts that were closer to the computer network application area, like [4], either were focused on a different objectives or overly simplified the problem, employing tabular approaches to compute network complexity scores that were too limited to capture subtleties of connectivity between nodes and only allowed a linear type of model, so these methods were insufficient for our purposes. Other methods focused only on a narrow sub-set of aspects of complexity that impact defensive difficulty, ignoring many other important factors, so, again, the limitations of other approaches forced us to explore a new technique to model network complexity. However, despite the various shortcoming enumerated above, many of these antecedent approaches have influenced our efforts.

This paper introduces a hybrid complexity modeling approach that treats the network in a multimodal fashion, encapsulating certain parameter like numbers or operating systems or number of device types hyper edges of a hypergraph, abstracting them as attributes of the associated subnets or the network itself, but maintaining flexibility to model more complicated network properties. Many global network and subnet-specific single-parameter attributes are captured with a tabular method. Concepts of complexity that are distributed in nature, related to connectivity, or describe the balance of a specific property across the network are

 ISBN: 978-1-61208-512-8

analyzed using information theoretic and related approaches that better address those concepts. For example, certain aspects of subnet and router topology are better described with information theoretic model and corresponding complexity analysis.

To demonstrate the efficacy and overall utility of our complexity model, we developed numerous networks, some devised on paper and other actual cyber range networks, each emphasizing different network attributes. By analyzing this collection of varied networks, we were able to explore the model behaviors as we vary scoring parameters and confirm that the model parameters and network properties interact in the way that the algorithms were designed. As one of the scored networks, we also examine a virtual, large financial network used for cyber range training of cyber security defenders. For the large financial network case, we developed parsing routines to collect network attribute values from configuration files for the cyber range, demonstrating the potential for automating the complexity computations. This exercise directly supports a future in which the input accumulation, analysis, and complexity scoring can be accomplished by automated tools.

The remainder of this manuscript describes the proposed network complexity model in greater detail and is organized as follows. Section II describes the technical details of our model and how it describes the complexity of a network that pertains to cyber security defensive difficulty. Section III describes the performance and provides results of applying our network complexity model to multiple networks that possess sufficient variety of the values of the attributes that we deemed crucial to network complexity. Section V offers our conclusions. Finally, the acknowledgment and reference sections complete the manuscript.

II. TECHNICAL DETAILS

We begin this section by describing the fundamental design of the network complexity model and the notations that we will be using throughout the manuscript. The primary purpose of the network complexity model is to distinguish between different networks in a manner that agrees with intuitive notions of cyber defensive difficulty. As mentioned previously, we have adopted a hybrid approach that includes both tabular and information theoretic components to incorporate contributions from both global single-value attributes and distributed attributes, such as connectivity. Thus, the model is designed with the flexibility to accommodate both linear, weighted combinations of attributes and as well as more complicated functions to describe attribute contributions.

Attributes were selected based on several primary premises. We first considered attributes that describe the scale of the network, including numbers of devices and numbers of accounts. Then, we incorporated attributes that capture complexity in the structure or topology of the network, including organization of subnetworks, router connectivity, multiple security zones, and other similar concepts. Finally, we included attributes that directly impact the level of defensive effort or increase the attack surface.

Table 1, below, enumerates the attributes that comprise our network complexity model. It contains a list of network complexity attributes that contribute to the network complexity algorithm and a rating for the differentiation enabled by that attribute. In addition, table 1 provides a differentiation rating for that attribute's relative contribution to the network complexity algorithm.

TABLE I. NETWORK COMPLEXITY ATTRIBUTES

Attribute	Differentiation
User accounts	1
Machines	1
Operating System	2
Device Types	2
Firewalls	1
Protocols	2
Administrative Domains	2
Key Business Systems	1
External Interfaces	1
Router Connectivity	2
Subnet Size Distribution	1

The differentiation ratings have 2 values, 1 or 2, and indicate the relative differentiation provided by that attribute. Thus, attributes with differentiation rating level 1 contribute to the complexity score in a way that distinguishes between network to a greater degree than level 2 attributes. There are numerous other potential attributes that were deemed of less significance and would provide level 3 or lower of differentiation, and therefore not critical to include in this stage of the network complexity algorithm design.

There is no standard accepted definition of cyber defensive network complexity, so our fundamental goal was simply to design an algorithm that best suits our intended applications, in this case the evaluation of cyber defensive performance of cyber security teams, operators, or tools operating on a network, measured against the complexity of the network. As such, attribute contribution to network cyber complexity was designed to match notions of cyber defensive impact. To accomplish this, multiple iterations of functions were explored to capture the contribution of each attribute. Primarily, three types of functions were utilized to represent complexity attribute contributions. For some attributes, a linear, tabular approach best was able to achieve the desired contribution. In other cases, non-linear functions featuring the product of 2 attributes was most suited to capture the complexity. For yet other attributes, an information theoretic approach is used to map a custom probability function analog to information content, which is used to represent complexity. In most cases, a log scale, either base 2 or 10, is used to transform raw attribute quantities of very different orders of magnitude to similar magnitude ranges and yet retain the desired differentiation.

 ISBN: 978-1-61208-512-8

One of our sub-goals is to assess the structural or topological complexity of a modern day computer network as an element of our overall complexity model. The density of edges or connectivity are concepts that come to mind. They are related to measures of the importance or centrality of the nodes. Centrality measures are commonly used for this purpose, as described in [9] and [5]. Degree centrality, Bonacich centrality, closeness or path centrality, betweenness centrality, eigenvector centrality are well known examples. The degree of nodes in a network is useful to describe connectivity but that may not correspond to defensive complexity. Towards that end, we developed several new concepts for complexity, including hop complexity, subnet complexity, and others.

To measure hop complexity, we adapted an approach utilized in [8]. First, we specify that the hop count corresponds to the number of hops for traffic to traverse the shortest path of the network between a particular vertex and each other vertex in the network. In this complexity sub-model, the aforementioned functional is defined as:

$$g_F(r_i) = \beta^{c_1|S_1(r_i)|+c_2|S_2(r_i)|+\cdots+c_{\rho_r}|S_{\rho_r}(r_i)|} \quad (1)$$

The router, r_i, is one of N_r routers in the network. The parameter beta, β,is selected empirically to help differentiate between multiple networks based on hop complexity. The functions $S_m(r_i)$, in the exponent, enumerate the number of routers that can be reached within m hops of router r_i and scale each such hop value tally by a corresponding combining coefficient, c_m. Utilizing the g_f function we can construct a related function p for router r_i as follows:

$$p(r_i) = \frac{g_F(r_i)}{\sum_{j=1}^{N_r} g_F(r_j)} \quad (2)$$

Thus, for beta greater than one, the value will be greater for a router which has more routers that can be reached within a certain hop count, all else remaining the same, and we will use that function p to compute information content [10] of the network, based on the hop complexity, as follows:

$$g_{HC}(\bar{x}) = -\sum_{i=1}^{N_r} p(r_i) log\big(p(r_i)\big) \quad (3)$$

We observe that the protocol complexity is a complementary attribute to hop complexity and measures the number of types of traffic traversing the network, so we combine these two attributes in one measure by forming the product. Our intuition suggests that this contribution element is potentially an analog for flow complexity.

The subnet complexity is computed in a similar manner except that the probability function is defined as a simpler and more intuitive ratio of the number of nodes in the subnet divided by the total number of nodes in the network.

$$p(S_i) = \frac{N_{ni}}{N_T} \quad (4)$$

The information content or associated complexity computation is the identical formula as previously used to compute the router connectivity.

$$g_{SNC}(\bar{x}) = -\sum_{i=1}^{N_n} p(S_i) log\big(p(S_i)\big) \quad (5)$$

Other attributes contribute to our complexity model through straightforward, tabular functions. For example, this approach is used to capture the complexity associated with the diversity of technology deployed on the network. The distribution of operating systems and device types across the subnets are scaled and summed. A slightly more complicated contribution results from pairs of attributes, such as administrative domains and user accounts, that contribute as a scaled product of the direct values.

Employing the approaches described above, to capture complexity contributions from all the attributes enumerated in table 1, we have developed a network complexity model that provides strong differentiation between networks to enhance scoring of cyber range defensive and offensive testing and war-gaming. In the next section, we will discuss the results of applying the model to a variety of networks of different size, structure, and technological diversity.

III. RESULTS

To explore the ability of the model to distinguish key differences in network structure or other attributes that impact network complexity, we developed more than 10 networks that we modeled and analyzed. The largest network, named the large financial network (LFN) has the largest size, most intricate structure, and generally the high degree of all contributing complexity factors. Following that are three similar networks with 8 routers but differing distributions of machines across subnets and router nodes. Clearly demonstrated throughout the results in this section, the scoring model was designed such that evenly distributed networks contribute the most defensive complexity, whereas the networks with the greatest degree of clustering of machines within fewer subnets pose slightly less defensive complexity. Finally, there are multiple 5 node and 4 node networks arranged with three different connection graphs: all the routers in a line, all the routers connected in a ring, and full connectivity between all the routers. These example networks demonstrate that the networks which require the greatest number of hops for traffic to traverse the network tend to incur a higher score in our cyber defensive complexity model, which matches the design objectives.

In this section, we explore the attribute contributions to the overall cyber network complexity score for each of the test networks. Figure 1 illustrates the hop complexity sub-score generated by our model, measured for each of the test networks.

 ISBN: 978-1-61208-512-8

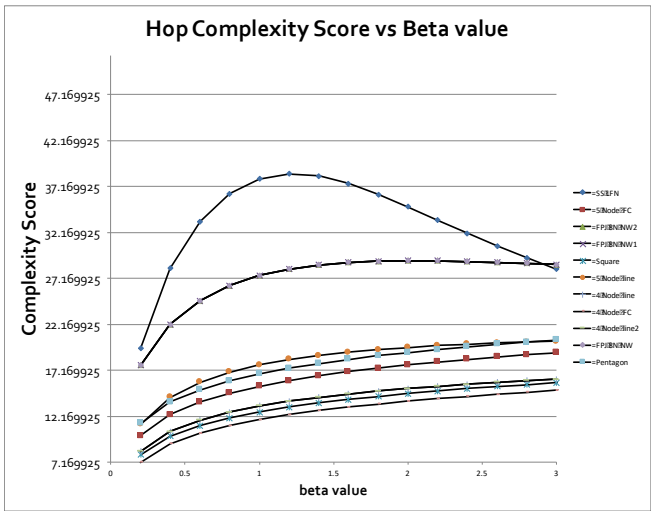

Figure 1: This scatter graph plots the complexity scores, assigned to the selected networks based on the network topology, as measured by hop counts for the various routers, versus the parameter beta. Note that the value 1.2 provides the greatest differentiability.

Since the hop complexity, shown in figure 1, captures topological and structural complexity of the network, captured through a measure of router interconnectivity, we decided to combine hope complexity with the complexity resulting from the number of protocols that must be supported on that very network with the described topology. Thus, figure 2 illustrates the product of hop complexity with protocol complexity

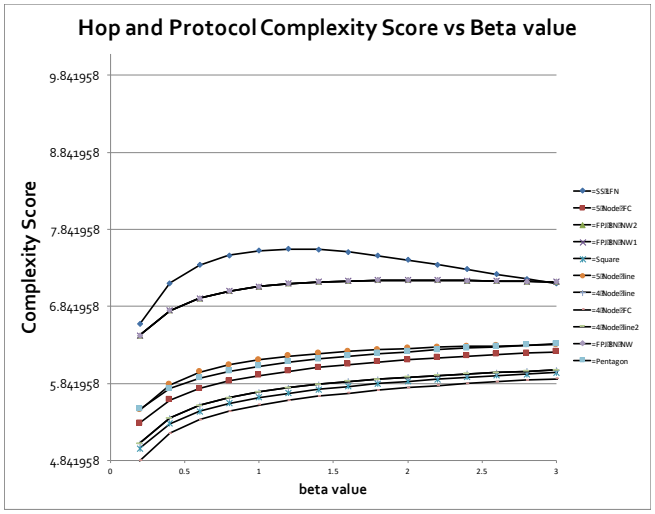

Figure 2: This scatter graph plots the complexity derived from the hop complexity and the protocol traffic traversing the network. It is plotted versus the parameter beta. Note that a beta value of approximately 1.2 provides the maximum differentiability.

As described in the technical details section, the router hop and network protocol derived complexity measures scale with the degree of asymmetry in the distribution of hop complexity across the network routers as well as directly with the number of protocols traversing the network. This relationship is dependent on the parameter beta and as we can see the value 1.2 seems to maximize the differentiability of this complexity measure.

In figures 3 and 4, we can see the influence of the distribution of device types and operating systems in the network. These complexity measures are directly linked to numbers of device types and operating systems, as well as the subnet distribution complexity for these quantities.

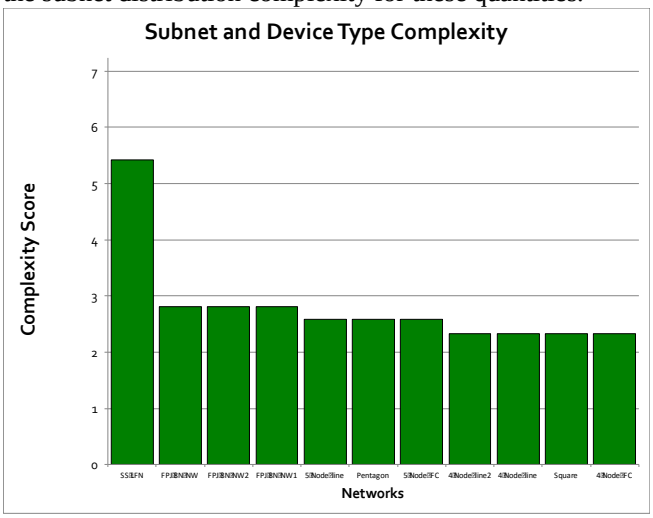

Figure 3: This bar chart shows how subnet and device type complexity impact overall complexity scores for each of the different networks.

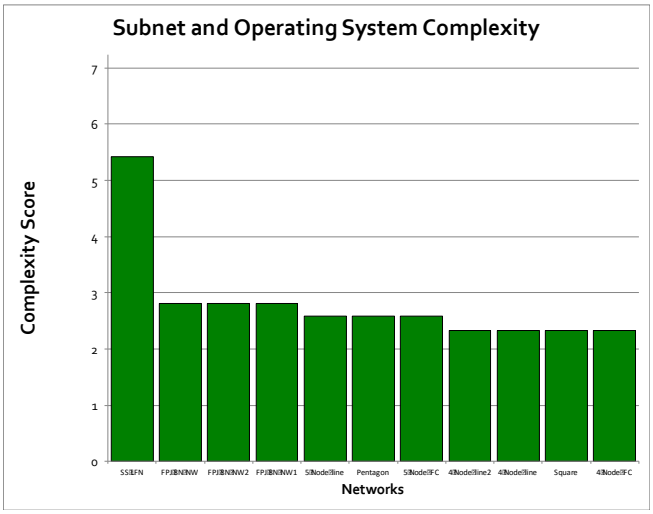

Figure 4: This chart depicts the complexity due to operating systems deployed on the various subnetworks of the network.

The fifth figure presents the complexity associated with firewalls and external interfaces in the network. The subnet complexity sub-score generated by our cyber defensive complexity model is illustrated in figure 6. Notice how the

subnet complexity scores scale with the distribution of subnets and machines across the various test networks.

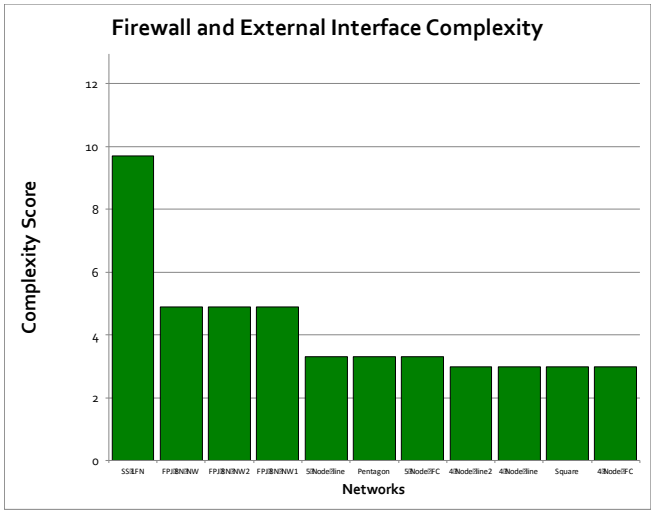

Figure 5: This bar chart shows the complexity derived from firewalls and external interfaces of the overall network.

For example, we see that FPJ 8N NW2 incurs the highest complexity score since it has the most nodes by a significant factor, and those nodes are distributed fairly evenly. Then, FPJ 8N NW1 has the second highest complexity sub-score because it has significantly fewer nodes, but those nodes are arranged evenly. Finally, FPJ 8N NW has the lowest sub-score, because, although it contains the same number of

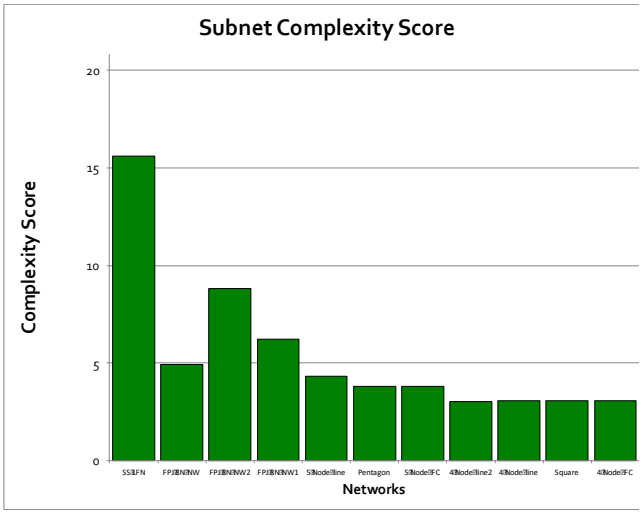

Figure 6: This plot shows scores generated by the subnet functional designed to measure complexity due to topology and machine distribution, based on an information theoretic approach.

nodes as NW1, those nodes are distributed asymmetrically across the subnets of the network, reducing the complexity slightly relative to NW1.

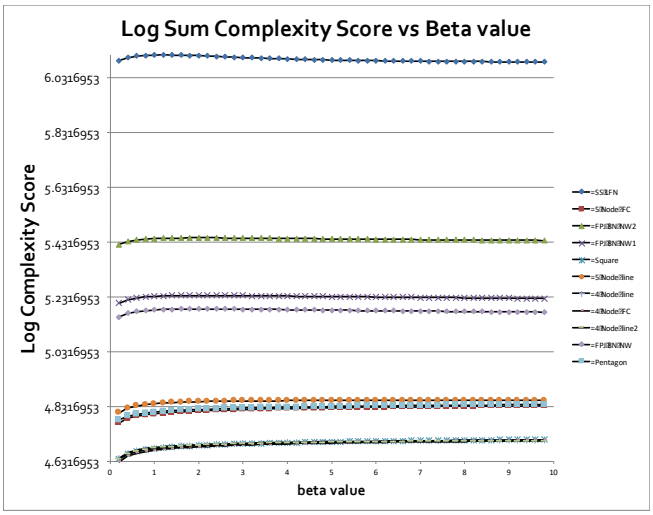

Figure 7: This scatter plot captures the total complexity of the network, summing each of the contributing elements, including hops, protocols, device types, operating systems, administrative domains, user accounts, administrative domains, subnet distribution, and numbers of machines.

Finally, in figure 7, we see a scatter plot showing the overall cyber defensive complexity score computed by superposition of all the complexity model attribute sub-scores. Since the overall scoring model retains the parameter beta, utilized in the hop and protocol complexity sub-model, the overall score is also a function of beta. However, as discussed earlier, a value of approximately 1.2 produced the greatest differentiation between the various test networks. Thus, the overall complexity score would be approximately 6.11 for the LFN network, 5.45 for the FPJ 8N NW2 network, 5.24 for the FPJ 8N NW1 network, 5.19 for the FPJ 8N NW network, 4.85 for the 5 node line network, 4.81 for the 5 node ring network, 4.81 for the 5 node fully connected network, and 4.68 for the 4 node networks.

IV. CONCLUSION

In this research, we have explored the efficacy of using a hybrid approach involving network theory, information theory, and tabular functions to model the cyber defensive complexity of various test networks. The results presented in this paper demonstrate that our model is able to differentiate between a selection of networks with varying attributes. Future work will involve incorporation of new attributes, such as supported applications, inclusion of cloud services, and other features that will enhance the model.

ACKNOWLEDGMENT

The authors would like to thank the SimSpace corporation for presenting us with the opportunity and means to conduct this research and solve this difficult problem and the leadership for sharing insights as to challenges facing the marketplace that we addressed in this research.

References

[1] M. Behringer, "Classifying network complexity", Proceedings of the 2009 workshop on Re-architecting the internet, ACM 978-1-70668-749-3/09/12, pp. 13-18, 2009

[2] D. Bonchev and G. Buck, "Quantitative measures of network complexity," Complexity in Chemistry, Biology, and Ecology, pp. 191-192, 2011.

[3] J. Carlson and J Doyle, "Complexity and robustness", Proceedings of National Academy of Sciences (PNAS), www.pnas.org/cgi/doi/10.1073/pnas.012582499, Vol. 99, Suppl. 1, pp 2538-2545, 2002.

[4] B. Chun, S. Ratnasamy, E. Kohler, "NetComplex: a complexity metric for networked system designs", Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, ISBN: 111-999-5555-22-1, pp. 393-406, 2008.

[5] V. Karyotis, E. Stai, S. Papavassiliou, Evolutionary Dynamics of Complex Communication Networks, CRC Press Taylor & Francis Group, 2014.

[6] T. Klemas and D. Rajchwald, "Evolutionary clustering analysis of multiple edge set networks used for modeling Ivory Coast mobile phone data and sensemaking", Proceedings of Third International Conference on Data Analytics, IARIA, ISBN: 978-1-61208-358-2, pp. 100-104, 2014

[7] M. Mitchell, "Complex systems: network thinking", Elsevier B.V., Artificial Intelligence Vol 170, Iss. 18, pp. 1194-1212, 2006.

[8] T. Moscriboda and R. Wattenhofer, "The Complexity of Connectivity in Wireless Networks", Proceedings of 25th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, 2006.

[9] M. Newman, Networks, An Introduction. Oxford: Oxford University Press, 2010.

[10] C. Shannon, "A mathematical theory of communication", The Bell Systems Technical Journal, Vol 27, pp. 379–423; 623–656, Oct. 1948.

[11] Wikipedia Community, "Complexity", Wikipedia, https://en.wikipedia.org/wiki/Complexity, 2002.

 ISBN: 978-1-61208-512-8