

# Impact of Topological Characteristics on Consensus Building in Multiagent Systems

Hicham Hatime

**Abstract**—This paper represents a network of nodes as a graph and investigates the impact of topological characteristics on consensus building. The properties evaluated are the algebraic connectivity ratio, average path, average cluster coefficient, average matching index, modularity, and the average participation. The consensus is measured by the average number of state updates that the system requires to get the global scope of the task at hand. Factor analysis and regression analysis are applied to analyze the impact of the topological characteristics, to understand the dependencies among them and to reduce the dimensionality of the analysis. A theoretical study is provided to support the results obtained by the statistical method. The insights from the analysis are discussed.

**Index Terms**—Communication system, consensus building, topology characteristics.

## I. INTRODUCTION

**I**N MANY fields, multiagent systems are often characterized by the interactions that take place among the agents. In biology, for example, the collective behavior of cells and tissues is a direct consequence of the properties of their molecular constituents. In computer networks, the overall performance is measured by each station's throughput, security, scalability of protocol, etc. In manufacturing, the quality of a product is evaluated by the functionality, reliability and maintainability of each component. The performance of such a complex system is affected by the connectivity and placement (topology) of its constituents.

Complex systems of collaborative multiagent systems have to address coordination problem as well. Constructing a communication topology for randomly dispersed nodes is necessary for exchanging information. In a mobile ad hoc network, the initial scope of information available to each node is usually limited by the information collected locally. Increasing this scope of information requires either collecting more data or requesting it from other nodes. The iterative process of exchanging local information among nodes followed by inferencing in order to gain global perspective of a problem is referred to as consensus building. There are many factors that influence the consensus building process, communication topology being the most

important of all. The relationship between the communication topology and the consensus process, at present, seems to be a black box. Studies aimed at evaluating the impact of a communication topology on the process of consensus building do not exist in the literature as of now. The purpose of this study is to explore this relationship and expose some of its inner structure. Random topologies are generated. Few topology metrics, namely, the algebraic connectivity ratio, average path, average cluster coefficient, average matching index, modularity, and the average participation are evaluated. The objective is to extract the most important metrics influencing the consensus building.

In order to investigate the intricate connectivity and coordination, the system has to be translated into an abstract concept that can be mathematically modeled. The model is then subjected to a systematic characterization which seeks to uncover its characteristics. These characteristics, in turn, provide insights into the design and organization of the network that affect the consensus building process. Statistical techniques such as regression analysis and factor analysis can then be applied to show the interactions and interdependencies among the topological characteristics, and the impact of these structural properties on consensus building. The most significant properties can be included in the model for consensus building and the performance of such a model can be analyzed in detail.

## II. BACKGROUND

Communication topology in small-world networks (SWN) has been an attractive area of research. They exhibit the desired characteristics of a multiagent system such as short path length, high clustering coefficient, and adaptive communication links. In addition, they are simple to model. Graph theory is usually used to model the topology of SWNs. A network is represented by an undirected graph  $G = (V, E)$  where nodes represent the set of vertices ( $V$ ) and the communication links are described by the set of edges ( $E$ ). In this paper, generating random SWN is accomplished through different stages [1] (see Fig. 1). The first stage utilizes a combination of ART and Maxnet neural network concepts to cluster nodes into separate groups. The second stage is the election of a cluster-head for each cluster to facilitate intracluster communication.

The election is performed using the minimum dominant set concepts. The third stage consists of establishing intercluster communication through gateways. The selection of a gateway node is evaluated by a willingness function that combines node degree, location to cluster center, available power, and transmission energy. The last stage consists of establishing connections between clusters.

Manuscript received June 07, 2011; accepted June 07, 2011. Date of publication June 23, 2011; date of current version November 02, 2011. The associate editor coordinating the review of this paper and approving it for publication was Prof. Evgeny Katz.

The author is with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260 USA (e-mail: hxh-hatime@wichita.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSEN.2011.2160393

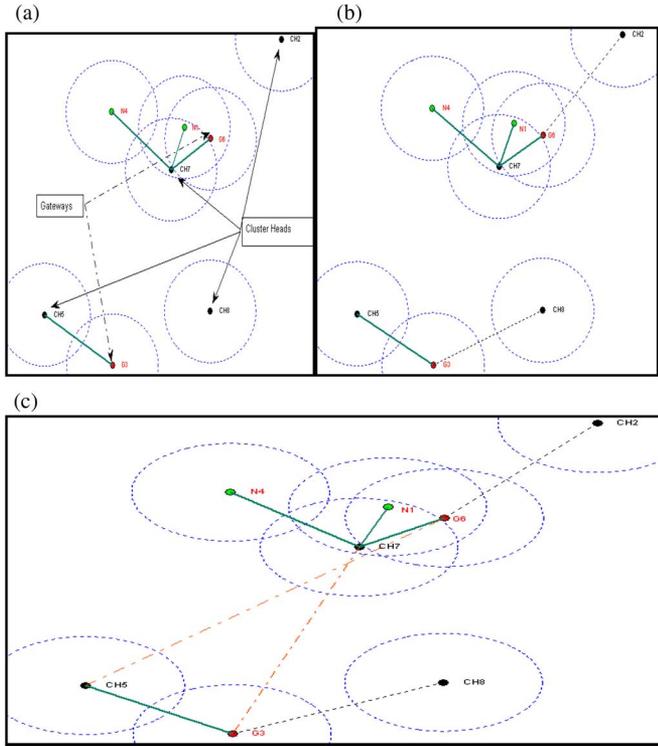


Fig. 1. Three phases in creating a communication topology among randomly dispersed wireless nodes: (a) creation of clusters and intracluster links, election of cluster-heads and selection of gateways; (b) connecting node clusters to closest gateways; and (c) establishing intercluster links. (CH = Cluster Head, G = Gateway, N = Node).

### A. Topological Metrics

The objective, here, is the selection of a set of topological metrics that need to be evaluated [6], [8]–[15]. Several structural properties and metrics were defined in the literature for networks based on intuitive concepts [2]–[5]. Due to the difficulty in enumerating all properties and metrics, it becomes a necessity to choose a suitable subset. The challenging question then is to identify the most relevant properties and metrics. In this study, the focus is on introducing metrics that were not investigated extensively in the literature, while keeping some of the well investigated metrics (average path and cluster coefficient). Another goal is establishing a relationship between the new set of parameters and other parameters that are commonly used in the existing models in the literature.

### B. Average Path

In graph theory, the minimum distance  $d_{ij}$  between two vertices  $i$  and  $j$  is measured by the length of the shortest path between the vertices or the minimal number of edges that can be traversed. The average path length  $d = \langle d_{ij} \rangle$  is defined as the average distance between all pairs of vertices.

Computing the shortest path between two vertices is not straightforward. The commonly used algorithm is Floyd–Warshall’s [7], which runs in  $\theta(n^3)$  and returns shortest path between all pairs of vertices.

### C. Average Cluster Coefficient

Cluster coefficient is another characteristic that relates to the internal structure of a network, more precisely its cohesiveness. It measures the probability that two vertices with a common neighbor are connected. Given a vertex  $i$  with  $n_i$  neighbors,  $n_i(n_i - 1)/2$  possible edges could exist between the neighbors. The clustering coefficient  $C_i$  could be defined as the ratio of the actual number of edges  $E_i$  over all the possible edges

$$C_i = \frac{2E_i}{n_i(n_i - 1)}. \quad (1)$$

The average cluster coefficient is defined as  $C = (1/n) \sum_{i=1}^n C_i$ , where  $n$  is the number of nodes.

### D. Average Matching Index

Matching index is a basic measure for neighborhood analysis. It quantifies the topological overlap between two nodes by measuring their similarity in terms of common shared neighbors. The similarity is given by the following expression:

$$\begin{aligned} M_{ij} &= \frac{\sum \text{common neighbors}}{\sum \text{total number of neighbors}} \\ &= \frac{\sum_{k,l} a_{ik} a_{jl}}{n_i + n_j - \sum_{k,l} a_{ik} a_{jl}} \end{aligned} \quad (2)$$

where  $a_{ij}$ ,  $a_{jl}$  are elements of the adjacency matrix and  $n_i$ ,  $n_j$  are the respective number of neighbors of vertices  $i$  and  $j$ . The average matching index is defined as:  $AMI = \langle M_{ij} \rangle = (1/n^2) \sum_{i=1}^n \sum_{j=1}^n M_{ij}$ .

### E. Average Participation Coefficient

Participation coefficient is a characteristic that defines the role of each node in a complex network. The role fulfilled by a vertex is an important factor since it decides the node’s connections. A clusterhead or a gateway has more connections than a peripheral node since they act as facilitators in addition to their default role. This leads to the interpretation that the participation coefficient provides information about the link distribution among different clusters. A higher value indicates that the network has a higher ratio of nodes to clusters and *vice versa*. Computing the participation coefficient  $P_i$  of vertex  $i$  is given by

$$P_i = 1 - \sum_{c=1}^n \left( \frac{k_{ic}}{k_i} \right)^2 \quad (3)$$

where  $k_{ic}$  denotes the degree of  $i$  within its cluster  $c$  and  $k_i$  is the total degree of  $i$ . The average participation coefficient is defined as  $P = \langle P_i \rangle$ .

### F. Algebraic Connectivity Ratio

Another commonly used property in measuring the connectivity in a graph is the algebraic connectivity parameter. It is an important factor in consensus building since it directly affects the network synchronization. This parameter also provides information about the robustness of a topology in terms of link and

node failure. The higher the value the more failure tolerant is the network. In a given topology, this parameter is evaluated by the second-smallest eigenvalue of the Laplacian matrix  $L$ . The ratio  $\gamma$  of the second-smallest eigenvalue over the last eigenvalue defines the algebraic connectivity ratio

$$\gamma = \frac{\lambda_2(L)}{\lambda_n(L)}. \quad (4)$$

The eigenvalues of  $L$  are ordered such that they satisfy:  $0 = \lambda_1(L) < \lambda_2(L) < \dots < \lambda_n(L)$ .

### G. Modularity

Cluster coefficient, as defined previously, measures the bonding between nodes. Modularity, on the other hand, quantifies the strength of clusters and evaluates the quality of the clustering. When “good” clustering occurs, intracluster connections are dense and intercluster links are sparse, whereas, when “bad” clustering occurs, nodes form a cluster on their own. One of the definitions of modularity is the difference between the fraction of intracluster edges over the total network edges and the fraction of extremities that connect two clusters and is given by

$$Q = \sum_{i=1}^n (e_{ii} - b_i^2) \quad (5)$$

where  $e_{ij}$  represents the fraction of within-cluster edges,  $b_i = e_{ii} + (1/2) \sum_{i \neq j} e_{ij}$  and  $e_{ij}$  represent the fraction of edges having one extremity in cluster  $i$  and the other in cluster  $j$ .  $Q$  is proportional to quality and tends to be negative for “bad” clustering.

## III. CONSENSUS ALGORITHM

In the previous sections, the steps involved in building the communication network were presented and a brief description of some structural properties to be extracted from the communication topology was provided. This section, presents a description of the distributed implementation of the consensus building algorithm along with the mathematical formulation of the information flow that will be used as the main metric to evaluate the average state update.

Before presenting the consensus building algorithm, we need to define initial target awareness for each node. At this level, nodes will have partial or no information about the targets location. The objective of each node is to build the list of all targets through the consensus algorithm. Since target discovery is not within the scope of this paper, two simple approaches were considered. The first one assigns randomly a target to a node. The advantage of this method is its fairness in target awareness. However, it fails on proximity criteria (target assigned to the closest node). The second approach uses a self organizing map principle when the winner (closest) takes all. It has the advantage of being close to real-world scenario. However, its shortcoming is the lack of fairness. Many nodes end up with no initial information about targets. To remedy this issue, targets were also randomly positioned to increase the chance of partial knowledge. Having built a target scope for each node, the

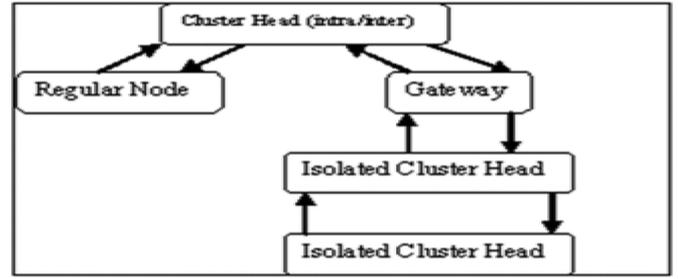


Fig. 2. Diagram showing the flow of information between different types of nodes.

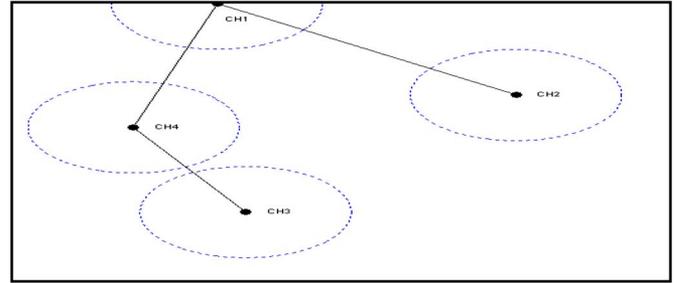


Fig. 3. Nodes forming clusters on their own where CH3 is the root.

next step is to exchange these initial scopes in order to build a consensus (final list of targets). The task is performed through an iterative process that is divided into two data flows processes: bottom up and top down (Fig. 2). Bottom up is used by a child to update its parent node. The top down flow is utilized by a parent to update its child. However, due to randomness in the placement of the nodes, there are cases where nodes form clusters on their own. Fig. 3 illustrates an example of isolated clusterheads (ICHs), where CH3 is the root and CH2 is the last descendant. To accommodate this case and reach a consensus among ICHs, the algorithm includes communication data between ICHs. Thus, the following.

#### Bottom up flow:

- Each ICH that has a parent ICH sends its target scope to it.
- Each ICH or Cluster head (CH) that has a parent Gateway sends its target scope to it.
- Each gateway or regular node sends its target scope to the parent cluster head.
- During this bottom up data forwarding, each node receiving a target scope updates its scope.

#### Top down flow:

- Each CH that has a child gateway or regular node sends its target scope to it.
- Each gateway that has a child ICH or CH sends its target scope to it.
- Each ICH that has a child ICH sends its target scope to it.
- Similarly, each node receiving data, updates its scope.

In summary, before the “target discovery” process, all the nodes in the network are in a state of zero knowledge  $S_i(t=0) = 0$ , where  $S_i(t)$  is the target scope of node  $i$  at time  $t$ . After

the initial discovery of targets, the state of some nodes changes to reflect the partial awareness  $0 \leq S_i(t=1) \leq n$  assuming that there are  $n$  targets to discover. At the end of the iterative communication process and exchange of information, the nodes reach the saturation state, where they all have a complete and identical view of the targets list.

In general, the state of a node  $i$  changes whenever it receives a state update from either its parent or descendant(s). This can be formulated by the following equation:

$$S_i(t) = S_i(t-1) + \sum_{j=1, j \neq i}^m a_{ij} s_j(t) \quad (6)$$

where  $m$  is the number of nodes,  $a_{ij}$  denotes the adjacency of node  $i$  in the network topology and emphasizes that only the parent or descendant(s) are allowed to share their state  $s_j$  with it.

However, not all exchange of information between a node  $i$  and its parent or descendant(s) results in updates of  $S_i$ . Sometimes, it takes a few iterations for new information to travel from one extremity of the network to the other, meanwhile, nodes in the middle exchange the same state. To incorporate this into (6), let us define an activation function

$$\Phi(x) = \begin{cases} x, & \text{If } x \text{ is a new target.} \\ 0, & \text{Otherwise} \end{cases}$$

Equation (6) then becomes

$$S_i(t) = S_i(t-1) + \sum_{j=1}^m \Phi(a_{ij} S_j(t)). \quad (7)$$

At any given time, a node must check whether it reached the saturation state or not. To accomplish that, it evaluates its state through the following saturation function:

$$\sigma(s_i) = \frac{2}{1 + e^{\left(\frac{n-S_i}{an}\right)}} \quad (8)$$

where  $a$  is a constant. If the returned value is 1, the node's target list is complete. The iterative update process ends when all the nodes reach their saturation.

To measure the consensus building, the average state changes could be defined as the total number of necessary updates performed by each node to reach saturation over the number of nodes as follows:

$$ASU = \frac{\sum_{i=1}^m \sum_{t=0}^T S_{i,t}}{m} \quad (9)$$

where  $m$  is the number of nodes,  $S_{i,t}$  is the state update of node  $i$  at time  $t$ , and  $T$  is the saturation time.

Practically, in small world or random networks, a node can verify whether it reached the saturation state or not without knowing in advance the number of targets to discover. It can be accomplished by evaluating the network size through its diameter, which increases logarithmically with the number of nodes [16]. Thus, if the number of unchanged messages coming from one direction exceed  $\log(m)$  ( $m$  number of nodes), it is an indication that saturation had been reached.

## IV. METHODOLOGY

In any scientific field, studying the nature of parameters is necessary, but not sufficient. The study should also account for the relationship among the variables. The idea of variables being interrelated is a well-accepted concept that can be mathematically represented. It helps in better understanding the contribution of each parameter, by itself and through interaction, to the overall result. Hence, the evaluation of parameter is an important factor in deciding whether to account for it or ignore it in the study.

Among the commonly used multivariable statistical methods that help scientists in analyzing the relationship between variables is factor analysis.

The aim of factor analysis [17], [18] is to reduce the set of observable variables to few hypothetical variables called factors. These factors are supposed to contain the essential information in the original variable set and are constructed in a way that reduces the overall complexity of the data by taking advantage of inherent interdependencies.

In addition to studying the interrelationships between the variables, the objective of this section is to partition data into mutually exclusive sets with different weights. The importance assigned to each variable within a set will also help in reducing the dimensionality of data.

The factor analysis method involves the following major steps.

- Collecting data and computing the correlation matrix.
- Extracting the unrotated factors.
- Rotating and interpreting the factors.

### A. Data Collection and Correlation Matrix

The selection of the variables to investigate has a critical bearing on the output of the study. The problem becomes more complex when the conducted investigation is an exploratory one. The lack of knowledge about the underlying nature of variables can lead to the discovery of higher orders of overlapping or similarities among the variables that requires a reassessment of the variables under consideration.

As mentioned previously in Section II-A, the objective is to establish a bridge between newly introduced variables and some already studied parameters. The topological variables we intend to investigate using factor analysis are the following:

- average path;
- average cluster coefficient;
- average matching index;
- average participation coefficient;
- algebraic connectivity ratio;
- modularity.

To perform a study using factor analysis method, there are some constraints that need to be satisfied. The first constraint deals with the measurement characteristics of the variables. Ideally, factor analysis requires the data to be continuous and normally distributed (bell-shape). This requirement is important because it has a direct effect on the reliability of the correlation coefficient which is an indicator of the degree of correlation between variables. In practice, a bell shape is seldom satisfied. Prior to analysis, a preprocessing stage is sometimes required to clean up data from noisy, missing and outlier values. Since it is an exploratory phase, the preprocessing stage should be limited

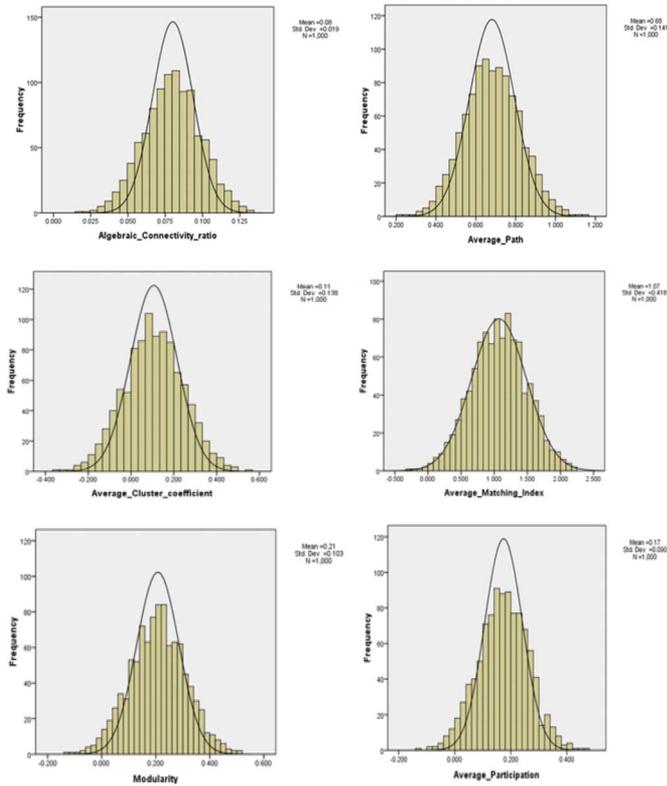


Fig. 4. Histograms and normal distribution plots for the six variables using a sample size of 1000.

to normalizing data to fit a bell shape distribution. The initial test of normality on raw data revealed that none of the variables followed a normal distribution. The best goodness-of-fit value obtained was for a four-parameters beta distribution. The histograms and distribution fitting of the six normalized variables are presented in Fig. 4.

The second constraint deals with size of the sample. As the size increases the more stability is introduced in the evaluation of the correlation coefficients. Factor analysis recommends a sample size between 500 and 1000. Below that range, spurious variance distortion increases, and above that range no major gains are obtained.

After the data preprocessing step, the next crucial point in factorial analysis is the evaluation of correlation coefficients. This step helps in determining the relationship strength between the set of measures upon which the final decision will rely on. A major concern at this stage is the accuracy of the correlation coefficients. A result is statistically significant if it is unlikely to have occurred by chance. Usually, a statistical test of significance is devised for each coefficient. It computes the probability  $p$  of a result being larger and compares it to the confidence level  $\alpha$ . The result is said to be statistically significant if  $p \leq \alpha$ . However, a statistical significant result is not always of practical significance. Since the conducted study is an exploratory one, the only option available is to rely on the statistically significant results. To compute the correlation matrix for the six topological metrics, Pearson’s correlation coefficients is used. Pearson’s technique requires the data to be normally distributed and uses the covariance of the two variables over their standard

deviation product to compute the elements of the correlation matrix [21]. The coefficients can take on values between  $-1$  and  $+1$ . The result is a correlation matrix for each pair of variables along with a p-value for the test of significance for each coefficient. To interpret the correlation coefficients, statisticians recommend squaring the coefficients. The new coefficient is called a coefficient of determination. It provides information about the contribution of each variable to the total variance in one variable. For example, if the correlation between  $V_1$  and  $V_2$  is 0.5, then 25% of the variance in  $V_1$  can be explained by the linear relationship between  $V_1$  and  $V_2$ . The other 75% remains unexplained.

*B. Factor Extraction*

The general goal of factor analysis is to define and extract clusters of highly interrelated variables from the correlation matrix [19]. These clusters are called factors and can be written as a linear combination of the variables. Different techniques for factor extraction have been proposed. Each has its possibilities, limitations, and advocates [20]. The techniques include principal component, maximum likelihood, alpha, image, centroid, minimum residual, etc. The most widely used is principal component. It has the characteristic of being suitable for exploratory factor analysis. However, some assessments and evaluations of these methods suggest that the best method to use when data is normally distributed is maximum likelihood because it evaluates a wide range of goodness of fit indexes for the model [21].

Choosing an extraction technique is only the first step in factor extraction. The next important step is to decide on the number of factors to retain. The main objective is not to over-extract or under-extract. But there is a general consensus among researchers that it is better to extract too many factors than too few. At this stage, different criteria are proposed to stop the extraction process [19]. The most commonly used are the eigenvalues which are the sum of the factors squared weights. The cutoff point is a value above one. Another criterion based on discontinuity is the scree test [21]. It involves locating the point where the sudden drop of eigenvalues occurs on a scree plot. The points above it are generally the factors to retain. A third criterion relies on the desired percentage of variation to be explained by the factors. Extraction in this case, is subjective and relies on the user requirements on when to stop extracting.

In this investigation, the idea is to experiment with principal components and maximum likelihood. The objective is to compare their results in order to choose the best method to use for the analysis. Maximum likelihood has the characteristic of providing the correct number of factors to extract due mainly due to the definition of the degree of freedom for the chi-square test which depends on the set of variables  $df = ((z - k)^2 - (z + k))/2$ , where  $z$  denotes the number of variables and  $k$  is the number of factors to extract. Knowing the number of factors we will then use the method that is more accurate.

*C. Factor Rotation and Interpretation*

The output of factor extraction phase is a factor matrix that contains the weights of each variable on the extracted factors also called loadings. These coefficients are unrotated [20]. They lack meaningful interpretation, although they account

for maximum variability. The objective of factor rotation is to preserve the amount of the variability, while enhancing the interpretability. The improvement is obtained by rotating the axes, thus resulting in a clear separation of variable clusters on each factor. In other words, each cluster of variables will be associated with one factor which simplifies the interpretation.

There are two methods for factor rotation [19], [21]. Each offers a variety of choices. The orthogonal method, in which the factors are kept at a 90° angle of each other, employs varimax, equimax, and orthomax as a rotation method. The oblique method, on the other hand, departs from the right angle rule and offers promax, oblimin, and quartimin as options. Orthogonal rotation produces factors that are uncorrelated and easy to interpret. The oblique method takes into consideration the possible existence of correlation between the extracted factors. Forcing orthogonality on correlated factors, results in a loss of information. In exploratory data, it is recommended to implement the oblique method since both methods yield same results on uncorrelated factors.

To rotate the factors in this study, the oblique method with the “promax” option was implemented.

#### D. Multiple Regression Analysis

Multiple regression analysis is a statistical tool aimed at uncovering and understanding the relationship between a variable to be explained, called a dependent variable, and explanatory variables, called independent variables [22]. Multiple regression analysis tests whether the independent variables contribute to the variance in the dependent variable or not. It is used in many fields to measure the magnitude of the effect, if present, and forecast its value.

Multiple regression analysis involves many steps, among them, the characterization of the relationship between the dependent variable and the set of predictors by a mathematical representation. The model can adopt a linear or a nonlinear form, depending on the nature of the relationship. It is a crucial step to determine this nature before proceeding with the analysis itself. Another step that can substantially bias the results is the choice of the predictors. Failure to select the appropriate independent variables can lead to the collinearity effect [23]. This happens when two or more independent variables contain much of the same information, making it difficult for regression analysis to determine which one is more important. Another concern, at this step, is the precision of the measurements when irrelevant variables are included, which tend to reduce the precision of the results.

Multiple regression analysis offers a set of tools that helps in the interpretation of the results. The most important are the following.

- The R-square represents the portion of variance in the dependent variable explained by all the predictors when grouped together.
- The ANOVA (analysis of variance) table, which can be thought of as a significance value for the whole model, confirms the statistical significance of the R-square value.
- The coefficients table indicates both the weight of each variable in the equation and informs about the severity of collinearity.

TABLE I(a)

SUMMARIZES THE PERCENTAGE CONTRIBUTION OF EACH VARIABLE TO THE TOTAL VARIANCE OF A VARIABLE. THE SHADED AREAS ARE THE CONTRIBUTION THAT FAILED THE STATISTICAL TEST OF SIGNIFICANCE. THE LAST COLUMN INDICATES THE COMBINED CONTRIBUTIONS OF ALL VARIABLES TO THE VARIANCE OF ONE VARIABLE

Topology metrics	1	2	3	4	5	6	Total Contribution
Algebraic Connectivity Ratio (1)	~	L	L	L	L	L	Medium
Average Path (2)		~	L	L	L	L	Low
Average Cluster Coefficient (3)			~	L	L	L	Low
Average Matching Index (4)				~	M	L	Medium
Modularity (5)					~	L	Medium
Average Participation (6)						~	Medium

TABLE I (b)

SHOWS THE PERCENTAGE CONTRIBUTION OF EACH VARIABLE TO THE TOTAL VARIANCE OF A VARIABLE

Topology metrics	1	2	3	4	5	6	Total
Algebraic Connectivity Ratio(1)	~	4.7	0.3	4.6	10.6	13	39
Average Path (2)	4.7	~	0.2	11.1	4	0.8	14
Average Cluster Coefficient (3)	0.3	0.2	~	0.1	0	1.5	3.7
Average Matching Index (4)	4.6	11.1	0.1	~	36.7	27	62
Modularity (5)	10.6	4	0	36.7	~	4	41
Average Participation (6)	13	0.8	1.5	27	4	~	52
Total	39	14	3.7	62	41	52	

## V. STATISTICAL RESULTS

Series of experiments were conducted in a MATLAB environment where scores of communication topologies were generated using eight nodes randomly dispersed in a small geographic area. The choice of eight nodes was dictated by the desire of obtaining a balance between the number of single-node clusters and multiple-node clusters. The number of targets to be assigned to the nodes was also fixed to 8. Hopping that in the best case the initial scope of each node will be 1. The coordinates of targets, however, remained unchanged, thus allowing one parameter (topology) to change while targets positions are fixed.

Collecting data was performed in two phases. First, the six structural metrics were evaluated for each topology generated. The same topology was then used, in the second stage, as a communication mean for the nodes to update their target scope and evaluate the average state update. The data was then fed to the Statistical Package for the Social Sciences (SPSSs) for factor analysis and regression modeling.

After normalizing the data, the next step in the methodology described in Section IV-A is the evaluation of the correlation matrix for the six metrics. Using the coding listed below, Table I(a) summarizes the contribution percentage of each variable to the overall variance of a topology metric. The actual values are presented in Table I(b). The cases that failed the statistical test of significance are represented by the shaded area. Since the matrix is symmetric, only the upper diagonal is presented.

- Low (L):  $0\% \leq v < 30\%$ ;
- Medium (M):  $30\% \leq v < 70\%$ ;
- High (H):  $70\% \leq v < 100\%$ ;

TABLE II  
SHOWS THE LOADINGS FOR EACH FACTOR EXTRACTED, THE COMMUNALITIES VALUES AND THE VARIANCE ACCOUNTED FOR. *Note.* LOADINGS AND COMMUNALITIES (C) < .5 ARE OMITTED

Methods / Factors Variables	Maximum Likelihood			Principal Components		
	Factor Loading					
	1	2	C	1	2	C
Algebraic Connectivity Ratio		.65	.52		.76	.79
Average Path				.55		
Average Cluster Coefficient						
Average Matching Index	-.91		.87	-.88		.83
Modularity	.68			.80		.65
Average Participation		.78	.84		.85	.86
Eigenvalues	2.1	1.5		2.1	1.5	
% of variance	35	24		35	24	

From Table I(b), all variables exhibit little influence over each other variance. Thus, suggesting the existence of a weak correlation between the variables. However, the last column suggests that the variability in one variable can be explained by the combined variance in all other metrics. Table I(a), also, reveals that average cluster coefficient failed the test of significance suggesting that its contribution is minimal if not inexistent.

Combining factor extraction and rotation stage, maximum likelihood and principal components analysis, with “promax” rotation, were conducted to assess the underlying structure for the six variables. Several assumptions were tested, beforehand, to verify the existence of a factor analytic solution. The Kaiser-Meyer-Olkin and Bartlett’s tests were performed to measure the sampling adequacy and sphericity [23]. The first test reveals whether the distribution of data is acceptable for conducting a factor analysis. The level attributed to the data was on the border between middling and mediocre. The second test measures whether data is normally distributed and correlated enough to provide a reasonable basis for factor analysis. This condition was also met since the significance value for the test was far below the 0.05 value required. When experiencing with factor extraction, maximum likelihood did not allow the extraction of more than two factors. The two extracted factors were designed to represent two constructs: intracluster structure (factor 1) and intercluster structure (factor 2). After rotation the first factor accounted for 35% of variance and the second factor for 24%. Table II displays the variables and factor loadings for the rotated factors along with their communalities (C) which represent the amount of variance in each variable accounted for by the factor. A high value indicates that the extracted factors are a good representation of the variables. Values less than 0.5 were omitted to improve clarity.

Few observations can be made from Table II. The communalities evaluated by the principal components indicate that the factors extracted represent well the original variables. Both methods reveal that the average cluster coefficient is not a determinant variable in variance changes. Another observation, in which both methods differ, is related to the average path variable. Principal components report it as a variable to be considered with a low loading, whereas maximum likelihood suggests ignoring it.

In a second phase, a multiple regression assessment was conducted in order to determine the best linear combination of the five structural variables for predicting the consensus variable

TABLE III  
SIGNIFICANCE TEST OF LINEARITY VERIFIED FOR ALL THE INDEPENDENT VARIABLES.  $p < 0.05$

Correlation						
	Average_Update	Algebraic_Connectivity_Ratio	Average_Path	Average_Matching_Index	Modularity	Average_Participation
Pluribus_Cohesion	1.000	-.555	.121	-.551	.472	.487
Average_Update		1.000	-.217	.216	-.326	.300
Algebraic_Connectivity_Ratio			1.000	-.334	.168	.092
Average_Path				1.000	-.205	-.519
Average_Matching_Index					1.000	.200
Modularity						1.000
Average_Participation						
Sig. (1-tailed)		.000	.000	.000	.000	.000
Average_Update			.000	.000	.000	.000
Algebraic_Connectivity_Ratio				.000	.000	.000
Average_Path					.000	.000
Average_Matching_Index						.000
Modularity						
Average_Participation						
N	1000	1000	1000	1000	1000	1000
Average_Update	1000	1000	1000	1000	1000	1000
Algebraic_Connectivity_Ratio	1000	1000	1000	1000	1000	1000
Average_Path	1000	1000	1000	1000	1000	1000
Average_Matching_Index	1000	1000	1000	1000	1000	1000
Modularity	1000	1000	1000	1000	1000	1000
Average_Participation	1000	1000	1000	1000	1000	1000

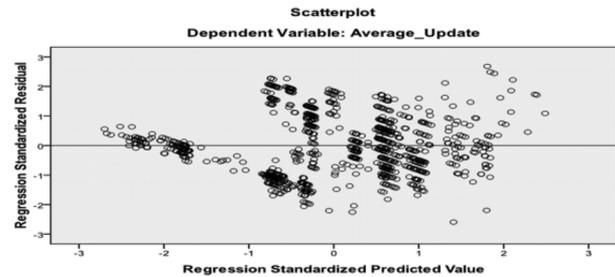


Fig. 5. Residual scatter-plot indicating that the errors are normally distributed and no pattern is generated.

(average state update). Prior to the assessment, there are some assumptions that need to be considered. The linear regression requires the relationship between each of the independents to the dependent variable to be linear. There are different ways to evaluate the linearity. The graphical method uses scatter-plots to visualize the relationship. It is the commonly recommended although difficult to interpret when the sample size is large. Another technique, statistical method, relies on the test of significance of the linearity hypothesis. The relationship is linear if the correlation coefficient between the independent variables and the dependent variable is statistically significant. This condition was met for all the five considered variables.

Table III shows that the tests are below the 0.05 level of significance. Another assumption that was also checked is the residual (predicted minus observed values) being normally distributed. The scatter-plot in Fig. 5 shows that dots are scattered and no patterns are created.

The ANOVA analysis of the regression model, in Table IV, showed that the combination of variables significantly predict the average update ( $sig < 0.05$ ). The model summary in Table V exhibited a multiple correlation coefficient R equal to 0.71 and the adjusted R-square is 0.5 meaning that 50% of the variance in average state update variable can be predicted by algebraic connectivity ratio, average matching index, modularity and average participation combined. According to Cohen’s effect size, this is a considerable effect [24]. The beta weights represented in Table VI suggested that the intracluster structure contributed more to the prediction of average update. The tolerances in the same table indicate a low collinearity (Tolerance > 0.5). This is due to the “orthogonality” of the two extracted factors.

TABLE IV  
SHOWS THAT THE REGRESSION MODEL SIGNIFICANTLY PREDICTS THE  
CONSENSUS BUILDING VARIABLE

ANOVA <sup>b</sup>						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	76.975	2	38.487	498.550	.000 <sup>a</sup>
	Residual	76.967	997	.077		
Total		153.942	999			

TABLE V  
THE ADJUSTED R SQUARE INDICATED THE PRESENCE OF A  
FAIRLY GOOD MODEL EXPLAINING 50% OF THE  
VARIANCE IN THE CONSENSUS BUILDING VARIABLE

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.707 <sup>a</sup>	.500	.499	.277847

TABLE VI  
SHOWS THAT INTRACLUSTER PARAMETERS ARE MORE IMPORTANT, THE  
COLLINEARITY EFFECT IS MINIMAL SINCE TOLERANCE IS > 0.5 = (1-0.499)  
AND THE MODEL COULD BE REPRESENTED BY THE FOLLOWING EQUATION

Coefficients <sup>a</sup>								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VF
1	(Constant)	1.817	.009		206.821	.000		
	Intra-cluster	.271	.009	.848	28.917	.000	.998	1.002
	inter-cluster	.135	.010	.311	13.880	.000	.998	1.002

Average – State – update = 0.271 \* intra – cluster + 0.135 \*  
intercluster + 1.817.

## VI. THEORETICAL ANALYSIS

So far, the statistical approach revealed that intra and intercluster links contribute to some degree in the consensus and synchronization among a set of nodes. The collected data for OCTOPUS [1] communication network emphasized the importance of intracluster links over intercluster links in the synchronization process. The subjectivity of these results is a concern. Collecting different data could lead to different results. The question is to prove mathematically the validity of the statistical results and uncover any other condition(s) on the role that intra and intercluster links play in the consensus.

In this section, we analyze the importance of intra and intercluster links in the consensus building for OCTOPUS topology. We proceed by defining a ratio for intracluster, intercluster, and overall network synchronization. We investigate the impact of intra and intercluster links on these ratios.

The consensus process is a dynamic system by nature. It involves node's state update. Some of its features, in OCTOPUS underlying lattice, are time and state variables. Modeling the dynamics of OCTOPUS consensus involves defining the so called coupled map lattice which represents the dynamics of a node with respect to the state of other nodes at a given time. The general form of the coupling equation is [25]

$$s_i(t+1) = f(s_i(t)) + \alpha \sum_{j=1}^n a_{ij} w_{ij} [f(s_j(t)) - f(s_i(t))] \quad (10)$$

where  $s_i(t)$  is the state of node  $i$  at time  $t$ ,  $\alpha$  is the coupling strength,  $w_{ij}$  is the weight of the edge between node  $i$  and  $j$ ,  $A = [a_{ij}]$  is the adjacency matrix ( $a_{ij} w_{ij} \neq 0$  if  $a_{ij} \neq 0$ ) and  $f$  is a mapping function.

In this analysis, we focus on deriving a general form of the coupling equation for intra and intercluster convergence and define a new condition for synchronization. Similar as in [1], we define a graph  $G = (V, E)$  where  $V$  is the set of vertices or nodes and  $E$  is the set of edges or links. The set of nodes  $V$  is divided into  $m$  separate clusters.  $V = \cup_{k=1}^m C_k$ , where  $C_k$  is a cluster and  $C_k \cap C_l = \phi$  for  $k \neq l$ . In a cluster, we define  $N(i)$  as the adjacent neighbors of node  $i$ . (10) then becomes

$$s_i^k(t+1) = f_k(s_i^k(t)) + \alpha \sum_{k'=1}^m \sum_{j=1}^{|C_{k'}|} a_{ij} w_{ij} [f_{k'}(s_j^{k'}(t)) - f_k(s_i^k(t))] \quad (11)$$

Within a cluster  $k$  the coupling equation is

$$s_i^k(t+1) = f_k(s_i^k(t)) + \alpha \sum_{j=1}^{|C_k|} a_{ij} w_{ij} [f_k(s_j^k(t)) - f_k(s_i^k(t))] \quad (12)$$

We also define  $\beta_{ij} = a_{ij} w_{ij}$  which restricts the dynamic of updates to the network organization. For a given cluster,  $\beta_{ij}$  informs about the links between the cluster head, dominant node, and its neighbor(s), dominated node(s). In a cluster,  $\beta_{ij}$  is only dependent of  $w_{ij}$  ( $a_{ij} = 1$ ). It can take many forms to represent the cluster lattice. One solution is to make  $\beta_{ij}$  independent of  $j$  and define it as a ratio of a node degree ( $\gamma_i^{\text{intra}} = |N_{\text{intra}}(i)|$ ), intralinks only, over the highest node degree in any cluster ( $\Delta_k^{\text{intra}}(G) = \max\{\gamma_i^{\text{intra}} : i \in V(G)\}$ ). Equation (12) becomes

$$s_i^k(t+1) = f_k(s_i^k(t)) + \alpha \sum_{j=1}^{|C_k|} \frac{\gamma_i^{\text{intra}}}{\Delta_k^{\text{intra}}} [f_k(s_j^k(t)) - f_k(s_i^k(t))] \quad (13)$$

Similarly, we define a coupling equation for nodes that play a role in intercluster communication. In OCTOPUS, these nodes are the gateways and the cluster heads. The ratio comprises the node degree, in terms of the number of intercluster links, over the total number of intercluster links. Equation (13) becomes

$$s_i^k(t+1) = f_k(s_i^k(t)) + \alpha \sum_{k'=1, k' \neq k}^m \sum_{j=1}^{|\Gamma|} a_{ij} \frac{\gamma_i^{\text{inter}}}{\Delta^{\text{inter}}} \times [f_{k'}(s_j^{k'}(t)) - f_k(s_i^k(t))] \quad (14)$$

where  $\Gamma$  is a set comprising one cluster head and one gateway within a cluster.

To define the Laplacian of the system, we set  $l_{ij}^{\text{intra}} = \gamma_i^{\text{intra}} / \Delta_k^{\text{intra}}$  and  $l_{ij}^{\text{inter}} = a_{ij} (\gamma_i^{\text{inter}} / \Delta^{\text{inter}})$  for intra, respectively, intercluster connection, and  $l_{ij} = 0$  otherwise. The Laplacian representing the cluster connections (intra) will have one row and one column of identical values. One known characteristic of a Laplacian matrix is that the sum of each row is equal to zero. For this to hold, we set  $l_{ii}^{\text{intra}} = -\sum_{j=1}^{|C_k|} (\gamma_i^{\text{intra}} / \Delta_k^{\text{intra}})$  and

$l_{ii}^{inter} = -\sum_{j=1}^{|\Gamma|} a_{ij}(\gamma_i^{inter}/\Delta^{inter})$ . Equations (13) and (14) become

$$s_i^k(t+1) = f_k(s_i^k(t)) + \alpha \sum_{j=1}^{|C_k|} l_{ij}^{intra} [f_k(s_j^k(t)) - f_k(s_i^k(t))]. \quad (15)$$

Respectively

$$s_i^k(t+1) = f_k(s_i^k(t)) + \alpha \sum_{k'=1, k' \neq k}^m \sum_{j=1}^{|\Gamma|} l_{ij}^{inter} [f_{k'}(s_j^{k'}(t)) - f_k(s_i^k(t))]. \quad (16)$$

The purpose is to derive a convergence ratio and boundary condition(s) for (15) and (16). They will help in investigating the impact of intra and interlinks on the overall convergence and synchronization process. This could be accomplished by deriving a general form for the two equations and evaluating the impact of a small perturbation on the compacted equation.

When a system approaches the synchronous state  $s(t)$ , all the nodes exhibit similar state behavior  $s_i^k(t+1) = f(s_i^k(t)) = f(s(t))$ , which is independent of the state of the neighboring node  $s_j^{k'}(t)$ . By denoting  $S(t) = [s_1(t), \dots, s_n(t)]^T$  and  $F(S) = [f(s_1), \dots, f(s_n)]^T$ , (15) and (16) take the general form

$$S(t) = (I_n - \alpha L)F(S(t)) \quad (17)$$

where  $n$  is the number of nodes. Since the sum of each row in the Laplacian matrix  $L$  is equal to zero, diagonalizing  $L$  leads to a set of nonnegative eigenvalues  $\lambda_i, i = 1, \dots, n$ . This set can be sorted as  $\lambda_1 = 0 \leq \lambda_2 \leq \dots \leq \lambda_n$ . Normalizing  $L$  using the corresponding eigenvectors  $e_i, i = 1, \dots, n$ , results in

$$S(t) = (1_n - \alpha \lambda)^T I_n F(S(t)) \quad (18)$$

At this stage, a small perturbation  $\delta s_i$  of the state is dictated by the equation  $\delta s_i(t+1) = f'(s(t))\delta s_i(t)$  [26], [27]. Along the  $i^{th}$  eigenvector of (18), the perturbation is given by

$$\delta s_i(t+1) = (1 - \alpha \lambda_i) f'(s(t)) \delta s_i(t). \quad (19)$$

One solution for (19) is [27]

$$\delta s_i(t+1) = \delta s_i(t_0) \prod_{l=0}^t (1 - \alpha \lambda_i) f'(s(l)). \quad (20)$$

A system of more than two nodes reaches a local asymptotic stability if [28]–[30]

$$\begin{aligned} & \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \log \frac{|\delta s_i(l+1)|}{|\delta s_i(l)|} < 0 \\ \Leftrightarrow & \lim_{t \rightarrow \infty} \frac{1}{t} \log \left| \prod_{l=0}^{t-1} (1 - \alpha \lambda_i) f'(s(l)) \right| < 0 \\ \Leftrightarrow & \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} (\log |1 - \alpha \lambda_i| + \log |f'(s(l))|) < 0 \\ \Leftrightarrow & \log |1 - \alpha \lambda_i| + \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \log |f'(s(l))| < 0. \quad (21) \end{aligned}$$

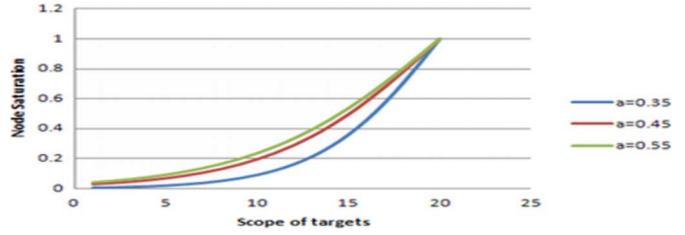


Fig. 6. Represents the evolution of a node’s saturation state for different mapping functions.

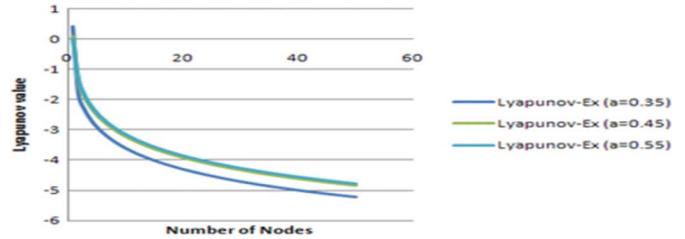


Fig. 7. Represents Lyapunov exponent values for different number of nodes and different mapping functions.

In a dynamical system with an evolution equation, the second term of the addition is identified as the Lyapunov exponent  $\mu$  [29]. The more negative the exponent, the greater the stability of the system

$$\mu \equiv \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{l=0}^{t-1} \log |f'(s(l))|. \quad (22)$$

Equation (21) becomes

$$|1 - \alpha \lambda_i| < e^{-u} \quad (23)$$

which leads to the following solutions:

$$\lambda_2 > \frac{1 - e^{-u}}{\alpha}, \quad \lambda_n < \frac{1 + e^{-u}}{\alpha} \text{ and } \frac{\lambda_2}{\lambda_n} > \frac{1 - e^{-u}}{1 + e^{-u}}.$$

For any communication network, the smaller the ratio  $\lambda_2/\lambda_n$ , the better is the convergence and synchronization.

During simulation, the saturation function  $f(x) = 2/(1 + e^{(n-x/an)})$  given in (8) was selected as a mapping function.

Fig. 6 indicates that all nodes will reach the saturation state. As the exchange of target list occurs between nodes, each node will update its target scope. The update stops when all targets are visible to a node. Fig. 7 shows that using 4000 iterations, the increase of the number of nodes generates a higher negative Lyapunov exponent. A higher negative value is a sign of greater system stability. Fig. 7 reveals also, that the mapping function used is conducive to a stability point for the communication system. The theoretical ratio  $\lambda_2/\lambda_n$  (synch-ratio) obtained from (23) defines a lower bound for the communication system stability region and informs about the shape and behavior of the synchronization curve. The system will synchronize and stabilize as long as its calculated ratio stays above the theoretical bound and follows its shape.

In general, by saturating the area with a higher number of nodes, the number of clusters tends to get smaller. More precisely, increasing the number of nodes leads to an increase, followed by a decrease of the number of clusters [1]. This is illustrated in Fig. 8, where 20 nodes form one cluster. Moreover,

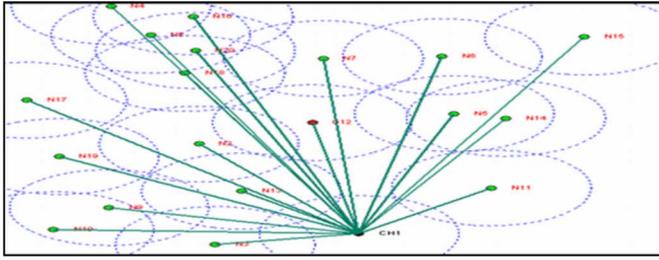


Fig. 8. Formation of one cluster due to a higher number of nodes.

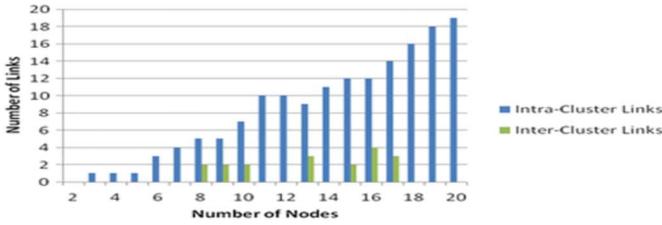


Fig. 9. Number nodes in the network versus density of intracluster and inter-cluster links.

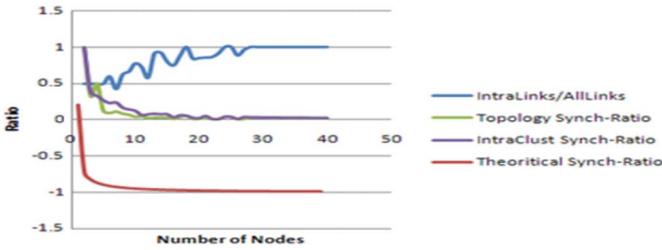


Fig. 10. Network and intracluster synchronization ratio when the number of nodes increases.

the clustering technique is designed to favor dense intracluster links over sparse intercluster links when the number of nodes increases. Fig. 9 shows that the ratio of intracluster links to intercluster links increases with an increase of number of nodes. This result forms the basis in the study of the importance of intra and intercluster links in the convergence process.

Series of communication topologies were generated by increasing the number of nodes. For each topology, a number of measures were evaluated: The ratio of intracluster links over the total number of links, intraclusters synchronization ratio [(15)], interclusters synchronization ratio [(16)], and the overall system synchronization ratio.

In Fig. 10, the calculated topology and intracluster synchronization ratios are above the theoretical boundary and follow its trend. This confirms that the generated topologies will converge to a stable point. As the number of node increases, the system synchronization is dictated by intracluster synchronization. The intracluster convergence curve merges with the overall system synchronization curve.

In contrast to intracluster synchronization, Fig. 11 shows that smaller intercluster synchronization ratios are obtained for smaller number of intercluster links. In other words, a better synchronization is obtained when less intercluster links are present.

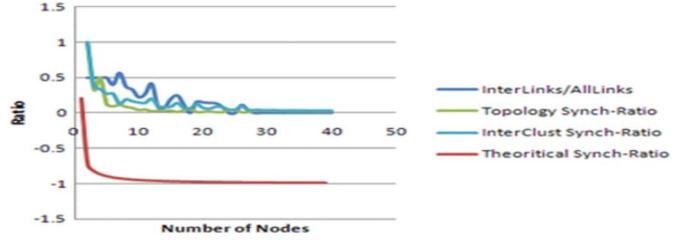


Fig. 11. System and intercluster synchronization ratio when the number of nodes increases.

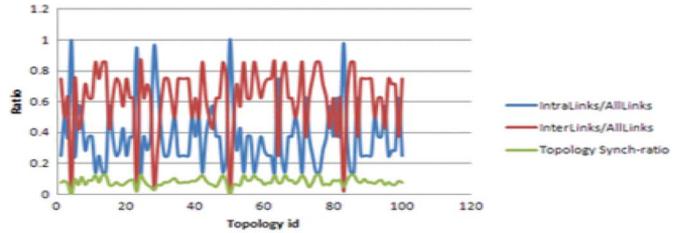


Fig. 12. Variation of topology synchronization rate when intra and intercluster links vary.

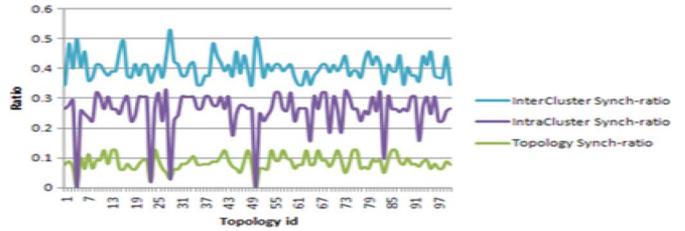


Fig. 13. Variation of topology synchronization rate when intra and intercluster synchronization rates vary.

A better visualization of the impact of intra and intercluster links on the convergence of a topology is illustrated in Fig. 12. Better system synchronization (low value) is obtained when the ratio of intralinks is higher. Bad system synchronization (high value) is obtained when the ratio of interlinks is higher. Similarly, in Fig. 13, a better system synchronization is obtained when intracluster synchronization dominates intercluster synchronization. Low system synchronization ratios are obtained for low intracluster synchronization ratios.

## VII. CONCLUSION AND DISCUSSION

This paper investigated the effect of few topology structural properties on the consensus building among nodes. Six metrics, algebraic connectivity ratio, average path, average cluster coefficient, average matching index, modularity and average participation, were investigated using factor analysis technique. Only algebraic connectivity ratio, average matching index, modularity and average participation proved to have a major impact on the consensus building. The regression analysis proved that up to 50% in consensus variability could be explained by these four parameters. It also showed that the consensus building could be predicted and that the metrics describing the intracluster connection have more weight in the model. This role was confirmed by the theory part. Better system synchronization is conditioned by good intracluster

synchronization. This indicates that a good communication topology has more intracluster connections and less intercluster connections. It also suggests that achieving better synchronization requires topology creation techniques that focus on generating dense clusters.

In this study, only six parameters were investigated resulting in a linear model and 50% explained variance. This result could be improved if more structural and nonstructural metrics were incorporated. Furthermore, a model describing the consensus building for any given topology with a random number of nodes could be devised.

## REFERENCES

- [1] H. Hatime, K. Namuduri, and J. M. Watkins, "OCTOPUS: An on-demand communication topology updating strategy for mobile sensor networks," *IEEE Sensors J.*, vol. 11, no. 4, pp. 1004–1012, Apr. 2011.
- [2] A. Reka and A. L. Barabasi, "Statistical mechanics of complex networks," *Rev. Modern Phys.*, vol. 74, no. 1, pp. 47–97, Jan. 2002.
- [3] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topology generators: Degree-based vs. Structural," in *Proc. ACM SIGCOMM*, 2002, pp. 147–150.
- [4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, pp. 175–308, Feb. 2006.
- [5] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, pp. 167–256, 2002.
- [6] B. H. Junker and F. Schreiber, *Analysis of Biological Networks*. New York: Wiley, Mar. 2008.
- [7] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications*. Berlin, Germany: Springer-Verlag, 2001.
- [8] A. Li and S. Horvath, "Network neighborhood analysis with the multi-node topological overlap measure," *Bioinformatics*, Nov. 2006.
- [9] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 23, no. 2, pp. 22–231, Feb. 2005.
- [10] A. Krishnan, A. Giuliani, J. P. Zbilut, and M. Tumita, "Implications from a network-based topological analysis of ubiquitin unfolding simulations," *Plos One*, vol. 3, no. 5, p. e2149, May 2008.
- [11] S. Kar and J. M. F. Moura, "Topology for global average consensus," in *Proc. 40th Asilomar Conf. Signals, Syst. Comput.*, Oct. 2006, pp. 276–280.
- [12] A. Jamakovic and S. Uhlig, "On the relationship between the algebraic connectivity and graph's robustness to node and link failures," in *Proc. 3rd EuroNGI Conf. Next Generation Internet Networks*, May 2007, pp. 96–102.
- [13] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, pp. 1–15, 2004.
- [14] E. Ziv, M. Middendorf, and C. H. Wiggins, "Information-theoretic approach to network modularity," *Phys. Rev. E*, vol. 71, no. 4, p. e046117, Apr. 2005.
- [15] E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, *Complex Networks*. New York: Springer, Sep. 2004.
- [16] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," in *Proc. Nat. Acad. Sci.*, Jul. 2000, vol. 97, no. 2, pp. 11149–52.
- [17] R. Reyment and K. G. Joreskog, *Applied Factor Analysis in the Natural Science*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
- [18] A. L. Comrey and H. B. Lee, *A First Course in Factor Analysis*. Mahwah, NJ: Lawrence Erlbaum, 1992.
- [19] D. F. Polit and C. T. Beck, *Nursing Research: Principles and Methods*. Baltimore, MD: Williams & Wilkins, Jun. 2000.
- [20] R. L. Gorsuch, *Factor Analysis*. Mahwah, NJ: Lawrence Erlbaum, Nov. 1983.
- [21] A. B. Costello and J. W. Osborne, "Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis, practical assessment," *Res. Evaluation*, vol. 10, no. 7, pp. 1–9, Jul. 2005.
- [22] D. L. Rubinfield, "Reference guide on multiple regression," Reference manual on Scientific Evidence, 2000.
- [23] N. L. Leech, K. C. Barrett, and G. A. Morgan, *SPSS for Intermediate Statistics Use and Interpretation*. Mahwah, NJ: Laurence Erlbaum, Jul. 2007.
- [24] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum, 1988.
- [25] K. Kaneko, *Theory and Applications of Coupled Map Lattices*. New York: Wiley, 1993.
- [26] T. Kato, *Perturbation Theory for Linear Operators*. New York: Springer-Verlag, 1976.
- [27] F. M. Atay, T. Biyikoglu, and J. Jost, "Network synchronization: Spectral versus statistical properties," *Physica*, vol. 224, no. 1–2, pp. 35–41, 2006.
- [28] O. Edwards, *Chaos in Dynamical Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
- [29] N. Hiroyuki, *Introduction to Chaos*. Philadelphia, PA: Institute of Physics, 1999.
- [30] K. Kaneko and I. Tsuda, *Complex Systems: Chaos and Beyond*. New York: Springer, 2001.

**Hicham Hatime** received the B.S. degree in computer science from High-Tech School, Rabat, Morocco, in 1995, and the M.S. degrees in computer science and in electrical and computer engineering from Wichita State University, Wichita, KS, in 1998 and 2001 respectively. Currently, he is working towards the Ph.D. degree in electrical engineering and computer science at Wichita State University.

He was with Morocco's Real Estate Department from 1994 to 1995 and with Prestige Informatique, Morocco, from 1995 to 1996. From 1998 to 2007, he was with LSI Logic working on SCSI, fiber channels, and iSCSI protocols. His areas of research include mobile sensors networks and robotics.