

Simplicial Complexes of Networks and Their Statistical Properties

Slobodan Maletić, Milan Rajković*, and Danijela Vasiljević

Institute of Nuclear Sciences Vinča, Belgrade, Serbia

*milanr@vin.bg.ac.yu

Abstract. Topological, algebraic and combinatorial properties of simplicial complexes which are constructed from networks (graphs) are examined from the statistical point of view. We show that basic statistical features of scale free networks are preserved by topological invariants of simplicial complexes and similarly statistical properties pertaining to topological invariants of other types of networks are preserved as well. Implications and advantages of such an approach to various research areas involving network concepts are discussed.

Keywords: Networks, statistical mechanics, complex systems, topology, simplicial complexes, homology, betti numbers.

1 Topological Features of Simplicial Complexes

In this section we give a short introduction to the subject of simplicial complexes and related topological terminology [8].

Any subset of V , $\{v_{\alpha_0}, v_{\alpha_1}, \dots, v_{\alpha_n}\}$ determines an n -simplex denoted by $\langle v_{\alpha_0}, v_{\alpha_1}, \dots, v_{\alpha_n} \rangle$. The elements v_{α_i} of V are the vertices of the simplex denoted by $\langle v_{\alpha_i} \rangle$, and n is the dimension of the simplex. Any set of simplices with vertices in V is called a simplicial family and its dimension is the largest dimension of its simplices. A q -simplex σ_q is a q -face of an n -simplex σ_n , denoted by $\sigma_q \lesssim \sigma_n$, if every vertex of σ_q is also a vertex of σ_n . A simplicial complex represents a collection of simplices. More formally, a simplicial complex K on a finite set $V = \{v_1, \dots, v_n\}$ of vertices is a nonempty subset of the power set of V , so that the simplicial complex K is closed under the formation of subsets. Hence, if $\sigma \in K$ and $\rho \in \sigma$, then $\rho \in K$.

Two simplices σ and ρ are q -connected if there is a sequence of simplices $\sigma, \sigma_1, \sigma_2, \dots, \sigma_n, \rho$, such that any two consecutive ones share a q -face, implying that they have at least $q+1$ vertices in common. Such a chain is called a q -chain. The complex K is q -connected if any two simplices in K of dimensionality greater or equal to q are q -connected. The dimension of a simplex σ is equal to the number of vertices defining it minus one. The dimension of the simplicial complex K is the maximum of the dimensions of the simplices comprising K . In Fig. 1 we show an example of a simplicial complex and its matrix representation.

In this example $V = \{1, 2, \dots, 11\}$, and the simplicial complex K consists of the subsets $\{1, 2, 3, 4, 5\}$, $\{2, 3, 6, 7\}$, $\{6, 7, 8, 9\}$ and $\{10, 11\}$. Its dimension is

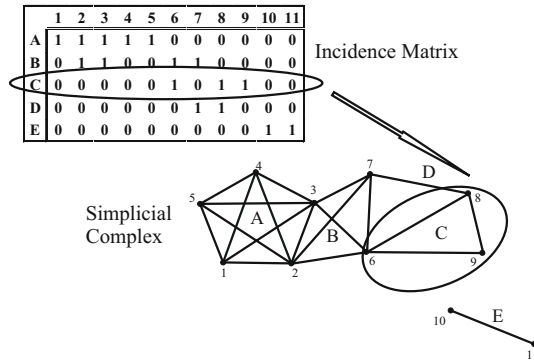


Fig. 1. An example of a simplicial complex and its representation

4, as there is a 4-dimensional simplex, in addition to two 3-dimensional ones attached to it and one 1-dimensional simplex. A convenient way to represent a simplicial complex is via a so called incidence matrix, whose columns are labeled by its vertices and whose rows are labeled by its simplices, as shown also in Fig. 1. The multifaceted property (algebraic, topological and combinatorial) of simplicial complexes makes them particularly convenient for modelling complex structures and connectedness between different substructures.

1.1 The Method

Our approach is to encode a network into a simplicial complex, construct vector valued quantities representing topological or algebraic invariants and examine statistical properties of vector valued measures. More precisely, we inspect the distribution of vector components as a function of dimension and compare these distributions for different types of networks which should reflect the basic statistical properties of the networks under study (scale-free, random etc.). Furthermore, by extracting topological properties of substructures the characterization of networks may go beyond the degree distribution, and give insight into higher levels of connectedness of networks. This implies that some topological properties of networks may be distinct even if they have the same degree distribution.

1.2 Construction of Simplicial Complexes from Graphs

Simplicial complexes may be constructed from directed graphs (digraphs) in several different ways. Here we only consider construction of the so called neighborhood complex $N(G)$ from the graph G , with vertices $\{v_1, \dots, v_n\}$. For each vertex v of G there is a simplex containing the vertex v , along with all vertices w corresponding to directed edges $v \rightarrow w$. The neighborhood complex is obtained by including all faces of those simplices and in terms of matrix representation, the incidence matrix is obtained from the adjacency matrix of G by increasing all diagonal entries by 1.

The second method associated to digraphs (or undirected graphs) has the complete subgraphs as simplices. The complete subgraph complex $C(G)$ has the vertices of G as its vertices, while the maximal simplices are given by the collection of vertices that make up the cliques of G .

Naturally, these two methods are not the only ones that may be used for constructing simplicial complexes from graphs. Actually, any property of the graph G that is preserved under deletion of vertices or edges may be used for construction purposes. A detailed account of the methods for obtaining simplicial complexes from graphs may be found in [9].

1.3 Invariants of Simplicial Complexes

Simplicial complexes may be considered from three different aspects: (1) a combinatorial model of a topological space; (2) a combinatorial object; (3) an algebraic model. Consequently, the invariants of simplicial complexes may be defined based on their different aspects and each aspect provides completely different measures of the complex and, by extension, of the graph from which the complex was constructed. In the first case various algebraic topological measures may be associated such as homotopy and homology groups [6]. In the second case several invariants may be defined and numerically evaluated. The first is the dimension of the complex. The next one is the so called an f -vector (also known as the *second structure vector*) [2], [3], [4], [5], which is an integer vector with $\dim(K) + 1$ components, the i -th one being equal to the number of i -dimensional simplices in K . An invariant is also a Q -vector (*first structure vector*), an integer vector of the same length as the f -vector, whose i -th component is equal to the number of i -connectivity classes. The structure vector, illustrated in Fig. 2, provides information about connected components at each level of connectivity with initial level equal to the dimension of the complex. In terms of the topological framework of the Q-analysis, we denote the q -level at which the simplex first appears its top q (symbol \hat{q}), and the q -level at which it first joins another simplex in a local component its bottom q (symbol \check{q}). These symbols are then used to define *eccentricity* of a particular simplex as

$$ecc = \frac{\hat{q} - \check{q}}{\hat{q} + 1},$$

so that this expression scales eccentricity from 0 to 1 [7]. Eccentricity quantifies the way the simplex is integrated into the complex such that high values reflect low levels of integrity, while low values result from high integrity levels. Also important is the so called *vertex significance* defined as

$$\rho = \frac{\Delta_i}{\max(\Delta_k)},$$

where Δ_i represents the sum of the node weights which the simplex shares with other simplices, $\max(\Delta_k)$ represents the maximal value of Δ_i and the node weight is the number of simplices formed by that node.

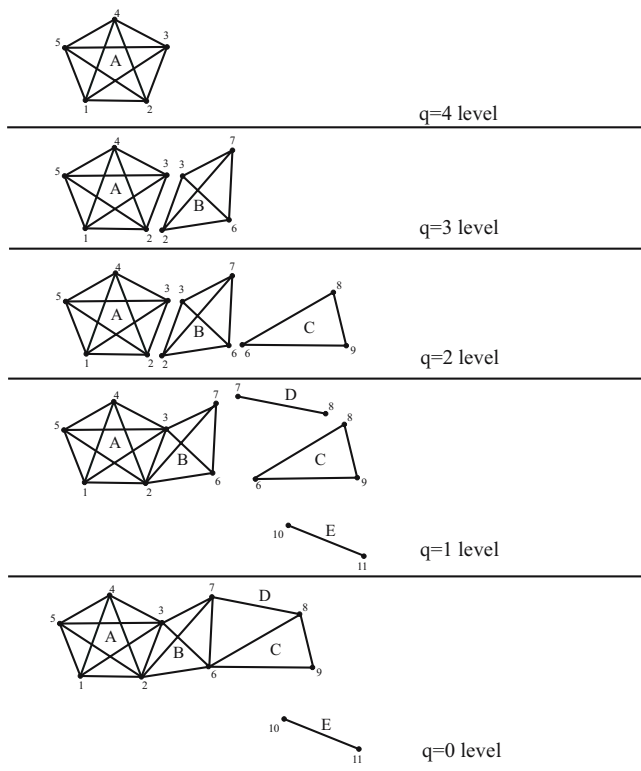


Fig. 2. Structure vector (Q-vector) disclosing connectivity of the complex at various levels

1.4 The Betti Numbers

Betti numbers may be associated to the simplicial complex when the aspect of abstract algebra is applied. In simple terms the Betti numbers are a topological invariant allowing to measure either the number of holes (simplices representing holes) of various dimensions present in a simplicial complex, or equivalently, the number of times the simplex loops back upon itself. Hence Betti numbers form an integer vector where each component corresponds to a distinct dimension.

1.5 Clustering Coefficients of Simplicial Complexes

In analogy to networks [1], one may define the clustering coefficients of simplices. We can define this variable between the reference simplex and its neighbors in the following way. Suppose we have two simplices σ^i and σ^j which share face of dimension f_{ij} . The dimensions of the simplices σ^i and σ^j are q_i and q_j ,

respectively. Since they share face whose dimension is f_{ij} , they also share faces with dimensions $f_{ij} - 1, f_{ij} - 2, \dots, 0$. The clustering coefficient is defined as

$$C'_{ij} = \frac{\text{Number of shared faces}}{(1/2)(\text{Overall number of faces that each of simplices can have})},$$

or

$$C'_{ij} = \frac{\sum_{f=0}^{f_{ij}} \alpha_{f_{ij},f}}{\frac{1}{2} \left(\sum_{f=0}^{q_i} \alpha_{q_i,f} + \sum_{f=0}^{q_j} \alpha_{q_j,f} \right)}, \tag{1}$$

where $\alpha_{x,f} = \frac{(x+1)!}{(f+1)!(x-f)!}$ and $x = f_{ij}, q_i, q_j$. If we sum over all neighbors of the simplex, we get the variable which characterizes the simplex i.e. $C'_i = \sum_{j \in nn} C'_{ij}$ where nn stands for nearest-neighbor.

Simplifying equation (1) one gets

$$C'_{ij} = \frac{2(-1 + 2^{1+f_{ij}})(1 + f_{ij})!}{\Gamma(2 + f_{ij}) \left(\frac{(-1+2^{1+q_i})(1+q_i)!}{\Gamma(2+q_i)} + \frac{(-1+2^{1+q_j})(1+q_j)!}{\Gamma(2+q_j)} \right)}, \tag{2}$$

where $\Gamma(z)$ is gamma-function. We can further simplify equation (2) and get

$$C'_{ij} = \frac{2^{1+f_{ij}} - 1}{2^{q_i} + 2^{q_j} - 1}, \tag{3}$$

and for the simplex

$$C'_i = \sum_{j \in nn} \frac{2^{1+f_{ij}} - 1}{2^{q_i} + 2^{q_j} - 1}. \tag{4}$$

2 Random (Erdős-Rényi) Network

The network constructed in order to illustrate the concepts mentioned previously consists of 2000 nodes with the probability of two nodes having a link equal to $p = 0.05$.

As is well known, a random network has a characteristic scale in its node connectivity reflected by the peak of the distribution which corresponds to the number of nodes with the average number of links. A careful reader has certainly noticed that the number of links connecting a specific node in a network corresponds to the dimension of the encoded simplex. Hence, distribution with respect to the dimension of simplices, shown in Fig 3 (left), is equivalent to the degree distribution of random networks and follows a bell-shaped curve. The distribution of vector valued measures is illustrated by distributions of the first and second structure vectors (Q -vector and f -vector), shown in Fig. 3 (right). The distribution in each case assumes dimension dependence.

As in the dimension (network degree) distribution, characteristic dimension of a simplicial complex may be noticed, and both vectors have a similar Poissonian-like shape, as expected for a random network.

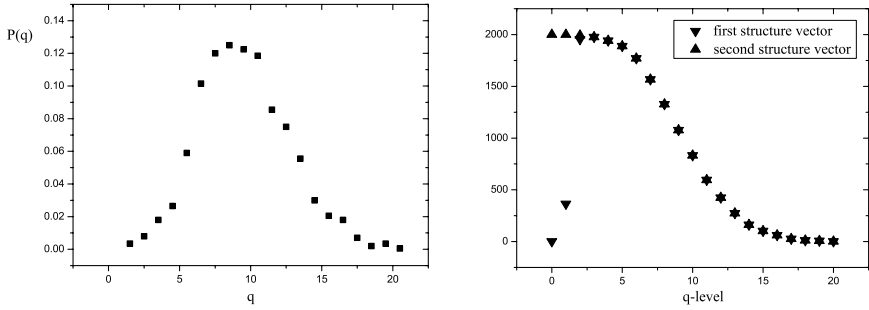


Fig. 3. Dimension distribution of a random simplicial complex (left); Q-vector and f-vector as a function of dimension (right)

3 A Network with Exponential Degree Distribution

A US power grid network of the western United States consists of 4941 nodes [11]. As an illustration of vector-valued network measures and topological invariants of the corresponding simplicial complexes we present in the Fig. 4 the number of simplices of a given dimension as a function of dimension, first and second structure vectors. The distribution of vertex significance also follows an exponential dependence (not shown).

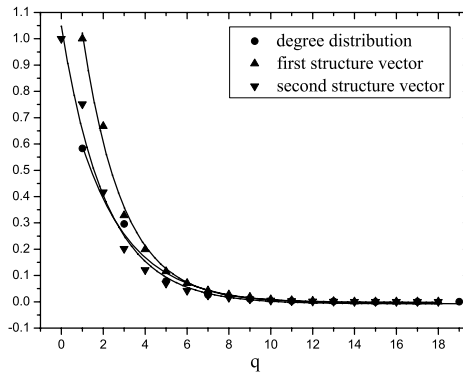


Fig. 4. Degree distribution and distributions of first and second structure vector for the US power grid network

4 Scale-Free Network

The absence of a peak in a power-law degree distribution of scale-free networks implies absence of characteristic scale (characteristic node). This property is

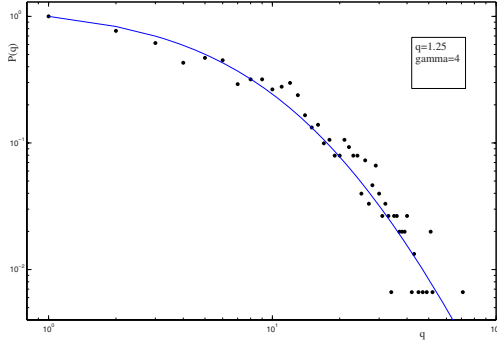


Fig. 5. Distribution of simplex dimensions. Fitting parameters to the q -exponential function are shown in the inset.

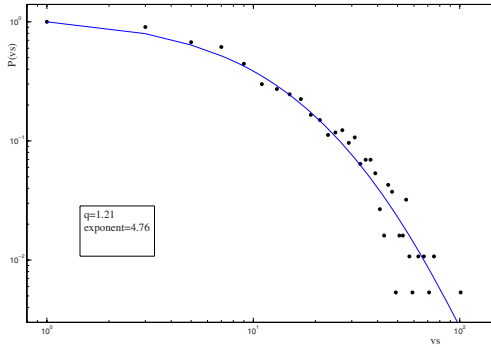


Fig. 6. Distribution of vertex significance with respect to dimension. The parameters of the fit to the q -exponential function are in the inset.

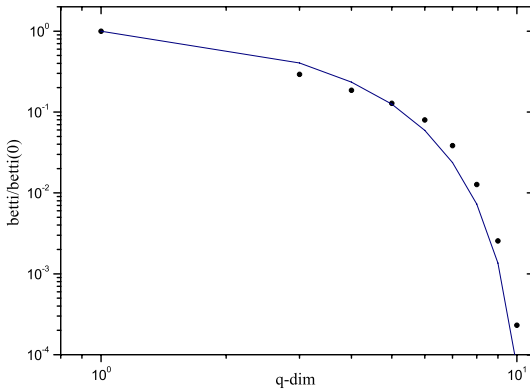


Fig. 7. Distribution of normalized Betti numbers

reflected in the distribution of dimensions of simplices forming the simplicial complex obtained from the scale-free network (Fig. 5). The distribution fits very well to the q -exponential function, with parameters shown in the inset ($q = 1.25$, $\gamma = 4$). The distribution of the vertex significance along with the q -exponential fit is presented in Fig. 6, while in Fig. 7 the distribution of the Betti numbers exhibits very good fit to the q -exponential function.

5 Final Remarks

Based on the results pertaining to random and scale-free networks, and to networks showing exponential degree distribution, it is clear that simplicial complexes encoded from these networks generate remarkable properties characterizing their topological, algebraic and combinatorial features. Based on this short exposition it is clear that conventional network approaches lack the diversity and abundance of information offered by the simplicial complex approach. Moreover, advanced algebraic topology methods enable dynamic analysis of simplicial complexes along with the dynamic updating of topological properties, a topic to be discussed elsewhere [10]. Route-link structure, so important in many applications of network models, may obtain more suitable description using the simplicial complex approach while it is evident that this method may be able to address a wider class of problems than network theory. In a straightforward manner traffic (in the most general sense) may be developed for these structures so that a theory and modeling of multidimensional traffic on a multidimensional topological background may be realized.

References

1. Albert, R., Barabasi, A.L.: Statistical mechanics of complex networks. *Reviews of Modern Physics* 74, 47–97 (2002)
2. Atkin, R.: From cohomology in physics to q -connectivity in social sciences. *Int. J. Man-Machines Studies* 4, 341–362 (1972)
3. Atkin, R.: *Mathematical Structure in Human Affairs*, London, Heinemann (1974)
4. Atkin, R.: An Algebra of Patterns on a Complex I. *Int. J. Man-Machine Studies* 6, 285–307 (1974)
5. Atkin, R.: An Algebra of Patterns on a Complex II. *Int. J. Man-Machine Studies* 8, 483–498 (1974)
6. Barcelo, H., Kramer, X., Laubenbacher, R., Weaver, C.: Foundations of connectivity theory for simplicial complexes. *Adv. Appl. Math.* 26, 97–128 (2001)
7. Gould, P., Johnson, J., Chapman, G.: *The Structure of Television*. Pion Limited, London (1984)
8. Hatcher, A.: *Algebraic Topology*. Cambridge University Press, Cambridge (2002)
9. Jonsson, J.: *Simplicial Complexes from Graphs*. Lecture Notes in Mathematics. Springer, Heidelberg (2008)
10. Vasiljević, D., Maletić, S., Rajković, M.: (in preparation)
11. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. *Nature* 393, 440–442 (1998)