# DESCRIPTION AND STATISTICAL ANALYSIS OF THE SOD1 NETWORK AND ITS CORRELATION WITH ITS GENETIC PROPERTIES

**Morán Titla C.D., Atenco Rosas J.E., Robles E.V., Fortiz R.M., Morales G.C., Hernández Santiago A.A., Arzola Flores J.A.**

Meritorious Autonomous University of Puebla

*av. San Claudio, S/N Col. San Manuel, Puebla, Pue. C.P. 71590, Mexico; e-mail: ximikad09@mail.ru*

Received: 10.07.2019

**Abstract.** In living organisms are produced reactive oxygen species (ROS) which are oxidizing elements of cells. They are commonly produced because of cellular metabolism, as well as the consequence of toxic agents, such as tobacco smoke, ionizing radiation and carcinogens, to name a few. The negative effect of ROS on the biochemical pathways of cells is called oxidative stress. The methodology used in this study could guide the study of different genetic regulatory networks, in order to identify possible therapeutic targets. The co-expressed genetic complex network of SOD1 in Homo sapiens was obtained from the database Functional protein association networks, which is made up of the following genes: ATP5H, ATP5O, PSMB3, PSMB1, PSMA5, PSMB4, PSMB6, ATP5D, SOD1, MDH1 and HSD17B10. For each gene in the network the following functional properties were obtained: central score (hub), authority score, and grade, clustering coefficient and betweenness centrality. The properties were calculated using two types of network, directed and non-directed. When carrying out the study of the genetic network of SOD1 using the Theory of complex networks, it was shown that possibly its functional properties of the SOD1 gene are not completely related to its structural chemical properties, that is, the function of the SOD1 gene within the Genetic network should be studied from a holistic approach and not from a reductionist perspective.

***Key words:*** *genetic complex network, statistical analysis.*

## INTRODUCTION

In living organisms are produced reactive oxygen species (ROS) which are oxidizing elements of cells. They are commonly produced because of cellular metabolism, as well as the consequence of toxic agents, such as tobacco smoke, ionizing radiation and carcinogens, to name a few. These elements can become harmful depending on the concentration and the biochemical process they trigger [1]. ROS are extremely reactive molecules, have a limited time of existence, capable of oxidative modifying any biomolecule, which results in a series of chain reactions causing damage. In pathological conditions or not, when the quantity of these molecules increases, the essential macromolecules can be modified oxidative and consequently affect the signaling pathways that are controlled by the cellular redox state. The negative effect of ROS on the biochemical pathways of cells is called oxidative stress [2]. The organisms exposed to an atmosphere with a large amount of oxygen, developed macromolecules of enzymatic origin, able to regulate the amount of ROS and inhibit their reaction with biological molecules [3]. The components of the antioxidant system of the cells are diverse, it can be mentioned the primary antioxidants that eliminate ROS or prevent the formation of these molecules. The primary function of primary antioxidants is to convert ROS into less reactive species. For example, some enzymes that are part of the primary antioxidants are the enzymes of the superoxide dismutase (SOD) family. SOD1 is a soluble protein of 32 kDa, discovered for the first time in 1969 by Mc Cord and Fridovich [4]. It can be in the cytosol, in the nucleus and in the mitochondrial intermembrane space. Immunocytochemical studies have shown that it can also be found in lysosomes and peroxisomes [5]. The importance of this enzyme is such that in genetic diseases such as Down syndrome (DS) and the familial form of amyotrophic lateral sclerosis (ALS) have been described alterations in the functions of one of the components of the family of SOD, in Particularly Cu-Zn cytosolic superoxide dismutase (Cu / Zn-SOD1 or SOD1), so that currently many investigations have been developed with the aim of deepening the role of this antioxidant enzyme in the pathogenesis of these genetic diseases [4]. An alternative to study the importance of these enzymes at the cellular level, is by analyzing the interactions between enzymes, proteins and genes from the computational context, analyzing the networks of chemical interactions, which can be modeled using the Theory of complex networks. With this approach, the chemical species participating in a series of chemical reactions can be modeled as nodes and the connections between them can represent physical, chemical or functional interactions between the biomolecules [6].You can compare networks that are derived from a macromolecule, that is, modeling changes in the topology of the network, for example, removing nodes or modifying the direction of the information flow (directed or unguided network).

Because of its importance in pathologies and its function as an antioxidant, the SOD1 enzyme has been studied in different aspects, from the thermodynamic aspect at the molecular level [7], its relation with diseases of the nervous system [9] and also in studies at the cellular level [8]. However, from within the context of the Theory of complex networks, until now an analysis for the gene that encodes the enzyme has not been implemented. Therefore, the objective of this paper is to statistically describe and analyze the SOD1 genetic network and correlate its functional properties with its genetic properties, in order to study whether the functional properties of a gene are actually related to its structural chemical properties.

## METHODOLOGY

**Construction of a complex SOD1 network.** The co-expressed genetic complex network of SOD1 in Homo sapiens was obtained from the database Functional protein association networks, which is made up of the following genes: ATP5H, ATP5O, PSMB3, PSMB1, PSMA5, PSMB4, PSMB6, ATP5D, SOD1, MDH1 and HSD17B10. For each gene in the network the following functional properties were obtained: central score (hub), authority score, and grade, clustering coefficient and betweenness centrality. The properties were calculated using two types of network, directed and non-directed.

Subsequently, to encompass all the properties of the genes and obtain a representative value, the model of Bickerton et al. [10], which consists of an attractiveness index that allows identifying the most important nodes in the network by weighing on all the structural properties of the network, the model is as follows:

$$Index = exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln d_i\right),$$

where $n$ is the number of structural properties of the network, $d_i$ corresponds to each of the statistical properties of the network for each i-th node.

**Obtaining the genetic properties of the SOD1 network.** The genetic properties of the SOD1 network of the GeneCard database were obtained; the obtained properties are the following: gene of origin, beginning of the sequence in pb (pairs base), term of the sequence (pb), size of the gene (pb), gift and score. On the other hand, the number of nucleotides in each gene of the National Center for Biotechnology Information database was obtained, the nucleotides used were: Adenine (A), Thymine (T), cytosine (C) and Guanine (G). He again evaluated the attractiveness index.

**Statistical analysis.** The statistical analysis was carried out using the programming language R V3.6.0. To obtain the construction and modification of the network, the following libraries were used: dplyr, tidyr, data.table, igraph, bipartite, network and textclean. Because not all the genetic and functional properties of the genes had normal distribution errors, nonparametric statistical tests and a generalized linear model were used. To compare the index between the two types of SOD1 network (directed and non-directed network), a wilcoxon analysis was performed for dependent groups. To compare the values of the index by removing the most important nodes both in the directed network and in the non-directed network, a generalized linear model of two factors with Gamma error was performed for independent groups, since the data had a high coefficient of variation, on the other hand, to evaluate the devolution, the car package was used. To study the similarity between the functional properties of the SOD1 networks (directed and non-directed, without modifications) and the genetic properties, a clustering with Euclidean scale was generated and of the average type with the vegan library. To correlate the attractiveness indexes obtained from the matrices with functional properties of the SOD1 networks (directed and non-directed without modification) and the genetic properties, a multiple Spearman correlation was performed using the libraries: Hmisc and corrplot. An alpha value of 0.05 was considered.

## RESULTS

**Analysis of the complex network of SOD1.** Figures 1 and 2 show the value of the attractiveness index for each of the nodes of the directed and non-directed networks, showing that in both cases the node with the highest value of the index in SOD1, because it shows a greater degree of functional importance in both networks. On the other hand, the average value of the attractiveness index for the directed network is 1.39 ± 0.03 and the median value was 1.40. The non-directed network had on average a value of 1.41 ± 0.04 and a median with a value of 1.39, no statistical differences were found between the networks (fig. 3). The genes with the highest value of the attractiveness index in the directed and non-directed network were SOD1 and PSMB4 (directed: 1.71 and 1.44, non-directed: 1.84 and 1.48, respectively), which confirms the importance of the SOD1 gene in the network. By removing both genes from the networks, their structure is modified (fig. 4 and 5), because the SOD1 gene functions as a link between two groups of nodes in the network, showing the formation of two modules in the network. On the other hand, the value of the index also changes, due to the removal of the nodes with the highest value of the attractiveness index (tab. 1). Marginally significant differences were found in the value of the indexes when removing nodes (tab. 2 and fig. 6).

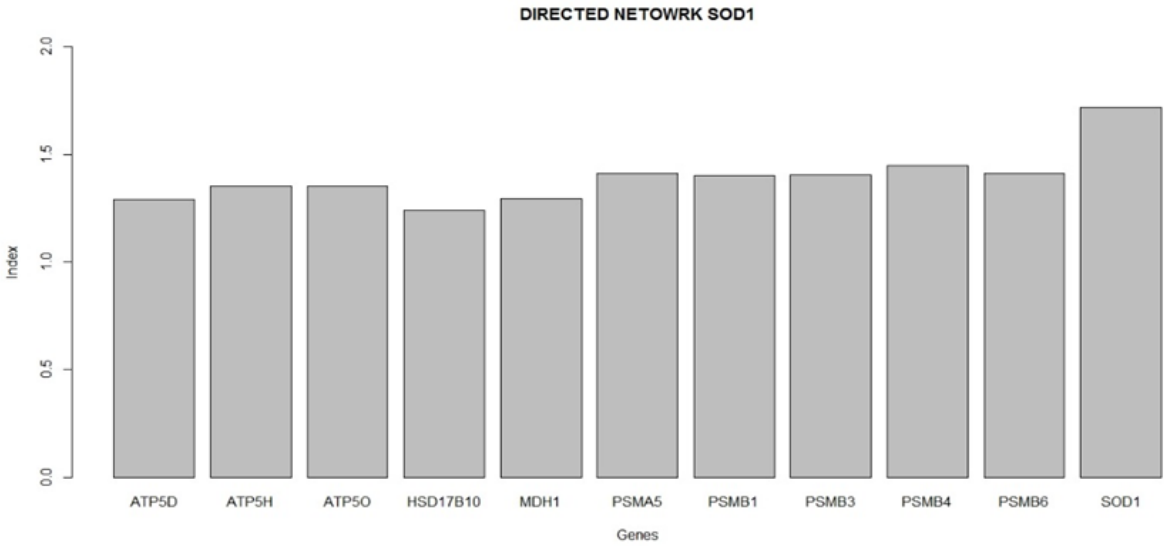


**Figure 1.** Values of the attractiveness index of the nodes (genes) in the directed SOD1 network. The highest values are observed: SOD1 and PSMB4 and the lowest value: HSD17B10 (1.24)

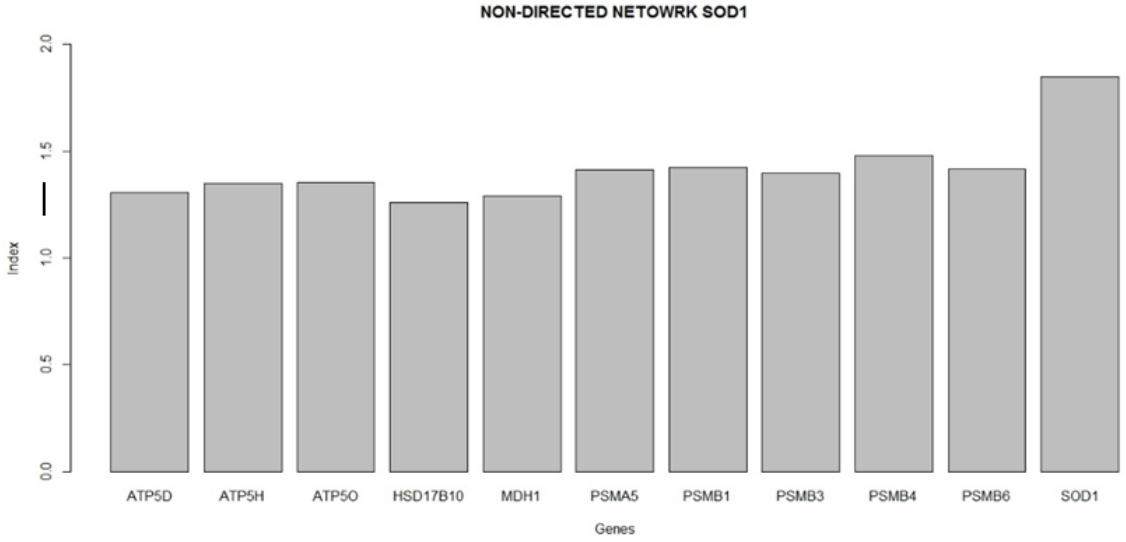


**Figure 2.** Values of the attractiveness index of the nodes (genes) in the directed SOD1 network. The highest values are observed: SOD1 and PSMB4 and the lowest value: HSD17B10 (1.25)

**Table1.** The descriptive statistics are shown when modifying the SOD1 network. The highest values are observed in the networks without changes (original). The lowest values in both networks are observed when removing the nodes SOD1 and PSMB4

| **Network type** | **Directed** | | | **Non-directed** | | |
|---|---|---|---|---|---|---|
| **Chance** | Nodes | Mean ± SE | Median | Nodes | Mean ± SE | Median |
| **Original** | 11 | 1.39 ± 0.03 | 1.40 | 11 | 1.41 ± 0.04 | 1.39 |
| **SOD1** | 10 | 1.30 ± 0.03 | 1.32 | 10 | 1.32 ± 0.04 | 1.33 |
| **PSMB4** | 10 | 1.35 ± 0.04 | 1.36 | 10 | 1.38 ± 0.05 | 1.37 |
| **PSMB4 and SOD1** | 9 | 1.29 ± 0.02 | 1.30 | 9 | 1.30 ± 0.02 | 1.38 |

**Table2.** Generalized linear model devolution table with Gamma distribution error. It is observed that when removing nodes there is a tendency to find differences in the value of the index. *% = Percentage of variation explained in the model

| Deviance analysis | | | | |
|---|---|---|---|---|
| **Factor** | DF | $X^2$ | P | %* |
| **Net type** | 1 | 0.4161 | 0.51 | 0.56 |
| **Remove nodes** | 3 | 7.7208 | 0.052 | 10.40 |
| **Interaction** | 3 | 0.0081 | 0.99 | 0.01 |
| **Residuals** | 70 | 66.21 | | 89.24 |
| **Total** | 77 | 74.19 | | |

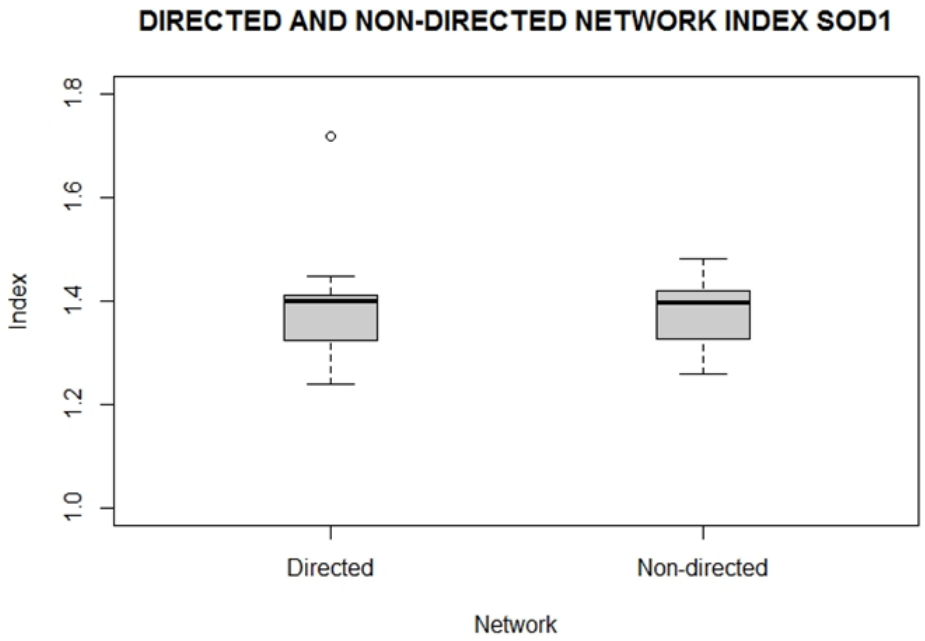


**Figure 3.** Graph showing the median with the minimum and maximum values of the attractiveness indexes in the types of the SOD1 network (directed and non-directed). No significant differences were found W = 16, P = 0.14

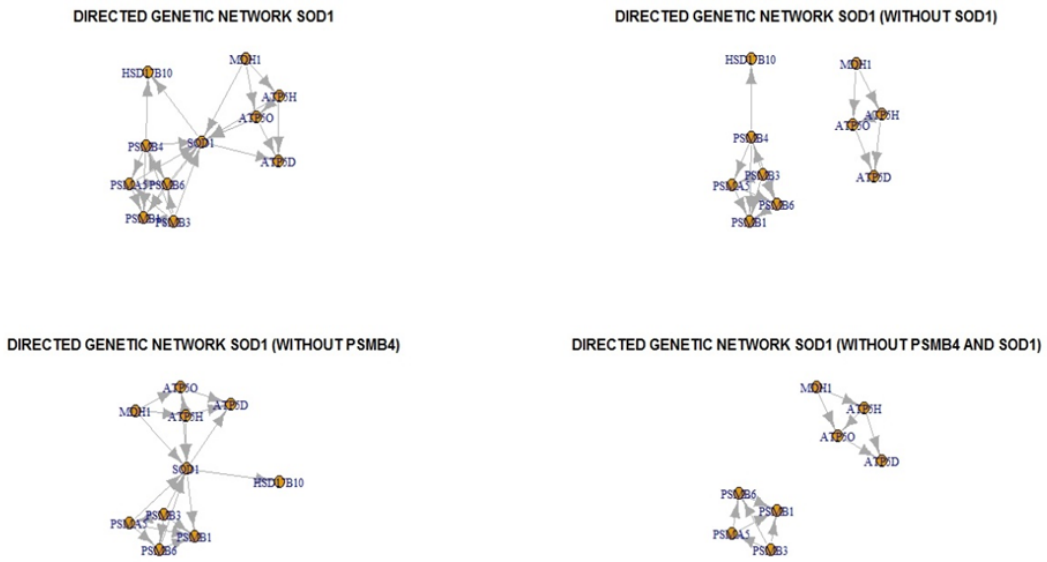


**Figure 4.** The different topologies of the SOD1 network of directed type are shown. By removing the nodes (SOD1 and PSMB4), the HSD17B10 gene node disappears

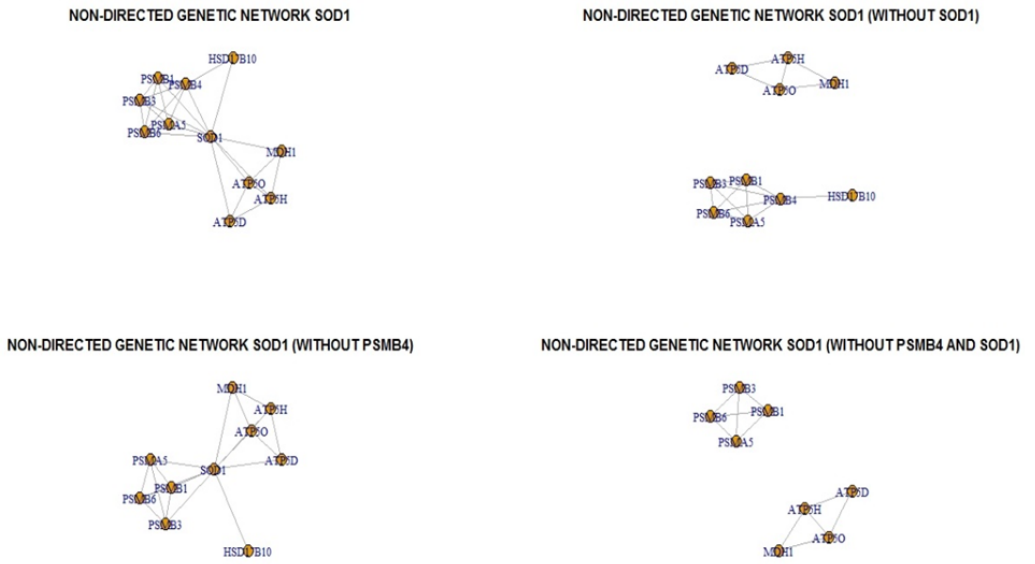


**Figure 5.** The different topologies of the SOD1 network of non-directed type are shown. By removing the nodes (SOD1 and PSMB4), the HSD17B10 gene node disappears

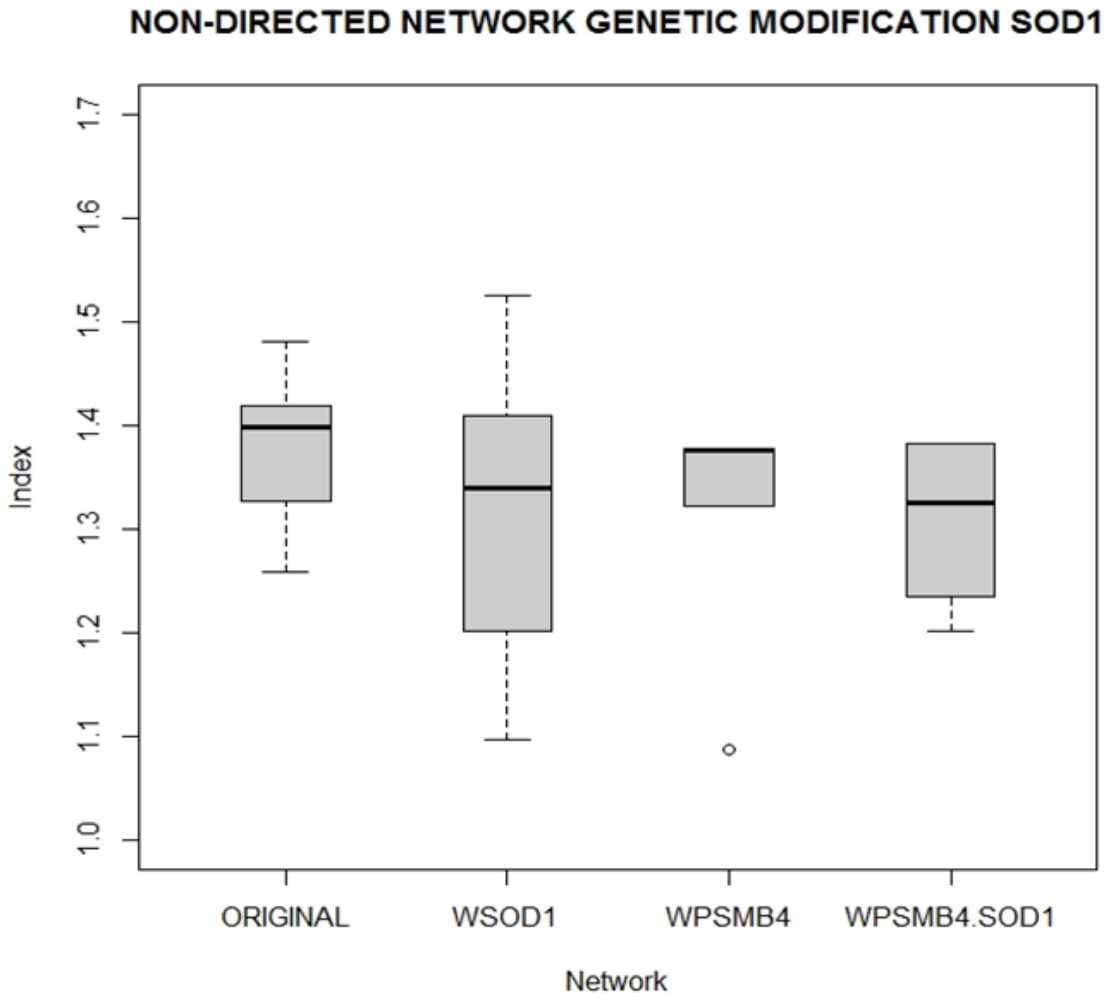


**Figure 6.** Graph showing the median with the minimum and maximum values of the attractiveness indexes when modifying the SOD1 network. The lowest values can be seen when removing SOD1 and PSMB4

**Correlation of the SOD1 network and its genetic properties.** When calculating the attractiveness index obtained from the genetic characteristics of the nodes (genes) that make up the SOD1 network, we obtained an average value of 6.18 ± 0.25 and a median of 6.35. When clustering, only a similarity pattern is observed between the directed and non-directed network (both without modifications), on the other hand, the genetic network differs from these two (fig. 7). A high correlation was found between the two SOD1 networks (directed and non-directed). However, none is correlated with the genetic properties (fig. 8), this difference is possibly due to the fact that the particular characteristics (physicochemical properties) of the genes do not reflect their functional properties within genetics, that is, it is necessary to study the genes from a holistic approach and not from the reductionist perspective.

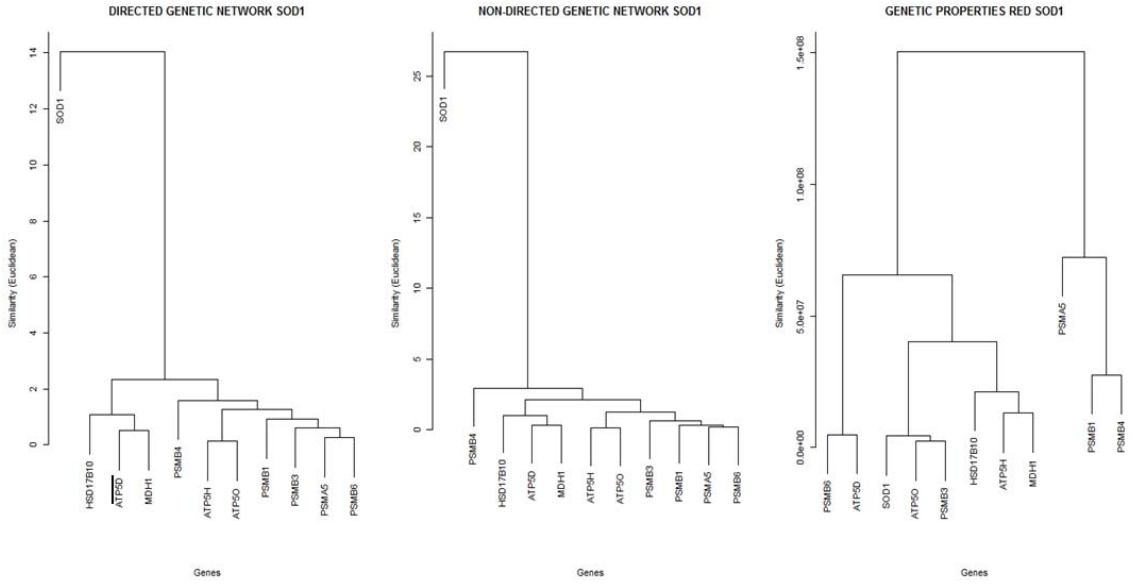


**Figure 7.** Graph showing the dendograms corresponding to the unmodified network of SOD1 (directed and non-directed) and their genetic properties. A pattern of two groups is observed in general for the dendograms

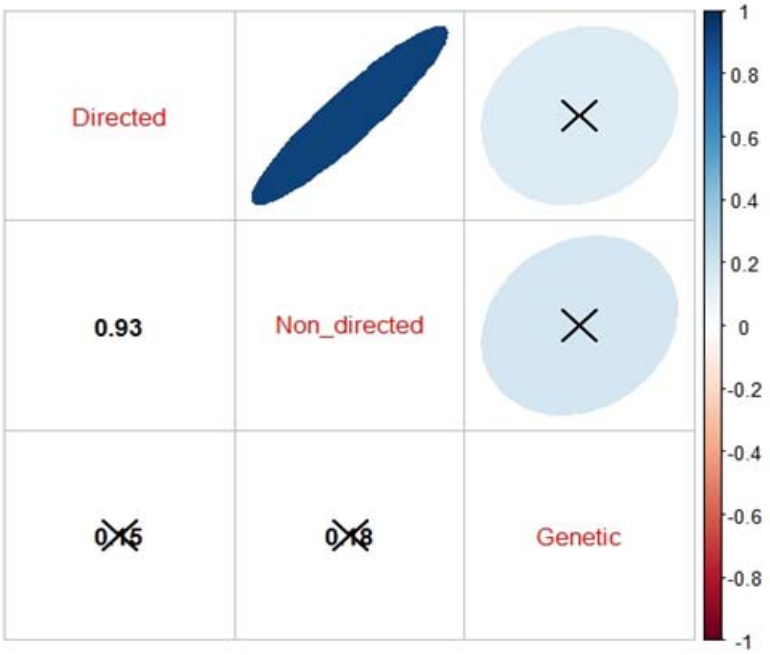


**Figure 8.** Graph showing the multiple correlation coefficients of attractiveness indices between the two SOD1 networks (directed and non-directed) and their genetic properties. SOD1 networks have a high correlation ($P < 0.001$). However, none is correlated with its genetic properties

## CONCLUSIONS

When carrying out the study of the genetic network of SOD1 using the Theory of complex networks, it was shown that possibly its functional properties of the SOD1 gene are not completely related to its structural chemical properties, that is, the function of the SOD1 gene within the Genetic network should be studied from a holistic approach and not from a reductionist perspective. The methodology used in this study could guide the study of different genetic regulatory networks, in order to identify possible therapeutic targets.

***References:***

1. Dröge W. Free Radicals in the Physiological Control of Cell Function. *Physiol. Rev*., 2002, vol.82, no. 1, pp. 47-95. DOI: 10.1152/physrev.00018.2001.

2. Jones D.P. Redefining Oxidative Stress. *Antioxidants & Redox Signaling*, 2006, vol. 8, pp. 1865-79. DOI: 10.1089/ars.2006.8.1865.

3. Rani P.K, Meena U., Karthikeyan J. Evaluation of antioxidant properties of berries. *Indian Journal of Clinical Biochemistry*, 2004, vol. 19, no. 2, pp. 103-10.

4. Castillo-Casaña Y., Riverón-Forment G. Superóxido dismutasa citosólica y enfermedades genéticas. *Rev. Cubana Genet Comunit*., 2014, vol. 8, no. 1, pp. 5-11.

5. Milani P., Gagliardi G., Cova E., Cereda C. SOD1 Transcriptional and Posttranscriptional Regulation and Its Potential Implications in ALS. *Neurology Research International*, 2011, pp. 1-9. DOI: 10.1155/2011/458427.

6. Tamari Y., Nawata H., Inoue E., Yoshimura A., Yoshii H., Kashino G., Seki M. et al. Protective roles of ascorbic acid in oxidative stress induced by depletion of superoxide dismutase in vertebrate cells. *Free Radical Research*, 2013, vol. 47, no. 1, pp. 1-7. DOI: 10.3109/10715762.2012.734916.

7. Prz˘ulj N. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 2006, pp. 177-183. DOI: 10.1093/bioinformatics/btl301.

8. Scardoni G., Petterlini M., Laudanna C. Analyzing biological network parameters with CentiScaPe. *Bioinformatics*, 2006, pp. 177-183. DOI: 10.1093/bioinformatics/btp517.

9. Andersen P.M. Amyotrophic lateral sclerosis associated with mutations in the Cu-Zn superoxide dismutase gene. *Current Neurology and Neuroscience Reports*, 2006, vol. 6, no. 1, pp. 37-46,

10. Banks C.J., Rodriguez N.W., Gashler K.R., Pandya R.R., Mortenson J.B., Whited M.D., Soderblom E.J., Thompson J.W., Moseley M.A., Reddi A.R., Tessem J.S., Torres M.P., Bikman B.T., Andersen J.L. Acylation of superoxide dismutase 1 (SOD1) at K122 governs SOD1- mediated inhibition of mitochondrial respiration. *Mol Cell Biol.*, 2017, vol. 37, pp. e00354-17. DOI: 10.1128/MCB.00354-17.