# Sensitivity of complex networks measurements

P. R. Villas Boas, F. A. Rodrigues, G. Travieso, and L. da F. Costa
*Instituto de Física de São Carlos, Universidade de São Paulo,*
*PO Box 369, 13560-970, São Carlos, SP, Brazil*

Information about real-world networks is often characterized by incompleteness and noise, which are consequences of the lack of complete data as well as artifacts in the acquisition process. Because the characterization, analysis and modeling of complex systems underlain by complex networks are critically affected by the quality of the respective structures, it becomes imperative not only to improve the quality of data, but also to devise methodologies for identifying and quantifying the effect of such sampling problems on the characterization of complex networks. In this article we report such a study, involving 10 different measurements, 4 complex networks models and 5 real world networks. We evaluate the sensitiveness of the measurements to perturbations in the topology of the network. Three particularly important types of progressive perturbations to the network are considered: edge suppression, addition and rewiring. The obtained results allowed conclusions with important practical consequences including the identification that edge removal is less critical than rewiring, followed by edge addition. The measurements allowing a better balance of stability (smaller sensitivity to perturbations) and discriminability (possibility of identification of different network topologies) were also identified. 

## I. INTRODUCTION

Complex networks theory has been largely applied to model natural and artificial systems, such as the Internet, the World Wide Web, protein interactions, airlines, roads, food webs and society [1, 2]. Its success is mainly a consequence of two recent developments: increase of computational power and availability of several databases. In the former case, computers allowed processing of networks with thousand or even million of vertices. In the latter, many maps of interactions, raging from biology to social science, have become available since the 90's. However, most of these maps are not complete and methods should be developed to treat these databases [3].

Sampling is a fundamental issue in complex network research, because incomplete data can yield biased outcomes. Indeed, the connection structures of many studied real-world networks may differ substantially from the original complex systems from which they were derived. This effect results in biased models, inaccurate characterization, classification and modeling of complex systems. Besides, since many dynamical processes, such as resilience to random and target attacks [4], spreading process [5], synchronization [1], random walk [6] and flow [7] are highly related to the structure of the network, accurate samplings become particularly critical. A way to study sampling effects is to perform perturbations on the networks through alterations of their topology and analyze their consequences.

A variety of sampling methods can be considered to map the interactions in a system into a network. The sampling issue has been recently considered in the analysis of different cross-section approaches to construct biological, information, technological and social networks. For instance, the available protein-protein interactions cover only a fraction of the complete interactome map. As a matter of fact, the high-throughput "yeast two-hybrid" assay provides high number of false positives, i.e. interactions identified in the experiment but that never take place in the cell [8, 9]. Sprinzak et at. [10] suggested that the reliability of the high-throughput Y2H is about 50%. Generally, it is assumed that the incomplete maps can be extrapolated to the complete interactome, so that limited sampling would not affect the topological structure of the network [11]. This assumption is based on the scale-free structure of protein interaction networks. However, the subnetworks obtained by sampling of scale-free networks are not guaranteed to be scale-free [3]. In addition, limited sampling can result in scale-free structures irrespective of the original network topology [12, 13]. In order to overcome these limitations, efforts have been developed to obtain more accurate databases of protein interactions [14].

In case of the World Wide Web, the network structures depend on the web crawler applied for sampling and the chosen domain [15]. Different sampling strategies can induce bias and influence in many ways the resulting recovered structure [16]. Some crawlers tends to overestimate the average number of connections of pages. A possible solution for this limitation is to start from as large a set of pages as possible [16].

Accurate topologies of the Internet are fundamental for routing strategies and to forecast its growth. Internet sampling is generally based on *tracerouters* — packets are sent through the network in order to obtain the IP address of the routers in the path. However, it is often assumed that these packets follow the shortest paths in the network [17], implying a large set of connections to be missed because of the possible presence of redundant links among routers. Moreover, in the traceroute strategy edges close to the root are more visible, i.e. the probability to obtain a edge far from the root decreases with the hierarchical level [18]. It has also been observed that the traceroute sampling of random networks leads

to networks with power-law degree distribution [18].

Social networks are also incomplete. Generally, these networks are restricted to a special class of human activity (e.g. music, sports, casting and collaborations in science) or are constructed by considering human relations (e.g. friendship and businesses). The way in which these networks are obtained can often result in biased data, such as the boundary specification problem, inaccuracy in questionnaire application and inaccessibility of subjects [19]. Moreover, depending of the considered type of personal relationship, it becomes particularly hard to define the links. Indeed, it may be difficult to estimate the effects of missing data in social networks.

In the light of the above discussion, it becomes clear that the sampling bias might induce properties not representative of the actual complex networks. In this way, it is imperative to develop accurate sampling methods in order to pave the way to sound understating of many complex systems. Some strategies have been developed to overcome the incomplete sampling problem, such as to develop remedial techniques in order to minimize the effect of missing data [15]. One of these strategies consists in identifying the influence of bias on complex networks measurements, as we propose in the current article. Measurements that are too sensitive to perturbations in the network may not be adequate to characterize incomplete or noisy networks. Moreover, measurements that do not reflect differences in network structures are of reduced value because of the implied lack of discriminability [20]. In this paper, we analyze the most traditional measurements used for networks characterization by considering three important classes of perturbations: (i) edge removal, (ii) edge inclusion, and (iii) edge rewiring. By inspecting the behavior of the measurements under these perturbations, we identify the candidate measurements most suitable for analysis and characterization of networks constructed with incomplete data or in the presence of noise. These measurements correspond to those which are not too sensitive to the perturbations while providing high differentiation among different network structures, i.e. different measurements are obtained for different topologies. We considered 10 different measurements, 4 complex networks models and 5 real world networks.

The article starts by presenting the basic concepts and methodology and proceeds by presenting the investigation of the effect of the 3 types of perturbations over theoretical and real-world networks.

## II. BASIC CONCEPTS AND METHODOLOGY

An undirected complex network (or graph) $G$ is defined as $G = (V, E)$, where $V$ is the set of $N$ nodes and $E$ is the set of $M$ undirected edges of the type $\{i, j\}$, indicating that the nodes $i$ and $j$ are connected. An undirected complex network can be represented in terms of its adjacency matrix $A$, whose elements $a_{ij}$ and $a_{ji}$ are equal to one whenever there is a connection between the vertices $i$ and $j$; and equal to 0 otherwise. Since most-real world networks are composed by thousand or even million of vertices, the analysis of their structure cannot be performed by visual inspection. In this way, a set of measurements are considered in order to describe the network topologies. These measurements can reflect different features on the network, such as connectivity, assortativity, centrality and hierarchies. In this work, we considered the following set of measurements in order to characterize the network structures [20].

- *Average degree*: the degree of a node $i$ is given by its number of connections. The resulting global measurement that quantify the density of connection in the network is given by the respective average.

- *Average degree of nearest neighbors*: The average neighbor connectivity $\langle k_{nn} \rangle$ measures the average degree of the neighbors of the the vertices in the network.

- *R square*: A metric generally considered to characterize the network structure is the degree distribution. Depending of the uniformity or inhomogeneity of connections in a network, it can present a Poisson like degree distribution or a power law, respectively. In the latter case, the distribution is characterized by a straight line in the log-log scale plot. Thus, quantifying how much a given distribution is close to a power law is a way to classify different network structures. A possible measurement to this goal is the R square, which is equal to the square of the Pearson correlation coefficient of the distribution in the log-log plot. For scale-free networks, this value should close to 1, and for random networks, smaller than 0.6 [12].

- *Assortativity*: The assortativity coefficient, represented by $r$, is the Pearson correlation coefficient between the degree of pairs of connected nodes [21]. It quantifies the connection between nodes of similar degree, indicated by positive values of this coefficient, while negative values indicate relationships between nodes of different degrees (highly connected nodes tend to be connected with few connected ones).

- *Average clustering coefficient*: The clustering coefficient of a node $i$ $cc_i$ is defined by the proportion of links between the vertices within its neighborhood, $l_i$, divided by the number of links that could possibly exist between them $(k_i(k_i - 1)/2)$. The average clustering coefficient $\langle cc \rangle$ is computed by taking the average of the clustering coefficient over the whole network.

- *Hierarchical measurements*: Hierarchical measurements are defined by considering the successive neighborhoods around each node [22**?** ]. The ring

of vertices $R_d(i)$ (or hierchical/concentric level) is formed by those vertices distant $d$ edges from the reference vertex $i$. The *hierarchical degree* at level $d$ $hk_d(i)$ is defined as the number of edges connecting the rings $R_d(i)$ and $R_{d+1}(i)$. The *hierarchical clustering coefficient* $hcc_d$ is given by the number of edges among nodes in the respective $d$-ring ($m_d(i)$), divided by the total number of possible edges between the vertices in that ring. The *convergence ratio* $hcr_d$ corresponds to the ratio between the hierarchical node degree at level $d$ and the number of vertices in the ring at level $d+1$. The reciprocal of the convergence ratio is the *divergence ratio* $hdr_d$.

- *Average shortest path length*: The average shortest path length $\ell$ is calculated by taking into account the shortest distance between each pair of vertices in the network. Disconnected nodes are not taken into account for the calculation

- *Average betweenness centrality*: The betweenness centrality of a vertex $i$ quantifies the fraction of shortest paths between each pair of nodes in the network that pass through this vertex. The average betweenness centrality $\langle B \rangle$ is computed considering the whole set of vertices in the network [20].

- *Central point dominance*: The central point dominance is defined in terms of the betweenness as $c_D = \sum_i (B_{\max} - B_i)/(N-1)$, where $B_{\max}$ represents the maximum betweenness found in the network [20].

### A. Perturbation methods

Three basic and important edge perturbation types that can be applied to a networks include:

- Edge removal: Edges are selected at random and removed from the network.

- Edge addition: Two vertices are selected at random and, if they are not connected, a connection is established.

- Edge rewiring: Two pairs of connected vertices are chosen and their connections are interchanged.

We considered perturbations ranging from 0 to 10% of the edges in the network. Perturbations involving vertices could also be considered. However, the addition of vertices should depend on the model considered to generate the network. In order to make our analysis simpler and more robust, we considered only edge perturbations. The behavior of the measurements were studied with respect to this type of perturbation.

### B. Principal component analysis

Principal component analysis (PCA) is a multivariate statistical approach used for dimensionality reduction. PCA allows to optimally reduce the dimensionality $M$ os such measurements by removing the correlations between them [20, 23, 24]. Let each perturbed network be represented by a sequence of measurements $f_v = (m_1, m_2, \ldots, m_M)$ and the complete sequence of measurements be represented by the vector $\overrightarrow{f}$. In order to perform the PCA, it is necessary to compute the covariance between each pair of measurements $i$ and $j$,

$$C(i,j) = \frac{1}{N-1} \sum_{v=1}^{N} (f_v(i) - \mu_i)(f_v(j) - \mu_j), \quad (1)$$

where $\mu_i$ is the average of $f_v(i)$ over the set of $N$ networks, $\mu_i = \sum_{v=1}^{N} f_v(i)/N$. From each combination of $i$ and $j$, it is defined the $M \times M$ dimensional covariance matrix $C$. The eigenvalues of $C$, sorted in decreasing order $(\lambda_i, i = 1 \ldots M)$, with respective eigenvectors $(\overrightarrow{v_i})$, define the matrix:

$$G = \begin{bmatrix} \uparrow & \uparrow & \uparrow \\ \vec{v_1} & \vec{v_2} & \vec{v_3} \\ \cdots & \cdots & \cdots \\ \downarrow & \downarrow & \downarrow \end{bmatrix}. \quad (2)$$

The new coordinate system is organized such that the greatest variance by any projection of the data appears along the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

The transformed vector of the whole set of perturbed networks is given by:

$$\overrightarrow{g} = G\overrightarrow{f}. \quad (3)$$

The obtained new set of variables are linear combinations of the original set of measurements. We considered the PCA approach to determine the separation between different networks considering each measurement. The most suitable measurement for network study are those that provide good discrimination between networks with different topologies.

### C. Analyzed Networks

In order to study the effects of perturbations on networks, we considered structures generated by four network models and five real-world networks. The consideration of models is fundamental for investigating the outcomes of perturbations in networks with different number of edges and vertices. Also, since the four considered models have distinct structure, it is possible to analyze the discrimination in networks classification by network measurements.

### 1. Theoretical models

Complex networks models have been developed to reproduce the topology and growth of many complex systems. In this work, we considered the following network models [1].

- *Erdős-Rényi random graph (ER)*: This model generates networks with random distribution of connections. The network is constructed connecting each pair of vertices in the network with a fixed probability $p$ [25]. This model generates a Poisson like degree distribution [20].

- *Small-world model of Watts and Strogatz (WS)*: To construct this type of small-word network, one starts with a regular lattice of $N$ vertices in which each vertex is connected to $\kappa$ nearest neighbors in each direction. Each edge is then randomly rewired with probability $q$ [26]. In this work, we considered $q = 0.3$.

- *Barabási-Albert scale-free model (BA)*: This model generates networks with power law degree distribution. The network is generated by starting with a set of $m_0$ vertices and, at each time step, the network grows with the addition of a new vertices with $m$ links. The vertices which receive the new edges are chosen following a linear preferential attachment rule, i.e. the probability of the new vertex $i$ to connect with an existing vertex $j$ is proportional to the degree of $j$, $\mathcal{P}(i \to j) = k_j / \sum_u k_u$ [27].

- *Waxman geographical model (WG)*: Geographical networks can be constructed by distributing $N$ vertices at random in a 2D space and connecting them according to the distance [28]. This model is created by randomly distributing $N$ vertices in a square of length $L = \sqrt{N}$ and connecting them with probability $p = e^{-\lambda d}$, where $d$ is their geographic distance, and $\lambda$ is a constant adjusted to achieve the desirable average degree.

### 2. Real-world networks

In order to analyze the effect of perturbations in real-world networks, we considered networks of three different classes, namely social, technological and biological.

- **Email network** were obtained by analyzing e-mail interchanges between members of the University Rovira i Virgili (Tarragona) [29]. This is a type of social network, as it represents relationships among people.

- **The US airlines transportation network** is formed by US airports in 1997 connected by flights [30].

- **The Western States Power Grid** represents the topology of the electrical distribution grid [26]. Vertices represent generators, transformers and substations, and edges the high-voltage transmission lines that connect them.

- **Neural network** of the worm *Caenorhabditis elegans*, composed by neurons connected according to synapses [26].

- **Protein-protein interaction network** of the yeast *Saccharomyces cerevisiae* is formed by proteins connected according to identified directed physical interactions [10].

Most of these networks are affected by sampling problems. The air transportation and power grid networks are relatively complete. The e-mail and neural networks can be quite complete depending on the methodology and accuracy adopted to obtain the database. On the other hand, protein interaction networks are far from complete [12].

## III. RESULTS AND DISCUSSIONS

The analyzed theoretic models had the following parameters: $N = 2000$ vertices; average degree 6; in case of WS model, the probability $q$ of reconnection was 0.3; and $\lambda$ was 1.0 for WG model. More details about their properties as well the properties of the real networks are given in Table I. The directed networks where transformed into undirected ones by applying the symmetry operation [20], that is, considering two vertices as connected if there is at least one connection between them (disregarding direction).

The perturbations were performed from 0.2% up to 10% of the total number of edges of each network in steps of 0.2%. Also, for each network and for each type of perturbation, an ensemble of 50 networks was obtained. The trajectories defined by the evolution of perturbations can be seen in Figures 1 and 2. For each measurement, a set of nine plots was obtained, where the first row of plots corresponds to the values (average of the 50 networks generated) of the respective measurements in each step of perturbation; the second row is the corresponding variation of the first row in relation to the original value (without perturbation); and the third is the PCA projections (how they were obtained will be explained below). The columns of each measurement correspond, respectively, to the types of perturbations: edge addition, rewiring, and removal.

In order to obtain the PCA projections for each measurement and for each type of perturbation, we considered each step of perturbation as one variable in the feature vector and all perturbed networks as the samples. As the networks were perturbed from 0.2% up to 10% in steps of 0.2%, we obtained 50 variables in the feature vector, 200 samples for the theoretic models (50 for each

TABLE I: Properties of the analysed networks before the perturbations, where $\langle k \rangle$ is the average degree; $\langle k_{nn} \rangle$, the average degree of nearest neighbors; $\langle hk_2 \rangle$, the average hierarchical degree of level 2; $R$ square, a statistical measurement of how the accumulated degree distribution approaches to a straight line; $r$, the Pearson correlation coefficient of degree of both extremities of edges, also known as assortativity; $\langle cc \rangle$, the average clustering coefficient; $\langle hcc_2 \rangle$, the average hierarchical clustering coefficient of level 2; $\langle hdr_2 \rangle$, the average hierarchical divergence ratio of level 2; $\ell$, the average shortest path length; $\langle B \rangle$, the average betweenness; and $c_D$, the central point dominance.

| Network | Size | $\langle k \rangle$ | $\langle k_{nn} \rangle$ | $\langle hk_2 \rangle$ | $R$ square | $r$ | $\langle cc \rangle$ | $\langle hcc_2 \rangle$ | $\langle hdr_2 \rangle$ | $\ell$ | $\langle B \rangle$ | $c_D$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ER | 2000 | 6.02 | 6.98 | 208.5 | 0.599 | -0.018 | 0.003 | 0.003 | 0.989 | 4.25 | 0.002 | 0.008 |
| WS | 2000 | 6.00 | 6.27 | 97.5 | 0.698 | -0.044 | 0.213 | 0.037 | 0.881 | 4.88 | 0.002 | 0.007 |
| BA | 2000 | 5.99 | 17.37 | 831.4 | 0.974 | -0.054 | 0.017 | 0.012 | 0.977 | 3.54 | 0.001 | 0.145 |
| GN | 2000 | 6.03 | 6.86 | 90.6 | 0.638 | 0.199 | 0.151 | 0.079 | 0.808 | 7.78 | 0.004 | 0.080 |
| Email | 1133 | 9.62 | 17.90 | 933.6 | 0.751 | 0.078 | 0.220 | 0.078 | 0.804 | 3.33 | 0.002 | 0.037 |
| US airports | 332 | 12.81 | 49.83 | 449.4 | 0.862 | -0.208 | 0.625 | 0.176 | 0.607 | 2.47 | 0.005 | 0.202 |
| Power grid | 4941 | 2.67 | 3.97 | 13.5 | 0.911 | 0.003 | 0.080 | 0.062 | 0.969 | 15.90 | 0.004 | 0.285 |
| Neural network | 297 | 14.46 | 32.00 | 618.3 | 0.839 | -0.163 | 0.292 | 0.072 | 0.579 | 2.25 | 0.005 | 0.299 |
| Protein interaction | 4134 | 4.21 | 29.63 | 497.9 | 0.978 | -0.127 | 0.082 | 0.033 | 0.928 | 4.78 | 0.001 | 0.173 |

of the four models) and 250 for the real-world networks (50 for each of the five real networks). The PCA plots, then, indicate the projections which correspond to the maximum variation in the first two eigenvalues. As can be seen in Figures 1 and 2, only the first projection is enough to separate the network classes for several cases.

The variations of the measurements presented in Figures 1 and 2 are summarized in Tables II and III, which show, for theoretical models and real networks, respectively, the average and standard deviation of the percentage of variations of each measurement for a perturbation of 10% of the total number of edges.

TABLE II: Average and standard deviation of the percentage of variation of each measurement for the theoretic models and for 10% of edge perturbation. The numbers between parenthesis represent the standard deviation. Symbols are the same used in Table I.

| Measurement | Edge addition | Edge rewiring | Edge removal |
|---|---|---|---|
| $\langle k_{nn} \rangle$ | 8.03 ± 3.38 | 0.87 ± 1.34 | 8.07 (0.43) |
| $\langle hk_2 \rangle$ | 42.24 ± 24.7 | 23.71 ± 30.1 | 21.38 (3.15) |
| $r$ | 34.38 ± 30.5 | 18.22 ± 11.2 | 8.32 (8.09) |
| $R$ square | 2.77 ± 2.06 | 0.00 ± 0.00 | 7.74 (9.07) |
| $\langle cc \rangle$ | 13.35 ± 3.62 | 24.64 ± 22.8 | 9.58 (0.78) |
| $\langle hcc_2 \rangle$ | 19.70 ± 14.5 | 29.40 ± 27.4 | 6.14 (1.99) |
| $\langle hdr_2 \rangle$ | 1.81 ± 2.03 | 4.61 ± 5.62 | 1.3 (1.46) |
| $\ell$ | 13.30 ± 16.7 | 11.22 ± 18.9 | 4.88 (0.57) |
| $\langle B \rangle$ | 17.81 ± 23.1 | 15.33 ± 25.9 | 5.61 (1.18) |
| $c_D$ | 26.20 ± 39.2 | 26.76 ± 40.4 | 11.54 (8.09) |

The main results obtained in Figures 1 and 2 and in Tables II and III are identified as follows:

- The relative variation of all measurements is generally greater than the percentage of perturbed edge for all types of perturbations and for all networks. This result implies that most of the measurements are sensitive to small perturbations.
- The random edge removal causes smaller variations in the measurements than the other two types of perturbations. From this result, we can conclude that it is better not to include edges which we are uncertain, as the inclusion of an unexistent edge implies larger deviations of the measurements than the removal of an existing one.

TABLE III: Average and standard deviation of the percentage of variation of each measurement for the real networks and for 10% of edge perturbation. The numbers between parenthesis represent the standard deviation. Symbols are the same used in Table I.

| Measurement | Edge addition | Edge rewiring | Edge removal |
|---|---|---|---|
| $\langle k_{nn} \rangle$ | 7.49 ± 5.07 | 2.03 ± 1.71 | 10.28 (1.43) |
| $\langle hk_2 \rangle$ | 22.88 ± 13.3 | 17.79 ± 10.4 | 13.82 (9.17) |
| $r$ | 26.96 ± 13.1 | 31.87 ± 39.2 | 4.66 (10.0) |
| $R$ square | 0.90 ± 0.81 | 0.00 ± 0.00 | 0.72 (0.59) |
| $\langle cc \rangle$ | 16.50 ± 5.39 | 30.64 ± 12.3 | 12.53 (2.93) |
| $\langle hcc_2 \rangle$ | 16.19 ± 10.1 | 23.11 ± 18.6 | 7.86 (5.38) |
| $\langle hdr_2 \rangle$ | 2.00 ± 1.10 | 2.66 ± 2.19 | 3.04 (1.79) |
| $\ell$ | 11.95 ± 16.8 | 10.14 ± 16.7 | 10.69 (10.3) |
| $\langle B \rangle$ | 14.81 ± 20.0 | 13.72 ± 20.4 | 2.25 (2.00) |
| $c_D$ | 26.79 ± 34.6 | 22.46 ± 35.2 | 5.08 (3.01) |

- Networks with geographic structure, e.g. WG model and power grid, present great variations in measurements related to shortest path in the edge addition and rewiring perturbations (see Figures 1 and 2). This result can be explained by the fact that adding or rewiring edges can connect vertices which are far away, reducing the average shortest path length, that is, it is an effect of the randomness of the rewiring, as opposed to the geographic nature of the existing edges.

The main motivation for studying perturbations in networks is to find measurements allowing an acceptable compromise between stability and discriminability. Therefore the measurements which satisfy these requirements are those with small variations (see Tables II and III) and good separability of respective PCA projections (see Figures 1 and 2). According to these criterions, the best measurements are, in sequence from best

to worst: average hierarchical divergence ratio of level 2, average shortest path length, average betweenness, average hierarchical clustering coefficient of level 2, and average clustering coefficient. The other measurements have poor separability. Interestingly, the hierarchical measurements for level 2 tended to present better performance in presence of the perturbations than the more traditional measurements considering only the immediate neighborhood of each node. This property is possibly related to fact that the hierarchical measurements take into account a larger portion of the original network, therefore providing a more stable and informative quantification of the local topology.

## IV. CONCLUSIONS

Much of the success of complex network research relies on accurate modeling of complex phenomena. To reach this goal, the main efforts should be concentrated in developing methods able to obtain accurate databases and measurements that can characterize networks structures with accuracy. Thus, the development of accurate sampling techniques and analysis of the behavior of measurements with respect to incomplete networks or networks with biased connections are fundamental for complex networks research. In this paper, we reported an analysis of network measurements of progressively perturbed networks. The perturbations were performed at the edges level, considering random removal, addition and rewiring. Our results suggest that the considered complex networks measurements are less sensitive to edges removal, with greater sensibility to rewiring and addition of edges. This result is interesting because when a network is sampled, if there is no certainty of a connection, it is better not include it.

The measurements most suitable to analyze perturbed network were identified as: (i) the average hierarchical divergence ratio of level 2, (ii) the average shortest path length, (iii) the average betweenness, (iv) the average hierarchical clustering coefficient of level 2, and (v) the average clustering coefficient. The other measurements have poor discrimination between networks with different structures. In particular, networks with geographical structure are very sensitive to perturbations. Unfortunately, many real networks such as the Internet, power grid distribution, air connections, road connections, street networks, subway networks and even proteins interactions (e.g. [20, 31]) present some level of geographical organization. We suggest as future works the consideration of other complex network measurements as well as other types of perturbations, such are node removal or perturbation with preferential rules. The consideration of multivariate statistical methods (e.g. Manova [32]) and data mining techniques can also help complementing the respective analysis.

**Acknowledgments**

Luciano da F. Costa is grateful to FAPESP (05/00587-5), CNPq (301303/06-1) for financial support. Francisco A. Rodrigues acknowledges FAPESP sponsorship (07/50633-9), Paulino R. Villas Boas acknowledges CNPq sponsorship (141390/2004-2).

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chaves, and D.-U. Hwang. Complex networks: structure and dynamics. *Physics Reports*, 424:175–308, 2006.

[2] L. da F. Costa, O.N. Oliveira Jr, G. Travieso, F.A. Rodrigues, P.R. Villas Boas, L. Antiqueira, M.P. Viana, and L.E.C. da Rocha. Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications. *eprint arXiv: 0711.3199*, 2007.

[3] M.P.H. Stumpf, C. Wiuf, and R.M. May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *PNAS*, 102(12):4221–4224, 2005.

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

[5] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Physical Review Letters*, 90(2):028701–028703, 2003.

[6] L. da F. Costa and G. Travieso. Exploring complex networks through random walks. *Physical Review E*, 75(1):16102, 2007.

[7] B. Tadić, G. J. Rodgers, and S. Thurner. Transport on complex networks: Flow, jamming and optimization. *International Journal of Bifurcation and Chaos*, 17(7):2363–2385, 2007.

[8] R. Mrowka, A. Patzak, and H. Herzel. Is there a bias in proteome research? *Genome Research*, 11:1971–1973, 2001.

[9] R. Saito, H. Suzuki, Y. Hayashizaki, and O. Journals. Interaction generality, a measurement to assess the reliability of a protein–protein interaction. *Nucleic Acids Research*, 30(5):1163–1168, 2002.

[10] E. Sprinzak, S. Sattath, and H. Margalit. How reliable are experimental protein–protein interaction data. *J Mol Biol*, 327(5):919–23, 2003.

[11] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

[12] J. D. Han, D. Dupuy, N. Bertin, M. E. Cusick, and M. Vidal. Effect of sampling on topology predictions of protein-protein interaction networks. *Nature Biotechnology*, 23(7):839–844, 2005.

[13] R. Khanin and E. Wit. How scale-free are biological networks. *J. Comput. Biol*, 13:810–818, 2006.

[14] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, 440:637–643, 2006.

[15] M.Á. Serrano, A. Maguitman, M. Boguñá, and A. Vespignani. Decoding the structure of the WWW: A comparative analysis of Web crawls. *ACM Transactions on the Web*, 1(2), 2007.

[16] L. Becchetti, C. Castillo, D. Donato, and A. Fazzone. A comparison of sampling techniques for web graph characterization. In *Proceedings of the Workshop on Link Analysis (LinkKDD06)*, Philadelphia, PA, 2006.

[17] J. Leguay, M. Latapy, T. Friedman, and K. Salamatian. Describing and simulating Internet routes. *Computer Networks*, 51(8):2067–2085, 2007.

[18] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *INFOCOM 2003*, volume 1. IEEE, 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies.

[19] G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28(3):247–268, 2006.

[20] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167 – 242, 2007.

[21] M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):26126, 2003.

[22] L. da F. Costa and F. N. Silva. Hierarchical characterization of complex networks. *Journal of Statistical Physics*, 125(4):841–872, 2006.

[23] I. T. Jolliffe. Principal component analysis. *Springer Series in Statistics, Berlin: Springer, 1986*, 1986.

[24] L. da F. Costa and R. M. Cesar Jr. *Shape Analysis and Classification: Theory and Practice*. CRC Press, 2001.

[25] P. Erdős and A. Rényi. On the evolution of random graphs. *Bulletin of the International Statistical Institute*, 38(4):343–347, 1960.

[26] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[27] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.

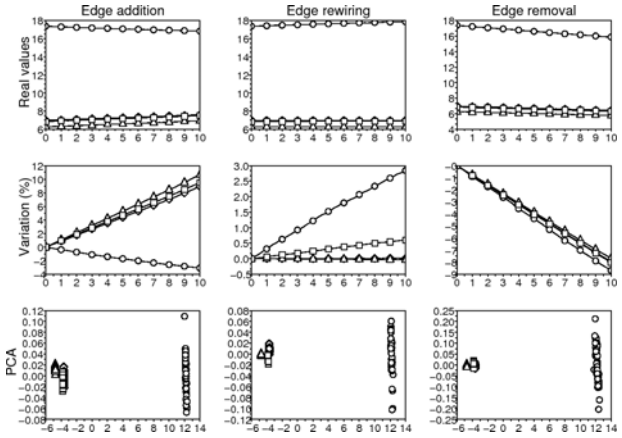
[28] B. M. Waxman. Routing of multipoint connections. *Selected Areas in Communications, IEEE Journal on*, 6(9):1617–1622, 1988.

[29] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):65103, 2003.

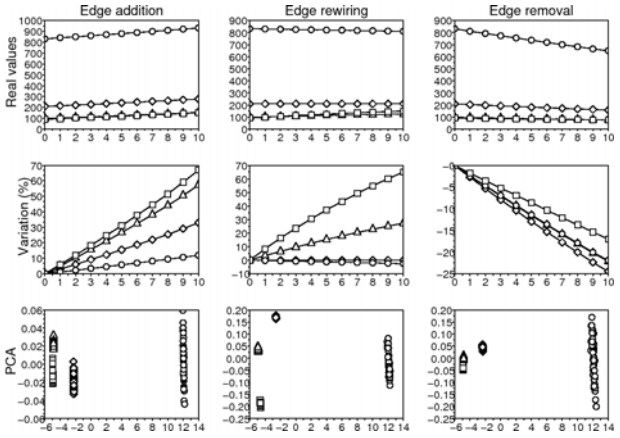
[30] V. Batagelj and A. Mrvar. Pajek datasets, 2006. `http://vlado.fmf.uni-lj.si/pub/networks/data`.

[31] N. Przulj, DG Corneil, and I. Jurisica. Modeling interactome: scale-free or geometric?, 2004.

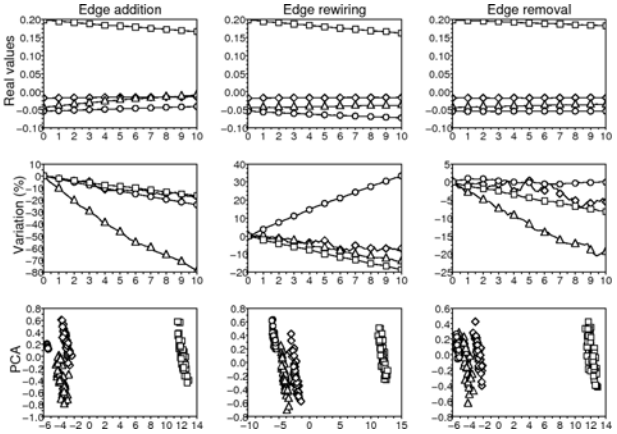
[32] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2001.



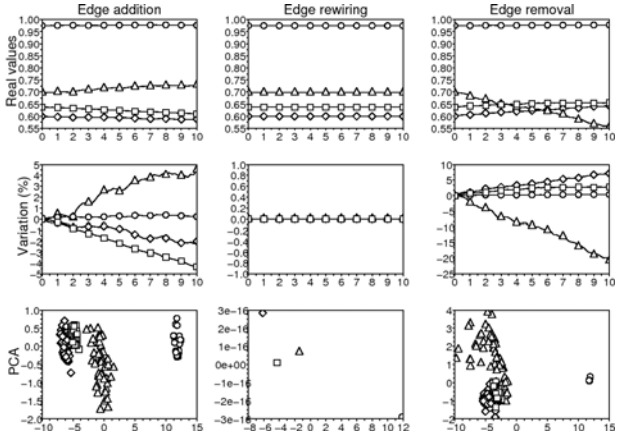

**(a)** Average degree of nearest neighbors.



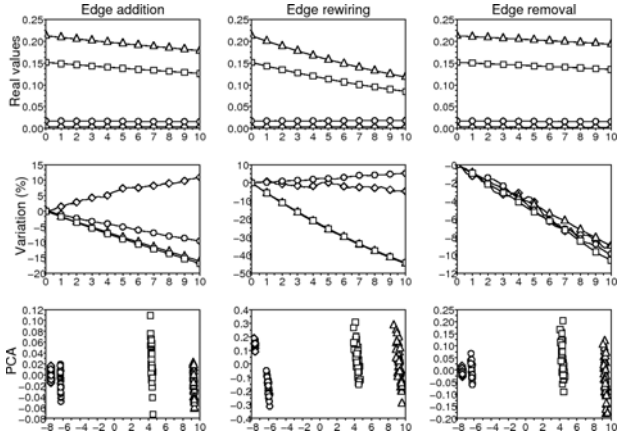

**(b)** Average hierarchical degree of level 2.



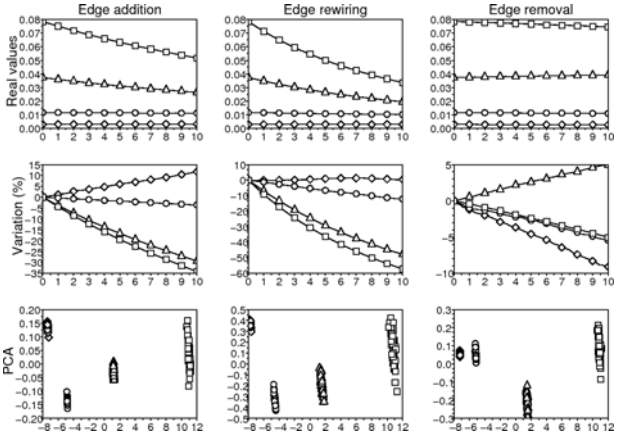

**(c)** Assortativity.



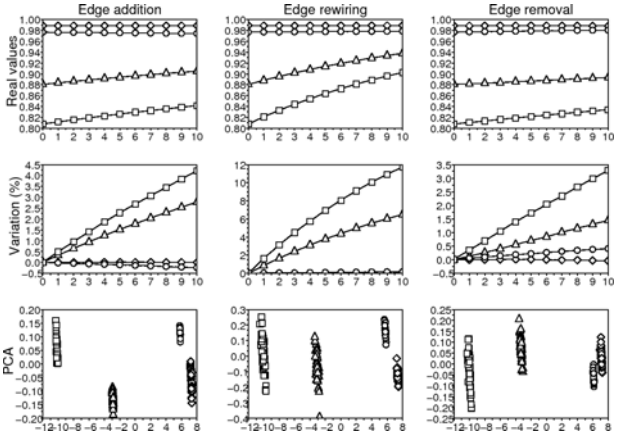

**(d)** $R$ square.



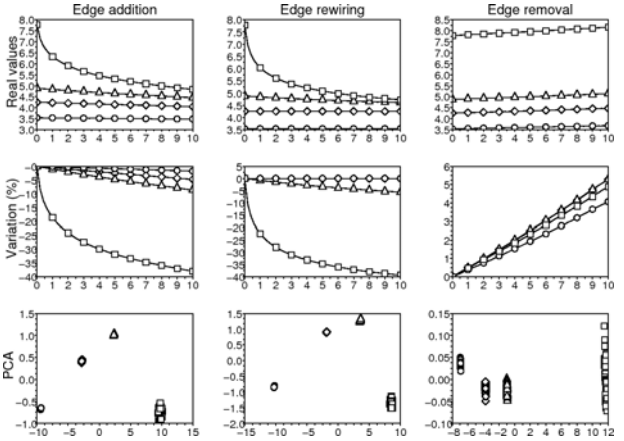

**(e)** Average clustering coefficient.



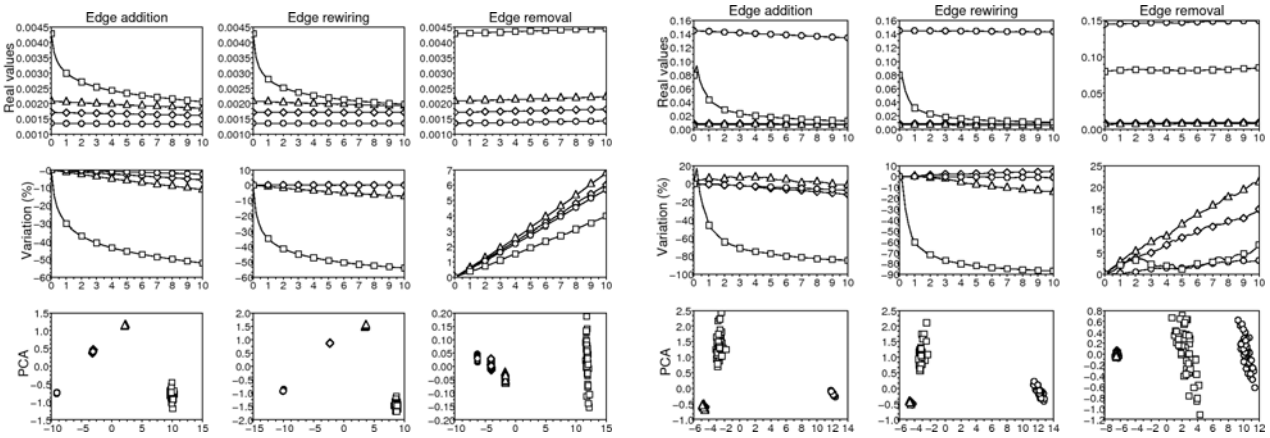

**(f)** Average hierarchical clustering coefficient of level 2.




**(g)** Average hierarchical divergence ratio of level 2.

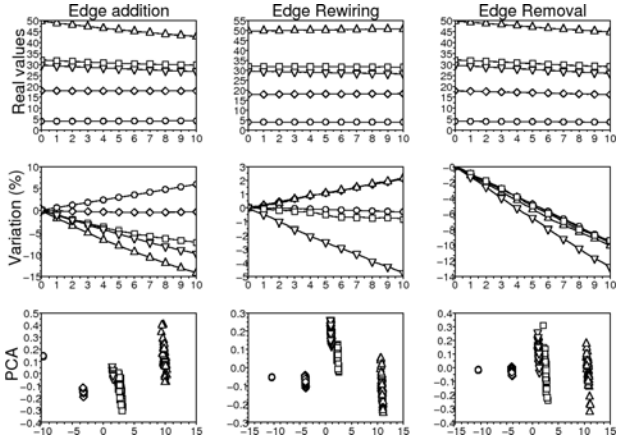



**(h)** Average shortest path length.

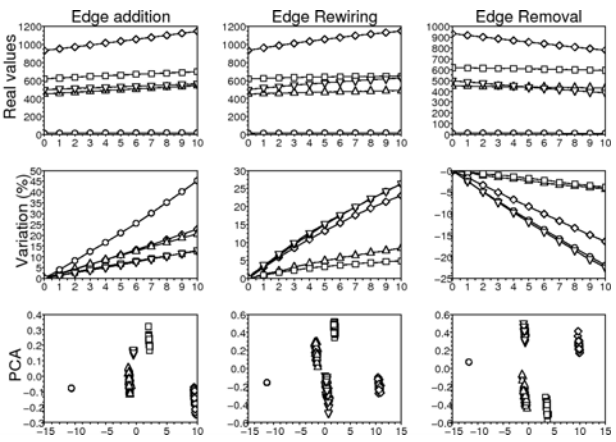



**(i)** Average betweenness. **(j)** Central point dominance.

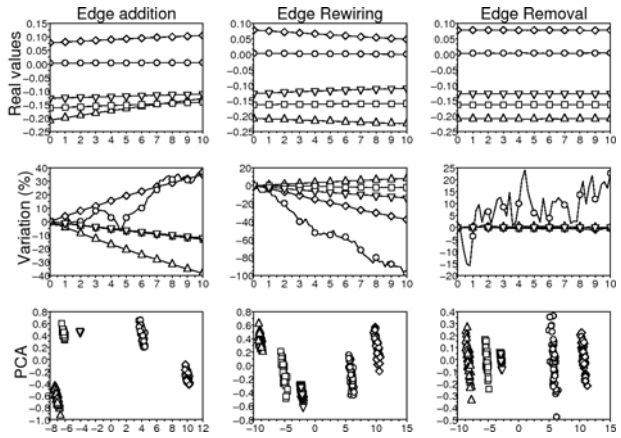
FIG. 1: Measurements for the theoretical models: ♢ – ER, △ – WS, ◯ – BA, and □ – WG. For each measurement, the first column corresponds to randomly adding edges; the second column, to randomly rewiring edges; and the third, to randomly removing edges; the first row corresponds to the real values; the second row, to the normalized values (each real value is divided by the not-perturbed corresponding value); and the third row is the projections of the PCA (refer to text for explanation). For the first and second rows, the $x$-axis is the percentage of added, removed, or rewired edges.



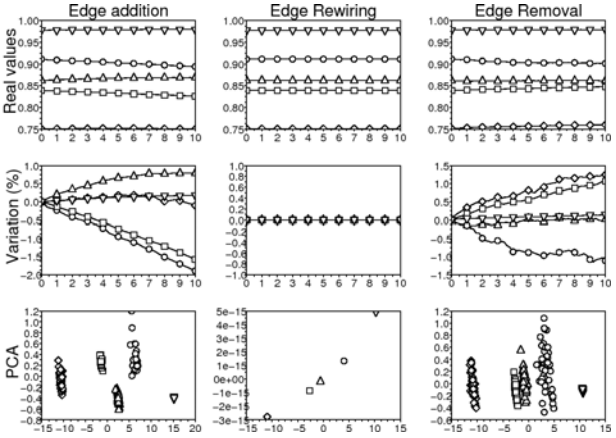

**(a)** Average degree of nearest neighbors.



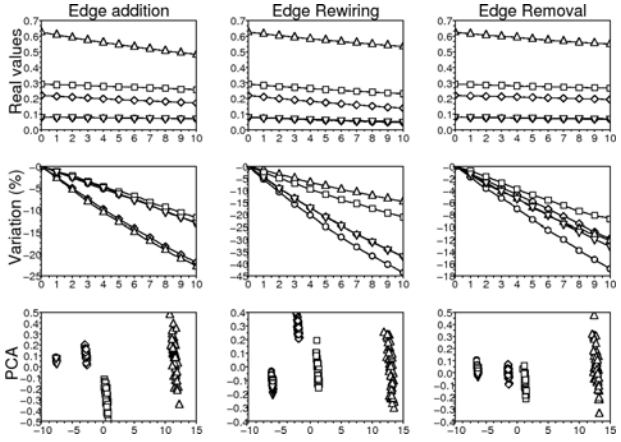

**(b)** Average hierarchical degree of level 2.



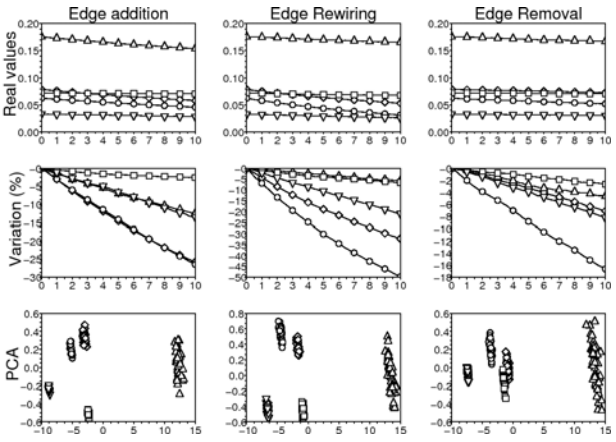

**(c)** Assortativity.



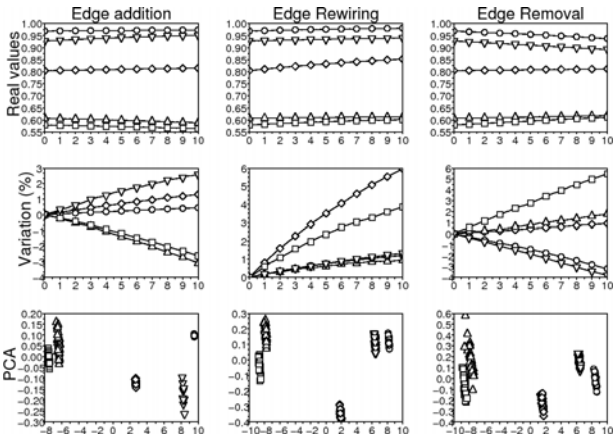

**(d)** $R$ square.



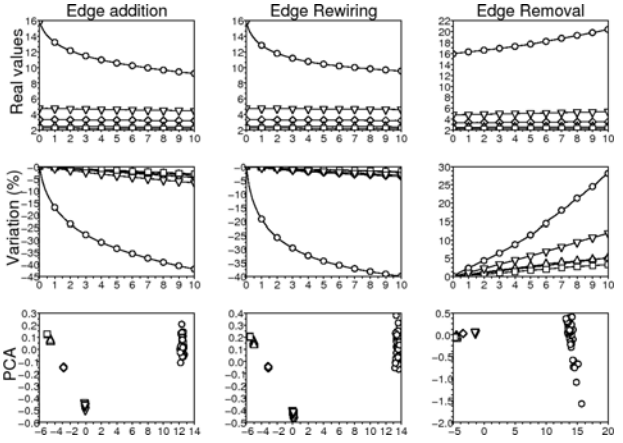

**(e)** Average clustering coefficient.



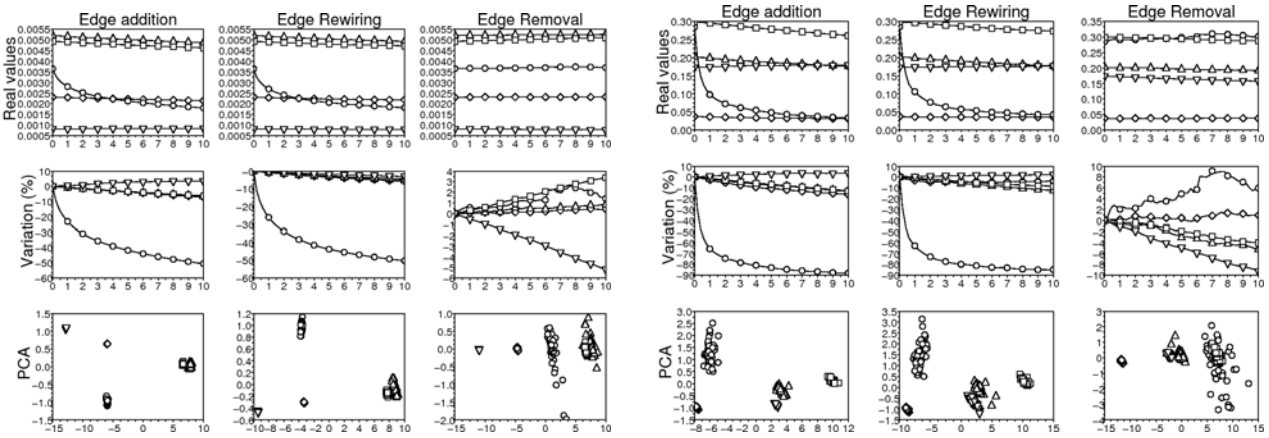

**(f)** Average hierarchical clustering coefficient of level 2.




**(g)** Average hierarchical divergence ratio of level 2.




**(h)** Average shortest path length.




**(i)** Average betweenness. **(j)** Central point dominance.

FIG. 2: Measurements for the real networks ◊ – email, △ – US airports, ◯ – power grid, □ – neural network, ▽ – protein interaction. For each measurement, the first column corresponds to randomly adding edges; the second column, to randomly rewiring edges; and the third, to randomly removing edges; the first row corresponds to the real values; the second row, to the normalized values (each real value is divided by the not-perturbed corresponding value); and the third row is the projections of the PCA (refer to text for explanation). For the first and second rows, the $x$-axis is the percentage of added, removed, or rewired edges.