s-core network decomposition: A generalization of *k*-core analysis to weighted networks

Marius Eidsaa and Eivind Almaas*

Department of Biotechnology, NTNU - Norwegian University of Science and Technology, N-7491 Trondheim, Norway

(Received 14 March 2013; revised manuscript received 15 November 2013; published 30 December 2013)

A broad range of systems spanning biology, technology, and social phenomena may be represented and analyzed as complex networks. Recent studies of such networks using *k*-core decomposition have uncovered groups of nodes that play important roles. Here, we present *s*-core analysis, a generalization of *k*-core (or *k*-shell) analysis to complex networks where the links have different strengths or weights. We demonstrate the *s*-core decomposition approach on two random networks (ER and configuration model with scale-free degree distribution) where the link weights are (i) random, (ii) correlated, and (iii) anticorrelated with the node degrees. Finally, we apply the *s*-core decomposition approach to the protein-interaction network of the yeast *Saccharomyces cerevisiae* in the context of two gene-expression experiments: oxidative stress in response to cumene hydroperoxide (CHP), and fermentation stress response (FSR). We find that the innermost *s*-cores are (i) different from innermost *k*-cores, (ii) different for the two stress conditions CHP and FSR, and (iii) enriched with proteins whose biological functions give insight into how yeast manages these specific stresses.

DOI: 10.1103/PhysRevE.88.062819 PACS number(s): 89.75.Hc, 05.10.−a, 05.40.Fb, 87.18.−h

I. INTRODUCTION

Weighted complex networks, where the importance or strength of links is nonuniform, are frequently used as representations of a wide range of systems, spanning biology, technology, and social phenomena [1–3]. A quest in the analysis of complex networks is the identification of sets of nodes with special topological properties. Often, highly connected nodes (hubs) are considered *a priori* to be of central importance for the network, and the importance of the node-specific property of "degree" has been supported by many studies (see, e.g., [4–6]).

The method of *k*-core analysis [7] was proposed to uncover regions in a network that are tightly connected: sets of nodes with high degree connected to other such nodes. It has recently received some attention for its ability to identify locally and globally central clusters of nodes in biological networks more robustly than simply through the ranking of nodes according to degree [8,9], which is a local property. *k*-core analysis has also become an efficient tool in the visualization of large-scale networks [10,11], and is used in the analysis of Internet maps in relation to different network models [12].

A variant of *k*-core decomposition, called the *k*-shell [13], was shown to effectively identify a nucleus set of nodes of size ∼100 in a map of the Internet at the autonomous systems level consisting of 20 000 nodes. Additionally, *k*-core analysis has been used to evaluate the structure of the air transportation networks of the seven largest carriers in USA [14] and to identify influential spreaders in networks [15].

Part of the attraction of *k*-core decomposition lies in its simplicity and the possibility for implementing fast computational algorithms in sparse networks [16]. However, there also exist efficient algorithms for *k*-core analysis in dense, massive networks tested on systems of sizes up to 52.9×10^6 nodes and 1.65×10^9 links [17]. The *k*-core decomposition is additionally of interest because of its close relationship with bootstrap percolation [18]. Notably, a recent analysis of *k*-core percolation on randomly damaged, uncorrelated networks found that it consists of a hybrid phase transition with characteristics typically present in either first order or continuous phase transition [19,20]. The sizes of *k*-cores have also been calculated for random networks with tunable clustering [21].

While the majority of network-analysis methods have been developed for systems where the links are of uniform strength, several network measures have recently been generalized [1,2,22] to networks where the links are heterogeneous. Here, we present a generalization of the *k*-core method to weighted networks: The *s*-core method takes node strength into account while following a heuristic similar to that of *k*-core. For networks with uniform link weights, the *s*-core decomposition becomes the same as *k*-core. To investigate the effect of link strength on a network's-core structure, we apply the *s*-core decomposition to the ER and the configuration network model [23–25] with a scale-free degree distribution and with link strengths that are (i) random, (ii) correlated, and (iii) anticorrelated with node degrees. Finally, we conduct a comparative analysis of the *s*-core structure of two networks: the protein interaction network (PIN) of the yeast *Saccharomyces cerevisiae* using gene-expression correlation data calculated from (A) an oxidative stress experiment [26] and (B) a 15-day wine fermentation experiment [27] as link strengths.

II. *s*-CORE DECOMPOSITION METHOD

We will use the following notation: A network consisting of N nodes and M links is described by the adjacency matrix $A = [a_{ij}]$, where $a_{ij} = 1$ when nodes i and j are connected, zero otherwise. The number of nearest neighbors of node i is $k_i = \sum_j a_{ij}$. The k-core of a network consists of all nodes i with degree $k_i > (k-1)$, and we identify the k-core by the iterative removal of all nodes i with degrees $k_i \leqslant (k-1)$ when $k > 0$. From the definition, we see that the $k = 0$-core consists of the whole network. The illustration in Fig. 1(a) demonstrates the *k*-core structure in a simple network. Here, we find three *k*-cores, with the innermost core (black nodes) corresponding

*eivind.almaas@ntnu.no

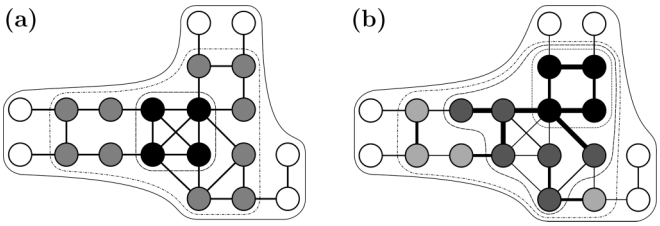

FIG. 1. Illustration of network decomposition: (a) k-core decomposition using node degree. (b) s-core decomposition using node strength. All nodes inside the continuous lines belong to the same core. The link thickness indicates connection strength. Nodes in the same k- or s-shell have identical shading. From outermost towards the core: white, gray, dark-gray (only in the s-core), and black.

to the most tightly connected part of the network. Note that the $(k+1)$-core is included in the k-core.

However, for many systems the links are not of equal strength, and it is important to have the capability to capture the effect of link heterogeneity. For instance, taking into account the number of people traveling between two cities in the world airport network (WAN), or the number of times two scientists have co-authored a paper, would possibly change the identification of central nodes in a network when compared with an analysis based solely on network topology. We introduce the s-core decomposition, the generalization of k-core to weighted networks, by defining the s-core as consisting of all nodes with node strengths $s_i > s$, where s is a threshold value. Following Barrat *et al.* [1], we use the link-weight matrix $W = [w_{ij}]$, where w_{ij} is the strength of the link between nodes i and j, and calculate the strength of node i as $s_i = \sum_j a_{ij} w_{ij}$. This approach is motivated by the observation that node degree, the basis for k-core decomposition, is defined as a node strength in a network where all link strengths are unity $w_{ij} = 1$.

We define the threshold value of the s_n-core as $s_{n-1} = \min_i s_i$, where i is only among the nodes in the s_{n-1}-core network. The s_n-core is thus identified by the iterative removal of all nodes with strengths $s_i \leqslant s_{n-1}$. Similar to k-core analysis where node degrees are recalculated for every node removal, the strengths of nodes neighboring a removed node must be recalculated. Thus, the threshold value of the s_1-core is determined by the smallest s_i value in the whole network, and the s_0-core consists of all the nodes in the whole network. Note that for the case when $w_{ij} = w$ is constant, s-core analysis becomes identical with k-core analysis.

Here, we have introduced s-core analysis in the context of weighted networks where the weights are distributed on the links. However, we would like to point out that in many systems, possible measures of node weight are not limited to being *inferred* from the links, as with node strength s_i, but may instead be a property directly associated with the nodes. For instance, an airport in the WAN may be characterized by its number of runways or the number of people employed. In such situations, where node weight is not related to network connectivity, s-core analysis will not be useful.

Figure 1(b) shows a change in core structure of the example network when taking a set of possible link weights (indicated by varying link thickness) into account by using the s-core decomposition. In this case we find four cores in the network, and the innermost core consists of nodes that were in the 2-core (3 nodes) and the 3-core (1 node) of the unweighted network [see Fig. 1(a)]. Note that this figure illustrates that membership in the innermost core is determined both by network topology and link strength, because several of the nodes connected by the strongest links are not part of the innermost k-core.

For k-core decomposition, the k-shell is defined as the subset of the k-core nodes that are not members of the $(k+1)$-core. Similarly for s-core decomposition, we define the s_n-shell as consisting of all nodes in the s_n-core that are not members of the s_{n+1}-core. For cases where very few nodes share the same strength value, it is not surprising if the majority of s-shells will consist of only a few nodes. Consequently, we could expect the index of the innermost s-core, $n_{\max} \lesssim N$; i.e., the number of s-cores to be of the order of nodes in the network. However, if a large number of nodes share strength values, we could expect $n_{\max} \ll N$. The latter situation is the case of k-core analysis. Finally, the concept of a k-crust [13], i.e., the union of all shells with $k_i \leqslant k$, simply becomes the union of all s-shells with $s_n \leqslant s$.

The fact that all the nodes in the s_{n+1}-core are members of the s_n-core opens the possibility for several approximate approaches in s-core calculation. First, it is possible to calculate a sequence of s-cores based on a predetermined set of threshold values s_n^P. An s_n^P-core would be identical to the s_n-core if they shared threshold value. The s^P-cores could be further analyzed, e.g., by the above described s-core analysis between chosen s_n^P threshold values. Alternatively, one could employ a scheme for repeated subdivision of select s^P intervals $[s_n^P, s_{n+1}^P]$. Second, for cases where few nodes share strength values, one could discretize the node strength spectrum to reduce the number of cores. Such modifications of the s-core approach could have practical implications for the s-core analysis of large weighted networks.

III. APPLYING s-CORE ANALYSIS TO MODEL NETWORKS

In this section, we illustrate s-core analysis and its function by applying it to two different network models. The configuration network model (CNM) [23–25] is an uncorrelated random network model where it is possible to specify the degree sequence of the nodes. The Erdös-Rényi network model (ER) consists of a prespecified number of nodes N connected through a random selection process giving rise to a binomial degree distribution [28–31]. In the following, we used the CNM with a scale-free [5,32] degree distribution ($\gamma = 2.5$), and ER with probability of keeping a link set to $p = 3 \times 10^{-5}$ to investigate effects on the s-core network structure when link strengths are *correlated* with node degrees. For the computer simulations, we used $N = 10^5$ nodes and average over either 100 or 1000 network realizations, as specified below.

First, we used the s-core approach to analyze CNM networks with homogeneous link weights $w_{ij} = 1$, i.e., the standard k-core analysis. Figure 2(a) shows the average size of the various k-cores, decreasing as a power-law with exponent $\alpha \approx -3.5$. We find the innermost core to be the $k = 17$-core for this number of network realizations. Figure 2(b)(left) shows the average degree per k-core, $\langle k \rangle_{\text{core}}$. We also

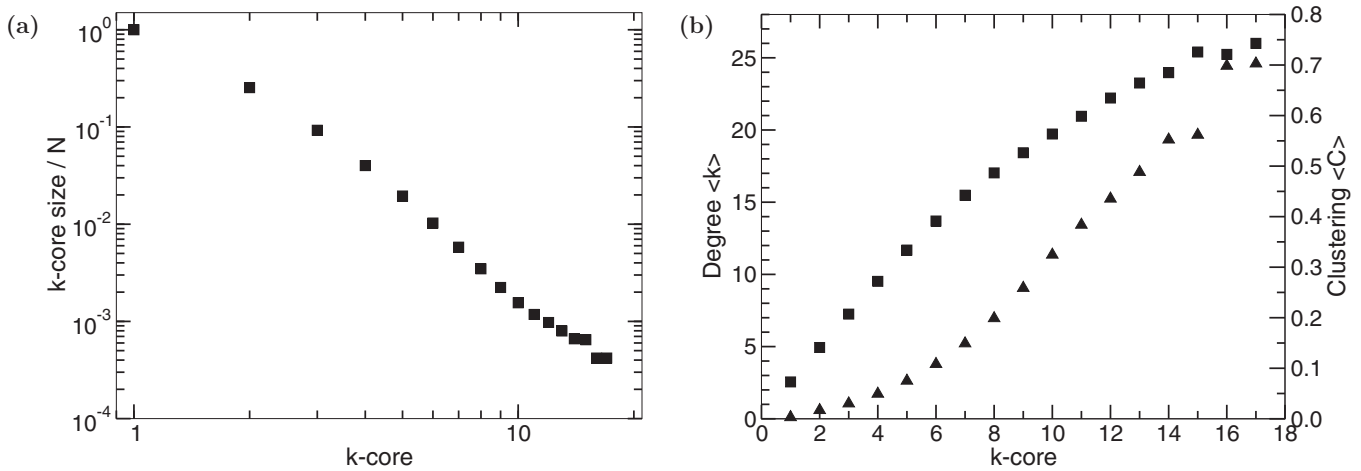

FIG. 2. Properties of k-cores in scale-free configuration model network (CNM) with $N = 10^5$ and power-law exponent $\gamma = 2.5$, averaged over 1000 realizations. (a) Average size and (b) average degree $\langle k\rangle_{\rm core}$ (left, boxes) and average clustering $\langle C\rangle_{\rm core}$ (right, triangles) of the k-cores.

calculated $\langle C\rangle_{\rm core}$, the average clustering [33] of the k-cores [Fig. 2(b)(right)], where the clustering of a node is a measure of its propensity to participate in triangles with its nearest neighbors. As expected, we find that both $\langle k\rangle_{\rm core}$ and $\langle C\rangle_{\rm core}$ display a clearly increasing trend towards the central core of the network.

To study the effect of link weights on this scale-free CNM, we considered three cases of link-weight correlations [34]: (i) uniformly random, $w_{ij} \in (0,1]$, (ii) positively correlated with node degrees, $w_{ij} \sim (1+r)k_i k_j$, and (iii) negatively correlated with node degrees, $w_{ij} \sim (1+r)/(k_i k_j)$, where r is uniformly random $r \in (-0.1,0.1)$. Figure 3(a) shows that the largest number of s-cores is found in networks with positive correlation between weights and node degree, and the smallest number of cores is found when the weights are anticorrelated with node degree. This is not surprising since scale-free networks are known to rapidly fracture when hubs are targeted for removal [35]: When the weights are positively correlated with degree, low-degree nodes will be removed first and the hubs last. When the weights are anticorrelated with degree, the hubs will be removed at a much earlier stage than for case (ii), and the network fractures.

Furthermore, we observe that the curves for s-core size [Fig. 3(a)] for both cases (ii) and (iii) display a marked change of slope. In the correlated case (triangles), the slope is -2 for the s-core index range $x = n/N \in [0,0.1]$ and -1 for $x \in [0.1,0.9]$. Since this figure shows the scaled size of the s-cores, a slope of -2 means that on average only two nodes are removed between consecutive cores. Similarly, the slope of -1 is caused by the removal of a single node between two consecutive cores. Thus, when link weights are correlated with node degrees on both sides of the link, nodes with the smallest node strength are found in network components of size 2. Consequently, network components of size 2 will be removed first, giving rise to the initial slope of -2. Due to the nondegenerate nature of the node strengths, further node removals tend not to generate cascades: A cascade will take place only if the node strength of *a nearest neighbor* of the removed node falls below the node strength threshold. Since the link weights are correlated with node degrees and the network is uncorrelated scale free with neutral assortativity [$\langle s\rangle(k) \sim k^2$], the vast majority of the following node removals give rise to s-shells of size 1.

In link-weight scheme (iii) [circles in Fig. 3(a)], the slope is initially -1 (for $x \in [0,0.25]$) before it changes to -2 for $x \in [0.35,0.7]$. Here, nodes with degree $k = 1$ that are connected to a hub node possess the smallest node strengths and are removed first. These removals tend not to lead to cascades since the only node that is perturbed in the network is a hub. At $x \approx 0.25$, the pool of single-degree nodes connected to hub nodes is exhausted, and the hub nodes are beginning to be removed. At $x > 0.35$, the network mostly consist of pairs of connected nodes or small stars, giving rise to the -2 slope.

For case (i), with random link weights [squares in Fig. 3(a)], the slope is somewhere in between the two correlated cases. Since we can expect $\langle s(k)\rangle \sim k$, the average strength of a node to be proportional to its degree, we should expect a behavior that shows stronger similarities with the correlated case than the anticorrelated one.

In support of these arguments, Fig. 3(b) shows the s-shell size distribution for the three cases of link-weight correlations. Here, we find that CNM networks with positive correlations between weights and node degree have the biggest fraction of the smallest s-shell (size 1), and the smallest fraction for all other s-shell sizes. Approximately 10% of the s-shells have size 2, which is in good agreement with Fig. 3(a) where almost 20% of the network has been removed at $x = 0.1$. For the anticorrelated case, ~40% of the s-shells are size 1 and ~30% are size 2, which is reasonable since the transition from slope -2 to -1 is not very sharp in Fig. 3(a).

Figures 3(c) and 3(d) show the average degree $\langle k\rangle_{\rm core}$ and the average clustering $\langle C\rangle_{\rm core}$ per s-core, respectively, for the three weight schemes. For both cases (i) and (ii), we find a markedly increasing trend towards the innermost cores. This is in clear contrast with case (iii), where both $\langle k\rangle_{\rm core}$ and $\langle C\rangle_{\rm core}$ display a clear decrease towards the innermost cores: Up to $x \sim 0.25$, the nearly flat shape of $\langle k\rangle_{\rm core}$ is in agreement with removal of degree 1 nodes. At $x \sim 0.25$, $\langle k\rangle_{\rm core}$ drops sharply towards 1 when the hubs are removed from the system. In this region, the network consists of a rapidly increasing fraction of

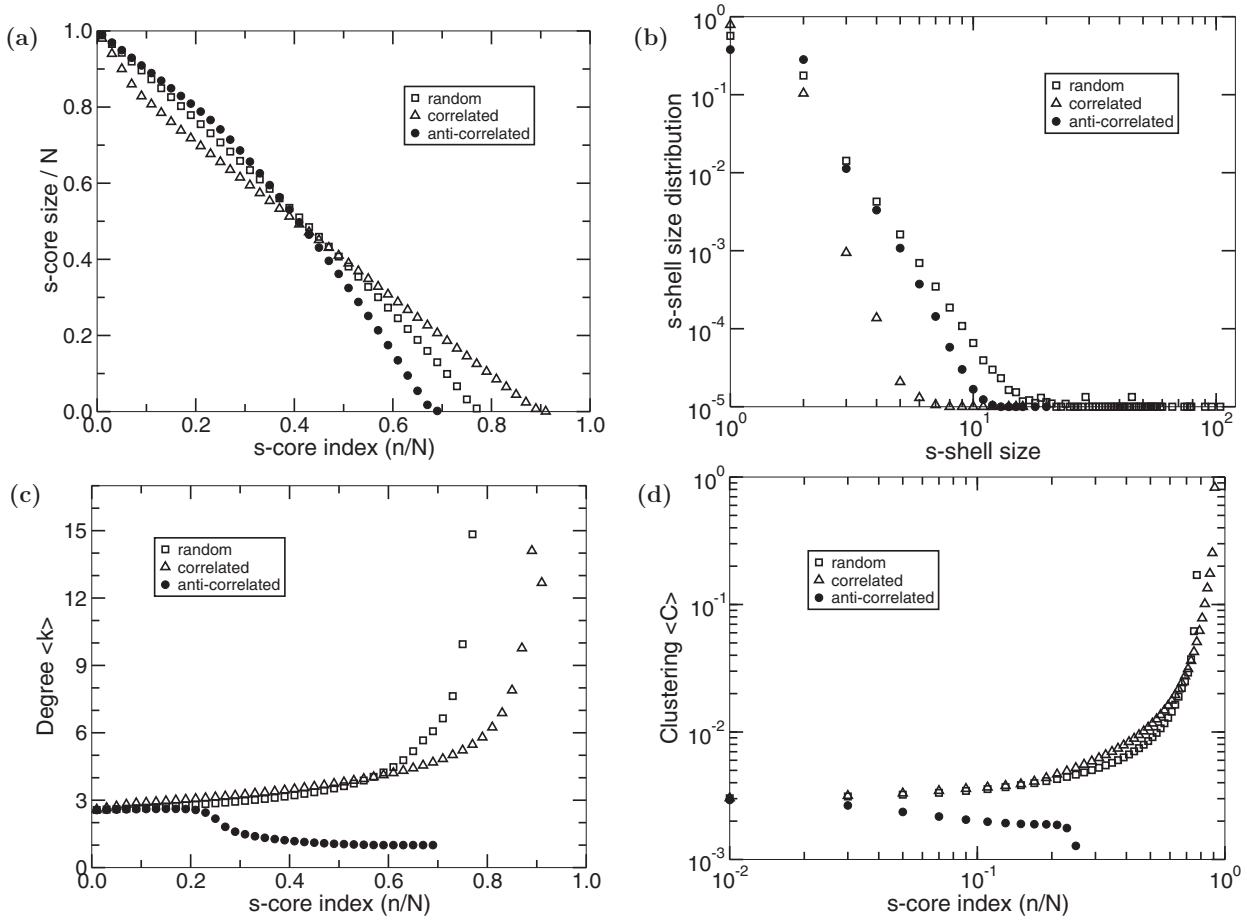

FIG. 3. Properties of s-cores in weighted, scale-free configuration model network (CNM) with $N = 10^5$ and power-law exponent $\gamma = 2.5$, averaged over 100 realizations. (a) Average s-core size, (b) average s-shell size, (c) average degree $\langle k \rangle_{\text{core}}$, and (d) average clustering $\langle C \rangle_{\text{core}}$ as functions of scaled s_n-core index n/N for the three link-weight schemes.

dyads (with increasing x), with dyads dominating the network at $x = 0.35$.

In the above section, we have applied s-core analysis to CNM scale-free networks and studied the effect of different link-weighting schemes. However, some systems are better characterized by a bell-shaped connectivity distribution, and in the following, we use the same analysis approach as for scale-free CNM on the Erdös-Rényi model network (ER) [28–31]. The connectivity distribution for an ER network with $N = 10^5$ and $p = 3 \times 10^{-5}$ is bell shaped, thus with a majority of the nodes with very similar node strengths due to our link-weighting approaches, we expect the s-core structure to be notably different from that of the scale-free case.

Figure 4(a) shows the average s-core size as a function of scaled s-core index $x = n/N$. Initially, weighting schemes (i)–(iii) all display a slope of -1, corresponding to single node removals between consecutive s-cores. From our argumentation in connection with Fig. 3(a), we are not surprised that weight schemes (i) and (ii) behave similarly, and that s-cores for scheme (i) collapse at a smaller x value than those of (ii). For a large range of x values, the resulting s-shells consist only of a single node. In contrast to the CNM model with scale-free connectivity distribution, the transition from the slope -1 region is not abrupt with increasing x. For both approaches (i) and (ii), we instead find a gradual transition to steeper slopes. Detailed investigations of the node removal dynamics confirms that this corresponds to increasing s-shell size from the removal of the (originally) more connected nodes.

The main difference between cases (i) and (ii) is that the network with random link-weight assignments [case (i)] fracture completely at a smaller x value than for case (ii). This is because case (ii) has node strengths $s(k) \sim k^2$, resulting in the node-strength distribution being broader for case (ii) than for case (i), where $s(k) \sim k$.

For the anticorrelated link weights [case (iii)] low-degree nodes connected to the the most connected nodes in the ER network are removed first since they possess the smallest node strengths. This is also what we observed for scale-free CNM. However, in contrast to the scale-free CNM case, we find that for $x \in [0.05,0.10]$, the size of the s-cores drop rapidly (slope approximately -10). At the smallest x value for this behavior, about 5% of the nodes have been removed from the network. Here, due to the shape of the connectivity distribution, the pool of low-degree nodes connected to high-degree nodes has already been exhausted and the most connected nodes are being removed. Since the connectivity distribution is narrow, the strength of many nodes is close enough to initiate cascades. At the largest x value for this region, the network

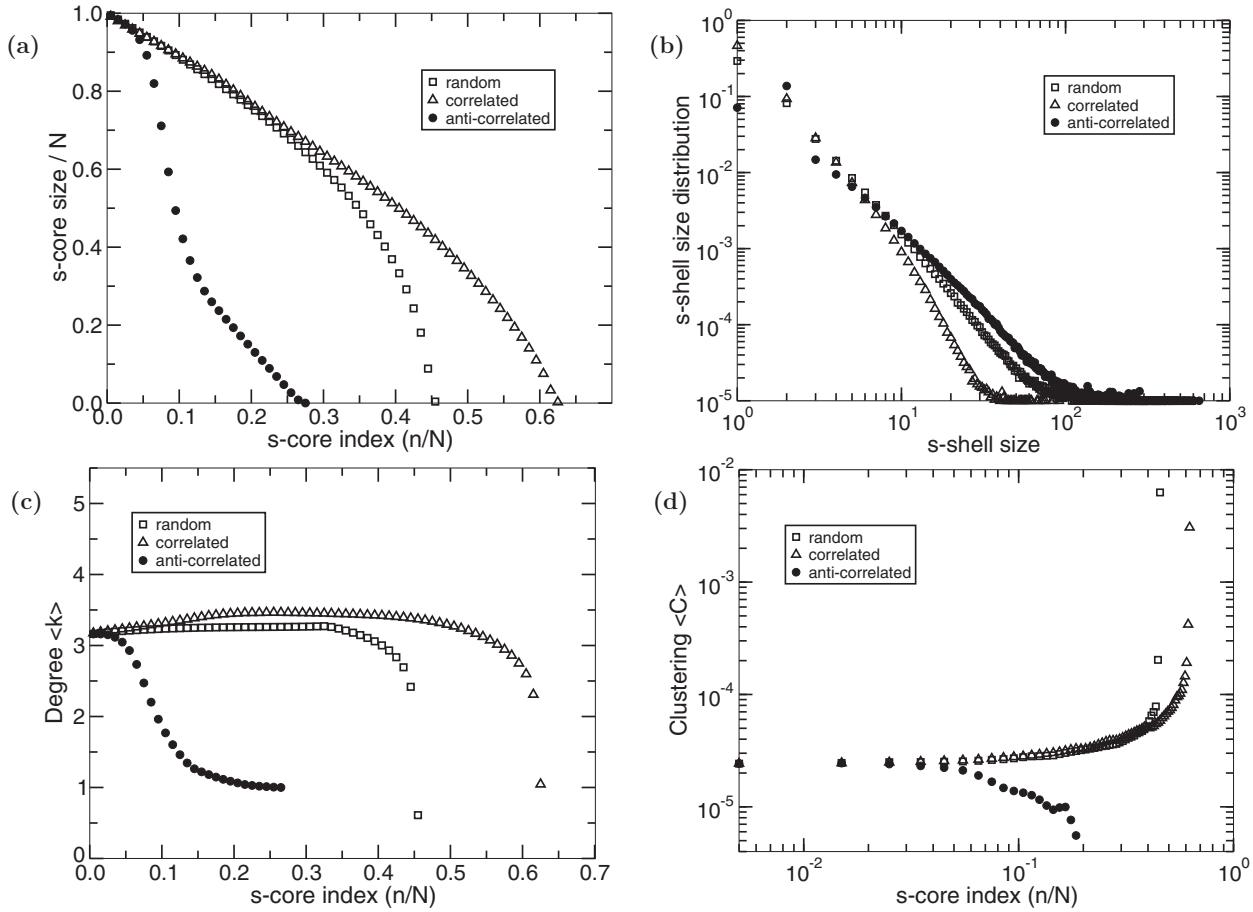

FIG. 4. Properties of s-cores in weighted, Erdös-Rényi model network (ER) with $N = 10^5$ and $p = 3 \times 10^{-5}$, averaged over 100 realizations. (a) Average s-core size, (b) average s-shell size, (c) average degree $\langle k \rangle_{\text{core}}$, and (d) average clustering $\langle C \rangle_{\text{core}}$ as functions of scaled s_n-core index n/N for three link-weight schemes.

is mostly fractured into dyads. Thus, the slope for $x > 0.1$ is near -2.

Figures 4(b)–4(d) support these explanations: in Fig. 4(b) we find that the correlated scheme has the largest number of s-shells with size 1 and the narrowest s-shell distribution. In contrast, the anticorrelated weight scheme (iii) shows a peak at s-shell size 2. In Fig. 4(c), we find that the average degree drops rapidly toward 1 for case (iii) in $x \in [0.05,0.10]$, whereas both cases (i) and (ii) demonstrate neutral to a weak increase with increasing x before a rapid drop near their maximum x. This is in clear contrast to the behavior in Fig. 3(c) for the scale-free model, where cases (i) and (ii) demonstrated a monotonically increasing trend with a rapid increase near their maximal x values. The behavior of the average clustering $\langle C \rangle_{\text{core}}$ in Fig. 4(d) is similar to what we found in Fig. 3(d), with an abrupt decrease in $\langle C \rangle_{\text{core}}$ for case (iii) near the transition point for slope -2 in Fig. 4(a). Also, cases (i) and (ii) show a marked increase in $\langle C \rangle_{\text{core}}$ toward their maximal x values.

IV. APPLYING s-CORE ANALYSIS TO THE YEAST PROTEIN INTERACTION NETWORK

In the following, we will apply the s-core analysis to the yeast protein interaction network (PIN). Our aim is to demonstrate that s-core analysis is capable of uncovering distinct regions of the PIN that are of biological importance in different stress conditions. First, we will analyze the s-core structure of the system, then we will discuss some biological insights gained from looking at a set of central s-cores.

The PIN consists of experimentally determined pair-wise interactions (physical binding) between proteins [36]. Here, we used a high-quality version of the PIN based on extensive data quality experiments and manual curation [37], consisting of 1278 proteins and 1809 links. We calculate the Spearman correlation matrix from gene-expression time-course experiments, and we introduce the absolute values of the correlations as weights in the PIN network whenever the corresponding protein interactions exist. We find that $N = 1181$ proteins in the network match entries on the gene expression array (GEO [38] platform GPL90) used by both experiments, leaving a total of $M = 1566$ links.

We selected two time-course gene-expression data sets measuring yeast response to (i) oxidative stress from exposure to cumene hydroperoxide (CHP) [26] (GEO access ID GSE7645 [38]) and (ii) fermentation stress response (FSR) due to increasing extracellular ethanol levels from 0.5% to 10% (GEO access ID GSE8536 [38]) naturally occurring during the fermentation process [27]. This allows the investigation of how the s-core structure of the yeast PIN changes in response to two

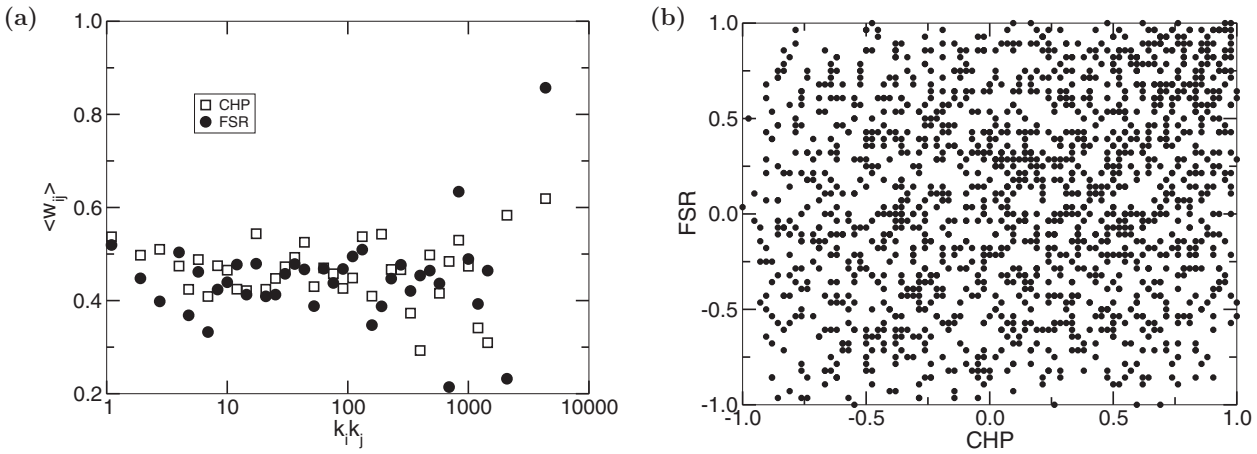

FIG. 5. Link-weight correlations: (a) with network structure $\langle w_{ij}\rangle$ as function of end-node degrees $k_i k_j$, and (b) scatter plot of the link weights for the gene-expression experiments with oxidative (CHP) and fermentation (FSR) stress.

different types of stress, and thus, how the yeast *S. cerevisiae* reorganizes its proteome in response to these two kinds of stress. While we expect that many of the genes will be common in the stress responses for these two different experiments [39], we should also find condition-specific genes.

Figure 5(a) shows average link weight $\langle w_{ij}\rangle$ as a function of the end-node degree $k_i k_j$ for the stresses induced by CHP and FSR. The absence of a trend suggests that the gene correlations associated with the links are independent of the (connectivity) role the proteins have in the network. It is further interesting to note that the scatter plot of the link weights from the two experiments demonstrates there are also no correlations between the link weights in the two stress experiments [Fig. 5(b)]. Thus, while the PIN topology is identical for the two experiments (since the gene array platforms are the same) the weights associated with the links are (i) uncorrelated with the topology and (ii) not correlated with each other. Consequently, these link weights are most similar to link-weight case (i) in the previous section.

In Fig. 6, we calculate the properties of the s-cores of the network when using the CHP and the FSR stress data as link weights. While the average s-core size [Fig. 6(a)] and average degree $\langle k\rangle_{\text{core}}$ Fig. 6(c)] appear different, the insets demonstrate that scaling by the innermost s-core index $n_{\max}$, there is a data collapse. This is caused by (i) the time series consisting of seven (FSR) and eight (CHP) time points giving rise to a discrete spectrum of correlation values. (ii) When the node removal threshold $s_n > 1$, it is no longer possible for nodes with connectivity $k = 1$ to be present in the network, causing the discontinuity in the slope of the curves in Fig. 6(a). (iii) The network topology is identical in the CHP and FSR cases and the link weights are uncorrelated with the topology.

However, the shape of both the s-shell size distribution and the average clustering $\langle C\rangle_{\text{core}}$ are clearly different. Thus, while some of the large-scale s-core network properties are similar, others are different. First, with FSR link weights the network has a reduced number of s-cores [Fig. 6(a)]. This is reasonable, since the s-shell size distribution Fig. 6(b)] shows that CHP conditions give rise to an increased number of smaller s-cores. Additionally, there are eight FSR cores larger than the biggest CHP core with up to 18% more nodes than in the biggest CHP core.

Second, in contrast to the scale-free CNM model with random link weights, where the average degree per core $\langle k\rangle_{\text{core}}$ and average clustering per core $\langle C\rangle_{\text{core}}$ displayed a systematic increase towards the central core, the weighted PINs demonstrate an initial increase followed by a plateau in $\langle k\rangle_{\text{core}}$ before a rapid dropoff at $n/N \gtrsim 0.04$ (0.07) for FSR (CHP). This behavior is not unlike that of the ER network (Fig. 4). For $\langle C\rangle_{\text{core}}$ we see a similar trend, with the FSR weighted network reaching a maximum clustering of $\langle C\rangle_{\text{core}} = 0.17$ at $n/N = 0.33$ before rapidly dropping towards the whole network average value. The CHP weighted PIN reaches a maximum of $\langle C\rangle_{\text{core}} = 0.19$, but displays a less clear trend for $n/N > 0.05$.

Taken together, these results suggest that (i) the CHP and FSR crusts associated with s-core index $n/n_{\max} = 0.5$ ($n/N \approx 0.025$ for FSR and $n/N \approx 0.035$ for CHP) are quite similar, and (ii) the inner cores of CHP and FSR show different network properties and are likely associated with different genes and genetic functions. In support of (i) we find that 74.8% of the genes in the two $n/n_{\max} = 0.5$ crusts are identical.

Next we investigate point (ii) by comparing the properties of two central cores, one from CHP ($n/N = 0.0627$, consisting of 59 nodes) and one from FSR ($n/N = 0.0415$, consisting of 62 nodes). We chose these s-cores since the innermost k-core ($k = 4$) of the PIN network consists of 62 nodes. In contrast to the crust comparison above, the overlap between these two s-cores is reduced to 49.2% [see Fig. 7(a)]. Also note that, while there is a marked overlap between either s-core with the k-core, the s-cores' node content is far from determined by the k-core structure of the underlying PIN network.

Calculating the enrichment of annotated biological functions separately in the sets of CHP, FSR, and k-core proteins, we find a strong dominance of functional categories directly related to membrane, transmembrane, and cytoplasmic functions (Bonferroni, $P < 0.01$) [40]. These functions are central in combating a wide variety of stress responses [39], thus being the type of biological activities we would hope to uncover in stress experiments.

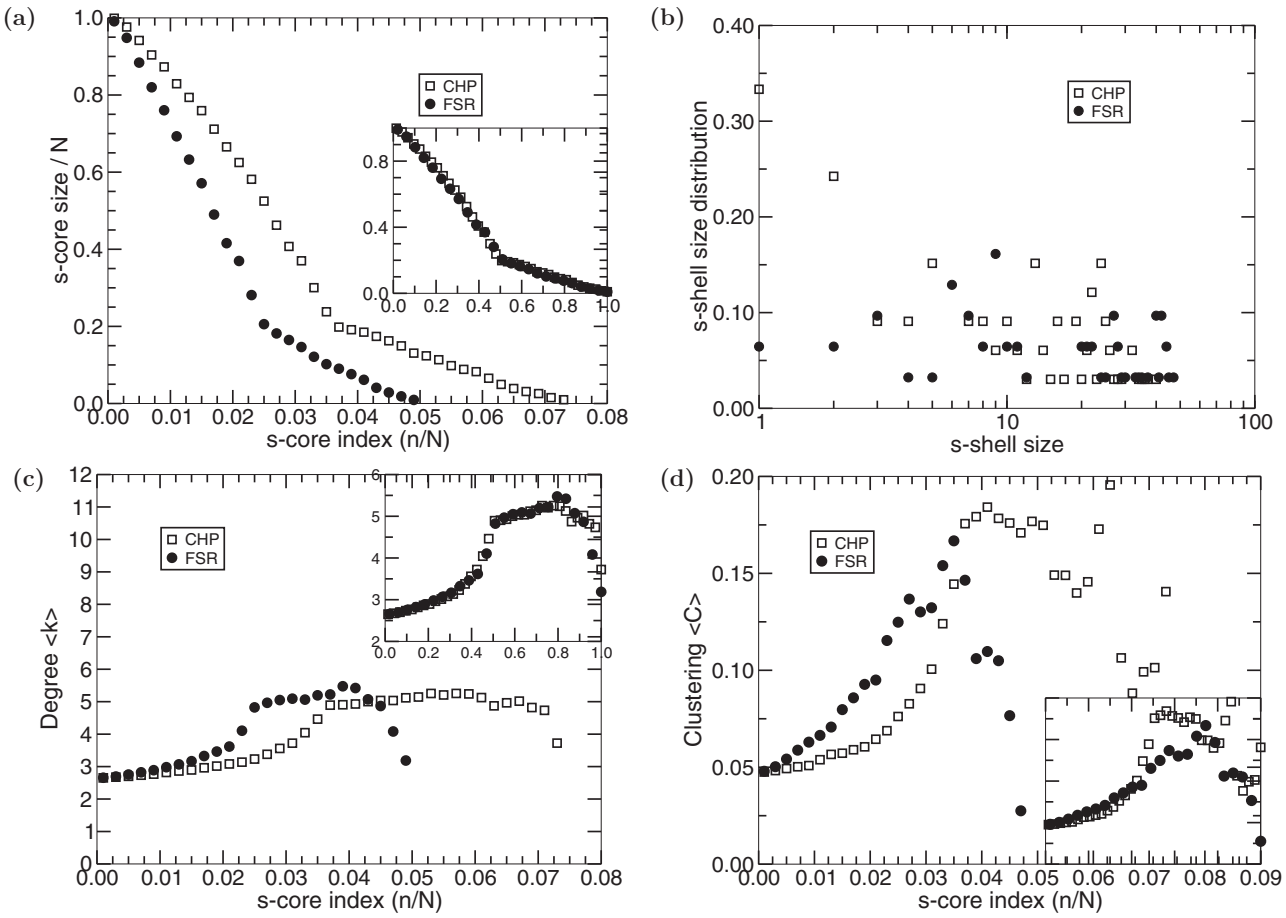

FIG. 6. Properties of s-cores in the yeast PIN for two types of stress, CHP (□) and FSR (●). (a) Average s-core size, (b) average s-shell size, (c) average degree $\langle k \rangle_{\text{core}}$, and (d) average clustering $\langle C \rangle_{\text{core}}$ per core as functions of scaled s-core index n/N. The insets show the same data, only with the x axis scaled by the s-core index $n_{\max}$ of the innermost core of each data set.

However, focusing only on the protein sets of the s-cores unique to either CHP (20 proteins) or FSR (16 proteins) while also not members of the k-core (colored red and yellow in Fig. 7), we find that the CHP unique proteins demonstrate a ∼10-fold enrichment for multiple functional annotations associated with actin and cortical cytoskeleton. This is in agreement with established research on the relationship between oxidative stress and the yeast cytoskeleton [41,42]. For FSR, we find a

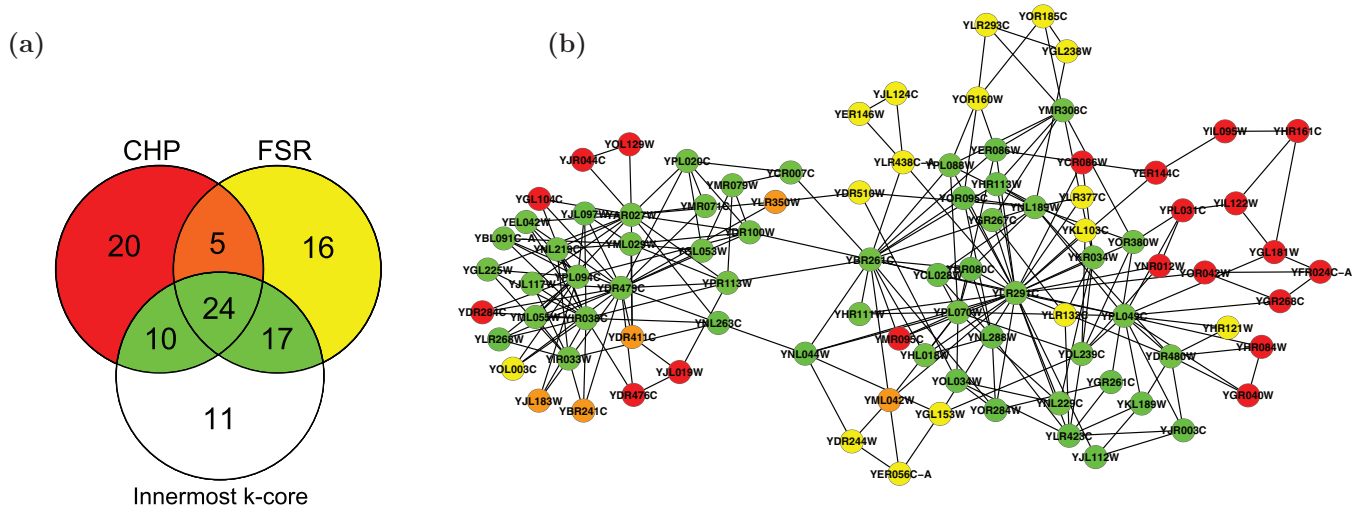

FIG. 7. (Color online) Comparison of inner s-cores. (a) Venn diagram of gene overlap between two selected s-cores and the PIN k-core. (b) The s-core of the two data sets: green - nodes shared with k-core; red - nodes unique to CHP; yellow - nodes unique to FSR; orange - nodes shared only by CHP and FSR.

∼7-fold enrichment for nuclear transport, nucleocytoplasmic transport, and cellular protein complex assembly; also typical responses to stress [39].

V. DISCUSSION

The above s-core analysis of a scale-free CNM network and the ER model for three different link-weight schemes has demonstrated that (1) network topology and (2) how the link weights are correlated with degree are of importance for determining the s-core structure of a network. The effect of link weights is, perhaps, most dramatically demonstrated for the case where link weights are anticorrelated with node degrees, leading to central s-cores only consisting of nodes with a very small number of nearest neighbors.

The discontinuous transition observed in k-cores, where the innermost core collapses, is also a possibility with s-core decomposition. For the s-cores, however, this possibility depends on the network topology, the link-weight correlation, and the link-weight distribution. For example, the CNM networks discussed in this article have an innermost k-core of size 67.1 (averaged over 1000 realizations). The innermost s-cores for this network are of sizes 43.9 (random link weights), 6.4 (correlated), and 2 (anticorrelated). For the ER network, these numbers are 30.9, 2.1, and 2 respectively. Additionally, we have analyzed gene-correlation networks (data not shown) where the innermost s-core consists of ∼2000 nodes (nearly 15% of the total network).

The application of s-core analysis on two gene-expression experiments in the yeast *S. cerevisiae* [26,27] demonstrates the utility of our approach: s-core analysis is able to identify distinct parts of the PIN associated with FSR or CHP. This is a discovery that is not possible with the standard k-core approach. In sum, the yeast PIN analysis uncovered a network of highly interconnected proteins involved in coordinating yeast stress responses. Different parts of the network are enriched for a variety of gene functions known to partake in stress responses, both general and biological functions with known associations to the two stress environments. These results suggest that the s-core analysis is a useful approach to identifying and ranking genes of relevance in, e.g., interaction networks when combined with condition specific data.

In general, we hypothesize that s-core analysis will be an effective approach in ranking the importance of nodes, as well as selecting sets of nodes, in large-scale (link) weighted networks: The application of k-core analysis to large-scale unweighted networks for identifying central sets of nodes has recently led to a deeper understanding of their organization [10–15]. Simply by incorporating known link strengths in these networks, s-core would be directly applicable. For the cases where link weights are discrete (or have been discretized) and, as a consequence, give rise to a node-strength distribution with an appreciable number of nodes with the same strength values, we would expect a lower number of s-cores than found in our analyses in Sec. III.

ACKNOWLEDGMENT

We are very grateful to Prof. Lisa Stubbs for discussions and suggestions.

[1] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).
[2] B. Zhang and S. Horvath, Stat. Appl. Genet. Mol. Biol. **4**, 17 (2005).
[3] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, Proc. Natl. Acad. Sci. USA **106**, 21484 (2009).
[4] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001).
[5] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[6] J. Han, N. Bertin, T. Hao, D. Goldberg, G. Berriz, L. Zhang, D. Dupuy, A. Walhout, M. Cusick, F. Roth, and M. Vidal, Nature **430**, 88 (2004).
[7] S. B. Seidman, Social Networks **5**, 269 (1983).
[8] S. Wuchty and E. Almaas, Proteomics **5**, 444 (2005).
[9] S. Wuchty and E. Almaas, BMC Evol. Biol. **5**, 24 (2005).
[10] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, Advances in Neural Information Processing Systems **18**, 41 (2006).
[11] J. I. Alvarez-Hamelin, L. Dall'Ásta, A. Barrat, and A. Vespignani, in *Advances in Neural Information Processing Systems*, Vol. 18, edited by Y. Weiss, B. Schölkopf, and J. Platt (MIT, Cambridge, MA, 2006), p. 41.
[12] J. I. Alvarez-Hamelin, L. Dall'Ásta, A. Barrat, and A. Vespignani, Networks and Heterogeneous Media **3**, 395 (2008).
[13] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, Proc. Natl. Acad. Sci. USA **104**, 11150 (2007).
[14] D. R. Wuellner, S. Roy, and R. M. D'Souza, Phys. Rev. E **82**, 056101 (2010).
[15] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Nat. Phys. **6**, 888 (2010).
[16] V. Batagelj and M. Zaversnik, arXiv:cs/0310049.
[17] J. Cheng, Y. Ke, S. Chu, and T. Ouzsu, in *Data Engineering (ICDE), 2011 IEEE 27th International Conference*, Vol. 27 (IEEE, New York, 2011), p. 51.
[18] J. Chalupa, P. L. Leath, and G. R. Reich, J. Phys. C. **12**, L31 (1979).
[19] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. Lett. **96**, 040601 (2006).
[20] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **73**, 056101 (2006).
[21] J. P. Gleeson and S. Melnik, Phys. Rev. E **80**, 046121 (2009).
[22] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, Phys. Rev. E **71**, 065103(R) (2005).
[23] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[24] M. Molloy and B. Reed, Combinator. Probab. Comput. **7**, 295 (1998).
[25] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[26] L. Tuli, A. Martins, W. Sha, P. Mendes, and V. Shulaev, Mol. Cell. Proteomics **6** (2007).

[27] V. Marks, S. Sui, D. Erasmus, G. van der Merve, J. Brumm, W. Wasserman, J. Bruan, and H. van Vuuren, FEMS Yeast Res. **8**, 35 (2008).
[28] P. Erdos and A. Rényi, Publicationes Mathematicae **6**, 290 (1959).
[29] P. Erdos and A. Rényi, Bull. Inst. Int. Stat. **38**, 343 (1961).
[30] E. N. Gilbert, Ann. Math. Stat. **30**, 1141 (1959).
[31] M. E. J. Newman, *Networks: An Introduction* (Oxford University, Oxford, 2010).
[32] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[33] D. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[34] P. J. Macdonald, E. Almaas, and A.-L. Barabási, Europhys. Lett. **72**, 308 (2005).
[35] R. Albert, H. Jeong, and A.-L. Barabási, Nature **406**, 378 (2000).
[36] B. Schwikowski, P. Uetz, and S. Fields, Nat. Biotechnol. **18**, 1257 (2000).
[37] H. Yu, P. Braun, M. A. Yildirim *et al.*, Science **322**, 104 (2008).
[38] T. Barrett, D. Troup, S. Wilhite *et al.*, Nucl. Acid Res. **37**, D885 (2009).
[39] S. Hohmann and W. H. Mager, *Yeast Stress Responses* (Springer-Verlag, Berlin, 2003).
[40] D. Huang, B. Sherman, and R. Lempicki, Nat. Protocol. **4**, 44 (2009).
[41] I. Dalle-Donne, R. Rossi, A. Milzani, P. Di Simplicio, and R. Colombo, Free Radic. Biol. Med. **31**, 1624 (2001).
[42] J. Moseley and B. Goode, Microbiol. Mol. Biol. Rev. **70**, 605 (2006).