Centers of Complex Networks

Stefan Wuchty[†,‡] **and Peter F. Stadler**[#,¶,§]

[†]European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany

[‡]Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, IN 46556, U.S.A.

[*,#]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria

[¶]Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany

[§]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

[*]Address for correspondence.
Tel: ++43 1 4277 52737, Fax: ++43 1 4277 52793, Email: `studla@tbi.univie.ac.at`

Abstract. The central vertices in complex networks are of particular interest because they might play the role of organizational hubs. Here, we consider three different geometric centrality measures, eccentricity, status, and centroid value, that were originally used in the context of resource placement problems. We show that these quantities lead to useful descriptions of the centers of biological networks which often, but not always, correlate with a purely local notion of centrality such as the vertex degree. We introduce the notion of local centers as local optima of a centrality value "landscape" on a network and discuss briefly their role.

1. Introduction

Complex networks occur in diverse areas from metabolic and gene regulation networks in each cell, food webs in ecology, transportation networks, economic interactions and the organization of the internet, just to mention a few examples. Starting with the seminal paper by (Watts & Strogatz, 1998), it has been recognized that these real life network differ qualitatively from the classical random graph models (Erdős & Rényi, 1960; Bollobás, 1985) by the so-called *small-world* property: while the graphs are very sparse on average, the mutual distances between their vertices are nevertheless much shorter than expected.

The recent review by (Albert & Barabási, 2002) indicates that current research focuses on the one hand evolving graphs with various preferential attachment rules and, on the other hand, on characterizing new empirically determined graphs in terms of a small number of parameters, in particular their degree distribution, clustering coefficient, and average path length. The vertex degree is typically used as a measure of centrality in these networks. In the graph evolution models, high vertex degrees usually indicate "old" vertices. (Fell & Wagner, 2000; Wagner & Fell, 2000) indeed show that the metabolites with the highest connectivity are part of the oldest "core" metabolism.

It is a bit surprising, however, that classical graph-theoretical properties of such large real-life networks so far have not been studied systematically, although at least some of them are easily within the reach of present-day computer facilities. In this contribution, we consider three notions of centrality that were originally designed for resource location problems. As such, these measure seem to be particularly appropriate for the study of metabolic and signaling networks, which after all have probably evolved to solve the tasks of efficiently allocating resources to required metabolites and of controlling a cell's biochemistry with as little time delay as possible.

This contribution is organized as follows: In section 2 we provide some background on the global structure of networks and introduce centrality measures and their basic properties. These measures are compared in the subsequent section, where we also introduce the notion of a local center in a graph. Applications to three different types of biological networks are discussed briefly in section 4, namely metabolic networks, protein interaction networks, and protein domain networks.

2. Network Structure

2.1. Basic Definitions. A network is conveniently modeled as a graph G which consists of a set V of vertices and a set E of edges which we regard as un-ordered pairs of distinct vertices. Hence we consider only *simple undirected graphs* in the language of (Berge, 1985). A path in G is an alternating sequence $(x_0, e_1, x_1, \ldots, e_\ell, x_\ell)$ of vertices and edges, where the $e_i = \{x_{i-1}, x_i\}$ are the edges connecting subsequent vertices. The length of a path is its number edges. The set of neighbors of x is denoted by $\partial\{x\} = \{y \in V | \{x, y\} \in E\}$.

The *degree* of a vertex x is the number of edges that contain x, i.e., the number of neighbors of x:

$$\deg(x) = |\{e \in E | x \in e\}| = |\{y \in V| \{x, y\} \in E\}| = |\partial\{x\}| \, , \tag{1}$$

where $|A|$ denotes the cardinality (number of elements) of the set A. Equivalently, we may define $\deg(x)$ as the number of edges incident with x.

The *distance* $d(x, y)$ is the length of the shortest path in G connecting x with y. If a path connecting x and y does not exist we set $d(x, y) = \infty$. Thus, the graph G is connected if and only if $d(x, y)$ is finite for all $x, y \in V$.

We remark that our approach can trivially be extended to weighted graphs. So, simply define $\deg(x)$ as the sum of the weight of the edges that contain x and define the length of a path as the sum of the weights of each edges.

2.2. Degree Distributions. (Amaral *et al.*, 2000) showed that there are (at least) three structurally different classes of networks that are distinguished by the distribution $P(k)$ of the vertex degrees $k = \deg(x)$:

(a) *Single-Scale Networks* with a sharp distribution of vertex degrees exhibiting exponential or Gaussian tails. This class includes also the Erdős-Rényi model of uncorrelated random graphs (Erdős & Rényi, 1960; Bollobás, 1985).
(b) *Scale-Free Networks* with a power law distribution $P(d) \sim d^{-\gamma}$. A simple model for this type of networks was introduced recently by Barabási *et al.* (Barabási & Albert, 1999; Barabási *et al.*, 1999). Metabolic networks (Wagner & Fell, 2000; Jeong *et al.*, 2000) and food-webs (Montoya & Solé, 2002) belong to this class.
(c) *Broad-Scale Networks* for which $P(d)$ has a power-law regime followed by a sharp cut-off, e.g. exponential or Gaussian decay of the tail. An example is the movie-actor network described in (Watts, 1999)

The Erdős-Rényi model (ER) (Erdős & Rényi, 1960) assumes a fixed number $n = |V|$ of vertices and assigns edges independently with a certain probability p. For details see the book by (Bollobás, 1985). In many cases, ER random graphs turn out the be quite different from a network of interest. The Watts-Strogatz (SW) (Watts & Strogatz, 1998) model of small-world networks starts with a deterministic graph, usually a circular arrangement of vertices in which each vertex is connected to k nearest neighbors on each side. Subsequently, edges are "rewired" (in the original version) or added (Newman & Watts, 1999; Newman *et al.*, 2000) with probability p. Both ER and SW graphs exhibit an approximately Gaussian degree distributions.

The other extreme is the scale-free model (BA) (Barabási & Albert, 1999; Barabási *et al.*, 1999) with a degree distribution of the form $P(d) \sim k^{-3}$. Starting from a small core graph, at each time step a vertex is added together with m edges that are connected to each previously present vertex k with probability

$$\Pi(k) = d(k) \Big/ \sum_k d(k) \, , \tag{2}$$

where $d(k)$ is the degree of vertex k. A recent extension of the model allows the tuning of the scaling exponent γ in the range $2 \le \gamma \le 3$ (Albert & Barabási, 2000a).

The vertex degrees are an intrinsically local characterization of a graph. Consequently, they allow a meaningful interpretation only when the graph is a typical instance of a *known* statistical ensemble such as the ER model or the BA model.

It is therefore necessary to consider additional characteristics of G that are preferably not closely related to the degree distribution. A quantity that is commonly used in the literature on the small-world networks is the clustering coefficient that measures how close the neighborhood of a each vertex comes on average to being a complete subgraph (clique) (Herzel, 1998; Barrat & Weigt, 2000; Watts & Strogatz, 1998). Again, this measure is intrinsically local. A more global measure is the average length $\ell(G)$ of a path between two vertices, see e.g. for an extensive discussion (Newman *et al.*, 2000). The distribution of short cycles, i.e., detours, may be regarded as an intermediate case (Gleiss *et al.*, 2001).

2.3. Geometric Centrality. Geometric notions of centrality are closely linked to facility location problems. Suppose, we are given a graph G representing, say, a traffic network. We may then ask questions such as the following:

(A) What is the optimal location of a hospital such that the worst case response time of an ambulance is minimal?
(B) What is the optimal location of a shopping mall so that the average driving time to the mall is minimal?
(C) What is the optimal location of a shop if customers buy at the nearest shop, and there will be a competitor placing its shop after we have placed ours?

These three classical facility local problems can be recast as optimization problems based on the distance matrix $\mathbf{D} = (d(x, y))$ of G. Their solutions define three different notions of "central" vertices. The distance matrix $\mathbf{D}$ can be computed rather efficiently e.g. using Dijkstra's algorithm with time complexity $\mathcal{O}(|V|^2 \ln |V|)$, see e.g. (Cormen *et al.*, 1990).

The *excentricity* of a vertex x in G and the radius $\rho(G)$, respectively, are defined as

$$e(x) = \max_{y\in V} d(x, y) \qquad \text{and} \qquad \rho(G) = \min_{x\in V} e(x) \tag{3}$$

The *center* of G is the set

$$\mathcal{C}(G) = \{x \in V | e(x) = \rho(G)\} . \tag{4}$$

$\mathcal{C}(G)$ is the center to the "emergency facility local problem" (A) which is always contained in a single block of G (Harary & Norman, 1953).

The *status* $d(x)$ of a vertex (Harary, 1959) and the status $\sigma(G)$ of the graph G, respectively, are defined as

$$d(x) = \sum_{y\in V} d(x, y) \qquad \text{and} \qquad \sigma(G) = \min_{x\in V} d(x) . \tag{5}$$

The *median* (Slater, 1980) of G is the set

$$\mathcal{M}(G) = \{x \in V | d(x) = \sigma(G)\} . \tag{6}$$

The median is the solution of the "service facility location problem" (B). Both the center and the median of a graph were already considered by (Jordan, 1869). Instead of the status, one may of course use the average distance $\ell(x)$ of a vertex from x. Clearly, $d(x) = (|V| - 1)\ell(x)$. The *Wiener index* (Wiener, 1947) is

$$W(G) = \frac{1}{2}\sum_{x\in V} d(x) = \frac{1}{2}\sum_{x,y\in V} d(x,y) = \binom{n}{2}\ell(G)\,, \tag{7}$$

where $\ell(G)$ is the mean path length in G. It provides an important characteristic of molecular graphs. For details, see (Gutman *et al.*, 1996).

For any pair of distinct vertices $u, v \in V$, $u \neq v$, define

$$V_{xy} = \{w \in V | d(x,w) < d(y,w)\}\,, \tag{8}$$

i.e., V_{uv} is the set of vertices that are closer to u than to v. The competitive location problem (C), which was first considered by (Slater, 1975), is then solved by the vertices x that maximize $|V_{xy}| - |V_{yx}|$ over all possible locations of the competitor y. The following identity

$$d(x) + |V_{xy}| = d(y) + |V_{yx}| \tag{9}$$

holds for all connected graphs (Entringer *et al.*, 1976). Following (Slater, 1975) we define *centroid value* of a vertex and the graph G itself as

$$f(x) = d(x) - \min_{y\neq x} d(y) \qquad \text{and} \qquad \varphi(G) = \min_{x\in V} f(x)\,. \tag{10}$$

The *centroid* of G is the set

$$\mathcal{Z}(G) = \{x \in V | f(x) = \varphi(G)\}\,. \tag{11}$$

We have inverted the sign of $f(x)$ compared to the discussion in Slater's work (Slater, 1999) as we prefer the centrality measure $f(x)$ to be minimal at the most central vertices in analogy to $d(x)$ and $e(x)$.

The mutual location of the three types of "central" vertices is of obvious interest. The median $\mathcal{M}(G)$ and the centroid $\mathcal{Z}(G)$ are always contained in the same block of a connected graph G (Smart & Slater, 1999). Both, the center and the centroid may serve as the root of a distance preserving spaning tree (Barefoot *et al.*, 1997).

The centroid value $f(x)$ may, perhaps surprisingly, become 0 or even negative. If this is the case, then $d(x) = \min_y d(y) = \sigma(G)$. It follows that $f(x) < 0$ for at most one vertex x^*, in which case $d(x^*)$ is the unique minimum of $d(x)$, hence $\mathcal{Z}(G) = \mathcal{M}(G) = \{x^*\}$. If f is non-negative then $f(x) = 0$ iff $d(x)$ is minimal and there are at least two distinct vertices minimizing d. Again we have $\mathcal{Z}(G) = \mathcal{M}(G)$. Conversely, if $\varphi(G) > 0$ then the minima of $d(x)$ do not minimize $f(x)$ and hence median and centroid are disjoint. Such graphs are called *secure* graphs. It is shown in (Slater, 1976) that there are no secure graphs with $|V| < 9$ vertices. An example with $|V| \geq 9$ is given (Smart & Slater, 1999, Fig.4).

It is shown in (Smart & Slater, 1999) that $\mathcal{C}(G)$, $\mathcal{M}(G)$, and $\mathcal{Z}(G)$ may be pairwise disjoint and even separated by arbitrary distances if G is large enough (Slater, 1999).

A slightly different, much less studied notion of centrality is introduced in (Nieminen, 1984). An induced subgraph H of G is convex if it contains a shortest path (in G) between any two of its vertices. A branch of G at a vertex x is a maximal convex induced subgraph that does not contain x. The *branch weight* $b(x)$ is the maximum number of vertices in a branch of G at x. The branches of a tree T at a vertex x are thus the connected components of the forest obtained by deleting the vertex x. The *convex center* or *branch weight center* $\mathcal{B}(G)$ is the set of vertices that minimize $b(x)$. (Zelinka, 1968) showed that for any tree T the convex center and the median coincide. We shall not consider the branch weight center in this contribution because there does not seem to be a convenient way to compute $b(x)$ in large graphs.

Betweenness centrality (Freeman, 1977) is a distant relative of the resource placement centralities discussed above. Originally designed to measure a person's influence in a society, it is quantified in terms of the number of shortest paths that run through a given vertex. Most recently, a classification of scale-free networks based on the scaling of betweenness centrality has been proposed (Goh *et al.*, 2002). A comparison of this measure with resource placement centralities will be described elsewhere.

3. Properties of Centrality Measures

3.1. Correlations. Both the vertex degree and the three geometric measures $d(x)$, $e(x)$, and $f(x)$ are defined with a notion of centrality in mind. It is not surprising, therefore, that there are strong correlations between them. Intuitively, there should also be a strong correlation between a large vertex degree and a short average distance to the other nodes of the network, and hence to various measures of centrality. In Fig. 1, we compare the behavior of a ER random graph with a scale-free graph with the same average degree. The results are essentially the same for different numbers of vertices and average vertex degrees (data not shown). The scale-free networks are connected by construction. This is not true in general for the ER graphs where we consider only the giant component.

We observe that the centrality measures correlated linearly with the logarithm of the vertex degree. For both status and centroid value the regression coefficients are about 0.99. The lower quality of the fit for the excentricity is a consequence of the granularity of this measure. The random graphs fit slightly better to the exponential distributions and have smaller standard deviations.

Interestingly, the relative location of ER and BA models differs between status $d(x)$ and centroid value $f(x)$: Surprisingly, the centroid values are smaller in ER graph than in scale-free networks. Both status $d(x)$ and excentricity $e(x)$ directly measure path length which we expect and observe to be significantly smaller in the scale-free case. In contrast, $f(x)$ tries to optimize also the distance from a competitor due to the longer path lengths in scale-free graphs compared to ER random graphs, we observe larger values of $f(x)$ as well.

3.2. Landscapes on Graphs. The concept of a *fitness landscape* originated in the 1930s in theoretical biology (Wright, 1932; Wright, 1967) as a mean of visualizing evolutionary adaptation. The same abstract concept arise naturally in many other areas of scientific study, for instance the physics of disordered systems, combinatorial optimization, protein folding, and evolutionary computation; for a recent review, see (Reidys & Stadler, 2002). In the simplest case, a *landscape* can be defined as function $F : V \to \mathbb{R}$ that assigns a number to each vertex of a graph G. In this sense, we can

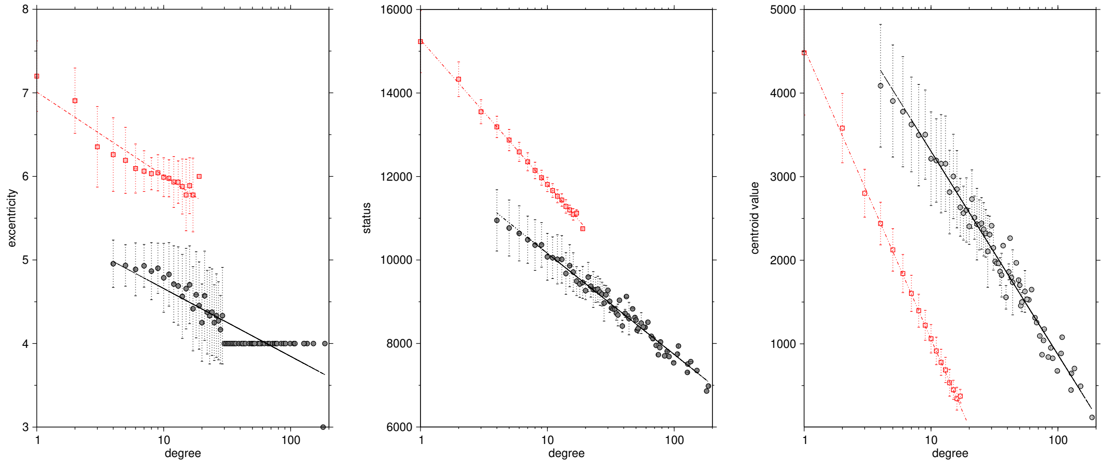

Figure 1. Correlation of excentricity (left), status (middle), and centroid value (right) with the degree of nodes. We compare a scale-free graph (○) and a corresponding random graph (□) with 3000 vertices and 12000 edges, respectively. The scale-free network is generated by means of *preferential attachment* (Barabási & Albert, 1999). Best fits are obtained in the form $y = A - B \ln x$ with the following parameters:
Excentricity: $A = 5.47$, $B = 0.35$ (scale-free); $A = 7.01$, $B = 0.94$ (random).
Status: $A = 12585$, $B = 1053$ (scale-free); 15276, $B = 1505$ (random).
Centroid Value: $A = 5736$, $B = 1056$ (scale-free); $A = 4537$, $B = 1512$ (random).

regard the excentricity $e(x)$, the status $d(x)$, and the centroid value $f(x)$ as landscapes on G.

This idea suggest the definition of *local* centers, medians, and centroids as local minima $\hat{x}$ of the cost function, i.e.,

$$g(\hat{x}) \leq g(y) \qquad \text{for all } y \in \partial\{\hat{x}\} \tag{12}$$

where $g = d, e, f$. Furthermore, we may consider the hierarchical structure of the local centers, and one may ask for connecting paths between local centers that are themselves as central as possible. The program `barriers`[1] that was originally designed to investigate the structure of the energy landscape of RNA molecules (Flamm *et al.*, 2000; Flamm *et al.*, 2002) can be used to solve exactly this problem. Starting from a list of vertices that is sorted by the value of the cost function $g(x)$, `barriers` identifies all local minima and the saddle points that connect them. In Fig. 2 we show the barrier trees for the three centrality measures $d(x)$, $e(x)$, and $f(x)$ for the ER random graph in Fig. 1.

Interestingly, with all three measures we do not find local centers in the scale-free BA graphs. This indicates that the BA graphs are arranged around a unique central structure.

An alternative approach to describe the geometry of a landscape $g(x)$ is the consideration of its *level sets*

$$V_g(u) = \{x \in V | g(x) \leq u\} \tag{13}$$

[1] Available from `http://www.tbi.univie.ac.at/~ivo/RNA/Barriers/`.

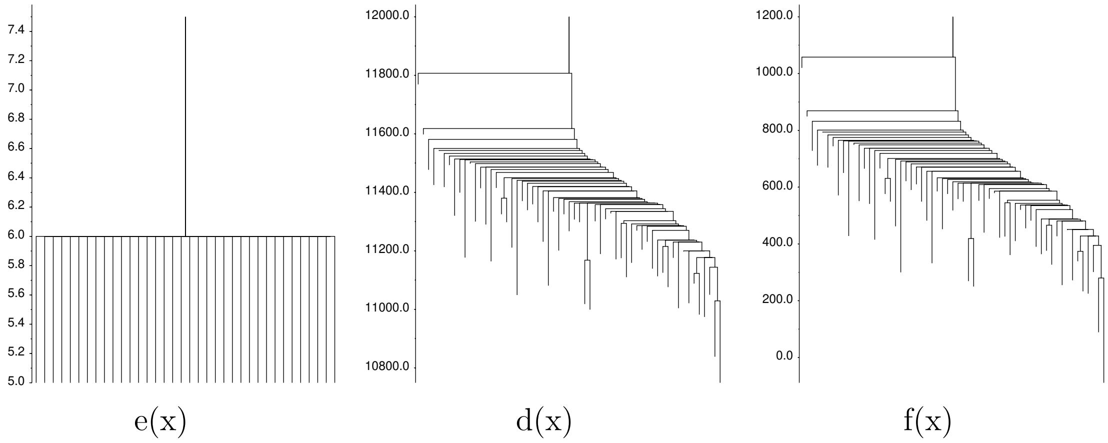

Figure 2. Barrier trees for excentricity (left), status (middle), and centroid value (right) for an ER random graph with 3000 vertices and 12000 nodes.

or the associated induced subgraph $G[V_g(u)]$ which consists of the vertex set $V_g(u)$ and all edges of G connecting two vertices of $V_g(u)$. Note, that $G[V_g(u)]$ is connected for levels u if the landscapes does not have no local minima. On the other hand, if there are local mimima, then $G[V_g(u)]$ decomposes into disconnected components for some values of u. This behavior is illustrated for the centroid value on the r.h.s. of Fig. 3.

The behavior of $|V_g(u)|$ as a function of the level u can be used to quantify the robustness of the center, median, or centroid. The level sets with small centroid

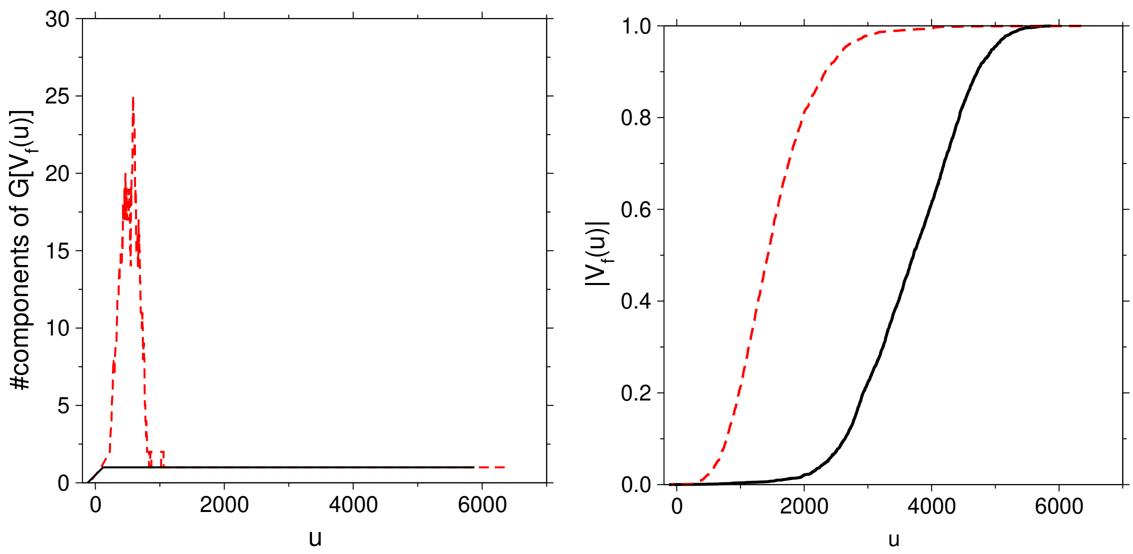

Figure 3. Level sets $V_f(u)$ of the centroid value. L.h.s.: Number of connected components of the level set graph $G[V_g(u)]$ as a function of the centroid value u. R.h.s.: Size of the level sets $|V_g(u)|$ as a function of the centroid value u. We compare an ER graph (dashed line) and an AB scale-free graph (full line) with 3000 vertices and 12000 edges each.

values are smaller in the AB model indicating a tighter near central subgraph than in the ER model. Not surprisingly, the ER model is more homogeneous on average.

4. Applications

4.1. Metabolic network. In metabolic networks, the meaning of centrality is obvious: The central metabolites are the crossroads of the networks and, in the spirit of the graph evolution models, also the historically oldest ones. Centrality therefore should reflect both age and importance. Local centers, if they exists, therefore are likely the remnants of a previous merging, while the absence of local centers indicated a continuous, step-wise, growth the network.

(Fell & Wagner, 2000) assembled the central routes of the energy metabolism and small-molecule building block synthesis in *E. coli* and constructed a *substrate graph* with the metabolites as vertices and edges connecting any two metabolites that appear in the same reaction. In their analysis, ATP and H_2O was excluded. (Jeong *et al.*, 2000) considered metabolic networks of 43 different organisms including *E. coli*. Unlike (Fell & Wagner, 2000), a bipartite graph was used in which both the substrates and the reactions are vertices, and edges connect substrates with the reactions they are taking part in. Both studies report a power-law degree distribution characteristic for scale-free networks.

In Table 1, we list the center $\mathcal{C}(G)$, median $\mathcal{C}(G)$, and centroid $\mathcal{Z}(G)$. We find that the results are identical for the substrate graph of (Fell & Wagner, 2000) and the bipartite reaction graph of (Jeong *et al.*, 2000). Following the proceedure of (Fell & Wagner, 2000), we deleted ATP and H_2O from this graph because they are connected with almost every other metabolite. Not surprisingly, we are left with inorganic phosphate P_i and ADP at the next most connected vertices.

Interestingly, the most coarse grained measure, the center $\mathcal{C}(G)$ yields a good intuitive estimate of the most centered substrates. Among these are ATP, ADP and AMP which are obviously the most central substrates to the energetics and signaling pathways of the cell. In this regard, the frequent occurrence of P_i and PP_i fits the picture perfectly. Furthermore, NADP and NADPH appear in a similar context. Among the most popular metabolites, succinate SUCC, glutamate GLU, pyruvate PYR, and coenzyme A COA appear in the center. These metabolites indeed play a "central" role in well known pathways emphasizing glycolysis and the citrate cycle.

A comparison of the nodes degree with their corresponding centrality measures shows that a highly linked vertex has small numbers of the geometric centrality measures (see Table 2). It is striking that all three centrality measures $e(x)$, $d(x)$, and $f(x)$ yield very similar rank orders close to the central vertices, despite the fact that these measures may disagree significantly for non-central vertices, although the three notions of geometric centrality are conceptually quite different from each other. In fact, the rank of all metabolites listed in table 2 are the same for $d(x)$ and $f(x)$.

4.2. Protein Networks. A second class of networks that have received particular attention recently are networks of (direct) protein interaction (Jeong *et al.*, 2001; Wagner, 2001). The likelihood that the elimination of a protein from the genome is

Table 1. Central nodes in the metabolic network of *E. coli.* For the full network, the results from the substrate graph and the bipartite reaction graph coincide. For comparison, we also give the centers for the complete networks without removing any vertex.

Network G	$\mathcal{C}(G)$	$\mathcal{M}(G)$	$\mathcal{Z}(G)$
(Fell & Wagner, 2000) (Jeong *et al.*, 2000)	AMP, ADP, T3P1 P_i, PP_i, CO_2, NH3, NAD, NADH, NADP, NADPH,PYR, SUCC, AC, HEXT, GLU, THR, GLY, NMN	P_i	P_i
complete	ATP, H_2O	ATP, H_2O	ATP, H_2O

Table 2. Nodes in the metabolic network of *E. coli* (Fell & Wagner, 2000) arranged to increasing numbers of their degree $deg(x)$. For comparison, we also added the respective numbers of excentricity, $e(x)$, status, $d(x)$, and excentricity value, $f(x)$.

Substrate	$deg(x)$	$e(x)$	$d(x)$	rank	$f(x)$	rank
P_i	451	4	1151	1	-4	1
ADP	451	4	1155	2	4	2
PP_I	229	4	1375	9	224	9
CO_2	217	4	1323	4	172	4
GLU	192	4	1340	5	189	5
HEXT	186	4	1285	3	134	3
NAD	174	4	1352	7	201	7
NADH	162	4	1379	10	228	10
NADP	155	4	1387	11	236	11
PYR	154	4	1344	6	193	6
NADPH	151	4	1387	11	236	11
CoA	117	5	1462	21	311	21
NH_3	106	4	1367	8	216	8
GLN	100	5	1461	20	310	20
AMP	96	4	1478	26	327	26

lethal for the organism has been reported to correlate with its degree in the protein network (Jeong *et al.*, 2001). This finding is easy to interpret: It does not come as a surprise that the interruption of a large number of direct interactions increases the probability of breakdown. A recent re-evaluation of the available data indicated, however, that lethal proteins cannot be cleanly distinguished from viable ones by their degree alone (Wuchty, 2002). In the same vein, (Hahn *et al.*, 2002) show that connectivity is not related to robustness against aminoacid substitutions in protein networks. This poses the question whether the correlation of lethality and connectivity is a local or an organizational phenomenon.

Fig. 4 shows that, somewhat surprisingly, there is almost no difference in the distribution of excentricities between essential and non-essential proteins, i.e., essential proteins are organizationally not more central than others. In other words, the long-range impact of the loss of a given protein apparently can be compensated by the rest of the network, but this robustness tends to break down at the level of direct interaction partners. The question whether a deletion of a protein is lethal or leads to

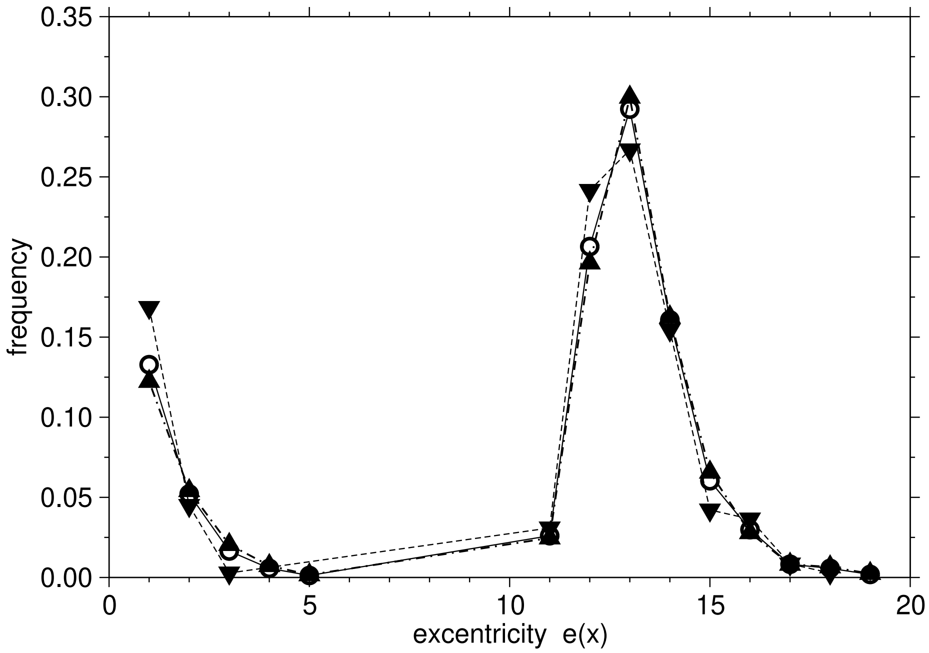

Figure 4. Frequency proteins with given excentricity values for ○ all, ▲ non-lethal, and ▼ lethal deletion mutants. There is no significant difference between the three datasets. The protein network graph is not connected, hence the distribution of the eccentricity, which we compute as the superposition of the eccentricities of the individual components is bi-modal: One peak reflects the small components and the other, larger peak refers to the main (giant) component of the network.

a viable phenotype therefore will require a gradual answer. A promising approach in this direction was recenly undertaken by (Jeong *et al.*, 2002), who found correlations of phenotypic effects of single gene deletions in *Yeast* with fluctuations in the corresponding mRNA expression levels, functional classification of gene products and the number of interactions in the underlying protein-protein interaction network. Based on these qualitative measurements, they were able to predict gradual phenotypic effects that are in good agreement with already known experimental results. The dependence of the degree of lethality on multiple factors, including the functional class of the protein, highlights the limitations of purely structural approaches to network analysis.

4.3. Domain Sequence Network. Many proteins consist of a number of recognizable domains that appear in oftentimes many different proteins. A graph G can be constructed that has the domains as its vertices and edges between them whenever two domains co-occur in a protein. Essentially, they give a tentative insight into the structure of the proteome since they were found to exhibit scale-free behavior. Thus, domains which prove to be highly connected since they frequently occur in multidomain proteins shape the backbone of the proteome of the underlying organism. Similar to the metabolic networks, highly connected domains might have shaped an evolutionary core of proteomes.

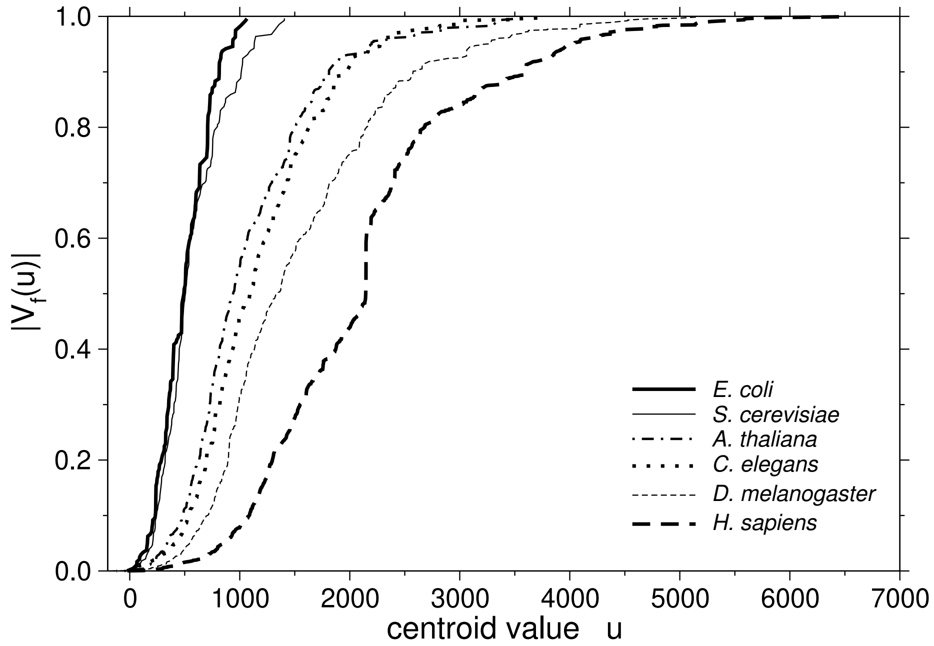

Figure 5. Level set sizes $|V_f(u)|$ of the centroid value for protein interaction networks of *E.coli*, *S.cerevisiae*, *A.thaliana*, *D.melanogaster*, *C.elegans* and *H.sapiens*.

The connectivity of domains that are involved in cell-cell contacts and signal transduction increases with the complexity of the organism. In Fig. 5, we compare the level sets $V_f(u)$ of the centroid value of the domain networks of a number of prokaryotic and eukaryotic organisms. The fraction of "more central" domains, indicated e.g. by the position of the inflection point of the sigmoidal curves, increases significantly from unicellular organism to vertebrates. These data are in good agreement with the results of (Wuchty, 2001), which were obtained using vertex connectivity.

In Table 3, we list the ten domains with the smallest centroid value, i.e., the ones that are most central, for six different organism. Again, the data are in good agreement with the previous study (Wuchty, 2001) and show that the central-most domains are predominantly those that take part in cell-cell contacts and signal transduction. We find here a stronger emphasis on signaling domains with kinases and zinc-fingers.

5. Discussion

Here, we considered the application of three different geometric centrality measures, excentricity, status, and centroid value to various kinds of biological networks. These three measures arise in the context of slightly different resource placement problems. By construction, they measure geometric or structural centrality in contrast to the vertex degree, which is a purely local measure. There is a significant correlation between vertex degree and the geometric centrality measures; in specific cases, however, they may differ significantly. This allows us to distinguish between purely local effects (such as the vertex degrees of essential proteins) from organizational effects (such as the identification of central metabolites).

Table 3. Centroids $\mathcal{Z}(G)$ and the 9 next most centroidal domains from different organisms.

E.coli	*S.cerevisiae*	*A. thaliana*	*D.melanogaster*	*C.elegans*	*H.sapiens*
AAA	WD-40	ZF-C3HC4	PRICHEXTENSN	ZF-C3HC4	PRICHEXTNSN
PAS	SER/THR-PKINASE	EUK-PKINASE	EGF	EUK-PKINASE	EUK-PKINASE
CBS	EUK-PKINASE	SER/THR-PKINASE	ANK	ANK	LRR
HTH-FIS	ZF-C3HC4	PRICHEXTENSN	ZF-C2H2	PH	IG-MHC
SIG54	HEAT	HELICASE C	ZF-C3HC4	EF-HAND	PH
RES-REG	FHA	PPR	SH3	ZF-C2H2	C2
GAF	C2	TYR-PKINASE	PDZ	BTB/POZ	EGF
FMN	DEAD/DEAH	DEAD/DEAH	SER/THR PKINASE	PRICHEXTENSN	FN3
DEAD/DEAH	CLATHRIN	TPR	LDL-RECEPTOR	DAG-PE BIND	TYR-PKINASE
			EF-HAND	EGF	ZF-C3HC4

Since measures of centrality can be assigned to individual vertices of the networks, we can define not only global centers as the vertices that minimize (or maximize) a centrality measures, but we obtain as a natural definition of local centers in terms of local optima of the centrality value landscape on the network. In the connected examples that we have considered here, as well as the generic examples of scale-free networks, we did not encounter local centers; on the other hand, local centers are abundant in random networks.

Another important implication of local centers is that they are indicative of a modular organization and possibly also of a modular origin of the network. If the network in question does not resemble a ER random graph, then local centers could be interpreted as the centers building blocks that were merged together. The absence of local centers is then a strong indication for the stepwise growth of the network.

Acknowledgments. We thank Hawoong Jeong for supplying his protein-protein interaction data.

References

Albert, R. & Barabási, A.-L. (2000a). Topology of evolving networks: local events and universality. *Phys. Rev. Lett.* , 5234–5237.

Albert, R. & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97.

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small world networks. *Proc. Natl. Acad. Sci. USA,* **97**, 11149–11152.

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science,* **286**, 509–512.

Barabási, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A,* **173-187**, 272.

Barefoot, C. A., Entringer, R. C., & Székely, L. A. (1997). Extremal values for ratios of distance trees. *Discr. Appl. Math.* **80**, 37–56.

Barrat, A. & Weigt, M. (2000). On the properties of small-world network models. *Europ. Phys. J. B,* **13**, 547–.

Berge, C. (1985). *Graphs.* Amsterdam, NL: North-Holland.

Bollobás, B. (1985). *Random Graphs.* London UK: Academic Press.

Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1990). *Introduction to Algorithms.* Cambridge, MA: MIT Press.

Entringer, R. C., Jackson, D. E., & Snyder, D. A. (1976). Distance in graphs. *Czechoslowak Math. J.* **26**, 283–296.

Erdős, P. & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci., Ser. A,* **5**, 17–61.

Fell, D. A. & Wagner, A. (2000). The small world of metabolism. *Nature Biotech.* **189**, 1121–1122.

Flamm, C., Fontana, W., Hofacker, I. L., & Schuster, P. (2000). RNA folding at elementary step resolution. *RNA,* **6**, 325–338.

Flamm, C., Hofacker, I. L., Stadler, P. F., & Wolfinger, M. T. (2002). Barrier trees of degenerate landscape. *Z. Phys. Chem.* **216**, 155–173.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry,* **40**, 35–41.

Gleiss, P. M., Stadler, P. F., Wagner, A., & Fell, D. A. (2001). Relevant cycles in chemical reaction network. *Adv. Complex Syst.* **4**, 207–226.

Goh, K.-I., Oh, E. S., Jeong, H., Khang, B., & Kim, D. (2002). Classification of scale-free networks. Technical Report 0205232 arXiv:cond-mat.

Gutman, I., Klavžar, S., & Mohar, B., eds (1996). *Fifty Years of the Wiener index* volume 36 of *MATCH.*

Hahn, M. W., Conant, G., & Wagner, A. (2002). Molecular evolution in large genetic networks: connectivity does not equal importance. Technical Report 02-08-039 Santa Fe Institute.

Harary, F. (1959). Status and contrastatus. *Sociometry,* **22**, 23–43.

Harary, F. & Norman, R. Z. (1953). The dissimilarity characteristic of Husimi trees. *Ann. Math.* **58**, 134–141.

Herzel, H. (1998). How to quantify "small world networks"? *Fractals,* **6**, 301–303.

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

Jeong, H., Oltvai, Z. N., & Barabási, A.-L. (2002). Prediction of protein essentiality based on genomic data. in press, ComPlexUs.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature,* **407**, 651–654.

Jordan, C. (1869). Sur les assemblages de lignes. *J. Reine Angew. Math.* **70**, 185–190.

Montoya, J. M. & Solé, R. V. (2002). Small world patterns in food webs. *J. Theor. Biol.* . in press, Santa Fe Institute preprint 00-10-059.

Newman, M. E. J., Moore, C., & Watts, D. J. (2000). Mean-field solution of the small-world network model. *Phys. Rev. Lett.* **84**, 3201–3204.

Newman, M. E. J. & Watts, D. J. (1999). Renormalization group analysis of the small-world network model. *Phys. Lett. A,* **263**, 341–346.

Nieminen, J. (1984). Centrality, convexity and intersections in graphs. *Bull. Math. Soc. Sci. Math. Répub. Soc. Roum., Nouv. Sér,* **28(76)**, 337–344.

Reidys, C. M. & Stadler, P. F. (2002). Combinatorial landscapes. *SIAM Review,* **44**, 3–54.

Slater, P. J. (1975). Maximum facility location. *J. Res. Natl. Bur Stand. B,* **79**, 107–115.

Slater, P. J. (1976). Central vertices in a graph. *Congr. Num.* **17**, 487–487. Proc. 7th Southeast. Conf. Comb., Graph Theory, Comput.; Baton Rouge.

Slater, P. J. (1980). Medians of arbitrary graphs. *J. Graph Theory,* **4**, 389–392.

Slater, P. J. (1999). A survey of sequences of central subgraphs. *Networks,* **34**, 224–249.

Smart, C. & Slater, P. J. (1999). Center, median, and centroid subgraphs. *Networks,* **34**, 303–311.

Wagner, A. (2001). The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* **18**, 1283–1292.

Wagner, A. & Fell, D. A. (2000). The small world inside large metabolic networks. Technical Report 00-07-041 Santa Fe Institute.

Watts, D. J. (1999). *Small Worlds.* Princeton NJ: Princeton University Press.

Watts, D. J. & Strogatz, H. S. (1998). Collective dynamics of "small-world" networks. *Nature,* **393**, 440–442.

Wiener, H. (1947). Structural determination of paraffine boiling points. *J. Amer. Chem. Soc.* **69**, 17–20.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeeding and selection in evolution. In: *Proceedings of the Sixth International Congress on Genetics*, (Jones, D. F., ed) volume 1 pp. 356–366, New York: Brooklyn Botanic Gardens.

Wright, S. (1967). "surfaces" of selective value. *Proc. Nat. Acad. Sci. USA,* **58**, 165–172.

Wuchty, S. (2001). Scale-free behavior in protein domain networks. *Mol. Biol. Evol.* **18**, 1694–1702.

Wuchty, S. (2002). Interaction and domain networks of yeast. *Proteomics,* . in press.

Zelinka, B. (1968). Medians and peripherans of trees. *Arch. Math. (Brno),* **4**, 87–95.