Visual Reasoning about Social Networks Using Centrality Sensitivity

Carlos D. Correa, *Member*, *IEEE*, Tarik Crnovrsanin, and Kwan-Liu Ma, *Senior Member*, *IEEE*

Abstract—In this paper, we study the sensitivity of centrality metrics as a key metric of social networks to support visual reasoning. As centrality represents the prestige or importance of a node in a network, its sensitivity represents the importance of the relationship between this and all other nodes in the network. We have derived an analytical solution that extracts the sensitivity as the derivative of centrality with respect to degree for two centrality metrics based on feedback and random walks. We show that these sensitivities are good indicators of the distribution of centrality in the network, and how changes are expected to be propagated if we introduce changes to the network. These metrics also help us simplify a complex network in a way that retains the main structural properties and that results in trustworthy, readable diagrams. Sensitivity is also a key concept for uncertainty analysis of social networks, and we show how our approach may help analysts gain insight on the robustness of key network metrics. Through a number of examples, we illustrate the need for measuring sensitivity, and the impact it has on the visualization of and interaction with social and other scale-free networks.

Index Terms—Social network visualization, centrality, sensitivity analysis, eigenvector and Markov importance.

1 Introduction

Social network analysis and visualization have become increasingly important with the growing popularity of Websites such as Facebook and Flickr. Although statistical analysis is often used for discovering patterns and formulating hypotheses about the social interaction, visual analysis can provide better overviews and reveal patterns missed via quantitative measures alone. Recently, Perer and Shneiderman argued for the tight integration of social network statistics and visualization as a fundamental tool toward effective exploration of social networks [42].

One of the most studied statistical metrics for social and other scale-free networks is *centrality*. Central nodes in a graph are often deemed as important hubs through which social interaction is conducted and are good indicators of the relative popularity of individual nodes and clusters. Centrality has also been recognized as an important statistic for biological networks. For instance, Jeong et al. found a significant relation between lethality and centrality in protein networks [29]. As a consequence, it is important to not only enhance visualizations of social networks with centrality metrics, but also to understand the factors involved in the centrality of a given node.

In this paper, we study an aspect of centrality often ignored in visualization: its *sensitivity*. In general, the sensitivity of a function refers to the change in the output values in terms of changes in its inputs. In the case of a social network, we can consider the centrality of nodes as a function of structural variables, such as degree, or, more generally, as a function of the adjacency matrix of a network. Centrality is in fact a multivariate function, affected by each individual node in a network. Therefore, we can think of the centrality sensitivity of a node with respect to another as the change in the centrality metric of the first after a change in the second. The study of sensitivity helps us answer questions such as: if we add one or more edges to a node, how would the centralities of the other nodes change? Do nodes in a subnetwork increase importance together, i.e., is it a collaborative network, or do nodes compete in importance with each other? To support these queries, we visualize the sensitivity values directly on the social network visualization. An overview of the network, enhanced with sensitivity parameters, helps us gain insight on the global distribution of importance. Overviews help us answer questions such as: to what group of nodes can we associate the importance of a given node? Are all nodes surrounding it equally responsible for its centrality? What are the most important nodes relative to a single focal node?

To this end, we approach the problem from the perspective of calculus of variations. In general, computing the sensitivity of a multidimensional data set is a challenging task, as the number of possible variations grows exponentially with the number of variables. Common approaches include analytical differentiation, local methods, which approximate the sensitivity in a neighborhood along each variable at a time, and Montecarlo simulations, which use stochastic searches of the subspace of variations. In our case, we follow a hybrid approach using local neighborhoods and analytic derivation, which computes sensitivity of a function as its partial derivative with respect to each of the variables. We describe a general method for computing these derivatives for centralities that can be expressed as functions of the adjacency matrix, such as the

- *C.D. Correa is with Lawrence Livermore National Laboratory, Box 808, L-422, Livermore, CA 94551-0808. E-mail: correac@llnl.gov.*
- *T. Crnovrsanin and K.-L. Ma are with the University of California, Davis, 2063 Kemper Hall, One Shields Avenue, Davis, CA 95616-8562. E-mail: {tecrnovrsanin, ma}@ucdavis.edu.*

Manuscript received 9 Dec. 2009; revised 23 June 2010; accepted 1 Nov. 2010; published online 7 Dec. 2010.
Recommended for acceptance by A. MacEachren.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-2009-12-0283.
Digital Object Identifier no. 10.1109/TVCG.2010.260.

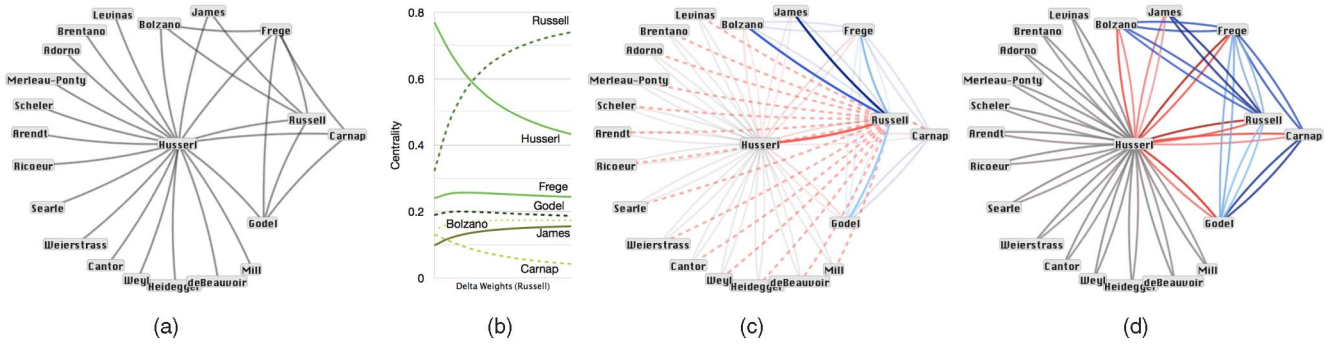

Fig. 1. Centrality sensitivity. (a) Subnetwork of the genealogy of influence data set [25]. (b) Plot of changes in centrality for some key nodes as the degree of *Russell* changes. We see that some nodes become less important while others improve. The rate of change of these functions, their derivatives, are the sensitivity parameters. (c) We can color code the sensitivity (red/blue indicates negative/positive influence) to understand how centrality is propagated for a change in a given node. (d) A full visualization of the sensitivity parameters.

eigenvector and Markov centralities [5], [48]. To better understand the notion of sensitivity and derivatives of centrality, let us analyze the problem for a small network.

1.1 An Illustrative Example

Consider the network depicted in Fig. 1a, a subset of a network of intellectual influence among great thinkers in History [25]. In this network of renowned artists, mathematicians and philosophers, a link is made if a person's work has influenced the works of another. We see this subnetwork as a combination of a star network rooted by *Husserl* and a cluster formed by the nodes on the right, including *Russell*, *Frege*, *Godel*, and others. To understand sensitivity, we perform the following experiment. Take the node *Russell*, and start increasing its degree. Because there are so many combinations that lead to the same increment in degree, let us assume that this change is stochastic. In this example, we assume that all links incident to *Russell* have the same probability of $1/5$ (since there are five edges). Then, we start increasing all these edges by $1/5$, then $2/5$, and so on. At each step, we measure the centrality of all nodes. The result is depicted in Fig. 1b. The x-axis represents the change in the weights associated to the edges incident to *Russell* and the y-axis is centrality. Each line corresponds to the centrality of one of the nodes in the cluster. We clearly see that the changes applied to *Russell* have both positive and negative effects, e.g., it boosts the centralities of *James*, *Bolzano*, *Frege* and *Godel*, but also hinders the centralities of *Husserl* and *Carnap*. Notice also that in the latter case, the impact is indirect, since *Russell* and *Carnap* are not directly connected. Notice also that the rate at which the change occurs is not uniform. This rate, the derivative of those curves, is the sensitivity of centrality, and can be computed analytically for some centrality functions, as described in Section 3.

An example visualization of these derivatives is shown in Fig. 1c, where color denotes the sign and strength of the sensitivity. Red and blue links denote negative and positive sensitivity, respectively, while the saturation of color indicates strength. For example, *Frege*'s sensitivity to *Russell* is smaller than that of *James*'. Dashed lines denote indirect sensitivities, which occur between pairs of nodes not directly connected. These are useful to visualize the "region" of influence of a node. If we repeat the experiment for all nodes, we end up with a pairwise matrix of sensitivities. In Fig. 1d, we visualize the pairwise sensitivity between all connected nodes using the same color scheme. Note that sensitivity is, in general, not symmetric. For example, the sensitivity between *Husserl* and *Frege* is asymmetric.

1.2 Contributions

In this paper, we provide an analytical mechanism for computing sensitivities of centrality and show their practical value for visual reasoning about social networks. In particular, we provide:

1. a general strategy for computing the variation of centrality as an analytical expression for eigenvector and Markov centralities, and a numerical approximation for centrality functions in general,
2. sensitivity overviews in node-link diagrams as a mechanism for characterizing and filtering complex social networks,
3. a network simplification strategy that preserves the centrality distribution of the original network, and
4. a mechanism for assessing uncertainty in networks and its application in understanding the robustness of network metrics.

The study of sensitivity in social networks is important to characterize networks that are seemingly similar, to understand the sources of variability in metrics such as centrality, and to gain insight on the social dynamics of a network. To the best knowledge of the authors, this is the first variational study of social networks from the perspective of visualization.

2 Related Work

2.1 Network Centralities

The issue of centrality has been widely explored in numerous settings, including sociometry, biology and information systems [43]. One of the most obvious ways of measuring centrality of a node is via its degree, as first noted by Shaw [44]. However, this simple definition of centrality may not suffice to capture the complex structural relationships in a graph. Harary and Hage [19] proposed a centrality based on eccentricity, defined as the maximum distance of any node to other nodes in the network. Other metrics are

defined in terms of the total distance to other nodes in the network. Nodes with small total distance are said to be central. Numerous closeness metrics have been proposed, including the information centrality and radiality of a node. For a closer look at these metrics, refer to Jacob et al. [28] and Newman's survey [38].

Other more popular methods have been proposed for applications such as the analysis of social networks and Website ranking. These methods can be broadly categorized as centralities based on *shortest paths*, *feedback* and *random-walks* [7]. The most common metric based on shortest paths is the *betweenness* centrality of a node, introduced by Anthonisse [2] and Freeman [14], as an alternative for closeness centrality in disconnected graphs. Feedback-based metrics define the centrality of a node in terms of the centralities of other nodes. Bonacich introduced a metric based on the eigenvectors of the adjacency matrix of a network [5]. Before Bonacich, Hubbell formulated the problem in a similar fashion, where the centrality of a node is a linear combination of the centralities of others, whose solution can be found from the ensuing system of linear equations [27]. These feedback-based methods became more popular for Webpage indexing and are the core of search algorithms such as PageRank [9], HITS [32], and SALSA [37]. Unlike previous eigenvector centralities, algorithms such as PageRank made the matrix stochastic, ensuring that the corresponding Markov chain converges to a stationary distribution. Finally, the idea of using random processes to represent a network led to Markov centralities, as proposed by White et al. [48]. Their metric is defined as the mean first-passage time of the Markov chain derived from the adjacency matrix of the network.

Several comparisons of these centrality metrics have been performed. Freeman presents an exhaustive treatise of these methods in his seminal paper [14]. Dwyer et al. performed a visual analysis to compare different centrality metrics [12]. They present a series of conventional visual analysis methods and hierarchical views to correlate the centralities of nodes under different metrics. Koschützki and Schreiber present a comparison of centrality measures for biological networks [35]. Although no method was particularly better than the others, the authors recognized that each centrality method provided interesting insight on how proteins interact.

Inspired by these results, we saw a need to understand the behavior of centralities. In this paper, we extract sensitivity as a visual quantity that helps users gain additional insight on the distribution and evolution of centrality metrics, and consequently, on the structure and dynamics of the social network. Similar studies have been carried out to measure the sensitivity of centralities to small perturbations in the network. Langville and Meyer [36] studied the numerical stability of the eigenvector centrality in the context of Web search. Ng et al. [40] were able to provide bounds of the difference magnitude between old and new centralities after a perturbation. These bounds were later improved by Bianchini et al. [4]. In their study, they were concerned mostly with the stability of the centrality vector given a perturbation in the network. In our paper, we have a similar goal, although we discriminate these perturbations as changes in the degree of a node. Therefore, the difference in centrality can be understood as the partial derivative of the centrality with respect to the degree. Naturally, these derivatives can be combined to provide a bound (although not necessarily tight) of the stability of the centrality of a node. A deeper analysis of the stability of centralities is performed by Koschützki et al. [34]

2.2 Network Visualization

The literature in network and graph visualization is extensive [13]. One of the most widely studied topics is the issue of graph layout. Although force-directed placement is popular and easy to implement, other more sophisticated approaches have been proposed, such as GRIP [15], ACE [33] and FM3 [18]. To improve the exploration of such networks, a number of tools have emerged, such as yEd [26], GUESS [23] and JUNG [24], which provide a number of layouts, overview+detailed views, magnifying glasses and color encoding of graph properties. Heer and Boyd presented Vizster, a system for visualizing social networks [21]. In addition to the clustering effect of forced-directed layouts, they also provide an explicit visualization of communities.

Recently, there has been particular interest in guiding the visualization of social and scale-free networks using centrality. Perer and Shneiderman argue that an effective social network system must tightly couple statistics and visualization to provide a more effective exploration [42]. Brandes and Wagner discuss visone [8], a system for visualizing social networks based on centrality, which includes layered and radial layouts, similar to the Pajek system [3]. To improve the layout of large graphs, Girvan and Newman [16] propose edge filtering based on the betweenness centrality of edges. By removing the edges with high BC, they obtain simpler layouts that capture the structure of the network. A similar approach was explored by van Ham and Wattenberg [46]. The minimum spanning tree retains these clusters. When the highest BC edges are added back to the tree, the result is a filtered, but structurally meaningful, network. A different approach is taken by Jia et al. [30], who used the highest BC edges to construct the tree. They based their approach on the observation that scale-free networks are mostly minimally connected. Using the highest BC nodes, they extract the communication channels that are most important in the network. In our paper, we show that centrality sensitivity also provides a ranking of the edges, and that, when used to compute a minimum spanning tree, the result maintains the centrality of the important nodes. This property is important to ensure the correct interpretation of a simplified network diagram.

3 Centrality Sensitivity

A graph $G = \{V, E\}$ consists of a set of *nodes* V and a set of edges E. A node centrality is a function $C : V \mapsto R$, that assigns a real value to each node in V. The larger the value $C(v)$ is, the more important a node v is. One of the simplest ways to measure the centrality of a node is via its *degree*. The degree of a node is the number of edges incident to that node. In a more general sense, for weighted graphs we can define

the degree or valency of a node as the sum of weights of all the edges incident to that node. Unweighted graphs are just a special case where the weight of an edge is 1. To this end, it becomes convenient to represent a graph via its adjacency matrix A of size $n \times n$, where n is the number of nodes.

For directed graphs, it is often common to divide this metric as indegree and outdegree, corresponding to the sums of weights for incoming and outgoing edges of a node, respectively. During our discussion, we will derive centralities in terms of the adjacency matrix, regardless of whether the matrix denotes a directed or an undirected graph.

We can see that the degree of a node is a somewhat local metric of centrality. Other measures, such as betweenness and eigenvector centralities, as discussed below, act globally, and the weights associated to the edges incident to a node can potentially affect the centrality of other nodes throughout the network. The measure of how much a node can affect the centrality of others is called *sensitivity*, within the context of sensitivity analysis [10]. One mechanism for computing sensitivity is via function derivatives [17].

In a general sense, sensitivity analysis explores the variation of a function in terms of the variation of its inputs. For social networks, we can consider the centrality as a multidimensional multivariate function, which takes an adjacency matrix as input and its output is an n-dimensional vector, where each of its components is the centrality of a node. To find the sensitivity of this function, we must first define the variable with respect to which we compute the derivative. In principle, it is possible to compute the sensitivity of centrality with respect to each edge, which is analogous to computing the derivative of the centrality function with respect to each entry in the adjacency matrix. However, the resulting sensitivity space is astronomical, since it must consider all possible variations of variables. Even in the simplest case, where the sensitivity is computed with respect to each edge independently, the number of sensitivity parameters would grow cubically with the number of nodes. Instead, we define one variable per node, which is not only computationally less expensive, but it results in a measure of sensitivity easy to understand. The definition of these variables and the associated derivatives is formalized below.

3.1 A Variational Definition of a Social Network

A variational definition of a social network describes its structure, typically the adjacency matrix, and subsequent metrics, as functions of variables associated to its elements. In this paper, we define variables associated to each node. Let us define n independent variables representing a parameterized space for the weighted degree of each node $t_1, t_2, \ldots, t_n$. Therefore, we can think of the adjacency matrix as a function of these parameters, and consequently, a centrality metric as a composite function in terms of the adjacency matrix.

We, therefore, write centrality as a function:

$$C(t_1, t_2, \ldots, t_n), \tag{1}$$

with partial derivatives with respect to these parameters

$$\frac{\partial C(t_1, \ldots, t, \ldots, t_n)}{\partial t} = \lim_{h \to 0} \frac{C(t_1, \ldots, t+h, \ldots, t_n) - C(t_1, \ldots, t, \ldots, t_n)}{h}. \tag{2}$$

Therefore, the derivatives of centrality can be represented as a matrix S where each element

$$s_{ij} = \frac{\partial C_i(t_1, \ldots, t_n)}{\partial t_j}, \tag{3}$$

encodes the sensitivity of node i with respect to node j. To find an infinitesimal change in the centrality, and therefore, its derivative, we observe that many centrality metrics are algebraic operations on the adjacency matrix. Therefore, we can expand the derivative in terms of the derivatives of the adjacency matrix, using the chain rule of differentiation:

$$\frac{\partial C(A)}{\partial t} = \frac{dC}{dA}\frac{\partial A}{\partial t}, \tag{4}$$

where the derivative of the adjacency matrix is, analogously,

$$\frac{\partial A(t_1, \ldots, t, \ldots, t_n)}{\partial t} = \lim_{h \to 0} \frac{A(t_1, \ldots, t+h, \ldots, t_n) - A(t_1, \ldots, t, \ldots, t_n)}{h}. \tag{5}$$

This means that, if we know the closed form of both the centrality function and the adjacency matrix, we can readily compute the sensitivities of centrality via symbolic differentiation. However, adjacency matrices are seldom, if ever, defined analytically in terms of a set of parameters. Instead they are defined discretely as a collection of edges that may change over time. For this reason, we must approximate the variation of the adjacency matrix for a given change in one of the variables t_i. Because we define t_i as a variable that models the degree distribution, an infinite number of adjacency matrices can result in the same variation. Say, for example, that we want to measure the variation in the adjacency matrix that results from adding 1 to the degree variable t_i. Naturally, this can be obtained by adding a new edge incident to i of weight 1, or increasing the weights of, say, 10 edges incident to i by 0.1, and so on. We adopt an stochastic approach and define the variation of the adjacency matrix as a probabilistic change in all the edges incident to that node. The probabilities are given by the edge weights. Formally, we can define the variation matrix with respect to node k as follows:

$$A_{ij}(t_1, \ldots, t+h, \ldots, t_n) \approx A_{ij}(t_1, \ldots, t, \ldots, t_n) w_{ij}(t, h) \tag{6}$$

$$w_{ij}(t, h) = \begin{cases} 1 + \frac{h}{deg(t)}, & i = k \text{ or } j = k \\ 1 & \text{otherwise,} \end{cases} \tag{7}$$

where $deg(t)$ is the degree function, so that $deg(t_i) = degree(v_i)$, for a node v_i. It can be seen that $deg(t+h) = deg(t) + h$, for all t. This equation simply states that the adjacency matrix is updated to "report" a change in degree as the change of the edge weights proportional to the probability of each edge.

This definition does not assume anything particular about the adjacency matrix, such as symmetry. Therefore, this variational approach can be applied to directed and undirected networks alike. Fig. 2 shows the result of

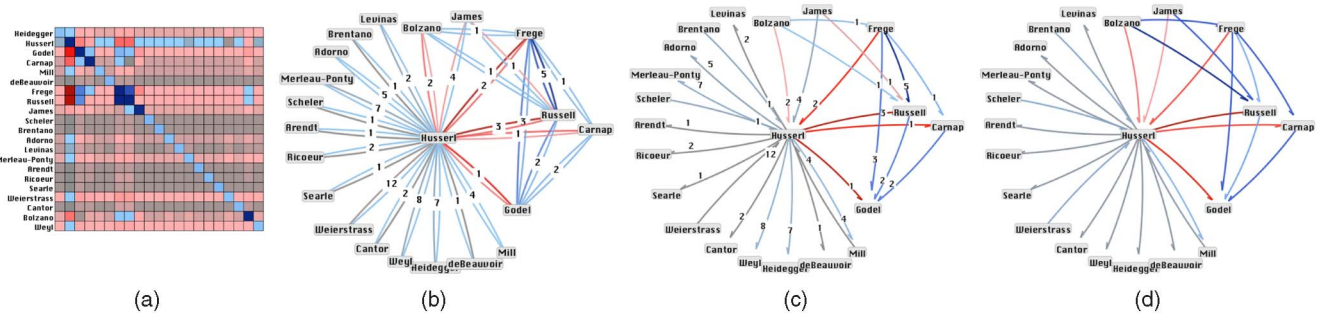

Fig. 2. Visualization of sensitivities for different graph types. (a)-(b) Weighted undirected graph as an adjacency matrix (a) and a node-link diagram (b). Notice the imbalance in the sensitivity between *James* and *Russell*. (c)-(d) Weighted and unweighted directed graph. Although the overall distribution of sensitivity is preserved (two subnetworks separated by red links), we notice the sensitivities of *James* are reversed when weights are added.

extracting sensitivities for different variants of the network depicted in Fig. 1. From left to right, we show the sensitivities for a weighted undirected network (edges are bidirectional with the same weight, i.e., A is symmetric), for a weighted (Fig. 2c) and an unweighted (Fig. 2d) directed network. Overall, the sensitivities exhibit similar behavior, since the network is small, but subtle relationships emerge. For example, *James* is more sensitive to Russell than *Bolzano* is, when considering uniform edges. But this ranking is reversed when weighting is used.

In the following sections, we follow this variational approach to derive formulas for the sensitivity of both eigenvector and Markov centralities, based on feedback and random walks of the adjacency matrix, for which there is a continuous function. Then, we describe a general approach for approximating sensitivity via forward differences.

3.2 Eigenvector Centrality

For the case of eigenvector centrality, whose variants are at the core of PageRank [9] and HITS [32], the centrality of a node can be formulated as a linear combination of the scores of the other nodes, which results in the eigenvector equation:

$$A\mathbf{x} = \mathbf{x}, \tag{8}$$

where A is the adjacency matrix of the network. The solution can be found as the eigenvector corresponding to the eigenvalue $\lambda = 1$. Alternatively, one can normalize the adjacency matrix A so that the sum of columns is 1. In this case, the solution to the problem is the eigenvector corresponding to the largest eigenvalue (which equals 1). The eigenvector centrality of a node is therefore $C_E(v) = \mathbf{x}_v$.

To find the derivative with respect to a degree variable t_i, let us denote $Q = A - I$, so that $Q\mathbf{x} = 0$. Differentiating at both sides, we have:

$$\frac{\partial(Q\mathbf{x})}{\partial t_i} = 0$$
$$\frac{\partial Q}{\partial t_i}\mathbf{x} + Q\frac{\partial \mathbf{x}}{\partial t_i} = 0,$$

from which it follows our formula for the eigenvector centrality sensitivity:

$$\frac{\partial \mathbf{x}}{\partial t_i} = -Q^{+}\frac{\partial Q}{\partial t_i}\mathbf{x}, \tag{9}$$

where Q^{+} is the pseudoinverse of Q (since Q is, in general, a singular matrix). This pseudoinverse can be computed using the singular value decomposition of Q or the least squares pseudoinverse: $Q^{+} = (Q^{\top}Q)^{-1}Q^{\top}$.

3.3 Markov Centrality

This centrality interprets the network as a Markov process, and can be understood intuitively as the amount of time an imaginary token performing random walks spends on each node. According to White et al. [48], this can be computed as the mean first-passage time in the Markov chain [31]:

$$m_{rt} = \sum_{n=1}^{\infty} n f_{rt}^{(n)}, \tag{10}$$

where $f_{rt}^{(n)}$ is the probability that the chain first returns to node t in exactly n steps. According to Schaffer et al., this can be computed as a matrix M,

$$M = (I - Z + EZ_{dg})D, \tag{11}$$

where I is the identity matrix, E is a matrix containing all ones, and D is a diagonal matrix where each element in the diagonal is the reciprocal of the stationary distribution $\mathbf{x}(v)$ of a node v. Z is the fundamental matrix of the Markov Chain, and Z_{dg} is a matrix consisting of the diagonal elements of the fundamental matrix. Z is defined as:

$$Z = (I - A - e\mathbf{x}^{\mathrm{T}})^{-1}, \tag{12}$$

where A is the Markov transition probability matrix and $\mathbf{x}$ is a column vector of the stationary probabilities, which are the same ones computed for the eigenvector centrality as the solution to $Q\mathbf{x} = 0$.

The Markov centrality of a node v (among n nodes) can therefore be extracted from M as [7], [48]:

$$C_M(v) = \frac{n}{\sum_{s \in V} M_{sv}}. \tag{13}$$

The derivatives of M now can be found analytically as:

$$\frac{\partial M}{\partial t_i} = (I - Z + EZ_{dg})\frac{\partial D}{\partial t_i} + \left(-\frac{\partial Z}{\partial t_i} + E\frac{\partial Z_{dg}}{\partial t_i}\right)D, \tag{14}$$

where the derivative of the fundamental matrix, being the inverse of another matrix, is

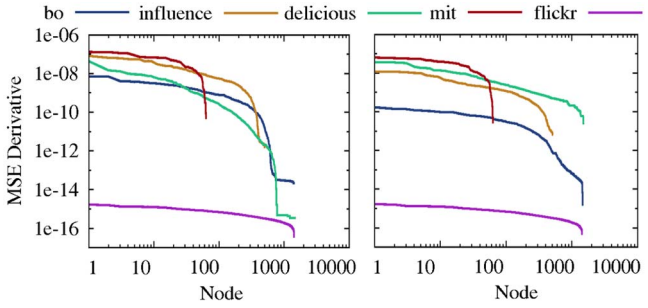

Fig. 3. Error plot of finite difference approximation and analytic derivation for eigenvector (left) and Markov (right) centralities. The x-axis plots nodes in decreasing order of the error. The y-axis denotes the mean square error (MSE) of the linear fit between the approximated and the analytic derivative. Although an approximation, it serves for validating the correctness of our derivation. Notice the difference in accuracy of the flickr data with respect to the others. This is due to the tight connectivity in this data set, which makes the derivatives more accurate as they influence each other directly.

$$\frac{\partial Z}{\partial t_i} = -Z\left(\frac{\partial Q}{\partial t_i} - e\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial t_i}\right)Z. \tag{15}$$

The derivative $\partial D/\partial t_i$ is a diagonal matrix containing the inverse of the derivatives of the stationary probabilities $\mathbf{x}$, and $\partial Z_{dg}/\partial t_i$ is the diagonal of the derivatives of Z. The partial derivatives of Q and $\mathbf{x}$ are computed as defined above for the case of eigenvector centrality.

3.4 Sensitivity via Forward Differences

As described above, it makes sense to compute the sensitivity as a derivative of the centrality function, since it can be defined as a continuous function in terms of the adjacency matrix A. Other types of centrality, such as closeness and betweenness are usually defined in terms of a count of shortest paths or other metrics. This makes it difficult to obtain a derivative. However, a sensitivity metric can still be obtained applying finite differences.

For a given centrality metric, we can approximate the derivatives as the change in centrality induced by the variation matrix. That is:

$$\frac{\partial C(A)}{\partial t} \approx \frac{C(A(t_1,\ldots,t+h,\ldots,t_n)) - C(A(t_1,\ldots,t,\ldots,t_n))}{h}, \tag{16}$$

where $C(A(t_1,\ldots,t+h,\ldots,t_n))$ is the centrality function of the graph that results from a variation matrix along parameter t.

3.5 Validation

To evaluate the validity of our approach, we approximate the derivatives via finite differences, as described in Section 3.4. We then compute the mean square error of the linear fit between the approximated and analytical values for the eigenvector and Markov centralities. Fig. 3 shows the error of the finite difference approximation for both the eigenvector and Markov centralities. Although an approximation, it helps us validate the results of our analytic derivation, since, in theory, the derivatives should represent the amount of change in the centrality of a node when the degree of another node changes. Note the logarithmic scale on number of nodes. Similar to the distribution of centrality of small world networks, the error exhibits an exponential fall off. This means that highly central nodes are more sensitive to approximation error than other relatively unimportant nodes. We see that the linear approximation is quite good for both types of derivatives. This is important as many centralities may be difficult to differentiate analytically. The most dramatic difference can be seen between the flickr data set and the other networks. We believe this is due to the tight connectivity in the flickr data set compared to the rest. Since each node practically influences directly every other single node, there are less chances of introducing numerical error in the approximation. As the length of random walks between any pairs increases, the approximation of their relative influence (partial derivative) becomes less accurate.

4 Visual Reasoning

Here we discuss some of the applications of centrality derivatives in the visualization of social networks to improve the analysis and understanding of interaction between nodes in a scale-free network. Some of the questions that arise when analyzing social networks, which centrality derivatives help answer via visual means, are:

1. What is the distribution of sensitivity in a social network? Do links represent friendship or enmity relationships? These questions can be answered, at a glance, in a visualization of *mutual* sensitivity.
2. Can we simplify the network representation to its core elements? Is the simplification meaningful? *Sensitivity-based simplification* can be used to obtain more manageable graph layouts that have a similar centrality distribution to the original network.
3. What are the most important nodes in relation to a given search, outside their immediate neighbors? Is the range of sensitivity of a node large or local? With a search-and-expand approach, we can provide efficient means of social network navigation.
4. How reliable are centrality metrics?

This type of reasoning, about the analysis itself, is seldom answered in typical visualization applications. Sensitivity analysis is an essential tool for measuring the robustness and uncertainty of centrality and related metrics.

4.1 Visualizing Friendship and Enmity

The natural application of sensitivities to visualization is the generation of overviews. These overviews, where we encode the sensitivity as a visual property, help understand the distribution of importance and the types of relationships represented by the links. One of the properties of sensitivity parameters is that they can be characterized by their magnitude and sign. This generates a signed network that is essential for analyzing the social balance of a network. In this context, we can refer to positive links as representing "friendship," while negative links represent "enmity." This analogy, widely used within the context of social dynamics, helps us understand the evolution of social networks in terms of the balance of the signed edges in triads [1]. For example, a balanced network is likely to evolve into an

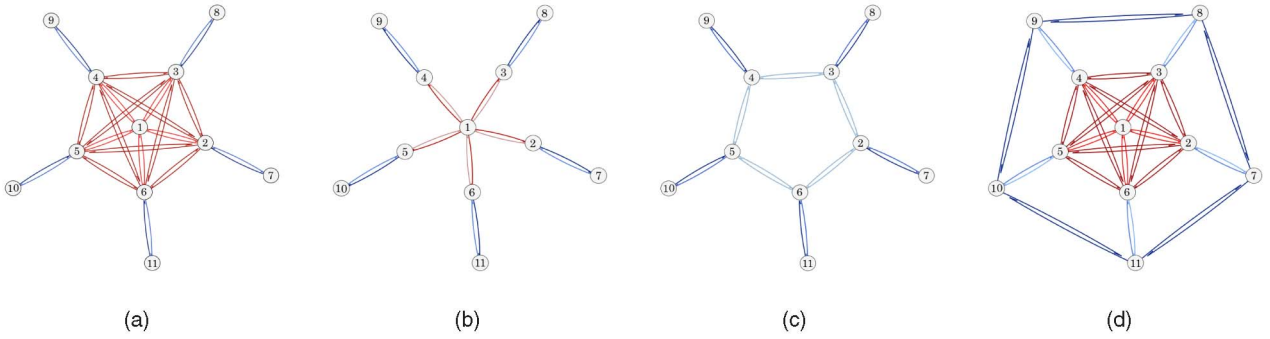

Fig. 4. Friendship and enmity on simple networks. Nodes in complete (a) and star (b) networks compete for importance, and therefore, exhibit strong negative sensitivities, while nodes arranged in a ring (c) exhibit collaboration. Hybrid networks exhibit the two behaviors (d). (a) Complete, (b) Star, (c) Ring, and (d) Ring+Complete.

"utopia," formed by all positive links. Because this analogy is easy to comprehend, we use the same terminology here.

We have experimented with a number of data sets with different structure, and have identified key properties as a result of the visualization of the sensitivity parameters. Fig. 4 contrasts several simple networks consisting of a core network of six nodes and five peripheral nodes. Depending on the topology of the core network, we observe different relationships. For a complete graph, nodes are competing for importance and we observe a large negative sensitivity among all nodes. A similar behavior is seen for a ring network, although we observe an asymmetric relationship. Nodes are more sensitive to changes on the center of the star (node 1), while having little impact on that node. For a cycle, we see a more collaborative network, where each node is "friends" with their own neighbors. Hybrid networks, like seen in Fig. 4d show the two behaviors, as it is formed by a complete graph connected to a wheel. In real networks, we often find a combination of these types.

Fig. 5 summarizes our study with the selection of three types of networks we have encountered in our experiments. Fig. 5a shows a typical sparse network, often found in hierarchies and exhibiting a number of subnetworks in a star pattern. This particular example shows the core network of the Friendster social network and the main connections of these core nodes. We notice a predominancy of negative links between clusters. This is expected, since each cluster center has roughly an equal chance of becoming the most important node. Therefore, any change in a cluster center will impact negatively the importance of another. But the visualization also characterizes the magnitude of this competitive relationship. We see that the root of the largest cluster (toward the right) has a larger sensitivity to the clusters in the middle (a darker red edge) than the other cluster centers. Fig. 5b shows the highly interconnected core of the cocitation ArXiv network. A different pattern emerges. We do not see the individual skeletal negative links, but we see them all clustered within a single region. By contrast, other interconnected groups, such as the one toward the right, exhibits collaboration (all positive links) rather than competition (all negative links). This indicates a separation of groups that may not be evident in an overview and seems typical of tightly connected networks, such as in certain specialized cocitation or proximity networks. Section 5.1 describes another example of this type of behavior. Fig. 5c shows a hybrid network, where tightly coupled subnetworks are connected via a few links, resembling both the star-shaped clusters and the tightly connected group. This example is the core subnetwork of the del.icio.us graph, a Web tagging social networking site. In this case, we see a similar behavior, but negative links are not exclusive of intercluster links. We also see a collection of groups of varying connectivity. Specifically,

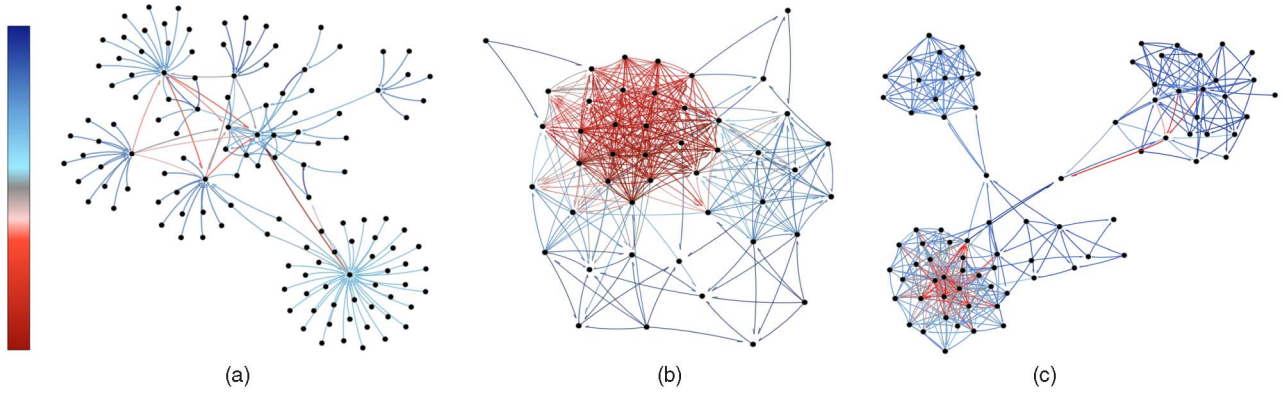

Fig. 5. Color encoding of sensitivities helps identify different types of interaction. (a) Collection of star-shaped networks, intercluster links exhibit negative sensitivity (Friendster). (b) Tightly connected networks with either a strong positive or a strong negative sensitivity. Here, we see one competitive and one collaborative network (Astrophysics). (c) Hybrid case with tightly-connected groups linked via a core network (Del.icio.us).

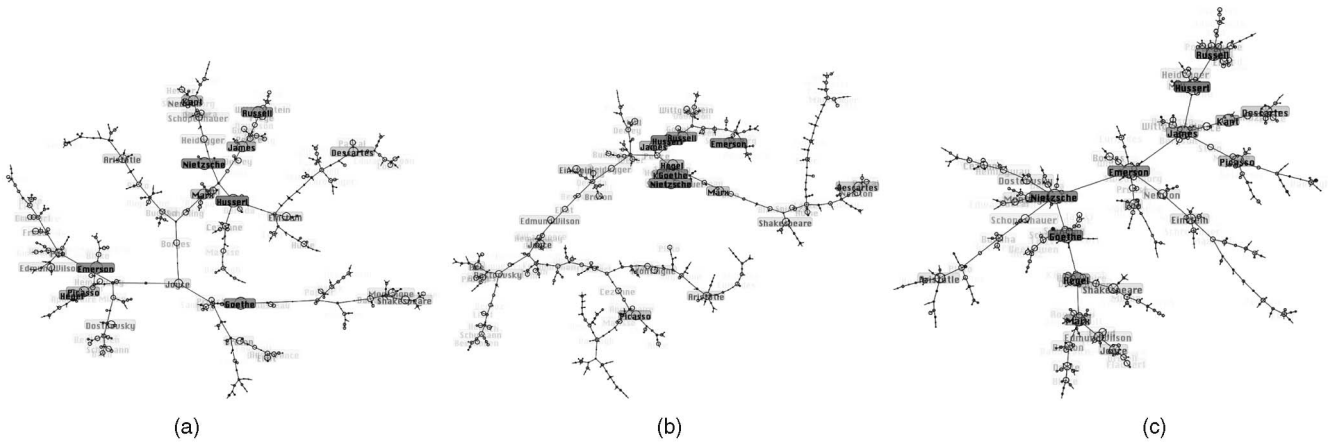

Fig. 6. Comparison of edge simplification methods for the influence data set [25]. Size and opacity denote the centrality of a node. We compare several edge weighting strategies: uniform, edge betweenness, betweenness sensitivity and eigenvector sensitivity. The latter is the one that preserves node centralities the most. (a) Unweighted, (b) Edge betweenness, and (c) Eigenvector sensitivity.

the group in the lower part suggests that it is formed by smaller subclusters, one of them highly competitive (red links) while others, less tightly connected and more collaborative (blue links). Notice how this cluster resembles the distribution of sensitivity of the ArXiv network.

4.2 Sensitivity-Guided Simplification

Another important application of centrality derivatives is the ranking of edges for filtering and simplifying complex networks. This idea was suggested by Girvan and Newman [16] and van Ham and Wattenberg [46]. Both use the betweenness centrality of edges, analogous to the betweenness centrality of nodes, to rank the edges. By removing the highest BC edges, van Ham and Wattenberg obtain a minimum spanning tree that preserves the structural properties of the network, in particular, the presence of clusters of interest. In a diametrically opposite direction, Jia et al. [30] consider the highest BC edges first to construct a maximum spanning tree. This metric preserves the communication paths that form the network.

Sensitivity, although not equivalent to edge betweenness centrality, also provides an edge ranking. Therefore, it is only natural to study the implications of this metric in simplifying a social network. We follow the general strategy laid out by previous simplification approaches and compute the minimum spanning tree of the graph, where each edge is weighted, in our case, by the centrality sensitivity. We first performed a qualitative analysis where we obtained simplifications of a number of data sets under different weighting schemes, including uniform weighting, where all edges are equally important, weighting based on edge betweenness and weighting based eigenvector sensitivities. Fig. 6 compares the results for the network of genealogy of influence, a network that relates great thinkers in History depending on the influence of one thinker on the works of another [25]. In this visualization, we use circles to represent each node, with sizes proportional to their centrality. Labels of the most important nodes are also highlighted. At a glance, it is difficult to judge which method is better. However, we were able to identify structural properties that were retained during the simplification. In this paper, we are interested in how centrality was preserved. For the cases of uniform and edge betweenness weighting, centrality of nodes is barely preserved. For the latter, a global notion of centrality remains, since the most important nodes remained clustered together. However, we see that important nodes often appear at the edges of the network rather than at its center. This is somewhat solved with the sensitivity, and we now see the highlighted names at the center of local groups. For betweenness sensitivity (not shown), we found that groups are often connected through unimportant nodes, since shortest paths can short-circuit through relatively unimportant paths. For the eigenvector centrality sensitivity, the nodes retained the centrality better, and important nodes appear connected in a single skeletal path, highlighting what Jia et al. identified as important communication paths.

In general, selecting an appropriate simplification scheme depends on the task at hand and no single method can be said to be superior to others. Simplification based on edge betweenness retains most local clusters, but they appear connected in a rather arbitrary way. Eigenvector derivatives, on the other hand, preserve centrality, but may break apart some loosely coupled clusters. We show an example of such a case in Section 5.3. This observation prompted us to perform a systematic evaluation of the result of simplification from a structural point of view.

4.2.1 *Preservation of Centrality*

To validate how effective is a simplification, we must first measure the quality of a simplification in a meaningful way. At a higher level, the individual task, the semantics of the data and the context are factors that determine whether a specific layout is useful or not. These are difficult to measure and isolate, and, to the knowledge of the authors, there has not been a convincing study that helps reveal these issues. On the other hand, from an algorithmic point of view, one can study simplification algorithms in terms of their performance to preserve structural properties. In previous simplification approaches, the main goal is to preserve clusters, but evaluation is performed mostly as a qualitative assessment. Here, we performed a systematic evaluation and analyzed the relationship between the network statistics before and after simplification. In particular, we were interested in the degree to which centrality is preserved

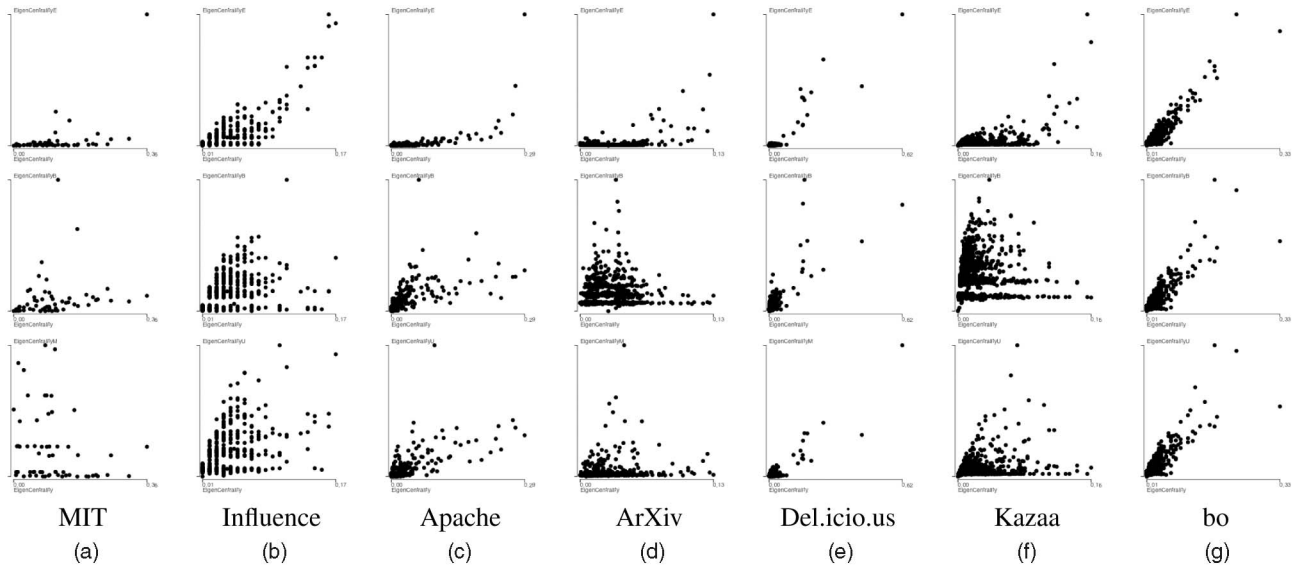

Fig. 7. Measuring centrality preservation of simplification strategies for a number of networks. Top: eigenvector sensitivity. Middle: Edge betweenness. Bottom: Uniform. The eigenvector sensitivities consistently preserve the centralities of important nodes (closer to diagonal).

after simplification. We argue that a good network simplification should *retain the centrality distribution of the original graph*. This property is important to avoid misleading visualizations. Node link diagrams often convey the notion that, if a node is connected to many other nodes, it is regarded as important. When the network is described as a hierarchy, it is, therefore, expected to see important nodes at the higher levels. We studied the impact of simplification on the centrality metric for a number of data sets. Fig. 7 summarizes the results for seven networks and three weighting schemes for simplification: eigenvector sensitivity, uniform weighting and edge betweenness. Each graph plots the original centrality of each node in the x-axis versus the new centrality in the simplified network (y-axis). A centrality-preserving simplification should result in a distribution of points near the diagonal. Notice how eigenvector sensitivities consistently result in a better preservation of centrality than the other two schemes. For example, note that the few important nodes remain important for the most part (points in the upper right corner of plots). In addition, we gain insight about the nature of the network by looking at these plots. We can identify four types, Fig. 7a, Fig. 7b, Figs. 7c, 7d, 7e, and 7f, and Fig. 7g. The first one corresponds to a proximity network, which behaves different to social networks in that there is no preferential attachment. Everyone has roughly the same probability of being in proximity to others. The second is a synthetic data set. The third group corresponds to real social networks from online dynamics or cocitation patterns. Last, Fig. 7g is a protein network, which appears more consistent across the different metrics.

4.3 Search and Explore

Centrality derivatives are also useful for bottom-up visualization approaches, where we begin with a given node, possibly as a result of a search, and then expand the context and navigate around the network to discover important relationships of that node. This approach has been shown to be effective when exploring large graphs [45]. An example is shown in Fig. 8, for the core network of the del.icio.us data set. To support effective exploration of the network, centrality derivatives can be used to visualize the magnitude (saturation) and sign (hue, red for negative and blue for positive) of the influence of a given node. On top, we show the distribution of sensitivity to a selected node (center of

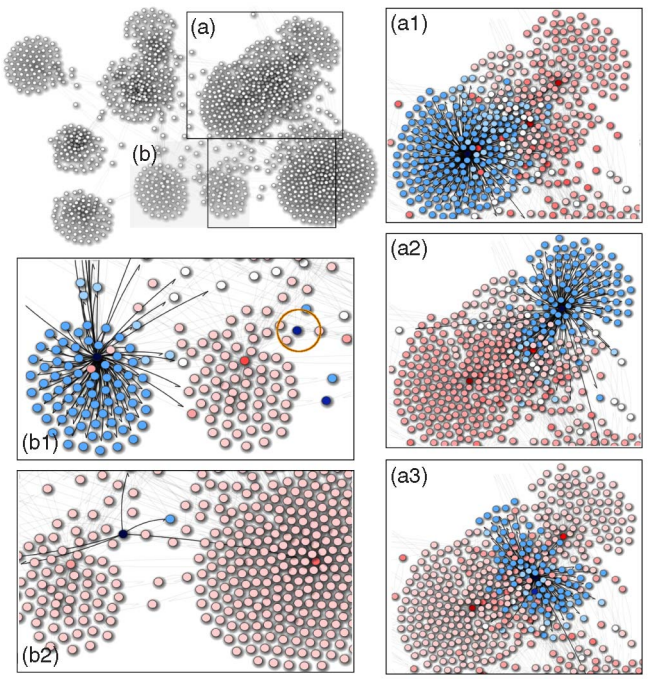

Fig. 8. Sensitivity-guided search and exploration of the del.icio.us network. (a) A large group in the network appears as a connected cluster. (a1-3) Encoding sensitivity as color (blue for positive, red for negative sensitivity) helps the user recognize three distinct clusters intertwined together. (b) Sensitivity-guided navigation. (b1) After selecting a node (center of blue cluster), we see a salient node (circled) which has a strong sensitivity even when not directly connected to the selected node. This prompts the user to navigate further in the network centering at that node (b2), to discover that the node connects to an important cluster.

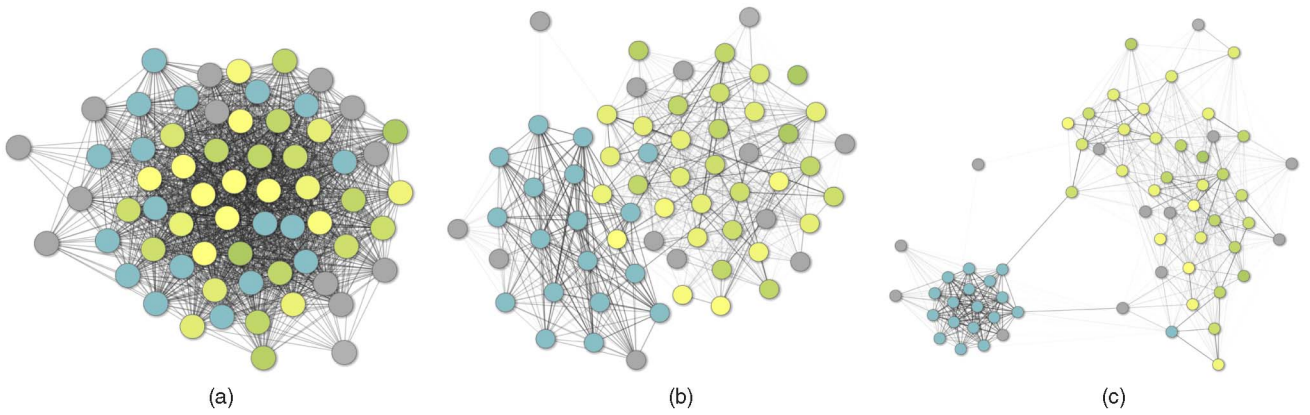

Fig. 9. Exploration of MIT Reality Data set (a) Force-directed layouts do not convey groups, except via semantic attributes (blue for Sloan members, green for MediaLab members) (b) Considering only positive sensitivities, we see a better separation between the two groups. (c) This becomes clear when filtering edges based on sensitivity magnitude. The Sloan group persists, while the MediaLab splits into a small number of subclusters.

leftmost cluster highlighted in blue). All other nodes are color-coded based on the sensitivity to the selected node. We see two salient nodes in dark blue, which indicate a high positive sensitivity, even though they are not directly connected to the selected node. This prompts the user to follow the links of these nodes (circled), and we observe that the node is connected to the center of a cluster, therefore, acting as a critical bridge between the two clusters. At the bottom, sensitivity-guided navigation helps us visualize different clusters which are otherwise obscured by the layout. After selecting one of the cluster centers, we see that the other (possible) cluster centers exhibit a large negative sensitivity (since they are "competitors" for importance). Selecting one of these nodes highlights its local structure (seen as blue nodes) and also highlights the other two cluster centers in red. Without this interaction, the boundaries between these intertwined clusters are difficult to define. As suggested by this example, one can define a heuristic for navigating large graphs, which dictates that one should follow the nodes with highest sensitivity, either positive or negative, in order to quickly traverse full regions without getting stuck in local structures.

5 Examples

Here we illustrate how we use centralities sensitivities to improve the insight gained about three social networks.

5.1 MIT Reality Data Set

The MIT Reality data set collects information about one hundred subjects from the MIT Media Lab and the School of Management, using a series of communication devices, powered with Bluetooth chips, throughout a period of about two years. The data set contains several ways in which a social network can be extracted, such as call and texting logs, as well as proximity data. Here, we focus on the proximity data. In this part of the data set, a link is created between two actors if they were in close proximity to each other for a period of time. One of the key questions that the MIT Reality group wants to answer with the compilation and analysis of this data set is whether the topology of the network can be inferred from the proximity data alone, since it provides information that may not be captured by tracking calls.

Indeed, when we plot the network using a force-directed layout, we get the inevitable hairball. This is not surprising, as most nodes are in close proximity to each other for some period of time, and the vast amount of links forces the nodes to clump together. This is depicted in Fig. 9a. Color coding denotes the position held by the person, a simple identifier that helps find clusters. Along this dimension, we clearly see a big group in blue, corresponding of students of the Sloan School of Management, and various groups in green, corresponding to first year, graduate and senior students, as well as faculty and researchers from the Media Lab. Unidentified actors are represented in gray. Although we already know there is a clear separation of roles (Sloan versus Media Lab), the goal of the analysis and the collection of the data is to find out if this can be extracted by structural properties alone. Evidently, it is difficult to observe such separation visually without the semantic clues. We performed a centrality analysis, and found that the two main clusters could be characterized by the sign and magnitude of their sensitivities. Actors within each group had positive influence to each other, while the sensitivity with respect to actors in the other group was mostly negative. Fig. 9b shows the result after laying out the graph considering only those edges that represent a positive sensitivity. Certain layout algorithms, such as those based on LinLog energy models, often extract clusters better [41]. In this example, a LinLog layout results in an equivalent separation to that in Fig. 9b, validating the capability of sensitivities to retrieve clusters.

We explored the data set further for those links with larger sensitivity. If we "weaken" the edges with low sensitivity, a graph layout algorithm starts revealing a hidden structure. In Fig. 9c, we see that the two clusters behave differently. The Sloan cluster remains tightly connected while the Media Lab cluster begins to separate into three or four groups, one of them consisting predominantly of first year students. This strategy can be used to determine how tightly connected is a visible cluster in a network and provides a simple, yet robust, methodology for social network exploration.

5.1.1 *Reasoning about Uncertainty*

We expanded our analysis of this network to understand more how centrality is distributed. First, we observed that

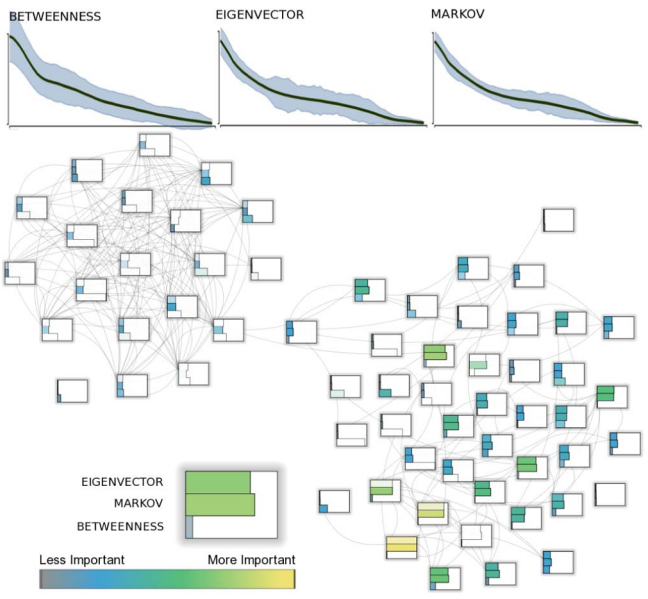

Fig. 10. Uncertainty visualization for centrality. Left: overview of centrality metrics and uncertainty as area error for Betweenness, eigenvector, and Markov centralities. Right: Detail uncertainty view, where each node shows the scores for the three centralities, mapped as size and color. Transparency encodes uncertainty. We see a consistent trend of high-uncertainty nodes in the cluster on the left, with a mix of low- and high-uncertainty nodes in the cluster on the right.

several centrality metrics provide different, often contradicting results. This is not at all surprising, since centrality metrics are, in general, defined differently. A node with a high degree may have low betweenness if no shortest paths go through it. To this end, we study the aggregate effect of sensitivities in the centrality of each node. This can be approached from the perspective of uncertainty analysis. As pointed out by Wasserman [47], social network representations may not be an accurate depiction of the underlying social structure. Moreover, biological networks often include an inherent measurement error that cannot guarantee complete accuracy. Therefore, every edge between two nodes carries an inherent uncertainty that is propagated through network operations, such as clustering, filtering and, naturally, and centralities. Let us define the uncertainty of a node v_i as the variance σ_i of its corresponding variable t_i. We can think of this uncertainty as the inverse probability of increasing the degree of a node by one degree (i.e., adding or removing a node). The uncertainty of the centrality of a node $\sigma_C(t_i)$ is a linear combination of the uncertainties of these variables, using the law of propagation of uncertainties:

$$\sigma_C^2 = \sum_{i=1}^{N}\left(\left(\frac{\partial C}{\partial t_i}\right)^2 \sigma_i^2 + \sum_{j=1}^{N} COV_{ij}\frac{\partial C}{\partial t_i}\frac{\partial C}{\partial t_j}\right), \qquad (17)$$

where σ_i^2 is the variance in the degree of a node and COV_{ij} is the covariance of the degrees of nodes i and j. If we simplify the uncertainty modeling to describe each node as an independent variable, the resulting uncertainty is just the linear combination of the variances of each node.

Fig. 10 shows an uncertainty analysis of the MIT proximity data set for three centralities: betweenness, eigenvector and Markov centralities. On top, we plot the distribution of centrality for all nodes (in descending order) as a uncertainty area curve, where the area represents variance. We first notice that Markov centrality is more robust than the other two metrics, since it looks at the long term stability of random walks, more sensitive to variation than betweenness, which is prone to short-circuit errors. This confirms the observation by Carpenter et al. [11] about betweenness. Fig. 10-bottom shows a detailed view of uncertainty. Color indicates centrality, while transparency indicates uncertainty. More uncertain nodes are more transparent. We notice that the sloan cluster on the left has a consistent behavior of relatively unimportant nodes with high uncertainty. We also identify certain nodes with low uncertainty in the MediaLab cluster (right). Although not the most important, they are the most reliable. These uncertainty views are useful to predict the expected behavior of nodes in a dynamic network. As nodes disconnect and reconnect from their neighbors, the centrality of certain nodes will undoubtedly change. With these plots, we can predict where these changes are most likely to occur. For example, according to the eigenvector centrality, a change in the degree of nodes is likely to change the centrality of the nodes in the bottom part of the plot (which happen to be members of the "Sloan" group), while the central nodes in the rightmost cluster ("Media Lab") are more likely to remain constant. We see a few exceptions in the Media Lab cluster, where the variance of the eigenvector centrality is high. Further inspection shows a different role of these people (they are not students or "new grads"), which may hint at the source of this disparity.

5.2 Genealogy of Influence

The genealogy of influence is a network compiled by Mike Love [25], that describes the intellectual influence among the works of great thinkers in History including renowned artists, writers, mathematicians, philosophers and scientists. Although the network was built synthetically after studying what are deemed to be the most influential works for each person, the network has traits of a social network. This is manifested by the typical hairball in Fig. 11a. Identifying meaningful clusters is practically impossible. We have applied our approach to look at meaningful relationships in terms of sensitivity. We then applied the filtering approach based on the minimum spanning tree of the eigenvector centrality. The resulting visualization is shown in Fig. 11b, and we begin to see emergent clusters. To retain the original edge connectivity, we use a hierarchical edge bundling strategy to group links together and avoid excessive clutter [22]. The bundles are routed through the hierarchy computed in the minimum spanning tree. We highlight two clusters. One of them, in orange, is a cluster of *Russell* and highlights the well-known connections with mathematicians and logicians such as *Godel*, *Quine* and *Whitehead*. We even see the connection to *Vico*, a philosopher from the seventeen hundreds that inspired mathematicians such as *Russell*. This connection is hard to identify from the original visualization (Fig. 11d). Although this cluster can be found in close proximity in an unaugmented visualization, it would not stand out visually as a single coherent group. When we consider the cluster highlighted

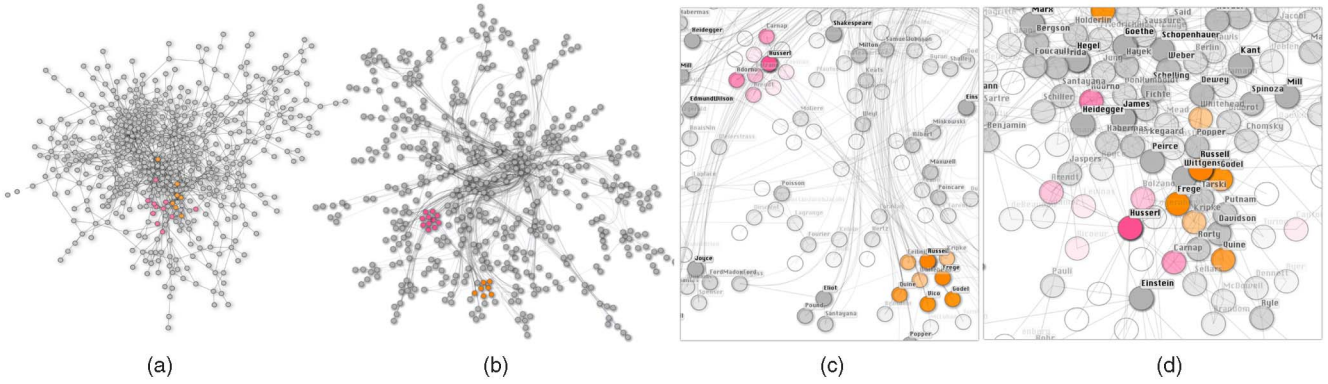

Fig. 11. Visualizing the genealogy of influence network. (a) Traditional visualization prevents users from identifying any meaningful group. (b) A simplified network using eigenvector centralities allows us to see individual clusters. (c) Close-up views of two highlighted clusters. (d) The same clusters, in an unfiltered visualization, are hard to identify.

in magenta (*Husserl* group), these connections are even more difficult to identify together without filtering. The overwhelming amount of nodes connected to the most central ones creates a hairball where the locations of nodes become increasingly arbitrary.

5.3 Astrophysics Collaboration Network

This data set contains the collaboration network of scientists publishing abstracts on the astrophysics e-print archive (arXiv) between 1995 and 1999 [39]. A link between authors is created if they are coauthors of an abstract. Fig. 12 shows a sensitivity-guided visualization of the network. We follow the general strategy of simplifying the network in terms of sensitivity. First, we obtained the minimum spanning tree of the network using the derivatives of eigenvector centrality as weights. The resulting tree is visualized using a radial layout, where nodes higher in the MST hierarchy are positioned closer to the center of a circle, while leaf nodes are farther from the center. After this, we represent the original edges from the graph using a hierarchical bundling technique similar to that of Holten et al. [22]. The result is shown in Fig. 12a, with edges representing those links with higher positive sensitivity. We see the emergence of interconnected clusters. In Fig. 12b, we visualize the links with high negative sensitivity, which helps us see the core network, formed by highly central nodes. In traditional layouts, without considering the effects of sensitivity, these nodes are inevitably collapsed together at the center of the layout, as seen in Fig. 12c for the linlog (top) and force-directed layouts (bottom). We also see that the more evident clusters are well represented in all three types of layouts (clusters in purple and magenta), while other groups of nodes, for example the co-authorship networks of *Filippenko* (orange) and *Stetson* (yellow), are not evident

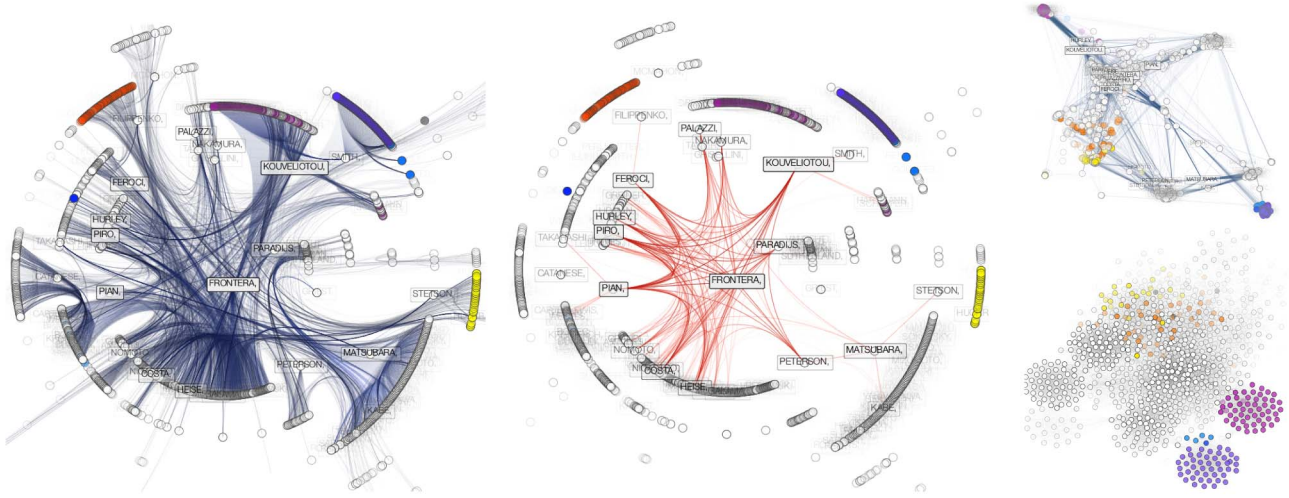

Fig. 12. Visual reasoning for the astrophysics collaboration network. (a) Sensitivity-guided radial layout showing only positive sensitivities helps us identify clusters and their connections. The use of hierarchical edge bundling helps us make sense of the multitude of connections. (b) The same layout showing the links with negative sensitivity highlights the core network formed by the nodes with higher coauthorship. (c) Layouts without a sensitivity augmentation often produce visualizations where clusters cannot be easily identified, as seen for the orange and yellow groups (top: Linlog layout bottom: mass-spring layout).

TABLE 1
Network Properties for the Data Sets used in This Paper and Comparative Timing for Several Techniques (Seconds)

Graph	Num. nodes	Num. edges	Edge Betweenness	Betweenness Derivs.	Eigenvector Derivs	Eigenvector Derivs. MT
friendster (core)	129	161	0.012258	0.281917	0.036193	0.017957
apache (core)	51	856	0.006538	0.110089	0.00884	0.009227
influence	514	2132	0.176664	24.1369	0.656055	0.527134
MIT	64	2964	0.01434	0.284446	0.011967	0.015476
apache	511	1796	0.190487	27.62803	0.629286	0.468059
flickr	1425	2848	1.089283	276.8418	16.351770	12.328721
bo	1458	3896	1.382473	494.611816	17.584723	12.983744
del.icio.us	1503	6032	1.454253	517.889404	18.858734	14.469767
astro-ph	955	37554	1.460922	610.64679	5.552344	3.567967
kazaa	1550	8028	2.104503	1069.565674	20.876854	16.481899
complete200	200	19900	0.202467	19.190643	0.203148	0.095947
complete500	500	124750	3.570290	752.1272	3.043994	1.348568

at all in force-directed layouts, and not clearly separable in linlog layouts. In a sensitivity-guided visualization, these clusters appear as separate groups. By following the bundled edges, we can still make sense of the connectivity of these two clusters with other groups. Note that sensitivity-guided strategies can be applied to other layouts and they are not intended as a replacement of a good layout. However, these strategies suggest to us that we should exploit the implicit hierarchy given by MST and the centrality ranking of nodes. Layouts that exploit these properties, such as the radial layout, are likely to produce better diagrams. Although simplification is not new here, the use of sensitivity provides a robust mechanism to obtain the critical links in terms of the dynamics of the network. Preservation of centrality is ideal to identify the representative actors in a cluster and it turns out to be useful when using visual representations intended for hierarchical structures, such as radial layouts and edge bundling.

6 Discussion

We have shown a number of applications of the centrality derivatives for the visualization of social networks. The networks used in this paper are summarized in Table 1.

Centrality operations, however, are often costly. For example, betweenness centrality can be computed in $O(\|V\|\|E\| + \|V\|^2)$ time for unweighted graphs, or, when using the Floyd-Warshall algorithm, in time $O(\|V\|^3)$ [7]. Approximating the derivative using finite differences implies increasing the cost by a factor proportional to $O(\|V\|)$. Brandes presented a fast approximation of betweenness centrality [6] that runs in $O(\|V\|\|E\|)$ for unweighted graphs. Using such an implementation, the evaluation of derivatives using finite difference approximations is more feasible. For the case of eigenvector centralities, the costlier operation is the solution to the eigenvector problem. A number of acceleration techniques have been proposed, as surveyed by Langville and Meyer [36]. The derivatives only imply an additional matrix multiplication, or equivalently, solving the linear system of equations in (9). Table 1 shows a comparison of timing among different techniques for computing sensitivity, including betweenness derivatives using finite differences and eigenvector sensitivities using analytic derivatives. We see that for moderate graphs, numerical approximation becomes impractical and analytic derivatives can be computed an order of magnitude faster. As a way to compare the complexity of eigenvector sensitivities, we also show the time complexity of edge betweenness using Brandes' fast algorithm [6]. For sparse graphs, this fast implementation proves much faster than computing sensitivities. However, edge betweenness does not account for indirect influences between nodes, for which a complete graph is required, e.g., one computing all pairwise distances. We see that as graphs become complete, edge betweenness and the eigenvector analytic derivative can be computed at the same cost. Nonetheless, analytic derivation is highly parallelizable. A simple multithreaded version of the algorithms proves faster than other alternatives for dense graphs. Markov centralities are probably the most expensive, which are $O(\|V\|^3)$, due to the computation of an inverse matrix during the estimation of the fundamental matrix. In this sense, the use of analytical derivatives become advantageous, since they can be computed as a constant sequence of matrix products. In contrast, a numerical approximation using central differences would require time $O(\|V\|^4)$, which is impractical even for relatively small networks. Harrison and Knottenbelt describe a method for approximating the first passage time and its derivatives using a novel Laplace transform [20]. We believe methods like this would greatly improve the scalability of our approach. Given the complexity of social networks, neither the analytic expressions for centrality derivatives nor their linear approximations can be applied directly to large networks. Although they are useful for local analysis of the social network (considering subnetworks at a time), more effective means are necessary. One may perform a similar analysis as the one presented in this paper for known approximations of the centrality measures or apply a hierarchical solution that works locally in subnetworks at a time and progressively expands to larger portions of the network.

7 Summary and Future Work

Several researchers have stressed the importance of coupling statistics and visualization to improve the exploration of large networks. This has been capitalized upon by recent attempts to produce effective visualizations based on statistical measures such as centrality. These efforts resulted in insightful color and shape encodings of nodes in social networks, radial layouts and graph simplifications. However, none has considered the variational aspects of these centrality metrics, essential for understanding the process by which a given node becomes important. In this paper, we

presented a general methodology to extract the sensitivity of centrality and apply it to typical visualizations of social networks. The quantification of sensitivity is addressed in this paper as an analytical derivative, following our variational description of the social network. We show that multiple tasks in visual reasoning can be supported with this new type of information. Overviews show *friendship and enmity* relationships, useful for characterizing the cooperation or competition within networks. Filtering can be supported in a more effective manner, as sensitivity provides a robust mechanism to simplify the network, and bottom-up approaches, such as search and expand on demand, can be improved by representing the relative importance of actors with respect to a given focal node.

While we have shown important applications of this work for social and other scale-free networks, our analysis can be applied to network analysis in general. Since the notion of sensitivity is based on the immediate change in the degree of a node, we believe our approach can provide insight on the behavior of dynamic graphs as well. Just as the addition or removal of edges may change the centrality of a node, sensitivity coefficients provide hints about how drastic are those changes and let users find the most *reliable* nodes or communication channels. Our approach is not restricted to the particular metrics used throughout this paper. We focused on common centrality metrics, which span a vast selection of algorithms, based on shortest-paths, feedback and Markov processes, but our general notion of sensitivity applies to other centrality metrics, such as closeness centrality and radiality.

ACKNOWLEDGMENTS

This work was sponsored in part by the US National Science Foundation (NSF) through grants CCF-0811422, CCF-0808896, CCF-0938114, and CCF-1025269.

REFERENCES

[1] T. Antal, P.L. Krapivsky, and S. Redner, "Social Balance on Networks: The Dynamics of Friendship and Enmity," *Physica D,* vol. 224, pp. 130-136, May 2006.
[2] J. Anthonisse, "The Rush in a Directed Graph," technical report, Stichting Mathetmatisch Centrum, 2005.
[3] V. Batagelj and A. Mrvar, *Pajek—Analysis and Visualization of Large Networks.* pp. 477+, Springer, 2002.
[4] M. Bianchini, M. Gori, and F. Scarselli, "Inside Pagerank," *ACM Trans. Internet Technology,* vol. 5, no. 1, pp. 92-128, 2005.
[5] P. Bonacich, "Factoring and Weighting Approaches to Status Scores and Clique Identification," *J. Math. Sociology,* vol. 2, no. 1, pp. 113-120, 1972.
[6] U. Brandes, "A Faster Algorithm for Betweenness Centrality," *J. Math. Sociology,* vol. 25, pp. 163-177, 2001.
[7] U. Brandes and T. Erlebach, *Network Analysis: Methodological Foundations.* Springer, 2005.
[8] U. Brandes and D. Wagner, "Visone-Analysis and Visualization of Social Networks," *Proc. Graph Drawing Software,* pp. 321-340, 2004.
[9] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems,* vol. 30, no. 1, pp. 107-117, 1998.
[10] D. Cacuci, *Sensitivity and Uncertainty Analysis: Theory Vol. 1.* Chapman and Hall/CRC, 2003.
[11] T. Carpenter, G. Karakostas, and D. Shallcross, "Practical Issues and Algorithms for Analyzing Terrorist Networks." 2002.
[12] T. Dwyer, S.-H. Hong, D. Koschützki, F. Schreiber, and K. Xu, "Visual Analysis of Network Centralities," *Proc. Asia-Pacific Symp. Information Visualisation,* pp. 189-197, 2006.
[13] L. Freeman, "Visualizing Social Networks," *J. Social Structure,* vol. 1, no. 1, 2000.
[14] L.C. Freeman, "Centrality in Social Networks: Conceptual Clarification," *Social Networks,* vol. 1, no. 3, pp. 215-239, 1979.
[15] P. Gajer and S.G. Kobourov, "Grip: Graph Drawing with Intelligent Placement," *Proc. GD '00: Eighth Int'l Symp. Graph Drawing,* pp. 222-228, 2000.
[16] M. Girvan and M.E. Newman, "Community Structure in Social and Biological Networks," *Proc. Nat'l Academy of Sciences of USA,* vol. 99, no. 12, pp. 7821-7826, 2002.
[17] A. Griewank, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* SIAM, 2000.
[18] S. Hachul and M. Jünger, "Drawing Large Graphs with a Potential-Field-Based Multilevel Algorithm," *Graph Drawing,* pp. 285-295, Springer, 2004.
[19] F. Harary and P. Hage, "Eccentricity and Centrality in Networks," *Social Networks,* vol. 17, pp. 57-63, 1995.
[20] P.G. Harrison and W.J. Knottenbelt, "Passage Time Distributions in Large Markov Chains," *SIGMETRICS Performance Evaluation Rev.,* vol. 30, no. 1, pp. 77-85, 2002.
[21] J. Heer and D. Boyd, "Vizster: Visualizing Online Social Networks," *Proc. IEEE Symp. Information Visualization,* pp. 32-39, 2005.
[22] D. Holten, "Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 741-748, Sept./Oct. 2006.
[23] http://graphexploration.cond.org/, GUESS, the graph exploration system, 2007.
[24] http://jung.sourceforge.net/, JUNG java universal network/graph framework, 2008.
[25] http://mike-love.net, Genealogy of influence, 2008.
[26] http://www.yworks.com/en/products_yed_about.html, yed graph ed., 2009.
[27] C.H. Hubbell, "An Input-Output Approach to Clique Identification," *Sociometry,* vol. 28, no. 4, pp. 377-399, 1965.
[28] R. Jacob, D. Koschützki, K.A. Lehmann, L. Peeters, and D. Tenfelde-Podehl, "Algorithms for Centrality Indices," *Network Analysis: Methodological Foundations,* U. Brandes and T. Erlebach, eds., pp. 62-82, Springer, 2005.
[29] H. Jeong, S.P. Mason, A.L. Barabasi, and Z.N. Oltvai, "Lethality and Centrality in Protein Networks," *Nature,* vol. 411, p. 6833, 2001.
[30] Y. Jia, J. Hoberock, M. Garland, and J. Hart, "On the Visualization of Social and Other Scale-Free Networks," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 6, pp. 1285-1292, Nov./Dec., 2008.
[31] J. Kemeny and J. Snell, *Finite Markov Chains.* Springer Verlag, 1976.
[32] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *J. ACM,* vol. 46, no. 5, pp. 604-632, 1999.
[33] Y. Koren, L. Carmel, and D. Harel, "Drawing Huge Graphs by Algebraic Multigrid Optimization," *Multiscale Modeling and Simulation,* vol. 1, no. 4, pp. 645-673, 2003.
[34] D. Koschützki, K.A. Lehmann, D. Tenfelde-Podehl, and O. Zlotowski, "Advanced Centrality Concepts," *Network Analysis: Methodological Foundations,* U. Brandes and T. Erlebach, eds., pp. 83-104, Springer, 2005.
[35] D. Koschützki and F. Schreiber, "Comparison of Centralities for Biological Networks," *Proc. German Conf. Bioinformatics (GCB '04),* pp. 199-2006, 2004.
[36] A.N. Langville and C.D. Meyer, "Deeper Inside Pagerank," *Internet Math.,* vol. 1, no. 3, pp. 335-380, 2003.
[37] R. Lempel and S. Moran, "Salsa: The Stochastic Approach for Link-Structure Analysis," *ACM Trans. Information Systems,* vol. 19, no. 2, pp. 131-160, 2001.
[38] M. Newman, "The Structure and Function of Complex Networks," *Soc, for Industrial and Applied Math. Rev.,* vol. 45, no. 2, pp. 167-256, 2003.
[39] M.E.J. Newman, "The Structure of Scientific Collaboration Networks," *Proc. Nat'l Academy of Sciences of USA,* vol. 98, no. 2, pp. 404-409, Jan. 2001.
[40] A.Y. Ng, A.X. Zheng, and M.I. Jordan, "Link Analysis, Eigenvectors and Stability," *Proc. Int'l Joint Conf. Artificial Intelligence (IJCAI),* pp. 903-910, 2001.
[41] A. Noack, "Energy Models for Graph Clustering," *J. Graph Algorithms Applications,* vol. 11, no. 2, pp. 453-480, 2007.
[42] A. Perer and B. Shneiderman, "Integrating Statistics and Visualization: Case Studies of Gaining Clarity during Exploratory Data Analysis," *Proc. SIGCHI Conf. Human Factors in Computing Systems,* pp. 265-274, 2008.

[43] J.P. Scott, *Social Network Analysis: A Handbook.* SAGE Publications, 2000.
[44] M. Shaw, "Group Structure and The Behavior of Individuals in Small Groups," *J. Psychology,* vol. 38, pp. 139-149, 1954.
[45] F. van Ham and A. Perer, "Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest," *IEEE Trans. Visualization and Computer Graphics,* vol. 15, no. 6, pp. 953-960, Nov./Dec. 2009.
[46] F. van Ham and M. Wattenberg, "Centrality Based Visualization of Small World Graphs," *Computer Graphics Forum,* vol. 27, no. 3, pp. 975-982, 2008.
[47] S. Wasserman, K. Faust, and M. Granovetter, *Social Network Analysis: Methods and Applications.* Cambridge Univ. Press, 1994.
[48] S. White and P. Smyth, "Algorithms for Estimating Relative Importance in Networks," *Int'l Conf. Knowledge Discovery and Data Mining,* pp. 266-275, 2003.

Carlos D. Correa received the BS degree in computer science from EAFIT University, Colombia, in 1998, and the MS and PhD degrees in electrical and computer engineering from Rutgers University in 2003 and 2007, respectively. Currently, he is a postdoctoral researcher at the Lawrence Livermore National Laboratory. Prior to joining LLNL, he was a postdoctoral researcher in the Department of Computer Science, University of California, Davis. His research interests include computer graphics, visualization and user interaction. He is a member of the IEEE.

Tarik Crnovrsanin received the BSc and MSc degrees in computer science from the University of California, Davis. His research interests include Social Network and Movement Data Visualization.

Kwan-Liu Ma received the PhD degree in computer science from the University of Utah in 1993. Currently, he is working as a professor of computer science and the chair of the Graduate Group in Computer Science (GGCS) at the University of California, Davis. He leads the VIDi research group and directs the DOE SciDAC Institute for Ultrascale Visualization, which involves researchers from three other universities and two DOE national laboratories. His research interests include visualization, high-performance computing, and user interface design. He was a paper chair of the IEEE Visualization Conference in 2008 and 2009, and he was the founder of the IEEE Pacific Visualization Symposium. He presently serves on the editorial boards of the *IEEE Computer Graphics and Applications*, the *IEEE Transactions on Visualization and Computer Graphics*, and the *Journal of Computational Science and Discoveries*. He is a senior member of the IEEE.